The ~~first rainfall erosivity database in~~ Rainfall and Erosivity Database for Mexico ~~: facing challenges of leveraging legacy climate data~~(1968-2017)

Viviana Marcela Varón-Ramírez[1,2], Douglas A. Gómez-Latorre[3], Carlos Eduardo Arroyo-Cruz[1,2], Alberto Gómez-Tagle[4], Blanca Lucía Prado Pano[5], Ronald R. Gutierrez[6], Deyanira Lobo-Luján[7], and Mario Guevara[1]

[1]Instituto de Geociencias, Universidad Nacional Autónoma de México, Campus UNAM Juriquilla, Blvd Juriquilla 3001, Juriquilla, Querétaro, CP 76230, México

[2]Posgrado en Ciencias de la Tierra, Instituto de Geociencias, Universidad Nacional Autónoma de México, Campus UNAM Juriquilla, Blvd Juriquilla 3001, Juriquilla, Querétaro, CP 76230, México.

[3]Corporación Colombiana de Investigación Agropecuaria - AGROSAVIA, Centro de Investigación Tibaitatá, Mosquera-Cundinamarca, Colombia

[4]Instituto de Investigaciones sobre los Recursos Naturales - INIRENA, Universidad Michoacana de San Nicolás de Hidalgo. Morelia, Mexico

[5]Instituto de Geología - Universidad Nacional Autónoma de México. Campus UNAM Ciudad Universitaria. Ciudad de Mexico. Mexico.

[6]Centro de Investigación y Tecnología del Agua (CITA), Universidad de Ingeniería y Tecnología (UTEC), Lima, Perú

[7]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)~~ worldwide. Rainfall erosivity (also known as the R factor) quantifies the potential 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~~ soil water erosion and is regarded as a dominant factor determining it. This contribution presents the first annual R factor database across the continental Mexico 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~~(1968–1997, 1978–2007, and 1988–2017). 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 controland assurance~~comprised three main steps. Firstly, a harmonaized daily rainfall time series dataset was obtained by compiling and harmonizing (through quality control, homogeneity analysis ~~(using the normal homogeneity test), and the data~~ and gap-filling~~process (using the proportion method) . Then, we tested~~) 5,410 raw rainfall time series distributed across Mexico. Secondly, 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~~ were tested to estimate daily R factors using three validation databases ~~(~~at global, national, and local scales~~). Finally, we estimated the annual rainfall erosivity~~ . Thirdly, a validated and continuously distributed annual R factor for all three ~~CNs with multiple combinations of α~~ climate normal was obtained. Thus, the database presented herein encompasses 1,369 (climate normal 1968–1997) , 1,678 (climate normal 1978–2007), 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 **?** 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~~ 1,676 (climate normal 1988–2017) rainfall time series and the corresponding R factor. Our results indicate that the median values of the R factor for the three climate normals were 3,245; 3,070; and 3,327 MJ mm ha^{-1} h^{-1} yr^{-1}, respectively. The statistical distribution of the ~~erosivity values was~~ R factor is right-skewed for the three ~~CNs~~climate normals, with high erosivity values reaching >12000 MJ mm ha^{-1} h^{-1} yr^{-1} in all ~~three CNs. The behavior throughout the year of the rainfall erosivity~~ cases. The R factor across Mexico showed a strong ecoregional differentiation, with the highest values concentrated in the Tropical Rain Forest and Tropical Dry Forest ecoregions, whereas the North American Deserts and Mediterranean California exhibited the lowest values. The annual behavior of the R factor was similar for the three ~~CNs~~climate normals. 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~~ annual R factor. By identifying areas with higher susceptibility to soil loss due to rainfall action and providing ~~a more spatially dense~~ spatially distributed and well-documented ~~rainfall erosivity database . Following the FAIR principles (Findability, Availability, Interoperability, and Reproducibility) for scientific data, this~~ estimates, we believe that the present R factor database has the potential to support the study of soil water erosion in the continental Mexico at country scale. Said database is available from a ~~scholarly accepted repository~~ scholarly-accepted repository at https://doi.org/10.6073/pasta/~~e0de8bd3501f8e19bb750e853e3289eb (**?**)~~ dd2b30e28ee25ff2d60d8a9f436951d2 (**?**) for public consultation.

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

1

Soil water erosion (SWE) refers to the soil displacement from its original location due to water action, such as rainfall, overland flow, and irrigation (**?**). SWE represents the dominant soil degradation issue at the global scale because it affects nearly 33% of the World's surface (**?**). ~~The impact of SWE is not just~~ SWE has both on-site ~~but also has~~ and off-site effects~~on distant locations~~. On-site, soil loses its natural ~~fertility and~~ capacity to store water, nutrients, and organic carbon (**?**), affecting, for instance, 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 ~~(**?**)~~(**???**). These widespread impacts underscore the urgent need to study SWE at national scales to guide effective land and water management~~.~~, especially in tropical countries (??).

~~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 (?).~~ Rainfall erosivity is the potential of rainfall to cause SWE (?). ~~Furthermore,~~ Research evidence shows that rainfall erosivity is the ~~first~~ dominant factor influencing SWE~~and is crucial for land and water conservation planning. Rainfall erosivity could increase its harmful effects on Mexican soils because, as~~. As extreme rainfall events are expected to increase in ~~tropical zones due to climate change, SWE will likely increase as well (?). Furthermore, in Mexico~~ frequency and intensity in tropical zones, rainfall erosivity is also expected to rise (?). Mexico is exhibiting this pattern as the temporal distribution of rainfall has become more extreme, with ~~more extended periods of drought and increasingly extreme~~ longer droughts and increasingly severe rainfall events (?). Thus, understanding rainfall erosivity patterns across Mexico and ~~analysing~~ analyzing them over a multi-year period is essential for enhancing soil sustainability and informed decision-making in soil conservation at national scale.

Rainfall erosivity—often represented as the R ~~factor— quantifies the potential of rainfall to cause SWE (?). The~~ ~~R factor captures the~~ factor—captures the combined effect of raindrop impact and water flow. ? ~~determined~~ defined 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~~. The R factor is commonly expressed as the $EI30$ ~~. Also,~~ index, and over multi-year period, the R factor ~~is presented as~~ represents the mean annual rainfall erosivity~~over a multi-year period~~, and the yearly rainfall erosivity (R_y) ~~is the sum of the~~ results from totaling the rainfall erosivities for ~~the total~~ all 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~~ Several studies have shown that higher-resolution data (e.g., minute-resolution) yield more accurate erosivity estimates for a specific time period (?). High-resolution rainfall data is however challenging to obtain, especially ~~at the national scale. A common solution to estimate the~~ ~~R~~ ~~factor in a data scarcity framework was constructing empirical relationships between erosivity from limited~~ for national-scale analyses in developing countries (?). Consequently, a common approach in these contexts is to construct empirical relationships of erosivity based on insufficient finer-resolution ~~and~~ rainfall data and/or long 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 (?). 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~~ lacks a public database of ~~high-temporal resolution (~~high-temporal-resolution (i.e., 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~~ rainfall time series. The existing data is incomplete and has breakpoints pitfalls that limit studying climatic-related processes at national scales (?). Thus, SWE-related studies have computed the R factor ~~from~~ using coarse-resolution ~~RTS~~ rainfall time series (monthly or annual) for ~~some~~ specific regions (??) ~~; however, there is an imperative need to address the rainfall effects at a national scale~~

~~in Mexico (?). On the other hand, the National Meteorological Service (SMN, by its Spanish acronym) has~~ and/or developed empirical relationships using daily rainfall to approximate $EI30$. The Mexican Meteorological Service provides a rainfall database ~~for public consultation (from 1900 to 2017) at a daily resolution of about 5454~~ (daily records from approximately 5,454 weather stations across ~~Mexican territory. Empirical relationships using daily rainfall~~ Mexico, with time series spanning heterogeneous periods from 1900 up to 2020), which opens up the opportunity to study the national rainfall erosivity patterns.

Several models have been developed to ~~approximate the rainfall erosivity $EI30$.~~ estimate the R factor worldwide. For instance, **?** proposed a power law model by assuming that a daily rainfall could be interpreted as a single storm~~; this~~. This model has been used in ~~many countries and has been demonstrated to be helpful as an indicator of~~ several countries to describe rainfall erosivity patterns (**?**). ~~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 (?). Thus, climatology studies, such as erosivity, need a complete and reliable rainfall time series database (?)~~Said model nevertheless requires sufficiently long time rainfall series (e.g., at least 20 years) to capture seasonal variability and mitigate cyclical rainfall biases (**?**). Therefore, ~~a whole scheme of quality assurance, gap-filling, and homogenization process of the rainfall time series is needed (?). 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 (?). 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~~access to harmonized, quality-controlled, and sufficiently long rainfall datasets is critical for users aiming to perform consistent soil erosion modeling based on the R factor and environmental assessments at regional and national scales.

~~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 (?) 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 (?). 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.~~

~~The lack of reliable and complete daily climate databases has not allowed the study of the~~ The main objective of this contribution is to address the requirements for consistent and reliable model inputs to assess the impact of precipitation on ~~the SWE process at a~~ SWE at national scale in Mexico. ~~Hence, the main objectives of this research are 1) to develop the first Mexican rainfall-series~~ To this end, two specific objectives are defined, namely (a) to construct the first harmonized rainfall database for three climate normals (~~1968-1997, 1978-2007, and 1988-2017) , leveraging on the availability of~~ 1968–1997, 1978–2007, and 1988–2017) derived from legacy climate data, ~~2) to estimate rainfall erosivity~~ and (b) to construct the first rainfall erosivity estimates across continental Mexico~~by using daily rainfall time series. The following section describes the three methodological moments of this research:~~ . Section 2 of this article elaborates on the methodological aspects of 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. ~~Then, the results are presented accordingly. We present a discussion section with a critical analysis of results against~~ The results (Section 3) are then presented accordingly, and, subsequently, discussed through a critical analysis based on recent scientific literature ~~.~~ in Section 4. We finally present our conclusions ~~section, and the availability of all the~~ in Section 5. All the computational codes and resulting databases of this research ~~. The new knowledge allows a better understanding and prediction~~ are freely available from a scholarly-accepted repository. We believe that these products provide a basis for improving the understanding of rainfall distribution and its ~~associated processes in different regions of~~ role in soil erosion processes across Mexico.

2 Methodology

The study area corresponds to ~~the conterminous~~ continental Mexico (1,948,170 km^2)~~. The country~~, which is located between latitudes 14° ~~W~~ N and 32°N and longitudes 86°W and 118°W. ~~Because of its geographical location, the region exhibits complex topographic and climate features~~ Mexico exhibits complex climatic features influenced by subtopical high-pressure systems over the North Atlantic and northeastern Pacific, as well as its position relative to the Intertropical Convergence Zone (**?**). Mexico has been clustered into seven first-level ecoregions—which represent geographical units with characteristic biodiversity (**?**)—, 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~~ Data processing and quality control of rainfall time series ~~(RTS) database development~~

~~The RTS database~~ Developing a rainfall time series and, subsequently, 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 (**?**) 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 (**?**). The CNs are widely used to compare recent observations, create anomaly-based datasets, and provide context for future

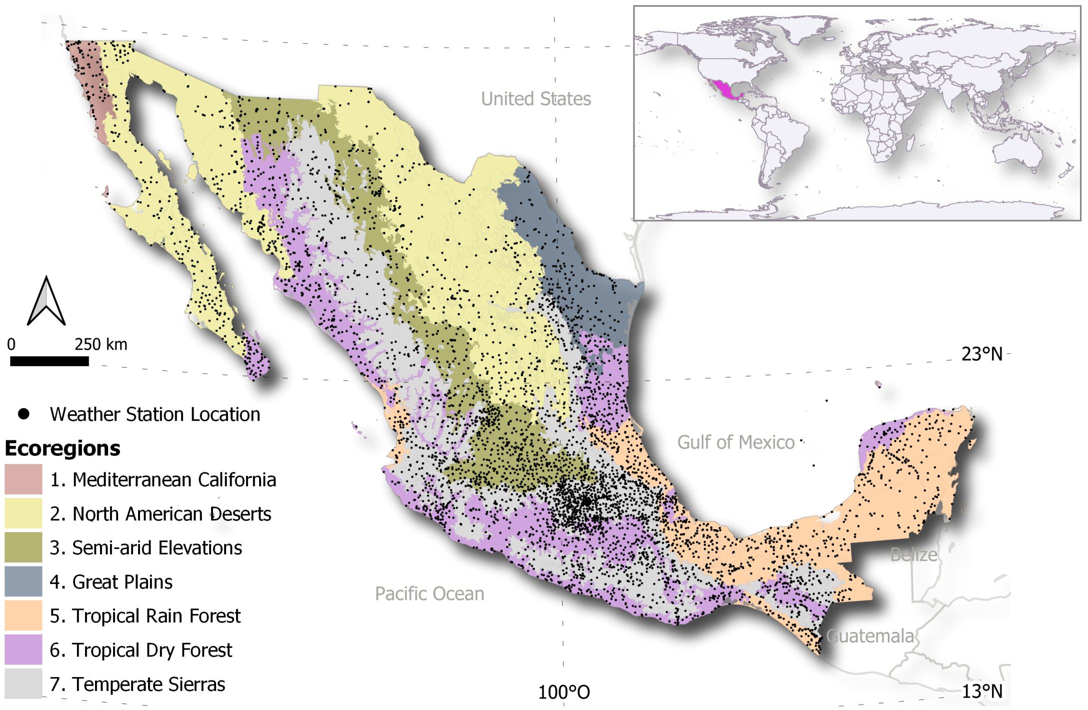


Figure 1. Ecoregions defined by the **?** and locations of the weather station available from the National Meteorological Service (SMN)

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.

On the other hand, climatology studies, such as erosivity, need a complete and reliable rainfall time series database (**?**). Therefore, a whole scheme of quality assurance, gap-filling, and homogenization process of the rainfall time series is needed (**?**) . 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 (**?**). 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.

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

2.1.1 Compilation, selection, quality assurance of ~~RTS~~rainfall time series

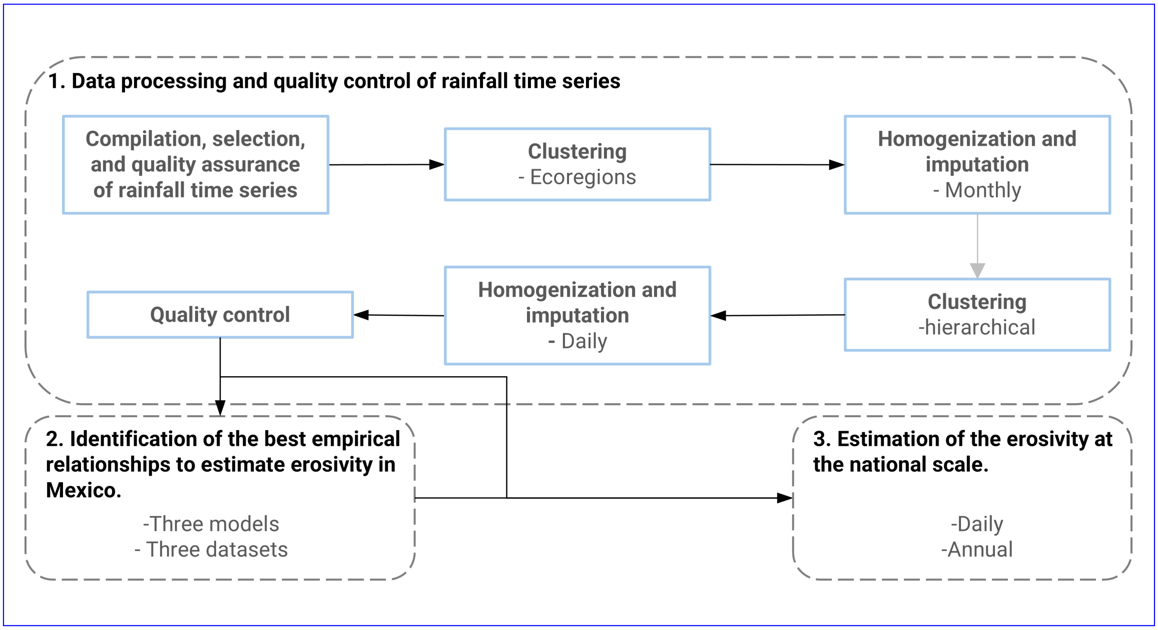


Figure 2. Workflow summarizing the three key methodological stages: (1) Development of the rainfall series database, (2) Identification of the best empirical relationships to estimate erosivity in Mexico, and (3) Estimation of the erosivity at the national scale.

~~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~~ The data were downloaded from the official National Meteorological Service ~~site~~ website (https://smn.conagua.gob.mx/es ~~. In the case of~~) in January 2022. For this project, ~~5454 files in~~ 5,454 plain-text ~~format~~ files were downloaded, corresponding to the daily database of the entire network 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.~~ . After a harmonization process, 44 plain-text files without rainfall data were discarded. Finally, a set of 5,410 daily rainfall time series was obtained for use in this study.

~~All data processing, including file downloading, was implemented in R (?). 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 (%).~~

~~To select the useful RTS, we identified the standard period of registers for each RTS between 1961 and 2018.~~ Although the database includes records with an emission date extending to June 2020, inspection of the rainfall time series indicates that the period with the most complete and consistent data spans from 1961 to 2017. On the other hand, for erosivity estimations, it is recommended to use a historical ~~RTS~~ rainfall time series of ≥20 years (**??**). Therefore, this ~~article looked over the RTS~~ study analyzed the rainfall time series in three climate normals: CN1 (1968 to 1997), CN2 (1978 to 2007), and CN3 (1988 to 2017).

Subsequently, we selected ~~those RTS~~ rainfall time series with less than 20% ~~of~~ missing values (**?**)~~for each CN.~~

~~Calculating the monthly cumulative rainfall, we identified that~~ , resulting in 1489, 1785, and 1728 rainfall series for CN1, CN2, and CN3, respectively (Table A1). However, long sequences of zeros and ~~NAs were common. For quality assurance, we identified those years with no rainfall and replaced zero values~~ missing values (NAs) were frequently identified within the selected rainfall series. Therefore, as part of the quality-assurance procedure, zero sequences ranging from three to six consecutive years were considered suspicious and replaced 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~~Additionally, rainfall time series containing zero sequences longer than six years were discarded from the analysis (10, 11, and 7 series for CN1, CN2, and CN3, respectively). After applying these filters, the final dataset comprised 1479, 1774, and 1721 rainfall time series for CN1, CN2, and CN3, respectively.

2.1.2 Clustering of rainfall time series~~(RTS)~~

Clustering analysis is highly recommended when gap-filling a large set of ~~RTS~~ rainfall time series (**?**). Clustering similar ~~RTS allows~~ rainfall time series allows us 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 (**?**). In this workflow, we performed ~~the~~ data gap-filling by clustering ~~the RTS according to~~ rainfall time series according to (1) ~~the geographical space, and the environmental characteristics of Mexico, and~~ ecoregions and (2) ~~the data dissimilarity among diverse~~ data dissimilarity across different environments (Figure 2). The first clustering was performed using the ~~ecological regions~~ ecoregions of North America (**??**). The second ~~separation was performed after the data~~ clustering was applied after gap-filling and homogenization of the monthly series ~~by ecoregion. A~~ within each ecoregion, using a hierarchical cluster analysis ~~groups the data~~ to group time series with similar seasonality and rainfall volume patterns (**?**). The number of clusters (k) ranged from 2 to $sqrt(n)$, where n is the total number of stations in the dataset (**?**). The better k values for each ecoregion were found using the Hartigan cluster validation index (**?**). This index was identified by **?** 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~~ rainfall time series available for those ecoregions.

2.1.3 Homogenization~~and~~ , data gap-filling and quality control of ~~RTS~~the rainfall time series

We followed three steps to get a complete ~~RTS~~ rainfall time series for each CN: quality assurance, homogeneity analysis, and data gap-filling (?). 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~~ filled during the data gap-filling ~~processes~~process.

Homogeneity analysis removes the biases caused by some artificial breaks in the ~~RTS~~ rainfall time series (?). These breaks result from common issues such as reading or ~~instrumental mistakes~~measurement errors, instrumental changes, or ~~special~~ atypical situations at the weather station location ?. We used the standard normal homogeneity test (?) to ~~analyze~~ analyse 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 (?). 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` R package (?).

We summarized the homogenization parameters used for each step. Table A2 shows the parameters used to homogenize the monthly series, while Table A3 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 (?) 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~~ rainfall time series with a difference greater than 10% related to its original ~~were~~ was discarded.

Finally, to evaluate the effect of the gap-filling procedure on rainfall climatology, we calculated the root mean square error (RMSE) between the mean monthly precipitation of the original incomplete series and the corresponding completed series after gap filling. RMSE was calculated for each weather station and subsequently summarized by ecoregion and climate normal. This metric quantifies the deviation in the monthly climatological behavior introduced by the gap-filling procedure. Therefore, RMSE provides an indicator of the sensitivity of rainfall regimes to missing data completion across different climatic regions.

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 with EI30 information on a global (?), national **?** and local (EI30 calculated from sub-hourly rainfall time series) scale.

2.2.1 Empirical relationships to estimate daily erosivity

The empirical relationship follows the power-law model proposed by **?** when using daily precipitation to estimate the daily erosivity R_d (Equation 1). ~~It has been seen in experiments worldwide in tropical (?) and subtropical (?) zones that the power model has better performance than other models when calculating erosivity from daily data.~~ When comparing rainfall erosivity estimates using power and linear models, power models perform better when applied to daily rainfall time series in tropical and subtropical zones (**??**)

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

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** (?).

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** (?).

The value of α and β variate depending on the climate zone according to the Koppen-Greig Classification as follows: for Tropical (A), $\beta = 1.964 + 0.013 \times$ 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 (?).

-**Model III** (?).

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 **?**.

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

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.

2.2.2 Databases containing $EI30$ information

We used three databases ~~; two have information on~~ to evaluate the performance of the three models described earlier. Two compiled databases provide $EI30$ ~~on a global and national scale, and the other database was constructed using~~ values directly: a global dataset developed by **?** and a national dataset derived from the Master´s thesis of **?**. In addition, a local database was included to enable model performance at a finer scale in Mexico. This third database corresponds to the Michoacán Region and was constructed using sub-hourly ~~RTS to calculate~~ (15 min) rainfall time series from which $EI30$ ~~on a local scale in Michoacán state~~was calculated. These three validation databases have a different statistical distribution of the ~~$EI30$~~ $EI30$ factor (Figure A1~~a~~), 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 (**?**). They estimated the $EI30$ from ~~RTS~~ rainfall time series 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~~ rainfall time series from 2005 to 2015; however, not all the ~~RTS~~ rainfall time series 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, **?** estimated the $EI30$ factor using 54 ~~RTS~~ rainfall time series across Mexico (Figure 3b). The temporal resolution for those 54 ~~RTS~~ rainfall time series is 1 minute, with a temporal period from 1977 to 1987. However, not all ~~RTS~~ rainfall time series have registers for those ten years, so we selected 42 ~~RTS~~ rainfall time series 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

The Michoacán region is one of the main avocado-producing areas in Mexico, making it particularly relevant for studies of soil erosion in intensively managed agricultural systems. At the local scale, the ~~Michoacan~~ Michoacán mountain region has a database with 30 ~~RTS~~ rainfall time series (Figure 3c). These ~~RTS~~ rainfall time series 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 **?** by using the `RainfallErosivityFactor` ~~package (?) in R project~~R package (**?**). 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}.

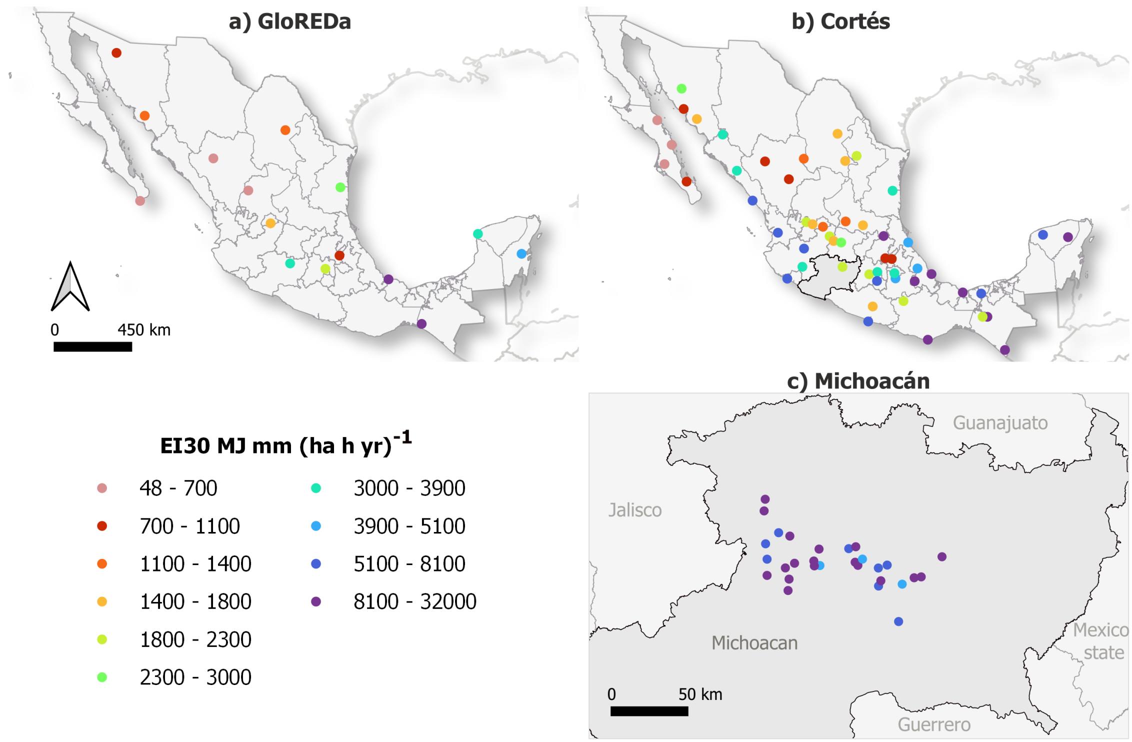


Figure 3. Spatial distribution of the validation datasets: a) GloREDa (n= 15), b) Cortés (n=44), and c) ~~Michoacan~~ Michoacán (n=23).

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~~ 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~~Afterwards, for each point in the GloREDa and Cortés databases, we identified the nearest 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~~ rainfall time series at a daily resolution to estimate the R factor with the three coefficient combinations.

For those points in ~~our~~ the 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 values. In this sense, the slope of the linear model quantifies the mean change ~~of~~ in the R factor ~~when~~ associated with a one-unit increase in $EI30$. Additionally, for each empirical model and validation database, we calculated the Mean Error (ME) and the Root Mean Square Error $(RMSE)$ to evaluate the magnitude and direction of the differences between the estimated R factors and the observed $EI30$ ~~increases by a unit.~~ values.

2.3 Rainfall erosivity estimation

We estimated the R factor for each weather station on the Mexican ~~RTS~~ rainfall time series database. First, we identified the days with erosive rainfall as those with cumulative precipitation greater than ~~12.5~~ 12.7 mm, an extension of the suggestion by (**???**). 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).

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 **?**, 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 (**?**). They found that in the study period, snowfall had never occurred in 93.8% of Mexico's municipalities. ~~The~~ Additionally, 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 (**?**)~~

Finally, to provide an overview of the sensitivity associated with the gap-filling and homogenization procedures, the complete rainfall time series were modified according to the percentage change in the mean value of each rainfall time series before and after these procedures. As established in the quality control section (2.1.3), the mean change in the rainfall time series did not exceed 10%. The modified series were subsequently used to recalculate daily rainfall erosivity values, from which new R factors were estimated for each station and climate normal. The percentage change between the recalculated R factor and the R factor estimated from the complete rainfall time series after the gap-filling and homogenization procedures was then computed to evaluate the sensitivity of rainfall erosivity estimates to alterations introduced during these procedures. This analysis provides a first-order assessment of the propagation of uncertainties associated with modifications in the rainfall time series into long-term erosivity estimates.

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~~ initial number of rainfall time series for the gap-filling, homogenization and quality control process was 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 A1. After the~~ After data gap-filling, homogenization, and quality control~~processes~~, we discarded 4.8% of the ~~RTS~~rainfall time series. Hence, Table 1 shows that the largest number of ~~RTS~~ rainfall time series corresponds to CN1 (1968-1997), with 7.4% (~~109 RTS~~110 rainfall time series), which is followed by CN2 (1978-2007), with 5.4% (~~95 RTS~~96 rainfall time series) and CN3 (1988-2017), with ~~1.9% (33 RTS~~2.6% (45 rainfall time series). Likewise, the most significant proportion of discarded ~~RTS~~ rainfall time series corresponded to Mediterranean California (3 ~~RTS~~rainfall time series, 21.4%) and North American Deserts (29 ~~RTS~~rainfall time series, 15.84%) in CN1; Mediterranean California (2 ~~RTS~~rainfall time series, 11.11%) and Great Plains (7 ~~RTS~~rainfall time series, 12.72%) in CN2; and Great Plains (5 ~~RTS~~rainfall time series, 10.86%) in CN3. Additionally, in the data gap-filling processes, we identified that none of the 5 ~~RTS~~ rainfall time series available for Mediterranean California, in CN3, 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; however, CN1 and CN2 provide a robust basis for assessing rainfall erosivity in this ecoregion. Finally, ~~the available RTS are 1370, 1676, and 1683~~ we obtained a database with 1369, 1678, and 1676 rainfall time series for CN1, CN2, and CN3, respectively. The available ~~RTSs~~ rainfall time series are distributed across the Mexican territory for the three CNs and represent the seven ecoregions (Figure 4).

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 the 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	29	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	16	6	5	179	231	233
Tropical Dry Forest	401	466	454	26	25	12	375	441	444
Temperate Sierras	400	411	415	28	26	9	373	386	409
Total	**1479**	**1774**	**1721**	**110**	**96**	**40**	**1369**	**1678**	**1676**

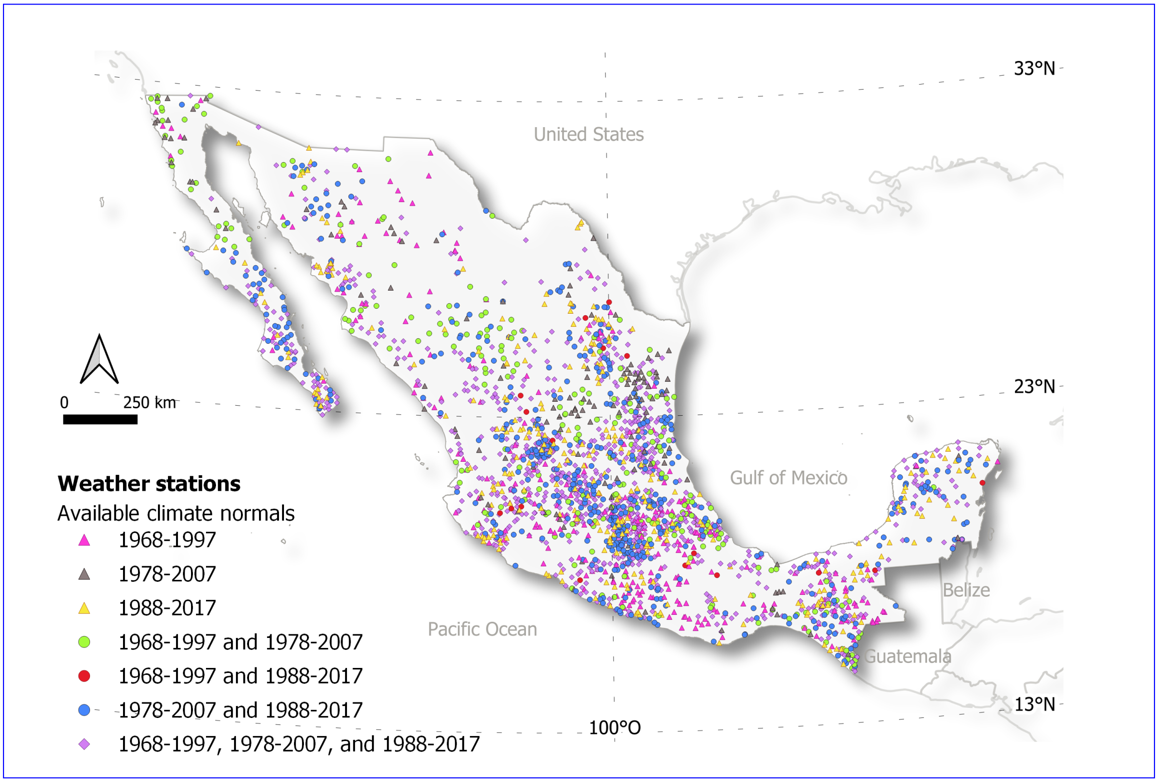


Figure 4. Spatial distribution of the available weather stations for each climate normal: CN1 (1968-1997), CN2 (1978-2007), and CN3 (1988-2017).

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 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).

Figure ~~??~~ A2 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~~ 1968–1997), CN2 ~~and~~ (1978–2007), 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,~~ (1988–2017). The results indicate that Mexico exhibits a unimodal rainfall regime, with a marked peak and a well-defined dry season across all ecoregions; however, the characteristics of the wet period vary among them. In the Tropical Rain Forest, rainfall increases rapidly from May to June and peaks in September, with monthly values exceeding 300 mm. The Tropical Dry Forest and the Temperate

Table 2. ~~Average percent (%) of~~ Mean percentage change ~~of~~ in the mean value of rainfall time series by ecoregion

	Ecoregion	~~Average change of the mean (%)~~ **Mean change in rainfall time series (%)**		
		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

Sierras show a similar annual pattern, with a persistent and evenly distributed wet season from June to September, reaching monthly values of around 200 mm. In the Semi-arid Elevations, ~~and Great Plains, the rainfalldoes not exceed 200 mmfor the month with the highest volume. However, rainfall in Tropical Rain Forest, Tropical Dry Forest, and Temperate Sierras can reach up to 800 mmin June and September. The~~ the wet period begins in May, peaks in July (approximately 150 mm), and declines gradually until October. The Great Plains ecoregion exhibits a wet season that begins in March, followed by a gradual increase in rainfall, sustained precipitation through July (around 100 mm), and a peak in September (approximately 150 mm). In contrast, the North American Deserts ecoregion is characterised by low rainfall throughout the year, with only a modest increase during the summer months (around 50 mm), and no clear peak. The Mediterranean California ecoregion shows a markedly different pattern, with rainfall concentrated between November and March, reaching monthly values of approximately 50 mm. Overall, the general behavior of rainfall distribution ~~by ecoregion is the same for the three CNs.~~ across ecoregions is consistent among the three climate normals. However, slight differences are observed in CN3, where the North American Deserts show an increase in accumulated precipitation in September, and the Great Plains show an increase in July precipitation.

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

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 5). In this context, a slope of 1 indicates perfect agreement between estimated R factors and $EI30$ values, whereas deviations from 1:1 slope reflect fewer agreement.

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 (~~with the best overall performance across the three datasets, with slopes closest to 1 ~~:1 perfect model) with~~ (1.07, 0.92, 0.83, and 0.43 for GloREDa vs. Mexico-

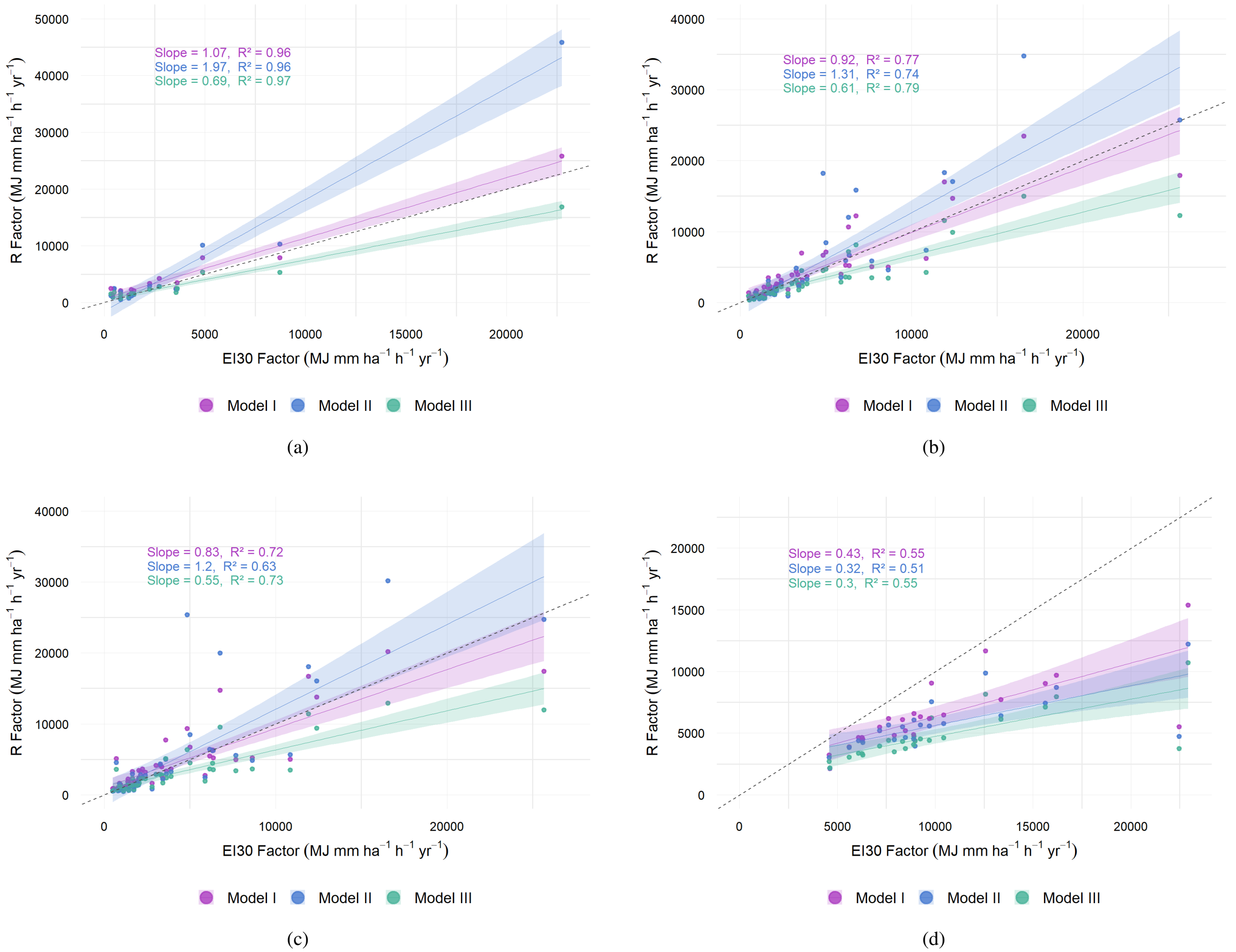


Figure 5. ~~Verification of rainfal~~ Comparison between observed $EI30$ factor and rainfall erosivity ~~index calculated with~~ estimates (R) factor obtained using the three evaluated models ~~: model I (?), Model II (?) and Model III (?). The verification was made with three erosivity databases at~~ across different ~~scales~~ datasets: (~~global, national, and local).~~ a) ~~Global dataset~~ global (GloREDa~~(?) and Mexico-CN3.~~), (b) ~~National dataset Erosivity-Cortés (?) and~~ national (Cortés; Mexico-CN1~~.~~), (c) ~~National dataset Erosivity-Cortés (?) and~~ national (Cortés; Mexico-CN2~~.~~), and (d) ~~Local dataset in Michoacan state: comparison of EI30~~ local (Michoacán). The x-axis represents observed $EI30$ values~~calculated using RTS at two temporal resolutions~~, ~~15-minute and~~ while the y-axis shows model-based rainfall erosivity estimates derived from daily ~~resolution~~rainfall time series. Solid lines indicate the fitted linear regression for each model, ~~respectively~~and the dashed line represents the 1:1 relationship (perfect agreement). The slope of each regression is used as an indicator of model performance.

Table 3. Metrics of performance of three models: Model I (?), Model II (?), and Model III (?) 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

CN3, Cortés vs. Mexico-CN1, Cortés vs. México-CN2 and ~~the two predictions for~~ for the 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~~ , respectively) 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~~the lower RMSE values (1524, ~~0.55, and 0.3, respectively. The model I had the fewest RMSE for all the four comparisons (1544,~~ 2438, 2628, ~~and~~ 4978 MJ mm ha^{-1} h^{-1} yr^{-1} ~~) (See Table 3). Model I also shows a general overestimation in predicting $EI30$ values in the GloREDa and~~ for GloREDa vs. Mexico-CN3, Cortés ~~(~~vs. Mexico-CN1~~and Mexico-CN2) databases with a ME of 956, 324, and 338~~ , Cortés vs. México-CN2 and for the Michoacán database, respectively). Whereas, Model II obtained the poorest performance for GloREDa vs Mexico-CN3, Cortés vs Mexico-CN1, and Cortés vs México-CN2 with the highest RMSE values (6176, 4088, and 4554 MJ mm ha^{-1} h^{-1} yr^{-1}, respectively). Model III ~~underestimated the~~ showed intermediate RMSE values, but consistently underestimated $EI30$ values in all ~~databases. Additionally, for~~ datasets (negative ME values). For the Michoacán ~~database~~dataset, all three models underestimated ~~the~~ $EI30$~~values.~~ , with Model I still providing the best performance (ME = -3699; RMSE = 4978 MJ mm ha^{-1} h^{-1} yr^{-1}), followed by Models II and III, which showed progressively larger errors.

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~~ present the results of ~~the rainfall erosivity estimations~~ rainfall erosivity estimates for the three ~~CNs studied~~climate normals considered. First, ~~a descriptive statistic for erosivity values, then a yearly distribution of rainfall erosivity~~ we describe the statistical distribution of erosivity values, followed by their annual distribution and, ~~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~~ rainfall for each day of the year ~~. For the three CNs, there is one erosive period throughout the year; however, for CN3,~~ for the three climate normals. In all CNs, erosive rainfall is concentrated within a single annual erosive season extending approximately from days 150 to 280, corresponding to the summer rainfall period in Mexico. For CN1, the erosive period is ~~marked by two peaks, first from day~~ relatively stable, with the highest number of locations with erosive rainfall occurring between days 170 and 260. Similarly, CN2 exhibits a well-defined erosive season, although with slightly greater variability and higher peaks than CN1. In contrast, CN3 is characterised by a more pronounced intra-seasonal variability, with two distinct peaks occurring between days 173~~to~~ –190 and ~~second from day~~ 229~~to 261; these days represent those with the most (10% higher)~~ –261. These peaks correspond to periods when the number of locations reporting erosive rainfall ~~. On the other hand~~exceeds the 90th percentile. Additionally, the 90th percentile threshold of the number of locations with erosive rainfall ~~for~~ increases from 211 in CN1 to 227 in CN2 and 238 in CN3~~has increased by~~ , representing increases of 7% ~~(238)~~ and 12%~~(227)~~, respectively, ~~compared~~ relative to CN1~~(212)~~.

The mean values of rainfall erosivity 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 4). The statistical distribution of the rainfall erosivity values was right-skewed with skewness of 2.7, 3.0, and 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).

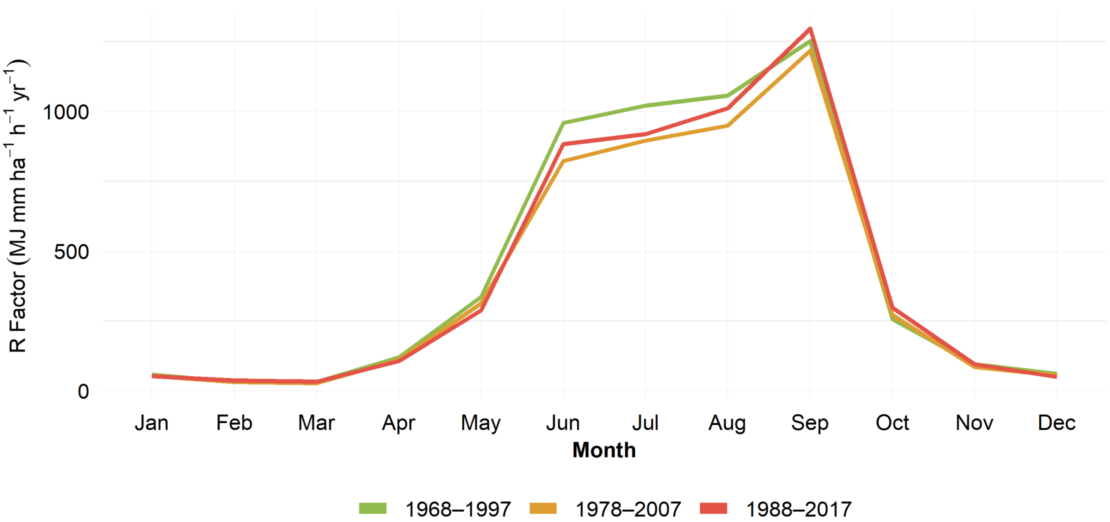
Table 4. Descriptive statistic of the erosivity values for three climate normals (1968-1997, 1978-2007, 1988-2017). Min: minimum, Max:maximum, SD: standard deviation, CV: coefficient of variation, skew: skewness, kurt: kurtosis

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

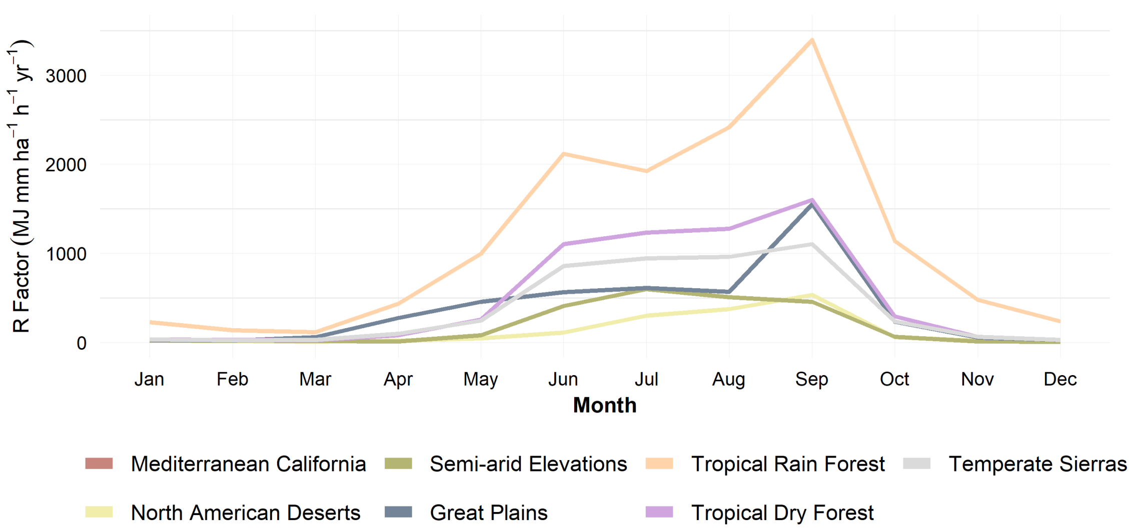
Monthly erosivity varies throughout the year in the three CNs (Figure 6a). The behavior of the erosivity in the three CNs is monomodal, indicating just one peak and one valley. The month with the highest erosivity is September, reaching values almost to 1300 MJ mm ha^{-1} h^{-1} yr^{-1}, followed by 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 6b the monthly erosivity for CN3 by ecoregion. ~~It can be seen that the~~ The Tropical Rain Forest ~~has~~ exhibits the highest erosivity values ~~for all months, with a minimum value~~ throughout the year, ranging from a minimum in March (114 MJ mm ha^{-1} h^{-1} yr^{-1}) ~~and~~ to a maximum in September (3395 MJ mm ha^{-1} h^{-1} yr^{-1}). ~~Notably, the ecoregion with the smallest monthly erosivity is the~~ In contrast, the North American Deserts ~~, with valueshardly reaching~~ show the lowest erosivity values, generally remaining below 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~~ Most ecoregions reach their maximum erosivity in September; however, the Semi-arid Elevations peak earlier, in July. It is important to note that CN3 does not include stations in the Mediterranean California ecoregion. ~~Still~~Nevertheless, 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.~~ this ecoregion shows the lowest erosivity values among all regions.

Regarding the spatial distribution, the erosivity values across Mexican territory look similar for the three CNs (Figure 7 ~~). 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~~ and A5). Low erosivity values (< 1000 MJ mm ha^{-1} h^{-1} yr^{-1}~~, which represents the 1st decile of the distribution. The 2nd decil (values between 1000 and~~) are mainly concentrated in the Baja California peninsula and northern Mexico, corresponding to arid regions such as the North American Deserts and Semi-arid Elevations. Intermediate values (approximately 1600~~–~~4300 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~~ are distributed across central Mexico, reflecting transitional climatic conditions. In contrast, the highest erosivity values (> 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~~ southern and southeastern regions of the country. These areas correspond to the upper deciles of the statistical distribution, indicating a strong spatial concentration of high erosive potential in humid tropical environments. This spatial pattern is also reflected in the mean erosivity values calculated for each ecoregion (Table A5). The Tropical Rain Forest consistently exhibited the highest mean erosivity values for the three climate normals, followed by the Tropical Dry Forest, Temperate Sierras, and Great Plains. In contrast, the Mediterranean California and ~~some places in the North American desert ecoregions because the rainfall series there had records until 2012.~~ North American Deserts ecoregions showed the lowest mean erosivity values. Across the three climate normals, the relative hierarchy among ecoregions remained similar, with tropical and humid regions presenting substantially higher erosivity values than arid and semi-arid regions.



(a)



(b)

Figure 6. Montly erosivity values estimated from daily rainfall time series and using the **?** power-law model. a) Monthly erosivity for the three climate normals; b) Monthly erosivity by ecoregion for the Mexico-CN3 (1988-2017).

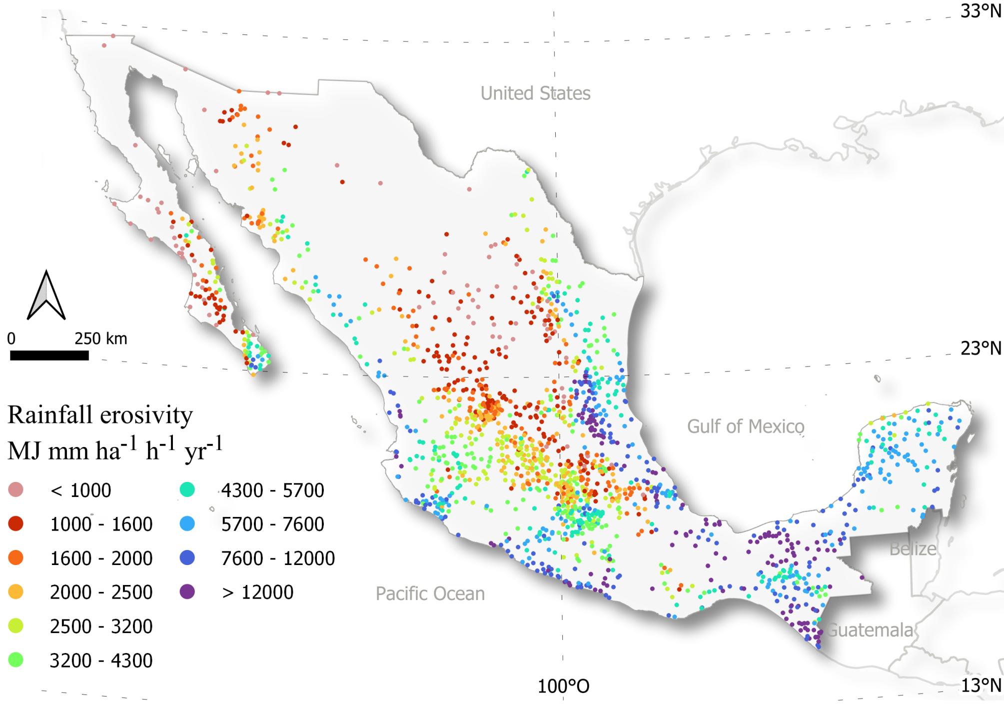


Figure 7. R factor values calculated with the power law equation proposed by (**?**) at daily resolution for the climate normal 1988 - 2017. The legend classes approximately represent deciles of the statistical distribution, although not exactly.

4 ~~Discussion~~

The spatial distribution of percentage changes in rainfall erosivity after the homogenization and gap-filling procedures was heterogeneous across the three climate normals, with positive and negative changes interspersed throughout Mexico (Figure 8a for the CN3 and Figure A6a for the CN1 and CN2). In all climate normals, most stations exhibited relatively small changes, generally between -5% and 5%. CN1 showed a slight predominance of positive changes in northern Mexico, particularly in the Baja California Peninsula and northwestern regions, whereas CN2 displayed a more balanced distribution of positive and negative changes across the country. In contrast, CN3 exhibited a greater occurrence of negative changes in northern and northwestern Mexico. Southern Mexico and the Yucatán Peninsula generally showed lower magnitudes of change across the three climate normals. Overall, the largest positive and negative changes were associated with isolated stations rather than coherent regional clusters.

~~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 scatterplots reveal a consistent inverse relationship between the percentage change in the mean rainfall series after the homogenization and gap-filling procedures and the corresponding percentage change in rainfall erosivity across the three climate normals (Figure 8b for the CN3 and Figure A6b for the CN1 and CN2). The three climate normals exhibited a similar funnel-shaped distribution, in which the dispersion of erosivity changes increased progressively as the magnitude of rainfall change became larger. Stations with precipitation changes close to 0% showed relatively small erosivity variations concentrated near zero, whereas stations with larger positive or negative rainfall changes exhibited a wider range of erosivity responses, reaching approximately ±20%.

Table A5 summarizes the percentage changes in rainfall erosivity by ecoregion for the ~~understanding of key ecosystem climate-related processes at different scales (?). 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.~~ three climate normals after the homogenization and gap-filling procedures. Overall, mean percentage changes remained low across all ecoregions and climate normals, with most values below 3%. For CN1, the largest mean positive percentage changes were observed in the North American Deserts (2.3%) and Mediterranean California (2.1%). In CN2, mean percentage changes were generally smaller, although the North American Deserts still exhibited the largest positive mean change (1.7%), while the Great Plains showed a slight negative mean change (-0.6%). Similarly, in CN3, mean percentage changes remained close to zero across ecoregions, ranging from -0.4% in the North American Deserts and Great Plains to 0.7% in the Tropical Dry Forest. Although the mean changes were comparatively small at the ecoregional scale, some individual stations exhibited larger positive and negative percentage changes. The maximum positive changes reached 21.2%, 19.3%, and 17.6% for CN1, CN2, and CN3, respectively, whereas the minimum changes reached -18.8%, -23.0%, and -19.3%. Overall, the range of percentage changes was similar among the three climate normals, with both positive and negative changes occurring within comparable magnitudes.

4 Discussion

This research addressed the data incompleteness and breakpoints in the legacy Mexican climate time series. Particularly, we compiled and ~~systematized~~ systematised a national dataset with daily rainfall and rainfall erosivity for three ~~CNs~~ climate normals (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~~ rainfall time-series and erosivity dataset at the national scale, based on legacy climate data, ~~at the national scale, following the quality control and inhomogeneity analysis~~ following the quality-control and inhomogeneity analyses 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 (?). The new information ensures the highest possible quality (as explained in the methods section), it includes a rigorous quality control, homogenization, and data~~ This discussion is organised into four principal components. First, we discuss the development and quality-control procedures of the first national rainfall time series database for Mexico, including

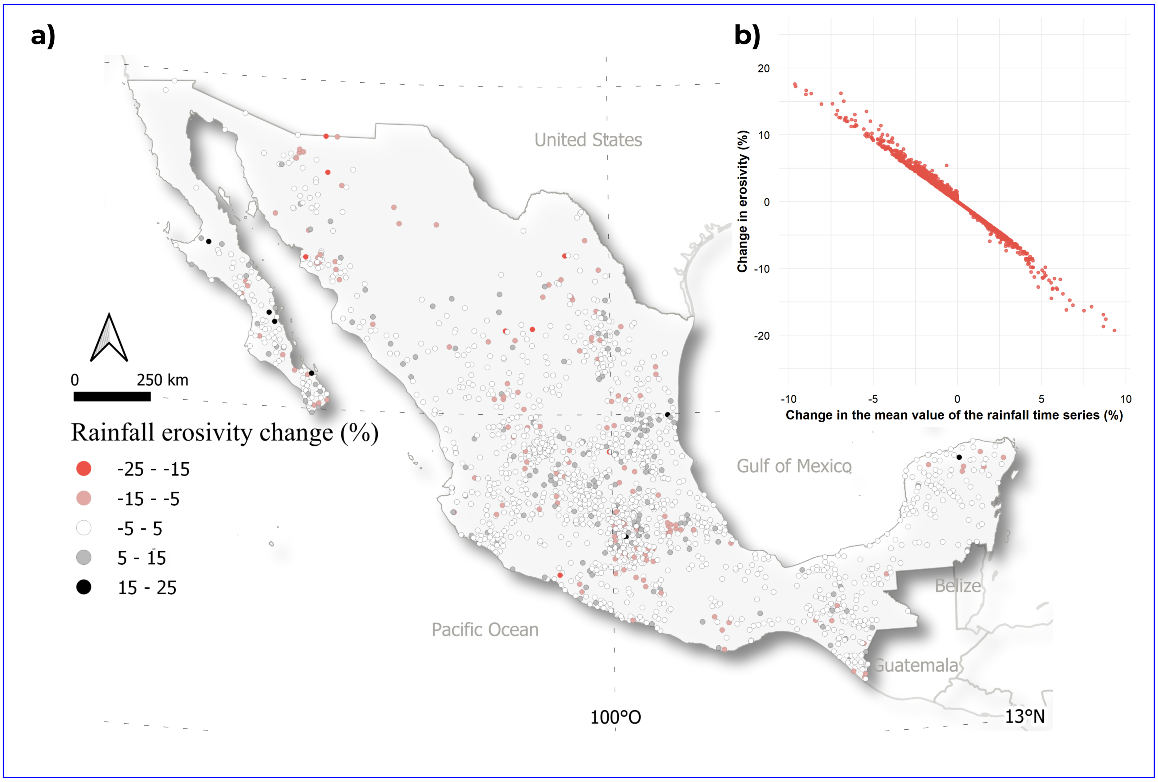


Figure 8. Percentage changes in rainfall erosivity after modifying precipitation time series according to the percentage change in the mean value of the rainfall time series before and after the gap-filling and homogenization processes for the 1988–2017 climate normal. a) Spatial distribution of percentage changes in rainfall erosivity. b) Relationship between percentage changes in the mean value of the rainfall time series and percentage changes in rainfall erosivity.

the implications of homogenization across contrasting climatic regions. Second, we analyse the rainfall erosivity estimates, their spatial patterns, and the performance of the empirical models used for daily erosivity estimation. Third, we evaluate the sensitivity of rainfall erosivity estimates to the homogenization and 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, ? ~~. Finally, we discuss the limitations, potential applications, and future research directions associated with the rainfall and erosivity databases

4.1 Development of the first Mexican rainfall time series database

- Improvement over the previous rainfall products in Mexico

Previous national-scale rainfall datasets for Mexico are those produced by **?** and **?**; they emphasise spatial completeness through interpolation. (**?**) updated the mean monthly rainfall for 100 years (~~1910-2009~~1910–2009) using around 5000 ~~RTS. Other studies have compared multiple gap-filling approaches at a regional scale (?) and conducted quality control and homogenization~~

~~process (?) 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 ??. 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.~~ rainfall time series and used thin-plate smoothing spline interpolation to generate gridded data at 30 arcsec spatial resolution for the continental area of Mexico. Similarly, (?) developed the Mexico High Resolution Climate Database (MexHiResClimDB), a gridded dataset that provides daily, monthly, and annual rainfall for the period 1951–2020, using up to 4000 rainfall records per day. This dataset was generated using Kriging with External Drift on a local neighbourhood, producing gridded data at 20 arcsec spatial resolution across Mexico.

~~Our methodology allowed us to reconstruct climate series with significant missing data gaps, as reported in previous studies (?). The methodology applied on an ecoregion basis, potentially increases spatial coherence in the completed and homogenized RTS, as previously recommended (?). The methodology is also recommended by ?. Other robust data gap-filling strategies leverage high computational capacity and big data analyses (i.e., machine learning) to obtain reliable RTS (??). These strategies build a predictive function to estimate values for missing data using the climate seriesdata itself or auxiliary data such as satellite products (?). Despite the performance predicting missing data values (??), 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~~ These approaches prioritise spatial completeness and provide valuable insights into the spatial distribution of climate variables across Mexico. However, previous gridded datasets either did not perform quality control (?) or applied only basic quality control procedures (?), such as selecting stations with more than 80% of recorded data for at least 10 consecutive years and discarding unrealistic values (e.g., daily rainfall exceeding 600 mm).However, climatological studies have shown that inhomogeneities in rainfall time series can introduce biases and artificial trends, potentially leading to erroneous interpretations and misleading conclusions (?). In consequence, **??** guidelines recommend that rainfall data undergo a comprehensive set of quality control tests, including constraint, consistency, rapid change, and domain analysis. These procedures are essential to ensure the reliability of climatic analyses, avoid biases due to measurement or recording errors, and maintain comparability across stations and regions, particularly in long-term rainfall time series.

Therefore, although gridded datasets are important for understanding large-scale ecological processes across Mexico, generating climate data from reliable and quality-controlled time series is equally important. This represents a key strength of our rainfall dataset compared to previous efforts. By providing a robust and quality-controlled rainfall time series database, we ensure greater reliability for climate analyses that depend on the accurate representation of extreme events and temporal variability, including rainfall erosivity estimation, hydrological modeling, and climate extremes assessment.

- Changes in the mean monthly rainfall by ecoregion

Since the homogenization process corrects outliers, artificial breaks, and inconsistencies in the raw records, larger RMSE values may reflect stronger corrections applied to low-quality or highly inhomogeneous rainfall series (?). In this sense, the

RMSE partly reflects the initial quality of the raw rainfall time series rather than solely the uncertainty introduced during the reconstruction process. Consequently, the RMSE also provides insight into the importance of homogenization and quality-control procedures across ecoregions, where breakpoints and outliers may affect rainfall series differently depending on the climatic regime. For example, in the Tropical Rain Forest ecoregion, the RMSE value for October was 12.11 mm, whereas the corresponding mean monthly precipitation was 230.9 mm, representing a relative deviation of only 5.4% with respect to the ~~supplementary material A4. Future improvements of this new RTS database must include a progressive exploration of multiple gap-filling data techniques~~raw series. In contrast, in the North American Deserts, the RMSE value for March was only 2.08 mm; however, because the corresponding mean monthly precipitation was 7.97 mm, this represented a relative deviation of 26%. These comparisons indicate that rainfall series without homogenization may produce proportionally larger distortions in climatic variability and long-term trend analyses in arid and semi-arid regions, where low rainfall magnitudes amplify the effect of outliers and artificial breaks. This is particularly relevant because arid and semi-arid regions represent approximately 41.7% of the continental area of Mexico.

- Rainfall patterns across Mexico

The ~~gap-filled dataset (considering the completed and homogenized RTS for the three CNs) is useful for analysis of changes in precipitation trends (?) across Mexico. This new RTS dataset shows that Mexico experiences~~ rainfall patterns observed across ecoregions are clearly influenced by both global climate dynamics and local topographic conditions. Overall, Mexico exhibits a unimodal rainfall regime~~with a cusp in July and September, depending on the ecoregion. This unimodal regime was also~~, with a wet period occurring during late spring and summer (May to October), as reported by **?**~~who analysed RTS at the national scale, reporting that Mexico's~~. However, the characteristics of this rainfall regime vary among ecoregions. For example, in the Tropical Dry Forest and Temperate Sierras (principally located over the Sierra Madre Occidental and the Eje Neovolcánico), the 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 ?? 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 ?. In addition, the new dataset also provides an earlier perspective of the spatial distribution of rainfall erosivity in Mexican territory.~~ concentrated between July and September. This pattern is largely driven by the North American Monsoon, which develops from July to September and originates along the western coast of Mexico (**?**). During this period, a pressure gradient draws warm, moist air inland from the Pacific Ocean and the Gulf of California, and winds transport this moisture toward the interior. As the air rises over the Sierra Madre Occidental, it cools and condenses, producing convective rainfall, typically in the afternoon (**?**). In addition to the monsoonal influence, the Tropical Rain Forest (located mainly in southeastern Mexico) exhibits a rainfall pattern associated with the northward migration of the Intertropical Convergence Zone, enhanced convective activity, and the contribution of tropical cyclones during late summer (August and September), resulting in increased monthly precipitation (**?**).

In contrast to the tropical and monsoonal processes that dominate rainfall across most of Mexico, the Mediterranean California rainfall pattern is influenced by different climate dynamics. This ecoregion is located in the northwesternmost part of Mexico (Figure 1) and is characterised by hot, dry summers and wet winters (**?**). Winter rainfall is associated with frontal storms originating in the Pacific Ocean (?), linked to the position of the region on the equatorward flank of mid-latitude storm tracks and the poleward edge of the Hadley cell. In contrast, dry summer conditions are related to descending air and atmospheric subsidence associated with the eastern flanks of subtropical anticyclones (?).

4.2 Rainfall erosivity estimates and model performance

- Evaluation of empirical models for daily erosivity estimation

Among the three coefficient combinations evaluated for estimating daily rainfall erosivity, the **?** 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~~ parameterisation of the α and β coefficients have substantial implications for model behavior under diverse rainfall regimes. The **?** equation was developed using ~~RTS~~ rainfall time series across the United States of America, including rainfall conditions of the southern states such as Georgia, Mississippi, and Texas. This equation defines different α ~~coefficient~~ coefficients 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 (?). Additionally, the β coefficient has no spatial or seasonal pattern, the same as found in the **?** equation and contrary to the **?** 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 **?** equation (1.81) is slightly higher than that for **?** and **?** (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 **?** model obtained the highest erosivity estimations for the Michoacán dataset (purple points always over the blue and green dots in Figure 5d), because β coefficients for all the Michoacán dataset were 1.81, 1.558 and 1.72 for the three models **?**, **?**, and **?**, 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 (?) and annual (?) resolution. This relationship has also been seen in our three validation datasets, where the~~

- Erosivity estimations

Our work reveals that the distribution of erosivity values in Mexico corresponds to the geographical distribution of rainfall and seasonal precipitation conditions. The areas with the highest erosivity values are concentrated in the Isthmus of Tehuantepec (i.e., the shortest distance between the Gulf of Mexico and the Pacific Ocean). This region corresponds to the Tropical Rain Forest ecoregion, where high moisture availability enhances extreme precipitation events, increasing rainfall erosivity (?). High values are also concentrated along the Pacific coastal zone and in low-elevation areas of the Sierra Madre Occidental, where previous studies have reported intense precipitation events at hourly and daily timescales (?). In contrast, lower erosivity values occur in the central and northern regions of Mexico, where severe droughts (e.g., those occurring from the 1990s to the beginning of the twenty-first century), associated with large-scale ocean–atmosphere circulation patterns, have reduced mean annual rainfall~~can explain 90, 70, and 40~~. We also highlight a hotspot in the southern part of the Baja California Peninsula (e.g., Sierra La Laguna), where erosivity values are higher than those estimated for the northern part of the peninsula. This local variation is associated with the influence of tropical cyclones, which contribute up to 50% of the ~~total variability of the~~

~~*EI*30 values in the GloREDa~~mean annual rainfall in this region (?). Overall, the spatial variability of erosivity across Mexico reflects the interplay between rainfall regimes and climatic events, underscoring the strong influence of regional atmospheric processes on soil erosion dynamics.

In agreement with these spatial patterns, our rainfall erosivity estimates are generally higher than those reported for Mexico by global products such as GloRESatE (?) and GloREDa (?). The GloRESatE grid reports erosivity values ranging from 49 to 49,~~Cortés,~~ 811 MJ mm ha^{-1} h^{-1} yr^{-1}, with the highest values located in southern Chiapas. Similarly, the GloREDa grid reports values between 51 and ~~Michoacán databases~~ 13,058 MJ mm ha^{-1} h^{-1} yr^{-1}, identifying the highest erosivity in the Isthmus of Tehuantepec and the coastal zone of the Gulf of Mexico. In contrast, our database reports maximum erosivity values exceeding 43,~~respectively (Figure A1a~~ 000 MJ mm ha^{-1} h^{-1} yr^{-1} in the tropical regions of southern Mexico. Additionally, **?** reported erosivity values ranging from 8,000 to 20,000 MJ mm ha^{-1} h^{-1} yr^{-1} for the Yucatán Peninsula, further supporting the occurrence of high erosivity conditions in tropical regions of Mexico.

Despite differences in magnitude, both global products and our dataset consistently identify tropical regions as the areas with the highest erosivity. This pattern is also consistent with the global distribution of EI30 values reported in the databases used to construct these products. For example, the GloRESatE database (?), which contains approximately 6200 sites worldwide, reports a maximum EI30 value of 39,124 MJ mm ha^{-1} h^{-1} yr^{-1} in Bangladesh. Likewise, the GloREDa database (?), based on 3940 sites globally, reports a maximum erosivity value of 58,288 MJ mm ha^{-1} h^{-1} yr^{-1} in Las Cruces, Costa Rica. These comparisons suggest that the high erosivity values observed in tropical regions of southern Mexico are consistent with the global tendency for extreme erosivity to occur in humid tropical environments.

Conversely, the lowest erosivity values in our database are associated with the arid regions of northern Mexico and are consistent with those reported by GloREDa and GloRESatE. Although there are no specific EI30 estimations available for these Mexican arid regions, comparable values have been reported for the southern United States, which shares similar climatic conditions. For example, **?** estimated EI30 values between 500 and ~~b). Despite the difference between the three datasets, it is also seen that the adjustment coefficients proposed by ? 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~~1000 MJ mm ha^{-1} h^{-1} yr^{-1} using CMORPH satellite data with a temporal resolution of 30 minutes and a spatial resolution of 8 km. Similarly, the GloRESatE database (?), which includes a large number of weather stations across the United States, reports erosivity values ranging from 400 to 1200 MJ mm ha^{-1} h^{-1} yr^{-1} near the Mexico–United States border. These comparisons indicate that the low erosivity values identified in northern Mexico are consistent with the broader climatic conditions of arid and semi-arid regions in North America.

- Improvement over previous erosivity datasets.

This new erosivity dataset is appealing due to two principal advantages: spatial and temporal coverage. It provides erosivity values for a substantially larger number of sites (1369, 1678, and 1676 for CN1, CN2, and CN3, respectively) compared to the dataset developed by **?**, which includes erosivity estimates for 15 sites in Mexico, and the dataset compiled by **?**, which estimated erosivity values for only 53 sites across the country. Additionally, this dataset improves the representation of rainfall erosivity conditions in several regions, particularly in the north of Mexico (e.g., Chihuahua and Baja California), the central region (e.g., Zacatecas and Querétaro), and the south (e.g., Campeche and Quintana Roo), where erosivity estimates were previously unavailable. Additionally, this database provides erosivity estimations for the most rainy regions in Mexico, such as the Pacific coast and the isthmus of Tehuantepec

On the other hand, this database is appealing for presenting multi-annual erosivity values for three climate normals covering the period from 1968 to 2017. As defined by **?**, estimating the R factor requires a long-term rainfall time series for reasonable estimations. Some studies have reported different minimum length required for a reliable estimations of rainfall erosivity: five years (**?**), 10 years (**?**), 15 years (**?**), and 20 years (**??**). In contrast, previous datasets have reported erosivity estimations from rainfall time series with a length between 1 and 11 years (**?**) and 5 and 10 years in **?**. This new database is appealing for the need of long-term rainfall time series to mitigate the seasonal and cyclical rainfall biases in the estimation of rainfall erosivity (**?**).

- Sensitivity of R-factor estimates to homogenization and gap filling process.

To evaluate the implications of the homogenization and gap-filling procedures on rainfall erosivity estimates, we analyzed both the sensitivity of erosivity responses to changes in the rainfall time series and the spatial distribution of these changes across Mexico. The results reveal that homogenization may locally modify erosivity estimates; however, the spatial structure of rainfall erosivity patterns remains consistent across ecoregions. The funnel-shaped distribution observed in the scatterplots (Figures 8b and A6b) indicates a high sensitivity in the erosivity estimations to the larger changes in the mean of the rainfall time series (**?**). This sensitivity is explained by the potential model implemented in erosivity estimation (**?**), but it could also partly be explained by changes in the number of erosive rainfall days. Since daily erosivity was calculated only for days exceeding the erosive rainfall threshold (12.7 mm), modifications introduced during homogenization and gap filling may cause some days to move above or below this threshold. As a result, changes in rainfall can affect both the magnitude of daily erosivity and the frequency of erosive days, leading to greater dispersion in erosivity responses as rainfall changes become larger.

The ~~models based on a coarser RTS could~~ absence of spatially coherent clusters of positive or negative erosivity change suggests that the effects of the homogenization and gap-filling procedures were not primarily controlled by regional climate conditions. This is particularly relevant because the gap-filling process was performed within ecoregions, using neighbouring stations in the same climatic context (**?**). Therefore, if the procedure had introduced a systematic regional bias, changes would be expected to appear spatially concentrated within specific ecoregions. Instead, the largest changes occurred at isolated stations, suggesting that they are more likely associated with the initial quality of individual non-homogenized rainfall time series. In this sense, larger percentage changes in the mean of the rainfall time series should be interpreted as indicators of stronger corrections applied to the raw rainfall time series, rather than as evidence of systematic distortion introduced by the homogenization process. Thus, the spatially scattered distribution of changes supports the interpretation that the homogenization procedure corrected station-specific inconsistencies while preserving the broader ecoregional rainfall structure.

4.3 Limitations and potential use of the study

- Limitations of the rainfall and erosivity database

Models based on coarser rainfall time series can underestimate and overestimate ~~the~~ $EI30$ values, as ~~seen~~ observed in this study. Although calculated at a coarser temporal resolution, underestimations have already been indicated by **??** while evaluating the effect of modifying the time interval for calculating $R(EI30)$ using 5, 15, 30, and 60-minute ~~RTS~~rainfall time series. The authors concluded that increasing the time interval leads to underestimating erosivity values. Similarly, **?** 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 (**?**). 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~~In this study, the best-performing coefficient combination was that proposed by **?**. 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~~ we did not differentiate parameter coefficients by ecoregion due to the limited availability of sites with $(EI30)$ values. In regions with high-intensity rainfall events, such as the Pacific coastal zone of Mexico, rainfall tends to exhibit greater erosivity than in regions with lower-intensity rainfall, even when total rainfall amounts are similar. Under these conditions, coefficient parameters are expected to vary across regions (**??**), reflecting differences in the erosive power of rainfall. Additionally, 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 (?). Regions with small amounts of~~ choice of model and its associated coefficients constitutes an important source of uncertainty in subsequent soil loss estimations, which should be taken into account by users (**?**).

- Potential uses of the rainfall ~~, for example those in the California Mediterranean and North American Deserts ecoregions, it is necessary to fit specific adjustment coefficients (?).~~ and erosivity database

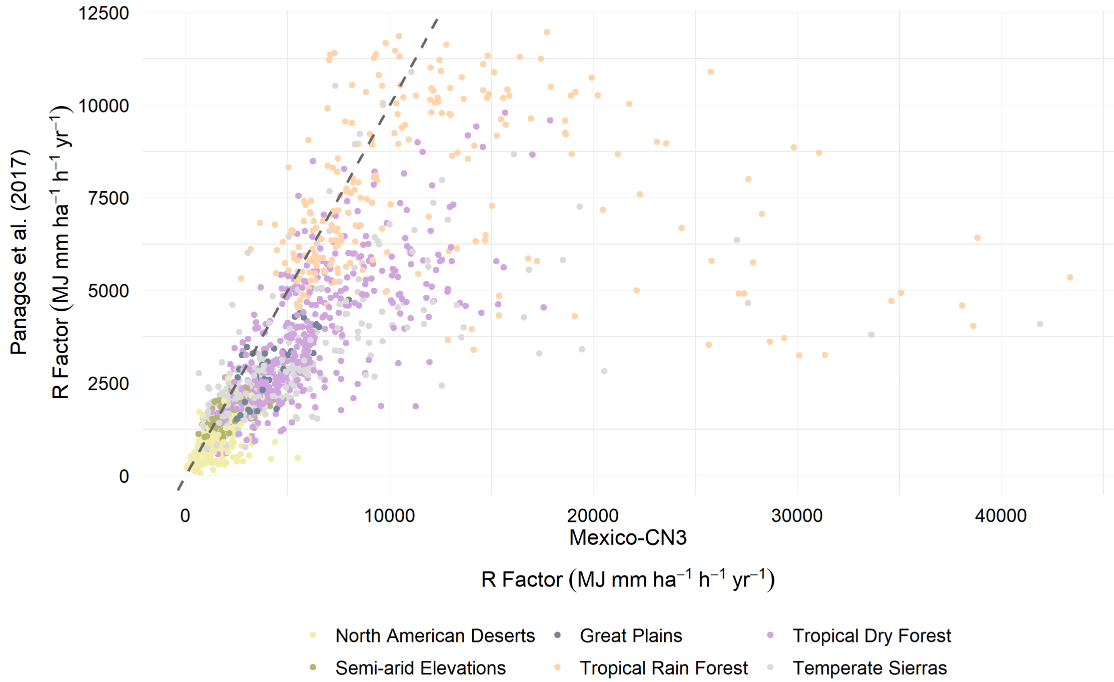
~~Our work reveals that the distribution of erosivity values in Mexico corresponds to~~ This new database will help as an indicator of rainfall erosivity patterns across Mexico. We report daily rainfall at a wide range of altitudes from 1 to 4283 meters above sea level. This is a valuable rainfall database because it is meeting the three primary requirements 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 (?). 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~~ 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 **?**.

In this context, this erosivity database represents a valuable tool for local, national, and global erosion studies. At the local scale, it can support the design of field experiments to validate rainfall erosivity estimates under different rainfall regimes, as well as the identification of the magnitude of underestimation or overestimation associated with different temporal resolutions of rainfall data (**???**). Finally, because this database spans a long period (1968–2017) and is divided into three climate normals, it enables users to analyse trends in rainfall and rainfall erosivity values and relate them to other erosion drivers, such as land use and soil cover dynamics (**?**).

At the national scale, this database can be used as input for erosion models to establish a baseline of soil loss rates across Mexico; indeed, the need for reliable erosivity values has been highlighted in previous studies (**?**). Additionally, within the framework of the United Nations Convention to Combat Desertification (UNCCD), countries have the opportunity to report their contributions to the Sustainable Development Goals (SDGs), including Indicator 15.3.1, which measures the proportion of degraded land relative to the total land area (**?**). In this context, Mexico could use this database to improve the assessment of such indicators by relying on nationally derived datasets, which better capture local conditions than global products. ~~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 (**?**). 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~~ At the global scale, this new erosivity database 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 **?**, where they used~~ , for example, the global rainfall erosivity databases such as GloREDa (**??**) or the Global Rainfall Erosivity database from Reanalysis and Satellite Estimates -GloRESatE- (**?**). The GloREDa is a gridded erosivity product (at 1 km spatial resolution) made with almost 4000 ~~RTS~~ rainfall time series at 30-min temporal resolution to estimate $EI30$ values~~and then map the erosivity distribution across the globe.~~ , and global bioclimatic covariates. Appealing for a validation of the use of this database, we compared our erosivity estimations with those published in **?**.

We identified that our rainfall erosivity predictions for the Mexico-CN3 database obtained greater values than those from **?** 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 **?** reports a lower mean annual rainfall (1383-2100 mm and an SD of 402.35 mm, those values calculated with three ~~RTS~~ rainfall time series 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 **?** (range of 814 to 2258 mm, in this ecoregion, GloREDa has just two ~~RTS~~rainfall time series). We therefore report a more complete description of ~~RTS~~ rainfall time series variance. We believe that our contribution could be useful in better representing global erosivity estimates.



(a)

Figure 9. Scatter plot of R factor values from the global rainfall erosivity surface from **?** and the R factor estimated with the **?** power law equation with the ~~RTS~~ rainfall time series of the Mexico-CN3 (1988-2017) database.

~~Aligned~~ Consistent with our results, **?** found the largest differences in annual rainfall erosivity values between the global product and a satellite-based approach (using sub-hourly ~~RTS) for~~ rainfall time series) in the rainiest regions (tropical and temperate climates) ~~around the world~~worldwide. 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 **?** product was found in Europe, where the density of rainfall gauges is the highest globally ~~(?)~~(**??**). 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~~

4.4 Future research directions

Now, we present our perspective on future directions for cost effective modeling and mapping rainfall erosivity patterns ~~in all seven ecoregions defined in Mexico . Also, we report rainfall erosivities at awide range of 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 ? 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~~and trends across Mexico along with four main interoperability barriers: a) transitioning into high temporal resolution and to sub hourly dynamics, b) climate change projections and non stationarity, c) integration with digital soil mapping and soil organic carbon sequestration models and d) uncertainty quantification and the role of citizen science.

~~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 (???). At the national scale, this database could serve as input in erosion models to generate a baseline of soil lossrates across Mexican territory. This necessity has been highlighted by ?. 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 (?) or the Global Rainfall Erosivity database from Reanalysis and Satellite Estimates -GloRESatE- (?). 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~~First, while the current database provides a robust baseline for the R-factor in Mexico, future research could prioritise the integration of sub-hourly rainfall intensity data. The increasing availability of automated weather station networks and high-resolution satellite products (such as GPM-IMERG) offers a pathway to characterise soil erosivity with increased accuracy (???). Transitioning from daily aggregates to event-based modeling will allow for a deeper understanding of how extreme, short-duration convective storms, which are common in Mexico's complex topography, contribute disproportionately to annual soil loss.

~~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) (?) 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 (?) 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 (?). Therefore, evaluating the performance of different adjusted models is appealing for future research when~~ Second, rainfall erosivity projections are commonly based on stationarity assumptions (?); a critical frontier involves the incorporation of non-stationarity into erosivity mapping. As climate change alters the frequency and intensity of tropical cyclones and North American Monsoon patterns, static historical averages may become less predictive (**???**). Future modeling efforts could couple the current R factor baseline with CMIP6 climate projections to develop digital scenarios of future erosivity patterns and trends. This will be essential for designing long-term sustainable land management strategies and adapting hydraulic infrastructure to a more aggressive erosive environment.

Third, there is a significant opportunity to integrate rainfall erosivity databases with digital soil mapping frameworks to evaluate the positive and negative feedbacks between erosion and soil organic carbon dynamics (**???**). Future studies could focus on the soil carbon losses in response to soil erosion and sediment transport across different geomorphic positions (**??**) . By coupling 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~~erosivity maps with national soil property grids, researchers can better quantify the role of water erosion in carbon redistribution, moving Mexico toward a more comprehensive National Soil Monitoring System (**???**).

Fourth, future iterations of the Mexican rainfall erosivity database should focus on rigorous uncertainty quantification e.g., using Bayesian frameworks or ensemble modeling. Given the uneven distribution of primary weather stations across Mexico's complex and diverse terrain, identifying regions with high predictive uncertainty is paramount. Furthermore, the integration of 'Citizen Science' and low-cost IoT (Internet of Things) recording rain gauge could serve as a useful validation tool. Engaging local communities in data collection would not only fill geographic gaps but also bridge the divide between high-level modeling and on-the-ground soil conservation practices (**???**).

5

We present ~~unprecedented~~ the first harmonized rainfall and rainfall erosivity ~~databases across Mexico. The research used~~ database for Mexico, developed from legacy climate data ~~to assess erosivity across Mexican territory. However, the~~ and covering the period from 1968 to 2017. Our workflow included complete quality control and homogenization processes to ensure data quality for 1369, 1678, and 1676 rainfall time series ~~(RTS) contained a relatively large number of missing values. A gap-filling procedure was performed to obtain gap-free estimates~~ data and rainfall erosivity 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 .~~ (1968-1997, 1978-2007, and 1988-2017), respectively. This is a climatological database that substantially improves the spatial representation and documentation of erosivity patterns compared to global and local products.

Results revealed a strong climatic and geographical control on erosivity patterns. Tropical humid regions, particularly the Isthmus of Tehuantepec, southern Mexico, and portions of the Pacific coastal zone, exhibited the highest erosivity values, whereas the arid and semi-arid regions of northern Mexico showed the lowest erosive potential. These patterns are consistent with the broader global tendency for extreme erosivity to occur in humid tropical environments.

The sensitivity analysis indicates that the homogenization and gap-filling procedures may locally modify rainfall erosivity estimates, particularly in stations with low-quality or highly inhomogeneous rainfall records. However, despite these local variations, the general spatial patterns and ecoregional contrasts of rainfall erosivity remained stable across the three climate normals. These results suggest that the homogenization process effectively corrected station-specific inconsistencies while preserving the broader climatic structure of rainfall erosivity across Mexico

The new database ~~is available for public consultation. This database is for researchersand students, technical assistants, decision-makers, and other users~~ provides an important foundation for national-scale soil erosion assessments and improves the representation of rainfall forcing for RUSLE-based applications, land degradation studies, hydrological analyses, and climate-related environmental research. In addition, the open availability of the database under FAIR principles facilitates its use by researchers, technical agencies, students, and decision-makers 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~~dynamics and soil erosion processes in Mexico.

Future efforts should focus on incorporating high-temporal-resolution rainfall observations from automatic weather stations, strengthening institutional data-sharing initiatives, and developing spatially explicit erosivity prediction models using geostatistical and machine learning approaches. These advances would further improve the accuracy of rainfall erosivity estimation and contribute to more robust soil erosion assessments under changing climatic conditions.

6

Rproject scripts to reproduce the workflow described in this research is available at: https://doi.org/10.5281/zenodo.15468097 (**?**)

7

Following the FAIR principles for scientific data, we published the resulting rainfall erosivity databases (Mexico-CN1 1968–1997, Mexico-CN2 1978–2007, and Mexico-CN3 1988–2017) together with the completed daily rainfall time series for the three climate normals in the Environmental Data Initiative (EDI) repository at https://doi.org/10.6073/pasta/dd2b30e28ee25ff2d60d8a9f436951d (**?**).

The rainfall erosivity databases contain station metadata, including weather station code, name, geographic coordinates, altitude, and ecoregion, as well as indicators associated with the homogenization and gap-filling procedures, such as the Root Mean Square Error (RMSE) and the percentage change in the rainfall time series. Additionally, the databases include mean annual rainfall, the accumulated number of erosive rainfall days, and rainfall erosivity estimates derived from the empirical models proposed by **?**, **?**, **?**.

The daily rainfall databases contain quality-controlled, homogenized, and gap-filled daily rainfall time series for 1369, 1678, and 1676 weather stations corresponding to the 1968–1997, 1978–2007, and 1988–2017 climate normals, respectively. Each rainfall time series is identified by a unique weather station code and includes flags associated with the quality-control and homogenization procedures.

Appendix A: Supplementary Tables and Figures

Table A1. ~~Homogenization parameters for monthly~~ Identification of the number of rainfall time series ~~. Eco: Ecoregion (~~with consecutive zero and NA values

Consecutive years	**Number 1968 - 1**
1 ~~: Mediterranean California,~~	794
2 ~~: North American Deserts,~~	344
3 ~~: Semi-arid Elevations,~~	155
4 ~~: Great Plains,~~	88
5 ~~: Tropical Rain Forest,~~	77
6 ~~: Tropical Dry Forest,~~	21
7 ~~: Temperate Sierras); dz.max and dz.min (upper and lower): standard deviations to consider suspicious and anomalous data~~	1
8	2
9	1
10	2
11	3
12	0
13	0
14	0
15	0
16	1
RS (NA menor 20%)	1489
RS removed	10
RS for data gap-filling process	1479

Table A2. 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); inht: threshold values used to identify break points in the Standard Normal Homogeneity Test (SNHT); dz.max and dz.min (upper and lower): standard deviations to consider suspicious and anomalous data.

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

Table A3. 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: threshold values used to identify break points in the Standard Normal Homogeneity Test (SNHT); dz.max and dz.min (upper and lower): standard deviations to consider suspicious and anomalous data

1968-1997

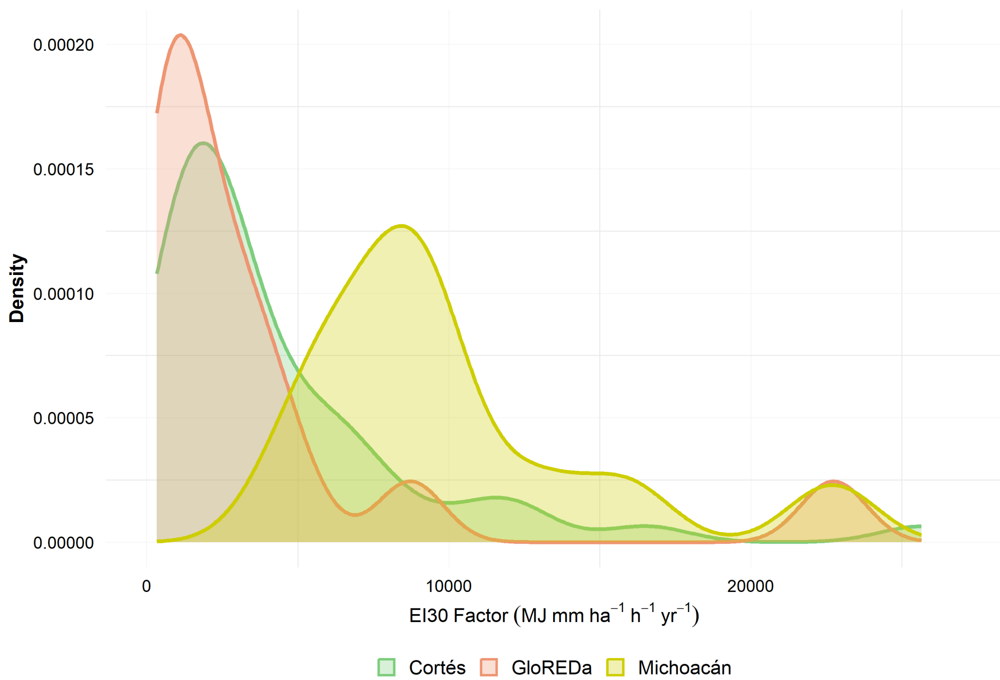
Eco	Group	RS	inht	dz.max lower	dz.max upper	dz.min lower	dz.min upper
2	1	94	50	40	45	-20	-20
	2	18	25	40	45	-25	-30
	3	36	20	40	45	-25	-30
	4	35	20	35	40	-20	-25
3	1	34	25	24	26	-10	-12
	2	87	40	26	28	-14	-16
	3	59	40	22	24	-12	-12
	4	52	40	26	28	-14	-14
	5	16	40	26	28	-10	-12
5	1	36	20	30	35	-15	-20
	2	11	5	20	22	-10	-10
	3	36	50	24	26	-14	-14
	4	58	50	30	32	-14	-16
	5	27	40	20	22	-10	-12
	6	27	40	20	22	-12	-14
6	1	118	25	35	40	-30	-30
	2	187	60	40	45	-25	-30
	3	77	25	40	40	-20	-25
	4	19	25	50	55	-30	-35
7	1	82	50	22	24	-14	-16
	2	49	50	35	35	-20	-25
	3	153	50	35	35	-25	-25
	4	116	50	24	26	-10	-12

1978-2007

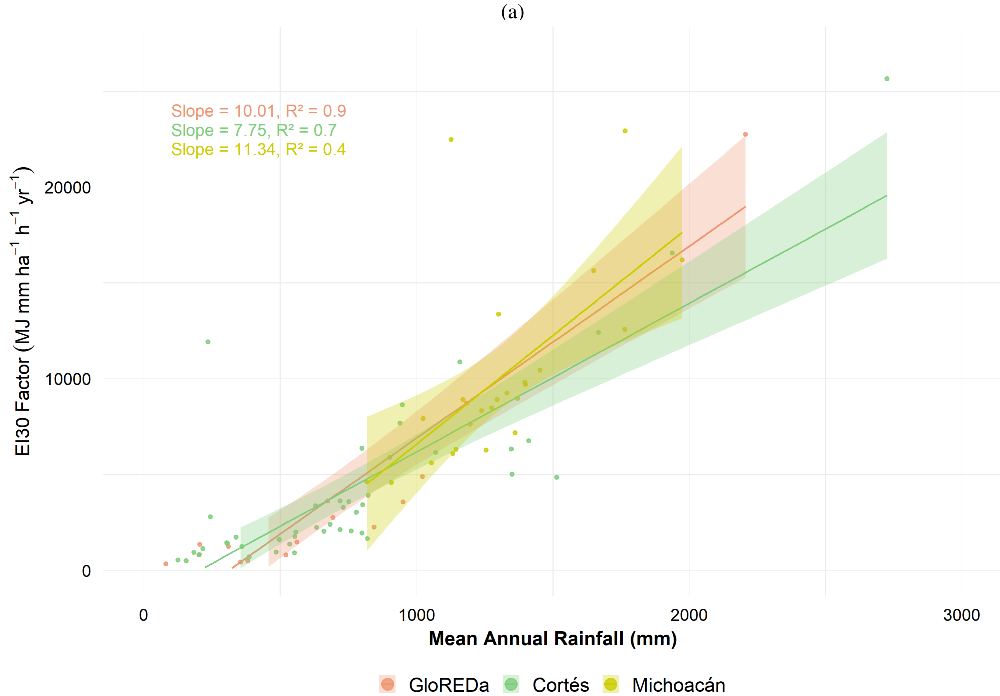
Eco	Group	RS	inht	dz.max lower	dz.max upper	dz.min lower	dz.min upper
2	1	142	60	32	34	-18	-18
	2	23	20	45	50	-25	-25
	3	68	12	35	40	-30	-35
	4	40	12	30	35	-20	-25
3	1	41	30	22	24	-12	-14
	2	55	25	22	24	-14	-14
	3	130	50	24	26	-14	-14
	4	71	40	24	26	-12	-14
	5	16	30	22	24	-10	-12
5	1	42	40	30	35	-25	-30
	2	13	50	18	20	-10	-12
	3	30	40	18	20	-12	-14
	4	32	40	18	20	-16	-18
	5	68	50	22	24	-16	-18
	6	52	30	20	22	-12	-14
6	1	62	20	30	30	-20	-25
	2	267	80	30	35	-25	-30
	3	117	25	30	30	-25	-30
	4	20	10	40	45	-35	-35
7	1	87	60	20	22	-12	-14
	2	62	60	35	40	-25	-30
	3	157	80	26	28	-12	-14
	4	105	70	26	28	-22	-24

1988-2017

Eco	Group	RS	inht	dz.max lower	dz.max upper	dz.min lower	dz.min upper
2	1	112	100	40	45	-20	-25
	2	12	10	35	40	-25	-30
	3	45	20	35	40	-20	-25
	4	67	50	45	40	-30	-35
3	1	115	60	26	28	-16	-16
	2	27	20	22	24	-14	-14
	3	180	40	26	28	-18	-18
5	1	80	60	22	24	-20	-22
	2	81	40	24	26	-16	-18
	3	75	100	24	26	-16	-18
6	1	290	200	35	40	-25	-30
	2	46	20	30	35	-20	-25
	3	85	80	35	40	-30	-35
	4	33	15	40	45	-35	-40
7	1	242	250	26	28	-14	-14
	2	59	100	35	40	-25	-25
	3	114	200	30	35	-25	-30



(a)

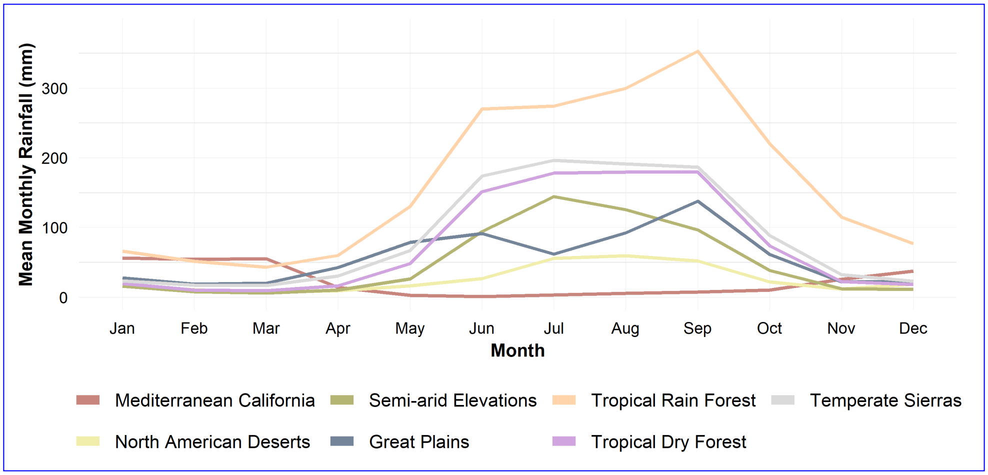


(b)

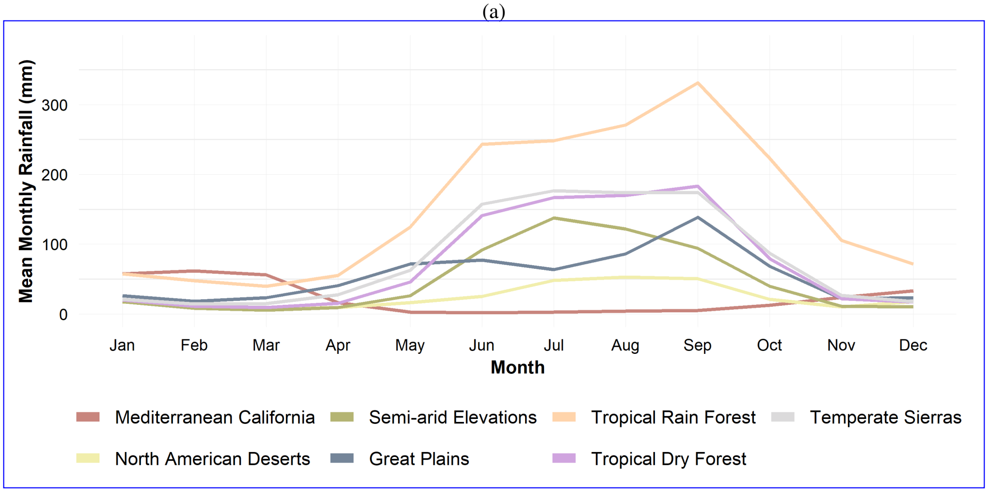
Figure A1. Characteristics of the three validations databases: GloREDa, Cortés, and Michoacán. a) Density plot of $EI30$ values; b) Linear relationship between $EI30$ and the mean annual rainfall

Table A4. ~~Identification of number of rainfall time series with consecutive zero~~ Monthly climatological deviation introduced by the gap-filling process by month and ~~NA values~~ecoregion (metric RMSE in mm). 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)

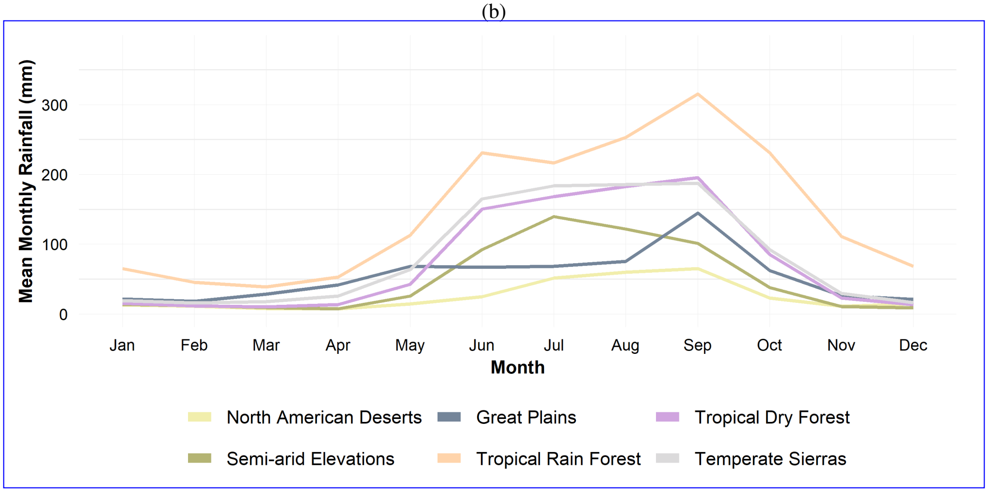
Climate Normal	Eco	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sept	Oct	Nov	Dec	Total
	1	6.84	6.71	6.10	1.36	0.33	0.16	0.67	0.58	0.60	0.78	1.74	3.43	2.44
	2	1.91	1.39	1.02	1.31	1.37	2.20	6.32	3.42	3.55	2.84	1.11	2.41	2.40
	3	1.83	0.96	0.93	0.78	1.42	3.08	3.64	3.15	3.06	1.85	0.74	0.83	1.86
1968-1997	4	1.72	2.08	1.06	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
1988-2017	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**
	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**



(a)

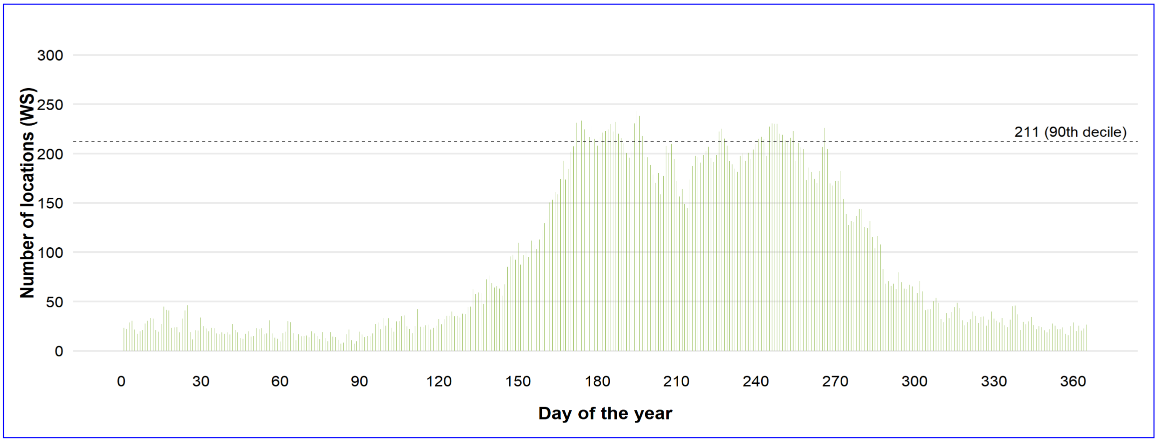


(b)

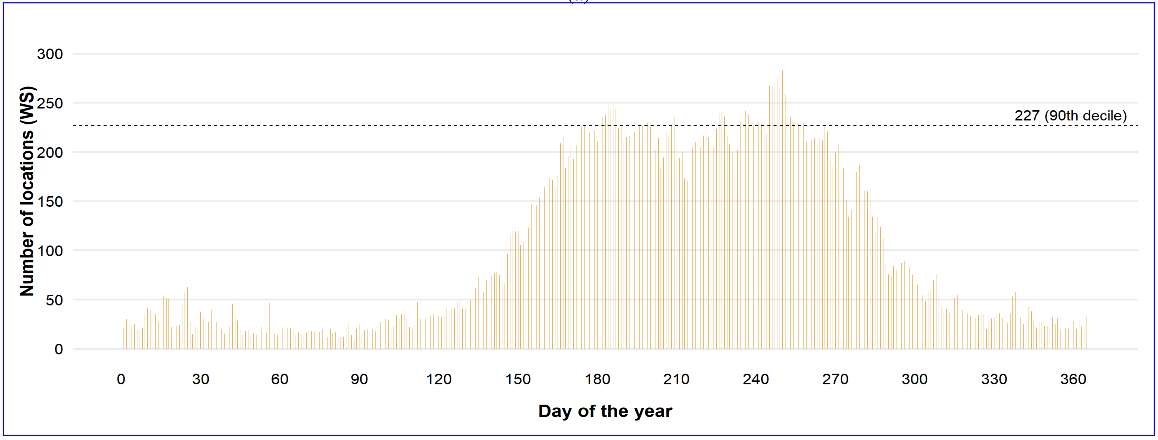


(c)

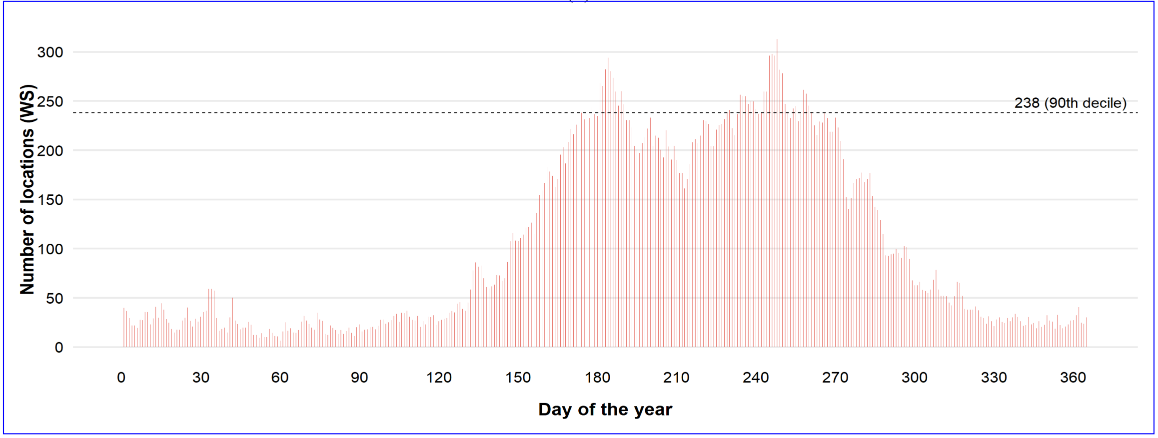
Figure A2. Mean monthly rainfall for all seven ecoregions. a) Climate normal (CN1) 1968-1997; b) Climate normal (CN2) 1978-2007; and c) Climate normal (CN3) 1988-2017



(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

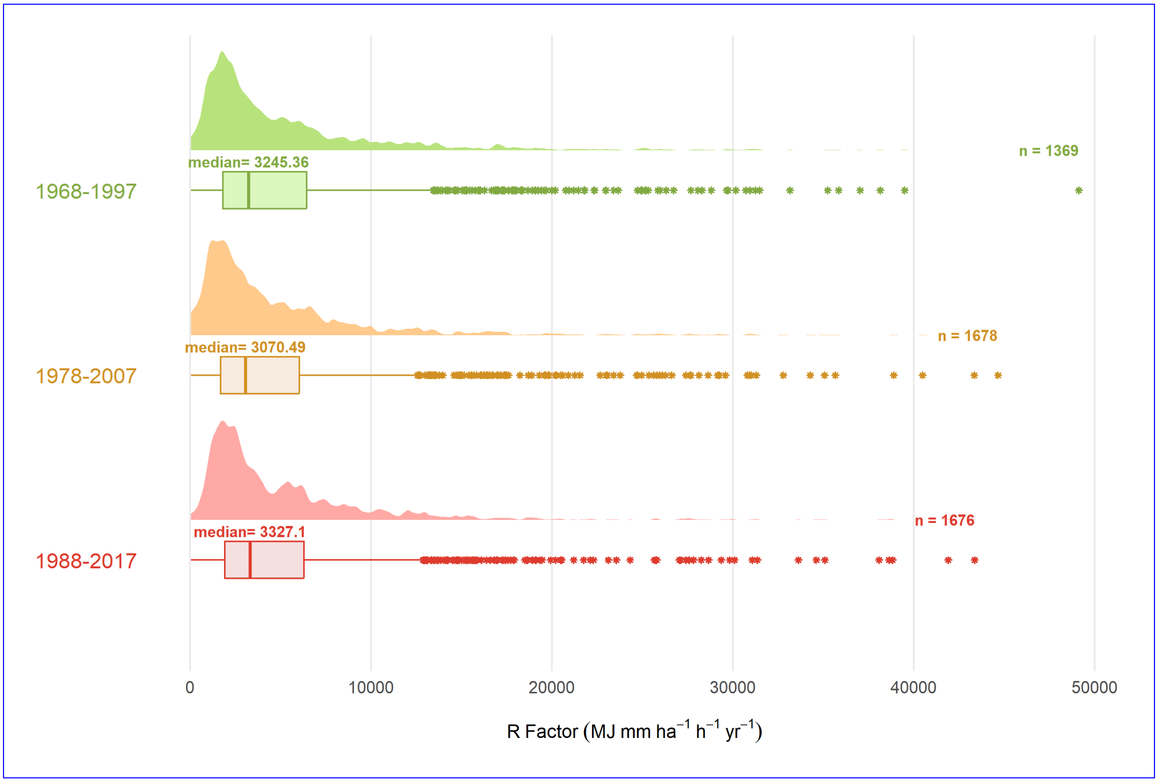


Figure A4. Density plot and Box-Plot of erosivity factor (R) calculated from daily rainfall time series for the three climate normals

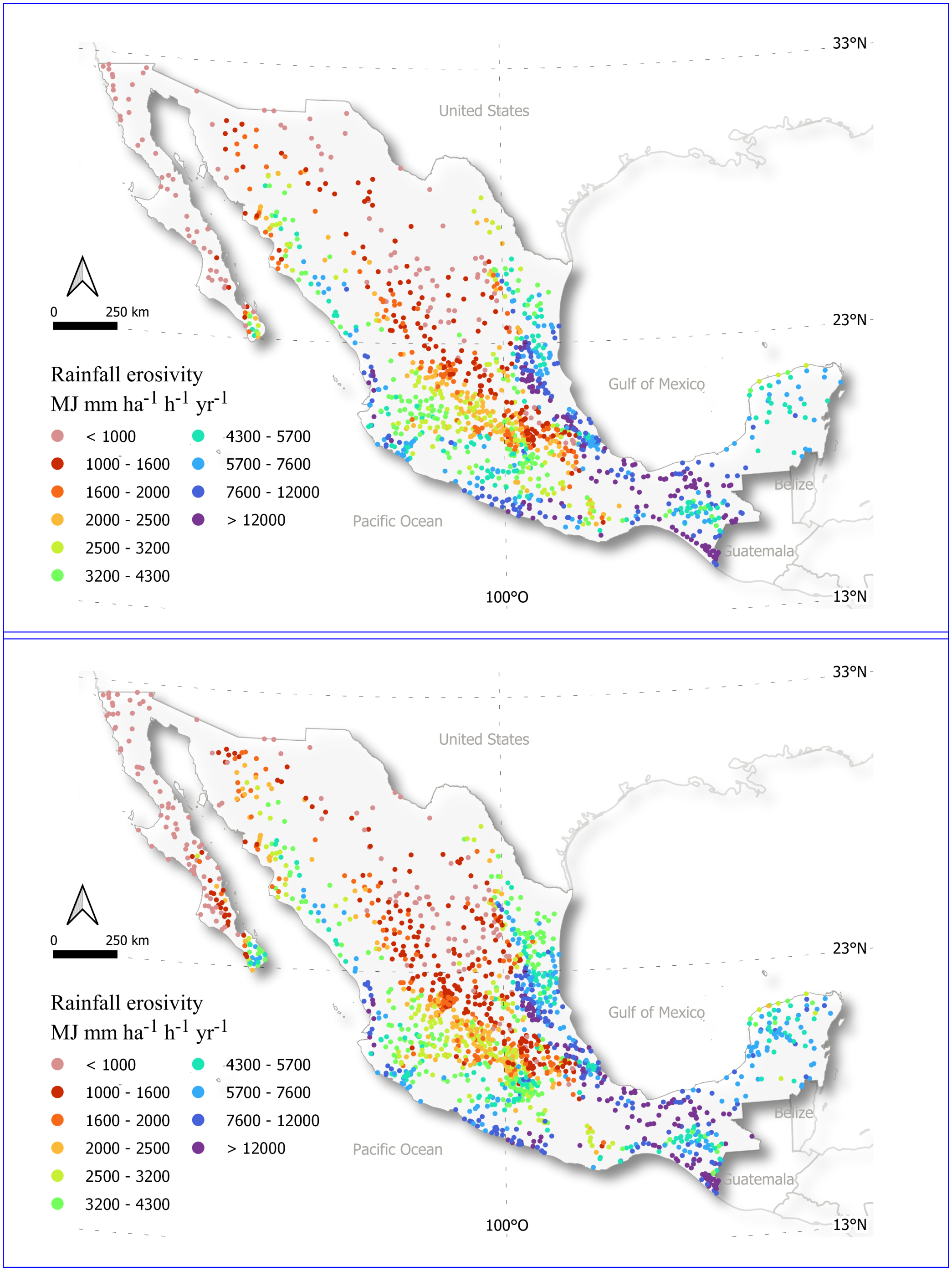


Figure A5. *R* factor values calculated with the power law equation proposed by (?) at daily resolution. Upper panel: climate normal 1968–1997. Bottom panel: climate normal 1978–2007. The legend classes approximately represent deciles of the statistical distribution, although not exactly

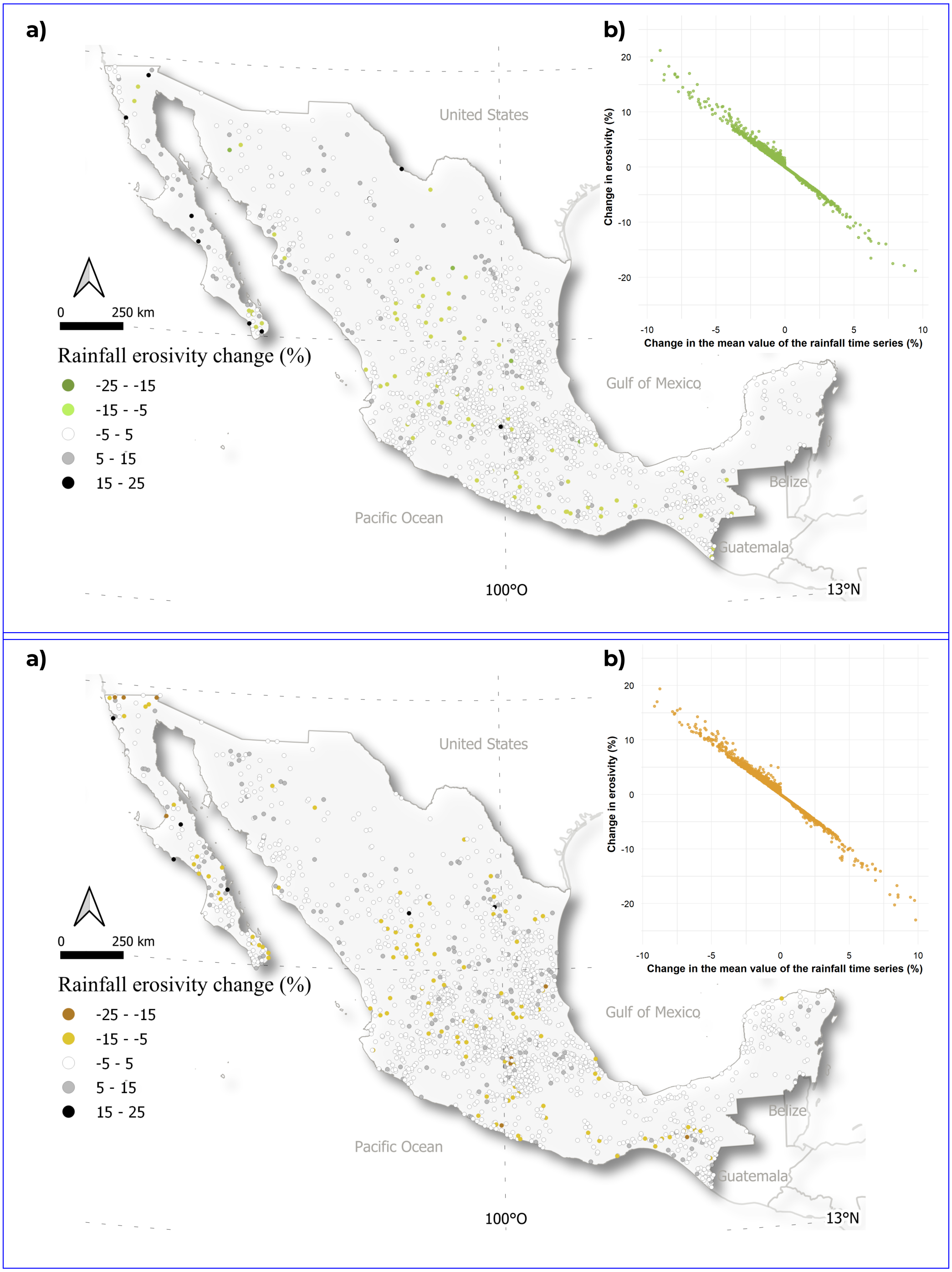


Figure A6. Percentage changes in rainfall erosivity after modifying precipitation time series according to the percentage change in the mean value of the rainfall time series before and after the gap-filling and homogenization processes. Upper panel: climate normal 1968–1997. Bottom panel: climate normal 1978–2007. a) Spatial distribution of percentage changes in rainfall erosivity. b) Relationship between percentage changes in the mean value of the rainfall time series and percentage changes in rainfall erosivity.

Table A5. Mean rainfall erosivity and mean percentage change in erosivity by ecoregion for the three climate normals. Percentage changes were estimated from modifications to the rainfall time series resulting from the gap-filling and homogenization processes. 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	Mean erosivity ($MJ\ mm\ ha^{-1}\ h^{-1}\ yr^{-1}$)	Mean change ($MJ\ mm\ ha^{-1}\ h^{-1}\ yr^{-1}$)	Mean change (%)	Minimum change (%)	Maximum change (%)
1968-1997	1	562	6.0	2.1	-4.6	16.9
	2	1133	16.0	2.3	-17.9	21.2
	3	2072	9.6	0.5	-13.1	12.4
	4	4349	54.5	1.4	-2.1	6.5
	5	14033	8.1	0.3	-11.5	6.5
	6	5627	26.3	0.6	-14.0	16.8
	7	4743	9.7	0.5	-18.8	17.0
1978-2007	1	541	-0.1	0.7	-20.3	17.0
	2	1122	20.1	1.7	-18.9	19.3
	3	2020	8.9	0.3	-13.3	12.5
	4	4240	-23.0	-0.6	-13.8	8.8
	5	12755	78.7	0.6	-10.3	11.2
	6	5628	23.1	0.4	-19.5	13.2
	7	4075	15.8	0.7	-23.0	15.2
1988-2017	2	1455	-2.8	-0.4	-18.7	17.2
	3	2134	13.1	0.3	-16.3	13.0
	4	4184	-6.9	-0.4	-17.6	9.7
	5	12019	26.4	0.1	-6.7	16.6
	6	5942	60.9	0.7	-16.9	17.6
	7	4566	36.5	0.6	-19.3	16.2

Appendix B: Description of the methods and materials used by ?

This supplementary material presents translated excerpts from the Materials and Methods section of the master's thesis by ?, Caracterización de la erosividad de la lluvia en México utilizando métodos multivariados, developed at the Colegio de Posgraduados (Montecillo, Mexico) in 1991. These excerpts describe the procedures used to construct the rainfall erosivity dataset employed in this study as a validation dataset. Because the original thesis is only available in physical format and written in Spanish, the most relevant sections were translated into English to facilitate understanding and reproducibility for non-Spanish-speaking readers. The translated material includes the procedures for pluviogram reading, rainfall event definition, and calculation of rainfall intensities and the EI30 index. A physical copy of the thesis is available at the Colegio de Posgraduados library, and the catalog information can be accessed at (http://catalogo.colpos.mx/cgi-bin/koha/opac-detail.pl?biblionumber= A complete scanned PDF version of the original thesis is available from the authors upon request.

-Pages 11, 12 and 13

III. LITERATURE REVIEW

3.3.2.1. Wischmeier Index ($EI30$)

The $EI30$ index was proposed by Wischmeier (1959) and is defined as the product of the total kinetic energy of rainfall (E) and the maximum 30-minute intensity ($I30$). It measures the effect in which erosion caused by splash detachment and flow turbulence combine with runoff to remove detached soil particles from the land surface. This process is known as sheet erosion. Its calculation is performed using the following equation:

$EI30 = (E)(I30)$

where $EI30$ is the erosivity index for a rainfall event (MJ mm ha^{-1} h^{-1}). E is the total kinetic energy of rainfall (MJ ha^{-1}). $I30$ is the maximum rainfall intensity in 30 minutes (mm h^{-1}).

The kinetic energy of rainfall is obtained using the equation:

$E = \sum_{j=1}^{n} e_j P_j$

where e_j is the kinetic energy for the time interval j (MJ ha^{-1} mm^{-1}). P_j is the amount of rainfall during the time interval j (mm). n is the number of intervals with different intensity during the same event.

The calculation of e_j in units of the International System is carried out using the equation (Wischmeier and Smith, 1978; Foster et al., 1981):

$e_j = 0.119 + 0.0873 \log_{10} I_j; \quad I_j \leq 76 \text{ mm hr}^{-1}$

$e_j = 0.283; \quad I_j > 76 \text{ mm hr}^{-1}$

where I_j is the rainfall intensity during interval j (mm h^{-1}); $I_j = \frac{P_j(60)}{t_j}$, t_j is the duration of interval j (minutes) and e_j and P_j were already defined.

The multiplication by (60) converts the data into hourly units. The value of e_j becomes constant for $I > 76$ mm h^{-1} because it is considered that, at higher intensities, the size of raindrops no longer increases (Wischmeier and Smith, 1978).

The sum of the $EI30$ values during one year forms the annual rainfall erosivity factor (R) of the Universal Soil Loss Equation (USLE), proposed by Wischmeier and Smith (1978) and expressed as:

$A = RKLSCP$

where A is the average annual soil loss (Ton ha^{-1}), R is the rainfall erosivity factor (MJ ha^{-1} h^{-1}), K is the soil erodibility factor (Ton ha h MJ^{-1} mm^{-1} ha^{-1}), L is the slope length factor (dimensionless), S is the slope steepness factor (dimensionless), C is the crop management factor (dimensionless), and P is the support practice factor for erosion control (dimensionless)

The algebraic expression of R is therefore:

$R = \sum_{i=1}^{m} (EI30)_i$

where R is the rainfall erosivity factor or annual erosivity index (MJ mm ha^{-1} h^{-1} yr^{-1}) and m is the number of rainfall events during the year.

The $EI30$ has proven to be an efficient rainfall erosivity index for estimating soil loss in different parts of the world, although originally developed for the region east of the Rocky Mountains in the United States. In other cases, however, (Lal, 1979; Roose, 1979; Hudson, 1981), the correlations between $EI30$ and soil loss have been low, and better results have been found when considering other approaches, thus obtaining other erosivity indices.

-Pages 40, 43, 44 and 45

IV. MATERIALS AND METHODS

4.1 Materials

The materials used were: daily-recording pluviograms, digitizing tablets, microcomputers, and floppy disks for microcomputers. The statistical analysis of the information was carried out using the SAS, STATGRAPHICS, and GEOEAS statistical packages. The latter, perhaps less widely known than the previous two, is a geostatistical package (Geostatistical Environmental Assessment Software: GEOEAS) developed by the U.S. EPA Environmental Monitoring Systems Laboratory in Las Vegas, Nevada, in cooperation with the Department of Applied Earth Sciences at Stanford University. It is useful for interpolation studies because, in addition to allowing the calculation of basic statistics and scatterplots of variables, it generates variograms for one or more variables in bidimensional space, enables validation of such variograms, and based on them, kriging can be performed (estimating values for a grid of points). It also has the capability to generate contour maps of isovalues for the variable of interest. Harvard Graphics was also used in the preparation of some figures and graphs. The programs used for data acquisition and information processing were written in BASIC and PASCAL programming languages.

4.2 Methodology

4.2.1 Data acquisition, organization, and classification

The development of the present work was based on the analysis of pluviograms provided by the National Meteorological Service (SMN). The number and names of the stations are presented in Table 4.2.1, and their location can be seen in Figure 4.2.1.

The selection of stations was based on the assumption that a greater number of years analyzed would produce more reliable results, as shown by the experience of studies conducted in different parts of the world related to rainfall erosivity. Thus, it was found that although there is no number of years defined as a minimum, there is consensus (Stocking and Elwell, 1976; Moldenhauer, 1980) that a period equal to or greater than 20 years would be desirable.

Finally, the determining factor appears to be data availability, and as expected, developing countries generally do not have the same number and equipment of meteorological stations available in more developed countries. While Koolhaas (1977) presents an isoerodent map (lines connecting points with equal mean annual rainfall erosivity) for Uruguay based on the analysis of pluviograms from a single meteorological station (and estimated values for another 100 stations using a regression equation), the isoerodent map for the United States (Koolhaas, 1977) was obtained from the analysis of 181 pluviographs (estimating values for another 1700 locations using a regression equation).

Based on the above, and also considering that the work to be carried out is laborious, it was decided to analyze an uninterrupted 11-year period of rainfall data from 1977 to 1987, inclusive. However, the number of stations in Mexico meeting this requirement is quite limited (approximately 20 were found within the SMN), and their spatial distribution across the national territory is deficient. Consequently, it was necessary to make considerations regarding the minimum number of years of available information (pluviograms) that each station should fulfill. Therefore, it was decided to accept 51 stations that had a minimum of four consecutive years of data records. Finally, the Ceicades-CP station (53) in Cárdenas, Tabasco, with only one year of information, and Ocozocuatla, Chiapas (52), with six non-consecutive years of information, were included because the sample coverage in that area was very poor and it was desirable to have some reference data for the region. It was also found that the stations of Santa Rosalía, Baja California Sur; Cuernavaca, Morelos; Puerto Ángel, Oaxaca; and Tapachula, Chiapas, did not strictly meet the aforementioned restrictions, since they lacked one month of data (Cuernavaca lacked two months) within the four-year historical period. Nevertheless, despite the above, these stations were also included in the analysis.

In total, information from 53 meteorological stations was used. Table A.1 in the Appendix presents the years analyzed for each station.

4.2.2. Reading of pluviograms

For the reading of pluviograms, all rainfall events were considered, regardless of how small the precipitation depth was; therefore, every event recorded by the pluviograph was taken into account in the study.

To perform the reading process, the PLUVIOGRAMAS program in Applesoft BASIC for the Apple II microcomputer was used. In combination with a digitizing tablet, the program allowed the coordinates of the pluviograph chart to be read and subsequently performed the required calculations (the program can be consulted in the work of Cortés, 1987).

To define an individual rainfall event or storm, the criterion proposed by Wischmeier (1959) was followed. In developing the EI30 erosivity index, he found better results when a rainfall event was considered independent from another only if there was a rainless interval of 6 or more hours between them.

4.2.3. Procedure for calculating maximum rainfall intensities and erosivity indices

The calculation of maximum rainfall intensities for different duration periods (5, 10, 15, 30, 45, 60, 90, and 120 minutes) and erosivity indices was performed using the previously mentioned PLUVIOGRAMAS program. The procedure used to determine maximum rainfall intensities is presented below.

The different intensities within a particular rainfall event were determined using the time (t) and rainfall depth (h) coordinates at the points where the slope of the curve traced by the pluviograph changed. Using these intensities (I), an array was generated to obtain the maximum intensities for durations of 5, 10, 15, 30, 45, 60, 90, and 120 minutes. The calculation is described by the following algorithm.

The maximum intensity $(Imax)$ for a duration of T_i minutes is:

$$Imax_{T_i} = \frac{1}{T_i}\left[\sum_{i=1}^{n-1}(I_i d_i) + I_n\left(T_i - \sum_{i=1}^{n-1} d_i\right)\right]$$

Where i is the position of the intensities (I) selected together with their respective duration (d). Example: $i = 1$ is assigned to the highest intensity, $i = 2$ to the second highest intensity, and so on in descending order of magnitude, searching above and below the highest intensity. n is the position of the last intensity that must be selected because its respective duration added to the previous one exceeds T_i (or is equal to T_i). Thus, the duration taken from this last intensity is the number of minutes necessary to complete T_i.

For the calculation of erosivity indices, the corresponding equations are incorporated into the program, as described in Section 3.3.2 of the Literature Review. The indices are calculated for each storm event; therefore, by summing the values occurring within a month, the monthly value is obtained; if all the values for a year are summed, the annual value is obtained.

~~Root Mean Square Error (mm) of the data gap-filling process by month and ecoregion. Eco: Ecoregion (1: Mediterranean California, 2:~~

~~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~~

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

~~R factor values calculated with the power law equation proposed by (?) at daily resolution for the climate normal 1968 - 1997 and 1978-2007.~~

. 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.

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

. VMVR and CEAC acknowledged support from SECIHTI scholarship (PhD level,CVU: 1240028, 1268803, respectively). VMVR, BLPP, and MG acknowledge support from grant CF-2023-I-1846 of the National Council of Humanities, Sciences, and Technologies; abbreviated CONAHCYT. VMVR acknowledges the Corporación Colombiana de Investigación Agropecuaria (AGROSAVIA) for granting a study leave. Authors acknowledge support from grant IGCP-765 and from UNAM PAPIIT IG101325. 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~~.