



***SoilErosionDB*: A global database for surface runoff and soil erosion evaluation**

Jinshi Jian¹★, Xuan Du²★, Ryan D. Stewart³, Zeli Tan⁴, Ben Bond-Lamberty¹

¹Pacific Northwest National Laboratory, Joint Global Change Research Institute at the University of Maryland–College Park, 5825 University Research Court, Suite 3500, College Park, MD 20740, USA

²Department of Hydraulic Engineering, Yangling Vocational & Technical College, Yang Ling, Shaanxi, China

³School of Plant and Environmental Sciences, Virginia Tech, Blacksburg, VA, USA

⁴Pacific Northwest National Laboratory, Richland, WA, USA

Correspondence to: Jinshi Jian (jinshi@vt.edu)

★ These authors contributed equally to this work

Abstract

Soil erosion is a major threat to soil resources, continuing to cause environmental degradation and social poverty in many parts of the world. Many field and laboratory experiments have been performed over the past century to study spatio-temporal patterns of soil erosion caused by surface runoff under different environmental conditions. However, these historical data have never been integrated together in a way that can inform current and future efforts to understand and model soil erosion at different scales. Here, we designed a database (*SoilErosionDB*) to compile field and laboratory measurements of soil erosion caused by surface runoff, with data coming from sites across the globe. The *SoilErosionDB* includes 18 columns for soil erosion related indicators and 73 columns for background information that describe factors such as latitude, longitude, climate, elevation, and soil type. Currently, measurements from 99 geographic sites and 22 countries around the world have been compiled into *SoilErosionDB*. We provide examples of linking *SoilErosionDB* with an external climate dataset and using annual precipitation to explain annual soil erosion variability under different environmental conditions.

All data and code to reproduce the results in this study can be found at: Jian, J., Du, X., Stewart, R., Tan, Z. and Bond-Lamberty, B.: *jinshijian/SoilErosionDB*: First release of *SoilErosionDB*, Zenodo, doi:10.5281/zenodo.4030875, 2020b. All data are also available through GitHub: <https://github.com/jinshijian/SoilErosionDB>.

Keywords: Surface runoff, soil erosion, nutrient leaching, database



30 1. Background

Soil is an essential natural resource for human sustainable development that is continually threatened by erosion and related land degradation processes (Borrelli et al., 2017; Poesen, 2017). Soil erosion is a geomorphic process that occurs when soil particles, soil aggregates, organic matter, and rock fragments detach from their original positions and become transported to other
35 locations (Morgan, 1988; Toy et al., 2002). Erosion is a naturally occurring process affected by both abiotic (e.g., rainfall, runoff, wind, and snow avalanches) and biotic drivers (e.g., animal trampling and tree fall), and may have significant consequences for carbon cycling (Berhe et al., 2007; Ito, 2007; Lal, 2003; Tan et al., 2020; Yue et al., 2016).

Erosion rates have been substantially increased by human activities such as cropland cultivation,
40 mining, building construction, and deforestation (Borrelli et al., 2017; García-Ruiz et al., 2015; Poesen, 2017; Vanwalleghem et al., 2017). For instance, large-scale farmland development in forested and grass-covered prairies of the United States of America led to the Black Dust Bowl of the 1930s, resulting in widespread ecosystem damage and economic losses (Schubert et al., 2004). In China, population and demographic shifts have increased demand for food and other resources,
45 leading to increases in cultivated croplands and concomitant soil erosion problems (Guo et al., 2015; Xu et al., 2010; Zhang et al., 2012; Zhao et al., 2013).

Many field and lab experiments have characterized soil erosion under different environmental conditions. Some have utilized soil erosion runoff plots, with nations and regions such as the USA (>10,000 plots), Europe (>8,000 plots), Brazil (3,525 plots) and Brazil (>2,500 plots) all having
50 widespread installations (Poesen, 2017). These experiments have led to more than 31,000 different peer-reviewed studies (searched at <https://www.sciencedirect.com> using “Erosion and runoff” as keyword). However, there has not yet been a successful effort to compile the data from those studies into a single and coherent dataset.

To bridge this research gap, we developed a global soil erosion database (*SoilErosionDB*) for
55 standardizing and compiling historical soil erosion related measurements together. The database can be used to support evaluation and parameterization of global soil erosion models, statistical modeling, non-point pollution evaluation, as well as cropland management recommendation. It can also be used to perform synthesis analyses, such as meta-analyses, and may inspire future efforts to better understand spatial and temporal patterns of soil erosion.



60 2. Methods

We designed the *SoilErosionDB* following FAIR protocols, i.e., Findable, Accessible, Interoperable, and Reusable (Wilkinson et al., 2016). All data, quality assurance/quality control (QA/QC) code, and analysis code are immediately available through a GitHub repository (<https://github.com/jinshijian/SoilErosionDB>), and each release will be issued a DOI through
65 Zenodo to ensure reusability. The version format follows an “x.y.z” format, where *x* is the major version number, *y* is the minor version number, and *z* is the patch number. We update the major version number only if the database changes its structure; we expect this to happen at an approximately decadal time step. We update the minor version number whenever the database has a significant data update; this usually happens at annual time steps. The patch number will be
70 updated relatively often, whenever the database has an important documentation update or data correction. We also made efforts to ensure interoperability so that *SoilErosionDB* could easily link to external datasets. For an example of linking *SoilErosionDB* to an external climate dataset please see “4. Linkages to external data sources” section below.

2.1 Publication collection

75 Publications were collected during an online literature search using “runoff, erosion” as keyword in ScienceDirect (<https://www.sciencedirect.com/>). We had no restrictions on literature types, i.e., both peer-reviewed articles and no-peer-reviewed articles such as theses, dissertations, and conference collections were included. We initiated our search on January 10, 2020 and found 31,235 papers, with an increasing number of published papers by year (**Figure 1a**). The following
80 criteria were used to determine whether a article should be included in the *SoilErosionDB*: (1) measurements were measured in the field, at the laboratory with rainfall simulation experiment, or from indirect methods (**Table 1**); (2) soil erosion was reported in units that could be converted to $\text{t ha}^{-1} \text{ year}^{-1}$ or $\text{g m}^{-2} \text{ hour}^{-1}$; and (3) articles were published in English or Chinese language journals after 1960. We included no other filtering criteria or restrictions to the literature. Note that we did
85 not include a constraint for leaching data because a variety of leaching types (e.g., soil organic carbon loss, organic matter losses, and total nitrogen loss) were reported in papers, with those measurements reported in different units.

2.2 Database structure design

The *SoilErosionDB* (i.e., “*SoilErosionDB.xlsx*” file in the GitHub repository) has 12 data sheets, and the core part is the “*SoilErosionDB*” data sheet, with 18 columns for soil erosion, surface runoff, and nutrient leaching records (**Table 2**) and 73 columns for background information (**Table 3**). The “DataBase_fields” sheet describes all 91 columns in the “*SoilErosionDB*” data sheet. The



“UnitsConverter” sheet contains a ‘units converter’ to standardize all surface runoff and soil erosion measurements into the same unit (i.e., soil erosion in units of t/ha/yr or g/m²/hr; runoff in units of mm/yr or mm/hr). The “CountryCode” sheet holds the international country code for the usage of Site_ID. The “Slope” sheet provides the converter of transforming slope from % to °. The “Quality_flag” sheet describes the quality control flag of measurements collected from papers (see **Table 4** for details). The “Meas_method” sheet describes soil erosion measurement methods reported in papers (see Table 1 for details). The “Biome” sheet describes different biome types. The “IGBP” sheet describes all 20 International Geosphere–Biosphere Programme (IGBP) (Townshend, 1992) vegetation types reported in papers. The “Manipulation” sheet includes description and comments about 17 manipulation types used in the *SoilErosionDB* (see **Table 5** for details). The “ReferenceList” sheet holds all reference details for all papers we compiled into the *SoilErosionDB*; and the “LiteratureSearch” sheets describes literature search details for the *SoilErosionDB*, such that users can reproduce the literature search results based on the description.

We read through each publication and compiled measurements and background information into *SoilErosionDB*. Currently we have collected and processed data from 124 papers that included measurements taken between 1980 and 2017 (**Figure 1b**). Each column in *SoilErosionDB* corresponds to either background information, surface runoff, soil erosion, or nutrient leaching indicator. When sites’ location (latitude and longitude) was not reported, we estimated site coordinates according to the site name or the maps provided in the paper. For the soil erosion-related indicators, i.e., surface runoff, soil erosion, and nutrient leaching, data were either directly read from tables or digitized from figures. We used Data Thief (version III) (Flower et al., 2016) whenever we had to obtain the values from figures. Replications and standard deviation (SD) information were usually directly obtained from the original papers, however, sometimes confidence interval (CI), coefficient of variation (CV), or standard error (SE) was reported rather than SD, we calculated SD using equations 1-3 of (Jian et al., 2020a).

2.3 Surface runoff, soil erosion, and nutrient leaching measurements

The field, unit, and explanation about the surface runoff, soil erosion, and nutrient leaching measurements are presented in **Table 2**. It should be noted that *SoilErosionDB* has been designed to hold surface runoff, leaching and soil erosion measurements in terms of both annual amounts and instantaneous rates. However, nutrient leaching was organized in a different way, where the “Leaching” column holds values, the “Leaching_unit” holds the unit of measurement reported in the original paper, and the “Leaching_type” column records the leaching type reported in the original paper.

2.4 Background information



Background information (**Table 3**) includes descriptive data about sites and experimental design. Soil erosion measurement methods, quality control flag, and manipulation are further described in **Tables 1, Table 4, and Table 5**. Specifically, Table 1 describes 16 soil erosion measurement
130 methods reported in literature; Table 4 describes 10 quality control flags to help the developer record necessary information for quality control. Table 5 describes manipulation information, which is useful for the further analysis about how treatment affects surface runoff and soil erosion.

3 Technical validation

We carefully checked the data with the original paper to ensure the fidelity. We used the Mendeley
135 bibliography management software (<https://www.mendeley.com>) to ensure papers were not compiled into the database multiple times by different contributors. Each paper was first carefully read by the data collector, and any useful records were compiled into *SoilErosionDB*. Then a data quality checker compared the data in the database against the original paper. Specially, we paid attention to the methods sections, figures, and tables, where most of the surface runoff, soil erosion,
140 nutrient leaching, and background information were located.

In addition, we developed an R markdown file (*ErosionDB_validation.Rmd* in the Github repository) to examine the data quality of *SoilErosionDB*. The file was created using R Version 3.6.1 (R Core Team, 2020). For the latitude and longitude inputs, we plotted sites by individual country (currently a total of 22 countries, **Figure 2**), then compared the sites with that country's
145 boundaries to ensure that no sites fell outside. For any sites that appeared to be mislocated, we went back to the original paper and corrected the coordinates in the database. For all numeric columns in the *SoilErosionDB* (except "*Unique_ID*" and "*Study_number*"), we plotted histograms for each column, and checked whether extreme values were included in the database. **Figure 3** shows an example using the histograms of annual surface runoff and annual soil erosion.

150 4. Linkages to external data sources

Potentially important climate data (e.g., temperature and precipitation) are important factors affecting surface runoff and soil erosion; however, many papers did not report that information. Therefore, we linked the *SoilErosionDB* with a $0.5^\circ \times 0.5^\circ$ resolution global climate data product (Willmott and C. J., 2000) to obtain annual temperature, mean annual temperature (MAT), annual
155 precipitation, and mean annual precipitation (MAP) based on site latitude and longitude. The MAT and MAP were calculated based on records between 1961 and 2015.

The results showed that temperature and precipitation data from the global climate dataset are highly correlated with that reported in the literature (**Figure 4**). Furthermore, we analyzed whether annual precipitation obtained from the external climate dataset can be used to explain annual soil



160 erosion variability in *SoilErosionDB*. We found that annual precipitation from the global climate dataset can explain ~7% of variability in annual soil erosion (**Figure 5**, $R^2 = 0.07$, $p = 0.01$). We presume that linking *SoilErosionDB* with other external data sources, e.g., leaf area index, vegetation type, climate type, and soil properties, can lead to increased explanatory power for spatial and temporal variability of soil erosion.

165 **5. Data and code availability**

The data and source code are available through GitHub (<https://github.com/jinshijian/SoilErosionDB>) and Zenodo (Jian et al., 2020b) (<http://doi.org/10.5281/zenodo.4030875>). The code is described in details with instructions for users. Generally, a markdown file (*SoilErosionDB.Rmd*) were created, which generated all figures
170 (Figure 1 to Figure 5) and described the analysis for this study. All the data processing and data visualization were conducted using R (version 3.6.1).

6. Usage notes

We suggest users download the data and code directly from Zenodo (<http://doi.org/10.5281/zenodo.4030875>), as Zenodo provides DOI and generates the same results
175 for all users. Another advantage of using data and code from Zenodo is that it avoids any run errors caused due to adding new measurements during database updating. On the other hand, the data and code in the GitHub are for development purposes. In addition, as new records are added to the database, output results may differ from those generated using older versions, and may even cause run errors. The users are encouraged to contact the *SoilErosionDB* development team before
180 using the data from Zenodo for analysis. We recommend that users contribute as a data quality checker is a great first step to understand the data; with the provided R code, users could explore the database as the code explained the analysis and the data in details.

7. Future directions and contribution notes

We have decided to share this work at this initial database development stage for two reasons: 1)
185 we want to receive feedback from the community about how to improve the data structure to ensure optimal usage; 2) the large number of potentially relevant papers that have been or will be published makes it important to expand the development team. Thus, we welcome and invite scientists and data users who are interested in developing *SoilErosionDB* to download the dataset and consider contributing published or unpublished data. Our long-term goal is to update
190 *SoilErosionDB* by including measurements from newly published papers every year.



Acknowledgements

Xuan Du was supported by the Yangling Vocational & Technical College, under grant number: A2019009. Jinshi Jian was supported by the US Department of Energy, Office of Science,
195 Biological and Environmental Research as part of the Terrestrial Ecosystem Sciences Program, under contract DE-AC05-76RL01830. Ryan Stewart was supported by the Virginia Agricultural Experiment Station and the Hatch Program of the National Institute of Food and Agriculture, U.S. Department of Agriculture. All the data and code to support this analysis can be found at:
200 <https://github.com/jinshijian/SoilErosionDB> and <http://doi.org/10.5281/zenodo.4030875>. This database can be used for research, academic, individual, or commercial usage, and can be sold or repackaged without written permission.

Author contributions

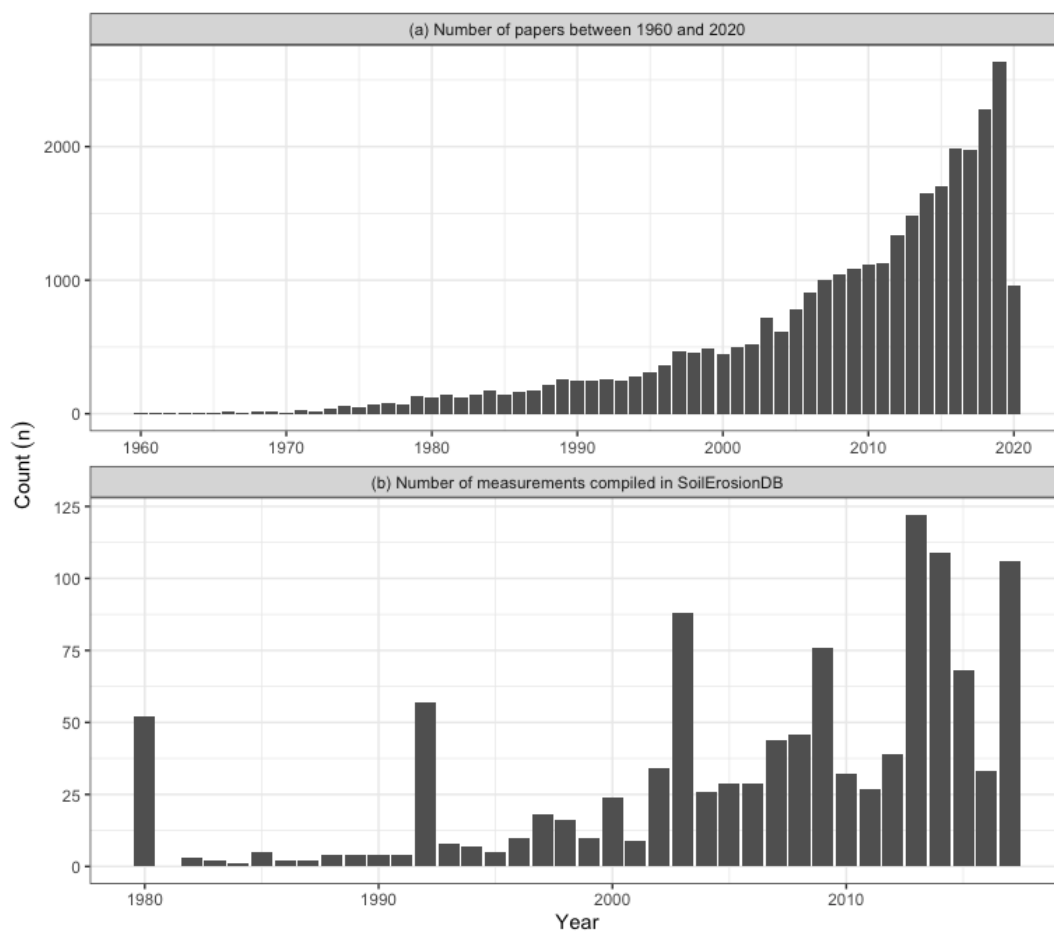
Xuan Du and Jinshi Jian conceived the design of the data framework, compiled the data from papers to the *SoilErosionDB*. Xuan Du and Jinshi Jian wrote the manuscript, and all authors revised
205 and approved the manuscript.

Competing interests

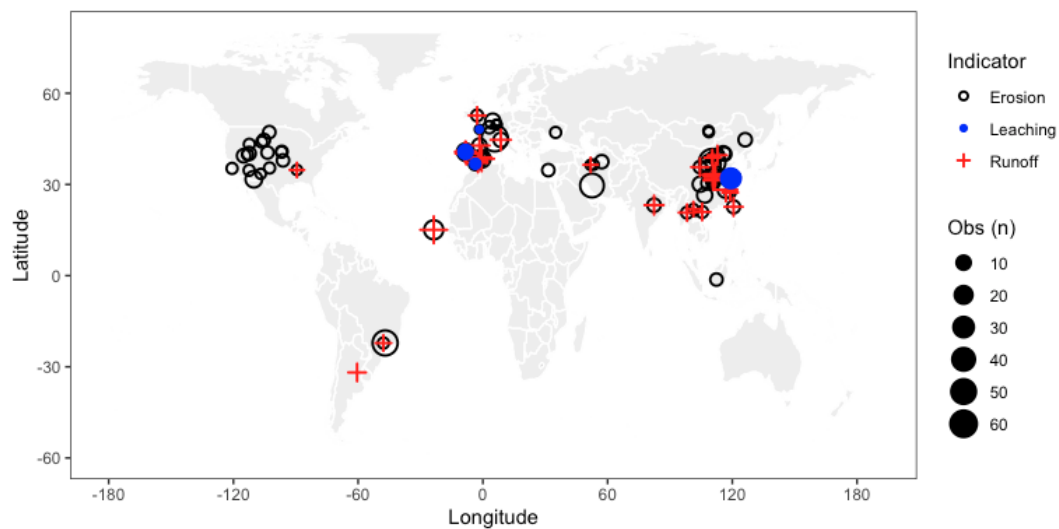
The authors declare no conflicts of interest.



Figures and Tables



210 **Figure 1.** Summary of (a) studies published on the topics of surface runoff, soil erosion, and leaching between 1960 and 2020; and (b) temporal distribution of soil erosion measurements from 124 papers compiled into *SoilErosionDB*. The trend shows that more and more runoff and erosion related studies are published.



215 **Figure 2.** Spatial distribution of surface runoff, soil erosion, and leaching sites. The size of circles represents the sample size at each measurement site (i.e., bigger circles represent more data).

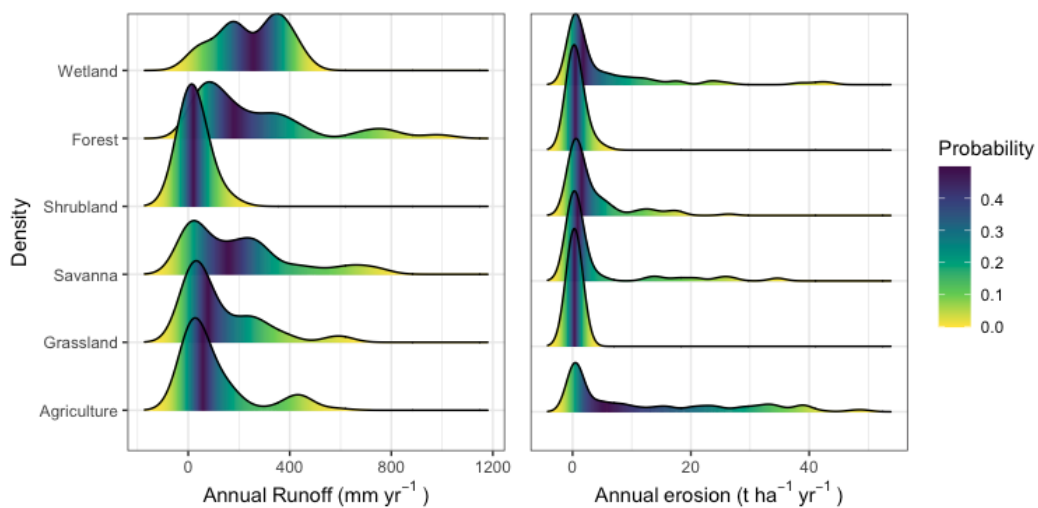
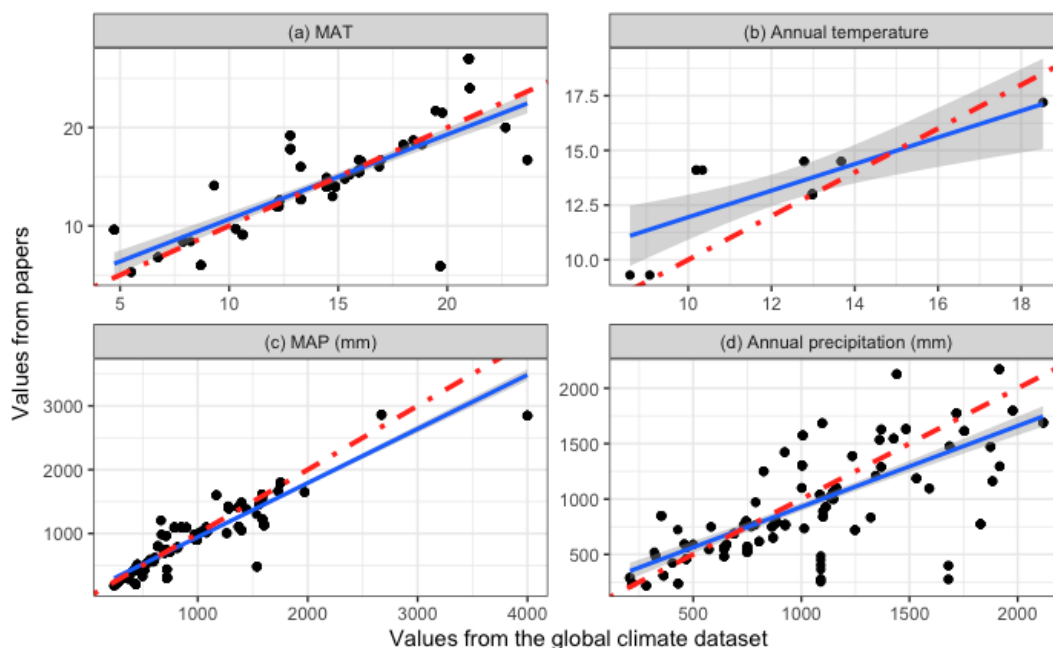


Figure 3. Distribution of surface runoff and soil erosion values in the global soil erosion database (*SoilErosionDB*). A few soil erosion data points with rates $> 50 \text{ t ha}^{-1} \text{ yr}^{-1}$ were not shown.



220

Figure 4. Relationship between (a and c) mean annual temperature/precipitation (between 1960 and 2017, MAT or MAP) reported in the papers and that from the global climate dataset (Willmott and C. J., 2000); and (b and d) annual temperature/precipitation obtained from papers vs. values obtained from the global climate dataset (Willmott and C. J., 2000). The linear regression lines (solid blue) are very close to the 1:1 lines (dashed red), indicating a good agreement between air temperature (precipitation) reported in papers and values obtained from the global climate dataset. Note that the shaded regions around the regression lines indicate 95% confidence intervals.

225

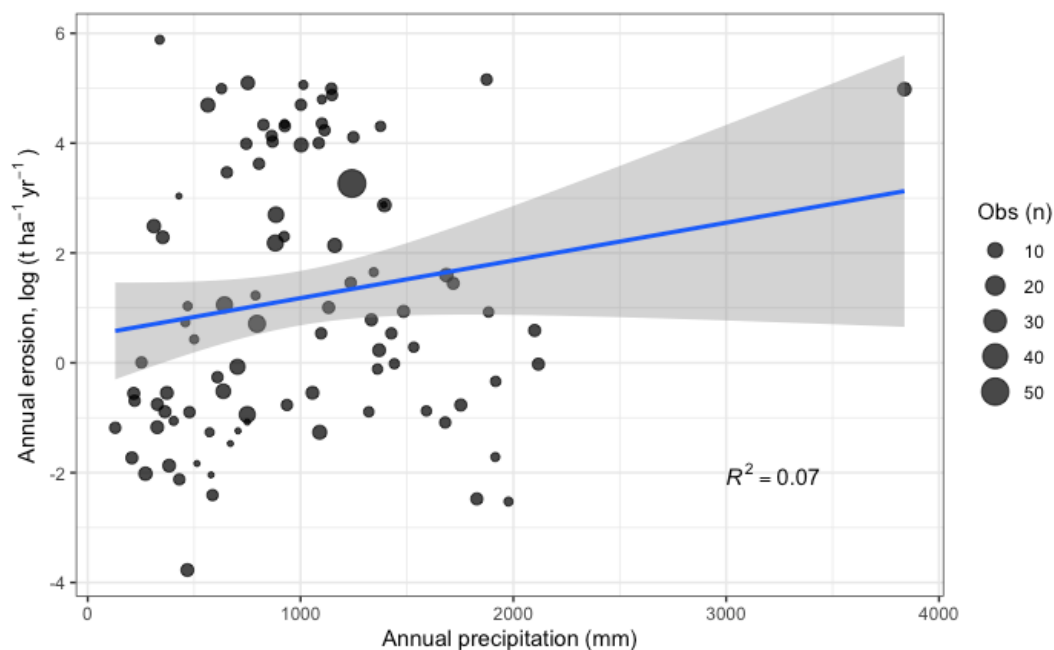


Figure 5. Relationship between annual soil erosion (log transferred, $\text{t ha}^{-1} \text{yr}^{-1}$) and annual precipitation (mm). The linear regression (solid blue line) showed that annual precipitation can explain $\sim 7\%$ annual soil erosion variability (R^2 of the regression = 0.07 and p of slope = 0.01). Note that the shaded region around the regression line indicates 95% confidence intervals. We used a Cook's distance with threshold of 0.5 to detect any influential outliers, but no data points were found to be outliers.



235 **Table 1.** Soil erosion measurement methods included in the global soil erosion database (*SoilErosionDB*).

Method_Group	Method	Explanation
Field measurement	Plot	E.g., standard runoff plot (20m × 5m)
	Morphology	Morphological transects
	Erosion pins	Erosion pins
	Profilemeters	Profilemeters
	Caesium 137	Measuring soil erosion rate using Caesium137
	Be7	Measuring soil erosion rate using Be7
	Catchment	Experimental catchment/watershed
	Reservoir surveys	Bathymetrical surveys of reservoirs
Field rainfall simulation	Field rainfall simulation	Rainfall simulation experiment in the field
Lab rainfall simulation	Lab rainfall simulation	Rainfall simulation experiment in the laboratory
Indirect methods	Model	Modelling, e.g., soil erosion rate calculated by the Revised Universal Soil Loss Equation (RUSLE) model (Foster et al., 2000; Nam et al., 2003)
	Remote sensing	Soil erosion estimated from remote sensing
	GIS	Soil erosion estimated based on GIS technology
	Topography	Topographic benchmarks related to vegetation
Other	Other	Measure methods not mentioned above



Table 2. Description and categories of surface runoff and soil erosion related metrics in the global soil erosion database (SoilErosionDB).

Field	Unit	Explanation
ER_annual	t/ha/yr	Annual soil erosion amount in unit of t/ha/yr
ER_annual_err	t/ha/yr	Annual soil erosion error (plot-to-plot)
ER_interann_err	t/ha/yr	Interannual error reported for annual soil erosion. This is sometimes reported in the article
ER_SD	t/ha/yr	Standard deviation (typically plot-to-plot) for annual soil erosion
ER_max	g/m ² /h	Maximum soil erosion rate
ER_maxday	No units	Day of year maximum soil erosion happen
ER_M_Area_h	g/m ² /hr	Soil erosion rate in unit of g/m ² /hr
ER_M_Volume	g/L	Soil erosion rate in unit of g/L
Runoff_annual	mm/yr	Annual Runoff amount in unit of mm/yr
Runoff_err	mm/yr	Error (typically plot-to-plot) for annual runoff
Runoff_SD	mm/yr	Annual runoff standard deviation (plot-to-plot)
Runoff_max	mm/h	Maximum runoff rate
Runoff_max_day	No units	Day of year when maximum runoff happen
Runoff_mm_h	mm/h	Runoff rate in unit of mm/h
Leaching	Units vary	Annual leaching amount or rate
Leaching_unit	Units vary	Annual leaching error (typically plot-to-plot) for
Leaching_SD	Units vary	Annual leaching standard deviation (plot-to-plot)
Leaching_type	No units	Leaching type (e.g., phosphorus, organic matter, organic carbon)



Table 3. Description and categories of background fields in the global soil erosion database (*SoilErosionDB*).

Field	Description	Comments
Unique_ID	Unique ID	A numeric unique ID number
Entry_date	Data entry date	When data was inputted, YYYY/MM/DD
Study_number	Paper identical ID	A numeric ID for each paper in <i>SoilErosionDB</i>
Author	First author's family name	
Duplicate_record	Duplicate sigh	If this study is a known duplicate, give the study number it repeated to
Quality_flag	Quality flag	Q0-Q10, see Table 4 for details
Contributor	Initial of data collector	Initial of data collector, e.g., JJ for Jinshi Jian
Checker	Initial of data checker	Initial of data quality check person, e.g., XD for Xuan Du
Country	Data from which country	
Region	Data from which region	Usually, province or state
Site_name	Site name of the experiment	
Site_ID	Unique Site ID within a study	It is a combination of country code (see countrycode sheet), region code, and identify code (could be site name code, manipulation code, disturbance code, etc)
Paper_year	Paper published year	
Study_midyear	Year of data measured	
YearsOfData	Number of years	Data averaged from how many years, e.g., 3 for average from 3 years, usually it is 1 year
Annual_coverage	Annual coverage	0-1, 0.01 means ≤ 1 day of data, 1 means covered at least a whole year, with at least 12 months of data
Latitude	Latitude of site	°
Longitude	Longitude of site	°



Elevation	Elevation of site	m
MAT	Mean annual temperature	°C
MAP	Mean annual precipitation	mm
Study_temp	Annual temperature	°C
Study_precip	Annual average precipitation	mm
MPET	Annual potential evapotranspiration	mm
Biome	Biome classification	Tropic, subtropic, temperate, Mediterranean, boreal, arctic etc.
STIR	Soil tillage intensity rating	
Manipulation_age	Years since manipulation	How many years since the manipulation
Manipulation	Manipulation	See Table 5 for details
Manipulation_level	Management level	
Ecosystem_age	Year	Age of ecosystem
Species	Plant species of site	Latin name of dominant species
Leaf_habit	Leaf habit of site	Evergreen, deciduous, or mixed
Stage	Stage of site	Aggrading, mature, subjective
IGBP	IGBP classification	The International Geosphere–Biosphere Programme (IGBP) classification: https://climatedataguide.ucar.edu/climate-data/ceres-igbp-land-classification
Ecosystem_stage	Ecosystem stage	Ecosystem stage (natural, managed, unmanaged). Subjective
LAI	Leaf area index	Hemispheric (one-sided) if possible
Soil_family	Soil classification	Soil family, US classification if available
Soil_texture	Soil classification	Soil texture, US classification if available
Soil_sand	Sand percentage	%



Soil_silt	Silt percentage	%
Soil_clay	Clay percentage	%
Soil_rock	Rock fragment	%
Soil_BD	Soil bulk density	g cm^{-3}
Soil_pH	Soil pH	Unitless
Soil_SAR	Sodic description	Unitless
Soil_C_%	Soil carbon concentration	%
Soil_C_stock	Soil carbon stock	g/m^2
Soil_C_Depth	Soil carbon depth	cm
Soil_N	Soil nitrogen	%
Soil_CN	Soil carbon to nitrogen ratio	Unitless
Crust%	Soil crust	Unitless
Ksat	Soil saturated conductivity	cm/h
E	Revised Universal Soil Loss Equation (RUSLE) model rainfall factor, energy (Foster et al., 2000; Nam et al., 2003)	$\text{MJ ha}^{-1} \text{ year}^{-1}$
I30	RUSLE rainfall factor (Foster et al., 2000; Nam et al., 2003)	mm h^{-1}
EI30_R	RUSLE rainfall factor (Foster et al., 2000; Nam et al., 2003)	$E \times I30, \text{ MJ mm h}^{-1} \text{ ha}^{-1} \text{ year}^{-1}$
K	RUSLE soil erodibility factor (Foster et al., 2000; Nam et al., 2003)	$\text{t-ha-h ha}^{-1} \text{ MJ}^{-1} \text{ mm}^{-1}$
L	RUSLE slope length factor (Foster et al., 2000; Nam et al., 2003)	m
Slope	RUSLE slope factor (Foster et al., 2000; Nam et al., 2003)	$^{\circ}$



Plant_cover_C	RUSLE plant coverage factor (Foster et al., 2000; Nam et al., 2003)	%
P	RUSLE plant cropland management factor (Foster et al., 2000; Nam et al., 2003)	0-1, unitless
Meas_method	Soil erosion measure method	See Table 1 for details
Field_scale	Scale of experiment field area	Catchment, watershed, plot etc.
Field_area	Area of experiment field	m ²
Measure_time	Time	When data was measured
Measure_interval	Frequency of measurement	How many minutes per measurement
Replication	Number of replications	
Rainfall_intensity	Rainfall intensity	Could be both in the field or in the lab, mm h ⁻¹
Rainfall_length	Rainfall length	How long the rainfall last, minute
Rainfall_amount	Rainfall amount	How much rainfall during the simulation, mm h ⁻¹
Notes	Notes	Other important notes about the data in this row
Data_source	Data source	Tips for where the data can be find in the paper
Other_comments	Other comments	Other important comments



245 **Table 4.** Description for the quality flag field in the global soil erosion database (*SoilErosionDB*).

Quality_flag	Description
Q0	Default/none
Q1	Estimated from figure
Q2	Data not reported in the original paper, but could be found from another study
Q3	Values estimated from a figure, with potential quality problem
Q4	Potentially useful values in the future
Q5	Values with potential problem
Q6	Data need to double check with original authors
Q7	Known problem
Q8	Duplicate
Q9	Inconsistency
Q10	Lack of useful data



Table 5. Description for the manipulation field in the global soil erosion database (*SoilErosionDB*).

Manipulation Group	Manipulation	Comments
Control	Control	default, if no manipulation can be identified
Changes in precipitation	Precipitation pattern change	e.g., same precipitation amount as control, but different intensity and duration time
	Precipitation amount change	More or less precipitation comparing with control
Fertilization	Fertilization	e.g., N addition, slurry addition, compost
Agriculture cultivation	Conventional tillage	Traditional tillage (using plough, destroy soil structure a lot)
	Cover crop	One type of conservation management
	No-till	One type of conservation management
	Reduced tillage	One type of conservation management
	Mulch	One type of conservation management
	Conservation tillage	Other conservation management other than cover crop, no-till, reduced tillage, mulch, and ridge tillage
Pollution	Pollution	Human activities related pollution, such as acid rain
PAM	PAM	Polyacrylamide application
Fire	Fire	Disturbance by wildfire or artificial fire
Multiple factors	Multiple factors	Interactive and relative effects of two or more than two manipulations
Others	Others	Other treatment not included above



250 References

- Berhe, A. A., Harte, J., Harden, J. W. and Torn, M. S.: The Significance of the Erosion-induced Terrestrial Carbon Sink, *Bioscience*, 57(4), 337–346, doi:10.1641/B570408, 2007.
- Borrelli, P., Robinson, D. A., Fleischer, L. R., Lugato, E., Ballabio, C., Alewell, C., Meusburger, K., Modugno, S., Schütt, B., Ferro, V., Bagarello, V., Oost, K. V., Montanarella, L. and Panagos, P.: An
255 assessment of the global impact of 21st century land use change on soil erosion., *Nat. Commun.*, 8(1), 2013, doi:10.1038/s41467-017-02142-7, 2017.
- Flower, A., McKenna, J. W. and Upreti, G.: Validity and reliability of graphclick and datathief III for data extraction., *Behav. Modif.*, 40(3), 396–413, doi:10.1177/0145445515616105, 2016.
- Foster, G. R., Yoder, D. C., McCool, D. K., Weesies, G. A., Toy, T. J. and Wagner, L. E.: Improvements
260 in science in RUSLE2., *Improvements in science in RUSLE2.*, 2000.
- García-Ruiz, J. M., Beguería, S., Nadal-Romero, E., González-Hidalgo, J. C., Lana-Renault, N. and Sanjuán, Y.: A meta-analysis of soil erosion rates across the world, *Geomorphology*, 239, 160–173, doi:10.1016/j.geomorph.2015.03.008, 2015.
- Guo, Q., Hao, Y. and Liu, B.: Rates of soil erosion in China: A study based on runoff plot data,
265 *CATENA*, 124, 68–76, doi:10.1016/j.catena.2014.08.013, 2015.
- Ito, A.: Simulated impacts of climate and land-cover change on soil erosion and implication for the carbon cycle, 1901 to 2100, *Geophys. Res. Lett.*, 34(9), doi:10.1029/2007GL029342, 2007.
- Jian, J., Du, X. and Stewart, R. D.: A database for global soil health assessment., *Sci. Data*, 7(1), 16, doi:10.1038/s41597-020-0356-3, 2020a.
- 270 Jian, J., Du, X., Stewart, R., Tan, Z. and Bond-Lamberty, B.: jinshijian/SoilErosionDB: First release of SoilErosionDB, *Zenodo*, doi:10.5281/zenodo.4030875, 2020b.
- Lal, R.: Soil erosion and the global carbon budget., *Environ. Int.*, 29(4), 437–450, doi:10.1016/S0160-4120(02)00192-7, 2003.
- Morgan, R. P. C.: Soil erosion and conservation, *Soil Sci.*, 145(6), 461, doi:10.1097/00010694-
275 198806000-00011, 1988.
- Nam, P. T., Yang, D., Kanae, S., Oki, T. and Musiak, K.: Global soil loss estimate using RUSLE model: the use of global spatial datasets on estimating erosive parameters., *GEOINFORMATICS*, 14(1), 49–53, doi:10.6010/geoinformatics.14.49, 2003.
- Poesen, J.: Soil erosion in the Anthropocene: Research needs, *Earth Surf. Process. Landforms*, 43(1), 64–
280 84, doi:10.1002/esp.4250, 2017.
- R Core Team: R: A language and environment for statistical computing. Computer software, R Foundation for Statistical Computing, Vienna, Austria., 2020. <http://www.r-project.org/index.html>.
- Schubert, S. D., Suarez, M. J., Pегion, P. J., Koster, R. D. and Bacmeister, J. T.: On the cause of the 1930s Dust Bowl., *Science*, 303(5665), 1855–1859, doi:10.1126/science.1095048, 2004.
- 285 Tan, Z., Leung, L. R., Li, H.-Y., Tesfa, T., Zhu, Q. and Huang, M.: A substantial role of soil erosion in the land carbon sink and its future changes., *Glob. Chang. Biol.*, 26, 2642–2655, doi:10.1111/gcb.14982, 2020.



- 290 Townshend, J. R. G. (ed. .: Improved global data for land applications. A proposal for a new high resolution data set. Report of the Land Cover Working Group of IGBP-DIS, Global Change Report (Sweden), 1992.
- Toy, T. J., Foster, G. R. and Renard, K. G.: Soil Erosion: Processes, Prediction, Measurement, and Control, illustrated., John Wiley & Sons., 2002.
- 295 Vanwalleghem, T., Gómez, J. A., Infante Amate, J., González de Molina, M., Vanderlinden, K., Guzmán, G., Laguna, A. and Giráldez, J. V.: Impact of historical land use and soil management change on soil erosion and agricultural sustainability during the Anthropocene, *Anthropocene*, 17, 13–29, doi:10.1016/j.ancene.2017.01.002, 2017.
- 300 Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., Gonzalez-Beltran, A. and Mons, B.: The FAIR Guiding Principles for scientific data management and stewardship., *Sci. Data*, 3, 160018, doi:10.1038/sdata.2016.18, 2016.
- Willmott and C. J.: Terrestrial Air Temperature and Precipitation: Monthly and Annual Time Series (1950-1996), WWW URL: http://climate.geog.udel.edu/~climate/html_pages/README.ghcn_ts.html, 2000.
- 305 Xu, X. Z., Xu, Y., Chen, S. C., Xu, S. G. and Zhang, H. W.: Soil loss and conservation in the black soil region of Northeast China: a retrospective study, *Environmental Science & Policy*, 13(8), 793–800, doi:10.1016/j.envsci.2010.07.004, 2010.
- 310 Yue, Y., Ni, J., Ciais, P., Piao, S., Wang, T., Huang, M., Borthwick, A. G. L., Li, T., Wang, Y., Chappell, A. and Van Oost, K.: Lateral transport of soil carbon and land-atmosphere CO₂ flux induced by water erosion in China., *Proc Natl Acad Sci USA*, 113(24), 6617–6622, doi:10.1073/pnas.1523358113, 2016.
- Zhang, S., Li, Q., Zhang, X., Wei, K., Chen, L. and Liang, W.: Effects of conservation tillage on soil aggregation and aggregate binding agents in black soil of Northeast China, *Soil and Tillage Research*, 124, 196–202, doi:10.1016/j.still.2012.06.007, 2012.
- 315 Zhao, G., Mu, X., Wen, Z., Wang, F. and Gao, P.: Soil erosion, conservation, and eco-environment changes in the loess plateau of China, *Land Degrad. Dev.*, doi:10.1002/ldr.2246, 2013.