land

MDPI

Article

Deciphering Spatiotemporal Patterns and Drivers of Surface Soil Moisture in Gannan Prefecture (2000–2022) Using Interpretable Machine Learning

Xuhu Wang [1,*], Jianhao Chen [2], Xiaowei Zhang [1], Furong Niu [1], Xiaolei Zhou [1], Weibo Du [1] and Songsong Lu [1]

[1] College of Forestry, Gansu Agricultural University, Lanzhou 730070, China; zhangxw@gsau.edu.cn (X.Z.); zhouxl@gsau.edu.cn (X.Z.); 13843@gsau.edu.cn (W.D.); luss@gsau.edu.cn (S.L.)
[2] State Key Laboratory of Soil and Water Conservation and Desertification Control, College of Soil and Water Conservation Science and Engineering, Northwest A&F University, Yangling 712100, China
* Correspondence: wangxuh@gsau.edu.cn; Tel.: +86-188-0931-5897

Abstract

As a critical alpine transition zone linking the Qinghai–Tibet Plateau and the Loess Plateau, Gannan Prefecture acts as an important water conservation area in the upper Yellow River basin of China. Based on GLDAS-2.1 surface soil moisture (SSM) datasets spanning 2000–2022 and interpretable machine learning tools (SHAP and ALE), this paper analyzes the spatiotemporal evolution, future trend sustainability, and nonlinear statistical associations between environmental predictors and SSM. The main results were as follows: (1) SSM exhibited a significant upward trend with an annual growth rate of 0.18 $kg \cdot m^{-2} \cdot a^{-1}$ ($p < 0.001$), and an abrupt turning point occurred in 2017. The spatial pattern of high SSM in the southeast and low SSM in the northwest remained relatively stable, with the centroid migration distance being less than 1.81 km; most regions presented statistically significant moistening trends ($p < 0.05$). (2) Natural environmental predictors jointly carried 95.79% of the total statistical explanatory weight for modeled SSM variability. Precipitation possessed the highest explanatory proportion (37.93%), followed by temperature (27.30%), potential evapotranspiration (ET_p, 12.26%), elevation (10.44%), and fractional vegetation cover (FVC, 7.77%). One-dimensional ALE curves identified sample-limited statistical breakpoints: SSM gradually plateaued when precipitation reached 650–700 mm, while modeled SSM decreased substantially once ET_p exceeded 800 $mm \cdot a^{-1}$. Two-dimensional ALE further characterized combined statistical correlations among precipitation, temperature, and ET_p. Model outputs also indicated that FVC above 0.45 corresponded to enhanced soil water retention within the observed sample range, which only reflects statistical patterns captured in this dataset rather than universal regulatory standards. This study offers quantitative statistical understanding of SSM variations across alpine transition zones.

Keywords: Qinghai–Tibet Plateau; Gannan Prefecture; water conservation; SSM; nonlinear driving mechanisms; SHAP-ALE

Check for updates

Academic Editor: Xun Wang

Received: 7 May 2026
Revised: 1 July 2026
Accepted: 2 July 2026
Published: 5 July 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

As a critical interface between terrestrial, ecological and hydrological processes, SSM exerts profound controls on evapotranspiration, runoff generation, and land–atmosphere climate feedbacks [1–3]. Acting as a key hub linking precipitation and surface runoff, SSM

regulates water and energy exchanges across the soil–atmosphere boundary and modulates regional precipitation recycling, thereby shaping climate patterns at both regional and global scales [4]. SSM also mediates climate–vegetation interactions and strongly affects plant growth, ecosystem productivity, and carbon–water coupling dynamics [5]. Recent studies highlight that interannual SSM variability drives approximately 90% of the fluctuations in the global terrestrial carbon sink, underscoring its fundamental role in biogeochemical cycling and ecohydrological equilibrium [6]. Moreover, land–atmosphere coupling reinforces positive feedback loops: depleted SSM suppresses evapotranspiration and intensifies atmospheric aridity, increasing the frequency of heatwaves and drought extremes [7]. Therefore, quantifying the spatiotemporal dynamics of SSM and identifying its underlying drivers is essential for advancing the understanding of water cycling and climate system responses.

Under global warming, SSM exhibits substantial spatial heterogeneity worldwide [8]. A growing body of evidence indicates widespread drying trends in global SSM over recent decades, particularly in historically humid regions where elevated evaporative demand accelerates moisture loss [9]. While climatic factors are widely recognized as primary drivers [3,9], anthropogenic activities also exert non-negligible impacts on long-term SSM variations [10]. Attribution studies suggest that greenhouse gas emissions and related human pressures have significantly exacerbated topsoil drying in mid-latitude regions during summer and autumn [11]. Benefiting from advances in remote sensing and reanalysis products, integrated multi-source observations now enable robust investigations into the climatic and anthropogenic controls on SSM dynamics across regional scales.

Substantial efforts have been devoted to characterizing SSM dynamics using ground measurements, satellite-based microwave retrievals, and model assimilation products [12]. A suite of analytical methods—including trend detection, breakpoint identification, wavelet analysis, Hurst exponent forecasting, and machine learning—has been widely applied to quantify SSM trends and driver contributions [10,11]. Globally, SSM displays divergent trends: areas with significant drying and wetting are roughly balanced over the past four decades, with topsoil drying more pronounced than deep-soil moistening. Significantly dried regions are concentrated in densely populated arid and semi-arid zones, while moistening dominates high-latitude, alpine, and tropical humid areas [13]. Enhanced evaporative demand driven by rising vapor pressure deficit under warming has been identified as a major cause of widespread topsoil desiccation [14]. Climatic forcing dominates long-term SSM evolution [11], whereas intensive irrigation can locally offset climatic drying in agricultural regions [13]. However, inconsistencies remain among SSM products, including discrepancies between remote sensing retrievals and land surface simulations, divergent trends between topsoil and root-zone moisture, and coarse spatial resolution in large-scale products. These uncertainties call for the cautious interpretation of long-term trends and climate sensitivity [15]. Although global studies have outlined general drying–wetting patterns, fine-grained driver attribution and localized process mechanisms remain understudied, especially in transitional mountain zones.

In China, marked disparities in SSM dynamics exist between the Qinghai–Tibet Plateau (QTP) and the Loess Plateau (LP). The QTP exhibits strong spatial heterogeneity: western alpine zones experience soil drying due to atmospheric aridification, while eastern sectors show moistening driven by increasing precipitation [16]. Recurrent droughts have triggered grassland degradation across the QTP, with approximately 40% of the plateau exhibiting growing-season SSM moistening despite an overall declining trend during 2000–2020 [17]. In contrast, the LP is a semi-arid drought-prone region with severe historical soil erosion and ecological degradation, creating an urgent demand for SSM research [18,19]. Large-scale ecological restoration programs have substantially increased vegetation greening, yet

enhanced transpiration has exacerbated deep-soil water depletion [14,20]. Up to 87.8% of the spatial variability in evapotranspiration trends on the LP is driven by vegetation greening, with precipitation playing a secondary role [21]. Notably, most existing studies focus on either the QTP or the LP independently, leaving their transitional zones severely understudied. As a typical alpine transition zone between the two plateaus and a headwater source of the Yellow River, Gannan Prefecture serves as a critical ecological barrier and water conservation base [22]. However, the long-term evolution of SSM and its response to climatic and anthropogenic forcing in this region remain poorly characterized, creating a critical knowledge gap regarding regional ecohydrological mechanisms.

Given that previous studies have focused predominantly on the Qinghai–Tibet Plateau and the Loess Plateau in isolation, the transitional zone between them—Gannan Prefecture—has remained severely understudied, particularly regarding the long-term evolution of SSM and its response to climatic and anthropogenic drivers. This knowledge gap hinders our understanding of ecohydrological processes in this critical alpine transition zone, which also serves as a key water conservation area in the upper Yellow River basin. To address this gap, the present study aims to answer two overarching questions: (1) How has SSM evolved in terms of trend magnitude, abrupt changes, and spatial pattern stability over the past two decades? (2) What are the relative statistical explanatory weights of climatic–topographic predictors and anthropogenic predictors for SSM variability, and what nonlinear statistical patterns and joint correlations exist between these predictors and SSM? To answer these questions, we quantify the spatiotemporal dynamics of SSM in Gannan Prefecture from 2000 to 2022 using trend analysis, abrupt change detection, centroid migration, and wavelet analysis. Furthermore, we employ interpretable machine learning tools (SHAP and ALE) to quantify the relative statistical explanatory proportion of natural and anthropogenic predictors, extract sample-limited statistical breakpoints of nonlinear responses, and characterize joint statistical correlations between variables. This study delivers a systematic quantitative analysis of SSM spatiotemporal variation and its correlated environmental predictors across this alpine transition zone, providing descriptive statistical insights into regional alpine soil moisture dynamics.

2. Materials and Methods

2.1. Study Area

Gannan Prefecture is located in the southwestern part of Gansu Province, China, spanning 100°45′45″–104°45′30″ E and 33°06′30″–35°34′00″ N (Figure 1). It lies at the critical transitional zone between the northeastern Qinghai–Tibet Plateau and the northwestern Loess Plateau. The region is characterized by a cold, dry plateau continental climate, with an average elevation of approximately 3000 m, annual mean temperature of 1–3 °C, and annual precipitation of 400–800 mm. Precipitation is concentrated in the warm seasons and shows strong spatiotemporal heterogeneity. The terrain slopes from west to east, with elevation ranging from 1200 m to 4700 m, forming a typical alpine transition zone with distinct hydrothermal gradients and ecological zonation [5]. As a key headwater and recharge region of the upper Yellow River, Gannan serves as an irreplaceable ecological barrier and water conservation zone connecting the Qinghai–Tibet Plateau and Loess Plateau [22].

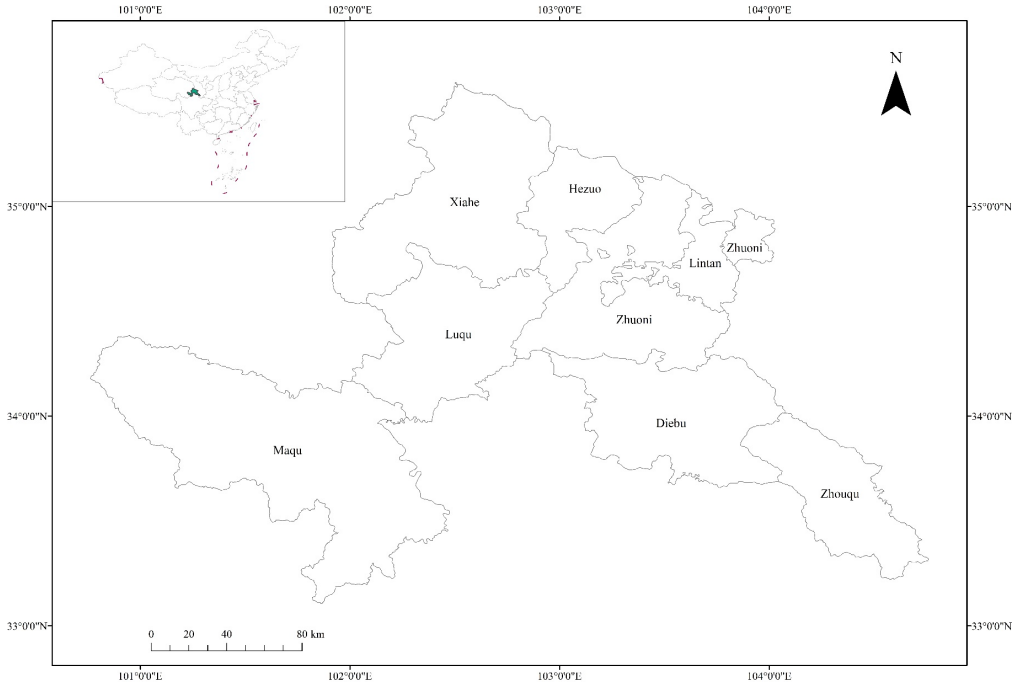

Figure 1. Location of the study area (Gannan Prefecture).

2.2. Data Sources and Preprocessing

Monthly SSM in the 0–10 cm layer during 2000–2022 was obtained from the GLDAS-2.1 dataset [23]. The original spatial resolution of this product is 0.25° (approximately 27.8 km). Previous in situ validation in the Western Sichuan Plateau [24], an adjacent alpine region with similar terrain and climate conditions, has proven that GLDAS series data performs well for 0–10 cm surface soil moisture simulation. It shows an extremely significant correlation with ground observations ($p < 0.01$) and captures temporal variations in soil moisture driven by precipitation reliably.

Seven environmental and anthropogenic variables were selected as driving factors: precipitation (PRE), potential evapotranspiration (ET_p), fractional vegetation cover (FVC), air temperature (TEM), digital elevation model (DEM), population density (LS), and land cover type (CLCD). Among anthropogenic factors, population density (LS) reflects the intensity of human activities such as residential construction and agricultural production, while land cover type (CLCD) represents long-term land use changes driven by ecological policies and grazing management. The data sources and spatial resolutions are listed in Table 1.

Table 1. Data sources and specifications.

Type	Factors	Spatial Resolution	Data Sources
Natural factors	ET_p	1 km	https://data.tpdc.ac.cn/
	TEM	1 km	https://data.tpdc.ac.cn/
	DEM	500 m	https://www.gebco.net/
	FVC	250 m	https://data.tpdc.ac.cn/
	PRE	1 km	https://data.tpdc.ac.cn/
Human factors	CLCD	30 m	https://zenodo.org/
	LS	1 km	https://landscan.ornl.gov/

All data were preprocessed into a unified spatial–temporal framework aligned with the native GLDAS resolution of 0.25°. First, all auxiliary driving variables originally at a 1 km resolution were aggregated to the 0.25° grid of GLDAS-2.1, via zonal averaging within each coarse pixel, to eliminate spatial scale mismatch between SSM and the predictors.

This aggregation operation preserves the native spatial information of GLDAS without generating artificial fine-scale values through upsampling. All rasters were then clipped to the administrative boundary of Gannan Prefecture. Missing values were filled using spatial interpolation, and outliers were detected and removed using a threshold-based method. Finally, monthly time series were aggregated to annual averages to guarantee consistent interannual comparisons across all variables at the unified 0.25° spatial scale.

All auxiliary predictors were aggregated to the native 0.25° grid of GLDAS-2.1 to build a unified-resolution dataset for modeling and trend diagnosis. Accordingly, all statistical analyses in this study were performed on native 0.25° GLDAS-2.1 grids, and all quantitative results and significance evaluations originate solely from these original coarse-resolution datasets. Direct visualization with raw 0.25° grids yields blocky, fragmented maps that obscure clear recognition of regional SSM spatial patterns. To improve figure readability, the classified raster generated from native grids was resampled to 1 km using the nearest-neighbor method only for plotting. Nearest-neighbor resampling simply subdivides original grid cells and does not create new values or modify the trend and significance classes of each 0.25° pixel. The 1 km raster serves purely as a visual base layer and is never adopted for any statistical calculation or quantitative interpretation.

2.3. Research Methods

2.3.1. Mann–Kendall Trend Test and Sen's Slope Estimator

The Mann–Kendall (MK) test is a nonparametric method for detecting monotonic trends in time series. It is robust to non-normal distributions and outliers and is widely used in hydroclimatic studies [25]. Sen's Slope estimator calculates the median slope of all pairwise data points to quantify the trend magnitude [26]. In this study, the MK test was used to determine the significance of SSM trends, and Sen's Slope was applied to estimate the rate of change. The calculation formula for the test statistic S is as follows:

$$\mathbf{S} = \sum_{j=1}^{n-1} \sum_{i=j+1}^{n} \operatorname{sgn}(x_j - x_i) \tag{1}$$

In Formula (1), "n" represents the number of samples; "x_j" and "x_i" respectively denote the values of the sequence in the j-th year and the i-th year (where j > i). The function $\operatorname{sgn}(x_j - x_i)$ is an indicator function that takes a value based on the sign of $x_j - x_i$. The calculation formula is as follows:

$$\operatorname{sgn}(x_j - x_i) = \begin{cases} 1, x_j - x_i > 0 \\ 0, x_j - x_i = 0 \\ -1, x_j - x_i < 0 \end{cases} \tag{2}$$

S approximately follows a normal distribution with a mean of 0, and the variance S is calculated as follows:

$$\operatorname{Var}(S) = \frac{n(n-1)(2n+5)}{18} \tag{3}$$

When n > 10, the standard normal statistical variable begins to be calculated using formula Z, as follows:

$$Z = \begin{cases} \frac{S-1}{\sqrt{\operatorname{Var}(S)}}, S > 0 \\ 0, \quad S = 0 \\ \frac{S+1}{\sqrt{\operatorname{Var}(S)}}, S < 0 \end{cases} \tag{4}$$

In this context, formula Z represents the statistic of the standardized test. If Z > 0, it indicates that the overall SSM is on an upward trend; conversely, it signifies a downward

trend. n denotes the number of time series data points, which in this case is 23. This article employs a 95% confidence level test, meaning that $Z > 1.96$ signifies a significant upward trend, while $Z < -1.96$ indicates a significant downward trend.

The formula for calculating the median slope estimated by Sen's Slope is as follows:

$$\beta = \text{Median}\left(\frac{x_j - x_i}{t_j - t_i}\right) \quad (5)$$

In this context, the initial formula Beta represents the median slope of all point pairs. If the initial formula Beta is greater than 0, it indicates an increasing trend in SSM; conversely, it indicates a decreasing trend in SSM. The capitalized X in the initial formula represents the estimated trend value of SSM. Subscript j of X and subscript i of X represent the SSM values of Gannan Prefecture for the initial formula j year and the initial formula i year, respectively. Similarly, subscript j of t and subscript i of t represent the initial formula j year and the initial formula i year, respectively.

Prior to pixel-wise trend detection, trend-free pre-whitening (TFPW) was adopted to eliminate temporal autocorrelation in each SSM time series. The strict Benjamini–Hochberg False Discovery Rate (FDR) correction with backward monotonic constraint was applied to all raw *p*-values of valid 0.25° GLDAS pixels, and statistical significance was identified when the FDR-adjusted q-value < 0.05. Full pixel-level MK Z statistics, raw *p*-values, FDR-adjusted q-values, and corresponding significance labels for all 58 native coarse grid cells are listed in Supplementary Table S1 for a transparent record of the trend test.

2.3.2. Abrupt Change Detection

Two methods were used to identify abrupt changes:

Moving *t*-test: Detects significant mean shifts between two adjacent subseries within a sliding window (set to 3 years in this study) [27].

Pettitt test: A rank-based nonparametric method for detecting a single change point, which is robust to non-normality and outliers [28].

Consistent results from both tests were used to confirm the occurrence and timing of abrupt changes in SSM.

2.3.3. Centroid Migration Model

The centroid migration model quantifies the spatiotemporal shift in the geographic center of SSM by treating pixel values as weights [29]. Annual centroids were calculated to characterize shifts in position, direction, and distance, revealing the stability and evolution of SSM spatial patterns. The calculation formula is as follows:

$$X_t = \frac{\sum_{i=1}^{n} P_{it} X_{it}}{\sum_{i=1}^{n} P_{it}} \quad (6)$$

$$Y_t = \frac{\sum_{i=1}^{n} P_{it} Y_{it}}{\sum_{i=1}^{n} P_{it}} \quad (7)$$

In Formulas (6) and (7), X_t and Y_t represent the longitude and latitude coordinates of the centroid of the SSM distribution, respectively; P_{it} denotes the SSM content of the i-th pixel in year t; and X_{it} and Y_{it} represent the longitude and latitude coordinates of the centroid of the SSM of the i-th pixel in year t, respectively.

The formula for calculating the SSM centroid shift distance is as follows:

$$D_{s\text{-}k} = C\sqrt{(Y_s - Y_k)^2 + (X_s - X_k)^2} \quad (8)$$

where S and k represent the start and end time of the migration, respectively; Ds-k denotes the migration distance of the SSM centroid during this time period; (X_s, Y_s) and (X_k, Y_k) are the spatial geographical coordinates of the SSM centroid for the corresponding time period; C is a constant of 111.111, which is the coefficient for converting geographical coordinates (1°) to planar distance (km).

2.3.4. Wavelet Analysis

The Morlet wavelet was used for continuous wavelet transform to decompose the SSM time series into time–frequency space and identify dominant periodicities [30]. This method captures both temporal and frequency characteristics and reveals multi-scale oscillatory behavior.

2.3.5. Machine Learning Modeling

Seven machine learning models were compared: Gradient Boosting Decision Tree (GBDT), K-Nearest Neighbor (KNN), Kernel Ridge Regression (KRR), Multi-Layer Perceptron (MLP), Random Forest (RF), Support Vector Regression (SVR), and eXtreme Gradient Boosting (XGB) [31], to select the optimal nonlinear modeling tool for SSM driving factors.

All datasets for modeling, including SSM and environmental predictors, shared a consistent spatial resolution of 0.25°, covering the study period from 2000 to 2022. After masking water bodies, bare rock, and other invalid land areas within Gannan Prefecture, only 58 valid 0.25° grid pixels were retained. All covariates were uniformly clipped to the study boundary, resampled to the unified 0.25° grid, screened to eliminate cloud-contaminated observations, missing values, and extreme outliers. Min–max normalization was applied to all predictor variables prior to modeling to eliminate dimensional differences. Pixel-year paired time series samples were constructed by matching annual SSM and environmental variable records for each valid pixel over the 23-year period. All samples were randomly partitioned into three mutually independent subsets: the training set for model fitting, the validation set for hyperparameter iterative optimization, and the test set dedicated to evaluating within-sample fitting performance.

Grid search combined with inner five-fold random cross-validation was performed exclusively on training subsets to optimize hyperparameters and reduce overfitting risks. Three unified metrics were adopted to assess model fitting performance: coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE). A critical caveat: Random pixel-year sample splitting mixes spatially adjacent grids and consecutive multi-year observations, leading to spatiotemporal leakage and overoptimistic fitting results. Such metrics merely reflect within-sample fitting effects and cannot represent genuine generalization performance for unseen spatial regions or independent years. Spatially blocked cross-validation and leave-one-year-out temporal cross-validation are effective solutions to avoid leakage bias [32].

2.3.6. Model Interpretability

(1) SHapley Additive exPlanations (SHAP)

SHAP quantifies feature contributions based on cooperative game theory, providing consistent and unbiased attribution for each predictor [33]. It was used to identify driver importance, directional effects, and nonlinear impacts on SSM. The calculation formula is as follows:

$$\phi_j = \sum_{S \subseteq N \setminus \{j\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!} [f_x(S \cup \{j\}) - f_x(S)] \quad (9)$$

where N = {1, 2, …, M} is the complete set of M features; S represents any subset of features excluding feature j; $f_x(S)$ denotes the model output under the feature subset S; |S|

represents the number of elements in the set S; ϕ_j represents the SHAP value of feature j, which is the average marginal contribution of this feature to the model f output across all possible combinations of feature subsets.

(2) Accumulated Local Effects (ALE)

ALE curves visualize the marginal effect of each feature on model predictions while minimizing bias from correlated features [33]. ALE was used to identify nonlinear responses and critical thresholds of SSM to key drivers. The calculation formula is as follows:

$$ALE_s = \int_{z_{s-1}}^{z_s} E[f(x)|X_j = z] - E[f(x)]dz \quad (10)$$

where z_s denotes the upper bound of the feature partition interval s, while z_{s-1} represents the lower bound of that interval.

3. Results

3.1. Spatiotemporal Dynamics of SSM in Gannan Prefecture

3.1.1. Temporal Variation and Abrupt Change

As shown in Figure 2, the annual mean SSM in Gannan Prefecture during 2000–2022 was 24.70 kg·m^{-2}, and exhibited a statistically significant increasing trend of 0.18 kg·m^{-2}·a^{-1} ($p < 0.001$). The maximum SSM (28.34 kg·m^{-2}) appeared in 2020, whereas the minimum (21.97 kg·m^{-2}) occurred in 2002. Notably, a distinct turning point was identified around 2017 based on the 3-year moving average: the mean SSM was 23.91 kg·m^{-2} before 2017 but increased substantially to 27.22 kg·m^{-2} afterward. Consistently, both the moving *t*-test and Pettitt test statistically validated an abrupt change in 2017 (Figure 3). Furthermore, wavelet analysis detected two dominant fluctuation timescales (22-year primary and 11-year secondary signals; Figure 4). However, the 23-year limited annual time series cannot verify these as stable multi-decadal periodic cycles. These signals merely reflect short-to-medium oscillating wet–dry variations within the study window, and we avoid definitive extrapolations of long-term moisture evolution based on this periodic feature.

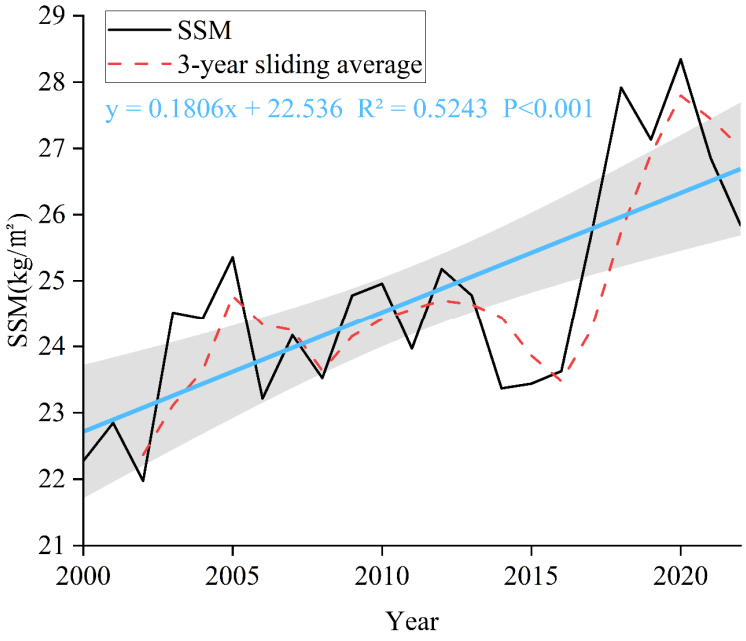

Figure 2. Temporal variations in SSM in Gannan Prefecture during 2000–2022. The gray shaded band denotes the 95% confidence interval of the linear regression trend line for annual SSM variations.

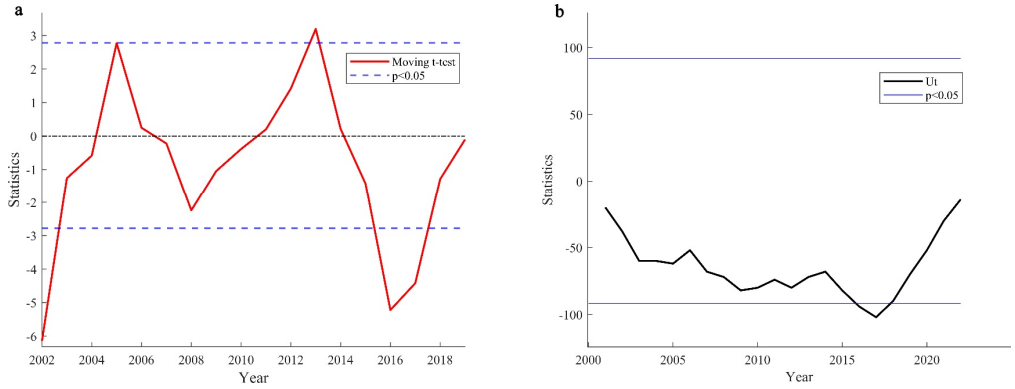

Figure 3. Abrupt change detection of SSM during 2000–2022. (**a**) Moving *t*-test; (**b**) Pettitt test.

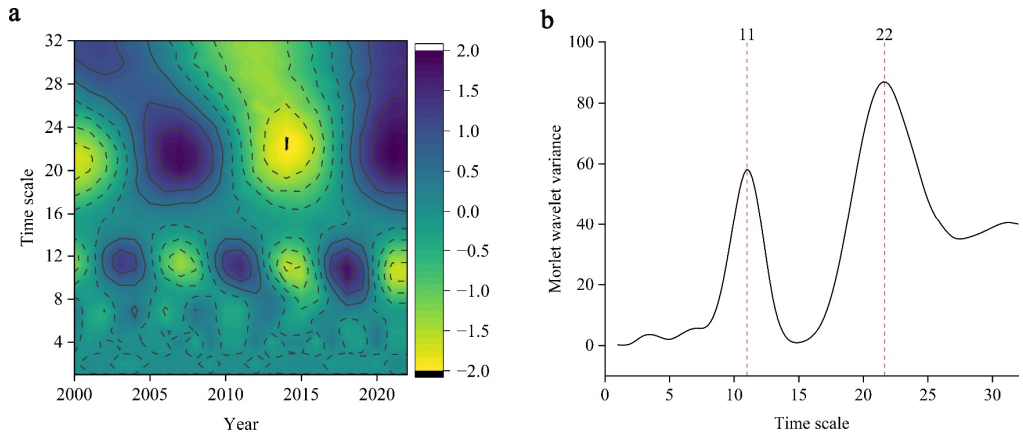

Figure 4. Wavelet analysis of SSM during 2000–2022. (**a**) Contour map of wavelet coefficients; (**b**) wavelet variance.

3.1.2. Spatial Pattern and Temporal Dynamics of SSM

The spatial distribution of multi-year mean SSM exhibited a south–high and north–low pattern (Figure 5), with values ranging from 22 to 25 kg·m^{-2} in the north to 25–27 kg·m^{-2} in the south, which is highly consistent with the spatial pattern of precipitation.

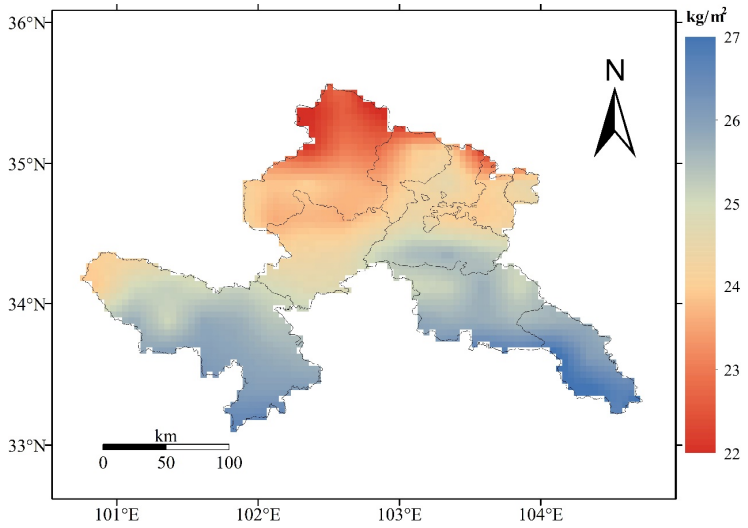

Figure 5. Spatial distribution of multi-year mean SSM across Gannan Prefecture (2000–2022). The 1 km background is resampled from native 0.25° grids via nearest-neighbor interpolation only for plotting, and all quantitative analyses use original GLDAS pixels.

After TFPW preprocessing and strict BH-FDR multiple testing correction based on the metrics in Supplementary Table S1, the results of MK test (Figure 6a) indicated that almost all grids in Gannan Prefecture had a significantly rising SSM trend; only fragmented small areas showed non-significant SSM variations. Sen's slope results (Figure 6b) showed annual SSM growth rates ranging −0.01 to 0.27 kg·m^{-2}·a^{-1}. High increasing rates (0.17–0.27 kg·m^{-2}·a^{-1}) were concentrated in central and northeastern Gannan, while western and southeastern regions had weaker rising trends below 0.17 kg·m^{-2}·a^{-1}. Minor western Maqu areas presented near-zero or slight drying signals, corresponding to the non-significant patches in Figure 7a.

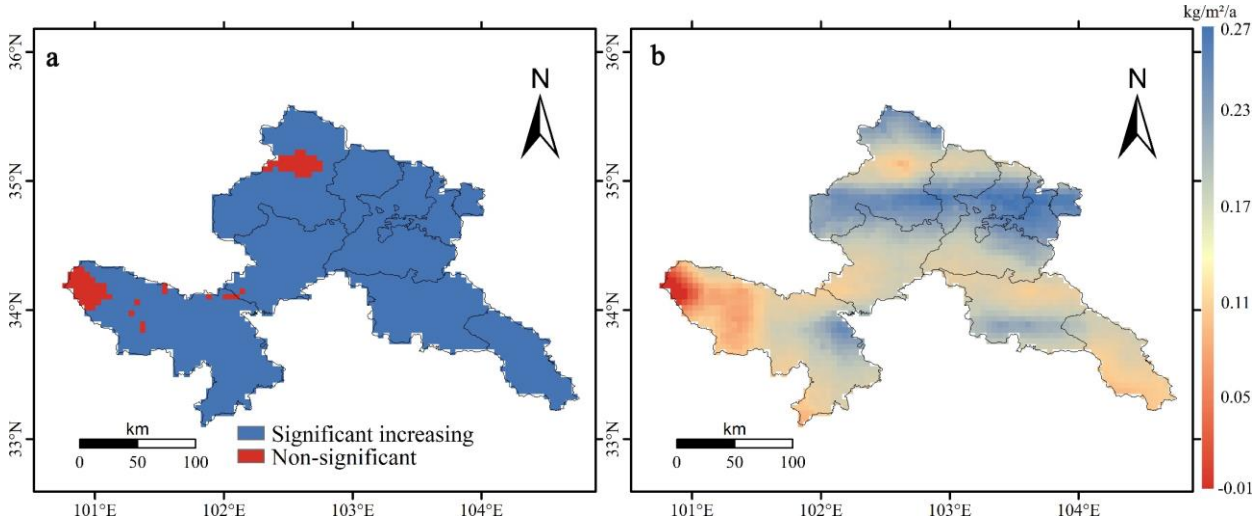

Figure 6. Spatial pattern of annual SSM trends over 2000–2022. (**a**) Spatial distribution of significance for interannual SSM trend; (**b**) Spatial variation of SSM annual trend magnitude. Trend significance classification follows FDR-adjusted q-values in Supplementary Table S1; the 1 km background raster is obtained by nearest-neighbor resampling of native 0.25° GLDAS grids only for visual clarity; all quantitative analyses are based on original coarse pixels.

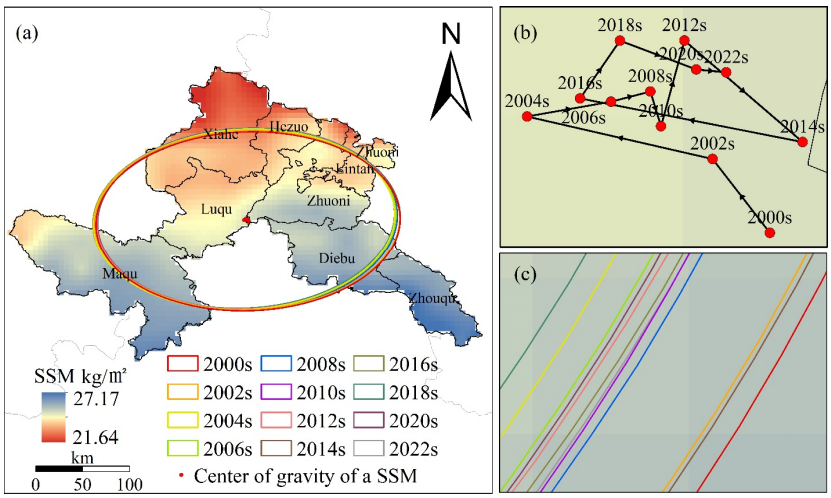

Figure 7. Centroid shift in SSM in Gannan Prefecture during 2000–2022. (**a**) General centroid distribution; (**b**) centroid migration trajectory; (**c**) directional distribution of spatial dynamics. The 1 km background raster is obtained by nearest-neighbor resampling of native 0.25° GLDAS grids only for visual clarity; all quantitative analyses are based on original coarse pixels.

Regarding spatial dynamic evolution (Figure 7 and Table 2), the SSM centroid remained stably located in the central part of Gannan Prefecture (102.76°–102.78° E, 34.29°–34.31° N)

during 2000–2022, with a migration distance of only 0.45–1.81 km per interval. Meanwhile, the standard deviational ellipse maintained a persistent east–west long axis and north–south short axis, with only minor fluctuations in semi-axis lengths. Collectively, these results demonstrate that the spatial pattern of SSM was stable yet increasingly moistened, without large-scale structural rearrangement.

Table 2. Centroid migration of SSM in Gannan Prefecture during 2000–2022.

Year	2002	2006	2010	2014	2018	2022
Center of gravity coordinate/°	102.77 34.29	102.76 34.30	102.76 34.30	102.78 34.29	102.76 34.31	102.77 34.30
Direction of migration/°	328.81	64.98	351.84	256.7	33.31	76.90
Migration distance/km	1.20	1.28	1.28	1.81	1.00	0.45
Long semi-axis/km	143.50	141.73	141.68	142.73	140.19	140.90
Short semi-axis/km	84.47	85.14	85.37	85.23	84.72	84.79

3.2. Statistical Associations Between Environmental Predictors and SSM

To quantify statistical correlations between environmental variables and SSM across Gannan Prefecture, seven predictors were incorporated into subsequent modeling. Figure 8 presents the spatial distribution of the multi-year average values of the seven variables over 2000–2022. ET_p was spatially higher in the east and lower in the west; DEM increased from east to west; FVC remained high over most regions, with localized low-value zones; PRE was higher in the south and lower in the north; the LS factor showed higher values in the east and lower values in the west; grassland dominated the western part of the study area, while forest covered most eastern zones for CLCD; and TEM was lower in the west and higher in the east.

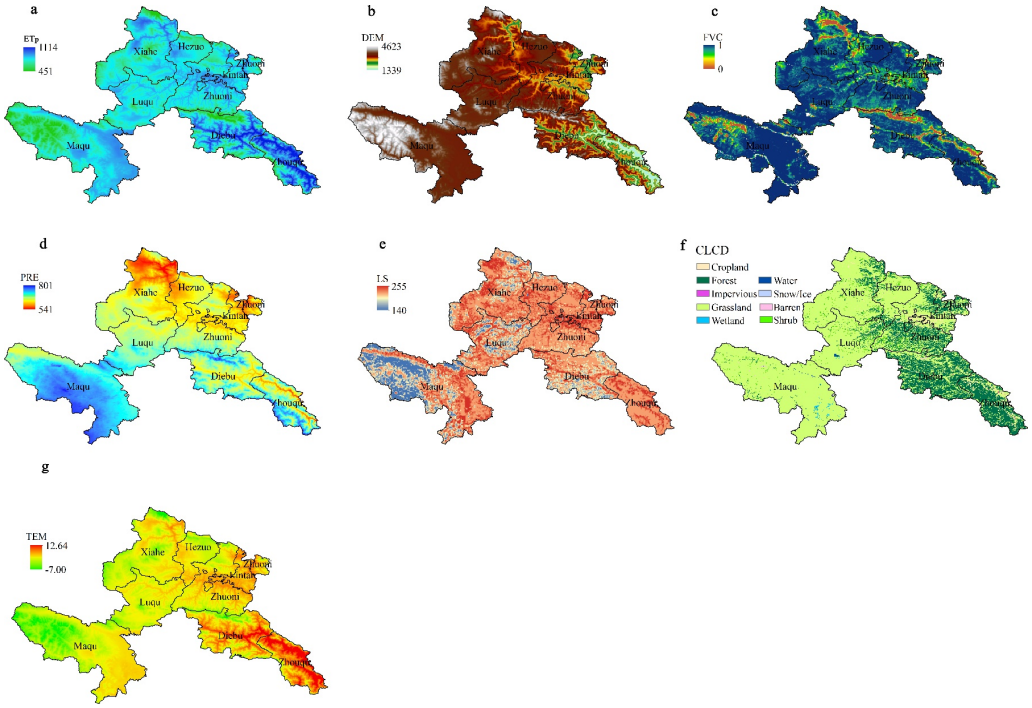

Figure 8. Spatial distribution of driving factors in Gannan Prefecture. (**a**) ET_p; (**b**) DEM; (**c**) FVC; (**d**) PRE; (**e**) LS; (**f**) CLCD; (**g**) TEM.

3.2.1. Selection of Optimal Machine Learning Model for Statistical Correlation Analysis of SSM

To characterize the nonlinear statistical associations between SSM and environmental predictors and quantify the statistical explanatory weight of each variable to SSM variations, we compared seven widely used machine learning algorithms: GBDT, KNN, KRR, MLP, RF, SVR, and XGB. The purpose of model comparison was to identify the optimal model for describing complex nonlinear statistical relationships between variables and SSM, since conventional linear approaches fail to capture such intricate nonlinear patterns. The dataset used for model comparison covers 2000–2022 at 0.25° resolution, with 58 valid grid pixels after masking non-vegetated areas. All pixel-year samples were randomly separated into three subsets for fitting assessment (see Section 2.3.6 for full preprocessing details).

As shown in Table 3, grid search combined with inner five-fold random cross-validation was applied exclusively on training subsets to tune hyperparameters and avoid overfitting. Key tuned parameters included the number of decision trees and maximum tree depth for RF, kernel type, and regularization coefficient for SVR, as well as learning rate and tree depth for GBDT and XGB. Model fitting performance was assessed via three unified metrics: R^2, RMSE, and MAE.

Table 3. Hyperparameter ranges for machine learning algorithms.

Algorithms	Hyperparameters	Search Range	Step
GBDT	n_estimators learning_rate	[100, 200] [0.05, 0.1]	100 0.05
XGB	n_estimators learning_rate	[100, 200] [0.05, 0.1]	100 0.05
RF	n_estimators min_samples_leaf	[100, 200] [1, 5]	100 4
KRR	alpha kernel	[0.1, 1] ['linear', 'rbf']	0.9 -
SVR	C kernel	[10, 100] ['rbf']	90 -
KNN	n_neighbors metric	[5, 7] ['minkowski']	2 -
MLP	hidden_layer_sizes alpha learning_rate_init solver	[(50,), (100,), (100, 50)] [0.0001, 0.001, 0.01] [0.001, 0.01] ['adam']	- 0.0099 0.009 -

Table 4 summarizes the cross-validation fitting performance of all models. R^2 reflected the statistical explanatory power of the model for SSM variability within the randomly split sample set, while RMSE and MAE quantified prediction bias; higher R^2 alongside lower RMSE and MAE indicated better fitting performance on mixed pixel-year samples.

It should be noted that all fitting metrics are derived from randomly mixed pixel-year samples, which may suffer from spatiotemporal leakage, as elaborated on in Section 2.3.6. These values only reflect within-sample fitting effects rather than unbiased generalization performance for unseen regions or years.

The RF algorithm exhibited the most stable fitting performance among all seven models on the randomly partitioned sample set, with an average R^2 of 0.67, average RMSE of 1.21, and average MAE of 0.93. Compared with XGB, GBDT, and KNN, RF delivered balanced and consistent fitting results on the mixed training and test samples from the 23-year dataset covering extreme drought and heavy rainfall events. RF is widely

adopted as a reliable modeling tool for soil moisture simulation across alpine regions of the Tibetan Plateau in previous regional studies [34,35]. For our modeling workflow, this stable performance only holds for randomly mixed samples within the study period and cannot be interpreted as robust predictive capacity across independent spatial blocks or unseen annual time slices.

Table 4. Performance metrics of machine learning models.

Algorithms	Train			Test			Validation		
	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE
GBDT	0.63	1.29	1.01	0.62	1.31	1.02	0.62	1.33	1.04
XGB	0.72	1.13	0.87	0.66	1.24	0.96	0.66	1.25	0.97
RF	0.71	1.14	0.88	0.67	1.21	0.93	0.67	1.23	0.94
KRR	0.64	1.27	0.99	0.63	1.3	1.01	0.63	1.3	1.02
SVR	0.65	1.26	0.94	0.63	1.29	0.98	0.63	1.3	0.99
KNN	0.71	1.14	0.88	0.61	1.32	1.01	0.61	1.33	1.02
MLP	0.65	1.25	0.97	0.64	1.27	0.99	0.65	1.28	0.99

Therefore, the RF model was selected as the optimal candidate for subsequent statistical attribution analysis based on its comparative fitting performance on our sample dataset. Readers should interpret all RF-derived statistical patterns cautiously due to the limitation of the random splitting validation scheme.

3.2.2. SHAP-Based Statistical Explanatory Weight of Environmental Predictors for SSM

The SHAP results (Figure 9) revealed strong statistical correlations between natural environmental predictors and model-predicted SSM, with these variables accounting for a cumulative statistical explanatory weight of 95.79%. This high cumulative proportion only reflects sample-based statistical dependence rather than definitive causal control relationships between climate, terrain, and SSM within the study area.

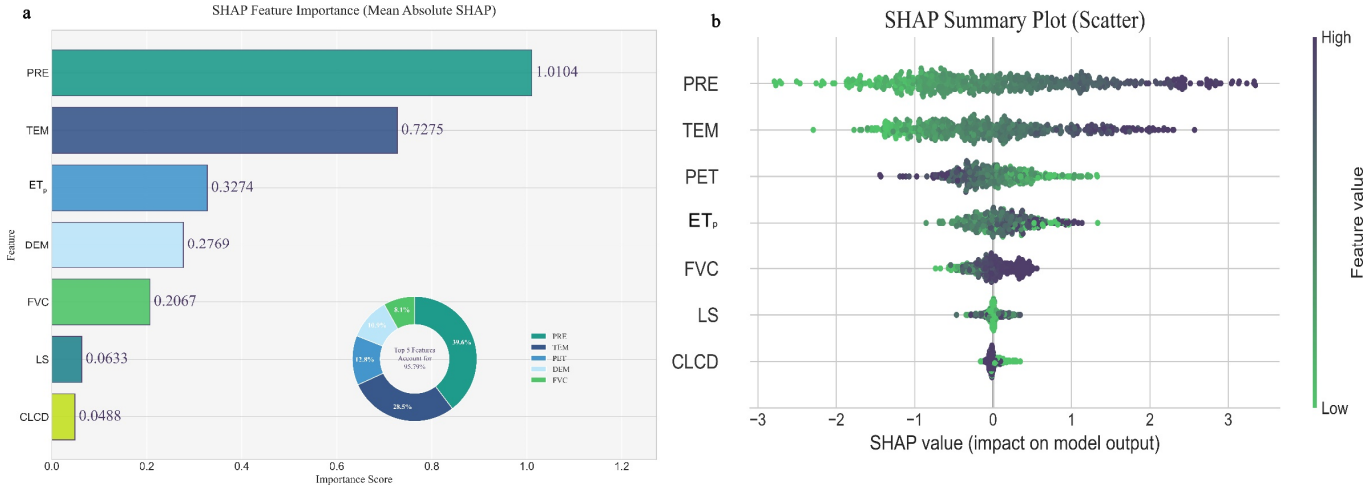

Figure 9. SHAP analysis based on the RF model. (**a**) Feature importance; (**b**) SHAP summary plot.

In terms of SHAP-based explanatory weight, variables were ordered as PRE (37.93%) > TEM (27.30%) > ET_p (12.26%) > DEM (10.44%) > FVC (7.77%). Precipitation exhibited positive statistical associations with SSM, whereas elevated ET_p corresponded to reduced SSM values. TEM, DEM, and FVC displayed nonlinear patterns across SHAP curves: they correlated with lower SSM at low variable values and higher SSM above specific sample-based breakpoints. Such segmented patterns merely describe statistical response features

limited to the observed dataset, rather than universal ecological management thresholds applicable across the entire alpine study region.

Anthropogenic predictors (LS and CLCD) together contributed the remaining 4.21% of the total explanatory weight. Despite their low overall statistical contribution, evident spatial heterogeneity was observed. Regions with intensive human settlement tended to show lower SSM, which may be linked to impervious surface expansion and anthropogenic water consumption [17]. Forest and grassland areas generally corresponded to higher SSM, while built-up and bare land exhibited relatively low soil moisture [36]. Human activities can indirectly shape SSM conditions by modifying vegetation coverage and landscape water retention capacity, as captured by the statistical patterns in our dataset.

3.2.3. ALE-Based Analysis of SSM Nonlinear Response and Critical Thresholds

To further characterize the nonlinear statistical relationships between environmental predictors and model-estimated SSM, we performed ALE analysis (Figure 10). The ALE curves displayed distinct sample-limited statistical patterns for each variable. Model-estimated SSM increased rapidly with rising precipitation and gradually plateaued when precipitation reached 650–700 mm. The positive correlation between temperature and SSM weakened once temperature exceeded 8–9 °C. A notable statistical breakpoint occurred at ET_p = 800 mm·a^{-1}, above which the predicted SSM decreased substantially. Elevation exhibited a V-shaped statistical trend, with minimum SSM values near 3000 m. When FVC surpassed 0.45, the model output indicated elevated soil water retention within the observed sample range. All the segmented patterns above only reflect statistical variation features captured by our dataset and cannot be generalized as unified regulatory standards for regional vegetation and water conservation management.

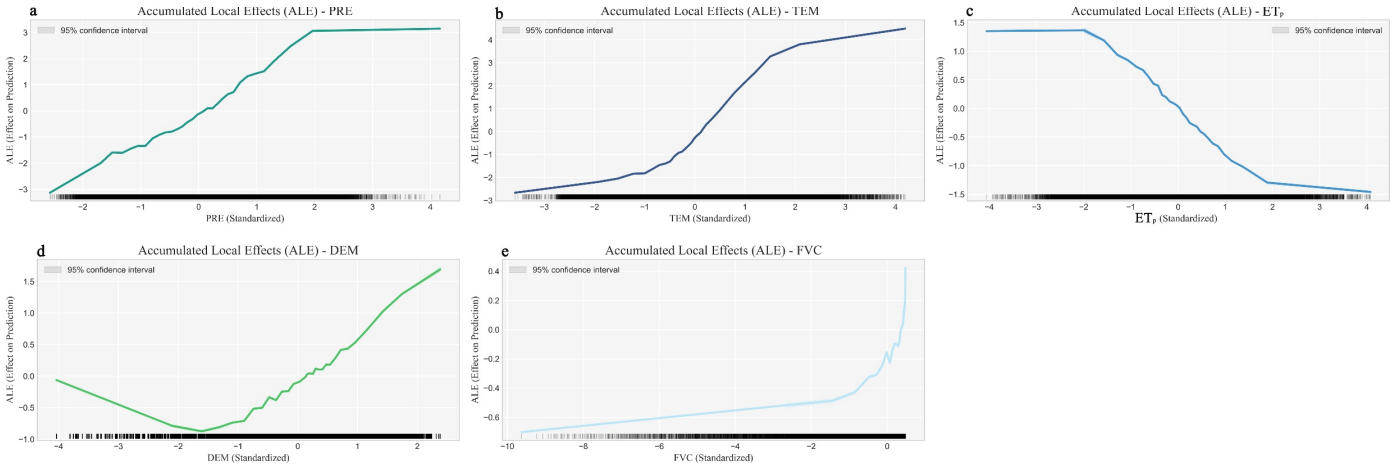

Figure 10. ALE Statistical Response Curves of Main Environmental Predictors Derived from the RF Model. (**a**) PRE; (**b**) TEM; (**c**) $ET_{p;}$ (**d**) DEM; (**e**) FVC.

3.2.4. 2D ALE Statistical Interaction Patterns Between Environmental Predictors and SSM

Building on single-variable ALE results, two-dimensional ALE was used to characterize joint statistical associations between environmental predictors (Figure 11), complementing the outcomes of SHAP and one-dimensional ALE analysis. The results revealed complex nonlinear statistical correlations among variables. For the PRE–TEM combination, higher temperature corresponded to larger model-estimated SSM under low precipitation, while this association weakened when precipitation was abundant. Under low PRE and low ET_p, the model produced the highest SSM values; the statistical link between precipitation and SSM diminished at high ET_p levels. Low temperature counteracted the SSM reduction associated with high ET_p, whereas rising temperature eliminated this offset statistical

pattern. The combination of moderate elevation and low temperature corresponded to the highest SSM, while extreme elevation coupled with high temperature was associated with lower SSM values. All these joint statistical patterns only reflect relationships captured within the sample dataset and do not represent definitive interactive causal mechanisms among environmental variables.

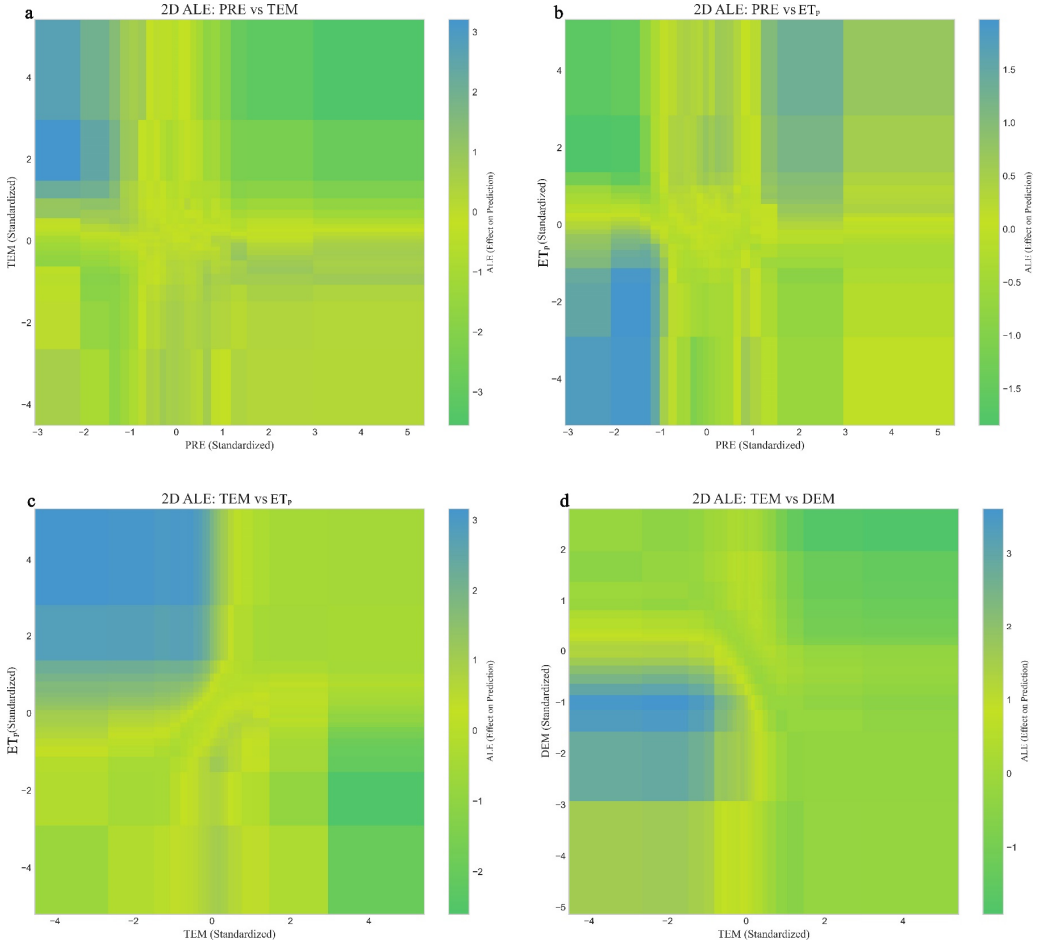

Figure 11. Two-dimensional ALE statistical interaction patterns between environmental predictors. (**a**) 2D ALE statistical correlation: PRE vs. TEM; (**b**) 2D ALE statistical correlation: PRE vs. ET_p; (**c**) 2D ALE statistical correlation: TEM vs. ET_p; (**d**) 2D ALE Statistical Correlation: TEM vs. DEM.

4. Discussion

4.1. Spatiotemporal Patterns and Regional Comparisons of SSM in the Gannan Alpine Transition Zone

From 2000 to 2022, SSM in Gannan Prefecture exhibited a stable southeast–high to northwest–low spatial pattern, with high values concentrated in the eastern and southern humid mountainous areas and low values in the western semi-arid alpine meadows. Temporally, the regional average SSM showed a significant increasing trend (0.18 $kg \cdot m^{-2} \cdot a^{-1}$, $p < 0.001$) with a clear abrupt change in 2017. However, strong spatial heterogeneity was evident: most of the central and eastern regions experienced significant moistening, while only a small portion of the western area showed a weak drying trend.

Previous studies have noted that arid regions globally have become drier, and humid regions wetter over recent decades, with transition zones often exhibiting amplified wet–dry contrasts [10,37,38]. In the arid and semi-arid regions of northwest China, increased precipitation has driven a general rise in soil moisture [39]; on the eastern Qinghai–Tibet Plateau, sufficient precipitation has similarly led to soil moistening [40]. In contrast, the

western part of Gannan Prefecture (a semi-arid zone) has experienced only a weak drying trend, rather than the pronounced drying predicted by a simple "dry gets drier" pattern. This deviation may be attributed to the buffering effect of local precipitation increases or the moderated evaporative demand at high elevations in this alpine transition zone, which has partially offset the warming-induced drying pressure [8].

The abrupt change in SSM in 2017 is particularly noteworthy. This turning point coincides with a regional shift in hydroclimatic conditions: meteorological records indicate that annual precipitation in the upper Yellow River basin increased markedly after 2016, while temperature exhibited a slight stabilization or slower rise after 2017 [41]. Consequently, SSM entered a sustained moistening phase after 2017, confirming the dominant role of precipitation over temperature in controlling SSM interannual variability in this region [42].

Overall, the interannual variations in SSM over the study period has been relatively gentle, with no large-scale monotonic drought or moistening, but clear spatial differentiation and phased temporal evolution. Such spatial heterogeneity and the 2017 mutation reflect the combined impact of climate fluctuations and local underlying surface conditions [8]. These findings are consistent with reports for the entire Yellow River basin, where overall soil moisture has slightly increased but decreased significantly in local areas [10]. Nevertheless, uncertainties remain among SSM products (e.g., discrepancies between remote sensing retrievals and reanalysis data), warranting a cautious interpretation of long-term trends [38]. From the perspective of data quality and limitations, GLDAS-2.1 is a widely used global reanalysis dataset for long-term hydrological research. Its relatively coarse native resolution restricts its ability to depict fine-scale spatial heterogeneity, and a slight overestimation of soil moisture has been reported in alpine areas, especially under rainless conditions [24]. To reduce such errors, we adopted a series of strict preprocessing steps, including outlier elimination, missing value interpolation, and annual averaging of monthly data. Combined with existing field verification results in similar alpine plateau regions, GLDAS-2.1 is sufficiently reliable to support the analysis of long-term SSM trends and regional patterns in this study. In summary, SSM in this alpine transition zone shows "overall stability with significant local changes", highlighting its sensitivity to hydrothermal gradients. This understanding sets the stage for the following section, which examines the nonlinear driving mechanisms underlying these spatiotemporal patterns using SHAP and ALE analyses.

4.2. Statistical Associations Between Environmental Predictors and SSM Variation in Gannan Prefecture

4.2.1. Statistical Explanatory Weight of Natural Environmental Predictors

Based on model outputs, the spatiotemporal variations in SSM exhibited strong statistical associations with multiple environmental predictors. Natural predictors accounted for the majority of total statistical explanatory weight (95.79%) according to SHAP results, with precipitation carrying the largest relative explanatory proportion (37.93%). This finding is consistent with previous regional studies reporting close statistical links between precipitation and interannual SSM fluctuations [42].

Globally, precipitation variability explains most long-term soil moisture trends, and the influences of rising temperature are frequently counterbalanced by precipitation shifts at interdecadal scales [42]. Lal et al. (2023) demonstrated that climate variables correlate with more than 90% of global long-term soil moisture variability [43]. Our ALE curves captured analogous sample-based statistical patterns at the regional scale: model-estimated SSM gradually plateaus once precipitation exceeds 650–700 mm, while precipitation deficits correspond to reduced SSM values within the observed dataset.

Temperature ranked second in explanatory weight (27.30%) and displayed bidirectional statistical associations in SHAP plots. ET_p showed strong negative statistical correla-

tion with modeled SSM. These statistical patterns align with prior field observations [44]. The ALE curves captured obvious sample-limited breakpoints: the statistical link between temperature and SSM weakens above 8–9 °C, and modeled SSM declined markedly at ET_p of 800 mm·a^{-1}. The co-occurrence of low precipitation and high temperature was statistically associated with reduced SSM and increased likelihood of drought conditions [17,38]. Two-dimensional ALE further reflected comparable joint statistical patterns across predictors.

This study also captured typical land–atmosphere statistical correlation features unique to the study area. Combined with 2D ALE outputs, modeled SSM reached relatively high levels under simultaneous low precipitation and low ET_p. This seemingly counterintuitive pattern was widespread in alpine regions, where low ET_p usually accompanied low temperature and high elevation, suppressing evapotranspiration loss and facilitating soil water retention.

DEM and FVC also showed distinct statistical trends. DEM presented a V-shaped statistical pattern, with minimum modeled SSM values near 3000 m. Model outputs indicated improved soil water retention within the sample range when FVC exceeded 0.45. FVC exhibited dual statistical characteristics: dense vegetation may elevate transpiration and correspond to lower SSM in water-limited environments.

4.2.2. Local Effects and Dual Nature of Human Activities

Although our SHAP analysis showed anthropogenic factors (CLCD and LS) accounted for less than 5% of variations in model-estimated SSM, their local impacts across the study area cannot be ignored. Major human disturbances since 2000 include urban expansion, agricultural activities, grazing and a series of ecological projects, such as the Grain for Green Program, nature forest protection, and grassland compensation policies [41,42]. These activities and policies affect SSM mainly via two pathways: direct modification of land surface properties and indirect changes in vegetation conditions.

In terms of direct impacts, land use and human activity intensity clearly alter soil moisture. Urban and rural construction create impervious surfaces and reduce infiltration, leading to lower SSM in populated areas. Grassland and woodland generally maintain higher soil moisture, while cropland and built-up land have lower values [45,46].

Indirect impacts are mainly mediated by FVC, which is shaped by both climate and human management. Large-scale vegetation restoration has continuously improved regional FVC over the past two decades. According to our ALE results, model-estimated water retention rises with increased FVC [47,48]. The strengthened implementation of ecological policies around 2017 was an important factor for the abrupt rise in regional SSM. Policy effects vary spatially: vegetation restoration has promoted SSM growth in eastern and southern forest–grassland areas, while standardized grazing has prevented grassland degradation and soil drying in western pastoral zones [34]. By contrast, uncontrolled overgrazing reduces vegetation coverage and weakens water retention, resulting in localized SSM decline [49,50].

Consistent with previous studies on the Qinghai–Tibet Plateau and Loess Plateau, vegetation greening driven by policies has dual effects on SSM [34]. Improved vegetation enhances water conservation, while overly dense vegetation increases transpiration and reduces SSM under water-limited conditions. Artificial irrigation can alleviate drought, but excessive groundwater extraction may also cause sustained SSM loss.

Human activities and local policies do not dominate the overall long-term trend of SSM, but they profoundly reshape its spatial pattern and interact with natural factors at local scales. Rational ecological governance and land use planning are essential for maintaining regional water conservation functions.

Large-scale land use changes have altered soil moisture in many regions [51]. For instance, the Grain for Green Program on the Loess Plateau increased vegetation but raised transpiration and reduced soil moisture; vegetation restoration there even led to a decline in the 0–10 m soil layer [52]. Globally, 40%–500% of greening areas experience soil drying due to higher transpiration [1]. On the Qinghai–Tibet Plateau, rising vegetation cover increased evapotranspiration by about 2.9% from 1982 to 2020 [53]. These findings indicate that policy-driven vegetation restoration may bring risks of soil moisture deficit and drought.

In Gannan Prefecture, human activities such as forest-grass expansion and overgrazing alter soil–vegetation–atmosphere water exchange. Existing studies show climate change acts as the primary factor for grassland SSM decline on the Qinghai–Tibet Plateau, with human activities bringing additional local drying effects [17]. Our SHAP results also show that human factors make a small contribution to model-estimated SSM compared with natural variables. Human activities impose dual effects on SSM. Specifically, irrigation increases SSM levels, and the ALE curves identify a statistical precipitation breakpoint ranging from 650 to 700 mm that differentiates the divergent impacts of precipitation on SSM. In contrast, excessive groundwater extraction reduces deep-soil moisture and amplifies the adverse effects of low precipitation or high ET [38,54]. These segmented statistical breakpoints only reflect patterns captured within our sample dataset rather than universal ecological critical thresholds, but human interventions including irrigation, groundwater pumping, and land use change may push the regional water–energy balance toward these sample-limited transition ranges and correspond to subsequent SSM reduction.

4.3. Discussion on SHAP and ALE Methodologies

Combining SHAP and ALE algorithms adopted in this study, we further discuss their performance in identifying critical thresholds and complex interactions among variables. As a mainstream interpretable tool for black-box machine learning, SHAP quantifies the independent contribution of each variable and ranks their overall importance [55]. This method effectively distinguishes major variables correlated with SSM variations, while it has inherent limitations in directly characterizing nonlinear relationships and extracting explicit ecological thresholds.

One-dimensional ALE compensates well for this shortcoming. This method removes the interference from correlated variables and reliably captures the nonlinear features of model-estimated SSM [33]. In our analysis, key thresholds including precipitation at 650–700 mm and the abrupt decline of ET_p at 800 mm·a^{-1} were accurately identified. Two-dimensional ALE was further applied to analyze synergistic and antagonistic relationships among precipitation, temperature, and potential evapotranspiration, which cannot be fully revealed by single-factor analyses.

Overall, the integrated framework of SHAP, one-dimensional ALE, and two-dimensional ALE constructs a complete analytical workflow for studying nonlinear correlations. This combined method has been widely applied to explore soil moisture dynamics and related ecohydrological processes in alpine watersheds [15]. It should also be noted that the thresholds and interaction characteristics obtained from these data-based methods are region-specific. Long-term field monitoring and cross-regional investigations are needed in future studies to further support these statistical findings.

Likewise, the validation framework for assessing model predictions should also account for spatial and temporal dependencies. The cross-validation in this study adopted a random pixel-year splitting strategy, which may lead to overly optimistic estimates of model accuracy [56]. To obtain more reliable generalization error estimates, future studies should employ spatial or temporal blocked cross-validation. For instance, the blockCV R package [15] can generate independent training and test sets through spatial blocking,

while the CAST package [57] provides a more comprehensive framework for spatial model validation and assessment.

5. Conclusions

This study quantitatively analyzed the spatiotemporal characteristics of SSM in Gannan Prefecture, a critical alpine water conservation zone in the upper Yellow River basin. From 2000 to 2022, SSM exhibited a significant upward trend across most territories, while its spatial pattern of high values in the southeast and low values in the northwest remained stable. Wavelet analyses depicted the long-term temporal evolution patterns of SSM.

Natural environmental predictors accounted for the majority of statistical explanatory weight for modeled SSM variability, with precipitation and temperature showing the strongest statistical associations, followed by ET_p, DEM, and FVC. ALE curves captured nonlinear sample-limited statistical patterns and joint correlations between variables. Modeled SSM gradually plateaued once precipitation exceeded 650–700 mm; SSM decreased markedly at a ET_p of 800 mm·a^{-1}. The statistical link between temperature and SSM weakened above 8–9 °C, and elevated FVC values corresponded to enhanced soil water retention within the observed dataset. Two-dimensional ALE further illustrated combined statistical relationships among precipitation, temperature, and ET_p under diverse hydrothermal backgrounds.

The overall moistening trend observed in this study reflects favorable background conditions for local ecological restoration. The segmented statistical breakpoint of FVC only describes variation characteristics captured by our dataset and cannot be treated as a unified regulatory benchmark for regional water conservation. The statistical associations between environmental predictors and SSM derived from machine learning offer descriptive insights into alpine soil moisture dynamics in the upper Yellow River basin.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/land15071202/s1, Table S1. Statistical metrics of pixel-wise Mann–Kendall test and strict Benjamini–Hochberg FDR correction for all valid native 0.25° GLDAS-2.1 pixels within Gannan Prefecture. Trend significance was determined by FDR-adjusted q-value threshold of 0.05 after trend-free pre-whitening (TFPW).

Author Contributions: Conceptualization, X.W. and J.C.; methodology, J.C. and X.W.; software, J.C.; validation, J.C., X.W. and X.Z. (Xiaowei Zhang); formal analysis, S.L. and F.N.; investigation, S.L. and W.D.; resources, F.N., S.L. and X.Z. (Xiaowei Zhang); data curation, F.N. and W.D.; writing—original draft preparation, J.C. and X.W.; writing—review and editing, X.W.; visualization, J.C.; supervision, X.W. and X.Z. (Xiaolei Zhou); project administration, X.W. and X.Z. (Xiaolei Zhou); funding acquisition, F.N. and X.Z. (Xiaolei Zhou). All authors have read and agreed to the published version of the manuscript.

Funding: This research work was supported by grants from the Science and Technology Innovation Fund of Gansu Agricultural University (GAU-KYQD-2022-20) and Evaluation of Multi-functional Ecosystem Service Values and Improvement Strategies for Forest and Grassland Resources in Gansu Province (GSAU-JSFW-2025-27-01).

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Liu, Y.; Li, Z.; Chen, Y.; Jin, L.; Li, F.; Wang, X.; Long, Y.; Liu, C.; Kayumba, P.M. Global Greening Drives Significant Soil Moisture Loss. *Commun. Earth Environ.* **2025**, *6*, 600. [CrossRef]

2. Tong, Y.; Wang, Y.; Zhou, J.; Guo, X.; Wang, T.; Xu, Y.; Sun, H.; Zhang, P.; Li, Z.; Lauerwald, R. Dataset of soil hydraulic parameters in the Yellow River Basin based on in situ deep sampling. *Sci. Data* **2024**, *11*, 740. [CrossRef] [PubMed]
3. Yao, L.; Leng, G.; Yu, L.; Li, H.; Tang, Q.; Python, A.; Hall, J.W.; Liao, X.; Li, J.; Qiu, J. Emergent constraints on global soil moisture projections under climate change. *Commun. Earth Environ.* **2025**, *6*, 39. [CrossRef]
4. Sun, J.; Yang, K.; He, X.; Wang, G.; Wang, Y.; Yu, Y.; Lu, H. Causal Pathways Underlying Global Soil Moisture–Precipitation Coupling. *Nat. Commun.* **2025**, *16*, 8935. [CrossRef] [PubMed]
5. Zhao, W.; Yu, X.; Xu, C.; Li, S.; Wu, G.; Yuan, W. Dynamic traceability effects of soil moisture on the precipitation–vegetation association in drylands. *J. Hydrol.* **2022**, *615*, 128645. [CrossRef]
6. Hao, Y.; Mao, J.; Bachmann, C.M.; Apple, M.; Shi, M.; Jin, M.; Zhu, Q.; Kannenberg, S.; Shi, X.; Zhang, X.; et al. Soil Moisture Controls over Carbon Sequestration and Greenhouse Gas Emissions: A Review. *npj Clim. Atmos. Sci.* **2025**, *8*, 16. [CrossRef]
7. Qiao, L.; Zuo, Z.; Zhang, R.; Piao, S.; Xiao, D.; Zhang, K. Soil moisture–atmosphere coupling accelerates global warming. *Nat. Commun.* **2023**, *14*, 4908. [CrossRef] [PubMed]
8. Guan, Y.; Gu, X.; Slater, L.J.; Li, J.; Kong, D.; Zhang, X. Spatio-Temporal Variations in Global Surface Soil Moisture Based on Multiple Datasets: Intercomparison and Climate Drivers. *J. Hydrol.* **2023**, *625*, 130095. [CrossRef]
9. Qing, Y.; Wang, S.; Yang, Z.-L.; Gentine, P.; Zhang, B.; Alexander, J. Accelerated soil drying linked to increasing evaporative demand in wet regions. *npj Clim. Atmos. Sci.* **2023**, *6*, 205. [CrossRef]
10. Hu, L.; Xie, Y.; Liu, Y.; Chen, S.; Yu, H.; Bie, Q.; Zhao, F.; Zhao, Y. Mapping the primary factors driving spatiotemporal variations of surface soil moisture in the Yellow River Basin of China. *Sci. Rep.* **2025**, *15*, 12359. [CrossRef] [PubMed]
11. Wang, Y.; Mao, J.; Hoffman, F.M.; Bonfils, C.J.W.; Douville, H.; Jin, M.; Thornton, P.E.; Ricciuto, D.M.; Shi, X.; Chen, H.; et al. Quantification of human contribution to soil moisture-based terrestrial aridity. *Nat. Commun.* **2022**, *13*, 6848. [CrossRef] [PubMed]
12. Pan, Y.; Hu, S.; Gao, L.; Qiu, J.; Peng, Z.; Hu, L.; Zhao, T.; Dong, J.; Tan, C. Capabilities of microwave-based soil moisture products in capturing extreme hydrological conditions. *J. Hydrol.* **2025**, *664*, 134392. [CrossRef]
13. Liu, Y.; Yang, Y.; Song, J. Variations in Global Soil Moisture During the Past Decades: Climate or Human Causes? *Water Resour. Res.* **2023**, *59*, e2023WR034915. [CrossRef]
14. Gebrechorkos, S.H.; Sheffield, J.; Vicente-Serrano, S.M.; Funk, C.; Miralles, D.G.; Peng, J.; Dyer, E.; Talib, J.; Beck, H.E.; Singer, M.B.; et al. Warming accelerates global drought severity. *Nature* **2025**, *642*, 628–635. [CrossRef] [PubMed]
15. Liu, Y.; Zha, Y.; Ran, G.; Zhang, Y.; Shi, L. SMRFR: A global multilayer soil moisture dataset generated using Random Forest from multi-source data. *Sci. Data* **2025**, *12*, 1170. [CrossRef] [PubMed]
16. Cao, X.; Zhao, Z.; Zheng, Y.; Su, C.; Lei, Q.; Li, W.; Li, C. Climate change threatens water resources over the Qinghai-Tibetan plateau. *Sci. Rep.* **2025**, *15*, 21996. [CrossRef] [PubMed]
17. Xiao, J.; Yu, C.; Fu, G. Influences of human activity and climate change on growing-season soil moisture in the Qinghai–Tibet grasslands from 2000 to 2020. *Front. Ecol. Evol.* **2023**, *11*, 1264870. [CrossRef]
18. Zhang, L. Study Advances of Soil Moisture Regime on the Loess Plateau. *Hans. J. Agric. Sci.* **2017**, *7*, 412–420.
19. Zhang, C.; Guo, S.; Guan, Y.; Cai, D.; Bian, X. Temporal Stability of Vegetation Cover across the Loess Plateau Based on GIMMS during 1982–2013. *Sensors* **2021**, *21*, 315. [CrossRef] [PubMed]
20. Tan, Q.; Hua, T.; Zhao, H.; Zhou, P. Vegetation greening and climate change respectively regulates the long-term trend and interannual variability in evapotranspiration over the Loess Plateau since the 21st century. *Front. Ecol. Evol.* **2025**, *13*, 1513189. [CrossRef]
21. Wang, S.; Cui, C.; Dai, Q. Contributions of Vegetation Greening and Climate Change to Evapotranspiration Trend after Large-Scale Vegetation Restoration on the Loess Plateau, China. *Water* **2021**, *13*, 1755. [CrossRef]
22. Liu, Z.; Cao, W. Spatiotemporal Variation of Water Conservation Function and Its Driving Factors in Gannan Prefecture of Gansu Province During 2005—2020. *Bull. Soil Water Conserv.* **2024**, *44*, 231–240. [CrossRef]
23. Rodell, M.; Houser, P.R.; Jambor, U.; Gottschalck, J.; Mitchell, K.; Meng, C.-J.; Arsenault, K.; Cosgrove, B.; Radakovich, J.; Bosilovich, M.; et al. The Global Land Data Assimilation System. *Bull. Am. Meteorol. Soc.* **2004**, *85*, 381–394. [CrossRef]
24. Lu, C.Y.; Feng, W.L.; Wang, Y.Q.; Wang, H.F.; Tian, P.; Lu, X.N. Application of soil moisture assimilation products at different depths in Western Sichuan Plateau. *Bull. Soil Water Conserv.* **2021**, *41*, 173–181. (In Chinese) [CrossRef]
25. Yang, Y.; Sun, H.; Zhang, Y.; Yu, Z.; Gu, H.; Gu, X. Regional patterns and long-term persistence of extreme precipitation in China (1959–2015). *J. Hydrol. Reg. Stud.* **2025**, *61*, 102625. [CrossRef]
26. Qiu, Z.; Gong, D.; Zhao, M.; Dong, D. Spatiotemporal Dynamics and Driving Mechanisms of Soil Conservation Services (SCS) in Zhejiang Province, China: Insights from InVEST Modeling and Machine Learning. *Remote Sens.* **2025**, *17*, 2865. [CrossRef]
27. Zhong, X.; Gong, H.; Chen, B.; Zhou, C.; Xu, M. Study on the evolution of shallow groundwater levels and its spatiotemporal response to precipitation in the Beijing Plain of China based on variation points. *Ecol. Indic.* **2024**, *166*, 112466. [CrossRef]
28. Conte, L.C.; Bayer, D.M.; Bayer, F.M. Bootstrap Pettitt test for detecting change point in hydroclimatological data: A case study for Itaipu hydroelectric plant in Brazil. *Hydrol. Sci. J.* **2019**, *64*, 1312–1326. [CrossRef]

29. Zeng, B.; Wen, B.; Zhang, X.; Zhao, S.; Shang, G.; An, S.; Li, Z. Analysis on Spatiotemporal Variation in Soil Drought and Its Influencing Factors in Hebei Province from 2001 to 2020. *Agriculture* **2025**, *15*, 1109. [CrossRef]
30. Wang, X.; Wang, B.; Xu, X.; Liu, T.; Duan, Y.; Zhao, Y. Spatial and temporal variations in surface soil moisture and vegetation cover in the Loess Plateau from 2000 to 2015. *Ecol. Indic.* **2018**, *95*, 320–330. [CrossRef]
31. Ren, J.; Zhou, H.; Tao, Z.; Ge, L.; Song, K.; Xu, S.; Li, Y.; Zhang, L.; Zhang, X.; Li, S. Long-term monitoring chlorophyll-a concentration using HJ-1 A/B imagery and machine learning algorithms in typical lakes, a cold semi-arid region. *Opt. Express* **2024**, *32*, 16371. [CrossRef] [PubMed]
32. Bergstra, J.; Bengio, Y. Random Search for Hyper-Parameter Optimization. *J. Mach. Learn. Res.* **2012**, *13*, 281–305. Available online: http://jmlr.org/papers/volume13/bergstra12a/bergstra12a.pdf (accessed on 1 July 2026).
33. Apley, D.W.; Zhu, J. Visualizing the Effects of Predictor Variables in Black Box Supervised Learning. *Model. J. R. Stat. Soc. Ser. B Stat. Methodol.* **2016**, *82*, 1059–1086. [CrossRef]
34. Wang, X.; Yao, W.; Feng, X.; Jia, Z.; Zhang, X.; Ma, J.; Zhou, J.; Tu, Y.; Sun, Z. Changes and driving factors of ecosystem services supply and demand on the Tibetan plateau. *Acta Ecol. Sin.* **2023**, *43*, 6968–6982. (In Chinese) [CrossRef]
35. Ma, Y.; Hou, P.; Zhang, L.; Cao, G.; Sun, L.; Pang, S.; Bai, J. High-Resolution Quantitative Retrieval of Soil Moisture Based on Multisource Data Fusion with Random Forests: A Case Study in the Zoige Region of the Tibetan Plateau. *Remote Sens.* **2023**, *15*, 1531. [CrossRef]
36. Xu, G.; Huang, M.; Li, P.; Li, Z.; Wang, Y. Effects of Land Use on Spatial and Temporal Distribution of Soil Moisture within Profiles. *Environ. Earth Sci.* **2021**, *80*, 128. [CrossRef]
37. McColl, K.A.; Alemohammad, S.H.; Akbar, R.; Konings, A.G.; Yueh, S.; Entekhabi, D. The global distribution and dynamics of surface soil moisture. *Nat. Geosci.* **2017**, *10*, 100–104. [CrossRef]
38. Wang, L.; Xie, Z.; Jia, B.; Xie, J.; Wang, Y.; Liu, B.; Li, R.; Chen, S. Contributions of climate change and groundwater extraction to soil moisture trends. *Earth Syst. Dyn.* **2019**, *10*, 599–615. [CrossRef]
39. Guo, L.; Zhu, B.; Jin, H.; Zhang, Y.; Min, Y.; He, Y.; Shi, H. Spatial-Temporal Variation Characteristics and Influencing Factors of Soil Moisture in the Yellow River Basin Using ESA CCI SM Products. *Atmosphere* **2022**, *13*, 962. [CrossRef]
40. Liu, Y.; Yang, Y. Multi-depth evolution characteristics of soil moisture over the Tibetan Plateau in the past 70 years using reanalysis products. *Front. Environ. Sci.* **2022**, *10*, 979853. [CrossRef]
41. Zhang, Y.N.; Song, X.N.; Leng, P.; Liang, G.; Yin, D. Temporal and spatial variation of summer soil moisture and its driving factors in the Yellow River Basin in recent 20 years. *J. Univ. Chin. Acad. Sci.* **2024**, *41*, 477–489. (In Chinese) [CrossRef]
42. Vargas Zeppetello, L.R.; Trevino, A.M.; Huybers, P. Disentangling contributions to past and future trends in US surface soil moisture. *Nat. Water* **2024**, *2*, 127–138. [CrossRef]
43. Lal, P.; Shekhar, A.; Gharun, M.; Gharun, M.; Das, N.N. Spatiotemporal evolution of global long-term patterns of soil moisture. *Sci. Total Environ.* **2023**, *867*, 161470. [CrossRef] [PubMed]
44. Fan, K.; Slater, L.; Zhang, Q.; Sheffield, J.; Gentine, P.; Sun, S.; Wu, W. Climate warming accelerates surface soil moisture drying in the Yellow River Basin, China. *J. Hydrol.* **2022**, *615*, 128735. [CrossRef]
45. Li, X.; Liu, X.; Sun, B.; Wu, D.; Jin, Z.; Gao, L. Study on Grassland Degradation in Gannan Tibetan Autonomous Prefecture. *Pratacultural Sci.* **2026**, *43*, 43–51. (In Chinese)
46. Wang, Z.; Huang, L.; Shao, M.; Pei, Y. Soil Water Holding Capacity under Different Land Use Patterns in the Qinghai Alpine Region. *Arid. Zone Res.* **2021**, *38*, 1722–1730. (In Chinese) [CrossRef]
47. Jia, H.; Zhang, J.; Liu, L.; Liu, J.; Yang, S. Variation and Multi-Scenario Simulation of Habitat Quality in the Gonghe Basin, China. *J. Desert Res.* **2025**, *45*, 29–36. (In Chinese) [CrossRef]
48. Hu, J.; Zhou, Q.; Lü, Y.; Hu, J.; Chen, Y.; Gou, X. Comparison study to the effectiveness of typical ecological restoration measures in semi-humid sandy land in eastern Qinghai-Tibetan Plateau, China. *Acta Ecol. Sin.* **2020**, *40*, 7410–7418. (In Chinese) [CrossRef]
49. Seneviratne, S.I.; Corti, T.; Davin, E.L.; Hirschi, M.; Jaeger, E.B.; Lehner, I.; Orlowsky, B.; Teuling, A.J. Investigating soil moisture–climate interactions in a changing climate: A review. *Earth-Sci. Rev.* **2010**, *99*, 125–161. [CrossRef]
50. Shi, M.; Wang, X.; Li, X.; Duojie, Z.; Zhou, B.; Luo, Z.; Chen, S. Change in Seasonally Frozen Ground and Its Effects on Soil Moisture and Vegetation in an Alpine Meadow on the Eastern Edge of the Qinghai-Xizang Plateau. *Chin. J. Appl. Environ. Biol.* **2025**, *31*, 1111–1119. (In Chinese)
51. Xu, C.; Zhang, L.; Du, J.; Guo, Y.; Wu, Z.; Xu, Y.; Li, F.; Wang, F. Impact of alpine meadow degradation on soil water conservation in the source region of three rivers. *Acta Ecol. Sin.* **2013**, *33*, 12. (In Chinese) [CrossRef]
52. Lu, J.; Carbone, G.J.; Grego, J.M. Uncertainty and hotspots in 21st century projections of agricultural drought from CMIP5 models. *Sci. Rep.* **2019**, *9*, 4922. [CrossRef] [PubMed]
53. Wang, Y.; Hu, W.; Sun, H.; Zhao, Y.; Zhang, P.; Li, Z.; Zhou, Z.; Tong, Y.; Liu, S.; Zhou, J.; et al. Soil moisture decline in China's monsoon loess critical zone: More a result of land-use conversion than climate change. *Proc. Natl. Acad. Sci. USA* **2024**, *121*, e2322127121. [CrossRef] [PubMed]

54. Han, P.; Ren, H.; Zhao, Y.; Zhao, N.; Wang, Z.; Wang, Z.; Liu, Y.; Wang, Z. Quantifying the Impact of Vegetation Greening on Evapotranspiration and Its Components on the Tibetan Plateau. *Remote Sens.* **2025**, *17*, 1658. [CrossRef]
55. Omer, A.; Zhuguo, M.; Yuan, X.; Zheng, Z.; Saleem, F. A hydrological perspective on drought risk-assessment in the Yellow River Basin under future anthropogenic activities. *J. Environ. Manag.* **2021**, *289*, 112429. [CrossRef]
56. Gholami, H.; Mohammadifar, A.; Song, Y.; Li, Y.; Rahmani, P.; Kaskaoutis, D.G.; Panagos, P.; Borrelli, P. An Assessment of Global Land Susceptibility to Wind Erosion Based on Deep-Active Learning Modelling and Interpretation Techniques. *Sci. Rep.* **2024**, *14*, 18951. [CrossRef] [PubMed]
57. Meyer, H.; Reudenbach, C.; Wöllauer, S.; Nauss, T. Importance of spatial predictor variable selection in machine learning applications – Moving from data reproduction to spatial prediction. *Ecol. Model.* **2019**, *411*, 108815. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.