SPACEBORNE GNSS- R RETRIEVING ON GLOBAL SOIL MOISTURE APPROACHED BY SUPPORT VECTOR MACHINE LEARNING

Aung Lwin^{1,2*}, Dongkai Yang¹, Xuebao Hong³, Sara Cheraghi Shamsabadi⁴, Wasiu Akande Ahmed¹

¹ School of Electronic and Information Engineering, Beihang University, 100083, China -(aunglwin, edkyang, ahmedwasiu) @buaa.edu.cn

² Remote Sensing and GIS Research Centre, Yangon Technological University, Myanmar - ual@ytu.edu.mm ³ School of Electronic and Information Engineering, Beihang University, 100083, China - joyce_hong2008@yeah.net ⁴School of Electronic and Information Engineering, Beihang University, 100083, China - sara.cheraghi@outlook.com

Commission III, WG III/6

KEY WORDS: TDS-1; DDM, NDVI, SMOS, SVM, Soil Moisture

ABSTRACT:

GNSS Reflectometry system is an excellent to sense soil moisture content. In recent, GNSS-R technique could be aided to detect soil moisture contents but still have many difficulities issues, most especially vegetation impact. Soil moisture observing is a major concept for enhancing the sustainability of the earth's system and process. On retrieving soil moisture from spaceborne GNSS-R technology has been challenging to the system, retrieving model and geophysical parameters. In this research, we use the Support Vector Machine (SVM) method to retrieve global soil moisture, the TDS-1 Delay Doppler Map (DDM) and the AVHRR Normalized Difference Vegetation Index (NDVI) imagery as inputs and the Soil Moisture and Ocean Salinity (SMOS) soil moisture data as a reference to retrieve global SM daily basis. The results have shown that the squared correlation coefficient (R) values are much higher in TDS-1 fused with NDVI than using DDM alone, which indicates that vegetation impact has effectively weakened. The feasibility of this approach could provide the performance for spaceborne GNSS-R retrieving to soil moisture analysis.

1. INTRODUCTION

GNSS-R technology has tremendously succeeded in earth observations since the recent decades. Spaceborne GNSS-R platforms as TDS-1 and CYGNSS have been successfully launched and aided in retrieving useful information on the areas of sea and land applications. In this GNSS-R concept, both together with transmitter and receiver becomes bistatic radar by adequately retrieving the reflected signals. This configuration can be used for both of an altimeter or a scatterometer, to investigate the interesting characteristics as surface roughness, or dielectric properties and sub-surface features. GNSS- R technology proved to be highly productive and efficient for monitoring the earth's environments, measuring significant wave heights, wind speeds, surface roughness, ice and snow thicknesses and soil moisture conditions.

However, despite these advantages, there are still challenges in retrieve techniques on data assimilation, verification and accuracy assessments in line with a temporal and spatial capsule. Many soil moisture techniques determination employed, ranging from multi-disciplinary data sources as well as machine learning approaches. GNSS- R for soil moisture observation has utilized in-ground and air-based. Camps, A., et al. (2016) initiated spaceborne GNSS- R sensitivity to soil moisture using TechDemoSat-1 data preliminary results showed a good correlation with soil moisture. Recently, Senyurek, V., et al. (2020) calculated with NASA CYGNSS data are compared and analyzed for the soil moisture retrieval through utilizing multiple validation strategies including SVM has RMSE values 0.065 cm3/cm3. Observing soil moisture is a major factor in hydrological practice, which influence on evapotranspiration, run-off and infiltration on the land surface. It has significantly improved the prediction of soil moisture with short, medium and huge changes in climate, and hazardous crisis such as in disaster mitigated land. Raghu Garg et al. (2019) process using with several machines techniques figured out to extract knowledge from big data learning methods for sustainability on plant-related studied. In this research, attempts have using SVM machine learning methods to retrieve global soil moisture, taking the TDS-1 DDM input and the SMOS SMC as reference. This paper is categorized as follows: section 2 describes the specific data preparation; section 3 states the data processing and next is the discussion; finally, the conclusion in section 4.

2. DATA PREPARATION

In this research, we used UK TDS 1 data and AVHRR NDVI data as inputs and SMOS data which is an ESA Explorer Opportunity Science Mission as reference. All data are prepared monthly from January to March 2017 and collocated.

2.1 Soil Moisture and Ocean Salinity (SMOS)

SMOS is one of ESA's missions, and the satellite brings radiometer that operates in the microwave L band range to detect brightness temperature images. The images over land used to derive global maps of SMC every three days, achieving an accuracy of 4% at a spatial resolution of about 50 km. A new reprocessed level of SMOS SSM product released from the CATDS Centre, i.e. SMOS level 3 products (SMOSL3), which provided as global gridded maps of SSM. It contains surface soil moisture conditions and land information in a daily product. Soil moisture data generated with two products per day: one for ascending and one for descending orbits. Finally, the three months products aggregated, and it provides retrieving to soil moisture from spaceborne observation.

Corresponding author



Figure 1. SMOSL3 monthly data from January to March 2017

2.2 Normalized Differential Vegetation Index (NDVI)

NOAA Climatic satellites with AVHRR sensor record the NDVI daily product. The NDVI produces daily output and obtains time-series data for vegetation observation. NDVI is ubiquitous as an index of vegetation. Based on remote sensing observation, the monitoring of vegetation occurrences takes place via times series data analysis and big data processing systems. These systems may use a pixel or object-based algorithms to examine vegetation health, evapotranspiration, and other ecosystem functions as agricultural concerned. NDVI observation can make sustainability of earth's vegetation and enhancing the climatic condition as well.



Figure 2. AVHRR NDVI monthly data of January to March 2017

2.3 UK Test DemoSat (TDS-1)

The spaceborne GNSS receiver received the reflected signals, which is flying on-board TechDemoSat-1satellite. The receiver can process various simultaneous tracks, from multiple GNSStransmitters. GNSS reflections are not only sensitive to the ocean, but also for the land information, triggering to the other potential new opportunities for remote sensing and estimating the thickness of sea ice, snow depth, soil moisture levels, and the classification of vegetative foliage.

Camps, A., et al. (2016) pointed out the presence of vegetation indicates that it attenuates the GNSS signal, reducing the reflection coefficient, and the sensitivity to soil moisture, which is still enough for remote sensing from space. They detail analysed vegetation impacts on SM as the more vegetation cover (NDVI increases), the reflective SNR and the sensitivity to soil moisture and the Pearson correlation coefficient decreases; however, it is still significant.

Camps, A., et al. (2018) analysed retrieval to surface and subsurface soil moisture by using four years TDS-1 data and SMOS with different spatial scales at global and regional observation. They found that GNSS-R observation to soil moisture retrieving performance is ~0.09 dB %. However, this is relying on the spatial scale used for the ground-truth and the selected region. Park, H., et. al., (2018) developed the DDM simulator using TDS-1 and CYGNSS data shown that the DDM over land varies according to the glistening zone characteristics, e.g., the mixture of surface type, topography, soil moisture, and vegetation, etc.



Figure 3. Averaged SNR of DDM for TDS L1B monthly data

For the sufficiently diffuse scattering of the GNSS signal from a surface roughness, V. U. Zavorotny et al. initiated the following bistatic radar equation holds for the correlation power as a function of the time delay and the frequency offset, namely the Delay-Doppler Map (DDM) as below.

$$\langle |Y(\tau, f)|^2 \rangle = \frac{\lambda^2 T_i^2}{(4\pi)^3} P_t G_t \iint \frac{G_r}{R_t^2 R_r^2} \chi^2(\tau, f) \delta_0 \, dS \qquad (1.1)$$

Where, $T_i =$ the coherent integration time,

 P_tG_t = the transmitter's capable isotropic radiated power G_r = the receiver antenna gain pattern,

 R_t , R_r = distances between the nominal specular point and the transmitter/receiver,

 χ^2 = the Woodward Ambiguity Function (WAF), which describes the range and Doppler selectivity of the coherent radar.

 σ_0 = the normalized bistatic radar cross-section (BRCS) of the rough surface.



Figure 4. SNR and Antenna Gain Histogram for the DDM data

In the above figure, Delay Doppler Map obtained by the collocated three months data of UK TDS-1 from January to March 2017. SNR DDM observation can vary how the reflected signal received from the receiver relies on the ground surface topography and other geophysical parameters influenced.

2.4 Support Vector Machine (SVM)

SVM supervised learning technique used for data regression and classification analysis. In other terms, SVM is a prediction tool that maximizes estimated accuracy while automatically omitting overfit to the data. SVM trained with a learning algorithm that executes a learning bias derived from statistical theory. SVM is very popular utilized in machine learning research around the world. SVM becomes popular and, it gives exact correctness comparable to complex neural networks with further details features. A proposed explanation that performed the accurate classification of the training samples, and new learning algorithms designed to find such precise fit to the data.



Figure 5. Maximum margin hyperplane (Bahga & Madisetti)

SVM training involves deciding the maximum margin hyperplane that divide the two classes. The maximum margin hyperplane is one which has the most significant separation from the nearest training data point. Figure 5 shows the margins for an SVM. Given a training data set (x_i, y_i) where x_i is an ndimensional vector and $y_i = 1$. If x_i is in class 1 and $y_i = -1$ if x_i is in class 2. A standard SVM finds a hyperplane w.x-b = 0, which correctly separates the training data points. It has a maximum margin which is the distance between the two hyperplanes w.x-b =1 and w.x-b = -1 as shown in figure 5. The optimal hyperplane which a maximum margin can obtain with the following quadratic programming problem,

$$Min\{w, b\} \frac{1}{2} ||w||^2 + C \sum_{i=1}^{l} \xi_i$$
(1.2)

subject to $y_i(w.x_i-b)=\geq 1-\xi_i$, $\xi_i > 0$, 1 < i < l where *C* is the soft margin parameter and ξ is a slack variable for the non-separable case. The optimal hyperplane to be obtained as,

$$f(x) = sign(c\sum_{i=1}^{l} \alpha_i y_i K(x_i, x) - b)$$
(1.3)

Where, α_i is the Lagrange multiplier, and $K(x_i, x)$ is the kernel function. A standard SVM is a two-class classier where the outcome is 1 or -1. When sets are non-linear separable, the data points in the initial finite-dimensional space classed to an important dimensional feature where it can be separated easily. The accuracy performance of an SVM classifier depends on the selection of kernel, the kernel's parameters, and soft margin parameter *C*. There are two kinds of Kernel RBF functions; Gaussian and Exponential. In this research, we use Gaussian Radial Basis Function with a Gaussian form,

$$k(x_i, x_j) = exp\left(-\gamma \|x_i - x_j\|^2\right)$$
(1.4)

After the model has been built, the predict method is used to make the predictions.

2.5 Controlling Complexity

SVM is a robust technique to evaluate any training samples and generalizes preferable on given datasets. The complexity in terms of kernel affects the accomplishment of new datasets. For controlling complexity, SVM supports parameters should be able to determine by cross-validation on the given datasets. Choosing the right kernel-related with the problem or application would enhance SVM's performance. The following diagram gives a better understanding of how to control in complexity process carried out.



Figure 6. Controlling Complexity (Andrew W. Moore)

2.5.1 SVM Classification: Although, it can consider that neural networks are more applicable to use than SVM technique, however, sometimes unsatisfactory results obtained. A classification scheme has usually involved training and testing data involving some data instances. Each occurrence in the training set concluded one of the target values and different attributes. SVM intends to build a model that shows the target value of information instances in the testing set, which gives only the data attributes. SVM classification is an example of one of the supervised learning techniques. Known labels help to show whether the system is performing accurately or not. This information points guides to the desired feedback, validating the correctness of the system, and improve the network learn to act accuracy performance.

2.5.2 SVM prediction on Soil Moisture: In this research, Support Vector Machines (SVMs) are used to develop these predictive soil moisture models and then drive soil moisture retrieving on time series data analysis. Vapnik (1995) developed the statistical learning strategy theory that SVMs can be used to predict a quantity forward in time with the results of "training" data. There are two critical factors to enhance the generalization capability of the learning machine. The training error rate is the first factor, and the second one is the potentiality of the learning machine estimated according to Vapnik–Chervonenkis (VC) dimensions.

Ahmad, S., et al. (2010) pointed out, and compared soil moisture observation with SVM RMSE value was 0.34 to 0.77 tested in selected sites 2% less and more predicted than other techniques such as an Artificial Neural Network (ANN) and Linear Regression model (MLR).

Ren, C., et al. (2019) figured out GNSS Interferometric Reflectometry data to the correlation with received SNR and soil moisture estimation using least square- SVM method. They concluded that LS-SVM based on multi-satellite fusion results is a more accurate estimation on retrieving soil moisture than single satellite means single station. Garg, R., et al. (2019) extracted big data for the sustainability of soil nutrition composition comparatively analyzed with machine learning techniques as support vector machine (SVM) using the polynomial function, radial basis function (RBF) methods and others.

3. DATA PROCESSING AND RESULTS DISCUSSION

3.1 Data Processing

The following figure presents the processing flow of this work. On the one hand, SMC calculated using both TDS-1 data and NDVI data. On the other hand, the retrieval also carried out using TDS-1 data alone as a comparison. The data used in this work collected from January to March 2017.



Figure 7. Data Processing Flow of Retrieving Soil Moisture

The data processing carries out as follows. First, TDS-1 DDM SNR data has filtered according to these principles: (1) The data whose corresponding antenna gain is below 10dB is filtered out to guarantee the data quality; (2) the information whose coordinates located in the ocean is filtered out. Secondly, the TDS-1DDM and SMOS SMC data are matched in both time and space dimensions to ensure the rationality of the retrieval. The next step is to fuse TDS-1 DDM and the NDVI imagery as input and the SMOS soil moisture data as a reference. Again, TDS-1DDM SNR only used for input and to give room for results comparison with SMOS data as a reference. They were retrieving SM, a better accuracy result with the SVM method based on input data achieved. These involve removing features from the usual which have low weights to gain a specified level of data sparsity.

Lastly, using only features retained after choosing the feature selection process is done, a representation of the full training set of documents created. The Radial Basis Kernel Function (RBF) SVM in the fixed feature space classify the test data retained. This method planned to take advantage of the memory freed as a result of increased data sparsity and include more big data training sets while keeping the memory consumption constant. Concurrently, the performance controlling the possible negative impact of the reduced feature space calibrated. We here trained 19701 and tested 4925 out of 24626 samples data.

3.2 Results and Discussion

UK TDS 1 and AVHRR NDVI data as inputs and SMOS data as reference datasets used to train SVM-based machine learning classifier.

3.2.1 Training with SVM Classification (without NDVI): The following figures present the data processing results obtained using TDS-1 along and SMOS as reference dataset according to training and testing tasks.



Figure 8. SVM classification with a training dataset



Figure 9. SVM classification with a testing dataset

In the above figures 8 and 9, the data optimization finished with iteration 34214 times and the respective values obtained are as follows: nu = 0.996238, obj value= 4068.650054, rho = 0.027386, nSV = 19640 and nBSV = 19614. After the prediction task has done, the result shows that the mean square error is 0.0827153, and the squared correlation coefficient is 0.102284 for the training dataset. And the mean square error is 0.0779197, and the squared correlation coefficient is 0.0960314 for the testing dataset.

3.2.2 Training with SVM Classification (with NDVI):

The following figures present the data processing results obtained using TDS-1 fused with NDVI data as input and SMOS as reference according to training and testing tasks.





Figure 11. SVM classification with a testing dataset

In the above figure 10 and 11, the data optimization finished with iteration 45525 and the respective values obtained are as follows: nu = 0.995864, obj = -3830.046966, rho = -0.126590, nSV = 19642 and nBSV = 19603. After the prediction task has done, the result shows that the mean square error is 0.0764551, and the squared correlation coefficient is 0.177215 for the training dataset. And the mean square error is 0.0727246, and the squared correlation coefficient is 0.16478 for testing dataset. It is showed that the squared correlation coefficient (R) values with NDVI much higher than without NDVI dataset. Higher correlation coefficient means more correlation to retrieve soil moisture. Specifically, for each calculation, its training performance is superior to its testing stage.

4. CONCLUSION

This work focus on the TDS-1 DDM data and SMOS data to remotely sense the global soil moisture using SVM learning approach. Support Vector Machines acts as one of the best methods for data classification performance. It combines generalization as a technique to control dimensionality. In classification problems, generalization control obtained by maximizing the margin, which corresponds to the minimization of the weight vector in a legal framework. The solution obtained as a set of support vectors that can be sparse. However, retrieval performance is still to be improved. Although ground and airbased observation had shown good results of (0-5) cm soil moisture or vegetation water contents, it had assumed that spaceborne GNSS- R observation would be too weak and uncertainty on soil moisture sensing. Also, in this research, space-borne GNSS-R data to retrieve on soil moisture is not very significant. Data range should extend on yearly analysis. Data, features and models are the essential cores of the multidisciplinary data source fusion driven by the SVM classification method, in terms of time series data analysis. GNSS R technology boost to complement the results of the existing space-based earth sensing techniques, such as SAR or other space-borne data.

ACKNOWLEDGEMENTS

The authors are much grateful to the space-related organizations for their data that freely provided by UK's TDS, NOAA and ESA SMOS mission.

REFERENCES

Ahmad, S., Kalra, A., Stephen, H., 2010: Estimating soil moisture using remote sensing data: A machine learning approach. Advances in Water Resources, 33, 69-80.

Andrew, W.M., 2014: Support Vector Machine tutorial; http://www.cs.cmu.edu /~awm/tutorials.html.

Arshdeep, B., Vijay, M., 2019: Big Data Analytics: A Hands-On Approach. ISBN: 978-1-949978-00-1, pp. 458-460.

Nandintsetseg, B., Shinoda. M., 2011: Seasonal change of soil moisture in Mongolia: its climatology and modelling. Int. J. Climatol., 31 pp. 1143-1152.

Brien, A.O., Johnson, J.T., 2017: Comparing the cygnss simulator forward scattering model with TDS-1 and cygnss onorbit DDMS. 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). Cai, Y., Zheng W., Zhang, X., Zhangzhong, L., Xue, X., 2019: Research on soil moisture prediction model based on deep learning. PLoS ONE 14(4): e0214508.

Camps, A., et al., 2016: Soil moisture and vegetation impact in GNSS-R TechDemosat-1 observations. 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS).

Camps, A., et al., 2016: Sensitivity of GNSS-R Spaceborne Observations to Soil Moisture and Vegetation. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 9(10): 4730-4742.

Camps, A., et al., 2018: Sensitivity of TDS-1 GNSS-R Reflectivity to Soil Moisture: Global and Regional Differences and Impact of Different Spatial Scales. Remote Sensing 10(11): 1856.

Chew, C. C., Small, E. E., 2018: Soil Moisture Sensing Using Spaceborne GNSS Reflections: Comparison of CYGNSS Reflectivity to SMAP Soil Moisture. Geophysical Research Letters, 45(9), 4049-4057. doi.org/10.1029/2018 GL077905

Duda, R., Hart, P., Pattern Classification and Scene Analysis. Wiley, New York 1973.

Deutsch, E., Bork, E., Willms, W., 2010: Soil moisture and plant growth responses to litter and defoliation impacts in Parkland grasslands. Agric. Ecosyst. Environ., 135, pp. 1-9.

Garg, R., et al., 2019: Extracting Knowledge from Big Data for Sustainability: A Comparison of Machine Learning Techniques. Sustainability 11(23).

Han, M., Zhu, Y., Yang, D., Hong, X., Song, S., 2018: A Semi-Empirical SNR Model for Soil Moisture Retrieval Using GNSS SNR Data. MDPI Remote Sens. 10, 280.

Janez, B., Marko, G., Nataša, M.F., Dunja, M., 2002: Feature Selection Using Linear Support Vector Machine. Technical Report MSR-TR-2002-63.

Lwin, A., Yang, D., Hong, X., 2020: Leveraging Bayesian Deep Learning for Spaceborne GNSS-R Retrieval on Global Soil Moisture. International Conference on Artificial Intelligence in Information and Communication (ICAIIC), Fukuoka, Japan, pp. 352-355.

Martin-Neira, M., 1993: A passive reflectometry and interferometry system (PARIS): Application to ocean altimetry. ESA J. 17, 331–355.

MERRByS Product Manual, 2018: GNSS Reflectometry on TDS-1 with the SGRReSI. @SSTL Ref: 0248366 Rev:005.

Milzow, C., Krogh, P., Gottwein, P. B., 2011: Combining satellite radar altimetry, SAR surface soil moisture and GRACE total storage changes for hydrological model calibration in a large poorly gauged catchment. Hydrol. Earth Syst. Sci, 15 pp. 1729-1743.

Nello, C., John., Shawe, T., 2000: An Introduction to Support Vector Machines and Other Kernel-based Learning Methods. Cambridge University Press.

Park, H., et al., 2018: Spaceborne GNSS-R End-To-End Simulator: Topography and Vegetation Effects. IGARSS 2018 –

IEEE International Geoscience and Remote Sensing Symposium.

Reichstein, M., et al., 2019: Deep learning and process understanding for data-driven Earth system science. Nature 566(7743): 195-204.

Ren, C., et al., 2019. "Research on the soil moisture sliding estimation method using the LS-SVM based on multi-satellite fusion." International Journal of Remote Sensing 40(5-6): 2104-2119.

Rodriguez-Alvarez, N.; Aguasca, A.; Valencia, E.; Bosch-Lluis, X.; Ramos-Pérez, I.; Park, H.; Camps, A.; Vall-Llossera, M., 2011: Snow Monitoring Using GNSS-R Techniques. In Proceedings of the International Geoscience and Remote Sensing Symposium (IGARSS), Vancouver, BC, Canada, pp. 4375–4378.

Camici, S., Tarpanelli, A., Brocca, L., Melone, F., Moramarco, T., 2011: Design soil moisture estimation by comparing continuous and storm-based rainfall-runoff modelling. Water Resour. Res., 47 (5), p. W05527.

Senyurek, V.; Lei, F.; Boyd, D.; Kurum, M.; Gurbuz, A.C.; Moorhead, R. Machine Learning-Based CYGNSS Soil Moisture Estimates over ISMN sites in CONUS. *Remote Sens.* 2020, *12*, 1168.

Soisuvarn, S.; Jelenak, Z.; Said, F.; Chang, P.S.; Egido, A., 2016: The GNSS Reflectometry Response to the Ocean Surface Winds and Waves. IEEE J. STARS, 9, 4678–4699.

Srivastaa, P. K., Han, D., Ramirez, M. A., Islam. T., 2013: Appraisal of SMOS Soil Moisture at a Catchment Scale in a Temperate Maritime Climate. Hydrology 498: 292–304.

Xu, H., et al., 2018: Quality Improvement of Satellite Soil Moisture Products by Fusing with In-Situ Measurements and GNSS-R Estimates in the Western Continental U.S. Remote Sensing 10(9).

Vapnik, V., Cherkassky, V., 1971: On the uniform convergence of relative frequencies of events to their probabilities. Theory Probab Appl 16: 264–80.

Vapnik, V., 2000: The Nature of Statistical Learning Theory. Second Edition, Springer, New York, pp. 138-143.

Yin, C., et al., 2019: Intercomparison of Soil Moisture Retrieved from GNSS-R and from Passive L-Band Radiometry at the Valencia Anchor Station. MDPI Sensors 19(8).

Zavorotny, V. U., et al., 2014: Tutorial on Remote Sensing Using GNSS Bistatic Radar of Opportunity. IEEE Geoscience and Remote Sensing Magazine 2(4): 8-45.

Zhu, Y.; Yu, K.; Zou, J.; Wickert, J., 2017: Sea Ice Detection Based on Differential Delay-Doppler Maps from UK TechDemoSat-1. MDPI Sensors 17, 1614.