A BAYESIAN METHODOLOGY FOR SOIL PARAMETERS RETRIEVAL FROM SAR IMAGES

M. Barber, P. Perna, C. Bruscantinni, F. Grings, H. Karszenbaum*

Grupo de Teledetección Cuantitativa,
Instituto de Astronomía y Física del Espacio (IAFE),
Argentina

M. Piscitelli

Fac. de Agronomía,
Univ. Nac. del Centro
de la Pcia. de Buenos Aires
(UNICEN), Argentina

J. Jacobo-Berlles

Depto. de Computación, Fac. de Cs. Exactas y Naturales (FCEN),
Univ. de Buenos Aires (UBA), Argentina.

ABSTRACT

Soil moisture retrieval from SAR data presents two main sources of uncertainty: terrain heterogeneity and speckle noise. In this paper, these issues will be addressed by using a Bayesian approach. Such a Bayesian approach (1) needs only a forward model (no retrieval model required), (2) gives the optimal unbiased estimator for the soil moisture and its error and (3) can include as many error sources as required. Through numerical simulations, a standard Oh retrieval procedure and the Bayesian approach were tested for different number of looks ($n = 3$ and $n = 64$). The results indicate that for a large number of looks the region of validity of both approaches are similar. Furthermore, contrary to the Oh model retrieval procedure which is only valid in a bounded region of the (hh, vv, hv)-space, the Bayesian approach gives an estimation of soil moisture and its error for any combination of hh, vv and hv, so enlarging the region where the retrieval is possible.

***Index Terms*—** Soil moisture, Bayesian retrieval approaches, radar remote sensing

1. INTRODUCTION

Orbiting microwave synthetic aperture radar (SAR) systems offer the opportunity of monitoring soil moisture content (SMC) at different scales and under any kind of weather conditions, through the known sensibility that the backscattered signal exhibits to soil parameters, including soil moisture. In this framework, soil moisture retrieval can be considered an inference problem, where one essentially wants to infer soil condition given a set of measured backscatter coefficients and ancillary information.

A wide range of forward models, ranging from experimental relationships to physically-based models have been developed in order to assess the dependency of soil parameters to the backscattered signal. These models are important to understand the soil backscattering physics, but they are also a key tool to the retrieval of soil condition from SAR measurements.

One of the limiting problems of SAR-based soil moisture retrieval is the unsatisfactory performance of retrieval models. Reasons for mismatches between model parameters and measured data include instrument error, the heterogeneity of the target's surface, the difficult to measure in the field the roughness parameters input to the models [1][2][3]. In the case of semi-empirical models, the standard modeling approach based on scatterometers is to describe the average behaviour of the signal as a function of soil parameters, disregarding the spread around the average value and its causes [3]. This is true for all the semi-empirical models published, and give rise to artifacts characterized by several soil moisture estimates that correspond to the same soil moisture measurement [3].

Furthermore, another phenomenon degrades SAR-based soil moisture retrieval: speckle noise. It is a multiplicative noise that leads to a grain-like appearance of SAR images that decreases their contrast and therefore their quality [4]. It is characteristic of SAR images, and it is usually reduced in a post-processing stage by averaging neighboring pixels (multi-looking process) at the expense of spatial resolution. Nevertheless, averaging implicitly assumes that soil properties inside the average window are constant, which is usually not the case. Therefore, a tradeoff between multilooking and soil heterogeneity is usually accepted.

In the classical approaches aforementioned, the retrieval model and the speckle noise are considered as independent problems, whereas they are indeed part of the same retrieval problem. In this investigation, we analyze a Bayesian retrieval methodology which incorporates in a natural way the speckle and the terrain heterogeneity as a source of uncertainty that degrades the output value predicted by the forward models. This approach will allow us to investigate the total uncertainty in estimated soil moisture (associated both to terrain hetergeneity and speckle) as a function of the number of looks. In this paper, only the effect of speckle is being considered. Such a methodology will be presented using a simplified version of the Oh model [5] as the forward model.

2. RETRIEVAL APPROACH

2.1. Oh Model

The most widely accepted semi-empirical soil scattering model is the one developed by Oh [5], where model expressions are physically-based, but model parameters are derived from an extense database of polarimetric radar scatterometer measurements. Such a model relates backscattering returns and certain soil properties through a set

*Corresponding author: mbarber@iafe.uba.ar

of three analitycal functions f_i, and can be symbolically expressed as [5, eqs. (1),(2) and (4)],

$$z_i = f_i(m, ks) \qquad (i = 1, 2, 3), \tag{1}$$

where z_i is the backscattering (measured) return and the subscript $i = 1, 2, 3$ stands respectively for the hh, vv and hv-polarizations. The returns z_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 incidence angle, which is a known parameter. The Oh model is constrained to the range $0.04 \leq m \leq 0.297$ and $0.13 \leq ks \leq 6.98$, although the later has a better agreement between the model and the experimental results for $ks \leq 3.5$ [5].

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

$$f_1 = \tilde{f}_1(m, ks) f_2 \tag{2}$$

$$f_3 = \tilde{f}_3(m, ks) f_2, \tag{3}$$

Through a minimization procedure, Oh established an algorithm for retrieving soil moisture and roughness from a set of measured returns z_1, z_2 and z_3 (hh, vv and hv respectively). Assuming a certain value for $z_3 = -25dB$ (hv), when applying to the entire z_1, z_2-space such a procedure gives rise to the contour lines depicted on Fig. 1.

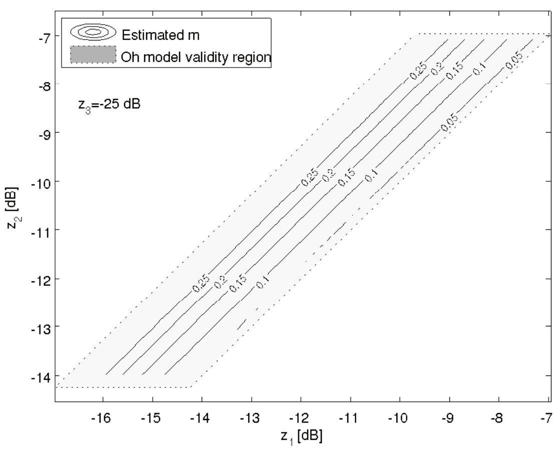

Fig. 1. Soil moisture m estimated on the (z_1,z_2)-plane (at fixed $z_3 = -25dB$) from Oh model. The light gray area encloses the pairs (z_1,z_2) where the model is valid.

2.2. Bayesian Approach

The deterministic forward model developed by Oh can be extended to a stochastic model following [3]. In doing so, we can include in the forward model both the terrain heterogeneity and speckle. The model that naturally incorporates both the terrain backscattering and the speckle is the multiplicative model,

$$Z_i = X_i Y_i \qquad (i = 1, 2, 3), \tag{4}$$

where Z_i is the random variable which represent the return z_i and again the subscript i stands for the different polarizations. 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 $X_i = f_i(M, KS)$, where f_i represents 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 and ks 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 of (m, 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. 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 of the returns Z_i in terms of the Oh's forward model since $E[Z_i] = f_i$ under the assumption of independence of X and Y.

From the set of equations (4) 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,

$$P(m, ks|z_1, z_2, z_3) = \frac{P_{Z_1Z_2Z_3}(z_1, z_2, z_3|m, ks) P_{MKS}(m, ks)}{P_{Z_1Z_2Z_3}(z_1, z_2, z_3)}, \tag{5}$$

where $P_{Z_1Z_2Z_3}(z_1, z_2, z_3|m, ks)$ 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 joint density function is exact, the optimal unbiased estimator of m that has the minimum variance is the mean of (5) [6],

$$m_{mean} = \iint_D m P(m, ks|z_1, z_2, z_3) dksdm \tag{6}$$

and similarly the standard deviation of this estimator will be:

$$m_{std} = \iint_D (m - m_{mean})^2 P(m, ks|z_1, z_2, z_3) dksdm \tag{7}$$

where an explicit expression for (5) must be found in order to calculate m_{mean} and m_{std}. The integration domain D in (6) and (7) spans the same range of (m, ks) where the model was originally constrained, except for ks which is taken to be ≤ 3.5.

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

$$P_{Z_1Z_2Z_3}(z_1, z_2, z_3|m, ks) = P_{Z_1}(z_1) P_{Z_2|Z_1=z_1}(z_2) \times \\ \times P_{Z_3|Z_1=z_1, Z_2=z_2}(z_3) \tag{8}$$

where in the right term the given m and ks were suppressed for simplicity. In (8), $P_{Z_1}(z_1)$ is calculated using the change of variables theorem upon (4) ($i = 1$) and the assumption of independence between X and Y,

$$P_{Z_1}(z_1) = \int_0^{\infty} P_{X_1}(w) P_{Y_1}(\frac{z_1}{w}) \frac{1}{w} dw. \tag{9}$$

In order to calculate the remaining two terms in (8), it might be noted that replacing m by M and ks for KS in (2) and (3) the following relationships concerning X_i hold

$$X_1 = \tilde{f}_1(M, KS)X_2 \tag{10}$$

$$X_3 = \tilde{f}_3(M, KS)X_2 \tag{11}$$

Replacing this set of equation in (4) and then equating for Z_2 and Z_3 one obtains

$$Z_2 = \frac{1}{\tilde{f}_1(M, KS)} \frac{Y_2}{Y_1} Z_1 \tag{12}$$

$$Z_3 = \tilde{f}_3(M, KS) \frac{Y_3}{Y_2} Z_2 \tag{13}$$

Finally, using again the change of variables theorem upon (12) and (13) the given m and ks the remaining conditional probabilities are

$$P_{Z_2|Z_1=z_1}(z_2|m, ks) = \frac{\tilde{f}_1(m, ks)}{z_1} P_{\frac{Y_2}{Y_1}}\left(\frac{\tilde{f}_1(m, ks)z_2}{z_1}\right), \tag{14}$$

$$P_{Z_3|Z_1=z_1,Z_2=z_2}(z_3|m, ks) = \frac{1}{\tilde{f}_3(m, ks)z_2} P_{\frac{Y_3}{Y_2}}\left(\frac{z_3}{\tilde{f}_3(m, ks)z_2}\right), \tag{15}$$

where $P_{\frac{Y_i}{Y_j}}$ $(i \neq j)$ is the joint distribution which corresponds to the ratio of two multilooked random variables which are affected by speckle.

3. MATHEMATICAL MODELING

From [4], the probability density function of a single multilooked polarization sample measured in intesity is:

$$P_Y(y) = \frac{n^n}{\Gamma(n)} y^{n-1} e^{-ny} \tag{16}$$

where n is the equivalent number of looks. Furthermore, the PDF of the ratio of two multilooked samples measured in intensity was derived by [refLeeCocientes] in terms of a Kibble's bivariate gamma:

$$P_U(u) = \frac{\Gamma(2n)}{\Gamma(n)\Gamma(n)} \frac{\tau^n(1 - |\rho_c|^2)^n(\tau + u)u^{n-1}}{[(\tau + u)^2 - 4\tau|\rho_c|^2 u]^{n+1/2}} \tag{17}$$

where $U = \frac{Y_i}{Y_j}$ $(i \neq j)$, 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 mean value of Y_i and Y_j. As expected, both these density functions depend on n: when n increases, the distributions becomes narrower. As we assumed that speckle only affects with an multiplicative noise, then $\tau = 1$. The ratio distribution also depends on the correlation between the numerator and the denominator ρ_c. This is very important, since when numerator/denominator correlation increases, the variance of the distribution decreases [4].

The distribution P_X is derived using the forward model $f(m, ks)$ and assuming a prior joint distribution for the random variable M and KS, i.e. $P_{M,KS}$. The receipt is given in [7],

$$F_X(x) = \iint_{A_x} P_{M,KS}(m, ks)\, dm\, dks, \tag{18}$$

where the $F_X(x)$ 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 obtained by deriving eq. 18 with respect to x. In what follows, it would be assumed that M and KS are uncorrelated and gaussian random variables, so that $P_{M,KS} = P_M P_{KS}$ where $P_M \sim N(\mu_m, \sigma_m)$ and $P_{KS} \sim N(\mu_{ks}, \sigma_{ks})$.

Up to this point, we presented all the mathematics necessary for a Bayesian retrieval scheme.

4. NUMERICAL SIMULATION

We are now able to perform a comparison between the retrieval approachs discussed so far: Oh model and Bayesian model. In order to test the goodness of the approach, an uniform prior is used. This kind of prior represent no knowledge about soil condition. M is an uniform $U[0.01 - 0.35]$ and KS is an uniform $U[0.1 - 4.0]$. Fig. 2 shows the isolines of soil roughness with $n = 3$, for $z_3 = -25dB$. The light shadow area represents the validity region of the Oh Model, where the contour lines of soil moisture derived from the Oh Model are also shown.

When using the Bayesian methodology, the retrieved soil moisture values cover the entire (hh,vv,hv)-space, although the extreme value (the ones that are far away from Oh model validity region (shadow area)) will present a very low probability associated. The high spread showed by the contour lines is consistent with a high speckle noise for these small number of looks ($n = 3$).

Fig. 3 shows the contour lines retrieved after increasing the number of looks to $n = 64$. 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 for low values of z_1 and z_2 (hh and vv respectively), the contour lines of both retrieval approaches are convergent.

In the same way, the contour lines of one-sigma standard deviation of the estimated m can be calculated by means of eq. 7. The results are not shown.

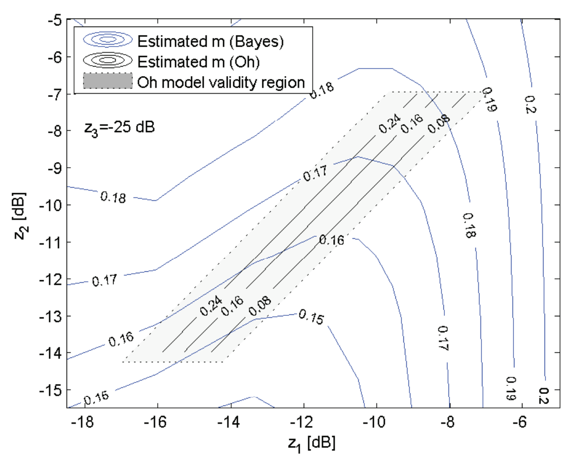

Fig. 2. Comparison between the soil moisture estimated using Oh model and the Bayesian retrieval approach. The values adopted by the simulation are: $n = 3$, $\sigma_m = 0.005$, $\sigma_{ks} = 0.01$.

Fig. 4 depicts the asymptotic behaviour of estimated m as n increases from 3 to 128 for $z_1 = -13dB$, $z_2 = -12dB$, $z_3 = -24dB$ and $\sigma_m = 0.005$, $\sigma_{ks} = 0.01$. The error bars correspond to the one-sigma error calculated by (7). The dotted line is the estimated m derivated from Oh model, which it does not depend on n since Oh model does not take into account speckle. At $n = 128$, the m estimate differs in about 2% (absolute error) with respect to the limiting value given by the Oh model.

This remaining bias is related to the terrain heterogeneity (σ_m, σ_{ks}), which contributes to the total soil moisture error even when n is large.

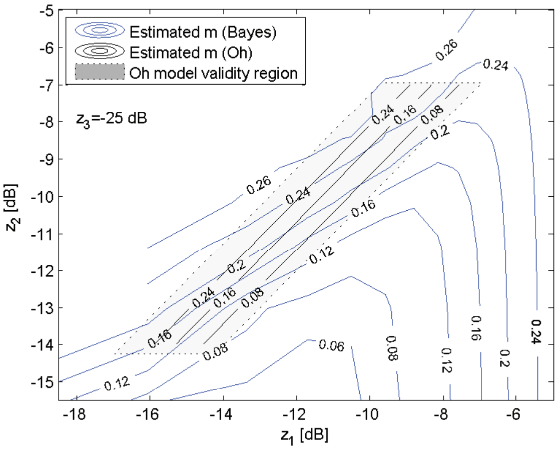

Fig. 3. Comparison between the soil moisture estimated using Oh model and the Bayesian retrieval approach. The values adopted by the simulation are: $n = 64$, $\sigma_m = 0.005$, $\sigma_{ks} = 0.01$.

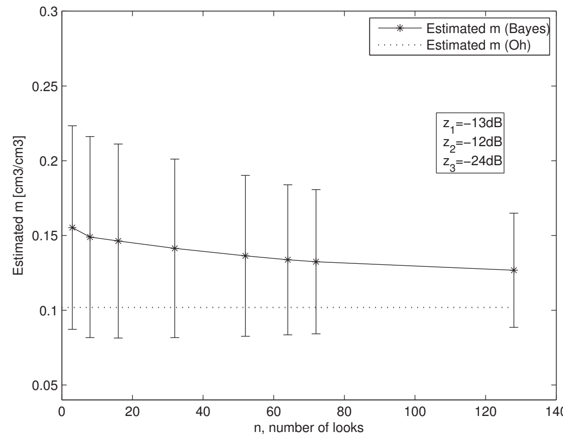

Fig. 4. Asymptotic behaviour of estimated m, as n increases from 3 to 128 ($\sigma_m = 0.01$, $\sigma_{ks} = 0.01$). The dotted line is the estimated m that corresponds to the Oh model.

5. CONCLUSIONS

Soil moisture retrieval from SAR data presents two main sources of uncertainty: terrain heterogeneity and speckle noise. These issues are being addressed by using a Bayesian approach, that rigorously models these two phenomena. Such a Bayesian approach: (1) needs only a forward model (no retrieval model required), (2) gives the optimal unbiased estimator for the soil moisture and its error and (3) can include as many error sources as required. The Bayesian was tested for different number of looks ($n = 3$ and $n = 64$). One of the expected results is that for a large number of looks the contour lines of both the Bayesian method and the standard Oh model become closer. Furthermore, contrary to the Oh model retrieval procedure, which is only valid in a bounded region of the (hh, vv, hv)-space, the Bayesian approach gives an estimation of soil moisture for any conbination of hh, vv and hv. Bayesian model also shows that, for those (hh, vv, hv) values that are far from the region of validity of the Oh model, the probability associated to the estimation is very low. Therefore, the Bayesian approach has the possibility of enlarging the region of (hh, vv, hv) space where the retrieval is possible, also providing the quality of this estimation. In addition, since every Bayesian approach includes a prior, these information can be used to avoid confusion related to landcover types. Work in progress is also addressing a functional relation between number of looks and terrain heterogeneity, in order to determine the best multilooking strategy.

6. REFERENCES

[1] N. E. C. Verhoest, H. Lievens, W. Wagner, J. Alvarez-Mozos, M. S. Moran, and F. Mattia, "On the soil roughness parameterization problem in soil moisture retrieval of bare surfaces from synthetic aperture radar," *Sensors*, vol. 8, pp. 4213–4248, Jul. 2008.

[2] M. Callens, N. E. C. Verhoest, and M. W. J. Davidson, "Parameterization of tillage-induced single-scale soil roughness from 4-m profiles," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 4, pp. 878–888, 2006.

[3] Z. S. Haddad, P. D. Dubois, and J. J. van Zyl, "Bayesian estimation of soil parameters from radar backscatter data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 34, no. 1, pp. 76–82, Jan. 1996.

[4] J. S. Lee, K. W. Hoppel, S. A. Mango, and A. R. Miller, "Intensity and phase statistics of multilook polarimetric and interferometric sar imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 32, no. 5, pp. 1017–1028, Sep. 1994.

[5] Y. Oh, "Quantitative retrieval of soil moisture content and surface roughness from multipolarized radar observations of bare soil surfaces," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 3, pp. 596 – 601, Mar. 2004.

[6] D. J. C. MacKay, *Information Theory, Inference, and Learning Algorithms*, Cambridge University Press, 2003.

[7] L. Wasserman, *All of Statistics: A Concise Course in Statistical Inference*, Springer Texts in Statistics. Springer, 2004.