

Responses to the Reviewer

General comments

I would first like to thank the authors for the effort invested in compiling and ‘harmonizing’ multiple historical data sources, which are not always easily accessible, particularly for researchers who do not work in Mexico. I also appreciate that the manuscript focuses not so much on a purely historical climatic analysis, but rather on how the R factor (rainfall erosivity) is estimated, whose utility—and potential future users of this dataset—is a key aspect of the work.

Response: We thank the reviewer for the revision of our work and for highlighting the value of the new dataset as a solid starting point for rainfall erosivity studies in Mexico.

The manuscript “*The first rainfall erosivity database in Mexico: facing challenges of leveraging legacy climate data*” by Viviana Marcela Varón-Ramírez and colleagues provides a detailed dataset of historical precipitation time series for Mexico, applied to the estimation of rainfall erosivity. The dataset itself, as well as the calibration using several available empirical models, is interesting and represents a solid starting point. However, the applicability of the dataset would benefit from being presented more clearly, particularly in terms of its potential users and intended applications.

Response: We appreciate the recognition of the value of the dataset and its potential as a solid starting point for rainfall erosivity studies in Mexico. We agree that the applicability of the dataset could be presented more clearly. In the revised manuscript, we have strengthened the description of its potential users and applications, particularly in the Introduction and Discussion sections, to better highlight its relevance for soil erosion modelling, climate studies, hydrological applications, and environmental monitoring.

I do not have major concerns regarding the core content of the manuscript, and I have provided specific comments throughout the text that I hope the authors will find helpful and intuitive to address. My main concern—which may require more substantial work—does not relate to the dataset itself or its calibration, but rather to the discussion section. In its current form, the discussion is unsatisfactory and does not allow the reader to properly assess the potential usefulness or relevance of the dataset.

Response: We agree that the discussion section can be improved. Therefore, we have restructured it to enhance its clarity and to better highlight the usefulness and relevance of this new dataset. Taking into account the reviewer’s comments, we propose an improved scheme for the discussion section as follows:

Firstly, we briefly discuss the results and highlight their scientific and societal implications. We also discuss the interoperability barriers to achieve the main outcome.

Secondly, in a separate subsection (i.e., 4.1 Development of the first Mexican rainfall time series database), a critical analysis of the rainfall time series dataset is presented. Here, we describe the Improvements over previous datasets (higher temporal resolution, namely, daily Vs monthly), quality control, the interpretation of RMSE, and the identification of ecoregions with higher sensitivity to gap-filling. We also present a discussion of the spatial distribution of rainfall in Mexico

(Ecoregions-based) and the contribution of this new dataset (reproducible, standardised, and compliant with FAIR principles).

To avoid confusion among the multiple moments of our research, in the revised discussion section we added a subsection (i.e., 4.2 Rainfall erosivity estimates and model performance) to discuss the rainfall erosivity estimates. In this section, we also discuss the spatial and temporal patterns of rainfall erosivity revealed by the new database (e.g., differences among climate normals, spatial gradients across ecoregions, and the influence of regional climate drivers such as tropical systems and seasonal rainfall regimes) under the light of previous studies. Thus, we discuss the evaluation of empirical models for daily erosivity estimation (e.g., comparison of multiple model parameterisations, interpretation of model performances using global, national, and local validation datasets described in previous studies). We finally discuss how our results compare with national and global erosivity datasets.

Subsequently, in the revised discussion, the following section (i.e., 4.3 Limitations and potential use of this new database) is dedicated to highlighting the main applications, the scope, and the main limitations of the newly generated knowledge on rainfall erosivity. We address the use of daily rainfall instead of sub-hourly data and the potential under/overestimation of the EI30 factor. We extend the discussion on the dependence of the results on the empirical relationships used to estimate rainfall erosivity, particularly the parameters α and β and how the spatial distribution and density of available data vary across ecoregions. Here, we also highlight the main applications of the newly generated knowledge (e.g., use as input for soil erosion and land degradation assessments, contribution to environmental monitoring systems) and its potential use for validating global erosivity datasets and climate change and Earth system models.

The final subsection of the Discussion section (i.e., 4.4 Future research directions) focuses on introducing the main conclusion and the perspectives for future work. Here, we discuss the potential of integrating higher temporal resolution rainfall data to estimate rainfall erosivity. We also explore different approaches to gap-filling data and machine-learning possibilities for the digital mapping of rainfall erosivity in continental Mexico.

With this proposed structure, the results will be critically interpreted, the limitations will be clearly defined, and the valuable contribution of this new dataset will be highlighted.

The discussion needs to be completely restructured in a more organized and focused manner, selecting and developing the strongest points of the article (some suggestions are provided in the annotated manuscript). In this sense, I consider and expect that this manuscript will be accepted subject to the revisions (they fall between minor and majors) so if authors handle the chaotic way in which they currently present their results I believe this paper can make in through and be a valuable asset for people in need of R-factor data/maps, etc.

Response: Thanks for your suggestion. We agree that the clarity and focus of the Discussion section need to be improved. As described in the response to your preceding comment, we have markedly improved the Discussion section. It now highlights the key results, their implications, and the limitations of the new database.

If possible, it would be highly valuable for the authors to incorporate the suggestions provided in the manuscript so that the significant effort invested in compiling historical data for Mexico can be communicated more clearly and effectively. Revising the discussion may also require supporting it with a broader range of references than are currently included, depending on the final focus the authors choose to adopt.

Response: We have revised and fully addressed each comment in the manuscript. Likewise, we thank the reviewer for the thorough review. We are glad that the reviewer recognises the value of the new database for the national to international scientific communities. We believe that the reviewer's feedback has significantly increased the quality of our revised manuscript.

Specific comments:

Authors can follow up -in my opinion- my comments and suggestions in a better way when they check their original manuscript with an annotated comments. I hope the editor finds this suitable given this manuscript structure of a 'dataset-paper-like'.

typing errors are shortlisted within the original manuscript

Response: We also acknowledge the reviewer for the detailed revision of this manuscript. We have replied to all specific comments and questions below. All typographical errors are corrected in the revised version of the manuscript.

Title: Why don't you go for a much stronger title like? The rainfall erosivity database of Mexico (1968-2017)

Response: We have considered the reviewer's suggestion and revised the title to better emphasise the dataset and its temporal coverage. We have modified the title to "The first rainfall and erosivity database in México (1968-2017)"

Line 4. so if you are using existing historical rainfall records then you are compiling it rather than developing it, right? Is this data really developed here or rather you compile it from different sources?

Response: We agree that the term "development" may suggest primary data generation, whereas our study relies on existing historical rainfall records. However, our workflow involves not only the compilation of these records but also substantial processing steps, including quality control, gap filling, and temporal harmonisation.

We have rewritten the required sentences throughout the manuscript to ensure that we are not generating new data; instead, we are developing a new harmonised and engineered database.

Line 5. Again, writing. I suggest following a structure similar to: Because of a) (historical rainfall data) we compute b) (rainfall erosivity products).

Response: In the revised manuscript, we follow the reviewer's proposed writing structure to link the rainfall data and the resulting erosivity products, thereby improving the overall flow and comprehension of the text.

Line 6. I think writing needs to be improved for clarity. What is a methodological moment? You mean steps?

Response: Yes, we refer to methodological steps. In the revised manuscript, we have replaced the term “methodological moments” with “methodological steps” for the sake of clarity.

Line 8. I am strongly against the overusing of acronyms. This is a clear example of a bad use of acronyms in this text. Text gets harder to read and decoding them is an unnecessary effort in such cases.

Response: We agree that the use of the acronym “RTS” (rainfall time series) is unnecessary and may hinder readability. We have removed this acronym throughout the manuscript and replaced it with “rainfall time series.”

Line 8. how you define well distributed? based on what?

Response: We underscore that weather stations are distributed throughout the Mexican territory, and there are no large areas without coverage. To avoid confusion, we have removed this adjective.

Line 10. first time you mention "three authors". This is not recommended for an abstract. You need to rewrite and rethink the whole abstract and do not dedicate to much detail.

Response: We agree that the abstract should emphasize the main results and the dataset itself, rather than alluding the methodological details. In the revised version, we will rewrite the abstract to improve clarity and reduce unnecessary detail.

Line 21-24. In my opinion the outcomes -which you make available to the people in your datasets - are the strongest points to show in the abstract. Of course you do well by informing about how robust they are but these are clearly the outcomes people will be using your dataset for ... I strongly recommend you to emphasize that.

Response: We agree. We have modified the abstract accordingly to highlight the study's principal outcome, potential uses, and limitations.

Line 24. Is this really necessary to be included in the abstract??? I suggest removing it from the abstract!

Response: We agree that it is not necessary to include it in the abstract. We have removed it and provided the link to the repository directly.

Line 27. is this really a keyword of your manuscript?

Response: We have removed it and added "climate normals" and “R factor”.

Lines 32-35. more references here are necessary

Response: We have included two new key references:

Borrelli, P., Ballabio, C., Yang, J. E., Robinson, D. A., and Panagos, P.: GloSEM: High-resolution global estimates of present and future soil displacement in croplands by water erosion, *Scientific Data*, 9, 406, <https://doi.org/10.1038/s41597-022-01489-x>, 2022.

Guo, C., Chen, Y., Xia, W., Qu, X., Yuan, H., Xie, S., and Lin, L.-S.: Eutrophication and heavy metal pollution patterns in the water supplying lakes of China's south-to-north water diversion project, *Science of The Total Environment*, 711, 134 543, <https://doi.org/https://doi.org/10.1016/j.scitotenv.2019.134543>, 2020.

Line 62. you have already mentioned the need above. Rewrite or reorganize accordingly.

Response: We agree that this sentence is redundant. We have rephrased it accordingly.

Line 64. is this accurate? There is a 117-yr time series for the whole 5454 meteorological stations? Maybe you need to rephrase for accuracy.

Response: The Mexican meteorological service provides daily records from approximately 5,454 weather stations across Mexico, with time series spanning different periods from 1900 to 2020. We have rewritten this sentence for accuracy.

Line 70-87. I find that such detail is not relevant for the introduction of your manuscript. You can move this or reuse it within your methods section accordingly. I recommend to structure the introduction with the main shortcomings (you do it more-or-less already) but also to support how rainfall erosivity depends on the rainfall data in a way you provide a clear statement of what would be the uses / applications of your database when published.

Response: Thanks for your suggestion. In the revised manuscript, we have clearly stated the potential uses and applications of the new database, highlighting direct implications for estimating erosivity as a latent variable, which has not been described with sufficient detail before, to improve local-to-national estimates.

Line 100. In regard to what? Too vague. Expand.

Response: We agree that it is too vague. We wanted to highlight that Mexico exhibits complex climatic features influenced by subtropical high-pressure systems over the North Atlantic and northeastern Pacific, as well as its position relative to the Intertropical Convergence Zone. This has been clarified in the revised manuscript.

Line 119. previously you defend CNs till 2017.

Response: The data were downloaded from the official National Meteorological Service website in January 2022. Although the database includes records with an emission date extending to June 2020, an inspection of the rainfall time series indicates that the period with the most complete and consistent data spans from 1961 to 2017. We have rephrased this sentence for the sake of clarity.

L122-L126: You maybe want to put emphasis into what you deliver or make available rather than how you compiled the data. I.e..

Response: We agree, a great detail is not necessary. We have shortened this paragraph.

Line 126. please clearly state that from XXXX weather station you selected XXXX because of A,B, C...

Response: We accessed 5,454 files, each corresponding to a weather station's records; however, we discarded 354 files that lacked rainfall data. Finally, we obtained a set of 5,410 daily rainfall

time series, which served as the input for the first step of our methodological workflow. We have clarified the number of rainfall time series used in this initial step.

Line 138. What are NAs. if you refer to missing values you need to clarify.

Response: Yes, we refer to missing values. It has been clarified in the revised manuscript.

Line 146. what is the geographical space. Please clarify or find another chose of words.

Response: We agree; this was somewhat confusing. The first clustering refers to the grouping by ecoregion. We have rewritten this sentence and modified Figure 2 to avoid confusion.

Line 173. you don't really need a new subsection here.

Response: We agree on this point. This subsection is part of the quality control process. We have removed this subsection from the revised manuscript.

Line 184. clarify which type of experiments you refer to. odd writing. So, you have already mentioned that the power-law model proposed by Richardson is the way to go but now you mention other models without providing clear specifications nor references to them. I would suggest rewording and providing references accordingly.

Response: We refer that 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. We have rewritten this sentence and provided some references in the revised manuscript.

Line 200. you will explore further how this classification works with the one you presented in figure 1?

Response: No. The Köppen–Geiger classification is based on climatic conditions, whereas ecoregions are defined by multiple factors, including climate, topography, vegetation, and soils. In this study, we used the climate classification only to estimate rainfall erosivity following Liu et al. (2020). Subsequently, we compared the results across the three models

Line 209. Would you mind please expand on why you do more detail in this state/region?

Response: Data availability in this region is supported by previous research, which facilitated access to high-quality sub-hourly rainfall datasets. This high-temporal-resolution rainfall enables the direct calculation of EI30 following the standard definition. In addition, this region is of particular interest due to its agricultural importance (e.g., avocado production), making it a relevant case study for assessing soil erosion. We will briefly describe the interest in this state in the methodology section.

Line 262. irrelevant here

Response: We have revised the paragraph to improve clarity. Nevertheless, we consider it important to emphasise that snow cover in Mexico is minimal (0.004% of the total area of Mexico) and falls outside the scope of this study.

Line 264 - 265. already said above

Response: We agree, this is redundant. We have removed it from the revised manuscript.

Line 279. It is worth to mention that drylands always face such problems. References would help here.

Response: Indeed, this kind of problem has also been reported by <https://doi.org/10.5194/essd-17-6911-2025> in the Mexican territory. We have added references to support this statement.

Line 283. again, typical of drylands... So maybe worth mention that you are no capable of providing R estimates to such regions of Mexico? You have declined areas with snow so maybe here same strategy? Worth mention it specifically ... I'd say...

Response: We agree that the Mediterranean California ecoregion, which can be considered part of Mexico's dryland environments, presents limitations in data availability for the most recent climate normal (CN3, 1988–2017). In this case, no rainfall time series met the selection criteria, and therefore, rainfall erosivity estimates are not provided for this period. However, this ecoregion represents a small proportion of the Mexican continental area (approximately 1.3%), and for the earlier climate normals (CN1 and CN2), a sufficient number of rainfall series (11 and 16, respectively) were available to estimate rainfall erosivity and characterise its temporal behaviour.

Therefore, the limitation in Mediterranean California is primarily related to data availability for CN3 rather than a general lack of representation of dryland regions. In contrast, other arid and semi-arid ecoregions, such as the North American Deserts, Semi-arid elevations, and Great Plains, are well represented in the dataset, with a substantial number of weather stations. However, we will clarify these aspects in the discussion section and explicitly state that erosivity estimates for CN3 in Mediterranean California are not provided, but users can use CN1 and CN2 to assess rainfall erosivity in this region.

Among the other comments in the PDF, we will revise the writing structure to improve specificity and clarity. We have strengthened statements that require additional supporting citations. We have ensured that table and figure captions are complete and clearly describe all variables presented

Figure 4. can you make points of the legend bigger thanks! Also is hard to find light blue points so maybe find another color. Also 1988-2017, 1968--199 and 1978-2007 are not easy to identify. Maybe combine circles with other geometries?

Response: We have improved this figure. We have made the points in the legend bigger, and we have used a different shape (triangle) to describe the weather station with rainfall time series for the periods 1988-2017, 1968--199 and 1978-2007

Figure 4. if its alsways 0' you can easy get rid of minutes here. sampe for the following maps.

Response: We have changed the coordinate format for all maps.

Figure 5. I think its okey to keep this in the manuscript but I would move fig5 to the supplementary materials.

Response: We agree, this figure has been moved to the supplementary material.

L304-L318. This section is super hard to read. It's not clear to me whether the models have different slopes for different datasets or for the same dataset. Which model performs worst for each dataset and why? Can you expand also what "the slope furthest from the perfect model" means? This is so ambiguous. Probably need complete rewriting for clarity.

Response: We agree. This paragraph could be improved by organising its principal ideas. The main objective for this paragraph is to clarify which model performs better for each dataset. We built linear models to identify the rate of change between the EI30 factor and the R factor estimated with the three models. In this context, a slope of 1 indicates perfect agreement between estimated R factors and EI30 values, whereas deviations from 1:1 slope reflect less agreement. We have rewritten this section improving its clarity.

Line 321. This is based on what. Provide the reference, no?

Response: This threshold is based on Wischmeier and Smith (1958). However, as mentioned in previous responses on the revised methodology section, we have removed it from the results section.

Figure 6. But this you don't mention it on Lines 242-244... I need to find them in figure 6? That's odd. You need to reorganise the manuscript so these references are previously reported in the right places.

Response: The figure caption has been rewritten. We already explained each model in the methodology section (Lines 190 to 205).

Line 345. Odd writing. In the same sentence all ecoregions and also except for the semi-arid ... You really need to make a better effort to explain in a clear and simple way your results. Rewrite L340-LL346 for clarity please

Response: We agree that the original sentence is unclear and difficult to follow. In the revised manuscript, we have rewritten this section to improve clarity and readability by simplifying the sentence structure.

Line 347. Ok I do a similar comment here. Between lines 347 and 356 you need to restructure the information and make very clear your point. Now is a bunch of numbers and decils that are mixed up with regions and R values. I suggest you rethink a little how you want to present this because in its present way is hard to read.

Response: We recognise this paragraph is hard to read. We have rewritten this paragraph to make it clearer.

Figure 8. Make these circles larger please

Response: We have enlarged the circles in the legend of Figure 8.

“L358–L381: This section is not necessary in its current form and should be removed or substantially revised. The discussion should focus on the key outcomes of the analysis and the limitations of the proposed approach. It could be structured, for example, by ecoregions, climate normals, or another logical framework, at the authors’ discretion; however, the discussion requires a complete rewrite.”

Response: Thanks for your comment. We substantially revise the mentioned lines and, in the revised version, focused them on the main outcome, considering the logical order proposed by the reviewer. We also highlight the main limitations of our work in the new discussion section, following the critical analysis of the results.

This L372-L381 can be itself a subsection within discussion where you focus on previous work on R estimates... and why/how your results differ or are similar to...

Response: We agree. As mentioned above, in the revised manuscript, our first subsection of the discussion section is focused on the improvements of our new dataset over the previous one. So, we have extended and discussed in depth the statements presented in lines 372-381.

L385 to L 392 can complement the previous comment / subsection within discussion.

Response: We agree. These lines, which refer to a comparison of filling-gap methods, can complete the subsection on improvement over previous datasets. We have rephrased these sentences accordingly.

Line 399. ok there is a potentially new subsection here for the discussion where you discuss spatial rainfall patterns (historically) with other/previous works? If yes you will need to list all references not stay only with the one you have now.

Response: In the revised discussion section, we have included a subsection that describes (4.1 Development of the first Mexican rainfall time series database) spatial rainfall patterns and compares them with previous studies.

Line 401. You see... this discussion section is so confusing... Now you evaluate models. If so, okey, but provide a new subsection that focus on that... you decide what suits you the best for structuring it but ... in its present form is very very hard to read.

Response: We recognise that the order of ideas in the current version of the discussion section has abrupt changes. In the revised manuscript, the discussion sections are clearly defined and logically ordered to improve the text clarity and cohesion. We have considered all constructive feedback from the reviewer and improved the overall quality of our study.

Line 424. ok now a bit of limitations... so you have it here and there but you need to restructure the discussion ...clearly.

Response: In the discussion section of the revised manuscript, we have included a subsection (4.3 Limitations and potential use of this new database) that clearly describes the principal limitations of our datasets.

Lines 425-429. which author you refer to now? Please clarify

Response: We have rephrased this paragraph and clarified the statements' corresponding references.

Line 441. OK this is a result not a discussion... You provide clues of areas/regions with high R values that's fine but thats a result. Here (in the discssusion) you should either focus on the

methodological constraints into your results or provide interpretations to your results... Using references, previous works to compare to... etc.

Response: We agree. We have removed this sentence from the discussion section. With the proposed discussion structure, we go deeper into the methodological constraints and the interpretation of previous results.

Line 506. Conclusions. Rewrite based on all previous comments and possibly new or clearer focus of the publication.

Response: In the revised manuscript, we have rewritten the Conclusions to improve their clarity and alignment with the paper's new structure.