https://doi.org/10.5194/essd-2025-306
Preprint. Discussion started: 5 August 2025

Open Access Earth System Science Data Discussions

The first rainfall erosivity database in Mexico: facing challenges of leveraging legacy climate data

Viviana Marcela Varón-Ramírez[1,2], Douglas A. Gómez-Latorre[2], Carlos Eduardo Arroyo-Cruz[1], Alberto Gómez-Tagle[3], Blanca Lucía Prado Pano[4], Ronald R. Gutierrez Llantoy[5], Deyanira Lobo-Luján[6], and Mario Guevara[1]

[1]Instituto de Geociencias - Universidad Nacional Autónoma de México - UNAM. Campus Juriquilla, Qro. Mexico.
[2]Corporación Colombiana de Investigación Agropecuaria - AGROSAVIA, Centro de Investigación Tibaitatá, Mosquera-Cundinamarca, Colombia
[3]Instituto de Investigaciones sobre los Recursos Naturales - INIRENA, Universidad Michoacana de San Nicolás de Hidalgo. Morelia, Mexico
[4]Instituto de Geología - Universidad Nacional Autónoma de México - UNAM. Mexico City. Mexico.
[5]Departamento de Ingeniería (GERDIS, GEOSED), Pontificia Universidad Católica del Perú. Lima, Peru
[6]Instituto de Edafología - Universidad Central de Venezuela. Maracay, Aragua. Venezuela

Correspondence: Viviana Marcela Varón-Ramírez (viviana.varon@geociencias.unam.mx) and Mario Guevara (mguevara@geociencias.unam.mx)

Abstract. Soil water erosion (SWE) is the dominant soil degradation driver on a global scale. For quantifying SWE, erosivity is an index that reflects the potential (i.e., the energy) of rainfall to cause SWE. To support large-scale SWE studies and the assessment of the SWE process at the national scale in Mexico, the objectives of this research are a) to develop the first Mexican rainfall time series database for three climate normals CNs (1968-1997, 1978-2007, and 1988-2017) leveraging legacy climate data, and b) to estimate rainfall erosivity across continental Mexico by using daily rainfall time series. The workflow has three methodological moments: 1) development of the daily rainfall time series database, 2) identification of the best empirical relationship to estimate daily rainfall erosivity, and 3) estimation of the rainfall erosivity across Mexican territory. We compiled and harmonized 5410 rainfall time series (RTS) well distributed across the Mexican territory. We perform quality control and assurance, homogeneity analysis (using the normal homogeneity test), and the data gap-filling process (using the proportion method). Then, we tested three combinations of the α and β coefficients, proposed by three authors, in a power model to estimate rainfall erosivity; in this step, we used three validation databases (global, national, and local scales). Finally, we estimated the annual rainfall erosivity for all three CNs with multiple combinations of α and β coefficients. As principal results, the new database includes 1370, 1678, and 1676 RTS for each CN and its corresponding rainfall erosivity. The best parameter combination is the one proposed by Richardson et al. (1983) for all three validation databases. For the global and national databases, we observe a positive bias (Mean error of 956 and 324 MJ mm ha^{-1} h^{-1} yr^{-1}, respectively); in contrast, for the local database, results show a negative and higher bias (Mean error of -3699 MJ mm ha^{-1} h^{-1} yr^{-1}). About the erosivity estimation across the Mexican territory, the median values for rainfall erosivity for the three CNs were 3245, 3070, and 3327 MJ mm ha^{-1} h^{-1} yr^{-1}, respectively. The statistical distribution of the erosivity values was right-skewed for the three CNs, with high erosivity values reaching >12000 MJ mm ha^{-1} h^{-1} yr^{-1} in all three CNs. The behavior throughout the

https://doi.org/10.5194/essd-2025-306
Preprint. Discussion started: 5 August 2025

Open Access Earth System Science Data Discussions

year of the rainfall erosivity was similar for the three CNs. However, September had the highest contribution to the rainfall erosivity. The new database provides daily climatological data and analysis across Mexican territory through a multi-year period (1968 to 2017). Rainfall erosivity results support the study of SWE at the national scale by identifying areas with higher susceptibility to soil loss due to rainfall action and providing a more spatially dense and well-documented rainfall erosivity database. Following the FAIR principles (Findability, Availability, Interoperability, and Reproducibility) for scientific data, this database is available from a scholarly accepted repository https://doi.org/10.6073/pasta/e0dc8bd3501f8c19bb750e853c3289cb (Varón-Ramírez et al., 2025) for public consultation.

Keywords: legacy climate data, gap filling of climate series, daily rainfall erosivity, climatol package.

1 Introduction

Soil water erosion (SWE) refers to the soil displacement from its original location due to water action, such as rainfall, overland flow, and irrigation (Nearing, 2013). SWE represents the dominant soil degradation issue at the global scale because it affects nearly 33% of the World's surface (Pennock, 2019). The impact of SWE is not just on-site but also has off-site effects on distant locations. On-site, soil loses its natural fertility and capacity to store water, nutrients, and organic carbon (Hatfield et al., 2017), affecting food security. Off-site, the eroded soil triggers environmental issues such as water pollution, dam siltation, eutrophication of water bodies, contamination of coastal and marine ecosystems, and overall environmental damage (Feng et al., 2023). These widespread impacts underscore the urgent need to study SWE at national scales to guide effective land and water management.

When other erosion factors (e.g., erodibility, soil coverage and management, and topography) are constant, regions with frequent rainfall experience more soil loss than areas with limited rainfall (Ke and Zhang, 2021). Rainfall erosivity is the potential of rainfall to cause SWE (Nearing et al., 2017). Furthermore, rainfall erosivity is the first factor influencing SWE and is crucial for land and water conservation planning. Rainfall erosivity could increase its harmful effects on Mexican soils because, as extreme rainfall events are expected to increase in tropical zones due to climate change, SWE will likely increase as well (Borrelli et al., 2020). Furthermore, in Mexico, the temporal distribution of rainfall has become more extreme, with more extended periods of drought and increasingly extreme rainfall events (Porrúa et al., 2020). Thus, understanding rainfall erosivity patterns across Mexico and analysing them over a multi-year period is essential for enhancing soil sustainability and informed decision-making in soil conservation.

Rainfall erosivity—often represented as the R factor— quantifies the potential of rainfall to cause SWE (Nearing et al., 2017). The R factor captures the combined effect of raindrop impact and water flow. Wischmeier and Smith (1958) determined the rainfall erosivity of a storm as energy · times · intensity (EI). The E term refers to the total storm energy, and the I term refers to the maximum 30-min intensity $I30$; so, it is usually to find the notation $EI30$. Also, the R factor is presented as the mean annual rainfall erosivity over a multi-year period, and the yearly rainfall erosivity (R_y) is the sum of the rainfall erosivities for the total storms in a year. However, calculating $EI30$ for each storm requires high temporal resolution rainfall data, often at a minute temporal resolution, which can be challenging to obtain, especially at the national scale. A common

https://doi.org/10.5194/essd-2025-306
Preprint. Discussion started: 5 August 2025

solution to estimate the R factor in a data scarcity framework was constructing empirical relationships between erosivity from limited finer-resolution and coarse-resolution rainfall data (e.g., daily, monthly, and annual). However, studies also showed that higher-resolution data gave more accurate erosivity estimates for a specific time period (Yin et al., 2015). For this reason, the finer the time interval of the rainfall time series (RTS) used to calculate the R factor, the closer the estimations are to the $EI30$ values. Therefore, rainfall erosivity calculations relied on rainfall data with limited temporal resolution to help approximate the $EI30$ values.

Mexico does not have a public database of high-temporal resolution (sub-hourly) RTS; however, there is an opportunity to understand the national rainfall erosivity patterns using Mexico's finest available legacy climate data. Due to the lack of sub-hourly RTS across Mexico, the SWE studies have calculated the R factor from coarse-resolution RTS (monthly or annual) for some specific regions (Benites et al., 2020; González et al., 2016); however, there is an imperative need to address the rainfall effects at a national scale in Mexico (Varón-Ramírez and Guevara, 2024). On the other hand, the National Meteorological Service (SMN, by its Spanish acronym) has a rainfall database for public consultation (from 1900 to 2017) at a daily resolution of about 5454 weather stations across Mexican territory. Empirical relationships using daily rainfall have been developed to approximate the rainfall erosivity $EI30$. Richardson et al. (1983) proposed a power law model by assuming that a daily rainfall could be interpreted as a single storm; this model has been used in many countries and has been demonstrated to be helpful as an indicator of rainfall erosivity patterns (Rutebuka et al., 2020). Therefore, leveraging high-quality daily rainfall records is essential to reliably estimate erosivity and support national-scale soil erosion assessments in Mexico.

As in many countries around the world, there are some issues in the available Mexican rainfall time series, such as missing values, short measurement periods, and series inhomogeneity (breaks due to station relocation and measurement mistakes), which further compound the challenge of using climate data (McKinnon, 2022). Thus, climatology studies, such as erosivity, need a complete and reliable rainfall time series database (Yozgatligil et al., 2013). Therefore, a whole scheme of quality assurance, gap-filling, and homogenization process of the rainfall time series is needed (WMO, 2023). This quality control and homogenization process has been widely applied before rainfall erosivity analysis, in order to avoid incoherent rainfall amounts that could affect the long-term rainfall erosivity estimations (Rutebuka et al., 2020). Consequently, with a reliable rainfall database, it is possible to represent the actual rainfall characteristics in a particular region and allow soil erosion monitoring at the local and national scales in Mexico.

Developing a rainfall time series and, consequently, a rainfall erosivity database is challenging in both spatial and temporal terms. The large diversity of topographic conditions (i.e., two principal mountain ranges and a large latitudinal extent) and proximity to large water bodies from the Pacific Ocean and the Gulf of Mexico make Mexico a contrasting scenario of rainfall patterns (Carrera et al., 2024) and its hydrological-related processes(e.g., rainfall erosivity). Accurate benchmarks for understanding typical climate conditions and characterizing climate trends require a rainfall database long enough to represent its corresponding climate normal (CN), i.e., a statistical product computed over 30 years of rainfall time series (WMO, 2017). The CNs are widely used to compare recent observations, create anomaly-based datasets, and provide context for future climate projections. Considering local patterns across different CNs, these characteristics will contribute to an unprecedented rainfall time series dataset to estimate rainfall erosivity in Mexico.

https://doi.org/10.5194/essd-2025-306
Preprint. Discussion started: 5 August 2025

The lack of reliable and complete daily climate databases has not allowed the study of the impact of precipitation on the SWE process at a national scale in Mexico. Hence, the main objectives of this research are 1) to develop the first Mexican rainfall-series database for three climate normals (1968-1997, 1978-2007, and 1988-2017), leveraging on the availability of legacy climate data, 2) to estimate rainfall erosivity across continental Mexico by using daily rainfall time series. The following section describes the three methodological moments of this research: development of the daily rainfall time series database, identification of the best empirical relationship to estimate daily rainfall erosivity, and estimation of the rainfall erosivity across Mexican territory. Then, the results are presented accordingly. We present a discussion section with a critical analysis of results against recent scientific literature. We finally present our conclusions section, and the availability of all the codes and resulting databases of this research. The new knowledge allows a better understanding and prediction of rainfall distribution and its associated processes in different regions of Mexico.

2 Methodology

The study area corresponds to the conterminous Mexico (1,948,170 km^2). The country is located between latitudes 14°W and 32°N and longitudes 86°W and 118°W. Because of its geographical location, the region exhibits complex topographic and climate features (de Anda Sánchez, 2020). Mexico has been clustered into seven first-level ecoregions—which represent geographical units with characteristic biodiversity (Commission for Environmental Cooperation, 1997)—, namely: Mediterranean California, North American Deserts, Semi-arid Elevations, Great Plains, Tropical Rain Forest, Tropical Dry Forest, and Temperate Sierras. Each ecoregion occupies 1.3, 28.6, 11.8, 5.5, 14.2, 16.4, and 22.3 % of the total country area, respectively (Figure 1).

This research followed a workflow of three methodological stages (Figure 2). First, we developed a rainfall time series database with a daily resolution. Second, we identified the best empirical relationship to estimate erosivity using daily rainfall data. Third, we estimated the rainfall erosivity by using daily rainfall time series across the entire Mexican territory.

2.1 Rainfall time series (RTS) database development

The RTS database was developed through four steps: first, compilation, selection, and quality assurance of RTS. Second, the clustering of the RTS following its geographical and data attributes. Third, homogenization and data gap-filling of monthly and daily RTS. Fourth, quality control of the data gap-filling process.

2.1.1 Compilation, selection, quality assurance of RTS

In Mexico, the climate time series database results from the continuous effort of measuring, compiling, transcribing, and analyzing data reported by weather stations distributed throughout the entire Mexican territory. This effort is led by the National Meteorological Service, which makes the raw data collected since 1900 available to the public.

The data can be downloaded from the official National Meteorological Service site [https://smn.conagua.gob.mx/es/]. In the case of this project, 5454 files in plain-text format were downloaded, corresponding to the daily database of the entire network

https://doi.org/10.5194/essd-2025-306
Preprint. Discussion started: 5 August 2025

Open Access Earth System Science Data Discussions

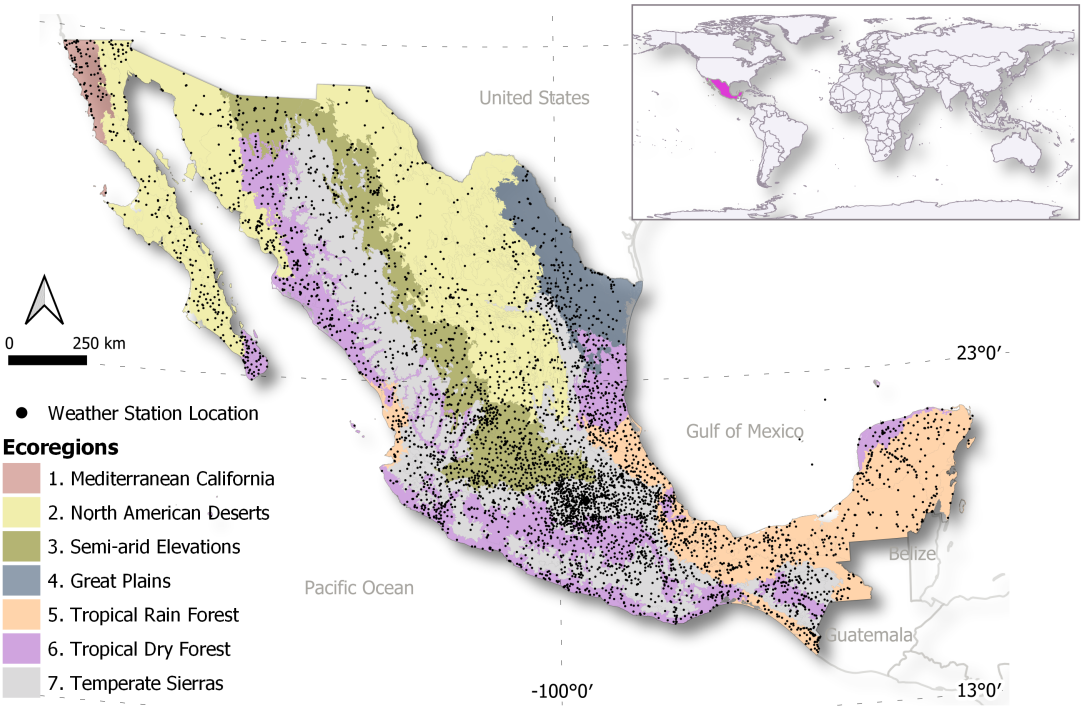


Figure 1. Ecoregions defined by the Commission for Environmental Cooperation (1997) and locations of the weather station available from the National Meteorological Service (SMN)

of weather stations (Figure 1) available up to June 2022. Each file contains a header in which the weather station identification is reported and a series of data on its geographical location. Additionally, the files have daily data for rainfall, evaporation, and maximum and minimum temperature.

All data processing, including file downloading, was implemented in R (R Core Team, 2022). The download process was automated using the `download.file` function implemented in the `utils` package. A preparation pre-process facilitates data handling by separating the header's information and the daily data. The header data were extracted and converted to shapefile (sic) (.shp) format for spatial management. Subsequently, the file is converted into a comma-delimited values format (*.cvs).

Finally, there were 5410 RTS at a daily temporal resolution with a unique format, and we considered those potentially useful RTS for our study. Consequently, there were two principal products in this step. First, a *.csv file with the information of the weather station, where each column corresponds to each of the data provided in the header (e.g., station ID, name, state, municipality, current situation, institution in charge, longitude, latitude, altitude, and report date), and each row corresponds to a weather station. Second, a set with 5410 *.csv files with the climate series of each weather station with four columns: date, rainfall (mm), temperature (°C), and relative humidity (%).

https://doi.org/10.5194/essd-2025-306
Preprint. Discussion started: 5 August 2025

Open Access Earth System Science Data Discussions

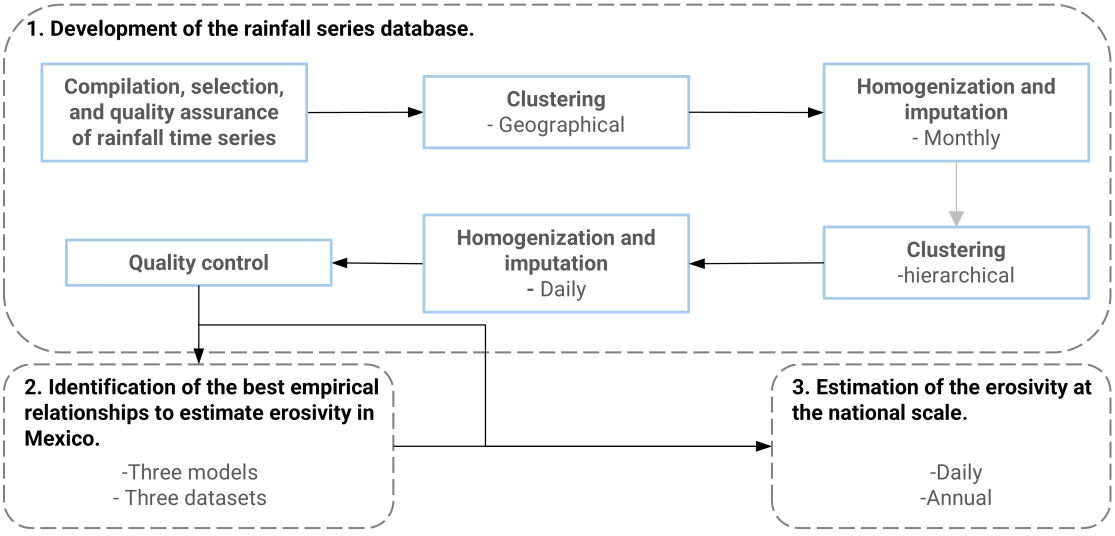


Figure 2. Workflow summarizing the three key methodological stages: (1) Development of the rainfall series database, (2) Identification of the best empirical relationships to estimate erosivity in Mexico, and (3) Estimation of the erosivity at the national scale.

To select the useful RTS, we identified the standard period of registers for each RTS between 1961 and 2018. On the other hand, for erosivity estimations, it is recommended to use a historical RTS of $\geq$20 years (Vantas et al., 2019; Renard and Freimund, 1994). Therefore, this article looked over the RTS in three climate normals: CN1 (1968 to 1997), CN2 (1978 to 2007), and CN3 (1988 to 2017). Subsequently, we selected those RTS with less than 20% of missing values (WMO, 2023) for each CN.

Calculating the monthly cumulative rainfall, we identified that long sequences of zeros and NAs were common. For quality assurance, we identified those years with no rainfall and replaced zero values with NAs. Those RTS with less than 20% missing values after the replacement were included in the resulting rainfall dataset. However, we used those RTS as a reference in the data gap-filling step.

2.1.2 Clustering of rainfall time series (RTS)

Clustering analysis is highly recommended when gap-filling a large set of RTS (Guijarro, 2014). Clustering similar RTS allows to use information from related series to fill in gaps. Rainfall patterns often exhibit spatial and temporal correlations, so data gap-filling from a group of similar series can result in more accurate estimates (Fransiska et al., 2024). In this workflow, we performed the data gap-filling by clustering the RTS according to 1) the geographical space, and the environmental characteristics of Mexico, and 2) the data dissimilarity among diverse environments (Figure 2). The first clustering was performed using the ecological regions of North America (Commission for Environmental Cooperation, 1997; INEGI-CONABIO-INE,

https://doi.org/10.5194/essd-2025-306
Preprint. Discussion started: 5 August 2025

2008). The second separation was performed after the data gap-filling and homogenization of the monthly series by ecoregion. A hierarchical cluster analysis groups the data series with similar seasonality and rainfall volume patterns (Gómez-Latorre et al., 2022). The number of clusters (k) ranged from 2 to $sqrt(n)$, where n is the total number of stations in the dataset (Rohlf, 1974). The better k values for each ecoregion were found using the Hartigan cluster validation index (Hartigan, 1975). This index was identified by Todeschini et al. (2024) to perform better when evaluating 68 cluster validation indexes (CVIs) over 21 different datasets. However, we did not perform a clustering analysis for Mediterranean California and Great Plains ecoregions due to the small size of the ecoregions and, therefore, the amount of the RTS available for those ecoregions.

2.1.3 Homogenization and data gap-filling of RTS

We followed three steps to get a complete RTS for each CN: quality assurance, homogeneity analysis, and data gap-filling (WMO, 2020). In this step, quality assurance involved verifying the physical and statistical consistency of the series, discarding outliers whose standardized anomaly was outside a predefined threshold and was unrelated to any climate variability events. Outliers were removed and replaced with NA values to be completed in the data gap-filling processes.

Homogeneity analysis removes the biases caused by some artificial breaks in the RTS (Yan et al., 2014). These breaks result from common issues such as reading or instrumental mistakes, instrumental changes, or special situations at the weather station location Guijarro (2014). We used the standard normal homogeneity test (Alexandersson, 1986) to analyze homogeneity.

To fill in missing data, we used the proportions method. This method estimates the missing information based on neighboring stations, considering the distance between each station (Paulhus and Kohler, 1952). The procedure used the three precipitation series with the highest correlation coefficient to the series that will be filled, with the condition of having been previously normalized. Then, we estimated $N_1 = \frac{A_1}{3}(\frac{N_a}{A_a} + \frac{N_b}{A_b} + \frac{N_c}{A_c})$; where N_a, N_b and N_c are the precipitation data for each of the stations with the highest correlation, while A_a, A_b and N_c are their corresponding normal average. All of the data gap-filling and homogenization process was made with the `homogen` function of the `climatol` package (Guijarro, 2024).

We summarized the homogenization parameters used for each step. Table A1 shows the parameters used to homogenize the monthly series, while Table A2 shows the parameters used to homogenize daily series by ecoregions 2, 3, 5, 6, and 7 and their corresponding subgroups.

2.1.4 Quality control of the data gap-filling processes

A quality validation was performed using the McCuen test (McCuen, 2016) to ensure consistency during the data gap-filling process of the rainfall time series. McCuen test compares the differences between the aggregated rainfall of the original multi-annual monthly means and the final series. The generated RTS with a difference greater than 10% related to its original were discarded.

https://doi.org/10.5194/essd-2025-306
Preprint. Discussion started: 5 August 2025

Open Access Earth System Science Data Discussions

2.2 Identification of the best empirical relationships to estimate erosivity in Mexico

In this step, we identified the best empirical relationship to calculate the R factor according to the erosivity characteristics of Mexico. To achieve this, we evaluated three combinations of parameters $alpha$ and $beta$ in a power law model and used three databases.

2.2.1 Empirical relationships to estimate daily erosivity

The empirical relationship follows the power-law model proposed by Richardson et al. (1983) when using daily precipitation to estimate the daily erosivity R_d (Equation 1). It has been seen in experiments worldwide in tropical (Rutebuka et al., 2020) and subtropical (Karami et al., 2012) zones that the power model has better performance than other models when calculating erosivity from daily data.

$$R_d = \alpha P_d^{\beta} \tag{1}$$

where P_d is the daily precipitation, α, and β are the adjusted coefficients. We tested three combinations (Model I, II, and III) of adjusted coefficients as follows:

- **Model I** (Richardson et al., 1983).

The value of α is equal to 0.18 for the cool season (October to March) and 0.41 for the warm season (April to September). The value of β is equal to 1.81 and constant throughout the year.

-**Model II** (Liu et al., 2020).

The value of α and β variate depending on the climate zone according to the Koppen-Greig Classification as follows: for Tropical (A), $\beta = 1.964 + 0.013 \times \text{Latitude}$, and $\alpha = 10^{(2.363-1.561\times\beta)}$. For Arid-steppe (BS), $\beta = 1.73$ and $\alpha = 0.3296$. For Arid-desert (BW), $\beta = 1.514$ and $\alpha = (2.123 - 0.04 \times \text{Latitude}) \times 10^{(1.781-1.341\times\beta)}$. For Temperate with dry summer (Cs), $\beta = 1.563$ and $\alpha = 0.2735$. For Temperate with dry winter (Cw), $\beta = 1.558$ and $\alpha = 0.817$. Finally, for Temperate with dry winter (Cf), $\beta = 1.5$ and $\alpha = (3.792 - 0.012 \times \text{Longitude} - 0.037 \times \text{Latitude}) \times 10^{(3.016-2.079\times\beta)}$. To identify the climate classification at each location of the datasets, we used the Koppen-Greig Classification for the present (1980-2016) at 1 km of spatial resolution made by (Beck et al., 2018).

-**Model III** (Xie et al., 2016).

The value of α is equal to 0.2686. The value of β is equal to 1.7265. This model also includes a sinusoidal relationship to describe the annual cycle of the coefficient of the power law function to represent seasonal differences in rainfall characteristics (Equation 2) as proposed by Yu and Rosewell (1996).

$$R_d = \alpha[1 + \eta cos(2\pi f j - \omega)]P_d^{\beta} \tag{2}$$

where f is the monthly frequency (1/12); ω is $7\pi/6$; η is 0.5412; and j is the j-month of the year.

https://doi.org/10.5194/essd-2025-306
Preprint. Discussion started: 5 August 2025

2.2.2 Databases

We used three databases; two have information on $EI30$ on a global and national scale, and the other database was constructed using sub-hourly RTS to calculate $EI30$ on a local scale in Michoacán state. These three databases have a different statistical distribution of the $EI30$ factor (Figure A1a), displaying a wide range of values from 333 to almost 26000 MJ mm ha^{-1} h^{-1} yr^{-1}. Additionally, all validation databases show a strong linear relationship between the $EI30$ values and the mean annual rainfall (Figure A1b).

-GloREDa

On the global scale, the GloREDa database was built using data from almost 4000 weather stations worldwide (Panagos et al., 2023). They estimated the $EI30$ from RTS with a resolution from 1 to 60 min. In Mexico, the GloREDa database has 15 locations with information on $EI30$ across continental territory (Figure 3a). The $EI30$ factor in these 15 locations was calculated from 5-minute RTS from 2005 to 2015; however, not all the RTS have registered for the multiyear period of ten years. The $EI30$ factor in this database shows a mean value of 3700.7 MJ mm ha^{-1} h^{-1} yr^{-1}, a standard deviation of 5719.2 MJ mm ha^{-1} h^{-1} yr^{-1}, and a range of 333.5 to 22743.6 MJ mm ha^{-1} h^{-1} yr^{-1}.

-Cortés

At the national scale, Cortés (1991) estimated the $EI30$ factor using 54 RTS across Mexico (Figure 3b). The temporal resolution for those 54 RTS is 1 minute, with a temporal period from 1977 to 1987. However, not all RTS have registers for those ten years, so we selected 42 RTS presenting more than five years of registers. The $EI30$ factor in this database shows a mean value of 4347 MJ mm ha^{-1} h^{-1} yr^{-1}, a standard deviation of 4827 MJ mm ha^{-1} h^{-1} yr^{-1}, and a range of 504 to 25654 MJ mm ha^{-1} h^{-1} yr^{-1}.

-Michoacán

At the local scale, the Michoacan mountain region has a database with 30 RTS (Figure 3c). These RTS have a 15-minute temporal resolution, and the period of registers is from 2011 to 2017. The weather stations belong to the Association of Producers, Packers, and Exporters of Avocado from Mexico (APEAM). However, as finer time resolution is available in this case, we estimated the $EI30$ factor as defined in Wischmeier and Smith (1978) by using the `RainfallErosivityFactor` package (Cardoso et al., 2020) in R project. The $EI30$ factor in this database shows a mean value of 10092 MJ mm ha^{-1} h^{-1} yr^{-1}, a standard deviation of 4928 MJ mm ha^{-1} h^{-1} yr^{-1}, and a range from 4580 to 22928 MJ mm ha^{-1} h^{-1} yr^{-1}.

2.2.3 Comparison of the empirical relationships

We compared the three $EI30$ databases against the R factor estimations by using the three coefficients (α and β) combinations called Model I, Model II, and Model III. We identified the direct comparison against the validation databases and our estimated Mexican databases (Mexico-CN1, Mexico-CN2, and Mexico-CN3) according to the overlapping between periods. Thus, the GloREDa database was compared against the R factor calculated by using Mexico-CN3 (1988-2017). the Cortés database was compared against Mexico-CN1 (1968-1997) and Mexico-CN2 (1978-2007). and the Michoacán database was compared against Mexico-CN3 (1988-2017). Afterward, for each point in the GloREDa and Cortés databases, we identified the nearest

https://doi.org/10.5194/essd-2025-306
Preprint. Discussion started: 5 August 2025

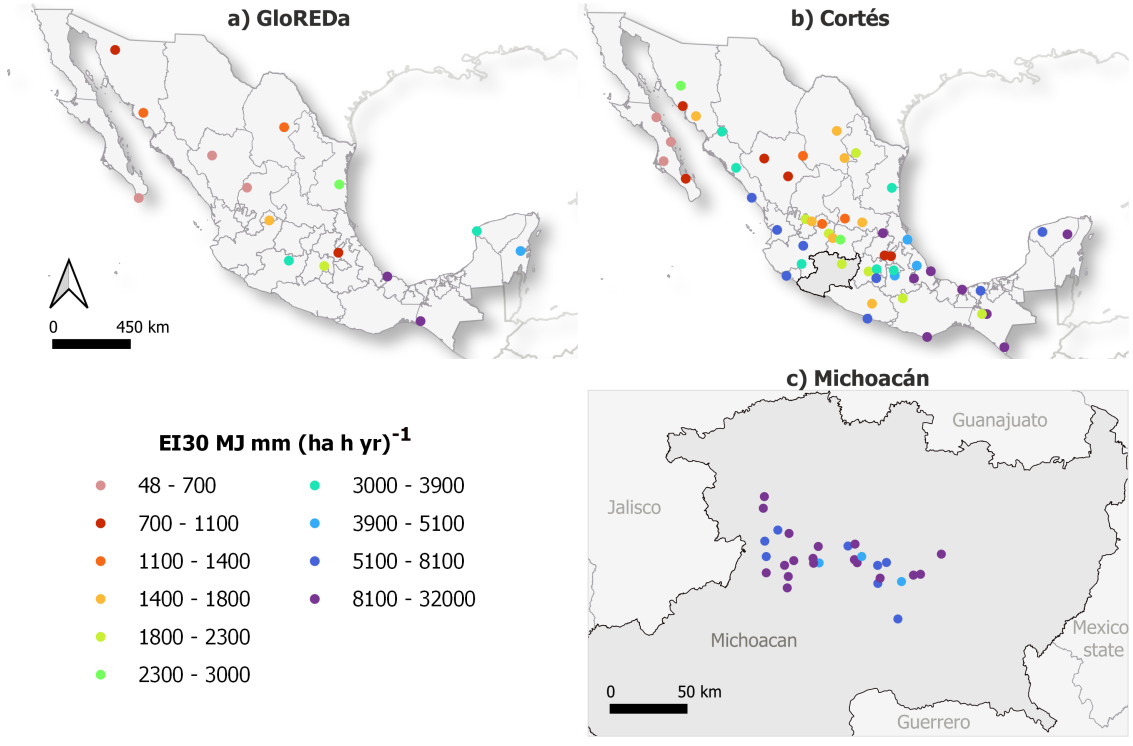


Figure 3. Spatial distribution of the validation datasets: a) GloREDa (n= 15), b) Cortés (n=44), and c) Michoacan (n=23).

point in the Mexican database produced in this study. Moreover, as the Michoacán database is a 15-minute temporal resolution RTS, we aggregated these RTS at a daily resolution to estimate the R factor with the three coefficient combinations.

For those points in our Mexican databases, we calculated the daily erosivity R_d factor by using Model I, Model II, and Model III. Following, we calculated the annual erosivity R_y (MJ mm ha^{-1} h^{-1} yr^{-1}) as the sum of the daily erosivity values in a year time (R_y=$\sum_{i=1}^{ed} R_{d_i}$). Finally, the R factor corresponds to the mean annual rainfall erosivity values for a multi-year period (R=$\sum_{j=1}^{30} R_{y_j}$).

We performed a linear regression to identify the relationship between the $EI30$ and R factor. In this sense, the slope of the linear model quantifies the mean change of the R factor when $EI30$ increases by a unit.

2.3 Rainfall erosivity estimation

We estimated the R factor for each weather station on the Mexican RTS database. First, we identified the days with erosive rainfall as those with cumulative precipitation greater than 12.5 mm, an extension of the suggestion by (Wischmeier and Smith, 1978; Shin et al., 2019; Efthimiou, 2018). Second, we calculated daily erosivity R_d using the best coefficient combination identified in the previous step (sec. 2.2). Finally, we estimated the mean annual rainfall erosivity, R factor, as described in the previous step (sec. 2.2.3).

https://doi.org/10.5194/essd-2025-306
Preprint. Discussion started: 5 August 2025

At this point, it is essential to highlight that we did not consider the erosivity due to the snowmelt because there is a pint-sized area covered with snow (or risk of snowfall), and no monitoring system for snowfall is publicly available in Mexico. To estimate the surface covered by snow, we explored the climate classification developed by García (1998), a modified Köppen classification system to better fit Mexico's climate conditions. According to the Köppen classification, only 83 km^2 (0.004% of the total area of Mexico) can be classified as E climates in the format of ET (tundra: temperature of warmest month greater than 0 °C but less than 10 °C) and EF (snow/ice: temperature of the warmest month 0 °C or below), both only produced by high altitudes. Additionally, the National Center for Disaster Prevention of Mexico developed a national snowfall danger index at a municipality level based on the occurrence of snowfall during the centuries XV to XXI (Jiménez Espinosa et al., 2012). They found that in the study period, snowfall had never occurred in 93.8% of Mexico's municipalities. The municipality with the highest snowfall frequency is Juarez (Chihuahua state), with just 30 events over five centuries. On the other hand, in Mexico, there is no public monitoring system to account for snowfall. Indeed, some research efforts have tried to relate the characteristics of snowfall with other variables such as ground surface temperature (Soto and Delgado-Granados, 2023).

At the end of this step, we obtained three datasets with R factor values for the three CNs: Mexico-CN1 (1968-1997), Mexico-CN2 (1978-2007), and Mexico-CN3 (1988-2017).

3 Results

This section presents the results of the three principal methodological stages outlined in the workflow. First, we developed a daily-resolution rainfall time series database, which provided a basis for further analysis. Second, we identified the best empirical relationship to estimated daily erosivity. Third, based on that best empirical relationship, we estimated rainfall erosivity values using the daily-resolution rainfall time series resulted in the first step.

3.1 Rainfall time series database development

The resulting database includes 1479, 1774, and 1721 RTS for CN1, CN2, and CN3, respectively (Table 1). The RTS with less than 20% of missing values and the identification of the number of series with NA and zero consecutive values are summarized in Appendix A3. After the data gap-filling, homogenization, and quality control processes, we discarded 4.8% of the RTS. Hence, Table 1 shows that the largest number of RTS corresponds to CN1 (1968-1997), with 7.4% (109 RTS), which is followed by CN2 (1978-2007), with 5.4% (95 RTS) and CN3 (1988-2017), with 1.9% (33 RTS). Likewise, the most significant proportion of discarded RTS corresponded to Mediterranean California (3 RTS, 21.4%) and North American Deserts (29 RTS, 15.84%) in CN1; Mediterranean California (2 RTS, 11.11%) and Great Plains (7 RTS, 12.72%) in CN2; and Great Plains (5 RTS, 10.86%) in CN3. Additionally, in the data gap-filling processes, we identified that none of the 5 RTS available for Mediterranean California had a rainfall register for three consecutive years, which did not allow us to carry out the data gap-filling process for the complete study period. Finally, the available RTS are 1370, 1676, and 1683 for CN1, CN2, and CN3, respectively. The available RTSs are distributed across the Mexican territory for the three CNs and represent the seven ecoregions (Figure 4).

https://doi.org/10.5194/essd-2025-306
Preprint. Discussion started: 5 August 2025

Open Access Earth System Science Data Discussions

Table 1. Number of available rainfall time series before and after data gap-filling process, as well as the number of discarded rainfall time series. The frequency of rainfall time series (RTS) is shown by ecoregion and climate normal.

Ecoregion	RTS before data gap-filling process			RTS with change greater than 10%			RTS after the data gap-filling process		
	1968-1997	1978-2007	1988 - 2017	1968-1997	1978-2007	1988 - 2017	1968-1997	1978-2007	1988 - 2017
Mediterranean California	14	18	5	3	2		11	16	
North American Deserts	183	273	243	9	18	4	154	255	239
Semi-arid Elevations	248	314	322	5	12	5	243	302	317
Great Plains	38	55	46	3	7	5	35	48	41
Tropical Rain Forest	195	237	236	6	6	3	179	231	233
Tropical Dry Forest	401	466	454	6	25	10	375	441	444
Temperate Sierras	400	411	415	27	25	6	373	386	409
Total	**1479**	**1774**	**1721**	109	**95**	**33**	**1370**	**1679**	**1683**

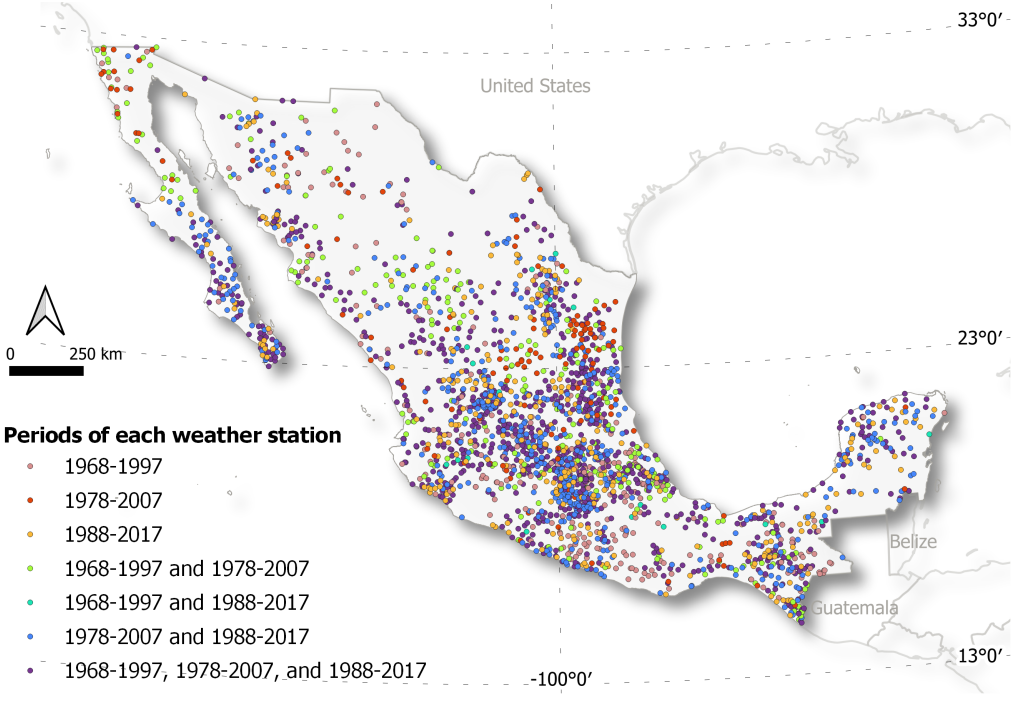


Figure 4. Spatial distribution of the available weather stations for each climate normal: CN1 (1968-1997), CN2 (1978-2007), and CN3 (1988-2017).

The percentage of variation of the series mean by ecoregion and CN using the two-step clustering is shown in Table 2. The CN2 (1978-2007) showed the highest average change in the mean, with -1.80%, where it is also noted that four of the seven ecoregions present relatively high average changes. Notably, in CN1 (1968-1997), for North American Deserts, the highest

https://doi.org/10.5194/essd-2025-306
Preprint. Discussion started: 5 August 2025

average change in the mean is -1.50%. In the CN3 (1988-2017), for Tropical Dry Forest and for Temperate Sierras, it is -3.38% and -1.61%, respectively. Additionally, Table A4 shows the Root Mean Square Error (mm) of the monthly accumulated rainfall by ecoregion. It can be seen that the highest RMSE was found for Tropical Rain Forest in the three CNs (6.12, 5.72, 6.78 mm, respectively), followed by Great Plains in CN2 and CN3 (5.72, 6.78 mm, respectively).

Table 2. Average percent (%) of change of the mean of rainfall time series by ecoregion

	Ecoregion	Average change of the mean (%)		
		1968-1997	**1978-2007**	**1988 - 2017**
1	Mediterranean California	-0.84	3.24	
2	North American Deserts	-1.50	-1.36	-0.14
3	Semi-arid Elevations	-0.25	-1.37	0.03
4	Great Plains	-0.69	3.23	-0.29
5	Tropical Rain Forest	-0.06	0.20	0.10
6	Tropical Dry Forest	-0.54	-0.49	-3.38
7	Temperate Sierras	-0.20	-3.73	-1.61
	Total	-0.28	**-1.80**	-0.98

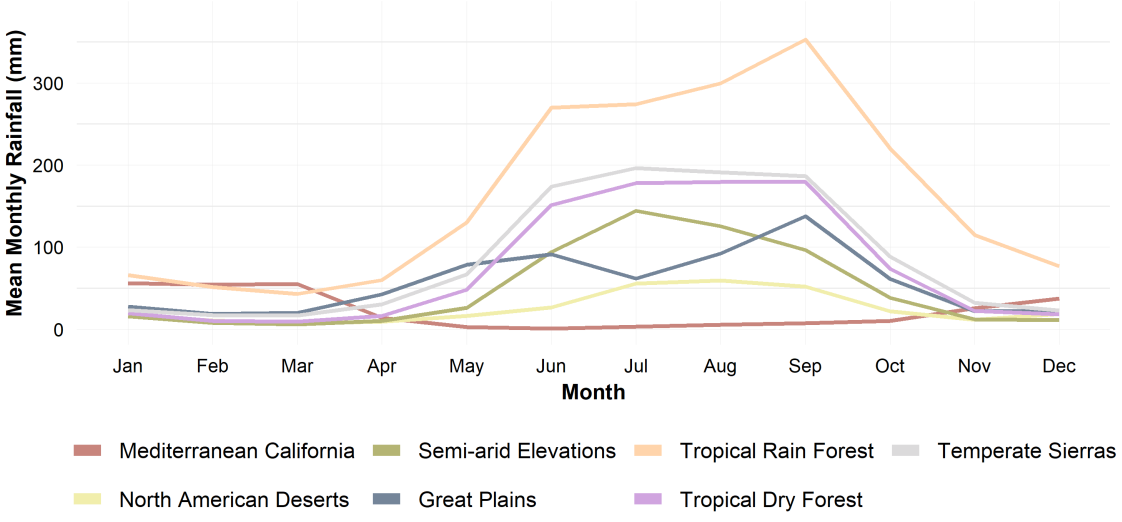
Figure 5 shows the monthly rainfall distribution for the seven ecoregions for CN1 (1968-1997). In Figure A2 a and b, there are the monthly rainfall distributions for the CN2 and CN3, respectively. Generally, rainfall is concentrated between May and November in all ecoregions, except in Mediterranean California, where high values occur between November and April. For Mediterranean California, North American Deserts, Semi-arid Elevations, and Great Plains, the rainfall does not exceed 200 mm for the month with the highest volume. However, rainfall in Tropical Rain Forest, Tropical Dry Forest, and Temperate Sierras can reach up to 800 mm in June and September. The general behavior of rainfall distribution by ecoregion is the same for the three CNs.

3.2 Identification of the best empirical relationships to estimate erosivity in Mexico

This section presents the results of identifying the best empirical relationship between $EI30$ and the R factor calculated from three models. To achieve this, we built linear models to identify the rate of change between the $EI30$ factor and the R factor estimated with the three models (Figure 6).

Model I was the coefficient combination that performed better in the erosivity estimations compared to the $EI30$ values estimated with RTS at sub-hourly resolution. The model I is that with a closer slope to one (1:1 perfect model) with 1.07, 0.92, 0.83, and 0.43 for GloREDa vs. Mexico-CN3, Cortés vs. Mexico-CN1, Cortés vs. México-CN2 and the two predictions for Michoacán database. Model II is the one with the slope furthest from the perfect model (1:1 line) for the GloREDa database, with a slope of 1.97, and it is the second in order, with a slope of 1.31 and 1.2 for Cortés Mexico-CN1 and Mexico CN2, respectively. Model III has the slope furthest from the perfect model (1:1 line) for Cortés (Mexico-CN1 and Mexico-CN2) and Michoacán databases with slopes of 0.61, 0.55, and 0.3, respectively. The model I had the fewest RMSE for all the four

https://doi.org/10.5194/essd-2025-306
Preprint. Discussion started: 5 August 2025



(a)

Figure 5. Mean monthly rainfall for all seven ecoregions in the climate normal (CN1) 1968-1997.

comparisons (1544, 2438, 2628, and 4978 MJ mm ha^{-1} h^{-1} yr^{-1}) (See Table A5). Model I also shows a general overestimation in predicting $EI30$ values in the GloREDa and Cortés (Mexico-CN1 and Mexico-CN2) databases with a ME of 956, 324, and 338 MJ mm ha^{-1} h^{-1} yr^{-1}, respectively. Model III underestimated the $EI30$ values in all databases. Additionally, for the Michoacán database, all three models underestimated the $EI30$ values. On the other hand, the distance of each point of the GloREDa and Cortés (Mexico-CN1 and Mexico-CN2) database to the nearest weather station varied in a range of 0.6 to 42.0 km (mean:11.6 km), 0.3 km to 73.4 km (mean: 12.17 km), and 0.3 to 76.5 km (mean: 10.7), respectively. It is important to highlight that there is no correlation between error and the distances between points in the validation datasets and the weather stations of our Mexican databases.

3.3 Rainfall Erosivity estimations

In this section, we will display the results of the rainfall erosivity estimations for the three CNs studied. First, a descriptive statistic for erosivity values, then a yearly distribution of rainfall erosivity, and finally, their spatial distribution.

The erosive days were considered as those with rainfall greater than 12.5 mm. Figure A3 shows the mean number of locations with erosive rain for each day of the year. For the three CNs, there is one erosive period throughout the year; however, for CN3, the erosive period is marked by two peaks, first from day 173 to 190 and second from day 229 to 261; these days represent those with the most (10% higher) locations reporting erosive rainfall. On the other hand, the 90th percentile number of locations with erosive rainfall for CN2 and CN3 has increased by 7% (238) and 12% (227), respectively, compared to CN1 (212).

The mean values for CN1, CN2, and CN3 were 5276 (SD 5662), 4832 (SD 5266), and 5067 (SD 5071) MJ mm ha^{-1} h^{-1} yr^{-1}, respectively (Table 3). The statistical distribution of the erosivity values was right-skewed with skewness of 2.7, 3.0, and

https://doi.org/10.5194/essd-2025-306
Preprint. Discussion started: 5 August 2025

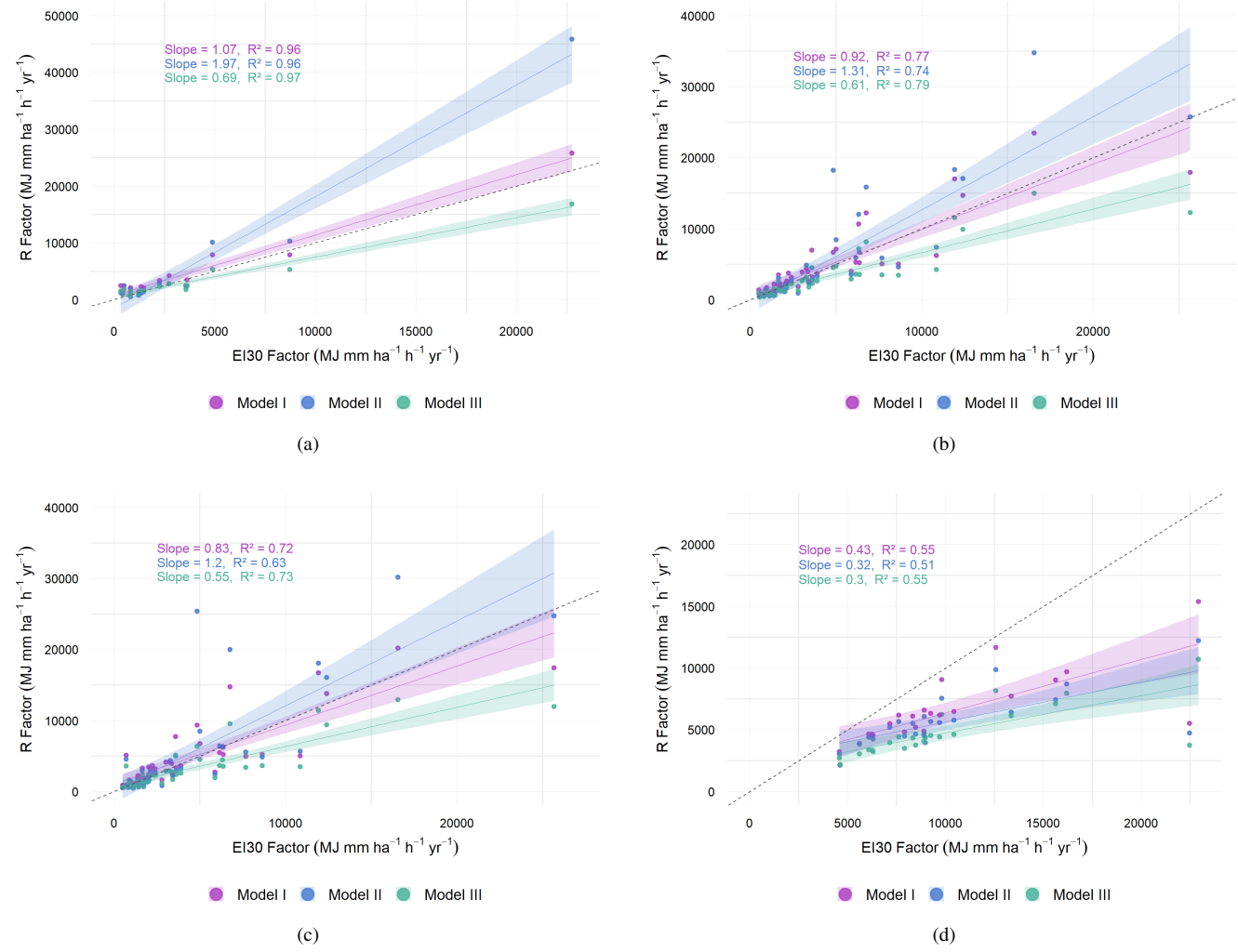


Figure 6. Verification of rainfal erosivity index calculated with three models: model I (Richardson et al., 1983), Model II (Liu et al., 2020) and Model III (Xie et al., 2016). The verification was made with three erosivity databases at different scales (global, national, and local). a) Global dataset GloREDa (Panagos et al., 2023) and Mexico-CN3. b) National dataset Erosivity-Cortés (Cortés, 1991) and Mexico-CN1. c) National dataset Erosivity-Cortés (Cortés, 1991) and Mexico-CN2. d) Local dataset in Michoacan state: comparison of EI30 values calculated using RTS at two temporal resolutions, 15-minute and daily resolution, respectively.

https://doi.org/10.5194/essd-2025-306
Preprint. Discussion started: 5 August 2025

2.9 for CN1, CN2, and CN3, respectively. So the median was less than the mean with values of 3245, 3070, and 3327 MJ mm ha^{-1} h^{-1} yr^{-1} for CN1, CN2, and CN3, respectively. The Krustal-Wallis test showed that the erosivity median value for CN2 differed from CN1 and CN3 at a 95% confidence level. The kurtosis values of 12.6, 15.1, and 15.3 for CN1, CN2, and CN3, respectively, indicate large tails with the presence of outliers. In this case, the outliers are high erosivity values reaching more than 12000 MJ mm ha^{-1} h^{-1} yr^{-1} for the three CNs (Figure A4).

Table 3. Descriptive statistic of the erosivity values for three climate normals (1968-1997, 1978-2007, 1988-2017). Min: minimum, Max:maximum, SD: standard deviation, CV: coefficient of variation, skew: skewness, kurt: kurtosis

Climate normal	n	Mean	Median	SD	Min	Max	CV	Skew	Kurt
CN1 (1968-1997)	1369	5276	3245a	5662	65.5	49129	1.1	2.7	12.6
CN2 (1978-2007)	1678	4832	3070b	5265	56.8	44660	1.1	3.0	15.1
CN3 (1988-2017)	1676	5067	3327a	5071	64.2	43368	1.0	2.9	15.3

Monthly erosivity varies throughout the year in the three CNs (Figure 7a). The behavior of the erosivity in the three CNs is monomodal, indicating just one peak and one valley. The month with the highest erosivity is September, reaching values almost to 1300 MJ mm ha^{-1} h^{-1} yr^{-1}, followed by August, July, and June with erosivity values around 1000 MJ mm ha^{-1} h^{-1} yr^{-1}. It is important to highlight that for June, July, and August, the CN1 (1968-1997) had the highest monthly erosivity values; however, for September, the highest values are found in the CN2 (1988-2017). The months with the smallest erosivity values are February, followed by March with values less than 40 MJ mm ha^{-1} h^{-1} yr^{-1}.

As the three CNs had the same behavior throughout the year, we displayed in Figure 7b the monthly erosivity by ecoregion. It can be seen that the Tropical Rain Forest has the highest erosivity values for all months, with a minimum value in March (114 MJ mm ha^{-1} h^{-1} yr^{-1}) and a maximum in September (3395 MJ mm ha^{-1} h^{-1} yr^{-1}). Notably, the ecoregion with the smallest monthly erosivity is the North American Deserts, with values hardly reaching 500 MJ mm ha^{-1} h^{-1} yr^{-1}. However, the figure shows the erosivity values for the CN3 (1988-2017), and it does not have a weather station in the Mediterranean California ecoregion. Still, in CN1 and CN2, Mediterranean California had the smallest erosivity values. All ecoregions have the highest erosivity in September except for the Semi-arid elevations, which the peak is in July.

Regarding the spatial distribution, the erosivity values across Mexican territory look similar for the three CNs (Figure A5). The ranges plotted in the legend correspond (not precisely) to the deciles of the three statistical distributions. In the California peninsula, there are concentrated those fewer erosivities (peach dots) with less than 1000 MJ mm ha^{-1} h^{-1} yr^{-1}, which represents the 1st decile of the distribution. The 2nd decil (values between 1000 and 1600 MJ mm ha^{-1} h^{-1} yr^{-1}) is concentrated in the northern region and extends a little to the central region (red dots). These areas represent the North American deserts and Semiarid elevations ecoregions. On the other hand, the 9th (dark blue dots) and 10th (dark purple dots) with values between 7600 to 12000 and greater than 12000 MJ mm ha^{-1} h^{-1} yr^{-1}, respectively, are concentrated in the tropical rain forest ecoregion in the southwest region, and some places in the Tropical Dry Forest ecoregion in the southeast of Mexico. It is clear that for the CN3 (1988-2017), there were no RTS for Mediterranean California and some places in the North American desert ecoregions because the rainfall series there had records until 2012.

https://doi.org/10.5194/essd-2025-306
Preprint. Discussion started: 5 August 2025

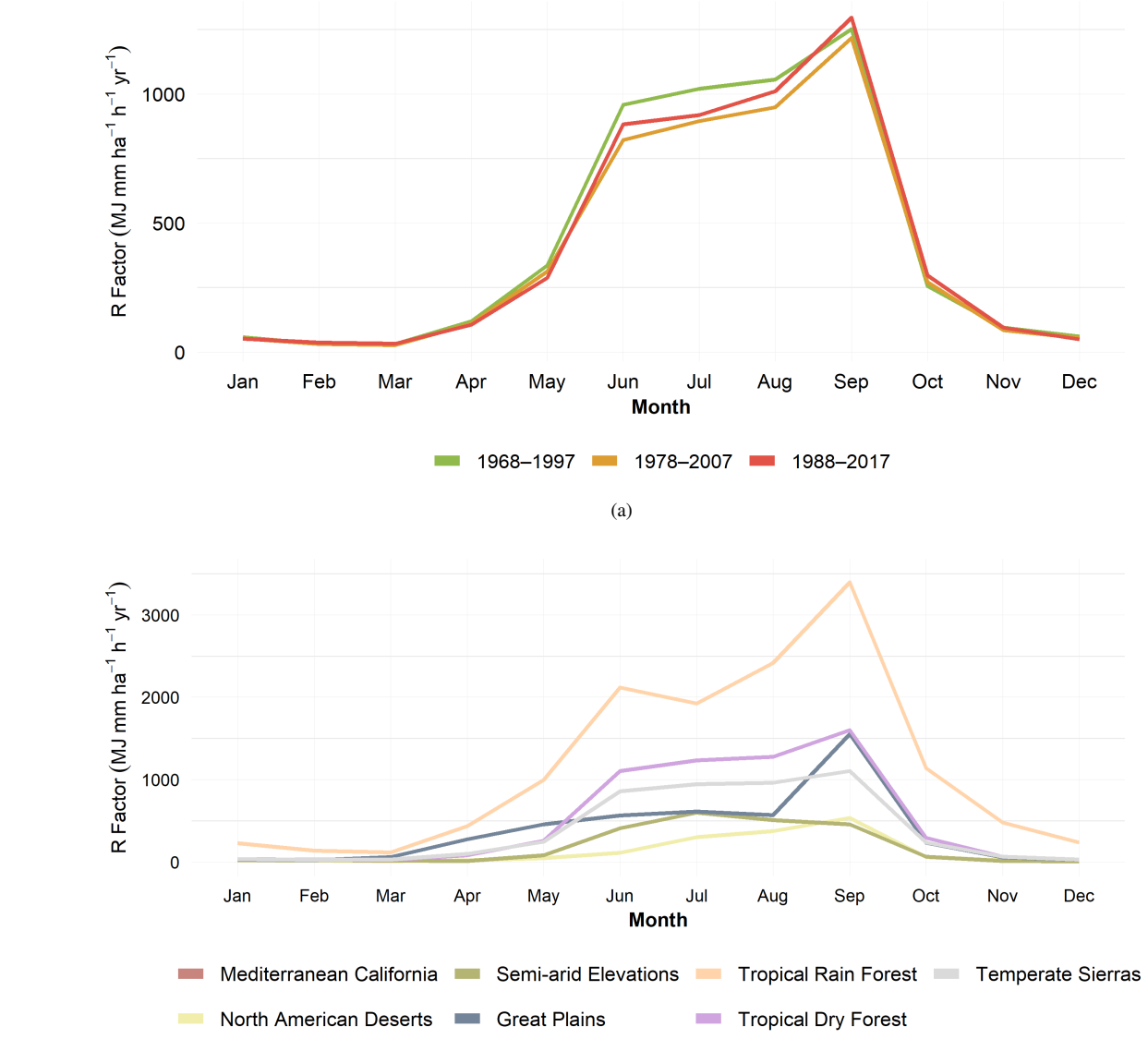


Figure 7. Montly erosivity values estimated from daily rainfall time series and using the Richardson et al. (1983) power-law model. a) Monthly erosivity for the three climate normals; b) Monthly erosivity by ecoregion for the Mexico-CN3 (1988-2017).

https://doi.org/10.5194/essd-2025-306
Preprint. Discussion started: 5 August 2025

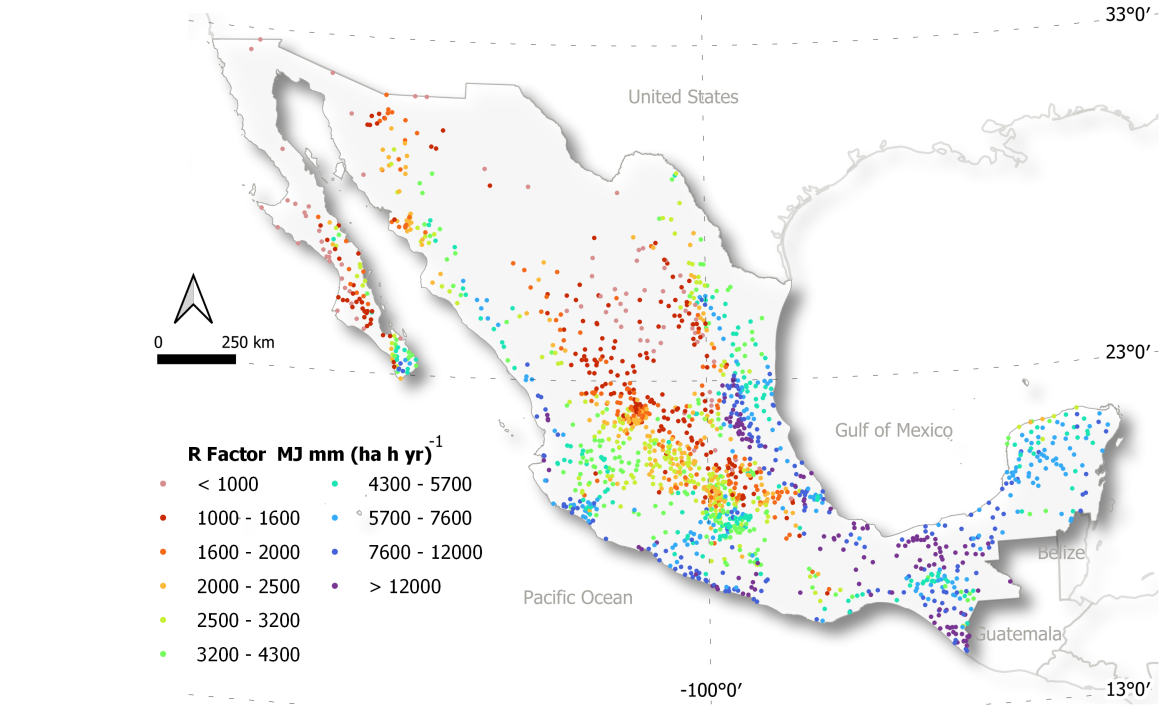


Figure 8. R factor values calculated with the power law equation proposed by (Richardson et al., 1983) at daily resolution for the climate normal 1988 - 2017.

4 Discussion

This research addressed the data incompleteness and breakpoints in the available Mexican climate time series, resulting in a comprehensive and unprecedented rainfall time series (RTS) database. The new information primarily applies to analyzing rainfall erosivity, which is required for studying soil erosion at the national scale in Mexico. The new information is appealing for diverse users, as the insufficient availability of climate information (i.e., publicly funded data) represents an interoperability barrier limiting the development of climate services and the understanding of key ecosystem climate-related processes at different scales (Vaughan et al., 2016). Consequently, scientists search for new ways to create national climate-related datasets containing state-of-the-art information for various applications such as climate modeling, yield prediction, or ecological forecasting. Particularly, we compiled and systematized a national dataset with daily rainfall and rainfall erosivity for three CNs (CN1:1968-1997, CN2:1978-2007, and CN3:1988-2017) across Mexico. To the best of our knowledge, this research is the first effort to develop a large daily RTS dataset based on legacy climate data, at the national scale, following the quality control and inhomogeneity analysis proposed by the World Meteorological Organization.

We address the incompleteness and breakpoints in the Mexican climate time series available data, as indicated in previous reports (Cuervo-Robayo et al., 2020). The new information ensures the highest possible quality (as explained in the methods

https://doi.org/10.5194/essd-2025-306
Preprint. Discussion started: 5 August 2025

section), it includes a rigorous quality control, homogenization, and data gap-filling procedures for 1370, 1679, and 1683 RTS for the CNs 1968-1997, 1978-2007, and 1988-2017, respectively. Other extensive studies have been conducted at the national scale, but they use a coarser time resolution compared to those studied here. For example, Cuervo-Robayo et al. (2014) updated the mean monthly rainfall for 100 years (1910-2009) using around 5000 RTS. Other studies have compared multiple gap-filling approaches at a regional scale (Cespedes et al., 2023) and conducted quality control and homogenization process (García-Cueto et al., 2019) using daily RTS. Yet, those studies used RTS datasets representing shorter periods of time than ours. Studies using RTS across comparable periods of time to our study have considered RTS with the fewest missing data, without considering a data gap-filling process Pineda-Martínez and Carbajal (2017); Mateos et al. (2016). Thus, our research represents a significant contribution by offering a more comprehensive and higher-quality daily RTS dataset, addressing limitations related to the dataset size, representativeness, and the rigorous homogenization and data gap-filling process. We contribute a new standard for climate data analysis nationally in Mexico.

Our methodology allowed us to reconstruct climate series with significant missing data gaps, as reported in previous studies (Guijarro, 2014). The methodology applied on an ecoregion basis, potentially increases spatial coherence in the completed and homogenized RTS, as previously recommended (Adeyeri et al., 2022). The methodology is also recommended by World Meteorological Organization (2020). Other robust data gap-filling strategies leverage high computational capacity and big data analyses (i.e., machine learning) to obtain reliable RTS (Hırca and Türkkan, 2024; Lupi et al., 2023). These strategies build a predictive function to estimate values for missing data using the climate series data itself or auxiliary data such as satellite products (Duarte et al., 2022). Despite the performance predicting missing data values (Hırca and Türkkan, 2024; Lupi et al., 2023), many machine learning methods are still experimental, and a well-known and WMO-recommended methodology (as a reference point) is needed to benchmark more complex data-driven approaches. Our effort is not error-free, thus we present an error estimation in our gap-filling approach in the supplementary material A4. Future improvements of this new RTS database must include a progressive exploration of multiple gap-filling data techniques.

The gap-filled dataset (considering the completed and homogenized RTS for the three CNs) is useful for analysis of changes in precipitation trends (Yan et al., 2014) across Mexico. This new RTS dataset shows that Mexico experiences a unimodal rainfall regime with a cusp in July and September, depending on the ecoregion. This unimodal regime was also reported by Carrera-Hernández (2025) who analysed RTS at the national scale, reporting that Mexico's rainy season is from May to October and the wettest months are July, August, and September. Additionally, the regional variation in rainfall patterns is markedly latitudinal (Figure 5 and A2), from Tropical Rain Forests across the southeast, to the water-limited environments of North American Deserts. These results are consistent with rainfall patterns previously reported by de Anda Sánchez (2020). In addition, the new dataset also provides an earlier perspective of the spatial distribution of rainfall erosivity in Mexican territory.

Among the three coefficient combinations evaluated for estimating daily rainfall erosivity, the Richardson et al. (1983) proposal consistently demonstrated superior performance across the national and local datasets. Although all three models adopt a similar functional structure of the form $R = \alpha P^{\beta}$, differences in the parameterization of the α and β coefficients have substantial implications for model behavior under diverse rainfall regimes. The Richardson et al. (1983) equation was developed using RTS across the United States of America, including rainfall conditions of the southern states such as Georgia,

https://doi.org/10.5194/essd-2025-306
Preprint. Discussion started: 5 August 2025

Mississippi, and Texas. This equation defines different α coefficient for the cool and warm seasons that occur from October to March and from April to September, respectively; it is very similar for the Mexican climate, where it is seen that the warmest months are those from May to October (Carrera-Hernández, 2025). Additionally, the β coefficient has no spatial or seasonal pattern, the same as found in the Xie et al. (2016) equation and contrary to the Liu et al. (2020) one, where the β coefficient is affected by the latitude in the Tropical (A) climate classification according to Köppen–Geiger. However, it is notable that the β coefficient for Richardson et al. (1983) equation (1.81) is slightly higher than that for Xie et al. (2016) and Liu et al. (2020) (1.72 and between 1.5 and 2 depending on the latitude, respectively) placing greater weight on high rainfall amounts which may better capture the erosive potential of intense storms. Indeed, this is the reason why the Richardson et al. (1983) model obtained the highest erosivity estimations for the Michoacán dataset (purple points always over the blue and green dots in Figure 6d), because β coefficients for all the Michoacán dataset were 1.81, 1.558 and 1.72 for the three models Richardson et al. (1983), Liu et al. (2020), and Xie et al. (2016), respectively.

On the other hand, different works have reported that the rainfall erosivity has a strong relationship with the accumulated rainfall, as seen at monthly (Rutebuka et al., 2020) and annual (Mikhailova et al., 1997) resolution. This relationship has also been seen in our three validation datasets, where the mean annual rainfall can explain 90, 70, and 40 % of the total variability of the $EI30$ values in the GloREDa, Cortés, and Michoacán databases, respectively (Figure A1a and b). Despite the difference between the three datasets, it is also seen that the adjustment coefficients proposed by Richardson et al. (1983) were those that performed best. Therefore, it is clear that the wettest months (from April to October, depending on the ecoregion) were the months with the highest rainfall erosivity when calculated from daily RTS.

The models based on a coarser RTS could underestimate and overestimate the $EI30$ values, as seen in this study. Although calculated at a coarser temporal resolution, underestimations have already been indicated by Tu et al. (2023); Yin et al. (2015) while evaluating the effect of modifying the time interval for calculating $R(EI30)$ using 5, 15, 30, and 60-minute RTS. The authors concluded that increasing the time interval leads to underestimating erosivity values. Similarly, Li et al. (2022) identifies that using a monthly model underestimates $R(EI30)$. However, the same author found an overestimation of the R values regarding $R(EI30)$ using annual models. However, no matter under or overestimation, it has been reported that when coarser time intervals are used instead of high temporal resolution, the relationship between E and $I30$ remains consistent, but a larger calibration coefficient is needed (Tu et al., 2023). Therefore, having more detailed information is arguably the best way to estimate the erosivity factor with greater certainty and to know which model explains the greatest variance of $R(EI30)$.

Our results highlight the advantage of having detailed information for the mountain region in Michoacan. However, trying to find adjusted coefficients for a power law model with the RST (15 min of temporal resolution) of the Michoacán database is not enough to represent the conditions of Mexican precipitation. The adjustment coefficient will apply only to the rainfall conditions of the mountain region in Michoacán. For other areas, it has been shown that the fit of a model is not the same for different rainfall patterns (Li et al., 2022). Regions with small amounts of rainfall, for example those in the California Mediterranean and North American Deserts ecoregions, it is necessary to fit specific adjustment coefficients (Chen et al., 2020).

https://doi.org/10.5194/essd-2025-306
Preprint. Discussion started: 5 August 2025

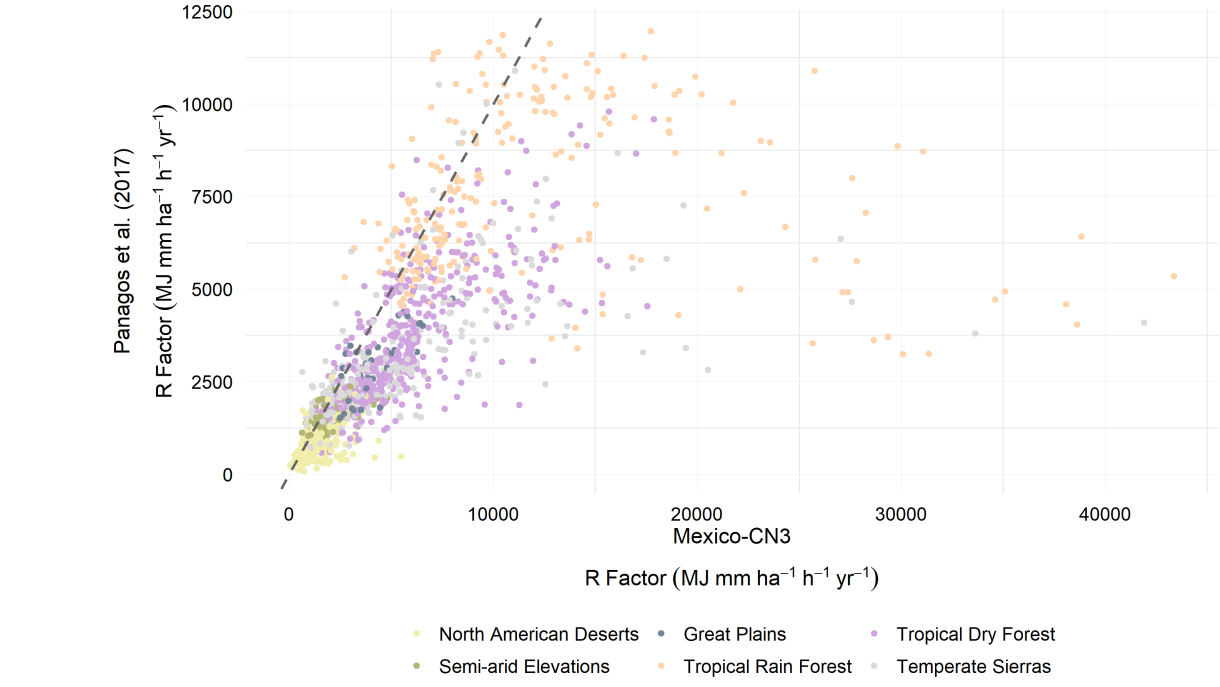
Our work reveals that the distribution of erosivity values in Mexico corresponds to the geographical distribution of the rainfall and seasonal rainfall conditions. The areas with the major erosivity values are concentrated in the isthmus of Tehuantepec (i.e., the shortest distance between the Gulf of Mexico and the Pacific Ocean), where the highest mean annual rainfall values in Mexico have also been reported (de Anda Sánchez, 2020). In contrast, lower erosivity values are in Mexico's central and northern regions, where severe droughts (e.g., that from the 1990s until the beginning of the twenty-first century), due to the large-scale changes in ocean-atmospheric circulation patterns, have affected the mean annual rainfall. On the other hand, we highlight a spot in the south of the California peninsula (e.g., Sierra La Laguna), where erosivity values are higher than those estimated in the north. This local variation is due to the influence of tropical cyclones, which contribute up to 50% of the mean annual rainfall in this region (Agustín Breña-Naranjo et al., 2015). Overall, the spatial variability in erosivity across Mexico reflects the interplay between rainfall patterns and climatic events, underscoring the significant influence of regional weather phenomena on soil erosion processes.

The new dataset is appealing for validating global datasets that are generally used to evaluate soil erosion by water at large scales when no more detailed information is available. One of those global products is that made by Panagos et al. (2017), where they used almost 4000 RTS at 30-min temporal resolution to estimate $EI30$ values and then map the erosivity distribution across the globe. We identified that our rainfall erosivity predictions for the Mexico-CN3 database obtained greater values than those from Panagos et al. (2017) product (dots under the 1:1 dashed line in Figure 9). The North American Deserts and the Semi-arid elevation ecoregions obtained similar erosivity estimations (yellow and green dots); in contrast, the Tropical Rain Forest and the Temperate Sierras were those ecoregions with the most significant differences between estimations (peach and gray dots). This difference in the Tropical Rain Forest (the wettest ecoregion) is evident because the ecoregion has higher mean annual rainfall and standard deviation (862 to 4823 mm and an SD of 758 mm). In comparison, the Panagos et al. (2017) reports a lower mean annual rainfall (1383-2100 mm and an SD of 402.35 mm, those values calculated with three RTS in the ecoregion). The same pattern is observed for Temperate Mountains, where Mexico-CN3 has a wider probability distribution of mean annual rainfall (range of 318-4009 mm and an SD of 500 mm) than Panagos et al. (2017) (range of 814 to 2258 mm, in this ecoregion, GloREDa has just two RTS). We therefore report a more complete description of RTS variance. We believe that our contribution could be useful in better representing global erosivity estimates.

Aligned with our results, Fenta et al. (2023) found the largest differences in annual rainfall erosivity values between the global product and a satellite-based approach (using sub-hourly RTS) for the rainiest regions (tropical and temperate climates) around the world. The best agreement between satellite-based rainfall erosivity (using the satellite precipitation estimates corrected and reprocessed with the Climate Prediction Center Morphing Technique - CMORPH) and the Panagos et al. (2017) product was found in Europe, where the density of rainfall gauges is the highest globally (Bezak et al., 2022). Furthermore, as the global product has in its database just 15 rainfall erosivity values in the Mexican area, it is important to identify the regions with high discrepancies between our national approach (with more than 1300 erosivity values for each CN) and the global product, to understand their limitations and use them in places with scarce rainfall erosivity information.

This new database will help as an indicator of rainfall erosivity patterns across Mexico, because we are analyzing rainfall and rainfall erosivity patterns in all seven ecoregions defined in Mexico. Also, we report rainfall erosivities at a wide range of

https://doi.org/10.5194/essd-2025-306
Preprint. Discussion started: 5 August 2025



(a)

Figure 9. Scatter plot of R factor values from the global rainfall erosivity surface from Panagos et al. (2017) and the R factor estimated with the Richardson et al. (1983) power law equation with the RTS of the Mexico-CN3 (1988-2017) database.

altitudes from 1 to 4283 meters above sea level. These estimations, based on daily rainfall data, are also frequently employed because daily rainfall records are widely available. Moreover, according to Yin et al. (2017) daily-scale data meet the three primary requirements of the Universal Soil Loss Equation (USLE) and its derived models: (1) estimation of average annual rainfall erosivity for predicting long-term soil loss; (2) construction of seasonal erosivity curves to reflect the interaction between rainfall distribution and crop management practices; and (3) calculation of daily or 10-year return period erosivity values to assess the impact of extreme events on runoff generation and the effectiveness of soil conservation measures, such as terracing.

The new database is a potential tool for local, national and global erosion studies. At the local scale, this database could serve as a tool to design field experiments to validate rainfall erosivity estimations in different rainfall patterns, as well as to identify the magnitude of the underestimations or overestimations when using different time-resolutions of the RTS (Meng et al., 2021; Zhao et al., 2019; Dunkerley, 2019). At the national scale, this database could serve as input in erosion models to generate a baseline of soil loss rates across Mexican territory. This necessity has been highlighted by Bolaños González et al. (2016). The authors emphasize the necessity to estimate soil and organic carbon loss rates. Additionally, this could help to support the development of environmental monitoring systems in Mexico. At the global scale, the results could serve as validation benchmarks and increase the representativeness of global rainfall erosivity databases such as GloREDa (Panagos et al., 2023)

https://doi.org/10.5194/essd-2025-306
Preprint. Discussion started: 5 August 2025

or the Global Rainfall Erosivity database from Reanalysis and Satellite Estimates -GloRESatE- (Das et al., 2024). The new information is appealing for the aforementioned efforts, as Mexican territory does not have enough data representation due to the lack of primary information. The new dataset and its possible improvements would allow for a more accurate estimation of erosivity and, thus, the annual amount of soil lost due to rainfall. This implies a better understanding of soil resources, better territorial planning, better agricultural and land use management, and better land conservation programs.

As potential limitations to this study, we selected just one methodology for gap-filling daily data. Multiple remote sensing products could help to improve gap-filling efforts. These products include the Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS) (Funk et al., 2015) with daily temporal resolution; the National Oceanic and Atmospheric Administration (NOAA) data presenting hourly time resolution, and Climate Prediction Center Morphing Technique (CMORPH), among others. Using various data sources improves the robustness of the resulting datasets (Bessenbacher et al., 2023) because those products enhance the completeness of the RTS and the accuracy of the gap-filling data. On the other hand, the selection of the model for calculating erosivity is a source of uncertainty in subsequent models estimating soil loss rates (Li et al., 2022). Therefore, evaluating the performance of different adjusted models is appealing for future research when high-resolution time series data are available. Additionally, at the national scale, it is important to test other daily models and estimate the variation of the predictions by ecoregions. Additionally, if finer RTS were available for the entire national territory, it would be possible to calibrate a potential equation for Mexico.

5 Conclusions

We present unprecedented rainfall and rainfall erosivity databases across Mexico. The research used legacy climate data to assess erosivity across Mexican territory. However, the rainfall time series (RTS) contained a relatively large number of missing values. A gap-filling procedure was performed to obtain gap-free estimates for three climate normals to increase RTS completeness. The new database provides a more detailed insight into rainfall erosivity with respect to global models. This study reveals that the North American deserts and Mediterranean California are regions where the rainfall has less erosive power, while tropical rainfall forests have the highest rainfall erosivity. The new database is available for public consultation. This database is for researchers and students, technical assistants, decision-makers, and other users interested in rainfall erosivity patterns and trends. Additionally, all environmental studies in Mexico, where the rainfall process is needed at the daily resolution, may benefit from this dataset.

6 Code availability

Rproject scripts to reproduce the workflow described in this research is available at: https://doi.org/10.5281/zenodo.15468097 (Varón-Ramírez, 2025)

https://doi.org/10.5194/essd-2025-306
Preprint. Discussion started: 5 August 2025

7 Data availability

Following the FAIR principles for scientific data, we published our resulting databases (Mexico-CN1 1968-1997, Mexico-CN2 1978-2007, and Mexico-CN3 1988-2017) and the completed daily rainfall time series for the three climate normals in the Environmental Data Initiative (EDI) at https://doi.org/10.6073/pasta/e0dc8bd3501f8c19bb750e853c3289cb (Varón-Ramírez et al., 2025) .

Rainfall erosivity databases contain eleven columns with the weather station location (code, coordinates, altitude, name, and ecoregion), the root means squared error (RMSE) of the data gap-filling process, rainfall erosivity, the accumulated number of days with erosive rainfall, and the multiyear mean rainfall.

Daily rainfall time series databases contain 1369, 1678, and 1676 columns for the climate normal 1968-1997, 1978-2007, and 1988-2017, respectively. Each column corresponds to one rainfall time series.

https://doi.org/10.5194/essd-2025-306
Preprint. Discussion started: 5 August 2025

Open Access Earth System Science Data Discussions

Appendix A

Table A1. Homogenization parameters for monthly series. Eco: Ecoregion (1: Mediterranean California, 2: North American Deserts, 3: Semi-arid Elevations, 4: Great Plains, 5: Tropical Rain Forest, 6: Tropical Dry Forest, 7: Temperate Sierras); dz.max and dz.min (upper and lower): standard deviations to consider suspicious and anomalous data

Climate normal	Ecoregion	inht	dz.max lower	dz.max upper	dz.min lower	dz.min upper
1968 - 1997	1	15	8	10	-8	-10
	2	25	12	13	-7	-8
	3	40	10	11	-8	-9
	4	20	6	7	-5	-6
	5	30	8	9	-6	-7
	6	30	14	16	-10	-10
	7	35	12	13	-7	-8
1968 - 1997	1	15	10	10	-10	-10
	2	35	12	13	-8	-9
	3	35	11	12	-8	-9
	4	20	8	9	-7	-8
	5	50	8	9	-7	-8
	6	30	13	14	-10	-11
	7	40	9	10	-8	-8
1968 - 1997	1	–	–	–	–	–
	2	50	14	14	-10	-10
	3	60	14	14	-8	-8
	4	15	7	7	-6	-6
	5	55	8	8	-7	-7
	6	60	14	14	-10	-10
	7	100	12	12	-8	-8

https://doi.org/10.5194/essd-2025-306
Preprint. Discussion started: 5 August 2025

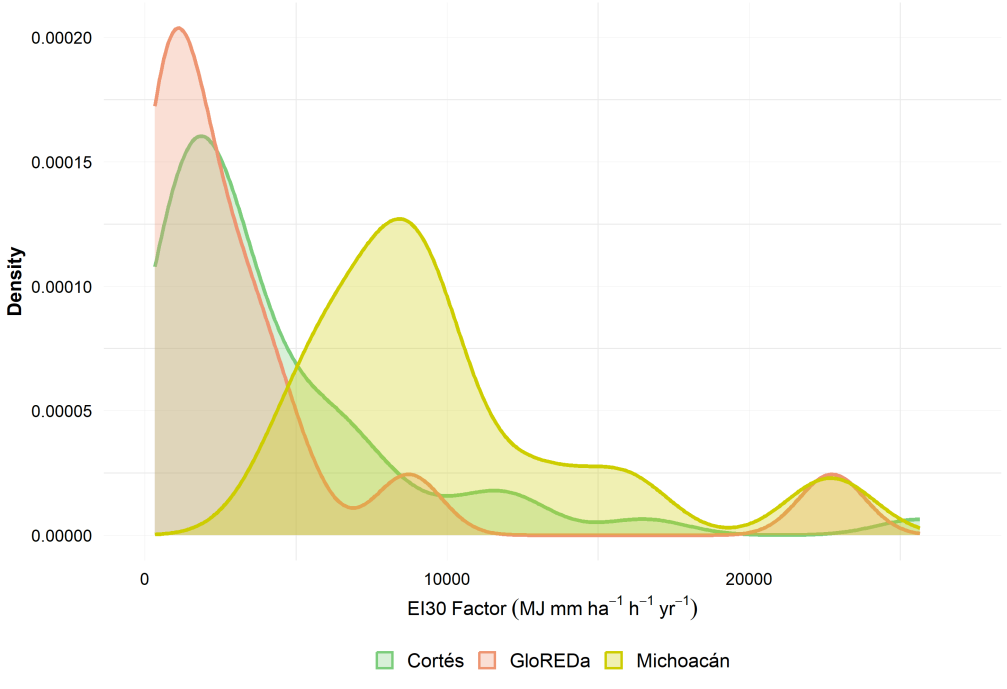
Open Access Earth System Science Data Discussions

Table A2. Homogenization parameters for daily rainfall time series by ecoregion and group. Eco: Ecoregion (2: North American Deserts, 3: Semi-arid Elevations, 5: Tropical Rain Forest, 6: Tropical Dry Forest, 7: Temperate Sierras); RS: Number of rainfall time series; inht: ; dz.max and dz.min (upper and lower): standard deviations to consider suspicious and anomalous data

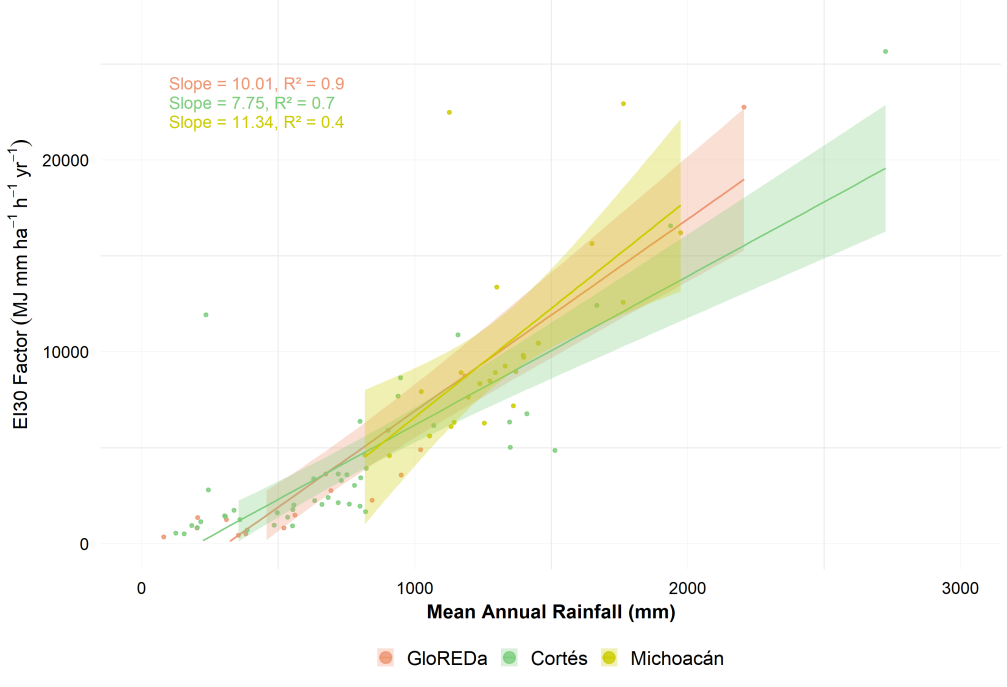
1968-1997								1978-2007								1988-2017							
Eco	Group	RS	inht	dz.max lower	dz.max upper	dz.min lower	dz.min upper	Eco	Group	RS	inht	dz.max lower	dz.max upper	dz.min lower	dz.min upper	Eco	Group	RS	inht	dz.max lower	dz.max upper	dz.min lower	dz.min upper
2	1	94	50	40	45	-20	-20	2	1	142	60	32	34	-18	-18	2	1	112	100	40	45	-20	-25
	2	18	25	40	45	-25	-30		2	23	20	45	50	-25	-25		2	12	10	35	40	-25	-30
	3	36	20	40	45	-25	-30		3	68	12	35	40	-30	-35		3	45	20	35	40	-20	-25
	4	35	20	35	40	-20	-25		4	40	12	30	35	-20	-25		4	67	50	45	40	-30	-35
3	1	34	25	24	26	-10	-12	3	1	41	30	22	24	-12	-14	3	1	115	60	26	28	-16	-16
	2	87	40	26	28	-14	-16		2	55	25	22	24	-14	-14		2	27	20	22	24	-14	-14
	3	59	40	22	24	-12	-12		3	130	50	24	26	-14	-14		3	180	40	26	28	-18	-18
	4	52	40	26	28	-14	-14		4	71	40	24	26	-12	-14	5	1	80	60	22	24	-20	-22
	5	16	40	26	28	-10	-12		5	16	30	22	24	-10	-12		2	81	40	24	26	-16	-18
5	1	36	20	30	35	-15	-20	5	1	42	40	30	35	-25	-30		3	75	100	24	26	-16	-18
	2	11	5	20	22	-10	-10		2	13	50	18	20	-10	-12	6	1	290	200	35	40	-25	-30
	3	36	50	24	26	-14	-14		3	30	40	18	20	-12	-14		2	46	20	30	35	-20	-25
	4	58	50	30	32	-14	-16		4	32	40	18	20	-16	-18		3	85	80	35	40	-30	-35
	5	27	40	20	22	-10	-12		5	68	50	22	24	-16	-18		4	33	15	40	45	-35	-40
	6	27	40	20	22	-12	-14		6	52	30	20	22	-12	-14	7	1	242	250	26	28	-14	-14
6	1	118	25	35	40	-30	-30	6	1	62	20	30	30	-20	-25		2	59	100	35	40	-25	-25
	2	187	60	40	45	-25	-30		2	267	80	30	35	-25	-30		3	114	200	30	35	-25	-30
	3	77	25	40	40	-20	-25		3	117	25	30	30	-25	-30								
	4	19	25	50	55	-30	-35		4	20	10	40	45	-35	-35								
7	1	82	50	22	24	-14	-16	7	1	87	60	20	22	-12	-14								
	2	49	50	35	35	-20	-25		2	62	60	35	40	-25	-30								
	3	153	50	35	35	-25	-25		3	157	80	26	28	-12	-14								
	4	116	50	24	26	-10	-12		4	105	70	26	28	-22	-24								

https://doi.org/10.5194/essd-2025-306
Preprint. Discussion started: 5 August 2025

Open Access Earth System Science Data Discussions



(a)



(b)

Figure A1. Characteristics of the three validations databases: GloREDa, Cortés, and Michoacán. a) Density plot of $EI30$ values; b) Linear relationship between $EI30$ and the mean annual rainfall

https://doi.org/10.5194/essd-2025-306
Preprint. Discussion started: 5 August 2025

Open Access Earth System Science Data Discussions

Table A3. Identification of number of rainfall time series with consecutive zero and NA values

Consecutive years	Number of RS with sequences of NA and zeros			Decision
	1968 - 1997	1978 - 2007	1988 - 2017	
1	794	787	611	Not changed
2	344	478	613	Not changed
3	155	286	277	Replaced with NA and used as reference
4	88	136	109	Replaced with NA and used as reference
5	77	78	82	Replaced with NA and used as reference
6	21	11	31	Replaced with NA and used as reference
7	1	1	0	Removed
8	2	0	1	Removed
9	1	2	2	Removed
10	2	0	0	Removed
11	3	2	1	Removed
12	0	0	0	Removed
13	0	1	1	Removed
14	0	1	0	Removed
15	0	0	0	Removed
16	1	2	0	Removed
RS (NA menor 20%)	1489	1785	1728	
RS removed	10	9	5	
RS for data gap-filling process	1479	1776	1723	

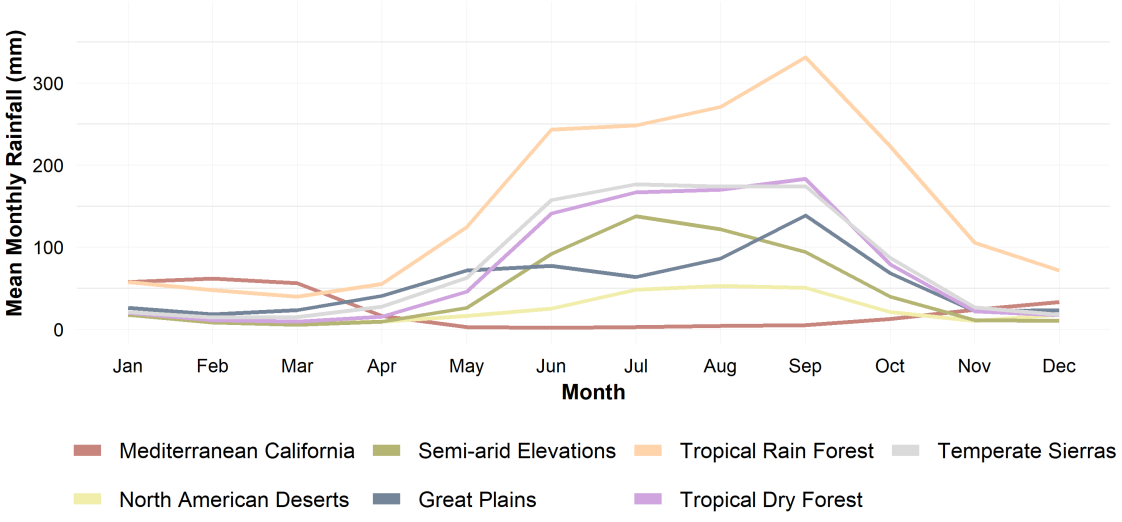
https://doi.org/10.5194/essd-2025-306
Preprint. Discussion started: 5 August 2025

Open Access Earth System Science Data Discussions

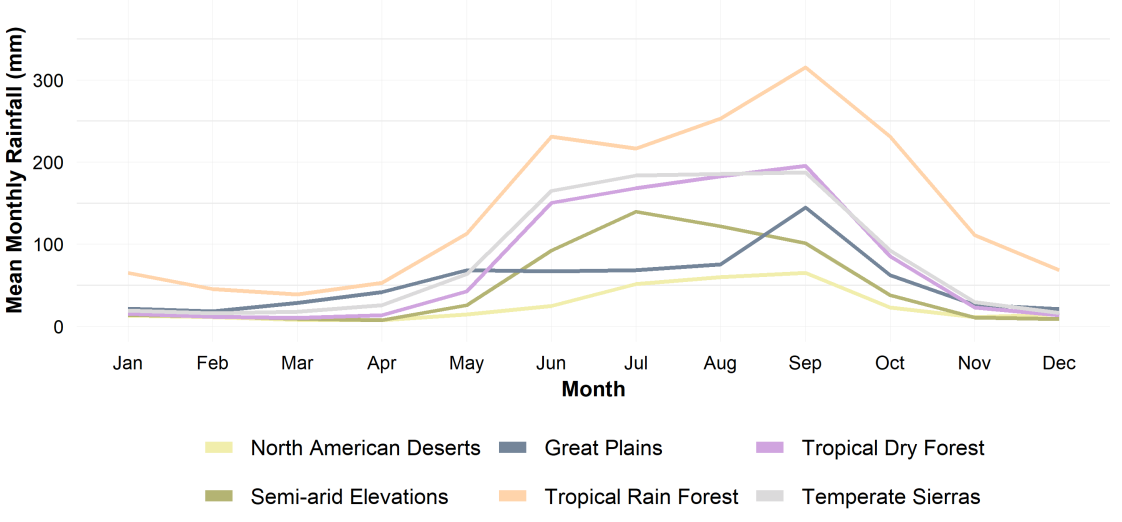
Table A4. Root Mean Square Error (mm) of the data gap-filling process by month and ecoregion. Eco: Ecoregion (1: Mediterranean California, 2: North American Deserts, 3: Semi-arid Elevations, 4: Great Plains, 5: Tropical Rain Forest, 6: Tropical Dry Forest, 7: Temperate Sierras)

Climate Normal	Eco	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sept	Oct	Nov	Dec	Total
	1	6.84	6.71	6.10	1.36	0.33	0.16	0.67	0.58	0.60	0.78	1.74	3.43	2.44
	2	1.91	1.39	1.02	1.31	1.37	2.20	6.32	3.42	3.55	2.84	1.11	2.41	2.40
	3	1.83	0.96	0.93	0.78	1.42	3.08	3.64	3.15	3.06	1.85	0.74	0.83	1.86
1968-1997	4	1.72	2.08	1.06	2.25	2.66	3.61	1.74	4.40	3.75	2.62	1.02	1.63	2.38
	5	3.07	3.57	4.28	3.20	6.06	7.92	8.39	8.45	9.79	8.32	5.66	4.78	**6.12**
	6	2.47	1.17	1.49	1.46	2.75	6.26	6.76	5.84	7.13	3.90	2.34	1.89	**3.62**
	7	2.31	1.07	1.53	2.06	2.75	4.55	5.11	5.00	8.03	3.93	2.11	2.06	3.38
	1	8.16	5.81	7.87	1.63	0.86	2.53	0.41	1.40	0.70	1.73	3.27	5.69	3.34
	2	2.13	1.13	0.86	1.38	1.73	1.98	5.26	3.95	3.84	2.10	1.06	2.28	2.31
	3	1.66	0.93	0.75	0.72	1.38	3.54	3.73	3.21	3.49	1.81	0.75	0.95	1.91
1978-2007	4	2.77	1.95	1.97	3.44	3.27	5.06	6.15	5.68	9.01	5.08	1.55	3.40	**4.11**
	5	4.04	3.82	3.15	3.02	6.01	7.71	6.34	7.07	8.66	9.14	5.22	4.51	**5.72**
	6	3.19	1.12	0.92	1.68	2.89	6.83	5.71	5.71	8.58	4.85	2.38	2.38	3.85
	7	2.37	1.31	1.17	2.35	4.08	5.61	6.34	5.54	6.92	6.92	2.72	2.27	3.97
	2	1.59	1.45	2.08	1.11	2.02	2.14	6.21	3.96	4.88	2.53	1.89	2.69	2.71
	3	1.67	1.68	1.23	0.77	1.41	3.23	4.29	4.56	3.21	1.99	1.05	1.11	2.18
	4	2.75	2.53	4.08	6.08	8.74	5.68	13.17	9.87	12.52	6.27	5.05	6.27	**6.92**
1988-2017	5	5.25	3.48	3.21	3.42	6.81	7.79	8.51	9.69	9.06	12.11	7.34	4.73	**6.78**
	6	3.56	1.68	1.57	2.38	3.48	6.35	6.55	6.78	8.28	5.42	2.41	2.36	4.24
	7	2.73	2.28	2.55	2.21	6.34	6.79	8.65	7.59	8.34	6.08	4.37	3.20	5.09
Total		3.10	2.31	2.39	2.13	3.32	4.65	**5.70**	**5.29**	**6.17**	4.51	2.69	2.94	**3.77**

https://doi.org/10.5194/essd-2025-306
Preprint. Discussion started: 5 August 2025



(a)



(b)

Figure A2. Mean monthly rainfall for all seven ecoregions. a) Climate normal (CN2) 1978-2007; and b) Climate normal (CN3) 1988-2017.

https://doi.org/10.5194/essd-2025-306
Preprint. Discussion started: 5 August 2025

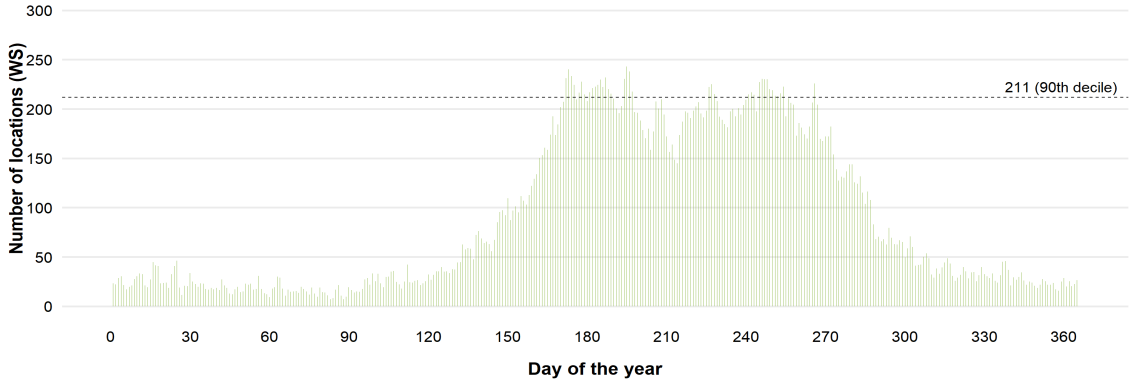
Open Access Earth System Science Data Discussions

Table A5. Metrics of performance of three models: Model I (Richardson et al., 1983), Model II (Liu et al., 2020), and Model III (Xie et al., 2016) in predicting the $EI30$ values of three databases: GloREDa, Cortés and Michoacán. ME: Mean error and RMSE: Root Mean Square Error

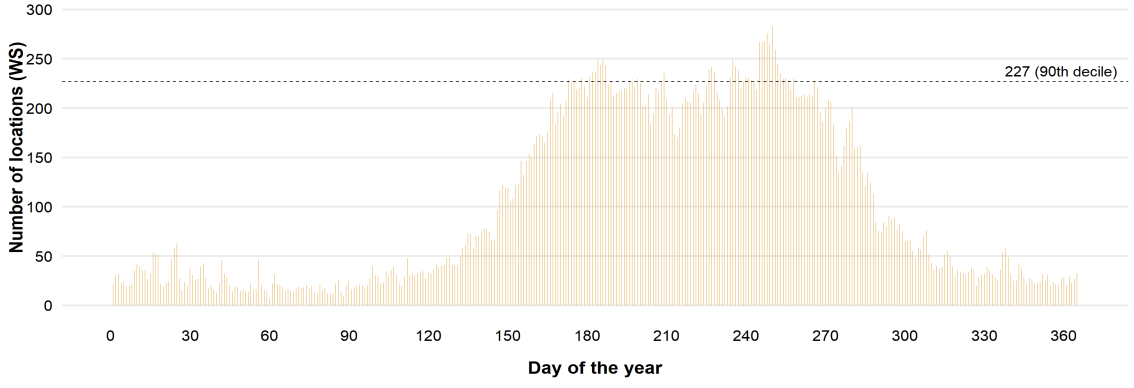
Databases	Models	Performance metrics MJ mm ha^{-1} h^{-1} yr^{-1}	
		ME	RMSE
GloREDa vs. Mexico-CN3	I	956	1524
	II	2077	6176
	III	-566	1921
Cortés vs. Mexico-CN1	I	324	2438
	II	910	4088
	III	-1171	2667
Cortés vs. Mexico-CN2	I	338	2628
	II	959	4554
	III	-1153	2900
Michoacán	I	-3699	4978
	II	-4434	5717
	III	-5328	6438

https://doi.org/10.5194/essd-2025-306
Preprint. Discussion started: 5 August 2025

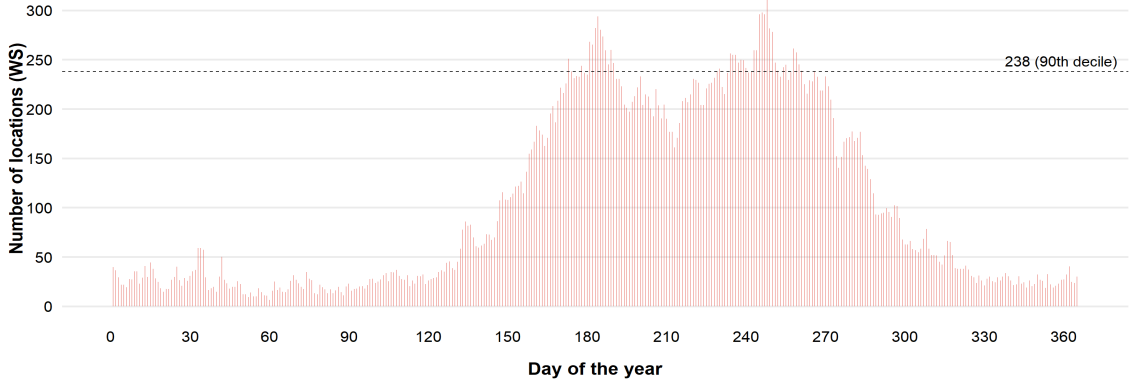
Open Access Earth System Science Data Discussions



(a)



(b)



(c)

Figure A3. Number of locations with erosive rainfall of each day of the year for the three climate normal, a) 1968-1997, b) 1978-2007, and c) 1988-2017

https://doi.org/10.5194/essd-2025-306
Preprint. Discussion started: 5 August 2025

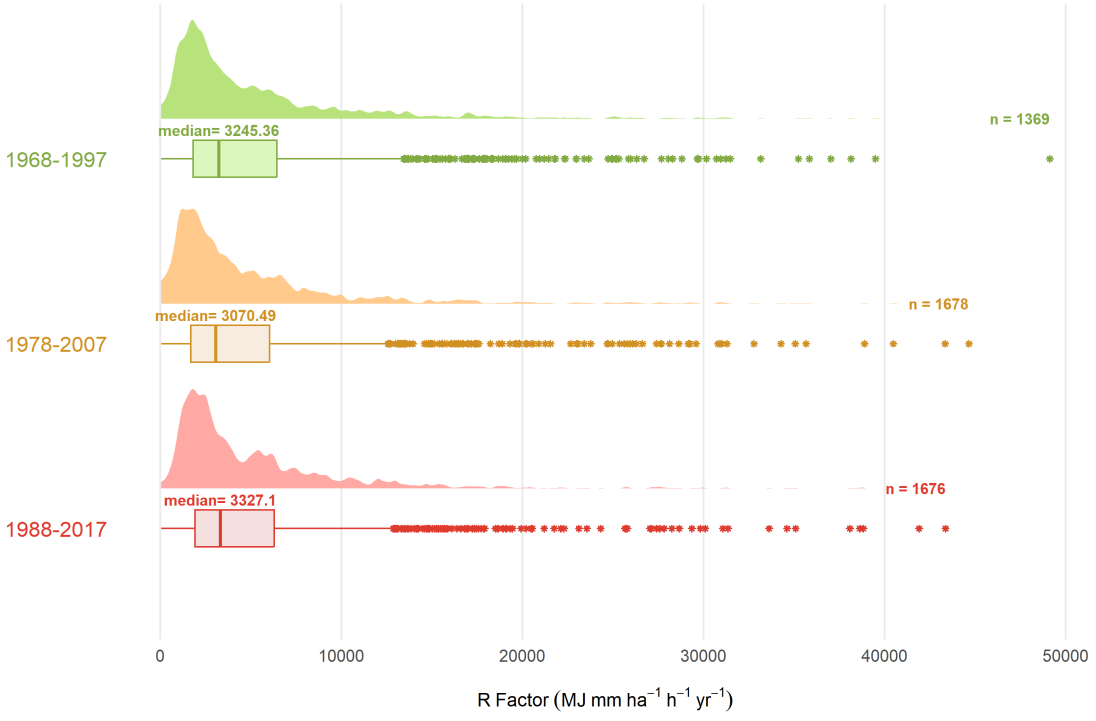


Figure A4. Density plot and Box-Plot of erosivity factor (R) calculated from daily rainfall time series for the three climate normals

https://doi.org/10.5194/essd-2025-306
Preprint. Discussion started: 5 August 2025

Open Access Earth System Science Data Discussions

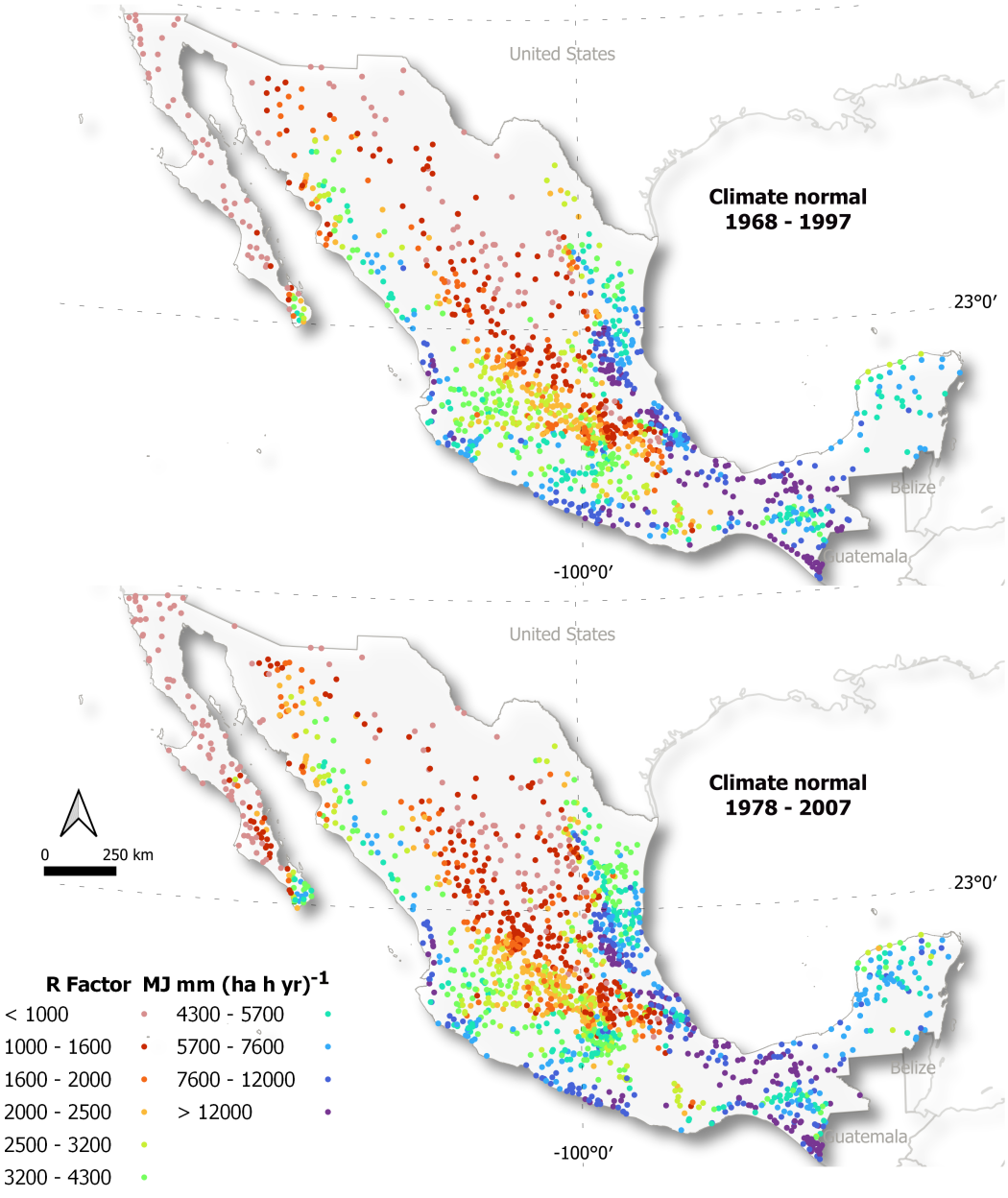


Figure A5. *R* factor values calculated with the power law equation proposed by (Richardson et al., 1983) at daily resolution for the climate normal 1968 - 1997 and 1978-2007.

https://doi.org/10.5194/essd-2025-306
Preprint. Discussion started: 5 August 2025

Author contributions. VMVR, DAGL, and MG contributed with conceptualization, formal analysis, Methodology, and visualization. VMVR, DAGL, and CEAC contributed to the data curation and writing - original draft preparation. CEAC, AGT, BLPP, DLL, RRGLL, and MG contributed to Project administration, writing, review, and editing. VMVR, BLPP, and MG contributed to funding acquisition.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. The authors want to thank to the National Meteorological Service (SMN by its initials in Spanish) and National Water Commission (CONAGUA by its initials in Spanish) for making public available the national climate data. VMVR, BLPP, and MG acknowledges support from grant CF-2023-I-1846 of the National Council of Humanities, Sciences, and Technologies; abbreviated CONAHCYT. Mario Guevara acknowledges support from grant IGCP-765-UNESCO.

https://doi.org/10.5194/essd-2025-306
Preprint. Discussion started: 5 August 2025

References

Adeyeri, O., Laux, P., Ishola, K., Zhou, W., Balogun, I., Adeyewa, Z., and Kunstmann, H.: Homogenising meteorological variables: Impact on trends and associated climate indices, Journal of Hydrology, 607, 127 585, https://doi.org/https://doi.org/10.1016/j.jhydrol.2022.127585, 2022.

Agustín Breña-Naranjo, J., Pedrozo-Acuña, A., Pozos-Estrada, O., Jiménez-López, S. A., and López-López, M. R.: The contribution of tropical cyclones to rainfall in Mexico, Physics and Chemistry of the Earth, Parts A/B/C, 83-84, 111–122, https://doi.org/https://doi.org/10.1016/j.pce.2015.05.011, emerging science and applications with microwave remote sensing data, 2015.

Alexandersson, H.: A homogeneity test applied to precipitation data, Journal of Climatology, 6, 661–675, https://doi.org/https://doi.org/10.1002/joc.3370060607, 1986.

Beck, H. E., Zimmermann, N. E., McVicar, T. R., Vergopolan, N., Berg, A., and Wood, E. F.: Present and future Köppen-Geiger climate classification maps at 1-km resolution, Scientific Data, 5, 180 214, https://doi.org/10.1038/sdata.2018.214, 2018.

Benites, E. T., Becerra, J. C., Gil, J. U., Cedillo, L. T., and Torres, P. S. R.: Predicción de la erosión hídrica en la cuenca del Cañón del Sumidero, Chiapas, Revista Mexicana de Ciencias Agrícolas, 11, 1903–1915, https://doi.org/10.29312/remexca.v11i8.2747, epub 13 de diciembre de 2021, 2020.

Bessenbacher, V., Schumacher, D. L., Hirschi, M., Seneviratne, S. I., and Gudmundsson, L.: Gap-Filled Multivariate Observations of Global Land–Climate Interactions, Journal of Geophysical Research: Atmospheres, 128, e2023JD039 099, https://doi.org/https://doi.org/10.1029/2023JD039099, e2023JD039099 2023JD039099, 2023.

Bezak, N., Borrelli, P., and Panagos, P.: Exploring the possible role of satellite-based rainfall data in estimating inter- and intra-annual global rainfall erosivity, Hydrology and Earth System Sciences, 26, 1907 – 1924, https://doi.org/10.5194/hess-26-1907-2022, cited by: 26; All Open Access, Gold Open Access, Green Open Access, 2022.

Bolaños González, M. A., Paz Pellat, F., Cruz Gaistardo, C. O., Argumedo Espinoza, J. A., Romero Benítez, V. M., and de la Cruz Cabrera, J. C.: Mapa de erosión de los suelos de México y posibles implicaciones en el almacenamiento de carbono orgánico del suelo, Terra Latinoamericana, 34, 271–288, http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0187-57792016000300271&lng=es&tlng=es, recuperado en 03 de septiembre de 2024, 2016.

Borrelli, P., Robinson, D., Fleischer, L. R., Lugato, E., Ballabio, C., Alewell, C., Meusburger, M., Modugno, P., Schütt, M., Tenuta, K. E., and Panagos, P.: Land use and climate change impacts on global soil erosion by water (2015-2070), Proceedings of the National Academy of Sciences, 117, 21 994–22 001, https://doi.org/10.1073/pnas.2001403117, 2020.

Cardoso, D. P., Silva, E. M., Avanzi, J. C., Muniz, J. A., Ferreira, D. F., Silva, M. L. N., Acuña-Guzman, S. F., and Curi, N.: RainfallErosivityFactor: An R package for rainfall erosivity (R-factor) determination, CATENA, 189, 104 509, https://doi.org/https://doi.org/10.1016/j.catena.2020.104509, 2020.

Carrera, J. J., Levresse, G. P., and Hernández-Espriú, J. A.: Geostatistical Analysis of Yearly Precipitation at the National Level: Stratification and Anisotropy Considerations in Mexico, pre-print available at SSRN: https://ssrn.com/abstract=4820011 or http://dx.doi.org/10.2139/ssrn.4820011, 2024.

Carrera-Hernández, J. J.: Mexico's High Resolution Climate Database (MexHiResClimDB): a new daily high-resolution gridded climate dataset for Mexico covering 1951–2020, Earth System Science Data Discussions, 2025, 1–30, https://doi.org/10.5194/essd-2025-100, 2025.

https://doi.org/10.5194/essd-2025-306
Preprint. Discussion started: 5 August 2025

Cespedes, J. M. N., Anguiano, J. H. H., Concepción, P. C. A., Martínez, J. L. M., Aguilera, G. C., and Padilla, F.: A comparison of missing value imputation methods applied to daily precipitation in a semi-arid and a humid region of Mexico, Atmosfera, 37, 33 – 52, https://doi.org/10.20937/ATM.53095, cited by: 3; All Open Access, Green Open Access, Hybrid Gold Open Access, 2023.

Chen, Y., Xu, M., Wang, Z., Chen, W., and Lai, C.: Reexamination of the Xie model and spatiotemporal variability in rainfall erosivity in mainland China from 1960 to 2018, CATENA, 195, 104 837, https://doi.org/https://doi.org/10.1016/j.catena.2020.104837, 2020.

Commission for Environmental Cooperation: Ecological Regions of North America -Toward a Common Perspective, Commission for Environmental Cooperation, Montreal, Canada, ISBN 2-922305-20-1, 1997.

Cortés, H. G.: Caracterización de la erosividad de la lluvia en México utilizando métodos multivariados, Master's thesis, Colegio de Posgraduados, Montecillo, Mexico, centro de Edafología, 1991.

Cuervo-Robayo, A. P., Téllez-Valdés, O., Gómez-Albores, M. A., Venegas-Barrera, C. S., Manjarrez, J., and Martínez-Meyer, E.: An update of high-resolution monthly climate surfaces for Mexico, International Journal of Climatology, 34, 2427 – 2437, https://doi.org/10.1002/joc.3848, cited by: 153, 2014.

Cuervo-Robayo, A. P., Ureta, C., Gómez-Albores, M. A., Meneses-Mosquera, A. K., Téllez-Valdés, O., and Martínez-Meyer, E.: One hundred years of climate change in Mexico, PLOS ONE, 15, 1–19, https://doi.org/10.1371/journal.pone.0209808, 2020.

Das, S., Jain, M. K., Gupta, V., McGehee, R. P., Yin, S., de Mello, C. R., Azari, M., Borrelli, P., and Panagos, P.: GloRESatE: A dataset for global rainfall erosivity derived from multi-source data, Scientific Data, 11, 926, https://doi.org/10.1038/s41597-024-03756-5, 2024.

de Anda Sánchez, J.: Precipitation in Mexico, pp. 1–14, Springer International Publishing, Cham, ISBN 978-3-030-40686-8, https://doi.org/10.1007/978-3-030-40686-8_1, 2020.

Duarte, L. V., Formiga, K. T. M., and Costa, V. A. F.: Comparison of Methods for Filling Daily and Monthly Rainfall Missing Data: Statistical Models or Imputation of Satellite Retrievals?, Water, 14, https://doi.org/10.3390/w14193144, 2022.

Dunkerley, D. L.: Rainfall intensity bursts and the erosion of soils: an analysis highlighting the need for high temporal resolution rainfall data for research under current and future climates, Earth Surface Dynamics, 7, 345–360, https://doi.org/10.5194/esurf-7-345-2019, 2019.

Efthimiou, N.: Evaluating the performance of different empirical rainfall erosivity (R) factor formulas using sediment yield measurements, CATENA, 169, 195–208, https://doi.org/https://doi.org/10.1016/j.catena.2018.05.037, 2018.

Feng, Z., Zhang, Z., Zuo, Y., Wan, X., Wang, L., Chen, H., Xiong, G., Liu, Y., Tang, Q., and Liang, T.: Analysis of long term water quality variations driven by multiple factors in a typical basin of Beijing-Tianjin-Hebei region combined with neural networks, Journal of Cleaner Production, 382, 135 367, https://doi.org/https://doi.org/10.1016/j.jclepro.2022.135367, 2023.

Fenta, A. A., Tsunekawa, A., Haregeweyn, N., Yasuda, H., Tsubo, M., Borrelli, P., Kawai, T., Sewale Belay, A., Ebabu, K., Liyew Berihun, M., Sultan, D., Asamin Setargie, T., Elnashar, A., and Panagos, P.: Improving satellite-based global rainfall erosivity estimates through merging with gauge data, Journal of Hydrology, 620, https://doi.org/10.1016/j.jhydrol.2023.129555, cited by: 15; All Open Access, Hybrid Gold Open Access, 2023.

Fransiska, H., Agustina, D., Setyorini, D., Sumartajaya, I. M., and Kurnia, A.: Time Series Clustering Analysis Using Dynamic Time Warping Technique of Daily Rainfall in Bengkulu Province, IOP Conference Series: Earth and Environmental Science, 1359, 012 026, https://doi.org/10.1088/1755-1315/1359/1/012026, 2024.

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A., and Michaelsen, J.: The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes, Scientific Data, 2, 150 066, https://doi.org/10.1038/sdata.2015.66, 2015.

García, E.: Climas (clasificación de Köppen, modificado por García), Map, scale 1:1,000,000, méxico, 1998.

https://doi.org/10.5194/essd-2025-306
Preprint. Discussion started: 5 August 2025

García-Cueto, O. R., Santillán-Soto, N., López-Velázquez, E., Reyes-López, J., Cruz-Sotelo, S., and Ojeda-Benítez, S.: Trends of climate change indices in some Mexican cities from 1980 to 2010, Theoretical and Applied Climatology, 137, 775–790, https://doi.org/10.1007/s00704-018-2620-4, 2019.

González, O. N., Serrano, J. I. B., Vílchez, F. F., Núñez, R. M. M., and García-Sancho, A. G.: Riesgo de erosión hídrica y estimación de pérdida de suelo en paisajes geomorfológicos volcánicos en México, Cultivos Tropicales, 37, 45–55, http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S0258-59362016000200006&lng=es&tlng=es, recuperado en 02 de septiembre de 2024, 2016.

Guijarro, J. A.: Quality control and homogenization of climatological series, https://doi.org/10.1201/b15625, cited by: 5, 2014.

Guijarro, J. A.: climatol: Climate Tools (Series Homogenization and Derived Products), https://CRAN.R-project.org/package=climatol, r package version 4.1.0, 2024.

Gómez-Latorre, D. A., Araujo-Carrillo, G. A., and Leguizamón, Y. R.: Regionalización de patrones de lluvias para períodos multianuales secos y húmedos en el Altiplano Cundiboyacense de Colombia, Revista de Climatología, 22, 162–177, https://rclimatol.eu/2022/12/31/regionalizacion-de-patrones-de-lluvias-para-periodos-multianuales-secos-y-humedos-en-el-altiplano-cundiboyacense-de-colombia/, 2022.

Hartigan, J. A.: Clustering algorithms, John Wiley & Sons, Inc., 1975.

Hatfield, J. L., Sauer, T. J., and Cruse, R. M.: Chapter One - Soil: The Forgotten Piece of the Water, Food, Energy Nexus, Advances in Agronomy, 143, 1–46, https://doi.org/https://doi.org/10.1016/bs.agron.2017.02.001, 2017.

Hırca, T. and Türkkan, G. E.: Assessment of Different Methods for Estimation of Missing Rainfall Data, Water Resources Management, https://doi.org/10.1007/s11269-024-03936-3, published online on 2024/07/31, 2024.

INEGI-CONABIO-INE: Ecorregiones Terrestres de México, Tech. rep., Instituto Nacional de Estadística, Geografía e Informática (INEGI) and Comisión Nacional para el Conocimiento y Uso de la Biodiversidad (CONABIO) and Instituto Nacional de Ecología (INE), México, escala 1:1000000, 2008.

Jiménez Espinosa, M., Baeza Ramírez, C., Matías Ramírez, L. G., and Eslava Morales, H.: Mapas de Índices de Riesgo a Escala Municipal por Fenómenos Hidrometeorológicos, Informe técnico, Sistema Nacional de Protección Civil, Centro Nacional de Prevención de Desastres (CENAPRED), México, http://www.atlasnacionalderiesgos.gob.mx/descargas/Metodologias/Hidrometeorologico.pdf, subdirección de Riesgos Hidrometeorológicos, 2012.

Karami, A., Homaee, M., Neyshabouri, M. R., Afzalinia, S., and Basirat, S.: Large scale evaluation of single storm and short/long term erosivity index models, Turkish Journal of Agriculture and Forestry, 36, 207 – 216, https://doi.org/10.3906/tar-1102-24, cited by: 9; All Open Access, Bronze Open Access, 2012.

Ke, Q. and Zhang, K.: Patterns of runoff and erosion on bare slopes in different climate zones, CATENA, 198, 105 069, https://doi.org/https://doi.org/10.1016/j.catena.2020.105069, 2021.

Li, J., Sun, R., and Chen, L.: Assessing the accuracy of large-scale rainfall erosivity estimation based on climate zones and rainfall patterns, CATENA, 217, 106 508, https://doi.org/https://doi.org/10.1016/j.catena.2022.106508, 2022.

Liu, Y., Zhao, W., Liu, Y., and Pereira, P.: Global rainfall erosivity changes between 1980 and 2017 based on an erosivity model using daily precipitation data, CATENA, 194, 104 768, https://doi.org/https://doi.org/10.1016/j.catena.2020.104768, 2020.

Lupi, A., Luppichini, M., Barsanti, M., Bini, M., and Giannecchini, R.: Machine learning models to complete rainfall time series databases affected by missing or anomalous data, Earth Science Informatics, 16, 3717–3728, https://doi.org/10.1007/s12145-023-01122-4, 2023.

https://doi.org/10.5194/essd-2025-306
Preprint. Discussion started: 5 August 2025

Mateos, E., Santana, J.-S., Montero-Martínez, M. J., Deeb, A., and Grunwaldt, A.: Possible climate change evidence in ten Mexican watersheds, Physics and Chemistry of the Earth, Parts A/B/C, 91, 10–19, https://doi.org/https://doi.org/10.1016/j.pce.2015.08.009, appropriate Technology and Climate Change Adaptation for Water Resources Management, 2016.

McCuen, R. H.: Hydrologic Analysis and Design, Pearson, 4th edition edn., ISBN 978-0134313122, 2016.

McKinnon, K. A.: Discussion on "A combined estimate of global temperature", Environmetrics, 33, e2721, https://doi.org/https://doi.org/10.1002/env.2721, 2022.

Meng, X., Zhu, Y., Yin, M., and Liu, D.: The impact of land use and rainfall patterns on the soil loss of the hillslope, Scientific Reports, 11, 16 341, https://doi.org/10.1038/s41598-021-95819-5, 2021.

Mikhailova, E., Bryant, R., Schwager, S., and Smith, S.: Predicting rainfall erosivity in Honduras, Soil Science Society of America Journal, 61, 273 – 279, https://doi.org/10.2136/sssaj1997.03615995006100010039x, cited by: 85, 1997.

Nearing, M. A.: Soil Erosion and Conservation, p. Chapter page numbers if available, Wiley, 2nd edn., https://digitalcommons.unl.edu/usdaarsfacpub/1290/, 2013.

Nearing, M. A., qing Yin, S., Borrelli, P., and Polyakov, V. O.: Rainfall erosivity: An historical review, CATENA, 157, 357–362, https://doi.org/https://doi.org/10.1016/j.catena.2017.06.004, 2017.

Panagos, P., Borrelli, P., Meusburger, K., Yu, B., Klik, A., Jae Lim, K., Yang, J. E., Ni, J., Miao, C., Chattopadhyay, N., Sadeghi, S. H., Hazbavi, Z., Zabihi, M., Larionov, G. A., Krasnov, S. F., Gorobets, A. V., Levi, Y., Erpul, G., Birkel, C., Hoyos, N., Naipal, V., Oliveira, P. T. S., Bonilla, C. A., Meddi, M., Nel, W., Al Dashti, H., Boni, M., Diodato, N., Van Oost, K., Nearing, M., and Ballabio, C.: Global rainfall erosivity assessment based on high-temporal resolution rainfall records, Scientific Reports, 7, 4175, https://doi.org/10.1038/s41598-017-04282-8, 2017.

Panagos, P., Hengl, T., Wheeler, I., Marcinkowski, P., Rukeza, M. B., Yu, B., Yang, J. E., Miao, C., Chattopadhyay, N., Sadeghi, S. H., Levi, Y., Erpul, G., Birkel, C., Hoyos, N., Oliveira, P. T. S., Bonilla, C. A., Nel, W., Al Dashti, H., Bezak, N., Van Oost, K., Petan, S., Fenta, A. A., Haregeweyn, N., Pérez-Bidegain, M., Liakos, L., Ballabio, C., and Borrelli, P.: Global rainfall erosivity database (GloREDa) and monthly R-factor data at 1 km spatial resolution, Data in Brief, 50, 109 482, https://doi.org/https://doi.org/10.1016/j.dib.2023.109482, 2023.

Paulhus, J. L. H. and Kohler, M. A.: Interpolation of missing precipitation records, Monthly Weather Review, 80, 129–133, https://doi.org/10.1175/1520-0493(1952)080<0129:IOMPR>2.0.CO;2, 1952.

Pennock, D.: Soil erosion: the greatest challenge for sustainable soil management, FAO, Rome, Italy, ISBN 978-92-5-131426-5, https://openknowledge.fao.org/items/6c070e1e-6533-4b7e-ba5f-a2f21a0e59ff, 2019.

Pineda-Martínez, L. F. and Carbajal, N.: Climatic analysis linked to land vegetation cover of Mexico by applying multivariate statistical and clustering analysis, Atmósfera, 30, 233–242, https://doi.org/10.20937/atm.2017.30.03.04, 2017.

Porrúa, F. E., Hidalgo, J. Z., Arroyo, A. M., Raga, G., and García, C. G.: Estado y perspectivas del Cambio Climático en México: un punto de partida, Tech. rep., Universidad Nacional Autónoma de México, Mexico, ISBN 978-607-30-8172-6, https://cambioclimatico.unam.mx/estado-y-perspectivas-del-cambio-climatico-en-mexico/, 2020.

R Core Team: R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, https://www.R-project.org/, 2022.

Renard, K. G. and Freimund, J. R.: Using monthly precipitation data to estimate the R-factor in the revised USLE, Journal of Hydrology, 157, 287–306, https://api.semanticscholar.org/CorpusID:5030732, 1994.

https://doi.org/10.5194/essd-2025-306
Preprint. Discussion started: 5 August 2025

Richardson, C. W., Foster, G. R., and Wright, D.: Estimation of Erosion Index from Daily Rainfall Amount, Transactions of the ASABE, 26, 153–0156, https://api.semanticscholar.org/CorpusID:108883838, 1983.

Rohlf, F. J.: Methods of comparing classifications, Annual Review of Ecology and Systematics, 5, 101–113, 1974.

Rutebuka, J., De Taeye, S., Kagabo, D., and Verdoodt, A.: Calibration and validation of rainfall erosivity estimators for application in Rwanda, Catena, 190, https://doi.org/10.1016/j.catena.2020.104538, cited by: 19, 2020.

Shin, J.-Y., Kim, T., Heo, J.-H., and Lee, J.-H.: Spatial and temporal variations in rainfall erosivity and erosivity density in South Korea, CATENA, 176, 125–144, https://doi.org/https://doi.org/10.1016/j.catena.2019.01.005, 2019.

Soto, V. and Delgado-Granados, H.: Occurrence and characteristics of snowfall on the highest mountain of Mexico (Citlaltépetl volcano) through the ground's surface temperature, Atmósfera, 38, 35–54, https://doi.org/10.20937/ATM.53204, 2023.

Todeschini, R., Ballabio, D., Termopoli, V., and Consonni, V.: Extended multivariate comparison of 68 cluster validity indices. A review, Chemometrics and Intelligent Laboratory Systems, p. 105117, https://doi.org/https://doi.org/10.1016/j.chemolab.2024.105117, 2024.

Tu, A., Xie, S., Li, Y., Liu, Z., and Shen, F.: Effect of fixed time interval of rainfall data on calculation of rainfall erosivity in the humid area of south China, CATENA, 220, 106 714, https://doi.org/https://doi.org/10.1016/j.catena.2022.106714, 2023.

Vantas, K., Sidiropoulos, E., and Evangelides, C.: Rainfall Erosivity and Its Estimation: Conventional and Machine Learning Methods, in: Soil Erosion, edited by Hrissanthou, V. and Kaffas, K., chap. 2, IntechOpen, Rijeka, https://doi.org/10.5772/intechopen.85937, 2019.

Varón-Ramírez, V., Gómez-Latorre, D., Arroyo-Cruz, C., Guevara Santamaría, M., and Prado Pano, B.: Daily rainfall series and rainfall erosivity in Mexico for three climatic normals (1968-1997, 1978-2007, and 1988-2017) ver 2, https://doi.org/10.6073/pasta/e0dc8bd3501f8c19bb750e853c3289cb, accessed 2025-05-20, 2025.

Varón-Ramírez, V. M.: VimiVaron/Rainfall-Erosivity-Mexico: Rainfall- Erosivity-Mexico, https://doi.org/10.5281/zenodo.15468097, 2025.

Varón-Ramírez, V. M. and Guevara, M.: Advances in the study of soil erosion by water in Mexico published in Spanish, European Journal of Soil Science, 75, e13 458, https://doi.org/https://doi.org/10.1111/ejss.13458, 2024.

Vaughan, C., Buja, L., Kruczkiewicz, A., and Goddard, L.: Identifying research priorities to advance climate services, Climate Services, 4, 65–74, https://doi.org/https://doi.org/10.1016/j.cliser.2016.11.004, 2016.

Wischmeier, W. H. and Smith, D. D.: Rainfall energy and its relationship to soil loss, Eos, Transactions American Geophysical Union, 39, 285–291, https://doi.org/https://doi.org/10.1029/TR039i002p00285, 1958.

Wischmeier, W. H. and Smith, D. D.: Predicting rainfall erosion losses : a guide to conservation planning, https://api.semanticscholar.org/CorpusID:129088976, 1978.

WMO: WMO Guidelines on the Calculation of Climate Normals, https://library.wmo.int/es/records/item/55797-wmo-guidelines-on-the-calculation-of-climate-normals, 2017.

WMO: Guidelines on Homogenization. WMO-No. 1245, https://library.wmo.int/idurl/4/57130, 2020.

WMO: Guide to Climatological Practices. WMO-No. 100, https://library.wmo.int/idurl/4/60113, 2023.

World Meteorological Organization: Guidelines on Homogenization, https://dgm-meteo.github.io/wmo-clino/Docs/WMO2020_1245_en_Guidelines%20on%20Homogenization.pdf, wMO-No. 1245, 2020.

Xie, Y., qing Yin, S., yuan Liu, B., Nearing, M. A., and Zhao, Y.: Models for estimating daily rainfall erosivity in China, Journal of Hydrology, 535, 547–558, https://doi.org/https://doi.org/10.1016/j.jhydrol.2016.02.020, 2016.

Yan, Z., Li, Z., and Xia, J.: Homogenization of climate series: The basis for assessing climate changes, Science China Earth Sciences, 57, 2891–2900, https://doi.org/10.1007/s11430-014-4945-x, 2014.

https://doi.org/10.5194/essd-2025-306
Preprint. Discussion started: 5 August 2025

Yin, S., Xie, Y., Liu, B., and Nearing, M. A.: Rainfall erosivity estimation based on rainfall data collected over a range of temporal resolutions, Hydrology and Earth System Sciences, 19, 4113–4126, https://doi.org/10.5194/hess-19-4113-2015, 2015.

Yin, S., Nearing, M. A., Borrelli, P., and Xue, X.: Rainfall Erosivity: An Overview of Methodologies and Applications, Vadose Zone Journal, 16, vzj2017.06.0131, https://doi.org/https://doi.org/10.2136/vzj2017.06.0131, 2017.

Yozgatligil, C., Aslan, S., Iyigun, C., and Batmaz, I.: Comparison of missing value imputation methods in time series: the case of Turkish meteorological data, Theoretical and Applied Climatology, 112, 143–167, https://doi.org/10.1007/s00704-012-0723-x, 2013.

Yu, B. and Rosewell, C. J.: Rainfall erosivity estimation using daily rainfall amounts for South Australia, Soil Research, 34, 721–733, https://api.semanticscholar.org/CorpusID:128543482, 1996.

Zhao, B., Zhang, L., Xia, Z., Xu, W., Xia, L., Liang, Y., and Xia, D.: Effects of Rainfall Intensity and Vegetation Cover on Erosion Characteristics of a Soil Containing Rock Fragments Slope, Advances in Civil Engineering, 2019, 7043428, https://doi.org/https://doi.org/10.1155/2019/7043428, 2019.