ISSN: 2414-1895 DOI: 10.6919/ICJE.202307_9(7).0036

# Research on GNSS-IR Soil Moisture Inversion Method based on Machine Learning

Jie Zhang, Xiaofang Liu*

School of Computer Science and Engineering, Sichuan University of Science & Engineering, Sichuan 643000, China

*Corresponding author: lxf1969@163.com

## Abstract

**Global Navigation Satellite System-Interferometric Reflectometry (GNSS-IR) is an emerging technology with high potential for soil moisture retrieval. In this paper, a soil moisture retrieval method combining wavelet transform and machine learning is proposed. The method replaces traditional polynomial fitting with wavelet transform to remove noise interference and obtain high-quality inverted soil moisture signals. Three machine learning models, Least Squares Support Vector Machine, Random Forest, and XGBoost, are used for comparative validation. The experimental results show that this method can effectively improve the quality of the reflected signal, suppress the impact of surface roughness, and enhance the accuracy of soil moisture retrieval. The model is established and evaluated using GNSS observations from the PBO P037 station in 2015 and 2016, with root mean square errors of 0.0086, 0.0083, and 0.0106 respectively.**

## Keywords

**GNSS-IR; Soil Moisture; Wavelet Transform; Machine Learning.**

## 1. Introduction

Soil moisture refers to the proportion of water in the soil by weight. It is an important parameter that affects plant growth and soil moisture management,and is of great significance in fields such as agriculture, forestry, environmental protection, and water resources management [1].Traditional soil moisture measurement methods include weight method, volume method, and resistance method [2], but these methods can only be used for small-scale measurement, are time-consuming, labor-intensive, and costly. With the development of technology, emerging technologies such as Global Navigation Satellite System-Interferometric Reflectometry (GNSS-IR) provide a new way for rapid and accurate measurement of soil moisture. This technology uses the observation values of signals reflected by the satellite propagation process received by the Global Navigation Satellite System (GNSS) receiver to monitor the surface environment. Compared with traditional methods, GNSS-IR technology can effectively compensate for the shortcomings of insufficient spatial and temporal resolution. At the same time, the interferometric remote sensing method based on GNSS-IR technology has become a new technical means, providing new ideas and methods for soil moisture measurement. This method can not only measure soil moisture over a large area, but also achieve non-contact, non-destructive measurement, greatly facilitating the practical application of soil moisture measurement [3].

GNSS-IR technology has attracted extensive research by scholars at home and abroad in soil moisture inversion. Since 2008, when Professor Larson [4] of the University of Colorado first proposed that the reflected signal component in GNSS signals' signal-to-noise ratio could be used for soil moisture inversion, this technology has been widely experimentally validated. Researchers have found that the three interference parameters (phase, frequency, and effective reflector height) are affected by

ISSN: 2414-1895 DOI: 10.6919/ICJE.202307_9(7).0036

changes in water content at the top 5cm of soil, with phase showing a linear correlation with surface soil moisture through experiments using PBO data [5-6]. Scholars such as Du Rui used the SVRM method to explore the relationship between characteristic parameters such as amplitude and phase of the interference signal and soil moisture, and to simultaneously estimate soil moisture through inputting characteristic parameters as the input values[7]. Liang Yueji and others used Plate Boundary Observatory (PBO) data to establish a multi-satellite linear regression inversion model for soil moisture, improving the inversion accuracy by increasing the number of satellites jointly inverted [8]. Through these studies, we can see that GNSS-IR technology has great potential in soil moisture inversion. The advantages of its rapid, accurate, and high-resolution will provide important support for soil moisture monitoring in agriculture, forestry, environmental protection, and water resource management.

Scholars have been exploring how to use the GNSS-IR method to obtain soil moisture, but traditional univariate linear regression models can usually only be used under ideal conditions. In the process of extracting reflection signals, scholars usually use low-order polynomial fitting methods, but the reflection signals obtained by this method often contain a lot of noise and are not ideal. To solve this problem, wavelet transform [9] can be used as an analysis method that represents the time and frequency of signals. It has the advantages of multi-resolution analysis and can express local features of time and frequency simultaneously. Therefore, combining wavelet transform and machine learning can improve the GNSS-IR soil moisture inversion model. The main idea of this method is to separate the reflection signal into a direct signal and a noise term using wavelet transform and then extract the signal trend term. Soil moisture inversion is then performed by combining these characteristics with machine learning methods. Compared with traditional methods, this method can effectively reduce noise interference and improve inversion accuracy and stability. This method has high practicality and broad application prospects and can provide important support for soil moisture monitoring in fields such as agriculture, forestry, environmental protection, and water resources management.

## 2. Basic Principles and Methods of Experiment

### 2.1 GNSS-IR Traditional Method for Soil Moisture Retrieval

This article compares and analyzes the observed data from the PBO monitoring network with the actual soil moisture data measured by nearby climate stations. The GNSS-IR method for inverting soil moisture utilizes the mechanism of interaction between GNSS signals reflected on the surface and underground soil. This method typically uses signal-to-noise ratio (SNR) observation data received by the receiver, including direct and reflected signals. The mathematical formula for SNR is [10]:

$$SNR^2 = A_c^{\ 2} = A_d^2 + A_r^2 + 2A_d A_r \cos\psi \tag{1}$$

where Ac is the composite signal amplitude; Ad and Ar are the amplitudes of the direct and reflected signals, respectively, and ψ is the phase difference between the direct and reflected signals, which can be separated from the obtained SNR observation values through multipath effects to obtain the multipath variation information caused by surface reflection in SNR. Since the value in the above equation is usually much larger than , a low-order polynomial fitting method is used to eliminate the trend term. Figure 1 shows the SNR residual sequence after removing the trend term. This article proposes a GNSS-IR soil moisture inversion model that combines wavelet transform and machine learning to overcome the problem of high noise in the reflected signal obtained by traditional methods.

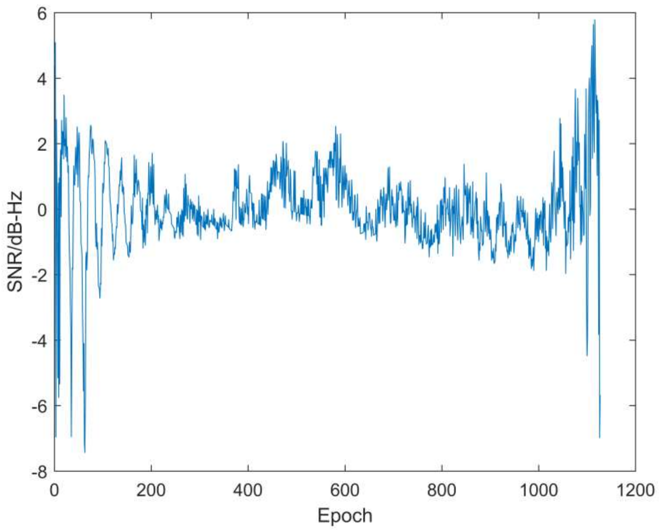


**Figure 1.** Signal-to-noise Ratio Residual Sequence after Removing the Trend Term

After removing the trend component, the remaining part of the satellite reflection signal can be approximated by a cosine model, as shown in Equation (2) [11]:

$$SNR_R = A\cos(\frac{4\pi H}{\lambda}\sin(\theta)+\varphi) \tag{2}$$

where SNRR is the reflection signal component, A is the amplitude, H is the equivalent reflector height, λ is the signal wavelength, θ is the satellite elevation angle, and φ is the phase. According to references [12-13], there is a strong correlation between the characteristics of the satellite reflection signal (including amplitude, phase, and frequency) and soil moisture, which can be used for soil moisture retrieval. The equivalent reflector height H can be obtained through L$omb$−$Scargle$ analysis (non-uniform periodic sampling) [14], as Equation (3) shows:

$$f = \frac{4\pi H}{\lambda} \tag{3}$$

where π and λ are constants. Therefore, H can be used to represent the reflection signal frequency in the inversion process. Then, the amplitude A and phase φ of the reflection signal can be obtained by fitting the time series of the reflection signal using the least squares method.

### 2.2 Wavelet Transform Theory

This paper uses the wavelet transform method to remove the trend term instead of the traditional low-order polynomial fitting method. Wavelet transform is a commonly used multiscale signal analysis method for extracting the characteristics of signals at different time and frequency scales by decomposing and reconstructing signals [15]. This paper uses the orthogonal symmetrical wavelet (sym) as the basis function, considering the oscillation and smoothness of the initial signal. Figures 2 and 3 show the SNR fitting result and the detrended SNR data of PRN01 satellite at the P037 station on DOY 100 in 2016. The horizontal axis represents the equally spaced observation epochs, the vertical axis represents the change of the SNR index in dB-Hz, and the data sampling rate is 15s. Traditional polynomial fitting performs poorly in the low-frequency part of satellite signals and is difficult to accurately fit the signal trend. Since the required reflection signal for final inversion is obtained by subtracting the fitted value from the original signal, accurately fitting the original SNR is crucial. The wavelet transform method can well express the signal trend and separate the high-frequency and low-frequency parts of the signal to extract a high-quality SNR residual sequence.

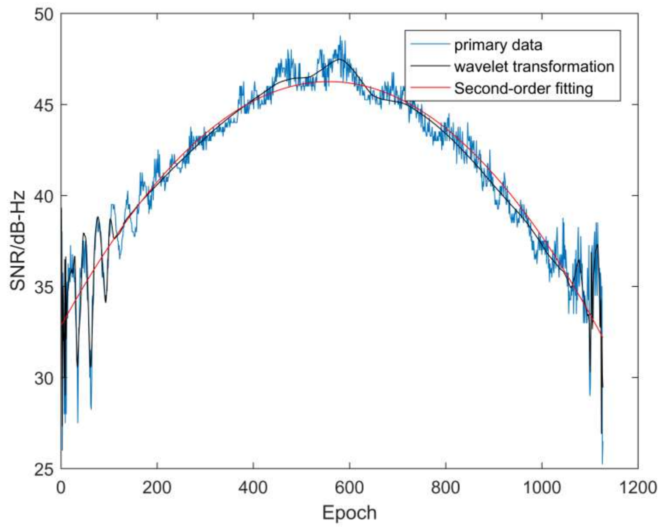


**Figure 2.** Fitting Diagram of Signal-to-noise Ratio of PRN 01 Satellite Signal

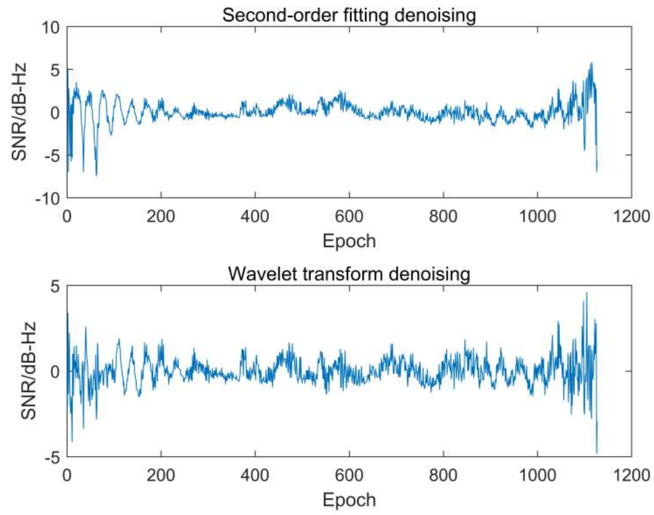


**Figure 3.** Trend-removing Diagram of PRN 01 Satellite by Two Methods.

## 2.3 Machine Learning Model

By using machine learning models such as Least Squares Support Vector Machines (LS-SVM), Random Forest, and XGBoost, various signal features can be trained and tested to infer soil moisture near the monitoring station.

### 2.3.1 Least Squares Support Vector Machines (LS-SVM)

Least Squares Support Vector Machines (LS-SVM) is an extension of Support Vector Machines (SVM), which is a supervised learning algorithm used for binary classification and regression problems. Similar to traditional SVM, LS-SVM is also constructed based on the principle of maximum margin. However, LS-SVM uses the method of minimizing the regularized squared error instead of maximizing the margin in traditional SVM. In LS-SVM, we need to find a hyperplane that separates the data into two classes. Unlike traditional SVM, we want the model to predict the minimum training error. During this process, the parameters of the hyperplane need to be optimized, and the optimal solution is obtained by solving a linear equation system. The advantage of LS-SVM

ISSN: 2414-1895 DOI: 10.6919/ICJE.202307_9(7).0036

is that it can handle nonlinear problems without explicitly mapping the data, because it uses the kernel function for mapping. In addition, it performs well in solving small sample problems [16-17].

#### 2.3.2 Random Forest

Random Forest (RF) uses the idea of ensemble learning to generate multiple independent sub-samples of the sample data through the Bootstrap algorithm, and each sub-sample generates a weak decision tree independently based on a certain feature set and splitting rule. Finally, the final prediction result is obtained by integrating the prediction results of multiple weak decision trees [18].

#### 2.3.3 XGBoost

XGBoost is a commonly used gradient boosting tree-based machine learning algorithm and ensemble learning algorithm that combines multiple weak classifiers to build a strong classifier. Specifically, it is an ensemble learning algorithm of decision trees that combines multiple decision trees to build a more powerful model. The core idea of XGBoost is to use gradient boosting technology to construct multiple decision tree models and combine them into a strong classifier [19].

## 3. GNSS-IR Soil Moisture Retrieval based on Machine Learning

### 3.1 Experimental Data Acquisition

To validate the reliability and accuracy of the LS-SVM, random forest, and XGBoost models in inverting soil moisture, this study downloaded data from the P037 station in the PBO for the years 2015 and 2016. The station is located near Chaffee County, Colorado, at an altitude of approximately 1600 meters. It was established in 2004 and uses a TRIMBLE NETRS receiver and a TRM29659.00 antenna, providing L1 observations every 15 seconds. The data from this station are representative and were collected relatively early for soil moisture analysis. The terrain around the station is flat and has no significant undulations, making it an ideal location for soil moisture experiments. In this study, the reference values for soil moisture were obtained from the PBO H2O station, with a sampling interval of one day. Since there were missing values in the soil moisture reference data, only 570 days of data were selected. Figure 4 shows the surface environment around the P037 station.

**Figure 4.** P037 Surface Environment

### 3.2 Experimental Data Processing

Firstly, RTKLIB software was used to obtain data from each satellite, including satellite elevation angle, azimuth angle, and signal-to-noise ratio (SNR) data in the L1 band ($\lambda$=19.05 cm). Then, single satellite data for the continuous observation time was selected, and the wavelet transform method was used to remove the trend component dominated by the direct component, instead of the traditional low-order polynomial fitting method, to obtain the SNR residuals required for the experiment, i.e., the reflected signal component.Next, the reflected signal component was subjected to a spectral analysis transformation to obtain the frequency f, and then a least-squares fit of equation (2) was performed to obtain the amplitude A and phase $\phi$. The three soil moisture characteristic vectors obtained were used as input data for the least-squares support vector machine model, random forest model, and XGBoost model, and the actual soil moisture values were used as output data for model construction. After multiple adjustments of the number of groups in the training set and test set, the 570 data sets were divided into 520 training sets and 50 test sets, with no overlap between the training set and test set data. In the traditional method, it was found that the phase had the highest correlation with soil moisture among the three attributes, so a linear model between soil moisture and phase was established, as shown in equation (4).

$$y = 0.0316x + 0.1165 \tag{4}$$

Where y represents the soil moisture value and x represents the signal phase value.

### 3.3 Analysis of Experimental Results

After testing on the training and testing sets, the performance of the traditional phase linear regression model, the least squares support vector machine model, the random forest model, and the XGBoost model on the testing set are shown in Figure 5. From Figure 5, it can be seen that the traditional phase linear regression model performs poorly when the soil moisture value is small, with a certain deviation, but the other three models perform well. The inversion effect of the least squares support vector machine model and the XGBoost model is not as ideal as that of the random forest model.

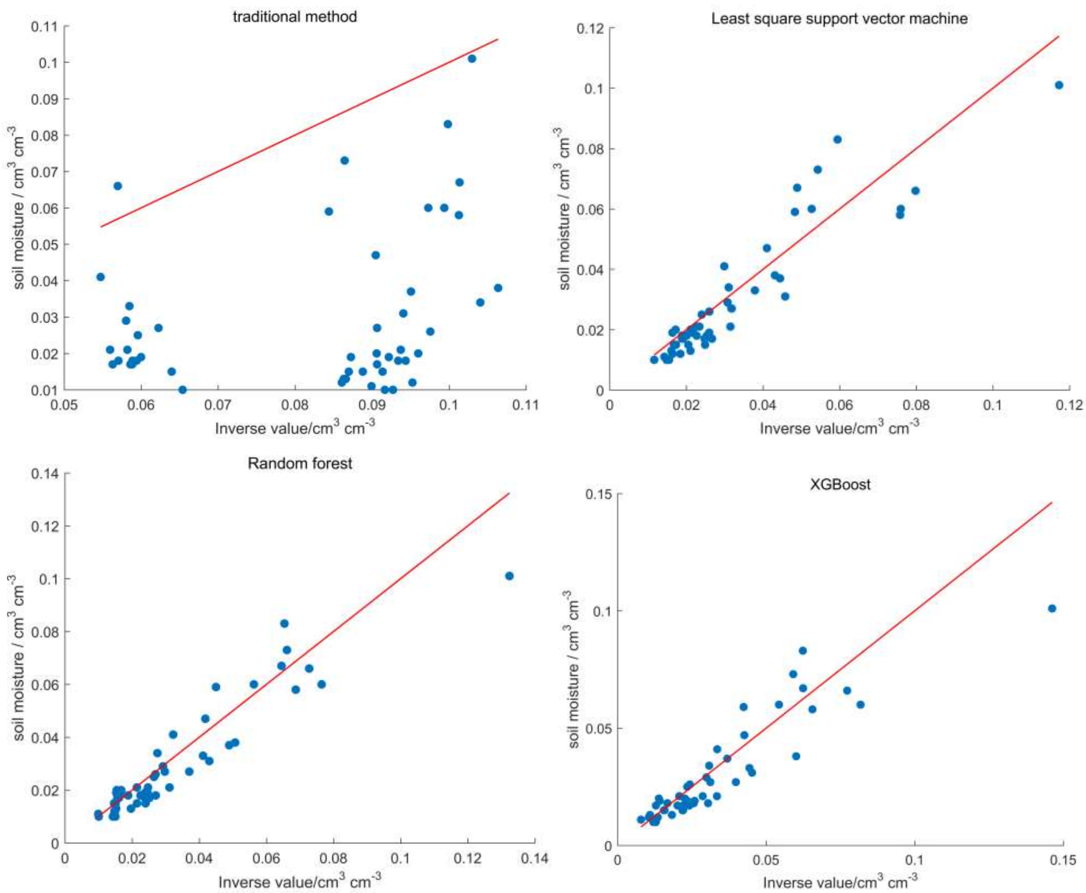


**Figure 5.** Analysis Diagram of Results of Four Models

Volume 9 Issue 7, 2023
DOI: 10.6919/ICJE.202307_9(7).0036

From Figure 6, it can be seen that the least squares support vector machine model, random forest model, and XGBoost model can generally fit the true soil moisture well. The traditional phase linear regression model performs well in trend changes, but has significant deviations on some days, especially when the true soil moisture is small. It can be clearly seen that when the true soil moisture varies greatly, the inversion results of the random forest model are closer to the true soil moisture. When the variation is slow, the least squares support vector machine model performs better. However, compared with the traditional phase linear regression model, all three models have significantly improved inversion accuracy.

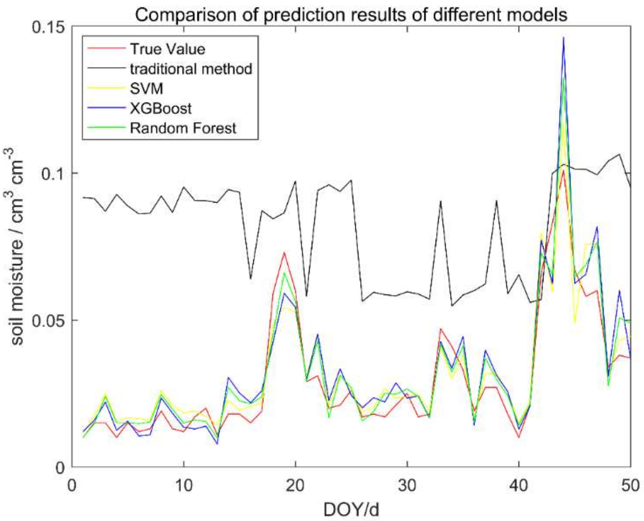


**Figure 6.** Comparison Between Inversion Results and True Values of Four Models

Table 1 presents a comparison of the inversion accuracy of the three models. The XGBoost model has a root-mean-square error (RMSE) and mean absolute error (MAE) of 0.0106 and 0.0073, respectively, and a coefficient of determination (R2) of 0.750, which is a significant improvement over the traditional phase-linear regression model. The RMSE and MAE of the least-squares support vector machine (LS-SVM) model and random forest model are 0.0086 and 0.0083, 0.0066 and 0.0063, respectively, with R2 values of 0.832 and 0.850, indicating higher accuracy compared to the above two models. This suggests that the wavelet transform method is more effective than traditional methods in removing trends. Figure 7 shows the absolute error between the four models and the true soil moisture values. The error of the traditional phase-linear regression model is controlled at around 0.008 cm3/cm-3, while the error of the LS-SVM model, random forest model, and XGBoost model is controlled within 0.02 cm3/cm-3, 0.03 cm3/cm-3, and 0.045 cm3/cm-3, respectively. Overall, the latter three models exhibit significantly improved accuracy.

**Table 1.** Comparison of Inversion Accuracy of Four Models

| *Model* | RMSE | MAE | $R^2$ |
|---|---|---|---|
| *Traditional phase regression* | 0.0568 | 0.0523 | 0.643 |
| *Least square support vector machine* | 0.0086 | 0.0066 | 0.832 |
| *Random forest* | 0.0083 | 0.0063 | 0.850 |
| XGBoost | 0.0106 | 0.0073 | 0.750 |

Volume 9 Issue 7, 2023
DOI: 10.6919/ICJE.202307_9(7).0036

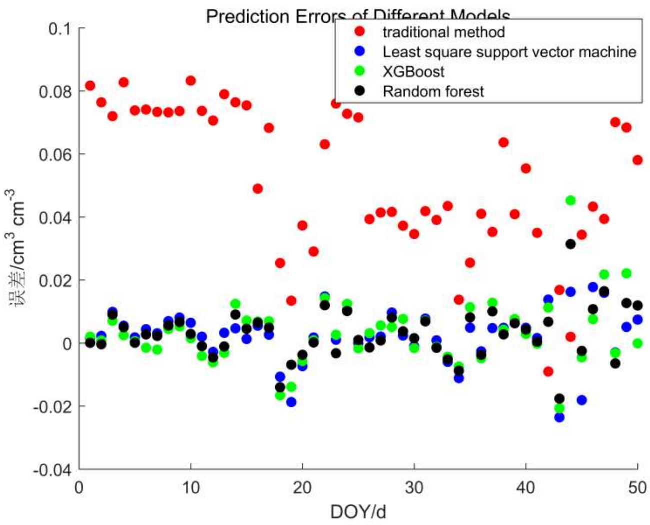


**Figure 7.** Error Comparison Diagram of Four Models

# 4. Conclusion

This article addresses the problems encountered in soil moisture inversion research, such as poor separation of reflected signals, unstable satellite signals, and the impact of surface roughness and vegetation coefficients. A model that combines wavelet transform and machine learning is proposed, and the following conclusions are drawn:

1) Improving the signal-to-noise ratio (SNR) processing method can improve the accuracy of soil moisture inversion to a certain extent, but cannot effectively avoid the occurrence of abnormal jumps during periods of drastic changes in soil moisture. Wavelet transform can effectively improve the separation performance of multipath signals by reducing the proportion of noise in the reflected signal.

2) The model combining wavelet transform and machine learning improves the inversion accuracy and effectively corrects errors compared to the traditional phase-linear regression model. The model not only reduces the satellite noise ratio but also effectively suppresses the influence of surface roughness and vegetation cover. This indicates that wavelet transform has significant advantages in removing satellite noise, while machine learning can effectively suppress the impact of surface roughness and vegetation coefficients.

In the future, the author will continue to explore how to achieve satellite joint inversion of soil moisture between different satellite systems and further reduce the impact of multipath environments to further improve the accuracy of GNSS-IR soil moisture inversion.

# Acknowledgments

Funding: Supported by The Innovation Fund of Postgraduate, Sichuan University of Science & Engineering under Grant Y2022186.

# References

[1] Yan Songhua, Gong Jianya, Xun Zhang Xie, et al. Ground-based experiment of measuring surface soil moisture by GNSS-R [J]. Journal of Geophysics, 2011, 54(11): 2735-2744.

[2] Luo Chenxi. Study on Inversion of Soil Moisture by GPS Signal-to-Noise Ratio [D]. Southwest Jiaotong University, 2018.

[3] Oscar Chan, Shen Fei, Cao Xinyun, Zhu Yifan. GNSS-IR Soil Moisture Retrieval Based on Deep Confidence Network [J]. Bulletin of Surveying and Mapping, 2020(09):100-105.

[4] Larson K M, Small E E, Gutmann E D, et al. Use of GPS receivers as a soil moisture network for water cycle studies[J]. Geophysical Research Letters, 2008, 35(24):L24405-n/a.

[5] Chew C C, Small E E, Larson K M, et al. Effects of near-surface soil moisture on GPS SNR data: Development of a retrieval algorithm for soil moisture[J]. IEEE Transactions on Geoscience and Remote Sensing, 2013, 52(1): 537-543.

[6] Chew C C, Small E E, Larson K M, et al. Vegetation sensing using GPS-interferometric reflectometry: Theoretical effects of canopy parameters on signal-to-noise ratio data[J]. IEEE Transactions on Geoscience and Remote Sensing, 2014, 53(5): 2755-2764.

[7] Duan Rui, Zhang Bo, Han Moutian, et al. Single antenna GNSS-R soil moisture retrieval using SVRM method [J]. Journal of Navigation and Positioning, 2018, 6(1): 34-39.

[8] Liang Yueji, Ren Chao, Huang Yibang, et al. Multi-satellite linear regression inversion model for monitoring soil moisture using GPS-IR [J]. 2020,46(6):1089-1096.

[9] Liu Guoyong. Research on Face Recognition Based on Gabor Wavelet Transform [D]. Northeastern University, 2008.

[10] Larson K M, Gutmann E D, Zavorotny V U, et al. Can we measure snow depth with GPS receivers?[J]. Geophysical research letters, 2009, 36(17):L17502.

[11] Guo Fei, Chen Weijie, Zhu Yifan, Zhang Xiaohong. A GNSS-IR soil moisture retrieval method integrating phase, amplitude and frequency [J/OL]. Journal of Wuhan University (Information Science Edition): 1-11 [April 29, 2023].

[12] Chew C, Small E E, Larson K M. An algorithm for soil moisture estimation using GPS-interferometric reflectometry for bare and vegetated soil[J]. GPS solutions, 2016, 20: 525-537.

[13] Chew C C, Small E E, Larson K M, et al. Effects of near-surface soil moisture on GPS SNR data: Development of a retrieval algorithm for soil moisture[J]. IEEE Transactions on Geoscience and Remote Sensing, 2013, 52(1): 537-543.

[14] Glynn E F, Chen J, Mushegian A R. Detecting periodic patterns in unevenly spaced gene expression time series using Lomb–Scargle periodograms[J]. Bioinformatics, 2006, 22(3): 310-316.

[15] Sidney. Overview on the application of wavelet transform in remote sensing [J]. Western Resources, 2013(01):178-180.

[16] Leong W C, Bahadori A, Zhang J, et al. Prediction of water quality index (WQI) using support vector machine (SVM) and least square-support vector machine (LS-SVM)[J]. International Journal of River Basin Management, 2021, 19(2): 149-156.

[17] Zhang Z, Ren C, Liang Y, et al. Inversion of Soil Moisture by GPS-IR Combined with Wavelet Analysis and LS-SVM[C]//China Satellite Navigation Conference (CSNC) 2020 Proceedings: Volume I. Springer Singapore, 2020: 27-37.

[18] Sun Bo, Zhang Chi, Yin Shichao, et al. Research on GNSS-IR Soil Moisture Inversion Method Based on PSO-RF[J].Radio Engineering, 2021, 51(10):1080-1085

[19] Jia Y, Jin S, Savi P, et al. GNSS-R soil moisture retrieval based on a XGboost machine learning aided method: Performance and validation[J]. Remote sensing, 2019, 11(14): 1655.