UNIVERSIDADE FEDERAL DO CEARÁ

CENTRO DE CIÊNCIAS AGRÁRIAS

DEPARTAMENTO DE ENGENHARIA AGRÍCOLA

PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA AGRÍCOLA

PEDRO HENRIQUE LIMA ALENCAR

TOO MUCH, TOO LITTLE – ENTROPY AND STATISTICS AS TOOLS FOR HYDROLOGY: A SCIENCE OF IRREGULAR DISTRIBUTION OF DATA AND EVENTS

FORTALEZA

2022

PEDRO HENRIQUE LIMA ALENCAR

TOO MUCH, TOO LITTLE – ENTROPY AND STATISTICS AS TOOLS FOR
HYDROLOGY: A SCIENCE OF IRREGULAR DISTRIBUTION OF DATA AND
EVENTS

Tese apresentada ao Programa de Pós-Graduação em Engenharia Agrícola do Departamento de Engenharia Agrícola da Universidade Federal do Ceará, como parte dos requisitos para obtenção do título de Doutor em Engenharia Agrícola. Área de concentração: Manejo e Conservação de Bacias Hidrográficas no Semiárido.

Orientador: Prof. Dr. José Carlos de Araújo
Orientadora: Profa. Dra. Eva Nora Paton

FORTALEZA

2022

PEDRO HENRIQUE LIMA ALENCAR

TOO MUCH, TOO LITTLE – ENTROPY AND STATISTICS AS TOOLS FOR HYDROLOGY: A SCIENCE OF IRREGULAR DISTRIBUTION OF DATA AND EVENTS

> Tese apresentada ao Programa de Pós-Graduação em Engenharia Agrícola do Departamento de Engenharia Agrícola da Universidade Federal do Ceará, como parte dos requisitos para obtenção do título de Doutor em Engenharia Agrícola. Área de concentração: Manejo e Conservação de Bacias Hidrográficas no Semiárido.

Aprovado em: 08/02/2022.

BANCA EXAMINADORA

_____
Profa. Dra. Galina Churkina
Technische Universität Berlin (TUB)

_____
Profa. Dra. Suzana Maria Giro Lima Montenegro
Universidade Federal de Pernambuco (UFPE)

_____
Prof. Dr. Iran Eduardo Lima Neto
Universidade Federal do Ceará (UFC)

_____
Prof. Dr. José Carlos de Araújo (Orientador)
Universidade Federal do Ceará (UFC)

_____
Profa. Dra. Eva Nora Paton (Orientadora)
Technische Universität Berlin (TUB)

TOO MUCH, TOO LITTLE

ENTROPY AND STATISTICS AS TOOLS FOR HYDROLOGY: A SCIENCE OF IRREGULAR DISTRIBUTION OF DATA AND EVENTS

PEDRO H. L. ALENCAR

—— Too much, too little ——

Entropy and Statistics as Tools for Hydrology: A Science of Irregular Distribution of Data and Events

vorgelegt von
M.Sc.
Pedro Henrique Lima Alencar
ORCID: 0000-0001-6221-8580

from the Faculty VI - Institute of Ecology
and the Department of Agricultural Engineering
Technischen Universität Berlin
and Universidade Federal do Ceará
zur Erlangung des akademischen Grades

Doktor rerum naturalium
-Dr. rer nat-

genehmigte Dissertation

Berlin 2022

Zusammenfassung

In den letzten Jahrzehnten hat die Anzahl von Extremwetter-Ereignissen im gleichen Tempo zugenommen wie die Häufigkeit von Regen- und Trockenwetterperioden. Extremes Wetter verursacht zahlreiche Naturgefahren wie Überschwemmungen, Murgänge und Kolmation, sowie auch Ernteausfälle und Waldbrände. Während Hydrologen sich im Rahmen der Dringlichkeit der Klimakrise mit komplexen Prozessen auseinandersetzen müssen, können die Daten für ihre Modelle einerseits knapp sein, wie dies der Fall ist in der Region des Globalen Südens, oder andererseits zu zahlreich und verrauscht. Dieses Szenario ungleich verteilter Daten und (Wasser-)Ereignisse war die Motivation für die vorliegende Dissertation.

Der Text untersucht die folgenden fünf Themen: (1) Erosionsmodellierung von Gullys, (2) Bewertung der Sedimentausbeute in nicht vermessenen Becken, (3) Blitzdürre – Definition und Schlüsselfaktoren, (4) Identifizierung solcher Ereignisse und (5) die Werkzeuge, mit denen Hydrologen Probleme im Zusammenhang mit extremen Wetterbedingungen und deren Auswirkungen angehen können. Diese fünf Fragen führten zur Erstellung von vier wissenschaftlichen Arbeiten und einer interaktiven App zur Identifizierung von Blitzdürren.

Die ersten beiden Veröffentlichungen befassen sich mit der Modellierung von Gullyerosion. Im ersten Artikel untersuchen wir die Schlüsselvariablen, um die langfristige Erosion durch Gullys erfolgreich zu modellieren. Wir präsentieren auch ein physikalisch-basiertes Modell für kleine permanente Gullys, die einen weit verbreiteten Erosionsprozess in der Semiarid-Region Brasiliens darstellen. In der zweiten Abhandlung schlagen wir eine neuartige Gleichung für die Schubspannungsverteilung in Kanalbetten und Wänden vor. Diese Gleichung führt, zusammen mit dem im ersten Papier vorgelegten Modellrahmen, zu einem neuartigen entropiebasierten Gullyerosionsmodell, das die Simulation der Gullyerosion und die Bewertung der Gesamterosion für Beitragsflächen von bis zu 8 Hektar ermöglicht.

In der dritten Publikation stellen wir eine neuartige Methode zur Berechnung der Sedimentausbeute und des Abgabeverhältnisses in einem ereignisbasierten Zeitrahmen vor. Die Technik kombiniert eine temporale Downscaling-Methode und ein gut veröffentlichtes Sedimenttransportmodell, die beide auf dem Prinzip der maximalen Entropie basieren, einem Werkzeug der Informationstheorie, das es dem Modellierer ermöglicht, die beste Wahrscheinlichkeitsdichtefunktion zu identifizieren, ohne dabei unbewiesene Hypothesen zuzulassen. Die vorgeschlagene Methode lieferte vielversprechende Ergebnisse: Sie bewertete die Sedimentausbeute mit einer Nash-Sutcliffe-Effizienz von 0,96.

Im letzten Artikel untersuchen wir die Definition und die Schlüsselfaktoren für das Auftreten von Blitzdürre im mitteleuropäischen Ackerland. Wir haben die Identifizierung von Blitzdürren mithilfe von fünf verschiedenen, gut veröffentlichten Methoden und einer von den Autoren vorgeschlagenen Methode, inspiriert durch unsere Definition von Blitzdürren, verglichen. Es wurde ein hohes Maß an Synchronität einzelner Blitzdürre-Ereignisse festgestellt, aber auch eine gewisse Divergenz in Dürreperioden. Um die Stärken und Schwächen dieser Methoden auszugleichen schlugen wir vor, einen Ensemble-Ansatz zur Ereignisidentifikation zu verwenden. Alle Methoden wurden in einem R-Paket implementiert und stehen der Öffentlichkeit als Shiny-App zur Verfügung.

Abstract

During the past decades, the number of extreme weather events has increased at the same pace as the growing frequency of both unusual wet and dry weather periods. Extreme weather causes multiple natural hazards, such as floods, debris flow and oversiltation, as well as crop losses and forest fires. While hydrologists have to deal with complex processes under the urgency of the climate crisis, data to power their models can be both scarce, as in regions of the Global South, or too numerous and noisy. This scenario of unequally distributed data and (water) events was the motivation for the present dissertation.

The text s driven by five questions regarding (1) gully erosion modelling, (2) sediment yield assessment in ungauged basins, (3) the key factors of flash drought, its definition and (4) how to identify such events, and (5) the tools available for hydrologists to tackle problems related to extreme weather and their impacts. These five questions led to the preparation of four scientific papers and an interactive application for flash drought identification.

The first two papers deal with gully erosion modelling. In the first paper, we explore the key variables to successfully model long term erosion by gullies. We also present a physically-based model for small permanent gullies, a common erosion process in the Brazilian Semiarid Region. In the second paper, we propose a novel equation for shear stress distribution in channel beds and walls. This equation, together with the model framework submitted in the first paper, leads to a novel entropy-based gully erosion model that allows simulation of gully erosion and assessment of total erosion for contribution areas up to 8 hectares.

In the third paper, we introduce a novel methodology to assess sediment yield and delivery ratio in an event-based timeframe. The technique couples a temporal downscaling method and a well-published sediment transport model, both based on the principle of maximum entropy, a tool from Information Theory which allows the modeller to identify the best probability density function without admitting any unproven hypotheses. The proposed method presented promising results: it assessed sediment yield at a Nash-Sutcliffe Efficiency of 0.96.

In the last paper, we study the definition and key factors of flash drought occurrence in Central European croplands. We compared the identification of flash droughts using five different well-published methods and one proposed by the authors under the light of our definition of flash droughts. A large degree of synchronicity of individual flash drought events was identified, but also some divergence in drought periods. So as to balance out the strengths and weaknesses of those methods, we suggested using an ensemble approach for

event identification. All methods were implemented in an R package and are also available in a Shiny app for the public.

Resumo

Nas últimas décadas, o número de eventos climáticos extremos aumentou no mesmo ritmo que a crescente frequência de períodos incomuns de clima úmido e seco. As condições meteorológicas extremas causam vários riscos naturais, como inundações, enxurradas e assoreamento, bem como perdas de safra e incêndios florestais. Enquanto os hidrólogos têm que lidar com processos complexos sob a urgência da crise climática, os dados para alimentar seus modelos podem ser escassos, como nas regiões do Sul Global, ou muito numerosos e cheios de ruído. Este cenário de dados e eventos (água) desigualmente distribuídos foi a motivação para a presente tese.

O texto é orientado por cinco questões relativas a (1) modelagem de erosão por voçorocas, (2) avaliação de produção de sedimentos em bacias não monitoradas, (3) os principais fatores causadores de *flash droughts*, sua definição e (4) como identificar tais eventos, e (5) as ferramentas disponíveis para hidrólogos para lidar com problemas relacionados a condições meteorológicas extremas e seus impactos. Essas cinco questões levaram à preparação de quatro artigos científicos e um aplicativo interativo para identificação de *flash droughts*.

Os primeiros dois artigos tratam da modelagem de erosão em voçorocas. No primeiro artigo, exploramos as variáveis-chave para modelar com sucesso a erosão de longo prazo por voçorocas. Apresentamos também um modelo de base física para pequenas voçorocas permanentes, um processo erosivo comum no Semiárido Brasileiro. No segundo artigo, propomos uma nova equação para a distribuição da tensão de cisalhamento em leitos e paredes de canais. Esta equação, associada a estrutura do modelo submetida no primeiro artigo, leva a um novo modelo de erosão de voçorocas baseado em entropia que permite a simulação da erosão de voçorocas e avaliação da erosão total para áreas de contribuição de até 8 hectares.

No terceiro artigo, apresentamos uma nova metodologia para estimar produção e taxa de aporte de sedimentos baseado em eventos. A técnica acopla um método de redução de escala temporal e um modelo de transporte de sedimentos da literatura, ambos baseados no princípio da entropia máxima, uma ferramenta da Teoria da Informação que permite identificar a melhor função de densidade de probabilidade sem admitir hipóteses não comprovadas. O método proposto avaliou a produção de sedimentos com eficiência Nash-Sutcliffe de 0,96.

No último artigo, estudamos a definição e os principais fatores da ocorrência de *flash droughts* em áreas agrícolas da Europa Central. Comparamos a identificação de *flash droughts* usando cinco métodos diferentes bem publicados e um proposto pelos autores considerando nossa definição de *flash droughts*. Um grande grau de sincronicidade de eventos individuais

de *flash drought* foi identificado, mas também alguma divergência nos períodos de seca. Para equilibrar os pontos fortes e fracos desses métodos, sugerimos o uso de uma abordagem conjunta de múltiplos métodos e critérios para identificação de eventos. Todos os métodos foram implementados em um pacote R e também estão disponíveis em um Shiny App.

Acknowledgements

Table of Contents

List of Figures

List of Tables

<div style="text-align: right; font-size: 4em; color: #999;">1</div>

Introduction

1.1 Hydrology and extreme weather

We live in a changing climate and a changing world. Recent events, as the intense floods in Western Germany of 2021, the recent droughts in Europe (2018-2019) and in the Brazilian Northeast (2012-2017) are a call for action, not only to stop greenhouse gasses emissions, but also to advance in our understanding of causes and effects of such extreme weather.

There is evidence that both extremely wet and extremely dry events are increasing in frequency (Herold et al., 2021). To illustrate this statement, we built Figures 1.1 and 1.2, as an approach to observe trends in extremely wet and dry weather (*wetness* and *dryness*, respectively). As an example, we took data from the centennial weather station of Potsdam (Germany), in the periphery of Berlin, that has been providing continuous monitoring since 1893 (Fenner et al., 2018). We built *ridge plots* to analyse long term (120 years) trends in the severity and duration of wet and dry periods, assessed with the Standardized Precipitation Index (SPI – McKee, Doesken, Kleist, et al., 1993).

We can observe a clear trend of intensification of both severity and duration, with dry and wet periods becoming longer and harsher. This process happens simultaneously for both wet and dry events, suggesting a stronger alternance (and, therefore, higher frequency; Herold et al., 2021) between extremes. In both Figures 1.1 and 1.2, we observe the emergence of a longer tail towards extreme values in the last decades, indicating that new extremes, rarely observed before, are now significant. These changes have been evolving over the decades, with the last 40 years presenting the most skewed data.

Extreme climate can affect the ecology of regions (Easterling et al., 2000; Stenseth et al., 2002, e.g., populations abundance, distribution, and behaviour) and impose economic and social challenges (Lindner et al., 2010). Understanding the dynamics involved in extreme

<div style="text-align: center;">1</div>
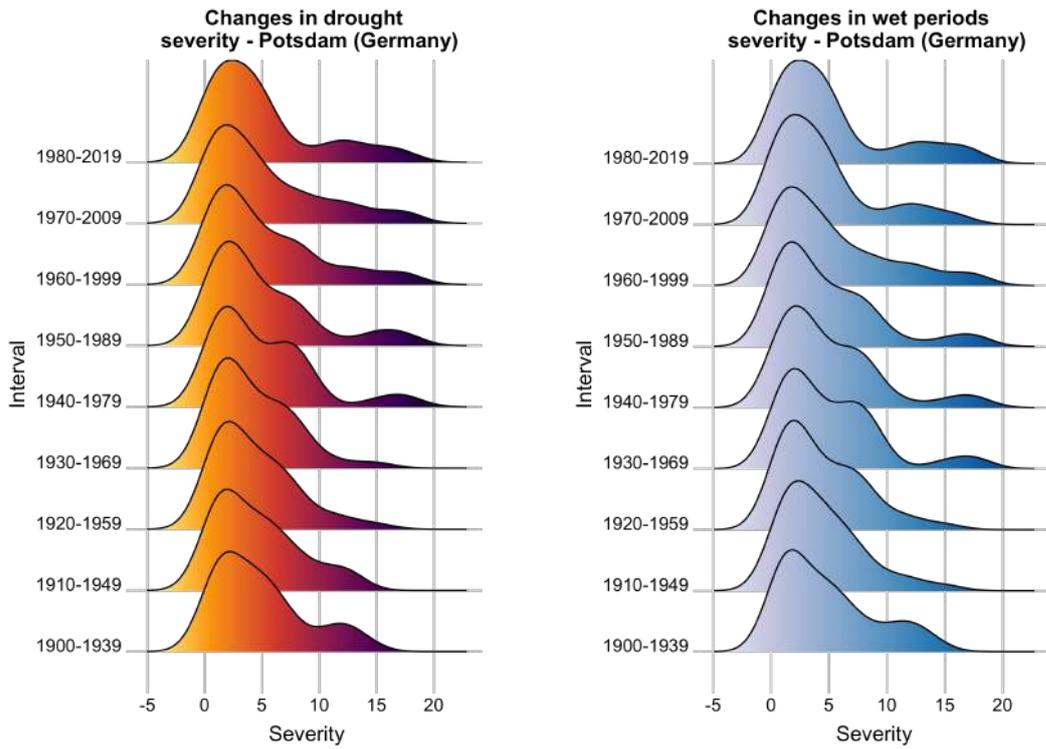
Figure 1.1: Changes in severity (ΣSPI) distribution of both wet and dry periods in Potsdam. The distributions are for 40-year long intervals, ten years apart. Note that the drought severity had the sign inverted for visualization help. By definition, droughts are periods with negative SPI, and severity.
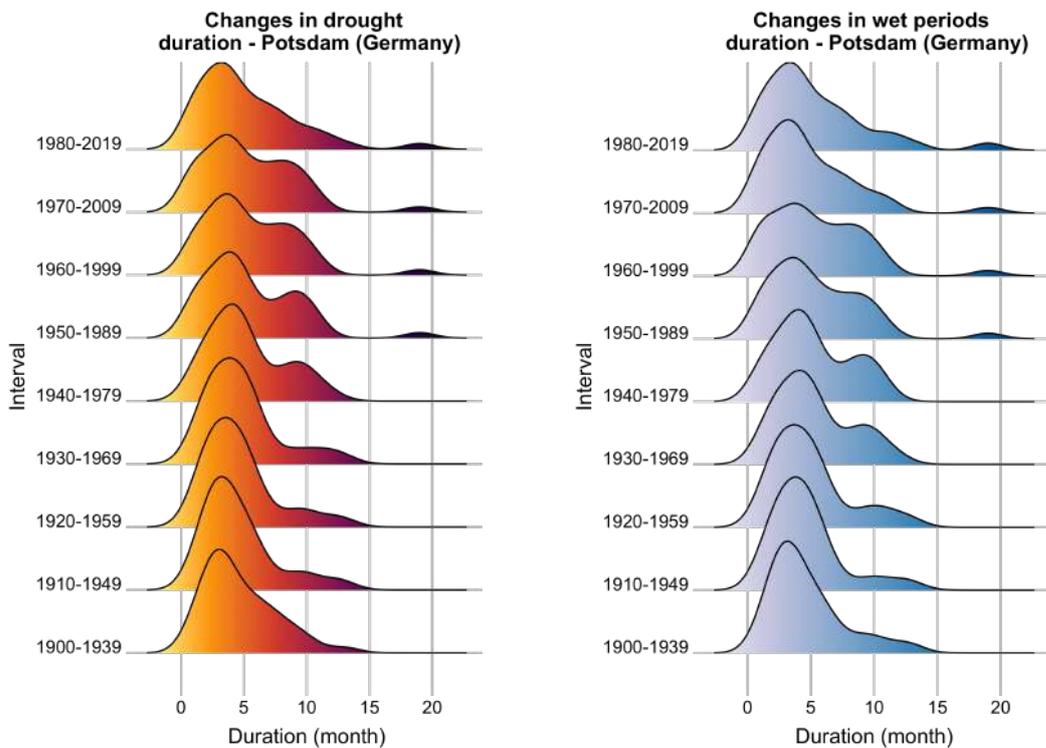
Figure 1.2: Changes in duration distribution of both wet and dry periods in Potsdam. The distributions are for 40-year long intervals, ten years apart.

climate conditions and their impacts in ecosystems and landscapes is a necessary step towards preservation and sustainability.

1.2   Hydrology and data

Hydrology, according to Chow, Maidment, and Mays (1988), is the physical science that studies the hydrologic cycle, encompassing the distribution and circulation of water, its physical and chemical properties and its interactions with the environment and all living things, particularly humans. Water is the conveyance of extreme weather[1], either through its abundance in floods, debris flows, and storms, or its scarcity as in dry spells, droughts, and flash droughts.

To understand the hydrological cycle, hydrologists make use of models, to represent quantitative or qualitatively the processes observed in nature (Singh, 1995). Those hydrologic models are powered by data, used to calibrate and/or validate the model. Nevertheless, data is often unevenly distributed, being abundant in some regions while very little data is available in other regions. As illustrated in Figure 1.3, which maps rain gauge stations distribution in the globe, the majority of regions with data scarcity is located in the Global South.

Number of Centennial Stations per country
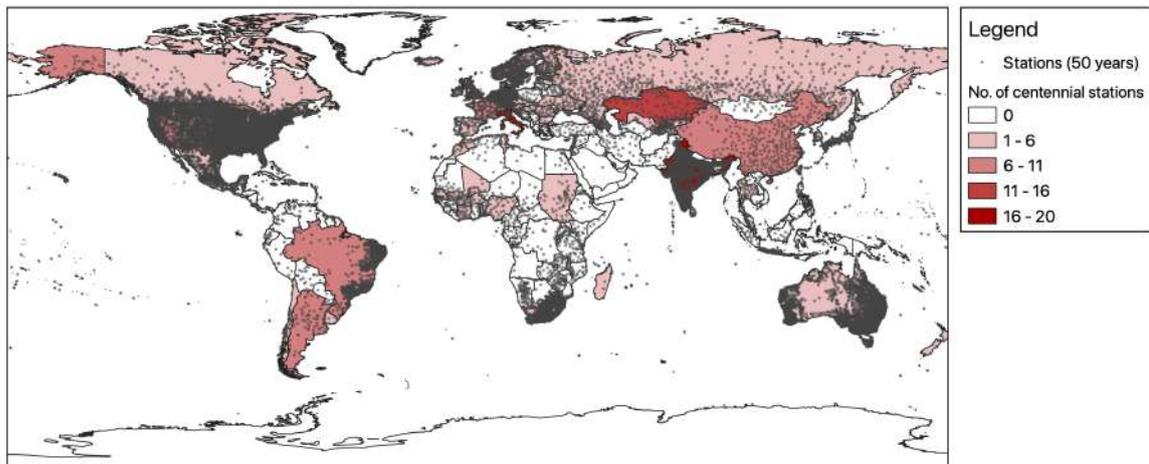
Source: World Meteorological Organization (2021) and US National Centers for Environmental Information (2021)
Projection: WGS84

Figure 1.3: Location of the Centennial Observing Stations. The data from the World Meteorological Organization. The WMO data accounts for 291 stations with over one hundred years of constant monitoring. The Global South has a considerable lack of such data. The dots indicate stations from the NCEI-NOAA with more than 50 years of data, also concentrated in Europe and USA, besides some regions of Australia, Brazil, India, and South Africa.

Figure 1.3 shows in colours the distribution of the centennial meteorological observation stations catalogued by the World Meteorological Organization. Such stations are considered an excellent source of data, being operated uninterruptedly and in the same place for more than 100 years. Only 61 countries have at least one centennial station catalogued, and the

---

[1]Or, as stated by Leonardo da Vinci, "the driving force of all nature" (Pfister, Savenije, Fenicia, et al., 2009).

majority are located in North America and Europe. Europe alone accounts for 49% of all catalogued stations, while the entire African Continent has only 10%. Figure 1.3 also shows the distribution of the rain stations catalogued by the NCEI-NOAA that have more than 50 years of continuous register and are still in operation. From the over 115 thousand stations around the globe compiled in the database, only 2979 satisfy those criteria.

In contrast to data scarcity, we have scenarios when hydrologists have too much data, and it is important to filter noise from meaningful information. This is hardly an easy task, as the boundaries between those two categories can be blurred (Klemes et al., 1982).

In the era of the data revolution, hydrologists have taken advantage of new sources of data (Vogel, Lall, et al., 2015), while the ever more powerful computers allow global models, as the ERA5 (the 5$^{\text{th}}$ generation of the European Centre for Medium-Range Weather Forecasts Re-Analysis - Hersbach et al., 2020) and GLEAM (Global Land Evaporation Amsterdam Model Martens et al., 2017). The ERA5 captures over 24 million data points per day and delivers over a hundred climatological variables, from precipitation to soil moisture (Hersbach et al., 2020), with three to five days of delay. The GLEAM uses a broad set of satellites and models to globally provide soil moisture in multiple depths, actual and potential evapotranspiration, among other variables (Martens et al., 2017).

Yet another rich data source is the FLUXNET2015 data set (Pastorello et al., 2020), containing a total of 212 sites with an eddy-covariance monitoring system and over sixty variables with hourly resolution. Some advantages of such a dataset are the direct measurement of the data, and a consistency test performed on all variables. In Figure 1.4, we observe the significant difference between the three datasets. Assuming the values provided by FLUXNET2015 as the ground-truth, both ERA5 and GLEAM fail to approximate the values of soil moisture from the station. GLEAM is more successful in representing the decrease during the summer, although it does not capture the quick-drying events in January and March. ERA5 presents overall a poor performance in this particular station, with wide variability.
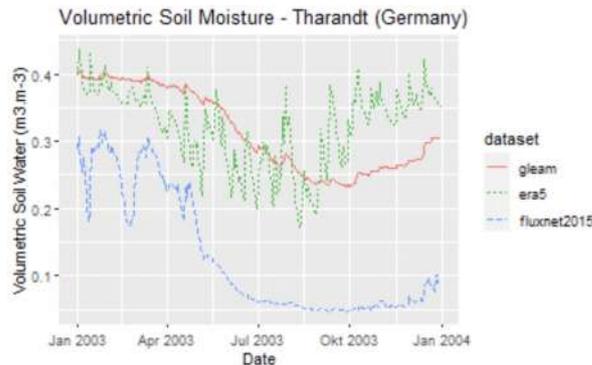
Figure 1.4: Comparison of Volumetric Soil Water values from three different datasets in Tharandt (Germany) during the year 2003.

It is part of the researcher responsibility to use the appropriate tools to select and analyse the data, convert it into meaningful information, and understand how the data and process interact (Wagener et al., 2010; Slater et al., 2019).

1.3   Motivation

In the perspective of climate change (Paglia and Parker, 2021), intensification of extreme weather (Herold et al., 2021) and the sustainable development goals (SDGs; UN, 2015), three research areas were selected: Erosion, Sediment Transport, and Droughts. Those three areas are strongly related to the SDGs and the water-energy-food nexus (WEF-Nexus; Simpson and Jewitt, 2019).

Erosion and Sediment transport are strongly related to "wet extremes", i.e., most erosion and sediment transport occur during extreme events (Coppus and Imeson, 2002). Also, in those areas, we deal with processes that usually present lack of data. We make use of Information and Entropy theories (Shannon, 1948; Kullback and Leibler, 1951; Jaynes, 1957a; Jaynes, 1957b; Kullback, 1978a) to help understand these processes using the available data.

Borrelli et al. (2017), and Poesen (2018) state that soil erosion is a key challenge for the 21$^{\text{st}}$ century and the achievement of the SDGs, being linked to multiple goals. The FAO (2019) has also pointed out erosion as a great threat to soil conservation and agriculture, being a threat to the WEF-Nexus, responsible for crop losses of up to 33.7 million tons and additional 48 km^3 of water usage yearly.

Erosion and sediment transport are natural processes of landscape formation, however, with intense land-use change, the loss of topsoil causes loss of arable land (Panagos, Borrelli, and Robinson, 2019). With the intensification of construction of dams in the last century (Zarfl et al., 2015), for energy production, flood control, or water supply, sediment that would leave the catchment now is trapped in reservoirs that work as sediment sinks. Such a process threatens the sustainability of those structures and hinders their benefits (Landwehr, Schomberg, and Pahl-Wostl, 2021), while current methods are not successful in estimating the amount of sediment that reaches these reservoirs (de Araújo, 2007). In this context, the selected problems are the modelling of erosion by gullies (Chapters 2 and 3), and the modelling of sediment yield (Chapter 4).

The third area (droughts) is related to "dry extremes". Droughts are ranked first place among natural hazards in terms of the number of affected people (Mishra and Singh, 2010). FAO (2021) lists the impacts of droughts and the risks imposed by them on food security and environmental conservation. In the last decade alone, the ten largest wildfires, all related to dry conditions, consumed 36 million hectares. Additionally, drought induces crop losses that cost over 37 billion US dollars yearly, and the consequent loss in dietary energy production (calories) leading to food and nutritional insecurity (FAO, 2021).

Despite the substantial literature on droughts (McKee, Doesken, Kleist, et al., 1993; Mishra and Singh, 2010; Serrano, Begueria, and Moreno, 2010; Mishra and Singh, 2011; Marengo, Chou, et al., 2011; Samaniego, Kumar, and Zink, 2013; de Araújo and Bronstert, 2015; Svoboda, Fuchs, et al., 2016; Brito et al., 2017; Hao et al., 2018; Krakauer, Lakhankar, and Hudson, 2019; Tijdeman and Menzel, 2021), there is a gap in our understanding of Flash

Droughts (Otkin, Svoboda, et al., 2018), a different kind of drought, contrasting with the conventional drought, that is characterized by slow onset, long duration and large affected areas (Mishra and Singh, 2010). The interest in flash drought has increased recently, after some extreme events in the last decade (Christian, Basara, Otkin, and Hunt, 2019; Christian, Basara, Hunt, et al., 2020; Basara et al., 2019). However, there is still no consensus on a definition and measurable metrics for flash droughts.

1.4 State of the Art

In this dissertation, we deal with three research areas, divided into two parts. In the first part, we deal with Erosion and Sediment Transport related to "wet extremes". Also, in this first part, we usually deal with a lack of data. We make use of Information and Entropy theories (Shannon, 1948; Kullback and Leibler, 1951; Jaynes, 1957a; Jaynes, 1957b; Kullback, 1978a) to help with the lack of available data. The two research processes from those areas are Gully Erosion Modelling (Chapters 2 and 3), and Sediment Yield (Chapter 2).

In the second part, we tackle a novel area regarding *Flash Droughts* (Chapter 5), therefore related to "dry extremes". Furthermore, in this second part, we deal with an abundance of data from stations and remote sensing, but a lack of data in some particular variables, such as soil physical characteristics and water content. To deal with such an amount of data, we use statistical computing (Slater et al., 2019) and big data analysis applied to hydrological sciences (Hampton et al., 2013; Chen and Han, 2016).

1.4.1 Linear erosion

Erosion is the natural process of abrasion of rocks that built (and keeps on building) the landscapes we see today. This process can be slow and gradual or fast and destructive, depending on the involved mechanisms (Haan, Barfield, and Hayes, 1994; Garcia, 2008).

Usually, erosion processes are divided into two big groups. (1) The laminar (or inter-rill) erosion, that occurs in areas with sheet flow (shear stress $\tau \to 0$) and the main source of detachment force is the *raindrop splash*; (2) the linear (rill or gully) erosion, with shear stress as the main source of detachment forces, and occurring due to concentrated flow in channel incisions (Haan, Barfield, and Hayes, 1994). The boundary between rills and gullies is blurred, with multiple definitions, usually based on depth, width or cross-section (Brice, 1966; Imeson and Kwaad, 1980; Poesen, 1993); being the most common the threshold of one square foot (929 cm^2) for the cross-section area (Poesen, Torri, and Van Walleghem, 2011). Rill formation is generally controlled by microrelief caused, for example, by tillage and landforming operations, while gullies are connected to macrotopography and catchment relief (Poesen, Nachtergaele, et al., 2003). After remediation, actions are applied, rills do not show up in the same place and configuration, while (ephemeral) gullies tend to do so (Nortcliff, 1987; Grissinger, 1996; Poesen, Nachtergaele, et al., 2003).

There is a shortage of gully-erosion monitoring, as pointed by Cerdan et al. (2006). They identified the amount of site-year observations of linear to be over 40 times lower than for its counterpart laminar erosion. There is also a lack of publications, with articles focusing on laminar erosion being three to four times more frequent than on linear erosion (Alencar, 2018), despite their relevance as a sediment source (McCool et al., 1987). Gullies alone produce between 10 and 100 ton.ha^{-1}.yr^{-1} in affected areas, besides causing changes to water and sediment connectivities and productivity/biodiversity loss (Avni, 2005; Verstraeten et al., 2006; Poesen, Torri, and Van Walleghem, 2011). Therefore, it is important to comprehend how gullies behave (Poesen, 2018).

Despite their influence in hydro-sedimentological processes on a catchment scale, erosion models overlook gully erosion assessment (Poesen, 2018). Gullies are complex systems with the superposition of multiple processes, as shear stress, head-cutting, jet flow, piping, and wall failure. Therefore, gully erosion is a process with the interaction of many variables, many of which are difficult to model (Bernard et al., 2010; Castillo and Gómez, 2016; Alencar, de Araújo, and Teixeira, 2020). Therefore, no model has ever been proposed to explain the governing forces controlling gully initiation and growing (Bennett and Wells, 2019). In Figure 1.5, we have some examples of how different gullies can look like. In the first photo (left), we have a gully head in shallow soil, that tends to be wider. In the second (centre), we have a deep incision with a wall almost vertical, stabled by cohesion forces and roots. The last photos (right and bottom) show deep gullies with strong activity in the wall, due to the high concentration of silt. Gullies can vary widely in shape, scale and governing processes (Starkel, 2011).

Nevertheless, some models have been proposed to simulate gully erosion, with two distinct approaches, empirical (e.g, Thompson, 1964; Watson and Laflen, 1986; Woodward, 1999; Nachtergaele, Poesen, Steegen, et al., 2001; Nachtergaele, Poesen, Sidorchuk, et al., 2002; Poesen, Vandekerckhove, et al., 2002; Yao et al., 2008; Wells et al., 2013), and physical models (e.g., Foster and Lane, 1983; Storm, Barfield, and Ormsbee, 1990; Hairsine and Rose, 1992a; Ascough et al., 1997; Sidorchuk, 1999; Dabney et al., 2015; Alonso, Bennett, and Stein, 2002), both with limited success. The empirical methods are restricted to small areas and climatic conditions, being hard to replicate or adapt to other areas. Physical models present better performance and are more flexible, however require multiple calibrations and some variables (as critical shear stress and shear stress distribution) are based on empirical equations themselves, or yet, require so much data that is not feasible to implement (Douglas-Mankin et al., 2020).

More recently, the interest on gully erosion has grown, specially in identification (Marzolff and Poesen, 2009; Agüera-Vega, Carvajal-Ramírez, and Martínez-Carricondo, 2017) and risk and susceptibility assessment (Arabameri, Rezaei, et al., 2018; Arabameri, Pradhan, et al., 2019; Conoscenti, Agnesi, et al., 2018; Bonakdari, Qasem, et al., 2020), using machine learning and artificial neural networks.

In our literature review, we observed some gaps: the key variables of gully erosion modelling (Hairsine and Rose, 1992a; Storm, Barfield, and Ormsbee, 1990; Bennett and Wells,
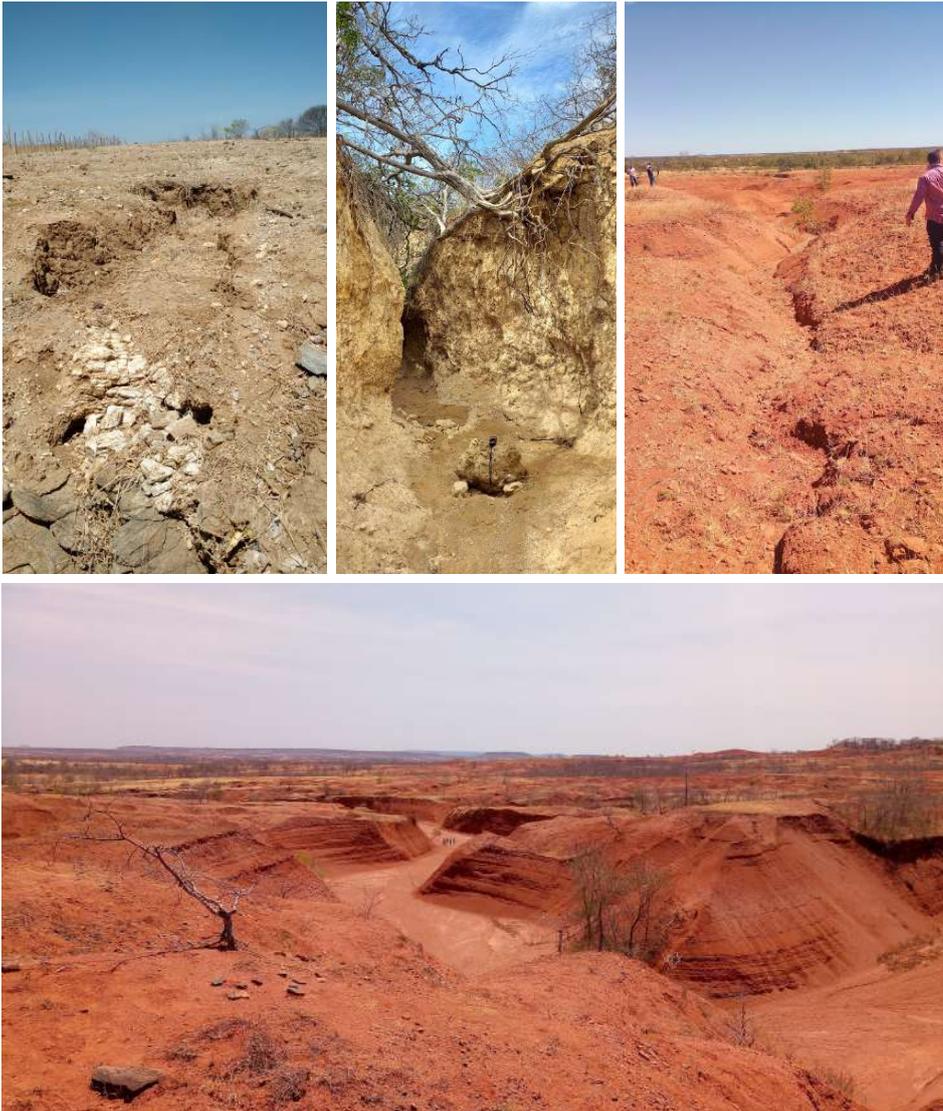
Figure 1.5: The three upper photos row show gullies of different scales in Madalena (Ceará), Campo Formoso (Bahia) and Gilbués (Piauí), from left to right. Lower, the register of a large gully (50-metre wide and 20-metre deep) in Gilbués. Each one, at a different level, disturbs the local hydrological functioning and the land use.

2019) and of a model for gully erosion that can be implemented in multiple temporal and spatial scales (Sidorchuk, 2009; Bennett and Wells, 2019).

1.4.2   Sediment yield and sediment delivery ratio

Erosion and sediment transport lead to siltation, which causes water quality and quantity depletion (de Araújo, 2003; Coelho et al., 2017). It also affects energy production (Landwehr, Schomberg, and Pahl-Wostl, 2021; Zarfl et al., 2015) and increases overspill and consequent floods (Mamede et al., 2012). That is particularly harmful in regions as the Brazilian semiarid, where most water supply is provided via artificial surface reservoirs (Medeiros and Sivapalan, 2020). The Brazilian Semiarid Region has suffered from water scarcity for centuries (Gaiser et al., 2003), with most of the rivers in the region being ephemeral or intermittent, and suffering from frequent and long-lasting droughts (Aragão Araújo, 1990; Brito et al., 2017). Additionally,

the region has its precipitation concentrated in a few months, with rainfall intensities over 100 mm.h^{-1} (Figure 1.6) (de Figueiredo et al., 2016; Alencar, de Araújo, and Teixeira, 2020). Under this scenario, artificial surface reservoirs in the State of Ceará alone accumulate over 22 hm^3 of sediment per year, with a siltation rate of approximately 2.7 t.ha^{-1}.yr^{-1} (de Araújo, 2003) (Figure 1.6).

Figure 1.6: On the left, a small reservoir in Auiaba (Ceará) with a high concentration of sediments, rendering the water undrinkable. On the right, the register of an intense convective rainfall near the municipality of Caridade (Ceará). Such events, with short duration and high intensity and erosivity, are common in the area (author's collection, 2017).

Sediment is a mixture of multiple materials, as primary mineral particles, aggregates, organic matter and chemicals and has a defined life cycle of erosion, transport, and deposition (Haan, Barfield, and Hayes, 1994). All these materials are originated in a catchment and transported by erosive forces throughout the catchment. In the course of transport these materials might settle in sink areas in the catchment or reach a water body (river, lake, reservoir de Araújo, 2007), either being deposited on the bottom of the water body, in a process called siltation (de Araújo, Güntner, and Bronstert, 2006), or being spilt and transported to outside the catchment (Brune, 1953).

The hydraulic transport of sediments is the mechanism of change from potential and kinetic energy from water into morphological alterations in the landscape (Aagaard and Hughes, 2021) and occurs when water moves towards the watershed outlet. The fraction of sediment that leaves the catchment or is trapped in reservoirs is called the catchment's sediment yield (López-Vicente and Guzmán, 2021) and is defined by Equation 1.1.

$$SY = \overline{\varepsilon} \times SDR \tag{1.1}$$

where SY is the sediment yield, $\overline{\varepsilon}$ the total mobilized soil (or gross erosion) and SDR the sediment delivery ratio.

The Sediment Delivery Ratio (SDR) is therefore the ratio of sediment yield and gross erosion (i.e., all mobilized soil by erosive forces in the catchment) (Haan, Barfield, and Hayes, 1994). It can be also interpreted as the probability of a particle of sediment reaching the catchment outlet (de Araújo, 2007). For the sediment to be carried by runoff, it is necessary a quantity of energy (or stream power) (Bagnold, 1966; Hairsine and Rose, 1992b; Beckers, 2018). The available energy for sediment transport, particularly in ephemeral water paths and

on a hillslope scale, are directly dependent on precipitation intensity (de Figueiredo et al., 2016).

There are multiple methods to estimate both $\bar{\varepsilon}$ and SDR. Gross erosion, is often estimated via the Universal Soil Loss Equation (USLE; by Wischmeier and Smith, 1978) and its variations (e.g., the Revised and the Modified Universal Soil Loss Equations – RUSLE and MUSLE Renard et al., 1991; Williams, 1975), that take into account the rainfall energy (function of the 30-minute intensity), as well as topographic conditions, land use and soil properties. To assess the sediment delivery ratio, even though the process is dependent on the same variables (e.g., available energy, land use, and topography), the multiple methods available are often based only on topographical variables (de Araújo, 2003; Maner, 1958; Roehl, 1962; Williams and Berndt, 1972).

There are some strategies to deal with over-siltation in reservoirs, as sediment management (e.g., sediment washing) and sediment control (e.g., upstream infrastructure to capture sediment, as check dams; Peng, Yonggang, and Yongming, 2011; Kondolf et al., 2014). Another successful strategy is sediment reuse (Braga et al., 2019), particularly for small reservoirs that are frequently dry. However, stakeholders still lack the necessary tools to efficiently plan such actions (Landwehr, 2021).

de Araújo (2007) proposed a method to assess an event-based sediment yield model, by deriving a SDR equation that is based not only on the topography of the hillslope but also on the stream power (Bagnold, 1966; Bagnold, 1977) generated by the runoff. The model presents good results for simulations in catchments of multiple scales, however, it requires a large amount of data, rarely available, such as sub-hourly features of rainfall events. Automatic stations that register sub-hourly precipitation with time resolutions ranging from 5 to 30 minutes are expensive to buy and maintain, making the monitoring of such precipitation features hard, particularly in poor or developing countries. Figure 1.7 shows the contrast between the distribution of automatic stations and conventional *Ville de Paris* rain gauges in the Brazilian Semiarid Region, area of study of Chapter 3.

The review of the literature uncovered some gaps, as the lack of an event-based equation for sediment delivery ratio that takes into consideration sub-daily features of precipitation, besides topography and land use conditions.

1.4.3 Droughts and Flash Droughts

Drought is often defined as a long period of shortage of water availability. The primary causes are insufficient precipitation, or excess on evapotranspiration and demand (Bullock, Haddow, and Coppola, 2018). Differing from other natural hazards, droughts have three particularities: (1) its onset and end are difficult to identify because they often accumulate slowly and last even after the shortage end; (2) there is no universally accepted definitions of drought nor what conditions constitute it; and (3) droughts are events that usually affect large, heterogeneous areas and landscapes (Wilhite, 2000; Mishra and Singh, 2010; Bullock, Haddow, and Coppola,
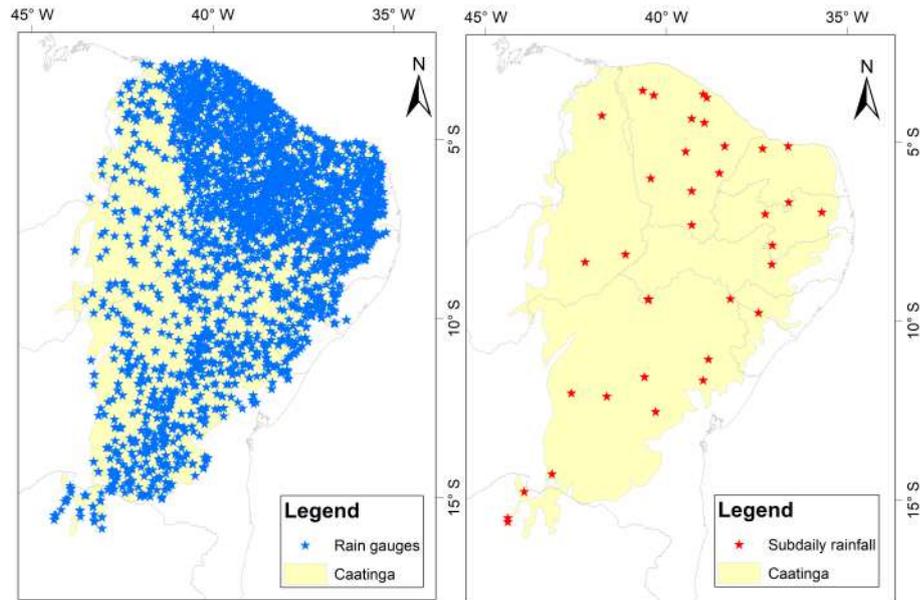
Figure 1.7: Maps of *Ville de Paris* (left) and Automatic (tipping bucket, right) gauging stations in the Caatinga, the Brazilian dry forest located in the semiarid region and managed by the Brazilian National Water Agency (ANA, 2019).

2018). Once drought conditions are established in a region, there is the initiation of a positive feedback loop (Bravar and Kavvas, 1991). Low soil moisture in the upper layers reduces evapotranspiration and relative air humidity, making it harder to reach saturation. Only when external disturbances carry enough moisture from other areas into the drought-affected area, there will be sufficient precipitation to end the drought (Wang and Asefa, 2019).

Some other classical definitions of drought are:

- **Linsley (1959):** a sustained period of time without significant rainfall.

- **Palmer (1965):** a significant deviation from the normal hydrologic conditions of an area

- **World Meteorological Organization (1986):** a sustained, extended deficiency in precipitation.

- **UN (1994):** the naturally occurring phenomenon that exists when precipitation has been significantly below normal recorded levels, causing serious hydrological imbalances that adversely affect land resource production systems.

- **Schneider, Root, and Mastrandrea (2011):** an extended period – a season, a year, or several years – of deficient rainfall relative to the statistical multi-year mean for a region.

These slow-evolving, long-lasting, widespread droughts, henceforth addressed as conventional droughts, are also classified into four categories: Meteorological, Hydrological, Agricultural and Socio-economical droughts. Each classification indicates what kind of effect the drought event caused, its temporal extent and cause (Mishra and Singh, 2010).

Another kind of drought that has gained attention from hydrologists and climatologists is the Flash Drought (Otkin, Anderson, et al., 2013; Otkin, Svoboda, et al., 2018; Pendergrass et al., 2020). This kind of drought, differently from its counterpart (classical droughts) has a short duration (Mo and Lettenmaier, 2016), rapid intensification (Christian, Basara, Otkin, Hunt, et al., 2019), and can affect both large and small areas (Li, Wang, et al., 2020).

Flash Droughts were firstly documented by Peters et al. (2002), and the first publication dedicated exclusively to the topic only in 2013 (Otkin, Anderson, et al., 2013). Flash droughts have drawn more attention from climatologists and hydrologists (Lisonbee, Woloszyn, and Skumanich, 2021), particularly after exceptional events in the United States of America (Basara et al., 2019; Christian, Basara, Otkin, and Hunt, 2019; Otkin, Zhong, Hunt, Basara, et al., 2019), Russia (Christian, Basara, Hunt, et al., 2020), India (Mahto and Mishra, 2020), and Spain (Noguera, Domınguez-Castro, and Vicente-Serrano, 2021). Despite the interest from the scientific community, particularly in the last five years, there is no consensus of what should be a definition of flash drought and how to identify an event.

There are multiple attempts to define flash droughts (Mo and Lettenmaier, 2016; Ford and Labosier, 2017; Christian, Basara, Otkin, Hunt, et al., 2019; Pendergrass et al., 2020; Noguera, Castro, and Serrano, 2020; Li, Wang, et al., 2020; Osman et al., 2021), however, no definition has been universally accepted.

We understand Flash Droughts as a process of rapid (accelerated) and unusually large depletion of soil moisture in comparison with "average" conditions *of the growing season* and land-use related water demand, due to the simultaneous or concurrent occurrence of two or more atmospheric and/or weather conditions over a short time frame of several weeks. When combined, these conditions should lead to an extreme (rare) and dry setup when compared to their expected values for the period of the year. Furthermore, flash droughts may or may not have a direct obvious impact on biota or society, and may become a long-term drought.

Being a new research field, there is not much literature on the topic as of today (Otkin, Svoboda, et al., 2018), but the interest has grown. Lisonbee, Woloszyn, and Skumanich (2021) identified 20 papers with definitions of flash droughts and methods to identify them. The papers, however, present elemental differences, e.g., if flash droughts are rapid-onset events (11 papers), short-term droughts (9 papers) or both (1 paper). In Figure 1.8 we present a bibliometric network[2] of the papers indexed on the Web of Science™. We can observe some trends in the interest for flash droughts, with the initial paper focusing on terms as *rapid onset* (a basic feature of flash drought), *comparison* (to conventional droughts), and the *US drought monitor*. More recent papers are interested in frequency, impact and trend.

Despite the multiple definitions, even when looking exclusively at those with rapid-onset, authors showed a substantial difference (Osman et al., 2021) on flash drought identification. Nevertheless, those definitions share some common features:

---

[2]Network built with assistance of VOSviewer and information extracted from the Web of Science™ on 10.10.2020.
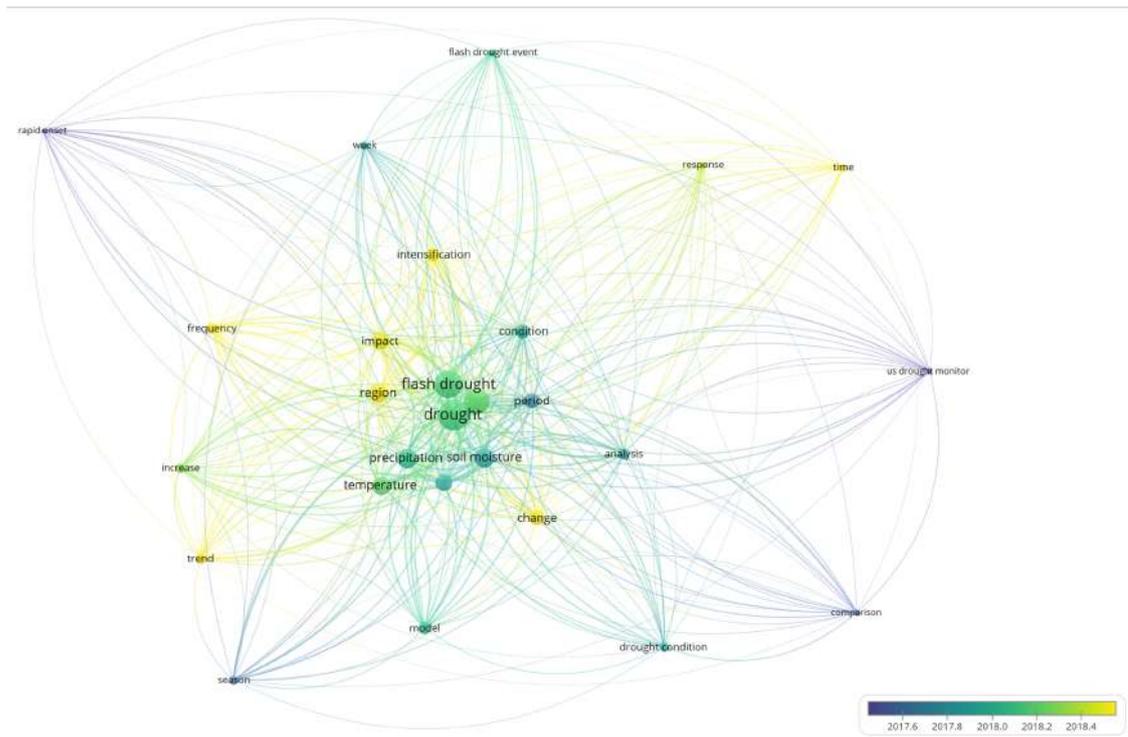
Figure 1.8: Bibliometric network of 81 papers indexed on the Web of Science™containing "flash drought" in their topic (title, keywords, or abstract). The network shows the most common related terms and how they connect. The larger the circle and font size of the term, the more frequent it is. the colour scale indicates the average year of publications containing this term.

- Flash droughts evolve rapidly, with an intensification period lasting a few weeks;

- The final condition should lean towards extreme hot/dry conditions (high temperature, or evapotranspiration, low precipitation or soil moisture, or etc.);

- Flash droughts are seasonal processes and relative to expected values. The weather should be in an extreme hot/dry condition compared to usual values for that time of the year;

- Flash drought is a threshold process that is correctly identified if, and only if, environmental conditions meet a set of predefined rules.

The differences in identification might come from the different variables, statistics, indexes and intervals selected by each method (Lisonbee, Woloszyn, and Skumanich, 2021). The most common variables used to define and identify flash drought are precipitation, temperature, evapotranspiration (potential and actual), and soil moisture. Authors use either combinations of multiple variables (Mo and Lettenmaier, 2015; Mo and Lettenmaier, 2016; Zhang, You, et al., 2017; Hunt et al., 2009; Zhang, Wu, and Hu, 2019; Zhang, Wu, Yeh, et al., 2020; Ran et al., 2020), tracking a single variable (Osman et al., 2021; Ford, McRoberts, et al., 2015; Ford and Labosier, 2017; Yuan et al., 2019; Li, Wang, et al., 2020; Liu, Zhu, et al., 2020; Zhang, Chen, et al., 2019), or adapted indexes from the literature (Pendergrass et al., 2020; Noguera, Castro, and Serrano, 2020; Noguera, Domınguez-Castro, and Vicente-Serrano, 2021; Christian, Basara, Otkin, Hunt, et al., 2019; Christian, Basara, Otkin, and Hunt, 2019; Christian, Basara,

Hunt, et al., 2020; Otkin, Anderson, et al., 2013; Nguyen, Wheeler, Otkin, et al., 2019). The data is usually summarized in short time steps of pentads (73 per year) or weeks (52 per year) and data is compared against the values for each variable/index for the same interval of previous years. Some exceptions to that rule are Noguera, Castro, and Serrano (2020) that divides each month into 4 irregular intervals ranging from 7 to 9 days (with a total of 48 intervals per year) and Osman et al. (2021) that compares the variable value to the values of the same year, using an approach called *volatility*, borrowed from stock market analysis.

1.4.4 Entropy Theory

A relevant novelty of our project is that the designed equations should contain being not only physical but also a statistical basis. The statistical approach is based on the Entropy Theory (Shannon, 1948), a branch of the Information Theory (Kullback and Leibler, 1951; Kullback, 1978a). In Information Theory, the entropy indicates the level of knowledge (or uncertainty) about a process or event (Singh, 2013).

Established by Jaynes (1957a) and Jaynes (1957b), the Principle of Maximum Entropy (POME) allows selecting carefully the probability density function (pdf), which does not admit unproven hypotheses (pdf*). The pdf* is obtained by maximizing the entropy function of the Information Theory, subject to the constraints which, in this case, represent the prior knowledge of the modelled system. According to the Principle of Maximum Entropy, the function pdf* is the one that should be used in physical formulations. The principle has been used successfully in several areas of knowledge, including Hydrology, Hydraulics and Sedimentology (de Araújo and Chaudhry, 1998; de Araújo, 2007; Ghoshal, Kumbhakar, and Singh, 2018; Chiu and Tung, 2002; Chiu, 1987; Chiu, 1988; Singh, 2014a; Singh, 2014b; Singh, 2018; Bonakdari, Sheikh, and Tooshmalani, 2014a; Cui, Sivakumar, and Singh, 2018). The entropy of Shannon is described in Equation 1.2.

$$H = -\int_{\Omega} f(x) \ln f(x) \; dx \tag{1.2}$$

where H is the information entropy and $f(x)$ the function we are looking for, that maximizes entropy H. Ω is the domain of function $f(x)$. This maximization problem can be solved algebraically with the method of the Lagrange Multipliers (Eq. 1.3) using up to three constraints (de Araújo, 2007).

$$\mathfrak{L} = -\int_{\Omega} f(x) \ln f(x) \; dx \; - (\lambda_0 - 1) \left[\int_{\Omega} f(x) dx \right] \; - \sum_{u=1}^{n} \lambda_i \left[\int_{\Omega} f(x) g_i(x) dx - c_i \right] \tag{1.3}$$

λ_i are the Lagrange multipliers. C_i and $g_r(x)$ are t mathematical expression of the constraints (Eq. 1.4). The constraints are previous knowledge that the modeller has on the process or function. They are often statistical moments (e.g., mean and standard deviation) or constitutive

equations (e.g., energy and mass balance). Because $f(x)$ is a probability distribution function, there is always a trivial constraint that $\int_\Omega f(x)dx = 1$.

$$c_i = \int_\Omega g_i(x)f(x)dx, \;\; i = 1, \; 2, \; ..., \; n \tag{1.4}$$

The Principle of Minimum Cross-Entropy (POMCE) is based on the relative entropy of information proposed by Kullback (1978a) and is a generalization of the Principle of Maximum Entropy of Jaynes (1957a). Both principles use all information available, however, the Principle of Minimum Cross-Entropy minimizes the relative entropy against a *prior* function (Eq. 1.5). The *prior* function is, ultimately, a new constraint, representing the previous knowledge or educated guess from the modeller about the expected behaviour of the variable or process. When the modeller has no knowledge, a natural decision is to use a uniform distribution as the prior, which presents the same result as the POME (Kumbhakar, Ghoshal, and Singh, 2019). Therefore, the POME can be understood as a particular case of the POMCE, when the *prior* function is a uniform probability distribution[3]. The minimization problem can be solved again with the method of Lagrangian multipliers, in the form of Equation 1.6.

$$D(f,q) = \int_\Omega f(x) \left[\ln \frac{f(x)}{q(x)}\right] dx \tag{1.5}$$

$$\mathfrak{L} = \int_\Omega p(x) \ln \frac{p(x)}{q(x)} \, dx \; + \; \lambda_0 \left[\int_\Omega p(x)dx - 1\right] \; + \; \sum_{i=1}^{n} \lambda_i \left[\int_\Omega g_i(x)\,f(x)dx - c_i\right] \tag{1.6}$$

1.5 Research questions and objectives

From the literature review, four gaps concerning our focus subjects were identified and selected. The gaps led to the proposition of five questions listed below.

1. What are the key variables of erosion by gullies and how to model such processes?

2. How can we introduce relevant variables as precipitation patterns and features to assess sediment yield and delivery ratio in ungauged basins?

3. What is the definition of Flash Droughts, and what are their key variables?

4. How can flash drought events be identified?

5. What are the available tools from Information Theory and Data Science to help hydrologists to solve problems related to extreme event modelling and/or its impacts?

---

[3]Although the literature does not mention, functions with the form $f(x) = ae^{bx} + c \; \forall \; (a,b,c) \in \mathbb{R}$ behave the same, due to the general solution of the Lagrangian function.

Based on those questions, six objectives were drawn:

- Using field experiments and modelling, identify the key variables to model long term gully erosion and build a framework that allows the addition of multiple sources of energy and sediment;

- Using the principle of minimum cross-entropy, model the key variables related to gully erosion, namely shear stress in walls and bed;

- Using the principle of maximum entropy, propose a methodology to assess sub-daily precipitation features (duration and 30-minute intensity) that allow SDR and sediment yield estimations in ungauged catchments;

- Propose a new definition of flash droughts and a novel method, based on multiple variables and thresholds, to identify them;

- Compare methods from the literature and their performance on identifying events in Central Europe, building a visual platform to easily observe discrepancies between methods;

- Implement all selected methods in R-language and build an open-source package to be available to the community.

1.6   Data structure

To accomplish our objectives presented in Section 1.5 we use multiple data sources and formats. All tabular data and code are publicly available[4]. Imagery, which comprises mainly of image acquisition with UAV (Unmanned Aerial Vehicles, or drones) is typically too heavy for such public online directories.

During the doctorate work, all data was handled and backed up periodically (every one to three months) with the conventional 3-2-1 scheme (three copies, two different media, one out of the office)[5]. All data used in the elaboration of our work is registered as public and/or under CC BY 4.0, which allows copy, reuse, modification, and redistribution, provided reference of the original creator.

For Chapters 2 to 4, precipitation data from conventional and automatic rain gauges provided by the Brazilian National Water Agency and research groups were used to build time series of precipitations (See Figure 1.7). Soil samples were collected and analysed for the obtention of soil properties, as erodibility, bulk density, and critical shear stress. This data was used to assess siltation in reservoirs and validate the gully erosion models. Imagery from UAV was acquired under the proper regulations from each region (Ceará, Bahia, and Piauí – all

---

[4]Please refer to my GitHub page.

[5]One in the computer's hard drive, one in the cloud and a second hard drive kept out of the office.

of them, Brazilian Federal States). Structure from Motion (James, Robson, et al., 2017) was used to generate Digital Surface Models and topography from reservoirs and gully systems.

For Chapter 5, data from the FLUXNET2015 monitoring stations were used. FLUXNET2015 data[6] is also scarce and not well distributed. Out of the 212 stations, only 160 have soil water content measurements (only three in Latin America, and four in Africa), and yet with an average duration of 6.6 years (range from 1 to 18 years). In total, only 47 stations (or 22% of all stations) have a data series for soil water with 10 years or more.

1.7   Dissertation structure and articles overview

This dissertation comprises six chapters. Chapter 1 is the Introduction. Chapters 2 to 5 are articles published or submitted, and follow the order of the objectives presented in Section 1.5. Research Question 1, related to gully erosion, resulted in two research papers, a first that presents an physically-based model and key variables (Chapter 2) and a second with a entropy-based approach (Chapter 3). Research Question 2, related to sediment yield, resulted in one paper (Chapter 4). Research Questions 3 and 4 resulted in a paper (Chapter 5), as well as in the elaboration of an R-package and interactive visualization tool. In Chapter 6, we present a short technical description of the flash drought visualization tool, prepared in the context of the research on flash drought definitions. Finally, Chapter 7 presents the conclusions and outlook of the dissertation. All manuscripts have been drafted by Pedro Alencar. All figures, tables, and calculations were also produced by the author except when otherwise noted.

Chapter 2, page 21:

Title: Physically based model for gully simulation: application to the Brazilian semiarid region

Authors: Pedro Henrique Lima Alencar[1,2], José Carlos de Araújo[2], and Adunias dos Santos Teixeira[2]

[1] Institute of Ecology, Technical University of Berlin

[2] Department of Agricultural Engineering, Federal University of Ceará

Journal: Hydrology and Earth System Sciences (Hydrol. Earth Syst. Sci.)

Year and status: 2020, published

DOI: https://doi.org/10.5194/hess-24-4239-2020

P. Alencar contributed with fieldwork, programming, and laboratory analysis. A. S. Teixeira carried out image acquisition and processing. J. C. de Araújo acted as supervisor

---

[6]Data available at FLUXNET2015 Dataset website (Pastorello et al., 2020).

of the work and in its conceptualization. All authors collaborated in the preparation of the manuscript.

Chapter 3, page 45:

Title: Entropy-based Model for Gully Erosion – a Combination of Probabilistic and Deterministic Components

Authors: Pedro Henrique Lima Alencar[1,2], Antonio Alisson Fernandes Simplício[3], and José Carlos de Araújo[2]

[1] Institute of Ecology, Technical University of Berlin

[2] Department of Agricultural Engineering, Federal University of Ceará

[3] Federal Institute of Science, Technology and Education of Maranhão

P. Alencar contributed with fieldwork, programming, image acquisition and laboratory analysis. A. Simplício carried out image acquisition and processing. J. C. de Araújo acted as supervisor of the work and in its conceptualization. All authors collaborated in the preparation of the manuscript.

Chapter 4, page 67:

Title: Entropy-Based Temporal Downscaling of Precipitation as Tool for Sediment Delivery Ratio Assessment

Authors: Pedro Henrique Lima Alencar[1,2], Eva Nora Paton[1] and José Carlos de Araújo[2]

[1] Institute of Ecology, Technical University of Berlin

[2] Department of Agricultural Engineering, Federal University of Ceará

For this paper, P. Alencar contributed with programming, data processing, analysis, and visualization. J. C. de Araújo contributed with programming and data processing, and E. N. Paton with data analysis. All authors collaborated in the preparation of the manuscript.

Chapter 5, page 87:

Title: How do we identify flash droughts? A case study in Central European Croplands

Authors: Pedro Henrique Lima Alencar[1,2], Eva Nora Paton[1]

[1] Institute of Ecology, Technical University of Berlin

[2] Department of Agricultural Engineering, Federal University of Ceará

Journal: Hydrology Research (Hydrol. Res.)

Year and status: 2021, submitted

P. Alencar contributed programming, data preparation, analysis and visualization. E. N. Paton contributed with data analysis and supervision. Both authors collaborated in the preparation of the manuscript.

2

Physically-based model for gully simulation: Application to the Brazilian Semiarid Region

Abstract

Gullies lead to land degradation and desertification as well as increasing environmental and societal threats, especially in arid and semiarid regions. Despite this fact, there is a lack of research initiatives. As an effort to better understand soil loss in these systems, we studied small permanent gullies, a recurrent problem in the Brazilian Northeastern semiarid region. The increase of sediment connectivity and reduction of soil moisture, among other deleterious consequences, endanger this desertification-prone region and reduce its capacity to support life and economic activities. Thus, we propose a model to simulate gully-erosion dynamics, derived from the previous physically-based models by Foster and Lane and by Sidorchuk. The models were adapted so as to simulate long-term erosion. A threshold area shows the scale dependency of gully erosion internal processes (bed scouring and wall erosion). To validate the model, we used three gullies ageing over six decades in an agricultural basin in the State of Ceará. The geometry of the channels was assessed using Unmanned Aerial Vehicle and Structure-from-Motion technique. Laboratory analyses to obtain soil properties were performed. Local and regional rainfall data were gauged to obtain sub-daily rainfall intensities. The threshold value (cross-section area of 2 m^2) characterises when erosion in the walls, due

to loss of stability, becomes more significant than sediment detachment in the wet perimeter. The 30-minute intensity can be used when no complete hydrographs from the rainfalls are available. Our model could satisfactorily simulate the gully-channel cross-section area growth over time, yielding Nash-Sutcliffe efficiency of 0.85 and R^2 of 0.94.

2.1   Introduction

On our way to sustainable development and environmental conservation, soil erosion by water was pointed out as a key problem to be faced in the 21st century (Borrelli et al., 2017; Poesen, 2018). The impact of water-driven soil erosion, on the economy and food supply alone, represents an annual loss of US\$ 8 to 40 billion, a cut in food production of 33.7 million tonnes and an increase in water usage by 48 km[3]. These effects are felt more severely in countries like Brazil, China and India as well as low-income households worldwide (Nkonya et al., 2016; Sartori et al., 2019). Estimates on total investments to mitigate land-degradation effects on site (e.g. productivity losses) and their off-site effects (e.g. biodiversity losses, water body siltation) lead to more alarming values, averaging US\$ 400 billion yr[-1] (Nkonya et al., 2016). Nonetheless, these values were obtained by studies of soil loss using USLE (Universal Soil Loss Equation) or similar methods, with none considering gully erosion. Thus, the real economical and social impacts of soil erosion are not completely comprehended as long as we do not better understand gully erosion and how to model it.

Notwithstanding, soil degradation had already been an issue since the early 20th-century, having been, for instance, reported by the USDA and the National Conservation Congress, with over 44 thousand km[2] of abandoned land due to intense erosion. By the end of the 1930s, this number had increased to over two hundred thousand km[2] (Montgomery, 2007). Among soil erosion mechanisms, gully erosion plays a relevant role in sedimentological processes in watersheds since it frequently is the major source of sediment displacement (Vanmaercke et al., 2016). Ireland, Sharpe, and Eargle (1939) observed the effect of intense land-use change on gully formation early, mainly due to changes of land-cover and flow path direction. These landscape modifications were connected to runoff acceleration and/or concentration, therefore, triggering gullies.

Gully erosion consists of a process that erodes one (or a system of) channel(s) that starts mainly due to the concentration of surface water discharge erosion during intense rainfall events (Bernard et al., 2010). The concentrated flow causes a deep topsoil incision and may reach the groundwater table and sustain the process (Starkel, 2011). Gullies are a threshold-controlled process (Conoscenti and Rotigliano, 2020) and their initiation is connected to anthropogenic landscape modifications as well as land use and land cover changes, as observed in the other tropical biomes (Katz, Daniels, and Ryan, 2014; Hunke et al., 2015; Poesen, 2018). On the other hand, the presence of vegetation may prevent soil erodibility both by increasing cohesion forces and enhancing soil structure (Vannoppen et al., 2017). Maetens et al. (2012) suggested that land-use changes lead to runoff changes and, hence, directly affect erosive processes. Gully erosion can also be affected by climate change, e.g., an increase in rainfall intensity could lead

to higher erosive potential (de Figueiredo et al., 2016; Panagos, Ballabio, et al., 2017). Gullies are strongly dependent on landscape factors. With the advance of machine-learning techniques and the use of large data sets, some of the factors that mostly influence gully formation were identified, such as lithology, land use and slope. Some indexes were also pointed out as relevant to indicate gully initiation, such as the Normalized Difference Vegetation Index, Topography Wetness Index and Stream Power Index (Arabameri, Pradhan, et al., 2019; Azareh et al., 2019; Conoscenti and Rotigliano, 2020).

Gullies play a relevant role in the connectivity of catchments (Verstraeten et al., 2006), allowing more sediment to reach water bodies and, thus, increase siltation (de Araújo, Güntner, and Bronstert, 2006). For being particularly relevant among the erosion processes, gullies execute a great pressure on landscape development: they change the water-table height, alter sediment dynamics and increase runoff (Valentin, Poesen, and Li, 2005; Poesen, 2018; Yibeltal et al., 2019). They represent an increasing risk to society and environment for affecting land productivity, water supply, floods, debris flow and landslides (Liu, Tang, et al., 2016; Wei et al., 2018). Gullies also have a large impact on the economy due to high mitigation costs, a reduction of arable fields, a decrease of groundwater storage, an increase of water and sediment connectivity and more intense reservoir siltation (Verstraeten et al., 2006; Pinheiro, Metselaar, et al., 2016). The assessment of gully impacts on production costs in an arid region of Israel showed that costs of gully mitigation represent over 5% of total investments, and production losses are as large as 37 % (Valentin, Poesen, and Li, 2005).

The State of Ceará, located in the semiarid region, has its total area (over 148,000 km^2) included in the risk zone of desertification. From this total, about 11.5% is also under advanced land degradation conditions, including the formation of Badlands and Gullies, a similar condition to the one found in other desertification hotspots in the semiarid (Mutti et al., 2020). The region is also especially vulnerable to climate change (Gaiser et al., 2003), and both degradation and desertification can be accelerated by gullies (Zweig et al., 2018). The Brazilian semiarid region is also characterised by shallow crystalline rock bed with scarce groundwater and baseflow, which makes its population rely almost exclusively upon superficial reservoirs for water supply (Coelho et al., 2017). Therefore, gullies are a two-way threat, first, by depleting the already scarce groundwater and second, by increasing sediment connectivity, causing siltation and resulting in loss of storage capacity and water quality (Verstraeten et al., 2006).

Despite their relevance to hydro-sedimentological processes, gullies are often neglected in models (Poesen, 2018), and should be directly addressed (Paton et al., 2019). However, gully erosion is a process with the interaction of many variables, with several of them difficult to assess (Bernard et al., 2010; Castillo and Gómez, 2016). According to Bennett and Wells (2019), for instance, no model has ever been presented to clearly explain the process of gully formation. Among the models that do consider gully erosion, the use of empirical approach prevails (e.g. Thompson, 1964; Woodward, 1999; Nachtergaele, Poesen, Sidorchuk, et al., 2002; Wells et al., 2013); whereas others focus primarily on physically-based algorithms (e.g. Foster and Lane, 1983; Hairsine and Rose, 1992b; Sidorchuk, 1999; Dabney et al., 2015).

It is, therefore, an important milestone to understand how gully erosion starts and develops (Poesen, 2018). The objective of this work is to propose a physically-based model that predicts growing dynamics and sediment production in small permanent gullies on a hillslope scale. In order to achieve this, we tested two models – Foster and Lane (1983) and Sidorchuk (1999) – and two adapted models, one being the modification of the model of Foster and Lane and the other the coupling of both models. To validate the model, we selected three small permanent gullies in the State of Ceará. The gullies' geometry was assessed using UAV (Unmanned Aerial Vehicle) and the soils were sampled and characterised.

We understand small permanent gullies to be the result of active erosive processes that form channels by concentrated flow and do not interact with groundwater. Normally, these gullies could be remedied by regular tillage processes, but in abandoned or unclaimed land,they usually remain untreated for long periods. Although the land where they develop is usually unused for economic activities besides livestock grazing in open range, the development of such gullies threatens the ecosystem and community.

2.2 Materials and methods

2.2.1 Study area

The Brazilian Semiarid Region (1 million km^2) is mainly covered by the Caatinga biome with vegetation characterised by bushes and broadleaf deciduous trees (Pinheiro, Costa, and Araújo, 2013). The region is prone to droughts and highly vulnerable to water scarcity (Coelho et al., 2017). More than 25 million people live in this region where agriculture (maize, beans, cotton) and livestock in the open range are of utmost socio-economic relevance. Usually, rural communities use deleterious practices, such as harrowing and field burning, which enhance the risk of intense erosive processes. These characteristics lead to a scenario of soil erosion and water scarcity with high social, economic and environmental consequences (Sena et al., 2014). Erosion in general (and gullies in particular) increases local water supply vulnerability due to reservoir siltation (de Araújo, Güntner, and Bronstert, 2006) and water-quality depletion (Coelho et al., 2017).

The study area is located in the Madalena Representative Basin (MRB, 75 km^2, State of Ceará, Northeastern Brazil; see Figure 2.1), located in the Caatinga biome, a dry environment with a semiarid hot BSh climate, according to the Köppen classification. The annual precipitation averages 600 mm, concentrated between January and June (Figure 2.2), and the potential evapotranspiration totals 2,500 mm yr^{-1}. Geologically, the basin is located on top of the crystalline bedrock with shallow soils and limited water storage capacity. The rivers are intermittent and runoff is low, typically ranging from 40 to 60 mm yr^{-1} (Gaiser et al., 2003). The basin is located within a land reform settlement with 20 inhabitants per km^2, whose main economic activities are agriculture (especially maize), livestock and fishing (Coelho et al., 2017; Zhang, Foerster, Medeiros, de Araújo, et al., 2018).
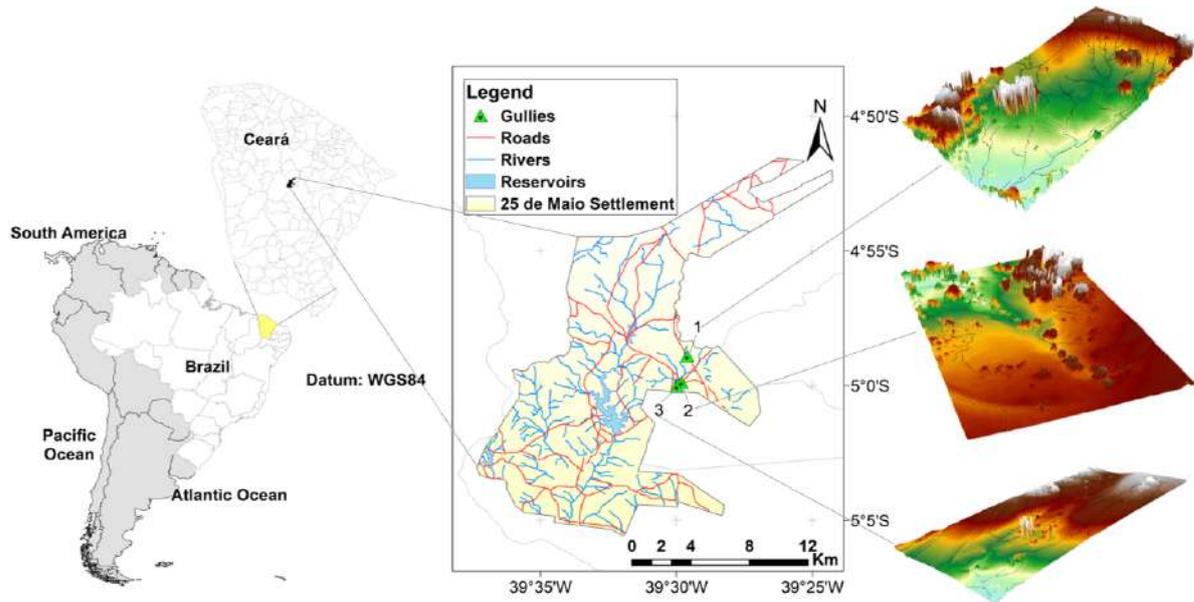
Figure 2.1: Location of the study area and the gully sites (gullies 1, 2 and 3) and the digital elevation models. The roads, rivers and reservoirs were mapped by Silva, Gorayeb, and de Araújo (2015).

Three gullies were selected for this study, all located on the eastern portion of the basin. The studied gullies have the following dimensions (average $\pm$ standard deviation): projection area (317 ± 165 m^2), length (38 ± 6 m), volume (42 ± 25 m^3), depth (0.44 ± 0.25 m) and eroded mass (61 ± 36 Mg). The coordinates are presented in Table 2.1. Despite their small sizes, they possess a significant impact on the landscape for reducing productive areas and soil fertility. According to the information obtained from local villagers, gully erosion started immediately after the construction of a country road in 1958 (Figure 2.3). Before the construction, the sites were covered by Caatinga vegetation (Pinheiro, Costa, and Araújo, 2013). The road modified the natural drainage system and does not provide for any side nor outlet ditches, therefore generating a concentrated runoff at its side. This has caused excessive runoff on the hillslopes and triggered gully erosion.
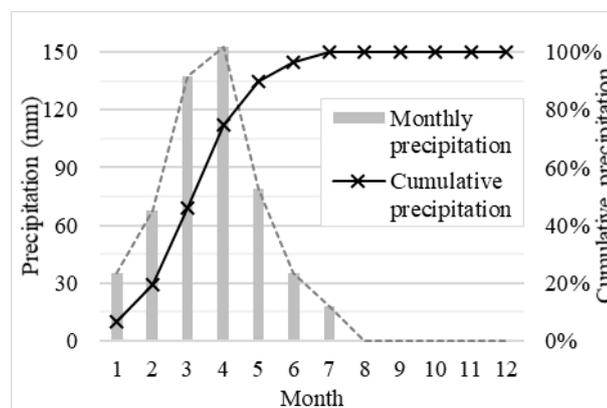
Figure 2.2: Monthly precipitation (median) and cumulative precipitation at MRB from 1958 to 2015.
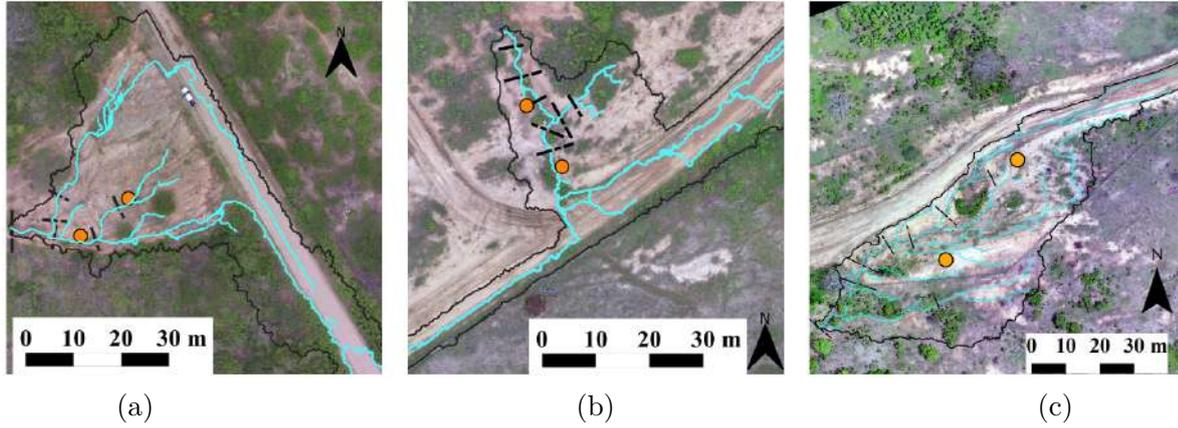
Figure 2.3: Aerial photogrammetry of the studied gullies. Note that they are at the margin of the road, receiving the concentrated flow diverged from it. The continuous black line represents the catchment boundaries; the blue line represents the flow paths; the dashed black lines are the cross-sections used on the validation of the model; and the orange dots are the soil sampling points - (a) gully 1; (b) gully 2; (c) gully 3.

Table 2.1: Coordinates of the three gullies used in this study (Datum: WGS84).

Area	Latitude	Longitude
Gully 1	04°58'54.3"S	39°29'36.4"W
Gully 2	04°59'53.1"S	39°29'49.4"W
Gully 3	05°00'02.4"S	39°29'59.4"W

2.2.2 Topography survey

The assessment of the gully data was achieved using an Unmanned Aerial Vehicle (UAV), a technique applied in other regions as well (Stöcker, Eltner, and Karrasch, 2015; Wang, Zhang, et al., 2016; Agüera-Vega, Carvajal-Ramírez, and Martínez-Carricondo, 2017). A UAV equipped with a 16 MP camera (4000 x 4000 pixels) and field of view of 94 % was used. The flight was at 50 m altitude with a frontal overlap of 80 % and lateral overlap of 60 %. For the geo-reference of the mosaic, five ground control points were deployed which were evenly distributed in each area, both in the high and low ground. The coordinates were collected using a stationary GNSS – RTK (L1/L2) system with centimetre-level accuracy.

The Digital Surface Model was produced using the Structure from Motion technique. This process consists of a three-dimensional reconstruction of the surface, derived from images and the generation of a dense cloud of 3D points based on the matching pixels of different pictures and Ground Control Points (GPCs); the processing result is a model as accurate as one obtained by laser surveys (e.g., Light detection and ranging - LiDAR), but cheaper and less time consuming (Stöcker, Eltner, and Karrasch, 2015; Agüera-Vega, Carvajal-Ramírez, and Martínez-Carricondo, 2017). The ground sample distance (pixel size) obtained is four to five centimetres and the digital models have high precision, with a vertical position error of one centimetre and horizontal error of six millimetres. The vegetation, yet sparse, was an

obstacle to increasing the quality of the survey. However, as the focus of this study was the gully cross-section geometry, vegetation interference was acceptably low.

2.2.3 Soil data

Due to the scale of this experiment and the homogeneity of the soil-vegetation components (Güntner and Bronstert, 2004), we divided the areas into two sets based on grain-size distribution, organic matter and bulk density. Gully 1 (G1) has specific features and comprises the first soil (S1), whereas gullies G2 and G3, close to each other, are represented by the second soil (S2).

At the gully sites, soil surveys were carried out to assess the properties and parameters required to implement the model; undisturbed soil samples were collected (see Figure 2.3) at depths of 0.10 m, 0.30 m and 0.50 m (two sites, three depths, three samples per depth, totalling 18 samples collected). At the depth of 0.50 m, a well-defined horizon C, rich in rocks and soil under formation, was identified. The maximum depth of the non-erodible layer ranged from 60 to 75 cm in all gullies. We performed grain-size distribution, sedimentation, organic matter, bulk density and particle density analysis.

The soils are loamy, with clay content ranging from 6 % to 37 %. The particle density is 2580 kg m^{-3}. The soils are Luvisols and have a typical profile, with the top layer relatively poor in clay when compared to the layers below and with the regular occurrence of gravel at the surface. Furthermore, Luvisols are rich in active clay, which makes them prone to form cracks and macropores when dry (dos Santos et al., 2016), a process also documented in soils with similar texture in a semiarid area in Spain (van Schaik et al., 2014). Rill erodibility values (K_r) and critical shear stress (τ_c) for the soils were obtained using the Equations 2.1 and 2.2 (Alberts, Nearing, et al., 1995) and are also presented in Table 2.2.

$$K_r = 0.00197 + 0.00030\%VFS + 0.038633e^{(-1.84\%OM)} \tag{2.1}$$

$$\tau_c = 2.67 + 0.065\%C - 0.058\%VFS \tag{2.2}$$

where $\%VFS$ is the percentage of very fine sand, $\%C$ is the percentage of clay and $\%OM$ is the percentage of organic matter.

2.2.4 Rainfall data

Daily rainfall data for the location spanning the entire period was provided by the Foundation of Meteorology and Water Resources of Ceará (Funceme). We used five rain-gauge stations in the region, covering the period from 1958 to 2015. The annual rainfall in the area averages 613 mm (Supplementary material - Fig. S1) with a coefficient of variation of 43%, being typical

2. Physically-based model for gully simulation: Application to the Brazilian Semiarid Region

Table 2.2: Grain-size distribution, organic matter for both soils (S1 - for the gully 1 - and S2 - for the gullies 2 and 3) at three depths (10, 30 and 50 centimetres) and the respective texture classification (USDA); and the estimated (in italic) rill erodibility (K_r) and critical shear stress (τ_c of the site soils) obtained using Equations 2.1 and 2.2.

Soil and layer	Gravel > 2 mm	FCS[a] > 0.1 mm	VFS[b] > 0.05 mm	Silt > 0.002 mm	Clay < 0.002 mm	Organic Matter	Bulk density (kg m^{-3})	Soil Class	K_r (s.m^{-1})	τ_c (Pa)
S1-10	13 %	45 %	21 %	11 %	10 %	3.1 %	1699	Sandy Loam	*0.015*	*2.102*
S1-30	6 %	46 %	16 %	14 %	18 %	3.3 %	1677	Sandy Loam	*0.016*	*2.912*
S1-50	4 %	63 %	20 %	7 %	6 %	2.2 %	1765	Loamy Sand	*0.020*	*1.900*
S2-10	17 %	33 %	22 %	11 %	17 %	4.9 %	1509	Sandy Loam	*0.012*	*2.499*
S2-30	8 %	29 %	6 %	20 %	37 %	5.7 %	1572	Clay Loam	*0.011*	*4.611*
S2-50	2 %	28 %	15 %	20 %	25 %	1.4 %	1643	Loam	*0.014*	*3.425*

[a] Fine to Coarse Sand; [b] Very Fine Sand.

values for the Brazilian Semiarid Region (de Araújo and Piedra, 2009). The double mass method was employed to check data consistency (Supplementary material - Fig. S2). The gaps in the measurements (January 1958 and September 1960) were filled by the nearest gauging station.

The modelling of the gullies is based on peak discharge, which demands sub-daily rainfall data, but only daily precipitation was available inside the study basin covering the whole experiment period. To proceed with the modelling, correlation curves relating total daily precipitation and rainfall intensity were used. In order to define which was the best intensity to be used in the modelling, four were tested (Figure 2.4) as input for the model: average (I_{av}), sixty-minute maximum (I_{60}), thirty-minute maximum (I_{30}) and fifteen-minute maximum (I_{15}) intensities.

To build such curves, we used data from the Aiuaba Experimental Basin's (AEB). This basin has been monitored since 2005 (de Figueiredo et al., 2016) and has detailed hydrographs, with 5-minute temporal resolution. This experimental basin is located 190 km south of MRB and both basins are climatically homogeneous (Mendes, 2010). In addition, Figure 2.4 shows the rainfall data for the MRB collected during the rainy season in 2019 (January to July). We can observe that the data has similar behaviour but is constantly on the lower area of the plot. It is relevant to note that in the year of 2019 in MRB it was dry (total annual precipitation of 402 mm, over 30 % lower than the average) and such behaviour was expected.

To obtain discharge values from intensity, we used the SCS-CN method (Chow, 1959). For the models, the main rainfall related variables are the event peak discharge and its respective duration. Because the gully catchment areas were small, their respective concentration time was negligible compared with the intense-rainfall duration in the region (de Figueiredo et al., 2016), yielding a uniform pattern of peak discharge.
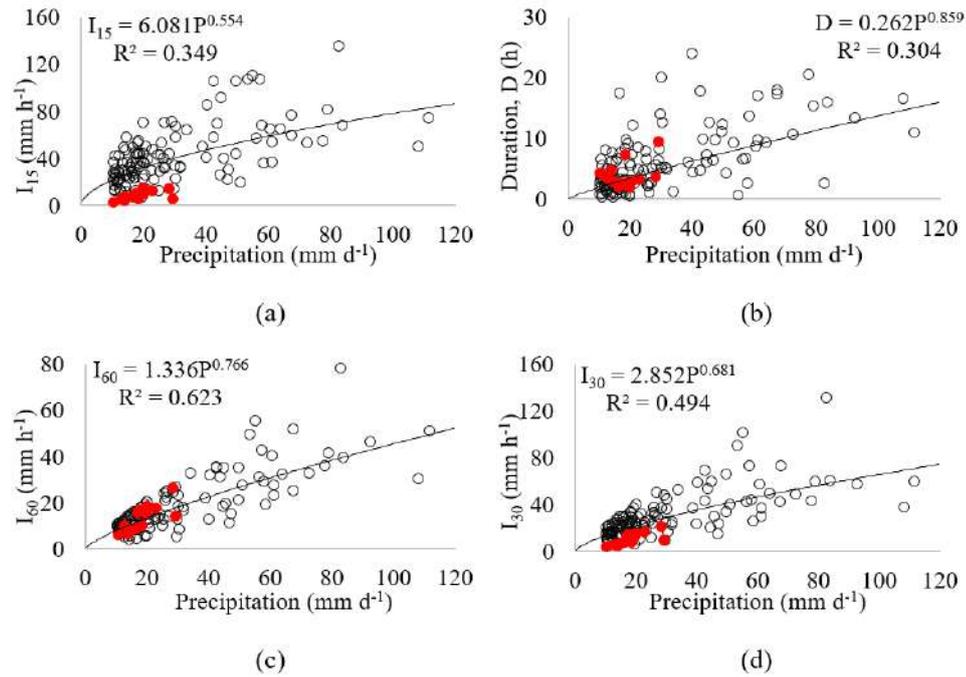
Figure 2.4: Correlation between daily precipitation and sub-daily variables at the Aiuaba Experimental Basin (de Figueiredo et al., 2016). (a) daily precipitation versus event duration (D); (b) daily precipitation versus 60-minute maximum intensity (I_{60}); (c) daily precipitation versus 30-minute maximum intensity (I_{30}); and (d) daily precipitation versus 15-minute maximum intensity (I_{15}). The white circles indicate data obtained in Aiuaba from 2005 to 2014. The red dots indicate precipitations measured in the MRB from January to July 2019 (rainy season).

2.2.5 Gully modelling

To model small permanent gullies, we propose two models based in classical formulations from the literature of Foster and Lane (1983) and Sidorchuk (1999). The Foster and Lane Model (FLM) is one of the most used models of gully erosion based on net shear stress and transport capacity. The FLM assumes a rectangular cross-section and was originally designed for single rainfall ephemeral gully modelling. The Sidorchuk Model (SM) considers the mass balance of sediments, shear stress (in terms of critical velocity), soil cohesion and the Manning equation to estimate the cross-section geometry and channel slope. It also uses empirical equations based on field measurement to estimate the flow depth and width. This model gives special attention to the processes involving gully wall transformation. A description of both models is available in the literature and in the supplement materials of this paper.

The proposed models are the Adapted Foster and Lane Model (FLM-λ) and the coupled model Foster and Lane & Sidorchuk Model (FL-SM). The key difference between the two proposed models is the amount of data required to use each one. The models are presented below.

2.2.5.1   Adapted Foster and Lane Model (FLM-λ)

The Foster and Lane Model (FLM), as proposed by its authors, considers a single source of erosion: the soil detached from the channels bed and walls due to shear stress. Field observation and literature (Blong and Veness, 1982), however, show that wall instability and failure can represent a significant source of sediment. To estimate the effect of wall erosion at the studied site, we proposed an empirical parameter (λ – Eq. 2.3) to correct the effect of lateral flow and wall erosion. This multiplicative parameter was calibrated and validated as a function of the catchment shape based on two coefficients: the Gravelius coefficient (K_G – Eq. 2.4) and the Form coefficient (K_F – Eq. 2.5). Both coefficients describe the geometry of the catchment area and can be interpreted as how compact the distribution of area is. Commonly linked to flood proneness, these parameters also relate to the transversal distance of the catchment area, which influences the amount of lateral inflow into the mainstream.

$$\lambda = \frac{A_o}{A_m} \tag{2.3}$$

$$K_G = 0.28 \cdot \frac{C_P}{C_A} \tag{2.4}$$

$$K_F = \frac{C_A}{C_L^2} \tag{2.5}$$

In Equation 2.3, the terms A_o and A_m are the observed and measured cross-section area and in Equations 2.4 and 2.5, C_P, C_A and C_L stand for the catchment perimeter, area and length, respectively.

The plots of λ versus both parameters are presented in Figure 2.5. Two equations ($\lambda(K_G)$ and $\lambda(K_F)$, see Figure 2.5) were calibrated using data from 14 randomly selected sections out of the 21 assessed by the DEM. The remaining data were used to validate the equations. The model FLM-λ consists of processing the FLM as originally proposed and, afterwards, multiplying the output area by the λ correction parameter. Since $\lambda \geq 1.0$, the multiplication simulates the effect of wall erosion.
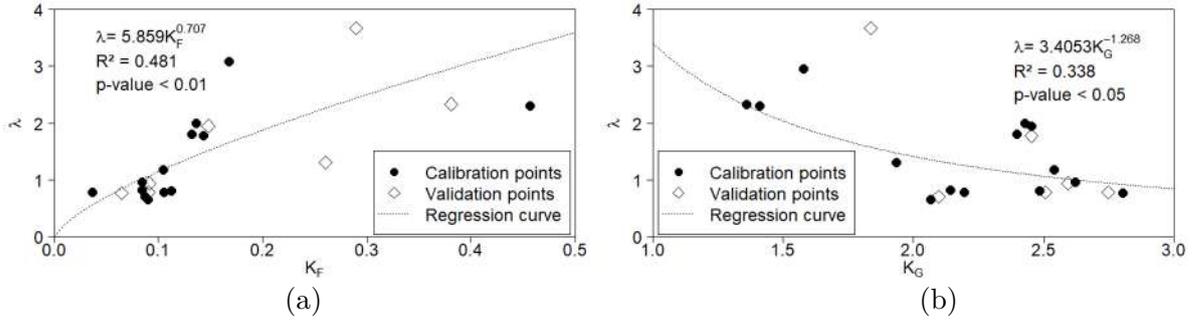
Figure 2.5: Correlations between the ratio ($\lambda = A_o/A_m$) and (a) the Gravelius coefficient (K_G) and (b) the form factor (K_F) for 21 monitored cross-sections at MRB. Black dots refer to calibration cross-sections and white diamonds refer to validation cross-sections. The values of R^2 indicated in the plots are for the calibration. The validation R^2 were 0.10 for K_G and 0.54 for K_F.

$$\lambda = \max\left(5.859 K_F{}^{0.707}; 1.0\right) \tag{2.6}$$

The coefficient (K_F) yielded a positive Nash-Sutcliffe Efficiency value and smaller RMSE (0.17 and 0.67, respectively) which did not occur with the Gravelius coefficient (-2.43 and 0.84, respectively). In the revised model, hereafter addressed as FLM-λ, the FLM area output is multiplied by the calibrated parameter λ (Equation 2.6), yielding the eroded area. Applying this factor caused a significant improvement in model efficiency, with NSE increasing from 0.557 to 0.757. The incremental area produced by the multiplication of λ is assumed to increase the width of the upper half of the cross-section, keeping bottom width and the orthogonality of the walls unchanged.

2.2.5.2 Coupled Model – Foster and Lane & Sidorchuk Model (FL-SM)

The previous model was produced due to the necessity of considering wall failure as a sediment source and used an empirical approach. Another way, using a physically-based concept, is to include the specific routine of the Sidorchuk model that tackles wall failure in the Foster and Lame model. This can be achieved by simply including a test after each event, checking if the depth of the channel causes wall instability given the current angle of the bank.

By analysing the data, however, we identified that for small cross-sections, even after the critical had been reached, the section remained rectangular. This implies an additional threshold mechanism related to the geometry of the channel and/or catchment. Therefore, the FL-SM requires the determination of a threshold value for the implementation of the wall erosion equations. Such a threshold controls when the wall erosion becomes significant for the total amount of eroded soil. In the model, it represents the limit stage, above which Sidorchuk (1999) equations are used. It also represents the scale when solely the channel bed erosion, described by the Foster and Lane (1983) equations, start to consistently underestimate the measured area. In this study, we used the Foster and Lane model to identify this scale where both processes (channel bed and wall erosion) switch relevance.

A flow chart of the FL-SM is presented below in Figure 2.6. The core of the model is the Foster and Lane Model, processed for every runoff event. When the cross-section reaches the threshold condition and satisfies the criteria for wall failure as described in Sidorchuk (1999), a new step is included that calculates wall transformation.
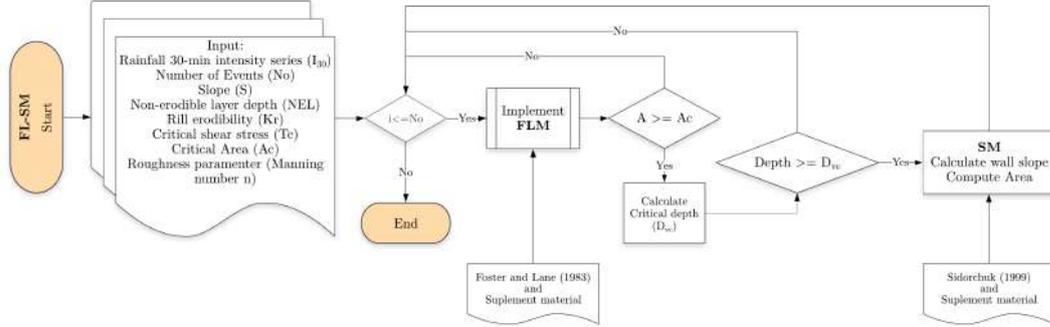
Figure 2.6: Flow chart of the Coupled model FL-SM.

2.2.6 Model fitness evaluators

To assess the goodness-of-fit, the Nash-Sutcliffe Efficiency coefficient (NSE), the root mean squared error (RMSE) and the percent bias (PBIAS) were used (see Moriasi et al., 2007). Additionally, the methodology proposed by Ritter and Muñoz-Carpena (2013) asserts statistical significance to the evaluators. The proposed model is based on Monte Carlo sample techniques to reduce subjectivity when assessing the goodness-of-fit of models.

$$NSE = 1 - \frac{\sum_{i=1}^{n}(X_{o,i} - X_{m,i})^2}{\sum_{i=1}^{n}(\bar{X}_o - X_{o,i})^2} \tag{2.7}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(X_{o,i} - X_{m,i})^2}{n}} \tag{2.8}$$

$$PBIAS = \frac{\sum_{i=1}^{n}(X_{o,i} - X_{m,i})}{\sum_{i=1}^{n}(X_{o,i})} \tag{2.9}$$

In Equations 2.7, 2.8 and 2.9, $X_{o,i}$ is an observation and $X_{m,i}$ a modelled value, with n being the total of observations and simulations. $\bar{X}_o$ is the average of the observed values.
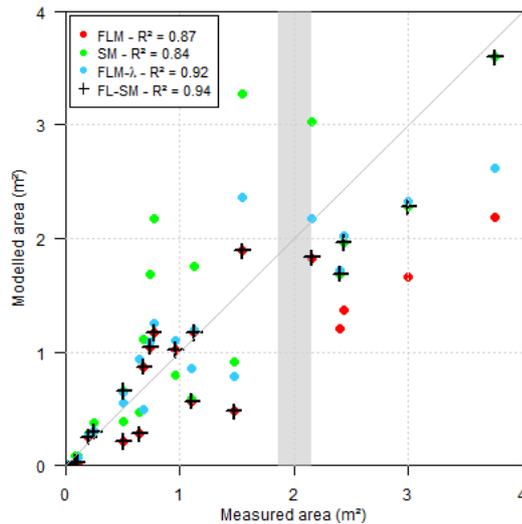
Figure 2.7: Performance of the coupled model (FL-SM), Foster and Lane Model (FLM and FLM-λ) and the Sidorchuk model (SM). p-value < 0.001 for all sets. The grey bar indicates the identified threshold area where there is a change, and SM becomes consistently better than the FLM.

2.3 Results

2.3.1 Gully modelling

From the three gullies measured, twenty-one cross-sections with different dimensions were selected and used to validate and compare the model's quality. Figure 2.7 presents the scatter of modelled and measured data for the models implemented. The FL-SM presented a Nash-Sutcliffe Efficiency coefficient of 0.846 when using a threshold for the area of the cross-section of 2.2 m^2. In Figure 2.8, some output examples for the sections above the threshold are presented.

In terms of geometry, sections models with the Foster and Lane Model with its original formulation present output cross-sections similar to Figure 2.8 [(a), (b), (c)], with rectangular-live shape. When the parameter λ is introduced (FLM-λ), the output cross-sections are modelled with piled rectangles as in Figure 2.8 [(d), (e), (f)]. Using the FL-SM, when the area surpasses the threshold value, sections have mainly trapezoidal geometry, as illustrated in Figure 2.8 [(g), (h), (i)]. It is important to highlight that the model can produce triangular geometry, but none was obtained in this study.

2.3.1.1 Threshold analysis

The interpretation of the threshold for implementation of the wall erosion routine can be based on (a) the cross-section area or (b) the catchment geometry, as illustrated in Figure 2.9. The first approach considers a critical area that once reached, marks when the wall erosion is truly significant with respect to the other processes. In this study, the threshold identified
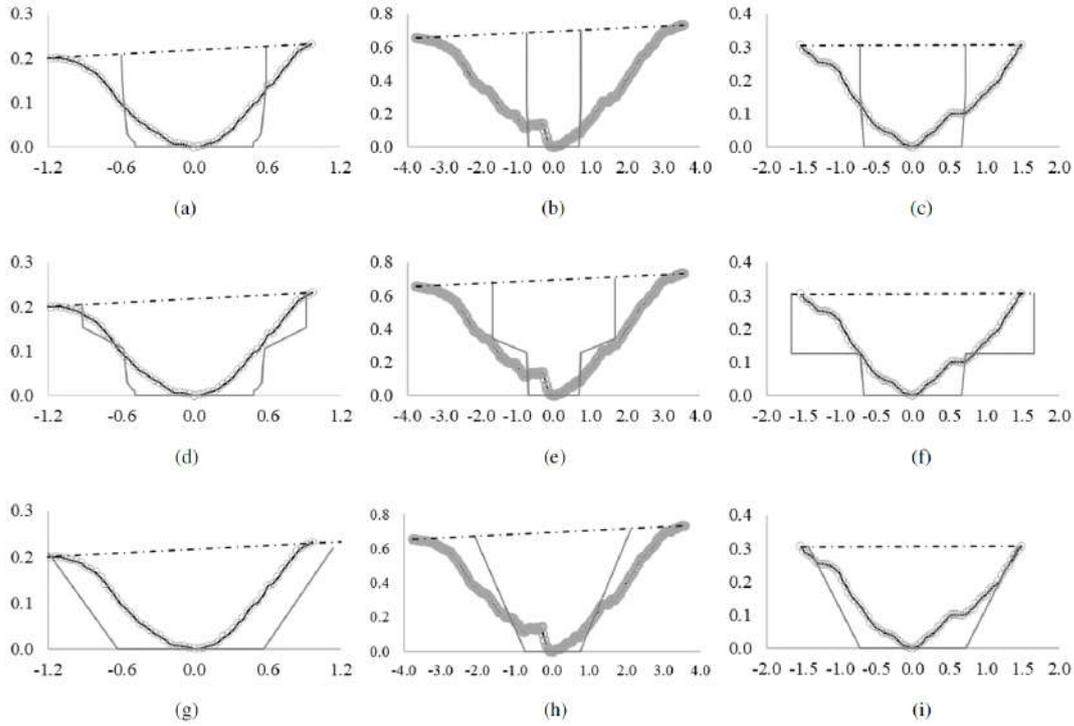
Figure 2.8: Some examples of gully cross-sections measured (black line with circles) and the modelled (dark grey line) geometry. figures (a), (b) and (c) show the output for the model of Foster and Lane; figures (d), (e) and (f) the output for the model of Foster and Lane adapted with the parameter λ and figures (g), (h) and (i) the result from the Sidorchuk Model (SM). Distances in metres. The section in (a, d and g) is a section obtained from gully 1, (b, e and h) from gully 2, and (c, f and i) from gully 3.

was at an area of 2.2 m^2. After that, the model calculates the effect of sidewall erosion and reaches the critical final area for the analysed section. The presence of a threshold for applying the sidewall erosion routine indicates a change of relevance among processes on a given scale. Although the threshold is addressed as an area, this is only a consequence of more complex interactions among discharge, flow erosivity, cohesion and gravitational forces.

The second interpretation is related to the catchment geometry, as the approach given to the parameter λ also related to the K_F. From the distribution of the cross-sections, we can observe sections that are better modelled by SM even below the threshold. By analysing the values of the form coefficient (K_F) of each set (Figure 2.9b), we observed that set 1 has K_F of 0.08 ($\pm$ 0.02) and set 2 has 0.22 ($\pm$ 0.12). Higher values of K_F indicate a more compact catchment, with more lateral flux into the channel, therefore producing more erosion in the soil. By sorting the model results of FLM and SM based on the form coefficient, using the threshold of $K_F = 0.15$, we obtained an NSE of 0.79.

Given the obtained efficiencies, in this study, we adopted a threshold based on the cross-section area. However, the use of a catchment-based threshold should not be discarded and could be promising, since there are reports in literature of the relation between lateral flow and gully erosion (Blong and Veness, 1982).
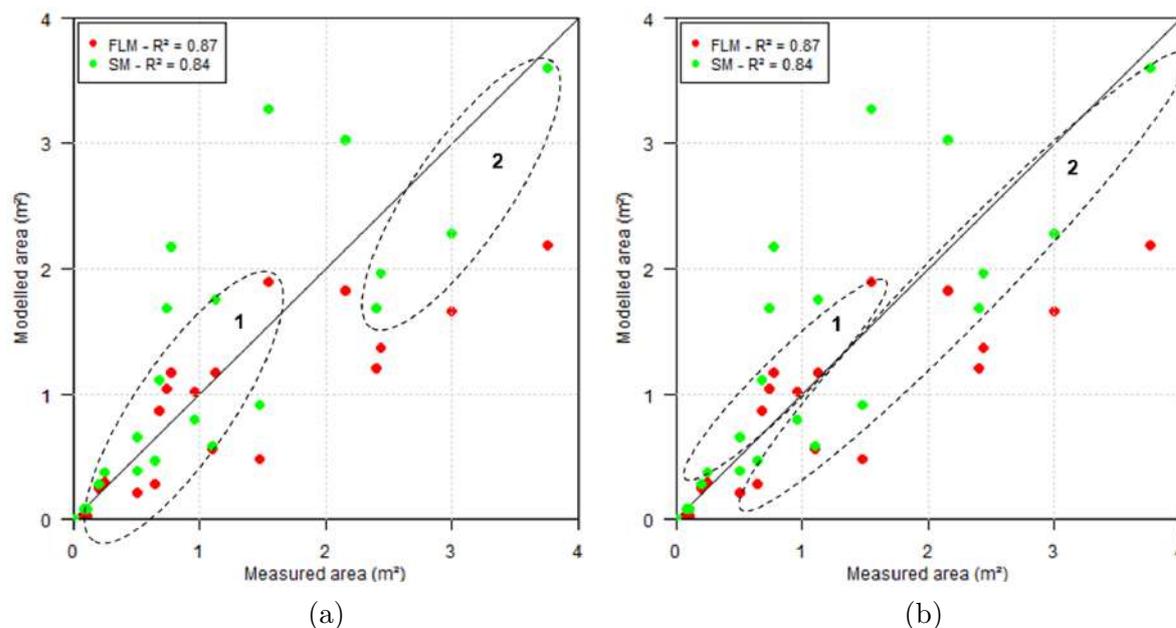
Figure 2.9: Thresholds for wall erosion: (a) based on the cross-section area; (b) based on the catchment geometry and K_F. In both plots, the set 1 indicates the domain of bed erosion and Foster and Lane equations and set 2 indicates the domain of wall erosion and Sidorchuk equations. p-value < 0.001.

2.3.1.2 Rainfall intensity

From the three gullies, twenty-one cross-sections with no interference of bushes or trees were selected from the Digital Elevation Model. The FLM was tested for the 60-minute, 30-minute, 15-minute and average intensities [FLM(I_{60}), FLM(I_{30}), FLM(I_{15}), FLM(I_{av})]. The best response was shown when using the thirty-minute intensity [FLM(I_{30}); NSE = 0.557]. Figure 2.10 presents the plot of the model outputs for the cross-section area compared with the measured data. Moreover, the Foster and Lane Model did not show good responses to the cross-section geometry, regardless of the intensity tested. This may indicate a flaw in the model concerning the side-ward erosive process.

The FLM considers rectangular-shaped cross-sections, but the field survey showed that the sections were rather trapezoidal or triangularly shaped (Figure 2.8). Among the factors that can shape gully walls, others include seepage, angle of internal friction, and the slope angle itself (Sidorchuk, 1999; Bingner et al., 2016). Besides this, gully walls can be shaped by lateral discharge (Blong and Veness, 1982) which depends directly on the morphology of the cross-section catchment area. Figure 2.10 also shows a tendency of the model to underestimate the cross-section area, which implies that the model does not consider all the relevant erosive processes. The sidewall erosion has proven to be a relevant source material, often representing over 50 % of the eroded mass (Crouch, 1987) whereas the FLM only assumes the vertical sidewall morphology. Therefore, in this study, we adopted the 30-minute intensity as the standard intensity and duration to assess peak-discharge.
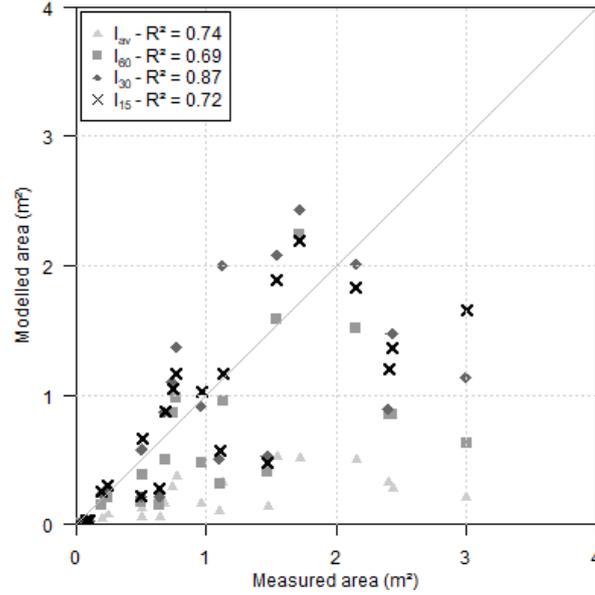
Figure 2.10: Results of the FLM for cross-section area using different intensities (I_{av} – average; I_{60} – 60-min; I_{30} – 30-min; and I_{15} – 15-min) to generate the peak discharge.

2.3.2   Model evaluators

The coupled model, FL-SM, presented the highest performance of goodness-of-fit evaluators (Figure 2.11). The model yielded a PBIAS value below 10 %, which is very good. The coupled-model RMSE was also low (0.397), whereas the NSE reached a value of 0.846, being classified as good (Ritter and Muñoz-Carpena, 2013) or very good (Moriasi et al., 2007).
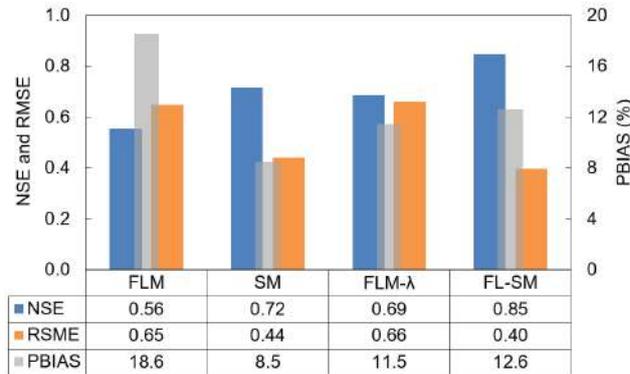
Figure 2.11: Model evaluators *NSE*, *RMSE* and *PBIAS*. The bar plot shows the performance of all tested models – values of *PBIAS* in percentage.

Figure 2.11 shows the evolution of NSE values with more details to allow for conclusions be drawn. The Foster and Lane model with λ parameter (FLM-λ) was calibrated with 14 cross-sections out of 21 and performed as well as the Sidorchuk model (SM), which considers the sidewall effect. For the coupled model, there is no efficiency gain when applying the calibrated parameter (λ) to sections below the threshold which indicates that the lateral inflow is only relevant for larger sections. Figure 2.12 presents the Taylor diagram for comparison of the four models. In this diagram, the closer a model is to the reference (measured data) the better. The FL-SM presented the highest correlation and the lowest RMSE.
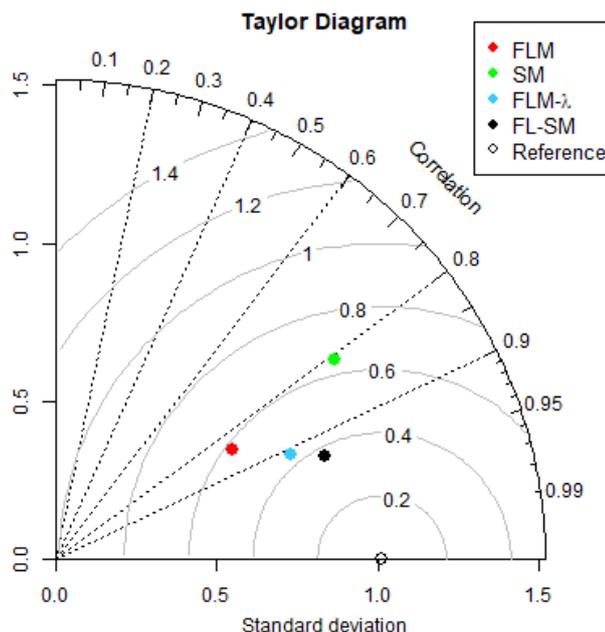
Figure 2.12: Taylor diagram for the model performance. The azimuthal distance gives the correlation (R - Pearson). The distance to the origin is proportional to the standard deviation of the model values and the distance to the reference (measured data) is proportional to the *RMSE*.

We also used the strategy of Ritter and Muñoz-Carpena (2013), the FITEVAL. The concept behind this strategy is a Monte Carlo approach to the Nash-Sutcliffe Efficiency computation. Using repeated re-sampling from the dataset, their method delivers a probability density function to the NSE. This allows for an uncertainty analysis for the evaluator. The FL-SM presented an NSE $\in [0.66;0.95]$ for a p-value of 0.05, being classified as acceptable to very good. A conservative interpretation of this result implies considering the lowest values as the minimum state of information, or as the one that contain (almost) no unproven hypothesis. As a consequence, and according to Ritter and Muñoz-Carpena (2013), the FL-SM can be classified as acceptable to very good. The detailed output of the FITEVAL analysis can be found in the supplementary material (Fig. S6).

2.3.3 Gully growing modelling

Gully growth is commonly described as being a fast process in the first years which progressively slows down its enlargement. In our model, the mechanism that produces this dynamic is event piling. It could be observed that, after a particularly intense event, the channel is sufficiently wide. Therefore, following less intense events produce only shallow flow and low shear stress, producing fewer or no sediment. Only with a more intense event than the last erosive one, there is further erosion.

Our model mimics this growing dynamic and its periods between extreme events. Such behaviour is widely explored in literature (Vanwalleghem et al., 2005; Poesen, Vanwalleghem, et al., 2006; Poesen, 2018) and illustrated in Fig. 2.13. Vanwalleghem et al. (2005), using several data-sets from previous studies, found a strong correlation between *GT* (the percentage

of the gully age over the total) and GV (the percentage of the gully volume over the total), given by a function as expressed in Eq. 2.10. The parameters α and β were calibrated by Vanwalleghem et al. (2005) as 96.5 and -0.07 with the coefficient of determination (R^2) equal to 0.99.

$$GV = \alpha[1 - \exp(\beta\,GT)] \tag{2.10}$$
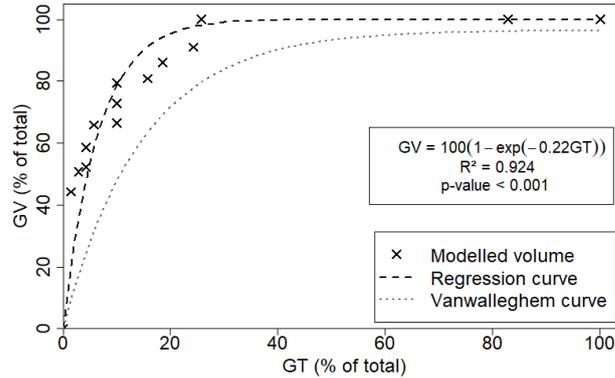
Figure 2.13: The behaviour of gully growing rate as proposed by literature (Vanwalleghem et al., 2005; Poesen, Vanwalleghem, et al., 2006) and modelled (data from Gully 1). GV is the gully volume in percentage and GT the gully age in percentage.

The parameters α (equal to 100) and β (equal to -0.22) obtained in our study differ from the values in literature. While the difference in α is due to a numerical formulation (GV_{total} is equal to the measured volume), the parameter β brings us some insights. Its absolute value for our data set is three times larger than that calibrated by Vanwalleghem et al. (2005). A larger $|\beta|$ indicates a fast initial growth, possibly caused by the intensive rainfall regime of the region, with convective intense events and high erosivity (Medeiros and Araújo, 2014). This is a different condition from Belgium and Russia, where most studies that lead to Vanwalleghem et al.'s equation were carried out. Therefore, although gully growth behaviour is similar in different regions, local conditions such as climate and land use should be taken into account.

2.3.4 Landscape development impacts on gully erosion

Gullies are scale-dependent phenomena and frequently related to thresholds due to their initiation, which is based on catchment area and slope (Torri and Poesen, 2014; Poesen, 2018). Both characteristics are directly linked to shear stress and stream power when using physical gully models. Montgomery and Dietrich (1992) argue that changes in the landscape and the drainage system can lead to a larger occurrence of channelisation and its impacts can be noticed faster. Torri and Poesen (2014) suggest a threshold for head development in gullies as conveyed in Equation 2.11.

$$S\,C_A^{0.38} > k \tag{2.11}$$

where S (m m^{-1}) is the slope, C_A (ha) is the catchment area and k is a parameter for channel and gully initiation.

For croplands in tropical conditions, the proposed value of k is 0.042 (Torri and Poesen, 2014). For the areas in the present study, we have channel initiation for values lower than half ($k = 0.020$) and systematically lower than the field data of Vandaele et al. (1996) could be observed. These findings suggest the vulnerability of the region gullying. Considering that the three experimental sites were located next to a road, this disturbance triggered gully initiation and other actions may cause similar problems in the region, such as deforestation and forest fire. The roads have not only enlarged the total catchment area but have also increased its length. While relations between catchment length and area are well-established ($L = b\,c_A^{0.49}$) with values of b varying from 1.78 to 2.02 (Montgomery and Dietrich, 1992; Sassolas-Serrayet, Cattin, and Ferry, 2018), the present experiments found b equalling 3.17. With a smoother surface and almost no meandering, road construction caused modifications that promoted more energetic flows on the gully head. Road construction has also been identified as a potential factor for gully initiation in other areas of the Brazilian Semiarid Region, as in the Salitre Catchment, where large gullies started after construction of an unpaved rural road (da Silva and Rios, 2018).

2.4   Discussion

2.4.1   Model limitations

The proposed models, especially the FL-SM, presented a significant improvement, reaching an efficiency over 0.8. Yet, some reflections can be made to understand when the models fail, as well as understand where new advances can be pursued.

2.4.1.1   Foster and Lane Model (FLM)

The FLM requires a peak discharge duration input. Given the lack of such data in the region, our first step in this study was to identify which was the best peak and duration of rainfall to be considered, based on rainfall intensity. Therefore, a relevant result from this work is the confirmation that the 30-minute intensity is the one that provides the most information about gully erosion. Wischmeier and Smith (1978) proposed the product of total storm energy and the 30-minute intensity to "predict the long-time average soil loss in runoff". The use of I_{30} for estimating event-related gully erosion was previously experimentally tested by Han, li Zheng, and meng Xu (2017). The authors had monitored a gully in the Loess Plateau in China for 12 years, registering 115 erosive rainfall events. They concluded that the product of 30-minute intensity and total precipitation ($P\,I_{30}$) was the key parameter to estimate total soil loss. Our results corroborate this.

Furthermore, by applying the I_{30} in this study in order to estimate peak discharge and duration, it is implied that all the energy necessary to initiate and develop a gully channel comes from the most intense 30 minutes. Due to the limited number of gullies, it is not straightforward that the I_{30} could be the most representative index for any situation. Peak discharge and critical rainfall duration are often central variables in gully models (Foster and Lane, 1983; Hairsine and Rose, 1992b; Sidorchuk, 1999; Gordon et al., 2007) and are related to erosion initiation parameters and thresholds, such as shear stress and stream power. This second factor has frequently been reported in literature as being more correlated with both laminar and linear erosion (Bennett and Wells, 2019). Figure 2.10a shows the performance of the tested intensities. Although the model using the 15-minute intensity presented smaller PBIAS and RMSE, the results indicated a large scatter around average.

Finally, the Foster and Lane Model also considers a fixed shear distribution, which is often unrealistic (Bonakdari, Sheikh, and Tooshmalani, 2014a) and has a fixed rectangular shape that, although frequently accurate for ephemeral gullies, does not agree with field data and literature (Fig. 2.8; Starkel, 2011).

2.4.1.2 Sidorchuk Model (SM)

The SM produced good results in this study which were similar to those obtained by inserting a calibrated factor (λ) in FLM. It is important to note that the original model used empirical correlations to determine width (Sidorchuk, 1999; Nachtergaele, Poesen, Sidorchuk, et al., 2002) and these were obtained using data from the Yamal basin. In the present study, we substituted this approximation for the width estimated by the FLM model, which permitted a more physical approach and increased the quality of the SM. The model was also capable to predict the sidewall slope well.

The model, however, showed a trend of overestimating smaller cross-sections (see Figure 2.7) mainly due to section geometry. When applied, the bottom width of the channel is considered to be the final width obtained by FLM. In larger sections, this hypothesis holds once the discharge is large enough to carry all soil produced by sidewall erosion. In smaller sections, part of the soil is deposited and produces a V or U-shaped cross-section (Starkel, 2011).

2.4.1.3 Proposed Models

The FLM was further improved by the addition of the calibrated parameter λ (FLM-λ). This parameter was included to predict the effect of lateral discharges over wall erosion. Due to the significant improvement produced by its insertion, it could be understood that the original FLM fails to tackle this source of material (Blong and Veness, 1982; Crouch, 1987).

The FL-SM considers two sediment sources: channel bed and sidewall. Gullies are, however, complex systems with many sources and interactions. Headcut, sidewall erosion due

to raindrops, flow jets and piping were not considered in our modelling approach. Processes of infiltration, subsurface flow and transport capacity were also neglected and should be properly addressed in future works. Nevertheless, the FL-SM assumptions managed to mimic the field measurements well, which implies that, at least in this study, the neglected processes are of lower relevance or were considered indirectly. For instance, sidewall erosion by raindrops can be considered insignificant over the wall failure process considered by Sidorchuk (1999). In addition, it is important to notice that, by selecting the 30-minute intensity, a less intense interval might be overlooked that also produces erosive discharge, and can, therefore, explain the remaining processes.

One advantage of the FLM-λ over the FL-SM is that the former requires less data than the latter. The Sidorchuk model, and consequently the FL-SM require extra fieldwork and laboratory analysis to assess root mass and plasticity index.

Despite the good results obtained from the modelling, the use of stochastic approaches and introduction of other sources of sediment(Sidorchuk, 2005) should improve the performance of the model. This is also relevant for generalisation and modelling of classical gullies. In the same way, the introduction of processes such as armouring and energy losses, as proposed by Hairsine and Rose (1992b), can be interpreted as probabilistic terms.

Comparatively, with other models, either physical or empirical (Hairsine and Rose, 1992b; Woodward, 1999; Wells et al., 2013; Dabney et al., 2015), our proposed model (FL-SM) requires a similar or less amount of data, little calibration (one parameter – the threshold) and is more versatile. Most models fail to account for multiple rainfall events (Foster and Lane, 1983; Woodward, 1999; Nachtergaele, Poesen, Sidorchuk, et al., 2002) and to consider multiple sources of sediment (Foster and Lane, 1983; Hairsine and Rose, 1992b; Dabney et al., 2015). The FL-SM model ($R^2 = 0.94$) presented a better performance index than empirical models [e.g. R^2: 0.55 and 0.12 for Woodward (1999) and Wells et al. (2013) respectively] and physical models [e.g. R^2: 0.87 and 0.84 for Foster and Lane (1983) and Sidorchuk (1999) respectively]. This enhancing in the performance can be accounted for by the more detailed modelling, considering wall failure and non-rectangular cross-section.

2.4.2 Data limitations

2.4.2.1 Topographic data

In terms of accuracy and agility, a topographic survey with UAV permits to measure sites within a few minutes. Conventional measurements, such as those with total station or profilometer, are more time consuming and do not grant better resolutions. The UAV accuracy, however, can be enhanced by performing flights in lower heights and with more GCPs (Agüera-Vega, Carvajal-Ramírez, and Martínez-Carricondo, 2017; James, Robson, et al., 2017), as well as by using high-end equipment, such as more robust UAVs and stabilizers. Total stations can also be used to improve the accuracy of ground control points (Mesas-Carrascosa et al., 2016). Given the scale of this study and the presented results of the models, the four-centimetre pixel

represents a good resolution, since it combines good precision with affordable computational costs (Wang, Zhang, et al., 2016). The solution of UAV-based volume assessment is a good option for monitoring gully evolution (Stöcker, Eltner, and Karrasch, 2015), allowing frequent surveys, e.g., after every intense rainfall event.

However, trees and bushes obscure topographic measurements if too close to the gully channel and/or too dense in the catchment. Thus, UAV monitoring is more reliable for gully sites in non- or meagre-vegetated areas and meadows, which combines with the conditions of this study, except for gully 3 (G3), where it was impossible to accurately measure the total erosion volume due to relatively dense vegetation. It was, however, possible to select a large enough number of sections (eight) at G3 to assess the total erosion volume.

The topographic survey showed that all gullies had a significant portion of their watersheds occupied by the road, indicating a modification of the drainage system and change in the catchment boundaries – both causes of gully initiation foreseen by Ireland, Sharpe, and Eargle (1939) – due to road construction, which promoted intense runoff and triggered gullies. Impacts of road construction on gully formation were also observed in Ethiopia (Nyssen et al., 2002) and the USA (Katz, Daniels, and Ryan, 2014). Considering such previous records in literature and the information collected from locals, the modelling considered 1958 as the start of gully erosion, coinciding with road construction.

2.4.2.2 Soil data

Though the three studied gullies are located in the same mesoscale basin, the Caatinga biome is known for its soil variability (Güntner and Bronstert, 2004) and soil properties do differ among the gullies. However, only small changes of texture were observed in different depths, allowing an analysis based on average properties. Nevertheless, for deeper and/or more variable soils, the discretization of soil properties, and therefore parameters such as rill erodibility (K_r) and critical shear stress (τ_c), can easily be taken into account. The good performance of the final model (FL-SM) also indicates that the WEPP equations for critical shear stress and rill erodibility (Eq. 2.1 and 2.2) can be used for the soils of the region. These equations were obtained via regression curves from data collected on 34 plot areas in the USA with a wide range of textures, slopes, land use and land cover. The areas from the WEPP model possess different geological and climatic conditions from the soils in the Brazilian semiarid region; this is why local studies should be carried out, given that empirical equations frequently have strong local character (Ghorbani-Dashtaki, Homaee, and Loiskandl, 2016; Dionizio and Costa, 2019).

2.4.2.3 Rainfall data

This study shows that sub-daily information of rainfall is of crucial importance for gully modelling. In this study, we used correlation curves based on long-term time series of a similar catchment in the region. However, such analysis might introduce an averaged and

monotonic behaviour for the intensities, as presented in Figure 2.4, and is, therefore, unrealistic. Stochastic models should be tested to estimate sub-daily information from daily rainfall. The estimation of discharge from rainfall can also be improved by considering water content in the soil and modelling its evolution over the studied period using water balance models.

2.5 Conclusions

In this study, we proposed and tested two new gully models based on two previous models (Foster and Lane, 1983 and Sidorchuk, 1999). We also investigated which rainfall intensity is best suit for gully modelling when sub-daily rainfall data is not available, finding the 30-minute intensity (I_{30}) to be the most appropriate.

The models present a significant improvement when compared to other models in literature. While the FLM-λ requires less calibration, the FL-SM presented better results, not only in terms of total gross erosion, but also in terms of gully growing dynamic. Through modelling and fieldwork, it was also possible to identify the effects of landscape change and climatic conditions on gully development in the region. Gully is an erosion related to many processes and it is scale-dependent. The attempt of proposing a generalist model for gullies should also consider these different scales and mechanisms involved in different stages of the gully development. Catchment shape and lateral flow have a central role in gully erosion and their influence should be further investigated. Infrastructure construction, like roads, changes conditions for gully initiation and was the trigger for the studied gullies.

Nonetheless, further efforts are required to fully model gully erosion, such as include the multiple other sources including headcut, pipping, channel shear stress and flow jets. Also, stochastic modelling should be implemented in order to tackle inherent uncertainties of many sediment sources and lack of data.

3

Entropy-based Model for Gully Erosion – a Combination of Probabilistic and Deterministic Components

Abstract

Gullies are a major threat to ecosystems, potentially leading to land degradation, groundwater depletion, crop loss, debris flow, and desertification. Gullies are also characterized by having a fast development and turning into primary sediment sources. Despite their impact, we have but scarce understanding of how gully erosion evolves and how to model it. In this paper, we propose a new gully erosion model that is based on the classical premise of net shear stress, i.e., hydraulic shear stress *minus* critical (resistant) shear stress, to calculate detachment rates. In order to calculate hydraulic shear stress, we developed a new equation derived from the principle of minimum cross-entropy; it was validated with laboratory measures from the literature with a Nash-Sutcliffe Efficiency of 0.95. Soil samples were analysed in the laboratory to assess critical shear stress and other soil properties. The novel gully erosion model was implemented in three gully impacted locations with catchment areas ranging from 10^{-2} to 10^{+1} hectares. To assess channel geometry and eroded volumes, we used Unmanned Aerial Vehicle and Structure-from-Motion technique. The model successfully estimated long-term erosion rates, its efficiency was 0.77, and it is recommended for catchments up to 8 ha. Therefore, the

new model provides planners and stakeholders with a tool to assess gully erosion, sediment yield and geometry in most areas.

Keywords: linear erosion; open channel; entropy theory; minimum cross entropy

3.1 Introduction

As far as achieving sustainable development is concerned, soil erosion has been pointed out as a central problem to be confronted in the 21st century(Borrelli et al., 2017; Poesen, 2018). Additionally, the Food and Agriculture Organization named erosion as one of the most pressing threats to soil conservation and agriculture (FAO, 2019). Soil erosion is already responsible for crop losses up to 33.7 million tonnes and additional 48 km^3 of water usage yearly, affecting more severely countries like Brazil, China and India, and low-income households worldwide (Nkonya et al., 2016; Sartori et al., 2019). Among the erosion processes, gullies are critical for the functioning of hydrosedimentological processes in a catchment scale Vanwalleghem et al. (2005) and Bingner et al. (2016).

Gullies are linear erosion, that develop rapidly and are connected to land degradation, crop losses, and desertification (Valentin, Poesen, and Li, 2005); they are, however, frequently overlooked by erosion models (Poesen, Vandekerckhove, et al., 2002; Poesen, 2018; Bennett and Wells, 2019). Gullies are channels carved by concentrated rainwater overflow. The runoff usually concentrates in narrow paths due to natural micro-relief (Poesen, Torri, and Van Walleghem, 2011). This process causes an increase in sediment transport capacity, connectivity, stream power and flow shear stress, which promote the deepening of narrow channels (Valentin, Poesen, and Li, 2005). Erosion by gully, thus increase sediment yield (Bingner et al., 2016). The resulting channels foster multiple hydrological modification at local (hillslope) and regional (catchment) scale. Locally, gullies further a reduction in groundwater storage and fertility (Valentin, Poesen, and Li, 2005); on a catchment scale, gullies increase water and sediment connectivity from the upper areas to the outlet, causing water-body siltation and pollution (de Vente and Poesen, 2005). Due to the increase in water connectivity, gullies may also aggravate debris flow and floods (Liu, Tang, et al., 2016; Wei et al., 2018).

Such considerable hydrological changes in a region have a direct impact on the economic and social activities of the respective areas. Productive lands affected by gullies present, on average, a 37% loss in agricultural production and land recuperation costs in the order of 5.2% of total expenses (Valentin, Poesen, and Li, 2005). Gullies are impediments for human, animal and machinery movement on farmlands and may represent a risk to security of infrastructure (Alencar, de Araújo, and Teixeira, 2020). The occurrence of gullies is usually related to changes in land use, land cover and drainage conditions, and these transformations are commonly associated with human activities such as deforestation and infrastructure development (Poesen, 2018). Gullies alone produce between 10 and 100 ton.ha^{-1}.yr^{-1} in affected areas, besides causing changes to water and sediment connectivities and productivity/biodiversity loss (Avni, 2005; Verstraeten et al., 2006; Poesen, Torri, and Van Walleghem, 2011). Despite their strong

impact on ecological, hydrological and social impacts (Poesen, 2018), there is a shortage of gully-erosion monitoring, as pointed by Cerdan et al. (2006), who identified a number of site-year observations of linear to be over 40 times lower than for its counterpart laminar erosion.

Despite their influence in hydro-sedimentological processes in a catchment scale, erosion models overlook gully erosion assessment (Poesen, 2018). Gullies are complex systems with the superposition of multiple processes, as shear stress, head-cutting, jet flow, piping, and wall failure. Therefore, gully erosion is a process with the interaction of many variables, many of which are difficult to model (Bernard et al., 2010; Castillo and Gómez, 2016; Alencar, de Araújo, and Teixeira, 2020). Therefore, no model has ever been proposed to explain the governing forces controlling gully initiation and growing (Bennett and Wells, 2019), as gullies can vary widely in shape, scale and governing processes (Starkel, 2011).

Many models have been proposed to estimate soil loss in gullies (Thompson, 1964; Woodward, 1999; Nachtergaele, Poesen, Sidorchuk, et al., 2002; Wells et al., 2013; Foster and Lane, 1983; Hairsine and Rose, 1992b; Sidorchuk, 1999; Dabney et al., 2015; Alencar, de Araújo, and Teixeira, 2020). Among them, shear stress is the most used variable to describe soil detachment due to gully erosion. Nevertheless, the shear stress distribution in the boundary layer, which is decisive to assess erosive processes in channels, has not been well described mathematically.

In the Foster and Lane model (1983, one of the most cited and used gully erosion models – Alencar, de Araújo, and Teixeira, 2020), the soil detachment rate is directly proportional to the difference $\tau - \tau_c$. τ is the acting shear stress promoted by the discharge in the channel and τ_c the critical shear stress, a measure of soil resistance. Detachment occurs whenever $\tau \geq \tau_c$. The constant of proportionality is known as Rill Erodibility (K_r). This approach has been successfully implemented in multiple studies (Storm, Barfield, and Ormsbee, 1990; Woodward, 1999; Casalı, López, and Giraldez, 2003; Dabney et al., 2015), yet these net-shear-stress methods pose two challenges: (1) how to estimate soil-based parameters, i.e., critical shear stress and rill erodibility; and (2) how to estimate shear stress distribution over the channel's boundary layer (wet perimeter).

Assessing soil-based parameters such as $_c$ and K_r is still an open problem (He et al., 2021). They are frequently estimated with pedotransfer functions (Alberts, Laflen, et al., 1989; Lal, 1994; Alberts, Nearing, et al., 1995). In order to assess $_c$ and K_r it is common to employ percentages of sand, clay, organic matter, roots and the plasticity index; pedotransfer functions (PTF) have been implemented successfully in multiple studies (Watson and Laflen, 1986; Ascough et al., 1997; Sidorchuk, 1999; Dabney et al., 2015; Alencar, de Araújo, and Teixeira, 2020; Luquin et al., 2021).

With respect to assessing shear stress distribution on the boundary layer, several models have been proposed, such as empirical equations (Foster and Lane, 1983; Storm, Barfield, and Ormsbee, 1990), geometric methods (Khodashenas and Paquier, 1999; Ikeda, 1982), physical-based equations (Prandtl, 1925; Smart, 1999; Yang and McCorquodale, 2004) and

maximum-entropy-based equations (Chiu, 1988; Sterling and Knight, 2002; Bonakdari, Sheikh, and Tooshmalani, 2014b).

The objective of this paper is to propose a novel gully erosion model based on the successful net-shear-stress approach, and to validate it for gully-affected areas. To achieve that, we propose a new shear stress distribution equation that was derived by applying the principle of minimum cross entropy. The new shear stress distribution was validated with experimental data from literature. Subsequently, the equation was included in the EBGEM (Entropy-Based Gully Erosion Model) frame to assess net shear stress and after that, the EBGEM was validated for three gully systems of different scales.

3.2 Materials and Methods

3.2.1 Entropy-Based Gully Erosion Model (EBGEM)

The EBGEM is an advance with respect to the model by Alencar, de Araújo, and Teixeira (2020), which uses the model by Foster and Lane (1983) as the basis for a long-term modelling of small gullies with an area-based threshold parameter to control wall geometry: cross-sections with areas above 2 m^2 have triangular or trapezoidal shape, otherwise the shape would be rectangular.

In the EBGEM, we propose applying the entropy theory to derive a novel shear stress distribution over the channel. Some of the advantages of this model are that it does not require any previous calibration (such as the threshold parameter for wall erosion) and is not limited to small gullies (drainage area up to 1 hectare and depth up to 1 meter), while using the same input data as the model in Alencar, de Araújo, and Teixeira (2020). Moreover, the EBGEM does not depend on empirical relations and the conveyance functions defined in Foster and Lane (1983). The only exception is the initial incision, which uses the Watson and Laflen (1986) equation to calculate the event-based width and depth. A complete EBGEM flowchart is presented in Figure 3.1.

One of the great advances of the proposed model is the use of an accurate shear-stress distribution in the channel, which is obtained by employing the Principle of Minimum Cross-Entropy (Kullback, 1978a). Appendix A extensively describes how to deduce the shear stress distribution function.

3.2.1.1 The Principle of Minimum Cross-Entropy (POMCE)

Kullback and Leibler (1951) first presented the concept of relative information, which was further explored in Kullback (1978a). It late became known as the Kullback-Leibler Divergence (D_{KL}).
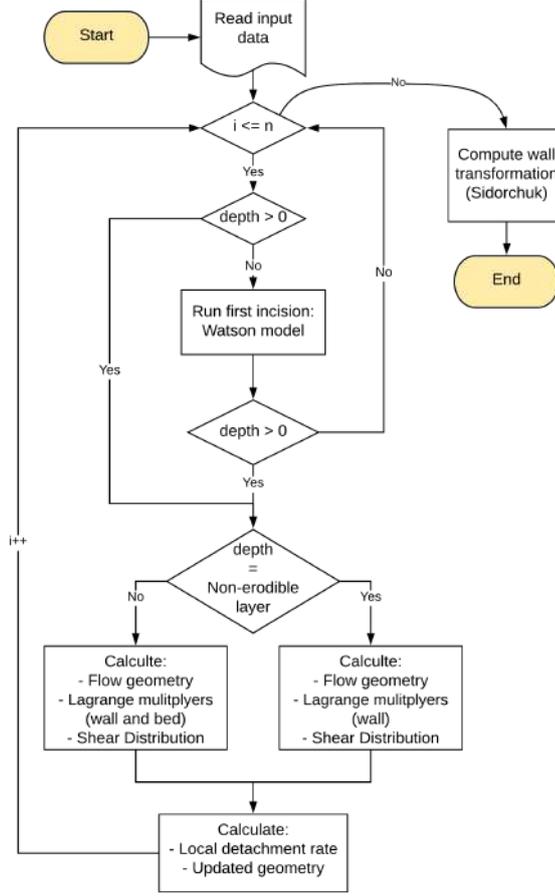
Figure 3.1: Flowchart of the Entropy-Based Gully Erosion Model (EBGEM).

$$D_{KL}(P \,||\, Q) = \int_\Omega p(x) \ln \frac{p(x)}{q(x)} \, dx \qquad (3.1)$$

where P and Q are probability distributions following $p(x)$ and $q(x)$. Ω is the space (domain) of both functions.

The POMCE follows the assumption that, given a chosen function (also known as *prior function*) $q(x)$, either guessed or obtained from previous knowledge about the data, the less unbiased function $p(x)$ to represent the variable, given a set of constraints $c_i(x)$, is the one that minimizes Divergence (D_{KL}). In other words, $p(x)$ is the most similar function to *prior* $q(x)$ given the new knowledge gained by the constraints $c_i(x)$ and without assuming any unproven hypothesis.

Using the Euler-Lagrange method to solve the minimization of Equation 3.1 subject to a set of constraints $c_i(x)$, one finds the Lagrangian $\mathfrak{L}$ (Eq. 3.2).

$$\mathfrak{L} = \int_\Omega p(x) \ln \frac{p(x)}{q(x)} \, dx \;+\; \lambda_0 \left[\int_\Omega p(x) dx - 1 \right] \;+\; \sum_{i=1}^{n} \lambda_i \left[\int_\Omega c_i(x) \, f(x) dx - \overline{c_i(x)} \right] \qquad (3.2)$$

49

where λ_0 and λ_i are Lagrange multipliers. From the method of Lagrange multipliers, one finds the following equation (Eq. 3.3).

$$p(x) = q(x) \exp \left[-1 - \lambda_0 - \sum_{i=1}^{n} \lambda_i c_i(x) \right] \tag{3.3}$$

The constraints can be expressed as functions of x and statistical moments are frequently used (Eq. 3.4).

$$c_0(x) = \int_\Omega p(x)\,dx = 1 \tag{3.4a}$$

$$c_1(x) = \int_\Omega x\,p(x)\,dx = \bar{x} \tag{3.4b}$$

Selecting a suitable *prior function* is essential for obtaining a good performance. Whenever a uniform distribution is adopted as prior function, the POMCE is reduced to the maximization of the Shannon entropy (Kumbhakar, Ghoshal, and Singh, 2019).

3.2.2 Cross-entropy shear stress equation

Using the principle of minimum cross entropy (Kumbhakar, Ghoshal, and Singh, 2019), we derived a novel equation to assess shear stress distribution in the boundary layer of open channels (Eq. 3.5; Appendix 3.5).

$$e^{-\lambda x}(\lambda x + 1) = 1 - (1 - e^{-\lambda}(\lambda + 1))\frac{y}{L} \tag{3.5}$$

where $x = \frac{\tau}{\tau_{max}}$, y is the position over the wall or bed length (L), and λ is a Lagrange multiplier that can be obtained through the constraints (Equation 3.4. Also, Equation 3.16 in the Appendix A.1).

Equation 3.5 was validated with measured data (Knight, Demetriou, and Hamed, 1984; Tominaga et al., 1989; Knight and Sterling, 2000) and compared with the following two equations from literature: Sterling and Knight Equation (SKE – Eq. 3.6; Sterling and Knight, 2002) and Ford and Labosier Equation (FLE – Eq. 3.7; Foster and Lane, 1983).

$$\tau = \frac{1}{\lambda} \ln \left\{ 1 + (\exp\left(\lambda\,\tau_{max}\right) - 1)\frac{y}{L} \right\} \tag{3.6}$$

$$\tau = 1.35\bar{\tau} \left[1 - \left(1 - 2\frac{X}{P}\right)^{2.9} \right] \tag{3.7}$$

where $\bar{\tau} = \rho g R S$. X is the distance measured from the intersection of the water surface following the wetted perimeter (P).

3.2.3 Experimental Areas

To validate the EBGEM we surveyed three gully affected areas in Northeastern Brazil (Madalena, Gilbués and Campo Formoso – Figure 3.2).

The Northeast of Brazil covers an area of over 10^6 km². Its climate is predominantly hot semi-arid (BSh, by the Köppen classification), while its vegetation mainly consists of Caatinga forest, one of the largest dry forests in the world and formed mostly by broad-leaf deciduous trees and bushes (da Silva and Rios, 2018). The average precipitation and potential evapotranspiration are 800 and 2500 mm.yr^{-1}, respectively (Pinheiro, Costa, and Araújo, 2013). Housing over 25 million people, the main activities in the region are cash-crop agriculture (sugar cane, maize, soy beans and cotton), open range cattle and fishery (Zhang, Foerster, Medeiros, Carlos, et al., 2018; Coelho et al., 2017). The area is mostly situated on top of crystalline bedrock with shallow soils and limited groundwater resources (de Araújo, Güntner, and Bronstert, 2006). Therefore, rivers are intermittent and most water supplies rely on surface reservoirs which, due to siltation, have reduced storage capacity and water quality (de Araújo, Güntner, and Bronstert, 2006). This problem is aggravated by gully erosion.
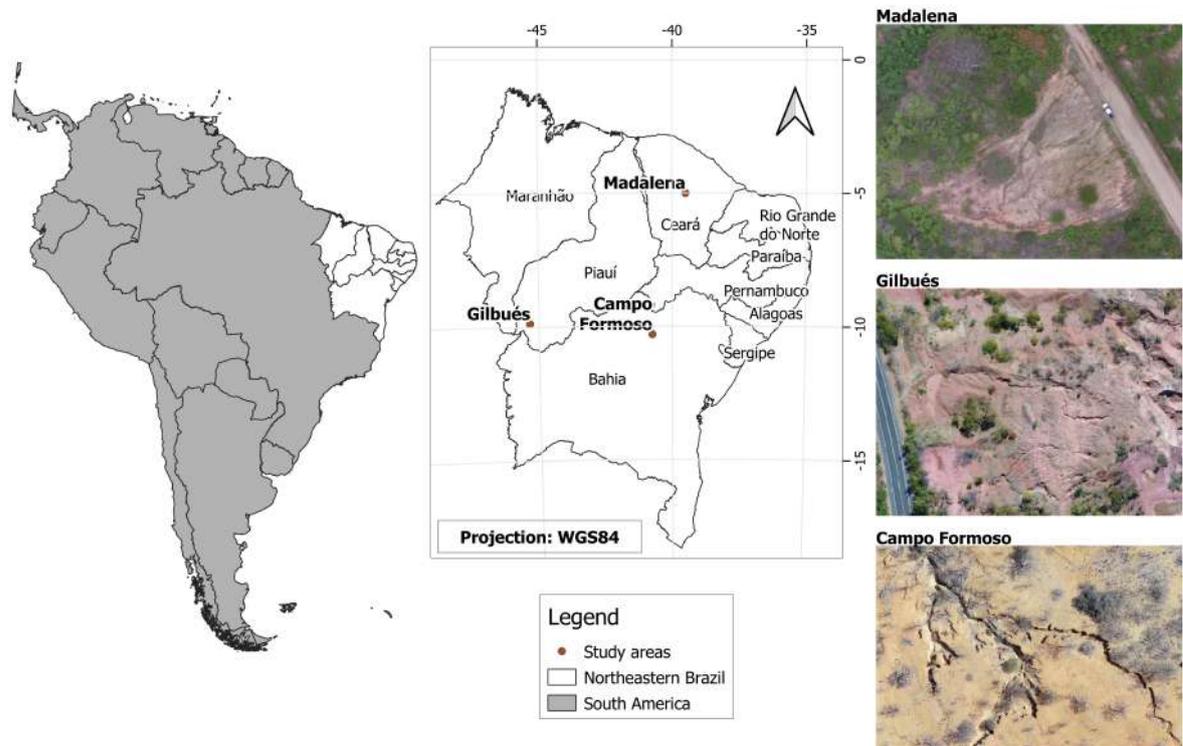
Figure 3.2: Study areas in the Northeast of Brazil. On the right side, we present pictures of the landscape of each area.

The first area is located in the municipality of Madalena, State of Ceará, in the Madalena Representative Basin (MRB). The MRB houses a land reform settlement with 550 families who live from agriculture (maize, beans and vegetables), cattle raising (mainly for milk) and fishery (Coelho et al., 2017). The 75 km^2 basin has numerous gullies distributed along a road constructed in 1958 (Figure 3.2). The construction of this road caused the initiation of the gully erosions due to deforestation and modification of the drainage system (Alencar, de Araújo, and Teixeira, 2020).

The second area is located in Gilbués, State of Piauí. The Gilbués Experimental Basin (GEB) measures 6200 km^2 and is the largest desertification area in Brazil. The region has numerous badlands and gullies, some reaching dozens of meters of depth and width (Simplıcio et al., 2020). The main economic activities are agriculture (soy beans) and livestock.

The third area is located in Campo Formoso, State of Bahia. The Representative Basin of Campo Formoso (RBCF) stretches out over 16 hectares and is also under desertification, a process that was initiated after intensive deforestation, when the native Caatinga forest was removed to produce *Agave sisalana* on the hillslopes (Jesus, 2021). *Agave* production is still the main economic activity in the region, besides livestock (da Silva and Rios, 2018).

3.2.4   Topography and soil survey

In the three regions, a detailed topographic survey was performed with the aid of an unmanned aerial vehicle (UAV) carrying a camera with a 16 megapixels resolution (4000 x 4000 pixels) and a field of view of 94 %. Flight altitude was 50 meters with frontal overlap of 80 % and lateral overlap of 60 %. Geo-referencing the mosaic was possible due to eight ground control points evenly distributed in each area, both in high and low ground. The coordinates of the ground control points were collected using a stationary GNSS-RTK (L1/L2) system with centimetre-level accuracy (Alencar, de Araújo, and Teixeira, 2020).

Thanks to the application of the Structure from Motion technique (Schonberger and Frahm, 2016), it was possible to produce a detailed digital surface model (DSM) of each study area. The Structure from Motion (SfM) is founded on a three-dimensional reconstruction of the surface, it bases itself on images and the generation of a dense cloud of points obtained by matching pixels of different images and ground control points. This process results in a DSM as accurate as if obtained by through laser survey (e.g. LiDAr), while cheaper and less time-consuming (Agüera-Vega, Carvajal-Ramírez, and Martínez-Carricondo, 2017). The DSM pixel size (ground sample distance) ranges from two to five centimetres, while average vertical position accuracy amounts to one centimetre and average horizontal accuracy to five millimetres.

At multiple depths and locations of each study area, soil samples were extracted with an Uhland sampler. The samples were analysed in laboratory in order to investigate the physical properties of the soil which are necessary to assess critical shear stress and rill erodibility (Eq. 3.8 and 3.9). The detailed soil analyses results are available in the supplementary data of this

paper. Table 3.1 presents the properties and classification of the soils in each region, following the Brazilian Soil Classification System (Dos Santos et al., 2018).

Table 3.1: Soil properties and classification. Each study area was subdivided according to its soil properties. The values of K_r and τ_C were calculated with Equations 3.8 and 3.9

Study area	Soil texture	Soil class[a]	K_r (s m^{-1})	τ_c (Pa)
Madalena				
Area 1	Sandy Loam	*TC17 (Luvisol)*	0.017	2.30
Area 2	Clay Loam		0.012	3.50
Gilbués				
Area 1	Silt Loam	*RQ47 (Neosol)*	0.013	3.50
Area 2			0.009	3.50
Campo Formoso				
Area 1			0.011	2.74
Area 2	Sandy Loam	*CX33 (Cambisol)*	0.012	2.75
Area 3			0.013	2.75

[a] Brazilian Soil Classification System (Dos Santos et al., 2018).

The critical shear stress was estimated through pedotransfer functions (Alberts, Laflen, et al., 1989; Lal, 1994; Alberts, Nearing, et al., 1995) that use mostly the content of sand (S_a), clay(C_l), organic matter (O_m), and root index (R_i) (Eq. 3.8 and 3.9).

$$K_r = 0.0017 + 0.0024\,C_l - 0.0088\,O_m - 0.00088\left[\frac{\rho}{1000}\right] - 0.00048\,R_i \qquad (3.8)$$

$$\tau_c = 3.23 - 5.6\,S_a - 24.4\,O_m + 0.9\left[\frac{\rho}{1000}\right] \qquad (3.9)$$

rho is the density of water, 1000 kg m$_{-3}$.

3.2.5 Rainfall data

Daily rainfall data for each location throughout the whole period of analysis was provided by the Brazilian Water Agency (ANA) and the Foundation of Meteorology and Water Resources of Ceará (Funceme). Table 3.2 indicates the periods of analysis and rain data availability. The complete rainfall time series for all the study areas is available in the supplementary data. The rain data was tested for consistency through the double mass method, and measurement gaps ($< 0.05\%$ of the series duration) were filled with data provided by the nearest gauging station.

Since peak discharge is a measurement which is not available in any of the reseach areas, it was assessed using the thirty-minute maximum intensity (I_{30}) instead (Alencar, de Araújo, and Teixeira, 2020). In order to obtain sub-daily rainfall intensity, a correlation curve between total daily precipitation and 30-minute intensity was drawn (Figure 3.3). This was done on

Table 3.2: Daily rainfall data availability. The start year reveals the beginning of erosion processes in all the three, coinciding with infrastructural or land use changes, e.g. road construction.

Study area	Start	End	Data source*
Madalena	1958	2015	Funceme[a]
Gilbués	1946	2020	ANA[b]
Campo Formoso	1967	2020	ANA[b]

* Rain gauges ville de Paris (daily total precipitation)
[a] FUNCEME: Foundation of Meteorology and Water Resources of Ceará
[b] ANA: National Water Management Agency

the basis of data from the Experimental Basin of Aiuaba (de Figueiredo et al., 2016) with continuous monitoring of sub-daily rainfall since 2005.

I30 = 2.852 P^0.681
R^2 = 0.49 P < 0.001
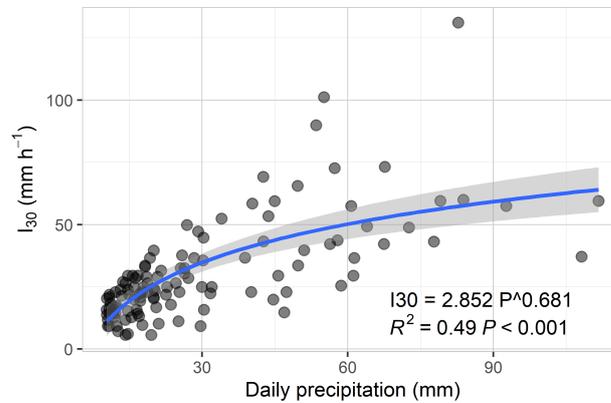
Figure 3.3: Correlation between daily precipitation (P) and 30-minute maximum intensity (I_{30}) at the Aiuaba Experimental Basin (de Figueiredo et al., 2016). The blue line indicates the power law regression and the shaded area the confidence interval at 95 %.

3.3 Results

3.3.1 Cross-entropy shear stress equation

The shear stress equation (Eq. 3.5) was derived by applying the Principle of Minimum Cross-Entropy, and was validated using results from 13 experiments from Knight, Demetriou, and Hamed (1984), Tominaga et al. (1989), and Knight and Sterling (2000). We evaluated the gain in quality provided by the new equation when we compare t with the POME-derived equation (SKE – Sterling and Knight, 2002) and with the equation in Foster and Lane (1983, FLE). Figure 3.4 exhibits the absolute error distribution of the equations compared to 181 shear stress measurements in 13 sections.

Figure 3.4 conveys, a remarkable performance difference when comparing the FLE (Eq. 3.7) with SKE (Eq. 3.6) and Eq. 3.5. The two Equations 3.5 and 3.6 performed very similarly, although the Cross-Entropy equation (Eq. 3.5) proved to be marginally better. Multiple

<center>(a)   (b)</center>
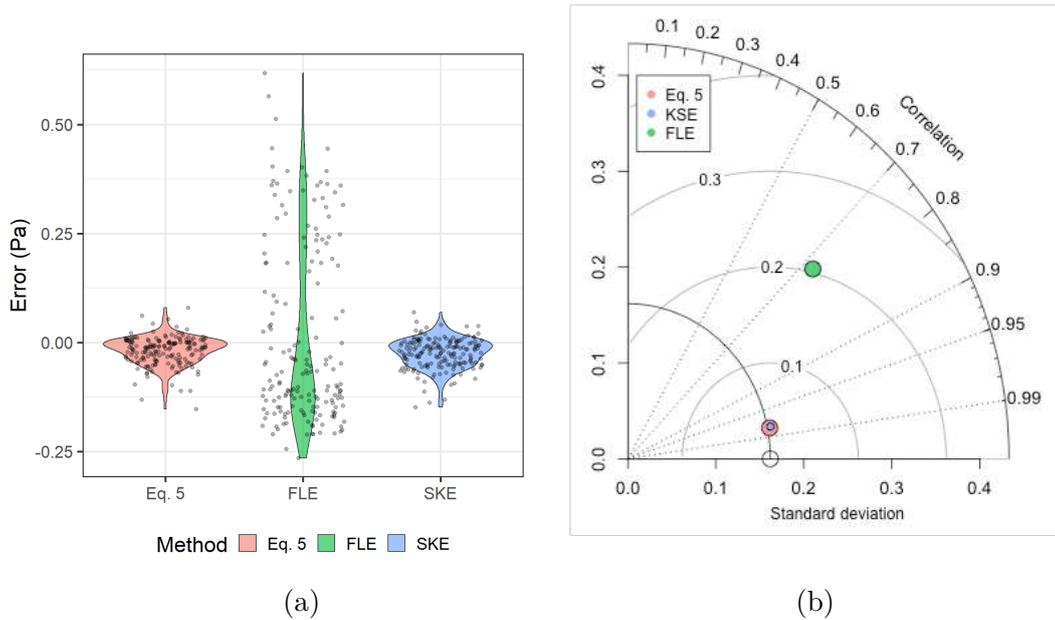
Figure 3.4: (a) Absolute error distribution for the Cross-Entropy Shear Stress Equation (Eq. 3.5), Foster and Lane (1983, FLE, Eq. 3.7) and Sterling and Knight (2002, SKE, Eq. 3.6). Black dots indicate the actual values of error for each equation. (b) The Taylor diagram presents the Pearson correlation (azimuthal angle), standard deviation of results (radial distance) and root-mean-square error (RMSE – grey contour lines) for the three equations. Note that Eq. 3.5 and SKE have very similar results, with Eq. 3.5 being marginally closer to the reference (white circle over the x-axis).

performance evaluators for the three equations are listed in Table 3.3 which, at the same time, displays that Eq. 3.5 performed marginally better than SKE in all criteria.

3.3.2   Entropy-based gully erosion model (EBGEM)

In this study, we measured and modelled nine gully systems from three catchments areas in the Brazilian Northeast. In total, 65 cross-sections were modelled with the EBGEM. The complete list of cross-sections and their characteristics can be found in Appendix 3.5 (Table 3.4). The drainage area of the gullies ranged from 10^{-3} to 10^{2} hectares and the slope from 3 to 28 %. The dimensions of carved channels also differed, with eroded area ranging from 10^{-3} to 10^{2} m^{2}. In Figure 3.5 we show the scatter plot of measured and modelled values for all the cross-sections.

Especially in the Campo Formoso sections we can notice that the model tends to overestimate, where no particular trend can be observed in the other catchment areas. The increase in marker size reveals an augmented squared-error.

In Figure 3.6 we present some model evaluators for the cross-sections. The model has an overall efficiency of 0.77 when dealing with gullies with drainage areas of up to 3 hectares. When considering all the sections (with drainage area up to 16 ha), efficiency drops to 0.58.

In Figure 3.6 and using the classification by Moriasi et al. (2007) we can see a good EBGEM performance in Madalena (NSE = 0.69) and Gilbués (NSE = 0.72), and an unsatisfactory one

Table 3.3: Multiple performance evaluators for the three models: EBGEM (Eq. 3.5), SKE (Eq. 3.6), and FLE (Eq. 3.7). The modelled values were compared with experimental results (Knight, Demetriou, and Hamed, 1984; Tominaga et al., 1989; Knight and Sterling, 2000) of 13 cross-sections with 181 point measurements.

	Eq. 3.5	SKE	FLE
NSE[a]	0.950	0.930	-0.594
RMSE[b]	0.037	0.042	0.204
PBIAS[c]	3.3%	4.3%	-2.1%
ρ_P[d]	0.981	0.979	0.728
ρ_S[e]	0.961	0.958	0.772
τ_K[f]	0.839	0.829	0.601

[a] NSE: Nash-Sutcliffe Efficiency
[b] RMSE: Root Mean Square Error
[c] PBIAS: Percentage of Bias
[d] ρ_P: Pearson correlation coefficient – parametric measure of linear correlation
[e] ρ_S: Spearman's rank correlation coefficient – non-parametric measure of rank correlation
[f] τ_K: Kendall rank correlation coefficient – non-parametric ordinal association coefficient

in Campo Formoso (NSE = 0.43), a locality with the largest gullies and catchment areas. This result indicates a relation between model performance and catchment area, which might be the key to multiple processes that are not considered in the EBGEM. Furthermore, Figure 3.5 shows that the model tends to overestimate eroded area in larger sections, suggesting that processes not taken into account are protective responses to the erosion, either by human action (e.g. simple structures of sediment retention) or natural event (e.g. shielding of top layer by coarse material).

3.3.3   Gully growth modelling

Gully channel growth is usually described as being fast in the early stages of development, and gradually slowing down thereafter (Vanwalleghem et al., 2005; Poesen, 2018; Alencar, de Araújo, and Teixeira, 2020). Figure 3.7 shows the erosive-process evolution according to the EBGEM. As comparison with the model by Vanwalleghem et al. (2005) displays that the proposed model captures correctly the expected dynamic.

The equation proposed by Vanwalleghem et al. (2005; Fig. 3.7) has a very good agreement with measured data, presenting $R^2 = 0.97$. Nevertheless, it is interesting to note a significant deviation in the early stages ($GT < 5\%$), when most of the gullies present a faster development than the one predicted by the Vanwalleghem et al. curve. Immediately after this intense initial erosion, there is a growth deceleration relative to the Vanwalleghem et al. curve. This fact might be explained by the regime of precipitation in the Brazilian Northeast, which differs from the one of the countries where the equation was calibrated (Belgium and Russia). In our study area, precipitation tends to be convective, with 30-minute intensities of up to 130 mm h[-1] and a high inter-annual variability (Medeiros and Araújo, 2014).
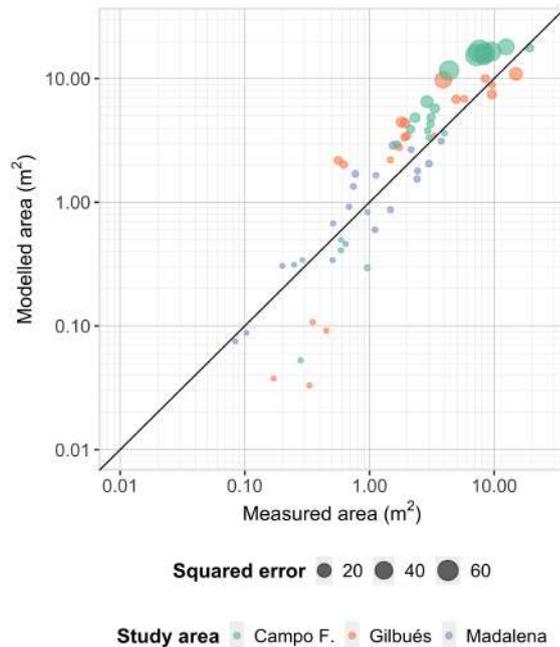
Figure 3.5: Measured and modelled area of the 65 modelled cross-sections in the three study areas in Brazil: Campo Formoso, Gilbués, and Madalena.

3.4 Discussion

Soil erosion by water was pointed out as a key problem to be dealt with in order to achieve the 21st century sustainable development goals (Borrelli et al., 2017; Poesen, 2018). Gullies, as presented in Section 3.1 are a major threat to environments, reducing available water for vegetation and accelerating desertification processes (Valentin, Poesen, and Li, 2005). This is particularly harmful to arid and semiarid regions, where studies on gully susceptibility mapping showed that, over 30% of the catchments in those regions have high or very high risk of gully erosion (Javidan et al., 2019; Azareh et al., 2019), potentially making gullies the main source of sediment (Bocco, 1991; Bennett and Wells, 2019). Understanding the behaviour and processes in gullies is relevant to guarantee effective actions of control and prevention (Castillo and Gómez, 2016; Bennett and Wells, 2019).

3.4.1 Entropy-based gully erosion model (EBGEM)

The EBGEM main driver of gully erosion is shear stress, and this fact explains why the novel entropy-based shear stress distribution is so advanced. Still, gully erosion is a highly complex and non-linear process in terms of time and scale (Sidorchuk, 2005). and there are also unforeseen processes that gain relevance as the drainage area grows and, consequently, a gully system becomes more complex. In Figure 3.8 we present the direct effect of a drainage area over model performance.

It can clearly been seen in Figure 3.8 that, a model experiences a monotonic performance decrease when the drainage area increases. This is most likely due to the increasing complexity
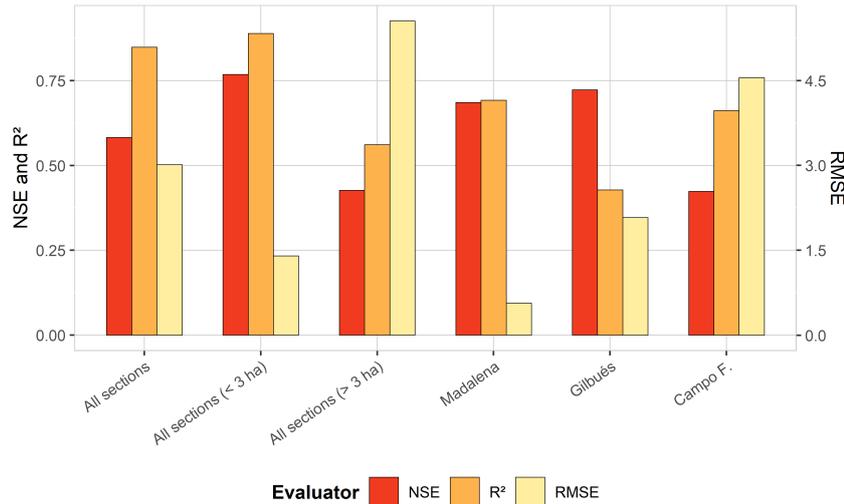
Figure 3.6: Performance evaluators (NSE, R^2 and RMSE) for the model and the three research areas.
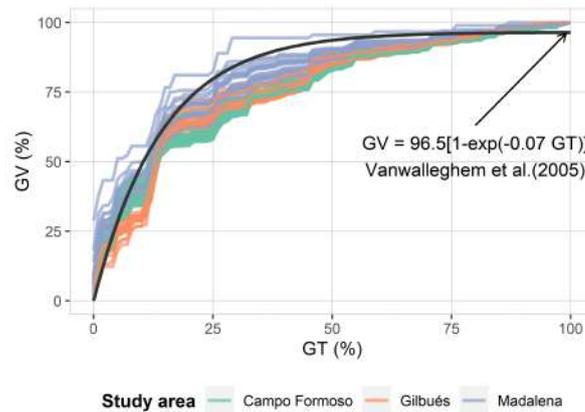
Figure 3.7: Growth dynamic of all the gullies. The black continuous line represents the model proposed by Vanwalleghem et al. (2005). *GT* is the percentage of gully age (time) over the total and *GV* the percentage of gully volume over the total.

of gully formation in larger areas, where processes that before had apparently been negligible begin to represent important processes and influence on gully erosion. Some of such processes are, among others: shielding (Mohr et al., 2021), energy dissipation due to turbulent flow processes (Tominaga et al., 1989) and turbulent bursts (Nearing, 1991). It is interesting to highlight that the larger drainage areas in Campo Formoso have steep slopes in the head of the catchment ($S > 20\%$) and low slopes in the channel ($S < 6\%$). This leads to the erosion of coarser material from higher regions of the catchment and to its posterior deposition in the channel, as observed in the field (see respective photographs in the supplementary data to this paper). The overestimation bias observed for large sections and/or drainage areas (Fig. 3.5) could be explained by an overestimation of the runoff and transport capacity.

In Figure 3.9 we compare the results of the coupled Foster-Lane-Sidorchuk Model (Alencar, de Araújo, and Teixeira, 2020) with those of the proposed Entropy-Based Gully Erosion Model (EBGEM). The FL-S Model is based on the Foster and Lane (1983) model, and uses its shear stress distribution (FLE - Eq. 3.7) to estimate erosion. It moreover includes a routine to
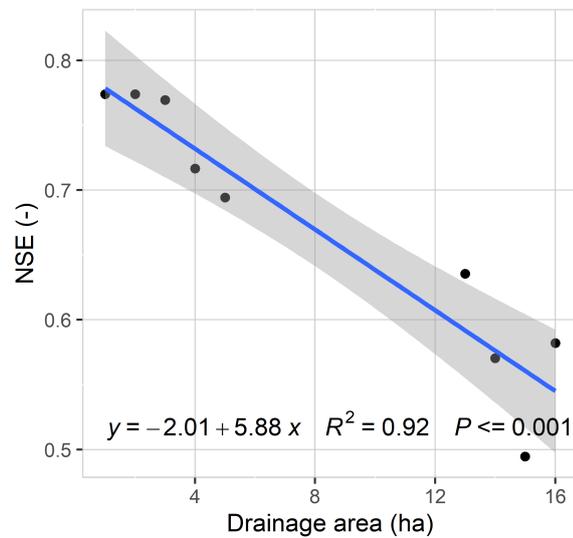
Figure 3.8: Effect of drainage area on model efficiency. Values in the x-axis indicate the maximum area considered (i.e., for 5 ha it means that all sections up to 5 ha were used in the calculation). The blue line stands for the linear regression between the area and the Nash-Sutcliffe efficiency coefficient (NSE), while the grey-shaded area represents the 95% confidence interval. Note that there is a gap in the plot because no areas between 6 and 12 ha were measured.

allow simulation of long term erosion and the Sidorchuk et al. (2003) algorithm to estimate wall-failure (Alencar, de Araújo, and Teixeira, 2020).
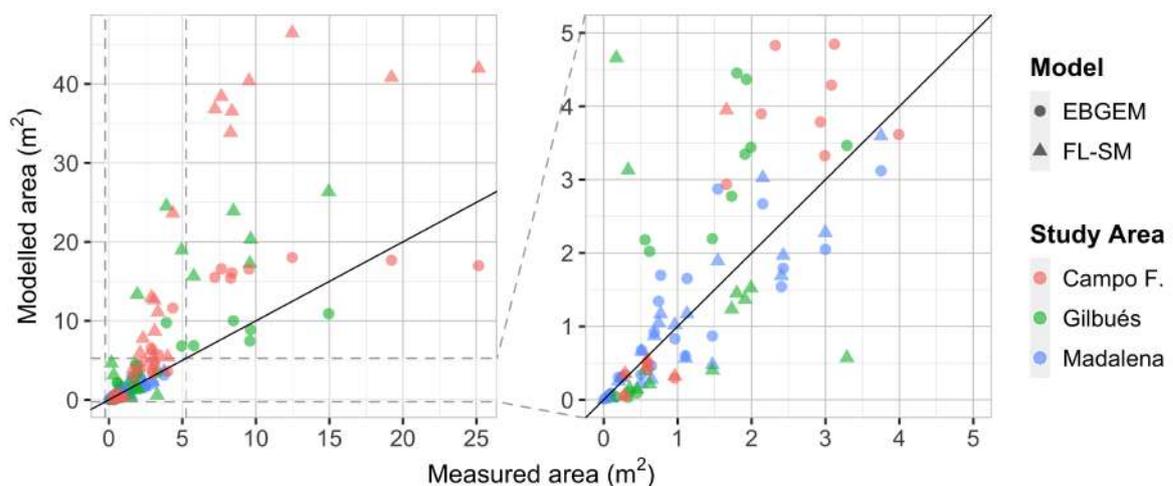
Figure 3.9: Comparison between model outputs of FL-SM (Alencar, de Araújo, and Teixeira, 2020) and EBGEM. The straight black line indicates the identity. It is easy to note that EBGEM results are better distributed around this line.

We can observe in Figure 3.9 that the FL-SM significantly overestimates the cross-section area for sections above 5 m^2 (usually associated with large catchments). This increasing difference in larger catchments and/or sections derive from the scope of the model itself. Alencar, de Araújo, and Teixeira (2020) propose the FL-SM model be employed to simulate *small permanent gullies*. The model relies on (and is highly sensitive to) an additional parameter related to the cross-section area which, for larger sections, simulates a stronger response to wall erosion. Additionally, the model based on Foster and Lane (1983) assumes that the entire section is affected by erosion when net shear stress available.

One of the improvements of the new EBGEM is that it does not depend on this particular critical area parameter. This makes the EBGEM easier to implement and suitable also for larger gully systems. Both Moriasi et al. (2007) and Ritter and Muñoz-Carpena (2013) propose a threshold of NSE = 0.65 as good/acceptable for hydrological models. Borrowing this threshold and using the results in Figure 3.8 makes the EBGEM model acceptable to be implemented in catchments measuring up to 8 hectares.

The identified upper limit of catchment area (8 hectares) is an evidence that models that rely solely on bed and wall erosion (Foster and Lane, 1983; Storm, Barfield, and Ormsbee, 1990; Sidorchuk, 1999; Torri and Borselli, 2003; Dabney et al., 2015; Alencar, de Araújo, and Teixeira, 2020) are not sufficient to model more complex gully systems, that occur in larger catchments and which channels may reach dozens of meters of depth and width. In order to model such processes, more energy and sediment inputs are required (Alonso, Bennett, and Stein, 2002; Sidorchuk, 2015; Bennett and Wells, 2019).

Nevertheless, an advantage of the EBGEM is that it is not bounded by regional features and can be easily implemented in other regions. Shear stress is based on channel geometry, differently from more conventional approaches (Storm, Barfield, and Ormsbee, 1990; Torri and Borselli, 2003; Casalı, López, and Giraldez, 2003). Critical shear stress and rill erodibility are estimated from pedotransfer functions obtained using Water Erosion Prediction Project (WEPP) equations (Alberts, Nearing, et al., 1995; Ascough et al., 1997), that were obtained from experiments in a wide range of soils. Nevertheless, when available, local values can be used. Furthermore, the limit of 8 hectares established in this paper comprises most gully systems. Multiple studies that investigated the relation of catchment area and gully erosion identified that most gullies are within this limit (Vandekerckhove et al., 2000; Nachtergaele, Poesen, Steegen, et al., 2001; Knapen and Poesen, 2010; Yibeltal et al., 2019).

Additionally, the proposed model includes a novel shear stress equation, that presents better perforce than other well published approaches (Sterling and Knight, 2002; Bonakdari, Sheikh, and Tooshmalani, 2014b) and a shorter computational time (Bonakdari, Sheikh, and Tooshmalani, 2014b; Khozani and Bonakdari, 2018) is of interest not only to gully erosion modelling, but to channel engineering in general. Shear stress distribution in channel boundary layer is important also to estimate channel durability and flow resistance (Knight, Yuen, and Al Hamid, 1994), velocity distribution (Tominaga et al., 1989), erosion and deposition (Storm, Barfield, and Ormsbee, 1990), sediment transport (Chiu, 1988), and river morphology (Khodashenas and Paquier, 1999).

3.4.2   Cross-entropy shear stress equation

Entropy-based equations constitute a significant improvement compared to the empirical equation used by Foster and Lane (1983). A disadvantage of the latter is that the distribution shape never changes and that it assumes a continuous differentiable function. Such functions, however, can only be observed in a cross-section with that has a wetted perimeter following a path that is differentiable itself (e.g. round and parabolic), as the round sections studied

by Knight, Yuen, and Al Hamid (1994). In rectangular, trapezoidal or any other section with vertexes, the shear stress distribution on the boundary layer presents a discontinuity (Chiu, 1988; Tominaga et al., 1989; Knight and Sterling, 2000). Equation 3.5 provides a distinct parameterization for each part (edge) of the section, and this significantly improves its performance.

Equation 3.5 also performed better than the maximum entropy-based equation of Sterling and Knight (2002). The Principle of Maximum Entropy (POME) is a particular case of the Principle of Minimum Cross-Entropy (POMCE), and occurs when the prior function used in POMCE is a uniform probability distribution (Kullback, 1978a). By using the POMCE and a prior function that resembles more the probability distribution of shear stress values, we are able to obtain a better fitted function while maintaining the same number of unknown parameters (Lagrange multiplier λ).

Equation 3.5 (RMSE = 0.033) also presented a better performance when compared to other entropy based models from the literature that improve upon the Sterling and Knight equation, such as Bonakdari, Sheikh, and Tooshmalani (2014b, RMSE = 0.052) and Khozani and Bonakdari (2018, RMSE = 0.072), two models based on the Tsallis and the Renyi entropy theories, respectively. Our equation has the advantage of requiring the calibration of only one parameter (Lagrange multiplier) for each stretch of the wetted perimeter (bed and wall), while both Bonakdari, Sheikh, and Tooshmalani (2014b) and Khozani and Bonakdari (2018) require the calibration of two. Therefore, less computational time was needed for our equation, and this was particularly important to this study, since it was necessary to recalculate the shear stress profile at every time step.

3.5 Conclusions

We propose a novel gully erosion model driven by net shear stress. For the assessment of shear stress exerted by water flow, our suggestion is a novel cross-entropy-based shear stress distribution function. It is a high-performing new function in comparison to independent laboratory results, and achieves a Nash-Sutcliffe Efficiency of 0.95. Our new gully erosion model (EBGEM) was tested in three gully-affected areas drainage areas ranging from 10^{-2} to 10^{+1} hectares, and cross-sections area ranging from 0.08 to 25 m^2. The model presented good results and attained NSE levels from 0.77 to 0.49. The drainage area plays a relevant role for model efficiency. While the model performs well for small catchments, its efficiency steadily drops for larger areas due to unforeseen processes. This is why it is recommended to implement it in catchment areas up to eight hectares. The Entropy-based Gully Erosion Model is easy to create and requires little parameterization, being well suited for modelling and simulation in small catchments and small permanent gullies. Future outcomes should tackle additional sediment and energy sources related to gully erosion, such as head cut, pipping, and flow jets. The entropy theory has great potential to deal with uncertainties and may be used to further advance gully modelling.

Appendix 1 - Mathematical foundation

Proof of the entropy-based shear stress equation

The Gamma distribution is very flexible and frequently used in hydrological and ecological processes (Singh, 1998). Its formulations are:

$$q(x) = Gamma(x, \alpha, \beta) = \frac{1}{\Gamma(\alpha)} \beta^{\alpha} x^{\alpha-1} e^{-\beta x} \tag{3.10}$$

where α and β are parameters. By setting α and β as 2 and 1, respectively, one finds a rather simpler distribution that was used as our prior function.

$$q(x) = Gamma(x, 2, 1) = \gamma x \, e^{-x} \tag{3.11}$$

where γ is a scaling factor $\left(\gamma = \frac{e}{e-2}\right)$ so that $\int_0^1 q(x)dx = 1$.

We can solve equation 3.2 with the Lagrange multipliers method (Eq. 3.3) and the help of selected constraints (Eq. 3.4).

(1) From the first constraint (Eq. 3.4a) we obtain:

$$\int_0^1 \frac{e}{e-2} x e^{-x} e^{-1-\lambda_0 - \lambda_1 x} dx = 1; \quad \text{let} \quad \frac{e^{-\lambda_0}}{e-2} = \psi \quad \text{and} \quad 1 + \lambda_1 = \lambda \tag{3.12a}$$

$$\psi = \frac{\lambda^2}{1 - e^{-\lambda}(\lambda + 1)} \tag{3.12b}$$

(2) And from this first constraint (Eq. 3.4b) we obtain:

$$\int_0^1 \frac{e}{e-2} x^2 e^{-x} e^{-1-\lambda_0 - \lambda_1 x} dx = \bar{x} \tag{3.13a}$$

$$\psi \int_0^1 x^2 e^{-x(1+\lambda_1)} dx = \bar{x} \tag{3.13b}$$

$$\psi \left[\frac{2 - e^{-\lambda}((\lambda + 1)^2 + 1)}{\lambda^3} \right] = \bar{x} \tag{3.13c}$$

From Equations 3.12b and 3.13c:

$$\frac{2}{\lambda} - \frac{\lambda}{e^{\lambda} - \lambda - 1} = \bar{x} \tag{3.14}$$

Finally, from the connection with the spatial domain:

$$\frac{1}{L}\frac{dy}{dx} = p(x) \tag{3.15a}$$

$$\int_0^y \frac{1}{L}dy = \int_0^x \frac{e}{e-2}xe^{-x}e^{-1-\lambda_0-\lambda_1 x}dx \tag{3.15b}$$

$$\frac{y}{L} = \psi\left[\frac{-e^{-\lambda x}(\lambda x + 1)}{\lambda^2}\right]_0^x \tag{3.15c}$$

$$e^{-\lambda x}(\lambda x + 1) = 1 - (1 - e^{-\lambda}(\lambda + 1))\frac{y}{L} \tag{3.15d}$$

Equation 3.15d cannot be further simplified, hence it becomes necessary to use numerical methods to find an explicit expression for $t(y)$.

A solution is to use the Newton-Raphson method for each y. Good approximations are expected, once the left side of Eq. 3.15d grows monotonically and has a single solution.

Obtaining constraint values

Following the same premises as Sterling and Knight (2002), Knight and Sterling (2000) and Khozani and Bonakdari (2018) we assumed a division of the cross-sectional wetted perimeter in two zones (wall and bed) and also a similar behaviour of the two zones, although controlled by different parameters, yet calibrated by using the same empirical equations (3.16) of Knight, Yuen, and Al Hamid (1994).

$$\frac{\overline{\tau_w}}{\rho g R S} = 0.01\% SF_w\left(1 + \frac{P_b}{P_w}\right) \tag{3.16a}$$

$$\frac{\overline{\tau_b}}{\rho g R S} = (1 - 0.01\% SF_w)\left(1 + \frac{1}{P_b/P_w}\right) \tag{3.16b}$$

$$\frac{\tau_{max,w}}{\rho g R S} = 0.01\% SF_w\left[2.0372\left(\frac{P_b}{P_w}\right)^{0.7108}\right] \tag{3.16c}$$

$$\frac{\tau_{max,b}}{\rho g R S} = (1 - 0.01\% SF_w)\left[2.1697\left(\frac{P_b}{P_w}\right)^{-0.3287}\right] \tag{3.16d}$$

$$\% SF_w = C_{sf}\exp(\alpha) \tag{3.16e}$$

$$\alpha = -3.23\log_{10}\left(\frac{P_b}{P_w C_2} + 1\right) + 4.6052 \tag{3.16f}$$

$$C_{sf} = \begin{cases} 1.0 \quad \forall \; P_b/P_w < 4.374 \\ 0.6603(P_b/P_w)^{0.28125} \quad \forall \; P_b/P_w \geq 4.374 \end{cases} \tag{3.16g}$$

where P_b and P_w are the wetted perimeter in bed and wall, respectively; $C_2 = 1.38$. This is how we solve the Eq. 3.15d for wall and bed zones.

On the validation of the *unique solution* of Equation 3.15d

For each $y \in [0, L]$ in Equation 3.15d, let $1 - (1 - e^{-\lambda}(\lambda + 1))\frac{y}{L}$ be equal to $-w$, where $w \in \mathbb{R}$, then:

$$f(x) = e^{-\lambda x}(\lambda x + 1) + w = 0 \tag{3.17a}$$

$$\frac{df}{dx}(x) = -\lambda^2 x e^{-\lambda x} \tag{3.17b}$$

$$\frac{d^2 f}{dx^2}(x) = \lambda^2 e^{-\lambda x}(\lambda x - 1) \tag{3.17c}$$

For a negative value of λ, the roots of equation 3.17b are 0 and $-\infty$, therefore:

$$\frac{d^2 f}{dx^2}(0) = -\lambda^2 e^0, \quad \text{which is negative} \tag{3.18a}$$

$$\frac{d^2 f}{dx^2}(-\infty) = \lambda^2 e^{+\infty}(+\infty), \quad \text{which is positive} \tag{3.18b}$$

It is also important to note that Equation 3.17b, for $x \to -\infty$ is indeterminate $(-\infty \times 0)$. Moreover, the roots of equation 3.17c are $1/\lambda$ and $+\infty$. Therefore, the function has a maximum at $x = 0$ and monotonically decreases for all $x > 0$. For $x < 0$, the function reaches an asymptote at $y = w$ as $x \to -\infty$, as illustrated below.
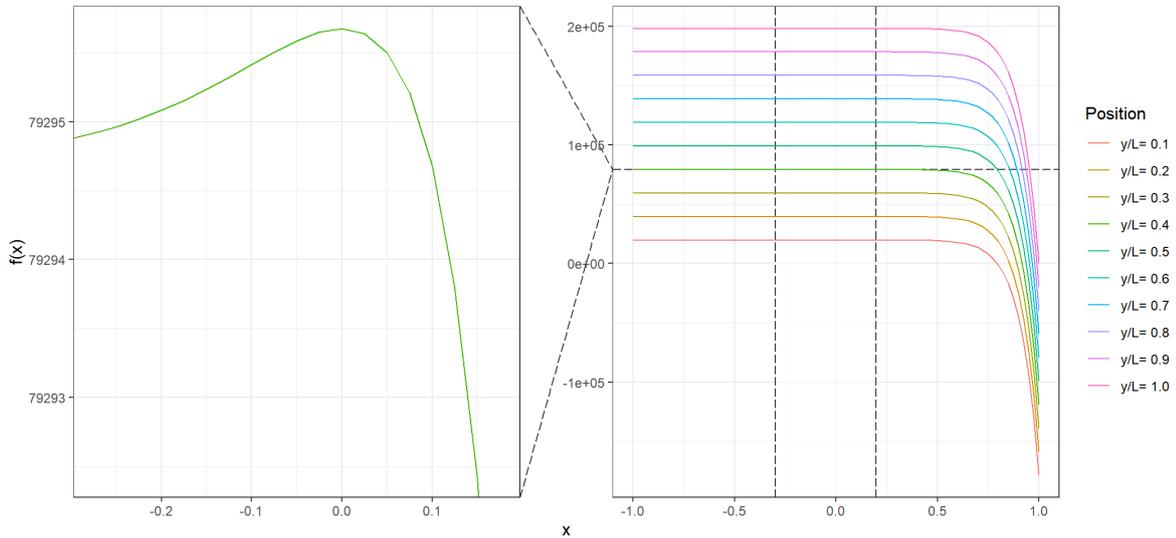
Figure 3.10: Numerical example of equation 3.17a for $\lambda = -10$. The figure illustrates the uniqueness of the solution of $f(x) = 0$.

For general cases, $\bar{x} > 0.75 x_{max}$ (Foster and Lane, 1983), therefore Eq. 3.14 yields $\lambda < 0$. In the experimental cases, $\bar{x} \in (0.87, 0.98)$. Although physically inconsistent, it is easy to show that for $\lambda > 0$ the equations still hold with a unique solution between zero and one.

Appendix 2 - List of cross-sections and their characteristics

Table 3.4: Cross-sections and their characteristic.

Site	Gully	Sec.	DA[a] (m)	P[b] (m)	L[c] (m)	K_c[d]	K_f[e]	Shape [f]	Slope	Measured A*	Measured W*	Modelled A*	Modelled W*
Madalena	Gully 1	S1	3200	433	195	2.14	698.25	v	13%	0.25	2.26	0.31	1.98
		S2	3080	416	188	2.10	700.88	v	15%	0.74	3.19	1.34	3.41
		S3	1183	337	179	2.75	156.81	v	16%	0.68	3.56	0.92	2.82
		S4	491	108	36	1.36	265.10	t	18%	0.50	3.76	0.34	1.82
		S5	78	50	23	1.58	31.29	t	12%	0.08	1.50	0.07	1.16
		S6	3060	408	185	2.07	717.47	t	15%	0.77	2.18	1.70	3.71
	Gully 2	S1	2000	383	123	2.40	347.81	t	3%	3.00	8.93	2.05	4.14
		S2	1870	379	114	2.45	310.52	t	3%	2.43	7.32	1.79	3.89
		S3	1470	336	100	2.45	244.14	t	3%	1.11	6.89	0.60	2.62
		S4	252	80	24	1.41	126.88	t	5%	0.64	2.95	0.46	2.04
		S5	40	44	12	1.93	10.69	t	6%	0.01	0.48	0.01	0.27
		S6	1930	381	119	2.43	327.30	t	3%	2.40	7.44	1.54	3.60
		S7	1425	320	92	2.37	252.94	t	3%	1.47	6.16	0.87	2.83
	Gully 3	S1	3500	554	204	2.62	509.10	v	4%	1.13	6.32	1.65	4.19
		S2	3400	540	193	2.59	505.65	t	4%	0.96	5.54	0.83	3.39
		S3	1000	248	105	2.20	207.39	v	4%	0.20	2.18	0.31	2.05
		S4	56	49	14	1.84	16.59	t	8%	0.10	1.30	0.09	1.19
		S5	2230	473	186	2.80	283.51	v	4%	0.51	3.02	0.67	2.88
		S6	4800	620	213	2.51	764.51	v	4%	3.75	6.06	3.12	5.12
		S7	4964	639	218	2.54	769.74	t	4%	2.15	4.68	2.67	4.86
		S8	4712	609	205	2.48	763.59	v	4%	1.54	5.33	2.87	4.96
Gilbués	Gully 1	S1	2388	434	147	2.49	385.74	t	16%	5.77	6.51	6.85	7.21
		S2	2238	384	129	2.27	434.41	t	16%	4.95	5.84	6.80	7.18
		S3	2156	359	120	2.16	460.30	t	17%	3.91	3.97	9.76	8.16
		S4	1783	322	114	2.14	390.41	t	17%	1.93	2.69	4.37	6.30
		S5	1440	285	105	2.11	324.85	t	18%	1.99	2.94	3.44	6.00
		S6	916	234	83	2.16	195.91	t	20%	1.91	4.14	3.35	5.82
		S7	517	168	62	2.07	120.98	t	23%	1.80	2.68	4.45	6.07
	Gully 2	S1	277	165	68	2.77	36.07	v	8%	0.56	2.35	2.18	5.57
		S2	201	133	59	2.63	29.10	t	7%	0.35	2.17	0.11	0.34
		S3	150	160	49	3.66	11.16	v	6%	0.62	2.64	2.02	5.70
		S4	96	113	37	3.23	9.15	v	6%	0.45	2.03	0.09	0.16
		S5	64	68	22	2.40	10.99	v	6%	0.17	1.86	0.04	0.16
	Gully 3	S1	4414	407	169	1.72	1499.41	t	7%	14.95	11.68	10.90	8.64
		S2	3531	352	149	1.66	1286.78	t	7%	9.64	10.35	8.86	7.96
		S3	3065	328	139	1.66	1114.93	v	7%	9.59	9.53	7.44	7.46
		S4	2754	300	121	1.60	1078.37	v	8%	8.47	10.83	10.01	8.28
		S5	2008	97	106	0.61	5413.43	t	8%	1.73	4.61	2.77	6.52
		S6	393	261	55	3.68	28.92	v	8%	1.47	4.55	2.20	6.22
		S7	53	131	22	5.01	2.13	v	10%	0.33	3.68	0.03	0.22
		S8	157	69	29	1.53	66.74	v	24%	3.29	6.11	3.46	5.58
	Gully 1	S1	27631	1022	220	1.72	9319.94	v	7%	3.08	6.47	4.29	6.78
		S2	29194	1037	231	1.70	10105.06	v	7%	2.93	7.91	3.79	6.54
		S3	33257	1103	254	1.69	11597.34	v	6%	3.99	9.55	3.62	6.63
		S4	37631	1189	295	1.72	12770.22	v	6%	3.34	5.56	5.74	7.76
		S5	35756	1144	279	1.69	12467.05	v	6%	3.12	6.79	4.85	7.20
		S6	37787	1196	311	1.72	12724.11	v	6%	2.89	2.80	6.50	8.33
		S7	137562	2891	738	2.18	28879.38	v	7%	7.65	7.24	16.58	12.89
		S8	147719	3024	796	2.20	30428.23	v	6%	8.28	7.67	15.41	12.47
		S9	159125	3169	892	2.22	32160.02	v	6%	19.22	13.96	17.66	13.28
		S10	141781	2912	771	2.17	30237.01	v	7%	8.37	7.04	16.08	12.69
		S1	129219	1659	665	1.29	77391.74	t	5%	25.15	16.67	16.98	13.28
		S2	124000	1580	610	1.26	78562.11	t	5%	7.21	9.47	15.50	12.58
		S3	126406	1607	650	1.27	78959.70	t	5%	12.47	8.70	18.03	13.94

3. Entropy-based Model for Gully Erosion – a Combination of Probabilistic and Deterministic Components

Site	Gully	Sec.	DA[a] (m)	P[b] (m)	L[c] (m)	K_c[d]	K_f[e]	Shape [f]	Slope	Measured A*	Measured W*	Modelled A*	Modelled W*
Campo Formoso	Gully 2	S4	128125	1625	630	1.27	79343.67	t	5%	9.53	10.48	16.55	13.06
		S5	122344	1554	590	1.24	79047.76	v	5%	4.35	5.43	11.62	11.02
		S6	43125	1024	448	1.38	22609.34	v	7%	2.32	5.75	4.83	7.47
		S7	40781	969	420	1.34	22582.82	v	7%	2.13	5.36	3.90	7.17
		S8	38438	918	380	1.31	22371.65	v	7%	1.66	6.60	2.94	7.09
		S9	23906	688	273	1.25	15409.29	t	9%	2.99	8.55	3.33	6.21
	Gully 3	S1	1922	272	124	1.73	638.85	t	8%	0.29	3.52	0.34	2.50
		S2	2611	297	136	1.63	987.53	v	7%	0.59	5.76	0.50	2.73
		S3	1768	228	105	1.52	766.36	v	8%	0.59	3.59	0.41	2.49
		S4	1194	176	79	1.42	590.82	t	9%	0.96	6.19	0.30	2.23
		S5	75	37	14	1.21	51.25	v	10%	0.28	2.84	0.05	1.35

[a]DA: Drainage area in square metres

[b]P: Perimeter of the drainage area in metres

[c]L: Length of the drainage area in metres

[d]K_g: Gravelius coefficient $\left(0.28\frac{P}{DA}\right)$

[e]K_f: Form coefficient $\left(t\frac{DA}{P}\right)$

[f]Shape: Indicate the general cross-section shape. "v" indicates V-shaped sections and "t" trapezoidal-shaped cross-sections.

*Area (A) and width (W) of cross-sections in m^2 and m respectively.

4

Entropy-based method for temporal downscaling of precipitation to improve sediment delivery ratio assessment

Abstract

Many regions around the globe are subjected to precipitation-data scarcity that often hinders the capacity of hydrological modeling. The entropy theory and the principle of maximum entropy can help hydrologists to extract useful information from the scarce data available. In this work, we propose a new method to assess sub-daily precipitation features such as duration and intensity based on daily precipitation using the principle of maximum entropy. Particularly in arid and semiarid regions, such sub-daily features are of central importance for modeling sediment transport and deposition. The obtained features were used as input to the SYPoME model (sediment yield using the principle of maximum entropy). The combined method was implemented in seven catchments in Northeast Brazil with drainage areas ranging from 10^{-3} to 10^{+2} km^2 in assessing sediment yield and delivery ratio. The results show significant improvement when compared with conventional deterministic modeling, with Nash–Sutcliffe efficiency (NSE) of 0.96 and absolute error of 21% for our method against NSE of -4.49 and absolute error of 105% for the deterministic approach.

Keywords: maximum entropy; sediment transport; sediment yield; hydrology

4.1 Introduction

Climate change challenges our capacity to preserve natural resources, such as clean water and productive soil. The Food and Agriculture Organization named erosion as one of the most relevant threats to soil conservation and agriculture (FAO, 2019). Climate change is blamed for erosion rates increasing by nearly 17% in the USA and Europe until 2050 due to higher rainfall erosivity (Nearing, Pruski, and O'neal, 2004; Panagos, Ballabio, et al., 2017). This is why soil erosion turned into a key challenge for the Sustainable Development Goals of the UN (Keesstra et al., 2016; Borrelli et al., 2017). Soil erosion also imposes a threat to water supply, as pollutants and heavy metals are transported along with sediment, augmenting toxicity, turbidity and eutrophication in aquatic environments (Coelho et al., 2017; Li, Peng, et al., 2020).

In addition, 30% of all land on Earth has an arid or a semiarid climate (Sivakumar, Das, and Brunini, 2005), which causes some places to be especially vulnerable to climate change and soil erosion (Huang et al., 2015). Special attention is required for semiarid regions, since they house and sustain over 14% of the global population and around 70% of the dry-land population (Huang et al., 2015). Arid and semiarid areas are commonly affected by data scarcity, particularly in Africa, Asia and South America (Sanyal, Densmore, and Carbonneau, 2014; Worqlul et al., 2017; Rezende de Souza et al., 2021). It is necessary to improve sedimentological and other models in order to better estimate the amount of sediment reaching water bodies. Modelers normally have information only on daily precipitation data, yet sub-daily processes play a crucial role in sediment transport, as a substantial amount occurs during high-intensity storms (Srinivasan and Galvão, 2003; Shrestha et al., 2019). Therefore, we need a methodology to downscale precipitation duration and to improve erosion models at the sub-daily scale.

Diverse branches of water sciences point out the use of stochastic methods in hydrology as being the next generation of models (Sidorchuk, 2009; Singh, 2018). In this context, a powerful tool deployed in several studies over the last decades is the principle of maximum entropy (PoME—Shannon, 1948; Jaynes, 1957a). The first applications of the PoME in water sciences were proposed by Chiu (1987) and by Singh and Chowdhury (1985) for modeling velocity distribution in open channels. Since then, several other applications in hydrology, hydraulics and sedimentology have been presented (Sterling and Knight, 2002; Singh, 2011; Cui and Singh, 2013; Chen, Singh, and Xiong, 2017; Kumbhakar, Ghoshal, and Singh, 2020).

de Araújo (2007) proposed a PoME-based model to assess sediment yield and reservoir siltation. The model (sediment yield using the principle of maximum entropy—SYPoME), however, requires sub-daily data, such as rainfall duration and intensity measurements, which are often unavailable in arid and semiarid regions (Pilgrim, Chapman, and Doran, 1988), such as the Brazilian northeast region. According to the Brazilian Water Management Agency (ANA, 2019), the country's semiarid region has 2163 operating rainfall stations connected to the national weather monitoring system, which averages one rain gauge per 462 km². Most of those instruments are standard Ville de Paris gauges, providing only daily precipitation. Only

36 are active and reliable automatic stations providing sub-daily precipitation data—one every 27,800 km^2, on average (Figure S1—Supplementary material). The gauging station density is much lower than in other regions (e.g., the density of automatic stations is one per 3600 km^2 in the United States and 77 km^2 in Italy—NOAA, 2013; Baldassarre, Castellarin, and Brath, 2006). The data series are also not long; only 16 stations have more than 15 years of continuous data.

The Brazilian northeast (10^6 km^2) has an average annual temperature varying between 20 and 28 °C and is characterized by a high temporal and spatial rainfall variability (Medeiros and Araújo, 2014), with average annual rainfall between 400 mm and 800 mm (increasing towards the coast—Cadier, 1994; Andrade et al., 2020) and evapotranspiration between 2000 and 2600 mm per year (de Figueiredo et al., 2016). The vegetation is mainly Caatinga, formed by deciduous broadleaf bushes. The largest part of the region is placed over Precambrian crystalline bedrock with shallow soils. In these areas, groundwater is scarce and usually salty (Gaiser et al., 2003; Marengo, Alves, et al., 2013). The simultaneous occurrence of such geological features, concentrated precipitation patterns and high evaporation rates leads to a scenario where rivers are predominantly intermittent (Montenegro and Ragab, 2012). As a result, water for over twenty million people living in the Brazilian northeast region is mainly supplied by reservoirs (Coelho et al., 2017). The region has a concentration of reservoirs as high as one per 5 km^2 (Mamede et al., 2012). Due to excessive erosion and eutrophication, however, reservoir siltation is one of the key threats to the water supply in the region (Coelho et al., 2017).

Our objectives are as follows: (1) to propose a temporal down-scaling method to estimate sub-daily precipitation data from daily precipitation data based on the principle of maximum entropy (MEDRID); (2) to assess the method quality when implemented on ungauged regions (spatial-scalability); and (3) to evaluate the effect of the method on the performance of long-term sediment yield modeling.

In order to achieve these objectives, measured data of high-resolution precipitation were used to calibrate and validate the MEDRID method, and the statistical distance measures after Kullback (1978b) and Fedotov, Harremoes, and Topsoe (2003) were used to assess spatial scalability. Measured sediment yield data of seven catchments of different sizes and series durations were employed to test and validate the improved sediment yield modeling using scaled precipitation together with the model by de Araújo (2007), which is based on entropy equations and quantifies gross erosion by means of the universal soil loss equation (USLE).

4.2 Materials and Methods

Sediment yield can be quantified by multiplying gross erosion and sediment delivery ratio (SDR—Maner, 1958; Sharda and Ojasvi, 2016; Llena et al., 2021). These terms are highly nonlinear, and deterministic models do not always account for their uncertainties (Sidorchuk, 2009; Royall and Kennedy, 2016; Llena et al., 2021). Therefore, such processes need to be

modeled stochastically and event-wise (Sidorchuk, 2009; Gupta et al., 2020). In this study, the sediment yield of sub-daily events was quantified using the principle of maximum entropy (PoME). To incorporate sub-daily rainfall information, we developed temporal-downscaling equations to assess the effective rainfall duration (D) and its respective 30-minute intensity (I_{30}). As proposed by de Araújo (2007), the rainfall duration was drawn on to calculate the SDR, and the I_{30} to calculate the erosivity factor of the universal soil loss equation (Wischmeier and Smith, 1978), so as to assess gross erosion.

A new method (Figure 4.1) was proposed to estimate sediment yield: it consists of an entropy-based approach to downscale rainfall duration and intensity (the MEDRID—maximum entropy distribution of rainfall intensity and duration method). We coupled MEDRID with the SYPoME model to determine an event-wise SDR (de Araújo, 2007).
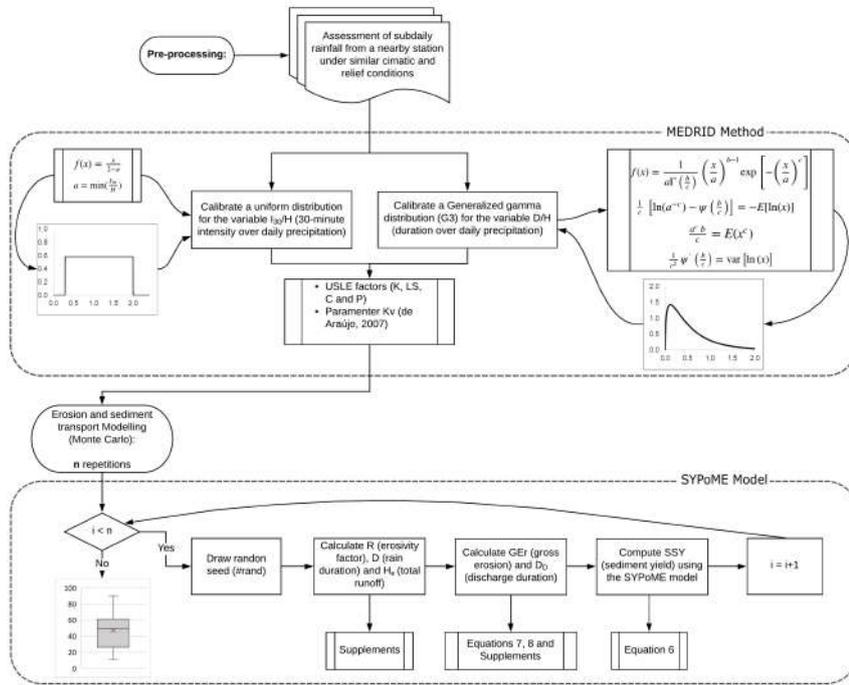
Figure 4.1: Flowchart of the proposed model. The processing is divided in two main parts, the MEDRID method and the SyPOME model. The two parts are coupled by a Monte Carlo process with multiple random seeds generated.

4.2.1 Maximum Entropy Distribution of Rainfall Intensity and Duration— MEDRID Method

Two sub-daily variables were selected to be assessed from daily rainfall data: (1) the duration–precipitation ratio D/H (D for duration and H for total daily precipitation) and (2) intensity–precipitation ratio I_{30}/H (where I_{30} stands for 30-minute intensity). Three probability density functions were tested to fit D/H frequencies: the beta (B3), the gamma (G2) and the generalized gamma (G3) distributions (Stacy, 1962; Chen, Singh, and Xiong, 2017). For the intensity–precipitation ratio (I_{30}/H), two probability density functions were tested: the beta (B3) and the uniform distribution. After calibrating the equations using the principle of

maximum entropy (Singh, 1998), we tested the best fitting equations to measured data, as well as spatial scalability.

Table 4.1 presents the three probability density functions (PDFs—beta, gamma and generalized gamma), their constraints and the respective system of equations for parameterization. $\Psi(\cdot)$ is the digamma function, the first derivative of $\Gamma(\cdot)$, the gamma function. $\Psi'(\cdot)$ is the tri-gamma function, the second derivative of $\Gamma(\cdot)$. The terms a, b and c in the three distributions are parameters obtained maximizing entropy using the Lagrange multipliers method (Kumbhakar, Ghoshal, and Singh, 2020). The systems of equations in Table 4.1 can be solved using empirical data (e.g., rain gauge readings, as for this study—Singh and Chowdhury, 1985). The parameter r in the beta distribution (B3) is a scale factor. For this specific distribution, the random variable $X \in [0, 1]$. The systems of equations were solved with help of the software Octave (v. 5.1.0.0).

Additionally, sub-daily data are scarce and stations may cover a large area. It is important to assess the loss in performance of the method when using data from a distant station. This loss of performance can be measured as the difference between the calibrated PDF for the weather station and the expected PDF, if the region of study had such a station. In this study we compared the variations among four stations with sub-daily data (Aiuaba, Sobral, Sumé and Gilbués) using the Kullback–Leibler divergence (Kullback and Leibler, 1951) and the Kolmogorov–Smirnov distance (Kolmogorov, 1933; Smirnov, 1939). These statistical measures allow us to find similarities between the areas, and therefore to determine which areas can be modeled with which calibrated PDF without a significant performance loss.

Let m and n be two populations (sets)—in our study, automatic stations—each with an associated PDF p_m and p_n. Kullback and Leibler (1951) present a measure that allows us to compare how different those two distributions are. Known as the Kullback–Leibler divergence, the D_{KL} is an asymmetric measure, given by Equation (4.1).

$$D_{KL}(P_m \parallel P_n) = I(m:n) = \int_0^{+\infty} p_m(x)\ln\left[\frac{p_m(x)}{p_n(x)}\right] dx \tag{4.1}$$

$$J(m,n) = \frac{I(m:n) + I(n:m)}{2} \tag{4.2}$$

where p_m and p_n are continuous probability distributions. $I(m:n)$ can be understood as the loss of information if the population m is modeled using p_n instead of p_m. Furthermore, ref. Kullback, 1978b introduces a symmetric measure, given by Equation (4.2). $J(m,n)$ is also a measure of divergence between the distributions p_m and p_n and can be interpreted as how easily we can distinguish the two distributions, henceforth called symmetric divergence.

The Kolmogorov–Smirnov distance (δ—Equation (4.3)) is the maximum distance between two distributions in their domain and is related to the Kullbach–Leibler divergence by Pinsker's inequality (Equation (4.4)).

$$\delta(P_m, P_n) := sup\left\{\left|\int_0^x p_m(x)dx - \int_0^x p_n(x)dx\right|\right\} \tag{4.3}$$

Table 4.1: Parameterization of PDFs Beta (B3), Gamma (G2) and Generalized Gamma (G3). We present the list of constrains used for each equation and the obtained system after solving with the Lagrange multipliers method.

Equation		B3	G2	G3
PDF		$f(x) = \dfrac{\Gamma(a+b)}{\Gamma(a)+\Gamma(b)} \left(\dfrac{x}{r}\right)^{a-1} \left(1-\dfrac{x}{r}\right)^{b-1}$	$f(x) = \dfrac{1}{a\Gamma(b)} \left(\dfrac{x}{a}\right)^{b-1} \exp\left(-\dfrac{x}{a}\right)$	$f(x) = \dfrac{c}{a\Gamma\left(\frac{b}{c}\right)} \left(\dfrac{x}{a}\right)^{b-1} \exp\left[-\left(\dfrac{x}{a}\right)^c\right]$
constrains	$i.$	$\displaystyle\int_0^1 f\left(\dfrac{x}{r}\right) d\left(\dfrac{x}{r}\right) = 1$	$\displaystyle\int_0^{+\infty} f(x)dx = 1$	$\displaystyle\int_0^{+\infty} f(x)dx = 1$
	$ii.$	$\displaystyle\int_0^1 \dfrac{x}{r} f\left(\dfrac{x}{r}\right) d\left(\dfrac{x}{r}\right) = E\left[\ln\left(\dfrac{x}{r}\right)\right]$	$\displaystyle\int_0^{+\infty} x f(x)dx = E(x)$	$\displaystyle\int_0^{+\infty} x^q f(x)dx = E(x^q)$
	$iii.$	$\displaystyle\int_0^1 \ln\left(1-\dfrac{x}{r}\right) f\left(\dfrac{x}{r}\right) d\left(\dfrac{x}{r}\right) = E\left[\ln\left(1-\dfrac{x}{r}\right)\right]$	$\displaystyle\int_0^{+\infty} \ln(x) f(x)dx = E[\ln(x)]$	$\displaystyle\int_0^{+\infty} \ln(x) f(x)dx = E[\ln(x)]$
system	$i.$	$E\left[\ln\left(\dfrac{x}{r}\right)\right] = \psi(a) - \psi(a+b)$	$ab = \bar{x}$	$\dfrac{1}{c}\left[\ln(a^{-c}) - \psi\left(\dfrac{b}{c}\right)\right] = -E[\ln(x)]$
	$ii.$	$E\left[\ln\left(1-\dfrac{x}{r}\right)\right] = \psi(b) - \psi(a+b)$	$\psi(b) - \ln(b) = E[\ln(x)] - \ln(\bar{x})$	$\dfrac{a^c b}{c} = E(x^c)$
	$iii.$			$\dfrac{1}{c^2}\psi'\left(\dfrac{b}{c}\right) = \text{var}[\ln(x)]$

$$\delta(P_m, P_n) \leq \sqrt{\frac{1}{2} D_{KL}(P_m || P_n)} \tag{4.4}$$

It is also important to note that J is not an actual distance, while δ is. The PDFs obtained for each of the four stations will be compared pairwise. The lower the values of D_{KL} and δ are, the more alike are the two distributions and the lower the loss of information is between the areas.

4.2.1.1 Other literature approach

de Araújo (2017) also attempted to assess event duration using stochastic modeling using Equations (4.5 to 4.7). D is duration and H daily precipitation. $S_\bullet$ is the standard deviation of the sample. j is a counter index (j-th event). χ is a random number such that $\chi^j \in [0, \chi_{max}]$. χ_{max} is calibrated for each watershed. The author proposes that for each event j, at least 20 values of χ^j should be drawn. The simulated duration D would be the arithmetic average of the 20 produced results.

$$D^j = \bar{D} + k^j S_D \tag{4.5}$$

$$k^j = \frac{H^j - \bar{H}}{S_H} \chi^j \tag{4.6}$$

$$\frac{\bar{D} - D^j}{\bar{H} - H^j} = \frac{S_D}{S_H} \chi^j \tag{4.7}$$

4.2.2 Sediment Yield-PoME – SYPOME Method

de Araújo (2007) proposed an entropy-based model for event-based SDR (Equation (4.8)) and sediment yield (SSY – Mg km^{-1} yr^{-1}). $\bar{\varepsilon}$ (Mg km^{-1} yr^{-1}) is the gross erosion obtained, for example, by using the universal soil loss equation –USLE (Wischmeier and Smith, 1978), L_0 the hill slope length (m), L_m the maximum sediment travel distance (m), x_0 the initial position of erosion in the hillslope and λ a Lagrange multiplier.

$$SSY = \bar{\varepsilon} \times \text{SDR} = \bar{\varepsilon} \times \frac{e^{\lambda L_m} (L_0 - x_0) \lambda - \left(e^{\lambda(L_0 - x_0)} - 1\right)}{\lambda L_0 \left(e^{\lambda(x_0 + L_m)} - 1\right)} \tag{4.8}$$

The SDR is the ratio of sediment yield (SSY) and mobilized sediment ($\bar{\varepsilon}$). The SDR is physically constrained to a closed interval ($SDR \in [0, 1]$), and it can be interpreted as the average probability of a detached particle reaching the river system (de Araújo, 2007).

The SYPoME model uses as input the duration of the sub-daily precipitation which, in our case, is not known. The MEDRID method can solve this gap, based on daily precipitation.

4.2.3   Monte Carlo and MEDRID-SYPoME coupling

A Monte Carlo approach was used to adapt the SYPoME model (de Araújo, 2007) and its output to an interval of possible values of sediment yield associated to a probability function (Vrugt et al., 2008). The results were compared with measured data from seven catchments (Figure 4.2 and Table 4.2) and values from the literature model (Maner, 1958).

Using the MEDRID method we can find the probability distribution function (PDF) for the duration–precipitation ratio $\frac{D}{H}$. To model the inherent uncertainty of the duration–precipitation ratio we used the Monte Carlo approach. For each event in the time interval Δt, a large number of random seeds ($\#_{rand} \in [0,1]$—Equation (4.9)) are generated and used as input in the calibrated PDF to assess the duration (Figure 4.1).

$$\#_{rand} = F\left(x \leq \frac{D}{H}\right) = \int_0^{D/H} f(x)dx \tag{4.9}$$

where f is the calibrated PDF according to Table 4.1 and F the associated cumulative distribution function of x. Solving Equation (4.9) for D/H, with known H, we can obtain the rainfall duration for each random seed $\#_{rand}$. The set of pairs (D, H) is used as input for the SYPoME model.

4.2.4   Gross erosion and siltation assessment

To estimate gross erosion in the catchments we used the universal soil loss equation (Equation (4.10)—Wischmeier and Smith, 1978; Bagarello, Ferro, and Pampalone, 2020). A more detailed description of each factor and the values for the study areas can be found in the supplements to this paper. Siltation (ΔV) and sediment yield are proportional and related according to Equation (4.11).

$$\bar{\varepsilon} = R\,K\,LS\,C\,P \tag{4.10}$$

$$SSY = \frac{\Delta V\,\rho_s}{\eta\,A\,\Delta t} \tag{4.11}$$

where ΔV is the volumetric siltation, or the reservoir capacity loss (in m^3), ρ_s is the bulk density of the silted sediment (in Mg m^{-3}), η the trap efficiency of the reservoir (using, e.g., the method by Brune, 1953), A is the catchment area in hectares and Δt the interval of time in analysis.

In order to assess the performance gain by using the MEDRID+SYPoME model, we compared the measured data with empirically based SDR equations (Sharda and Ojasvi, 2016).

Gaiser et al. (2003) found that, for the Brazilian northeast region, the most fit among those equations is the one by Maner Maner, 1958, (hereafter Equation (4.12)). Simplıcio et al. (2020) had the same result for the dry Cerrado region of Gilbués (Figure 4.2).

$$\text{SDR} = \exp\left[2.943 - 0.824\log_{10}\left(\frac{F_L}{F_R}\right)\right] \qquad (4.12)$$

F_L (m) is the length factor, measured as the maximum distance in the catchment with a straight line from the outlet to the water divide approximately parallel to the main river. F_R (m) is the relief factor, calculated as the difference between the outlet altitude and the average altitude of the water divide.

4.2.5 Study area

We selected seven catchments in three different states of the Brazilian northeast, all under dry conditions (Figure 4.2) to test the method approach for precipitation downscaling (MEDRID) and the sediment yield assessment model (SYPoME). The catchments vary widely in area and availability of data (number of years in a time series). They also vary in terms of land use and land cover. The characteristics of the studied catchments are listed in Table 4.2.

The Brazilian northeastern region houses the country's semiarid region (BSh climate, according to the Köpper Classification—Gaiser et al., 2003) and the Caatinga Biome. The Caatinga is the largest tropical dry forest in the world and houses the highest endemic genera of all (Miles et al., 2006; Silva and Souza, 2018). The main economic activities in the region are agriculture (especially maize, beans and soybeans), livestock and fishing (Coelho et al., 2017). Due to deleterious practices in agriculture and overgrazing, the degraded area surpassed 72,000 km^2 in the Brazilian Drylands (ca. 8% of its original area—Tomasella et al., 2018).

As presented in Section 4.1, the Brazilian northeast region suffers with data scarcity concerning sub-daily rainfall events. Therefore the selection is restricted to the existing (and operating) stations. The stations in Gilbués, Aiuaba and Sumé (Figure 4.2) were maintained by research groups (Simplıcio et al., 2020; de Figueiredo et al., 2016; Srinivasan and Galvão, 2003) and only the station of Sobral is maintained by the Brazilian Water Management Agency (ANA). Those four stations presented consistent measurements over at least two years without gaps. Another constraint for the selection of stations was the proximity to the sediment control equipment. Again, the stations in Gilbués, Aiuaba and Sumé were installed to monitor experimental basins and are inside the catchment areas. The Sobral station was chosen because it is in the Várzea da Volta catchment and is the closest to Acarape under the same climate conditions. For a detailed map of stations in the region please refer to the Supporting Materials.

Experimental data were used to estimate sediment yield (Morris and Fan, 1998). We used bathymetric assessments from different years of the reservoirs of five catchments (Canabrava, Aiuaba, Várzea da Volta, Acarape and Gilbués) to estimate the total siltation (ΔV—see

Equation (4.11)). Direct data for sediment yield (SSY) were available at the micro-basins in Sumé, where monitoring is carried out eventwise (Srinivasan and Galvão, 2003). Table 4.2 lists the type and timing of available sediment yield data. For each catchment we obtained the time series of daily rainfall from FUNCEME (2019). Sub-daily measurements are scarce and available for the whole study period only in one station in Gilbués (Simplıcio et al., 2020) and one in Aiuaba (de Figueiredo et al., 2016), the basins with the shortest and most recent time series. Assuming similar climatic and environmental conditions, we used the data from the Aiuaba station for the analysis of Canabrava, and from Várzea da Volta for Acarape.
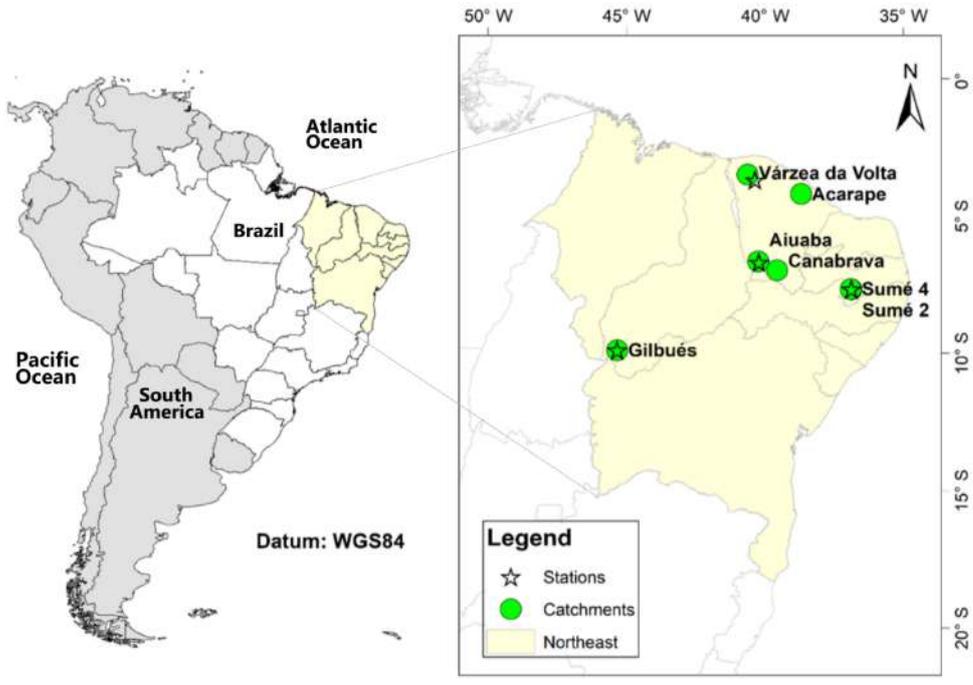
Figure 4.2: Location of study areas (catchments) and automatic rain gauges. All areas are located in the Brazilian northeast.

4.3 Results

4.3.1 Probability distributions functions - MEDRID

Table 4.3 presents the entropy-based calibrated parameters for B3 (beta distribution), G2 (gamma distribution) and G3 (generalized gamma distribution). Those values were obtained by solving the systems of equations in Table 4.1. In Figure 4.3 we present the model evaluators of distributions at the four stations. From the method evaluators we can observe that B3 represents poorly the distribution when compared with the gamma distributions (Figure 4.3). G3 performs slightly better than G2. From Table 4.3 we see that the parameter c of the generalized gamma does not sufficiently approach the unit (when $c = 1$, the gamma and generalized gamma are equal). The strict two-parameter gamma distribution (G2) does not quite represent the process, but less skewed function G3 does.

Table 4.2: Study areas information. Lines with same colour indicate areas that share automatic rain gauge data.

Basin	Area (km²)	Control system	Land use	Location	Catchment position		Bathymetry		Time Series	Automatic weather station				
					Lon	Lat	First[a]	Second		Name	Position Lon	Lat	Recording Start	End
Canabrava	2.9	Reservoir	Agriculture and open range cattle rasing	Ceará	39.56 W	6.97 S	1944	2000	57	Aiuabá	40.22 W	6.69 S	2004	2014
Aiuaba	11.53	Reservoir	Conservation area with native vegetation (Caatinga)	Ceará	40.24 W	6.65 S	2003	2009	7	Aiuabá	40.22 W	6.69 S	2004	2014
Várzea da Volta	155	Reservoir	Agriculture and open range cattle raising	Ceará	40.62 W	3.50 S	1917	1997	81	Sobral	40.36 W	3.69 S	2017	2019
Acarape	208	Reservoir	Agriculture and open range cattle raising	Ceará	38.69 W	4.20 S	1924	1999	74	Sobral	40.36 W	3.69 S	2017	2019
Sumé 2	0.0107	Sediment load	Experimental area - Preserved vegetation	Paraíba	36.88 W	7.67 S	-	-	10	Sumé	36.88 W	7.67 S	1982	1991
Sumé 4	0.0048	Sediment load	Experimental area - Degraded land without vegetation	Paraíba	36.9 W	7.66 S	-	-	10	Sumé	36.88 W	7.67 S	1982	1991
Gilbués	0.0004	Check dam	Abandoned land under desertification process without vegetation	Piauí	45.34 W	9.88 S	2018	2019	1	Gilbués	45.34 W	9.88 S	2018	2019

[a] The first bathymetry corresponds to the topography in the year of construction.

4. Entropy-based method for temporal downscaling of precipitation to improve sediment delivery ratio assessment

Table 4.3: Equation parameters for the D/H distribution. a, b and c are the parameters as described in Table 4.1. The data used to calibrate the parameters are available in the supplementary material.

	B3		G2		G3		
	a	*b*	*a*	*b*	*a*	*b*	*c*
Sobral	1.124	4.316	0.250	1.525	0.066	2.114	0.678
Aiuaba	1.584	10.686	0.138	1.855	0.004	3.306	0.488
Gilbués	0.696	2.691	0.777	0.953	0.390	2.099	0.812
Sumé	0.955	5.398	0.740	0.911	0.269	1.410	0.818

Two probability distribution functions were tested for the ratio I_{30}/H. The beta distribution (B3) and uniform distribution allow an explicit definition of lower and upper boundaries. For the Sobral, Aiuaba and Gilbués stations the uniform distribution presented much better results, with Nash–Sutcliffe efficiency (NSE) as high as 0.98, while the beta distribution had an efficiency lower than 0.50 (Figure 4.4). In the Sumé station both B3 and uniform distributions had similar performance with NSE of 0.98 and 0.99, respectively. In this work we used the uniform distribution for the modeling in all regions.

Additionally, using statistical measures, we calculated the information loss resulting from using the PDF calibrated for one region into another (Equations (4.2) and (4.3)). We compared the four stations with sub-daily data among themselves. The measures (symmetric divergence and Kolmogorov–Smirnov distance) for the variable D/H are given in Table 4.4.

Table 4.4: Values of symmetric divergence and Kolmogorov–Smirnov distance for the generalized gamma distribution of D/H. The higher the value, the greater the difference between the probability distributions.

(a) Symmetric Divergence

	Sobral	Aiuaba	Gilbués	Sumé
Sobral	0	0.198	1.210	0.097
Aiuaba	0.198	0	2.494	0.536
Gilbués	1.210	2.494	0	0594
Sumé	0.097	0.536	0.594	0

(b) Kolmogorov–Smirnov Distance

	Sobral	Aiuaba	Gilbués	Sumé
Sobral	0	0.242	0.550	0.152
Aiuaba	0.242	0	0.719	0.365
Gilbués	0.550	0.719	0	0.404
Sumé	0.152	0.365	0.404	0

These measures indicate that there is a considerable difference in the duration–precipitation (D/H) distribution in Gilbués over the other three regions.

Sobral and Sumé also appear to be very similar, despite the distance between them. Located in the Brazilian Semiarid Region, the stations in Sobral, Sumé and Aiuaba are under the same major atmospheric process for rainfall formation (the Inter-Tropical Convergence Zone—ITCZ) and have a similar rainfall regime (more than 70% of the annual precipitation concentrated in three months) and amount (500–600 mm yr^{-1}). Gilbués has a higher precipitation rate (1200 mm yr^{-1}) and better temporal distribution. Therefore, based on statistical distances (Table 4.4) and regional characteristics, Sobral and Sumé are most similar and have the lowest information loss when (quality) data from one station are used for the other region. Aiuaba is also similar to Sumé and (especially) to Sobral. Gilbués has particular PDF parameters, with both D_{KL} and δ significantly higher when compared with the other three stations.
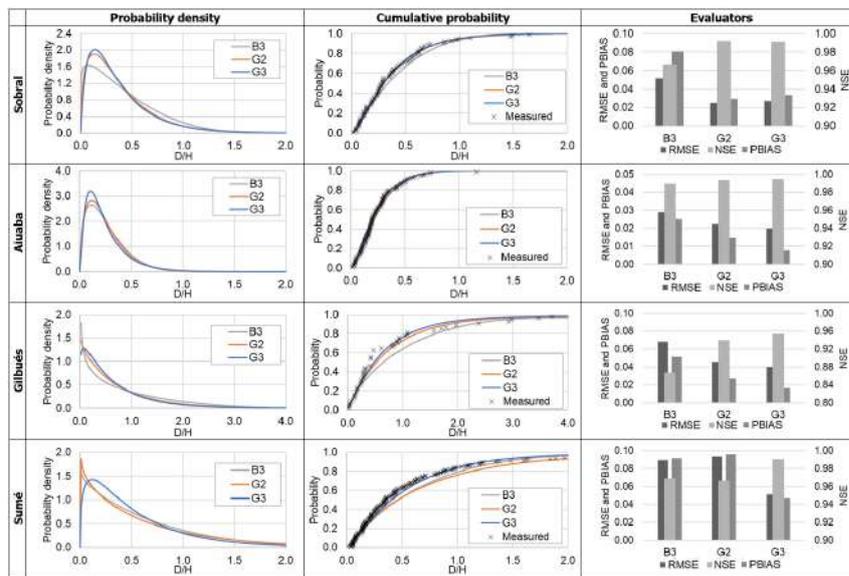
Figure 4.3: Probability distributions and the performance evaluators for the variable D/H.
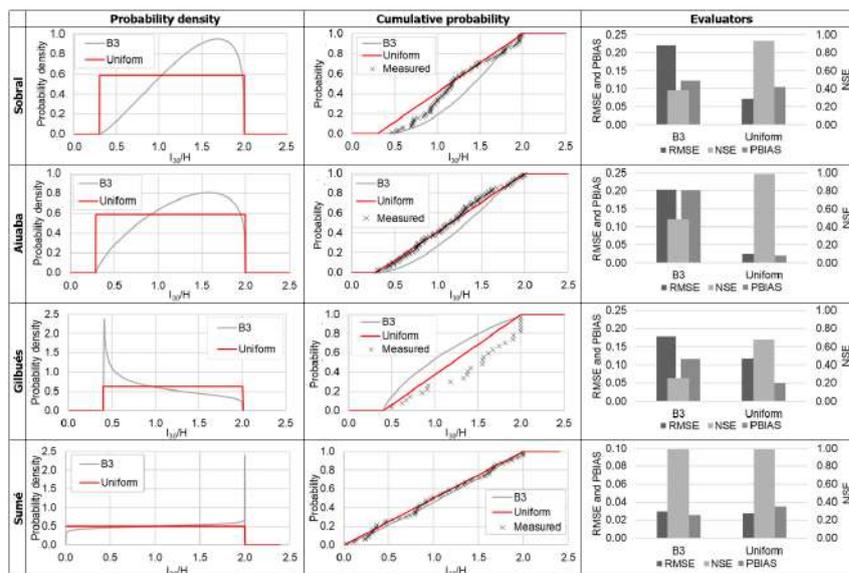
Figure 4.4: Probability distributions and the performance evaluators for the variable I_{30}/H.

4.3.2   Sediment yield modelling

Two models were tested to assess sediment yield: a classic model consisting of the multiplication USLE gross erosion ($\bar{\varepsilon}$) and empirically based SDR (Maner, 1958), hereby called model M1, and the proposed MEDRID+SYPoME model (M2).

In Table 4.5 we present the output of the combination of the MEDRID method and SYPoME model (M2) for the seven study areas. Average modeled sediment yield at the outlet varied between 5 (Aiuaba) and 2346 (Sumé 4) Mg km^{-2} yr^{-1} and SDR between 5.9% (Várzea da Volta) and 29.7% (Gilbués). The outputs for sediment yield and SDR of model M2 passed the normality test Shapiro and Wilk, 1965 and we obtained the confidence interval ($p = 0.01$) using a Gaussian distribution. M1 is a deterministic model, thus it has only one single output, presented in Figure 4.5.

Table 4.5: Modeled values (M2) of sediment yield and SDR for the study areas. The values are shown in terms of average (μ), standard deviation (σ) and coefficient of variation (CV). Confidence intervals (CIs) of the average calculated for p = 0.01.

Basin	Sediment yield (Mg km^{-2} yr^{-1})				SDR (%)			
	μ	σ	CV	CI	μ	σ	CV	CI
Canabrava	664.5	24.9	4%	12.5	13.9	0.2	1.4%	1.04
Aiuaba	5.0	1.2	25%	0.6	14.8	4.2	28.4%	2.12
Várzea da Volta	418.2	20.2	5%	10.1	5.9	0.4	7.3%	0.22
Acarape	189.5	9.1	5%	3.1	8.3	0.7	8.1%	0.23
Sumé 2	13.1	1.8	14%	0.9	23.5	2.6	11.1%	1.32
Sumé 4	2345.6	264.1	11%	132.9	20.4	3.0	14.6%	1.50
Gilbués	2141.7	540.5	25%	272.0	29.7	8.9	29.9%	4.47

In Figure 4.5, we present two plots. Figure 4.5a shows modeled (M1 and M2) and measured values of siltation rate (siltation rate per unit of area) and Figure 4.5b the modeled (M1 and M2) values of SDR. The siltation rates generated by our approach (M2) clearly outperform those based on deterministic methods (M1). When assessing average sediment yield for each area, our model also outperforms the deterministic model for all experimental basins, with an error reduction by a factor of at least 2 and as high as 20 (Table 4.6). In addition, the new methodology (M2: MEDRID+SYPoME) presented better performance evaluators (NSE = 0.96 and RMSE = 608.6 ton km^{-2} yr^{-1}) than the conventional (M1) approach (NSE = -4.49 and RMSE = 3286 Mg km^{-2} yr^{-1}).
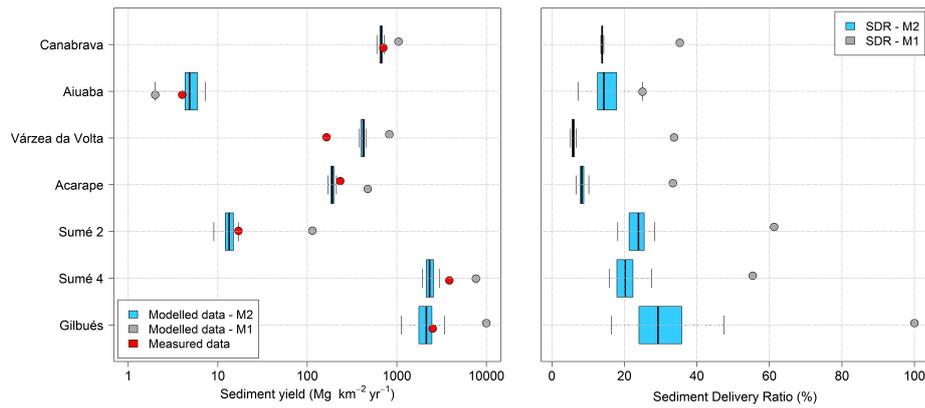
Figure 4.5: M1 and M2 outputs of (**a**) sediment yield and (**b**) SDR. Red dots in (**a**) indicate the measured values of sediment yield.

By comparing the values of siltation rate in Figure 4.5a with land use and land cover (Table 4.2) we can draw a strong correlation between them. Catchments with preserved vegetation, such as Aiuaba and Sumé 2, have the lowest siltation rate, over two orders of magnitude lower than degraded regions, such as Sumé 4 and Gilbués. Basins with the presence of agriculture (Canabrava, Várzea da Volta and Acarape) presented intermediary rates, although ten times larger than preserved regions.

Figure 4.5b shows the modeled average SDR (for the whole time series) of the basins obtained by M2 and M1 (Equation (4.12)). Considering the area of the basins (Table 4.2), we can observe a dependency of the SDR to the catchment area. Although M2 also showed a similar tendency, its values of SDR are systematically lower than M1's. It is interesting to note that for the catchments Canabrava, Acarape and Várzea da Volta there is almost no dispersion of SDR values. This is due to the long time series for those experimental areas. With a long temporal series, the averaging of the SDR of all events tends to a narrow range of values that can be understood as the basin SDR. Additionally, the Maner equation (Equation (4.12)) allows values of SDR numerically larger than 100 %, which is inconsistent with the physical interpretation of SDR. Whenever the calculated SDR was larger than the physical limit, the value was limited to 100 %, as is the case of Gilbués.

4.4 Discussion

The complexity of hydrological processes can be better modeled with the help of stochastic approaches (Sidorchuk, 2009; Singh, 2011). Ref. Sidorchuk, 2009 proposed a path for sedimentological models relying on the combination of deterministic and probabilistic models in a so-called third-generation erosion model, to which our method belongs. By introducing stochastic routines and calibrating parameters with the principle of maximum entropy, we extracted from the scarce data more valuable information than by employing deterministic models, and even preserved the local characteristics of each region. The method performed well across a large range of time series and catchment-area scales.

Table 4.6: Measured and modeled values of siltation rate (Mg km^{-2} yr^{-1}). M1 represents the classic model using empirically based SDR (Maner, Equation (4.12)) and M2 the proposed MEDRID+SYPoME model.

Name	Brune Coefficient	Sediment Yield (Mg km^{-2} yr^{-1})			Relative Error (%)	
		Measured	Modeled M1	Modeled M2	Modeled M1	Modeled M2
Canabrava	0.98	704	1042	664	*48.0%*	*−5.6%*
Aiuaba	1.00	4	2	5	*−50.0%*	*27.5%*
Várzea da Volta	0.95	164	824	418	*402.3%*	*155.1%*
Acarape	0.98	233	473	191	*102.9%*	*−18.1%*
Sumé 2	1.00	17	114	13	*570.6%*	*−21.8%*
Sumé 4	1.00	3857	7644	2314	*98.2%*	*−40.0%*
Gilbués	1.00	2518	10305	2142	*309.3%*	*−14.9%*
NSE			**−4.49**	**0.96**		

4.4.1 Probability distributions functions - MEDRID

In the literature (Singh, 1998; Bhunya et al., 2007; Brigandì and Aronica, 2019; Martinez-Villalobos and Neelin, 2019) many probability distribution functions are related to precipitation processes (e.g., gamma, power-law, exponential); especially concerning its duration (e.g., gamma, Weibul, lognormal). From Figure 4.3, we conclude that, although the gamma distribution (G2) does reproduce the D/H ratio, the generalized gamma distribution yields the best results in all study areas. Its better fit to the measured data appears to be related to the high complexity (uncertainty/entropy) involved in rainfall events, when many factors interact simultaneously. In such conditions, a less constrained distribution such as the G3 allows for more flexibility and calibration. With one additional parameter, the function becomes more adaptable to the peculiarities of each region in comparison with G2. This is confirmed by the values obtained for the parameter *c*, which never approximate to one (Table 4.3). Table 4.1 shows that a parameter c equal to one reduces a generalized gamma distribution to a conventional one (G2).

Information entropy is a measure of uncertainty (Jaynes, 1957a). Therefore, the PoME delivers the probability distribution function that maximizes the uncertainty under a set of constraints and avoids unproven assumptions (Chiu, 1991). It can be proven that the uniform distribution, such as the one obtained for $\frac{I_{30}}{H}$, has the highest uncertainty (see Jaynes, 1957a).

In the selection of the best distribution using the PoME, additionally to the constraints listed in Table 4.1, there is an implicit assumption taken: that the data follow a specified distribution (i.e., beta, gamma, uniform, etc.). Silva Filho, de Araújo, and Raabe (2020) pointed out that the selected constraints of the PoME have to be relevant to the studied variable and that additional constraints do not necessarily lead to better results. Therefore, as we see in Figures 4.3 and 4.4, the narrowest distribution does not necessarily best suit the model. The constraints-quality trade-off problem becomes clear in the modeling of rain intensity (Section 4.3.1), where the most suitable distribution is the uniform one. Such a result occurs because the unproven implicit constraint (the distribution itself) is shown not to be valid.

The use of a uniform distribution for intensity implies that a stochastic approach is more valid than regression curves, as previously proposed by Avila and Avila, 2015; Alencar, de Araújo, and Teixeira, 2020; Dash, Das, and Adhikary, 2019. Therefore, in stochastic models, a more realistic approach to be adopted is the uniform distribution, as expressed in Equation (4.13). The value of 30-min intensity (I_{30}) can vary between 0—in the case of $H \to 0$—and 2H (for a precipitation with duration lower than 30 min). Equation (4.13) is a general equation and does not depend on calibration. Nevertheless, the implementation of Equation (4.13) also requires a Monte Carlo approach, as presented in Section 4.2.3, with drawing of multiple random seeds ($\#_{rand}$).

$$I_{30} = \frac{H}{D} + H\left(2 - \frac{1}{D}\right)\#_{rand} \text{ such that } \frac{I_{30}}{H} \in (0,\, 2] \qquad (4.13)$$

In terms of regionalization of the MEDRID method, equations calibrated using data from a gauged catchment can be used in ungauged regions, provided that they have similar relief and climatic conditions, thus reducing the loss of information. It is important to note that geographic proximity between the station and the application site is not enough to guarantee better parameter homogeneity and, thus, good model performance. The equations from Sumé and Sobral are remarkably similar, although they are more distant from each other than to Aiuaba. Nevertheless, the conditions of the Aiuaba catchment, which is higher and prone to orographic precipitation, may explain its distinction from the others. Finding the causes of similarities between areas, however, surpasses the scope of this work. Still, from analysis of relief and climate of the studied areas and based on the statistical distances (Table 4.4), we can build a map of possible factors that influence such similarity (Figure 4.6). The relative position of each area in Figure 4.6 is based on geographical location. The connecting lines indicate how similar the areas are to each other.
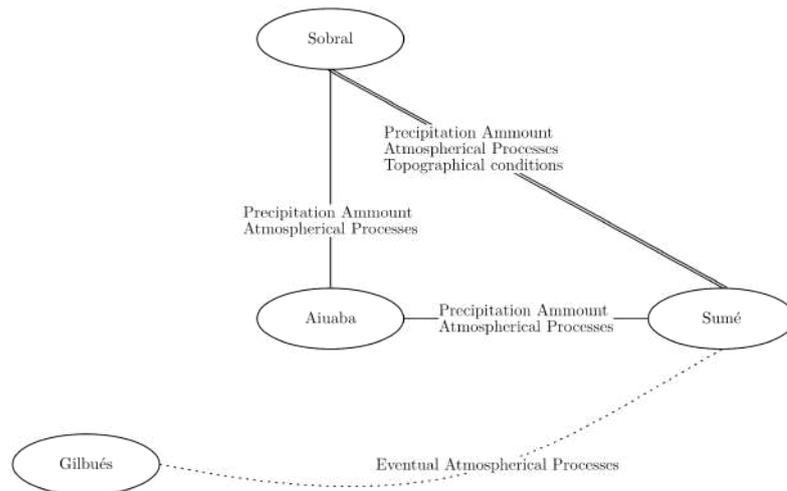
Figure 4.6: Clustering (connections) of regionalized PDFs and possible influencing factors for the similarities, based on relief and climate conditions. Note that nodes are positions to roughly match the geographical location of each study area (no scale—Figure 4.2).

De Araújo's (2017, see Section 4.2.1.1 of this paper) method of precipitation down-scaling, although simpler, has two problems. Firstly, each precipitation event is processed by the model only once, using an averaged duration as input. This reduces the freedom of the model to simulate extreme cases. The model by de Araújo (2017) also tends to represent the process by a linear function, after the averaging (Figure 4.7). Secondly, the author's approach assumes a normal distribution of duration and daily precipitation. It is also assumed that both distributions are related by an unknown scaling factor χ (Equation (4.7)). None of these assumptions could be confirmed by experimental data.
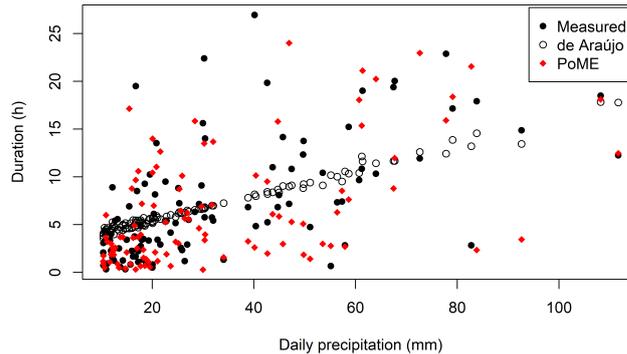
Figure 4.7: Scatter plot of daily precipitation and duration for Aiuaba. Note that both methods depend on random seeds, therefore the points' position in the plot is not fixed, but rather an example. Other examples are available in the supplementary material.

4.4.2 Sediment yield modelling

In all cases the MEDRID+SYPoME model (M2) performed better than the deterministic model (M1) with empirically based SDR. As shown in Figure 4.5 and in Table 4.6, the relative error was reduced nine-fold, on average. Except for Várzea da Volta, the average error was 21%, five times smaller than the average error for M1. When also excluding Várzea da Volta, the performance of M1 was similar to values obtained from the literature, see (Risse et al., 1993). The Nash–Sutcliffe efficiency of event-wise sediment yield calculated for the catchments of Sumé 2 and 4 (0.52 and 0.47, respectively) can be classified as satisfactory since its efficiency is marginally equal to 0.50 (Moriasi et al., 2007). These are, nonetheless, important results, especially considering the little information required to achieve them. The efficiency of the model for total siltation rate is 0.96 (Table 4.6); its classification ranges from very good (Moriasi et al., 2007) to good (Ritter and Muñoz-Carpena, 2013). This supports the argument that stationary parameters such as relief (in our temporal analysis scale) play a relevant role for sediment delivery mechanisms (Simplicio et al., 2020); they therefore increase the performance of the model over time.

Both models perform poorly in the assessment of siltation of the Várzea da Volta reservoir (see also Gaiser et al., 2003). This is mainly caused by the peculiarity of its catchment topography and lithology. As illustrated in Figure 4.8, the upper (southern) part of the watershed is formed by a plateau ending in a cliff of over 500 meters in depth formed by soil that is prone to erosion (USLE parameter K = 0.032 Mg h MJ^{-1} mm^{-1}—Gaiser et al.,

2003). The lower portion of the watershed is mostly flat, and its soil has a higher permeability, promoting an interruption of connectivity and therefore reducing the SDR, similar to the process identified by Medeiros and Araújo (2014) in a flat area upstream of the Benguê Reservoir in north-eastern Brazil. Our model (M2) was not able to describe such behavior, although it significantly reduces the error when compared to the conventional methodology (M1).

One limitation of this study is the use of the universal soil loss equation to assess gross erosion. The USLE does not directly address gully erosion (Wischmeier and Smith, 1978). Nevertheless, gullies may be major sediment sources (Bennett and Wells, 2019), especially in degraded areas such as Sumé 4 and Gilbués (Srinivasan and Galvão, 2003; Simplıcio et al., 2020).
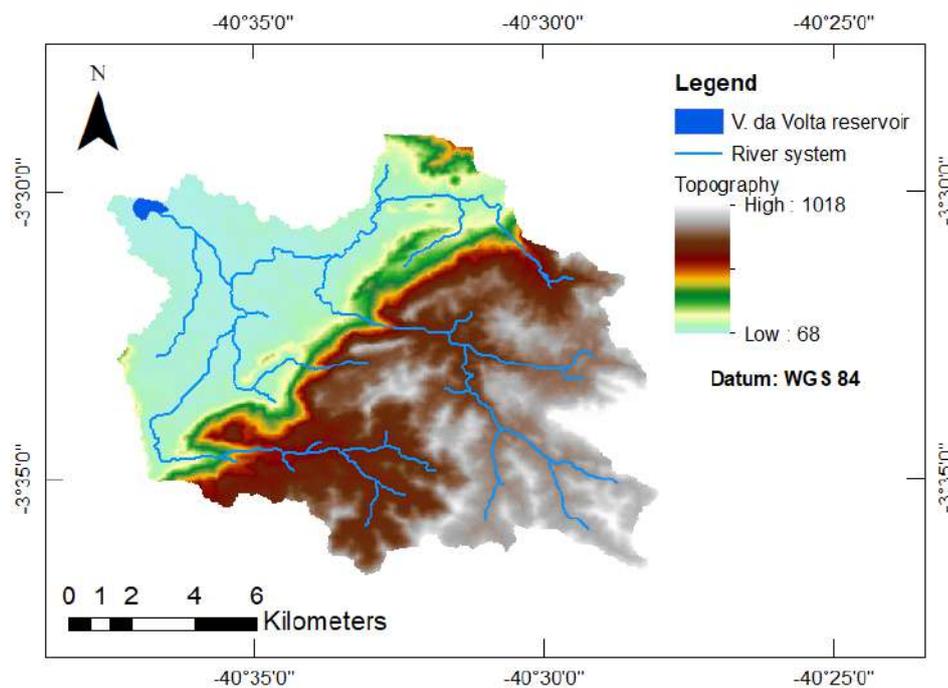
Figure 4.8: Topography and river system of the Várzea da Volta Catchment area.

4.5 Conclusions

We have proposed a novel method to downscale duration and intensity of precipitation for erosion modeling based on daily data. The best probability distribution function for the duration–precipitation ratio (D/H) is the generalized gamma distribution (NSE = 0.98). For the ratio I_{30}/H, the uniform distribution (NSE = 0.47) performs best. The MEDRID method presents resilience to regionalization, therefore demanding fewer climatological stations to cover a large area and allowing the implementation of the model in regions with data scarcity.

Using the downscaled duration and I_{30} intensity generated by MEDRID, we are able to assess sediment yield with a higher accuracy than conventional USLE and relief-based SDR. The coupling MEDRID+SYPoME model allowed assessment of event-wise sediment yield

and presented errors that were six times smaller than the ones from conventional models. The new model (MEDRID+SYPoME), based on the combination of deterministic and entropy-based components, improved substantially the performance of assessment of sediment yield (NSE = 0.96) when compared with deterministic modeling.

Additional studies should be carried out to test and assess the most suited probability distribution families to precipitation data, especially 30-min intensity. Efforts are still necessary to validate the method's potential concerning regionalization. It is not at all a trivial matter to determine which factors (relief, climate, position, etc.) influence homogeneity between regions, and therefore produce similar PDFs.

The MEDRID method can be used to assess rainfall sub-daily features (duration and 30-min intensity). When coupled as MEDRID+SYPoME, the novel model provides accurate results for sediment yield across a wide range of catchment areas in catchments with areas of different orders of magnitude (from 10^{-3} to 10^{+2} km^2) and land use.

5

How do we identify flash droughts? A study case in Central Europe Croplands

Abstract

Many definitions and delineation methods exist for identifying flash droughts, which are events of rapid and unusually large depletion of root-zone soil moisture, in comparison to average moisture conditions, due to climatic compound conditions over a short period of several weeks. Six flash drought identification methods were compared to analyse their functioning using data from several experimental cropland sites across Central Europe. Co- and misidentification of the flash drought time series were assessed using confusion and synchronicity metrics. Even though a large degree of synchronicity of individual flash drought events was observed, some divergence in drought periods was detected, which was related to four intrinsic differences in the underlying flash drought definitions: (1) type of critical variable, (2) velocity of drought intensification, (3) pre-set threshold values for final depletion, and/or (4) minimum length of the duration of flash droughts. To balance the strengths and weaknesses of those methods that are not soil moisture-based, we suggest using an ensemble approach for event identification. In doing so, the current unclearly defined sub-types of flash droughts can be detected, regardless of the different combinations of compound drivers and differences in intensification dynamics. All methods were implemented in an R package and are available as a Shiny app for the public.

5.1   Introduction

Droughts are among the most extreme climatic weather events that threaten food security (FAO, 2021). They have negative impacts on the global food-energy-water nexus and the Sustainable Development Goals (D'Odorico et al., 2018). Droughts are generally characterised by unusually high levels of rainfall deficit, runoff deficit or soil moisture deficit (Palmer, 1965; Mishra and Singh, 2010) and are expected to increase in many regions of the world in terms of frequency, severity, and duration under current and future climate change conditions (Lesk, Rowhani, and Ramankutty, 2016; Samaniego, Thober, et al., 2018).

Flash droughts (FDs) are a special form of drought. In contrast to the descriptions of classical droughts (de Araújo and Bronstert, 2015; Oikonomou et al., 2020), FDs are characterized by a rapid onset and relatively short durations (Otkin, Svoboda, et al., 2018; Lisonbee, Woloszyn, and Skumanich, 2021). They are associated with severe and immediate soil moisture depletion, resulting in plant water stress and mortality (Ford and Labosier, 2017; Osman et al., 2021; Liang and Yuan, 2021)

One of the first flash drought studies was by Peters et al. (2002), who were among the first to study single short drought events in late summer, characterised by the concurrence of low antecedent moisture and unusually high temperature. Interest in FDs has increased over the last few years, motivated by extreme flash drought occurrences in the USA, Russia, and China, which caused extreme impacts on managed vegetation, disruption to the global food supply, and increased wildfires (Christian, Basara, Otkin, and Hunt, 2019; Otkin, Anderson, et al., 2013; Mo and Lettenmaier, 2016; Christian, Basara, Otkin, Hunt, et al., 2019; Liang and Yuan, 2021). A flash drought in Australia during the spring of 2019 is thought to have played a central role in the massive forest fires that consumed over 1.6 million hectares (Nguyen, Wheeler, Hendon, et al., 2021).

Over the last two decades, multiple methods have been proposed to identify flash drought events (Lisonbee, Woloszyn, and Skumanich, 2021), yet there is no consensus on what a flash drought entails and how they may be defined in terms of onset, duration, velocity of intensification, and absolute or relative changes (Osman et al., 2021). In their recent literature review, Lisonbee, Woloszyn, and Skumanich (2021) identified as many as 20 studies with different definitions using climate variables or indexes related to soil moisture, air temperature, precipitation, and actual and potential evapotranspiration. Eleven of these definitions included an interval of intensification or rapid onset as part of flash flood delineation, whereas nine studies merely considered short-term drought events as flash droughts. A method comparison study by Osman et al. (2021) identified four core types where the definition of a heatwave flash drought (Mo and Lettenmaier, 2015) was based on temperature anomalies, rapid soil drying (Ford and Labosier, 2017; Yuan et al., 2019), actual and/or potential evapotranspiration

anomalies (Christian, Basara, Otkin, and Hunt, 2019; Pendergrass et al., 2020), and multi-criteria indexes (Chen, Gottschalck, et al., 2019). For the United States, they showed that flash drought frequency, spatial extent, and onset would vary significantly depending on which definition is used. They also suggested that a root-zone soil moisture-based method effectively captures flash drought onset in both humid and semi-arid regions. Other than the study by Osman et al., which focused on the identification of spatial differences in delineated flash droughts, there have been no systematic studies on the temporal divergence and synchronicity of delineated flash droughts. At the same time, little is known about flash drought dynamics in Central Europe and it is not known which flash drought method would apply to this region, given that the present methods have so far been used for flash drought identification mainly outside Europe.

Due to the lack of a general definition, we define a flash drought as the process of rapid, accelerated, and unusually large depletion of root-zone soil moisture, in comparison with "average" moisture conditions, due to the simultaneous or concurrent occurrence of two or more atmospheric and/or weather conditions over a short period of several weeks during the main growing season.

The objective of this study was to compare the functioning of six recently developed flash drought identification methods with data from four well-monitored experimental cropland sites in Central Europe, by assessing co- and misidentification of flash drought time series using similarity and synchronicity metrics. We selected two soil moisture-based methods (Osman et al., 2021; Ford and Labosier, 2017) and four indirect methods that used single or multiple climatic variables or indices for flash drought delineation (Noguera, Castro, and Serrano, 2020; Pendergrass et al., 2020; Christian, Basara, Hunt, et al., 2020, and a Multi-Criteria method by the authors). The methods were implemented in an R package and a Shiny App available to the public.

5.2   Materials and Methods

5.2.1   Flash Drought Identification Methods

The following six flash drought identification methods were selected on the basis that they used station data as input and, following our definition, included a clear definition of the rapid onset of water limitation. The first two methods are soil moisture-based and the other four used indirect proxies of drought conditions, such as anomalies of rainfall, temperature, and the ratio of actual and potential evapotranspiration. The methods used are described as follows:

M1: Osman et al. (2021) borrowed the concept of volatility from stock market analysis techniques for the analysis of rapid soil moisture changes. According to them, a flash drought occurs when the one-pentad (5 d) running average for root-zone soil moisture content falls below the four-pentad (20 d) running average for a period of at least four pentads, with soil

moisture at the end of this period dropping below the 20th percentile for that time of year (Figure 5.1a).

M2: Ford and Labosier (2017) identified flash droughts as periods when the pentad-average 0–40 cm volumetric water content declines, from at least the 40th percentile to below the 20th percentile, in four pentads or less (Figure 5.1b).

M3: The Multi-criteria method is a new method that uses a set of 10 anomalies and indexes derived from weekly precipitation, temperature, and potential evapotranspiration data. It calculates a score for each week equivalent to the proportion of indicators that meet or surpass the respective pre-set thresholds. Weeks with a score higher than 0.65 and an absolute change of score ($\Delta score$) higher than 0.25 over up to 3 weeks are classified as flash droughts (Figure 5.1c). The event duration is computed as the time from the beginning of intensification until the score is below 0.65. A full description of this method is provided in Appendix 5.4.

M4: Christian et al. (2020) used the standardized evaporative stress ratio (SESR), which is derived as the z-score of the quotient of actual to potential evapotranspiration rate values for a specific pentad. They used four criteria which flash drought events were required to have: 1) a minimum length of five SESR changes, equivalent to a length of six pentads (30 d minimum length), 2) a final SESR value below the 20th percentile of SESR values, 3) SESR changes must be at or below the 40th percentile between individual pentads and no more than one SESR change above the 40th percentile following the previous criterion, and 4) an overall mean change in SESR during the entire length of the flash drought must be below the 25th percentile in SESR (Figure 5.1d).

M5: Noguera et al. (2020) used the standard precipitation evapotranspiration index (SPEI) on a short timescale (1 month) and performed calculations based on a temporal frequency of 1 week (four per month). To identify the rapid onset of a drought event, the change in the SPEI for each week, in periods of 4 weeks, was calculated and the onset of a flash drought was defined as involving a change in SPEI equal to or less than –2 SPEI units (z-values) over an intensification period of 4 weeks. Further, final SPEI values had to be equal to or less than –1.28 SPEI units (Figure 5.1e).

M6: Pendergrass et al. (2020) utilised the evaporative demand drought index (EDDI) following Hobbins et al. (2016) and Lukas, Hobbins, and Rangwala (2017), which is calculated based only on the potential evapotranspiration using the Penman–Monteith equation. The method identifies a flash drought when a 50% increase in EDDI over 2 weeks is sustained for at least another 2 weeks (Figure 5.1f).

Workflow and key characteristics are summarised in Figure 5.1 and Table 5.1. Readers interested in a more detailed description of the percentiles and thresholds are referred to the original papers.

The six methods used one or more climate variables (rainfall, temperature, soil moisture, and actual and potential evapotranspiration). Further, they all share an underlying set of characteristics:

A. Flash droughts evolve rapidly, with an intensification period lasting between 2 and 4 weeks.

B. The final conditions at the end of a flash drought lean toward extreme values, often characterised by the variable reaching values under the 20th percentile or, in some, a *z-score* value over ± 1.

C. Flash droughts are considered seasonal phenomena and are identified based on the expected values of climatic variables for each specific time of the year subdivided either in pentads or weeks

D. Flash droughts depend on crossing certain thresholds and are thought to be correctly identified if, and only if, environmental conditions meet a set of predefined rules.

The key differences between the methods are how the flash drought variables and durations are defined. Methods M1-Osman et al. and M2-Ford Labosier take a direct approach to assess plant water availability using soil water data, whereas all other methods use proxy variables, which are likely to be less accurate, while simultaneously overcoming severe data limitations. Additionally, differences exist in the definition of the onset of the flash drought periods, the time resolution used (weeks and pentads), the minimum period over which it should be sustained, and the maximum duration beyond which it might be considered a "normal" drought (Table 5.1).

All six methods were implemented in R programming language, and are organised in an R-package named "fdClassify"[1] and in a Shiny-App named "FD-Viz"[2].

---

[1]Link to fdClasify git repository: `https://github.com/pedroalencar1/fdClassify`
[2]Link to Flash Drought Visualization tool (FD-Viz): `https://pedroalencar.shinyapps.io/FD-Viz/`
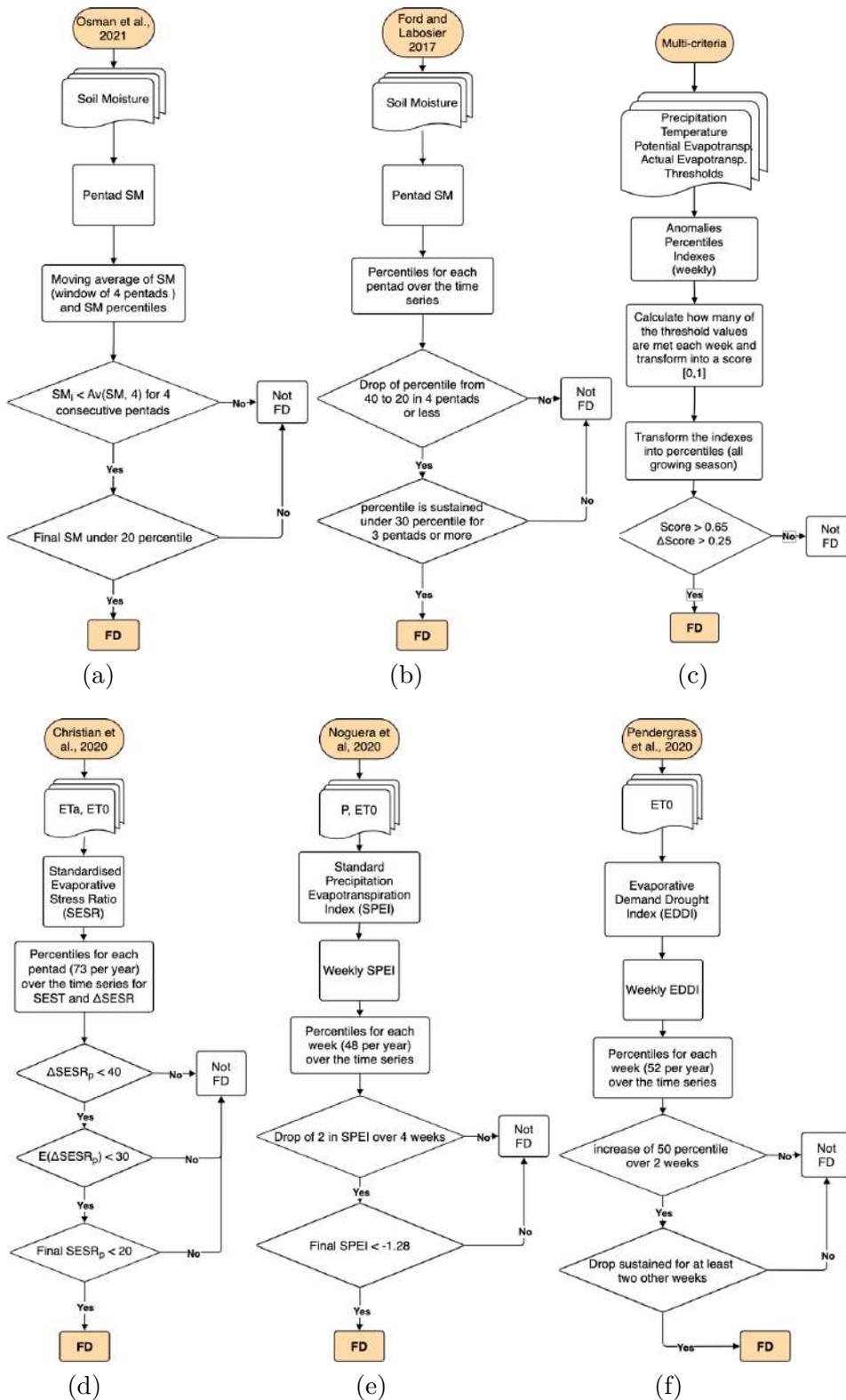
Figure 5.1: Flowcharts of the six methods for flash drought identification: (a) M1: Osman et al. (2021); (b) M2: Ford and Labosier (2017); (c) M3: Novel multi-criteria method; (d) M4: Christian, Basara, Hunt, et al. (2020); (e) M5: Noguera, Castro, and Serrano (2020); (f) M6: Pendergrass et al. (2020). The implementation of all methods is available in the supplements of this paper as an R package.

Table 5.1: Comparison of key variables, statistics, and threshold criteria for the detection of flash droughts in all six methods

Method	Variables	Original dataset type[1]	Statistics	Onset rate	Total duration
M1 Osman et al. (2021)	Soil Moisture	Grid (CONUS)	Soil Moisture Volatility Index (SMVI) Running averages (RA)	4 weeks	Onset duration + while SM lower than 4-pentad running average
M2 Ford and Labosier (2017)	Soil Moisture	Station (Eastern USA)	Percentiles	4 pentads	Onset duration + until SM >30th p.
M3 Multi-criteria	Precipitation Temperature Actual Evapotransp. Potential Evapotransp.	Station (Central Europe)	Anomalies Indexes Percentiles	4 weeks	Onset duration + until 2 consecutive weeks with Score <0.65
M4 Christian, Basara, Otkin, Hunt, et al. (2019) Christian, Basara, Hunt, et al. (2020)	Actual Evapotransp. Potential Evapotransp.	Grid (Eastern USA[2] and SW Russia)	Standardized Evaporative Stress Ratio (SESR)	6 pentads	Onset duration + until 2 consecutive weeks with increasing SESR
M5 Noguera, Castro, and Serrano (2020)	Precipitation Potential Evapotransp.	Station (Spain)	Standard Precipitation Evapotranspiration Index (SPEI)	4 weeks	After onset, until SPEI >-1.28
M6 Pendergrass et al. (2020)	Potential Evapotransp.	GRID (CONUS)	Evaporative Demand Drought Index (EDDI)	2 weeks	Onset duration + until EDDI increases towards wet conditions

1 The data type used in the original application/publication of the method. Grid indicates methods that were implemented with reanalysis or remote sensing data. Station indicates methods that were originally implemented with weather station (direct measure) data.

2 The Christian et al. method studied multiple areas in the continental USA: the Great Plains, Corn Belt, and Great Lake regions, in the states of Georgia, Kansas, Iowa, and Minnesota.

5.2.2 Comparison Metrics

Two types of metrics were used to compare the six FD methods: synchronicity metrics and the confusion matrix. The method of Osman et al. was used as a reference method as it was considered the method which most closely followed our definition of a flash drought as given in Section 5.1 (using rapid soil moisture decline as a key variable for FD identification). This method was evaluated against measured soil moisture data (see Section 5.2.3) in the root zone and, therefore, appeared to be particularly suited to reproduce the flash drought dynamics of croplands.

5.2.2.1 Synchronicity metrics

The concept of synchronicity based on the work of Kemter et al. (2020) was employed to compare the rate of identification of flash drought events and the intervals, which were correctly identified as intervals with no flash drought, for a weekly resolution. Synchronicity metrics were originally developed to analyse whether extreme floods occur concurrently with the same timing in larger basins (Kemter et al., 2020). It is based on the two synchronicity metrics $sync_1$ and $sync_0$, which are defined as the average proportion of successful identification of flash drought events and no flash drought events, respectively. Here, the Osman reference method was compared separately with the other methods. The metrics can take values between zero and one, therefore, a perfect agreement between one method and the Osman et al. reference method occurs if both metrics are equal to one. If only $sync_1$ is close to one and $sync_0$ is close to zero, it would mean that all FD events were identified as in the reference method, but the interval times without flash droughts were all identified incorrectly (by identifying many more FD events, where the reference method did not find one). The two metrics $sync_1$ and $sync_0$ are further summarised by their harmonic mean $\overline{sync}$ (which is given by the reciprocal of the arithmetic mean of their reciprocals) to represent the trade-offs between the two synchronicities.

5.2.2.2 Confusion Matrix

A confusion matrix is a tool borrowed from data science (James, Witten, et al., 2013), from which several metrics can be derived to evaluate the rate of true and false identifications of an event using a weekly resolution. Using the Osman et al. method as the reference method, for all weeks the matrix assesses the following scores for each flash drought identification method individually (Figure 5.2 presents as an example of the comparison of M1 with M2):

- A true positive score (TP) for a true identification of a flash drought: both methods identified a flash drought,

- a true negative score (TN) for a correct non-identification: both methods did not identify a flash drought,

- a false positive score (FP) for a wrong identification: the Osman et al. method did not identify a flash drought but the other method did and,

- a false negative score (FN) for an incorrect identification: the Osman et al. method did identify a flash drought but the other method did not.

The four scores are then summarised into several confusion metrics that describe the degree of similarity (equations in Table 5.2):

A. The true positive rate (TPR) and positive prediction value (PPV), which evaluate the performance of the tested method to correctly replicate a flash drought.

B. The negative rate (TNR) and negative prediction value (NPV), which evaluate the performance of the tested method to correctly replicate the intervals between flash droughts.

C. The Matthews correlation coefficient (MCC), which summarises resemblance and considers imbalanced datasets for which one class of events, in this case, the flash droughts, is much smaller than the other class, that is, no flash drought periods (Matthews, 1975; Delgado and Tibau, 2019; Chicco, Tötsch, and Jurman, 2021).
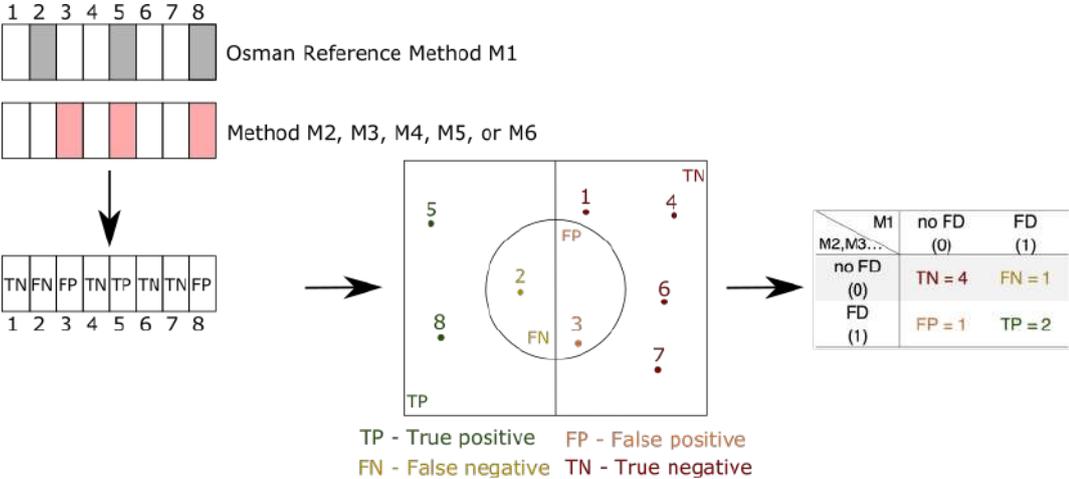
Figure 5.2: Graphic representation of the confusion matrix for an example data set of 8 weeks. The M1 Osman reference method is compared with method M2 Ford (or any other method) for true negative, true positive, false negative, and false positive identification of flash droughts. Grey intervals: a flash drought occurred according to M1, red intervals: a flash drought occurred according to M2, white interval: no flash drought occurred according to either method. The plot in the middle visualizes the same identifications (TP and FN) on the left side and the right side shows opposing identifications (FP and TN).

The metrics range from 0 to 1, with values close to 1 indicating a high degree of similarity and values close to 0 indicating little resemblance.

Table 5.2: Metrics derived from confusion matrices used in this study.

Metric	Equation
True negative rate (TNR)	$TNR = \frac{TN}{TN+FP}$
Negative predictive value (NPV)	$NPV = \frac{TN}{TN+FN}$
True positive rate (TPR)	$TPR = \frac{TP}{TP+FN}$
Positive predictive value (PPV)	$PPV = \frac{TP}{TP+FP}$
Matthews Correlation Coefficient (MCC)	$\frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$

5.2.3 Data and study areas

The climate series from cropland stations of the FLUXNET2015 dataset (Pastorello et al., 2020) was used for flash drought identification, as this dataset provided high-quality measured data for most of the required climate variables. Daily station data of the variables of precipitation, temperature, soil water content, latent and sensitive heat fluxes, wind speed, and relative humidity, with durations ranging between 11 and 14 years, were available for four stations in Central Europe (Figure 5.3), where soil water data was the most limiting criterion for station selection. Table 5.3 contains information on soil type and location.

Potential evapotranspiration was calculated using the standard method described by Allen (1998). The actual evapotranspiration (ET) was calculated using Equation 5.1, which relates evaporation to latent heat and temperature:

$$ET = \frac{H_l}{l_v \rho_w} \tag{5.1}$$

where H_l is the latent heat flux, ρ_w is the density of water, and l_v is the latent heat of vaporisation, which in turn is a function of air temperature calculated following the procedures Rogers (1989).

Table 5.3: Experimental cropland stations with information on location, duration of available soil moisture series, soil type, and climate zone

Station	Dur. (years)	Lat.	Lon.	Elev. (m.a.s.l.)	Temp. (ºC)	Prec. (mm)	Cli.[*]	Soil	Reference (DOI)
BE-Lon	11	50.55	4.75	167	11.4	766	Cfb	Luvisol (silt loam)	10.18140/FLX/1440129
DE-Geb	14	51.10	10.91	161.5	9.7	531	Cfb	Chernozem (-)[2]	10.18140/FLX/1440146
DE-Kli	11	50.89	13.52	478	7.8	811	Cfb	Podsol (silty loam)	10.18140/FLX/1440149
IT-BCi	11	40.52	14.96	20	17.9	1199	Csa	(-)[2]	10.18140/FLX/1440166

[1] Data source: fluxnet.org/data/fluxnet2015-dataset/
[2] Data not available
[*] Climate classification according to the Köppen classification system.
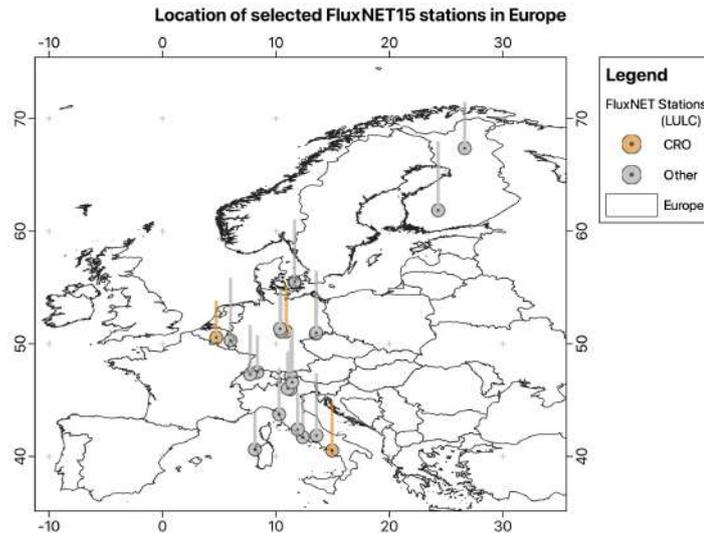
Figure 5.3: Location of selected FLUXNET2015 stations in Europe. The colours indicate land cover classes. The bar size is proportional to the length of the data series, varying from 11 to 14 years.

5.3 Results and Discussion

Station-by-station comparison

Flash droughts identified by all six methods for the four stations showed many commonalities as well as disparities (Figure 5.4a-d). Each time series plot in Figure 5.4 shows the daily soil moisture content in a black line and the critical variable of each method in red: the 20th percentile for M1 Osman et al. and M2 Ford Labosier, the summary score for M3 the Multi-criteria method, SESR for M4 Christian et al., SPEI for M5 Noguera et al., and EDDI for M6 Pendergrass et al. The green bars mark the duration of identified flash droughts and the yellow bars indicate periods in which a method missed identification by a small fraction of its respective score or threshold.

In the majority of the 11-year study period, all methods detected at least one flash drought per year, except in 2005 and 2008. Method M6 Pendergrass et al. yielded the lowest number of events at all stations (4-7), whereas the other methods identified considerably more events varying between 8-10 (M2), 13-14 (M3), 5-15 (M4), and 5-15 (M5). M1 Osman et al., with 11-18 events, identified the largest number of events.
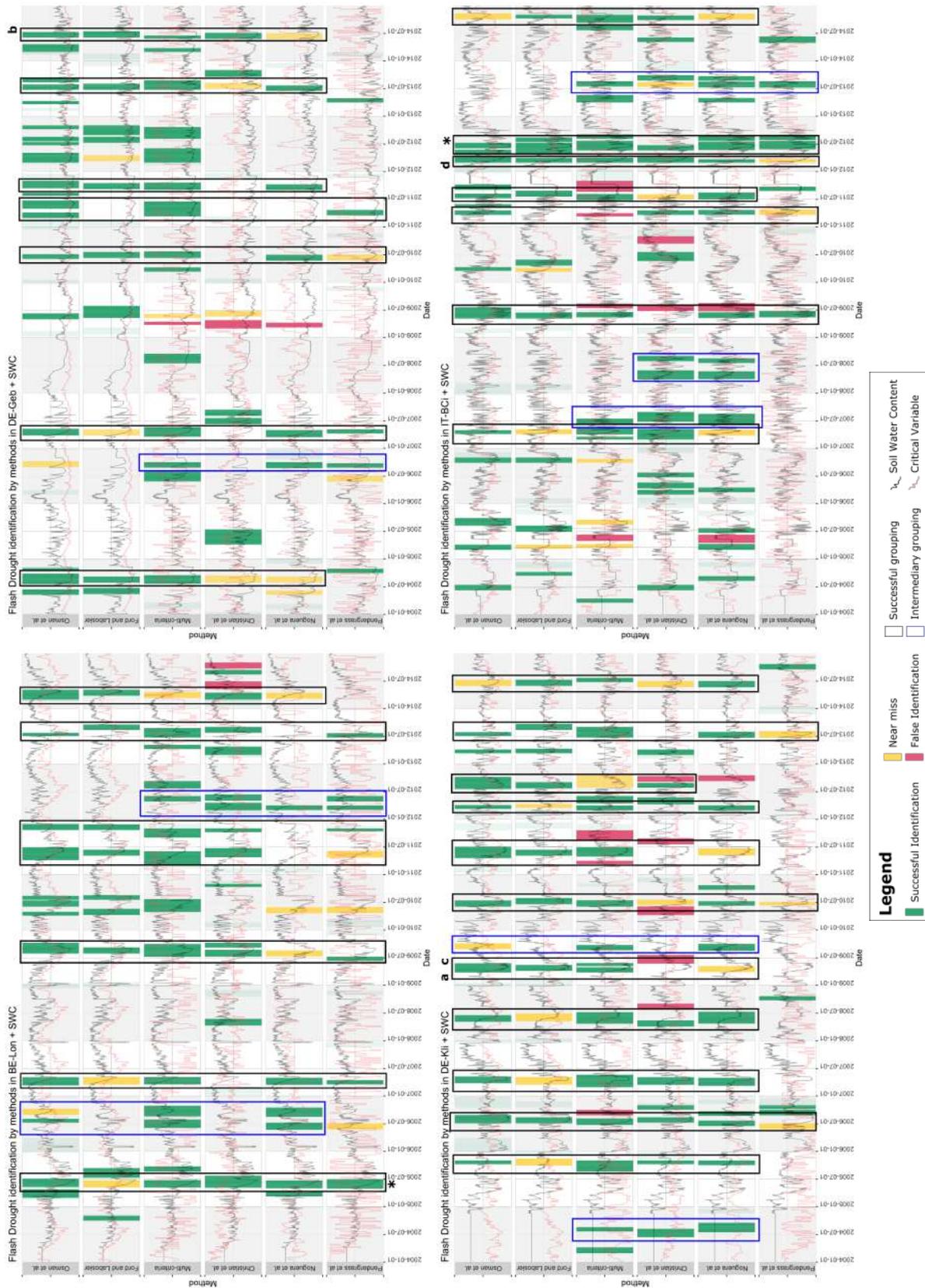
Figure 5.4: Flash drought events for six methods and four stations (A: BE-Lon, B: DE-Geb, C: DE-Kin, D: IT-BCi) for the time period 2004-2014. Colour coding: green bars show start and end of flash droughts, yellow bars show near misses and red bars show clear false identifications.

Concurrent flash drought identification of five methods, marked with black frames, (three methods, marked with blue frames) occurred 6 (2) times at Be-Lon, 10 (2) times at DE-Kli, 7 (1) times at DE-Geb, and 6 (3) times at IT-Bci, with some small variations in the exact start time and duration. Station DE-Geb showed the overall best overlap for all methods, followed by Be-Lon. De-Geb and IT-Bci were characterised by longer periods, with no clear overlap between 2009 and 2012 for De-Geb. There was also no clear overlap between 2004-2006 and 2010 for IT-Bci.

The two soil moisture-based methods, M1 and M2, showed the highest similarities overall. Every single event of M1 as the reference method was assessed for consistency with our flash drought definition. Other events that were not detected by the reference method were also evaluated to determine if they should have been delineated. It is noteworthy to mention that there were no concurrent flash drought identifications that did not include the reference method M1.

Concurrent identification among multiple methods would considerably increase if the just-missed events (yellow coding in Figure 5.4) were also considered, suggesting that some of the thresholds might have been set too rigidly. The red bars in Figure 5.4 mark obvious false identifications due to artefacts of the percentile approaches. This signifies some threshold crossing, but at the same time, the corresponding time series does not show any rapid decrease in absolute terms. Methods M3 to M5 showed 24 false events altogether, with most of them in the early spring or autumn seasons. Instances of false identification may be reduced if additional rules regarding absolute changes in climate variables are introduced into the methods.

Method comparison using synchronicity metrics

The metrics of synchronicity $sync_1$ (two methods identifying the same flash drought periods) and $sync_0$ (two methods identifying the same no flash drought intervals), and their harmonic mean $\overline{sync}$ are given in Figure 5.5 for methods M2 to M6, in comparison to M1 as the reference method, for all four stations. None of the methods showed perfect agreement with the reference method, for which all three metrics would have equalled 1.

Methods M2 to M5 presented intermediate values for the harmonic mean $\overline{sync}$ indicating similar behaviour regarding flash drought and no drought identification, in comparison to the reference method M1, with $\overline{sync}$ values ranging between 0.47 (M2: Ford Labosier) and 0.39 (M5: Noguera et al.). The Multi-criteria method M3 showed the highest $sync_1$ values overall, particularly for station BE-Lon, directly followed by M2. For methods M2 to M5, the $sync_1$ value at station DE-Geb was significantly smaller than that of the other stations, indicating that the co-identification of flash droughts in comparison with method M1 was the smallest at this station.

Method M6 (Pendergrass et al.) showed the largest differences from the reference method, with a $\overline{sync}$ value of only 0.2. However, M6 had large values for $sync_0$ (0.9), thus showing a

great agreement in not identifying periods of no flash drought. Moreover, the small value of $sync_1$ (0.1) clearly shows that M6 rarely co-identified a flash drought simultaneously with M1.
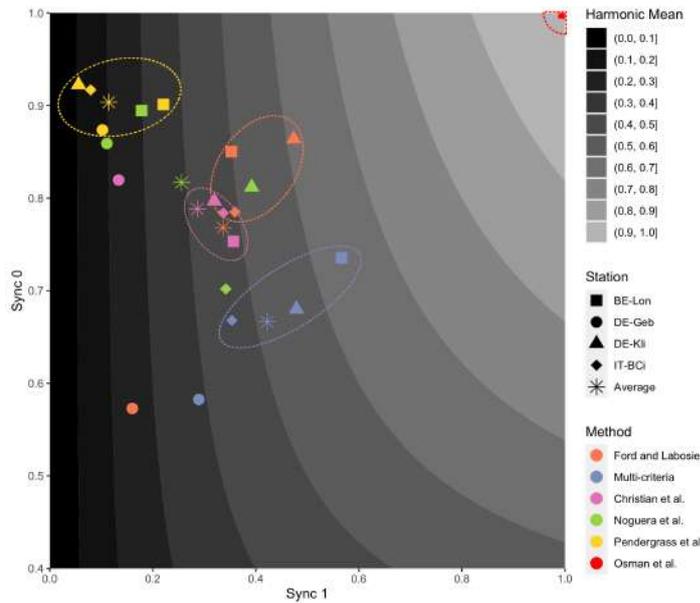
Figure 5.5: Synchronicity metrics of co-identification of flash drought events ($sync_1$) and intervals between flash drought events ($sync_0$) for Methods 2 to 6, compared to M1 Osman et al. as the reference method. The grey surface plot indicates the possible values of the harmonic mean $\overline{sync}$.

Method comparison using confusion matrix metrics

The metrics given by the confusion matrix in Figure 5.6 provide a slightly different picture compared to the synchronicity metrics. The MCC describing the general resemblance of co-identification between reference method M1 Osman et al. and the other methods was highest for M2 Ford Labosier (median value of 0.43), followed by the Multi-criteria method (0.28). The other three methods had small values close to zero indicating little overall resemblance.

All methods had high values for the TNR and NPV, with M2 Ford Labosier and M6 Pendergrass et al. having the highest TNR (above 0.9), and all NPV metrics varying between 0.7 (Pendergrass et al.) and 0.8 (Ford Labosier). Hence, the methods did well in identifying intervals with no flash drought, relative to the reference method. However, there were significantly larger discrepancies for the TPR and PPV, M2 Ford Labosier had reasonably high values (0.75 and 0.43, respectively), directly followed by the Multi-criteria method (0.55, 0.51) and the method of Christian et al. (0.45, 0.41), whereas Noguera et al. and Pendergrass et al. had very low values, particularly for the TPR metric, with values below 0.3. Thus, method identification deviated significantly in two ways: the identification of multiple events that were not identified by the reference method or the omission of multiple events (e.g. M3 Noguera et al. and M6: Pendergrass et al., both examples are also illustrated in Figure 5.4).

Both comparison metrics showed that even though there is a considerable resemblance in how certain methods identify flash drought events, there are just as many disparities. By
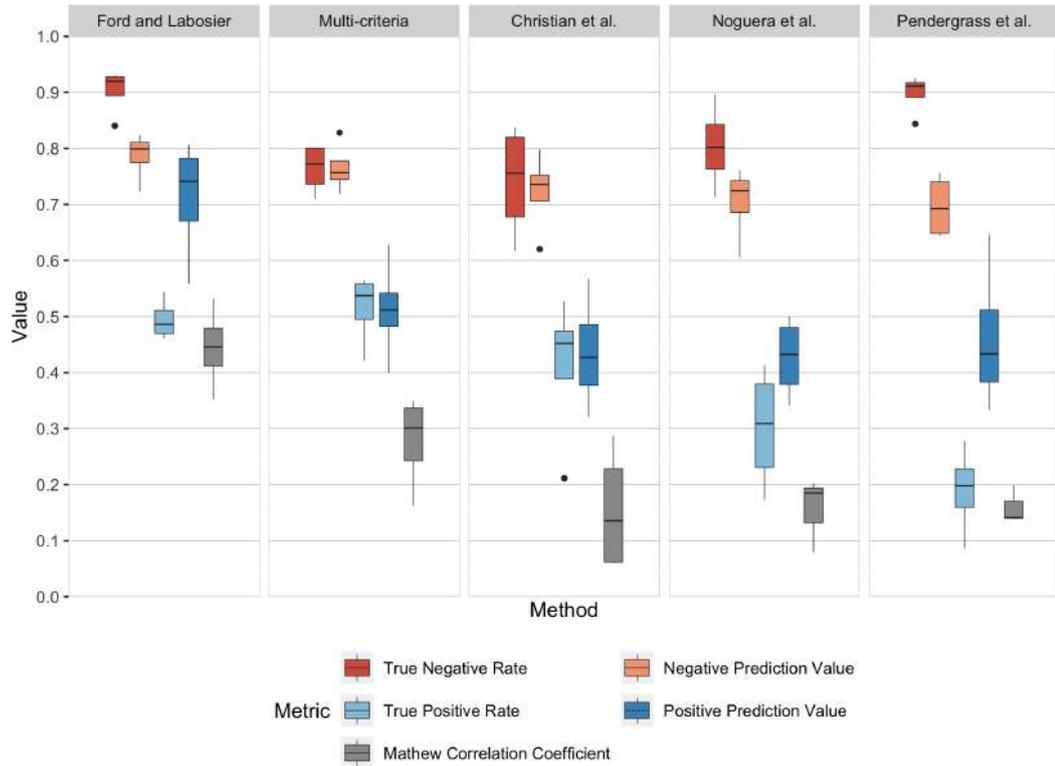
Figure 5.6: Confusion matrix with multiple metrics evaluating true and false identifications of flash drought and no flash drought weeks of Methods M2 to M6 compared to those of M1 Osman et al., which was the reference method. Box plots represent all metrics including the data from all four stations and all weeks during the study period 2004-2014.

studying each time series in Figure 5.4 in detail, four key factors were identified that explain the differences in flash drought detection as visualised in Figure 5.7:

A. Critical proxy variable opposed soil moisture dynamics

An example of this behaviour is shown in Figure 5.7A, where monitored soil water content increased due to a longer rainfall period. However, M3 Christian et al. identified a flash drought event using the SESR index, which increased considerably due to an on-going increase in the potential evapotranspiration rate.

B. Velocity of depletion was too small

As shown in Figure 5.7B, Method M4 Noguera et al. showed a decrease in their proxy variable SPEI during a period of rapid depletion of soil moisture. However, the time period over which the intensification occurred was not long enough to trigger a flash drought with M4 (but the reference method M1 was able to identify it).

C. Threshold for final depletion was not exceeded

In this case, the critical final threshold for flash drought identification is not exceeded. For example, the score of the M3 Multi-criteria method did not surpass the set value of 0.65 during an actual incidence of rapid soil moisture loss (Figure 5.7C).

D. Duration of the event too short:

Finally, the event in Figure 5.7D shows an event missed by the EDDI index of M6

101
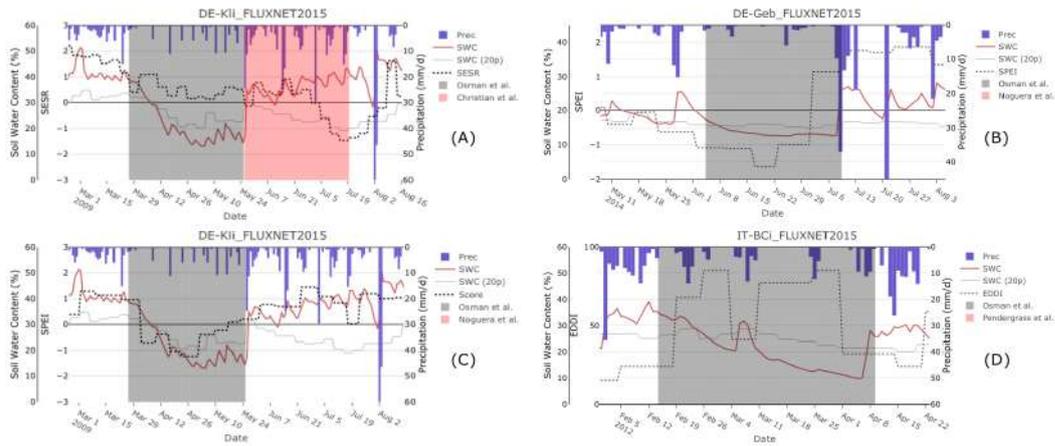
Figure 5.7: Differences in flash drought detection due to four factors, illustrated with an example of misidentification: A) Critical proxy variable SESR of M3 Christian et al. opposed moisture dynamics; B) Velocity of depletion was too small for M5 Noguerra et al.; C) Threshold for final depletion of M4: Multi-Criteria score was not exceeded; and D) Duration of the event was too short for M6 Pendergrass et al.

Pendergrass et al., where a significant amount of precipitation caused a short recuperation, which did not break the course of rapid and extreme drying visible in the soil moisture series. However, the two intervals with high EDDI were too short to be classified as flash droughts by M6.

Misidentification and near-misses (as illustrated in Figure 5.7) might have been due to just one of the factors, but were often due to a combination of them.

Commonalities and disparities due to the lack of definition – the way forward: the ensemble approach

So far, we established that all six methods showed clear commonalities, but also disparities, in the detection and delineation of individual flash drought periods. The differences in the detected periods can be directly linked to the use of different critical variables and threshold values, and the different minimum durations and absolute or relative changes that are required for each method to detect a flash drought (Table 5.1).

We showed that the choice of the critical variable plays a major role in explaining the differences in identifications (Figure 5.7A). Different critical variables are already used in the analysis of 'normal' droughts, where the choice of variables such as rainfall, runoff, and soil moisture deficits determines the type of drought form to be analysed. Normal droughts are commonly categorised into meteorological, hydrological, and agricultural droughts (Mishra and Singh, 2010). There are no comparable sub-definitions for flash droughts, although several studies have stated that some flash droughts are mainly influenced by heatwave dynamics and others state that they are more strongly influenced by rainfall deficits (Mo and Lettenmaier, 2016; Pendergrass et al., 2020), but currently no clear definitions exist for such sub-types. Flash droughts are known to be triggered or exacerbated by compound extreme climatic events

102

(Pendergrass et al., 2020; Otkin, Svoboda, et al., 2018; Lisonbee, Woloszyn, and Skumanich, 2021), but the relative magnitude and associated impact of each extreme have not yet been established (Mo and Lettenmaier, 2016). At the same time, the methods differ in the extent to which they include compound dynamics in their detection procedure. Notably, method M6 Pendergrass et al. does not include any information on rainfall; therefore, it may be considered as a flash drought method that is solely driven by evaporative demand. Methods M3-5 include a mixture of rainfall deficit, heatwave, and evapotranspiration information, whereas M1 and M2 use soil moisture as an integrative response variable for all vertical water exchange processes. We propose that by using multiple methods, one can identify different flash drought types, and future research efforts should concentrate on disentangling the extent to which different climatic components are responsible for the respective droughts.

The focus of this method comparison was placed on flash droughts in temperate croplands, where unusually rapid declining soil moisture during the growing season is considered the most direct indicator of severe plant water stress (Ford and Labosier, 2017; Samaniego, Thober, et al., 2018). However, methods using monitored, root-zone soil moisture data (M1 and M2) are severely limited by data restrictions, whereas indirect methods (M3-M6) using other proxy variables, such as rainfall or temperature, allow the use of much longer data series and more stations. Directly connected to the data limitation of the soil-moisture-based methods are uncertainties in the derivation of the required percentiles. While they are impressive for soil moisture data, time series that are 11 or 14-years long, as the ones used here, are likely to result in skewed misrepresentation of the upper and lower percentiles required for the soil moisture-based methods. This fact highlights the dilemma of soil moisture-based approaches. While they should be preferred over other methods to correctly reproduce flash drought dynamics for croplands, the shortness of their series potentially leads to large uncertainties in their threshold values.

This study did not aim to determine whether some methods are better than others, but rather to verify whether the methods compare well and thus evaluate whether any method could be used to assess flash drought dynamics for croplands in Central Europe. However, detailed synchronicity analysis does not confirm the latter. Similar to the comparison study of Osman et al. (2021) for the United States, we found different frequencies depending upon which definition was used. Nevertheless, three or more methods often detected the same flash drought (Figure 5.4). We propose that to balance the strengths and weaknesses of all methods, an ensemble approach for event identification be used as a way forward; thereby not just one, but several methods may be employed, which can be easily implemented with our R Shiny app. Multiple co-identifications would thus diminish the uncertainty of incorrect identifications and might give a more comprehensive picture when different types of flash droughts occur (Wang and Yuan, 2018).

Several questions remain unanswered. Although a rapid depletion in soil moisture was detected multiple times during the investigated growing periods, the impact on vegetation health remains unclear and needs to be assessed with information other than climate data (plant mortality or remotely sensed imagery, as done by Peters et al. (2002) and Otkin, Zhong,

Hunt, Basara, et al. (2019)). The impact of different soil types, their water-holding and storage capacities, as well as the impact of different crop and vegetation species on flash drought identification, requires future assessment. Finally, a thorough long-term analysis of flash drought dynamics using an ensemble approach, as suggested here, is pending. However, the preliminary analysis showed that flash droughts on croplands in Central Europe are no rare events, but occurred on average once every 1 to 2 years during the investigated time period.

5.4 Conclusions

Defining and delineating flash droughts pose new challenges for hydrologists and climatologists. In this study, we compared six different delineation methods and observed a large degree of synchronicity, but also some divergence, in the identified flash drought periods depending on which definition was used. The disparities of one method for detecting different drought intervals in comparison to others were narrowed down to four factors: (1) the opposing behaviour of proxy climate variables in comparison to prevailing soil moisture dynamics, (2) differences in the estimated velocity of drought intensification, (3) not exceeding pre-set threshold values for final depletion, and/or (4) differences in pre-set minimum drought lengths. Rather than seeing the detected divergence in identifying drought periods among the various methods as a weakness, we suggest using an ensemble approach for event identification to ensure that flash droughts of different sub-types are detected. These include different combinations of compound drivers, with differences in intensification dynamics.

Compound events, such as compound warm spells and rainfall deficit, have become more frequent in past decades (Vogel, Paton, et al., 2021; Zscheischler et al., 2020) and are expected to increase. However, it is not clear to what degree the different components that are relevant for flash drought development will develop. Thus, a single, commonly accepted definition for flash drought delineation may be the wrong goal, as it would take away the flexibility that an ensemble approach using multiple methods has

Appendix - Multi-criteria method characteristics

The Multi-criteria method uses weekly data for precipitation amount, mean temperature, and potential evapotranspiration amount, and is calculated following the method of Allen (1998). It calculates a set of indices, including the drought index SPEI (Serrano, Begueria, and Moreno, 2010), the EDDI following the procedure of Hobbins et al. (2016) and Lukas, Hobbins, and Rangwala (2017), and anomalies of temperature, precipitation, and potential evapotranspiration on a weekly and 3-week basis (Table 5.4, Equation 5.2). The formula for which is as follows:

$$X_{an_{ij}} = \frac{X_{ij} - \overline{X_i}}{\sigma_{X_i}} \qquad (5.2)$$

where X is the studied variable (P, T, or PET). The index an indicates an anomaly. The index i is the week (from 1 to 52) and the index j the year (from 1 to n, where n is the number of years in the time series). X_i is the average value of variable X in week i, and σ_{X_i} the standard deviation.

For each indicator a threshold value was set as given in Table 5.4. For SPEI, the proposed limits by Noguera, Castro, and Serrano (2020) were used and for EDDI the proposed limits by Pendergrass et al. (2020) were used.

Table 5.4: All criteria used in the Multi-criteria method in this study.

Id	Indicator name	Metric	Threshold	Weight	Source
1	Temperature	Anomaly	1	1	Mo and Lettenmaier (2016)
2	Precipitation	Anomaly	-1	1	Mo and Lettenmaier (2016)
3	Potential Evapotransp. (ET_0)	Anomaly	1	1	-
4	4-week acc. temperature	Acc. Anomaly	3	1	-
5	4-week acc. precipitation	Acc. Anomaly	-3	1	-
6	4-week acc. ET_0	Acc. Anomaly	3	1	-
7	SPEI	Index value	-1.28	1	Noguera, Castro, and Serrano (2020)
8	4-week SPEI variation	Absolute	-2	1	Noguera, Castro, and Serrano (2020)
9	EDDI	Index value	85	1	Pendergrass et al. (2020)
10	2-week EDDI variation	Absolute	50	1	Pendergrass et al. (2020)

Once the 10 *indicators* were computed, we would then calculate the score of each week by comparing the *indicator* values with the defined *threshold*, as presented in Equation 5.3.

$$Score = \frac{1}{\sum_{i=1}^{10} W_i} \sum_{i=1}^{10} (I_i \geq T_i) \times W_i \in [0, 1] \qquad (5.3)$$

where I denotes each *indicator*. The index i represents the number of indicators (from 1 to 10). T represents the *thresholds* of each indicator and W is the *weight* of each indicator in the assessment of FD events. In this study, we kept the values of all W equal to 1. The values for each *threshold* are listed in Table 5.4.

A flash drought event is identified when, over a period of up to 3 weeks, the change in score ($\Delta Score$) is 0.25 or higher and the final score is 0.65 or more. This describes the intensification period and flash drought onset. The flash drought lasts until the score decreases to below 0.65 for 2 consecutive weeks.

6

Flash Drought Visualization: Shiny App

In this chapter, we present a technical description of the Shiny App built to assist Flash Drought visualization and identification. The application is available at https://pedroalencar.shinyapps.io/FD-Viz/. Also, the complete code for all images, GUI (graphical user interface), as well as data used are publicly available at https://github.com/pedroalencar1/FD-Viz.

6.1 Shiny applications

Shiny®is an open-source and collaborative R package primarily developed and maintained by R-Studio®. It can be used to build dynamic apps on webpages or Markdown, and to build dashboards.

The syntax of Shiny is object-oriented with multiple pre-built classes, but allows the programmer to build their own classes and objects using other languages, such as R, Java, and JScript.

The minimal structure of a Shiny app is composed of three parts:

Pre-processing: It consists of loading libraries, and defining datasets and functions that can be used throughout the App.

```
1 # Load packages
  library(shiny)
  library(ggplot2)
  library(lubridate)
  library(plyr)
6 library(dplyr)
```

6. Flash Drought Visualization: Shiny App

```r
library(plotly)

#loads data set
complete_series <- read.csv('series_gebesee.csv') %>%
                        .[,-c(1)]

#rename columns
colnames(complete_series) <- c("Date","Mo and Lettenmeier","Ford and Labosier",
                                "Pendergrass et al.","Noguera et al." ,
                                "Christian et al.", "Osman et al.",
                                "swc","et0", "eta", "temperature",
                                "precipitation","swc_p20" ,"spei","sesr" , "eddi")
#select subset
models <- names(complete_series[2:7])
```

User interface (UI):   In this part, the programmer defines all parameters and designs of the GUI, including menus, messages to the user, location of graphs, etc.

```r
ui <- fluidPage(
    # App title
    titlePanel('Flash Drought - Interactive Visualisation'),

    #Position menus
    sidebarLayout(
        # Sidebar panel for inputs
        sidebarPanel(
            HTML("<h3>Input parameters</h3>"),

            selectInput("Method 1", "Choose Method 1", choices = models,
            selected = 'Ford and Labosier'),
            selectInput("Method 2", "Choose Method 2", choices = models, selected = NULL),
            sliderInput("year","Choose year",min=min(year(complete_series$Date)),
                        max=max(year(complete_series$Date)),value=2004,step = 1),
        ),

        # Position and order of Plotly graphs
        mainPanel(
            plotlyOutput("graph1"),
            plotlyOutput("graph2")
        )
    )
)
```

Processing (server):   In this part, the programmer defines all operations and visualizations that will be processed on the server. In this part, the App will receive the information provided by the user in the menus at the GUI, process the data set accordingly and generate the figures, graphs and maps.

```r
server <- function(input, output, session){

    output$graph1 <- renderPlotly({
        plot_ly(complete_series[year(complete_series$Date) == input$year,], x = ~Date) %>%
            add_bars(y = ~precipitation, name='Prec',yaxis = 'y2',
            marker = list(color = 'rgba(106,90,205,1)', width = 1) %>%
            add_lines(y = ~temperature, name='Temp', line = list(color = 'rgba(255,0,0,1)')) %>%
            add_bars(y = ~get(input$`Method 1`), name='Method 1',yaxis = 'y3',
            marker = list(color = 'rgba(0,0,0,0.3)', width = 1) %>%
            add_bars(y = ~get(input$`Method 2`), name='Method 2',yaxis = 'y3',
            marker = list(color = 'rgba(255,0,0,0.3)', width = 1) %>%
            layout(
                title=paste('Fluxnet station Gebesee: ',input$year, sep = ''),
                xaxis = list(title = "Date", domain = c(0,0.95), dtick = 14),
                yaxis = list(title = 'Temperature ( C )', side = "left", color = "black",
                position = 0, anchor = 'free', range = c(-10,35), dtick = 10),
                yaxis2 = list(title = 'Precipitation (mm/d)', side = "right", color = "black",
                            overlaying = "y", anchor = 'free', position = 0.95,
                            range = c(45,0), dtick = 10),
                yaxis3 = list(title = '', side = "right", color = "white",
```

```
21                                 overlaying = "y", anchor = 'free', position = 1,
                                   range = c(0,1), dtick = 10),
                       showlegend = T
                   )
           })
26
       output$graph2 <- renderPlotly({
           plot_ly(complete_series[year(complete_series$Date) == input$year,], x = ~Date) %>%
               add_bars(y = ~precipitation, name='Prec',yaxis = 'y2',
               marker = list(color = 'rgba(106,90,205,1)', width = 1) %>%
31             add_lines(y = ~et0, name='ET0', line = list(color = 'rgba(255,0,0,1)')) %>%
               add_lines(y = ~eta, name='ETa', line = list(color = 'rgba(0,255,0,1)')) %>%
               add_bars(y = ~get(input$'Method 1'), name='Method 1',yaxis = 'y3',
               marker = list(color = 'rgba(0,0,0,0.3)', width = 1) %>%
               add_bars(y = ~get(input$'Method 2'), name='Method 2',yaxis = 'y3',
36             marker = list(color = 'rgba(255,0,0,0.3)', width = 1) %>%
               layout(
                   title=paste('Fluxnet station Gebesee: ',input$year, sep = ''),
                   xaxis = list(title = "Date", domain = c(0,0.95), dtick = 14),
                   yaxis = list(title = 'Evapotranspiration', side = "left", color = "black",
41                 position = 0, anchor = 'free', range = c(0,6.75), dtick = 1.5),
                   yaxis2 = list(title = 'Precipitation (mm/d)', side = "right", color = "black",
                                 overlaying = "y", anchor = 'free', position = 0.95,
                                 range = c(45,0), dtick = 10),
                   yaxis3 = list(title = '', side = "right", color = "white",
46                               overlaying = "y", anchor = 'free', position = 1,
                                 range = c(0,1), dtick = 10),
                   showlegend = T
               )
           })
51 }

   # Creates the app, linking ui and server
   shinyApp(ui, server)
```

Application output: With the previous three steps, we can create a functional application. The application above is available at https://pedroalencar.shinyapps.io/Shiny$_G$ebesee/
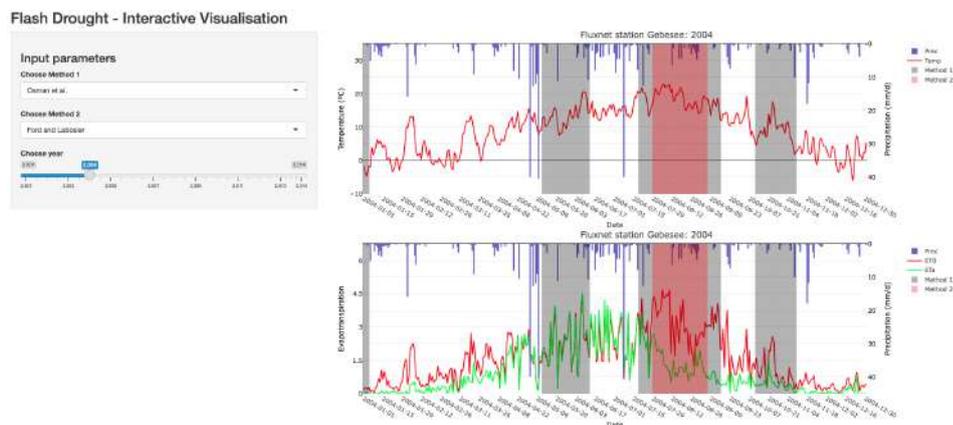
Figure 6.1: Application Shiny_Gebesee, comparing annual flash drought identification by multiple methods in Gebesee (Germany), using data from the FLUXNET2015 station DE-Geb.

6.2   FD-Viz

Our application allows the user to compare the identification of six different methods at 22 different FLUXNET2015 stations in Central Europe. It is structured in 4 pages:

A. **Events:** presents five graphs that show simultaneously up to 2 method overlaying multiple variable and indicators

B. **Metrics:** shows three bar plots with metrics derived from confusion matrix (Chapter 5)

C. **Summary:** presents seven plots with counts and averages of events by station, year, and method

D. **Co-Identification:** shows three plots with identification of events by methods and co-identification over time and stations

Below we present some of the plots generated in the FD-Viz.
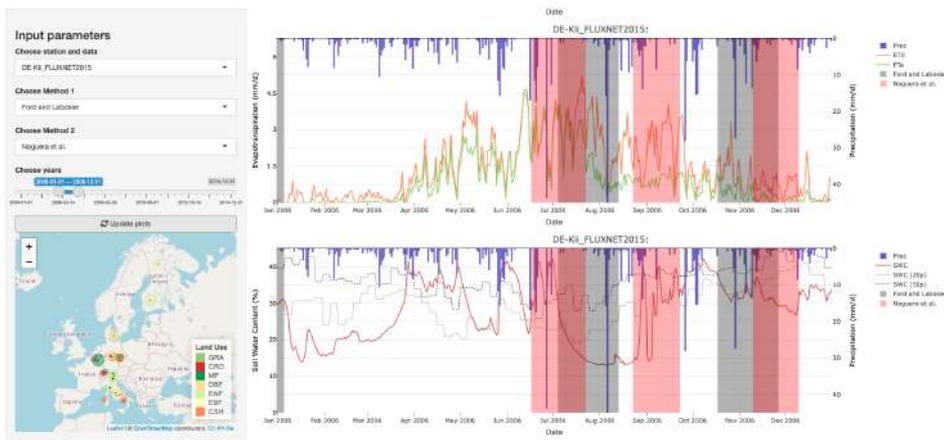
Figure 6.2: View of the first page of FD-Viz. On the left we have the menu where the user can select the station, methods, and duration of the analysis, along with an interactive map that shows the location, land use and time series duration of each station.
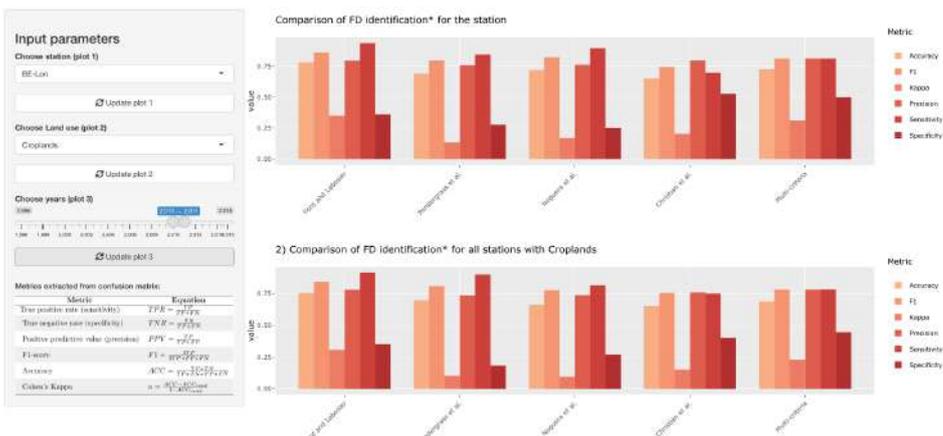
Figure 6.3: View of the second page of FD-Viz. On the left we have the menu where the user can select the station, land use and duration of the analysis, together with a table containing a short description of the metrics presented in the bar plots.
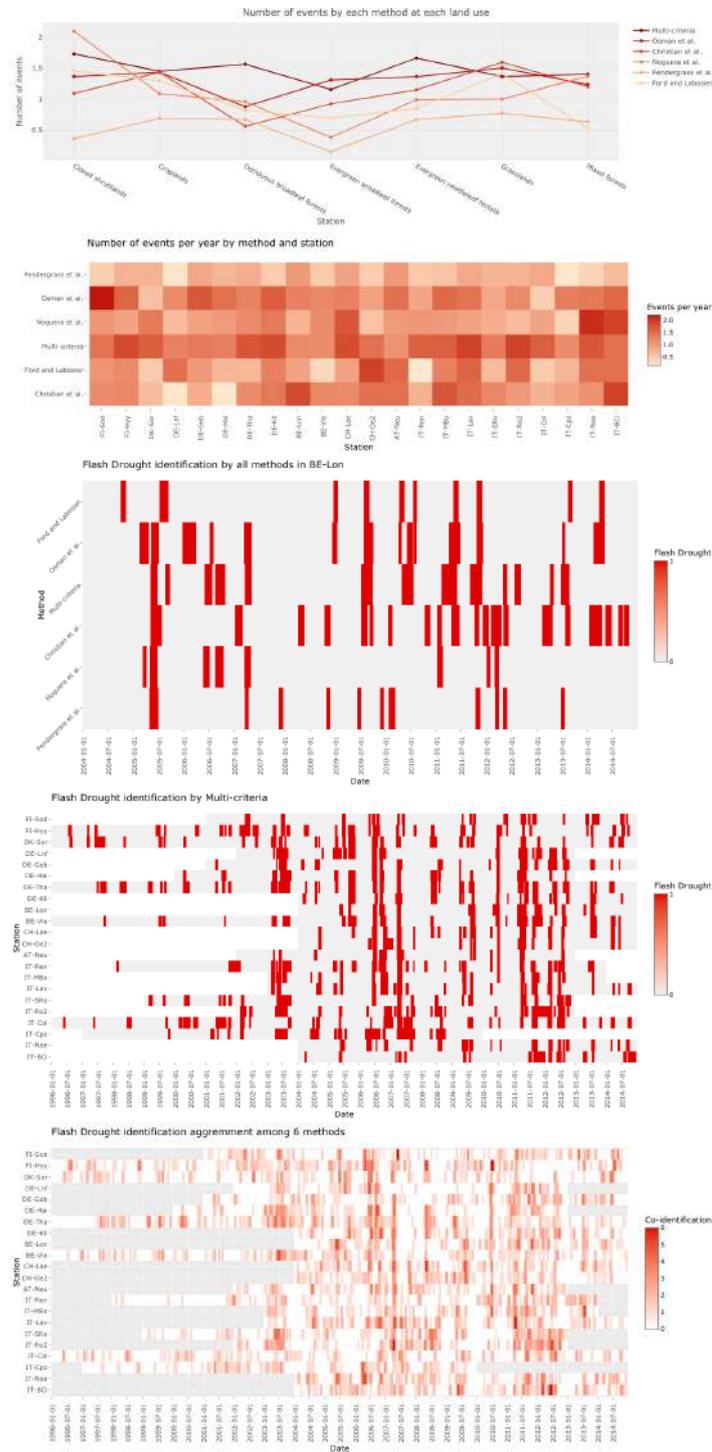
Figure 6.4: From top to bottom, (1) the average number of events per year by method and land use; (2) the average number of events per year by method and station; (3) the co-identification of all methods in a selected station; (4) the identification of events by one method on all stations; and (5) the co-identification of events by all methods in all stations, colour-coded according to the number of methods that identify the same period as a flash drought. In graphs (4) and (5), the y-axis with stations is organized by latitude.

7

Summary, conclusions, and outlook

This chapter summarizes the findings of Chapters 2 to 6 and relate them to the research questions and objectives presented in Section 1.5 of the Introduction. This chapter is structured following the three main topics of the research (Section 1.3): Gully Erosion, Sediment Yield, and Flash Droughts. It also brings a view on potential future advances in those fields.

7.1 Gully erosion

> **Objectives:**
>
> - Using field experiments and modelling, identify the key variables to model long term gully erosion and build a framework that allows the addition of multiple sources of energy and sediment;
>
> - Using the principle of minimum cross-entropy, model the key variables related to gully erosion, namely shear stress in walls and bed.

To answer the research question 1, presented in Section 1.5, two research papers were prepared. In the first paper (Chapter 2; Alencar, de Araújo, and Teixeira, 2020) we presented a physical approach to assessing long term erosion, identifying shear stress in the channel bed and walls, caused by large precipitation and runoff, to be the key variable to simulate gully erosion growing over time and to estimate its total soil loss.

The authors propose a physically-based method with the combination of two previous models (Foster and Lane, 1983, and Sidorchuk, 1999), and allow the simulation of multiple precipitation events in a long term analysis. The new model (FL-SM) was validated to small permanent gullies in the Brazilian Semiarid Region. Such gullies are a common problem in

the region. Unlike the conventional permanent gullies, small permanent gullies are limited in size due to shallow soils and lack of groundwater (Poesen, Torri, and Van Walleghem, 2011). Additionally, they are permanent because they remain unremediated for long periods, particularly when occurring in unclaimed land (Alencar, de Araújo, and Teixeira, 2020). Although small permanent gullies do not impose a significant obstacle to movement and life (de Vente and Poesen, 2005), they cause a significant change in sediment connectivity, are relevant sources of sediment and decrease land productivity and biodiversity (De Vente et al., 2013; Bagnold, 1977; Poesen, 2018).

A second paper, presented by the authors in Chapter 3, a new model for gully erosion is presented. In this new Entropy-Based Gully Erosion Model (EBGEM), a new shear stress distribution is defined, with the assistance of the principle of minimum cross-entropy (Kullback, 1978a). The equation was validated with published experimental data (Knight, Yuen, and Al Hamid, 1994; Tominaga et al., 1989; Bonakdari, Sheikh, and Tooshmalani, 2014a; Khozani and Bonakdari, 2018).

The new gully erosion model, as its predecessor (FL-SM), allows the simulation of long-term erosion processes, but on larger scales. The model was validated using data from three gully affected sites with different scales, with gully cross-sections ranging from 10^{-1} to 10^{+1} m^2 and presented good efficiency (Moriasi et al., 2007; NSE > 0.65) when modelling gullies with catchment areas up to 8 hectares.

From those promising results, the EBGEM equip hydrologists with a tool to assess and gully erosion on a larger scale than previous models (Bennett and Wells, 2019; Douglas-Mankin et al., 2020). Coupled with gully erosion risk assessment models (Javidan et al., 2019; Arabameri, Pradhan, et al., 2019; Arabameri, Cerda, and Tiefenbacher, 2019), the EBGEM can also be used as a planning tool to estimate sediment yield originated in gullies and automatically update sediment connectivity in the catchment.

Looking back to our research question *"What are the key variables of erosion by gullies and how to model such processes?"*, we are able to conclude that shear stress is a key variable for assessing erosion in gullies and that a better shear stress distribution, as that obtained in Chapter 3 improved significantly the assessment of erosion in gully systems of multiple scales. However, there are other variables that request further exploration as the critical shear stress and rill erodibility, which were assumed constant over the time of the study (that comprises decades). There are also some uncertainties to be considered, as the multiple processes not taken under consideration in the study that might play a relevant role in gully erosion, as head-cut retreat and armouring.

7.2   Sediment Yield

Objective:

- Using the principle of maximum entropy, propose a methodology to assess sub-daily precipitation features (duration and 30-minute intensity) that allow SDR and sediment yield estimations in ungauged catchments.

Relative to research question 2 (Section 1.5), we presented in Chapter 4 a novel method to assess sub-daily precipitation features. The method allows estimation of the 30-minute intensity, necessary to estimate gross erosion (Wischmeier and Smith, 1978), and of the event duration, necessary to assess the sediment delivery ratio (SDR - de Araújo, 2007).

In their model, de Araújo (2007) successfully introduced precipitation patterns and vegetation cover to the assessment of sediment yield via the novel parameter ϕ, that is the average travel distance of a sediment particle and is a function of vegetation cover, stream power (Ω), topography, and soil properties. Stream power is calculated as a function of precipitation intensity, and therefore required the knowledge of sub-daily features of each rainfall event, which is the main disadvantage of the model, that was successfully validated in gauged basins.

To make the model viable also in ungauged basins, we proposed the use of the principle of maximum entropy and a Monte Carlo approach of successive sampling of random seeds to obtain a third-generation model, as proposed by Sidorchuk (2009), on an approach that involves deterministic and stochastic variables (Sidorchuk, 2015; Singh, 2018). By using available data from five gauged stations in the Northeastern Region of Brazil, we identified that the best distribution of probabilities for precipitation duration follows the shape of a gamma distribution. We calibrated and validated the parameters of the distribution and assess sediment yield in the studied catchments. We also assessed the loss of information caused by the transference of those parameters/distributions to other regions. We concluded that the loss of information is connected not necessarily to the region, but rather to rain formation processes (convective, stratiform, orographic), topography and total precipitation.

The temporal downscaling of precipitation data, based on the principle of maximum entropy (Jaynes, 1957a; Jaynes, 1957b), coupled with the sediment yield model proposed by de Araújo (de Araújo, 2007; de Araújo, 2017) improved significantly the performance of sediment yield assessment, to a Nash-Sutcliffe Efficiency of 0.96.

This encouraging result indicates that the coupled methodology, proposed in Chapter 2 (Alencar, Paton, and de Araújo, 2021), can be used to accurately assess sediment yield and siltation in ungauged basins. The methodology required a low level of calibration, with the parameters of the *Gamma distribution* easily obtained using the available script in the Appendix 3.5, and the parameter ϕ calculated by the methodology proposed by de Araújo (2007).

The novel methodology provides hydrologists and stakeholders with a powerful tool to plan sediment management (Peng, Yonggang, and Yongming, 2011; Kondolf et al., 2014; Morris and Fan, 1998; Braga et al., 2019) and increase the lifetime of infrastructure, guaranteeing the sustainability of their benefits and welfare of its users (Landwehr, 2021; Zarfl et al., 2015). Conventional methods of sediment yield that relied only on topographical relations (Maner, 1958; Roehl, 1962) could not take into account the effect of single and extreme events. Our methodology also allows the study of the impacts of such events and forecast siltation for design events, e.g., how much sediment would be carried in the precipitation event with a 1000-year recurrence interval.

As an answer to our question, *"How can we introduce relevant variables as precipitation patterns and features to assess sediment yield and delivery ratio in ungauged basins?"*, we can say that entropy theory is an efficient path to assess sediment yield in ungauged basins. Using the POME we can estimate precipitation duration and intensity using probability density functions derived with data from gauged basins that have similar rain formation processes and total annual precipitation. In the study, some uncertainties could not be cleared. The effect of using parameters from another region was not completely explores since most catchments with available siltation data are ungauged. Also, the effect of topography as in local discontinuities and sediment sinks within the catchment area is still unclear.

7.3 Flash drought

> **Objectives:**
>
> - Propose a new definition of flash droughts and a novel method, based on multiple variables and thresholds, to identify them;
>
> - Compare methods from the literature and their performance on identifying events in Central Europe, building a visual platform to easily observe discrepancies between methods;
>
> - Implement all selected methods in R-language and build an open-source package to be available to the community.

Relative to the research questions 3 and 4 (Section 1.5), we presented in Chapter 5, a study with the comparison of multiple well-published methods in identifying flash drought events in croplands in Central Europe. From the selected methods and literature review (Mo and Lettenmaier, 2016; Ford and Labosier, 2017; Christian, Basara, Otkin, Hunt, et al., 2019; Noguera, Castro, and Serrano, 2020; Pendergrass et al., 2020; Lisonbee, Woloszyn, and Skumanich, 2021; Otkin, Svoboda, et al., 2018), we propose a new definition for flash droughts.

> "We define a flash drought as the process of **rapid**, **accelerated** and **unusually large** depletion of **soil moisture** in comparison with 'average'

> moisture conditions due to the **simultaneous or concurrent** occurrence
> of two or more atmospheric and/or weather conditions over a **short time**
> frame of several weeks during the main **growing season**."

In our definition, we summarize the main features usually associated with flash drought, rapid onset and intensification, soil moisture depletion, compound drivers and short duration. The definition brings the plant (crop, grass, or forest) as the main user of soil moisture and main affected subject when it points that flash drought should be considered in terms of the growing season. None of the studied methods in Chapter 5 had such temporal constraint. Nevertheless, it is our understanding that depletion of soil moisture during periods of low vegetation activity, as winter in Central Europe, should not be considered as a flash drought event, as no impact on vegetation health is expected.

By implementing and comparing the multiple methods in Chapter 5, we also observed a significant disagreement on the identification among different methods. This disagreement can be summarized into four key factors: (1) variables show opposite behaviour; (2) velocity of intensification is lower than required; (3) final value of intensification is marginally lower (or higher) than required; and (4) the event does not last as long as required.

The studied methods were compared to the reference method by Osman et al. (2021) and the co- and misidentification were measured using synchronicity, and confusion matrix derived metrics. These metrics made explicit that, for methods that do not rely on soil moisture, i.e., *proxy* methods, an ensemble approach that used multiple variables and indexes, better balances out the strength and weaknesses of each variable than single-variable based methods, allowing different kinds of flash droughts that occur due to combination of different compound drivers to be identified.

In Chapter 6, we present the visualization tool built and publicly available. It allows the user to compare different methods in various regions in Central Europe and explore the drivers to Flash Drought occurrence, co- and misidentification.

To answer our questions *"What is the definition of flash droughts, what are their key variables, and how to identify them?"*, we gave a definition of what a flash drought is, concerning particularly croplands in temperate climate as in Central Europe (Section 5.1). The key variable related to flash droughts and their identification is soil moisture in the root zone. Nevertheless, this data is often scarce and proxy variables can be used, as precipitation, evapotranspiration, and temperature. In our work, some uncertainty sources cannot be discarded, namely the short term analysis (11 to 14 years) due to lack of soil moisture data.

7.4   Outlook

Our last research question addressed the multiple tools available for hydrologists to study extreme events and their impacts. We explored the use of two of those tools and successfully implemented them to problems of different nature, scale, timeframe, and location. From

Information Theory, we borrowed concepts as Entropy (maximum and minimum cross-entropy), cross information, divergence, and statistical distance (Chapters 3 and 4). From Data Science we explored big data, visualization, supervised classification (confusion matrix), and time series analysis (synchronicity).

Nevertheless, this dissertation is not sufficient to close the studied processes, and new questions have been stated based on the advances presented. In this section, we present these questions and give some perspective of future developments.

Regarding sediment yield and the use of entropy theory to assess event-based sediment delivery ratio, additional studies should be carried out to test and assess the most suited probability distribution families to precipitation data, especially 30-minute intensity. Efforts are still necessary to validate the method's potential concerning regionalization. It is not at all a trivial matter to determine which factors (relief, climate, position, etc.) influence homogeneity between regions, and therefore produce similar PDFs. The method was also validated only in the Brazilian Semiarid Region. Although the region has over one million square kilometres, there is still some homogeneity, especially regarding annual precipitation and rainfall formation processes. Precipitation in the region is predominantly convective, which has as features being intense and short, with precipitation rarely exceeding 24 hours (de Figueiredo et al., 2016). The temporal downscaling method should be implemented in regions with different rain processes (particularly stratiform) and other, more temperate climates.

From the results in Chapter 4 (Figure 4.5) we observe that catchments with longer time series present a narrower distribution for SDR values. However, such time series are not often available. Although the use of confidence intervals in modelling has advantages, as providing an easy-to-read measure of uncertainty, a single value of SDR could be easier to be implemented in existing models. The use of synthetic series, as with the use of bootstrapping (Ritter and Muñoz-Carpena, 2013; Ng et al., 2019), could provide modellers with a single-value SDR. Also, the use of Bayesian metrics and statistics (e.g., Bayesian evidence for model comparison) and transfer entropy could provide a better insight into the effects of using probability density function parameters in a gauged basin in an ungauged one.

Concerning gully erosion modelling, the new EBGEM presented encouraging advances in model efficiency and applicability. Nevertheless, there are still various limitations that should be explored. The upper limit of 8 hectares for the contribution area is a piece of evidence that not all energy and sediment sources are taken under consideration by the model. Processes such as head cut (Alonso, Bennett, and Stein, 2002), pipping (Bernatek-Jakiel and Wrońska-Wałach, 2018), armouring (Bennett and Nordin, 1977), and flow jets (Nearing, 1991) should be explored further studied. These processes have also a strong stochastic component (Sidorchuk, 2005; Sidorchuk, 2015) and could benefit from a similar approach as the one used in Chapters 3 and 4, combining entropy theory and Monte Carlo methods (Binder et al., 1993; Vrugt et al., 2008; Hayas et al., 2017).

Droughts, and more specifically flash drought research, as a new topic, is an exciting and rapidly evolving topic. A definition of the basic features of a flash drought, as the one

presented in Chapter 5 should be debated by the scientific community in forums. Also, the definition of measurable indexes (Lisonbee, Woloszyn, and Skumanich, 2021), not only for flash drought identification, but for assessment of the severity of events (similar to SPI and SPEI; McKee, Doesken, Kleist, et al., 1993; Serrano, Begueria, and Moreno, 2010) is still missing (Otkin, Zhong, Hunt, Christian, et al., 2021). There is still no discussion on the definition of different kinds of rapid-intensification flash droughts (Mishra and Singh, 2010; Mo and Lettenmaier, 2016). The use of entropy and copula theory (AghaKouchak, 2014; Singh and Zhang, 2018), besides the already implemented big-data analysis, has the potential to advance our understanding of the multiple compound events associated with flash droughts. The same approach can be used to understand changes in extreme events, as illustrated in Chapter 1 (Figures 1.1 and 1.2).

Finally, extreme wet and dry weather events are not separated phenomena and, as stated by Herold et al. (2021), the frequency of alternance between the two is increasing. Dry weather leads to a reduction of vegetation cover and consequent increment in potential erosion. When followed by very intense precipitation, an increase in both laminar and linear erosion are expected, given that the precipitation surpasses infiltration. There is very little research on the effects of such drought-flood compound events over erosion and siltation (Fraticelli, 2006), particularly on gullies.

In this new moment of data evolution and changing climate, hydrologists have the task of using the available tools (and when not available, develop new ones) to help the global community in achieving our sustainable development goals. Only by improving our understanding of the processes that lead to extremely dry and wet events and their impacts, and therefore a better understanding of the water cycle (the driver of all nature and change), we can improve life in the uncertain future.

References

Aagaard, T. and M.G. Hughes (2021). "Sediment Transport". In: *Reference Module in Earth Systems and Environmental Sciences.* Elsevier.

AghaKouchak, Amir (Dec. 2014). "Entropy–Copula in Hydrology and Climatology". In: *Journal of Hydrometeorology* 15.6, pp. 2176–2189. DOI: 10.1175/jhm-d-13-0207.1.

Agüera-Vega, Francisco, Fernando Carvajal-Ramírez, and Patricio Martínez-Carricondo (2017). "Assessment of photogrammetric mapping accuracy based on variation ground control points number using unmanned aerial vehicle". In: *Measurement* 98, pp. 221–227.

Alberts, E E, M A Nearing, M A Weltz, L M Risse, F B Pierson, X C Zhang, J M Laflen, and J R Simanton (1995). *Soil Component.* Tech. rep. July. USDA. Chap. Chapter 7, pp. 7.1–7.47.

Alberts, E. E., J. M. Laflen, W. J. Rawls, J. R. Simanton, and M. A. Nearing (1989). "Chapter 6: Soil Component". In: *USDA-Water Erosion Project: Hillslope profile model documentation (LJ Lane and MA Nearing, eds.), National Soil Erosion Research Laboratory, Report* 2.

Alencar, Pedro Henrique Lima (2018). "Medição e Modelagem de Voçorocas no Bioma Caatinga: O Caso da Bacia Representativa e Madalena, CE". Master's thesis. Universidade Federal do Ceará, p. 80.

Alencar, Pedro Henrique Lima, J. C. de Araújo, and A. S. Teixeira (Aug. 2020). "Physically based model for gully simulation: application to the Brazilian semiarid region". In: *Hydrology and Earth System Sciences* 24.8, pp. 4239–4255. DOI: 10.5194/hess-24-4239-2020.

Alencar, Pedro Henrique Lima, Eva Nora Paton, and José Carlos de Araújo (Dec. 2021). "Entropy-Based Temporal Downscaling of Precipitation as Tool for Sediment Delivery Ratio Assessment". In: *Entropy* 23.12, p. 1615. DOI: 10.3390/e23121615.

Allen, R. G. (1998). *Crop evapotranspiration : guidelines for computing crop water requirements.* Rome: Food and Agriculture Organization of the United Nations. ISBN: 9251042195.

Alonso, Carlos V., Sean J. Bennett, and Otto R. Stein (2002). "Predicting head cut erosion and migration in concentrated flows typical of upland areas". In: *Water Resources Research* 38.12, pp. 39–1–39–15. DOI: 10.1029/2001wr001173.

ANA (2019). *Rede Hidrometeorológica Nacional.* https://metadados.ana.gov.br/geonetwork/. Visited on 20/03/2020.

Andrade, Eunice Maia, Maria João Simas Guerreiro, Helba Araújo Queiroz Palácio, and Diego Antunes Campos (Feb. 2020). "Ecohydrology in a Brazilian tropical dry forest: thinned vegetation impact on hydrological functions and ecosystem services". In: *Journal of Hydrology: Regional Studies* 27, p. 100649. DOI: 10.1016/j.ejrh.2019.100649.

REFERENCES

Arabameri, Alireza, A Cerda, and John P. Tiefenbacher (2019). "Spatial Pattern Analysis and Prediction of Gully Erosion Using Novel Hybrid Model of Entropy-Weight of Evidence". In: *Water* 11.6, p. 1129. ISSN: 2073-4441.

Arabameri, Alireza, Biswajeet Pradhan, Khalil Rezaei, and Christian Conoscenti (2019). "Gully erosion susceptibility mapping using GIS-based multi-criteria decision analysis techniques". In: *Catena* 180.May, pp. 282–297. ISSN: 03418162. DOI: 10.1016/j.catena.2019.04.032.

Arabameri, Alireza, Khalil Rezaei, Hamid Reza Pourghasemi, Saro Lee, and Mojtaba Yamani (2018). "GIS-based gully erosion susceptibility mapping: a comparison among three data-driven models and AHP knowledge-based technique". In: *Environmental Earth Sciences*. DOI: 10.1007/s12665-018-7808-5.

Aragão Araújo, José Amaury (1990). *Dams in the Northeast of Brazil*. DNOCS.

Ascough, J C, C Baffaut, M A Nearing, and B Y Liu (1997). "The WEPP watershed model .1. Hydrology and erosion". In: *Transactions of the Asae* 40.4, pp. 921–933.

Avila, Humberto Fabián and Boris Daniel Avila (Aug. 2015). "Spatial and temporal estimation of the erosivity factor R based on daily rainfall data for the department of Atlántico, Colombia". In: *Ingenieria e Investigación* 35.2, pp. 23–29. DOI: 10.15446/ing.investig.v35n2.47773.

Avni, Y (2005). "Gully incision as a key factor in desertification in an arid environment, the Negev highlands, Israel". In: *Catena* 63.2-3, pp. 185–220.

Azareh, Ali, Omid Rahmati, Elham Rafiei-Sardooi, Joel B. Sankey, Saro Lee, Himan Shahabi, and Baharin Bin Ahmad (2019). "Modelling gully-erosion susceptibility in a semi-arid region, Iran: Investigation of applicability of certainty factor and maximum entropy models". In: *Science of the Total Environment*. ISSN: 18791026. DOI: 10.1016/j.scitotenv.2018.11.235.

Bagarello, Vincenzo, Vito Ferro, and Vincenzo Pampalone (Jan. 2020). "A comprehensive analysis of Universal Soil Loss Equation-based models at the Sparacia experimental area". In: *Hydrological Processes* 34.7, pp. 1545–1557. DOI: 10.1002/hyp.13681.

Bagnold, R. A. (1966). *An Approach to the Sediment Transport Problem from General Physics*. Tech. rep. U. S. Departament of the Interior, p. 42. DOI: 10.1017/S0016756800049074.

— (Apr. 1977). "Bed load transport by natural rivers". In: *Water Resources Research* 13.2, pp. 303–312. DOI: 10.1029/wr013i002p00303.

Baldassarre, G. Di, A. Castellarin, and A. Brath (Aug. 2006). "Relationships between statistics of rainfall extremes and mean annual precipitation: an application for design-storm estimation in northern central Italy". In: *Hydrology and Earth System Sciences* 10.4, pp. 589–601. DOI: 10.5194/hess-10-589-2006.

Basara, Jeffrey B, Jordan I Christian, Ryann A Wakefield, Jason A Otkin, Eric H Hunt, and David P Brown (July 2019). "The evolution, propagation, and spread of flash drought in the Central United States during 2012". In: *Environmental Research Letters* 14.8, p. 084025. DOI: 10.1088/1748-9326/ab2cc0.

Beckers, Felix (2018). "Uncertainty analysis of a 2D sediment transport model : an example of the Lower River Salzach". In: pp. 3133–3144. DOI: 10.1007/s11368-017-1816-z.

Bennett, James P and Carl F Nordin (1977). "Simulation of Sediment Transport and Armouring". In: *Hydrological Sciences-Journal-des Sciences Hydrologiques* xxii.4, pp. 555–569.

Bennett, Sean J. and Robert R. Wells (2019). "Gully erosion processes, disciplinary fragmentation, and technological innovation". In: *Earth Surface Processes and LandformsBeschuldigung* 44.1, pp. 46–53. ISSN: 10969837. DOI: 10.1002/esp.4522.

Bernard, Jerry, Jerry Lemunyon, Bill Merkel, Fred Theurer, Norm Widman, Ron Bingner, Seth Dabney, Eddy Langendoen, Robert Wells, and Glenn Wilson (2010). "Ephemeral Gully Erosion — A National Resource Concern." In: *U.S. Department of Agriculture. NSL Technical Research Report No. 69* 69, 67 pp.

Bernatek-Jakiel, Anita and Dominika Wrońska-Wałach (June 2018). "Impact of piping on gully development in mid-altitude mountains under a temperate climate: A dendrogeomorphological approach". In: *CATENA* 165, pp. 320–332. DOI: 10.1016/j.catena.2018.02.012.

Bhunya, P K, R Berndtsson, C S P Ojha, and S K Mishra (2007). "Suitability of Gamma, Chi-square, Weibull, and Beta distributions as synthetic unit hydrographs". In: *Journal of Hydrology* 334.1-2, pp. 28–38. ISSN: 00221694. DOI: 10.1016/j.jhydrol.2006.09.022.

Binder, Kurt, Dieter Heermann, Lyle Roelofs, A John Mallinckrodt, and Susan McKay (1993). "Monte Carlo simulation in statistical physics". In: *Computers in Physics* 7.2, pp. 156–157. DOI: 10.1063/1.4823159.

Bingner, R. L., R. R. Wells, H. G. Momm, J. R. Rigby, and F. D. Theurer (2016). "Ephemeral gully channel width and erosion simulation technology". In: *Natural Hazards* 80.3, pp. 1949–1966. DOI: 10.1007/s11069-015-2053-7.

Blong, R. J. and J. A. Veness (1982). "The Role of Sidewall Processes in Gully Development". In: *Earth Surface Processes and Landforms* 7, pp. 381–385.

Bocco, Gerardo (1991). "Gully erosion: Processes and models". In: *Progress in Physical Geography* 15.4, pp. 392–406.

Bonakdari, Hossein, Sultan Noman Qasem, Isa Ebtehaj, Amir Hossein Zaji, Bahram Gharabaghi, and Marjan Moazamnia (Feb. 2020). "An expert system for predicting the velocity field in narrow open channel flows using self-adaptive extreme learning machines". In: *Measurement* 151, p. 107202. DOI: 10.1016/j.measurement.2019.107202.

Bonakdari, Hossein, Zohreh Sheikh, and Mohtaram Tooshmalani (Sept. 2014a). "Comparison between Shannon and Tsallis entropies for prediction of shear stress distribution in open channels". In: *Stochastic Environmental Research and Risk Assessment* 29.1, pp. 1–11. DOI: 10.1007/s00477-014-0959-3.

— (Sept. 2014b). "Comparison between Shannon and Tsallis entropies for prediction of shear stress distribution in open channels". In: *Stochastic Environmental Research and Risk Assessment* 29.1, pp. 1–11. DOI: 10.1007/s00477-014-0959-3.

Borrelli, Pasquale, David A. Robinson, Larissa R. Fleischer, Emanuele Lugato, Cristiano Ballabio, Christine Alewell, Katrin Meusburger, Sirio Modugno, Brigitta Schütt, Vito Ferro, Vincenzo Bagarello, Kristof Van Oost, Luca Montanarella, and Panos Panagos (2017). "An assessment of the global impact of 21st century land use change on soil erosion". In: *Nature Communications* 8.1. ISSN: 20411723. DOI: 10.1038/s41467-017-02142-7.

Braga, Brennda Bezerra, Thayslan Renato Anchieta de Carvalho, Arlena Brosinsky, Saskia Foerster, and Pedro Henrique Augusto Medeiros (June 2019). "From waste to resource: Cost-benefit analysis of reservoir sediment reuse for soil fertilization in a semiarid catchment".

REFERENCES

In: *Science of The Total Environment* 670, pp. 158–169. DOI: 10.1016/j.scitotenv.2019.03.083.

Bravar, L and ML Kavvas (1991). "On the physics of droughts. I. A conceptual framework". In: *Journal of Hydrology* 129.1-4, pp. 281–297.

Brice, James C. (1966). "Erosion and Deposition in the Loess-Mantled Great Plains , Medicine Creek Drainage Basin , Nebraska Erosion and Deposition in the Loess-Mantled Great Plains , Medicine Creek Drainage Basin , Nebraska". In: *Geological Survey Professional Paper 352 H*.

Brigandì, Giuseppina and Giuseppe T. Aronica (2019). "Generation of sub-hourly rainfall events through a point stochastic rainfall model". In: *Geosciences (Switzerland)* 9.5, pp. 11–14. ISSN: 20763263. DOI: 10.3390/geosciences9050226.

Brito, S. S. B., A. P. M. A. Cunha, C. C. Cunningham, R. C. Alvalá, J. A. Marengo, and M. A. Carvalho (Aug. 2017). "Frequency, duration and severity of drought in the Semiarid Northeast Brazil region". In: *International Journal of Climatology* 38.2, pp. 517–529. DOI: 10.1002/joc.5225.

Brune, Gunnar M (1953). "Trap efficiency of reservoirs". In: *Eos, Transactions American Geophysical Union* 34.3, pp. 407–418.

Bullock, Jane A., George D. Haddow, and Damon P. Coppola (2018). "3 - Hazards". In: *Homeland Security (Second Edition)*. Ed. by Jane A. Bullock, George D. Haddow, and Damon P. Coppola. Second Edition. Butterworth-Heinemann, pp. 45–66.

Cadier, Eric (1994). *Hidrologia das pequenas bacias do Nordeste semi-árido: transposição hidrológica*. Ed. by SUDENE/ORSTOM. Hidrologia 31. Recife: Sudene, p. 448.

Casalı, J, JJ López, and JV Giraldez (2003). "A process-based model for channel degradation: application to ephemeral gully erosion". In: *Catena* 50.2-4, pp. 435–447.

Castillo, C. and J. A. Gómez (2016). "A century of gully erosion research: Urgency, complexity and study approaches". In: *Earth-Science Reviews* 160, pp. 300–319. ISSN: 00128252. DOI: 10.1016/j.earscirev.2016.07.009.

Cerdan, Olivier, Jean Poesen, Gérard Govers, Nicolas Saby, Yves Le Bissonnais, Anne Gobin, Andrea Vacca, John Quinton, Karl Auerswald, Andreas Klik, Franz F P M Kwaad, and M J Roxo (2006). "Sheet and Rill Erosion". In: *Soil Erosion in Europe*. Ed. by John Boardman and Jean Poesen. John Wiley & Sons Ltd. Chap. 38, pp. 501–513. ISBN: 9780470859100. DOI: 10.1002/0470859202.ch38.

Chen, L. Gwen, Jon Gottschalck, Adam Hartman, David Miskus, Rich Tinker, and Anthony Artusa (Aug. 2019). "Flash Drought Characteristics Based on U.S. Drought Monitor". In: *Atmosphere* 10.9, p. 498. DOI: 10.3390/atmos10090498.

Chen, Lu, Vijay P. Singh, and Feng Xiong (2017). "An entropy-based generalized gamma distribution for flood frequency analysis". In: *Entropy* 19.6. ISSN: 10994300. DOI: 10.3390/e19060239.

Chen, Yiheng and Dawei Han (2016). "Big data and hydroinformatics". In: *Journal of Hydroinformatics* 18.4, pp. 599–614.

Chicco, Davide, Niklas Tötsch, and Giuseppe Jurman (Feb. 2021). "The Matthews correlation coefficient (MCC) is more reliable than balanced accuracy, bookmaker informedness, and

markedness in two-class confusion matrix evaluation". In: *BioData Mining* 14.1. DOI: `10.1186/s13040-021-00244-z`.

Chiu, Chao Lin (1988). "Entropy and 2-D velocity distribution in open channels". In: *J. Hydraul. Eng.* 114.7, pp. 738–756. DOI: `https://doi.org/10.1061/(ASCE)0733-9429(1988)114: 7(738)`.

Chiu, Chao-Lin (May 1987). "Entropy and Probability Concepts in Hydraulics". In: *Journal of Hydraulic Engineering* 113.5, pp. 583–599. DOI: `10.1061/(asce)0733-9429(1987)113: 5(583)`.

— (May 1991). "Application of Entropy Concept in Open-Channel Flow Study". In: *Journal of Hydraulic Engineering* 117.5, pp. 615–628. DOI: `10.1061/(asce)0733-9429(1991)117: 5(615)`.

Chiu, Chao-Lin and Ning-Chien Tung (Apr. 2002). "Maximum Velocity and Regularities in Open-Channel Flow". In: *Journal of Hydraulic Engineering* 128.4, pp. 390–398. DOI: `10.1061/(asce)0733-9429(2002)128:4(390)`.

Chow, Ven Te (1959). *Open-Channel Hidraulics*. New York: McGraw-Hill, p. 728.

Chow, Ven Te, David R Maidment, and Larry W Mays (1988). *Applied Hydrology*. ISBN: 0070108102.

Christian, Jordan I, Jeffrey B Basara, Eric D Hunt, Jason A Otkin, and Xiangming Xiao (Sept. 2020). "Flash drought development and cascading impacts associated with the 2010 Russian heatwave". In: *Environmental Research Letters* 15.9, p. 094078. DOI: `10.1088/1748-9326/ab9faf`.

Christian, Jordan I, Jeffrey B Basara, Jason A Otkin, and Eric D Hunt (Nov. 2019). "Regional characteristics of flash droughts across the United States". In: *Environmental Research Communications* 1.12, p. 125004. DOI: `10.1088/2515-7620/ab50ca`.

Christian, Jordan I., Jeffrey B. Basara, Jason A. Otkin, Eric D. Hunt, Ryann A. Wakefield, Paul X. Flanagan, and Xiangming Xiao (May 2019). "A Methodology for Flash Drought Identification: Application of Flash Drought Frequency across the United States". In: *Journal of Hydrometeorology* 20.5, pp. 833–846. DOI: `10.1175/jhm-d-18-0198.1`.

Coelho, Christine, Birgit Heim, Saskia Foerster, Arlena Brosinsky, and José Carlos de Araújo (2017). "In situ and satellite observation of CDOM and chlorophyll-a dynamics in small water surface reservoirs in the brazilian semiarid region". In: *Water (Switzerland)* 9.12. DOI: `10.3390/w9120913`.

Conoscenti, Christian, Valerio Agnesi, Mariaelena Cama, Nathalie Alamaru Caraballo-Arias, and Edoardo Rotigliano (2018). "Assessment of Gully Erosion Susceptibility Using Multivariate Adaptive Regression Splines and Accounting for Terrain Connectivity". In: *Land Degradation and Development* 29.3, pp. 724–736. ISSN: 1099145X. DOI: `10.1002/ldr. 2772`.

Conoscenti, Christian and Edoardo Rotigliano (June 2020). "Predicting gully occurrence at watershed scale: Comparing topographic indices and multivariate statistical models". In: *Geomorphology* 359, p. 107123. DOI: `10.1016/j.geomorph.2020.107123`.

Coppus, R and AC Imeson (2002). "Extreme events controlling erosion and sediment transport in a semi-arid sub-Andean valley". In: *Earth Surface Processes and Landforms: The Journal of the British Geomorphological Research Group* 27.13, pp. 1365–1375.

REFERENCES

Crouch, R. J. (1987). "The relationship of gully sidewall shape to sediment production". In: *Australian Journal of Soil Research* 25.4, pp. 531–539.

Cui, Huijuan and Vijay P. Singh (Mar. 2013). "Two-Dimensional Velocity Distribution in Open Channels Using the Tsallis Entropy". In: *Journal of Hydrologic Engineering* 18.3, pp. 331–339. DOI: 10.1061/(asce)he.1943-5584.0000610.

Cui, Huijuan, Bellie Sivakumar, and Vijay P. Singh (2018). *Entropy applications in environmental and water engineering.* Vol. 20. 8. ISBN: 9783038972228. DOI: 10.3390/e20080598.

D'Odorico, Paolo, Kyle Frankel Davis, Lorenzo Rosa, Joel A. Carr, Davide Chiarelli, Jampel Dell'Angelo, Jessica Gephart, Graham K. MacDonald, David A. Seekell, Samir Suweis, and Maria Cristina Rulli (July 2018). "The Global Food-Energy-Water Nexus". In: *Reviews of Geophysics* 56.3, pp. 456–531. DOI: 10.1029/2017rg000591.

Dabney, S. M., D. A. N. Vieira, D. C. Yoder, E. J. Langendoen, R. R. Wells, and M. E. Ursic (2015). "Spatially Distributed Sheet, Rill, and Ephemeral Gully Erosion". In: *Journal of Hydrologic Engineering* 20.6, p. C4014009. ISSN: 1084-0699.

Dash, Ch. Jyotiprava, N. K. Das, and Partha Pratim Adhikary (June 2019). "Rainfall erosivity and erosivity density in Eastern Ghats Highland of east India". In: *Natural Hazards* 97.2, pp. 727–746. DOI: 10.1007/s11069-019-03670-9.

Da Silva, Alisson Jadavi Pereira and Márcio Lima Rios (2018). *Alerta de desertificação no médio curso do Rio Salitre, afluente do Rio São Francisco.* Tech. rep. Senhor do Bonfim, Bahia: IFBA, p. 48.

de Araújo, J C (2003). "Assoreamento em Reservatórios do Semi-árido : Modelagem e Validação". In: *RBRH - Revista Brasileira de Recursos Hídricos* 8, pp. 39–56.

— (2007). "Entropy-based equation to assess hillslope sediment production". In: *Earth Surface Processes and Landforms* 32.1, pp. 2005–2018. ISSN: 01851101. DOI: 10.1002/esp.1502.

De Vente, Joris, Jean Poesen, Gert Verstraeten, Gerard Govers, Matthias Vanmaercke, Anton Van Rompaey, Mahmood Arabkhedri, Carolina Boix-fayos, Joris De Vente, Jean Poesen, Gert Verstraeten, Gerard Govers, Matthias Vanmaercke, Anton Van Rompaey, Mahmood Arabkhedri, and Carolina Boix-fayos (2013). "Predicting soil erosion and sediment yield at regional scales: Where do we stand?" In: *Earth-Science Reviews* 127, pp. 16–29. DOI: 10.1016/j.earscirev.2013.08.014.

De Araújo, José Carlos (2017). "Entropy-based equation to assess sediment yield: application to ungauged basins". In: *Proceeding of the IASWS 2017 Symposium.* Taurmina, Italy, pp. 2005–2018.

De Araújo, José Carlos and Axel Bronstert (2015). "A method to assess hydrological drought in semi-arid environments and its application to the Jaguaribe River basin, Brazil". In: *Water International* 41.2, pp. 213–230. ISSN: 02508060. DOI: 10.1080/02508060.2015.1113077.

De Araújo, José Carlos and Fazal H. Chaudhry (Oct. 1998). "Experimental Evaluation of 2-D Entropy Model for Open-Channel Flow". In: *Journal of Hydraulic Engineering* 124.10, pp. 1064–1067. DOI: 10.1061/(asce)0733-9429(1998)124:10(1064).

De Araújo, José Carlos, Andreas Güntner, and Axel Bronstert (2006). "Loss of reservoir volume by sediment deposition and its impact on water availability in semiarid Brazil / Perte de volume de stockage en réservoirs par sédimentation et impact sur la disponibilité en eau au

Brésil semi-aride Loss of reservoir volume by se". In: *Hydrological Sciences Journal* 51.1, pp. 157–170. DOI: `10.1623/hysj.51.1.157`.

De Araújo, José Carlos and Julio Iván González Piedra (2009). "Comparative hydrology: analysis of a semiarid and a humid tropical watershed". In: *Hydrological Process* 23, pp. 1169–1178. DOI: `10.1002/hyp`.

De Figueiredo, José Vidal, José Carlos de Araújo, Pedro Henrique Augusto Medeiros, and Alexandre C. Costa (2016). "Runoff initiation in a preserved semiarid Caatinga small watershed, Northeastern Brazil". In: *Hydrological Processes* 30.13, pp. 2390–2400. ISSN: 10991085. DOI: `10.1002/hyp.10801`.

Delgado, Rosario and Xavier-Andoni Tibau (2019). "Why Cohen's Kappa should be avoided as performance measure in classification". In: *PLOS ONE* 14.9. Ed. by Quanquan Gu, e0222916. DOI: `10.1371/journal.pone.0222916`.

De Vente, Joris and Jean Poesen (2005). "Predicting soil erosion and sediment yield at the basin scale: Scale issues and semi-quantitative models". In: *Earth-Science Reviews* 71.1-2, pp. 95–125. ISSN: 00128252. DOI: `10.1016/j.earscirev.2005.02.002`.

Dionizio, Emily Ane and Marcos Heil Costa (2019). "Influence of Land Use and Land Cover on Hydraulic and Physical Soil Properties at the Cerrado Agricultural Frontier". In: *Agriculture* 9.24, p. 14. DOI: `10.3390/agriculture9010024`.

Dos Santos, HG, PK T Jacomine, LHC Dos Anjos, VÁ De Oliveira, JF Lumbreras, MR Coelho, JA De Almeida, JC de Araujo Filho, JB De Oliveira, and TJF Cunha (2018). "Brazilian soil classification system." In: *Embrapa Solos-Livro técnico (INFOTECA-E)*.

Dos Santos, Julio Cesar Neves, Eunice Maia de Andrade, Maria João Simas Guerreiro, Pedro Henrique Augusto Medeiros, Helba Araújo de Queiroz Palácio, and José Ribeiro de Araújo Neto (2016). "Effect of dry spells and soil cracking on runoff generation in a semiarid micro watershed under land use change". In: *Journal of Hydrology* 541, pp. 1–10. DOI: `10.1016/j.jhydrol.2016.08.016`.

Douglas-Mankin, Kyle R., Swapan K. Roy, Aleksey Y. Sheshukov, Asim Biswas, Bahram Gharabaghi, Andrew Binns, Ramesh Rudra, Narayan Kumar Shrestha, and Prasad Daggupati (Dec. 2020). "A comprehensive review of ephemeral gully erosion models". In: *CATENA* 195, p. 104901. DOI: `10.1016/j.catena.2020.104901`.

Easterling, David R., Gerald A. Meehl, Camille Parmesan, Stanley A. Changnon, Thomas R. Karl, and Linda O. Mearns (Sept. 2000). "Climate Extremes: Observations, Modeling, and Impacts". In: *Science* 289.5487, pp. 2068–2074. DOI: `10.1126/science.289.5487.2068`.

FAO (2019). *Global Symposium on soil erosion, 15-17 May 2019, FAO, Rome. Outcome Document*. Food and Agriculture Organization of the United Nations, p. 28.

— (Mar. 2021). *The impact of disasters and crises on agriculture and food security: 2021*. FAO. DOI: `10.4060/cb3673en`.

Fedotov, A.A., P. Harremoes, and F. Topsoe (June 2003). "Refinements of Pinsker's inequality". In: *IEEE Transactions on Information Theory* 49.6, pp. 1491–1498. DOI: `10.1109/tit.2003.811927`.

Fenner, Daniel, Achim Holtmann, Alexander Krug, and Dieter Scherer (Dec. 2018). "Heat waves in Berlin and Potsdam, Germany – Long-term trends and comparison of heat wave

REFERENCES

definitions from 1893 to 2017". In: *International Journal of Climatology* 39.4, pp. 2422–2437. DOI: 10.1002/joc.5962.

Ford, Trent W. and Christopher F. Labosier (Dec. 2017). "Meteorological conditions associated with the onset of flash drought in the Eastern United States". In: *Agricultural and Forest Meteorology* 247, pp. 414–423. DOI: 10.1016/j.agrformet.2017.08.031.

Ford, Trent W., D. Brent McRoberts, Steven M. Quiring, and Ryann E. Hall (Nov. 2015). "On the utility of in situ soil moisture observations for flash drought early warning in Oklahoma, USA". In: *Geophysical Research Letters* 42.22, pp. 9790–9798. DOI: 10.1002/2015gl066600.

Foster, G R and L J Lane (1983). "Erosion by concentrated flow in farm fields". In: *D.B. Simons Symposium on Erosion and Sedimentation*, p. 21.

Fraticelli, C. M. (Sept. 2006). "Climate Forcing in a Wave-Dominated Delta: The Effects of Drought-Flood Cycles on Delta Progradation". In: *Journal of Sedimentary Research* 76.9, pp. 1067–1076. DOI: 10.2110/jsr.2006.097.

FUNCEME (2019). *Foundation of Meteorology and Water Resources of Ceará, Daily total precipitation.* http://www.funceme.br/?page$_i d$ = 2694. Visited on 08/06/2020.

Gaiser, Th., M Krol, H Frischkorn, and J C de Araujo, eds. (2003). *Global Change and Regional Impacts.* Berlin: Springer-Verlag, p. 422. DOI: 10.1007/978-3-642-55659-3_9.

Garcia, Marcelo, ed. (May 2008). *Sedimentation Engineering.* American Society of Civil Engineers. DOI: 10.1061/9780784408148.

Ghorbani-Dashtaki, Shoja, Mehdi Homaee, and Willibald Loiskandl (2016). "Towards using pedotransfer functions for estimating infiltration parameters". In: *Hydrological Sciences Journal* 61.8, pp. 1477–1488. ISSN: 21503435. DOI: 10.1080/02626667.2015.1031763.

Ghoshal, Koeli, Manotosh Kumbhakar, and Vijay P. Singh (Sept. 2018). "Suspended Sediment Concentration and Discharge in Open Channels Using Rényi Entropy". In: *Journal of Hydrologic Engineering* 23.9, p. 04018038. DOI: 10.1061/(asce)he.1943-5584.0001687.

Gordon, L. M., S. J. Bennett, R. L. Bingner, F. D. Theurer, and C. V. Alonso (2007). "Simulating Ephemeral Gully Erosion in AnnAGNPS". In: *Transactions of the ASAE* 50.3, pp. 857–866.

Grissinger, E. H. (1996). "Rill and gullies erosion". In: *Soil erosion, conservation and rehabilitation*, pp. 153–167.

Güntner, Andreas and Axel Bronstert (2004). "Representation of landscape variability and lateral redistribution processes for large-scale hydrological modelling in semi-arid areas". In: *Journal of Hydrology* 297.1-4, pp. 136–161. DOI: 10.1016/j.jhydrol.2004.04.008.

Gupta, Sushindra Kumar, Pushpendra K. Singh, Jaivir Tyagi, Gunwant Sharma, and Ajay Singh Jethoo (June 2020). "Rainstorm-generatedsediment yield model based on soil moisture proxies (SMP)". In: *Hydrological Processes* 34.16, pp. 3448–3463. DOI: 10.1002/hyp.13789.

Haan, C. T., B. J. Barfield, and J. C. Hayes (1994). *Design hydrology and sedimentology for small catchments.* Amsterdam: Academic Press, p. 587. DOI: 10.1016/s0022-1694(96)90037-2.

Hairsine, P. B. and C. W. Rose (1992a). "Modeling water erosion due to overland flow using physical principles: 2. Rill flow". In: *Water Resources Research* 28.1, pp. 245–250. ISSN: 19447973. DOI: 10.1029/91WR02381.

— (1992b). "Modeling water erosion due to overland flow using physical principles: 1. Sheet flow". In: *Water resources research* 28.1, pp. 237–243.

Hampton, Stephanie E, Carly A Strasser, Joshua J Tewksbury, Wendy K Gram, Amber E Budden, Archer L Batcheller, Clifford S Duke, and John H Porter (2013). "Big data and the future of ecology". In: *Frontiers in Ecology and the Environment* 11.3, pp. 156–162.

Han, Yong, Fen li Zheng, and Xi meng Xu (2017). "Effects of rainfall regime and its character indices on soil loss at loessial hillslope with ephemeral gully". In: *Journal of Mountain Science* 14.3, pp. 527–538. ISSN: 16726316. DOI: `10.1007/s11629-016-3934-2`.

Hao, Zengchao, Fanghua Hao, Vijay P Singh, and Xuan Zhang (Dec. 2018). "Changes in the severity of compound drought and hot extremes over global land areas". In: *Environmental Research Letters* 13.12, p. 124022. DOI: `10.1088/1748-9326/aaee96`.

Hayas, Antonio, Tom Vanwalleghem, Ana Laguna, Adolfo Penã, and Juan V. Giráldez (2017). "Reconstructing long-term gully dynamics in Mediterranean agricultural areas". In: *Hydrology and Earth System Sciences* 21.1, pp. 235–249. ISSN: 16077938. DOI: `10.5194/hess-21-235-2017`.

He, Cheng, James Nodwell Taylor, Quintin Rochfort, and David Nguyen (2021). "A new portable in situ flume for measuring critical shear stress on river beds". In: *International Journal of Sediment Research* 36.2, pp. 235–242. ISSN: 1001-6279. DOI: `https://doi.org/10.1016/j.ijsrc.2020.08.004`.

Herold, N., S. M. Downes, M. H. Gross, F. Ji, N. Nishant, I Macadam, N. N. Ridder, and K. Beyer (Jan. 2021). "Projected changes in the frequency of climate extremes over southeast Australia". In: *Environmental Research Communications* 3.1. ISSN: 2515-7620. DOI: `10.1088/2515-7620/abe6b1`.

Hersbach, Hans, Bill Bell, Paul Berrisford, Shoji Hirahara, András Horányi, Joaqun Muñoz-Sabater, Julien Nicolas, Carole Peubey, Raluca Radu, Dinand Schepers, Adrian Simmons, Cornel Soci, Saleh Abdalla, Xavier Abellan, Gianpaolo Balsamo, Peter Bechtold, Gionata Biavati, Jean Bidlot, Massimo Bonavita, Giovanna Chiara, Per Dahlgren, Dick Dee, Michail Diamantakis, Rossana Dragani, Johannes Flemming, Richard Forbes, Manuel Fuentes, Alan Geer, Leo Haimberger, Sean Healy, Robin J. Hogan, Elıas Hólm, Marta Janisková, Sarah Keeley, Patrick Laloyaux, Philippe Lopez, Cristina Lupu, Gabor Radnoti, Patricia Rosnay, Iryna Rozum, Freja Vamborg, Sebastien Villaume, and Jean-Noël Thépaut (June 2020). "The ERA5 global reanalysis". In: *Quarterly Journal of the Royal Meteorological Society* 146.730, pp. 1999–2049. DOI: `10.1002/qj.3803`.

Hobbins, Michael T., Andrew Wood, Daniel J. McEvoy, Justin L. Huntington, Charles Morton, Martha Anderson, and Christopher Hain (June 2016). "The Evaporative Demand Drought Index. Part I: Linking Drought Evolution to Variations in Evaporative Demand". In: *Journal of Hydrometeorology* 17.6, pp. 1745–1761. DOI: `10.1175/jhm-d-15-0121.1`.

Huang, Jianping, Mingxia Ji, Yongkun Xie, Shanshan Wang, Yongli He, and Jinjiang Ran (May 2015). "Global semi-arid climate change over last 60 years". In: *Climate Dynamics* 46.3-4, pp. 1131–1150. DOI: `10.1007/s00382-015-2636-8`.

Hunke, Philip, Eva Nora Mueller, Boris Schröder, and Peter Zeilhofer (2015). "The Brazilian Cerrado: Assessment of water and soil degradation in catchments under intensive agricultural use". In: *Ecohydrology* 8.6, pp. 1154–1180. DOI: `10.1002/eco.1573`.

REFERENCES

Hunt, Eric D., Kenneth G. Hubbard, Donald A. Wilhite, Timothy J. Arkebauer, and Allen L. Dutcher (Apr. 2009). "The development and evaluation of a soil moisture index". In: 29.5, pp. 747–759. DOI: 10.1002/joc.1749.

Ikeda, Syunsuke (1982). "Lateral bed load transport on side slopes". In: *Journal of the Hydraulics Division* 108.11, pp. 1369–1373.

Imeson, A.C. and F.J.P.M. Kwaad (1980). "Gully types and gully prediction". In: *Geografisch Tijdschrift* 14.5, p. 13.

Ireland, H.A., C F S Sharpe, and D H Eargle (1939). *Principles of Gully Erosion in the Piedmont of South Carolina*. Tech. rep. Washington: USDA, p. 185.

James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani (2013). *An introduction to statistical learning*. Vol. 112. Springer.

James, M R, S Robson, S Oleire-Oltmanns, and U Niethammer (2017). "Geomorphology Optimising UAV topographic surveys processed with structure-from-motion : Ground control quality , quantity and bundle adjustment". In: *Geomorphology* 280, pp. 51–66. ISSN: 0169-555X.

Javidan, Kavian, Pourghasemi, Conoscenti, and Jafarian (Nov. 2019). "Gully Erosion Susceptibility Mapping Using Multivariate Adaptive Regression Splines—Replications and Sample Size Scenarios". In: *Water* 11.11, p. 2319. DOI: 10.3390/w11112319.

Jaynes, E. T. (May 1957a). "Information Theory and Statistical Mechanics". In: *Physical Review* 106.4, pp. 620–630. DOI: 10.1103/physrev.106.620.

— (Oct. 1957b). "Information Theory and Statistical Mechanics II". In: *Physical Review* 108.2, pp. 171–190. DOI: 10.1103/physrev.108.171.

Jesus, D. S. (2021). "Erosion in a microbasin under desertification in the Middle Salitre River, Bahia (in Portuguese)". M.Sc. Thesis. Cruz das Almas, Brazil: Universidade Federal do Recôncavo da Bahia – UFRB.

Katz, Harry Alexander, J. Michael Daniels, and Sandra Ryan (2014). "Slope-area thresholds of road-induced gully erosion and consequent hillslope-channel interactions". In: *Earth Surface Processes and Landforms* 39.3, pp. 285–295. ISSN: 01979337. DOI: 10.1002/esp.3443.

Keesstra, Saskia D., Johan Bouma, Jakob Wallinga, Pablo Tittonell, Pete Smith, Artemi Cerdà, Luca Montanarella, John N. Quinton, Yakov Pachepsky, Wim H. van der Putten, Richard D. Bardgett, Simon Moolenaar, Gerben Mol, Boris Jansen, and Louise O. Fresco (Apr. 2016). "The significance of soils and soil science towards realization of the United Nations Sustainable Development Goals". In: *Soil* 2.2, pp. 111–128. DOI: 10.5194/soil-2-111-2016.

Kemter, Matthias, Bruno Merz, Norbert Marwan, Sergiy Vorogushyn, and Günter Blöschl (Apr. 2020). "Joint Trends in Flood Magnitudes and Spatial Extents Across Europe". In: *Geophysical Research Letters* 47.7. DOI: 10.1029/2020gl087464.

Khodashenas, Saeed R. and André Paquier (May 1999). "A geometrical method for computing the distribution of boundary shear stress across irregular straight open channels". In: *Journal of Hydraulic Research* 37.3, pp. 381–388. DOI: 10.1080/00221686.1999.9628254.

Khozani, Zohreh Sheikh and Hossein Bonakdari (Jan. 2018). "Formulating the shear stress distribution in circular open channels based on the Renyi entropy". In: *Physica A: Statistical Mechanics and its Applications* 490, pp. 114–126. DOI: 10.1016/j.physa.2017.08.023.

Klemes, Vit et al. (1982). "Empirical and causal models in hydrology". In: *Scientific basis of water resource management* 95, p. 104.

Knapen, A and J Poesen (2010). "Soil erosion resistance effects on rill and gully initiation points and dimensions". In: *Earth Surface Processes and Landforms* 35.2, pp. 217–228. DOI: 10.1002/esp.1911.

Knight, Donald W, John D Demetriou, and Mohammed E Hamed (1984). "Boundary shear in smooth rectangular channels". In: *Journal of hydraulic engineering* 110.4, pp. 405–422.

Knight, Donald W. and Mark Sterling (Apr. 2000). "Boundary Shear in Circular Pipes Running Partially Full". In: *Journal of Hydraulic Engineering* 126.4, pp. 263–275. DOI: 10.1061/(asce)0733-9429(2000)126:4(263).

Knight, DW, KWH Yuen, and AAI Al Hamid (1994). "Boundary shear stress distributions in open channel flow". In: Beven, K, P Chatwin, and J Millbank. *Physical mechanisms of mixing and transport in the environment.* Wiley. Chap. 4, pp. 51–88.

Kolmogorov, Andrey (1933). "Sulla determinazione empirica di una lgge di distribuzione". In: *Inst. Ital. Attuari, Giorn.* 4, pp. 83–91.

Kondolf, G. Mathias, Yongxuan Gao, George W. Annandale, Gregory L. Morris, Enhui Jiang, Junhua Zhang, Yongtao Cao, Paul Carling, Kaidao Fu, Qingchao Guo, Rollin Hotchkiss, Christophe Peteuil, Tetsuya Sumi, Hsiao-Wen Wang, Zhongmei Wang, Zhilin Wei, Baosheng Wu, Caiping Wu, and Chih Ted Yang (May 2014). "Sustainable sediment management in reservoirs and regulated rivers: Experiences from five continents". In: *Earth's Future* 2.5, pp. 256–280. DOI: 10.1002/2013ef000184.

Krakauer, Nir Y., Tarendra Lakhankar, and Damien Hudson (Sept. 2019). "Trends in Drought over the Northeast United States". In: *Water* 11.9, p. 1834. DOI: 10.3390/w11091834.

Kullback, S. and R. A. Leibler (Mar. 1951). "On Information and Sufficiency". In: *The Annals of Mathematical Statistics* 22.1, pp. 79–86. DOI: 10.1214/aoms/1177729694.

Kullback, Solomon (1978a). *Information Theory and Statistics.* Dover, p. 409.

— (1978b). *Information Theory and Statistics.* New York: Dover, p. 409.

Kumbhakar, Manotosh, Koeli Ghoshal, and Vijay P Singh (2019). "Application of relative entropy theory to streamwise velocity profile in open-channel flow: effect of prior probability distributions". In: *Zeitschrift fur Angewandte Mathematik und Physik* 70.3. ISSN: 00442275. DOI: 10.1007/s00033-019-1124-0.

— (Apr. 2020). "Two-dimensional distribution of streamwise velocity in open channel flow using maximum entropy principle: Incorporation of additional constraints based on conservation laws". In: *Computer Methods in Applied Mechanics and Engineering* 361, p. 112738. DOI: 10.1016/j.cma.2019.112738.

Lal, Rattan (1994). *Soil erosion research methods.* CRC Press.

Landwehr, Tobias (2021). "Is Hydro Power Forever? A Research on the Sustainable Management of Water Dependent Electricity Generation with a Focus on Reservoir Sedimentation". Thesis. Osnabrück Universität, p. 142.

Landwehr, Tobias, Anna Schomberg, and Claudia Pahl-Wostl (2021). "A Holistic and Globally Applicable Indication System for Regional Electric-Energy-Water Security". In: *Ecosystem Health and Sustainability* 0.ja, p. 1901611.

REFERENCES

Lesk, Corey, Pedram Rowhani, and Navin Ramankutty (Jan. 2016). "Influence of extreme weather disasters on global crop production". In: *Nature* 529.7584, pp. 84–87. DOI: 10.1038/nature16467.

Li, Jun, Zhaoli Wang, Xushu Wu, Jie Chen, Shenglian Guo, and Zhenxing Zhang (Nov. 2020). "A new framework for tracking flash drought events in space and time". In: *CATENA* 194, p. 104763. DOI: 10.1016/j.catena.2020.104763.

Li, Zhongwu, Hao Peng, Binggeng Xie, Chun Liu, Xiaodong Nie, Danyang Wang, Mei Huang, Haibing Xiao, Lin Shi, Xuqin Zhang, and Jieyu Jiang (Feb. 2020). "Dissolved organic matter in surface runoff in the Loess Plateau of China: The role of rainfall events and land-use". In: *Hydrological Processes* 34.6, pp. 1446–1459. DOI: 10.1002/hyp.13660.

Liang, Miaoling and Xing Yuan (2021). "Critical role of soil moisture memory in predicting 2012 central USA flash drought". In: *Frontiers in Earth Science* 9, p. 46.

Lindner, Marcus, Michael Maroschek, Sigrid Netherer, Antoine Kremer, Anna Barbati, Jordi Garcia-Gonzalo, Rupert Seidl, Sylvain Delzon, Piermaria Corona, Marja Kolström, et al. (2010). "Climate change impacts, adaptive capacity, and vulnerability of European forest ecosystems". In: *Forest ecology and management* 259.4, pp. 698–709.

Linsley, Kohler (1959). "Applied Hydrology". In: *Mc Grow Hill Book Company.*

Lisonbee, Joel, Molly Woloszyn, and Marina Skumanich (2021). "Making sense of flash drought: Definitions, indicators, and where we go from here". In: *J. Appl. Serv. Climatol* 2021, pp. 1–19. DOI: 10.46275/joasc.2021.02.001.

Liu, X L, C Tang, H Y Ni, and Y Zhao (2016). "Geomorphologic analysis and a physico-dynamic characteristics of Zhatai-Gully debris flows in SW China". In: *Journal of Mountain Science* 13.2, pp. 137–145.

Liu, Yi, Ye Zhu, Liliang Ren, Jason Otkin, Eric D. Hunt, Xiaoli Yang, Fei Yuan, and Shanhu Jiang (Apr. 2020). "Two Different Methods for Flash Drought Identification: Comparison of Their Strengths and Limitations". In: *Journal of Hydrometeorology* 21.4, pp. 691–704. DOI: 10.1175/jhm-d-19-0088.1.

Llena, M., R.J. Batalla, MW. Smith, and D. Vericat (Mar. 2021). "Do badlands (always) control sediment yield? Evidence from a small intermittent catchment". In: *CATENA* 198, p. 105015. DOI: 10.1016/j.catena.2020.105015.

López-Vicente, Manuel and Gema Guzmán (2021). "Chapter 13 - Measuring soil erosion and sediment connectivity at distinct scales". In: *Precipitation*. Ed. by Jesús Rodrigo-Comino. Elsevier, pp. 287–326.

Lukas, Jeff, Mike Hobbins, and Imtiaz Rangwala (2017). *The EDDI User Guide v1.0.* National Integrated Drought Information System (NIDIS).

Luquin, Eduardo, Miguel A. Campo-Bescós, Rafael Muñoz-Carpena, Ronald L. Bingner, Richard M. Cruse, Henrique G. Momm, Robert R. Wells, and Javier Casalí (2021). "Model prediction capacity of ephemeral gully evolution in conservation tillage systems". In: *Earth Surface Processes and Landforms* n/a.n/a. DOI: https://doi.org/10.1002/esp.5134.

Maetens, W., M. Vanmaercke, J. Poesen, B. Jankauskas, G. Jankauskiene, and I. Ionita (2012). "Effects of land use on annual runoff and soil loss in Europe and the Mediterranean: A meta-analysis of plot data". In: *Progress in Physical Geography* 36.5, pp. 599–653. DOI: 10.1177/0309133312451303.

Mahto, Shanti Shwarup and Vimal Mishra (Oct. 2020). "Dominance of summer monsoon flash droughts in India". In: *Environmental Research Letters* 15.10, p. 104061. DOI: `10.1088/1748-9326/abaf1d`.

Mamede, George, Nuno A.M. Araujo, Christian Schneider, Jose Carlos de Araujo, and Hans J. Herrmann (2012). "Overspill Avalanching in a Dense Reservoir Network". In: *PNAS* 109, pp. 1–5. ISSN: 0027-8424. DOI: `10.2139/ssrn.2191634`.

Maner, Sam B. (1958). "Factors affecting sediment delivery rates in the red hills physiographic area". In: *Transactions, American Geophysical Union* 39.4, p. 669. DOI: `10.1029/tr039i004p00669`.

Marengo, Jose A., Lincoln M. Alves, Wagner R. Soares, Daniel A. Rodriguez, Helio Camargo, Marco Paredes Riveros, and Amelia Diaz Pabló (Nov. 2013). "Two Contrasting Severe Seasonal Extremes in Tropical South America in 2012: Flood in Amazonia and Drought in Northeast Brazil". In: *Journal of Climate* 26.22, pp. 9137–9154. DOI: `10.1175/jcli-d-12-00642.1`.

Marengo, Jose A., Sin Chan Chou, Gillian Kay, Lincoln M. Alves, José F. Pesquero, Wagner R. Soares, Daniel C. Santos, André A. Lyra, Gustavo Sueiro, Richard Betts, Diego J. Chagas, Jorge L. Gomes, Josiane F. Bustamante, and Priscila Tavares (Aug. 2011). "Development of regional future climate change scenarios in South America using the Eta CPTEC/HadCM3 climate change projections: climatology and regional analyses for the Amazon, São Francisco and the Paraná River basins". In: *Climate Dynamics* 38.9-10, pp. 1829–1848. DOI: `10.1007/s00382-011-1155-5`.

Martens, Brecht, Diego G. Miralles, Hans Lievens, Robin van der Schalie, Richard A. M. de Jeu, Diego Fernández-Prieto, Hylke E. Beck, Wouter A. Dorigo, and Niko E. C. Verhoest (May 2017). "GLEAM v3: satellite-based land evaporation and root-zone soil moisture". In: *Geoscientific Model Development* 10.5, pp. 1903–1925. DOI: `10.5194/gmd-10-1903-2017`.

Martinez-Villalobos, Cristian and J. David Neelin (Nov. 2019). "Why Do Precipitation Intensities Tend to Follow Gamma Distributions?" In: *Journal of the Atmospheric Sciences* 76.11, pp. 3611–3631. DOI: `10.1175/jas-d-18-0343.1`.

Marzolff, I and J Poesen (2009). "Geomorphology The potential of 3D gully monitoring with GIS using high-resolution aerial photography and a digital photogrammetry system". In: *Geomorphology* 111.1-2, pp. 48–60. ISSN: 0169-555X. DOI: `10.1016/j.geomorph.2008.05.047`.

Matthews, B.W. (Oct. 1975). "Comparison of the predicted and observed secondary structure of T4 phage lysozyme". In: *Biochimica et Biophysica Acta (BBA) - Protein Structure* 405.2, pp. 442–451. DOI: `10.1016/0005-2795(75)90109-9`.

McCool, D. K., L. C. Brown, G. R. Foster, C. K. Mutchler, and L. D. Meyer (1987). "Revised Slope Steepness Factor for the Universal Soil Loss Equation." In: *Transactions of the American Society of Agricultural Engineers* 30.5, pp. 1387–1396. ISSN: 00012351. DOI: `10.13031/2013.30576`.

McKee, Thomas B, Nolan J Doesken, John Kleist, et al. (1993). "The relationship of drought frequency and duration to time scales". In: *Proceedings of the 8th Conference on Applied Climatology.* American Meteorological Society Boston, MA, pp. 179–183.

REFERENCES

Medeiros, Pedro and Murugesu Sivapalan (2020). "From hard-path to soft-path solutions: slow–fast dynamics of human adaptation to droughts in a water scarce environment". In: *Hydrological Sciences Journal* just-accepted.

Medeiros, Pedro Henrique Augusto and José Carlos De Araújo (2014). "Temporal variability of rainfall in a semiarid environment in Brazil and its effect on sediment transport processes". In: *Journal of Soils and Sediments*, pp. 1216–1223. DOI: 10.1007/s11368-013-0809-9.

Mendes, Francisca Juscivânia (2010). "Uma Proposta de Reclassificação das Regiões Pluviometricamente Homogêneas do Estado do Ceará". Dissertação. Universidade Estadual do Ceará, p. 142.

Mesas-Carrascosa, Francisco Javier, María Dolores Notario García, Jose Emilio Meroño De Larriva, and Alfonso García-Ferrer (2016). "An analysis of the influence of flight parameters in the generation of unmanned aerial vehicle (UAV) orthomosaicks to survey archaeological areas". In: *Sensors (Switzerland)* 16.11.

Miles, Lera, Adrian C. Newton, Ruth S DeFries, Corinna Ravilious, Ian May, Simon Blyth, Valerie Kapos, and James E Gordon (Mar. 2006). "A global overview of the conservation status of tropical dry forests". In: *Journal of Biogeography* 33.3, pp. 491–505. DOI: 10.1111/j.1365-2699.2005.01424.x.

Mishra, Ashok K and Vijay P Singh (2010). "A review of drought concepts". In: *Journal of Hydrology* 391.1-2, pp. 202–216. ISSN: 0022-1694. DOI: 10.1016/j.jhydrol.2010.07.012.

— (June 2011). "Drought modeling – A review". In: *Journal of Hydrology* 403.1-2, pp. 157–175. DOI: 10.1016/j.jhydrol.2011.03.049.

Mishra, Surendra Kumar and Vijay P Singh (2003). *Soil Conservation Service Curve Number (SCS-CN) Methodology*. Vol. 42. New York: Springer. DOI: 10.1007/978-94-017-0147-1.

Mo, Kingtse C. and Dennis P. Lettenmaier (2015). "Heat wave flash droughts in decline". In: *Geophysical Research Letters* 42.8, pp. 2823–2829. DOI: https://doi.org/10.1002/2015GL064018.

— (Apr. 2016). "Precipitation Deficit Flash Droughts over the United States". In: *Journal of Hydrometeorology* 17.4, pp. 1169–1184. DOI: 10.1175/jhm-d-15-0158.1.

Mohr, Henning, Scott Draper, David J. White, and Liang Cheng (2021). "The effect of permeability on the erosion threshold of fine-grained sediments". In: *Coastal Engineering* 163, p. 103813. ISSN: 0378-3839. DOI: https://doi.org/10.1016/j.coastaleng.2020.103813.

Montenegro, Suzana and Ragab Ragab (2012). "Impact of possible climate and land use changes in the semi arid regions: A case study from North Eastern Brazil". In: *Journal of Hydrology* 434-435, pp. 55–68. DOI: 10.1016/j.jhydrol.2012.02.036.

Montgomery, D.R. and W.E. Dietrich (1992). "Channel Initiation and Problem of Landscape Scale". In: *Science* 255.February, pp. 826–830.

Montgomery, David R. (2007). *Dirt: The Erosion of Civilizations*. Berkeley: universit y of california press, p. 297.

Moriasi, D. N., J. G. Arnold, M. W. Van Liew, R. L. Bingner, R. Daren Harmel, and T. L. Veith (2007). "Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations". In: *Transactions of the ASABE* 50.3, pp. 885–900.

Morris, Gregory L and Jiahua Fan (1998). *Reservoir sedimentation handbook: design and management of dams, reservoirs, and watersheds for sustainable use.* McGraw Hill Professional.

Mutti, Pedro R., Paulo S. Lúcio, Vincent Dubreuil, and Bergson G. Bezerra (Dec. 2020). "NDVI time series stochastic models for the forecast of vegetation dynamics over desertification hotspots". In: *International Journal of Remote Sensing* 41.7, pp. 2759–2788. DOI: 10.1080/01431161.2019.1697008.

Nachtergaele, J., J. Poesen, A. Sidorchuk, and D. Torri (2002). "Prediction of concentrated flow width in ephemeral gully channels". In: *Hydrological Processes* 16.10, pp. 1935–1953. ISSN: 08856087. DOI: 10.1002/hyp.392.

Nachtergaele, J., J. Poesen, A. Steegen, I. Takken, L. Beuselinck, L. Vandekerchove, and G. Govers (2001). "The value of a physically based model versus an empirical approach in the prediction of ephemeral gully erosion for loess-derived soils". In: *Geomorphology.*

Nearing, MA (1991). "A probabilistic model of soil detachment by shallow turbulent flow". In: *Transactions of the ASAE* 34.1, pp. 81–0085.

Nearing, MA, FF Pruski, and MR O'neal (2004). "Expected climate change impacts on soil erosion rates: a review". In: *Journal of soil and water conservation* 59.1, pp. 43–50.

Ng, Jing Lin, Samsuzana Abd Aziz, Yuk Feng Huang, Majid Mirzaei, Aimrun Wayayok, and M. K. Rowshon (2019). "Uncertainty analysis of rainfall depth duration frequency curves using the bootstrap resampling technique". In: *Journal of Earth System Science* 128.5, pp. 1–15. ISSN: 0973774X. DOI: 10.1007/s12040-019-1154-1.

Nguyen, Hanh, Matthew C Wheeler, Jason A Otkin, Tim Cowan, Andrew Frost, and Roger Stone (June 2019). "Using the evaporative stress index to monitor flash drought in Australia". In: *Environmental Research Letters* 14.6, p. 064016. DOI: 10.1088/1748-9326/ab2103.

Nguyen, Hanh, Matthew C. Wheeler, Harry H. Hendon, Eun-Pa Lim, and Jason A. Otkin (June 2021). "The 2019 flash droughts in subtropical eastern Australia and their association with large-scale climate drivers". In: *Weather and Climate Extremes* 32, p. 100321. DOI: 10.1016/j.wace.2021.100321.

Nkonya, Ephraim, Weston Anderson, Edward Kato, Jawoo Koo, Alisher Mirzabaev, Joachim von Braun, and Stefan Meyer (2016). "Global Cost of Land Degradation". In: *Economics of Land Degradation and Improvement - A Global Assessment for Sustainable Development.* Ed. by Ephraim Nkonya, Alisher Mirzabaev, and Joachim von Braun. Cham: Springer Open. Chap. 6, pp. 117–165. DOI: 10.1007/978-3-319-19168-3.

NOAA (2013). *NCDC/NOAA Land-Based Datasets and Products.* URL: https://www.ncdc.noaa.gov/data-access/land-based-station-data (visited on 04/28/2020).

Noguera, Iván, Fernando Domınguez Castro, and Sergio M Vicente Serrano (2020). "Characteristics and trends of flash droughts in Spain, 1961–2018". In: *Annals of the New York Academy of Sciences.*

Noguera, Iván, Fernando Domınguez-Castro, and Sergio M. Vicente-Serrano (Jan. 2021). "Flash Drought Response to Precipitation and Atmospheric Evaporative Demand in Spain". In: *Atmosphere* 12.2, p. 165. DOI: 10.3390/atmos12020165.

REFERENCES

Nortcliff, Stephen (1987). *Soil Conservation: Assessing the National resources inventory—Volume 2.* Vol. 23. 2, pp. 153–155. ISBN: 0309568102. DOI: `10.1016/0308-521x(87)90092-8`.

Nyssen, Jan, Jean Poesen, Jan Moeyersons, Edith Luyten, Maude Veyret-Picot, Jozef Deckers, Mitiku Haile, and Gerard Govers (2002). "Impact of Road Building on Gully Erosion Risk: a Case of Study from the Northern Ethiopian Highlands". In: *Earth Surface Processes and Landforms* 27, pp. 1267–1283.

Oikonomou, Panagiotis D., Christos A. Karavitis, Demetrios E. Tsesmelis, Elpida Kolokytha, and Rodrigo Maia (Oct. 2020). "Drought Characteristics Assessment in Europe over the Past 50 Years". In: *Water Resources Management* 34.15, pp. 4757–4772. DOI: `10.1007/s11269-020-02688-0`.

Osman, Mahmoud, Benjamin F. Zaitchik, Hamada S. Badr, Jordan I. Christian, Tsegaye Tadesse, Jason A. Otkin, and Martha C. Anderson (Feb. 2021). "Flash drought onset over the contiguous United States: sensitivity of inventories and trends to quantitative definitions". In: *Hydrology and Earth System Sciences* 25.2, pp. 565–581. DOI: `10.5194/hess-25-565-2021`.

Otkin, Jason, Yafang Zhong, Eric Hunt, Jordan Christian, Jeff Basara, Hanh Nguyen, Matthew Wheeler, Trent Ford, Andrew Hoell, and Mark et al. Svoboda (2021). "Development of a flash drought intensity index". In: *Geophysical Research Abstracts.* EGU General Assembly. DOI: `10.5194/egusphere-egu21-1418`.

Otkin, Jason A., Martha C. Anderson, Christopher Hain, Iliana E. Mladenova, Jeffrey B. Basara, and Mark Svoboda (Aug. 2013). "Examining Rapid Onset Drought Development Using the Thermal Infrared–Based Evaporative Stress Index". In: *Journal of Hydrometeorology* 14.4, pp. 1057–1074. DOI: `10.1175/jhm-d-12-0144.1`.

Otkin, Jason A., Mark Svoboda, Eric D. Hunt, Trent W. Ford, Martha C. Anderson, Christopher Hain, and Jeffrey B. Basara (May 2018). "Flash Droughts: A Review and Assessment of the Challenges Imposed by Rapid-Onset Droughts in the United States". In: *Bulletin of the American Meteorological Society* 99.5, pp. 911–919. DOI: `10.1175/bams-d-17-0149.1`.

Otkin, Jason A., Yafang Zhong, Eric D. Hunt, Jeff Basara, Mark Svoboda, Martha C. Anderson, and Christopher Hain (Mar. 2019). "Assessing the Evolution of Soil Moisture and Vegetation Conditions during a Flash Drought–Flash Recovery Sequence over the South-Central United States". In: *Journal of Hydrometeorology* 20.3, pp. 549–562. DOI: `10.1175/jhm-d-18-0171.1`.

Paglia, Eric and Charles Parker (2021). "The intergovernmental panel on climate change: guardian of climate science". In: *Guardians of Public Value.* Palgrave Macmillan, Cham, pp. 295–321.

Palmer, Wayne C (1965). *Meteorological drought.* Vol. 30. US Department of Commerce, Weather Bureau.

Panagos, Panos, Cristiano Ballabio, Katrin Meusburger, Jonathan Spinoni, Christine Alewell, and Pasquale Borrelli (2017). "Towards estimates of future rainfall erosivity in Europe based on REDES and WorldClim datasets". In: *Journal of Hydrology* 548, pp. 251–262. DOI: `10.1016/j.jhydrol.2017.03.006`.

Panagos, Panos, Pasquale Borrelli, and David Robinson (Nov. 2019). "FAO calls for actions to reduce global soil erosion". In: *Mitigation and Adaptation Strategies for Global Change.* DOI: 10.1007/s11027-019-09892-3.

Pastorello, Gilberto, Carlo Trotta, Eleonora Canfora, Housen Chu, Danielle Christianson, You-Wei Cheah, Cristina Poindexter, Jiquan Chen, Abdelrahman Elbashandy, Marty Humphrey, Peter Isaac, Diego Polidori, Alessio Ribeca, Catharine van Ingen, Leiming Zhang, Brian Amiro, Christof Ammann, M. Altaf Arain, Jonas Ardö, Timothy Arkebauer, Stefan K. Arndt, Nicola Arriga, Marc Aubinet, Mika Aurela, Dennis Baldocchi, Alan Barr, Eric Beamesderfer, and Luca Belelli ... Marchesini (July 2020). "The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data". In: *Scientific Data* 7.1, p. 225. ISSN: 2052-4463. DOI: 10.1038/s41597-020-0534-3.

Paton, Eva Nora, Anna Smetanová, Tobias Krueger, and Anthony Parsons (2019). "Perspectives and ambitions of interdisciplinary connectivity researchers". In: *Hydrology and Earth System Sciences* 23.1, pp. 537–548. ISSN: 16077938. DOI: 10.5194/hess-23-537-2019.

Pendergrass, Angeline G., Gerald A. Meehl, Roger Pulwarty, Mike Hobbins, Andrew Hoell, Amir AghaKouchak, Céline J. W. Bonfils, Ailie J. E. Gallant, Martin Hoerling, David Hoffmann, Laurna Kaatz, Flavio Lehner, Dagmar Llewellyn, Philip Mote, Richard B. Neale, Jonathan T. Overpeck, Amanda Sheffield, Kerstin Stahl, Mark Svoboda, Matthew C. Wheeler, Andrew W. Wood, and Connie A. Woodhouse (Mar. 2020). "Flash droughts present a new challenge for subseasonal-to-seasonal prediction". In: *Nature Climate Change* 10.3, pp. 191–199. DOI: 10.1038/s41558-020-0709-0.

Peng, Cui, Ge Yonggang, and Lin Yongming (2011). "Soil Erosion and Sediment Control Effects in the Three Gorges Reservoir Region, China". In: *Journal of Resources and Ecology* 2.4, pp. 289–297.

Peters, Albert J, Elizabeth A Walter-Shea, Lei Ji, Andres Vina, Mlchael Hayes, and Mark D Svoboda (2002). "Drought monitoring with NDVI-based standardized vegetation index". In: *Photogrammetric engineering and remote sensing* 68.1, pp. 71–75.

Pfister, Laurent, Hubert HG Savenije, Fabrizio Fenicia, et al. (2009). *Leonardo Da Vinci's water theory: on the origin and fate of water.* Iahs Press.

Pilgrim, D. H., T. G. Chapman, and D. G. Doran (Aug. 1988). "Problems of rainfall-runoff modelling in arid and semiarid regions". In: *Hydrological Sciences Journal* 33.4, pp. 379–400. DOI: 10.1080/02626668809491261.

Pinheiro, E A R, C A G Costa, and J C De Araújo (2013). "Effective root depth of the Caatinga biome". In: *Arid Environments* 89, p. 4.

Pinheiro, Everton Alves Rodrigues, Klaas Metselaar, Quirijn de Jong van Lier, and José Carlos de Araújo (2016). "Importance of soil-water to the Caatinga biome, Brazil". In: *Ecohydrology* 9.7, pp. 1313–1327. ISSN: 19360592. DOI: 10.1002/eco.1728.

Poesen, J (1993). "Gully typology and gully control measures in the European loess belt". In: *Farm Land Erosion: In Temperate Plains Environment and Hills* June, pp. 221–239. DOI: 10.1016/B978-0-444-81466-1.50024-1.

Poesen, J, L Vandekerckhove, J Nachtergaele, D Oostwoud Wijdenes, G Verstraeten, and B Van Wesemael (2002). "Gully Erosion in Dryland Environments". In: *Dryland Rivers :*

REFERENCES

Hydrology and Geomorphology of Semi-Arid. January. John Wiley & Sons, Ltd. Chap. 8, pp. 229–248.

Poesen, J., J. Nachtergaele, G. Verstraeten, and C. Valentin (2003). "Gully Erosion and Environmental Change: Importance and Research Needs". In: *Catena* 50, pp. 91–133. DOI: 10.1016/S0341-8162(02)00143-1.

Poesen, J. W.A., D. B. Torri, and T. Van Walleghem (2011). "Gully Erosion: Procedures to Adopt When Modelling Soil Erosion in Landscapes Affected by Gullying". In: *Handbook of Erosion Modelling* May 2016, pp. 360–386. DOI: 10.1002/9781444328455.ch19.

Poesen, Jean (Jan. 2018). "Soil erosion in the Anthropocene: Research needs". In: *Earth Surface Processes and Landforms* 43.1, pp. 64–84. ISSN: 0197-9337. DOI: 10.1002/esp.4250.

Poesen, Jean, Tom Vanwalleghem, Joris De Vente, Anke Knapen, Gert Verstraeten, and José A Martínez-Casasnovas (2006). "Gully Erosion in Europe". In: *Soil Erosion in Europe.* John Wiley & Sons Ltd. Chap. 39, pp. 515–536. DOI: 10.1002/0470859202.ch39.

Prandtl, Ludwig (1925). "Bericht über Untersuchungen zur ausgebildeten Turbulenz". In: *ZAMM-Journal of Applied Mathematics and Mechanics/Zeitschrift für Angewandte Mathematik und Mechanik* 5.2, pp. 136–139.

Ran, Hui, Jing Li, Zixiang Zhou, Cheng Zhang, Chengyan Tang, and Yuyang Yu (July 2020). "Predicting the spatiotemporal characteristics of flash droughts with downscaled CMIP5 models in the Jinghe River basin of China". In: *Environmental Science and Pollution Research* 27.32, pp. 40370–40382. DOI: 10.1007/s11356-020-10036-3.

Renard, Kenneth G, George R Foster, Glenn A Weesies, and Jeffrey P Porter (1991). "RUSLE: Revised universal soil loss equation". In: *Journal of soil and Water Conservation* 46.1, pp. 30–33.

Rezende de Souza, Gabriela, Venkatesh Merwade, Luiz Fernando Coutinho de Oliveira, Marcelo Ribeiro Viola, and Matheus de Sá Farias (2021). "Regional flood frequency analysis and uncertainties: Maximum streamflow estimates in ungauged basins in the region of Lavras, MG, Brazil". In: *CATENA* 197, p. 104970. ISSN: 0341-8162. DOI: https://doi.org/10.1016/j.catena.2020.104970.

Risse, L. M., M. A. Nearing, J. M. Laflen, and A. D. Nicks (May 1993). "Error Assessment in the Universal Soil Loss Equation". In: *Soil Science Society of America Journal* 57.3, pp. 825–833. DOI: 10.2136/sssaj1993.03615995005700030032x.

Ritter, Axel and Rafael Muñoz-Carpena (2013). "Performance evaluation of hydrological models : Statistical significance for reducing subjectivity in goodness-of-fit assessments". In: *Journal of Hydrology* 480, pp. 33–45. ISSN: 0022-1694. DOI: 10.1016/j.jhydrol.2012.12.004.

Roehl, JE (1962). "Sediment source areas, and delivery ratios influencing morphological factors". In: *Int. Assoc. Hydro. Sci.* 59, pp. 202–213.

Rogers, R. R. (1989). *A short course in cloud physics.* Oxford, United Kingdom Boston, Massachusetts: Butterworth-Heinemann. ISBN: 9780750632157.

Royall, Dan and Linda Kennedy (Aug. 2016). "Historical erosion and sedimentation in two small watersheds of the southern Blue Ridge Mountains, North Carolina, USA". In: *CATENA* 143, pp. 174–186. DOI: 10.1016/j.catena.2016.03.020.

Samaniego, L., S. Thober, R. Kumar, N. Wanders, O. Rakovec, M. Pan, M. Zink, J. Sheffield, E. F. Wood, and A. Marx (Apr. 2018). "Anthropogenic warming exacerbates European soil

moisture droughts". In: *Nature Climate Change* 8.5, pp. 421–426. DOI: 10.1038/s41558-018-0138-5.

Samaniego, Luis, Rohini Kumar, and Matthias Zink (Feb. 2013). "Implications of Parameter Uncertainty on Soil Moisture Drought Analysis in Germany". In: *Journal of Hydrometeorology* 14.1, pp. 47–68. DOI: 10.1175/jhm-d-12-075.1.

Sanyal, Joy, Alexander L. Densmore, and Patrice Carbonneau (2014). "Analysing the effect of land-use/cover changes at sub-catchment levels on downstream flood peaks: A semi-distributed modelling approach with sparse data". In: *CATENA* 118, pp. 28–40. ISSN: 0341-8162. DOI: https://doi.org/10.1016/j.catena.2014.01.015.

Sartori, Martina, George Philippidis, Emanuele Ferrari, Pasquale Borrelli, Emanuele Lugato, Luca Montanarella, and Panos Panagos (2019). "A linkage between the biophysical and the economic: Assessing the global market impacts of soil erosion". In: *Land Use Policy* 86.5, pp. 299–312. ISSN: 02648377.

Sassolas-Serrayet, Timothée, Rodolphe Cattin, and Matthieu Ferry (2018). "The shape of watersheds". In: *Nature Communications* 9.1, pp. 1–8. ISSN: 20411723.

Schneider, Stephen H., Terry L. Root, and Michael D. Mastrandrea (2011). *Encyclopedia of Climate and Weather*.

Schonberger, Johannes L. and Jan-Michael Frahm (June 2016). "Structure-From-Motion Revisited". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4104–4113.

Sena, Aderita, Christovam Barcellos, Carlos Freitas, and Carlos Corvalan (2014). "Managing the health impacts of drought in Brazil". In: *International Journal of Environmental Research and Public Health* 11.10, pp. 10737–10751. DOI: 10.3390/ijerph111010737.

Serrano, Sergio M. Vicente, Santiago Begueria, and Juan I. López Moreno (Apr. 2010). "A Multiscalar Drought Index Sensitive to Global Warming: The Standardized Precipitation Evapotranspiration Index". In: *Journal of Climate* 23.7, pp. 1696–1718. DOI: 10.1175/2009jcli2909.1.

Shannon, Claude Elwood (1948). "A mathematical theory of communication". In: *Bell system technical journal* 27.3, pp. 379–423. DOI: 10.1002/j.1538-7305.1948.tb01338.x.

Shapiro, S S and M B Wilk (1965). "An Analysis of Variance Test for Normality (Complete Samples)". In: *Biometrika* 52.3/4, p. 591. DOI: 10.2307/2333709.

Sharda, Vishwa N. and Prabhat R. Ojasvi (June 2016). "A revised soil erosion budget for India: role of reservoir sedimentation and land-use protection measures". In: *Earth Surface Processes and Landforms* 41.14, pp. 2007–2023. DOI: 10.1002/esp.3965.

Shrestha, Narayan Kumar, Nabil Allataifeh, Ramesh Rudra, Prasad Daggupati, Pradeep K. Goel, and Trevor Dickinson (Feb. 2019). "Identifying threshold storm events and quantifying potential impacts of climate change on sediment yield in a small upland agricultural watershed of Ontario". In: *Hydrological Processes* 33.6, pp. 920–931. DOI: 10.1002/hyp.13374.

Sidorchuk, A. (1999). "Dynamic and static models of gully erosion". In: *Catena* 37.3-4, pp. 401–414. DOI: 10.1016/S0341-8162(99)00029-6.

Sidorchuk, Aleksey (2005). "Stochastic components in the gully erosion modelling". In: *Catena* 63.2-3, pp. 299–317.

Sidorchuk, Aleksey (2009). "Geomorphology A third generation erosion model : The combination of probabilistic and deterministic components". In: *Geomorphology* 110.1-2, pp. 2–10. ISSN: 0169-555X. DOI: 10.1016/j.geomorph.2008.12.019.

— (2015). "Stochastic Modelling of Soil Erosion and Deposition". In: January 2005.

Sidorchuk, Aleksey, Michael Märker, Sandro Moretti, and Giuliano Rodolfi (2003). "Gully erosion modelling and landscape response in the Mbuluzi River catchment of Swaziland". In: *Catena* 50.2-4, pp. 507–525. DOI: 10.1016/S0341-8162(02)00123-6.

Silva, Augusto C. and Alexandre F. Souza (Apr. 2018). "Aridity drives plant biogeographical sub regions in the Caatinga, the largest tropical dry forest and woodland block in South America". In: *PLOS ONE* 13.4. Ed. by RunGuo Zang, e0196130. DOI: 10.1371/journal.pone.0196130.

Silva, EV, A Gorayeb, and JC de Araújo (2015). "Atlas socioambiental do Assentamento 25 de Maio-Madalena-Ceará". In: *Fortaleza: Ed. Expressão Gráfica.*

Silva Filho, Antonio Viana da, José Carlos de Araújo, and Armin Raabe (2020). "Trade-off between number of constraints and primary-statement robustness in entropy models: the case of the open-channel velocity field". In: *Anais da Academia Brasileira de Ciências* 92.3. DOI: 10.1590/0001-3756202020200594.

Simplıcio, Antonio Alisson Fernandes, Carlos Alexandre Gomes Costa, Joaquın Navarro-Hevia, and José Carlos Araújo (Oct. 2020). "Erosion at hillslope and micro-basin scales in the Gilbués Desertification Region, North-eastern Brazil". In: *Land Degradation & Development.* DOI: 10.1002/ldr.3809.

Simpson, Gareth B and Graham PW Jewitt (2019). "The development of the water-energy-food nexus as a framework for achieving resource security: a review". In: *Frontiers in Environmental Science* 7, p. 8.

Singh, V P and P K Chowdhury (June 1985). "On Fitting Gamma Distrubution to Synthetic Runoff Hydrographs". In: *Hydrology Research* 16.3, pp. 177–192. DOI: 10.2166/nh.1985.0014.

Singh, V. P. (1998). *Entropy-based Parameter Estimation in Hydrology.* Dordrecht: Springer Science, p. 365. DOI: 10.1007/0-306-48065-4.

— (2013). *Entropy theory and its application in environmental and water engineering.* New York: Wiley-Blackwell. ISBN: 9781119976561.

Singh, Vijay P (1995). *Computer models of watershed hydrology.* Water Resources Publications.

— (2014a). *Entropy theory in hydrologic science and engineering.* McGraw Hill Professional.

— (2018). "Hydrologic modeling: progress and future directions". In: *Geoscience Letters* 5.1. ISSN: 21964092. DOI: 10.1186/s40562-018-0113-z.

Singh, Vijay P and Lan Zhang (2018). "Copula – entropy theory for multivariate stochastic modeling in water engineering". In: *Geoscience Letters.* ISSN: 2196-4092. DOI: 10.1186/s40562-018-0105-z.

Singh, Vijay P. (May 2011). "Hydrologic Synthesis Using Entropy Theory: Review". In: *Journal of Hydrologic Engineering* 16.5, pp. 421–433. DOI: 10.1061/(asce)he.1943-5584.0000332.

— (June 2014b). *Entropy Theory in Hydraulic Engineering.* American Society of Civil Engineers. DOI: 10.1061/9780784412725.

Sivakumar, M. V. K., H. P. Das, and O. Brunini (2005). *Impacts of Present and Future Climate Variability and Change on Agriculture and Forestry in the Arid and Semi-Arid Tropics.* Dordrecht: Springer Netherlands, pp. 31–72. DOI: `10.1007/1-4020-4166-7_4`.

Slater, L. J., G. Thirel, S. Harrigan, O. Delaigue, A. Hurley, A. Khouakhi, I. Prosdocimi, C. Vitolo, and K. Smith (2019). "Using R in hydrology: a review of recent developments and future directions". In: *Hydrology and Earth System Sciences* 23.7, pp. 2939–2963.

Smart, Graeme M. (Feb. 1999). "Turbulent Velocity Profiles and Boundary Shear in Gravel Bed Rivers". In: *Journal of Hydraulic Engineering* 125.2, pp. 106–116. DOI: `10.1061/(asce)0733-9429(1999)125:2(106)`.

Smirnov, Nikolai V (1939). "Estimate of deviation between empirical distribution functions in two independent samples". In: *Bulletin Moscow University* 2.2, pp. 3–16.

Snyder, Franklin F (1938). "Synthetic unit-graphs". In: *Eos, Transactions American Geophysical Union* 19.1, pp. 447–454.

Soil Conservation Service (1972). "Soil Conservation Service National Engineering Handbook, section 4, Hydrology". In: *Washington DC.*

Srinivasan, VS and CO Galvão (2003). *Bacia experimental de Sumé: descrição e dados coletados.* Ed. by UFCG. CAmpina Grande: UFCG/CNPq. 129 pp.

Stacy, Edney W (1962). "A generalization of the gamma distribution". In: *The Annals of mathematical statistics* 33.3, pp. 1187–1192.

Starkel, L. (2011). "Paradoxes in the development of gullies". In: *Landform Analysis* 17, pp. 11–13.

Stenseth, Nils C., Atle Mysterud, Geir Ottersen, James W. Hurrell, Kung-Sik Chan, and Mauricio Lima (Aug. 2002). "Ecological Effects of Climate Fluctuations". In: *Science* 297.5585, pp. 1292–1296. DOI: `10.1126/science.1071281`.

Sterling, M. and D. Knight (Apr. 2002). "An attempt at using the entropy approach to predict the transverse distribution of boundary shear stress in open channel flow". In: *Stochastic Environmental Research and Risk Assessment (SERRA)* 16.2, pp. 127–142. DOI: `10.1007/s00477-002-0088-2`.

Stöcker, Claudia, Anette Eltner, and Pierre Karrasch (2015). "Measuring gullies by synergetic application of UAV and close range photogrammetry - A case study from Andalusia, Spain". In: *Catena* 132, pp. 1–11. DOI: `10.1016/j.catena.2015.04.004`.

Storm, D. E., B. J. Barfield, and Lindell E. Ormsbee (1990). *Hydrology And Sedimentology Of Dynamic Rill Networks - Volume I - Erosion Model For Dynamic Rill Networks.* Tech. rep. Lexington: University of Kentucky, p. 130.

Svoboda, Mark D, Brian A Fuchs, et al. (2016). *Handbook of drought indicators and indices.* World Meteorological Organization Geneva, Switzerland.

Thompson, James R. (1964). "Quantitative Effect of Watershed Variables on Rate of Gully-Head Advancement". In: *Transactions of the ASAE* 7.1, pp. 0054–0055.

Tijdeman, Erik and Lucas Menzel (Apr. 2021). "The development and persistence of soil moisture stress during drought across southwestern Germany". In: *Hydrology and Earth System Sciences* 25.4, pp. 2009–2025. DOI: `10.5194/hess-25-2009-2021`.

Tomasella, Javier, Rita M. Silva Pinto Vieira, Alexandre A. Barbosa, Daniel A. Rodriguez, Marcos de Oliveira Santana, and Marcelo F. Sestini (Dec. 2018). "Desertification trends in

REFERENCES

the Northeast of Brazil over the period 2000–2016". In: *International Journal of Applied Earth Observation and Geoinformation* 73, pp. 197–206. DOI: 10.1016/j.jag.2018.06.012.

Tominaga, Akihiro, Iehisa Nezu, Kazuhiro Ezaki, and Hiroji Nakagawa (Jan. 1989). "Three-dimensional turbulent structure in straight open channel flows". In: *Journal of Hydraulic Research* 27.1, pp. 149–173. DOI: 10.1080/00221688909499249.

Torri, Dino and Lorenzo Borselli (2003). "Equation for high-rate gully erosion". In: *Catena* 50, pp. 449–467. ISSN: 03418162. DOI: 10.1016/S0341-8162(02)00126-1.

Torri, Dino and Jean Poesen (2014). "A review of topographic threshold conditions for gully head development in different environments". In: *Earth Science Reviews* 130, pp. 73–85. DOI: 10.1016/j.earscirev.2013.12.006.

UN (2015). "Transforming our world: The 2030 agenda for sustainable development". In.

UN, General Secretariat (1994). "United Nations convention to combat drought and desertification in countries experiencing serious droughts and/or desertification, particularly in Africa". In: *United Nations convention.*

USDA (1987). *Soil Mechanics Level I. Module 3–USDA Textural Soil Classification. Study Guide.*

Valentin, C., J. Poesen, and Yong Li (2005). "Gully erosion: Impacts, factors and control". In: *Catena* 63.2-3, pp. 132–153.

Vandaele, Karel, Jean A W Poesen, Gerard Govers, and Bas van Wesemael (1996). "Geomorphic threshold conditions for ephemeral gully incision". In: *Geomorphology* 16, pp. 161–173.

Vandekerckhove, Liesbeth, Jean Poesen, Dirk Oostwoud Wijdenes, Jeroen Nachtergaele, C Kosmas, MJ Roxo, and T De Figueiredo (2000). "Thresholds for gully initiation and sedimentation in Mediterranean Europe". In: *Earth Surface Processes and Landforms* 25.11, pp. 1201–1220.

Vanmaercke, Matthias, Jean Poesen, Bert Van Mele, Matthias Demuzere, Amber Bruynseels, Valentin Golosov, José Fernando Rodrigues Bezerra, Sergey Bolysov, Aleksandr Dvinskih, Amaury Frankl, Yulia Fuseina, Antônio José Teixeira Guerra, Nigussie Haregeweyn, Ion Ionita, Fils Makanzu Imwangana, Jan Moeyersons, Itshack Moshe, Aliakbar Nazari Samani, Lilian Niacsu, Jan Nyssen, Yoshinori Otsuki, Maria Radoane, Ivan Rysin, Yuri V. Ryzhov, and Oleg Yermolaev (2016). "How fast do gully headcuts retreat?" In: *Earth-Science Reviews* 154, pp. 336–355. ISSN: 00128252. DOI: 10.1016/j.earscirev.2016.01.009.

Vannoppen, W., S. De Baets, J. Keeble, Y. Dong, and J. Poesen (2017). "How do root and soil characteristics affect the erosion-reducing potential of plant species?" In: *Ecological Engineering* 109.August, pp. 186–195. DOI: 10.1016/j.ecoleng.2017.08.001.

Van Schaik, N. L.M.B., A. Bronstert, S. M. de Jong, V. G. Jetten, J. C. van Dam, C. J. Ritsema, and S. Schnabel (2014). "Process-based modelling of a headwater catchment in a semi-arid area: The influence of macropore flow". In: *Hydrological Processes* 28.24, pp. 5805–5816. ISSN: 10991085.

Vanwalleghem, T, J Poesen, J Nachtergaele, J Deckers, and M Van Den Eeckhaut (2005). "Reconstructing rainfall and land-use conditions leading to the development of old gullies". In: *The Holocene* 15.3, pp. 378–386.

Verstraeten, Gert, Paolo Bazzoffi, Adam Lajczak, Mariá Rãdoane, Freddy Rey, Jean Poesen, and Joris De Vente (2006). "Reservoir and Pond Sedimentation in Europe". In: *Soil Erosion in Europe*, pp. 757–774. DOI: `10.1002/0470859202.ch54`.

Vogel, Johannes, Eva Paton, Valentin Aich, and Axel Bronstert (June 2021). "Increasing compound warm spells and droughts in the Mediterranean Basin". In: *Weather and Climate Extremes* 32, p. 100312. DOI: `10.1016/j.wace.2021.100312`.

Vogel, Richard M., Upmanu Lall, Ximing Cai, Balaji Rajagopalan, Peter K. Weiskel, Richard P. Hooper, and Nicholas C. Matalas (June 2015). "Hydrology: The interdisciplinary science of water". In: *Water Resources Research* 51.6, pp. 4409–4430. DOI: `10.1002/2015wr017049`.

Vrugt, Jasper A., Cajo J. F. ter Braak, Martyn P. Clark, James M. Hyman, and Bruce A. Robinson (Dec. 2008). "Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with Markov chain Monte Carlo simulation". In: *Water Resources Research* 44.12. DOI: `10.1029/2007wr006720`.

Wagener, Thorsten, Murugesu Sivapalan, Peter A. Troch, Brian L. McGlynn, Ciaran J. Harman, Hoshin V. Gupta, Praveen Kumar, P. Suresh C. Rao, Nandita B. Basu, and Jennifer S. Wilson (May 2010). "The future of hydrology: An evolving science for a changing world". In: *Water Resources Research* 46.5. DOI: `10.1029/2009wr008906`.

Wang, Hui and Tirusew Asefa (2019). "Chapter 36 - Drought monitoring, mitigation, and adaptation". In: *Extreme Hydrology and Climate Variability*. Ed. by Assefa M. Melesse, Wossenu Abtew, and Gabriel Senay. Elsevier, pp. 457–474.

Wang, Linying and Xing Yuan (Oct. 2018). "Two Types of Flash Drought and Their Connections with Seasonal Drought". In: *Advances in Atmospheric Sciences* 35.12, pp. 1478–1490. DOI: `10.1007/s00376-018-8047-0`.

Wang, Ranghu, Shuwen Zhang, Luoman Pu, Jiuchun Yang, Chaobin Yang, Jing Chen, Cong Guan, Qing Wang, Dan Chen, Bolin Fu, and Xuejia Sang (2016). "Gully Erosion Mapping and Monitoring at Multiple Scales Based on Multi-Source Remote Sensing Data of the Sancha River Catchment, Northeast China". In: *ISPRS International Journal of Geo-Information* 5.11, p. 200. ISSN: 2220-9964. DOI: `10.3390/ijgi5110200`.

Watson, D.A. and J.M. Laflen (1986). "Soil Strength, Slope, and Rainfall Intensity Effects on Interrill Erosion". In: *Transactions of the ASAE* 29.1, pp. 0098–0102.

Wei, Rongqiang, Qingli Zeng, Tim Davies, Guangxiang Yuan, Kaiyang Wang, Xinyu Xue, and Qianfeng Yin (2018). "Geohazard cascade and mechanism of large debris flows in Tianmo gully, SE Tibetan Plateau and implications to hazard monitoring". In: *Engineering Geology* 233, pp. 172–182. ISSN: 00137952.

Wells, Robert R., Henrique G. Momm, James R. Rigby, Sean J. Bennett, Ronald L. Bingner, and Seth M. Dabney (2013). "An empirical investigation of gully widening rates in upland concentrated flows". In: *Catena* 101, pp. 114–121. ISSN: 03418162. DOI: `10.1016/j.catena.2012.10.004`.

Wilhite, DA (2000). *Drought: A Global Assessment*. Vol. 1 and 2. New York: Routledge. ISBN: 9781317854234.

Williams, Jimmy R (1975). "Sediment-yield prediction with universal equation using runoff energy factor". In: *Present and prospective technology for predicting sediment yield and sources* 40, p. 244.

REFERENCES

Williams, Jimmy R and Harold D Berndt (1972). "Sediment yield computed with universal equation". In: *Journal of the Hydraulics Division* 98.Hy 12.

Wischmeier, W. H. and D. D. Smith (1978). *Predicting rainfall erosion losses - a guide to conservation planning.* Tech. rep. Hyattsville: USDA, p. 69.

Woodward, D. E. (1999). "Method to predict cropland ephemeral gully erosion". In: *Catena* 37.3-4, pp. 393–399.

World Meteorological Organization (1986). *Report on Drought and Countries Affected by Drought During 1974–1985.* English. Ed. by WMO. Geneva: WMO. 118 pp.

Worqlul, Abeyou W., Haw Yen, Amy S. Collick, Seifu A. Tilahun, Simon Langan, and Tammo S. Steenhuis (2017). "Evaluation of CFSR, TMPA 3B42 and ground-based rainfall data as input for hydrological models, in data-scarce regions: The upper Blue Nile Basin, Ethiopia". In: *CATENA* 152, pp. 242–251. ISSN: 0341-8162. DOI: https://doi.org/10.1016/j.catena.2017.01.019.

Yang, Shu-Qing and John A. McCorquodale (May 2004). "Determination of Boundary Shear Stress and Reynolds Shear Stress in Smooth Rectangular Channel Flows". In: *Journal of Hydraulic Engineering* 130.5, pp. 458–462. DOI: 10.1061/(asce)0733-9429(2004)130:5(458).

Yao, C, T Lei, W J Elliot, D K Mccool, J Zhao, and S Chen (2008). "Critical Conditions for Rill Initiation". In: *Soil & Water Division of ASABE* 51.1, pp. 107–114.

Yibeltal, Mesenbet, Atsushi Tsunekawa, Nigussie Haregeweyn, Enyew Adgo, Derege Tsegaye Meshesha, Dagnachew Aklog, Tsugiyuki Masunaga, Mitsuru Tsubo, Paolo Billi, Matthias Vanmaercke, Kindiye Ebabu, Mekete Dessie, Dagnenet Sultan, and Mulatu Liyew (2019). "Analysis of long-term gully dynamics in different agro-ecology settings". In: *Catena* 179.April, pp. 160–174. ISSN: 03418162. DOI: 10.1016/j.catena.2019.04.013.

Yuan, Xing, Linying Wang, Peili Wu, Peng Ji, Justin Sheffield, and Miao Zhang (Oct. 2019). "Anthropogenic shift towards higher risk of flash drought over China". In: *Nature Communications* 10.1. DOI: 10.1038/s41467-019-12692-7.

Zarfl, Christiane, Alexander E Lumsdon, Jürgen Berlekamp, Laura Tydecks, and Klement Tockner (2015). "A global boom in hydropower dam construction". In: *Aquatic Sciences* 77.1, pp. 161–170.

Zhang, Haoyue, Chuanhao Wu, and Bill X. Hu (May 2019). "Recent intensification of short-term concurrent hot and dry extremes over the Pearl River basin, China". In: *International Journal of Climatology* 39.13, pp. 4924–4937. DOI: 10.1002/joc.6116.

Zhang, Haoyue, Chuanhao Wu, Pat J.-F. Yeh, and Bill X. Hu (Apr. 2020). "Global pattern of short-term concurrent hot and dry extremes and its relationship to large-scale climate indices". In: *International Journal of Climatology* 40.14, pp. 5906–5924. DOI: 10.1002/joc.6555.

Zhang, Minxia, Shulin Chen, Hong Jiang, Yong Lin, Jinmeng Zhang, Xinzhang Song, and Guomo Zhou (2019). "Water-Use Characteristics and Physiological Response of Moso Bamboo to Flash Droughts". In: *International Journal of Environmental Research and Public Health* 16.12.

Zhang, Shuping, Saskia Foerster, Pedro Medeiros, José Carlos, De Araújo, and Bjoern Waske (2018). "Effective water surface mapping in macrophyte-covered reservoirs in NE Brazil

based on TerraSAR-X time series". In: *Int J Appl Earth Obs Geoinformation* 69.August 2017, pp. 41–55. ISSN: 0303-2434. DOI: 10.1016/j.jag.2018.02.014.

Zhang, Shuping, Saskia Foerster, Pedro Medeiros, José Carlos de Araújo, and Bjoern Waske (2018). "Effective water surface mapping in macrophyte-covered reservoirs in NE Brazil based on TerraSAR-X time series". In: *International Journal of Applied Earth Observation and Geoinformation* 69, pp. 41–55. ISSN: 03032434. DOI: 10.1016/j.jag.2018.02.014.

Zhang, Yuqing, Qinglong You, Changchun Chen, and Xin Li (2017). "Flash droughts in a typical humid and subtropical basin: A case study in the Gan River Basin, China". In: *Journal of Hydrology* 551, pp. 162–176.

Zscheischler, Jakob, Bart van den Hurk, Philip J. Ward, and Seth Westra (2020). "Multivariate extremes and compound events". In: *Climate Extremes and Their Implications for Impact and Risk Assessment*. Elsevier, pp. 59–76. DOI: 10.1016/b978-0-12-814895-2.00004-5.

Zweig, Rachel, Sagi Filin, Yoav Avni, Amir Sagy, and Amit Mushkin (June 2018). "Land degradation and gully development in arid environments deduced by mezzo- and micro-scale 3-D quantification – The Negev Highlands as a case study". In: *Journal of Arid Environments* 153, pp. 52–65. DOI: 10.1016/j.jaridenv.2017.12.006.

Appendix A

In this appendix, we present the supplementary material of chapter 2 -Physically-based model for gully simulation: application to the Brazilian Semiarid Region.

Rain data

Daily rainfall data from the Foundation of Meteorology and Water Resources of Ceará (FUNCEME) were used in the study. The interval of the study was nearly sixty years and to cover the whole period, data from three stations were used:

A. From 1958 to 1973: Station Coroatá (5.03º S; 39.33º W)

B. From 1974 to 1987: Station Boa Viagem (5.12º S; 39.73º W)

C. From 1988 onward: Station Madalena (4.85º S; 39.57º W)

D. Validation: Station Uruquê (5.14º S; 39.18º W)

E. Validation: Station Paus Ferro (4.99º S; 39.48º W)

All stations were equipped with Pluviometer Ville de Paris with a standard opening of 400 cm^2. The final rain series is presented in Figure A.1. To assess the quality of the data series, double mass analysis with other stations in the vicinity was performed, as ilustrated in Figure A.2, from which we can see that there is no relevant bias in the data.
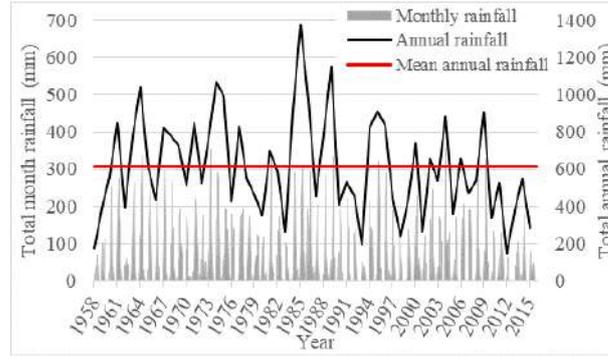
Figure A.1: Rain series with monthly (left vertical axes) and yearly (right vertical axes) rainfall. The red line indicates the average anual rainfall (over 600 mm yr^{-1})
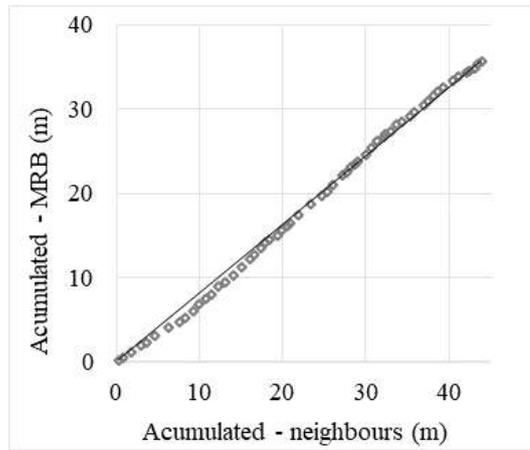
Figure A.2: Double mass diagram. Neighbour stations in a radius of 50 km.

Gully models from literature

Foster and Lane Model (FLM)

The Foster and Lane (1983) ephemeral-gully model aims at explaining "erosion by concentrated flow in farm fields" for single runoff intensive events. The gullies are considered ephemeral as productive farmlands usually provide periodic tillage to diminish or remove the gullies generated by previous events. The model is physically based and uses the Manning equation, mass balance, and shear stress mobilisation; it assumes an equilibrium channel width and the gully evolution in two steps. The first step is the vertical incision when the concentrated overflow starts digging the channel with a constant width. The second step starts after the bottom of the channel reaches a non-erodible layer. Then, the section starts a sideward erosive process, widening until the end of the effective runoff, i.e., with a shear stress below the critical stress. Detachment ratio (D_r) and shear stress (τ) are given by the Equations (A.1) and (A.2).

$$D_r = K_r \left(\tau - \tau_c \right) \tag{A.1}$$

$$\tau(X) = 1.35\gamma R_h S \left[1 - \left(1 - 2\frac{X}{WP}\right)^{2.9}\right] \tag{A.2}$$

In Equation A.2, X is the position of a point on the channel bed, varying from zero to WP (wet perimeter). S is the longitudinal slope of the channel; r_h the hydraulic radius; and γ is the specific gravity of water (assumed 9.81 kN m^{-3}).

In order to model long-term gullies using Foster and Lane (1983) equations, the following assumptions were made: First, all mobilised sediment is carried away by the discharges, i.e., there is no sediment deposition on the channel bed. This assumption was confirmed by field surveys in many sections where very little loose sediment was identified. Secondly, in the long run, the effect of each intense runoff can be piled in a cumulative model of widths/depths layers. This implies that each erosive event does not suffer significant influence of the previous, and the total eroded soil is related only with the energy of each event. The piling process considered all events with runoff. The following figures display the flow charts for the original Foster and Lane Model and how it allows to model multiple events by piling area available in the supplementary material (Fig. A.3 and A.4).
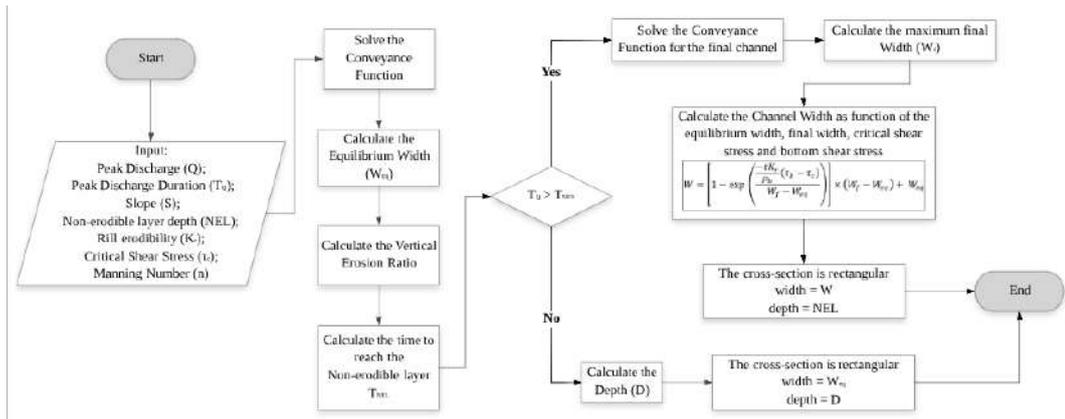
Figure A.3: Flowchart of the original Foster and Lane Model (1983)
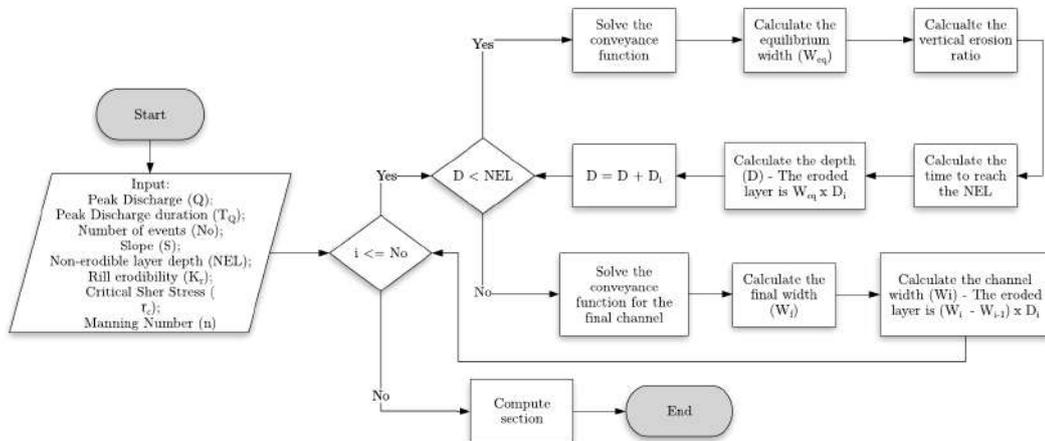
Figure A.4: Flowchart of the modified Foster and Lane Model, in order to allow multiple successive events.

Sidorchuk Model (SM)

The Sidorchuk model (Sidorchuk, 1999) is both physically and empirically based. It considers mass balance of sediment, shear stress (in terms of critical velocity), soil cohesion and the Manning equation to estimate the cross-section geometry and channel slope. It also uses empirical equations based on field measurement to estimate the flow depth and width. The model gives special attention to the processes involving gully wall transformation, as shown in Equations (A.3) and (A.4).

$$D_{vcr} = \frac{2C_h}{g\ \rho_s}\cos(\varphi)\sin^{-2}\left[\frac{1}{2}\left(\varphi + \frac{\pi}{2}\right)\right] \tag{A.3}$$

$$\frac{C_h}{g\ \rho_s\ D_v} = \frac{\rho_s - w\rho}{\rho}\tan\varphi\cos^2\phi - \frac{\sin 2\phi}{2} \tag{A.4}$$

In Equations (A.3) and (A.4), C_h is soil cohesion; D_v the depth incision; D_{vcr} the critical value of depth for wall failure; w the volumetric soil water content; ρ_s the bulk density of the soil; ρ the density of water; g gravity's acceleration; φ the soil internal friction angle; and ϕ the wall slope, in degrees. A flow chart of the model is available in the supplementary material (Fig. A.5).
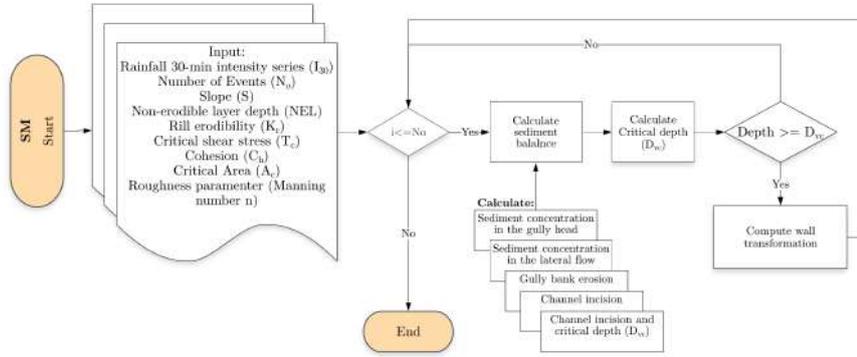
Figure A.5: Flowchart of the Sidorchuk Model (1999).

Model evaluators

To allow us to assess a more realistic, concerning uncertainty, evaluation of the model's quality, we implemented the routine proposed by Ritter and Muñoz-Carpena (2013). The routine, called FITEVAL, allows the modeller to obtain a confidence interval for the Nash-Sutcliffe Efficiency (NSE). The common values of NSE is, therefore, interpreted as the centre of this interval and of the probabilistic distribution of its values depending on how the data are resampled. The method classifies the model as Acceptable to Very good – NSE $\in$ [0.66, 0.95]; (p-value = 0.05). A conservative interpretation of these results is to understand the lowest values as the minimum state of information, i.e., the one that contains (almost) no unproved

hypothesis and does not rely on luck. Therefore, the model can be classified, in the worst case scenario, as Acceptable. The output of the FITEVAL is presented in (Figure A.6). On the left we see a scatter of points Computed (modelled) and Observed (measured) points, corresponding to the cross-section areas. Below, we have a list of the percentiles of each category. On the right we have the cumulative probabilistic distribution of values of efficiency and below that, a plot showing the observed (diamonds) and computed (continuous line) in decreasing order.
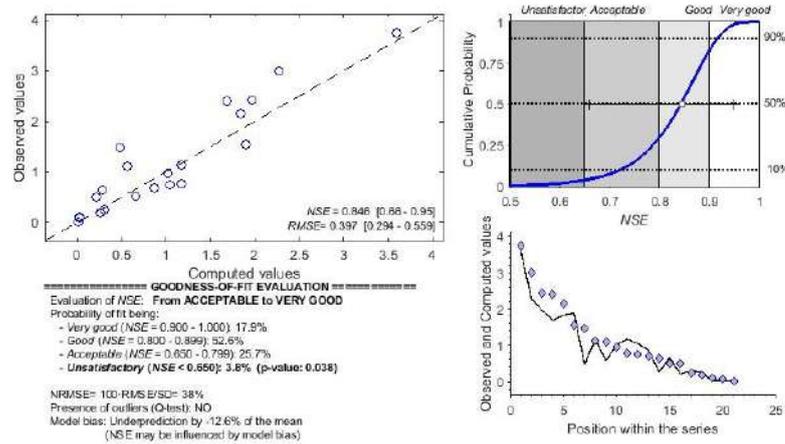
Figure A.6: Output of FITEVAL (Ritter and Muñoz-Carpena, 2013), indicating the model proposed as, in the worst case scenario, acceptable. On the left we have a scatter plot of modelled and measured data, on the right the cumulative probability distribution of NSE (top) and the values of measured (black line) and modelled (diamond) in decreasing order (bottom).

General data and code

Following we present basic data and results used in the model and the code used. For more, please access the following link for the permanent repository:

github.com/PedroAlencarTUB/GullyModel-FLSM

Gully	Volume (Total Station) (m³)	Volume (UAV) (m³)	Error volume (%)	Length (m)	Max. Width (m)	Max. Depth (m)	Channel area (m²)	Coordenates UTM (zone: 23S) X (m)	Y (m)
1	20	22.5	11%	45.5	3.7	0.709	150	445297	9449332
2	38	33.5	13%	33.1	5.8	0.694	320	444614	9447240
3	69	-	-	37.4	8.4	0.844	480	444883	9447479

Table A.1: Comparison of gullies dimensions.

Land-use description	Condition	Soil Class A	B	C	D
Cultuivated land	with conservation treatment	62	71	78	81
	without conservation treatment	72	81	88	91
Pasture	good	39	61	74	80
	bad	68	79	(86)	89

Table A.2: CN values by land-use (Chow, Maidment, and Mays, 1988). In parentesis the CN adopted in this study.

Gully	Section	Basin Area (m²)	Perimeter (m)	Length (m)	Kc	Kf	Conc. Time (s)	Measured Area (m²)	Perimeter (m)	Width (m)	Prof
Gully 1	S1	3200	432.5	195	2.141	0.084	25	0.248	2.322	2.264	0.215
	S2	3080	415.5	187.8	2.096	0.087	23	0.742	3.451	3.188	0.582
	S3	1183	337.4	179.3	2.747	0.037	22	0.684	3.734	3.558	0.498
	S4	491	107.7	35.9	1.361	0.381	6	0.505	3.826	3.763	0.288
	S5	78	49.8	22.9	1.579	0.149	5	0.084	1.520	1.498	0.103
	S6	3060	408	184.5	2.065	0.090	23	0.770	2.707	2.180	0.709
Gully 2	S1	2000	383	123	2.398	0.132	31	0.721	2.235	1.853	0.479
	S2	1870	379	114.3	2.454	0.143	29	2.428	7.559	7.322	0.694
	S3	1470	336	99.7	2.454	0.148	28	1.105	6.950	6.891	0.307
	S4	252	79.9	23.5	1.409	0.456	7	0.644	3.085	2.951	0.356
	S5	40	43.7	12.4	1.935	0.260	4	0.008	0.488	0.479	0.033
	S6	1930	381	119	2.428	0.136	30	2.403	7.633	7.438	0.653
	S7	1425	320	92.2	2.374	0.168	25	1.470	6.263	6.163	0.451
Gully 3	S1	3500	554	204	2.622	0.084	40	1.126	6.459	6.321	0.521
	S2	3400	540	193	2.593	0.091	39	0.962	5.596	5.543	0.306
	S3	1000	248	105	2.196	0.091	25	0.200	2.239	2.181	0.198
	S4	56	49.1	13.9	1.837	0.290	4	0.103	1.351	1.303	0.117
	S5	2230	473	185.7	2.805	0.065	38	0.511	3.098	3.016	0.306
	S6	4800	620	213	2.506	0.106	43	3.751	6.485	6.057	0.897
	S7	4964	639	218	2.539	0.104	45	2.151	5.744	4.679	0.785
	S8	4712	609	205	2.484	0.112	41	1.543	5.935	5.327	0.844

Table A.3: Paper data compilation part 1 - Topography results

Gully	FLM-Iav		FLM-I60		FLM-I30		FLM-I15		FLM-	SM	FL-SM
	Area (m²)	Width (m)	Area (m²)	Width (m)	Area (m²)	Width (m)	Area (m²)	Width (m)	Area (m²)	Area (m²)	Area (m²)
Gully 1	0.071	0.299	0.208	0.970	0.301	1.193	0.288	1.309	0.306	0.380	0.301
	0.283	0.476	0.862	1.423	1.047	1.720	1.106	1.877	1.094	1.685	1.047
	0.165	0.331	0.500	0.992	0.871	1.215	0.866	1.325	0.494	1.118	0.871
	0.050	0.173	0.172	0.542	0.216	0.663	0.210	0.726	0.641	0.388	0.216
	0.006	0.061	0.023	0.222	0.028	0.274	0.039	0.298	0.043	0.082	0.028
	-	-	-	-	1.170	1.669	-	-	1.248	2.173	1.170
Gully 2	0.207	0.431	0.624	1.303	1.359	1.592	1.142	1.738	2.324	2.276	2.276
	0.278	0.401	0.855	1.199	1.369	1.467	1.474	1.556	2.029	1.968	1.968
	0.100	0.326	0.315	0.984	0.567	1.206	0.505	1.318	0.860	0.590	0.567
	0.049	0.137	0.153	0.431	0.279	0.532	0.211	0.581	0.940	0.473	0.279
	0.002	0.060	0.005	0.154	0.006	0.187	0.007	0.203	0.014	0.006	0.006
	-	-	-	-	1.202	1.287	-	-	1.721	1.690	1.690
	-	-	-	-	0.477	1.081	-	-	0.791	0.917	0.477
Gully 3	0.328	0.641	0.961	1.877	1.171	2.288	2.003	2.493	1.192	1.752	1.171
	0.160	0.522	0.478	1.562	1.021	1.912	0.914	2.086	1.101	0.803	1.021
	0.048	0.242	0.152	0.740	0.254	0.909	0.234	0.991	0.273	0.289	0.254
	0.008	0.067	0.022	0.189	0.028	0.239	0.031	0.268	0.068	0.086	0.028
	0.120	0.392	0.379	1.183	0.660	1.450	0.582	1.584	0.558	0.661	0.660
	-	-	-	-	2.193	2.275	-	-	2.626	3.596	3.596
	-	-	-	-	1.832	2.318	-	-	2.173	3.025	1.832
	-	-	-	-	1.891	2.250	-	-	2.359	3.279	1.891

Table A.4: Paper data compilation part 2 - Model results

Fortran Code

```fortran
1  program FLSM

   IMPLICIT NONE
   CHARACTER arquivo*30,arquivo1*30,arquivo2*30
   integer No,i
6  REAL Q(1500),n,S,Kr,Ds,Pa,Tc,Xnc,Gc,Rn,WP,Rh,Ta,Wn,Weq,Le,Delta,Param,Tmax,tm(1500),t(1500),tn
   REAL NEL,Tb,Gcf,Xncf,Xnc1,Xnc2,Gc1,Gc2,Rn1,Rn2,WP1,Lim,Ch,Phi,Ang,Dvcr,g,pi,es,Ws
   REAL WP2,Rh1,Rh2,Ta1,Ta2,Dg1,Dg2,Tnc1,Tnc2,Xncf1,Xncf2,gcf1,gcf2,tnf1,tnf2,dgf1,dgf2
   REAL wi(1500),wf,dp(1500),depth,DeltaW,PerimF,AreaI,AreaF,Rh_S,Max_wi,Max_W,dW
   COMMON /grand1/Gcf,Xncf1,Xncf2
11 COMMON /grand2/Ch,g,Ds,es,Pa,Phi,depth,Ang,pi
   !Phi = internal friction angle of the soil (degrees) > the data can be obtained in the NAVFAC 7.02,
   !based in texture; in Ortiz et al 1989 based in texture, Plasticity intex and porosity or
   !via laboratory experiments
   !Ch = Cohesion (Pa) > the data can be obtained in the NAVFAC 7.02, based in texture;
16 !in Ortiz et al 1989 based in texture, Plasticity intex and porosity or via laboratory experiments
   !Lim = a threshold for aplication of sidorchuk routine
   !Q = peak discharge (m3/s)
   !tm  = discharge's duration (min)
   !n = Manning's number
21 !S = Slope (abs) − (not in degrees or percentage)
   !Kr = rill erodibility factor (s/m)
   !Tc = Critical shear stress (Pa)
   !Xnc = Normalized distance in the WP where T=Tc (abs)
   !Gc = Conveyance function at Xnc (abs)
26 !Rn = Normalized Hydraulic Radius (abs)
   !WP = Wet Perimeter (m)
   !Rh = Hydraulic Radius (m)
   !Ta = Average shear stress (Pa)
   !Wn = Normalized width (abs)
31 !Weq = equilibrium width (m)
   !Er = Erosion rate (kg/m.s)
   !Ve = Velocity of movement down (m/s)
   !Pse = Weigth of soil eroded in the event (kg)
   !Le = Sheet thickness of eroded soil(m)
36 !Param = A parameter that is repeated throughout the calculations
   !t = time (s)
   !NEL = Depth of nonerodible layer(m)
   !Tb = Tension when the erosion reach the NEL (Pa)
   !Gcf = conveyance funcion final (abs)
41 !Wf = final width of the channel (m)
   !tne = time to reach  the NEL (s)
           WRITE(*,*)'****************************************'
           WRITE(*,*)'          Foster_Lane program          '
           WRITE(*,*)'                                        '
46         WRITE(*,*)'    Calculates the final total of soil  '
           WRITE(*,*)'      eroded in an ephemeral gully      '
           WRITE(*,*)'       for a time−serie rainfall        '
           WRITE(*,*)'                                        '
           WRITE(*,*)'      Universidade Federal do Ceara     '
51         WRITE(*,*)'    Departamento de Engenharia Agricola  '
           WRITE(*,*)'      Doutorado em Engenharia Agricola      '
           WRITE(*,*)'                                        '
           WRITE(*,*)'          Pedro Alencar, 2019           '
           WRITE(*,*)'                                        '
56         WRITE(*,*)'****************************************'

   !1. READS INPUT DATA
   write(*,'(a)')'Insert the name of the file containing the runoff data (Discharge e Duration):
   ' !in the absense of measured data we sugest use the 30−min intensity
61 Read(*,'(a30)')arquivo1
   write(*,'(a)')'Insert the name of the file containing the hillslope and soil data: '
   Read(*,'(a30)')arquivo2
   WRITE(*,'(a)',advance='no') 'Insert the name of the output file: '
   READ(*,'(a30)')arquivo
66 open(50,file=arquivo1)
       do i=1, No
           read(50,*) Q(i),tm(i)
           write(*,*) Q(i),tm(i)
           t(i)=tm(i)*60.
71     end do
   close(50)
   open(60,file=arquivo2)
       read(60,*)n,S,Tc,kr,Ds,es,NEL,Ch,Phi,Lim
       write(*,*)'The number of Manning of the channel is....',n
76     write(*,*)'The declivity of the hillslope is..........',S
       write(*,*)'The critical shear stress is...............',Tc
       write(*,*)'The rill erodibility coefficient is........',Kr
       write(*,*)'The soil Bulk density is...................',Ds
       write(*,*)'The porosity of the soil is................',es
81     write(*,*)'The depth of the nonerodible layer is......',NEL
```

```fortran
      write(*,*)'The soil cohesion is .......................',Ch
      write(*,*)'The internal friction angle is .............',Phi
      write(*,*)'The threshold for wall erosion is ..........',Lim
    close(60)
 86 Pa = 9803. !N/m
    g = 9.803 !gravity
    i = 1
    depth = 0
    AreaI = 0
 91 Max_wi = 0
    Max_W  = 0
    pi = 3.1415926536
    Phi = Phi*pi/180.
    Ws = 0.
 96 i=1

    ! 2. FOSTER AND LANE EQUATIONS FOR INCISION BEFORE REACHING THE NEL
    open(40,file=arquivo)
    do while(i.le.No)
101      Param = (n*Q(i)/(S**(0.5)))**0.375

         if (depth.lt.NEL) then
             Gc = Pa*S*Param/Tc
             if (Gc.lt.1.79) then
106              i=i+1
                 else
                     Delta = sqrt(3.9429**2 - 4*6.9594*(1/Gc))
                     Xnc1 = (3.9429-Delta)/(2*6.9594)
                     Xnc2 = (3.9429+Delta)/(2*6.9594)
111                  Rn1 = -0.8834*Xnc1+0.1395*Xnc1+0.151
                     Rn2 = -0.8834*Xnc2+0.1395*Xnc2+0.151
             end if
             if (Rn1.gt.0) then
                 WP1 = Param/Rn1**(0.625)
116              Rh1 = Rn1*WP1
                 Ta1 = Pa*Rh1*S
                 Tnc1 = Tc/Ta1
                 Gc1 = 1./(Tnc1*(Rn1**0.375))
                 Dg1 = abs(Gc-Gc1)
121              else
                     Gc1 = 0
                     Dg1 = abs(Gc-Gc1)
             end if
             if (Rn2.gt.0) then
126              WP2 = Param/Rn2**(0.625)
                 Rh2 = Rn2*WP2
                 Ta2 = Pa*Rh2*S
                 Tnc2 = Tc/Ta2
                 Gc2 = 1/(Tnc2*(Rn2**(0.375)))
131              Dg2 = abs(Gc-Gc2)
                 else
                     Gc2 = 0
                     Dg2 = abs(Gc-Gc2)
             end if
136          if (Dg1.gt.Dg2) then
                 Xnc = Xnc2
                 else
                     Xnc = Xnc1
             end if
141          Write(*,*)'Xnc ',Xnc
             Wn = -1.4873*Xnc+0.7436
             Rn = -0.8834*Xnc**2+0.1395*Xnc+0.151
             WP = Param/Rn**(0.625)
             Weq = WP*Wn
146          Rh = Rn*WP
             Ta = Pa*Rh*S
             Tmax = 1.35*Ta
             if (Tmax.lt.Tc) then
                 Write(*,*)'The event didn t cause erosion.'
151              wi(i) = 0.
                 dp(i) = 0.
                 else
                     wi(i) = Weq
                     Le = Kr*(Tmax-Tc)*t(i)/Ds
156                  if (Le.gt.NEL) then
                         dp(i) = NEL
                         else
                             dp(i) = Le
                     end if
161                  depth = depth + dp(i)
                     if (Max_wi.lt.Weq) then
                         Max_wi = wi(i)
                         Max_W  = wi(i)
                     end if
```

A. Appendix A

```fortran
166                    AreaI = AreaI+wi(i)*dp(i)
                    Write(40,*)'Fase1',i,wi(i),dp(i),Max_wi,AreaI
                end if
        end if

!3. FOSTER AND LANE EQUATIONS FOR INCISION AFTER REACHING THE NEL
        if (depth.ge.NEL) then
            Gcf = Pa*S*Param/Tc
            if (Gcf.lt.1.78) then
                i=i+1
            else
                    CALL Newton_NEL
            end if
            tnf1 = 1.35*(1-(1-2.*xncf1)**2.9)
            gcf1 = 1/(tnf1*(xncf1*(1-2*xncf1))**(0.375))
            dgf1 = abs(gcf-gcf1)
            tnf2 = 1.35*(1-(1-2.*xncf2)**2.9)
            gcf2 = 1/(tnf2*(xncf2*(1-2*xncf2))**(0.375))
            dgf2 = abs(gcf-gcf2)
                if (dgf1.gt.dgf2) then
                    xncf = xncf2
                    else
                        xncf = xncf1
                    end if
            Tb = 1.35*Ta*(1-(1-2*xncf)**2.9)*Pa*S*Param*(xncf*(1-2*xncf))**0.375
            if (Tb.le.Tc)then
                dW = 0
                else
                dW = Kr*(Tb-Tc)/Ds
            end if

            wf = Param*((1-2.*xncf)/(xncf**1.667))**0.375
            if (Max_W.ge.Wf) then
                tn = 0.
                else
                    tn = t(i)*dW/(Wf-Max_W)
            end if
            Max_W = (1-exp(-tn))*(Wf-Max_W)+Max_W
            write(40,*)'Fase2',i, Max_W, NEL, dw
        end if
        i = i+1
    end do
    DeltaW = Max_W-Max_Wi
    AreaF = AreaI + DeltaW*depth
    PerimF = Max_W + 2.*depth
    Rh_S = AreaF/PerimF

    !4. WALL TRANSFORMATION
    if (AreaF.gt.Lim) then
        Dvcr = (2.*Ch*cos(Phi)/(g*Ds))/(sin(0.5*(Phi+pi/2.)))**2.
        if (Dvcr.le.NEL) then
            CALL Newton_PHI
            Ws = Max_W +2.*NEL/tan(Ang)
            AreaF = NEL*(Max_W+Ws)/2.
        end if
    end if
    write(40,*)'The cross section area is: ',AreaF,' m2'
    if (depth.lt.NEL) then
        write(40,*)'The depth is: ',depth,'m'
        write(40,*)' '
        write(40,*)'The erosion didn t reach the non erodible layer.'
        else
            write(40,*)'The depth is: ',NEL,'m'
            write(40,*)' '
            write(40,*)'The erosion reached the non erodible layer.'
        end if
    if (Ws.gt.0.) then
        write(40,*)'There is erosion of the walls.'
        write(40,*)'The top width is',Ws,'m and the wall slope is ',100.*tan(Ang),'%'
        write(40,*)Ang*180/pi
    end if
    write(*,*)' '
    write(*,*)'Please, see the output file'
    end program

    Subroutine Newton_NEL
    !Subrotina para calcular as raizes da equa  o de Xncf
    implicit none
    integer i,j
    real x,xa,xb,f,df,erro,gcf,xncf1,xncf2
    COMMON /grand1/Gcf,Xncf1,Xncf2
    Xa = 0.1
    Xb = 0.4
    i=0
```

156

```fortran
      j=0
251   erro=1000000.
      X=Xa
      do while (erro.gt.0.000001)
          f = -321.29*x**6+249.02*x**5+9.059*x**4-58.033*x**3+13.012*x**2+1.7199*x-0.004883*COS(x) -1/Gcf
          df = -321.29*6*x**5+249.02*5*x**4+9.059*4*x**3-58.033*3*x**2+13.012*2*x+1.7199+0.004883*sin(x)
256
          x = xa - f/df
          erro = abs(x-xa)/xa
          xa=x
          i=i+1
261   end do
      erro=1000000.
      x=Xb
      do while (erro.gt.0.000001)
          f = -321.29*x**6+249.02*x**5+9.059*x**4-58.033*x**3+13.012*x**2+1.7199*x-0.004883*COS(x) -1/Gcf
266       df = -321.29*6*x**5+249.02*5*x**4+9.059*4*x**3-58.033*3*x**2+13.012*2*x+1.7199+0.004883*sin(x)

          x = xb - f/df
          erro = abs(x-xb)/xb
          xb=x
271       j=j+1
      end do
      Xncf1 = xa
      Xncf2 = xb
      end subroutine
276
      Subroutine Newton_PHI
      !Subroutine to calculate wall stable angle
          implicit none

281       integer i
          real k1,k2,f1,df1,x,xa,erro,Ch,g,Ds,es,Pa,Phi,depth,Ang,pi
          COMMON /grand2/Ch,g,Ds,es,Pa,Phi,depth,Ang,pi

          i=0
286       k1 = Ch/(g*Ds*depth)
          k2 = tan(Phi)*(Ds - es*1000.)/1000.

          erro=1000000.
          x = pi/4.
291           do while (erro.gt.0.00000001)

              f1 = k2*(cos(x))**2 - sin(2*x)/2. - k1
              df1 = -2*k2*cos(x) - cos(2*x)

296           xa = x - f1/df1
              erro = abs(x-xa)
              x=xa
              i=i+1
          end do
301   Ang = x
      end subroutine
```

157

Appendix B

In this appendix, we present the supplementary material of chapter 3 - Entropy-based model for gully erosion – a combination of probabilistic and deterministic components

Soil data

Table B.1: Soil properties

Area	Sections	Gravel	FCS[a]	VFS[b]	Silt	Clay	Organic Matter	Bulk Density	Soil Texture[c]	K_r	τ_c
Madalena	1	6 %	46 %	16 %	14 %	18 %	3.3 %	1677	Sandy Loam	0.017	2.30
	2, 3	8 %	29 %	6 %	20 %	37 %	5.7 %	1572	Clay Loam	0.012	3.50
Gilbués	1, 2	7 %	38 %	11 %	17 %	28 %	4.5 %	1290	Silt Loam	0.013	3.50
	3	8 %	33 %	9 %	19 %	32 %	5.1 %			0.009	3.50
Campo Formoso	3	22 %	24 %	2 %	47 %	4 %	0.7 %	1750	Sandy Loam	0.011	2.74
	1	26 %	26 %	1 %	43 %	4 %	0.6 %			0.012	2.75
	2	14 %	18 %	1 %	64 %	4 %	0.6 %			0.013	2.75

[a] Fine to Coarse Sand;
[b] Very Fine Sand;
[c] Soil texture classification and grain size distribution following the USDA textural classification manual (USDA, 1987).

Photos from study areas

Madalena: below we present some photos from the studied sites in Madalena (Ceará). Note the fine sediment deposited in the channel bed and the geometry of the channel, with trapezoidal shape with wide surface and angles.

Figure B.1: Gullies in Madalena. In the figures, it is possible to observe the dry conditions of the region and vegetation. It is also possible to identify the fine sediment deposited in the bottom of the channel, in contrast to the coarse bedrock natural to the region's deeper soils.

Gilbués: here we present some gullies in the study area in Gilbués (Piaui). The region, in advanced desertification process, presents a fine and deep soil. There is also intense processes of pipping and rilling. Gullies cross-section areas range from a few to dozens of square meters.

Figure B.2: Gullies in Gilbués. The region, that has its name from the native language for *week land*, presents advanced desertification and land degradation.

Campo Formoso: the gullies in Campo Formoso (Bahia) are mainly caused by deforestation for the installation of Agave farms in the area. The gullies have intermediate cross-section area and depth, reaching up to 3 meters depth and areas of up to a couple dozen square meters. The shape of the cross-sections is usually rectangular. Again, it is possible to observe fine to coarse sand deposited on the channel bed, a phenomenon that may lead to shielding.

Figure B.3: Gullies in Campo Formoso. The region is under desertification process, mainly caused by bad land use management, that lead to spread deforestation and substitution of native vegetation by Agave farmers. the gullies have usually rectangular shape and depths up to 3 meters in the lower areas of the catchment.

Code - EBGEM

```fortran
! Entropy based gully erosion model V 1.0
! Initial incision using Watson (1986) equations
! Detachment rate calculated by proporcionality od net shear stress
! Shear stress calculated using the Principle of Minimum Cross-Entropy
! Wall erosion by Siorchuk (1999)

prograg ebgem_v1
    implicit none
    character file1*30, file2*30, file3*30
    integer No, i, i0, io, test_depth2, j
    logical test_depth1
    real Q(1500), qi, t, n, S, Kr, tauC, rhoB, rhoW, g, Gw, NEL, ch, phi, pi, Lim
    real par1, tauA, w0, Mr, width, depth, area, flow_w, flow_d, w_step, b_step
    real dt_w(21), dt_b(21), Dr_w(21), Dr_b(21), da_w, da_b, new_area, new_depth, new_width
    real flow_a, flow_p, Rh, T0, Lb, Lw, Lr, Cfs, SFw, Tw_a, Tb_a, Tw_m, Tb_m, tau
    real points, erro_s2m, erro_s2w, erro_s2b, kw, kb, x_w, x_b, x_new, gy, fx, dfx
    real T_w(21), T_b(21), tau_w(21), tau_b(21), max_increase_depth, area_SM, width_SM
    real k1, k2, x, xa, f1, df1, Ang, lbd_w, lbd_b, Dvcr, es

    !common variables
    common /grands/ flow_w, flow_d, S, Gw, Tw_a, Tb_a, Tw_m, Tb_m, lbd_w, lbd_b, tau_w, tau_b
    common /NewtonPhi/Ch, g, rhoB, es, Phi, depth, Ang, pi

    write(*,*)'*****************************************'
    write(*,*)'        Entropy-based Gully erosion      '
    write(*,*)'              model - V 1.0              '
    write(*,*)'                                         '
    write(*,*)'                                         '
    write(*,*)'        TUB - Institut fur Okologie      '
    write(*,*)'         UFC - PPGEA - Hidrosed          '
    write(*,*)'                                         '
    write(*,*)'                                         '
    write(*,*)'           Pedro Alencar, 02.2021        '
    write(*,*)'                                         '
    write(*,*)'*****************************************'

!1. Load input data
    write(*,'(a)')'Insert the name of the file containing the runoff data (Discharge e Duration): '
    read(*,'(a30)')file1 !in the absense of measured data we sugest use the 30-min intensity
    write(*,'(a)')'Insert the name of the file containing the hillslope and soil data: '
    read(*,'(a30)')file2
    write(*,'(a)',advance='no') 'Insert the name of the output file: '
    read(*,'(a30)')file3


!
!1.1 count number of events
    open(50, file = file1, iostat=io, status='old')
    if (io/=0) stop 'Cannot open file!'
    No = 0
    do
        read(50,*,iostat=io)
        if (io/=0) exit
        No = No + 1
    end do
    write(*,*)No
    close(50)

!1.2 Read and load all discharges (30-min intensities)
    open(50,file=file1,status='old')
        do i=1, No
            read(50,*) Q(i)
        end do
    close(50)
    t = 1800. ! in seconds - I30
    open(60,file=file2, status='old')
    open(70,file=file3,status='new')

!1.3 Read parameterisation file
    read(60,*)n,S,tauC,kr,rhoB,es,nel,ch,phi,Lim
    write(*,*)'The number of Manning of the channel is....',n
    write(*,*)'The declivity of the hillslope is..........',S
    write(*,*)'The critical shear stress is...............',tauC
    write(*,*)'The rill erodibility coefficient is........',Kr
    write(*,*)'The soil Bulk density is...................',rhoB
    write(*,*)'The porosity of the soil is................',es
    write(*,*)'The depth of the nonerodible layer is......',nel
    write(*,*)'The soil cohesion is.......................',ch
    write(*,*)'The internal friction angle is.............',phi
    write(*,*)'The threshold for wall erosion is..........',Lim
```

```
      close(60)
82
98        format(A43, 10X, I4)
99        format(A43, 4X, F10.4)
!!!!!!!!!!!!!!!! OUTPUT PREAMBLE !!!!!!!!!!!!!!!!!!!!
          write(70,*)'*****************************************'
87        write(70,*)'        Entropy-based Gully erosion          '
          write(70,*)'              model - V 1.0                  '
          write(70,*)'                                             '
          write(70,*)'        TUB - Institut fur Okologie          '
          write(70,*)'           UFC - PPGEA - Hidrosed            '
92        write(70,*)'                                             '
          write(70,*)'             Pedro Alencar, 2021             '
          write(70,*)'                                             '
          write(70,*)'*****************************************'
          write(70,*)' '
97        write(70,99)'The number of Manning of the channel is....',n
          write(70,99)'The declivity of the hillslope is..........',S
          write(70,99)'The critical shear stress is...............',TauC
          write(70,99)'The rill erodibility coefficient is........',Kr
          write(70,99)'The soil Bulk density is...................',rhoB
102       write(70,99)'The porosity of the soil is................',es
          write(70,99)'The depth of the nonerodible layer is......',nel
          write(70,99)'The soil cohesion is.......................',ch
          write(70,99)'The internal friction angle is.............',phi
          write(70,99)'The threshold for wall erosion is..........',Lim
107       write(70,98)'The number of rainfall events is...........',No
          write(70,*)' '
          write(70,*)' event    discharge (m3)    depth (m)     width (m)
                        area (m2)     lambda_wall      lambda_bed'
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
112 100       format(2X,I4,4X,F10.4,4X,A12)
101       format(2X,I4,4X,F10.4,4X,F10.4,4X,F10.4,4X,A10,4X,A10)
102       format(2X,I4,4X,F10.4,4X,F10.4,4X,F10.4,4X,F10.4,4X,F10.4,4X,F10.4)


!1.4 Definition of important constants
117   g = 9.7804 !m.s-2 - gravity, see extended documentation
      rhoW = 1000. !kg.m-3
      Gw = 9780.4 !N.m-3
      pi = 3.1415926536
      phi = phi*pi/180 !convert from degrees to radians
122
!1.5 Initialize important variables
      i = 1 !i is a counter
      depth = 0.
      area = 0.
127
!2. First erosion event
      do while (i .le. No)
          qi = Q(i)
          par1 = (n*qi/(S**(0.5)))**0.375 !flow parameter defined by Foster and Lane 1983
132       tau = 4867*par1*S !effective shear stress by Watson (1986), used in the initial incision
          if (tau .le. tauC) then
              write(70,100) i, qi,'NO EROSION'
              i = i+1
              else
137
              !initial incision's width:
              w0 = 2.66 * (qi**0.396) * (n**0.387) * (S**(-0.16)) * (tauC**(-0.24))

              Mr = Kr * (tau - tauC) / rhoB !downward moviment rate
142
              !lower is a function that compares two numbers  and delivers the lower one:
              depth = min(nel, Mr*1800)

              i0 = i !i0 is the event that causes the inicial incision
147           write(*,*)i,qi, depth, w0, tau, 'watson1'
              write(70,101)i, qi, depth, width, area, '-', '-'
!
              exit !end the loop, the value of i is preserved!
          end if
152   end do

      width = w0 !of the channel
      area = width*depth !of the channel

157   i = i0+1
      !downward erosion
      do while (i .le.No)
          qi = Q(i)
          par1 = (n*qi/(S**(0.5)))**0.375
162       flow_w = width
          flow_d = flow_depth(par1, flow_w)!function that calculates the depth using newton-raphson
```

```fortran
            if (flow_d .gt. depth) then
167               tau = 4867*par1*S

                  !initial incision's width:
                  w0 = 2.66 * (qi**0.396) * (n**0.387) * (S**(-0.16)) * (tauC**(-0.24))

172               Mr = Kr * dim(tau, tauC) / rhoB !downward moviment rate

                  !lower is a function that compares two numbers  and delivers the lower one:
                  depth = min(nel, depth + Mr*1800)

177               width = max(w0, width) !of the channel
                  area = width*depth !of the channel
                  write(70,101)i, qi, depth, width, area, '-', '-'
                  i = i+1
            else
182
            call shear_const


            call calib_ldb
187         write(*,*) i,Tw_a, Tb_a, Tw_m, Tb_m, lbd_w, lbd_b !############################
            w_step = flow_d/20 !shear step on the wall
            b_step = flow_w/40 !shear step on the bed. the shear stress in computed on half the section
            call dist_shear
            test_depth1 = (depth .lt. nel) test if the erosion reached the non-erosive layer
192         test_depth2 = test_depth1

            j = 1
            do while (j .le. 21) !calculate the available shear stress
                dt_w(j) =  dim(tau_w(j), tauC)
197             dt_b(j) =  dim(tau_b(j), tauC)
                !dim(X,Y) returns the difference X-Y if the result is positive; otherwise returns zero.


                j = j+1
202         end do

            Dr_w = Kr*dt_w
            Dr_b = Kr*dt_b
            max_increase_depth = Dr_b(21)*1800/rhoB
207
            da_w = w_step*sum(Dr_w)*1800/rhoB !\delta t is fixed at 1800 seconds
            da_b = b_step*sum(Dr_b)*1800/rhoB * test_depth2

            new_area = area + da_w + da_b
212         new_depth = depth + max_increase_depth
            new_depth = min(new_depth, NEL)
            new_width = new_area/new_depth

            width = new_width
217         depth = new_depth
            area = new_area
            write(*,*)i, qi, flow_w, flow_d, width, depth, area, 'entropy'   !########
            write(70,102)i, qi, depth, width, area, lbd_w, lbd_b
            i = i+1
222         end if
        end do

   if (area.gt.Lim) then
        Dvcr = (2.*ch*cos(phi)/(g*rhoB))/(sin(0.5*(phi+pi/2.)))**2.
227     if  (depth .gt. Dvcr) then
            CALL Newton_PHI
            width_SM = width +2.*depth/tan(Ang)
            area = depth*(width+width_SM)/2.
            write(70,102) 9999, qi, depth, width_SM, area, -999., -999.
232         write(70,*)''
            write(70,*)''
            write(70,*)'**ATENCION: The eroded channel has wall erosion (Sidorchuk, 1999)'
            write(70,*)'----------- Width displayed in the last line (9999) is the top with!'
            write(70,*)''
237
        end if
    end if


242 !Declaration of functions
    contains
    !F1. function to select the lower value (substituted by function min)
    function lower(a1,b1)
        real a1, b1, lower
247     if (a1.le.b1) then
            lower = a1
```

165

```fortran
              else
                  lower = b1
          end if
252  end function lower

     !F2. function to calculate flow depth using
     ! manning equation and newton-raphson optimisation method
     function flow_depth(a2,b2)
257      real a2, b2, flow_depth, erro2, erro_M2, f2, df2, x2, xn2
         erro_M2 = 0.0001

         x2 = 0.01
         erro2 = 1000.
262      do while (erro2 .gt. erro_M2)
              f2 = b2**5 * x2**5
              f2 = f2/(b2 + 2*x2)**2 - a2**8

              df2 = (b2**5)*(x2**4)*(5*b2 + 6*x2)
267           df2 = df2/((b2+ 2*x2)**3)

              xn2 = x2 - f2/df2
              erro2 = abs(xn2-x2)
              x2 = xn2
272      end do

         flow_depth = x2

       end function flow_depth
277
     end program


     !subroutines
282
     !S1. Subroutine to calculate the shear stress parameters (max and avg) by Knight 1994
     subroutine shear_const
       implicit none
       real flow_w, flow_d, flow_a, flow_p, Rh, T0, S, Gw, Lb, Lw, Lr, Csf, Sfw
287    real Tw_a, Tb_a, Tw_m, Tb_m

     common /grands/ flow_w, flow_d, S, Gw, Tw_a, Tb_a, Tw_m, Tb_m

       flow_a = flow_w*flow_d
292    flow_p = flow_w + 2.*flow_d

       Rh = flow_a/flow_p
       T0 = Gw*S*Rh

297    Lb = flow_w/2.
       Lw = flow_d

       !1.1 Calculating shear stress parameters based on Knight and Sterling (2000)
       Lr = Lb/Lw
302
       if (Lr .lt. 4.374) then
              Csf = 1.
              else
                  Csf = 0.6603*(Lr**0.28125)
307    end if

       Sfw = -3.23*log10(Lr/1.38 + 1) + 4.6052
       Sfw = 0.01*Csf*exp(Sfw)

312    Tw_a = T0 * Sfw * (1+Lr)
       Tb_a = T0 * (1-Sfw) * (1 + 1/Lr)
       Tw_m = T0 * Sfw * 2.0372 * (Lr**0.7108)
       Tb_m = T0 * (1-Sfw) * 2.1697 * (Lr**(-0.3287))
     end subroutine
317
     !S2. Subroutine to calibrate the lagrange multipliers to assess shear stress
     !    in open channels' boundaries

     subroutine calib_ldb
322
     !References - Sterling and Knight (2002)
     !            - Bonakdari et al. (2014)
     !            - Nocedal and Wright (2006)
       implicit none
327    integer cont1w, cont1b
       real Tw_a, Tb_a, Tw_m, Tb_m, lbd_w, lbd_b
       real Tr_w, Tr_b, erro_s1m, erro_s1w, erro_s1b, xs1, xs1_n,xs1s, xs1s_n
       real fp, dfp,lbd1_w, lbd1_b
       real flow_w, flow_d, S, Gw, tau_w, tau_b
332    common /grands/ flow_w, flow_d, S, Gw, Tw_a, Tb_a, Tw_m, Tb_m, lbd_w, lbd_b, tau_w, tau_b
```

```fortran
Tr_b = min(0.99,Tb_a/Tb_m) !For a Tr equal or larger than 1 there is no solution.
Tr_w = min(0.99,Tw_a/Tw_m)

erro_s1m = 1E-5 !Defines precision of approximation

!Sterling's routine was removed
! 2. Calculates POMCE's lambdas

        !2.1 Calibrating lambda for the wall

        if (Tr_w .lt. 0.97) then !check if it is possible a direct solution

            cont1w = 0
            xs1 = -10
            erro_s1w = 1000.0
            do while(erro_s1w .gt. erro_s1m)
                fp = exp(xs1) - xs1 - 1.
                fp = 2/xs1 - xs1/fp - Tr_w

                dfp = (exp(xs1) - xs1 - 1)
                dfp = (xs1*exp(xs1) - xs1)/dfp**2 - 2/(xs1**2) - 1/dfp

                xs1_n = xs1 - fp/dfp      !Newton-Raphson method

                erro_s1w = abs(xs1 - xs1_n)
                xs1 = xs1_n
                cont1w = cont1w+1
            end do
            lbd1_w = xs1

            else !for indirect solution
                xs1 = ((Tr_w + 2.)**2) - 8.
                xs1 = (Tr_w - 2) - sqrt(xs1)
                xs1 = xs1/(2 - 2*Tr_w)

                erro_s1w = -1
                cont1w = 1
                lbd1_w = xs1
        end if

        !2.2 Calibrating lambda for the bed
        if (Tr_b .lt. 0.97) then
            cont1b = 0
            xs1= -10 !initial guess
            erro_s1b = 1000.0

            do while(erro_s1b .gt. erro_s1m)
                fp = exp(xs1) - xs1 - 1.
                fp = 2/xs1 - xs1/fp - Tr_b

                dfp = (exp(xs1) - xs1 - 1)
                dfp = (xs1*exp(xs1) - xs1)/dfp**2 - 2/(xs1**2) - 1/dfp

                xs1_n = xs1 - fp/dfp       !m todo de newton

                erro_s1b = abs(xs1 - xs1_n)
                xs1 = xs1_n

                cont1b = cont1b+1
            end do
            lbd1_b = xs1

            else !for indirect solution
                xs1 = ((Tr_b + 2.)**2) - 8.
                xs1 = (Tr_b - 2) - sqrt(xs1)
                xs1 = xs1/(2 - 2*Tr_b)

                erro_s1b = -1
                cont1b = 1
                lbd1_b = xs1
        end if

    lbd_w = lbd1_w
    lbd_b = lbd1_b

end subroutine

!S3. Subroutine to distribute the shear stress
subroutine dist_shear
  implicit none
  integer points, kw, kb
  real lbd_w, lbd_b, T_w(21), T_b(21), tau_w(21), tau_b(21), Tw_m, Tb_m !SS variables
```

B. Appendix B

```fortran
417    real gy, x_w, x_b, x_new, fx, dfx, erro_s2m, erro_s2w, erro_s2b, frac
       real flow_w, flow_d, S, Gw, Tw_a, Tb_a
       common /grands/ flow_w, flow_d, S, Gw, Tw_a, Tb_a, Tw_m, Tb_m, lbd_w, lbd_b, tau_w, tau_b

       points = 21 !number of points for calculation of shear stress in each sector
422    erro_s2m = 1E-5 !for calibration of T in the PoMCE method

     !1. distribution on the wall
         if (lbd_w .gt. -38) then
     !          write(*,*)'here'
427

             kw = 1
             do while (kw .le. points)
                 frac = (kw*1.)/(points*1.) !convert from integer to real
                 gy = 1. - exp(-lbd_w) * (lbd_w + 1.)
432              gy = 1. - gy * frac
                 erro_s2w = 1000.
                 x_w = 0.9
                 do while (erro_s2w .gt. erro_s2m)
                     fx = exp(-lbd_w * x_w) * (lbd_w * x_w + 1.) - gy
437                  dfx = -1. * (lbd_w**2) * x_w * exp(-lbd_w * x_w)
                     x_new = x_w - fx/dfx
                     erro_s2w = abs(x_new - x_w)
                     x_w = x_new
                 end do
442              T_w(kw) = x_w

                 kw = kw + 1
             end do
             else
447

             kw = 1
                 do while (kw .le. points)
                 frac = (kw*1.)/(points*1.)
                 gy = log(-lbd_w*frac) - lbd_w
452              erro_s2w = 1000.
                 x_w = 0.9
                     do while (erro_s2w .gt. erro_s2m)
                         fx = log(-lbd_w*x_w) - lbd_w*x_w - gy
                         dfx = 1/x_w - lbd_w
457                      x_new = x_w - fx/dfx
                         erro_s2w = abs(x_new - x_w)
                         x_w = x_new
                     end do
                     T_w(kw) = x_w
462                  kw = kw + 1
                 end do
         end if


467      !3.1 Shear stress distribution in the bed
         if (lbd_b .gt. -38) then
             kb = 1
     !           cont2 = 0
             do while (kb .le. points)
472              frac = (kb*1.)/(points*1.)
                 gy = 1. - exp(-lbd_b) * (lbd_b + 1.)
                 gy = 1. - gy * frac
                 erro_s2b = 1000.
                 x_b = 0.9
477              do while (erro_s2b .gt. erro_s2m)
                     fx = exp(-lbd_b * x_b) * (lbd_b * x_b + 1.) - gy
                     dfx = -1. * (lbd_b**2) * x_b * exp(-lbd_b * x_b)
                     x_new = x_b - fx/dfx
                     erro_s2b = abs(x_new - x_b)
482                  x_b = x_new
                 end do
                 T_b(kb) = x_b
                 kb = kb + 1
             end do
487

             else !lbd_b <= -38
                 do while (kb .le. points)
                     frac = (kb*1.)/(points*1.)
                     gy = log(-lbd_b*frac) - lbd_b
492                  erro_s2b = 1000.
                     x_b = 0.9
                     do while (erro_s2b .gt. erro_s2m)
                         fx = log(-lbd_b*x_b) - lbd_b*x_b - gy
                         dfx = 1/x_b - lbd_b
497
                         x_new = x_b - fx/dfx
                         erro_s2b = abs(x_new - x_b)
                         x_b = x_new
```

```fortran
                    end do
                    T_b(kb) = x_b
                    kb = kb + 1


                end do
            end if


            tau_w = T_w*Tw_m
            tau_b = T_b*Tb_m
    end subroutine

!S4. Subroutine to calculate wall stable angle
    subroutine Newton_PHI
        implicit none
        integer i
        real k1,k2,f1,df1,x,xa,erro,Ch,g,rhoB,es,Pa,Phi,depth,Ang,pi
        COMMON /NewtonPhi/Ch,g,rhoB,es,Phi,depth,Ang,pi

        i=0
        k1 = Ch/(g*rhoB*depth)
        k2 = tan(Phi)*(rhoB - es*1000.)/1000.

        erro=1000000.
        x = pi/4.
            do while (erro.gt.0.00001)

            f1 = k2*(cos(x))**2 - sin(2*x)/2. - k1
            df1 = -2*k2*cos(x) - cos(2*x)

            xa = x - f1/df1
            erro = abs(x-xa)
            x=xa
            i=i+1


        end do
    Ang = x
    end subroutine



    !Description of variables:
    !All units in SI!
    !
    !integer
    !No = Number of events
    !i = general counter of evetns (first loop layer)
    !j = specific counter of events (second loop layer)
    !i0 = event that causes initial incision
    !test_depth2 = auxiliar variable to test if depth > nel
    !
    !logical
    !test_depth1 = auxiliar variable to test if depth > nel
    !
    !real - general variables
    !Q(i) = peak discharge (m3/s) > assumed equal to the I30 of the rainfall event - vector
    !q = event peak discharge (m3/s)
    !t = event duration - fixed at 30 minutes > t = 1800. (Alencar et al, 2019 - HESS)
    !n = Maner coefficient > see data in Chow (1959) - Table5-6 page 110
    !S = Slope in m/m
    !Kr = Rill erodibility (s/m) > see equation from WEPP model (Flanagan 1995) and Alberts 1989
    !tauC = Critical shear stress (Pa) > see equation from WEPP model (Flanagan 1995) and Alberts 1989
    !rhoB = Bulk density (kg/m3)
    !rhoW = Density of water (1000 kg/m3)
    !g = gravity acceleration > g = 9.7804 m/s2 > from WGS84 model - g = g45 - 0.5(g90-g0)*cos(2*Lat)
    ![g0 = 9.780; g45 = 9.806; g90 = 9.832; Lat = 5]

    !Gw = Specific weight of water > Da*g
    !NEL = depth of the Non-Erodible Layer (m) > obtained from measurements
    !ch = soil cohesion (Pa)
    !phi = internal friction angle (in degrees)
    !pi = 3.1415926536
    !
    !real - event's variables
    !par1 = flow parameter = (n Q/S^0.5)^0.375
    !tauA = event's average shear stress (Pa) - used on the first incision, in the watson equations
    !w0 = width of initial incision (in meters --- watson equations)
    !Mr = downward movement rate (m.s-1)
    !width = channel width (m)
    !depth = channel depth (m)
    !area = channel area (m2)
    !flow_w = flow width (m)
    !flow_d = flow depth (m)
    !w_step, b_step = length of the resolution for the calculation of shear stress; by defaut,
    !the section is divided in 80 points.
```

```
      !dt_w, dt_b = shear stress distributions (vector)
      !Dr_w, Dr_b = detachement rate (vectors)
587   !da_w, da_b = total displaced area (eroded area) due to the detachement rate.
      !new_area, new_depth, new_width = geometric properties of the altered (eroded) section.
      !Equations keep rectangular geometry.

      !max_increase_depth = test for maximum depth
592   !area_SM, width_SM = variables if the Sidorchuk model subroutine is triggered
      !variables in subroutine *shear_const*
      !flow_a = flow area (m2)
      !flow_p = flow wet perimeter (m)
      !Rh = hydraulic radius (m)
597   !T0 = hydraulic average shear stress = g*rhoW*S*Rh (Pa)
      !Lb = bed length (m)
      !Lw = wall lengh (m)
      !Lr = Length ratio (-)
      !Cfs, SFw = auxiliar variables (Knight 2000)
602   !Tw_a = average shear stress on the wall (Pa)
      !Tb_a = average shear stress on the bed (Pa)
      !Tw_m = max shear stress on the wall (Pa)
      !Tb_m = max shear stress on the bed (Pa)

607   !variables in subroutine *calib_lbd*
      !Tr_b = shear stress ratio a/m for the bed
      !Tr_w = shear stress ratio a/m for the wall
      !erro_s1m, erro_s1w, erro_s1b = errors test and max
      !cont1w, cont1b = counters
612   !xs1, xs1_n = variable (lbd1)
      !fp, dfp = calibration equations (from pomce)
      !lbd1_w, lbd1_b = auxiliar lbds
      !lbd_w, lbd_b = calibrated lambdas

617   !variables in subroutine *dist_shear*
      !points = number of points on the wall and (half-) bed where the shear stress will be calculated
      !erro_s2m, erro_s2w, erro_s2b = errors test and max
      !kw, kb = position control
      !x_w, x_b, x_new, gy, fx, dfx
622   !T_w, T_b = auxiliar (intern) vectors
      !tau_w, tau_b = vectors of shear stress (length by defaut = 21; extern)

      !variables in subroutine *Newton_PHI*
      !i = counter
627   !k1, k2 = auxiliars
      !x, xa, f1, df1 = newton variables and functions
      !Ang = anglw of stability
      !
      !auxiliars
632   !a1, b1, lower = in funcion *lower*
      !a2, b2, flow_depth2, erro2, erro_M2, f2, df2, x2, xn2 = in function *flow_depth*
```

C

Appendix C

In this appendix, we present the supplementary material of chapter 4 - Entropy-based temporal downscaling of precipitation as tool for sediment delivery ratio assessment

Introduction

In Figure C.1 we present the map of pluviometric stations with daily and sub-daily data in the Brazilian Northeastern Region. As discussed in the main text, there are few stations monitoring sub-daily precipitations in the region. Additionally, most stations in the left of the plot have a short time series and/or with gaps.

Materials and methods

The Principle of Maximum Entropy

The Principle of Maximum Entropy (PoME) is grounded on the concept of entropy as a measure of uncertainty or information, as proposed by (Shannon, 1948). Based on abstraction, Jaynes (1957a) and Jaynes (1957b) proposed the PoME to obtain the least-biased probability function on the basis of known information represented as constrains. The Shannon entropy equation is expressed as (Eq. C.1):

$$h_x = - \int f(x) \ln f(x) dx \qquad (C.1)$$
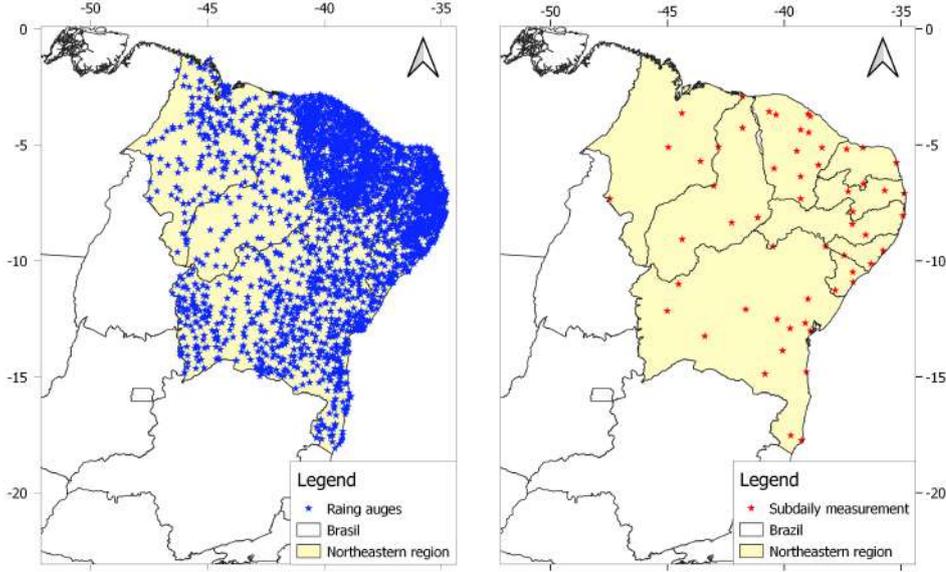
Figure C.1: Map of Ville de Paris and Automatic stations in the Brazilian Northeast Managed by the Brazilian National Water Agency (ANA, 2019).

h_x is the total entropy for the variable x. The function $f(x)$ that maximises h_x is the one that does not consider any non-proved hypothesis. To maximize Eq. C.1, subjected to the constrains, we can formulate the Lagrangian function $\mathcal{L}$ (Eq. C.2) and differentiate in respect to f and equals the derivative to zero (Eq. C.3).

$$\mathcal{L} = - \int_{x_0}^{x_1} f(x) \ln f(x) dx - \sum_{r=0}^{n} \lambda_r \left[\int_{x_0}^{x_1} f(x)\, g_r(x) dx - C_r \right] \tag{C.2}$$

$$\frac{\partial \mathcal{L}}{\partial f} = 0 \rightarrow \frac{\partial \mathcal{L}}{\partial f} = -1 - \ln f(x) - \sum_{r=0}^{n} \lambda_r\, g_r(x) = 0 \tag{C.3}$$

$\lambda_0, \lambda_1, ..., \lambda_n$ are the Lagrange multipliers. $g_r(x)$ are functions of x related to the constraints. n is the number of restrictions besides the trivial ($r = 0 \rightarrow \int f(x)dx = 1$). Solving Equation C.3 for $f(x)$ one finds the probability distribution in terms of the Lagrange multipliers as in Eq C.4.

$$f(x) = \exp \left[- \sum_{r=0}^{n} \lambda_r\, g_r(x) \right] \tag{C.4}$$

The SYPoME Model

Proposed by de Araújo, 2007, the SYPoME (Sediment Yield Model based on the Principle of Maximum Entropy) allows the user to assess the hillslope sediment production of each event and is given by Equation C.5:

$$Q_s = \bar{\varepsilon} \, A \, SDR = \bar{\varepsilon} \, A \frac{e^{\lambda L_m} \left(L_0 - x_0\right) \lambda - \left(e^{\lambda(L_0 - x_0)} - 1\right)}{\lambda L_0 \left(e^{\lambda(x_0 + L_m)} - 1\right)} \tag{C.5}$$

$\bar{\varepsilon}$ (Mg ha^{-1} yr^{-1}) is the gross erosion obtained, for example, by using the Universal Soil Loss Equation (USLE – Wischmeier and Smith, 1978), A (ha) the hillslope contribution area, L_0 the hill slope length (m), L_m the maximum sediment travel distance (m), x_0 is the initial position of erosion in the hillslope and λ is a Lagrange multiplier. The ratio of the sediment portion that reaches rivers and promotes siltation (Q_s) and all mobilised sediment ($\bar{\varepsilon} \, A$). The SDR is restricted to a closed interval ($SDR \in [0, 1]$).

The parameters λ and L_m can be obtained by solving the systems of equations derived with the PoME (Eq. C.6)

$$\begin{cases} \dfrac{1}{L_m} = \dfrac{e^{\lambda \, (x_0 + L_0)/2}}{e^{\lambda \, (x_0 + L_m)} - 1} \\ \dfrac{e^{\lambda \, (x_0 + L_m)}[\lambda \, (L_m + x_0) - 1] - e^{\lambda \, x_0}(\lambda \, x_0 - 1)}{\lambda \left(e^{\lambda \, (x_0 + L_m)} - 1\right)} = K_v \left(\dfrac{\rho_s}{\rho_s - \rho}\right) \dfrac{\Omega \, L_0}{g \, \bar{\varepsilon} \, v_s} \end{cases} \tag{C.6}$$

g (m s^{-2}) is the gravity, ρ (kg m^{-3}) is the density of water, ρ_{s} (kg m^{-3}) is sediment density, Ω (J s^{-1} m^{-2}) the stream power (Eq. C.7) according to Bagnold (1977), v_s (m s^{-1}) is the sediment settling velocity and K_v is the delivery parameter related to surface conditions, which be calibrated or obtained as function of the parameters CP of the USLE. The system of Equations C.6 allows us to obtain the two parameters necessary to calculate the SDR.

$$\Omega = \rho \, g \, S_0 \, R_H \, U \tag{C.7}$$

S_0 (m m^{-1}) is the slope; R_H (m) the hydraulic radius that can be approximated to the flow depth for wide hills; and U (m s^{-1}) is the flow velocity. In his original work, de Araújo, 2007 achieved good results (average absolute error 20%) with the model by using the average velocity for each event, given by Equation C.8.

$$\bar{U} = \left(\frac{D}{H_e}\right)^{-1} \tag{C.8}$$

H_e (mm) is the effective precipitation or total runoff and D (s) the total duration of the event. Hence, instead of requiring the knowledge of the complete hydrograph, we only need the information on the effective precipitation initiation and on its end, usually unavailable.

Gross Erosion Assessment

The Universal Soil Loss Equation (Wischmeier and Smith, 1978) is an empirical equation with simple implementation as expressed by the product below (Eq. C.9):

$$\bar{\varepsilon} = R \, K \, LS \, C \, P \tag{C.9}$$

R (rainfall and runoff factor or erosivity factor) represents the total energy of an event or a series of events which may produce erosion; K (erodibility factor) indicates how much the soil in the studied area is prone to be mobilised by the rain energy; LS (topographic factor) is the length factor and S the slope factor, directly connected to the topography; C (cover and management factor) is a measure of the effect of all cover and management variables, such as type and condition of vegetation and tillage practices; and P (management practice factor) accounts for good practices to reduce erosion, as contouring and terracing.

Erosivity Factor (R)

In order to calculate the gross erosion by employing the USLE we need to assess the erosivity value (R – MJ ha^{-1} h^{-1}). We used two approaches:

i.   Probabilistic approach

Based on measured data concerning sub-daily precipitation, we studied the best probabilistic distribution (uniform, gaussian, two-parameter gamma and beta distributions were tested) for the variable I_{30}/H. Using an estimated I_{30} (mm h^{-1}) we calculated the event erodibility using Equation C.10

$$R = E \, I_{30} \tag{C.10}$$

$$E = \begin{cases} 11.9 + 8.73 \log_{10} \bar{I} & \forall \, H < 76.2mm \\ 28.3 & \forall \, H \geq 76.2mm \end{cases} \tag{C.11}$$

where E is a storm's kinetic energy, given by the Equation C.10. In Eq. C.11 above, H (mm) is the total precipitation and $\bar{I}$ (mm h^{-1}) the average intensity. Note that we obtain $\bar{I}$ as the ratio H/D.

ii.   Regional approach

Using measured data of rainfall intensities in a semiarid region (de Figueiredo et al., 2016), an equation for the monthly erosivity was calibrated. Event erosivity was obtained by distributing the month's erosivity proportionally to the event's total precipitation within the month (Eq. C.12).

$$\begin{cases} R_m = \alpha \left(\frac{H_m^2}{H_a} \right)^{\beta} \\ R_{i,m} = \frac{R_m \, H_{i,m}}{H_m} \end{cases} \tag{C.12}$$

R_m is the month's total erosivity, H_m the month's total precipitation, H_a the average annual precipitation, and $R_{i,m}$ the erosivity of the $i - th$ event of the month m, and $H_{i,m}$ the precipitation of the $i - th$ event of the month m; α and β are regional calibrated parameters equalling 565 and 0.42 respectively.

Erodibility factor (K)

Soil erodibility was estimated using the Soil Classification Maps of Brazil (IPECE, 2007) and the correspondent erodibility factor as obtained experimentally by Silva (1978).

Topography factor (LS)

The topography factor was calculated applying equation C.13.

$$LS = 0.00984\, L_r^{0.63}\, S^{1.18} \tag{C.13}$$

$$L_r = \frac{A_q}{4 \sum L_{dem}} \tag{C.14}$$

where S is the slope in percentage and L_r is the average slope length, given by equation C.14. A_q is the area of the pixel, sub-basin, or landscape unity and $\sum L_{den}$ the sum of all water paths within A_q.

Cover and management practice factor (CP)

We used satellite images (LandSat 8) and field surveys in order to identify the land use. From land use maps the parameter C was mapped using the values of table 8.8 of Haan, Barfield, and Hayes (1994, p. 266). The practice factor P was assumed equals the unity, since no management practices where identified in the areas.

Runoff

To estimate the total runoff per event we used the Soil Conservation Service, 1972 Curve Number method (Eq. C.15). The *CN* value was estimated on the basis of land use, soil properties and antecedent moisture (Mishra and Singh, 2003). I_a accounts for all initial abstractions and S for the potential maximum retention of the catchment, all in millimetres. I_a is often represented as a fraction ϕ of S. In this study, ϕ was assumed equals 0.20 for all study areas. S is a function of *CN* (Eq. C.16).

$$H_e = \frac{(H - I_a)^2}{(H - I_a + S)} \tag{C.15}$$

$$S = 25.4 \left(\frac{1000}{\text{CN}} - 10 \right) \tag{C.16}$$

The duration of the runoff was assumed to be equal to the duration of rainfall for the small catchments (< 10 hectares). For the medium, such as Aiuaba and Canabrava, field measurements suggest a duration, on average, 2.5 times longer than the rainfall (de Figueiredo et al., 2016) and for the larger catchments we used the Snyder, 1938 Unit Hydrograph.

USLE Data

In Table C.1 we present the values of the USLE parameters obtained accordingly to Wischmeier and Smith (1978). The parameter P was assumed equal to one, for no management practices were identified in the regions.

Table C.1: Average characteristics of the study areas - LULC, arae and USLE parameters

Name	Land Use	Area (km²)	Slope (%)	K[a]	L (-)	S (-)	C (-)
Canabrava	Agriculture and open range cattle rasing	2.9	6.6%	0.032	3.252	0.606	0.01
Aiuaba	Conservation area with native vegetation (Caatinga)	11.53	18.0%	0.015	3.16	1.944	0.0005
Várzea da Volta	Agriculture and open range cattle raising	155	22.1%	0.028	3.766	2.364	0.028
Acarape	Agriculture and open range cattle raising	208	10.1%	0.037	2.766	1.115	0.015
Sumé 2	Experimental area - Preserved vegetation (Caatinga)	0.0107	6.1%	0.021	1.126	0.523	0.008
Sumé 4	Experimental area - Degraded land without vegetation	0.0048	6.8%	0.021	0.848	0.64	1.000
Gilbués	Abandoned land under desertification process without vegetation	0.0004	15.6%	0.007	1.083	1.698	0.771

[a] K in (Mg h MJ⁻¹ mm⁻¹)

Code - SYPoME

```fortran
1          PROGRAM SYPoME

!       PROGRAM TO SIMULATE SEDIMENT YIELD USING POME-EQUATION
!          1. VARIABLES DECLARATION

6          INTEGER nprec,iprec,ncell,icell,nev,iev,irep,i
           CHARACTER arquivo1*20,arquivo2*20
           CHARACTER*8, DIMENSION(10,3000) :: dia
           INTEGER, DIMENSION(10,3000) :: id
           REAL, DIMENSION(10,3000) :: D,dur,R
11         COMMON /EVENTOS/ id,D,dur,R
           INTEGER, DIMENSION(100) :: igauge
           REAL ds,vs,A,K,CP,S0,S,w0,fL,L0,Kv
           COMMON /CELULAS/ ds,vs,A,K,CP,S0,S,w0,fL,L0,Kv,igauge

16 !         2. MAIN PROGRAM

           CALL ABERTURA(arquivo1,arquivo2,nprec,ncell,nev,irep)

!        read rainfall-related data of the events
21         iprec = 0
           DO WHILE (iprec.lt.nprec)
              iprec=iprec+1
              READ(20,*) i
              IF (i.ne.iprec) THEN
26               WRITE(*,*)  ' Incompatibility between indexes of gauge stations !!!'
                 WRITE(21,*) ' Incompatibility between indexes of gauge stations !!!'
              ENDIF
              WRITE(21,201) ' Precipitation gauge number ..................... ',iprec
              WRITE(21,*)    '————————————————————————————————————————————————————'
31            WRITE(21,*)    '    i     id     date        D(mm)     Dur(min)     R(MJ.mm/ha/h)       '
              WRITE(21,*)    '————————————————————————————————————————————————————'
              WRITE(*,201)   ' Precipitation gauge number ....1................. ',iprec
              WRITE(*,*)     '————————————————————————————————————————————————————'
              WRITE(*,*)    '    i     id     date        D(mm)     Dur(min)     R(MJ.mm/ha/h)       '
36            WRITE(*,*)     '————————————————————————————————————————————————————'
              iev = 0
              DO WHILE (iev.lt.nev)
                 iev=iev+1
                 READ(20,*) id(iprec,iev),dia(iprec,iev),D(iprec,iev),dur(iprec,iev),R(iprec,iev)
41               WRITE(21,202) iev,id(iprec,iev),dia(iprec,iev),D(iprec,iev),dur(iprec,iev),R(iprec,iev)
              ENDDO
           WRITE(21,*)      '————————————————————————————————————————————————————————————'
           WRITE(*,*)       '————————————————————————————————————————————————————————'
           ENDDO
46
!        read physiographic-related data of the cells and compute sediment yield
           SSY = 0
           SGEr = 0
           icell = 0
51         DO WHILE (icell.lt.ncell)
              icell = icell+1
              READ (20,*) icell,ds,vs,A,K,CP,S0,w0,Kv,igauge(icell)
              CALL CALCSY(icell,nev,SY,GEr,irep)
              SSY = SSY + SY
56            SGEr = SGEr + GEr
           ENDDO

!        close program
           WRITE(21,203) ' Watershed gross erosion (kg) ............ ',SGEr
61         WRITE(21,203) ' Watershed sediment yield (kg) ........... ',SSY
           WRITE(21,204) ' Watershed average delivery ratio ........ ',SSY/SGer
           WRITE(21,*)
           WRITE(21,*)    ' Program concluded successfully.'
           WRITE(*,203) ' Watershed gross erosion (kg) ............ ',SGEr
66         WRITE(*,203) ' Watershed sediment yield (kg) ........... ',SSY
           WRITE(*,204) ' Watershed average delivery ratio ........ ',SSY/SGer
           WRITE(*,*)
           WRITE(*,*)    ' Program concluded successfully.'
           CLOSE(20)
71         CLOSE(21)
   201 FORMAT (a50,i4)
   202 FORMAT (i5,2x,i5,2x,a8,2x,f6.2,5x,f8.1,5x,f7.1)
   203 FORMAT (a44,e10.4)
   204 FORMAT (a44,f5.3)
76         END

!         3. SUBROUTINE THAT OPENS PROGRAM

           SUBROUTINE ABERTURA(arquivo1,arquivo2,nprec,ncell,nev,irep)
```

```fortran
81
          CHARACTER arquivo1*20,arquivo2*20,title*20
          INTEGER    nprec,ncell,nev,irep

          WRITE(*,*)' SEDIMENT-YIELD ESTIMATION - SYPOME3'
86        WRITE(*,*)
          WRITE(*,*)' * Version 3'
          WRITE(*,*)' * SY equation based on the principle of maximum entropy'
          WRITE(*,*)' * Program can only compute up to 3000 events'
          WRITE(*,*)'                                '
91        WRITE(*,*)' Universidade Federal do Ceara'
          WRITE(*,*)' Jose Carlos de Araujo'
          WRITE(*,*)' Technische Universitat Berlin'
          WRITE(*,*)' Pedro Alencar'
          WRITE(*,*)' 2019'
96        WRITE(*,*) '----------------------------------------------------------------'

          WRITE(*,*)'Type the name of the input file:'
          READ(*,302) arquivo1
          OPEN(20,file=arquivo1,status='old')
101       READ(20,*)title
!         OPEN(20,file='in.txt',status='old')
          WRITE(*,*)
          WRITE(*,*)'Type the name of the output file:'
          READ(*,302) arquivo2
106       OPEN(21,file=arquivo2,status='new')
          WRITE(*,*)
!         OPEN(21,file='out.txt',status='new')
          WRITE(*,*)'Do you need a complete (1) or a simplified (2) report?'
          READ(*,*) irep
111       IF(irep.ne.1.and.irep.ne.2) THEN
              WRITE(*,*) 'The number is not an option. Default (complete) report will be provided'
              irep = 1
          ENDIF

116       WRITE(21,*)' SEDIMENT-YIELD ESTIMATION - SYPOME3'
          WRITE(21,*)
          WRITE(21,*)' * Version 3'
          WRITE(21,*)' * SY equation based on the principle of maximum entropy'
          WRITE(21,*)' * Program can only compute up to 3000 events'
121       WRITE(21,*)'                                '
          WRITE(21,*)' Universidade Federal do Ceara'
          WRITE(21,*)' Technische Unibversitat Berlin'
          WRITE(21,*)' Jose Carlos de Araujo'
          WRITE(21,*)' Pedro Alencar'
126       WRITE(21,*)' 2019'
          WRITE(21,*)
          WRITE(21,*) '----------------------------------------------------------------'
          WRITE(21,*) 'Title: ', title
          WRITE(*,*) 'Title: ', title
131       WRITE(21,*) '----------------------------------------------------------------'
          WRITE(21,301) ' Input file .................................... ',"defaut"
          WRITE(21,301) ' Output file ................................... ',"defaut"
          WRITE(*,*) '----------------------------------------------------------------'
          WRITE(*,301) ' Input file .................................... ',"defaut"
136       WRITE(*,301) ' Output file ................................... ',"defaut"

          READ(20,*) nprec
          READ(20,*) ncell
          READ(20,*) nev
141       WRITE(21,303) ' Number of precipitation gauges ................ ',nprec
          WRITE(21,303) ' Number of cells ............................. ',ncell
          WRITE(21,303) ' Number of events ............................ ',nev
          WRITE(21,*)   '----------------------------------------------------------------'
          WRITE(*,303)  ' Number of precipitation gauges ................ ',nprec
146       WRITE(*,303)  ' Number of cells ............................. ',ncell
          WRITE(*,303)  ' Number of events ............................ ',nev
          WRITE(*,*)    '----------------------------------------------------------------'
    301   FORMAT (a50,a20)
    302   FORMAT (a20)
151 303   FORMAT (a50,i4)
          END

!      4. SUBROUTINE THAT COMPUTES SEDIMENT YIELD

156       SUBROUTINE CALCSY(icell,nev,SY,GEr,irep)

          INTEGER icell,nev,iev,irep
          REAL SY,SYi,GEr,GEri,beta
          INTEGER, DIMENSION(100) :: igauge
161       REAL ds,vs,A,K,CP,S0,S,w0,fL,L0,Kv
          COMMON /CELULAS/ ds,vs,A,K,CP,S0,S,w0,fL,L0,Kv,igauge

          L0 = 10000*A/(2*w0)
```

```fortran
         IF (S0.lt.0.090) THEN
            S = 10.8*SIN(ATAN(S0))+0.03
            ELSE
            S = 16.8*SIN(ATAN(S0))-0.50
         ENDIF
         beta = 11.16*SIN(ATAN(S0))/(3*(SIN(ATAN(S0))**0.8)+0.56)
         fL = (L0/22.1)**(beta/(beta+1))

         WRITE(*,400) ' Cell number................................',icell
         WRITE(*,401) ' Area (ha).................................',A
         WRITE(*,401) ' Soil density  (-)..........................',ds
         WRITE(*,401) ' Sedimentation velocity  (m/s)..............',vs
         WRITE(*,403) ' Drainage length w0 (m).....................',w0
         WRITE(*,403) ' Slope length L0 (m)........................',L0
         WRITE(*,401) ' Soil erodibility (ton.h/MJ/mm)............',K
         WRITE(*,402) ' Land-use factor CP  (-)....................',CP
         WRITE(*,402) ' Average slope S0 (-) ......................',S0
         WRITE(*,401) ' Slope factor S  (-)........................',S
         WRITE(*,401) ' Slope length factor L  (-)................',fL
         WRITE(*,401) ' Vegetation parameter Kv ...................',Kv

         WRITE(21,400) ' Cell number .............................',icell
         WRITE(21,401) ' Area (ha).................................',A
         WRITE(21,401) ' Soil density  (-).........................',ds
         WRITE(21,401) ' Sedimentation velocity  (m/s)..............',vs
         WRITE(21,403) ' Drainage length w0 (m).....................',w0
         WRITE(21,403) ' Slope length L0 (m)........................',L0
         WRITE(21,401) ' Soil erodibility (ton.h/MJ/mm)............',K
         WRITE(21,402) ' Land-use factor CP  (-)....................',CP
         WRITE(21,402) ' Average slope S0 (-) ......................',S0
         WRITE(21,401) ' Slope factor S  (-)........................',S
         WRITE(21,401) ' Slope length factor L  (-)................',fL
         WRITE(21,401) ' Vegetation parameter Kv ...................',Kv
         WRITE(21,400) ' Number of rainfall station ...............',igauge(icell)
         IF (irep.eq.1) THEN
             WRITE(21,*)   '_____',
             WRITE(21,*)    ' id   gross-er(kg) Stream-pw(J/s/m2) Lambda(1/m) Lm(m) SDR SY(kg/ha) '
             WRITE(21,*)   '_____',
         ENDIF
         SY = 0
         GEr = 0
         iev = 0
         DO WHILE (iev.lt.nev)
             iev=iev+1
             CALL EVENT(icell,iev,GEri,SYi,irep)
             GEr = GEr + GEri
             SY = SY + SYi
         ENDDO
         WRITE(*,*) '_____',
         WRITE(*,404) ' Total gross erosion (kg) in this cell ....',GEr
         WRITE(*,404) ' Total sediment yield (kg) ................',SY
         WRITE(*,405) ' Global sediment delivery ratio ...........',SY/GEr
         WRITE(*,*) '_____',
         WRITE(21,*) '_____',
         WRITE(21,404) ' Total gross erosion (kg) in this cell ....',GEr
         WRITE(21,404) ' Total sediment yield (kg) ................',SY
         WRITE(21,405) ' Global sediment delivery ratio ...........',SY/GEr
         WRITE(21,*) '_____',
     400 FORMAT (a44,i6)
     401 FORMAT (a44,f9.3)
     402 FORMAT (a44,f9.4)
     403 FORMAT (a44,f9.2)
     404 FORMAT (a44,e10.4)
     405 FORMAT (a44,f5.3)
         END

!      5. SUBROUTINE THAT PROCESSES DATA FROM EACH EVENT

         SUBROUTINE EVENT(icell,iev,GEri,SYi,irep)

         INTEGER iev,irep,icell
         REAL Lm,SDR,eps,erosion,streamp,f2,L2
         REAL GEri,SYi
         INTEGER, DIMENSION(100) :: igauge
         REAL ds,vs,A,K,CP,S0,S,w0,fL,L0,Kv
         COMMON /CELULAS/ ds,vs,A,K,CP,S0,S,w0,fL,L0,Kv,igauge
         INTEGER, DIMENSION(10,3000) :: id
         REAL, DIMENSION(10,3000) :: D,dur,R
         COMMON /EVENTOS/ id,D,dur,R

         eps = R(igauge(icell),iev)*K*CP*S*fL/10
         erosion = 10000*A*eps
         GEri = erosion
         streamp = 9807*L0*S0*(D(igauge(icell),iev)/1000)/(60*dur(igauge(icell),iev))
```

```fortran
          f2 = Kv*(ds/(ds-1))*streamp*L0/(9.807*eps*vs)

251       CALL PARAM(L0,Lm,L2,SDR,f2)
          SYi = erosion*SDR
          IF (irep.eq.1) THEN
              WRITE(21,501) id(igauge(icell),iev),erosion,streamp,L2,Lm,SDR,SYi/A
          ENDIF
256   501 FORMAT(i5,2x,e9.3,5x,e8.3,8x,e8.2,4x,f9.2,2x,f5.3,3x,f10.4)


          END


!       6. SUBROUTINE TO COMPUTE VARIABLE SDR AND PARAMETERS Lm & L2
261
      SUBROUTINE PARAM(L0,Lm,L2,SDR,f2)

      INTEGER i1
      LOGICAL run1
266   REAL L0,Lm,L2,SDR,f2
      REAL Lm1,Lm2,Lm3,tol1,err1,nmax1
      REAL*8 h1,h2,h3,a,b,aux_log

      Lm1 = L0/100.
271   x0 = L0-Lm1
      CALL Lambda(L0,f2,x0,Lm1,L2)
      a = L2*(Lm1+x0)
      b = L2*x0
      h1 = log(a - 1. - (b - 1.)*exp(-L2*Lm1)) - log(f2*L2) - log(1. - exp(-a))
276
      !print*,L2, f2, f2*L2, log(f2*L2)

      Lm2 = 50*L0
      x0 = 0.
281   CALL Lambda(L0,f2,x0,Lm2,L2)
      a = L2*(Lm2+x0)
      b = L2*x0
      alfa = a/b
      h2 = log(a - 1. - (b - 1.)*exp(-L2*Lm2)) - log(f2*L2) - log(1. - exp(-a))
286

      i1 = 0
      tol1 = 0.001
      nmax1 = 100.
291   run1 = .TRUE.
      DO WHILE (run1)
          i1 = i1+1

          Lm3 = (ABS(h1)*Lm2+ABS(h2)*Lm1)/(ABS(h1)+ABS(h2))
296       x0 = MAX(0.,L0-Lm3)
          CALL Lambda(L0,f2,x0,Lm3,L2)
          a = L2*(Lm3+x0)
          b = L2*x0

301       aux_log = (b-1)*exp(-L2*Lm3)
          aux_log = (a-1) - aux_log
          aux_log = abs(aux_log)

          h3 = log(aux_log) - log(f2*L2) - log(1. - exp(-a))
306       IF(h3*h2.le.0.) THEN
              Lm1 = Lm3
              h1 = h3
          ELSE
              IF(h3*h1.le.0.) THEN
311               Lm2 = Lm3
                  h2 = h3
              ELSE
                  IF(ABS(h1).le.ABS(h2)) THEN
                      Lm2 = Lm3
316                   h2 = h3
                  ELSE
                      Lm1 = Lm3
                      h1 = h3
                  ENDIF
321           ENDIF
          ENDIF
          err1 = ABS(h3)
          IF (err1.le.tol1.or.i1.ge.nmax1) THEN
              run1 = .FALSE.
326       ENDIF
      ENDDO
      Lm = Lm3
      x0 = max(0.,L0-Lm)
      CALL Lambda(L0,f2,x0,Lm,L2)
331
      SDR = (L0-x0)/L0
```

180

```fortran
        SDR = SDR*(fexp(L2*Lm)-L2*(L0-x0)/2.)
        SDR = SDR/fexp(L2*(x0+Lm))

336     END

!       7. SUBROUTINE TO COMPUTE PARAMETER LAMBDA-2, GIVEN L0, f2 AND Lm

        SUBROUTINE Lambda(L0,f2,x0,Lm,L2)
341
        INTEGER i2
        LOGICAL run2
        REAL L0,f2,x0,Lm,L2
        REAL xm,L21,L22,L23,tol2,err2,nmax2
346     REAL*8 g1,g2,g3,c1, c2, c3

        i2 = 0.

        nmax2 = 100.
351     tol2 = 0.001
        xm = (x0+L0)/2.

        L21 = (5E-8)/Lm
        c1 = L21*(0.5*(L0-x0) + Lm)
356     g1 = log(L21*Lm) + c1 - log(1. - exp(-L21*(x0+Lm)))

        L22 = 0.01
        c2 = L22*(0.5*(L0-x0) + Lm)
        g2 = log(L22*Lm) + c2 - log(1. - exp(-L22*(x0+Lm)))
361
        i2 = 0.
        run2 = .TRUE.
        DO WHILE (run2)
            i2 = i2+1.
366         L23 = (ABS(g1)*L22+ABS(g2)*L21)/(ABS(g1)+ABS(g2))
            L23 = MAX(L23,(5E-8)/Lm)
            c3 = L23*(0.5*(L0-x0) + Lm)
            g3 = log(L23*Lm) + c3 - log(1. - exp(-L23*(x0+Lm)))

371         IF(g3*g2.le.0.) THEN
                L21 = L23
                g1 = g3
            ELSE
                IF(g3*g1.le.0.) THEN
376                 L22 = L23
                    g2 = g3
                ELSE
                    IF(ABS(g1).le.ABS(g2)) THEN
                        L22 = L23
381                     g2 = g3
                    ELSE
                        L21 = L23
                        g1 = g3
                    ENDIF
386             ENDIF
            ENDIF
            err2 = ABS(g3)
            IF (err2.le.tol2.or.i2.ge.nmax2) THEN
                run2 = .FALSE.
391         ENDIF
        ENDDO
!       requirement due to numerical stability
        IF(L23*Lm.lt.5E-8) THEN
            L23 = (5E-8)/Lm
396     ELSE
            IF (L23*Lm.gt.1.) THEN
                L23 = 1./Lm
            ENDIF
        ENDIF
401         L2 = L23
    END

!       Function that computes approximation of exp(x) - 1 using McLaurin series
        REAL FUNCTION fexp(x)
406     REAL x
        fexp = x+(x**2)/2+(x**3)/6+(x**4)/24+(x**5)/120
        END
```

D

Appendix D

In this appendix, we present the structure of the package "fdClassify" (Table D.1) and usage of available functions of pre-processing, identification and visualization. The source code and package are available at https://github.com/pedroalencar1/fdClassify. To install the package in R, use the following command

```
1  devtools::install_github("pedroalencar1/fdClassify")
   library(fdClassify)
```

Table D.1: List of functions available in the package fdClassity

	Function	Description
1	actual_evap_day	Calculate daily actual evaporation
2	alencar2021	FD identification method proposed by Pedro Alencar
3	Christian2020	FD identification method from Christian et al. (2020)
4	Christian_clean_data_week	Function to clean the ESR data
5	conf_matrix	Auxiliar function for calculating and exporting confusion matrix
6	eddi	Calculate EDDI
7	eddi_percentile	Function to calculate EDDI in percentiles on a week acc. time
8	f.anomaly	Function to calculate anomalies
9	f.pentad	Accumulation into pentads
10	f.percentile	Percentile function
11	f.spei	Standard Precipitation Evaporation Index calculation
12	f.week	Accumulation into weeks
13	f.year	Separation by years - Internal function
14	FordLabosier2017	Method of Ford and Labosier (2017)
15	FordLabosier_gs	Modified method of Ford and Labosier (2017)
16	get.nc.data	Function to extract data from ERA5 raw data
17	get_df_era5	Generate DF from ERA5 raw (netcedf) data
18	get_df_fluxnet	Generate DF from FLUXNET raw data
19	hargreaves_day	Hargreaves-Samany daily-ET0 function
20	multicriteria_fd	Multiple criteria method of FD classification
21	Noguera2020	FD identification based on Noguera et al. (2020)
22	Osman2021	FD identification based on Osman et al. (2020)
23	param_loglogist	Calibrate parameters of multiple log-logist functions
24	Pendergrass2020	FD identification by Pendergrass et al. (2021)
25	penman_day	penman_day - Daily ET0 using Penman-Monteith
26	prepare.nc	Prepare .nc files
27	process_all	Process all methods with default values

`actual_evap_day`	*Calculate daily actual evaporation*

Description

Calculate daily actual evaporation

Usage

```
actual_evap_day(vtime, vlatent_heat, vtemperature = 20)
```

Arguments

`vtime`	data frame column or vector containing date data
`vlatent_heat`	data frame column or vector containing daily latent heat flux (W.m-2.day-1)
`vtemperature`	data frame column or vector containing daily average tempeature (Celcius)

Details

The calculation is based on the description in "A short course in Cloud Physics" (Rogers and Yau, 1989)

Value

Data frame containing dates (daily) and evaporation (mm/day)

Examples

```
ETa <- actual_evap_day(vtime = de_tha_d$time,
                       vlatent_heat =de_tha_d$latent_heat,
                       vtemperature = de_tha_d$temperature)
```

`alencar2021`	*FD identificaiton method proposed by the Author (Pedro Alencar)*

Description

FD identificaiton method proposed by the Author (Pedro Alencar)

Usage

```
alencar2021(
  vtime,
  vprecipitation,
  vet0,
  crit = c(4, 2, 8, 2),
  months = c(3, 10)
)
```

Arguments

`vtime`	data frame column or vector containing date data
`vprecipitation`	data frame column or vector containing daily precipitation (mm.day-1)
`vet0`	data frame column or vector containing daily potential evapotranspiration (Penma Monteith - mm.day-1)
`crit`	a vector with four elements indicating the thresholds (more in details).
`months`	a vector with two elements, the first and last months in the analysis (definition of growing season)

Details

The method tracks the variation (slope) of the precipitation and potential evapotranspiration. It uses intra-year limits and the anomaly to the time series to identify exceptionally dry events.

The `crit` vector contains 4 elements: c_1 = accumulation time for slope (in weeks) - Default = 4 c_2 = threshold for combined anomaly (sum) - Default = 2 c_3 = latency period of negative SPEI (in weeks) - Default = 8 c_4 = allows positive SPEI over latency period - Default = 2

Value

The function returns a list with two data frames. One with weekly and detailed values from the function and a second with a summary of all evetns identified.

Examples

```
fd_Alencar <- alencar2021(vtime = df_d$time,
                          vprecipitation = df_d$precipitation,
                          vet0 = ET0$et0)
```

`Christian_clean_data_week`

Function to clean the ESR data (necessary before running Christian et al. method)

Description

Function to clean the ESR data (necessary before running Christian et al. method)

Usage

```
Christian_clean_data_week(vtime, vET0, vETa, threshold = 1)
```

Arguments

vtime	data frame column or vector containing date data
vET0	data frame column or vector containing daily potential evapotranspiration (mm.d 1)
vETa	data frame column or vector containing daily actual evapotranspiration (mm.day 1)
threshold	a positive number, indicates the maximum ratio ETa/ET0 allowed.

Details

Due to data and modelling quality, eventualy ETa » ET0, causing errors, particularly in cold days when ET0 0. This function allows the reduction o distortions. Pentads discarded are interpolated.

Value

It returns a list with two elements. A data frame with time stamps and pentad ESR and a matrix with ESR data distributed in a 73 lines (pentads).

Examples

```
Christian2020_clean_data(vtime = df_d$time,
                         vET0 = ET0$et0, vETa = ETa$eta,
                         threshold = 1)
```

`Christian2020`

FD identification method from Christian et al. (2020)

Description

FD identification method from Christian et al. (2020)

Usage

```
Christian2020(esr_list)
```

Arguments

esr_list	The list output from the function `Christian2020_clean_data`

Value

The function returns a list with two data frames. One with weekly and detailed values from the function and a second with a summary of all evetns identified.

Examples

```
#' fd_Christian <- Christian2020_clean_data(vtime = df_d$time,
vET0 = ET0$et0, vETa = ETa$eta,
threshold = 2) %>% Christian2020()
```

---

conf_matrix   *Auxiliar (intern) function for calculating and exporting confusion matrix*

Description

Auxiliar (intern) function for calculating and exporting confusion matrix

Usage

```
conf_matrix(df, ref)
```

Arguments

df   complete dataset with all events (by month)

ref   the reference model - Mod. Ford and Labosier

Details

This function is supposed to be used only in the context of the FD-Viz app

Value

melted data frame to draw plot. The function return some confusion matrix common statistics: - Accuracy - Cohen's Kappa - Specificity - Sensitivity - Precision - F1

Examples

```
all_fd <- process_all(de_tha_d,include_variables = T, data = 'station')$Series

confusion_matrix <- conf_matrix(all_fd, all_fd$`Osman et al.`)
```

---

eddi_percentile   *Function to calculate EDDI in percentiles on a week accumulation time*

Description

Function to calculate EDDI in percentiles on a week accumulation time

Usage

```
eddi_percentile(vtime, vet0, dist = "tukey")
```

Arguments

vtime   a data.frame column or vector with daily time stamps (Date type)

vet0   a data.frame column or vector with ordered daily ET0 values. They can be obtained directly from reanalysis/models or from the functions penman_day or hargreaves_day

dist   string containing the distribution used to calibrate EDDI. It can either be 'tukey' following the original method from Hobbings et al (2016) or log-logist' according to Noguera et al. (2021)

Examples

```
ET0 <- penman_day(vtime = df_d$time, vwind = df_d$wind_speed,
                  vvpd = df_d$vapor_p_def, vtemp = df_d$temperature,
                  vheatflux = (df_d$sensible_heat + df_d$latent_heat))

percentiles_eddi <- eddi_percentile(vtime = de_tha_d$time, ET0$et0, dist = 'tukey')
```

---

eddi	*Calculate EDDI*

Description

Process all methods with default values

Usage

```
eddi(vet0)
```

Arguments

vet0	data frame column or vector containing daily ET0 (potential evapotranspiration). It can be obtained with the function penman_day

Details

function to calculate EDDI (Evaporative Demand Drought Index) following Hobbins et al., 2016

Value

vector with EDDI values

Examples

```
eddi_values <- penman_day(vtime = de_tha_d$time, vwind = de_tha_d$wind_speed,
                          vvpd = de_tha_d$vapor_p_def, vtemp = de_tha_d$temperature,
                     vheatflux = (de_tha_d$sensible_heat + de_tha_d$latent_heat))[,2] %>%
                          eddi()
```

---

f.anomaly	*Function to calculate anomalies*

Description

Calculates the anomaly (number of standard deviations). This is an intern function.

Usage

```
f.anomaly(vector)
```

Arguments

vector	A general vector or data frame column

Value

vector of anomalies.

---

f.pentad	*Accumulation into pentads*

Description

Internal function to accumulate data into pentads (5-day long periods) using different accumulation functions (mean, max, min, sum, etc.)

Usage

```
f.pentad(vtime, vvalue, na_rm = F, f = mean)
```

Arguments

vtime	data frame column or vector containing date data
vvalue	data frame column or vector containing the analysed data
na_rm	boolean (should NA values be removed? Defaulf = F)
f	R function to be applied (default = mean)

Value

The function return a list with two elements. One data frame with time stamped pentad values and a matrix with the 73 pentads organized in lines and years in columns.

---

f.percentile   *Percentile function*

Description

Calculates the percentile. This is an intern function.

Usage

f.percentile(vector)

Arguments

vector   data frame column or vector containing the analysed data

Value

The function return a vector with percentiles

---

f.spei   *Standard Precipitation Evaporation Index calculation*

Description

Internal funciton to calculate the SPEI

Usage

f.spei(vtime, vdeficit, n)

Arguments

vtime   a data.frame column or vector with daily time stamps (Date type)

vdeficit   a data.frame column or vector with daily hydrological deficit obtained by the
                difference of precipitation and potential evapotranspitation (P - ET0)

n   a natural number that indicates the accumulation time (pentad, week, month,
                etc)

Value

The function return a list with two elements. One data frame with time stamped pentad values and
a matrix with years organized in columns.

---

f.week   *Accumulation into weeks*

Description

Internal function to accumulate data into weeks using different accumulation functions (mean, max,
min, sum, etc.)

Usage

f.week(data.var, na_rm = F, f = mean, kind = "standard")

Arguments

data.var   data frame with time stamps

na_rm   boolean (should NA values be removed? Defaulf = F)

f   R function to be applied (default = mean)

kind   String indicating what is the size of the week. The default value is "standard"
                (conventional 7 day week). Other option is "noguera", that devides each month
                in 4 weeks (1 to 8; 9 to 15; 16 to 22; and 23 to 29/30/31)

Value

The function return a list with two elements. One data frame with time stamped pentad values and
a matrix with the 52 (or 48) weeks organized in lines and years in columns.

Description

Internal function to separate the data into years

Usage

```
f.year(i, day.var, year.var = NULL)
```

Arguments

i	a numeric index
day.var	data frame with time stamps (input)
year.var	data frame with time stamps (output - default = NULL)

Value

return a vector with all data pertaining to a year

FordLabosier_gs   *Modified method of Ford and Labosier (2017)*

Description

Besides the additional duration criteria (at least 3 pentads under 30th p.) we also added two new criteria: 1. Events with onset out of the growing season (MAMJJASO) are not considered as flash droughts 2. Evetns that are 3 pentads or less apart are considers the same event.

Usage

```
FordLabosier_gs(vtime, vswc, crit = c(40, 20, 30))
```

Arguments

vtime	data frame column or vector containing date data
vswc	data frame column or vector containing soil water content values
crit	a vector of three value (default c(20,40,30)) indicating the model thresholds for lower, upper and persistance limits for SWC percentiles.

Value

A list with two data frames, one a time series with all data for FD identification, and the second with a summary of FD events.

Examples

```
FD_events <- FordLabosier_modif(de_tha_d$time, de_tha_d$soil_water, crit = c(40,20,30))
```

FordLabosier2017   *Method of Ford and Labosier (2017)*

Description

This function follows the description contained in the original paper. We have as additional criterion that FD should have at least 3 pentads with SM lower than 30th percentile (as proposed by Dr. Ford in personal communication).

Usage

```
FordLabosier2017(vtime, vswc, crit = c(40, 20, 30))
```

Arguments

vtime	data frame column or vector containing date data
vswc	data frame column or vector containing soil water content values
crit	a vector of three value (default `c(20,40,30)`) indicating the model thresholds for lower, upper and persistance limits for SWC percentiles.

Value

A list with two data frames, one a time series with all data for FD identification, and the second with a summary of FD events.

Examples

```
FD_events <- FordLabosier2017(de_tha_d$time, de_tha_d$soil_water, crit = c(40,20,30))
```

---

get_df_era5   *get_df_era5 - Generate DF from ERA5 raw (netcedf) data*

Description

The ERA5 provides 4 soil water content layers: Layer 1: 0 - 7cm, Layer 2: 7 - 28cm, Layer 3: 28 - 100cm, Layer 4: 100 - 289cm.

Usage

```
get_df_era5(list_files, lat, lon, soil_layers = c(1))
```

Arguments

list_files	vector containing the name of file(s)
lat	latitude of study area (decimal degrees)
lon	longitude of study area (decimal degrees)
soil_layers	vector with on or more values between 1 and 4. This indicates which are the soil layers desired in the analysis. More information in the function description.

Value

The function returns a data frame with all variables of interest: Precipitation (tp) Temperature (t2m) Potential evapotranspiration (pev) actual evapotranspiration (e) volumetric soil water content (vswl1, vswl2, vswl3, vswl4)

---

get_df_fluxnet   *get_df_fluxnet - Generate DF from FLUXNET raw data*

Description

Builds data frame to be used as input in all functions from the raw FLUXNET data (.csv)

Usage

```
get_df_fluxnet(filename, timestep, soil_level = 1)
```

Arguments

filename	name of fluxnet file
timestep	what is the analysed timestep, if daily or hourly
soil_level	Which is the sensor index (depends on the station, please pay attention to the .csv file)

```
get.nc.data                 get.nc.data - Function to extract data from ERA5 raw data
```

Description

get.nc.data - Function to extract data from ERA5 raw data

Usage

```
get.nc.data(my_lon, my_lat, my_filename, vname, file = T)
```

Arguments

my_lon	Longitude of study area (decimal degrees)
my_lat	Latitude of study area (decimal degrees)
my_filename	File name
vname	Variable to be extracted
file	boolean, indicate if a csv file should be generated (default = True)

Value

The function returns a data frame containing one variable and time stamps

```
hargreaves_day              Hargreaves-Samany daily-ET0 function
```

Description

Hargreaves-Samany daily-ET0 function

Usage

```
hargreaves_day(vtime, vtemperature, my_lat)
```

Arguments

vtime	data frame column or vector containing date data
vtemperature	data frame column or vector containing temperature
my_lat	Latitude of study area

Value

The function return a data frame with two columns, time stamps (daily) and ET0

Mo2016   *FD identification based on Mo & Lettenmeier (2015, 2016)*

Description

FD identification based on Mo & Lettenmeier (2015, 2016)

Usage

```
Mo2016(
  vtime,
  vprecipitation,
  vtemperature,
  vsoil_water,
  vlatent_heat = NULL,
  vevap = NULL,
  flux_data = T
)
```

Arguments

vtime	data frame column or vector containing date data
vprecipitation	data frame column or vector containing daily precipitation
vtemperature	data frame column or vector containing daily temperature
vsoil_water	data frame column or vector containing soil water content (soil moisture)
vlatent_heat	data frame column or vector containing daily total latent heat flux (to calculate actual evapotranspiration)
vevap	data frame column or vector containing daily actual evapontranspiration (if already available)
flux_data	boolean, indicates if the data is from eddy covariance stations.

Value

The function returns a list with two data frames. One with pentad and detailed values from the function and a second with a summary of all events identified.

Examples

```
fd_Mo <- Mo2016(vtime = df_d$time, vprecipitation = df_d$precipitation,
                vtemperature = df_d$temperature, vsoil_water = df_d$soil_water,
                vlatent_heat = df_d$latent_heat, flux_data = T)
```

multicriteria_fd	*Multiple criteria method of FD classification*

Description

Multiple criteria method of FD classification

Usage

```
multicriteria_fd(
    vtime,
    vtemp,
    vprec,
    vet0,
    veta,
    score = 0.6,
    d_score = 0.2,
    thresholds = c(1, 0, 0, 0, -2, 50, 10, 30)
)
```

Arguments

vtime	data frame column or vector containing date data
vtemp	data frame column or vector containing daily temperature
vprec	data frame column or vector containing daily precipitation
vet0	data frame column or vector containing daily ET0 (potential evapotranspiration). It can be obtained with the function penman_day
veta	data frame column or vector containing daily ETa (actual evapotranspiration). It can be obtained with the function actual_evap_day
score	a number ranging from 0 to 100. The percentile above which events will be considered flash droughts.
thresholds	a vector with multiple threshold values. View details for more information. We advise not to change from the default values.

Details

The thresholds vector contains 8 elements respectively:

- The threshold used for cleaning data in Christian et al. method (Christian_clean_data_week) It indicates the maximum value accepted for the ration ETa/ET0. The default is 1.1, allowing ETa marginally higher than ET0. This reduces significatly the number o missing data.

- General threshold for anomalies. Indicates the number of standard deviations to the average for the anomalies. The default value is 1.

- SPEI threshold. The value to be considered under drought conditions. The default value is the same froim Noguera et al. (2020), i.e., -1.28

- General threshold for accumulated anomalies. Indicated the sum of weekly anomalies accumulated over a period of 4 weeks. The default value is 3.

- SESR and SPEI intensifications. The accumulated difference over 4 weeks. The default value is -2.

- EDDI intensification and recuperation. The values are the same from Pendergrass2020. Default: 50 and 10 respectively

- Delta_SESR percentile. Indicates the percentile of the difference over 4 weeks. We used the same threshold from Christian et al. 2020 for default (30th percentile.)

Value

The function returns a list with two data frames. One with weekly and detailed values from the function and a second with a summary of all events identified.

Examples

```
ET0 <- penman_day(vtime = de_tha_d$time,
                  vwind = de_tha_d$wind_speed, vtemp = de_tha_d$temperature,
                  vvpd = de_tha_d$vapor_p_def,
                  vheatflux = (de_tha_d$sensible_heat + de_tha_d$latent_heat))[,2]

ETa <- actual_evap_day(vtime = de_tha_d$time,
                       vlatent_heat =de_tha_d$latent_heat,
                       vtemperature = de_tha_d$temperature)[,2]

multi_criteria_fd <- multicriteria_fd(vtime = de_tha_d$time, vtemp = de_tha_d$temperature,
                                      vprec = de_tha_d$precipitation, vet0 = ET0, veta = ETa)
```

Noguera2020   *FD identification based on Noguera et al. (2020)*

Description

The function uses the SPEI to identify a FD. The intentisication period is defined by a 4 week (one month) period with a SPEI reduction of 2. The SPEI value must go equal or below the threshold. If threshold is not provided -1.28 (10

Usage

Noguera2020(vtime, vprecipitation, vet0, threshold = NA)

Arguments

vtime	data frame column or vector containing date data
vprecipitation	data frame column or vector containing daily precipitation
vet0	data frame column or vector containing daily ET0 (potential evapotranspiration). It can be obtained with the function penman_day
threshold	numeric, indicate the lower limit to identify a FD (see description)

Value

The function returns a list with two data frames. One with weekly and detailed values from the function and a second with a summary of all events identified.

Osman2021   *FD identification based on Osman et al. (2020)*

Description

FD identification based on Osman et al. (2020)

Usage

Osman2021(vtime, vswc, threshold = 20)

Arguments

vtime	data frame column or vector containing date data
vswc	data frame column or vector containing soil moisture data
threshold	a numeric value (default = 20) indicating the lower limit of SM percentile for FD identification

Value

The function returns a list with two data frames. One with pentad and detailed values from the function and a second with a summary of all events identified.

Examples

```
fd_Osman  <- Osman2021(vtime = df_d$time,
                       vswc = df_d$soil_water,
                       threshold = 20)
```

---

param_loglogist   *Calibrate parameters of multiple log-logist functions using Probability Weighted Moments (PWM) according to Singh (1998) "Entropu-based parameter estimation in Hydrology" (Ch. 18)*

Description

Auxiliar function or calculate SPEI and EDDI.

Usage

param_loglogist(var, n_param = 3)

Arguments

var	A matrix or data frame with the variable (hydrologic deficit for SPEI, potential evapotranspitation for EDDI). The data has to be organized with weeks (or desired interval) in rows (e.g. 52 rows) and years in columns.
n_param	Number of parameters of the Log-logist distribution (either 2 or 3).

194

Description

Identifies Flash Drougt evetns using EDDI variations, as described in Pendergrass et al. "Flash droughts present a new challenge for subseasonal-to-seasonal prediction", 2021.

Usage

```
Pendergrass2020(vtime, vet0, limit.down = 10)
```

Arguments

vtime	a data.frame column or vector with daily time stamps (Date type)
vet0	a data.frame column or vector with ordered daily ET0 values. They can be obtained directly from reanalysis/models or from the functions penman_day or hargreaves_day
limit.down	a numeric value, indicating the limit of recuperation after onset. This criterion is additional, to flexibilise the original method. We set it's value to 10 as default. TO run the original method (more restrict) define it to 0.

Details

This function is based on EDDI calculated according to Hobbings et al. (2016) DOI: 10.1175/jhm-d-15-0121.1 It uses a simple empirical Tukey plotting position to assess EDDI percentiles.

Value

Function Pendergrass2020 retuns a list with two data frames. 1) a complete time stamped series containing relevant variables and FD events; 2) a summary of each event with its duration and interval of occurance.

Examples

```
fd_Pendergrass <- Pendergrass2020(vtime = de_tha_d$time,
                                  vet0 = ET0$et0, limit.down = 10)
```

---

penman_day   *penman_day - Daily ET0 using Penman-Monteith*

Description

penman_day - Daily ET0 using Penman-Monteith

Usage

```
penman_day(vtime, vwind, vtemp, vvpd, vheatflux, altitude = 0)
```

Arguments

vtime	data frame column or vector containing date data
vwind	data frame column or vector containing daily wind velocity
vtemp	data frame column or vector containing daily temperature
vvpd	data frame column or vector containing daily vapor pressure
vheatflux	data frame column or vector containing daily total heat flux (latent and sensible heat flux)
altitude	study area altitude in meters above sea level

Value

The function returns a data frame with time stamped daily potential evapotranspiration

Examples

```
ET0 <- penman_day(vtime = de_tha_d$time, vwind = de_tha_d$wind_speed,
                  vvpd = de_tha_d$vapor_p_def, vtemp = de_tha_d$temperature,
                  vheatflux = (de_tha_d$sensible_heat + de_tha_d$latent_heat))
```

---

prepare.nc   *Prepare .nc files*

Description

Title

Usage

```
prepare.nc(data, period = "day", f = function(x) x)
```

Arguments

data	data frame with time stamps
period	string with interval of accumulation (default = 'day')
f	R function to be applied (mean, max, sum, etc)

---

process_all   *Process all methods with default values*

Description

Process all methods with default values

Usage

```
process_all(df_d, include_variables = T, data = "station")
```

Arguments

df_d	a data frame with all relevant variables. It can be obtained using the functions get_df_era5 or get_df_fluxnet
include_variables	
	Boolean,informs with all variables should be included in the output
data	string value, either 'station' or 'reanalysis'.

Details

The function run all methods with their default thresholds

Value

The function returns a list with two data frames. One with daily and detailed values for all methods and a list with one data frame for each method, each containing a summary of all identified events.

Examples

```
all_methods <- process_all(de_tha_d,include_variables = T, data = 'station')
```

Appendix E

In this appendix, we present additional publications in international conferences.

A. Alencar PHL, de Araújo JC, Paton EN. Gully-walls soil loss effect on gully modelling: a Brazilian Semiarid case of study. EGU2018 (2018)

B. Alencar PHL, de Araújo JC, Teixeira AS. Small permanent gullies: modelling and application to a semiarid region. EGU2020 (2020)

C. Alencar PHL, de Araújo JC, Paton EN. Flash Drought identification – a comparison of definitions across different datasets. EGU2021 (2021)

D. Alencar PHL, de Araújo JC, Paton EN. Entropy-based temporal downscaling of precipitation as tool for SDR assessment. MaxEnt21 - 40th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering (2021)

E. Paton EN, Alencar PHL, Vogel J, Nehls T, Kluge B. Durren und ihre Compounds: Extrema, Synchronizität un Trendverhaltend. Tag der Hydrologie (2021)

E. Appendix E

Gully-walls soil loss effect on gully modelling: a Brazilian Semiarid case of study

Pedro Henrique Lima Alencar (1), Eva Nora Paton (2), and José Carlos de Araújo (3)

(1) Universidade Federal do Ceará, CCA, Agricultural Engineering, Fortaleza, Brazil (pedrohlalencar@gmail.com), (2) Institut für Ökologie, TU Berlin, Berlin, German (eva.paton@tu-berlin.de), (3) Universidade Federal do Ceará, CCA, Agricultural Engineering, Fortaleza, Brazil (jcaraujo@ufc.br)

In the Brazilian semiarid region (a one-million km^2 area that coincides with the Caatinga biome), due to the shallow soils gully erosion processes are limited to small dimensions, being less representatives than sheet erosion, concerning total sediment yield. Nevertheless, gullies, even with small sizes, have a high influence on sedimentological processes, changing the sediment dynamics inside the watershed. Due to the land use change in the Caatinga, agricultural automation, deforestation for extensive cattle rising and development of infrastructure, as construction of new roads, gully occurrence has been becoming more frequent. This study focuses on the Madalena representative basin (124 km2, state of Ceará, Brazil), a land-reform settlement with 20 inhabitants per km^2, whose main economic activities are agriculture (especially Zea mays), livestock and fishing. Topographic surveys were performed using Total Station and UAV (Unmanned Aerial Vehicle), to obtain digital terrain models and assess the volume of soil eroded in the channel. Soil samples were collected in order to estimate their erodibility and the critic shear stress. We used the Model Foster and Lane (MFL; 1983) for ephemeral gullies to model the measured gullies, adjusting it to fit the model to the Caatinga conditions. We could observe that the MFL was not able to predict properly the geometry and area of the gully once it has as basic premise the verticality of walls, which was not observed in any of the cases. Thus, a statistical correlation was proposed in order to consider the loss of stability of the gully walls and the bank erosion. With this new factor we obtained positive responses from the modelling (NSE = 0.75; for the Cross section area), implying that the model can be applied as tool of volume prediction of gully volume in the semiarid, even for small classical gullies, once the analysed gullies began around 58 years ago, due to the construction of a new road. As secondary result, we identified that the 30-minute intensity is the most representative for the gully process, the same for sheet erosion proposed by Wischmeier and Smith (1978).

Small permanent gullies: modelling and application to a semiarid region

Pedro Henrique Lima Alencar[1], José Carlos de Araújo[2], and Adunias dos Santos Teixeira[2]
[1]Technische Universität Berlin, Institut für Ökologie, Ökohydrologie & Landschaftsbewertung, Germany
(pedro.alencar@campus.tu-berlin.de)
[2]Federal University of Ceará, Department of Agricultural Engineering, Brazil

Gullies are key drivers of land degradation, are important sources of sediment and increase sediment and pollutant connectivity in the catchment. They also play an important role in desertification areas, changing the water-table height and in farmlands, reducing productive areas. In this study, we attempted to model small permanent gullies, common in the Brazilian Semiarid Region, where the shallow soils limit the size of gullies cross-sections to a depth of no more than one meter. To model this process, we coupled the models of Foster and Lane (1983) and Sidorchuk (1999), in order to consider the effect of permanent gullies not considered in the first. Both models need as input the discharge peak and its duration, however, these data are frequently not available. We tested four different rain intensities (average, 60-minute, 30-minute and 15-minute), finding that the most intense 30 minutes represent the best the effects of the storms over gully erosion. The coupling of the two models is defined by a threshold that indicates when the equations for sidewall erosion proposed by Sidorchuk should be applied. To validate the model, we measured three gullies in the Brazilian Semiarid Region. The gullies were initiated in 1958 after the construction of a country road and have drainage area below 1 ha. The model yielded a Nash-Sutcliffe coefficient of 0.85.

Flash Drought identification – a comparison of definitions across different datasets.

Pedro Henrique Lima Alencar[1,2], José Carlos de Araújo[2], and Eva Nora Paton[1]
[1]Ecohydrology and Landscape Evaluation, Institute of Ecology, Technische Universität Berlin, Germany
[2]Agricultural Engineering, Agricultural Sciences, Universidade Federal do Ceará, Brazil

Flash droughts recently started to draw a larger curiosity to its occurrence and, therefore, its features. Differently from the slow development of droughts (months to years), flash droughts evolve over a short time (weeks) of a rapid intensification. Over the last few years, multiple methods for flash drought identification were proposed. Those methods, although sharing some characteristics, as tracking of soil water content and/or evapotranspiration (actual and potential), end up not flagging the same periods under flash drought events. We compared six well-known flash drought identification methods from the literature and used two different datasets. The datasets are: (1) the FluxNET15 dataset (Pastorello et al, 2020), a collection of worldwide, quality-controlled measurements of several hydroclimatic variables, such as soil water content, precipitation, temperature, and wind speed; and (2) the ECMWF Reanalysis 5 (ERA5 – Hersbach et al., 2019) provides over three hundred different data including soil water content in multiple levels, evapotranspiration, precipitation, and temperature. Ten stations from FluxNET15 were selected and the data from the ERA5 on the respective pixels was acquired. The aim of this work is to compare the event identification of different methods using different datasets as input (direct measures and reanalysis based).

Entropy-based temporal downscaling of precipitation as tool for sediment delivery ratio assessment

Alencar, P.H.L.[1,2], de Araújo, J.C.[2], Paton, E.N.[1]

(1) Technische Universität Berlin, Institut für Ökologie,
Berlin, Germany

(2) Universidade Federal do Ceará, Departamento de Engenharia
Agrícola, Fortaleza, Brazil

(pedro.alencar@campus.tu-berlin.de)

Abstract

Many regions around the globe are subjected to precipitation-data scarcity that often hinders the capacity of hydrological modelling. The entropy theory and the principle of maximum entropy can help hydrologists to extract useful information from the scarce data available. In this work we propose a new method to assess sub-daily precipitation features such as duration and intensity based on daily precipitation using the Principle of Maximum Entropy. Particularly in arid and semiarid regions, such sub-daily features are of central for modelling sediment transport and deposition. The obtained features were used as input to the SYPoME model (Sediment Yield using the Principle of Maximum Entropy – [1]). The combined method was implemented in seven catchments in Northeast Brazil with drainage areas ranging from 10^{-3} to 10^{+2} km^2 in assessing sediment yield and delivery ratio. The results showed significant improving when comparing with conventional deterministic modelling, with Nash-Sutcliffe Efficiency (NSE) of 0.96 and absolute error of 21% for our method against NSE of -4.49 and absolute error of 105% for the deterministic approach.

References:

[1] Araújo, José Carlos de. Entropy-based equation to assess hillslope sediment production. Earth Surface Processes and Landforms, pg 2005-2018 (2007).

Key Words: SDR; Sediment Yield; Downscaling

SESSION D2: Hydrologische Extreme / Dürren

D2-1 Dürren und ihre Compounds: Extrema, Synchronizität und Trendverhalten

Eva Paton[1], Pedro Alencar[1], Johannes Vogel[1,2], Thomas Nehls[1], Björn Kluge[1]
[1]TU Berlin, Deutschland; [2]Uni Potsdam, Deutschland; eva.paton@tu-berlin.de

Zusammenfassung

In der derzeitigen Wahrnehmung werden die Sommer länger, stärker und heißer – sowohl im ländlichen als auch im urbanen Raum, wo im dicht bebauten Gebiet Hitzeinseln den Effekt verstärken. Um das Ausmaß der Dürre bewerten zu können, wurde ein Klimadatensatz für ganz Deutschland für den Zeitraum 1950-2019 bezüglich Niederschlagsdefiziten, niederschlagsfreien Phasen, Hitzewellen und Compounds einschließlich Blitzdürren ausgewertet. Die Analyse zeigt eine große Heterogenität innerhalb von Deutschland im urbanen Raum: in den meisten Städten trat 2018 eine schwere und lange Dürre auf, gleichzeitig war das Jahr 2018 nur bei einem Drittel der Städte unter den drei Jahren mit den längsten Dürren. Bei einigen mitteldeutschen Städten kann man eine klare Zunahme an Dürremonaten verzeichnen, andere Städte eher im Norden und Nordwesten zeigen nur in den letzten 2 Dekaden eine Zunahme oder gar keinen Trend. Die Compoundanalyse lässt für die meisten Städte eine starke Zunahme erkennen, besonders in den letzten zwei Dekaden, was hauptsächlich auf die deutschlandweit verbreitete Zunahme von Hitzewellen zurückzuführen ist. Ein ähnliches Bild ergibt sich bei der Compoundanalyse von langen, niederschlagsfreien Phasen und dem gleichzeitigen Auftreten von Hitzeperioden. Bei der Untersuchung von Blitzdürren zeigte ein Ensemble-Ansatz den konsistentesten Weg ihrer Identifizierung; ihr Trendverhalten ist jedoch wegen Datenlimitierung nur bedingt bestimmbar.