

Responses to the Reviewer

1. Significance

This manuscript presents the first national-scale daily rainfall time series and rainfall erosivity database for Mexico, covering three climate normals (1968-1997, 1978-2007, and 1988-2017). Mexico lacks publicly available sub-hourly rainfall data at the national level, which makes this contribution genuinely unique: no comparable effort exists in terms of spatial coverage, daily temporal resolution, rigorous quality control, and length of the study period for this country.

Response: We thank the reviewer for this positive and encouraging assessment of our work.

The database has broad potential uses: as input to the USLE/RUSLE model for national-scale soil loss estimates, as a reference for validating global products (GloREDa, GloRESatE), for climate trend studies, hydrological modelling, territorial planning, and soil conservation. The authors explicitly identify these applications, which is appropriate.

Response: We appreciate your comment on the novelty of this dataset, particularly in terms of the rigorous quality control applied to legacy climate data in a context where sub-hourly observations are not available. We also appreciate the reviewer's acknowledgement of the broad range of potential applications of our database. Moreover, we believe that it has the potential to guide and encourage similar efforts in other developing countries.

Two aspects deserve attention regarding significance:

The coverage of the Mediterranean California ecoregion is very limited (absent in CN3, very sparse in CN1 and CN2). The authors acknowledge this, but it is not clear to what extent this limits the database's utility for that region.

Response: We agree that the coverage of the Mediterranean California ecoregion requires clarification regarding its implications for dataset usability. In our case, the limitation is restricted to CN3 (1988–2017), for which no rainfall time series met the selection criteria. Therefore, rainfall erosivity estimates are not provided for this period. However, for CN1 and CN2, there are sufficient rainfall series to support the estimation of rainfall erosivity and to characterise its spatial and temporal patterns (from 1968 to 2007). Thus, the limitation in Mediterranean California is primarily temporal (2007 to 2017) rather than spatial, and does not affect the applicability of the dataset for the earlier climate normals. We have clarified this distinction in the revised Discussion section to explicitly indicate that while CN3 estimates are not available for this ecoregion, CN1 and CN2 provide a robust basis for assessing rainfall erosivity.

The most recent period available ends in 2017. Given that the SMN source data were available up to 2022, the authors should explain in more detail why the climate normals are not extended or why a more recent supplementary period is not included.

Response: Although we accessed data from the SMN in 2022; the data did not cover the year 2020. In the original file (format: .csv, one for each station), we present the raw information of each

station. For example, “location”, “name”, and “emission date”, which corresponds to the date when the rainfall time series was uploaded to the SMN platform. These emission dates were no later than June 2020. In conclusion, the rainfall time series has a delay with respect to the date of publication.

Likewise, the last date is June 2020; many stations lack continuous or sufficiently complete records. So, extending the analysis beyond 2017 would have substantially reduced the number of stations meeting the selection criteria (less than 20% of NA and no NA or zero values sequences longer than 5 years).

In the revised methodology section, we have clarified the selection of climate normals and the dates of record for the weather stations.

2. Data Quality

The methodological workflow is robust: quality control, homogeneity analysis (Alexandersson standard normal homogeneity test), data gap-filling (proportions method using the climatol package), and subsequent validation (McCuen test, 10% threshold). The use of a WMO-recommended framework (WMO 2020, 2023) is a significant strength. RMSE estimates for the gap-filling process are provided by ecoregion and month (Table A4).

Response: Thank the reviewer for pointing out that using the WMO-recommended framework is a significant strength of our work. In this contribution, we are providing a new database with quality control, homogeneity analysis and data gap-filling to promote the correct use of legacy climate data in Mexico.

However, the following issues must be addressed:

2.1 Uncertainty propagation

The highest RMSE values correspond to the Tropical Rain Forest ecoregion (up to 12.11 mm in October for CN3) and the Great Plains in CN3 (13.17 mm in July). These values are considerable. The authors should discuss their implications for the quality of erosivity estimates in those ecoregions, given that erosivity depends non-linearly on precipitation (power model with $\beta \approx 1.81$). No uncertainty estimates propagated from the gap-filling error to the final R factor are provided. This would be an important addition for end users.

Response: The RMSE presented in Table A4 reports an aggregated monthly climatological deviation after the gap-filling procedure. This RMSE was calculated to show the discrepancy between the mean monthly precipitation from the incomplete series and the mean monthly precipitation from the complete series computed by station and then summarised by ecoregion. Then, Table A4 reports an aggregate consistency metric at the monthly scale. It does not represent a direct uncertainty for daily rainfall imputation. Likewise, because rainfall erosivity is calculated on a daily scale, it is not possible to estimate the implications of the RMSE values for rainfall erosivity.

Although the RMSE might seem a bit high for both regions, it must be interpreted within the specific context of the homogenization process of climatic series, and not as an indicator of the quality of the dataset. The imputation, within the quality control framework for daily precipitation climate series, is based on the method implemented in CLIMATOL, which estimates missing values by comparing them with a reference series constructed from the weighted average, using the inverse of the distance, of the adjacent stations. So, the RMSE is calculated against raw values with inhomogeneities, not against the absolute truth.

We have included these modifications in the revised manuscript:

- 1) **Methodology:** We have included a description of the calculation of this RMSE and its meaning.
- 2) **Discussion:** We have provided a contextualization metric because high RMSE does not necessarily imply high relative distortion as the impact depends on the monthly rainfall magnitude in each ecoregion. Thus, we are going to include a relative monthly deviation (equal to RMSE/mean monthly rainfall) to show it. For example, the RMSE value in October for the Tropical Rain Forest ecoregion is 12.11 mm; the mean monthly precipitation for this month is 230.9mm, representing a deviation of 5.4% with respect to the original precipitation series.
- 3) **Appendix:** We have changed the caption of Table A4 to “Monthly climatological deviation introduced by the gap-filling process (metric RMSE)” to avoid confusion.

2.2 Numerical inconsistencies

Several numerical inconsistencies in the RTS counts must be resolved:

The abstract reports 1,678 RTS for CN2, but Table 1 shows a total of 1,679, and Section 3.1 states 1,676. This inconsistency appears in multiple locations and must be corrected and unified throughout the manuscript.

For CN3, the abstract cites 1,676 RTS, while Table 1 indicates 1,683, and Section 3.1 also states 1,683. The authors must review and reconcile these figures.

Table A3 shows "RS for data gap-filling process" as 1,479 / 1,776 / 1,723, whereas Table 1 shows 1,479 / 1,774 / 1,721 as "RTS before data gap-filling". This discrepancy of 2 units for CN2 and CN3 is unexplained.

Response: We recognise some inconsistencies in the presentation of the total number of rainfall time series across methodological moments. The correct reduction sequence in the number of available rainfall time series for each climate normal (CN) is as follows:

Methodological moment	CN1 (1968-1997).	CN2 (1978-2007)	CN3 (1988-2017)
Rainfall time series with less than 20% of missing values	1489	1785	1728

Rainfall time series after removing those with sequences of NA and zero values longer than five years	1479	1774	1721
Rainfall time series after the filling gap procedure, homogenization and quality control (McQueen test: no changes in the Rainfall time series mean value greater than 10%)	1369	1678	1676

Those final rainfall time series are already available in the EDI repository for the precipitation time series as well as the rainfall erosivity estimations. We have corrected and ensured consistency in the rainfall time series numbers throughout the manuscript.

2.3 Data accessibility

The data are deposited in the Environmental Data Initiative (EDI) repository with a permanent DOI (<https://doi.org/10.6073/pasta/e0dc8bd3501f8c19bb750e853c3289cb>), and the R code is available via Zenodo. This complies with the FAIR principles. It is strongly recommended that the authors:

Include a detailed README file in the EDI repository describing variables, units, missing value conventions, and known limitations.

Response: Thanks for your recommendation. We agree. We will update the EDI repository, highlighting the information indicated by the reviewer. The current data repository includes a detailed description of variables, units and missing values. The EDI initiative is different from other repositories (e.g., Zenodo), where no feedback is provided by the repository staff, ensuring more complete information for appropriate use and for avoiding misinterpretation by users.

We will improve our data availability section in the revised manuscript by precisely indicating how to identify variable descriptions, units, and missing values in the EDI repository. Additionally, we are going to update our EDI repository, and we are going to include a README file.

Clarify whether gap-filled values are distinguished from original observations in the released files (e.g., via a flag column indicating imputed vs. observed values). This is essential for users conducting trend analyses.

Response: Thanks for your comment. We agree. We will include a data gap-filling report (imputed vs observed) in the new version of the data repository. We will label the gap-filled data in a new column as requested. We are working on updating the EDI repository.

3. Presentation Quality

The manuscript is well organised and follows a logical structure. The workflow diagram (Figure 2) is

helpful for understanding the methodological process. The figures are generally of good quality. Figure 8 (R factor map for CN3) and Figure 9 (comparison with Panagos et al., 2017) are informative. However, the following aspects need attention:

Figures:

Figure 6 (verification of the three models) has four panels. The caption states that panel (d) corresponds to the Michoacán database. However, the x-axis is labelled "EI30 Factor" when the values shown are actually EI30 estimated from daily-resolution data — not from high-resolution sub-hourly data. This may cause confusion and the legend/caption should be clarified.

Response: We thank the reviewer for this comment. In panel (d), the x-axis represents EI30 values calculated from high-resolution (15-minute) RTS, as described in the methodology section (Section 2.2.2, Michoacán database). The y-axis corresponds to the R factor estimated from daily rainfall using empirical relationships.

We recognise that this distinction was not sufficiently explicit in the figure caption and may lead to confusion when interpreting the figure independently of the text. We have modified the figure's caption to remember the temporal resolution of the Michoacan database.

Figures A3a, A3b, and A3c show the number of locations with erosive rainfall for each day of the year across the three climate normals, but the text only discusses CN3 in detail. A more explicit comparative discussion among the three normals would be beneficial.

Response: We agree that a more explicit comparison among climate normals would be beneficial for the discussion. We will expand the comparison of climate normals, highlighting common patterns and similarities in their temporal distributions.

Language and abbreviations:

Lines 99-100: "The country is located between latitudes 14°W and 32°N and longitudes 86°W and 118°W" -latitude values cannot be expressed in °W. This appears to be a typographical error (should read °N).

Response: This mistake has been corrected in the revised manuscript.

The abbreviations "RTS" and "RS" are used interchangeably throughout the manuscript to refer to the same concept (rainfall time series). These must be unified.

Response: We have standardised and minimised the abbreviations throughout the revised manuscript.

References:

The reference list is broad and appropriate. However, Cortés (1991) -a master's thesis- serves as the national-scale validation database and plays a central role in the study. Given its importance,

the authors should ensure that this source is either fully accessible or that the methods used to obtain the data from it are described in greater detail.

Response: We thank the reviewer for this suggestion. In the revised manuscript, we have included, as supplementary material, the data table digitised from the thesis (i.e., the national validation dataset), along with a detailed description of the methods used by the author to generate it. We believe that providing this information enhances transparency and enables readers to replicate the validation exercise. In addition, we note that the original document is available upon request from the lead author, and we provide details regarding its availability in the university library.

4. Re-usability of the Dataset

Based on the information provided in the manuscript and the data deposited in EDI, a user with basic knowledge of R and climatology would be able to re-use the database. The code available in Zenodo facilitates reproducibility. The database columns are described in Section 7. Nonetheless, the following improvements are recommended to facilitate re-use:

Include a detailed README file in the EDI repository.

Clearly flag imputed versus observed values in the gap-filled daily time series files.

Response: We are working on updating the EDI repository by highlighting the information indicated by the reviewer. We will include a README file describing the variables, units, and known limitations. In addition, we will update the rainfall time series tables by adding a new flag column indicating which data were filled or replaced.

Summary of Recommendations

Major (must be addressed before acceptance):

Resolve all numerical inconsistencies in RTS counts across the abstract, main text, and tables.

Response: We have ensured the consistency of the number of rainfall time series at each methodological step.

Discuss uncertainty propagation from gap-filling errors to the R factor, especially for ecoregions with high RMSE.

Response: This has been included in the revised discussion section.

Correct the typographical error in latitude/longitude notation at lines 99–100.

Response: This has been corrected.

Minor (recommended):

Include a detailed README file in the EDI repository.

Response: We are working on adding a README file that describes the variables, units, and known limitations to the EDI repository.

Clarify the availability and accessibility of the Cortés (1991) dataset.

Response: We will provide the Cortés (1991) dataset in the supplementary material, along with the methods used by the author to generate the national-validation dataset.

Unify the use of abbreviations (RTS vs. RS) throughout the manuscript.

Response: As suggested by reviewer 1, we have removed the RTS and RS acronyms and replaced them with “rainfall time series”.

Expand the comparative discussion among climate normals in Appendix Figures A3a–c.

Response: We have expanded the discussion section by comparing the rainfall erosivity days among the climate normals shown in Figures A3a-c.

Improve the legend of Figure 6d to avoid ambiguity regarding temporal resolution.

Response: We have improved the figure caption of Figure 6d.