ESTIMATING SOIL MOISTURE USING POLSAR DATA: A MACHINE LEARNING APPROACH

E. Khedri a, M. Hasanlou a, A. Tabatabaeenejad b

a School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Tehran, Iran - (esmailkhedri, hasanlou@ut.ac.ir)
b Department of Electrical Engineering – Electrophysics, University of Southern California, Los Angeles, California, USA - (alirezat@usc.edu)

KEYWORDS: Soil moisture, Sequential Forward Selection, sequential backward selection, support vector regression

ABSTRACT:

Soil moisture is an important parameter that affects several environmental processes. This parameter has many important functions in numerous sciences including agriculture, hydrology, aerology, flood prediction, and drought occurrence. However, field procedures for moisture calculations are not feasible in a vast agricultural region territory. This is due to the difficulty in calculating soil moisture in vast territories and high-cost nature as well as spatial and local variability of soil moisture. Polarimetric synthetic aperture radar (PolSAR) imaging is a powerful tool for estimating soil moisture. These images provide a wide field of view and high spatial resolution. For estimating soil moisture, in this study, a model of support vector regression (SVR) is proposed based on obtained data from AIRSAR in 2003 in C, L, and P channels. In this endeavor, sequential forward selection (SFS) and sequential backward selection (SBS) are evaluated to select suitable features of polarized image dataset for high efficient modeling. We compare the obtained data with in-situ data. Output results show that the SBS-SVR method results in higher modeling accuracy compared to SFS-SVR model. Statistical parameters obtained from this method show an R² of 97% and an RMSE of lower than 0.00041 (m3/m3) for P, L, and C channels, which has provided better accuracy compared to other feature selection algorithms.

1. INTRODUCTION

Due to the abundance of aerial measurement data from different sources, application of remote sensing data has gained a great deal of domain. Vast data distribution in remote sensing in comparison to field data has helped develop remote sensing technology in different fields (Tabatabaeenejad et al., 2015; Narvekar et al., 2015; Oh et al., 1992). One of such fields is an estimation of soil moisture from radar data. Scientists have proven that there is a significant effect in soil moisture anomaly and local weather condition, thus, providing accurate soil moisture data helps better understanding of the local weather condition.

Radar potential with synthetic aperture radar (SAR) for calculating soil parameters has been known for more than thirty years (Tabatabaeenejad et al., 2015; Narvekar et al., 2015; Oh et al., 1992). Therefore, finding better solutions in predicting soil moisture based on radar data with synthetic aperture is of significant importance. Polarimetric synthetic aperture (PolSAR) images in both vast territory and high spatial resolution has made it an effective tool for calculating soil moisture compared to passive data (Narvekar et al., 2015). Effects of vegetation and intensity of signals affect the intensity of active data more significantly than passive data, and this characteristic is used for predicting soil moisture (Narvekar et al., 2015). Besides, lack of databases for these parameters on calculating soil moisture data, soil effects decomposition, and vegetation on redistribution coefficient (D₀) is considered as a major hindrance for varied applications (Ahmad et al., 2010). To achieve accurate soil moisture estimation and avoid above-mentioned difficulties, a data-driven model is needed that can efficiently relate the inputs to the desired output and that is not computationally intensive (Ahmad et al., 2010).

Artificial neural networks (ANN) are models assimilated based on humans learning the ability. These models are more powerful than noise data and they are able to provide a nonlinear multivariable relation between variables (Twarakavi et al., 2006).

Lately, another data-oriented model such as support vector machine (SVM), which has a considerable popularity in the area of ANN, has gained popularity among researchers (Lin et al., 2009; Kalra and Ahmad, 2009). The SVM has been recognized as a basic kernel learning method and has been utilized broadly after successful application in the model identification and regression in different fields such as bioinformatics and artificial intelligence (Barh et al., 2015).

Lyn et al. used SVM model for predicting windstorms and hourly rains in catchment basins in northern Taiwan and compared obtained data with ANN model. They demonstrated the superiority of SVM model over the ANN model (Lin et al., 2009). Carla and Ahmad used SVM model for predicting long-term sailing ship guidance with regards to oceans fluctuations in Colorado Riverfront (Kalra and Ahmad, 2009).

Gail et al. applied SVM model to data over four and a half days for predicting soil moisture based on aerial variables, and compared the results with ANN model. Also, in their research, they concluded that SVM model has obtained a higher degree of accuracy in prediction compared to ANN model (Gill et al., 2006).

Ahmad et al. have studied soil moisture downstream of Colorado River in the western United States based on data obtained from remote sensing using SVM regression technique for ten sites and compared results with forwarding and backward ANN model. They demonstrated that SVM model provides a better prediction for soil moisture compared to ANN and multivariate linear regression (MLR) models (Ahmad et al., 2010).

Sarti et al. also calculated soil moisture with high accuracy. They applied polarimetric extraction technique of AIRSAR in C and L channels. This technique was applied after surface rigidity filtration and differentiation of vegetation from the surface of the earth (Sarti and Mascolo, 2012).

On the other hand, many polarized properties for classifying land use/cover are obtainable from PolSAR images (Lardeux et al.,...
2.2 Support vector evaluation method

Original support vector machine (SVM) algorithm was proposed by Viping (Viping, 1987). SVM is an assimilated method of classification, regression and it is used for other training procedures. SVM take data to a new space based on predetermined category, in which data are dispersed and categorized in a linear fashion. Then a linear equation is meant to be found after searching for support vectors, the equation supports the highest margin between categories. Obtained data from these methods have high stability besides high accuracy (Viping, 1987). One type of SVM is support vector regression (SVR) which utilizing assimilated data for proposing a model and utilizing such models for predicting test data. The quality of SVM and SVR models is dependent on proper regulation of modeling of support vector machine (Nikraftar and Hasanlou, 2015). SVR has been identified as a reliable method for the last two decades (Ahmad et al., 2010; Nikraftar and Hasanlou, 2015). In this method, the aim is to evaluate an unknown function based on a limited number of samples. In SVR, entering $X$ in a space with $M$ number varied features is nonlinear and then based on these features a linear model designed from the following equation (Eq.1).

$$f(x,w) = \sum_{i=1}^{m} w_i g_i(x) + b$$  \hspace{1cm} (1)

In Eq.1 $g_i(x)$ refers to a set of linear transition and $b$ is term bias. The legitimacy of regression function is dependent on good gamma and epsilon ($\gamma$, $\epsilon$) parameters selection in kernel function. In recent years there are various methods for determining optimal features in SVM by scientists (Ahmad et al., 2010; Nikraftar and Hasanlou, 2015). In this research we, use a grid search (GS) model to determine suitable features for modeling. In order to choose gamma and epsilon parameters, the $2^n$ range is assigned for each parameter in which the power $n$ is a numeral range from minus to positive value. For evaluating $n$ number, the quality of all sets is evaluated and those parameters with the lowest error are selected as optimal parameters.

2.3 Feature selection method

Feature selection (FS) is one of the important subjects in machine learning and it is considered in statistical pattern recognition (Bf and Ap, 2005). This is essential in many applications (such as in classification) since in these applications there are many features which are impractical or without informational value (Bf and Ap, 2005). In fact, forming a category is of transition and delivery property. This decline in data and regression algorithms helps easier and faster application. In some cases, determination coefficient ($R^2$) can be improved, in other cases, it helps more vivid and concise results from the subject matter (Bf and Ap, 2005). On the other hand, we know that the quality of SVR model depends on selected features. Hence in this research. Choosing unique and suitable features using optimization algorithms is the aim of this study. These features help the model to determine soil surface parameters on radar features to provide the best core base model for evaluating soil moisture. In this study two algorithms, including sequential forward selection (SFS) and sequential backward selection (SBS) (Guo et al., 2011), are utilized for varied feature selection of PolSAR features.
In SFS which is highly simple search algorithm, all features are evaluated based on one single property after the designation of null for all sets of (Guo et al., 2011). Then in proximity to this feature with the best function sample is chosen. Then this feature is evaluated in wider sizes and this cycle continues unless there is a progress in the feature and the lowest quantity remains. The most important drawback in SFS is that the added feature is not deleted from sets of answers when proven unrelated (Aha and Bankert, 1996). In comparison, SBS acts differently and starts its work a set including all features and in each algorithm repetition, the feature chosen by evaluating function is deleted from the set. This continues until the number of features equal a fixed number. Likewise, the other method the main drawback here is the deletion of added feature even if it is suitable (Aha and Bankert, 1996).

3. PROPOSED METHOD

Modeling is based on PolSAR imaging that common methods for predicting soil moisture in remote sensing community (Sarti and Mascolo, 2012). So far there has been some research in the area of soil moisture based on radar features (Sarti and Mascolo, 2012). As mentioned previously the quality of SVR is dependent on proper regulation of SVR parameters. It is essential in this study to optimize SVR parameters in order to model soil moisture. Hence based on input data as mentioned in Table 1, two scenarios are suggested. In the first scenario, all SVR parameter are optimized based on using all PolSAR features then soil moisture estimation is done. In the second scenario, based on incorporating two selection methods (SFS and SBS,) the best features are chosen. In other parallel procedures, soil moisture model is done without optimization the SVR parameters but with selecting suitable features. Therefore, general procedures in this study follow four main parts, (1) feature generation from AIRSAR dataset, (2) soil moisture modelling based on SVR model and according to optimized features of SFS and SBS without optimizing SVR parameters, (3) soil moisture modelling based on SVR model and according to optimized features of SFS and SBS with optimizing SVR parameters and (4) comparing accuracy of different presented models (Figure 1).

4. STUDY AREA AND DATA-SETS

4.1 Study area

The study area is located in the south of Oklahoma in United State and also is covered with vegetation (Figure 2). This area is hot and dry in summer and moderate in winter (Miralles et al., 2010). Data obtained from these areas are gathered from the airborne platform and in-situ during soil moisture test in 2003 (Table 2).

![](image)

Figure 2. Study area in south Oklahoma.

| Study site | Location | Climate | Vegetation |
|------------|----------|---------|------------|
| South of Oklahoma | Top 35.0239N | Hot and dry | Grass |
| South of Oklahoma | Right 34.8693W | Sub-humid | Grass |
| South of Oklahoma | Left 97.7197W | | |

4.2 Remotely sensed datasets

For incorporating in-situ data in this study, information gathered from SMEX03 campaign and was obtained in 2003 July 10(’http://nsidc.org/data/amsr_validation/soil_moisture’). The amount of soil moisture was measured and analyzed in fourteen locations in Oklahoma. In this study among surface points, eight points were chosen randomly for model training and testing was done for remaining points. Also, PolSAR dataset acquired by AIRSAR instrument from this area. This dataset has a resolution of 6.6 meters in the angular direction. Every pixel in AIRSAR shows radar backscattering for every obtained vertical and horizontal (VV, HH, VH, and HV) polarizations. Each pixel includes backscattering information in three channels, C (5.31 GHZ), L (1.26 GHZ) and P (0.45 GHZ) as illustrated in Figure 3.

![](image)

Figure 3. Location of in-situ data points and their distribution in AIRSAR scene.

5. EXPERIMENT AND RESULTS

Since soil moisture modeling depends on radar features and on SVR parameters, hence in this area we evaluate soil moisture considering all features, features with optimization of SFS and features with optimization of SBS in two states of SVR parameter optimization and no optimization of SVR.

5.1 Soil moisture estimation with all features

As mentioned previously, the quality of SVR model results is dependent on suitably selected features. In this regards,
optimized SVR model features based on all features obtained from radar imaging according to Table 1. Then estimation of soil moisture for three microwave channels AIRSAR image is done. In this research, there is 26 field point from which 18 points are for SVR model training and 8 point is for testing of this model. Obtained results for three microwave channel images AIRSAR are in agreement with Table 3.

Table 3. Results of SVR model with all features with optimizing SVR parameters

| band | R² | RMSE | SSE |
|------|----|------|-----|
| C    | 91.31 | 94.39 | 0.0161 | 0.0153 | 0.3884 | 0.3835 |
| L    | 95.32 | 93.80 | 0.0140 | 0.0147 | 0.3376 | 0.3698 |
| P    | 98.99 | 97.90 | 0.00041 | 0.00049 | 0.0104 | 0.0120 |

As it clears from obtained results, the proposed SVR model with considering P channel AIRSAR soil moisture image provide better modeling comparing to that C and L AIRSAR images.

5.2 Estimation by optimal features with FS

5.2.1 SFS algorithm

In this scenario, four suitable features are chosen according to Table 4 are chosen for all selected features for proposing a support vector model based on SFS algorithm.

Table 4. Selected features using optimization algorithms SFS

| Number | Name Feature                  |
|--------|------------------------------|
| X₁     | Freeman_Db1                  |
| X₂     | beta                         |
| X₃     | Huynen_T33                   |
| X₄     | Yamaguchi3_Odd               |

Results of support vector model without optimizing SVR parameters are as Table 5 considering features that are chosen with optimized SFS algorithm and core model results are as for Table 6 with optimizing SVR model.

Table 5. Results of SVR model with optimal features selection with SFS algorithm without optimizing SVR parameters

| band | R² | RMSE | SSE |
|------|----|------|-----|
| C    | 62.62 | 53.59 | 0.0301 | 0.0299 | 0.7278 | 0.7507 |
| L    | 85.96 | 85.24 | 0.0147 | 0.0167 | 0.3542 | 0.4200 |
| P    | 40.78 | 39.79 | 0.0314 | 0.0319 | 0.8554 | 0.8580 |

Table 6. Results of SVR model with optimal features selection with SFS algorithm with optimizing SVR parameters

| band | R² | RMSE | SSE |
|------|----|------|-----|
| C    | 71.26 | 58.00 | 0.026 | 0.0277 | 0.6284 | 0.6949 |
| L    | 90.27 | 86.29 | 0.0137 | 0.0152 | 0.3301 | 0.3826 |
| P    | 95.53 | 95.09 | 0.0116 | 0.0092 | 0.2800 | 0.2323 |

As it is conceivable, better soil moisture modeling obtains when SVR parameters are optimized compared to the state in which SVR parameters are not optimized based on optimal values SFS algorithm proposed. Also, obtained results here show that proposed SVR model with considering P-channel AIRSAR soil moisture image provide better modeling that C and L AIRSAR images.

5.2.2 SBS algorithm

In this scenario, ten suitable features are chosen according to Table 7 are chosen for all selected features for proposing a support vector model based on SBS algorithm.

Table 7. Selected features using optimization algorithms SBS

| Number | Name Feature                  | Number | Name Feature                  |
|--------|------------------------------|--------|------------------------------|
| X₁     | Barnes1_T22                  | X₂     | TSVM_alpha_s                 |
| X₃     | Barnes1_T33                  | X₄     | TSVM_psi1                    |
| X₅     | Barnes2_T22                  | X₆     | Yamaguchi4_Y4O_Vol           |
| X₇     | Cloude_T22                   | X₈     | Holm2_T22                    |
| X₉     | Yamaguchi4_YAO_Hlx           |        |                              |

Results of support vector model without optimizing SVR parameters are as Table 8 considering features that are chosen with optimized SBS algorithm and core model results are as for Table 9 with optimizing SVR model.

Table 8. Results of SVR model with optimal features selection with SBS algorithm without optimizing SVR parameters

| band | R² | RMSE | SSE |
|------|----|------|-----|
| C    | 84.94 | 88.51 | 0.0154 | 0.0151 | 0.3725 | 0.3791 |
| L    | 80.29 | 73.38 | 0.0253 | 0.0258 | 0.6109 | 0.6480 |
| P    | 73.64 | 66.72 | 0.0224 | 0.0236 | 0.5803 | 0.5939 |

Table 9. Results of SVR model with optimal features selection with SBS algorithm with optimizing SVR parameters

| band | R² | RMSE | SSE |
|------|----|------|-----|
| C    | 98.24 | 98.97 | 0.0058 | 0.0047 | 0.1404 | 0.1149 |
| L    | 91.26 | 79.27 | 0.0204 | 0.0221 | 0.4918 | 0.5549 |
| P    | 97.93 | 97.40 | 0.00041 | 0.00047 | 0.0115 | 0.0105 |

As it is conceivable, better soil moisture modeling obtains when SVR parameters are optimized compared to the state in which SVR parameters are not optimized based on optimal values SBS algorithm proposed. Also, obtained results here show that proposed SVR model with considering P-channel AIRSAR soil moisture image provide better modeling that C and L AIRSAR images.

6. CONCLUSION

In this study tried to utilize data obtained from two optimal selection algorithms of SFS and SBS in three channels C, L and P which are related to 2003 soil moisture project in Oklahoma region in the United States. In the first scenario, optimal feature selection algorithms for soil moisture core model are compared. Among these models, support vector model has proven to more accurately evaluate soil moisture modeling. It has used SBS algorithm and considered all features and other features that selected by SFS algorithm. In the second scenario, optimal input features for support vector model when SVR parameter are not optimized and when SVR parameter are optimized are compared. Evaluating soil moisture and results show that the SVR model with optimized parameters has modeled soil moisture more accurately compared to the case in which SVR parameters are not optimized. In the third scenario, different soil moisture evaluation models in different channels are compared. Among three channels in AIRSAR data, P-band more accurately model soil moisture in all utilized models and L and C bands come later. The C-band has shown to provide inaccurate estimations comparing to other channels, it also provides a weaker result in optimal selective algorithm SBS among another algorithm for soil moisture modeling.

REFERENCES

Aha, D.W., Bankert, R.L., 1996. A Comparative Evaluation of Sequential Feature Selection Algorithms, in: Fisher, D.,
Lenz, H.-J. (Eds.), Learning from Data, Lecture Notes in Statistics. Springer New York, pp. 199–206. doi:10.1007/978-1-4612-2404-4_19
Ahmad, S., Kalra, A., Stephen, H., 2010. Estimating soil moisture using remote sensing data: A machine learning approach. Adv. Water Resour. 33, 69–80. doi:10.1016/j.advwatres.2009.10.008
Attema, E.P.W., Ulaby, F.T., 1978. Vegetation modeled as a water cloud. Radio Sci. 13, 357–364. doi:10.1029/RS013i002p00357
Barh, D., Khan, M.S., Davies, E., 2015. PlantOomics: The Omics of Plant Science. Springer.
Bf, S., Ap, C., 2005. Gene selection based on multi-class support vector machines and genetic algorithms. Genet. Mol. Res. GMR 4, 599–607.
Cloude, S.R., Pottier, E., 1996. A review of target decomposition theorems in radar polarimetry. IEEE Trans. Geosci. Remote Sens. 34, 498–518. doi:10.1109/36.485127
Freeman, A., Durden, S.L., 1998. A three-component scattering model for polarimetric SAR data. IEEE Trans. Geosci. Remote Sens. 36, 963–973. doi:10.1109/36.673667
Gill, M.K., Asea, T., Kembelowski, M.W., McKee, M., 2006. Soil Moisture Prediction Using Support Vector Machines. JAWRA J. Am. Water Resour. Assoc. 42, 1033–1046. doi:10.1111/j.1752-1688.2006.tb04512.x
Guo, L., Rivero, D., Dorado, J., Munteanu, C.R., Pazos, A., 2011. Automatic feature extraction using genetic programming: An application to epileptic EEG classification. Expert Syst. Appl. 38, 10425–10436. doi:10.1016/j.eswa.2011.02.118
Haddadi G., A., Reza Sahebi, M., Mansourian, A., 2011. Using oceanic radar data for land use discrimination based on SVM classification. Water Resour. Res. 33, 69–80. doi:10.1007/11053-006-9013-6
Jafari, M., Maghsoudi, Y., Zoeg, M.I.V., 2015. A New Method for Land Cover Characterization and Classification of Polarimetric SAR Data Using Polarimetric Signatures. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 8, 3595–3607. doi:10.1109/JSTARS.2014.2387374
JONG-SEN LEE, ERIC POTTIER, “Polarimetric Radar Imaging: From Basics to Applications,” 1st edition, February 2009, pp 422-ISBN: 978-142005492
Kalra, A., Ahmad, S., 2009. Using oceanic-atmospheric oscillations for long lead time streamflow forecasting. Water Resour. Res. 45, W03413. doi:10.1029/2008WR008655
Lardeux, C., Frison, P.-L., Tison, C., Deleflie, D., Souyris, J.-C., Rudant, J.-P., Stoll, B., 2007. Comparison of Compact Polarimetric with Full Polarimetric Radar Data for Land Use Discrimination Based on SVM Classification. Presented at the Science and Applications of SAR Polarimetry and Polarimetric Interferometry, p. 19.
Lin, G.-F., Chen, G.-R., Wu, M.-C., Chou, Y.-C., 2009. Effective forecasting of hourly typhoon rainfall using support vector machines. Water Resour. Res. 45, W08440. doi:10.1029/2009WR007911
Mao, K.Z., 2004. Orthogonal forward selection and backward elimination algorithms for feature subset selection. IEEE Trans. Syst. Man Cybern. Part B Cybern. 34, 629–634. doi:10.1109/TSMCB.2002.804363
Miralles, D.G., Crow, W.T., Cosh, M.H., 2010. Estimating Spatial Sampling Errors in Coarse-Scale Soil Moisture Estimates Derived from Point-Scale Observations. J. Hydrometeorol. 11, 1423–1429. doi:10.1175/2010JHM1285.1
Narvekar, P.S., Entekhabi, D., Kim, S.B., Njoku, E.G., 2015. Soil Moisture Retrieval Using L-Band Radar Observations. IEEE Trans. Geosci. Remote Sens. 53, 3492–3506. doi:10.1109/TGRS.2014.2377714
Oh, Y., Sarabandi, K., Ulaby, F.T., 1992. An empirical model and an inversion technique for radar scattering from bare soil surfaces. IEEE Trans. Geosci. Remote Sens. 30, 370–381. doi:10.1109/36.134086
Salehi, M., Sahebi, M.R., Maghsoudi, Y., 2014. Improving the Accuracy of Urban Land Cover Classification Using Radarsat-2 PolSAR Data. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 7, 1394–1401. doi:10.1109/JSTARS.2013.2273074
Sarti, M., Mascolo, L., 2012. An investigation of different polarimetric decomposition techniques for soil moisture estimation, in: 2012 Tyrrhenian Workshop on Advances in Radar and Remote Sensing (TyWRRS). Presented at the 2012 Tyrrhenian Workshop on Advances in Radar and Remote Sensing (TyWRRS), pp. 209–213. doi:10.1109/TyWRRS.2012.6381131
SMEX03 Data | National Snow and Ice Data Center [WWW Document]. n.d. URL http://nsidc.org/data/amsr_validation/soil_moisture/smex03/index.html (accessed 1.3.17).
Support Vector Machines for Classification and Regression - SVM.pdf, n.d.
Tabatabaeenejad, A., Burgin, M., Duan, X., Moghaddam, M., 2015. P-Band Radar Retrieval of Subsurface Soil Moisture Profile as a Second-Order Polynomial: First AirMOSS Results. IEEE Trans. Geosci. Remote Sens. 53, 645–658. doi:10.1109/TGRS.2014.2326839
Twarakavi, N.K.C., Misra, D., Bandopadhyay, S., 2006. Prediction of Arsenic in Bedrock Derived Stream Sediments at a Gold Mine Site Under Conditions of Sparse Data. Nat. Resour. Res. 15, 15–26. doi:10.1007/s11053-006-9013-6
Zahir, N., Mahdi, H., 2015. Snow Depth Estimation Using Time Series Passive Microwave Imagery via Genetically Support Vector Regression (case Study Urmia Lake Basin). ISPRS - Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. 15, 555–558. doi:10.5194/isprsarchives-XL-1-W5-555-2015
Zhang, T., Fu, X., Goh, R.S.M., Kwoh, C.K., Lee, G.K.K., 2009. A GA-SVM feature selection model based on high performance computing techniques, in: 2009 IEEE International Conference on Systems, Man and Cybernetics. Presented at the 2009 IEEE International Conference on Systems, Man and Cybernetics, pp. 2653–2658. doi:10.1109/ICSMC.2009.5346120