A COMPARISON OF TREE-BASED REGRESSION MODELS FOR SOIL MOISTURE ESTIMATION USING SAR DATA

Z. Akhavan [1] [*], M. Hasanlou [1], M. Hosseini[2]

[1] School of Surveying and Geospatial Engineering, University of Tehran, Tehran 1417414418, Iran – (zeinab.akhava.h, hasanlou)@ut.ac.ir

[2] Dept. of Geographical Sciences, University of Maryland, College Park, USA – (mhoseini)@umd.edu

Commission IV, WG IV/3

KEY WORDS: Soil Moisture, Dual Polarimetric Decomposition, Gradient Boosted Regression Tree, eXtreme Gradient Boosted, Random Forest, SAR, Remote Sensing

ABSTRACT:

Soil moisture content plays a pivotal role in biomass development of vegetation coverage at various growth stages. Moisture content of the soil is considered as a crucial parameter for agricultural studies which directly leads to higher fertility rate. Remote sensing techniques, specifically Synthetic Aperture Radar (SAR) sensors, provides suitable opportunity for continuous soil moisture monitoring at various spatial and temporal resolutions. In this study, field campaigns conducted to measure soil surface parameters, including soil moisture and roughness, synchronized with Sentinel-1 pass over an agricultural region near Mohammadshahr, Iran. Fieldwork for soil moisture sampling have done during plants' (canola and winter wheat) growth stages. The Gradient Boosted Regression Tree (GBRT), eXtreme Gradient Boosted (XGB), and Random Forest (RF) machine learning algorithms were employed to model the relationship between the ground measured soil moisture and polarimetric SAR derived features from Sentinel-1 imageries. The results showed promising results obtained for soil moisture estimation using the dual-polarized SAR dataset over crop-covered agricultural fields with $R^2 = 0.95$ and RMSE $= 0.023$ m^3 m^{-3} using the GBRT regression model.

1. INTRODUCTION

The moisture content of the soil plays a crucial role in a broad contexts, such as irrigation management, crop growth study, and climate change analysis (Thi et al. 2019; Ranjbar, Zarei, et al. 2021) . In the last few decades, the agricultural industry has widely used new technologies, including remote sensing tools and techniques, to boost agricultural productivity because of today's growing demand (Akhavan, Hasanlou, Hosseini, and Becker-Reshef 2021; Reisi-Gahrouei et al. 2019). Among remote sensing sensors, active microwave sensors, specifically Synthetic Aperture Radar (SAR) are greatly affected by the characteristics of the target, including the surface roughness and vegetation coverage, making the SAR-based soil moisture retrieval a challenging process (Hajj et al. 2017; Paloscia et al. 2013; Ranjbar, Akhoondzadeh, et al. 2021). Soil moisture estimation over vegetated agricultural regions performed in X-, C-, and L-bands using airborne and spaceborne fully polarimetric and dual-polarized remotely sensed datasets, including the Airborne Synthetic Aperture Radar (AIRSAR), Uninhabited Aerial Vehicle Synthetic Aperture Radar (UAVSAR), Advanced land observing satellite/ phased array L-band SAR (ALOS/PALSAR), RADARSAT, and Sentinel-1. Recently, Copernicus Sentinel-1 dataset has sparked interest in using free C-band SAR datasets for soil moisture studies with an acceptable spatial and temporal resolution.

Several backscattering models, including physical, empirical, and semi-empirical, have developed and widely employed to reduce the associated uncertainties by taking some factors into consideration, and also studying their effects on backscattered

signals (Baghdadi et al. 2016; Ma, Li, and McCabe 2020; Oh 2004; Dubois and Engman 1995). To overcome limitations, some researchers developed theoretical, empirical, and semi-empirical models, including Integral Equation Model (IEM), Oh, and Dubois models for soil moisture retrieval (Dubois and Engman 1995; Ezzahar et al. 2020; Fung, Li, and Chen 1992; Oh 2004; Sekertekin, Marangoz, and Abdikan 2020). Despite these models' good performances over bare and sparsely vegetated regions, they demonstrated limitations in well-developed vegetated areas; hence, for soil moisture retrieval over vegetated areas, Attema and Ulaby proposed a model, known as Water Cloud Model (WCM), that is capable of modelling backscattered signals from vegetation's canopy coverage (Attema and Ulaby 2016). Models as mentioned earlier have selected and tested in several studies based on their research objectives (Zarei, Hasanlou, and Mahdianpari 2021). Furthermore, over the last few years, several regression-based models, especially machine learning-based algorithms, have been widely used for remote sensing-based environmental studies (M. Ansari and Akhoondzadeh 2019; Mohsen Ansari and Akhoondzadeh 2020; Dev et al. 2016).

Field measurements for gathering soil samples are not feasible over large scales due to some factors, like the quick variability of moisture content over time and space, and the time-consuming process involved. Also, soil moisture retrieval using remote sensing technologies is known as a demanding task in the absence of the adequate number of soil moisture samples; Accordingly, the use of tools and skills for soil parameter (moisture and roughness) sampling provides a suitable context for more accurate surface soil moisture monitoring using remote

---

[*] Corresponding author

sensing science (El Hajj, Baghdadi, and Zribi 2019). Numerous field campaigns conducted to assess the potential of remote sensing technologies for continuous soil water content monitoring (Jackson et al. 2005; Mcnairn et al. 2017; McNairn et al. 2015). Among soil parameters, moisture content and roughness parameters sampling are weighted among the substantial variables for agricultural-related studies (Akhavan, Hasanlou, Hosseini, and McNairn 2021).

In-situ data collection procedure, applied regression methods, and type of the study region all together affect the final result. In this study, ground soil parameter measurements conducted over a permanent agricultural field during plants' growth stages, from September 2020 to February 2021 (see Figure 1). The main purpose of this study is to evaluate the potential of boosting and bagging tree-based GBRT, XGB, and RF machine learning algorithms for soil moisture monitoring over the vegetated region. For this end, SLC and GRD formats of the Sentinel-1 dataset used to extract dual-polarized SAR features (Table 1).

GRD feature	SLC feature
GLCM Contrast	Alpha
GLCM Correlation	Entropy
GLCM Dissimilarity	VH
GLCM Entropy	VV
GLCM Homogeneity	
GLCM Mean	
GLCM Variance	

Table 1. Sentinel-1 GRD- and SLC-derived features

2. STUDY AREA AND DATASET

2.1 Study Area

The study region located in the Mohammadshahr (central district of Karaj country), Alborz province, Iran (Figure 1). The agricultural research farm of the University of Tehran, with an area of approximately 206 km^2 and an altitude of 1160 meters above sea level, is located centrally in latitude = 35°48′25″ and longitude= 50°57′11″.
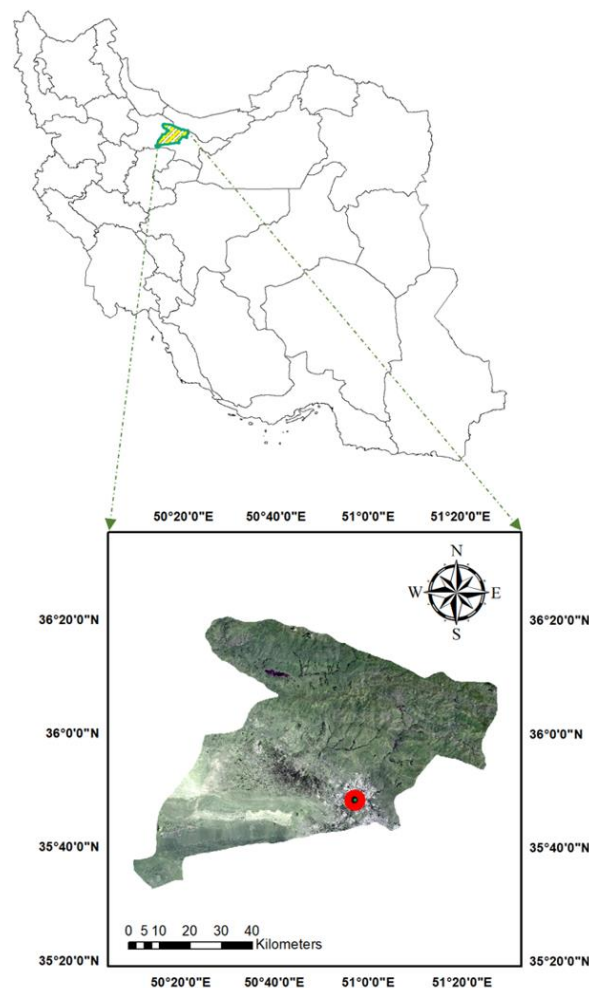

Figure 1. Location of sample points and the study region over canola and winter wheat fields

2.2 Data Collection

2.2.1 In-situ data: Fieldworks were conducted over canola and winter wheat fields during plants' growth stages using a TDR-350 probe and a man-made surface roughness profilometer to gather useful soil information. Figure 2 shows the landscape of the study area and the TDR-350 probe. During field data collection, soil moisture samples were collected three times in each field during crop's growth stages over pre-designed control points that were designed in Google Earth Pro. For accurate soil moisture retrieval, sampling points located within 30 meters distance. Each sample location was replicated four times, and the average of those values were considered as the final values. Based on the sampling strategy, the total number of 305 and 195 samples measured over winter wheat and canola fields, respectively. The collected dataset was divided into 75% training and 25% test categories for the training and evaluation phases. Figure 3 demonstrates moisture content ranges of the soil during ground measurements. To ensure the accuracy of TDR-350, gravimetric soil sampling was also conducted (Figure 4). In this procedure, the specific amount of the moist samples was weighted separately. The samples were then dried in an oven at 105°C for 24 hours. In the next step, dried samples were weighed, and moisture percent was calculated using (1). The results of gravimetric-measured soil moisture values are demonstrated in Table 2.

$$\theta_g = \frac{m_{wet} - m_{dry}}{m_{dry}} \qquad (1)$$

where
θ_g = weight of moisture content
m_{wet} = weight of moist soil
m_{dry} = weight of the dried soil

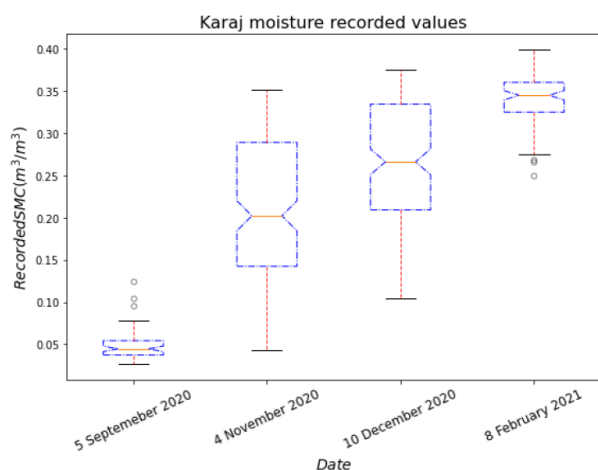**Figure 2.** TDR-350 soil moisture probe and agricultural (canola (left) and winter wheat (right)) fields

Figure 3. Ranges of the soil moisture content collected during In-situ measurement

Figure 4. Gravimetric soil water content sampling

Station No.	Gravimetric	In-situ
409	10.3	10.43
501	1.6	1.5
553	1.9	2.2

Table 2. Gravimetric vs. TDR-measured soil water content

A portable profilometer with a length of 150 centimeters used to measure surface roughness parameters. The profilometer was aligned in two directions, and photos were taken with a digital camera to capture the roughness status. Then, two measurements were taken at two different locations in each field at the beginning and end of the season. Finally, manual alignment was used to ensure the vertical status of the roughness profilometer, and also photos were captured and post-processed using the Webplotdigitizer software and Python Programming Language (Figure 5).

Figure 5. Surface roughness profilometer

2.2.2 Sentinel-1: The European Space Agency Copernicus program provides Sentinel-1 SAR imageries with six days revisit cycle in dual- and single polarization mode (Bahrami et al. 2022). In the present study, the backscatter coefficients, the Entropy and Alpha Dual-Polarized information were extracted from the Side Looking Complex (SLC) format of the Sentinel-1 C-band dataset. In order to extract information, the C_2 matrix was extracted after implementing radiometric calibration and TopSAR-Deburst steps. Finally, the backscatter coefficients beside Entropy-Alpha parameters were extracted after multi-looking and implementing a polarimetric speckle filter. Besides, the Ground Range Detected (GRD) format of the Sentinel-1 dataset have been used for Gray Level Co-occurrence Matrix (GLCM) feature extraction.

3. METHODS

In the present investigation, various tree-based models as the ensemble learning machine learning algorithms, including GBRT, XGB, and RF, were deployed. To evaluate the performance of these algorithms, models were trained and tested using a split of 75/25 ratio. The tree-based algorithms, used in this study, are described briefly in the following subsections.

3.1 Gradient Boosted Regression Trees

Gradient Boosted Regression Trees (GBRT) takes advantage of boosting as a statistical approach to aggregate weak learners and convert them to a single strong learner model. This model

minimizes the residuals, diminishes the loss function, and optimizes the prediction by generating new decision trees and adding them in sequential steps (Bahrami et al. 2021). Furthermore, the GBRT algorithm is sensitive to its parameters; hence, appropriate parameter selection, including the number of estimators, maximum depth, learning rate, loss function, and so forth assists in reaching the best algorithm implementation and final result.

3.2 eXtreme Gradient Boosted

The eXtreme Gradient Boosted algorithm uses an additive training process to develop strong learners using the additive learning procedure. The two-phases additive training process compensates for the drawbacks of other methods by fitting the learning phase to all input data, followed by an adjustment phase for residuals until reaching the stopping criterion (Pham et al. 2020). This algorithm tries to control the fitting problem by avoiding biases, under- and over-fitting problems by considering several hyperparameters identical to GBRT's hyperparameters, as well as gamma parameter.

3.3 Random Forest

Random Forest (RF), as a robust ensemble learning model, combines several decision trees in a parallel structure in order to determine the optimum nonlinear relationship between target and input features. This model employs the bagging procedure and uses a bootstrap sampling strategy for random subset input variable selection of each decision tree (Akhavan, Hasanlou, Hosseini, and McNairn 2021; Mao et al. 2019). This approach assists in training several models independently, which reduces the variance and prevents the overfitting phenomenon. The final outcome shows the average performance of all decision trees that created the structure of this ensemble learning algorithm.

4. RESULTS AND DISCUSSION

This section demonstrates the results of GBRT, XGB, and RF algorithm implementation using SAR-derived features extracted from SLC and GRD formats of the Sentinel-1 dataset. Corresponding to each In-situ sampling point, the features listed in Table 2 extracted to train and evaluate the performance of GBRT, XGB, and RF regression models over test dataset. Table 3, and Figure 6 summarizes and demonstrates the performance of these algorithms using the RMSE and R-squared parameters. Additionally, soil moisture maps produced by these regression models are demonstrated in Figure 7. These maps produced based on the combination of soil moisture datasets over canola and winter wheat fields

Algorithm	Canola	Winter Wheat
GBRT	$R^2 = 0.95$ RMSE = 0.023	$R^2 = 0.64$ RMSE = 0.042
XGB	$R^2 = 0.93$ RMSE = 0.028	$R^2 = 0.60$ RMSE = 0.044
RF	$R^2 = 0.95$ RMSE = 0.025	$R^2 = 0.59$ RMSE = 0.047

Table 3. Results of soil moisture retrieval over canola and winter wheat agricultural region using GBRT, XGB, and RF algorithms
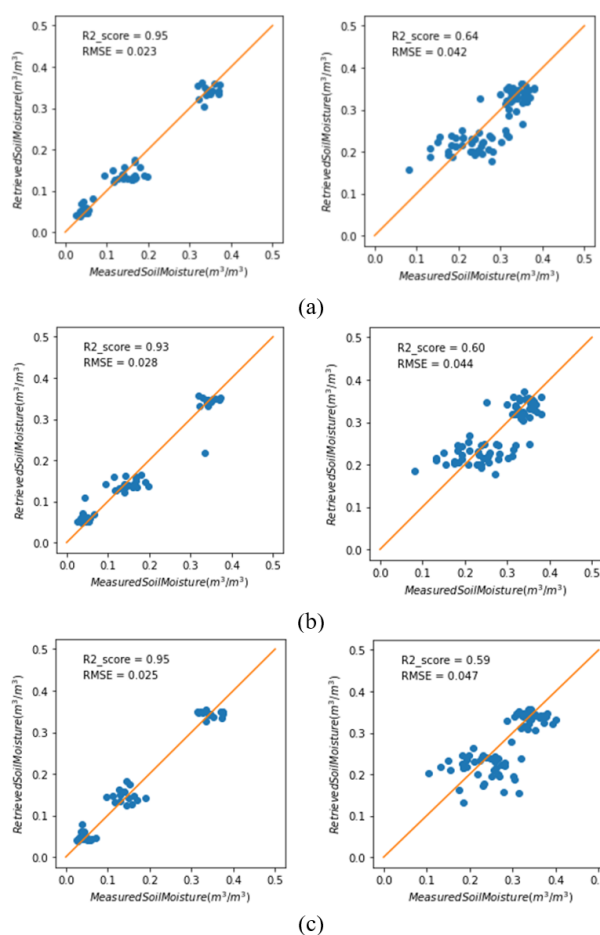

(a)

(b)

(c)

Figure 6. Results obtained using (a) GBRT, (b) XGB, and (c) RF regression models over canola (left column) and winter wheat (right column) fields

Key Words As illustrated in Table 3, the results confirm the relatively close performance of GBRT, XGB, and RF for soil moisture retrieval over canola and winter wheat fields. All algorithms performed better for soil moisture retrieval over canola fields. The best result was obtained using the GBRT model with $R^2 = 0.95$ and the least error content of RMSE = 0.023 m^3 m^{-3}. The reason behind the better performance of these models over canola field and also the relatively poor performance of algorithms over winter wheat fields compared with canola field would probably be due to winter wheat's biomass development, and also plant's structure that affect the backscattered SAR signal.
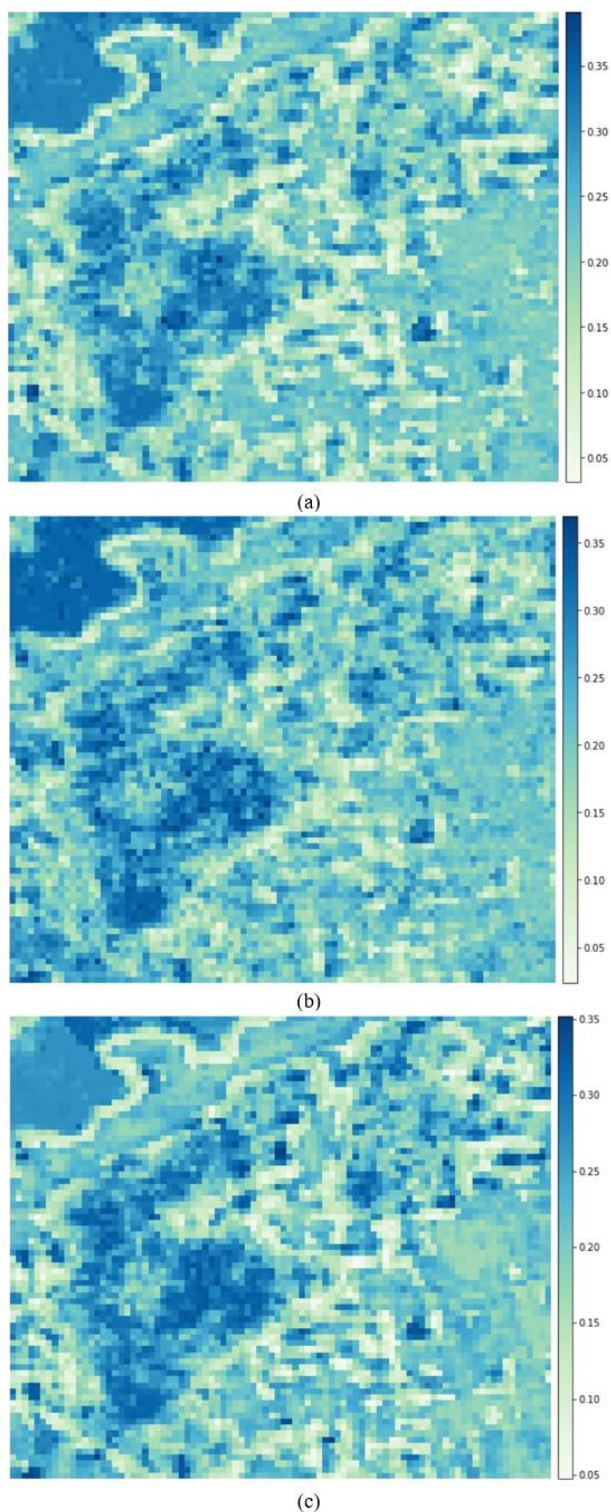
Figure 7. Predicted soil moisture maps using tree-based (a) GBRT, (b) XGB, and (c) RF algorithms.

5. CONCLUSION

The principal purpose of the present study was authentic soil moisture investigation over an agricultural region by conducting fieldwork (soil parameter (moisture and roughness) sampling) coincident with ESA Sentinel-1 satellite pass. For this purpose, several SAR features were extracted from the SLC and GRD formats of Sentinel-1 imageries besides field collected samples. Then, tree-based machine learning algorithms employed for accurate soil moisture retrieval. In this study, several features were considered for soil moisture monitoring. Moreover, machine learning algorithms obtained relatively identical accuracies and the most accurate result with the lowest RMSE value was obtained by the GBRT model over the canola field ($R^2 = 0.95$ and RMSE = 0.023 m^3 m^{-3}). This study showed that dual-polarized SAR, as well as machine learning approaches, were effective for soil moisture estimation over winter wheat and canola fields. However, to have a solid and comprehensive conclusion, more tests and experiments over other areas are required. It is recommended to use SAR-derived vegetation features like Radar Vegetation Index (RVI) besides features extracted from optical sensors like Sentinel-2 to evaluate multispectral-derived indexes for soil moisture retrieval.

REFERENCES

Akhavan, Zeinab, Mahdi Hasanlou, Mehdi Hosseini, and Inbal Becker-Reshef. 2021. "Soil Moisture Retrieval Improvement over Agricultural Fields by Adding Entropy–Alpha Dual-Polarimetric Decomposition Features." *Journal of Applied Remote Sensing* 15 (03): 1–16. https://doi.org/10.1117/1.jrs.15.034516.

Akhavan, Zeinab, Mahdi Hasanlou, Mehdi Hosseini, and Heather McNairn. 2021. "Decomposition-Based Soil Moisture Estimation Using Uavsar Fully Polarimetric Images." *Agronomy* 11 (1): 1–25. https://doi.org/10.3390/agronomy11010145.

Ansari, M., and M. Akhoondzadeh. 2019. "Water Salinity Mapping of Karun Basin Located in Iran Using the SVR Method." *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives* 42 (4/W18): 97–101. https://doi.org/10.5194/isprs-archives-XLII-4-W18-97-2019.

Ansari, Mohsen, and Mehdi Akhoondzadeh. 2020. "Mapping Water Salinity Using Landsat-8 OLI Satellite Images (Case Study: Karun Basin Located in Iran)." *Advances in Space Research* 65 (5): 1490–1502. https://doi.org/10.1016/j.asr.2019.12.007.

Attema, E P W, and Fawwaz T Ulaby. 2016. "Prevalence of Anemia in Patients with Diabetic Kidney Disease." *Abstract Book – 29th European Diabetic Nephropathy Study Group Meeting* 13 (2): 357–64.

Baghdadi, Nicolas, Mohammad Choker, Mehrez Zribi, Mohammad El Hajj, Simonetta Paloscia, Niko E.C. Verhoest, Hans Lievens, Frederic Baup, and Francesco Mattia. 2016. "A New Empirical Model for Radar Scattering from Bare Soil Surfaces." *Remote Sensing* 8 (11): 1–14. https://doi.org/10.3390/rs8110920.

Bahrami, Hazhir, Saeid Homayouni, Heather McNairn, Mehdi Hosseini, and Masoud Mahdianpari. 2022. "Regional Crop Characterization Using Multi-Temporal Optical and Synthetic Aperture Radar Earth Observations Data." *Canadian Journal of Remote Sensing* 48 (2): 258–77. https://doi.org/10.1080/07038992.2021.2011180.

Bahrami, Hazhir, Saeid Homayouni, Abdolreza Safari, Sayeh Mirzaei, Masoud Mahdianpari, and Omid Reisi-Gahrouei. 2021. "Deep Learning-Based Estimation of Crop Biophysical Parameters Using Multi-Source and Multi-Temporal Remote Sensing Observations." *Agronomy* 11 (7). https://doi.org/10.3390/agronomy11071363.

Dev, Soumyabrata, Bihan Wen, Yee Hui Lee, and Stefan Winkler. 2016. "Ground-Based Image Analysis: A Tutorial on Machine-Learning Techniques and Applications." *IEEE Geoscience and Remote Sensing Magazine* 4 (2): 79–93. https://doi.org/10.1109/MGRS.2015.2510448.

Dubois, Pascale C., and Ted Engman. 1995. "Measuring Soil Moisture with Imaging Radars." *IEEE Transactions on Geoscience and Remote Sensing* 33 (4): 915–26. https://doi.org/10.1109/36.406677.

Ezzahar, Jamal, Nadia Ouaadi, Mehrez Zribi, Jamal Elfarkh, Ghizlane Aouade, Said Khabba, Salah Er-Raki, Abdelghani Chehbouni, and Lionel Jarlan. 2020. "Evaluation of Backscattering Models and Support Vector Machine for the Retrieval of Bare Soil Moisture from Sentinel-1 Data." *Remote Sensing* 12 (1): 1–20. https://doi.org/10.3390/RS12010072.

Fung, Adrian K., Zongqian Li, and K. S. Chen. 1992. "Backscattering from a Randomly Rough Dielectric Surface." *IEEE Transactions on Geoscience and Remote Sensing* 30 (2): 356–69. https://doi.org/10.1109/36.134085.

Hajj, Mohammad El, Nicolas Baghdadi, and Mehrez Zribi. 2019. "Comparative Analysis of the Accuracy of Surface Soil Moisture Estimation from the C- and L-Bands." *International Journal of Applied Earth Observation and Geoinformation* 82 (March): 101888. https://doi.org/10.1016/j.jag.2019.05.021.

Hajj, Mohammad El, Nicolas Baghdadi, Mehrez Zribi, and Hassan Bazzi. 2017. "Synergic Use of Sentinel-1 and Sentinel-2 Images for Operational Soil Moisture Mapping at High Spatial Resolution over Agricultural Areas." *Remote Sensing* 9 (12): 1–28. https://doi.org/10.3390/rs9121292.

Jackson, Thomas J., Rajat Bindlish, Albin J. Gasiewski, Boba Stankov, Marian Klein, Eni G. Njoku, David Bosch, Tommy L. Coleman, Charles A. Laymon, and Patrick Starks. 2005. "Polarimetric Scanning Radiometer C- and X-Band Microwave Observations during SMEX03." *IEEE Transactions on Geoscience and Remote Sensing* 43 (11): 2418–29. https://doi.org/10.1109/TGRS.2005.857625.

Ma, Chunfeng, Xin Li, and Matthew F. McCabe. 2020. "Retrieval of High-Resolution Soil Moisture through Combination of Sentinel-1 and Sentinel-2 Data." *Remote Sensing* 12 (14): 1–28. https://doi.org/10.3390/rs12142303.

Mao, Huihui, Jihua Meng, Fujiang Ji, Qiankun Zhang, and Huiting Fang. 2019. "Comparison of Machine Learning Regression Algorithms for Cotton Leaf Area Index Retrieval Using Sentinel-2 Spectral Bands." *Applied Sciences (Switzerland)* 9 (7). https://doi.org/10.3390/app9071459.

McNairn, Heather, Thomas J. Jackson, Grant Wiseman, Stéphane Bélair, Aaron Berg, Paul Bullock, Andreas Colliander, et al. 2015. "The Soil Moisture Active Passive Validation Experiment 2012 (SMAPVEX12): Prelaunch Calibration and Validation of the SMAP Soil Moisture Algorithms." *IEEE Transactions on Geoscience and Remote Sensing* 53 (5): 2784–2801. https://doi.org/10.1109/TGRS.2014.2364913.

Mcnairn, Heather, Tom J Jackson, Jarrett Powers, Stephane Bélair, Aaron Berg, Andreas Colliander, Michael H Cosh, et al. 2017. "SMAPVEX16 Database Report," 1–56.

Oh, Yisok. 2004. "Quantitative Retrieval of Soil Moisture Content and Surface Roughness from Multipolarized Radar Observations of Bare Soil Surfaces." *IEEE Transactions on Geoscience and Remote Sensing* 42 (3): 596–601. https://doi.org/10.1109/TGRS.2003.821065.

Paloscia, S., S. Pettinato, E. Santi, C. Notarnicola, L. Pasolli, and A. Reppucci. 2013. "Soil Moisture Mapping Using Sentinel-1 Images: Algorithm and Preliminary Validation." *Remote Sensing of Environment* 134: 234–48. https://doi.org/10.1016/j.rse.2013.02.027.

Pham, Tien Dat, Nga Nhu Le, Nam Thang Ha, Luong Viet Nguyen, Junshi Xia, Naoto Yokoya, Tu Trong To, Hong Xuan Trinh, Lap Quoc Kieu, and Wataru Takeuchi. 2020. "Estimating Mangrove Above-Ground Biomass Using Extreme Gradient Boosting Decision Trees Algorithm with Fused Sentinel-2 and ALOS-2 PALSAR-2 Data in Can Gio Biosphere Reserve, Vietnam." *Remote Sensing* 12 (5). https://doi.org/10.3390/rs12050777.

Ranjbar, Sadegh, Mehdi Akhoondzadeh, Brian Brisco, Meisam Amani, and Mehdi Hosseini. 2021. "Soil Moisture Change Monitoring from C and L-Band SAR Interferometric Phase Observations." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 14: 7179–97. https://doi.org/10.1109/JSTARS.2021.3096063.

Ranjbar, Sadegh, Arastou Zarei, Mahdi Hasanlou, Mehdi Akhoondzadeh, Jalal Amini, and Meisam Amani. 2021. "Machine Learning Inversion Approach for Soil Parameters Estimation over Vegetated Agricultural Areas Using a Combination of Water Cloud Model and Calibrated Integral Equation Model." *Journal of Applied Remote Sensing* 15 (01): 1–17. https://doi.org/10.1117/1.jrs.15.018503.

Reisi-Gahrouei, Omid, Saeid Homayouni, Heather McNairn, Mehdi Hosseini, and Abdolreza Safari. 2019. "Crop Biomass Estimation Using Multi Regression Analysis and Neural Networks from Multitemporal L-Band Polarimetric Synthetic Aperture Radar Data." *International Journal of Remote Sensing* 40 (17): 6822–40. https://doi.org/10.1080/01431161.2019.1594436.

Sekertekin, Aliihsan, Aycan Murat Marangoz, and Saygin Abdikan. 2020. "ALOS-2 and Sentinel-1 SAR Data Sensitivity Analysis to Surface Soil Moisture over Bare and Vegetated Agricultural Fields." *Computers and Electronics in Agriculture* 171 (October 2019): 105303. https://doi.org/10.1016/j.compag.2020.105303.

Thi, Dinh Ngo, Nguyen Thi Thu Ha, Quy Tran Dang, Katsuaki Koike, and Nhuan Mai Trong. 2019. "Effective Band Ratio of Landsat 8 Images Based on VNIR-SWIR Reflectance Spectra of Topsoils for Soil Moisture Mapping in a Tropical Region." *Remote Sensing* 11 (6). https://doi.org/10.3390/RS11060716.

Zarei, Arastou, Mahdi Hasanlou, and Masoud Mahdianpari. 2021. "A Comparison of Machine Learning Models for Soil Salinity Estimation Using Multi-Spectral Earth Observation Data." *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 5 (3): 257–63. https://doi.org/10.5194/isprs-annals-V-3-2021-257-2021.