Dieses Dokument ist eine Zweitveröffentlichung (Verlagsversion) /

This is a self-archiving document (published version):

Philipp Goebes, Karsten Schmidt, Werner Härdtle, Steffen Seitz, Felix Stumpf, Goddert von Oheimb, Thomas Scholten

Rule-based analysis of throughfall kinetic energy to evaluate biotic and abiotic factor thresholds to mitigate erosive power

Diese Version ist verfügbar / This version is available on:

https://nbn-resolving.org/urn:nbn:de:bsz:14-qucosa2-353826

SLUB
Wir führen Wissen.

TECHNISCHE UNIVERSITÄT DRESDEN

Qucosa
Quality Content of Saxony

Article

Rule-based analysis of throughfall kinetic energy to evaluate biotic and abiotic factor thresholds to mitigate erosive power

Philipp Goebes
University of Tübingen, Germany

Karsten Schmidt
University of Tübingen, Germany

Werner Härdtle
Leuphana University Lüneburg, Germany

Steffen Seitz
University of Tübingen, Germany

Felix Stumpf
University of Tübingen, Germany

Goddert von Oheimb
Technische Universität Dresden, Germany

Thomas Scholten
University of Tübingen, Germany

Abstract
Below vegetation, throughfall kinetic energy (TKE) is an important factor to express the potential of rainfall to detach soil particles and thus for predicting soil erosion rates. TKE is affected by many biotic (e.g. tree height, leaf area index) and abiotic (e.g. throughfall amount) factors because of changes in rain drop size and velocity. However, studies modelling TKE with a high number of those factors are lacking.

This study presents a new approach to model TKE. We used 20 biotic and abiotic factors to evaluate thresholds of those factors that can mitigate TKE and thus decrease soil erosion. Using these thresholds, an optimal set of biotic and abiotic factors was identified to minimize TKE. The model approach combined recursive feature elimination, random forest (RF) variable importance and classification and regression trees (CARTs). TKE was determined using 1405 splash cup measurements during five rainfall events in a subtropical Chinese tree plantation with five-year-old trees in 2013.

Corresponding author:
Philipp Goebes, Department of Geosciences, University of Tübingen, Rümelinstraße 19–23, 72070 Tübingen, Germany.
Email: philipp.goebes@uni-tuebingen.de

Our results showed that leaf area, tree height, leaf area index and crown area are the most prominent vegetation traits to model TKE. To reduce TKE, the optimal set of biotic and abiotic factors was a leaf area lower than 6700 mm^2, a tree height lower than 290 cm combined with a crown base height lower than 60 cm, a leaf area index smaller than 1, more than 47 branches per tree and using single tree species neighbourhoods. Rainfall characteristics, such as amount and duration, further classified high or low TKE. These findings are important for the establishment of forest plantations that aim to minimize soil erosion in young succession stages using TKE modelling.

I Introduction

Soil erosion by water is a major threat to natural ecosystems and agricultural land in many regions of the world (Cao et al., 2013; Cerdá et al., 2009; Lieskovský and Kenderessy, 2014; Seutloali and Beckedahl, 2015). Besides slope, slope length, soil erodibility and vegetation, rainfall erosivity is another important driver in predicting soil erosion rates by empirical (Renard et al., 1997) or process-based models (Morgan et al., 1998). Higher rainfall and rainfall erosivity are negatively related to soil conservation and thus soils can lose important ecosystem services, e.g. filtering water (Keestra et al., 2012), secure food production and plant diversity (Brevik et al., 2015), while conversely plant diversity can also affect soil conservation (Berendse et al., 2015). Rainfall erosivity is most commonly expressed by the EI_{30}, which combines rainfall energy (E) and rainfall intensity per 30 minute interval (I_{30}). While there are numerous studies investigating rainfall intensity and related processes (van Dijk et al., 2002), research on the determining processes of rainfall energy is limited. Few studies deal with the discussion of a proper erosivity index of rainfall energy (Goebes et al., 2014), while others investigate seasonal and temporal trends of rainfall energy (Nunes et al., 2014; Taguas et al., 2013). This lack of studies is particularly true when rainfall energy is examined below tree canopies as throughfall kinetic energy (TKE). Here, the size distribution of rain drops is changed because of biotic factors (e.g. leaf traits), potentially resulting in higher TKE than rainfall energy at open field sites (Geißler et al., 2010; Geißler et al., 2012; Nanko et al., 2004; Nanko et al., 2015). In addition, rain drop size is positively related to rainfall intensity (Cerdá, 1997). This strengthens the influence of TKE on inducing soil erosion processes below tree canopies. Hence, if a litter cover at the soil surface is missing, TKE is directly influencing soil erosion (Seitz et al., 2015) indicating the definite role of vegetation for soil erosion control (Cerdá, 1998).

Reflecting the relevance of TKE for soil erosion, TKE has been measured in different regions, under different rainfall conditions and below different vegetation in the past 15 years (Nanko, 2007; Nanko et al., 2008, 2011; Sanchez-Moreno et al., 2012; Zhou et al., 2002). In addition, several studies investigated the influence of biotic (single leaf and tree architectural traits) and abiotic factors (rainfall characteristics) on TKE separately. For instance, a positive effect on TKE has been reported for leaf area (Goebes et al., 2015a), tree height (Foot and Morgan, 2005; Geißler et al., 2013), crown area (Brandt, 1988; Nanko et al., 2008), crown base height (Brandt, 1990; Nanko et al., 2008) and throughfall amount (Brandt, 1988; Geißler et al., 2012; Scholten et al., 2011). TKE is negatively influenced by leaf area index (LAI) (Nanko et al., 2006; Nanko et al., 2008) and the number of branches (Herwitz, 1987).

In addition, TKE shows spatial variability (Finney, 1984; Nanko et al., 2011) and deciduous tree species can cause higher TKE than evergreens (Goebes et al., 2015a).

There are some studies that modelled TKE with biotic and abiotic factors to evaluate its role in erosion processes. However, these studies are limited in their number of biotic and abiotic factors. For instance, Moss and Green (1987) reported a maximum crown base height of 30 cm below which TKE is non-erosive. Brandt (1990) developed a model incorporating tree height as the most important vegetation variable while Calder (1996) used interception processes to model TKE by evaluating the drop size distribution. Foot and Morgan (2005) suggested to model TKE by only using tree height and canopy area. Type and intensity of a rainfall event determine whether TKE is erosive or not (Brandt, 1989; Zhou et al., 2002). Furthermore, several studies used modelling approaches to determine the role of rainfall kinetic energy in soil erosion at open sites in different regions of the world (Assouline, 2009; Assouline and Mualem, 1989; Salles and Poesen, 2000; van Dijk et al., 2002). As a consequence, literature on modelling TKE patterns and potential thresholds for a variety of biotic and abiotic factors in the context of erosivity are scarce. It also remains unclear if thresholds exist for biotic and abiotic factors that lead to a specific TKE. This motivated us to model TKE by using a variety of biotic and abiotic predictor variables to clarify their influence, interaction and importance. This, in turn, helps to better understand mechanisms that underlie and mediate soil erosion processes.

In the past decades, statistical and machine-learning methodologies have made huge progress. Random forest (RF) is such a machine-learning technique, representing an ensemble of randomized classification and regression trees (CART) (Breiman, 2001). The final estimation is derived by aggregating the individual trees. A single CART uses a set of binary rules to compute a target variable. The binary rules are based on independent variables and the observed response variable (Breiman et al., 1984). In RF, estimations are derived from multiple CART-like trees, adapted by using randomized subsets of the input data (Grimm et al., 2008). As a consequence, RF is increasingly applied in ecological studies. Peters et al. (2007) estimated the occurrence of vegetation types, while Kuz'min et al. (2011) estimated aquatic toxicity. With regard to soil erosion research, Märker et al. (2011) used RF to model erosional response units and to identify major controlling factors of soil erosion. While RF provides a variable importance measure, the estimations exhibit limited interpretability. Since in RF the final estimation is derived from aggregated results of multiple decision tree models, the relation between predictors and estimations cannot be assessed easily. This can, however, be accomplished by single CART models (Breiman et al., 1984; Cutler et al., 2007).

In this study we propose a step-wise decision tree approach to establish a rule-based system for estimating TKE. We combined the RF feature importance measure and recursive feature elimination (RFE) to determine a feature subset as input for estimating TKE using a single CART modelling approach. Subsequently, we analysed the CART with regard to biotic and abiotic factors to detect erosion-relevant thresholds of those factors in the context of TKE. We used this methodological frame to evaluate three objectives:

1. To describe and model TKE with a distinct set of biotic and abiotic factors.
2. To identify relevant biotic and abiotic factor thresholds for predicting TKE in order to find an optimal predictor subset that minimizes TKE.
3. To evaluate those predictions using a literature comparison.

Table 1. Rainfall characteristics of five rainfall events. Rainfall amount (R_A), intensity (I) and duration (D) were measured at the climate station of BEF-China using a tipping bucket. Mean throughfall (TF) was measured at each TKE measurement position using rainfall gauges.

Rainfall events	R_A (mm)	D (h)	I (5 min peak intensity, mm h^{-1})	I (total event, mm h^{-1})	TF (mm)	Mean TKE [J m^{-2} mm^{-1}] (standard deviation)
Event 1	23.3	10.16	12.1	2.29	28.3	11.00 (7.90)
Event 2	39.3	11.50	22.8	3.42	47.9	9.02 (7.66)
Event 3	61.2	14.50	44.4	4.25	73.8	9.05 (5.15)
Event 4	6.6	2.33	25.2	2.83	5.0	11.93 (8.36)
Event 5	185.7	30.58	127.2	6.07	192.7	6.96 (3.23)

II Data collection and modelling

1 Study site and experimental design

The study was conducted within the framework of the large-scale biodiversity-ecosystem functioning experiment 'BEF-China' (Bruelheide et al., 2014) at Xingangshan, Jiangxi Province, PR China (N29°08-11, E117°90-93). The climate in Xingangshan is typical of subtropical summer monsoon regions with a mean annual temperature of 17.4°C and an average annual rainfall of 1635 mm. The experimental area holds 70 ha with a plot-based tree diversity treatment including 24 tree species on 261 plots. Tree individuals were planted after harvest of the previous stand in 2009 and they were five years old at the time of TKE measurements. For this study, 40 plots were selected at random, including 17 monocultures, 10 2-species mixtures, six 4-species mixtures, four 8-species mixtures, one 16-species mixture and two 24-species mixtures to cover a wide range of different species richness levels and compositions. Within one plot, eight measurements were taken by selecting eight different positions in order to cover a wide range of spatial variability (Goebes et al., 2015b). Positions (1), (4), (6) and (8) were influenced by one tree individual (1, 15 cm from the stem; 4, 45 cm from the stem; 6, first branch; 8, 30 cm from the stem), (2), (5) and (7) were influenced by two tree individuals (2, middle of two; 5, 45 × 120 cm intersection; 7, 75 × 75 cm intersection) and (3) was influenced by four tree individuals (3, middle of four).

2 Measurement of TKE and rainfall

TKE was measured using Tübingen Splash Cups (Scholten et al., 2011) filled with uniform fine sand (diameter 0.125 mm). Sand loss in grams (ds) in splash cups (sc) was used to calculate TKE (standardized by gross rainfall; J m^{-2} mm^{-1}) by the function given by Scholten et al. (2011) with a modified slope, a correction to 1 m^2 and the gross rainfall amount in mm (rf) of each rainfall event

$$KE_{rainfall}\left[\frac{J}{m^2 mm}\right] = \frac{ds_{sc}[g] * 0.1455 * \left(\frac{10,000 cm^2}{\pi r_{sc}^2}\right)}{rf_{event}}.$$

(1)

In total, 1600 splash cups were measured during five rainfall events from May to July 2013. Table 1 shows rainfall characteristics. These rainfall events covered a broad range of all rainfall events. In 2013, our climate station registered 33 erosive events (Renard et al., 1997; Wischmeier and Smith, 1978) ranging from 13 mm to 185 mm with a total rainfall amount of 1205 mm. In 2012, 49 erosive events ranging from 13 mm to 211 mm were measured. Mean rainfall amount per event was 40 mm in 2012 and 30 mm in 2013.

By reviewing literature on TKE measurements (measured in J m^{-2} mm^{-1}) of the past

Table 2. Mean, minimum and maximum throughfall kinetic energy (TKE in J m^{-2} mm^{-1}) measured in different studies. Rainfall characteristics show amount of annual precipitation or simulated rainfall intensity and type of rainfall. Abbreviations: TF = throughfall, FF = freefall, art = artificial, SD = standard deviation.

Study	Rainfall characteristics of either study site or experiment	Mean	Min	Max
Sanchez-Moreno et al., 2012	300–500 mm, FF	∼21	4	70
Nanko, 2007	2300 mm, TF+(FF)	27 (11)	23.8	31.2
Finney, 1984	61 mm h^{-1} (art.), TF	7	0.4	10.5
Brandt, 1987	n/a mm, TF	21.8	3	40
Nanko et al., 2008	39.8 mm h^{-1} (art.), TF+(FF)	17.5 (12.7)	15.9	20.7
Nanko et al., 2011	40 and 85 mm h^{-1} (art.), TF	16.2	11.8	21.2
van Dijk et al., 2002	0.4–372 mm h^{-1}, FF	21	3.4	36.8
Brandt, 1988	2478 mm, TF+(FF)	27 (18)	13.6	40.2
Zhou et al., 2002	1454 mm, TF	28	21	33
All nine studies combined		20.7 (SD 6.6)	0.4	70
Present study	**1642 mm a^{-1}, TF**	**9.6**	**0.3**	**54.8**

30 years (Table 2), and classifying those results into four different categories using k-means clustering with 1000 iterations (MacQueen, 1967), we evaluated our TKE measurements according to these categories. The cluster means appeared in a multiplicative way using standard deviations (SD) from the mean TKE across all studies (20.7 J m^{-2} mm^{-1}, Table 2). Thus, category 1 was calculated by subtracting 2 SD from mean (hereafter referred to as low TKE, range = 0–11.3 and mean = 7.5), category 2 by subtracting 1 SD from mean (moderate TKE, range = 11.3–17.4 and mean = 14.1), category 3 by representing the mean (average TKE, range = 17.5–24.0 and mean = 20.7) and category 4 by adding 1 SD to the mean (high TKE, range = 24.1–70 and mean = 27.3). The studies cover a wide range of rainfall amounts (300–2478 mm a^{-1}) and intensities (0.4–372 mm h^{-1}). They confirm that rainfall characteristics of our study (rainfall amount of 1635 mm a^{-1} and intensities of 12–127 mm h^{-1}) are close to the mean of the literature review and thus can be considered representative. This allows the comparison and categorization of our TKE measurements to the categories resulting from the literature review.

3 Measurement of biotic and abiotic factors

With regard to biotic factors, plot-level diversity was evaluated based on the experimental design. Neighbourhood diversity was specified by the composition of direct neighbouring tree individuals of a measurement position. In addition, we used the binary contrast mono-mixture to differ between monoculture plots and mixture plots. Tree height, LAI, crown area, ground coverage, number of branches, ground diameter, crown base height, leaf habit (deciduous, evergreen and in mixtures both), leaf area (mean leaf area per one leaf of one species) and specific leaf area (Goebes et al., 2015b; Kröber et al., 2014; Kröber and Bruelheide, 2014; Li et al., 2014) were measured as biotic factors.

As abiotic factors, we measured throughfall at each TKE measurement position using rainfall gauges. The number of individuals was determined by counting direct tree neighbours that were influencing one splash cup. Spatial variability was assessed using the different positions of the sampling design. All splash cup positions were covered by vegetation. If a splash cup was influenced by more than one tree individual, mean values of biotic factors of the respective tree individuals were used. Tree

Table 3. Predictors used as independent variables in the CART models. Mean values (and standard deviation, SD) were calculated using all five rainfall events. c = categorical variable and n = numerical variable.

	Indicators	Abbr.	Mean (SD)	Min	Max	Unit
Biotic factors	Tree species richness (c and n)	A, B	4.00 (5.63)	1	24	–
	Neighbourhood tree species richness	C	1.24 (0.55)	1	4	–
	Mono-mixture contrast	D	–			–
	Tree height	E	271.60 (156.42)	30	831	cm
	Leaf area index (LAI)	F	1.43 (1.07)	0.02	4.56	–
	Crown area	G	24,132 (26,462)	192	173,590	cm^2
	Ground coverage	H	0.62 (0.28)	0.01	0.99	–
	Number of branches	I	21 (16)	1	110	–
	Ground diameter	J	3.92 (2.03)	0.65	12.60	cm
	Crown base height	K	56.74 (75.10)	0.5	603.5	cm
	Leaf habit (deciduous vs evergreen)	L	–			–
	Leaf area	M	13,898 (13,214)	1121	37,038	mm^2
	Specific leaf area	N	11.61 (1.27)	8.61	15.23	g mm^{-2}
Abiotic factors	Throughfall amount	O	69.55 (73.97)	0.8	303.5	mm
	Position	P	–			–
	Number of individuals (c and n)	Q, R	1.73	1	4	–
	Rainfall event intensity	S	46.3 (46.7)	12.1	127.2	mm h^{-1}
	Rainfall event duration	T	13.8 (10.4)	2.3	30.6	h

species richness and the number of individuals were included as categorical and continuous predictors to avoid under parameterization of categorical predictors. Altogether we used a set of five categorical and 15 continuous predictors to model TKE (Table 3).

4 Data modelling

Leaf and tree architectural thresholds on which TKE was evaluated were finally derived by using CART. Instead of pruning the final CART, we decided to use RFE followed by variable importance selection of RF to decrease the number of input variables before the construction of the final CART. This (i) allows us to reduce noise in the CART if we exclude less important features prior to the CART, (ii) enables a rule-based interpretation of the constructed trees and (iii) limits over-fitting. For instance, noise could be reduced as a result of exclusion of unimportant input variables if a very large number of uninformative predictors

were collected and one such predictor would randomly correlate with the outcome.

Recursive feature elimination. RFE with incorporated resampling was used to identify model performance related to the numbers of input variables (Kuhn, 2014). The model approach is based on the following steps (Kuhn, 2014):

1. Split data in training and validation set.
2. Train the model on the training set using all predictors.
3. Calculate model performance.
4. Calculate variable importance.
5. For each subset size S_i, i = 1 . . . S do.
6. Keep the S_i most important variables.
7. Train the model on the training set using Si predictors.
8. Calculate model performance.
9. Calculate the performance profile over all S_i.
10. Determine the appropriate number of predictors.

11. Determine the final ranks of each predictor.
12. Fit the final model based on the optimal S_i.

Variables occurring after the optimal input variable number were dismissed in the subsequent RF models.

This approach leads to a distinct number of input variables for CART. Therefore, it limits input variables in the final CART and simplifies subsequent rule-based model interpretation. However, RFE cannot give information on what the most important variables have been and thus a second approach is needed.

Variable importance using RFs. The variable importance of RFs was used to detect the most important variables. RFs are optimally suited to identify relevant features (Breiman, 2001) based on mean increased modelling performance (%IncMSE) via randomized feature and instance sampling. This is calculated by using the inherent structure of the RF approach as an ensemble of multiple decision trees where each individual tree is based on a bootstrap sample (random sampling with replacement; Efron and Tibshirani, 1994) of the data. Additionally, at each split only a random subset of all features is tested to find the parameter, which is best suited to further split the node (see Rule construction using CART).

All single trees are evaluated using the out-of-the-bag data. OOB is the portion of the data that is left out in each bootstrap replicate to build one tree of the ensemble. For the mean increased modelling performance each feature is randomly permuted at each split and the rate of change of the mean square error, compared with the original feature, is used as an indicator for its importance (Breiman, 2001; Grimm et al., 2008). This measure does not over-fit because it is tested against the independent OOB data (Prasad et al., 2006).

As a consequence, RF allows for analysis of non-parametric and non-linear effects and gives no need to transform data before modelling. They provide high prediction accuracy by fitting an ensemble of CARTs to a data set and combining the predictions from all CARTs (Cutler et al., 2007). The major drawback is that the resulting models are often black boxes and not able to obtain leaf and tree architectural thresholds for specific TKE measurements. Therefore, the variable importance of RF was only used to dismiss all input variables that do not lead to a better model performance based on the results of the RFE.

Rule construction using CART. Classification rules to evaluate biotic and abiotic factor thresholds on TKE were constructed using CART. CARTs build rules by splitting the continuous response into two groups (resulting in two nodes, which are the sample means of each group) by using an optimal threshold of a predictor (splitting) variable. The optimal split (threshold) is defined as the largest drop to reduce the residual sum of squares between the two groups of the target variable fitted with an ANOVA to the predictor evaluated at this split. The splitting process is iterated in a recursive way for each of the two sub-regions and for each of the predictor variables (Breiman et al., 1984). The vertical location of a predictor defines its importance in predicting the target variable TKE. CARTs were constructed using the ANOVA method. Because of the simplification of the model structure by dismissing none/or less relevant input variables, no tree pruning was applied.

Modelling setups and validation. We used TKE as dependent target variable and the variables listed in Table 3 as independent variables according to RFE and RF results. Six models were constructed for each approach: one model of each single rainfall event to obtain rainfall-specific TKE models and one model of all rainfall events to obtain rainfall-independent TKE models. Rainfall event
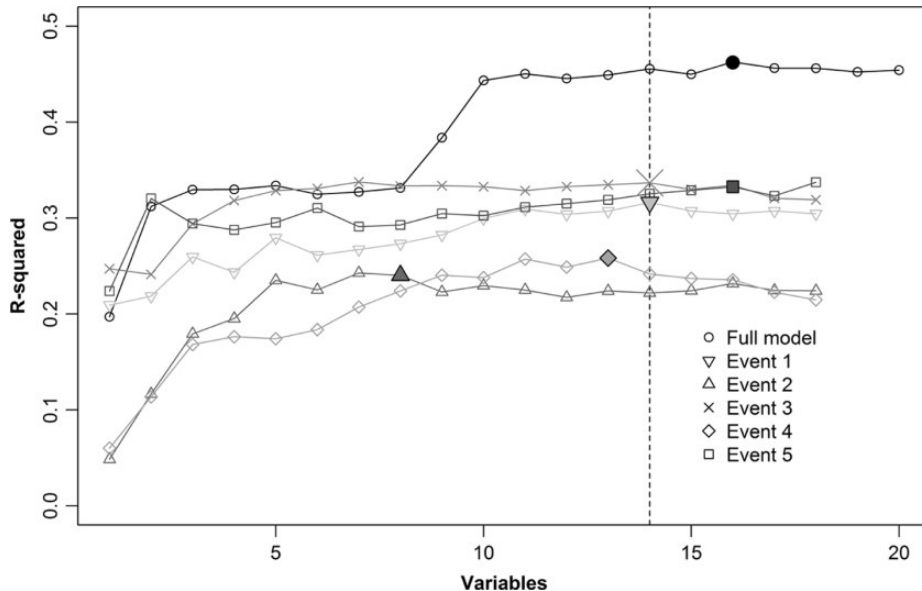
Figure 1. Results of the recursive feature elimination (RFE) with data from each event and combining all events (full model, which had two additional variables characterizing the rainfall event). Large symbols indicate the best variable set for each subset (five single events and one combining all events). Dashed line indicates the best variable set by calculating the weighted mean of all subsets.

intensity and duration were used as input variable only in the models constructed out of all rainfall events. Model performance of the RFE was evaluated using the root mean square error (RMSE) and the explained variance (R^2). To evaluate the optimal number of input variables based on RFE, we calculated the weighted mean of all six models (the model combining all rainfall events was double-weighted). We only used one number of dismissed variables so that every rainfall event was treated identically with the same number of input variables, resulting in equal CART starting positions considering tree growth and importance evaluation. This equal number of input variables allows a comparison between different models. Mean increased modelling performance (%Inc-MSE) was used to obtain the most important variables within the RFs. The number of randomly selected predictors to test at each node (m_{try}) and the number of instances/data points in the final node ($n_{nodesize}$) were tested with 1, 2, 3 and 4 and finally set to 3. We constructed 1500 trees per model using regression. Five-fold repeated 10-fold cross-validation was used to validate the CARTs by RMSE and R^2, as well as the model stability/robustness. All models were analysed using R 2.15.3 (R Core Team, 2013) with the packages randomForest (Liaw and Wiener, 2002) and rpart (Therneau et al., 2013) and were validated using the caret package (Kuhn, 2014).

III Results

RFE resulted in dismissing the least important four variables (mean of dismissed variables of the six models; Figure 1).

Variable importance of all input predictors of all single rainfall events and the model combining all rainfall events is shown in Figure 2. The least five important variables of each model were dismissed in further analysis. A detailed list of the dismissed variables may be found in Table A1.
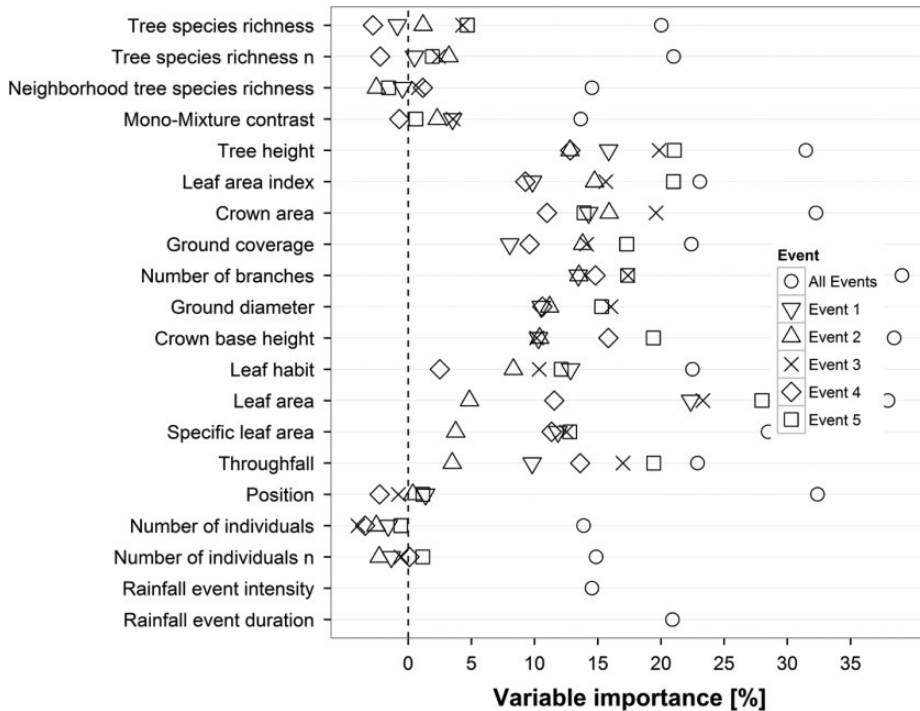
Figure 2. Variable importance (%IncMSE) of 20 biotic and abiotic factors on throughfall kinetic energy for six rainfall event models. For statistical descriptions of the factors, see Table 3.

The final CART model including all rainfall events is displayed in Figure 3 (Figures A1, A2, A3, A4 and A5 of rainfall event 1, 2, 3, 4 and 5, respectively; see Appendix). Considering non-standardized TKE, CART model performance was $R^2 = 0.65$, 0.45, 0.37, 0.46, 0.41 and 0.43 and RMSE = 32.0, 16.0, 25.7, 26.5, 4.9 and 52.7 for the model including all rainfall events and single rainfall events 1, 2, 3, 4 and 5, respectively. Considering standardized TKE, CART model performance was $R^2 = 0.30$, 0.27, 0.21, 0.32, 0.25 and 0.31 and RMSE = 6.09, 7.31, 7.30, 4.64, 7.85 and 2.83 for the model including all rainfall events and single rainfall events 1, 2, 3, 4 and 5, respectively.

Leaf area and throughfall amount occurred in all six CARTs. Tree height and LAI were second prominent with five times occurrence. Ground coverage, specific leaf area, ground diameter and neighbourhood diversity occurred

only once though. Leaf area was the most prominent variable in first splits. Throughfall amount, tree height and LAI were most prominent in second splits, while leaf area, throughfall amount, LAI, number of branches and crown base height were most prominent in third splits.

The thresholds of each biotic and abiotic predictor varied slightly between different rainfall events. Summarizing biotic and abiotic thresholds of CARTs of all single rainfall events and the CART including all rainfall events (for details see Figures 3 and A1–A5), leaf area showed prominent thresholds of approximately 35,000 mm^2 and 6,700 mm^2. Throughfall amount splits were found at 2.8 mm, 24 mm, 70 mm and 220 mm. Tree height showed prominent thresholds at 289 and 330 cm. Crown base height showed the most prominent thresholds at 60 cm. Thresholds for the number of branches were found at 14 and 47. LAI showed prominent
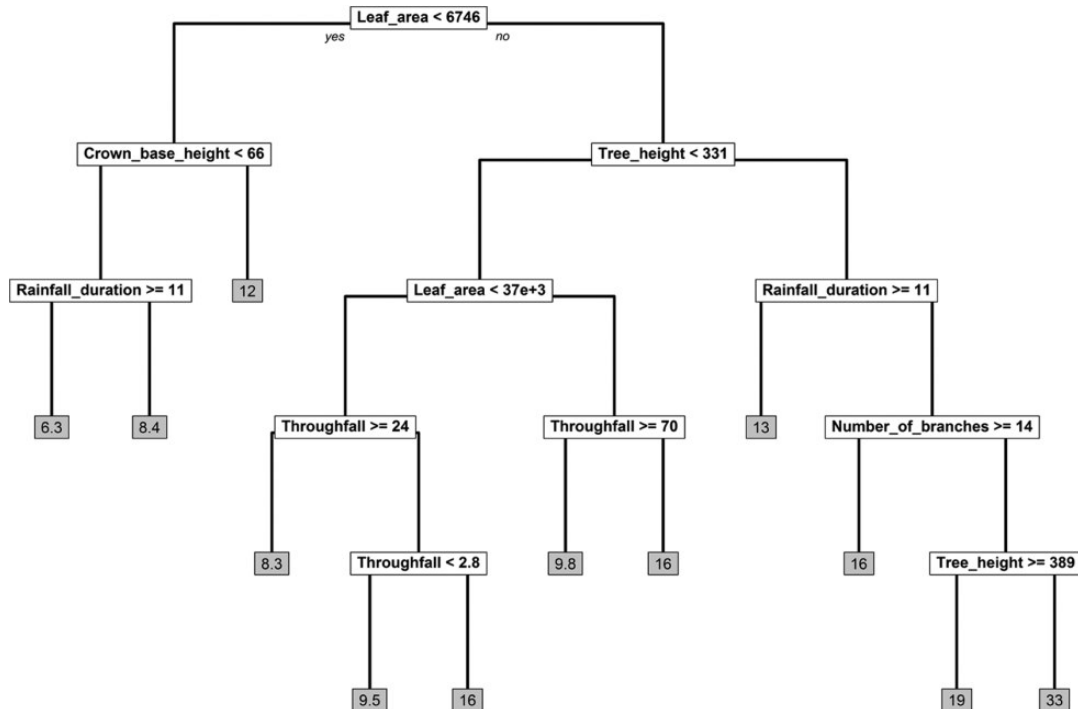
Figure 3. CART across all events. Target variable was throughfall kinetic energy (TKE) and predictor variables are listed in Table 3. TKE was measured as J m^2 mm^{-1} (n = 1405).

thresholds at 1 and 1.8, while crown area splits were found at 37,000 cm^2.

To monitor low TKE, thresholds were set by leaf area, throughfall and tree height as the most prominent variables. Leaf area, throughfall, LAI and crown area were most prominent in building splits to yield moderate TKE, while thresholds of leaf area, throughfall, crown area, number of branches and crown base height led to average TKE. High TKE was monitored with splits occurring by leaf area, throughfall and LAI.

The CART model including all rainfall events showed six different predictor variables, five split levels and 12 terminal nodes (Figure 3). Similarly, the CARTs of rainfall event 2 and 4 showed eight different variables, five split levels and 13 and 12 terminal nodes, respectively.

IV Discussion

We investigated the influence of 20 biotic and abiotic factors on TKE using a step-wise approach of RF and CART. We showed rules induced by those factors to obtain low, moderate, average or high TKE compared to nine studies which investigated TKE in different regions. Leaf area, throughfall, tree height and LAI affected TKE as most prominent variables in the CART models.

1 Ensemble approach using RF variable importance and CART to predict TKE

We detected effects of biotic and abiotic factors on TKE that are consistent with previous studies (objective 1). CARTs showed the influence of leaf area (Goebes et al., 2015a), throughfall

amount (Brandt, 1988; Geißler et al., 2012; Scholten et al., 2011), tree height (Foot and Morgan, 2005; Geißler et al., 2013), LAI (Nanko et al., 2006; Nanko et al., 2008), crown area (Brandt, 1988; Nanko et al., 2008), number of branches (Herwitz, 1987), crown base height (Brandt, 1990; Nanko et al., 2008), spatial variability (Finney, 1984; Nanko et al., 2011) and duration as well as intensity of rainfall event (Brandt, 1989; Zhou et al., 2002) on TKE. Furthermore, feature elimination and selection before using CART left no need for pruning or modifying the final trees. A typical pruned CART has 3–12 terminal nodes (Cutler et al., 2007), which was in the range of 9 to 14 terminal nodes in this study. Prediction results of $R^2 = 0.68$ for the non-standardized models emphasized the suitability of this approach. In addition, the approach was able to detect a non-linear effect of throughfall on TKE due to interactions with biotic factors such as leaf area.

2 Thresholds of biotic and abiotic factors to model TKE

In general, results obtained from data across all rainfall events can be found in the results of each rainfall event, though in less detail (Figures A1–A5). Thus, we used the CART that combined all rainfall events as a major source of interpretation in the following discussion. Since TKE was standardized using rainfall amount, rainfall duration was the major rainfall event characteristic that changed the optimal set of biotic and abiotic factors and their thresholds.

Leaf area was the most important predictor in our CARTs to describe different TKE. Leaf area was of major importance to yield low, moderate, average or high TKE. Leaf areas beyond 35,000 mm^2 caused average to high TKE whereas leaf areas below 6700 mm^2 led to low TKE (Figures 3, A1–A5). The latter size was most prominent for all species and showed that species with large leaf area cannot function as erosion

inhibitors. A higher leaf area might create a larger surface for rain drop gathering as well as confluence, and hence a release of larger rain drops (Herwitz, 1987). For instance, leaves of *Schima superba* (38,090 mm^2) increased sand loss in splash cups by 30% compared to leaves of *Castanopsis eyrei* (12,920 mm^2), which led to TKE (converted out of sand loss with a linear function by Scholten et al. (2011)) within the range of 1 SD of natural rainfall (Geißler et al., 2012). This shows that the erosion potential below vegetation can be distinctly reduced compared to that of natural rainfall using small leaf sizes.

Our study showed a non-linear effect of throughfall on TKE and thus contradicts a positive linear effect reported in previous studies (Brandt, 1988; Levia and Frost, 2006; Scholten et al., 2011). Throughfall amount as abiotic factor was second prominent to describe TKE differences, but with a positive and negative effect on TKE (see Figure 3). Our approach was particularly dedicated to investigate non-linear relationships that can be caused by interaction with other factors. In this case, throughfall mainly interacted with leaf area (see Figure 3). Throughfall amounts (see Figure A5) below 229 mm led to moderate TKE, whereas throughfall amounts higher or lower than 185 mm led to low TKE during high rainfall amounts per event. However, low throughfall amounts such as 2.8 mm can also lead to high, average or moderate TKE. It is likely that biotic factors emerged as a result of the standardization of TKE by rainfall amount at each event, suggesting the importance of interaction effects between biotic and abiotic factors with regard to TKE. The non-linear effect of throughfall on TKE was especially visible when data of all events entered the analyses (Figure 2).

A tree height below 290 cm resulted in low to moderate TKE (7.5–14.1 J m^{-2} mm^{-1}, Figures 3 and A3) because of shorter falling heights and, hence, reduced rain drop velocities (Gunn and Kinzer, 1949). This threshold led to TKE of

about 2 J m^{-2} mm^{-1}, which is below values reported by Brandt (1990) and Nanko et al. (2008). Brandt (1990) emphasized in her model that effects on TKE were more pronounced for tree height shifts of small trees. Figure 3 indicates that only tree heights above 389 cm led to high TKE, while lower heights (of about 60 cm) led to low to moderate TKE. This suggests that there is a 'critical tree height', at approximately 330 cm, above which TKE becomes highly erosive. However, this height is close to the mean of all species and indicates that young tree individuals in particular are non-erosive.

Crown base height was the fourth most important predictor of TKE. Rain drops falling from trees with crown base heights below 60 cm had a low to moderate TKE (Figures 3, A2 and A4). Moss and Green (1987) showed that the height–velocity relationship for rain drops increased rapidly over the first two metres, and that under drop heights of 30 cm no soil erosion took place. This threshold represented the mean crown base height of trees in the present study, and is a further argument to consider slow- and low-growing tree species in plantations that aim to minimize soil erosion.

The importance of the number of branches in affecting TKE was moderate. While fewer than 14 branches at low rainfall amounts (events 3 and 4) led to average or high TKE, more than 47 branches led to low and moderate TKE. We ascribe this negative effect to the higher probability for raindrops to split up at branches thus decreasing drop size and velocity, resulting in low TKE (Herwitz, 1987).

A LAI larger than 1 led to average or high TKE, whereas a lower LAI caused a low or moderate TKE (Figures A2, A4 and A5). This threshold resulted in a positive effect of LAI on TKE, which is contrary to previous studies (Geißler et al., 2013; Nanko et al., 2008). However, these studies dealt with LAI ranging from 1.5 to 11. Therefore, the positive effect of LAI might occur only for low LAI, when values are closely related to canopy openness or crown area. Within these low values, a higher LAI represents a higher coverage and throughfall creation without creating more rainfall interception and breaking points by different canopy layers. LAI did not influence TKE variation across all rainfall events.

A crown area below 37,000 cm^2 always led to low or moderate TKE and thus indicates an upper threshold below which TKE can be seen as less-erosive (Figures A2 and A3). We ascribe this positive effect on TKE to rain drop gathering and the creation of a higher area at which throughfall occurred. However, low rainfall intensities (rainfall event 1) counteract this effect when TKE is analysed at distances of 15 cm, 30 cm and 60 cm from the tree stem (see Figure A1). Nanko et al. (2008) showed this negative effect of crown area on TKE by investigating crown areas larger than 85,000 cm^2. Nevertheless, the effect shift remains nonpredictable and crown area did not influence TKE variation across all rainfall events.

The effect of spatial variability on TKE remains inconclusive as its importance in the CART was low and effects became evident only in combination with crown area. Thus, it remains unclear below which spatial positions low or moderate TKE appeared. This absence of a spatial variability of TKE is in agreement with findings of Nanko et al. (2011). Nevertheless, at a stem-distance of 30 cm high TKE may appear below or at margins of the canopy (Finney, 1984).

If all neighbouring trees belong to one species, low TKE occurred. Species mixtures, however, led to moderate TKE. A diverse neighbourhood might lead to more complex tree structures, which can positively affect throughfall by creating different canopy layer height at which drops can confluence (Getzin et al., 2008; Schröter et al., 2012). Nevertheless, a classification by neighbourhood tree diversity as well as ground diameter and specific leaf area was not prominent (low importance in CART and not occurring in CART of all rainfall events).
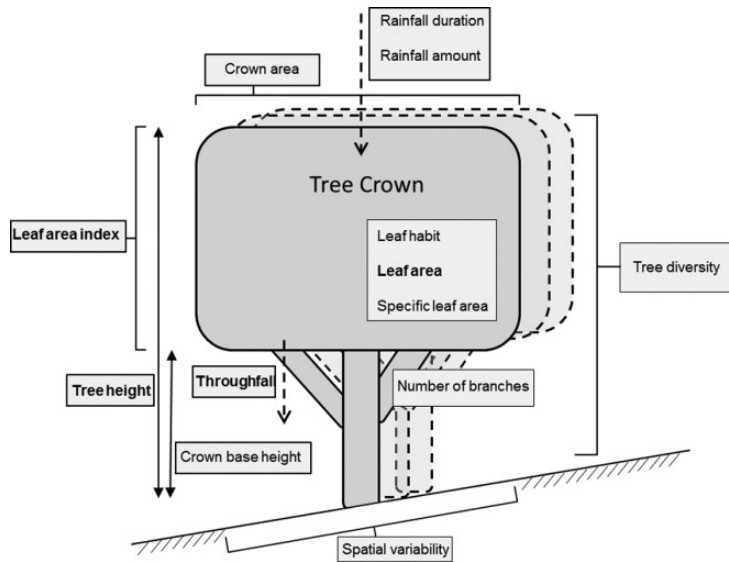
Figure 4. Graphical compilation of relevant biotic and abiotic factors that affect TKE based on CART. Factors in bold were most important in explaining TKE.

3 TKE comparisons with previous studies

In this study, TKE was two-fold lower compared to the mean of other studies investigating rainfall kinetic energy in open fields and below vegetation (see Table 2). The age of the subtropical tree plantation can be considered as the main reason for this finding. Many tree individuals have not yet reached full tree height, which leads to low fall velocities and thus lower TKE (Gunn and Kinzer, 1949). Furthermore, a dense and thick crown cover was not developed in some plots in the previous six years that this plantation existed. LAI and number of branches as major predictors for high TKE emphasized the importance of a dense crown cover (see Figure A3). To our knowledge, only one study measured similar TKE (Finney, 1984); compared to our study, the relatively low vegetation heights there prevented rain drops from achieving their terminal velocity. In contrast, Nanko et al. (2008), Nanko et al. (2011) and Sanchez-Moreno et al. (2012) measured average to high TKE, which might be caused by high-intensity rainfall above 40 mm h^{-1}. These intensities

exceed those of four events measured in our study. Since throughfall amounts are similar to or lower than our measurements, rainfall intensity might function as the major abiotic factor leading to high TKE throughout all studies (Levia and Frost, 2006). However, TKE can be stable among different rainfall intensities ranging from 1 to 46 mm h^{-1} (Zhou et al., 2002). In this case, throughfall amount might be a better predictor for TKE differences.

V Conclusions

We successfully applied a rule-based analysis to model TKE and to compare our findings with literature results. The present study linked biotic and abiotic factors to TKE and set thresholds below which low TKE and above which high TKE occurred (Figure 4). Planting new forests or plantations, these factors should be considered as they constrain the extent of soil erosion. With the set of species and the biotic and abiotic factors used in this study the erosive potential of TKE can be mitigated by: a smaller leaf area than

$6700\,mm^2$, a lower tree height than 290 cm combined with a crown base height lower than 60 cm, a LAI smaller than 1, more than 47 branches and by using a single tree species neighbourhood, while the amount of throughfall can vary. Although these models have been calibrated with data of a young tree plantation, they are, nevertheless, another step towards identifying the importance of biotic and abiotic factors and most of all, setting thresholds for erosion occurrence based on TKE. However, further research is needed in mature forests.

Appendix
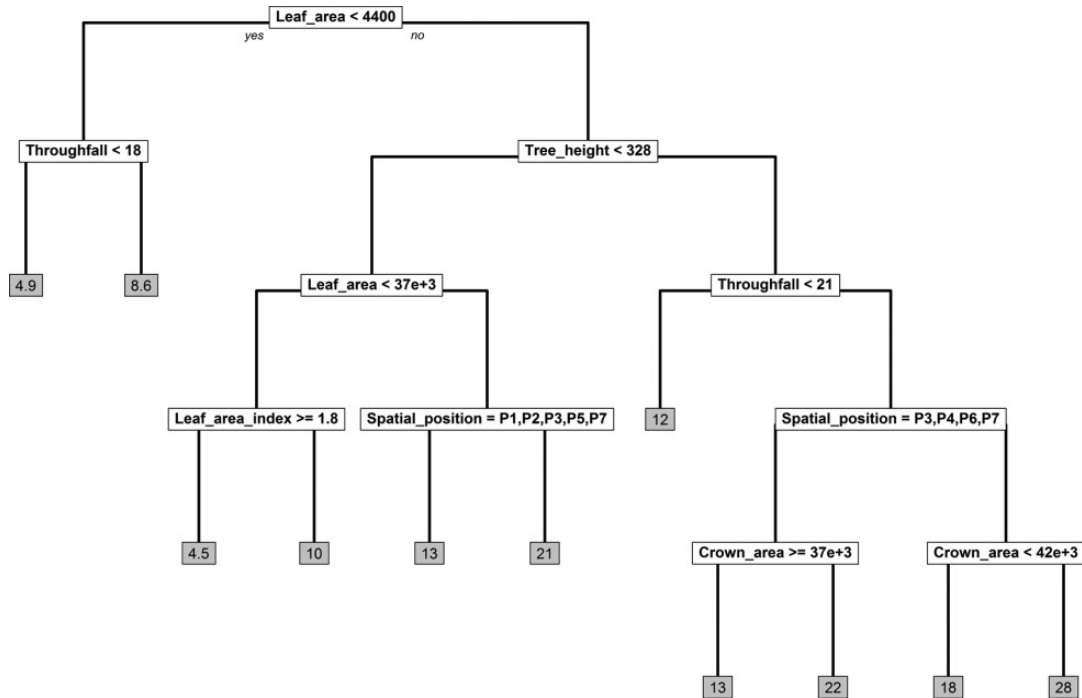
Figure A1. CART of rainfall event 1. Target variable was throughfall kinetic energy (TKE) and predictor variables are listed in Table 3. TKE was measured as $J\,m^{-2}\,mm^{-1}$ (n = 279).
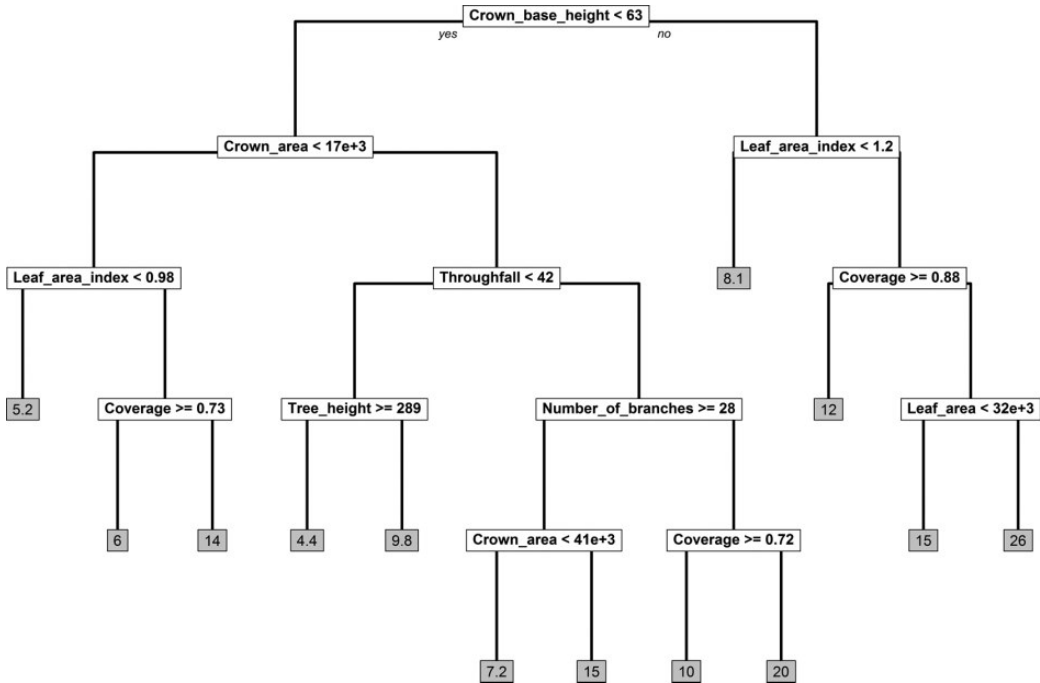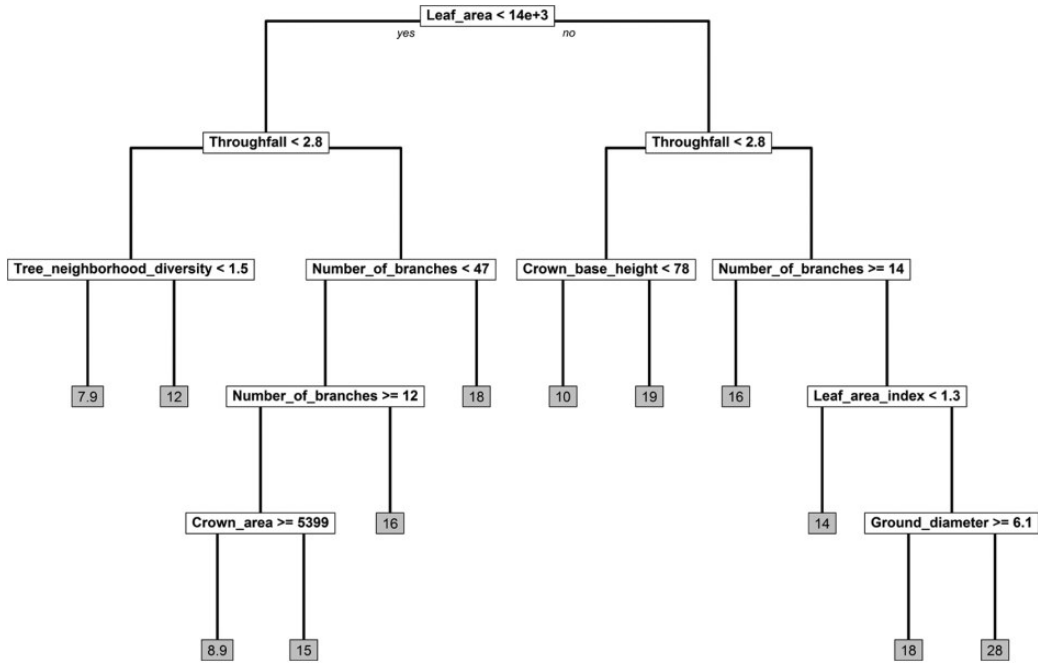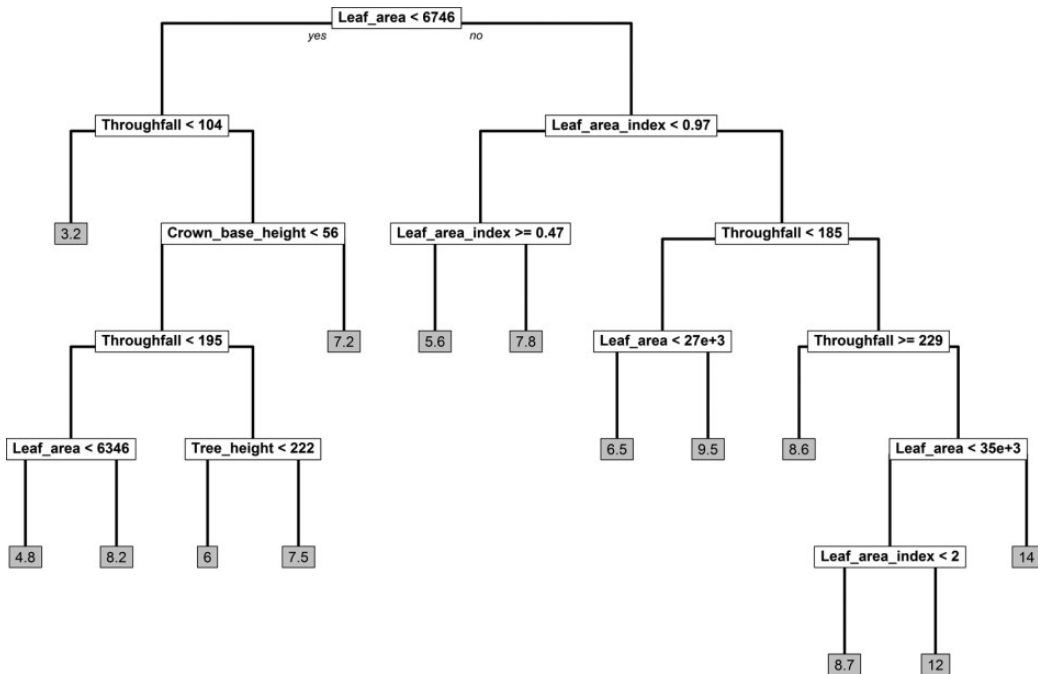
Figure A2. CART of rainfall event 2. Target variable was throughfall kinetic energy (TKE) and predictor variables are listed in Table 3. TKE was measured as $J\,m^{-2}\,mm^{-1}$ (n = 281).
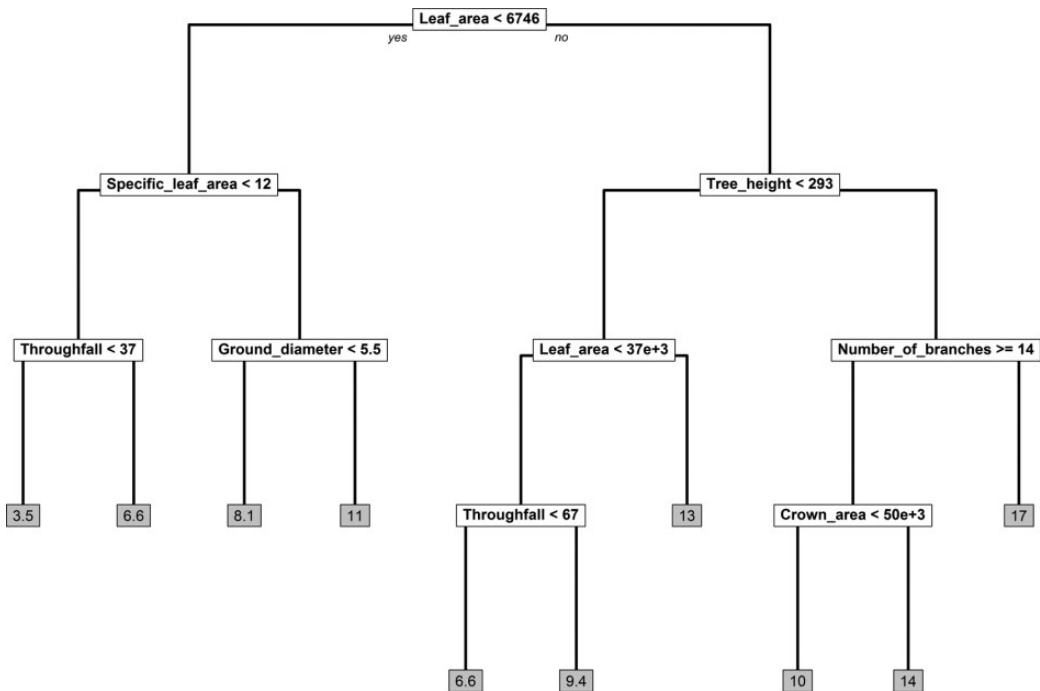
Figure A3. CART of rainfall event 3. Target variable was throughfall kinetic energy (TKE) and predictor variables are listed in Table 3. TKE was measured as $J\,m^{-2}\,mm^{-1}$ (n = 282).

Figure A4. CART of rainfall event 4. Target variable was throughfall kinetic energy (TKE) and predictor variables are listed in Table 3. TKE was measured as $J\,m^{-2}\,mm^{-1}$ ($n = 281$).

Figure A5. CART of rainfall event 5. Target variable was throughfall kinetic energy (TKE) and predictor variables are listed in Table 3. TKE was measured as $J\,m^{-2}\,mm^{-1}$ ($n = 282$).

Table A1. Results of the random forest (RF) feature importance. 'Variables selected' mark the variables on which final classification and regression trees (CART) were built, and 'Variables dismissed' mark variables that have been dismissed after recursive feature elimination (RFE) combined with RFs feature importance. Abbreviations of variables are defined in Table 3.

Model	Variables selected	Variables dismissed
Complete	A, B, E, F, G, H, I, J, K, L, M, N, O, S, T	C, D, P, Q, R
Event 1	D, E, F, G, H, I, J, K, L, M, N, O, P	A, B, C, Q, R
Event 2	A, D, E, F, G, H, I, J, K, L, M, N, O	B, C, P, Q, R
Event 3	A, D, E, F, G, H, I, J, K, L, M, N, O	B, C, P, Q, R
Event 4	C, E, F, G, H, I, J, K, L, M, N, O, Q	A, B, D, P, R
Event 5	A, B, E, F, G, H, I, J, K, L, M, N, O	C, D, P, Q, R

Acknowledgements

Declaration of Conflicting Interests

Funding

References

Assouline S and Mualem Y (1989) The similarity of regional rainfall: a dimensionless model of drop size distribution. *Transaction of the ASAE* (32): 1216–1222.

Assouline S (2009) Drop size distributions and kinetic energy rates in variable intensity rainfall. *Water Resources Research* 45(11): W11501.

Berendse F, van Ruijven J, Jongejans E, et al. (2015) Loss of plant species diversity reduces soil erosion resistance. *Ecosystems* 18(5): 881–888.

Brandt J (1987) The effect of different types of forest management on the transformation of rainfall energy by the canopy in relation to soil erosion. In: *Proceedings of the forest hydrology watershed management conference*. Vancouver, BC (1987), IAHS Publication. No. 167, 213–222.

Brandt J (1988) The transformation of rainfall energy by a tropical rain forest canopy in relation to soil erosion. *Journal of Biogeography* 15(1): 41–48.

Brandt J (1989) The size distribution of throughfall drops under vegetation canopies. *Selected papers of the Fourth Benelux Colloquium on Geomorphological Processes* 16(4–5): 507–524.

Brandt J (1990) Simulation of the size distribution and erosivity of raindrops and throughfall drops. *Earth Surface Processes and Landforms* 15(8): 687–698.

Breiman L (2001) Random forests. *Machine Learning* 45(1): 5–32.

Breiman L, Friedman J, Stone CJ, et al. (1984) *Classification and Regression Trees*. Chapman & Hall/CRC Press, Wadsworth.

Brevik EC, Cerdà A, Mataix-Solera J, et al. (2015) The interdisciplinary nature of *SOIL*. *SOIL* 1(1): 117–129.

Bruelheide H, Nadrowski K, Assmann T, et al. (2014) Designing forest biodiversity experiments: general considerations illustrated by a new large experiment in subtropical China. *Methods in Ecology and Evolution* 5(1): 74–89.

Calder IR (1996) Dependence of rainfall interception on drop size: 1. development of the two-layer stochastic model. *Journal of Hydrology* 185(1–4): 363–378.

Cao L, Zhang K, Dai H, et al. (2013) Modeling inter-rill erosion on unpaved roads in the loess plateau of China. *Land Degradation & Development*: n/a. DOI: 10.1002/ldr.2253.

Cerdá A (1997) Rainfall drop size distribution in the Western Mediterranean basin, València, Spain. *CATENA* 30(2–3): 169–182.

Cerdá A (1998) The influence of geomorphological position and vegetation cover on the erosional and hydrological processes on a Mediterranean hillslope. *Hydrological Processes* 12(4): 661–671.

Cerdá A, Morera AG and Bodí MB (2009) Soil and water losses from new citrus orchards growing on sloped soils in the western Mediterranean basin. *Earth Surface Processes and Landforms* 34(13): 1822–1830.

Cutler DR, Edwards JR, Thomas C, et al. (2007) Random forests for classification in ecology. *Ecology* 88(11): 2783–2792.

Efron B and Tibshirani RJ (1994) *An Introduction to the Bootstrap*. New York: CRC Press.

Finney HJ (1984) The effect of crop covers on rainfall characteristics and splash detachment. *Journal of Agricultural Engineering Research* 29(4): 337–343.

Foot K and Morgan R (2005) The role of leaf inclination, leaf orientation and plant canopy architecture in soil particle detachment by raindrops. *Earth Surface Processes and Landforms* 30(12): 1509–1520.

Geißler C, Kühn P, Böhnke M, et al. (2012) Splash erosion potential under tree canopies in subtropical SE China. *CATENA* 91(0): 85–93.

Geißler C, Kühn P, Shi X, et al. (2010) Estimation of throughfall erosivity in a highly diverse forest ecosystem using sand-filled splash cups. *Journal of Earth Science* 21(6): 897–900.

Geißler C, Nadrowski K, Kühn P, et al. (2013) Kinetic energy of throughfall in subtropical forests of SE China: effects of tree canopy structure, functional traits, and biodiversity. *PLoS ONE* 8(2): e49618.

Getzin S, Wiegand K, Schumacher J, et al. (2008) Scale-dependent competition at the stand level assessed from crown areas. *Large-scale Experimentation and Oak Regeneration* 255(7): 2478–2485.

Goebes P, Bruelheide H, Härdtle W, et al. (2015a) Species-specific effects on throughfall kinetic energy in subtropical forest plantations are related to leaf traits and tree architecture: accepted. *PLoS ONE*. DOI: 10.1371/journal.pone.0128084.

Goebes P, Seitz S, Geißler C, et al. (2014) Momentum or kinetic energy: how do substrate properties influence the calculation of rainfall erosivity? *Journal of Hydrology* 517(0): 310–316.

Goebes P, Seitz S, Kühn P, et al. (2015b) Throughfall kinetic energy in young subtropical forests: investigation on tree species richness effects and spatial variability. *Agricultural and Forest Meteorology*: 148–159.

Grimm R, Behrens T, Märker M, et al. (2008) Soil organic carbon concentrations and stocks on Barro Colorado Island: digital soil mapping using random forests analysis. *Geoderma* 146(1–2): 102–113.

Gunn R and Kinzer GD (1949) The terminal velocity of fall for water droplets in stagnant air. *Journal of Meteorology* 6(4): 243–248.

Herwitz SR (1987) Raindrop impact and water flow on the vegetative surfaces of trees and the effects on stem flow and throughfall generation. *Earth Surface Processes and Landforms* 12(4): 425–432.

Keesstra SD, Geissen V, Mosse K, et al. (2012) Soil as a filter for groundwater quality. *Current Opinion in Environmental Sustainability* 4(5): 507–516.

Kröber W and Bruelheide H (2014) Transpiration and stomatal control: a cross-species study of leaf traits in 39 evergreen and deciduous broadleaved subtropical tree species. *Trees* 28(3): 901–914.

Kröber W, Heklau H and Bruelheide H (2014) Leaf morphology of 40 evergreen and deciduous broadleaved subtropical tree species and relationships to functional ecophysiological traits. *Plant Biology*: n/a. DOI: 10.1111/plb.12250.

Kuhn M, et al. (2014) *Caret: classification and regression training*. R package version 3.08.

Kuz'min VE, Polishchuk PG, Artemenko AG, et al. (2011) Interpretation of QSAR models based on random forest methods. *Molecular Informatics* 30(6–7): 593–603.

Levia DF and Frost EE (2006) Variability of throughfall volume and solute inputs in wooded ecosystems. *Progress in Physical Geography* 30(5): 605–632.

Li Y, Härdtle W, Bruelheide H, et al. (2014) Site and neighborhood effects on growth of tree saplings in subtropical plantations (China). *Forest Ecology and Management* 327(0): 118–127.

Liaw A and Wiener M (2002) Classification and regression by random forest. *R News* 2(3): 18–22.

Lieskovský J and Kenderessy P (2014) Modelling the effect of vegetation cover and different tillage practices on soil erosion in vineyards: a case study in Vráble (Slovakia) using watem/sedem. *Land Degradation & Development* 25(3): 288–296.

MacQueen J (1967) Some methods for classification and analysis of multivariate observations. In: *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*. Vol. 1. No. 14. Oakland, CA.

Märker M, Pelacani S and Schröder B (2011) A functional entity approach to predict soil erosion processes in a small Plio-Pleistocene Mediterranean catchment in Northern Chianti, Italy. *Geomorphology* 125(4): 530–540.

Morgan RP, Quinton JN, Smith RE, et al. (1998) The European soil erosion model (EUROSEM): a dynamic approach for predicting sediment transport from fields and small catchments. *Earth Surface Processes and Landforms* 23(6): 527–544.

Moss AJ and Green TW (1987) Erosive effects of the large water drops (gravity drops) that fall from plants. *Soil Research* 25(1): 9–20.

Nanko K (2007) *Studies on the process of throughfall drop generation in forest canopies*: PhD thesis. University of Tokyo.

Nanko K, Giambelluca TW, Sutherland RA, et al. (2015) Erosion potential under *Miconia calvescens* stands on the island of Hawai'i. *Land Degradation & Development* 26(3): 218–226.

Nanko K, Hotta N and Suzuki M (2004) Assessing raindrop impact energy at the forest floor in a mature Japanese cypress plantation using continuous raindrop-sizing instruments. *Journal of Forest Research* 9(2): 157–164.

Nanko K, Hotta N and Suzuki M (2006) Evaluating the influence of canopy species and meteorological factors on throughfall drop size distribution. *Journal of Hydrology* 329(3–4): 422–431.

Nanko K, Onda Y, Ito A, et al. (2008) Effect of canopy thickness and canopy saturation on the amount and kinetic energy of throughfall: an experimental approach. *Geophysical Research Letters* 35(5): L05401.

Nanko K, Onda Y, Ito A, et al. (2011) Spatial variability of throughfall under a single tree: experimental study of rainfall amount, raindrops, and kinetic energy. *Agricultural and Forest Meteorology* 151(9): 1173–1182.

Nunes AN, Lourenço L, Vieira A, et al. (2014) Precipitation and erosivity in southern Portugal: seasonal variability and trends (1950–2008). *Land Degradation & Development*: n/a. DOI: 10.1002/ldr.2265.

Peters J, Baets BD, Verhoest, et al. (2007) Random forests as a tool for ecohydrological distribution modelling. *Ecological Modelling* 207(2–4): 304–318.

Prasad A, Iverson L and Liaw A (2006) Newer classification and regression tree techniques: bagging and random forests for ecological prediction. *Ecosystems* 9(2): 181–199.

R Core Team (2013) *R: A Language and Environment for Statistical Computing*: R Foundation for Statistical Computing, Vienna.

Renard KG, Foster GR, Weesies GMD, et al. (1997) Predicating soil erosion by water: a guide to conservation planning with the Revised Universal Soil Loss Equation (RUSLE). *USDA Agriculture Handbook* (703): 384.

Salles C and Poesen J (2000) Rain properties controlling soil splash detachment. *Hydrological Processes* 14(2): 271–282.

Sanchez-Moreno JF, Mannaerts CM, Jetten V, et al. (2012) Rainfall kinetic energy–intensity and rainfall momentum–intensity relationships for Cape Verde. *Journal of Hydrology* 454–455: 131–140.

Scholten T, Geißler C, Goc J, et al. (2011) A new splash cup to measure the kinetic energy of rainfall. *Journal of Plant Nutrition and Soil Science* 174(4): 596–601.

Schröter M, Härdtle W and von Oheimb G (2012) Crown plasticity and neighborhood interactions of European beech (*Fagus sylvatica* L.) in an old-growth forest. *European Journal of Forest Research* 131(3): 787–798.

Seitz S, Goebes P, Zumstein P, et al. (2015) The influence of leaf litter diversity and soil fauna on initial soil erosion in subtropical forests. *Earth Surface Processes and Landforms*: n/a. DOI: 10.1002/esp.3726.

Seutloali KE and Beckedahl HR (2015) Understanding the factors influencing rill erosion on road cuts in the south eastern region of South Africa. *Solid Earth* 6(2): 633–641.

Taguas EV, Carpintero E and Ayuso JL (2013) Assessing land degradation risk through the long-term analysis of erosivity: a case study in southern Spain. *Land Degradation & Development* 24(2): 179–187.

Therneau T, Atkinson B and Ripley B (2013) *rpart: recursive partitioning*. R package version 3: 1–46.

van Dijk A, Bruijnzeel L and Rosewell C (2002) Rainfall intensity–kinetic energy relationships: a critical literature appraisal. *Journal of Hydrology* 261(1–4): 1–23.

Wischmeier WH and Smith DD (1978) *Predicting Rainfall Erosion Losses: a guide to conservation planning*. USDA Agriculture handbook, 537. Washington.

Zhou G, Wei X and Yan J (2002) Impacts of a eucalyptus (*Eucalyptus exserta*) plantation on sediment yield in Guangdong Province, Southern China: a kinetic energy approach. *CATENA* 49(3): 231–251.