

Comment on **essd-2022-87**

Pia Benaud (Referee)

Referee comment on "*AWESOME: Archive for Water Erosion and Sediment Outflow MEasurements*" by Jinshi Jian et al., Earth Syst. Sci. Data Discuss.,
<https://doi.org/10.5194/essd-2022-87-RC1>, 2022

General Comments:

Soil erosion by water is a complex process, resulting from the interaction between a number of environmental variables and land management decisions, making it challenging to predict rates of soil erosion. Testing soil erosion models against empirical data is therefore essential to improving modelling accuracy. However, accurately measuring rates of soil erosion at local or national scales is incredibly costly. Collating existing data into an open access database is therefore a very useful endeavour and the authors have clearly gone to considerable effort compiling data, amassing results from 1985 different geographic sites, along with useful records of environmental and study variables. However, the usefulness of such a database for modelling applications will inevitably be limited by the reliability of the data, so rigorous quality control is crucial. Consistently extracting data from studies is incredibly challenging, particularly when there is a lot of inconsistency in the methods used to collect the empirical data. Unfortunately, there needs to be some improvements to the database in that regard and revisions to the manuscript before the article and database should be accepted for publication.

For full transparency, I recently compiled and published a similar database, however it was focussed on soil erosion observations in the UK (Benaud et al., 2020). Accordingly, I focussed my assessment on the quality of the UK observational data, and noticed a number of discrepancies. Brazier et al., (2001), for example, is not a first-hand soil erosion study. Through not seeking the original research, they have mis-reported details. For example, records 3832-3848 are not runoff plot data, they were volumetrically estimated from overflight surveys of the regions listed. The work of Chambers and Garwood (2000) is also incorrectly reported – the measurement method was a field survey, not a gauging station, and the results are therefore not sediment yields. Some of Walling et al. (2002) is indeed based on Caesium-137 observations, but 4526-4538 are suspended sediment yields. More work needs to be done to properly interrogate the data sources to match the claim on line 250.

The manuscript needs more information on the decisions surrounding how results were

standardised for inclusion in the database. For example, Walling et al. (2002) report 'gross' and 'net' erosion rates, as frequently done with Caesium-137 studies, and the database contains the gross rate. While, Chambers and Garwood (2000) report both the mean and median 'net' erosion rates, and the database contains the mean rate. The "UnitsConverter" tab suggests an assumed bulk density of 1.5 g cm^{-3} was used to convert volumetric measurements, why was this decision taken? Some of these types of discrepancies are inevitable when compiling such a large dataset, however there needs to be a clear description of the rationale behind these decisions in the manuscript. The manuscript would also benefit from the results being contextualised with other soil erosion studies – either modelling or other databases.

Specific comments:

Introduction

The introduction does a good job explaining the background and need for the database, but is currently too long – paragraphs 3 and 4 could be condensed. At the risk of promoting citation of my own work, I'd suggest it would be appropriate to cite Benaud et al. (2020) as a key aim of the paper was to collate all available, empirically-derived soil erosion datasets into a spatially explicit and open access resource, albeit with a UK focus. AWESOME is the first database that I know of that does so on a global scale, with as much detail. It would be useful to outline the aims/objectives of the paper.

Methods

This contains lots of useful information, however, as above, more information is needed on the decisions taken when selecting or standardising soil erosion measurements. I assume you log-transformed (line 284) the data to account for the skewed nature of soil erosion observations?

Results

Figure 2 needs to be updated – it's trying to display too much information. Number of observations binned by region (see Fig1 Garcia-Ruiz et al., 2015) or whichever factor you think is most important would be more appropriate.

Figure 3 – boxplots would be more appropriate here to show the true distribution of the data. I also don't think it is fair/useful to compare erosion pins with runoff plot measurements here.

Line 341 – Given how skewed soil erosion data is, it would be appropriate to also report median rates.

3.4 – This section seems a strange inclusion in the manuscript – it's not introduced elsewhere in the manuscript. Details need to be added to the methods, and you need to explain why you have done this work. You also need to quantify "the two datasets matched each [other] well"

Discussion

The discussion covers some interesting points, but needs more work to improve its quality/value. At present, it misses the context provided important existing soil erosion research papers. Some key points:

400 – Plot studies are known to create bias due to their short length and boundaries (see Parsons et al., 2006, and subsequent papers)

415 – I suggest you have a look at the original papers in more details, and consider the environment where they were tested.

Paragraphs between 425 and 435 – it's also important to consider channel bank erosion.

435 – vegetation cover also plays a significant role and the distribution of slopes in your data.

440 – this needs a better explanation

450 – Yes, the database is great for integrating with spatial datasets. The thing to consider is what the climatic data in the database represents i.e. is it say long-term average precipitation or is it precipitation during the erosion experiment.

Section 4.2 – How do your results match global models? A lot of the content here is not relevant.

Table 6 should be in the results

4.3 Great ambition! Though this section could do with being a little more formal.

Thanks for sharing the data. The R code could use some editing to make it fully reproducible – some of the data links are to local drives, for example.

The conclusion would benefit from stated aims/objectives in the introduction.

Author contributions needs some clarification.

In summary, the database represents significant effort from the authors, and could be a very useful and gratefully received resource for the soil erosion modelling community. However, there needs to be greater quality control carried out on the data – particularly looking at the original sources – and the manuscript needs to be improved to better support the database and its application.

References:

Benaud, P., Anderson, K., Evans, M., Farrow, L., Glendell, M., James, M. R., Quine, T. A., Quinton, J. N., Rawlins, B., Jane Rickson, R., & Brazier, R. E. (2020). National-scale geodata describe widespread accelerated soil erosion. *Geoderma*, 371, 114378. <https://doi.org/10.1016/j.geoderma.2020.114378>

Parsons, A. J., Brazier, R. E., Wainwright, J., & Powell, D. M. (2006). Scale relationships in hillslope runoff and erosion. *Earth Surface Processes and Landforms*, 31(11), 1384–1393. <https://doi.org/10.1002/esp.1345>