

Inversion of Soil Moisture from Backscattering Coefficient Using LS-SVM

ZHIHUI CHENG, QINHE ZHANG

School of Science
China Three Gorges University
YiChang, China
zhcheng1211@126.com

JIANING LI, WEI LU

School of Science
China Three Gorges University
YiChang, China
705893089@qq.com

Abstract—Inverting soil moisture from remote sensing data is difficult for the problem is usually nonlinear and ill-posed. To enhance the accuracy of this inversion problem and reduce the effect of surface roughness, a least square-support machine (LS-SVM) based inversion approach is used to retrieve the soil moisture from the radar backscattering coefficients. Starting from the generation of data set by using Integral Equation Model (IEM), wide range of soil moisture and surface roughness are simulated. The kernel and capacity parameter of LS-SVM are optimized through the training process. Then, to assess the effectiveness of the proposed approach, testing data added with Gaussian distributed noise is processed by the suitably defined model. Concerning the robustness of the approach, selected training data is applied when the model is established, and the soil moisture is inverted again. Along this process, the comparison between BP neural network and LS-SVM based method is conducted.

Index Terms—Soil moisture, inversion, remote sensing, backscattering coefficient, LS-SVM.

I. INTRODUCTION

Soil moisture is fundamentally important to land activities, especially those involving agriculture, hydrology, forestry, meteorology, and climate change. To estimate the soil moisture in field scale, numerous approaches have been conducted via using active radar system and passive microwave imaging radiometers [1]-[5]. Among a variety of remote sensing methods, radar backscattering coefficient relate closely to the soil dielectric constant, which is sensitive to 0%-35% soil moisture volumetric content. The integral equation method [6] is the most widely used theoretical radar scattering model for bare surface or sparsely vegetated landscape. According to Mametsa *et al.* [7], the IEM had a much wider applicable region compared to the Kirchhoff approximation (KA) [2], [8] and the small perturbation model (SPM) [2], [8]. It is indicated in the IEM that the backscattering coefficient is also affected by root mean square of surface height and correlation length. Hence, it is indispensable to inverse soil moisture under the effect of various soil roughness properties. To overcome this difficulty, empirical [9-10] and probabilistic [11] approaches have been presented. However, those methods have several disadvantages that hinder them from further application,

which include limited valid span and accuracy, subjecting to a specific region, etc.

This study is trying to approach the inversion problem by using LS-SVM combined with IEM, and aiming to enhance the accuracy of the retrieval process. The IEM model is act as the theoretical direct model, whereas the LS-SVM is for the model based retrieval method. The overall accuracy of this method would also be verified by the noisy and noiseless testing data set. In order to cover wider range of surface roughness condition, this approach is model-based so that it is not site-dependent, and it can be easily adapted to different experimental conditions.

II. BACKGROUND AND METHOD

A. Integral Equation Method (IEM)

For a randomly rough dielectric surface, the backscattering coefficient is mainly determined by soil moisture, surface roughness, soil texture. In the condition of natural terrains, mostly with small slope ($k\sigma > 3$), the single scattering will dominate over the multiple scattering in polarized scattering calculation. Hence, the like polarization backscattering coefficients are given below [6]

$$\sigma_{pp}^0 = \frac{k^2}{2} \exp\{-2k_z^2 \sigma^{2n}\} \sum_{n=1}^{\infty} |I_{pp}^n|^2 \frac{W^n(-2k_x, 0)}{n!} \quad (1)$$

where $p = h$ (horizontal) or v (vertical) polarization and

$$I_{pp}^n = (2k_z)^n f_{pp} \exp\{-k_z^2 \sigma^2\} + k_z^n \frac{[F_{pp}(-k_x, 0) + F_{pp}(k_x, 0)]}{2} \quad (2)$$

where σ is the surface root mean square(rms) height, k is the wave number, $k_z = k \cos \theta$, θ is the incidence angle, f_{pp} and F_{pp} are field coefficients which can be found in [6]. $W^{(n)}(k_x, k_y)$ is the Fourier transform of the n th power of a known surface correlation function which can be calculated by

$$W^{(n)}(k_x, k_y) = \frac{1}{2\pi} \iint \rho_n(x, y) \exp\{-jk_x x - jk_y y\} dx dy \quad (3)$$

where $\rho_n(x, y)$ is the surface correlation function.

B. Least Square Support Vector Machine (LS-SVM)

LS-SVM, put forward by J. A. K. Suykens [12], is different from the original support vector machine. Specifically, LS-SVM using least-square linear system as loss function, thus inequality constraints can be converted to equal constraints. Its advantages over traditional SVM can be shown in the following aspects: a) Simpler operation; b) fast convergence; c) high accuracy; d) small training sample is needed.

The corresponding algorithm goes as follows [12]:

Assume k to the training sample

$$(x_i, y_i) \quad x_i \in R^n, y_i \in R, i = 1, \dots, k$$

Provides the optimization problems as follows:

$$\min_{\omega, \alpha, \varepsilon} J(\omega, \zeta) = 0.5(\|\omega\|^2 + C \sum_{i=1}^k \zeta^2) \quad (4)$$

subject to equality constraint condition:

$$y_i [\omega^T \phi(x_i) + b] = 1 - \varepsilon_i, i = 1, \dots, k \quad (5)$$

According to the above description, the corresponding Lagrange function can be constructed as

$$L(\omega, b, \zeta, \alpha) = J(\omega, \zeta) - l(\omega, b, \zeta, \alpha), i = 1, \dots, k \quad (6)$$

$$\text{where } l(\omega, \alpha, \zeta, \beta) = \sum_{i=1}^k \alpha_i [y_i (\omega^T \phi(x_i) + b) - 1 + \varepsilon_i] \quad ,$$

α_i are the Lagrangian multipliers corresponding to (6). The saddle point is obtained from

$$\max_{\alpha} \min_{\omega, b, \zeta} L(\omega, b, \zeta, \alpha) \quad (7)$$

This yields the Karush-Kuhn-Tucher optimality condition:

$$\begin{cases} \frac{\partial L}{\partial \omega} = 0 \Rightarrow \omega = \sum_{i=1}^k \alpha_k y_k \phi(x_i) \\ \frac{\partial L}{\partial b} = 0 \Rightarrow \sum_{i=1}^k \alpha_k y_k = 0, i = 1, \dots, k \\ \frac{\partial L}{\partial \zeta_i} = 0 \Rightarrow \alpha_k = C \zeta_i, i = 1, \dots, k \\ \frac{\partial L}{\partial \alpha_k} = 0 \Rightarrow y_i [\omega^T \phi(x_i) + b] = 1 - \varepsilon_i \end{cases} \quad (8)$$

And then the optimization problem is transformed into solving linear equations

$$\begin{bmatrix} 0 & I^T \\ I & \Omega + \gamma^{-1} I \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix} \quad (9)$$

$$\text{with } y = [y_1, \dots, y_k]^T, I = [1, \dots, 1]^T$$

$$\Omega = \{\Omega_{ij}\}_{k \times k}, \Omega = \phi(x_i)^T * \phi(x_j), i, j = 1, \dots, k. \quad (10)$$

According to Mercer condition, there are mapping function ϕ and kernel function K , make:

$$K(x_i, x_j) = \phi(x_i) \phi(x_j) \quad (11)$$

Then obtain LS-SVM decision function

$$y(x) = \sum_{i=1}^k \beta_i K(x, x_i) + \alpha \quad (12)$$

Among them, the b and α can be determined by formula (8).

III. DATA SET AND SIMULATION

Generally, the LS-SVM based inversion model need three successive steps: 1) train the model with the data simulating from the IEM; 2) determine the two parameters, C and δ^2 , respectively. The former is the capacity paramount that control the tradeoff between the empirical risk and the model complexity to avoid over-fitting. The latter is the RBF kernel width if RBF kernel is used; 3) test the LS-SVM on a different data set.

A. DATA SET

Considering the fact that LS-SVM does not need huge training patterns as neural network do, it is reasonable to use large interval during the data generating process. The range of soil moisture and surface rough parameters cover the most common field and thus can adapt to different experimental conditions.

As far as the training set is concerned, Table 1 includes various incidence angles that have been involved in the range of the satellite-based SAR, such as ENVISAT-ASAR, and other airborne microwave sensing parameter.

TABLE I. SOIL MOISTURE AND SURFACE PARAMETERS USED IN SIMULATION

Parameters	Min	Max	Interval	Unit
Moisture	2.0	40	2	% by volume
RMS height	0.2	2.0	0.2	cm
Correlation length	2.5	20	2.5	cm
Incidence angle	20	50	15	degree
Correlation function	Exponential			

Altogether 1260 points is included. Similarly, the testing set are generated with moisture ranging from 2.0 to 40.5 with step length of 2.5, RMS height ranging from 0.2 to 2.0 with step length of 0.45, and correlation length ranging from 2.5 to 20 with step length of 4.5, while the incidence angle are the same. The testing data set contains 480 points.

B. SIMULATION

During the process of simulation, the soil moisture was taken as the output and the backscattering simulated by IEM was used as the input, which include the information of vv and hh polarization and different incidence angle. Here only two parameters of the LS-SVM was to be determined, C and δ^2 . More in detail, C varied in range $[10, 10^4]$ and δ^2 from 0.001 up to 3. The best value of the two parameters was determined by its overall error between the retrieval and the reference value of soil moisture. After such a process, the optimal values of the parameters turned out to be $C=1000$ and $\delta^2=0.003$. In the following retrievals, the same parameters have been used.

IV. ANALYSIS OF RESULTS

First of all, it is necessary to find out how the single or multiple frequency bands exert the effect on the accuracy of

moisture inversion. As was shown in Fig. 1, the error of single C band (5.3GHz) is larger than the multi-band of C band and X band (9.3GHz).

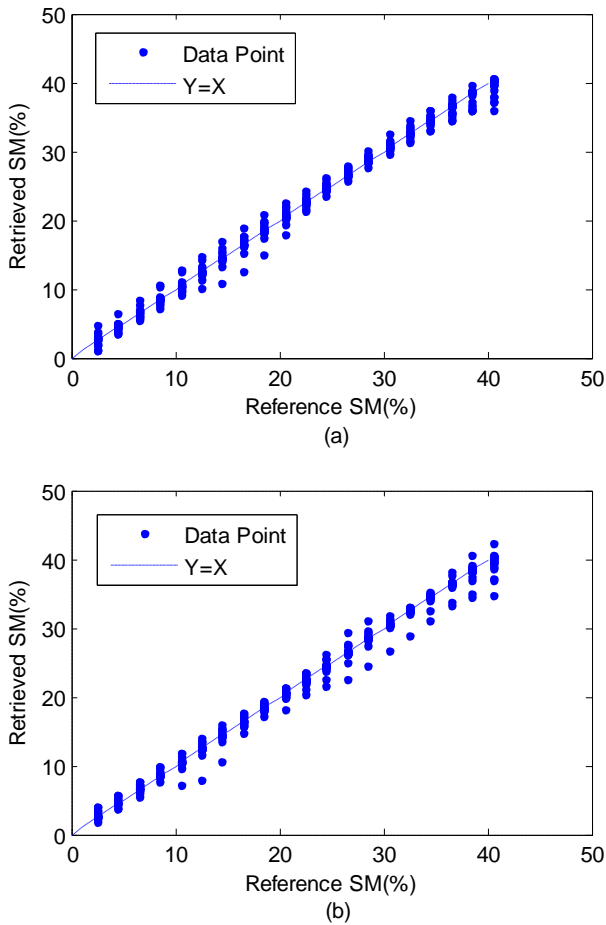


Fig.1. The comparison of inversion: C band-only data (a) and both C and X band data (b).

The C band-only trial shows the root mean squared error (RMSE) and the correlation coefficient between the retrieved soil moisture and the reference is 0.87% and 0.9971 respectively. However, when both C and X band data been used, RMSE and R turned out to be 0.89% and 0.9971. This result shows that multiple frequencies might not make a big difference to the accuracy like expected. Thus, in the following inversion scheme, only C band frequency is conducted. The same scheme was applied to the BP neural network, which was trained under the function 'trainlm' and with the same hidden layer is 6 and 10, respectively. The result of the BP method shows that 1.06% in the C band-only inversion and 1.07% in the C and X band inversion process. It is also worth noting that the BP method yield different result every time and quite time-consuming, over 6.5 seconds compared to the LS-SVM which is less than 1 seconds in the same computing platform. This suggests that the LS-SVM based method is more stable and accurate than BP-method.

To further verify the ability of the approach to estimate soil moisture in noisy condition, Gaussian distributed noises

with standard deviation of 0.01 and 0.001 were added to the C band scheme. In here, the Gaussian distributed noise is taken as the instrumental noise when the data was gathered and processed.

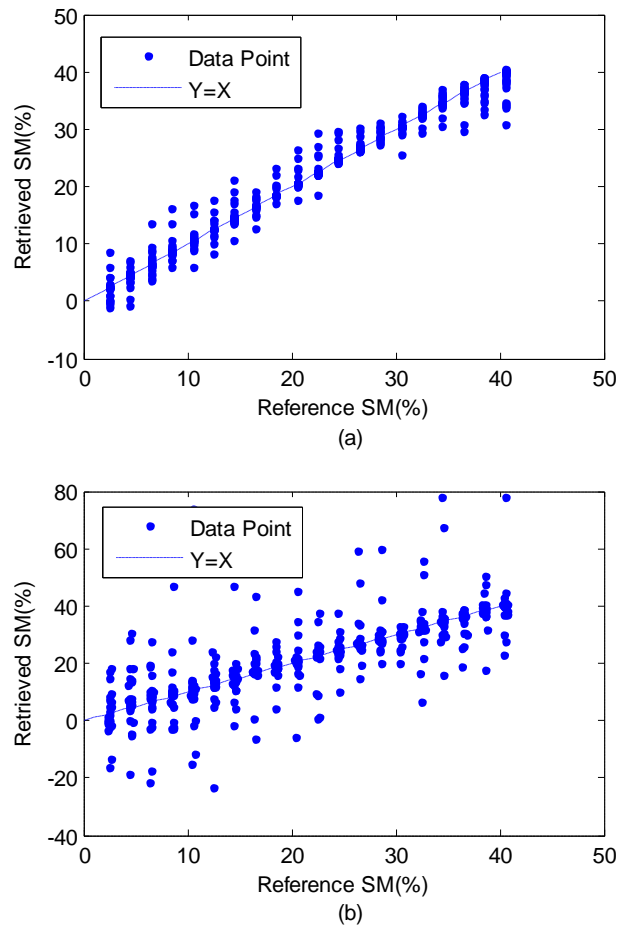


Fig.2. Added with Gaussian distributed noise: standard deviation of 0.001 (a) and 0.01 (b)

Affected by the noise of backscattering data, the RMSE and R turned out to be 1.96% and 0.9854. When the standard deviation of Gaussian noise increase to 0.01, the RMSE and R increase to 9.75% and 0.7217. This was due to the fact that the instrumental noise is comparable to the backscattering coefficient. In the context of application, the data is worthless when the noise is comparable to the signal, which should be averted. Moreover, it is essential to find out to what extend the noise is tolerable. Thus we conducted the inversion scheme when the standard range from 0.001 to 0.01, we find that when the deviation is larger than 0.05, the RMSE is larger 4.81% and the R is less than 0.9170. And it is necessary to point out that the best capacity parameter should be 50 at the noisy condition.

In order to assess the robustness of the method, we only use the former 700 samples of the training data set in the training phase, while the testing data is identical to the formal inversion process. The same scheme is applied to the BP neural network method. The RMSE and R of the LS-SVM and

BP method is 5.16% and 0.9596, 6.15% and 0.9461, respectively.

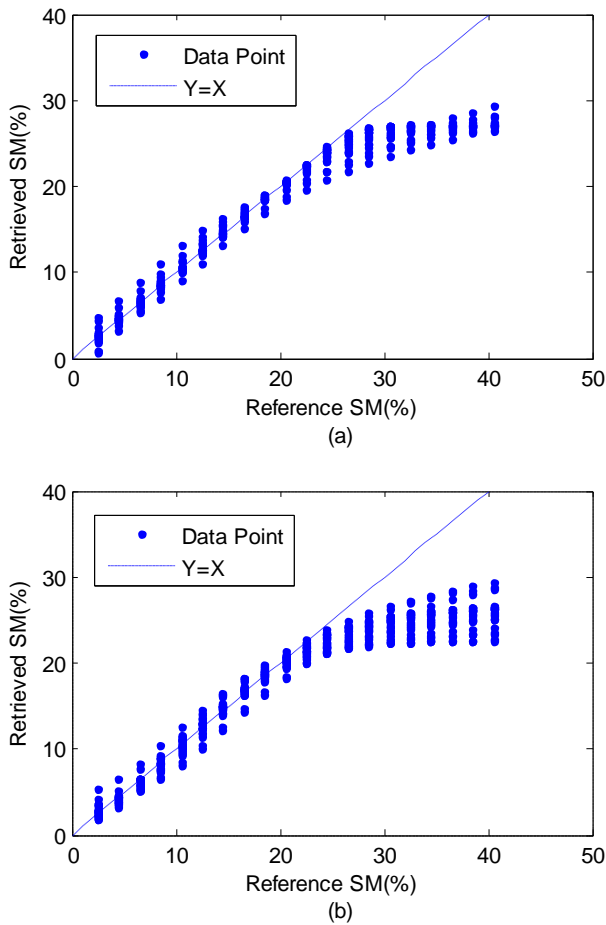


Fig.3. Train the SVM method (a) and BP neural network method (b) with 700 samples to assess their robustness.

V. CONCLUSION

In this paper, an inversion approach has been proposed to retrieve the soil moisture from simulated radar backscattering data. A suitable LS-SVM based strategy has been developed to eliminate the effect of roughness during the inversion process. The effectiveness of the approach has been assessed by considering the noiseless as well as noisy condition. Comparison between LS-SVM and BPNN is also been made in terms of the accuracy and robustness. The obtained results confirmed the advantage of the method in estimating the soil moisture.

Nevertheless, it is essential to add the real data to train and test this approach in the near future. Also, the information derived from both active and passive remote sensing could be integrated into the method to further learn the relation between remote sensing data and soil moisture as well as other parameters. After this, it might meet the need of the upcoming satellite SAR data processing and the future near-real-time soil moisture estimation.

ACKNOWLEDGMENT

Support was provided by “QiuSuo” College Students Innovation Foundation of China Three Gorges University and National Nature Science Foundation of China, No. 61179025.

REFERENCES

- [1]. List F.T.Ulaby, R.K.Moore, and A.K.Fung, *Microwave Remote Sensing: Active and Passive. Radar Remote Sensing and Surface Scattering and Emission Theory*, vol. 2. Reading, MA: Addison-Wesley, 1982.
- [2]. F.T.Ulaby, R.K.Moore, and A.K.Fung, *Microwave Remote Sensing: Active and Passive. From Theory to Application*, vol. 3. Norwood, MA: Artech House, 1986.
- [3]. T.J.Jackson, D.M.Le Vine, C.T.Swift, T.J.Schmugge, and F.R.Schiebe, “Large area mapping of soil moisture using the ESTAR passive microwave radiometer in Washita’92,” *Remote Sens. Environ.*, vol. 54, no. 1, pp.27-37, oct. 1995.
- [4]. E.G.Njoku, T.J.Jackson, V.Lakshmi, T.K.Chan, and S.V.Nghiem, “Soil moisture retrieval from AMSR-E,” *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 2, pp. 215-229, Feb. 2003.
- [5]. N.Baghdadi, M.Zribi, C.Loumagne, P.Ansart, and T.P.Anguela. “Analysis of TerraSAR-X data and their sensitivity to soil surface parameters over bare agriculture fields,” *Remote Sens. Environ.*, vol. 112, no. 12, pp. 4370-4379, Dec.2008.
- [6]. A.K.Fung, Z.Li, and K.S.Chen, “Backscattering from a randomly rough dielectric surface,” *IEEE Trans. Geosci. Remote Sens.*, vol. 30, no. 2, pp. 356-369, Mar. 1992.
- [7]. H. Mametsa, F. Koudogbo, and P. Conbes, “Application of IEM and radiative transfer formulation for bistatic scattering of rough surfaces,” in *Proc. IGARSS, 2002*, pp. 662-664.
- [8]. L. Tsang, J. A. Kong, and K. H. Ding, *Scattering of Electromagnetic Waves*. New York: Wiley, 2000.
- [9]. P. C. Dubois, J. Van Zyl, and T. Engman, “Measuring soil moisture with imaging radars,” *IEEE Trans. Geosci. Remote Sens.*, vol.33, no.4, pp. 915-926, Jul. 1995.
- [10]. K. Song, X. Zhou, and Y. Fan, “Empirically adopted IEM for retrieval of soil moisture from radar backscattering coefficients,” *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no.6, pp. 1662-1672, Jun. 2009.
- [11]. N. Verhoest, B. De Baets, F. Mattia, G. Satalina, C. Lucau, and P. Defourny, “A possibilistic approach to soil moisture retrieval from ERS synthetic aperture radar backscattering under soil roughness uncertainty,” *Water Resources Res.*, vol.43, no. 7, p. W07435, Jul. 2007.
- [12]. J. A. K. Suykens, *Least Squares Support Vector Machines*. World Scientific, Singapore, 2002.
- [13]. Li, Zongqian, Yuhua Tu, and Ning Liu. “Inversion of the soil moisture based upon neural network.” *Antennas, Propagation and EM Theory, 2000. Proceedings. ISAPE 2000. 5th International Symposium on. IEEE, 2000.*