VALIDATING SMOS ON THE WETLAND POLESIE, IN POLAND

Jan Słomiński (1), Ewa Słomińska (1), Bogusław Usowicz (2), Jerzy Usowicz (3), Mateusz Łukowski (2), Anna Wójciga (2), Wojciech Mareczewski (1)

(1) Space Research Centre (SRC PAS), 00-716 Warsaw, ul. Bartycka 18A, Poland, Email: wmar@cbk.waw.pl
(2) Institute of Agrophysics (IA PAS), Lublin, Poland
(3) University of Nicolaus Copernicus (UMK), Toruń, Poland

ABSTRACT

Results of first approaches to validating L1C SMOS data are presented on example test sites at the wetland Biebrza and Polesie, in Poland. The discussion starts arguing that a validation is not for simple comparing values, but a way of referring two multi variable data sets by means of the CMEM model in L-MEB loop, under necessity of behaving the radiation transfer equation. The said arguments try explaining why SMOS principle of observing manages better with finding areal representation in variables than the ground data. First effects of assessing optical thickness are presented. In conclusion the task of validation is a kind of statistical processing. The importance statistical means is demonstrated on another thread, the dielectric constant modelling for soils. Another statistical aspect is approached in modelling physical properties of soils, for including relations between dielectric properties and the porosity. A believe that the statistical-physical model on dielectric properties, may include the porosity to validation, and improve soil moisture retrieval from SMOS data, is demonstrated.

1. TEST SITES AND RFI IN POLAND

This paper takes examples for validating SMOS data, for two selected test sites in Poland, in the Cal/Val project AO 3275. They are two largest wetlands in the country – Biebrza, and Polesie. The L1C SMOS data for these sites are available for us in 26-42 pixel subsections. Firstly, the aim was to catch manners and purposes of possible validations, with the data taken in May 2010. The Fig. 1, shows few subsections of May data overlaid on a full snapshot taken in February 2010. One can see how the selected patches from May, match the background in spatial distributions. It is also visible well how they fit one to another, pixel by pixel, due DGG (Digital Global Grid), and how these sub-areas emerge from the background, by BT values. However, the background covers RFI sources, contaminating the test sites. There are at least six RFI sources persisting in Poland. Inspection by Stokes-1 element, was done for this example because then RFI sites are disclosed better, looking under apodization. This is believed that higher elements of Stokes can aid in further recognition of error components better, but till now that is not clear why some of Stokes elements behave in the SMOS Toolbox differently than it would be expected by values. Having the larger area of a full snapshot, one can inspect more and state whether a particular ROI site falls, or not under the line of ripple errors propagated along three directions, properly to the three arm structure of antenna on the orbit.

Now the collected SMOS data from these test sites in Poland cover the period from April to June, and is still delivered in NRT, even nearly twice a day (!) in morning and evening passes. The differences between morning and evening passes, and presence of RFI, made that the way of synthetic presenting evolution of the target area in data, was not possible yet. Each third day the sites fall more or less centrally into FOV (Field of View). So there are data files with different quality, best quality in central passes, and lower quality in more frequently collected every day. Those more frequent passes are aside to the central FOV. Generally, we we already have data collection rich but most contaminated by RFI, and not elaborated in time evolution. First questions for elaborating was interpreting in BT components, and defining a path of validating by ground based values.

2. RADIATION TRANSFER

SMOS L2 data shall be released soon after the commissioning, but till now we work on L1C. A use of L1C requires behaving the Radiation Transfer Equation
alternated states. SMOS must alternate the polarization components, may be determined, in at least three polarization components, and two other circular polarizations exclusively or commonly. That way two linear Only two orthogonal components (X,Y) can be measured. Presently the paper can't demonstrate such potential effects in mitigating RFIs.

SMOS works with polarizations in alternating manner. Only two orthogonal components (X,Y) can be measured exclusively or commonly. That way two linear polarization components, and two other circular polarization components, may be determined, in at least three alternated states. SMOS must alternate the polarization states in triples, and provide a control of the polarization states by a system of flags. There is a potential difficulty with taking a triple improperly to determine snapshot components, but now it is not described. Validation can be done on the Stokes elements, but also on linear polarization components BT (H,V). More advanced ways of employing Stokes elements, especially for the second pair, is to be postponed on further.

Validating L1C data requires confronting BT received on the orbit to BT known on the ground, in pairs (X,Y or H,V). The question of alternating polarizations is now needed to conclude that snapshots, determined by alternated triples, are available in a time series. One snapshot is wide on about 1000 km in the ground view, and subsequent snapshots are taken approximately once per second, in time. Subsequent snapshots image the same area, capturing a new part coming in the front, and loosing a part behind. The same current area is watched by the spacecraft moving, through some interval of time. It means that a particular pixel, fixed in DGG, is shot many times till it falls out of the view. This way the time series for a pixel represents some number of independent looks, each are taken independently and with another incident angle from the range of approximately 5-65 degree. That is a great potential of SMOS observations, and the property of keeping consistence in time and space distribution, in the snapshot data content. Data in a 1000 km wide snapshot covers emission from the entire area, and none of the contributing target element is missed, as far as the optical thickness, and the boundary conditions in relation to the wavelength, allow. It is worth emphasizing that such consistence is not available for ground based techniques, and this is the reason for which one should take SMOS observations for a real valid reference for observing long distance trends, even if the precision in radiometric and spatial imaging is limited. Consistency in time and space, lays fundamentally in the principle of observations. That aspect of spatio-temporal consistence is not possible for keeping with comparable quality in any statistical estimations from ground based data. The consequence of the polarization regime in SMOS, is that it provides the polarimetric (or angular) signature and combines contributions from all existing emission sources in the target, into a common representative value of BT, over each large pixel in its total area, and with the consistency to all (!) other pixels in the snapshot. However, the signature is specific to the pixel content, and it does not enhance the BT variable image spatially. Not details in images are in concern, but total representative BT values in polarization components. For that purpose of radiometric observations, the constrained aperture of SMOS on the orbit, is not an obstacle, though for imaging the details it is. That differs SMOS from InSAR, and POLinSAR in principles and abilities. That techniques need the aperture
for precise imaging, watching geometry of boundary conditions, etc., but SMOS reduces all the heterogeneity of the target in representativeness and relevance to water over large areas.

Understanding that mechanism of alternating polarizations, was needed to behave the order of polarizations while employing data, that is when data is extracted to a sub-section and passed to further processing. It can be found in L1C data, that a particular pixel can be “sampled” by SMOS for BT even approximately 220-230 times, for a single polarization state, till it is in view of FOV. That corresponds to about maximum 72 snapshots available in a pass.

SMOS integrates the measured emission over the pixel area, taking into account the entire existing environment even with respect to geometrical dimensions, boundaries, and optical thickness, respecting the wavelength. Ground measurements remain being specific to the measurement site, while the entire surrounding environment is simply ignored until it is not included into a statistical estimation by spatio-temporal modelling, or into some other thematic model developed over the area, like for instance a hydrological model of the catchment. Even a very good grid of the test sites, cannot keep consistency so perfectly in spatio-temporal aspects, as the SMOS instrument performs it. Besides that, any great number of samples and measurements taken from the ground is still poor and less than sufficient for statistical approaches. SMOS delivers data massively.

We demonstrate a believe that the principle of SMOS provides de facto a statistical observation based not only on the physics of emission and instrumental specific, but also on statistical aspects of further processing the polarization signature. The instrumental principle was developed with with great respect of fundamental postulates statistics. If that was not an aim taken directly, it came in effect of technical principles of interferometric radiometry employing polarizations.

One needs comparing the ground based data in the areal representative value, to BT values from SMOS in a temporal series ordered by varying incident angles, when all values from L1C are representative for a particular pixel. There is not a dilemma of choosing the best BT value, or comparing to a particular place on the ground, but is the task of comparing two values finally representing the entire pixel area from L1C data, and another representing the entire area on the ground in ground data, both obtained independently. One of them should be taken for a reference. We claim that it is just L1C pixel from SMOS, not the ground data. Some apparent trouble is that we have not a single value of BT but a series of angular signature. A corresponding series of the ground based BT values, in polarimetric components H, V, can be obtained from the Common Microwave Emission Model (CMEM) [2], which was elaborated at ECMWF for SMOS. It converts a set of ground data to BT, respecting RTE equation, composed from the values of tau (optical thickness), omega (albedo in proper microwave band L), and split on BT(H), BT(V) for nadir. One can create a series of the BT components in dependence on the incident angle (theta) by dividing it over costheta. That way the question of representativeness is left for the order of multiple use of CMEM, when required is smaller or larger aggregated pixels. However, the values of BT (H,V) obtained from the ground, under comparing to L1C BT angular series, must be expressible in a from combined to well understood and usable parameters. Two series can be taken for a product of two RTE equations differing by parameter values, firstly tau, and omega. This is the question how to obtain them. We get them from CMEM converting ground based data on BT, and then may fit these BT in series to the series of BT in L1C, in a process of iterations under L-MEB, being the Levenberg-Marquardt Algorithm (LMA) known from many sources, or as L-band Microwave Emission of the Biosphere.

In practice the task of fitting BT obtained from the ground data, to BT from L1C data, requires accepting that SMOS data is the reference to which the ground based data is drawn for best available fitting. This is not obvious for a first approach, especially that L1C data can be spread, contaminated from different contributions, etc., while the ground based data is apparently consistent in the angular series. However, that consistency has nothing to the representativeness of the ground data to the entire area of the SMOS pixel. The most precise and honest measurements are made in-situ, and in always a limited number being not enough for achieving areal representativeness. In the contrary, the SMOS data is inherently highly consistent in spatio-temporal measures not only in a pixel, but in an entire snapshot, or at least inside its central part of FOV.

Proceeding that way seems to be difficult for realizing a sense when it is disturbed by a background question why the ground data is questioned on a base of so much spread L1C data under fitting? The quality of ground based data matters however much, because it is a source of the starting estimate in iterations. Convergence under fitting may depend on it. The other side of that way is that a user provide weighting between contributions in ground data, for example by weighting fractions of low vegetation in high vegetation classes, or fractions of deciduous trees in coniferous forests, or a contribution of open waters to some area which does not cover lakes but temporarily may contain a plenty of water pools after the rain, what may be sensed in L1C data. CMEM contains a hierarchical system of entering ground based
data, ordered from detailed indexes of vegetation and meteorological conditions, to final variables (tau, omega) determining RTE equation parameters and values. CMEM employs several models of relations between internal variables, what may create a temporary trouble for a user. CMEM is a package capable for satisfying approx. 200 kinds of ecosystems, depending on a set of input data content, and that may be not easy to extract from it an analytical record of a function needed for current instances, and enabling evaluating not only a direct way of coming to BT, but also an inverse way of determining a needed parameter value. However, that is possible to find such a function for a particular sort of the ecosystem proper to a selected pixel and collected ground data.

Some results obtained on that way are presented below. The Fig.2, 3 show L1C raw data (Full pol.) before rotating the system from X,Y to B,H, for two wetland areas: Biebrza, and Polesie respectively.

The two components proper to cross polarization XY, sticks more or less to zero BT values, and now are not employed yet, but the remaining components corresponding to linear polarizations X, Y, are meaningfully different between the areas. The first area of Biebrza is less, while the second Polesie is more covered by forests. Now we present only the way of proceeding and particular values for the land cover, and other input parameters are not quoted. It is somehow expected that the polarimetric angular signature for more homogeneous area should be observed in the incident angle interval more smooth. Then the data is rotated from the antenna plane coordinate system X,Y, to the ground based system H,V, as shown on the Fig. 3 (Biebrza), and Fig.4 (Polesie).
One can conclude that rotating the coordinate system made the linear polarization components (H,V) departed, at large incident angles, while remaining properly equal for nadir. The effect of departing, is more clear for Biebrza than Polesie, but now we refrain fro further concluding till the process of fitting ground based data (continuous lines) is not precisely controlled by several fits according to several variables being subjective for fitting. These first approaches to fitting, were taken only for optical depth (τ) with the achieved results quoted on the figures. A process of fitting, and in effect being an action of affecting input data to the RTE parameters, should be performed several times choosing several parameters., and performing it one by one under supervision of a user. All that large set of input parameters to CMEM, is hierarchically combine into two, τ and ω, but it is still a set of many variables determining parameters of RTE. And among them some are more or less sensiti-vely contributing to final effect of fitting. Now we stay with the choice of τ, that is in the last highest output level of the hierarchy of variables. Each process of fitting converges to shallow local extrema, and can be performed variable by variable separately, if they are more than less independent. Therefore, it is expected that results of fitting may become more significant when the supervising decisions for particular variables shall be done, not only on the highest level of τ, but also on the lower level of fractions, particular physical proper-ties etc. Effects of validations depend on the quality of supervision, and finally on the knowledge of the ground based data relevance to the land cover, physical characteristics and biosphere parameters. Extents of possible proceeding seems to be not constrained, but must be constrained statistically, by the quality of L1C data. Now that is not recognized. May be that more experience with similar processing at a homogeneous super-site in ecosystem classes proper to that kind of ecosystems in middle latitudes shall provide more answers.

The other type of concluding is from the area of statistics. Assimilating values of the polarimetric angular signature is a kind of statistical processing. BT, Stokes elements are provided in power-energy terms. L1C values are expressed in a way constrained to the measures being as far independent on the instrumental specific as it was possible in the method. We mean here similarly like that is in fundamental motivations of radiometry, justifying that the measure of BT is proper because it does not depend essentially on the aperture.

One can use the aperture larger or smaller but the brightness temperature must be proper to the radiation source, not to the instrument. Aperture, and other instrumental condition may determine possible sensitivity and angular ranges but not the measures.

That is a very good cause to think about large potential of information covered in SMOS data. Statistics employs similar measures, and even similar tools, though the system was determined on the way of engineering, respecting physics. Also other processing with the iterative L-MEB loop converges under estimating total error measures by the cost function being a sum of squares from all departures between ground based data representation in a pixel, from L1C data. That criteria are also very usual in statistics. On that way the RTE and CMEM combine a complex superposition of several or many exponential components, properly to the layered media. This is a kind of a predictive function, and it was proved in the literature that there are very good examples of successful modelling very fine effects under sensing such aspects which were hardly detectable before. J. Walker demonstrated effects of rocky soils, Ch. Maetzler proved many subtle but fundamental effects in vegetation forms which were not expected for observing that way by microwave radiometric means. Now, L1C users search for effects of dew. Detectability of fine effects, is in fact a problem of recovering signal buried in noise, what is also a domain of statistics. If SMOS respects physics so much as it was possible in the observational principle, then we expect the existing potential is not disclosed yet.
Figure 7. The area of the test site Bagno Bubnów, in the ASAR APP image interpreted on water content bound to the land cover vegetation classes. The black frame is in correspondence to the Fig. 6, the black cross marks the kernel of maximum soil moisture.

These exercises with L1C data, CMEM, and L-MEB, rebuilt our simply minded imagination about validation. The aim of validating SMOS data was taken with the research background in soil physics at IA PAS, Lublin, Poland. Previous experience included modelling thermal properties of soil and investigating water related spatial distributions in the field by statistical methods. Also, the most significant achievement was in modelling soil physical properties by statistical means [3]. That evoked a concept of continuing investigating spatial distributions of properties, and adopting this interests to the project SWEX (Soil Water and Energy Exchange) on contributing to Cal-Val SVRT program. The concept was much too simple. It was expected that spatial distribution patterns, like that one on the Fig. 6, should agree to other distributions related to water, retrieved from other data sources, like ENVISAT ASAR.

And indeed, the agreement was evident but only in concentration kernels of a retrieved property, Soil Moisture (SM) for instance. However other parts of lower values of the property, that is having less probability for taking the values lower than in the kernel, were disappointing. One cannot rely on other ASAR data if that source is available rarely, when it cannot be in good temporal agreement. The more, it soon occurred that statistically determined distributions may also depend on the method (kriging and co-kriging were used). Waiting for comparing to patterns in larger scales seemed to be in least hopes. Now we see the problem differently, in use, what does not depreciate the value of statistical means. The validating task occurred not so hopelessly poor, though it is a bit more complex than it was desired.

3. STATISTICAL MODELING OF SOIL PROPERTIES

B. Usowicz, IA PAS, Lublin, developed the concept of statistical modeling physical properties of the soil as a mixture of compounds defined by particle fractions. Fig. 8 is a general geometrical depiction of the concept. A sample of mixture is a unitary volume filled by equal radius spheres, representing particular fractions. The spheres contacts one another, create conductive paths and substitute a capacity, for electrical or thermal storage of charge or heat, and conductivity. The spheres are only to define starting conditions for a combinatorial play for a necessary number of spheres, to model a desired characteristics of the property in relation to the content of gas, liquid and solid compounds, defined by fractions as it is practiced in soil researches. The spheres are only for establishing a general expression on the property value in term of the probability for existing contact points and conductive paths. The concept was developed firstly for thermal properties, and is explained in several papers available as [3]. Now it is extended on dielectric properties of soil.

Figure 8. A general configuration of the Usowicz model and corresponding representation by analogue electrical R and C meshes [2]

A general expression on the property value is the following:

$$\varepsilon = \frac{4\pi}{\sum_{j=1}^{L} x_{ij} \varepsilon_{1} (T) r_{i} + \ldots + x_{ik} \varepsilon_{k} (T) r_{k}}$$

(1)

It corresponds to the analogue meshes for behaving Kirchoff's and Ohm's laws, but the values under modeling are a subject of the number of spheres \( u \), and the number of possible contacts \( L \). One of elementary configuration of spheres is shown on the Fig. 9, for the case \( u=3 \) and \( u=7 \), when the spheres are to exemplify that all the state phase involved are substituted by equal
spheres but differing in number. That example was to generate possible characteristics in dependence on water content.

Figure 9. Fig. 8 Number of the required parallel connections "u" as a function of soil water saturation (theta/phi) [3]

In effect of assuming the largest number of spheres, the entire range of the domain is discretized, and the following types of characteristics are accessible for modelling, as on the Fig. 10. The clue of the concept, is to provide a standard of available functions and then to feed them by the granulometric fractions of compounds before final calibration to physical values. The concept is a mechanism for controlling a mixture by fractions on the input, and by internal conditions on the play with the sphere numbers for achieving a desired variability versus water content.

That way any desired mixture can be defined on formal way for other modelling purposes. The concept was very intensively checked in physical cases of many soil types. Currently the interest is in using that statistical model for CMEM, for that purposes when one could need investigating more than is enabled by the models of Dobson, or Mironov, in relation to the porosity. The model was developed mainly for investigating porosity of soils.

We expect that a use of CMEM with the couple of this model can open a possibility on matching the optical thickness tau to the porosity. The area of a SMOS pixel is not intended for investigating the spatial distribution of the porosity of soils, but SMOS more rather then less indirectly senses the optical thickness. Thus the porosity should gain the access to investigating its contribution to SMOS data. It is believed that it shall improve SM retrieval. Fig. 12, displays the proof on agreement of the statistical-physical model on the dielectric constant property to few other models. That reference models do not cover Dobson or Mironov, however. The comparison was developed for other purposes related to calibrating TDRs.

Figure 11. Available types of the dielectric constant versus the water content, for different numbers of u [3]

4. CONCLUSIONS

I) The process of validation provide Ground Supported Values (GSV) which are based on the data from the orbit.
II) Validations for L1C data lead to statistical aspects of data processing with a use of CMEM and L-MEB.
III) Other statistical experience in modelling physical properties of soils, shows that there are possibilities for including contribution from the porosity to validations and GSV values.

ACKNOWLEDGEMENTS

The project is conducted in a part, in the frames of the program PECS realized between ESA and partners from Poland, and in another ground based part by national grants N305 046 31/1707, and N305 107 32/3865. The EO data from ENVISTA ASAR, and SMOS data were provided by the ESA EOPI Cat-1 project SMOS Cal/Val 3275.
5. References

1. SMOS level 2 processor Soil moisture ATBD, SO-TN-ESL-SM-GS-0001, Issue 3.d, Date 13/06/2010
2. CMEM: ECMWF Community M-ve Emission Model
   http://www.ecmwf.int/research/ESA_projects/SMOS/cmem/cmem_index.html
3. B. Usowicz, „Statystyczno-fizyczne modele przepływu masy i energii w ośrodku porowatym” (Statistical-physical models for mass and energy transfer in porous media), Acta Agrophysica, IA PAN, Lublin 2000 (in Polish), English paper is also available as item 78 in the web pages http://users.ipan.lublin.pl/~usowicz/active.html