PEDRO VELLOSO GOMES BATISTA

A CRITICAL EVALUATION OF SOIL EROSION MODELS:

CASE STUDIES FROM SOUTHEAST BRAZIL

LAVRAS – MG

2019

PEDRO VELLOSO GOMES BATISTA

A CRITICAL EVALUATION OF SOIL EROSION MODELS: CASE STUDIES FROM SOUTHEAST BRAZIL

Tese submetida à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Ciência do Solo, área de Concentração de Recursos Ambientais e Uso da Terra, para obtenção do título de Doutor.

Orientador

Dr. Marx Leandro Naves Silva (Universidade Federal de Lavras)

Co-orientadores

Dr. John N. Quinton (Lancaster University)
Dr. Jessica Davies (Lancaster University)

LAVRAS – MG
2019

Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca Universitária da UFLA, com dados informados pelo(a) próprio(a) autor(a).

Batista, Pedro Velloso Gomes.
A critical evaluation of soil erosion models: Case studies from Southeast Brazil / Pedro Velloso Gomes Batista. - 2019.
189 p. : il.

Orientador(a): Marx Leandro Naves Silva.

Tese (doutorado) - Universidade Federal de Lavras, 2019.
Bibliografia.

1. Uncertainty. 2 sediment fingerprinting. 3. MMF model. 4 SEDD model. 5 Generalized Likelihood Uncertainty Estimation. I. Universidade Federal de Lavras. II. Título.

PEDRO VELLOSO GOMES BATISTA

A CRITICAL EVALUATION OF SOIL EROSION MODELS: CASE STUDIES FROM SOUTHEAST BRAZIL

Tese submetida à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Ciência do Solo, área de Concentração de Recursos Ambientais e Uso da Terra, para obtenção do título de Doutor.

Aprovada em 19 de dezembro, 2019.

Dr. John N. Quinton Lancaster	LEC/LU
Dr. Jessica Davies	LEC/LU
Dr. Nilton Curi	DCS/UFLA
Dr. Teotonio Soares de Carvalho	DCS/UFLA
Dr. Olivier Evrard	CEA/LSCE
Dr. Patrick Laceby	AE&P/ Science Division

Dr. Marx Leandro Naves Silva
Orientador

LAVRAS – MG
2019

ACKNOWLEDGEMENTS

I decided to write this section while taking a break from finishing the last article of this thesis. The break ended up lasting a couple of hours.

Mostly I thank my family. My parents, Antônio and Wanessa, not only for their love, but – in the lack of a better term – for the working ethics they have transmitted. My little sister, Titila, for her friendship and care. And also for her insights on unknown knows and semiotics: she is so much smarter than I'll ever be! My brother, Dum, for being the best mate one could ask for.

Still in the family tree (it's a big family, so bear with me), I thank my grandparents, Rosa and Laércio. For the love, the fun, and support. My granddad was particularly important in my decision to become an agronomist and a soil scientist. He was always teaching me about plants and telling me stories about his mining days. I miss him very much. My uncles, Lu, Paulo, João, and Zé, for the different kinds of knowledge they shared with me – from rocks to football, from cars to Schopenhauer. It all came in handy at some point of this PhD.

I thank my hometown friends, who have been nothing short of a family for as long as I can remember. I thank my under and post grad mates, from Lavras and Lancaster. They have been an inspiration and a huge part of this thesis. I have to thank Diego Tassinari specifically, as he's been carrying me around since the first term of our undergraduate course.

I am also in great debt to Patrick Laceby. Pat took an interest in my work out of the kindness of his heart. I would have never been able to do the sediment fingerprinting work presented here if it wasn't for him. Pat has been a mentor and a proper friend, really.

I thank my long time supervisor, Marx, for all the incentive and support. My pedology professor, Nilton, for the amazing classes, the soil knowledge, and the friendship.

I thank my supervisors, Jess and John. Learning from them was an inspiration and an absolute pleasure. They taught me so much about so many things, but most of all about what it means to be scientist.

I thank the University of Lavras and the Department of Soil Science. In special Pézão, Téo, and Dirce. I would be lost without these three.

I thank Lancaster University and the Pentland Centre for receiving me in the UK.

Speaking of a UK reception, I thank Birdy, Callum, Deano, and the lads for being my family away from home.

Finally, I acknowledge this study was funded 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 number 140131/2016-7), and the Minas Gerais State Research Foundation – FAPEMIG (process numbers GAG-APQ-01053-15 and APQ-00802-18).

“*But then ...*” I venture to remark, “*you are still far from the solution. ...*”

“*I am very close to one*,” William said, “*but I don’t know which.*”

“*Therefore you don’t have a single answer to your questions*?”

“*Adso, if I did I would teach theology in Paris*.”

“*In Paris do they always have the true answer*?”

“*Never*,” William said, “*but they are very sure of their errors*.”

“*And you*,” I said with childish impertinence, “*never commit errors*?”

“*Often*,” he answered. “*But instead of conceiving only one, I imagine many, so I become the slave of none*.”

Dialogue between Adso of Melk and William of Baskerville

Umberto Eco, The Name of the Rose, 1983

GENERAL ABSTRACT

Soil erosion models are a potentially powerful tool. They tell us where and when erosion and deposition occur, along with their magnitude. They simulate erosion and sediment transport responses to land use and climate changes. They identify erosion hot spots in large areas while mathematically describing complex non-stationary processes. But how confident are we in the capability of erosion models to fulfill their potential, and how do we establish such trust? As any model of real-world phenomena, soil erosion models must be tested against empirical evidence to have their performance evaluated. However, evaluating soil erosion models is complicated due to the uncertainties involved in the estimation of model parameters and system responses. Hence, in paper 01 of this thesis I studied some of the theoretical and practical issues regarding the evaluation of soil erosion models. I undertook a scientometric analysis to investigate how model evaluation has been approached in soil erosion research and performed a meta-analysis of model performance to understand the mechanisms that influence model predictive accuracy. I reviewed how soil erosion models have been evaluated at different temporal and spatial scales, focusing on the methods and sources of data used for model testing. I presented a case study to illustrate how model realizations can be tested as hypotheses in face of the epistemic uncertainty in models and the observational data. I concluded that model evaluation topics are often neglected in soil erosion research and that calibration seems to be the main mechanism of improvement of model performance. Finally, I discussed some philosophical aspects of hypothesis testing in environmental modelling. I disputed the notion that erosion models could be validated and called for change of attitude about model evaluation; highlighting the importance of pursuing multiple lines of evidence to investigate the usefulness, consistency, and fit-for-purposeness of a model. In paper 02 I performed a methodological investigation of sediment fingerprinting as a tool for modelling sediment provenance in large river catchments. I studied how pedogenetic processes can lead to the expression of the geochemical signals used for sediment fingerprinting source apportionments, and how this expression is controlled by sediment particle size. I argued that understanding the underlying processes leading to source signal development could improve the selection of geochemical tracers for sediment fingerprinting purposes. I demonstrated how this could be achieved by testing my approach for source stratification and element selection against a set of artificial mixtures. Moreover, I described how particle size affects sediment transport on large river systems, concluding that different types of sampling strategies and source stratification methods might be necessary for modelling the provenance of particular size fractions. In paper 03 I build on some of the methods developed in paper 02 in order to create framework for evaluating soil erosion models with sediment fingerprinting source apportionments. I applied the Generalized Likelihood Uncertainty Estimation (GLUE) methodology to the Sediment Delivery Distributed (SEDD) model at a large catchment (~6600 km^2) in Southeast Brazil. I assessed the model usefulness for identifying the main sediment sources in the catchment by evaluating behavioral model realizations against sediment fingerprinting source apportionments. From a falsificationist perspective, the SEDD model could not be rejected, as many model realizations yielded adequate system representations of catchment sediment loads. Moreover, I found that the partial agreement between fingerprinting and SEDD results provided some conditional corroboration of the models capability to simulate sediment provenance in the catchment – at least with some degree of spatial aggregation. However, this approach led to highly uncertain grid-based estimates of soil erosion and sediment delivery rates, which brought me to question the model usefulness for quantifying sediment transport dynamics. I concluded that multiple sources of data can – and should be – incorporated into the evaluation of spatially-distributed soil erosion models. Finally, I argued that although my results were case-specific, similar levels of error could be expected in soil erosion models elsewhere. This thesis demonstrated how uncertainty permeates all facets of soil erosion modelling and the very things we call observational data. Any deterministic "validation" of soil erosion models should be strongly refuted, and modelers such be made accountable for translating uncertainty to decision-makers.

Keywords: Uncertainty. Sediment fingerprinting. RUSLE model. SEDD model. Morgan-Morgan-Finey model. Generalized Likelihood Uncertainty Estimation.

RESUMO GERAL

Modelos de erosão do solo são ferramentas potencialmente úteis. Tais modelos descrevem onde e quando ocorrem o desprendimento, transporte e deposição de partículas do solo, e com qual magnitude. Modelos podem também identificar locais propensos à erosão enquanto descrevem matematicamente uma interação complexa de processos não-estacionários. Porém, até que ponto é possível confiar na capacidade de modelos de erosão do solo em cumprir seu potencial e como estabelecemos tal confiança? Como qualquer modelo representativo de fenômenos do mundo real, modelos de erosão precisam ser testados contra evidências empíricas para que sua capacidade preditiva seja avaliada. Porém, avaliar modelos de erosão é uma tarefa complicada devido às incertezas associadas à estimativa de parâmetros e a medição de respostas dos sistemas. Dessa forma, no primeiro artigo dessa tese foram abordados alguns dos problemas teóricos e práticos relativos à avaliação de modelos de erosão do solo. Além disso, uma análise cientométrica foi utilizada para investigar como a avaliação de modelos tem sido abordada por pesquisadores. Uma meta-análise sobre a exatidão preditiva de modelos de erosão foi realizada para identificar os mecanismos que influenciam o seu desempenho. Neste artigo também foi realizada uma revisão sobre a avaliação de modelos em diferentes escalas espaços-temporais, com foco nos métodos e fontes de dados usados para o teste de modelos de erosão. Foi demonstrado como a avaliação de modelos é um tópico negligenciado na pesquisa de erosão do solo e como a calibração de parâmetros é o principal mecanismo responsável pelo aumento da exatidão preditiva. Finalmente, foram discutidos aspectos filosóficos sobre o teste de hipóteses em modelos nas ciências naturais, refutando-se a noção de que modelos de erosão podem ser validados e salientando-se a necessidade de múltiplas linhas de evidência para avaliar a utilidade, consistência e a adequação à finalidade desses modelos. No segundo artigo da tese, foi realizada uma investigação metodológica sobre o uso de técnicas de rastreamento como ferramentas para modelagem da proveniência de sedimentos em grandes bacias hidrográficas. Avaliaram-se também como processos pedogenéticos podem levar à expressão de sinais geoquímicos utilizados no rastreamento de sedimentos, e como essa expressão é influenciada pelo tamanho de partículas. Argumentou-se que a compreensão dos processos controlando o desenvolvimento de sinais de fontes pode facilitar a escolha de traçadores geoquímicos. Para demonstrar esta suposição, uma abordagem processual para escolha de traçadores e delimitação de fontes de sedimentos foi testada contra misturas artificiais, em diferentes frações texturais de sedimentos. No terceiro artigo da tese, alguns dos métodos desenvolvidos no artigo anterior foram usados para criar uma estrutura para avaliar modelos de erosão a partir de técnicas de rastreamento de sedimentos. Uma metodologia de estimativa de incertezas foi aplicada em um modelo espacial de erosão e entrega de sedimentos em uma grande bacia hidrográfica (~6000 km^2) localizada no sudeste do Brasil. A utilidade do modelo para identificar as principais fontes de sedimentos na bacia foi avaliada comparando-se as predições do modelo contra os resultados obtidos por meio do rastreamento geoquímico de sedimentos. Dentro de uma perspectiva falsificacionista, o modelo não pode ser rejeitado, uma vez que várias realizações geraram respostas quantitativas aceitáveis da produção total de sedimentos na bacia. Ademais, uma concordância parcial entre as predições do modelo e os resultados do rastreamento geoquímico corroborou condicionalmente a capacidade do modelo em simular a dinâmica hidrossedimentológica na bacia – ao menos com certo grau de agregação espacial. Porém, as estimativas espacialmente contínuas de perdas de solo e entrega de sedimentos foram altamente incertas, o que gerou um questionamento quanto à utilidade do modelo para quantificar espacialmente o transporte de sedimentos na bacia. Apesar de estes resultados serem específicos ao estudo de caso, níveis de incerteza semelhantes podem ser esperados em modelos de erosão aplicados em outras situações. De forma geral, esta tese demonstrou como a incerteza permeia todos os aspectos da modelagem da erosão e das próprias coisas que consideramos dados observacionais. Qualquer "validação" determinística de modelos de erosão do solo deve, portanto, ser peremptoriamente refutada. Ademais, analistas e cientistas devem ser responsabilizados por traduzir a incerteza associada a estes modelos para usuários finais.

Palavras-chave: Incerteza. Rastreamento de sedimentos. Modelo RUSLE. Modelo SEDD. Modelo Morgan-Morgan-Finey. Generalized Likelihood Uncertainty Estimation.

SUMMARY

GENERAL INTRODUCTION

Keith Beven (2009) wrote that his research career has been an attempt to cope with the failure of the hydrological model he worked on during his PhD. In a sense, I believe that this thesis has been my attempt to cope with the success of the soil erosion models I worked with during my Master's degree. This may sound counter intuitive: as scientists, we want our models to be successful; we want them to provide insightful descriptions of phenomena and, by doing so, we want them to make accurate estimates of system responses. Then why did I need to cope with the success of the models I was using?

The cause of the uneasiness I felt about my results at the time now seem obvious and formally definable: given the errors involved in the characterization of a complex system, such as the soil and the processes driving its redistribution by water, there are many acceptable representations of reality. This is defined and equifinality or non-singularity (Beven, 2009, 2006). To put it in practical terms, I felt like any model, given enough freedom, would be able to make good predictions of sediment transport rates at the outlet of the catchment I was monitoring. So the problem was not with the success of the models, the problem was with the data and the tests I was using to evaluate them.

While writing my Master's thesis, I could hardly translate my frustration into words, and I had no idea about how to deal with it in a scientifically sound manner. Of course, I was not the first one to be bothered about the issues I am describing. But it sure felt like it, considering that the vast majority of papers I was reading did not mention these issues and did not offer a methodology to address them. This all changed when John Quinton, who is now one of my supervisors, came to visit Lavras in late 2015. John supplied me with the literature and the concepts that would help me formalize the questions I wanted to ask in what would become my PhD research. This is the result.

Originally, the objective of this thesis was to develop a sediment fingerprinting study that would improve the evaluation of large-scale spatially-distributed soil erosion models. As sediment fingerprinting provides quantitative information about sediment provenance (Collins et al., 2017; Koiter et al., 2013; Laceby et al., 2019), my idea was to use this technique to create extra lines of evidence to test the models capability to represent sediment dynamics within catchments. Although I understand I have accomplished this task, the focus of the thesis ultimately shifted to a more general investigation about model evaluation and data uncertainty. Since the first paper of this thesis provides an in depth introduction to these topics, as well as a literature review, I will refrain from prolonging this general introduction and focus on an overview of the studies here presented.

The main objective of this thesis is to investigate methods and sources of data that allow us to assess the usefulness and consistency of soil erosion models, at different temporal and spatial scales. Moreover, it aims to examine what makes a useful soil erosion model and to scrutinize the uncertainty in models and the observational data of system responses.

In Paper 01, I studied how erosion models have been tested, and discussed the advantages and limitations of previously employed methodologies. Paper 01 provides a scientometric investigation of erosion modeling research topics, as well as a meta-analysis of model predictive accuracy. Taking into account the outcomes of these analyses, I presented a way forward for improving the evaluation of soil erosion models, discussing some philosophical aspects of hypothesis testing in environmental models. In practical terms, I presented a case study that demonstrates how process-based erosion models can be evaluated at plot-scale while considering the uncertainty in model structures and the observational data. Therefore, Paper 01 established the theoretical background from which the rest of the thesis is structured upon.

Paper 02 provides a more applied sediment fingerprinting study, although still focusing on methodological advances that would ultimately enable me to create an uncertainty-based framework for incorporating fingerprinting data into erosion model testing. For this research, I investigated how pedogenetic processes can lead to the expression of the geochemical signals used for sediment fingerprinting source apportionments, and how this expression is controlled by sediment particle size. I argued that understanding the underlying processes leading to source signal development could improve the selection of geochemical tracers for sediment fingerprinting purposes. I demonstrated how this could be achieved by testing my approach to source stratification and element selection against a set of artificial mixtures. Moreover, I analyzed how particle size affects sediment transport on large river systems and developed a bootstrapping method to simulate sediment fingerprinting un-mixing model solutions.

In Paper 03 I focused on representing the uncertainty in commonly used empirical spatially-distributed soil erosion and sediment delivery models. Namely, the Revised Universal Soil Loss Equation (RUSLE) (Renard et al., 1997) and the Sediment Delivery Distributed model (SEDD) (Ferro and Minacapilli, 1995; Ferro and Porto, 2000). I applied the models within Generalized Likelihood Uncertainty Estimation (GLUE) (Beven and Binley, 1992) methodology at a large catchment (~6600 km^2) in Southeast Brazil. I conditioned the model parameters based on measured sediment loads and tested the behavioral model realizations against sediment fingerprinting sources apportionments. For this research, I developed a hierarchical tributary sampling design for sediment fingerprinting that aimed to facilitate a comparison against erosion/sediment delivery models. This was the largest sediment fingerprinting research ever conducted in Brazilian river basins. To the best of my knowledge, it was also the first study to provide a method for incorporating sediment fingerprinting source apportionments into soil erosion model testing that fully represents the uncertainty in model structures, forcing observational data, and the independent evaluation data.

REFERENCES

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. https://doi.org/10.1016/j.jhydrol.2005.07.007

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

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. 194. https://doi.org/10.1016/j.jenvman.2016.09.075

Ferro, V., Minacapilli, M., 1995. Sediment delivery processes at basin scale. Hydrol. Sci. J. 40, 703–717. https://doi.org/10.1080/02626669509491460

Ferro, V., Porto, P., 2000. Sediment Delivery Distributed (SEDD) Model. J. Hydrol. Eng. 411–422.

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. https://doi.org/10.1016/j.earscirev.2013.05.009

Laceby, J.P., Gellis, A.C., Koiter, A.J., Blake, W.H., Evrard, O., 2019. Preface — evaluating the response of critical zone processes to human impacts with sediment source fingerprinting 3245–3254.

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).

PAPER 01

Standards of the journal – Earth-Science Reviews (Published)

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

Pedro V. G. Batista[1,2], 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

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.

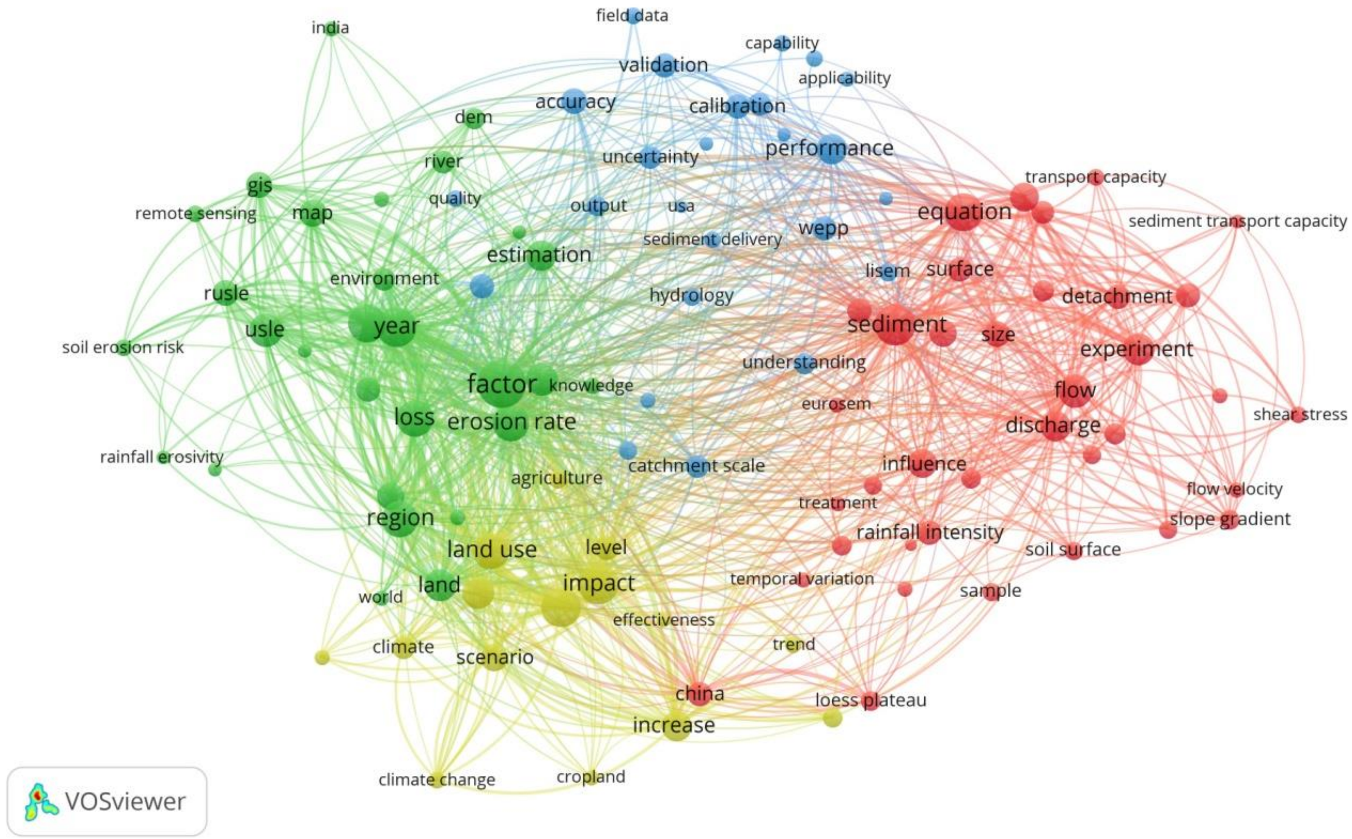


Fig. 1. Term co-occurrence network map. Clusters are identified by color (Cluster 1: green; Cluster 2: red; Cluster 3: yellow; Cluster 4: blue). Labels and circle sizes are proportional to the number of occurrences. Lines indentify major links between terms, and line thickness represents association 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 Erosion Deposition	Mitasova et al. (1996)
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.

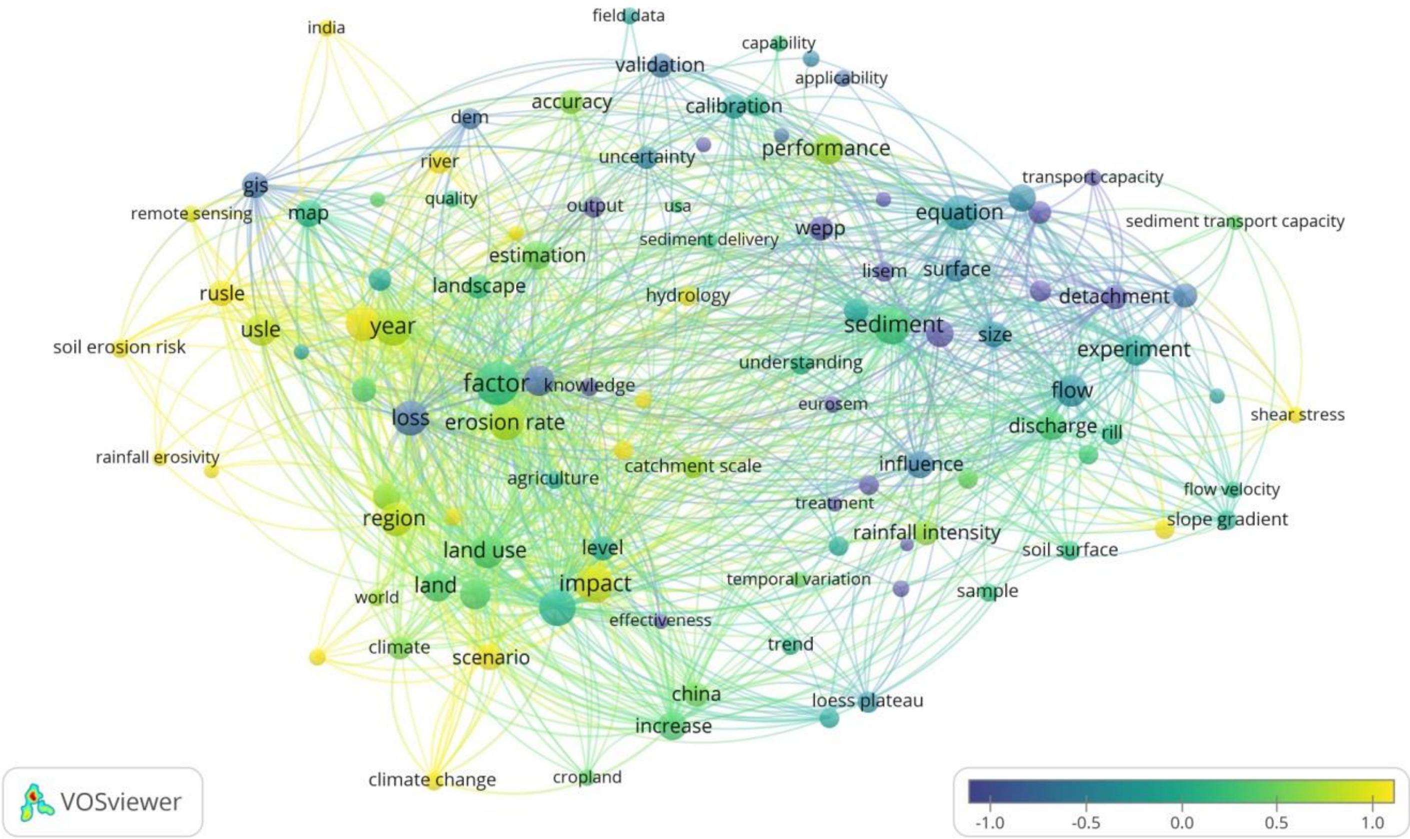


Fig. 2. Overlay visualization of the term co-occurrence network map. Colors are rendered according to the normalized average year of publication of the articles in which the terms occur. Normalization was performed by subtracting the term average by the overall mean and dividing it by the standard deviation.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

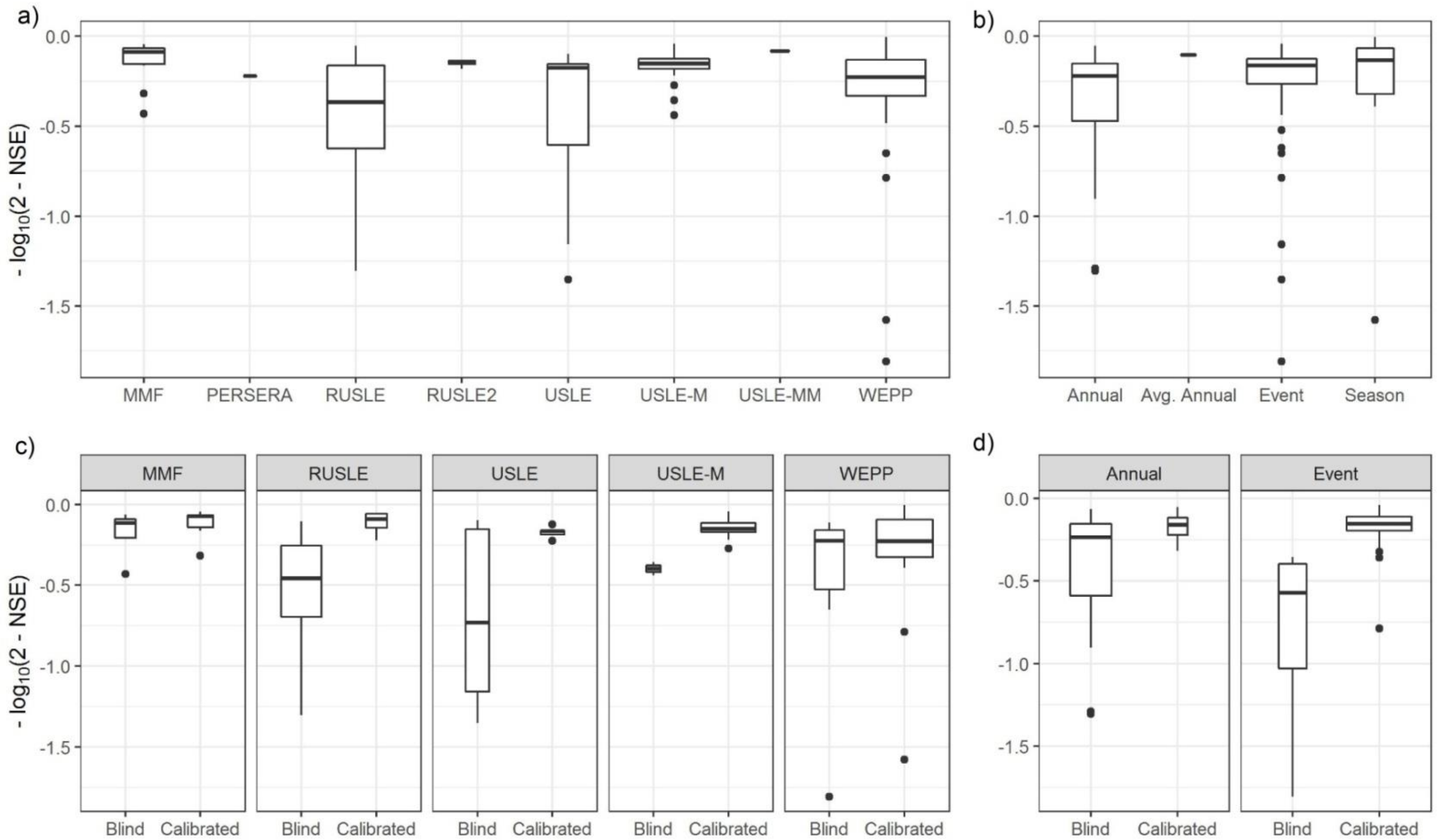


Fig. 3. NSE values reported in erosion modelling studies grouped by: a) model; b) temporal scale of model application; c) model and the use or not of calibration; d) temporal scale of model application and use or not of calibration. The width of the boxes is scaled according to the size of the datasets 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

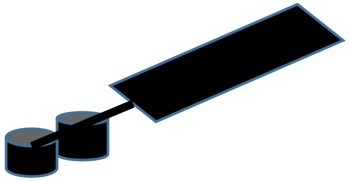
The advent of GIS, the accessibility of computing, and the popularization of remote sensing images had a great impact on erosion modelling: models can now be applied at large scales and in a distributed manner with relative ease. Contrary to earlier lumped model results, the grid-based outputs of spatially 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.

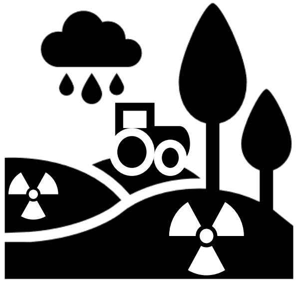
Hence, a great effort has been put into adapting and scaling erosion models into a GIS framework (e.g., Desmet and Govers, 1997; Mitasova et al., 1996; Renschler, 2003; Renschler and Harbor, 2002), and some models, such as LISEM and WaTEM/SEDEM, were developed in an explicitly distributed, rater-based structure.

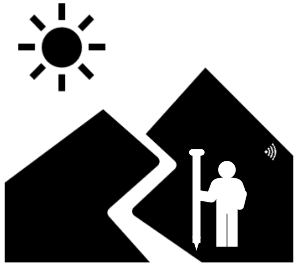
However, evaluating distributed erosion models, where catchments are the predominant spatial scale of application, is problematic: the previously discussed issues regarding model evaluation are exacerbated, as parameterization becomes even more uncertain and equifinality more likely. Moreover, the outlet-based 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).

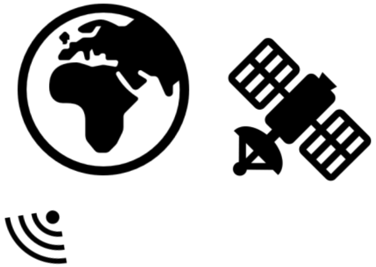
Spatially distributed data suitable for model evaluation are generally acquired by I) field-based monitoring, in which erosion and depositional features are mapped and often combined volumetric measurements of rills, gullies, and sediment deposits (e.g., Desmet and Govers, 1997; Evans and Brazier, 2005; 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 	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 km^2). 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 km^2) 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)	$R_e = R(1 - A)$	R = rainfall (mm) A = proportion of rainfall intercepted by vegetation
Leaf drainage (mm)	$L_d = R_e C_c$	C_c = proportion of canopy cover
Direct throughfall (mm)	$D_t = R_e - L_d$	
Kinetic energy of direct throughfall for tropical climates (J m^{-2})	$K_{et} = D_t(29 - \frac{127.5}{I})$	I = typical rainfall intensity value for erosive rain (mm h^{-1})
Kinetic energy of leaf drainage (J m^{-2})	$K_{el} = L_d(15.8\, P_h^{0.5}) - 5.87$	P_h = plant canopy height (m)
Total kinetic energy (J m^{-2})	$K_e = K_{et} + K_{el}$	
Annual runoff (mm)	$Q = R_e^{\frac{-R_c}{R_o}}$	R_o = mean rain per day (mm)
Soil moisture storage capacity (mm)	$R_c = 1000\, M_c\, B_d\, H_d(\sqrt{\frac{E_t}{E_o}})$	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})	$F = 0.001\, K\, K_e$	K = soil detachability index (g J^{-1})
Soil particle detachment by runoff (kg m^{-2})	$H = Z\, Q^{1.5} \sin S\, (1 - G_c)10^{-3}$	S = slope steepness (°) G_c = proportion of ground cover
Resistance of the soil	$Z = \frac{1}{0.5\sigma}$	σ = soil cohesion (kPa)

Runoff transport capacity (kg m^{-2})	$T_c = C\, Q^2 \sin 10^{-3}$	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 ($Rdiff_{occ}$):

$$Rdiff_{occ} = m\, log_{10}(M) + \ b$$

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.

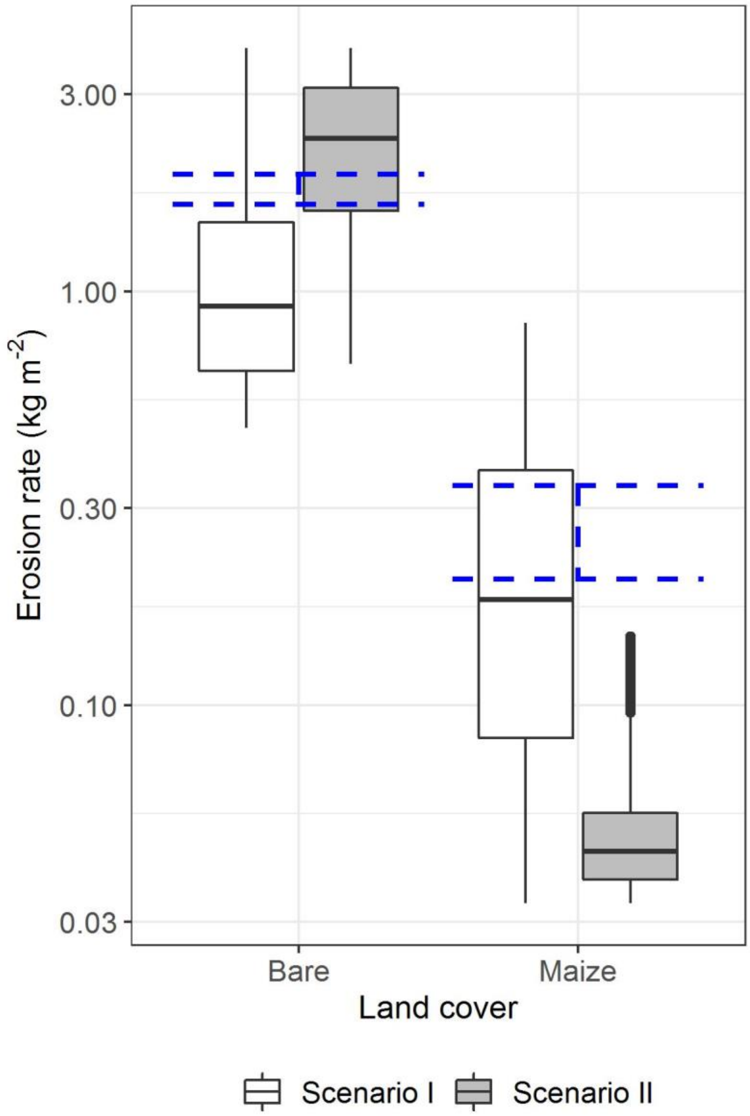


Fig. 4. Estimated erosion rates of behavioral realizations of the MMF model for the bare and maize plots. Blue dashed lines represent the range of observed soil losses for the replicate plots within each treatment.

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).

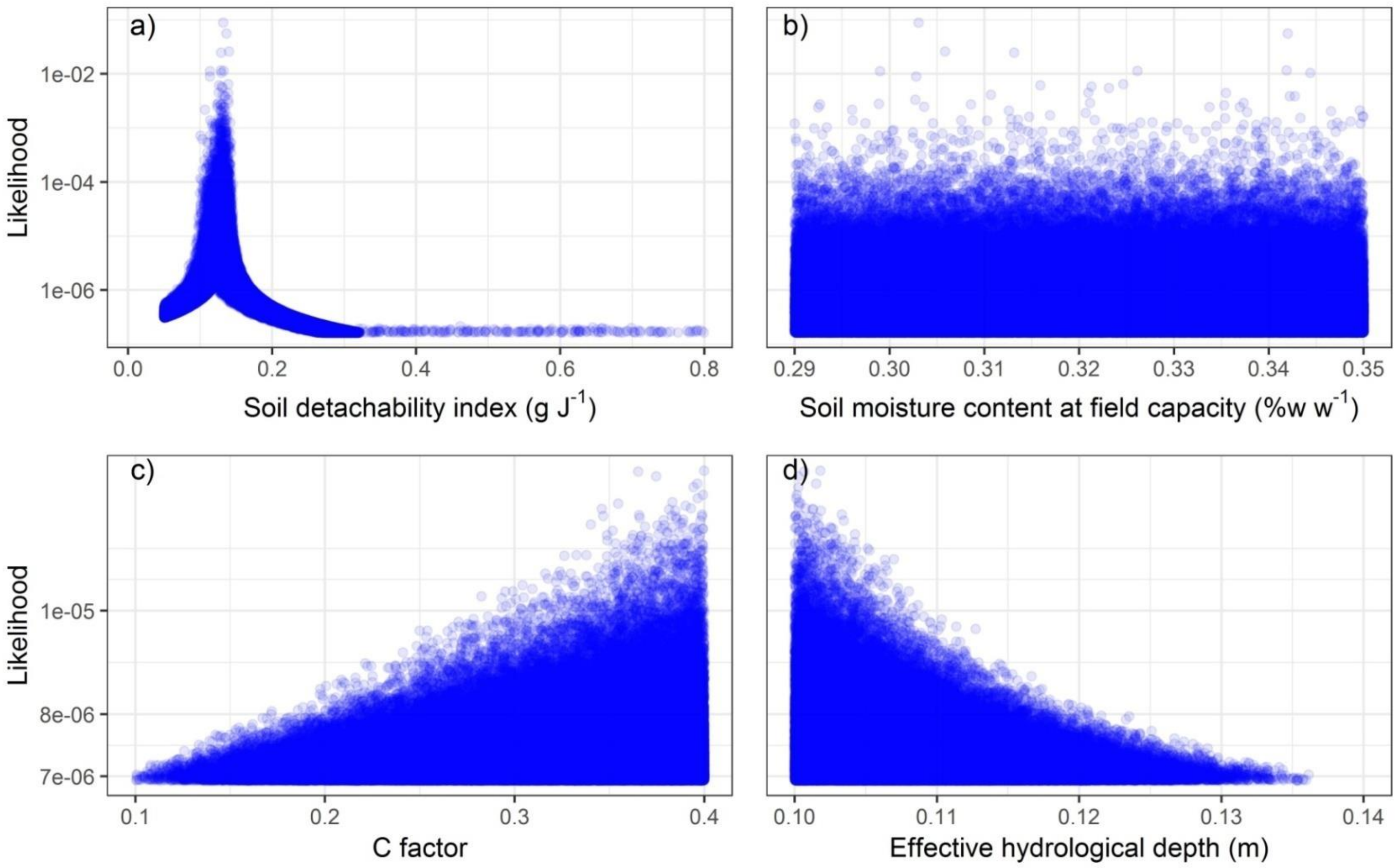


Fig. 5. Dotty plots of behavioral model realizations for the simulations from Scenario II in the bare (a,b) and maize plots (c,d). Each point relates a sampled parameter value to the rescaled likelihood of the model realization. High-sensitivity parameters, such as the soil detachability index, have 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 Harkin 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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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.
https://doi.org/https://doi.org/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.
https://doi.org/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.
https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/10.2136/sssaj1999.6361829x

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

Oreskes, N., 1998. Evaluation (not validation) of quantitative models. Environ. Health Perspect. 106, 1453–1460. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/10.2307/4003812

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

Stroosnijder, L., 2005. Measurement of erosion: Is it possible? Catena 64, 162–173. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/doi: 10.13031/2013.27570

PAPER 02

Standards of the journal – Journal of Soils and Sediments (Published)

SEDIMENT FINGERPRINTING IN THE CRITICAL ZONE

Using pedological knowledge to improve sediment source apportionment in tropical environments

Pedro V. G. Batista[1, 2] •J. Patrick Laceby[3] • Marx L. N. Silva[1] • Diego Tassinari[1] • Diêgo F. A. Bispo[1] • Nilton Curi[1] • Jessica Davies[2] • John N. Quinton[4]

[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, United Kingdom

[3] Alberta Environment and Parks, Environmental Monitoring and Science Division, Calgary, Canada

[4] Lancaster Environment Centre, Lancaster University, Lancaster, United Kingdom

✉Pedro V.G. Batista

p.batista@lancaster.ac.uk

Abstract

Purpose Soils are important regulators of Critical Zone processes that influence the development of geochemical signals used for sediment fingerprinting. In this study, pedological knowledge of tropical soils was incorporated into sediment source stratification and tracer selection in a large Brazilian catchment.

Materials and methods In the Ingaí River basin (~ 1200 km²), Brazil, three source end-members were defined according to the interpretation of soil and geological maps: the upper, mid, and lower catchment. A tributary sampling design was employed, and sediment geochemistry of three different size fractions was analyzed (2 – 0.2 mm; 0.2 – 0.062 mm, and< 0.062 mm). A commonly used statistical methodology to element selection was compared to a knowledge-based approach. The mass balance un-mixing models were solved by a Monte Carlo simulation.

Modeled source contributions were evaluated against a set of artificial mixtures with known source proportions.

Results and discussion For the coarse fraction (2 – 0.2 mm) both approaches to element selection yielded high errors compared to the artificial mixtures (23.8 % and 17.8 % for the statistical and the knowledge-based approach, respectively). The knowledge-based approach provided the lowest errors for the intermediate (0.2 – 0.062 mm) (10.9 %) and fine (< 0.062 mm) (11.8 %) fractions. Model predictions for catchment outlet target samples were highly uncertain for the coarse and intermediate fractions. This is likely the result of the spatial scale of the source stratification not being able to represent sediment dynamics for these fractions. Both approaches to element selection show that most of the fine sediments (median >90 %) reaching the catchment outlet were derived from Ustorthents in the lower catchment.

Conclusions The different element selection methods and the artificial mixtures provide multiple lines of evidence for evaluating the fingerprint approaches. Our findings highlight the importance of considering pedogenetic processes in source stratification, and demonstrate that different sampling strategies might be necessary to model specific sediment fractions.

Keywords Erosion processes • Geochemical fingerprinting • Sediment particle size • Sediment sources • Sediment tracing • Tropical soils

1 Introduction

Soil forming processes and ecosystem services provided by the pedosphere are central to the Critical Zone (Lin 2010; Banwart 2011). Soil erosion reduces soil quality by reducing soil depth, degrading soil structure, and reducing organic carbon and nutrient contents. In addition to these on-site effects, increased sediment delivery due to accelerated soil erosion can lead to pollution and eutrophication of downstream water bodies (Zamparas and Zacharias 2014; Yang et al. 2017). Moreover, high sedimentation rates reduce dam and reservoir storage capacity, compromising water supply and hydroelectric power generation (Hu et al. 2009; Zhao et al. 2017). These off-site consequences of soil erosion are often experienced at significant distances downstream. Knowledge of sediment transport processes and identifying the origin of sediments in river catchments is therefore necessary to understand, predict, and remediate off-site erosion impacts.

Sediment fingerprinting techniques are often used to identify sediment sources within a catchment. As the properties of the material being transported through river networks essentially reflect biogeochemical processes occurring in the Critical Zone (Amundson et al. 2007), the fingerprinting approach is based on the similarity of physical or biogeochemical properties between target sediment and their potential upstream sources (Klages and Hsieh 1975; Yu and Oldfield 1989; Walling and Woodward 1995; Collins et al. 1996). The relative source contribution is estimated through parameter optimization of mass balance un-mixing models, which are typically either stochastically solved in a Monte Carlo simulation (Collins et al. 2013; Wilkinson et al. 2015; Tiecher et al. 2016) or in a Bayesian framework (Cooper et al. 2014; Cooper and Krueger 2017).

Although many different sediment properties have been used to identify sources, sediment elemental composition has been commonly used in fingerprinting studies to distinguish source contributions according to landuse (Collins et al. 2010; Voli et al. 2013; Cooper et al. 2015;

Pulley et al. 2017), geological units (Olley and Caitcheon 2000; Wilkinson et al. 2013; Laceby and Olley 2015), and, less frequently, soil classes (Evrard et al. 2013; Lepage et al. 2016; Le Gall et al. 2017). In addition to aiding catchment management, Koiter et al. (2013b) argue that the information obtained in such studies can be used to understand the underlying processes that regulate sediment transport and generate the individual geochemical signatures within sources.

Large catchments present particular problems for fingerprinting studies. The long distances between potential upstream sources and the catchment outlet often lead to increased residence times, which may intensify fluvial sorting processes and particle size selectivity (Koiter et al. 2013a, b). Moreover, large catchments often have a diversity of landuses, parent materials, and soil classes. In these settings, a landuse based source apportionment may be unsuitable for geochemical fingerprinting, due to within landuse soil variability (Pulley et al. 2017). In such cases, lithological and/or confluence-based source stratifications might be more effective (Collins et al. 2017). While lithology has been proven to be a main control of sediment geochemistry in catchments with contrasting felsic/mafic geological units (Laceby et al. 2015), pedogenetic processes may provide an important insight to source signal development in catchments with less dissimilar parent materials, as demonstrated by Bajard et al. (2017). These processes might be particularly relevant in tropical environments, where intense weathering-leaching may have considerable influence on soil, and ultimately, sediment properties.

The selection of sediment geochemical properties prior to modeling has received much attention in fingerprinting studies, and recent work has brought to question the validity of widely used statistical approaches (Smith et al. 2018). To address this, Koiter et al. (2013a) and Laceby et al. (2015) have proposed a combination of statistical and process/knowledge-based methods, which increases interpretation possibilities of modeling estimates. Ideally, fingerprint properties should be conceptually relatable to upstream processes regarding sediment transport and geochemical source signals (Koiter et al. 2013a). Given that the soil is an important regulator of these processes, pedological knowledge can offer valuable information regarding geochemical tracer selection.

Furthermore, understanding the relationship between sediment particle size and elemental concentration is imperative to improve the knowledge of sediment tracer predictability (Laceby et al. 2017). Fluvial processes typically have a sorting effect on sediment particles, which usually decrease in median grain size with travelled distance as a result of selective transportation and deposition (Walling et al. 2000). Given that soil elemental composition is strongly related to particle size, transport selectivity can affect geochemical fingerprinting properties (Koiter et al. 2013b). Moreover, different processes regulate sediment transport in varying size fractions. While coarser fractions have a greater interaction with channel bed, finer loads are controlled primarily by catchment sediment supply and are therefore less influenced by river transport capacity (Walling and Collins 2016). Hence, sediment source contributions can display contrasting patterns across different size fractions (Haddadchi et al. 2016). Although the influence of particle size on sediment source signals is widely recognized, relatively few studies have focused on tracing different particle size fractions (e.g. Motha et al. 2002; Hatfield and Maher 2009; Haddadchi et al. 2016).

The evaluation of sediment fingerprinting approaches is crucial to enable informed decision making based on modeled source apportionments. However, gathering independent data to test the outputs of fingerprinting models is problematic, as reliable alternative techniques to quantify source contributions (i.e. suspended sediment yield measurements from multiple sub-catchments or source unit end-members) can be operationally complex and expensive (Collins et al. 2017). Therefore, artificial mixtures with known proportions of sediment source groups have been increasingly used to, at the very least, test the accuracy of un-mixing model estimates (Haddadchi et al. 2014; Sherriff et al. 2015; Pulley et al. 2017; Cooper and Krueger 2017). With this approach, the robustness of the models is assessed by a comparison of calculated source contributions and known mixture proportions (Haddadchi et al. 2014).

The goal of this research is to develop a tributary tracing technique that incorporates pedological knowledge of tropical soil formation/erosion processes into sediment source apportionment and tracer selection across multiple particle size fractions. The study is conducted in the Ingaí River basin (~1200 km²), Brazil, which has a complex geological and pedological heterogeneity. We compare a knowledge-based element tracer selection to a statistical methodology, which are both evaluated against a set of artificial mixtures. While others have incorporated knowledge-based criteria to the selection of fingerprinting properties, our approach is the first to be comprehensively grounded on pedological reasoning, highlighting the role of soils as regulators of the processes leading to source signal development. Multiple particle size fractions are analyzed to understand the relationship between particle size and source signal, as well as their interaction with fluvial transport processes. The outcomes of this research will help develop appropriate strategies for sediment fingerprinting and management in tropical environments, while also contributing to our knowledge of processes affecting sediment geochemistry and transport across different particle sizes.

2 Materials and methods

2.1 Catchment description

The Ingaí River basin (~ 1200 km²) is located within the upper Grande River basin, in southeastern Brazil (Fig. 1c). The Ingaí River is formed by sources in the Mantiqueira mountain range and flows into the Capivari River, which is dammed near its confluence with the Grande River, at the Funil hydroelectric power plant reservoir. Altitude ranges from approximately 1780 m in the headwaters to 900 m at the catchment outlet. The predominant climate type according to Köppen's climatic classification is humid subtropical with dry winters and warm summers (Cwb) with an average annual precipitation of ~ 1500 mm (Hijmans et al. 2005; Alvares et al. 2013).

The Ingaí River basin is set upon old surfaces, mostly made of metamorphosed Proterozoic and Archean rocks (Fig. 1a). The upper catchment is dominated by both paragneiss (38 %) and orthogneiss (32 %) (CODEMIG - CPRM 2014) (Table 1). The remaining area contains biotite-schists of the same formation as the paragneiss, though with a less intense metamorphic facies. Although the main soil class is Paleudult (48 %), there are also areas of Hapludoxes (20 %) and Ustorthents (16 %) (FEAM 2010) (Fig. 1b). Landuse consists mainly of extensive, minimally managed, pastures (64 %), found on the slightly more fertile blocky structured Paleudults (Fig. 2a). Erosion is typically only evident where cattle trails create preferential water pathways that tend to evolve to rills and small gullies. Also, cropland located on steep slopes in the absence of soil conservation practices often results in isolated erosion hotspots.

In the mid-catchment, the relief is gentler and the river valley widens enough to generate some clastic Quaternary sediment deposits (CODEMIG - CPRM 2014). The surface is again very old, with a predominance of orthogneiss (65 %). Cropland is more widespread, despite the major occurrence of Dystrudepts (54 %), which are shallow and highly erodible soils. Gullies are a common feature, often associated with degraded pastures and unpaved roads, some of which have been used since colonial times in the early 18th century (Fig. 2b).

In the lower area of the catchment, the Ingaí River crosses a Proterozoic ridge formation dominated by quartzite, mica-schist, and phyllite (CODEMIG - CPRM 2014). These same rocks establish the northern boundary of the watershed. The steeper slopes contain Ustorthents and rock outcrops (46 %) (FEAM 2010). Soils are very shallow because of naturally high erosion rates, which remove the surface soil layer before pedogenetic processes take place at greater depths (Resende et al. 2014) (Fig. 2c). The environment restricts agriculture to eucalyptus stands and extensive cattle grazing. In addition, mine pits for commercial quartzite exploration are found in the region. In the last decade, some of these mines have been fined or had their activities suspended due to irregularities regarding deforestation and waste disposal (Borges 2011; G1 Sul de Minas 2016). The remaining area of the lower catchment is dominated by biotite-schist, metagraywacke, and orthogneiss (Table 1), upon which Hapludoxes (54 %) have developed, favored by the gentler landscape. These soils have the most intense agricultural use in the watershed: soybean followed by maize and wheat or oats are a common no-till crop rotation scheme.

Accordingly, three geographical source units were established: i) the upper catchment (S1), comprised predominantly of Paleudults derived from gneiss; ii) the mid catchment (S2), where Dystrudepts are widespread and are also developed from a gneissic parent material; and iii) the lower catchment (S3), comprised of a mixture of Ustorthents, that occur in association to quartzite/phyllite/mica-schist ridge formations, and Hapludoxes, which are found in more gentle slopes formed above biotite and schist-metagraywacke bedrocks. These three geographical source end-members will be modeled as the potential contributors of target sediment sampled at the catchment outlet.

2.2 Sampling design and sample collection

A tributary sampling design (Laceby et al. 2015; Le Gall et al. 2016; Vale et al. 2016) was utilized within the catchment hydrological network to stratify potential sediment sources based on contributing area soil classes and their underlying parent material (Fig. 1). In the Ingaí catchment, the heterogeneity of lythotypes and soil classes makes it difficult to sample sources directly. The basic foundation of our approach is that a set of tributaries can be considered a specific spatial sediment source. Tributary tracing designs do not rely on hillslope connectivity assumptions, given that source samples are retrieved from the riverine system. Moreover, potential particle size selectivity during sediment transport is restricted to in-stream processes (Laceby et al. 2017).

Sediment sampling was conducted from July 2017 to February 2018. Composite samples were collected from lag deposits, which consisted of sediment drapes located on riverbanks or floodplains formed as water level receded after recent floods (Laceby and Olley 2015; Theuring et al. 2015). The uppermost sediment layer (1-2 cm) was scraped with a non-metallic trowel. Each sample was composed of approximately 15 scrapes. In total 69 source samples (*n* S1 = 29, S2 = 21, S3 = 19) and 10 target sediment samples from the catchment outlet were collected.

2.3 Laboratory analysis

Samples were oven dried at 60 °C before being dry sieved into three particle size fractions: 2-0.2 mm, 0.2-0.062 mm, and <0.062 mm. Sediment elemental composition was determined by X-ray fluorescence (XRF), using a portable XRF spectrometer equipped with a 50 kV/100 µA X-ray tube. XRF technology has been increasingly used for quantifying soil geochemistry, given that it provides a non-destructive method with rapid results and no chemical waste generation (Ribeiro et al. 2017; Silva et al. 2017). The analysis allows for the quantification of the following 45 elements: Ag, Al_2O_3, As, Au, Ba, Bi, CaO, Cd, Ce, Cl, Co, Cr, Cu, Fe, Hf, Hg, K_2O, La, MgO, Mn, Mo, Nb, Ni, P_2O_5, Pb, Pd, Pt, Rb, Rh, S, Sb, Se, SiO_2, Sn, Sr, Ta, Th, Ti, Tl, U, V, W, Y, Zn, Zr. Each sample was measured in triplicates, and the average element concentration was used. Elements below detection limits on all tributary source samples were excluded from subsequent analyses (Electronic Supplementary Material Table 1). P_2O_5 was not considered as a possible tracer due to potential biogeochemical transformations during transport in aquatic environments (Koiter et al. 2013b; Cooper et al. 2015; Sherriff et al. 2015). Unfortunately, the portable XRF spectrometer broke down near the end of analyses. Accordingly, for the intermediate particle size fraction, one source sample from the mid catchment and two catchment outlet samples were not analyzed.

2.4 Artificial mixtures

To test the accuracy and precision of the un-mixing models, a set of 10 artificial mixtures with different known relative source contributions were produced for each sediment size fraction (Table 2). Sub-samples of equal mass were retrieved from each of the individual dried/sieved composite samples. The sub-samples from the same source units were then combined in a source pool, which was later used to create mixtures with known source mass proportions. Elemental composition of the artificial mixtures was used to solve the un-mixing models as if the artificial mixtures comprised the outlet target sediment. Similar approaches to model testing have been adopted by Cooper et al. (2014), Haddadchi et al. (2014), and Pulley et al. (2017).

2.5 Element selection

In this study, widely used statistical procedures to tracer selection were compared to a process-based methodology, where prior knowledge of soil geochemistry is used to identify elements that are expected to provide source discrimination. For the statistical approach, a commonly used three-step method to element selection was employed. First, box-plots were used to evaluate if elements on target samples plotted within the mixing polygon defined by element concentrations on individual source types. Elements on target sediments with a range of variation plotting outside the source ranges were excluded, as tracer properties outside mixing polygons violate numerical modeling assumptions and may lead to spurious results (Collins et al. 2013). Box-plot range of variation is defined as the 25th and 75th percentiles ± extreme values within 1.5 times the interquartile range (IQR). The use of these ranges helps to select elements which are well bounded by the distributions of the mixing polygon. If only minimum and maximum values are taken into account, element distributions from target sediments may plot outside all but potentially one of the source samples. This would bias the un-mixing model solutions in the Monte Carlo simulation, which samples parameter values from data distributions. Elements within the source range were grouped by source and then tested for normality with a Shapiro-Wilk test. When the null hypothesis that the data comes from a normal distribution was rejected ($p < 0.05$), the elements were analyzed with a Kruskal-Wallis *H*-test. Otherwise, elements were analyzed with an ANOVA. Elements that provided significant discrimination between sources ($p < 0.05$) were analyzed with a forward step-wise linear discriminant analysis (LDA) (niveau = 0.1) in order to select a minimum set of variables that maximizes source discrimination (Collins et al. 2010). All statistical analyses were performed with R software (R Core Team 2017). Packages MASS (Venables and Ripley 2012) and klaR (Weihs et al. 2005) were used for the multivariate analyses.

The knowledge-based approach to element selection essentially relies on the interpretation of the theoretical source apportionment and sampling design. While the upper and mid catchment areas have a similar parent material, soil classes may provide an adequate stratification: Paleudults from the upper area are more weathered-leached than Dystrudepts from the mid catchment, which means that the first soils are deeper, have higher clay content and higher residual concentration of Al- and Fe-oxides than the latter (Kämpf and Curi 2012). The lower catchment provides more of a challenge, given that the soil map presents an association of Ustorthents and Hapludoxes. However, a greater tributary density is associated to shallow headwaters (Fig. 1), which allows us to assume that sediments from the lower area will have a greater connection to the least weathered-leached soils in the catchment. Hence, it is expected that this sediment source will be characterized by much higher contents of SiO_2, due to minimal dessilification. Mica-inherited K_2O may also be found in greater quantities than it should be expected in the mid and upper areas. Ti and Zr are some of the most resistant elements in the soil (Marques et al. 2004; Koiter et al. 2013a) and would be expected to occur at reduced concentrations in the lower catchment sediments, reflecting the younger parent material and the underdeveloped soils. Accordingly, Al_2O_3, Fe, K_2O, SiO_2, Ti, and Zr were proposed as potential knowledge-based tracers. Elements from target samples plotting outside the source range of variation were excluded from modeling, similarly to the statistical approach, for each sediment particle size fraction. The selected knowledge-based tracers were also analyzed with a LDA to compare the reclassification accuracy of the element selection methods.

2.6 Modeling

Source contributions were estimated by minimizing the sum of squared residuals (SSR) of the mass balance un-mixing model:

$$SSR = \sum_{i=1}^{n} [(C_i - \sum_{s=1}^{m} P_s S_{si})/C_i]^2 \tag{1}$$

where n is the number of elements used for modeling, C_i is the concentration of element i in the target sediment, m is the number of sources, P_s is the optimized relative contribution of source

s, and S_{si} is the concentration of element i in source s. Optimization constraints were set to ensure that source contributions P_s were non-negative and that their sum equaled 1.

The un-mixing model was solved by a Monte Carlo simulation with 2500 iterations. In each iteration, target and source element concentrations were sampled from a multivariate normal distribution, which preserves correlations between variables (Cooper et al. 2014). Prior to modeling the multivariate distributions, element concentrations were log transformed to ensure a near normal distribution and to avoid possible negative concentration values. During the Monte Carlo simulation, element concentrations were back-transformed by an exponential function. R packages foreach (Calway et al. 2017) and Rsolnp (Ghalanos and Theussl 2015) were used to script the simulations and the optimization functions, respectively. Modeling results are presented as the median and the IQR of possible un-mixing model solutions based on the Monte Carlo simulations. The IQR is a more adequate measure of variability for highly skewed data than the standard deviation, as it is not influenced by extreme values (Sainani 2012). Local optimization functions typically produce heavily skewed data, as some model realizations lead to best fit scenarios where one source provides 100 % of the sediments and others 0 % (Cooper et al. 2014). Accordingly, the IQR may provide a more informative representation of parameter distributions than broader confidence intervals.

Model accuracy was evaluated against artificial mixtures according to their Mean Absolute Error (MAE):

$$MAE = \sum_{s=1}^{m} \frac{|X_s - P_s|}{m} \tag{2}$$

where: m is the number of sources, X_s is the known proportion of source s on the artificial mixture, and P_s is the median of modeled relative contribution of source s. Sediment geochemical data and R un-mixing model scripts are included as Electronic Supplementary Material.

3 Results

3.1 Element selection and source analysis

Of all the 45 analyzed elements, 19 (42 %) were below detection limit on all source samples for the coarse (2 mm – 0.2 mm) and intermediate (0.2 mm – 0.062 mm) fractions, whereas 13 (29 %) elements were not detected for the fine fraction (< 0.062 mm) (Electronic Supplementary Material Table 1). Of the detected elements, only 13 (52 %) plotted within the mixing polygons for the coarse fraction, mainly because of higher element concentrations in the outlet target sediments (all element selection results are displayed in Table 3). Concentrations of major (e.g. K_2O and CaO) and trace elements (e.g. Y and Sr) were enriched in the outlet sediment when compared to source samples. For the intermediate and fine fractions, 22 (88 %) and 30 (97 %) elements plotted within the source mixing polygons, respectively.

Of the elements plotting within the mixing polygon for the coarse and intermediate fractions, five (38%) and six (27 %) elements respectively, failed to provide significant discrimination between sources according to the Kruskal-Wallis H-test (or ANOVA for normally distributed elements) (Electronic Supplementary Material Tables 2-4). For the fine fraction, only four elements (13 %) failed to reject the null hypothesis of the employed statistical tests (Electronic Supplementary Material Table 5).

The forward step-wise LDA selected four elements for modeling the coarse fraction (Fe, Cl, SiO_2, and V), which were able to correctly reclassify only 64 % of the samples according to a cross-validation (Fig. 3). For the intermediate fraction, nine elements (Al_2O_3, CaO, Fe, K_2O, Mo, Ti, V, Y, and Zn) selected by the LDA correctly reclassified 84 % of the samples. For the fine fraction, eight selected elements (Al_2O_3, Ba, Ce, K_2O, Nb, Pb, Y, and Zr) yielded 90 % reclassification accuracy.

For the knowledge-based elements proposed for modeling (Al_2O_3, Fe, K_2O, SiO_2, Ti, and Zr); only Al_2O_3, Fe, SiO2, Ti and Zr plotted within the mixing polygon for the coarse fraction. No elements were outside the source range for the intermediate fraction, and therefore all proposed elements were used in the un-mixing model. For the fine fraction, a depletion of Fe contents on target samples led to the exclusion of this element from analysis.

The LDA reclassification accuracy was on average 9 % lower for the knowledge-based element selection method in comparison to the statistical approach, which could be expected. However, a similar trend of increasing accuracy was observed with a decrease of particle size, as the percentage of correctly reclassified samples ranged from 58 % for the coarse fraction to 78 % and 80 % on the intermediate and fine fractions, respectively.

Overall, the behavior of the knowledge-based proposed elements for all size fractions was in accordance with the anticipated scenario used to stratify sediment sources in the Ingaí catchment: sediments from catchment headwaters (S1) are derived from more weathered-leached soils (mainly Paleudults), with a higher residual concentration of Fe, Al_2O_3, Ti, and Zr (Fig. 4). Samples from the lower catchment (S3) display decreased Fe, Al_2O_3, Ti, and Zr contents and a higher concentration of SiO_2, which confirms that these sediments were generated from younger soils (mainly Ustorthents). Samples from the mid catchment (S2), where Dystrudepts are the main soil class, have intermediate concentrations of the discussed elements in comparison to S1 and S3. Also expectedly, K_2O contents were higher overall on S3 samples, except for the coarse fraction.

3.2 Artificial mixtures and model evaluation

The comparison between modeled source contribution and actual mixture proportions demonstrate that modeling the coarse fraction yielded the poorest results, with a MAE error of 23.8% on the statistical variable selection model (M1) and 17.8 % on the knowledge-based variable selection model (M2) (Table 4) (Fig. 4). On the intermediate fraction, model error decreased from 22.6 % on M1 to 10.9 % on M2. Results from the fine fraction had the lowest errors and a more similar model performance between M1 (MAE = 12.9 %) and M2 (MAE = 11.8%).

Considering all size fractions, models were more effective at estimating the source contributions of artificial mixtures 1-4 (MAE = 9.8%), in which source proportions varied from 25 to 50 %. Results from artificial mixtures 5-10, in which source proportions ranged from 0 to 75 %, had increased error (MAE = 21.2%). Overall, the models had a greater difficulty distinguishing contributions from S2 (MAE = 18.25%) than from S1 (MAE = 15.0 %) and S3 (MAE = 16.9 %). Such behavior is particularly evident for the fine fraction, where the MAE of M2 decreased from 15.5 % on S2 to 7.2 % on S3.

3.3 Model results for the Ingaí catchment

Source proportions estimated by M1 and M2 for the coarse fraction are highly uncertain, as demonstrated by the prediction intervals on Fig. 6, and no inference can be made based on the data. Moreover, considering the median source proportions estimates, the models display contrasting results: M1 indicates that target sediments are derived mainly from S2 (median = 40 %; IQR = 0-87 %), whereas M2 signals a higher contribution from S1 (median = 39 %; IQR = 5-76 %) (Electronic Supplementary Material Table 6).

Results from the Monte Carlo simulations again demonstrate a high degree of uncertainty for the intermediate fraction source apportionments, which are contrasting between models. For the intermediate fraction, M1 estimates that the contribution to outlet sediments are dominated by S2 (median = 57 %, IQR = 8-100 %), whereas M2 estimates reveal a greater contribution from S3 (median = 60 %, IQR = 0-94 %) and S1 (median = 16 %, IQR = 0-61%).

For the fine fraction, the simulation results display much narrower source apportionment estimates. M1 indicates that contributions from S1 (median = 0 %, IQR = 0-3 %) and S2 (median = 0 %, IQR = 0-18 %) are negligible, with target sediments being almost completely derived from S3 (median = 93 %, IQR = 71-100 %). M2 results are nearly identical, estimating that S3 (median = 96 %, IQR = 77-100 %) is again the dominant source, with insignificant contributions from S1 (median = 0 %, IQR = 0-2 %) and S2 (median = 0 %, IQR = 0-11 %).

4 Discussion

Source signal development in the Ingaí catchment is controlled primarily by pedogenetic processes, which display different degrees of expression across particle sizes. Such behavior was reflected throughout this research, starting with the elements indentified by XRF analysis. Fewer elements were detected for the coarse and intermediate fractions (Table 3), which could be expected, since trace elements are retained in greater quantities in finer particles (Antoniadis et al. 2017). Moreover, a greater proportion of detected elements for the coarse and intermediate fractions were outside the source range. We deliberately avoided using the term conservative behavior to describe this process, as we do not have evidence that the elements failing to plot within source range were depleted or enriched during sediment transport due to biogeochemical mechanisms or to changes in physical properties, including grain size distributions. Nevertheless, the greater number of elements plotting outside source mixing polygons, particularly for coarse sediments, may indicate that there has been particle size selectivity occurring during mobilization, transportation and deposition processes or there could be a missing/unsampled source of coarse material near the catchment outlet (Smith and Blake 2014; Laceby et al. 2015).

By comparing the composition of target and source samples, it can be observed that unlike the source sediments, in which Al_2O_3 increased with decreasing particle size, the highest Al_2O_3 contents on the catchment outlet target sediments were associated with the intermediate and coarse size fractions (Fig. 7). Moreover, the coarse fraction had the highest Fe and the lowest SiO_2 concentrations, which is also inconsistent with the tributary source sample patterns. Within soils derived from a same parent material, elements found in stable clay minerals (e.g. Al_2O_3 and Fe) usually occur in greater residual concentrations on finer particles, as demonstrated by Silva et al. (2018). Contrarily, SiO_2 decreases with particle size, due to dessilification and of the lower stability of quartz in the clay fraction (Fontes 2012). The higher concentration of Al_2O_3 and Fe for the coarse and intermediate fractions of the target sediments may therefore suggest that these fractions have received a greater contribution of sediments derived from a contrasting parent material compared to the sources influencing the fine fraction. Such parent material is likely to have been un- or under-sampled, which may explain the number of elements plotting outside the source range for the coarser sediments.

Results from the analyses of variance and the LDA also demonstrate contrasting patterns regarding the geochemical composition of sediments across particle sizes. Fewer elements provided statistical discrimination between sources for the coarse and intermediate sediments compared to the finer fraction, according to the employed tests (i.e. ANOVA or Kruskal-Wallis) (Table 3). These results demonstrate that the source stratification was more effective for the fine fraction, likely because the geochemical source signal in the Ingaí catchment is mainly associated to pedogenetic processes (e.g. dessilification, residual accumulation of Al and Fe in pedogenetic oxides). These processes are more clearly expressed on finer, more weathered-leached particles, and particularly on clay minerals (Kämpf et al. 2012). Conversely, the coarser particles may be more representative of the parent material (Curi and Kämpf 2012), which is less contrasting among the sources in the catchment. These findings may also reflect on the poor reclassification accuracy of the LDA for the coarse fraction (64 % and 58 % for the statistical and knowledge-based approach, respectively) when compared to the intermediate (84 % and 78 % for the statistical and knowledge-based approach, respectively) and fine fractions (90 % and 80 % for the statistical and knowledge-based approach, respectively). Interestingly, the forward step-wise LDA selected elements that were also proposed by the knowledge-based approach for all size fractions (SiO_2 and Fe for the coarse, Al_2O_3, Fe, K_2O, and Ti for the intermediate, and Al_2O_3, K_2O and Zr for the fine). This demonstrates that these elements provide both statistical and pedological discrimination between sources.

The model evaluation against artificial mixtures corroborates the lack of source discrimination for the coarse fraction, in which the MAE for both statistical (M1) (23.8 %) and knowledge-based (M2) (17.8 %) models is higher than what is usually reported on similar studies (e.g. Haddadchi et al. 2014; Pulley et al. 2017; Cooper and Krueger 2017). For the intermediate fraction, although M2 yielded the lowest MAE (10.9 %) among the analyzed models and particle sizes, high errors were again associated to M1 (22.6%). In contrast, M1 (MAE = 12.9 %) and M2 (11.8 %) displayed a similar performance for the fine sediments, and a greater confidence can be ascribed to model predictions for this fraction

The modeling results for the catchment outlet target sediments for the coarse fraction again demonstrate poor source discrimination, given the uncertainty of the estimates (Fig. 6). Moreover, the relative contributions from the upper and mid catchment represented by model predictions seem unlikely considering the results for the finer fraction, which predict with little uncertainty that target outlet sediments are derived almost entirely from the lower catchment. As coarser material is often transported as bed or saltating bed load, at slower rates than the finer wash load (Collins and Walling 2016), proximal sources are usually the major contributors of coarse sediment particles (Haddadchi et al. 2016). Therefore, estimated source contributions from the mid and upper catchment for the coarse fraction are more likely to have been derived from other downstream sources, probably in close proximity to the outlet sediment sampling location, with a similar soil parent material as the mid and upper regions of the watershed.

In a similar way, modeling the intermediate fraction indicated a considerable, although also very uncertain, contribution from the mid and upper catchment for both models (Fig. 6). Again, such contributions seem unlikely to represent sediment dynamics in the catchment, and a missing or under-sampled source located proximately to catchment outlet might be biasing model predictions.

A possible provenance of sediments identified as derived from the upper and mid catchment by the un-mixing models may be related to a strip of orthogneiss located near the outlet of the Ingaí River (Fig. 1). This lithotype comprises only 3 % of the lower catchment and a single composite sample was retrieved from a tributary draining the area. The concentrations of Al_2O_3 (13.9 %), Fe (3.6 %), and SiO_2 (37.0 %) for the coarse sediments from this particular sample were different to the average concentrations of these elements in the other lower catchment samples (Al_2O_3 = 6.1 %, Fe = 2.2 %, SiO_2 = 51.6 %). The sample concentrations are however similar to the average contents of Al_2O_3 (13.8 %), Fe (4.3 %) and SiO_2 (34.0 %) for the coarse fraction of the target outlet sediments. Nevertheless, this interpretation of the modeling results remains speculative, and the most important inference from the data is that the spatial scale of the source stratification was not appropriate for fingerprinting the coarse and intermediate size fractions.

Contrarily to the coarser fractions, the source contributions estimated for the fine sediments are consistent among the employed models (Fig 6). The similarity between model results increases the confidence in the predictions, which are also corroborated by the small errors of the estimated source proportions of the artificial mixtures. Moreover, the results fit with our understanding of erosion and sediment transport dynamics in the catchment.

According to model predictions, the fine sediments collected at the watershed outlet are almost entirely derived from the lower catchment. These sediments are primarily associated with the shallow and underdeveloped Ustorthents from the quartzitic/mica-schistic ridges within the lower catchment, as demonstrated by the higher SiO_2 and K_2O contents and the lower Al_2O_3, Fe, Ti, and Zr concentrations. This Entisols region is erosion prone: the solum is shallow and the underlying C horizon is situated right below the A horizon, decreasing water infiltration and increasing runoff propensity (Araújo 2006). These soils are also located on steep slopes and have elevated contents of silt and fine sand in relation to clay (Curi et al. 1990). Hence, a large sediment supply from these soils in the lower catchment is plausible. Furthermore, the lower catchment is much closer to the Ingaí River outlet than the mid and upper areas. Fine sediments originated from these upstream sources have a greater probability of being stored on floodplains and lower-gradient sections.

Results reported by Le Gall et al. (2017) also show that the contribution of fine sediments from farther upstream sources on large catchments is minor, at least considering the sediments that effectively reach the catchment outlet. Such behavior must be analyzed with caution, as fingerprinting the origin of outlet sediments does not necessarily represent overland and fluvial transport processes elsewhere in the catchment (Koiter et al. 2013a). These considerations might be particularly important in large watersheds, where sediment yield components are likely to be subjected to a variety of travel times and transport energies, which will also vary with particle size (Parsons 2011), as illustrated by our results.

The Ingaí River drains approximately 60 % of the Capivari River basin, which is estimated to supply over 480,000 t yr^{-1} of sediment to the Funil hydroelectric power plant reservoir (Batista et al. 2017). Accordingly, fine sediment from the Ingaí River may contribute significantly to reservoir sedimentation. Soil conservation practices targeting the lower Ingaí Entisols may therefore help minimize fine sediment delivery to the Funil reservoir. According to RUSLE-based estimates (Batista et al. 2017), average erosion rates were the highest in the mid catchment area. Therefore, future research should monitor erosion dynamics across multiple scales and different particle size fractions in the Ingaí catchment. Ultimately it is important to understand how different Critical Zone processes regulate sediment connectivity throughout the catchment in order to help target the implementation of best management practices that limit the deleterious off-site effects of soil erosion.

Overall, our results demonstrate that source stratification and geochemical element selection for sediment fingerprinting can be carried out based on the knowledge of pedogenetic processes that develop source signals in tropical soils. However, such an approach might be less effective for coarse sediment particles, particularly if parent material has few geochemical contrasts. In this sense, a soil-based source stratification might be more powerful for fine sediment fingerprinting than a geological approach, given that pedogenetic processes and soil forming factors other than parent material are also able to generate contrasting source signals, particularly in tropical soils. Nevertheless, for modeling coarse sediment provenance a geological source stratification may be more appropriate, as pronounced lithological dissimilarity might dominate the source signal generation for coarse material.

The comparison between the element selection methods demonstrated that the commonly used three-step statistical approach does not necessarily yield more accurate model predictions, which is corroborated by the results of Smith et al. (2018). However, a valuable outcome of using both methods is that different model predictions can be compared. If similar results are achieved with a different set of variables, a greater confidence can be ascribed to model estimates (Laceby et al. 2015).

A significant advantage of a knowledge-based element selection is that subsequent modeling results are more easily relatable to known source characteristics. In the knowledge-based approach, processes occurring in the Critical Zone that drive source signal development, erosion and sediment transport, can be conjointly analyzed. This contributes to a more comprehensive understanding of these processes, and generates multiples lines of evidence to corroborate or falsify model assumptions and predictions. The use of the knowledge-based approach encourages researchers to understand the fundamental Critical Zone processes driving erosion and sediment geochemistry across multiple scales. This increased understanding of fundamental processes is instrumental to improve catchment sediment management strategies, particularly in erosion-prone tropical environments.

5 Conclusions

In this research, the pedological knowledge of tropical soils was incorporated into source stratification and geochemical element selection in a fingerprinting study across three particle size fractions. Our approach provided source discrimination for the fine and intermediate size fractions, as demonstrated by the comparison of the un-mixing model estimates and artificial mixture proportions. However, the source stratification was unable to provide sufficient geochemical discrimination for the coarse sediments. This probably stems from the fact that pedogenetic processes are the main drivers of geochemical contrast and source discrimination between fine sediment sources, whereas geological background may be more likely to drive these contrasts for the coarser material. Model evaluation against the artificial mixtures also indicated that the commonly used three-step statistical approach to variable selection may not always provide the most accurate estimates.

The spatial scale of the source stratification was however unable to represent the coarse and intermediate size sediment dynamics in the catchment, which seems to be controlled by very proximal sources – at least in the temporal scale of the analysis. Hence, different field sampling approaches might be necessary to model specific size fractions in the Ingaí catchment, and potentially in other catchments.

For the fine sediments, both knowledge-based and the statistical methods to geochemical element selection yielded very similar results: Ustorthents from the lower catchment ridges are by far the main sediment source reaching the Ingaí River outlet. The consistent model results increase confidence in the predictions. Moreover, the knowledge-based method facilitates the interpretation of the results, as the selected fingerprinting properties can be explicitly related to upstream processes regarding source signals and behavioral characteristics of the soils comprising each end-member source. This enhanced interpretation of fingerprint models provides a framework for an integrated assessment of Critical Zone dynamics, linking soil and parent material geochemistry to soil erosion and sediment transport processes in river catchments.

The source stratification procedure and the knowledge-based element selection for sediment fingerprinting described in this study have potential to improve sediment management strategies across Brazil and around the world. This approach would be particularly useful in large catchments where soil parent materials have similar geochemistry, and source signal development of fine sediments is controlled by pedogenetic processes. Ultimately, understanding the fundamental pedogenetic processes driving the formation of source signatures will likely aid in the management of the dominant Critical Zone processes driving erosion in Brazil and in other tropical regions where intense weathering-leaching leads to unique expressions of soil forming processes.

Acknowledgements This research was funded by the Coordination of Improvement of Higher Level Education Personnel – CAPES – Finance Code 001 (Process number 88881.190317/2018-01), the National Counsel of Technological and Scientific Development – CNPq (Process number 306511-2017-7, 202938/2018-2 and 150.689/2017-8), and the Minas Gerais State Research Foundation – FAPEMIG (Process number CAG-APQ01053-15). The authors are thankful for the support from the undergraduate students Tom de Lima, Fabio Quaresma, and Wesley Soares during laboratory work. The statistical and R scripting advices from Teotonio de Carvalho are highly appreciated. We also thank two anonymous reviewers for their valuable suggestions.

References

Alvares CA, Stape JL, Sentelhas PC et al (2013) Köppen's climate classification map for Brazil. Meteorol Zeitschrift 22:711–728. doi: 10.1127/0941-2948/2013/0507

Amundson R, Richter DD, Humphreys GS (2007) Biota and earth materials in the Critical Zone.

Elements 3:327–332. doi: 10.2113/gselements.3.5.327

Antoniadis V, Levizou E, Shaheen SM et al (2017) Trace elements in the soil-plant interface: Phytoavailability, translocation, and phytoremediation–A review. Earth-Science Rev 171:621–645. doi: 10.1016/j.earscirev.2017.06.005

Araújo ARDE (2006) Solos da bacia do Alto Rio Grande (MG): Base para estudos hidrológicos e aptidão agrícola. Lavras Federal University. Brazil.

Bajard M, Poulenard J, Sabatier P et al (2017) Progressive and regressive soil evolution phases in the Anthropocene. Catena 150:39–52. doi: 10.1016/j.catena.2016.11.001

Banwart S (2011) Save our soils. Nature 474:151–152. doi: 10.1038/474151

Batista PVG, Silva MLN, Silva BPC et al (2017) Modelling spatially distributed soil losses and sediment yield in the Upper Grande River basin, Brazil. Catena 157:139–150. doi: 10.1016/j.catena.2017.05.025

Borges FP (2011) Atividade de 18 mineradoras é suspensa no Sul de Minas Gerais. Estado de Minas. Retrieved from https://www.em.com.br. Brazil.

Calway R, Microsoft, Weston S (2017) foreach: Provides foreach looping construct for R. R package version 1.4.4.

CODEMIG - CPRM (2014) Mapa geológico de Minas Gerais. Belo Horizonte, Comanhia de Desenvolvimento Econômico de Minas Gerais. Brazil.

Collins AL, Pulley S, Foster IDL et al (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 194:86–118 doi: 10.1016/j.jenvman.2016.09.075

Collins AL, Walling DE, Leeks GJL (1996) Composite fingerprinting of the spatial source of fluvial suspended sediment: a case study of Exe and Severn River basins, United Kingdom. Geomorphol Relief, Process Environ 2:41–54.

Collins AL, Walling DE, Webb L, King P (2010) Apportioning catchment scale sediment sources using a modified composite fingerprinting technique incorporating property weightings and prior information. Geoderma 155:249–261. doi: 10.1016/j.geoderma.2009.12.008

Collins AL, Zhang YS, Hickinbotham R, et al (2013) Contemporary fine-grained bed sediment sources across the River Wensum Demonstration Test Catchment, UK. Hydrol Process 27:857–884. doi: 10.1002/hyp.9654

Cooper RJ, Krueger T (2017) An extended Bayesian sediment fingerprinting mixing model for the full Bayes treatment of geochemical uncertainties. Hydrol Process 31:1900–1912. doi: 10.1002/hyp.11154

Cooper RJ, Krueger T, Hiscock KM, Rawlins BG (2014) Sensitivity of fluvial sediment source apportionment to mixing model assumptions: A Bayesian model comparison. Water Resour Res 50:9031–9047. doi: 10.1002/2014WR016194.

Cooper RJ, Krueger T, Hiscock KM, Rawlins BG (2015) High-temporal resolution fluvial sediment source fingerprinting with uncertainty: A Bayesian approach. Earth Surf Process Landforms 40:78–92. doi: 10.1002/esp.3621

Curi N, Kämpf N (2012) Caracterização do solo. In:Ker JC, Curi N, Schaefer CEGR, Vidal-Torrado P (eds) Pedologia fundamentos. Sociedade Brasileira de Ciência do Solo, Viçosa, pp 147-170. Brazil.

Curi N, Lima JM De, Andrade H, Gualberto V (1990) Geomorfologia, física, química e mineralogia dos principais solos da região de Lavras (MG). Ciência e Agrotecnologia 14:297–307. Brazil.

Evrard O, Poulenard J, Némery J et al (2013) Tracing sediment sources in a tropical highland catchment of central Mexico by using conventional and alternative fingerprinting methods. Hydrol Process 27:911–922. doi: 10.1002/hyp.9421

FEAM - Fundação Estadual Do Meio Ambiente (2010) Mapa de solos de Minas Gerais: legenda expandida. FEAM/UFV/CETEC/UFLA, Belo Horizonte. Brazil.

G1 Sul de Minas (2016) Jazidas de quartzito tem atividades suspensas em Luminárias, MG. Retrieved from http://g1.globo.com/mg/sul-de-minas. Brazil.

Ghalanos A, Theussl S (2015) Rsolnp: General non-linear optimization using augmented lagrange multiplier method. R package version 1.16

Haddadchi A, Olley J, Laceby JP (2014) Accuracy of mixing models in predicting sediment source contributions. Sci Total Environ 497–498:139–152. doi: 10.1016/j.scitotenv.2014.07.105

Haddadchi A, Olley J, Pietsch T (2016) Variable source contributions to river bed sediments across three size fractions. Hydrol Process 30:1609–1623. doi: 10.1002/hyp.10732

Hatfield RG, Maher BA (2009) Fingerprinting upland sediment sources: particle size-specific magnetic linkages between soils, lake sediments and suspended sediments. Earth Surf Process Landforms 34:155–161. doi: 10.1002/esp

Hijmans RJ, Cameron SE, Parra JL et al (2005) Very high resolution interpolated climate surfaces for global land areas. Int J Climatol 25:1965–1978. doi: 10.1002/joc.1276

Hu B, Yang Z, Wang H, Sun X, Naishuang B, Li G (2009) Sedimentation in the Three Gorges Dam and the future trend of Changjiang (Yangtze River) sediment flux to the sea. Hydrol Earth Syst Sci 13:2253–2264

Kämpf N, Curi N (2012) Formação e evolução do solo (pedogênese). In:Ker JC, Curi N, Schaefer CEGR, Vidal-Torrado P (eds) Pedologia fundamentos. Sociedade Brasileira de Ciência do Solo, Viçosa, pp 207-302. Brazil.

Kämpf N, Marques JJ, Curi N (2012) Mineralogia de solos brasileiros. In:Ker JC, Curi N, Schaefer CEGR, Vidal-Torrado P (eds) Pedologia fundamentos. Sociedade Brasileira de Ciência do Solo, Viçosa, pp 81-146. Brazil.

Klages MG, Hsieh YP (1975) Suspended solids carried by the Gallatin River of Southwestern Montana: II. Using mineralogy for inferring sources. J Environ Qual 4:68–73.

Koiter AJ, Lobb DA, Owens PN, Petticrew EL, Tiessen KHD, Li S (2013a) Investigating the role of connectivity and scale in assessing the sources of sediment in an agricultural watershed in the Canadian prairies using sediment source fingerprinting. J Soils Sediments 13:1676–1691. doi: 10.1007/s11368-013-0762-7

Koiter AJ, Owens PN, Petticrew EL, Lobb DA (2013b) 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

Laceby JP, Evrard O, Smith HG, et al (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

Laceby JP, McMahon J, Evrard O, Olley J (2015) A comparison of geological and statistical approaches to element selection for sediment fingerprinting. J Soils Sediments 15:2117–2131. doi: 10.1007/s11368-015-1111-9

Laceby JP, Olley J (2015) An examination of geochemical modelling approaches to tracing sediment sources incorporating distribution mixing and elemental correlations. Hydrol Process 29:1669–1685. doi: 10.1002/hyp.10287

Le Gall M, Evrard O, Dapoigny A, et al (2017) Tracing sediment sources in a subtropical agricultural catchment of Southern Brazil cultivated with conventional and conservation farming practices. Land Degrad Dev 28:1426–1436. doi: 10.1002/ldr.2662

Le Gall M, Evrard O, Foucher A et al (2016) Quantifying sediment sources in a lowland agricultural catchment pond using137Cs activities and radiogenic87Sr/86Sr ratios. Sci Total Environ 566–567:968–980. doi: 10.1016/j.scitotenv.2016.05.093

Lepage H, Laceby JP, Bonté P et al (2016) Investigating the source of radiocesium contaminated sediment in two Fukushima coastal catchments with sediment tracing techniques. Anthropocene 13:57–68. doi: 10.1016/j.ancene.2016.01.004

Lin H (2010) Earth's Critical Zone and hydropedology: Concepts, characteristics, and advances. Hydrol Earth Syst Sci 14:25–45. doi: 10.5194/hess-14-25-2010

Marques JJ, Schulze DG, Curi N, Mertzman SA (2004) Major element geochemistry and geomorphic relationships in Brazilian Cerrado soils. Geoderma 119:179–195. doi: 10.1016/S0016-7061(03)00260-X

Motha JA, Wallbrink PJ, Hairsine PB, Grayson RB (2002) Tracer properties of eroded sediment and source material. Hydrol Process 16:1983–2000. doi: 10.1002/hyp.397

Olley J, Caitcheon G (2000) Major element chemistry of sediments from the Darling-Barwon River and its tributaries : implications for sediment and phosphorus sources. Hydrol Process : 14:1159–1175

Parsons AJ (2011) How useful are catchment sediment budgets ? Prog Phys Geogr 36:60–71. doi: 10.1177/0309133311424591

Pulley S, Foster I, Collins AL (2017) The impact of catchment source group classification on the accuracy of sediment fingerprinting outputs. J Environ Manage 194:16–26. doi: 10.1016/j.jenvman.2016.04.048

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

Resende M, Curi N, Rezende SB et al (2014) Pedologia: base para distinção de ambientes. Universidade Federal de Lavras, Lavras. 322 p. Brazil.

Ribeiro BT, Silva SHG, Silva EA, Guilherme LRG (2017) Portable X-ray fluorescence (pXRF) applications in tropical soil science. Cienc. e Agrotecnologia 41:245–254. doi: 10.1590/1413-70542017413000117

Sainani KL (2012) Dealing With Non-normal Data. PM R 4:1001–1005. doi:

10.1016/j.pmrj.2012.10.013

Sherriff SC, Franks SW, Rowan JS, Fenton O, Ó'hUllacháin D (2015) Uncertainty-based assessment of tracer selection, tracer non-conservativeness and multiple solutions in sediment fingerprinting using synthetic and field data. J Soils Sediments 15:2101–2116. doi: 10.1007/s11368-015-1123-5

Silva SHG, Hartemink AE, Teixeira AFS et al (2018) Soil weathering analysis using a portable X-ray fluorescence (PXRF) spectrometer in an Inceptisol from the Brazilian Cerrado. Appl Clay Sci 162:27–37. doi: 10.1016/j.clay.2018.05.028

Silva SHG, Silva EA, Poggere GC et al (2017) Tropical soils characterization at low cost and time using portable X-ray fluorescence spectrometer (PXRF): Effects of different sample preparation methods. Cienc e Agrotecnologia 42:80–92. doi: 10.1590/1413-70542018421009117

Smith HG, Blake WH (2014) Sediment fingerprinting in agricultural catchments: A critical re-examination of source discrimination and data corrections. Geomorphology 204:177–191. doi: 10.1016/j.geomorph.2013.08.003

Smith HG, Karam DS, Lennard AT (2018) Evaluating tracer selection for catchment sediment fingerprinting. J Soils Sediments 18:3005-3019. doi: 10.1007/s11368-018-1990-7

Theuring P, Collins AL, Rode M (2015) Source identification of fine-grained suspended sediment in the Kharaa River basin, northern Mongolia. Sci Total Environ 526:77–87. doi: 10.1016/j.scitotenv.2015.03.134

Tiecher T, Caner L, Minella JPG et al (2016) Tracing sediment sources in a subtropical rural catchment of southern Brazil by using geochemical tracers and near-infrared spectroscopy. Soil Tillage Res 155:478–491. doi: 10.1016/j.still.2015.03.001

Vale SS, Fuller IC, Procter JN et al (2016) Characterization and quanti fi cation of suspended sediment sources to the Manawatu River, New Zealand. Sci Total Environ 543:171–186.

doi: 10.1016/j.scitotenv.2015.11.003

Venables WN, Ripley BD (2002) Modern applied statistics with S. Fourth Edition. Springer, New York.

Voli MT, Wegmann KW, Bohnenstiehl DR, Leithold E, Osburn CL, Polyakov V (2013) Fingerprinting the sources of suspended sediment delivery to a large municipal drinking water reservoir: Falls Lake, Neuse River, North Carolina, USA. J Soils Sediments 13:1692–1707. doi: 10.1007/s11368-013-0758-3

Walling DE, Owens PN, Waterfall BD et al (2000) The particle size characteristics of fluvial suspended sediment in the Humber and Tweed catchments, UK. Sci Total Environ 251–252:205–222. doi: 10.1016/S0048-9697(00)00384-3

Walling DE, Woodward JC (1995) Tracing sources of suspended sediment in river basins: a case study of the River Clum, Devon, UK. Mar Freshw Res 46:327–336.

Weihs C, Ligges U, Luebke K, Raabe N (2005) klaR analyzing German business cycles. In Baier D, Decker R, Schmidt-Thieme L. (Eds.). Data analysis and decision support, Springer-Verlag, Berlin, pp 335-343.

Wilkinson SN, Hancock GJ, Bartley R et al (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 SN, Olley JM, Furuichi T, Burton J, Kinsey-Henderson AE (2015) Sediment source tracing with stratified sampling and weightings based on spatial gradients in soil erosion. J Soils Sediments 15:2038–2051. doi: 10.1007/s11368-015-1134-2

Yang Y, Gao B, Hao H et al (2017) Nitrogen and phosphorus in sediments in China: A national-scale assessment and review. Sci Total Environ 576:840–849. doi: 10.1016/j.scitotenv.2016.10.136

Yu L, Oldfield F (1989) A multivariate mixing model for identifying sediment source from

from magnetic measurements. Quat Res 32:168–181.

Zamparas M, Zacharias I (2014) Restoration of eutrophic freshwater by managing internal nutrient loads. A review. Sci Total Environ 496:551–562. doi: 10.1016/j.scitotenv.2014.07.076

Zhao G, Mu X, Han M, et al (2017) Sediment yield and sources in dam-controlled watersheds on the northern Loess Plateau. Catena 149:110–119. doi: 10.1016/j.catena.2016.09.010

TABLES

Table 1. Percentage area distribution of soil classes, lithological units, and land use in the Ingaí River basin and source groups.

	Ingaí River basin	S1	S2	S3
Soil classes		area %		
Usorthents and rock outcropts	27	16	18	46
Dystrudepts	24	16	54	-
Paleudults	16	48	3	-
Hapludoxes	33	20	25	54
Lithology		area %		
Paragneiss	20	38	-	-
Biotite-schist	14	22	20	22
Quartzite, phyllite, mica-schist	15	-	1	44
Schist-metagraywacke	14	8	3	31
Orthogneiss	34	32	65	3
Clastic sediments	3	-	11	-
Landuse		area %		
Pasture	49	64	53	31
Forest	27	26	28	27
Rupestrian fields*	11	-	1	31
Cropland	9	7	12	7
Eucalypt	4	3	5	3
Other	-	-	-	1

S1: upper catchment; S2: mid catchment; S3: lower catchment. *Grassland herbaceous/ sub-shrubby formation usually associated to quartzitic ridges.

Table 2. Artificial mixtures with known source contributions used for model evaluation.

Artificial Mixtures	Sources		
	S1	S2	S3
	Relative contributions (%)		
1	33	33	33
2	50	25	25
3	25	50	25
4	25	25	50
5	75	25	0
6	75	0	25
7	25	75	0
8	0	75	25
9	0	25	75
10	25	0	75

Table 3. Selected elements for modeling after each step of the statistical procedure for each size fraction.

Size fraction (mm)	Selection step	Selected elements	% correctly classified samples
2 - 0.2	1	Al_2O_3, Bi, Cl, Fe, Mo, Nb, Rh, S, SiO_2, Ta, Ti, V, Zr	
	2	Al_2O_3, Cl, Fe, Nb, SiO_2, Ti, V, Zr	
	3	Fe, SiO_2, Cl, V	64
0.2 - 0.062	1	Al_2O_3, Bi, CaO, Cl, Cr, Cu, Fe, K_2O, Mn, Mo, Nb, Ni, Pb, Rb, S, SiO_2, Ta, Ti, V, Y, Zn, Zr	
	2	Al_2O_3, Bi, Cl, Cu, Fe, K_2O, Mo, Nb, Ni, SiO_2, Ti, V, Y, Zn, Zr	
	3	Al_2O_3, CaO, Fe, K_2O, Mo, Ti, V, Y, Zn	84
< 0.062	1	Ag, Al_2O_3, As, Ba, Bi, CaO, Ce, Cl, Cr, Cu, Hf, K_2O, Mn, Mo, Nb, Ni, Pb, Rb, Rh, S, SiO_2, Sr, Ta, Th, Ti, Tl, V, Y, Zn, Zr	
	2	Ag, Al_2O_3, As, Ba, Bi, CaO, Ce, Cl, Cu, Hf, K_2O, Mo, Nb, Ni, Pb, Rb, Rh, SiO_2, Sr, Ta, Th, Ti, V, Y, Zn, Zr	
	3	Al_2O_3, Ba, Ce, K_2O, Nb, Pb, Y, Zr	90

Step 1: Range of variation; Step 2: Kruskal-Wallis *H*-test or ANOVA; Step 3: step-wise LDA.

Table 4. Mean absolute errors (MAE) of the statistical variable selection model (M1) and the knowledge-based variable selection model (M2) for the three analyzed sediment size fractions.

Artificial Mixture	MAE (%)					
	M1			M2		
	Size Fraction (mm)					
	2 – 0.2	0.2 – 0.062	< 0.062	2 – 0.2	0.2 – 0.062	< 0.062
1	29.3	17.0	4.0	9.3	1.7	3.0
2	5.7	24.3	14.3	6.7	5.0	11.7
3	16.3	8.0	10.0	7.7	6.3	1.3
4	6.0	15.7	5.0	11.0	9.0	6.0
5	39.3	20.3	14.7	13.7	19.7	22.3
6	15.7	31.3	20.0	28.7	14.7	22.0
7	32.0	20.0	15.3	23.0	16.0	12.7
8	15.3	22.7	20.3	25.3	9.0	17.0
9	30.7	36.3	8.0	29.7	10.7	7.3
10	47.3	30.3	17.3	22.7	17.3	14.7
Mean	23.8	22.6	12.9	17.8	10.9	11.8

FIGURES AND CAPTIONS

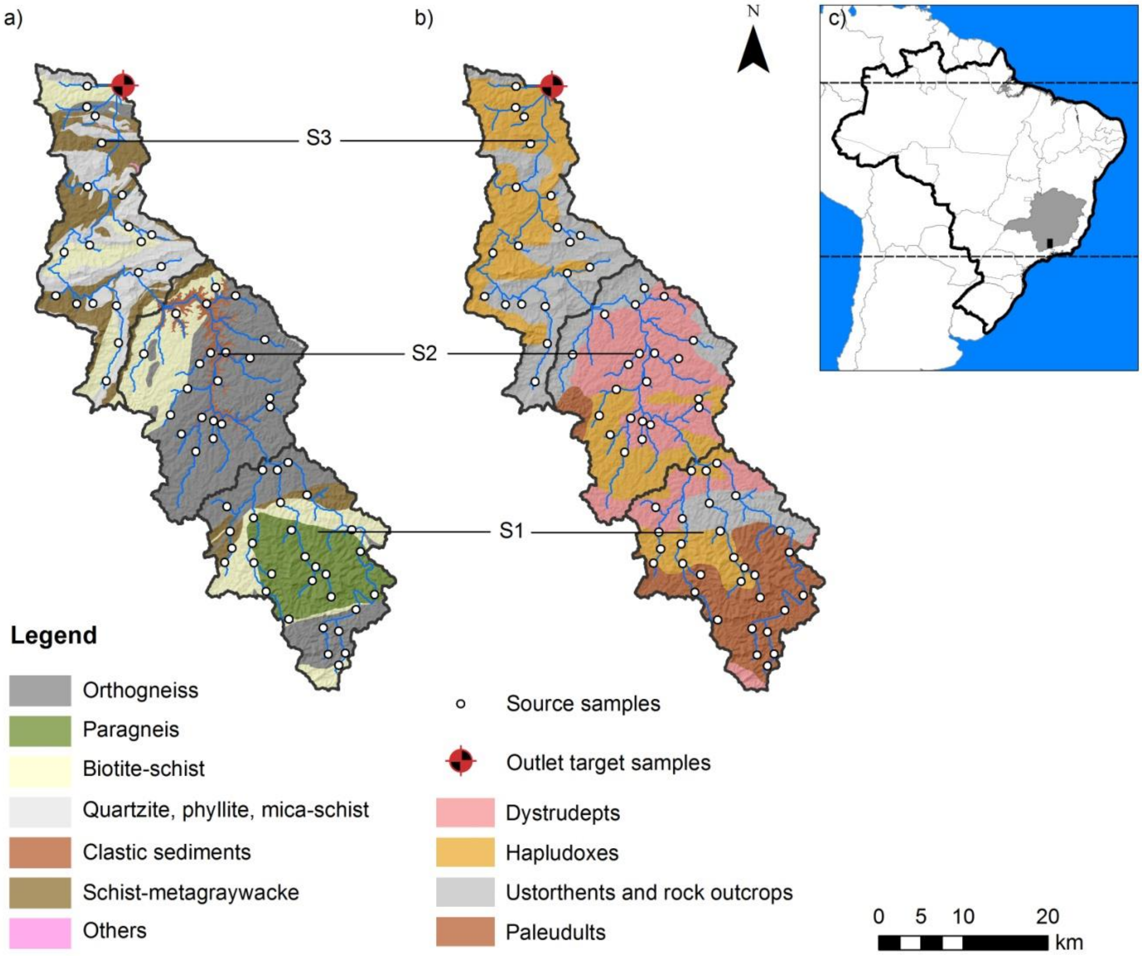


Figure 1 Geological (a) and pedological (b) map of the Ingaí River basin, Brazil (c). S1: upper catchment; S2: mid catchment; S3: lower catchment. Adapted from CODEMIG – CPRM (2014) and FEAM (2010).

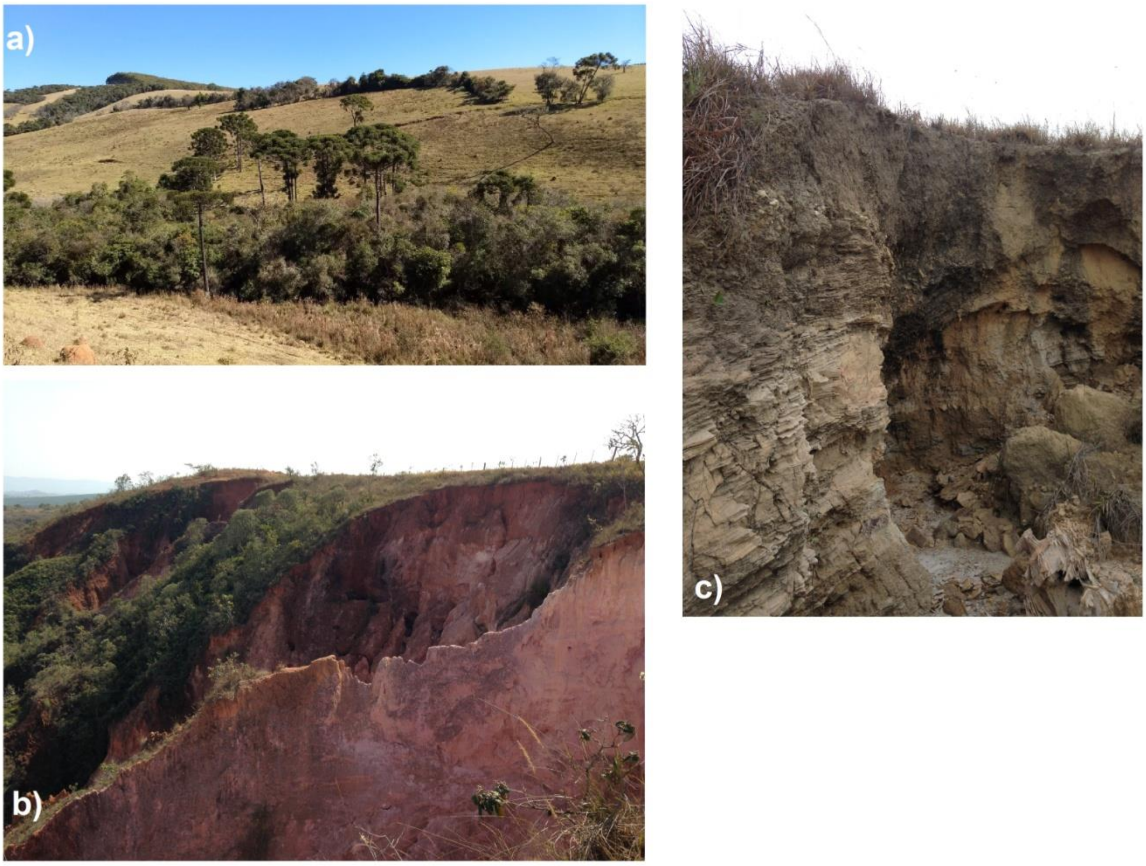


Figure 2 a) Characteristic landscape of the upper catchment; b) gully erosion formed in the intermediate region of the Ingaí basin; c) shallow soils derived from the quartzitic/mica-schistic ridges of the lower catchment.

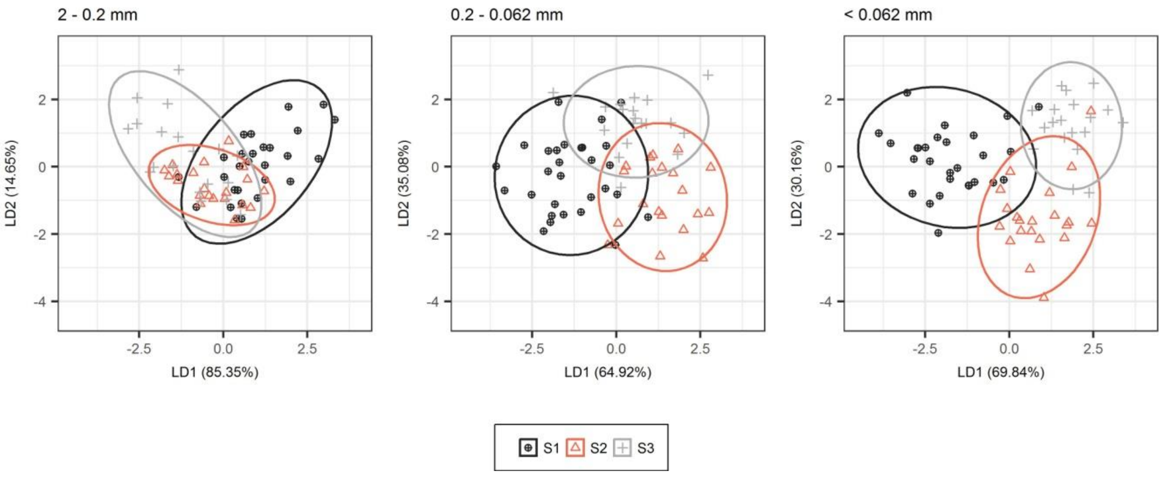


Figure 3 LDA bi-plots of source classification using the selected elements from the statistical approach. Ellipses represent 90 % confidence intervals.

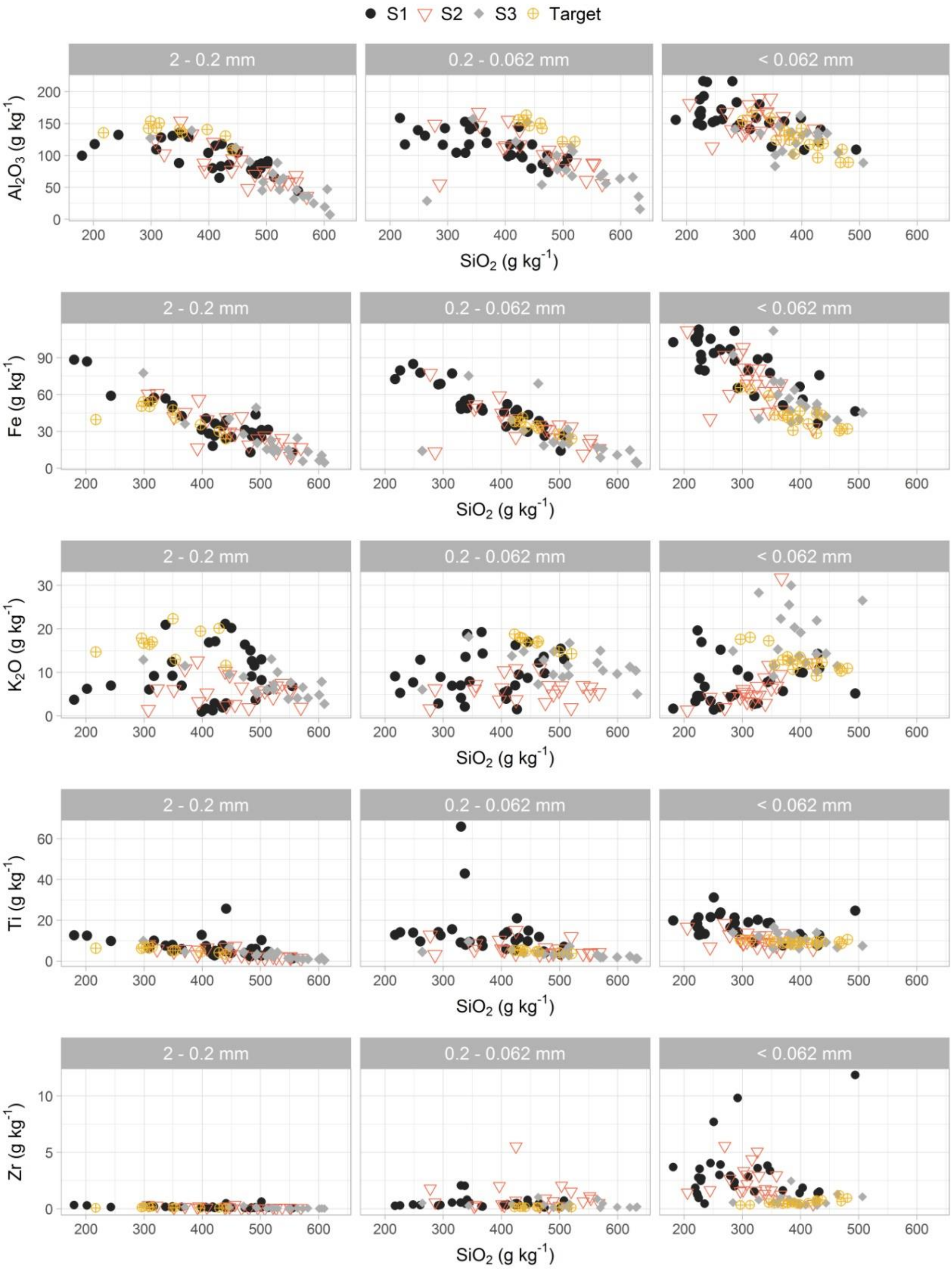


Figure 4 Scatter plots of the knowledge based proposed elements for each sediment size fraction. S1: upper catchment; S2: mid catchment; S3: lower catchment.

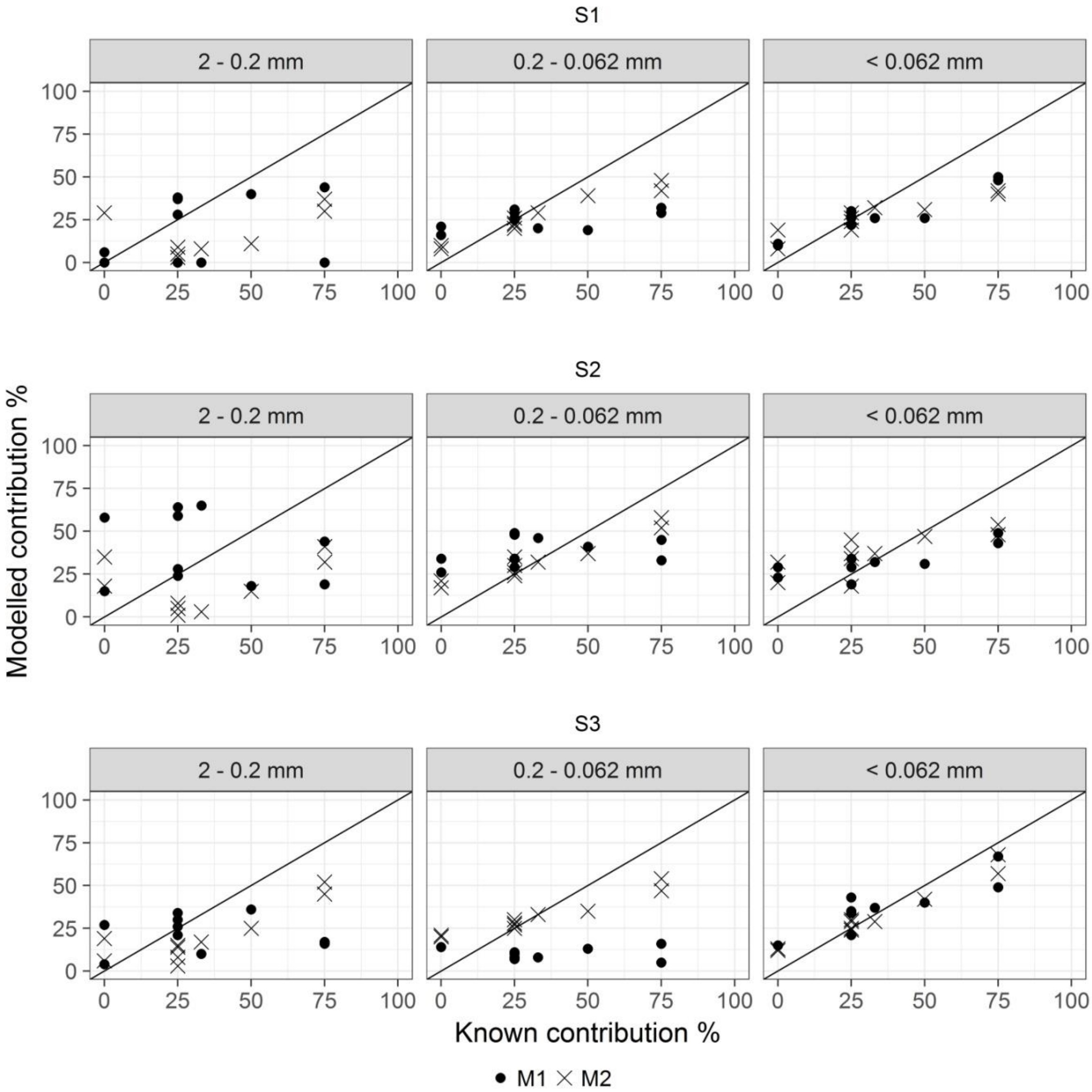


Figure 5 Scatter plots of known and modeled source contributions of the artificial mixtures for each sediment size fraction. S1: upper catchment; S2: mid catchment; S3: lower catchment; M1: statistical element selection model; M2: knowledge-based element selection model.

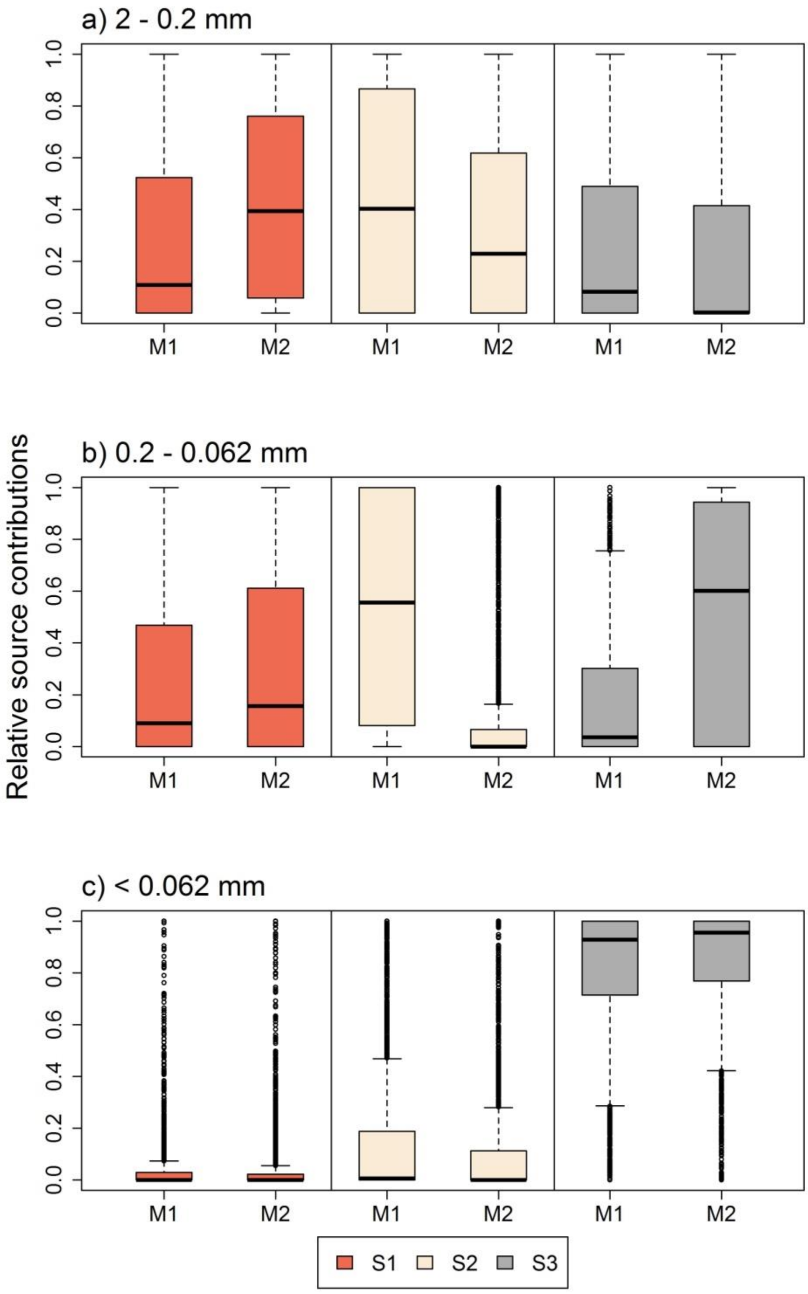


Figure 6 Box plots of estimated source contributions based on the 2500 iterations of the Monte Carlo simulations. a) coarse fraction (2 – 0.2 mm); b) intermediate fraction (0.2-0.062 mm); c) fine fraction (<0.062 mm). S1: upper catchment; S2: mid catchment; S3: lower catchment; M1: statistical element selection model; M2: knowledge-based element selection model.

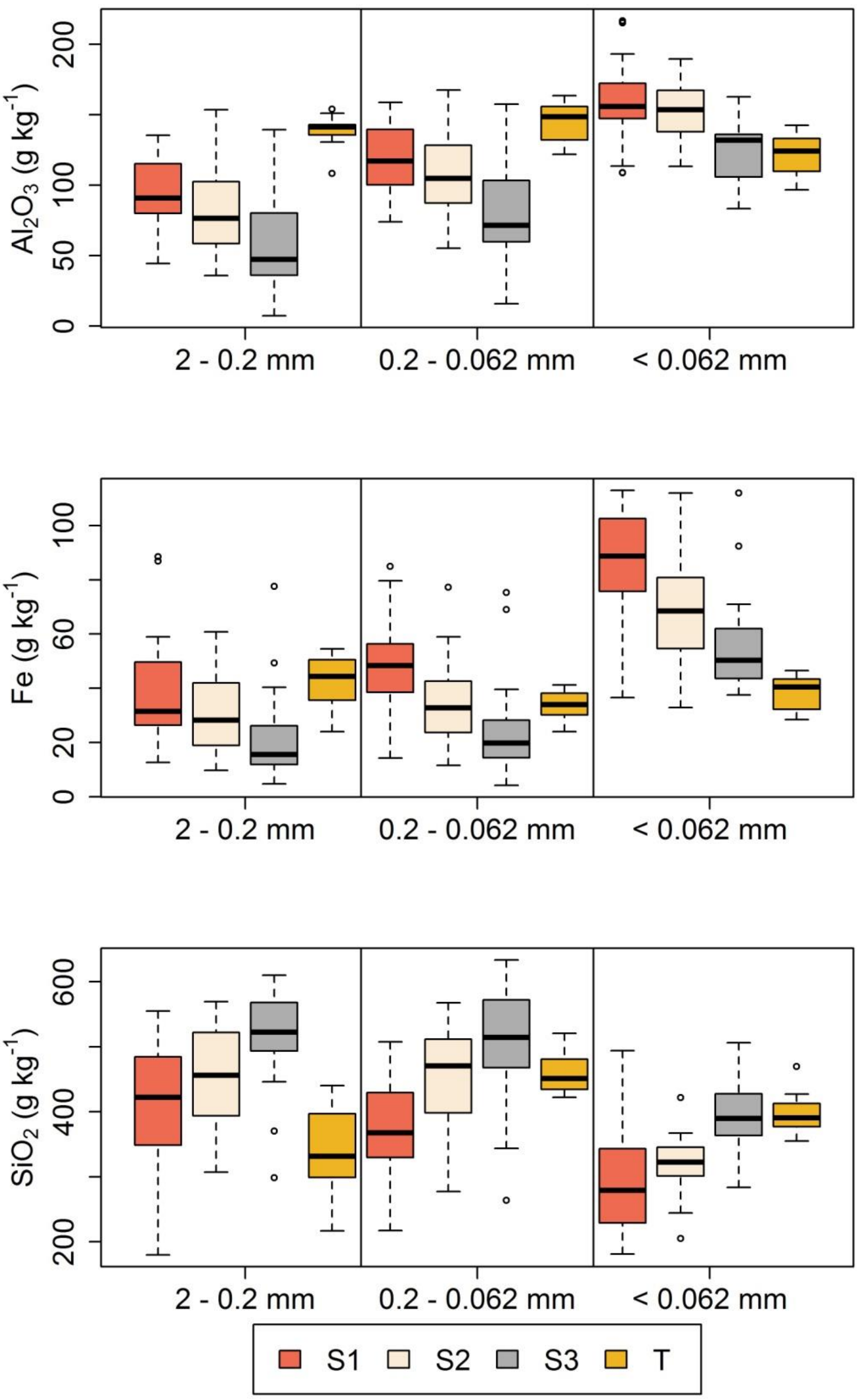


Figure 7 Al_2O_3, Fe, and SiO_2 contents on source (S1: upper catchment; S2: mid catchment; S3: lower catchment) and target (T) sediments.

PAPER 03

Standards of Journal – Environmental Modelling and Software

Running title: **A framework for testing large-scale distributed sediment transport models: dealing with epistemic uncertainty in models and the forcing data**

Pedro V. G. Batista[1]*, J. Patrick Laceby[2], Jessica Davies[3], Teotônio S. Carvalho[1], Diego Tassinari[1], Marx L. N. Silva[1,4], Nilton Curi[1], John N. Quinton[4]

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

2 Alberta Environment and Parks, Environmental Monitoring and Science Division, Calgary, Canada

3 Pentland Centre for Sustainability in Business, Lancaster Environment Centre, Lancaster University, Lancaster, UK

4 Lancaster Environment Centre, Lancaster University, Lancaster, UK

*Corresponding author; email: pbatista.ufla@gmail.com

Keywords: soil erosion models, sediment loads, sediment tracing, RUSLE, SEDD, GLUE.

Abstract

Evaluating the usefulness of spatially-distributed soil erosion models is inherently difficult. Complications stem from the uncertainty in models and measurements of system responses, as well from the scarcity of commensurable spatial data for model testing. Here we present a novel approach for evaluating distributed soil erosion and sediment delivery models, which incorporates sediment source fingerprinting into model testing within a stochastic framework. We applied the Generalized Likelihood Uncertainty Estimation (GLUE) methodology to the Sediment Delivery Distributed (SEDD) model for a large catchment (~6600 km²) in Southeast Brazil. Sediment concentration measurements were used to estimate long-term sediment loads with a sediment rating curve. Regression uncertainty was propagated with posterior simulations of model coefficients. A Monte Carlo simulation was used to generate SEDD model realizations, which were compared against limits of acceptability of model errors derived from

the uncertainty in the curve-estimated sediment loads. Given that SEDD calculations of gross erosion rates are RUSLE-based, we also performed a forward error analysis of RUSLE outputs. The models usefulness for identifying the sediment sources in the catchment was assessed by evaluating behavioral model realizations against sediment fingerprinting source apportionments. Accordingly, we developed a hierarchical tributary sampling design, in which sink sediments were sampled from multiple nodes in the main river channel. The relative contributions of the main sub-catchments in the basin were subsequently estimated by solving the fingerprinting un-mixing model with a Monte Carlo simulation. Results indicate that gauging station measurements of sediment loads were fairly uncertain (average annual specific sediment yields = 0.47 – 11.95 ton ha^{-1} yr^{-1}). This led to 23.4 % of SEDD model realizations being considered behavioral system representations. Spatially-distributed estimates of sediment delivery to water courses were also highly uncertain, as grid-based absolute errors of SEDD results were hundredfold the median of the predictions. Such uncertainty was influenced by the large errors (median = 588 %) associated to RUSLE simulations. A comparison of SEDD outputs and fingerprinting source apportionments revealed an overall agreement between modeled contributions from individual sub-catchments to sediment loads, although some large discrepancies were found in a specific tributary. From a falsificationist perspective the SEDD model could not be rejected, as many model realizations were behavioral. The partial agreement between fingerprinting and SEDD results provide some conditional corroboration of the models capability to identify the sources of sediments in the catchment, at least with some spatial aggregation. However, the uncertainty in the grid-based outputs might dispute the models usefulness for actually quantifying sediment dynamics. The same can be said about RUSLE outputs, which highlights how modeled erosion rates under similar conditions should be interpreted with extreme caution. Ultimately, we have shown how multiple sources of data can – and should be – incorporated into the evaluation of spatially-distributed soil erosion models.

Moreover, we have demonstrated that testing such models requires being explicit about the uncertainty in both models and observational data. Although our results are case-specific, similar levels of error can be expected in erosion models elsewhere. Failing to represent such errors is at best naïve.

1 INTRODUCTION

Spatially-distributed soil erosion and sediment delivery models have received a great attention from the erosion modelling community, arguably due to their potential usefulness for identifying erosion-prone areas and the main sediment sources within large catchments. However, evaluating the usefulness of such models is inherently challenging: measurements of model parameters and system responses are necessarily uncertain, the spatial and temporal resolution of models and observational data are frequently divergent, and the definition of what is a useful model is often subjective (Oreskes and Belitz, 2001). Moreover, our ability to measure erosion rates across landscapes is limited and methods for doing so are known be to at some level flawed (Parsons, 2019). Since model evaluation is an essential step to recognize model failure and to consequently gain knowledge about the modeled phenomena; how should we proceed in face of these aforementioned challenges?

Currently, the most common approach for testing distributed erosion models at the catchment scale is based on a comparison between observed and modeled outlet sediment loads. The estimation of observed loads usually rely on I) suspended solid measurements and/or sediment rating curves (Didoné et al., 2015; Duraes et al., 2016; Jain and Ramsankaran, 2018; Krasa et al., 2019; Vigiak et al., 2015); II) temporally-spaced bathymetric surveys or excavations of ponds and reservoirs (de Vente et al., 2008; Eekhout et al., 2018; Tanyaş et al., 2015); or III) radiometric dating of lake sediment cores (Smith et al., 2018b). Although a comparison against sediment loads can give an indication of a models capability to simulate sediment transport rates at the outlet of a catchment, it provides no information on the adequacy with which models simulate erosion patterns or identify sediment sources. Moreover, models have been known to reproduce observed outlet sediment loads for the wrong reasons, through misrepresenting internal catchment processes (see Pontes, 2018 for an example).

Therefore, the outlet-based approach for testing distributed erosion models has received criticism (Favis-Mortlock et al., 2001; Govers, 2011; Jetten et al., 2003; Parsons et al., 2009), and modelers have pursued other sources of data to evaluate internal process representations. For instance, field monitoring of erosion features combined with volumetric measurements of rills, gullies, and sediment deposition drapes can provide spatially referenced information of internal erosion dynamics that are commensurate with model simulations (Evans and Brazier, 2005; Takken et al., 1999; Van Oost et al., 2005). Alternatively, tracing techniques have been used to estimate medium to long-term soil redistribution rates, which are also comparable to distributed erosion model outputs (Lacoste et al., 2014; Porto and Walling, 2015; Walling et al., 2003; Warren et al., 2005). More recently, Fischer et al. (2018) demonstrated how aerial images taken after prominent erosion events could be used to visually classify the severity of erosion features, and how this classification was appropriate to assess the capability of a spatially distributed model to relatively rank erosion-prone areas.

While the previously described sources of data for model testing are useful for evaluating simulations of on-site erosion, they offer little information about sediment transport to water courses and subsequent off-site erosion impacts. Therefore, they cannot be used to test the sediment delivery or routing components of distributed erosion models. Models such as WaTEM/SEDEM (Van Oost et al., 2000; Van Rompaey et al., 2001; Verstraeten et al., 2010), Morgan-Morgan-Finey (MMF) (Morgan, 2001; Morgan et al., 1984), and the Sediment Delivery Distributed model (SEDD) (Ferro and Minacapilli, 1995; Ferro and Porto, 2000) represent hillslope connectivity to the stream network either by routing sediment transport capacity along the flowpath or by estimating a topography-based sediment delivery ratio. These models are therefore not only able to simulate how much sediment is delivered to water courses, but also to identify where it comes from. To evaluate the quality of such simulations, quantitative data of sediment provenance is necessary.

A technique that provides quantitative apportionments of sediment provenance is sediment source fingerprinting. In this approach, physical and biogeochemical attributes of sink sediments are used to trace their origin from potential upstream sources (Klages and Hsieh, 1975; Yu and Oldfield, 1989; Walling and Woodward, 1995). Relative source contributions are then calculated by solving end-member un-mixing models based on source and sink sediment tracer concentrations (Collins et al., 1997; Cooper et al., 2014; Laceby and Olley, 2015). Such estimates are conceivably comparable to the outputs of distributed soil erosion models with a sediment routing/delivery component. However, a meaningful comparison requires fingerprinting source stratifications to be reasonably analogous to model outputs.

For sediment fingerprinting, potential upstream sources have been stratified in various manners, depending on the purpose, characteristics, and scale of the investigation. Common approaches include geological (Laceby et al., 2015; Olley and Caitcheon, 2000), land use (Pulley et al., 2016; Tiecher et al., 2016), and tributary-based stratifications (Habibi et al., 2019; Nosrati et al., 2018; Theuring et al., 2015). But what kind of source apportionments provide meaningful comparisons against distributed soil erosion models?

An interesting example is presented by Wilkinson et al. (2013), in which sediment fingerprinting was used to model the contributions of different erosion processes (i.e. surface and subsurface) to sediment loads in the Burdekin River basin, Australia (130,000 km^2). The resulting source apportionments were compared to SedNet model outputs (Wilkinson et al., 2009). Since SedNet calculates sediment budgets by differentiating inputs from different erosion processes (i.e. gullies, sheetwash), results provided a useful analogy. Likewise, Borrelli et al. (2018) were able to compare land use source apportionments from Alewell et al. (2016) to WaTEM/SEDEM model outputs in a 41 km^2 catchment on the Swiss Plateau.

However, a difficulty when testing erosion models in particular and environmental models in general arises from the epistemic uncertainties in model structures, parameter estimation, and the forcing/testing data (Beven, 2019). That is, uncertainty is a result of a lack of knowledge about I) the modeled phenomena: models are inherently flawed approximations of reality; II) model parameters: we cannot measure model parameters in every point in space and even if we could, parameters are often empirical abstract aggregations that require calibration; and III) the observational data: erosion is a highly variable phenomenon and our methods for measuring it are somewhat inadequate. Testing models as hypotheses therefore requires representing the uncertainties in both models and the things we call observational data or systems responses (Beven, 2018). It also requires a clear definition of model purpose and of the limits of acceptability of model error (Beven, 2009, 2006). These concepts provide the foundation of the Generalized Likelihood Uncertainty Estimation (GLUE) (Beven and Binley, 1992) methodology, in which Monte Carlo simulations are used to create a large number of possible model realizations by sampling uncertain model parameters. If the response surface does not produce acceptable realizations of the observational data, then the model itself can be rejected as not useful for prediction – at least under the testing conditions (Beven, 2009).

Although sediment fingerprinting models are now consistently applied in stochastic structures, usually relying on Monte Carlo simulations (Evrard et al., 2013; Pulley et al., 2016; Smith and Blake, 2014) or Bayesian inference (Blake et al., 2018; Cooper and Krueger, 2017), soil erosion models are more frequently used in a deterministic fashion. Moreover, outlet sediment loads, which are the common forcing/testing data with which models are evaluated, are also represented deterministically. Therefore, an uncertainty-based framework for incorporating sediment fingerprinting into soil erosion model testing is lacking.

In this study we present a novel approach to evaluate spatially distributed soil erosion/sediment delivery models that represents the uncertainties in both models and observational data. Since we understand that the purpose of spatially distributed sediment transport models is to not only provide acceptable simulations of outlet transport rates, but also to represent sediment dynamics within a catchment, we use sediment loads and sediment fingerprinting source apportionments as model evaluation data. By use of the GLUE methodology, we apply the SEDD model to a ~6600 km² river basin in southeast Brazil. Limits of acceptability of model error are defined according to the uncertainty in the outlet sediment load data. Behavioral model simulations are then evaluated against sediment fingerprinting source apportionments, which have been stratified based on a hierarchical tributary design that facilitates model comparisons along different stages of sediment transport. Our approach is implemented on free GIS software and programming languages, being fully reproducible and/or adaptable elsewhere. The outcomes of this research therefore provide a much needed open source framework for incorporating uncertainty analysis into distributed soil erosion models applications. Moreover, it demonstrates how sediment fingerprinting, and potentially other sources of data, can be assimilated into model testing within a stochastic structure.

2 METHODS

2.1 Catchment description

The Mortes River drains an area of approximately 6600 km² in the south of the State of Minas Gerais, Brazil (Figure 1). The river's headwaters are in Mantiqueira Mountain Range and it flows until it's confluence with the Grande River, at the Funil hydroelectric power plant reservoir. Elevation within the basin ranges from 1414 m to 807 m. According to Köppen's classification, the climate in the area is predominantly humid subtropical with dry winters and warm summers (Cwb) (Alvares et al., 2013). Average annual rainfall is approximately 1500 mm (Fick, 2017), which is almost entirely concentrated in the spring and summer months.

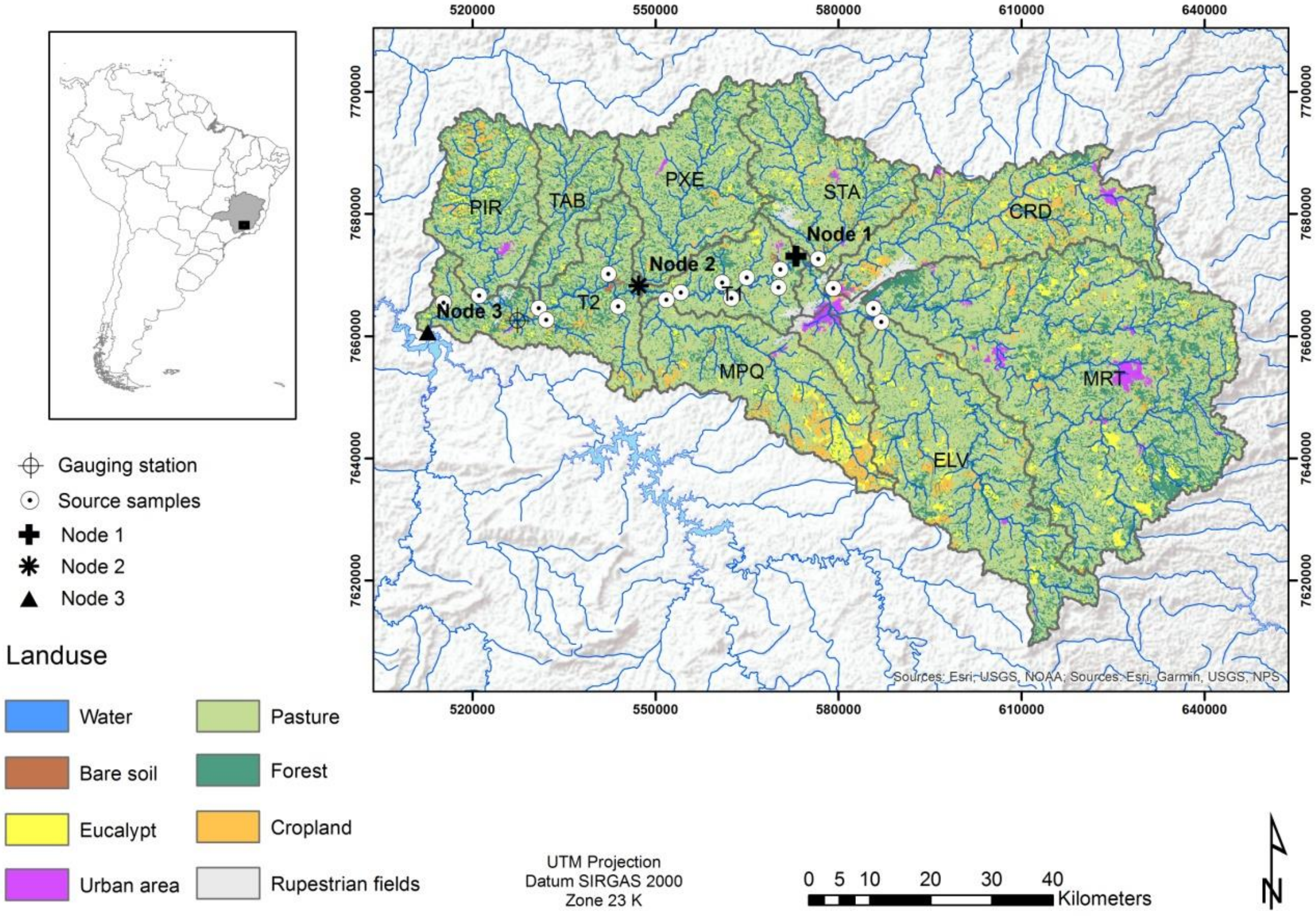


Figure 1. Location of the Mortes River basin and the land use map of the catchment. Sub-catchments and sampling locations for sediment fingerprinting are also displayed. Legend: MRB: Mortes River basin; CRD: Carandái River sub-catchment; ELV: Elvas River sub-catchment, MPQ: Mortes Pequeno River sub-catchment; MRT: Mortes River sub-catchment; PIR: Pirapetinga River sub-catchment; PXE: Peixe River sub-catchment; STA: Santo Antônio River sub-catchment; T1: mid catchment area; T2: lower catchment area; TAB: Tabões River sub-catchment.

Hapludoxes (48 %) and Dystrudepts (35 %) are the main soil classes in the basin (Table 1). The first are deep, highly weathered oxidic soils, while the latter are less developed, shallow, and erosion-prone. Most of the catchment is occupied by pastures (66 %), often degraded by over-grazing and/or a lack of management. Remaining forest strips (22 %) are mostly found on ridges and buffer strips along the stream network. Croplands, which are mostly composed of maize fields for silage production, occupy a small portion of the catchment area (5 %). Eucalypt forests (5 %) are commonly planted for charcoal manufacturing. Most of the agricultural areas, notably in the Carandaí, Mortes Pequeno, and Pirapetinga sub-catchments, are associated to the occurrence of Hapludoxes (Figure 1, Table 1). Dystrudepts support extensive pastures for raising dairy cattle and/or eucalypt plantations.

The Mortes River basin was chosen for this study due to the availability of continuous sediment concentration and discharge data from the Ibituruna gauging station (Figure 1). Although water discharge records are frequently made available by the Brazilian Water Agency, sediment concentration data are difficult to obtain. Moreover, field observations and bathymetric surveys have shown that the Mortes River delta is the main sedimentation zone in the Funil reservoir. Although the reservoir was filled in 2003, the high sedimentation rates in Mortes River already impede navigation near its delta.

Table 1. Physiographic attributes of the Mortes River basin and its sub-catchments.

	MRB	CRD	ELV	MPQ	MRT	PIR	PXE	STA	T1	T2	TAB
Soil class	Area (%)										
Dystrudepts	35.2	24.3	30.2	37.6	49.5	0.0	36.7	38.6	94.8	6.5	0.0
Acrudoxes	0.7	0.0	0.0	0.3	0.0	1.0	0.0	0.0	0.0	3.3	10.0
Hapludoxes	48.2	74.0	61.0	55.8	18.9	89.0	47.4	60.5	1.0	70.9	60.0
Rhodudults	4.6	0.0	0.0	0.0	0.0	10.0	15.1	0.0	0.5	18.3	30.0
Paleudults	10.2	0.0	8.8	5.7	30.9	0.0	0.0	0.0	0.0	0.0	0.0
Ustorthents	0.5	1.7	0.0	0.0	0.7	0.0	0.7	0.8	0.0	0.0	0.0
Rocky outcrops	0.6	0.0	0.0	0.5	0.0	0.0	0.0	0.0	3.8	1.0	0.0
Land use	Area (%)										
Bare soil	0.1	0.0	0.1	0.0	0.0	0.2	0.0	0.0	0.6	0.4	0.0
Cropland	4.6	11.6	3.9	11.1	2.7	5.0	1.4	3.6	0.7	4.8	1.3
Eucalypt	5.2	5.8	5.9	8.5	5.8	6.2	3.0	5.1	2.3	2.5	2.3
Forest	21.6	18.0	18.6	14.2	25.1	25.5	22.1	20.4	22.8	27.5	24.4
Pasture	66.2	60.0	71.2	65.3	64.3	61.9	73.1	68.5	70.5	62.6	71.3
Rupestrian fields*	1.0	2.7	0.0	0.5	0.2	0.6	0.0	2.0	1.6	1.1	0.5
Urban area	1.1	1.8	0.1	0.3	1.8	0.7	0.3	0.4	0.5	0.1	0.0
Water	0.2	0.2	0.1	0.1	0.1	0.0	0.0	0.0	1.0	0.8	0.1
Slope	(θ)										
Mean	9.1	8.3	8.9	7.8	9.9	8.4	9.1	9.9	9.7	8.3	9.7
Std. Dev.	5.2	4.8	5.0	4.4	5.7	4.5	5.0	5.3	5.5	4.9	5.4
Elevation	(m)										
Min.	807	890	892	869	892	826	865	890	865	807	864
Max.	1414	1407	1412	1191	1414	1209	1339	1312	1246	1239	1205
Mean	1035	1073	1061	996	1091	971	1035	1035	956	931	988
	(km^2)										
Area	6608	676	875	566	1817	424	510	509	281	526	259

Legend: MRB: Mortes River basin; CRD: Carandái River sub-catchment; ELV: Elvas River sub-catchment, MPQ: Mortes Pequeno River sub-catchment; MRT: Mortes River sub-catchment; PIR: Pirapetinga River sub-catchment; PXE: Peixe River sub-catchment; STA: Santo Antônio River sub-catchment; T1: mid catchment area; T2: lower catchment area; TAB: Tabões River sub-catchment.

* Grassland herbaceous/sub-shrubby formation that is commonly found on quartzitic ridges and other rocky outcrops.

2.2 Sediment load data

Suspended sediment concentration (mg L^{-1}) and water discharge (m3 s^{-1}) were monitored in the Ibituruna gauging station (Figure 1) from March 2008 to December 2012 (Batista et al., 2017). Measurements were taken on an approximately monthly basis, resulting in 44 observations. In order to estimate long-term sediment loads, we fitted a sediment rating curve relating suspended solid concentration to water discharge by ordinary least squares. Both variables were log-transformed, as the relationship between sediment concentration and discharge in the log-scale is approximately linear (Vigiak and Bende-michl, 2013). The goodness-of-fit of the linear model was visually assessed with residual and Quantile-Quantile plots. These and all other statistical analyses here presented were performed with the R programming language (R Core Team, 2019).

In order to propagate the error of the fitted model, 10^4 posterior simulations of the model coefficients were generated by an informal Bayesian inference function of the R package arm (Gelman and Hill, 2007). This function uses the model residual standard errors to create multivariate normal distributions of model coefficients, thus preserving their correlation when estimating posterior simulations. Next, daily sediment concentrations values were calculated based on continuous discharge records from the Brazilian Water Agency on the Ibituruna gauging station (1992-2013) and the simulations of model coefficients. Concentrations were back-transformed and used to estimate daily sediment loads (ton day^{-1}), which were subsequently aggregated into monthly, annual, and average annual transport rates. In summary, the 10^4 simulations of daily sediment concentrations were used to propagate the rating curve uncertainty when calculating long-term sediment loads. As pointed out by Vigiak and Bende-michl (2013), this approach only quantifies the regression uncertainty, and the actual errors associated to sediment load calculations might be underestimated. More detailed descriptions of Bayesian and bootstrapping methods for propagating the uncertainty of sediment rating curves can be found in Rustomji and Wilkinson (2008) and in Vigiak and Bende-michl (2013).

2.3 Sediment fingerprinting data

2.3.1 Sampling design and sample collection

In order to facilitate a comparison between SEDD model outputs and fingerprinting source apportionments, we established a tributary sampling design, in which sub-catchments of the Mortes River basin were treated as end-member sediment sources. In addition, we adopted a hierarchical approach for sink sediment sampling (Blake et al., 2018; Boudreault et al., 2019; Koiter et al., 2013). That is, considering the disconnectivity of the sediment cascade on large river basins due to the variability of residence times of sediment storage (Koiter et al., 2013), we understood it was important to sample sink sediments at different nodes of the main river channel (Figure 1). As a result of our sampling design, three nodes with four potential upstream sources each were stratified within the catchment.

Node 1, the most upstream sink sediment sampling location, has four main tributaries: the Mortes River (MRT) itself before its confluence with the Elvas River (ELV), the Carandaí River (CRD), and the Santo Antônio River (STA). Due to our hierarchical approach, Node 1 sediments become a potential source of the next downstream node. Hence, Node 2 sources are comprised by the Mortes Pequeno River (MPQ), the Peixe River (PXE), the set of small tributaries in the mid catchment (T1), and Node 1. Similarly, Node 3 on the catchment outlet receives sediments from the Pirapetinga River (PIR), the Tabões River (TAB), the set of small tributaries in the lower catchment (T2), and Node 2.

Sediment sampling was conducted in two different periods to represent transport dynamics during the well-defined seasons of the local climate. During September 2017 (dry season), all nodes and sources were sampled. In February 2018, during the rainy season, we retrieved extra samples from the sink sediment nodes.

Source samples were taken from lag-deposits of tributaries near their confluence with the main river channel. The uppermost layer (1-2 cm) of freshly deposited sediments from river margins was scrapped with a non-metallic trowel, and approximately 15 scrapes were combined into one individual sample. We collected a total of 20 composite samples per each tributary, except for sources T1 and T2. These sources are comprised by a set of small tributaries that drain directly to the Mortes River (Figure 1). Hence, 25 and 17 samples were retrieved in T1 and T2, respectively (4-5 samples from each small tributary).

Sampling sink sediments from Nodes 1 and 2 followed the same methods described above. During the dry season 20 samples were collected from each of these nodes, whereas during the rainy season, 12 and 20 samples were retrieved from Nodes 1 and 2, respectively. Given that the Mortes River flows into the Funil reservoir, samples from Node 3 were taken from the bottom of the shallow river delta, before its confluence with the Grande River. At the node 3 site, 26 and 12 composite samples were collected during the dry and rainy season, respectively.

2.3.2 Laboratory analyses

Sediment samples were oven dried at 60 ºC and dry-sieved with a 0.2 mm mesh. Subsequently, total concentration of the 21 following elements was determined by inductive coupled plasma optical emission spectrometry (ICP OES): Al, As, Ba, Ca, Cd, Ce, Co, Cr, Cu, Fe, K, La, Mg, Mn, Ni, Pb, Se, Ti, V, Zn, Zr.

2.3.3 Element selection

The first step of tracer selection is to investigate the composition of source and sink sediments. In this study, we started with an exploratory analysis by visually examining box-plots of element concentrations. Next, a range test was performed to verify if sink element concentrations were well bounded by the source mixing polygon. That is, if element contents on sink sediments are enriched or depleted in relation to source samples, then there is evidence that elements might not be behaving conservatively during sediment transport or there is a missing source (Laceby et al., 2017; Smith and Blake, 2014). Moreover, a mismatch of element concentrations on source and sink sediments may compromise the numerical solutions of the un-mixing models (Collins et al., 2013). The range test therefore aims not only to eliminate elements plotting outside the mixing polygon from further analyses, but also to provide an initial insight into the quality of the geochemical data.

Different approaches have been employed for analyzing conservative behavior and for performing range tests (Smith et al., 2018a; Wilkinson et al., 2015). Although earlier research might have focused on maximum and minimum tracer values, distribution-based un-mixing models (Bayesian or frequentitst) requires an examination of the distributions of tracer concentrations. Considering the structure of the bootstrapping approach we employed for solving our un-mixing model, we adopted a mean and standard deviation range test. That is, we assumed that means and standard deviations of log-transformed tracer concentrations on sink sediments should plot within the means and standard deviations of the source log-transformed tracer concentrations. This ensures that, during the Monte Carlo simulation, sampled sink element contents will always be within the source range. The means and standard deviation range test was performed locally for each node and sampling season.

Given the heterogeneity of land uses and geological/pedological backgrounds of the sub-catchments comprising sediment sources in the Mortes River basin, a process-based approach to element selection (Batista et al., 2019; Koiter et al., 2013; Laceby et al., 2015) was not appropriate to this research. Hence, we adopted a more common statistical procedure, in which elements passing the range test were submitted to a step-wise forward Linear Discriminant Analysis (LDA) (niveau = 0.1). This approach aims to define a minimum set of tracers that maximize source discrimination, and elements selected by the LDA were used for modelling. Again, the procedure was repeated for all nodes and sampling seasons.

2.3.5 Un-mixing modelling

Relative sediment source contributions were calculated by minimizing the sum of squared residuals (SRR) of the un-mixing model:

$$SSR = \sum_{i=1}^{n}[(C_i - \sum_{s=1}^{m} P_s S_{si})/C_i]^2 \tag{1}$$

where n is the number of elements used for modeling, C_i is the concentration of element i in the target sediment, m is the number of sources, P_s is the optimized relative contribution of source s, and S_{si} is the concentration of element i in source s. Optimization constraints were set to ensure that source contributions P_s were non-negative and that their sum equaled 1.

In order to quantify the uncertainty in the un-mixing model source apportionments, we employed the bootstrapping methods described in Batista et al. (2019). The model was solved by a Monte Carlo simulation with 2500 iterations. For each iteration, log-transformed element concentrations were sampled from multivariate-normal distributions generated from the source and sink geochemical data. During the Monte Carlo simulation, values were back-transformed by an exponential function. Log-transformation was applied to avoid sampling negative element concentrations and to force a near-normal distribution on the typically skewed sediment geochemistry data. The optimization function was scripted with the R package Rsolnp

(Ghalanos and Theussl, 2015), whereas the Monte Carlo simulation (here and elsewhere in this study) with the package foreach (Calway et al., 2017).

2.4 SEDD model description

The SEDD model calculates a spatially distributed sediment delivery ratio *SDRi* that expresses the proportion of eroded sediments that are delivered to the stream network (Ferro and Minacapilli, 1995; Ferro and Porto, 2000). The model does not represent channel erosion or deposition processes, and sediments reaching the stream network are assumed to reach the catchment outlet. Following a grid based structure, the *SDRi* is calculated as:

$$SDR_i = \exp(-\beta \frac{l_i}{s_i}) \quad (2)$$

where: SDR_i is the soil delivery ratio of a grid cell i; β is a catchment specific empirical parameter (m^{-1}), l_i is the flow length from cell i to the nearest stream channel (m) along the flow path, and s_i is the slope of cell i ($m\ m^{-1}$).

Typically, the empirical parameter β is calibrated to minimize the errors of sediment load predictions (Fernandez et al., 2003; Fu et al., 2006; Lin et al., 2016), whereas the flow length and slope parameters can be derived from DEM processing.

The *SDRi* grid is used to calculate area specific sediment yields (*SSYi*) ($ton\ ha^{-1}\ yr^{-1}$), which quantifies the amount of sediments that are delivered from cell i to the stream network:

$$SSY_i = SDR_i A_i \quad (3)$$

where: SSY_i is the specific sediment yield for a grid cell i; SDR_i is the soil delivery ratio for a grid cell i and A_i is the annual soil loss computed by Revised Universal Soil Loss Equation (RUSLE) for a grid cell i.

RUSLE estimates average annual erosion rates by the following empirical equation (Renard et al., 1997):

$$A = R * K * LS * C * P \tag{4}$$

where: A is soil loss per unit area (t ha^{-1} yr^{-1}); R is the rainfall and runoff erosivity factor (MJ mm ha^{-1} h^{-1} yr^{-1}); K is soil erodibility factor (t ha h ha^{-1} MJ^{-1} mm^{-1}); LS is the topographic factor, representing slope length and steepness (dimensionless); C is cover management factor (dimensionless), and P is support practice factor (dimensionless).

As the SEDD model neglects channel deposition, the total sediment yield at the catchment outlet can be calculated as the sum of $SSYi$ values. Equivalently, the mean of $SSYi$ values correspond to the area specific sediment yield in the catchment, and the same calculations can be employed at sub-catchment scale. With this approach, sub-catchment relative contributions can estimated based on SEDD model results.

2.6 GLUE

The basic structure of the GLUE methodology is summarized by Beven (2009) in five decision steps:

1. Decide on a likelihood measure to evaluate model realizations.
2. Decide on the rejection criteria for non-behavioral model realizations.
3. Decide which parameters are uncertain.
4. Decide on a prior distribution to characterize the uncertainty of the chosen parameters.
5. Decide on a simulation method for generating model realizations.

Here we did not establish a formal likelihood measure to evaluate model realizations, as the rejection criteria for non-behavioral simulation was set according to an actual range of system responses. That is, all model realizations which produced sediment load responses within the 95% prediction interval of the sediment load rating curve were considered behavioral. Since the SEDD temporal scale is inherited from the RUSLE, the model simulates long-term average annual sediment yields. Therefore, for comparison purposes the sediment rating curve estimates were aggregated into a 22 years average.

Model realizations were generated by a Monte Carlo simulation with 1000 iterations. SEDD parameter β was sampled from a log-uniform distribution, with minimum and maximum parameters retrieved from typical values reported in the literature (min = 0.000001 m^{-1}, max = 0.1 m^{-1}) (e.g. Porto and Walling, 2015; Taguas et al., 2011). We used a log-uniform distribution to ensure that the extreme values of this broad range were sampled during the simulation. The threshold for stream definition, which affects drainage density and therefore distance to streams (l_i), was sampled from a uniform distribution (min = 50000 m^2, max = 5000000 m^2). To represent the uncertainty in the DEM derived model variables, we created a pseudo-random error surface for each model iteration. Mean and standard deviation of DEM errors were retrieved the NASA SRTM report (Rodriguez et al., 2006)(μ = 1.7 m, σ = 4.1 m) and used to create a normally distributed error field, which was added to the original DEM. All terrain attributes used in the models were then calculated within the Monte Carlo simulation. All herein described spatial analyses were supported by SAGA GIS (Conrad et al., 2015) and the R package RSAGA (Brenning and Bangs, 2015).

Since we understood that RUSLE factors were not parameters requiring calibration or conditioning, but instead uncertain model variables, we performed a forward uncertainty analysis, similarly to Biesemans et al. (2000) and Van Rompaey and Govers (2002). Although this can be seen as a separate analysis, RUSLE error was propagated into SEDD simulations, as we explain in the following.

The forward error analysis was performed with a Monte Carlo simulation with 1000 iterations. In order to represent the uncertainty in the RUSLE R factor, we first calculated a deterministic rainfall erosivity map. This was carried out with average monthly and annual rainfall grids from WorldClim (Fick, 2017) and the regression equation developed by Aquino et al. (2014). This regression equation estimates annual (or average annual, in this case) EI_{30} index values, and it was originally fitted using detailed rainfall data from the Municipality of Lavras. For each iteration of the Monte Carlo simulation, we added a normally distributed error surface to the deterministic rainfall erosivity map, with mean equal zero and a standard deviation equal to 10 % of mean deterministic R factor for the catchment.

For the K factor, we created truncated normal distributions for each soil class occurring in the catchment soil map (FEAM, 2010). The discrete soil map was rasterized and, for each simulation, a grid cell erodibility value was sampled according to its corresponding soil class. Distribution parameters were set according to published K factor values for Brazilian soils. Although in general there were not enough different estimations of K factor values for individual soil classes to create data-based probability distributions, we used the available published data and our own interpretation to infer distribution parameters (Table 2).

Table 2. Parameters of the truncated normal distribution of each soil class in the Mortes River basin.

Soil class	Mean	Standard dev.	Minimum	Maximum
Dystrudepts	0.035	0.01	0.01	1
Acrudoxes	0.012	0.01	0.001	1
Hapludoxes	0.015	0.01	0.001	1
Rhodudults	0.017	0.01	0.005	1
Paleudults	0.02	0.01	0.005	1
Ustorthents	0.05	0.01	0.03	1
Rocky outcrops	0.00001	0.00001	0.00001	1

Uncertainty in the LS factor was represented following the DEM error propagation described above. Slope (rad) and catchment area (m^2) grids were created for each model iteration. These grids were subsequently used to calculate the LS factor with the equation of Desmet and Govers (1996). A maximum threshold of 10800 m² was enforced to the catchment area grid, which corresponds to maximum flow length of 360 m for a 30m resolution DEM. This was performed to avoid spuriously high LS factor values in flow concentration areas, as usually carried out in RUSLE applications (Panagos et al., 2015; Schmidt et al., 2019).

Similarly to the K factor, errors in the C factor estimation were propagated by creating truncated normal distributions for individual land use classes (Table 3). The land use grid was produced using 30 m resolution Landsat 8 Surface Reflectance images from 2013 and the methods described in Batista et al. (2017). Since no widespread support management practices are found in the catchment agricultural areas, no specific procedure was applied to represent P factor uncertainty. However, the C factor distribution parameters for cropland and eucalypt were set to reflect occasional contour cropping and/or crop residue management.

Table 3. Parameters of the truncated normal distribution of each land use class in the Mortes River basin.

Land use	Mean	Standard dev.	Minimum	Maximum
Bare	0.8	0.2	0.6	1
Cropland	0.088	0.045	0.02	1
Eucalypt	0.015	0.03	0.0005	1
Forest	0.001	0.003	0.0001	1
Pasture	0.01	0.02	0.001	1
Rupestrian vegetation	0.001	0.005	0.0001	1

The resulting RUSLE model realizations were used as input for the SEDD model simulations. Moreover, we performed a sensitivity analysis by fixing each model factor and sampling the remaining variables in new Monte Carlos simulations, each with 1000 iterations. This enabled us to evaluate the proportion of model variance explained by each factor.

It should be highlighted that forward error propagation is essentially subjective, given its total dependence on the assumptions made by the modeler about potential sources of uncertainty (Beven, 2009). Our approach presents a rather conservative estimate of model uncertainty, basically representing the errors involved in parameter estimation. This is because we could not describe all the sources of error in the model structure. Moreover, we wanted to constrain model realizations based on choices of factor values that modelers are expected to make. That is, we did not want to give the models full freedom: if all parameters and variables are allowed to vary beyond a range of physical meaning, models are capable of reproducing almost any answer – usually for the wrong reasons (see Batista et al., 2019a).

2.6.1 Spatial representation of model uncertainty

In order to represent the spatial uncertainty the final SEDD model predictions, we first filtered the behavioral model simulations according to the criterion previously described. Next, we calculated the 2.5 %, 50 %, and 97.5 % quantiles for each grid cell SSY_i estimates. Absolute error grids were then calculated by subtracting the 97.5 % grid by the 2.5 % grid. Relative errors were determined as:

$$RE_i\,(\%) = \left(\frac{AE_i}{M_i}\right) * 100 \tag{5}$$

where: AE_i is the absolute error for a grid cell i, M_i is the simulation median for grid cell i.

The filtered behavioral model realizations were also used to calculate total sediment yields from the sub-catchments described in Table 1. These calculations were used to estimate the relative contribution of the sub-catchments to aggregated sediment yields at each sink sediment sampling location (i.e. Nodes 1, 2, and 3). The SEDD-estimated relative contributions were then evaluated against fingerprinting source apportionments. The same approach was employed for creating RUSLE error maps, except in this case all model simulations were considered when calculating grid-cell quantiles.

3 RESULTS

3.1 Discharge curve

The error propagation method used to represent the uncertainty in the sediment rating curve resulted in a broad estimate of average annual specific sediment yields, with a 95 % prediction interval of 0.47 – 11.95 ton ha^{-1} yr^{-1}(mean = 3.45 ton ha^{-1} yr^{-1}; median = 2.52 ton ha^{-1} yr^{-1}) (Figure 2 a). As expected, annual estimates of sediment loads were more uncertain for the years with greater discharge and sediment transport (Figure 2 c). Monthly calculations revealed that over 85 % of the annual sediment load is transported from November to March. The monthly relative contributions to annual sediment yield showed less uncertainty than annual and average annual estimates (Figure 2 b).

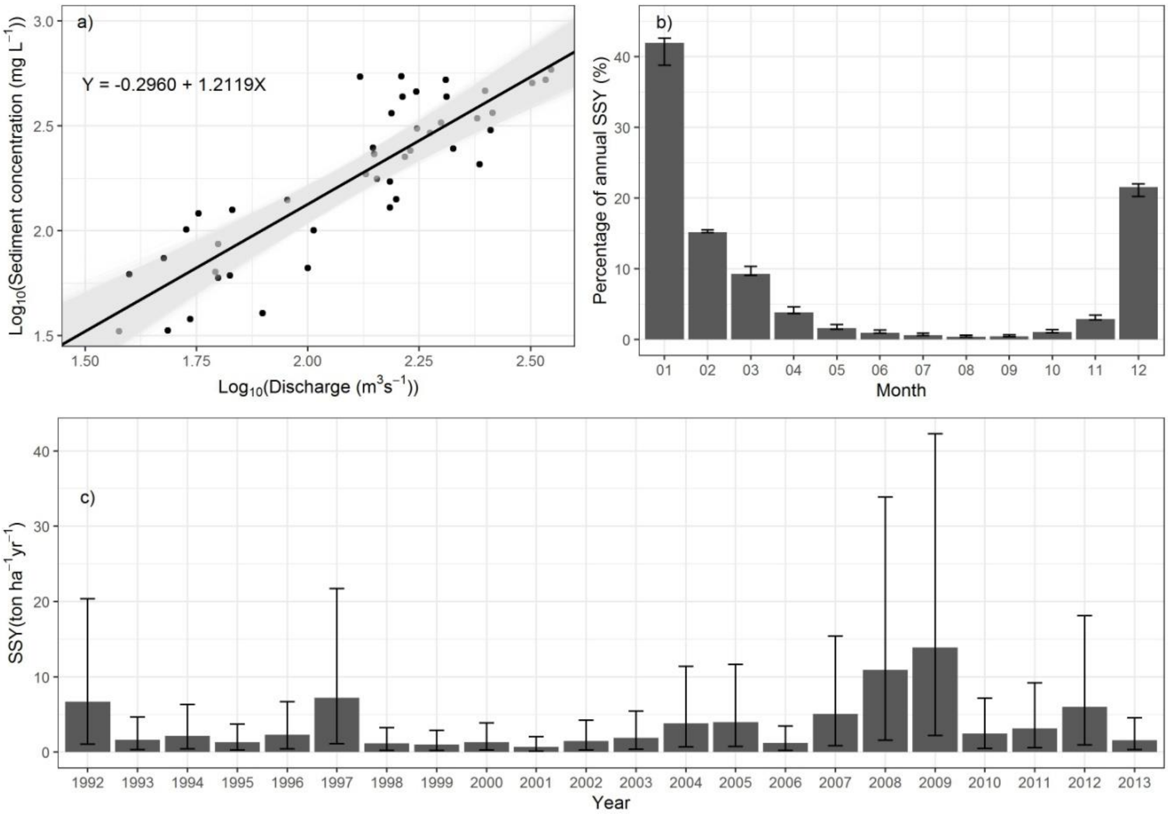


Figure 2. a) Sediment rating curve: dark line represents a deterministic mode fit and faded gray lines represent the 1000 simulations used to propagate the regression uncertainty; b) Monthly percentage of area specific sediment yields (SSY); c) Annual estimates of area specific sediment yields (SSY). Error bars represent 95 % prediction intervals.

3.2 Sediment fingerprinting

3.2.1 Element selection

Our exploratory analysis demonstrated that Cu and Zn displayed spurious within-source variability, as most sample concentrations were below detection limit. These elements were therefore omitted from further scrutiny.

Of the remaining 19 elements, 16 (84 %), 17 (87 %), and 9 (47 %) plotted within the source mixing polygons for Nodes 1, 2, and 3, respectively, for the dry season (Table 4). For the rainy season, 18 (95 %) elements passed the range test for Nodes 1 and 2, whereas only seven (37 %) elements were within source range for Node 3 sediments.

Table 4. Selected elements by the range test and the forward LDA for each node and season, along with the LDA reclassification accuracy.

Node	Season	Selection step	Selected elements	% of correctly classified samples
1	Dry	Range test	Al, As, Ba, Cd, Ce, Co, Cr, Fe, K, La, Mg, Mn, Pb, Se, Ti, V	
		LDA	Al, Ba, Cd, Ce, Co, Cr, Fe, K, Mg, Mn, Ti, V	100
	Rainy	Range test	Al, As, Ba, Ca, Ce, Co, Cr, Fe, K, Ka, Mg, Mn, Ni, Pb, Se, Ti, V, Zr	
		LDA	Al, Ba, Ca, Co, Cr, K, Mg, Ni, Se, V, Zr	100
2	Dry	Range test	Al, As, Ba, Ca, Cd, Ce, Co, Cr, Fe, K, La, Mg, Ni, Se, Ti, V, Zr	
		LDA	As, Ca, Ce, K, La, Mg, Ni, Se, Ti, Zr	100
	Rainy	Range test	Al, As, Ba, Ca, Cd, Ce, Co, Cr, Fe, K, La, Mn, Ni, Pb, Se, Ti, V, Zr	
		LDA	As, Ca, Ce, Co, K, La, Mn, Pb, Se, Ti, Zr	100
3	Dry	Range test	Ba, Ca, Cd, Ce, Cr, Fe, K, La, Ni	
		LDA	Ca, Cd, Ce, Cr, K, La	91
	Rainy	Range test	Ba, Ce, Co, Fe, K, La, Mn	
		LDA	Ba, Ce, Co, Fe, K, La	82

For Node 1, the forward step-wise LDA selected 12 elements for the dry season sediments, whereas 11 elements were selected for the rainy season (Table 4). The LDA for both seasons showed a reclassification accuracy of 100 %. For Node 2, the discriminant analysis selected 10 elements for the dry season and 11 for the rainy season. Again, all samples were correctly reclassified during the LDA cross-validation. As fewer elements passed the range test for Node 3, only six elements were selected by the LDA for both seasons. Reclassification accuracy was lower in this case, with 91 % and 82 % for the dry and rainy seasons, respectively. The largest errors associated to the LDA reclassification for Node 3 source samples can be visualized in the bi-plots displayed in Figure 3.

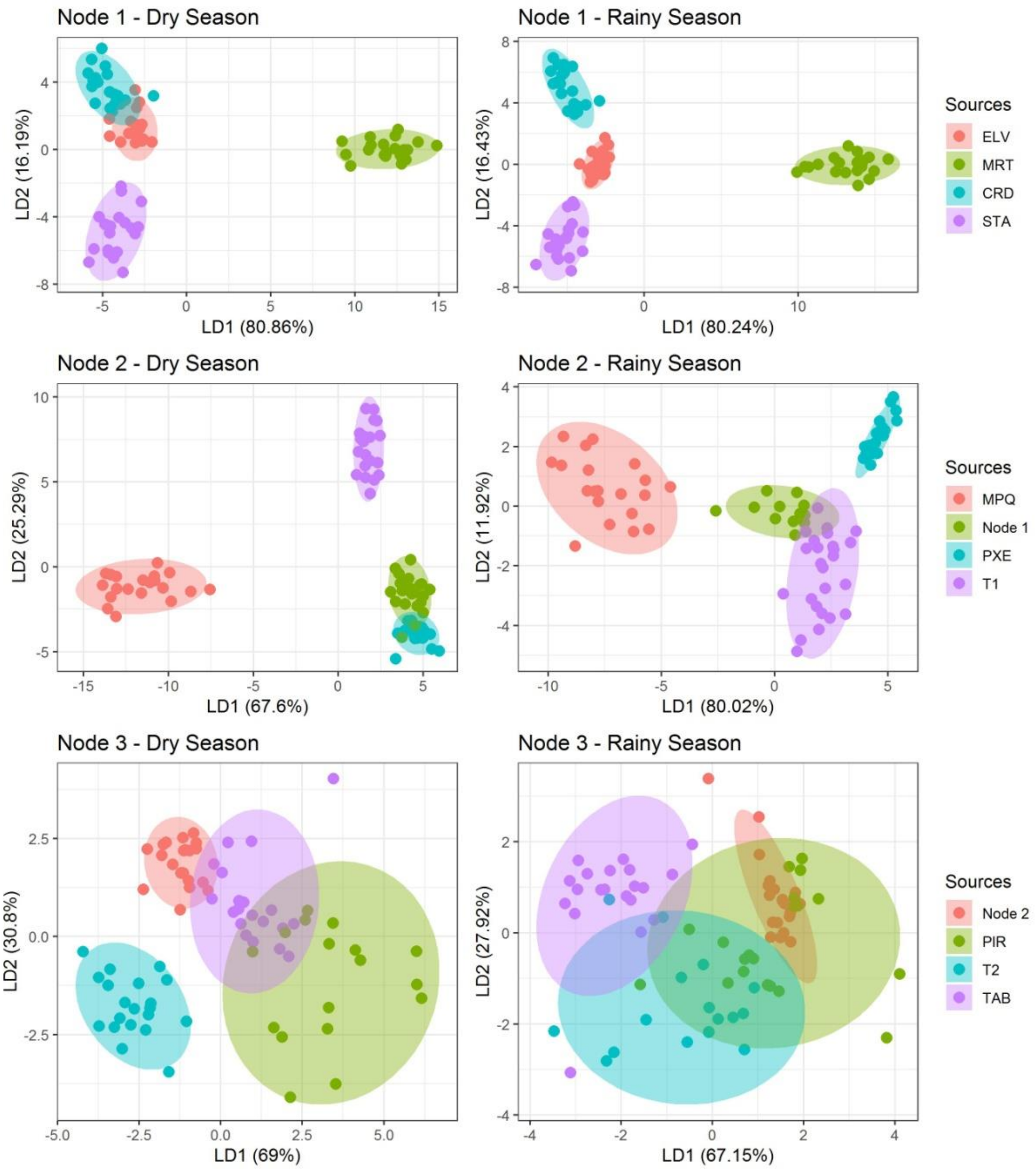


Figure 3. LDA bi-plots of source reclassification based on selected element concentrations. Ellipses represent 90 % confidence intervals.

3.2.2 Un-mixing model results

Un-mixing model solutions for Node 1 were highly uncertain for both seasons, as demonstrated by the broad density curves displayed in Figure 4. According to model estimates, sources CRD and ELV seem to dominate sediment contributions in relation to MRT and STA – at least considering the median and interquartile (IQR) values of the simulated source apportionments (Table 5).

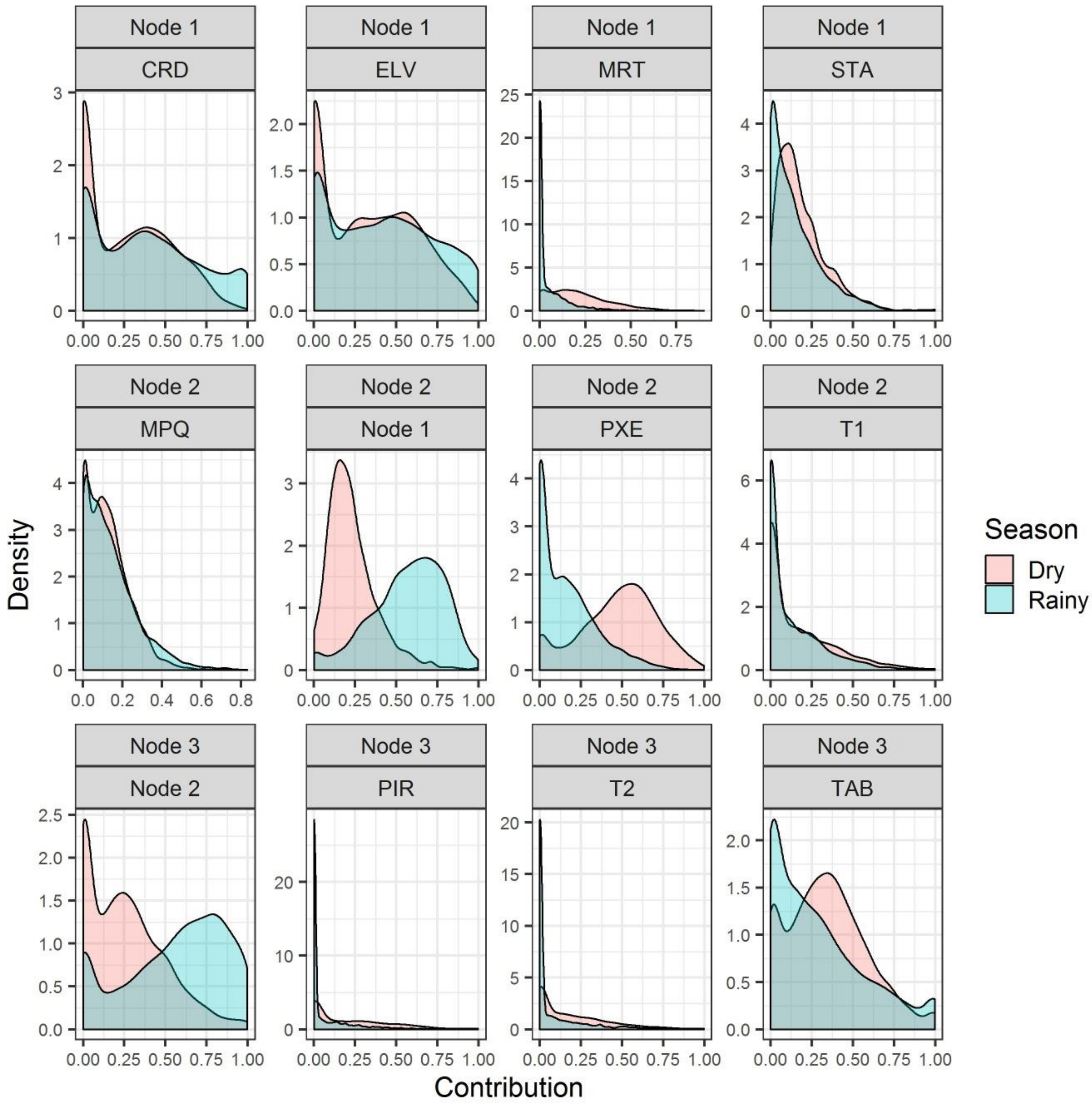


Figure 4. Probability density functions of estimated relative source contributions.

Table 5. Results of the un-mixing models source apportionments based on the Monte Carlo simulations.

Node	Source	Season	2.5 % quantile	25 % quantile	50 % quantile	75 % quantile	97. 5 % quantile
1	CRD	Dry	0.00	0.00	0.25	0.47	0.78
		Rainy	0.00	0.08	0.36	0.62	1.00
	ELV	Dry	0.00	0.00	0.31	0.56	0.88
		Rainy	0.00	0.12	0.41	0.66	1.00
	MRT	Dry	0.00	0.08	0.18	0.31	0.61
		Rainy	0.00	0.00	0.00	0.07	0.33
	STA	Dry	0.00	0.08	0.15	0.26	0.53
		Rainy	0.00	0.02	0.11	0.22	0.55
2	MPQ	Dry	0.00	0.04	0.11	0.18	0.38
		Rainy	0.00	0.04	0.11	0.20	0.47
	Node 1	Dry	0.02	0.13	0.20	0.30	0.62
		Rainy	0.04	0.44	0.60	0.74	0.92
	PXE	Dry	0.00	0.32	0.50	0.64	0.89
		Rainy	0.00	0.00	0.12	0.26	0.60
	T1	Dry	0.00	0.00	0.08	0.28	0.72
		Rainy	0.00	0.00	0.04	0.21	0.58
3	Node 2	Dry	0.00	0.05	0.23	0.41	0.81
		Rainy	0.00	0.34	0.61	0.80	1.00
	PIR	Dry	0.00	0.00	0.13	0.36	0.75
		Rainy	0.00	0.00	0.00	0.07	0.51
	T2	Dry	0.00	0.00	0.12	0.31	0.73
		Rainy	0.00	0.00	0.00	0.09	0.60
	TAB	Dry	0.00	0.17	0.33	0.50	0.88
		Rainy	0.00	0.04	0.22	0.45	1.00

Results for Node 2 were less uncertain and revealed a greater contrast between seasonal sediment transport dynamics. During the dry season, model estimates indicate that a significant part of sediments reaching Node 2 are derived from PXE (median = 50 %, IQR = 32 – 64 %). However, such contributions decrease during the rainy season, for which the models suggest a large apportion of sediments from Node 1 (median = 60 %, IQR = 44 – 74 %). Modeled source contributions from MPQ and T1 were relatively low for both seasons (Table 5).

Model solutions for Node 3 displayed a similar pattern to Node 2 regarding the seasonal variation of source contributions. During the dry season, a greater proportion of sediments were estimated to derive from the sources proximally located to the catchment outlet, particularly TAB (median = 33 %, IQR = 15 – 50 %). However, rainy season source apportionments indicate that most of the sediments reaching the Funil reservoir are originated on the upstream areas of the catchment, which are represented by Node 2 (median = 61 %, IQR = 34 – 80 %). This illustrates how even in the relative short time-period represented by our study, sediments from the upper and mid catchment area are transported throughout the river network. Given that most of the Mortes River sediment load is transported during the rainy season, it is plausible to assume that upstream sediments are important contributors to reservoir sedimentation.

3.3 RUSLE uncertainty

The results of the forward error analysis revealed that RUSLE estimates were highly uncertain in spite of the moderately conservative assumptions made about sources of model error. The median of grid cell absolute errors was of 29.0 ton ha^{-1} yr^{-1}, which translated to a median relative error of 588 %. The highest absolute errors in the RUSLE estimates were associated to the areas with higher erosion rate predictions, as expected (Figure 5 b, c). Contrarily, relative errors were higher on the areas with lower soil loss estimates. This is possibly a result of small variations on sampled parameter values leading to a large relative fluctuation on the low erosion predictions (Figure 5 a). Considering the median of the simulations as a point-based estimate of erosion rates, the influence of soil erodibility on soil loss predictions was evident on Figure 5 c. Upper and mid catchment areas, where Dystrudepts are widespread, had overall higher erosion rates, according to the model simulations. Moreover, modeled erosion hot-spots are visibly associated to areas with high flow accumulation and more intensive land uses (e.g. cropland, eucalypt).

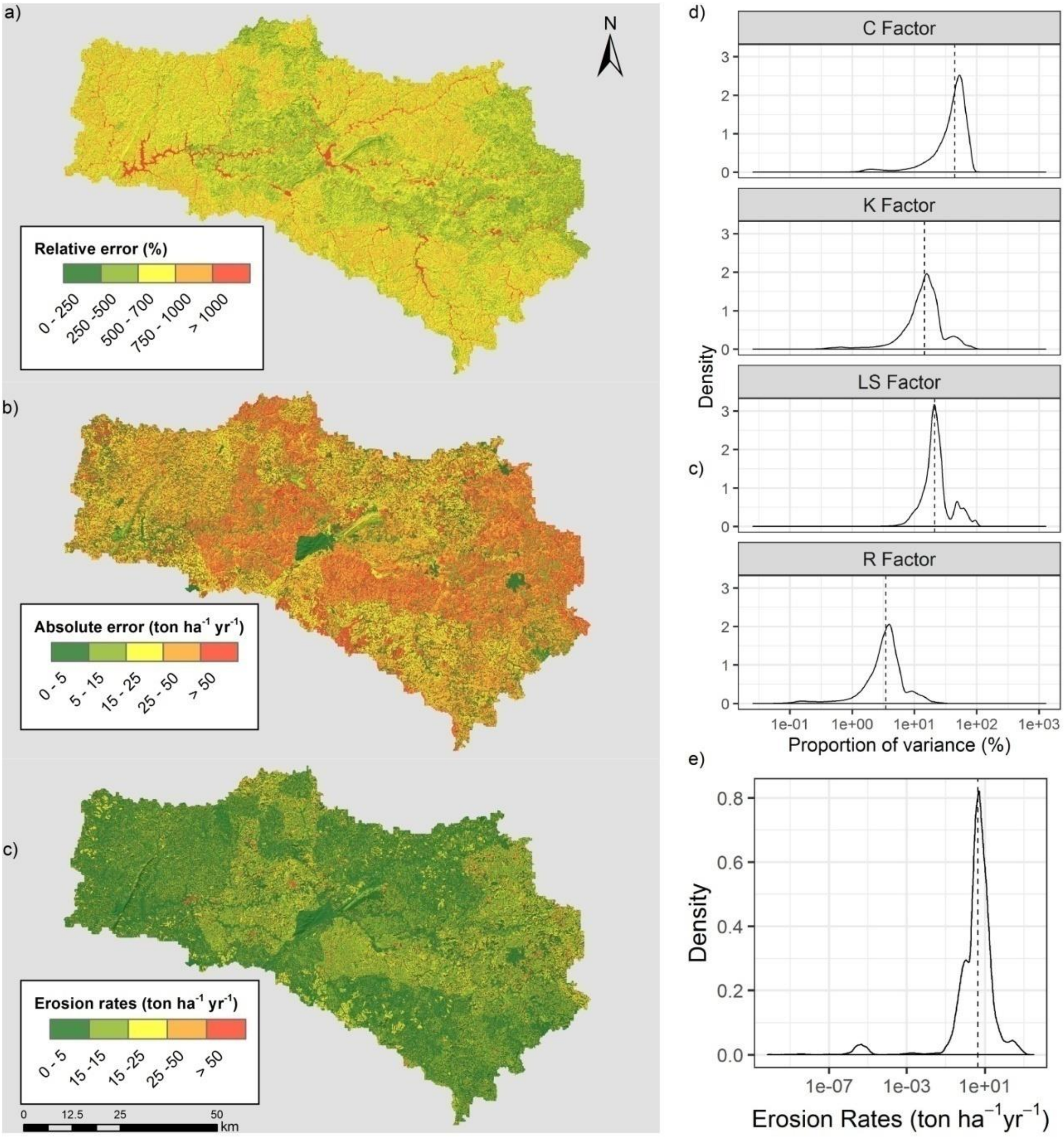


Figure 5. a) RUSLE relative error map; b) RUSLE absolute error map; c) RUSLE prediction map, based on the median simulation values; d) density curves of the proportion of variance of model error explained by individual RUSLE factors; e) density curve of grid cell erosion rate predictions, based on the median simulation values. Dashed vertical lines represent median values.

The sensitivity analysis demonstrated that the C factor was the largest source uncertainty in the model predictions. The proportion of model variance explained by the C factor had a median value of 45 %. (IQR = 30 – 56 %). The LS (median = 21 %; IQR = 17 – 27 %) and K factors (median = 15 %; IQR = 10 – 20 %) also contributed significantly to the propagated model errors. The R factor had a small influence on overall model uncertainty (median = 3 %, IQR = 2 – 5 %).

3.4 SEDD results

From the 1000 SEDD model realizations generated by the Monte Carlo simulation, 234 were behavioral considering the established limits of acceptability. That is, 234 model realizations provided estimates of outlet-based SSY within the 95 % prediction interval of the sediment rating curve (0.47 – 11.95 ton ha^{-1} yr^{-1}). Most of the non-behavioral model response surface was associated to an overestimation of the curve-calculated sediment yields (Figure 6 a).

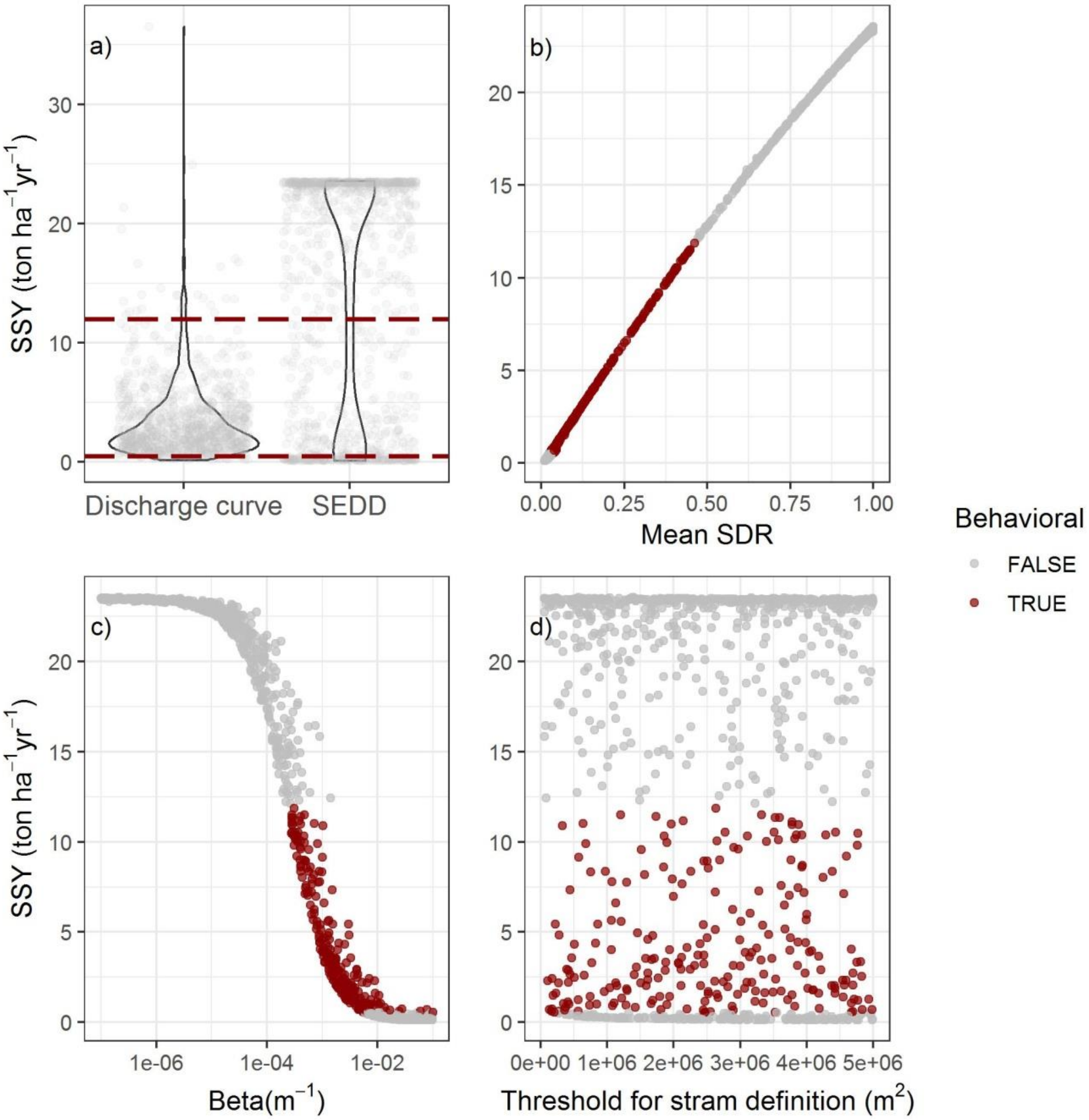


Figure 6. a) violin plots of catchment-lumped SSY values estimated by the discharge curve and the SEDD model; b) scatter plot of simulated mean grid cell SDR_i and resulting catchment-

lumped SSY values; c) dotty plot of sampled β values; d) dotty plot of sampled threshold for stream definition values.

By analyzing the dotty plots of sampled parameter values, it was clear that the empirical parameter β had a preponderant influence on the model results (Figure 6 c). Behavioral model realizations are concentrated within a relatively narrow range of β values, whereas acceptable system representations are spread throughout the sampled values of stream definition thresholds. The fluctuation of mean catchment SDR_i values in the catchment led to a linear increment of estimated SSY (Figure 6 b), indicating a little influence of RUSLE simulation results in the outlet-aggregated SEDD model predictions. Behavioral model realizations had mean SDR_i values between 5-50 %, which illustrates the uncertainty in the model predictions. Considering the median of the behavioral model realizations, grid-cell SSY_i estimates had a median value of 0.06 ton ha^{-1} yr^{-1}, whereas the median of analogous absolute error values was of 6.64 ton ha^{-1} yr^{-1}. Although outlet-lumped model results seem to be little influenced by the uncertainty in the RUSLE or in the stream definition threshold, the errors derived from such input variables/parameters are explicit when the uncertainty of spatially distributed SSY estimates are presented in Figure 7 a. Areas with large absolute errors in the SSY map clearly match the RUSLE errors displayed in Figure 6 b. Moreover, the influence of stream definition threshold uncertainty is visible in the surroundings of lower order streams.

.

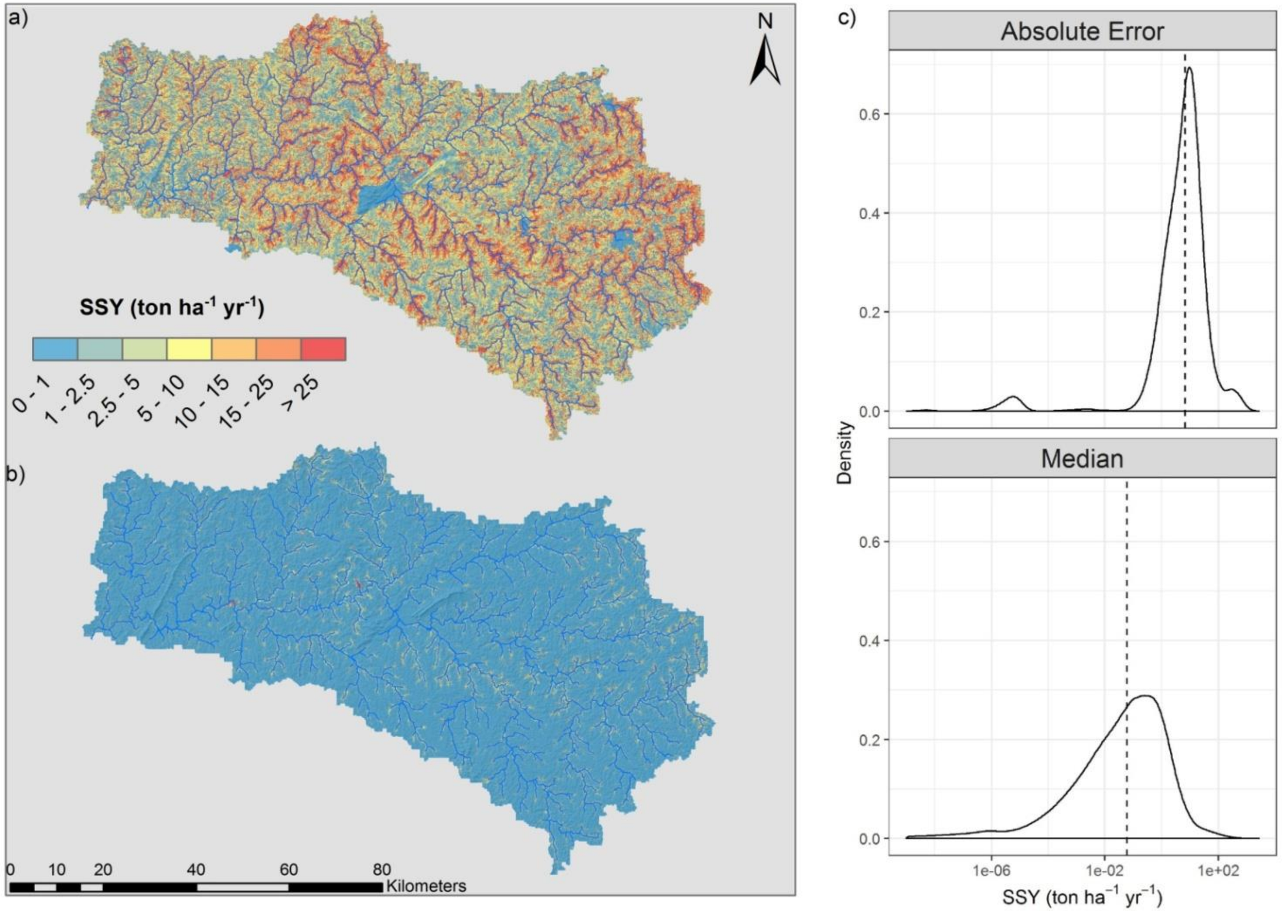


Figure 7. a) Absolute error of behavioral SEDD simulations of SSY b) Median of behavioral SEDD simulations of SSY; c) density curves of grid-cell values of absolute model error and median SSY simulations. Dashed lines represent the median of the distributions.

3.4.1 Evaluation of SEDD results against fingerprinting source apportionments

Distributions of relative source contributions estimated by the SEDD model overall displayed a similar pattern to the rainy season fingerprinting source apportionments, except for Node 1 (Figure 8). Opposite to the fingerprinting results, SEDD simulations indicated that MRT was the main source of sediments (IQR = 52.9 – 53.4 %) reaching the main river channel.

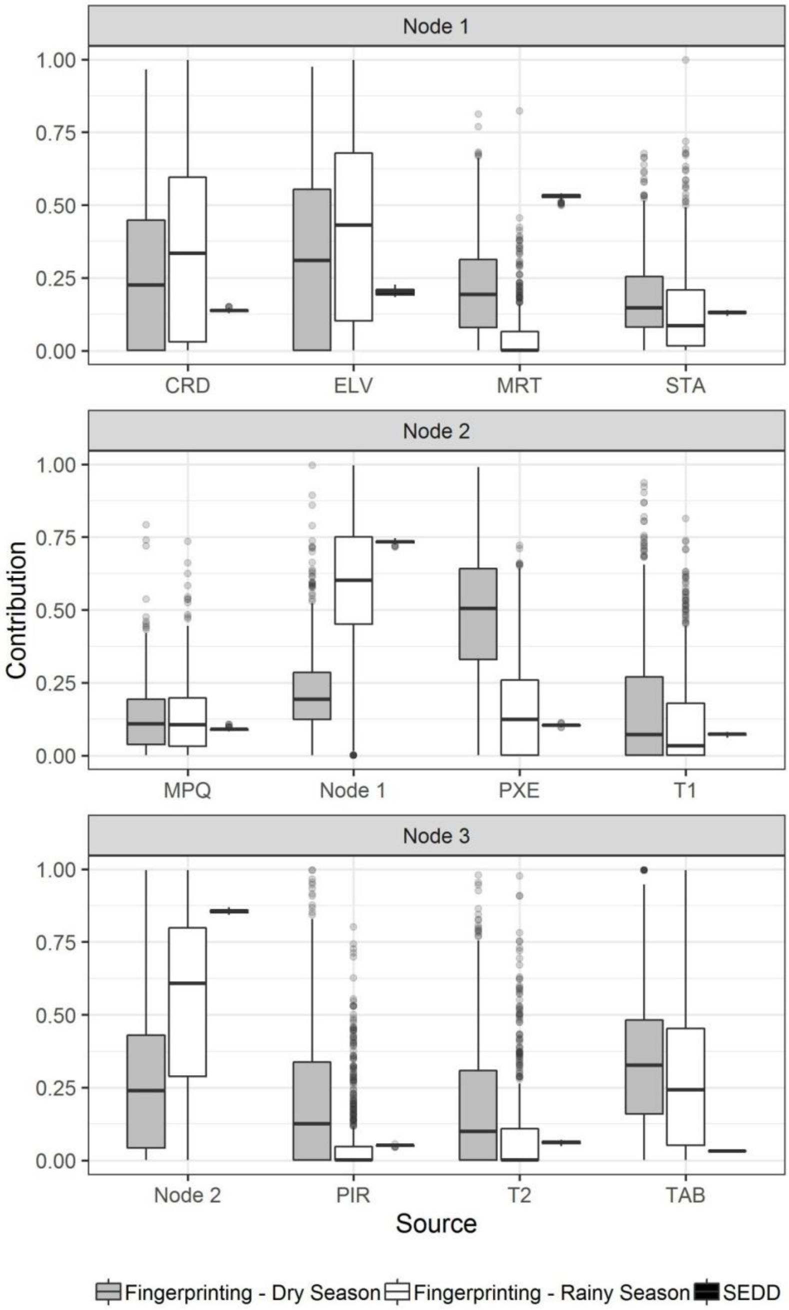


Figure 8. Relative source contributions estimated by SEDD and fingerprinting un-mixing models.

Node 2 results revealed a large agreement between rainy season fingerprinting and SEDD-estimated relative source contributions, as all SEDD model realizations were bound by the IQR of the fingerprinting apportionments. However, SEDD simulations indicate an even larger contribution of Node 1 sediments (IQR = 72.6 – 73.0 %). Similarly, fingerprinting results for the rainy season for Node 3 showed a similar pattern to the SEDD simulations. Both models indicate that Node 2 sediments are the largest contributors to outlet sediment loads, although SEDD results again suggest a greater contribution of upstream sediments (Node 2 IQR = 84.3 – 85.7 %). Moreover, SEDD-estimated TAB contributions (IQR = 3.2 – 3.4 %) were considerably lower than the ones estimated by the sediment fingerprinting un-mixing models.

4 DISCUSSION

The model testing framework presented here has demonstrated how uncertainty permeates all facets of soil erosion models and the things we call observational data. The error propagation method used to represent the uncertainty in the sediment rating curve resulted in such broad estimates of average annual sediment loads that many different SEDD realizations were able to encompass the forcing data. Similar results have been reported by other soil erosion modelers (Banis et al., 2004; Janes et al., 2018), and a logical conclusion is that we need better data in order to reject non-behavioral model realizations.

However, even if more accurate and precise sediment load data were available, our approach has demonstrated how very different spatial model representations can produce similar outlet responses. Despite the fact that we only considered behavioral simulations while calculating the uncertainty of grid cell SSY_i estimates, absolute model errors were almost hundred-fold the median of the predictions. This brings to question if the numerical spatial predictions are at all useful. Furthermore, it demonstrates how misleading it can be to neglect model and observational data uncertainty in soil erosion and sediment transport models.

The fact that the SEDD results were mostly driven by the empirical and somewhat abstract parameter β raises some concerns about the quality of process representation in the model. The common deterministic parameter optimization method for calibrating β should be therefore disputed. If SEDD model simulations are to provide meaningful system representations, alternative methods for deriving β values should be encouraged (e.g. Ferro and Stefano, 2003; Porto and Walling, 2015). Nonetheless, given the sensitivity of the model to parameter β, representing the uncertainty associated to the parameter estimation is paramount.

In spite of the large errors associated to grid cell SSYi estimates, aggregated sub-catchment relative contributions calculated from the SEDD simulations were precise, as shown by the narrow uncertainty bands in Figure 8. This indicates that spatially aggregated results were consistent, as the model repeatedly identified the same sub-catchments as the main sediment sources in the catchment. The accuracy of these estimates is difficult to assess, although some insight can be gained by an evaluation against fingerprinting source apportionments.

The bootstrapping method for solving the un-mixing models resulted in uncertain sediment fingerprinting estimates of relative source contributions, particularly for Node 1. Bootstrapping methods are known to produce somewhat spurious uncertainty bands of un-mixing model results, as local optimization functions frequently yield numerical solutions where one source provides 100 % of source contributions (Cooper et al., 2014). This is illustrated by the bi-modal density curves in Figure 4.

Nonetheless, the uncertainty of Node 1 un-mixing model solutions indicates there might be an issue with the data. Moreover, the negligible contributions from MRT (by far the largest sub-catchment in the basin) during the rainy season bring to question the consistency of the model results as a narrative. It might be the case that there was an issue of particle size incommensurability between MRT and Node 1 sink sediments. MRT element concentrations were overall higher than in the remaining Node 1 sources and sink samples, which might indicate MRT sediments were composed by smaller-sized particles. An alternative hypothesis is that sediment storage in the MRT sub-catchment is influenced by small hydroelectric plants in the Mortes River or by civil engineering works in the cities the river flows through before its confluence with the Elvas River. Regardless, fingerprinting and SEDD model outputs shown a contrasting pattern for Node 1, and we have no supporting evidence to corroborate either of the system representations.

On the other hand, the overall correspondence of fingerprinting un-mixing model solutions and SEDD simulations of relative source contributions for Nodes 2 and 3, while considering the uncertainty in both system representations, provide some conditional corroboration of the methods. Although the SEDD model simulates long-term sediment transport dynamics and the fingerprinting approach was limited by the temporal scale of our sampling, both modelling exercises designated that most of the sediments reaching Nodes 2 and 3 are originated from farther upstream sources. That is, at least under the reasonable assumption that rainy season fingerprinting results represent the bulk of the sediment transport dynamics in the catchment. For management purposes, the convergence of model results is an important outcome of this research. Different models and sources of data have indicated that the sediments reaching the Funil reservoir by the Mortes River come from the mid and upper catchment areas, even during a short temporal scale. Hence, reducing reservoir sedimentation rates requires widespread soil conservation efforts throughout the catchment, instead of local/proximal interventions.

Another valuable outcome of this research was demonstrating how uncertain common large-scale distributed RUSLE applications are. Although RUSLE is the most widely used soil erosion model in the world (Alewell et al., 2019), studies which have attempted to quantify model error are scarce (e.g. Tetzlaff et al., 2013). Our results indicate that numerical RUSLE predictions of spatially distributed erosion rates were practically meaningless, given the uncertainty in the model outputs (see Figure 5). Of course, these results are case specific and entirely determined by the assumptions made about potential sources of model error, which we understand, were cautious. That is, the uncertainty in the rainfall erosivity regression equation was not properly assessed, let alone in the equations relating rainfall intensity to kinetic energy (Wilken et al., 2018). Moreover, errors in the soil and land use map classifications were not represented, nor were the potential errors in the plot-based experiments used to generate RUSLE factors (Nearing, 2000; Parsons, 2019). Hence, similar or larger errors should be expected in comparable spatially distributed RUSLE applications elsewhere, unless otherwise demonstrated.

Overall, results from our forward error analysis indicate that RUSLE-modeled spatially distributed erosion rates should be viewed with extreme caution, particularly when actual numerical model outputs are used to project the influence of climate and land use changes on future erosion rates (e.g., Teng et al., 2018; Yang et al., 2003). Due to the difficulties involved in large-scale model parameterization, the costs of plot-based experiments for developing empirical model factors, and the multiplicative structure of the RUSLE (and USLE-family models), we suspect that model applications should remain largely uncertain. This might be particularly true for developing countries such as Brazil, where data scarcity further complicates model parameterization. Under such conditions, model testing should hereon focus on evaluating if the models are at least consistently capable of relatively ranking erosion-prone areas, as in Fischer et al. (2018).

5 Conclusions

Soil erosion models and the measurements of system responses we call observational data are necessarily uncertain. A failure to represent such uncertainty is at best naïve. Here we provided a framework for incorporating the uncertainty of sediment rating curves, sediment fingerprinting un-mixing models, and soil erosion/sediment delivery models into the GLUE methodology. More specifically, the framework was applied to the RUSLE-based SEDD model at a large catchment in Southeast Brazil.

Our results have shown how common large-scale spatially-distributed RUSLE applications are highly uncertain. This means model applications of such type cannot afford to disregard uncertainty analysis, and that modeled erosion rates should be interpreted with upmost caution. SEDD simulations of catchment sediment yields were also highly uncertain, mostly due to the errors in the rating curve forcing data and the sensitivity of the model to the empirical parameter β. Spatially distributed simulations of area specific sediment yields were even more uncertain, which meant the grid-based numerical model outputs were of little utility. However, when the SEDD model outputs were lumped into sub-catchment relative contributions, results were consistent and far less uncertain.

The comparison between SEDD model outputs the fingerprinting source apportionments presented here was facilitated by the hierarchical tributary sampling design we employed. Moreover, the uncertainty-based framework enabled us to compare distributions of model realizations of relative source contributions. The comparison revealed an overall similarity of fingerprinting and SEDD-modeled distributions of source apportionments, although large discrepancies were found in part of the catchment.

Ultimately, we found that under the testing conditions, the SEDD model might be useful for identifying the sub-catchments that contribute to most of the sediment load in the Mortes River basin. On the other hand, the uncertainty in the simulations questions the model's usefulness for calculating actual erosion and sediment delivery rates. From a falsifacationist perspective, the model could not be rejected, as multiple model realizations produced acceptable system representations. However, this was largely facilitated by the uncertainty in the forcing data. One of the most important conclusions from this research is that we need better data in order to reject models and therefore to improve our understanding of soil erosion and sediment transport in large catchments. This will require honest representations of the uncertainty in models and observational data. Moreover, multiple sources of data are necessary to evaluate model usefulness and consistency, as we have shown.

6 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).

7 REFERENCES

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. https://doi.org/10.5194/bg-13-1587-2016

Alewell, C., Borelli, P., Meusburger, K., Panagos, P., 2019. Using the USLE: Chances, challenges and limitations of soil erosion modelling. Int. Soil Water Conserv. Res. https://doi.org/10.1016/j.iswcr.2019.05.004

Alvares, C.A., Stape, J.L., Sentelhas, P.C., de Moraes Gonçalves, J.L., Sparovek, G., 2013. Köppen’s climate classification map for Brazil. Meteorol. Zeitschrift 22, 711–728. https://doi.org/10.1127/0941-2948/2013/0507

Aquino, F., Silva, N., Leandro, M., Freitas, F. De, Antonio, D., Rogério, C., Avanzi, C., Aquino, R.F., Curi, N., 2014. Erosividade das chuvas e tempo de recorrência para Lavras , Minas Gerais.

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. https://doi.org/10.1002/hyp.1447

Batista, P.V.G., Davies, J., Silva, M.L.N., Quinton, J.N., 2019a. On the evaluation of soil erosion models : Are we doing enough ? Earth-Science Rev. 197, 102898. https://doi.org/10.1016/j.earscirev.2019.102898

Batista, P.V.G., Laceby, J.P., Silva, M.L.N., Tassinari, D., Bispo, D.F.A., Curi, N., Davies, J., Quinton, J.N., 2019b. Using pedological knowledge to improve sediment source apportionment in tropical environments. J. soils a 19, 3274–3289.

Batista, P.V.G., Silva, M.L.N., Silva, B.P.C., Curi, N., Bueno, I.T., Acérbi Júnior, F.W., Davies, J., Quinton, J.N., 2017. Modelling spatially distributed soil losses and sediment yield in the upper Grande River Basin - Brazil. Catena 157, 139–150. https://doi.org/10.1016/j.catena.2017.05.025

Beven, K.J., 2019. How to make advances in hydrological modelling 1–14. https://doi.org/10.2166/nh.2019.134

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

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. https://doi.org/10.1016/j.jhydrol.2005.07.007

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

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.

Blake, W.H., Boeckx, P., Stock, B.C., Smith, H.G., Bodé, S., Upadhayay, H.R., Gaspar, L., Goddard, R., Lennard, A.T., Lizaga, I., Lobb, D.A., Owens, P.N., Petticrew, E.L., Kuzyk, Z.Z.A., Gari, B.D., Munishi, L., Mtei, K., Nebiyu, A., Mabit, L., 2018. A deconvolutional Bayesian mixing model approach for river basin sediment source apportionment 1–12. https://doi.org/10.1038/s41598-018-30905-9

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. https://doi.org/10.1002/ldr.2898

Boudreault, M., Koiter, A.J., Lobb, D.A., Liu, K., Benoy, G., Owens, P.N., Li, S., 2019. Comparison of sampling designs for sediment source fingerprinting in an agricultural watershed in Atlantic Canada. J. Soils Sediments. https://doi.org/10.1007/s11368-019-02306-6

Calway, R., Weston, S. 2017. foreach: Provides foreach looping construct for R. R package version 1.4.4.

Collins, A.L., Walling, D.E., Leeks, G.J.L., 1997. Source type ascription for fluvial suspended sediment based on a quantitative composite fingerprinting technique. Catena 29, 1–27.

Collins, A.L., Zhang, Y.S., Hickinbotham, R., Bailey, G., Darlington, S., Grenfell, S.E., Evans, R., Blackwell, M., 2013. Contemporary fine-grained bed sediment sources across the River Wensum Demonstration Test Catchment, UK. Hydrol. Process. 27, 857–884. https://doi.org/10.1002/hyp.9654

Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., Böhner, J., 2015. System for Automated Geoscientific Analyses (SAGA) v . 2 . 1 . 4 1991–2007. https://doi.org/10.5194/gmd-8-1991-2015

Cooper, R.J., Krueger, T., 2017. An extended Bayesian sediment fingerprinting mixing model for the full Bayes treatment of geochemical uncertainties. Hydrol. Process. 31, 1900–1912. https://doi.org/10.1002/hyp.11154

Cooper, R.J., Krueger, T., Hiscock, K.M., Rawlins, B.G., 2014. Sensitivity of fluvial sediment source apportionment to mixing model assumptions: A Bayesian model comparison. Water Resour. Res. 50, 9031–9047. https://doi.org/10.1002/2014WR016194.Received

de Vente, J., Poesen, J., Verstraeten, G., Van Rompaey, A., Govers, G., 2008. Spatially distributed modelling of soil erosion and sediment yield at regional scales in Spain. Glob. Planet. Change 60, 393–415. https://doi.org/10.1016/j.gloplacha.2007.05.002

Didoné, E.J., Paolo, J., Minella, G., Merten, G.H., 2015. Quantifying soil erosion and sediment yield in a catchment in southern Brazil and implications for land conservation Quantifying soil erosion and sediment yield in a catchment in southern Brazil and implications for land conservation. J. Soils Sediments. https://doi.org/10.1007/s11368-015-1160-0

Duraes, M.F., de Mello, C.R., Beskow, S., 2016. Sediment yield in Paraopeba River Basin – MG, Brazil. Int. J. River Basin Manag. 14, 367–377. https://doi.org/10.1080/15715124.2016.1159571

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. https://doi.org/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. https://doi.org/10.1016/j.envsci.2005.04.009

Evrard, O., Poulenard, J., Némery, J., Ayrault, S., Gratiot, N., Duvert, C., Prat, C., Lefèvre, I., Bonté, P., Esteves, M., 2013. Tracing sediment sources in a tropical highland catchment of central Mexico by using conventional and alternative fingerprinting methods. Hydrol. Process. 27, 911–922. https://doi.org/10.1002/hyp.9421

Favis-Mortlock, D., Boardman, J., MacMillan, V., 2001. Landscape Erosion and Evolution Modeling. Springer US, Boston, MA. https://doi.org/10.1007/978-1-4615-0575-4

FEAM - Fundação Estadual Do Meio Ambiente. 2010. Mapa de solos de Minas Gerais: legenda expandida. FEAM/UFV/CETEC/UFLA, Belo Horizonte. Brazil.

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., Minacapilli, M., 1995. Sediment delivery processes at basin scale. Hydrol. Sci. J. 40, 703–717. https://doi.org/10.1080/02626669509491460

Ferro, V., Porto, P., 2000. SEDIMENT DELIVERY DISTRIBUTED (SEDD) MODEL. J. Hydrol. Eng. 411–422.

Ferro, V., Stefano, C.D.I., 2003. Calibrating the SEDD model for Sicilian ungauged basins.

Fick, S.E., 2017. WorldClim 2 : new 1-km spatial resolution climate surfaces for global land areas 4315, 4302–4315. https://doi.org/10.1002/joc.5086

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.

https://doi.org/10.1002/esp.4216

Fu, G., Chen, S., McCool, D.K., 2006. Modeling the impacts of no-till practice on soil erosion and sediment yield with RUSLE, SEDD, and ArcView GIS. Soil Tillage Res. 85, 38–49. https://doi.org/10.1016/j.still.2004.11.009

Gelman, A., Hill, J., 2007. Data Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge Univeristy Press, New York.

Ghalanos, A., Theussl, S. 201. Rsolnp: General non-linear optimization using augmented lagrange multiplier method. R package version 1.16

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.

Habibi, S., Gholami, H., Fathabadi, A., Jansen, J.D., 2019. Fingerprinting sources of reservoir sediment via two modelling approaches. Sci. Total Environ. 663, 78–96. https://doi.org/10.1016/j.scitotenv.2019.01.327

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. https://doi.org/10.1002/esp.4372

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. https://doi.org/10.1002/esp.4149

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

Klages, M.G., Hsieh, Y.P. 1975. Suspended solids carried by the Gallatin River of Southwestern Montana: II. Using mineralogy for inferring sources. J Environ Qual 4:68–73.

Koiter, A.J., Lobb, D.A., Owens, P.N., Petticrew, E.L., Tiessen, K.H.D., Li, S., 2013. Investigating the role of connectivity and scale in assessing the sources of sediment in an agricultural watershed in the Canadian prairies using sediment source fingerprinting. J. Soils Sediments 13, 1676–1691. https://doi.org/10.1007/s11368-013-0762-7

Krasa, J., Dostal, T., Jachymova, B., Bauer, M., Devaty, J., 2019. Soil erosion as a source of sediment and phosphorus in rivers and reservoirs – Watershed analyses using WaTEM/SEDEM. Environ. Res. 171, 470–483. https://doi.org/10.1016/j.envres.2019.01.044

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. https://doi.org/10.1016/j.earscirev.2017.04.009

Laceby, J.P., McMahon, J., Evrard, O., Olley, J., 2015. A comparison of geological and statistical approaches to element selection for sediment fingerprinting. J. Soils Sediments 15, 2117–2131. https://doi.org/10.1007/s11368-015-1111-9

Laceby, J.P., Olley, J., 2015. An examination of geochemical modelling approaches to tracing sediment sources incorporating ... An examination of geochemical modelling approaches to tracing sediment sources incorporating distribution mixing and. https://doi.org/10.1002/hyp.10287

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.

https://doi.org/10.1016/j.catena.2014.03.004

Lin, C., Wu, Z., Ma, R., Su, Z., 2016. Detection of sensitive soil properties related to non-point phosphorus pollution by integrated models of SEDD and PLOAD. Ecol. Indic. 60, 483–494. https://doi.org/10.1016/j.ecolind.2015.07.023

Morgan, R.P.C., 2001. A simple approach to soil loss prediction: A revised Morgan-Morgan-Finney model. Catena 44, 305–322. https://doi.org/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. https://doi.org/10.1016/S0021-8634(84)80025-6

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. https://doi.org/10.1002/1096-9837(200008)25:9<1035::AID-ESP121>3.0.CO;2-B

Nosrati, K., Collins, A.L., Madankan, M., 2018. Fingerprinting sub-basin spatial sediment sources using different multivariate statistical techniques and the Modified MixSIR model Catena Fingerprinting sub-basin spatial sediment sources using di ff erent multivariate statistical techniques and the Modi. Catena 164, 32–43. https://doi.org/10.1016/j.catena.2018.01.003

Olley, J., Caitcheon, G., 2000. Major element chemistry of sediments from the Darling ± Barwon River and its tributaries : implications for sediment and phosphorus sources Abstract : Hydrol. Process. 14, 1159–1175.

Oreskes, N., Belitz, K., 2001. Philosophical Issues in Model Assessment, in: Anderson, M.G., Bates, P.D. (Eds.), Model Validation: Perspectives in Hydrological Science. John Wiley & Sons, pp. 23–41.

Panagos, P., Borrelli, P., Meusburger, K., 2015. A New European Slope Length and Steepness Factor (LS-Factor) for Modeling Soil Erosion by Water. Geosciences 5, 117–126. https://doi.org/10.3390/geosciences5020117

Parsons, A.J., 2019. How reliable are our methods for estimating soil erosion by water? Sci. Total Environ. 676, 215–221. https://doi.org/10.1016/j.scitotenv.2019.04.307

Parsons, A.J., Wainwright, J., Brazier, R.E., Powell, D.M., 2009. Is sediment delivery a fallacy? Earth Surf. Process. Landforms 34, 155–161. https://doi.org/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. https://doi.org/10.1061/(ASCE)HE.1943-5584.0001058

Pulley, S., Foster, I., Collins, A.L., 2016. The impact of catchment source group classification on the accuracy of sediment fingerprinting outputs. J. Environ. Manage. https://doi.org/10.1016/j.jenvman.2016.04.048

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

Rodriguez, E., Morris, C., Belz, J., 2006. An assessment of the SRTM topographic products. Photogramm. Eng. Remote Sensing 72, 249–260. https://doi.org/0099-1112/06/7203–0249/$3.00/0

Rustomji, P., Wilkinson, S.N., 2008. Applying bootstrap resampling to quantify uncertainty in fluvial suspended sediment loads estimated using rating curves 44, 1–12. https://doi.org/10.1029/2007WR006088

Schmidt, S., Tresch, S., Meusburger, K., 2019. MethodsX Modi fi cation of the RUSLE slope length and steepness factor (LS-factor) based on rainfall experiments at steep alpine grasslands. MethodsX 6, 219–229. https://doi.org/10.1016/j.mex.2019.01.004

Smith, H.G., Blake, W.H., 2014. Sediment fingerprinting in agricultural catchments: A critical re-examination of source discrimination and data corrections. Geomorphology 204, 177–191. https://doi.org/10.1016/j.geomorph.2013.08.003

Smith, H.G., Karam, D.S., Lennard, A.T., 2018a. Evaluating tracer selection for catchment sediment fingerprinting. J. soils. https://doi.org/10.1007/s11368-018-1990-7

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

Stefano, C. Di, Ferro, V., 2007. Evaluation of the SEDD model for predicting sediment yield at the Sicilian experimental SPA2 basin 32, 1094–1109. https://doi.org/10.1002/esp

Taguas, E.V., Moral, C., Ayuso, J.L., Pérez, R., Gómez, J. a., 2011. Modeling the spatial distribution of water erosion within a Spanish olive orchard microcatchment using the SEDD model. Geomorphology 133, 47–56. https://doi.org/10.1016/j.geomorph.2011.06.018

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. https://doi.org/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. https://doi.org/10.1016/j.jhydrol.2015.06.048

Teng, H., Liang, Z., Chen, S., Liu, Y., Viscarra, R.A., Chappell, A., Yu, W., Shi, Z., 2018. Current and future assessments of soil erosion by water on the Tibetan Plateau based on RUSLE and CMIP5 climate models 635, 673–686. https://doi.org/10.1016/j.scitotenv.2018.04.146

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. https://doi.org/10.1016/j.catena.2011.08.001

Theuring, P., Collins, A.L., Rode, M., 2015. Source identification of fine-grained suspended sediment in the Kharaa River basin, northern Mongolia. Sci. Total Environ. 526, 77–87. https://doi.org/10.1016/j.scitotenv.2015.03.134

Tiecher, T., Caner, L., Minella, J.P.G., Bender, M.A., dos Santos, D.R., 2016. Tracing sediment sources in a subtropical rural catchment of southern Brazil by using geochemical tracers and near-infrared spectroscopy. Soil Tillage Res. 155, 478–491. https://doi.org/10.1016/j.still.2015.03.001

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. https://doi.org/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.

https://doi.org/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. https://doi.org/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. https://doi.org/10.1002/esp.275

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. https://doi.org/10.1111/j.1475-2743.2002.tb00257.x

Vigiak, O., Bende-michl, U., 2013. Estimating bootstrap and Bayesian prediction intervals for constituent load rating curves 49, 8565–8578. https://doi.org/10.1002/2013WR013559

Vigiak, O., Malagó, A., Bouraoui, F., Vanmaercke, M., Poesen, J., 2015. Adapting SWAT hillslope erosion model to predict sediment concentrations and yields in large Basins. Sci. Total Environ. 538, 855–75. https://doi.org/10.1016/j.scitotenv.2015.08.095

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. https://doi.org/10.1016/S0167-1987(02)00126-5

Walling, D.E., Woodward, J.C. 1995. Tracing sources of suspended sediment in river basins: a case study of the River Clum, Devon, UK. Mar Freshw Res 46:327–336.

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. https://doi.org/10.1016/j.catena.2005.08.010

Wilken, F., Baur, M., Sommer, M., Deumlich, D., Bens, O., Fiener, P., 2018. Uncertainties in rainfall kinetic energy-intensity relations for soil erosion modelling. Catena 171, 234–244. https://doi.org/10.1016/j.catena.2018.07.002

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. https://doi.org/10.1016/j.agee.2012.02.002

Wilkinson, S.N., Olley, J.M., Furuichi, T., Burton, J., Kinsey-Henderson, A.E., 2015. Sediment source tracing with stratified sampling and weightings based on spatial gradients in soil erosion. J. Soils Sediments 15, 2038–2051. https://doi.org/10.1007/s11368-015-1134-2

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. https://doi.org/10.1016/j.envsoft.2008.09.006

Yang, D., Kanae, S., Oki, T., Koike, T., Musiake, K., 2003. Global potential soil erosion with reference to land use and climate changes 2928, 2913–2928. https://doi.org/10.1002/hyp.1441

Yu, L., Oldfield, F. 1989. A multivariate mixing model for identifying sediment source from from magnetic measurements. Quat Res 32:168–181.

CONCLUDING REMARKS

So that science that was to teach me everything ends up in a hypothesis, that lucidity founders in metaphor, that uncertainty is resolved in a work of art. What need had I of so many efforts? […]

I realize that if through science I can seize phenomena and enumerate them, I cannot, for all that, apprehend the world.

Albert Camus, The Myth of Sisyphus, 1942

This thesis addressed the methodological issues regarding the evaluation of soil erosion models. I focused on investigating the methods and sources of data that should enable modelers to analyze the usefulness and consistency of their models according to the purpose and scale of their applications. Moreover, I tried to advance some of the approaches for representing the uncertainty in soil erosion models and observational testing data, as well as for establishing limits of acceptability of model error. These advancements are built upon the theoretical and methodological foundations established by Quinton (1997, 1994) and Beven (2019, 2018, 2009, 2006). I particularly focused on sediment fingerprinting as a mean for acquiring sediment provenance data for evaluating spatially-distributed soil erosion models.

I have shown how model evaluation is currently a neglected topic in soil erosion modelling research (Paper 01). Based on a meta-analysis of model performance, I demonstrated how different erosion 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. I argued that erosion models should not be calibrated based on a deterministic optimization of model parameters. Instead, I demonstrated how a criteria for establishing limits of acceptability of model error based on the variability of erosion measurements (Nearing, 2000) could be used to filter behavioral parameter sets and model realizations within the GLUE methodology (Beven and Binley, 1992).

Based on a literature review, I concluded that spatially-distributed soil erosion models frequently compare poorly to independent spatial estimates of soil redistribution rates, even while making accurate predictions of outlet sediment transport rates. Therefore, evaluating spatial models requires spatial testing data, and outlet measurements of sediment loads are insufficient to characterize model performance. Moreover, the epistemic uncertainties associated to model structures and measurements of system responses impose that any model evaluation methodology should be established upon an explicit representation of the uncertainty in models and observational data (Beven, 2019)

Focusing on sediment fingerprinting as a potential source of testing data for soil erosion models, I investigated how pedological knowledge could be incorporated into source stratification and geochemical tracer selection for analyzing sediment provenance in tropical catchments (Paper 02). Moreover, I studied how pedogenetic processes can lead to the development of geochemical source signals on soils developed from similar parent materials, and how the expression of these signals is controlled by sediment particle size. I concluded that the proposed knowledge-based element selection method facilitated the interpretation of fingerprinting un-mixing model results, and that different sampling strategies and source stratification methods might be necessary to model sediment dynamics in large river catchments. On a management level, the fingerprinting results indicated that the fine sediments reaching the outlet of the Ingaí River basin (~1200 km^2) are predominantly derived from Ustorthents located on lower catchment. This highlights how this Entisol region is environmentally sensitive and erosion-prone.

Furthermore, I have shown how sediment fingerprinting source apportionments can be incorporated into spatially-distributed soil erosion model testing within the GLUE framework (Paper 03). This was performed by applying the RUSLE-based Sediment Delivery Distributed model (Ferro and Minacapilli, 1995) in the Mortes River basin (~6600 km²). I demonstrated that when the uncertainty of sediment load estimates is considered, multiple parameter sets and model realizations provide acceptable simulations of catchment sediment yields. Hence, I concluded that the SEDD model could not be falsified under the testing conditions, and that better data are necessary in order to reject non-behavioral model realizations. This approach led me to further conclude that the uncertainty in the model outputs was so large that actual numerical predictions of erosion and sediment delivery rates were of little use for quantifying sediment dynamics. On the other hand, a comparison between the tributary-based fingerprinting source apportionments and spatially-aggregated SEDD results indicated that the model might be useful for identifying the sub-catchments that contributed to most of the sediment loads from the Mortes River. For management purposes, the outcomes of this research indicate that the sediments reaching the Mortes River delta in the Funil reservoir are mainly derived from the mid and upper catchment areas. Hence, reducing reservoir sedimentation requires widespread soil conservation efforts, instead of local/proximal interventions.

Ultimately, this thesis has shown how uncertainty permeates all facets of soil erosion modelling and the measurements with which such models are tested – from small plots to large river basins. Although I can't say this is a novel finding (e.g., Brazier et al., 2000; Quinton, 1997; Van Rompaey and Govers, 2002), I hope the methods developed here can help other modelers to deal with uncertainty – which will always be there, as "hypotheses always remain hypotheses, that is, presuppositions whose complete certainty we can never attain" (Kant, 1988).

Accordingly, I have made an effort to only use free software and open source programming languages for my analyses, and shared all my raw data and codes on the publications presented here (with the exception of Paper 03, which has not yet been submitted to a journal). This is because I believe that we, soil erosion modelers, should hold ourselves to the highest standards of documentation, and that models should be fully reproducible and open to criticism (Sterman, 2002). This is necessary to advance knowledge, to make meaningful and useful predictions, and to provide honest descriptions of the shortcomings of our models.

The deleterious effects of soil erosion, as well as methods to reduce them, have been known for a long time (see Montgomery, 2007). Still, while collecting samples for this research, I saw large gullies eating away the scarce pastures that some impoverished farmers had left. I saw crystalline streams becoming muddy brown rivers. And I wondered: what good are my models for? Would my government-funded research grants not be better spent in public policies and agricultural extension services? Honestly, I do not have an answer to these questions. But I know that if we expect that soil erosion models are to have any real-world impact, we should first let go of tests and sources of data that are ultimately designed promote model acceptance and the status/authority of the modeler (Sterman, 2002). Instead, we should focus on purpose-oriented critical model evaluation approaches, which scrutinize model deficiencies, encompass multiple sources of data, and fully acknowledge uncertainty and equifinality. This might lead to model improvements and more responsible decision-making.

REFERENCES

Beven, K.J., 2019. How to make advances in hydrological modelling 1–14. https://doi.org/10.2166/nh.2019.134

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

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. https://doi.org/10.1016/j.jhydrol.2005.07.007

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

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.

Ferro, V., Minacapilli, M., 1995. Sediment delivery processes at basin scale. Hydrol. Sci. J. 40, 703–717. https://doi.org/10.1080/02626669509491460

Kant I. 1988. Logic Courier. Dover Publications: New York.

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

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. https://doi.org/10.1002/1096-9837(200008)25:9<1035::AID-ESP121>3.0.CO;2-B

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. Cranfield University.

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

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. https://doi.org/10.1080/13658810210148561