We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Chapter

Rainfall Erosivity and Its Estimation: Conventional and Machine Learning Methods

Konstantinos Vantas, Epaminondas Sidiropoulos and Chris Evangelides

Abstract

Rainfall erosivity concerns the ability of rainfall to cause erosion on the surface of the earth. The difficulty in modeling the distribution, the size, and the terminal velocity of raindrops in relation to the detachment of soil particles led to the use of more tractable rainfall indices. Thus, in the universal soil loss equation (USLE), the coefficient of rainfall erosivity, *R*, was introduced. This coefficient is based on the product of the rainfall kinetic energy of a storm and its maximum 30-minute intensity. An important problem in the application of USLE and its revisions in various parts of the world concerns the computation of *R*, which requires pluviograph records with a length of at least 20 years. For this reason, empirical equations have been developed that are based on coarser rainfall data, such as daily, monthly, or yearly, which are available on larger spatial and temporal extents. However, the lack of denser data is dealt more effectively by means of machine learning methods. Computational systems for this purpose were recently developed based on feed-forward neural networks, yielding significantly better results.

Keywords: rainfall, erosivity, machine learning, erosivity density, universal soil loss equation, nonlinear regression, neural networks

1. Introduction

Soil erosion is the detachment, transport, and deposition of soil particles, and it takes place in the course of one or more processes. These processes are natural, such as rainfall with the energy that it carries, surface water runoff, wind, and gravity. The observed processes may be combined with human activities and works, such as intensive cultivations, deforestation, construction of public works, and mining activities [1].

Earth terrain is strongly influenced by erosion, which progresses in geological time, in case it is a purely natural process. Otherwise, when human activities and works are involved, the phenomenon of erosion is accelerated. This accelerated erosion may cause uncontrollable soil loss with negative consequences for human health, the natural ecosystems, the climate, as well as the economy.

The universal soil loss equation (USLE) with its revisions (RUSLE, RUSLE2) can be used to predict the average rate of soil erosion by grouping the numerous parameters that affect soil loss into a set of factors and is the prediction equation

most widely used in the world [2]. According to Nearing et al. [3], "Soil loss refers to the amount of sediment that reaches the end of a specified area on a hillslope that is experiencing net loss of soil by water erosion." Development of soil loss equations began in 1940, and the universal soil loss equation was developed at the National Runoff and Soil Loss Data Center established in 1954 by the Science and Education Administration [4].

The USLE represents an erosion model developed for the prediction of soil losses in an average long-term sense. It is based on the knowledge of the physical characteristics of the field area under study, along with the prevailing cropping and management system. USLE has been widely tested in field conditions, and therefore its validity has been established.

USLE consists of the product of six factors, whose numerical values can be specified depending on a particular location. There is a considerable variation in the resulting erosion values, if observations are limited within short periods. However, long-term averages correspond more satisfactorily to predictions.

The soil loss equation is

$$A = R \cdot K \cdot L \cdot S \cdot C \cdot P \tag{1}$$

where the meaning of the symbols is given in [5] exactly as follows:

- *A is the computed soil loss per unit area, expressed in the units selected for K and for the period selected for R.*
- *R, the rainfall and runoff factor, is the number of rainfall erosion index units, plus a factor for runoff from snowmelt or applied water where such runoff is significant.*
- *K, the soil erodibility factor, is the soil loss rate per erosion index unit for a specified soil as measured on a unit plot, which is defined as a 72.6-ft length of uniform 9% slope continuously in clean-tilled fallow.*
- *L, the slope length factor, is the ratio of soil loss from the field slope length to that from a 72.6-ft length under identical conditions.*
- *S, the slope-steepness factor, is the ratio of soil loss from the field slope gradient to that from a 9% slope under otherwise identical conditions.*
- *C, the cover and management factor, is the ratio of soil loss from an area with specified cover and management to that from an identical area in tilled continuous fallow.*
- *P, the support practice factor, is the ratio of soil loss with a support practice like contouring, strip cropping, or terracing to that with straight-row farming up and down the slope.*

Out of the six factors, only two have units, the erosivity *R* and the erodibility *K*. The remaining four, i.e., slope steepness, slope length, cropping management, and components of control practice, are dimensionless factors, because they represent ratios in relation to the unit plot [4].

Regarding *R*, it needs to be observed that its units stem from its definition as the energy multiplied by maximum 30-minute intensity term. In contrast, the erodibility factor values were determined empirically by calibration against measured erosion data. This is important due to the fact that if there is a change on the definition of the *R* factor, then the values of the *K* factor should be recalculated [3].

2. Computation of the *R* factor

The *R* factor is computed for time periods greater than 20 years, so that wet or dry rainfall periods will be incorporated, eliminating any bias. It is set equal to the average of the sum of the erosivity values for every year's rainfalls. The *R* coefficient is defined as the product of the kinetic energy of an erosive rainfall event by the maximum intensity of a 30-minute duration rainfall, during the rainfall event:

$$R = \frac{1}{n}\sum_{j=1}^{n}\sum_{k=1}^{m_j}(EI_{30})_k \tag{2}$$

where n is the number of years in the record, m_j the number of erosive events during year j, and EI_{30} (MJ mm ha^{-1} h^{-1}) the rainfall erosivity for event k.

The erosivity of a particular event is

$$EI_{30} = \left(\sum_{r=1}^{m} e_r \cdot v_r\right) \cdot I_{30} \tag{3}$$

where e_r is the kinetic energy per unit of rainfall (MJ ha^{-1} mm^{-1}); v_r is the rainfall depth (mm) for the time interval r of the hyetograph, which has been divided into $r = 1, 2, \ldots, m$ subintervals; and I_{30} is the maximum rainfall intensity for a 30-minute duration.

For the computation of e_r, numerous empirical relations (**Figure 1**) involving rain intensity have been proposed [6]. Thus, in USLE, the empirical relation of Wischmeier and Smith [4] was used:

$$e_r = 0.119 + 0.0873 log_{10}(i) \tag{4}$$

with the upper limit of 0.283 MJ ha^{-1} mm^{-1} if i>76 mm h^{-1}, where i is rainfall intensity.

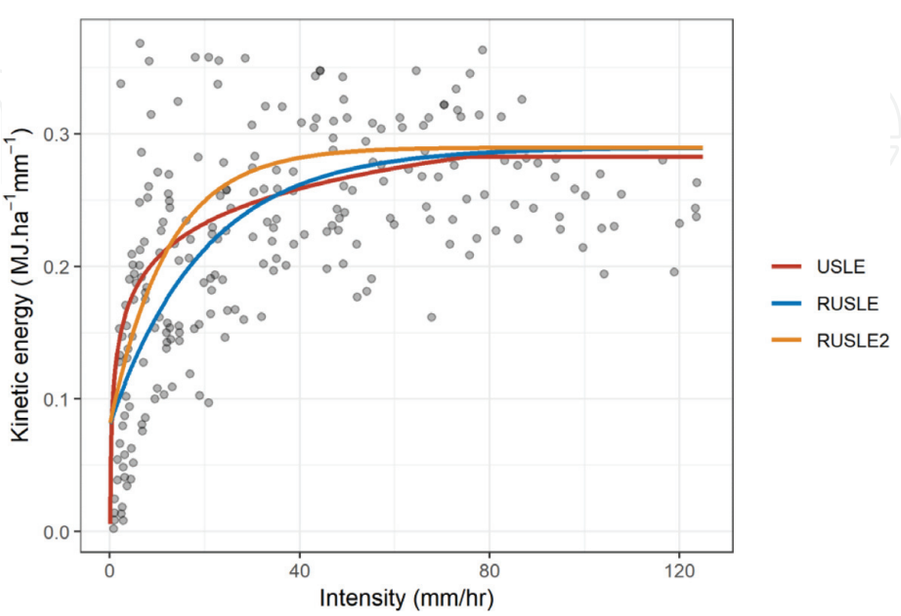


Figure 1.
The three different kinetic energy equations used in USLE and its revisions. The points are actual data, showing the relation between intensity and the kinetic energy of the rainfall, converted to SI units and coming from Haan et al. [61].

Later, in RUSLE [5], the exponential relation of Brown and Foster [7] was used:

$$e_r = 0.29\left(1 - 0.72\ e^{-0.05i}\right) \tag{5}$$

Upon comparison of the equations used in USLE and RUSLE, McGregor and Mutchler [8] found out that for rain intensities up to 35 mm/h, Eq. (5) yields results by 12% less than those of Eq. (4), and they proposed a modification in the value 0.05 that controls the rate of change of e_r with i. Thus, in RUSLE, the following relation has been adopted:

$$e_r = 0.29\left(1 - 0.72\ e^{-0.082i}\right) \tag{6}$$

Eq. (5) was developed for an application concerning reordered rainfall intensity data and not natural rainfall data [6]. The systematic underestimation of R has been shown in many other studies [3, 7–9], so Eq. (5) should not be used for calculations. The rules that apply in order to single out the storms causing erosion and to divide rainfalls of large duration in RUSLE2 [10, 11] are:

1. A rainfall event is divided into two parts, if its cumulative depth for duration of 6 hours at a certain location is less than 1.27 mm.

2. A rainfall is considered erosive if it has a cumulative value greater than 12.7 mm.

3. All rainfalls with extreme EI_{30} values and a return period greater than 50 years are deleted.

The current revised version of USLE, RUSLE2 [10], introduced the erosivity density (ED), as a measure of rainfall erosivity per unit rainfall to develop erosivity values for the USA, because ED requires shorter record lengths, as 10 years leads to acceptable results, allows more missing data than R, and is independent of the elevation:

$$ED_j = \frac{\sum_{k=1}^{m_j} (EI_{30})_k}{P_j} \tag{7}$$

where m_j is the number of storms during a time period j, $(EI_{30})_k$ the erosivity of storm k, and P_j the total precipitation height for the period j (usually monthly or annual).

Vantas et al. [12] using a numerical scheme with data from Greece showed that ED is more robust against the presence of missing precipitation values, as reflected in the following results:

- Using only 5% of the data, annual R values are underestimated on average by 85%, when the average estimation error of ED values is 50%.

- In the presence of 50% of the data, R values are underestimated by 49%, while at the same time, the estimation error of ED is 20%.

The use of ED permits the utilization of data consisting of daily rainfall values that are more abundant in comparison to data from pluviographs. The details of the method, which was employed in the USA in combination with kriging, may be found in the RUSLE2 Science Documentation [10].

3. Methods for the estimation of *R*

The lack of dense time series from pluviographs gives rise to many difficulties in the application of USLE and of its revisions in many countries. In order to cope with these difficulties, many models have been developed, based on rainfall depth values for various time steps (daily, monthly, and yearly), under specific spatial parameters and climatological data. Thus, a number of researchers reported good correlation results between *R* and yearly rainfall in many countries of the world by using various schemes ranging from simple parametric equations to geostatistical models.

These methods were applied in the countries of West Africa [13]; Switzerland taking account of elevation, aspect, longitude, and altitude [14]; the USA [15]; India [16]; Spain [17]; Italy [18]; China taking account of maximal rainfalls of 1-hour and 1-day durations per year [19]; Korea [20]; Mexico [21]; Honduras in combination with elevation [22]; Rhodesia [23]; and Hawaii [24] as well as for the development of a European [25] and Universal model [26].

A different model followed in several countries concerns the correlation, by the use of parametric equations, of the yearly values of *R* to the monthly values of rainfall depth. Those equations were applied in Venezuela [27], Germany [28], the USA [15], Italy [29], Iran in combination with the yearly and the maximum daily rainfall per year [30], North Africa [31], Morocco [32], Nigeria [33], South Italy and Southeast Australia [34], Uruguay [35], and Sudan [36].

Finally, for the estimation of the yearly value of *R*, several researchers used parametric equations for the initial determination of the daily *R* value and the corresponding daily rainfall depths. Those equations were applied in the Eastern USA [37], Australia [38], Spain [17], Kenya [39], China [40], Nigeria [41], North Africa [42], South Italy [43], Peru [44], and Slovenia [45].

Indicatively, some of the empirical equations of the literature are given below in dimensionless form, representing various parametrizations and methodologies. In West African countries, Roose [13] developed a simple linear relation between yearly values of *R* and rainfall *P*:

$$R = \alpha_1 + \beta_1 \cdot P \tag{8}$$

where α_1 και β_1 are linear regression parameters.

In the USA, Renard and Freimund [15] proposed the following nonlinear equations for continental regions of the country, in which no rainfall intensity data exist, also for yearly values of *R* and rainfall *P*:

$$R = \alpha_2 \cdot P^{\beta_2} \tag{9}$$

or

$$R = \alpha_3 + \beta_3 \cdot P^{\gamma_3} \tag{10}$$

where α_2, β_2, α_3, β_3, and γ_3 are nonlinear regression parameters.

In Morocco, Arnoldus [32] used the modified Fournier index, which is equal to

$$\text{MFI} = \frac{\sum_{i=1}^{12} p_{m,i}}{P} \tag{11}$$

where *P* is the yearly rainfall and $p_{m,i}$ the monthly rainfall from January to December, and developed the formula

$$\mathrm{R} = \alpha_5 \cdot MFI^{\beta_5} \tag{12}$$

where α_5 and β_5 are nonlinear regression parameters.

Richardson et al. [37] used the daily rainfall values for the estimation of the corresponding daily erosivity values R_d:

$$\mathrm{R_d} = \alpha_{m,s} \cdot P_d^{\beta_{m,s}} \tag{13}$$

where the parameters $\alpha_{m,s}$ and $\beta_{m,s}$ correspond to station s and to month m and are evaluated by means of nonlinear regression.

A serious issue with the applicability of Eq. (13) is the difficulty to compute its coefficients in dry areas or months due to small number of erosive events. Another issue, even in wet periods, is the amount of missing values that also leads to a small set of observations and the same problems [46].

In order to reduce the computation burden of Eq. (13), Yu and Rosewell [38] proposed the following empirical equation:

$$R_d = \alpha_s \cdot \left[1 + \beta_s \cdot \cos\left(\frac{\pi}{6}m - \omega\right)\right] P_d^{\gamma_s} \tag{14}$$

where the parameters α_s, β_s, and γ_s are different only for the different stations s and ω is the month with the highest average of daily R values.

It must be noted that the logarithmic transformation and the subsequent application of linear regression must be avoided in the above equations, because in that case the minimization is applied to the average of the differences of the logarithms of the aggregated daily EI_{30}values (coming from pluviograph data) and the logarithms of the estimated R_d values (coming from daily rainfall data) and not of the corresponding absolute values themselves. Specifically, the application of the log transformation to Eq. (13) resulted in a −10% systematic error, when data from RUSLE2 was used, as noted on p. 59 of [10].

A different approach for the estimation of R, in the presence of missing values, was followed by Vantas and Sidiropoulos [47]. They compared empirical equations to machine learning methods, extracting for the latter methods better results, in terms of statistical significance, compared to the former methods.

4. Utilization of neural networks for the estimation of *R*

4.1 Data

The data utilized in the analysis were taken from the Greek National Bank of Hydrological and Meteorological Information [48] and came from 84 meteorological stations (**Figure 2**). By adding up the years of record registered on the various stations, the time series comprised a total of 2425 years of pluviograph records, with an average of 28.9 years per station. The time step was 30 min for the time period from 1953 to 1997, and the data coverage was equal to 56%.

For the above described data, it was deemed useful to change the time step and aggregate the data into weekly values, because 57% of the recordings were associated with storms occupying time periods covering parts of more than one calendar day, although only 17% of the storms had duration of more than 24 hours.

Under the time step of 1 week, it was found out that 80% of the values emanated from a single storm. The storms that were crossed temporally by 2 consecutive weeks were assigned to the first of the 2 weeks, and they comprised only 7% of the

data. Thus, through the use of weekly instead of daily values, divisions of storms due to time step were prevented, and a measure of rainfall duration in days has been added to the observations.

4.2 Exploratory data analysis

Table 1 gives the concise statistics of the cumulative rainfall, the duration, and the erosivity EI_{30} of the erosive events that resulted following RUSLE2 rules.

In the sequel, a series of three figures is given that contain:

1. The relation of the precipitation height of the erosive events to the EI_{30} value in **Figure 3**, where a linear relation appears with a wide opening between the logarithms of rainfall values and the logarithms of EI_{30}.

2. The relation of EI_{30} to the months in **Figure 4**, where it turns out that in July its maximum appears with an almost sine-like variation of the median values per month.

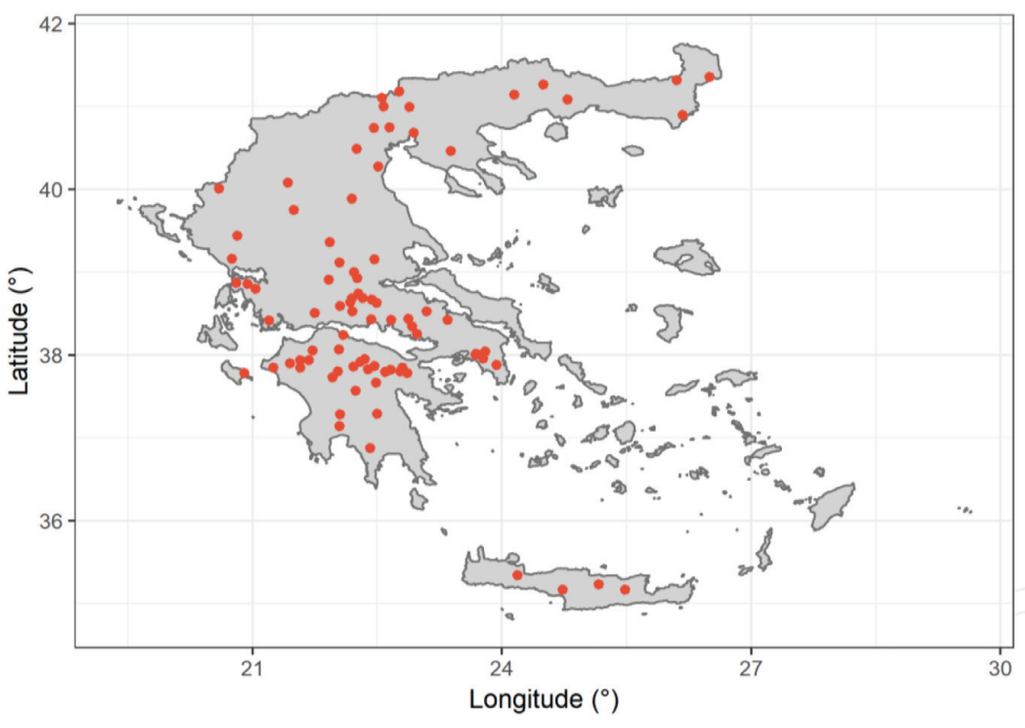


Figure 2.
Stations' location coming from Greece used in the analysis.

Variable	Min	Mean	Median	Max	SD	SW	KR	CV
EI_{30} (MJ mm ha^{-1} h^{-1})	0.7	89.1	44.2	3845.0	157.4	7.6	101.0	1.7
Precipitation (mm)	12.7	29.4	22.2	322.4	21.9	3.4	19.8	0.7
Duration (h)	0.5	13.0	10.5	152.5	10.5	2.5	13.6	0.8

SD is an abbreviation for standard deviation, SW for skew, KR for kurtosis, and CV for coefficient of variation.

Table 1.
Average statistical properties of cumulative rainfall, the duration and EI_{30} of the erosive events.

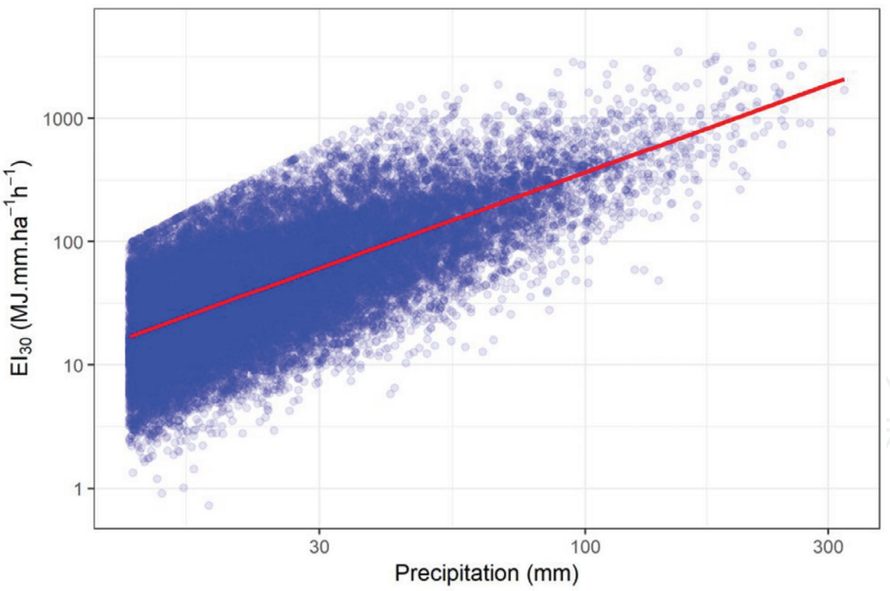


Figure 3.
Relation of cumulative precipitation of the erosive events to the EI_{30}. The red line is the LOESS line.

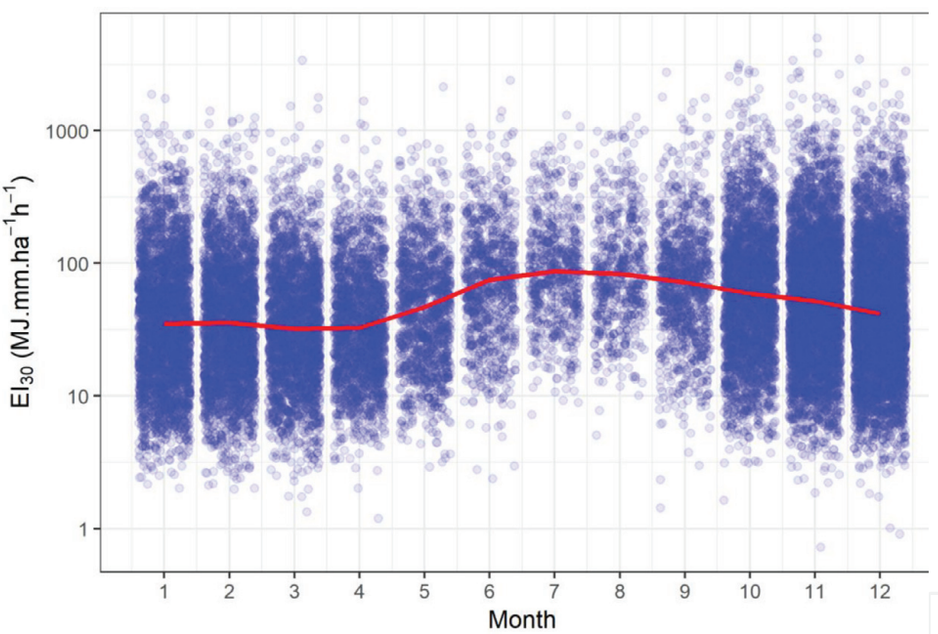


Figure 4.
Relation of cumulative precipitation of the erosive events to the EI_{30}. The red line is the LOESS line that is fitted using the median values of EI_{30} per month.

3. The relation of the duration of erosive events to EI_{30} in **Figure 5**, where it is shown that for the selected duration widths, the relation to the median values of EI_{30} is parabolic.

In these three figures, local polynomial regression fitting (LOESS, [49]) was used as a nonparametric method to produce smooth curves through the plotted variables. In that method, the fitting is done locally for a given point x using all the points in the neighborhood of x and the distances of these points from x as weights.

From the above diagrams and from the statistics of the erosive events, it can be seen that the estimation of EI_{30} values is a nonlinear problem with skewed data and with expected results of a relatively low accuracy. This point is particularly reinforced by **Figure 3**, in which it is shown that for a specific rainfall depth, such as

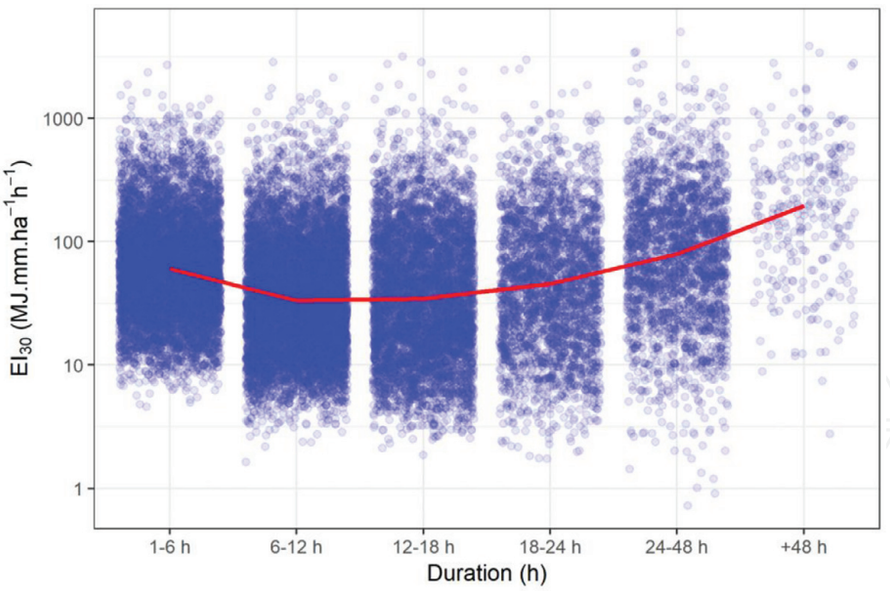


Figure 5.
Relation of cumulative precipitation of the erosive events to the EI_{30}. The red line is the LOESS line that is fitted using the median values of EI_{30} per duration width.

30 mm, a very large range of EI_{30} values, from about 10 to 700 MJ mm ha^{-1} h^{-1}, is to be assigned, corresponding to the erosive events connected to 30 mm.

4.3 Methods

The outcomes, denoted as $y^{(i)}$, where (i) is the index of a sample consisting of 19,688 values, represented weekly cumulative rainfall erosivity as calculated from pluviograph data. The features included the weekly cumulative rainfall of erosive events, the month to which the above individual m values are referred, the altitude of the meteorological stations, and the number of the days of the week for which rainfall is recorded. The matrix of the features is denoted as $x^{(i)}$, while $h(x^{(i)})$ denotes the output vector corresponding to the hypothesis in question (i.e., as computed either from the neural network or the empirical equations). The subscripts *test* and *train* mark quantities that belong to the testing or the training set, respectively.

The erosivity estimation problem is set up as a scheme of machine learning. The data were split using 70% of them as the training set and 30% as a testing one. As measures of the out-of-sample error were used:

(a) The coefficient of determination:

$$R^2 = 1 - \frac{\sum_{i=1}^{m} \left(h\left(x_{test}{}^{(i)}\right) - y_{test}{}^{(i)}\right)^2}{\sum_{i=1}^{m} \left(\hat{y}_{train} - y_{test}^{(i)}\right)^2} \tag{15}$$

with this form of R^2:

- $R^2 = 0$ means that there is no improvement comparing to a simplistic model that returns the average of the training set values $\hat{y}_{train}$.
- $R^2 = 1$ means a perfect algorithm.
- $R^2 < 0$ means an algorithm that makes prediction worse than the base model.

(b) The root mean-squared error:

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}\left(h(x_{test}{}^{(i)}) - y_{test}{}^{(i)}\right)^2} \tag{16}$$

(c) The mean absolute error:

$$MAE = \frac{1}{m}\sum_{i=1}^{m}\left|h\left(x_{test}{}^{(i)}\right) - y_{test}{}^{(i)}\right| \tag{17}$$

In Eqs. (15)–(17), m is the number of the samples in the test set, $h\left(x_{test}{}^{(i)}\right)$ is the estimations from a trained algorithm (i.e., either from a trained neural network or from a fitted empirical equation) used as inputs, the testing set x_{test} and y_{test} are the outcomes from calculations by means of pluviograph data, and $\hat{y}_{train}$ is the average of the training set erosivity values coming also from pluviograph data.

In order to compare neural network performance to that of empirical equations, two alternative exponential models discussed in the previous section were tried, namely, those given by Eqs. (13) and (14) and leading to optimal adjustments of two respective hypotheses. The latter were determined by minimizing the following objective function by the trust-region-reflective method [50, 51]:

$$J(\theta) = \frac{1}{2m}\sum_{i=1}^{m}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)^2 \tag{18}$$

where θ denotes the vector of the respective parameters and $y^{(i)}$ are the outcomes, as defined in the beginning of this section.

Neural networks aim at mimicking the function of the brain. They constitute powerful nonlinear regression methods. The result or exit of the neural network is produced via a series of intermediate nodes arranged in successive layers, characterized as hidden. These layers mediate between an initial layer of input nodes and a final layer of exit nodes. The hidden nodes define linear combinations of the data contained in the initial input layer. These linear combinations are further transformed by a nonlinear function, which possesses a continuous derivative, such as the hyperbolic tangent function:

$$g(x) = \tanh(x) \tag{19}$$

The example in **Figure 6** displays three layers, of which the one to the left is the input layer, the middle one is the hidden layer, and the third one is the output layer. In the neural network, $\alpha_i^{(j)}$ denotes the activation of node i in layer j, and $\Theta^{(j)}$ denotes the matrix of weights that serve as coefficients of the linear transformation from layer j to layer $j+1$. The values at the output nodes in the example of **Figure 6** are:

$$\begin{aligned}
\alpha_1^{(2)} &= g\left(\Theta_{10}^{(1)}\cdot x_0 + \Theta_{11}^{(1)}\cdot x_1 + \Theta_{12}^{(1)}\cdot x_2 + \Theta_{13}^{(1)}\cdot x_3\right)\\
\alpha_2^{(2)} &= g\left(\Theta_{20}^{(1)}\cdot x_0 + \Theta_{21}^{(1)}\cdot x_1 + \Theta_{22}^{(1)}\cdot x_2 + \Theta_{23}^{(1)}\cdot x_3\right)\\
\alpha_3^{(2)} &= g\left(\Theta_{30}^{(1)}\cdot x_0 + \Theta_{31}^{(1)}\cdot x_1 + \Theta_{32}^{(1)}\cdot x_2 + \Theta_{33}^{(1)}\cdot x_3\right)\\
\alpha_4^{(2)} &= g\left(\Theta_{40}^{(1)}\cdot x_0 + \Theta_{41}^{(1)}\cdot x_1 + \Theta_{42}^{(1)}\cdot x_2 + \Theta_{43}^{(1)}\cdot x_3\right)\\
\alpha_1^{(3)} &= g\left(\Theta_{10}^{(2)}\cdot \alpha_0^{(2)} + \Theta_{11}^{(2)}\cdot \alpha_1^{(2)} + \Theta_{12}^{(2)}\cdot \alpha_2^{(2)} + \Theta_{13}^{(2)}\cdot \alpha_3^{(2)} + \Theta_{14}^{(2)}\cdot \alpha_4^{(2)}\right)
\end{aligned} \tag{20}$$

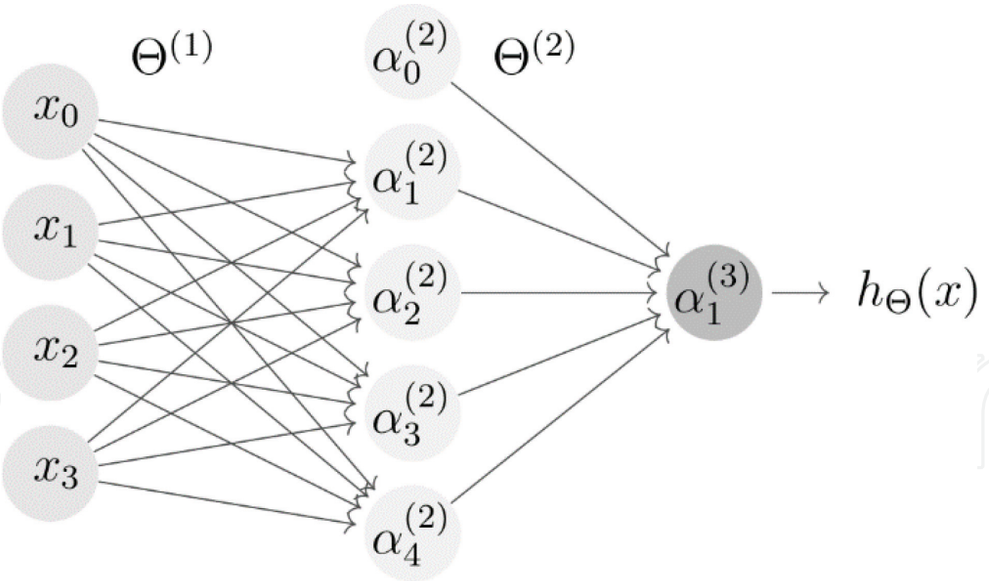


Figure 6.
A neural network with a single hidden layer.

The optimization of the $\Theta^{(j)}$ weights is executed in such a way as to minimize the sum of squares of the differences between computed and measured results:

$$J(\Theta) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\Theta\left(x^{(i)}\right) - y^{(i)} \right)^2 \tag{21}$$

where m is the number of values $h_\Theta\left(x^{(i)}\right)$ obtained from the network using $x^{(i)}$ as input data and $y^{(i)}$ as the calculated values (i.e., the EI_{30} values as computed from the pluviographic data).

This minimization is a difficult optimization problem, since there are no constraints related to these parameters. The parameters are usually initialized by random values, and in the sequel, specialized algorithms are used for their determination. A survey about neural networks and their forms as shallow and deep can be found in Schmidhuber [52] and Goodfellow et al. [53].

The architecture of the neural network used in the analysis included two hidden layers that had 32 and 16 neurons, respectively. In order to keep $h_\Theta\left(x^{(i)}\right)$ nonnegative, the second hidden layer of the neural network used the rectifier activation function [54]:

$$g(x) = \max(0, x) \tag{22}$$

The training set values x_{train} were used to compute the averages $\overline{x}$ and standard deviation $sd(x)$ and normalize both x_{train} and x_{test} using the normalizing transformation:

$$N\left(x^{(i)}\right) = \frac{x^{(i)} - \overline{x}}{sd(x)} \tag{23}$$

The training of the networks was performed by the method of early stopping [55] by utilizing a random validation set consisting of 10% of the training data, so as to avoid over fitting of the neural network.

The data importing, calculation of EI_{30} values, and analysis were done using the R language for statistical computing and graphics [56] using the packages:

Method	RMSE	MAE	R^2
Richardson et al.	141.94	65.46	0.59
Yu and Rosewell	140.36	64.33	0.60
Neural network	116.83	54.33	0.73

R^2 values are unitless and RMSE and MAE values are in MJ mm ha^{-1} h^{-1}. The metrics refer to the weekly erosivity time step.

Table 2.
Estimation of the out-of-sample error metrics for the two empirical equations and the neural network.

Method	RMSE	MAE	R^2
Richardson et al.	261.73	144.63	0.76
Yu and Rosewell	259.80	141.48	0.76
Neural network	223.01	124.24	0.82

R^2 values are unitless and RMSE and MAE values are in MJ mm ha^{-1} $h^{-1}y^{-1}$. The values are calculated using the estimation of the annual erosivity values.

Table 3.
Estimation of the out-of-sample error metrics for the two empirical equations and the neural network.

hydroscoper [57], hyetor [58], and ggplot2 [59]. The open source software library TensorFlow [60] was used in order to train the neural network.

4.4 Results

The comparison of the algorithms was based on their performance on the testing set using the three different error metrics and two time steps: weekly and annual.

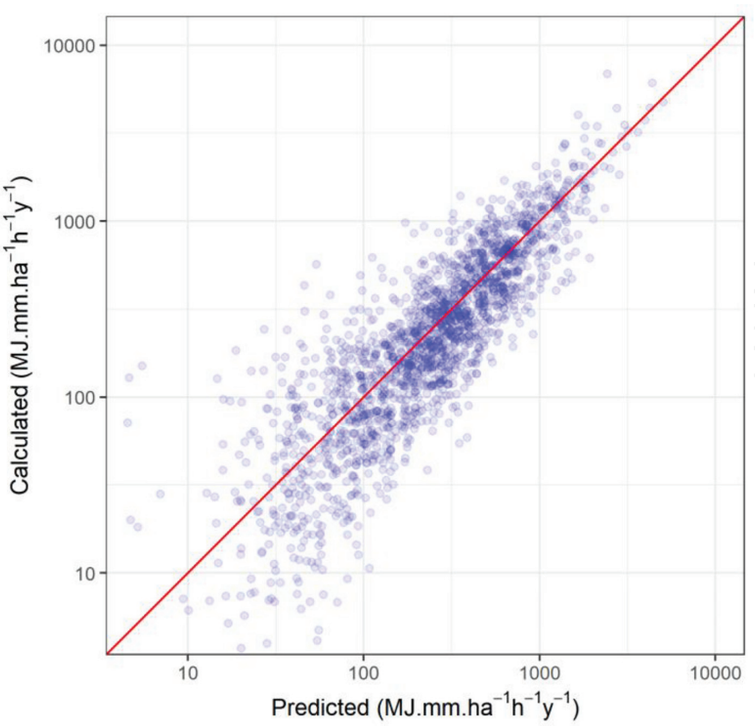


Figure 7.
Calculated annual erosivity values coming from pluviograph data versus predicted values using the Yu and Rosewell model on the testing set. With red line, the identity function $f(x) = x$ is symbolized.

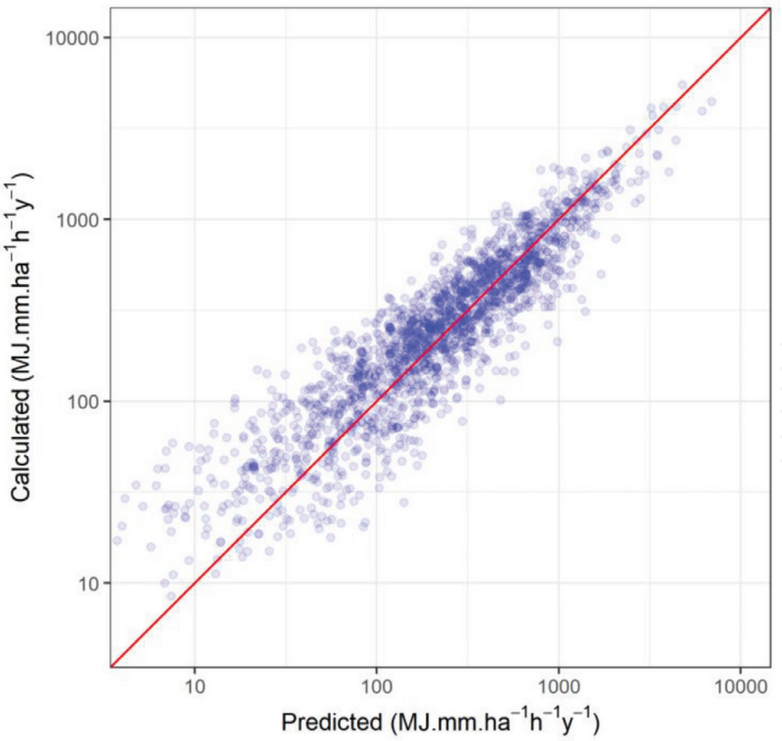


Figure 8.
Calculated annual erosivity values coming from pluviograph data versus predicted values using the neural network on the testing set. With red line, the identity function f(x) = x is symbolized.

The weekly and the annual calculated values were produced from the aggregated erosivity values coming from the pluviograph data. The annual predicted values are also the aggregated values coming from the output of the corresponding algorithm (the predicted weekly erosivity values in our case). The annual time step was used in order to examine if there is a systematic error on the estimations of the algorithms, which was not the case.

These results are given in **Tables 2** and **3**, where it is shown that the neural network outperformed the two parametric methods, which both had similar results. Specifically, the machine learning method had better performance that ranged from 8 to 22%, depending on the error metric and time step.

In **Figures 7** and **8**, the annual erosivity values are presented as calculated in the test set versus the values estimated from the 84 fitted Yu and Rosewell equations for each station and the single neural network. In these two diagrams, the neural network's estimations are closer to the calculated ones than those coming from the empirical equations.

5. Conclusions

The universal soil loss equation with its revisions represents the established model and the means for estimating annual long-term average soil loss. An important factor of this equation is the rainfall erosivity, *R*, which is closely connected to the energy carried by rainfall. Rainfall data are essential for the estimation of *R*. In particular, data from pluviographs are more suitable for that task. However, such data are scarce, and various empirical methods have been presented for estimating R from coarser data. These methods are reviewed in this chapter.

Also, various difficulties are discussed that are associated with the issues of missing values and the separation of erosive events. In addition, indications are given of the adverse nature of the relation of erosivity to precipitation characteristics. All these facts raise objections to the suitability of empirical models and lead to the use of data-driven methods that will be able to handle more effectively the problematic nature of the data and also to discover more reliably the hidden relations between erosivity, precipitation, seasonality, and local characteristics. This is achieved, because in machine learning procedures, no human intervention or bias is involved in the selection or in the forms of these hidden nonlinearities.

More specifically, neural networks, as machine learning tools, are employed in this chapter for the estimation of erosivity, in comparison to two different empirical equations of the literature. The results demonstrate, first of all, the superior performance of the machine learning method and its ability for generalization, so as to perform equally well upon new data. Another advantage of the proposed learning algorithm is that it produces a single model accessing all available data, in contrast to empirical equations, where a researcher must fit and handle as many nonlinear equations as the number of available stations or the product of the number of the stations by the number of the used seasons, making the analysis cumbersome and prone to errors.

The latter fact is exemplified on the country-wide data from Greece, used in this chapter. The single neural network, which was produced, outperformed 84 different empirical nonlinear equations, each one of which was fitted using data from each station. The comparison of the performance was made using test data sets to which neither the empirical equations nor the neural network had any access during their training.

In conclusion, the present chapter gives an indicative and characteristic sample of the use of machine learning methods in the problems associated with erosivity. There are still many related problems, such as the spatial and temporal distribution of erosivity, with significant potential for the application of machine learning methods. The same is true of many areas of hydrology and water resources.

Conflict of interest

The authors declare no conflict of interest.

IntechOpen

IntechOpen

Author details

Konstantinos Vantas, Epaminondas Sidiropoulos* and Chris Evangelides
Faculty of Engineering, Aristotle University of Thessaloniki, Thessaloniki, Macedonia, Greece

*Address all correspondence to: nontas@topo.auth.gr

IntechOpen

© 2019 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. CC BY

References

[1] Boardman J, Poesen J. Soil erosion in Europe: Major processes, causes and consequences. In: Soil Erosion in Europe. Wiley-Blackwell; 2006. pp. 477-487

[2] Kinnell PIA. Event soil loss, runoff and the universal soil loss equation family of models: A review. Journal of Hydrology. 2010;**385**(1):384-397

[3] Nearing MA, Yin S, Borrelli P, Polyakov VO. Rainfall erosivity: An historical review. Catena. 2017;**157**: 357-362

[4] Wischmeier WH, Smith DD. Predicting Rainfall Erosion Losses— A Guide to Conservation Planning. USDA, Science and Education Administration; 1986

[5] Renard KG, Foster GR, Weesies GA, McCool DK, Yoder DC. Predicting Soil Erosion by Water: A Guide to Conservation Planning with the Revised Universal Soil Loss Equation (RUSLE). Vol. Agriculture Handbook No. 703. Washington, DC: United States Department of Agriculture; 1997

[6] Brown L, Foster G. Storm erosivity using idealized intensity distributions. Transactions of the ASAE. 1987; **30**(2):379-386

[7] Ramon R, Minella JP, Merten GH, de Barros CA, Canale T. Kinetic energy estimation by rainfall intensity and its usefulness in predicting hydrosedimentological variables in a small rural catchment in southern Brazil. Catena. 2017;**148**:176-184

[8] McGregor KC, Mutchler CK. Status of the R factor in northern Mississippi. In: Soil Erosion: Prediction and Control. Soil Conservation Society of America; 1976. pp. 135-142

[9] Van Dijk A, Bruijnzeel LA, Rosewell CJ. Rainfall intensity–kinetic energy relationships: A critical literature appraisal. Journal of Hydrology. 2002; **261**(1–4):1-23

[10] USDA-ARS. Science Documentation, Revised Universal Soil Loss Equation Version 2 (RUSLE2). USDA-Agricultural Research Service; 2013

[11] Yin S, Nearing MA, Borrelli P, Xue X. Rainfall Erosivity: An overview of methodologies and applications. Vadose Zone Journal. 2017;**16**(12):0

[12] Vantas K, Sidiropoulos E, Loukas A. Temporal and elevation trend detection of rainfall erosivity density in Greece. In: 3rd International Electronic Conference on Water Sciences (ECWS-3); 2018

[13] Roose E. Erosion et ruissellement en Afrique de l'Ouest, vingt années de mesures en parcelles expérimentales. O. R.S.T.R.O.M. 1977;**78**:30-33

[14] Meusburger K, Steel A, Panagos P, Montanarella L, Alewell C. Spatial and temporal variability of rainfall erosivity factor for Switzerland. Hydrology and Earth System Sciences. 2012; **16**(1):167-177

[15] Renard KG, Freimund JR. Using monthly precipitation data to estimate the R-factor in the revised USLE. Journal of Hydrology. 1994;**157**:287-306

[16] Sudhishri S, Patnaik U. Erosion index analysis for Eastern Ghat high zone of Orissa. Indian Journal of Dryland Agricultural Research and Development. 2004;**19**:42-47

[17] Angulo-Martínez M, Beguería S. Estimating rainfall erosivity from daily precipitation records: A comparison among methods using data from the

Ebro Basin (NE Spain). Journal of Hydrology. 2009;**379**(1):111-121

[18] Diodato N. Estimating RUSLE's rainfall factor in the part of Italy with a Mediterranean rainfall regime. Hydrology and Earth System Sciences Discussions. 2004;**8**(1):103-107

[19] Zhang G, Nearing M, Liu B. Potential effects of climate change on rainfall erosivity in the Yellow River Basin of China. Transactions of the ASAE. 2005;**48**(2):511-517

[20] Lee J-H, Heo J-H. Evaluation of estimation methods for rainfall erosivity based on annual precipitation in Korea. Journal of Hydrology. 2011;**409**(1):30-48

[21] Millward AA, Mersey JE. Adapting the RUSLE to model soil erosion potential in a mountainous tropical watershed. Catena. 1999;**38**(2):109-129

[22] Mikhailova EA, Bryant R, Schwager S, Smith S. Predicting rainfall erosivity in Honduras. Soil Science Society of America Journal. 1997;**61**(1):273-279

[23] Stocking M, Elwell H. Rainfall erosivity over Rhodesia. Transactions of the Institute of British Geographers. 1976;**2**:231-245

[24] Lo A, El-Swaify S, Dangler E, Shinshiro L. Effectiveness of EI30 as an erosivity index in Hawaii. Soil Conservation Society of America. 1985;**1**:384-339

[25] Panagos P, Ballabio C, Borrelli P, Meusburger K, Klik A, Rousseva S, et al. Rainfall erosivity in Europe. Science of the Total Environment. 2015;**511**: 801-814

[26] Yang D, Kanae S, Oki T, Koike T, Musiake K. Global potential soil erosion with reference to land use and climate changes. Hydrological Processes. 2003; **17**(14):2913-2928

[27] Lujan DL, Gabriels D. Assessing the rain erosivity and rain distribution in different agro-climatological zones in Venezuela. Sociedade & Natureza. 2005;**1**(1):16-29

[28] Sauerborn P, Klein A, Botschek J, Skowronek A. Future rainfall erosivity derived from large-scale climate models—Methods and scenarios for a humid region. Geoderma. 1999;**93**(3): 269-276

[29] Lastoria B, Miserocchi F, Lanciani A, Monacelli G. An estimated erosion map for the Aterno-Pescara River basin. European Water. 2008;**21**(22):29-39

[30] Sepaskhah A, Sarkhosh P. Estimating storm erosion index in southern region of Iran. Iranian Journal of Science and Technology, Transaction B: Engineering. 2005;**29**(3):357-363

[31] Smithen A, Schulze R. The spatial distribution in Southern Africa of rainfall erosivity for use in the universal soil loss equation. Water SA. 1982;**8**(2): 74-78

[32] Arnoldus H. Methodology used to determine the maximum potential average annual soil loss due to sheet and rill erosion in Morocco. FAO Soils Bulletins (FAO). 1977;**34**:8-9

[33] Igwe C, Mbagwu J, Akamigbo F. Application of SLEMSA and USLE models for potential erosion hazard mapping in south-eastern Nigeria. International Agrophysics. 1999;**13**:41-48

[34] Ferro V, Porto P, Yu B. A comparative study of rainfall erosivity estimation for southern Italy and Southeastern Australia. Hydrological Sciences Journal. 1999;**44**(1):3-24

[35] Munka C, Cruz G, Caffera RM. Long term variation in rainfall erosivity in Uruguay: A preliminary Fournier approach. GeoJournal. 2007;**70**(4): 257-262

[36] Elagib N. Changing rainfall, seasonality and erosivity in the hyper-arid zone of Sudan. Land Degradation & Development. 2011;**22**(6):505-512

[37] Richardson C, Foster G, Wright D. Estimation of erosion index from daily rainfall amount. Transactions of the ASAE. 1983;**26**(1):153-156

[38] Yu B, Rosewell C. An assessment of a daily rainfall erosivity model for New South Wales. Soil Research. 1996;**34**(1): 139-152

[39] Angima S, Stott D, O'Neill M, Ong C, Weesies G. Soil erosion prediction using RUSLE for central Kenyan highland conditions. Agriculture, Ecosystems & Environment. 2003; **97**(1):295-308

[40] Xie Y, Yin S, Liu B, Nearing MA, Zhao Y. Models for estimating daily rainfall erosivity in China. Journal of Hydrology. 2016;**535**:547-558

[41] Salako F. Development of isoerodent maps for Nigeria from daily rainfall amount. Geoderma. 2010; **156**(3):372-378

[42] Le Roux J, Morgenthal T, Malherbe J, Pretorius D, Sumner P. Water erosion prediction at a national scale for South Africa. Water SA. 2008;**34**(3):305-314

[43] Capolongo D, Diodato N, Mannaerts CM, Piccarreta M, Strobl R. Analyzing temporal changes in climate erosivity using a simplified rainfall erosivity model in Basilicata (Southern Italy). Journal of Hydrology. 2008; **356**(1):119-130

[44] Elsenbeer H, Cassel D, Tinner W. A daily rainfall erosivity model for Western Amazonia. Journal of Soil and Water Conservation. 1993;**48**(5): 439-444

[45] Petkovšek G, Mikoš M. Estimating the R factor from daily rainfall data in the sub-Mediterranean climate of Southwest Slovenia. Hydrological Sciences Journal. 2004;**49**(5)

[46] Hollinger SE, Angel JR, Palecki MA. Spatial Distribution, Variation, and Trends in Storm Precipitation Characteristics Associated with Soil Erosion in the United States. Prepared for the United State Department of Agriculture; 2002. p. 103

[47] Vantas K, Sidiropoulos E. Imputation of erosivity values under incomplete rainfall data by machine learning methods. European Water. 2017;**57**:193-197

[48] Vafiadis M, Tolikas D, Koutsoyiannis D. HYDROSCOPE: The new Greek national database system for meteorological, hydrological and hydrogeological information. In: WIT Transactions on Ecology and the Environment. Vol.3. WIT Press; 1994. pp. 1-8

[49] Cleveland WS, Grosse E, Shyu WM. Local regression models. Statistical models in S. 1992;**2**:309-376

[50] Coleman T, Li Y. On the Convergence of Reflective Newton Methods for Large-Scale Nonlinear Minimization Subject to Bounds. Vol. 67. Ithaca, NY, USA: Cornell University; 1994

[51] Coleman TF, Li Y. An interior trust region approach for nonlinear minimization subject to bounds. SIAM Journal on Optimization. 1996;**6**(2): 418-445

[52] Schmidhuber J. Deep learning in neural networks: An overview. Neural Networks. 2015;**61**:85-117

[53] Goodfellow I, Bengio Y, Courville A, Bengio Y. Deep Learning. Vol. 1. Cambridge: MIT Press; 2016

[54] Glorot X, Bordes A, Bengio Y. Deep sparse rectifier neural networks. In:

Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics; FL, USA: Fort Lauderdale; 2011. pp. 315-323

[55] Wang C, Venkatesh SS, Judd JS. Optimal stopping and effective machine complexity in learning. Advances in Neural Information Processing Systems. 1994;**6**:303-310

[56] R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2018

[57] Vantas K. Hydroscoper: R interface to the Greek National Data Bank for hydrological and meteorological information. Journal of Open Source Software. 2018;**3**(23):625

[58] Vantas K. hyetor: R package to analyze fixed interval precipitation time series. 2018. Available from: https://kvantas.github.io/hyetor/

[59] Wickham H. ggplot2: Elegant Graphics for Data Analysis. New York: Springer-Verlag; 2009. p. 213

[60] Abadi M, Barham K, Chen J, Chen Z, Davis A, Dean J, et al. Tensorflow: A system for large-scale machine learning. In: OSDI. 2016. pp. 265-283

[61] Haan CT, Barfield BJ, Hayes JC. Design Hydrology and Sedimentology for Small Catchments. Academic Press; 1990. pp. 238-246