

Interactive comment on “SoilErosionDB: A global database for surface runoff and soil erosion evaluation” by Jinshi Jian et al.

Karl Auerswald (Referee)

auerswald@wzw.tum.de

Received and published: 24 February 2021

Soil erosion is a process of high temporal variability where single events may differ by several orders of magnitude. Given that the number of events per year is limited, long periods of measurement are required for obtaining reliable averages. These periods of time are usually much longer than what can be reasonably financed and maintained. Most studies on soil erosion thus suffer from too short measuring periods. Furthermore, erosion is highly variable in space. This calls for many measuring sites which also cannot be afforded. Both problems may be attenuated by combining the data of many studies. It is thus applaudable that Jian et al. took the effort to compile such a database, which presently contains data from 124 publications but which is intended to grow in the future. The key property of such a database, however, is not the number of

C1

studies included but the reliability of the data, which is difficult to achieve for different reasons originating already from the individual studies (e. g. inconsistent and incomplete variables, use of different units, differing spatial and temporal resolution) but also from insufficiencies by the compilers that often result from the tradeoff between the number of studies and the time remaining for the individual study. I will assess the quality of the database below but first I will assess the manuscript.

The title of the manuscript is rather unspecific as many types of erosion exist (to name some: water erosion, wind erosion, tillage erosion, coastal erosion, genetic erosion, bone erosion and many more). Neither the title nor the abstract tells, which one is addressed. The term runoff lets me speculate that water erosion is addressed but early in the Background the dust bowl is mentioned pointing to wind erosion. The Science Direct search (erosion, runoff) implies that all types of erosion including those that have no relation to soil are meant (note: this is not clear because in one place the authors wrote that they searched for 'erosion and runoff' while in two other places including the database they searched for 'erosion, runoff'). From the text and the database itself I got the impression that the authors looked for sheet and rill erosion (but not for gully erosion). The authors have to be precise!

The manuscript does not follow the structure of scientific manuscripts with Introduction, Material and methods, Results and Discussion but these parts are intermingled. This makes it difficult locating a specific information and likely is the reason why some information is repeated several times despite the short length of the manuscript (e. g. the address of the database is given in four places).

The introduction is called background and briefly describes erosion (many types like that caused by avalanches or animal trampling, which doesn't help to clarify the topic). Surprisingly, it does not summarise the existing data collections on sheet and rill erosion that already exist and could have easily incorporated into the new database. Likely this would also have helped to identify some problems associated with data compilation. Even without reading and reporting the existing studies, the authors conclude out

C2

of the blue that there was no successful effort to compile data from several studies into a single and coherent dataset.

I appreciate that the authors attempted to make the Chinese data available. However, from the database I learned that the authors had considerable problems of correctly understanding European studies. Russian studies were completely missing. I would recommend expanding the consortium to include a wider regional experience, which is indispensable for correctly interpreting the data. Also, given the multitude of sciences that work together and which are required (e. g. geomorphology, meteorology, agronomy in very different regional settings, soil science, vegetation science), I recommend to widen the consortium regarding the sciences as well. For instance, from the database it became clear that the authors seem not to be aware of the difference of SOM and SOC and that soil carbon is an ambiguous term because it may be organic, mineral or total carbon. These differences would be well aware to any soil scientist.

The authors decided to include data without any quality control (e.g., also papers without peer review). This invalidates the entire database even when excellent studies are included. Bad data don't become better when many of them are compiled or when they are mixed with good data but they spoil the good data.

Assessment of database: Due to my specific knowledge of the German situation, I looked at the entries for Germany. Only two studies are incorporated. This is surprising because several compilations and databases exist for Germany, which compile already the data of many studies on natural-rain plots (27 studies covering 1076 plot years compiled by Auerswald et al. 2009), rainfall simulations studies (726 simulations compiled by Fiener et al. 2011) and small-watershed data (112 watershed years with daily resolution compiled by Fiener et al. 2019).

Which two studies were selected for Germany is not clear because the database does not report the sources. Hence there is no chance to complete the data, identify errors or look for an interpretation of results that are not self-explaining. It appears indispensable

C3

that the database reports the full source information. This is already necessary to acknowledge the hard work of the experimenters. The authors of the database would also not want their database to be used without acknowledging their work.

I concentrate on the second German publication in the database, which likely is Rodrigo Comino et al. (2019); at least the truncated name of the first author in the database, the year and some data agree and support this assumption. I found no other publication that would fit to these data. The database reports two sites while the publication reports only the first one. The reason for the discrepancy remains unclear.

MAP and MAT are given in the publication but the database assigns these as study precipitation and study temperature (= wrong columns).

The database reports the biom Mediterranean, which does not exist in Germany.

The publication mentions only nine rainfall simulations while the database reports twelve replications at five sites (= 60 rainfall simulations). Only the values of the first data set agree with the publication. The origin of the other four datasets is unclear.

The publication reports SOM. The identical value is reported in the database but as SOC.

Intensity and duration of the rain is given but not the amount. Why? This makes data selection and retrieval trickier than necessary.

The variable names and the units of many variables in the database are unclear. Some (very few) examples:

Is ER_annual a multi-year mean or an individual year?

What is an interannual error (standard error, standard deviation...)?

What is ER_max? From the different unit that is used here, I expect that this is on a shorter time scale than years but which? From ER_max day I speculate that this may be the maximum of daily soil loss. All variables need much better description.

C4

ER_M_Area_h is an especially confusing variable because also ER_annual is the mass lost per area and time. Is this variable only obtained by unit conversion or is it something different?

Leaching? What is meant with leaching? Nitrate leaching to groundwater? This was already mysterious in the manuscript.

MAP and MPET carry the unit mm/yr!

From the unit of Study_precip I speculate that this is event precipitation. Or is it annual precipitation and the temporal unit was wrongly omitted as for MAP. In any case, the database user should not be forced to speculate.

Stage: is subjective also a stage?

LAI: what does 'if possible' mean?

Sand and other variables: the unit % is meaningless unless the base is given (e.g. percent of volume or of dry or wet weight, percent of bulk soil or of fine earth fraction). The definition of the soil particle sizes differs among countries. Which definition is used here? Is it consistently used? Soil carbon concentration: which carbon identity?

These were just some striking examples. Virtually the description of all variables has to be improved. I wonder how data compilation was done with such imprecise definition of variables.

In summary: I highly applaud the effort by the authors but given the many deficits, particularly in the database itself, they should better start from the scratch.

References: Auerswald, K., Fiener, P., Dikau, R. (2009): Rates of sheet and rill erosion in Germany – a meta-analysis. *Geomorphology* 111: 182–193. <http://dx.doi.org/10.1016/j.geomorph.2009.04.018>

Fiener, P.; Seibert, S.; Auerswald, K. (2011): A compilation and meta-analysis of rainfall simulation data on arable soils. *Journal of Hydrology* 409: 395–406,

C5

<http://dx.doi.org/10.1016/j.jhydrol.2011.08.034>

Fiener P., Wilken F., Auerswald K. (2019): Filling the gap between plot and landscape scale – eight years of soil erosion monitoring in 14 adjacent watersheds under soil conservation at Scheyern, Southern Germany. *Advances in Geosciences* 48, 31–48, <https://doi.org/10.5194/adgeo-48-31-2019>

Rodrigo Comino J, Iserloh T, Morvan X, Malam Issa O, Naisse C, Keesstra SD, Cerdà A, Prosdocimi M, Arnáez J, Lasanta T, Ramos MC, Marqués MJ, Ruiz Colmenero M, Bienes R, Ruiz Sinoga JD, Seeger M, Ries JB. Soil Erosion Processes in European Vineyards: A Qualitative Comparison of Rainfall Simulation Measurements in Germany, Spain and France. *Hydrology*. 2016; 3(1):6. <https://doi.org/10.3390/hydrology3010006>

Interactive comment on Earth Syst. Sci. Data Discuss., <https://doi.org/10.5194/essd-2020-283>, 2020.

C6