



The rainfall erosivity factor in the Czech Republic and its uncertainty

Martin Hanel^{1,2}, Petr Máca¹, Petr Bašta¹, Radek Vlnas¹, and Pavel Pech¹

¹Faculty of Environmental Sciences, Czech University of Life Sciences, Kamýcká 1176, Prague 6, Czech Republic

²T. G. Masaryk Water Research Institute, Podbabská 30, Prague 6, Czech Republic

Correspondence to: Martin Hanel (hanel@fzp.czu.cz)

Received: 6 April 2016 – Published in Hydrol. Earth Syst. Sci. Discuss.: 18 April 2016

Revised: 11 July 2016 – Accepted: 9 October 2016 – Published: 25 October 2016

Abstract. In the present paper, the rainfall erosivity factor (R factor) for the area of the Czech Republic is assessed. Based on 10 min data for 96 stations and corresponding R factor estimates, a number of spatial interpolation methods are applied and cross-validated. These methods include inverse distance weighting, standard, ordinary, and regression kriging with parameters estimated by the method of moments and restricted maximum likelihood, and a generalized least-squares (GLS) model. For the regression-based methods, various statistics of monthly precipitation as well as geographical indices are considered as covariates. In addition to the uncertainty originating from spatial interpolation, the uncertainty due to estimation of the rainfall kinetic energy (needed for calculation of the R factor) as well as the effect of record length and spatial coverage are also addressed. Finally, the contribution of each source of uncertainty is quantified. The average R factor for the area of the Czech Republic is $640 \text{ MJ ha}^{-1} \text{ mm h}^{-1}$, with values for the individual stations ranging between 320 and $1520 \text{ MJ ha}^{-1} \text{ mm h}^{-1}$. Among various spatial interpolation methods, the GLS model relating the R factor to the altitude, longitude, mean precipitation, and mean fraction of precipitation above the 95th percentile of monthly precipitation performed best. Application of the GLS model also reduced the uncertainty due to the record length, which is substantial when the R factor is estimated for individual sites. Our results revealed that reasonable estimates of the R factor can be obtained even from relatively short records (15–20 years), provided sufficient spatial coverage and covariates are available.

1 Introduction

Erosion is a natural geological phenomenon resulting from the removal and transportation of soil particles by water, wind, ice, and gravity. Soil erosion by water is a widespread problem throughout Europe (Van der Knijff et al., 2000; Panagos et al., 2015). Erosion is usually driven by a combination of factors like climate (e.g. long dry periods followed by heavy rainfall), topography (steep slopes), inappropriate land use, land cover patterns (e.g. sparse vegetation), ecological disasters (e.g. forest fires), and soil characteristics (e.g. a thin layer of topsoil, a silty texture, or low organic matter content).

Although measurements of soil erosion exist, they are often used as a basis for development, modification, or verification of soil erosion models (applicable at larger scales), which relate soil loss to indicators of relevant factors. A classic example is the Universal Soil Loss Equation (USLE, Wischmeier and Smith, 1978) or the Revised USLE (Renard et al., 1997). Both methods express the long-term average annual soil loss as a product of rainfall erosivity factor (R factor), soil erodibility factor, slope length factor, slope steepness factor, cover-management factor, and support practice factor. The experimental data indicate that when factors other than rainfall are held constant, soil losses from cultivated fields are directly proportional to an erosivity index (EI30) calculated as the total rainfall kinetic energy times the maximum 30 min intensity (Renard et al., 1997). The R factor is then obtained as a long-term average annual rainfall erosivity index. In addition to a point estimate of the R factor, a

spatial information (map) is often required for practical purposes. A number of such maps have been released recently at national (Lu and Yu, 2002; Yin et al., 2007; Bonilla and Vidal, 2011; Oliveira et al., 2013; Borrelli et al., 2016; Panagos et al., 2016a; Meddi et al., 2016) or larger (Panagos et al., 2015) scales.

Because of large spatiotemporal variability of rainfall, long records from a dense network of stations are in general required in order to provide reliable estimates of the R factor and/or to develop rainfall erosivity maps. For instance, Wischmeier and Smith (1978) or Verstraeten et al. (2006) recommend at least 22 years of data to be used for estimation of the R factor. On the other hand, lack of high-resolution rainfall data in combination with a need for soil erosion risk assessment often leads to situations where the R factor has to be estimated from shorter records. For instance, a number of stations used recently for the derivation of rainfall erosivity maps for Europe (Panagos et al., 2015) were shorter than 20 years and several records shorter than 10 years were considered. Similarly, the comparison of the spatial interpolation methods in the Ebro Basin (Angulo-Martínez et al., 2009) was based on 10 years of data (for a large number of stations) and Catari et al. (2011) assessed the uncertainty in the estimated R factor considering a 13-year record for eight stations. In general, the uncertainty in the estimated R factor increases as the length of the available record decreases, but can be reduced by combination of data from different sites (Catari et al., 2011) or by considering covariates that are better sampled and/or whose variation over space and time is smaller (Goovaerts, 1999). For the spatial interpolation of the R factor, variables like longitude, latitude, and elevation (Goovaerts, 1999; Angulo-Martínez et al., 2009) or long-term precipitation (Lee and Lin, 2014) are often considered.

In addition to spatial and temporal variability, the expressions for the rainfall kinetic energy (needed for estimation of the erosivity index) and spatial interpolation are also relevant sources of uncertainty for the development of an R factor map. The rainfall kinetic energy can be estimated by a number of expressions (see e.g. van Dijk et al., 2002). Therefore, several authors assessed the effect on the estimated R factor. For instance, Catari et al. (2011) mentioned a variation due to kinetic energy calculation of about 10 %. Similarly, many spatial models can be used to predict the R factor values over the area. The differences between several spatial interpolation methods have been reported e.g. by Angulo-Martínez et al. (2009). Catari et al. (2011) compared the contribution of different sources of variability to the overall uncertainty in the estimated basin average R factor in north-eastern Spain, concluding that while the uncertainty in the annual erosivity index is dominated by the temporal variability (explaining more than 40 % of variation) for the long-term R factor, the kinetic energy calculation becomes more important.

Several maps of the R factor for the whole of Europe have been released in recent decades. For instance, Van der Kni-

jff et al. (2000) applied simple relationships between seasonal and annual rainfall and the R factor and Panagos et al. (2015) used many high-resolution station datasets (with various record lengths and observation periods) to derive an R factor map based on a Gaussian process model. The values derived for the Czech Republic from these maps range between 350 and 700 MJ ha⁻¹ mm h⁻¹.

A number of estimates of the R factor have been also published specifically for the Czech Republic. See e.g. the overview by Krása et al. (2014), who also mention results from a number of national projects. Official guidelines for soil erosion risk assessment recommend the value 200 MJ ha⁻¹ mm h⁻¹ to be used for all agricultural land in the Czech Republic. Only recently was this value increased to 400 MJ ha⁻¹ mm h⁻¹ (Janeček et al., 2007, 2012a). These values are relatively small with respect to the neighbouring countries and some of the research published for the area. For instance, Janeček et al. (2006) published values of the R factor in the range of 430–1060 MJ ha⁻¹ mm h⁻¹ and concluded that considerably larger values of the R factor (450–600 MJ ha⁻¹ mm h⁻¹) should be used for practical application of USLE in the Czech Republic (instead of 200 MJ ha⁻¹ mm h⁻¹). Even larger values of the R factor are reported by Krása et al. (2015). On the other hand, Janeček et al. (2012b), using a regression between the daily erosion index (EI30) and daily precipitation in order to predict annual EI30, report values of the R factor between 150 and 1200 MJ ha⁻¹ mm h⁻¹, with an average for arable land of 300–400 MJ ha⁻¹ mm h⁻¹. Similarly, Janeček et al. (2013) derived an R factor for the Czech Republic from daily data considering the fraction of erosive events in each year for each station and the areal-average annual sum of the erosivity index. In addition, they excluded years with the largest and years with the smallest erosivity index from the analysis. This resulted in a recommendation to use 400 MJ ha⁻¹ mm h⁻¹ for all agricultural land in the Czech Republic. The trends in annual rainfall erosivity index were studied by Hanel et al. (2016), who found a significant increasing trend ($\approx 4\%$ per decade) in 51-year records for 11 stations (more than half of the considered stations). However, recent values of the erosivity index were not exceptional when compared to a longer (91-year) record available at a single station. Considerable projected increases in the rainfall erosivity index in an ensemble of regional climate model simulations for the Czech Republic were reported by Svoboda et al. (2016b).

The present paper compares several methods of spatial interpolation of the R factor over the Czech Republic and evaluates the bias and uncertainty due to expression for the kinetic energy, spatial model, record length, and spatial coverage. The paper is structured as follows. The study area and data considered for calculation and spatial interpolation of the R factor are given in Sect. 2. Section 3 describes the methods used for estimation of the R factor, spatial interpolation, and uncertainty assessment. The results are presented

and discussed in Sect. 4. The paper is closed with concluding remarks (Sect. 5).

2 Study area and data

2.1 Study area

Because of relatively complex orography and a combination of Atlantic, Mediterranean, and continental effects, the precipitation patterns over the area of the Czech Republic are rather variable (see Fig. 1). The precipitation is mostly due to enhanced westerly flows in winter (Atlantic influence), except for the eastern part of the Czech Republic, which is typically influenced by the Mediterranean Sea. The Mediterranean influences are also dominant over the whole area in summer (e.g. Brádka, 1972; Brázdil, 1980).

The mean annual total precipitation varies from about 400 mm in the western part of the Czech Republic up to more than 1400 mm in the mountains to the north (Tolasz et al., 2007). Almost two-thirds of the annual total falls in the warm half of the year (April–September). The maximum precipitation amounts can be quite large at short timescales, thus contributing considerably to the annual total precipitation. Štekl et al. (2001) point to historical records of 237 mm in 1 h (25 May 1872), 345 mm in 1 day (29 July 1897), 537 mm in 3 days (6–8 July 1997), and 617 mm in 5 days (4–8 July 1997).

2.2 Pluviograph records

A database of 10 min digitized pluviograph records was used in our study. Since a considerable amount of precipitation falls as snow during winter, the data are in general available only for mid-May to mid-September (further referred to as the year). This is, however, the period when heavy precipitation usually occurs. The data were digitized by the Czech Hydrometeorological Institute (CHMI). Detailed identification and reconstruction of unreadable, damaged, or missing records comprised part of the digitization process (see Květoň et al., 2004, for details). Consistency of the aggregated 10 min data with daily records from control ombrometers was further checked by Hanel and Máca (2014). Identified inconsistent days were marked as unreliable and the years with more than 10 % of unreliable data were excluded from the analysis. A set of 96 stations covering the study area for the period 1989–2003 was selected (based mainly on data availability) in order to provide reasonable spatial coverage and record length for the spatial interpolation of rainfall erosivity. The distance between neighbouring stations is 5–42 km (17 km on average) and the stations are located at altitudes from 150 to 1118 m (450 m on average). Note that although a significant positive trend in the annual erosivity index has been reported by Hanel et al. (2016) for several stations in the Czech Republic, no clear trend is observed (not shown) in the considered period (1989–2003). Detailed information

on the considered stations is included in the Supplement to this paper.

To analyse the temporal variability and the effect of the number of stations used for the spatial interpolation, the longest available record (C2TREB01 – Třeboň), with 80 years of data and an additional 40 stations (usually with a slightly different time period than 1989–2003), were also considered (see Sect. 3.3). The location of the stations is indicated in Fig. 1.

The database of 10 min digitized pluviograph records was also used for the estimation of average at-site rainfall event characteristics (event depth, event duration, maximum 10 min intensity, mean event rate, time to peak, depth to peak, and number of events), which were later linked with selected spatial interpolation models (see Hanel and Máca, 2014, for details of the considered rainfall event characteristics).

2.3 Gridded precipitation data

Gridded (1 km resolution) statistics of monthly precipitation for the period 1989–2003 (the same as for the station data) provided by the CHMI were considered as covariates for the spatial interpolation models. Specifically, we used the average (May–September) precipitation (r_{mea}), the coefficient of variation of monthly (May–September) precipitation (r_{cv}), and the mean fraction of precipitation above the 95 % quantile of monthly (May–September) precipitation (r_{p95}). Please note that these gridded statistics are based on much larger numbers of stations (usually more than 750 stations were available for each month and year) than were used in our study for spatial interpolation of the R factor.

3 Methods

3.1 Rainfall erosivity factor

A standard methodology for the assessment of the at-site rainfall erosivity factor (Wischmeier and Smith, 1978) was applied. A continuous 6 h interval without precipitation is used to separate individual rainfall events. To be considered erosive, the cumulative rainfall of an event should be greater than 12.7 mm, or the event should have at least one peak greater than 6.35 mm in 15 min. The latter criterion was modified to 8.5 mm in 20 min in order to match the temporal resolution of our dataset. A similar modification was reported by Meusburger et al. (2012). The proportion of erosive events that are included on the basis of the intensity criterion is, however, small.

The rainfall erosivity factor is defined as a long-term average annual sum of the event erosivity index (EI30 ($\text{MJ mm ha}^{-1} \text{ h}^{-1} \text{ yr}^{-1}$), Wischmeier and Smith, 1978),

$$\text{EI30} = I_{30} \sum_i e_i v_i, \quad (1)$$

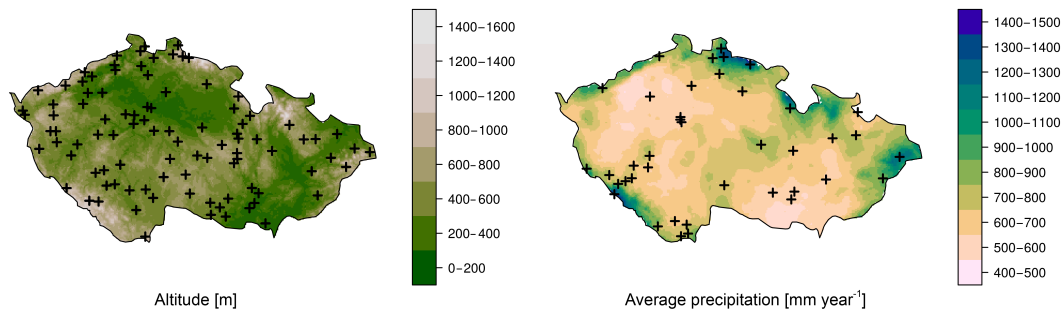


Figure 1. The study area with altitudes and locations of 96 stations used for derivation of the R factor map (left panel) and mean annual precipitation together with locations of additional stations used for uncertainty assessment (right panel).

where

$$e_i = 28.3 \left[1 - 0.52 \exp(-0.042r_i) \right] \quad (2)$$

is the unit rainfall energy ($\text{MJ ha}^{-1} \text{mm}^{-1}$) (van Dijk et al., 2002), v_i and r_i are the rainfall volume (mm) and intensity (mm h^{-1}) during a time interval i , respectively, and I_{30} is the maximum rainfall intensity during a period of 30 min in the event (mm h^{-1}).

The unit rainfall energy in Eq. (2) is from van Dijk et al. (2002), who assessed many expressions for its calculation. To assess the related uncertainty, the R factor was also estimated using additional 14 expressions for e_i (see Appendix A for their definition).

3.2 Spatial interpolation models

An at-site dataset of rainfall erosivity factors was further explored using spatial interpolation models. Following the classification of spatial interpolation models for rainfall erosivity provided by Angulo-Martínez et al. (2009), selected local, geostatistical, mixed, and global models were tested.

The local models, which described the spatial distribution of rainfall erosivity using local at-site information of the R factor, were represented by the model based on inverse distance weighting (IDW; see e.g. Angulo-Martínez et al., 2009; Meusburger et al., 2012). The IDW model predicts the R factor as a weighted average of R factor estimates from a selected number of stations. Weights are inversely proportional to the distance (between the interpolated location and the corresponding station) raised to the power r . The number of considered stations and the value of parameter r were estimated during the model cross-validation (see Sects. 3.2.1 and 4).

The geostatistical models were represented by simple kriging (SK), ordinary kriging (OK), simple cokriging (SC), and ordinary cokriging (OC) models (Goovaerts, 1997, 1999). The arithmetic mean of at-site R factor values was used as the estimator of mean values for simple kriging and cokriging models. The cokriging models were tested using a set of seven rainfall event characteristics (see Sect. 2.2) as

a cokriging variate. Their spatial covariance structures were described by the Matérn model (Minasny and McBratney, 2005; Haskard, 2007; Pardo-Iguzquiza and Chica-Olmo, 2008).

The mixed models were represented using a set of different regression kriging models, further denoted as RK (Hengl et al., 2004, 2007). Their spatial covariance structures of residuals were described using the Matérn model. The spatially varied means of R factors were estimated using three types of inputs:

1. location information represented by longitude (x), latitude (y), and altitude (z);
2. spatial rainfall information expressed by the combinations of r_{mea} , r_{cv} , and r_{p95} ; and
3. information consisting of combinations of both previous types of spatial covariates.

The global spatial interpolation models were represented by the set of generalized linear models (GLS). The deterministic components were estimated using the same types of inputs as in the case of mixed models. Their random error terms were represented using the Matérn and exponential models; the heteroscedasticity of the GLS model residuals was studied by means of the exponential variance function (see p. 206 in Pinheiro and Bates, 2000). The parameters of all tested GLS models were estimated with the restricted maximum likelihood (REML) method (Kitanidis, 1993; Pinheiro and Bates, 2000; Minasny and McBratney, 2007).

3.2.1 Model selection

A standard leave-one-out procedure (Minasny and McBratney, 2007; Angulo-Martínez et al., 2009) was used for the validation of spatial interpolation models. The following indices were considered:

- Willmott's agreement index (WI)

$$WI = 1 - \frac{\sum_{i=1}^n [R_S(x_i) - R_I(x_i)]^2}{\sum_{i=1}^n [|R_S(x_i) - \bar{R}_S(x_i)| + |R_I(x_i) - \bar{R}_I(x_i)|]^2}, \quad (3)$$

- Mean absolute error (MAE)

$$MAE = \frac{1}{n} \sum_{i=1}^n |R_S(x_i) - R_I(x_i)|, \quad (4)$$

- Relative mean absolute error (rMAE)

$$rMAE = \frac{1}{n} \sum_{i=1}^n \frac{|R_S(x_i) - R_I(x_i)|}{R_S(x_i)}, \quad (5)$$

- Root mean square error (RMSE)

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n [R_S(x_i) - R_I(x_i)]^2}, \quad (6)$$

where $R_S(x_i)$ is the rainfall erosivity calculated for station i , $R_I(x_i)$ is the spatially interpolated value of the rainfall erosivity for the same station, and $\bar{R}_S(x_i)$ and $\bar{R}_I(x_i)$ are the mean rainfall erosivities calculated for the station data and obtained from spatial interpolation, respectively.

The choice of the validation criteria follows the discussion about RMSE and MAE presented by Willmott and Matsuura (2005), who recommended the use of MAE for the estimation of average model error over the RMSE, since the RMSE can be influenced by outlying observations. The WI was also considered for consistency with other relevant studies (for example, see Angulo-Martínez et al., 2009).

The model selection aimed at identification of robust spatial interpolation methods from the considered groups and also at identification of optimal structures and parameters of the particular models, including various combinations of covariates, covariance models, and the number of nearest stations considered for local, geostatistical, and mixed models.

3.3 Uncertainty assessment

The uncertainty in the derived R factor map originates from the formulation of the kinetic energy term, the spatial interpolation model, and spatial and temporal variability. While the effect of the former two can be assessed by direct comparison of results for different expressions of rainfall kinetic energy and different spatial interpolation models, the effect of spatial and temporal variability is evaluated here by simple bootstrap resampling procedures, which are briefly summarized in the rest of this section.

3.3.1 Temporal variability

Record length influences the width of the confidence interval around the R factor estimate. Due to the temporal variability, sufficient record length is required in order to provide an estimate of the R factor such that the long-term average R factor (further denoted the “true R factor”) would be covered by the estimated confidence interval. Therefore, we derived the confidence intervals for various record lengths together with the probability that the “true R factor” will lie within the corresponding confidence interval (further denoted the “coverage probability”) using our longest available record, i.e. station C2TREB01 (Třeboň) with 80 years of data. This can be done using a nested bootstrap procedure in which a sample of required length (e.g. 10, 20, or 80 years) is drawn (with replacement) from the original record and further resampled. This allows for determination of the confidence interval and examination whether the “true R factor” is included. To obtain the coverage probability, the whole procedure has to be repeated many times. Here we evaluated the coverage probabilities for the record lengths of 10–80 years. For details, see Appendix B.

3.3.2 Record length and spatial coverage

Long records from a dense network of stations should be ideally available to derive an R factor map. In reality, however, long records are often available only for a relatively small number of stations and a balance between record length and spatial coverage has to be found. It is then not clear whether longer records or better spatial coverage should be preferred.

Specifically, we considered (a) what the relationship of the error in the estimated R factor with the spatial and temporal coverage is and how spatial and temporal coverage influences it, (b) the width of the confidence interval around the estimated R factor, and (c) the coverage probability (i.e. the probability that the estimated confidence interval will include the “true R factor”).

To be able to assess these questions, a simulation study was conducted. The procedure is fully described in Appendix C, and here we provide only a general overview. As a reference, a synthetic dataset of the monthly (May–September) erosivity index (EI30) was created by permutation of 10 years of data available for 120 stations as follows: first, a 100-year long sequence of the months May–September was created and a random year (from the available period) was assigned to each month in each year. Data for each of the 120 stations were then rearranged according to this year–month sequence and the data were aggregated by years. This resulted in a dataset of 100 years for 120 stations. This procedure preserves the annual cycle of the erosivity index and its spatial variability, while it assumes independence of the erosivity index between individual months. This dataset is further denoted the “full dataset”. Please note that although many different replications of this “full dataset” can

be obtained, only one replication is used in this study. We do not expect the results (presented further) to vary significantly when different replication is considered.

The R factor in this “full dataset” was estimated and a simple GLS model of the form $R \sim \text{NAVY} + r_{\text{mea}} + Y$ was fitted. This model was used to predict the R factor for an additional 62 locations (coincident with real stations, but independent of the “full dataset”). This is further denoted the “validation dataset”. The assessment of the effects of spatial and temporal coverage was based on a repetitive resampling of the subsets of the “full dataset”, fitting a GLS model and predicting the R factor for the “validation dataset” (for details, refer to Appendix C).

Within the simulation study, we evaluated the RMSE, the width of the 90 % confidence interval, and the coverage probability for record lengths of 5, 10, 15, 20, 30, . . . , 100 years for 10, 20, . . . , 120 stations. Since it has been frequently noted that areal averaging is in general preferred to spatial interpolation in the Czech Republic (Janeček et al., 2006, 2012b, 2013), at least for agricultural areas (in general at low altitudes), we also assessed this “areal-average model” (i.e. a constant R factor for all locations). The “areal-average model” was applied for the whole “validation dataset” and also considering only the stations with altitudes below 600 m.

3.3.3 Comparison of uncertainties

To compare the contribution of different sources of uncertainty, the R factor was predicted for the “validation data” using

- the best GLS model of the R factor considering 15 different expressions for kinetic energy (see Appendix A);
- selected classes of spatial interpolation models;
- GLS models based on different numbers of stations and years (see Sect. 3.3.2).

The coefficient of variation (CV) for each set of the estimated R factors was calculated to summarize the variability due to different sources, similarly as done by Catari et al. (2011) or Panagos et al. (2015). For comparison, we also evaluated the CV for the R factor estimates, considering various record lengths for the C2TREB01 (Třeboň) station (see Sect. 3.3.1). Finally, the CV for the R factor estimate for each of the 96 stations (representing natural variability) was evaluated using a simple bootstrap resampling of the annual erosivity values.

4 Results and discussion

4.1 R factor

The estimated R factor (considering the kinetic energy relationship proposed by van Dijk et al., 2002, Eq. 2) for

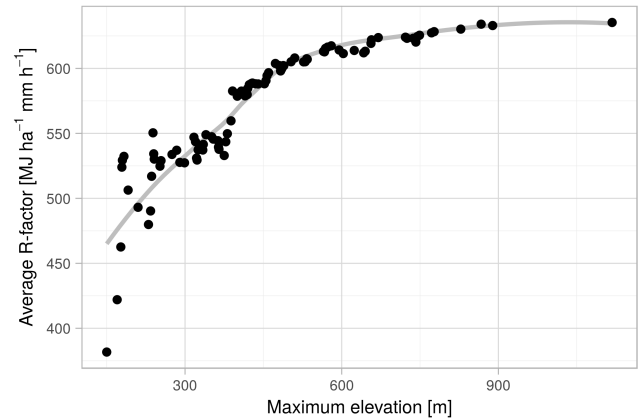


Figure 2. Average R factor for subsets of stations with smaller elevations than that plotted on the horizontal axis. The leftmost point corresponds to a station with the lowest altitude, the rightmost point to the overall average R factor.

the 96 stations used for spatial interpolation ranges between 320 (U1KOPI01 – Kopisty, north-western Czech Republic) and 1520 (O1RASK01 – Raškovice, north-eastern Czech Republic) $\text{MJ ha}^{-1} \text{mm h}^{-1}$, and averages $640 \text{ MJ ha}^{-1} \text{mm h}^{-1}$. Note that the second-largest R factor (station H2DEST01 – Deštné v Orlických horách) equals $1108 \text{ MJ ha}^{-1} \text{mm h}^{-1}$. R factor values for individual stations can be found in the Supplement to this paper. Figure 2 shows average R factor values for subsets of stations based on the maximum station elevation included in the subset. For instance, the average R factor for the elevations up to 300 m is slightly less than $550 \text{ MJ ha}^{-1} \text{mm h}^{-1}$ and for elevations up to 600 m slightly more than $600 \text{ MJ ha}^{-1} \text{mm h}^{-1}$. The average contribution of individual months to the annual total R factor is 17 % for May, 19 % for June, 28 % for July, 26 % for August, and 10 % for September.

The at-site R factor estimates correspond well to those published by Janeček et al. (2006) and Krása et al. (2015), but are considerably larger than values recommended by the official guidelines for the Czech Republic (Janeček et al., 2007, 2012a) and those published by Janeček et al. (2012b, 2013). In part, differences might be due to the modifications to the standard USLE methodology considered by Janeček et al. (2012b, 2013), e.g. calculation of the R factor as a trimmed mean (excluding years with the two smallest and two largest annual erosivity values) of the annual erosivity index. Perhaps in small measure, the differences can also be attributed to the different time period used for R factor assessment (here 1989–2003; in other studies, the series from 1960 and earlier were often considered). The estimated average R factor, after reduction for temporal resolution (from 10 to 30 min; Panagos et al., 2016b), becomes $525 \text{ MJ ha}^{-1} \text{mm h}^{-1}$, which almost equals the average R factor from the rainfall erosivity map of Europe – $524 \text{ MJ ha}^{-1} \text{mm h}^{-1}$ for the Czech Republic (Panagos et al.,

Table 1. Cross-validation indices for the best variants of each spatial interpolation model. AVst – at-site average; SDst – at-site standard deviation; AVsp – spatial average; SDsp – standard deviation of the spatial R factor. The best cross-validation indices are marked in bold font.

	WI	RMSE	MAE	rMAE	AVst	SDst	AVsp	SDsp
IDW	0.78	164.33	123.68	0.20	625.69	162.15	649.46	165.37
SK	0.79	163.01	124.78	0.21	626.98	163.73	647.33	174.49
OK	0.79	165.96	126.48	0.21	630.72	174.62	656.59	135.70
SC	0.73	169.72	131.15	0.22	631.26	139.77	644.53	151.57
OC	0.79	158.78	121.86	0.21	632.86	158.90	656.32	138.13
RK _{XYZ}	0.79	171.44	131.34	0.22	636.86	191.49	653.69	187.78
RK _{$r_{\text{mea}}, r_{\text{cv}}$}	0.89	126.89	93.59	0.15	636.87	185.67	626.22	168.13
RK _{$xyz, r_{\text{mea}}, r_{\text{cv}}$}	0.90	121.91	93.13	0.15	636.18	183.74	632.77	158.78
GLS _M	0.90	115.38	90.57	0.15	635.34	175.12	628.27	149.29
GLS _E	0.91	114.37	89.95	0.14	635.64	177.47	627.80	153.35

2015). In addition to different temporal resolutions, Panagos et al. (2015) also considered different stations and kinetic energy formulation (i.e. that of Brown and Foster, 1987; see Fig. 5 for a comparison with van Dijk et al., 2002, used here). Although the spatial distribution of the R factor over the Czech Republic appears rather homogeneous in Panagos et al. (2015), the range of the at-site R factors (after correction for temporal resolution) corresponds well to that from Panagos et al. (2015) when the station with the maximum rainfall erosivity (O1RASK01) is left out. The effect of this station on the resulting R factor map is further discussed in the following section.

4.2 Model selection

The estimated at-site R factor is positively correlated with average precipitation (r_{mea}), coefficient of variation of monthly precipitation (r_{cv}), and the mean fraction of precipitation above the 95 % quantile of monthly precipitation (r_{p95}) derived from the gridded data (see Sect. 2.3), with correlation coefficients 0.75, 0.44, and 0.54, respectively. Further, a positive correlation was also found with altitude (0.32) and longitude (0.47), and a weak negative correlation (-0.25) with latitude. These variables have been primarily used also as covariates in the relevant spatial interpolation methods. Note that correlation may be substantially weaker or that different variables may be relevant in other regions, especially regions with complex topography (see e.g. Capra et al., 2015 or Porto, 2016).

To explore these relationships further, we analysed a set of GLS models with various combinations of fixed component covariates and spatial stochastic covariance structures. The number of fixed-term covariates ranged from one to five. Two theoretical models of spatial covariance and a heteroscedastic error model were considered (see Sect. 3.2). The best GLS model according to the cross-validation results of WI, RMSE, and rMAE had four fixed-term covariates: r_{mea} , r_{p95} , altitude, and longitude, exponential spatial correlation struc-

ture, and exponential heteroscedastic error. The coefficient of determination for regression between at-site R factor values and R factor values predicted by this GLS model equals 0.7, which is comparable to results of Meusburger et al. (2012). This was followed by a GLS model with fixed-term covariates longitude, latitude, altitude, r_{mea} , and r_{cv} and a Mathérn spatial covariance structure, which had the lowest value of MAE. These two GLS models are further referred to as GLS_E and GLS_M, respectively. The GLS_E model with estimated coefficients reads as

$$R = 10.246r_{\text{mea}} + 2902.462r_{\text{p95}} - y - 0.130z + 3649.792, \quad (7)$$

with y the latitude (km) (in the Czech S-42–Pulkovo 1942/Gauss–Kruger zone 3 projection) and z the altitude (m). The parameters are transformed such that the equation can be directly applied to observed covariates. Please note that r_{mea} and r_{p95} are based on (1 km) gridded data (see Sect. 2.3). Calculating these indices using station data would likely lead to a positive bias in the resulting R factor. This is due to area reduction effects influencing areal-average rainfall event characteristics (see e.g. Svoboda et al., 2016a, for a discussion).

Figure 3 further demonstrates the effect of excluding station(s) with a large R factor on the interpolated R factor in the case of the GLS_E model. Although the mean and maximum R factors decrease and local R factor patterns change slightly when stations with a large R factor are excluded, the differences are small (especially with respect to other sources of uncertainty; see further), suggesting our model is rather robust.

In addition to GLS models, other spatial interpolation methods were considered. Table 1 presents the cross-validation indices, spatial average, and standard deviation of the interpolated R factor. The estimated average R factor ranges from 626 to 657 MJ ha⁻¹ mm h⁻¹. Comparing the R factor estimated by different spatial interpolation models, the largest similarities were found between the GLS_E, GLS_M, and RK _{$xyz, r_{\text{mea}}, r_{\text{cv}}$} . The GLS_M and RK _{$xyz, r_{\text{mea}}, r_{\text{cv}}$} models had

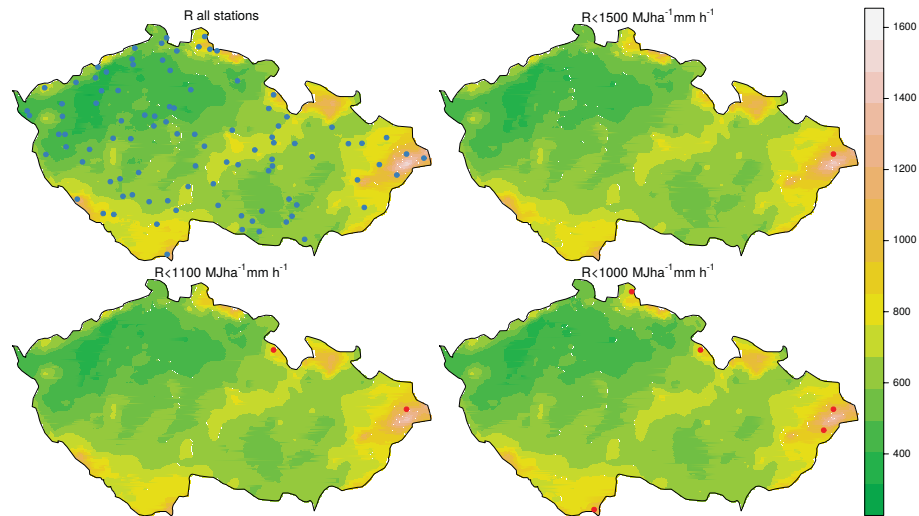


Figure 3. R factor estimated with the GLS_E model using all available stations (top left panel) and the GLS_E model fitted considering only stations with an R factor below $1500 \text{ MJ ha}^{-1} \text{ mm h}^{-1}$ (top right panel), $1100 \text{ MJ ha}^{-1} \text{ mm h}^{-1}$ (bottom left panel), and $1000 \text{ MJ ha}^{-1} \text{ mm h}^{-1}$ (bottom right panel). Blue points in the top left panel indicate the full set of 96 stations; red dots in the other panels correspond to stations excluded from the fitting procedure.

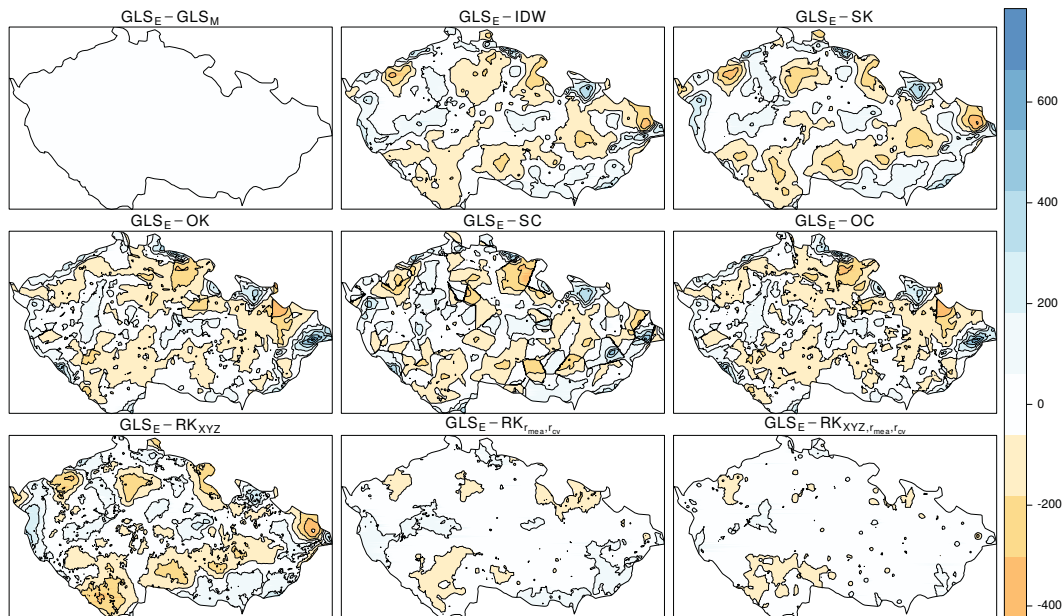


Figure 4. Differences ($\text{MJ ha}^{-1} \text{ mm h}^{-1}$) between the GLS_E and other spatial interpolation models. GLS_E – generalized linear model with exponential covariance structure; GLS_M – generalized linear model with Matérn covariance structure; IDW – inverse distance weighting; SK – simple kriging; OK – ordinary kriging; SC – simple cokriging; OC – ordinary cokriging; RK – regression kriging.

fixed-term inputs formed from longitude, latitude, altitude, r_{mea} , and r_{cv} . The correlation coefficient between estimates of GLS_E and GLS_M equals 0.99, and between the GLS_E model and $RK_{xyz,r_{\text{mea}},r_{\text{cv}}}$ models equals 0.97. These similarities were confirmed on the rasters of differences between values of the GLS_E model and the remaining spatial interpolation models (see Fig. 4). The median abso-

lute difference between the GLS_E and GLS_M models was $20.0 \text{ MJ ha}^{-1} \text{ mm h}^{-1}$, with the standard deviation of differences $19.2 \text{ MJ ha}^{-1} \text{ mm h}^{-1}$. For the differences between the GLS_E and the best RK model, the median absolute difference and standard deviation were roughly double.

The largest differences were found between the GLS_E model and the spatial interpolation models, which did

not take into account the long-term rainfall characteristics. The largest difference in R factor estimates ($741 \text{ MJ ha}^{-1} \text{ mm h}^{-1}$) was found between the GLSE and OK models. Large values of MAE, rMAE, and RMSE and small values of WI for the IDW, SK, OK, SC, and OC models show that the spatial distribution of the R factor could not be sufficiently described by the models, which emphasized the stochastic component of the R factor, or explain the R factor using local information. The spatial interpolation models with fixed-term covariates based on r_{mea} or r_{cv} or r_{p95} were superior to the models without a fixed component linked to the long-term rainfall characteristics (cf. the cross-validation results of RK_{XYZ}).

Including the stochastic information obtained from rainfall event characteristics in simple cokriging and ordinary cokriging models also did not improve the spatial interpolation of the R factor. The presented SC and OC models were selected from seven different types of cokriging models. They differed according to the rainfall event characteristic, which was used for the second cokriging variate. The best SC and OC models were those which linked their stochastic component with the maximum 10 min intensity and on-site R factor. Including these rainfall event characteristics did not, however, improve the spatial interpolation of the R factor over the spatial models based on the long-term rainfall event characteristics (see Table 1).

The success of models including long-term precipitation characteristics might be surprising in part because daily or subdaily rainfall data are in general preferred for calculation of the R factor over monthly or annual data (Angulo-Martínez et al., 2009). The potential of long-term rainfall characteristics for R factor estimation is also stressed by Lee and Lin (2014), who explored relationships between rainfall and erosivity indices at daily, monthly, and annual timescales and concluded that the relationship between the annual erosivity index and annual rainfall is closer than that for the other timescales.

The average R factor varies considerably when different formulas for kinetic energy are considered (see Fig. 5). The smallest average R factor is obtained by the RWa relationship ($500 \text{ MJ ha}^{-1} \text{ mm h}^{-1}$) and the largest value by the standard USLE ($790 \text{ MJ ha}^{-1} \text{ mm h}^{-1}$). The average ($640 \text{ MJ ha}^{-1} \text{ mm h}^{-1}$) corresponds well to the value estimated using the van Dijk formula (the difference is $6 \text{ MJ ha}^{-1} \text{ mm h}^{-1}$). The range between estimates for individual stations is proportional to the estimated R factor (corresponding roughly to 40%). This also has an effect on the estimated spatial distribution of the R factor values.

4.3 Temporal variability

Wischmeier and Smith (1978) used a 22-year record to derive the rainfall erosivity factor because of “apparent cyclical patterns in rainfall data”. The same is repeated by Renard et al. (1997) with a remark that longer records are advis-

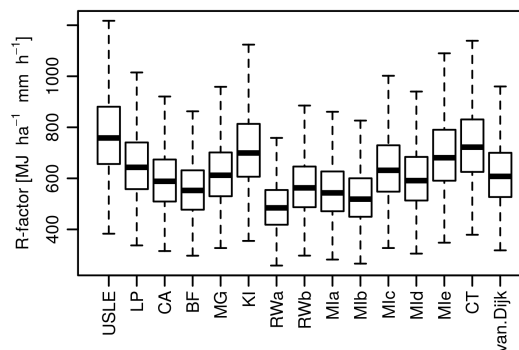


Figure 5. Estimated R factor ($\text{MJ ha}^{-1} \text{ mm h}^{-1}$) considering different kinetic energy formulations (see Appendix A).

able especially in the case that the coefficient of variation of annual precipitation is large. This recommendation is often mentioned in rainfall erosivity studies. However, due to data availability, shorter records are often considered, at least in addition to longer records (e.g. Angulo-Martínez et al., 2009; Meusburger et al., 2012; Oliveira et al., 2013; Lee and Lin, 2014; Panagos et al., 2015). For a 105-year record from Belgium, Verstraeten et al. (2006) tested whether rainfall erosivity derived from running 10- and 22-year averages is significantly different than that from using the overall (105-year) mean. They concluded that while a 22-year period is sufficient, reliable estimates of the R factor cannot be based on 10 years of data. Apart from their study, the actual effect of the sample size on the estimate of the R factor was seldom investigated.

Figure 6 (left panel) gives the estimated confidence intervals (grey area) together with the coverage probability, i.e. the probability that the long-term mean R factor ($669 \text{ MJ ha}^{-1} \text{ mm h}^{-1}$) will lie within the confidence intervals for record lengths between 10 and 80 years for the Třeboň station (C2TREB01). The 90% confidence interval for record length 10 years ranges from 400 to $950 \text{ MJ ha}^{-1} \text{ mm h}^{-1}$ (i.e. $\pm 40\%$), narrows to $500\text{--}870 \text{ MJ ha}^{-1} \text{ mm h}^{-1}$ ($\pm 25\text{--}30\%$) for 20 years and $530\text{--}830 \text{ MJ ha}^{-1} \text{ mm h}^{-1}$ ($\pm 20\text{--}25\%$) for 30 years, and remains relatively wide, i.e. $580\text{--}770 \text{ MJ ha}^{-1} \text{ mm h}^{-1}$ ($\pm 15\%$), even for an 80-year record. Note that the temporal variability of the erosivity index is considerably larger than in the case of the annual total or maxima. For instance, the standard deviation of the annual erosivity index is more than double that of the annual precipitation total and more than 40% larger than for annual 1 h precipitation maxima (estimated for the same station). The coverage probability (red line in Fig. 6, left panel) of the 90% confidence interval is around 75% for the 10-year record, and from 82% for the 15-year record it increases only slowly, with an increasing record length of up to 87% for the 80-year record.

The width of the confidence interval as well as the coverage probability are influenced by the variability of the at-

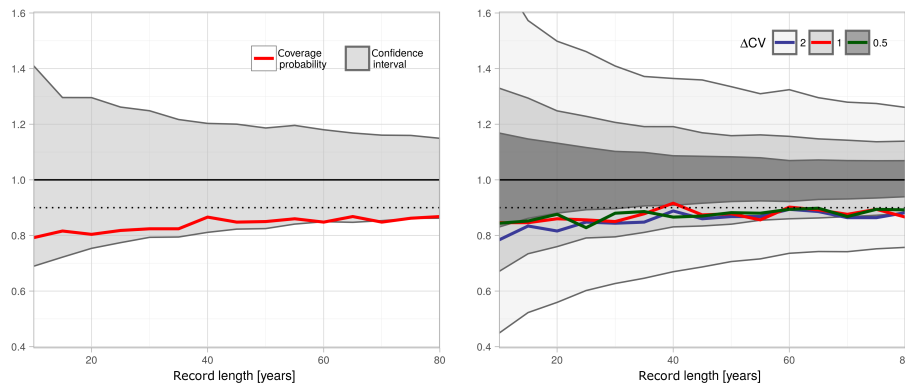


Figure 6. The average confidence intervals (relative to the long-term mean R factor, $669 \text{ MJ ha}^{-1} \text{ mm h}^{-1}$; grey area) and the coverage probability (thick lines) for different record lengths based on station C2TREB01 – Třeboň (left panel) and simulated data (right panel). The dotted line corresponds to the coverage probability of 90 %.

site erosivity index. This is demonstrated in Fig. 6 (right panel) for synthetic data generated from a Gamma distribution with parameters estimated from the C2TREB01 record (denoted $\Delta\text{CV} = 1$) and with parameters modified such that the coefficient of variation of the modified distribution is half and double that of the C2TREB01 record ($\Delta\text{CV} = 0.5$ and $\Delta\text{CV} = 2$, respectively), and the mean R factor remains constant. Note that the Gamma distribution was not rejected by the Anderson–Darling test at the 0.05 significance level at most of the stations, including C2TREB01. It is clear that the confidence intervals as well as the coverage probability for the erosivity index simulated with parameters estimated from C2TREB01 (i.e. $\Delta\text{CV} = 1$) correspond reasonably to those based on observed data. It is also clear (and expected) that the width of the confidence interval increases with the coefficient of variation. For instance, for a 15-year record and doubling of the coefficient of variation, it ranges from 0.5 to 1.57. For increasing record lengths the coverage probability increases and the width of the confidence interval decreases. Note that the confidence interval for the erosivity index with a large coefficient of variation remains relatively large ($> 50\%$), even for 80 years of data. The coverage probability, on the other hand, decreases only slightly with the coefficient of variation.

4.4 Spatial and temporal coverage

Figure 7 shows the average RMSE, coverage probability, and relative width of the 90 % confidence interval for the validation data in the cases of the GLS model (solid line), “areal-average” model (dotted line), and “areal-average” model considering only stations with an altitude of less than 600 m (dashed line). In the case of the GLS model, the RMSE for small numbers of stations and years is relatively large ($\approx 300 \text{ MJ ha}^{-1} \text{ mm h}^{-1}$). However, it quickly drops to $\approx 100 \text{ MJ ha}^{-1} \text{ mm h}^{-1}$ for 25 stations and then decreases almost linearly to $\approx 50 \text{ MJ ha}^{-1} \text{ mm h}^{-1}$ for 100 stations. The RMSE for both “areal-average” models de-

pends on the number of stations only up to 25 stations and remains almost constant for larger numbers of stations ($\approx 110 \text{ MJ ha}^{-1} \text{ mm h}^{-1}$ when validated against all stations and $\approx 100 \text{ MJ ha}^{-1} \text{ mm h}^{-1}$ when only stations below 600 m are used). The RMSE depends only slightly on the number of years used for estimation of the R factor in the case of all models (GLS and both “areal-average” models). With respect to RMSE, the “areal-average” model is beneficial only when fewer than 20 stations are available. The RMSE for individual stations might be considerably larger. The 90 % quantile RMSE (not shown) is around $600 \text{ MJ ha}^{-1} \text{ mm h}^{-1}$ for both “areal-average” models (mostly independently of the number of stations) and from $2000 \text{ MJ ha}^{-1} \text{ mm h}^{-1}$ (10 stations) to $450 \text{ MJ ha}^{-1} \text{ mm h}^{-1}$ (30 stations) and $250 \text{ MJ ha}^{-1} \text{ mm h}^{-1}$ (100 stations) for the GLS model with 15 years of data.

The width of the 90 % confidence interval (Fig. 7, middle panel) for the GLS model drops from 120 % (5 stations) to 50 % (25 stations) and 30 % (100 stations). This value corresponds well to the width of the confidence interval for the full record of the C2TREB01 (Třeboň) station (see Fig. 6). The confidence interval for both “areal-average” models is approximately half that for the GLS model. The impact of the record length is small.

The coverage probability (Fig. 7, right panel) for the GLS model increases from 70 to 90 % for 5 and 10 stations, respectively. The coverage probability further increases only when more than 20 years of data are used. For shorter records, the coverage probability decreases to $\approx 80\%$ for 100 stations. This is a consequence of faster reduction of the confidence interval width when compared to the decrease in the RMSE. Similarly, the coverage probability decreases for both “areal-average” models, since the RMSE is almost constant for more than 25 stations, while the width of the confidence interval decreases. As with the RMSE, the coverage probability might be considerably lower for individual stations, especially in the case of both “areal-average” models (for a number of stations close to 0), while the coverage prob-

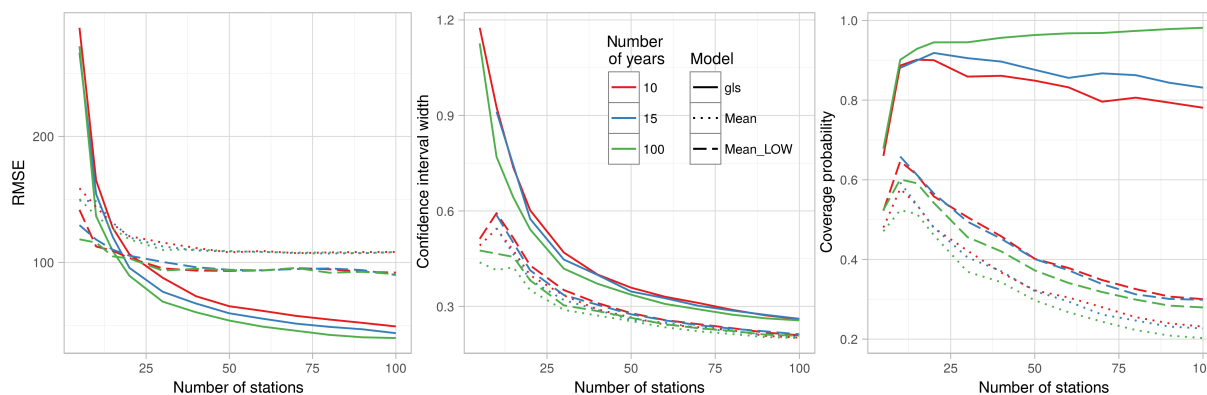


Figure 7. The RMSE, width of the confidence interval, the coverage probability for a GLS model (solid lines), the “areal-average” model considering all stations (dotted lines), and only those below 600 m (dashed lines) based on simulated data with various record lengths and number of stations.

ability is larger than 70 % for most of the stations in the case of the GLS model.

The results for all evaluated characteristics indicate that appropriate spatial coverage is more important than the length of the record, at least in situations when other relevant information to build a spatial model is available. However, at least 15–20 years of data should be considered (if possible) to provide reasonable coverage probabilities.

4.5 Comparison of different sources of uncertainty

Using the “validation data” and the GLS model, we calculated the coefficient of variation (CV) associated with formulation of the kinetic energy, spatial interpolation, and spatial and temporal coverage (Table 2). In addition, the CV was also calculated for estimates of the *R* factor based on different record lengths for the C2TREB01 (Třeboň) station and for the set of 96 stations considered for spatial interpolation. For the latter, the estimated CV was 23 % on average (9–43 % for all stations). This value can be interpreted as an indicator of natural variability of the *R* factor based on a 15-year record. Almost the same value (21 %) is estimated for the C2TREB01 (Třeboň) station and 15 years of data. The contribution of the kinetic energy formulation ($\approx 13\%$) and spatial interpolation ($\approx 9\%$) is about half of this value. As expected, the CV of the estimates for the C2TREB01 (Třeboň) station decreases with increasing record length (38, 16, and 12 % for 5, 50, and 80 years, respectively).

The same applies for the GLS model, for which, in addition, the CV also decreases with an increasing number of stations. When comparing corresponding record lengths, the CV is considerably smaller for the GLS model than for individual stations, providing a sufficient number of stations is considered in the model. For instance, for the *R* factor based on 15 years of data, the average CV is 31, 10, 7, and 6.4 % for 10, 30, 50, and 100 stations, respectively, while for the station data and the same record length the average CV was

Table 2. The coefficient of variation (CV) for the *R* factor estimated by considering different formulas for kinetic energy, spatial interpolation methods, record length, and spatial coverage. In the second column, the average CV for all considered stations is given; the last two columns indicate the range of the CV from all stations in the “validation dataset”.

Source	Coefficient of variation (%)	
Kinetic energy	12.7	(11.8–14.4)
Spatial interpolation	9.4	(1.5–29.7)
Number of stations (GLS model, 15 years)		
10 stations	31.0	(19.6–69.1)
30 stations	10.3	(7.6–21.4)
50 stations	7.6	(6.0–13.9)
Number of years (GLS model, 100 stations)		
5 years	13.0	(10.9–22.3)
15 years	6.4	(4.9–11.8)
50 years	4.9	(3.4–9.5)
Number of years (C2TREB01 – Třeboň)		
5 years	37.6	–
15 years	21.6	–
50 years	16.1	–
80 years	11.8	–
Natural variability (96 stations, 15 years)		
	23.3	(9.1–43.0)

23 %. In addition, considering 100 stations, the CV is only 13 (5) % for 5 (50) years. From Table 2 it is evident that not only does the average CV decrease, but that the same holds also for the stations with maximum variation. For instance, using 100 stations and 15 years, the maximum CV in the “validation set” is 12 % for a GLS model and 43 % for the station data. As the CV decreases with more stations and longer records, the relative importance of the expression

for the kinetic energy and spatial interpolation increases. The assessment of the sources of uncertainty could be done more formally using an analysis of variance (ANOVA) model (e.g. Yip et al., 2011) or a slightly more flexible linear mixed-effects model (see e.g. Hanel and Buishand, 2015).

5 Summary and conclusions

In the present paper we estimated the rainfall erosivity factor (R factor) for the area of the Czech Republic. The at-site values of the R factor based on a 15-year record for 96 stations were considered in several spatial models in order to provide estimates of the R factor for the whole area of the Czech Republic.

The spatial interpolation models included inverse distance weighting, simple and ordinary kriging, simple and ordinary cokriging, regression kriging with parameters estimated by the method of moments, and the GLS models. Several covariates have been considered to explain the spatial variation of the R factor over the area.

In addition, uncertainty due to kinetic energy formulation, spatial models, and spatial and temporal coverage was assessed by direct comparison of different methods and simulation studies.

The most important findings can be summarized as follows:

- the average R factor in the period 1989–2003 for the considered stations is $640 \text{ MJ ha}^{-1} \text{ mm h}^{-1}$, with values for the individual stations between 320 and $1520 \text{ MJ ha}^{-1} \text{ mm h}^{-1}$;
- the at-site R factor is considerably correlated with average precipitation, coefficient of variation of monthly precipitation, the mean fraction of precipitation above the 95 % quantile of monthly precipitation, and longitude, while the correlation with altitude and latitude is weak;
- from the considered spatial models, a GLS model with altitude, latitude, mean precipitation, and the mean fraction of precipitation above the 95 % quantile of monthly precipitation provided the best performance according to three of four cross-validation indices;

- with respect to the cross-validation statistics, the spatial interpolation models that included long-term rainfall characteristics performed considerably better than those based on local interpolation and/or geographical information only;
- the resulting map based on the GLS model does not change considerably when stations with the largest R factor are excluded;
- when the number of stations and years available for interpolation is small, the relative contribution of the uncertainty due to the kinetic energy estimate and the spatial interpolation method is small compared to that due to the choice of the stations and time period;
- although the RMSE and confidence interval width decrease and coverage probability in general increases with record length and number of stations, reasonable estimates of the R factor may be obtained from relatively short records (e.g. 15–20 years) provided a sufficient number of stations are available and appropriate covariates can be found;
- the confidence intervals around the R factor estimates remain relatively wide even for long records, especially at locations with large natural variability of the annual erosivity index, and are considerably wider than those for annual rainfall total or precipitation maxima;
- the spatial model should in general be preferred over the areal-averaging of R factor estimates from individual stations unless only very short records for a small number of stations are available or appropriate covariates cannot be used.

6 Data availability

The original 10 min precipitation data as well as the gridded precipitation characteristics cannot be published due to licence of the Czech Hydrometeorological Institute. The R factor for 96 available stations calculated using 15 models for rainfall kinetic energy (Appendix A) is available at doi:10.6084/m9.figshare.c.3521319.

Table A1. Coefficients for calculation of rainfall kinetic energy in Eq. (A2). The acronyms used throughout the paper are given in the first column.

	e_{\max}	a	b	
LP	28.9	0.54	0.059	Laws and Parsons (1943)
CA	28.0	0.76	0.090	Carter et al. (1974)
BF	29.0	0.72	0.050	Brown and Foster (1987)
MG	29.0	0.72	0.082	McGregor et al. (1995)
KI	29.3	0.28	0.018	Kinnell (1981)
RWa	26.4	0.67	0.035	Rosewell (1986)
RWb	28.1	0.60	0.040	Rosewell (1986)
MIa	24.6	0.46	0.037	McIsaac (1990)
MIb	29.2	0.51	0.011	McIsaac (1990)
MIc	28.8	0.45	0.033	McIsaac (1990)
MId	25.1	0.40	0.045	McIsaac (1990)
MIe	26.8	0.29	0.049	McIsaac (1990)
CT	35.9	0.56	0.034	Coutinho and Tomás (1995)

Appendix A: Considered formulations of rainfall kinetic energy

For the application of the Universal Soil Loss Equation, Wischmeier and Smith (1978) derived a logarithmic relationship between rainfall intensity and kinetic energy of the form (converted to metric units)

$$e_i = 210 + 89 \log(r_i/10), \quad (\text{A1})$$

where (as in Eq. 2), r_i is the rainfall intensity (mm h^{-1}) during time interval i . Note that slightly different coefficients are provided by Renard et al. (1997). The logarithmic relationship implies that there is no upper limit to kinetic energy, whereas research has suggested that a maximum value does exist (see van Dijk et al., 2002, for references). Therefore, Wischmeier and Smith (1978) considered constant rainfall kinetic energy for intensities greater than 76 mm h^{-1} . Other authors used a relationship of the form

$$e_i = e_{\max} [1 - a \exp(-br_i)], \quad (\text{A2})$$

where e_{\max} denotes the maximum kinetic energy contents and a and b are empirical constants. Many different combinations of the parameters e_{\max} , a , and b have been published. van Dijk et al. (2002) therefore proposed a relationship given in Eq. (2) as one providing estimates that are close to the average of many formulas for calculation of e_i . In the present paper, in addition to formulas given in Eqs. (2) and (A1), which are further referred to as “van Dijk” and “USLE”, respectively, we calculated the e_i considering a set of coefficients for Eq. (A2) given in Table A1.

Appendix B: A resampling scheme for the assessment of temporal variability

Here we describe a nested bootstrap procedure in which samples of required length l (e.g. 10, 20, 80 years) are repeatedly drawn from the original (observed) record and resampled to obtain the R factor estimate with the corresponding confidence interval. Finally, the probability that the “true R factor” (here the estimate based on the full record) lies within the confidence interval is estimated.

The resampling is performed in the following steps:

1. choose the number of bootstrap samples for derivation of the confidence intervals (n_{CI}) and the number of bootstrap samples for the assessment of the coverage probability (n_{CP}); in our study we set $n_{\text{CI}} = n_{\text{CP}} = 500$;
2. draw a sample of length l with replacement from the original series of annual erosivity and denote this sample s ;
3. draw a sample of length l with replacement from s and use it to calculate the average erosivity (i.e. the R factor);
4. repeat the previous step n_{CI} times;
5. calculate the 90 % confidence interval from the n_{CI} estimates of the R factor from step 3 and check whether this interval includes the true R factor;
6. repeat steps 2–5 n_{CP} times;
7. calculate the coverage probability associated with the record length l as the proportion of cases when the confidence interval from step 5 included the true R factor;
8. repeat the whole process for different record lengths l .

Please note that the described procedure provides n_{CP} estimates of the confidence intervals for specific l , and only their average is presented in the paper. Setting $n_{\text{CP}} = 1$ might be sufficient in the situation that only the confidence intervals would be of interest.

Appendix C: A resampling scheme for the assessment of spatial and temporal coverage

The following scheme describes a nested bootstrap procedure for assessment of the RMSE, coverage probability, and the width of the 90 % confidence interval for a GLS model considering different lengths of the precipitation data and different numbers of stations. The assessment is based on resampling of a synthetic dataset of 100 years for 120 stations (denoted the “full dataset”) and validated against the independent “validation dataset” (see Sect. 3.3.2). The procedure is summarized as follows:

1. draw a sample of n_{yr} years for n_{sta} stations from the “full dataset” and calculate the R factor for each station;
2. fit a GLS model of the form $R \sim \text{NAVY} + r_{\text{mea}} + Y$ using the sample from the previous step;
3. simulate data for the given n_{sta} stations from the fitted model (see e.g. Pinheiro and Bates, 2000);
4. refit the model and use this refitted model to predict the R factor for the “validation dataset”;
5. repeat the previous two steps 500 times;
6. calculate the 90 % confidence interval around the estimated R factor and the RMSE for each station of the “validation dataset” from the 500 samples obtained in steps 2–5;
7. repeat the previous steps (1–6) 500 times;
8. repeat the whole procedure for different n_{yr} and n_{sta} .

The RMSE, confidence interval, and coverage probability for the “areal-average” model used for comparison were derived by replacing the estimates from the refitted model in step 4 by the areal average of the simulated data from step 3.

The Supplement related to this article is available online at doi:10.5194/hess-20-4307-2016-supplement.

Acknowledgements. The research has been conducted within the framework of the project “Erosion runoff – increased risk of the residents and the water quality exposure in the context of the expected climate change” (VG20122015092) sponsored by the Ministry of the Interior of the Czech Republic. Data have been kindly provided by the Czech Hydrometeorological Institute. All calculations and plotting were done in R.

Edited by: M. Mikos

Reviewed by: three anonymous referees

References

- Angulo-Martínez, M., López-Vicente, M., Vicente-Serrano, S. M., and Beguería, S.: Mapping rainfall erosivity at a regional scale: a comparison of interpolation methods in the Ebro Basin (NE Spain), *Hydrol. Earth Syst. Sci.*, 13, 1907–1920, doi:10.5194/hess-13-1907-2009, 2009.
- Bonilla, C. A. and Vidal, K. L.: Rainfall erosivity in central Chile, *J. Hydrol.*, 410, 126–133, 2011.
- Borrelli, P., Diodato, N., and Panagos, P.: Rainfall erosivity in Italy: a national scale spatio-temporal assessment, *Int. J. Digit. Earth*, 9, 835–850, 2016.
- Brádka, J.: Srážky na území ČSSR při jednotlivých typech povětrnostní situace, Hydrometeorological Institute, Prague, 1972.
- Brázdil, R.: Vliv atlantského oceánu a středozemního moře na srážkové poměry léta na území ČSSR, J. E. Purkyne University, Brno, 1980.
- Brown, L. and Foster, G.: Storm erosivity using idealized intensity distributions, *T. ASAE*, 30, 0379–0386, 1987.
- Capra, A., Porto, P., and La Spada, C.: Long-term variation of rainfall erosivity in Calabria (Southern Italy), *Theor. Appl. Climatol.*, doi:10.1007/s00704-015-1697-2, in press, 2015.
- Carter, C., Greer, J., Braud, H., and Floyd, J.: Raindrop characteristics in South Central United States, *T. ASAE*, 17, 1033–1037, 1974.
- Catari, G., Latron, J., and Gallart, F.: Assessing the sources of uncertainty associated with the calculation of rainfall kinetic energy and erosivity – application to the Upper Llobregat Basin, NE Spain, *Hydrol. Earth Syst. Sci.*, 15, 679–688, doi:10.5194/hess-15-679-2011, 2011.
- Coutinho, M. A. and Tomás, P. P.: Characterization of raindrop size distributions at the Vale Formoso Experimental Erosion Center, *Catena*, 25, 187–197, 1995.
- Goovaerts, P.: Geostatistics for natural resources evaluation, in: Applied geostatistics series, Oxford University Press, Incorporated, Oxford, 1997.
- Goovaerts, P.: Using elevation to aid the geostatistical mapping of rainfall erosivity, *Catena*, 34, 227–242, 1999.
- Hanel, M. and Buishand, T. A.: Assessment of the sources of variation in changes of precipitation characteristics over the Rhine basin using a linear mixed-effects model, *J. Climate*, 28, 6903–6919, 2015.
- Hanel, M. and Máca, P.: Spatial variability and interdependence of rain event characteristics in the Czech Republic, *Hydrol. Process.*, 28, 2929–2944, 2014.
- Hanel, M., Pavlásková, A., and Kyselý, J.: Trends in characteristics of sub-daily heavy precipitation and rainfall erosivity in the Czech Republic, *Int. J. Climatol.*, 30, 1833–1845, 2016.
- Haskard, A.: Anisotropic Matérn spatial covariance model: REML estimation and properties, University of Adelaide, Adelaide, 2007.
- Hengl, T., Heuvelink, G., and Stein, A.: A generic framework for spatial prediction of soil variables based on regression-kriging, *Geoderma*, 120, 75–93, 2004.
- Hengl, T., Heuvelink, G., and Rossiter, D.: About regression-kriging: From equations to case studies, *Comput. Geosci.*, 33, 1301–1315, 2007.
- Janeček, M., Kubátová, E., and Tipl, M.: Revised determination of the rainfall-runoff erosivity factor R for application of USLE in the Czech Republic, *Soil Water Res.*, 1, 65–71, 2006.
- Janeček, M., Bečvář, M., Bohuslávka, J., Dufková, J., Dumbrovský, M., Dostál, T., Hula, J., Kadlec, V., Krása, J., Kubátová, E., Novotný, I., Podhrázská, J., Tipl, M., Toman, F., Vopravil, J., and Vrána, K.: Guidelines on protection of agricultural land against soil erosion, Research Institute of Agricultural Engineering, Prague, 2007.
- Janeček, M., Dostál, T., Kozlovsky Dufková, J., Dumbrovský, M., Hula, J., Kadlec, V., Kovář, P., Krása, J., Kubátová, E., Kobzová, D., Kudrnáčová, M., Novotný, I., Podhrázská, J., Pražan, J., Procházková, E., Středová, H., Toman, F., Vopravil, J., and Vlasák, J.: Guidelines on protection of agricultural land against soil erosion, Czech University of Life Sciences, Prague, 2012a.
- Janeček, M., Květoň, V., Kubátová, E., and Kobzová, D.: Differentiation and regionalization of rainfall erosivity factor values in the Czech Republic, *Soil Water Res.*, 7, 1–9, 2012b.
- Janeček, M., Květoň, V., Kubátová, E., Kobzová, D., Vošmerová, M., and Chlupsová, J.: Values of rainfall erosivity factor for the Czech Republic, *J. Hydrol. Hydromech.*, 61, 97–102, 2013.
- Kinnell, P.: Rainfall intensity-kinetic energy relationships for soil loss prediction, *Soil Sci. Soc. Am. J.*, 45, 153–155, 1981.
- Kitanidis, P.: Generalized covariance functions in estimation, *Math. Geol.*, 25, 525–540, 1993.
- Krása, J., Středová, H., Dostál, T., and Novotný, I.: Rainfall erosivity research on the territory of the Czech Republic, in: Mendel a bioklimatologie, <http://www.cbks.cz/SbornikBrno14/Krasa.pdf> (last access: 20 October 2016), 2014.
- Krása, J., Středová, H., Štěpánek, P., Hanel, M., Dostál, T., and Novotný, I.: Recent and future rainfall erosivity on the territory of the Czech Republic, EGU General Assembly Conference Abstracts, vol. 17, p. 7714, 2015.
- Květoň, V., Zahradníček, J., and Žák, M.: Quality control and digitising of pluviographic measurements in the Czech Hydrometeorological Institute, *Meteorologické zprávy*, 57, 47–52, 2004.
- Laws, J. O. and Parsons, D. A.: The relation of raindrop-size to intensity, *Eos T. Am. Geophys. U.*, 24, 452–460, 1943.
- Lee, M.-H. and Lin, H.-H.: Evaluation of Annual Rainfall Erosivity Index Based on Daily, Monthly, and Annual Precipitation Data of Rainfall Station Network in Southern Taiwan, *Int. J. Distrib. Sensor Netw.*, 11, 1–15, doi:10.1155/2015/214708, 2014.
- Lu, H. and Yu, B.: Spatial and seasonal distribution of rainfall erosivity in Australia, *Soil Res.*, 40, 887–901, 2002.

- McGregor, K., Bingner, R., Bowie, A., and Foster, G.: Erosivity index values for northern Mississippi, *T. ASAE*, 38, 1039–1047, 1995.
- McIsaac, G.: Apparent geographic and atmospheric influences on raindrop sizes and rainfall kinetic energy, *J. Soil Water Conserv.*, 45, 663–666, 1990.
- Meddi, M., Toumi, S., and Assani, A. A.: Spatial and temporal variability of the rainfall erosivity factor in Northern Algeria, *Arabian J. Geosci.*, 9, 1–13, 2016.
- Meusburger, K., Steel, A., Panagos, P., Montanarella, L., and Alewell, C.: Spatial and temporal variability of rainfall erosivity factor for Switzerland, *Hydrol. Earth Syst. Sci.*, 16, 167–177, doi:10.5194/hess-16-167-2012, 2012.
- Minasny, B. and McBratney, A. B.: The Matérn function as a general model for soil variograms, *Geoderma*, 128, 192–207, 2005.
- Minasny, B. and McBratney, A. B.: Spatial prediction of soil properties using EBLUP with the Matern covariance function, *Geoderma*, 140, 324–336, 2007.
- Oliveira, P. T. S., Wendland, E., and Nearing, M. A.: Rainfall erosivity in Brazil: A review, *Catena*, 100, 139–147, 2013.
- Panagos, P., Ballabio, C., Borrelli, P., Meusburger, K., Klik, A., Rousseva, S., Tadić, M. P., Michaelides, S., Hrabalíková, M., Olsen, P., Aalto, J., Lakatos, M., Rymaszewicz, A., Dumitrescu, A., Beguería, S., and Alewell, C.: Rainfall erosivity in Europe, *Sci. Total Environ.*, 511, 801–814, 2015.
- Panagos, P., Ballabio, C., Borrelli, P., and Meusburger, K.: Spatio-temporal analysis of rainfall erosivity and erosivity density in Greece, *Catena*, 137, 161–172, 2016a.
- Panagos, P., Borrelli, P., Spinoni, J., Ballabio, C., Meusburger, K., Beguería, S., Klik, A., Michaelides, S., Petan, S., Hrabalíková, M., Olsen, P., Aalto, J., Lakatos, M., Rymaszewicz, A., Dumitrescu, A., Tadić, M. P., Diodato, N., Kostalova, J., Rousseva, S., Banasik, K., and Alewell, C.: Monthly Rainfall Erosivity: Conversion Factors for Different Time Resolutions and Regional Assessments, *Water*, 8, 119, doi:10.3390/w8040119, 2016b.
- Pardo-Iguzquiza, E. and Chica-Olmo, M.: Geostatistics with the Matern semivariogram model: A library of computer programs for inference, kriging and simulation, *Comput. Geosci.*, 34, 1073–1079, 2008.
- Pinheiro, J. C. and Bates, D. M.: *Linear Mixed-Effects Models: Basic Concepts and Examples*, Springer, New York, NY, USA, 2000.
- Porto, P.: Exploring the effect of different time resolutions to calculate the rainfall erosivity factor R in Calabria, southern Italy, *Hydrol. Process.*, 30, 1551–1562, 2016.
- Renard, K. G., Foster, G. R., Weesies, G. A., McCool, D., and Yoder, D.: *Predicting soil erosion by water: a guide to conservation planning with the revised universal soil loss equation (RUSLE)*, Agriculture Handbook 703, US Government Printing Office, Washington, D.C., USA, 1997.
- Rosewell, C. J.: Rainfall kinetic energy in eastern Australia, *J. Clim. Appl. Meteorol.*, 25, 1695–1701, 1986.
- Štekl, J., Brázdil, R., Kakos, V., Jež, J., Tolasz, R., and Sokol, Z.: Extreme daily precipitation on the territory of the Czech Republic in the period 1879–2000 and their synoptic causes, *National Climatic Programme of the Czech Republic*, 31, 140, 2001.
- Svoboda, V., Hanel, M., Máca, P., and Kyselý, J.: Characteristics of rainfall events in RCM simulations for the Czech Republic, *Hydrol. Earth Syst. Sci. Discuss.*, doi:10.5194/hess-2016-283, 2016a.
- Svoboda, V., Hanel, M., Máca, P., and Kyselý, J.: Projected changes of rainfall event characteristics for the Czech Republic, *J. Hydrol. Hydromech.*, 64, 415–425, doi:10.1515/johh-2016-0036, 2016b.
- Tolasz, R., Brázdil, R., Bulfř, O., Dobrovolný, P., Dubrovský, M., Hájková, L., Halášová, O., Hostýnek, J., Janouch, M., Kohut, M., Kveton, V., Lepka, Z., Lipina, P., Macková, J., Metelka, L., Míková, T., Mrkvica, Z., Močný, M., Nekovář, J., Němec, L., Pokorný, J., Reitschläger, D., Richterová, D., Rožnovský, J., Řepka, M., Semerádová, D., Sosna, V., Stríž, M., Šercl, P., Škáchová, H., Štěpánek, P., Štěpánková, P., Trnka, M., Valeriánová, A., Valter, J., Vaníček, K., Vavruška, F., Voženílek, V., Vráblík, T., Vysoudil, M., Zahradníček, J., Zusková, I., Žák, M., and Žalud, Z.: *Atlas podnebí Česka (Climate atlas of the Czech Republic)*, Český hydrometeorologický ústav, Universita Palackého, Palackého, 2007.
- Van der Knijff, J., Jones, R., and Montanarella, L.: *Soil erosion risk assessment in Europe*, European Soil Bureau, Joint Research Centre, Ispra, Italy, 2000.
- van Dijk, A., Bruijnzeel, L., and Rosewell, C.: Rainfall intensity-kinetic energy relationships: a critical literature appraisal, *J. Hydrol.*, 261, 1–23, 2002.
- Verstraeten, G., Poesen, J., Demarée, G., and Salles, C.: 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, *J. Geophys. Res.-Atmos.*, 111, 1–11, 2006.
- Willmott, C. and Matsuura, K.: Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance, *Clim. Res.*, 30, 79–82, 2005.
- Wischmeier, W. H. and Smith, D. D.: *Predicting rainfall erosion losses – A guide to conservation planning*, Agriculture Handbook 537, US Department of Agriculture, Maryland, USA, 1978.
- Yin, S., Xie, Y., Nearing, M., and Wang, C.: Estimation of rainfall erosivity using 5- to 60-minute fixed-interval rainfall data from China, *Catena*, 70, 306–312, 2007.
- Yip, S., Ferro, C. A., Stephenson, D. B., and Hawkins, E.: A simple, coherent framework for partitioning uncertainty in climate predictions, *J. Climate*, 24, 4634–4643, 2011.