
Jialei Li¹,², Hongbin He¹, Liding Chen¹,², Ranhao Sun¹,²,*

¹State Key Laboratory of Urban and Regional Ecology, Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences, Beijing, 100085, China
²College of Resources and Environment, University of Chinese Academy of Sciences, Beijing, 100190, China

Correspondence to: Ranhao Sun (rhsun@rcees.ac.cn)
Abstract. Soil conservation capacity (SC) is defined as the ability of the ecosystem to control soil erosion and protect soil function. A long-term and high-resolution estimation of soil conservation is urgent for ecological assessment and land management on a large scale. Here, a 300-m resolution SC dataset in China is established from 1992 to 2019 based on the Revised Universal Soil Loss Equation (RUSLE) model. The RUSLE modelling was conducted based on five key parameters, including the rainfall erosivity (interpolation of daily rainfall), land cover management (provincial data), conservation practices (weighted by terrain and crop types), topography (30 m), and soil properties (250 m). The dataset agrees with previous measurements ($R^2 > 0.5$ in all the basins) and other regional simulations. The results show that China's SC decreased before 2003 and then increased up to now. The SC change exhibits the ecological effects of soil and water conservation policies in China, such as the Conversion of Farmland to Forests and Grass (Grain-for-Green), which unfolded many movements after 2000. Compared with current studies, the dataset has long-term, large-scale, and relatively high-resolution characteristics. This dataset will serve as a base to open out the mechanism of SC variations in China and could help assess the ecological effects of land management policies. This dataset is available at https://doi.org/10.11888/Terre.tpdc.272668 (Li et al., 2022).

1 Introduction

Excessive soil erosion can negatively impact crop production, carbon transfer, soil organisms, and soil nutrition (Wang et al., 2017; Borrelli et al., 2018; FAO, 2019; Zhang et al., 2022a). To prevent soil erosion, many countries have attempted to improve soil conservation services, including China. China governments have made a lot of plans and policies to control soil erosion for recent years, for example, legislation for soil and water conservation, building agricultural terraces (Chen et al., 2017; Cao et al., 2021), 3-North Shelter Forest Program (since 1978) (SC, 1978), and Conversion of Farmland to Forests and Grass (Grain-for-Green) (since 1999) (SC, 2002). Recent research shows that China has led in a quarter of the global increase in greening after 2000 (Chen et al., 2019). However, China supports one-fifth of the world population with 9% of the total global cropland (SCIO, 2019), which means that the reduction of cropland has a potential to affect China grain security (Lu et al., 2013; Li et al., 2021). To weight the relationship between human and the land, a reliable dataset of soil conservation capacity in China is urgent.

Over the past few decades, China has experienced an expansion of farmland, a massive growing total population (from 987 million people in 1980 to 1.413 billion people in 2020), and a large urbanization process (Liu et al., 2005; Liu et al., 2020; Zhang et al., 2022b). Although these increasing trends have slowed down in recent years, anthropogenic activities, along with climate change, have brought about non-negligible changes in soil erosion risk and soil conservation capacity. To support the huge population, agriculture may face the challenges to be expanded or intensified (d'Amour et al., 2017). Either approach puts pressure on the land, including soil fertility and water conservation. Assessment of soil erosion and soil conservation can better understand and develop measures in the context of global change. The Chinese government has decided to conduct the third national soil census from 2022 to 2025 (SC, 2022); however, this will be an arduous task.
Development of soil erosion models based on remote-sensing data and GIS technology makes it possible to assess soil conservation capacity and its dynamic changes. By far the most widely used empirical model to simulate soil conservation services is based on the Revised Universal Soil Loss Equation (RUSLE), which predicts longtime average annual soil loss in specified experimental fields (Wischmeier and Smith, 1978; Renard et al., 1991). They describe the relationship between soil loss rate and the following effects: soil properties, topography, vegetation cover, land management, and rainfall and runoff (Montgomery, 2007). In order to apply the USLE-family model on a large scale, many researchers have made improvements and enhancements to the parameters of the model (Lufafa et al., 2003; De Asis and Omasa, 2007; Naipal et al., 2015; Panagos et al., 2015b; Panagos et al., 2015c; Barao et al., 2019). Yang et al. (2003) develop a GIS-based RUSLE model to offer a pioneering overview of global soil erosion on cell grids. Now the USLE-based models are wildly used on regional or national scale (Vallebona et al., 2016; Bamutaze et al., 2021). However, there is an uncertainty in the accuracy of global soil erosion estimation based on USLE-family models (FAO and ITPS, 2015; Batista et al., 2019; Quine and Van Oost, 2020). Considering the availability of data, model mechanism, and China's geographical conditions, we finally chose the RUSLE model.

High-precision, long-term soil conservation capacity data are requested in China. However, accurate and high-precision soil conservation capacity datasets based on field measurement are unfeasible and money-consuming on a national scale. Modelled soil conservation capacity data seem to meet demands and be more feasible. Currently, soil conservation and soil erosion data in China are mostly based on regional or watershed scales (Kong et al., 2018; Duan et al., 2020; Ma et al., 2020), only few studies are based on national scales (Rao et al., 2014; Teng et al., 2019). The current large scale is only for one year or average year data, lacking long-term series data.

Based on high-precision and long-term series data, this study calibrated the input factors and obtained the database of soil conservation capacity and soil erosion rate in China from 1992 to 2019. These results can provide data support for assessing soil erosion risk in China and drive analysis of past changes to enhance regional management. The main goal of this data is to analyze the interannual changes in soil conservation in China and provide a basis for identifying potential soil erosion hotspots. In addition, this dataset can be regarded as a comparison of soil conservation for other studies on different scales in the future. The results can provide a reference for further improvement of model parameters and lay a foundation for identifying driving factors of soil erosion changes.

2 Data and methods

2.1 Data sources

The list of the input data is shown in Table 1, and the spatial distribution of the raw data are shown in Fig. 1.
<table>
<thead>
<tr>
<th>Data name</th>
<th>Spatial resolution</th>
<th>Format</th>
<th>Temporal period</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>NASA SRTM digital elevation</td>
<td>30 m</td>
<td>Raster</td>
<td>Static (2000)</td>
<td>Shuttle Radar Topography Mission digital elevation model (SRTM DEM) data</td>
</tr>
<tr>
<td>Normalized difference vegetation index (NDVI)</td>
<td>8 km</td>
<td>Raster</td>
<td>Half-monthly 1992-2015</td>
<td>Global Agricultural Monitoring System (GIMMS) AVHRR NDVI3g datasets</td>
</tr>
<tr>
<td>Land temperature</td>
<td>0.5°</td>
<td>Raster</td>
<td>Monthly 1992-2019</td>
<td>Climatic Research Unit (CRU TS)</td>
</tr>
<tr>
<td>Land cover</td>
<td>300 m</td>
<td>Raster</td>
<td>Annual 1992-2019</td>
<td>European Space Agency Climate Change Initiative Land Cover (ESA CCI LC)</td>
</tr>
<tr>
<td>Soli properties</td>
<td>250 m</td>
<td>TIFF</td>
<td>Static</td>
<td>International Soil Reference and Information Centre (ISRIC) (<a href="http://data.isric.org/geonetwork/srv/chi/catalog.search;jsessionid=A887E5B4">http://data.isric.org/geonetwork/srv/chi/catalog.search;jsessionid=A887E5B4</a>)</td>
</tr>
</tbody>
</table>
Figure 1: Raw data to develop the soil conservation dataset. NDVI: Normalized difference vegetation index. LULC: Land use and land cover types.

2.2 Methods

We used the RUSLE (Revised Universal Soil Loss Equation) model to estimate the soil erosion rate in China during 1992-2019, and certain adjustments were made in the factor calculation according to the actual backgrounds in China. The model equations are as follows:

\[ SE_p = R \cdot L \cdot S \cdot K, \]  
\[ SE_a = R \cdot L \cdot S \cdot K \cdot C \cdot P, \]  

where \( SE_p \) is the predicted potential annual soil erosion on bare land \( (t \text{ ha}^{-1} \text{ a}^{-1}) \), \( SE_a \) is the predicted actual annual soil erosion \( (t \text{ ha}^{-1} \text{ a}^{-1}) \) on land with vegetation cover and erosion control practices, \( R \) is the rainfall erosivity factor \( (\text{MJ mm ha h}^{-1} \text{ a}^{-1}) \), \( LS \) is the topographic factor (dimensionless) with \( L \) being the slope length factor and \( S \) being the slope factor, \( K \) is the soil erodibility factor \( (t \text{ h MJ}^{-1} \text{ mm}^{-1}) \), \( C \) is the vegetation cover and management factor (dimensionless), and \( P \) is the water ...
and soil conservation measure factor (dimensionless). Every factor was calculated according to the original resolution of the input data (for example, DEM data for 30 m × 30 m, soil properties data for 250 m × 250 m), then all the factors were resampled to a resolution of 300 m × 300 m by bilinear interpolation and multiplied to obtain the soil erosion rate in each year.

In this study, soil conservation capacity (SC, t ha⁻¹ a⁻¹) is defined as water erosion prevented by vegetations and practice measures, equal to potential soil erosion minus actual soil erosion:

\[ SC = SE_p - SE_a = R \cdot L \cdot S \cdot K \cdot (1 - C \cdot P), \]

(3)

2.2.1 Estimation of the R-factor

The calculation of the R-factor is essential in the RUSLE model, because it can reflect the impact of natural rainfall on soil erosion, which is especially sensitive to climate change. Large-scale studies generally use monthly or annual rainfall data to calculate the R-factor due to the feasibility. A recent study uses hourly rainfall data to estimate the R-factor (Yue et al., 2022), which improve the accuracy of the erosivity map in China. However, it is difficult to obtain such high-precision and complete rainfall records in the long-term estimations. Here, in order to better describe China's annual rainfall erosivity, we used a daily rainfall erosivity model developed according to climate characteristics in China (Zhang et al., 2002):

\[ R = \sum_{i=1}^{24} R_i, \]

(4)
where $R$ is the annual rainfall erosivity, and $R_i$ is the half monthly rainfall erosivity. $P_j$ is the daily erosive rainfall amount on the $j$-th day during the half-month (only select the days with $P_j \geq 12$ mm, which is the threshold of a rainfall erosivity event in China (Yun et al., 2000). $P_{d12}$ represents the average daily rainfall (mm) with a daily rainfall of 12 mm or more, and $P_{y12}$ represents the average annual rainfall with a daily rainfall of 12 mm or more.

However, due to the special geological background of the karst areas in southwest China, the model overestimated rainfall erosion there (Zhu et al., 2021), so we implemented another daily rainfall erosivity model for karst areas (Xie et al., 2016). It has proven to be that the performance of this model is good in karst areas of China (Zhu et al., 2021). The equation is as follows:

$$R_d = 0.2686 \left[ 1 + 0.5412 \cos \left( \frac{\pi}{6} \left( \frac{7}{6} \right) \right) \right] P_d^{1.7263},$$

where $R_d$ is the daily rainfall erosivity in the month $j$ (MJ·mm·ha$^{-1}$·h$^{-1}$), $P_d$ is the daily rainfall (mm) in the day $d$, and the sum of the $R_d$ in a year is the annual rainfall erosivity.

The source of the daily rainfall data and the distribution of meteorological stations are shown in Table 1 and Fig. 1a, respectively. Annual values of the $R$-factor were firstly calculated for each station, then the values were interpolated to erosivity maps employing the method of Universal Kriging, which has proven to be an effective method for the spatial interpolation of the rainfall erosivity in China (Li et al., 2011; Yue et al., 2022).

### 2.2.2 Modification of the $C$-factor

The $C$-factor is closely related to the types of vegetation and crops (Panagos et al., 2015b; Borrelli et al., 2017); therefore, the values of the $C$-factor in arable and non-arable land in China were calculated separately. Some adjustments were made to the classification of crop types and the reference $C$-values of these crop types in arable land according to the actual agricultural conditions in China (Table 2). The released crops were classified into ten categories according to the main crop types in each province issued by the National Bureau of Statistics. The values of the $C$-factor in arable land ($C_d$) were calculated by the following equation:

$$C_d = \sum_{n=1}^{10} C_{cropn} \times Region_{Cropn},$$

where $C_{cropn}$ represents the value of the $C$-factor of the crop type $n$, and $Region_{Cropn}$ represents the share of the sown areas of this crop type $n$ in the total agricultural land in a province.
Table 2. The values of the C-factor in cropland ($C_{crop}$)

<table>
<thead>
<tr>
<th>Farm crops</th>
<th>n</th>
<th>$C_{crop}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grain crops</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cereal</td>
<td>1</td>
<td>0.15</td>
</tr>
<tr>
<td>Corn</td>
<td>2</td>
<td>0.38</td>
</tr>
<tr>
<td>Other cereals</td>
<td>3</td>
<td>0.2</td>
</tr>
<tr>
<td>Beans</td>
<td>4</td>
<td>0.32</td>
</tr>
<tr>
<td>Tubers</td>
<td>5</td>
<td>0.34</td>
</tr>
<tr>
<td>Sugar crops</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sugarcane and Beetroots</td>
<td>6</td>
<td>0.15</td>
</tr>
<tr>
<td>Oil-bearing crops</td>
<td>7</td>
<td>0.25</td>
</tr>
<tr>
<td>Fiber crops</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cotton</td>
<td>8</td>
<td>0.4</td>
</tr>
<tr>
<td>Other fiber crops</td>
<td>9</td>
<td>0.28</td>
</tr>
<tr>
<td>Tobacco</td>
<td>10</td>
<td>0.5</td>
</tr>
<tr>
<td>Vegetables</td>
<td>11</td>
<td>0.25</td>
</tr>
<tr>
<td>Medicinal materials</td>
<td>12</td>
<td>0.15</td>
</tr>
<tr>
<td>Succulence</td>
<td>13</td>
<td>0.1</td>
</tr>
<tr>
<td>Orchards</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grapes</td>
<td>14</td>
<td>0.35</td>
</tr>
<tr>
<td>Melons and fruits orchards</td>
<td>16</td>
<td>0.25</td>
</tr>
<tr>
<td>Other orchards</td>
<td>15</td>
<td>0.32</td>
</tr>
<tr>
<td>Other farm crops</td>
<td>17</td>
<td>0.15</td>
</tr>
</tbody>
</table>

The values of the $C$-factor in non-arable land ($C_{NA}$) were estimated by vegetation coverage (based on the NDVI data) and reference values of the $C$-factor of various land cover types. The GIMMS NDVI products were the main source of the NDVI data from 1992 to 2015 for data integrity; however, they were unavailable after 2015. To form a temporally continuous SC dataset covering 1992–2019, we decided to search for other the monthly NDVI data covering 2016–2019. We compared MODIS NDVI with GIMMS NDVI, of which the temporal span overlaps from January 2010 to December 2015. We found that the temporal fluctuations of the monthly NDVI were consistent, and their correlation was obvious ($R^2=0.95$, p<0.01) (Fig. 3). Therefore, the MODIS NDVI was selected to supplement the NDVI data after 2015. The monthly NDVI was then normalized using the following equation:

$$F_{cover} = \frac{NDVI_i - \min(NDVI_{in})}{\max(NDVI_{in}) - \min(NDVI_{in})},$$

(10)

where $F_{cover}$ is the vegetation coverage, which is the monthly average value of the NDVI during the growing season (the months with a mean surface air temperature $\geq 0^\circ\text{C}$) (Fig. 1b) (Pan et al., 2018). $NDVI_i$ means the value of the NDVI in grid cell $i$, and $\max(NDVI_n)$ and $\min(NDVI_n)$ represent the maximum and minimum values of the NDVI in year $n$, respectively.

The values of the $C_{NA}$ were calculated by (Panagos et al., 2015b; Borrelli et al., 2017):

$$C_{NA} = \min(C_{NA0}) + [\max(C_{NA0}) - \min(C_{NA0})] \times (1 - F_{cover}),$$

(11)
where $C_{NAn}$ is the reference values of the C-factor in land cover type $n$ (Table 3) (Fig. 1c), and $\text{max}(C_{NAn})$ and $\text{min}(C_{NAn})$ are the maximum and minimum value of the $C_{NAn}$, respectively.

![Figure 3: Comparison of the NDVI from GIMMS with MODIS based on temporal fluctuations (a) and correlative relationship (b).](image)

Table 3. The values of the C-factor in non-cropland ($C_{NAn}$)

<table>
<thead>
<tr>
<th>n</th>
<th>Land cover type</th>
<th>$C_{NAn}$ (maximum-minimum)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Broadleaf evergreen forest</td>
<td>0.0001-0.003</td>
</tr>
<tr>
<td>2</td>
<td>Broadleaf deciduous forest</td>
<td>0.0001-0.003</td>
</tr>
<tr>
<td>3</td>
<td>Needleleaf evergreen forest</td>
<td>0.0001-0.003</td>
</tr>
<tr>
<td>4</td>
<td>Needleleaf deciduous forest</td>
<td>0.0001-0.003</td>
</tr>
<tr>
<td>5</td>
<td>Mixed forest</td>
<td>0.0001-0.003</td>
</tr>
<tr>
<td>6</td>
<td>Tree open</td>
<td>0.01-0.15</td>
</tr>
<tr>
<td>7</td>
<td>Shrub</td>
<td>0.01-0.15</td>
</tr>
<tr>
<td>8</td>
<td>Herbaceous</td>
<td>0.01-0.15</td>
</tr>
<tr>
<td>9</td>
<td>Herbaceous with sparse tree/shrub</td>
<td>0.01-0.15</td>
</tr>
<tr>
<td>10</td>
<td>Sparse vegetation</td>
<td>0.1-0.5</td>
</tr>
<tr>
<td>15</td>
<td>Wetland</td>
<td>No data</td>
</tr>
<tr>
<td>17</td>
<td>Bare land</td>
<td>0.1-0.5</td>
</tr>
<tr>
<td>18</td>
<td>Urban</td>
<td>No data</td>
</tr>
<tr>
<td>19</td>
<td>Snow and ice</td>
<td>No data</td>
</tr>
</tbody>
</table>

2.2.3 Improvement of the $P$-factor

Support practices have significant control effects on soil erosion; however, the $P$-factor is difficult to quantify in large-scale modeling of soil erosion (Xiong et al., 2019a). The differences in $P$ values of arable land with different slopes and different
soil and water conservation measures on the basis of numerous literatures. In this study, based on Xiong et al.’s assignment method of $P$-factor and the land use types, different $P$ values were assigned. The horizontal paddy field was assigned a value of 0.2, and $P$-factor values in other arable areas were assigned according to the slope (Table 4). We only apply this method on the terraces, because support practices were mostly applied on terraces in mountainous regions (Xiong et al., 2019a; Cao et al., 2021). For other non-arable land, the value of the $P$-factor is 1.

Table 4. The values of the $P$-factor for cropland on terraces.

<table>
<thead>
<tr>
<th>Slope gradient</th>
<th>The values of the $P$-factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>$&lt; 4.6^\circ$</td>
<td>0.45</td>
</tr>
<tr>
<td>$4.6^\circ \sim 7^\circ$</td>
<td>0.52</td>
</tr>
<tr>
<td>$7^\circ \sim 9^\circ$</td>
<td>0.61</td>
</tr>
<tr>
<td>$9^\circ \sim 14^\circ$</td>
<td>0.70</td>
</tr>
<tr>
<td>$14^\circ \sim 24^\circ$</td>
<td>0.92</td>
</tr>
<tr>
<td>$&gt; 24^\circ$</td>
<td>1</td>
</tr>
</tbody>
</table>

2.2.4 Calculation of the $LS$- and $K$-factor

The $LS$-factor reflects the natural geographic influence of topography on soil erosion. The $LS$-factor consists of a slope length ($L$) factor and a slope steepness ($S$) factor, which are calculated by DEM data in GIS-based RUSLE model (Wischmeier and Smith, 1978; Renard et al., 1991). The resolution of the DEM data influences the accuracy of the simulation of the $LS$-factor, as it has significant scaling effect (Panagos et al., 2015a; Lu et al., 2020; Wang et al., 2020). Considering the trade-off between the accuracy of simulations and feasibility of calculation, we used the DEM data with a resolution of 30 m $\times$ 30 m (Fig. 1h). The formulas for calculating the $L$-factor are as follows (Wischmeier and Smith, 1978; Liu et al., 2000):

$$L = \left( \frac{l_i}{22.13} \right)^m, \quad (12)$$

$$l_i = \sum \frac{D_i}{\cos \theta_i}, \quad \sum_{i=1}^{i-1} \frac{D_i}{\cos \theta_i} = \frac{D_i}{\cos \theta_i}, \quad (13)$$

$$m= \begin{cases} 
0.2, & \beta < 1\% \\
0.3, & 1\% \leq \beta < 3\% \\
0.4, & 3\% \leq \beta < 5\% \\
0.5, & \beta \geq 5\% 
\end{cases}, \quad (14)$$

where $m$ is slope length exponent, $l_i$ is the slope length of a grid $i$, $D_i$ is the horizontal projection distance of the slope length of each grid along the runoff direction, $\theta_i$ is the angle of the slope of a grid $i$, and $\beta$ is the slope gradients in %.

The $S$-factor was calculated by following the method in the CSLE model, which is improved according to the different slope degree in China (Liu et al., 2002):

$$S= \begin{cases} 
10.80 \sin \theta + 0.03, & \theta < 5^\circ \\
16.80 \sin \theta - 0.5, & 5^\circ \leq \theta < 10^\circ, \\
21.91 \sin \theta - 0.96, & \theta \geq 10^\circ 
\end{cases}, \quad (15)$$
where $\theta$ is the slope angle in degrees.

180 Soil erodibility ($K$-factor) reflects soil erosion affected by soil characteristics, which is originally measured by establishing unit plot for each soil type (Wischmeier and Smith, 1978). Sharpley and Williams (1990) establish regression equations between the measured data of plot experiments and soil properties in the EPIC model (Sharpley and Williams 1990), which make it feasible to calculate the $K$-factor on a large scale (Pham et al., 2001; Yang et al., 2003; Scherer and Pfister, 2015). The formula is as follows:

\[
K = 0.2 + 0.3 \exp \left( -0.0256 \times \frac{SAN}{100} \right) \left( \frac{SIL}{CLA + SIL} \right)^{0.3} \left( 1 - \frac{0.25 \times OC}{C + \exp(3.72 - 2.95 \times OC)} \right) \left( 1 - \frac{0.7 \times SN1}{SN1 + \exp(-5.51 + 22.9 \times SN1)} \right),
\]

(16)

\[
SN1 = 1 - \frac{SAN}{100},
\]

(17)

where $SAN$, $SIL$, $CLA$ and $OC$ are the percentage content of sand, silt, clay and organic carbon from topsoil, respectively (Fig. 1d, 1e, 1f, and 1g).

2.2.5 Change detection

A linear regression was employed to quantify the rate of change of the $R$-factor, $C$-factor, and SC from 1992 to 2019. The least squares were applied to calculate the rate of change, and the equation is as follows:

\[
\theta_{\text{slope}} = \frac{n \times \sum_{i=1}^{n} A_i \sum_{i=1}^{n} \left( \sum_{j=1}^{n} A_j \right)^2 - \left( \sum_{i=1}^{n} A_i \right)^2 \left( \sum_{i=1}^{n} \left( \sum_{j=1}^{n} A_j \right)^2 \right)}{n \times \sum_{i=1}^{n} A_i^2 \left( \sum_{j=1}^{n} A_j \right)^2},
\]

(18)

where $\theta_{\text{slope}}$ is the regression coefficient (slope of the linear regression), representing the changing rate of the data. $A_i$ represents the data in the year of $i$. The significance of $\theta_{\text{slope}}$ was analyzed by t-test, where $p < 0.1$ indicated a significant change and $p < 0.05$ indicated an extremely significant change. The change trends of the results were classified according to their change rates and significance levels.

2.2.6 Methods for data validation

Since soil conservation capacity is calculated based on soil erosion rates estimated by the RUSLE model, its reliability can be indirectly validated by verifying simulated soil erosion (Rao et al., 2014; Xiao et al., 2017; Ma et al., 2020). However, it is different to obtain abundant measured soil erosion rates from plot experiments for large-scale validation of modelled soil erosion, so we applied a cross-comparison of the assessments with substituted observations and other regional simulations of soil erosion (Borrelli et al., 2017). Substituted observations refer to observed sediment runoff data for hydrologic stations in eight major river basins in China from 2010 to 2019 (Fig. 1i) (MWR, 2010-2015), which were compared with modeled soil erosion rates by computing the correlation coefficients ($R^2$) between the mean values of them in every basin. High values of $R^2$ usually mean satisfactory estimations of soil erosion (Xiong et al., 2019b; Jin et al., 2021; Wang et al., 2021). We also compared assessments of soil erosion rates with other regional simulations using the USLE-based models, which can verify whether our results are in a reliable range. We also presented a correlation analysis between our estimated soil conservation
capacity and other regional estimations that applying USLE-based models. The sources of collected studies are in the supplementary Table S1 and S2.

3 Results

3.1 Parameters of soil conservation in China

The distributions of the input factors are shown in Fig. 4. The values of $R$-factor in China ranges from 13.52 to 20 011.31 MJ mm ha h$^{-1}$ a$^{-1}$, with a mean value of $1 743.70$ MJ mm ha h$^{-1}$ a$^{-1}$. The high values of the $R$-factor ($>8000$ MJ mm ha h$^{-1}$ a$^{-1}$) are mainly distributed in southeast China, while low values ($<200$ MJ mm ha h$^{-1}$ a$^{-1}$) are distributed in north-western China (Fig. 4a). The average $C$-factor is 0.198 in China (Fig. 4c), and the range of the $C$-factor is from $3.00 \times 10^{-4}$ to 0.50. The $C$-factor with high value is located in regions with sparse vegetation, such as the Tarim Basin in northwest China. Regions covered by thick forests, such as hilly areas in southern China and the Greater Khingan Mountains, have lower values of the $C$-factor. The mean value of the $P$-factor is 0.90 in China (Fig. 4g). Areas assigned $P$-factor value (value $< 1$) account for 15.6% of the total study area. The mean value of the $LS$-factor is 7.62, the maximum is 947.10, and the minimum is $3.97 \times 10^{-2}$. Regions with low values of the $LS$-factor are located in plains, for example, the North China Plain, Northeast China Plain, Tarim Basin, and Inner Mongolian Plateau (Fig. 4h). The high values of the $LS$-factor occur on steep hills, such as the south-eastern Himalayas, Kunlun Mountains, Tianshan Mountains, and Qinling Mountains. The values of the $K$-factor in China range from $1.54 \times 10^{-2}$ to $6.36 \times 10^{-2}$, and the mean value is $3.60 \times 10^{-2}$ t h MJ$^{-1}$ mm$^{-1}$. High values of the $K$-factor occur in regions with loose soil texture, such as the Loess Plateau and desert regions in northwest China (Fig. 4i).

Variations of the $R$-factor and $C$-factor are drivers of the change in soil conservation capacity (Fig. 4b, 4c, 4e, and 4f). The change trend of the $R$-factor is classified into significant decreasing, decreasing, stable, increasing, and significant increasing (Fig. 4b), accounting for 1.2%, 22.0%, 17.4%, 35.1%, and 24.2% of the total study area, respectively. The average increasing rate of the $R$-factor is 7.16 MJ mm ha h$^{-1}$ a$^{-1}$ per year ($0.05 < p < 0.1$). However, we found a decreasing trend of the $R$-factor before 2005, then the $R$-factor increase fast after 2005 (Fig. 4c). The change trend of the $C$-factor is divided into seven classes (Fig. 4e), including extremely significant decreasing (accounting for 5.6% of the total study area), significant decreasing (14.4%), decreasing (3.6%), stable (45.1%), increasing (3.7%), significant increasing (8.8%), and extremely significant increasing (18.8%). The $C$-factor also shows a significant increasing trend generally from 1992 to 2019, with an average annual change rate of $1.27 \times 10^{-4}$ ($p < 0.05$). Nevertheless, after around 2007 or 2008, the $C$-factor shows a decreasing trend (Fig. 4f), which means that vegetation management has improved in China in recent decade.

https://doi.org/10.5194/essd-2022-222
Preprint. Discussion started: 19 July 2022
© Author(s) 2022. CC BY 4.0 License.
Figure 4: The input factors of the RUSLE model to estimate SEp, SEa, and SC (a, c, g, h, and i) and the change trend of the rainfall erosivity factor (R-factor) (b, c) and vegetation management factor (C-factor) (e, f) from 1992 to 2019. The change trend of the R-factor was classified into five degrees (b): significant decreasing (slope $\leq -1$, $p < 0.1$), decreasing (slope $\leq -1$, $p \geq 0.1$), stable (-1 < slope $\leq 1$), increasing (slope $> 1$, $p \geq 0.1$), significant increasing (slope $> 1$, $p < 0.1$). The change trend of the C-factor was classified into seven degrees (e): extremely significant decreasing (slope $\leq -0.0005$, $p < 0.05$), significant decreasing (-0.0005 < slope $\leq -0.0001$, $p < 0.05$), decreasing (slope $\leq -0.0001$, $p \geq 0.05$), stable (-0.0001 < slope $\leq 0.0001$), increasing (slope $> 0.0001$, $p \geq 0.05$), significant increasing (0.0005 < slope $\leq 0.0001$, $p < 0.05$), extremely significant increasing (slope $> 0.0005$, $p < 0.05$).

3.2 Spatial patterns of soil conservation capacity in China

The average annual soil conservation capacity in China was 542.09 t·ha$^{-1}$·a$^{-1}$ from 1992 to 2019, ranging from 0 to 55 316.03 t ha$^{-1}$ a$^{-1}$. Most of the regions in China (over 87%) were characterized by SC $> 1$ t ha$^{-1}$ a$^{-1}$ (Fig. 5). Regions with a great soil conservation capacity are mostly in southern and eastern China, especially in the Yangtze River Basin, Pearl River Basin, and Southeast Basin. Regions with SC $> 1$ 000 t ha$^{-1}$ a$^{-1}$ counted for 41.4% in the Southeast Basin, which was the
highest among the basins. The average SC was the highest in the Pearl River Basin (2182.64 t ha\(^{-1}\) a\(^{-1}\)). Correspondingly, SC capacity was relatively low in northern and western China. Regions with SC < 1 t ha\(^{-1}\) a\(^{-1}\) were mostly located in the Continental Basin, especially in the Tarim Basin, accounting for half of all regions with SC < 1 t ha\(^{-1}\) a\(^{-1}\). The average SC was 9.48 t ha\(^{-1}\) a\(^{-1}\) in the Continental Basin, which was also the lowest among the basins.

Figure 5: (a) Spatial patterns of soil conservation capacity during 1992-2019. (b) Soil conservation capacity among nine basins. The bar graphs depict the proportion of different classes of soil conservation capacity among different basins. The rhombus in the center of bar graphs depict the mean values of soil conservation capacity in each basin.

3.3 Temporal changes of soil conservation capacity

Soil conservation capacity shows an increasing trend in China from 1992 to 2019, with an increasing rate of 0.40 t ha\(^{-1}\) a\(^{-1}\) per year. 17.8% of the regions in China show significant changes in soil conservation capacity (\(p < 0.05\)), where 17.2% of the regions significantly increase (slope > 0, \(p < 0.05\)), while 0.6% of the regions significantly decrease (slope < 0, \(p < 0.05\)). The variation trend of soil conservation capacity is divided into seven types (Fig.6a), of which increasing trend classes account for 31.6%, and decreasing trend classes account for 9.5%. Although soil conservation capacity generally increases from 1992 to 2019, it shows a slightly decreasing trend before 2003 (slope=-0.86) (Fig. 6b). After 2003, soil conservation capacity increases rapidly in China (slope=10.05, \(p < 0.1\)). Most basins show an increasing trend, while decreasing trends are showed in the Pearl River Basin and Haihe River Basin (Fig. 6c). Regions with a significant decreasing trend of SC are also mostly distributed in the Pear River Basin. The significant decreasing trend of SC also occurs in the southern part of the Yangtze River Basin, mainly in the Yunnan-Guizhou Plateau. The significant increasing trend of SC mostly occurs in the Southwest Basin and west of the Yangtze River Basin, accounting for 23.5% and 11.2% of the basin, respectively.
The variation trend of soil conservation capacity is classified into seven types, which are significant decreasing (slope < -1, p < 0.05), decreasing (slope < -10, p ≥ 0.05), slight decreasing (-10 ≤ slope < -1, p ≥ 0.05), stable (-1 ≤ slope ≤ 1), slight increasing (1 < slope ≤ 10, p ≥ 0.05), increasing (slope > 10, p ≥ 0.05), significant increasing (slope > 1, p < 0.05).

The variation trend of soil conservation capacity among nine basins. The bar graphs depict the proportion of different types of the variation trend of soil conservation capacity among different basins. The rhombus in the center of bar graphs depict the average change rate of soil conservation capacity in each basin.

We found annual soil conservation capacity in some basins undergo divergent change trends in China from 1992 to 2019 (Fig. 7). The Pearl River Basin has the largest soil conservation capacity among all basins in China (Fig. 6b); however, its soil conservation capacity has a decreasing trend (Fig. 7a). Although the inter-annual trend of SC in the Haihe River Basin can be concluded as monotonically decreasing (Fig. 6c), it can also be divided into two phases: a non-significant decreasing phase between 1992 and 2001 and a slight but highly significant increasing phase between 2002 and 2019 (Fig. 7b). The inter-annual trend of SC in Songhua and Liaohe River Basin shows two different stages: a non-significant decrease from 1992 to 2004 and a significant increase from 2005 to 2019 (Fig. 7c). The inter-annual trend of SC for the Continental Basin (Fig. 4d) is stable before 2011, while after 2012, SC has a slightly increasing trend (Fig. 7d). Annual SC in the Yellow River Basin shows a significant decrease before 2001; however, it becomes an increasing trend after 2002 (Fig. 7e). In the Yangtze River Basin, SC increases significantly (Fig. 7h). Annual SC of other basins shows a non-significant increasing trend (Fig. 7f, 7h, and 7i). In general, except for the Pearl River Basin, soil conservation capacity of most basins has shown a rising trend in recent years in China.
Figure 7: Changes of annual soil conservation capacity among nine basins in China from 1992 to 2019. \( p < 0.05 \) indicates a highly significant trend, and \( p < 0.1 \) indicates a significant trend.

3.4 Data accuracy of the dataset

The comparison between the plot data and the modeled data (Fig. 8a) shows that the simulation of soil erosion can be accepted (all \( R^2 \geq 0.5 \)). The simulation in the Pearl Basin has the highest correlation with the sediment runoff data (\( R^2 = 0.91 \)), while the correlation is lowest in the Haihe River Basin (\( R^2 = 0.50 \)). In addition, modeled soil erosion results in this study were compared with those in other studies representing seven basins in China (Fig. 8b and Table S1), which indicates that our simulations of soil erosion are within a reliable range. However, our modeled soil erosion rates are overestimated compared with those from collected literature in the Pearl River Basin. We also collected some SC data from other studies on a regional scale (Table S2). Results of comparison with other regional SC (Fig. 8c) indicates that our SC data in these regions are consistent with these results. The model uncertainty is mainly derived from different algorithms of the parameters and data sources (Saltelli et al., 2019), which is also the reason for the differences between our results and other results.
Figure 8: (a) Correlation ($R^2$) between measurements of sediment runoff and simulations of soil erosion. (b) Comparison of simulated annual water erosion rates among different studies using USLE-based models. PLR: Pearl River Basin, SEB: Southeast Basin, YZR: Yangtze River Basin, SLR: Songhua and Liaohe River Basin, CNB: Continental Basin, YLR: Yellow River Basin, HAR: Haihe River Basin, HUR: Huaihe River Basin, SWB: Southwest Basin. (c) Comparison of soil conservation capacity from this study with those from other studies. TGR: Three Gorges Reservoir region, APE: Agro-pastoral ecotone of northern China.

4 Data availability

The dataset of soil conservation capacity in China (1992–2019) is available at: https://doi.org/10.11888/Terre.tpdc.272668 (Li et al., 2022). The data include average annual soil conservation capacity in China from 1992 to 2019, the annual change rate of soil conservation capacity from 1992 to 2019, and key input factors that lead to the variation of SC ($C$-factor and $R$-factor). Soil conservation capacity for a certain year or a certain region can be obtained by contacting the author.

5 Discussion

This study presented soil conservation capacity dataset in China, which were calculated from potential and actual soil erosion in China based on the RUSLE model. This dataset has several advantages. First, it is a long-term dataset covering
1992–2019, filled a gap of the lack of long-time sequenced field measurements of soil conservation capacity. Compared with other regional simulations, our dataset also has the most comprehensive time coverage (Table S1 and S2). Second, it is a large-scale (national) dataset, which contains more regional information. Although it is less detailed than some regional scale studies (Jin et al., 2021), our dataset has wider applicability for more regions. Third, these data have a relatively high precision (300 m) compared with other national-scale studies (≥ 1 km) (Teng et al., 2019; Zhuang et al., 2021). Although there is a large-scale study (Rao et al., 2013) of which simulations has higher precision than ours (90 m vs 300 m), our dataset has longer time series (one year vs 28 years), updated data (2009 vs 2019), and more detailed input data (for example, 90 m DEM vs 30 m DEM). Last, the data accuracy was improved by modifying algorithms of the input factors in a distributed way. Specifically, the $C$-factor values were designated according to different natural vegetation and crop types. The $R$-factor was calibrated according to different regions in China based on daily rainfall data. The $P$-factor was improved according to management measures in crop types and terrain characteristics.

The potential applications of this dataset are as follows. First, this dataset can be used to further clarify the mechanisms of the pattern and the variation of soil conservation in China, including the effects of climate change and human activities (Li and Fang, 2016; Eekhout and de Vente, 2022). Second, this dataset could provide the data basis for comprehensive assessments and trade-offs of other ecosystem service functions (Schroter et al., 2005; Ouyang et al., 2016). Third, it can help evaluate the ecological benefits of soil conservation measures and policies, which could provide scientific basis for policy making and land management on regional and national scales (Kong et al., 2018). Last, this dataset can also serve as the data references and comparisons for other regional soil conservation capacity, global soil conservation capacity, and even future soil conservation assessment.

Although the simulation results have been improved by using high-resolution input data and calibrating model factors, there are still some limitations in this study. First, there are certain difficulties in the trade-off between the feasibility of calculation and the accuracy of the simulations in large-scale modelling (Yang et al., 2021). The use of the RUSLE model has a spatial scale effect in the precision of the input data (Alewell et al., 2019; Wang et al., 2020), which is one of the reasons that lead to the differences between large-scale model estimations (our study) and small-scale estimations (other studies). The finer input data can produce more reliable results. We can employ finer input data in further study. Second, more different algorithms of input factors can be applied according to their performance in different regions. We can also call this a “distributed” way to improve the algorithm of the model. For example, estimations in the Pearl River Basin are higher than other simulations (Fig.9b). The reason may be that the methods we used in this region are different from other studies. Further exploration of the methods that are applicable to this region are still needed. Last, this dataset still needs further validations. However, the lack of sufficient measurement data for comparison disturbs the subsequent verification of the dataset.
Author contribution

JL designed the algorithm and wrote the article. HH and JL developed the model code and performed the simulations. LC offered critical review. RS formulated overarching research goals and aims.

Competing interests

The authors declare that they have no conflict of interest.

References


FAO, and ITPS. Status of the World’s Soil Resources (SWSR) – Main Report, Rome, Italy, Food and Agriculture Organization of the United Nations and Intergovernmental Technical Panel on Soils. 2015


SCIO. Food Security in China, State Council of the People's Republic of China r. 2019


