Estimating the rainfall erosivity factor from monthly precipitation data in the Madrid Region (Spain)

David Hernando*, Manuel G. Romana

Department of Civil Engineering – Transport, Technical University of Madrid, Profesor Aranguren 3, 28040 Madrid, Spain.
Corresponding author. Tel.: +34 616 585 256. Fax: +34 916 633 354. E-mail: davhernando@gmail.com

DOI: 10.1515/johh-2015-0003

J. Hydrol. Hydromech., 63, 2015, 1, 55–62

Abstract: The need for continuous recording rain gauges makes it difficult to determine the rainfall erosivity factor (R-factor) of the Universal Soil Loss Equation in regions without good spatial and temporal data coverage. In particular, the R-factor is only known at 16 rain gauge stations in the Madrid Region (Spain). The objectives of this study were to identify a readily available estimate of the R-factor for the Madrid Region and to evaluate the effect of rainfall record length on estimate precision and accuracy. Five estimators based on monthly precipitation were considered: total annual rainfall (P), Fournier index (F), modified Fournier index (MFI), precipitation concentration index (PCI) and a regression equation provided by the Spanish Nature Conservation Institute (RICONA). Regression results from 8 calibration stations showed that MFI was the best estimator in terms of coefficient of determination and root mean squared error, closely followed by P. Analysis of the effect of record length indicated that little improvement was obtained for MFI and P over 5-year intervals. Finally, validation in 8 additional stations supported that the equation \( R = 1.05 \cdot MFI \) computed for a record length of 5 years provided a simple, precise and accurate estimate of the R-factor in the Madrid Region.

Keywords: Rainfall erosivity; R-factor; Universal Soil Loss Equation; Modified Fournier index; Soil erosion.

INTRODUCTION

Erosion models are a powerful tool for soil loss evaluation and land management. One of the most widely-used models is the Universal Soil Loss Equation (USLE) developed by Wischmeier and Smith (1961, 1965, 1978). Compared to more complex physically-based erosion models, USLE is a simplistic empirical model. Despite the simplicity of USLE, determination of the rainfall erosivity factor (R-factor) is no easy task. The R-factor is the product of the total kinetic energy (\( E \)) multiplied by the maximum 30-minute intensity (\( I_{30} \)). The R-factor at a particular location is then obtained as the average of annual \( E \cdot I_{30} \) values over long time intervals (over 20 years) to include apparent cyclical rainfall patterns (Wischmeier and Smith, 1978).

Although there are several equations for computing the kinetic energy of a storm (Brown and Foster, 1987; Wischmeier and Smith, 1978), all of them require continuous recording rain gauges with time resolution of at least 15 minutes. This need for continuous recording makes it difficult to determine the R-factor in many regions where good spatial and temporal data coverage is scarce. This is the case for the Madrid Region (Spain), where the R-factor has been only computed by the Spanish Nature Conservation Institute (ICONA) at 16 rain gauge stations based on rainfall data recorded from 1950 to 1985 (ICONA, 1988).

There have been many attempts worldwide to establish correlations between the R-factor calculated by the prescribed method and more readily available rainfall data, such as daily and monthly precipitation (Angulo-Martínez and Bugeura, 2009; Bonilla and Vidal, 2011; Colotti, 2004; Diodato, 2004; Diodato and Bellochi, 2007; Lee and Heo, 2011; Loureiro and Coutinho, 2001; Petkovšek and Mikosi, 2004; Renard and Freimund, 1994; Salako, 2008; Yu and Rosewell, 1996; Yu et al., 2001). Nevertheless, most of the obtained equations have limited application out of the areas in which they were developed without a thorough validation analysis.

Therefore, the objectives of this study were: (1) to identify a readily available estimate of the rainfall erosivity factor for the Madrid Region, and (2) to evaluate the effect of rainfall record length on the precision and accuracy of the estimates.

MATERIAL AND METHODS

Rainfall erosivity estimators

Based on the literature review conducted, five estimators of rainfall erosivity were selected for this study: total annual rainfall (P), Fournier index (F), modified Fournier index (MFI), Oliver’s precipitation concentration index (PCI) and a regression model proposed by the Spanish Nature Conservation Institute (RICONA). Other factors such as Hudson’s \( KE > 25 \) index (Hudson, 1971), Lal’s \( A_{FE} \) index (Lal, 1976), Onchev’s \( P/Si \) universal index (Onchev, 1985), and Burst factor (Smithen and Schulze, 1982) were not considered since they still require continuous recording. A further description of selected estimators is provided below.

Fournier (1960) conducted a regression analysis between sedimentation in rivers and several rainfall variables, finding a high correlation between the total annual erosion and the distribution coefficient of rainfall or, most commonly named, Fournier index:

\[
F = \frac{p^2}{P}
\]

where \( F \) is the Fournier index, \( p \) is the precipitation in the wettest month and \( P \) is the total annual rainfall.

Due to the simplicity in calculating the Fournier index, Arnoldus (1980) attempted to correlate this index with the known values of the rainfall erosivity factor for 164 rainfall stations in the United States and 14 stations in West Africa. The results were not satisfactory because of the different behavior of each index: while the R-factor adds all erosive storms, the Fournier index only captures those storms out of the month with
the highest precipitation within the denominator (total annual rainfall). Thus, Arnoldus (1980) proposed a modification of the Fournier index in which the storms that occur outside the month of maximum rainfall increase the overall value of the index, obtaining significantly higher coefficients of determination \((r^2 > 0.80)\):

\[
MFI = \sum_{i=1}^{12} \frac{P_i^2}{P}
\]

(2)

in which \(MFI\) is the modified Fournier index, \(p_i\) is the monthly rainfall and \(P\) is the total annual rainfall. Colotti (2004) reported that the Food and Agriculture Organization (FAO) of the United Nations used the modified Fournier index as an erosion estimate according to the following general equation:

\[
R = a \cdot MFI + b
\]

(3)

where \(R\) is the rainfall erosivity factor, \(MFI\) is the modified Fournier index, and \(a\) and \(b\) are two regional fitting parameters.

Ferro et al. (1991) stated that the \(F_i\) index, which represents the average value of the modified Fournier index for an interval of \(N\) years, is better correlated with the \(R\)-factor than both the mean annual precipitation and the modified Fournier index:

\[
F_F = \frac{1}{N} \sum_{j=1}^{N} \left( \sum_{i=1}^{12} \frac{P_i^2}{P_j} \right) = \frac{1}{N} \sum_{j=1}^{N} MFI_j
\]

(4)

where \(p_{ij}\) is the rainfall (mm) in month \(i\) of year \(j\), \(P_j\) is the annual precipitation (mm) of year \(j\), and \(N\) is the number of years considered. The \(F_F\) index was also successfully used by Ferro et al. (1999) to estimate the \(R\)-factor in southern Italy and southeastern Australia.

Oliver (1980) proposed an index of rainfall concentration in order to estimate the aggressiveness of storms from the temporal variability of monthly precipitation:

\[
PCI = 100 \sum_{i=1}^{12} \frac{P_i^2}{P^2}
\]

(5)

where \(PCI\) is the precipitation concentration index, \(p_i\) is the monthly rainfall and \(P\) is the total annual rainfall.

In Spain, ICONA (1988) proposed the following regression equation to estimate the rainfall erosivity factor in the Madrid Region:

\[
R_{ICONA} = e^{-0.834 \cdot PMEX^{1.314} \cdot MR^{-0.388} \cdot F_{24}^{0.563}}
\]

(6)

where \(R_{ICONA}\) is the rainfall erosivity factor (MJ cm ha\(^{-1}\) h\(^{-1}\) year\(^{-1}\)) as estimated by ICONA, \(PMEX\) is the maximum monthly precipitation (mm), \(MR\) is the total rainfall from October to May (mm), and \(F_{24}\) is the ratio of the square of the maximum annual rainfall in 24 hours (mm) to the sum of the maximum monthly rainfall in 24 hours (mm):

\[
F_{24} = \frac{(P_{24,\text{annual}})^2}{\sum_{i=1}^{12} P_{24,\text{ai}}}
\]

(7)

Annual values for the period covered by each rainfall station were calculated for the five estimators. A series of values for each estimator were then obtained by averaging annual values over time intervals. These series were used to evaluate the effect of record length through regression analyses. The effect of record length on each estimator was studied using the following time intervals: 1, 2, 5, 10, 15 and 20 years, described by Eq. 8:

\[
X_N = \frac{1}{N} \sum_{i=1}^{N} X_i
\]

(8)

where \(X_N\) represents the values of the estimator (\(R_{ICONA}, P, F, MFI\) and \(PCI\)) for a record length of \(N\) consecutive years, and \(X_i\) is the annual value of the estimator in year \(i\). Eq. 8 was applied to all consecutive 1, 2, 5, 10, 15 and 20-year intervals within the period covered by each rainfall station.

**Study area**

The Madrid Region is located in central Spain (Figure 1), between the Atlantic Ocean and the Mediterranean Sea, in a high plateau around 600 m above sea level (a.s.l.). The study area lies between 39º53’N–41º10’N and 3º03’W–4º34’W, covering a total extension of 8022 km\(^2\). The region is in the shape of a triangle (Figure 1), with the Central System mountain range on the NW side (elevations above 2400 m a.s.l.). From the base of the Central System, it begins a grade that ends in the Tagus valley (SE corner), with elevations below 500 m a.s.l. The area is dominated by a Mediterranean climate, characterized by seasonal temperatures, summer drought and erratic rainfall. The average annual rainfall of the area ranges from 1500 mm in the NW to 400 mm in the SE.

**Database**

The rain gauge network provided by the Spanish Meteorological Agency (AEMET) in the Madrid Region consists of more than 170 rainfall stations. However, ICONA (1988) only determined the \(R\)-factor according to the prescribed method at 16 stations. To date, no more attempts to determine the \(R\)-factor in additional stations have been published.

In this study, 8 stations were used for regression analysis (model calibration) between the single computed \(R\)-factor reported by ICONA (1988) and the five estimators presented previously (\(R_{ICONA}, P, F, MFI\) and \(PCI\)). The equations obtained from these 8 calibration stations were then validated in 8 additional stations. Locations of the 16 stations are shown in Figure 1. Table 1 presents the station name, elevation, available data period covered and number of complete years. For each station, monthly rainfall and maximum monthly rainfall in 24 hours were provided by AEMET.

**Statistical models**

Two statistical models were selected for this study. First, a simple linear regression with intercept term, as defined by equation 9:

\[
R = \beta_0 + \beta_1 \cdot X + \varepsilon
\]

(9)

where \(R\) is the rainfall erosivity factor, \(\beta_0\) is the intercept term, \(\beta_1\) is the slope, \(X\) is the estimator (\(R_{ICONA}, P, F, MFI\) and \(PCI\)) and \(\varepsilon\) represents the error.

It should be noted that the intercept term \(\beta_0\) is just a fitting parameter which has no physical meaning since no erosion should occur for zero rainfall. Therefore, a simple linear
Estimating the rainfall erosivity factor from monthly precipitation data in the Madrid Region (Spain)

Regression through the origin (no intercept term) was defined as second statistical model:

\[ R = \beta_2 \cdot X + \varepsilon \]  

(10)

in which \( R \) is the rainfall erosivity factor, \( \beta_2 \) is the slope, \( X \) is the estimator and \( \varepsilon \) represents the error.

The assumptions for both models were that errors are independent of each other and normally distributed with a mean of zero and constant variance. Based on available rainfall data, a series of values for each estimator (\( R_{ICONA}, P, F, MFI \) and \( PCI \)) and record length (1, 2, 5, 10, 15 and 20 years) were determined. These values were then correlated with the single computed \( R \)-factor reported by ICONA (1988) for each station.

Fig. 1. Location of the study area and selected rainfall stations.

Table 1. Characteristics of rainfall stations selected for calibration and validation.

<table>
<thead>
<tr>
<th>Code</th>
<th>Station name</th>
<th>Elevation (m.a.s.l.)</th>
<th>( R )-factor (MJ·cm⁻¹·ha⁻¹·h⁻¹·year⁻¹) (ICONA, 1988)</th>
<th>Period covered</th>
<th>Complete years</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Calibration</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2462</td>
<td>Navacerrada</td>
<td>1890</td>
<td>194</td>
<td>1947–2007</td>
<td>61</td>
</tr>
<tr>
<td>3112</td>
<td>Puentes Viejas Dam</td>
<td>960</td>
<td>89</td>
<td>1947–2005</td>
<td>58</td>
</tr>
<tr>
<td>3119</td>
<td>Fuente El Saz</td>
<td>645</td>
<td>63</td>
<td>1947–2007</td>
<td>61</td>
</tr>
<tr>
<td>3190</td>
<td>Hoyo de Manzanares</td>
<td>1100</td>
<td>124</td>
<td>1956–2007</td>
<td>52</td>
</tr>
<tr>
<td>3195</td>
<td>Madrid Retiro</td>
<td>667</td>
<td>65</td>
<td>1947–2007</td>
<td>61</td>
</tr>
<tr>
<td>3196</td>
<td>Madrid (Cuatro Vientos Airport)</td>
<td>687</td>
<td>74</td>
<td>1947–2007</td>
<td>61</td>
</tr>
<tr>
<td>3200</td>
<td>Getafe (Air Base)</td>
<td>617</td>
<td>53</td>
<td>1951–2007</td>
<td>57</td>
</tr>
<tr>
<td>3341</td>
<td>San Juan Dam</td>
<td>540</td>
<td>105</td>
<td>1952–1999</td>
<td>47</td>
</tr>
<tr>
<td><strong>Validation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3116</td>
<td>El Atazar Dam</td>
<td>960</td>
<td>85</td>
<td>1970–2005</td>
<td>35</td>
</tr>
<tr>
<td>3121E</td>
<td>El Vellón Dam</td>
<td>850</td>
<td>87</td>
<td>1970–2007</td>
<td>38</td>
</tr>
<tr>
<td>3129</td>
<td>Madrid (Barajas Airport)</td>
<td>582</td>
<td>65</td>
<td>1951–2007</td>
<td>57</td>
</tr>
<tr>
<td>3169</td>
<td>Alcalá de Henares (Canaleja)</td>
<td>600</td>
<td>62</td>
<td>1955–2004</td>
<td>50</td>
</tr>
<tr>
<td>3182E</td>
<td>Arganda</td>
<td>530</td>
<td>66</td>
<td>1972–2007</td>
<td>35</td>
</tr>
<tr>
<td>3193O</td>
<td>Majadahonda (MAFRE)</td>
<td>725</td>
<td>99</td>
<td>1976–2005</td>
<td>27</td>
</tr>
<tr>
<td>3223</td>
<td>Pezuela de las Torres</td>
<td>852</td>
<td>69</td>
<td>1968–2007</td>
<td>36</td>
</tr>
<tr>
<td>3274</td>
<td>San Lorenzo de El Escorial (Royal Seat)</td>
<td>1028</td>
<td>130</td>
<td>1976–2005</td>
<td>27</td>
</tr>
</tbody>
</table>
RESULTS AND DISCUSSION

Regression analysis (model calibration)

A preliminary analysis showed that regression results barely changed whether the period covered was that employed by ICONA (1950–1985) or a broader period. Thus, the decision was to consider the complete rainfall data set that was available for each rainfall station. Regression results obtained from the 8 calibration stations are presented in Table 2. These results include regression equation, coefficient of determination ($r^2$) and root mean squared error (RMSE) for each estimator and record length. PCI was found to be poorly correlated with $R$-factor so no results are presented hereafter. Assumptions of the statistical models were successfully validated. It was observed that the regression model through the origin ($β_0 = 0$) provided $r^2$ and RMSE values extremely close to those provided by the regression model with intercept term ($β_0 ≠ 0$), especially for record lengths over 5 years. In fact, the intercept term was found not to be statistically significant in the regression analysis, so the simpler regression model without intercept term was proposed in Table 2 for record lengths of 5 years or more.

Overall, rather good results were provided by $R_{\text{ICONA}}$, $P$, $F$ and $MFI$, especially for record lengths over 5 years. $MFI$ was the estimator with the best results in terms of $r^2$ and RMSE, closely followed by $P$. It should be pointed out that the equation obtained for the total annual rainfall ($R = 0.15 \cdot P$) closely agreed with previous results proposed by Van der Knijff et al. (2000) for Southern Europe ($R = 0.13 \cdot P$). Lastly, the slope of 0.84 obtained for $R_{\text{ICONA}}$ when long record lengths were considered suggested that the available regression equation proposed by ICONA (Eq. 6) may overpredict the $R$-factor by approximately 16%. A possible explanation is that Eq. 6 was developed for a broad area of Spain (not only the Madrid Region) and it may not perfectly fit the climatic pattern of the Madrid Region.

Effect of record length on estimate precision and accuracy

As can be inferred from Table 2, record length had a direct effect on regression models: as record length increased, $r^2$ increased and RMSE decreased. The reason is that as record length increased, the annual values of the estimators were averaged over a longer time interval as defined by Eq. 8. Thus, this ‘smoothing’ effect translated into less dispersion and, consequently, a better fitting.

In order to further analyze the effect of record length on the precision and accuracy of the estimates, two additional statistics were evaluated: the coefficient of variation (CV) and the mean absolute percentage error (MAPE). CV is defined as the ratio of the standard deviation to the mean of a sample, expressed as a percentage. This statistic represents the variability of an index
about its mean value and represents the precision of the estimator. MAPE is a measure of the error estimating the R-factor (accuracy) and is determined as follows:

\[
\text{MAPE} (\%) = \frac{1}{N} \sum_{j=1}^{N} \frac{|R_j - \hat{R}_j|}{R_j}
\]  

(11)

where \( R \) is the known value of the rainfall erosivity factor (Table 1), \( \hat{R}_j \) is the estimated value from the regression model, and \( N \) is the number of data points for a given record length.

\( CV \) results are presented in Figure 2. As can be seen in the figure, \( P \) and \( MFI \) were the estimators with the lowest \( CV \) for any record length. \( CV \) for \( R_{ICONA} \) and \( F \) was almost twice that for \( P \) and \( MFI \). It can also be observed that \( CV \) values barely changed for \( P \) and \( MFI \) over 5 years of record length whereas \( R_{ICONA} \) and \( F \) required at least 10 years to become relatively constant.

Figure 3 shows that \( MFI \) was overall the estimator with the lowest MAPE. Again, small reductions in MAPE were obtained for record lengths over 5 to 10 years. For \( P \) and \( MFI \) (the estimators with the lowest MAPE values), \( MAPE \) was below 13% for a record length as short as 5 years. In addition, a Tukey-Kramer honest significance difference (HSD) test revealed that \( P \) and \( MFI \) were statistically different from \( R_{ICONA} \) and \( F \) for a 5% significance level when a 5-year record length was considered.

The above results showed that record length definitely increased both the precision and accuracy of the estimates when time intervals up to 10 years were considered, but there was almost no improvement beyond 10 years. In fact, both \( P \) and \( MFI \) provided quite good results (in terms of \( CV \) and MAPE) for a record length of only 5 years. According to these results, a record length of 5 years could be considered adequate for \( P \) and \( MFI \), while a record length of 10 years could be proposed for \( R_{ICONA} \) and \( F \). Therefore, regression results in Table 2 were reduced to one single equation for each estimator, as shown in Table 3.

**Table 3. Simplified regression models obtained from the 8 calibration stations.**

<table>
<thead>
<tr>
<th>Estimator</th>
<th>Proposed record length</th>
<th>Regression equation</th>
<th>( r^2 )</th>
<th>RMSE (MJ·cm·ha(^{-1})·h(^{-1})·year(^{-1}))</th>
<th>( CV ) (%)</th>
<th>MAPE (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( R_{ICONA} )</td>
<td>10 years</td>
<td>( R = 0.83 \cdot R_{ICONA,10} )</td>
<td>0.80</td>
<td>20</td>
<td>14</td>
<td>12</td>
</tr>
<tr>
<td>( P )</td>
<td>5 years</td>
<td>( R = 0.15 \cdot P_5 )</td>
<td>0.87</td>
<td>16</td>
<td>12</td>
<td>13</td>
</tr>
<tr>
<td>( F )</td>
<td>10 years</td>
<td>( R = 2.51 \cdot F_{10} )</td>
<td>0.87</td>
<td>16</td>
<td>13</td>
<td>14</td>
</tr>
<tr>
<td>( MFI )</td>
<td>5 years</td>
<td>( R = 1.05 \cdot MFI_5 )</td>
<td>0.91</td>
<td>13</td>
<td>12</td>
<td>11</td>
</tr>
</tbody>
</table>

**Fig. 3.** Mean absolute percentage error for each estimator and record length (model calibration).

**Fig. 2.** Coefficient of variation for each estimator and record length (model calibration).

The simplified regression models obtained from the 8 calibration stations (Table 3) were used to estimate the R-factor in 8 additional stations. Results are shown in Figure 4, in which the y-axis represents measured R-factor reported by ICONA (Table 1) and the x-axis represents predicted R-factor. As can be seen in the figure, the four estimators provided fairly good results for the proposed record lengths, observing only some dispersion for the station with the highest R-factor (San Lorenzo de El Escorial, \( R = 130 \) MJ·cm·ha\(^{-1}\)·h\(^{-1}\)·year\(^{-1}\)). RMSE and MAPE were also evaluated for the validation stations. Figure 5 shows that RMSE values obtained from validation were rather close to those previously obtained from calibration, \( P \) and \( MFI \) being the estimators with the lowest RMSE. With respect to MAPE, Figure 6 indicates that validation results were similar to those observed from calibration, except for \( R_{ICONA} \), which experienced an important increase in estimate error due to the differences observed in one of the stations (San Lorenzo de El Escorial).
Fig. 4. Scatter plot of $R$-factor measured by ICONA vs. $R$-factor predicted by estimates for proposed record lengths in MJ·cm·ha$^{-1}$·h$^{-1}$·year$^{-1}$ (model validation).

Fig. 5. Comparison of root mean squared error (RMSE) results for calibration and validation.
CONCLUSIONS

A detailed regression analysis of 16 rainfall stations throughout the Madrid Region resulted in the identification and validation of a readily available estimate of the rainfall erosivity factor. The findings and conclusions of this study can be summarized as follows:

- The modified Fournier index (MFI) provided the best estimates among the assessed indices for any record length, in terms of coefficient of determination ($r^2$) and root mean squared error (RMSE) obtained from 8 calibration stations.
- Regression models through the origin resulted in $r^2$ and RMSE values extremely close to those obtained when intercept term was used, especially for record lengths of 5 years or more. In fact, the intercept term was found not to be statistically significant. Therefore, estimates with no intercept were selected for record lengths of 5 years or more.
- The analysis of the effect of record length on estimate precision and accuracy concluded that little improvement was obtained for $P$ and MFI when time intervals over 5 years were considered. For $R_{ICONA}$ and Fournier index, no improvement was observed over 10 years. Therefore, regression models were reduced to one single equation for each estimator.
- A Tukey-Kramer HSD test revealed that $P$ and MFI were statistically different from $R_{ICONA}$ and $F$ for a 5% significance level when a 5-year record length was considered.
- Validation results from 8 additional stations confirmed that the equation $R = 1.05 \text{ MFI}_5$, in which $R$ is expressed in MJ cm ha$^{-1}$ h$^{-1}$ year$^{-1}$ and $\text{MFI}_5$ represents the modified Fournier index obtained for a record length of 5 years (in mm), provided a simple, precise and accurate estimate of the rainfall erosivity factor in the Madrid Region.
- The slope obtained for $R_{ICONA}$ suggested that the regression equation previously proposed by ICONA (Eq. 6) may overpredict the $R$-factor in the Madrid Region by approximately 16%.

Results from this study support that the rainfall erosivity factor can be successfully approximated at a local level (regional scale) by using readily available estimators based on monthly precipitation. Furthermore, the effect of record length can be assessed through the novel methodology described in this paper, which is based on the analysis of different statistics (namely coefficient of determination, root mean squared error, coefficient of variation and mean absolute percentage error) for different time intervals.

REFERENCES


Fig. 6. Comparison of mean absolute percentage error (MAPE) results for calibration and validation.
Alimentación, Madrid. (In Spanish.)


Received 5 March 2014
Accepted 21 August 2014