A BAYESIAN APPROACH FOR A SAC-D/AQUARIUS SOIL MOISTURE PRODUCT

C. A. Bruscantini, Student Member, IEEE, F. Grings, M. Barber, P. Perna, H. Karszenbaum*

Institute of Astronomy and Space Physics (IAFE), Argentina

ABSTRACT

In this work, several retrieval algorithms were implemented to retrieve soil moisture (sm) and optical depth (τ) from Aquarius/SAC-D observations. Currently used sm retrieval algorithms (H- and V-pol Single Channel Algorithm, SCAH and SCAV; Microwave Polarization Difference Algorithm, MPDA) were computed over Pampas Plains, Argentina. The methodology of a novel Bayesian algorithm developed is also presented, and its results are contrasted with the previous algorithms. Finally, performance metrics for each algorithms were derived using SMOS Level-2 sm and τ as benchmark products. The new Bayesian approach provide the sm retrieval algorithm that exhibited the lowest ubRMSE ($0.115\, m^3/m^3$), though very close to USDA SCA and SCAV ubRMSE ($0.116\, m^3/m^3$). Nevertheless, some improvements are discussed in Section 4 that might increase significantly the Bayesian algorithm performance.

***Index Terms*—** Aquarius; soil moisture; Bayesian inference; Markov Chain Monte Carlo.

1. INTRODUCTION

The SAC-D/Aquarius (launched on June 2011) is a cooperative international mission between CONAE (Comisión Nacional de Actividades Espaciales), Argentina, and NASA, USA. Its primary goal is to monitor weekly global sea surface salinity to help understanding both climate change and the global water cycle [1]. The Aquarius is an integrated L-band radiometer (1.413 GHz) and scatterometer (1.26 GHz). In this paper, land Aquarius observations were used for monitoring soil moisture over the Pampas Plains region in Argentina.

Several retrieval algorithms were developed to retrieve soil moisture from passive remote sensing data. The most commonly used are the Single Channel Algorithm (SCA), the Dual Channel Algorithm (DCA) and the Land Parameter Retrieval Model (LPRM). All these algorithms rely on the omega-tao model to link brightness temperature (Tb) and surface dielectric and geometric properties, and differ among them on the polarization channels they use and the minimization scheme implemented [2]. LPRM and DCA make use of H-pol Tb (TbH) and V-pol Tb (TbV) to retrieve soil moisture and optical depth. One disadvantage of both previous algorithms is their sensitivity to noise in both TbH and TbV channels (specially uncorrelated noise between channels). On the other hand, SCAH (SCAV) uses only TbH (TbV) to retrieve soil moisture using optical depth as an auxiliary input to the retrieval algorithm (usually derived from an optical proxy). The main disadvantage of relying on optical depth to retrieve soil moisture is that if optical depth is not well known, SCA will have poor performance. In practice, accurate knowledge of optical depth is tricky. In general, optical depth is obtained through the vegetation parameter b (a land cover dependent parameter, empirically derived, not unique values found on literature) and vegetation water content, VWC (derived from different proxies and models that result in different VWC values). All these retrieval implementations also need other ancillary parameters as necessary auxiliary inputs. In this paper, a novel retrieval algorithm (BRA, Bayesian Retrieval Algorithm) is developed, which uses Bayesian inference to retrieve soil moisture and optical depth from both H & V channels. Bayesian likelihood is derived in a non parametric manner, in such a way to be a function of ancillary parameters uncertainties (uncertainties in the parameters needed for the retrieval). As a major advantage, prior knowledge for soil moisture and optical depth can be directly included as inputs to BRA to improve the retrieval.

A first version of the operative algorithm customized for the Pampas Plains is under evaluation. The Argentinas Pampas region (approximately 83 million hectares) is located in the center-east of Argentina where the main agricultural activities are cereal production and cattle-raising. It accounts for more than 90% of the national grain production. Soybean, wheat, maize and sunflower are the main crops. Weather is among the most important and uncontrollable elements affecting agriculture in this region. Customized inputs for the area of interest includes specific ancillary data (landcover, local VWC information priors). Moreover, several proxies to vegetation optical depth are being tested (RVI, NDVI and SMOS optical depth). Argentinean radiometer MWR data is used as proxy of skin temperature over vegetated areas.

*Corresponding author: cintiab@iafe.uba.ar

This work was funded by MinCyT-CONAE-CONICET SACD/Aquarius Project 12.

2. METHODOLOGY

2.1. Bayesian Algorithm

Current sm retrieval algorithms such as LPRM and SCA make use of the zero-order radiative transfer model (RT-0) [3] as a deterministic forward model that should be minimized in order to retrieve sm (and τ in the case of LPRM). Further analyses can extend the model to a stochastic model that includes both forward model ancillary parameters (θ) uncertainties and instrumental noise. Furthermore, forward model uncertainties can be added as well by making use of an ensemble of forward model instead of implementing solely RT-0. In this work, θ uncertainties and Aquarius instrumental noise where modeled by random variables with Gaussian probability density function (pdf). However, forward model uncertainties where not considered.

Using Bayes' theorem, the conditional (*posterior*) probability of having terrain condition $\bar{sm}$ and $\bar{\tau}$ given measurements of H- & V-pol brightness temperature (TbH_m and TbV_m) and the ancillary parameters $\theta = \tilde{\theta}$ can be expressed as follows:

$$P_Z(\bar{sm}, \bar{\tau} | TbH_m, TbV_m, \tilde{\theta}) = \frac{P_L(TbH_m, TbV_m | \bar{sm}, \bar{\tau}, \tilde{\theta}) P_P(\bar{sm}, \bar{\tau})}{\int\int P_L(TbH_m, TbV_m | \bar{sm}, \bar{\tau}, \tilde{\theta}) P_P(\bar{sm}, \bar{\tau}) dsm\, d\tau} \quad (1)$$

where $P_L(TbH_m, TbV_m | \bar{sm}, \bar{\tau}, \tilde{\theta})$ is the (*likelihood*) probability of measuring TbH_m and TbV_m given the terrain state $sm = \bar{sm}$, $\tau = \bar{\tau}$ and $\theta = \tilde{\theta}$, $P_P(\bar{sm}, \bar{\tau})$ is the *prior* joint density function of $\bar{sm}$ and $\bar{\tau}$ (that includes previous knowledge of $\bar{sm}$ and $\bar{\tau}$), and the double integral is a normalization factor that computes the probability of measuring TbH_m and TbV_m.

The *likelihood* pdf is related to both the forward model (in this case, RT-0) and the distribution of the random variables used as inputs to the model. Moreover, the *prior* used in this analysis was chosen to be uniform ranging from 0 to $0.5\, m^3/m^3$ for the sm variable (meaning no previous knowledge of sm is available), and Gaussian distribution for the τ variable. The Gaussian pdf was centered on the τ value derived from MODIS NDVI, with a variance being a function of the difference between τ obtained from MODIS NDVI (τ_{NDVI}) and Aquarius RVI (τ_{RVI}). When both τ_{NDVI} and τ_{RVI} are similar, then it is assumed that the derived value completely is reliable. Accordingly, the *prior* pdf enhances the *likelihood* pdf on the area of the domain where τ values are close to the τ_{NDVI}. Therefore, in this case, the *prior* pdf highly restricts the *posterior* pdf, thus strongly lowering its variance. On the other hand, if τ_{NDVI} and τ_{RVI} are very different, then the τ_{NDVI} is not reliable, and the *posterior* pdf is likely to resemble the *likelihood*. Furthermore, if the uncertainties on the ancillary parameters are low, then the BRA approach is presumably to encounter sm and τ values similar to the ones retrieved by the DCA.

Given the *posterior* pdf in (1), two estimators were derived. One of the estimators is the minimum variance estimator (BRA Mean). It is derived as the expectation value of the posterior pdf, and the estimated sm from it is of the form:

$$\hat{sm}_{mean} = \int\int_D sm P_Z(sm, \tau | TbH_m, TbV_m, \tilde{\theta}) dsm\, d\tau \quad (2)$$

and the variance of the estimation will be:

$$\sigma^2_{\hat{sm}_{mean}} = \int\int_D (sm - \hat{sm}_{mean})^2 P_Z(sm, \tau | TbH_m, TbV_m, \tilde{\theta}) dsm\, d\tau \quad (3)$$

being D the sm and τ domain where the forward model spans. Previous variance is an indicator of the error in the estimated $\hat{sm}_{mean}$.

The other estimator implemented was the maximum a posteriori (BRA MAP), which is the mode of the *posterior* pdf, and can be expressed as:

$$\hat{sm}_{map} = \arg\max_{sm} P_Z(sm, \tau | TbH_m, TbV_m, \tilde{\theta}) \quad (4)$$

in this case, the variance can be computed as follows:

$$\begin{aligned}\sigma^2_{\hat{sm}_{map}} &= \int\int_D (sm - \hat{sm}_{map})^2 P_Z(sm, \tau | TbH_m, TbV_m, \tilde{\theta}) dsm\, d\tau \\ &= \sigma^2_{\hat{sm}_{mean}} + (\hat{sm}_{mean} - \hat{sm}_{map})^2\end{aligned} \quad (5)$$

Both estimators were also used to estimate τ with its corresponding functional form.

The advantages of BRA are: (i) errors on the retrieved variables can be estimated in an univocal way, (ii) it gives the possibility to use prior information about the retrieved variables (provided by other sensors or in situ historical data), (iii) it can handle uncertainties on the ancillary parameters. The main disadvantage of BRA is its time performance. In order to improve the runtime, a Markov Chain Monte Carlo was implemented.

2.1.1. *Markov Chain Monte Carlo*

At first, the BRA approach was computed on a regular grid spanning the Bayes pdfs domain (limited mainly by the *prior* pdf). In this scheme, the precision of the estimations are related to the grid resolution. Therefore, the grid was then refined where the *posterior* pdf displays significant values. An even better sampling approach involves the implementation of the Markov Chain Monte Carlo (MCMC) method, which consists in random paths sampling the distributions. MCMC resulted in a 10x speedup using an 8-cores CPU. However, MCMC issues need to be addressed and carefully overcomed, such as initial *burn in* iterations and convergence criteria.

2.2. Other Retrieval Algorithms Implemented

In addition to implementing the BRA approach for Aquarius Tb, other algorithms were computed and evaluated over the area of interest for a few days on August 2012 (austral winter). The algorithms include SCAH & SCAV, Microwave Polarization Difference Algorithm (MPDA [4]) and DCA.

Both MPDA and DCA make use of H- and V-pol Tb to retrieve sm and τ and they differ on the cost function they use to minimize. However retrievals obtained from both algorithms were exactly the same for both sm and τ, therefore DCA results are not *explicitly* shown hereafter.

Ancillary parameters for all the algorithms were selected to be consistent. Land cover dependent parameters were selected following the Look Up Table of algorithm parameters on the SMAP ATBD [5].

3. RESULTS

The sm and τ products derived from the BRA approach (Mean and MAP), SCAH, SCAV and MPDA were evaluated through several performance metrics (correlation, bias, root mean square error RMSE, unbiased RMSE). The Aquarius Level-2 sm estimates derived by United States Department of Agriculture (USDA) [6] were also evaluated. SMOS Level-2 sm and τ products were used as benchmark products because, for the time period selected, SMOS sm spatial pattern was in good agreement with the product Soil Available Water (derived from a water balance model [7]). Nevertheless, absolute SMOS L2 values are not necessarily the *ground truth*. Performance metrics results are shown in Table 1. As shown in Table 1, MPDA exhibited the lowest correlation with SMOS sm, whereas SCAV displayed the highest correlation. On the other hand, USDA sm showed the lowest bias. This is probably due to the fact that USDA is performing a linear fit between co-located Aquarius and SMOS observations to recalibrate Aquarius brightness temperatures. Finally, Bayesian algorithms Mean and MAP exhibited the lowest ubRMSE, though very close to USDA and SCAV ubRMSE.

	R	**Bias**	**RMSE**	**ubRMSE**
Mean	0.714	-0.056	0.128	0.115
Map	0.700	-0.063	0.131	0.115
MPDA	0.597	-0.072	0.148	0.130
USDA	0.743	-0.010	0.116	0.116
SCAH	0.715	0.153	0.271	0.224
SCAV	0.762	-0.020	0.118	0.116

Table 1: Soil Moisture Algorithms Performance Metrics

4. DISCUSSION

Several sm and τ retrieval algorithms for Aquarius/SAC-D were implemented and contrasted over the Pampas Plains region in Argentina. Furthermore, a new sm and τ retrieval algorithm that makes use of Bayesian inference was proposed and its performance was also evaluated. As major advantages, the BRA approach provides errors on the estimated variables, enable to enter prior knowledge of the variables to be retrieved and can manage uncertainties on the ancillary parameters. However, the main concern of this approach is its time performance, that though it can be ease through sampling methodologies such as MCMC, the runtime is still large to drive operational global sm and τ retrievals.

Performance metrics for each retrieval algorithm were derived using SMOS Level-2 datasets as benchmark products. One of the main issues that was observed during this study was the sm dynamic range of each algorithm. Whereas SMOS, MPDA, SCAH and SCAV displayed sm values as high as $1\, m^3/m^3$, USDA and Bayesian approaches saturate at sm around $0.5\, m^3/m^3$. USDA algorithm saturates sm taking into account the field capacity (around $0.55\, m^3/m^3$ depending on the soil texture), and Bayesian approach was *manually* saturated at $0.5\, m^3/m^3$ by assigning zero probability to sm higher than $0.5\, m^3/m^3$ on the *prior* pdf. This saturation of sm can be removed on the Mean and MAP estimators by extending the *prior* domain of the MCMC. Accordingly, correlation, RMSE and bias of Mean and MAP sm retrievals might improve significantly. Not only the saturation might be removed, but also when ground truth sm values are high (before saturation), restricting the pdfs domain of the Bayesian approach results in biases toward the center values of sm (around $0.25\, m^3/m^3$).

Another important subject to point out is the low performance of SCAH in contrast to SCAV. SCAH sm exhibited a significant high bias and ubRMSE. If τ values input to SCA are overestimated, then Tb changes are amplified on sm changes, thus overestimating sm. Indeed, when comparing τ values derived from SMOS, MPDA and Bayesian approaches with τ values used as inputs to SCA (derived from MODIS NDVI), τ_{NDVI} displayed highly increased values. Therefore, polarized vegetation parameter b value should be considered, probably using $bH \neq bV$ ($bH < b_{actual}; bV = b_{actual}$) to lower SCAH bias and errors. This will also result in changes in the performance metrics of MPDA, Mean and MAP.

5. REFERENCES

[1] David M. Le Vine, Gary S. E. Lagerloef, Simon Yueh, Fernando A. Pellerano, Emmanuel P. Dinnat, and Frank Wentz, "Aquarius mission technical overview.," in *IGARSS*. 2006, pp. 1678–1680, IEEE.

[2] T.J. Jackson, M.H. Cosh, R. Bindlish, P.J. Starks, D.D. Bosch, M. Seyfried, D.C. Goodrich, M.S. Moran, and

Jinyang Du, "Validation of advanced microwave scanning radiometer soil moisture products," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 48, no. 12, pp. 4256–4272, Dec 2010.

[3] T.J. Jackson and T.J. Schmugge, "Vegetation effects on the microwave emission of soils," *Remote Sensing of Environment*, vol. 36, no. 3, pp. 203 – 212, 1991.

[4] R.A.M. de Jeu, T.R.H. Holmes, R. Panciera, and J.P. Walker, "Parameterization of the land parameter retrieval model for l-band observations using the nafe'05 data set," *Geoscience and Remote Sensing Letters, IEEE*, vol. 6, no. 4, pp. 630–634, Oct 2009.

[5] P. ONeill, S. Chan, E. Njoku, T. Jackson, and R. Bindlish, "Algorithm Theoretical Basis Document (ATBD) SMAP Level 2 & 3 Soil Moisture (Passive)," *Initial Release v.1*, 2012.

[6] R. Bindlish and T. Jackson, "Aquarius level-2 swath single orbit soil moisture. version 2," *Boulder, Colorado USA: NASA DAAC at the National Snow and Ice Data Center*, 2013.

[7] Argentina Facultad de Agronomía, Universidad de Buenos Aires, "Centro de información agroclimática (CIAg)," `http://www.agro.uba.ar/centros/ciag`.