Article:
Tong, Y., Wang, Y., Li, Z. et al. (2025) Deciphering spatial pattern and environmental drivers of deep soil moisture security: insights from integrating spatial non-stationarity in large-scale analysis. Catena, 261. 109516. ISSN: 0341-8162

https://doi.org/10.1016/j.catena.2025.109516

Deciphering spatial pattern and environmental drivers of deep soil moisture security: Insights from integrating spatial non-stationarity in large-scale analysis

Yongping Tong[1, 2, 3, 4], Yunqiang Wang[1, 2, 4, 5, *], Zimin Li[1, 2, 4], Pingping Zhang[1, 2, 4], Wei Hu[6], Jingxiong Zhou[1, 2, 4], Xiangyu Guo[1, 4, 7], Hui Sun[1, 4, 7], Alexis Comber[8], Ronny Lauerwald[3]

[1] State Key Laboratory of Loess Science, Institute of Earth Environment, Chinese Academy of Sciences, Xi'an 710061, China

[2] University of Chinese Academy of Sciences, Beijing 100049, China

[3] Université Paris-Saclay, INRAE, AgroParisTech, UMR Ecosys, Palaiseau 91120, France

[4] National Observation and Research Station of Earth Critical Zone on the Loess Plateau, Xi'an 710061, China

[5] Department of Earth and Environmental Sciences, Xi'an Jiaotong University, Xi'an 710049, China

[6] The New Zealand Institute for Plant and Food Research Limited, Private Bag 4704, Christchurch 8140, New Zealand

[7] Xi'an Institute for Innovative Earth Environment Research, Xi'an 710061, China

[8] School of Geography, University of Leeds, Leeds, LS2 9JT, UK

*** Corresponding author:** Yunqiang Wang (wangyq@ieecas.cn)

Address: State Key Laboratory of Loess Science, Institute of Earth Environment, Chinese Academy of Sciences, Xi'an, Shaanxi 710061, China.

Abstract

Soil moisture (SM) profoundly influences global ecosystem services and climate change, creating the broad needs for SM management. However, existing single-scale studies have limited representativeness and overlook the spatial non-stationarity, while treating regions with diverse SM conditions as a whole may obscure interactions between environment and different SM levels. Therefore, we expanded the study scale both horizontally and vertically and established an "SM security" framework, which classifies SM conditions based on their impacts on evapotranspiration, vegetation growth, and soil quality. SM was measured to 500 cm depth across the Yellow River Basin (YRB) (795,000 km^2) and categorized into three security zones including "wet zone" that SM ≥ 80% * field capacity (FC), "transitional zone" that permanent wilting point (PWP) < SM < 80% * FC, and "dry zone" that SM ≤ PWP. In the YRB, the transitional zone was predominant (61.58%), followed by wet (20.24%) and dry (18.17%) zones. In relatively stable layers (110–500 cm), the wet zone expanded with depth (17.04% to 25.86%) while dry zone contracted (20.76% to 14.99%). The mean relative SM (dimensionless) and available SM storage, indicating relative soil saturation and vegetation water availability, were 0.84 and 32.14 cm, respectively. Considering spatial non-stationarity, environmental drivers shaping the SM security pattern were: slope in wet zone, vegetation in dry zone, and clay content, vegetation, and slope in transitional zone, with notable coupling among these factors. These findings help characterizing SM security and developing targeted SM management measures in the YRB and similar regions worldwide.

Keywords

Deep soil layer; Soil moisture security; Spatial heterogeneity; Hyper-local GWR; Machine learning

1. Introduction

Soil moisture (SM) affects soil-vegetation-atmosphere interactions in various ways (Huang and Shao, 2019; Liu et al., 2020; Zuo et al., 2024), and plays a pronounced role in regulating climate processes and ecosystem service functions (Fu et al., 2022; Seneviratne et al., 2010). In climate transitional zones, which are considered as a hotspot of land-atmosphere coupling (Koster et al., 2004), SM is one of the most important factors influencing climate through both local and non-local effects (Seneviratne et al., 2010). In water limited zones such as dryland ecosystems, SM directly constrains evapotranspiration and controls most of ecosystem processes (Zhou et al., 2021). Strong and persistent SM deficits may transform terrestrial ecosystems from “carbon sinks” to “carbon sources” (Green et al., 2019). Indeed, diverse SM conditions can lead to varied climate and ecological states. Analyzing spatial regions with diverse SM conditions as a unified whole may obscure and confound the interaction mechanisms between environmental factors and different SM levels. To address this, it is necessary to establish an SM grading scheme based on SM levels to characterize SM-environment interactions in specific SM zones. This concept underpins the “SM security” framework proposed here, which is a classification system that categorizes SM conditions based on their impact on evapotranspiration, vegetation growth and soil quality. By applying this SM security classification system, we can map SM security zones, identify their spatial patterns, and evaluate the environmental control mechanism of these patterns. Importantly, this classification system creates a monitorable baseline that allows for identifying SM changes and evaluating their availability to meet potential future ecological or hydrological demands in specific security zones. Proposing targeted SM management measures based on these insights can contribute to the stability and sustainability of ecosystems.

Driven by heterogeneity in soil properties, topography, vegetation, climate, human activities, and their interactions (Huang and Shao, 2019), SM exhibits variability both vertically and horizontally (Seneviratne et al., 2010). Vertically, shallow SM (≤ 100 cm) shows greater vertical and horizontal heterogeneity and temporal variability, while deep SM (> 100 cm) remains relatively stable (Cao et al., 2018; Yang et al., 2012). Deep SM serves as an important water source for plant growth and plays a key role in drought resistance (Miguez-Macho and Fan, 2021). In fact, water deficits in deep soil layers have a more persistent negative impact on gross primary production than dry shallow soil (Fu et al., 2022). Recent researches also revealed that deep SM may affect future climates and ecosystems in ways distinct form surface SM (Fu et al., 2022; Hirschi et al., 2014; Schlaepfer et al., 2017). This underscores the necessity to investigate SM security in both shallow and deep soils to yield more representative outcomes. Horizontally, numerous studies have demonstrated that the extent of spatial variability of SM and its dominant factors largely depend on the scale of the investigation (Cao et al., 2018). Due to the heterogeneity of geographical elements affecting soil hydrological processes, the representativeness of small-scale SM studies is inherently limited. As a result, generalizing findings from small watersheds to broader regional scales is often challenging (Gruber et al., 2018). Additionally, spatially adjacent geographic factors often exhibit spatial correlation; however, when geographic locations vary, the relationships or structures among spatial variables may differ due to the phenomenon known as spatial non-stationarity (Brunsdon et al., 1996). Large-scale studies often overlook spatial non-stationarity, potentially missing variations in interaction mechanisms across different geographic units (Tran et al., 2022). Whether at small or large scales, single-scale analysis limits the ability to achieve a general, cross-scale understanding of SM security assessment. Remote sensing enables rapid, cost-effective acquisition of large-scale geospatial data (Kucukpehlivan et al., 2023; Cevik Degerli and Cetin, 2023; Zeren Cetin et al., 2023). However, current large-scale SM studies based on direct remote sensing data (e.g., SMAP, SMOS) generally face limited detecting depth (approximately the top 5 cm) (Chen et al., 2021). Although data assimilation that integrate *in situ* observations and model simulations can extend the effective retrieval depth to approximately 1 m (Baldwin et al., 2017; Das et al., 2008), reliable large-scale observations of deeper SM remain challenging to obtain. To date, the large-scale *in situ* SM studies are also generally restricted to shallow soils (Gruber et al., 2013), while many deep SM studies mainly focused on slopes and small catchment scales (Cao et al., 2018; Wang et al., 2013). Therefore, there is a pressing need for SM security studies that account for spatial non-stationarity over wide spatial extents and deeper soil profiles to generate a comprehensive understanding of SM security.

Here we present an SM security study across the Yellow River Basin (YRB). The Yellow River is the fifth longest river in the world and the YRB covers many of China's most important ecological barriers, making it of great ecological and climatic significance (Yin et al., 2019). Previous SM research in the basin has mainly focused on near-surface layers. At the basin scale, satellite-based products and reanalysis dataset have been used to map surface SM, revealing pronounced spatial heterogeneity in SM itself and its interaction with environment (Hu et al., 2025; Zhang et al., 2024). In the YRB, numerous ecohydrological studies have assessed the impacts of large-scale ecological restoration programs such as "Grain for Green". While these programs significantly improved vegetation cover, they also increased vegetation water use, reduced SM availability, and led to widespread deep soil desiccation or dried-soil layers (Chen et al., 2019; Ge et al., 2020; Jia et al., 2017; Zhang et al., 2018). Other studies suggest that SM declines are more strongly linked to land-use conversion than to climate variability, underscoring the complexity of ecohydrological trade-offs in this region (Li et al., 2025; Wang et al., 2024b). Attribution analyses have further shown that vegetation dynamics are the dominant controls of SM variability, with temperature and wind playing secondary roles (Guo et al., 2022). Despite these advances, most investigations are limited to surface or shallow soils, and systematic deep-profile (> 1 m) studies at the basin scale remain scarce. Meanwhile, the YRB itself presents a strong need to improve SM availability: extensive ecological restoration has altered water cycle processes (Wang et al., 2024a), and the basin faces severe water stress due to the imbalance between high demand and limited supply (Liu et al., 2024). Together with its diverse climatic zones, vegetation types, and soil properties, these factors make the YRB a highly representative area for assessing SM security. Moreover, insights gained here may provide references for other regions worldwide with similar environmental backgrounds.

Hence, we included the YRB as research area to identify and evaluate the situation and spatial pattern of SM security in the YRB, and decipher environmental drivers that shapes the spatial patterns of deep SM security in large-scale area. To address the limitation of single-scale and shallow-depth studies, our first objective was to identify and assess SM security across broad spatial scales and sufficiently deep soil profiles. To this end, we conducted high-density *in situ* deep soil sampling (0–500 cm) using drilling machines across the YRB (795,000 km^2). This sampling campaign spatially encompassed diverse climatic, vegetative, edaphic, and topographic conditions, and vertically captured both the typical depth of rainfall infiltration and most of the rooting zone. To further address potential errors caused by spatial non-stationarity, our second objective was to design a research path that incorporates the *in situ* sampling, random forest (RF) modeling, and Hyper-local geographically weighted regression (GWR). This approach allows us to explore the environmental drivers of large-scale, deep-profile SM security while accounting for spatial non-stationarity of environmental variables and balancing study scale with cost. Based on these items, we hope to provide insights into deciphering the environmental control mechanisms of deep SM security in large-scale studies, thereby serving the broad needs for SM management.

2. Materials and methods

2.1. Soil sampling and data collection

The study area is the entire YRB (95°53′–119°5′E, 32°10′–42°50′N) with a total area of approximately 795,000 km^2 (Wang et al., 2020). A 100 km × 100 km sampling grid covering the YRB was constructed to evenly subdivide the entire basin. For each grid, we chose a sampling site (generally at the center of the grid) to represent the overall landscape of the grid. Both soil and water data were collected *in situ* during two field campaigns—22 September 2018 to 31 December 2018 and 8 October 2019 to 6 November 2019. Sampling was completed after the rainy seasons to avoid uneven rainfall event effects on SM, especially influencing the results of shallow SM. A total of 84 sites were sampled, encompassing all four different precipitation levels of the YRB (Figure 1).

A handheld vibrating drilling machine (CHPD78, Christie Engineering Pty Ltd, New South Wales, Australia) was used to collect soil cores to a maximum depth of 500 cm, which covered the typical rooting depth of the main vegetation types in the study area (Fang et al., 2016). The soil samples were sampled at an interval of 20 cm. Soil particle composition was measured using a Mastersizer 3000 laser particle size analyser (Malvern Panalytical, Malvern, UK). Prior to measurement, soil samples were passed through a 1 mm sieve and then soaked in purified water for 24 hours. Purified water was also used as the dispersant, and the samples were dispersed by ultrasonic vibration for 28 seconds. The optical parameters were set with a refractive index of 1.52 and an absorption value of 0.10. Volumetric SM content was measured *in situ* in the boreholes at 10 cm depth intervals using a neutron probe (CNC503, Beijing Chaoneng Technology Co. Ltd., Beijing, China). The probe directly outputs neutron counts and need to be converted to volumetric SM through a neutron count transform curve. Since curves from other regions may not be applicable to the YRB due to potential biases introduced by soil property differences, the existing transform curve of our probe was calibrated prior to field measurements. For calibration, soil samples were collected by excavating soil profiles at representative sites within the YRB. Their gravimetric SM and bulk density were firstly measured, and gravimetric SM was converted to volumetric SM based on bulk density. These converted volumetric SM values were then used to establish the locally valid transform curve relating neutron counts to volumetric SM. The R^2 of the calibrated curve for the YRB was 0.92 and its equation was:

$$Q = 60.31 \times (R/Rs) + 2.36 \quad (1)$$

where Q is volumetric SM, R is neutron count directly read from the probe, Rs is standard neutron count (829) of our probe. The gravimetric SM was only used for calibration, while all subsequent analyses in this study were conducted using volumetric SM data derived from the neutron probe based on the calibrated curve for the YRB.

The meteorological data were obtained from the Annual Dataset of China Surface Climate Data provided by the China National Meteorological Service Center. To minimize potential bias, we further applied the following measures: (1) any missing or inconsistent values in the annual dataset were infilled using the Monthly Datasets of China Surface Climate Data and Daily Datasets of China Surface Climate Data (V3.0), and (2) for each sampling site, the closest meteorological station was selected to represent local conditions. Seven meteorological factors (from 1982 to 2018) were selected for further analysis, including mean wind speed (WSmean), mean temperature (Tmean), mean minimum temperature (Tminmean), mean maximum temperature (Tmaxmean), mean annual precipitation (Precipitation), mean relative humidity (RHmean), and annual sunshine duration (H). Normalized difference vegetation index (NDVI) data was derived from the third-generation Global Inventory Monitoring and Modelling System (GIMMS) NDVI Dataset (3g.v1), with a time series of 1982 to 2015 and a spatial resolution of 1/12° (NCAR, 2018). China's Monthly NDVI Spatial Distribution Dataset NDVI data was obtained from the Vegetation (VGT) sensor mounted on the SPOT4 satellite, with the time series of 2016 to 2018 and a spatial resolution of 1 km (Xu, 2018). Monthly NDVI composites in each year were averaged to obtain the annual NDVI composites. Field capacity (FC) and permanent wilting point (PWP) data in the YRB were obtained from existing datasets (Dai et al., 2013), with a spatial resolution of 30 arc seconds and a depth of 0–138.30 cm. Over the whole basin, the FC varied from 12.80% to 42.23% with a mean of 25.94%, while the PWP varied from 3.40% to 28.82% with a mean of 12.30%.

2.2. Vertical stratification of SM profile

Due to the distinct hydrological processes in shallow and deep soil layers, the spatial distribution characteristics of SM at different depths, as well as their interaction mechanisms with environmental factors under long-term environmental stress, may also differ. Therefore, it is essential to first determine whether there is vertical stratification in SM and then conduct subsequent calculations and analysis within the identified characteristic layers. Agglomerative hierarchical clustering (group average) based on Euclidean distance was employed to delineate the vertical SM characteristic layers. To ensure the reliability of the results, clustering analysis was performed within five subdivided depth ranges (0–100 cm, 0–200 cm, 0–300 cm, 0–400 cm, and 0–500 cm), thereby forming five clustering groups (Figure S1). This design minimized potential bias from sample exclusion: conducting clustering only over the full 0–500 cm range would have required removing sites with missing samples at certain depths, thereby discarding valid SM information from other depths at those excluded sites. As shallow soil samples obtained through drilling were more intact than deeper samples, clustering groups with shallower depth ranges contained information from more sites. When discrepancies arose between the characteristic layers derived from different clustering groups, the results from shallower depth ranges were prioritized. The final vertical SM characteristic layers were then derived from a synthesis of the clustering results from these five clustering groups.

2.3. Spatial prediction of SM distribution

In order to obtain the spatial distribution of SM with higher resolution, the RF model was performed in all characteristic layers. The RF model is not sensitive to multiple collinearities, it is able to predict the effects of multiple explanatory variables and its results are relatively robust with unbalanced data (Breiman, 2001). It is worth noting that although the RF model is robust, it functions as a data-driven “black box” and does not explicitly represent the physical mechanisms through which environmental factors control SM, as process-based models do. In this study, a total of

14 factors influencing SM were pre-selected to construct the RF model, including factors related to soil texture (clay, silt, and sand), topography (elevation, slope, and aspect), meteorology data from 1982 to 2018 (WSmean, Tmean, Tminmean, Tmaxmean, Precipitation, RHmean, and H) and vegetation data from 1982 to 2018 (annual averaged NDVI). The RF model was implemented in R (version 3.6.3) using the "randomForest" package (Liaw and Wiener, 2002). In the RF model of each layer, 80% of the sampled site data (n = 68) was randomly selected and used as the training set, with the remaining 20% of the sampled site data (n = 16) used in the testing set. The sampling and modelling were repeated 1000 times as cross validation to reduce the impact of random partitioning and obtain stable estimates of model performance metrics. This procedure minimized bias caused by spatial clustering of training and testing samples and ensured robust evaluation of the RF model. The model assessment was then based on the full set of 1000 results. The R^2, root mean square error (RMSE), and mean absolute error (MAE) were used to evaluate the model accuracy, and the p-value (p) was used to evaluate the model reliability. The calculation methods were:

$$R^2 = \frac{\left[\sum_{i=1}^{N}(m_i - \bar{m})(\bar{s_i} - \bar{s})\right]^2}{\sum_{i=1}^{N}(m_i - \bar{m})^2 \sum_{i=1}^{N}(s_i - \bar{s})^2} \tag{2}$$

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(m_i - s_i)^2} \tag{3}$$

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|m_i - s_i| \tag{4}$$

where m_i and s_i are measured and predicted SM of the i sample, respectively; $\bar{m}$ and $\bar{s}$ are mean measured and predicted SM, respectively; and N is the sample number. The model was considered to be reliable when $p < 0.05$. Upon validating of the RF model's accuracy, the topographic, soil, and climate factors from all sites were interpolated into rasterized data using ordinary kriging. While these factors exhibit nonnegligible heterogeneity at regional scale, kriging remains a widely used and effective method for interpolating environmental variables with local spatial dependence, enabling reliable spatial predictions from site-based data (Liu et al., 2025; Zhou et al., 2019). These data, along with the vegetation data, was subsequently resampled to a resolution of 1 km. The resampled 14 variables were then fed into the RF model to calculate the spatial distribution of SM at a resolution of 1 km.

The relative SM (RSM) is a dimensionless indicator defined as the ratio of SM to FC. It characterizes the relative degree of soil saturation in field situation, and has been widely used to evaluate vegetation water availability, agricultural drought severity and provides an important reference for irrigation strategies (Gardin et al., 2021; Zhu et al., 2022). By normalizing SM to FC, RSM also enables direct comparison across soils with different backgrounds. Through introducing the PWP, the available SM storage (ASMS) quantifies the absolute volume of water that can be used by plants. Compared to soil moisture content alone, ASMS is also directly linked to vegetation water availability and provides a more practical basis for guiding vegetation restoration and regional water resource management (Zhao et al., 2021). To gain a preliminary understanding of the situation regarding deep SM resources in the YRB, the RSM and ASMS at diverse depths across the basin were calculated based on the estimated raster data as follows:

$$RSM = \frac{\theta_i}{\theta_{FC}} \tag{5}$$

$$ASMS = (\theta_i - \theta_{PWP}) \times d_i \tag{6}$$

where θ_i, θ_{FC}, θ_{PWP}, and d_i is the SM, FC, PWP, and soil depth of the *i* sample, respectively. To avoid potential bias in the assessment of shallow SM resulting from one-off sampling, subsequent analysis was focused on the relatively stable soil layers (110–500 cm) (Yang et al., 2012).

2.4. Identification of SM security zones

Seneviratne et al. (2010) defined three SM/climate regimes based on the influence of SM on evapotranspiration: (1) a dry regime, when SM < PWP, evapotranspiration is strongly moisture-limited, vegetation cannot utilize soil water, and the impact on climate is very weak; (2) a wet regime, when SM > (50–80%) * FC, evapotranspiration is primarily energy-limited and changes in SM exert little influence on local climate; and (3) a transitional regime, when PWP < SM < (50–80%) * FC, in which SM has a pronounced control on evapotranspiration and exerts strong land-atmosphere feedbacks, concurrently influencing vegetation growth and climate (Zhou et al., 2021). In addition, Wang et al. (2010) suggested that when SM falls below 60% of FC, dried soil layers (DSLs) may develop, leading to soil quality degradation and poor vegetation growth. Considering the heterogeneous soil-environment background of the YRB and such ecological thresholds, we adopted a stricter criterion by setting the boundary of the “wet” regime at 80% of FC to avoid overestimation of wet conditions.

Building upon these concepts, we extend the regime-based classification into a “soil moisture (SM) security” framework. In this framework, SM conditions are categorized into three security zones based on their potential impacts on evapotranspiration, vegetation growth and soil quality: (1) wet zone (SM ≥ 80% * FC), representing high-security states with minimal water limitation and high potential to meet future demands for vegetation restoration and irrigation; (2) transitional zone (PWP < SM < 80% * FC), representing medium-security states where SM strongly influences evapotranspiration and land-atmosphere coupling, thereby playing a key role in both vegetation growth and climate regulation; and (3) dry zone (SM ≤ PWP), representing low-security states where vegetation suffers severe water stress and soil quality degrades. This classification provides a monitorable baseline for evaluating SM changes and their ecological or hydrological implications. For clarity, the classification thresholds are summarized in Table 1. The FC and PWP data in diverse SM characteristic layers were calculated from the existing dataset (Dai et al., 2013) (Figures S2 and S3) and detailed calculation methods are shown in Table S1.

2.5. Evaluation of dominant factors affecting SM security

2.5.1. Pre-screening of environmental drivers of SM security

To eliminate the potential multicollinearity among the factors influencing SM, a principal component analysis (PCA) was used to preliminary screen the 14 factors described above in each SM characteristic layer. The principal components (PC) were determined according to the two following principles: (1) Select sufficient PCs so that an adequate percentage of the variation is explained (i.e., ≥ 70%); and (2) The eigenvalue λ corresponding to each PC is > 1 (Arias et al., 2018). Based on this, factors with large weight coefficients of each PC in the component matrix were used (i.e., factors with load coefficients greater than 90% of the maximum load coefficient of the PC) and then screened by PCA (Mandal et al., 2008).

2.5.2. Evaluating overall influence intensity of environmental factors on SM

security

Pearson correlation analysis quantified the basin-wide influence intensity of the 11 PCA-screened factors on SM. In each characteristic layer of 110–500 cm, correlation analysis was performed between the mean values of SM and 11 factors.

2.5.3. Evaluating localized influence intensity and scope of environmental factors on SM security

Unreasonable assumptions on spatial non-stationarity may triggered by whole map statistical models (Comber et al., 2018). Therefore, the GWR method was used to explore the spatial non-stationarity of regression relationships. In simple terms, "geographically weighting" is achieved by assigning weights based on the distance to the kernel center, such that data farther from the center contribute less to the overall model. During this, the key parameter called kernel bandwidth determines the spatial scale of each local model. In a standard GWR, kernel bandwidth and regression model selection could fail to recognize the importance of variables in different locations, making it challenging to reliably assess the localized influence of different factors (Chen et al., 2024). To better address this, "Hyper-local GWR" was used to enable local variable selection under local kernel bandwidths (Comber et al., 2018). Compared with the traditional correlation analysis and the GWR method, Hyper-local GWR allows the nature of the spatial relationship to be investigated in greater depth.

Specifically, for each sampling site, we tested a series of bandwidths from 200 km to the maximum longitudinal-latitudinal extent of the study area, in steps of 50 km. The minimum bandwidth (200 km) was chosen to ensure that at least two sampling sites were included in each local model, given the 100 km sampling resolution in this study. The 50 km step size was adopted to avoid overly coarse increases in bandwidth. At each bandwidth setting, a corresponding local regression model was constructed, and the stepwise Akaike information criterion (AIC) method was applied to evaluate model fit. The t-values were used to indicate which corresponding covariates had significant impact on SM (Figures S4–S10). This resulted in an AIC score for every possible local model at each location, from which the optimal model was selected based on the minimum AIC value. The covariates (i.e., influencing factors) in the GWR were the 11 factors pre-screened by PCA. The above processes were implemented in R (version 3.6.3). Following these steps, the Hyper-local GWR was able to identify the factors influencing SM across different regions and characteristic layers (both horizontal and vertical) of the YRB. The size of the standardized covariate coefficients reflected the influence intensity of environmental factors on SM. The total number that a specific covariate occurred in all regression formulas indicated the spatial scope of influence from this factor on SM. In a traditional GWR analysis, under-fitting of the true nonstationary relationship is not uncommon, whereas Hyper-local GWR may result in overfitting at some locations (Comber et al., 2018). For regressions not intended for predictive purposes, this is beneficial for identifying the most critical factors contributing to the spatial pattern of SM security at each location. However, it also carries the risk of overestimating the importance of certain factors. Therefore, the identification of dominant factors through Hyper-local GWR was augmented with partial correlation analysis, which also served to assess the accuracy of the dominant factors selected by Hyper-local GWR and to detect any remaining synergistic effects among these factors. The core code for understanding and applying the algorithm of Hyper-local GWR can be found in the Loess Science Data Center (Tong and Wang, 2025). The overall methodological framework is summarized in Figure 2.

3. Results

3.1. Overview of SM, environmental factors, and predicted SM distribution

The character of SM was firstly evaluated. Across all layers from 0 to 500 cm, SM ranged from 2.94% to 50.95%, with most observations concentrated between 15% and 25%, and an overall mean of 19.08%. The mean SM across sites showed a decline from the surface to about 100 cm, followed by irregular fluctuations at greater depths (Figure 3a, b). We then assessed the horizontal coefficient of variation (CV) of SM across all sites within each layer (Figure 3b) and the spatial distribution of vertical CV at each site (Figure 3c). The mean horizontal CV (45.74%) exceeded the mean vertical CV (20.85%), indicating greater horizontal than vertical variability of SM across the YRB. The wide range of SM values at several depths in Figure 3a further reflects the nonnegligible inter-site spatial heterogeneity, which has also been reported in other large-scale studies (Wang et al., 2012). Environmental factors (soil texture, meteorological, and vegetation factors) also exhibit obvious spatial heterogeneity, both vertically and horizontally (Figure S11 and Figure S12). Comparing to the spatial heterogeneity of environmental factors, most of the interannual variability of meteorological factors are characterized by weak to mild variation (Figure S12 and Table S2), indicating a relatively stable long-term climate across the basin (Nielsen and

Bouma, 1985).

Considering the spatial heterogeneity of SM and associated environmental factors, we explored the potential stratified characteristics of SM before further analysis. Within the 0–100 cm interval, discrepancies were observed between the 0–100 cm and 0–200 cm clustering groups and the other three groups. According to the stratification rule, the results from the 0–100 cm group were adopted, yielding characteristic layers at 10 cm and 30 cm. Similarly, the characteristic layers in deeper intervals were determined from the corresponding clustering groups: in the 100–200 cm interval, the characteristic layer was at 110 cm (from the 0–200 cm group); in the 200–300 cm interval, at 200 cm (from the 0–300 cm group); in the 300–400 cm interval, at 330 cm (from the 0–400 cm group); in the 400–500 cm interval, at 420 cm (from the 0–500 cm group). The vertical depth of SM characteristic layers in 0–500 cm profile of the YRB were finally determined to be 10 cm, 30 cm, 110 cm, 200 cm, 330 cm, and 420 cm (Figure S13).

Subsequently, the RF model was used to obtain higher spatial resolution of SM distribution, and the simulation accuracy of RF was acceptable at the YRB scale (Table 2). The RF model generated statistically significant predictions with almost all $p < 0.05$. The relatively small MAE and RMSE values indicated that the model was able to reproduce the magnitude and broad spatial distribution of SM across the YRB. However, the explained variance remained modest with R^2 ranging from 0.39 to 0.57, mainly due to fine-scale heterogeneity driven by factors such as soil structure, microclimatic variations, and human activities that were not fully captured by the available predictors. It should also be acknowledged that practical constraints associated with deep drilling still exist. To obtain soil parameters and moisture data down to 5 m, we used *in situ* drilling, which entailed high field costs. Despite applying a grid-based sampling design to ensure spatial representativeness, the number of sites was inevitably limited, thereby constraining further improvements in the variance explained by the RF model. Future work could enhance predictive performance by expanding the borehole network, maintaining regular observations, incorporating additional local-scale predictors, and coupling data-driven with process-based models to better represent SM variability.

Although meteorological conditions in the YRB are spatially heterogeneous, their long-term interannual variability is relatively low (Table S2), suggesting a stable climatic background that supports the feasibility of using one-off sampling to characterize general SM patterns. Nevertheless, previous studies have shown that SM within 0–100 cm layer exhibits high temporal variability and generally considered as an active layer due to frequent disturbances from climatic fluctuations, vegetation dynamics, and human activities (Cao et al., 2018). In contrast, deep SM (>100 cm) is relatively buffered against short-term fluctuations and less directly affected by surface disturbances (Gou and Zhu, 2021; Li et al., 2021; Liu and Shao, 2016; Yang et al., 2012). Besides, the time-lagged nature of infiltration further reduces the influence of occasional rainfall events prior to sampling on deep SM. Therefore, to mitigate potential biases from one-off measurement and yield more reliable insights, we focused on the relatively stable layers (110–500 cm) in the following analyses.

Based on SM distributions from RF (Figure S14), RSM and ASMS were calculated for each sub-layer (Figure 4). The mean RSM and ASMS in 110–500 cm of the entire YRB were 0.84 and 32.14 cm, respectively. Horizontally, the spatial distribution of RSM and ASWS exhibited obvious regional differences, without continuous uniform zones. Vertically, noticeable differences in RSM also existed across diverse characteristic layers at the same location. Both RSM and ASMS demonstrated an obvious spatial heterogeneity. Considering that RSM and ASWS were computed from SM, FC, and PWP, simultaneously reflecting SM and its security classification thresholds, this indicates the potential heterogeneity in the spatial patterns of SM security across different characteristic layers. Therefore, stratified identification of the SM security patterns is necessary.

3.2. Classification of SM security zones

The average proportions of the wet, transitional, and dry zones in the whole relatively stable layers (110–500 cm) were 20.24%, 61.58% and 18.17%, respectively (Figure 5). As the depth increased from 110 to 500 cm, the proportion of the wet zone increased from 17.04% to 25.86%, the transitional and dry zone decreased from 62.21% to 59.15%, and 20.76% to 14.99%, respectively. In the relatively stable layers in the YRB, deep SM showed a higher availability than shallow SM.

Mapping of the spatial distribution of SM security is required before taking action to regulate SM availability at regional scale. Considering the high cost of demarcating SM security zones based on *in situ* measured SM data, we reported the ranges of environmental factors within each security zone as a guideline to estimate the spatial demarcation of these security zones based on existing geodata representing these factors (Figure S15). The interquartile range of environmental factors in different zones showed that slope is the best predictor to delineate the wet zone. Subsequently, NDVI, Precipitation, WSmean, and temperature, factors could be used to further distinguish the transitional and dry zones. However, it should be noted that the interquartile of some environmental factors in transitional and dry zones still overlap over a considerable value range. It remains thus challenging to delineate all three SM security zones with no use of SM data at all. While the availability of these explorations is limited, certain indicators such as slope can find utility in specific scenarios, especially in assisting to reduce the cost of large-scale SM security classification.

3.3. Considering localized effects in regional-scale dominant factors identification

The factors that shape spatial pattern of SM security at the basin-wide scale were first quantified through the overall-perspective analysis (Figure 6). In terms of correlation coefficient (Figure 6a), the interaction between soil- and temperature-related factors and SM exhibited an opposite trend with depth, with a decreasing influence of soil texture and an increasing influence of temperature. Slope and NDVI significantly affected the SM within the soil layers from 110–500 cm. When localized effects are not considered, slope, NDVI, clay, WSmean, and sand were identified as the most important factors affecting deep SM at the entire basin scale (Figure 6b).

Considering the obvious spatial heterogeneity in SM and environmental factors, hyper-local GWR models were then established for each characteristic layer to capture both overall and localized perspectives and all models were acceptable ($R^2 > 0.9$). Comparing the size of standardized regression coefficients in the Hyper-local GWR, the dominant factors of SM and the proportion these factors can explain within the 110–

500 cm layers were, in order, slope (31.07%), NDVI (25.60%), clay (18.17%), and Precipitation (9.32%) (Figure 7a). The Hyper-local GWR also suggested that Tmean, clay, Tmaxmean, sand, NDVI, and Tminmean were main factors that exist a larger spatial range of influence on the SM within the 110–500 cm layers (Figure 7b). Variations between the influence intensity and scope of different factors further reflected the complexity of the factors affecting large-scale SM security. This analysis suggested that the importance of meteorological factors in terms of the scope of influence was greater than that in terms of the intensity of influence. Vegetation and soil factors were of equally high importance in the scope of influence, while terrain factor effect was more localized.

Both the correlation analysis and Hyper-local GWR analysis indicated that slope, NDVI, and clay were the most important factors affecting SM throughout the entire YRB. However, the two methods varied in the order of importance of other main dominant factors (Figures 6b and 7a). The discrepancies between the two methods are related to Hyper-local GWR considered the local effect at each sample point. It uses a local kernel bandwidth to identify specific dominant factors at each sample location. This directly demonstrates the necessity to conduct a consistent localized analysis across the YRB. Therefore, factors that dominate the spatial pattern of SM security within each of the three security zones were further analyzed. (Figure 8). The dominant factors and the proportion explained by these factors both changed after considering specific security zones. In the wet zone, the dominant factors and their contribution rates were slope (29.33%), clay (15.04%), NDVI (14.50%), and Tmaxmean (14.09%).

In the transitional zone, the dominant factors and contribution rates were slope (27.48%), NDVI (21.32%), and clay (12.81%), while it became NDVI (26.37%), clay (18.26%), slope (16.01%) and RHmean (9.10%) in the dry zone. In attributing the environmental drivers of SM security patterns across the entire basin and within each zones using Hyper-local GWR, the required data were spatially discrete, with both SM and environmental factors obtained directly from observations, while the RF-predicted SM distribution was used solely to identify SM security zones. This approach minimized the potential biases from predicted SM and ensured the reliability of the Hyper-local GWR results.

After accounting for localized effects, the partial correlation analysis ruled out potential synergistic effects and further confirmed the contributions of certain dominant factors to SM security patterns (Figure 9). In the wet zone, the influence of clay, NDVI, and Tmaxmean were no longer significant, while slope remained the dominant factors. In the transitional zone, slope, NDVI, and clay remained the dominant factors, but the importance ranking suggested an order of NDVI, clay and slope. In the dry zone, SM security remained most influenced by NDVI, while clay, slope and RHmean were no longer significant. Finally, slope and vegetation were identified as the dominant factors for classifying the wet and dry zones, respectively. The classification of the transitional zone was controlled by all three factors with a total contribution rate of 61.60%.

4. Discussion

4.1. Environmental control mechanism of spatial patterns of SM security

Identifying the associations between environmental factors and SM security spatial patterns helped uncover the mechanisms of environmental factors shape these patterns and the potential coupling effects among these factors. The influence of topographic factors, particularly slope, on SM security patterns is primarily linked to the infiltration. The source of deep SM mainly relies on recharge from precipitation and long-term infiltration, yet steep slopes accelerate runoff generation and shorten the contact time of water and soil, which limits infiltration and reduces water replenishment to deep layers (Huang et al., 2013; Morbidelli et al., 2015). This leads to spatial contrasts in SM recharge capacity under different slope conditions. Our analysis of average slope values for grids corresponding to wet, transitional, and dry zones revealed mean slopes of 2.54°, 7.03°, and 7.30°, respectively, indicating that areas with steeper slopes tend to have lower SM security. This finding underscores the impact of slope on SM security patterns. A regional-scale study conducted in Utah, USA also corroborated this (Wu et al., 2021). Steep slopes also intensify soil erosion, causing the loss of fine particles and organic matter, which reduces soil water holding capacity and affect vegetation growing. The resultant decrease in canopy interception and sparse vegetation may further limit infiltration and promote runoff, forming a feedback that ultimately restricts deep SM recharge (Sun et al., 2014). In addition, steeper slopes can amplify aspect-induced contrasts in solar radiation and microclimate, leading to stronger differences in surface temperature, vapor pressure deficit, and evapotranspiration (Bennie et al., 2008; Yin et al., 2023), thereby limiting SM accumulation (Qiu et al., 2001).

Soil texture forms a fundamental descriptor of natural soils and significantly influences soil structure as well as water movement and storage (Sun et al., 2020). In the YRB, clay was identified as the most important soil texture factor influencing SM security pattern with a positive effect. High clay content increases the proportion of fine particles and specific surface area, thereby enhancing capillary forces and matric potential and enabling soils to retain more water (Lehmann et al., 2021; Tuller and Or, 2005). It also promotes the formation of soil aggregates and compact pore structure that lower saturated hydraulic conductivity and slow water percolation, thereby prolonging water residence in the soil profile (Bargués-Tobella et al., 2024; Krause et al., 2018). Consistent with these mechanisms, Chen et al. (2020) evaluated the deep SM situation after long-term environmental processes in a 57 m-deep profile in the YRB and found soil layers with higher clay content generally stored more SM. At the global scale, Wankmüller et al. (2024) reported that clay-rich soils delay the onset of water stress in ecosystems during drought compared with coarse-textured soils, thereby providing higher potential for vegetation to adjust to water limitations. These findings also provide direct evidence for the positive role of clay in influencing SM security. It is noteworthy that, beyond the soil texture factors included in the direct attribution analysis, soil properties such as hydraulic thresholds represented by the FC and PWP are also fundamental in shaping SM security. These thresholds determine the bounds of plant-available water, thereby regulating SM storage capacity and availability and influencing vegetation-associated evapotranspiration. Comparison between SM and RSM at the same depths and locations showed that RSM exhibited a more discontinuous spatial distribution across most positions (Figures 4 and S14). This pronounced spatial heterogeneity suggests that even under similar environmental conditions, soils with different FC-PWP combinations can manifest distinct SM security states. This is also why FC and PWP were integrated as key thresholds into our SM security framework.

Vegetation significantly alters terrestrial water fluxes. Trees could increase soil porosity and organic carbon content, in turn increasing soil infiltration and water storage capacity, especially in previously treeless areas (Bargués-Tobella et al., 2014; Ellison et al., 2017; Hoek van Dijke et al., 2022). In addition, vegetation provides an important link with land-atmosphere interactions and they not only directly affect SM through root water uptake, but also indirectly regulates SM via atmospheric moisture recycling (Hoek van Dijke et al., 2022) from local to continental scales (Lawrence and Vandecar, 2015; Meier et al., 2021). The positive correlation coefficient and standardized Hyper-local GWR regression coefficient between NDVI and SM were associated with the use of multi-year averaged climatological NDVI and SM data. While this study focuses on the environmental control mechanisms of SM spatial patterns rather than temporal dynamics, it does not determine whether this relationship is positive or negative over time. However, regional studies of deep SM in the YRB have identified that vegetation can deplete SM in the deep layer below 100 cm (Qiu et al., 2021). Wang et al. (2024b) provided direct evidence from 37 years of continuous observations in the Loess Plateau (LP), which covers approximately 640,000 km^2 and constituting the main body of the YRB, demonstrating a significant decline in SM storage in profiles as deep as 10 m, primarily due to vegetation changes rather than long-term climate variability.

It is important to note that the aforementioned dominant factors do not independently influence the SM security spatial patterns but can also have synergistic effects through multi-factor coupling. For instance, slope can directly affect precipitation's impact by regulating water infiltration: the finite depth of water infiltration in soil may be further limited by steeper slope and leads to a diminished direct influence of precipitation on SM in deep layers (Guo, 2020; Wang et al., 2008). The effect of slope is more significant in wet and transitional zone with better SM supply conditions further supporting this. Additionally, topographic factors such as slope can modulate climate wetness, thereby weakening or enhancing the periodicity of synchronized fluctuations between rainfall and SM (Li et al., 2020; Wu et al., 2021). Regarding vegetation, other factors may also work synergistically with vegetation to influence SM security patterns. In deep soil layers, the influence of temperature on SM may be closely related to vegetation. This is because high temperature may increase the vapor pressure deficit, and the generation of precipitation becomes uncertain, but the water consumption by transpiration increased, leading to a reduction in the SM (Yang et al., 2012). The influence of wind speed on deep SM may also connect with root water consumption, as winds reduce surface temperature, helping to limit deep water consumption via evapotranspiration (Will et al., 2013).

Moreover, due to spatial heterogeneity in specific locations, the influence of some environmental factors may be obscured when multiple factors jointly influence SM. For example, a comparison of Figure 1 and Figures 5d–g shows that precipitation within main area of the wet zone mostly falls in the similar threshold range, while Figure S11a shows that in the same region, soil, particularly clay content, exhibits exceptional horizontal and vertical spatial variability. The spatial variability of soil partially obscures the impact of climate forcing within the wet zone. Similar phenomena have been observed in other regional studies, where increasing spatial variability in climate factors amplifies their influence while correspondingly weakening the impact of soil factors (Wang et al., 2017a; Wu et al., 2020). This synergistic mechanism among multiple factors, along with its dependence on the location and scale of the study area, underscores the need for large-scale studies that encompass diverse environmental backgrounds.

Currently, studies suggested the influence of local factors such as topography, soil, and vegetation on SM is primarily present at small scales (Seneviratne et al., 2010; Vereecken et al., 2014), we also confirmed the significance of these factors in large-scale study. Several factors may explain this seemingly divergent conclusion: firstly, the spatial variability of these factors within the scales of existing studies may be relatively small, insufficient to dominate (Wang et al., 2017a). Secondly, in previous studies, the lack of detailed soil information based on field measurements limits a comprehensive assessment of local factors' influence on soil hydrological processes (Wang et al., 2017a). Besides, by employing hyper-local GWR, this study accounted for the spatial non-stationarity in the interaction mechanisms between SM spatial patterns and environmental factors, which helps to reveal the role of local factors on a larger scale (Comber et al., 2018). Recent studies have increasingly found that the influence of local factors on regional SM spatial patterns may be stronger than previously thought (Dong and Ochsner, 2018; Wang and Franz, 2015; Wang et al., 2017b; Wu et al., 2020), while the impact of meteorological forcing at the regional scale may be weaker than expected (Wang et al., 2017a). Although a particularly prominent influence of meteorological factors was not found this study, as suggested by Seneviratne et al. (2010), the impact of meteorological forcing on SM security may become more significant in regions with greater climate gradient variation or more uniform soil texture. Therefore, in studies involving the temporal dynamics of SM security, it is crucial to comparatively analyze the relative roles of climate and other time-varying factors alongside soil, topography, and other time-invariant factors.

4.2. Implications for SM management

This study identifies the SM security at different layers separately, aiming to provide more comprehensive insights into the SM management across the entire profile. SM in the wet zone had the highest utilization potential. The utilization of SM and the maintenance of water availability in this region should focus on ecosystem impact. Considering that the border of the transitional zone was close to the border of the dried soil layers (DSLs), the exploitation of SM in this area (especially in deep layers) should tailored to avoid the formation of DSLs, resulting further soil quality degradation and poor vegetation growth (Wang et al., 2010). The impacts of this process on both ecosystem and climate should also be considered when regulating SM availability in the transitional zone. The SM in the dry zone was unavailable to vegetation and it is important to implement water conservation rather than extract in these zones.

The SM management measures proposed in this study were derived based on the attribution analysis results. Using Hyper-local GWR, we identified the dominant factors shaping the SM-security patterns within the wet, transitional, and dry zones and then focused on those factors that are practically controllable by human interventions. The results indicated that slope, vegetation and soil are all controllable factors, although with different associated costs. Changing these controllable factors provides a direct way for targeted SM management measures for each zone. When considering management of SM, measures should not be considered in isolation at a specific layer. Instead, the flow and replenishment of moisture between different layers within the entire profile should be taken into account, and combined with the depth range affected by different measures, to formulate appropriate strategies.

Slopes can be adjusted by terrain engineering, such as gully land consolidation (GLC), which has been implemented in a series of watersheds in the YRB. Studies in the wet and transitional zones of the YRB have showed that GLC increases the deep SM and shortens the hysteresis for precipitation to recharge the deep SM (Zhao et al., 2022). A large-scale survey also demonstrated the increase in NDVI after GLC (He et al., 2020). The absolute value of the regression coefficient of slope in the wet and transitional zone indicates that implementing GLC in these zones would have a more obvious impact on SM security. However, it should be noted that in relatively flat regions or areas with minimal topographic variation, slope may play a less significant role, and slope-based interventions such as GLC may have limited applicability.

A significant relationship between deep SM security spatial pattern and vegetation was identified in this study. Many researches also confirm the profound impact of vegetation restoration on ecosystem services (Griscom et al., 2017; Wang et al., 2024a). Therefore, vegetation management is a considerable way to adjust SM effects on climate and ecosystem security. At present, vegetation restoration has occurred at the large scale all over the world, especially in northern high latitude areas, and the climatic and ecological impacts of vegetation restoration has attracted widespread attention (Piao et al., 2020). Notably, vegetation restoration (including afforestation) may have bidirectional effects on SM (Hoek van Dijke et al., 2022; Li et al., 2018; Wang et al., 2024a; Zuo et al., 2024). When considering enhancing SM security through vegetation management in the dry zone, shallow-rooted vegetation (such as grass) may prevent water consumption in the deep layers, increasing the SM conservation capacity in the surface layer (Huang et al., 2019; Li et al., 2021). In the transitional zone, selective vegetation restoration is a viable option to regulate the SM availability. In areas with well SM recharge, trees with shallow roots should be planted as the water consumed by trees can be replenished by precipitation, and the water storage and water holding capacity of soil can be enhanced by vegetation restoration. In areas with poor SM recharge, planting shallow-rooted shrubs and grasses should be implemented. In addition to vegetation restoration, rational thinning and adjusting the vegetation species structure are alternatives to regulate SM availability, especially in dry and transitional zones. It may be feasible to alter the vegetation structure by shifting deep-rooted to shallow-rooted species. Research in the LP area has shown that converting artificial forests to grass can alleviate deep soil drought (Bai et al., 2021).

For the clay factor, it influences SM security by enhancing the SM holding capacity. Considering the economy cost and limited impact area of directly altering soil texture, SM availability can be regulated through the similar idea of enhancing soil water-holding capacity by adding soil amelioration such as Biochar (Wei et al., 2023). Furthermore, we suggest that soil amelioration may be limited to areas with economic returns, such as farmland and orchards in transitional zones.

Based on these controllable factors, three main SM management measures are suggested. Changing terrain (slope) conditions through engineering activities could be applied in wet and transitional zones. In transitional and dry zones, SM conditions could be managed by adjusting vegetation configuration, such as implementing selective restoration, rational thinning, or shifting species composition. Additionally, in transitional zones where agricultural or economic returns are expected, soil amelioration (e.g., adding biochar) could help increase soil water-holding capacity.

4.3. Limitations

This study examined both the horizontal and vertical dimensions, and identified the dominant factors of large-scale deep SM security patterns by integrating both regional and local perspectives. However, several limitations should be acknowledged due to time and cost constraints. First, the large-scale sampling of deep soil data was based on a one-off sampling approach rather than repeated sampling. Although main analyses were focused on the relatively stable deep layers (110–500 cm) to mitigate potential biases, it should be acknowledged that deep SM is not entirely static and may shift over longer time periods due to deep-root plant growth, land use changes, and human activities, particularly. This limitation restricts the comprehensive understanding of SM security over long-term scales. Future work may incorporate time-series SM data obtained through continuous *in situ* measurement, satellite monitoring, and model simulation approaches. Such data would help better capture temporal dynamics and support the evaluation of long-term SM security in both shallow and deep soil profile. Second, uncertainties may exist in the ancillary datasets. For meteorological data, although the nearest station was selected for each sampling site to reduce spatial bias, potential uncertainties remain. These include possible interpolation errors in areas with sparse station coverage, as well as local-scale climatic variability that may not be captured by station-based data. In addition, the limited depth of the FC and PWP data obtained from existing datasets may overlook the reduced FC and increased PWP caused by soil compaction in deeper layers, thereby resulting in a potential underestimation of the wet zone and overestimation of the dry zone. Third, although the sampling sites were evenly distributed across the entire YRB to ensure spatial representativeness and the RF model produced statistically significant predictions with small MAE and RMSE, its ability to explain the variance of SM remained modest. This limitation likely reflects the strong fine-scale heterogeneity of SM driven by factors such as soil structure, microclimatic variations, and human disturbances that were not fully captured by the available predictors. Moreover, the number of deep soil observations was inevitably constrained by the high cost of *in situ* drilling, which restricted the potential to further improve the model’s generalization ability. These constraints highlight that, while the current dataset provides valuable insights into basin-scale deep SM patterns, the RF model may still be limited in capturing fine-scale variability. Future research could address these limitations by expanding the borehole network, maintaining long-term repeated observations, incorporating more fine-scale environmental predictors, coupling data-driven models with process-based models to better represent SM dynamics, and adopting transfer learning approaches to leverage external datasets from broader regions as source-domain training samples to enhance model accuracy and generalizability while reducing field costs. Finally, given the SM flow and replenishment between different layers, management measures for SM availability should be considered in the entire soil profile rather than a specific layer. To address these limitations, process-based models that couple the compensating effects of water uptake and dynamic vegetation responses could be used. Despite these limitations, we believe the integrated approach and findings of this study provide a basis for characterizing deep SM security pattern and its driving factors, supporting SM management, and enhancing ecosystem stability and sustainability in the YRB and similar regions.

5. Conclusions

By integrating *in situ* sampling, RF modeling, and the Hyper-local GWR, this study introduced a novel integration to investigate the large-scale, deep-profile spatial pattern of SM security and its environmental drivers. This integrative approach enabled a more nuanced understanding of deep SM security with improved spatial resolution, better accounted for the spatial non-stationarity of environmental factors, and maintained a balance between study scale and cost. The results demonstrated the vertical stratification characteristics of SM at the regional scale and revealed seven vertical characteristic layers (0–10 cm, 10–30 cm, 30–110 cm, 110–200 cm, 200–330 cm, 330–420 cm, and 420–500 cm) based on SM content. In accordance with the varying impact of SM on ecology and climate, SM security was categorized into wet, transitional, and dry zones within these seven characteristic layers. In relatively stable layers (110–500 cm), the transitional zone accounted for the largest average area proportion (61.58%), followed by the wet (20.24%) and dry zone (18.17%). The mean relative SM and available SM storage were 0.84 and 32.14 cm, respectively. As the depth increased in relative stable layers, the proportion of the wet zone area increased from 17.04% to 25.86%, the transitional and dry zone area decreased from 62.21% to 59.15%, and 20.76% to 14.99%, respectively. This highlights a higher SM availability of deep soil layers than shallow layers. Furthermore, our statistical analysis of the environmental factors within three SM security zones underscores the challenges of identifying SM security zones solely based on environmental factors without relying on SM values. Through Hyper-local GWR analysis that considers localized effects, we further revealed the dominant factors shaping SM security spatial patterns in large scale: slope in the wet zone, vegetation in the dry zone, while the transitional zone was dominated by clay content, vegetation, and slope. Besides, notable coupling effects exist among these factors. In wet and transitional zones, terrain (slope) can be adjusted by ecological engineering activities to regulate SM availability. In transitional and dry zones, vegetation management including restoration, thinning, and adjusting planting structure, can regulate SM availability. In the transitional zone, soil amelioration is also a feasible approach in areas with economic benefits. These findings will provide insights for characterizing spatial patterns of deep SM security and SM management in the YRB and similar regions worldwide.

Declaration of competing interest

The authors declare that they have no competing interests. This includes, but is not limited to, financial interests, non-financial interests, patents, or any other interests that may be perceived as influencing the research or its presentation. The corresponding author affirms that this information has been discussed and agreed upon by all authors listed.

Author contributions

Yongping Tong: Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing – original draft, Writing – review &editing. **Yunqiang Wang:** Conceptualization, Project administration, Supervision, Writing – review & editing. **Zimin Li:** Writing – review & editing. **Pingping Zhang:** Writing – review & editing. **Wei Hu:** Writing – review & editing. **Jingxiong Zhou:** Investigation. **Xiangyu Guo:** Investigation. **Hui Sun:** Data curation. **Alexis Comber:** Methodology, Writing – review & editing. **Ronny Lauerwald:** Conceptualization, Supervision, Writing – review & editing.

Acknowledgments

This research was supported by National Key Research and Development Program of China (No. 2019YFA0607303), National Natural Science Foundation of China (Nos. 41977422 and 42177306), and Innovation Cross Team-Key Laboratory project of the Chinese Academy of Sciences. We thank Professor Piao for his valuable assistance with this manuscript’s scientific aspects. We also appreciate the reviewers for their insightful suggestions on improving the quality of the manuscript.

References

Arias, O.V., Garrido, A., Villeta, M., Tarquis, A.M., 2018. Homogenisation of a soil properties map by principal component analysis to define index agricultural insurance policies. Geoderma 311, 149-158.

Bai, X., Jia, X., Zhao, C., Shao, M.a., 2021. Artificial forest conversion into grassland alleviates deep-soil desiccation in typical grass zone on China's Loess Plateau: Regional modeling. Agric. Ecosyst. Environ. 320, 107608.

Baldwin, D. et al., 2017. Predicting root zone soil moisture with soil properties and satellite near-surface moisture data across the conterminous United States. J. Hydrol. 546, 393-404.

Bargués-Tobella, A. et al., 2014. The effect of trees on preferential flow and soil infiltrability in an agroforestry parkland in semiarid Burkina Faso. Water Resour. Res. 50, 3342-3354.

Bargués-Tobella, A., Winowiecki, L.A., Sheil, D., Vågen, T.-G., 2024. Determinants of Field-Saturated Soil Hydraulic Conductivity Across Sub-Saharan Africa: Texture and Beyond. Water Resour. Res. 60, e2023WR035510.

Bennie, J., Huntley, B., Wiltshire, A., Hill, M.O., Baxter, R., 2008. Slope, aspect and climate: Spatially explicit and implicit models of topographic microclimate in chalk grassland. Ecol. Model. 216, 47-59.

Breiman, L., 2001. Statistical Modeling: The Two Cultures. Statist. Sci. 16, 199-231.

Brunsdon, C., Fotheringham, A.S., Charlton, M.E., 1996. Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity. Geogr. Anal. 28, 281-298.

Cao, R. et al., 2018. Deep soil water storage varies with vegetation type and rainfall amount in the Loess Plateau of China. Sci. Rep. 8, 12346.

Cevik Degerli, B., Cetin, M., 2023. Evaluation of UTFVI index effect on climate change in terms of urbanization. Environ. Sci. Pollut. Res. 30, 75273-75280.

Chen, C. et al., 2019. China and India lead in greening of the world through land-use management. Nat. Sustain. 2, 122-129.

Chen, F., Leung, Y., Wang, Q., Zhou, Y., 2024. Spatial non-stationarity test of regression relationships in the multiscale geographically weighted regression model. Spat. Stati.-Neth. 62, 100846.

Chen, Y., Feng, X., Fu, B., 2021. An improved global remote-sensing-based surface soil moisture (RSSSM) dataset covering 2003–2018. Earth Syst. Sci. Data 13, 1-31.

Chen, Y., Qian, H., Hou, K., Zhang, Q., Zhang, Y., 2020. Vertical distribution characteristics of soil moisture with different strata in deep profile in Guanzhong Basin, China. Environ. Earth Sci. 79, 103.

Comber, A., Wang, Y., Lü, Y., Zhang, X., Harris, P., 2018. Hyper-local geographically weighted regression: extending GWR through local model selection and local bandwidth optimization. J. Spat. Inf. Sci. 17, 63-84.

Dai, Y. et al., 2013. Development of a China Dataset of Soil Hydraulic Parameters Using Pedotransfer Functions for Land Surface Modeling. J. Hydrometeorol. 14, 869-887.

Das, N.N. et al., 2008. Modeling and assimilation of root zone soil moisture using remote sensing observations in Walnut Gulch Watershed during SMEX04. Remote Sens. Environ. 112, 415-429.

Dong, J., Ochsner, T.E., 2018. Soil Texture Often Exerts a Stronger Influence Than Precipitation on Mesoscale Soil Moisture Patterns. Water Resour. Res. 54, 2199-2211.

Ellison, D. et al., 2017. Trees, forests and water: Cool insights for a hot world. Glob. Environ. Change 43, 51-61.

Fang, X. et al., 2016. Variations of deep soil moisture under different vegetation types and influencing factors in a watershed of the Loess Plateau, China. Hydrol. Earth. Syst. Sci. 20, 3309-3323.

Fu, Z. et al., 2022. Atmospheric dryness reduces photosynthesis along a large range of soil water deficits. Nat. Commun. 13, 989.

Gardin, L. et al., 2021. Simulation of soil water content through the combination of meteorological and satellite data. Geoderma 393, 115003.

Ge, J., Pitman, A.J., Guo, W., Zan, B., Fu, C., 2020. Impact of revegetation of the Loess Plateau of China on the regional growing season water balance. Hydrol. Earth Syst. Sci. 24, 515-533.

Gou, Q., Zhu, Q., 2021. Response of deep soil moisture to different vegetation types in the Loess Plateau of northern Shannxi, China. Sci. Rep. 11, 15098.

Green, J.K. et al., 2019. Large influence of soil moisture on long-term terrestrial carbon uptake. Nature 565, 476-479.

Griscom, B.W. et al., 2017. Natural climate solutions. Proc. Natl. Acad. Sci. 114, 11645-

11650.

Gruber, A., Crow, W.T., Dorigo, W.A., 2018. Assimilation of Spatially Sparse In Situ Soil Moisture Networks into a Continuous Model Domain. Water Resour. Res. 54, 1353-1367.

Gruber, A., Dorigo, W.A., Zwieback, S., Xaver, A., Wagner, W., 2013. Characterizing Coarse-Scale Representativeness of in situ Soil Moisture Measurements from the International Soil Moisture Network. Vadose Zone J. 12, vzj2012.0170.

Guo, L. et al., 2022. Spatial-Temporal Variation Characteristics and Influencing Factors of Soil Moisture in the Yellow River Basin Using ESA CCI SM Products. Atmosphere 13, 962.

Guo, Z., 2020. Estimating Method of Maximum Infiltration Depth and Soil Water Supply. Sci. Rep. 10, 9726.

He, M. et al., 2020. Evaluation of the environmental effects of intensive land consolidation: A field-based case study of the Chinese Loess Plateau. Land Use Policy 94, 104523.

Hirschi, M., Mueller, B., Dorigo, W., Seneviratne, S.I., 2014. Using remotely sensed soil moisture for land–atmosphere coupling diagnostics: The role of surface vs. root-zone soil moisture variability. Remote Sens. Environ. 154, 246-252.

Hoek van Dijke, A.J. et al., 2022. Shifts in regional water availability due to global tree restoration. Nat. Geosci. 15, 363-368.

Hu, L. et al., 2025. Mapping the primary factors driving spatiotemporal variations of surface soil moisture in the Yellow River Basin of China. Sci. Rep. 15, 12359.

Huang, J., Wu, P., Zhao, X., 2013. Effects of rainfall intensity, underlying surface and slope gradient on soil infiltration under simulated rainfall experiments. Catena 104, 93-102.

Huang, L., Shao, M.a., 2019. Advances and perspectives on soil water research in China's Loess Plateau. Earth Sci. Rev. 199, 102962.

Huang, Z. et al., 2019. Natural grasslands maintain soil water sustainability better than planted grasslands in arid areas. Agric. Ecosyst. Environ. 286, 106683.

Jia, X., Shao, M.a., Zhu, Y., Luo, Y., 2017. Soil moisture decline due to afforestation across the Loess Plateau, China. J. Hydrol. 546, 113-122.

Koster, R.D. et al., 2004. Regions of Strong Coupling Between Soil Moisture and Precipitation. Science 305, 1138-1140.

Krause, L. et al., 2018. Microaggregate stability and storage of organic carbon is affected by clay content in arable Luvisols. Soil Till. Res. 182, 123-129.

Kucukpehlivan, T. et al., 2023. Determination of the impacts of urban-planning of the urban land area using GIS hotspot analysis. Comput. Electron. Agric. 210, 107935.

Lawrence, D., Vandecar, K., 2015. Effects of tropical deforestation on climate and agriculture. Nat. Clim. Change 5, 27-36.

Lehmann, P. et al., 2021. Clays Are Not Created Equal: How Clay Mineral Type Affects Soil Parameterization. Geophys. Res. Lett. 48, e2021GL095311.

Li, B. et al., 2021. Deep soil moisture limits the sustainable vegetation restoration in arid and semi-arid Loess Plateau. Geoderma 399, 115122.

Li, C. et al., 2021. Drivers and impacts of changes in China’s drylands. Nat. Rev. Earth Environ. 2, 858-873.

Li, L. et al., 2020. A Causal Inference Model Based on Random Forests to Identify the Effect of Soil Moisture on Precipitation. J. Hydrometeorol. 21, 1115-1131.

Li, X. et al., 2025. Clarifying the impacts of climate and vegetation changes on hydrological drought in the Yellow River Basin, China. J. Hydrol. Reg. Stud. 61, 102678.

Li, Y. et al., 2018. Divergent hydrological response to large-scale afforestation and vegetation greening in China. Sci. Adv. 4, eaar4182.

Liaw, A., Wiener, M., 2002. Classification and regression by randomForest. R news 2.3, 18-22.

Liu, B., Shao, M.a., 2016. Response of soil water dynamics to precipitation years under different vegetation types on the northern Loess Plateau, China. J. Arid Land 8, 47-59.

Liu, B. et al., 2024. Exacerbating water shortage induced by continuous expansion of surface artificial water bodies in the Yellow River Basin. J. Hydrol. 633, 130979.

Liu, L. et al., 2020. Soil moisture dominates dryness stress on ecosystem production globally. Nat. Commun. 11, 4892.

Liu, X. et al., 2025. Impacts of various soil moisture replenishment patterns on plant water use strategies in the Third Pole region. Catena, 254, 108942.

Mandal, U.K. et al., 2008. Evaluating impact of irrigation water quality on a calcareous clay soil using principal component analysis. Geoderma 144, 189-197.

Meier, R. et al., 2021. Empirical estimate of forestation-induced precipitation changes in Europe. Nat. Geosci. 14, 473-478.

Miguez-Macho, G., Fan, Y., 2021. Spatiotemporal origin of soil water taken up by vegetation. Nature 598, 624-628.

Morbidelli, R. et al., 2015. Infiltration on sloping surfaces: Laboratory experimental evidence and implications for infiltration modeling. J. Hydrol. 523, 79-85.

Nielsen, D.R., Bouma, J., 1985. Soil spatial variability: proceedings of a workshop of the ISSS (Int. Society of Soil Science) and the SSSA (Soil Science Society of America), Las Vegas, USA/Pdc296, Center Agricultural Pub and Document, Pudoc, Wageningen, The Netherlands.

NCAR, 2018. Global GIMMS NDVI3g v1 dataset (1981-2015) [dataset]. National Tibetan Plateau/Third Pole Environment Data Center. https://data.tpdc.ac.cn/zh-hans/data/9775f2b4-7370-4e5e-a537-3482c9a83d88

Piao, S. et al., 2020. Characteristics, drivers and feedbacks of global greening. Nat. Rev. Earth Environ. 1, 14-27.

Qiu, L. et al., 2021. Quantifying spatiotemporal variations in soil moisture driven by vegetation restoration on the Loess Plateau of China. J. Hydrol. 600, 126580.

Qiu, Y., Fu, B., Wang, J., Chen, L., 2001. Soil moisture variation in relation to topography and land use in a hillslope catchment of the Loess Plateau, China. J. Hydrol. 240, 243-263.

Schlaepfer, D.R. et al., 2017. Climate change reduces extent of temperate drylands and intensifies drought in deep soils. Nat. Commun. 8, 14196.

Seneviratne, S.I. et al., 2010. Investigating soil moisture–climate interactions in a changing climate: A review. Earth Sci. Rev. 99, 125-161.

Sun, H., Lee, J., Chen, X., Zhuang, J., 2020. Estimating soil water retention for wide ranges of pressure head and bulk density based on a fractional bulk density concept. Sci. Rep. 10, 16666.

Sun, W., Shao, Q., Liu, J., Zhai, J., 2014. Assessing the effects of land use and topography on soil erosion on the Loess Plateau in China. Catena 121, 151-163.

Tong, Y., Wang, Y., 2025. The Hyper local geographically weighted regression calculation code for the Yellow River Basin dataset [dataset]. Loess Science Data Center, Institute of Earth Environment, Chinese Academy of Sciences. https://doi.org/10.12262/lsdc.61099.2512.

Tran, D.X. et al., 2022. Quantifying spatial non-stationarity in the relationship between landscape structure and the provision of ecosystem services: An example in the New Zealand hill country. Sci. Total Environ. 808, 152126.

Tuller, M., Or, D., 2005. Water films and scaling of soil characteristic curves at low water contents. Water Resour. Res. 41, W09403

Vereecken, H. et al., 2014. On the spatio-temporal dynamics of soil moisture at the field scale. J. Hydrol. 516, 76-96.

Wang, Q. et al., 2024a. Will large-scale forestation lead to a soil water deficit crisis in China's drylands? Sci. Bull. 69, 1506-1514.

Wang, T., Franz, T.E., 2015. Field Observations of Regional Controls of Soil Hydraulic Properties on Soil Moisture Spatial Variability in Different Climate Zones.

Vadose Zone J. 14, vzj2015.02.0032.

Wang, T. et al., 2017a. Evaluating climate and soil effects on regional soil moisture spatial variability using EOFs. Water Resour. Res. 53, 4022-4035.

Wang, T. et al., 2017b. Spatial patterns of soil moisture from two regional monitoring networks in the United States. J. Hydrol. 552, 578-585.

Wang, W., Zhang, Y., Tang, Q., 2020. Impact assessment of climate change and human activities on streamflow signatures in the Yellow River Basin using the Budyko hypothesis and derived differential equation. J. Hydrol. 591, 125460.

Wang, X. et al., 2008. Effects of rainfall characteristics on infiltration and redistribution patterns in revegetation-stabilized desert ecosystems. J. Hydrol. 358, 134-143.

Wang, Y. et al., 2024b. Soil moisture decline in China’s monsoon loess critical zone: More a result of land-use conversion than climate change. Proc. Natl. Acad. Sci. 121, e2322127121.

Wang, Y., Shao, M.a., Liu, Z., 2010. Large-scale spatial variability of dried soil layers and related factors across the entire Loess Plateau of China. Geoderma 159, 99-108.

Wang, Y., Shao, M.a., Liu, Z., 2013. Vertical distribution and influencing factors of soil water content within 21-m profile on the Chinese Loess Plateau. Geoderma 193, 300-310.

Wang, Y., Shao, M.a., Liu, Z., Warrington, D.N., 2012. Investigation of Factors Controlling the Regional-Scale Distribution of Dried Soil Layers Under Forestland on the Loess Plateau, China. Surv. Geophys. 33, 311-330.

Wankmüller, F.J.P. et al., 2024. Global influence of soil texture on ecosystem water limitation. Nature, 635, 631-638.

Wei, B. et al., 2023. Drivers of biochar-mediated improvement of soil water retention capacity based on soil texture: A meta-analysis. Geoderma 437, 116591.

Will, R., Wilson, S., Zou, C., Hennessey, T., 2013. Increased vapor pressure deficit due to higher temperature leads to greater transpiration and faster mortality during drought for tree seedlings common to the forest-grassland ecotone. New Phytol. 200, 366-374.

Wu, D., Di, C., Wang, T., Wang, L., Chen, X., 2021. Characterization of the Coherence Between Soil Moisture and Precipitation at Regional Scales. J. Geophys. Res.-Atmos. 126, e2020JD034340.

Wu, D., Wang, T., Di, C., Wang, L., Chen, X., 2020. Investigation of controls on the regional soil moisture spatiotemporal patterns across different climate zones. Sci. Total Environ. 726, 138214.

Xu, X., 2018. The spatial distribution dataset of monthly Vegetation Index (NDVI) in China (In Chinese) [dataset]. Data Registration and Publishing System of the Data Center for Resources and Environmental Sciences, Chinese Academy of Sciences, v1. http://www.resdata.com.cn/DataProduct/Detail/115.

Yang, Z. et al., 2012. Temperature effect on transpiration response of maize plants to vapour pressure deficit. Environ. Exp. Bot. 78, 157-162.

Yin, G. et al., 2023. Aspect Matters: Unraveling Microclimate Impacts on Mountain Greenness and Greening. Geophys. Res. Lett. 50, e2023GL105879.

Yin, L. et al., 2019. Trade-offs and synergy between ecosystem services in National Barrier Zone. Geogr. Res. 38, 2162-2172.

Zeren Cetin, I., Varol, T., Ozel, H.B., 2023. A geographic information systems and remote sensing–based approach to assess urban micro-climate change and its impact on human health in Bartin, Turkey. Environ. Monit. Assess. 195, 540.

Zhang, K. et al., 2024. Spatiotemporal interactions between soil moisture and water availability across the Yellow River Basin, China. J. Hydrol. Reg. Stud. 54, 101874.

Zhang, S. et al., 2018. Excessive Afforestation and Soil Drying on China's Loess Plateau. J. Geophys. Res-Biogeo.123, 923-935.

Zhao, C., Jia, X., Shao, M.a., Zhu, Y., 2021. Regional variations in plant-available soil water storage and related driving factors in the middle reaches of the Yellow River Basin, China. Agric. Water Manag. 257, 107131.

Zhao, Y., Wang, Y., Zhou, J., Sun, H., Jin, Z., 2022. Spatiotemporal soil water storage variation comparison between newly formed and untreated gully land sites under a land restoration project and associated implications on land management. Ecol. Eng. 180, 106670.

Zhou, S. et al., 2021. Soil moisture–atmosphere feedbacks mitigate declining water availability in drylands. Nat. Clim. Change 11, 38-44.

Zhou, Y. et al., 2019. Heavy metal distribution, relationship and prediction in a wheat-rice rotation system. Geoderma, 354, 113886.

Zhu, P., Jia, X., Zhao, C., Shao, M., 2022. Long-term soil moisture evolution and its driving factors across China's agroecosystems. Agric. Water Manag. 269, 107735.

Zuo, Z. et al., 2024. Importance of soil moisture conservation in mitigating climate change. Sci. Bull. 69, 1332-1341.

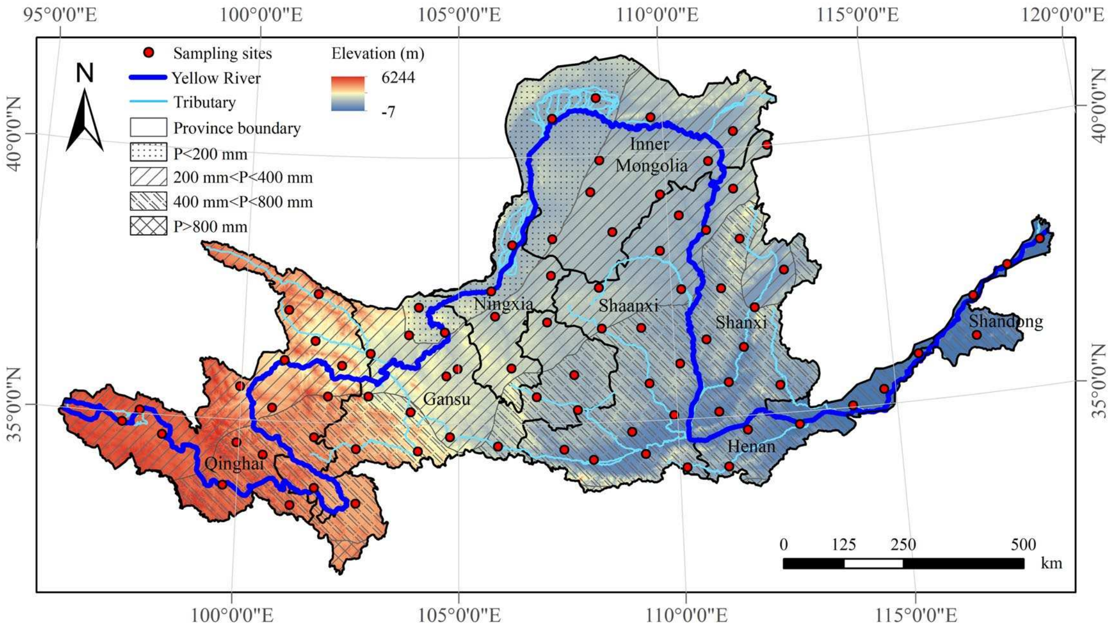

Figure 1. The location of the sampling sites in the Yellow River Basin with mean precipitation (P).

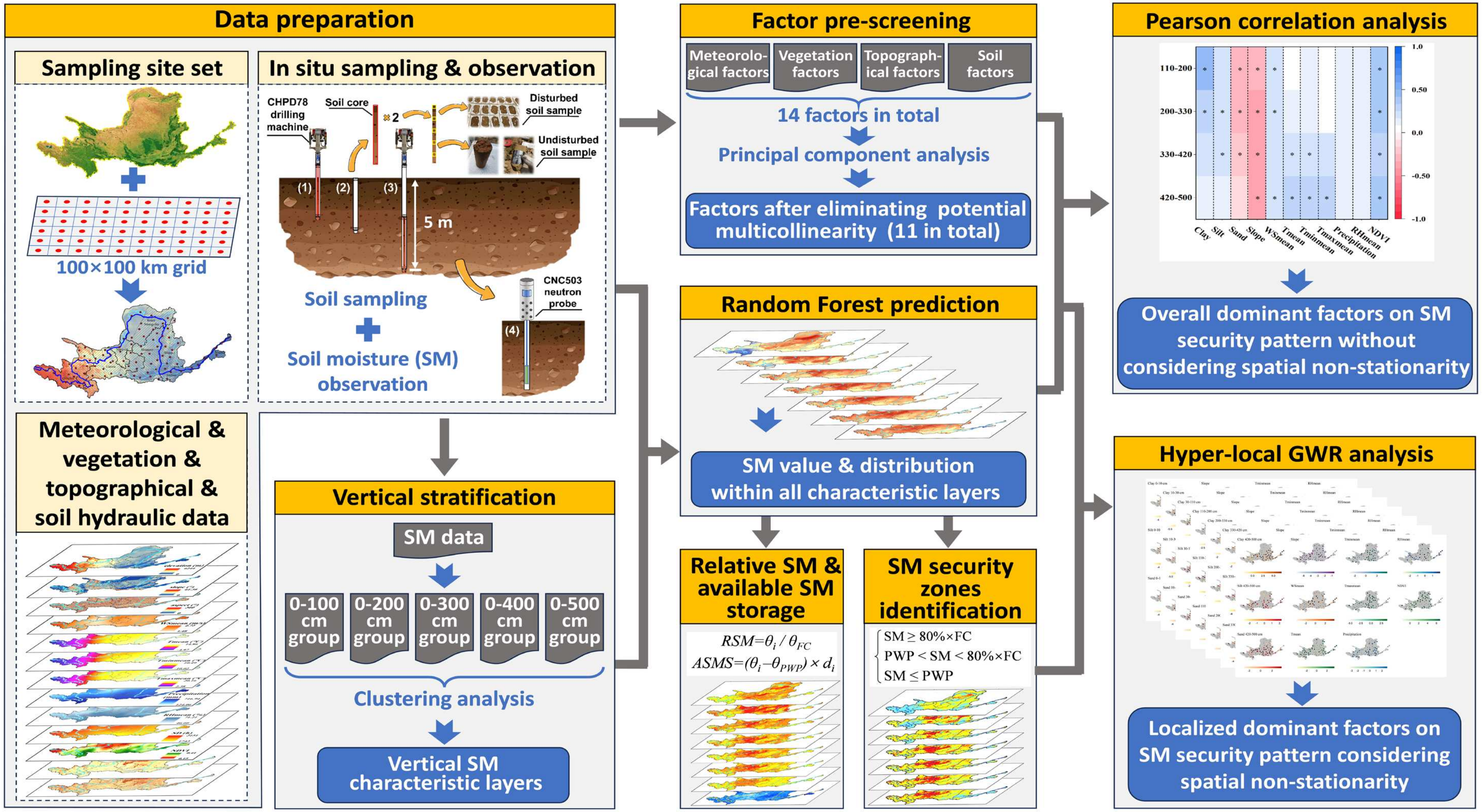

Figure 2. The schematic diagram of the methodological framework.

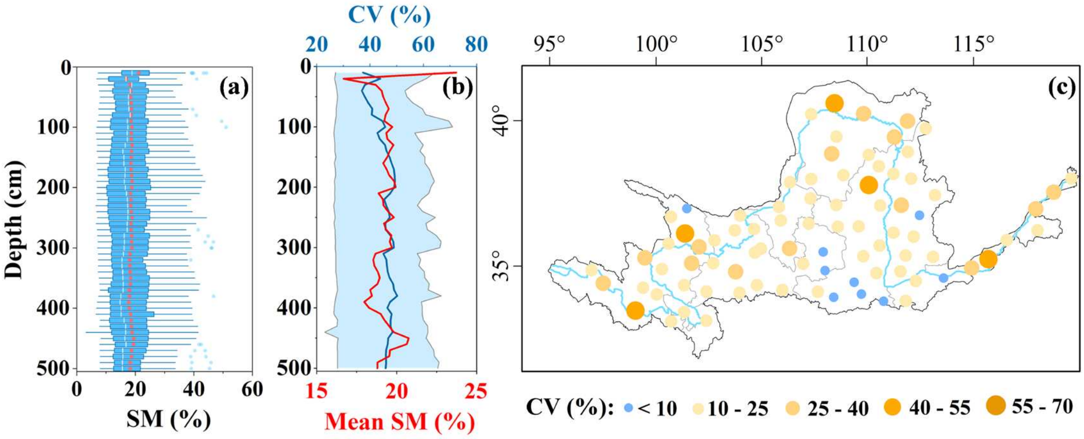

Figure 3. The distribution character of soil moisture (SM) and its coefficient of variation (CV) in the Yellow River Basin. In subgraph (a), the red dots are mean values; the blue boxes show the interquartile range (IQR; Q1–Q3); the center line denotes the median; the whiskers extend to the most extreme points within 1.5×IQR from the quartiles; and points beyond are outliers. In subgraph (b), the blue shaded area represents the CV distribution range.

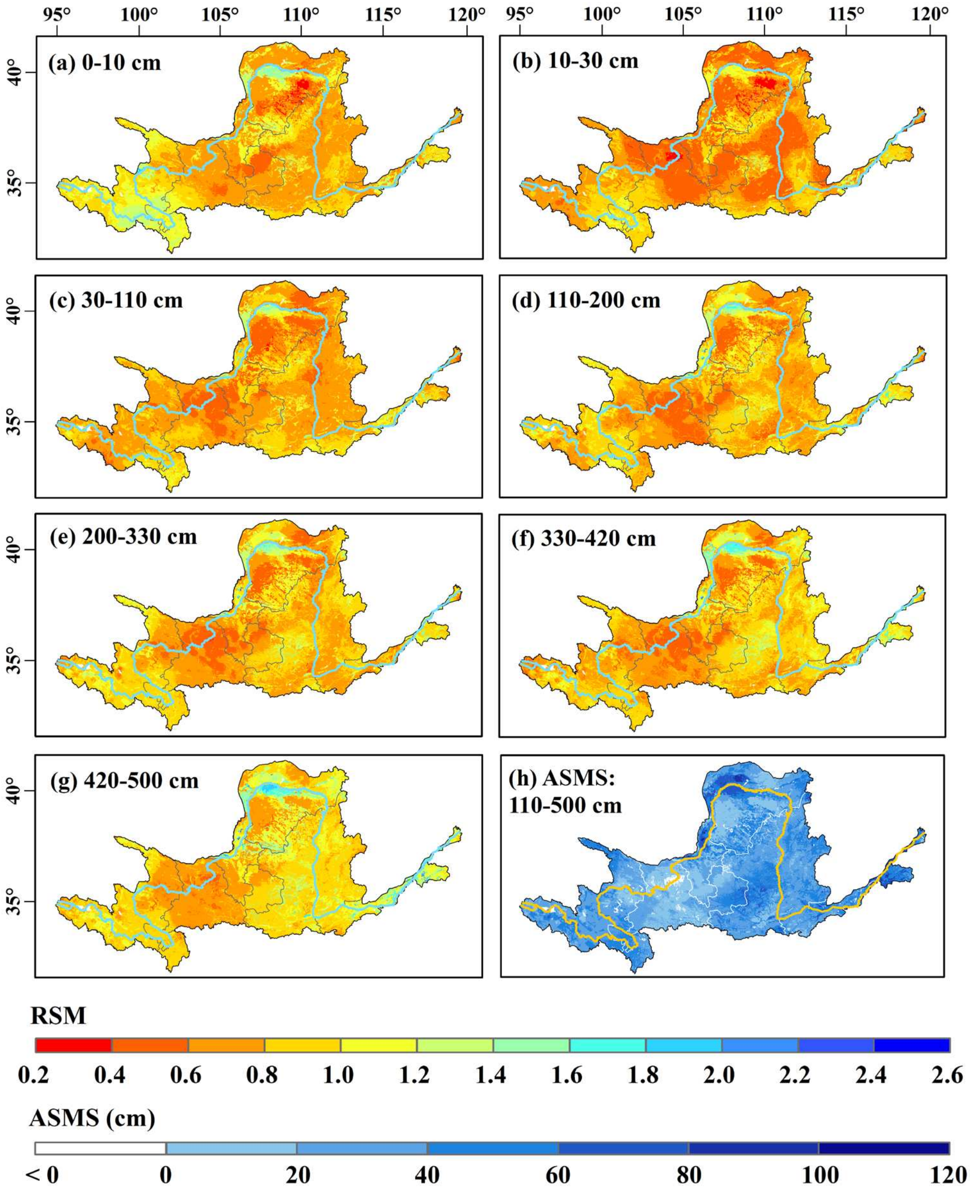

Figure 4. Spatial distribution of the relative soil moisture (RSM) within seven characteristic layers and the available soil moisture storage (ASMS) in the soil layer of 110–500 cm across the Yellow River Basin.

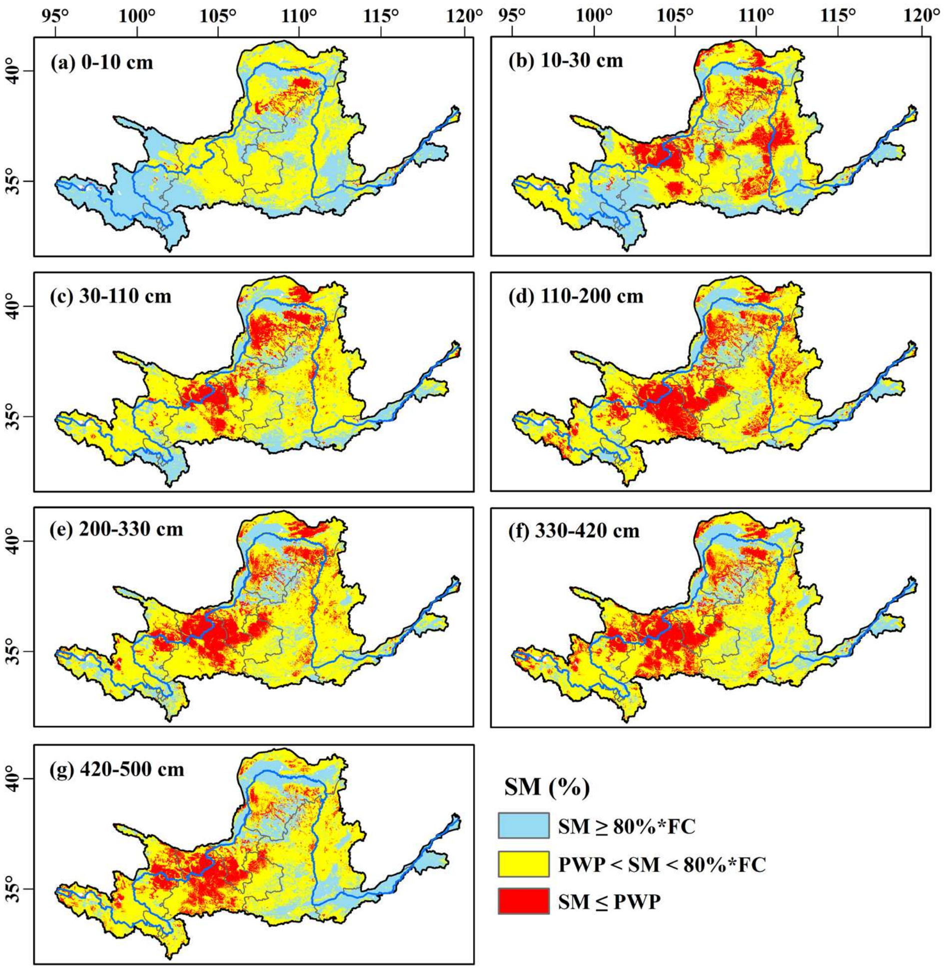

Figure 5. Soil moisture (SM) security spatial patterns within seven characteristic layers in the Yellow River Basin, where FC represents the field capacity and PWP represents the permanent wilting point. Blue represents the wet zone; yellow represents the transitional zone; and red represents the dry zone.

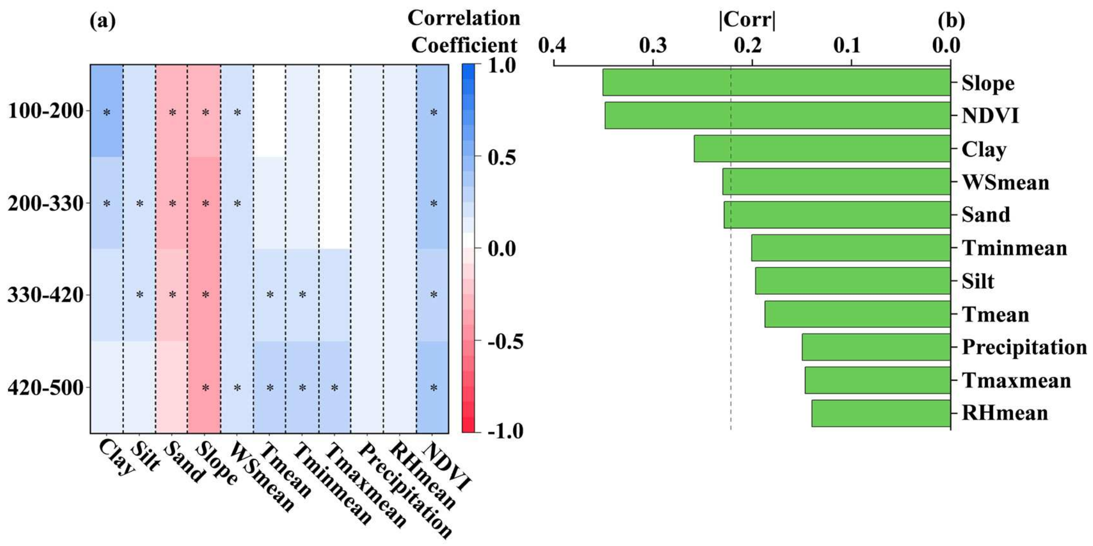

Figure 6. The Pearson correlation analysis within 110–500 cm across the whole Yellow River Basin. Specifically, the correlation coefficients between the soil moisture and the soil texture, topographic, meteorological, and vegetation factors in each soil layer (a) and the ranking of the absolute values of correlation coefficients (b). Significant correlations are marked with black dots. The grey dashed line in (b) represents the average of all factors. Corr is coefficient of correlation. WSmean is mean wind speed; Tmean is mean temperature; Tminmean is mean minimum temperature; Tmaxmean is mean maximum temperature; Precipitation is mean annual precipitation; RHmean is mean relative humidity; and NDVI is normalized difference vegetation index.

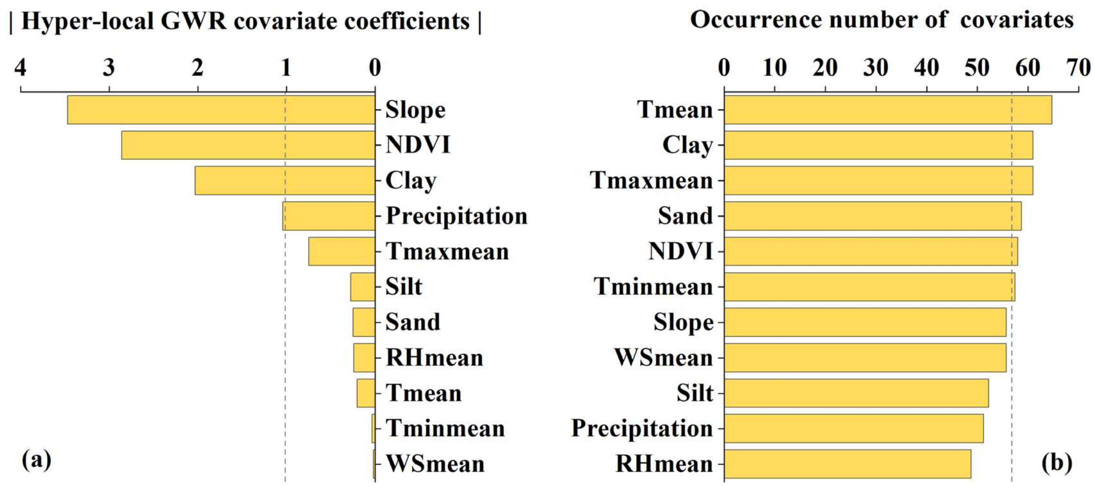

Figure 7. The results of the Hyper-Local geographically weighted regression (GWR) analysis within 110–500 cm across the whole Yellow River Basin. Specifically, the ranking of absolute values of the averaged standardized covariate coefficients of all regression equations in Hyper-Local GWR (a) and the total occurrence number of different covariates in all regression formulas (b). WSmean is mean wind speed; Tmean is mean temperature; Tminmean is mean minimum temperature; Tmaxmean is mean maximum temperature; Precipitation is mean annual precipitation; RHmean is mean relative humidity; and NDVI is normalized difference vegetation index. The grey dashed line represents the average of all factors. All covariate coefficients in the figure are standardized and all factors involved in the analysis are pre-screened by the principal component analysis.

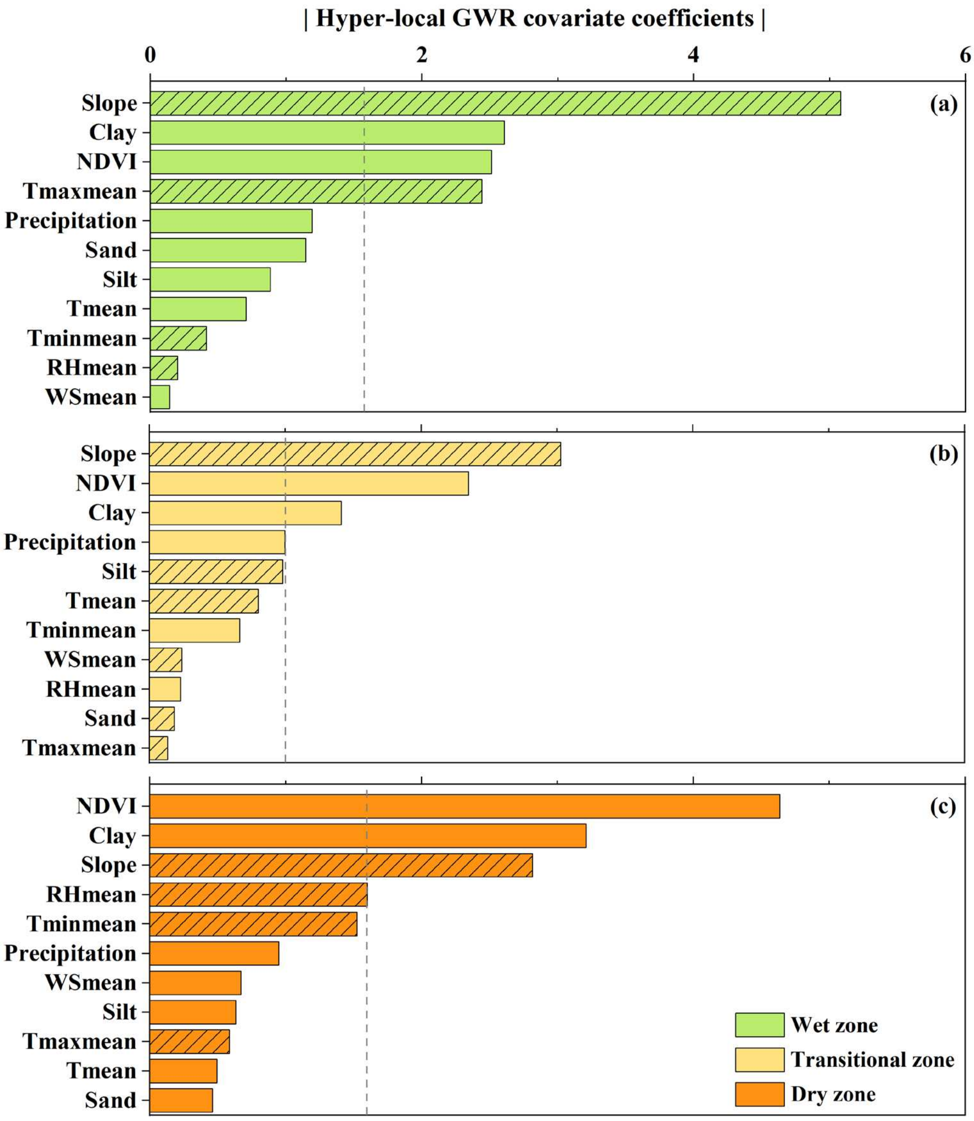

Figure 8. The order of the absolute values of average Hyper-local geographically weighted regression (GWR) covariate coefficient corresponding to 110–500 cm soil layers in three soil moisture (SM) security zones. Each subgraph is arranged from top to bottom in descending order. WSmean is mean wind speed; Tmean is mean temperature; Tminmean is mean minimum temperature; Tmaxmean is mean maximum temperature; Precipitation is mean annual precipitation; RHmean is mean relative humidity; and NDVI is normalized difference vegetation index. The grey dashed line represents the average of all factors. The shaded areas with diagonal lines indicate negative average regression coefficients.

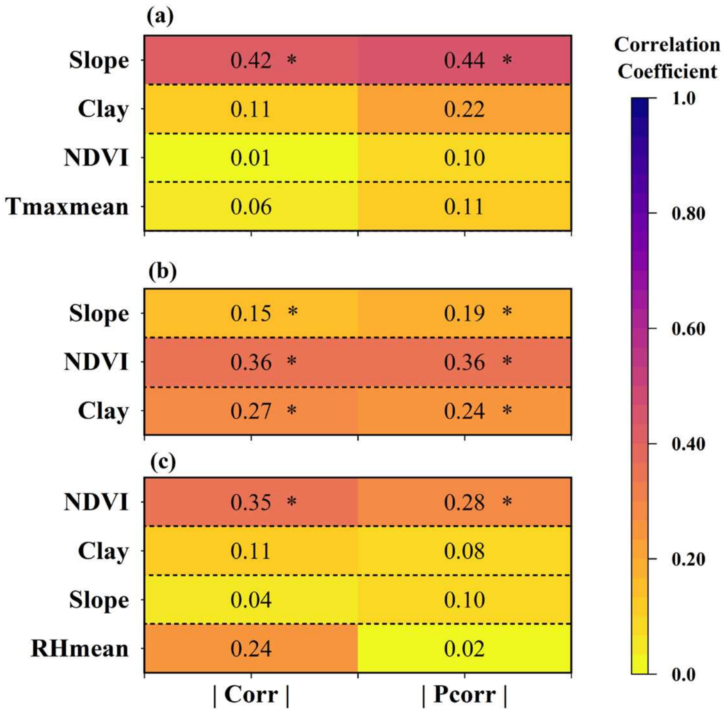

Figure 9. The absolute value of the coefficient of correlation (Corr) and partial correlation analysis (Pcorr) between soil moisture and soil moisture security dominating factors screened by Hyper-local geographically weighted regression in the wet zone (a), transitional zone (b), and dry zone (c). Significant correlations are highlighted with black dots. NDVI is normalized difference vegetation index; Tmaxmean is mean maximum temperature; and RHmean is mean relative humidity.

Table 1. The thresholds that define the soil moisture (SM) security zones based on their relationship to field capacity (FC) and permanent wilting point (PWP).

SM security zones	Thresholds
Wet zone	SM ≥ 80% * FC
Transitional zone	PWP < SM < 80% * FC
Dry zone	SM ≤ PWP

Table 2. Statistical evaluation indicators of the random forest model.

SM	Statistic	0-10 cm		10-30 cm		30-110 cm		110-200 cm		200-330 cm		330-420 cm		420-500 cm		0-500 cm	
		Med.	Mean	Med.	Mean	Med.	Mean	Med.	Mean	Med.	Mean	Med.	Mean	Med.	Mean	Med.	Mean
Training dataset	R^2	0.93	0.93	0.95	0.95	0.95	0.95	0.94	0.94	0.94	0.94	0.95	0.95	0.94	0.94	0.95	0.95
	RMSE	0.11	0.11	0.12	0.12	0.13	0.12	0.16	0.16	0.16	0.16	0.17	0.17	0.17	0.17	0.14	0.14
	MAE	0.09	0.09	0.10	0.10	0.10	0.10	0.13	0.13	0.13	0.13	0.13	0.13	0.14	0.14	0.11	0.11
	p	< 0.05	< 0.05	< 0.05	< 0.05	< 0.05	< 0.05	< 0.05	< 0.05	< 0.05	< 0.05	< 0.05	< 0.05	< 0.05	< 0.05	< 0.05	< 0.05
Testing dataset	R^2	0.46	0.45	0.57	0.56	0.50	0.49	0.42	0.41	0.39	0.39	0.42	0.42	0.41	0.41	0.47	0.46
	RMSE	0.27	0.27	0.29	0.29	0.30	0.30	0.37	0.38	0.39	0.39	0.40	0.40	0.41	0.41	0.33	0.33
	MAE	0.21	0.21	0.24	0.24	0.25	0.25	0.31	0.31	0.32	0.32	0.33	0.33	0.34	0.34	0.27	0.27
	p	< 0.05	< 0.05	< 0.05	< 0.05	< 0.05	< 0.05	< 0.05	< 0.05	< 0.05	0.05	< 0.05	< 0.05	< 0.05	< 0.05	< 0.05	< 0.05

Note: All evaluation indicators came from statistical results of 1000 times of random sampling. R^2, coefficient of determination; RMSE, root mean square error; MAE, mean absolute error; p, significance level; Med., median value.

Supporting Information for

Deciphering spatial pattern and environmental drivers of deep soil moisture security: Insights from integrating spatial non-stationarity in large-scale analysis

Yongping Tong[1, 2, 3, 4], Yunqiang Wang[1, 2, 4, 5, *], Zimin Li[1, 2, 4], Pingping Zhang[1, 2, 4], Wei Hu[6], Jingxiong Zhou[1, 2, 4], Xiangyu Guo[1, 4, 7], Hui Sun[1, 4, 7], Alexis Comber[8], Ronny Lauerwald[3]

[1] State Key Laboratory of Loess Science, Institute of Earth Environment, Chinese Academy of Sciences, Xi'an 710061, China

[2] University of Chinese Academy of Sciences, Beijing 100049, China

[3] Université Paris-Saclay, INRAE, AgroParisTech, UMR Ecosys, Palaiseau 91120, France

[4] National Observation and Research Station of Earth Critical Zone on the Loess Plateau, Xi'an 710061, China

[5] Department of Earth and Environmental Sciences, Xi'an Jiaotong University, Xi'an 710049, China

[6] The New Zealand Institute for Plant and Food Research Limited, Private Bag 4704, Christchurch 8140, New Zealand

[7] Xi'an Institute for Innovative Earth Environment Research, Xi'an 710061, China

[8] School of Geography, University of Leeds, Leeds, LS2 9JT, UK

*** Corresponding author:** Yunqiang Wang (wangyq@ieecas.cn)

Address: State Key Laboratory of Loess Science, Institute of Earth Environment, Chinese Academy of Sciences, Xi'an, Shaanxi 710061, China.

Contents of this file

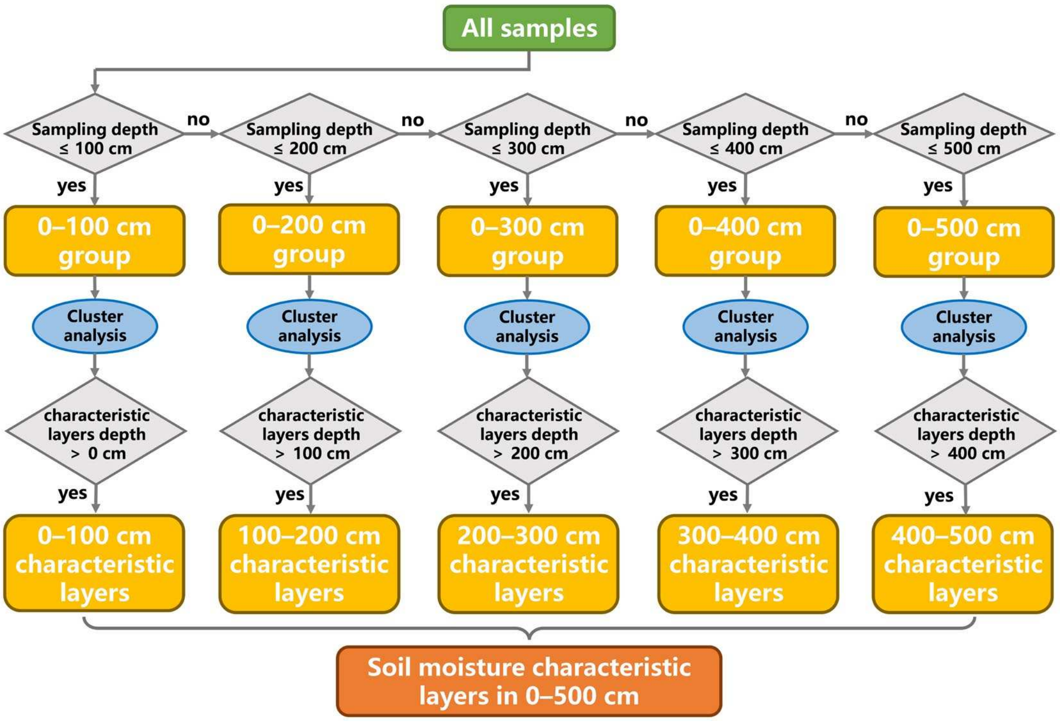

Figure S1. Schematic diagram of dividing the soil moisture profile into characteristic layers based on cluster analysis.

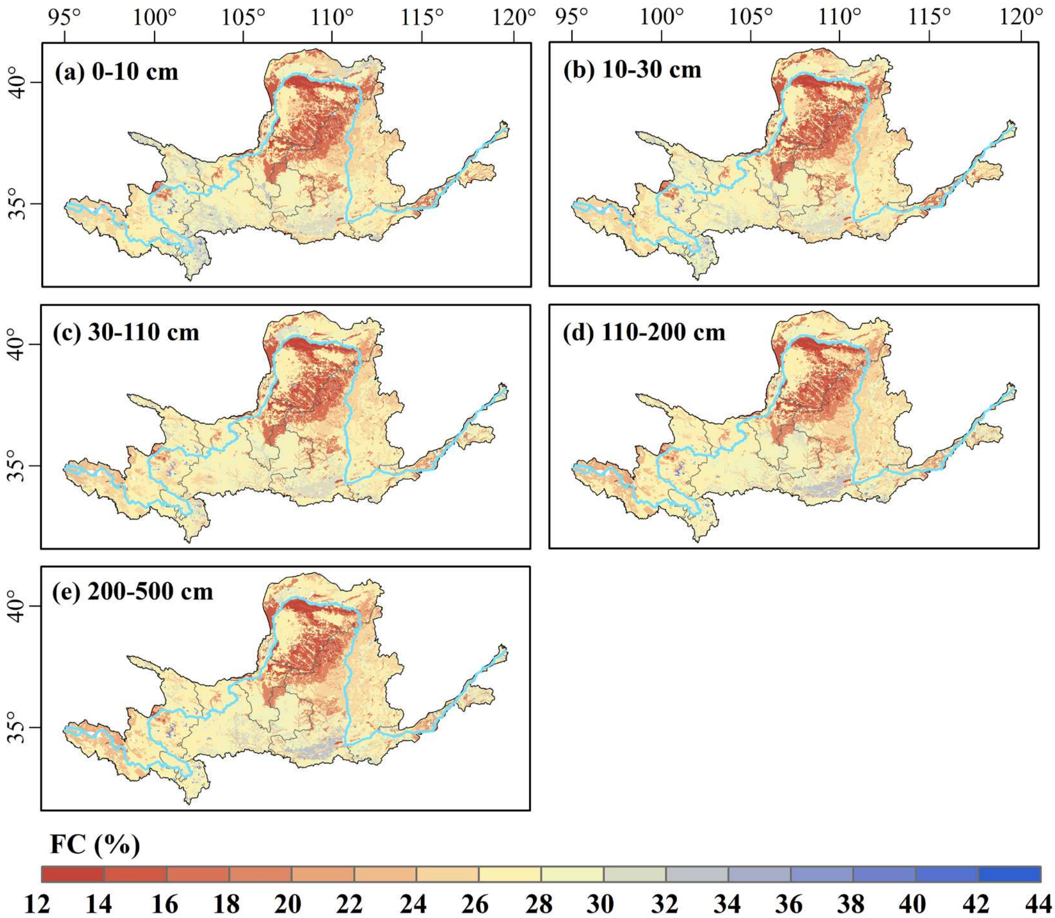

Figure S2. Spatial distributions of the field capacity (FC) in the various layers from 0–500 cm. The original data of this figure are sourced from an existing dataset (Dai et al., 2013), and the calculation methods of these five specified layers were detailed in Table S1.

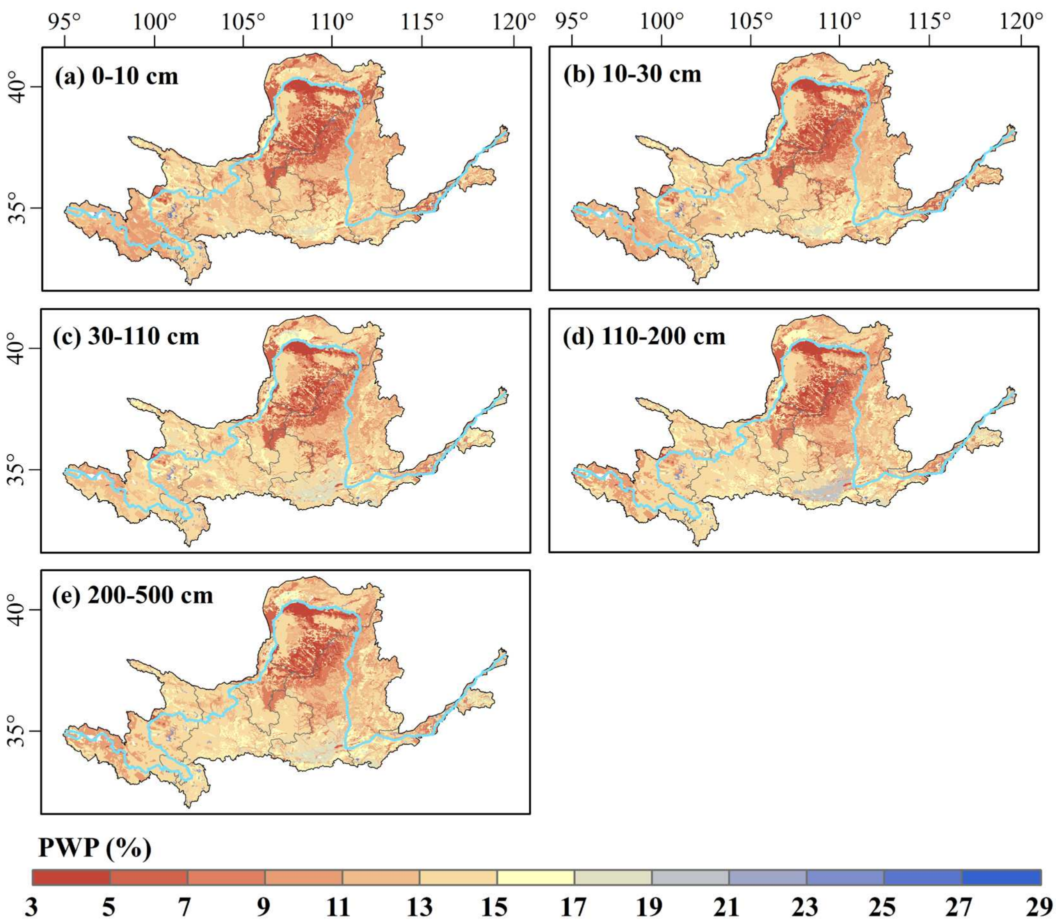

Figure S3. Spatial distributions of the permanent wilting point (PWP) in the various layers from 0–500 cm. The original data of this figure are sourced from an existing dataset (Dai et al., 2013), and the calculation methods of these five specified layers were detailed in Table S1.

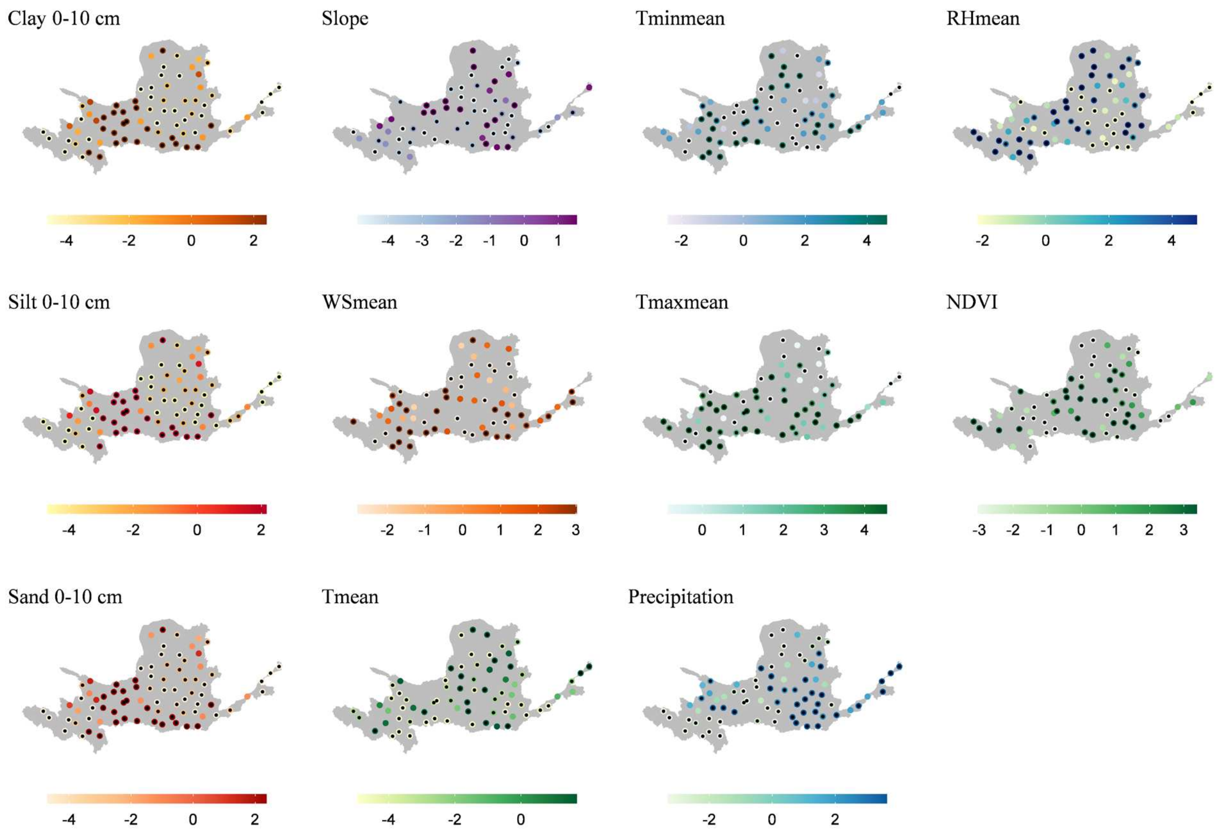

Figure S4. Standardized regression coefficients and their spatial distributions in the Hyper-local geographically weighted regression model of soil moisture in the 0–10 cm soil layer. Significant t-values are indicated by black dots. The legend displays the inter-quartile ranges of the data with extreme values set at the 25th or 75th quartile value. WSmean is mean wind speed; Tmean is mean temperature; Tminmean is mean minimum temperature; Tmaxmean is mean maximum temperature; Precipitation is mean annual precipitation; RHmean is mean relative humidity; and NDVI is normalized difference vegetation index.

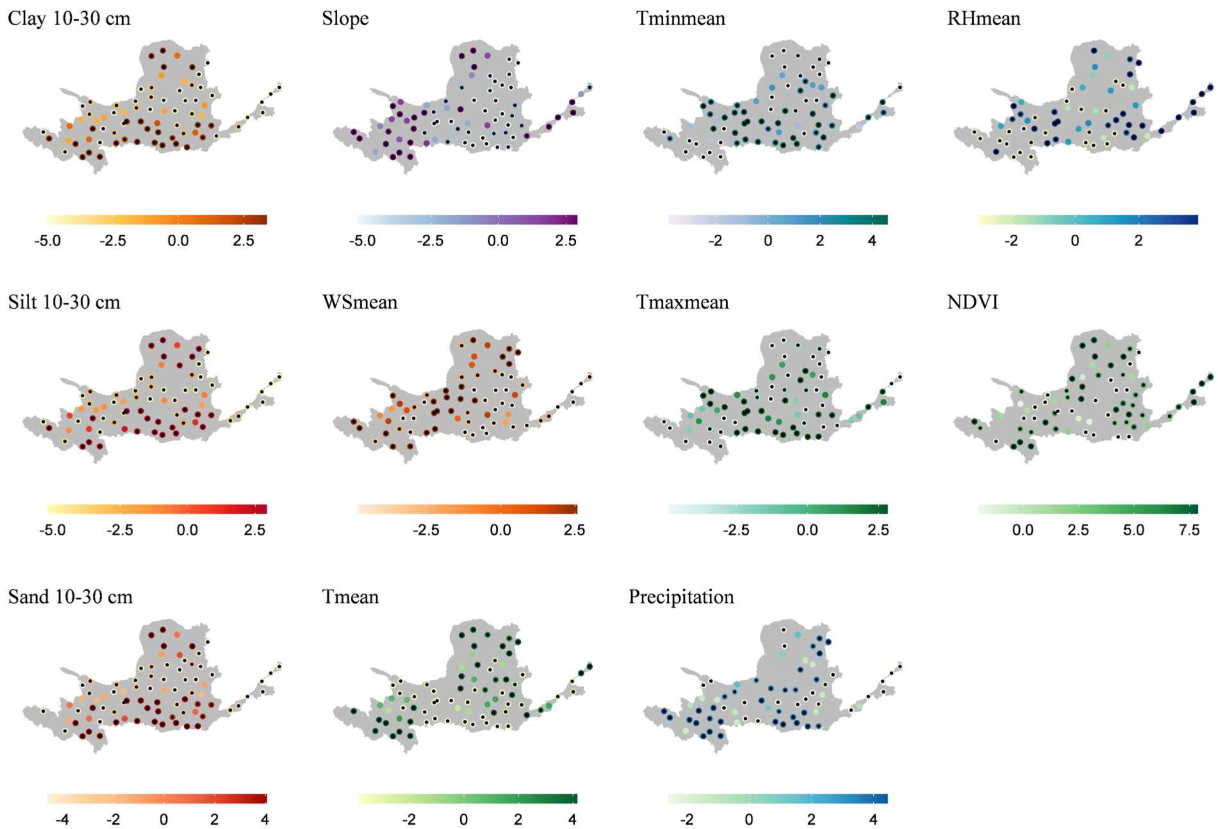

Figure S5. Standardized regression coefficients and their spatial distributions in the Hyper-local geographically weighted regression model of soil moisture in the 10–30 cm soil layer. Significant t-values are indicated by a black dot. The legend displays the inter-quartile range of the data with extreme values set at the 25th or 75th quartile value. WSmean is mean wind speed; Tmean is mean temperature; Tminmean is mean minimum temperature; Tmaxmean is mean maximum temperature; Precipitation is mean annual precipitation; RHmean is mean relative humidity; and NDVI is normalized difference vegetation index.

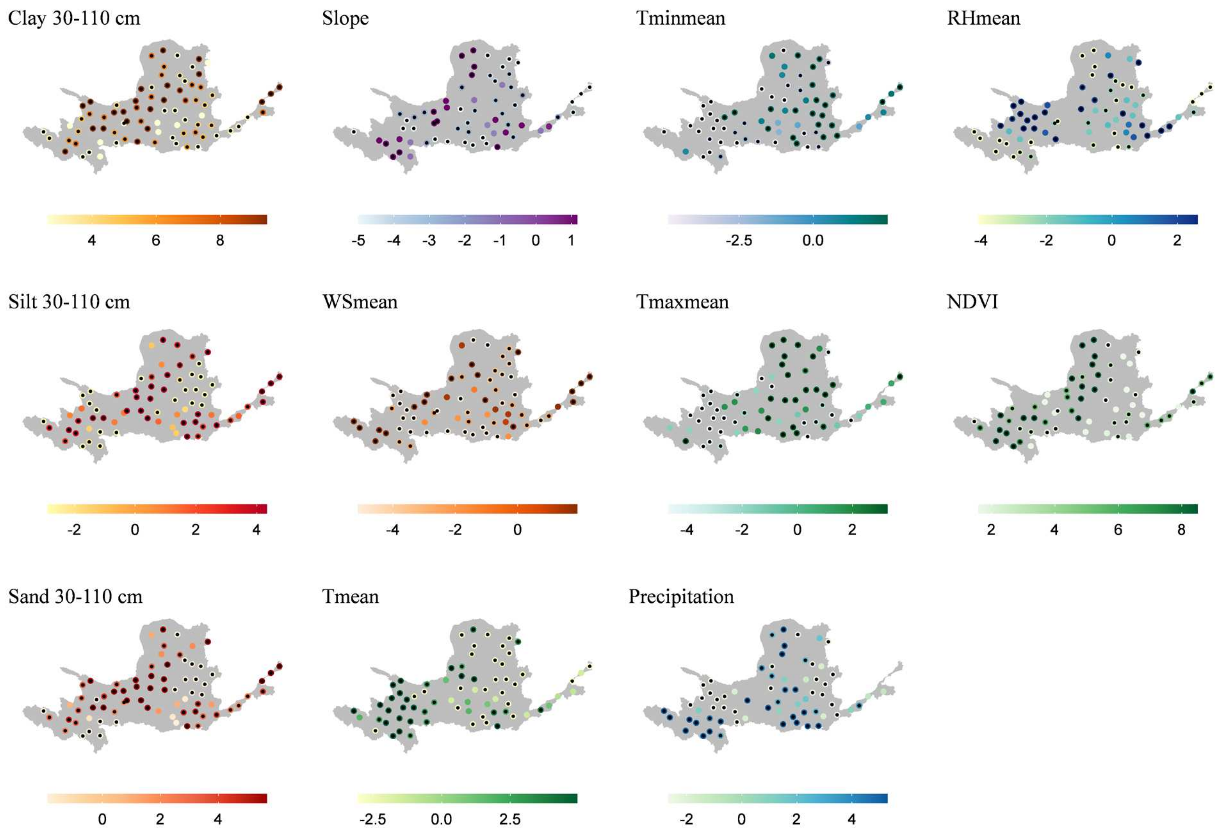

Figure S6. Standardized regression coefficients and their spatial distributions in the Hyper-local geographically weighted regression model of soil moisture in the 30–110 cm soil layer. Significant t-values are indicated by a black dot. The legend displays the inter-quartile range of the data with extreme values set at the 25th or 75th quartile value. WSmean is mean wind speed; Tmean is mean temperature; Tminmean is mean minimum temperature; Tmaxmean is mean maximum temperature; Precipitation is mean annual precipitation; RHmean is mean relative humidity; and NDVI is normalized difference vegetation index.

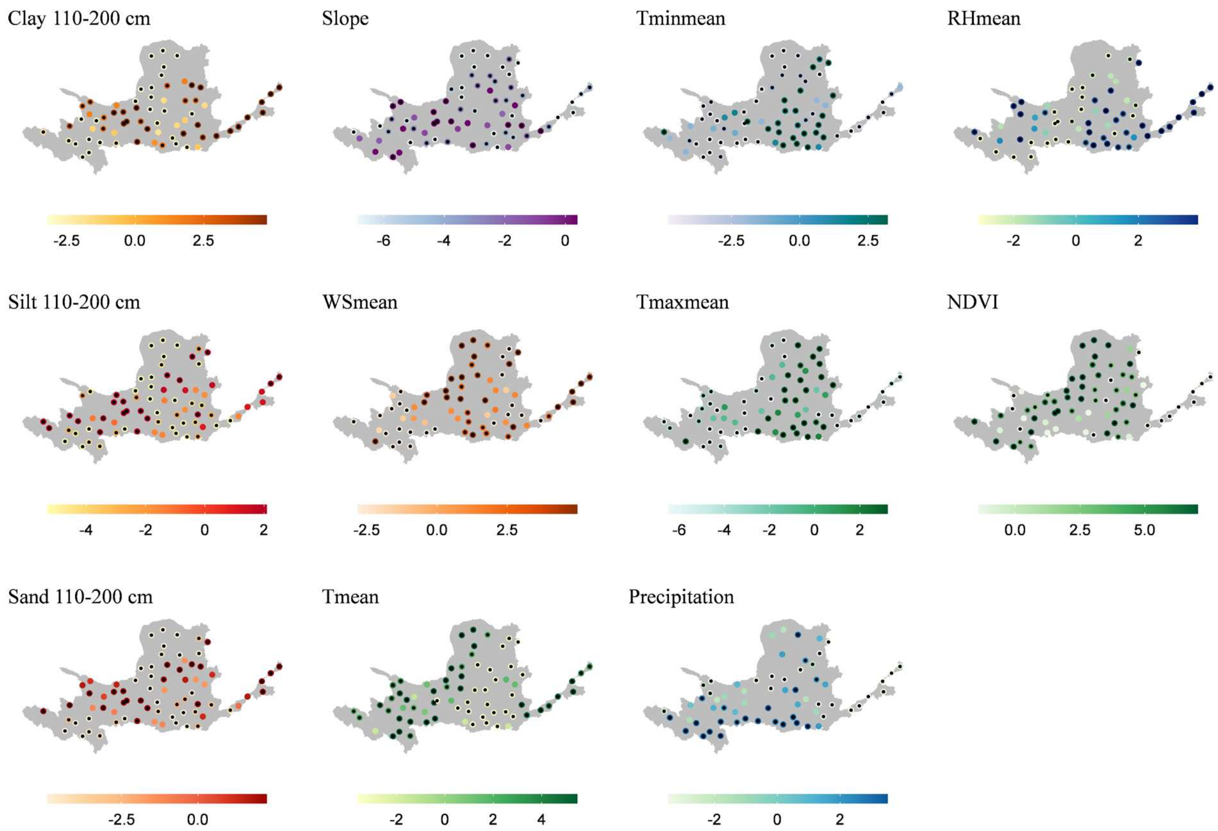

Figure S7. Standardized regression coefficients and their spatial distributions in the Hyper-local geographically weighted regression model of soil moisture in the 110–200 cm soil layer. Significant t-values are indicated by a black dot. The legend displays the inter-quartile range of the data with extreme values set at the 25th or 75th quartile. WSmean is mean wind speed; Tmean is mean temperature; Tminmean is mean minimum temperature; Tmaxmean is mean maximum temperature; Precipitation is mean annual precipitation; RHmean is mean relative humidity; and NDVI is normalized difference vegetation index.

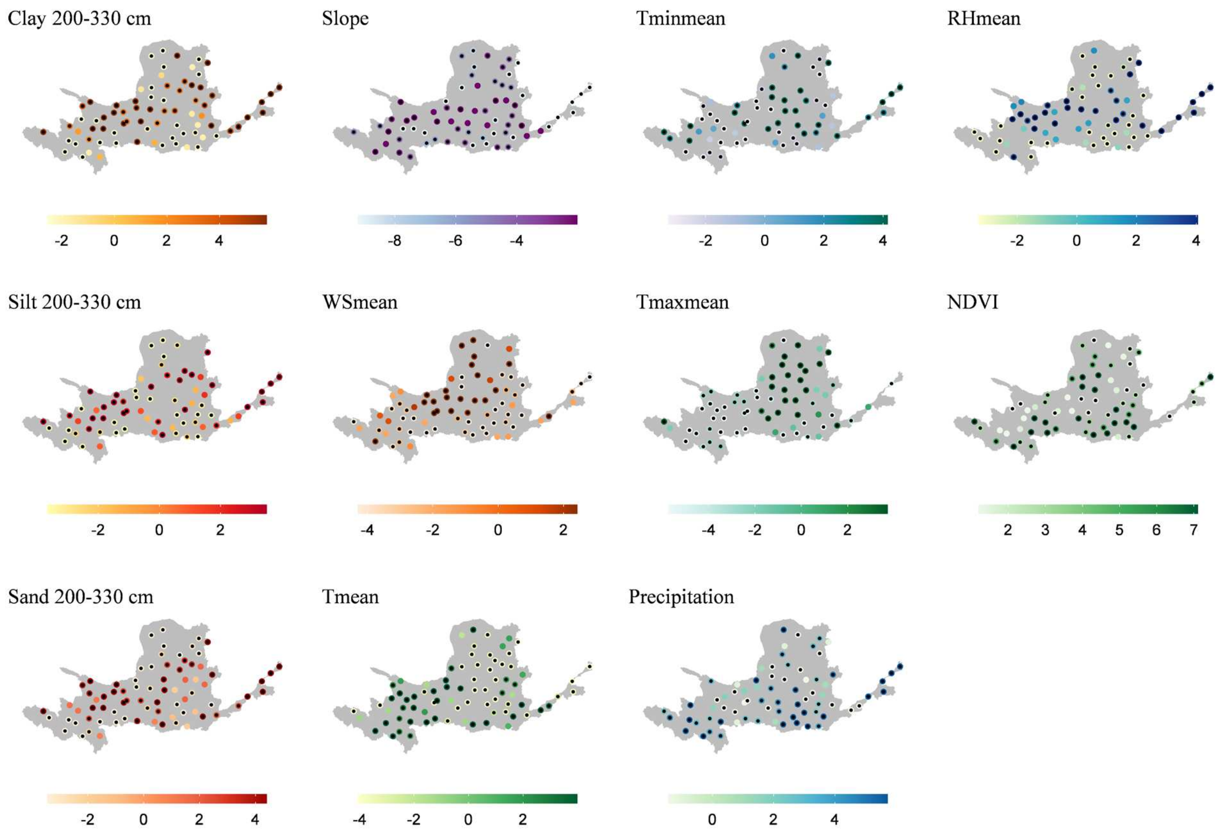

Figure S8. Standardized regression coefficients and their spatial distributions in the Hyper-local geographically weighed regression model of soil moisture in the 200–330 cm soil layer. Significant t-values are indicated by a black dot. The legend displays the inter-quartile range of the data with extreme values set at the 25th or 75th quartile value. WSmean is mean wind speed; Tmean is mean temperature; Tminmean is mean minimum temperature; Tmaxmean is mean maximum temperature; Precipitation is mean annual precipitation; RHmean is mean relative humidity; and NDVI is normalized difference vegetation index.

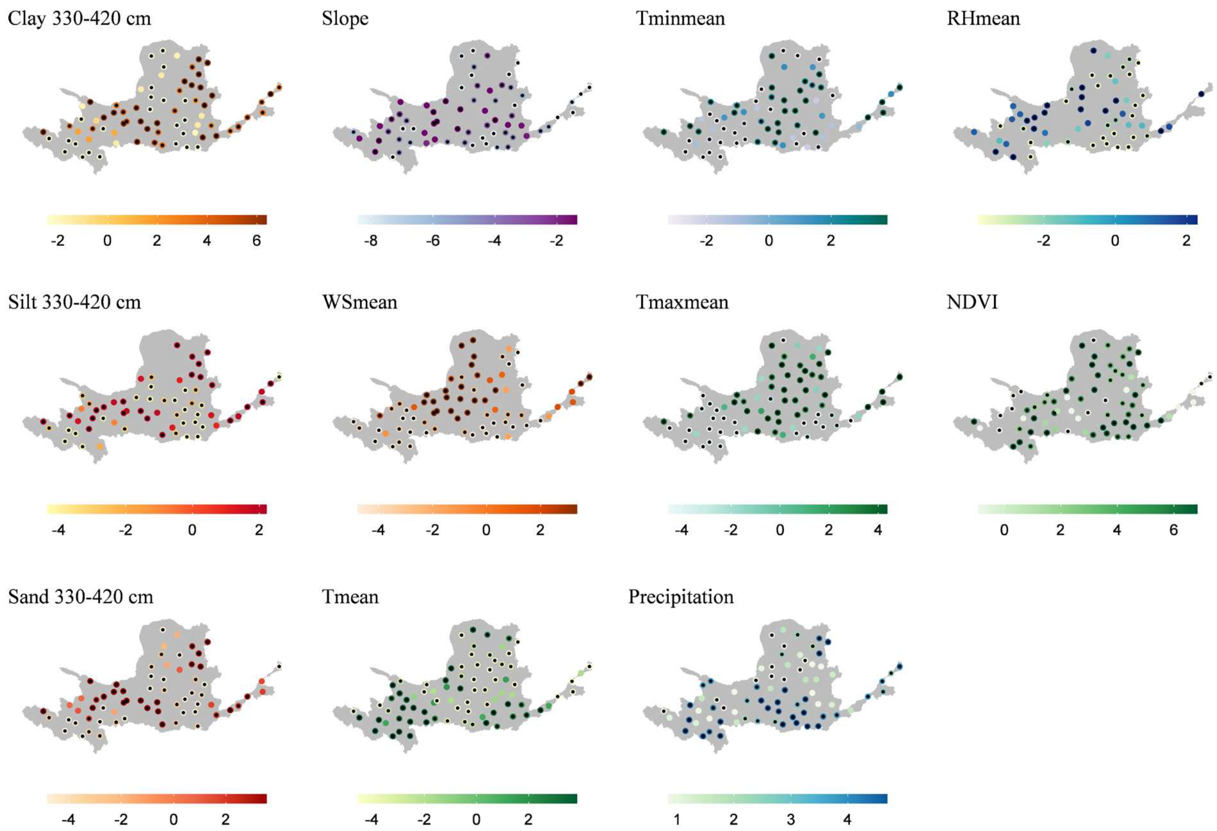

Figure S9. Standardized regression coefficients and their spatial distributions in the Hyper-local geographically weighted regression model of soil moisture in the 330–420 cm soil layer. Significant t-values are indicated by a black dot. The legend displays the inter-quartile range of the data with extreme values set at the 25th or 75th quartile value. WSmean is mean wind speed; Tmean is mean temperature; Tminmean is mean minimum temperature; Tmaxmean is mean maximum temperature; Precipitation is mean annual precipitation; RHmean is mean relative humidity; and NDVI is normalized difference vegetation index.

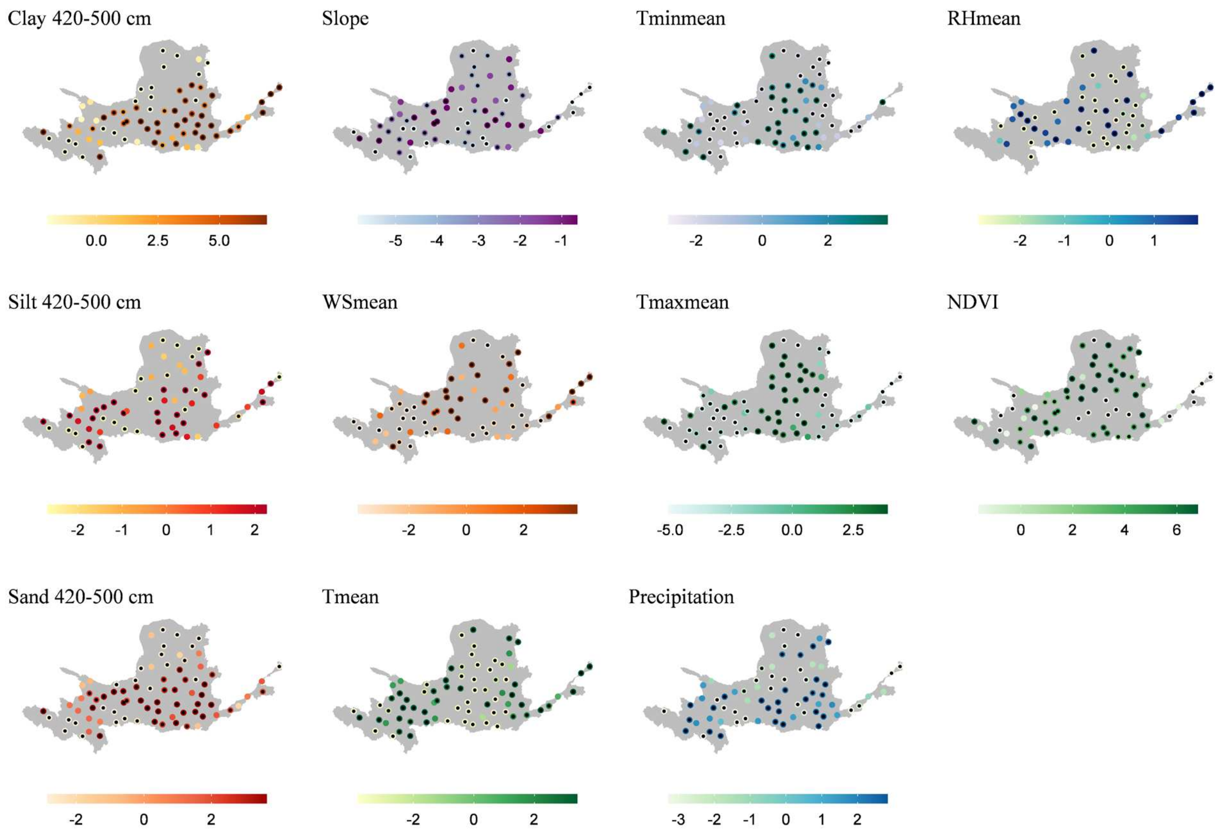

Figure S10. Standardized regression coefficients and their spatial distributions in the Hyper-local geographically weighted regression model of soil moisture in the 420–500 cm soil layer. Significant t-values are indicated by a black dot. The legend displays the inter-quartile ranges of the data with extreme values set at the 25th or 75th quartile value. WSmean is mean wind speed; Tmean is mean temperature; Tminmean is mean minimum temperature; Tmaxmean is mean maximum temperature; Precipitation is mean annual precipitation; RHmean is mean relative humidity; and NDVI is normalized difference vegetation index.

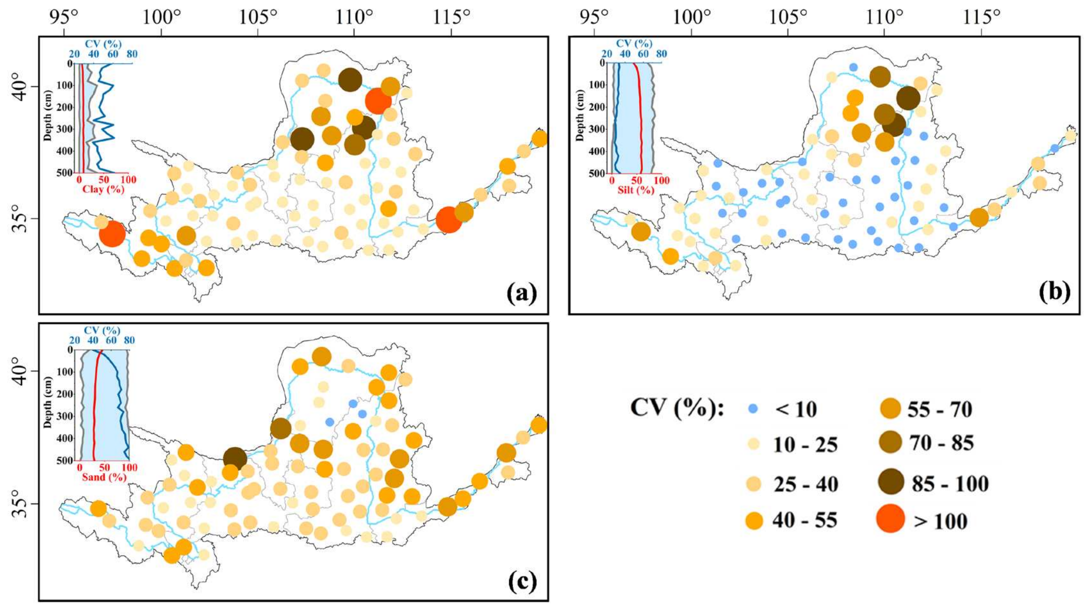

Figure S11. Spatial distribution of the vertical coefficient of variation (CV) of the soil factors in the Yellow River Basin (YRB). Specifically, clay content (a), silt content(b), and sand content (c) from the 0–500 cm profile at each sampling sites. The size of the dots represents the magnitude of the CV, where CV less than 10% indicates weak variability, 10% to 100% indicates moderate variability, and greater than 100% indicates strong variability. The inset figures in (a), (b), and (c) describe the vertical distribution of the horizontal mean value (red line) and CV (blue line) of all sampling sites in the YRB and the blue shaded area represents the data distribution range.

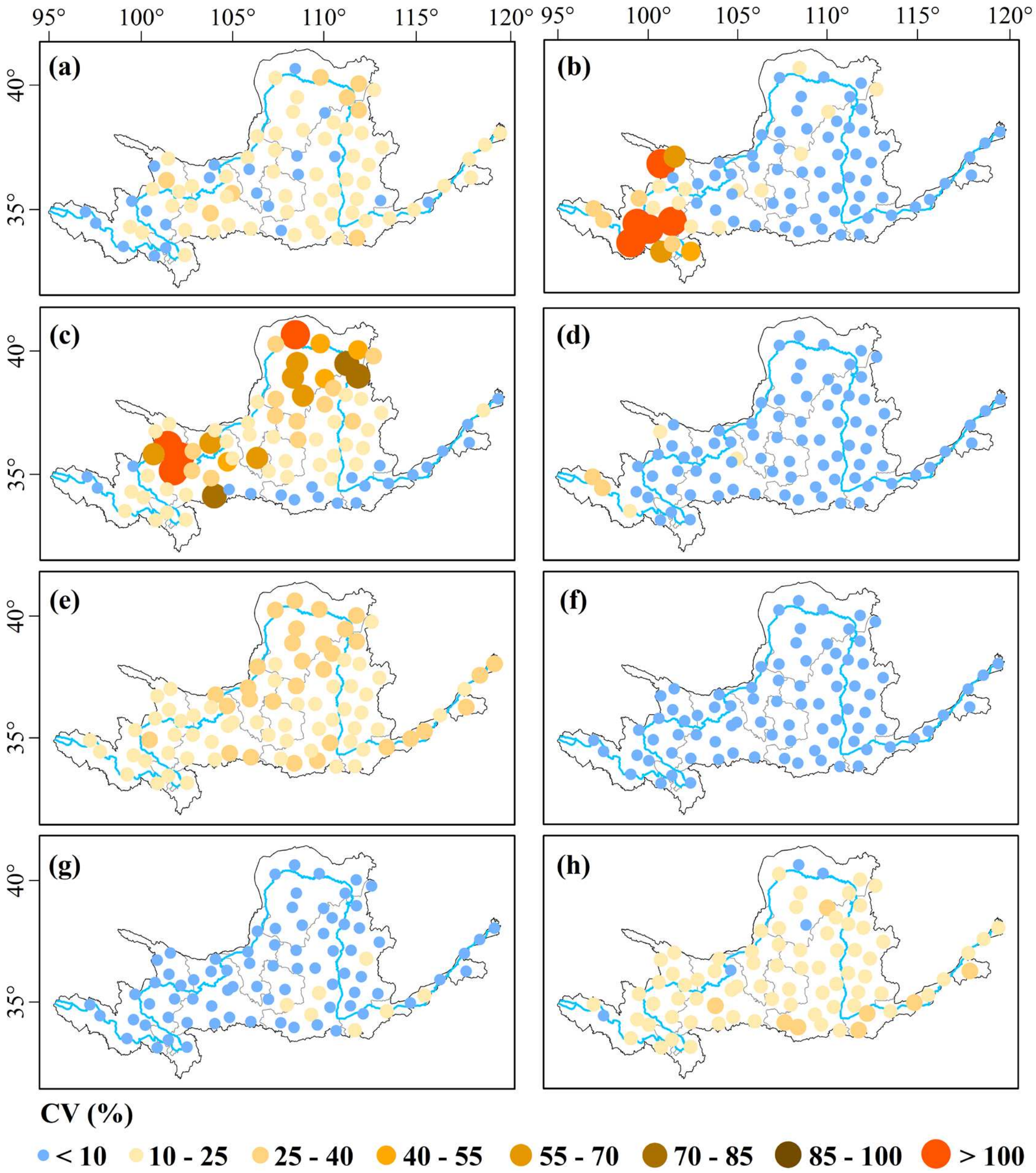

Figure S12. Spatial distributions of the interannual coefficient of variation (CV) of the environment factors in the Yellow River Basin (YRB). Specifically, the mean wind speed (a), mean temperature (b), mean minimum temperature (c), mean maximum temperature (d), mean annual precipitation (e), mean relative humidity (f), mean annual sunshine duration (g), and the normalized difference vegetation index (h). The size of the dots represents the magnitude of the CV, where CV less than 10% indicates weak variability, 10% to 100% indicates moderate variability, and greater than 100% indicates strong variability.

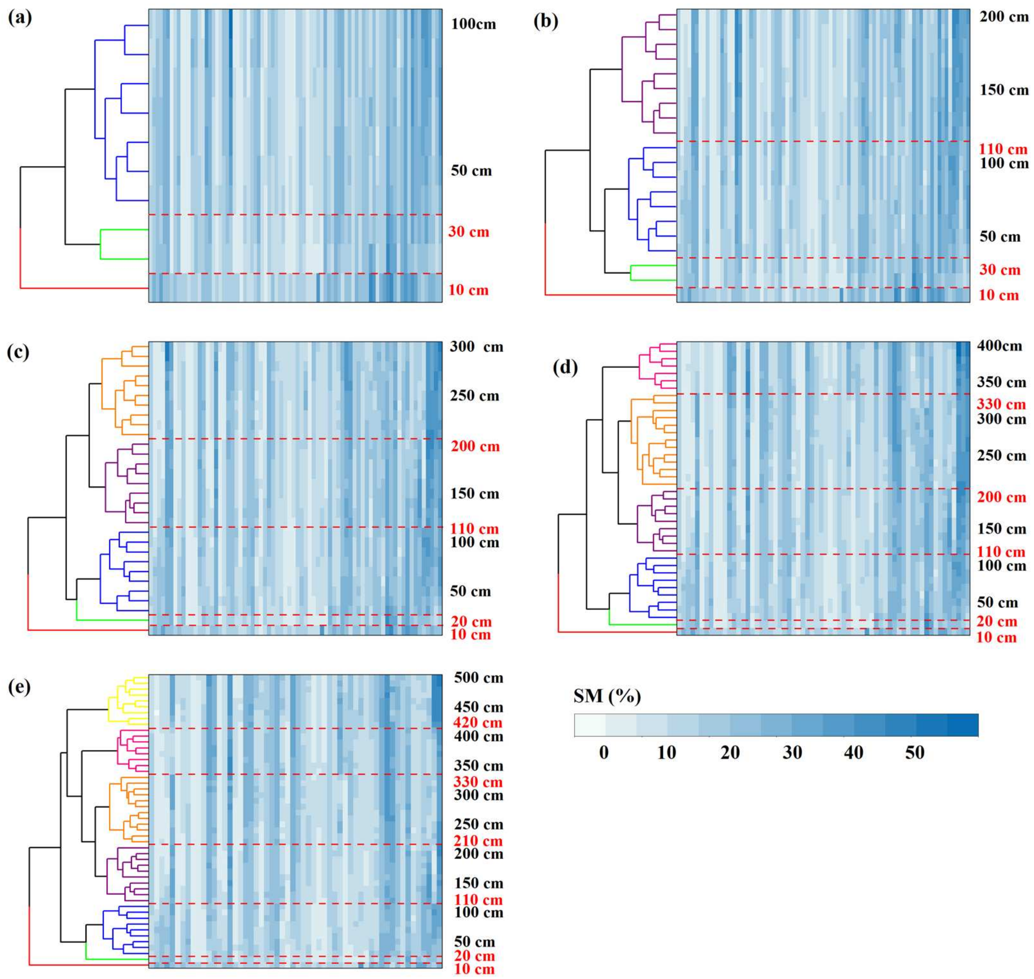

Figure S13. Vertical cluster results of soil moisture (SM) in the Yellow River Basin. The sample sites are divided into five cluster groups according to the sampling depth, including: 0–100 cm (a), 0–200 cm (b), 0–300 cm (c), 0–400 cm (d), and 0–500 cm (e). The cluster groups with larger depth ranges contained more sites (and SM data). The red dashed lines represent the locations of the SM characteristic layers identified in different cluster groups.

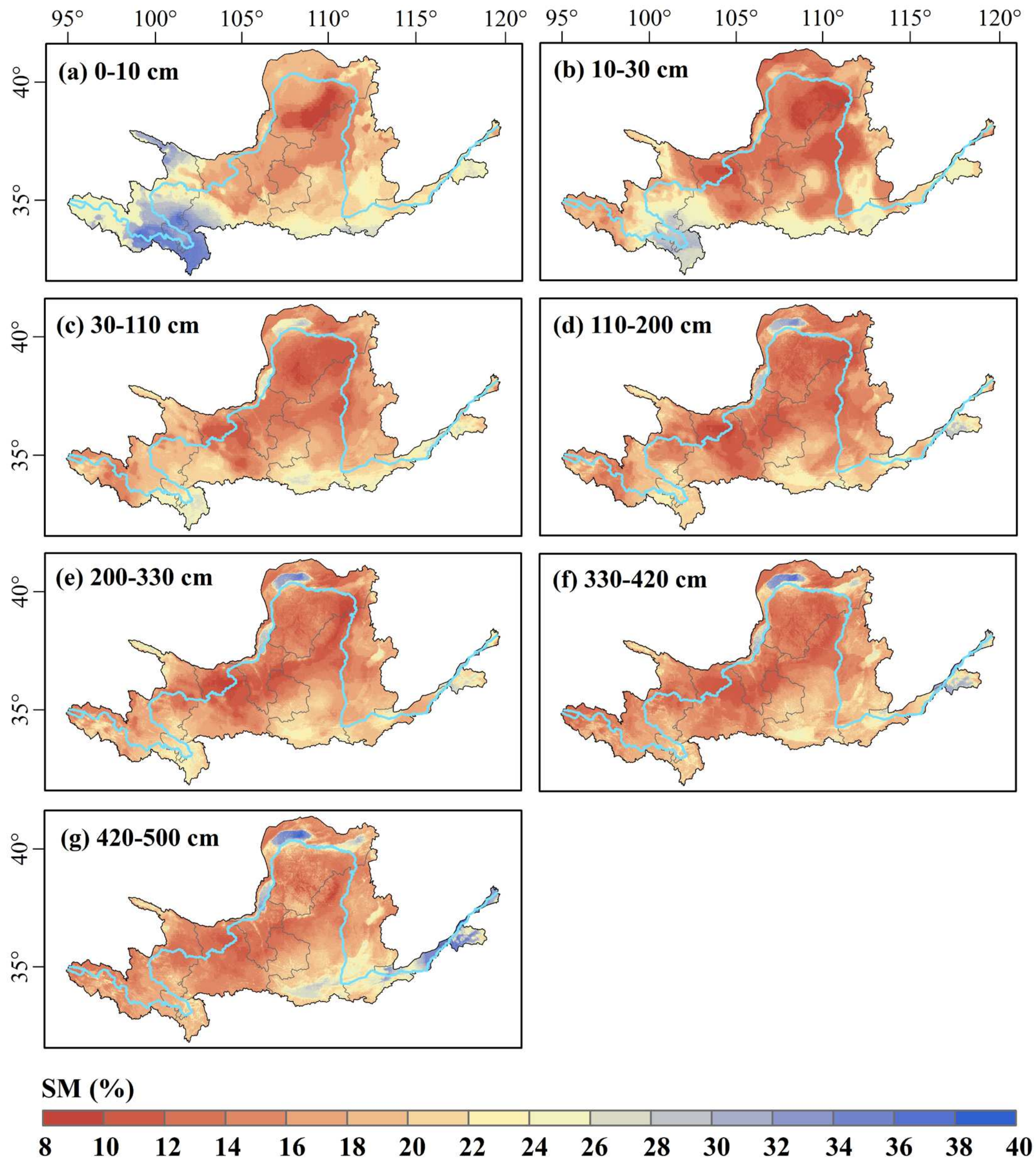

Figure S14. Spatial distributions of soil moisture (SM) in the seven SM characteristic layers of the Yellow River Basin. The maps were developed using the random forest model.

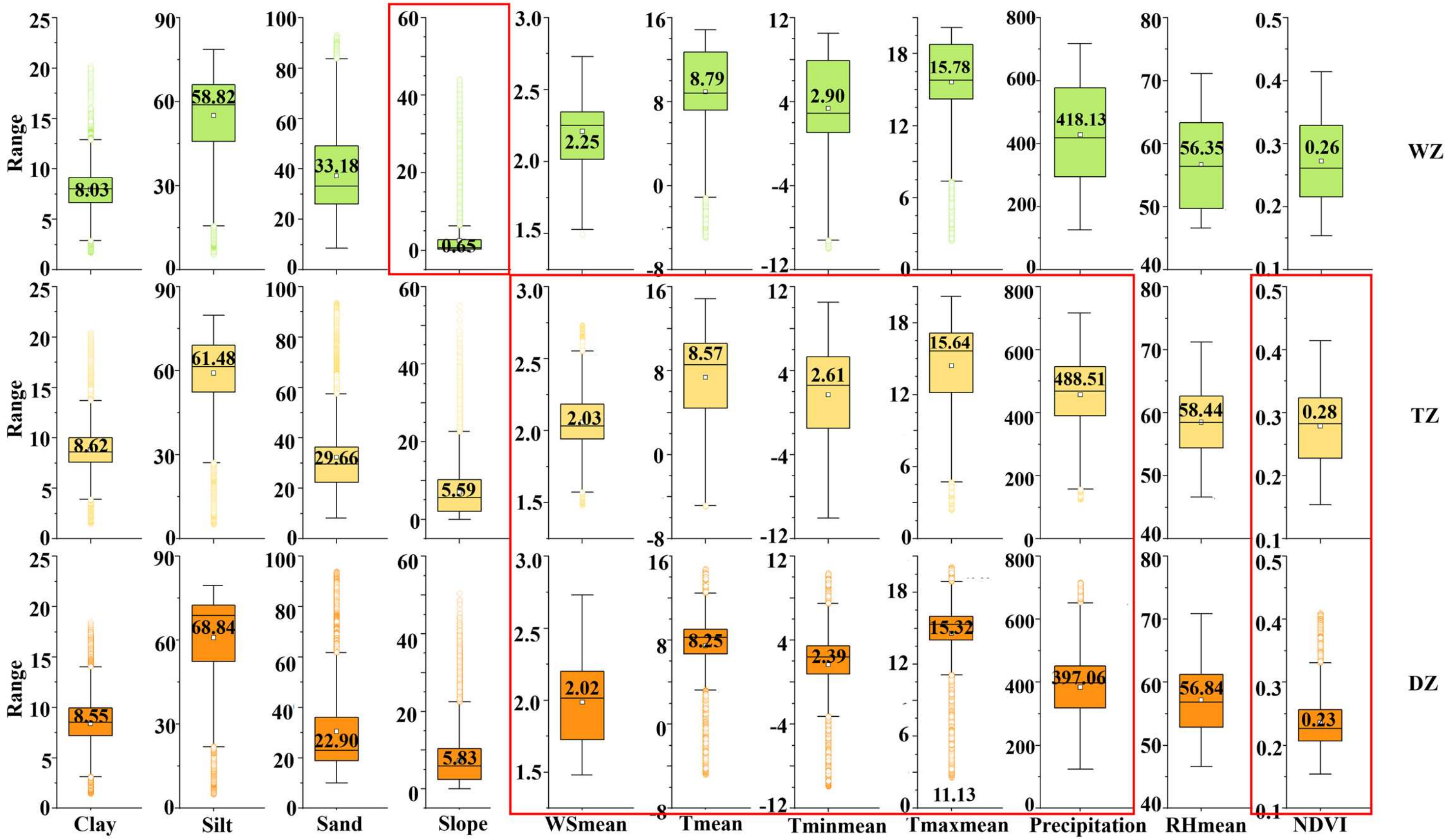

Figure S15. Box plot of distribution range of environmental factor data in different soil moisture (SM) security zones. The box plots were generated for each security zones from 110–200 cm, 200–330 cm, 330–420 cm and 420–500 cm. WZ, TZ, and DZ represents the wet, transitional, and dry zone, respectively. WSmean is mean wind speed; Tmean is mean temperature; Tminmean is mean minimum temperature; Tmaxmean is mean maximum temperature; Precipitation is mean annual precipitation; RHmean is mean relative humidity; and NDVI is normalized difference vegetation index. It's important to note that the data corresponding to the box plots did not account for localized effect.

Table S1. The data sources and calculation methods for field capacity (FC) and permanent wilting point (PWP) in soil moisture (SM) characteristic layers. The original data are derived from published datasets (Dai et al., 2013). This dataset comprises seven layers, ranging from 0–4.50 cm, 4.50–9.10 cm, 9.10–16.60 cm, 16.60–28.90 cm, 28.90–49.30 cm, 49.30–82.90 cm, and 82.90–138.30 cm. Given the depth range in the original dataset, for this study, when computing and interpolating FC and PWP within each SM characteristic layer, data from the nearest available layer in the original dataset was chosen for analysis.

SM characteristic layer (cm)	Original dataset layer (cm)	Calculation method	Reference
0–10	0–4.50 & 4.50–9.10	average	Dai et al. (2013)
10–30	9.10–16.60 & 16.60–28.90	average	
30–110	28.90–49.30 & 49.30–82.90	average	
110–200	82.90–138.30	original	
200–330	82.90–138.30	original	
330–420	82.90–138.30	original	
420–500	82.90–138.30	original	

Table S2. The averaged interannual coefficient of variation (CV) of the meteorological factors across all sampling sites in the Yellow River Basin.

Items	WSmean (m/s)	Tmean (℃)	Tminmean (℃)	Tmaxmean (℃)	Precipitation (mm)	RHmean (%)
CV (%)	3.56	7.49	23.18	4.62	11.39	3.49

Note: WSmean is mean wind speed; Tmean is mean temperature; Tminmean is mean minimum temperature; Tmaxmean is mean maximum temperature; Precipitation is mean annual precipitation; RHmean is mean relative humidity; meteorological factors in this table are factors that has been pre-screened by principal component analysis.