Speckle noise and soil heterogeneities as error sources in a Bayesian soil moisture retrieval scheme for SAR data

Matias Barber, *Student Member, IEEE,* Francisco Grings, Pablo Perna, Marcela Piscitelli, Martin Maas, Cintia Bruscantini, Julio Jacobo-Berlles and Haydee Karszenbaum.

Abstract—Soil moisture retrieval from SAR images is always affected by speckle noise and uncertainities associated to soil parameters, which impact negatively on the accuracy of soil moisture estimates. In this paper a Bayesian model is proposed to address these issues. A soil moisture Bayesian estimator from polarimetric SAR images is presented. This estimator is based on a set of stochastic equations for the polarimetric soil backscattering coefficients, which naturally includes models for the soil scattering, the speckle and the soil spatial heterogeneity. Since it is a Bayesian estimator, it may extensively use a priori information about soil condition, enhancing the performance of the retrieval. The Oh model is used as scattering model, although it presents a limiting range of validity for retrieving. After fully stating the mathematical modeling, numerical simulations are presented. First, traditional minimization-based retrieval using Oh model is investigated. The Bayesian retrieval scheme is then compared with Oh's retrieval. The results indicate that Bayesian model enlarge the validity region of Oh's retrieval. Moreover, as speckle effects are reduced by multilooking, Bayesian retrieval approachs to Oh's retrieval. On the other hand, an improvement in the accuracy of the retrieval is achieved by using a precise prior when speckle effects are large. The proposed algorithm can be applied to investigate which are the optimum parameters regarding multi-loking process and prior information required to perform a precise retrieval in a given soil type/condition.

Index Terms—Soil moisture, radar applications, Bayesian methods, synthetic aperture radar, inverse problems.

I. INTRODUCTION

Surface soil moisture content plays a key role in the interaction between the land surface and the atmosphere, and accurate knowledge about this variable is of interest for a variety of reasons. First, it is strongly related to vegetation development. Second, it is a predictor of the partitioning between rainfall into infiltration and runoff, which is strongly related to erosion of top soil through leaching. Third, when soil moisture is high, infiltration decreases and the risk of floods due to rainfall increases. And finally, soil evaporation and transpiration depends on soil moisture and therefore it influences the heat and mass transfers between the Earth and the atmosphere [1] .

M. Barber, F. Grings, P. Perna, M. Maas, C. Bruscantini and H. Karszenbaum are with the Instituto de Astronomía y Física del Espacio (IAFE), Argentina (e-mail: mbarber@iafe.uba.ar).

M. Piscitelli is with the Cátedra de Conservación y Manejo de Suelos, Facultad de Agronomía, Universidad Nacional del Centro de la Pcia. de Buenos Aires (UNICEN), Argentina.

J. Jacobo-Berlles is with the Departamento de Computación, Facultad de Ciencias Exactas y Naturales (FCEN), Universidad de Buenos Aires (UBA), Argentina.

Following this demand of information, there is a systematic effort to develop maps of soil moisture of the Earth's surface. Orbiting microwave synthetic aperture radar (SAR) systems offer the opportunity of monitoring soil moisture content at different scales and under any weather condition, through the known sensibility that the backscattered signal exhibits to soil parameters, including, among others, soil moisture and soil roughness [2]. Polarimetric SAR systems are able to transmit and receive radiation that is linearly polarized in the horizontal (h) and vertical (v) planes (relative to the plane defined by the wave vector and the normal to the surface being illuminated), giving rise to four intensity images hh, hv, vh, and vv of the target of interest [3].

However, the relation between backscattered signal and soil parameters is not straightforward at all, and consequently there are still no operational SAR-derived soil moisture products. This has two main reasons: (1) the scattering processes that relate backscattering to soil properties (moisture, roughness, and others) are difficult to model [4], and (2) the necessary input parameters are difficult to measure in the field [5], [6]. The former is mainly related to the SAR imaging system whereas the latter to soil parameters heterogeneity.

Moreover, typically there exist many combinations of surface parameters producing the same SAR observations. As a consequence, any retrieving scheme is an ill-posed inverse problem. Accordingly, soil parameters retrieval remains challenging, and soil moisture products derived from remotely-sensed SAR data are still poorly accurate [7].

Restricting our study to bare soils, surface soil moisture presents a high degree of spatial variability at differents scales, even for relatively small areas. This is associated to water-routing processes, radiative effects and heterogeneity in vegetation and soil characteristics [5]. On the other hand, heterogeneity of surface roughness arises from both man-made and natural factors: tillage system, soil texture, soil type, among others [6].

When using SAR images for retrieving soil properties, the speckle phenomenon, characteristic of SAR images, further hinders soil moisture retrieval. Speckle leads to a grain-like appearance of SAR images decreasing their contrast and radiometric quality [3]. It is characteristic of the coherent nature of the SAR imaging system, can be modeled as a multiplicative noise and it is usually reduced in a post-processing stage by: (1) averaging neighboring pixels (multi-looking process) at the expense of spatial resolution [8] or (2) using adaptive

filters [9], to reduce radiometric uncertainties without losing spatial resolution, but at the expense of introducing artifacts. It is important to note that the process of averaging to reduce radiometric uncertainties implicitly assumes that soil properties within the average window are constant, which is usually not the case in common bare soils. Therefore, a trade-off between averaging and soil properties heterogeneity is usually accepted. However, heterogeneity of soil properties and speckle are usually considered as independent problems, whereas they are indeed a part of the same inference problem.

In this general framework, soil moisture retrieval over bare soils from SAR images can be considered an inference problem, where one essentially wants to infer soil condition given a set of measured backscatter coefficients and ancillary information. Polarimetric [10], [11], possibilistic [12], radar backscatter modeling (theoretical and semi-empirical) [13], [14], [15], [4] and Bayesian approaches [16], [17], [18] are among the retrieval methodologies offered in the literature.

Polarimetric methods are based on modeling the backscatter response in terms of a certain polarimetric matrix decomposition (see [19] for a review) and taking into account the amplitude as well as the phase difference of the measured backscattering coefficients. Although polarimetry looks promising, a major effort should be still done to achieve an operational soil moisture retrieval algorithm using these techniques. Such algorithm was only developed in closed form for the Small Pertubation Model [11], which has a highly restrictive range of validity for the normalized RMS height ($ks \ll 0.3$), limiting operational soil parameter retrieval to very smooth surfaces. Therefore, this method is not suitable for real applications, where it is usually found values of $ks \sim 0.3$ for L-band (i.e. $s(RMS) = 1$ cm). In addition, speckle noise is not taken into account, although a polarimetric SAR speckle noise model was developed in [10].

The possibilistic methods make use of an alternative axioms set called fuzzy logic. As an advantage, they enable and required the use of prior information, which is used to improve the retrieving of soil parameters. On the other hand, they do not take into account speckle and they are computationally intensive [12].

Regarding radar backscatter modeling approaches, a wide range of forward models, ranging from semiempirical to theoretical, physically-based models have been developed in order to assess the dependency of soil parameters to backscattered signal. These models are important to understand the physics related to soil backscattering, but they also play a key role in the retrieval of soil condition from SAR measurements, since many retrieval algorithms only need a forward model.

Physical Optics model (PO), Geometrical Optics model (GO), the first-order Small Pertubation Model (SPM) and the Integral Equation Model (IEM) [4] with its further improvements and updates [20], [21], [22] are the analytical electromagnetic backscattering models available. Their strength lies in the fact that they are derived from the well-established electromagnetic theory. However, the first three of them have been derived considering some specific assumptions and therefore have a limited applicability in terms of surface roughness. Although IEM is valid for a wider range of surface roughness

conditions, the complexity of the model and the implicit relationship between soil parameters and soil backscattering make difficult to perform a direct retrieval.

Semi-empirical models are the most popular for soil moisture retrieval applications. This is related to their simple algebraic formalism, that allows a straightforward retrieval scheme being the usual ones the direct inversion [14] and minimization (look-up table) procedures [15], [23]. The standard approach for the development of these models is to measure soil backscattering at different polarizations, incidence angles and soil conditions using scatterometers, for thus deriving a model. In all the semi-empirical models [13], [14], [15] available, only the mean value of the backscattering coefficient as a function of soil parameters is modeled, disregarding the spread around the average value and its causes. This gives rise to characteristic artifacts where several values of soil moisture estimated from scatterometer data correspond to the same soil moisture measured on the field [16]. Reasons for mismatches between model estimations and measured data include system measurement errors, the inhomogeneity of soil parameters within a given system resolution cell (or from one cell to the next) and the difficult to measure soil parameters on the field [6], [24], [16]. Regarding this, the most difficult parameter to measure and to interpret is the correlation length [6]. Concerning the Oh model [15], a simplified alternative version was modeled ignoring the correlation length, because of the insensitivity of the $vh - vv$ ratio on the roughness parameter.

To the authors' knowledge, it was not until Haddad *et. al* [16] that a systematic way to include uncertainties in the formalism of forward model based on Bayes' theorem was presented for soil parameters retrieving. Bayesian approaches have the main feature of potentially include many sources of uncertainty as well as many sources of information about the variables involved in the retrieval. Whereas the radar backscatter models give rise to several combinations of surface parameters that map the same SAR observations, the Bayesian algorithm appropriately assimilates a priori information on geophysical parameters in order to constrain the inversion of forward models. Despite these outstanding features, in his original paper Haddad *et. al* [16] only included a term as error source related to model uncertainties and used only uniform distributions as prior. In addition, the potential of such Bayesian methodology is pointed out in [17] where data from active and passive sensors were merged in order to retrieve soil moisture. Nevertheless, up to date there is no model that incorporates multi-looking speckle noise as an error source, despite of the fact that Mattia *et al.* [18] included a rather simple speckle noise model that works only for one-look imagery.

In this paper, we propose a Bayesian retrieval methodology which incorporates in a natural way soil parameters heterogeneity and speckle as sources of uncertainty that degrade the estimated soil moisture. Such a Bayesian approach (1) needs only a forward model (no retrieval model is required), (2) gives the optimal unbiased estimator for the soil moisture and its error, (3) can include as many error sources as required and (4) can include a priori information in a systematic way.

The methodology will be presented using a simplified version of the Oh model [15] as the forward model, in which the correlation length is disregarded.

The present paper has been divided as follows. In Section II a brief description of the general properties of scatterometer-based semi-empirical forward models is presented, focused on Oh model and the multiplicative model. Section III is devoted to present the statistical model, and Bayesian estimators are derived. Numerical results are reported in Section IV. Finally, Section V present the main conclusions derived from the study presented in this paper.

II. SCATTEROMETER-BASED SEMI-EMPIRICAL FORWARD MODELS

A. Oh Model

The most widely used semi-empirical soil scattering model is the one developed by Oh [15], where model expressions are physically-based, but model parameters are derived from an extense database of polarimetric radar scatterometer measurements over bare soils. In its simplified version, where the correlation length is disregarded, the Oh model relates backscattering coefficients and certain bare soil properties through a set of three analitycal functions f_i, that can be symbolically expressed as [15, eqs. (1),(2) and (4)],

$$x_i = f_i(m, ks) \qquad (i = 1, 2, 3), \qquad (1)$$

where x_i is the backscattering (measured) coefficients and the subscript $i = 1, 2, 3$ stands respectively for the hh-, vv- and vh-polarizations. The backscattering coefficients x_i are functionally related to the volumetric soil moisture content m (cm^3/cm^3) and the normalized surface soil RMS height ks (where $k = 2\pi/\lambda$ is the wavenumber and s the RMS height) throughout the functions f_i. This model also depends on the system incidence angle θ, which is a known parameter. The Oh model is constrained to the range $0.04 \leq m \leq 0.291$ and $0.13 \leq ks \leq 6.98$, although the latter has a better agreement between the model and the experimental results for $ks \leq 3.5$ [15]. Explicitly from [15, eqs. (1),(2) and (4)],

$$f_3 = 0.11 m^{0.7}(cos\theta)^{2.2}[1 - exp(-0.32(ks)^{1.8})], \quad (2)$$

$$f_2 = \frac{f_3(m, ks)}{0.095(0.13 + sin(1.5\theta)^{1.4}[1 - exp(-1.3(ks)^{0.9})]}, \quad (3)$$

$$f_1 = f_2(m, ks)[1 - (\frac{\theta}{90°})^{0.35m^{-0.65}} exp(-0.4(ks)^{1.4})]. \quad (4)$$

Concerning the f_i functions, it is worth mentioning that they are not independent of each other, since by (1) there are three equations and only two variables. Then, providing that m and ks are given, it always holds

$$f_1 = \tilde{f}_1(m, ks)f_2 \qquad (5)$$

$$f_3 = \tilde{f}_3(m, ks)f_2, \qquad (6)$$

thus indicating that both hh and vh backscattering coefficients are a rescaled version of vv, where the derivation of functions $\tilde{f}_1$ and $\tilde{f}_2$ is straightforward from (4) and (3), respectively. This is a consequence of the deterministic nature of the Oh model. From (2), (3) and (4), it is easy to show that the backscattering coefficients for hh, vv and vh increase as m and ks increase. However, they increase with different growth rates each other; for a bare soil, vv is always greater than hh and the latter greater than vh. Any retrieval scheme using Oh model is based on the differential sensitivity exhibited by the backscattering coefficients to m and ks. The dynamic range in dB of the backscattering coefficients (eqs. (2-4)) is given in a nested way from the simplified formulation of the Oh model, constrained to $ks \leq 3.5$:

$$-42.2dB < vh < -15.6dB, \qquad (7)$$

$$vh + 10.8dB < vv < vh + 18.0dB \qquad (8)$$

and

$$vv - 2.7dB < hh < vv. \qquad (9)$$

The limiting values allowed by the inequations (7), (8) and (9) bound a general validity region where Oh model is valid inside. Only points (hh, vv, vh) within this region may be used to retrieve (m, ks) using Oh model.

Considering the aim of this work, it is relevant to consider the differences between the backscattering coefficients measured from SAR systems and the ones measured from scatterometers. First, the scatterometer footprint is small; the actual size varies for different experiments and sensors, but it is always of the order of a few squared meters. This justifies assuming that the soil properties on which measured microwave backscattering depends (soil moisture and roughness) are constant inside the sampled area. Therefore, it is reasonable to assume that the backscattering coefficient of the study area is a function of a single soil moisture value and roughness profile. In other words, the terrain scattering properties within the footprint can be considered constant. Second, it is easy to average several measurements upon the same surface's target and thus reduce the speckle noise.

On the other hand, SAR system resolution is larger (of the order of hundreds of m^2) and even larger if we want to average and increase the number of looks to reduce speckle. Therefore, any retrieval scheme based on SAR data that uses scaterometer-based models should deal with the heterogeneity of soil properties and the speckle. This will lead to non-constant soil scattering properties in the averaging window and/or non-negligible speckle noise, which in any case will degrade soil moisture retrieval.

B. Multiplicative Model

The multiplicative model is generally used to model the SAR response of a target as a function of the combined effect of terrain backscattering and speckle noise. Specifically, the model states that the observed intensity value in every pixel of a SAR image is the outcome of a random variable Z, called return, defined as the product between the random variables X and Y, where X represents the random variable modeling the variations of terrain backscattering properties and Y represents the random variable modeling the speckle noise; i.e. $Z = XY$ [3].

Different probability density distributions (PDF) for X and for Y yield different models for the observed data Z. For homogeneous regions, the terrain scattering properties are assumed constant. Therefore, the distribution of Z is a rescaled version of the distribution of Y, which is usually assumed as Gamma-distributed with parameters (n, n) and mean value $E[Y] = 1$ [3],

$$P_Y(y) = \frac{n^n}{\Gamma(n)} y^{n-1} e^{-ny} \qquad (10)$$

where n is the equivalent number of looks and $\Gamma(n)$ is the Gamma function. Since $Var[Y] = \frac{1}{n}$, as n approaches to infinity, the radiometric uncertainities related to speckle becomes negligible.

The basic hypothesis that governs the modeling of inhomogeneous regions ($X \neq constant$) is that their scattering properties are not constant, though they can be modeled by a convenient distribution. In our case, we will propose a PDF for X that arises as the result of inter-pixel soil parameters heterogeneity. Indeed, if soil parameters changes from pixel to pixel, soil backscattering (which is a function of soil parameters) will also change accordingly. In order to map soil parameters heterogeneity into backscattering heterogeneity, we will use Oh model. In this way, soil moisture and soil roughness PDFs can be mapped into soil backscattering PDF for all the polarizations.

By means of the multiplicative model, we can include two independent sources of SAR image inhomogeneity: soil spatial variability and speckle. This idea can be formalized as follows. First, we will assume that X and Y are independent. Second, the average properties of the return Z will be determined through the average properties of both X and Y, since by virtue of the multiplicative model,

$$E[Z] = E[X]E[Y]. \qquad (11)$$

Suitable distributions for X and Y will be introduced in sections III-C and III-D.

III. STATISTICAL MODEL

A. Bayesian Approach

The deterministic forward model developed by Oh can be extended to a stochastic model following [16]. In doing so, we can include in the forward model both the terrain heterogeneity and speckle through the multiplicative model,

$$Z_i = X_i Y_i \qquad (i = 1, 2, 3), \qquad (12)$$

where Z_i is the random variable which represents the return z_i and the subscript i stands for the different polarizations, as stated before. X_i and Y_i are independent random variables that model the heterogeneity of the target backscattering and the speckle noise, respectively.

From the point of view of the radar backscattering signal, we assumed that the target response to the backscatter is modeled through the Oh model by $X_i = f_i(M, KS)$ ($i = 1, 2, 3$), where the f_i are the same as in (1) and represent here the deterministic "typical" or average way in which the random variable X depends on the random variables M and KS

(which represent the m's and ks's of the target). In other words, an heterogeneous soil will produce a wide range of possible outcomes x of X, provided a wide range of soil moisture and roughness values were presented in the soil. On the other hand, an extremely homogeneous soil (i.e. a certain mean value $(\bar{m}, \bar{ks})$ with a very low standard deviation) will produce a very narrow probability density function for X. So it is reasonably to state that $E[X_i] = f_i(\bar{m}, \bar{ks})$, for all $i = 1, 2, 3$, where $\bar{m}$ and $\bar{ks}$ are the expected or mean values of M and KS within the resolution cell. In addition, we assume that the speckle adds only a multiplicative noise so that $E[Y_i] = 1$ ($i = 1, 2, 3$). This approach leads into a proper average behaviour for the returns Z_i in terms of the forward Oh model since $E[Z_i] = f_i(\bar{m}, \bar{ks})$ under the assumption of independence of X and Y.

From the set of equations (12) and using Bayes' theorem, an expression for the conditional ("posterior") probability of measuring m and ks given measurements of returns z_1, z_2 and z_3 is,

$$P(m, ks|z_1, z_2, z_3) = \frac{P_{Z_1 Z_2 Z_3}(z_1, z_2, z_3|m, ks) P_{MKS}(m, ks)}{P_{Z_1 Z_2 Z_3}(z_1, z_2, z_3)}, \qquad (13)$$

where $P_{Z_1 Z_2 Z_3}(z_1, z_2, z_3|m, ks)$ is the probability of measuring a certain set (z_1, z_2, z_3) of returns given measurements of m and ks (the "likelihood"), P_{MKS} is the prior joint density function of m and ks (where it is included all the a priori information about m and ks) and $P(z_1, z_2, z_3)$ works as a normalizing factor and it is the probability of a certain set (z_1, z_2, z_3) to be measured. Then, providing the conditional density function (13) is exact, the optimal unbiased estimator of m that has the minimum variance is the mean of (13) [25],

$$m_{est}^{Bayes} = \iint_D m P(m, ks|z_1, z_2, z_3) dks dm \qquad (14)$$

and similarly the standard deviation of this estimator will be:

$$m_{std}^2 = \iint_D (m - m_{est}^{Bayes})^2 P(m, ks|z_1, z_2, z_3) dks dm \qquad (15)$$

where an explicit expression for (13) must be found in order to calculate m_{est}^{Bayes} and m_{std}. The integration domain D in (14) and (15) spans the same range of (m, ks) where the forward Oh model was originally constrained, except for ks which is taken to be ≤ 3.5 as discussed in Section II-A. The standard deviation m_{std} can be used as a measure of the error of the estimate m_{est}^{Bayes}.

B. Derivation of the Likelihood

The posterior distribution $P(m, ks|z_1, z_2, z_3)$ in (13) can be computed as follows. First, using recursively the definition of conditional probability yields

$$P_{Z_1 Z_2 Z_3}(z_1, z_2, z_3|m, ks) = P_{Z_1}(z_1) P_{Z_2|Z_1 = z_1}(z_2) \times$$
$$\times P_{Z_3|Z_1 = z_1, Z_2 = z_2}(z_3) \qquad (16)$$

where in the right term the given m and ks were suppressed for simplicity. In (16), $P_{Z_1}(z_1)$ is calculated using the change

of variables theorem upon (12) ($i = 1$) and the assumption of independence between X and Y,

$$P_{Z_1}(z_1) = \int_0^\infty P_{X_1}(w) P_{Y_1}(\frac{z_1}{w}) \frac{1}{w} dw. \qquad (17)$$

In order to calculate the remaining two terms in (16), it might be noted that replacing m by M and ks for KS in (5) and (6) the following relationships concerning X_i hold

$$X_1 = \tilde{f}_1(M, KS) X_2 \qquad (18)$$

$$X_3 = \tilde{f}_3(M, KS) X_2 \qquad (19)$$

Replacing this set of equation in (12) and then equating for Z_2 and Z_3 one obtains

$$Z_2 = \frac{1}{\tilde{f}_1(M, KS)} \frac{Y_2}{Y_1} Z_1 \qquad (20)$$

$$Z_3 = \tilde{f}_3(M, KS) \frac{Y_3}{Y_2} Z_2 \qquad (21)$$

Finally, given m and ks and using again the change of variables theorem upon (20) and (21) separately, the remaining conditional probabilities are

$$P_{Z_2|Z_1=z_1}(z_2|m, ks) = \frac{\tilde{f}_1(m, ks)}{z_1} P_{\frac{Y_2}{Y_1}}(\frac{\tilde{f}_1(m, ks) z_2}{z_1}), \quad (22)$$

$$P_{Z_3|Z_1=z_1, Z_2=z_2}(z_3|m, ks) = \frac{1}{\tilde{f}_3(m, ks) z_2} P_{\frac{Y_3}{Y_2}}(\frac{z_3}{\tilde{f}_3(m, ks) z_2}), \qquad (23)$$

where $P_{\frac{Y_i}{Y_j}}$ ($i \neq j$) is the joint distribution of the ratio of two multilooked random variables which are affected by speckle. The likelihood in (16) is constructed then by multiplying (17), (22) and (23).

C. Modeling the Terrain Backscatter X

Natural variability of soil moisture are always present at different scales, in particular, at SAR systems scale [26]. In general, this implies that soil moisture inside a field cannot be considered constant; i.e. the field is heterogeneous in terms of soil moisture. Soil roughness can also be framed within this description. In agricultural fields, roughness is generated artificially by tilling and naturally by wind and water erosion. Moreover, soil surface roughness is very dependent on tillage operations and soil type [27].

In order to map the randomness in soil parameters to a randomness in soil backscattering and thus giving rise of the inter-pixel heterogeneity of both soil moisture and roughness, a forward model must be included. This mapping will be completely defined by the functions f_i from (1) that associates soil backscattering with soil parameters (i.e. the forward model). To compute this mapping, we will use a three-step procedure given in [28, §2.12]. Such a procedure allows to find the distribution P_R of a general function $r(u, v)$ which depends on two random variables U and V of known distribution. In our case, we are interested on the computation of the distribution of $x_1 = f_1(m, ks)$ used in (17) as P_{X_1} when the soil moisture m and roughness ks are considered random variables M and

KS, respectively. Ignoring the subscript 1, such computation states that

$$F_X(x) = \iint_{A_x} P_{MKS}(m, ks) dm dks, \qquad (24)$$

where $F_X(x)$ is the cumulative distribution function of the random variable X and the integration domain is $A_x = \{(m, ks) : f(m, ks) \leq x\}$. Then $P_X(x)$ is readily obtained by deriving (24) with respect to x. In what follows, it would be assumed that M and KS are uncorrelated and gaussian random variables, so that $P_{MKS} = P_M P_{KS}$ where $P_M \sim N(\bar{m}, \sigma_m)$ and $P_{KS} \sim N(\bar{ks}, \sigma_{ks})$. Therefore, the heterogeneity of the soil parameters within a (one-look) SAR pixel is controlled throughout the variance σ_m and σ_{ks}. The Gaussian assumption is not restricting or fundamental in any way, and the procedure can be also applied to different distributions for m and ks, even empirical ones. On the other hand, under this assumption the computation of (24) can be only performed numerically.

D. Modeling the Speckle Noise Y

Statistical properties of multi-look polarimetric data are quite different from those of single-look data [8]. Therefore, in order to model the expected speckle phenomena, we need to know the probability density function of polarimetric data as a function of the number of looks n. In the case of multi-look intensity values, the corresponding distribution P_Y is that of (10) and is used in (17). On the other hand, the probability density function of the ratio of two multi-look polarimetric data P_{Y_i/Y_j} which are not independent are required in (22) and (23). Such a distribution was derived by Lee *et al.* [8]:

$$P_U(u) = \frac{\Gamma(2n)}{\Gamma(n)\Gamma(n)} \frac{\tau^n (1 - |\rho_u|^2)^n (\tau + u) u^{n-1}}{[(\tau + u)^2 - 4\tau|\rho_u|^2 u]^{n+1/2}} \qquad (25)$$

where $U = \frac{Y_i}{Y_j}$ ($i \neq j$), n is the number of looks, ρ_c is the correlation between the numerator and the denominator and $\tau = \frac{E[Y_i]}{E[Y_j]}$ is the ratio of the expected value of Y_i and Y_j. In order to the expected value of the returns to be determined only by the expected value of the forward model, we stated that $E[Y_i] = 1$ ($i = 1, 2, 3$) and then $\tau = 1$. Thus the expected value of Z is determined only by X as follows from (11). The ratio distribution also depends on the correlation between the numerator and the denominator ρ_u. This is very important, since when numerator/denominator correlation increases, the variance of the distribution decreases [8]. As expected, when n increases the distribution becomes narrower and thus the variance of the estimates decreases, leading to a more precise retrieving. Up to this point, we presented all the mathemathics necessary for a Bayesian retrieval scheme. In the next sections we present the results of numerical simulations.

IV. NUMERICAL RESULTS

A. Minimization Estimate

Since Oh model is not directly invertible, he [15] established an algorithm for retrieving soil moisture and roughness from a set of measured backscattering coefficients hh, vv and vh through a minimization procedure. Such a procedure is based on the simultaneous solution of model equations (2), (3) and

(4), leading to the following non-linear expression [15, eq. (6)],

$$1 - (\frac{\theta}{90°})^{0.35m^{-0.65}} exp(-0.4(ks(\theta, m, vh))^{1.4}) - hh/vv = 0$$

$$(26)$$

where $ks(\theta, m, vh)$ is directly obtained after solving (2). For a given (hh, vv, vh), the estimated value of m is the one that minimizes this expression, namely m_{est}^{Oh}. It is important to note that (26) can be solved only for the values of (hh, vv, vh) that are allowed by the forward model, specifically those values that lie within the region bounded by inequations (7), (8) and (9). This means that this approach is not robust to high statistical fluctuations in the backscattering coefficients, that are commonly found in real applications.

Assuming a certain value for vh ($vh = -25dB$) and $\theta = 35°$, when applying to the entire (hh, vv)-space a root-finding procedure applied on (26) gives rise to the contour lines depicted on Fig. 1. Although the levels of the contour lines spans the entire range of Oh model ($0.04 - 0.291$ cm^3/cm^3), only the levels corresponding to 0.05, 0.10, 0.15, 0.20 and 0.25 (in units of cm^3/cm^3) are drawn. The linear trend of the contour lines is consistent with the fact that at fixed vh, the dynamical range of the minimization estimates from (26) is governed by the ratio hh/vv, which takes constant values over lines in the entire (hh, vv)-space. To corroborate the inversion, the exact values of m were computed using the deterministic forward Oh model (1), constrained to the assumption that $vh = -25dB$ ('+' marks in Fig. 1). The levels of the exact values agree with those of the minimization estimates.
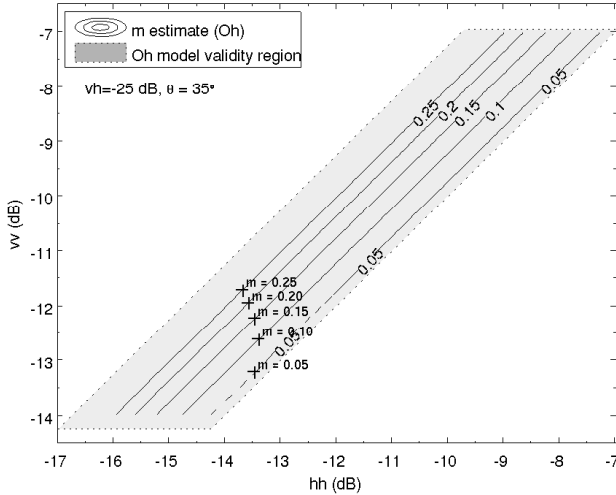
Fig. 1. Soil moisture m_{est}^{Oh} estimated on the (hh,vv)-plane (at fixed $vh = -25dB$) from Oh model, in units of cm^3/cm^3. The light gray area encloses the pairs (hh,vv) where the model is valid.

Every value of (hh, vv, vh) yield in a value of m_{est}^{Oh} inside Oh model validity region, as expected, whereas for the values of (hh, vv, vh) lying outside that region the inversion technique cannot produce a retrieval. The latter situation could be related to landcover uncertainties (i.e. the target is not completely bare soil), speckle noise and/or system fluctuations. In an operational implementation, the spurious

estimations related to the landcover can be reduced using ancillary information about landcover status. Nevertheless, it is important to remark that even bare soil can produce values of (hh, vv, vh) outside the Oh model validity region, due to speckle and system fluctuations.

The estimation procedure from (26) produce a single value of m_{est}^{Oh} given a set of measured values (hh, vv, vh). No ancillary information about soil status (previous or estimated by other means) is allowed. Moreover, it is implicitly assumed that image radiometric uncertainties are very low, since small fluctuation of measured values can produce strong variations in soil moisture estimation. Therefore, in order to successfully use this kind of retrieval, a speckle reduction technique is mandatory.

B. Bayesian Estimate

An alternative method for the estimation of m, which is suitable for taking into account the speckle, arises when using the expressions (14) and (15). In order to test the goodness of the Bayesian approach, an uniform prior is used as P_{MKS} in (13). This kind of prior represent no knowledge about soil condition. Specifically, it is taken $P_M \sim U(0.04, 0.35)$ and $P_{KS} \sim U(0.13, 3.5)$ as reasonable priors.

Fig. 2 shows a contour plot of the estimate m_{est}^{Bayes} for soil moisture, as a function of the measured values of hh and vv with $n = 3$, for $vh = -25dB$ and $\theta = 35°$. The light shaded area represents Oh model validity region, where the contour lines of soil moisture derived from the Oh model are also shown. The remaining model parameters are $\sigma_m = 0.005$ cm^3/cm^3, $\sigma_{ks} = 0.01$, $\rho_{vv/hh} = 0.7$ and $\rho_{vh/vv} = 0.1$. When using the Bayesian methodology, the retrieved soil moisture values cover the entire (hh,vv,vh)-space, although the extreme values (the ones that are far away from Oh model validity region (shaded area)) will present a very low probability of occurence associated. The high spread showed by the contour lines is consistent with a high speckle noise for this small number of looks ($n = 3$).

In Fig. 2, the results of both estimations (minimization and Bayesian) are compared. It is readily seen that m_{est}^{Oh} and m_{est}^{Bayes} do not coincide. Since the prior is uniform, this discrepancy is related to the chosen values of model parameters σ_m, σ_{ks} and n. The election $\sigma_m = 0.005$ cm^3/cm^3, $\sigma_{ks} = 0.01$ corresponds to a very homogeneous soil, which corresponds to low variance is the soil backscattering X. However, $n = 3$ corresponds to a high variance in the speckle Y, which ultimate leads to a poor soil moisture estimation. This statement is reflected in the contour lines of one-sigma standard deviation of m_{est}^{Bayes} depicted on Fig. 3 and calculated by means of (15). The standard deviation reaches a relative high value (about 2/5 of the dynamic range for soil moisture) of ~ 0.07 cm^3/cm^3 in everywhere.

Fig. 4 shows the contour lines retrieved after increasing the number of looks to $n = 256$. When significant multilooking is present, the Bayesian retrieval looks more compact around the contour lines of Oh model indicating, to some extent, a correct asymptotical behaviour. It could be seen that the '+' marks and the Bayes' contour lines agrees, especially for the
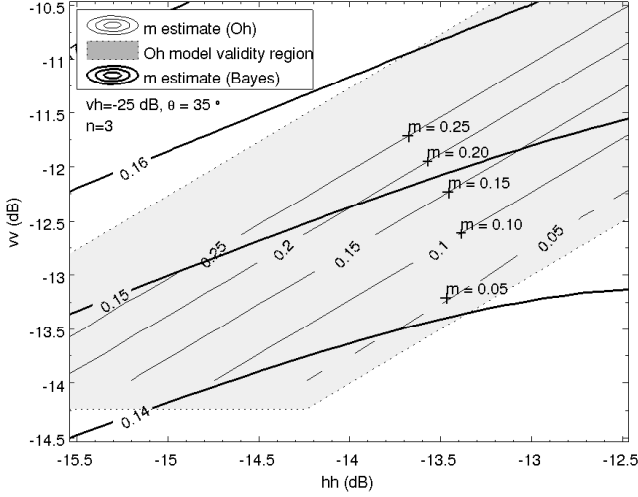
Fig. 2. Comparison between the soil moisture estimated using Oh model and the Bayesian retrieval approach, in units of cm^3/cm^3. The parameters adopted by the simulation are: $n = 3$, $\sigma_m = 0.005\ cm^3/cm^3$, $\sigma_{ks} = 0.01$, $\rho_{vv/hh} = 0.7$ and $\rho_{vh/vv} = 0.1$.
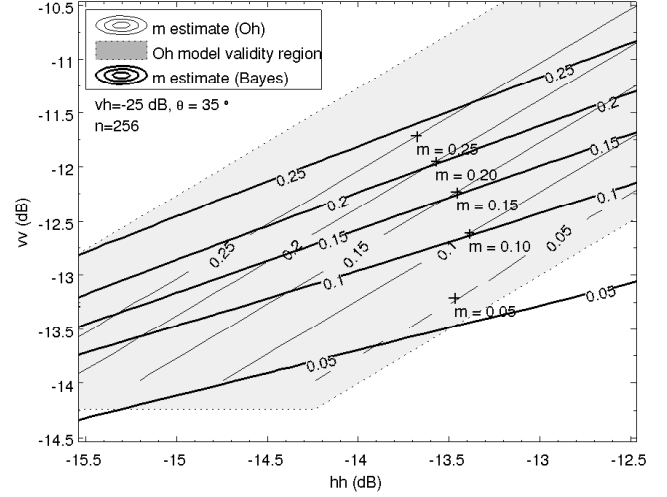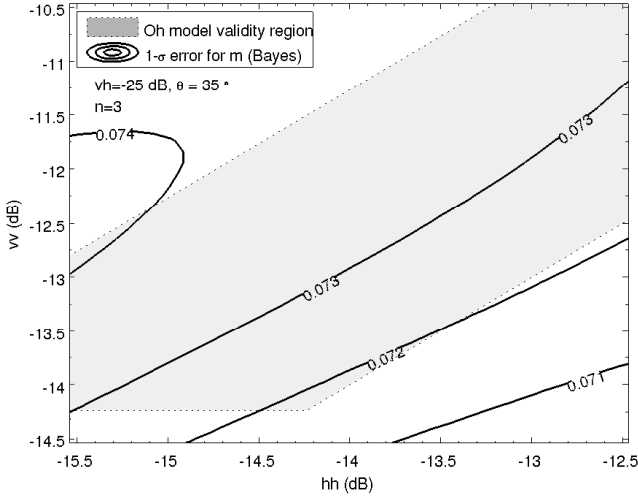

Fig. 4. Comparison between the soil moisture estimated using Oh model and the Bayesian retrieval approach, in units of cm^3/cm^3. The parameters adopted by the simulation are: $n = 256$, $\sigma_m = 0.005\ cm^3/cm^3$, $\sigma_{ks} = 0.01$, $\rho_{vv/hh} = 0.7$ and $\rho_{vh/vv} = 0.1$.

Fig. 3. One-sigma standard deviation of m_{est}^{Bayes} for a number of looks $n = 3$, in units of cm^3/cm^3. The parameters adopted by the simulation are: $\sigma_m = 0.005\ cm^3/cm^3$, $\sigma_{ks} = 0.01$, $\rho_{vv/hh} = 0.7$ and $\rho_{vh/vv} = 0.1$.
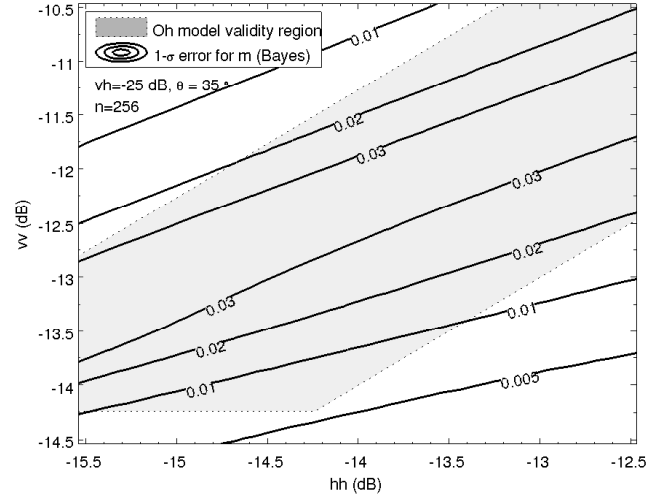
Fig. 5. One-sigma standard deviation of m_{est}^{Bayes} for a number of looks $n = 256$, in units of cm^3/cm^3. The parameters adopted by the simulation are: $\sigma_m = 0.005\ cm^3/cm^3$, $\sigma_{ks} = 0.01$, $\rho_{vv/hh} = 0.7$ and $\rho_{vh/vv} = 0.1$.

C. Including A Priori Information

If a priori information is on hand, the Bayesian retrieval scheme can include it straightforwardly. A priori information can be available from historical records, estimations from other sensors, in situ data and/or contextual information about soil texture/use. Using this information, suitable distributions for the prior distributions of soil moisture and roughness can be estimated.

As an example, we now assumed that the prior distribution for soil roughness in the study area is Gaussian distributed $N(\mu_{ks}, \sigma'_{ks})$ and we will assess the performance of the retrieval as a funcion of the number of looks. We start using a (hh, vv, vh) simulated from $m = 0.20\ cm^3/cm^3$ and

levels of 0.10, 0.15 and 0.20 in units of cm^3/cm^3. Of course, since the minimization and Bayesian estimators are different, an overlap of the contourn lines are not expected. In the same way, Fig. 5 depicts the contour lines of one-sigma standard deviation of m_{est}^{Bayes} calculated by means of (15) for $n = 256$. In this case, the standard deviation ranges between a minimum of $\sim 0.005\ cm^3/cm^3$ and reaches a maximum value of $\sim 0.03\ cm^3/cm^3$. The relative improvement regarding the case showed in Fig. 3 is due to the increasing of the number of looks, which is a way to reduce the uncertainties due speckle in soil moisture estimation.

$ks = 0.66$ through the functions f_i (eqs. (2-4)). In the following paragraph, the behaviour of the retrieval when the precision (spread) σ'_{ks} of the prior varies will be analyzed.

Fig. 6 depicted the estimated m for the Bayesian retrieval using uniform and Gaussian distributions as priors for soil roughness, where the latter distributions are centered at the true value $ks = 0.66$ and the precision takes values of 0.05, 0.1 and 0.25. Uniform prior $U(0.04, 0.35)$ is used for soil moisture. For number of looks $n > 300$, all the estimates tend to the true $m = 0.20 \ cm^3/cm^3$ within the $0.005 \ cm^3/cm^3$ value, which is the intrinsic heterogeneity of the soil given by σ_m. This true value is also the estimated m derivated from Oh model (m_{est}^{Oh}), which it does not depend on n since Oh model does not take into account speckle. As expected, the retrieval schemes weights the likelihood using the prior, and different rates of convergency are reached. However, two regions are readily determined. On one hand, a region for large n, where it is observed that the retrieval with uniform prior converges faster than the case when gaussian prior is used. On the other hand, a precise prior is preferable for low n ($n < 50$), where it is observed that $N(0.66, 0.05)$ approachs to the true $m = 0.20 \ cm^3/cm^3$ value faster (i.e. with a higher slope) than an imprecise one ($N(0.66, 0.25)$) and even faster than the case when uniform prior $U(0.13, 3.5)$ is used. In other words, when variance from speckle is significant (low values of n), a precise prior improves the retrieval by strongly restricting the possible values of m, whereas for large n any prior performs equally well, specially the uniform one. For $n > 300$, the error m_{std} is less than $0.03 \ cm^3/cm^3$, where for $n < 50$ the error is about $0.06 - 0.07 \ cm^3/cm^3$.
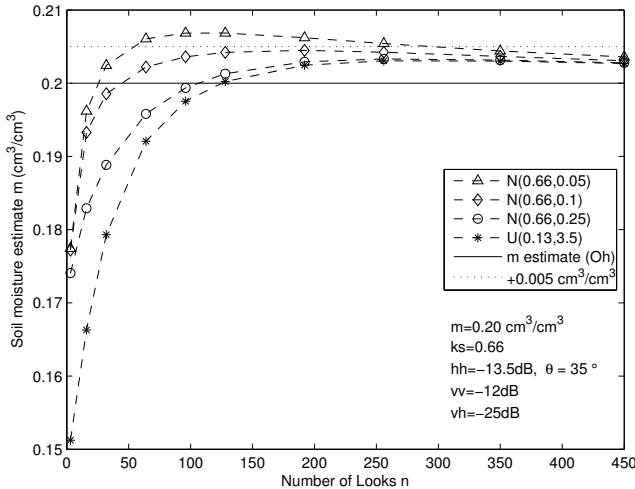
Fig. 6. Bayesian retrievals of soil moisture using Gaussian and uniform distributions as prior information for soil roughness, as a function of the number of looks n, in units of cm^3/cm^3. Uniform prior $U(0.04, 0.35)$ is used for soil moisture. The true value from Oh retrieval is shown along with the σ_m-line, thus indicating the minimum uncertainty that every retrieval might have. The parameters adopted by the simulation are: $\sigma_m = 0.005 \ cm^3/cm^3$, $\sigma_{ks} = 0.01$, $\rho_{vv/hh} = 0.7$ and $\rho_{vh/vv} = 0.1$.

V. CONCLUSION

Surface soil moisture estimation from SAR data is a complex task. This is related to many issues, but the spatiotemporal dynamics of soil moisture and the low dynamic ranges of soil backscatter involved are among the most important ones. Solutions to this complex problem should include better and more tested forward and inverse models. However, it is important to understand that inverse models should address in some way the two phenomena that most degrade the retrieval: speckle and soil spatial heterogeneities. In order to address these issues, a Bayesian methodology has been proposed.

In this methodology, a model for the soil backscattering and a model for the speckle are combined using the framework of the multiplicative model and Bayes' theorem. Therefore, this methodology is able to take into account terrain features as well as speckle noise to achieve a robust retrieval of soil parameters from SAR data. This Bayesian methodology: (1) needs only a forward model (no retrieval model is required), (2) gives an estimation of soil parameters as well as their associated error, (3) can include as many error sources as necessary, and (4) can include a priori information in a systematic way.

To illustrate the retrieval scheme, a simplified formulation of Oh's model was used throughout this work. Furthermore, the speckle was modeled using appropriate distributions. Using reasonable hypothesis about functions and model parameters, the retrieval scheme was tested in different scenarios using numerical simulations.

For any soil condition, when the number of looks n is low and uniform priors for soil parameters are used, the retrieval errors are large. However, when significant multilooking ($n = 256$) is present, the retrieval error decreases. The relative improvement due to the increasing of n is displayed by the one-sigma contour lines, where error decreases from $\sim 0.07 \ cm^3/cm^3$ to $\sim 0.03 \ cm^3/cm^3$.

The effect on the retrieval of different prior distributions was also studied. Comparing Gaussian and uniform priors gives rise to two well-defined behaviour for the m estimates in terms of the number of looks n. For large n ($n > 300$), a uniform prior works well as a Gaussian one (i.e. convergency is assured within the intrinsic variance of the soil roughness). For low values of n ($n < 50$), a precise prior (i.e. $\sigma'_{ks} = 0.05$) determines a rate of convergence higher than an imprecise one (i.e. $\sigma'_{ks} = 0.25$).

In summary, the proposed soil moisture retrieval scheme takes as inputs the measured soil backscattering coefficients, soil ancillary parameters and the number of looks, among others. Soil ancillary parameters are related to the expected distribution of soil parameters within SAR pixel. So defined, soil moisture estimation converges to the expected behaviour when $\sigma_m \to 0$, $\sigma_{ks} \to 0$ and $n \to \infty$, so confirming that the standard Oh's model regime is reached.

Due to its construction, the model presented here is able to study different retrieval schemes for different kinds of soils and/or different soil moisture spatial distributions. Furthermore, since soil variance increases with scale, multilooking will reduce speckle variance but increase observed soil parameters variance (σ_m, σ_{ks}), thus ultimately degrading the