

Soil moisture estimation using TerraSAR-X data and empirical models upgraded with SOMs

Matej Kseneman, *Member, IEEE*, Dušan Gleich, *Member, IEEE*, Jože Mohorko, *Member, IEEE*,
Žarko Čučej, *Member, IEEE*

Abstract — For public safety reasons is essential early prediction of natural disasters like floods and landslides, which can be based on soil moisture estimation. Our goal is to evaluate quality of the Dubois and Shi empirical models used in soil moisture parameters retrieved from TerraSAR-X satellite data. In order to compare and validate estimated results, volumetric soil moisture measurements were done with a Pico64 sensor at the time of SAR image capture. Our measurements show superiority of the Shi model by a factor 4 with regard to the Dubois model. Results are further improved using Self-organizing map, which is a type of a neural network in order to file out artefacts and fields with inaccurate soil moisture estimation.

Keywords — TerraSAR-X, soil moisture estimation, empirical model, self-organizing maps, MBD, GMRF, CUDA

I. INTRODUCTION

MANY studies have shown, that microwave radiation data depends on many natural surface parameters, such as dielectric constant [1] and surface roughness. Dielectric constant highly depends on soil moisture. Because of a huge difference in the relative dielectric constant for dry bare soil (typical value is around 2-3) and pure water (typical value is about 80) [2], there is a chance of moisture estimation through remote sensing. But there are a lot of factors that affect dielectric parameters retrieval, such as molecular orientation, soil type and aggregation. Soil moisture estimation from the remote sensing techniques has been very popular in the past few years, mainly because of a unique range resolution and the ability of global coverage.

For the purpose of soil moisture estimation two empirical models for dielectric parameters retrieval were used. Those models were Dubois [3] and Shi [4], which are based on theoretical models, mainly on the integral equation model (IEM) [5]. Those models are extended and modified to obtain more accurate estimated data on specific terrains compared to field sensor measurements.

This paper is organized in sections. In the section II two empirical models for obtaining dielectric parameters are presented. Section III gives a fundamental description

regarding SOM neural networks used to filter results from soil moisture estimation. Section IV concludes our proposed algorithm for soil moisture parameters retrieval. In the section V an experimental results are shown followed by a brief discussion.

II. EMPIRICAL MODELS

Earth's natural surfaces are considered to be rough. Surface roughness has an effect on radar backscattering, while incidence angle has a key role in a radar reflection.

A. Dubois empirical model and it's inverse algorithm

Dubois empirical algorithm [3] is a simplification of the original Oh empirical model [6]. Soil moisture and surface roughness can be estimated from two different polarizations (horizontal and vertical) of a reflected radar beam. This empirical model is constructed from the POLARSCAT [7] data and only describes co-polarized backscattering as a function of surface roughness, incidence angle and frequency. However, the dielectric constant depends on soil volumetric moisture and backscattering coefficients, which are empirically derived as is written by Eq. (1-2) [3]

$$\sigma_{hh}^0 = 10^{-2.75} \frac{\cos^{1.5} \theta}{\sin^5 \theta} 10^{0.028\varepsilon \tan \theta} (ks \cdot \sin \theta)^{1.4} \lambda^{0.7} \quad (1)$$

$$\sigma_{vv}^0 = 10^{-2.35} \frac{\cos^3 \theta}{\sin^3 \theta} 10^{0.046\varepsilon \tan \theta} (ks \cdot \sin \theta)^{1.1} \lambda^{0.7} \quad (2)$$

where θ is local incidence angle, ε is a real part of complex dielectric constant, h is RMS height, k is wave number and λ is wavelength in cm. Those two relations are valid in the frequency range of 1.5–11 GHz; with incidence angle between 30 and 65° (typical frequency for a German Earth observation satellite TerraSAR-X [8] that uses an X-band SAR is 9.65 GHz).

The reason why this model uses two co-polarized channels is because such solution, as opposed to fully-polarized channels, is less sensitive to system noise and crosstalk. Also their calibration is easier to perform. However, we are interested in resolving equations for unknown dielectric constant and surface roughness parameters for a given TerraSAR-X radar backscatter product file, where a little modification [8] is used (conversion from β^0 to σ^0). Inverse algorithm is described by equations (3-8).

$$A_{hh} = 1.5 \log(\cos \theta) - 5 \log(\sin \theta) + 0.7 \log(\lambda) - 2.75 \quad (3)$$

$$A_{vv} = 3 \log(\cos \theta) - 3 \log(\sin \theta) + 0.7 \log(\lambda) - 2.35 \quad (4)$$

$$B_{hh} = 0.028 \quad B_{vv} = 0.046 \quad C_{hh} = 1.4 \quad C_{vv} = 1.1 \quad (5)$$

M. K. Author, Faculty of Electrical Engineering - University of Maribor, Slovenia (e-mail: matej.kseneman@gmail.com).

D. G. Author, Faculty of Electrical Engineering - University of Maribor, Slovenia (e-mail: dusan.gleich@uni-mb.si).

J. M. Author, Faculty of Electrical Engineering - University of Maribor, Slovenia (e-mail: mohorko@uni-mb.si).

Ž. Č. Author, Faculty of Electrical Engineering - University of Maribor, Slovenia (e-mail: zarko.cucej@uni-mb.si).

$$\Sigma_{hh} = \log(\sigma_{hh}^0) \quad \Sigma_{vv} = \log(\sigma_{vv}^0) \quad (6)$$

$$\varepsilon = \frac{C_{vv}(\Sigma_{hh} - A_{hh}) - C_{hh}(\Sigma_{vv} - A_{vv})}{\tan \theta (B_{hh}C_{vv} - B_{vv}C_{hh})} \quad (7)$$

$$ks = \sigma_{HH}^0{}^{1/1.14} 10^{2.75/1.4} \frac{\sin^{2.57} \theta}{\cos^{1.07} \theta} 10^{-0.02\varepsilon \tan \theta} \lambda^{-0.5} \quad (8)$$

This model is robust also in vegetated areas, but with limitation that the ratio $\sigma_{hh}^0/\sigma_{vv}^0$ has to be less than -11dB.

B. Shi empirical model and it's inverse algorithm

The foundation of Shi model [4] represents the IEM model [5], which includes the effect of surface power spectrum. IEM works well with parameter estimation of σ_{hh}^0 and σ_{vv}^0 for AIRSAR and SIR-C measurements, which operate in the L-band. However, for other SAR data this model remains quite complex.

In literature [5] has been shown, that the IEM can be represented as two functions dependent only on ε_s , θ and ks . These functions can be written as:

$$\alpha_{hh} = \frac{(\varepsilon_s - 1)}{(\cos \theta + \sqrt{\varepsilon_s - \sin^2 \theta})^2} \quad (9)$$

$$\alpha_{vv} = \frac{(\varepsilon_s - 1)(\sin^2 \theta - \varepsilon_s(1 + \sin^2 \theta))}{(\cos \theta + \sqrt{\varepsilon_s - \sin^2 \theta})^2} \quad (10)$$

Shi combined this with a numerical regression equation derived from a simulated data. The inverse equation for soil moisture estimation is given by:

$$10 \log_{10} \left[\frac{|\alpha_{vv}|^2 + |\alpha_{hh}|^2}{\sigma_{vv}^0 + \sigma_{hh}^0} \right] = a_{vh}(\theta) + b_{vh}(\theta) 10 \log_{10} \left[\frac{|\alpha_{vv}| |\alpha_{hh}|}{\sqrt{\sigma_{vv}^0 \sigma_{hh}^0}} \right] \quad (11)$$

$$a_{vh}(\theta) = e^{-12.37 + 37.206 \sin(\theta) - 41.187 \sin^2(\theta) + 18.898 \sin^3(\theta)} \quad (12)$$

$$b_{vh}(\theta) = 0.649 + 0.659 \cos(\theta) - 0.306 \cos^2(\theta) \quad (13)$$

III. SELF-ORGANIZING MAPS

For further improvement of results self-organizing maps (SOM) were used to fill out artefacts and fields with inaccurate soil moisture estimation. The SOM is a type of artificial neural network, trained using unsupervised learning to produce a low-dimensional (typically two-dimensional), discrete representation of the input space of the training samples, called a map. The self-organizing maps are different than other artificial neural networks in the sense that they use a neighbourhood function to preserve topological properties of the input space [9].

The SOM consists of components called nodes or neurons. Each node is associated with a weight vector of the same dimension as the input data vectors and a position in the map space. The usual arrangement of nodes is a regular spacing in a hexagonal or rectangular grid. The SOM describes a mapping from a higher dimensional input space to a lower dimensional map space. The undergoing procedure for a given input sample vector is to find a node, which has a weight vector closest to an input vector. This node is called a winning neuron. Neurons are arranged in the grid topology and for every input vector a winning neuron is selected. Input weight is then updated for winning neuron and its closest neighbouring nodes. For updat-

ing nearest neuron neighbours a Kohonen's rule [9] is used, which is given by (14)-(15). By this rule for every neuron i a certain neighbourhood is updated, and by changing parameter α , a distance of affected neighbourhood can be changed.

$${}_i w(q) = {}_i w(q-1) + \alpha(p(q) - {}_i w(q-1)) \quad (14)$$

$${}_i w(q) = (1 - \alpha) {}_i w(q-1) + \alpha p(q) \quad (15)$$

IV. ALGORITHM CONSTRUCTION

The proposed algorithm is composed of many partial steps, which includes:

1. Sub-sampling of TerraSAR-X image by factor 2. This step has to be done, because raw image with speckle noise greatly affects the performance of soil moisture estimation.
2. Speckle noise removal with Model-Based Despeckling (MBD) algorithm [10].
3. Data conversion to from β^0 to σ^0
4. Conversion to soil volumetric moisture using selected empirical model.
5. Removing artefacts using SOM classification.

A. Speckle noise removal

Radar beams can interact with each other constructively or destructively, which causes bright or dark pixels named speckle noise. Speckle noise is well known phenomenon in radar remote sensing systems, even though it can appear at any type of coherent radiation.

Bayes approach to speckle noise removal from synthetic aperture radar (SAR) is used for image quality enhancement and various techniques for information extraction. A first order Bayes inference is used for maximum a posterior (MAP) estimation. Prior in Bayes formula is modelled with Gauss-Markov random fields (GMRF). To find the best model parameters a second order Bayes inference is used. This method is proven to be good at speckle noise removal and texture estimation [10]. Bayes inference is given by

$$p(x|y, \theta) = \frac{p(y|x, \theta) p(x|\theta)}{p(y|\theta)} \quad (16)$$

where y is image with a speckle noise, x is a noise-free image, θ are model parameters, $p(y|x, \theta)$ represents likelihood, $p(x|\theta)$ is prior and $p(y|\theta)$ is evidence. Because evidence does not play any role in maximization process over x , it can be neglected from further model derivation.

$$\hat{x}(y) = \arg \max_x p(y|x, \theta) p(x|\theta) \quad (17)$$

Speckle noise in the original SAR image is modelled as multiplicative noise $y = xn$, where n represents noise. Probability density function (pdf) of likelihood is modelled by gamma distribution.

$$p(y_s | x_s) = 2 \left(\frac{y_s}{x_s} \right)^{2L-1} \frac{L^L}{x \Gamma(L)} \cdot \exp \left(-L \left(\frac{y_s}{x_s} \right)^2 \right) \quad (18)$$

where L is equivalent number of looks, s is current pixel location and Γ is gamma function.

GMRF belong to a Gibbs family of models, which are suitable for describing SAR images. GMRF are given by

$$p(y|\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x - \sum_{r \in N_s} \theta_r (x_{s+r} + x_{s-r})}{2\sigma^2}\right) \quad (19)$$

where N_s is pixel neighbourhood, r defines neighbouring pixels and θ_r are texture parameters. Solution of first derivative is given by Eq. 18 and Eq. 20, where we used Hessian approximation in logarithmic space to derive Eq. 21 and 22.

$$x_s^4 - x_s^3 \sum_{r \in N_s} \theta_r (x_{s+r} + x_{s-r}) + 2L\sigma^2 x_s^2 - 2L\sigma^2 y_s^2 = 0 \quad (20)$$

$$\log p(y|\theta) \approx \sum_{i=1}^{N \times N} \frac{1}{2} (\log 2\pi - \log h_{ii}) + \log p(y_i | \hat{x}_i) + \log p(\hat{x}_i | \theta) \quad (21)$$

$$h_{ii} = \frac{6Ly_s^2}{x_{sMAP}} - \frac{2L}{x_{sMAP}^2} + \frac{1}{\sigma^2} \left(1 + \sum_{j \in N_s} \sigma_j^2\right) \quad (22)$$

This algorithm is very complex, so an implementation was made using CUDA technology [11] on graphic card's GPU, where a speed-up up to 30x was achieved.

V. EXPERIMENTAL RESULTS

For testing purposes a TerraSAR-X SAR image of hydro plant region Zlatoličje near Maribor was used as shown on Fig. 1. The image was taken on 6.6.2009 with 49° inclination angle and in dual polarization mode, with ground range 1.56 m and azimuth range 2.2 m.



Fig. 1. Quick-look image of the testing area.

Parameters for MBD algorithm were: larger window with a size of 13x13, within this window a smaller window moves with the size of a single pixel (because of multithreading on GPUs). Model order was set to 2. Fig. 2 shows the sample image and an output of this algorithm.



Fig. 2. Result of the MBD de-speckling algorithm.

The next step was to assess, which of the two dielectric parameter empirical models performs best on a given

TerraSAR-X data in whole soil moisture estimation algorithm described in section IV. For inversion process of empirical models, a series of equations were used; for Dubois model Eq. (3-7) and for Shi model Eq. (9-13), were used. Results of this assessment are presents on Fig. 3, where can be clearly seen, that Shi model outperforms Dubois model. This fact also proves Table 1, where comparison between estimated and measured values with their deviations is given.

TABLE 1: COMPARISON BETWEEN DIFFERENT VALUES.

	<i>Field</i>	<i>Dubois</i>	<i>Shi</i>	Δx_D	Δx_S
1.	0.26	0.1904	0.2511	26.7	3.4
2.	0.34	0.2422	0.2901	28.7	14.6
3.	0.33	0.1864	0.3220	43.5	2.4
4.	0.33	0.2396	0.3285	27.3	0.4
5.	0.29	0.2311	0.2875	20.3	0.8
6.	0.30	0.2823	0.2804	5.9	6.5
7.	0.26	0.3687	0.2749	41.8	5.7

From a volumetric soil moisture point of view the overall values lie in the interval 10-40%. Given results are better with the Shi model and maximal deviation is almost 4 times smaller than in the case of the Dubois model. Shi model also has better data fitting in the bare soil areas or even over the small vegetated areas.

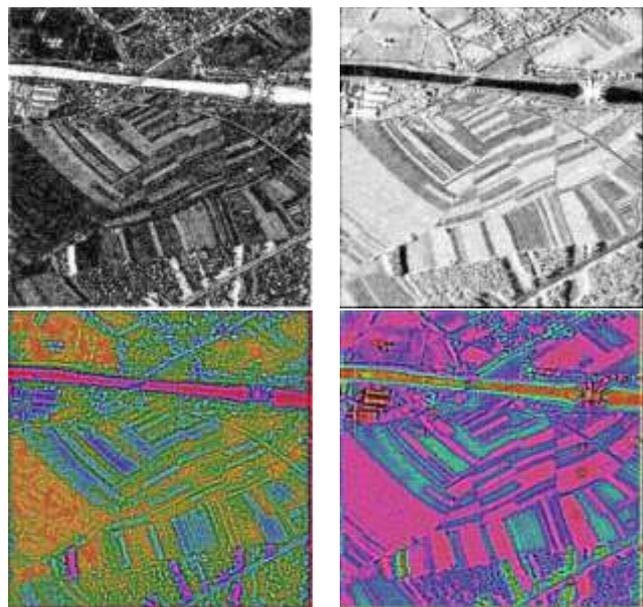


Fig. 3. Volumetric moisture estimations based on empirical models; Dubois (left) and Shi (right), Gray level images (up) artificial coloured images (down).

The last step is to apply SOM in order to remove any disturbances, which are present in various topological forms, such as roads, woods, settlements and volumetric scattering. Therefore a SOM was constructed with an input feature vector consisting of gray-scale value, mean and standard deviation value of 5x5 neighbouring pixels, Sobel edge detector of the same neighbouring size and certain masks, which determine the four most important directions (horizontal, vertical and two diagonals). These masks are represented as zeros on the main direction, values 1 over

the main direction and -1 elsewhere.

Learning samples were manually selected from three independent images, for instance 30 right and 30 false samples. There are used two SOMs consisting of 20x20 neurons in grid topology, one for right and one for false classified samples. The algorithm is designed in a way, that the user can select a number of k-fold neighbours and by selecting this option it greatly affects the produced results. Output sample is marked as positive or with number 1 in cases, where the smallest Euclidian distance occurs between input vector and positive SOM input weights and vice versa. Each of these SOMs is therefore learned with 90x8 input matrixes and with 5000 epochs. Fig. 4 shows an example of the input samples selection and corresponding edge detection image obtained by the Sobel edge operator.

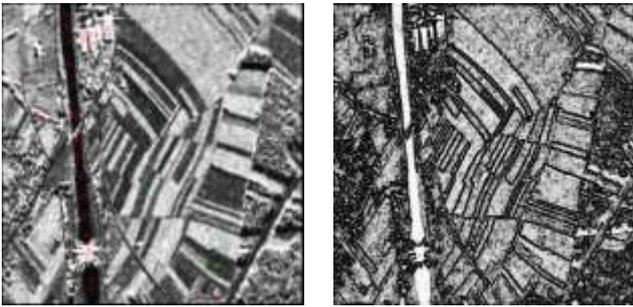


Fig. 4. Manual selection of the input samples (left), edge detection using Sobel operator (right)

Fig. 5 shows SOM output, where the image on the left is a result of SOM with 20x20 neurons, meanwhile the image on the right is a result of SOM composed of 15x15 neurons and 3-fold neighbouring neurons. The left image also contains information about the pixel location used as element of feature vector, while the image on the right does not contain such information. We had shown that pixel location is irrelevant for feature detection algorithm.

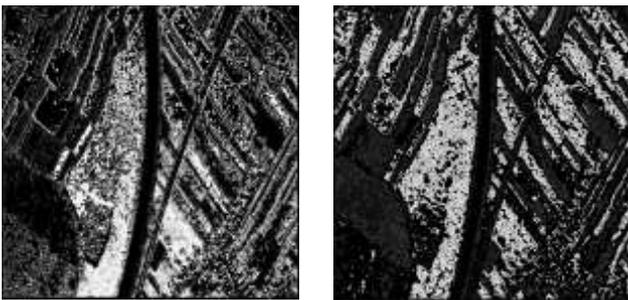


Fig. 5. SOM output; left 20x20, right 15x15 neurons.

For method validation we used energy of quantization noise, which is given by Eq. 23. Table 2 shows the results of SOMs with different number of neurons.

TABLE 2: ENERGY OF QUANTIZED NOISE.

SOM	$E [20 \times 20]$	$E [15 \times 15]$
True	0.099253	0.051228
False	0.094436	0.044901

$$E = \frac{1}{N} \sum_{w_s} \sum_{x_i \in R_s} \|x_i - w_s\|^2 \quad (23)$$

From Table 2 follows that even though a smaller num-

ber of neurons is used, the output performance is nearly two times better when more neurons are captured.

VI. CONCLUSION

In this paper we have shown how to construct volumetric soil moisture estimation algorithm using dual polarization TerraSAR-X data. By our opinion it is better to use Shi model than Dubois. It is true, that Shi model is more complex and takes more computational time, but on the other hand it gives better results.

In order to enhance quality of results the SOMs were introduced, which gave better results in the case if no information regarding pixel position is used as element of feature vector. We have to emphasize that the overall accuracy of SOMs is quite good also in condition of very low percentage of learning set, which was only $30/512^2 = 0,1144 \%$ per image. We can conclude that the SOM usage has enhanced algorithm and almost completely removed undesired artefacts such as roads, rivers and woods. Sobel edge detector is quite influential feature and is therefore seen in the resulting images. This effect can be suppressed with post-processing step like Lee filter [12]. SOM algorithm has turned out to be robust and will achieve even better data fitting in case of appropriate further input feature vector expansion.

Black spots those are visible on Fig. 5 can be removed with simple filling algorithm. Further research will show if this additional step is needed.

APPENDIX

At this point the authors would like to thank Slovenian company Dravske elektrarne Maribor (DEM) for financing project regarding soil moisture estimation in surroundings of the river Drava.

REFERENCES

- [1] E. T. Engman, "Applications of microwave remote sensing of soil moisture for water resources and agriculture," *Remote Sens. Environ.*, vol. 35, 1991, pp. 213-226.
- [2] F. Ulaby, R. Moore, & A. Fung: *Microwave Remote Sensing: Active and Passive I - III*, Addison-Wesley Publication, 1981-1986
- [3] P. C. Dubois, J. J. van Zyl & T. Engman, "Measuring Soil Moisture with Imaging Radars," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 33, no. 4, 1995, pp. 915-926.
- [4] J. Shi, J. Wang, A. Y. Hsu, P. E. O'Neil and E. T. Engman, "Estimation of Bare Surface Soil Moisture and Surface Roughness Parameter Using L-band SAR Image Data," *IEEE Trans. Geosci. Remote Sens.*, vol. 35, no. 5, pp. 1254-1266, 1997.
- [5] A. K. Fung, Z. Li, and K. S. Chen, "Backscattering from a randomly rough dielectric surface," *IEEE Trans. Geosci. Remote Sensing*, vol. 30, 1992, pp. 356-369.
- [6] Y. Oh, Y. C. Kay "Condition for precise measurement of soil surface roughness," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 36, no. 2, 1998, pp. 691-695.
- [7] M. A. Tassoudji, K. Sarabandi, and F. T. Ulaby, "Design consideration and implementation of the LCX polarimetric scatterometer (POLARSCAT)," Rep. 022486-T-2, Radiation Lab., Univ. Michigan, Ann Arbor, June 1989.
- [8] T. Fritz: TerraSAR-X Basic Product Specification Document, DLR, February 2008
- [9] T. Kohonen: *Self-Organizing Maps*, Springer-Verlag, Berlin, 2001
- [10] M. Walessa and M. Datcu, "Model-based despeckling and information extraction from SAR images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 38, 2000, pp. 2258-2269.
- [11] CUDA, http://www.nvidia.com/object/cuda_home.html
- [12] J. Lee, "Refined filtering of image noise using local statistics," *Comput. Graph. Image Processing*, vol. 15, 1981, pp. 380-389.