Statistical evaluation of proxies for estimating the rainfall erosivity factor

Xiaoqing Ma2,3 & Mingguo Zheng1,4∗

Considering the high-temporal-resolution rainfall data requirements for calculating the Rainfall Erosivity factor (that is, the R-factor), studies have developed a large number of proxies for the R-factor (PR). This study aims to evaluate 15 widely used proxies, which were developed in various countries using daily, monthly, or yearly rainfall data, in terms of correlation and statistical equality with the R-factor by using the 6-min pluviographic data from 28 stations in Australia. Meng’s test was applied to rank the correlations. Although the Meng’s test indicated that the correlation between Rainfall Erosivity (R) and Rainfall Erosivity calculated by the proxy model (PR) generally increased with a finer time resolution of the rainfall data (in the order of year, month, and day), the 15 PRs under examination were all highly correlated with R (r > 0.62, p < 0.004), implying that all of them can be reasonably used as an R predictor. A direct estimation of the R-factor using PRs produced a mean relative error (MRE), root mean square error (RMSE), and Nash–Sutcliffe efficiency coefficient (NSE) with a mean of 50.0%, 1392 MJ mm ha⁻¹ h⁻¹ a⁻¹, and 0.17, respectively. The linear calibrations improved the accuracy of the estimation and produced an MRE, RMSE, and NSE with a mean of 36.0%, 887 MJ mm ha⁻¹ h⁻¹ a⁻¹, and 0.70, respectively. Finally, suitable proxies for instances where only daily, monthly, or yearly rainfall data are available were recommended.

Abbreviations

R  Rainfall erosivity calculated by RUSLE2, MJ mm ha⁻¹ h⁻¹ a⁻¹
PR  Rainfall erosivity calculated by proxy model, proxy for R, MJ mm ha⁻¹ h⁻¹ a⁻¹
MAP  Mean annual precipitation, mm
MRE  Mean relative error, %
RMSE  Root mean square error, MJ mm ha⁻¹ h⁻¹ a⁻¹
NSE  Nash–Sutcliffe efficiency coefficient
MFI  Modified Fournier index, mm
F  Fournier index, mm

Soil erosion can lead to land degradation and environmental deterioration1–3. Eroded sediments also have negative off-site effects, such as impaired water quality, altered aquatic habitats, limited channel navigability, reduced hydroelectric equipment lifespan, and increased flood risk4–6. Rainfall is a driver of water erosion. Unlike other natural factors affecting erosion, such as topography and vegetation, rainfall is less subject to human alteration or control. Rainfall thus represents an environmental contributor to soil loss and is of great importance in assessing or predicting soil erosion intensity2.

Rainfall Erosivity refers to the potential of a storm to erode the soil. The universal soil loss equation (USLE)8,9 and the revised universal soil loss equation (RUSLE, RUSLE2)10,11, the most widely used models for predicting soil erosion, applied equations (Eqs. 1 and 2) to quantify the average annual Rainfall Erosivity, R (MJ mm ha⁻¹ h⁻¹ a⁻¹). The discrepancy is reflected in the algorithm of the estimated unit rainfall kinetic energy (e, MJ mm⁻¹ ha⁻¹), and Eqs. (3–5) are utilized by USLE, RUSLE, and RUSLE2, respectively.

1National-Regional Joint Engineering Research Center for Soil Pollution Control and Remediation in South China, Guangdong Key Laboratory of Integrated Agro-environmental Pollution Control and Management, Guangdong Engineering Research Center for Non-point Source Pollution Control, Institute of Eco-environmental and Soil Sciences, Guangdong Academy of Sciences, Guangzhou 510650, People’s Republic of China. 2Key Laboratory of Water Cycle and Related Land Surface Processes, Institute of Geographic Sciences & Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, People’s Republic of China. 3College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100101, People’s Republic of China. 4International Institute of Soil and Water Conservation, Meizhou 514000, People’s Republic of China. ∗email: mgzheng@soil.gd.cn
\[ R = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m} (EI_{30})_j \]  

\[ E = \left( \frac{1}{\sum_{i=1}^{n} (e_r \cdot P_r)} \right) \]  

\[ e_r = \begin{cases} 0.119 + 0.0873 \log i_r, & i_r \leq 76.2 \text{ mm/h} \\ 0.283 i_r, & i_r > 76.2 \text{ mm/h} \end{cases} \]  

\[ e_r = 0.29 \left[ 1 - 0.72 \exp(-0.05i_r) \right] \]  

\[ e_r = 0.29 \left[ 1 - 0.72 \exp(-0.082i_r) \right] \]  

where \( n \) is the number of years of the record, \( m \) is the number of erosive events in each year, \( EI_{30} \) is the Rainfall Erosivity index of a single event, \( E \) is the total rainfall energy (MJ ha\(^{-1}\)), \( I_{30} \) is the maximum 30 min rainfall intensity (mm h\(^{-1}\)), and \( P_r \) and \( i_r \) are the rainfall amount (mm) and rainfall intensity of the \( r \)th time interval (mm h\(^{-1}\)), respectively.

Comparing the three equations for the \( R \)-factor, the results indicated that the RUSLE significantly underestimates the \( R \) value\(^1\)\(^2\)\(^\text{13-15. Nearing et al.}^14\) stated that the \( R \) value calculated by the RUSLE was 14% lower than that of the USLE and RUSLE2; however, the USLE and RUSLE2 results showed resemblances based on data from 56 stations from three countries: China, Italy, and the US. Studies show that the RUSLE used to estimate the \( R \)-factor is based on a conceptual mistake that results in underestimation of the \( R \) value; hence, Eq. (4) (RUSLE) should not be used for \( R \) calculation and Eq. (5) (RUSLE2) is recommended to be used instead\(^14\). The RUSLE2 is commonly utilized by the US government for conservation planning\(^16\) and has been used in recent studies on Rainfall Erosivity\(^17\)\(^18\).

The \( EI_{30} \) index is a standard algorithm for calculating Rainfall Erosivity. However, the algorithm requires pluviographic records at <15 min intervals\(^4\); moreover, the record period should be longer than 20 years for the purpose of containing dry and wet climatic regimes\(^15\). Such rainfall data are often unavailable, particularly in developing countries\(^24\). In addition, the algorithm is complicated and often laborious and time-consuming\(^1\).

Considering the limited data availability and tedious process, many proxies for the \( R \)-factor (PR) have been developed, using daily, monthly, and annual data instead of pluviographic data. Table 1 lists 15 proxies that have been extensively applied in the literature, three of which have been developed in the US, two in Spain, two in Australia. The \( R \)-factor was calculated using \( EI_{30} \) and the \( R \)-factor and the proxies\(^7\)\(^23-27\). However, the most suitable proxy was determined simply through a direct comparison of the obtained Nash–Sutcliffe coefficient and the relative error, which vary with the collected sample data and are random variables. Because the comparison among random variables should be done in the framework of statistics, a direct comparison of the obtained Nash–Sutcliffe coefficient or relative error is questionable.

To assess candidate proxies, studies have often used the Nash–Sutcliffe efficiency coefficient\(^22\) and the relative error to evaluate the discrepancies between the \( R \)-factor and the proxies\(^7\)\(^23-27\). However, the most suitable proxy was determined simply through a direct comparison of the obtained Nash–Sutcliffe coefficient and the relative error, which vary with the collected sample data and are random variables. Because the comparison among random variables should be done in the framework of statistics, a direct comparison of the obtained Nash–Sutcliffe coefficient or relative error is questionable.

To the best of our knowledge, no studies have conducted statistical evaluations of the established PR. The objective of this study was to statistically evaluate the 15 R-proxies listed in Table 1 using rainfall data from Australia. We ranked the correlations between the proxies and \( R \)-factor using Meng’s test. Based on the statistical evaluations, we recommended proxies for instances where only daily, monthly, or yearly rainfall data are available.

**Methodology**

**Data source.** We used 6-min pluviographic rainfall data from 28 stations in Australia (Table 2), which were provided by the Australian Government Bureau of Meteorology (http://www.bom.gov.au). Australia has a diverse climate, including tropical monsoon climate in the northern part, Mediterranean climate in the southwestern corner of the country, oceanic and humid subtropical climate in the southeastern part, and arid to semi-arid climate in the interior mainland. Rainfall has high spatial variation, ranging from <50 mm to >3000 mm, with a mean of 465 mm (Table 2). Detailed information on the selected stations is listed in Table 2, including location information, length of record, mean annual precipitation (MAP), and \( R \) value. The selected stations spanned from tropical to subtropical and temperate, and from humid to semi-arid and arid regions, with a MAP ranging from 195 to 1236.2 mm. Most of the selected stations (22 out of the 28 stations) had a MAP ranging between 400 and 800 mm because soil erosion is rather limited in very humid or very dry areas. The missing rainfall data for each station varied, and for the purpose of the quality of the rainfall data, the maximum missing rate of each year should not be greater than 90% for each station. Hence, the data for the selected period were unequal. Rainfall data from the 28 stations were for the period 1960–2011, and the data periods ranged from 20 to 31 years, with a mean of 23 years for the 28 selected stations.

**Calculation of the R-factor using the RUSLE2 and its proxies.** The \( R \) value is generally underestimated using the RUSLE2, as discussed in the Introduction section. Hence, the \( R \)-factor was calculated using Eqs. (1), (2), and (5) (RUSLE2), and the 15 PRs (the proxies for the \( R \)-factor; Table 1) were calculated using the models listed in Table 1. In the calculation, we defined two consecutive rainfall events as having a time interval longer
| Rainfall data | Equations | Countries where PRI was developed | References |
|--------------|-----------|----------------------------------|------------|
| Annual       | $PR_a = 0.04830 MAP^{0.454} P < 850$ mm  
$PR_a = 587.5 - 1.219 MAP + 0.004105 MAP^2 P > 850$ mm  
$PR_a = 699.3 + 7.0001 MAP - 2.7190H$ | US  
Honduras | 28  
29 |
| Monthly      | $PR_m = 9.8 \times \sum_{i=1}^{n} AR = \sum_{i=1}^{n} \left(1.735 \times 10^{-8} \left(\frac{L}{1000}\right)^{-0.8188}\right)$ | US | 40 |
|              | $PR_m = 0.07397 MFI^{0.474} MFI < 55$ mm  
$PR_m = 95.77 - 6.081 MFI + 0.4770 MFI^2 MFI > 55$ mm  
$MFI = \sum_{i=1}^{n} \left(\frac{P}{MAP}\right)$ | US | 28 |
|              | $PR_m = 21.56 MFI^{0.474}$ | Spain | 7 |
|              | $PR_m = 25.1F$ | Spain | 28 |
|              | $PR_m = 0.3589 MFI^{0.502}$ | China | 42 |
|              | $PR_a = 0.10 \times \sum_{j=1}^{12} \frac{0.0125}{2}d^{2.95}$ | China | 42 |
| Daily        | $PR_d = \sum_{i=1}^{n} \sum_{j=1}^{12} MR_j = \sum_{i=1}^{n} \sum_{j=1}^{12} 6.97 (rain_{ij}) - 11.23 \{day_{ij}\}$ | Malaysia | 32 |
|              | $PR_d = \sum_{i=1}^{n} \sum_{j=1}^{12} MR_j = \sum_{i=1}^{n} \sum_{j=1}^{12} 0.266 (rain_{ij})^{1.107} \{day_{ij}\}^{1.367}$ | Malaysia | 32 |
|              | $PR_{11} = \sum_{i=1}^{24} \left(\sum_{j=1}^{n} (HMR)_{ij} = \sum_{i=1}^{24} \sum_{j=1}^{n} \alpha \sum_{m=1}^{n} (P_{D12})^\beta_m \right.$  
$\alpha = 21.586\beta^{-0.2083}$  
$\beta = 0.8363 + \frac{18.144}{P_{D12}} + 24.450 \frac{P_{D12}}{P_{D12}}$ | China | 43 |
|              | $PR_{12} = \sum_{i=1}^{24} \left(\sum_{j=1}^{n} (HMR)_{ij} = \sum_{i=1}^{24} \sum_{j=1}^{n} \alpha \sum_{m=1}^{n} (P_{D12})^\beta_m \right.$  
$\alpha = 10^{0.124 - 0.495j + 0.00214D_{season}}$  
$\beta = 0.8363 + \frac{18.144}{P_{D12}} + 24.450 \frac{P_{D12}}{P_{D12}}$ | China | 44 |
|              | $PR_{13} = \sum_{i=1}^{24} \left(\sum_{j=1}^{n} (HMR)_{ij} = \sum_{i=1}^{24} \sum_{j=1}^{n} \alpha \sum_{m=1}^{n} (P_{D12})^{1.7285} \right.$  
$\alpha = \left\{0.3937 \text{ warm season (10-12months, 1-4months)}\right.$  
$0.3101 \text{ cold season (5-9months)}\right\}$ | China | 45 |
|              | $PR_{14} = \sum_{i=1}^{12} \sum_{j=1}^{n} MR_j = \sum_{i=1}^{12} \sum_{j=1}^{n} \left[1 + \eta \cos(2\pi f - \omega) \right] \sum_{m=1}^{n} (P_{D12})^\beta_m \right.$  
$f = \frac{1}{12}; \omega = \frac{\beta}{2}; \beta = 1.49; \eta = 0.29$  
$\alpha = 0.3591 + 0.0989\exp(3.265/MAP)$ | Australia | 46 |
|              | $PR_{15} = \sum_{i=1}^{12} \sum_{j=1}^{n} MR_j = \sum_{i=1}^{12} \sum_{j=1}^{n} \left[1 + \eta \cos(2\pi f - \omega) \right] \sum_{m=1}^{n} (P_{D12})^\beta_m \right.$  
$f = \frac{1}{12}; \omega = \frac{\beta}{2}; \beta = 0.3$  
$\alpha = 1.02 \times 10^{-2.081} - 1.508$  
$\alpha = 2.349 + 0.0404\alpha - 0.0002684H$ | Australia | 47 |

Table 1. Established proxies for the R-factor. $PR$ is the ith proxy for the R-factor, $AR_i$, $MR_i$, and $HMR_i$ (M) mm ha$^{-1}$ h$^{-1}$ a$^{-1}$ are the Rainfall Erosivity in the ith year, jth month, and kth half-month, respectively. $P$, $P_a$, and $R$ are the rainfall in month j, annual rainfall, average rainfall in month i, and highest monthly rainfall, respectively. $rain_{ij}$ (mm) represents monthly rainfall exceeding 10 mm for days, and $day_{ij}$ represents the monthly rainfall exceeding 10 mm for days. $(P_{D12})_{12,0}$ (mm) on the mth day. $P_{dmax}$ (mm) represents the maximum daily rainfall in an average year. $H$ (mm) and $L$ are the altitude and latitude in decimal degrees of the stations, respectively. $S$ (mm) is the average rainfall for the summer half of the year (November–April, Bureau of Meteorology, 1989). $MAP$ (mm) is the mean annual precipitation. $\alpha$, $\beta$, and $\eta$ are regional constants. $F$ and $MFI$ represent the Fournier index and the modified Fournier index, respectively.

than 6 h, as suggested by Wischmeier and Smith$^4$. The $< 12.7$ mm rainfall events, which have limited erosivity, were excluded by the USLE; in contrast, the RUSLE used all rainfall events while calculating Rainfall Erosivity. In Australia, exclusion of the $< 12.7$ mm rainfall events had negligible consequences. However, in the areas with $a < 400$ mm MAP, where all rainfall events should be included in calculating Rainfall Erosivity$^{28}$. Therefore, we used all rainfall events to calculate the R values.
Using Meng’s test to rank correlations. The correlation coefficient indicates agreement between two variables, and we applied the Pearson’s correlation coefficient in this study. The strength of the correlation between $R$ and $PR_i$ (the $i$th proxy) is very useful for evaluating the applicability of a $PR_i$. Methods to compare correlations differ depending on whether the correlations are dependent or independent. The correlations between $R$ and $PR_i$ are dependent because they have a common dependent variable, that is, the $R$-factor. This study used Meng’s test to compare correlations between $R$ and $PR_i$. Meng’s test was designed to compare the correlations and has been extensively applied in psychological research. Zheng and Chen used this test to determine the most accurate Rainfall Erosivity indices for single storms on the Chinese Loess Plateau.

Meng et al. designed a $Z$-test to compare two correlations ($r_1$ and $r_2$) that have a common dependent variable. The equations are as follows:

$$Z = (z_{r_1} - z_{r_2}) \sqrt{\frac{N - 3}{2(1 - r_1)r_2h}}$$

$$h = \frac{1-f r_i^2}{1-r_i^2}$$

$$f = \frac{1-r_x}{2(1-r_i^2)}$$

where $z_{r_i}$ is the Fisher $z$-transformed value for $r_i$ ($i = 1$ or $2$), $z_{r_1} = \frac{1}{2} \ln \frac{1+r_1}{1-r_1}$, $z_{r_2} = \frac{1}{2} \ln \frac{1+r_2}{1-r_2}$, $N$ is the sample size, $r_x$ is the correlation between the two independent variables, $r_i^2 = (r_1^2 + r_2^2) / 2$, and $f$ should be set to 1 if it is greater than 1.

To compare $k$ correlations with a common dependent variable ($r_1$, $r_2$, ..., $r_k$, $k > 2$), Meng’s test becomes...
\[ Z = r_{xz} \sqrt{\chi^2(k-1)} \]  \hspace{1cm} (9)

where \( r_{xz} \) represents the correlation coefficient between \( z_i \) and \( x_i \), \( \lambda_i \) is the contrast weight assigned to each \( x_i \) and \( z_i \), and \( k \) is the number of variables. \( \chi^2 \) is distributed with \( k-1 \) degrees of freedom.

\[ \chi^2(k-1) = \left( \frac{N - 3}{(1 - r_z h)^2} \right) \sum_{i=1}^{k} (z_{ri} - z_i)^2 \]  \hspace{1cm} (10)

where \( N \) is the sample size, \( z_i \) is the mean of \( z \), \( r_z \) is the median intercorrelation among the \( k \) independent variables, and \( r_z \) is the median intercorrelation among the \( k \) independent variables.

By comparing \( r \) and the average of the other \( k \) correlation coefficients, Meng's test determines whether \( r \) is a contrast in the two-tailed test. In the case of the one-tailed test, Meng's test determines whether \( r \) is a less- or better-correlated contrast. This study used a stepwise Meng's test procedure to rank the correlations between \( R \) and \( PR \). We first labeled each of the correlations between \( R \) and \( PR \) as better- or less-correlated contrast, less-correlated contrast, or non-contrast by means of the one-tailed Meng's Z-test, whereby the correlations under examination were grouped into two classes: the less-correlated class and the other class. Then, we repeated the procedure for each of the two classes, and the procedure did not terminate until no contrast was detected among any subclass. The significance level was set at 0.05. Bonferroni correction applies when several bivariate tests are performed simultaneously. We do not use Eq. (9) multiple times simultaneously, so that the Bonferroni correction is not necessary. In this way, we ranked the correlation between \( R \) and \( PR \).

In addition to the correlation coefficient, we also used the MRE (\%), the RMSE (MJ mm ha\(^{-1}\) h\(^{-1}\) a\(^{-1}\)), and NSE to assess the performance of the regression equation. The MRE, RMSE, and NSE are defined as follows:

\[ MRE = \frac{\sum_{i=1}^{n} 100 \times |R_i - PR_i|/R_i}{n} \]  \hspace{1cm} (11)

\[ RMSE = \sqrt{\frac{\sum_{i=1}^{n} (PR_i - R_i)^2}{n}} \]  \hspace{1cm} (12)

\[ NSE = 1 - \frac{\sum_{i=1}^{n} (R_i - PR_i)^2}{\sum_{i=1}^{n} (R_i - \bar{R})} \]  \hspace{1cm} (13)

where \( R_i \) and \( PR_i \) are the \( R \) and \( PR \) values of the \( i \)-th station, respectively, \( n \) is the number of stations, and \( \bar{R} \) is the average value of \( R \).

Results and discussion
Calculation results and \( R \) and \( PR \) deviations. The calculated \( R \)-factor of the 28 stations varied between 290.8 MJ mm ha\(^{-1}\) h\(^{-1}\) a\(^{-1}\) and 7153.7 MJ mm ha\(^{-1}\) h\(^{-1}\) a\(^{-1}\) with a mean of 2424.1 MJ mm ha\(^{-1}\) h\(^{-1}\) a\(^{-1}\). As shown in Fig. 1, all examined proxies were surprisingly found to be highly correlated with the \( R \)-factor (\( r > 0.62, p < 0.004 \)), implying that all of them can be reasonably used as a predictor of the \( R \)-factor in Australia. This finding is at odds with the common notion that regional differences in physiographic settings prevent a model of the \( R \)-factor from application to other areas. The reason might be that the \( R \) models selected were established in different regions and developed by the \( R \) value calculated by the USLE or RUSLE method. The \( R \) value is in fact an approximation of the \( R \)-factor calculation of the USLE or RUSLE. Therefore, the \( R \) value is significantly related to the \( R \) value, and the difference between \( PR \) and \( R \) reflects the correction coefficient, which is the regional difference, and therefore all \( PR \) values were highly correlated with the \( R \) values.

The linear regression equation in Fig. 1 can be applied to modify the deviation between the \( PR \) and \( R \). Table 3 lists the RMSE, MRE, and NSE for the estimation of \( R \) values that directly used \( PR \) and modified \( PR \). The direct estimation showed high deviation, the MRE ranged from 18.3 to 153.0%, the RMSE ranged from 561 to 2279 MJ mm ha\(^{-1}\) h\(^{-1}\) a\(^{-1}\), and the NSE ranged from –2.1 to 0.89. The revised calculations, as shown in Table 3, indicated the MRE ranged from 13.1 to 77.4%, the RMSE ranged from 302 to 1482 MJ mm ha\(^{-1}\) h\(^{-1}\) a\(^{-1}\), and the NSE ranged from 0.23 to 0.97. The RMSE (\( p = 0.002 \)) and the MRE (\( p = 0.01 \)) were notably reduced, and the NSE (\( p = 0.009 \)) significantly increased under the correction, which further proved that the error was string-reduced when the modified \( PR \) was used. However, not all \( PR \) showed similar expression, such as the RMSE, MRE, and NSE of \( PR_{11} \), \( PR_{12} \), \( PR_{13} \), and \( PR_{14} \), which changed slightly after calibration.

Correlation comparison between \( R \) and \( PR \) using Meng’s test. According to the step-by-step test method designed in this study, the flow chart is shown in Fig. 2, Eq. (9) divided the correlation coefficients of 15 \( PR \)-s and \( R \) into two classes: \( L_1 \) (\( r_{11}, r_{12}, r_{22}, r_{23} \)), and \( O_1 \) (\( r_{32}, r_{42}, r_{52}, r_{62}, r_{72}, r_{82}, r_{92}, r_{112}, r_{122}, r_{132}, r_{142}, r_{152}, r_{162}, r_{172}, r_{182}, r_{192} \)), and no significant difference was observed in the \( L_1 \) class (\( p > 0.05 \)). The correlation coefficients in the \( O_1 \) class were separated into \( L_2 \) (\( r_{32}, r_{42}, r_{52}, r_{62} \)) and \( O_3 \) (\( r_{72}, r_{82}, r_{92}, r_{112}, r_{122}, r_{132}, r_{142}, r_{152}, r_{162}, r_{172}, r_{182}, r_{192} \)) classes, and no contrast was detected in the \( L_2 \) class (\( p > 0.38 \)). The correlation coefficients in the \( O_2 \) class were divided into \( L_3 \) (\( r_{32}, r_{42}, r_{52} \)) and \( O_4 \) (\( r_{62}, r_{72}, r_{82}, r_{92} \)) classes, and the coefficients in the \( L_3 \) class showed statistical equality under Meng’s test. The correlation coefficients of \( PR_{14} \) and \( PR_{13} \) were notably greater than those of \( r_{11} \) and \( r_{13} \), and further statistical tests indicated no significant differences.
between $r_{11}$ and $r_{13}$, $r_{14}$, and $r_{15}$, respectively. After four rounds (Fig. 2), the stepwise Meng’s test procedure ranked the correlation between $PR_i$ and $R$, $r_i$, as follows:

$$r_1 = r_2 = r_6 = r_9 < r_3 = r_4 = r_8 = r_{10} < r_5 = r_{12} < r_{11} = r_{13} < r_{14} = r_{15}$$  \(14\)

Figure 1. Linear regression between $R$ and $PR$.

As expected, the correlation generally became stronger with finer time resolution of the rainfall data, and the daily proxies were stronger correlated with $R$ than the monthly ones, which in turn were stronger correlated
when estimating the RPR value by revised coefficients of the equations were significantly unequal to 1. Table 3 shows that directly using the monthly PR was both highly correlated with MAP and Hernando and Romana36 reported a stronger correlation between the MFI and Fournier index (F)

Table 3. MRE, RMSE and NSE of the selected PRs in estimating the R value.

| PR   | Direct estimation | Modified estimation |
|------|------------------|---------------------|
| MRE  | RMSE             | NSE                 | MRE  | RMSE             | NSE  |
| (MJ mm ha⁻¹ h⁻¹ a⁻¹) | (%) | (MJ mm ha⁻¹ h⁻¹ a⁻¹) | (%) | (MJ mm ha⁻¹ h⁻¹ a⁻¹) | (%) |
| PR₁  | 50.3             | 1599                | 0.10 | 56.5             | 1323 | 0.39 |
| PR₂  | 153.0            | 2279                | −0.82 | 77.4             | 1482 | 0.23 |
| PR₃  | 37.4             | 1214                | 0.48 | 37.6             | 955  | 0.68 |
| PR₄  | 62.4             | 2971                | −2.10 | 39.8             | 896  | 0.72 |
| PR₅  | 44.8             | 1732                | −0.05 | 33.6             | 862  | 0.74 |
| PR₆  | 77.1             | 2275                | −0.82 | 36.6             | 1027 | 0.63 |
| PR₇  | 50.7             | 1405                | 0.31 | 38.6             | 884  | 0.73 |
| PR₈  | 43.2             | 1433                | 0.28 | 40.8             | 927  | 0.70 |
| PR₉  | 54.4             | 1080                | 0.59 | 39.2             | 1008 | 0.64 |
| PR₁₀ | 59.6             | 1197                | 0.50 | 28.9             | 940  | 0.69 |
| PR₁₁ | 26.1             | 1732                | 0.78 | 25.0             | 730  | 0.81 |
| PR₁₂ | 29.6             | 926                 | 0.70 | 30.1             | 846  | 0.75 |
| PR₁₃ | 23.6             | 812                 | 0.77 | 24.8             | 708  | 0.82 |
| PR₁₄ | 19.6             | 561                 | 0.89 | 18.6             | 410  | 0.94 |
| PR₁₅ | 18.3             | 609                 | 0.87 | 13.1             | 302  | 0.97 |
| Average | 50.0             | 1392                | 0.17 | 36.0             | 887  | 0.70 |

Table 3. MRE, RMSE and NSE of the selected PRs in estimating the R value.

with R than the yearly ones. The MRE, RMSE, and NSE (Table 3) showed a very similar rank to the correlation in Eq. (14).

**Recommendation for estimating rainfall data with different resolutions.** Two yearly proxies, PR₁ and PR₂, showed strong correlation with R (Fig. 1a,b). The t-test indicated that the coefficient of PR₁ was equal to 1, and the constant term of PR₁ was, significantly, equal to 0. It showed that the R value in Australia can be directly estimated by PR₁, and Table 3 shows that no striking variation was observed among the MRE, RMSE and NSE when estimating the R value by revised PR₁. The coefficient and constant term of PR₁ were significantly unequal to 1 and 0, respectively. It showed that the estimation of the R value in Australia directly by PR₁ would result in a large discrepancy. Based on Eq. (14), no statistical difference was observed between r₁ and r₂. Hence, we suggest the use of PR₁ and revised PR₂ as predictors of the R-factor when only yearly rainfall data are available.

Among the monthly proxies, PR₆, PR₇, and PR₈ are based on the modified Fournier index (MFI), whereas PR₅ is based on the Fournier index (F) (Table 1). Both indices have been widely applied to estimate the R-factor31–34. Renard et al.30 and Yue et al. (2014) found that the MFI was stronger correlated with MAP than the F. Arnoldus35 and Hernando and Romana36 reported a stronger correlation between the MFI and R than between the F and R. The Australian data we used agree well with the observations above. Although the MFI and F were both highly correlated with MAP and R (Eq. (3)), Meng’s test indicated that the MFI was stronger correlated with both MAP and R (p < 0.048). The three MFI-based monthly proxies, PR₆, PR₇, and PR₈, even showed a stronger correlation with R than PR₁ (Eq. (14)), a daily proxy for the R-factor. Hence, we concluded that the MFI is superior to the F in developing a proxy for the R-factor. The t-test of the regression equation of PR₆–PR₈ indicated that the coefficients of the equations were significantly unequal to 1. Table 3 shows that directly using the monthly PRs to estimate the R-factor produced large errors. Therefore, when only monthly rainfall data are available, it is better to use the revised PR₂ and PR₃ to predict the R-factor.

Among the daily proxies, PR₁₃ and PR₁₄ were developed in Australia, and their correlation with R ranked first among the 15 proxies. PR₁₃ and PR₁₄, two Chinese proxies, ranked second in terms of their correlation with R. The reason for the superior performance of PR₁₃ and PR₁₄ may not simply be due to the fact that they were established using Australian data. Capturing the periodic variation within a year, the model developed by Yu and Rosewell37 (the original version of PR₁₄ and PR₁₃) performed best among the 8 R-proxies in northeastern Spain3.

The t-test of the regression equation of PR₁₀–PR₁₃ indicated that the coefficient and the constant were equal to 1 (p > 0.06) and 0 (p > 0.33), respectively. Table 3 illustrates that no remarkable improvement was observed in the deviation of the R value when estimated by the revised equations. Consequently, PR₁₀–PR₁₃ can be used to estimate the R value. The coefficient and constant term of PR₁₀ were significantly unequal to 1 and 0, respectively. The revised PR₁₀ was recommended when estimating the R-factor in Australia. The coefficient of PR₁₀ was significantly equal to 1 (p = 0.87) and the constant was notably greater than 0 (p = 0.02), which indicated that the equation for PR₁₀ in Fig. 1k probably leads to a large deviation in estimating R values for dry areas. The constant term of PR₁₅ was equal to 0 (p = 0.45) and the coefficient was significantly greater than 1 (p = 0.00), which indicated
that PR$_{15}$ underestimated the $R$ value. PR$_{15}$ was established using the RUSLE model, which may have led to this result. Nearing et al. (2017) concluded that the RUSLE underestimated 14%, compared with the RUSLE2 and the USLE, in estimating the $R$-factor. This result coincided with the coefficient of 0.19, which is shown in Fig. 1l.

We suggest the use of revised PR$_{14}$ and PR$_{15}$ as surrogates of the $R$-factor when daily rainfall data are available. However, with MREs of 26.1% and 23.6%, respectively (Fig. 1i,j), the accuracy remains satisfactory while using PR$_{11}$ and PR$_{13}$ as a surrogate of the $R$-factor.

**Conclusion**

This study statistically evaluated 15 developed PR using rainfall data from Australia. All proxies were found to be significantly correlated with $R$ ($p < 0.004$), and thus can be used to predict the $R$-factor by means of linear calibration models. Meng's test ranked the correlations between $R$ and PR (see Eq. (14)), indicating that the daily proxies were stronger correlated with $R$ than the monthly ones, which in turn were stronger correlated than the yearly ones. Furthermore, the t-test of the regression equations showed that five proxies ($PR_{1}$, $PR_{10}$, $PR_{11}$, $PR_{12}$, and $PR_{13}$) were statistically equal to the $R$-factor, implying that they can be directly used as an $R$ surrogate without calibration. A direct estimation of the $R$-factor using PRs produced an MRE between 18.3 and 153.0%, with a mean of 50.0%. The linear calibrations produced an MRE ranging from 13.1 to 77.4%, with a mean of 36%.

Based on the statistical evaluations above, we recommended proxies for instances where only daily, monthly, or yearly rainfall data are available. Our rainfall data were collected from varying climate zones (from tropical to temperate, and from humid to semi-arid and arid). Hence, although future work in other regions is needed, our results may be applicable to other regions around the world.
References

1. Pimentel, D., Harvey, C., Resosudarmo, P. & Sinclair, K. Environmental and economic costs of soil erosion and conservation benefits. Science 267, 1117–1123 (1995).
2. Jiang, C. et al. Impact of erosion-induced land degradation on rainfall infiltration in different types of soils under field simulation. J. Arid Environ. 30, 2 (2019).
3. Wei, Y. et al. Impact of erosion-induced land degradation on rainfall infiltration in different types of soils under field simulation. Land Degrad. Dev. 30, 2 (2019).
4. Kisi, O. River suspended sediment concentration modeling using a neural differential evolution approach. J. Hydrol. 389, 227–235 (2010).
5. García-Ruiz, J. M., Bequería, S., Lana-Renault, N., Nadal-Romero, E. & Cerdà, A. Ongoing and emerging questions in water erosion studies. Land Degrad. Dev. 28, 5–21 (2017).
6. Poersen, & Jean,. Soil erosion in the Anthropocene: Research needs. Earth Surf. Process. Landforms 43(1), 64–84 (2018).
7. Angulo-Martínez, M. & Bequería, S. Estimating rainfall erosivity from daily precipitation records: A comparison among methods using data from the Ebro Basin (NE Spain). J. Hydrol. 379(1), 111–121 (2009).
8. Wischmeier, W.H., & Smith, D.D. Predicting rainfall-erosion losses from cropland east of the Rocky Mountains: guide for selection of practices for soil and water conservation. Agriculture handbook/USDA, (1965).
9. Wischmeier, W.H., & Smith, D.D. Predicting rainfall erosion losses—a guide to conservation planning. United States of Agriculture Handbook, (1978).
10. Renard, K.G., Foster, G.R., Weesies, G., McCool, D. & Yoder, D. Predicting soil erosion by water: a guide to conservation planning with the Revised Universal Soil Loss Equation (RUSLE). US Government Printing Office Washington, (1997).
11. Foster, G. R. User’s Reference Guide: Revised Universal Soil Loss Equation (RUSLE2), US Department of Agriculture, Agricultural Research Service, Washington DC, (2004).
12. McGregor, K. C., Binger, R. L., Bowie, A. J. & Foster, G. R. Erosivity index values for northern Mississippi. Trans. Asae 38(4), 1039–1047 (1995).
13. VanDijk, A. J. M., Bruinzeel, L. A. & Rosewell, C. J. Rainfall intensity–kinetic energy relationships: A critical literature appraisal. J. Hydrol. 261(1), 1–23 (2002).
14. Nearing, M. A., Yin, S. Q. & Borrelli, P. Rainfall erosivity: An historical review. CATENA 157, 357–362 (2017).
15. Ramon, R., Minella, J. P. G., Merten, G. H., de Barros, C. A. P. & Canale, T. Kinetic energy estimation by rainfall intensity and its usefulness in predicting hydrosedimentological variables in a small rural catchment in southern Brazil. CATENA 148, 176–184 (2017).
16. USDA-Agricultural Research Service. Science Documentation Revised Universal Soil Loss Equation Version 2. https://www.ars.usda.gov/ARSwaterFiles/60060505/RUSLE/RUSLE2_Science_Doc.pdf (2013).
17. Xie, Y., Yin, S. & Liu, B. Models for estimating daily rainfall erosivity in China. J. Hydrol. 535, 547–558 (2016).
18. Wang, Y., Tan, S., Liu, B. & Yang, Y. Estimating rainfall erosivity by incorporating seasonal variations in parameters into the Richardson model. *J. Geogr. Sci.* 27(3), 275–296 (2017).
19. Wischmeier, W. H. Use and misuse of the universal soil loss equation. *J. Soil Water Conserv.* 31(1), 5–9 (1976).
20. Cohen, M. J., Shepherd, K. D. & Walsh, M. G. Empirical reformulation of the universal soil loss equation for erosion risk assessment in a tropical watershed. *Geoderma* 124, 235–252 (2004).
21. Lee, J. H. & Heo, I. H. Evaluation of estimation methods for rainfall erosivity based on annual precipitation in Korea. *J. Hydrol.* 409(1), 30–48 (2011).
22. Nash, J. E. & Sutcliffe, J. V. River flow forecasting through conceptual models part I—A discussion of principles. *J. Hydrol.* 10(3), 282–290 (1970).
23. Petkovšek, G. & Miko, M. Estimating the R factor from daily rainfall data in the sub-Mediterranean climate of southwest Slovenia. *Hydrol. Sci. J.* 49(5), 869–877 (2004).
24. Tian, G., Ling, Y., Chen, X. M. & Liu, T. Comparative study on several rainfall erosiveness models in Liashui Basin. *Acta Redol. Sin.* 47(1), 7–12 (2010) (in Chinese).
25. Wu, Q. J., Wu, F. Q., Zhai, Y. B. & Wu, G. Y. Comparative study of rainfall erosion model in loess Plateau in the North of Shannxi province. *J. Irrig. Drain.* 30(6), 138–140 (2011) (in Chinese).
26. Yu, B. F., Hashim, G. & Eusof, Z. Estimating the R-factor with limited rainfall data: A case study from peninsular Malaysia. *J. Soil Water Conserv.* 56(2), 101–105 (2001).
27. Yue, B. J., Shi, Z. H. & Fang, N. F. Evaluation of rainfall erosivity and its temporal variation in the Yanhe River catchment of the Chinese Loess Plateau. *Nat. Hazards* 74(2), 585–602 (2014).
28. Lu, H. & Yu, B. F. Spatial and seasonal distribution of rainfall erosivity in Australia. *Soil Res.* 40(6), 887–901 (2002).
29. Meng, X. L., Rosenthal, R. & Rubin, D. B. Comparing correlated correlation coefficients. *Psychol. Bull.* 111(1), 172–175 (1992).
30. Zheng, M. G. & Chen, X. A. Statistical determination of rainfall-runoff erosivity indices for single storms in the Chinese Loess Plateau. *PLoS ONE* 10(3), e0117989 (2015).
31. Silva, A. M. D. Rainfall erosivity map for Brazil. *CATENA* 57(3), 251–259 (2004).
32. Shamshad, A., Azhari, M. N., Isa, M. H., Hussin, W. M. A. W. & Parida, B. P. Development of an appropriate procedure for estimation of RUSLE EI30 index and preparation of erosivity maps for Pulau Penang in Peninsular Malaysia. *CATENA* 72(3), 423–432 (2008).
33. Ufoegbune, G. et al. Rainfall erosivity pattern of Ogun river basin area (Nigeria) using modified Fournier index. *Eur. Water* 35, 23–29 (2011).
34. Costea, M. Using the Fournier indexes in estimating rainfall erosivity. *Case study—the Secasal Mare Basin*, 313–320 (2012).
35. Arnoldus, H. Methodology used to determine the maximum potential average annual soil loss due to sheet and rill erosion in Morocco. *Nature* 414, 405–406 (1977).
36. Hernando, D. & Romana, M. G. Estimating the rainfall erosivity factor from monthly precipitation data in the Madrid Region (Spain). *J. Hydrol. Hydromech.* 63(1), 55–62 (2015).
37. Yu, B. F. & Rosewell, C. J. An assessment of daily rainfall erosivity model for New South Wales. *Aust. J. Soil Res.* 34(1), 139–152 (1996).
38. Renard, K. G. & Freimund, J. R. Using monthly precipitation data to estimate the R-factor in the revised USLE. *J. Hydrol.* 157(1–4), 287–306 (1994).
39. Mikhailova, E. A., Bryant, R. B., Schwager, S. J. & Smith, S. D. Predicting rainfall erosivity in Honduras. *Soil Sci. Soc. Am. J.* 61(1), 273–279 (1997).
40. Wischmeier, W. H. & Smith, D. D. Rainfall energy and its relationship to soil loss. *Trans. Am. Geophys. Union* 39(2), 285–291 (1958).
41. Zhang, W. B. & Fu, J. S. Rainfall erosivity estimation under different rainfall amount. *Resour. Sci.* 25(1), 35–41 (2003) (in Chinese).
42. Wu, S. Y. The simplified rainfall erosivity model and spatial and temporal distribution in Dabie Mountains in Anhui. *Soil Water Conserv. China* 4, 12–13 (1994) (in Chinese).
43. Zhang, W. B., Xie, Y. & Liu, Y. Rainfall erosivity estimation using daily rainfall amounts. *Sci. Geogr. Sin.* 22(6), 705–711 (2002) (in Chinese).
44. Men, M. X., Yu, Z. R. & Xu, H. Study on the spatial pattern of rainfall erosivity based on geostatistics in Hebei Province, China. *Front. Agric. China* 2(3), 281–289 (2007).
45. Liu, B. Y., Guo, S. Y. & Li, Z. G. Water erosion sample survey in China. *Soil Water Conserv.* 10, 26–34 (2013) (in Chinese).
46. Yu, B. F. Rainfall erosivity and its estimation for Australia’s tropics. *Aust. J. Soil Res.* 36(1), 143–165 (1998).
47. Yang, X. H. & Yu, B. F. Modelling and mapping rainfall erosivity in New South Wales, Australia. *Soil Res.* 53(2), 178–189 (2015).
48. Fournier, F. Climat et érosion: la relation entre féroson du sol par eau et les precipitations atmospheriques. *Presses universitaires de France*, (1960).

Author contributions
X.M. wrote the main manuscript text and M.Z. reviewed and revised the original draft. All authors have read and agreed to the published version of the manuscript.

Funding
The study was financially supported by the Guangdong Foundation for Program of Science and Technology Research (2020B1111530001), GuangDong Basic and Applied Basic Research Foundation (2021A1515011552), Guangdong Provincial Science and Technology Project (2021B1212050019), Meizhou Science and Technology Plan Project (2020B0204001) and GDAS’ Project of Science and Technology Development (2022GDASZH-2022010203).

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to M.Z.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
Open Access  This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2022, corrected publication 2022