Machine Learning-Based Soil Moisture Retrieval Using CYGNSS and Interpolated SMAP Data

Kun Zhang [1], Kegen Yu [1], Nianfu Xu [1], Changyang Wang [1], Yiruo Lin [1]

[1] School of Environment Science and Spatial Informatics, China University of Mining and Technology, Xuzhou, China - ts22160030a31ld@cumt.edu.cn, kegen.yu@cumt.edu.cn, nianfuxu@cumt.edu.cn, tb20160008b1@cumt.edu.cn, yiruo.lin@foxmail.com

Keywords: GNSS-R, CYGNSS, Soil Moisture, SMAP, Machine Learning, Interpolation.

Abstract

Soil moisture is a crucial component of the global terrestrial ecosystem water vapor cycle, and higher spatial-temporal soil moisture product has significant impacts on research in agriculture, hydrology, and ecology. Spaceborne Global Navigation Satellite System Reflectometry (GNSS-R) is a new remote sensing technology which can be used to retrieve soil moisture. Cyclone Global Navigation Satellite System (CYGNSS), launched by NASA in 2016 to observe ocean surface hurricanes, has also collected a substantial amount of data of GNSS signals reflected over the land. This paper focuses on soil moisture retrieval using CYGNSS and SMAP dataset. In this study, three machine learning methods (GSVM, BP and RF) are used for spatial interpolation of SMAP soil moisture data. Result shows RF interpolation is the best with the RMSE of 0.044 cm^3/cm^3 and the CC of 0.959 compared with SMAP truly SM. Then soil moisture data after RF interpolation and conventional linear interpolation are respectively used as response variables to train the retrieval model by ELM, BP and RF machine learning methods respectively. Results show that, among the six regression models, the RF-RF regression model performs the best with the RMSE of 0.012 cm^3/cm^3 and the CC of 0.986. Using this model, we get a global soil moisture product and have higher temporal and spatial resolution.

1. Introduction

Soil moisture (SM) is a physical quantity used to express the degree of soil dryness and moisture. It is an important component of the water vapor cycle in global terrestrial ecosystems (Vereecken et al., 2008). SM plays an indispensable role in research on climate change, global ecology, surface hydrology models, agricultural drought, and crop yield estimation. Previous measurement methods such as resistance method and time domain reflectometry have a small measurement range and can only reflect the local conditions of the measured area (Persson and Haridy, 2003). At the same time, they require a lot of manpower and material resources, have a long experimental cycle, and are greatly restricted by external environmental conditions. Currently, using microwave remote sensing satellites to obtain ground SM data has become a common method, but there are also some shortcomings. The lenses and sensors of optical remote sensing satellites are susceptible to the influence of clouds, fog and other weather conditions, and are expensive; the spatial and temporal resolution of infrared band remote sensing is small; microwave remote sensing has a relatively long wavelength, strong penetrating power, and can achieve all-weather observation, but the cost is relatively high (Price, 2017; Paloscia et al., 2001).

GNSS reflectometry (GNSS-R) is an emerging remote sensing technology, which is a derivative technology from GNSS. GNSS-R makes use of the GNSS signals which are always available on the Earth’s surface. Due to the fact that GNSS-R does not require any dedicated signal transmitter, it has the advantage of low cost. This technology also has the advantage of large coverage or higher spatial resolution. GNSS-R has achieved great advances over the past few decades, and it is still making fast progress. GNSS-R has been exploited to retrieve a series of physical and environmental parameters such as ocean surface height, sea surface wind speed, significant wave height, snow depth, soil moisture, above-ground biomass, and the sea ice detection, ship detection and flood detection (Yu, 2021). The SM experiment 2002 (SMEX02) is the first experiment in which both GPS reflected signals from land and in situ data were collected throughout the state of Iowa, USA in June-July 2002. In this series of experiments, it was found that the change of SM has a great correlation with precipitation events (Masters et al., 2003; Jackson et al., 2003).

CYGNSS (Cyclone Global Navigation Satellite System) mission successfully launched 8 microsatellites to establish a constellation for observing tropical cyclones with high spatial and temporal resolution, and successfully implemented the use of GNSS-R to monitor ocean hurricane intensity (Ruf et al., 2013a). Its coverage range is the mid- to low-latitude areas on both sides of the equator (about 40°S ~ 40°N), to cover the ocean areas of interest. The coverage area of the CYGNSS also includes a lot land area, so the recorded data can also be used to retrieve parameters. In 2018, Chew et al. found a strong positive correlation between CYGNSS surface reflectance and SMAP SM, proving that CYGNSS can be used to develop global SM products with high temporal resolution (perhaps every 6 hours) (Chew and Small, 2018). Many studies have also confirmed the correlation between CYGNSS data and SM (Carreno-Luengo et al, 2018; Al-Khaldi et al., 2019; Al-Khaldi et al., 2021). Volkan Senyurek evaluated the error of different spatial interpolation methods (Senyurek et al., 2021). Volkan Senyurek used three machine learning methods (ANN, RF, SVM) to retrieve SM (Senyurek et al., 2020). Orhan Eroglu used a fully connected artificial neural network to retrieve SM by learning the nonlinear relationship between SM and other terrestrial geophysical parameters and the observable data of CYGNSS (Eroglu et al., 2019).

In this study, we first propose to interpolate the low-resolution SM data using three machine learning algorithms, namely Gaussian Support Vector Machine (GSVM), Back-Propagation (BP) Neural Network, and Random Forest (RF). Traditional linear interpolation is also studied for performance comparison. The SM data generated by a machine learning-based interpolation and the traditional linear interpolation method are respectively employed are used as reference data for constructing machine learning-based SM retrieval models.

2. Dataset and ML Algorithm

2.1 Cyclone Global Navigation Satellite System

In this study, the Level 1 version 3.2 data of the cgy01 satellite in 2023 was used to retrieve SM. Daily observations for each satellite were stored in an NC file. DDM (Delay-Doppler Maps) is one of the key variables of analogue scattered power for each specular point. It is a two-dimensional map composed of 11 doppler frequency shift units and 17 time delay units. The reflection area of DDM associated with land surface specular reflection point is mostly coherent, and the reflected signal power is concentrated in a small and smooth area around the specular emission point (Ruf et al, 2013b). This part of the energy comes from the coherent scattering signal. In addition to DDMs, the NC files also provide some useful metadata of geometric and instrumental parameters. Those commonly used parameters include quality flags, error estimates, and bias estimates as well as a variety of orbital, spacecraft/sensor health, timekeeping, and geolocation parameters.

Compared with v3.1 L1 data, the correction for coarse quantization effects that was implemented in v3.1 for the signal portion of the DDM has been updated to include a correction to the noise floor portion of the DDM. This update is found to improve the sensitivity to SM over land and to have a minimal effect on the sensitivity to wind speed over ocean. L1 variables over land and ocean are now combined in common NC data files, with additional details added regarding the specular point calculation over land. Nadir antenna pattern and NBRCS rescaling have been updated to improve the inter-satellite consistency of the L1 calibration (CYGNSS., 2024). This is undoubtedly a huge step forward for SM retrieval.

2.2 SMAP

The SM products from the L3 level of the Soil Moisture Active Passive (SMAP) mission are used in this study as the "truth value" in the retrieval process. The soil moisture product obtained by this mission is also considered to be the most accurate product among the existing SM products. SMAP L3 level global SM data is divided into active and passive products. The SMAP L3 product currently used is 0-5 cm surface SM on a global scale. The SMAP data used here are L3 v.5 radiometer global daily Equal Area Expandable Earth Grid (EASE-Grid) data with a spatial resolution of 36 km × 36 km. They contain SM, quality flags and other ancillary information gridded on EASE-Grid v2.0, with ascending and descending channels being averaged together to form a single daily channel (Entekhabi et al., 2010). Besides the SM, additional variables extracted from the SMAP files and used in the algorithm such as quality flags, land roughness coefficient (LRC), vegetation water content (VWC), vegetation optical depth (VOD) and soil surface temperature (SST).

2.3 Selected Machine Learning Algorithms

The selection of machine learning algorithm and its hyperparameter has a significant impact on the performance of SM prediction. In this work, we use BP neural network, RF, Extreme Learning Machine (ELM) and GSVM methods, which are widely used for supervised regression problems. BP, RF and GSVM are used to build SM interpolation models. SM; ELM, BP and RF are used to build SM retrieval regression models.

2.3.1 BP Neural Network: BP neural network is a multi-layer feedforward neural network (Rumelhart et al., 1986). The model is so called because the error is subject to backpropagation while the information is propagated forward. In the propagation process, the model uses gradient descent method to constantly update the weight and bias of a single sample. BP neural network can be divided into three layers: input layer, hidden layer and output layer.

2.3.2 Random Forest: Based on Bootstrap self-help method, RF randomly selects dependent variables from the original sample as the training set, and randomly selects a certain amount of influence factors from the dependent variables in the original sample as the splitting point of tree nodes, so that each tree in the forest has its own different training data. Accordingly, random forest can randomly generate many different classification trees, from which the number with high repetition rate is selected or the average value is used as the final result of the model (Breiman, 2001).

2.3.3 Gaussian Support Vector Machine: GSVM is a classification algorithm, and its basic idea is to map the data into a high-dimensional space, and then find the best hyperplane in the high-dimensional space to classify. This mapping is usually achieved by kernel functions, of which Gaussian kernel functions are a common choice (Gonzalez-Abril et al., 2014).

2.3.4 Extreme Learning Machine: ELM is a machine learning algorithm used to solve supervised learning problems, especially regression and classification problems. The main idea of ELM is that in the neural network, the weight of the input layer and the bias of the hidden layer are randomly selected, while the weight of the output layer is calculated by the analytical method, which can greatly speed up the training speed (Huang et al., 2006).

3. Method

3.1 Data Quality Control

First, we need to extract DDM data and relevant parameters from the CYGNSS data by excluding those related to the ocean specular points (SP). After the land and ocean data are separated, in order to ensure the data quality, it is necessary to set up a number of rules to exclude the poor quality data such as antenna gain, signal-to-noise ratio, and altitude angle, thereby improving the accuracy of the model. In this work, if the data of SP reach any of the following receiver and experimental conditions, the data is eliminated:

1) the incident angle θ is higher than 65°.
2) the receiver gain Gr is less than 0 dB.
3) the signal-to-noise ratio is less than 2 dB.
4) the high elevation more than 2000m.

3.2 Data Matching

Firstly, since the CYGNSS land reflection point coordinates are inconsistent with the SMAP SM data point coordinates, space matching work must be carried out to obtain the SM and auxiliary data at the reflection point. In order to compare with the machine learning-based interpolation, we also use a commonly used linear interpolation method. In this work, SM during daily time is considered constant, so it only needs to perform day to day time matching. In other words, time matching is aligning CYGNSS and SMAP data on the same day. In space dimension, since SMAP's revisit time is 2-3 days, this means that all the land in the world cannot be covered in one day. In order to obtain reliable interpolation results, only CYGNSS specular reflection points within SMAP satellite coverage are considered to retrieve SM on the same day.

3.3 Machine Learning-based Interpolation

In this work, SMAP SM data is used to explore the interpolation accuracy of three machine learning methods. Elevation, longitude, and latitude of SMAP data were used as feature vectors of the interpolation model. SM of SMAP data was used as response variables. To obtain interpolated SM results on a global scale, three consecutive days of SMAP data were used. 70% of the data is used for the training of machine learning models and 30% was used to test the performance of the models. The best-performing machine learning method was used for spatial matching of CYGNSS and SMAP data. SMAP SM is the response variable of training data set. The best performing machine learning method is used to train and generate the model. The longitude, latitude and elevation of the CYGNSS specular reflection point are taken as the input of the model, and the output data is taken as the SM value of the CYGNSS reflection point. Other auxiliary parameters from the SMAP dataset are also processed by the same method. VWC, SST and LRC of CYGNSS reflection points will be used as input to retrieve the SM regression model (Clarizia et al., 2019).

3.4 Regression Model Construction

This work used three machine learning methods: RF, BP and ELM to build retrieval model with SM as the response variable. DDM snr is the main input vector. Furthermore, auxiliary parameters are also essential as feature vectors such as incident angle, inst gain, LRC, VWC, VOD, SST, SP longitude and latitude. All of them have great influence on SM retrieval. In order to get a model with good generalization ability, we used CYGNSS and SMAP data from January 1, 2023 to December 31, 2023 to train the machine learning regression models. To verify the generalization ability of these regression models, another test dataset was used, which was independent of the training dataset.

Input Variables	Long Name
ddm snr	DDM signal to noise ratio
incident angle	Specular point incidence angle
inst_gain	Instrument gain
longitude	Specular point longitude
latitude	Specular point latitude
LRC	Roughness coefficient
VWC	Vegetation water content
VOD	Vegetation opacity
SST	Surface temperature

Table 1. List of Input variables for the Regression Models

3.5 Accuracy Index

In this work, root-mean-square error (RMSE), mean absolute error (MAE) and correlation coefficient (CC) are used to verify the interpolation accuracy of the above interpolation method and the accuracy of the SM retrieval regression model. They are defined as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}, \quad (1)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i|, \quad (2)$$

$$CC = \frac{\sum_{i=1}^{n}(y_i-\bar{y})(\hat{y}_i-\bar{\hat{y}})}{\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2 \sum_{i=1}^{n}(\hat{y}_i-\bar{\hat{y}})^2}}, \quad (3)$$

where n = the number of data samples
y_i = the SM estimates by the regression model
$\hat{y}_i$ = the SM data of SMAP
$\bar{y}_i$ = the mean of y_i
$\bar{\hat{y}}$ = the mean of $\hat{y}_i$

4. Results

The results were analysed in two parts. Firstly, the interpolation results of different methods are compared. Secondly, the regression performance based on three machine learning models was analysed and the retrieval result of the RF model using RF interpolation was given.

4.1 Interpolation Accuracy

Figure 1 shows the test dataset of SMAP SM, treated as truth SM values, while Figure 2 shows the comparison of machine learning-based interpolation results with SMAP SM.

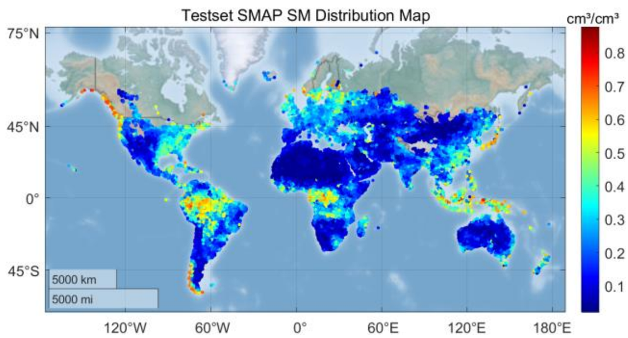


Figure 1. Test Set SMAP SM Distribution MAP.

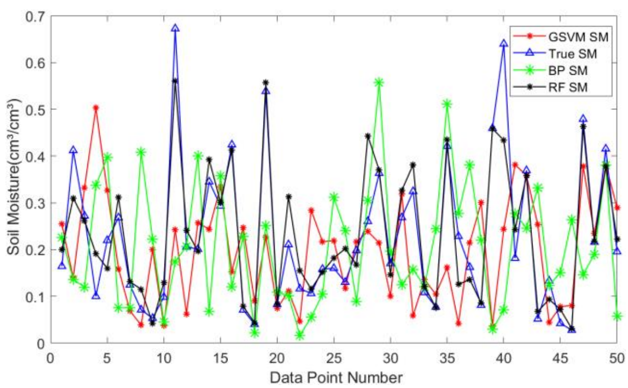


Figure 2. Soil Moisture after Different Interpolation.

Interpolation	CC	MAE (cm^3/cm^3)	RMSE (cm^3/cm^3)
GSVM	0.855	0.054	0.079
BP	0.84	0.059	0.082
RF	0.938	0.034	0.053

Table 2. Interpolation Performance of different Machine Learning Models

Figure 1 shows the distribution points of SMAP data of the test set on the global map. It can be seen that all land is covered in the range 45°S to 45°N; this also fully covers the coverage of CYGNSS. Figure 2 shows the SM interpolation errors by machine learning-based methods. Overall, the RF interpolation results are better than the BP and GSVM interpolation results. Table 2 shows the SM retrieval performance of GSVM, BP and RF model interpolation. RF model has the highest CC of 0.938, the lowest RMSE of 0.053 cm^3/cm^3 compared with SMAP truly SM. Performances of BP and GSVM model is worse than RF model. This also proves that the interpolation result using RF is more reliable so that we used daily SMAP SM for RF interpolation to obtain SM at CYGNSS specular reflection points on the same day.

4.2 Performance based on Different Machine Learning Regression Models

In this part, the SM data generated by machine learning-based interpolation are used as response variables to train the SM retrieval models based on ELM, BP and RF machine learning methods respectively. At the same time, we also use the linear SM interpolation results to train the three machine learning-based models as a contrast. According to different combinations of interpolation methods and machine learning regression methods, six models are named as Linear-ELM Model, RF-ELM Model, Linear-BP Model, RF-BP Model, Linear-RF Model, RF-RF Model for simplicity.

ELM	Performance Index		
Interpolation	CC	RMSE (cm^3/cm^3)	MAE (cm^3/cm^3)
Linear	0.814	0.076	0.058
RF	0.831	0.065	0.050

Table 3. Trainset Performance of ELM Regression Model

BP	Performance Index		
Interpolation	CC	RMSE (cm^3/cm^3)	MAE (cm^3/cm^3)
Linear	0.838	0.072	0.054
RF	0.854	0.060	0.047

Table 4. Trainset Performance of BP Regression Model

RF	Performance Index		
Interpolation	CC	RMSE (cm^3/cm^3)	MAE (cm^3/cm^3)
Linear	0.988	0.021	0.014
RF	0.99	0.014	0.009

Table 5. Trainset Performance of RF Regression Model

Table 3-5 shows the different model performances. In the same machine learning regression method, the RF interpolation model is superior to linear interpolation, with higher CC and lower RMSE. For RF-ELM regression model, the CC increases 2.1% and RMSE decreases 14.5% compared with Linear-ELM model. For RF-BP regression model, the CC increases 1.9% and RMSE decreases 16.7% compared with Linear-BP model. For RF-RF regression model, the CC increases 0.2% and RMSE decreases 35.7% compared with Linear-RF model. For three machine learning retrieval methods, RF regression method shows the best retrieval results. In general， RF-RF model has the highest CC of 0.99 and lowest RMSE of 0.009 cm^3/cm^3.

In order to verify the generalization ability of the above retrieval model, 300000 groups of data are selected as the test set. Here is the performance of the test set.

ELM	Performance Index		
Interpolation	CC	RMSE (cm^3/cm^3)	MAE (cm^3/cm^3)
Linear	0.813	0.076	0.058
RF	0.831	0.065	0.050

Table 6. Testset performance of ELM Regression Model

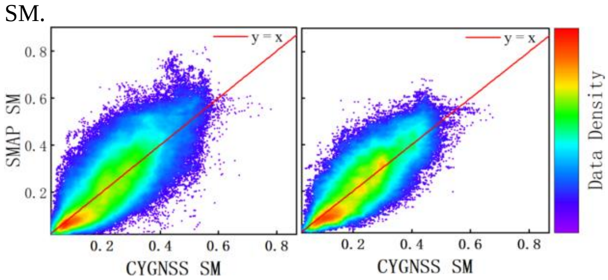
BP	Performance Index		
Interpolation	CC	RMSE (cm^3/cm^3)	MAE (cm^3/cm^3)
Linear	0.838	0.073	0.054
RF	0.854	0.060	0.047

Table 7. Testset performance of BP Regression Model

RF	Performance Index		
Interpolation	CC	RMSE (cm^3/cm^3)	MAE (cm^3/cm^3)
Linear	0.973	0.031	0.02
RF	0.986	0.019	0.012

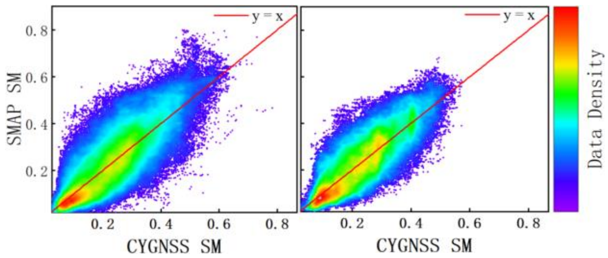
Table 8. Testset Performance of RF Regression Model

Table 6-8 shows the performance of different models using the test data. Each model shows good generalization ability. For two ELM regression models, the CC increases 2.1% and RMSE decreases 14.5%. For two BP regression models, the CC increases 2.1% and RMSE decreases 14.5%. The performance of the RF model has been slightly reduced, but it still shows the best results with higher CC and lower RMSE. Density maps are used to show the regression performance of predicted SM with true SM.



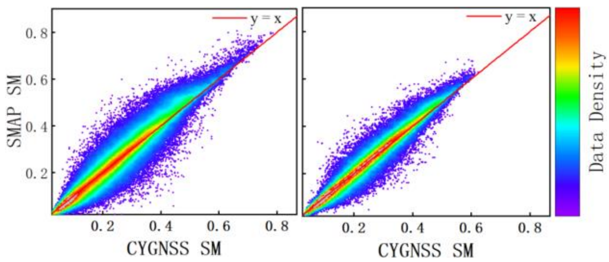
(a) Linear-ELM Model (b) RF-ELM Model

Figure 3. Density Plot of ELM Model.



(a) Linear-BP Model (b) RF-BP Model

Figure 4. Density Plot of BP Model.



(a) Linear-RF Model (b) RF-RF Model

Figure 5. Density Plot of RF Model.

Figure 3-5 shows the point density plots for the nine models. The performance of each model showed good fitting results, which basically conforms to the function relationship of y=x. The RF-RF model performs best. The Linear-ELM model has a large point distribution area and the worst performance. In summary, according to the data in Figure 2, Table 2, Table 8 and Figure 5, the RF-RF model shows the best retrieval performance. Moreover, in the same machine learning regression method, the model using the SM of RF interpolation result as the response variable performs better than the traditional linear interpolation method. According to this, we used the RF-RF model to generate a global SM product as follows:

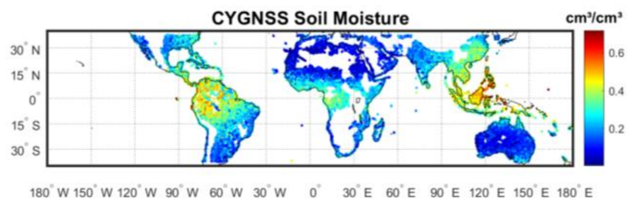


Figure 6. Global CYGNSS Soil Moisture Map.

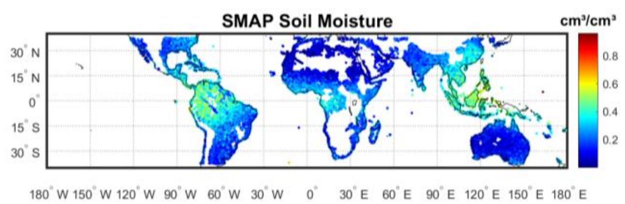


Figure 7. Global SMAP Soil Moisture Map.

Figures 6 and 7 show good consistency. However, in the area near the equator, it is obviously that the CYGNSS SM value is higher than the true SM of SMAP. However, on a global scale, SM retrieved by CYGNSS showed a good agreement with SMAP SM. Since the CYGNSS satellite revisit cycle time is one day and the space coverage is wider, the SM products generated by CYGNSS have high temporal and spatial resolution

5. Conclusion

In this study, we investigate interpolation of SM data using GSVM, BP, RF interpolation methods. The results show that the RF interpolation achieves the best results. Using SM data by traditional linear interpolation and RF-based interpolation respectively, six models for SM retrieval based on three machine learning methods were trained and used to retrieve SM. For the SM retrieval model with the same machine learning method, the model using RF interpolation result shows the best result. For different SM retrieval regression, the RF model using linear and RF interpolation methods has higher retrieval performance. Furthermore, the result shows that the RF-RF model has the best performance with the CC of 0.986 and the RMSE of 0.012 cm^3/cm^3. Finally, the RF-RF model is used to generate a global SM product. CYGNSS SM products have good consistency with SMAP SM on a global scale.

Acknowledgements

This work was supported by the National Natural Science Foundation of China under Grant 42174022 and A Project Funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions.

References

Al-Khaldi, M. M., Johnson, J. T., 2021. Soil moisture retrievals using CYGNSS data in a time-series ratio method: progress update and error analysis. *IEEE Geoscience and Remote Sensing Letters*, 19, 1-5.

Al-Khaldi, M. M., Johnson, J. T., O’Brien, A. J., Balenzano, A., & Mattia, F., 2019. Time-series retrieval of soil moisture using CYGNSS. *IEEE Transactions on Geoscience and Remote Sensing*, 57(7), 4322-4331.

Breiman, L., 2001. Random forests. *Machine learning*, 45, 5-32.

Carreno-Luengo, H., Luzi, G., Crosetto, M., 2018. Impact of the elevation angle on CYGNSS GNSS-R bistatic reflectivity as a function of effective surface roughness over land surfaces. *Remote Sensing*, 10(11), 1749.

Chew, C. C., and 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.

Clarizia, M. P., Pierdicca, N., Costantini, F., Floury, N., 2019. Analysis of CYGNSS data for soil moisture retrieval. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(7), 2227-2235.

CYGNSS., 2024. CYGNSS Level 1 Science Data Record Version 3.2. Ver. 3.2. PO. DAAC, CA, USA. doi.org/10.5067/CYGNS-L1X32.

Entekhabi, D., Njoku, E. G., O'neill, P. E., Kellogg, K. H., Crow, W. T., Edelstein, W. N., Van Zyl, J., 2010. The soil moisture active passive (SMAP) mission. *Proceedings of the IEEE*, 98(5), 704-716.

Eroglu, O., Kurum, M., Boyd, D., Gurbuz, A. C., 2019. High spatio-temporal resolution CYGNSS soil moisture estimates using artificial neural networks. *Remote sensing*, 11(19), 2272.

Gonzalez-Abril, L., Nunez, H., Angulo, C., Velasco, F., 2014. GSVM: An SVM for handling imbalanced accuracy between classes inbi-classification problems. *Applied Soft Computing*, 17, 23-31.

Huang, G. B., Zhu, Q. Y., Siew, C. K., 2006. Extreme learning machine: theory and applications. *Neurocomputing*, 70(1-3), 489-501.

Jackson, T. J., Bindlish, R., Klein, M., Gasiewski, A. J., Njoku, E. G., 2003, July. Soil moisture retrieval and AMSR-E validation using an airborne microwave radiometer in SMEX02. In *IGARSS 2003. 2003 IEEE International Geoscience and Remote Sensing Symposium*, 1, 401-403.

Masters, D., Katzberg, S., Axelrad, P., 2003, July. Airborne GPS bistatic radar soil moisture measurements during SMEX02. In *IGARSS 2003. 2003 IEEE International Geoscience and Remote Sensing Symposium*, 2, 896-898.

Paloscia, S., Macelloni, G., Santi, E., Koike, T., 2001. A multifrequency algorithm for the retrieval of soil moisture on a large scale using microwave data from SMMR and SSM/I satellites. *IEEE Transactions on Geoscience and Remote Sensing*, 39(8), 1655-1661.

Persson, M., Haridy, S., 2003. Estimating water content from electrical conductivity measurements with short time - domain reflectometry probes. *Soil Science Society of America Journal*, 67(2), 478-482.

Price, J. C., 1977. Thermal inertia mapping: A new view of the earth. *Journal of Geophysical Research*, 82(18), 2582-2590.

Ruf, C. S., 2013. DDM data compression and decimation algorithm. *SPRL Technical Memorandum* No. 148–0046-X1.

Ruf, C., Unwin, M., Dickinson, J., Rose, R., Rose, D., Vincent, M., Lyons, A., 2013. CYGNSS: Enabling the future of hurricane prediction [remote sensing satellites]. *IEEE Geoscience and Remote Sensing Magazine*, 1(2), 52-67.

Rumelhart, D. E., Hinton, G. E., Williams, R. J., 1986. Learning representations by back-propagating errors. *nature*, 323(6088), 533-536.

Senyurek, V., Gurbuz, A., Kurum, M., Lei, F., Boyd, D., Moorhead, R., 2021. Spatial and temporal interpolation of CYGNSS soil moisture estimations. In *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, 6307-6310.

Senyurek, V., Lei, F., Boyd, D., Kurum, M., Gurbuz, A. C., Moorhead, R., 2020. Machine learning-based CYGNSS soil moisture estimates over ISMN sites in CONUS. *Remote Sensing*, 12(7), 1168.

Vereecken, H., Huisman, J. A., Bogena, H., Vanderborght, J., Vrugt, J. A., Hopmans, J. W., 2008. On the value of soil moisture measurements in vadose zone hydrology: A review. *Water resources research*, 44(4), W00D06.

Yu, K., 2021. *Theory and practice of GNSS reflectometry*. Berlin: Springer.