On the evaluation of soil erosion models: are we doing enough?

Pedro V. G. Batista[1,2,*] pbatista.ufla@gmail.com, Jessica Davies[2], Marx L. N. Silva[1,3], John N. Quinton[3]

[1]Soil Science Department, Universidade Federal de Lavras, Lavras, Minas Gerais, Brazil

[2]Pentland Centre for Sustainability in Business, Lancaster Environment Centre, Lancaster University, Lancaster, UK

[3]Lancaster Environment Centre, Lancaster University, Lancaster, UK

[*]Corresponding author.

Abstract

As any model of real-world phenomena, soil erosion models must be tested against empirical evidence to have their performance evaluated. This is critical to develop knowledge and confidence in model predictions. However, evaluating soil erosion models is complicated due to the uncertainties involved in the estimation of model parameters and measurements of system responses. Here, we undertake a term co-occurrence analysis to investigate how model evaluation is approached in soil erosion research. The analysis illustrates how model testing is often neglected, and how model evaluation topics are segregated from current research interests. We perform a meta-analysis of model performance to understand the mechanisms that influence model predictive accuracy. Results indicate that different models do not systematically outperform each other, and that calibration seems to be the main mechanism of model improvement. We review how soil erosion models have been evaluated at different temporal and spatial scales, focusing on the methods, assumptions, and data used for model testing. We discuss the implications of uncertainty and equifinality in soil erosion models, and implement a case study of uncertainty assessment that enables models to be tested as hypotheses. A comment on the way forward for the evaluation of erosion models is presented, discussing philosophical aspects of hypothesis testing in environmental modelling. We refute the notion that soil erosion models can be validated, and emphasize the necessity of defining fit-for-purpose tests, based on multiple sources of data, that allow for a broad investigation of model usefulness and consistency.

Keywords: soil erosion models; model evaluation; model validation; model calibration; data uncertainty; term co-occurrence analysis.

1 Introduction

There is no shortage of soil erosion models, model applications, and model users. But just how useful are these models? How far can we trust them, and how do we establish such trust? Ideally, soil erosion models should be a valuable tool for scientists, policymakers, and stakeholders. For scientists, erosion models provide a framework to formalize their conceptual interpretation of the processes that regulate the detachment, transport, and deposition of soil particles. This interpretative description of a phenomenon is key to provide understanding and insight (Bailer-Jones, 2009), which is scientifically relevant on its own. Moreover, erosion models are used to make quantitative predictions and scenario-based simulations about how soil is redistributed in potentially complex landscapes, at multiple spatial and temporal scales (e.g., Eekhout et al., 2018; Panagos et al., 2015; Prasuhn et al., 2013; Shrestha and Jetten, 2018; Smith et al., 2018). Policymakers and stakeholders might find these predictions useful, as they may help substantiate environmentally sensitive decisions regarding soil, water, and food security.

With any model of real-world phenomena, it is critical that they are tested against observations if our conceptual understanding of how things work is to be evaluated, and thus continuously improved. Testing is also essential to ascertain the degree of confidence which can be attributed to model predictions under a given set of circumstances. However, gathering data to test soil erosion models is difficult. Erosion is a spatially and temporally variable phenomenon, potentially affected by non-stationary processes (Nearing, 2000; Quinton, 2004). Quantitative erosion measurements therefore require multiple observations in time and space. These measurements always carry a level of uncertainty, are expensive and time consuming (Stroosnijder, 2005). Nonetheless, erosion models must be tested: if

we fail to understand how far erosion models deviate from reality, then how useful can these models be – for scientists or decision-makers?

In this review paper we undertake a scientometric analysis to understand how model evaluation is approached in soil erosion modelling research. We analyze how erosion models have been evaluated, at different spatial and temporal scales, focusing on the concepts, methods, and the data used to test these models. We employ a meta-analysis to investigate model performance and present a case study describing how the uncertainties in both observational data and model structures can be incorporated into evaluation. While describing the advantages and limitations of previously employed approaches to model testing, we provide perspective on what is needed to improve the evaluation of soil erosion models.

2 Model evaluation in soil erosion research: a scientometric term co-occurrence analysis

Term co-occurrence is used in scientometrics to investigate conceptual structures in research fields (Mora-Valentín et al., 2018). The analysis is based on the premise that the relatedness of research topics can be established according to the frequency with which terms co-occur in research articles. Specifically, VOSviewer is a free software (Van Eck and Waltman, 2010) that allows for the construction of distance-based co-occurrence maps, where terms retrieved from titles and abstracts are clustered and mapped according to their relatedness in a similarity matrix.

In order to obtain data-based insight regarding how model evaluation concepts relate to conceptual structures in erosion modelling research, we performed a term co-occurrence

analysis using VOSviewer. We carried out a bibliographic research in October 2018 in the Web of Science database. The query “soil erosion model*” returned 550 articles, with publishing dates from 1985 to 2018. We chose this specific query because it provided an adequate filter of unrelated articles while still allowing for a broad, although not exhaustive, representation of erosion modelling research. Titles, abstracts, and bibliographic information from the returned articles were exported to a text file. A thesaurus file was used to merge synonyms and to exclude general expressions (i.e., aim, study area, and conclusion). A minimum of 15 occurrences was set as a threshold for including terms in the analysis. This process resulted in the inclusion of 178 terms, from which 106 were selected based on a relevance score calculated by VOSviewer. The relevance score is useful for filtering the more informative terms that better represent specific topics (Van Eck and Waltman, 2018). The resulting co-occurrence network map is displayed in Figure 1, and the text files for exploring the map in VOSviewer are provided as the supplementary material.

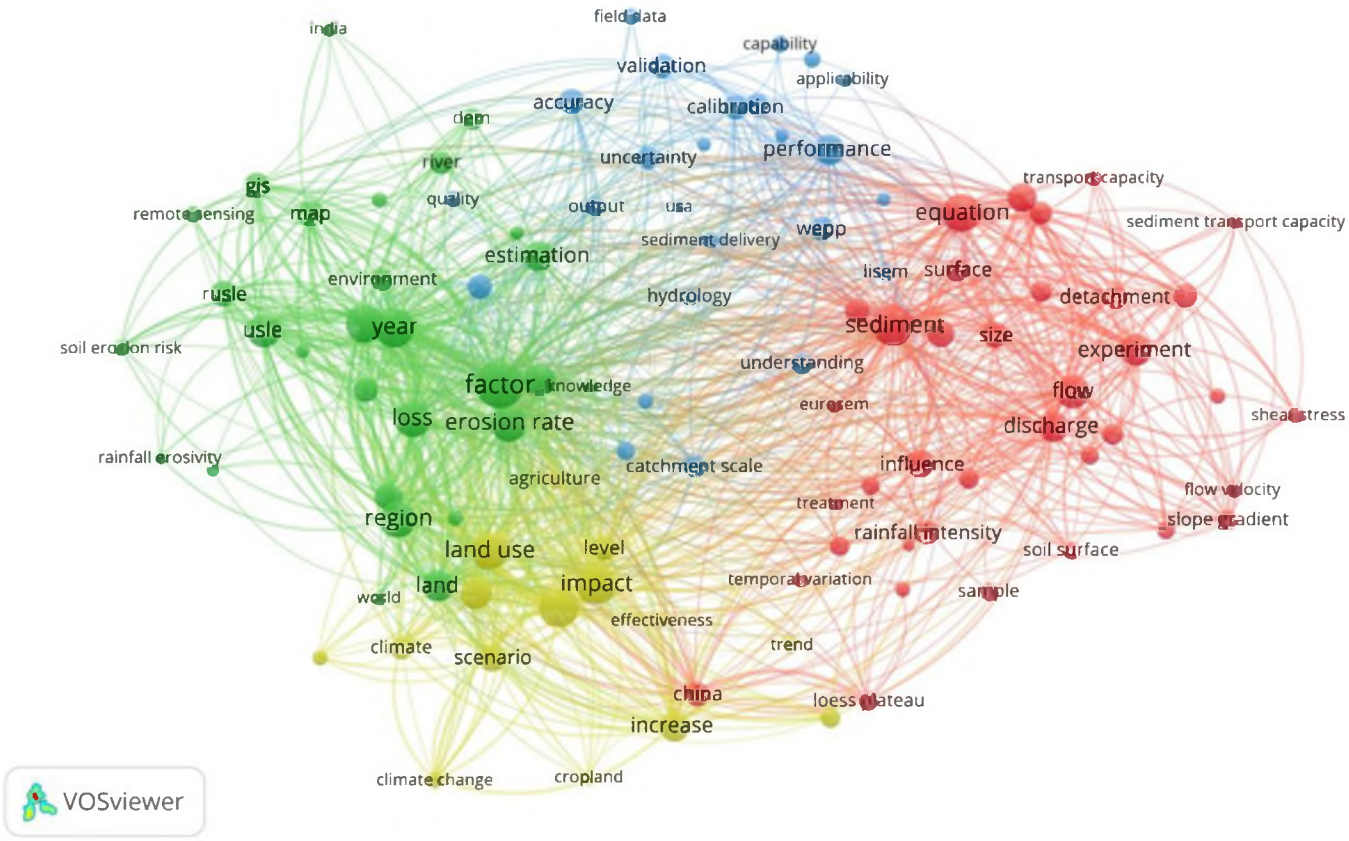


Fig. 1. Term co-occurrence network map. Clusters are identified by color (Cluster 1: green; Cluster 2: red; Cluster 3: yellow; Cluster 4: blue). Labels and circle sizes are proportional to the number of occurrences. Lines indentify major links between terms, and line thickness represents association strength. The distance between terms also reflects association strength. Some term labels are not displayed because of scale (e.g., the circle for the term “outlet” overlaps the one for the term “calibration"). We have provided text files for plotting the co-occurrence map in VOSviewer as supplementary material.

The co-occurrence map identifies four clusters that express different research fronts in erosion modelling. Cluster 1 is primarily driven by model application, as denoted by the presence of terms such as “assessment”, “estimation”, and “erosion rates” (Figure 1). The occurrence of the terms “GIS”, “map”, “remote sensing”, “DEM”, and “spatial patterns” demonstrates this research front is influenced by spatially distributed erosion modelling. These terms may also indicate an interest in large-scale model applications, which is corroborated by the co-occurrence of terms such as “world” and “region”. The temporal scale of model application is coarse, as the association to the term “year” shows. The model names USLE and RUSLE (all model names, acronyms, and their respective references are listed in Table 1) are grouped within Cluster 1, indicating these are the preferred models in this research front.

Table 1 Acronyms, model names, and references.

Acronym	Model name	Reference
AGNPS	A Non-Source Pollution Model	Young et al. (1989)
ANSWERS	Areal Nonpoint Source Watershed Environment Response Simulation	Beasley and Huggins (1982)
EUROSEM	European Soil Erosion Model	Morgan et al. (1998)
LISEM	LImburg Soil Erosion Model	De Roo et al. (1996a, 1996b)
MMF	Morgan-Morgan-Finey Model	Morgan (2001); Morgan et al. (1984)
PESERA	Pan European Soil Erosion Risk Assessment	Kirkby et al. (2008)
RUSLE	Revised Universal Soil Loss Equation	Renard et al. (1997)
SedNet	Sediment and River Network Model	Wilkinson et al. (2004)
SWAT	Soil and Water Assessment Tool	Arnold et al. (1998)
USLE	Universal Soil Loss Equation	Wischmeier and Smith (1978)
USLE-M	Modified Universal Soil Loss Equation	Kinnel and Risse (1998)
USLE-MM	Modified-Modified Universal Soil Loss Equation	Bagarello et al. (2008)
USPED	Unit Stream Power-based	Mitasova et al. (1996)

	Erosion Deposition	
WaTEM/SEDEM	Water and Tillage Erosion Model and Sediment Delivery Model	Van Oost et al. (2000); Van Rompaey et al. (2001); Verstraeten et al. (2010)
WEPP	Water Erosion Prediction Project	Flanagan and Nearing (1995)

On the opposite side of the network map, the research front depicted by Cluster 2 is concerned with process description (Figure 1). Most terms in Cluster 2 are related to erosion-driving processes (e.g. “overland flow”, “sediment transport”, “infiltration”, and “detachment”), mathematical description of these processes (e.g. “equation” and “coefficient”), and to experimental data (e.g. “treatment”, “experiment”, and “sample”). Moreover, Cluster 2 research front is focused on finer time scales, as indicated by the links with terms such as “rainfall event”, “min” and “temporal variation”. EUROSEM is the only model name grouped within Cluster 2.

On the bottom-left corner of the network map, Cluster 3 encompasses erosion modelling research related to scenario-based simulations (Figure 1). This is expressed by the occurrence of terms such as “scenario”, “trend”, “increase”, and “decrease”. The main themes appear to be land use and climate change scenarios. The location of Cluster 3 on the network map indicates it is more strongly related, and has more connections to Cluster 1, with fewer links to Clusters 2 and 4.

On the top of the network map, Cluster 4 clearly distinguishes research focused on model evaluation (Figure 1). Terms associated to the description of model efficiency (e.g., “performance”, “accuracy”, “capability”, “limitation”, and “applicability”) and important topics regarding model evaluation (e.g., “calibration”, “validation”, “uncertainty”, “sensitivity analysis”, and “field data”) are plotted within Cluster 4. The model names

WEPP and LISEM are grouped within this cluster, although overlapping Cluster 2 in the network map. This indicates that the use of these models is frequently associated to some form of model evaluation. Interestingly, the term “outlet” is also found within Cluster 4. “Outlet” also has a strong connection to terms like “discharge”, “sediment transport”, “calibration”, and “validation”. This demonstrates how erosion model testing commonly relies on system outlet measurements of sediment fluxes.

The fact that model evaluation topics are clustered separately from other fronts in erosion modelling research highlights two distinct trends. First, more optimistically, it demonstrates that there is a specific interest in model evaluation: researchers are trying to test their models, which is essential to develop knowledge and confidence in model predictions. Second, it illustrates that such interest is perhaps too specific: models are mostly tested in evaluation-oriented studies, and not in general model applications. The latter conclusion can be corroborated by the fact that the terms “validation”, “validate”, or “validated” only appear in 8 % of the titles and abstracts of the analyzed articles. Related words, such as “tested” or “verified” did not meet the occurrence threshold and/or the VOSviewer relevance score.

In Figure 2 we plotted the co-occurrence map using overlay visualization. Circle colors are rendered according to normalized average year of publication of the articles in which the labeled terms occur. Although the range of the average years of publication is relatively narrow (2003-2013), Figure 2 demonstrates a clear trend towards the outer regions of Clusters 1 and 3. This indicates that erosion modelling research has recently focused on model application and scenario-based simulations, possibly trying to understand the impacts of land use and climate change on future erosion rates. Terms such as

“assessment”, “impact”, “scenario”, “magnitude”, “land use change”, and “climate change” seem to be current popular topics. Figure 2 also indicates a growing interest in RUSLE (e.g., “RUSLE”, “soil erodibility”, and “rainfall erosivity”) and on large scale modelling (e.g., “region” and “remote sensing”). Overall, process description (Cluster 2) and model evaluation (Cluster 4) research articles have earlier publication dates.

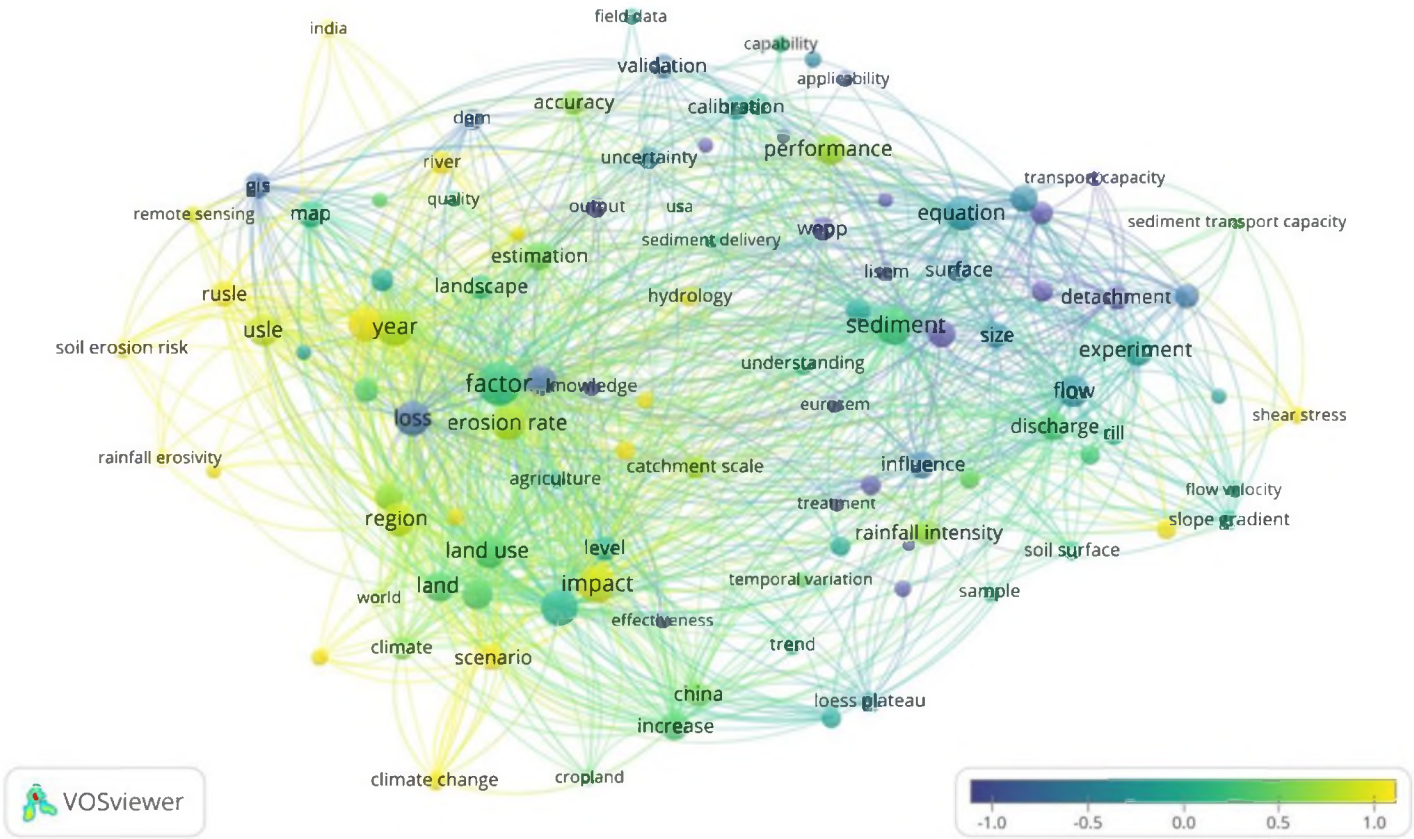


Fig. 2. Overlay visualization of the term co-occurrence network map. Colors are rendered according to the normalized average year of publication of the articles in which the terms occur. Normalization was performed by subtracting the term average by the overall mean and dividing it by the standard deviation. Earlier research topics are colored blue and more recent ones in yellow.

This recent publication trend may indicate that researchers are confident about the capacity of erosion models to estimate soil loss rates and sediment yields, to indentify erosion hot-spots in large catchments, and to simulate erosion responses to land use and climate change. However, comprehensive evaluation-oriented studies have demonstrated that the predictive accuracy of un-calibrated erosion models is often limited (de Vente et al., 2013; Jetten et al., 1999; Van Rompaey et al., 2003), that the variability of soil erosion measurements is large and poorly understood (Nearing, 2000), that the quality of spatial predictions is questionable (Evans and Brazier, 2005; Jetten et al., 2003; Takken et al., 1999), and that model outputs are considerably uncertain (Brazier et al., 2000; Quinton, 1997). Hence, what do we expect to achieve from increasingly complex, large scale and simulation-driven erosion model applications without further testing? What have we learned from previous attempts to evaluate soil erosion models? In the remainder of this review we will discuss different approaches to erosion model evaluation while trying to answer these questions.

3 Evaluation of soil erosion models

The basic approach to the evaluation of soil erosion models is testing their predictive accuracy against measured empirical data, which, as the term co-occurrence analysis demonstrates, are most often sediment transport rates at the outlet of a system. Transport rates are usually expressed in terms of mass area^{-1} time^{-1}. Although the use of these units has been criticized for not accounting for scale dependency (Parsons et al., 2009), it is perhaps the best available system for quantifying erosion (Boardman, 2006).

The use of the mass area^{-1} time^{-1} unit system and the outlet approach to erosion quantification are connected to the earliest and most widely used devices for measuring soil

losses and runoff: the erosion plots (Dotterweich, 2013). These plots operate by conducting runoff from a delimited upslope area to collection tanks, in which sediments are collected and quantified (see Kinnell, 2016). Soil loss measurements from erosion plots have therefore also been used to build/test erosion models (e.g., Morgan, 2001; Renard et al., 1997; Risse et al., 1993; Wischmeier and Smith, 1978; Zhang et al., 1996), and similar outlet-based approaches to model testing have been expanded to spatially distributed catchment scale model applications (e.g., Amore et al., 2004; Fernandez et al., 2003; Jain and Ramsankaran, 2018; Tanyaş et al., 2015). For distributed models, however, investigating the quality of the spatial predictions is an important part of model evaluation. Other issues regarding process representation and parameter estimation can have quite different ramifications according to the spatial scale of the model applications. Therefore, in the next sections we review separately how erosion models have been evaluated at I) plot scale model applications, and at II) larger scales spatially distributed applications (e.g., field, catchment, regional), with an emphasis on spatial data used for model testing in the latter case.

3.1 Evaluating soil erosion models at the plot scale

At first, testing erosion models at plot scale seems reasonably straightforward. As many models were initially developed to predict erosion rates from hillslope segments, model outputs were analogous to soil losses from erosion plots. Therefore, once models had been parameterized and run, their outputs could be directly compared to measured sediment transport rates at the outlet of erosion plots. Model efficiency could then be described by performance metrics such as the coefficient of determination (R^2) or the Nash-Sutcliffe efficiency index (NSE) (Nash and Sutcliffe, 1970). However, there are several approaches

to model evaluation, even at plot scale. Different approaches can be more or less useful according to the purpose of the evaluation, the structure of the models, and the robustness of the dataset.

The simplest approach is a “blind” evaluation. Models are parameterized, run, and tested against observed soil losses. In the case of empirical models, such as USLE-family models, parameterization is carried out based on plot characteristics and rainfall measurements that allow for the selection/calculation of appropriate parameter (i.e., factor) values (e.g. Rapp et al., 2001; Risse et al., 1993). For process-based models, measuring soil, plant, and rainfall/runoff properties is usually necessary. If these measurements are not feasible, parameter values can be retrieved from the literature, estimated by transfer functions or by knowledge-based approximations (e.g., Bulygina et al., 2018; Fernández et al., 2010; Flanagan and Frankenberger, 2012; Veihe et al., 2001). According to Quinton (1994), “blind” evaluation is useful to test model performance in a specific set of soil, topography, and land use characteristics. This can provide an indication of the confidence with which a model can be applied to these specific conditions.

However, the parameterization of erosion models, particularly process-based, can be challenging. Some parameters may not be directly measurable, and therefore might have to be estimated based on regression techniques and expert judgments (Brazier et al., 2001). Moreover, establishing initial conditions for continuous simulation models is problematic, as detailed temporal measurements of model parameters are rarely available (Beven, 2009; Quinton, 1997). Therefore, soil erosion models are often calibrated, meaning one or multiple parameters and/or boundary conditions are adjusted so that prediction error is minimized.

For calibrated erosion models, common approaches to evaluation rely on some kind of split-off sub-setting, in which a dataset is used for model calibration (or training) and another set is used for "validation" (or testing). This split-off can be I) temporal, in which soils losses observed during a certain period of time are used as the training dataset and analogous records from a different period are used for testing (e.g., Anache et al., 2018; Jetten et al., 1999; Licciardello et al., 2013; Veihe et al., 2001); or II) spatial, in which models are calibrated using data from a given plot, or set of plots, and are subsequently tested on different plots with similar conditions (e.g., Bagarello et al., 2013; Vieira et al., 2014). Although split-off sub-setting is commonly employed to test calibrated erosion models, some studies have used the same dataset for both calibration and testing (e.g., Kinnell et al., 2018; Mahmoodabadi and Cerdà, 2013).

Considering that models often have a large number of parameters, that parameter measurements are subject to considerable uncertainty and may therefore assume a wide range of values, calibrated erosion models are sometimes capable of reproducing the right answer for the wrong reasons (Govers, 2011; Jetten et al., 2003; Quinton, 1994). Hence, it can be argued that using the same dataset for calibration and testing is the least robust approach. Moreover, although temporal split-off tests can provide information on the capability of a calibrated model to simulate the responses of erosion rates to temporal changes in soil properties, plant growth, and rainfall events; such tests are restricted to the very specific systems used during calibration/testing. As demonstrated by Nearing et al. (1999) and Wendt et al. (1986), the variability of erosion rates on replicate plots is large and poorly explained by the differences in plot characteristics, at least considering our ability to measure them. Hence, even if a model is able to make perfect predictions of

erosion rates for one plot, such a model would always fail to provide the same efficiency for a replicate. As argued by Beven (2009), "an 'optimum' model can only be *conditionally* optimal", as the solution to an inverse problem will depend on the optimization function being used, the errors in the calibration data, and the evaluation criteria. Temporal split-off tests may therefore transmit an overestimated sense of confidence to model estimates, unless it is made clear that the reported model performance should not be expected elsewhere then in the calibrated system. In this sense, spatial split-off tests seem more powerful, as in this approach model performance will reflect some of the variability of erosion measurements in very similar systems. Successive interactions of temporal and spatial split-off tests, as in Klemes (1986) hierarchical scheme, can therefore provide a framework to evaluate the performance of calibrated models regarding their transferability in time and space, which is a desirable feature for erosion models (Beven and Young, 2013; Quinton, 1994).

A robust framework for incorporating the variability of erosion plot data into model evaluation is provided by Nearing (2000), who developed a criterion based on the difference of erosion rates between replicate plots. Nearing (2000) argued that "the replication of an individual plot may be considered a 'real-world' physical model of that plot". However, erosion rates on replicate plots can be quite variable, particularly for events of lower magnitude (Nearing et al., 1999; Wendt et al., 1986). This is most likely the result of the spatial variability of the soil properties and the underlying processes driving soil erosion, which we are unable to measure and to represent deterministically in model structures. Hence, Nearing (2000) stated that acceptable model errors could be defined according to the measured variability of erosion rates between replicates. That is, if the

differences between modeled and observed soil losses are within the 95 % occurrence interval of the differences between replicate measurements, then the model error should be considered acceptable. This is based on the premise that a mathematical model should not be expected to outperform a "real-world" physical model.

3.1.2 A meta-analysis of erosion model performance at the plot scale

Still building on the variability of replicate plot data, Govers (2011) argued that models have already achieved the upper limit of erosion predictability. Such limit would be equivalent to the predictability observed in replicate plots provided by Nearing (2006) (R^2 = 0.77 for erosion rates >75 ton ha^{-1}). Govers (2011) demonstrates that many evaluation studies have reported similar R^2 values to the ones obtained in replicate plots, particularly for annual and average annual erosion rates, and that sophisticated process-based models do not out perform more simple USLE-family models.

In order to investigate the performance of erosion models at plot scale, we compiled the results from several model evaluation studies which compared predicted and observed soil losses (Table 2). As the NSE was the preferred metric used to describe model efficiency by authors, our analysis focused on such index. This yielded 112 data entries, which were grouped by model, by the temporal scale of the application, and by the use or not of calibration. Results are displayed in Figure 3.

Table 2 References for the compiled NSE values on Figure 3.

Reference	Location	Data entries	Models
Amorim et al. (2010)	Brazil	3	RUSLE, USLE, WEPP
Anache et al. (2018)	Brazil	2	WEPP
Bagarello et al. (2013)	Italy	2	USLE-M, USLE-MM

Bulygina et al. (2018)	USA	1	WEPP
Di Stefano et al. (2017)	Italy	3	USLE-M, USLE-MM, USLE
Fernández et al. (2010)	Spain	7	MMF, RUSLE
Fernández et al. (2016)	Spain	4	PESERA, RUSLE
Fernández et al. (2018)	Spain	2	RUSLE, WEPP
Flanagan and Frankenberger (2012)	USA	4	WEPP
Kinnel (2017)	USA	43	RUSLE, RUSLE2, USLE, USLE-M, WEPP
Kinnel et al. (2018)	China	2	RUSLE, USLE-M
Larsen and MacDonald (2007)	USA	4	WEPP
Licciardello et al. (2013)	Spain	12	WEPP
Mahmoodabadi and Cerdà (2013)	Iran	3	WEPP
Morgan (2001)	Multiple	2	MMF
Rapp et al. (2001)	USA	2	RUSLE, USLE
Risse et al. (1993)	USA	2	RUSLE, USLE
Spaeth et al. (2003)	USA	6	RUSLE, USLE
Tiwari et al. (2000)	USA	2	WEPP
Vieira et al. (2014)	Portugal	6	MMF

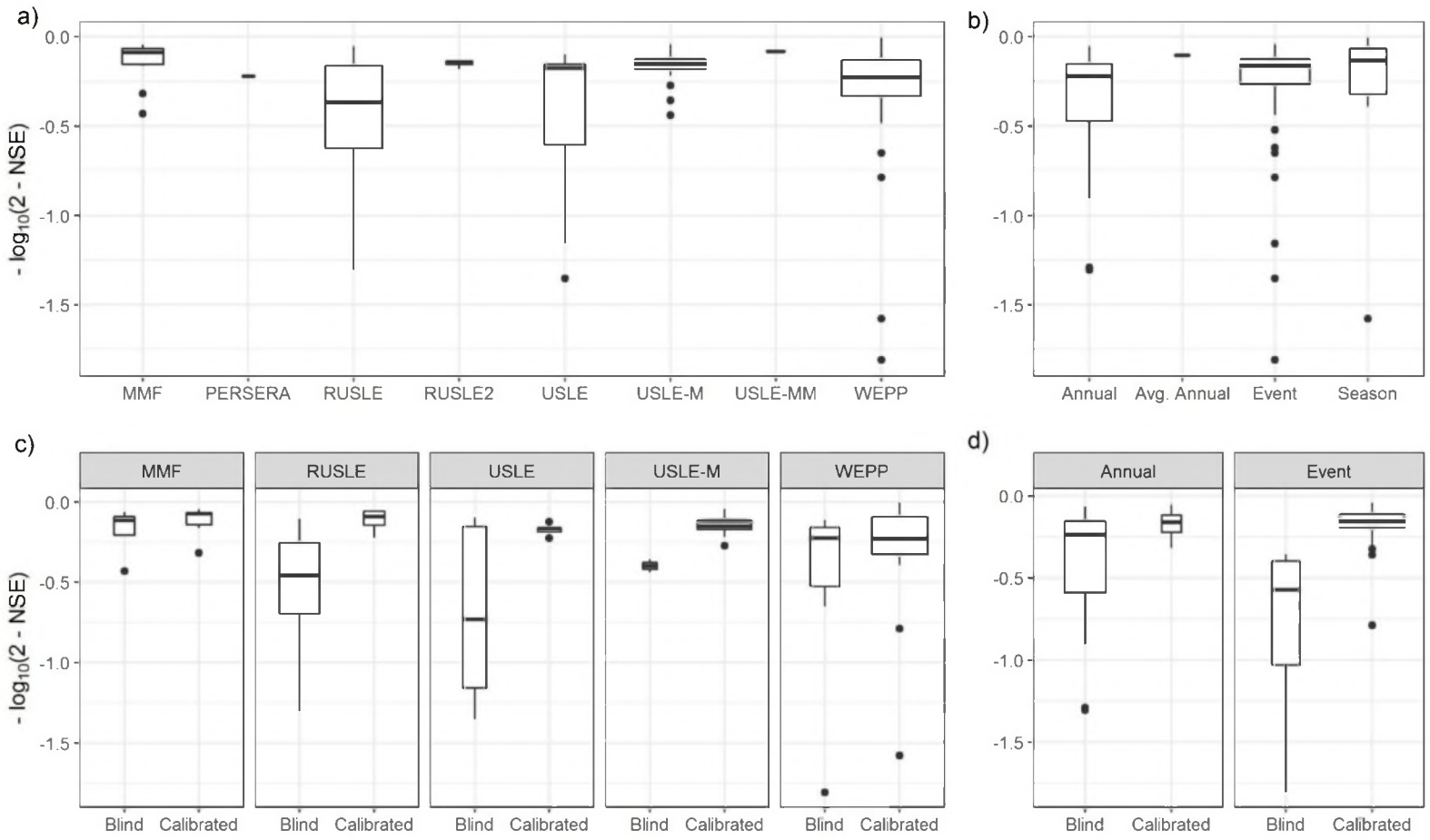


Fig. 3. NSE values reported in erosion modelling studies grouped by: a) model; b) temporal scale of model application; c) model and the use or not of calibration; d) temporal scale of model application and use or not of calibration. The width of the boxes is scaled according to the size of the datasets for each group. In figures 3c and 3d we only display models and temporal scales which were used both with and without calibration. For better visualization, NSE values have undergone log-linear transformation.

Our literature review corroborates part of Glovers (2011) conclusion: models do not systematically outperform each other regarding the accuracy of erosion predictions (Figure 3a). Moreover, we found that model performance is not necessarily linked to the temporal scale of the application (Figure 3b, d), and that, apparently, mathematical models are quite capable of outperforming the physical "real-world" models; at least considering the way they have been evaluated. For instance, Licciardello et al. (2009) achieved, after calibration, R^2 values of 0.90 for annual erosion predictions using PESERA. Anache et al. (2018) reported R^2 values of 1.00 and NSE values of 0.93 for seasonal calibrated WEPP estimates. Kinnel (2017) reports NSE values of 0.89 for event-based USLE-M predictions, also after calibration. Using event-based calibrated WEPP predictions, Mahmoodabadi and Cerdà (2013) reported NSE values of 0.90.

Hence, does this mean that mathematical models do a better job at estimating soil losses than "real-world" physical models? Probably not: if the mathematical models were applied to a wider range of replicates in a more robust evaluation scheme, their performance would be bounded by variability of erosion plot data and our inability to represent it deterministically.

Overall, the compilation of NSE values reported in erosion modelling studies displayed in Figure 3 seems to indicate that calibration is the main mechanism for improving model performance. This is made particularly clear when models and the temporal scale of model application are compared separately (Figures 3c, d). If calibration is really the main way of affecting model performance, we must come to the conclusion that different models or different model realizations can be equally accurate, or equally acceptable. This is because of the conditional nature of parameter optimization, as we previously discussed. Hence,

how can we ever reject a model? Moreover, how can we know if a model is making accurate estimates for the right reasons?

The concept that, given the errors involved in the characterization of a system, many representations of reality can be considered acceptable, is defined by Beven (2006) as equifinality. This seemly uncomfortable assertion has serious implications on the evaluation of environmental models, which are often ignored in erosion modelling research. If one is aware of the epistemic uncertainties necessarily embedded into model structures, as well as of the inevitable errors associated to the measurements of temporal and spatially variable parameters, it is hardly justifiable that model outputs should be presented in a deterministic fashion. Hence, Quinton (1994) argues that, even if a model is applied "blind", some sort of uncertainty measure should be provided. During calibration, dealing with uncertainty and equifinality is perhaps even more urgent. Without it, confidence in model predictions can be overestimated, as model deficiencies can be concealed by optimization. Moreover, as we discussed, (quite) different parameter sets can produce adequate representations of reality. If multiple model realizations are empirically equivalent, then why should one be preferable over another? For spatially distributed models, the degrees of freedom afforded by parameterization are even larger, as well as the uncertainties surrounding parameter estimation. Methods for incorporating equifinality and uncertainty analysis to erosion model evaluation will be discussed in section 4 of this review.

3.2 Evaluating spatially distributed erosion models: from field to regional scales

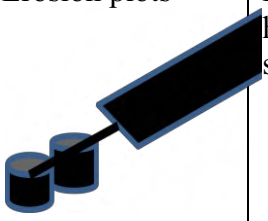
The advent of GIS, the accessibility of computing, and the popularization of remote sensing images had a great impact on erosion modelling: models can now be applied at large scales and in a distributed manner with relative ease. Contrary to earlier lumped model results, the grid-based outputs of spatially distributed erosion models make it possible to identify where erosion and deposition occur, together with their magnitude, at different temporal and spatial scales. This could ultimately help policymaking and resource allocations regarding soil conservation. Hence, a great effort has been put into adapting and scaling erosion models into a GIS framework (e.g., Desmet and Govers, 1997; Mitasova et al., 1996; Renschler, 2003; Renschler and Harbor, 2002), and some models, such as LISEM and WaTEM/SEDEM, were developed in an explicitly distributed, rater-based structure.

However, evaluating distributed erosion models, where catchments are the predominant spatial scale of application, is problematic: the previously discussed issues regarding model evaluation are exacerbated, as parameterization becomes even more uncertain and equifinality more likely. Moreover, the outlet-based approach to model evaluation – which seems reasonable at plot scale – is usually unsatisfactory to describe the performance of distributed erosion models. The main reasons for this is that I) at catchment scale, different processes which may not be described by model structures can considerably influence sediment yield dynamics (e.g., bank erosion, gully erosion, overbank sedimentation, and floodplain deposition) (Favis-Mortlock et al., 2001); and II) models can adequately simulate catchment sediment yield while misrepresenting the spatial patterns of erosion and deposition (Jetten et al., 2003; Takken et al., 1999; Van Oost et al., 2004). Therefore, data used for model evaluation must be compatible with model structure and process

representation (Govers, 2011). Moreover, evaluating distributed models requires spatial data, as erosion does not occur at discrete points in space (Boardman, 2006). Finally, incorporating the spatial errors of parameter estimation is necessary when describing the uncertainty of spatially distributed models. These issues have been recognized by erosion modelers, and the attempts made to address them – particularly by incorporating spatial data into model testing – will be reviewed in the following. For a discussion on outlet sediment yield predictions at catchment scale, covering lumped and distributed models, we refer to de Vente and Poesen (2005) and de Vente et al. (2013).

Spatially distributed data suitable for model evaluation are generally acquired by I) field-based monitoring, in which erosion and depositional features are mapped and often combined volumetric measurements of rills, gullies, and sediment deposits (e.g., Desmet and Govers, 1997; Evans and Brazier, 2005; Hessel et al., 2006; Prasuhn et al., 2013; Takken et al., 1999; Van Oost et al., 2004; Vigiak et al., 2005); II) tracing techniques, usually relying on fallout radionuclide inventories to model medium/long term soil redistribution rates (e.g., Bacchi et al., 2003; Banis et al., 2004; He and Walling, 2003; Lacoste et al., 2014; Porto and Walling, 2015; Walling et al., 2003; Walling and He, 1998) or fingerprinting techniques for identifying sediment sources (e.g., Borrelli et al., 2018; Wilkinson et al., 2013); and III) remote sensing, in which high resolution aerial images are used to assess erosion severity in a qualitative/ semi-quantitative manner by visual identification of erosion signs (e.g., Fischer et al., 2018) (Table 3).

Table 3 Characteristics and suitability of sources of data for evaluating soil erosion models according to the scale and purpose of the application.

Sources of data	Typical scale	Characteristics	Pros	Cons	Most useful for testing
Erosion plots 	Hillslope/ hillslope segment	• Quantitative soil loss measurements; • Point based (plot outlet) measurements; • Measurements reflect rill and interrill processes.	• Reasonably controlled experimental setting; • Direct sediment transport rate measurements.	• Requires constant monitoring/ maintenance; • Prone to edge effects; • Does not discriminate soil redistribution processes.	• Empirical and process-based models; • Model components and sub-routines; • Model responses to different land use/management, soil classes, and slopes.

Fallout radionuclide inventories 	Field/ catchment	• Quantitative; • Medium to long-term estimates; • Point-based measurements.	• Provides spatially referenced estimates erosion and deposition rates.	• Indirect method; • Uncertainty in conversion models; • Uncertainty in spatial interpolation; • Does not discriminate soil redistribution processes.	• Process-based erosion models; • Capability of models to simulate erosion rates/patterns;

Field-based monitoring 	Field/catchment	• Quantitative or semi-quantitative; • Cross-sectional rill/gully measurements • Deposition thickness/area measurements; • Visual identification of erosion signs.	• Direct volumetric measurements with explicitly spatial locations; • Recognition of soil redistribution processes (e.g., gully, rill, tillage).	• May not account for interrill erosion; • Requires constant monitoring; • Labor intensive and time consuming.	• Process-based erosion models; • Capability of models to simulate erosion rates/patterns;
Remote sensing*   	Regional	• Semi-quantitative; • Visual identification erosion signs.	• Low labor requirement, little time consuming; • Large areas are covered with relative ease.	• Restrictions due to temporal and spatial resolution of image acquirement; • Erosion rates are not measured; • May be unsuitable for non-arable land.	• Model-based erosion risk assessments; • Capability of models to identify relative rank of erosion-prone areas.

Sediment fingerprinting [redacted]	Catchment	• Quantitative; • Identification of in-stream sediment provenance.	• Represents multiple phases of sediment transport; • Provides insight into off-site erosion impacts.	• Indirect method, also model-based and uncertain; • Estimates relative contributions of sediment sources, not transport rates; • Sediment remobilization and non-stationarity of sources in time may complicate comparisons with models.	• Erosion models with sediment delivery/routing components; • Capability of models to simulate off-site erosion impacts and to identify sediment yield sources/components.

* Unmanned aerial vehicles and structure-from-motion techniques have shown promising results for reconstructing complex topographic features and measuring soil redistribution rates in recent studies (Balaguer-Puig et al., 2018; Fiener et al., 2018). Although to the authors' knowledge such techniques have not yet been used to test erosion models, such an approach might be able to combine some of the capabilities of remote sensing and field-

based surveys for monitoring soil erosion, and therefore might be useful for evaluating distributed models in a variety of scales.

3.2.1 Comparing soil erosion models to field-based monitoring schemes

Field surveys offer an interesting opportunity for evaluating spatially distributed erosion models, as their results often combine qualitative and quantitative data. For instance, the Ganspoel and Kindervel datasets (Van Oost et al., 2005) consist of two to three years of georeferenced measurements of internal erosion and deposition features, as well as outlet sediment transport rates from two Belgium catchments, with drainage areas of 117 ha and 250 ha. Although direct comparisons between distributed erosion models and field surveys are not always straight-forward – interrill erosion may not be accounted for in field monitoring (Evans and Brazier, 2005) and volumetric measurements can be considerably uncertain, particularly for sediment-deposition features (Castillo et al., 2012; Van Oost et al., 2004) – it is reasonable to assume that, in order to be useful, model estimates should compare well to field observations. That is, if a model depicts high soil losses for a given location, it should be expected that field surveys would also represent the erosion severity for the area (Evans and Brazier, 2005).

However, this is not always the case: in fact, many studies comparing field-based monitoring and distributed soil erosion models report a poor agreement between modeled and surveyed erosion patterns (e.g, Evans and Brazier, 2005; Hessel et al., 2006; Jetten et al., 2003; Takken et al., 1999; Vigiak et al., 2005). In such cases, models generally display a tendency to overestimate both the severity and the extent of erosion rates.

The poor performance of erosion models against observed field patterns is most commonly attributed to I) the uncertainties involved in spatial input parameter estimation, particularly for process-based models (Hessel et al., 2006; Jetten et al., 2003; Vigiak et al., 2006a) and

II) incomplete, incorrect, or unsuitable process descriptions embedded in model structures (Evans and Brazier, 2005; Vigiak et al., 2005). Both explanations for poor model performance provide insight into what is needed to improve the evaluated models. These conclusions would likely not be possible if model testing was restricted to catchment outlet responses. As argued by Quinton (1994), while successful tests can conditionally corroborate a model's capability to reproduce the behavior of a system, they do little to confirm the *veracity* (i.e., truthfulness) of model components. On the other hand, a failure will most likely lead to model improvements.

Although erosion and deposition patterns simulated by spatially distributed models frequently compare poorly to the ones observed in field surveys, erosion risk assessment maps – usually produced by USLE-type models or decision trees – have been reported to provide adequate identification of erosion-prone areas when evaluated against field data (e.g., Djuma et al., 2017; Prasuhn et al., 2013; Vigiak et al., 2006b; Vrieling et al., 2006; Waltner et al., 2018). In such cases, however, model testing is less rigorous, although arguably fit-for-purpose; as a more qualitative approach is employed by comparing modeled and observed erosion severity classes. When actual erosion rates are compared, results are not as encouraging (see Prasuhn et al. 2013).

3.2.2 Comparing soil erosion models soil/sediment tracers

An alternative to field surveys for acquiring spatially distributed data are tracing techniques, which are used to quantify soil redistribution rates across landscapes. Tracing usually relies on fallout radionuclides (FRN) (^{137}Cs, ^{210}Pb, ^{7}Be) inventories (see Guzmán et al., 2013 for a review). The technique is based on the premise that atmospheric input of

FRN is homogeneous within a given spatial unit (e.g., field, catchment), and that factors controlling FRN movement are the same as the physical processes regulating the redistribution of the soil particles to which they are adsorbed (Warren et al., 2005). Hence, when FRN inventories from point samples are compared to an undisturbed reference site inventory, the decrease or increase of tracer concentrations can indicate if an area has been subjected to erosion or deposition (Quine et al., 1994; Walling and He, 1998). Actual erosion/deposition rates are then estimated by conversion models (Walling and He, 1999), often combined with spatial interpolation procedures (e.g. Ferro et al., 1998; Porto and Walling, 2015).

FRN tracing offers an advantage over field surveys in the sense that medium to long term soil redistribution rates and patterns can be estimated based on a single sampling campaign, therefore not requiring constant monitoring. This can be more or less useful according to the time scale of the erosion model application involved in the testing procedure. On the other hand, the conversion of FRN inventories into erosion rates are a source of substantial uncertainty (Walling and He, 1999), as well as the interpolation methods used to spatialize point observations of tracer concentrations. Some researchers have even questioned the general applicability of FRN inventories for estimating soil redistribution rates (see Parsons and Foster, 2011 for a critical perspective). Another issue regarding the use of tracing techniques to evaluate distributed erosion models is that FRN inventories may reflect soil movement due to tillage and other farming operations (Bacchi et al., 2003; Lacoste et al., 2014; Quine et al., 1994), which are not always described in model structures.

Nonetheless, comparisons between tracing derived soil redistribution rates/patterns and erosion model outputs have provided insights into model performance. Some of the most

interesting comparisons have been achieved when multiple erosion models are evaluated, as different models often produce contrasting maps. For instance, He and Walling (2003) demonstrate how the ANSWERS and AGNPS models yielded quite different predictions of erosion and depositions patterns for a field in the UK. While ANSWERS-predicted soil redistribution rates failed to exhibit any relation with ^{137}Cs-estimated rates, AGNPS predictions showed a better visual agreement with the latter, although correlation between rates was still poor ($R^2 = 0.26$). Similarly, Bacchi et al. (2003) tested spatially distributed applications of the USLE and WEPP models against ^{137}Cs-derived soil redistribution rates for a sugar-cane field in Brazil. Results were again contrasting, as models yielded quite different spatial predictions and both compared poorly to the tracer-estimated patterns of erosion and deposition. Moreover, Warren et al. (2005) applied a 3-D enhanced version of the USLE (USPED) to a military training area in the USA. Their results demonstrate how the USPED model provided a better agreement with ^{137}Cs-estimated patterns of soil redistribution than older USLE versions which did not account for in-field sediment deposition. However, the model errors of erosion/deposition rates (tracer estimates were taken as observed values), were – according to the authors – still disappointing (RMSE = 7.96 ± 0.62 ton ha^{-1} yr^{-1}).

Overall, the evaluation of distributed erosion models by the use of tracing techniques indicate that while models can sometimes display a good agreement with tracer-estimated soil redistribution patterns, this is frequently not the case. Moreover, tracer-derived rates of soil erosion and deposition generally compare poorly to model estimates (Bacchi et al., 2003; Belyaev et al., 2004; He and Walling, 2003; Lacoste et al., 2014; Warren et al.,

2005). However, it is difficult to identify whether this is because of errors in the tracing techniques or because of modelling limitations.

Sediment fingerprinting studies, which aim to identify the origin of sediments rather than to model soil redistribution (Guzmán et al., 2013), have also been compared against erosion model estimates. The sediment fingerprinting approach allows for the quantification of the relative contribution of potential upstream sources to sediment yield (see Koiter et al., 2013 and Laceby et al., 2017 for reviews on sediment fingerprinting), which can provide a useful framework for distributed erosion model testing. This requires that sediment sources are stratified in comparable manner to model outputs.

For instance, Wilkinson et al. (2013) employed a fingerprinting approach to model the relative contributions of different erosion processes (i.e., surface or subsurface) to fine sediment loads in the Burdekin River basin, Australia (~130,000 km²). They also identified the spatial origin of the fine sediments reaching catchment outlet by use of a tributary/geological source stratification. The results were compared to a spatially distributed sediment budgeting model (SedNet), which had been previously tested against sediment yield measurements (Wilkinson et al., 2009). However, the fingerprinting and SedNet modelling outputs were contrasting, as the approaches identified different sub-catchments as the main contributors to sediment yield. Moreover, while SedNet results indicated that hillslope erosion (i.e. rill, sheetwash) was responsible for most of the fine sediments reaching catchment outlet, the fingerprinting data demonstrated that gully erosion was the dominant process controlling the basin sediment load. Similarly, Borrelli et al. (2018) compared WaTEM/SEDEM erosion predictions to a fingerprinting study carried out by Alewell et al. (2016). In this case sediment sources were stratified by land use, but

the comparison revealed once again a poor agreement between the independent estimates. Borrelli et al. (2018) supported the model over the fingerprinting data, concluding that "the modelling results seem to reject the validity of [fingerprinting] estimations". If anything, as argued by de Vente et al. (2013), the results from Wilkinson et al. (2013) and Borrelli et al. (2018) highlight how difficult it is for erosion models to identify where sediments originate from and to pinpoint what the dominant erosion processes are, within a catchment. Nevertheless, it is important to note that the fingerprinting approach is also uncertain and ultimately modeled-based. A comparison between erosion and fingerprinting models should explicitly consider the uncertainties present in both.

3.2.3 Comparing soil erosion models to remote sensing images

The approaches to distributed erosion model evaluation described so far have important limitations when these models are applied at a regional scale. This is because the extensive field sampling necessary for tracing techniques might be unattainable. Moreover, the assumption of homogeneous FRN input across large areas would be hardly justifiable. Also, although sediment fingerprinting is frequently applied at watersheds with over 1000 km^2 (see Collins et al., 2017 for some examples); this approach will not always be comparable to model outputs – particularly if model structures do not include a sediment routing component. Field monitoring schemes might also be restrictive at regional scale, considering the time and personnel that would be required to constantly survey potentially thousands of fields.

To overcome these issues, Fischer et al. (2018) developed a semi-quantitative evaluation approach based on the visual interpretation aerial imagery. The concept is similar to some

of the field monitoring approaches previously described (e.g., Prasuhn et al., 2013; Vigiak et al., 2005; Vrieling et al., 2006), as erosion severity classes are assigned according to the visual identification of erosion features. Although Evans and Brazier (2005) combined aerial photographs with field surveys on their evaluation of a spatially distributed version of WEPP, the study of Fischer et al. (2018) is perhaps the first to be fully based on the interpretation of remote sensing images. This enabled the authors to analyze 8100 eroding fields, from which aerial photographs were taken after prominent erosive events. Potentially erosion-causing events were identified based on daily rainfall maps and farmer reports. The assigned erosion severity classes were compared against USLE soil loss estimates for the Bavarian region, in Germany (~ 15,000 km²). Results were encouraging, as the visual erosion classes were highly correlated to predicted soil losses ($R^2 = 0.91$).

It should be highlighted that the model-based regional erosion risk assessment of Fischer et al. (2018) was supported by high resolution rainfall (1km, 5 min) and elevation (5 m) data. Sub-field soil texture measurements and site-specific cropping information were also available for model parameterization. Moreover, much effort has been put into adapting the USLE into German conditions (see Fiener and Auerswald, 2016) and therefore Fischer et al. (2018) were able to make use of suitable parameters to their particular regional settings. Hence, the results from the semi-quantitative approach to model evaluation performed by the authors indicate that simple USLE-type models seem to be capable of identifying eroding fields at regional scale, provided that adequate data is available for parameterization. Although studies such as of Prasuhn et al. (2013) and Fischer et al. (2018) are based on high resolution data, this is not the case for most erosion model applications at regional or large-catchment scale.

3.2.4 What have we learned from these comparisons?

Overall, the lessons learned about distributed erosion model performance based on the described attempts to evaluate them at field, catchment or regional scale can be summarized as: I) modeled-based erosion risk assessments are able to identify the relative rank of erosion-prone fields if high quality data are available for parameterization; II) actual erosion and deposition patterns/rates generally compare poorly to independent estimates; III) the capability of models to identify sediment sources is limited and very rarely evaluated; IV) acquiring independent spatial data for model evaluation is difficult and methods for doing so are subject to considerable uncertainty; V) the more rigorously a model is tested then the more likely poor performance is found.

The latter conclusion (V) might seem somewhat obvious: since all models are approximations, deficiencies will always become evident if models examined in enough detail (Beven and Young, 2013). Nonetheless, defining the type of tests and the sources of data by which a model will be evaluated, as well as the level of agreement one expects between models and observations, are important issues regarding model testing (see Beven and Young, 2013; Beven, 2018). That is, in order to declare a model conditionally useful, or fit-for-purpose, the tests involved in the evaluation approach must be also fit-for-purpose. However, testing erosion models as hypotheses is difficult because of the uncertainties necessarily associated to model structures, parameter estimation, and the observational data to which models are compared to (Beven, 2018). In the next section we review how uncertainty analysis has been incorporated into erosion model evaluation and hypothesis testing. It is our hope, however, that the methodologies described above will help erosion modelers choose sources of data and approaches to model evaluation that will

be more suitable to the purpose of their model application (see Table 3).

4 Uncertainty in soil erosion models

The discussions about model evaluation addressed in this review so far have made the case for the necessity of uncertainty analysis in erosion models. That is, given the limitation of our knowledge regarding the description of soil erosion processes, our inability to represent the variability of parameter values, and the errors associated to erosion measurements; uncertainty and equifinality are necessary consequences of any erosion modelling endeavor.

Still, uncertainty analysis is rarely undertaken. Beven and Brazier (2011) comprehensively reviewed the attempts made by erosion modelers to incorporate uncertainty analysis and declared that the "assessment of uncertainty in soil erosion models is in its infancy". This remains the case.

In order not to repeat or summarize the work of Beven and Brazier (2011), we decided to perform a case study of uncertainty estimation for a simple process-based erosion model. Since we believe one of the reasons not to perform uncertainty analysis stems from the misconception that they are too difficult to implement (see Pappenberger and Beven, 2006), we provided a clear explanation of our case study, along with a simple demonstration code, which has been scripted in the open source programming language R (R Core Team, 2017). But first, a brief description of uncertainty analysis tools that we believe are the most useful for common erosion model applications is warranted.

4.1 Uncertainty estimation methods for soil erosion models

The first step of uncertainty analysis is deciding on an estimation method. Detailed guidelines are provided by Beven (2009), but perhaps the main factor involved in the decision – particularly for erosion models – is the availability of quantitative data for model evaluation.

4.1.1 Forward uncertainty analysis

As we have shown, acquiring spatially distributed data for erosion model testing can be quite challenging. Moreover, outlet sediment fluxes are not always directly comparable to model outputs. Hence, it is frequently the case where no historical data are available for model evaluation. Lack of evaluation data will also be necessarily true for scenario-based simulations and future forecasts, for obvious reasons. In such circumstances, a forward uncertainty analysis can be employed to provide an initial estimate of input error. It is forward because feasible assumptions about model structures and parameter values must be "fed forward" by the modeler (Beven, 2009).

Forward uncertainty analysis of erosion models can be performed by Monte Carlo simulations. In this approach, distributions of uncertain model parameters must be defined a priori, based on replicate measurements, previously reported values, and/or expert judgments. Possible parameter values are then sampled throughout a large number of iterations, which in turn will produce a set of possible model realizations. The distribution of the resulting model realizations is then used to characterize model uncertainty, and the simulations can be extended to allow for sensitivity analysis (e.g. Quinton, 2004). Since forward uncertainty assessments are carried out in the absence of historical data for

evaluation, the estimated model errors will be totally dependent on the assumptions made about prior parameter distributions, parameter co-variation, and model structure (Beven, 2009; Beven and Brazier, 2011). This will necessarily lead to some degree of subjectivity.

Forward uncertainty analysis might be particularly useful for spatially distributed erosion models, which are often applied without any form of evaluation. At the very least, this will allow for some spatial representation of parameterization uncertainty. Although simulation-based error propagation is commonly employed in geostatistics and geoprocessing (e.g., Aerts et al., 2003; Hengl et al., 2010; Heuvelink, 1998; Oksanen and Sarjakoski, 2005; Wechsler and Kroll, 2006), very few studies have fully incorporated such an approach to distributed erosion modelling.

Noteworthy examples of forward uncertainty analysis within a distributed erosion model framework are provided by Biesemans et al. (2000), Van Rompaey and Govers (2002) and Tetzlaff et al. (2013). All studies focused on distributed RUSLE model applications, although in different scales and under different assumptions about parameterization uncertainty. These examples provide an illustration of the subjectivity embedded in forward uncertainty analyses, as we will demonstrate.

Biesemans et al. (2000) applied the RUSLE within a Monte Carlo framework in 1075 ha catchment in Belgium. The rainfall erosivity and the support practice factors (R and P factors of the RUSLE equation, respectively) were held constant, whereas the soil erodibility factor (K), the topographic factor (LS), and the cover management factor (C) were randomly re-sampled from predetermined distributions. This required spatial information on prior parameter distributions, which were acquired by: I) generating auto-

correlated DEM error surfaces for each iteration; II) a K factor kriging variance grid; and III) a land use map combined with minimum and maximum C factor values reported in the literature. The forward uncertainty analysis enabled the authors to provide a mean and a standard deviation soil loss map of the catchment. They also provided percentile error maps of each factor sampled during the simulation, which were used to calculate the contribution of each of these factors to the variance of estimated soil losses. Bisesemans et al. (2000) concluded that the LS factor was the main source of uncertainty in their model, which could be reduced by the use of a higher quality DEM. The authors further "validated" their model based on estimated catchment sediment yields, which were presumably obtained by summing the pixel-based soil loss estimates. The standard deviation of the simulated sediment yields was narrow, as to be expected considering that the sum of the pixel-based model realizations should somewhat converge. Nonetheless, the mean estimated sediment yield showed a good agreement with measured values.

A similar approach to uncertainty analysis was explored by Van Rompaey and Govers (2002) at a 250 ha catchment in Belgium. In this case, however, K factor values were derived from a discrete soil map and by the use of a regression equation which relates geometric mean particle size to soil erodibility. In order to represent the uncertainty of the model parameter, minimum and maximum grain sizes were assigned to specific textural classes in the soil map. For each iteration of the Monte Carlo simulation, a new K factor grid was created based on the sampled grain sizes. Results from the simulation revealed that the soil loss estimates had an average relative error of 111 %. Moreover, a sensitivity analysis performed by the authors indicated that the K factor was the main source of uncertainty in the model application.

The forward uncertainty analysis of Tetzlaff et al. (2013) is somewhat different to the ones previously described. The analysis was employed at a much larger catchment (~ 485 km²) in Germany, which meant that different sources of uncertainty were associated with model parameterization. The authors applied a Monte Carlo simulation to produce mean and standard deviation maps of each RUSLE factor, which were later used to propagate model error analytically. Tetzlaff et al. (2013) did not represent the uncertainty of spatial estimates of the R and K factors, which were assumed to be only associated to measurement errors of rainfall and soil texture. Moreover, the spatial auto-correlation of DEM errors was neglected. This approach led the authors to identify the LS factor as a main source of model uncertainty, and the reported mean relative error of soil loss estimates was of 34 %. These values are lower than the ones reported by Van Rompaey and Govers (2002), which raises the question if the narrower uncertainty bounds are a result of the higher quality of the input data or just a consequence of the different assumptions made about the sources of error.

Overall, the few studies which incorporated forward uncertainty analysis to distributed erosion model applications represent an improvement over the common deterministic approach. However, these studies also illustrate the variations in the uncertainty estimation method: forward error assessments rely entirely on the prior and subjective assumptions made by the modeler. This element of subjectivity could be somewhat attenuated if pessimistic and optimistic assumptions about sources of uncertainty were explored, and if the full distributions of possible model outputs were reported. Nonetheless, testing models against observed empirical data will always be preferable, as in this case the "true" uncertainty of model estimates can be assessed (Beven, 2009). As argued by Oreskes (1998), quantifying input error will not make a structurally flawed model reliable.

4.1.2 Uncertainty analysis in the presence of observational data

When historical quantitative data are available for model evaluation, the Generalized Likelihood Uncertainty Estimation (GLUE) (Beven and Binley, 1992) seems to be the preferred tool for dealing with the uncertainty of soil erosion models (e.g., Brazier et al., 2001, 2000; Cea et al., 2016; Krueger et al., 2012; Quinton, 1994; Quinton et al., 2011; Vigiak et al., 2006a). The GLUE methodology allows for an explicit representation of the uncertainties associated to model structures, parameterization, and to the observational data. For a detailed description of GLUE we will refer to some of the many studies of Beven (1993, 2006, 2012). The basis of the methodology, however, can be summarized in few decision steps (Beven, 2009):

I. Decide on a likelihood measure to evaluate model realizations.
II. Decide on the rejection criteria for non-behavioral model realizations (i.e. not acceptable reproductions of the observational data).
III. Decide which parameters are uncertain.
IV. Decide on a prior distribution to characterize the uncertainty of the chosen parameters.
V. Decide on a simulation method for generating model realizations.

In the GLUE methodology, calibration is not restricted to defining an optimum parameter set that minimizes model error against the observational data. Instead, multiple behavioral parameter sets and model realizations are retained to represent model uncertainty. A difficulty, however, is defining limits of acceptability to declare a model realization as behavioral or not, which is critical to enable models, or model realizations, to be tested as hypotheses (Beven, 2018, 2009). The definition of such limits should reflect our knowledge

about the errors and uncertainties in the observational data used for model evaluation (Beven, 2018). For erosion models being applied at plot scale, we argue that the evaluation criterion of Nearing (2000) provides a framework for defining the limits of acceptability for model errors within the GLUE methodology. This will be demonstrated in the following case study. Although recent erosion modelling efforts have focused on spatially distributed applications, testing models at plot scale is still desirable. Erosion plots provide a reasonably controlled experimental setting, allowing for more detailed parameterization and a greater scrutiny of process descriptions.

4.2 Case study: applying GLUE to the revised Morgan-Morgan-Finey model

The revised MMF (Morgan, 2001) is a simple, but still process-based model, and does not require as many inputs as models such as WEPP or EUROSEM. This makes it suitable for the straightforward uncertainty analysis we undertook with GLUE. Model parameters and operating equations are displayed in Table 4. The model implementation code in R (R Core Team, 2017) and all input data are provided as supplementary material. Full model descriptions are available in Morgan (2001, 2005).

Table 4 Parameters and operating equations for the revised MMF model.

Description	Operating equation	Parameter definitions
Effective rainfall (mm)		R = rainfall (mm) A = proportion of rainfall intercepted by vegetation
Leaf drainage (mm)		C_c = proportion of canopy cover
Direct throughfall (mm)		
Kinetic energy of direct throughfall for tropical climates (J m^{-2})		I = typical rainfall intensity value for erosive rain (mm h^{-1})
Kinetic energy of leaf drainage (J m^{-2})	$\left(15.8\, P_h^{0.5}\right) -$	P_h = plant canopy height (m)

Total kinetic energy (J m^{-2})		
Annual runoff (mm)	—	R_o = mean rain per day (mm)
Soil moisture storage capacity (mm)	$\sqrt{}$	M_c = soil moisture content at field capacity (% w w^{-1}) B_d = bulk density of the soil (Mg m^{-3}) H_d = effective hydrological depth (m) E_t/E_o = ratio of actual to potential evapotranspiration
Soil particle detachment by raindrop impact (kg m^{-2})		K = soil detachability index (g J^{-1})
Soil particle detachment by runoff (kg m^{-2})		S = slope steepness (˚) G_c = proportion of ground cover
Resistance of the soil	—	σ = soil cohesion (kPa)
Runoff transport capacity (kg m^{-2})		C = product of the C and P factors of the USLE

Sources: Morgan (2001, 2005)

The model was applied at two set of replicate plots, which were part of an erosion monitoring experiment at the Lavras Federal University, Brazil (Lima et al., 2018). Soils in the area are classified as Typic Hapludoxes (Soil Survey Staff, 2014) and the topsoil texture (20 cm) is sandy clay. According to the Köppen classification system, the climate is humid subtropical (Cwa), with dry winters and temperate summers. Average rainfall is ~ 1500 mm.

Soil losses were monitored during one cropping season, between December 2013 and April 2014. Three plots (4 m wide and 24 m long) were left bare and kept free of vegetation by manual hoeing. Another three plots (4 m wide and 12 m long) were cultivated with maize, which was sown manually and perpendicularly to the slope. Neither set of plots was ploughed or tilled. All plots were isolated by galvanized metal sheets, which transported runoff and sediments to collection tanks at the bottom of the slope. After each runoff event,

soil and water losses were determined.

The model application within the GLUE methodology was performed under two different scenarios. For scenario I, all parameters considered uncertain were allowed to vary across the full range of possible values reported in the MMF guidelines, regardless of a strict physical meaning. For instance, the possible values of land cover parameters, such as the percentage of canopy cover (C_C) or the percentage of ground cover (G_C), were set from zero to one even for the bare soil plots. This scenario represents model calibration, or conditioning, under a loose belief in the correctness of the physical equations represented by the model (Pappenberger and Beven, 2006). For scenario II, actual measurements of parameter values (e.g. bulk density, soil moisture at field capacity, canopy cover, and plant height) were used to construct prior parameter distributions. If measurements were unattainable (e.g. effective hydrological depth, soil cohesion, soil detachability index), minimum and maximum values were set according to our interpretation of model guidelines, but still allowing for some uncertainty in the estimates. This second scenario represents model conditioning under the assumption that parameter values should not be calibrated outside the range of a feasible physical meaning. It also represents an attempt to constrain model uncertainty.

Model realizations for both scenarios were generated by uniform random sampling, using uniform prior parameter distributions and a Monte Carlo simulation with 10^6 iterations. According to Beven (2009), this approach enables the identification of scattered regions of behavioral model realizations within the response surface.

Before the simulations were performed we decided on a rejection criterion for defining

model realizations as non-behavioral. Our criterion is the one of Nearing (2000), which states that "if the difference between the model prediction and the measured value lies within the population of differences between the measured data pairs, then the model reasonably reflects the erosion for that population". Nearing (2000) used a large number of replicate storm events (2061) and annual soil losses (797) to calculate the 95 % occurrence interval of the relative difference in soil losses between replicates ($Rdiff_{occ}$):

where:

m = 0.236 and b = - 0.641 for the lower limit of the 95 % interval;

m = -0.179 and b = 0.416 for the upper limit of the 95 % interval;

M = measured erosion rate (kg m^{-2}) (in our case this corresponds to the mean soil losses observed in the three replicate plots for each treatment – bare and maize).

Hence, if the relative difference of simulated and observed erosion rates laid outside the above defined occurrence intervals, the model realization was considered non-behavioral. This approach allows for a representation of the errors involved in soil loss measurements at plot scale, and also incorporates the variability of these errors according to the magnitude the measured erosion rates. Therefore, the approach enables model rejection: if none of the simulations are within the limits of acceptability then the model itself should be rejected as non-behavioral under the testing conditions.

Behavioral model realizations were assigned a likelihood measure according to the absolute error of the simulations in relation to the measured soil losses. Similarly to Brazier et al.

(2000), likelihoods were calculated by rescaling the absolute errors so that their sum would add up to one and that those simulations with the lowest errors were assigned a greater likelihood. Formulae are provided in the supplementary material code.

Results from the analysis indicate that Nearing's criterion for defining behavioral models were strict enough to eliminate poor simulations, but still retained a large number of acceptable model realizations. For the bare plots, 19 % and 33 % of sampled parameter sets in scenarios I and II, respectively, yielded behavioral model realizations. For the maize plots, these values changed to 48 % and 13 %. As the measured soil loss rates for the maize plots were lower than for the bare plots (mean bare = 1.774 kg m^{-2}, mean maize = 0.265 kg m^{-2}), thresholds of model acceptability were relatively wider in the first case. This is because equation 1 incorporates the higher uncertainty of low erosion rate measurements at plot scale.

Due to the degree of freedom afforded to the model, simulations from scenario I were able to encompass the observed data in both sets of plots, as expected (Figure 4). Model output realizations are spread throughout the behavioral response surface and part of them overlap the measured soil losses. Not much can be concluded from these results, and the obvious next step would be to evaluate the conditioned parameter sets against new observational data.

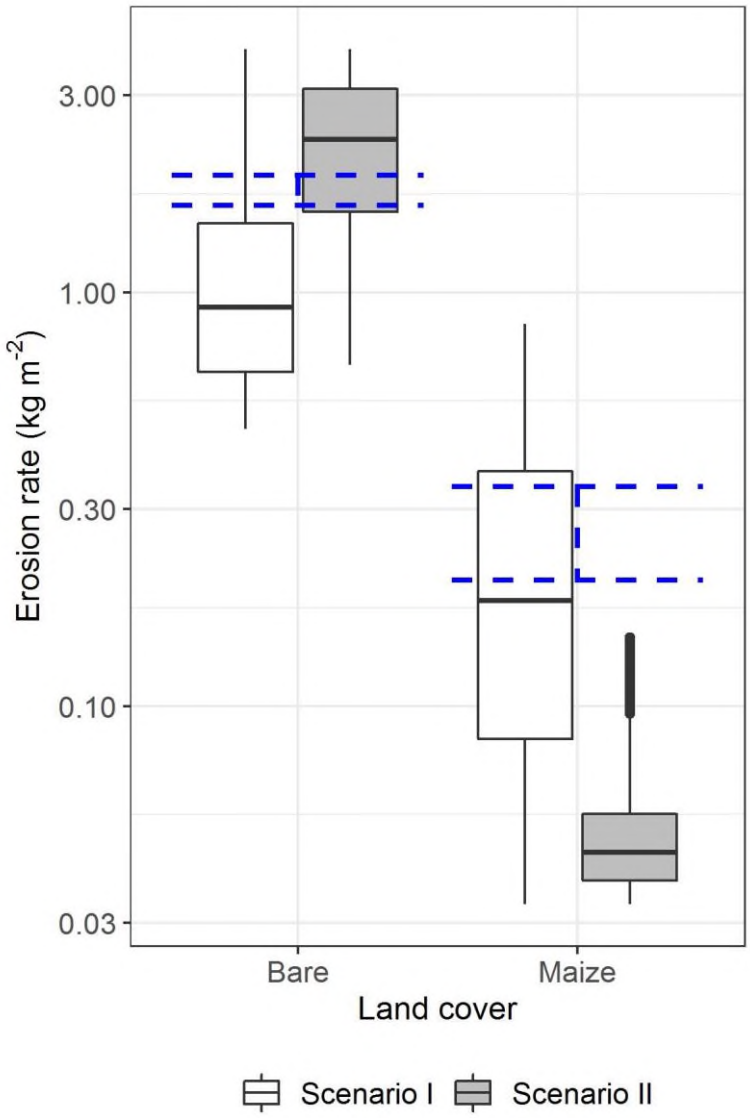


Fig. 4. Estimated erosion rates of behavioral realizations of the MMF model for the bare and maize plots. Blue dashed lines represent the range of observed soil losses for the replicate plots within each treatment.

Results from scenario II are more interesting. For the bare plots, simulations from the reduced parameter space do not systematically underestimate the observational data, as in the case of the Scenario I, and a greater part of behavioral models encompass the measured soil losses. By plotting individual parameter values against the rescaled likelihood measure, it was clear that more accurate results could be achieved if the range of sampled soil

detachability index (K) values was narrowed (Figure 5). Whether or not this would result in more accurate predictions for new observational data remains to be tested.

For the maize plots, the reduced parameter space from Scenario II considerably narrowed the spread of the behavioral models response surface. However, none of the simulations encompassed the observational data. That is, if model parameters were set according to actual measurements of soil properties and land cover characteristics, the model consistently underestimated the measured soil loss rates. The poor results appear to be caused by an underestimation of runoff transport capacity, as illustrated by the greater likelihoods associated to higher values of the USLE C and P factors, as well as to the lower values of parameters used in the calculation of soil moisture storage capacity (e.g., soil moisture, bulk density, and effective hydrological depth) (see Figure 5 and Table 2). Since estimated erosion rates seemed to be transport limited, model outputs were little sensitive to parameters associated the prediction of particle detachment (e.g., soil detachability index and rainfall intensity). Although the model application itself cannot be rejected, as many realizations were considered behavioral, this systematic underestimation within the conditioning period raises concerns about the potential usefulness of model predictions under the testing conditions (see Beven, 2009). These results illustrate how difficult it can be for erosion models to make accurate estimates while trying to constrain output uncertainty. Although these results are certainly case-specific, our experience indicates that similar problems might expected elsewhere (see Quinton, 1997).

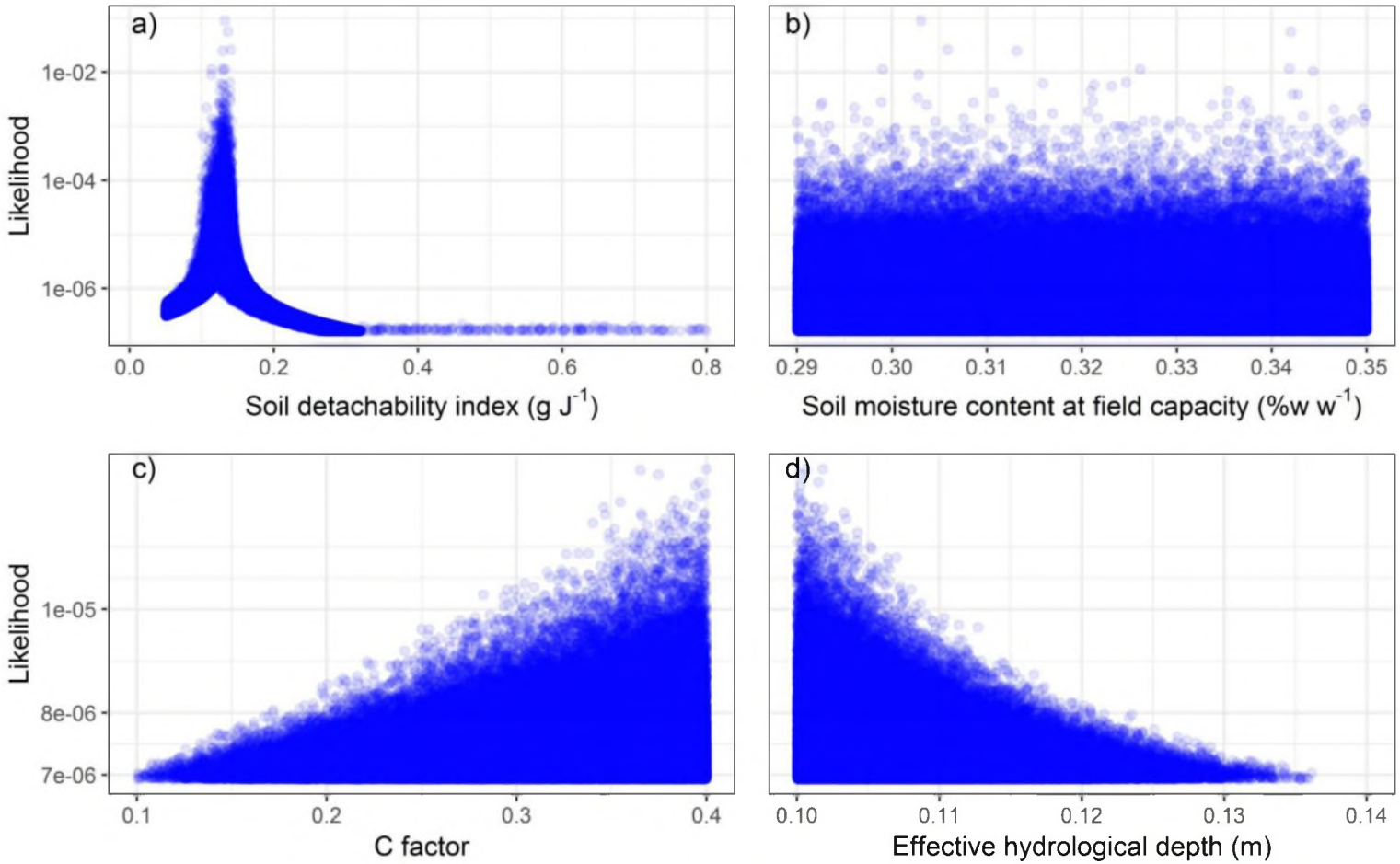


Fig. 5. Dotty plots of behavioral model realizations for the simulations from Scenario II in the bare (a,b) and maize plots (c,d). Each point relates a sampled parameter value to the rescaled likelihood of the model realization. High-sensitivity parameters, such as the soil detachability index, have higher likelihoods associated to a narrow parameter space. Contrarily, low-sensitivity parameters, such as soil moisture content at field capacity display variable likelihood values across the parameter space.

In summary, our case study demonstrates how Nearing's criterion can be incorporated into erosion model testing at plot scale within the GLUE methodology. This approach provides an objective definition of the limits of acceptability of model error, which is critical to enable models to be tested as hypotheses considering the uncertainties in both models and the observational data. We have provided a simple demonstration of erosion model conditioning while dealing with uncertainty and equifinality, which allows for a more realistic and forthright characterization of model performance than a single optimized parameter set. It is our sincere hope that the example herein implemented can be expanded and improved by other modelers, and that this review as a whole will be an incentive for model evaluation in face of the limitations of our knowledge.

5 A way forward for the evaluation of soil erosion models

This review has taken a somewhat critical perspective on the evaluation of soil erosion models and erosion modelling in general. This is not meant to discredit previous work, but instead to raise awareness about the necessity of continuous model testing. Moreover, we have focused on the limitations of the reviewed approaches to model evaluation. This is meant to enable modelers to make informed decisions about the tests and sources of data that should be more suitable for evaluating erosion models according to the context of their application.

It is our opinion that the way forward for erosion model evaluation involves pursuing fit-for-purpose tests according to the finality of the model applications (see Jakeman et al., 2006). Such tests should encompass multiple lines of evidence, should consider the uncertainties in model structures, parameter estimation, and the observational data.

Moreover, evaluation should allow for a broad investigation regarding the usefulness and consistency of the models, as we explain below.

When deciding on an evaluation methodology, the purpose of the modelling should be explicit. This will allow the modeler to pursue sources of data that will investigate the usefulness of the model according to the pre-defined application purpose (see Table 3). For instance, if a model is being used to simulate the impacts of land use changes on sediment yields at catchment scale, it is desirable that such model is not only able to make reliable quantitative predictions of sediment transport rates, but also to identify the spatial provenance of sediment sources. Moreover, catchment outlet responses should be sensitive to land use model parameters. Investigating the usefulness of a model for such purpose could involve a sensitivity analysis and a comparison between model outputs against sediment yield measurements and sediment fingerprinting source apportionments.

Erosion models are necessarily uncertain, and so are the observational data used for evaluation; and as such, models cannot be tested as hypotheses if uncertainty is not accounted for. Although a strict Popperian falsification of environmental models is somewhat useless, as all models are ultimately wrong, we feel the erosion modelling community would benefit by some degree of model rejection. That is, given the profusion of available soil erosion models, which are in theory able to accomplish the same task, how does one choose an appropriate model for a given purpose? Tests that allow for models to be rejected as not fit-for-purpose are therefore encouraged. We have supplied an example of how this can be achieved with GLUE, and further discussions on the matter can be found in Beven (2018).

Furthermore, we believe that taking a collaborative fit-for-purpose rejectionist approach is important from a public policy and decision-making perspective. Co-development of limits of acceptability and satisfactory uncertainty bands between modelers and decision-makers is necessary if we are to have tools and predictions that meet stakeholder needs whilst formally acknowledging observational errors (Beven and Binley, 2014). If an erosion model is required to support decision-making and no historical data are available for testing, it is still possible to provide a forward uncertainty analysis to give an initial assessment of model error. In this case, modelers should clearly justify the assumptions made about the sources of uncertainty.

Quantifying input errors will not lead to reliable predictions if the model itself is structurally flawed; however, it might help delineate what inferences can be made from model outputs. For instance, Alewell et al. (in press) have recently argued that large-scale erosion model applications should not strive to make accurate predictions of soil losses, but instead to explore scenarios and system reactions, focusing on understanding relative differences of erosion rates. Whether this premise is accepted or not, it is important to note that if models are applied deterministically, even simple conclusions regarding relative differences of erosion rates might be misleading. For example, policymakers might be prone to subsidize a given set of agricultural practices if a model depicts that this would lead to a 20 % decrease in regional gross erosion rates. However, they might want to consider different options if model results indicate there is a 50 % chance that adopting such practices will reduce soil losses in 20 %.The same policymakers might have even more concerns if it is made clear that these errors are only associated to parameter estimation, and that no case-specific quantitative/representative data are available to

corroborate model predictive accuracy. In summary, the modelling community needs to take responsibility for analyzing model limitations and uncertainties, and co-developing evaluation criteria that are fit-for-purpose with the end-user.

However, situations may arise in which the uncertainties in model estimates and in the observational data are so large that the response surface of model realizations will almost always overlap the empirical observations. This was somewhat illustrated in our case study, and similar outcomes have been reported by others (e.g., Banis et al., 2004; Janes et al., 2018). Then how to proceed? A logical conclusion would be to constrain uncertainty, by simplifying models and increasing measurement precision. But to what extent is this possible? Although technological developments continuously improve our ability to measure model parameters and system responses, the very things we call data are inference-laden signifiers of a reality we cannot fully access (Oreskes et al., 1994). In this sense, any real-life/open-system model test involves a number of embedded hidden assumptions, many of which are poorly understood or completely unknown (Baker, 2017; Oreskes, 1998). Hence, even when models are not rejected, is it possible to know if this is because of the quality of model process descriptions or to any of these assumptions?

A complement to model-testing-as-hypotheses is as an investigative/exploratory approach; in which hypotheses are pursued to generate knowledge, instead of to test theories (see Baker, 2017 for a complete philosophical discussion). This involves embracing uncertainty as a necessary motivation of science-as-seeking, and exploring observational data not as hard substitutes of phenomena, but as signs through which the world communicates to the investigator (Baker, 2000, 2017). In this approach, investigating the overall consistency of a model as a narrative is more important than testing individual hypotheses as propositions

(Baker, 2017).

According to Baker (2017), a hypothesis is consistent when it explains the cause of a system response without contradicting physical principles, spatial evidence of related phenomena, or other similar relationships. For instance, Pontes (2017) tested the SWAT model in a small mountainous catchment in Brazil. The model was applied in a stochastic framework, and estimates of outlet sediment transport rates were considered acceptable for both the conditioning and the evaluation period. However, a comparison against erosion plot measurements revealed that hillslope erosion rates were overestimated. Accurate sediment yield predictions were only possible because the model simulated a large sediment channel deposition. This was not *consistent* with the catchment characteristics or with the other lines of evidence investigated by the author.

Regardless of how testing models as hypotheses is perceived, it should be clear that environmental models cannot be verified or validated, and the use of such terminology is misleading. Semantics have been thoroughly discussed by others (e.g., Beven and Young, 2013; Oreskes et al., 1994; Oreskes, 1998), but the considerations made throughout this review have demonstrated how models are an incomplete descriptions of not fully accessible phenomena. Erosion models are therefore necessarily neither true nor free of apparent flaws, and therefore cannot be strictly valid. Although these issues have been recognized for a long time, the validation terminology still prevails, as demonstrated by our term co-occurrence analysis. As argued by Oreskes (1998), although the primary problems of model evaluation are not one of linguistic, “the language of validation buries uncertainty; as scientists, we should be doing the opposite”.

In a broader sense, changing the terms with which we describe model evaluation is a step towards to something we understand is necessary to improve soil erosion modelling, which is a change in attitude regarding model testing. As we have shown, erosion model evaluation is often neglected and/or restricted to a deterministic "validation" based on system outlet responses, even at catchment scale and regardless of the purpose of the application, in spite of the overwhelming criticism on the matter (Brazier et al., 2001; Favis-Mortlock et al., 2001; Fiener and Auerswald, 2016; Govers, 2011; Jetten et al., 2003; Takken et al., 1999). Although focusing on tests that are designed to prove a model right may promote acceptance and the status/authority of the modeler, "this [approach] makes learning difficult and ultimately erodes the impact of the model and the credibility of the modeler – and of all modelers" (Sterman, 2002). Instead, a purpose-oriented critical model evaluation approach, which focuses on model deficiencies, encompasses multiple sources of data, and fully acknowledges uncertainty and equifinality, will ultimately lead to model improvements and responsible decision-making.

6 Conclusions

If soil erosion models are to influence decision-making in matters of public interest, the level of disagreement between models and reality must be clear. Ultimately, comprehensive knowledge of model performance can only be acquired by rigorous evaluation, which means that erosion models must be thoroughly and continuously tested. Our term co-occurrence analysis demonstrates that currently they are not.

Moreover, the meta-analysis we undertook on erosion model performance indicated that different models do not systematically exceed each other regarding their predictive

accuracy. In fact, calibration appears to be the main mechanism of improvement of model performance for estimating soil losses. We have argued that results from calibrated models are only interpretable within the very specific systems they have been calibrated to. Given the conditional nature of parameter optimization and capability calibrated models to make accurate predictions for the wrong reasons, their results should be viewed with some caution. Hence, when dealing with erosion models that require calibration, modelers should formally recognize that equifinality is a necessary consequence of model conditioning in face of the uncertainties associated to models and observational data. We have provided an example of how this can be performed with GLUE.

We have also argued that evaluating spatially distributed models requires representative spatially distributed data. Our review has demonstrated that, in general, model-based estimates of erosion and deposition rates do not compare well to independent spatial data. However, we have shown how difficult and uncertain it is to measure soil redistribution rates across landscapes. Therefore, we stress that comparisons between model-based estimates and observational data requires being explicit about the uncertainties present in both. This literature review indicates that unless corroborative evidence is presented by modelers, results from spatially distributed soil erosion models should be perceived with a healthy dose of skepticism – even if they provide satisfactory estimates of catchment sediment yields. It is our opinion that corroborative evidence should be consistent with the purpose of the model application. Hence, we have provided guidelines that will help modelers to pursue sources of data to evaluate models according to the purpose, scale, and the structure of common erosion modelling applications.

Finally, we would like to remember why we are modelling soil erosion in the first place. Soil erosion is a threat to food and water security, and its deleterious effects in society have been well documented throughout the history of mankind (Montgomery, 2007). In face of the rising demands for agricultural production and the concerns regarding climate change (see Davies, 2017), models that enable us to understand how soil erosion, and all its negative consequences, will respond to the uncertain future ahead are increasingly necessary.

Although action is needed, informed decision-making requires being explicit about the limitations of our knowledge (see Sterman, 2002). This review has shown that we, soil erosion modelers, have all too often failed to communicate the uncertainties in our models and to provide sufficient evidence to corroborate their usefulness. Owning up to this failure, improving our attitude towards model evaluation, and changing the way we characterize and communicate model performance will ultimately lead to a better understanding of soil erosion. More importantly, it might help to build the much-needed confidence to solve real-world problems that affect real people – often the most vulnerable – and their livelihoods.

Acknowledgements

This study was funded in part by the Coordination of Improvement of Higher Level Education Personnel – CAPES (process number 88881.190317/2018-01), the National Counsel of Technological and Scientific Development – CNPq (process numbers 306511-2017-7 and 202938/2018-2), and the Minas Gerais State Research Foundation – FAPEMIG (process numbers CAG-APQ-01053-15 and APQ-00802-18). Comments from the editor and two anonymous reviewers were highly appreciated and improved the quality of this

article. We are also thankful to Trevor Page and Barry Hankin for taking a look at our dotty plots and sharing their valuable insights.

Data Availability

The supplementary material, including all raw data and model codes related to this article have been uploaded to the data repository.

References

Aerts, J.C.J.H., Heuvelink, G.B.M., Goodchild, M.F., 2003. Accounting for spatial uncertainty in optimization with spatial decision support systems. Trans. GIS 7, 211–230. doi:10.1111/1467-9671.00141

Alewell, C., Birkholz, A., Meusburger, K., Schindler Wildhaber, Y., Mabit, L., 2016. Quantitative sediment source attribution with compound-specific isotope analysis in a C3 plant-dominated catchment (central Switzerland). Biogeosciences 13, 1587–1596. doi:10.5194/bg-13-1587-2016

Alewell, C., Borrelli, P., Meusburger, K., Panagos, P. In press. Using the USLE: Chances, challenges and limitations of soil erosion modelling. Int. Soil Water Conserv. Res. doi:10.1016/j.iswcr.2019.05.004

Amore, E., Modica, C., Nearing, M. A, Santoro, V.C., 2004. Scale effect in USLE and WEPP application for soil erosion computation from three Sicilian basins. J. Hydrol. 293, 100–114. doi:10.1016/j.jhydrol.2004.01.018

Amorim, R.S.S., da Silva, D.D., Pruski, F.F., de Matos, A.T., 2010. Avaliação do desempenho dos modelos de predição da erosão hídrica USLE, RUSLE e WEPP para diferentes condições edafoclimáticas do Brasil. Eng. Agríc. Jaboticabal 30, 1046–1059.

Anache, J.A.A., Flanagan, D.C., Srivastava, A., Wendland, E.C., 2018. Land use and climate change impacts on runoff and soil erosion at the hillslope scale in the Brazilian Cerrado. Sci. Total Environ. 622–623, 140–151. doi:10.1016/j.scitotenv.2017.11.257

Arnold, J.G., Srinivasan, R., Muttiah, R.S., Williams, J., 1998. Large area hydrologic modeling and assessment part I: model development. J. Am. Water Resour. Assoc. 34, 73–89. doi:10.1016/S0899-9007(00)00483-4

Bacchi, O.O.S., Reichardt, K., Sparovek, G., 2003. Sediment spatial distribution evaluated by three methods and its relation to some soil properties. Soil Tillage Res. 69, 117–125. doi:10.1016/S0167-1987(02)00133-2

Bagarello, V., Ferro, V., Giordano, G., Mannocchi, F., Todisco, F., Vergni, L., 2013. Predicting event soil loss from bare plots at two Italian sites. Catena 109, 96–102. doi:10.1016/j.catena.2013.04.010

Bagarello, V., Piazza, G.A., Ferro, V., Giordano, G., 2008. Predicting unit plot soil loss in Sicily, south Italy V. Hydrol. Process. 22, 586–595. doi:10.1002/hyp

Bailer-Jones, D.M., 2009. Scientific Models in Philosophy of Science. University of Pittsburgh Press, Pittsburgh.

Baker, V.R., 2017. Debates - Hypothesis testing in hydrology: Pursuing certainty versus pursuing uberty. Water Resour. Res. 53, 1770–1778. doi:10.1002/2016WR020078.Received

Baker, V.R., 2000. Let Earth speak! In: Sneiderman, J.S. (Ed.), The Earth around us: Maintaining a livable planet, W. H. Freeman, New York, pp. 358-367.

Balaguer-Puig, M., Marqués-Mateu, Á., Lerma, J.L., Ibáñez-Asensio, S., 2018. Quantifying small-magnitude soil erosion: Geomorphic change detection at plot scale. L. Degrad. Dev. 29, 825–834. doi:10.1002/ldr.2826

Banis, Y.N., Bathurst, J.C., Walling, D.E., 2004. Use of caesium-137 data to evaluate SHETRAN simulated long-term erosion patterns in arable lands. Hydrol. Process. 18, 1795–1809. doi:10.1002/hyp.1447

Beasley, D.B., Huggins, L.F., 1982. ANSWERS – Users manual. In: EPA-905/9-82-001, USEPA, Region 5. IL: Chigago.

Belyaev, V.R., Wallbrink, P.J., Golosov, V.N., Murray, A.S., Sidorchuk, A.Y., 2004. A comparison of methods for evaluating soil redistribution in the severely eroded Stavropol region, southern European Russia. Geomorphology 65, 173–193. doi:10.1016/j.geomorph.2004.09.001

Beven, K.J., 2018. On hypothesis testing in hydrology: Why falsification of models is still a really good idea. WIREs Water 5, e1278. doi:10.1002/wat2.1278

Beven, K.J., 2012. Rainfall-Runoff Modelling, 2nd ed. John Wiley & Sons, Chichester.

Beven, K.J., 2009. Environmental Modelling: An Uncertain Future, Environmental Modelling : An Uncertain Future ? Routledge, Oxon.

Beven, K.J., 2006. A manifesto for the equifinality thesis. J. Hydrol. 320, 18–36. doi:10.1016/j.jhydrol.2005.07.007

Beven, K.J., 1993. Prophecy , reality and uncertainty in distributed hydrological modelling. Adv. Water Resour. 16, 41–51.

Beven, K.J., Binley, A., 2014. GLUE: 20 years on. Hydrol. Process. 28, 5897-5918.

Beven, K.J., Binley, A., 1992. The future of distributed models: Model calibration and uncertainty prediction. Hydrol. Process. 6, 279–298. doi:10.1002/hyp.3360060305

Beven, K.J., Young, P., 2013. A guide to good practice in modeling semantics for authors and referees. Water Resour. Res. 49, 5092–5098. doi:10.1002/wrcr.20393

Biesemans, J., Meirvenne, M. Van, Gabriels, D., 2000. Extending the RUSLE with the Monte Carlo error propagation technique to predict longlterm. J. Soil Water Conserv. 55, 35–42.

Boardman, J., 2006. Soil erosion science: Reflections on the limitations of current approaches. Catena 68, 73–86. doi:10.1016/j.catena.2006.03.007

Borrelli, P., Meusburger, K., Ballabio, C., Panagos, P., Alewell, C., 2018. Object-oriented soil erosion modelling: A possible paradigm shift from potential to actual risk assessments in agricultural environments. L. Degrad. Dev. 29, 1270–1281. doi:10.1002/ldr.2898

Brazier, R.E., Beven, K.J., Anthony, S.G., Rowan, J.S., 2001. Implications of model uncertainty for the mapping of hillslope-scale soil erosion predictions. Earth Surf. Process. Landforms 26, 1333–1352.

Brazier, R.E., Beven, K.J., Freer, J., Rowan, J.S., 2000. Equifinality and uncertainty in physically based soil erosion models: application of the GLUE methodology to WEPP - the Water Erosion Prediction Project - for sites in the UK and USA. Earth Surf. Process. Landforms 25, 825–845.

Bulygina, N.S., Nearing, M.A., Stone, J.J., Nichols, M.H., 2018. DWEPP: a dynamic soil erosion model based on WEPP source terms. Earth Surf. Process. Landforms 54, 171–178. doi:10.1002/esp

Castillo, C., Pérez, R., James, M.R., Quinton, J.N., Taguas, E. V., Gómez, J. A., 2012. Comparing the Accuracy of Several Field Methods for Measuring Gully Erosion. Soil Sci. Soc. Am. J. 76, 1319. doi:10.2136/sssaj2011.0390

Cea, L., Legout, C., Grangeon, T., Nord, G., 2016. Impact of model simplifications on soil erosion predictions: Application of the GLUE methodology to a distributed event-based model at the hillslope scale. Hydrol. Process. 30, 1096–1113. doi:10.1002/hyp.10697

Collins, A.L., Pulley, S., Foster, I.D.L., Gellis, A., Porto, P., Horowitz, A.J. 2017. Sediment source fingerprinting as an aid to catchment management: A review of the current state of knowledge and a methodological decision-tree for end-users. J Environ Manage 19, 86–118.

Davies, J., 2017. The business case for soil. Nature 543, 309–311. doi:10.1038/543309a

De Roo, A.P.J., Offermans, R.J.E., Cremers, N.H.D.T., 1996a. LISEM: a single-event, physically based hydrological and soil erosion model for drainage basins. II: sensitivity analysis, validation and application. Hydrol. Process. 10, 1119–1126. doi:10.1002/(SICI)1099-1085(199608)10:8<1119::AID-HYP416>3.0.CO;2-V

De Roo, A.P.J., Wesseling, C.G., Ritsema, C.J., 1996b. Lisem: a single-event physically based hydrological and soil erosion model for drainage basins. I: theory, input and output. Hydrol. Process. 10, 1107–1117. doi:10.1002/(SICI)1099-1085(199608)10:8<1107::AID-HYP415>3.0.CO;2-4

de Vente, J., Poesen, J., 2005. Predicting soil erosion and sediment yield at the basin scale: Scale issues and semi-quantitative models. Earth-Science Rev. 71, 95–125. doi:10.1016/j.earscirev.2005.02.002

de Vente, J., Poesen, J., Verstraeten, G., Govers, G., Vanmaercke, M., Van Rompaey, A., Arabkhedri, M., Boix-Fayos, C., 2013. Predicting soil erosion and sediment yield at regional scales: Where do we stand? Earth-Science Rev. 127, 16–29. doi:10.1016/j.earscirev.2013.08.014

Desmet, P.J.J., Govers, G., 1997. Two-dimensional modelling of the within-field variation in rill and gully geometry and location related to topography. Catena 29, 283–306.

Di Stefano, C., Ferro, V., Pampalone, V., 2017. Applying the USLE Family of Models at the Sparacia (South Italy) Experimental Site. L. Degrad. Dev. 28, 994–1004. doi:10.1002/ldr.2651

Djuma, H., Bruggeman, A., Camera, C., Zoumides, C., 2017. Combining Qualitative and Quantitative Methods for Soil Erosion Assessments: An Application in a Sloping Mediterranean Watershed, Cyprus. L. Degrad. Dev. 28, 243–254. doi:10.1002/ldr.2571

Dotterweich, M., 2013. The history of human-induced soil erosion: Geomorphic legacies, early descriptions and research, and the development of soil conservation—A global synopsis. Geomorphology 201, 1–34. doi:10.1016/j.geomorph.2013.07.021

Eekhout, J.P.C., Terink, W., de Vente, J., 2018. Assessing the large-scale impacts of environmental change using a coupled hydrology and soil erosion model. Earth Surf. Dyn. Discuss. 1–27. doi:10.5194/esurf-2018-25

Evans, R., Brazier, R., 2005. Evaluation of modelled spatially distributed predictions of soil erosion by water versus field-based assessments. Environ. Sci. Policy 8, 493–501. doi:10.1016/j.envsci.2005.04.009

Favis-Mortlock, D., Boardman, J., MacMillan, V., 2001. The limits of erosion modeling. In: Harmon, R.S., Doe, W.W. (Eds), Landscape Erosion and Evolution Modeling. Springer, Boston, pp. 477-516,

Fernández, C., Vega, J. A., 2016. Evaluation of RUSLE and PESERA models for predicting soil erosion losses in the first year after wildfire in NW Spain. Geoderma 273, 64–72. doi:10.1016/j.geoderma.2016.03.016

Fernández, C., Vega, J.A., 2018. Evaluation of the RUSLE and disturbed WEPP erosion models for predicting soil loss in the first year after wildfire in NW Spain. Environ. Res. 165, 279–285. doi:10.1016/j.envres.2018.04.008

Fernández, C., Vega, J.A., Vieira, D.C.S., 2010. Assessing soil erosion after fire and rehabilitation treatments in NW Spain: Performance of RUSLE and revised Morgan-Morgan-Finney models. L. Degrad. Dev. 21, 58–67. doi:10.1002/ldr.965

Fernandez, C., Wu, J.Q., Mccool, D.K., Stockle, C.O., 2003. Estimating water erosion and sediment yield with GIs , RUSLE , and SEDD. J. Soil Water Conserv. 58, 128–136.

Ferro, V., Di Stefano, C., Giordano, G., Rizzo, S., 1998. Sediment delivery processes and the spatial distribution of caesium-137 in a small Sicilian Basin. Hydrol. Process. 12, 701–711. doi:10.1002/(SICI)1099-1085(19980430)12:5<701::AID-HYP631>3.0.CO;2-L

Fiener, P., Auerswald, K., 2016. Comment on "The new assessment of soil loss by water erosion in Europe" by Panagos et al. (Environmental Science & Policy 54 (2015) 438–447). Environ. Sci. Policy 54. doi:10.1016/j.envsci.2015.12.012

Fiener, P., Wilken, F., Aldana-Jague, E., Deumlich, D., Gómez, J.A., Guzmán, G., Hardy, R.A., Quinton, J.N., Sommer, M., Van Oost, K., Wexler, R., 2018. Uncertainties in assessing tillage erosion – How appropriate are our measuring techniques? Geomorphology 304, 214–225. doi:10.1016/j.geomorph.2017.12.031

Fischer, F.K., Kistler, M., Brandhuber, R., Maier, H., Treisch, M., Auerswald, K., 2018. Validation of official erosion modelling based on high-resolution radar rain data by aerial photo erosion classification. Earth Surf. Process. Landforms 43, 187–194. doi:10.1002/esp.4216

Flanagan, D.C., Frankenberger, J.R., 2012. WEPP: model use, calibration and validation. Trans. ASABE 55, 1463–1477.

Flanagan, D.C., Nearing, M.A., 1995. USDA - Water Erosion Prediction Project: Hillslope profile and watershed model documentation. NSERL Report No . 10, USDA-ARS National Soil Erosion Research Laboratory, West Lafayette, Indiana.

Govers, G., 2011. Misapplications and misconceptions of erosion models. In: Morgan, R.P.C., Nearing, M.A. (Eds), Handbook of erosion modelling, Blackwell Publishing Ltd., Chichester, pp. 117-134.

Guzmán, G., Quinton, J.N., Nearing, M. A., Mabit, L., Gómez, J. A., 2013. Sediment tracers in water erosion studies: current approaches and challenges. J. Soils Sediments 13, 816–833. doi:10.1007/s11368-013-0659-5

He, Q., Walling, D.E., 2003. Testing distributed soil erosion and sediment delivery models using137Cs measurements. Hydrol. Process. 17, 901–916. doi:10.1002/hyp.1169

Hengl, T., Heuvelink, G.B.M., Van Loon, E.E., 2010. On the uncertainty of stream networks derived from elevation data: The error propagation approach. Hydrol. Earth Syst. Sci. 14, 1153–1165. doi:10.5194/hess-14-1153-2010

Hessel, R., van den Bosch, R., Vigiak, O., 2006. Evaluation of the LISEM soil erosion model in two catchments in the East African Highlands. Earth Surf. Process. Landforms 31, 469–486. doi:10.1002/esp.1280

Heuvelink, G.B.M., 1998. Error Propagation in Enviromental Modeling with GIS, Taylor & Francis.

Jain, P., Ramsankaran, R.A.A.J., 2018. GIS-based modelling of soil erosion processes using the modified-MMF (MMMF) model in a large watershed having vast agro-climatological differences. Earth Surf. Process. Landforms 43, 2064–2076. doi:10.1002/esp.4372

Jakeman, A.J., Letcher, R.A., Norton, J.P., 2006. Ten iterative steps in development and evaluation of environmental models. Environ. Model. Softw. 21, 602–614. doi:10.1016/j.envsoft.2006.01.004

Janes, V., Holman, I., Birkinshaw, S., O'Donnell, G., Kilsby, C., 2018. Improving bank erosion modelling at catchment scale by incorporating temporal and spatial variability. Earth Surf. Process. Landforms 43, 124–133. doi:10.1002/esp.4149

Jetten, V., de Roo, A., Favis-Mortlock, D., 1999. Evaluation of field-scale and catchment-scale soil erosion models. Catena 37, 521–541. doi:10.1016/S0341-8162(99)00037-5

Jetten, V., Govers, G., Hessel, R., 2003. Erosion models: Quality of spatial predictions. Hydrol. Process. 17, 887–900. doi:10.1002/hyp.1168

Kinnell, P.I.A., 2017. A comparison of the abilities of the USLE-M , RUSLE2 and WEPP to model event erosion from bare fallow areas. Sci. Total Environ. 596–597, 32–42. doi:10.1016/j.scitotenv.2017.04.046

Kinnell, P.I.A., 2016. A review of the design and operation of runoff and soil loss plots. Catena 145, 257–265. doi:10.1016/j.catena.2016.06.013

Kinnell, P.I.A., Wang, J., Zheng, F., 2018. Comparison of the abilities of WEPP and the USLE-M to predict event soil loss on steep loessal slopes in China. Catena 171, 99–106. doi:10.1016/j.catena.2018.07.007

Kinnel, P.I.A., Risse, L.M., 1998. USLE-M: empirical modelling rainfall erosion through runoff and sediment concentration. Soil Sci. Soc. Am. J. 62, 1667–1672.

Kirkby, M.J., Irvine, B.J., Jones, R.J.A., Govers, G., Boer, M., Cerdan, O., Daroussin, J., Gobin, A., Grimm, M., Le Bissonnais, Y., Kosmas, C., Mantel, S., Puigdefabregas, J., Van Lynden, G., 2008. The PESERA coarse scale erosion model for Europe. I. - Model rationale and implementation. Eur. J. Soil Sci. 59, 1293–1306. doi:10.1111/j.1365-2389.2008.01072.x

Klemes, V., 1986. Operational testing of hydrological simulation models. Hydrolog. Sci. J. 31, 13-24.

Koiter, A.J., Owens, P.N., Petticrew, E.L., Lobb, D.A., 2013. The behavioural characteristics of sediment properties and their implications for sediment fingerprinting as an approach for identifying sediment sources in river basins. Earth-Science Rev. 125, 24–42. doi:10.1016/j.earscirev.2013.05.009

Krueger, T., Quinton, J.N., Freer, J., Macleod, C., Bilotta, G.S., Brazier, R.E., Hawkins, J., Haygarth, P., 2012. Comparing empirical models for sediment and phosphorus transfer from soils to water at field. Eur. J. Soil Sci. 63, 211–223. doi:10.1111/j.1365-2389.2011.01419.x

Laceby, J.P., Evrard, O., Smith, H.G., Blake, W.H., Olley, J.M., Minella, J.P.G., Owens, P.N., 2017. The challenges and opportunities of addressing particle size effects in sediment source fingerprinting : A review. Earth-Science Rev. 169, 85–103. doi:10.1016/j.earscirev.2017.04.009

Lacoste, M., Michot, D., Viaud, V., Evrard, O., Walter, C., 2014. Combining137Cs measurements and a spatially distributed erosion model to assess soil redistribution in a hedgerow landscape in northwestern France (1960-2010). Catena 119, 78–89. doi:10.1016/j.catena.2014.03.004

Larsen, I.J., MacDonald, L.H., 2007. Predicting postfire sediment yields at the hillslope scale: Testing RUSLE and Disturbed WEPP. Water Resour. Res. 43, 1–18. doi:10.1029/2006WR005560

Licciardello, F., Govers, G., O, C., Kirkby, M., Vacca, A., Kwaad, F.J.P.M., 2009. Evaluation of the PESERA model in two contrasting environments. Earth Surf. Process. Landforms 34, 629–640. doi:10.1002/esp

Licciardello, F., Taguas, E. V, Barbagallo, S., Gómez, J.A., 2013. Application of the Water Erosion Prediction Project (WEPP) in olive orchards on vertic soil with different management conditions. Trans. ASABE 56, 951–961. doi:10.13031/trans.56.9880

Lima, P.L.T., Silva, M.L.N., Quinton, J.N., Batista, P.V.G., Cândido, B.M., Curi, N., 2018. Relationship among crop systems , soil cover , and water erosion on a Typic Hapludox. Rev. Bras. Ciência do Solo 42, e0170081. doi:doi:10.1590/18069657rbcs20170081

Mahmoodabadi, M., Cerdà, A., 2013. WEPP calibration for improved predictions of interrill erosion in semi-arid to arid environments. Geoderma 204–205, 75–83. doi:10.1016/j.geoderma.2013.04.013

Mitasova, H., Hofierka, J., Zlocha, M., Iverson, L.R., 1996. Modeling topographic potential for erosion and deposition using GIS. Int. J. GIS 10, 629–641.

Montgomery, D.R., 2007. Dirt: The erosion of civilizations. University of California Press, Berkeley.

Mora-Valentín, E.M., Ortiz-de-Urbina-Criado, M., Nájera-Sánchez, J.J., 2018. Mapping the conceptual structure of science and technology parks. J. Technol. Transf. 43, 1410–1435. doi:10.1007/s10961-018-9654-8

Morgan, R.P.C., 2005. Soil Erosion & Conservation, 3rd ed. Blackwell, Oxford.

Morgan, R.P.C., 2001. A simple approach to soil loss prediction: A revised Morgan-Morgan-Finney model. Catena 44, 305–322. doi:10.1016/S0341-8162(00)00171-5

Morgan, R.P.C., Morgan, D.D.V., Finney, H.J., 1984. A predictive model for the assessment of soil erosion risk. J. Agric. Eng. Res. 30, 245–253. doi:10.1016/S0021-8634(84)80025-6

Morgan, R.P.C., Quinton, J.N., Smith, R.E., Govers, G., Poesen, J.W.A., Auerswald, K., Chisci, G., Torri, D., Styczen, M.E., 1998. The European Soil Ersoion Model (EUROSEM): A dynamic approach for predicting sediment transport from fields and small catchments. Earth Surf. Process. Landforms 23, 527–544. doi:10.1002/(SICI)1096-9837(199806)23:6<527::AID-ESP868>3.0.CO;2-5

Nash, E., Sutcliffe, V., 1970. River flow forecasting through conceptual models Part I - A discussion of principles. J. Hydrol. 10, 282–290. doi:10.1016/0022-1694(70)90255-6

Nearing, M.A., 2006. Can soil erosion be predicted? In: Owens, P.N., Collins, A.J. (Eds.), Soil erosion and sediment redistribution in river catchments, CAB International, Wallingford, pp. 145-152.

Nearing, M.A., 2000. Evaluating Soil Erosion Models Using Measured Plot Data:Accounting for Variability in the Data. Earth Surf. Process. Landforms 25, 1035–1043. doi:10.1002/1096-9837(200008)25:9<1035::AID-ESP121>3.0.CO;2-B

Nearing, M.A., Govers, G., Norton, L.D., 1999. Variability in soil erosion data from replicated plots. Soil Sci. Soc. Am. J. 63, 1829–1835. doi:10.2136/sssaj1999.6361829x

Oksanen, J., Sarjakoski, T., 2005. Error propagation of DEM-based surface derivatives. Comput. Geosci. 31, 1015–1027. doi:10.1016/j.cageo.2005.02.014

Oreskes, N., 1998. Evaluation (not validation) of quantitative models. Environ. Health Perspect. 106, 1453–1460. doi:10.1289/ehp.98106s61453

Oreskes, N., Shrader-Frechette, K., Belitz, K., 1994. Verification, validation, and confirmation of numerical models in the Earth Sciences. Science 263, 641–646. doi:10.1126/science.263.5147.641

Panagos, P., Borrelli, P., Poesen, J., Ballabio, C., Lugato, E., Meusburger, K., Montanarella, L., Alewell, C., 2015. The new assessment of soil loss by water erosion in Europe. Environ. Sci. Policy 54, 438–447. doi:10.1016/j.envsci.2015.08.012

Pappenberger, F., Beven, K.J., 2006. Ignorance is bliss: Or seven reasons not to use uncertainty analysis. Water Resour. Res. 42, 1–8. doi:10.1029/2005WR004820

Parsons, A.J., Foster, I.D.L., 2011. What can we learn about soil erosion from the use of 137Cs? Earth-Science Rev. 108, 101–113. doi:10.1016/j.earscirev.2011.06.004

Parsons, A.J., Wainwright, J., Brazier, R.E., Powell, D.M., 2009. Is sediment delivery a fallacy? Earth Surf. Process. Landforms 34, 155–161. doi:10.1002/esp

Pontes, L.M., 2017. Hydrosedimentological modeling in the Jaguarí river basin. (PhD thesis) Universidade Federal de Lavras.

Porto, P., Walling, D.E., 2015. Use of caesium-137 measurements and long-term records of sediment load to calibrate the sediment delivery component of the SEDD model and explore scale effect: Examples from southern Italy. J. Hydrol. Eng. 20, C4014005. doi:10.1061/(ASCE)HE.1943-5584.0001058

Prasuhn, V., Liniger, H., Gisler, S., Herweg, K., Candinas, A., Clément, J.P., 2013. A high-resolution soil erosion risk map of Switzerland as strategic policy support system. Land use policy 32, 281–291. doi:10.1016/j.landusepol.2012.11.006

Quine, T.A., Desmet, P.J.J., Govers, G., Vandaele, K., Walling, D.E., 2004. A comparison of the roles of tillage and water erosion in landform development and sediment export on agricultural land near Leuven, Belgium. In: Olive, L.J., Loughran, R.J., Kesby, J.A. (Eds), Variability in stream erosion and sediment transport. IAHS publication no. 224, pp. 77-86.

Quinton, J.N., 2004. Erosion and sediment transport. In: Wainwright, J., Mulligan, M. (Eds.), Environmental modelling: Finding simplicity in complexity, John Wiley & Sons Ltd., pp. 187-196.

Quinton, J.N., 1997. Reducing predictive uncertainty in model simulations : a comparison of two methods using the European Soil Erosion Model (EUROSEM). Catena 30, 101–117.

Quinton, J.N., 1994. The validation of physically-based erosion models — with particular reference to EUROSEM. (PhD Thesis) Cranfield University.

Quinton, J.N., Krueger, T., Freer, J., Brazier, R.E., Bilotta, G.S., 2011. A case study of uncertainty: Applying GLUE to EUROSEM. In: Morgan, R.P.C., Nearing, M.A. (Eds), Handbook of erosion modelling, Blackwell Publishing Ltd., Chichester, pp. 80-97.

R Development Core Team, 2017. R: A Language and Environment for Statistical Computing.

Rapp, J.F., Lopes, V.L., Renard, K.G., 2001. Comparing soil erosion estimates from RUSLE and USLE on natural runoff plots. In: Ascough, J.C., Flanagan, D.C. (Eds.), Soil erosion research for the 21st century. American Society of Agricultural Engineers, St Joseph, pp. 24-27.

Renard, K.., Foster, G.R., Weesies, G.A., McCool, D.K., Yoder, D.C., 1997. Predicting soil erosion by water: A guide to conservation planning with the Revised Universal Soil Loss Equation (RUSLE). Agricultural Handbook 703, US Department of Agriculture.

Renscheler, C.S., 2003. Designing geo-spatial interfaces to scale process models: The GeoWEPP approach. Hydrol. Proc. 17, 1005–1017.

Renschler, C.S., Harbor, J., 2002. Soil erosion assessment tools from point to regional scales—the role of geomorphologists in land management research and implementation. Geomorphology 47, 189–209. doi:10.1016/S0169-555X(02)00082-X

Risse, L.M., Nearing, M.A., Nicks, A.D., Laflen, J.M., 1993. Error Assessment in the Universal Soil Loss Equation. Soil Sci. Soc. Am. J. 57, 825–833.

Shrestha, D.P., Jetten, V.G., 2018. Modelling erosion on a daily basis, an adaptation of the MMF approach. Int. J. Appl. Earth Obs. Geoinf. 64, 117–131. doi:10.1016/j.jag.2017.09.003

Smith, H.G., Peñuela, A., Sangster, H., Sellami, H., Boyle, J., Chiverrell, R., Schillereff, D., Riley, M., 2018. Simulating a century of soil erosion for agricultural catchment management. Earth Surf. Process. Landforms 43, 2089–2105. doi:10.1002/esp.4375

Soil Survey Staff, 2014. Keys to soil taxonomy. U.S. Department of Agriculture, Natural Resources Conservation Service, Washington DC.

Spaeth, K.E., Pierson, F.B., Weltz, M.A., Blackburn, W.H., 2003. Evaluation of USLE and RUSLE Estimated Soil Loss on Rangeland. J. Range Manag. 56, 234. doi:10.2307/4003812

Sterman, J., 2002. All models are wrong : Reflections on becoming a systems scientist. Syst. Dyn. Rev. 18, 501–531. doi:10.1002/sdr.261

Stroosnijder, L., 2005. Measurement of erosion: Is it possible? Catena 64, 162–173. doi:10.1016/j.catena.2005.08.004

Takken, I., Beuselinck, L., Nachtergaele, J., Govers, G., Poesen, J., Degraer, G., 1999. Spatial evaluation of a physically-based distributed erosion model (LISEM). Catena 37, 431–447. doi:10.1016/S0341-8162(99)00031-4

Tanyaş, H., Kolat, Ç., Süzen, M.L., 2015. A new approach to estimate cover-management factor of RUSLE and validation of RUSLE model in the watershed of Kartalkaya Dam. J. Hydrol. 528, 584–598. doi:10.1016/j.jhydrol.2015.06.048

Tetzlaff, B., Friedrich, K., Vorderbrügge, T., Vereecken, H., Wendland, F., 2013. Distributed modelling of mean annual soil erosion and sediment delivery rates to surface waters. Catena 102, 13–20. doi:10.1016/j.catena.2011.08.001

Tiwari, A.K., Risse, L.M., Nearing, M.A., 2000. Evaluation of WEPP and its comparison with USLE and RUSLE. Trans. ASAE 43, 1129–1135.

Van Eck, N.J., Waltman, L., 2018. VOSviewer Manual. Univeristeit Leiden, Leiden. Available from: http://www.vosviewer.com/documentation/Manual_VOSviewer_1.6.8.pdf.

Van Eck, N.J., Waltman, L., 2010. Software survey : VOSviewer , a computer program for bibliometric mapping. Scientometrics 84, 523–538. doi:10.1007/s11192-009-0146-3

Van Oost, K., Beuselinck, L., Hairsine, P.B., Govers, G., 2004. Spatial evaluation of a multi-class sediment transport and deposition model. Earth Surf. Process. Landforms 29, 1027–1044. doi:10.1002/esp.1089

Van Oost, K., Govers, G., Cerdan, O., Thauré, D., Van Rompaey, A., Steegen, A., Nachtergaele, J., Takken, I., Poesen, J., 2005. Spatially distributed data for erosion model calibration and validation: The Ganspoel and Kinderveld datasets. Catena 61, 105–121. doi:10.1016/j.catena.2005.03.001

Van Oost, K., Govers, G., Desmet, P.J.J., 2000. Evaluating the effects of changes in landscape structure on soil erosion by water and tillage. Landsc. Ecol. 15, 577–589. doi:10.1023/A:1008198215674

Van Rompaey, A.J.J., Govers, G., 2002. Data quality and model complexity for regional scale soil erosion prediction. Int. J. Geogr. Inf. Sci. 16, 663–680. doi:10.1080/13658810210148561

Van Rompaey, A.J.J., Verstraeten, G., Van Oost, K., Govers, G., Poesen, J., 2001. Modelling mean annual sediment yield using a distributed approach. Earth Surf. Process. Landforms 26, 1221–1236. doi:10.1002/esp.275

Van Rompaey, A.J.J., Vieillefont, V., Jones, R.J.A., Montanarella, L., Verstraeten, G., Bazzoffi, P., Dostal, T., Krasa, J., de Vente, J., Poesen, J., 2003. Validation of soil erosion estimates at European scale. European Soil Bureau Research Report 13, 26.

Veihe, A., Rey, J., Quinton, J.N., Strauss, P., Sancho, F.M., Somarriba, M., 2001. Modelling of event-based soil erosion in Costa Rica, Nicaragua and Mexico: Evaluation of the EUROSEM model. Catena 44, 187–203. doi:10.1016/S0341-8162(00)00158-2

Verstraeten, G., Van Oost, K., Van Rompaey, A.J.J., Poesen, J., Govers, G., 2010. Evaluating an integrated approach to catchment management to reduce soil loss and sediment pollution through modelling. Soil Use Manag. 18, 386–394. doi:10.1111/j.1475-2743.2002.tb00257.x

Vieira, D.C.S., Prats, S.A., Nunes, J.P., Shakesby, R.A., Coelho, C.O.A., Keizer, J.J., 2014. Modelling runoff and erosion, and their mitigation, in burned Portuguese forest using the revised Morgan-Morgan-Finney model. For. Ecol. Manage. 314, 150–165. doi:10.1016/j.foreco.2013.12.006

Vigiak, O., Okoba, B.O., Sterk, G., Groenenberg, S., 2005. Modelling catchment-scale erosion patterns in the East African Highlands. Earth Surf. Process. Landforms 30, 183–196. doi:10.1002/esp.1174

Vigiak, O., Sterk, G., Romanowicz, R.J., Beven, K.J., 2006a. A semi-empirical model to assess uncertainty of spatial patterns of erosion. Catena 66, 198–210. doi:10.1016/j.catena.2006.01.004

Vigiak, O., van Loon, E., Sterk, G., 2006b. Modelling spatial scales of water erosion in the West Usambara Mountains of Tanzania. Geomorphology 76, 26–42. doi:10.1016/j.geomorph.2005.09.002

Vrieling, A., Sterk, G., Vigiak, O., 2006. Spatial evaluation of soil erosion risk in the West Usambara Mountains, Tanzania. L. Degrad. Dev. 17, 301–319. doi:10.1002/ldr.711

Walling, D.E., He, Q., 1998. Use of fallout 137 Cs measurements for validating and calibrating soil erosion and sediment delivery models. IAHS Publ. 249, 267–278.

Walling, D.E., He, Q., Whelan, P.A., 2003. Using 137Cs measurements to validate the application of the AGNPS and ANSWERS erosion and sediment yield models in two small Devon catchments. Soil Tillage Res. 69, 27–43. doi:10.1016/S0167-1987(02)00126-5

Waltner, I., Pásztor, L., Centeri, C., Takács, K., Pirkó, B., Koós, S., László, P., 2018. Evaluating the new soil erosion map of Hungary—A semiquantitative approach. L. Degrad. Dev. 29, 1295–1302. doi:10.1002/ldr.2916

Warren, S.D., Mitasova, H., Hohmann, M.G., Landsberger, S., Iskander, F.Y., Ruzycki, T.S., Senseman, G.M., 2005. Validation of a 3-D enhancement of the Universal Soil Loss Equation for prediction of soil erosion and sediment deposition. Catena 64, 281–296. doi:10.1016/j.catena.2005.08.010

Wechsler, S.P., Kroll, C.N., 2006. Quantifying DEM Uncertainty and its Effect on Topographic Parameters. Photogramm. Eng. Remote Sensing 72, 1081–1090.

Wendt, R.C., Alberts, E.E., Hjelmfelt, A.T., 1986. Variability of runoff and soil loss from fallow experimental plots. Soil Sci. Soc. Am. J. 50, 730–736. doi:10.1016/j.jrp.2008.12.003

Wilkinson, S.N., Hancock, G.J., Bartley, R., Hawdon, A. A., Keen, R.J., 2013. Using sediment tracing to assess processes and spatial patterns of erosion in grazed rangelands, Burdekin River basin, Australia. Agric. Ecosyst. Environ. 180, 90–102. doi:10.1016/j.agee.2012.02.002

Wilkinson, S.N., Henderson, A., Chen, Y., Sherman, B., 2004. SedNet user guide, Version 2. Client report, CSIRO Land and Water, Canberra. Available from: www.toolkit.net.au/sednet.

Wilkinson, S.N., Prosser, I.P., Rustomji, P., Read, A.M., 2009. Modelling and testing spatially distributed sediment budgets to relate erosion processes to sediment yields. Environ. Model. Softw. 24, 489–501. doi:10.1016/j.envsoft.2008.09.006

Wischmeier, W.H., Smith, D., 1978. Predicting rainfall-erosion losses – A guide to conservation planning. Agriculture handbook No. 537, U.S. Dept. of Agric., Washington DC. 58 pp.

Young, R.R., Onstad, C.A., Bosch, D.D., Anderson, W.W., 1989. AGNPS, agricultural non-point-source pollution model for evaluating agricultural watersheds. J. Soil Water Conserv. 44, 168–173.

Zhang, X.C., Nearing, M.A., Risse, L.M., McGregor, K.C., 1996. Evaluation of {WEPP} runoff and soil loss predictions using natural runoff plot data. Trans. Am. Soc. Agric. Eng. 39, 855–863. doi:10.13031/2013.27570