

# Computation of rainfall erosivity from daily precipitation amounts

Santiago Beguería<sup>a,\*</sup>, Roberto Serrano-Notivoli<sup>b</sup>, Miquel Tomas-Burguera<sup>a</sup>

<sup>a</sup>*Estación Experimental de Aula Dei - Consejo Superior de Investigaciones Científicas (EEAD-CSIC), Zaragoza, Spain*

<sup>b</sup>*Barcelona Supercomputing Center (BSC), Barcelona, Spain*

---

## Abstract

Rainfall erosivity is an important parameter in many erosion models, and the *EI30* defined by the Universal Soil Loss Equation is one of the best known erosivity indices. One issue with this and other erosivity indices is that they require continuous breakpoint, or high frequency time interval, precipitation data. These data are rare, in comparison to more common medium-frequency data, such as daily precipitation data commonly recorded by many national and regional weather services. Devising methods for computing estimates of rainfall erosivity from daily precipitation data that are comparable to those obtained by using high-frequency data is, therefore, highly desired. Here we present a method for producing such estimates, based on optimal regression tools such as the Gamma Generalised Linear Model and universal kriging. Unlike other methods, this approach produces unbiased and very close to observed *EI30*, especially when these are aggregated at the annual level. We illustrate the method with a case study comprising more than 1500 high-frequency precipitation records across Spain. Although the original records have a short span (the mean length is around 10 years), computation of spatially-distributed upscaling parameters offers the possibility to compute high-resolution climatologies of the *EI30* index based on currently available, long-span, daily precipitation databases.

---

\*Corresponding author

*Email addresses:* [santiago.begueria@csic.es](mailto:santiago.begueria@csic.es) (Santiago Beguería),  
[roberto.serrano@bsc.es](mailto:roberto.serrano@bsc.es) (Roberto Serrano-Notivoli), [mtomas@eead.csic.es](mailto:mtomas@eead.csic.es) (Miquel Tomas-Burguera)

*Keywords:* soil erosion, rainfall erosivity, RUSLE, USLE, R factor, EI30

---

## 1. Introduction

Soil erosion by water is recognised as one of the major problems affecting agricultural productivity and food security, as well as impacting ecosystem services such as water quality and regulation, net primary production, biodiversity, and others. Most models for estimating soil erosion rates by water often include one term accounting for the ability of rainfall to cause erosion, i.e. a rainfall erosivity factor. The best known rainfall erosivity index is the R-factor used in the Universal Soil Loss Equation, USLE (Wischmeier, 1959; Wischmeier and Smith, 1978), and its two revised versions, RUSLE (Renard et al., 1997) and RUSLE2 (USDA-ARS, 2013). At many sites worldwide the R-factor has been shown to be highly correlated with soil loss (Van der Knijff et al., 2000; Diodato, 2004; Shi et al., 2004; Hoyos et al., 2005; Cruse et al., 2006; Onori et al., 2006; Romero et al., 2007), and it is without doubt the most commonly used rainfall erosivity index. A recent review on the R-factor can be found on Nearing et al. (2017).

The development of regional or even larger scale maps or rainfall erosivity (sometimes referred to as iso-erodent maps) is crucial for quantifying soil loss by water erosion and for understanding its spatial patterns. Recent efforts include the rainfall erosivity map of Europe at the annual and monthly time scales (Panagos et al., 2015a; Ballabio et al., 2017), or the global erosivity map (Panagos et al., 2017b), all based on the R-factor. A larger number of studies have been published that produced erosivity maps at the country or more local scales, as summarised in the review by Yin et al. (2017). The development of long time series is also desirable for understanding the temporal patterns of rainfall erosivity, allowing to characterise its inter-annual variation and return period curves, assess the existence of temporal trends, the role of atmospheric teleconnections, or even the effects of climate change (Angulo-Martínez and Beguería, 2012; Fiener et al., 2013; Hanel et al., 2016; Hoomehr et al., 2016;

Qin et al., 2016; Lai et al., 2016; Panagos et al., 2017a; Wang et al., 2017; Chen  
30 and Zha, 2017).

Reliable estimation of rainfall erosivity requires continuous recording rain  
gauge (hyetograph) data or, alternatively, high-frequency (sub-hourly) time in-  
terval precipitation records. There are fewer rain gauges recording at these  
high-frequency intervals, and the length of their record is usually shorter, than  
35 more accessible daily precipitation records. This poses a problem of spatial but  
also temporal representativity, since long time series are required in order to  
compute reliable average seasonal or annual values. As revealed by a selection  
of recent regional studies, usually the number of rain gauges used is low with  
respect to the study area covered, and the temporal length of the records is, in  
40 most cases, short (Table 1). It has been noted, however, that R-factor values  
vary largely from one year to the next, even much more than precipitation, so  
long data series are required for calculating mean erosivity values within an  
acceptable level of certainty. It is widely accepted that 30-year periods are the  
norm in most climatological studies, and Wischmeier and Smith (1978) proposed  
45 using 22 years of data for obtaining a representative estimate of the USLE R-  
factor. Foster et al. (2003) recommended at least 15 years of data for the same  
purpose.

Coarser precipitation records, such as daily or monthly data, are on the other  
hand much more frequent and cover longer time spans. A method for computing  
50 rainfall erosivity from coarse precipitation records is, therefore, highly desirable.  
The general idea is to fit an empirical relation between erosivity values computed  
from high-frequency and coarser resolution precipitation data, using typically  
a limited dataset of a few high-frequency rain gauges. Once such relationship  
is established, it can be used for estimating rainfall erosivity based on a much  
55 larger set of (coarse resolution) rain gauges. Developing this basic idea, there  
are approximations based on annual or monthly rainfall, despite the inaccuracies  
involved in using low-frequency precipitation data (Banasik and Górski,  
1994; Renard and Freimund, 1994; Ferro et al., 1999; Apaydin et al., 2006;  
Hernando and Romana, 2015); while other studies proposed estimating rainfall

Table 1: Some regional studies on rainfall erosivity and characteristics of their datasets: location, number of raingauges (N), temporal length of the series (Years; range and mean value, if available), and sampling frequency ( $\delta T$ ).

<b>Reference</b>	<b>Location</b>	<b>N</b>	<b>Years</b>	<b><math>\delta T</math></b>
Panagos et al. (2017b)	World	3625	5-30	3-60 min
Panagos et al. (2015a)	Europe	1541	7-40 (17)	5-60 min
Hanel et al. (2016)	Czech Republic	17	51	30 min
Xie et al. (2016)	East China	16	30-40	1 min
Yin et al. (2015)	East China	18	30-40	1 min
Lee and Lin (2015)	South Taiwan	55	10	10 min
Fiener et al. (2013)	West Germany	10	70	1-5 min
Klik et al. (2015)	New Zealand	35	4-16 (11)	10 min
Borrelli et al. (2016)	Italy	386	5-10 (8.8)	30 min
Lobo and Bonilla (2015)	Central Chile	30	3-28	60 min
Bonilla and Vidal (2011)	Chile	16	17-22	30 min
Ma et al. (2014)	Yunan, China	7	5	10 min
Sanchez-Moreno et al. (2014)	Cape Verde	2	2, 7	3 min, 15 min
Panagos et al. (2016a)	Greece	80	(29.7)	30 min
Risal et al. (2016)	South Korea	76	15-19 (18)	10 min

60 erosivity from daily precipitation data (Richardson et al., 1983; Bagarello and D’Asaro, 1994; Yu and Rosewell, 1996; Yu et al., 2001; Petkovšek and Mikoš, 2004; Angulo-Martínez et al., 2009; Yang and Yu, 2015). However, as we shall discuss later, these approaches are often sub-optimal since they tended to produce biased results that often result in underestimation of rainfall erosivity.

65 The main objective of this study is to develop a method for estimating daily rainfall erosivity,  $EI30_{day}$ , from records of daily precipitation,  $P_{day}$ , with the purpose of computing long-term mean annual and monthly erosivity values. Alternative modelling approaches are discussed, and optimal regression tools for the problem under study, such as the Gamma Generalised Linear Model  
70 and universal co-kriging, are presented. A large database of high-frequency precipitation records covering continental Spain is used to illustrate the method. The use of a better statistical approach is the main difference with respect to our previous work (Angulo-Martínez and Beguería, 2009), but also the analysis of a much larger dataset encompassing a greater climatic variability offers a  
75 much more robust validation of the methodology.

The article is structured as follows: section 2 describes the dataset; section 3 explains the computation of event erosivity values,  $EI30$ , and discusses technical issues related to using datasets with different time resolutions; section 4 describes the computation of daily erosivity,  $EI30_{day}$ , and shows that event-  
80 based and daily-based erosivity yield comparable results; section 5 develops the statistical model of the relationship between  $EI30_{day}$  and daily precipitation,  $P_{day}$ , and discusses different model configurations; section 6 discusses the seasonal variation of the relationship between  $EI30_{day}$  and  $P_{day}$ , while section 7 does the same but for the spatial variation; section 8 presents a summary of the  
85 process and explains how to use the results of the previous analyses for computing monthly and annual  $EI30$  aggregates based on daily precipitation data; section 9 discusses the practical consequences of some assumptions required in the previous steps; and section 10 concludes.

Table 2: Characteristics of the dataset: name of the data provider, dataset code, number of stations ( $N$ ), start and end of the data series (span), temporal length of the series (years), sampling frequency ( $\delta T$ ) and precision ( $\delta P$ ).

Data provider	Set	N	Span	Years	$\delta T$	$\delta P$
Confederación Hidrográfica del Cantábrico	CAN	36	2008–2016	9	5 min	0.1 mm
Servei Meteorològic de Catalunya	CAT	188	2005–2016	12	10 min	0.2 mm
Confederación Hidrográfica del Duero	DUE	168	2011–2016	6	5/15 min	0.1 mm
Confederación Hidrográfica del Ebro	EBR	224	1997–2016	20	15 min	0.1 mm
Meteo Galicia	GAL	80	2000–2014	15	10 min	0.1 mm
Confederación Hidrográfica del Guadalquivir	GUA	56	2008–2013	6	15 min	0.1 mm
Confederación Hidrográfica del Sur	HID	105	1999–2013	15	5 min	0.1 mm
Confederación Hidrográfica del Júcar	JUC	182	2000–2016	17	5 min	0.2 mm
Confederación Hidrográfica del Miño-Sil	MIN	89	2007–2016	10	5 min	0.1 mm
Gobierno de Navarra	NAV	54	1991–2013	23	10/30 min	0.1 mm
Confederación Hidrográfica del Segura	SEG	66	1996–2013	18	15 min	0.2 mm
Confederación Hidrográfica del Tajo	TAJ	187	2002–2014	13	15 min	0.2 mm

## 2. Description of the dataset

90 We collected a dataset comprising 1587 high-frequency precipitation records from 12 different data providers, including regional weather services and water basin authorities. The data varied in terms of their temporal length, sampling frequency  $\delta T$  and precision  $\delta P$  (Table 2).

The mean temporal span of the dataset was around 10 years, although some 95 sets were shorter (CAN, DUE, GUA, MIN) while others had more than 15 years of data (EBR, GAL, HID, JUC, NAV, SEG, TAJ). The sampling frequency of the data  $\delta T$  varied between 5 and 30 minutes, with most records having a frequency of 15 minutes. The sets also varied in their precision  $\delta P$ , ranging between 0.1 and 0.2 mm.

100 The dataset had an uneven spatial distribution, with some areas more densely represented than others (Figure 1). Most notably, no data could be collected for the Guadiana River basin.



Figure 1: Spatial distribution of the rain gauges used on this study, with indication of the data provider (set) and the number of stations (N). (Coordinate system ETRS89/UTM30N, EPSG:25830)

### 3. Computation of event erosivity, $EI30$

The USLE R-factor is the long-term mean annual sum of individual storm erosivities. The erosivity of a given rainfall event,  $EI30$  (MJ mm ha<sup>-1</sup> h<sup>-1</sup>) is calculated as the product of the storm total kinetic energy  $E$  (MJ ha<sup>-1</sup>) and the maximum rain intensity recorded within 30 consecutive minutes,  $I30$  (mm h<sup>-1</sup>):

$$EI30 = E \cdot I30. \quad (1)$$

$I30$  can be obtained directly from the precipitation record, while the storm energy  $E$  is:

$$E = \int_0^D e v dt, \quad (2)$$

where  $e$  is the unit rainfall energy (MJ ha<sup>-1</sup> mm<sup>-1</sup>),  $v$  is the rainfall volume (mm),  $D$  is the storm duration and  $t$  is time.

Empirical equations have been proposed for estimating the rainfall energy  $e$ , as reviewed by Angulo-Martínez et al. (2016). Most of these equations, including the original formulation by Wischmeier (1959), were based on data from specific locations or regions and hence have a local or regional validity, while others have been proposed for global use (Van Dijk et al., 2002). Here we will follow the relation developed for the USA by Brown and Foster (1987), which is also recommended in the RUSLE manual due to its superior functional form (Renard et al., 1997), and has been used in a large number of studies worldwide:

$$e = 0.29 [1 - 0.72 \exp(-0.05i)] \quad (3)$$

with  $i$ , the rainfall intensity, having units of mm h<sup>-1</sup>.

#### *Time resolution of the data*

Originally,  $EI30$  was calculated for individual storms based on breakpoint precipitation data from continuous recording rain gauge (i.e. hyetograph) charts, which are segmented into periods of constant rainfall intensity. However, this



process can not be performed without intensive human intervention and is very time consuming, and recording-raingage records are also very often not available or are hard to obtain. When precipitation data are available in the form of rainfall amounts recorded at fixed time intervals  $\delta T$ , equations 1 and 2 can  
130 be expressed in a discrete form:

$$EI30(\delta T) = E(\delta T) \cdot I30(\delta T) = \left( \sum_{j=1}^n e_j v_j \right) I30(\delta T), \quad (4)$$

where  $e_j$  and  $v_j$  denote the unit rainfall energy and rainfall volume recorded at each discrete time interval during the storm.  $e_j$  is computed from equation (3), after replacing  $i$  by  $i_j$ , i.e. the rainfall intensity recorded during time interval  $j$ . The term in parenthesis,  $\delta T$ , indicates that the values of  $EI30$ ,  $E$  and  $I30$   
135 are derived from time interval instead of breakpoint precipitation data. As compared with using breakpoint precipitation data, this implementation is also best suited to automated treatment by a computer.

Denoting in a non-ambiguous way the type of data (breakpoint or time interval data) that has been used and the sampling frequency  $\delta T$  of the data  
140 is a good practice, since it has a non negligible effect in the computed rainfall erosivity. It has been shown that rainfall intensity  $i$ , and hence  $e$ ,  $E$  and  $I30$  tend to decrease as  $\delta T$  increases. Practical consequences of this decision are discussed later in the article. In our case, all datasets were aggregated to the same  $\delta T = 30$  minutes basis, which was the largest common multiple in the  
145 dataset. Since it is clear that we only used interval data, the term ( $\delta T$ ) is omitted in the remainder of the article for the sake of simplicity.

#### *Definition of erosive events*

Very clear rules were set in the original USLE formulation (Wischmeier and Smith, 1978) for delimiting the start and end of an erosive rainfall event. A  
150 period of six hours or more with less than 1.27 mm (0.05 in) of rain was used to separate between two consecutive storms. Not all the events were used in computing  $EI30$ , since rainfall events with a total rain depth of 12.7 (0.5 in) or

less were discarded, unless at least 6.35 mm (0.25 in) fell in 15 minutes. These thresholds were set for convenience, since originally the calculations were done using breakpoint data and were very time consuming. It was determined that these events were very small and had a negligible effect on the annual  $EI30$ , so they could be safely ignored (Renard et al., 1997). However, there are no physical or mathematical reasons for neglecting any event, and with current automated processes the calculation time has greatly reduced. Thus, in many studies other (lower) thresholds have been proposed. Verstraeten et al. (2006), following (Laurant and Bollinne, 1976), used a threshold as low as 1.27 mm for Belgium, and Renard et al. (1997) used all the events for producing their erosivity map of the western US states.

It should be noted, however, that using a different threshold (or no threshold at all) can possibly lead to significant bias in the computation of  $EI30$ , which would make the obtained values not compatible with those obtained using the original USLE formulation. In particular, the soil erodibility factor in the USLE should be re-calibrated, since this factor does not have a physical basis and was calibrated empirically once the other factors were fixed Nearing et al. (2017). Therefore, any deviations in the computation of  $EI30$  with respect to the original method used in calibrating the USLE should be checked for unbiasedness.

For this reason, we assessed the bias in the calculation of  $EI30$  according to equations 4 and 3 and using all the events irrespective of their size, with respect to the original USLE approach. When no threshold was used and all the events were considered, we obtained an over-estimation of  $EI30$  of 3.67% on average (median value), with values higher than 10% in 131 stations (Figure 2). As shown by these results, there is a significant bias if all the events are used for computing  $EI30$ , so we followed the original USLE definition of erosive events.

#### 4. Computation of daily erosivity, $EI30_{day}$

Since the objective of this study was to develop a method for estimating rainfall erosivity based on daily rainfall amounts, we had to devise an approach

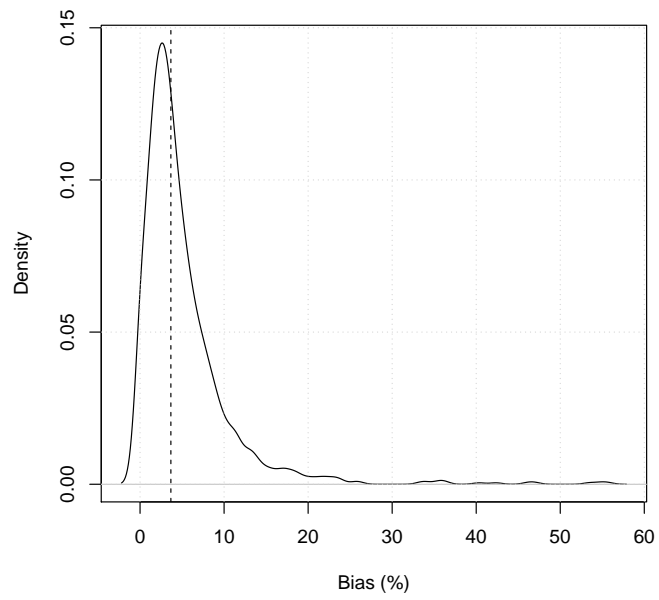


Figure 2: Kernel density over the whole station dataset of the percent bias in daily  $EI_{30}$  calculated with no rejection of events, with respect to the original USLE definition of erosive events.

that, using a daily time reference, resulted in  $EI30$  values that were unbiased with respect to those obtained using the original USLE method.

We followed a two-step approach. We started by considering each day with precipitation ( $P_{day} > 0$ ) as an potentially erosive event, and time series of daily  $EI30$  (hereinafter referred to as  $EI30_{day}$ ) were computed according to the following procedure:

1. a day was considered as potentially erosive if  $P_{day} > 0$  mm;
2. then,  $EI30_i$  were computed for that day using all high-frequency precipitation records  $P_i$ ,  $i = (1, 2, \dots, n)$  of that day, using equations 3 and 4;
3.  $EI30_{day} = \sum_{i=1}^n EI30_i$  values were obtained by aggregating over all the high-frequency records.

Once the daily time series of  $EI30_{day}$  were computed, including even the smallest daily amounts, an analysis was done in order to determine a threshold value  $\theta$  to reject very small daily events. We computed  $EI30$  values in all the stations rejecting all the daily events with precipitation amounts lower than ( $\theta = 1, 2, \dots 6$ ) mm, and we calculated the median bias with respect to the original, event based, USLE formulation.

The result is illustrated in Figure 3. When no threshold was used and all the daily events were considered, the mean bias was 3.17%, i.e. very similar to the mean bias of 3.65% obtained at the event scale when all the events were included. The bias was positive (i.e., indicating over-estimation) for values of the threshold up to approximately  $\theta = 4.5$  mm day<sup>-1</sup>, and then it became negative (i.e., indicating under-estimation). Therefore, a threshold value of  $\theta = 4.5$  mm day<sup>-1</sup> was established to reject small events.

Since the daily approach outlined here differs from the original, event-based, formulation of the index, annual aggregated values were compared between the two approaches. The median bias was negligible at 0.01%.

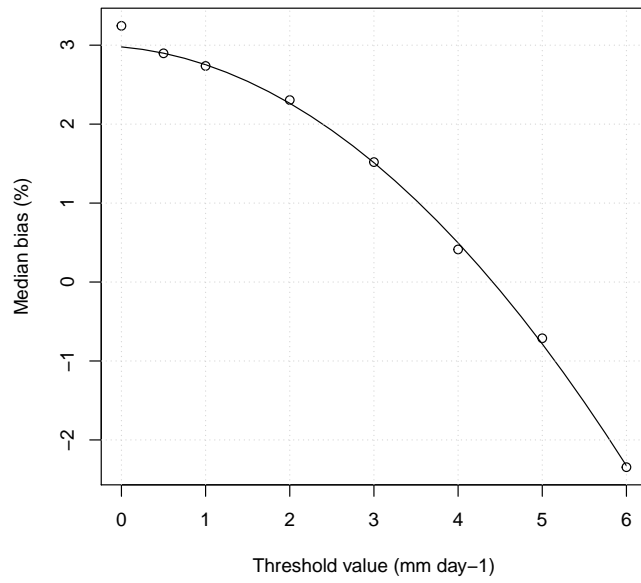


Figure 3: Median percent bias of daily  $EI30$  over the set of stations as a function of the threshold value used to reject very small events. The dots represent the values estimated from the data, and the line is a second-order polynomial model fit to the dots.

## 5. Modelling the relationship between $EI30_{day}$ and $P_{day}$

210 Early works by Athesian (1974) and Cooley et al. (1980) showed that event based  $EI30$  values could be conveniently related to event  $P$  by means of a power law:

$$EI30_{event} = \frac{\alpha P_{event}^{\beta}}{D_{event}^{\gamma}},$$

where  $D$  is the event duration and  $\alpha$ ,  $\beta$  and  $\gamma$  are coefficients. Since in our case the event duration is constant (one day), the former expression is reduced to:

$$EI30_{day} = \alpha P_{day}^{\beta}, \quad (5)$$

215 In order to estimate  $\alpha$  and  $\beta$ , very often this model is transformed to a linear model by taking logarithms, leading to the well known log-log linear model:

$$\ln y_i = \beta_0 + \beta_1 \ln x_i + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}(0, \sigma^2), \quad (6)$$

where  $\beta_0 = \ln \alpha$  is the intercept, and  $\beta_1 = \beta$  is the slope. This is a standard linear model that can be fit by ordinary least squares (OLS), and this approach has been used by some authors for estimating rainfall erosivity from daily precipitation records.  
220

An important assumption of the standard linear model is that the error term  $\epsilon$  is unbiased (it has zero expectancy), symmetric (since it is normally distributed), and homoskedastic (it has constant variance), so it is often neglected. Being an additive error term, it is trivial to derive that the expectancy of  $y_i$  is:

$$E(\ln y_i) = \mu_i = \beta_0 + \beta_1 \ln x_i$$

225 with variance:

$$\text{Var}(\ln y_i) = \sigma^2.$$

However, the interest here is in  $y_i$ , and not  $\ln y_i$ . When equation (6) is used to obtain estimates of  $y_i$  by replacing  $\alpha$  and  $\beta$  by OLS estimates of  $\beta_0$  and  $\beta_1$ , a

systematic underestimation occurs, as it has been pointed out by several authors (Richardson et al., 1983; Elsenbeer et al., 1993; Posch and Rekolainen, 1993).  
 230 The reason is that the logarithmic transformation affects also the structure of the error term, as it can be seen by back-transforming equation (6):

$$y_i = e^{\beta_0} x_i^{\beta_1} \epsilon'_i, \quad (7)$$

to see that the error term  $\epsilon'_i = e^{\epsilon_i}$  is now multiplicative, and thus biased ( $E(\epsilon'_i) > 1$ ) and not symmetric. The variance of  $y_i$  now becomes:

$$\text{Var}(y_i) = \mu_i^2 \sigma^2,$$

which is not constant but increases geometrically with  $\mu_i$ .

235 In order to solve the bias in the estimates of  $y_i$ , Richardson et al. (1983) proposed adding a random term  $\nu_i \sim N(0, \sigma_\nu^2)$  to the estimates of  $\ln y_i$  obtained using equation (6). Similarly, other bias-correction methods have been proposed based on the mean square error of the regression (e.g. Newman, 1993). Taking a different approach, Angulo-Martínez and Beguería (2009) proposed using  
 240 weighted OLS regression to counter-balance the excessive leverage of low observations in OLS arising from the incorrect modeling of the error term. However, the selection of the weighting scheme is not straightforward and there is no method assuring an optimum choice.

Summarizing, none of these options solved the problem of underestimating  
 245 *EI30* based on  $P$  data in a satisfactory way. Here we propose using the Generalised Linear Model (GLM) to fit the model in equation (5) with an adequate treatment of its error term. GLM is a generalisation of OLS regression which overcomes some of its assumptions, allowing for variables with error distribution differing from a Normal distribution (McCullagh and Nelder, 1989). A Gamma  
 250 GLM appears as a naturally suited model for the relationship between *EI30* and  $P$ . Under this model,  $y_i$  is treated as a Gamma distributed variate, which is a suitable choice for heteroskedastic positive data, as it is the case of *EI30*. A logarithmic link function was used in the formulation of the Gamma GLM in

Table 3: Goodness of fit and bias statistics of three alternative models of  $EI30_{day}$  over  $P_{day}$  at the ‘Aralar\_GN’ station (NAV dataset): log-log OLS ( $\mathcal{M}0$ ), weighted log-log OLS ( $\mathcal{M}1$ ) and Gamma GLM with a log link ( $\mathcal{M}2$ ). Degrees of freedom (df), Akaike’s Information Criterion (AIC), determination coefficient ( $R^2$ ), mean absolute error (MAE), mean error (ME), percent bias (PBIAS) and total relative error (TRE).

Model	df	AIC	R2	MAE	ME	PBIAS	TRE
$\mathcal{M}0$	3	3872.1	0.46	5.14	2.77	54.1%	0.65
$\mathcal{M}1$	3	6327.3	0.45	5.41	1.82	29.9%	0.77
$\mathcal{M}2$	3	-1267.0	0.47	6.10	0.10	1.3%	0.99

order to linearize the relationship between  $EI30_{day}$  and  $P_{day}$ .

255 To illustrate this, a comparison was made between log-log OLS ( $\mathcal{M}0$ ), log-weighted OLS ( $\mathcal{M}1$ ), and Gamma GLM with a log link ( $\mathcal{M}2$ ). Results are shown for one of the stations in Table (3) and Figure (4).

Despite the similar  $R^2$  (which, anyway, is not an adequate statistic for variables with non-Normal distributed errors), the Akaike Information Criterion (AIC) clearly favoured the GLM model as providing the best fit to the data. 260 While GLM provided virtually unbiased  $EI30_{day}$  estimates (ME = 0.10 and PBIAS = 1.3 %), the two models based on OLS had important biases (ME = 2.77 and 1.82, and PBIAS = 54.1 and 29.9 %, respectively). When the sum of all the daily estimates of  $EI30$  was compared with the observed ones, important 265 underestimation was found for OLS and weighted OLS (Total Relative Error, TRE = 0.65 and 0.77, respectively), while GLM yielded results very close to the observed ones (TRE = 0.99).

## 6. Seasonal variation of the relationship between $EI30_{day}$ and $P_{day}$

In most climates, precipitation characteristics vary significantly between seasons. It is reasonable to expect that the relationship between  $EI30_{day}$  and  $P_{day}$  270 will also vary. There are different ways to incorporate seasonality into the model in equation (5), such as allowing the parameter  $\alpha$  vary from month to month while keeping  $\beta$  constant (as seen, e.g., in Richardson et al., 1983; Bagarello and



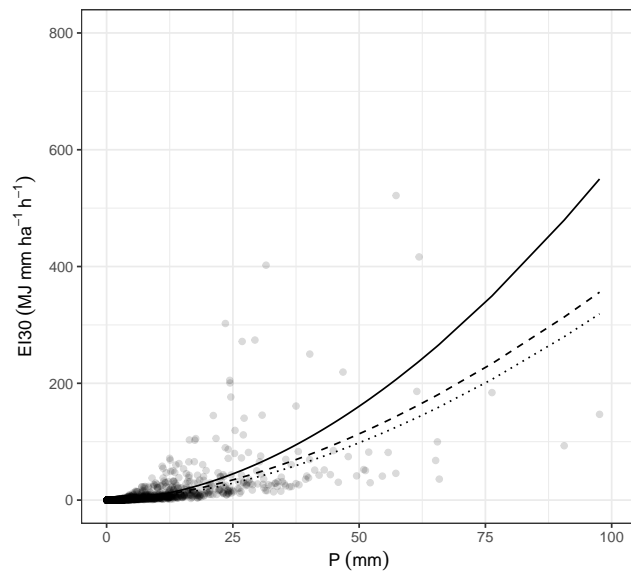


Figure 4: Relationship between  $EI30_{day}$  and  $P_{day}$  at the 'Aralar\_GN' station (NAV dataset): log-log OLS ( $\mathcal{M}0$ , dotted line), weighted log-log OLS ( $\mathcal{M}1$ , dashed line) and Gamma GLM with a log link ( $\mathcal{M}2$ , solid line).

D’Asaro, 1994; Petkovšek and Mikoš, 2004), varying  $\beta$  while keeping  $\alpha$  constant,  
 275 or allowing both parameters to vary (Verstraeten et al., 2006). These leads to  
 the following possible configurations:

$$EI30_{day} = \alpha P_{day}^{\beta} \quad (\mathcal{M}2.0)$$

$$EI30_{day} = \alpha_t P_{day}^{\beta} \quad (\mathcal{M}2.1)$$

$$EI30_{day} = \alpha P_{day}^{\beta_t} \quad (\mathcal{M}2.2)$$

$$EI30_{day} = \alpha_t P_{day}^{\beta_t} \quad (\mathcal{M}2.3)$$

where suffix  $t$  means that the parameter takes different values for each month  
 of the year.

A graphical illustration of these model configurations is shown on Figure (5)  
 280 for one random station. The selection of a time-varying model is not trivial.  
 Due to the high covariance between parameters  $\alpha$  and  $\beta$ , the full model  $\mathcal{M}2.3$   
 may not be the optimum, since the uncertainty of the estimated model param-  
 eters can be high, as the large error bars in Figure (5) suggest. However, the  
 comparison between models  $\mathcal{M}2.1$  and  $\mathcal{M}2.2$  seems more difficult.

285 In order to determine which model configuration was most suited to our  
 dataset, we conducted a comparison between the four models over the whole  
 station dataset. Goodness of fit and bias statistics are provided on Table 4.

The AIC statistic clearly favoured the most complex models over the base-  
 line stationary  $\mathcal{M}2.0$ , with only marginal differences between them (although  
 290  $\mathcal{M}2.2$  attained the overall lowest AIC). The best fit (highest R2 and lowest  
 MAE) corresponded to  $\mathcal{M}2.3$ , which is nothing strange since this model had  
 the highest number of parameters (as shown by the df), although the other  
 two non-stationary models came in close. The lowest bias (ME and PBIAS)  
 corresponded, however, to  $\mathcal{M}2.1$ , while the two models with the parameter  $\beta$   
 295 varying by month, that is  $\mathcal{M}2.2$  and  $\mathcal{M}2.3$ , had slightly high median bias (of  
 -5.4 and -4%). In addition to goodness-of-fit and bias statistics, we also looked  
 at the significance level of the model parameters, and we found that model  
 $\mathcal{M}2.3$ , in a majority of the stations, resulted in non-significant  $\beta$  parameters in

Table 4: Goodness of fit and bias statistics of four alternative models of  $EI30_{day}$  over  $P_{day}$  (median values over the set of stations): Gamma GLM with a log link and stationary parameters ( $\mathcal{M}2.0$ ), monthly variation of  $\alpha$  ( $\mathcal{M}2.1$ ), monthly variation of  $\beta$  ( $\mathcal{M}2.2$ ), and variation of both ( $\mathcal{M}2.3$ ). Statistic acronyms as in Table 3.

Model	df	AIC	R2	MAE	ME	PBIAS	TRE
$\mathcal{M}2.0$	3	1772.2	0.49	17.03	-0.74	-3.9%	1.04
$\mathcal{M}2.1$	14	1686.2	0.64	13.37	-0.11	-0.7%	1.01
$\mathcal{M}2.2$	14	1680.9	0.64	13.38	-1.12	-5.4%	1.06
$\mathcal{M}2.3$	25	1689.4	0.70	12.41	-0.89	-4.0%	1.04

most months, and even  $\alpha$  had problems for achieving significance in the summer  
 300 months.

In light of these results, we considered that  $\mathcal{M}2.1$  was the optimum model configuration for our dataset. Kernel density plots of goodness of fit and bias statistics over the set of stations is presented for this model in Figure 6. The estimated  $EI30_{day}$  values fitted well the observed ones, as shown by the scatterplot, and the median coefficient of determination ( $R^2$ ) of 0.641 is satisfactory  
 305 for the daily time resolution. The median MAE was 13.374, revealing a small deviation from the observed values. The kernel density of the observed and estimated values revealed a very similar distribution of the values, and the distribution of the ME is symmetric with respect to zero, indicating the absence of  
 310 bias. As a result, the median ratio between the modelled and observed summed  $EI30$  was also very close to one and symmetric, indicating no bias also in the cumulative values.

## 7. Spatial variation of the relationship between $EI30_{day}$ and $P_{day}$

Exploratory maps depicting the spatial distribution of  $\alpha$  and  $\beta$  parameters  
 315 resulting from applying model  $\mathcal{M}2.1$  over the set of stations suggest that their distribution is not homogeneous across the study area (Figures 7 and 8). Instead, smooth spatial gradients are apparent for both parameters, in addition to very noticeable monthly differences.

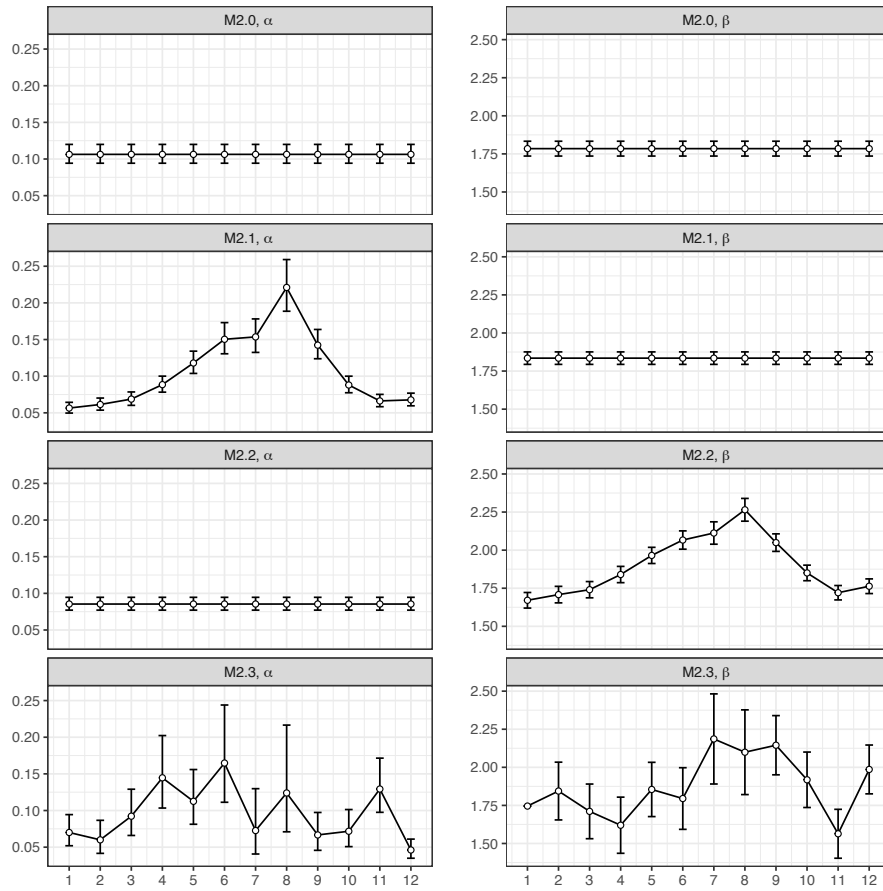


Figure 5: Different temporal models of  $\alpha$  and  $\beta$  at the ‘Aralar.GN’ station (NAV dataset): stationary ( $\mathcal{M}2.0$ ), monthly variation of  $\alpha$  parameter ( $\mathcal{M}2.1$ ), monthly variation of  $\beta$  parameter ( $\mathcal{M}2.2$ ), and monthly variation of both parameters ( $\mathcal{M}2.3$ ). The vertical error bars represent the standard error of the estimated parameters.

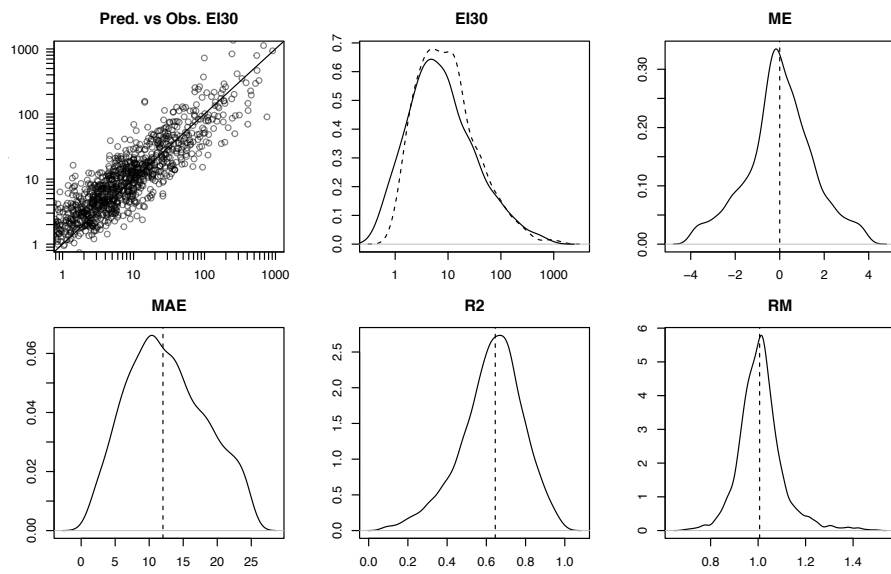


Figure 6: Goodness of fit and bias plots for  $EI30$  estimated using  $\alpha$  and  $\beta$  parameters fitted at each station by model  $\mathcal{M}2.1$ . Predicted and observed scatterplot and kernel density (dashed line: observed; solid line: modeled), and densities of several goodness of fit statistics: Mean Error, Mean Absolute Error,  $R^2$  and ratio of totals. The vertical lines indicates the median value of each statistic.

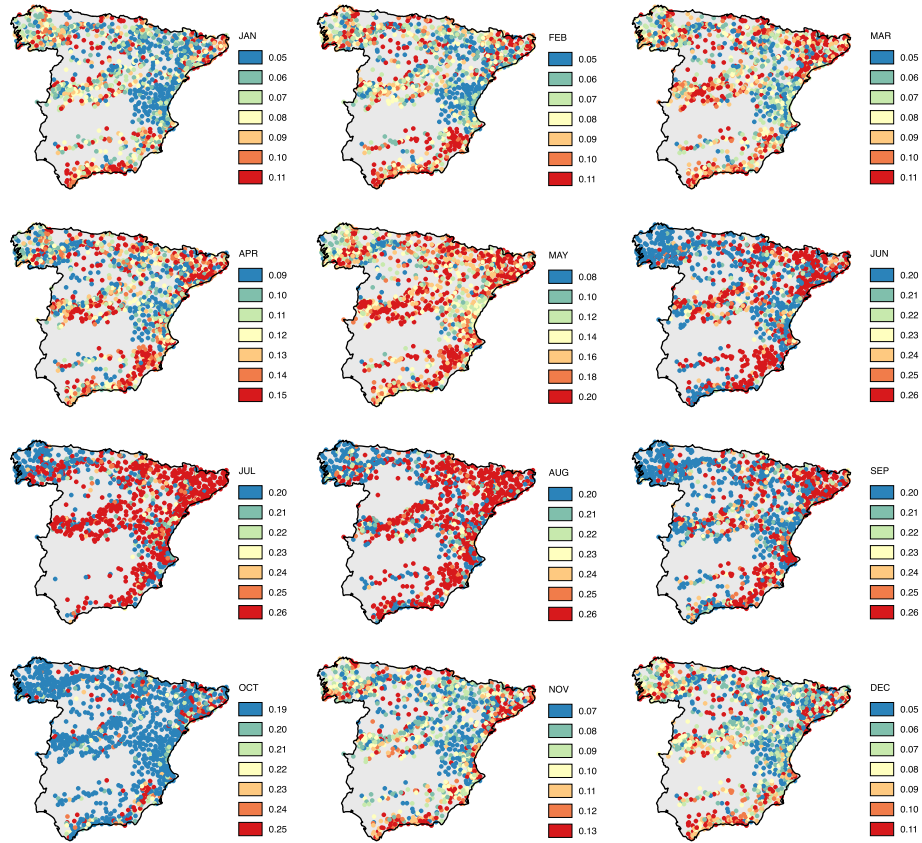


Figure 7: Monthly values of  $\alpha$  parameter over the set of stations, estimated by Gamma GLS regression with a log link and model configuration  $\mathcal{M}2.1$ .

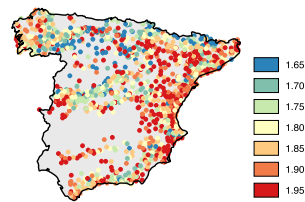


Figure 8: Monthly values  $\beta$  parameter over the set of stations, estimated by Gamma GLS regression with a log link and model configuration  $\mathcal{M}2.1$ .

In order to obtain continuous fields of both parameters over the study area,  
320 Kriging arises as a convenient approach. Kriging methods provide best linear  
unbiased predictions of the measured values (Cressie, 1993; Goovaerts, 1997).  
Kriging, or Gaussian Process Regression, is a method for spatial interpolation  
based on linear-weighted averaging of the values measured at given sampling  
points. The weights are based on a prior covariance model, which is based on a  
325 function of the distance between point pairs by the so-called semi-variogram.

Empirical and fitted semi-variograms for  $\alpha$  and  $\beta$  parameters are given in  
Figures 9 and 10. The spherical model was selected as it presented the best fit  
among a set of candidate models that included Gaussian, exponential and Bessel.  
Stronger and longer range spatial autocorrelation was found for  $\alpha$  during winter  
330 and spring, while it was weaker in summer. The range of the semi-variogram,  
which indicates the spatial extent of the spatial correlation, varied between 100  
km (in August) and 200 km (January). All the semi-variograms had a strong  
nugget effect, which can be interpreted as a sign of a high random variability at  
close locations.

335 While ordinary and simple Kriging approaches use only the information of  
the variable or variables that need be interpolated, universal Kriging facilitates  
incorporation of data on independent variables or predictors (Pebesma, 2006).  
Exploratory analysis revealed significant correlations between  $\alpha$  and  $\beta$  param-  
eters and several geographic and climatic variables (Table 5). Maps showing the  
340 spatial distribution of the covariables used are provided in Figure 11. Those in-  
cluded 5x5 km gridded elevation (spatially aggregated from SRTM 90m Digital  
Elevation Data), distance to the closed sea squared, annual precipitation and  
the average fraction of the annual precipitation falling on daily events above  
the 95th percentile. The last two variables were obtained from Serrano-Notivoli  
345 et al. (2017). Universal kriging was therefore the method selected for modelling  
the spatial variation of  $\alpha$  and  $\beta$ , using the above-mentioned environmental vari-  
ables as independent variables.

The resulting fields of  $\alpha$  and  $\beta$  parameters are shown in Figures 12 and  
13. The temporal variation was higher than the spatial variation for both param-

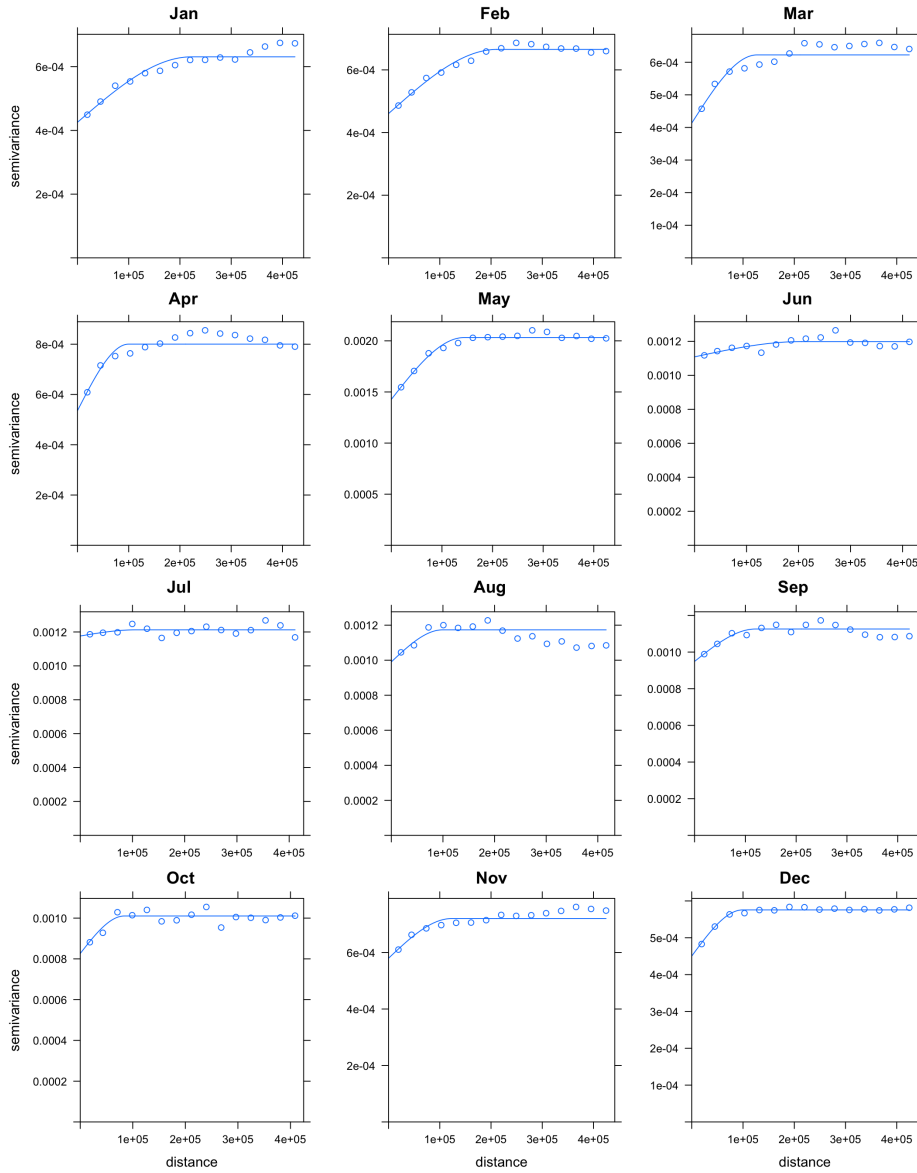


Figure 9: Empirical and fitted semivariogram of  $\alpha$  monthly parameters (distance units: m).



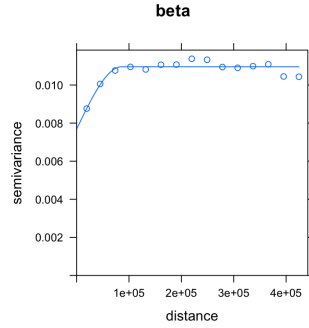


Figure 10: Empirical and fitted semivariogram of  $\beta$  parameter (distance units: m).

Table 5: Pearson's correlations between  $\alpha$  and  $\beta$  parameters and environmental variables: altitude above sea level (alt), distance to the sea squared (dis2), mean annual precipitation (panual) and fraction of the annual precipitation accumulated on daily events above the 95 percentile (r95).

Parameter	alt	dis2	panual	r95
$\alpha$ (Jan)	-0.077*	-0.130*	0.019	-0.047
$\alpha$ (Feb)	-0.077*	-0.104*	0.005	-0.003
$\alpha$ (Mar)	-0.090*	-0.045*	-0.002	0.047*
$\alpha$ (Apr)	-0.167*	-0.082*	-0.122*	0.125*
$\alpha$ (May)	-0.105*	0.116*	-0.177*	0.107*
$\alpha$ (Jun)	0.050	0.169*	-0.239*	0.179*
$\alpha$ (Jul)	0.014	0.251*	-0.377*	0.115*
$\alpha$ (Aug)	-0.012	0.109*	-0.301*	0.156*
$\alpha$ (Sep)	-0.065*	-0.010	-0.223*	0.152*
$\alpha$ (Oct)	-0.233*	-0.171*	-0.174*	0.223*
$\alpha$ (Nov)	-0.244*	-0.280*	-0.021*	0.149*
$\alpha$ (Dec)	-0.137*	-0.149*	-0.006	0.008
$\beta$	-0.246*	-0.129*	-0.199*	0.115*

\*  $p < 0.05$ .

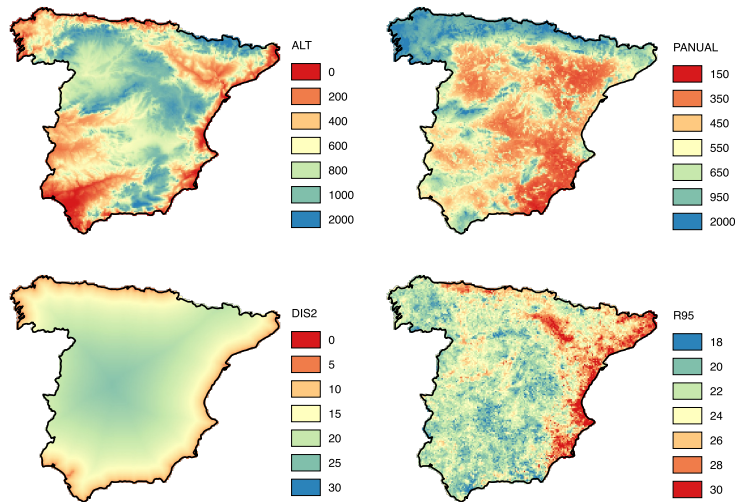


Figure 11: Covariables: elevation (alt), distance to the sea squared (dis2), annual precipitation (panual), and fraction of the annual precipitation on daily events above the 95th percentile (r95).

350 eters. The spatial pattern was complex and varied between months, although a coherence was apparent between correlative months. The associated marginal errors also had a stronger temporal variation, being higher in summer (between 15% relative error for  $\alpha$  in July, vs 5% in January; and between 5% and 3% for  $\beta$ ).

355 Ten-fold cross-validation statistics for kriged  $\alpha$  and  $\beta$  parameters are shown in Table 6.  $R^2$  values were quite high, above or near 0.95 for all months, and the MAE was reasonably low compared to the mean values of the parameters. The bias was almost zero, as shown by low values of ME and PBIAS.

## 8. Method summary and validation

360 A global view of the procedure described in the previous sections is outlined on Figure 14. Steps 1 through 4 represent the analysis part, which includes aggregating the original high-resolution precipitation series to a common  $\delta T$  of 30 minutes (1); computation of daily rainfall and erosivity events (2); applying

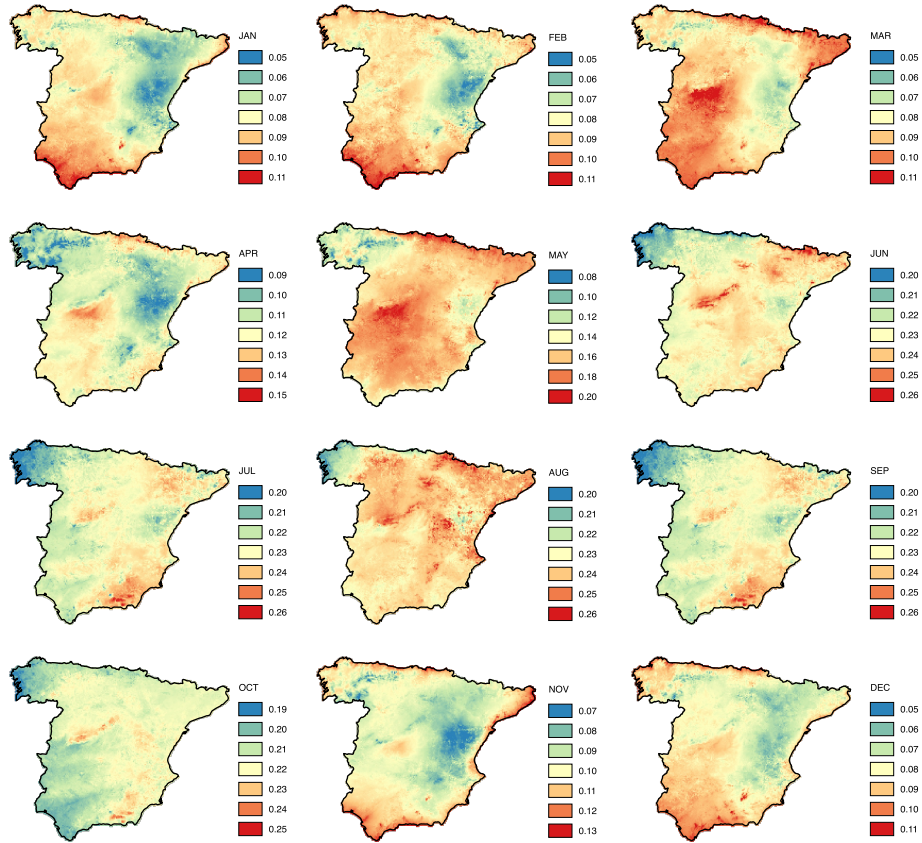


Figure 12: Kriged monthly values of  $\alpha$ .

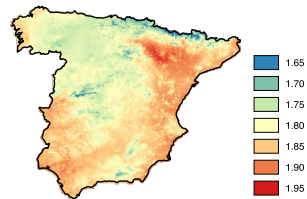


Figure 13: Kriged values of  $\beta$ .

Table 6: Cross-validation statistics for parameters  $\alpha$  and  $\beta$  over the complete station dataset: model M2.1 + universal kriging (UK).

Parameter	R2	MAE	ME	PBIAS %
$\alpha$ (Jan)	0.950	0.018	0.000	0.3
$\alpha$ (Feb)	0.960	0.019	0.001	0.8
$\alpha$ (Mar)	0.958	0.018	0.000	-0.3
$\alpha$ (Apr)	0.970	0.022	0.000	-0.1
$\alpha$ (May)	0.960	0.034	0.000	0.3
$\alpha$ (Jun)	0.987	0.030	0.001	0.3
$\alpha$ (Jul)	0.987	0.030	0.000	0.2
$\alpha$ (Aug)	0.980	0.028	0.000	0.2
$\alpha$ (Sep)	0.981	0.028	0.000	0.2
$\alpha$ (Oct)	0.989	0.027	-0.001	-0.3
$\alpha$ (Nov)	0.960	0.021	0.000	-0.2
$\alpha$ (Dec)	0.961	0.018	0.000	0.0
$\beta$	0.955	0.079	0.000	0.0

Gamma GLS regression to obtain  $\alpha$  and  $\beta$  parameters at each measuring site  
 365 (3); and using universal kriging to obtain a continuous surface model of both  
 parameters (4). From this last product, it is possible to compute daily time  
 series and, through time aggregation, monthly and annual  $EI30$  series (5 and  
 6). This can be done based on site (raingauge) data, or else using a grid of daily  
 precipitation.

370 For instance, we computed  $EI30_{day}$  time series at the locations in the dataset  
 based on observed  $P_{day}$  and  $\alpha$  and  $\beta$  kriged values from cross-validation. As a  
 means of validation, a comparison was made between observed and estimated  
 daily  $EI30_{day}$  (Figure 15), as well as for their annual aggregates (Figure 16).  
 The estimated values showed no signs of bias, and had a density distribution  
 375 which resembled quite well the observed one. The median coefficient of determi-  
 nation ( $R^2$ ) increased from 0.629 at the daily scale to 0.795 at the annual scale,  
 supporting the ability of the method for computing aggregated erosivity val-

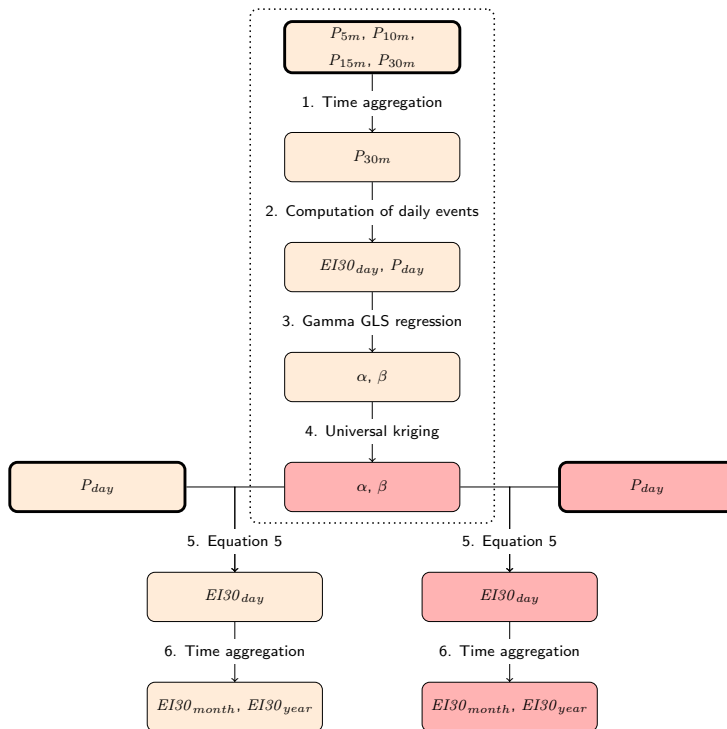


Figure 14: Flowchart of the method. Thick bordered boxes represent original (input) data. Orange color represents point data, while red represents spatially distributed (gridded) data. The analysis phase described on this article encompasses steps 1 to 4 (dotted rectangle), while steps 5 and 6 allow computation of monthly and annual erosivity values based on daily raingauge or gridded precipitation, plus the  $\alpha$  and  $\beta$  parameters from step 4.

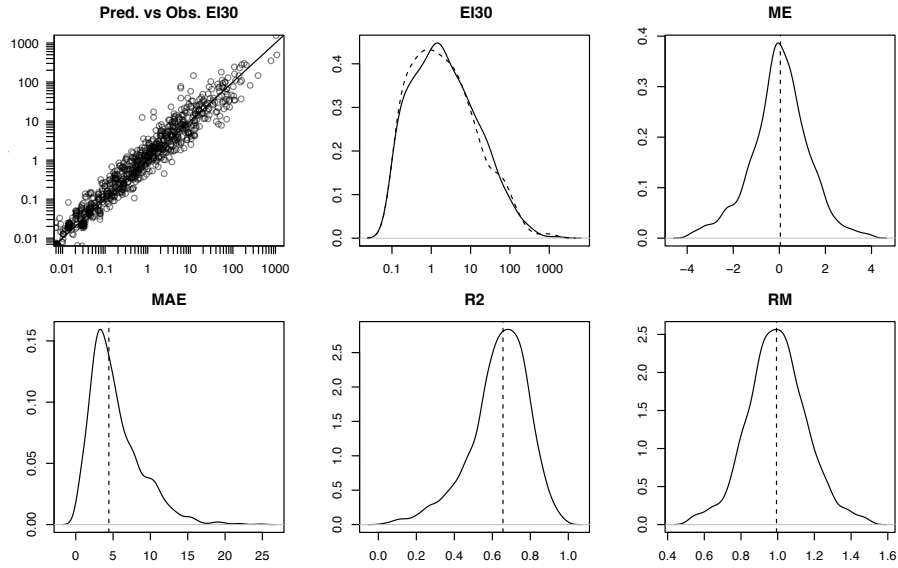


Figure 15: Goodness of fit and bias plots for daily *EI30* estimated using kriged  $\alpha$  and  $\beta$  parameters. Predicted and observed scatterplot and kernel density (dashed line: observed; solid line: modeled), and densities of several goodness of fit statistics (acronyms as in Figure 6).

ues. When annual averages were calculated considering all the complete years  
of record at each station, the  $R^2$  increased to 0.850, and this can be considered  
380 a conservative value since many station datasets had very few years of records  
which penalise the computation of a mean annual value. The percent bias was  
very close to zero (-0.5%).

## 9. Discussion

The results from the previous sections are indicative of a good ability of the  
385 proposed method for producing unbiased and accurate estimations of *EI30* from  
daily precipitation records. In the case of Spain, the gridded values of  $\alpha$  and  
 $\beta$  parameters can be used directly to produce estimates of rainfall erosivity at  
the daily level, and monthly or annual aggregates can be computed from those.  
The same methodology could be applied to other study areas for which high

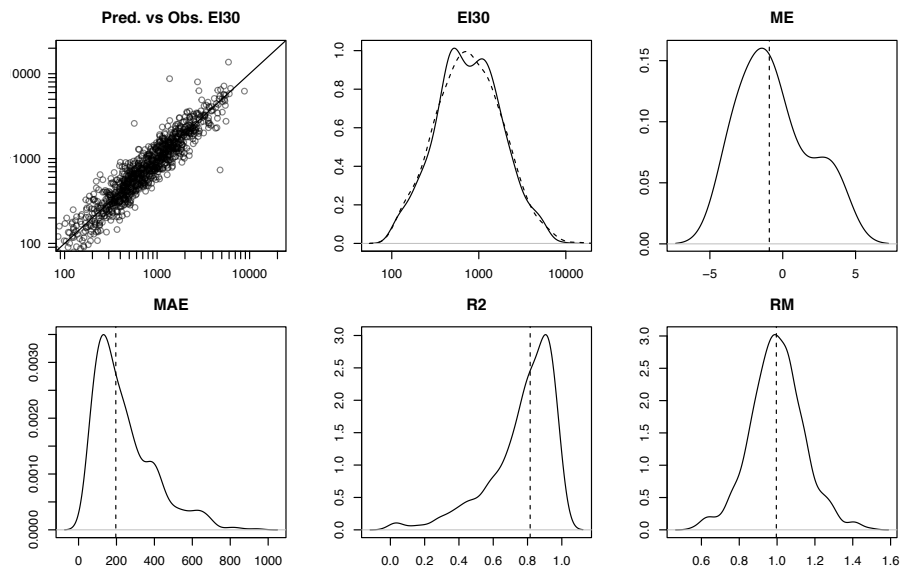


Figure 16: Goodness of fit plots for annual *EI30*, aggregated from daily values estimated using kriged  $\alpha$  and  $\beta$  parameters (only years with less than 5 missing days). Predicted and observed scatterplot and kernel density (dashed line: observed; solid line: modeled), and densities of several goodness of fit statistics (acronyms as in Figure 6).

390 frequency precipitation records exist. However, there are a number of issues  
that merit further discussion, since they can have an impact on the estimated  
erosivity values.

The computation of  $EI30$  requires estimating the unit rainfall energy,  $e$   
(equation 2). A number of empirical relationships relating  $e$  to the precipitation  
395 intensity  $i$  have been proposed, and the choice of one or another will affect the  
resulting  $EI30$ . Here we used the RUSLE relationship in equation 3 (Renard  
et al., 1997), which is one of the most commonly used ones. This expression,  
however, has been criticised and a new expression has been proposed in the  
RUSLE2 model (USDA-ARS, 2013) that provides a better fit to the original  
400 expression in Wischmeier (1959). In a recent study, however, a comparison  
was made between different relationships in their ability to estimate the rainfall  
energy measured by means of an optical disdrometer in the central Ebro Valley in  
NE Spain, and the RUSLE expression was among the best performers (Angulo-  
Martínez et al., 2016). Therefore, we decided to use the RUSLE expression,  
405 while future studies should determine the best approximation for the climatic  
conditions of Spain.

Another issue when time interval data are used for computing rainfall ero-  
sivity, as explained on section , is the existence of a bias related to the sampling  
frequency  $\delta T$ , that make erosivity values computed from data with different  
410 sampling frequencies not directly comparable. However, it has been shown that  
the relationship between  $EI30$  and  $\delta T$  follows nicely a linear relationship, so the  
following expression can be used to convert erosivity values obtained with data  
at different resolutions, based on the expression:

$$EI30_{\delta T} = \phi \cdot EI30_{\delta T'}, \quad (8)$$

with  $\phi$  being an empirical parameter that can be estimated, for instance, by  
415 OLS regression.

Values of  $\phi$  have been published to help converting between different  $\delta T$   
basis (Renard et al., 1997; Agnese et al., 2006; Yin et al., 2007; Panagos et al.,



2015b; Angulo-Martínez et al., 2016). One of the most comprehensive studies to date is that of Panagos et al. (2016b), who analysed a set of 1568 stations  
420 spanning over Europe.

On this work we had datasets with different sampling frequencies, but we decided to aggregate all data to a common sampling frequency of  $\delta T = 30$  minutes. This makes our results directly comparable with other rainfall erosivity estimations computed from 30-minutes precipitation records, as it is the case  
425 of the REDES map of rainfall erosivity over Europe (Panagos et al., 2015a). In order to compare with other datasets using higher or lower frequency data, however, linear correction factors need to be used. Generic published values such as those in previous references have to be used for that purpose, until specific values of  $\phi$  are published for Spain.

430 Another possible source of bias in the computation of  $EI30$ , and one that has been usually neglected, arises from the different numerical precision  $\delta P$  of the records. Different rain gauge types differ in their physical principles (e.g., weights vs. tipping buckets) and in their precision. For instance, in our dataset records coexisted with 0.1 and 0.2 mm precision. These differences  
435 are a potential source of bias, since data at coarser resolutions tend to have their values concentrated in fewer, but more intense, records, resulting in higher average erosivity. We checked for this effect by constructing new data series with 0.2 mm resolution from those that had an original resolution of 0.1 mm, and comparing the resulting daily  $EI30$  values. Equation 8 was applied to  
440 each station where data at both resolutions were available, and OLS regression was used for estimating the values of  $\theta$  required for translating between both resolutions. At the 15-minutes resolution, the results indicated a negligible effect ( $0.999 < \theta < 1$ ), so no correction was required for the different precisions.

## 10. Conclusions

445 The USLE/RUSLE  $EI30$  rainfall erosivity index is an important parameter in many erosion models, and a number of methods have been proposed for its

calculation. The original procedure (Wischmeier, 1959) based on breakpoint pluviometer bands is highly time-consuming, and breakpoint data are only seldom available over large regions. For that reason, a number of approaches  
450 have been developed for computing the *EI30* index from commonly available fixed-interval precipitation records. These have the problem that the obtained *EI30* values are not independent of the sampling interval,  $\delta T$ , at which the data were recorded. Linear scaling models, however, have been proposed in order to transform *EI30* values obtained with different  $\delta T$ , offering an opportunity for calculating *EI30* values from fixed-interval precipitation data that are  
455 compatible with those obtained from breakpoint data.

Current methods, however, require high-frequency precipitation data, i.e. typically in the range  $1 \text{ min} < \delta T < 15 \text{ min}$ . High-frequency precipitation data are seldom available with the desired spatial and temporal coverage for conducting regional studies, and in most cases the data sets have a short time span  
460 and cover only the most recent years since high-frequency records are usually related to the recent development of automatic pluviometers that were not used in operative networks until recently. The development of robust methods for computing *EI30* from medium-frequency precipitation data, such as daily time series, is therefore highly desirable, since it opens the possibility to use widely  
465 available daily precipitation databases which in most countries have a dense spatial density and a long temporal coverage. Being able to use daily precipitation data would allow the development of regional climatologies of monthly or annual *EI30* at very high spatial resolution. Most importantly, using several  
470 decades of daily rainfall data would allow computation of robust average *EI30* values, since large datasets spanning ideally over several decades is required for its calculation as rainfall erosivity shows a long-tailed statistical distribution and hence requires long time observations. Additionally, the ability to use time series spanning over several decades allows undertaking long-term trend analysis of rainfall erosivity, which is of the highest interest in a context of climate  
475 change.

In this study we present a robust methodology for computing monthly and

annual *EI30* from daily precipitation records. A set of high-resolution precipitation data is required for calibrating  $\alpha$  and  $\beta$  parameters that are used for  
480 computing daily *EI30* from daily precipitation, but once calibrated these parameters can be used with any series of daily precipitation available in the same study area. We used optimal regression tools for the problem under study, such Gamma GLS and universal co-kriging, and we discussed the consequences of the assumptions made in the process. Independent validation revealed that the  
485 method, applied to a case study covering most of continental Spain, produced very good estimates of monthly and annual *EI30* based on daily precipitation. This method could be used to produce similar regional models in other areas, provided that a sufficiently dense dataset of high-frequency, albeit not necessarily very long, precipitation data is available.

#### 490 **Acknowledgements**

This work has been supported by research projects CGL2014-52135-C3-1-R and CGL2017-83866-C3-3-R, financed by the Spanish Ministerio de Economía, Industria y Competitividad (MINECO) and EU-FEDER. The work of M. Tomas-Burguera was supported by a predoctoral grant under the FPU program 2013  
495 of the Spanish Ministerio de Educación, Cultura y Deporte. We want to express our gratitude to the organisms that provided the data: Confederación Hidrográfica del Cantábrico, Servei Meteorològic de Catalunya, Confederación Hidrográfica del Duero, Confederación Hidrográfica del Ebro, Meteo Galicia, Confederación Hidrográfica del Guadalquivir, Confederación Hidrográfica del  
500 Sur, Confederación Hidrográfica del Júcar, Confederación Hidrográfica del Miño-Sil, Gobierno de Navarra, Confederación Hidrográfica del Segura, Confederación Hidrográfica del Tajo. None of the authors have competing interests.

#### **References**

Agnese, C., Bagarello, V., Corrao, C., D'Agostino, L., D'Asaro, F., 2006. Influence of the rainfall measurement interval on the erosivity determinations in  
505

- the mediterranean area. *Journal of Hydrology* 329, 36–48.
- Angulo-Martínez, M., Beguería, S., 2009. Estimating rainfall erosivity from daily precipitation records: A comparison among methods using data from the ebro basin (ne spain). *Journal of Hydrology* 379, 111–121.
- 510 Angulo-Martínez, M., Beguería, S., 2012. Trends in rainfall erosivity in ne spain at annual, seasonal and daily scales, 1955-2006. *Hydrology and Earth System Sciences* .
- Angulo-Martínez, M., Beguería, S., Kyselý, J., 2016. Use of disdrometer data to evaluate the relationship of rainfall kinetic energy and intensity (ke-i). *Science*  
515 *of the Total Environment* 568, 83–94.
- Angulo-Martínez, M., Vicente-Serrano, S., Beguería, S., 2009. Trends in rainfall erosivity (1955-2006) over the ebro basin (ne spain), in: *EGU General Assembly Conference Abstracts*, p. 3183.
- Apaydin, H., Erpul, G., Bayramin, I., Gabriels, D., 2006. Evaluation of indices  
520 for characterizing the distribution and concentration of precipitation: A case for the region of southeastern anatolia project, turkey. *Journal of Hydrology* 328, 726–732.
- Athesian, J., 1974. Estimation of rainfall erosion index. *Journal of the Irrigation and Drainage Division, American Society of Civil Engineers* 100, 293–307.
- 525 Bagarello, V., D’Asaro, F., 1994. Estimating single storm erosion index. *Transactions of the ASAE* 37, 785–791.
- Ballabio, C., Borrelli, P., Spinoni, J., Meusburger, K., Michaelides, S., Beguería, S., Klik, A., Petan, S., Janeček, M., Olsen, P., et al., 2017. Mapping monthly rainfall erosivity in europe. *Science of the Total Environment* 579, 1298–1315.
- 530 Banasik, K., Górski, D., 1994. *Lectures in soil erosion control*. Silsoe College, Cranfield University, UK.

- Bonilla, C.A., Vidal, K.L., 2011. Rainfall erosivity in central chile. *Journal of Hydrology* 410, 126–133.
- Borrelli, P., Diodato, N., Panagos, P., 2016. Rainfall erosivity in italy: a national  
535 scale spatio-temporal assessment. *International Journal of Digital Earth* 9,  
835–850.
- Chen, S., Zha, X., 2017. Effects of the enso on rainfall erosivity in the fujian  
province of southeast china. *Science of The Total Environment* In press.
- Cooley, K.R., et al., 1980. Erosivity values for individual design storms. *Journal*  
540 *of the Irrigation and Drainage Division, American Society of Civil Engineers*  
106, 135–145.
- Cressie, N., 1993. *Statistics for Spatial Data: Wiley Series in Probability and  
Statistics*. Wiley-Interscience New York.
- Cruse, R., Flanagan, D., Frankenberger, J., Gelder, B., Herzmann, D., James,  
545 D., Krajewski, W., Kraszewski, M., Lafen, J., Opsomer, J., et al., 2006. Daily  
estimates of rainfall, water runoff, and soil erosion in iowa. *Journal of Soil  
and Water Conservation* 61, 191–199.
- Diodato, N., 2004. Estimating rusle’s rainfall factor in the part of italy with a  
mediterranean rainfall regime. *Hydrology and Earth System Sciences Discus-*  
550 *sions* 8, 103–107.
- Elsenbeer, H., Cassel, D., Tinner, W., 1993. A daily rainfall erosivity model for  
western amazonia. *Journal of Soil and Water Conservation* 48, 439–444.
- Ferro, V., Porto, P., Yu, B., 1999. A comparative study of rainfall erosivity es-  
timation for southern italy and southeastern australia. *Hydrological Sciences*  
555 *Journal* 44, 3–24.
- Fiener, P., Neuhaus, P., Botschek, J., 2013. Long-term trends in rainfall  
erosivity–analysis of high resolution precipitation time series (1937–2007)  
from western germany. *Agricultural and Forest Meteorology* 171, 115–123.

- Foster, G., Yoder, D., Weesies, G., McCool, D., McGregor, K., Bingner, R.,  
560 2003. Draft User's Guide, Revised Universal Soil Loss Equation Version 2  
(RUSLE-2). Technical Report. USDA-Agricultural Research Service, Wash-  
ington, DC.
- Goovaerts, P., 1997. Geostatistics for natural resources evaluation. Oxford  
University Press on Demand.
- 565 Hanel, M., Pavlásková, A., Kyselý, J., 2016. Trends in characteristics of sub-  
daily heavy precipitation and rainfall erosivity in the czech republic. Interna-  
tional Journal of Climatology 36, 1833–1845.
- Hernando, D., Romana, M.G., 2015. Estimating the rainfall erosivity factor  
from monthly precipitation data in the madrid region (spain). Journal of  
570 Hydrology and Hydromechanics 63, 55–62.
- Hoomehr, S., Schwartz, J.S., Yoder, D.C., 2016. Potential changes in rainfall  
erosivity under gcm climate change scenarios for the southern appalachian  
region, usa. Catena 136, 141–151.
- Hoyos, N., Waylen, P.R., Jaramillo, Á., 2005. Seasonal and spatial patterns of  
575 erosivity in a tropical watershed of the colombian andes. Journal of hydrolo-  
gyhydrology 314, 177–191.
- Klik, A., Haas, K., Dvorackova, A., Fuller, I.C., 2015. Spatial and temporal  
distribution of rainfall erosivity in new zealand. Soil Research 53, 815–825.
- Van der Knijff, J., Jones, R., Montanarella, L., 2000. Soil erosion risk assessment  
580 in Europe. European Soil Bureau, European Commission Belgium.
- Lai, C., Chen, X., Wang, Z., Wu, X., Zhao, S., Wu, X., Bai, W., 2016. Spatio-  
temporal variation in rainfall erosivity during 1960–2012 in the pearl river  
basin, china. Catena 137, 382–391.
- Laurant, A., Bollinne, A., 1976. L'érosivité des pluies à uccle (belgique). Bulletin  
585 de Recherche Agronomique Gembloux 11, 149–168.

- Lee, M.H., Lin, H.H., 2015. Evaluation of annual rainfall erosivity index based on daily, monthly, and annual precipitation data of rainfall station network in southern taiwan. *International Journal of Distributed Sensor Networks* 11, 214708.
- 590 Lobo, G.P., Bonilla, C.A., 2015. Effect of temporal resolution on rainfall erosivity estimates in zones of precipitation caused by frontal systems. *Catena* 135, 202–207.
- Ma, X., He, Y., Xu, J., van Noordwijk, M., Lu, X., 2014. Spatial and temporal variation in rainfall erosivity in a himalayan watershed. *Catena* 121, 248–259.
- 595 McCullagh, P., Nelder, J.A., 1989. *Generalized linear models* .
- Nearing, M.A., Yin, S.q., Borrelli, P., Polyakov, V.O., 2017. Rainfall erosivity: An historical review. *Catena* 157, 357–362.
- Newman, M.C., 1993. Regression analysis of log-transformed data: Statistical bias and its correction. *Environmental Toxicology and Chemistry* 12, 1129–  
600 1133.
- Onori, F., Bonis, P., Grauso, S., 2006. Soil erosion prediction at the basin scale using the revised universal soil loss equation (rusle) in a catchment of sicily (southern italy). *Environmental Geology* 50, 1129–1140.
- Panagos, P., Ballabio, C., Borrelli, P., Meusburger, K., 2016a. Spatio-temporal  
605 analysis of rainfall erosivity and erosivity density in greece. *Catena* 137, 161–172.
- Panagos, P., Ballabio, C., Borrelli, P., Meusburger, K., Klik, A., Rousseva, S., Tadić, M.P., Michaelides, S., Hrabalíková, M., Olsen, P., et al., 2015a. Rainfall erosivity in europe. *Science of the Total Environment* 511, 801–814.
- 610 Panagos, P., Ballabio, C., Meusburger, K., Spinoni, J., Alewell, C., Borrelli, P., 2017a. Towards estimates of future rainfall erosivity in europe based on redes and worldclim datasets. *Journal of Hydrology* 548, 251–262.

- Panagos, P., Borrelli, P., Meusburger, K., Yu, B., Klik, A., Lim, K.J., Yang, J.E., Ni, J., Miao, C., Chattopadhyay, N., et al., 2017b. Global rainfall erosivity assessment based on high-temporal resolution rainfall records. Scientific Reports 7, 4175.
- 615
- Panagos, P., Borrelli, P., Spinoni, J., Ballabio, C., Meusburger, K., Beguería, S., Klik, A., Michaelides, S., Petan, S., Hrabalíková, M., et al., 2016b. Monthly rainfall erosivity: Conversion factors for different time resolutions and regional assessments. Water 8, 119.
- 620
- Panagos, P., Meusburger, K., Ballabio, C., Borrelli, P., Beguería, S., Klik, A., Rymaszewicz, A., Michaelides, S., Olsen, P., Tadić, M.P., et al., 2015b. Reply to the comment on rainfall erosivity in europe by auerswald et al. Science of the Total Environment 532, 853–857.
- 625
- Pebesma, E.J., 2006. The role of external variables and gis databases in geostatistical analysis. Transactions in GIS 10, 615–632.
- Petkovšek, G., Mikoš, M., 2004. Estimating the r factor from daily rainfall data in the sub-mediterranean climate of southwest slovenia. Hydrological Sciences Journal 49, 869–877.
- 630
- Posch, M., Rekolainen, S., 1993. Erosivity factor in the universal soil loss equation estimated from finnish rainfall data. Agricultural Research Centre of Finland 2, 271–279.
- Qin, W., Guo, Q., Zuo, C., Shan, Z., Ma, L., Sun, G., 2016. Spatial distribution and temporal trends of rainfall erosivity in mainland china for 1951–2010. Catena 147, 177–186.
- 635
- Renard, K.G., Foster, G.R., Weesies, G.A., McCool, D., Yoder, D., et al., 1997. Predicting soil erosion by water: a guide to conservation planning with the revised universal soil loss equation (RUSLE). Technical Report 703. US Department of Agriculture (USDA).



- 640 Renard, K.G., Freimund, J.R., 1994. Using monthly precipitation data to estimate the r-factor in the revised usle. *Journal of Hydrology* 157, 287–306.
- Richardson, C., Foster, G., Wright, D., et al., 1983. Estimation of erosion index from daily rainfall amount. *Transactions of the ASAE* 26, 153–156.
- Risal, A., Bhattarai, R., Kum, D., Park, Y.S., Yang, J.E., Lim, K.J., 2016. Application of web erosivity module (werm) for estimation of annual and  
645 monthly r factor in korea. *Catena* 147, 225–237.
- Romero, L.D., Muñoz, J.L.A., Marín, A.P.G., 2007. Annual distribution of rainfall erosivity in western andalusia, southern spain. *Soil and Water Conservation Society* 62, 390–403.
- 650 Sanchez-Moreno, J.F., Mannaerts, C.M., Jetten, V., 2014. Rainfall erosivity mapping for santiago island, cape verde. *Geoderma* 217, 74–82.
- Serrano-Notivoli, R., Beguería, S., Saz, M.Á., Longares, L.A., de Luis, M., 2017. Spread: a high-resolution daily gridded precipitation dataset for spain—  
an extreme events frequency and intensity overview. *Earth System Science*  
655 *Data* 9, 721.
- Shi, Z., Cai, C., Ding, S., Wang, T., Chow, T., 2004. Soil conservation planning at the small watershed level using rusle with gis: a case study in the three gorge area of china. *Catena* 55, 33–48.
- USDA-ARS, 2013. Science documentation: Revised Universal Soil Loss Equation Version 2. Technical Report. US Department of Agriculture (USDA),  
660 Washington, DC.
- Van Dijk, A., Bruijnzeel, L., Rosewell, C., 2002. Rainfall intensity–kinetic energy relationships: a critical literature appraisal. *Journal of Hydrology* 261, 1–23.
- 665 Verstraeten, G., Poesen, J., Demarée, G., Salles, C., 2006. Long-term (105 years) variability in rain erosivity as derived from 10-min rainfall depth data

- for ukkel (brussels, belgium): Implications for assessing soil erosion rates. *Journal of Geophysical Research: Atmospheres* 111.
- Wang, Y., Cheng, C., Xie, Y., Liu, B., Yin, S., Liu, Y., Hao, Y., 2017. Increasing  
670 trends in rainfall-runoff erosivity in the source region of the three rivers, 1961–  
2012. *Science of The Total Environment* 592, 639–648.
- Wischmeier, W.H., 1959. A rainfall erosion index for a universal soil-loss equation. *Soil Science Society of America Journal* 23, 246–249.
- Wischmeier, W.H., Smith, D.D., 1978. Predicting rainfall erosion losses - A  
675 guide to conservation planning.
- Xie, Y., Yin, S.q., Liu, B.y., Nearing, M.A., Zhao, Y., 2016. Models for estimating daily rainfall erosivity in china. *Journal of Hydrology* 535, 547–558.
- Yang, X., Yu, B., 2015. Modelling and mapping rainfall erosivity in new south wales, australia. *Soil Research* 53, 178–189.
- 680 Yin, S., Nearing, M.A., Borrelli, P., Xue, X., 2017. Rainfall erosivity: An overview of methodologies and applications. *Vadose Zone Journal* 16.
- Yin, S., Xie, Y., Nearing, M., Wang, C., 2007. Estimation of rainfall erosivity using 5-to 60-minute fixed-interval rainfall data from china. *Catena* 70, 306–312.
- 685 Yin, S.q., Xie, Y., Liu, B., Nearing, M., 2015. Rainfall erosivity estimation based on rainfall data collected over a range of temporal resolutions. *Hydrology and Earth System Sciences* 19, 4113.
- Yu, B., Hashim, G., Eusof, Z., 2001. Estimating the r-factor with limited rainfall data: a case study from peninsular malaysia. *Journal of Soil and Water*  
690 *Conservation* 56, 101–105.
- Yu, B., Rosewell, C., 1996. Rainfall erosivity estimation using daily rainfall amounts for south australia. *Soil Research* 34, 721–733.