



Machine learning-based sediment connectivity surrogates for RUSLE and MUSLE in ungauged watersheds

Abdul Alim Mohammadi¹ · Sibel Canaz Sevgen² · Günay Erpul¹

Received: 27 June 2025 / Accepted: 28 December 2025
© The Author(s) 2026

Abstract

Sediment yield prediction is vital for sustainable watershed management, particularly in data-scarce regions. This study, conducted in the Göksun Çayı Karaahmet sub-basin, Türkiye, evaluated whether sediment connectivity indices can reproduce outputs from the Revised Universal Soil Loss Equation (RUSLE) and Modified Universal Soil Loss Equation (MUSLE). Sediment yield was modeled for 196 sub-catchments and 69 rainfall events over 10 years using GIS-based factors: rainfall erosivity, soil erodibility, slope length-steepness, land cover, and hydrological parameters. Despite different assumptions (rainfall erosivity versus runoff and peak discharge), RUSLE and MUSLE showed strong agreement ($R^2 = 0.87$ at the event scale; $R^2 = 0.93$ at the sub-catchment scale). Predicted sediment yields ranged from 0.02 to 16.46 t ha⁻¹ (MUSLE) and 0.04–10.63 t ha⁻¹ (RUSLE/SDR), with mean values of 0.89 and 0.96 t ha⁻¹, respectively. Sediment connectivity indices-including the Index of Connectivity (IC), Sediment Delivery Ratio (SDR), and Topographic Wetness Index (TWI), were applied as inputs to five machine learning (ML) models (XGBoost, Random Forest, k-NN, SVR, and ANN). XGBoost and Random Forest achieved the best performance ($R^2 = 0.912$ – 0.942 , RMSE=0.065–0.089, MAE=0.047–0.055), reproducing empirical outputs. IC, SDR, and TWI were dominant predictors. These results demonstrate that connectivity metrics integrated with ML can emulate empirical erosion models, offering a scalable, data-efficient alternative for ungauged basins. However, because the models were trained on RUSLE/MUSLE outputs from 69 events under static land use and climate, they may underpredict extreme sediment events and require field validation before operational use.

Keywords Sediment connectivity indices · Empirical erosion models · Sediment yield prediction · Index of connectivity (IC) · Sediment delivery ratio (SDR)

Introduction

Soil erosion remains a significant challenge for sustainable land management and water resource conservation, particularly in regions vulnerable to climate-induced changes and land-use modifications. Traditional models such as Revised Universal Soil Loss Equation (RUSLE) (Renard et al. 1997) and Modified Universal Soil Loss Equation (MUSLE) (Williams 1975), have been widely applied to predict soil erosion and sediment yield at various spatial and temporal scales (Merritt et al. 2003; Djoukbala et al. 2018; Reda et al. 2024; Nawazuzzoha et al. 2024; Alebachew et al. 2025; Lamane et al. 2025a, b). While these models provide a practical framework for erosion assessment, they rely on empirical parameters and generalized assumptions, which can limit their effectiveness in complex, heterogeneous landscapes (Kinnell 2010). In response to these limitations, recent development have integrated machine learning (ML)

✉ Sibel Canaz Sevgen
ssevgen@ankara.edu.tr

Abdul Alim Mohammadi
mohammadialim123@gmail.com

Günay Erpul
erpul@ankara.edu.tr

¹ Department of Soil Science and Plant Nutrition, Faculty of Agriculture, Ankara University, Dışkapı, Ankara 06110, Türkiye

² Department of Real Estate Development and Management, Faculty of Applied Sciences, Ankara University, Çankaya, Ankara, Türkiye

techniques with these models to enhance prediction accuracy by incorporating data-driven approaches (Ajibade et al. 2020; Chen et al. 2021; Khosravi et al. 2023; Avand et al. 2023; Seitzky et al. 2024; Lamane et al. 2025a, b).

Both RUSLE and MUSLE have been extensively used to estimate soil erosion and sediment yield, although their intended temporal applications differ (Benavidez et al. 2018; Kumar et al. 2019; Mohammadi et al. 2025). RUSLE was originally developed to predict long-term average annual rainfall-induced erosion, while MUSLE, which replaces the rainfall factor with runoff volume (Q) and peak discharge (qp) ($Q \times qp$), was specifically designed for event-based sediment yield estimation (Sadeghi et al. 2014; Shi et al. 2022; Özşahin 2023). Although RUSLE was developed for long-term predictions, recent studies have adapted it for event-scale sediment yield estimation by integrating the sediment delivery ratio (SDR) (Kelsey 2002; Kinnell 2010; Deviren Saygin et al. 2014; Madenoğlu et al. 2024; Mohammadi et al. 2025). Lamane et al. (2025a, b) assessed RUSLE and MUSLE performance over a 5-year continuous period in the Bouregreg Watershed, Morocco. They demonstrated that integrating SDR with RUSLE and using high-resolution soil datasets improved long-term sediment yield predictions, whereas MUSLE slightly underestimated sediment transport, emphasize the critical influence of soil data resolution and SDR integration on long-term sediment yield estimation reliability. Both models account for key factors influencing soil erosion and sediment transport, including rainfall erosivity (R factor), runoff volume and peak discharge (Runoff factor), soil erodibility factor (K), topographic parameters (LS), representing slope length and steepness, land use/land cover factor (C), and conservation practices factor (P) (Prasannakumar et al. 2011; Luvai et al. 2022; Kumar et al. 2022; Noori and Azeez 2023). MUSLE's integration of hydrological flow parameters such as runoff volume and peak discharge enhances its ability to quantify sediment transport to watershed outlets (Mishra et al. 2006), making it particularly suitable for catchments where hydrological variability plays a dominant role (Koç, 2014).

The combination of these empirical models with experimental data, remote sensing (RS), and geographical Information System (GIS-based) spatial analysis has significantly advanced soil erosion research, enabling high-resolution assessments across various timescales (Bouhadeb et al. 2018; Deviren Saygin et al. 2018; Hadou et al. 2024; Demirtaş and Erpul 2024). Recent advancements in GIS, geostatistics, and remote sensing have further expanded their applications from small-scale farm plots to larger watersheds while reducing costs and processing time (Djoukbal et al. 2019; Avand et al., 2023). These technologies enhance both spatial and temporal accuracy, allowing for more precise assessments beyond annual averages and

facilitating efficient analysis of multiple factors (Guche et al. 2024). In a GIS environment, a digital elevation model (DEM) is essential for generating and processing key input data for soil erosion modeling, including terrain attributes, slope, slope length and steepness for LS factor. Meanwhile, multi-temporal remote sensing data, such as satellite imagery, offer valuable information on land use and land cover changes (Ajibade et al. 2020). Similarly, GIS, remote sensing, and the RUSLE model were also integrated using rainfall, land cover, and DEM data to estimate an annual soil loss by Ghosh et al. (2022).

Connectivity evaluation has become a crucial component in hydrology, particularly for analyzing runoff and sediment transport within watersheds (Mohammadi et al. 2025). This approach is valuable for assessing sediment sources, soil loss distribution, and sediment transport, contributing significantly to watershed and river channel management (Fryirs et al. 2007). From a sediment delivery perspective, not all eroded materials reach to the watershed outlet, as buffers and barriers disrupt connectivity within drainage networks. Their impact on sediment transport depends on their shape, location, and interaction with the surrounding landscape (Hooke et al. 2021). Therefore, selecting appropriate sediment connectivity indices is critical for accurately evaluating hydrological responses in terms of both water and sediment fluxes. Previous studies have shown a strong positive relationship between sediment yield and sediment connectivity, where higher connectivity facilitating greater transport of eroded material to watershed outlets. This relationship has been confirmed in multiple studies, including observations across 11 sub-watersheds (Arabkhedri et al. 2021), analyses of sediment dynamics within sub-watersheds (Masselink et al. 2016; Fryirs 2013), and event-based assessments integrating sediment connectivity into erosion model (Hao et al. 2022).

However, assessing sediment connectivity remains a challenging task, as it cannot be directly measured, and multiple factors influence its dynamics (Bombino et al. 2020). Various frameworks have been proposed for its evaluation across different climatic and environmental settings (Borselli et al. 2008; Cavalli et al. 2013; Upadhayay et al. 2020). The Index of Connectivity (IC), initially proposed by Borselli et al. (2008) and later refined by Cavalli et al. (2013), integrates upslope and downslope factors to assess sediment connectivity in small watersheds. Recent advancements in coupling IC with soil erosion maps have demonstrated strong correlations with field data under various runoff conditions (Hao et al. 2022). When combined with field assessments, hydrological modeling, and land use change analysis, IC serves as a reliable tool for evaluating sediment transfer between hillslopes and channels across diverse landscapes (Crema and Cavalli 2018).

Further research has linked IC with the Hillslope Sediment Delivery Ratio (HSDR) to enhance erosion models for large watersheds and improve sediment yield predictions at the sub-catchment scale (Vigiak et al. 2012). Recent studies have coupled RUSLE with connectivity theory to enhance sediment transport modeling by integrating spatially explicit pathways contributing to stream networks. This approach improves sediment yield predictions by addressing structural and functional connectivity in watershed processes (Mahoney et al. 2021). Integrating IC with RUSLE allows for assessing spatiotemporal variations in sediment yield due to land use and dam construction (Zhao et al. 2020).

Additionally, the latest studies have demonstrated the effectiveness of machine learning models, such as Artificial Neural Networks (ANN), Convolutional Neural Networks (CNN), Boosted Regression Trees (BRT), Random Forests (RF), Support Vector Machines (SVM), and Classification and Regression Trees (CART), in improving soil erosion susceptibility assessments (Li et al. 2024). Beyond predictive performance, interpretability techniques such as SHapley Additive exPlanations (SHAP) have been applied to sediment modeling to quantify the influence of input variables on suspended sediment concentration, providing insights into the physical processes driving predictions (Lamane et al. 2025a, b). These models, often integrated with Geographic Information Systems (GIS), enhance prediction accuracy by capturing complex relationships between environmental factors, surpassing the limitations of traditional empirical models (Chen et al. 2021; Liu et al. 2023).

Several studies have applied machine learning to runoff and sediment discharge prediction, yet most have focused on either process separately, limiting comprehensive model evaluation. Additionally, ML models often lack transparency and interpretability, making it difficult to assess the contribution of different factors. Addressing these challenges requires evaluating multiple ML models for both runoff and sediment simulation while enhancing model interpretability (Miao et al., 2024). Among ML algorithms, support vector regression (SVR), artificial neural networks (ANN), and extreme learning machine (ELM) are widely used for hydrologic time series forecasting. Studies have shown that ANN performs well in predicting suspended sediment load (Kumar and Tripathi 2022), while SVR demonstrates high accuracy in estimating suspended sediment concentration (SSC). Additionally, k-nearest neighbor regression (kNNR) has been applied as a nonparametric method for handling sediment grade curve variations over time (Miao et al. 2024).

This study aims to evaluate whether sediment connectivity indices, when used as inputs to machine learning algorithms, can reproduce sediment yield estimates generated

by empirical models specifically RUSLE and MUSLE in the absence of observed sediment data. To the best of our knowledge, this is the first study to explore the potential of landscape-derived sediment connectivity metrics as surrogate predictors for event-based sediment yield, without relying on field-measured sediment data. By training ML models exclusively on landscape-derived connectivity metrics and comparing their outputs to RUSLE and MUSLE predictions, we assess the potential of sediment connectivity to serve as a surrogate modeling framework. This approach addresses a critical research gap: while previous studies have applied ML to sediment prediction or combined connectivity metrics with empirical models, few studies have tested whether connectivity metrics alone can replicate the behavior of established erosion models (Crema and Cavalli 2018; Mahoney et al. 2021). The findings provide a novel, data-light framework for sediment yield estimation in ungauged or data-scarce catchments, offering practical applications for watershed management and erosion risk assessment.

Materials and methods

Study area

This study was conducted in the Göksun Çayı Karaahmet sub-catchment (E20A006), a tributary of the Ceyhan River Basin, located in Kahramanmaraş, Türkiye (Fig. 1). The area covers 766 km², with an average elevation of 1315.22 m and a peak elevation of 2957.04 m at Berit Mountain. The region has a continental climate, with hot, dry summers and cold, snowy winters. Mean annual precipitation is 622 mm, and the average temperature is 8 °C. The predominant soil types, classified under the World Reference Base for Soil Resources (WRB) system, include Cambisols, Luvisols, Podzols, and Fluvisols. Land cover consists of forests, rangelands, croplands, urban areas, and bare land (Reis et al., 2016). The Göksun River (60 km) drains into the Ceyhan River, with hydrological records (1955–2000) indicating peak discharge in April (23.69 m³ s⁻¹) and minimum in September (3.34 m³ s⁻¹) (EIA Services, 2011).

Data collection and processing

This study applies RUSLE and MUSLE within a GIS-based framework to estimate event-based sediment yield in the Göksun Çayı Karaahmet sub-catchment. A total of 69 rainfall-runoff events recorded between 2009 and 2019 were used for the analysis, with each event representing a single rainfall episode producing measurable surface runoff, typically lasting from a few hours to one day. Input data were obtained from meteorological stations, soil survey

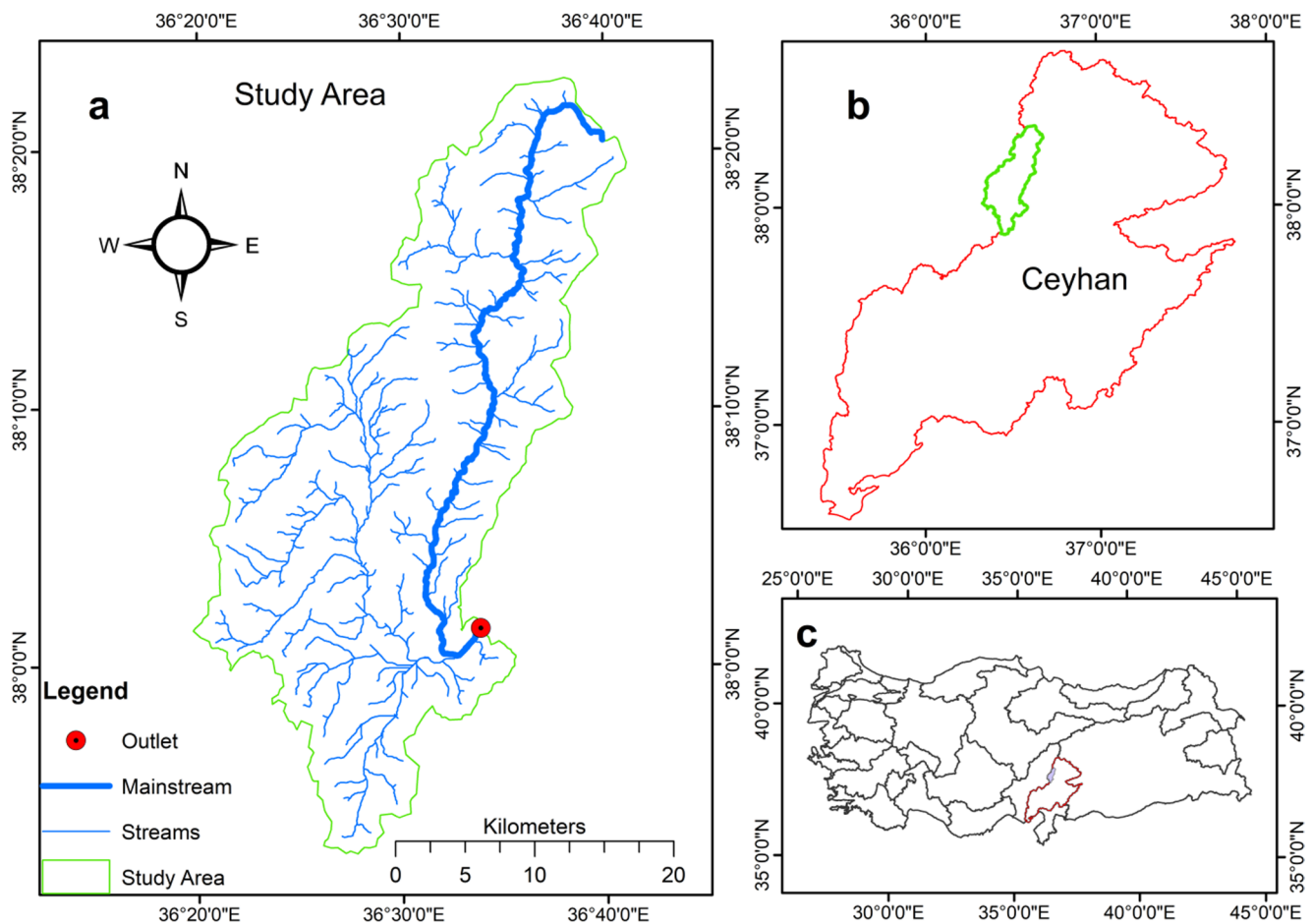


Fig. 1 Study area of the Göksun Çayı Karaahmet sub-catchment. **a** Sub-catchment boundary showing mainstream, streamlines, and outlet, **b** Ceyhan River Basin highlighting the main basin, **c** Location of the basin within Turkey

maps, and topographic data from a Digital Elevation Model (DEM) provided by the Ministry of Urbanization, Environment, and Climate Change of Türkiye. Additionally, Landsat 8 OLI satellite imagery was acquired from the United States Geological Survey (USGS). Spatial data processing and model implementation were conducted using ArcGIS Hydro-tools and Google Earth Engine (GEE).

Both RUSLE and MUSLE incorporate common factors such as K, LS, C, and P factors. RUSLE is a rainfall-driven model that employs the rainfall erosivity factor (R), calculated as the product of total rainfall energy (E) and the maximum 30-minute rainfall intensity (I30) ($E \times I30$). MUSLE, in contrast, is a runoff-driven model estimating sediment yield based on runoff volume and peak discharge. Since RUSLE does not directly predict event-based sediment yield, it was adapted by integrating the SDR, which quantifies the proportion of eroded soil transported to the watershed outlet based on main stream channel slope (Erpul et al. 2013; Das and Sarma 2017).

Multiple sediment connectivity indices were computed to characterize spatial variations in sediment transfer across 196

sub-catchments. A 10 m resolution DEM was used to derive slope, flow direction, flow accumulation, and flow path in a GIS environment, as these parameters drive sediment transport. C factor was integrated as a resistance factor in the connectivity analysis, influencing sediment detachment and retention. The sediment connectivity index (systematically evaluates each sub-catchment's contribution to sediment delivery by incorporating key hydrological and physical parameters).

To evaluate the potential of sediment connectivity metrics to approximate empirical erosion models, machine learning algorithms were trained exclusively on sediment yield outputs predicted by RUSLE and MUSLE, not on observed sediment data. The core objective was to test whether landscape connectivity metrics alone can replicate the behavior of these empirical models. Five ML models were applied: Random Forest (RF), Extreme Gradient Boosting (XGBoost), Support Vector Regression (SVR), Artificial Neural Networks (ANN), and k-Nearest Neighbors (k-NN). The models used twelve key sediment connectivity indices as input features, and their outputs were compared to RUSLE and MUSLE sediment yield predictions. Model

performance was evaluated using statistical metrics including coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE).

By integrating ML with sediment connectivity analysis, this study assesses the capacity of landscape connectivity metrics to serve as surrogates for empirical sediment yield models, capturing complex nonlinear relationships between landscape features and sediment transport processes. Further methodological details on RUSLE/SDR, MUSLE, sediment connectivity indices, and ML implementation are provided in the following sections. The datasets used in this study and the corresponding workflow of data processing and model integration are summarized in Table 1 and illustrated in Fig. 2.

Table 1 Summary of the datasets used for RUSLE and MUSLE-based sediment yield modeling and connectivity indices in the Göksun Çayı Karaahmet sub-catchment

Dataset	Source	Spatial/ Temporal Resolution	Purpose in the Study
Digital Elevation Model (DEM)	Ministry of Urbanization, Environment, and Climate Change of Türkiye	10 m spatial resolution	Used to derive slope, slope length (LS factor) and sub catchment delineation for RUSLE and MUSLE, as well as flow direction, flow accumulation and drainage line for generating connectivity indices.
Soil Map	Ministry of Urbanization, Environment, and Climate Change of Türkiye (General Directorate of Mapping)	1:25,000 scale	Used to determine the soil erodibility factor (K) in the RUSLE and MUSLE models.
Land Use/Land Cover (LULC)	Landsat 8 OLI imagery (USGS)	30 m spatial resolution	Classified to derive the cover-management factor (C) and runoff coefficient (Cr)
Rainfall/Runoff Data	Ministry of Urbanization, Environment, and Climate Change of Türkiye (Meteorology General Directorate) (MGM)	Event-based data (2009–2019)	Used to calculate the rainfall erosivity factor (R) in RUSLE and runoff-peak discharge in MUSLE.
Sub-catchment Boundaries	Derived using ArcGIS hydrological tools	—	Used to calculate area-weighted soil loss and sediment yield at the watershed outlet.

The datasets summarized in Table 1 were integrated within a GIS-based framework to derive model parameters and connectivity indices. The overall methodological workflow is illustrated in Fig. 2.

The final dataset used in this study, comprising 13,524, was derived from continuous spatial data (DEM, LULC) and complete meteorological records. Consequently, the dataset contained no missing values, eliminating the need for imputation. To ensure the models were not unduly influenced by extreme values, both the input features (the calculated SCIs) and the target sediment yield variables were analyzed for outliers. Extreme values, particularly in the RUSLE/MUSLE outputs, were found to correspond to the largest, most erosive rainfall events or steepest sub-catchments. As these represented physically significant extremes within the study area, they were retained to preserve the natural variability of the system. Their potential influence was mitigated by utilizing non-parametric tree-based models (such as XGBoost and Random Forest), which are inherently more robust to outliers compared to linear regression models. Furthermore, regarding the potential issue of overrepresentation or bias towards specific event types, the dataset's structure inherently addresses this by combining \$196\$ distinct spatial units with 69 unique temporal events, ensuring a wide representation of physical and meteorological conditions. To guarantee the robustness, transparency, and reproducibility of the machine learning results, a 5-fold cross-validation strategy was implemented across the entire dataset. This procedure ensures that model training and validation are performed on varied subsets of the data, confirming that the reported performance metrics (R^2 , RMSE, MAE) are generalized and not biased towards any single subset of the data. The inclusion of these detailed data processing and validation strategies significantly enhances the methodological rigor of our study.

Revised universal soil loss equation (RUSLE)

Following the methodological framework, the RUSLE and MUSLE models are employed for event-based sediment yield estimation. The corresponding equations and key parameters are presented below. Also, whole equations are summarized in the Tables 4 and 5 as appendix.

$$Ae = Re \cdot Ke \cdot L \cdot S \cdot Ce \quad (1)$$

$$Ye_{RUSLE} = Ae \cdot SDR \quad (2)$$

Where,

Ae : Event-based soil loss ($t \text{ ha}^{-1} \text{ event}^{-1}$)

Re : Event-based rainfall-runoff erosion factor ($\text{MJ mm ha}^{-1} \text{ h}^{-1} \text{ event}^{-1}$)

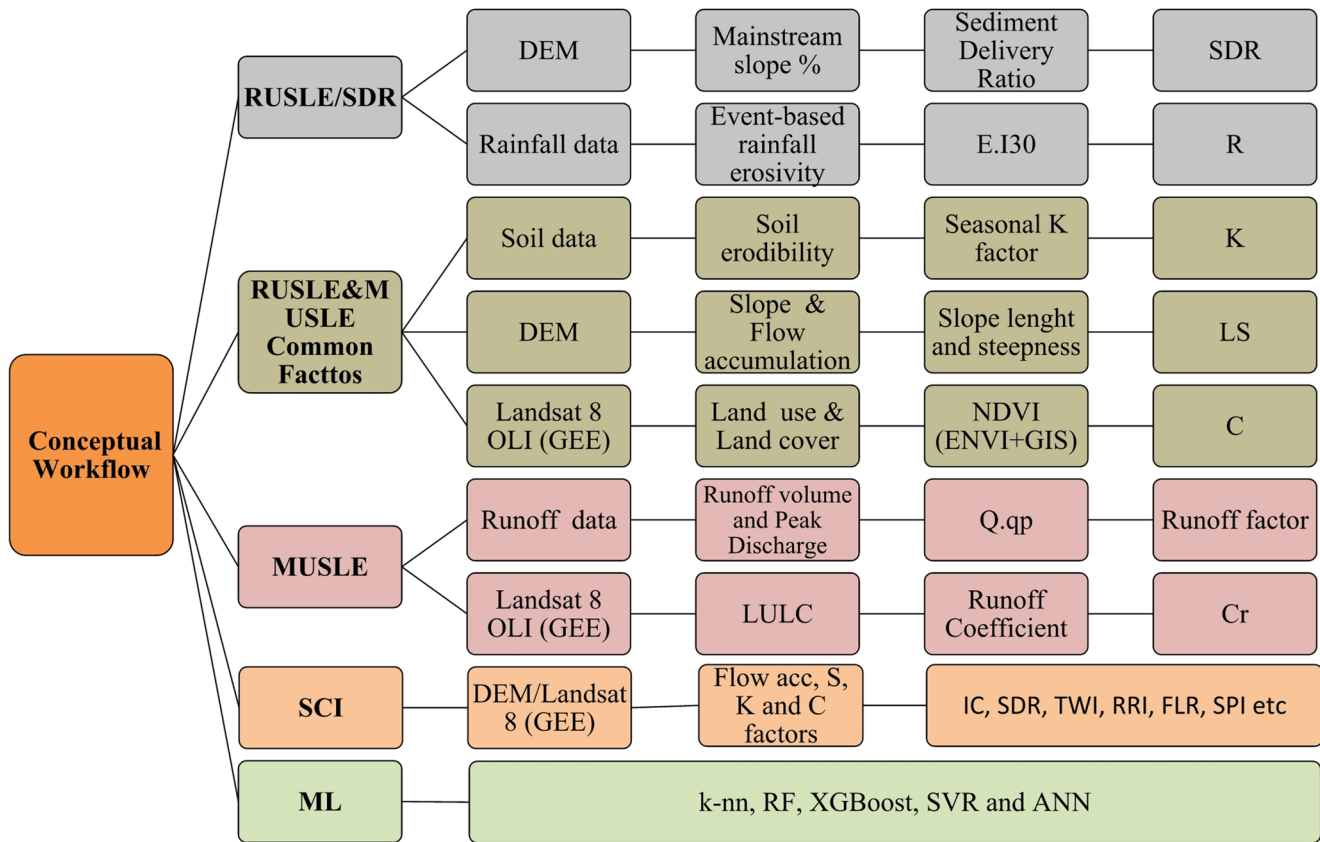


Fig. 2 Workflow for sediment yield prediction using connectivity-driven machine learning surrogates for RUSLE and MUSLE. Model parameters (R, K, LS, C, SDR, Q.qp, and Cr) were derived from rainfall, soil, DEM, and Landsat 8 data; SCI metrics were computed from

DEM and LULC; and ML models (k-NN, RF, XGBoost, SVR, ANN) were trained using SCI as predictors and RUSLE/MUSLE outputs as targets

Ke: Soil erodibility factor ($\text{t ha}^{-1} \text{MJ}^{-1} \text{mm}^{-1}$)

L: slope length factor (dimensionless)

S: slope steepness factor (dimensionless)

Ce: Cover management factor (dimensionless)

Ye: Event-based sediment yield ($\text{t ha}^{-1} \text{event}^{-1}$)

SDR: Sediment delivery ratio (dimensionless)

Note: The support practice factor (*P*) was not considered because no conservation practices were implemented in the study area.

Event-based rainfall erosivity factor (*Re*)

The event-based rainfall erosivity factor (*Re*) varies depending on the geographic location and storm characteristics. The erosion potential is directly influenced by a storm's intensity and duration, with higher intensity and longer duration leading to greater erosive potential (McKague 2023).

In this study, a total of 69 rainfall events over a 10-year period were analyzed to assess event-based rainfall erosivity. The *Re* factor is calculated following the approach defined by Wischmeier and Smith (1978), which involves multiplying the total kinetic energy of a storm by its maximum

30-minute rainfall intensity. This method effectively quantifies the storm's erosive potential (Eq. 3).

$$Re = \frac{1}{N} \sum_{j=1}^n (ET \cdot I_{30})_j \quad (3)$$

Where,

Re: is the rainfall erosivity factor ($\text{MJ mm ha}^{-1} \text{h}^{-1}$)

n: is the total number of storms in the dataset (62 storms)

E_T: is the total kinetic energy of a given storm ($\text{MJ ha}^{-1} \text{mm}$)

I₃₀: is the maximum 30-minute rainfall intensity (mm h^{-1}) for each storm

Event-based soil erodibility factor (*Ke*)

The soil erodibility factor (*Ke*) quantifies the susceptibility of soil to erosion, which is determined by various soil properties, including the relative proportions of sand, silt, and clay, as well as organic matter content (Djoukbala et al. 2018). In this study, the annual *K* factor dataset was obtained from the Dynamic Erosion Modeling and Monitoring

System (DEMIS) (Erpul et al. 2018), developed by the General Directorate of Combating Desertification and Erosion in Türkiye. This dataset includes all necessary soil properties required for K factor calculation.

The K factor was calculated on an annual basis using the equation proposed by Torri et al. (1997, 2002), which has been extensively validated in previous studies in Türkiye. This methodology played a key role in the development of the Türkiye Soil Erosion Map (Akgoz et al. 2022; Deviren Saygin et al. 2011, 2014), where, K represents the soil erodibility factor to erosion, DG is the geometric mean of soil particle size, C is the percentage of clay content, and OM refers to the organic matter content.

$$K = 0.0293.a.B \quad (4)$$

$$a = (0.65 - Dg + 0.24D_G^2) \quad (5)$$

$$B = \exp \left[-0.0021 \frac{OM}{C} - 0.00037 \left(\frac{OM}{C} \right)^2 - 4.02C + 1.72C^2 \right] \quad (6)$$

In this study, event-based K_e factor values were computed using seasonal Eqs. (7)–(13). Initially, the K factor was estimated on a semi-monthly (15-day) basis, incorporating key variables such as monthly temperature fluctuations, periods of soil freezing, and precipitation distribution (Eq. 17) and (Eq. 18). These factors play a crucial role in soil erodibility by influencing the duration and intensity of conditions that promote erosion. A critical aspect of these equations is the identification of periods corresponding to maximum (t_{max}) and minimum (t_{min}) soil erodibility throughout the calendar year, which reflects seasonal variations in soil vulnerability (Kiciman et al., 2024).

$$\text{If } t_{max} < t_i < t_{min} \text{ then} \quad (7)$$

$$K_i = K_{max} \left(\frac{K_{min}}{K_{max}} \right)^{\frac{t_i - t_{max}}{\Delta t}} \quad (8)$$

Where, K_i represents the soil erodibility factor at any time t_i (in calendar days), K_{max} and K_{min} are the soil erodibility factors at times t_{max} and t_{min} , respectively. Δt represents the growing season or frost-free period (≤ 183 d); and T_{av} is the average daily temperature.

$$\text{If } t_i < t_{max} \text{ or } t_i > t_{min} \text{ then} \quad (9)$$

$$K_i = K_{min} \exp [0.009 (t_i - t_{min} + 365\delta)] \quad (10)$$

$$\text{If } (t_i - t_{min}) \leq 0, \delta \text{ and if } t_i - t_{min} > 0, \delta = 0 \quad (11)$$

$$\text{If } T_{av} \leq -2.7^\circ C, K_i = t_{min} \quad (12)$$

$$\text{Where } K_i < K_{min} \quad K_i = K_{min} \quad (13)$$

The values of K_{max} , K_{min} , t_{max} and t_{min} to be used in the above relationships are determined with Eqs. (14)–(16), using event-based rainfall energies ($RUSLE-R$, EI) (Deviren Saygin 2021).

$$\frac{K_{max}}{K_{min}} = 8.6 - 0.19R \quad (14)$$

$$\frac{K_{max}}{K_{nom}} = 3.0 - 0.005R \quad (15)$$

$$T_{max} = 154 - 0.44R \quad (16)$$

Semi-monthly R factor distributions (15-day EI %) were computed using Eq. (17) and Eq. (18). This approach divided the monthly R factor into two 15-day periods based on percentage erosivity index (EI) values (Kiciman et al., 2024).

$$EI_1 = EI \frac{0.75 (EI - 1) + 0.25(EI + 1)}{(EI - 1) + (EI + 1)} \quad (17)$$

$$EI_2 = EI \frac{0.25 (EI - 1) + 0.75(EI + 1)}{(EI - 1) + (EI + 1)} \quad (18)$$

Where,

EI : the value for the desired month

$EI-I$: the value for before of the desired month

$EI+I$: The value for the after of the desired month

EI_1 : the value calculated for the first 15 days of the desired month

EI_2 : The value calculated for the second 15 days of the desired month

Slope length and steepness (LS-factor)

The LS factor represents the combined influence of slope length and steepness on soil erosion (Chang et al., 2016). It is defined as the ratio of soil loss from a given slope length and gradient to that from a standard unit plot (a 22.1 m long fallow plot with a 9% slope under up-and-down slope clean tillage) (Wischmeier and Smith 1978; Renard et al. 1997). As slope length and steepness increase, total surface runoff and its velocity also rise, intensifying soil loss (Handique et al. 2023). The LS factor is computed by multiplying the slope length and steepness values (Eq. 19). In this study, a 10-meter resolution Digital Elevation Model (DEM) was utilized to estimate LS values. These values were derived based on flow accumulation (χ) and slope degree (θ) using the Raster Calculator in ArcGIS. The flow accumulation

term (χ) represents the volume of water flow contributing to erosion along the slope length, while slope steepness (θ) quantifies the gradient of the terrain (Lee 2004). The LS factor was computed using Eq. (19) proposed by Moore and Burch (1986a, b), and cell-based average LS values for each sub-catchment were obtained using the Cell Statistics tool in ArcMap.

$$LS = \left(\frac{\chi \eta}{22.13} \right)^{0.4} \cdot \left(\frac{\sin \theta}{0.0896} \right)^{1.3} \quad (19)$$

where, χ ; flow accumulation derived from DEM using a GIS accumulation algorithm, η ; cell size, and θ ; slope steepness in degrees.

Event-Based land Use/Land cover (C_e)

In this study, event-based C_e factors were derived from Landsat 8 Operational Land Imager (OLI) satellite imagery. A selection of Landsat 8 images was made based on their temporal relevance to the event dates, considering appropriate months and seasons. The Normalized Difference Vegetation Index (NDVI) was processed to assess vegetation cover. The C factor was then calculated using Eq. (20) proposed by Van der Knijff et al. (2000), with coefficients $\alpha=2$ and $\beta=1$. These coefficients were specifically chosen based on their empirical validation for Mediterranean climatic conditions, effectively representing the influence of land cover on soil erosion. The coefficient ($\alpha=2$) accounts for the pronounced impact of vegetation changes in Mediterranean environments, while $\beta=1$ adjusts for the specific relationship between vegetation cover and erosion processes in these conditions. Cell-based average C values were computed using GIS cell statistic tools and applied to each sub-catchment within the study area.

$$C = \exp \left[-\alpha \cdot \frac{NDVI}{\beta - NDVI} \right] \quad (20)$$

Sediment delivery ratio (SDR)

The sediment delivery ratio quantifies the proportion of eroded soil (A) that is transported to the watershed outlet and is integrated with the RUSLE model to estimate sediment yield. SDR values are always less than 1, as sediment yield is invariably lower than total soil erosion. While SDR is influenced by multiple factors, including topography, drainage basin size, and vegetation cover (Kumar & Pandey, 2019). In this study, SDR values were estimated

solely based on the slope of the main stream within each sub-catchment, following the method proposed by Williams and Berndt (1977). This approach ensures that variations in slope across all sub-catchments were captured; however, it does not explicitly account for other factors that may affect sediment transport efficiency, such as basin size, micro-topography, or vegetation cover.

$$SDR = 0.627SLP^{0.403} \quad (21)$$

where, SDR refers to the Sediment Delivery Ratio, and SLP is the slope of the mainstream channel within of a sub-catchment.

This slope-based estimation provides a simple and consistent method for integrating SDR with RUSLE in the absence of detailed data on other controlling factors.

Modified universal soil loss equation model (MUSLE)

In this study, the Modified Universal Soil Loss Equation (MUSLE) (Williams 1975) was applied to estimate event-based sediment yield. Unlike RUSLE, which relies on rainfall erosivity (R), MUSLE replaces it with a runoff-based factor incorporating runoff volume and peak discharge. These parameters were computed using the unit hydrograph triangular method and the rational method, both widely used for small to medium-sized catchments with available runoff and rainfall data. Runoff volume and peak discharge values were derived based on the time of concentration concept, ensuring compatibility with the hydrological characteristics of the study area's sub-catchments (Wischmeier and Smith 1978). The event-based sediment yield was then estimated using Eq. (22).

$$Y = \frac{a(Q \times qp)^b K \times LS \times S \times C \times P}{100A} \quad (22)$$

Where,

Y : Sediment yield caused by the storm event ($t \text{ ha}^{-1}$)

Q : Runoff volume (m^3)

q_p : Peak discharge ($m^3 \text{ s}^{-1}$)

A : Drainage area (km^2)

The factors K , LS , C , and P are defined as in the RUSLE equation for the specific storm event. In Eq. (24), A is the drainage area in km^2 , multiplied by 100 to convert it to hectares. The resulting sediment yield will be in tons per hectare ($t \text{ ha}^{-1}$) (Shi et al. 2022). Equation (23) expresses the runoff factor for the MUSLE model:

$$MUSLE_{Runoff} = 11.8(Q \times qp)^{0.56} \quad (23)$$

Runoff volume (Q)

The runoff volume is calculated using the Triangular Unit Hydrograph method, developed by the Soil Conservation Service (SCS) (Eq. 24). This method assumes that the sub-catchments have a triangular shape, with the base equal to the sum of the time to peak and the time of recession, and the height equal to half of the peak discharge.

$$Q = \left(\frac{qp}{2} \right) \times (Tp + Tr) \quad (24)$$

where, Q represents the runoff volume (m^3), q_p refers to the peak runoff discharge ($m^3 s^{-1}$), T_p is the time to peak (minutes), T_r indicates the time of recession (minutes).

Runoff discharge rate (qp)

The runoff discharge rate ($m^3 s^{-1}$) for each micro-catchment is calculated using the rational method as proposed by Mulvaney (1851). In this method the surface flow coefficient is initially determined using land use land cover classification; along with the time of concentration using Kirpich method (Eq. 25) (Kirpich 1940) is employed in the following manner to calculate the flow discharge rate.

$$q_p = \frac{CrIA}{360} \quad (25)$$

Where, q_p represents the surface flow rate ($m^3 s^{-1}$), I refers to the rainfall intensity for a specific duration ($mm \text{ hour}^{-1}$), A indicates catchment area (ha) and C_r represents a surface flow coefficient that is dimensionless.

• Runoff Coefficient (Cr).

The runoff coefficient for each land use type was determined using Landsat8 OLI images from 2014, 2016 and 2018. The images were classified using a supervised classification method to create a detailed *LULC* map, categorizing areas into forest, agriculture, pasture, settlement, and bare land. The classified map was then converted to raster format and integrated with a slope spatial map to extract slope percentage for each land use type using ArcGIS spatial analysis tool. Runoff coefficient values were assigned based on the surface flow coefficient standard proposed by Schwab (1981), considering both slope percentage and soil texture within each sub-catchment. The weighted average runoff coefficient for each sub-catchment was calculated based on the proportional area of each land use type using Eq. (26).

$$Cr = \frac{1}{A} \sum_{i=1}^n (Cr_i \times A_i) \quad (26)$$

Where, C_r refers to the weighted runoff coefficient for a sub-catchment; A is the total area of a sub-catchment. C_{ri} is the runoff coefficient for each land use type i , and A_i is the area that occupied by a land use type i .

• Time of concentration (Tc).

The time of concentration, measured in minutes, is a key parameter in calculating the peak flow rate. To determine the peak flow using the Rational Method, it is first necessary to calculate the time of concentration. Several methods have been proposed for this calculation. In this study, the Kirpich (1940) equation is employed to estimate the time of concentration, as expressed in Eq. (27).

$$Tc = \frac{0.01947L^{0.77}}{S^{0.385}} \quad (27)$$

Where, T_c represents the time of concentration in minutes. L represents the maximum length of runoff in a given watershed, and is expressed in meters. S is the slope of the watershed expressed as $\Delta H/L$, where ΔH is the difference in elevation between the initial flow length point and the endpoint in the watershed.

Weighted average sediment yield for the entire watershed

The area-weighted average method was employed in both models to accurately estimate sediment yield for the entire watershed. This approach ensures that the contributions of individual sub-catchments are proportionally represented based on their respective areas. The study area consists of 196 sub-catchments, and the weighted average sediment yield for each storm event was computed using Eq. (28). Consequently, 69 sediment yield values were obtained for Ye-RUSLE and 69 for Ye-MUSLE, corresponding to the 69 erosive rainfall events analyzed.

$$Weighted_{Ave} = \frac{\sum_{i=1}^n (Y_{ei} \times A_i)}{\sum_{i=1}^n A_i} \quad (28)$$

Where,

Y_{ei} : Event-based Sediment yield from sub-catchment i .

A_i : Area of sub-catchment i

n : Total number of sub-catchments within the st

Connectivity analysis in sediment yield prediction

The study area was a large-scale watershed, subdivided into 196 sub-catchments where the Sediment Connectivity Index (SCI) was evaluated. The SCI for each sub-catchment was computed by integrating multiple spatial datasets to ensure a comprehensive assessment of sediment dynamics within the study area.

A Digital Elevation Model was used to derive slope, flow direction, flow accumulation, and flow path in a GIS environment. These parameters are crucial for computing the Connectivity Indices within the watershed. Slope, flow accumulation, and flow length are key drivers of sediment transport, with steeper slopes facilitating higher runoff velocities and increased sediment mobility. In the connectivity analysis, weighting factors such as land use/land cover factor were incorporated as resistance factors. Multiple connectivity indices were calculated to evaluate the contribution of each sub-catchment to sediment transport across the watershed.

Index of connectivity (IC)

The connectivity index, developed by Borselli et al. (2008), is utilized to assess structural connectivity and describe the transport of sediment from its source to its deposition area. The IC calculation is divided into upslope and downslope component (Eq. 29).

$$CI = \log_{10} \left(\frac{D_{up}}{D_{dn}} \right) \quad (29)$$

Where,

D_{up} : representing the upstream factor, which integrates flow distance (Flow length-downstream) to the outlet or stream with the slope mean and land use/land cover mean.

D_{dn} : represents the contributing downstream area, which is related to flow accumulation and combined with the slope percentage, land use/land cover mean.

The D_{up} was calculated using Eq. (30):

$$D_{up} = W \times S \times \sqrt{A} \quad (30)$$

W : Average weighting factor of the upslope contributing area, it was determined using the C factor in RUSLE, (dimensionless).

S : Average slope gradient of the upslope contributing area (m/m).

A : Upslope contributing area (m²).

The D_{dn} was calculated using Eq. (31):

$$D_{dn} = \sum_i \frac{d_i}{W_i \times S_i} \quad (31)$$

d_i : flow length to the downstream main channel along the i^{th} cell

W_i : Weighting factor related to the vegetation cover and soil erodibility factor of the i^{th} cell

S_i : Slope gradient of the i^{th} cell (m/m).

The Connectivity Index (CI) is defined within the range $(-\infty, +\infty)$ and was calculated using Eq. (32):

$$CI = \log_{10} \left(\frac{D_{up}}{D_{dn}} \right) = \log_{10} \left(\frac{W \times S \times \sqrt{A}}{\sum_i^n \frac{d_i}{W_i \times S_i}} \right) \quad (32)$$

Stream power index (SPI)

The Stream Power Index (SPI) is a topographical model designed to quantify the erosive potential of runoff within a watershed. In this study, the Stream Power Index (SPI) is calculated using GIS tools specifically the Raster Calculator integrating flow accumulation and slope degree (Eq. 33). (Moore et al. 1991)

$$SPI = A \cdot \tan(\beta) \quad (33)$$

Where, A , represents the flow accumulation, which quantifies the water volume contributing to specific location in the watershed, $\tan(\beta)$, represents the tangent of the slope angle at a given point, where the steeper slopes lead to increased water velocity, thereby enhancing erosive potential of the flow.

Proximity to outlet

Proximity to the outlet represents the relative distance sediment must travel from its origin within a sub-catchment to the main stream outlet, influencing sediment connectivity. It was computed using GIS-based analysis, where the centroid of each sub-catchment was identified, and its shortest distance to the outlet was determined. The proximity index (P_i) was calculated using Eq. (34).

$$P_i = 1 - \frac{d_i}{\max(d)} \quad (34)$$

Where, P_i , represents Proximity index for sub-catchment i , d_i , is the distance from the centroid of sub-catchment (i) to the outlet, while $\max(d)$, is the maximum distance among the all sub-catchments. This equation was implemented using the Field Calculator in Arc-GIS, with values closer to (1) showing higher proximity to the outlet.

Topographic wetness index (TWI)

The Topographic Wetness Index is a terrain-based index that estimates landscape wetness, providing insight into areas prone to water accumulation and potential sediment transport (Eq. 35) (Atalla et al. 2024).

$$TWI = \ln \left(\frac{A}{\tan(\beta)} \right) \quad (35)$$

Where, A is the flow accumulation, representing the water volume contributing to specific location in the watershed ($\text{m}^2 \text{m}^{-1}$) and $\tan(\beta)$ is the tangent of the slope angle at a given point.

Relative relief index (RRI)

This approach captures the influence of topographic relief and hydrological pathways on the potential for sediment transport. It is defined as the ratio of the elevation difference to the longest stream length within a sub-catchment (Eq. 36).

$$RRI = \frac{E_{max} - E_{min}}{L_{max}} \quad (36)$$

Where, E_{max} and E_{min} represent the maximum and minimum elevation, respectively, and L_{max} is the longest flow length within each sub-catchment.

Topographic connectivity index (TCI)

Topographic Connectivity Index is a measure of sediment connectivity that integrates contributing area (A) and slope (S) to estimate the ease with which sediment is transported down slope (Eq. 37).

$$TCI = A \cdot S \quad (37)$$

Where:

A : Contributing area (m^2)

S : Slope (%)

Drainage density (Dd)

Drainage density is a fundamental hydrological and geomorphological factor, calculated as the total length of streams and rivers divided by the drainage basin area (Eq. 38). This metric provides valuable insights into the degree of landscape dissection and runoff potential within a watershed.

Horton (1932) noted that drainage density is closely associated with other measures of landscape dissection.

$$D_d = \frac{\sum L}{A_{basin}} \quad (38)$$

$\sum L$: Total length of drainage channels or drainage lines (km^2)

A_{basin} : Drainage basin area (km^2)

Watershed shape factor (WSF)

Watershed Shape Factor is the ratio of the maximum watershed length to the watershed area, describing the compactness and elongation of the watershed, which affects sediment transport efficiency. The WSF was calculated using Eq. (39).

$$WSF = \frac{L_{max}}{A} \quad (39)$$

L_{max} : Maximum watershed length or longest flow length within a watershed (km).

A : Watershed area or drainage area (km^2).

Flow length ratio (FLR)

The Flow Length Ratio is a dimensionless index representing the proportion of the longest flow path within a given sub-catchment relative to the longest flow path in the entire watershed (Eq. 40).

$$FLR_i = \frac{L_i}{L} \quad (40)$$

FLR : Flow Length Ratio for sub-catchment i (dimensionless)

L_i : Longest flow length in sub-catchment i (km)

L : Longest flow length in the entire watershed (km)

Sediment delivery ratio (SDR) as connectivity indices in ML

The Sediment Delivery Ratio (SDR) also serves as a landscape connectivity indicator reflecting the efficiency of sediment transfer through the watershed. In this study, two empirical formulations were used to derive SDR-based connectivity indices: one based on slope (Eq. 41) and another based on drainage area (Eq. 42). These SDR indices, along with other topographic and hydrological factors, were incorporated into the machine learning (ML) framework to represent sediment transport potential and enhance predictive accuracy.

$$\text{SDR} = 0.627 (\text{SLP})^{0.403} \quad (41)$$

Where, SDR is the Sediment Delivery Ratio and *SLP*: slope percentage of the main flow length

$$\text{SDR} = 0.51 \times A^{-0.11} \quad (42)$$

Where, SDR is the Sediment Delivery Ratio and *A* is the drainage Area (km²).

Note: The slope-based SDR was applied in RUSLE to estimate sediment yield, whereas both slope- and area-based SDRs were used exclusively as sediment connectivity indices in the ML models and were not directly applied in the RUSLE or MUSLE yield computation.

Drainage area (DA)

Drainage Area used a connectivity that affects sediment delivery. the Sub-catchment area (km²) was also incorporated into the composite index, affecting sediment and flow contribution within the watershed.

Normalization of connectivity indices

To ensure comparability among the various connectivity indices, all indices were normalized to a common scale. The min-max normalization approach was applied to rescale the values between 0 and 1, as shown in Eq. (43).

$$X' = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \quad (43)$$

Where,

X': Normalized value

X: Original value

X_{min}: Minimum value in the dataset

X_{max}: Maximum value in the dataset

Machine learning (ML) models

Machine learning algorithms, widely used for their reliability across research domains, were applied to estimate RUSLE and MUSLE parameters. The selected models XGBoost, Random Forest, k-NN, SVR, and Artificial Neural Networks utilized 12 key factors influencing sediment transport dynamics within a watershed. To ensure robust evaluation, k-fold cross-validation was employed, where the dataset was divided into *k* subsets. Each subset served as a test set once, while the remaining *k-1* subsets were used for training. For instance, in 5-fold cross-validation, the model was trained and tested five times, each time with a different subset as the test set.

Extreme gradient boosting (XGBoost) algorithm

XGBoost algorithm is based on gradient boosting type of ML algorithm is an effective one. Like most of the ML algorithms, XGBoost also can handle with both classification and regression studies. The idea of XGBoost is the improvement of traditional gradient boosting algorithm (Chen and Guestrin 2016). XGBoost algorithm optimizes parallel computation. The algorithm corrects the errors of each previous tree. Thus, it is more accurate and faster. The optimization is conducted by a second-order Taylor expansion of the loss function. The following equation is the general objective function for XGBoost regression.

$$\mathcal{L}^{(t)} = \sum_{i=1}^n \mathcal{A}(y_i, \hat{y}_i) + f_t(x_i) + \Omega(f_t) \quad (44)$$

Where,

$\mathcal{L}^{(t)}$ is the objective function of iteration *t*,

$\sum_{i=1}^n \mathcal{A}(y_i, \hat{y}_i)$ is the loss function,

$f_t(x_i)$ is the new tree,

$\Omega(f_t)$ is the regularization term.

Random forest regression algorithm

RF algorithm is a type of decision tree and based on ensemble learning method. RF, introduced by Breiman in 2001, is an advanced version of the bagging method (Breiman 1996). The algorithm creates multiple decision trees to predict the final outcome. Because of many trees are used to reach the outcome, the name of the algorithm is called as “forest”. There are two significant parameters for the algorithms; the number of trees (*N*) and the number of variables (*m*) used for each node. Each tree is created randomly selected parameters. Therefore, the algorithm called as *Random Forest*. At the end of the process, the results from all *N* trees are aggregated (Eq. 46), and the variable importance scores are calculated to construct the final model.

$$\bar{y} = \frac{1}{M} \sum_{i=1}^M h_j(x) \quad (45)$$

Where,

$\bar{y}$ is the final predicted value,

M is the number of decision trees,

$h_j(x)$ is the *j*-th decision tree.

K-nn algorithm

The K-nearest neighbors (k-nn) algorithm is even though widely used for classification problems it is also applied in regression analysis as well. K-nn algorithm predicts value

based on the feature similarity principle of the value of a new data point. The k-nn consist of three main steps: calculation of the distances between data points, selection of the K closest points, and averaging these points to generated the final prediction. The main disadvantage of this algorithm is the usage of high memory and storage. On the other hand, the k-nn is requires only a few parameters: k value and a distance metric. K-nn calculations uses different distance metrics. Euclidian distance, which is computed as the square root of the sum of squared differences between two points, most commonly used distance for k-nn algorithm. The Manhattan distance, which sums the absolute differences between points, and the Minkowski distance, which generalizes both Euclidean and Manhattan metrics, are the other two distances used in k-nn predictions.

Support vector regression algorithm

Support Vector Regression (SVR) is an adaptation of Support Vector Machines (SVM) for regression tasks, achieved by modifying the loss function (Gunn 1997). Unlike classification, where SVM aims to find a hyperplane that best separates data points into distinct categories, SVR seeks to fit a function that best approximates the target values while allowing a certain margin of tolerance. SVR models the relationship between input variables and the target variable using a linear hyperplane function, which is defined as follows:

$$y = wx + b \quad (46)$$

The objective is to minimize the weight vector w to obtain the best prediction. This can be achieved by the optimization of the following function.

$$\min_w = \frac{1}{2} ||w||^2 \quad (47)$$

Artificial neural network algorithm

ANN is a ML algorithm inspired by the structure of the human brain. Biological neurons are modeled mathematically. ANN consist of three main layers: the input layer, hidden layer(s), and output layer. The hidden layer processes the information and transfers it to the output layer, which generates the model's predictions. The basic components of ANNs resemble biological neurons: synapses correspond to weights, the nucleus to the activation function, and dendrites to the summation function. The learning process of ANNs occurs through weight updates, optimized using training data.

The main challenge of the ANN during the training of the data is the overfitting and underfitting issues. If the model learns the training data very well than the overfitting occurs. In this case the preformation of the model is poor due to low generalization. On the contrary, underfitting occurs if the models fail to learn the data, and it yields poor predictive performance. In these cases, to ensure successful predictions for both training and test data, noise addition to data or dropout of the data can be performed. Eventually, an ideal model should achieve good fitting.

Performance evaluation metrics

ML algorithms performances are evaluated by widely used performance metrics: Means Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and the Coefficient of Determination (R^2). These metrics are essential to interpret the algorithm's performance. By observing and comparing these metrics, one can be analyzed how the model predictions are close the actual ones.

Mean squared error (MSE)

The Mean Squared Error is a widely used metric to evaluate particularly in regression problems. MSE calculates the average of the squared differences between the observed and the predicted values by any model. The lower the MSE, the better the predictions of the model. This metric is sensitive to large errors due to the squaring of the residuals, making it particularly useful for identifying models that may struggle with outliers. The equation of the MSE expressed as follows:

$$MSE = \frac{1}{N} (y_i - \hat{y}_i)^2 \quad (48)$$

Where,

N is the number of test samples,

y is the observed values,

$\hat{y}$ is the predicted values.

Root mean square error (RMSE)

RMSE is one of the most popular metrics for assessing the performance of the regression models. The RMSE is the square poor of the MSE. RMSE measures the standard deviation of the residuals. RMSE gives a sense of how much the model's predictions deviate from actual values. Lower RMSE means better performance of the models. The RMSE equation is given in Eq. (50).

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (49)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (51)$$

Mean absolute error (MAE)

The MAE penalizes larger errors more heavily due to the squaring of residuals, in contrast to RMSE. The MAE is more straightforward measure of the average absolute differences between the observed and predicted values. The MAE represents the average magnitude of the errors in the same units as the original data. Therefore, it is simpler to evaluate the models. If the MAE is large it means that the model have large errors, in contrast if the model's predictions is closer to actual values the MAE is smaller. The equation of the MAE is shown in Eq. (51).

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (50)$$

Coefficient of determination R-Square

The coefficient determination widely known and called as R^2 is performed to evaluate the performance and fit of regression models. R^2 is often used in regression analysis to assess how effectively the model's predictions correspond to the actual data. An R^2 value of 1 indicates a perfect fit, meaning the model predicts the target variable without any error. An R^2 value of 0 suggests that the model does not explain any of the variability in the target variable, and the predictions are no better than simply predicting the mean of the observed values Eq. (52).

Results

Event-based rainfall erosivity (R factor)

The event-based rainfall erosivity shows spatial variation across the study area, influenced by differences in rainfall intensity and frequency. Figure 3 illustrates the variations in rainfall erosivity patterns recorded over the 10-year study period. A total of 69 rainfall events over 2009–2019 were included in the analysis. Each event was defined as a rainfall occurrence separated by at least 24 h of dry conditions, following standard hydrological criteria. In other words, each event is treated as an independent rainstorm for modeling event-based sediment yield. This factor was derived from data collected at a single rain gauge within the study area. Due to the absence of multiple rainfall stations, rainfall was assumed to be uniformly distributed across the watershed. The event-based rainfall erosivity factor was computed using the $E \times I30$ method (Eq. 3), which considers the total kinetic energy of rainfall and the maximum 30-minute intensity for each of the 69 recorded events.

Soil erodibility factor (K) distribution

The spatial distribution of the event-based soil erodibility factor (K) across the study area is presented in Fig. 4.

Fig. 3 Temporal variation of event-based rainfall erosivity (R factor) for 69 rainfall events over 2009–2019 in the Göksun Çayı Karaahmet sub-catchment, calculated using the $E \times I30$ method

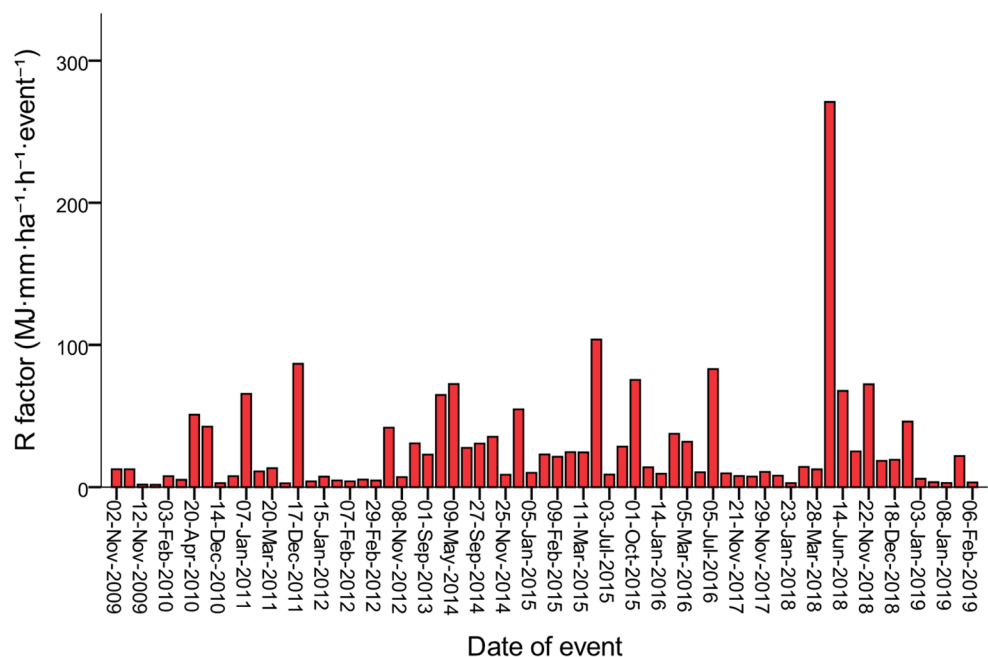


Fig. 4 Spatial distribution of the average soil erodibility factor (K) across the sub-catchments of the Göksun Çayı Karaahmet watershed

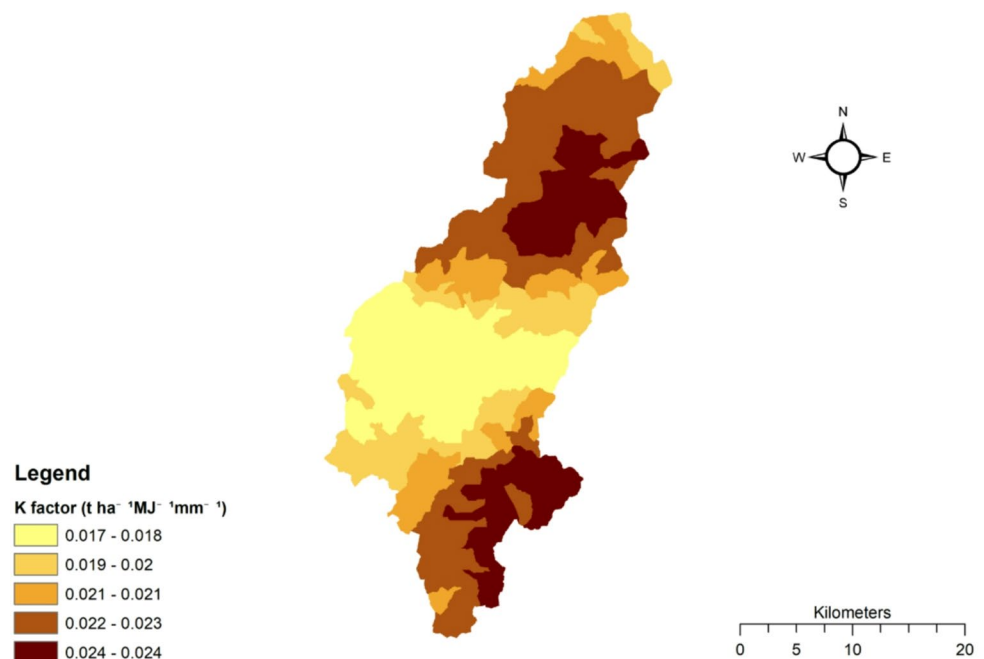
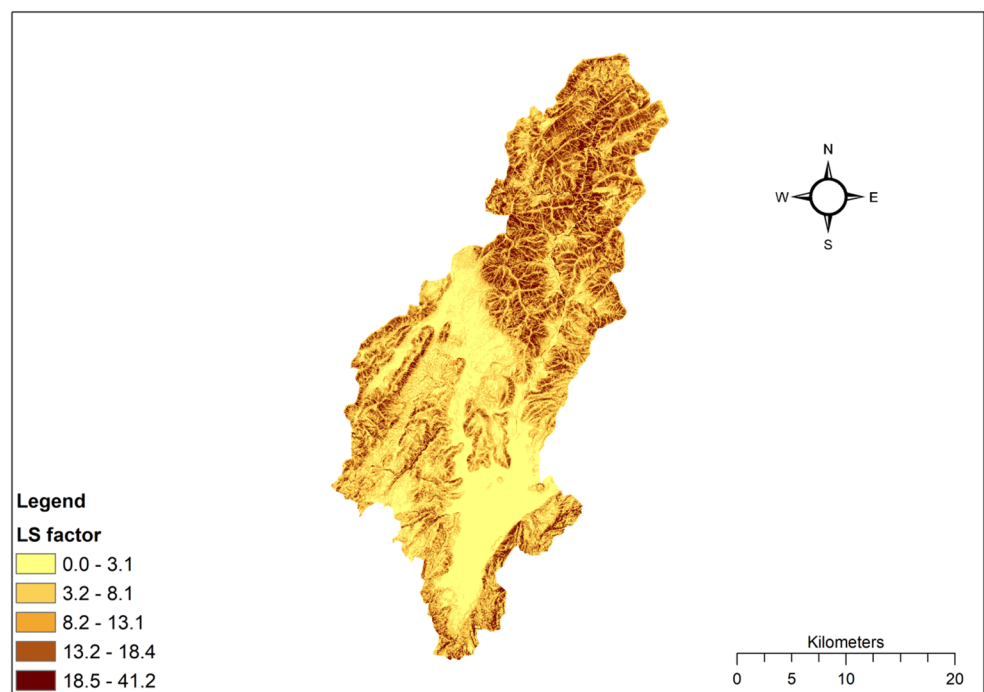


Fig. 5 Spatial distribution of the LS factor across sub-catchments in the study area, showing areas of high and low slope-length influence on soil erosion



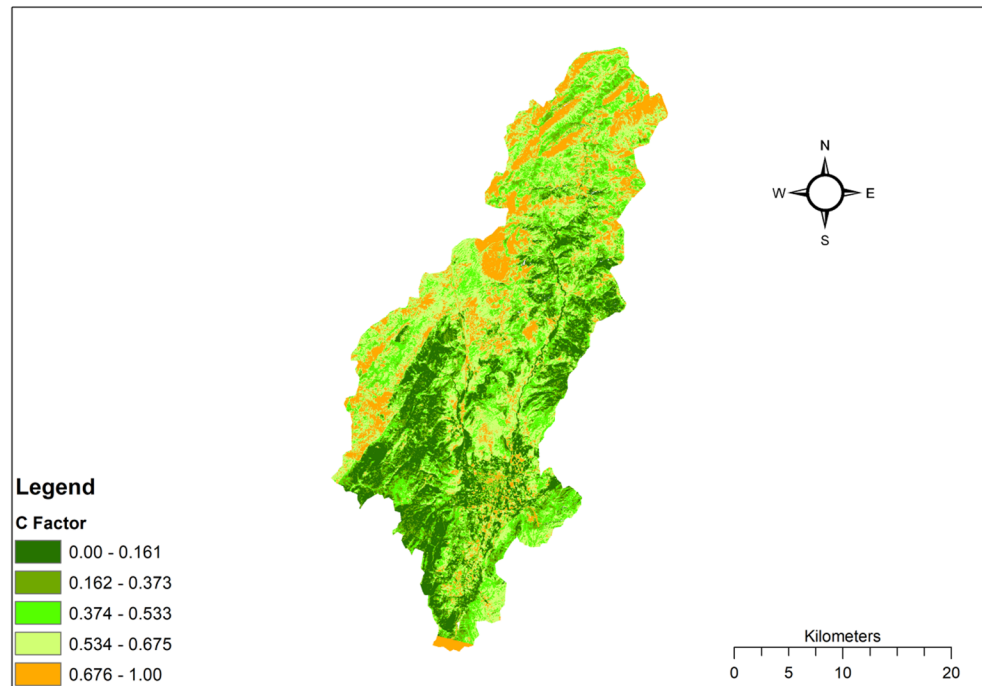
The map was generated by averaging event-based K values across all sub-catchments to illustrate their spatial variation. The values range from 0.017 to 0.024, indicating differences in soil susceptibility to erosion. Lighter shades represent areas with lower K values (0.017–0.018), suggesting more erosion-resistant soils, while darker shades indicate higher K values (0.022–0.024), reflecting soils that are more prone to detachment and transport by rainfall and runoff. These variations highlight the influence of soil texture, organic

matter content, and structure on erosion potential across the watershed.

Slope length and steepness (LS factor)

The spatial distribution of the LS factor was mapped using a GIS framework (Fig. 5), with values ranging from 0.0 to 41.182 across the study area. Average cell-based LS values were calculated for each of the 196 sub-catchments using

Fig. 6 A representative event-based land use/land cover factor (Ce) map, illustrating the spatial distribution of event-based C factor values across the study area, derived from satellite imagery



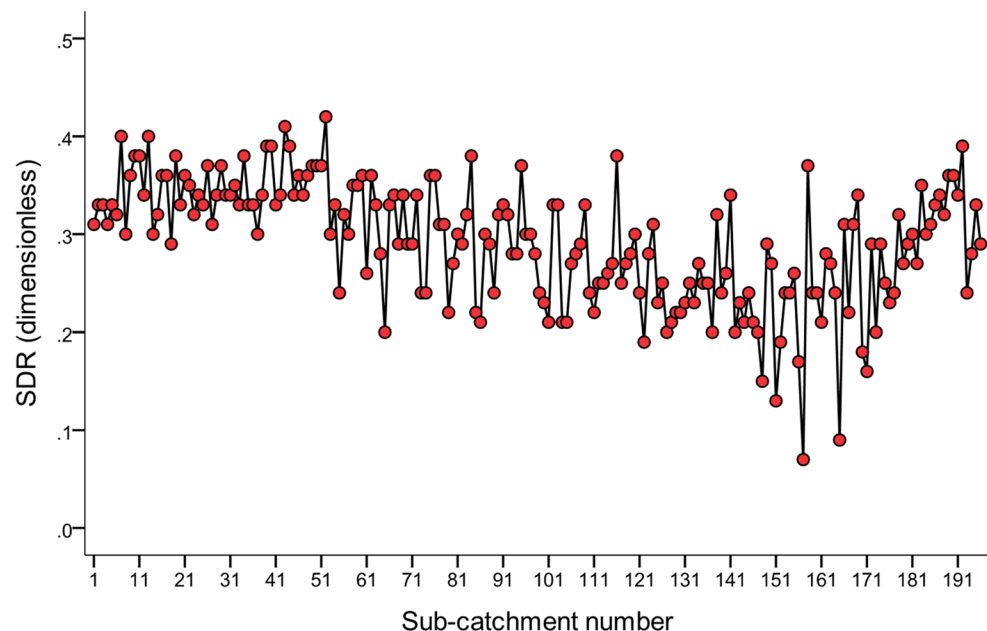
Eq. (19). The highest average cell-based LS value within a sub-catchment was 15.05, while the lowest was 0.06. The majority of LS values across sub-catchments ranged between 5 and 12.

Event-based land use/land cover factor (Ce)

A representative event-based C factor map (Fig. 6) illustrates the spatial distribution of C values across the study

area. The event-based C factor values range from 0.0 to 1.0 at the cell level, while sub-catchment averages vary between 0.155 and 0.869. These variations reflect changes in vegetation cover and land use conditions during different events, influencing soil erodibility and erosion potential. Higher C factor values indicate areas with minimal vegetation cover, making them more susceptible to erosion, whereas lower values correspond to well-vegetated regions that provide greater soil protection.

Fig. 7 Variation of the sediment delivery ratio in relation to the slope percentage of the mainstream within each sub-catchment



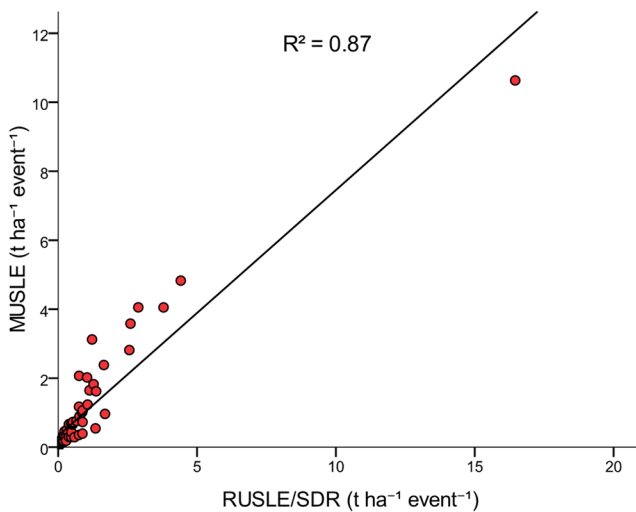


Fig. 8 Regression curve comparing event-based sediment yield predicted by MUSLE and MUSLE/SDR models (t ha^{-1}), showing the relationship between predicted values

Sediment delivery ratio (SDR)

The spatial variability of the sediment delivery ratio across the study area (Fig. 7) was calculated based on the slope percentage of the mainstream within each sub-catchment (Eq. 21). The results show that SDR values range from 0.072 to 0.416, with an average of 0.292, indicating variations in sediment transport potential from sub-catchments to the outlet. These variations emphasize the influence of slope and topographic features on sediment transport processes throughout the study area. It should be noted that this slope-based SDR was applied in MUSLE for sediment yield estimation, whereas a slope- and area-based SDR was later used as a connectivity index in the machine learning models to represent sediment transport potential.

Comparison of event-based predicted sediment yield using MUSLE and MUSLE/SDR

The event-based sediment yield predictions from MUSLE and MUSLE/SDR display noticeable differences due to the distinct conceptual approaches of the two models. MUSLE incorporates runoff volume and peak discharge as primary drivers of sediment transport, making it more responsive to variations in hydrological conditions. In contrast, MUSLE/SDR estimates soil loss based on

rainfall erosivity and adjusts with a sediment delivery ratio (SDR), resulting in predictions that are more rainfall-driven. MUSLE estimates ranged from 0.02 to 16.46 t ha^{-1} with an average of $0.89 \pm 2.09 \text{ t ha}^{-1}$, while MUSLE/SDR estimates varied between 0.04 and 10.63 t ha^{-1} with a mean of $0.96 \pm 1.59 \text{ t ha}^{-1}$. The results show that MUSLE tends to produce higher sediment yield estimates than MUSLE/SDR, particularly during high-runoff events. However, in some cases, the models exhibit close agreement, especially when sediment transport is dominated by rainfall-induced detachment. Discrepancies between the models can be attributed to differences in parameterization, such as the SDR and model sensitivity to hydrological factors. Overall, MUSLE captures a broader range of event-based hydrological variability, while MUSLE/SDR reflects more generalized erosion responses. These findings highlight the importance of model selection based on hydrological context in event-based sediment yield assessments. A strong correlation ($R^2 = 0.87$) between MUSLE/SDR and MUSLE predictions (Fig. 8) further supports their conceptual alignment and mutual consistency at the event scale.

Predicted sediment yield from MUSLE and MUSLE/SDR at the Sub-Catchment scale

To assess the consistency of sediment yield predictions across sub-catchments, estimates from the MUSLE model were compared with those from the MUSLE/SDR approach for 196 sub-catchments. Unlike previous event-based comparisons, this analysis focuses on aggregated sub-catchment values rather than individual storm events. MUSLE estimates ranged from 0.00 to 1.30 t ha^{-1} with an average of $0.34 \pm 0.30 \text{ t ha}^{-1}$, while MUSLE/SDR values ranged from 0.00 to 1.29 t ha^{-1} with a mean of $0.31 \pm 0.29 \text{ t ha}^{-1}$ (Table 2). The two models exhibited almost similar sediment yield dynamic across the study area. A regression analysis was conducted to evaluate the relationship between the two empirical models and determine the reliability for estimating event-based sediment yield at the sub-catchment scale. Figure 9 demonstrates a strong linear relationship ($R^2 = 0.93$) between sediment yield estimates from MUSLE/SDR and MUSLE at the sub-catchment scale. This high correlation indicates the reliability of both models in predicting event-based sediment yield.

Table 2 Descriptive statistics of predicted sediment yield (t ha^{-1}) from MUSLE/SDR and MUSLE at event and sub-catchment scales

Scale	Model	N	Min (t ha^{-1})	Max (t ha^{-1})	Mean (t ha^{-1})	SD (t ha^{-1})
Event-based	MUSLE/SDR	69	0.04	10.63	0.96	1.59
Event-based	MUSLE	69	0.02	16.46	0.89	2.09
Sub-catchment	MUSLE/SDR	196	0.00	1.29	0.31	0.29
Sub-catchment	MUSLE	196	0.00	1.30	0.34	0.30

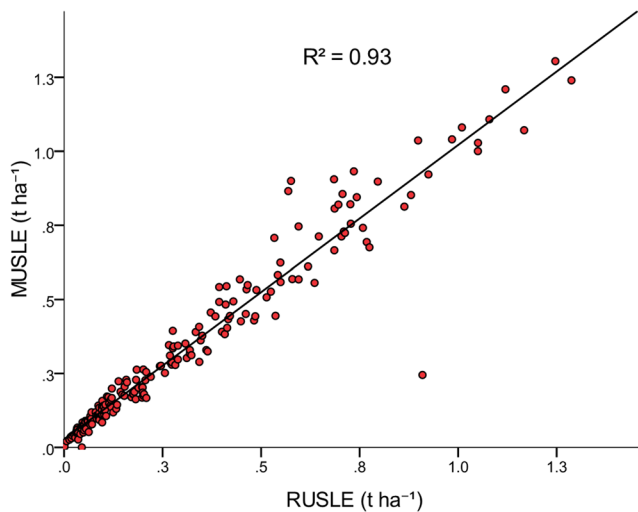


Fig. 9 Event-based sediment yield predicted by MUSLE and RUSLE/SDR (t ha^{-1}) for each sub-catchment, illustrating differences between the two modeling approaches

Spatial distribution of sediment yield across Sub-catchments

The spatial distribution of sediment yield across the sub-catchments, as estimated by the MUSLE and RUSLE/SDR models, shows considerable variation throughout the watershed (Fig. 10). Both models identified higher sediment yields predominantly in the northern and northeastern sub-catchments area with steepness, while lower values were concentrated in the south. The MUSLE model estimated event-based sediment yields ranging from 0.00 to 1.305 t ha^{-1} , whereas the RUSLE/SDR model predicted event-based sediment yield between 0.001 and 1.287 t ha^{-1} (Table 2). Despite general spatial agreement in identifying erosion-prone zones, MUSLE produced slightly broader high-yield areas, likely due to its sensitivity to hydrological factors such as runoff volume and peak discharge. In contrast, RUSLE/SDR results were more generalized, driven by energy-based rainfall erosivity and the sediment delivery ratio. These spatial patterns highlight the influence of topography, land use,

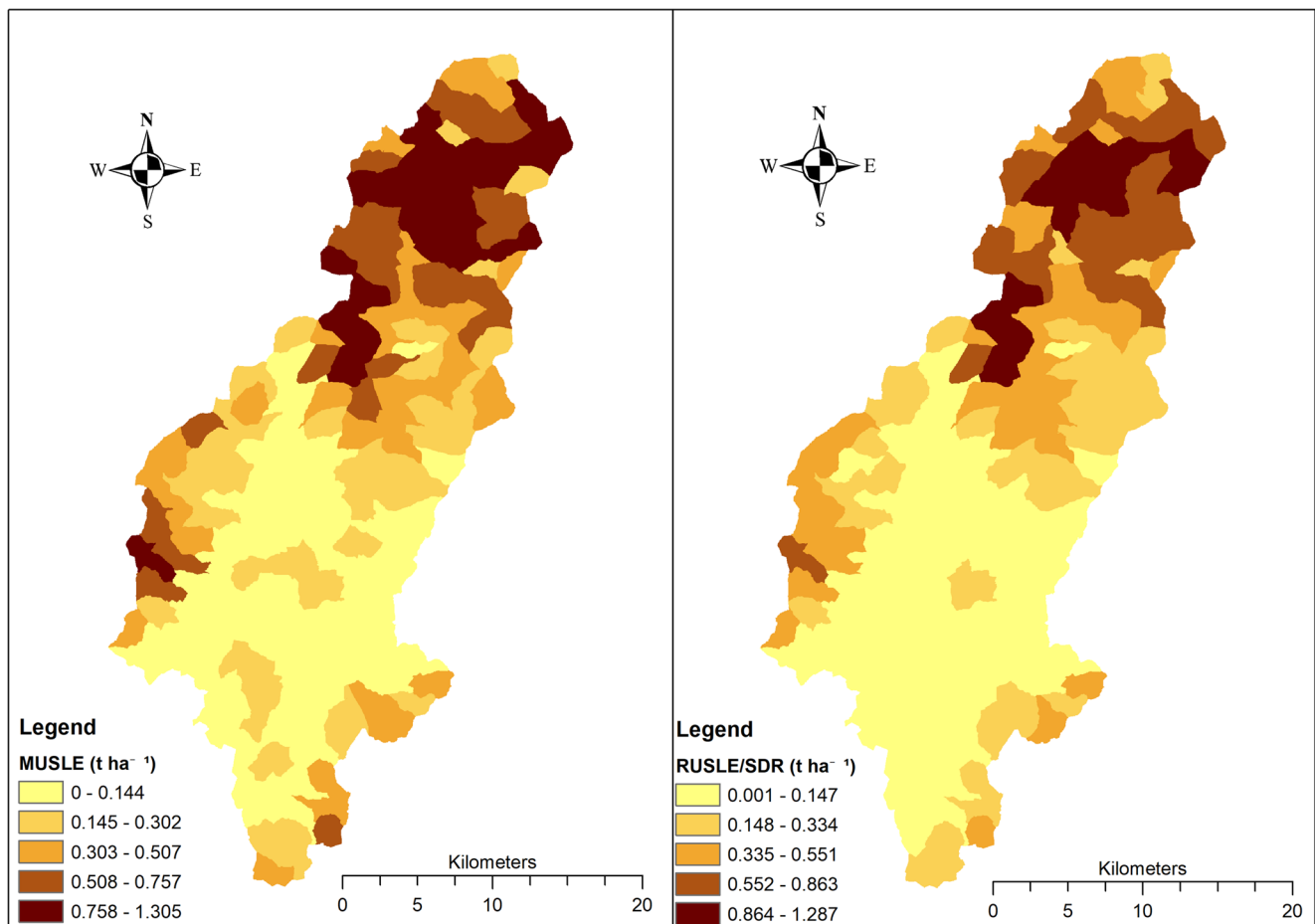


Fig. 10 Spatial distribution of predicted sediment yield across sub-catchments from MUSLE and RUSLE/SDR models, illustrating sediment transport patterns within the study area

and hydrological processes, and can inform prioritization of soil conservation efforts.

Alignment of machine learning models with RUSLE and MUSLE using sediment connectivity metrics

Five machine learning (ML) algorithms Random Forest (RF), Extreme Gradient Boosting (XGBoost), Support Vector Regression (SVR), Artificial Neural Networks (ANN), and k-Nearest Neighbors (k-NN) were used to reproduce sediment yield outputs generated by RUSLE and MUSLE models, based solely on sediment connectivity metrics and watershed descriptors. These ML models were not trained on observed sediment data, but rather on outputs derived from the empirical models (RUSLE and MUSLE).

The primary goal was to evaluate whether sediment connectivity indices could serve as reliable surrogates for empirical model behavior. Twelve input features were used, including the Sediment Connectivity Index (SCI), Sediment

Delivery Ratio (SDR), Relative Relief Index (RRI), and others. A 5-fold cross-validation approach was implemented to assess the internal consistency of each model. As shown in Table 3, XGBoost consistently achieved the best performance, yielding the lowest RMSE and MAE values and the highest alignment with RUSLE and MUSLE outputs ($R^2 = 0.942$). Random Forest also showed strong performance ($R^2 = 0.936$ for RUSLE, 0.912 for MUSLE). The remaining models ANN, k-NN, and SVR demonstrated reasonable model reproduction capability, with R^2 values above 0.77. Importantly, SCI, SDR, and RRI emerged as the most influential connectivity metrics, demonstrating their capacity to approximate the sediment yield responses of both empirical models (Table 3).

The performance analysis of the optimized machine learning models is visualized in Fig. 11 through scatter plots comparing the ‘Actual’ versus ‘Predicted’ sediment values. The term ‘Actual’ throughout this figure refers to the reference output values derived from the established

Table 3 Performance comparison of machine learning (ML) algorithms in reproducing sediment yield estimates from RUSLE and MUSLE models based on sediment connectivity indices (model-to-model comparison)

		ANN	Random Forest	k-NN	SVR	XGBoost
MSE	RUSLE	0.017	0.006	0.012	0.012	0.004
	MUSLE	0.015	0.008	0.015	0.012	0.004
RMSE	RUSLE	0.130	0.074	0.110	0.108	0.065
	MUSLE	0.122	0.089	0.121	0.108	0.065
MAE	RUSLE	0.097	0.049	0.087	0.090	0.047
	MUSLE	0.094	0.055	0.097	0.090	0.047
R^2	RUSLE	0.770	0.936	0.835	0.843	0.942
	MUSLE	0.813	0.912	0.816	0.843	0.942

The evaluation metrics (MSE, RMSE, MAE, and R^2) reflect the internal consistency between machine learning outputs and empirical model predictions (RUSLE and MUSLE), rather than accuracy against observed sediment data. Higher R^2 and lower error values indicate stronger alignment and reproduction capability

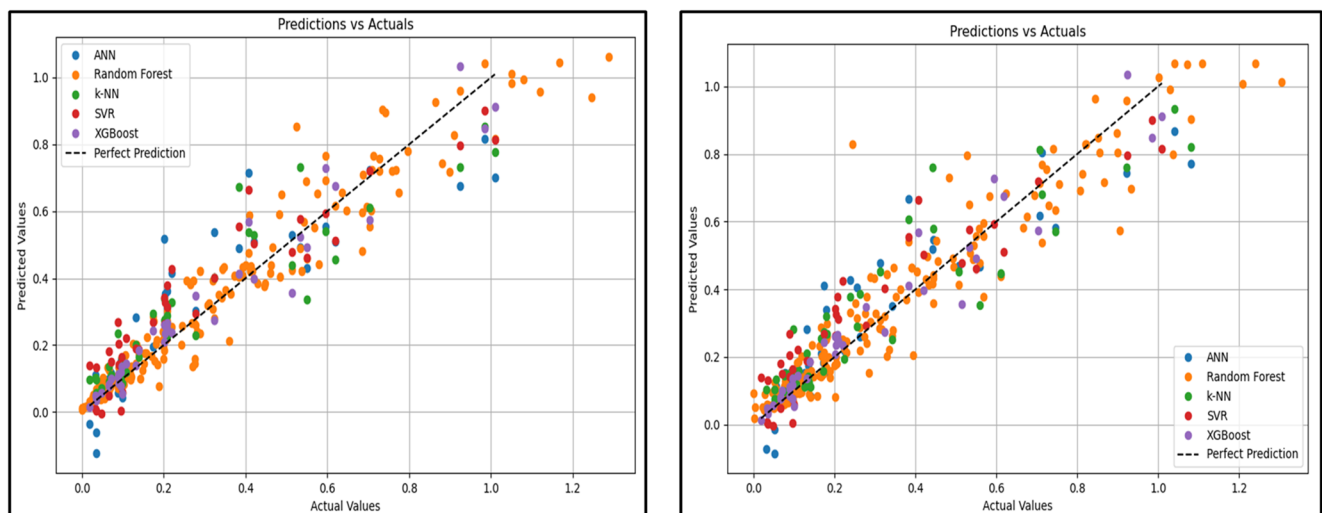


Fig. 11 Model-to-model comparison between predicted sediment yield from ML algorithms and empirical estimates from (a) MUSLE and (b) RUSLE

process-based models: the RUSLE model for average annual soil loss and the MUSLE model for event-based sediment yield. Specifically, Figure (a-left) displays the comparison for the RUSLE target variable (Annual Soil Loss), while Figure (b-right) displays the comparison for the MUSLE target variable (Event-Based Sediment Yield). The proximity of the data points to the 1:1 line in both graphs indicates the model's high predictive accuracy and its successful application as a surrogate for these physical models.

Spatial distribution of connectivity indices: IC and TWI

Among the twelve analyzed connectivity-related indices, the IC proposed by Borselli et al. (2008) and the TWI were selected for spatial visualization due to their relevance in interpreting sediment transport dynamics and hydrological flow accumulation (Fig. 12). The IC map illustrates the

spatial variability of sediment connectivity across the watershed, highlighting areas with high potential for sediment delivery toward the outlet. Conversely, the TWI map reflects zones of potential soil moisture accumulation, contributing to the understanding of runoff generation patterns. The spatial distribution of these indices is presented in Fig. 3.5, providing critical insights into topographic control on sediment and water flow processes across the basin. Areas with high connectivity (more negative IC values) were mainly located along the main drainage paths and steep slopes, suggesting greater potential for sediment to reach the outlet. In contrast, low connectivity zones (higher IC values) were observed in flatter or vegetated areas, where sediment is more likely to be trapped or deposited locally. These patterns help explain spatial differences in sediment yield and can guide targeted soil conservation practices. The connectivity map provides valuable spatial insights that complement the sediment yield predictions derived from the RUSLE and MUSLE models.

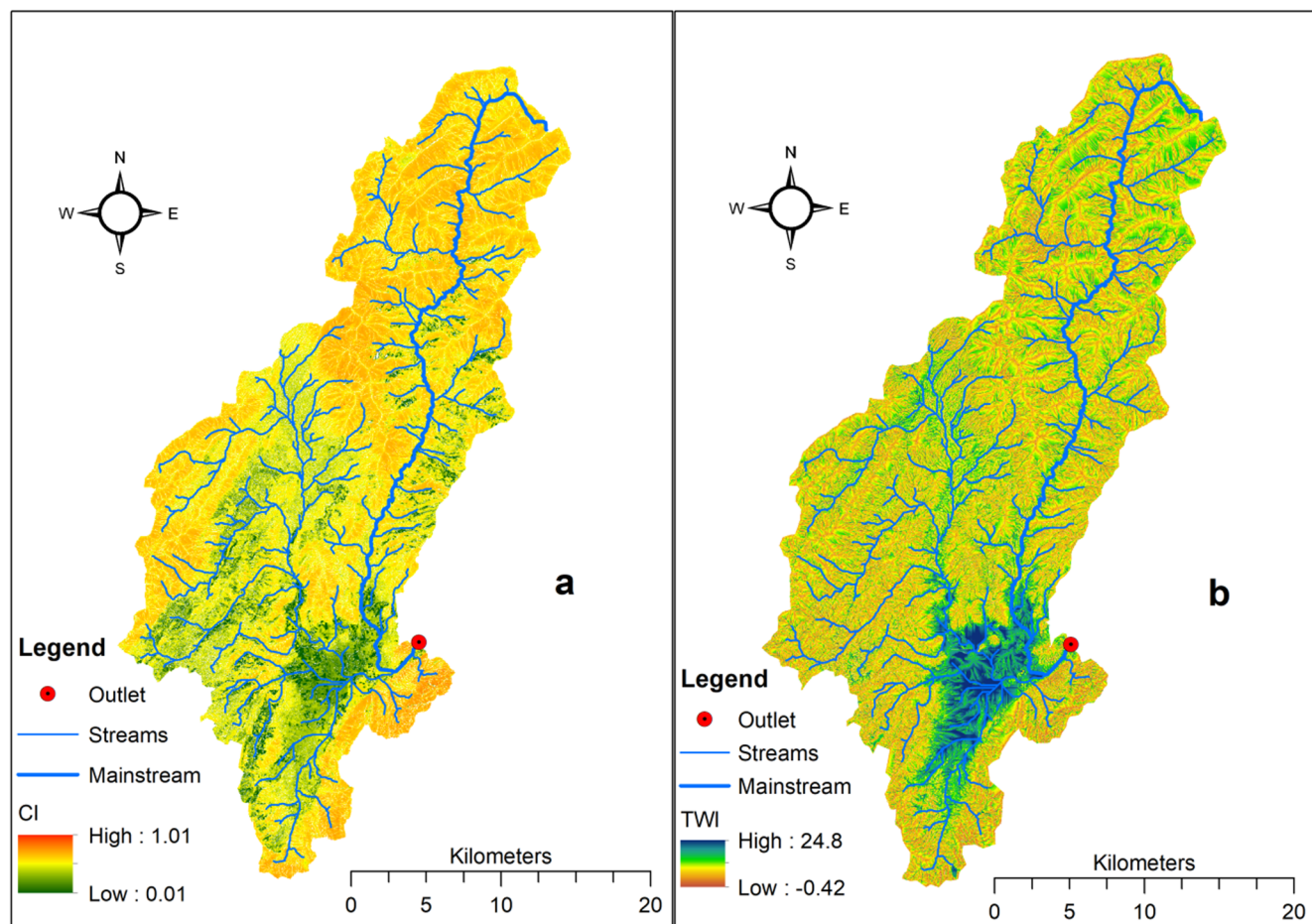


Fig. 12 Partial distribution of the two most effective connectivity indices across the study area: **a** Connectivity Index (IC; Borselli et al. 2008) and **(b)** Topographic Wetness Index (TWI)

Discussion

One of the most compelling findings of this study is the ability of sediment connectivity indices to reliably reproduce sediment yield estimates generated by the RUSLE and MUSLE models. Despite the absence of observed sediment data, the high predictive performance of ML models trained exclusively on connectivity metrics (e.g., SCI, IC, SDR, and TWI) demonstrates that these indices encapsulate key spatial and hydrological processes underpinning sediment transport. This suggests that empirical erosion models such as RUSLE and MUSLE may be well-represented by topography-driven connectivity patterns, enabling a parsimonious, spatially-explicit, and data-light approach to sediment yield estimation. While this does not replace the need for empirical validation, it highlights the explanatory power of landscape connectivity as a primary control on sediment flux across catchments. Similar conclusions were reported by Abebe et al. (2023), who demonstrated in a Mediterranean catchment that a combined RUSLE–IC–SDR approach could effectively identify spatial sediment source areas and capture sediment dynamics even in the absence of dense monitoring data.

These sediment connectivity indices, when used as inputs to machine learning models, successfully approximate sediment yield estimates, making them promising surrogates in ungauged or data-scarce watersheds. This finding is consistent with Pulley and Collins (2024), who highlighted that connectivity from hillslopes to channels, is key for linking local erosion processes to catchment-scale sediment delivery.

RUSLE and MUSLE, as widely accepted empirical and semi-empirical models, demonstrated strong potential in an event-based framework to capture temporal and spatial variations in sediment yield. They effectively represent the dynamic interactions among erosion-driving parameters such as rainfall intensity, slope, land cover, and soil erodibility at the scale of individual rainfall events. This is supported by Rai and Mathur (2008), who showed that ANN-based event models trained on rainfall, runoff, and sediment data can accurately reproduce temporal sediment yield variations, and by Shi et al. (2022), who demonstrated that an improved MUSLE model incorporating peak discharge and runoff accurately predicts event-based sediment yield at the watershed scale. These findings highlight the value of empirical and semi-empirical models as a foundation for integrating machine learning approaches that can capture additional nonlinearities and spatial heterogeneity in sediment connectivity.

The integration of machine learning complements these models by identifying complex, nonlinear interactions

among multiple controlling variables that may not be fully captured by deterministic frameworks. Notably, the high R^2 values achieved by XGBoost and Random Forest in reproducing RUSLE and MUSLE outputs suggest that ML algorithms can effectively reflect sediment connectivity dynamics. These models successfully integrate topographic descriptors (e.g., slope length and steepness), hydrological drivers (e.g., rainfall erosivity and runoff volume), and land surface properties (e.g., land use and land cover) to simulate sediment transport and delivery. This is consistent with Lamane et al. (2025), who demonstrated that SHAP frameworks enhance ML interpretability by quantifying the influence of individual predictors, aiding model refinement. This is further supported by Shi et al. (2022) and El Bilali et al. (2024), who highlighted that physics-informed ML models integrated with empirical frameworks like MUSLE improve predictive accuracy while maintaining physical consistency.

Their predictive performance demonstrates an enhanced capacity to represent the spatial organization of erosion processes, especially in large and heterogeneous watersheds subdivided into multiple sub-catchments. In such settings, the configuration of connectivity pathways exerts a critical influence on sediment mobilization. The dominance of SCI and SDR_A among the top-ranked ML features reinforces their importance within the RUSLE and MUSLE frameworks, validating both their physical relevance and statistical significance in predictive modeling. Similar emphasis on SDR-based indices as strong determinants of sediment delivery has been documented by Abebe et al. (2023), confirming the suitability of coupling topographic connectivity with empirical models for improved sediment yield estimation.

Despite the strong performance of ML algorithms in reproducing RUSLE and MUSLE predictions, an important limitation is the absence of observed sediment yield data. The machine learning models were trained solely on outputs from empirical models, meaning they reflect model behavior rather than actual sediment transport dynamics. Therefore, while the results demonstrate strong internal consistency, they cannot substitute for empirical validation in the field. Consequently, these ML-based estimates should be applied cautiously in operational sediment management until validated with field measurements. This limitation has been similarly addressed by recent studies developing physics-informed ML approaches, which integrate empirical models like MUSLE with ML algorithms to enhance predictive accuracy while maintaining physical consistency (Shi et al. 2022).

Furthermore, additional constraints should be acknowledged. The ML models were trained exclusively on RUSLE and MUSLE outputs, which may not fully capture

field-observed sediment dynamics. The analysis included only 69 rainfall events, limiting temporal coverage and potentially underrepresenting extreme hydrological events. This may reduce the model's ability to predict rare, high-magnitude sediment transport events, which are critical for watershed planning and hazard mitigation. Moreover, the models assumed static land use and climatic conditions, whereas changes in these factors could significantly affect sediment yield and connectivity patterns. Future studies should incorporate dynamic land use and climate scenarios to improve model applicability and capture temporal variations in sediment connectivity and yield. Recognizing these limitations provides context for the results and guides future research toward more comprehensive, field-validated approaches.

Moreover, by employing an event-based modeling approach, the integration of these connectivity parameters into ML algorithms enables more responsive and context-specific sediment yield estimates. This allows for the capture of short-term hydrological fluctuations and land surface changes, moving beyond traditional long-term averages toward temporally explicit and process-sensitive modeling. Such capabilities are reflected in the work of Rai and Mathur (2008), where event-based ANN models successfully captured short-term variations in sediment fluxes from small watersheds. This approach is particularly beneficial in data-scarce regions or rapidly changing landscapes, where static assumptions often fall short.

Ultimately, the convergence of empirical modeling and ML-driven prediction offers a dual advantage: it preserves the theoretical foundations of established hydrological models through physically meaningful parameters, while leveraging machine learning's flexibility to handle nonlinearities and spatial heterogeneity. This integrated framework not only improves prediction reliability but also enhances interpretability providing decision-makers with high-resolution, event-based insights essential for targeted watershed planning, sediment management, and erosion control in complex terrains.

This approach is particularly well-suited for application in ungauged or data-scarce catchments, where conventional modeling is constrained by limited field data. Studies have demonstrated that integrating ML with empirical models enhances predictive accuracy while maintaining physical consistency (El Bilali et al., 2024),

The results also affirm that XGBoost and Random Forest exhibit superior performance for both RUSLE and MUSLE

predictions, capturing over 94% and 93% of the variance in the data ($R^2 = 0.942$ and 0.936 , respectively). These outcomes highlight their strong generalization capacity and reliability across varying conditions.

Conclusion

This study demonstrates that sediment connectivity indices can serve as robust proxies for modeling sediment yield when coupled with machine learning (ML) algorithms. Sediment yields were first estimated using empirical models (RUSLE and MUSLE) across 196 sub-catchments and 69 rainfall events, capturing both rainfall- and runoff-driven processes. Despite their different assumptions, RUSLE and MUSLE showed strong agreement ($R^2 = 0.87$ at the event scale; $R^2 = 0.93$ at the sub-catchment scale), with event-based sediment yields ranging from 0.02 to 16.46 t ha^{-1} (MUSLE) and 0.04 – 10.63 t ha^{-1} (RUSLE/SDR), and mean values of 0.89 and 0.96 t ha^{-1} , respectively.

Using these empirical model outputs as targets, multiple sediment connectivity indices including the Index of Connectivity (IC), Sediment Delivery Ratio (SDR), Topographic Wetness Index (TWI), Stream Power Index (SPI), and Flow Length Ratio (FLR) were applied as input features in five ML algorithms (XGBoost, Random Forest, k-NN, SVR, and ANN). Among them, XGBoost and Random Forest achieved the highest performance, accurately reproducing empirical sediment yield estimates with $R^2 = 0.912$ – 0.942 , $\text{RMSE} = 0.065$ – 0.089 , and $\text{MAE} = 0.047$ – 0.055 . These results confirm that connectivity metrics alone can capture much of the variance in empirically modeled erosion, providing a scalable and data-efficient approach for ungauged or data-limited watersheds. It is important to note that these ML models were trained on RUSLE/MUSLE outputs (not measured sediment) from a limited number of events under static land use and climate. Therefore, practical application requires validation with field data and consideration of dynamic land-use and climate scenarios.

Overall, this study highlights the novelty and practical contribution of integrating sediment connectivity indices with ML to emulate empirical erosion models, offering a transferable framework for sediment yield prediction. Future work should incorporate dynamic land-use and climate scenarios, high-frequency monitoring, and field measurements to improve temporal resolution and evaluate model transferability across diverse watershed conditions.

Appendix 1

Equation Tables for RUSLE, MUSLE, and Connectivity Indices

The whole equations are summarized in Tables 4 and 5 as followings.

Table 4 Key equations for RUSLE and MUSLE models

Equation No.	Model/Component	Equation	Description
(1)	RUSLE Soil Loss	$Ae = Re \cdot Ke \cdot L \cdot S \cdot C \cdot P$	Event-based soil loss
(2)	RUSLE Sediment Yield	$Ye_{RUSLE} = Ae \cdot SDR$	Event-based sediment yield using sediment delivery ratio
(3)	Rainfall Erosivity (Re)	$Re = \sum (ET \cdot I_{30})$	Sum of kinetic energy and max 30-min intensity over storm events
(4–6)	Soil Erodibility (Ke)	Functions of soil texture (C), OM, and DG	Based on Torri et al. and seasonal variations
(7–16)	Seasonal K-factor Estimations	$K_i = f(K_{max}, K_{min}, T_{av}, t_{max}, t_{min})$	Time-dependent soil erodibility
(17–18)	Rainfall Erosivity Distribution	Semi-monthly $EI = f(EI_{month})$	Distributes EI over month halves
(19)	LS Factor	$LS = \left(\frac{X \cdot \eta}{22.13}\right)^{0.4} \cdot \left(\frac{\sin \theta}{0.0896}\right)^{1.3}$	Combines flow accumulation and slope
(20)	C Factor from NDVI	$C = Exp \left[-\alpha \cdot \frac{NDVI}{\beta - NDVI} \right]$	Remote sensing-based cover factor
(21)	SDR by Slope	$SDR = 0.627 SLP^{0.403}$	Empirical relationship with stream slope
(22)	MUSLE Sediment Yield	$Ye = 11.8(Q \times qp)^{0.56} K \times LS \times S \times C \times P$	MUSLE model using hydrological inputs
(23)	Runoff Factor	$11.8(Q \times qp)^{0.56}$	Product used in MUSLE model
(24)	Runoff Volume (Q)	$Q = 0.5 \times (Tp + Tr)$	Triangular hydrograph method
(25)	Peak Runoff (qp)	$qp = \frac{CrIA}{360}$	Rational method for peak discharge
(26)	Weighted Runoff Coefficient (Cr)	$Cr = \frac{1}{A} \sum_{i=1}^n (Cr_i \times A_i)$	Weighted by land use types
(27)	Time of Concentration (Tc)	$Tc = \frac{0.01947L^{0.77}}{S^{0.385}}$	Kirpich formula
(28)	Weighted Avg. Sediment Yield	$Ye_{Ave} = \frac{\sum_{i=1}^n (Ye_i \times A_i)}{\sum_{i=1}^n A_i}$	Area-weighted sediment yield

Table 5 Equations for sediment connectivity indices

Equation No.	Index/Factor	Equation	Description
(29)	Index of Connectivity (IC)	$CI = \log_{10} \left(\frac{D_{up}}{D_{dn}} \right)$	Log ratio of upslope and downslope components
(30)	Dup (Upslope Component)	$D_{up} = W \times S \times \sqrt{A}$	Slope, land use, and upslope contributing area
(31)	Ddn (Downslope Component)	$D_{dn} = \sum_i \frac{di}{W_i \times S_i}$	Downstream cells weighted by slope and land cover
(33)	Stream Power Index (SPI)	$SPI = A \times \tan(\beta)$	Combines flow accumulation and slope
(34)	Proximity to Outlet (Pi)	$P_i = 1 - \frac{d_i}{\max(d)}$	Relative distance to outlet
(35)	Topographic Wetness Index (TWI)	$TWI = \ln \left(\frac{A}{\tan(\beta)} \right)$	Wetness potential
(36)	Relative Relief Index (RRI)	$RRI = \frac{E_{max} - E_{min}}{L_{max}}$	Elevation difference over stream length
(37)	Topographic Connectivity Index (TCI)	$TCI = A \times S$	Area and slope product
(38)	Drainage Density (Dd)	$Dd = \frac{\sum L}{A_{basin}}$	Stream length per unit area
(39)	Watershed Shape Factor (WSF)	$WSF = \frac{L_{max}}{A}$	Compactness/elongation metric
(40)	Flow Length Ratio (FLR)	$FLR_i = \frac{L_i}{L}$	Local to global flow length ratio
(41)	SDR by Slope of mainstream	$SDR = 0.627 (SLP)^{0.403}$	Main stream slope-based SDR
(42)	SDR by Area	$SDR = 0.51 \times A^{-0.11}$	Area-based SDR
(43)	Normalization Equation	$X' = \frac{X - X_{min}}{X_{max} - X_{min}}$	Index standardization (Min-max Method)

Table 6 List of parameters and variables

Symbol/Term	Description	Unit/Notes
Ae	Event-based soil loss	t ha ⁻¹ event ⁻¹
Ye	Event-based sediment yield	t ha ⁻¹ event ⁻¹
R/Re	(Event-based) Rainfall erosivity factor	MJ mm ha ⁻¹ h ⁻¹ event ⁻¹
K/Ke	(Event-based) Soil erodibility factor	t ha ⁻¹ MJ ⁻¹ mm ⁻¹
L	Slope length factor	Dimensionless
S	Slope steepness factor	Dimensionless
LS	Topographic factor (product of L and S)	Dimensionless
C/Ce	Cover management factor	Dimensionless (derived from NDVI)
P	Support practice/Conservation practices factor	Dimensionless
SDR	Sediment Delivery Ratio	Dimensionless
Q	Runoff volume	m ³
qp	Peak discharge	m ³ s ⁻¹
A	Drainage area of watershed	km ²
χ (chi)	Flow accumulation	Derived from DEM
θ (theta)	Slope angle	Degrees
NDVI	Normalized Difference Vegetation Index	Remote sensing index
DG	Geometric mean of soil particle size	mm
OM	Organic matter content	%
TWI	Topographic Wetness Index	log-transformed index
SPI	Stream Power Index	Combines slope and flow accumulation
RRI	Relative Relief Index	(E _{max} - E _{min})/Longest Stream Length
IC	Index of Connectivity	Dimensionless
Dup	Upslope component of IC	Derived from slope, flow length, LULC
Ddn	Downslope component of IC	Derived from flow accumulation, LULC
Dd	Drainage density	km ⁻¹
FLR	Flow Length Ratio	Dimensionless
WSF	Watershed Shape Factor	L _{max} /Area
DA	Drainage Area	km ²
Pi	Proximity Index	Dimensionless (scaled 0–1)
X'	Normalized index value	Scaled 0–1
EI	Erosivity Index	% or absolute value
Cr	Runoff coefficient	Dimensionless
Tc	Time of concentration	Minutes
MSE	Mean Squared Error (model performance metric)	–
RMSE	Root Mean Squared Error	–
MAE	Mean Absolute Error	–
R ²	Coefficient of Determination	–

Acknowledgements All authors contributed to the study's conception and design. Abdul Alim Mohammadi: Conceptualization, Methodology, Literature Review/Introduction, Investigation, Data Interpretation, Data Analysis, Writing – Original Draft, Data Curation. Sibel Canaz Sevgen: Machine Learning (ML) Application, Validation, Visualization, Data Interpretation, Writing – Original Draft, Writing – Review & Editing. Gunay Erpul: Supervision, Conceptualization, Writing – Original Draft, Writing – Review & Editing. All authors have read and approved the final manuscript.

Author contributions All authors contributed to the study's conception and design. Abdul Alim Mohammadi: Conceptualization, Methodology, Literature Review/Introduction, Investigation, Data Interpretation, Data Analysis, Writing – Original Draft, Data Curation. Sibel Canaz Sevgen: Machine Learning (ML) Application, Validation, Visualization, Data Interpretation, Writing – Original Draft, Writing – Review & Editing. Gunay Erpul: Supervision, Conceptualization, Writing – Original Draft, Writing – Review & Editing. All authors have read and approved the final manuscript.

Funding Open access funding provided by the Scientific and Technological Research Council of Türkiye (TÜBİTAK). The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Data availability Data Availability: Data will be made available upon request.

Declarations

Competing interests The authors declare no competing interests.

Clinical trial Clinical trial number: not applicable.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the

source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abebe N, Eekhout J, Vermeulen B, Boix-Fayos C, De Vente J, Grum B, Hoitink T, Baartman J (2023) The potential and challenges of the 'RUSLE-IC-SDR' approach to identify sediment dynamics in a Mediterranean catchment. *CATENA* 233:107480. <https://doi.org/10.1016/j.catena.2023.107480>
- Ajibade FO, Nwogwu NA, Adelodun B, Abdulkadir TS, Ajibade TF, Lasisi KH, Fadugba OG, Owolabi TA, Olajire OO (2020) Application of RUSLE integrated with GIS and remote sensing techniques to assess soil erosion in Anambra State, South-Eastern Nigeria. *J Water Clim Change* 11(S1):407–422. <https://doi.org/10.2166/wcc.2020.222>
- Akgoz R, Deviren Saygin S, Erpul G, Tel S (2022) Monitoring seasonal and phenological variability of cover management factor for wheat cropping systems under semi-arid climate conditions. *Environ Monit Assess* 194(6):395. <https://doi.org/10.1007/s10661-022-10064-1>
- Alebachew ED, Abiye W, Dengiz O, Turan İD (2025) Soil erosion estimation and risk assessment based on RUSLE in Google Earth engine (Gee) in Türkiye. *Ann GIS* 31(1):123–141. <https://doi.org/10.1080/19475683.2025.2452262>
- Arabkhedri M, Heidary K, Parsamehr M-R (2021) Relationship of sediment yield to connectivity index in small watersheds with similar erosion potentials. *J Soils Sediments* 21(7):2699–2708. <https://doi.org/10.1007/s11368-021-02978-z>
- Atalla MA, Shebl A, Durin B, Kranjčić N, AlMetwaly WM (2024) Assessment of groundwater potential zones in Kuwait's semi-arid region: a hybrid approach of multi-criteria decision making, Google earth engine, and geospatial techniques. *Sci Rep* 14(1):29938. <https://doi.org/10.1038/s41598-024-76989-4>
- Avand M, Mohammadi M, Mirchooli F, Kaviani A, Tiefenbacher JP (2023) A new approach for smart soil erosion modeling: integration of empirical and machine-learning models. *Environ Model Assess* 28(1):145–160. <https://doi.org/10.1007/s10666-022-09858-x>
- Benavidez R, Jackson B, Maxwell D, Norton K (2018) A review of the (Revised) universal soil loss equation ((R)usle): with a view to increasing its global applicability and improving soil loss estimates. *Hydrol Earth Syst Sci* 22(11):6059–6086. <https://doi.org/10.5194/hess-22-6059-2018>
- Bombino G, Boix-Fayos C, Cataldo MF, D'Agostino D, Denisi P, De Vente J, Labate A, Zema DA (2020) A modified Catchment Connectivity Index for applications in semi-arid torrents of the Mediterranean environment. *River Res Appl* 36(5):735–748. <https://doi.org/10.1002/rra.3606>
- Borselli L, Cassi P, Torri D (2008) Prolegomena to sediment and flow connectivity in the landscape: a GIS and field numerical assessment. *CATENA* 75(3):268–277. <https://doi.org/10.1016/j.catena.2008.07.0066>
- Bouhadab CE, Menani MR, Bouguerra H, Derdous O (2018) Assessing soil loss using GIS-based RUSLE methodology: case of the Bou Namoussa watershed – North-East of Algeria. *J Water Land Dev* 36:27–35. <https://doi.org/10.2478/jwld-2018-0003>
- Breiman L (1996) Bagging predictors. *Mach Learn* 24:123–140
- Breiman L (2001) Random forests. *Mach Learn* 45:5–32
- Cavalli M, Trevisani S, Comiti F, Marchi L (2013) Geomorphometric assessment of spatial sediment connectivity in small Alpine catchments. *Geomorphology* 188:31–41. <https://doi.org/10.1016/j.geomorph.2012.05.007>
- Chang TJ, Zhou H, Guan Y (2016) Applications of erosion hotspots for watershed investigation in the Appalachian Hills of the United States. *J Irrig Drain Eng* 142(3):04015057
- Chen T, Guestrin C (2016) Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785–794)
- Chen W, Lei X, Chakraborty R, Chandra Pal S, Sahana M, Janizadeh S (2021) Evaluation of different boosting ensemble machine learning models and novel deep learning and boosting framework for head-cut gully erosion susceptibility. *J Environ Manage* 284:112015. <https://doi.org/10.1016/j.jenvman.2021.112015>
- Crema S, Cavalli M (2018) SedInConnect: a stand-alone, free and open source tool for the assessment of sediment connectivity. *Comput Geosci* 111:39–45. <https://doi.org/10.1016/j.cageo.2017.10.009>
- Das T, Sarma AK (2017) *Estimation of annual average soil loss and preparation of spatially distributed soil loss map: A case study of Dhansiri River basin* (Technical Report). Indian Institute of Technology Guwahati
- Demirtaş İ, Erpul G (2024) Determination of erosion statistics by land use type at national scale with the Dynamic Erosion Model and Monitoring System (DEMMS). *Journal of Environment, City, and Climate* 3(Special Issue on Desertification, Erosion, and Climate Change):136–155
- Deviren Saygin S (2021) Effects of season and phenology-based changes on soil erodibility and other dynamic RUSLE factors for semi-arid winter wheat fields. *J Agric Sci* 27(4):526–535. <https://doi.org/10.15832/ankutbd.749181>
- Deviren Saygin S, Basaran B, Ozcan AU, Dolarslan M, Timur OB, Yılman FE, Erpul G (2011) Land degradation assessment by geospatially modeling different soil erodibility equations in a semi-arid catchment. *Environ Monit Assess* 180:201–215. <https://doi.org/10.1007/s10661-010-1782-z>
- Deviren Saygin S, Ozcan AU, Basaran M, Timur OB, Dolarslan M, Yılman FE, Erpul G (2014) The combined RUSLE/SDR approach integrated with GIS and geostatistics to estimate annual sediment flux rates in the semi-arid catchment, Turkey. *Environ Earth Sci* 71:1605–1618
- Deviren Saygin S, Huang CH, Flanagan DC, Erpul G (2018) Process-based soil erodibility estimation for empirical water erosion models. *J Hydraul Res* 56(2):181–195. <https://doi.org/10.1080/00221686.2017.1312577>
- Djoukbala O, Mazour M, Hasbaia M, Benselama O (2018) Estimating water erosion in semiarid regions using the RUSLE equation under a GIS environment: case of Wadi El-Ham watershed in Hodna region, Algeria. *Environ Earth Sci* 77:345. <https://doi.org/10.1007/s12665-018-7532-1>
- Djoukbala O, Hasbaia M, Benselama O, Mazour M (2019) Comparison of the erosion prediction models from USLE, MUSLE and RUSLE in a mediterranean watershed, case of Wadi Gazouana (N-w of algeria). *Model Earth Syst Environ* 5(2):725–743. <https://doi.org/10.1007/s40808-018-0562-6>
- El Bilali A, Brouziyne Y, Attar O et al (2024) Physics-informed machine learning algorithms for forecasting sediment yield: an analysis of physical consistency, sensitivity, and interpretability. *Environ Sci Pollut Res Int* 31:47237–47257. <https://doi.org/10.1007/s11356-024-34245-2>
- Erpul G, Akgoz R, Özcan AU (2013) Comparison of soil loss predictions by the integrated RUSLE/SDR technology with the measured suspended sediment loads in the Buyuk Menderes Catchment, Turkey. In *M. De Boever, M. Khlosi, N. Delbecque,*

- J. De Pue, N. Ryken, A. Verdoodt, W. M. Cornelis, & D. Gabriels (Eds.), *Desertification and land degradation processes and mitigation* (pp. 112–117). UNESCO Chair of Eremology, Ghent University
- Erpul G, Şahin S, İnce K, Küçümen A, Akdağ MA, Demirtaş İ, Çetin E (2018) Türkiye Su Erozyonu atlası. Çölleşme ve Erozyonla Mücadele. Genel Müdürlüğü Yayınları, Ankara
- Fryirs K (2013) (Dis) connectivity in catchment sediment cascades: a fresh look at the sediment delivery problem. *Earth Surf Process Landforms* 38(1):30–46
- Fryirs KA, Brierley GJ, Preston NJ, Spencer J (2007) Catchment-scale (dis)connectivity in sediment flux in the upper Hunter catchment, new South Wales, Australia. *Geomorphology* 84(3–4):297–316. <https://doi.org/10.1016/j.geomorph.2006.01.044>
- Ghosh A, Rakshit S, Tikle S, Das S, Chatterjee U, Pande CB, Mattar MA (2022) Integration of GIS and remote sensing with RUSLE model for estimation of soil erosion. *Land* 12(1):116
- Guche MS, Geremew GB, Ayele EG (2024) Gis-based soil erosion dynamics modeling by rusle at watershed level in hare watershed, rift valley basin, ethiopia. *Appl Environ Soil Sci* 2024(1):9916448. <https://doi.org/10.1155/2024/9916448>
- Gunn SR (1997) Support vector machines for classification and regression. Technical report, image speech and intelligent systems research group, University of Southampton
- Haddou MA, Kabbachi B, Aydda A, Bouchriti Y, Mabrouki J (2024) RUSLE-based model for soil loss modeling and water erosion susceptibility mapping in the Issen basin (West-central of morocco). *Iran J Earth Sci* 16(4):1–14. <https://doi.org/10.57647/IJES.2024.1604.21>
- Handique A et al (2023) Application of revised universal soil loss equation (RUSLE) model for the Estimation of soil erosionand prioritization of erosion-prone areas in Majuli Island, Assam, India. *J Appl Nat Sci* 15(4):1667–1678. <https://doi.org/10.31018/jans.v15i4.51766>
- Hao R, Huang X, Cai ZW, Xiao HB, Wang J, Shi ZH (2022) Incorporating sediment connectivity index into MUSLE model to explore soil erosion and sediment yield relationships at event scale. *J Hydrol* 614:128579. <https://doi.org/10.1016/j.jhydrol.2022.128579>
- Hooke J, Souza J, Marchamalo M (2021) Evaluation of connectivity indices applied to a Mediterranean agricultural catchment. *CATENA* 207:105713. <https://doi.org/10.1016/j.catena.2021.105713>
- Horton RE (1932) Drainage-basin characteristics. *Transactions, American geophysical union* 13(1):350–361
- Kelsey KL (2002) *Use of the Revised Universal Soil Loss Equation (RUSLE) to predict event soil loss* (Master's thesis). University of Wisconsin-Stevens Point. Retrieved from <https://epapers.uwsp.edu/thesis/2002/kelsey.pdf>
- Khosravi K, Rezaie F, Cooper JR, Kalantari Z, Abolfathi S, Hatamifkoueieh J (2023) Soil water erosion susceptibility assessment using deep learning algorithms. *J Hydrol* 618:129229. <https://doi.org/10.1016/j.jhydrol.2023.129229>
- Kinnell PIA (2010) Event soil loss, runoff and the universal soil loss equation family of models: a review. *J Hydrol* 385(1–4):384–397. <https://doi.org/10.1016/j.jhydrol.2010.01.024>
- Kirpich Z (1940) Time of concentration of small agricultural watersheds. *Civ Eng* 10(6):339–342
- Kırcıman A, Akgoz R, Deviren Saygin S, Erpul G (2024) Seasonal modeling of soil erodibility in the central black sea region. *Rend Lincei Sci Fis Nat.* <https://doi.org/10.1007/s12210-024-01267-4>
- Koç, C (2014) A study on the development of hydropower potential in Turkey. *Renew Sust Energ Rev* 39:498–508
- Kumar A, Tripathi VK (2022) Capability assessment of conventional and data-driven models for prediction of suspended sediment load. *Environ Sci Pollut Res* 29:50040–50058. <https://doi.org/10.1007/s11356-022-18594-4>
- Kumar T, Jhariya DC, Pandey HK (2019) Comparative study of different models for soil erosion and sediment yield in Pairi watershed, Chhattisgarh, India. *Geocarto Int* 35(11):1245–1266. <https://doi.org/10.1080/10106049.2019.1576779>
- Kumar M, Sahu AP, Sahoo N, Dash SS, Raul SK, Panigrahi B (2022) Global-scale application of the RUSLE model: a comprehensive review. *Hydrol Sci J* 67(5):806–830. <https://doi.org/10.1080/02626667.2021.2020277>
- Lamane H, Mouhir L, Zouahri A, Baghdad B, El Bilali A, Moussadek R (2025) Modeling soil erosion and sediment yield under climate change: a comparison of RUSLE and MUSLE integrated with SDR using variable soil data resolutions. *Model Earth Syst Environ* 11(5):343. <https://doi.org/10.1007/s40808-025-02544-3>
- Lamane H, Mouhir L, Moussadek R, Baghdad B, Kisi O, El Bilali A (2025) Interpreting machine learning models based on SHAP values in predicting suspended sediment concentration. *Int J Sediment Res* 40(1):91–107. <https://doi.org/10.1016/j.ijsrc.2024.10.002>
- Lee S (2004) Soil erosion assessment and its verification using the universal soil loss equation and geographic information system: a case study at Boun. Korea. *Environ Geol* 45:457–465. <https://doi.org/10.1007/s00254-003-0897-8>
- Li H, Jin J, Dong F, Zhang J, Li L, Zhang Y (2024) Gully erosion susceptibility prediction using high-resolution data: evaluation, comparison, and improvement of multiple machine learning models. *Remote Sens* 16(24):4742. <https://doi.org/10.3390/rs16244742>
- Liu C, Fan H, Jiang Y, Ma R, Song S (2023) Gully erosion susceptibility assessment based on machine learning-a case study of watersheds in Tuquan County in the black soil region of Northeast China. *CATENA* 222:106798. <https://doi.org/10.1016/j.catena.2022.106798>
- Luvai A, Obiero J, Omuto C (2022) Soil loss assessment using the revised universal soil loss equation (Rusle) model. *Appl Environ Soil Sci* 2022:1–14. <https://doi.org/10.1155/2022/2122554>
- Madenoglu S, Pınar MÖ, Şahin S, Erpul G (2024) Sustainable land management for mitigating soil erosion at the catchment scale. *Türkiye Tarımsal Araştırmalar Dergisi* 11(2):176–190. <https://doi.org/10.19159/tutad.1434369>
- Mahoney D, Blandford B, Fox J (2021) Coupling the probability of connectivity and RUSLE reveals pathways of sediment transport and soil loss rates for forest and reclaimed mine landscapes. *J Hydrol* 594:125963. <https://doi.org/10.1016/j.jhydrol.2021.125963>
- Masselink RJ, Keesstra SD, Temme AJ, Seeger M, Giménez R, Casali J (2016) Modelling discharge and sediment yield at catchment scale using connectivity components. *Land Degrad Dev* 27(4):933–945
- McKague K ((2023)). Universal soil loss equation (USLE) vs. RUSLE2 (Factsheet #23–005). *Ministry of agriculture, Food and Rural Affairs Ontario*
- Merritt WS, Letcher RA, Jakeman AJ (2003) A review of erosion and sediment transport models. *Environ Model Softw* 18(8–9):761–799. [https://doi.org/10.1016/S1364-8152\(03\)00078-1](https://doi.org/10.1016/S1364-8152(03)00078-1)
- Miao J, Zhang X, Zhang G, Wei T, Zhao Y, Ma W, Chen Y, Li Y, Wang Y (2024) Applications and interpretations of different machine learning models in runoff and sediment discharge simulations. *CATENA* 238:107848. <https://doi.org/10.1016/j.catena.2024.107848>
- Mishra SK, Tyagi JV, Singh VP, Singh R (2006) SCS-CN-based modeling of sediment yield. *J Hydrol* 324(1–4):301–322. <https://doi.org/10.1016/j.jhydrol.2005.10.006>
- Mohammadi AA, Akgoz R, Saygin D (2025) Validation and comparison of event-based sediment yield estimations using MUSLE and RUSLE/SDR models: the case of çarşamba suyu sub-basin, Southwest central Anatolia, Türkiye. *Nat Hazards*. <https://doi.org/10.1007/s11069-025-07328-7>

- Moore ID, Burch GJ (1986a) Modelling erosion and deposition: topographic effects. *Trans ASAE* 29(6):1624–1630. <https://doi.org/10.13031/2013.30363>
- Moore ID, Burch GJ (1986b) Physical basis of the length-slope factor in the universal soil loss equation. *Soil Sci Soc Am J* 50(5):1294–1298. <https://doi.org/10.2136/sssaj1986.036159950050000500042x>
- Moore ID, Grayson RB, Ladson AR (1991) Digital terrain modeling: a review of hydrological, geomorphological and biological applications. *Hydrol Process* 5:3–30
- Mulvaney T (1851) On the use of self-registering rain and flood gauges in making observations of the relations of rainfall and flood discharges in a given catchment. *Proc Institution Civil Eng Irel* 4:18–33
- Nawazuzzoha M, Rashid Md. M., Mishra PK, Abdelrahman K, Fnais MS, Naqvi HR (2024) Empirical modeling of soil loss and yield utilizing rusle and syi: a geospatial study in south sikkim, teesta basin. *Land* 13(10):1621. <https://doi.org/10.3390/land13101621>
- Noori NE, Azeez DR (2023) Soil erosion risk assessment using revised universal soil loss equation rusle model: A case study in shwan sub basin, kirkuk, iraq. *IOP Conference Series: Earth and Environmental Science*, 1262(8):082007. <https://doi.org/10.1088/1755-1315/1262/8/082007>
- Özşahin E (2023) Climate change effect on soil erosion using different erosion models: a case study in the Naip Dam basin, Türkiye. *Comput Electron Agric* 207:107711. <https://doi.org/10.1016/j.compag.2023.107711>
- Prasannakumar V, Shiny R, Geetha N, Vijith H (2011) Spatial prediction of soil erosion risk by remote sensing, GIS and RUSLE approach: a case study of Siruvani river watershed in Attapady valley, Kerala, India. *Environ Earth Sci* 64(4):965–972. <https://doi.org/10.1007/s12665-011-0913-3>
- Pulley S, Collins AL (2024) Soil erosion, sediment sources, connectivity and suspended sediment yields in UK temperate agricultural catchments: discrepancies and reconciliation of field-based measurements. *J Environ Manage* 351:119810. <https://doi.org/10.1016/j.jenvman.2023.119810>
- Rai RK, Mathur BS (2008) Event-based sediment yield modeling using artificial neural network. *Water Resour Manage* 22:423–441. <https://doi.org/10.1007/s11269-007-9170-3>
- Reda Y, Moges A, Kendie H (2024) Application of the modified universal soil loss equation (Musle) for the prediction of sediment yield in agewmaria experimental watershed, tekeze river basin, northern ethiopia. *Heliyon* 10(15):e35052. <https://doi.org/10.1016/j.heliyon.2024.e35052>
- Reis M, Akay AE, Savaci G (2016) Erosion risk mapping using CORINE methodology for Goz watershed in Kahramanmaraş Region, Turkey
- Renard KG, Foster GR, Weesies GA, McCool DK, Yoder DC (1997) Predicting soil erosion by water—a guide to conservation planning with the Revised Universal Soil Loss Equation (RUSLE). *United States Department of Agriculture, Agricultural Research Service (USDA-ARS) Handbook No. 703*. United States Government Printing Office, Washington, DC
- Sadeghi SHR, Gholami L, Khaledi Darvishan A, Saeidi P (2014) A review of the application of the MUSLE model worldwide. *Hydrol Sci J* 59(2):365–375. <https://doi.org/10.1080/02626667.2013.866239>
- Schwab G (1981) *Soil and Water Conservation Engineering*. John Wiley Inc., New York
- Seitkazy M, Beisekenov N, Rakhimova M, Tokbergenova A, Zulpikharov K, Kaliyeva D, Taukebayev O, Levin E (2024) Soil erosion prediction in Western Kazakhstan through deep learning with a neural network approach to ls-factor analysis. *J Indian Soc Remote Sens*. <https://doi.org/10.1007/s12524-024-02080-0>
- Shi W, Chen T, Yang J, Lou Q, Liu M (2022) An improved MUSLE model incorporating the estimated runoff and peak discharge predicted sediment yield at the watershed scale on the Chinese Loess Plateau. *J Hydrol* 614:128598. <https://doi.org/10.1016/j.jhydrol.2022.128598>
- Torri D, Poesen J, Borselli L (1997) Predictability and uncertainty of the soil erodibility factor using a global dataset. *CATENA* 31(1–2):1–22. [https://doi.org/10.1016/S0341-8162\(97\)00036-2](https://doi.org/10.1016/S0341-8162(97)00036-2)
- Torri D, Poesen J, Borselli L (2002) Corrigendum to “predictability and uncertainty of the soil erodibility factor using a global dataset” [Catena 31 (1997) 1–22] and to “erratum to predictability and uncertainty of the soil erodibility factor using a global dataset” [Catena 32 (1998) 307–308]. *CATENA* 46(4):309–310. [https://doi.org/10.1016/S0341-8162\(01\)00175-8](https://doi.org/10.1016/S0341-8162(01)00175-8)
- Upadhayay HR, Lamichhane S, Bajracharya RM, Cornelis W, Collins AL, Boeckx P (2020) Sensitivity of source apportionment predicted by a Bayesian tracer mixing model to the inclusion of a sediment connectivity index as an informative prior: illustration using the Kharka catchment (Nepal). *Sci Total Environ* 713:136703. <https://doi.org/10.1016/j.scitotenv.2020.136703>
- Van der Knijff JM, Jones RJA, Montanarella L (2000) Soil erosion risk assessment in Europe, EUR 19044 EN. Office for Official Publications of the European Communities, Luxembourg
- Vigiak O, Borselli L, Newham LTH, McInnes J, Roberts AM (2012) Comparison of conceptual landscape metrics to define hillslope-scale sediment delivery ratio. *Geomorphology* 138(1):74–88. <https://doi.org/10.1016/j.geomorph.2011.08.026>
- Williams JR (1975) Sediment-yield prediction with universal equation using runoff energy factor. Present and prospective technology for predicting sediment yield and sources. ARS-S-40, U.S. Government Printing Office, Washington, DC, pp 244–252
- Williams JR, Berndt HD (1977) Sediment yield prediction based on watershed hydrology. *Trans ASAE* 20:1100–1104. <https://doi.org/10.13031/2013.35710>
- Wischmeier WH, Smith DD (1978) Predicting rainfall erosion losses. *Agriculture handbook No. 537*. Agricultural Research Service, U.S. Department of Agriculture, Washington D.C., USA
- Zhao G, Gao P, Tian P, Sun W, Hu J, Mu X (2020) Assessing sediment connectivity and soil erosion by water in a representative catchment on the Loess Plateau, China. *CATENA* 185:104284. <https://doi.org/10.1016/j.catena.2019.104284>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.