

SMSYNTH: AN IMAGERY SYNTHESIS SYSTEM FOR SOIL MOISTURE RETRIEVAL

Yun Cao¹, Linlin Xu^{1,*}, Junhuan Peng¹

¹Dept. of Land science and technology, China University of Geosciences, Xueyuan Road, Beijing, China
- cy12160019@163.com, beyond13031@126.com, pengjunhuan@163.com

Commission III, WG III/2

KEY WORDS: Image Synthesis, Soil Moisture, Bayesian Framework, Markov Random Field Model, Oh Model

ABSTRACT:

Soil moisture (SM) is an important variable in various research areas, such as weather and climate forecasting, agriculture, drought and flood monitoring and prediction, and human health. An ongoing challenge in estimating SM via synthetic aperture radar (SAR) is the development of the retrieval SM methods, especially the empirical models needs as training samples a lot of measurements of SM and soil roughness parameters which are very difficult to acquire. As such, it is difficult to develop empirical models using realistic SAR imagery and it is necessary to develop methods to synthesis SAR imagery. To tackle this issue, a SAR imagery synthesis system based on the SM named SMSynth is presented, which can simulate radar signals that are realistic as far as possible to the real SAR imagery. In SMSynth, SAR backscatter coefficients for each soil type are simulated via the Oh model under the Bayesian framework, where the spatial correlation is modeled by the Markov random field (MRF) model. The backscattering coefficients simulated based on the designed soil parameters and sensor parameters are added into the Bayesian framework through the data likelihood where the soil parameters and sensor parameters are set as realistic as possible to the circumstances on the ground and in the validity range of the Oh model. In this way, a complete and coherent Bayesian probabilistic framework is established. Experimental results show that SMSynth is capable of generating realistic SAR images that suit the needs of a large amount of training samples of empirical models.

1. INTRODUCTION

Soil moisture (SM) is an important variable in various research areas. Unfortunately, directly measuring the content of SM in a local scale with distributed, quantitative and accurate information is almost impracticable, due to the high spatial variability of SM. Local direct measurements of SM are also time-consuming, laboring, and expensive. The use of microwave remote sensing sensors, such as synthetic aperture radar (SAR), makes the estimation of SM in a large scale possible, and researches demonstrates that sensors operating in a low frequency of microwave spectrum (P- to L-band) are sensitive to the variations of SM. Many models have been developed using the microwave data which can be generally divided into physical model, semi-empirical model and empirical model. Empirical models, such as machine learning and statistics regression, are developing rapidly among these models. However, the development of empirical models needs as training samples a lot of measurements of SM and soil roughness parameters which are very difficult to acquire. As such, it is necessary to develop methods to synthesis SAR imagery according to known SM aiming at facilitating the training and development of empirical models.

There are physical-based forward electromagnetic models developed to simulate backscattering coefficients. The integral equation model (IEM) (Fung et al., 1992) has been used to produce radar signals (Baghdadi et al., 2012) and a total number of 268110 backscattering coefficients elements (C-band, HH, HV and VV) have been generated. The advanced integral equation model (AIEM) (Chen et al., 2003) and Oh model (Oh et al., 1992) have been used for simulating a set of radar signals (Paloscia et al., 2013) by using a pseudorandom function, drawn from the standard uniform distribution on the open interval (0,1) to cover

the range of each parameter. However, there has no spatial correlation among the set of backscattering coefficients while the real SAR images innately have spatial correlation effect. There have been some researches that synthesizing SAR images with spatial correlation in other areas. A novel SAR image synthesis system named IceSynth has been developed (Wong et al., 2009), which is capable of synthesizing SAR sea-ice textures for each ice type images via stochastic sampling. And IceSynth II is presented in (Wong et al., 2010) which assumed the Markov random field (MRF) model and used a conditional sampling approach. The MRF model is a classic and powerful method for modeling spatial information and has been used in hyperspectral imagery classification under the Bayesian framework (Xu and Li, 2013). An enhanced region-based probabilistic posterior sampling approach for synthesizing SAR imagery has been developed with both sea-ice and oil spills (Xu et al., 2017). To improve the consistency, accuracy, and comprehensiveness of the quantitative estimation of empirical models, the synthesized images should be capable of reflecting the spatial correlation effect.

To make the synthesis SM SAR imagery more similar to the real SAR images, the design of the soil parameters and sensor parameters should be both considered. Accurate simulation of SAR imagery requires an appropriate consideration of the soil characteristics, which include soil moisture, soil roughness, vegetation coverage, and soil types. Soil roughness and vegetation coverage are two main factors that influencing the estimation of SM from SAR images, which of contributions are more than the SM under certain conditions. As such, it is important to have a priori knowledge of the soil properties and surface coverage when simulating SAR SM imagery. The design of the soil parameters (SM and soil roughness parameters) based on soil types should be as realistic as possible to the circumstances on the ground and be suited to the scope of the forward electromagnetic models. Moreover, the

*Corresponding author

most sensitive microwave spectra to the variations of SM are P- to L- bands, while most SAR system sensors onboard are operating at C- to X- bands, which are not the best suited ones for the SM estimation under vegetation coverage. As such, the design of sensor parameters should be the P- to L-bands rather than C-band in (Paloscia et al., 2013). Therefore, it is necessary to simulate the backscattering coefficients that are both realistic and controllable, by integrating the prior information, i.e., the soil type-based soil parameters and sensor bands, into the simulating system.

This letter is the first effort to synthesis SAR imagery with the SM via the Oh model based on the Bayesian framework, named SMSynth. Under the Bayesian framework, the spatial correlation is addressed by the MRF label prior, which promotes the same soil type labels for the spatially close pixels and is implemented by multiple logistic (MLL) prior. The backscattering coefficients simulated by the designed soil parameters and sensor parameters are added into the Bayesian framework through the data likelihood. In the procedure of the parameter designs for the soil conditions, we consider three different kinds of soil characteristics corresponding to different soil parameters values that are as realistic as possible to the circumstances on the ground for three soil type labels and in the validity range of the Oh model. In this way, a complete and coherent Bayesian probabilistic framework that fully accounts for the backscattering coefficients and spatial information is established. Experiments demonstrate that SM-Synth is capable of generating relatively realistic SAR images that are well suited for the developing and testing of automatic SM mapping empirical models.

2. METHODOLOGY

2.1 Problem Formulation

Let a total of K kinds of soil types be denoted by L_1, L_2, \dots, L_K and the synthesized SAR imagery with the backscattering coefficients be denoted by X . The synthesized SAR imagery at site i that belongs to class k is represented by x_i^k .

We express the data generalization model as

$$x_i^k = f^k(\theta^k) + n \quad (1)$$

where $f^k(\theta^k)$ = forward electromagnetic model of class k
 θ^k = parameters in $f^k(\theta^k)$ of class k
 n = noise

The data simulation is to obtain X by maximizing the joint distribution $p(L, X)$ of X and L under the Bayesian framework as introduced in 2.2.

2.2 Bayesian Framework

The Bayesian framework is a probabilistic framework that enables a full exploitation of the information of the simulated backscattering coefficients and the spatial information. Under the Bayesian framework, the spatial information is addressed by the MRF label prior and the simulated backscattering coefficients are added by the data likelihood.

The data simulation of X is achieved by maximizing the joint distribution $p(L, X)$ of L and X , which can be expressed as

$$p(L, X) = p(X|L)p(L) \quad (2)$$

where $p(L)$ = label prior
 $p(X|L)$ = data likelihood

As such, $p(L, X)$ will be maximized when $p(L)$ and $p(X|L)$ are maximized, as introduced in 2.3 and 2.4 respectively.

2.3 Label Prior

The spatial correlation is modeled by the MRF model. The MRF model is a classical method, which can be placed the same label to the nearby pixels around the target pixel that have certain characteristics with the target pixel. The label prior $p(L)$ can be expressed as follows:

$$p(L) = \exp\left(-\sum_{i=1} \sum_{j \in N_i} \delta(L_i, L_j)\right) \quad (3)$$

where $N_i = 3 \times 3$ neighborhood centered at site i
 $\delta(L_i, L_j) = 0$ if $L_i = L_j$
 $\delta(L_i, L_j) = 1$ if $L_i \neq L_j$

The $p(L)$ will have a bigger value if $L_i = L_j$ since the exponential function is the reduced function, and the label of the target pixel i is more closer to the label of the pixel j .

2.4 Data Likelihood

The simulated backscattering coefficients are added into the Bayesian framework by the data likelihood $p(X|L)$, which is expressed by as follows:

$$p(X|L) = \exp\left(-\sum_{i=1} \|x_i^k - f^k(\theta^k)\|^2\right) \quad (4)$$

where x_i^k = value of the pixel at site i and class k
 $f^k(\theta^k)$ = forward electromagnetic model of class k
 θ^k = parameters in the $f^k(\theta^k)$ of class k

2.5 Oh Model

The backscattering coefficients are simulated by the Oh model in this paper. The Oh model is a physical-based forward electromagnetic model, and has been widely used in the radar signals simulation and the soil moisture retrieval. It is capable of simulating fully polarimetric backscattering coefficients, i.e., HH , HV , VH , and VV , where assumed $HV=VH$ with monostatic radars. The simulated backscattering coefficients $f_{hh}^k(\theta^k)$, $f_{hv}^k(\theta^k)$, and $f_{vv}^k(\theta^k)$ at the class k are given as follows:

$$f_{hh}^k(\theta_{hh}^k) = g\sqrt{p}\cos^3\theta(\Gamma_h + \Gamma_v) \quad (5)$$

$$f_{vv}^k(\theta_{vv}^k) = \frac{g}{\sqrt{p}}\cos^3\theta(\Gamma_h + \Gamma_v) \quad (6)$$

$$f_{hv}^k(\theta_{hv}^k) = g\sigma_{vv}^0 \quad (7)$$

where θ^k = all parameters in $f^k(\theta^k)$
 p = co-polarised backscatter ratio
 q = cro-polarised backscatter ratio
 g = function of w and s
 w = wavenumber
 s = RMS height
 Γ_h, Γ_v = Fresnel reflectivity of the surface
 at H and V polarisation respectively
 Γ_0 = Fresnel reflectivity of the surface at nadir

$$\sqrt{p} = \sqrt{\frac{\sigma_{HH}^0}{\sigma_{VV}^0}} = 1 - \left(\frac{2\theta}{\pi}\right)^{1/3} \Gamma_0 \exp[-w \cdot s] \quad (8)$$

$$q = \frac{\sigma_{HV}^0}{\sigma_{VV}^0} = 0.23\sqrt{\Gamma_0}(1 - \exp[-w \cdot s]) \quad (9)$$

$$g = 0.7[1 - \exp\{-0.65(w \cdot s)^{1.8}\}] \quad (10)$$

$$\Gamma_0 = \left| \frac{1 - \sqrt{\varepsilon}}{1 + \sqrt{\varepsilon}} \right|^2 \quad (11)$$

$$\Gamma_h = \frac{\cos\theta - \sqrt{\varepsilon - \sin^2\theta}}{\cos\theta + \sqrt{\varepsilon - \sin^2\theta}} \quad (12)$$

$$\Gamma_v = \frac{\varepsilon\cos\theta - \sqrt{\varepsilon - \sin^2\theta}}{\varepsilon\cos\theta + \sqrt{\varepsilon - \sin^2\theta}} \quad (13)$$

where ε = soil dielectric constant
 θ = incidence angle

The soil dielectric constant ε is generally used as the connection between the SAR backscattering coefficients and the SM. Next the SM and the ε can be determined from the Topp empirical model (Hallikainen et al., 2007).

3. EXPERIMENT AND DISCUSSION

In this section, the proposed method is used to synthesize SAR images, and the results are presented then. We start with describing the experimental design, followed by the results of the synthesized SAR images.

3.1 Experiments Design

Based on the Bayesian framework, the SAR images that are both equipped with the spatial information and the backscattering coefficients are synthesized. The synthesized SAR images are evaluated to see whether they have spatial correlation and precise backscattering coefficients.

3.1.1 Methods We use the MRF model to generate the label image and set three classes in the label image, and the simulated backscattering coefficients simulated by the Oh model are divided into three different kinds of soil characteristics that are the same as the number of classes in the label image. The three different kinds of soil characteristics are corresponding to different soil parameters values that are as realistic as possible to the circumstances on the ground for three soil type labels and in the validity range of the Oh model. The final images are synthesized with adding the simulated backscattering coefficients into the label image according to the label value by the data likelihood.

3.1.2 Parameter Setting For the MRF-based method, we adopt the grey tone values as features, set three classes, use 100 EM iterations and 40,000 simulated-annealing iterations. The generated label image have 1000×1000 pixels, as such the synthesized SAR images have 1000×1000 . For each soil type, the values of the soil parameters obey the Gaussian distribution. The first class is assumed as the relatively flat and bare soils, the soil RMS height s obeys the Gaussian distribution with a mean of 0.5 and a variance of 0.1, and the soil dielectric constant ε obeys the Gaussian distribution with a mean of 5.5 and a variance of 0.1. The second class is assumed as the wheat fields at seedling stage, the soil RMS height s obeys the Gaussian distribution with a mean of 1 and a variance of 0.1, and the soil dielectric constant ε obeys the Gaussian distribution with a mean of 15 and a variance of 0.2. The third class is assumed as a smooth surface such as sand, the soil RMS height s obeys the Gaussian distribution with a mean of 0.2 and a variance of 0.03, and the soil dielectric constant ε obeys the Gaussian distribution with a mean of 8 and a variance of 0.2. For the sensor parameters, the microwave spectrum of the sensor is L-band and the range of the incidence angle is 25 to 45.

3.2 Experimental Results

The label image produced using the MRF approach is shown in Figure 1.

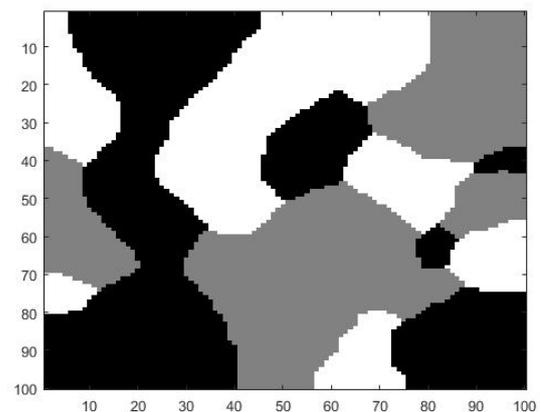


Figure 1. The label image

The synthesized SAR images (HH , HV and VV) produced by SMSynth are shown in Figure 2 (a) (c) (e). We drew the histogram of the data series contained in the synthesized SAR images, which are shown in Figure 2 (b) (d) (f).

As we can see, the synthesized SAR images have spatial correlation, and the simulated backscattering coefficients of the three classes of soil types in synthesized SAR images are realistic to the real SAR backscattering coefficients according to (Jagdhuber, 2014).

4. CONCLUSION

In this paper, a novel SAR image synthesis system based on SM named SMSynth has been presented for generating the simulated

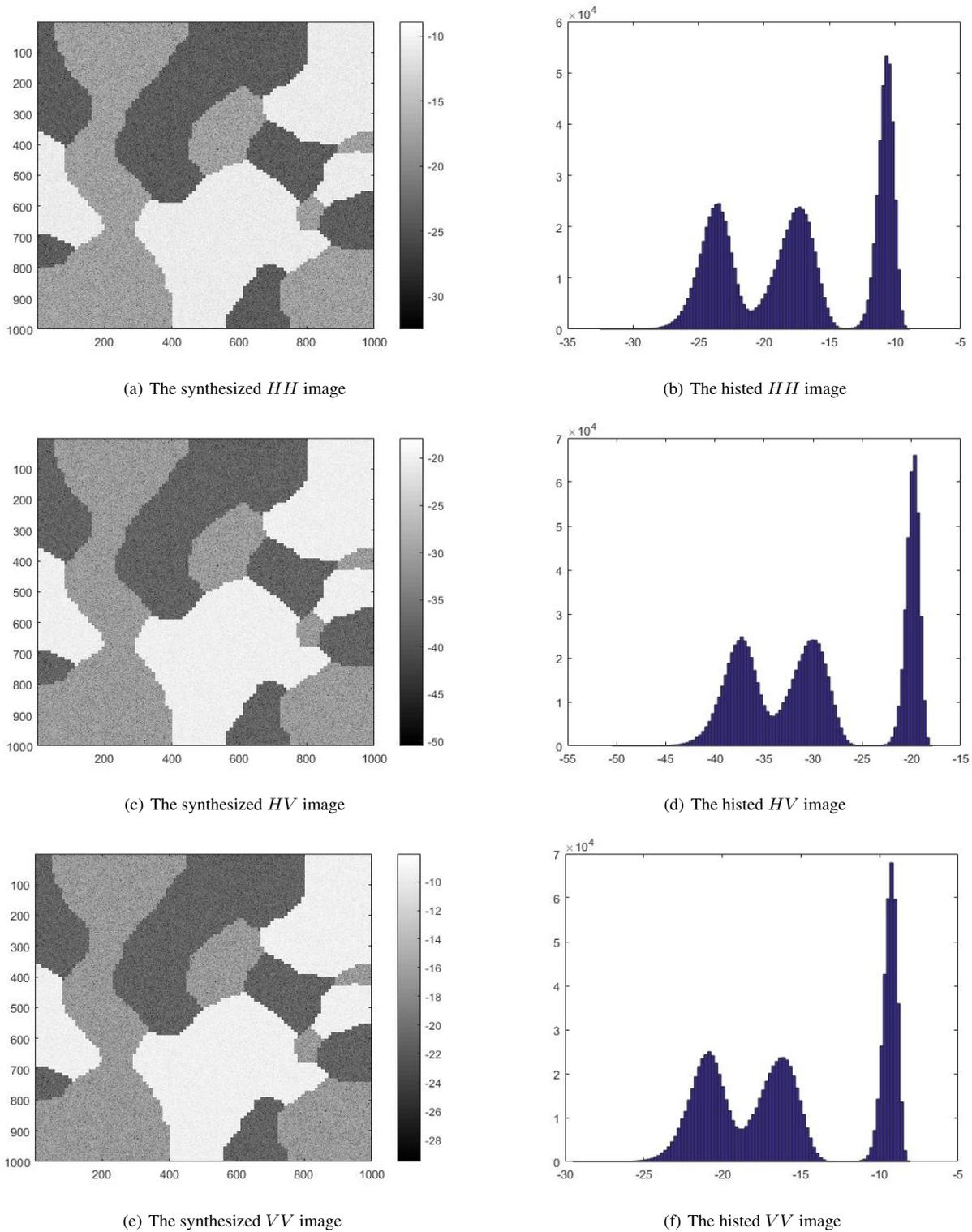


Figure 2. The synthesized SAR images (HH , HV , VV) using SMSynth

SAR images with known SM values. We firstly proposed to synthesize SAR images with SM that have spatial information modeled by the MRF approach and backscattering coefficients modeled by the Oh model based on the Bayesian framework. Under the Bayesian framework, the spatial correlation are established by the label consistency in the label image, and the backscattering coefficients simulated by the designed soil parameters and sensor parameters are adding into the label image according to the label of the soil types. Synthesis results show that the SMSynth has the ability of synthesizing SAR images which have spatial correlation and backscattering coefficients are realistic to the real SAR imagery. As such, the SMSynth makes it suitable for use in the systematic and reliable evaluation of the SM retrieval methods. Future work involves improving the SMSynth system to allow for developing more accurate empirical SM estimation methods, such as machine learning algorithm.

ACKNOWLEDGEMENTS

This paper was supported by the National Natural Science Foundation of China under Grant 41501410.

REFERENCES

- Baghdadi, N., Cresson, R., El Hajj, M., Ludwig, R. and La Jeunesse, I., 2012. Estimation of soil parameters over bare agriculture areas from c-band polarimetric sar data using neural networks. *Hydrology and Earth System Sciences* 16, pp. p–1607.
- Chen, K. S., Wu, T. D., Tsang, L. and Li, Q., 2003. Emission of rough surfaces calculated by the integral equation method with comparison to three-dimensional moment method simulations. *IEEE Transactions on Geoscience and Remote Sensing* 41(1), pp. 90–101.
- Fung, A. K., Li, Z. and Chen, K. S., 1992. Backscattering from a randomly rough dielectric surface. *IEEE Transactions on Geoscience and Remote Sensing* 30(2), pp. 356–369.
- Hallikainen, M. T., Ulaby, F. T., Dobson, M. C., Elrayes, M. A. and Wu, L., 2007. Microwave dielectric behavior of wet soil-part 1: Empirical models and experimental observations. *IEEE Transactions on Geoscience and Remote Sensing* GE-23(1), pp. 25–34.
- Jagdhuber, T., 2014. PolSAR. SAREDU.
- Oh, Y., Sarabandi, K. and Ulaby, F. T., 1992. An empirical model and an inversion technique for radar scattering from bare soil surfaces. *Geoscience and Remote Sensing IEEE Transactions on* 30(2), pp. 370 – 381.
- Paloscia, S., Pettinato, S., Santi, E., Notarnicola, C., Pasolli, L. and Reppucci, A., 2013. Soil moisture mapping using sentinel-1 images: Algorithm and preliminary validation. *Remote Sensing of Environment* 134(4), pp. 234–248.
- Wong, A., Yu, P., Zhang, W. and Clausi, D. A., 2010. Icesynth ii: Synthesis of sar sea-ice imagery using region-based posterior sampling. *IEEE Geoscience and Remote Sensing Letters* 7(2), pp. 348–351.
- Wong, A., Zhang, W. and Clausi, D. A., 2009. Icesynth: An image synthesis system for sea-ice segmentation evaluation. In: *Canadian Conference on Computer and Robot Vision*, pp. 178–183.
- Xu, L. and Li, J., 2013. Bayesian classification of hyperspectral imagery based on probabilistic sparse representation and markov random field. *IEEE Geoscience and Remote Sensing Letters* 11(4), pp. 823–827.
- Xu, L., Wong, A. and Clausi, D. A., 2017. An enhanced probabilistic posterior sampling approach for synthesizing sar imagery with sea ice and oil spills. *IEEE Geoscience and Remote Sensing Letters* 14(2), pp. 188–192.