MULTI-DIMENSIONAL VARIATIONS AND DRIVING FACTORS OF SOIL WATER EROSION USING MACHINE LEARNING MODELS: A CASE STUDY OF THE RED SOIL REGION OF SOUTHERN CHINA

XU, X. M.* – ZHANG, X. Y. – QIN, L. H. – ZHONG, K. Y. – JI, L. L. – WU, Y. – YANG, Y. W.

School of Geography and Environmental Engineering, Gannan Normal University, Ganzhou 341000, Jiangxi, China

**Corresponding author*
e-mail: xuxiangming@gnnu.edu.cn

(Received 14th Feb 2025; accepted 24th Mar 2025)

Abstract. The red soil region of southern China (RSRSC) faces substantial soil water erosion (SWE) challenges, however, previous studies have been found to lack sufficient exploration of multi-dimensional characteristics, as well as an in-depth understanding of the physical significance of driving factors of SWE. Addressing these gaps is essential to provide scientific evidence for soil and water conservation zoning and SWE prevention measures. The Revised Universal Soil Loss Equation (RUSLE) and the optimal machine learning (ML) model were utilized to investigate the driving factors. The findings are as follows: (1) The average SWE modulus in the RSRSC underwent a significant decrease. High-high clusters, predominantly located in the southern part of the study area, exhibited an expansion tendency. (2) The vertical distribution of SWE further demonstrated that low-altitude areas exhibited significantly higher SWE areas and intensities when compared to high-altitude areas, a similar trend was observed in areas with gentle to moderate slopes. (3) The eXtreme Gradient Boosting (XGBoost) and Shapley Additive Explanations (SHAP) models were utilized to calculate the contribution to model predictions, with the Fractional Vegetation Cover (FVC) identified as the most critical factor. Subsequent analysis of SHAP value distributions indicated that FVC exhibited a negative influence on SWE.
Keywords: *vertical spatial characteristics, spatial aggregation characteristics, SHAP model, RUSLE model*

Introduction

SWE is recognized as one of the primary forms of global soil degradation and a major environmental issue in China. It not only depletes land resources and reduces land productivity but also leads to sediment deposition in rivers, lakes, and reservoirs, exacerbating floods and droughts within watersheds. These effects pose severe threats to human survival and development (Xiao et al., 2015a,b). At larger spatial scales, such as national, regional, or watershed levels, quantitative estimation of SWE and analysis of its spatial distribution are critical for macro-level decision-making in soil and water conservation and erosion control (Ganasri and Ramesh, 2016). SWE modeling plays a key role in identifying areas prone to erosion and understanding the susceptibility of soils with high erosion potential (Tao et al., 2024). Empirical, physical and data-driven learning models are the main methods used to model SWE (Nguyen et al., 2022; Avand et al., 2023; Khosravi et al., 2023). Empirical models estimate SWE using predefined physical parameters that represent key factors influencing erosion (Cai et al., 2000). These include the Universal Soil Loss Equation (USLE; Wischmeier and Smith, 1965) and its derivatives, such as the Revised Universal Soil Loss Equation (RUSLE; Renard et al., 1997), and the modified Mediterranean Desertification and Land Use model (MEDALUS; Abuzaid et al., 2021). The RUSLE model effectively incorporates a wide range of factors

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2303_57655786
© 2025, ALÖKI Kft., Budapest, Hungary

influencing SWE and serves as a robust method for quantitatively studying erosion processes. The RUSLE model has been extensively applied in various studies, with numerous researchers employing it to predict SWE processes, facilitating the understanding of regional SWE patterns and their temporal evolution (Hateffard et al., 2021; Wu et al., 2022).

The study of the driving mechanisms has become a prominent research focus due to the complexity of the formation and development processes of SWE (Panagos et al., 2015). These processes are the result of the combined effects of natural and anthropogenic drivers. As a result, traditional research methods such as correlation analysis (Li et al., 2020), principal component analysis (Sun et al., 2018), and geographical detectors (Wang and Xu, 2017) face limitations. These limitations include the challenges associated with handling nonlinear interactions, elucidating the physical significance of driving factors, and revealing the directional and nonlinear response characteristics of these drivers. Consequently, there is an urgent need to develop new research methods that can comprehensively evaluate the importance, nonlinear features, and directional effects of driving factors and thoroughly elucidate the mechanisms of SWE dynamics and their ecological implications. In recent years, the application of ML algorithms in the field of SWE research has seen steady growth, providing novel tools for unraveling complex driving mechanisms and achieving precise predictions. ML has exhibited remarkable proficiency in efficiently processing large, multidimensional datasets and capturing intricate nonlinear relationships without requiring causal assumptions, a methodology that is particularly advantageous in dynamic simulations and spatial distribution predictions (Arabameri et al., 2019). The efficacy of ML algorithms, including bagged multivariate adaptive regression splines (bagged MARS) and random forest (RF), and the boosting method includes Cubist and gradient boosting machine (GBM), in elucidating the roles of driving factors and generating precise predictions has been well-documented (Nguyen et al., 2021). However, the "black-box" nature of ML models poses challenges for decision-makers in terms of interpretability, thereby hindering their capacity to intuitively grasp the relative importance of SWE driving factors (Lee et al., 2022). To address this issue, interpretable ML methods, such as Shapley Additive Explanations (SHAP), have been increasingly applied in studies (Zhao et al., 2022). These methods facilitate the discernment of model response characteristics, the elucidation of variable importance, and the clarification of their nonlinear effects (Ling et al., 2022; Senanayake et al., 2022).

The topography of southern China, characterized by undulating red soil in the tropical and subtropical regions, fosters favorable hydrothermal conditions and abundant natural resources, thereby presenting significant potential for agricultural development. However, the region is confronted with substantial challenges, including severe soil and water loss, due to its topography, degradation of ecological diversity, and vegetation destruction resulting from engineering activities and urbanization (Liang et al., 2010). Consequently, there is an imperative for enhanced research on SWE in this region to ensure effective soil and water conservation, ecological restoration, and the judicious utilization of land (Huang et al., 2019b). A review of the extant literature reveals notable lacunae in our understanding of SWE in this region. Firstly, while many studies have examined horizontal spatial changes in spatiotemporal patterns and SWE evolution, less attention has been given to vertical spatial variations and spatial clustering characteristics of soil water erosion. Secondly, the complex interplay of nonlinear factors, including topography, vegetation, precipitation, and human activities, contributes to the dynamic nature of SWE.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2303_57655786
© 2025, ALÖKI Kft., Budapest, Hungary

Conventional research methodologies often prove inadequate in providing a comprehensive understanding of these intricate mechanisms. Moreover, extant research on the driving mechanisms of SWE rarely addresses nonlinear interactions and the physical significance of driving factors. Consequently, there is an imperative for enhanced research in these domains to facilitate effective soil and water conservation management in complex hilly regions. Hence, this research study aims to explore the spatiotemporal distribution patterns of SWE intensity in the RSRSC and to reveal the driving forces of soil erosion.

This study incorporates digital elevation models (DEM) to analyze the multidimensional spatiotemporal changes over nearly two decades and employs spatial auto-correlation analysis to examine the spatial clustering characteristics at the county scale. Moreover, the study employs a suite of seven regression models, including eXtreme Gradient Boosting(XGBoost), Random Forest(RF), Gradient Boosting(GB), Decision Tree(DT), K-Nearest Neighbors(KNN), Support Vector Machine(SVM), and Adaptive Boosting(AdaBoost), to assess the spatiotemporal dynamics of SWE. The model that demonstrates the strongest performance is then integrated with SHAP methods to explore the driving factors of soil water erosion. The findings of this study offer significant practical implications for the prevention of soil and water loss and the improvement of the ecological environment in the RSRSC.

Materials and methods

General situation of the study area

The region of RSRSC encompasses an area of approximately 1.18 million km², constituting 12.3% of China's total land area. Of this, SWE affects approximately 340,000 km², with red soil comprising more than half of the soil types. The distribution of red soil is predominantly concentrated in hilly areas with an elevation below 500 meters (Zhao et al., 2013; Zhang and Cheng, 2014). As delineated in the "National Soil and Water Conservation Zoning of China", the RSRSC is subdivided into a total of nine subregions, numbered V-1 to V-9. The present study is specifically concerned with an examination of three of these subregions that experience the most severe soil and water loss (Hu et al., 2020): the Jiangnan Mountain and Hill Region (V-4), the Zhejiang-Fujan Mountain and Hill Region (V-5), and the Nanling Mountain and Hill Region (V-6). These regions exhibit the typical characteristics of the RSRSC (Ministry of Water Resources of the PRC, 2013).

The location of the study area is illustrated in *Figure 1*. Collectively, these regions encompass a total of eight provinces, including the eastern Guangxi Zhuang Autonomous Region, the entirety of Jiangxi and Fujian provinces, central and southern Hunan Province, southeastern Hubei Province, most of Zhejiang Province, southern Anhui Province, and central and northern Guangdong Province. These regions encompass a total of 486 counties, with a combined area of 792,000 square kilometers, accounting for 67.1% of the entire southern red soil hilly region. This region accounts for 8.3% of China's total land area. It is situated within China's tropical and subtropical monsoon climate zones, which are characterized by abundant hydrothermal resources that are highly conducive to vegetation restoration and recovery from SWE. Annual precipitation exceeds 800 mm, with average annual temperatures ranging from 11°C to 23°C. The region receives 1,500–2,200 hours of annual sunshine, with the lowest monthly average temperatures above 0°C and the highest monthly averages exceeding 22°C. The seasonal distribution of

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2303_57655786
© 2025, ALÖKI Kft., Budapest, Hungary

precipitation is highly pronounced, with 70%–80% of the annual rainfall occurring between April and September, characterized by high rainfall intensity and concentrated downpours (Hu et al., 2020).

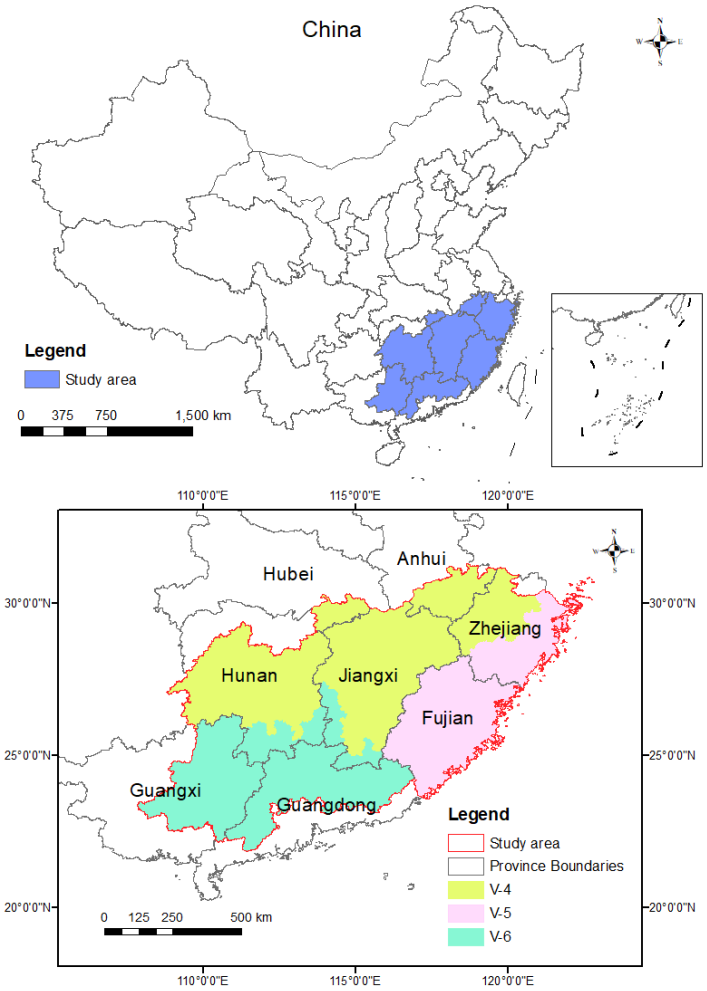

Figure 1. *The location of study area*

Data source

The data utilized in this study encompass the Normalized Difference Vegetation Index (NDVI), Digital Elevation Model (DEM), monthly average precipitation, soil texture, and land use data. Comprehensive details concerning the data sources are delineated in *Table 1*. To mitigate data errors and ensure spatial reference consistency, all datasets were standardized to the GCS_WGS_1984 coordinate system. Subsequently, calculations were conducted using the WGS_1984_UTM_Zone_50N projected coordinate system.

The research framework

The spatial modeling and driving factors of SWE in the RSRSC were performed by integrating RUSLE with ML models in four main steps: First, parameters and an erosion map using RUSLE in GIS were prepared. Second, the variations of SWE in multiple dimensions, temporal variation, horizontal and vertical spatial variation, and clustering

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2303_57655786
© 2025, ALÖKI Kft., Budapest, Hungary

variation were analyzed. Third, the parameters affecting erosion were prepared, and the ML models were compared. Fourth, the relative importance of the parameters causing SWE was determined using an XGBoost-SHAP model. Finally, targeted strategies for SWE prevention and control were proposed (*Figure 2*).

Table 1. *Data source*

Data Type	Data Sources	Data Accuracy	Year
Normalized Difference Vegetation Index	Resource and Environmental Science Data Platform (http://www.resdc.cn/DOI)	1 km	2000, 2010, 2020
Digital Elevation Model	Geospatial Data Cloud (https://www.gscloud.cn/)	30 m	2024
Monthly mean temperature	National Tibetan Plateau Data Center (https://data.tpdc.ac.cn/)	1 km	2000-2020
Monthly mean precipitation	National Tibetan Plateau Data Center (https://data.tpdc.ac.cn/)	1 km	2000-2020
Soil texture (Content of sand, silt,clay)	National Tibetan Plateau Data Center (https://data.tpdc.ac.cn/)	1 km	2009
Land use	Data Center for Resources and Environmental Sciences, Chinese Academy of Sciences (https://www.resdc.cn/)	30 m	2000, 2010, 2020

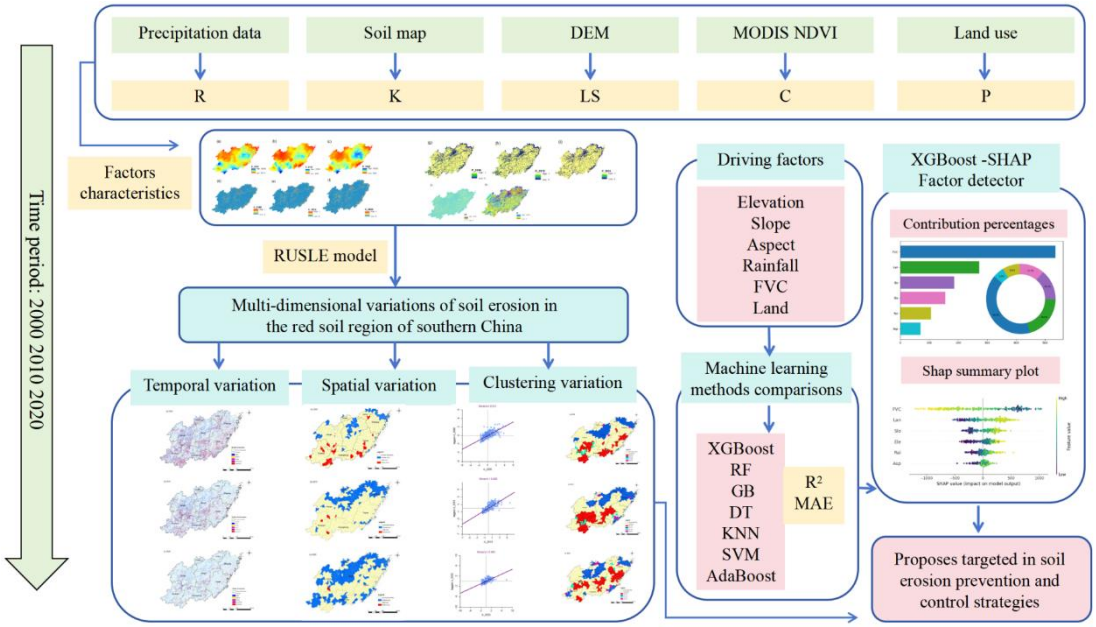

Figure 2. *The research framework*

RUSLE model

In this study, the RUSLE (Wischmeier and Smith, 1978; Renard et al., 1997) was employed to calculate the SWE modulus in the RSRSC. The calculations are shown in *Equation 1*.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2303_57655786
© 2025, ALÖKI Kft., Budapest, Hungary

$$A = R \times K \times L \times S \times C \times P \quad \text{(Eq.1)}$$

where A is the average soil loss of one year (Mg/(km^2·a)); K is soil erodibility (Mg·km^2·h/(km^2·MJ·mm)); R is rainfall erosivity (MJ·mm/(km^2·h·a)); L and S refers to slope length and steepness (dimensionless); C represents vegetation coverage and management (dimensionless); P stands for erosion control practices (dimensionless). The final outputs of SWE and its influencing factors were represented using a 1 km × 1 km grid cell, with a spatial resolution of 1 km.

Rainfall erosivity factor

Rainfall is a primary driving factor of SWE, with raindrop impact and runoff being the main manifestations of surface erosion. In this study, the rainfall erosivity factor (R) was estimated using the Wischmeier Equation (Wischmeier and Smith, 1978; Kite, 2001) based on monthly and annual average rainfall data from 1999–2001, 2009–2011, and 2019–2021. The calculations are shown in *Equation 2*.

$$R = \sum_{i=1}^{12} 1.735 \times 10^{(1.5\lg\frac{p_i^2}{p})-0.8188} \quad \text{(Eq.2)}$$

In *Equation 2*, P_i represents the monthly average rainfall (mm), and P represents the annual average rainfall (mm).

Soil erodibility factor

SWE is primarily influenced by the medium of soil, and the soil erodibility factor (K) is employed to evaluate the soil's susceptibility to erosion. In this study, the EPIC model proposed by Williams (1990) was utilized to simulate and calculate the soil erodibility factor based on a soil database. The calculation is predominantly contingent on the content of silt, sand, clay, and organic matter in the soil. The calculations are delineated in *Equation 3*.

$$K = 0.1317\left\{0.2 + 0.3\exp\left[-0.0256SAN\left(1-\frac{SIL}{100}\right)\right]\right\} \times \left(\frac{SIL}{SIL+CLA}\right)^{0.3} \times \left[1-\frac{0.25C}{C+\exp(3.72-2.95C)}\right] \times \left[1-\frac{0.7SN1}{SN1+\exp(-5.51+22.9SN1)}\right] \quad \text{(Eq.3)}$$

In *Equation 3*, 0.1317 is the coefficient for converting US units to international units. SAN, SIL, CLA, and C represent the sand (0.050–2.000 mm), silt (0.002–0.050 mm), clay (<0.002 mm), and organic matter content (%), respectively. SN1 =1- SAN/100.

Slope length and steepness factor

The slope length and steepness factor (LS) is a metric that reflects the impact of terrain features (slope length and steepness) on SWE. The precise relationship between these factors remains to be elucidated. However, within a certain range, it has been observed that the longer the slope, the greater the accumulation of flow, and the steeper the slope, the faster the runoff velocity. However, once a threshold is attained, the rate of SWE will no longer increase. The calculations for slope (S) are delineated in *Equation 4* (McCool et al., 1987; Liu et al., 1994; Liu et al., 1999).

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2303_57655786
© 2025, ALÖKI Kft., Budapest, Hungary

$$S = \begin{cases} 10.8\sin\theta + 0.03, \theta < 9\% \\ 16.8\sin\theta - 0.05, 18\% > \theta \geq 9\% \\ 21.91\sin\theta - 0.96, \theta \geq 18\% \end{cases} \quad \text{(Eq.4)}$$

In *Equation 4*, L represents the slope length factor, S represents the slope steepness factor, and θ is the slope value extracted from DEM data.

Using ArcGIS 10.8 software, the slope length factor (L) was calculated using the *Equation 5 and 6* (Qiu et al., 2018).

$$L = \left(\frac{\lambda}{22.13}\right)^m \quad \text{(Eq.5)}$$

$$\begin{cases} m = 0.5 & \theta \geq 9\% \\ m = 0.4 & 9\% > \theta \geq 3\% \\ m = 0.3 & 3\% > \theta \geq 1\% \\ m = 0.2 & \theta < 1\% \end{cases} \quad \text{(Eq.6)}$$

In *Equation 5.* and *Equation 6*, L represents the slope length factor; λ represents the slope length (m); m represents the slope length exponent; and θ represents the slope steepness (%).

Vegetation cover and management factor

According to the definition of the RUSLE, the Vegetation Cover and Management Factor(C) value represents the ratio of soil loss from land with vegetation cover or field management to that from bare soil under the same conditions, with a range from 0 to 1. The C value reflects the impact of vegetation cover and crop management practices on SWE. Given the significant correlation between the C value and vegetation cover, this study uses MODIS NDVI data as the basis and adopts the method proposed by Cai et al. (2000). The calculations are shown in *Equation 7 and 8*.

$$C = \begin{cases} 1 & f = 0 \\ 0.6508 - 0.3436\lg f & 0 < f \leq 78.3\% \\ 0 & f > 78.3\% \end{cases} \quad \text{(Eq.7)}$$

$$f = \frac{NDVI - NDVI_{\min}}{NDVI_{\max} - NDVI_{\min}} \quad \text{(Eq.8)}$$

In *Equation 7* and *8*, f represents the vegetation cover (%), and C represents the vegetation cover and management factor value. NDVI is the normalized difference vegetation index, with the multi-year monthly average used in this study. $NDVI_{max}$ refers to the NDVI value of pixels completely covered by vegetation, and $NDVI_{min}$ refers to the NDVI value of bare soil or areas without vegetation cover.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2303_57655786
© 2025, ALÖKI Kft., Budapest, Hungary

Soil conservation measure factor

The soil conservation measure factor (P) is defined as the ratio of SWE after implementing soil conservation measures relative to that of slope cultivation. This ratio ranges from 0 to 1. A value of 0 indicates no SWE, while 1 indicates no soil conservation measures were implemented. The present study utilizes land use data from 2000, 2010, and 2020, and refers to the assignment methods of Huang et al. (2019b); and Chen et al. (2014), to assign P values for different land use types. The P values assigned to various land use types, including dry land, paddy fields, forests, sparse forests, shrubland, other forest lands, grasslands, built-up areas, water bodies, and other land uses, are set to 0.4, 0.15, 1, 1, 1, 0.7, 1, 0, 0, and 1, respectively. The generation of the final P value raster layer is then undertaken.

Spatial auto-correlation of SWE modulus

Spatial auto-correlation methods are frequently employed to assess the spatial distribution characteristics of a designated attribute and its correlation with neighboring regions, thereby reflecting the disparities and interrelationships of spatial phenomena. This approach is a quantitative tool for analyzing spatial relationships and correlations between variables, encompassing global and local spatial auto-correlation (Pearson, 2002). To further explore the spatial correlation and heterogeneity of SWE modulus in the RSRSC, this study evaluates the SWE modulus of the RSRSC based on the average value of SWE modulus at the county level and conducts spatial auto-correlation analysis. Global spatial auto-correlation is employed to summarize the spatial distribution characteristics of correlated variables across the region. To describe the overall trend of spatial correlation of SWE modulus in the RSRSC, this study uses the global Moran's index (Moran's I) for analysis. The global Moran's index is a measure of spatial auto-correlation in regional aggregation.

Local spatial auto-correlation is defined as the correlation degree between a variable and the same variable in neighboring units within a small area. The spatial auto-correlation of county units is evaluated by the local Moran's index (Moran's I_i). The calculation formulae can be found in *Equations 9, 10, and 11* (Yao et al., 2015; Jin et al., 2020).

$$Moran'sI = \frac{n\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{(\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij})\sum_{i=1}^{n}(x_i - \bar{x})^2} \quad \text{(Eq.9)}$$

$$Moran'sI_i = \frac{(x_i - \bar{x})}{m_0}\sum_{j=1}^{n} w_{ij}(x_j - \bar{x}) \quad \text{(Eq.10)}$$

$$m_0 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n} \quad \text{(Eq.11)}$$

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2303_57655786
© 2025, ALÖKI Kft., Budapest, Hungary

where n represents the number of spatial units; x_i and x_j denote the water erosion values of the *i-th* and *j-th* spatial units, respectively; x is the mean water erosion value; w_{ij} is the spatial weight matrix; and m_0 denotes the variance.

Detection of SWE driving factors using ML methods

Comparison and optimization of ML models

First, a large dataset is constructed, with the 2000-2020 SWE modulus as the target variable and natural and human-induced factors closely related to SWE as the feature variables. The dataset is then divided into a training set and a test set, with 80% of the data allocated for training and 20% for testing. Subsequent to this, grid search (a hyperparameter tuning method) and Bayesian optimization (an intelligent search method) are employed to optimize the seven models. Consequently, a thorough evaluation and comparison of these models is conducted to ascertain the most optimal model and the most effective optimization method for subsequent driving factor detection. The assessment of the R^2 (a measure of how well the model fits the data, with values closer to 1 indicating a more effective explanation of data variation) and MAE (a measure of the model's mean prediction error, with smaller values indicating more accurate predictions) on the training and test sets across the seven ML models is conducted. The seven ML models that are employed in this study are as follows (Chen and Guestrin, 2016; Cristianint and Shawe-Taylor, 2000; Ke et al., 2017; Zhang et al., 2020):

(1) XGBoost

XGBoost is an efficient implementation based on Gradient Boosting Tree, which is widely used in classification and regression tasks. The primary advantages of XGBoost include its capacity to manage high-dimensional data, prevent overfitting, and attain high accuracy in a range of ML tasks.

(2) RF

RF is an ensemble learning method based on decision tree algorithms. It addresses classification problems through the implementation of the "majority voting" method and regression problems by means of averaging. The methodology entails the generation of numerous decision trees and the aggregation of their prediction outcomes, thereby facilitating enhanced generalization performance.

(3) GB

GB is a ML method that involves a stepwise optimization process, in which each iteration enhances the model by aligning residuals (errors from the preceding prediction). The underlying principle of this approach involves the integration of multiple weak learners, typically decision trees, with the objective of enhancing the prediction accuracy of the model.

(4) DT

DT is a simple supervised learning method that utilizes data partitioning to predict outcomes, gradually constructing a tree structure. This method finds common application in the domains of classification and regression.

(5) KNN

KNN is an instance-based learning algorithm. For the purposes of classification or regression, the KNN algorithm identifies the k nearest neighbors to the target data point.

(6) SVR

SVR is an extension of the Support Vector Machine (SVM) designed to solve regression problems. The methodology entails the optimization of a high-dimensional

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2303_57655786
© 2025, ALÖKI Kft., Budapest, Hungary

space through the identification of an optimal hyperplane, thereby facilitating the adjustment of the model's tolerance parameter.

(7) AdaBoost

AdaBoost is an ensemble learning method that iteratively optimizes weak learners (typically decision trees). The fundamental premise of AdaBoost is to augment the weight of misclassified samples, thereby compelling subsequent models to prioritize these samples. This process is intended to enhance the accuracy of the model.

Detection of driving factors

The optimal model and the most effective optimization method are applied to the trained model for driving factor detection. The SHAP method is employed to quantify the contribution of each feature factor to the model's prediction results, thereby generating feature importance values. This approach unveils the relative importance and direction of action of each driving factor on SWE. The SHAP method is a model interpretation technique grounded in game theory. It utilizes the Shapley value theory to quantify the contribution of each feature to the model output. The SHAP method ensures local accuracy and consistency, and it is capable of explaining various black-box models. The integration of XGBoost with SHAP capitalizes on the strengths of XGBoost in modeling complex nonlinear relationships and high-dimensional data while facilitating precise feature contribution analysis through SHAP. This method unveils the driving factors for individual sample predictions and conducts a global analysis of feature contributions, offering insights into the mechanisms driving factors in complex systems.

Results and analysis

Temporal variation trend of SWE in the RSRSC based on two kinds of scale

In this investigation, the parameters of the RUSLE model within the RSRSC exhibited considerable variability. The spatial distributions of these parameters were illustrated in *Figure 3*. The R factor in the study area underwent a 13.18% decrease from 2000 to 2010 and a 19.3% increase from 2010 to 2020, resulting in a net increase of 3.57% from 2000 to 2020. Analysis of the R factor revealed an upward trend in the northeastern sector of the RSRSC from the year 2000 to 2020 (*Figure 3a-c*). The C factor exhibited a decline of 5.04% from 2000 to 2010 and 0.38% from 2010 to 2020, resulting in a cumulative decrease of 5.40% from 2000 to 2020. The C factor demonstrated that the majority of the region was characterized by lower values (*Figure 3d-f*). The P factor exhibited a decline of 1.08% from 2000 to 2010, followed by an increase of 2.08% from 2010 to 2020, resulting in a net change of 0.98% over the entire period. It was observed that the values in the southern region surpassed those in the northern region (*Figure 3g-i*). The LS factor displayed a relatively uniform distribution across the study area (*Figure 3j*). Lastly, the K factor distribution suggested that the northeastern part of the RSRSC was predominantly associated with lower values (*Figure 3k*).

According to the SWE classification and grading standards established by the Ministry of Water Resources of the People's Republic of China, the SWE intensity calculated by the RUSLE model is divided into six levels: tolerable, slight, medium, strong, very strong, and destructive (with intensity ranges of <500 Mg/(km²·a), 500-2500 Mg/(km²·a), 2500-5000 Mg/(km²·a), 5000-8000 Mg/(km²·a), 8000-15000 Mg/(km²·a), and >15000 Mg/(km²·a), respectively.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2303_57655786
© 2025, ALÖKI Kft., Budapest, Hungary

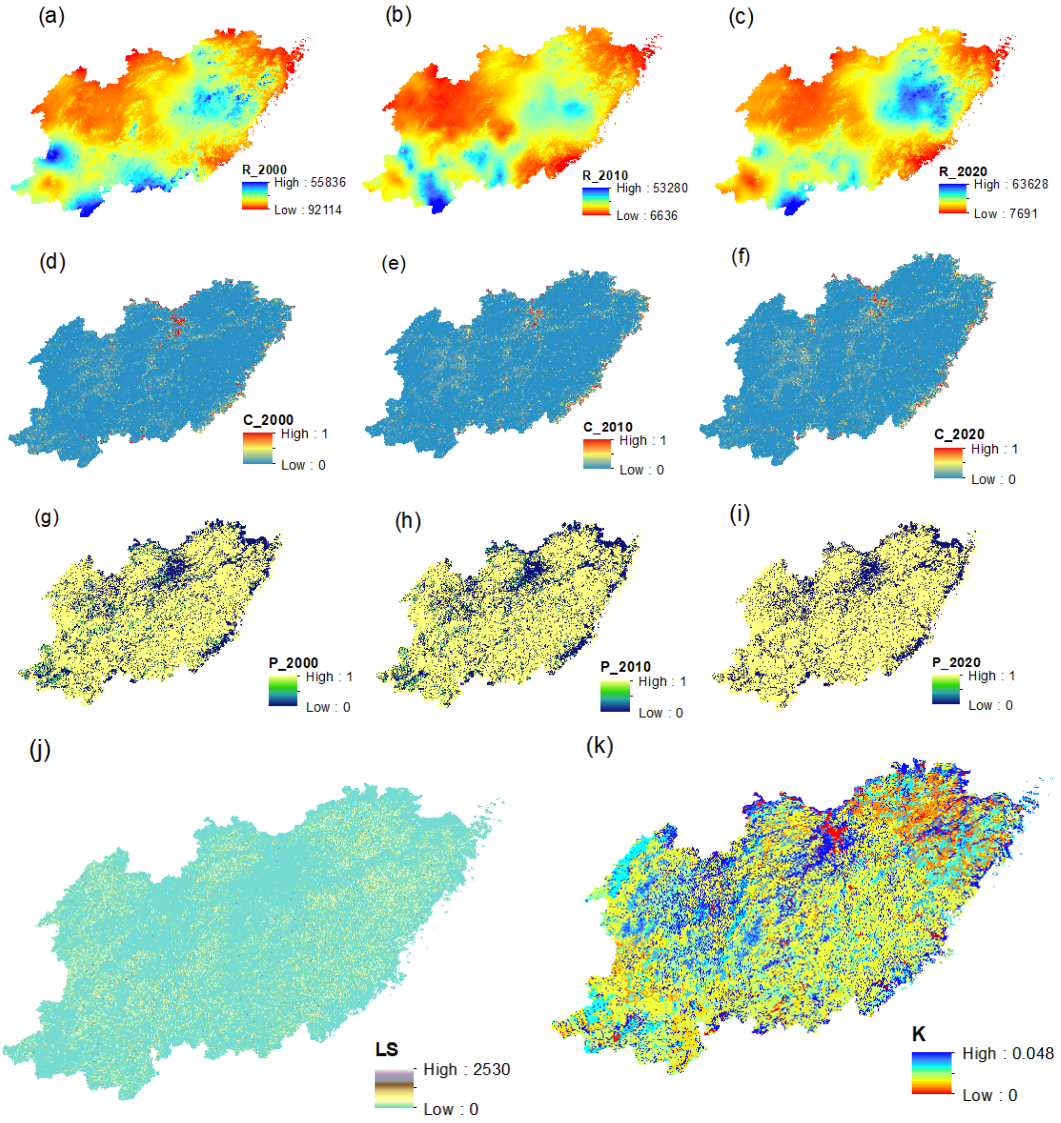

Figure 3. *Factors of RUSLE model distribution in the RSRSC. (a-c) rainfall erosivity factor 2000-2020, (d-f) vegetation cover and management factor 2000-2020, (g-i) soil conservation measure factor 2000-2020, (j) slope length and steepness factor, (k) soil erodibility factor*

Tolerable erosion intensity, defined as less than 500 Mg/(km²·a), is categorized as the tolerable soil loss amount. While regions experiencing slight or higher erosion intensity, categorized as >500 Mg/(km²·a), are designated as soil water erosion(SWE) areas. *Figure 4* illustrate the temporal variations based on raster scale within the RSRSC from 2000 to 2020. The results demonstrate that from 2000 to 2010, the slight, medium, strong, very strong, and destructive ratings decreased by 9.46%, 15.12%, 29.78%, 35.42%, and 40%, respectively. From 2010 to 2020, the ratings decreased by 36.08%, 40.02%, 36.9%, 32.8%, and 19.45%, respectively. The findings reveal that the average SWE modulus in 2000 was 828.99 Mg/(km²·a), which decreased to 532.24 Mg/(km²·a) in 2010, marking a reduction of 35.80%. By 2020, the average SWE modulus further decreased to 379.78 Mg/(km²·a), representing a reduction of 449.21 Mg/(km²·a), or 54.19%, compared to 2000. This substantial decrease in the average SWE modulus signifies a notable improvement in SWE conditions in the RSRSC. The SWE area diminished from

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2303_57655786
© 2025, ALÖKI Kft., Budapest, Hungary

163,003.54 km² in 2000 to 82,923.19 km² in 2020, representing a decline of 49.13%. The SWE conditions in the RSRSC have shown a gradual improvement, with a significant reduction in the SWE area, consistent with the previous findings (Wang et al., 2024).

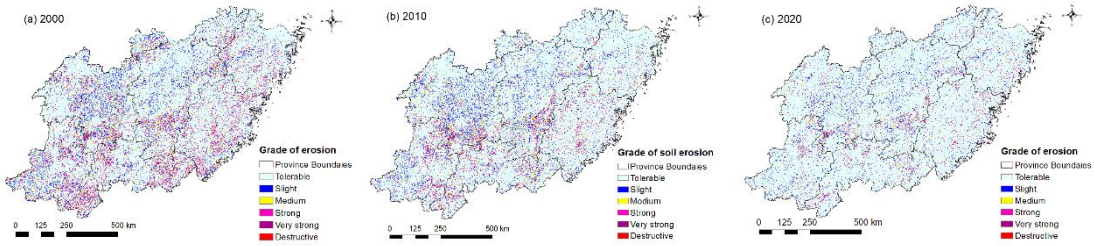

Figure 4. *Distribution of SWE modulus classes in 2000 (a), 2010 (b) and 2020 (c) based on raster scale*

In order to further illuminate the temporal variation trend of SWE in the RSRSC, this study examined the changes in SWE intensity levels at the county scale (486 counties totally). By calculating the difference in SWE intensity between two adjacent periods, the area changes in SWE intensity were analyzed for three time intervals (2000-2010, 2010-2020, and 2000-2020). These changes were classified into three categories: decreased erosion intensity, unchanged erosion intensity, and increased erosion intensity. The trend of SWE intensity changes from 2000 to 2020 is shown in *Figure 5*. The findings reveal that from 2000 to 2010, 78.07% of the SWE intensity levels remained constant, while 1.58% of the area exhibited an increase, with a limited extent and sporadic distribution. From 2010 to 2020, 1.22% experienced an increase, primarily concentrated in Anhui province. In the interval spanning from 2000 to 2020, 0.78% of the area experienced an escalation in intensity, characterized as sporadically distributed and minimal. The overall intensity of SWE in the RSRSC remained relatively stable during the study period. While areas exhibiting increased erosion intensity were limited in number and distributed sporadically, their significance as focal areas for SWE control and prevention should not be overlooked.

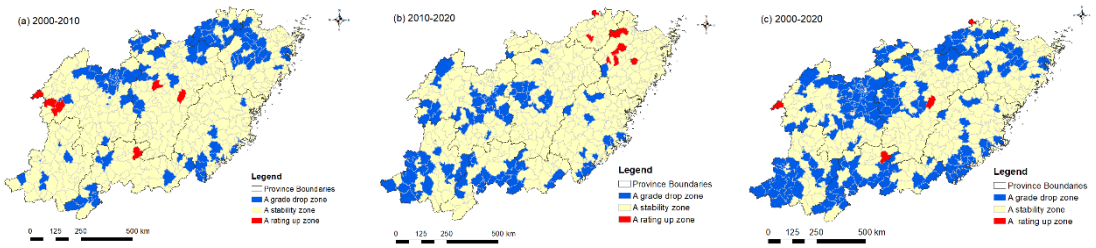

Figure 5. *Changes in SWE modulus classes in the RSRSC: 2000–2010 (a), 2010–2020 (b), and 2000–2020 (c), based on county-level scale*

Spatial variation trend of SWE in the RSRSC

Studies revealed that all 486 county units have SWE intensities below 5000 Mg/(km²·a). The SWE intensity can be classified into three categories: tolerable, slight, and medium. The horizontal spatial distribution characteristics of SWE at the

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2303_57655786
© 2025, ALÖKI Kft., Budapest, Hungary

county level in the RSRSC are shown in *Figure 6*. The findings reveal that in the year 2000, the predominant SWE intensity was light, yet 26 counties experienced medium intensity, predominantly situated in the southern regions of Guangdong, Guangxi, Fujian, and Hunan Province. By the year 2010, the SWE distribution remained primarily light, though the affected area had diminished, transitioning to permissible intensity. A mere four counties retained medium intensity: Daoxian and Beihu District in Hunan, Yunan County in Guangdong, and Yangshuo County in Guangxi. In 2020, the distribution of SWE was primarily composed of permissible and light erosion, each accounting for nearly 50%, with only Daoxian County in Hunan still experiencing medium erosion intensity. The SWE situation has further improved.

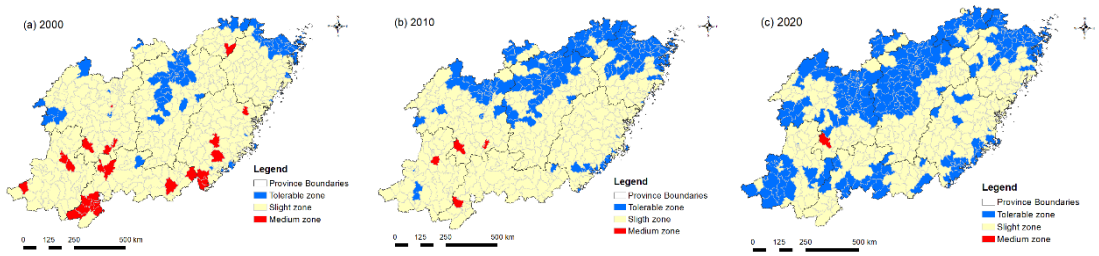

Figure 6. *Changes in SWE modulus classes based on county scale in the RSRSC*

This study analyzes the vertical distribution characteristics of SWE in the RSRSC, utilizing DEM data and focusing on two key factors: elevation and slope. The study area has been divided into six elevation classes, and the SWE area and its proportion within each class have been calculated (*Table 2*). The analysis reveals that the maximum SWE area is concentrated between elevations of 200 and 500 meters. This phenomenon is predominantly attributed to the topographical characteristics of the region, which predominantly comprises low mountains and hills with an average elevation below 500 meters. With respect to the proportion of SWE area, the highest percentage is observed in the 100-200 meter elevation range. From 2000 to 2020, a decreasing trend in SWE was observed across all elevations. The reduction in SWE is more pronounced in high-elevation areas (>500 m) compared to low-elevation areas (<500 m), with the smallest decrease observed in the 0-100 meter range. Despite the overall improvement in SWE, low-elevation areas (<500 m) continue to be a primary focus for SWE control efforts. The persistence of susceptibility in these regions is attributable to a combination of factors, including land use practices, topography, and terrain characteristics. The necessity for targeted interventions to avert further degradation is underscored by these findings.

Slopes were classified into six categories according to the SWE Classification and Grading Standards of the Ministry of Water Resources. The SWE area and its proportion within each slope category is delineated in *Table 3*. The results indicate that the largest areas and proportions of SWE occur at slopes of 8-15° (in 2000 and 2010) and 3-8° (in 2020). This phenomenon can be attributed to human activities, such as land reclamation, that are more prevalent in these slope ranges. In contrast, areas with slopes greater than 25° are typically protected by vegetation, which plays a crucial role in mitigating erosion. From 2000 to 2020, there was a substantial decrease in SWE on steep slopes (>15°), while SWE on moderate and gentle slopes (<15°) exhibited a more gradual decline. Despite these improvements, the need for reinforced SWE control measures persists in areas

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2303_57655786
© 2025, ALÖKI Kft., Budapest, Hungary

characterized by mild and moderate slopes (slopes less than 15°). These regions persist in their susceptibility to SWE, underscoring the necessity for targeted measures to forestall further degradation.

Table 2. *The SWE area and its proportion at different elevations*

Year	Statistics	0-100 m	100-200 m	200-500 m	500-800 m	800-1500 m	>1500 m
2000	Area(km^2)	35500.11	44056.68	61137.65	15300.9	6576.14	337.22
	Area(%)	20.83%	28.49%	21.62%	12.45%	11.09%	19.58%
2010	Area(km^2)	27731.76	34077.88	48353.93	13175.36	5526.56	319.08
	Area(%)	16.22%	22.02%	17.12%	10.69%	9.34%	18.39%
2020	Area(km^2)	22290.11	23083.27	28312.2	6614.9	2465.23	152.54
	Area(%)	13.04%	14.92%	10.03%	5.37%	4.17%	8.79%
2000-2010	Change(km^2)	-7768.35	-9978.8	-12783.72	-2125.54	-1049.58	-18.14
	Change(%)	-21.88%	-22.65%	-20.9%	-13.89%	-15.96%	-5.38%
2010-2020	Change(km^2)	-5441.65	-10994.61	-20041.73	-6560.46	-3061.33	-166.54
	Change(%)	-19.62%	-32.26%	-41.45%	-49.79%	-55.39%	-52.19%
2000-2020	Change(km^2)	-13210	-20973.41	-32825.5	-8686	-4110.91	-184.68
	Change(%)	-37.21%	-47.61%	-53.69%	-56.77%	-62.51%	-54.77%

Table 3. *The SWE area and its proportion at different slope*

Year	Statistics	0-3°	3-8°	8-15°	15-25°	25-35°	>35°
2000	Area(km^2)	11175.15	40967.31	46544.17	41608.76	17833.74	4862.85
	Area(%)	15.61%	23.64%	24.72%	19.97%	15.78%	12.91%
2010	Area(km^2)	8875.63	33083.54	37440.96	32302.74	13672.36	3805.86
	Area(%)	12.40%	19.09%	19.89%	15.50%	12.10%	10.11%
2020	Area(km^2)	7319.83	24561.59	23734.62	18245.15	7166.48	1888.91
	Area(%)	10.22%	14.17%	12.61%	8.76%	6.34%	5.02%
2000-2010	Change(km^2)	-2299.52	-7883.77	-9103.21	-9306.02	-4161.38	-1056.99
	Change(%)	-20.58%	-19.24%	-19.56%	-22.37%	-23.33%	-21.74%
2010-2020	Change(km^2)	-1555.8	-8521.95	-13706.34	-14057.59	-6505.88	-1916.95
	Change(%)	-17.53%	-25.76%	-36.61%	-43.52%	-47.58%	-50.37%
2000-2020	Change(km^2)	-3855.32	-16405.72	-22809.55	-23363.61	-10667.26	-2973.94
	Change(%)	-34.50%	-40.05%	-49.01%	-56.15%	-59.82%	-61.16%

Spatial clustering variation of SWE in the RSRSC

A global and local spatial auto-correlation analysis of SWE intensity was performed. The results are shown in *Figure 7*. From 2000 to 2020, the global Moran's I values for SWE intensity were 0.531, 0.620, and 0.463, respectively. The global Moran's I values for all three time periods were found to be greater than 0, and the significance levels were all below 0.01. This finding suggests that the spatial distribution of SWE intensity at the county scale exhibits significant positive auto-correlation, indicating that neighboring counties demonstrate a high degree of spatial similarity and clustering.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2303_57655786
© 2025, ALÖKI Kft., Budapest, Hungary

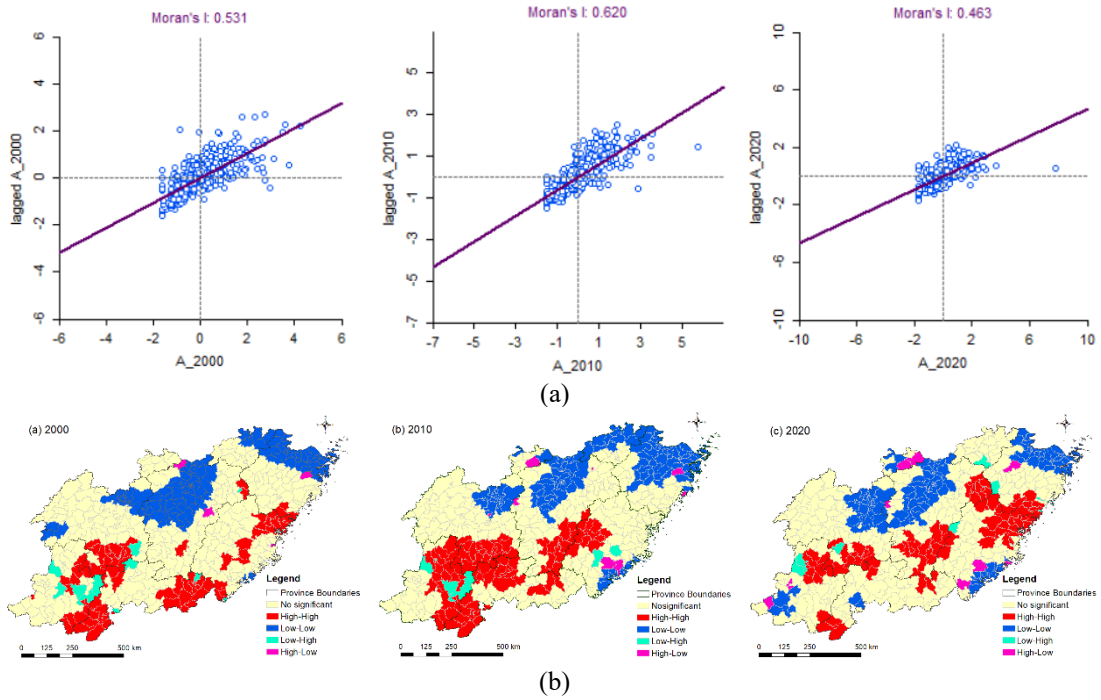

Figure 7. *Spatial clustering variation of SWE in the RSRSC a. Global spatial auto-correlation, b. Local spatial auto-correlation*

The objective of the local spatial auto-correlation analysis is to reveal the local distribution characteristics of SWE modulus in the study area and to visualize the clustering locations of these indices and their regional correlations. The LISA (Local Indicators of Spatial Association) distribution map is a visual representation of the local spatial relationship between events and their surrounding events. The results indicate that SWE in the RSRSC exhibited significant local spatial auto-correlation characteristics from 2000 to 2020. The spatial clustering of SWE intensity was primarily characterized by high-high and low-low value clusters. Specifically, high-high value clusters were predominantly located in the southern region, while low-low value clusters were predominantly situated in the northern region. Despite the enhancement in SWE control measures within the RSRSC, a meticulous observation of the proliferation of high-high value clusters is imperative.

Detection of driving factors

A total of seven ML models, including XGBoost, RF, GB, DT, KNN, SVR, and AdaBoost, were compared. The process involved three steps: dataset preparation, training and testing dataset splitting, and model optimization. Additionally, grid search and Bayesian optimization methods were evaluated for selecting the best optimization strategy. The R^2 and MAE values for the seven ML models and the two optimization methods are presented in *Table 4*. Under these specified parameters, the R^2 value on the training set reached 0.995, while the testing set achieved an R^2 of 0.7866 and an MAE of 187.6057. The performance of other models is outlined below: XGBoost (Grid Search) exhibited comparable performance to the Bayesian-optimized version, with slightly lower R^2 on the testing set, positioning it as the second-best model. GB (Bayesian Optimization and Grid Search) and RF (Bayesian Optimization) exhibited slightly lower testing R^2

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2303_57655786
© 2025, ALÖKI Kft., Budapest, Hungary

compared to XGBoost, with relatively higher MAE, indicating higher prediction errors. Overall, XGBoost with Bayesian Optimization demonstrated superior performance in terms of both R^2 and MAE on the testing set, combining high model fit and low prediction error. Therefore, it was identified as the optimal choice.

Table 4. *The R^2 and MAE scores among the seven ML models*

Model	Train R^2 (Grid)	Test R^2 (Grid)	Train R^2 (Bayes)	Test R^2 (Bayes)	Test MAE (Grid)	Test MAE (Bayes)
XGBoost	0.9902	0.7695	0.995	0.7866	189.6089	187.6057
Random Forest	0.7607	0.6593	0.8972	0.7233	244.7956	221.2769
Gradient Boosting	0.9624	0.7686	0.9853	0.7678	196.161	195.3195
Decision Tree	0.8248	0.6027	0.7603	0.5879	262.0749	267.6359
K-Nearest Neighbors	0.3409	0.1959	0.3013	0.2065	377.7701	367.687
Support Vector Regression	0.1352	0.163	0.1637	0.1706	357.2514	362.6473
Adaboost	0.5965	0.5718	0.5961	0.5604	302.6767	305.8517

The most suitable model, XGBoost combined with the SHAP method, was utilized to calculate the contribution of each feature factor to the model's prediction results. The factors employed in the study included elevation (Ele), slope (Slo), rainfall (Rai), fractional vegetation cover (FVC), aspect (Asp), and land use (Lan). The ensuing results are presented in *Figure 8*, elucidating the relative importance and direction of influence of each driving factor on SWE. *Figure 8* illustrates the Mean |SHAP| Value and the contribution percentage of each feature, thereby quantifying the importance of each feature to the model's output. The Mean |SHAP| Value is defined as the average absolute SHAP value of a feature across all samples, thereby reflecting its overall importance in the model's predictions. Higher values indicate a more significant influence on the prediction results, suggesting a more substantial role. The contribution percentages, on the other hand, signify the importance contribution of each feature. Higher values in these percentages indicate a more significant influence on the model's predictions. The results indicate that the Mean |SHAP| Value for FVC is 535.29, with a contribution percentage of 40.4%, making it the most significant feature factor and contributing the most to the ML model's prediction. The Mean |SHAP| Value for Lan is 272.87, with a contribution percentage of 20.6%, indicating its role as the second most significant feature factor. The third most salient feature factor, Slo, exhibits a Mean |SHAP| Value of 186.62 and a contribution percentage of 14.1%.

As illustrated in *Figure 8*, the impact of each feature factor on the model's prediction output is depicted based on SHAP analysis. The horizontal axis of the figure represents the SHAP values, with larger values indicating a greater positive contribution of the feature to the predicted value and smaller values denoting a larger negative contribution. The vertical axis enumerates the feature factors, with the color of the points ranging from purple (indicating smaller feature values) to yellow (indicating larger feature values), thereby representing the magnitude of the feature values. The results indicate that, based on the distribution of SHAP values, FVC exerts a significant negative impact on SWE modules, with higher vegetation coverage leading to a reduction in SWE modules. In contrast, other features, including Lan, Slo, Ele, Rai, and Asp, demonstrate a positive impact, with increased feature values leading to an exacerbation of SWE.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2303_57655786
© 2025, ALÖKI Kft., Budapest, Hungary

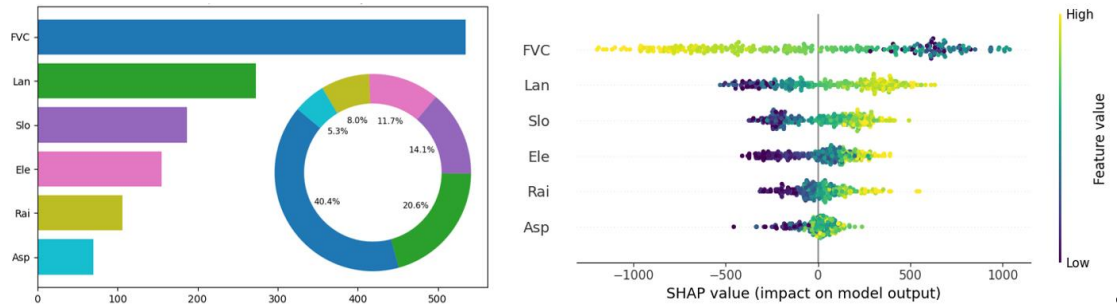

Figure 8. *Detection of driving factors. a. Bar plots of average absolute SHAP values and pie charts of contribution percentages. B. SHAP summary plot*

Discussion

Temporal and spatial variation of SWE in the RSRSC at different research scales

Existing studies on the spatial variation of SWE are typically conducted at one of several scales, such as raster (Wang et al., 2024), administrative (Xiao et al., 2015b; Chen et al., 2023), or watershed (Li and Zheng, 2012; Wu et al., 2019) levels. Raster scale studies, representing a micro-scale approach, are often based on grids of 1 km × 1 km, 5 km × 5 km, or 10 km × 10 km. In contrast, administrative or watershed scales are considered meso- or macro-scales. This study examined the temporal and spatial variations of SWE in the RSRSC at both 1 km × 1 km raster and county scales. The raster-based approach demonstrated a substantial decrease in the mean SWE modulus and a considerable reduction in SWE area. At the county scale, the results indicated that, although the majority of SWE intensity levels remained unchanged or decreased, a few counties experienced an increase in erosion intensity. The spatial auto-correlation analysis at the county scale revealed that high-high value clusters were predominantly located in the southern region, while low-low value clusters were mainly found in the northern region. Despite the overall improvement in SWE, high-high value clusters exhibited a tendency to expand. While micro-scale analysis can capture the differences among subsystems within larger regions, the results obtained at smaller scales are often difficult to generalize for ecological regulation applications at larger scales. The utilization of administrative units, such as counties, offers several advantages, including enhanced practicality for implementation, such as the collection of statistical data and the development of SWE control models (Huang et al., 2019a). Achieving this at the county scale is more straightforward than at the raster scale. For instance, the results of this study, which highlight counties where erosion intensity has increased and regions with expanding high-high value clusters, can serve as important references for the development of soil and water conservation regulations and planning. Consequently, integrating research across multiple scales is instrumental in comprehensively assessing SWE status and enhancing the objectivity and applicability of evaluation outcomes.

Vertical spatial variation of SWE in the RSRSC

The study of the vertical distribution characteristics of SWE in the RSRSC, in relation to elevation and slope, is a crucial aspect of understanding the spatial distribution of SWE. A review of extant literature reveals a divergence of conclusions among previous studies on the vertical distribution of SWE. Specifically, the study in Mekong River Basin

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2303_57655786
© 2025, ALÖKI Kft., Budapest, Hungary

revealed that the SWE modulus exhibited an initial increase with elevation up to 500–1000 m, followed by a subsequent decrease, reaching a peak within this specific range. Additionally, SWE increased with slope until it reached a peak at slopes of 8–15°, with most erosion occurring in the 0–15° slope range (Wu et al., 2019). In a separate study, Researchers examined SWE in the Qinling-Daba Mountains, where they reported a "increase-decrease-increase" pattern for SWE area with elevation, with the highest erosion occurring at elevations between 500 and 1500 m. For slope, they observed a "decrease-increase-decrease" pattern, with the largest erosion area occurring at slopes of 15°–35° (Wang et al., 2024). The study of the Hanjiang River Basin revealed that the 20–30° slope range exhibited the highest values of medium and above erosion intensity (Gao and Li, 2021). Concurrently, a similar range of slopes was identified, specifically 15°–25°, and elevations of 200–500 m, as the areas most affected by SWE in their study of SWE in the southern provinces (Chen et al., 2014).

The findings in this paper underscore the notion that SWE exhibits distinct vertical distribution patterns across different regions, which are influenced by environmental factors and human activities. In the RSRSC, characterized by hilly terrain, land use types such as orchards, dry land, and sparse forests are more prone to SWE, and these land uses are primarily found in the lower elevation (<500 m) areas. Agricultural and development activities, such as farming and orchard development, predominantly occur in gentle and medium slopes (<15°). According to the provisions outlined in the Soil and Water Conservation Law, agricultural activities are strictly prohibited on slopes exceeding 25°. The impact of vegetation cover on SWE is a pivotal consideration, with the influence of FVC changes being more pronounced in areas with steeper slopes. A study found that increasing vegetation cover was particularly effective in reducing SWE modulus in steep areas (Chen et al., 2023). Conversion of steep land to forestry has been identified as a measure to prevent further SWE. While current soil and water conservation laws and planning regulations strictly control erosion on steep slopes (above 25°), the erosion in areas with gentle and medium slopes (<15°) remains more severe and requires greater attention. This explains why the vertical distribution of soil erosion intensity in the RSRSC differs significantly from that in other areas.

Detection of SWE driving factors based on ML models

In addressing the reliability and practical application of methods for identifying the driving factors of SWE, the present study sought to contribute to the body of knowledge surrounding the risk detection of gully erosion in the Lhasa River Basin. To this end, four methods were employed: optimal scale regression, geodetector, and the combination of the two. These methods were selected based on their demonstrated efficacy in detecting the importance of factors affecting gully erosion. The analysis revealed that, in the context of optimal scale regression, the top three factor coefficients were elevation, soil type, and normalized vegetation index. Similarly, in the geodetic detector, elevation, soil type, and human footprints emerged as the top three influential factors (Li et al., 2022). To further explore the impact of fallow forest and geographic characteristics on SWE, a panel data fixed-effects model was employed. The findings indicated that geographic features exhibited a substantial moderating effect on the effectiveness of fallow forest projects in mitigating SWE. Specifically, slope and rainfall were found to have a significant negative moderating effect on the project's efficacy in preventing SWE, while sunshine duration demonstrated a positive moderating effect (Xie et al., 2022). The comparative assessment using boosted regression trees, binary logistic regression, frequency ratio, and numerical

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2303_57655786
© 2025, ALÖKI Kft., Budapest, Hungary

risk factor for gully erosion susceptibility modeling results showed that NDVI and lithology played a key role in gully occurrence (Arabameri et al., 2019).

In this study, an optimized XGBoost model was integrated with the SHAP model, leveraging the high fitting capability of the ML model and the interpretability of SHAP. The findings indicated that SHAP values in regions exhibiting high FVC (vegetation cover) were predominantly negative, signifying that increased vegetation cover exerts a mitigating effect on SWE. In contrast, other features, including Lan, Slo, Ele, Rai, and Asp, demonstrated a positive impact, with increased feature values leading to an exacerbation of SWE. The drivers of SWE in the region exhibit distinct characteristics, and the ML method employed in this study to identify these drivers differs from previous research. The findings suggest that the ML approach is effective in identifying the driving factors of southern red SWE. The identification of these factors provides a scientific foundation for the development of SWE prevention and control strategies in the southern red soil hilly region.

Proposes

The findings indicated a substantial decrease in the mean SWE modulus and a considerable reduction in SWE area in the RSRSC in 2020. However, a few counties experienced an increase in erosion intensity, and the high-high value clusters exhibited a tendency to expand. Additionally, the problem of SWE in low-altitude regions and gentle slopes of the RSRSC was significantly more serious compared to those in high-altitude and steeper slope regions. To address these concerns, several strategies have been proposed for the prevention and control of SWE. Primarily, there is a necessity to address the requirements for land use on gentle slopes within the framework of soil and water conservation zoning and control programs. Secondly, there is a need to allocate increased financial resources to soil and water conservation and control efforts to mitigate the tendency to expand. Lastly, bioremediation has been identified as an effective SWE control measure, with increased vegetation cover exerting a mitigating effect on SWE in the study area.

Conclusions

The study's primary conclusions are as follows:

From 2000 to 2020, the average SWE modulus underwent a significant decrease, with the SWE areas reducing by 49.13%. The analysis of spatial auto-correlation at the county scale demonstrated that high-high value clusters were predominantly located in the southern region. Despite the observed improvements in SWE, there is a discernible tendency for the expansion of high-high value clusters, a phenomenon that warrants close observation and potential intervention. (2) Despite improvements in SWE, the area and proportion of SWE in low-altitude regions (<500 m) of the RSRSC were significantly higher compared to those in high-altitude regions. Additionally, SWE was found to be more prevalent in areas characterized by gentle slopes and medium slopes compared to those with steeper slopes. (3) Applying the optimal model, XGBoost (with Bayesian optimization) plus SHAP method, the contributions of various feature factors (FVC, Lan, Slo, Ele, Rai, and Asp) to the model's predictions were quantified to explore the driving factors of SWE. The findings indicate that FVC was designated as the most significant feature factor contributing to the model's predictions. Lan and Slo were recognised as the other two important feature factors. A subsequent analysis of the direction of the driving

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2303_57655786
© 2025, ALÖKI Kft., Budapest, Hungary

effect reveals that FVC was predominantly negative on SWE, in contrast to the positive impact demonstrated by the other features.

Acknowledgements. This study was supported by the National Natural Science Foundation of China (42261015), the Science and Technology Foundation of the Education Department of Jiangxi Province (GJJ211426; GJJ201440).

REFERENCES

[1] Abuzaid, A. S., AbdelRahman, M. A. E., Fadl, M. E., Scopa, A. (2021): Land Degradation Vulnerability Mapping in a Newly-Reclaimed Desert Oasis in a Hyper-Arid Agro-ecosystem Using AHP and Geospatial Techniques. – Agronomy 11: 1426.

[2] Arabameri, A., Pradhan, B., Lombardo, L. (2019): Comparative assessment using boosted regression trees, binary logistic regression, frequency ratio and numerical risk factor for gully erosion susceptibility modelling. – Catena 183: 104223.

[3] Avand, M., Mohammadi, M., Mirchooli, F., Kavian, A., Tiefenbacher, J. P. (2023): A New Approach for Smart Soil Erosion Modeling: Integration of Empirical and Machine-Learning Models. – Environmental Modeling & Assessment 28: 145-160.

[4] Cai, C., Ding, S., Shi, Z., Huang, L., Zhang, G. (2000): Study of Applying USLE and Geographical Information System IDRISI to Predict Soil Erosion in Small Watershed. – Journal of Soil and Water Conservation 14: 19-24.

[5] Chen, S., Yang, X., Xiao, L., Cai, H. (2014): Study of Soil Erosion in the Southern Hillside Area of China Based on RUSLE Model. – Resour. Sci. 36: 1288-1297.

[6] Chen, T., Guestrin, C. (2016): XGBoost: A Scalable Tree Boosting System. – Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, August, pp. 785-794.

[7] Chen, M., Wang, X., Lin, J., Yue, H., Zhou, W., Jiang, H. (2023): Quantitative effects of land use and vegetation cover changes on soil erosion in Changting County in rencent 30 years. – Journal of Soil and Water Conservation 37: 168-177.

[8] Cristianint, N., Shawe-Taylor, J. (2000): An introduction to support vector machines and other kernel-based learning methods. – Cambridge university press.

[9] Ganasri, B. P., Ramesh, H. (2016): Assessment of soil erosion by RUSLE model using remote sensing and GIS - A case study of Nethravathi Basin. – Geoscience Frontiers 7: 953-961.

[10] Gao, Y., Li, H. (2021): Influence of landscape pattern change on soil erosion in Han River Basin. – Acta Ecologica Sinica 41: 2248-2260.

[11] Hateffard, F., Mohammed, S., Alsafadi, K., Enaruvbe, G. O., Heidari, A., Abdo, H. G., Rodrigo-Comino, J. (2021): CMIP5 climate projections and RUSLE-based soil erosion assessment in the central part of Iran. – Scientific Reports 11: 7273.

[12] Hu, X., Li, Z., Chen, J., Nie, X. (2020): Soil conservation benefits of grain for green program in the hilly red soil region of southern China. – Journal of Soil and Water Conservation 34: 95-100.

[13] Huang, M., Yue, W., Fang, B., Feng, S. (2019a): Scale response characteristics and geographic exploration mechanism of spatial differentiation of ecosystem service values in Dabie Mountain area, central China from 1970 to 2015. – Acta Geographica Sinica 74: 1904-1920.

[14] Huang, D., Yang, X., Cai, H., Xiao, Z., Han, D. (2019b): Identifying human-induced spatial differences of soil erosion change in a hilly red soil region of southern China. – Sustainability 11: 3103.

[15] Jin, S., Xiao, W., Yang, S., Zhao, D. (2020): Land Use Change and Its Ecosystem Response in the Yangtze River Economic Belt. – Economic Geography 40: 166-173.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2303_57655786
© 2025, ALÖKI Kft., Budapest, Hungary

[16] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T. (2017): Light GBM: A Highly Efficient Gradient Boosting Decision Tree. – Proceedings of the 31st International Conference on Neural Information Processing Systems, pp. 3149-3157.
[17] Khosravi, K., Rezaie, F., Cooper, J. R., Kalantari, Z., Abolfathi, S., Hatamiafkoueieh, J. (2023): Soil water erosion susceptibility assessment using deep learning algorithms. – Journal of Hydrology 618: 129229.
[18] Kite, G. (2001): Modelling the Mekong: hydrological simulation for environmental impact studies. – Journal of Hydrology 253: 1-13.
[19] Lee, C. Y., Hasegawa, H., Gao, S. C. (2022): Complex-valued neural networks: A comprehensive survey. – IEEE/CAA J. Autom. Sinica 9(8): 1406-1426.
[20] Li, T., Zheng, L. (2012): Soil erosion changes in the Yanhe watershed from 2001 to 2010 based on Rusle model. – Journal of Natural Resources 27: 1164-1175.
[21] Li, J., Bai, Y., Alatalo, J. M. (2020): Impacts of rural tourism-driven land use change on ecosystems services provision in Erhai Lake Basin, China. – Ecosyst. Serv. 42: 101081.
[22] Li, J., Chen, Y., Jiao, J., Chen, T., Chen, Y., Zhao, W., Zhao, C., Shang, T., Jian, J., Cao, X. (2022): Detecting gully occurrence risks using multivariate nonlinear spatial modeling in the Lhasa River Basin of China. – Transactions of the Chinese Society of Agricultural Engineering 38: 73-82.
[23] Liang, Y., Li, D., Lu, X., Yang, X., Pan, X., Mu, H., Shi, D., Zhang, B. (2010): Soil Erosion Changes over the Past Five Decades in the Red Soil Region of Southern China. – J. Mt. Sci. 7: 92-99.
[24] Ling, F., Luo, J., Li, Y., Tang, T., Bai, L., Ouyang, W., Yamagata, T. (2022): Multi-task machine learning improves multi-seasonal prediction of the Indian Ocean Dipole. – Nature Communications 13: 7681.
[25] Liu, B. Y., Nearing, M. A., Risse, L. M. (1994): Slope gradient effects on soil loss for steep slopes. – Transactions of the ASAE 37: 1835-1840.
[26] Liu, B. Y., Zhang, K. L., Jiao, J. Y. (1999): Soil erodibility and its application in erosion prediction. – Journal of Natural Resources 14: 345-350.
[27] McCool, D. K., Brown, L. C., Foster, G. R., Mutchler, C. K., Meyer, L. D. (1987): Revised slope steepness factor for the universal soil loss equation. – Transactions of the ASAE 30: 1387-1396.
[28] Ministry of Water Resources of the People's Republic of China. (2013): Bulletin on Soil and Water Conservation of the First National Water Conservancy Census. – Soil and Water Conservation in China 10: 2-3.
[29] Nguyen, K. A., Chen, W., Lin, B., Seeboonruang, U. (2021): Comparison of Ensemble Machine Learning Methods for Soil Erosion Pin Measurements. – ISPRS Int. J. Geo-Inf. 10: 42.
[30] Nguyen, T. C., Whelan, M., Nichols, J. D. (2022): Soil erosion response to land use change in a mountainous rural area of Son La Province of Vietnam. – Environ Monit Assess 194: 149.
[31] Panagos, P., Borrelli, P., Meusburge, K., Alewell, C., Lugato, E., Montanarella, L. (2015): Estimating the soil erosion cover-management factor at the European scale. – Land Use Policy 48: 38-50.
[32] Pearson, D. M. (2002): The application of local measures of spatial autocorrelation for describing pattern in north Australian landscapes. – J. Environ. Manag. 64: 85-95.
[33] Qiu, H., Cui, P., Regmi, A. D., Hu, S., Wang, X., Zhang, Y. (2018): The effects of slope length and slope gradient on the size distributions of loess slides: Field observations and simulations. – Geomorphology 300: 69-76.
[34] Renard, K. G., Foster, G. R., Weesies, G. A., Mccool, D. K., Yoder, D. C. (1997): Predicting soil erosion by water: a guide to conservation planning with the revised universal soil loss equation (RUSLE). – Agriculture Handbook, Washington DC: United States Department of Agriculture, 703.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2303_57655786
© 2025, ALÖKI Kft., Budapest, Hungary

[35] Senanayake, S., Pradhan, B., Alamri, A., Park, H. (2022): A new application of deep neural network (LSTM) and RUSLE models in soil erosion prediction. – Science of the Total Environment 845: 157220.
[36] Sun, B., Zhao, H., Lu, F., Wang, X. (2018): Spatial and temporal patterns of carbon sequestration in the Northeastern Forest Regions and its impact factors analysis. – Acta Ecologica Sinica 38: 4975-4983.
[37] Tao, W., Liu, S., Wang, Q., Su, L., Sun, W. (2024): Spatiotemporal Characteristics of Soil Erosion on the Chinese Loess Plateau and Strategies for Vegetation Management. – Journal of Soil Science and Plant Nutrition 24: 4439-4456.
[38] Wang, J., Xu, C. (2017): Geodetector; principle and prospective. – Acta Geographica Sinica 72: 116-134.
[39] Wang, L., Zhao, T., Zhu, W., Zhu, L. (2024): The spatiotemporal characteristics of soil erosion in the Qinling-Daba Mountains based on RUSLE model. – Journal of Soil and Water Conservation 38: 113-121.
[40] Williams, J. R. (1990): The erosion-productivity impact calculator (EPIC) model: A case history. – Philos. Trans. Biol. Sci. 329: 421-428.
[41] Wischmeier, W. H., Smith, D. D. (1965): Predicting rainfall-erosion losses from cropland east of the Rocky Mountains: Guide for selection of practices for soil and water conservation. – Agricultural Research Service, U. S. Dept. of Agriculture in cooperation with Purdue Agricultural Experiment Station.
[42] Wischmeier, W. H., Smith, D. D. (1978): Predicting rainfall erosion losses: a guide to conservation planning. – Agriculture Handbook No. 537. US Department of Agriculture Science and Education Administration, Washington DC.
[43] Wu, F., Zhu, Y., Xu, D., Shi, J., Jiang, Y. (2019): Assessment of soil erosion and prioritization for treatment at the catchment level in the Mekong Basin. – Acta Ecologica Sinica 39: 4761-4772.
[44] Wu, S., Zhang, B., Shi, X., Yuan, Z., Feng, H. (2022): Prediction of soil erosion under different land uses in the typical watershed of the Loess Plateau based on FLUS-CSLE model. – Transactions of the Chinese Society of Agricultural Engineering 38: 83-92.
[45] Xiao, L., Yang, X., Chen, S., Cai, H. (2015a): An assessment of erosivity distribution and its influence on the effectiveness of land use conversion for reducing soil erosion in Jiangxi, China. – Catena 125: 50-60.
[46] Xiao, Y., Ouyang, Z., Xu, W., Xiao, Y., Xiao, Q. (2015b): GIS-based spatial analysis of soil erosion and soil conservation in Chongqing, China. – Acta Ecological Sinica 35: 7130-7138.
[47] Xie, Y., Yao, S., Ding, Z., Hou, M., Deng, Y., Liu, G. (2022): The grain for green project, geographical features and soil erosion: taking 107 counties in Shaanxi province as example. – Acta Ecologica Sinica 42: 301-312.
[48] Yao, X., Zeng, J., Li, W. (2015): Spatial correlation characteristics of urbanization and land ecosystem service value in Wuhan Urban Agglomeration. – Transactions of the Chinese Society of Agricultural Engineering 31: 249-256.
[49] Zhao, Q., Huang, G., Ma, Y. (2013): The problems in red soil ecosystem in southern of China and its countermeasures. – Acta Ecologica Sinica 33: 7615-7622.
[50] Zhao, Y., Gao, G., Ding, G., Wang, L., Chen, Y., Zhao, Y., Yu, M., Zhang, Y. (2022): Assessing the influencing factors of soil susceptibility to wind erosion: A wind tunnel experiment with a machine learning and model-agnostic interpretation approach. – Catena 215: 106324.
[51] Zhang, P., Cheng, D. (2014): Interpretation on Technical Standards for Comprehensive Control of Water and Soil Erosion in the Red Soil Hilly Region in Southern China. – Journal of Yangtze River Scientific Research Institute 31: 28-34.
[52] Zhang, M., Shi, W., Xu, Z. (2020): Systematic comparison of five machine-learning models in classification and interpolation of soil particle size fractions using different transformed data. – Hydrol. Earth Syst. Sci. 24: 2505-2526.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2303_57655786
© 2025, ALÖKI Kft., Budapest, Hungary