



# OPEN Identifying monthly rainfall erosivity patterns using hourly rainfall data across India

Subhankar Das<sup>1✉</sup>, Manoj Kumar Jain<sup>1</sup>, Karl Auerswald<sup>2</sup>, Carlos Rogerio de Mello<sup>3</sup> & Peter Molnar<sup>4</sup>

Rainfall erosivity is a key dynamic factor of water erosion estimation, with a significant spatial and temporal variation. This study presents a comprehensive analysis of the spatial patterns and monthly distribution of rainfall erosivity across India, using data from 261 hourly and 2,525 monthly rainfall stations covering the period from 1969 to 2021. In India, monthly rainfall erosivity and related attributes—such as the kinetic energy of erosive rainfall, the number of erosive events, and peak hourly rainfall intensity—have been systematically examined for the first time. Monthly erosivity estimates derived from hourly data were linked with monthly rainfall, enabling a simplified and efficient estimation approach. To predict monthly erosivity based on rainfall, temperature, and topographic variables, we developed and evaluated three modeling approaches: linear regression, a machine learning-based XGBoost model, and an ensemble model. XGBoost outperformed the others, achieving a median coefficient of determination ( $R^2$ ) of 0.97, while the ensemble model also performed well with a median  $R^2$  of 0.96. Additionally, a Geographically Weighted Regression (GWR) approach was applied for spatial interpolation, yielding accurate high-resolution erosivity maps with a median  $R^2$  of 0.90. The results also demonstrate that erosivity peaks during the summer monsoon months (June to September), with July exhibiting the highest value due to intense rainfall and high kinetic energy. Notably, the analysis revealed that nearly 32% of India experiences monthly erosivity exceeding  $2,000 \text{ MJ mm ha}^{-1} \text{ h}^{-1} \text{ month}^{-1}$  in July alone. In contrast, non-monsoon months showed considerably lower erosivity levels across most of the country. A statistically significant long-term increase was detected in January, with an average rise of  $+0.86 \text{ MJ mm ha}^{-1} \text{ h}^{-1} \text{ month}^{-1}$  in total erosivity and  $+0.1 \text{ mm h}^{-1}$  in maximum 60-min rainfall intensity annually. While acknowledging certain limitations, this study provides valuable insights into erosive rainfall characteristics, enhances rain-driven erosion assessment, and supports the development of timely and location-specific soil conservation strategies across India.

**Keywords** Rainfall erosivity, R-factor, USLE, RUSLE, Soil erosion, India

Water erosion is one of the most significant forms of environmental degradation affecting sustainable economies, food production, and socio-economic development in developing countries<sup>1,2</sup>. The global water erosion assessment showed that developing countries with less effective erosion management are more susceptible to erosion than developed countries<sup>2,3</sup>. South Asian countries are among the most vulnerable regions, characterized by intense rainfall, a growing population, degradation of natural resources, and high rates of poverty and food insecurity<sup>4–7</sup>.

Predicting rainfall-induced water erosion using the Universal Soil Loss Equation (USLE)<sup>8</sup> and its revised version, i.e., Revised Universal Soil Loss Equation (RUSLE)<sup>9</sup> and RUSLE2<sup>10</sup>, requires spatially distributed and temporally highly resolved rainfall data. However, the limited availability of such rainfall datasets hinders accurate erosion estimation in many parts of the globe<sup>11–15</sup>. Erosion estimation using the USLE and its revised versions also requires an intra-annual dataset to account for the seasonality of rainfall and vegetation dynamics<sup>13,16</sup>. Notably, the seasonal and monthly distribution of rainfall erosivity is crucial for understanding soil loss dynamics and estimating the cover management factor<sup>9,17</sup>.

<sup>1</sup>Department of Hydrology, Indian Institute of Technology Roorkee, Roorkee, India. <sup>2</sup>School of Life Sciences, Technical University of Munich, Freising, Germany. <sup>3</sup>Water Resources Department, Federal University of Lavras, Lavras, Brazil. <sup>4</sup>Department of Civil, Environmental and Geomatic Engineering, ETH Zurich, Zurich, Switzerland. ✉email: sdas@hy.iitr.ac.in

Accurate estimation of rainfall erosivity requires high-resolution precipitation data at 1- to 60-min intervals, ideally spanning more than 20 years<sup>18–22</sup>. It has been recognized that the unavailability of such high-temporal resolution observed datasets has resulted in significant errors in erosivity estimation<sup>23–26</sup>. Highly resolved satellite and climate reanalysis datasets are a promising alternative for estimating erosivity. However, recent rainfall erosivity studies in India<sup>18,27,28</sup>, the USA<sup>29</sup>, Burkina Faso<sup>30</sup> and China<sup>31</sup> showed a significant bias in the satellite and reanalysis-based erosivity estimates. Furthermore, the global assessment of satellite and reanalysis-derived erosivity products also revealed that most datasets failed to estimate erosivity correctly<sup>11,14,32</sup>. Notably, a significant underestimation was observed in the satellite and reanalysis-derived rainfall erosivity in the tropics<sup>32</sup>.

Numerous empirical erosivity equations based on annual and monthly rainfall datasets are available<sup>17,17,33–41</sup>. Among these, one of the well-known empirical equations is based on the Fournier and modified Fournier index methods for estimating erosivity<sup>33,42</sup>. These equations have in common that they apply only regionally and may change in time due to climate change<sup>43</sup>. Various empirical erosivity equations have been applied for regional-scale erosivity estimation<sup>39,44–46</sup>, with some used for global-scale erosion studies<sup>47–51</sup>.

India's first national-level rainfall erosivity equation was developed almost four decades ago using rainfall datasets from 45 stations in different climate zones<sup>34</sup>. This equation was further used to develop seasonal and annual iso-erodent maps for the country. A linear regression was proposed to estimate erosivity based on the average annual rainfall of 44 stations, excluding Mahabaleshwar, due to its exceptionally high rainfall values<sup>34,52,53</sup>. Additionally, an erosivity equation for the monsoon season (June–September) was developed from these 44 stations to estimate erosivity for 180 additional locations distributed across the country. Finally, seasonal and annual iso-erodent maps were prepared from the 225 stations<sup>34,53</sup>. Notably, in 2004, Babu et al.<sup>35</sup> refined the annual and seasonal iso-erodent map by analyzing observed rainfall data from 123 stations. They developed new equations for estimating annual and seasonal erosivity from annual and monthly rainfall. These equations were applied to rainfall data from over 500 rain gauge stations, resulting in a comprehensive erosivity dataset from 623 stations for the updated iso-erodent map.

Tiwari et al.<sup>46</sup> utilized 101 years of monthly rainfall data from 52 stations across India to estimate erosivity using a modified Fournier index-based equation<sup>33,42</sup>. Notably, the equation applied for erosivity estimation was not explicitly developed for Indian conditions. Majhi et al.<sup>26</sup> pointed out that the equation is a modified version of the original erosivity equation by Arnoldus<sup>33</sup> developed for Morocco. Eventually, the same equation or different versions have been used in numerous studies in India<sup>54–59</sup>. Furthermore, Chen et al.<sup>24</sup> identified nearly 28 modified versions of Arnoldus's<sup>33</sup> equation being used globally. Chen et al.<sup>24</sup> also highlighted that China and India are the top two countries utilizing such equations despite not being developed for these regions. Despite Chen et al.<sup>24</sup> emphasizing the need to stop the misuse of such equations, their application continues in many regional studies<sup>54,55</sup>. Additionally, Majhi et al.<sup>26</sup> revealed that more than ten different methods, some developed for other regions<sup>36,60,61</sup>, have been employed for erosivity estimation in India using monthly and annual datasets. However, very few regional studies have utilized high-temporal resolution (1–60 min) rainfall datasets<sup>12,17,62,63</sup>, indicating a significant gap.

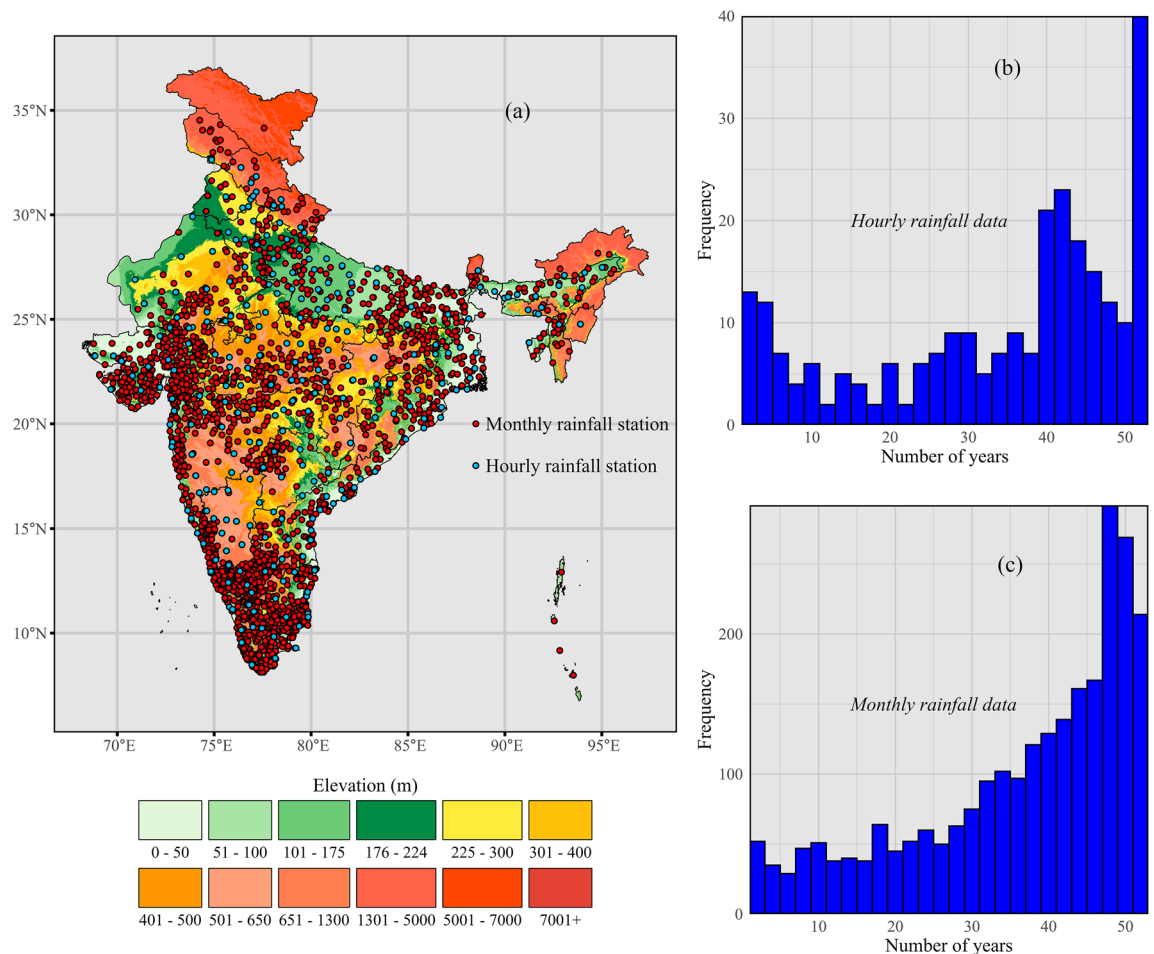
In summary, the unavailability of high-temporal resolution and spatially dense rainfall stations has led to unreliable and erroneous rainfall erosivity estimates in India, which further cascade into errors in erosion estimation, ranging from plot scale to national level studies<sup>26</sup>. Furthermore, the use of unreliable erosivity equations across the country has raised questions about the accuracy of erosion estimates. Given the high demand for accurate erosion assessment, rain-driven damage estimation, and water resource management, precise rainfall erosivity estimates and an understanding of the intra-annual variability of erosivity in India are crucial. Therefore, in this study, we leverage more than 30 years of long-term rainfall data with high-temporal resolution (60-min) from the India Meteorological Department (IMD) to understand the erosivity characteristics and achieve accurate erosivity estimation. Notably, such a long-term, high-temporal resolution observed rainfall dataset has not been utilized in previous studies in India. Moreover, by utilizing a large collection of hourly and monthly rainfall datasets, we aim to provide more reliable and useful rainfall erosivity distribution surfaces across the country. Advances have been made by combining machine learning techniques and an advanced Geographically Weighted Regression (GWR) interpolation scheme to prepare high-resolution monthly rainfall erosivity maps for India. Additionally, temporal trends and spatial relationships with geo-climatic variables influencing rainfall erosivity were analyzed on a monthly scale. Specifically, we aim to:

1. Identify the long-term spatial and intra-annual pattern of erosivity based on hourly resolved long-term rainfall data.
2. Prepare long-term, highly resolved monthly rainfall erosivity surfaces or maps for India.
3. Quantify the long-term trends or patterns in monthly erosivity and its attributes.

## Study area

India, located in the northern hemisphere, spans from approximately 8° 4' N to 37° 6' N latitudes and 68° 7' E to 97° 25' E longitudes, covering a vast area of approximately 3.287 million square kilometers, making it the seventh-largest country in the world (Fig. 1a). The country is bordered by China, Nepal, and Bhutan to the north, Pakistan to the west, and Bangladesh and Myanmar to the east. To the south, India is bounded by the Indian Ocean, with the Arabian Sea to its southwest and the Bay of Bengal to its southeast. A diverse geography, coupled with India's vast size, contributes to a wide range of climatic conditions. They range from the Himalayan cold deserts in the north (Trans-Himalaya; see Supplementary Fig. 1) to the tropical rainforests in the south (in particular the Western Ghats) and the arid regions of the west (Thar desert) to the fertile plains of temperate climate in the north (Gangetic plains) and the east (West Bengal).

India's climate is predominantly influenced by the monsoon seasons, with distinct wet and dry periods. The summer monsoon (June to September) brings heavy rainfall, particularly in the central highlands, northeast



**Fig. 1.** (a) Rainfall stations used for erosivity estimation from hourly and monthly rainfall data. (b) and (c) available temporal duration (years) of the hourly and monthly data. The map was generated using R version 4.4.3 with the *ggplot2*, *sf*, and *raster* packages.

India, and the Western Ghats regions. At the same time, the western Himalayas and southern regions (Deccan Plateau, Western and Eastern Ghats) experience more variation in precipitation during pre-monsoon (March to May) and post-monsoon periods (October to November)<sup>64</sup>. India's rainfall pattern shows marked spatial variability, with intense precipitation in the Western Ghats, northeast, and central regions, while arid zones like the Thar Desert receive as little as 100–500 mm annually. Notably, the northeastern state of Meghalaya hosts Mawsynram and Cherrapunji—among the wettest places on Earth—recording annual rainfall exceeding 10,000 mm<sup>65</sup>. This highly variable distribution is primarily shaped by orographic effects and the seasonal dynamics of the monsoon winds<sup>66,67</sup>.

With over 1.45 billion people, India is the most populous country in the world (<https://www.worldometers.info/>). The population is concentrated in urban centers and fertile agricultural regions, where large rural populations still rely on smallholder farming. The country's dependence on agriculture, coupled with its vulnerability to climate change, poses a serious threat to food security and rural livelihoods (IPCC<sup>5</sup>). The increasing trend in rainfall events, particularly during the monsoon season, can lead to accelerated soil erosion, threatening agricultural productivity. Given that over half of India's workforce is engaged in agriculture, such intense rainfall events have far-reaching socio-economic consequences, especially for marginal and smallholder farmers (FAO<sup>68</sup>).

### Datasets

This study incorporates an hourly rainfall dataset from 261 stations from 1969 to 2021, collected from the India Meteorological Department (IMD) (<https://dsp.imdpune.gov.in/index.php>). The start and end years of data records vary across stations. The collected dataset has an average coverage of nearly 35 years, with a median coverage of 41 years (Fig. 1b). Almost 80% of the stations have over 20 years of records, while approximately 90% of the stations have at least five years of hourly rainfall data. Additionally, a monthly rainfall dataset from 2,525 stations, spanning from 1969 to 2021, was collected from the IMD. The monthly dataset has an average duration of 37 years per station, with nearly 85% of the stations having more than 20 years of data, although some stations have less than 10 years of data (Fig. 1c).

Maximum and minimum temperature datasets from 1969 to 2021 were collected for 510 rainfall stations from the India Meteorological Department (IMD). However, temperature data for the remaining stations were unavailable, as many rain gauge stations are equipped only with non-recording rain gauges which do not capture temperature information. To fill this gap, the WorldClim (<https://www.worldclim.org/>)<sup>69</sup> monthly dataset was used for these stations. Additionally, a high-resolution gap-filled Shuttle Radar Topography Mission (SRTM) elevation dataset, with a ~90-m spatial resolution, was obtained from the Consultative Group on International Agricultural Research—Consortium for Spatial Information (CGIAR-CSI) (<https://srtm.csi.cgiar.org/>). The same dataset was also used for slope estimation. We also obtained historical long-term climate variables—solar radiation ( $\text{kJ m}^{-2} \text{day}^{-1}$ ), wind speed ( $\text{m s}^{-1}$ ), and water vapor pressure (kPa) at a 30 arc-second (~1  $\text{km}^2$ ) spatial resolution from WorldClim for the spatial mapping. Soil moisture ( $\text{m}^3 \text{m}^{-3}$ ) and surface runoff (m) data were obtained from the ERA5-Land<sup>70</sup> dataset ( $0.1^\circ \times 0.1^\circ$ ) of monthly averages (<https://cds.climate.copernicus.eu/cdsapp>). Soil type information (~250 m resolution) was sourced from the World Reference Base for Soil Resources 2006<sup>71</sup>, available at SoilGrids (<https://soilgrids.org/>). Land use data (~100 m resolution) were acquired from the Copernicus Land Monitoring Service (<https://land.copernicus.eu/en>). The ERA5-Land dataset was downscaled to the 30 arc-second resolution using a Random Forest-based downscaling method, as applied in our earlier study<sup>6</sup>. Soil and land use type datasets were upscaled using majority voting resampling<sup>72</sup>. The elevation (m) and slope (degree) estimated from the 90-m SRTM dataset were also converted to a 30 arc-second resolution using area-conservative regridding<sup>32,73</sup>. The coastline boundary was obtained from the Global Coast Line Dataset (GCL\_FCS30)<sup>74</sup>, which was used to estimate the distance from the sea and is freely available at <https://doi.org/10.5281/zenodo.13943679>.

## Methods

### Erosivity estimation from hourly data

The rainfall erosivity, or R-factor, in the Revised Universal Soil Loss Equation (RUSLE), is the product of the total kinetic energy and the maximum 30-min rainfall intensity of erosive storms, typically derived from high-resolution pluviograph rainfall data<sup>9,21,75,76</sup>. However, due to the unavailability of pluviograph data, we used hourly rainfall datasets for rainfall erosivity estimation, as used in many earlier studies<sup>12,29,77,78</sup>. We identified erosive rain events based on the criteria provided by Renard et al.<sup>9</sup>, which are widely applied for such assessments. An erosive rain event is defined as one with a total precipitation of at least 12.7 mm, or a maximum 30-min intensity exceeding  $12.7 \text{ mm h}^{-1}$ . The second criterion cannot be evaluated for hourly data. An erosive event is separated from the next rainfall event by at least a 6-h period of less than  $1.27 \text{ mm}$  rainfall. The kinetic energy for each unit of rainfall depth was calculated using the equation from RUSLE2<sup>10,79</sup>. The equation can be written as:

$$e = 0.29 \times [1 - 0.72 \times \exp(-0.082 \times i)] \quad (1)$$

where  $e$  is the rainfall kinetic energy per unit rainfall depth ( $\text{MJ ha}^{-1} \text{mm}^{-1}$ ), and  $i$  is the rainfall intensity in  $\text{mm h}^{-1}$ .

The total kinetic energy of erosive storms was estimated using Eq. 2.

$$E = \int_0^D (e \times \theta) dt \quad (2)$$

where  $E$  is the total kinetic energy ( $\text{MJ ha}^{-1}$ ) of the erosive storm.  $\theta$  is the rainfall depth (mm) for each time increment  $dt$  accumulated over rain duration  $D$ .

Equations 1 and 2 apply for increments of time with constant intensity, which can be assumed for short increments (minutes). For larger time increments like hours, intensity peaks are lost; the total kinetic energy and the maximum 30-min rainfall intensity become underestimated. This has to be corrected by correction factors<sup>80,81</sup>. This correction was applied when calculating the average monthly erosivity calculated from the event-based erosivity over the years of available data. The equation can be expressed as:

$$R_{monthk} = \frac{1}{n} \sum_{j=1}^m K_1 \times K_2 \times (EI_{60})_j \quad (3)$$

where,  $R_{monthk}$  is the long-term average monthly erosivity for month  $k$  in  $\text{MJ mm ha}^{-1} \text{h}^{-1} \text{month}^{-1}$ ,  $n$  is the number of years of available hourly rainfall,  $m$  is the total number of erosive storms that occurred in the month  $k$  in all  $n$  years. The  $I_{60}$  is the maximum 60-min rainfall intensity of an erosive storm derived from the hourly rainfall data. Because the starting and ending years differ among stations, a long-term average monthly erosivity was estimated separately for each station and then averaged over the entire period. Furthermore, two conversion factors  $K_1$  and  $K_2$  convert 60-min erosivity to 1-min erosivity. The first conversion factor  $K_1$  converts the rainfall erosivity from the 60-min rainfall resolution to the 15-min rainfall. A value of 1.678 was established by Das and Jain<sup>12</sup> for this factor for Indian conditions. A value of 1.15 for factor  $K_2$  was used based on Fischer et al.<sup>80</sup> that converts the 15-min resolution to 1-min resolution. The product of  $K_1$  taken from Das and Jain<sup>12</sup> and  $K_2$  from Fischer et al.<sup>80</sup> yields 1.93, which is slightly lower than the conversion factor 2.05 found by Fischer et al.<sup>80</sup> for Germany. In this study, we used a value of 2 for the product  $K_1 \times K_2$  (a trade-off of between 1.93 and 2.05). Thus, Eq. 3 simplifies to:



$$R_{monthk} = \frac{1}{n} \sum_{j=1}^m 2 \times (EI_{60})_j \quad (4)$$

### Monthly erosivity density

The concept of erosivity density (*ED*) was introduced in RUSLE2. *ED* quantifies how erosive rainfall is in a certain area or month. Small values of *ED* indicate gentle rainfall with little potential to cause erosion. In this case, erosivity is primarily driven by the total amount of rainfall. High values suggest that high-intensity rainstorms prevail<sup>39,82,83</sup>. *ED* is calculated as:

$$ED_{month} = \frac{R_{month}}{P_{month}} \quad (5)$$

where  $ED_{month}$  is the monthly erosivity density in  $\text{MJ mm ha}^{-1} \text{ h}^{-1} \text{ mm}^{-1}$ ,  $R_{month}$  is the long-term monthly rainfall erosivity in  $\text{MJ mm ha}^{-1} \text{ h}^{-1} \text{ month}^{-1}$ , and  $P_{month}$  is the long-term monthly precipitation in  $\text{mm month}^{-1}$ .

### Erosivity estimation for stations with monthly climate data

Transfer functions are required to utilize the much higher number of stations with monthly data compared to the stations with hourly data. Traditionally, linear regressions are used for this task<sup>39,84</sup>. We used XGBoost (Extreme Gradient Boosting), a machine learning model<sup>85,86</sup>, to estimate monthly erosivity. For comparison, we also applied linear regression, given its widespread use in previous studies<sup>34,35</sup>. Additionally, we explored a hybrid approach (ensemble modelling<sup>87</sup>) that combines both XGBoost and linear regression models. All approaches utilized three spatially resolved climate variables—maximum temperature, minimum temperature, and monthly rainfall—along with two geographical covariates, elevation and slope, to estimate monthly erosivity (Eq. 6). Notably, all datasets were synchronized based on the availability of the rainfall data, ensuring consistent temporal coverage across all variables. These variables were selected based on an extensive review of previous research and due to the long-term availability of consistent datasets across India. Similar predictors have been used for the erosivity estimation in several earlier studies<sup>15,39,88–90</sup>.

$$R_{month} = f(P_{month}, T_{min_{month}}, T_{max_{month}}, Elevation, Slope) \quad (6)$$

where  $R_{month}$  is the monthly erosivity,  $P_{month}$  is the monthly rainfall,  $T_{min_{month}}$  is the average minimum temperature, and  $T_{max_{month}}$  is the average maximum temperature of a particular month.

The function  $f$  can either be a linear regression model, similar to previous studies, for monthly erosivity estimation<sup>39</sup>:

$$R_{month} = \beta_0 + \beta_1 \times P_{month} + \beta_2 \times T_{min_{month}} + \dots + \beta_n \times Slope \quad (7)$$

where  $\beta_0, \beta_1, \dots, \beta_n$  are the regression coefficients that were obtained by using the *lm* function within the *Caret* package in R. Various combinations of independent parameters were tested to calibrate the most suitable multiple linear regression model based on the coefficient of determination for estimating monthly erosivity.

We also employed the XGBoost model to obtain  $f$ . Similar machine learning-based models have been found efficient in annual erosivity estimation<sup>12,90–93</sup>. XGBoost is known for its high efficiency and performance, especially with large datasets, due to parallel processing and tree-pruning techniques that reduce overfitting<sup>94,95</sup>. We used 70% of the erosivity dataset for calibration and 30% for the validation. We implemented tenfold cross-validation within the training dataset to prevent overfitting and ensure model generalizability. Each iteration used one of the 10 folds as a validation set for tuning the generalized model, while the remaining nine folds were used for training. The hyper-parameters of the XGBoost models were tuned to enhance model efficiency using a Bayesian optimization<sup>96</sup>. Specifically, hyper-parameters are external configurations or settings in machine learning that control the behavior of a learning algorithm and influence the model<sup>97</sup>. We used the *xgboost* package in R for modeling, and the hyper-parameters used in this study are provided in the Supplementary Table 1.

Additionally, an ensemble model was developed by combining linear regression and XGBoost models using the *SuperLearner* algorithm in R. This algorithm employs cross-validation to evaluate the performance of multiple machine learning models and determines the optimal weights for combining their predictions, resulting in an ensemble model. This approach was employed to assess whether combining the outputs of these two models leads to improved results. Similar methods have been applied in other research<sup>98,99</sup>. However, it should be noted that the model structures used in the individual models are not exactly the same as those in the combined model, as the *SuperLearner* algorithm operates with different hyper-parameters.

### Regionalisation of monthly erosivity

#### Geographically weighted principal component analysis (GWPCA)

The regionalization of station-based datasets was conducted using the Geographically Weighted Regression (GWR) and a Principal Component Analysis (PCA). Specifically, PCA was used to integrate efficient spatial covariates in the interpolation process. Traditional PCA often overlooks spatial heterogeneity among the factors<sup>100</sup>. However, recent advancements in integrating spatial information within PCA have revealed previously obscured details that may influence interpolation outcomes. The Geographically Weighted Principal Component Analysis

(GWPCA)<sup>101,102</sup>, an advanced extension of traditional PCA, derives spatially varying principal components by computing the local variance–covariance matrix. The weighted matrix can be expressed as follows:

$$\sum (u) = X^T \times W(u) \times X \quad (8)$$

The  $u$  represents the spatial location of the covariates,  $X$  denotes the original co-variate matrix and  $W(u)$  is the spatial location weight matrix. The weight matrix was computed using the Gaussian Kernel function.

The GW eigenvalues and eigenvectors were computed as:

$$L(u) \times V(u) \times L(u)^T = \sum (u) \quad (9)$$

where,  $L(u)$  is the matrix of eigenvectors and  $V(u)$  is the matrix of eigenvalues.

Finally, the GPWCA score matrix was calculated as:

$$S(u) = XL(u) \quad (10)$$

In this study, we utilized a 30-arc s spatial resolution dataset comprising rainfall (mm), elevation (m), slope (degree), solar radiation ( $\text{kJ m}^{-2} \text{day}^{-1}$ ), wind speed ( $\text{m s}^{-1}$ ), soil moisture ( $\text{m}^3 \text{m}^{-3}$ ), surface runoff (m), soil types, land use types and water vapor pressure (kPa) as spatial covariates for the GWPCA. The cumulative proportion of variance explained by the principal components was used to determine the number of components to retain, with a threshold of at least 85% variance explained. The resulting GWPCA score matrix was then used as input covariates in the Geographically Weighted Regression (GWR) model.

#### Geographically weighted regression (GWR)

Spatial interpolation can be performed by methods that rely solely on the spatial structure of the data under focus (e.g., Inverse Distance Weighting (IDW), Kriging, and Spline Interpolation)<sup>103</sup> or by approaches that additionally incorporate collocated covariates to enhance prediction accuracy<sup>104</sup>. The Geographically Weighted Regression (GWR)<sup>105</sup> takes co-variables into consideration by calculating ordinary least squares regressions of the local relationships between these co-variables and an outcome of interest, giving greater weight to observations close to the location under focus than to observations further away<sup>106</sup>. This makes GWR particularly effective for analyzing spatial characteristics of rainfall, as the regression coefficients capture the regionally varying influence of co-variables<sup>107</sup>. As a result, GWR is superior for the interpolation of spatially varying variables compared to ordinary least square methods, inverse-distance weighting, or spline functions, and is similarly suitable as kriging<sup>108,109</sup>. The GWR relationship can be expressed as:

$$y(u) = \beta_0(u) + \sum_{j=1}^n \beta_j(u) \times x_j(u) + \varepsilon(u) \quad (11)$$

where  $y(u)$  represents the dependent variables at location  $u$ , and  $x_j(u)$  are the independent variables at that same location. The  $\beta_j(u)$  represent the spatially varying parameters that need to be estimated for each location, and  $\varepsilon(u)$  is the random error term.

Local regression models were constructed using observations at a given site  $u$  and surrounding sites within a specific bandwidth by weighted least squares. The matrix can be expressed as:

$$\hat{\beta}(u) = (X^T \times W(u) \times X)^{-1} \times X^T \times W(u) \times y \quad (12)$$

where,  $y$  is the  $n \times 1$  vector of the dependent variable;  $X$  is the matrix of the independent variables and  $\hat{\beta}(u) = (\hat{\beta}_0(u) \dots \hat{\beta}_n(u))$  is the regression coefficient vector at site  $u$ . The weight matrix  $W(u)$  applies geographical weights to each observation for the regression at point  $i$ . It is calculated with a kernel function based on the regression point  $i$  and  $n$  data points around it within a specific bandwidth. The Akaike information criterion was selected to determine the optimal bandwidth with a Gaussian kernel function automatically. The Gaussian kernel  $w_{ij}$  can be written as:

$$w_{ij} = \exp\left(-\frac{1}{2} \left(\frac{d_{ij}}{b}\right)^2\right) \quad (13)$$

where  $w_{ij}$  is the weight between location  $i$  and  $j$ .  $d_{ij}$  is the distance between these locations, and  $b$  is the bandwidth in the Gaussian kernel.

#### Analyzing geo-climatic drivers of monthly erosivity

The spatial and temporal variation of rainfall erosivity is significantly influenced by geo-climatic variables (e.g. Mounirou et al.<sup>110</sup> in Africa, Chen et al.<sup>111</sup> in China). Therefore, a geo-climatological assessment of monthly rainfall erosivity was conducted to understand its spatial and temporal variability across India. This analysis integrated key climatic drivers—rainfall, solar radiation, soil moisture, wind speed, water vapor pressure, and surface runoff—with geographical factors such as elevation, slope, distance from the sea, soil type, and land

use. Monthly erosivity data from all stations were correlated with these eleven geo-climatic variables using the XGBoost model, as previously applied, to capture monthly variations and the underlying relationships among the variables.

To interpret the influence of each variable on the model output, we employed the SHAP (SHapley Additive exPlanations) approach using the *SHAPforxgboost* package in R. SHAP provides consistent and locally accurate explanations by decomposing model predictions into the additive contributions of individual input features<sup>94,112</sup>. This approach not only highlights the relative importance of each variable in explaining variations in rainfall erosivity but also reveals the direction (positive or negative) and nature of their influence across different months, offering a transparent and interpretable understanding of erosivity dynamics.

### Temporal trend analysis of erosivity and its attributions

The Modified Mann–Kendall test<sup>113–115</sup> is an enhanced non-parametric method used for detecting trends in environmental and climate-related time series data, particularly when the assumption of data independence is compromised. In this study, the Modified M–K test, proposed by Hamed and Rao (1998)<sup>115</sup>, and implemented using the *modifiedmk* package in R, was applied to assess long-term trends in monthly rainfall erosivity attributes. This test is preferred over the standard Mann–Kendall test because it adjusts the variance based on the autocorrelation structure of the dataset, thereby reducing the risk of Type I errors and false trend detection. Autocorrelation is a common feature in hydroclimatic time series, especially for high-resolution monthly or daily data, and can inflate the likelihood of detecting spurious trends if uncorrected.

The null hypothesis of the Modified M–K test assumes that there is no trend (i.e., the data are independent and randomly ordered), while the alternative hypothesis suggests the presence of a trend. The magnitude of the trend in the time series was estimated using Sen's slope estimator<sup>116</sup>. It provides a robust estimate of the median rate of change per unit time, with signs indicating the direction (positive or negative) of the trend.

Given that trend analysis was conducted on 96 combinations (12 months × 8 attributes), the risk of inflated Type I error due to multiple comparisons was addressed using the False Discovery Rate (FDR) correction to adjust the p-values. This approach helps control the expected proportion of false positives among the identified significant results, thereby enhancing the reliability and interpretability of the statistical inferences. By applying the FDR adjustment, we ensured that the detected trends reflect true underlying changes rather than random fluctuations.

### Metrics of error evaluation

We used three statistical metrics for accuracy assessment by applying the *hydroGOF* package in R: Percentage Error (PE), the coefficient of determination ( $R^2$ ), and the Root Mean Squared Error (RMSE). The description of these metrics can be found in previous studies<sup>117,118</sup>. They are calculated according to Eqs. 14 to 16. Additionally, we applied ANOVA (Analysis of Variance), a statistical method used to test whether the variation in the residuals can be explained by the variance in soil type, rainfall category, elevation, and land use category. A p-value of less than 0.05 was considered statistically significant, corresponding to a 95% confidence level in detecting an effect of the tested factor on the residual variability.

$$PE = \frac{(P_i - O_i)}{O_i} \times 100\% \quad (14)$$

$$R^2 = \left\{ \frac{\sum_{i=1}^N (O_i - \bar{O}) (P_i - \bar{P})}{\sqrt{\sum_{i=1}^N (O_i - \bar{O})^2} \sqrt{\sum_{i=1}^N (P_i - \bar{P})^2}} \right\}^2 \quad (15)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (O_i - P_i)^2}{N}} \quad (16)$$

where  $O$  is the observed value,  $P$  is the estimated value,  $\bar{O}$  is the mean of observed values,  $\bar{P}$  is the mean of estimated values, and  $N$  is the total number of values.

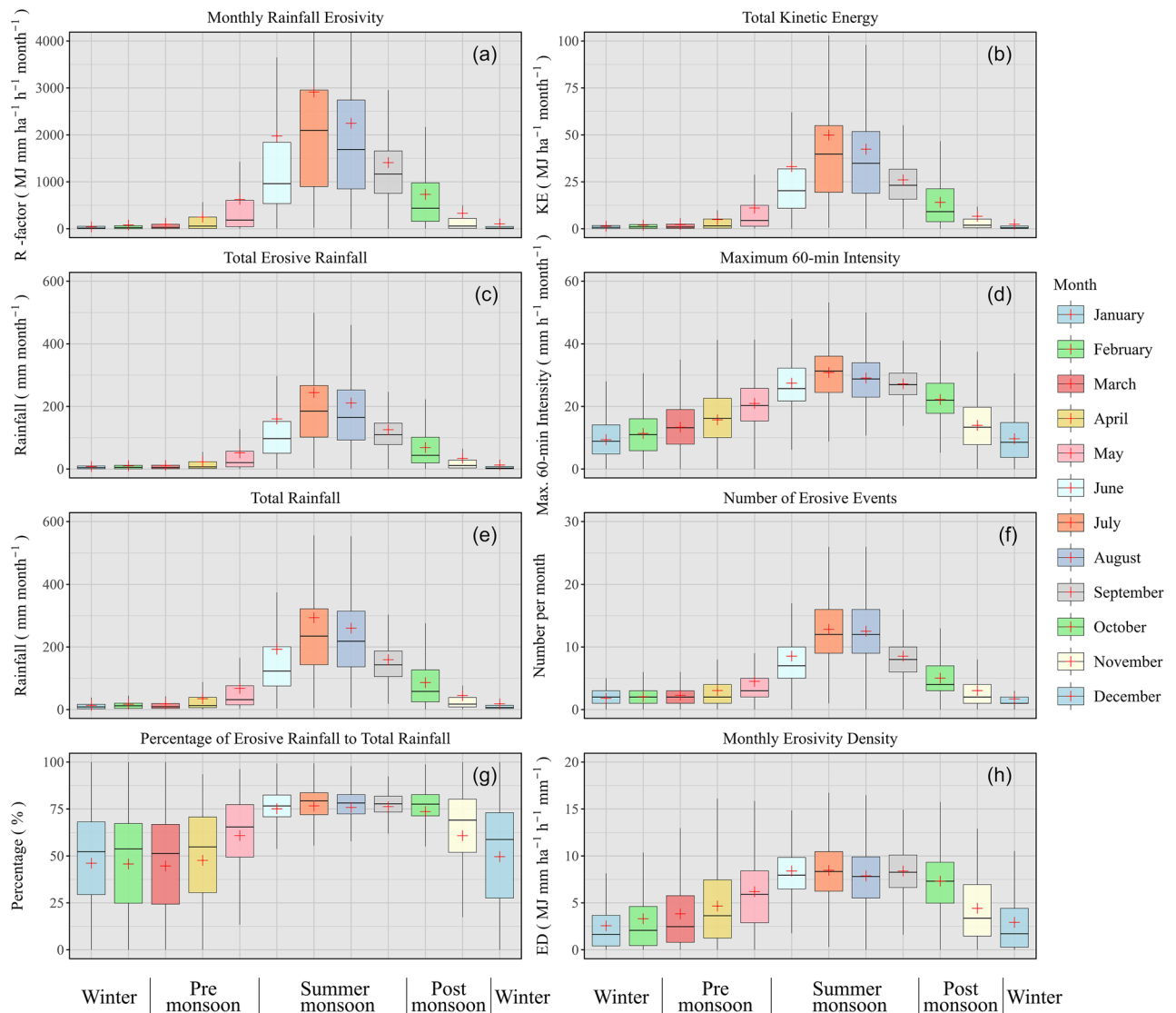
## Results

### Rainfall erosivity and attributes at stations with hourly rainfall data

The long-term mean rain erosivity at stations with hourly rainfall data exhibited significant spatial and temporal variation (Fig. 2). Monthly erosivity values exceeded 20,000 MJ mm ha<sup>−1</sup> h<sup>−1</sup> month<sup>−1</sup> at certain stations, while other stations recorded zero erosivity within the same months (for detailed minimum and maximum values, refer to Supplementary Table 2). Temporal variation was nearly as pronounced as spatial variation, with values approaching zero at most stations during the winter months (December and January).

When focusing on the majority of stations and disregarding the spatial pattern, three distinct seasonal patterns emerged for rain erosivity and its related attributes:

1. A marked peak in July and August was observed for rain erosivity (Fig. 2a), total kinetic energy (Fig. 2b), total erosive rainfall (Fig. 2c), total rainfall (Fig. 2e), and the number of erosive events (Fig. 2f). The two months preceding and following this peak showed either ascending or descending trends, while the remaining six months of the year exhibited very low values. This group of attributes, including rain erosivity, appeared to be primarily governed by total rainfall, which predominantly occurred in events large enough to be classified as erosive. Remarkably, the number of erosive events was larger than 15 for many stations between June and



**Fig. 2.** Monthly distribution of rainfall erosivity and its attributes for hourly rainfall stations in India, taking the long-term average of each station. The central line of the box plots indicates the median, the edges of the box represent the interquartile range, and the whiskers show the range within 1.5 times the interquartile range. The ‘+’ sign within the box indicates the mean value. For the minimum and maximum values, see Table 2 in the Supplementary. The plots were generated using R version 4.4.3 with the package *ggplot2*.

- August, meaning that an erosive event occurred at least every second day. This was true for July and August for one-quarter of all stations. Notably, storm intensity was significantly high during these months.
- A more gradual inter-month variation, also peaking in July, was found for the maximum 60-min intensity (Fig. 2d) and monthly erosivity density (ED) (Fig. 2h). The differences between months in this group were significantly smaller compared to the first group.
- The third group comprised only the percentage of erosive rainfall relative to total rainfall (Fig. 2g). This attribute also exhibited a gradual change between months, similar to the second group, though with different within-month variability. From June to October, approximately 75% of total rainfall was erosive at 50% of the stations, while in the other months, the interquartile range was notably large (~50%), indicating pronounced spatial variation. This spatial variability was further highlighted by the minimum and maximum values (Table 2 in the Supplementary). In almost all months, the values ranged from 0%—where no rainfall was erosive over the multi-year observation period at specific stations—to 100%, where all rainfall was erosive at other stations.

Overall, the values of rainfall erosivity and its attributes generally followed the seasonal cycle, with low values in winter, intermediate values during the pre-monsoon and post-monsoon periods, and highest values during the summer monsoon. However, the alignment of the monsoon season with rainfall and erosivity attributes was only approximate. The pre-monsoon season exhibited values nearly as low as those in winter. The summer monsoon

displayed considerable variation, with a continuous and pronounced increase starting low in May and peaking in August. The post-monsoon season began with values comparable to those observed in June, during the middle of the monsoon, before rapidly decreasing until November, when values were similar to the low levels recorded in December and January.

Validation of the rainfall erosivity model for utilizing monthly stations

A correlation analysis between monthly erosivity and independent climate parameters indicated a strong correlation to monthly precipitation, especially during the summer monsoon months ( $r > 0.70$ ; see Supplementary Figs. 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 and 13). Monthly minimum temperature showed a moderate correlation ( $r \sim 0.6$ ) during the post-monsoon months, while other parameters demonstrated weaker correlations. The XGBoost model outperformed linear regression by far in most metrics (Table 1); for variable importance plots, refer to Supplementary Figs. 14, 15 and 16. XGBoost consistently achieved a median percentage error within  $\pm 10\%$  across almost all months, with January being the only notable exception ( $-11.7\%$ ). In contrast, the linear regression model exhibited substantially larger errors, ranging from significant underestimations in February ( $-33.6\%$ ) and March ( $-18.4\%$ ) to overestimations in April ( $+14.1\%$ ). The  $R^2$  and RMSE values further confirmed that regional XGBoost models outperformed the globally used linear regression models and are better suited for estimating erosivity at stations with only monthly data.

Additionally, an ensemble model combining XGBoost and linear regression yielded better performance than linear regression alone. Although the ensemble model exhibited slightly higher RMSE values than XGBoost during certain winter months, it consistently outperformed linear regression across all months. Notably, the ensemble approach led to a significant reduction in RMSE (ranges from 30 to 633 MJ mm ha<sup>-1</sup> h<sup>-1</sup> month<sup>-1</sup>) and an improvement in  $R^2$  (ranges from 0.83 to 0.99) during the monsoon months. The XGBoost model, trained individually for each month with a significant number of hyper-parameters, demonstrated the lowest RMSE across all models, highlighting its robustness and effectiveness. Although the  $R^2$  values of the ensemble model were a little higher for some months, XGBoost consistently delivered superior overall performance. Therefore, this study adopts the XGBoost model for monthly erosivity estimation due to its superior predictive performance.

Validation of erosivity regionalisation

Initially, simple kriging interpolation of monthly rainfall erosivity was performed using only the spatial distribution of the dataset, without incorporating additional variables. This approach resulted in substantial interpolation errors, with low predictive performance ( $R^2 < 0.8$  for all months; Supplementary Table 3). To improve accuracy, we adopted a more advanced methodology combining Geographically Weighted Principal Component Analysis (GWPCA) with Geographically Weighted Regression (GWR). The performance metrics of the GWPCA-GWR interpolation of monthly rainfall erosivity revealed significant variation throughout the year (Table 2). The percentage error ranged from  $+11.5\%$  in January to  $-18\%$  in May, with the lowest error observed in September ( $-2.5\%$ ). The  $R^2$  values fluctuated between 0.72 in February and 0.98 in October, indicating variability in the model's predictive accuracy. Notably, a strong median  $R^2$  of 0.90 was observed across the months, reflecting overall good model performance. The Root Mean Squared Error (RMSE) also exhibited significant differences across months, with the highest value recorded in July (499 MJ mm ha<sup>-1</sup> h<sup>-1</sup> month<sup>-1</sup>) and the lowest in January (12 MJ mm ha<sup>-1</sup> h<sup>-1</sup> month<sup>-1</sup>). The high value in July is most likely due to the elevated rainfall erosivity across India during this month, as July typically experiences significant rainfall throughout the country. Conversely, the lower  $R^2$  values during the winter months, such as January and February, can be attributed to minimal or zero erosivity. The limited variability in erosivity values during these periods presents challenges for interpolation models, as it hinders their ability to accurately predict erosivity.

Months	XGBoost			Linear regression			Ensemble model		
	Median PE	R <sup>2</sup>	RMSE	Median PE	R <sup>2</sup>	RMSE	Median PE	R <sup>2</sup>	RMSE
January	-11.7	0.92	10	-8.4	0.45	62	-6.5	0.83	30
February	-3.7	0.94	11	-33.6	0.37	108	+2.1	0.90	50
March	-3.3	0.97	9	-18.4	0.60	170	-5.6	0.87	64
April	-1	0.98	9	+14.1	0.59	251	-2.7	0.97	105
May	-2.7	0.98	20	+7.5	0.87	680	-6.1	0.98	208
June	-1	0.97	67	-8.0	0.79	2267	-10.6	0.98	540
July	-5.4	0.97	134	-12.6	0.89	2266	-6.6	0.99	613
August	-1.5	0.98	100	+0.4	0.88	997	-6.1	0.96	633
September	+1	0.92	318	-8.9	0.66	478	-2.5	0.95	420
October	+2.4	0.97	169	-7.7	0.82	429	-1	0.94	202
November	-5.5	0.98	99	+12.1	0.81	193	+1	0.96	133
December	+4.1	0.98	16	-0.3	0.91	128	-8.8	0.99	44

**Table 1.** Evaluation metrics for the linear regression, XGBoost, and ensemble models used in monthly rainfall erosivity estimation. Metrics were derived by applying the transfer functions developed from the calibration dataset ( $n = 183$  hourly stations) to an independent test dataset ( $n = 78$  hourly stations). PE, Percentage Error (%);  $R^2$ , Coefficient of determination; and RMSE, Root Mean Squared Error (MJ mm ha<sup>-1</sup> h<sup>-1</sup> month<sup>-1</sup>).



Months	Median PE	R <sup>2</sup>	RMSE
January	+ 11.5	0.73	12
February	− 3.0	0.72	28
March	− 6.8	0.89	39
April	− 12.9	0.89	88
May	− 18	0.88	243
June	− 6.9	0.88	263
July	− 9.1	0.90	499
August	− 4.4	0.93	482
September	− 2.5	0.91	336
October	− 4.2	0.98	171
November	− 2.6	0.95	50
December	+ 4.9	0.92	23

**Table 2.** Spatial interpolation error metrics for the monthly erosivity mapping using a combination of Geographically Weighted Principal Component Analysis (GWPCA)-Geographically Weighted Regression (GWR) from the calibration dataset ( $n = 1950$  hourly and monthly stations) when applied to the test dataset ( $n = 836$  hourly and monthly stations). PE, Percentage Error (%);  $R^2$ , Coefficient of determination; and RMSE, Root Mean Squared Error ( $\text{MJ mm ha}^{-1} \text{ h}^{-1} \text{ month}^{-1}$ ).

Regionalisation of annual and monthly erosivity

Annual rainfall erosivity in India exhibited a highly variable spatial distribution (Fig. 3a), ranging from as low as  $500 \text{ MJ mm ha}^{-1} \text{ h}^{-1} \text{ year}^{-1}$  in narrow, rain-shadowed valleys of the Western Himalayas and the northeastern tip of the Eastern Himalayas to values exceeding  $20,000 \text{ MJ mm ha}^{-1} \text{ h}^{-1} \text{ year}^{-1}$  in regions such as the Western Ghats, the central highlands, and the southern slopes of the Khasi Hills. The spatial gradients of rain erosivity were particularly pronounced in the northeast, where the full range of values was observed within distances of  $\sim 200 \text{ km}$ . A significant portion of India, including the semi-arid northwest (mean rainfall erosivity values for different climatic conditions can be found in Supplementary Tables 4 and 5), the Deccan Plateau, and the western Gangetic Plain, experiences annual erosivities between  $2,000$  and  $5,000 \text{ MJ mm ha}^{-1} \text{ h}^{-1} \text{ year}^{-1}$ . In contrast, low erosivity values, ranging from  $400$  to  $700 \text{ MJ mm ha}^{-1} \text{ h}^{-1} \text{ year}^{-1}$ , are characteristic of the Thar Desert and the Trans-Himalayan region.

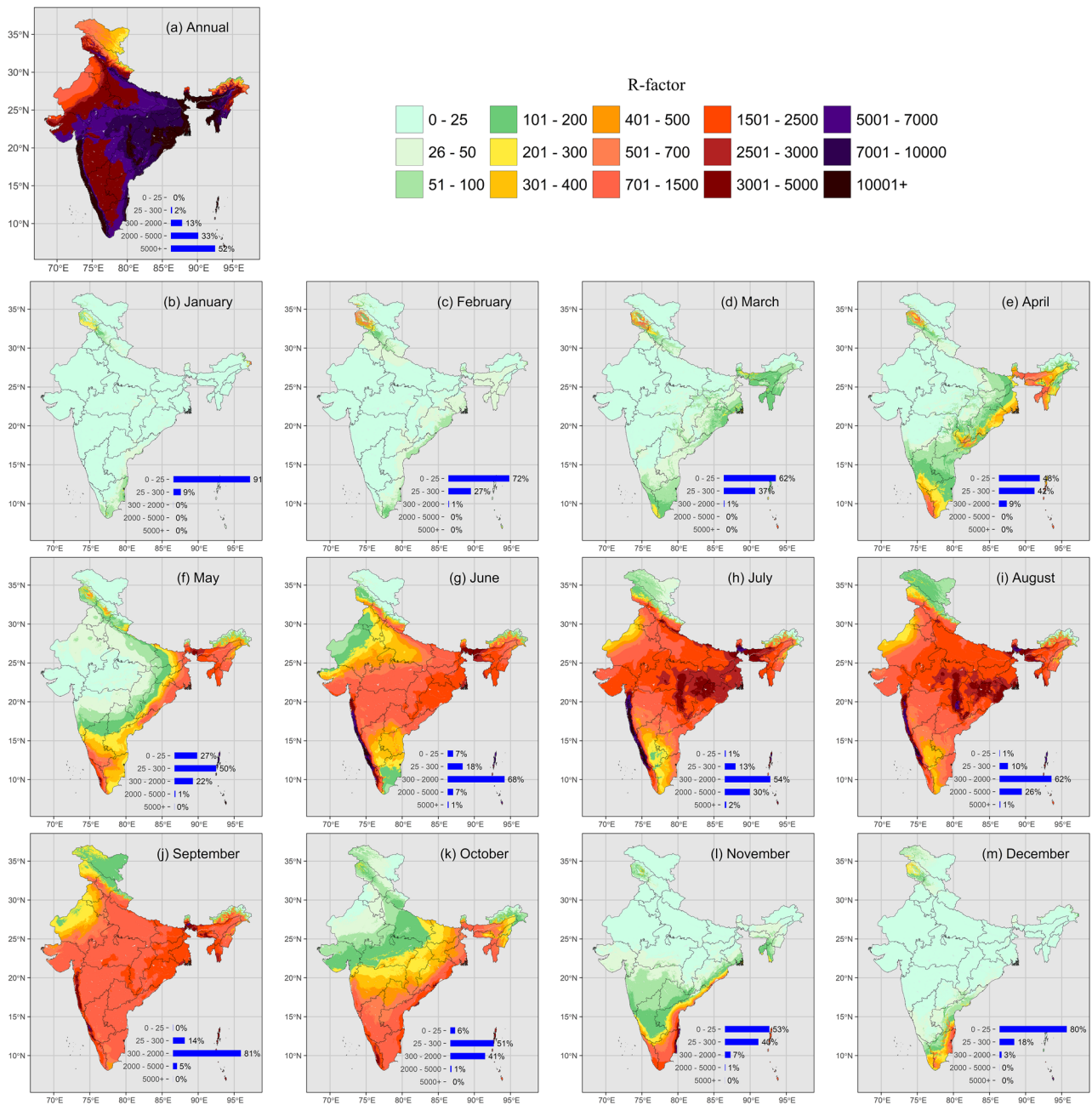
The spatial variability of erosivity became even more pronounced when examined on a monthly basis, as the temporal distribution of rain erosivity differed significantly across regions, partially averaging out in the annual totals. During January and February, rain erosivity was minimal across most of India, with the exception of the Trans-Himalayas (Fig. 3b, c). More than 70% of the country exhibited erosivities below  $25 \text{ MJ mm ha}^{-1} \text{ h}^{-1} \text{ month}^{-1}$  in January and February. By late March, as the high-sun season progresses toward the Equator, increasing atmospheric instability and convective rainfall cause erosivity to rise, particularly along the eastern parts of India (Fig. 3d), a trend that intensifies in April and May (Fig. 3e, f). Erosivities already reached values of  $2,000 \text{ MJ mm ha}^{-1} \text{ h}^{-1} \text{ month}^{-1}$ , and 2% of the country even exceeded this value in May. High erosivity values also emerged during this period in the Cardamom Hills, the Eastern Ghats, and the Northeast.

With the onset of the summer monsoon in early June, which advances from the south and reaches the northern regions, excluding the Thar Desert, by the end of the month, rain erosivity became uniformly high across most of India (Fig. 3g). Erosivity was within a range of  $300$  to  $2,000 \text{ MJ mm ha}^{-1} \text{ h}^{-1} \text{ month}^{-1}$  in two-thirds of India. By July, the monsoon—and the associated high erosivity—extends into the Thar Desert as well (Fig. 3h), a pattern that continues through August (Fig. 3i). This relatively uniform pattern persisted until September (Fig. 3j). During the peak of the monsoon season, from July to August, the increased monsoonal activity over the Arabian Sea significantly amplified erosivity in the Western Ghats, where orographic effects lead to spectacular rainfall as moist monsoon winds are blocked by the steep slopes. As a result, the Western Ghats experienced markedly higher erosivities during these months compared to other regions, even exceeding  $5,000 \text{ MJ mm ha}^{-1} \text{ h}^{-1} \text{ month}^{-1}$ .

By early October, the winter monsoon begins, lasting until December. It brings rain to the southeastern Deccan Plateau and the Eastern Ghats. In October, high erosivity was observed along the entire eastern part of India, but by December, it retreated to the southeastern tip of India (Fig. 3k–m). Consequently, the southern tip of India experienced a prolonged erosion season, spanning from April to December. The Western Himalayas also exhibited a comparably even temporal distribution in rain erosivity. In contrast, the Western Ghats experienced extremely high erosivity concentrated within just three months—June, July, and August—while the Thar Desert had an even shorter erosion season, limited to July and August.

Influence of geo-climatic variables on monthly erosivity

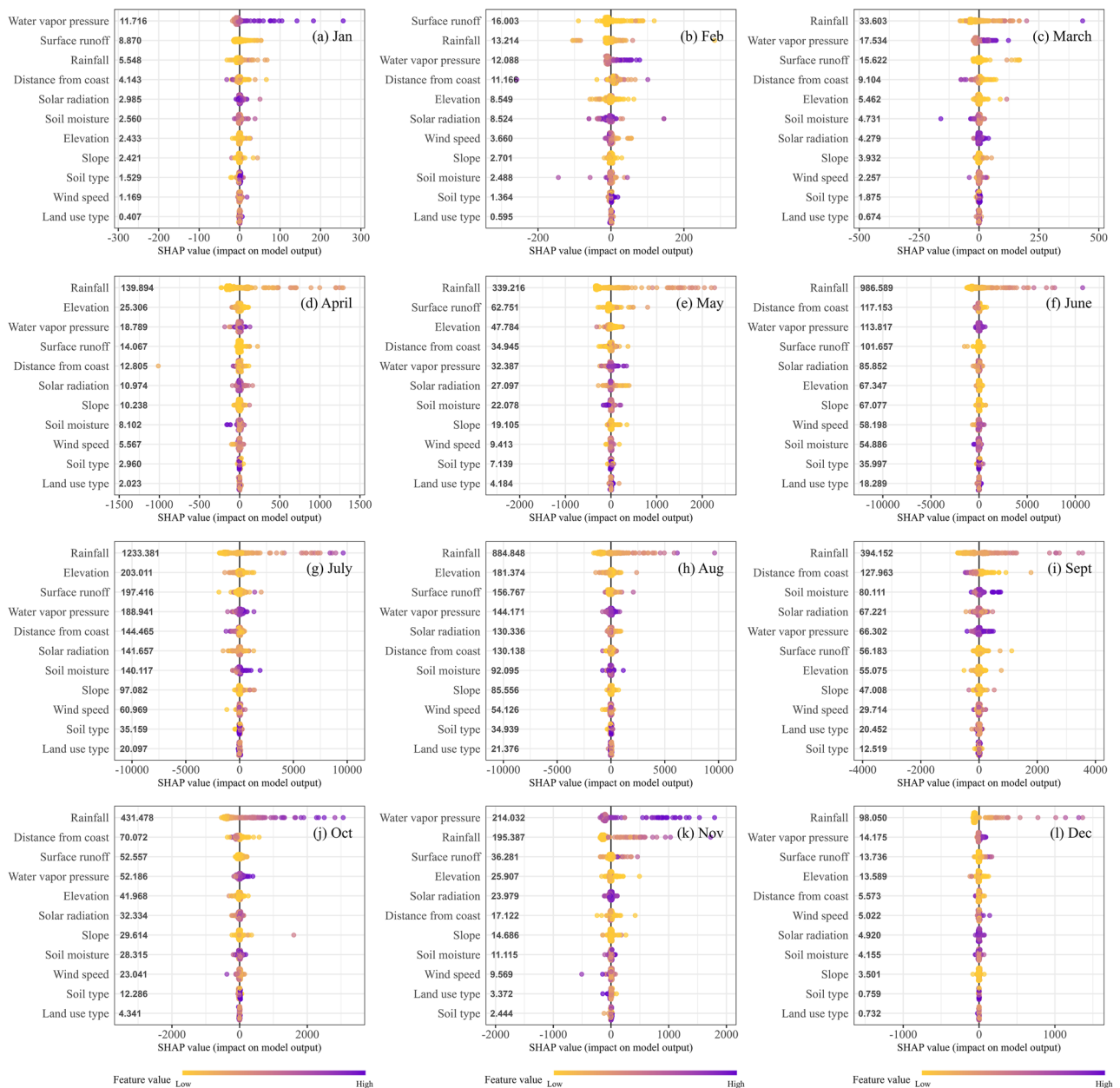
The SHAP (SHapley Additive exPlanations) analysis in Fig. 4 highlights the monthly contribution of geo-climatic variables to the estimation of rainfall erosivity across India. The results reveal notable seasonal variability in the influence of predictors, which reflects the dynamic nature of erosivity processes under varying climatic and geographic conditions. To better understand the role of individual variables throughout the year, we grouped the predictors based on the frequency of their relative importance across the months (details have been summarized



**Fig. 3.** Estimated long-term average (1969–2021) of annual and monthly erosivity across India. Units differ by panel: a)  $\text{MJ mm ha}^{-1} \text{h}^{-1} \text{year}^{-1}$  for annual erosivity; b) to m)  $\text{MJ mm ha}^{-1} \text{h}^{-1} \text{month}^{-1}$  for the monthly erosivity. Rounded areal percentages are shown as bar plots at the bottom of each panel; minor rounding discrepancies may result in slight deviations. For prediction errors associated with the maps, refer to Table 2. The maps were generated using R version 4.4.3 with the packages *ggplot2*, *sf*, and *raster*.

in the Supplementary Table 6). The rankings were categorized as high, medium, or low importance using their SHAP value magnitudes and positions.

1. High Importance (frequently among the top 3 contributors): Rainfall consistently emerged as the dominant driver of erosivity across all months, with its influence peaking during the summer monsoon period (June to September) (Fig. 4f–i); July recorded the highest importance score, underscoring rainfall's critical role during peak erosive months. Surface runoff also exhibited a strong influence, particularly in the pre-monsoon (Fig. 4c–e) and summer monsoon seasons, emphasizing its control over erosive energy and soil detachment. Additionally, water vapor pressure demonstrated significant importance, especially during the winter (Fig. 4a, b, and l) and post-monsoon (Fig. 4j and k) months, where it contributed notably to erosivity processes during drier periods.



**Fig. 4.** SHAP (SHapley Additive exPlanations) summary plot showing the importance of geo-climatic predictors contributing to the spatial variation of rainfall erosivity across India. The variables on the Y-axis are ranked by their importance, from high (top) to low (bottom), based on their average SHAP values. Each point represents a SHAP value for an individual observation, with color indicating the magnitude of the feature value—violet for high values and yellow for low values. The plot illustrates the contribution of each geo-climatic variable to rainfall erosivity, with wider distributions indicating greater variability in feature impact across the dataset. The plots were generated using R version 4.4.3 with the package *ggplot2*.

2. Medium Importance (intermittent or moderate influence): Distance from the coast showed moderate importance during the pre-monsoon and winter seasons, likely due to its impact on regional moisture availability and atmospheric dynamics. Solar radiation contributed moderately during several months, particularly in transitional periods, reflecting its role in influencing storm intensity and evaporation. Elevation also demonstrated a moderate influence, affecting rainfall distribution and erosivity gradients across varying terrains, especially during the pre-monsoon and monsoon months.
3. Low Importance (consistently low SHAP rankings): Soil moisture generally ranked low in importance, with only occasional influence observed in specific months (e.g., September). Wind speed also played a minor role throughout the year, likely due to its indirect or limited connection to monthly-scale erosivity processes. Similarly, slope consistently showed minimal importance across all months, suggesting a secondary role in influencing erosivity at the national scale. Additionally, soil type and land use exhibited consistently low

SHAP values, indicating that their contribution to monthly erosivity estimation was relatively minor across India.

Long-term trends in monthly erosivity attributes

There were significant changes in 33 out of 96 cases (12 months  $\times$  8 attributes) (Table 3) based on the Modified Mann–Kendall test (see p-values in Supplementary Table 7). However, this number must be interpreted cautiously, as the probability of false positives (Type I errors) is high due to multiple testing. Notably, all erosivity-related attributes with significant changes showed increasing trends, except for the number of erosive events, which exhibited a decreasing trend in six out of 12 months, with only one month (August) showing a statistically significant decline.

For all other attributes, increases were more common and were found to be significant in two to four months, suggesting that all attributes other than the number of events followed a similar trend (see Supplementary Table 7). A notable seasonal variation in the trends was observed—out of the 33 significant cases, 23 occurred between October and January, a period corresponding to the post-monsoon and winter seasons when overall erosivity is typically low. This was particularly evident in monthly erosivity, where increases during these months have limited relevance to the total annual erosivity.

However, after applying False Discovery Rate (FDR) corrections to control for multiple testing, 23 out of the 33 initially significant trends were found to be statistically insignificant. Only 10 cases remained significant (see adjusted p-values in Supplementary Table 7), and most of these were found in the post-monsoon and winter months, which are generally non-erosive.

This seasonal clustering of significant changes, mainly in low-erosivity months, may reflect subtle shifts in rainfall intensity or timing, potentially driven by changing climate dynamics. In contrast, the peak monsoon season (June to September)—which contributes the most to annual erosivity—remained relatively stable with fewer significant changes.

Discussion  
Comparison with the global dataset

We compared the estimated monthly erosivity at a  $\sim 1\text{ km}^2$  resolution with the Global Rainfall Erosivity Dataset version 1.2 (GloREDa v1.2)<sup>90</sup> for each month (Fig. 5). Large portions of India exhibit negative differences during the pre-monsoon months (March to May) (Fig. 5c–e), with this study’s erosivity estimates being lower than GloREDa by more than  $-50\%$  in Northeast India and the Eastern Ghats. However, certain regions, such as the Central Highlands, show positive differences, where estimated erosivity exceeds GloREDa estimates.

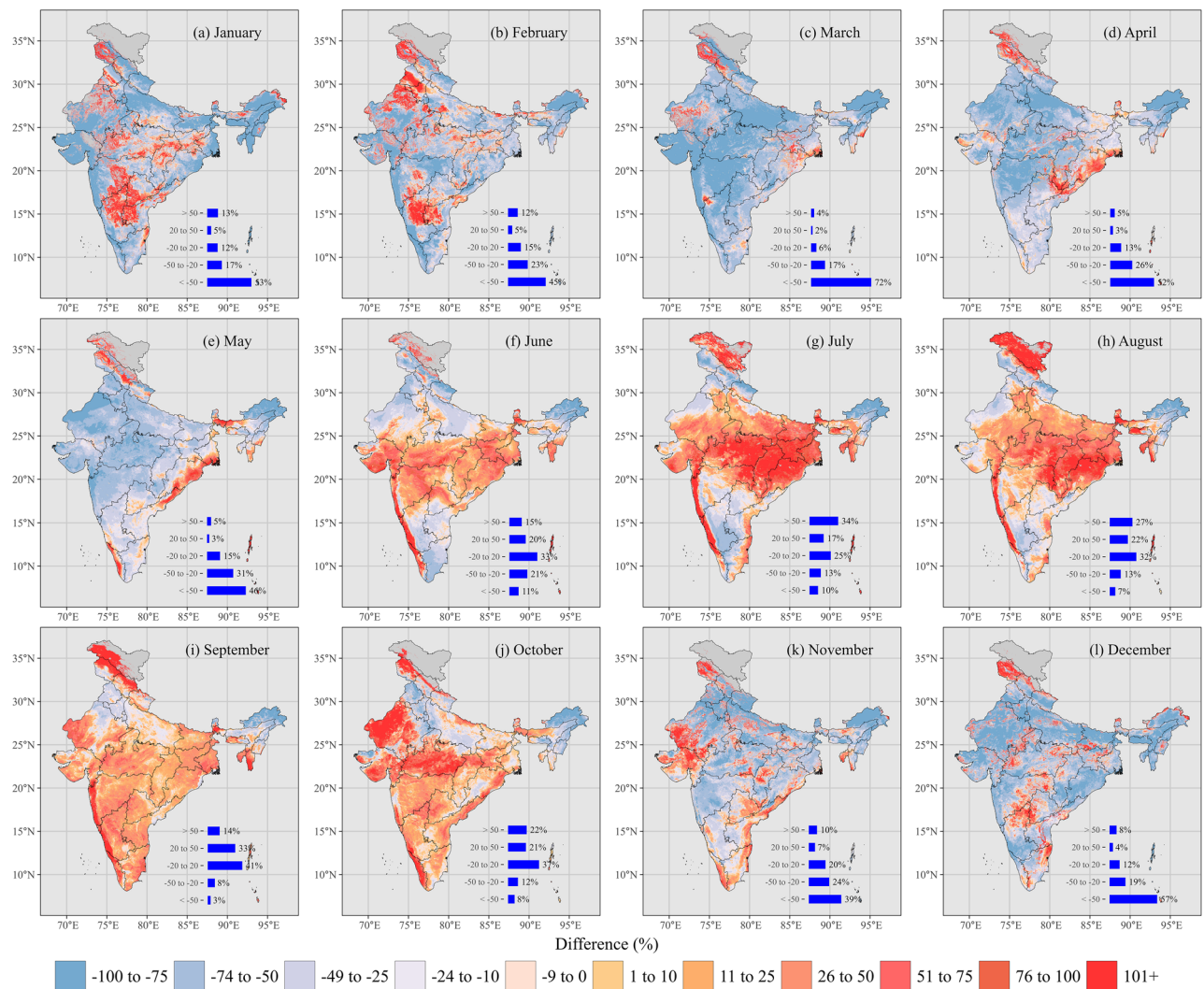
In May, negative differences persist in the Thar Desert region and central Highlands, ranging from  $-75\%$  to  $-25\%$ , while positive differences (up to  $+75\%$ ) are observed along the Eastern Ghats. Significant positive differences are seen during the peak monsoon months (June to September) (Fig. 5f–i), with erosivity in the study area exceeding GloREDa estimates by  $+20\%$  to more than  $+50\%$  in Central Highlands and the Western Ghats. By September and October (Fig. 5i and j), erosivity differences begin to decrease, showing a mix of negative and positive differences. In the winter season (Fig. 5a, b, and l), stronger negative differences appear in the plains, but positive differences are noted in the western Himalayas and the Deccan Plateau regions.

Overall, our results indicate that the rainfall erosivity estimates from this study tend to be lower during the dry months and higher during the monsoon months (Fig. 5c–k), when compared to GloREDa v1.2. This variation suggests that global datasets such as GloREDa may underestimate the local erosive power of intense monsoon rainfall, largely due to limited observational data (e.g., reported in earlier studies<sup>32,119–121</sup>).

Month	Total erosivity	Total kinetic energy	Total erosive rainfall	Maximum 60-min intensity	Total rainfall	Number of erosive events	Percentage of erosive rainfall to total rainfall	Monthly erosivity density
January	+0.86	+0.02	+0.10	+0.10	+0.10	−0.01	+0.25	+0.06
February	+0.34	+0.01	+0.04	+0.04	+0.03	−0.01	+0.21	+0.02
March	+0.98	+0.02	+0.07	+0.03	+0.09	−0.02	+0.15	+0.02
April	+0.58	+0.03	+0.10	+0.00	+0.13	−0.02	+0.05	0.00
May	+3.18	+0.03	+0.10	+0.06	+0.13	0.00	+0.06	+0.03
June	+0.15	+0.01	−0.03	+0.02	−0.01	0.00	+0.00	0.00
July	+7.40	+0.11	+0.44	+0.03	+0.33	−0.01	+0.06	+0.01
August	−0.34	−0.06	−0.39	+0.02	−0.45	−0.03	−0.02	+0.02
September	+3.47	+0.07	+0.34	+0.03	+0.36	0.00	+0.05	+0.01
October	+9.01	+0.15	+0.67	+0.10	+0.77	+0.04	+0.07	0.03
November	+3.79	+0.09	+0.48	+0.09	+0.52	+0.06	+0.07	+0.01
December	+3.32	+0.06	+0.32	+0.07	+0.36	0.00	+0.15	+0.05

**Table 3.** Sen’s slope estimates per year for monthly rainfall erosivity and its contributing components, derived from the spatial average of 261 hourly stations across India during the period 1969–2021. Total erosivity in MJ mm ha<sup>−1</sup> h<sup>−1</sup> month<sup>−1</sup>; Total kinetic energy in MJ ha<sup>−1</sup> month<sup>−1</sup>; Total erosive rainfall in mm month<sup>−1</sup>; Maximum 60-min intensity in mm h<sup>−1</sup> month<sup>−1</sup>; Total rainfall in mm month<sup>−1</sup>; Monthly erosivity density in MJ mm ha<sup>−1</sup> h<sup>−1</sup> mm<sup>−1</sup>.





**Fig. 5.** Percentage error (%) of long-term, highly resolved monthly erosivity estimates from this study compared to the Global Rainfall Erosivity Dataset version 1.2 (GloREDa v1.2)<sup>90</sup>. Negative values indicate that GloREDa v1.2 estimates are higher than those in this study, while positive values indicate lower estimates in GloREDa v1.2 relative to this study. The maps were generated using R version 4.4.3 with the packages *ggplot2*, *sf*, and *raster*.

Additionally, GloREDa v1.2, an extension of the first version of GloREDa, employs a different kinetic energy equation than that used in this study. Moreover, the rainfall dataset utilized in GloREDa is shorter in duration, potentially failing to capture long-term climatic variability<sup>78,122</sup>. This issue is particularly relevant in India, where significant changes in climate and rainfall patterns have occurred in recent decades<sup>123</sup>. Therefore, local and regional-scale studies are essential for accurately identifying and understanding erosivity patterns across both spatial and temporal scales.

### Comparison with past rainfall erosivity equations

The history of erosivity studies in India involves the use of various empirical equations. Most of them used the metric version of these equations until newer converted formulas were developed (Table 4). Alongside the erosivity equations of Babu et al.<sup>34,35</sup>, another popular approach based on the modified Fournier equation<sup>33,42</sup> has also been used for erosivity estimation in many studies. Initially formulated in metric units, these equations were converted to the SI unit system by Majhi et al.<sup>26</sup> and Chen et al.<sup>24</sup>. We estimated erosivity using these equations and compared the results with our observed in-situ erosivity values presented in the Supplementary Fig. 17. The results indicate that these equations tend to underestimate rainfall erosivity in most cases. Specifically, the equation of Babu et al.<sup>35</sup> even after converting to SI unit by Majhi et al.<sup>26</sup> shows a mean percentage error of almost -35% across India, with a median of almost -42%. Notably, the unit-converted equation of Arnoldus's<sup>33</sup> has a mean percentage error of -9%, with a median of -17%, though it results in both positive and negative errors ranging from -75% to +75%.



Eq. No	Original equation	Unit converted equation ( $R$ in MJ mm ha <sup>-1</sup> h <sup>-1</sup> per year or season) and rainfall range	Correlations	Study
Annual erosivity				
17	$R = 1.735 \times 10^{(1.5 \log_{10} \left( \sum_{i=1}^{12} \frac{P_i^2}{P} \right) - 0.8188)}$ $P_i$ : rainfall of month $i$ (mm month <sup>-1</sup> ); $R$ in metric units	$R = 17.02 \times 10^{(1.5 \log_{10} \left( \sum_{i=1}^{12} \frac{P_i^2}{P} \right) - 0.8188)}$	n.a	#1
18	$R = 79 + 0.363P$ $R$ in t-m cm ha <sup>-1</sup> h <sup>-1</sup> year <sup>-1</sup>	$R = 3.70P + 806$ Rainfall range: 350 to 3500 mm year <sup>-1</sup>	0.69	#2
20	$R = 81.5 + 0.380P$ $R$ in t-m cm ha <sup>-1</sup> h <sup>-1</sup> year <sup>-1</sup>	$R = 3.88P + 832$ Rainfall range: 340 to 3500 mm year <sup>-1</sup>	0.81	#3
22	$R = 14.41P - 6514$ $R$ in MJ mm ha <sup>-1</sup> h <sup>-1</sup> year <sup>-1</sup> ; Rainfall range: 226 to 8946 mm year <sup>-1</sup>		0.92	#4
Summer monsoon erosivity				
19	$R_s = 50 + 0.389P_s$ $R$ in t-m cm ha <sup>-1</sup> h <sup>-1</sup> season <sup>-1</sup>	$R_s = 3.97P_s + 510$ Rainfall range: 300 to 3000 mm season <sup>-1</sup>	0.77	#2
21	$R_s = 71.9 + 0.361P_s$ $R$ in t-m cm ha <sup>-1</sup> h <sup>-1</sup> season <sup>-1</sup>	$R_s = 3.68P_s + 734$ Rainfall range: 293 to 3190 mm season <sup>-1</sup>	0.82	#3
23	$R_s = 14.61P_s - 4693$ $R_s$ in MJ mm ha <sup>-1</sup> h <sup>-1</sup> year <sup>-1</sup> ; Rainfall range: 32 to 7448 mm season <sup>-1</sup>		0.92	#4

**Table 4.** Comparison of earlier used erosivity equations in India with this study.  $P$ , Long-term average annual rainfall;  $P_s$ , Long-term average summer-monsoon rainfall;  $R$ , Long-term average annual erosivity;  $R_s$ , Long-term average erosivity for summer monsoon season. Studies: #1 Original version: Arnoldus<sup>33</sup>, units converted by Chen et al.<sup>24</sup>, #2 Original version: Babu et al.<sup>34</sup> as cited in Singh<sup>53</sup> united converted using the procedure of Majhi et al.<sup>26</sup>, #3 Original version: Babu et al.<sup>35</sup>, units converted by Majhi et al.<sup>26</sup>, #4 This study

Additionally, when we applied a simplified rainfall erosivity equation based solely on rainfall data from 261 stations (Eqs. 22 and 23), we observed a low percentage error at the annual level; however, a high seasonal error was evident. Furthermore, the rainfall erosivity map prepared in this study showed a median percentage error of −20% even after interpolation at the annual scale, with a mean error of only −18%. Notably, the error is lower for the summer monsoon months, with a median percentage error of only −18%. Therefore, the combination of the machine learning model and GWR-based interpolation demonstrated strong performance compared to previous studies. In contrast, the empirical rainfall erosivity equations developed in earlier works are now outdated and tend to introduce significant bias across India. Several regional studies—for instance, Pandey et al.<sup>124</sup>, which estimated erosion in Indian forests, and Pal et al.<sup>125</sup>, which applied these equations for large-scale analyses—are likely to have incurred substantial errors in rainfall erosivity estimation and, consequently, in erosion estimation.

Hence, the use of such empirical equations should be approached with caution. Moreover, long-term erosivity estimates derived from these equations may vary depending on the temporal coverage of the data used to develop the equations and the period for which they are applied, which can potentially lead to significant differences in the results. We strongly recommend utilizing hourly rainfall data from the India Meteorological Department (IMD) for more accurate and reliable erosion assessments. Furthermore, despite the limited availability of hourly station data, the integration of machine learning and artificial intelligence with geo-climate variables or satellite datasets now makes it possible to estimate accurate rainfall erosivity in data-sparse regions or areas with complex topography. This approach will eventually eliminate the reliance on regionally developed empirical equations and facilitate more accurate estimation of rainfall erosivity for specific time periods and locations.

### Error analysis

There are notable sources of error that may affect this study's results. One is the use of WorldClim maximum and minimum temperature data for stations lacking observed records—excluding the 510 stations with observations. To assess the reliability of this method, we compared observed temperatures with WorldClim data at the same locations. The monthly median percentage error remains low—around ±1% for maximum temperature (Supplementary Fig. 18), and slightly higher for minimum temperature, with +3.77% in January and +3.11% in February (Supplementary Fig. 19). All errors remain within ±4%, which may partly explain the higher prediction errors noted in these months. Furthermore, similar error ranges in the WorldClim dataset have been reported in previous studies for Europe<sup>126</sup>, global scale<sup>127</sup>, and Nepal<sup>128</sup>. Given the unavailability of observed temperature records for all stations, we believe that using the WorldClim dataset is a reasonable and justifiable alternative despite these minor errors.

Another source of uncertainty is the use of correction factors to convert hourly rainfall erosivity to a 1-min scale, which inherently omits peak 30-min intensity. We compared our adopted factor with those from other studies (Supplementary Table 8), showing errors ranging from −40.67% to +73.61%. However, our results align well, showing only −3.89% error with Panagos et al.<sup>129</sup> (Europe), +6.89% with Yue et al.<sup>130</sup> (China), and −2.44% with Fischer et al.<sup>80</sup> (Germany). Given the lack of breakpoint pluviograph or 1-min rainfall datasets in many regions, our adopted correction factor offers a reliable and practical approach for utilizing hourly data in rainfall erosivity estimation.

Furthermore, we analyzed the variability in prediction errors (residuals) of erosivity in relation to key controlling factors—rainfall amount, elevation, land use classes, and soil types—using ANOVA. The detailed results are provided in Supplementary Table 9. The analysis revealed a clear seasonal pattern in the residuals. Specifically, rainfall significantly affected residuals during winter or dry periods ( $p < 0.05$ ), and similar higher rainfall erosivity variability under low rainfall conditions was observed in Europe<sup>84</sup>. Elevation and land use classes showed significant impacts during the winter and pre-monsoon months, while soil types influenced residuals primarily during the summer monsoon season. A similar topographic influence on rainfall erosivity has also been reported for the Tibetan Plateau<sup>131</sup>.

### Limitations and future scope

The study has several limitations due to the limited availability of high-resolution datasets in a tropical country like India. Variations in the start and end years of hourly and monthly rainfall across stations introduce inconsistencies in temporal coverage, which may affect the accuracy of erosivity estimates. Despite these challenges, we have attempted to incorporate long-term rainfall datasets to derive long-term average erosivity values. However, climate variability—such as shifts in rainfall patterns, frequency, and intensity—can significantly influence these results<sup>132</sup>. Additionally, recent studies have highlighted an increase in extreme rainfall events and changing monsoonal behavior in South Asia, driven by both natural variability and anthropogenic influences<sup>133</sup>. These changes may render historical erosivity equations outdated, especially those developed using short-term or localized datasets. Therefore, it is essential to rely on updated observed data and high-resolution gridded products to capture contemporary erosivity dynamics more accurately.

The reliance on WorldClim data to fill gaps in temperature records may introduce uncertainties due to resolution differences. While the error analysis shows that the median error is low, the error range is considerably larger, particularly in January and February. However, future studies could benefit from incorporating more accurate datasets, such as CHELSA<sup>127</sup>, or satellite-based and reanalysis products, or even combinations of these, which could help reduce interpolation errors and improve erosivity assessments.

Additionally, the conversion factors used to estimate 1-min erosivity from 60-min data are approximations that may not fully capture storm variability<sup>75,80,119</sup>. The lack of high temporal resolution datasets hinders the accurate estimation of maximum 30-min intensity, which can lead to under- or overestimation of rainfall erosivity values, particularly in regions with intense rainfall, such as the Western Ghats and Northeast India. We encourage researchers to derive region-specific correction factors using high-temporal resolution datasets to improve accuracy (e.g., Yue et al.<sup>130</sup>). Furthermore, future studies are encouraged to incorporate high-temporal resolution satellite and climate reanalysis datasets, along with limited gauge observations (e.g., Yonaba et al.<sup>30</sup>), to eliminate the need for such conversion factors in erosivity estimation.

Furthermore, the performance of the XGBoost machine learning model also depends on data availability, which could impact its generalizability<sup>134</sup>. Although we employed a two-way validation approach to improve the robustness of the model, XGBoost and similar machine learning models are still prone to overfitting and are highly sensitive to the structure of the input data. Future studies are recommended to explore advanced modeling techniques such as Deep Learning models<sup>135</sup> (e.g., Artificial Neural Networks, Long Short-Term Memory, and Convolutional Neural Networks), which are often more resilient to overfitting and capable of capturing complex spatio-temporal patterns.

While GWR captures spatial non-stationarity effectively, it is sensitive to data density, multicollinearity, and bandwidth selection<sup>134</sup>. The model may produce unstable estimates in sparsely gauged regions, and its results can be difficult to interpret due to spatially varying coefficients (e.g., high interpolation error for the winter months). Meusbürger et al.<sup>89</sup> also observed that in Switzerland, rainfall erosivity values were lowest during winter months, leading to increased uncertainty in spatial predictions. Future studies are encouraged to use recently available high temporal resolution radar remote sensing datasets (e.g., Dai et al.<sup>136</sup>, Auerwald et al.<sup>75</sup>) which can be applied on a broader scale and reduce the interpolation error.

Notably, the gap-filled SRTM dataset used in this study, obtained from the CGIAR-CSI, has certain limitations in accurately capturing ground elevation. As demonstrated in past studies<sup>137,138</sup>, although the CGIAR-CSI dataset offers improved vertical accuracy compared to standard SRTM data, the enhancements in slope and aspect measurements are significant only for slopes greater than 10°. For slopes less than 10°, however, the improvement is not as pronounced. Future studies are encouraged to use regionally available higher-resolution topographic information for event improvement in the DEM, such as ASTER-GDEM or CartoDEM, to achieve more accurate erosivity estimation.

The high-resolution mapping of rainfall erosivity in a traditional way requires more data, specifically at least one observation for each raster grid<sup>139</sup>. However, such detailed observations are likely unavailable for India. Even many countries in the Global South rarely have such high-temporal resolution and spatially dense datasets for rainfall erosivity and erosion estimation. We believe that limited observations of rainfall datasets, along with the help of machine learning and artificial intelligence, could significantly improve large-scale erosivity estimation. We strongly recommend that future studies address these limitations to improve accuracy. These limitations should be considered when interpreting the study's results, and future research could focus on overcoming these challenges.

### Conclusion

The analysis of monthly erosivity across India reveals a clear spatio-temporal pattern, with much of the country experiencing significantly high erosivity during the summer monsoon months (June to September). Notably, July sees the peak erosivity, primarily driven by total rainfall, with most of the rainfall occurring in large events classified as erosive. Interestingly, the number of erosive events exceeded 15 for many stations between June and

August, indicating that an erosive event occurred at least every second day. This pattern was particularly evident in July and August, with one-quarter of all stations experiencing such frequent erosive events.

In contrast, the remaining months (October to May) exhibit consistently lower erosivity values, except along the southern coast and Himalayan ranges, where erosive rainfall also occurs during the pre- and post-monsoon periods. The maximum 60-min rainfall intensity during the summer monsoon can reach up to  $50 \text{ mm h}^{-1}$ , while in the rest of the year, it typically remains below  $20 \text{ mm h}^{-1}$ .

The spatial patterns in rainfall erosivity during the summer monsoon months are clearly dominated by rainfall and elevation patterns. The SHAP analysis reveals that rainfall, surface runoff, and water vapor pressure are the primary drivers of monthly rainfall erosivity across India, with their influence varying seasonally. In contrast, factors like soil moisture, wind speed, slope, soil type, and land use play a minimal role in erosivity estimation at the national scale. Distance from the coast showed moderate importance, particularly during the pre-monsoon and winter seasons, due to its impact on regional moisture availability and atmospheric dynamics.

Long-term trends from 1969 to 2021 reveal a slight increasing trend in monthly erosivity for most months, except for a decreasing trend in August. However, only the trend for January ( $+0.86 \text{ MJ mm ha}^{-1} \text{ h}^{-1} \text{ month}^{-1}$  per year) was found to be statistically significant. Notably, among the 96 trend tests (12 months  $\times$  8 attributes), only 10 exhibited statistically significant trends. A statistically significant increase in the 60-min rainfall intensity was observed during the post-monsoon and winter months, with an annual increase rate of approximately  $+0.1 \text{ mm h}^{-1}$ .

Several sources of uncertainty in the current study warrant consideration—such as the reliance on WorldClim temperature data for stations lacking observations and the use of correction factors to estimate 1-min rainfall erosivity from hourly data. Although these approaches showed acceptable validation errors against observed and literature-based datasets, they may still affect estimation accuracy in months with high variability. Future research should prioritize the integration of high-temporal resolution satellite data, climate reanalysis products, and improved ground-based datasets such as CHELSA. This would help reduce the uncertainty and enhance the precision of erosivity estimates. These improvements will ultimately strengthen the basis for developing effective, region-specific soil conservation and erosion management strategies—particularly during the monsoon season when erosivity peaks.

## Data availability

The dataset set can be made available from the corresponding author on a reasonable request.

Received: 8 November 2024; Accepted: 14 July 2025

Published online: 31 July 2025

## References

- Ananda, J. & Herath, G. Soil erosion in developing countries: a socio-economic appraisal. *J. Environ. Manage.* **68**, 343–353 (2003).
- Borrelli, P. et al. Land use and climate change impacts on global soil erosion by water (2015–2070). *Proc. Natl. Acad. Sci. U. S. A.* **117**, 21994–22001 (2020).
- Wuepper, D., Borrelli, P. & Finger, R. Countries and the global rate of soil erosion. *Nat. Sustain.* **3**, 51–55 (2020).
- Sivakumar, M. V. K. & Stefanski, R. Climate change in South Asia. in *Climate change and food security in South Asia* 13–30 (Springer, 2010).
- IPCC. Land–climate interactions. In: *Climate Change and Land: an IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems*. in *Climate Change and Land* 131–248 (Intergovernmental Panel on Climate Change (IPCC), 2022).
- Das, S., Jain, M. K. & Gupta, V. An assessment of anticipated future changes in water erosion dynamics under climate and land use change scenarios in South Asia. *J. Hydrol.* **637**, 131341 (2024).
- Singh, J., Yadav, B. K., Schneidewind, U. & Krause, S. Microplastics pollution in inland aquatic ecosystems of India with a global perspective on sources, composition, and spatial distribution. *J. Hydrol. Reg. Stud.* **53**, 101798 (2024).
- Wischmeier, W. H. & Smith, D. D. *Predicting rainfall erosion losses: A guide to conservation planning*. (Department of Agriculture, Washington, D.C., USA, 1978).
- Renard, K. G. *Predicting soil erosion by water: A guide to conservation planning with the revised universal soil loss equation (RUSLE)*. (United States Government Printing, 1997).
- Foster, G. R. et al. User's Guide: Revised Universal Soil Loss Equation Version 2 (RUSLE2). *US Dep. Agric. Agric. Res. Serv.* **2**, 1–429 (2008).
- Bezak, N., Borrelli, P. & Panagos, P. Exploring the possible role of satellite-based rainfall data in estimating inter- and intra-annual global rainfall erosivity. *Hydrol. Earth Syst. Sci.* **26**, 1907–1924 (2022).
- Das, S. & Jain, M. K. Unravelling the future changes in rainfall erosivity over India under shared socio-economic pathways. *Catena* **232**, (2023).
- Fenta, A. A. et al. An integrated modeling approach for estimating monthly global rainfall erosivity. *Sci. Rep.* **14**, 8167 (2024).
- Fenta, A. A. et al. Improving satellite-based global rainfall erosivity estimates through merging with gauge data. *J. Hydrol.* **620**, 129555 (2023).
- Mello, C. R., Viola, M. R., Beskow, S. & Norton, L. D. Multivariate models for annual rainfall erosivity in Brazil. *Geoderma* **202–203**, 88–102 (2013).
- Wischmeier, W. H. Cropping-management factor evaluations for a universal soil-loss equation. *Soil Sci. Soc. Am. J.* **24**, 322–326 (1960).
- Dash, C. J., Das, N. K. & Adhikary, P. P. Rainfall erosivity and erosivity density in Eastern Ghats Highland of east India. *Nat. Hazards* **97**, 727–746 (2019).
- Das, S., Jain, M. K. & Gupta, V. A step towards mapping rainfall erosivity for India using high-resolution GPM satellite rainfall products. *CATENA* **212**, 106067 (2022).
- Nearing, M. A., Yin, S. qing, Borrelli, P. & Polyakov, V. O. Rainfall erosivity: An historical review. *Catena* **157**, 357–362 (2017).
- Panagos, P. et al. Rainfall erosivity in Europe. *Sci. Total Environ.* **511**, 801–814 (2015).
- Wischmeier, W. H. A rainfall erosion index for a universal soil-loss equation. *Soil Sci. Soc. Am. J.* **23**, 246–249 (1959).
- Yin, S., Nearing, M. A., Borrelli, P. & Xue, X. Rainfall Erosivity: An overview of methodologies and applications. *Vadose Zone J.* **16**, 1–16 (2017).

23. Benavidez, R., Jackson, B., Maxwell, D. & Norton, K. 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**, 6059–6086 (2018).
24. Chen, W., Huang, Y. C., Lebar, K. & Bezak, N. A systematic review of the incorrect use of an empirical equation for the estimation of the rainfall erosivity around the globe. *Earth-Sci. Rev.* **238**, 104339 (2023).
25. Ghosal, K. & Bhattacharya, S. D. A Review of RUSLE Model. *J. Indian Soc. Remote Sens.* **48**, 689–707 (2020).
26. Majhi, A., Shaw, R., Mallick, K. & Patel, P. P. Towards improved USLE-based soil erosion modelling in India: A review of prevalent pitfalls and implementation of exemplar methods. *Earth-Sci. Rev.* **221**, 103786 (2021).
27. Das, T. & Sarma, A. Kumar. A Step towards incorporating return period in rainfall erosivity of India using high temporal resolution Satellite precipitation product. (2023).
28. Raj, R., Saharia, M., Chakma, S. & Rafieinasab, A. Mapping rainfall erosivity over India using multiple precipitation datasets. *CATENA* **214**, 106256 (2022).
29. Kim, J., Han, H., Kim, B., Chen, H. & Lee, J. H. Use of a high-resolution-satellite-based precipitation product in mapping continental-scale rainfall erosivity: A case study of the United States. *CATENA* **193**, 104602 (2020).
30. Yonaba, R. et al. Exploring the added value of sub-daily bias correction of high-resolution gridded rainfall datasets for rainfall erosivity estimation. *Hydrology* **11**, 132 (2024).
31. Chen, Y., Xu, M., Wang, Z., Gao, P. & Lai, C. Applicability of two satellite-based precipitation products for assessing rainfall erosivity in China. *Sci. Total Environ.* **757**, 143975 (2021).
32. Das, S. et al. GloRESatE: A dataset for global rainfall erosivity derived from multi-source data. *Sci. Data* **11**, 926 (2024).
33. Arnoldus, H. M. J. Assessing soil degradation: Methodology used to determine the maximum average soil loss due to sheet and rill erosion in Morocco. In: Assessing soil degradation. *FAO Soil Bull.* **34**, 8–9 (1977).
34. Babu, R., Tejwani, K. G., Agarwal, M. C. & Bhushan, L. S. Distribution of erosion index and iso-erodent map of India. *Indian J. Soil Conserv.* (1978).
35. Babu, R., Dhyani, B. L. & Kumar, N. Assessment of erodibility status and refined Iso-Eroderent Map of India. *Indian J. Soil Conserv.* **32**, 171–177 (2004).
36. Renard, K. G. & Freimund, J. R. Using monthly precipitation data to estimate the R-factor in the revised USLE. *J. Hydrol.* **157**, 287–306 (1994).
37. Klik, A., Haas, K., Dvorackova, A. & Fuller, I. C. Spatial and temporal distribution of rainfall erosivity in New Zealand. *Soil Res.* **53**, 815–825 (2015).
38. Nakil, M. & Khire, M. Effect of slope steepness parameter computations on soil loss estimation: review of methods using GIS. *Geocarto Int.* **31**, 1078–1093 (2016).
39. Schmidt, S., Alewell, C., Panagos, P. & Meusburger, K. Regionalization of monthly rainfall erosivity patterns in Switzerland. *Hydrol. Earth Syst. Sci.* **20**, 4359–4373 (2016).
40. Dash, C. J. et al. Comparison of rainfall kinetic energy–intensity relationships for Eastern Ghats Highland region of India. *Nat. Hazards* **93**, 547–558 (2018).
41. Ma, X. & Zheng, M. Statistical evaluation of proxies for estimating the rainfall erosivity factor. *Sci. Rep.* **12**, 12092 (2022).
42. Arnoldus, H. M. J. An approximation of the rainfall factor in the Universal Soil Loss Equation. In De Boedt, M. and Gabriels, D., Eds., *Assessment of Erosion. John Wiley Sons N. Y.* 127–132 (1980).
43. Auerswald, K. & Fiener, P. Assessing the impact of climate change on soil erosion by water. In *Burleigh Dodds Series in Agricultural Science* 51–76 (Burleigh Dodds Science Publishing, 2024). <https://doi.org/10.19103/AS.2023.0131.05>.
44. Almagro, A., Oliveira, P. T. S., Nearing, M. A. & Hagemann, S. Projected climate change impacts in rainfall erosivity over Brazil. *Sci. Rep.* **7**, 1–12 (2017).
45. Maurya, N. K. & Tanwar, P. S. Estimation of temporal R-factor based on monthly precipitation data. *J. Phys. Conf. Ser.* **2070**, 012210 (2021).
46. Tiwari, H., Rai, S. P., Kumar, D. & Sharma, N. Rainfall erosivity factor for India using modified fourier index. *J. Appl. Water Eng. Res.* **4**, 83–91 (2016).
47. Chen, Y., Wei, T., Li, J., Xin, Y. & Ding, M. Future changes in global rainfall erosivity: Insights from the precipitation changes. *J. Hydrol.* **638**, 131435 (2024).
48. Li, J., Xiong, M., Sun, R. & Chen, L. Temporal variability of global potential water erosion based on an improved USLE model. *Int. Soil Water Conserv. Res.* <https://doi.org/10.1016/j.iswcr.2023.03.005> (2023).
49. Luo, X. et al. Increased precipitation weakens the positive effect of vegetation greening on erosion. *Geocarto Int.* **38** (1), (2023).
50. Naipal, V., Reick, C., Pongratz, J. & Oost, K. V. Improving the global applicability of the RUSLE model - Adjustment of the topographical and rainfall erosivity factors. *Geosci. Model Dev.* **8**, 2893–2913 (2015).
51. Yang, D., Kanae, S., Oki, T., Koike, T. & Musiak, K. Global potential soil erosion with reference to land use and climate changes. *Hydrol. Process.* **17**, 2913–2928 (2003).
52. Ali, S. & Sharda, V. N. Evaluation of the universal soil loss equation (USLE) in semi-arid and sub-humid climates of India. *Appl. Eng. Agric.* **21**, 217–225 (2005).
53. Singh, G. Soil loss prediction research in India. *Bull. Cent. Soil Water Conserv. Res. Train. Inst.* **70** (1981).
54. Dash, Ch. J., Shrimali, S. S., Madhu, M., Kumar, R. & Adhikary, P. P. Unveiling rainfall and erosivity dynamics in Odisha's varied agro-climatic zones for sustainable soil and water conservation planning. *Theor. Appl. Climatol.* **155**, 7557–7574 (2024).
55. Dash, S. S. & Maity, R. Effect of climate change on soil erosion indicates a dominance of rainfall over LULC changes. *J. Hydrol. Reg. Stud.* **47**, 101373 (2023).
56. Ganasri, B. P. & Ramesh, H. Assessment of soil erosion by RUSLE model using remote sensing and GIS - A case study of Nethravathi Basin. *Geosci. Front.* **7**, 953–961 (2016).
57. Gupta, A., Sawant, C. P., Kumar, M., Singh, R. K. & Rao, K. V. R. Assessment of rainfall erosivity for Bundelkhand region of central India using long-term rainfall data. *Mausam* **75**, 415–432 (2024).
58. Mondal, A., Khare, D. & Kundu, S. Change in rainfall erosivity in the past and future due to climate change in the central part of India. *Int. Soil Water Conserv. Res.* **4**, 186–194 (2016).
59. Sarkar, B. et al. Soil erosion and sediment yield estimation in a tropical monsoon dominated river basin using GIS-based models. *Geocarto Int.* **39**, 2309181 (2024).
60. El-Swaify, S. A., Gramier, C. L. & Lo, A. Recent advances in soil conservation in steep land in humid tropics. In *Proceedings of the international conference on steep land agriculture in the humid tropics. Kuala Lumpur, MARDI* 87–100 (1987).
61. Roose, E. J. Use of the universal soil loss equation to predict erosion in West Africa. In: soil erosion: prediction and control. Soil conservation society of America, Ankeny, Iowa. (1976).
62. Pandey, A., Chowdary, V. M. & Mal, B. C. Identification of critical erosion prone areas in the small agricultural watershed using USLE, GIS and remote sensing. *Water Resour. Manag.* **21**, 729–746 (2007).
63. Singh, G. & Panda, R. K. Grid-cell based assessment of soil erosion potential for identification of critical erosion prone areas using USLE, GIS and remote sensing: A case study in the Kapgari watershed. *India. Int. Soil Water Conserv. Res.* **5**, 202–211 (2017).
64. Kothiyari, U. C. & Singh, V. P. Rainfall and temperature trends in India. *Hydrol. Process.* **10**, 357–372 (1996).
65. Kuttippurath, J. et al. Observed rainfall changes in the past century (1901–2019) over the wettest place on Earth. *Environ. Res. Lett.* **16**, 024018 (2021).
66. Roy, P. D. & Singhvi, A. K. Climate variation in the Thar desert since the last glacial maximum and evaluation of the Indian monsoon. *TIP* **19**, 32–44 (2016).



67. Abish, B. & Arun, K. Resolving the weakening of orographic rainfall over India using a regional climate model RegCM 4.5. *Atmospheric Res.* **227**, 125–139 (2019).
68. FAO. India at a glance | FAO in India | Food and Agriculture Organization of the United Nations. <https://www.fao.org/india/fao-in-india/india-at-a-glance/en/>.
69. Fick, S. E. & Hijmans, R. J. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* **37**, 4302–4315 (2017).
70. Muñoz-Sabater, J. et al. ERA5-Land: A state-of-the-art global reanalysis dataset for land applications. *Earth Syst. Sci. Data* **13**, 4349–4383 (2021).
71. Baxter, S. World reference base for soil resources. World soil resources report 103. Rome: Food and Agriculture Organization of the United Nations (2006), pp. 132, US\$22.00 (paperback). ISBN 92-5-10511-4. *Exp. Agric.* **43**, 264–264 (2007).
72. Conrad, O. et al. System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. *Geosci. Model Dev.* **8**, 1991–2007 (2015).
73. Chen, C. J., Senarath, S. U. S., Dima-West, I. M. & Marcella, M. P. Evaluation and restructuring of gridded precipitation data over the Greater Mekong Subregion. *Int. J. Climatol.* **37**, 180–196 (2017).
74. Zuo, J. et al. GCL\_FCS30: a global coastline dataset with 30-m resolution and a fine classification system from 2010 to 2020. *Sci. Data* **12**, 129 (2025).
75. Auerswald, K., Fischer, F. K., Winterrath, T. & Brandhuber, R. Rain erosivity map for Germany derived from contiguous radar rain data. *Hydrol. Earth Syst. Sci.* **23**, 1819–1832 (2019).
76. Fischer, F. K., Winterrath, T. & Auerswald, K. Rain erosivity map for Germany derived from contiguous radar rain data. Preprint at <https://doi.org/10.5194/hess-2018-504> (2018).
77. Chang, Y., Lei, H., Zhou, F. & Yang, D. Spatial and temporal variations of rainfall erosivity in the middle Yellow River Basin based on hourly rainfall data. *CATENA* **216**, 106406 (2022).
78. Yue, T., Yin, S., Xie, Y., Yu, B. & Liu, B. Rainfall erosivity mapping over mainland China based on high-density hourly rainfall records. *Earth Syst. Sci. Data* **14**, 665–682 (2022).
79. Foster, G. P. Science documentation: Revised universal soil loss equation, Version 2 (RUSLE 2). USDA-Agric. Res. Serv. Wash. DC (2005).
80. Fischer, F. K., Winterrath, T. & Auerswald, K. Temporal- and spatial-scale and positional effects on rain erosivity derived from point-scale and contiguous rain data. *Hydrol. Earth Syst. Sci.* **22**, 6505–6518 (2018).
81. Yin, S., Xie, Y., Liu, B. & Nearing, M. A. Rainfall erosivity estimation based on rainfall data collected over a range of temporal resolutions. *Hydrol. Earth Syst. Sci.* **19**, 4113–4126 (2015).
82. Dabney, S. M., Yoder, D. C. & Vieira, D. A. N. The application of the Revised Universal Soil Loss Equation, Version 2, to evaluate the impacts of alternative climate change scenarios on runoff and sediment yield. *J. Soil Water Conserv.* **67**, 343–353 (2012).
83. Panagos, P., Ballabio, C., Borrelli, P. & Meusburger, K. Spatio-temporal analysis of rainfall erosivity and erosivity density in Greece. *CATENA* <https://doi.org/10.1016/j.catena.2015.09.015> (2016).
84. Ballabio, C. et al. Mapping monthly rainfall erosivity in Europe. *Sci. Total Environ.* **579**, 1298–1315 (2017).
85. Chen, T. et al. xgboost: Extreme Gradient Boosting. 1.7.8.1 <https://doi.org/10.32614/CRAN.package.xgboost> (2014).
86. Chen, T. & Guestrin, C. XGBoost: A scalable tree boosting system. in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* vols 13–17-Aug 785–794 (2016).
87. Dietterich, T. G. Ensemble Methods in Machine Learning. in *Multiple Classifier Systems* vol. 1857 1–15 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2000).
88. Mello, C. R. D., Viola, M. R., Owens, P. R., Mello, J. M. D. & Beskow, S. Interpolation methods for improving the RUSLE R-factor mapping in Brazil. *J. Soil Water Conserv.* **70**, 182–197 (2015).
89. Meusburger, K., Steel, A., Panagos, P., Montanarella, L. & Alewell, C. Spatial and temporal variability of rainfall erosivity factor for Switzerland. *Hydrol. Earth Syst. Sci.* **16**, 167–177 (2012).
90. Panagos, P. et al. Global rainfall erosivity database (GloREDa) and monthly R-factor data at 1 km spatial resolution. *Data Brief* **50**, 109482 (2023).
91. Azari, M., Oliaye, A. & Nearing, M. A. Expected climate change impacts on rainfall erosivity over Iran based on CMIP5 climate models. *J. Hydrol.* **593**, 125826 (2021).
92. Lee, J. et al. Evaluation of rainfall erosivity factor estimation using machine and deep learning models. *Water* **13**, 382 (2021).
93. Lee, S. et al. Estimation of rainfall erosivity factor in Italy and Switzerland using Bayesian optimization based machine learning models. *CATENA* **211**, 105957 (2022).
94. Mangukiya, N. K. & Sharma, A. Alternate pathway for regional flood frequency analysis in data-sparse region. *J. Hydrol.* **629**, 130635 (2024).
95. Nath, K., Nayak, P. C. & Kasiviswanathan, K. S. Soil volumetric water content prediction using unique hybrid deep learning algorithm. *Neural Comput. Appl.* **36**, 16503–16525 (2024).
96. Morita, Y. et al. Applying Bayesian optimization with Gaussian process regression to computational fluid dynamics problems. *J. Comput. Phys.* **449**, 110788 (2022).
97. Mangukiya, N. K. & Sharma, A. Flood risk mapping for the lower Narmada basin in India: A machine learning and IoT-based framework. *Nat. Hazards* **113**, 1285–1304 (2022).
98. Lee, S., Nguyen, N., Karamanli, A., Lee, J. & Vo, T. P. Super learner machine-learning algorithms for compressive strength prediction of high performance concrete. *Struct. Concr.* **24**, 2208–2228 (2023).
99. Dion, P., Martel, J.-L. & Arsenault, R. Hydrological ensemble forecasting using a multi-model framework. *J. Hydrol.* **600**, 126537 (2021).
100. Li, M., Yang, X., Wang, Y., Wang, Y. & Zhu, J. The use of the GWPCA-MGWR Model for studying spatial relationships between environmental variables and longline catches of Yellowfin Tunas. *J. Mar. Sci. Eng.* **12**, 1002 (2024).
101. Harris, P., Brunson, C. & Charlton, M. Geographically weighted principal components analysis. *Int. J. Geogr. Inf. Sci.* **25**, 1717–1736 (2011).
102. Harris, P., Clarke, A., Juggins, S., Brunson, C. & Charlton, M. Enhancements to a geographically weighted principal component analysis in the context of an application to an environmental data set. *Geogr. Anal.* **47**, 146–172 (2015).
103. Lloyd, C. D. Assessing the effect of integrating elevation data into the estimation of monthly precipitation in Great Britain. *J. Hydrol.* **308**, 128–150 (2005).
104. Goovaerts, P. Using elevation to aid the geostatistical mapping of rainfall erosivity. *CATENA* **34**, 227–242 (1999).
105. Zhang, Y., Hanati, G., Danierhan, S. & Hu, K. Application and assessment of a downscaled GPM dataset in the simulation of snowmelt runoff in alpine mountainous areas. *J. Hydrol. Reg. Stud.* **41**, 101107 (2022).
106. Tian, M. et al. Geographically weighted regression (GWR) and Prediction-area (P-A) plot to generate enhanced geochemical signatures for mineral exploration targeting. *Appl. Geochem.* **150**, 105590 (2023).
107. Yu, D. Spatial interpolation via GWR, a plausible alternative? in *2009 17th International Conference on Geoinformatics* 1–5 (IEEE, Fairfax, VA, 2009). <https://doi.org/10.1109/GEOINFORMATICS.2009.5293526>.
108. Huang, B., Wu, B. & Barry, M. Geographically and temporally weighted regression for modeling spatio-temporal variation in house prices. *Int. J. Geogr. Inf. Sci.* **24**, 383–401 (2010).
109. Wang, M. et al. Comparison of Spatial Interpolation and Regression Analysis Models for an Estimation of Monthly Near Surface Air Temperature in China. *Remote Sens.* **9**, 1278 (2017).



110. Mounirou, L. A. et al. Soil Erosion across Scales: Assessing Its Sources of Variation in Sahelian Landscapes under Semi-Arid Climate. *Land* **11**, 2302 (2022).
111. Chen, Y., Duan, X., Zhang, G., Ding, M. & Lu, S. Rainfall erosivity estimation over the Tibetan plateau based on high spatial-temporal resolution rainfall records. *Int. Soil Water Conserv. Res.* **10**, 422–432 (2022).
112. Nohara, Y., Matsumoto, K., Soejima, H. & Nakashima, N. Explanation of machine learning models using shapley additive explanation and application for real data in hospital. *Comput. Methods Programs Biomed.* **214**, 106584 (2022).
113. Mann, H. B. Nonparametric Tests Against Trend. *Econometrica* **13**, 245 (1945).
114. Praveen, B. et al. Analyzing trend and forecasting of rainfall changes in India using non-parametrical and machine learning approaches. *Sci. Rep.* **10**, 1–21 (2020).
115. Hamed, K. H. & Ramachandra Rao, A. A modified Mann-Kendall trend test for autocorrelated data. *J. Hydrol.* **204**, 182–196 (1998).
116. Sen, P. K. Estimates of the Regression Coefficient Based on Kendall's Tau. *J. Am. Stat. Assoc.* **63**, 1379–1389 (1968).
117. Moriasi, D. N. et al. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Trans. ASABE* **50**, 885–900 (2007).
118. Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. *Numerical Recipes 3rd Edition: The Art of Scientific Computing*. (Cambridge University Press, 2007).
119. McGehee, R. P. et al. An updated isocrotonal map of the conterminous United States. *Int. Soil Water Conserv. Res.* **10**, 1–16 (2022).
120. Delgado, D., Sadaoui, M., Ludwig, W. & Méndez, W. Spatio-temporal assessment of rainfall erosivity in Ecuador based on RUSLE using satellite-based high frequency GPM-IMERG precipitation data. *CATENA* **219**, 106597 (2022).
121. Johannsen, L. L. et al. An update of the spatial and temporal variability of rainfall erosivity (R-factor) for the main agricultural production zones of Austria. *CATENA* **215**, 106305 (2022).
122. Takhellambam, B. S. et al. Projected mid-century rainfall erosivity under climate change over the southeastern United States. *Sci. Total Environ.* **865**, 161119 (2023).
123. Varghese, S. J., Pentakota, S., Thadivalasa, P., Podapati, G. & Ashok, K. Changes in physical characteristics of extreme rainfall events during the Indian summer monsoon based on downscaled and bias-corrected CMIP6 models. *Sci. Rep.* **15**, 3679 (2025).
124. Pandey, R., Mehta, D., Kumar, V. & Prakash Pradhan, R. Quantifying soil erosion and soil organic carbon conservation services in indian forests: A RUSLE-SDR and GIS-based assessment. *Ecol. Indic.* **163**, 112086 (2024).
125. Pal, S. C. et al. Changing climate and land use of 21st century influences soil erosion in India. *Gondwana Res.* **94**, 164–185 (2021).
126. Marchi, M., Sinjur, I., Bozzano, M. & Westergren, M. Evaluating worldclim version 1 (1961–1990) as the baseline for sustainable use of forest and environmental resources in a changing climate. *Sustainability* **11**, 3043 (2019).
127. Karger, D. N., Wilson, A. M., Mahony, C., Zimmermann, N. E. & Jetz, W. Global daily 1 km land surface precipitation based on cloud cover-informed downscaling. *Sci. Data* **8**, 307 (2021).
128. Bobrowski, M., Weidinger, J. & Schickhoff, U. Is new always better? Frontiers in global climate datasets for modeling treeline species in the Himalayas. *Atmosphere* **12**, 543 (2021).
129. Panagos, P. et al. Monthly rainfall erosivity: Conversion factors for different time resolutions and regional assessments. *Water Switz.* **8**, 119 (2016).
130. Yue, T. et al. Effect of time resolution of rainfall measurements on the erosivity factor in the USLE in China. *Int. Soil Water Conserv. Res.* **8**, 373–382 (2020).
131. Cui, B. et al. Spatiotemporal Variation in Rainfall Erosivity and Correlation with the ENSO on the Tibetan Plateau since 1971. *Int. J. Environ. Res. Public Health* **18**, 11054 (2021).
132. Roxy, M. K. et al. A threefold rise in widespread extreme rain events over central India. *Nat. Commun.* **8**, 1–11 (2017).
133. Turner, A. G. & Annamalai, H. Climate change and the South Asian summer monsoon. *Nat. Clim. Change* **2**, 587–595 (2012).
134. Rao, P. et al. A comparison of multiple methods for mapping groundwater levels in the Mu Us Sandy Land. *China. J. Hydrol. Reg. Stud.* **43**, 101189 (2022).
135. Mangukiy, N. K. & Sharma, A. Deep Learning-Based Approach for Enhancing Streamflow Prediction in Watersheds With Aggregated and Intermittent Observations. *Water Resour. Res.* **61**, e2024WR037331 (2025).
136. Dai, Q. et al. Radar remote sensing reveals potential underestimation of rainfall erosivity at the global scale. *Sci. Adv.* **9**, eadg5551 (2023).
137. Rexer, M. & Hirt, C. Comparison of free high resolution digital elevation data sets (ASTER GDEM2, SRTM v2.1/v4.1) and validation against accurate heights from the Australian National Gravity Database. *Aust. J. Earth Sci.* **0**, 1–15 (2014).
138. Gorokhovich, Y. & Voustianiouk, A. Accuracy assessment of the processed SRTM-based elevation data by CGIAR using field data from USA and Thailand and its relation to the terrain characteristics. *Remote Sens. Environ.* **104**, 409–415 (2006).
139. Meng, J. Raster data projection transformation based-on Kriging interpolation approximate grid algorithm. *Alex. Eng. J.* **60**, 2013–2019 (2021).

## Acknowledgements

We sincerely acknowledge the India Meteorological Department (IMD), WorldClim, and the Consultative Group on International Agricultural Research—Consortium for Spatial Information (CGIAR-CSI) community for generously providing valuable open-access datasets.

## Author contributions

S.D., M.K.J., and K.A. conceived and designed the analysis. S.D. and K.A. conducted the analysis and drafted the initial manuscript. C.R.M. and P.M. reviewed and contributed to the writing of the manuscript. Finally, all authors reviewed and revised the manuscript.

## Funding

None.

## Declarations

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-11992-x>.

**Correspondence** and requests for materials should be addressed to S.D.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

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

**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/>.

© The Author(s) 2025