Title: Analysis of Radar Sounder Signals for the Automatic Detection and Characterization of Subsurface Features

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

Date of Publication: 2012

Author(s): Adamo Ferro, Lorenzo Bruzzone

Volume:50, Issue: 11

Page(s): 4333-4348.

DOI: 10.1109/TGRS.2012.2194500

Analysis of Radar Sounder Signals for the Automatic Detection and Characterization of Subsurface Features

Adamo Ferro, Lorenzo Bruzzone, *Fellow, IEEE*

***Abstract*—Radar sounders operating on satellite platforms (e.g., radar sounding missions at Mars) provide a huge amount of data that currently are mostly analyzed by means of manual investigations. This calls for the development of novel techniques for the automatic extraction of information from sounder signals that could greatly support the scientific community. Such a topic has not been addressed sufficiently in the literature. This paper provides a contribution to fill this gap by presenting both i) a study of the theoretical statistical properties of radar sounder signals, and ii) two novel techniques for the automatic analysis of sounder radargrams. The main goal of the study is the identification of statistical distributions that can accurately model the amplitude fluctuations of different subsurface targets. This is fundamental for the understanding of signal properties and for the definition of automatic data analysis techniques. The results of such a study drive the development of two novel techniques for i) the generation of subsurface feature maps, and ii) the automatic detection of the deepest scattering areas visible in the radargrams. The former produces for each radargram a map showing which areas have high probability to contain relevant subsurface features. The latter exploits a region-growing approach properly defined for the analysis of radargrams to identify and compose the basal scattering areas. Experimental results obtained on Shallow Radar (SHARAD) data acquired on Mars confirm the effectiveness of the proposed techniques.**

***Index Terms*—Radar sounding, ground penetrating radar (GPR), signal processing, statistical analysis, feature extraction.**

I. Introduction

PLANETARY radar sounders are orbiting ground penetrating radars (GPR) which operate at very low frequency (1-20 MHz) with a nadir looking geometry. Thanks to their characteristics, they are able to investigate the subsurface of planetary bodies by exploiting the radar signal propagation into the ground and measuring the backscattering from subsurface structures [1]. The output of a radar sounder is a radargram representing the vertical profile of the subsurface. Nowadays, two radar sounders are operating at Mars: the Mars Advanced Radar for Subsurface and Ionosphere Sounding (MARSIS) [2] on the Mars Express orbiter of the European Space Agency (ESA), and the Shallow Radar (SHARAD) [3] on-board the Mars Reconnaissance Orbiter of the US National Aeronautics and Space Administration (NASA). Both instruments were provided by the Italian Space Agency (ASI). On the one hand, MARSIS operates in a lower frequency range with respect to SHARAD and can penetrate the subsurface of Mars up to few kilometers with a vertical (range) free space resolution of 150 m. On the other hand, SHARAD has a maximum penetration of more than 1 km but with higher vertical resolution (15 m free space). These instruments are providing a new insight on the subsurface of Mars, both on its Polar Caps [4]–[7] and at mid-latitudes, where ice has been detected [8]. Indeed, radar sounders are particularly effective on glaciers and icy grounds because ice is the most transparent natural material in the range of frequencies in which they work. The success of MARSIS and SHARAD lead to the inclusion of a radar sounder in the study for the possible future missions to the exploration of the Jupiter system, where the icy moons Europa, Ganymede and Callisto are very important targets for this type of instrument [9], [10]. Radar sounding from space is also a possibility for the study of the Earth's subsurface and interest has been already shown by the scientific community [11].

The sounders currently operating at Mars are providing a huge amount of data. In general, the planetary scientific community which handles such data follows a manual investigation approach. Manual analysis of radargrams is a time-consuming task which leads to subjective interpretations of the data and limits their scientific return. This calls for the development of techniques for the automatic extraction of information from sounder data, which could greatly support the scientific community. On the one hand, the use of reliable techniques allows an objective and fast extraction of information from each radargram as soon as data become available. On the other hand, the exploitation of such techniques allows the joint analysis and the combination of many acquisitions, resulting in the possibility to analyze subsurface features at scales larger than a single radargram. This can highlight structures that are not visible from the measurements performed on single tracks. Automatic methods can also play a significant role in the integrated analysis of the radargrams with measurements obtained from other instruments. It is also worth noting that automatic methods developed for the analysis of orbiting radar sounders can be properly tuned for the analysis of sounding data acquired by airborne platforms on the Earth's subsurface. Finally, such methods can be also exploited for the processing of data acquired by possible future spaceborne sounding missions devoted to the observation of the Earth.

The automatic analysis of planetary radar sounder signals has not yet been addressed in the literature to a sufficient extent. The related works present in the literature regard

A. Ferro and L. Bruzzone are with the Department of Information Engineering and Computer Science, University of Trento, Via Sommarive 5, 38123, Trento, Italy, e-mail: adamo.ferro@disi.unitn.it, lorenzo.bruzzone@ing.unitn.it.

Manuscript received Month day, year; revised Month day, year.

the analysis of ground-based or airborne GPR signals (e.g., [12], [13]), which operate in different frequency ranges and achieve a better spatial resolution with respect to planetary radar sounders. Moreover, GPR campaigns are often reduced to well defined areas with limited extension, for which the interpretation of the radargrams can be performed manually, without the need of automatic techniques. An exception is represented by anti-mines and unexploded ordnance (UXO) detection campaigns, which make extensive use of GPR technology [14]. Different papers in the past decade proposed the use of pattern recognition approaches to the analysis of GPR signals (e.g., [15]). However, they are mainly devoted to the detection of specific buried objects, such as mines, pipes or tanks buried at small depths using ground-based GPR. Such objects present hyperbola-like signatures in the radargrams, which are completely different from the signatures of buried structures present in radar sounder images acquired by orbiting platforms. The radargrams obtained by airborne acquisitions over the Earth's polar areas show similarities with spaceborne radar sounder data acquired on icy bodies. The main features present in such images are subsurface echoes coming from the interfaces present between different subsurface ice layers and basal returns [16]. This is the typical situation shown in the radargrams related to the Mars' Poles [5], [6] and other areas of the Red Planet [17].

Another approach to the analysis of radar sounder measurements is to apply inversion techniques to the signals in order to estimate the dielectric characteristics of the subsurface [18], [19]. In this context, the correct understanding of the radargrams and the development of any information extraction technique need the knowledge of the propagation laws of the radar signal into the matter in order to avoid errors in the physical interpretation of the returns [20]. However, the inversion process is very complex and requires proper assumptions on the investigated domain, e.g., on the ground composition [21].

This paper provides a contribution to fill the gap present in the literature on the automatic analysis of planetary radar sounder data by presenting a study of the theoretical statistical properties of radar sounder signals. The goal of this study is the identification of a statistical distribution which can accurately model the amplitude fluctuations of different subsurface targets. On the basis of the results of this study, we then propose two novel techniques for i) the generation of subsurface feature maps, and ii) the automatic detection of the deepest scattering area visible in the radargrams. The former produces for each radargram a map showing which areas have high probability to contain relevant subsurface features. Such a map can be used to identify interesting radargrams in large datasets or to drive further signal processing steps on specific areas within single radargrams. The latter is based on an iterative procedure that exploits a region-growing method properly defined for the analysis of radargrams to identify and compose the basal scattering areas. The obtained regions are kept or discarded according to the statistical distribution of their samples. We tested both techniques on SHARAD data acquired on the North Polar Layered Deposits (NPLD) of Mars. Although in this paper we will focus on the analysis of the signals provided by the SHARAD instrument, the results obtained can be applied also to MARSIS data or to signals acquired by other radar sounder instruments after a proper tuning of the techniques.

The remaining of the paper is organized as follows. In Sec. II we address the problem of the statistical modeling of radar sounder signals. The models presented are then tested on real SHARAD data in Sec. III. Section IV presents an automatic technique for the generation of subsurface feature maps. Section V addresses the automatic detection of basal returns and its application to SHARAD radargrams of the NPLD of Mars. Finally, Sec. VI draws the conclusion of this paper and discusses possible future developments.

II. Statistical Modeling of Radar Sounder Signals

In order to develop effective information extraction techniques from radar sounder data, a precise knowledge of the statistics of the analyzed signals is necessary. In this section we review the main characteristics of the sounder signals and select three statistical models which are likely to be appropriate to model the signal fluctuations. The validity of such models will be tested on real SHARAD data in Sec. III.

A. Background and Motivation

The analysis of radar signals is historically linked to statistics. This is due to the coherent nature of the radar signals which makes the radar cross section (RCS) of targets fluctuate when even slightly changes in the viewing configuration or in the target orientation occur [22]. The effects of clutter and noise also greatly contribute to the fluctuations of the RCS. Radar signals are thus modeled using probability density functions (pdf) under the assumption that the signal amplitude (or intensity) is the realization of a random variable within each radar resolution cell. Many statistical models have been developed in order to fit the radar signals related to different target types. Such statistical models are based on theoretical descriptions of the scattering effects, or on empirical fitting to sample data. Examples of theoretical pdf commonly used in the analysis of radar signals are the Rayleigh, Rice, negative exponential, Gamma and K distributions. The most important empirical pdf are the Weibull and log-normal distributions [22].

The statistical approach has been extensively used in the analysis of synthetic aperture radar (SAR) images for the characterization of distributed targets such as agriculture fields, forests or water surfaces. For this type of targets a single resolution cell does not provide sufficient information about the scattering characteristics of the surface under investigation due to the signal fluctuations, which depend on intrinsic fluctuations of the target RCS and on the so-called *speckle*. In order to characterize the analyzed surface it is thus necessary to calculate statistical parameters of the distribution of the radar signals coming from the area of interest.

In this context, statistical tools can be also exploited for the analysis of radar sounder signals for the detection and characterization of different types of subsurface features. This can support the analysis of the radargrams, by automatically detecting the regions of interest and extracting information which can drive subsequent feature extraction algorithms. The goal of this section is thus to define a reference theoretical framework which can be used for a reliable statistical analysis of the signals, taking into account the physical characteristics of the targets.

B. Statistical Models

In order to perform an analysis of radar sounder signals, it is necessary to describe the signal statistical properties taking into account the physical processes involved in the scattering from subsurface features for a typical radar sounder instrument mounted onboard of an aerial or satellite platform. Our goal is to describe statistically the distribution of the signals coming from the subsurface by considering groups of adjacent samples in a predefined neighborhood system extended both in range (vertical) and along-track (azimuth) directions. Indeed, each radargram can be seen as a 2D image defined in the range and azimuth directions. The signals measured by the radar during each acquisition window (frames) correspond to the columns of the 2D image. Thus, pixels in the same neighborhood system describe the geologic features in a given position of the subsurface. According to this modeling, we can analyze radargrams with a 2D signal processing approach; this is important given that most of the subsurface features detected by a radar sounder are not spot features but show a certain extension, especially in the azimuth direction.

As a reference, Tab. I reports the main characteristics of the two radar sounders currently operating at Mars: MARSIS [2] and SHARAD [3]. f_c is the central frequency of the radar, λ depicts the wavelength (which is reported for the dielectric constant of the vacuum, $\varepsilon_r = 1$, and for an icy subsurface material, $\varepsilon_r = 3.15$), B_W represents the radar bandwidth, δ_v is the vertical (range) resolution in the subsurface, and δ_{al} and δ_{ac} are the along- and across-track resolutions, respectively. D_F is the theoretical Fresnel zone calculated at the surface. Fig. 1 shows the typical acquisition geometry of orbiting radar sounders.

The sizes of the radar footprints reported in Tab. I are comparable with the diameter of the corresponding Fresnel zone, from which the returns are supposed to be coherent. However, the surface and especially the subsurface, which is the target of our investigation, are far from being flat and always present a certain amount of roughness, which introduces a significant non-coherent component in the scattering [23]. Indeed, the amount of roughness drives the across-track resolution, which, for MARSIS and SHARAD, is controlled only by their dipole antenna pattern as no synthetic aperture processing is possible in the across-track direction. It is thus possible to consider the radar footprints sufficiently wide to assume that many different independent scatterers contribute to the scattering for each resolution cell.

In the following, we will focus on the statistical distribution of amplitude signals. The analysis of amplitude data is preferred here with respect to intensity data due to the large dynamic that characterize radar sounder acquisitions, which

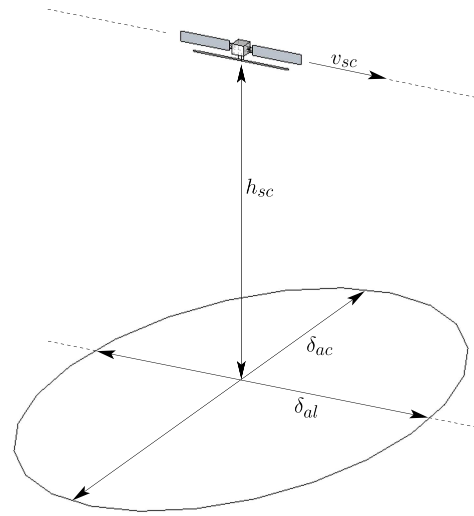


Fig. 1. Typical acquisition geometry of an orbiting radar sounder. v_{sc} is the pacecraft velocity, h_{sc} represents the orbit altitude, and δ_{al} and δ_{ac} are the ground resolution in the along-track and across-track directions, respectively.

is much more amplified in intensity data and may affect the stability of the analysis.

1) Rayleigh pdf: The simplest pdf that describes the amplitude x of the returns from a large number Q of independent scatterers is the Rayleigh distribution:

$$p_R(x) = \frac{2x}{\mu_z} \exp\left[-\frac{x^2}{\mu_z}\right], \tag{1}$$

where z indicates the signal power ($z = x^2$), and μ_z is the only parameter of the distribution and represents the mean power of the signal [22]. Eq. (1) is valid for $x \geq 0$ (this also holds for the other pdfs which will be presented in the following) and the mean value of x is given by $\mu_x = \sqrt{\pi\mu_z}/2$. The corresponding distribution in the power (intensity) domain is the negative exponential distribution. It is worth noting that the Rayleigh distribution is also the ideal theoretical model for the amplitude when a zero-mean additive white Gaussian noise (AWGN) affects the in-phase and quadrature signals received by the radar in areas of no subsurface scattering.

2) Nakagami pdf: The second model that we consider is the Nakagami pdf, which is a two-parameter function given by [24]:

$$p_N(x) = 2\left(\frac{v_N}{\mu_z}\right)^{v_N} \frac{x^{2v_N-1}}{\Gamma(v_N)} \exp\left[-\frac{v_N x^2}{\mu_z}\right], \tag{2}$$

where v_N is called *shape* or *order parameter* and $\Gamma(.)$ depicts the gamma function. The validity range of v_N is $(0; +\infty)$. The Nakagami pdf for amplitude data corresponds in the intensity domain to the Gamma pdf described by the shape parameter $v_\Gamma = v_N$ and the mean intensity μ_z [24]. The Gamma pdf has been widely used for the modeling of radar signals and is a generalization of other well-known distributions, such as the negative exponential and chi-square [25]. In particular, when v_Γ is an integer value, the Gamma pdf can be derived as the sum of v_Γ identical independent exponentially distributed random variables. Similarly, in the amplitude domain the

TABLE I
MAIN CHARACTERISTICS OF THE MARSIS AND SHARAD RADAR SOUNDERS OPERATING AT MARS.

Instrument	f_c	λ ($\varepsilon_r = 1$)	λ ($\varepsilon_r = 3.15$)	B_W	δ_v ($\varepsilon_r = 3.15$)	δ_{al}	δ_{ac}	D_F
MARSIS	1.8–5 MHz	167–60 m	94–34 m	1 MHz	85 m	5–10 km	10–30 km	∼10 km
SHARAD	20 MHz	15 m	8.5 m	10 MHz	8.5 m	0.3–1 km	3–7 km	∼3 km

Nakagami pdf is a generalization of the Rayleigh pdf, which can be obtained by setting $v_N = 1$ in (2).

3) K pdf: The last distribution that we consider is the K distribution, defined as [22]:

$$p_K(x) = \frac{4}{\Gamma(v_K)} \left(\frac{v_K}{\mu_z}\right)^{(v_K+1)/2} x^{v_K} K_{v_K-1}\left[2x\sqrt{\frac{v_K}{\mu_z}}\right], \tag{3}$$

where $K_{v_K-1}(.)$ is the modified Bessel function of the second kind of order $v_K - 1$. The parameter v_K is also called shape (or order parameter), and its validity range is $(0; +\infty)$. The K distribution has also been used for modeling sea clutter and distributed targets of different types in SAR images. It is derived by assuming that the number of scatterers within a resolution cell Q fluctuates being controlled by a birth-death-immigration process, i.e., Q is a random variable that follows a negative binomial distribution [22]. The assumption that the number of scatterers varies between different resolution cells is in agreement with the scenario represented by a radar sounder acquisition, where within each single radargram frame a different number of scatterers (e.g., subsurface interfaces) may contribute to the scattering measured in different time samples.

The K distribution is also obtained by modeling the radar intensity z as a compound pdf, also referred to as *product model*. This formulation expresses the radar intensity as the product of two uncorrelated processes with different spatial scales: an underlying RCS and a multiplicative speckle contribution. The mathematical representation of this formulation is:

$$p_K(z) = \int_0^\infty p_1(z/s)p_2(s)ds, \tag{4}$$

where $p_2(s)$ represents the pdf of the underlying RCS (which only depends on the physical characteristics of the scatterers) and $p_1(z/s)$ is the speckle contribution, which arises as a consequence of their random distribution and orientation. By assuming an underlying RCS which is Gamma distributed and a speckle contribution modeled by a negative exponential pdf, both the signal intensity and amplitude result K distributed [22].

The product model is thus suited to the modeling of spatially non-homogeneous targets. As proposed in [26] and [27], $p_1(z/s)$ can be interpreted as the density of the returns from an incremental area of a surface whose reflectivity varies spatially with mean s, while $p_2(s)$ describes the bunching of scatterers in terms of spatial variations of the underlying RCS, which are on a much larger scale than the variations described by $p_1(z/s)$. Such a formulation has been effectively used to model sea clutter, where scatterers are bunched by swell structure [27]. This situation to a certain extent resembles the measurements performed by a radar sounder in presence of subsurface layer stratigraphy, where the returns are bunched at each interface. The K distribution has thus physical basis which are in agreement with the characteristics of radar sounder acquisitions.

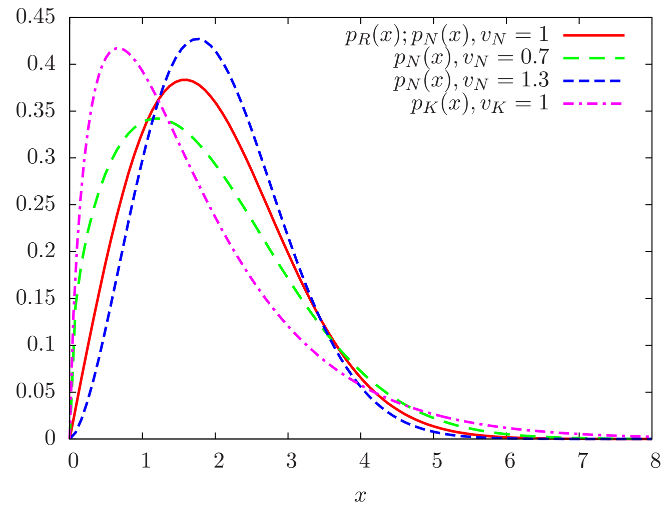


Fig. 2. Examples of pdf curves obtained using the models presented in Sec. II. For all the curves $\mu_z = 5$.

Fig. 2 shows a comparison between the Rayleigh, Nakagami and K distributions for a fixed μ_z and varying shape parameters.

Other pdfs can be used to model radar data, e.g., Rice, log-normal, Weibull [22]. In particular, for the analysis of radar sounder signals, the Rice distribution is suited to the modeling of surface returns from flat surfaces, allowing the estimation of the coherent scattering for inversion purposes [28]. However, the pdfs selected for the analysis reported in this paper cover the most important classes of theoretical distributions which are used for the modeling of radar data, and have the advantage to allow us to describe the scattering from subsurface features with a physical-based approach. As such, they represent generalizations or approximations of many other distributions proposed in the literature. It is worth noting that the research of the absolute best fitting pdf for radar sounder signals is out of the scope of this paper.

III. EMPIRICAL ANALYSIS OF THE STATISTICAL MODELS ON SHARAD RADARGRAMS

With the goal of studying the statistical distribution of real data, we analyzed different subsurface target types and studied the statistical distributions of their returns by fitting the theoretical pdfs described in Sec. II to the data. We selected as test data a set of SHARAD radargrams of the NPLD of Mars. Such radargrams show different target types, from very strong scattering linear interfaces (due to ice stratigraphy) to smooth returns from the base of the NPLD. An example of SHARAD radargram of the NPLD of Mars and its ground track are reported in Fig. 3.

A. Definition of Target Classes and Dataset Description

The target classes that we investigated are the following: *no target* (NT), *strong layers* (SL), *weak layers* (WL), *low returns* (LR), *basal returns* (BR). The class *no target* corresponds to areas of the radargram where no scattering is visible. These are the shallow part of the radargram, before any surface return, and the areas in the subsurface where no interfaces are detected. We define *strong layers* the areas of the radargram where dense and strong scattering layering is visible. This corresponds generally to areas in the shallow subsurface of the NPLD. The class *weak layers* corresponds to the subsurface scattering related to less dense and less strong scattering layering, which usually occurs below the areas described by the class *strong layers*. The class *low returns* includes the areas of the radargram containing very weak scattering coming from deep structures. When these are present, they are usually located between the areas of *weak layers* and *basal returns*. Finally, the class *basal returns* is related to the scattering coming from the base of the NPLD, which nature gives a diffuse scattering especially in correspondence of the so-called basal unit [7]. Fig. 4 highlights such classes on the test radargram of Fig. 3.

The analysis has been carried out on 7 SHARAD radargrams of the NPLD of Mars. The main characteristics of the SHARAD instrument are summarized in Tab. I. The radargrams were stored in the Reduced Data Record (RDR) format [30], and have been downloaded from the the Geosciences Node of NASA's Planetary Data System (PDS) [31]. We extracted the amplitude information and aligned in time the echoes using the information contained in the RDRs. As the data are highly oversampled in the along-track direction due to the high pulse repetition frequency (PRF) of the system, we applied a downsampling factor of 15 in the selection of the radargram frames. Each frame thus corresponds to an along-track step which can vary approximately between 270 and 500 m depending on the amount of presumming performed onboard the instrument. No multilooking has been performed in order to maintain the original statistics of the signals. In the range direction each frame is sampled every 75 ns. Therefore, each sample corresponds to a free-space distance of about 11.3 m, which scales to approximately 6.3 m in ice ($\varepsilon_r = 3.15$). The acquisitions have been cut in order to consider only the NPLD area. The resulting radargrams are made of a number of samples between 1,071,869 and 2,582,624. On each radargram we selected manually the areas corresponding to the classes defined in the previous subsection. In Tab. II we report for each analyzed acquisition its identification number and the number of samples per class we collected. It is worth noting that a very high number of samples for each class in each radargram is considered in order to have a reliable statistical analysis.

B. Procedure for the Estimation of pdf Parameters

For each class type we estimated the parameters of the Rayleigh, Nakagami and K distributions using a Maximum Likelihood (ML) estimation approach. For the Rayleigh distribution the ML estimate $\tilde{\mu}_z$ of the only parameter μ_z is given by the sample mean power [25]:

$$\tilde{\mu}_z = \frac{1}{n}\sum_{i=1}^{n} x_i^2, \tag{5}$$

where x_i depicts an amplitude sample, and n is the number of considered samples.

For the Nakagami distribution, the estimate $\tilde{\mu}_z$ is obtained as for the Rayleigh distribution, and is given by (5). The calculation of $\tilde{v}_N$ has been performed using the classical estimator proposed by Greenwood and Durand [32], which is considered in the literature an accurate estimator for the shape parameter of the Nakagami distribution [33]. Therefore, $\tilde{v}_N$ has been derived by:

$$\tilde{v}_N = \begin{cases} (0.5000876 + 0.1648852y - 0.0544274y^2)/y, \\ \qquad 0 < y \leq 0.5772 \\ \dfrac{8.98919 + 9.059950y + 0.9775373y^2}{y(17.79728 + 11.968477y + y^2)}, \\ \qquad 0.5772 < y < 17, \end{cases} \tag{6}$$

where

$$y = \ln\left(\frac{\tilde{\mu}_z}{F}\right) \tag{7}$$

and

$$F = \left(\prod_{i=1}^{n} x_i^2\right)^{\frac{1}{n}}. \tag{8}$$

The ML estimation of the K distribution has been obtained retrieving the $\tilde{v}_K$ and $\tilde{\mu}_z$ estimated values by the numerical maximization of the log-likelihood function, according to [34], i.e.,

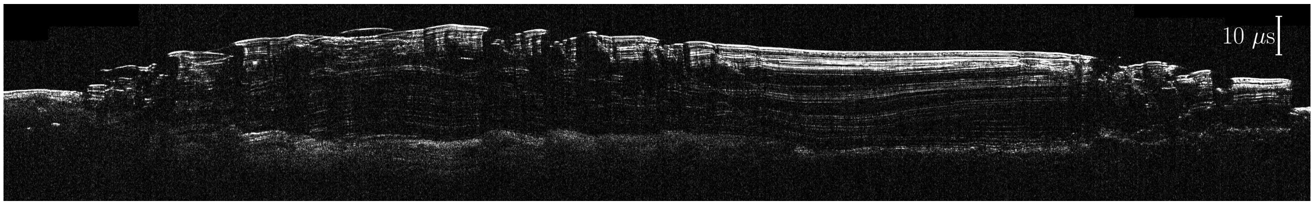
$$(\tilde{v}_K, \tilde{\mu}_z) = \arg\max_{(v_K,\mu_z)} \left\{\ln\left[l_n(v_K, \mu_z; x_1, x_2, \ldots, x_n)\right]\right\}, \tag{9}$$

where

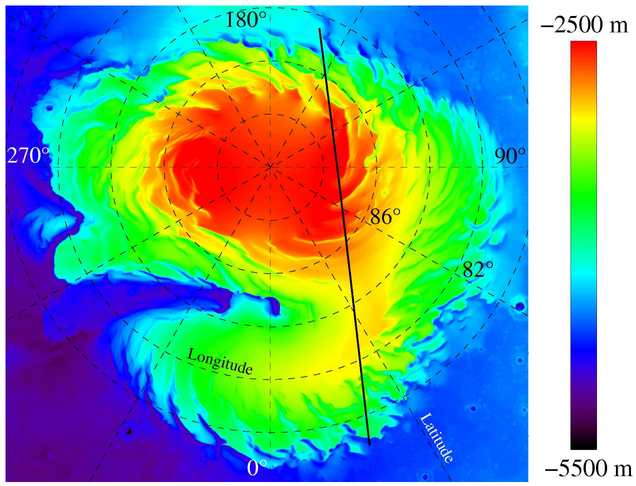
$$\begin{aligned} \ln\left[l_n(v_K, \mu_z; x_1, x_2, \ldots, x_n)\right] &= v_K \sum_{i=1}^{n} \ln x_i \\ &+ \sum_{i=1}^{n} \ln\left\{K_{v_K-1}\left[2x_i\sqrt{\frac{v_K}{\mu_z}}\right]\right\} \\ &+ n\left\{\frac{v_K+1}{2}\ln\left(\frac{v_K}{\mu_z}\right) + \ln 4 - \ln\Gamma(v_K)\right\}, \end{aligned} \tag{10}$$

TABLE II
SHARAD RADARGRAMS USED IN THE ANALYSIS AND NUMBER OF SAMPLES PER TARGET CLASS COLLECTED FOR EACH RADARGRAM.

Radargram number	NT	SL	WL	LR	BR
0371502	212,311	9,443	18,233	18,017	50,905
0385902	166,832	4,425	6,284	13,289	21,417
0681402	209,416	41,459	22,264	44,829	130,687
0794703	209,057	14,586	27,004	46,207	71,387
1292401	113,768	4,701	11,173	12,049	37,082
1312901	148,651	9,173	17,596	51,218	26,684
1319502	195,748	14,688	18,952	33,448	72,582



(a)



(b)

Fig. 3. (a) Portion of the SHARAD radargram 1319502, and (b) its acquisition track highlighted on an altimetric map of the NPLD of Mars. The altimetric map has been derived from Mars Orbiter Laser Altimeter (MOLA) [29] data. The radargram corresponds to the solid line.

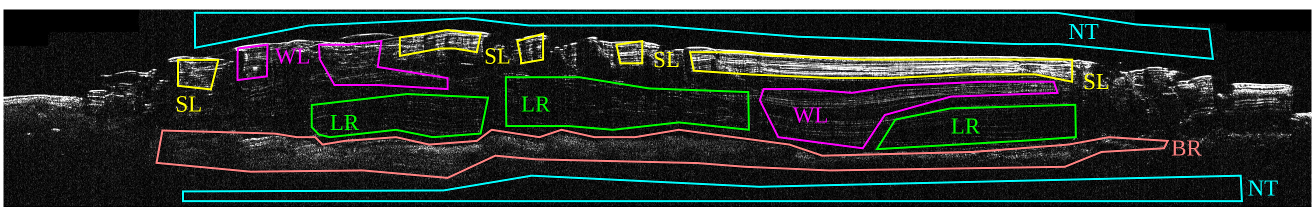


Fig. 4. Target classes used in the statistical analysis presented in this paper highlighted on the radargram showed in Fig. 3.

and $l_n(v_K, \mu_z; x_1, x_2, \ldots, x_n)$ is the likelihood function for the K distribution. Due to numerical constraints, the range of values of v_K has been limited between 0.1 and 50. However, this does not affect the generality of our analysis. Indeed, on the one hand, the characteristics of the signals never require values of v_K lower than 0.1. On the other hand, for $v_K \geq 50$ the K distribution becomes nearly Rayleigh [34]. Therefore, the use of values greater than 50 for v_K is not significant for the comparison between the fitting performance of the two pdfs. For the parameter μ_z we only imposed a lower limit at 0.1, which is well below the typical noise mean power of SHARAD data.

C. Results

Tab. III reports the fitting accuracies obtained for the different classes of targets for each analyzed radargram. Such accuracies have been evaluated in terms of root mean square error (RMSE) and Kullback-Leibler divergence (KL) between the normalized histogram of the data and the histogram obtained by the fitting of each distribution. The KL divergence is defined as [35]:

$$\mathrm{KL}(A, B) = \sum_{x_i} A(x_i) \log \frac{A(x_i)}{B(x_i)}, \tag{11}$$

where A and B represent the probability distribution of the samples and of the theoretical fit, respectively. The values of x_i depend on the size of the bins used for the computation of the histograms. This size has been calculated for each target class according to the method proposed in [36], which is suited for unknown distribution data values, and has already been used for the computation of histograms of SAR images [37]. As an example, Fig. 6 shows the histogram and the ML estimates for each target class for the test radargram of Fig. 3.

The results point out that the best fitting distribution is in almost all the cases the K distribution. Such results agree with the physical basis of the K distribution, which can describe effectively the cases where the scatterers are bunched (see Sec. II). Fig. 5 shows graphically the mean and the standard deviation of the parameters derived for the K distribution for each target class. It is possible to note that within each class the parameters of the distributions are quite stable.

Moreover, it is also worth noting that different targets are described by different parameters. The K distribution shows lower fitting performances for the *no target* case. This is due to the numerical limit imposed to $\tilde{v}_K$ (which leads to $\tilde{v}_K = 50$ for the *no target* class for all the test radargrams). However, as previously mentioned, the higher $\tilde{v}_K$ the more the distribution approximates the Rayleigh pdf. The Nakagami distribution provides almost always a more accurate fit than the Rayleigh distribution except for the case of the *no target* class. For the *no target* case, as expected from the theory, the Rayleigh distribution is an effective estimate as it provides accurate estimations using only one parameter. The Nakagami distribution has approximately the same fitting performance using two parameters, but $\tilde{v}_N$ is always nearly 1, i.e., it approximates the Rayleigh pdf. The Rayleigh pdf can thus be considered the best fitting distribution for the *no target* areas. This confirms that the background noise of the SHARAD data can be modeled as a zero mean AWGN in both the in-phase and quadrature components.

Let us now focus on the computational complexity of the ML estimation for the three considered distributions. Such issue becomes relevant when the statistical analysis of the signals is propaedeutic to other processing steps, e.g., filtering or feature-extraction algorithms. The calculations of the ML estimates for the Rayleigh and Nakagami pdfs are performed analytically and their computational time is negligible on a standard workstation. Instead, the maximization of (10) for the estimation of the parameters of the K distribution must be performed numerically. Although the computational time in our tests is still in the order of less than one minute, it may become not negligible when analyzing a large series of radargrams. When the computational time becomes a limit in practical analysis scenarios, one may consider to use the Nakagami distribution for the modeling of the signal statistics in order to speed up the processing, at the cost of slightly lower accuracies.

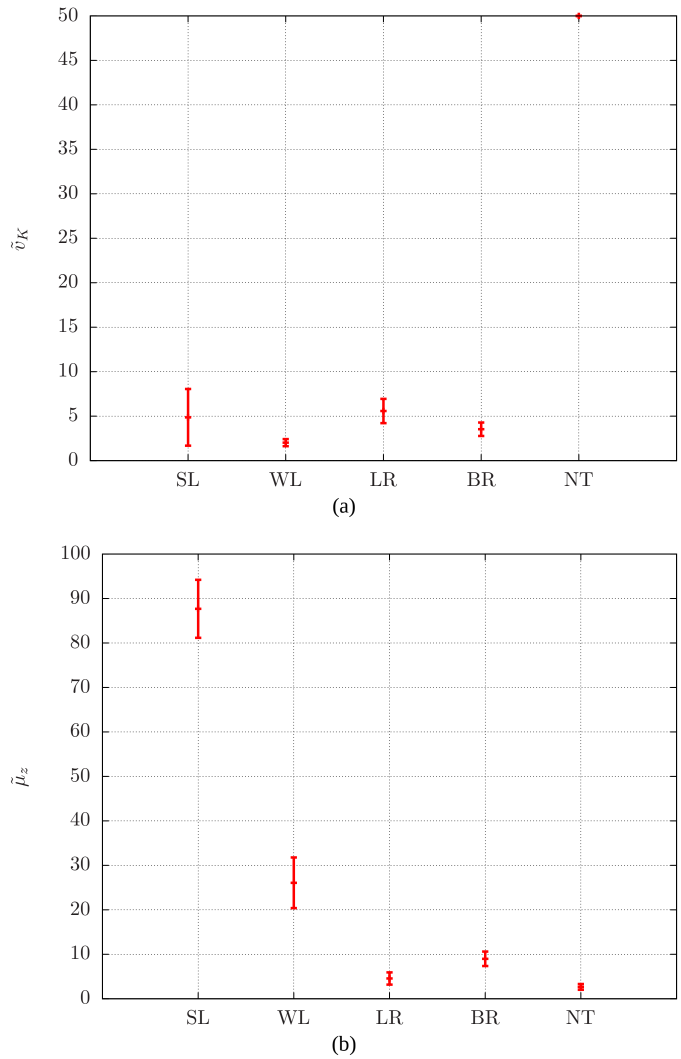


Fig. 5. Mean values and range of variation of the parameters of the fitted K distributions for each target class: (a) $\tilde{v}_K$; (b) $\tilde{\mu}_z$.

IV. Proposed Technique for the Generation of Subsurface Feature Maps

The results presented in the previous section can be used to study the radar sounder signals and analyze the scattering signatures of different types of targets. However, they also open to a wide range of applications for the automatic analysis of the radargrams. As mentioned in the introduction, planetary radar sounding missions have provided and are still providing a large amount of data, which have been studied mostly by means of manual investigations. In this framework, the automatic detection of radargrams containing subsurface features from the whole available set of radargrams, and the automatic identification of the subsurface areas containing relevant features within each radargram become important tasks that can greatly support scientific investigations. In this section we propose a novel automatic method for the generation of maps of the subsurface areas containing relevant features within a radargram by analyzing the statistical distributions of local parcels of the radargram.

A. Proposed Technique

As discussed in Sec. III, the background noise of SHARAD radargrams is Rayleigh distributed. The noise characteristics can be simply measured using the samples belonging to the free space region of the radargram, i.e., before any surface echo. Therefore, the statistical distribution of the noise can be determined precisely and in an automatic way. By measuring the statistical difference between the histograms of subsurface parcels and the noise distribution it is thus possible to discriminate in an unsupervised way the areas containing only noise from the regions which contain subsurface features. Several statistical indicators can be used to measure the difference between two distributions. Here, we propose the use of the KL divergence between the histogram of the samples H and the theoretical noise distribution N, i.e., $\mathrm{KL}_{HN} = \mathrm{KL}(H, N)$. The noise characteristics can vary between different acquisitions (see Fig. 5). This is mainly due to different conditions of acquisition, e.g., in terms of solar activity or spacecraft attitude, which may raise the background noise level. The proposed algorithm takes into account this issue and adapts its behavior to the variations of the background noise level by

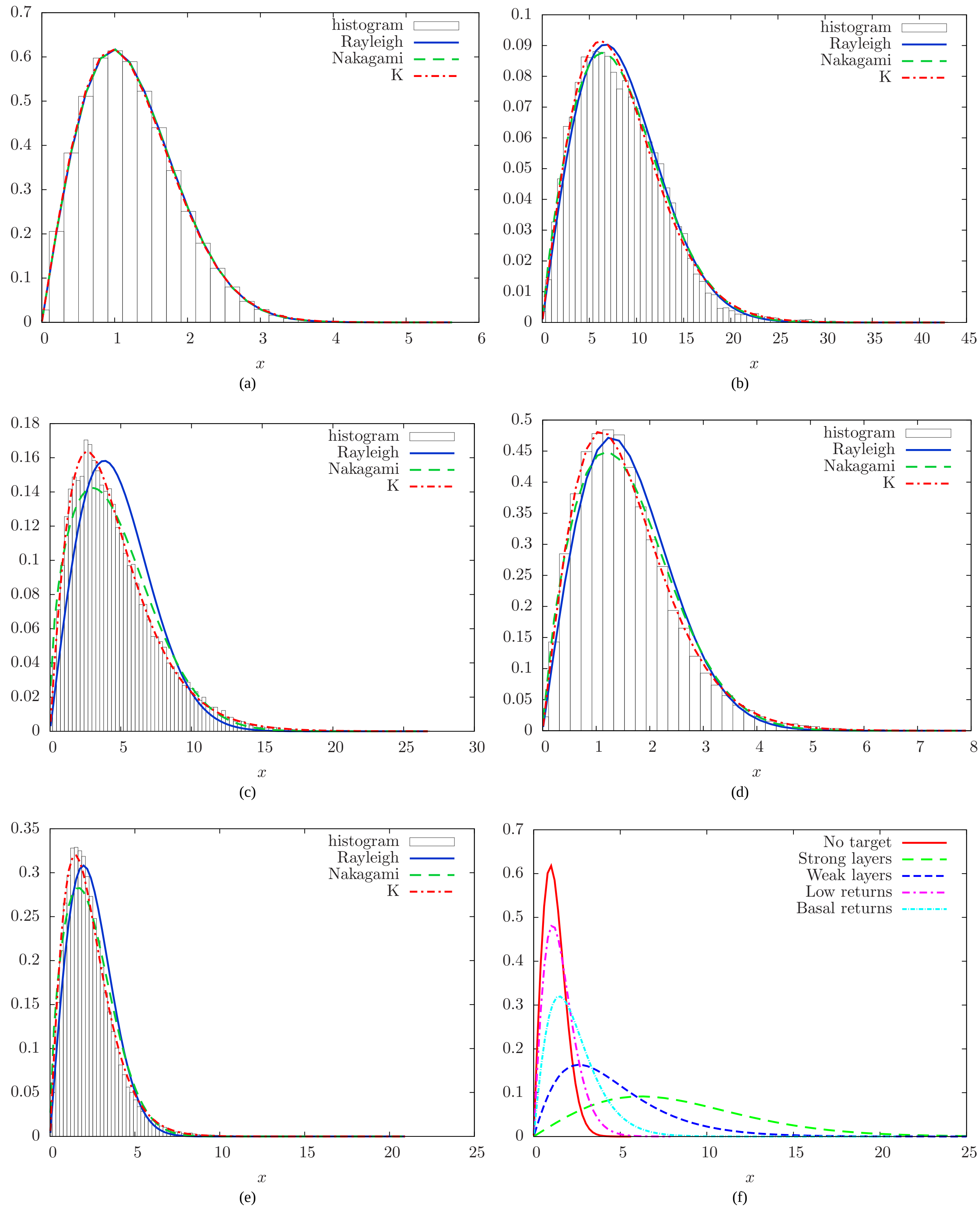


Fig. 6. Empirical and ML distributions for each target class for the SHARAD radargram 1319502 (see Fig. 3): (a) *no target*, (b) *strong layers*, (c) *weak layers*, (d) *low returns*, (e) *basal returns*, (f) summary of the fitted K distributions for each target class.

TABLE III
FITTING PERFORMANCES OF THE RAYLEIGH, NAKAGAMI AND K DISTRIBUTIONS TO THE SAMPLE AMPLITUDE DATA FOR EACH SCATTERING CLASS. THE BEST RESULTS ARE HIGHLIGHTED IN BOLD.

Radargram number	Distribution	NT		SL		WL		LR		BR	
		RMSE	KL	RMSE	KL	RMSE	KL	RMSE	KL	RMSE	KL
0371502	Rayleigh	0.0031	0.0067	0.0074	0.0381	0.0133	0.0516	0.0125	0.0108	0.0106	0.0243
	Nakagami	**0.0031**	**0.0067**	0.0032	0.0108	0.0075	0.0186	0.0085	0.0043	0.0079	0.0146
	K	0.0041	0.0068	**0.0028**	**0.0060**	**0.0018**	**0.0021**	**0.0046**	**0.0028**	**0.0024**	**0.0033**
0385902	Rayleigh	0.0032	**0.0029**	0.0118	0.1035	0.0147	0.0475	0.0161	0.0293	0.0108	0.0313
	Nakagami	**0.0031**	0.0030	0.0068	0.0418	0.0103	0.0249	0.0121	0.0153	0.0092	0.0214
	K	0.0047	0.0031	**0.0026**	**0.0067**	**0.0046**	**0.0056**	**0.0059**	**0.0042**	**0.0045**	**0.0058**
0681402	Rayleigh	0.0034	**0.0045**	0.0085	0.0707	0.0222	0.1258	0.0177	0.0247	0.0193	0.0675
	Nakagami	**0.0034**	0.0045	0.0054	0.0285	0.0141	0.0503	0.0139	0.0136	0.0149	0.0362
	K	0.0048	0.0046	**0.0014**	**0.0031**	**0.0044**	**0.0054**	**0.0054**	**0.0033**	**0.0060**	**0.0064**
0794703	Rayleigh	0.0041	0.0062	0.0027	0.0089	0.0188	0.0732	0.0122	0.0131	0.0155	0.0462
	Nakagami	**0.0040**	**0.0060**	0.0021	0.0052	0.0120	0.0293	0.0090	0.0068	0.0126	0.0283
	K	0.0052	0.0062	**0.0014**	**0.0033**	**0.0039**	**0.0028**	**0.0031**	**0.0036**	**0.0052**	**0.0048**
1292401	Rayleigh	0.0046	**0.0041**	0.0052	0.0288	0.0213	0.1016	0.0152	0.0108	0.0157	0.0343
	Nakagami	**0.0045**	0.0043	0.0043	0.0225	0.0140	0.0456	0.0116	0.0060	0.0124	0.0190
	K	0.0062	0.0042	**0.0034**	**0.0110**	**0.0051**	**0.0074**	**0.0087**	**0.0025**	**0.0053**	**0.0058**
1312901	Rayleigh	0.0058	0.0048	0.0039	0.0623	0.0253	0.1093	0.0174	0.0272	0.0178	0.0357
	Nakagami	**0.0058**	**0.0047**	0.0043	0.0500	0.0164	0.0452	0.0149	0.0157	0.0125	0.0189
	K	0.0068	0.0048	**0.0035**	**0.0252**	**0.0057**	**0.0061**	**0.0072**	**0.0065**	**0.0038**	**0.0026**
1319502	Rayleigh	**0.0053**	0.0091	0.0029	0.0135	0.0157	0.0540	0.0210	0.0202	0.0178	0.0585
	Nakagami	0.0053	**0.0089**	**0.0022**	0.0105	0.0079	0.0151	0.0166	0.0109	0.0140	0.0346
	K	0.0065	0.0091	0.0025	**0.0082**	**0.0027**	**0.0029**	**0.0073**	**0.0035**	**0.0056**	**0.0070**

automatically detecting and measuring the statistical characteristics of the free space region for each radargram.

A block scheme of the proposed technique is shown in Fig. 7. The main steps of the technique are explained in the following using the SHARAD radargram of Fig. 3 as a reference example.

1) *First return detection*: this step aims at automatically identifying the returns from the surface for then discriminating in the radargram the parts belonging to the free space and those associated with the subsurface. The former is used to estimate the radargram background noise signal distribution in the next step. For each frame (column) j of the radargram the algorithm detects the position of the first sample which is statistically different from the frame background noise. We denote such a position as $f(j)$ and calculate it as follows:

$$f(j) = \min\{i : x(i,j) > \mu_N + \gamma_1\sigma_N\} \qquad \forall j \qquad (12)$$

where $x(i,j)$ is the amplitude of the sample of the frame j at the time step i; $i \in [1, I]$; $j \in [1, J]$; $I = 667$ is the number of samples of a SHARAD frame; J is the number of frames of the radargram; μ_N and σ_N are the estimated frame noise mean amplitude and standard deviation, respectively; γ_1 is a multiplicative factor. The detected samples are in the ideal case representative of the nadir surface return. This is not true when lateral clutter echoes arrive to the receiver before the nadir return. The local statistics of the noise is estimated for each frame using its last 50 samples, which are in general free from subsurface features as the signal loss is very high at the corresponding depth. If no sample fulfills the condition, the value of γ_1 is decreased and the procedure is repeated. At each iteration e the value of γ_e is calculated using a positive damping factor $d < 1$, according to:

$$\gamma_e = d \cdot \gamma_{e-1} \qquad \forall e = 2, \ldots, E \qquad (13)$$

where E is the maximum number of iterations. E, the initial value γ_1 and the damping factor d are specified by the user. Note that from (12) the minimum signal level necessary to perform a detection cannot be lower than the frame noise mean μ_N. In the case that after E trials no sample fulfills the condition yet, the first return position of the considered frame is estimated using the average position of the first adjacent frames for which the detection was successfully. After the frame-based detection, a smoothing function is applied in order to reduce the effects of both outliers and missing detections. The smoothing function performs local regression using weighted linear least squares and a first degree polynomial model. Using this approach, for each frame the algorithm detects the most reliable first return at the first iteration (according to a user-defined minimum signal level dependent on γ_1). The reliability of the detection decreases as the number of iterations increases. By properly setting E and d the user can thus tune the reliability of the first return detection. The result of the first return detection applied to the test radargram using $E = 3$, $\gamma_1 = 4.5$, and $d = 0.9$ is shown in Fig. 8a. The first return line is detected with good accuracy for most of the frames composing the considered radargram. The only exception corresponds to a part of the radargram where no returns are visible until a certain depth for a relatively long series of frames. As the number of frames where no returns are visible is large, the smoothing procedure cannot recover the missing part of the first return line. In order to correct for this error, it would be necessary to use very large smoothing windows. However, this would compromise the accuracy of the detection of the first return on the whole radargram, as the detected line would result too much smoothed. For this reason, in this paper we preferred to give higher priority to achieve a good detection of the first return line than to correct for large missing parts.

2) *Estimation of the noise statistics*: in this step the algorithm uses all the samples of the radargram belonging to the free space region R_{fs} to estimate the parameter μ_z of a Rayleigh distribution, according to the ML approach (see Sec. III-B). R_{fs} is defined as the upper part of the radargram delimited by the line representing the first returns identified in the previous step, i.e.,

$$R_{fs} = \{(i,j) : 0 < i < f(j) - w_G\}, \qquad (14)$$

where w_G is a positive constant used in order to introduce a guard interval to take into account possible uncertainty in the detection of the first returns. The selection of the value of w_G should be made according to the level of reliability achieved by the first return detection. However, in our experiments the choice of the value of w_G has never been a critical issue. $w_G = 10$ has been used in all our tests. Such a value corresponds to a distance of approximately 112 m.

3) *Calculation of KL_{HN}*: a map of KL_{HN} is generated using a sliding window of $l_a \times l_r$ samples (azimuth $\times$ range), and a step of t_a and t_r samples in the azimuth and range direction, respectively. The distribution of N is the one estimated in the previous step. The value of KL_{HN} is averaged in the intersections of overlapping windows. This process is applied only to the subsurface part of the radargram, which is defined as the bottom part of the radargram delimited by the first return line. The choice of the size and of the steps of the sliding window should be driven by the characteristics of the considered targets, which are generally extended in the azimuth direction but can present sharp variations in the range direction. Fig. 8b shows the values of KL_{HN} obtained on the test radargram using $l_a = 40$, $l_r = 10$, $t_a = 8$, and $t_r = 10$. These values correspond to about 11–20 km, 63 m, 2.2–4 km, and 63 m, respectively (range distances have been calculated using $\varepsilon_r = 3.15$).

4) *Thresholding*: in this step the algorithm produces a binary map which discriminates between the presence and the absence of subsurface features by thresholding the image of KL_{HN} using the threshold thr_{KL}. The value of thr_{KL} can be chosen either manually or automatically [38]. Fig. 8c shows the binary map obtained from Fig. 8b by using $thr_{\mathrm{KL}} = 0.13$.

B. Results and Discussion

The results presented in Fig. 8b show a description of the characteristics of the subsurface in the radargram of Fig. 3 in terms of statistical difference from the background noise computed according to the values of the KL_{HN} distance. Such a difference may vary within the same scattering class. As an example, the statistical characteristics of the scattering coming from the basal area are not uniform. Fig. 8c shows the map of the subsurface features of the aforementioned radargram. A qualitative comparison between Fig. 3 and Fig. 8c points out the high accuracy obtained by the proposed technique in the detection of subsurface features. In order to measure quantitatively the performance of the proposed algorithm, for each tested radargram we selected randomly 3000 reference samples from the regions where it was possible to state clearly whether subsurface features were present or absent. Using such samples, we evaluated the number of missed and false detections yielded by the proposed algorithm with the parameters reported in the previous subsection. The obtained results (see Tab. IV) point out that most of the subsurface features present in the radargrams are correctly detected. As it is visible in Fig. 8c, the areas corresponding to the target classes *strong layers*, *weak layers* and *basal returns* are mostly correctly detected by the algorithm, whereas the class *low returns* is only partially detected. This can be explained by the simple sliding window and averaging model adopted in this paper. This acts as a low pass filtering and averages the statistical characteristics of the classes *low returns* and *no target*, which become similar from the statistical point of view. The choice of the parameters of the algorithm should be driven by both the sensitivity needed for the detection and the type of features which have to be detected.

The results of the proposed algorithm can be a starting point for a subsequent more detailed analysis of the detected targets, which can be achieved by estimating the statistical parameters of the local distributions, according to a given fitting model, e.g., the K distribution.

From a more general point of view, the performance obtained by the proposed technique allows one to assess with very high reliability whether a radargram contains or does not contain subsurface features. Thus, the technique can be effectively exploited to discriminate from the huge set of acquisitions the radargrams with significant subsurface features (which should be object of further analysis) from those that do not have subsurface features.

A possible extension of the proposed technique is to derive maps of the subsurface features by calculating the KL divergence between a theoretical distribution fitted to the local histogram (e.g., the K distribution) and the theoretical noise distribution. The use of the fitted distribution in place of the sample histogram can be seen as an implicit filtering of the signal aimed at discarding outliers.

It is worth noting that the presented approach cannot detect the difference between real subsurface features and clutter returns coming from the surface topography. The detection of clutter returns from single detected radargrams cannot be done automatically without the use of topographic data or clutter simulations. However, the proposed method can be simply integrated in a processing chain including a clutter detection step which masks the clutter areas in the radargram according to available clutter simulations.

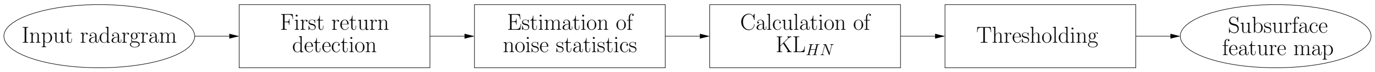
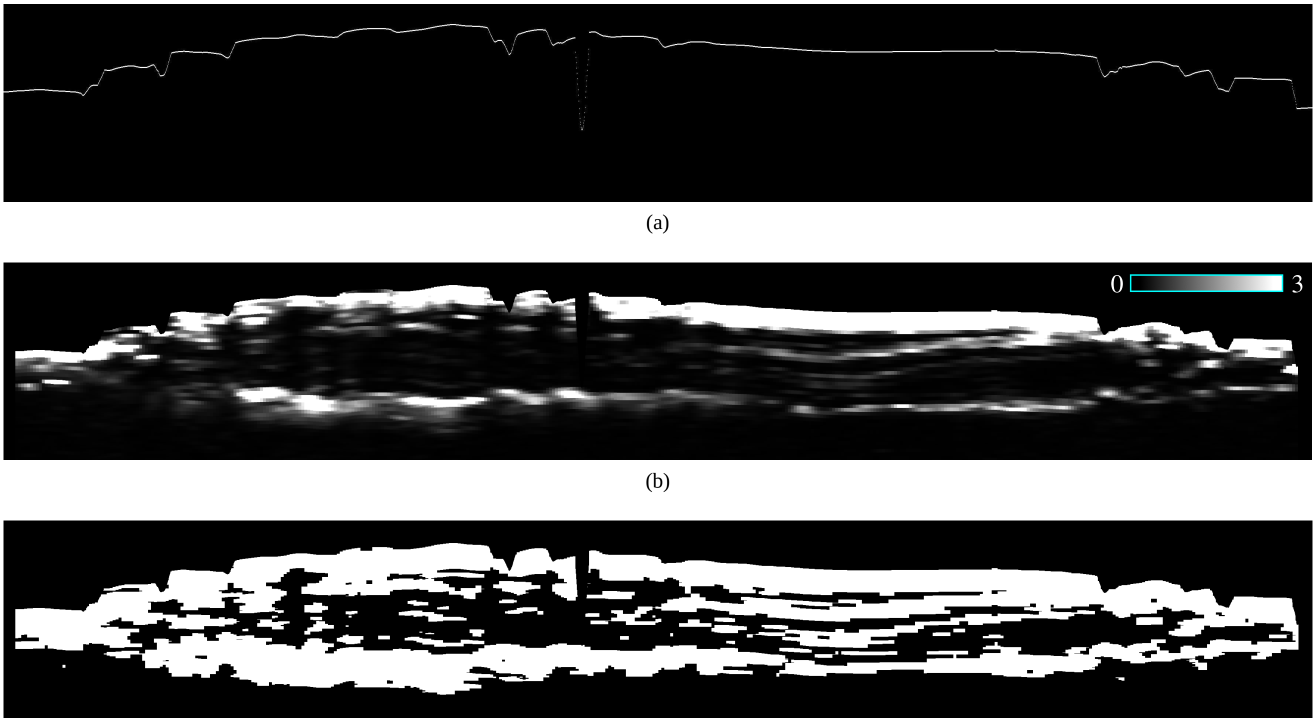


Fig. 7. Block scheme of the proposed technique for the generation of subsurface feature maps.



(a)

(b)

(c)

Fig. 8. (a) Detected first returns on SHARAD radargram 1319502 (see Fig. 3). (b) Map of KL_{HN} obtained on the same radargram. Values of $\mathrm{KL}_{HN} > 3$ have been saturated to 3 for visualization purposes. (c) Binary map obtained from (b) by thresholding KL_{HN} at $thr_{\mathrm{KL}} = 0.13$.

TABLE IV

ACCURACY PROVIDED BY THE PROPOSED TECHNIQUE FOR THE GENERATION OF SUBSURFACE FEATURE MAPS.

Radargram number	**Feature samples**	**Missed alarms**	**% missed alarms**	**Non-feature samples**	**False alarms**	**% false alarms**	**Total error**	**% total error**
0371502	492	28	5.69	2,508	240	9.57	268	8.93
0385902	515	50	9.71	2,485	189	7.61	239	7.97
0681402	830	44	5.30	2,170	305	14.06	349	11.63
0794703	718	8	1.11	2,282	362	15.86	370	12.33
1292401	491	9	1.83	2,509	277	11.04	286	9.53
1312901	625	21	3.36	2,375	304	12.80	325	10.83
1319502	657	34	5.18	2,343	318	13.57	352	11.73

V. PROPOSED TECHNIQUE FOR THE AUTOMATIC DETECTION OF BASAL RETURNS

In this section we propose an algorithm aimed at detecting the deepest scattering area of a radargram. We applied such a technique to the detection of the basal returns coming from the base of the NPLD in SHARAD radargrams. However, after proper tuning, it can be adapted to other operational conditions (e.g., to the detection of the bedrock returns in data acquired by airborne sounders on Earth's polar regions). As mentioned in Sec. III, basal returns in the NPLD include the scattering from the so-called basal unit. The basal unit is often described as sandy with varying amounts of volatiles [7]. Different hypotheses about its origin have been proposed in the literature [39]. SHARAD is able to penetrate the basal unit only to a certain extent. As shown in Sec. III, the returns coming from the basal unit in SHARAD radargrams are mostly diffuse. The mean amplitude of the signals varies spatially depending on the local geology of both the basal unit and the overlying ice stratigraphy. However, the results obtained in the following demonstrate that the statistical behavior of the signals is in average stationary at least in single SHARAD acquisitions.

A. Proposed Technique

A block scheme of the proposed technique is shown in Fig. 9. The technique is based on the statistical analysis carried out in Sec. III. It composes the basal scattering area using a region-growing approach. The obtained regions are kept or discarded according to the statistical distribution of their samples, which has to be similar to the expected distribution of the basal returns. The latter is estimated automatically by the algorithm. The technique is made up of two main phases: i) definition of an initial map of the basal scattering area, and ii) iterative refinement of the initial map. The two phases are described in detail in the following along with example images showing the main outputs (see Fig. 10). As test case we use the radargram shown in Fig. 3.

1) Definition of an initial map of the basal scattering area: the algorithm selects seed regions that have a high probability to belong to the basal scattering area. Then, it uses a region-growing approach which exploits a KL_{HN} map (calculated using the concepts introduced in Sec. IV) in order to produce a first initial map of the basal returns. In the following we describe in detail each step of this phase of the algorithm.

- *First return detection and calculation of KL_{HN}*: the radargrams are cut on the area of interest and the procedure described in Sec. IV is applied in order to detect the surface line, estimate the noise statistics, and calculate the KL distance between the local signal histogram and the estimated noise statistical distribution. The calculated image of KL_{HN} is used as basis for the next steps.
- *KL_{HN} thresholding*: the goal of this step is to extract the regions which have a statistical distribution significantly different from that for the noise distribution. Such regions will be used by the algorithm to select the seeds of the basal scattering area in the next step. Therefore, the map of KL_{HN} is thresholded using a threshold thr_1 in order to produce a binary image KL_1, defined as:

$$\mathrm{KL}_1(i,j) = \begin{cases} 1 & \text{if } \mathrm{KL}_{HN}(i,j) \geq thr_1 \\ 0 & \text{otherwise.} \end{cases} \tag{15}$$

  The value chosen for thr_1 should be high enough to identify only few small regions of the basal scattering area, besides strong scattering areas belonging mostly to the *strong layers* and *weak layers* classes. In our experiments a value equal to 1.2 fulfilled this condition. The image of KL_1 for the test radargram is shown in Fig. 10a.
- *BR seed selection*: the binary image KL_1 contains a set $R_{1,0}$ of disjoint regions. Only those which are likely to be related to the basal returns are kept. The selection is performed on the basis of geometrical criteria, which take into account the usual position of the basal returns in the radargrams, i.e., i) the regions should correspond to the maximum ranges (depths); ii) the regions must not belong to the neighborhood of the surface. Condition i) is verified by the subset of regions $R'_{1,0}$ defined as:

$$R'_{1,0} = \{r : r \in R_{1,0} \ \wedge \ \exists j : (i,j) \in r \\ \wedge \ i = \max\left\{\tilde{i} : \mathrm{KL}_1(\tilde{i},j) = 1\right\}\right\}. \tag{16}$$

  The subset $R''_{1,0}$ of regions of $R_{1,0}$ which fulfill condition ii) is defined as:

$$R''_{1,0} = \{r : r \in R_{1,0} \wedge r \cap R_s = \emptyset\}, \tag{17}$$

  where R_s is the subsurface neighborhood region of the first returns considering a distance w_{ss} from the first returns. Formally, it is given by:

$$R_s = \{(i,j) : f(j) < i < f(j) + w_{\text{ss}}\}. \tag{18}$$

  The selection of the value of w_{ss} should take into account the expected thickness of the area of the NPLD that is investigated. The final set of selected regions R_1 is composed by the regions of $R'''_{1,0} = R'_{1,0} \cap R''_{1,0}$ which fulfill the condition:

$$R_1 = \left\{r : r \in R'''_{1,0} \\ \wedge \ \bar{i}_{R'''_{1,0}} - w_{\text{up}} < \bar{i}_r < \bar{i}_{R'''_{1,0}} + w_{\text{down}}\right\}, \tag{19}$$

  where $\bar{i}_{R'''_{1,0}}$ is the weighted mean range position of the regions contained in $R'''_{1,0}$ (using the areas of the regions as weights), $\bar{i}_r$ is the mean range position of the region r, and w_{up} and w_{down} are tolerance thicknesses used to define the width of the range of the expected basal position. w_{up} is referred to the thickness toward the surface, while w_{down} represents the thickness towards the bottom of the subsurface. On the one hand, the choice of w_{down} is in general not critical as usually no returns are observed after the basal scattering area. On the other hand, similarly to the discussion about w_{ss}, the value of w_{up} should be chosen according to the expected thickness of the investigated NPLD region. The output of this step for the test radargram is shown in Fig. 10b.
- *Region growing*: the regions selected in the previous step are used as seeds for a level-set algorithm. Such an algorithm stretches their contour to fit the basal scattering area using the KL_{HN} image. The algorithm describes the contour as the zero level set of the function given by the following differential equation:

$$\frac{d}{dt}\psi = \left[-\alpha P(i,j) + \beta C\right] |\nabla \psi|, \tag{20}$$

  where $\alpha P(i,j)$ drives the expansion of the contour, and the term βC affects its curvature (and thus the "smoothness" of the detection), C is calculated as the mean curvature of the contour, and α and β are scalar values which define the weight of each term of the equation. In the proposed approach, the term $P(i,j)$ is calculated as

$$P(i,j) = \begin{cases} \mathrm{KL}_{HN}(i,j) - thr_L \\ \qquad \text{if } \mathrm{KL}_{HN}(i,j) < \frac{thr_U - thr_L}{2} + thr_L \\ thr_U - \mathrm{KL}_{HN}(i,j) \\ \qquad \text{otherwise,} \end{cases} \tag{21}$$

  where thr_U and thr_L define the upper and the lower thresholds of KL_{HN}, respectively, which limit the expansion of the contour. Using the definition in (21) the propagation term $P(i,j)$ is positive (expansion) only when $\mathrm{KL}_{HN}(i,j) \in (thr_U, thr_L)$. More details on level sets can be found in [40]. The choice of the values of

thr_U and thr_L depends on the limit values of KL_{HN} associated with the basal returns. The most important parameter is thr_L, as it defines the minimum statistical difference to the background noise which makes the contour expand.

At the end of the region growing step, the algorithm has produced an initial map of the basal scattering area composed by a set of regions $R_{1,\text{grow}}$.

2) Iterative refinement of the initial map: an $(M-1)$-step iterative procedure is started, which is aimed at detecting the weak scattering areas of the basal returns and refining the previous detection. The steps of the iterative loop performed for each iteration m $(m = 2, \ldots, M)$ are as follows.

- *Estimation of BR statistics*: in this step the algorithm uses the amplitudes of the samples belonging to the regions of $R_{m-1,\text{grow}}$ to estimate the parameters of a K distribution. The estimation is performed using an ML approach (see Sec. III). In this way, the algorithm estimates the statistical distribution of the basal returns, which will be exploited in the next steps.
- *KL$_{HN}$ thresholding*: a binary map is produced by considering only the samples of KL_{HN} which belong to the range $[thr_m, thr_{m-1})$. For each iteration m, the binary image KL_m is thus created according to:

$$\mathrm{KL}_m(i,j) = \begin{cases} 1 & \text{if } thr_m \leq \mathrm{KL}_{HN}(i,j) < thr_{m-1} \\ 0 & \text{otherwise.} \end{cases} \quad (22)$$

  The binary map contains a set $R_{m,0}$ of regions. The value of each thr_m and the number of iterations M should ensure that for each iteration the binary map contains regions with significant areas. Moreover, the value of thr_M must be greater than thr_L to assure that the level-set algorithm can expand the region contours also in the last iteration ($m = M$).
- *Selection of BR seeds and region growing*: similarly to the previous steps, the binary maps are used to select seed regions, which are likely to belong to the basal scattering area, and the level-set algorithm is run starting from such seeds. The subset of seed regions R_m is selected by means of geometrical constraints, i.e., the regions must belong to a range neighborhood of the estimated basal mean range. This is formally translated in a condition similar to (19):

$$R_m = \{r : r \in R_{m,0} \\ \wedge\ \bar{i}_{R_{m-1,\text{grow}}} - w_{\text{up}} < \bar{i}_r < \bar{i}_{R_{m-1,\text{grow}}} + w_{\text{down}}\}, \quad (23)$$

  where $\bar{i}_{R_{m-1,\text{grow}}}$ is the weighted mean range position of the regions contained in $R_{m-1,\text{grow}}$ (using their areas as weights), $\bar{i}_r$ is the mean range position of the region r, and w_{up} and w_{down} are the same tolerance thicknesses as those used in (19).
- *Region selection*: a subset of the regions obtained in the previous step is selected. The selection is made mainly on a statistical basis. For each region the histogram is computed and if its KL distance to the estimated basal return distribution is smaller than an user-defined threshold thr_G the region is kept, otherwise it is discarded. This step is performed to discard the regions which grew on areas which are not related to the basal scattering area. Therefore, the value of thr_G should be small (e.g., on the order of 0.1). Once the selection is performed, the provisional set of the basal return regions $R_{m-1,\text{grow}}$ is merged with the new regions obtaining the new set $R_{m,\text{grow}}$. Such a set will be the input for the next iteration.

The result of the iterative phase is thus a binary map composed by the merging of the whole set of regions produced during the different iterations.

Finally, small isolated regions are deleted and the final basal return map is created. The resulting basal return area detected on the test radargram is shown in Fig. 10c.

B. Results and Discussion

Fig. 11 reports the detected basal return areas of three radargrams. A qualitative analysis of the results points out that the proposed technique is able to detect with high accuracy the scattering areas related to the basal returns both in azimuth and in range direction. The worst performance is related to the detection of the NPLD base interface when layering is visible at close depths. As the subsurface layering is very close to the basal scattering area, the statistics of the two target types are very similar; thus, the algorithm may fail to discard the layered areas.

In order to measure quantitatively the performance of the algorithm, we followed an approach similar to that used in Sec. IV. For each of the 7 radargrams analyzed in the paper, we considered 3000 reference samples randomly taken in the areas of the radargrams for which it was possible to assess clearly the presence (or the absence) of basal returns. The results regarding the detection performance are reported in Tab. V in terms of number of missed and false alarms calculated using the selected samples. Taking into account that the proposed algorithm is automatic and unsupervised, the overall accuracy can be considered high.

An additional note should be made about the choice of the parameters of the proposed algorithm. As already discussed in the previous subsection, the algorithm is stable with respect to several parameters as their choice is not critical and the same values can be used for a large set of radargrams. The most sensitive parameters are thr_1 and thr_L. Indeed, thr_1 affects the definition of the initial seeds of the algorithm, while thr_L defines the minimum statistical difference that the basal returns must have with respect to the background noise. Therefore, such parameters should be chosen taking into account the average signal-to-noise ratio (SNR) of the analyzed radargram. This depends on the noise level, on the state of the subsurface materials (which affects the signal propagation), and on the spacecraft attitude (e.g., in certain configurations called *rolled acquisitions* the SHARAD antenna gain is greater than that for standard spacecraft attitude). From the practical viewpoint, this means that if the algorithm is run with the same parameters on a set of radargrams with similar SNR characteristics, its performances are almost constant on the whole set of radargrams. In addition, it is worth noting that almost all the parameters involved in the algorithm have a clear physical meaning that represents a guide for a proper tuning. For all the test radargrams considered in this paper we used the following algorithm parameters: $w_{\text{ss}} = 20$, $M = 3$, $thr_1 = 1.2$, $thr_2 = 0.7$, $thr_3 = 0.2$, $thr_U = 100$, $thr_L = 0.13$, $\alpha = 50$, $\beta = 10$, $w_{\text{up}} = 50$, $w_{\text{down}} = 100$, $thr_G = 0.10$. The values of w_{ss}, w_{up} and w_{down} correspond to approximately 127 m, 317 m, and 634 m in the subsurface using $\varepsilon_r = 3.15$. Using these parameters the computational time for a test radargram with 3500 frames is in the order of 5-7 minutes depending on the extension of the basal return region (the time includes the computation of the KL_{HN} map).

The output of the proposed algorithm can be used in scientific analysis for many purposes. A first application is the estimation of the NPLD thickness (assuming a reasonable dielectric constant for the icy materials of the NPLD) using a large set of acquisitions. Given the resolution of SHARAD radargrams, it is possible also to extrapolate from the detected basal topography local buried basins or impact craters. Another possible application is the measurement of the mean power scattered by the basal unit at a certain 3D position, which is useful to study local geology and radar bright (or dark) areas. Finally, the proposed technique can also be used to study seasonal variations of the signal propagation loss within the NPLD. This can be achieved by analyzing the amount of power scattered by the basal area during different seasons on the same areas, and relating such measurements to the absorption experienced by the signal within the NPLD.

VI. Conclusion

In this paper the problem of the automatic analysis of radar sounder signals acquired from orbiting platforms has been addressed. We presented both a study on the statistical properties of the sounder signals and two novel automatic techniques for the extraction of subsurface features from radargrams. In the study of the properties of sounder signals we analyzed different statistical models from a theoretical point of view and then empirically tested them on different real SHARAD data acquired on the NPLD of Mars. The obtained results show that the statistical distributions of the amplitude signals related to different types of targets can be modeled precisely using the K distribution, while, as expected, the background noise follows a Rayleigh distribution. Exploiting the results of the aforementioned study, we have then proposed two novel techniques for the automatic analysis of radargrams aimed at: i) producing maps of the subsurface areas showing relevant features; and ii) identifying and mapping the deepest scattering areas visible in the radargrams. The former is based on the comparison of the distributions of local subsurface parcels with that of noise adaptively estimated on each radargram. The latter exploits a specifically defined region-growing method implemented in an iterative technique based on the level-set algorithm. The results obtained by both the developed techniques are accurate and thus promising for operational applications.

The statistical analysis, the techniques and the results described in this paper are a first step to the definition of a general framework for the analysis of radar sounder data. The goal of such a framework is to extend the low-level processing chain currently applied to the downlinked data with information extraction steps. To this end, additional automatic techniques for the extraction of features and parameters from radargrams should be developed with respect to what was presented in this paper. This should be done by taking into account indications provided from scientists expert of the considered application and of the related requirements. The framework could be also extended to the use of input data coming from other sensors (e.g., optical images of the investigated area) or other information sources (e.g., a simulator for clutter cancellation).

Although human interpretation cannot be fully replaced by automatic algorithms, automatic methods can significantly help to overcome the subjectivity intrinsic in manual investigations by providing in a fast way numerical results obtained with predefined and fixed metrics. These results can then drive further manual refinements. This research field is also very important for future radar sounding missions. Indeed, the techniques developed for the analysis of present planetary radar sounder data represent a valuable starting point for the analysis of the data acquired by possible future missions that will investigate other planetary bodies (e.g., Europa and Ganymede) or the Earth.

As a future development, we will study novel methods for the generation of subsurface feature maps based on the local statistics using context-sensitive techniques for the adaptive determination of the local parcel size. Moreover, we plan to develop a procedure for the automatic and adaptive definition of the parameters of the proposed techniques. Finally, we will also focus on the identification of automatic methods for the detection and the filtering of surface clutter returns from the radargrams.

Acknowledgment

The authors would like to thank Dr. Mauro Dalla Mura for fruitful discussion. We also acknowledge Dr. Glen Davidson for freely distributing some of the routines used in this paper [41], and the Orfeo Toolbox (OTB) [42] and Insight Toolkit (ITK) [43] communities for publishing these libraries as open source products. We would also like to thank ASI for supporting this work.

References

[1] V. Bogorodsky, C. Bentley, and P. Gudmandsen, *Radioglaciology*. D. Reidel Publishing Co., 1985.

[2] R. Jordan et al., "The Mars Express MARSIS sounder instrument," *Planetary and Space Science*, vol. 57, pp. 1975–1986, 2009.

[3] R. Croci, R. Seu, E. Flamini, and E. Russo, "The SHAllow RADar (SHARAD) Onboard the NASA MRO Mission," *Proc. IEEE*, vol. 99, no. 5, pp. 794–807, 2011.

[4] J.J. Plaut et al., "Subsurface radar sounding of the South Polar Layered Deposits of Mars," *Science*, vol. 316, pp. 92–95, 2007.

[5] R. Seu et al., "Accumulation and erosion of Mars' South Polar Layered Deposits," *Science*, vol. 317, pp. 1715–1718, Sep. 2007.

[6] R. J. Phillips et al., "Mars North Polar Deposits: Stratigraphy, age, and geodynamical response," *Science*, vol. 320, pp. 1182–1185, May 2008.

[7] M. Selvans, J. Plaut, O. Aharonson, and A. Safaeinili, "Internal structure of Planum Boreum, from Mars advanced radar for subsurface and ionospheric sounding data," *J. Geophys. Res.*, vol. 115, p. E09003, 2010.

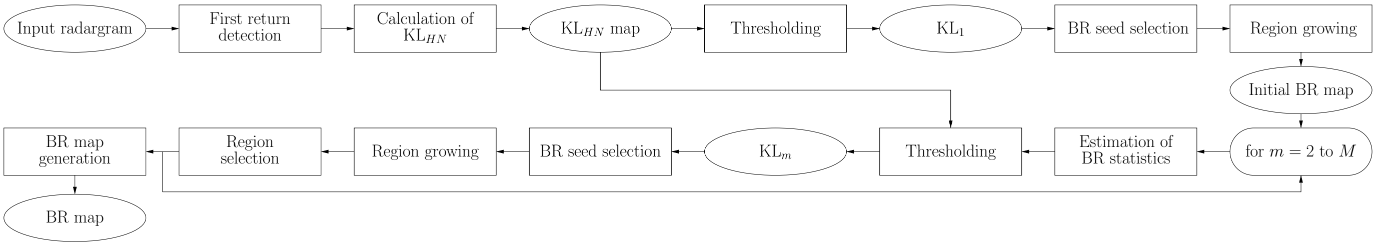
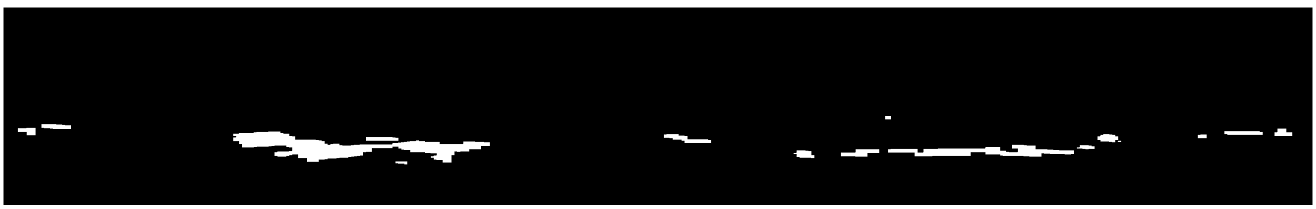
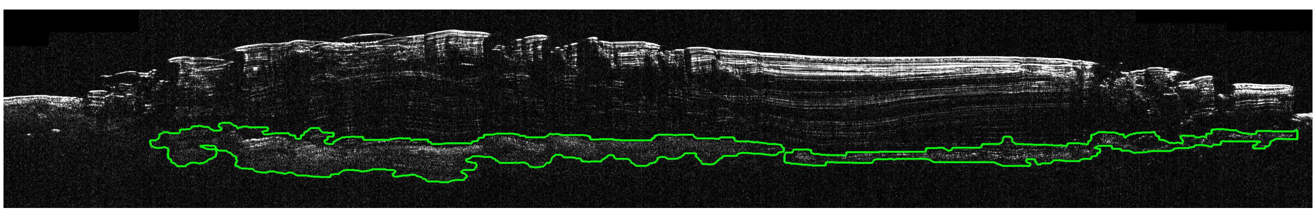


Fig. 9. Block scheme of the proposed technique for the automatic detection of basal returns.

(a)

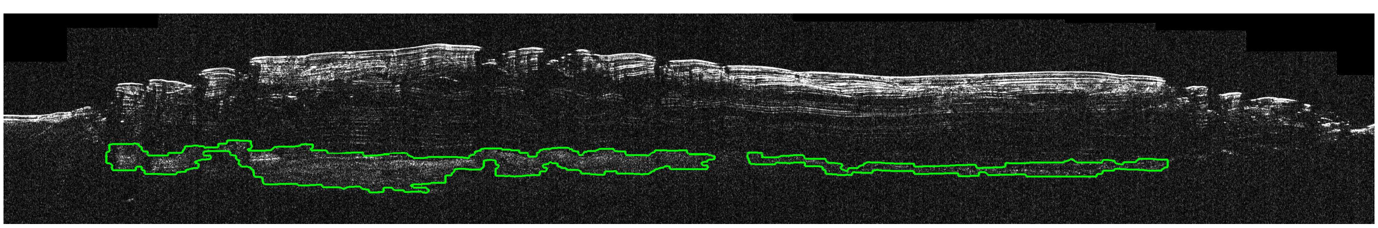


(b)

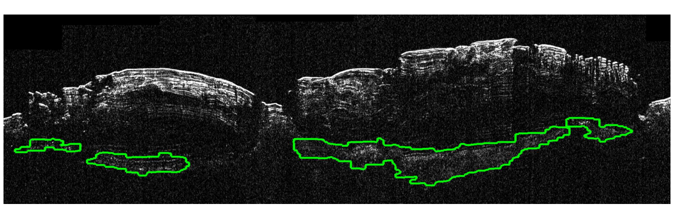


(c)

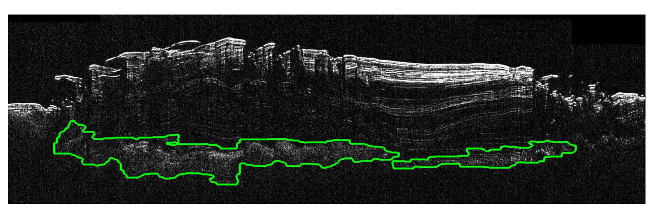
Fig. 10. Example of application of the proposed algorithm for the detection of the basal returns to SHARAD radargram 1319502 (see Fig. 3). (a) Areas remaining after the first thresholding ($thr_1 = 1.2$). (b) Selected starting seed regions. (c) Final detection.



(a)



(b)



(c)

Fig. 11. Detected basal scattering area on SHARAD radargrams (a) 0371502, (b) 1292401, and (c) 1312901.

TABLE V
ACCURACY PROVIDED BY THE PROPOSED TECHNIQUE FOR THE DETECTION OF BASAL RETURNS.

Radargram number	Feature samples	Missed alarms	% missed alarms	Non-feature samples	False alarms	% false alarms	Total error	% total error
0371502	250	30	12.00	2,750	37	1.35	67	2.23
0385902	281	51	18.15	2,719	30	1.10	81	2.70
0681402	340	61	17.94	2,660	59	2.22	120	4.00
0794703	282	19	6.74	2,718	71	2.61	90	3.00
1292401	124	9	7.26	2,876	90	3.13	99	3.30
1312901	240	5	2.08	2,760	93	3.37	98	3.27
1319502	271	25	9.23	2,729	80	2.93	105	3.50

[8] J. W. Holt, A. Safaeinili, J. J. Plaut, J. W. Head, R. J. Phillips, R. Seu, S. D. Kempf, P. Choudhary, D. A. Young, N. E. Putzig, D. Biccari, and Y. Gim, “Radar sounding evidence for buried glaciers in the southern mid-latitudes of mars,” *Science*, vol. 322, pp. 1235–1238, 2008.

[9] (2009, Jan.) Europa Jupiter System Mission NASA/ESA joint summary report. [Online]. Available: http://sci.esa.int/ejsm/

[10] L. Bruzzone, G. Alberti, C. Catallo, A. Ferro, W. Kofman, and R. Orosei, “Subsurface radar sounding of the Jovian Moon Ganymede,” *Proc. IEEE*, vol. 99, no. 5, pp. 837–857, 2011.

[11] IGOS, “Cryosphere theme report,” 2007. [Online]. Available: http://cryos.ssec.wisc.edu/docs/cryos_theme_report.pdf

[12] P. Gamba and S. Lossani, “Neural detection of pipe signatures in ground penetrating radar images,” *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 2, pp. 790 –797, Mar. 2000.

[13] M. Fahnestock, W. Abdalati, S. Luo, and S. Gogineni, “Internal layer tracing and age-depth-accumulation relationships for the northern Greenland ice sheet,” *J. Geophys. Res.*, vol. 106, no. 33, pp. 789–797, 2001.

[14] C.-C. Chen, M. B. Higgins, K. ONeill, and R. Detsch, “Ultrawide-bandwidth fully-polarimetric ground penetrating radar classification of subsurface unexploded ordnance,” *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 6, pp. 1221–1230, Jun. 2001.

[15] S. Delbo, P. Gamba, and D. Roccato, “A fuzzy shell clustering approach to recognize hyperbolic signatures in subsurface radar images,” *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 3, pp. 1447 –1451, May 2000.

[16] F. Heliere, C.-C. Lin, H. Corr, and D. Vaughan, “Radio echo sounding of Pine Island Glacier, West Antarctica: Aperture synthesis processing and analysis of feasibility from space,” *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 8, pp. 2573–2582, Aug. 2007.

[17] L. M. Carter et al., “Shallow radar (SHARAD) sounding observations of the Medusae Fossae Formation, Mars,” *Icarus*, vol. 199, no. 2, pp. 295–302, Feb. 2009.

[18] S. Lauro, E. Mattei, E. Pettinelli, F. Soldovieri, R. Orosei, M. Cartacci, A. Cicchetti, R. Noschese, and S. Giuppi, “Permittivity estimation of layers beneath the northern polar layered deposits, Mars,” *Geophys. Res. Lett.*, vol. 37, p. L14201, 2010.

[19] J. Boisson, E. Heggy, S. Clifford, A. Frigeri, J. Plaut, W. Farrel, N. Putzig, G. Picardi, R. Orosei, P. Lognonne, and D. Gurnett, “Sounding the subsurface of Athabasca Valles using MARSIS radar data: Exploring the volcanic and fluvial hypotheses for the origin of the rafted-plate terrain,” *J. Geophys. Res.*, vol. 114, p. E08003, 2009.

[20] D. J. Daniels, *Ground penetrating radar (2nd Edition)*. London, UK: Institution of Engineering and Technology, 2007.

[21] E. Pettinelli, P. Burghignoli, A. R. Pisani, F. Ticconi, A. Galli, G. Vannaroni, and F. Bella, “Electromagnetic propagation of GPR signals in Martian subsurface scenarios including material losses and scattering,” *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 5, pp. 1271–1281, May 2007.

[22] C. Oliver and S. Quegan, *Understanding Synthetic Aperture Radar Images*. Raleigh, NC, USA: SciTech Publishing, Inc., 2004.

[23] R. Orosei, R. Bianchi, A. Coradini, S. Espinasse, C. Federico, A. Ferriccioni, and A. Gavrishin, “Self-affine behavior of Martian topography at kilometer scale from Mars Orbiter Laser Altimeter data,” *J. Geophys. Res.*, vol. 108, no. E4, p. 8023, 2003.

[24] N. Nakagami, “The m-distribution, a general formula for intensity distribution of rapid fading,” in *Statistical Methods in Radio Wave Propagation*, W. G. Hoffman, Ed. Oxford, England: Pergamon, 1960.

[25] A. Papoulis, *Probability, Random Variables, and Stochastic Processes, 2nd ed.* New York, US: McGraw-Hill, 1984.

[26] D. J. Lewinski, “Nonstationary probabilistic target and clutter scattering models,” *IEEE Trans. Antennas Propag.*, vol. AP-31, no. 3, pp. 490–498, May 1983.

[27] K. Ward, “Compound representation of high resolution sea clutter,” *Electron. Lett.*, vol. 17, no. 16, pp. 561–563, Aug. 1981.

[28] C. Grima, W. Kofman, A. Herique, and R. Seu, “Physical parameters of the near-surface of Mars derived from SHARAD radar reflectivity: statistical approach,” in *38th COSPAR Scientific Assembly*, 18-28 July, Bremen, Germany, 2010.

[29] M. Zuber, D. Smith, S. Solomon, D. Muhleman, J. Head, J. Garvin, J. Abshire, and J. Bufton, “The Mars Observer Laser Altimeter investigation,” *J. Geophys. Res.*, vol. 97, no. E5, pp. 7781–7797, 1992.

[30] S. Slavney and R. Orosei, “Shallow radar Reduced Data Record software interface specification, ver. 1.0,” Jul. 2007.

[31] NASA’s Planetary Data System. [Online]. Available: http://pds-geosciences.wustl.edu

[32] J. Greenwood and D. Durand, “Aids for fitting the gamma distribution by maximum likelihood,” *Technometrics*, vol. 2, pp. 55–65, 1960.

[33] Q. Zhang, “A note on the estimation of Nakagami-m fading parameter,” *IEEE Commun. Lett.*, vol. 6, no. 6, pp. 237–238, Jun. 2002.

[34] I. Joughin, D. Percival, and D. Winebrenner, “Maximum likelihood estimation of K distribution parameters for SAR data,” *IEEE Trans. Geosci. Remote Sens.*, vol. 31, no. 5, pp. 989–999, Sep. 1993.

[35] J. Lin, “Divergence measures based on the shannon entropy,” *Information Theory, IEEE Transactions on*, vol. 37, no. 1, pp. 145 –151, Jan. 1991.

[36] H. Shimazaki and S. Shinomoto, “A method for selecting the bin size of a time histogram,” *Neural Computation*, vol. 19, no. 6, pp. 1503–1527, 2007.

[37] U. Stilla and K. Hedman, “Feature fusion based on bayesian network theory for automatic road extraction,” in *Radar Remote Sensing of Urban Areas*, ser. Remote Sensing and Digital Image Processing, U. Soergel, Ed. Springer Netherlands, 2010, vol. 15, pp. 69–86.

[38] L. Bruzzone and D. Fernandez Prieto, “Automatic analysis of the difference image for unsupervised change detection,” *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 3, pp. 1171–1182, 2000.

[39] K. E. Fishbaugh and J. W. H. III, “Origin and characteristics of the Mars north polar basal unit and implications for polar geologic history,” *Icarus*, vol. 174, no. 2, pp. 444 – 474, 2005.

[40] J. Sethian, *Level set methods and fast marching methods: evolving interfaces in computational geometry, fluid mechanics, computer vision, and materials science, 2nd ed.* Cambridge, UK: Cambridge University Press, 1999.

[41] Matlab Radar Toolbox v0.11. [Online]. Available: http://www.radarworks.com

[42] J. Inglada and E. Christophe, “The Orfeo Toolbox remote sensing image processing software,” in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2009.

[43] L. Ibanez, W. Schroeder, L. Ng, J. Cates, Consortium T.I.S., and R. Hamming, *The ITK Software Guide*. Kitware, Inc., 2003.