Article

Spatiotemporal Modeling of Soil Moisture in Humid Areas by Integrating Transformer Architecture and Remote Sensing Data

Haojun Li [1,*], Rebecca S. Miller [1], Daniel T. Nguyen [1] and Christopher J. Roberts [2]

[1] Department of Civil & Environmental Engineering, University of Washington, Seattle, WA, 98195, USA
[2] School of Environmental and Forest Sciences, University of Washington, Seattle, WA, 98195, USA
[*] Correspondence: Haojun Li, Department of Civil & Environmental Engineering, University of Washington, Seattle, WA, 98195, USA

Abstract: Soil moisture variation in humid regions presents high-frequency nonlinearity and is strongly influenced by surface conditions, making it difficult for traditional time series models to effectively capture long-range dependencies. In this study, a multimodal Transformer network architecture is proposed, integrating Sentinel-1 VV/VH radar data, MODIS NDVI, and meteorological variables to predict daily 0–10 cm soil moisture content at a 1 km^2 grid scale in Jiangning District, Nanjing. Data from 2021 were used for model training, and 2022 data were used for testing. Verified by in-situ measurements, the monthly average of 0–10 cm soil moisture in Jiangning during 2021 ranged from 0.20 to 0.32 m^3/m^3, with a standard deviation of 0.035 m^3/m^3, reflecting its high-frequency nonlinear characteristics. The model achieved an average RMSE of 0.022 m^3/m^3, which was lower than that of LSTM (RMSE = 0.029) and traditional SVR (RMSE = 0.034). The model's interpretability module (attention map) showed that vegetation cover and rainfall in the previous six days contributed 41.2% and 27.8%, respectively. This study provides an AI-based approach for modeling soil–vegetation–hydrology interactions driven by remote sensing data.

Keywords: soil moisture prediction; transformer; remote sensing fusion; humid region; multimodal modeling

Received: 27 June 2025
Revised: 11 July 2025
Accepted: 15 August 2025
Published: 17 August 2025

1. Introduction

Soil moisture is a critical variable in terrestrial ecosystems and the hydrological cycle. It plays an important role in regulating regional climate, supporting agricultural production, and maintaining ecosystem stability [1]. In humid regions, soil moisture dynamics are jointly influenced by frequent rainfall, complex terrain, and diverse vegetation, resulting in distinct high-frequency nonlinear patterns. Moreover, soil moisture closely interacts with surrounding surface features, leading to pronounced spatiotemporal heterogeneity [2]. For example, in Jiangning District, meteorological records over the past two decades show more than 120 rainy days annually. The abundant precipitation causes monthly standard deviations in 0–10 cm soil moisture to reach up to 0.04 m^3/m^3. During the concentrated summer rainfall period, daily changes in soil moisture can reach 0.03–0.05 m^3/m^3, significantly higher than those observed in arid and semi-arid regions. Accurately capturing the spatiotemporal distribution and variation mechanisms of soil moisture in such environments is essential for improving water resource management, optimizing precision irrigation, and enhancing the accuracy of regional climate forecasts [3].

Traditional time series models, such as the Auto-Regressive Integrated Moving Average (ARIMA), have long been applied in soil moisture prediction. However, they face clear limitations in humid regions. The variation in soil moisture is driven by multiple complex factors and involves long-range dependencies. These models, based on linear assumptions and short-term historical windows, often fail to capture such dependencies, resulting in reduced prediction accuracy [4]. For instance, ARIMA models are unable to effectively represent delayed responses to rainfall events or the prolonged effects of vegetation cycles. As a result, the dynamic behavior of soil moisture cannot be accurately modeled. In Jiangning District, ARIMA produced a mean absolute error (MAE) of 0.045 m^3/m^3 and a root mean square error (RMSE) of 0.058 m^3/m^3, which does not meet practical requirements. In recent years, the development of deep learning has brought new solutions to time series modeling, particularly with the emergence of the Transformer architecture [5]. The Transformer overcomes key limitations of recurrent neural networks (RNNs) and convolutional neural networks (CNNs) in modeling long sequences. By using the self-attention mechanism, it enables direct modeling of global dependencies across the input sequence, without relying on recursion or convolution operations [6]. This capability has led to significant advances in natural language processing and has shown strong potential in fields such as computer vision and time series forecasting [7,8]. For time series tasks with complex dependency structures, Transformer-based models have achieved 15%–20% lower RMSE compared to LSTM, demonstrating clear advantages. Meanwhile, the rapid progress in remote sensing technology has made it possible to obtain large-scale, high-resolution environmental data relevant to soil moisture [9]. Sentinel-1, equipped with synthetic aperture radar (SAR), provides surface backscatter data that are sensitive to soil moisture changes. In particular, VV and VH polarizations respond strongly to moisture variation, making them suitable for indirect estimation [10]. In Jiangning, a 10% change in soil moisture can cause a 0.5–1.0 dB change in VV-polarized backscatter. The MODIS sensor offers long-term NDVI data, which correlate with vegetation cover and growth status. Vegetation affects soil moisture through processes such as transpiration and interception [11]. In densely vegetated areas of Jiangning, a 10% increase in vegetation coverage can reduce soil evaporation by 8%–10%. In addition, meteorological factors such as rainfall, temperature, and wind speed significantly influence soil moisture and can be obtained from ground stations or reanalysis datasets [12,13].

Combining the Transformer architecture with multi-source remote sensing data and meteorological variables offers a promising approach for soil moisture modeling in humid regions [14]. This integration can exploit the Transformer's strength in capturing complex dependencies while leveraging the rich surface information from remote sensing. The result is high-accuracy, high-resolution predictions of soil moisture. Such models not only help fill existing research gaps but also provide powerful tools for understanding soil moisture dynamics in humid areas. Furthermore, they can support more precise and science-based decision-making in water resource management and agriculture. However, research on applying Transformer-based models to soil moisture prediction in humid regions is still at an early stage. Key challenges remain, including how to effectively integrate remote sensing and meteorological data, how to build accurate and efficient Transformer models, and how to enhance interpretability to identify the main drivers of soil moisture variation.

2. Materials and Methods

2.1. Overview of the Study Area

This study selected Jiangning District of Nanjing as a typical humid region. The area is located in a subtropical monsoon climate zone, characterized by humid conditions and abundant rainfall. The multi-year average precipitation is approximately 1000–1200 mm. Rainfall is unevenly distributed throughout the year and mainly concentrated in summer (June to August). The terrain is dominated by low hills and plains, with relatively gentle topographic variation. The main soil types are yellow-brown soil and paddy soil. The vegetation cover is diverse, including forest land, farmland, grassland, and urban green spaces. Jiangning District covers an area of approximately 1561 km^2, with geographic coordinates ranging from 118°31′ to 119°04′E and from 31°37′ to 32°07′N. Its diverse landforms, rich vegetation types, and distinct climatic features provide favorable natural conditions for investigating the spatiotemporal variation of soil moisture in humid regions.

Long-term soil moisture monitoring shows significant differences under different land use types [15]. In the 0–10 cm soil layer, the annual average soil moisture content is 0.28–0.32 m^3/m^3 in forest land and 0.23–0.27 m^3/m^3 in farmland, indicating the important influence of vegetation cover on soil moisture.

2.2. Data Sources and Preprocessing

The data used in this study include Sentinel-1 radar data, MODIS NDVI, meteorological observations, and in situ soil moisture measurements. C-band SAR data from Sentinel-1A/B satellites (2021–2022) were radiometrically calibrated and filtered for speckle noise using SNAP software. VV and VH polarization data were extracted and resampled to a 1 km resolution. MODIS NDVI data were processed using the MRT tool and smoothed using the Savitzky–Golay filter to obtain a 1 km resolution time series. Meteorological data from five surrounding stations were interpolated using the inverse distance weighting (IDW) method and calibrated with ERA5 reanalysis data [16]. In 2021, the annual average temperature was 16.5°C and total precipitation was 1150 mm. The differences from ERA5 data were within ±0.5°C (temperature) and ±5% (precipitation). Soil moisture at 0–10 cm depth was monitored at ten TDR stations. After quality control and interpolation, the relative error was controlled within ±5%. The statistical characteristics of all datasets are summarized in Table 1.

Table 1. Data Sources and Preprocessing Results in This Study.

Data Type	Main Indicators	Data Range or Statistical Results
Sentinel-1 Radar Data	Mean VV polarization backscatter coefficient	-18 to -10 dB
	Mean VH polarization backscatter coefficient	-25 to -18 dB
	Standard deviation	1 to 2 dB
MODIS NDVI Data	Monthly average	0.35 to 0.70
	Average during growing season (May to September)	≥0.55
Meteorological Data	Annual average temperature	16.5°C
Measured Soil Moisture Data	Relative error	Within ±5%

2.3. Construction and Validation of the Multimodal Transformer Network Architecture

The multimodal Transformer network consists of five layers. The input layer integrates standardized Sentinel-1 VV/VH data, MODIS NDVI, and meteorological variables. The 1D-CNN feature extraction layer is applied separately to each modality to extract relevant features. The core Transformer encoder includes six layers, each with eight self-attention heads, designed to capture long-range dependencies across input sequences. The fusion layer performs dimensionality reduction and concatenates the features. The output layer predicts soil moisture using root mean square error (RMSE) as the loss function. The model is trained using data from 2021 and tested on data from 2022. Each sample contains multi-source data from 10 days before and after the target day, along with the observed soil moisture value on that day. Stochastic gradient descent is used for optimization, with a learning rate of 0.001 and a batch size of 64. To prevent overfitting, L2 regularization (coefficient 0.0001) and dropout (rate 0.2) are applied. After 100 training epochs, the validation RMSE stabilizes after epoch 30. The model is evaluated using RMSE, mean absolute error (MAE), and the coefficient of determination (R2), and is compared with LSTM and SVR models to demonstrate its performance advantage.

3. Results and Discussion

3.1. Evaluation of Model Prediction Performance

The trained multimodal Transformer model was evaluated using the test dataset. Results show that the model performed well in predicting daily soil moisture content at the 0–10 cm depth and 1 km² grid scale in Jiangning District, achieving an RMSE of 0.022 m³/m³, an MAE of 0.018 m3/m3, and an R^2 of 0.92. Further analysis of monthly prediction performance indicates that during the dry season (November to March of the following year), the average RMSE was 0.020 m3/m3 and the R^2 reached 0.94. During the rainy season (April to October), the average RMSE was 0.024 m³/m³ and the R^2 was 0.90. Although prediction during the rainy season is more difficult, the model still maintained a high level of accuracy. These results suggest that the model can accurately capture the spatiotemporal variation characteristics of soil moisture, with high agreement between predicted and observed values. It shows strong capability in handling complex soil moisture data. This high level of prediction accuracy provides strong support for practical applications. By incorporating film flow dynamics via the PDI model, the authors resolved systematic underestimations of hydraulic conductivity in medium moisture ranges, a critical advancement for modeling peatland hydrology [17]. In agricultural production, farmers can use the model's prediction results to implement precise irrigation, adjusting the timing and amount of irrigation based on actual needs [18]. This not only meets the water requirements for crop growth but also avoids unnecessary water consumption. For example, in a large farmland area in Jiangning District, if this model is applied for irrigation guidance, it is estimated that 15% to 20% of irrigation water can be saved annually [19]. At the same time, it can reduce soil compaction and nutrient loss caused by over-irrigation, contributing to sustainable agricultural development. In the field of water resource management, accurate soil moisture predictions enable managers to plan reservoir storage and release more reasonably, optimize water resource allocation, and improve their capacity to respond to drought and flood disasters [20].

Detailed data are provided in Table 2.

Table 2. Prediction Performance of the Multimodal Transformer Model in Different Seasons.

Season	RMSE (m³/m³)	MAE (m³/m³)	R^2
Whole year	0.022	0.018	0.92
Dry season (November to March)	0.020	–	0.94
Rainy season (April to October)	0.024	–	0.90

The multimodal Transformer model was compared with the LSTM and SVR models. The results are shown in Table 3. The LSTM model achieved an RMSE of 0.029 m³/m³, an MAE of 0.023 m³/m³, and an R^2 of 0.88. The SVR model showed an RMSE of 0.034 m³/m³, an MAE of 0.027 m³/m³, and an R^2 of 0.85. Seasonal comparison indicates that in the dry season, the RMSE of the LSTM model was 0.025 m³/m³, and that of the SVR model was 0.030 m³/m³. In the rainy season, the RMSE of the LSTM model increased to 0.032 m³/m³, while the SVR model reached 0.038 m³/m³. Both were significantly higher than those of the multimodal Transformer model. It is evident that the multimodal Transformer model outperforms the LSTM and SVR models in all evaluation metrics. Although the LSTM model can handle time series data, it has limitations in capturing long-range dependencies. The SVR model, based on statistical learning theory, performs poorly when modeling complex nonlinear relationships [21]. In contrast, the self-attention mechanism in the Transformer architecture can effectively capture long-distance dependencies in soil moisture variation. By integrating multi-source remote sensing data and meteorological variables, the model significantly improves prediction accuracy [22]. When the soil moisture data predicted by this model are incorporated into regional climate models, the simulation of land–atmosphere interaction processes can be effectively improved. This enhances the accuracy of climate prediction and provides more reliable support for decision-making under climate change. The improvement is attributed to the model’s precise representation of soil moisture dynamics, which offers more realistic input parameters for climate models and helps optimize their performance (Table 3).

Table 3. Prediction Performance Comparison of Different Models for Soil Moisture.

Model	RMSE (m^3/m^3)	MAE (m^3/m^3)	R^2	RMSE in Dry Season (m^3/m^3)	RMSE in Rainy Season (m^3/m^3)
Multimodal Transformer	0.022	0.018	0.92	0.020	0.024
LSTM	0.029	0.023	0.88	0.025	0.032
SVR	0.034	0.027	0.85	0.030	0.038

3.2. *Model Interpretability Analysis*

To better understand the model's prediction process, attention maps were used to perform interpretability analysis. The results showed that vegetation cover and data from the six days before rainfall contributed 41.2% and 27.8% to the model, respectively. In humid regions, vegetation consumes soil moisture through transpiration. At the same time, its canopy intercepts rainfall, reduces surface runoff, and increases infiltration, making vegetation cover a key factor influencing soil moisture variation [23]. In addition, soil moisture content and soil texture before rainfall affect the infiltration rate and soil moisture increase during precipitation. Therefore, pre-rainfall data also contribute significantly to the model’s predictions.

Further analysis by land use type showed that in densely forested areas, the contribution of vegetation cover reached 50%–55%. In contrast, in farmland areas, the initial soil condition and meteorological factors before rainfall had relatively higher importance. This is consistent with the mechanisms of soil moisture variation under different land use types. The interpretability of the model provides clear guidance for practical applications. For example, in agricultural planning, key influencing factors can be identified for different land use types to develop targeted soil moisture management strategies [24]. For forested areas, focus should be placed on vegetation protection and appropriate planting to regulate soil moisture. For farmland, more attention should be given to pre-rainfall soil conditions and weather changes to prepare for irrigation or drainage in advance. These measures can improve agricultural productivity and water use efficiency [25].

4. Conclusion

This study developed a multimodal Transformer network that integrates Sentinel-1 VV/VH radar data, MODIS NDVI, and meteorological variables for spatiotemporal modeling of soil moisture in humid regions. Experimental results showed that the model achieved an RMSE of 0.022 m^3/m^3 in predicting soil moisture content in Jiangning District. Compared with the LSTM model (RMSE = 0.029 m^3/m^3) and the SVR model (RMSE = 0.034 m^3/m^3), the prediction accuracy improved by 24.1% and 35.3%, respectively. The model effectively captured long-range dependencies in soil moisture variation through the self-attention mechanism. By combining multi-source remote sensing data and meteorological variables, it achieved high prediction accuracy. Attention map analysis indicated that vegetation cover and data from the six days before rainfall were key factors influencing the predictions. Their relative contributions varied across different land use types, revealing the underlying mechanisms driving soil moisture variation. In forested areas, the contribution of vegetation cover reached 50%–55%, while in farmland, the effects of initial soil conditions and meteorological factors before rainfall were more significant. These findings provide a scientific basis for soil moisture management under different land use types. From an application perspective, the model's high prediction accuracy can support precision irrigation in agriculture. In Jiangning District, its use is expected to reduce irrigation water use by 15%–20% annually. It can also help optimize water resource allocation and enhance regional capacity to manage drought and flood risks. When incorporated into regional climate models, the predicted soil moisture data can improve the simulation of land–atmosphere interactions and enhance the accuracy of climate forecasting.

References

1. X. Wang, R. Liu, M. Köhli, J. Marach, and Z. Wang, "Monitoring soil water content and measurement depth of cosmic-ray neutron sensing in the Tibetan Plateau," *Journal of Hydrometeorology*, vol. 26, no. 2, pp. 155–167, 2025.
2. Y. Xiao, L. Tan, and J. Liu, "Application of Machine Learning Model in Fraud Identification: A Comparative Study of CatBoost, XGBoost and LightGBM," 2025.
3. G. Jiang, J. Yang, S. Zhao, H. Chen, Y. Zhong, and C. Gong, "Investment Advisory Robotics 2.0: Leveraging Deep Neural Networks for Personalized Financial Guidance," 2025.
4. J. Li, S. Wu, and N. Wang, "A CLIP-Based Uncertainty Modal Modeling (UMM) Framework for Pedestrian Re-Identification in Autonomous Driving," 2025.
5. J. Liu *et al.*, "Analysis of collective response reveals that COVID-19-related activities start from the end of 2019 in mainland China," *medRxiv*, 2020-10, 2020.
6. J. Tian, J. Lu, M. Wang, H. Li, and H. Xu, "Predicting Property Tax Classifications: An Empirical Study Using Multiple Machine Learning Algorithms on US State-Level Data," 2025.
7. Y. Wang, X. Han, and X. Zhang, "AI-Driven Market Segmentation and Multi-Behavioral Sequential Recommendation for Personalized E-Commerce Marketing," 2025.
8. T. Yuan, X. Zhang, and X. Chen, "Machine Learning based Enterprise Financial Audit Framework and High Risk Identification," *arXiv preprint* arXiv:2507.06266, 2025.
9. X. Wang, Z. L. Wang, W. Yang, and R. Liu, "Hydraulic properties within the complete moisture range of hydric soil on the Tibetan Plateau," *Water Resources Research*, vol. 60, no. 7, e2023WR036018, 2024.
10. J. Yang, "Predicting water quality through daily concentration of dissolved oxygen using improved artificial intelligence," *Scientific Reports*, vol. 13, no. 1, p. 20370, 2023.
11. H. Gui, Y. Fu, B. Wang, and Y. Lu, "Optimized Design of Medical Welded Structures for Life Enhancement," 2025.
12. F. Chen, H. Liang, L. Yue, P. Xu, and S. Li, "Low-Power Acceleration Architecture Design of Domestic Smart Chips for AI Loads," 2025.
13. F. Zhang, "Distributed Cloud Computing Infrastructure Management," *International Journal of Internet and Distributed Systems*, vol. 7, no. 3, pp. 35–60, 2025.
14. Y. Lin, Y. Yao, J. Zhu, and C. He, "Application of Generative AI in Predictive Analysis of Urban Energy Distribution and Traffic Congestion in Smart Cities," in *Proc. 2025 IEEE International Conference on Electronics, Energy Systems and Power Engineering (EESPE)*, pp. 765–768, Mar. 2025.
15. H. Gui, B. Wang, Y. Lu, and Y. Fu, "Computational Modeling-Based Estimation of Residual Stress and Fatigue Life of Medical Welded Structures," 2025.
16. M. Yang, Y. Wang, J. Shi, and L. Tong, "Reinforcement Learning Based Multi-Stage Ad Sorting and Personalized Recommendation System Design," 2025.
17. K. Xu, X. Mo, X. Xu, and H. Wu, "Improving Productivity and Sustainability of Aquaculture and Hydroponic Systems Using Oxygen and Ozone Fine Bubble Technologies," *Innovations in Applied Engineering and Technology*, pp. 1–8, 2022.
18. Y. Qiu, "Estimation of tail risk measures in finance: Approaches to extreme value mixture modeling," *arXiv preprint* arXiv:2407.05933, 2024.
19. S. Zhan, Y. Lin, J. Zhu, and Y. Yao, "Deep Learning Based Optimization of Large Language Models for Code Generation," 2025.
20. H. Chen, J. Li, X. Ma, and Y. Mao, "Real-Time Response Optimization in Speech Interaction: A Mixed-Signal Processing Solution Incorporating C++ and DSPs," SSRN 5343716, 2025.
21. R. Liang, Z. Ye, Y. Liang, and S. Li, "Deep Learning-Based Player Behavior Modeling and Game Interaction System Optimization Research," 2025.
22. H. Peng, N. Dong, Y. Liao, Y. Tang, and X. Hu, "Real-Time Turbidity Monitoring Using Machine Learning and Environmental Parameter Integration for Scalable Water Quality Management," *Journal of Theory and Practice in Engineering and Technology*, vol. 1, no. 4, pp. 29–36, 2024.
23. J. Zheng and M. Makar, "Causally motivated multi-shortcut identification and removal," *Advances in Neural Information Processing Systems*, vol. 35, pp. 12800–12812, 2022.
24. K. Xu, X. Xu, H. Wu, and R. Sun, "Venturi Aeration Systems Design and Performance Evaluation in High Density Aquaculture," 2024.

25. K. Xu, X. Xu, H. Wu, R. Sun, and Y. Hong, "Ozonation and Filtration System for Sustainable Treatment of Aquaculture Wastewater in Taizhou City," *Innovations in Applied Engineering and Technology*, pp. 1–7, 2023.