TECHNISCHE UNIVERSITÄT WIEN
Vienna University of Technology

Doctoral Thesis

Hydrological extremes in Brazil: large-scale patterns, mechanisms, and change

by

Vinícius Bogo Portal Chagas

Student registration number: 12144054

submitted in satisfaction of the requirements for the degree of

Doctor of Natural Sciences
of the Vienna University of Technology,
Faculty of Civil and Environmental Engineering,
as part of the Vienna Doctoral Programme on Water Resource Systems

under the supervision of

Univ.Prof. Dipl.-Ing. Dr.techn. Günter Blöschl
Institute of Hydraulic Engineering and Water Resources Management
Vienna University of Technology

and

Assoc.Prof. Pedro Luiz Borges Chaffe
Department of Environmental and Sanitary Engineering
Federal University of Santa Catarina

Vienna, August 2023

TECHNISCHE UNIVERSITÄT WIEN
Vienna University of Technology

Dissertation

Hydrologische Extremereignisse in Brasilien: Großräumige Muster, Mechanismen und Veränderungen

von

Vinícius Bogo Portal Chagas

Matrikelnummer: 12144054

Ausgeführt zum Zwecke der Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften
eingereicht an der Technischen Universität Wien,
Fakultät für Bau- und Umweltingenieurwesen,
als Teil des Doktoratskollegs Water Resource Systems

unter der Leitung von

Univ.Prof. Dipl.-Ing. Dr.techn. Günter Blöschl
Institut für Wasserbau und Ingenieurhydrologie
Technische Universität Wien

und

Assoc.Prof. Pedro Luiz Borges Chaffe
Abteilung Umwelt- und Sanitärtechnik
Föderale Universität von Santa Catarina

Wien, im August 2023

Examiners

Univ.Prof. Dipl.-Ing. Dr.techn. Günter Blöschl
Institute of Hydraulic Engineering and Water Resources Management,
Department of Civil Engineering,
Vienna University of Technology,
Karlsplatz, 13/222, 1040 Vienna, Austria

Assoc.Prof. Dr. Alberto Viglione
Department of Environment, Land and Infrastructure Engineering,
Polytechnic University of Turin,
Corso Duca degli Abruzzi, 24, 10129 Torino, Italy

Prof. Dr. Ralf Merz
Department of Catchment Hydrology,
Helmholtz-Centre for Environmental Research,
Theodor-Lieser-Str. 4, 06120 Halle (Saale), Germany

Assoc.Prof. Dr. Rodrigo Cauduro Dias de Paiva
Institute of Hydraulic Research,
Federal University of Rio Grande do Sul,
Av. Bento Gonçalves, 9500, 91501-970, Porto Alegre, Rio Grande do Sul, Brazil

Assist.Prof. Leonardo Hoinaski
Department of Sanitary and Environmental Engineering,
Federal University of Santa Catarina,
Campus Universitário UFSC/CTC, 88040-900, Florianópolis, Santa Catarina, Brazil

Abstract

Hydrological extremes are a growing concern worldwide. However, joint studies of floods and droughts are scarce, their generating mechanisms are unknown in many regions, and the causes of their recent changes remain elusive. The goal of this thesis is to investigate how floods and hydrological droughts are generated in Brazil, to detect and attribute their changes in the last 40 years, and to examine decadal their variabilities in the last 80 years. This study focuses on large-scale, data-driven approaches and comparisons of floods and droughts. In Chapter 2, we present a new data set for large-scale hydrological studies in Brazil. It includes information on streamflow, climate, soil, geology, land cover, and human interference on the water cycle. In Chapter 3, we investigate the main mechanisms of floods by examining flood seasonalities as a function of soil moisture and intense rainfall. We find that in most of Brazil, except for the south and southeast, the timing of floods is aligned mainly with that of soil moisture peaks. This shows that floods are modulated by antecedent soil wetness rather than by event rainfall magnitude, highlighting the importance of soil water storage capacities for flood generation. In Chapter 4, we explore drought flow generating mechanisms with a process-based conceptual model, enabling a multiscale spatial analysis and reducing the impact of a subjective choice of basin attributes. Results show that basin characteristics linked to water storage dynamics control the drought flow spatial variability on regional scales (smaller than 10^6 km^2), while both basin characteristics and climate govern drought flows on continental scales (larger than 10^7 km^2). Geology is the main control on every spatial scale, particularly bedrock type and formation, pointing to a higher importance of basin water storage than climatic variability. In Chapter 5, a joint analysis of flood and drought flow change shows that the water cycle has been accelerating in 29% of Brazil in 1980-2015, with increasing flood magnitudes but decreasing drought flows. The acceleration is aligned with deforestation and increased magnitudes of intense rainfall. Additionally, 42% of Brazil experienced decreasing flood and drought flows by up to 37% per decade, driven by reduced rainfall and rising water abstraction for irrigation. Chapter 6 explores how floods and droughts cluster in time at (multi)decadal scales from 1940 to 2020. Drought-rich periods are twice as likely as flood-rich periods, which is related to rainfall variability due to anomalies in the Pacific and Atlantic oceans. The 2000s and 2010s were marked by record low flows, associated with increased water abstractions. This thesis advances research on hydrological extremes in three directions. First, by providing evidence that both floods and hydrological droughts are controlled by basin characteristics representing water storage dynamics, which attenuate the climatic variability, with further insights possible by contrasting storage capacities in the soil (linked to floods) and groundwater compartments (linked to droughts). Second, by finding a consistent signal of the acceleration of the water cycle through joint analyses of floods and droughts, providing a new perspective on future changes in hydrological extremes. Third, by showing that the most intense floods and droughts of the century are often clustered into one or two decades, although twice as frequently for droughts, providing evidence of asymmetries in their long-term variabilities. Our findings highlight that exploring floods and droughts jointly is a promising approach for understanding the generating mechanisms of hydrological extremes.

Kurzfassung

Hydrologische Extremereignisse sind weltweit ein wachsendes Problem. Studien, die Hochwasser und Dürren gemeinsam untersuchen, sind jedoch selten, ihre Entstehungsmechanismen sind in vielen Regionen unbekannt, und die Ursachen für ihre jüngsten Veränderungen sind nach wie vor nicht klar. Ziel dieser Arbeit ist es zu untersuchen, wie Hochwasser und Dürren in Brasilien entstehen, ihre Veränderungen in den letzten 40 Jahren zu bestimmen und zu erklären und die dekadischen Schwankungen der letzten 80 Jahre zu untersuchen. Diese Arbeit konzentriert sich auf großmaßstäbliche, datengestützte Ansätze und Vergleiche von Hochwasser- und Dürreperioden. In Kapitel 2 stellen wir einen neuen Datensatz für großräumige hydrologische Studien in Brasilien vor. Er enthält Informationen über Abflüsse, Klima, Boden, Geologie, Bodenbedeckung und menschliche Einflüsse auf den Wasserkreislauf. In Kapitel 3 untersuchen wir die wichtigsten Mechanismen der Hochwasserentstehung, indem wir die saisonalen Schwankungen der Hochwasser in Abhängigkeit von Bodenfeuchtigkeit und intensiven Niederschlägen untersuchen. Die Ergebnisse zeigen, dass in den meisten Teilen Brasiliens, mit Ausnahme des Südens und Südostens, der Zeitpunkt der Hochwasser im Wesentlichen mit den Spitzenwerten der Bodenfeuchte übereinstimmt. Dies zeigt, dass Hochwasser eher durch die Vorbefeuchtung als durch die Niederschlagsmengen bestimmt werden, wodurch die Bedeutung der Wasserspeicherkapazität des Bodens für die Entstehung von Hochwasser verdeutlicht wird. In Kapitel 4 untersuchen wir die Mechanismen der Bildung von Trockenwetterabflüssen mit einem neuen prozessbasierten konzeptionellen Modell, das eine räumliche Analyse auf mehreren Skalen ermöglicht und die Auswirkungen der subjektiven Wahl der analysierten Einzugsgebietsmerkmale verringert. Die Ergebnisse zeigen, dass Einzugsgebietsmerkmale in Verbindung mit der Dynamik der Wasserspeicherung die räumliche Variabilität von Trockenwetterabflüssen auf regionaler Ebene (weniger als 106 km²) bestimmen, während sowohl Einzugsgebietsmerkmale als auch das Klima die Dürreflüsse auf kontinentaler Ebene (mehr als 107 km2) bedeutend sind. Die Geologie ist auf allen räumlichen Skalen der wichtigste Einflussfaktor, insbesondere die Art und Formation des Grundgesteins, was darauf hindeutet, dass die Wasserspeicherung in den Einzugsgebieten wichtiger ist als die klimatische Variabilität. In Kapitel 5 zeigt die gemeinsame Analyse der Veränderungen der Abflüsse bei Hochwasser und Trockenheit, dass sich der Wasserkreislauf in 29 % der brasilianischen Landesfläche im Zeitraum 1980-2015 beschleunigt hat, wobei die Hochwasser zunahmen, die Trockenwetterabflüsse jedoch abnahmen. Die Beschleunigung hängt mit der Zunahme des maximalen jährlichen Niederschlags und des Verlustes von Waldflächen. Darüber hinaus kam es in 42 % Brasiliens zu einem Rückgang der Hochwasser - und Trockenwetterabflüsse um bis zu 37 % pro Jahrzehnt, was auf geringere Niederschläge und eine steigende Wasserentnahme für die Bewässerung zurückzuführen ist. In Kapitel 6 wird untersucht, wie sich Hochwasser und Dürren im Zeitraum von 1940 bis 2020 auf dekadischer bis multidekadischer Skala häufen. Dürreperioden sind doppelt so häufig wie Hochwasserperioden, was vor allem mit der Variabilität der Niederschläge aufgrund von Anomalien im Pazifik und Atlantik zusammenhängt. Die 2000er und 2010er Jahre waren durch besonders niedrige Abflüsse gekennzeichnet, die mit einer erhöhten Wasserentnahme zusammenhängen. Diese Arbeit liefert drei

wesentliche Forschungsbeiträge zu hydrologischen Extremen. Erstens wird der Nachweis erbracht, dass sowohl Hochwasser als auch hydrologische Dürren durch die Merkmale von Einzugsgebieten in Verbindung mit der Wasserspeicherdynamik bestimmt werden, die sich der klimatischen Variabilität überlagert. Weitere Einblicke sind durch die Unterscheidung der Speicherkapazitäten in der Bodenkomponente (in Verbindung mit Hochwasser) und der Grundwasserkomponente (in Verbindung mit Dürren) möglich. Zweitens wird durch die gemeinsame Analyse von Hochwasser und Dürren ein konsistentes Signal der Beschleunigung des Wasserkreislaufs gefunden, das eine neue Sichtweise auf die möglichen künftigen Veränderungen der hydrologischen Extreme ermöglicht. Drittens wird gezeigt, dass Häufungen von Hochwassern und Dürren oft in Perioden von ein oder zwei Jahrzehnten auftreten, wobei Dürren doppelt so häufig vorkommen wie Hochwasser, ein Hinweis auf Asymmetrien in ihren langfristigen Schwankungen. Die Ergebnisse belegen auch, dass die gemeinsame Untersuchung von Hochwasser und Dürren ein vielversprechender Ansatz zum besseren Verständnis der Entstehungsmechanismen hydrologischer Extreme ist.

Funding information

This work was funded by the Brazilian National Council for Scientific and Technological Development (CNPq), grant number 141219/2019-0.

Contents

Chapter 1

Introduction

Floods and hydrological droughts are opposite extremes of river flows, yet both conditions have been a growing concern in many regions (Gudmundsson et al., 2021; Milly et al., 2015; Slater et al., 2021). For example, the coast of São Paulo in Brazil witnessed its largest floods in 2022 due to more than 700 mm of rainfall in less than 24 hours. In contrast, the same area, along with much of Brazil, has experienced record low flows in the past decade and a partial disruption of hydropower generation (Tomasella et al., 2022). Floods and droughts are usually investigated separately, but here we argue that contrasting differences in their generating mechanisms can further our understanding of hydrological functioning of river basins and how such hazards have been changing over time.

The generation of river floods depends on event precipitation, antecedent soil wetness, and snowmelt (Merz & Blöschl, 2003; Rosbjerg et al., 2013). The relative importance of each factor varies regionally as a function of climate and basin characteristics, such as soil properties, geology, and terrain morphology. For instance, in mountainous regions like central Europe and the eastern U.S., floods are linked to annual peaks in snowmelt and precipitation (Berghuijs, Harrigan, et al., 2019; Kemter et al., 2020; Stein et al., 2020). In colder climates, such as north-eastern Europe and northern U.S., floods are aligned with the timing of the warm season and snowmelt (Blöschl et al., 2017; Brunner et al., 2020; Kemter et al., 2020). In temperate flatlands of western Europe and central-eastern U.S., floods are aligned with soil moisture peaks that result from gradual precipitation accumulation over several months (Berghuijs, Harrigan, et al., 2019; Stein et al., 2020). In Brazil, floods are often examined in relation to extreme rainfall linked with phenomena like the El Niño-Southern Oscillation or mesoscale convective systems (e.g., Cavalcanti, 2012; Lima et al., 2017; Marengo & Espinoza, 2016; Sena et al., 2012). However, an explicit analysis of how basin characteristics and water storage dynamics regulate floods in Brazil has not yet been conducted on a large-scale, leaving the role of event rainfall and antecedent soil moisture unclear. Considering that tropical climates typically experience greater rainfall magnitudes than temperate and cold climates (Kidd & Huffman, 2011), often exceeding 1,500 mm yr^{-1}, a high relative importance of soil moisture for flood generation would suggest an exceptional water storage capacity and attenuation of intense rainfall events.

Drought flows, the opposite hydrological condition, result from the balance of climatic variability and basin characteristics although on longer timescales and flow paths (Laaha et al., 2013; Smakhtin, 2001). Climate determines potential groundwater recharge volumes and the length of dry spells, whereas basin characteristics govern the active water storage capacity that sustains streamflows in the dry season. However, the relative importance of each component for regulating drought flows is even more contradictory than for floods. Studies determining the predominance of climate generally involve large-scale analyses worldwide (e.g., Addor et al., 2018; Apurv & Cai, 2020; Beck et al., 2013, 2015). These suggest that most of the spatial variability in low flows is regulated by the aridity index (i.e., the ratio of mean annual precipitation to evaporation). Conversely, studies emphasizing a predominance of basin characteristics are

often based on local or regional scales (e.g., Bloomfield et al., 2009; Carlier et al., 2018; Floriancic et al., 2022; Pfister et al., 2017). These link low flows with an array of variables. Examples include basin fraction of colluvium, which promotes water removal and reduced low flows (Price, 2011); the age of the geological formation, implying deeper groundwater systems and increased low flows (Tague & Grant, 2004); and the fraction of permeable bedrocks, leading to higher water storage and increased low flows (Carlier et al., 2018; Floriancic et al., 2022). Thus, further understanding of drought flow generating mechanisms needs to capture the effects of the scale of analysis. For this, a new strategy is to conduct a multiscale analysis less sensitive to the choice of basin attributes, using conceptual models with components tied to hydrological processes like recharge, storage, and release.

An understanding of how hydrological extremes are generated can improve the detection and attribution of flood and drought changes. The spatiotemporal patterns of hydrological extremes have been changing considerably worldwide (Gudmundsson et al., 2021). Some models suggest joint increases in flood and drought magnitudes (Allan et al., 2020; Ward et al., 2020), a phenomenon termed the acceleration of the terrestrial component of the water cycle. However, evidence for such joint changes remains limited due to insufficient data, the confounding influence of human factors, and a reduced number of studies in many regions. In Brazil, local and regional streamflow changes have been linked with climate change (Barichivich et al., 2018), reservoir construction (Stevaux et al., 2009), and deforestation (Costa et al., 2003). Yet, a comprehensive large-scale analysis with multiple hypothesis testing has not been conducted, which provides an opportunity to investigate flood, drought, and joint changes.

Analyzing changes over a few decades is crucial due to the large data availability, but on larger timescales, streamflow often exhibits persistent decadal to multidecadal variabilities (Blöschl et al., 2020; Kundzewicz et al., 2019; Markonis & Koutsoyiannis, 2016). Variabilities include step changes and temporal clustering into flood-rich and drought-rich periods (Blöschl, Bierkens, et al., 2019; Lun et al., 2020), when extreme events cluster within one or two decades instead of being sparsely distributed over a century. Such persistence is particularly prominent in the southern hemisphere and, in South America, has been linked to long-term sea surface temperature anomalies at both centennial and millennial time scales (O'Connell et al., 2022). However, it is unclear how flood-rich and drought-rich periods are linked with one another and, in particular, if they are symmetric in terms of frequency and spatial extent.

The study of large-scale patterns of floods and droughts in Brazil is restricted by limited access to a comprehensive, open-source hydroclimatic data set. Streamflow, precipitation, and evaporation data require web-scraping techniques for access, followed by preprocessing, quality checks, and computation of relevant indices for river basins. These challenges hinder large-scale hydrological research in Brazil, leading to fragmented knowledge and often restricted to regions like Amazonia or the Paraná basin (e.g., Chagas & Chaffe, 2018; Levy et al., 2018; Melo et al., 2016; Tomasella et al., 2011). Establishing a comprehensive and easily accessible data set is the first step for new understanding of flood and drought generating mechanisms and their changes. Brazil includes a wide range of geological formations; soil types typically found in the tropics; typical tropical, temperate, and semiarid climates; and large river basins up to several million squared kilometers, such as the Amazon, Tocantins-Araguaia, São Francisco, and Paraná basins. However, the mechanisms in which this physiographic diversity regulates droughts and floods, the particularities of each region or large river basin, and the differences to other continents are not yet fully understood and present an opportunity for further hydrological studies.

The goal of this thesis is to investigate how hydrological extremes in Brazil are generated and how their spatiotemporal patterns have been changing over the past century. The focus is on data-driven, large-scale analyses. The contribution to research on hydrological extremes mainly

comes from contrasting floods and droughts patterns, considering the effects of multiple scale of analysis, human interference, and analyzing comprehensive data sets in an under-researched region. More specific goals are:

(i) To investigate the generating mechanisms of flood and drought flows, particularly the role of climate in contrast with that of basin characteristics.

(ii) To detect and attribute changes in flood and drought flow magnitudes in recent decades, including joint changes.

(iii) To examine decadal to multidecadal flood and drought persistence in the past century, especially time clustering and asymmetries between both extremes.

This thesis is organized as follows. The next chapters present scientific articles, including those already published (Chapters 2 to 4), under review (Chapter 5), and under preparation (Chapter 6). Chapter 2 introduces a new open data set for large-scale studies in Brazil, detailing the spatial distribution of a broad range of indices from hydrometeorology to basin characteristics and human interference. In Chapters 3 and 4, we investigate how flood and drought flows are generated respectively. Chapter 5 presents the detection of flood and drought flow change in the past four decades, including joint changes and their main causes. Chapter 6 extends the analysis of flood and drought change by examining time clustering at decadal to multidecadal scales over the past eight decades, including differences between both extremes. Finally, Chapter 7 summarizes the findings and conclusions of the thesis.

Chapter 2

CAMELS-BR: Hydrometeorological time series and landscape attributes for 897 catchments in Brazil

This chapter presents the following publication in its original form:

Chagas, V. B. P., Chaffe, P. L. B., Addor, N., Fan, F. M., Fleischmann, A. S., Paiva, R. C. D., & Siqueira, V. A. (2020). CAMELS-BR: hydrometeorological time series and landscape attributes for 897 catchments in Brazil. Earth System Science Data, 12(3), 2075–2096. https://doi.org/10.5194/essd-12-2075-2020

Abstract

We introduce a new catchment dataset for large-sample hydrological studies in Brazil. This dataset encompasses daily time series of observed streamflow from 3679 gauges, as well as meteorological forcing (precipitation, evapotranspiration, and temperature) for 897 selected catchments. It also includes 65 attributes covering a range of topographic, climatic, hydrologic, land cover, geologic, soil and human intervention variables, as well as data quality indicators. This paper describes how the hydrometeorological time series and attributes were produced, their primary limitations and their main spatial features. To facilitate comparisons with catchments from other countries, the data follow the same standards as the previous CAMELS (Catchment Attributes and MEteorology for Large-sample Studies) datasets for the United States, Chile, and Great Britain. CAMELS-BR complements the other CAMELS datasets by providing data for hundreds of catchments in the tropics and the Amazon rainforest. Importantly, precipitation and evapotranspiration uncertainties are assessed using several gridded products and quantitative estimates of water consumption are provided to characterize human impacts on water resources. By extracting and combining data from these different data products and making CAMELS-BR publicly available, we aim to create new opportunities for hydrological research in Brazil and to facilitate the inclusion of Brazilian basins in continental to global large-sample studies. We envision that this dataset will enable the community to gain new insights into the drivers of hydrological behavior, better characterize extreme hydroclimatic events, and explore the impacts of climate change and human activities on water resources in Brazil. The CAMELS-BR dataset is freely available at https://doi.org/10.5281/zenodo.3709337 (Chagas et al., 2020).

2.1 Introduction

Large-scale hydrological research relies on data from large samples of catchments to formulate general conclusions on hydrological processes and models (Addor et al., 2019; Gupta et al.,

2014). Hydrometeorological datasets with large spatial and temporal coverage are the basis to improve hydrological understanding with appropriate statistical robustness. For example, multiple studies used large-sample datasets to investigate the drivers of hydrological change (e.g., Blöschl et al., 2019; Gudmundsson et al., 2019; Slater et al., 2015), the impacts of anthropic activities on the water cycle (e.g., Hoekstra & Mekonnen, 2012; Milliman et al., 2008; Montanari et al., 2013), hydrological similarity and classification (e.g., Berghuijs et al., 2014; Knoben et al., 2018; Sawicz et al., 2014), predictions in ungauged basins (e.g., Ehret et al., 2014; Singh et al., 2014; Yadav et al., 2007), areas where extreme events are a concern (e.g., Van Lanen et al., 2013; Villarini, 2016; Woldemeskel & Sharma, 2016), and to predict future hydrological change (e.g., Luke et al., 2017; Zscheischler et al., 2018). Moreover, large-sample hydrology is needed for evaluation of continental to global hydrological models, to identify limitations in model structure, parameterization and forcing according to geographic and climatic regions (Beck, Van Dijk, et al., 2017; Gudmundsson et al., 2012; Haddeland et al., 2011; Siqueira et al., 2018; Veldkamp et al., 2018; Zhao et al., 2017), to estimate uncertainty in model estimates (e.g., Beck et al., 2016; Hirpa et al., 2018; Müller Schmied et al., 2014) and to make use of data assimilation techniques (e.g., Wongchuig et al., 2019). Better predictions in such models allow for the quantification of water resources availability over large scales and are fundamental for nationwide water resources planning and management (Alfieri et al., 2020; Bierkens, 2015; Döll et al., 2016; Schewe et al., 2014).

To uncover the hydrological functioning of a catchment, it is key to understand how it is controlled by climate, human interferences (Di Baldassarre et al., 2018; Wohl et al., 2012), and landscape attributes, such as vegetation, topography, soil, and lithology (Fan et al., 2019). For this, researchers must work with a multiple-hypotheses framework (Merz et al., 2012; Pfister & Kirchner, 2017), which frequently leads to processing massive amounts of data and often tedious, repetitive tasks. To deal with this problem, multiple datasets have been created, such as the Global Runoff Data Centre (GRDC, 2019), the Global Streamflow Indices and Metadata Archive (GSIM; Do et al., 2018; Gudmundsson et al., 2018); the HydroATLAS (Linke et al., 2019); and the Global Runoff Reconstruction (FRUN; Ghiggi et al., 2019). A noteworthy dataset, the Catchment Attributes and MEteorology for Large-sample Studies (CAMELS; Addor et al., 2017; Newman et al., 2015), produced a synthesis from multiple catchment attributes. It initially included only catchments in the United States, but later expanded to include Chile with the CAMELS-CL dataset (Alvarez-Garreton et al., 2018) and Great Britain with CAMELS-GB (Coxon et al., 2020). The CAMELS datasets facilitated hydrological research by addressing some of the major problems with large datasets, such as a lack of common standards across databases, absence of uncertainty estimation, and open accessibility of hydrological observations (Addor et al., 2019).

Even though there is a growing number of large-sample datasets worldwide (e.g., Addor et al., 2017; Do et al., 2018; Ghiggi et al., 2019; Gudmundsson et al., 2018; Linke et al., 2019), the access to open and readily available data in some regions like South America is still difficult and requires additional quality checks (Crochemore et al., 2019). Particularly in Brazil, large-sample hydrological studies lack a comprehensive dataset to rely on. Brazilian hydrometeorological information is currently collected, maintained and distributed by institutions such as the Brazilian National Water Agency (ANA – *Agência Nacional de Águas*; http://www.snirh.gov.br/hidroweb/) and the National Institute of Meteorology (INMET; http://www.inmet.gov.br/portal/). The creation of ANA in 2000 led to the release of open hydrological information, which prompted the growth of hydrological studies and fostered water resources management. However, the use of data provided for Brazilian catchments is challenging because (i) it requires either manual data

acquisition of one station at a time through the institutions' local repositories (e.g., ANA, 2019a) or web scraping techniques to access these data in an automated fashion, (ii) there is little consistency in data format across regions and stations, and (iii) current datasets do not systematically provide catchment attributes characterizing the hydroclimate, landscape, and anthropogenic influences. Further, the difficulty of accessing national meteorological daily time series has led users to compute them from other gridded global databases (e.g., Beck, Vergopolan, et al., 2017; Sun et al., 2018; Xavier et al., 2016). All these difficulties hinder large-sample hydrological studies in Brazil where, unsurprisingly, nationwide studies (e.g., Bartiko et al., 2019; Siqueira et al., 2018) are less common than in North America or Europe. Consequently, studies in Brazil generally include only a reduced number of stream gauges and catchment attributes, and are restricted to specific regions, such as the Amazon (e.g., Latrubesse et al., 2017; Levy et al., 2018; Paiva et al., 2013; Tomasella et al., 2011), or the La Plata basin (e.g., Chagas & Chaffe, 2018; Collischonn et al., 2001; Lima et al., 2017; Melo et al., 2016; Pasquini & Depetris, 2007).

To overcome these limitations, we produced and made publicly available a new dataset for large-sample hydrological studies in Brazil, CAMELS-BR. It includes daily streamflow time series from 3,679 stream gauges and, for a selected group of 897 catchments, daily meteorological time series and 65 catchment attributes from properties such as topography, climate, land cover, geology, soil, and human intervention. All catchment attributes and time series are in an easily readable file format and on a quickly accessible database. We follow standards defined by the previous CAMELS and CAMELS-CL datasets, thus allowing direct comparisons with them. Most attributes rely on data products that cover the whole of South America, so they are spatially consistent across Brazil. To reduce the risk of data misinterpretation, we describe the major limitations of the data sources and indices computed. By synthesizing hydrological information from thousands of catchments in Brazil into a single dataset, we allow researchers to skip the arduous task of collecting and preprocessing large quantities of disparate data.

2.2 Hydrometeorological daily time series

2.2.1 Streamflow data

We provide daily streamflow time series for two sets of gauges (Table 2.1). The first set comprises 3679 streamflow gauges and is provided by the Brazilian National Water Agency (ANA, 2019a). We refer to these as "raw streamflow" time series, as they are readily available from (ANA, 2019a). Their values are unchanged but, to ease their processing, we converted the native files (i.e., Excel files with daily streamflows not disposed in chronological order) to a new file format (i.e., text files with daily streamflow in chronological order). ANA estimates daily streamflow either by (i) taking two daily stream stage measurements, one in the morning (at 7 am) and another in the afternoon (at 5 pm), which are averaged and transformed into discharge using a stage-discharge relationship (rating curve); or (ii) resorting to regionalization methods when no stream stage measurements are available (no further details on the methods are provided by ANA). The raw streamflow time series cover different periods, ranging from a few days to more than a century. Additionally, although ANA performs data quality checks, these time series include inconsistencies such as typographical errors and days with missing data. The 3679 gauges are irregularly distributed throughout the country (Fig. 2.1a). Overall, their spatial distribution is denser and their time series longer in the Southern Atlantic, Southeastern Atlantic, and Paraná hydrographic regions (Fig. 2.1a and 2.1b).

Table 2.1. Summary of the data provided by CAMELS-BR.

	Variable	Description	N. of catchments
Daily time series	Raw streamflow	As read from the Brazilian National Water Agency[a], in $m^3 s^{-1}$, with varying coverage periods, with data quality flags (qual_control_by_ana and qual_flag) for each time step	3679
	Streamflow	Transformed to mm day^{-1}, covering from 1980-2018, with data quality flags (qual_control_by_ana and qual_flag) for each time step	897
	Precipitation	Catchment averages (mm day^{-1}) using three products: CHIRPS v2.0 (0.05º res., covering from 1981-2018)[b]; CPC Global Unified (0.5º res., 1980-2018)[c]; and MSWEP v2.2 (0.1º res., 1980-2016)[d].	897
	Potential evapotranspiration	Catchment averages (mm day^{-1}) using GLEAM v3.3a (0.25º res., 1980-2018)[e]	897
	Actual evapotranspiration	Catchment averages (mm day^{-1}) using two products: GLEAM v3.3a (0.25º res., 1980-2018)[e]; and MGB (0.5º res., 1980-2014)[f]	897
	Minimum, maximum, and average temperature	Catchment averages (degrees Celsius day^{-1}) using CPC Global Unified (0.5º res., 1980-2018)[g]	897
Catchment attributes	Location	5 attributes (Table 2.2)	3679
	Topography	4 attributes (Table 2.2)	897
	Data quality checks	2 attributes (Table 2.2)	897
	Climatic indices	13 attributes (Table 2.3)	897
	Hydrological signatures	13 attributes (Table 2.4)	897
	Land cover characteristics	11 attributes (Table 2.5)	897
	Geologic characteristics	7 attributes (Table 2.6)	897
	Soil characteristics	6 attributes (Table 2.7)	897
	Human intervention indices	4 attributes (Table 2.8)	897

Notes. [a] ANA (2019a). [b] Climate Hazards Group InfraRed Precipitation with Station data v2.0 (Funk et al., 2015). [c] Climate Prediction Center Global Unified Gauge-Based Analysis of Daily Precipitation (NOAA, 2019b). [d] Multi-Source Weighted-Ensemble Precipitation (Beck et al., 2019). [e] Global Land Evaporation Amsterdam Model v3.3a (Martens et al., 2017; Miralles et al., 2011). [f] Large-Scale Hydrological Model (Siqueira et al., 2018). [g] Climate Prediction Center Global Daily Temperature (NOAA, 2019a).

The second set of streamflow time series includes 897 gauges, and here we simply refer to them as "streamflow" time series (Table 2.1). This is the set of gauges used to compute the catchment attributes. It is a subset from the previous 3679 gauges, which resulted from two selection criteria. Firstly, we selected only gauges that have less than 5% of missing streamflow data between the water years 1990 (starting on September 1, 1989) and 2009 (ending on August 31, 2009). We chose the water years from 1990 to 2009 because (i) it is the period with the largest number of stream gauges with available data (Fig. 2.2), and (ii) it coincides with the period of analysis from other CAMELS datasets (Addor et al., 2017; Alvarez-Garreton et al., 2018), allowing for direct comparisons with them. Secondly, we only considered catchments for which boundaries have been delimited by (Do et al., 2018) and for which there is a good match with the area estimated by the data provider (see Section 2.3). Although the hydrological signatures introduced below were computed using data from 1990 to 2009, the time series for the 897 stream gauges include data from 1980 to 2018 when available, to enable complementary analyses by other users.

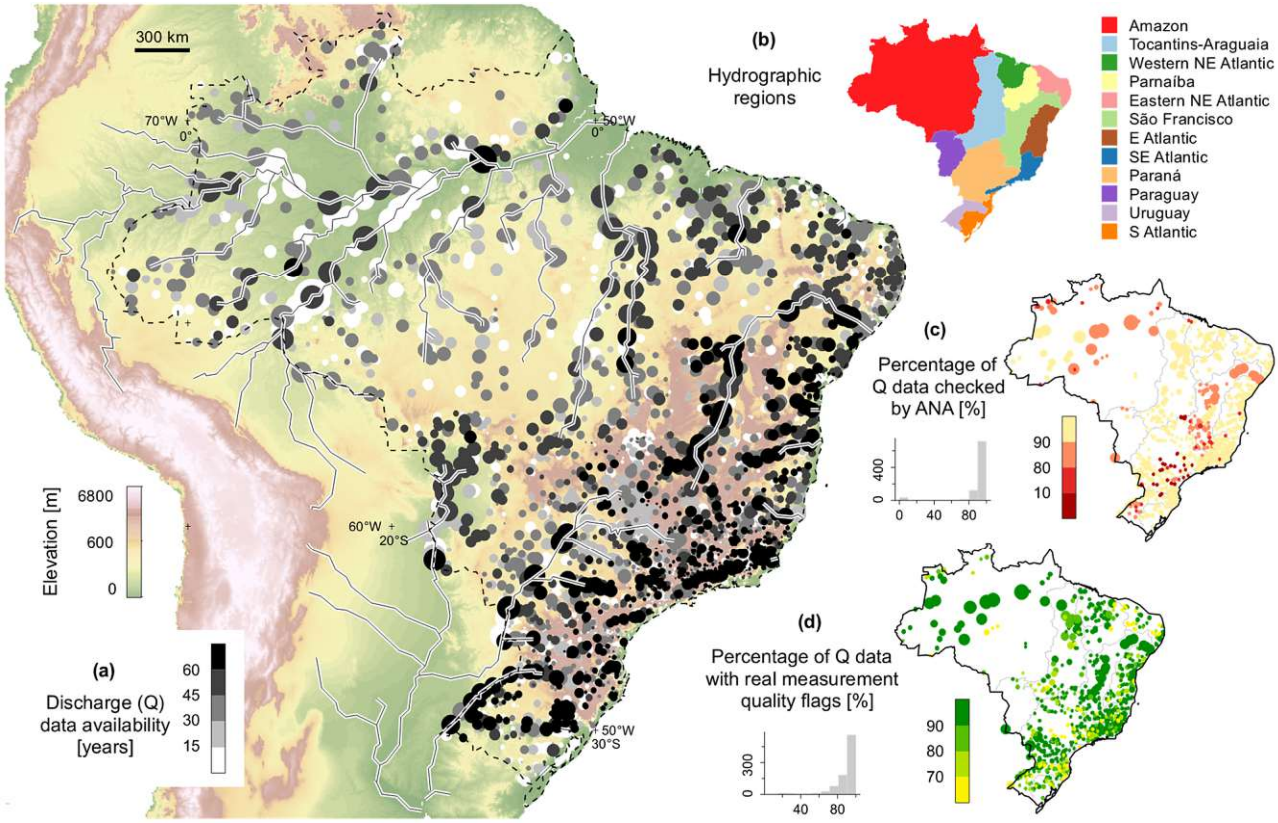

Figure 2.1. (a) South America and the total river discharge data availability of the 3679 stream gauges included in this study. The black line surrounded by a white line indicates rivers. The dashed line is Brazil's borders. (b) Hydrographic regions of Brazil according to ANA (2019a). (c) Percentage of streamflow data with quality control checks by ANA of the 897 selected catchments. (d) Percentage of streamflow data derived from stream stage measurements of the 897 selected catchments. The circles are located at the outlet of the catchments and their sizes are proportional to the sizes of the catchments. The grey line in (c) and (d) indicates the limits of hydrographic regions.

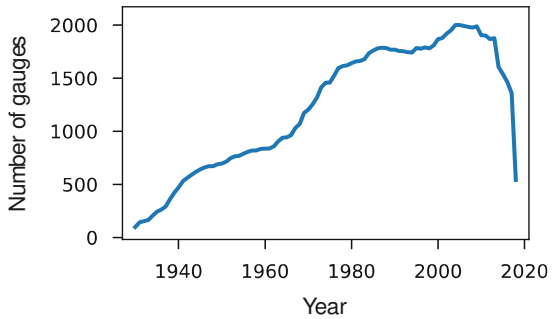

Figure 2.2. Time series with the number of streamflow gauges with at least one measurement for a given year in Brazil.

We individually screened the 897 selected streamflow time series between 1990 and 2009 for the following errors: zeroes or repeated values instead of missing values, abrupt changes resulting from changes in measurement instruments or rating curves, annual streamflow larger than annual precipitation, and unrealistic daily streamflow values (i.e., larger than 1,000

mm day^{-1}). Gauges affected by such errors were not included in the set of 897 catchments. In addition, we summarized the streamflow metadata provided by ANA as follows. For each daily streamflow measurement, we provide two pieces of information (Table 2.1). The first metadata variable, "qual_control_by_ana", was set to 1 if the data was quality checked by ANA and to 0 otherwise. The second metadata variable, "qual_flag", indicates the reliability of streamflow estimates. It is also provided by ANA and consists of the following quality flags: 0, when there is no description; 1, streamflow resulted from stream stage measurements and the rating curve; 2, streamflow estimated by ANA without stream stage measurements; 3, streamflow values marked as doubtful; and 4, when the stream water level falls outside the range of the stream stage, e.g., when the river ran dry. To summarize the ANA metadata (i.e., q_qual_control_perc and q_stream_stage_perc; Table 2.2), 80% of the 897 gauges had at least 90% of their data over 1990-2009 checked for inconsistencies (Fig. 2.1c). The Amazon, São Francisco, and Paraná regions have the lowest frequency of quality controls in Brazil. Furthermore, the streamflow estimates from 64% of the 897 catchments were derived from stream stage measurements for 90% of the days over 1990-2009 (Fig. 2.1d).

2.2.2 Meteorological data

Meteorological daily time series data are provided for 897 catchments (Table 2.1). These include (i) precipitation from CHIRPS v2.0 (Funk et al., 2015), CPC (NOAA, 2019b), and MSWEP v2.2 (Beck et al., 2019); (ii) potential evapotranspiration from GLEAM v3.3a (Martens et al., 2017; Miralles et al., 2011); (iii) actual evapotranspiration from GLEAM v3.3a and MGB South America (Siqueira et al., 2018); and (iv) minimum, maximum, and average temperature from CPC (NOAA, 2019a). The datasets were selected because of their high spatial resolution, their full coverage of South America allowing for consistency through all catchments, and because they are commonly used which enables comparisons with other studies. The daily values represent the average of all cells with their centroids intersected by the catchment, of which all cells contribute to the average equally, whether the catchment fully covers them or not. However, some catchments do not intersect the centroid of any cell. For those, we computed the daily values as the average of all cells partially covered by the catchment. A significant limitation of the meteorological data is that, because the cell grids of the adopted products have resolutions range from 0.05º (ca. 5.5 km^2 at the equator) to 0.5º (ca. 55 km^2 at the equator), some catchments are smaller than a single cell. This leads to the assumption that such a meteorological variable is homogeneous in catchments smaller than a single cell, even though this might not always be the case. This limitation has to be kept in mind particularly when using the CPC precipitation data (resolution of 0.5º; NOAA, 2019b), as precipitation is the meteorological variable with the highest spatial heterogeneity amongst those used in CAMELS-BR.

In addition to GLEAM v3.3a, estimates of actual evapotranspiration (ET) were obtained from the MGB model version for South America (Siqueira et al., 2018). The MGB is a conceptual, semi-distributed hydrologic-hydrodynamic model that discretizes the basin (or a set of basins) into irregular unit-catchments and further into hydrological response units by combinations of land use and soil types, where both water and energy balance are computed. The model calculates ET using the Penman-Monteith equation based on CRU meteorological data (i.e., temperature, pressure, radiation, and wind speed) and MSWEP v1.1 precipitation data (Beck, Vergopolan, et al., 2017). Surface resistance is adjusted according to the availability of water in the soil that is updated during the water budget. The MGB also computes the evaporation of flooded

areas and intercepted water from the canopy with the Penman equation. Regular ET cells of 0.5º resolution were generated by aggregating unit-catchments using their areas as weights.

The long-term water balance is accurate for most catchments, using either the estimated evapotranspiration from GLEAM (Fig. 2.3a) or MGB (Fig. 2.3b). Both evapotranspiration data sources indicate that the highest data uncertainties occur in the Amazon and smaller catchments in the Paraná and the Southeastern Atlantic regions since those catchments are further away from the 1:1 line in Fig. 2.3a-b. The same conclusions are derived from visualizing the runoff coefficient as a function of the humidity index (Fig. 2.3c). In addition, there are remarkable differences between GLEAM and MGB estimates, where evapotranspiration from GLEAM is substantially higher in the Amazon basin and substantially lower in the Eastern and the Western NE Atlantic regions.

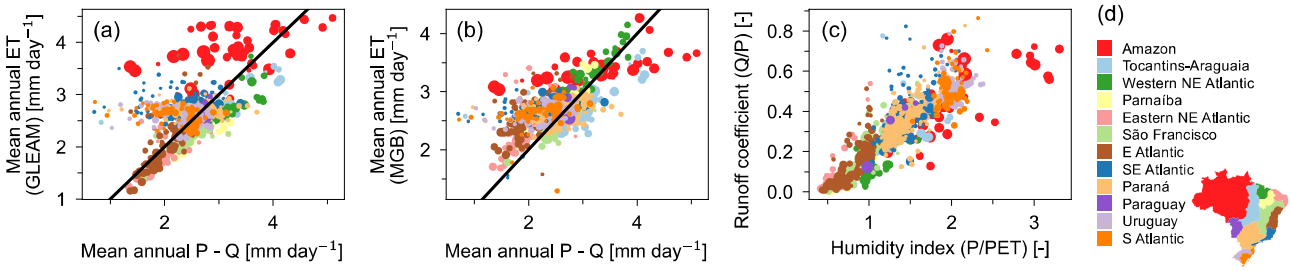

Figure 2.3. Long-term water balance of the 897 selected catchments for the water years 1990-2009. Mean annual evapotranspiration from (a) GLEAM or (b) MGB as a function of the difference between mean annual precipitation (P) from CHIRPS and streamflow (Q). (c) Runoff coefficient as a function of the humidity index. Line 1:1 is shown in black. The symbol size is proportional to the catchment area. The symbol color indicates the hydrographic region of the catchment (from panel d).

2.3 Topographic indices

Even though ANA (2019a) provides estimates of the areas of most gauged Brazilian catchments, the catchment boundaries are not publicly available. Hence, in this study, we used the catchment boundaries provided by Do et al. (2018), who used the HydroSHEDS 15 arc-sec resolution digital elevation model (DEM) and delineated the catchments with a procedure similar to (Lehner, 2012) for more than 3,000 gauges in Brazil. For each streamflow gauge, Do et al. (2018) positioned the outlet at the center of all the DEM grid cells within a radius of 5 km from the gauge coordinates indicated by the metadata. They then selected the grid cell (and associated catchment boundaries) leading to the catchment area most similar to the one indicated by ANA (2019a). The main limitation of the procedure of Do et al. (2018) is that catchment boundaries were not manually inspected.

Using those catchment boundaries, we computed four topographic attributes (Table 2.2), namely gauge elevation, catchment mean elevation, mean slope, and area. The area of the catchments ranged from 10.8 km^2 (i.e., in the upper São Francisco hydrographic region) to 4.7 million km^2 (i.e., the Amazon basin at Óbidos). Approximately 30% of the analyzed catchments are smaller than a thousand km^2, 43% are between 1 and 10 thousand km^2, and 27% are larger than 10 thousand km^2. The largest basins are in the Amazon and the Tocantins-Araguaia hydrographic regions (Fig. 2.4a). Combined with the Paraguay basin, those regions are usually characterized by low elevations (Fig. 2.4b), flat slopes (Fig. 2.4c), and large proportions of wetlands

(see Section 2.6.2). The smaller catchments are located along the mountain belts on the eastern coast of Brazil, particularly in the southern and southeastern parts of the country. Those are also the catchments with the steepest slopes. Additionally, many catchments with intermediate elevation ranges (i.e., between 500 and 900 m) are in the central part of the country, which comprises the Brazilian highlands. Note that, since we computed the average attribute value (unless otherwise noted) of each catchment, the attributes become less representative as the area of the catchment increases.

Table 2.2. Location, topographic characteristics, and data quality checks.

Attribute	Description	Units	Data source
gauge_id	Catchment identifier provided by ANA	-	ANA (2019a)
gauge_name	Gauge name provided by ANA	-	ANA (2019a)
gauge_region	Hydrographic region	-	ANA (2019a)
gauge_lat	Gauge latitude	ºN	ANA (2019a)
gauge_lon	Gauge longitude	ºE	ANA (2019a)
elev_gauge	Gauge elevation	m.a.s.l.	HydroSHEDS 15 arc-sec DEM
elev_mean	Catchment mean elevation	m.a.s.l.	HydroSHEDS 15 arc-sec DEM
slope_mean	Catchment mean slope	m km^{-1}	HydroSHEDS 15 arc-sec DEM
area	Catchment area	km²	Do et al. (2018)
q_quality_control_perc	Percentage of streamflow data (1990-2009) with quality control checks by ANA	%	ANA (2019a)
q_stream_stage_perc	Percentage of streamflow data (1990-2009) derived from stream stage measurements	%	ANA (2019a)

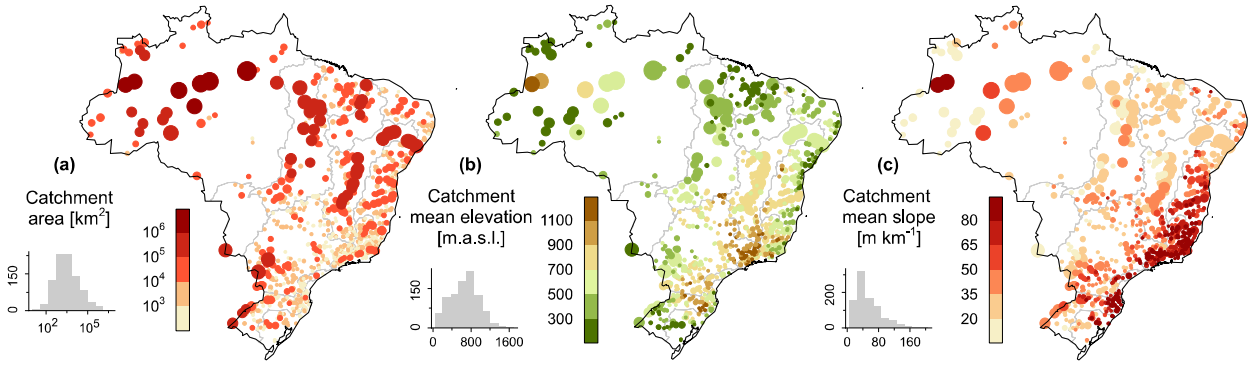

Figure 2.4. Topographic characteristics of the 897 selected catchments. The size of the circles is proportional to the size of the catchment. The grey line indicates the limits of hydrographic regions.

2.4 Climatic indices

2.4.1 Data and methods

We computed thirteen climatic indices (Table 2.3) over the same period (1990 to 2009, except for the asynchronicity index) as the ones in CAMELS (Addor et al., 2017) and CAMELS-Chile (Alvarez-Garreton et al., 2018). The first water year starts on 1st September 1989 and the last

one finishes on 31st August 2009. This is to facilitate inter-dataset comparability. We used precipitation data from CHIRPS v2.0 (Funk et al., 2015) to compute the indices since it has the highest spatial resolution among the three adopted precipitation products (i.e., CHIRPS v2.0, CPC, and MSWEP v2.2) and relies on both remote-sensing and gauge-based data.

Table 2.3. Climatic indices.

Attribute	Description	Units	Data source
p_mean	Mean daily precipitation	mm day^{-1}	CHIRPS v2.0
pet_mean	Mean daily potential evapotranspiration (PET)	mm day^{-1}	GLEAM v3.3a
et_mean	Mean daily actual evapotranspiration	mm day^{-1}	GLEAM v3.3a
aridity	Aridity, computed as the ratio of mean PET to mean precipitation	-	GLEAM v3.3a and CHIRPS v2.0
p_seasonality	Seasonality and timing of precipitation (estimated using sine curves to represent the annual temperature and precipitation cycles; values are typically between -1 (precipitation out of phase with temperature) and 1 (precipitation in phase with temperature, i.e., simultaneous peaks),; values close to 0 indicate uniform precipitation throughout the year). See equation 14 in Woods (2009)	-	CHIRPS v2.0
asynchronicity	Asynchronicity between the annual precipitation and PET cycles, where high values represent high relative magnitude and phase differences	-	Feng et al. (2009)
frac_snow	Fraction of precipitation falling as snow (i.e., on days colder than 0 ºC)	-	CHIRPS v2.0 and CPC
high_prec_freq	Frequency of high precipitation days (≥ 5 times the mean daily precipitation)	days yr^{-1}	CHIRPS v2.0
high_prec_dur	Average duration of high precipitation events (number of consecutive days ≥ 5 times the mean daily precipitation)	days	CHIRPS v2.0
high_prec_timing	Season during which most high precipitation days (≥ 5 times the mean daily precipitation) occur	season	CHIRPS v2.0
low_prec_freq	Frequency of dry days (< 1 mm day^{-1})	days yr^{-1}	CHIRPS v2.0
low_prec_dur	Average duration of dry periods (number of consecutive days < 1 mm day^{-1})	days	CHIRPS v2.0
low_prec_timing	Season during which most dry days (< 1 mm day^{-1}) occur	season	CHIRPS v2.0

The mean precipitation, mean potential evapotranspiration, and the aridity index are considered to capture long-term climatic conditions. The aridity index is the ratio of mean potential evapotranspiration to mean precipitation, which stands as a first-order control on the partitioning of precipitation into streamflow (Blöschl et al., 2013; Budyko, 1974). Those indices are complemented by the precipitation seasonality index (p_seasonality, Table 2.3), which relies on sine curves to approximate the monthly climatology of temperature and precipitation. While, for Brazil, the annual precipitation cycle is captured quite well, a sine curve provides a relatively rough approximation of the temperature cycle, particularly in the center of the country (around the state of Goiás; Berghuijs & Woods (2016). Hence, in addition to p_seasonality, we extracted the asynchronicity index proposed by (Feng et al., 2019), which relies on information theory and has the advantage of being non-parametric (in particular, it does not assume sinusoidality). The indices of extreme climatic conditions include the frequency, duration, and the most common season of high precipitation events and dry days. Dry days are defined as days with precipitation less than 1 mm, so that the index is not compromised by underdetected precipitation events (Haylock & Nicholls, 2000).

2.4.2 Spatial variability in climatic indices

The mean daily precipitation in Brazil is highest in the Amazon and in Southern Brazil, where it on average exceeds 5 mm day-1 (1825 mm year-1) (Fig. 2.5a). The lowest mean precipitation occurs in Northeastern Brazil, which is also where mean potential evapotranspiration exceeds the

mean precipitation (aridity index > 1, Fig. 2.5b). Northeastern Brazil (in particular, the states of Maranhão, Piauí, Ceará) also has the highest values of asynchronicity index in the country (not shown), which corresponds to Mediterranean climates. The precipitation regime is highly seasonal in most of the country, particularly in the central-west and southeastern Brazil (Fig. 2.5c). This seasonality is regulated by the South American Monsoon System (Carvalho et al., 2011; Raia & Cavalcanti, 2008), with peaks in the austral summer (Fig. 2.5f) and several dry months during the austral winter (Fig. 2.5i). Southern Brazil has a distinct regime, with uniform precipitation throughout the year caused by a combination of large-scale phenomena and a diversity of sources of atmospheric moisture (Martinez & Dominguez, 2014; Seager et al., 2010). The Amazon basin, which extends into both hemispheres, has contrasting precipitation regimes between the north (with a peak in austral winter) and the south (with a peak in austral summer) related to alternating warming of each hemisphere (Marengo & Espinoza, 2016). This seasonality is substantially diminished downstream in the Amazon.

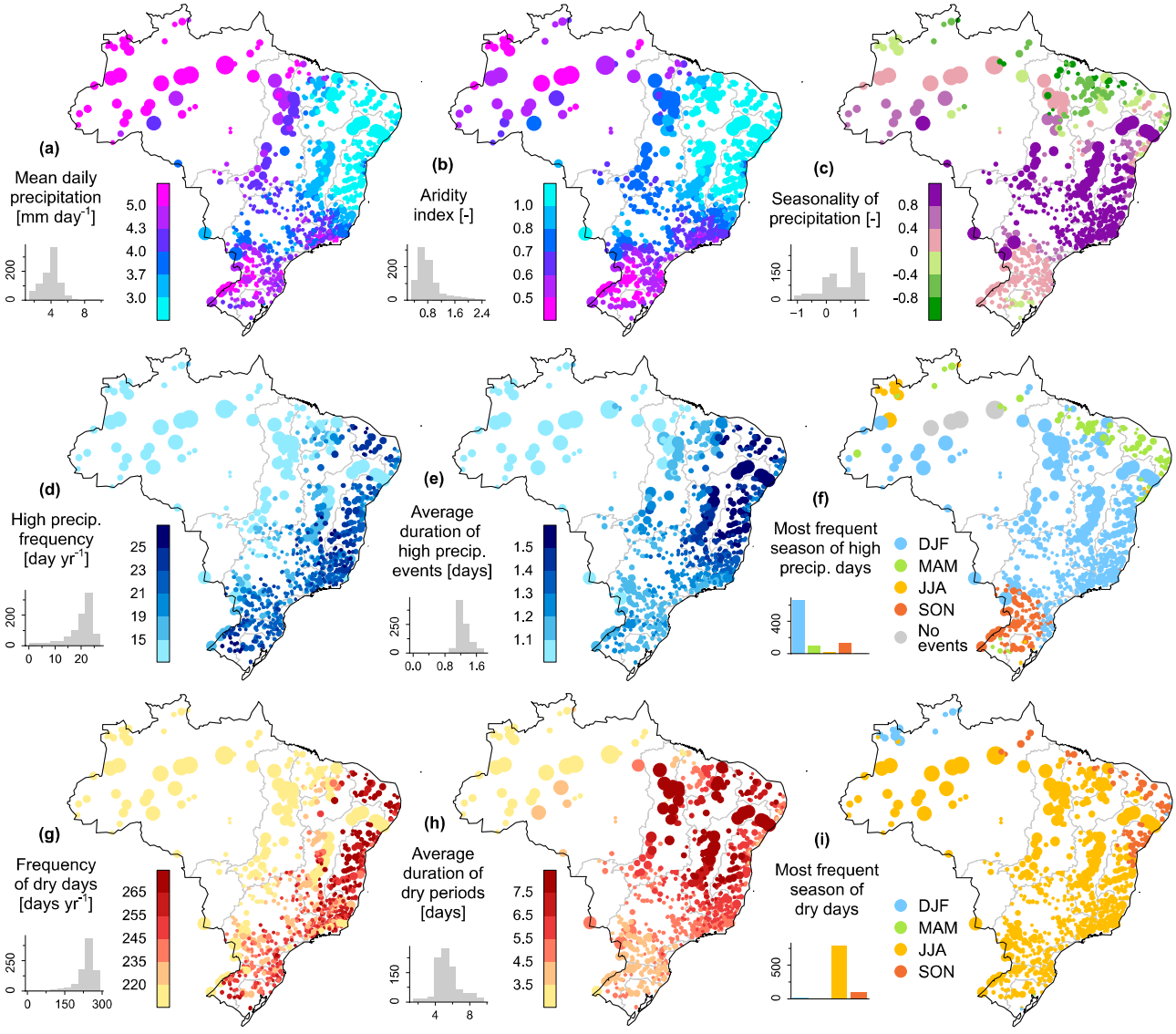

Figure 2.5. Climatic indices of the 897 selected catchments. The size of the circles is proportional to the size of the catchment. The grey line indicates the limits of hydrographic regions.

The number of high precipitation and dry days is highest along the catchments on the coast (Fig. 2.5d and 2.5g), which is also where the smallest catchments are located. Both indices are significantly correlated with catchment area ($p < 0.001$), so a regional analysis of both indices should be carried out with caution since large catchments are located in the Amazon and Tocantins-Araguaia basins. On the other hand, the duration of high precipitation (Fig. 2.5e) and dry day events (Fig. 2.5h) do not correlate with catchment area. Their spatial distribution is remarkably similar to the aridity index, except for the Tocantins-Araguaia basin, which has long dry periods but not necessarily long high precipitation events. Summer is the most common season of extreme precipitations in the majority of Brazil, with two main exceptions (Fig. 2.5f): (i) part of the coast of Northern Brazil; and (ii) Southern Brazil. This is possibly linked to mesoscale convective systems over Southeastern South America (Salio et al., 2007), to sea surface temperature anomalies in the Atlantic ocean (Liebmann et al., 2011), and the El Niño Southern Oscillation phenomenon, as those regions are particularly affected by it (Grimm, 2011; Tedeschi et al., 2013).

2.5 Hydrological signatures

2.5.1 Data and methods

We computed thirteen hydrological signatures (Table 2.4) that represent a wide range of hydrological information for the water years from 1990 to 2009. The hydrological signatures were computed in the same approach as in CAMELS, CAMEL-CL, and CAMELS-GB datasets. Intermediate streamflow conditions were evaluated with the mean daily flow and its ratio to mean daily precipitation. These were complemented by baseflow information, a fundamental component that sustains streamflow during dry periods (Smakhtin, 2001). The baseflow index is the ratio of long-term baseflow to long-term total streamflow. We used the digital filter from (Ladson et al., 2013) to separate the baseflow component from the hydrograph. The variability of streamflow was evaluated with the slope of the flow duration curve and the streamflow elasticity indices. The slope of the flow duration curve is defined as the slope between the log-transformed 33rd and 66th long-term percentiles of daily streamflow (Sawicz et al., 2014; Yadav et al., 2007). High values of that index suggest highly variable streamflow, caused either by a high seasonality of streamflow or by a flashy response to precipitation events (McMillan et al., 2017; Yokoo & Sivapalan, 2011). Streamflow elasticity is an indicator of the sensitivity of mean annual flow to changes in mean annual precipitation (Sankarasubramanian et al., 2001). For example, a streamflow elasticity value of 2 indicates that a 1% change in mean annual precipitation generates a 2% change in mean annual flow. Extreme streamflow conditions were analyzed using signatures based on the magnitude, frequency, and duration of high and low flow events. High and low flow events were defined through long-term thresholds, based on the median and mean flow, respectively (Olden & Poff, 2003). The magnitude of high and low flow events was characterized using the 5th and the 95th percentiles. There are two primary limitations to the hydrological signatures used in this study. First, several signatures might scale with catchment area. Since catchment area varies substantially among hydrographic regions, spatial analyses should be carefully conducted. Second, we did not check for temporal dependencies of consecutive high or low flow events, for example when two flood peaks occur within a couple of days from each other and both may be related to a single extreme precipitation event. Many criteria exist to identify

independent high flow events (Archfield et al., 2016; Hall et al., 2014) and low flow events (Fleig et al., 2006; Van Loon, 2015), which might lead to differences in the analyzed signatures.

Table 2.4. Hydrological signatures. Thresholds for high and low flow frequency and duration were obtained from Clausen & Biggs (2000) and Westerberg & McMillan (2015).

Attribute	Description	Units	Data source
q_mean	Mean daily discharge	mm day^{-1}	ANA (2019a)
runoff_ratio	Runoff ratio, computed as the ratio of mean daily discharge to mean daily precipitation	-	ANA (2019a)
stream_elas	Streamflow precipitation elasticity (i.e., the sensitivity of streamflow to changes in precipitation at the annual timescale, using the mean daily discharge as reference). See equation 7 in Sankarasubramanian et al. (2001), with the last element being $\bar{P}/\bar{Q}$ not $\bar{Q}/\bar{P}$	-	ANA (2019a)
slope_fdc	Slope of the flow duration curve between the log-transformed 33rd and 66th streamflow percentiles	-	ANA (2019a)
baseflow_index	Baseflow index, computed as the ratio of mean daily baseflow to mean daily discharge, with the hydrograph separation performed using the Ladson et al. (2013) digital filter	-	ANA (2019a)
hfd_mean	Mean half-flow date (i.e., the date on which the cumulative discharge since 1st September reaches half of the annual discharge)	day of the year	ANA (2019a)
Q5	5% flow quantile (low flow)	mm day^{-1}	ANA (2019a)
Q95	95% flow quantile (high flow)	mm day^{-1}	ANA (2019a)
high_q_freq	Frequency of high-flow days (> 9 times the median daily flow)	days yr^{-1}	ANA (2019a)
high_q_dur	Average duration of high-flow events (number of consecutive days > 9 times the median daily flow)	days	ANA (2019a)
low_q_freq	Frequency of low-flow days (< 0.2 times the mean daily flow)	days yr^{-1}	ANA (2019a)
low_q_dur	Average duration of low-flow events (number of consecutive days < 0.2 times the mean daily flow)	days	ANA (2019a)
zero_q_freq	Percentage of days with zero discharge	%	ANA (2019a)

2.5.2 Spatial variability in hydrological signatures

The spatial distribution of mean daily flows (Fig. 2.6a) and runoff ratio (Fig. 2.6b) closely resembles that of mean daily precipitation. These are notably high in Southern Brazil and parts of the Amazon, and low in Northeastern Brazil. The mean half-flow date (i.e., when the cumulative discharge since 1st September reaches half of the annual discharge) follows a gradient ranging from February and March in the Eastern Atlantic region to May in the Amazon and on the northern coast (Fig. 2.6c). Steep slopes of the flow duration curve occur especially in the tributaries of Southern Amazon, the Tocantins-Araguaia basin, the Eastern Atlantic hydrographic region and in parts of Southern Brazil (Fig. 2.6d). Some catchments have undefined values, meaning that they have zero flow for more than 33% of the time. Since the slopes of the flow duration curve indicate the overall streamflow variability, they are spatially similar to several other hydrological signatures. They are, most noticeably: (i) negatively correlated with the baseflow index (Fig. 2.6e), hence catchments with high baseflow may be highly resilient to dry periods (Fan, 2015); (ii) positively correlated with streamflow precipitation elasticity (Fig. 2.6f), which indicates variability at the interannual timescale; (iii) negatively correlated with the 5th percentile of streamflow (i.e., low flows; Fig. 2.6l); and (iv) positively correlated with the frequency and duration of low flow events (Fig. 2.6j and 2.6k). However, note that some regions do not follow those patterns. In particular, catchments in Southern Amazon and in the Tocantins-Araguaia basin have high baseflow indices despite steep slopes of the flow duration curve. It

possibly implies that the variability in those catchments is related to a high seasonality, rather than to a flashy response to precipitation events.

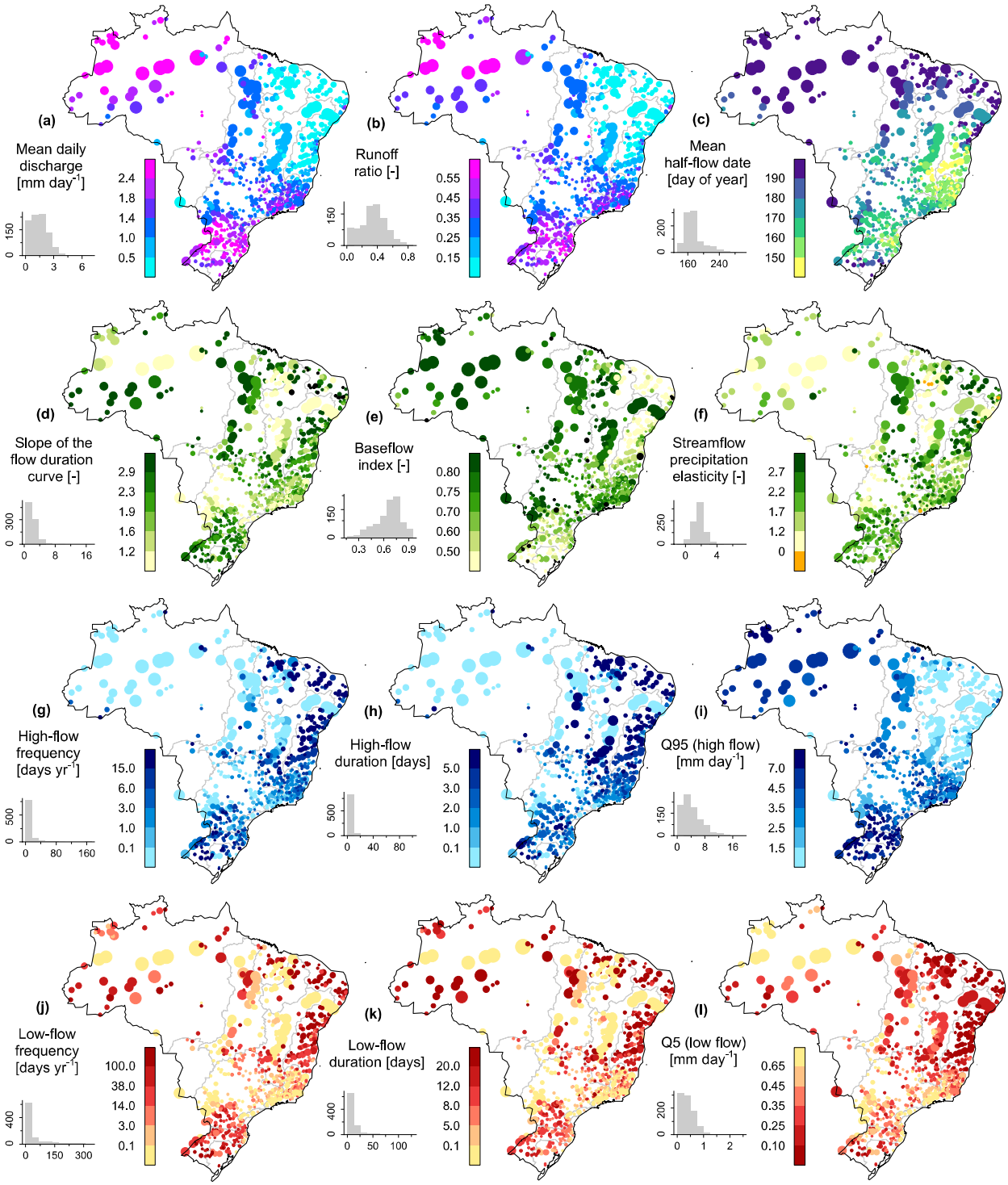

Figure 2.6. Hydrological signatures of the 897 selected catchments. The black circles are catchments with undefined values. The size of the circles is proportional to the size of the catchment. The grey line indicates the limits of hydrographic regions.

High flow days are more frequent and their events are longer in Southern Brazil, in the Eastern Atlantic region, and on the coast of northeastern Brazil (Fig. 2.6g and 2.6h). Those regions also have the most frequent and longest low flow events. This suggests that, in addition to the catchments in Southwestern Amazon and in the Tocantins-Araguaia basin, the extremes of both high and low flows might be related. Catchments seldom have high values in all three high flow signatures, except for Southern Brazil, revealing that this might be the region with the most problematic flood episodes. On the other hand, long and frequent low flows are found in nearly all hydrographic regions. The majority of catchments in the Eastern Atlantic and the East Northeastern Atlantic regions are characterized by long and frequent low flows, where nearly half of those have at least 100 days of low flows in the year.

2.6 Land cover characteristics

2.6.1 Data and methods

Each catchment was described using ten land cover classes (Table 2.5) based on GlobCover2009 (Arino et al., 2012). GlobCover2009 uses imagery from Envisat's Medium Resolution Imaging Spectrometer Instrument Fine Resolution (MERIS FR). The classification has a spatial resolution of 300 m, a global coverage every three days, and is based on images from January until December 2009. GlobCover2009 classification includes 22 land cover classes but, to simplify the dataset, we combined these into similar classes. In particular, the class "forests" is a combination of broad-leaved and needle-leaved forests, either evergreen or deciduous. Note that GlobCover 2009 does not differentiate between natural or planted forests.

There are three main limitations of the GlobCover2009 dataset. High confusion rates between croplands and grasslands show that the separation of crops, pastures, and meadows can be problematic, especially in Brazil where those land covers occur extensively. Identification of wetlands is also an issue (Arino et al., 2012) and flooded forests might be underrepresented in the classification. Lastly, GlobCover2009 used the Shuttle Radar Topography Mission (SRTM) Water Body Data, which is based on data from 2000 and does not coincide with the 2009 MERIS data.

2.6.2 Spatial variability in land cover characteristics

Croplands are widespread in Brazil, especially in the highlands, in Southern Brazil, and on the eastern coast of Northeastern Brazil (Fig. 2.7a and 2.7d). Out of the 897 CAMELS-BR catchments, 52.4% have croplands or mosaics of croplands and natural vegetation as the dominating land cover (Fig. 2.7c). Croplands are most noticeable particularly in the Uruguay and Paraná hydrographic regions. Even though GlobCover2009 does not cover the same period as the hydrological signatures (i.e., 1990-2009), croplands were already extensive in almost all states in Brazil in the 1980s and pastures in the 1960s (Dias et al., 2016; Leite et al., 2012). This is true except for Southern Amazon, where agricultural expansion has led to one of the highest deforestation rates in the world since the 1980s (Song et al., 2018).

Aside from the Amazon, catchments dominated by forests are located in mountain belts, i.e., in steep slope regions in Southern and Southeastern Brazil (Fig. 2.7b). Shrublands occur mainly in the driest regions of the country (Fig. 2.7e), but they are not the predominant land cover in these regions. Natural wetlands or water bodies are largely present in the Amazon, Tocantins-Araguaia, and Paraguay hydrographic regions (not shown). Some catchments in the

Paraná, Uruguay, and São Francisco basins are also substantially covered by water bodies. However, those are mainly artificial reservoirs (see Section 2.9.3). The CAMELS-BR catchments typically have a low fraction of their area considered to be "impervious areas", such as urban land covers; only 0.2% of the catchments have more than 5% of impervious areas (not shown). Besides, grasslands, bare soil areas, and permanent snow are rare in the CAMELS-BR catchments (not shown).

Table 2.5. Land cover characteristics.

Attribute	Description	Units	Data source
crop_perc	Percentage covered by croplands	%	ESA GlobCover2009
crop_mosaic_perc	Percentage covered by a mosaic of croplands and natural vegetation	%	ESA GlobCover2009
forest_perc	Percentage covered by broadleaved or needleleaved forests, either evergreen or deciduous	%	ESA GlobCover2009
shrub_perc	Percentage covered by shrublands	%	ESA GlobCover2009
grass_perc	Percentage covered by grasslands or areas with sparse (<15%) vegetation	%	ESA GlobCover2009
barren_perc	Percentage covered by barren areas	%	ESA GlobCover2009
imperv_perc	Percentage covered by artificial surfaces or urban areas	%	ESA GlobCover2009
wet_perc	Percentage covered by water bodies or wetlands	%	ESA GlobCover2009
snow_perc	Percentage covered by permanent snow or ice	%	ESA GlobCover2009
dom_land_cover	Dominant land cover	-	ESA GlobCover2009
dom_land_cover_perc	Percentage covered by the dominant land cover	%	ESA GlobCover2009

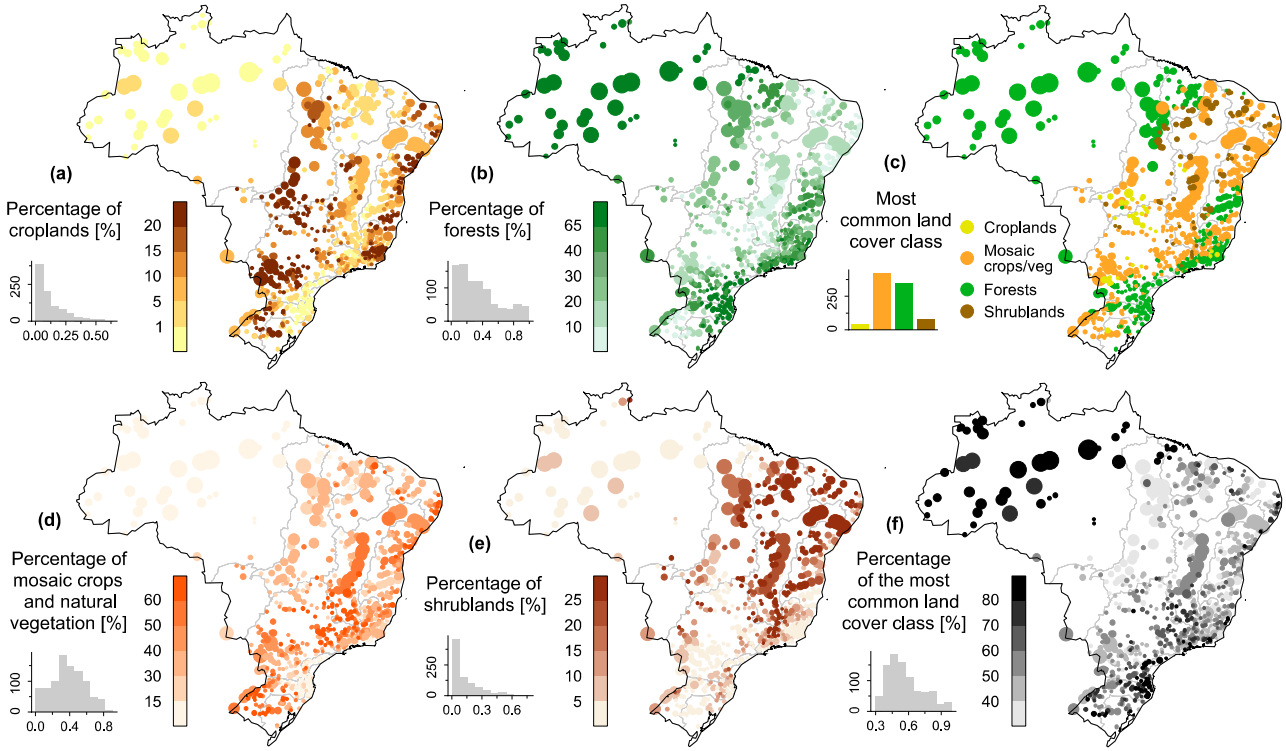

Figure 2.7. Land cover characteristics of the 897 selected catchments. The size of the circles is proportional to the size of the catchment. The grey line indicates the limits of hydrographic regions.

2.7 Geologic characteristics

2.7.1 Data and methods

The geology of the catchments was described using seven geologic attributes (Table 2.6). The first and second most common geologic class, their fractions, and the percentage of the catchment covered by carbonate rocks were extracted from the Global Lithological Map (GLiM; Hartmann & Moosdorf, 2012). GLiM was created by assembling information from 92 regional lithological maps. In the Brazilian territory, it relies on data from the Brazilian Geological Survey at the 1:1 million scale (Schobbenhaus et al., 2004). We considered only the first level of the GLiM geologic classes, which classifies lithology into 16 groups. The additional second and third levels provide more specific geologic information but were not included in this study. We note that two geologic classes cover a particularly broad variety of rocks. First, the "unconsolidated sediments" class is quite unspecific with regards to the sediment types and grain sizes (it includes sediments originated by areas as alluvial, swamp, and dune deposits). Second, catchments dominated by the "metamorphic rocks" class can have a wide range of lithologies, from shales to gneiss and quartzite.

Table 2.6. Geologic characteristics.

Attribute	Description	Units	Data source
geol_class_1st	Most common geologic class in the catchment	-	GLiM
geol_class_1st_perc	Percentage of the catchment covered by the most common geologic class	%	GLiM
geol_class_2nd	Second most common geologic class in the catchment	-	GLiM
geol_class_2nd_perc	Percentage of the catchment covered by the second most common geologic class	%	GLiM
carb_rocks_perc	Percentage of the catchment covered by carbonate sedimentary rocks	%	GLiM
geol_porosity	Subsurface porosity of the catchment	-	GLHYMPS v2.0
geol_permeability	Subsurface permeability (log10 scale) of the catchment, extract for each catchment using the geometric mean	m^2	GLHYMPS v2.0

We extracted the subsurface permeability and porosity indices from the GLobal HYdrogeology MaPS 2.0 (GLHYMPS; Gleeson et al., 2014; Huscroft et al., 2018), which is modeled based on information from the GLiM and the Global Unconsolidated Sediments Map (GUM; Börker et al., 2018). Subsurface permeability indicates how easily water can flow through the subsurface. GLHYMPS modeled it only for saturated conditions (Huscroft et al., 2018), so it is not adequate to characterize regions dominated by unsaturated processes, e.g., deeply weathered soils. The subsurface porosity indicates the fraction of void spaces in a material and controls the water storage capacity in the subsurface. A major caveat of GLHYMPS data is that it is only adequate for analyses at the regional scale, i.e., over spatial units greater than 5 km (Gleeson et al., 2014).

2.7.2 Spatial variability in geological characteristics

The catchments on the eastern coast have lithologies dominated by either metamorphic or acid plutonic rocks (Fig. 2.8a and 2.8b), related to high elevation and steep slopes in this region. These catchments also have low subsurface permeability (Fig. 2.8g) and the lowest subsurface

porosity rates in the country (Fig. 2.8f). In Southern Brazil, basic volcanic lithology is widespread, which encompasses basaltic rocks. Southern Brazil has the most homogeneous lithological types of the country (Fig. 2.8c and 2.8d), where more than 80% of the catchment areas are usually characterized by a single lithological type. However, subsuperficial porosity and permeability are highly heterogeneous, extending from middle-range to high porosity values and from middle-range to low permeability values.

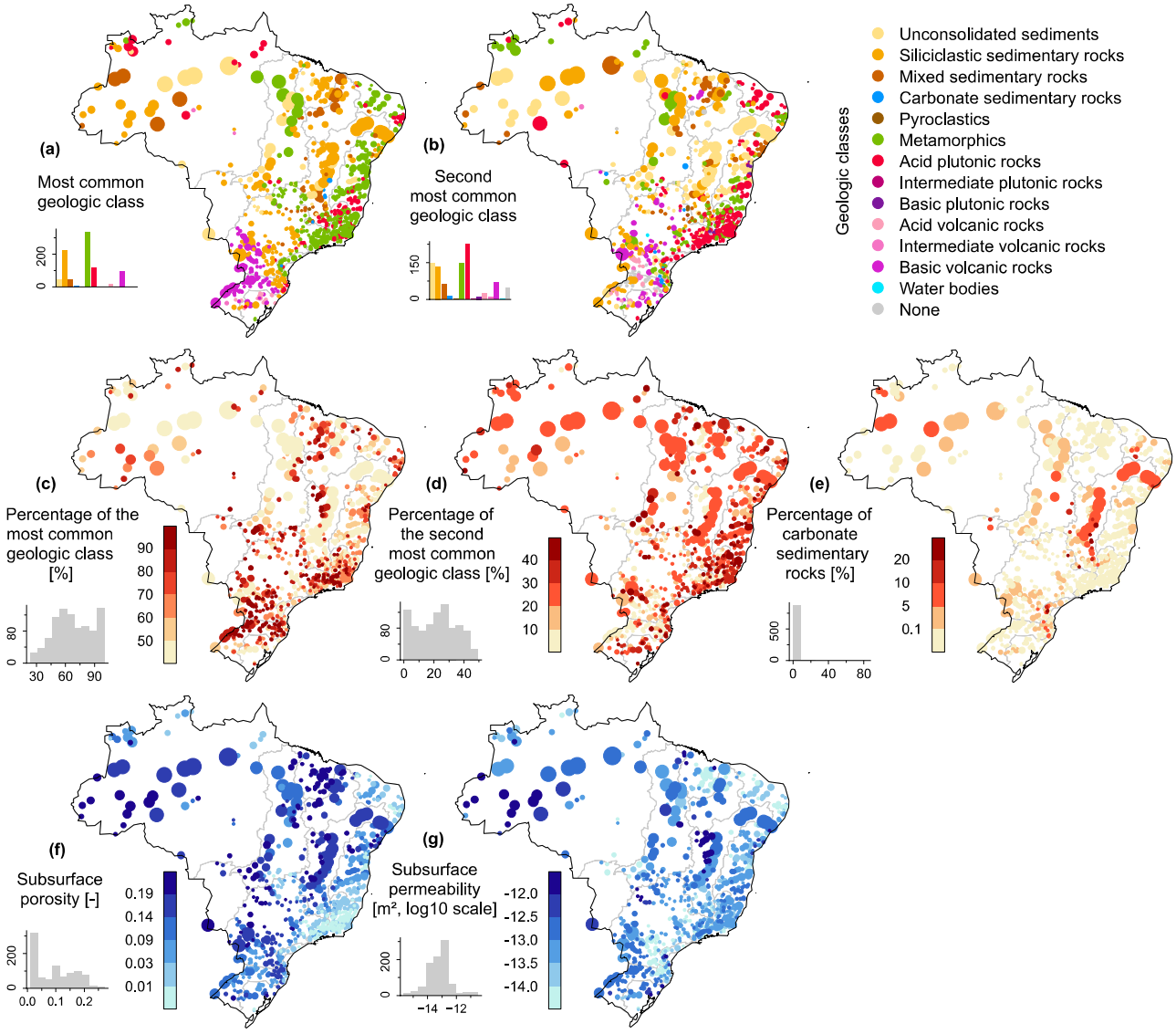

Figure 2.8. Geologic characteristics of the 897 selected catchments. The size of the circles is proportional to the size of the catchment. The grey line indicates the limits of hydrographic regions.

Sedimentary rocks occur on a large scale at São Francisco, Parnaíba, Western Northeast Atlantic, and part of the Amazon hydrographic regions. The Northern Amazon is characterized mostly by metamorphic or plutonic rocks, while the Western Amazon has either siliciclastic or mixed sedimentary lithologies. On the other hand, unconsolidated sedimentary lithologies occur particularly downstream in the Amazon, the Tocantins, and the Paraguay basins. These basins

also have flat slopes and large proportions of wetlands, which allows for alluvial particles to settle down. Most of the catchments with a high proportion of sedimentary rocks have high subsurface porosity, although their permeability varies according to the grain sizes of these rocks. Carbonate sedimentary rocks, such as karst or limestone, are more common in the São Francisco basin and in western Amazon (Fig. 2.8e). Those rocks are also present in some isolated and smaller catchments in Southern Brazil, in the Paraguay, and in the Tocantins-Araguaia basins.

2.8 Soil characteristics

2.8.1 Data and methods

We provide six soil characteristics (Table 2.7). Five of those were extracted from SoilGrids250m (Hengl et al., 2017; Shangguan et al., 2017), a collection of soil maps for the world at the 250 m resolution. SoilGrids250m maps are the result of a model using approximately 150,000 soil profiles, with predictions based on machine learning methods and 158 remote-sensing covariates including climate, vegetation, geomorphology, and lithology (Hengl et al., 2017). Although SoilGrids250m generated predictions for several soil depths, in this work we only computed soil characteristics over a depth of 30 cm. SoilGrids250m is based on a machine-learning model that explains large proportions of the variance of most observed variables, including 69% of the variance of organic carbon content and more than 70% of the soil textures (i.e., clay, silt, and sand content).

Table 2.7. Soil characteristics.

Attribute	Description	Units	Data source
sand_perc	Percentage of sand content of the soil material smaller than 2 mm at a depth of 30 cm	%	SoilGrids250m
silt_perc	Percentage of silt content of the soil material smaller than 2 mm at a depth of 30 cm	%	SoilGrids250m
clay_perc	Percentage of clay content of the soil material smaller than 2 mm at a depth of 30 cm	%	SoilGrids250m
org_carbon_content	Soil organic carbon content at a soil depth of 30 cm	$g\ kg^{-1}$	SoilGrids250m
bedrock_depth	Depth to bedrock	cm	SoilGrids250m
water_table_depth	Median water table depth	cm	Fan et al. (2013)

The soil characteristics might be highly correlated with other attributes from CAMELS-BR since they are modeled based on climatic and landscape covariates. Organic carbon content and clay content have modeled depth to bedrock as a predominant variable (Hengl et al., 2017). Other variables are also important, such as temperature and geomorphological characteristics (e.g., surface slope). The predictions of sand content are based primarily on depth to bedrock and precipitation, both at similar weights. Out of the five variables considered from the SoilGrids250m, predictions of depth to bedrock is the most problematic, with 59% of its variance explained by the model (Shangguan et al., 2017). It has precipitation as the predominant covariate, which accounts for the control of weathering rates and soil production. Other decisive covariates are vegetation dynamics and geomorphological characteristics, which accounts for factors such as soil erosion (Shangguan et al., 2017).

The sixth soil characteristic is the water table depth, based on a 1 km resolution global model by (Fan et al., 2013). Combined with depth to bedrock, water table depth can be an indicator of water storage potential in the catchment, which is related to baseflow and the supply of water for the vegetation during dry periods (Fan et al., 2013, 2019). The most important variables in the predictions of the water table depth of that model are, in decreasing order of importance, surface slope, elevation, precipitation, and temperature (Fan et al., 2013). Note that groundwater abstractions are not represented in the model, so water table depth data must be used with caution when analyzing catchments with intense anthropogenic intervention.

2.8.2 Spatial variability in soil characteristics

Soil texture in CAMELS-BR is characterized by (i) a predominance of clay content in Southern Brazil, in parts of Southeastern Brazil, particularly in higher elevations, and in northeastern Amazon (Fig. 2.9c); (ii) similar values of clay, sand, and silt content in the southern tributaries of Western Amazon (Fig. 2.9a to 2.9c); and (iii) a wide predominance of sand content in the rest of the country (Fig. 2.9a). As expected, the aridity index is closely related to the spatial distribution of the soil texture, since climatic attributes are important covariates in SoilGrids250m predictions (Hengl et al., 2017). The predominance of clay in Southern Brazil and in part of Southeastern Brazil might be linked to their lithological classes, i.e., with basic volcanic rocks in the former and acid plutonic rocks in the latter since they have coincidental spatial distributions.

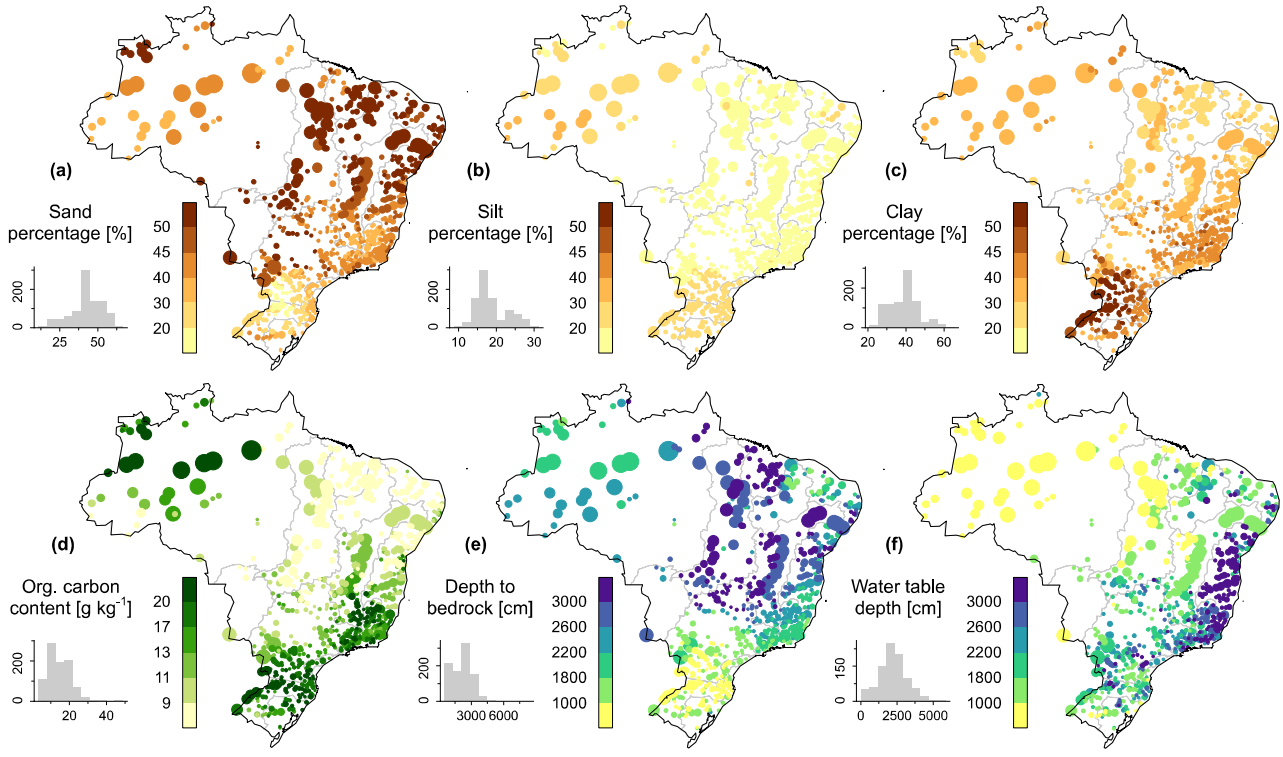

Figure 2.9. Soil characteristics of the 897 selected catchments. The size of the circles is proportional to the size of the catchment. The grey line indicates the limits of hydrographic regions.

Organic carbon content is most pronounced in parts of the Amazon and in regions with high clay content (Fig. 2.9d). The depth to bedrock is higher in Central and Northeastern Brazil, frequently above 30 m (Fig. 2.9e). On the other hand, only 14% of the catchments have depths to

bedrock lower than 10 m, all of them located in Southern Brazil. Regarding water table depths, there is a clear gradient with higher depths on the eastern coast of Brazil (i.e., exceeding 30 m deep) to lower depths towards the Amazon (i.e., less than 10 m deep). Amongst all six soil characteristics indices, water table depth has the lowest correspondence to climate, i.e., to mean precipitation and aridity index. It is mostly correlated with catchment slopes, as previously indicated by Fan et al. (2013).

2.9 Human intervention indices

2.9.1 Data and methods for consumptive water use

We computed four indices of human intervention in the catchments (Table 2.8). Two are the total consumptive water use in the catchment in 2017, one normalized by catchment areas and another normalized by mean annual streamflow. Consumptive water use refers to water withdrawals that do not return to the catchment, for example, by evaporating, transpiring, or being incorporated into manufactured products. The water uses are based on the Manual of Consumptive Water Use in Brazil (ANA, 2019c), which estimated the monthly water use of each municipality in Brazil. These estimates are the sum of water demands from six categories:

(i) Irrigation: demand based on water balance models that estimate the quantity of water needed by irrigated crops but not supplied by precipitation or soil moisture(ANA, 2019c). The spatial extent of irrigated croplands was characterized using the national censuses of agriculture (e.g., IBGE, 2007) and remote-sensing images from Landsat, Sentinel-2, and Moderate-Resolution Imaging Spectroradiometer (MODIS) (ANA, 2019b).

(ii) Livestock: demand estimated by multiplying the number of livestock units with their corresponding daily drinking requirements. The number and type of livestock of each municipality were mapped on the national censuses of agriculture.

(iii) Households: demand estimated by multiplying the number of people in a municipality by their per capita domestic water use.

(iv) Industry: demand estimated by multiplying the number of employees in several industrial categories from each municipality by its per capita water use.

(v) Mining: demand estimated by combining the water use coefficient with the annual production of several types of mineral extraction.

(vi) Thermoelectricity: demand estimated by applying a water use coefficient to the annual electricity production of each thermoelectric plant in the country.

These water demand estimates do not differentiate surface water from groundwater. Even though groundwater abstraction is extensive in the eastern part of Brazil (Fan et al., 2013), it is estimated that most of the water use in South America comes from surface water (Wada et al., 2014). To estimate the consumptive water use of each catchment, we divided the values of each municipality by its area. We assumed the water use to be spatially homogeneous throughout the municipality territory and transferred the data for each municipality onto a 500 m spatial resolution raster.

There are three major limitations of using ANA (2019c) estimated consumptive water use. First, evaporation from artificial reservoirs was not included in the computation. Thus, water use might be underestimated, particularly in the northeastern part of Brazil, i.e., in the driest part of the country. Second, the dataset comes on an irregular grid, since municipalities areas vary significantly. The smallest municipalities are usually within 500 km of the coast and their areas are mostly a few hundred km². In contrast, the western part of Brazil and the Amazon usually have municipalities larger than a thousand km². Hence, consumptive water use of small catchments in the western part of Brazil should be interpreted with caution because they are smaller than the input data. The third limitation is that the consumptive water use of South America outside Brazil was not estimated by ANA (2019c) and was not considered in this study. This affects particularly the basins in the Amazon since they cover large parts out of Brazil. That said, anthropic intervention in these basins is low: only three basins with international borders in the Amazon are more than 10% covered by croplands or croplands and natural vegetation mosaic, none has more than 0.05% of impervious land covers such as urban areas.

Table 2.8. Human intervention indices.

Attribute	Description	Units	Data source
consumptive_use	Total consumptive water use in 2017, normalized by catchment area	mm yr^{-1}	ANA (2019c)
consumptive_use_perc	Total consumptive water use in 2017, normalized by mean annual streamflow	%	ANA (2019c)
reservoirs_vol	Total maximum storage capacity of the reservoirs in the catchment	10^6 m^3	GRanD v1.3, ONS, and ANA (2018)
regulation_degree	Ratio of total reservoir storage capacity of the catchment to its total annual flow	%	GRanD v1.3, ONS, and ANA (2018)

2.9.2 Data and methods for reservoirs

The other two indices for human intervention are related to flow regulation (Table 2.8), i.e., the sum of the total storage capacity of all reservoirs in the catchment and its ratio to the total annual flow of the catchment (i.e., the degree of regulation). We worked with estimated storage capacities from 1406 reservoirs in South America. The reservoirs were mapped by combining three data sources: (i) the Global Reservoirs and Dam database v1.3 (GRanD; Lehner et al., 2011); (ii) the hydroelectric power plants database of the National Electrical System Operator (ONS, 2019); and (iii) the 2017 National Dam Safety Report (ANA, 2018) database. The GRanD database includes reservoirs throughout South America, while the other two provided data only for Brazil. The procedure for combining the three databases was:

(i) We included all reservoirs from GRanD v1.3 in South America.

(ii) For each GRanD reservoir, we visually compared the inundated area with the one indicated by the polygons from the water bodies maps from Pekel et al. (2016). When the inundated areas differed substantially, we substituted the former with the latter and updated the size of the inundated area.

(iii) Out of more than 24,000 reservoirs from (ONS, 2019) and (ANA, 2018) databases, we included only those that have their inundated areas (Pekel et al. (2016) visible at the

1:500,000 scale. Although our goal was to only include reservoirs larger than approximately 0.5 km^2, some smaller reservoirs were also included. We computed the size of the inundated areas of those reservoirs according to the polygons from Pekel et al. (2016).

(iv) To check for duplicates in the databases, we manually inspected all dam points and their inundated areas.

(v) Finally, the storage capacities of reservoirs updated in step (ii) or included in step (iii) were recalculated using their inundated areas and, when available, information on dam height. We applied two equations determined by (Lehner et al., 2011, Technical Document) with a statistical regression using data from 5824 reservoirs worldwide. When information on dam height was available, we applied Equation (2.1):

$$V = 0.678\,(A\,h)^{0.9229} \tag{2.1}$$

where V is the reservoir storage capacity in 10^6 m^3; A is the size of the inundated area in km^2; h is dam height in m. When information on dam height was not available, we used Equation (2.2):

$$V = 30.684\,A^{0.9578} \tag{2.2}$$

2.9.3 Spatial variability in human intervention indices

The spatial distribution of human interference indices reveals that, unlike the catchments in the original CAMELS for the United States, catchments in CAMELS-BR can be significantly impacted by human activities. There are 17.8% of catchments with annual consumptive water uses greater than 5% of the mean annual flow. Those are principally in the driest parts of the country, i.e., in the São Francisco, Eastern Atlantic, Eastern Northeast Atlantic, and upper Paraná hydrographic regions (Fig. 2.10b). Nevertheless, water uses greater than 20% of the mean annual flow are rare, occurring in only 3.9% of the catchments. The similarity encountered between arid climates and high consumptive water uses may be attributed to two main causes. First, in the most arid catchments, the mean annual flow is typically a third of that of the rest of the country, which, unsurprisingly, leads to higher water uses proportional to the annual flow. Second, crops in drier climates require frequent irrigation and considerable rates of water withdrawal. On the other hand, we observe that the central and southeastern regions of Brazil have the greatest values of water uses normalized by catchment area (Fig. 2.10a). Catchments in those regions are commonly occupied by either irrigated croplands or populous metropolitan areas, which are respectively the first and second categories with the highest water demands in Brazil (ANA, 2019c).

The degree of regulation is related to catchment area (Fig. 2.10c), meaning that the most regulated basins are downstream in the river basins. The main rivers with high regulations are the Paraná, Uruguay, São Francisco, Tocantins-Araguaia, Parnaíba, and Paraíba do Sul rivers. In those regions, 19.2% and 7.2% of the catchments have a degree of regulation greater than 10% and 50%, respectively. These values nearly double in the driest regions of the country (i.e., the Eastern Atlantic, São Francisco, Eastern Northeast Atlantic, and Parnaíba hydrographic regions): 37.6% and 22.1% of the catchments have a degree of regulation greater than 10% and 50%,

respectively. Therefore, the driest catchments of CAMELS-BR dataset have the highest human intervention rates, both in terms of consumptive water use and reservoir regulation

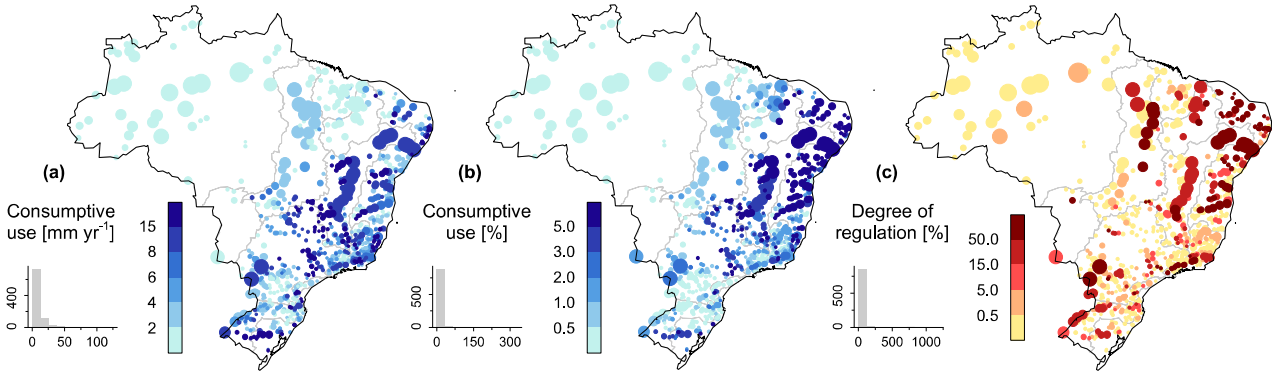

Figure 2.10. Human intervention indices of the 897 selected catchments. The size of the circles is proportional to the size of the catchment. The grey line indicates the limits of hydrographic regions.

2.10 Data availability

The CAMELS-BR dataset is freely available at https://doi.org/10.5281/zenodo.3709337 (Chagas et al., 2020b). The files provided are (i) the 65 attributes in a zip file, (ii) the daily time series in zip files, (iii) the catchment boundaries used to compute the attributes and extract the time series, computed by Do et al. (2018) and Gudmundsson et al. (2018), and (iv) a readme file.

2.11 Conclusions

So far, large-sample hydrological studies in Brazil lacked a comprehensive and easily accessible dataset. Here, we introduced the CAMELS-BR, a new dataset comprising streamflow time series for 3,679 catchments in Brazil and, for a selected quality-controlled set of 897 catchments, meteorological time series and 65 catchment attributes. The attributes cover a wide range of fundamental properties for large-sample hydrological research, such as topography, land cover, geology, soil, and human intervention characteristics. We strived to make CAMELS-BR as comparable as possible to the other CAMELS datasets (Addor et al., 2017; Alvarez-Garreton et al., 2018) by using common naming conventions, scripts, and datasets. We also discuss the major limitations of the data to limit the risk of misinterpretation and misuse.

Even though CAMELS-BR is a step forward for hydrological research in Brazil, there are several opportunities for expanding the dataset in the future. For example, future versions of CAMELS-BR could include additional catchment attributes critical to understand hydrological processes, such as drainage density and basin morphometry (Shen et al., 2017). Further, an updated version should better characterize heterogeneities within each catchment, both for the time series and attributes. Additionally, since data uncertainties are omnipresent (Addor et al., 2019; Blöschl, Hall, et al., 2019; Montanari, 2007), they should be further explored by including additional data sources.

By simplifying the access to hydrological data, we aim to encourage further large-sample hydrological studies in Brazil, to facilitate the inclusion of Brazilian catchments in global large-

sample studies, and to increase the transparency and reproducibility of these studies. We believe the data introduced here will, in particular, prove useful to explore the drivers of catchment behavior, to anticipate hydrological changes, and to study the impacts of human activities on the water cycle. We see CAMELS-BR as a resource designed to serve the broad water science community and to help with water resources management at regional, national, and continental scales.

Chapter 3

Process controls on flood seasonality in Brazil

This chapter presents a modified version of the following publication:

Chagas, V. B. P., Chaffe, P. L. B., & Blöschl, G. (2022). Process Controls on Flood Seasonality in Brazil. Geophysical Research Letters, 49(5). https://doi.org/10.1029/2021GL096754

Abstract

A coincidence in the timing of floods and their drivers can be used as a proxy for the causality of flood generation. Here, we investigate the relationship between the seasonality of floods, maximum annual rainfall, and maximum annual soil moisture data of 886 basins in Brazil for 1980-2015 to shed light on process controls of flood generation. Floods tend to occur at the same time of year as soil moisture peaks and lag behind rainfall peaks by three weeks. In Amazonia, central and northern Brazil, flood timing is more correlated with the timing of soil moisture peaks than with that of rainfall peaks, which is interpreted as resulting from high subsurface water storage capacities. In southern and southeastern Brazil, on the other hand, flood timing is highly correlated with both soil moisture and rainfall because of low subsurface water storage capacities. These findings can support flood forecasting and climate impact studies.

3.1 Introduction

River floods are usually generated by the interplay of event precipitation, antecedent soil wetness, and snowmelt (Merz & Blöschl, 2003; Rosbjerg et al., 2013; Tarasova et al., 2019). One way of exploring the relative importance of these drivers is by analyzing flood seasonality, defined as the day of the year that floods occur. A coincidence in the timing of floods and their drivers can be used as a proxy for the causality of flood generation (Parajka et al., 2010; Sivapalan et al., 2005; Tramblay et al., 2021).

The relative importance of the drivers of flood seasonality depends on climate, landscape properties, and has a large regional variability (Berghuijs, Woods, et al., 2016; Berghuijs, Harrigan, et al., 2019; Parajka et al., 2010; Tramblay et al., 2021; Wasko et al., 2020). In northeastern Europe, for example, floods are aligned with the onset of the warm season that leads to snowmelt (Blöschl et al., 2017; Kemter et al., 2020). In much of western Europe, floods are associated with soil moisture peaks in the winter because, even though rainfall peaks in the summer or autumn, it gets stored in the soil which slowly becomes wet over several months (Berghuijs, Harrigan, et al., 2019; Blöschl et al., 2017). On the other hand, in central Europe's mountain ranges, flood timing frequently coincides with annual rainfall peaks (Berghuijs, Harrigan, et al., 2019; Kemter et al., 2020). A similar pattern is found in the USA, where floods are linked with snowmelt in the coldest regions in the north, soil moisture peaks in the central-east, and extreme rainfall in the

mountain ranges in the west (Berghuijs, Woods, et al., 2016; Brunner et al., 2020; Stein et al., 2020).

In South America, more specifically Brazil, few studies have analyzed flood seasonality (Bartiko et al., 2019; Cassalho et al., 2019; Do et al., 2020). So far, no study has explored the process controls of flood seasonality by considering the interplay of rainfall and soil moisture. Even though soil moisture is crucial in determining if rain infiltrates or runs off (Bonell, 2004; Elsenbeer, 2001), floods in Brazil have usually been explained in terms of extreme rainfall and meteorological phenomena such as the South American monsoon, mesoscale convective systems, and the El-Niño-Southern Oscillation (Cavalcanti, 2012; Fleischmann et al., 2020; Lima et al., 2017; Marengo & Espinoza, 2016; Schöngart & Junk, 2007; Sena et al., 2012; Towner et al., 2021).

The objective of this paper is to explore the main process controls on flood generation in Brazil. We investigate whether annual floods are linked mainly with maximum annual rainfall or soil moisture considering the similarity of their seasonalities. We use circular statistics to analyze the mean dates of occurrence and interannual variabilities of hydrometeorological data of 886 basins in Brazil, from 1980 to 2015.

3.2 Materials and Methods

3.2.1 Hydrometeorological data

We use daily streamflow data of 886 hydrometric stations from the Brazilian National Water Agency (http://www.snirh.gov.br/hidroweb) as made available by the CAMELS-BR data set (Chagas et al., 2020a). The analysis period is from 1980 to 2015 because of its large data availability. The criteria for selecting the 886 hydrometric stations are: (i) at least 25 years of data from 1980 to 2015 with less than 5% missing; (ii) approved by a quality control similar to that of CAMELS-BR, on which stations with typographical errors or unrealistic high flows are discarded; (iii) basins with a ratio between total water storage in artificial reservoirs and annual flow lower than 25% (using a ratio of 5% we arrive at similar results – not shown); and (iv) basins with urban land cover lower than 10%, as we are interested in large-scale hydrological patterns and minimizing local impacts. The basin sizes range from 11 km^2 to 4.7 million km^2.

The daily rainfall data are from CHIRPS v2.0 (Funk et al., 2015), from 1981 to 2015. CHIRPS has a spatial resolution of 0.05º and uses data from meteorological stations and satellite sensors. We chose CHIRPS because it shows a good accuracy compared with other precipitation products (Beck, Vergopolan, et al., 2017; Wongchuig-Correa et al., 2017). Furthermore, we use surface soil moisture data from GLEAM v3.5a (Martens et al., 2017; Miralles et al., 2011), from 1980 to 2015, with a spatial resolution of 0.05º. GLEAM is based on satellite data and presents good performance compared with other soil moisture products (Beck et al., 2021). We use surface soil moisture data instead of moisture data from deeper layers because it has better accuracy (Beck et al., 2021; Brocca et al., 2017) and is more relevant for runoff processes (Bonell, 2004). We conducted alternative analyses using root zone soil moisture data from GLEAM and rainfall data from ERA5-Land (Hersbach et al., 2020; Muñoz-Sabater et al., 2021) and obtained similar conclusions (not shown). The rainfall and soil moisture time series are computed as basin averages. Snow data is not included because it is not a dominant hydrological variable in the analyzed basins (Chagas et al., 2020a).

3.2.2 Seasonality analysis

We investigate links between floods and their drivers by comparing the seasonalities of maximum annual streamflow (floods), maximum annual rainfall of a 7-day moving average series (maximum rainfall), and maximum annual surface soil moisture of a 7-day moving average series (maximum soil moisture). The 7-day time scale is relevant to both small and large basins and is widely used in flood studies (e.g., Berghuijs et al., 2016; Blöschl et al., 2017; Stein et al., 2020; Tramblay et al., 2021; Wasko et al., 2020). We conducted an alternative analysis with time scales of 1, 3, and 14 days. For the latter two we obtained similar conclusions and for the 1-day time scale neither driver is strongly linked to floods (Fig. A.4). We would like to mention, however, that the rainfall temporal scales analyzed are not representative particularly for large basins in the Amazon, where response times can reach a few months. Thus, large basins in the Amazon are a major limitation of this study.

We analyze the seasonality of floods and their drivers using circular statistics (Bayliss & Jones, 1993; Mardia & Jupp, 2009) because, unlike traditional statistics, it treats the first and last day of the year as temporally adjacent. We first calculate the mean dates of floods, maximum rainfall and soil moisture. The water year starts in September and the ordinal day D_i is transformed into an angular value θ_i with

$$\theta_i = \frac{D_i}{m_i} \cdot 2\pi \tag{3.1}$$

where m_i corresponds to the number of days in the year t_i and $i = 1, \ldots, n$, where n is the number of years of record. The mean date of occurrence $\bar{\theta}$ (in radians) is defined by (Bayliss & Jones, 1993; Mardia & Jupp, 2009)

$$\bar{\theta} = \begin{cases} \tan^{-1}\left(\frac{S}{C}\right) & c > 0, s \geq 0 \\ \tan^{-1}\left(\frac{S}{C}\right) + \pi & c \leq 0 \\ \tan^{-1}\left(\frac{S}{C}\right) + 2\pi & c > 0, s < 0 \end{cases} \tag{3.2}$$

with

$$s = \sum_{i=1}^{n} \sin(\theta_i) \tag{3.3}$$

$$c = \sum_{i=1}^{n} \cos(\theta_i) \tag{3.4}$$

where s and c are the sine and cosine components of the mean dates; n is the number of observations, in this case the number of years. Then, we calculate the seasonality strength with the mean resultant length R, defined by

$$R = \frac{1}{n}\sqrt{c^2 + s^2} \tag{3.5}$$

which ranges from 0 (no seasonality, flood peaks are evenly distributed throughout the year) to 1 (higher seasonality, flood peaks always occur on the same day of the year). We verify if the seasonality of each basin is statistically significant and unimodal with the Rayleigh test (Mardia & Jupp, 2009; Rayleigh, 1880). Out of the 886 basins analyzed, 750 have significant unimodal flood seasonalities ($p < 0.05$). Even though multimodal seasonalities might be present in Brazil, we do not analyze their strength and mean dates of occurrence as these metrics are only reliable for unimodal regimes. The circular histograms of the annual maxima for the basins analyzed are available in Chagas et al. (2022) so that their flood regimes can be checked visually

3.2.3 Importance of the flood seasonality drivers

We analyze links between the seasonality of floods, maximum rainfall and soil moisture with two methods. In the first, we examine differences between their mean dates as described in Section 3.2.2. This method may be affected by the differences in flood travel times in the basins (that is, the time an intense rainfall event takes to propagate through a basin), so we complement it with the second method.

In the second method, we analyze the interannual variability by correlating the time series of flood timing with maximum rainfall and soil moisture timing. For each basin, we compute the circular correlation (Jammalamadaka & Sengupta, 2001) with

$$\rho = \frac{\sum_{i=1}^{n} \sin(\alpha_i - \bar{\alpha}) \sin(\beta_i - \bar{\beta})}{\sqrt{\sum_{i=1}^{n} \sin^2(\alpha_i - \bar{\alpha}) \sin^2(\beta_i - \bar{\beta})}} \tag{3.6}$$

where α_i and β_i are the dates of occurrence (in angular values) of the correlated variables for the year t_i and $i = 1, \ldots, n$, where n is the number of observations; $\bar{\alpha}$ and $\bar{\beta}$ are mean dates of occurrence computed with Equation 3.2. For each basin, we analyze the correlation between the timing of floods and maximum rainfall and the correlation between the timing of floods and maximum soil moisture. The circular correlation is not affected by different mean dates in the variables, since it considers annual dates of occurrence as anomalies from their respective average by subtracting one from the other. The correlation is less affected by contrasting flood travel times in the basins as, for each basin, it removes the time gap between the mean dates of floods, maximum annual rainfall and soil moisture.

Furthermore, we explore the regional importance of each flood seasonality driver by spatially interpolating the circular correlations. We interpolate the correlations with ordinary block kriging sized 2º by 2º (approximately 222 by 222 km at the equator) using the best fit variogram model with the gstat R package (Gräler et al., 2016; Pebesma, 2004).

3.3 Results

3.3.1 Flood, rainfall, and soil moisture seasonality

Floods are remarkably seasonal in Brazil (Fig. 3.1). Flood seasonality is most pronounced in Amazonia (Fig. 3.1d, strength greater than 0.9, $p < 0.001$), where streamflow peaks in March on the southern tributaries and June on the northern tributaries (Fig. 3.1a). Floods are highly seasonal in central-eastern Brazil (strength between 0.7 and 0.9, $p < 0.001$) and occur in the summer (December-February). On the other hand, flood seasonality is less pronounced in the south (strength below 0.5, $p < 0.05$), with mean dates in the winter (July-September) but with frequent floods also in spring and autumn.

Similar to floods, rainfall and soil moisture peak in the summer in central Brazil (Fig. 3.1b-c), May-June in northern Amazonia, and are less seasonal in southern Brazil (Fig. 3.1e-f). However, in Amazonia, rainfall and soil moisture peaks are notably less seasonal than floods. In southern Brazil, rainfall seasonality is not statistically significant (significance level $\alpha = 0.05$) even though soil moisture and floods are. In addition, a visual comparison of the panels in Fig. 3.1 suggests that maximum rainfall tends to occur a few months before the floods, but this does not seem to be the case for maximum soil moisture.

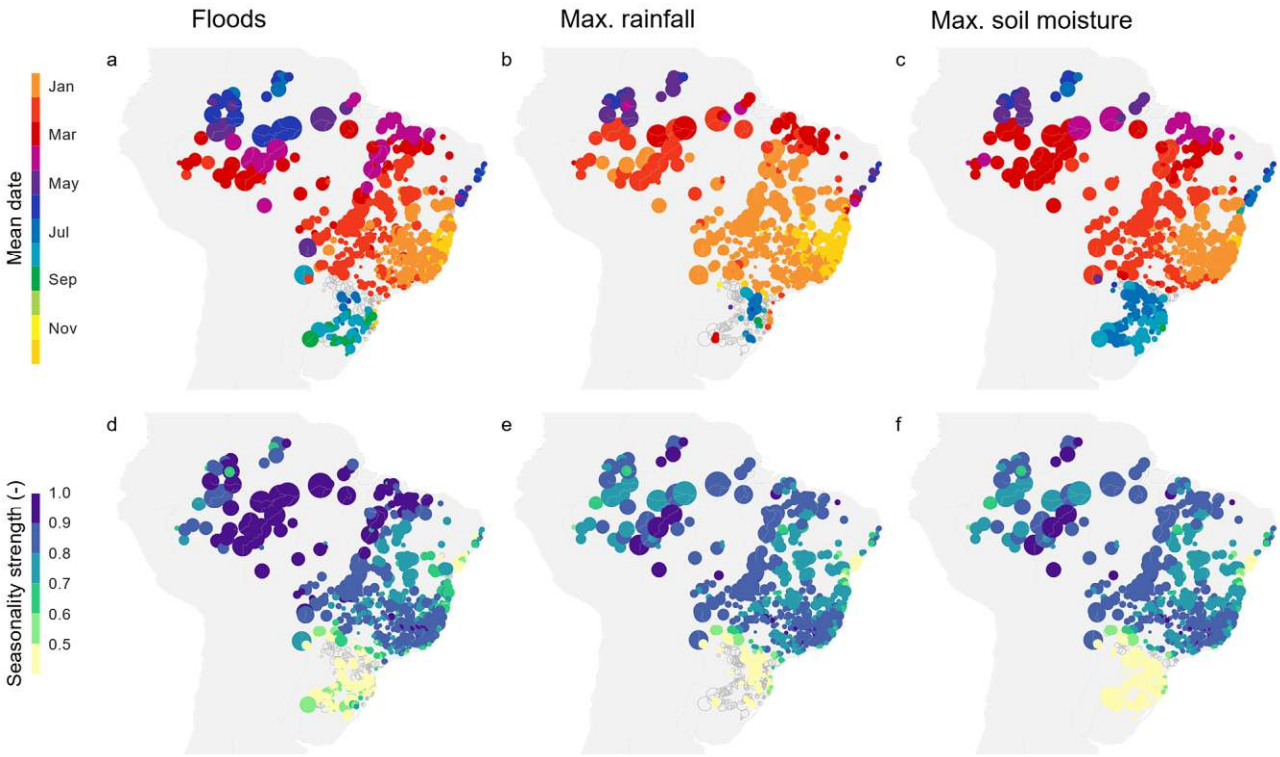

Figure 3.1. Mean dates and seasonality strength of (a, d) floods, (b, e) maximum annual rainfall, and (c, f) maximum annual soil moisture. Seasonality strength ranges from 0 (no seasonality) to 1 (floods always occur on the same day of the year). Gray open circles are non-significant or non-unimodal seasonalities ($\alpha = 0.05$). Circle size is proportional to the logarithm of the basin area.

A more detailed inspection shows that the mean dates of floods are closer to those of soil moisture than to those of rainfall (i.e., closer to the 1:1 line in Fig. 3.2). Floods occur on average

20 days after maximum rainfall and 5 days before maximum soil moisture. Floods usually lag behind maximum rainfall by 14 days or more in 60% of the basins (Fig. 3.2a). Floods and maximum rainfall are within 14 days of each other in 38% of the basins. On the other hand, floods and soil moisture occur within 14 days of each other in 65% of the basins, with the remaining 35% distributed equally between before and after the 14 days difference (Fig. 3.2b).

We explore the regional seasonality patterns with six hotspots with distinct flood regimes, climates, and located in the upstream areas of major rivers (i.e., Amazon, Araguaia, Uruguay, Parnaíba, Doce) (Fig. 3.2c). In the Southern Amazonia hotspot, flood and maximum soil moisture timings coincide but floods and maximum rainfall tend to occur more than 6 weeks apart. Such large differences are likely affected by flood travel times, as Southern Amazonia has the largest basins among the hotspots. A similar pattern is observed in the Central-West and North hotspots, where floods and maximum soil moisture coincide but lag behind maximum rainfall by 3 to 5 weeks. These findings are independent of basin size (Fig. A.1). In the South and Southeast hotspots, mean dates of floods, maximum rainfall and soil moisture are within 14 days of each other, suggesting that both drivers have similar importance for the process controls of flood seasonality.

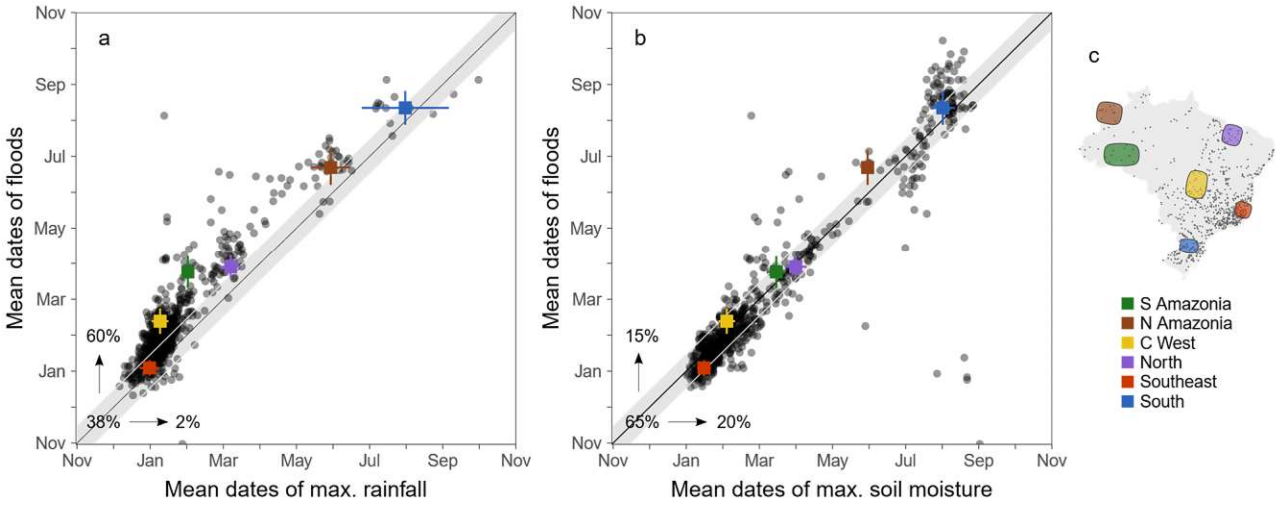

Figure 3.2. Mean dates of floods as a function of mean dates of (a) maximum annual rainfall and (b) maximum annual soil moisture. Points represent basins, squares indicate the median value of each hotspot shown in (c), and error bars show ± one standard deviation. The black line is a 1:1 function and the gray band and lines indicate ± 14 days. Points above the 1:1 line represent basins where floods generally occur after the mean dates of maximum annual rainfall in (a) and after the mean dates of maximum annual soil moisture in (b). The numbers indicate the percentage of basins within, above, or below the grey band. Basins with non-significant (α = 0.05) or non-unimodal seasonalities are not included.

3.3.2 Interannual variability of flood seasonality

The interannual variability analysis shows how between-year deviations from the mean dates of floods, maximum rainfall and soil moisture are correlated. As opposed to the method used in Section 3.3.1, the interannual variability is not affected by different mean dates in the variables. The interannual variability is less influenced by contrasting flood travel times in the basins as,

for each basin, it removes the time gap between the mean dates of floods, maximum annual rainfall and soil moisture.

The interannual variability of flood timing is more correlated with that of maximum soil moisture timing than with that of maximum rainfall timing in 87% of the study area (Fig. 3.3a-b). The mean correlation with soil moisture is higher than 0.4 in 36% of the study area and, with rainfall, in 15% of the study area (Fig. 3.3a-b). Floods and soil moisture correlations are high particularly in central, northern, southern, and eastern Brazil. Floods and rainfall correlations, on the other hand, are high in part of southern and eastern Brazil. The correlations of individual basins and their associated statistical significance are presented in Fig. A.2.

Most hotspots present either greater importance of soil moisture or similar importance of both drivers of flood seasonality (Fig. 3.3c). Greater importance of soil moisture is noticeable in the North (median correlations with soil moisture and rainfall of 0.67 and 0.45 respectively), Central-West (median correlations of 0.46 and 0.25 respectively), and Northern Amazonia hotspots (correlations of 0.38 and 0.20 respectively). On the other hand, similar importance of soil moisture and rainfall is observed in the South and Southeast hotspots. Median correlations in both hotspots are around 0.50, approximately 70% of which are significant (α = 0.05). Once again, as with the analysis of mean dates, Southern Amazonia is the hotspot with the weakest links between flood timing and its drivers, with correlations generally below 0.25 and mostly non-significant (α = 0.05).

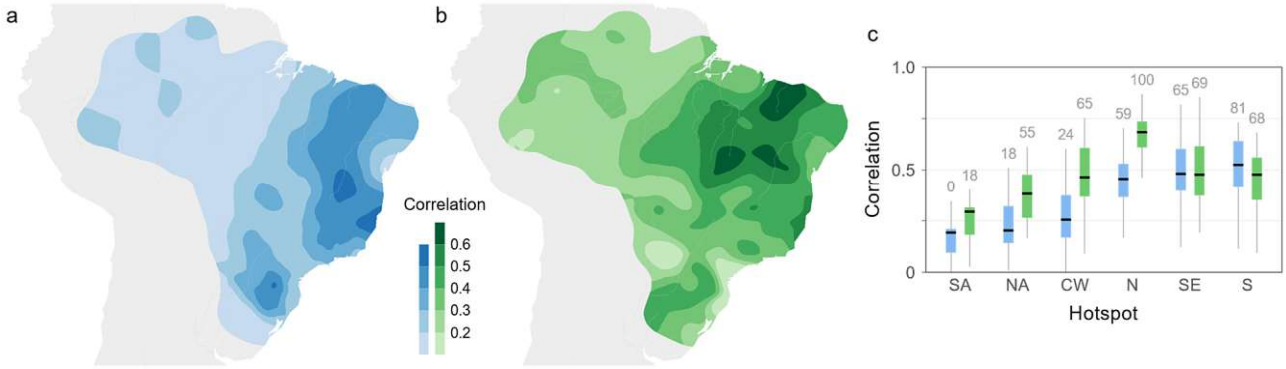

Figure 3.3. Circular correlation between the interannual variability of the timing of (a) floods and maximum annual rainfall, (b) floods and maximum annual soil moisture. Both (a) and (b) are obtained with interpolation using block kriging. (c) Spatial variability of the correlations with maximum rainfall (blue boxes) and maximum soil moisture (green boxes) over each hotspot. The numbers above the boxplots indicate the percentage of basins with significant correlations (α = 0.05). The hotspots are Southern Amazonia (SA, n = 11), Northern Amazonia (NA, n = 11), Central-West (CW, n = 34), North (N, n = 22), Southeast (SE, n = 65), and South (S, n = 37).

We complement the seasonality analysis with an examination of the magnitudes of the variables. For each basin, we analyze the correlations of the timing for two separate groups of events considering the lowest and highest 50% of the floods, respectively. Correlations of the timing of the highest floods to their drivers are substantially higher than those of the lowest floods (Fig. A.3, Table A.1). As with the previous analyses, flood seasonality is more closely associated with soil moisture in both groups. However, the correlation uncertainties and their spatial variability are greater. Furthermore, for each basin we correlate the interannual variability of flood magnitudes with maximum annual rainfall and soil moisture magnitudes, again indicating overall greater importance of soil moisture compared with rainfall (Fig. A.4, Table A.1).

3.4 Discussion

3.4.1 Process controls on flood seasonality

Our results suggest two main patterns of flood generation. In some parts of Brazil, flood timing is aligned mostly with maximum soil moisture timing, suggesting that on average floods are modulated mainly by antecedent soil wetness and less so by variations in event rainfall. In other parts of Brazil, flood seasonality is similar to both maximum rainfall and soil moisture seasonalities, indicating that both are relevant for flood generation.

Fig. 3.4a-b illustrates the first pattern of flood generation with the Candeias and Guamá river basins. In the Candeias basin, a tributary of the Madeira river in southern Amazonia, the most intense annual rainfall events usually occur in January before the soil is wet enough to become saturated and generate flood peaks. The soil gets wet during the summer and, when wetness peaks in March, other less intense rainfall events lead to the annual floods. A similar pattern is observed in the Guamá basin, on the northern coast of Brazil. The most intense annual rainfall events can occur at any time in the wet season (Jan-Apr), but the flood peaks are only generated when soil moisture peaks at the end of the wet season. This is comparable with those in parts of western Europe, such as southern England, where flood timing is aligned with annual soil moisture peaks (Berghuijs, Harrigan, et al., 2019; Blöschl et al., 2017).

Fig. 3.4c-d illustrates the second pattern of flood generation with the Doce and Içana river basins. In the Doce basin, on the southeastern coast, rainfall usually peaks in December and is soon followed by soil moisture peaks in late December or early January. Since peak rainfall and soil moisture are closely aligned, floods also usually occur in December or January. In the Içana basin, a tributary of the Negro river in northern Amazonia, rainfall rates are above 5 mm per day throughout the year without a dry season. Consequently, the soil is always close to saturation and, when intense rainfall events take place in May and April, floods occur.

3.4.2 Importance of water storage capacity

According to our interpretation, the patterns of flood generation are related to the root zone water storage capacity, that is, the maximum volume of hydrologically active soil water available for plant transpiration (de Boer-Euser et al., 2016; Wang-Erlandsson et al., 2016). The first pattern (greater importance of soil moisture) may occur particularly in basins with large water storage capacities. In these basins, rainfall in the wet season tends to get stored, increasing soil moisture and groundwater tables continuously until they peak at the end of the wet season when floods are generated. This phenomenon may be particularly noticeable in regions with long rainy seasons, as groundwater tables may rise over a longer period. The second pattern (similar importance of rainfall and soil moisture) may occur mainly in basins with low water storage capacities. Intense rainfall quickly saturates the soil, becoming runoff and generating floods.

The largest estimated water storage capacities are in southern Amazonia (median and standard deviation of 379 ± 83 mm year^{-1}), central-western (328 ± 105 mm year^{-1}) and northern Brazil (472 ± 251 mm year^{-1}) (Wang-Erlandsson et al., 2016) which is also where the first pattern of flood generation is most frequent. This does not come as a surprise, as these regions have several landscape properties associated with high water storage capacities. These include deeply weathered and highly permeable soils, commonly tens of meters deep (Hengl et al., 2017; Pelletier et al., 2016), and low topographic slopes with widespread floodplains and wetlands (Junk et al., 2014; Nardi et al., 2019). On the other hand, the lowest estimated storage capacities

are in southern (108 ± 38 mm year^{-1}) and southeastern (226 ± 92 mm year^{-1}) Brazil (Wang-Erlandsson et al., 2016), where the second pattern of flood generation is most frequent, associated with higher topographic slopes, mountain ranges, and shallower soils (Hengl et al., 2017; Pelletier et al., 2016).

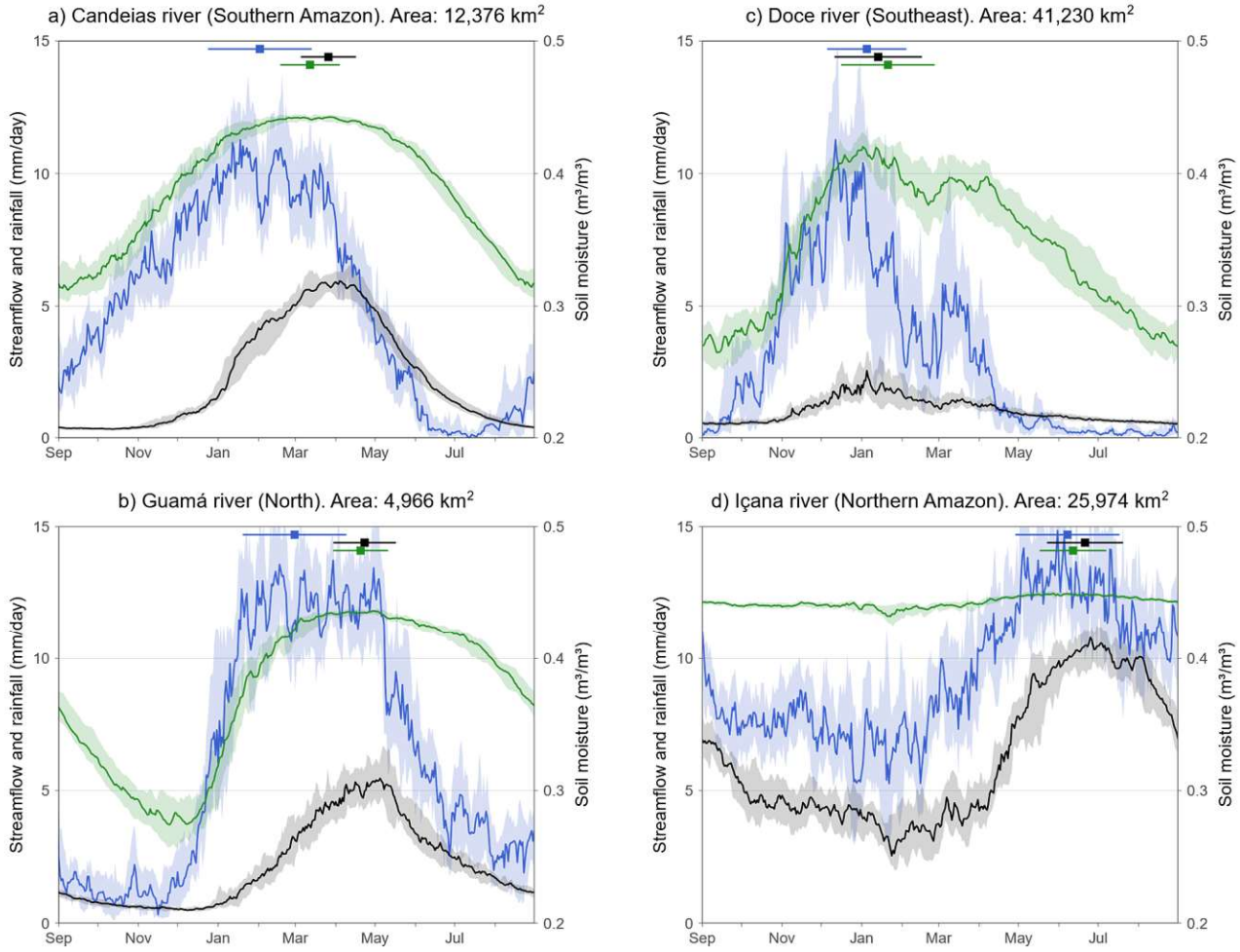

Figure 3.4. Examples of the two patterns of flood generation. (a, b) Flood seasonality associated mainly with the timing of maximum annual soil moisture. (c, d) Flood seasonality associated with the timing of maximum annual soil moisture and rainfall with similar importance. Lines indicate median daily streamflow (black), 7-day rainfall (blue), and 7-day soil moisture (green) of each day of the year (1980-2015). The bands indicate the percentiles 30 and 70. The squares and error bars on the top indicate mean dates and standard deviations. Streamflow values are normalized by basin area to facilitate inter-basin comparison.

The results suggest that water storage capacity is more relevant than basin size for regulating whether the interannual variability of floods depends on intense rainfall magnitudes or antecedent soil moisture. Larger river basins have longer flood wave travel times, leading to increased time lags between event rainfall and flooding and an attenuation of the magnitude of flood peaks. However, the correlations between the interannual variability of the timing of floods, soil moisture peaks, and rainfall peaks have a clear regional pattern and the absence of a changing pattern from upstream to downstream (Fig. A.2), which would be expected if basin size was the main control of floods. For example, in Amazonia, both the smaller and larger basins

have low correlations between the timing of floods and intense rainfall, both attributed to large water storage capacities. An impact of basin size is not evident even when stratifying the study area into regions, although here we investigate only the temporal variability of flood timing and contrasting results may be found for the spatial variability or flood magnitudes.

3.4.3 Links with meteorological characteristics

The flood timing found here is consistent with the meteorological characteristics of the region. The South American monsoon carries large amounts of atmospheric moisture to Amazonia, central and southeastern Brazil particularly in summer (Dec-Feb) (Grimm, 2019; Marengo et al., 2012). Cold fronts and transient systems favor convection and lead to frequent intense rainfall events, especially in the wet season. With the dry monsoon phase starting in April, the Intertropical Convergence Zone moves southward producing rainfall in the northeast (Cavalcanti, 2012; Grimm, 2019). In southern Brazil, intense rainfall can occur at any time of year with the wet monsoon phase in summer, mesoscale convective systems in spring or summer, and cold fronts particularly in winter (Cavalcanti, 2012; Durkee et al., 2009). These meteorological drivers play an extremely important role in flood generation through increases in soil moisture, particularly in Amazonia, central-western and northern Brazil.

3.5 Conclusions

In most of Brazil, flood seasonality is more closely associated with the seasonality of peak soil moisture than with the seasonality of peak rainfall. Peak soil moisture is associated with cumulative rainfall and evaporation, thus suggesting that flood seasonality is more closely aligned with rainfall variability at the monthly scale than with shorter intense rainfall events. We identified two patterns of flood generation. In Amazonia, central and northern Brazil, flood timing coincides with the timing of annual soil moisture peaks. This suggests that, on average, floods are modulated mainly by antecedent soil wetness and less so by variations in event rainfall, which we interpret as occurring mainly in basins with large water storage capacities. On the other hand, in southern and southeastern Brazil, flood timing coincides with the timing of both annual rainfall and soil moisture peaks. This indicates that both antecedent soil wetness and event rainfall are major modulators of flood generation, which we hypothesize as occurring mainly in basins with low water storage capacities. Because of the fast soil saturation, one would expect this pattern to be more sensitive to climate change induced increases in extreme rainfall, while the previous pattern would be more sensitive to cumulative (e.g., monthly) rainfall changes. The understanding of the major controls of flood generation provided here can support flood forecasting, risk management, and climate change and land cover impact studies.

Chapter 4

Regional low flow hydrology: Model development and evaluation

This chapter presents the following manuscript under preparation:

Chagas, V. B. P., Chaffe, P. L. B., & Blöschl, G. (2023). Regional low flow hydrology: Model development and evaluation. Currently under review at Water Resources Research.

Abstract

Low flows result from the interplay of climatic variability and basin storage characteristics, but it is not clear which of these variables is more relevant for explaining spatial low flow patterns. Here, we develop a new conceptual model that integrates process-based hydrological knowledge with statistics, and test it for 1,400 Brazilian river basins. Through comparative hydrology, we isolate the low flow generating mechanisms and estimate their components using tree models. The results show that the model explains 61% of the spatial variance in the observed 7-day minimum annual flows (Q_{min}). Basin characteristics are twice as important as climate in predicting Q_{min}, suggesting that low flows are governed by the basin's capacity to mitigate climatic variability through water storage. Geological properties are the most important basin characteristics, particularly bedrock type, lithology, and topographic slope, which determine streamflow recession rates in the dry season. Soil properties, primarily soil class and depth, are half as important as geology. Climate impacts Q_{min} mainly through mean annual rainfall minus evaporation, representing potential groundwater recharge, while dry-season length has the least impact. The primary Q_{min} controls vary depending on the spatial scale of analysis. Basin characteristics mainly govern Q_{min} up to the continental scale (10^7 km^2), where their relative importance matches that of climate. Low flow hydrology that combines statistics with process understanding offers a promising framework for understanding regional low flow generating mechanisms and could also support other estimation models than that presented here.

4.1 Introduction

Low river flows result from the interplay of climatic variability and basin characteristics related to water storage dynamics, such as soil and geological properties (Laaha et al., 2013; Smakhtin, 2001). Climate indicates precipitation rates in wet periods, when groundwater is recharged, and dry spell length in dry periods, when groundwater is released into the stream. Soil infiltration capacity controls the fraction of precipitation that recharges the aquifer (de Vries & Simmers,

2002), especially in wet regions, while bedrock permeability, fractures, and topography govern groundwater discharge into the stream (Sophocleous, 2002).

There is conflicting evidence on whether climate or basin characteristics regulate most of the regional pattern of low flows. One hypothesis is that climate controls the spatial variability of low flows (e.g., Beck et al., 2013, 2015; Addor et al., 2018; Peña-Arancibia et al., 2010; Apurv and Cai, 2020). It suggests that low flows are driven by precipitation and evaporation rates and timing, rather than by the basin's capacity to attenuate the climatic variability. For example, an analysis of over 3,000 basins worldwide shows that 47% of the spatial variability in the Q_{99} (flow exceeded 99% of the time) is explained by the aridity index (ratio of mean annual precipitation to evaporation), while soil and geology are relatively unimportant (Beck et al., 2015). Similar results were found for over 600 basins in the U.S. (Addor et al., 2018; Apurv & Cai, 2020). Additionally, the streamflow recession, which indicates flow decreasing rates in dry spells, is also driven by the aridity index in 167 basins in the global tropics (Peña-Arancibia et al., 2010). In middle latitudes, the flow recession is linked with mean snow depth and potential evaporation (Beck et al., 2013).

An alternative hypothesis proposes that basin characteristics control most of the low flow spatial variability (e.g., Bloomfield et al., 2009; Carlier et al., 2018; Floriancic et al., 2022; Pfister et al., 2017; Price et al., 2011; Sayama et al., 2011). It indicates that the basin's water storage largely mitigates the climatic variability. For instance, a study of 35 river basins in North Carolina and Georgia (U.S.) shows that low flow magnitudes are driven by the drainage density and the basin's fraction of colluvium (Price et al., 2011). It proposes that basins with dense river networks have greater connectivity between surface and subsurface water, promoting water removal and reduced low flows (Price et al., 2011). A study of 22 headwater basins underlain by basalts in western Oregon (U.S.) reveals that low flows are linked with the age of the geological formation (Tague & Grant, 2004). Older rocks have deeper groundwater systems that sustain streamflows for a longer period (Tague & Grant, 2004). Increased low flows have also been linked with highly permeable bedrocks, such as sandstones and alluvium in the Swiss Plateau (Carlier et al., 2018; Floriancic et al., 2022) and hydrogeological and soil types in the Thames basin (U.K.; Bloomfield et al., 2009).

From our perspective, the contradicting evidence on the controls of low flow spatial variability is related to two main challenges. The first challenge is that the results depend on the spatial scale of analysis, that is, the domain size of the study. At continental and global scales, low flows are generally linked with climate rather than basin characteristics (Beck et al., 2013, 2015; Addor et al., 2018; Peña-Arancibia et al., 2010; Apurv and Cai, 2020), with few exceptions (e.g., Van Lanen et al., 2013). At smaller scales, in contrast, low flows are typically associated with basin characteristics (Bloomfield et al., 2009; Carlier et al., 2018; Floriancic et al., 2022; Pfister et al., 2017; Price et al., 2011; Sayama et al., 2011). This scaling property is related to the fact that hydrological processes linked with basin characteristics occur at local to continental scales, while processes linked with the climate occur over large scales (Blöschl & Sivapalan, 1995; Western et al., 2002). However, a precise assessment of the impact of the scale of analysis on low flows remains to be determined.

The second challenge is that the basin characteristics connected to low flows are highly specific for each study area, leading to fragmented and place-specific hydrological knowledge. A particular basin attribute may be correlated with increased low flows in one region, but reduced low flows in another (Gnann et al., 2021). Thus, the lack of universally-relevant basin characteristics hinders the progress of large-scale low flow research (Addor et al., 2018; Beck et al., 2015; Gnann et al., 2021).

One way to address the scale-dependent and place-specific features of low flows is to build a process-based conceptual model before conducting a formal statistical analysis. Traditionally, low flow statistics rely on direct links between low flow indices, climate and basin characteristics (Laaha et al., 2013; Stedinger et al., 1993). But it is unclear if these "black box" models are right for the right reasons, as they prioritize parameter fitting over understanding the low flow generating mechanisms. Another approach is to first develop a conceptual model such that its components represent stages of low flow generation. The components point to universally-relevant hydrological functions such as recharge, storage, and release, which integrate the effects of climatic variability and water storage dynamics. Each component can be associated with a unique set of climatic and basin characteristics depending on the region. With this, the conceptual model can provide a process-based framework for other methods to build upon, such as machine learning techniques. We refer to this approach as "low flow hydrology", which merges statistics with hydrological information content, as inspired by the flood frequency hydrology approach (Merz & Blöschl, 2008a, 2008b; Viglione et al., 2013).

Here, we use comparative hydrology to develop a new conceptual model and investigate how climate and basin characteristics control low flow spatial variability across spatial scales. We analyze the 7-day minimum annual flow magnitude in 1,400 Brazilian river basins, a region where large-scale studies are critically needed to mitigate growing drought concerns (Bevacqua et al., 2021; Chagas et al., 2022a; Getirana et al., 2021; Tomasella et al., 2022). The model components are estimated with model-based recursive partitioning, a specific type of tree model. With this study, we aim to bridge the gap between local and large-scale studies by contrasting universally-relevant knowledge (captured by the conceptual model) with place-specific knowledge (captured by the tree model and selected basins).

4.2 Methods

4.2.1 Data Sets

We analyze basin characteristics and hydrometeorological data of 1,400 river basins in Brazil (Table 4.1). The basin characteristics include topography (elevation, slope), geology (bedrock type, formation, age, hydraulic conductivity), soil (class, depth, texture), and land cover. The hydrometeorological data are observed daily streamflow, precipitation, and estimated actual evaporation including transpiration. We include river gauges with at least three years with less than 5% missing data (1981-2020), as we are interested in the largest possible spatial coverage in data-scarce regions like Amazonia (Fig. 4.1a). We analyze only basins with reduced human interference by applying the selection criteria: (i) less than 5% of the basin covered by urban use; (ii) total upstream artificial water storage capacity lower than 10% of the annual river discharge; (iii) basins smaller than 100 km^2. We visually inspect the hydrometeorological time series and discard gauges with unrealistic values, such as flows higher than 1,000 mm d^{-1} or zero flow instead of missing data.

4.2.2 Study Area

The study area is covered mainly by floodplains and highlands (Fig. 4.1d). Floodplains include most Quaternary unconsolidated sediment lithologies (Fig. 4.1f), with high clay content in Amazonia and low clay content in the lower Paraguay and Araguaia basins (Fig. 4.1h). Conversely, the highlands can be roughly divided into three regions. First, highlands underlain by

metamorphic and plutonic rocks, mostly old formations (pre-Silurian) of the Brazilian shield (Fig. 4.1f). The soil is mostly highly weathered, deep (20 meters or more, Fig. 4.1g) and permeable ferralsols and acrisols (Fig. 4.1i). Second, highlands underlain by sedimentary siliciclastic rocks, ranging from sandstones with high hydraulic conductivities (e.g., Botucatú, Parecis, and Urucuia aquifers) to claystones with low hydraulic conductivities (e.g., southwestern Amazonia, southern Brazil). Third, highlands underlain by younger (post-Silurian) volcanic rocks, mostly basalts in the south, generally with the shallowest soils in Brazil (less than 10 meters deep) such as clay-rich regosols or cambisols.

Table 4.1. Description of the hydrometeorological and basin characteristics data sets used in this study.

Name	Description	Sources
Streamflow	Daily observed streamflow from 1981 to 2020 in 1,400 river gauges	CAMELS-BR (Chagas et al., 2020a), ANA (2019a)
Precipitation	Daily observed precipitation from 1981 to 2020 (basin aggregates)	CHIRPS (Funk et al., 2015)
Evaporation	Daily estimated actual evaporation (including transpiration) from 1981 to 2020 (basin aggregates)	GLEAM v3.5a (Martens et al., 2017; Miralles et al., 2011)
Elevation	Elevation at the 3 arc-second horizontal resolution (about 90 meters)	SRTM (Werner, 2001)
Topographic slope	Topographic slope for each 3 arc-second cell based on the elevation data	SRTM (Werner, 2001)
Flow wave travel time	Flow wave travel time from the furthest upstream river to the basin outlet	Allen et al. (2018)
Artificial water regulation	Ratio of the basin's total artificial reservoir storage capacity to its annual streamflow	CAMELS-BR (Chagas et al., 2020a), ANA (2019b), Lehner et al. (2011), ONS (2019)
Land cover	Basin fraction of forests, shrublands, grasslands, sparse vegetation, bare areas, wetlands, croplands, settlements, and water bodies in the year 2010	ESA (2020)
Bedrock type	Bedrock types including sedimentary, volcanic, plutonic and metamorphic rocks	GLiM (Hartmann & Moosdorf, 2012)
Lithology	Geological formation of the top layer	Schobbenhaus et al. (2004) and SGB (2021) for Brazil; Gómes Tapias et al. (2019) for the rest of South America
Geological age	Minimum and maximum age of the top layer geological formation	Schobbenhaus et al. (2004) and SGB (2021) for Brazil; Gómes Tapias et al. (2019) for the rest of South America
Groundwater hydraulic conductivity	Saturated hydraulic conductivity of the top layer geological formation	Diniz et al. (2014, 2015) for Brazil and the La Plata basin; Gleeson et al. (2014) for the rest of South America
Soil texture	Soil fraction of sand, clay, and silt	SoilGrids250m (Hengl et al., 2017; Poggio et al., 2021)
Soil depth	Depth to bedrock	Shangguan et al. (2017)
Soil class	Soil class according to the World Reference Base for Soil Resources (WRB, 2015)	EMBRAPA (dos Santos et al., 2011) for Brazil; SoilGrids250m (Hengl et al., 2017; Poggio et al., 2021) for the rest of South America
Water table depth	Water table depth	Fan et al. (2013)

The rainfall regime is predominantly wet and remarkably seasonal in most of the study area, including Amazonia, central and northern Brazil (Fig. 4.1b-c). The seasonality is associated with the South American Monsoon System (Grimm, 2019), with wet summers (Sep-Mar), dry winters (Apr-Aug) and in-phase rainfall and evaporation regimes (Feng et al., 2019). The highest rainfall rates (above 1,800 mm yr^{-1}) and shortest dry seasons (less than 3 months long, Jan-Mar) occur in northwestern Amazonia and temperate southern Brazil. The lowest rainfall rates (below 700 mm yr^{-1}) and longest dry seasons (more than 6 months long) occur in northeastern Brazil,

the only region with lower mean annual rainfall than potential evaporation. Snowfall is not relevant in the study area (Alvares et al., 2013).

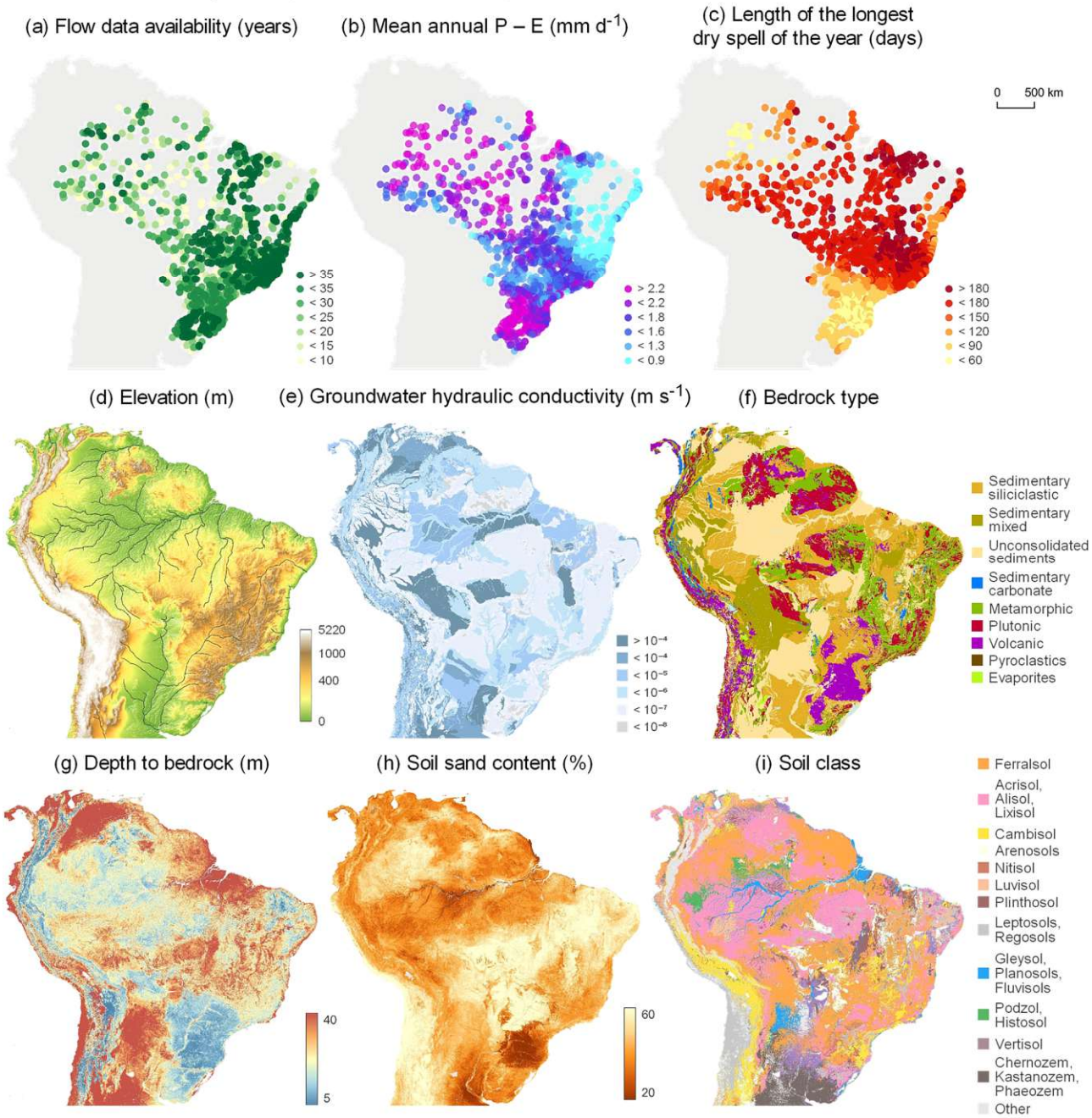

Figure 4.1. Streamflow, climate, geology, and soil properties in Brazil. (a) Streamflow data availability for 1,400 gauges. (b) Mean annual rainfall (P) minus evaporation (E) for each basin. (c) Length of the longest dry spell of the year for each basin. (d) Elevation. (e) Saturated groundwater hydraulic conductivity of geological formations. (f) Bedrock type. (g) Depth to bedrock. (h) Soil sand percentage. (i) Soil class.

4.2.3 Building a Perceptual Model with Comparative Hydrology

To estimate low flows regionally, instead of relying solely on statistical models, we develop a conceptual model that breaks down low flow generating mechanisms into components. This

approach aims to bring together process-based hydrological knowledge and statistics. We begin by examining the roles of climate and basin characteristics through comparative hydrology (Blöschl et al., 2013; Falkenmark & Chapman, 1989). We look for similarities between groups of basins, hypothesize the main high-level components that generate low flows, and use our interpretation to build a perceptual model (Gnann et al., 2021; Wagener et al., 2021). We then translate the perceptual model into a conceptual model and estimate its components using machine learning techniques.

We build the perceptual model by examining streamflow in three pairs of basins with analogous similarity measures in every aspect except for one (Fig. 4.2-4.3). The chosen similarity measures represent the interplay of climatic variability and basin water storage dynamics. For climate, we include mean annual rainfall (P) minus evaporation (E), and the length of the longest dry spell of the year (further explained in Section 4.2.4). For basin water storage dynamics, we include bedrock type and lithology, soil class and depth. Additionally, each selected basin represents a larger region with similar physiography, covering a wide range of climates and basin characteristics. For each pair of basins, we investigate disparities in low flow generation and link them to the contrasting similarity measure.

The first pair of basins illustrates how low flows in wet climates are affected by geologies with contrasting active storage capacities (Fig. 4.2a-b, Table 4.2). Both the Almas (Fig. 4.2a) and Sucuriú basins (Fig. 4.2b) have similarly wet climates (mean annual P – E of 1.52 and 1.71 mm d^{-1}, respectively) and long dry seasons (dry spell length of 168 and 159 days). Both basins are covered by ferralsols and arenosols, deep soils (more than 20 meters) with high permeabilities. However, each basin has a different lithology. The Almas basin is underlain by fractured aquifers with metamorphic rocks, mostly schists and metagranites with low hydraulic conductivity and low active storage. The Sucuriú basin is underlain by porous aquifers with sedimentary siliciclastic rocks, sandstones with high hydraulic conductivity and high active storage. When the dry spell begins, in both basins, streamflow magnitude is close to the mean annual P – E. However, streamflow decreases faster in the Almas basin than in the Sucuriú basin. Streamflow takes 59 days to halve in the Almas basin compared with 252 days in the Sucuriú basin (further explained in Section 4.2.4), suggesting that a smaller change in groundwater storage leads to a higher change in streamflow. Thus, the minimum flow is lower in the Almas basin (0.22 mm d^{-1}) than in the Sucuriú basin (1.16 mm d^{-1}), presumably driven by the contrasting active storage capacities of the sedimentary and metamorphic lithologies.

A similar interpretation is given for the second pair of basins, but for drier climates (Fig. 4.2c-d). The Mundaú (Fig. 4.2c) and Pandeiros basins (Fig. 4.2d) have similarly dry and seasonal climates, deep soils (more than 20 meters) and highly permeable ferralsols, arenosols and acrisols. However, the Mundaú basin is formed by metamorphic rocks (low hydraulic conductivity) while the Pandeiros basin is formed by sedimentary siliciclastic rocks (high hydraulic conductivity). When the dry spell begins, streamflow magnitude is close to the mean P minus E in both basins, but the flow recession is quicker in the Mundaú basin (71 days to half flow) than in the Pandeiros basin (196 days to half flow). Minimum flow is lower in the former (0.10 mm d^{-1}) than in the latter (0.26 mm d^{-1}), which we attribute to the contrasting active storage capacities in both lithologies.

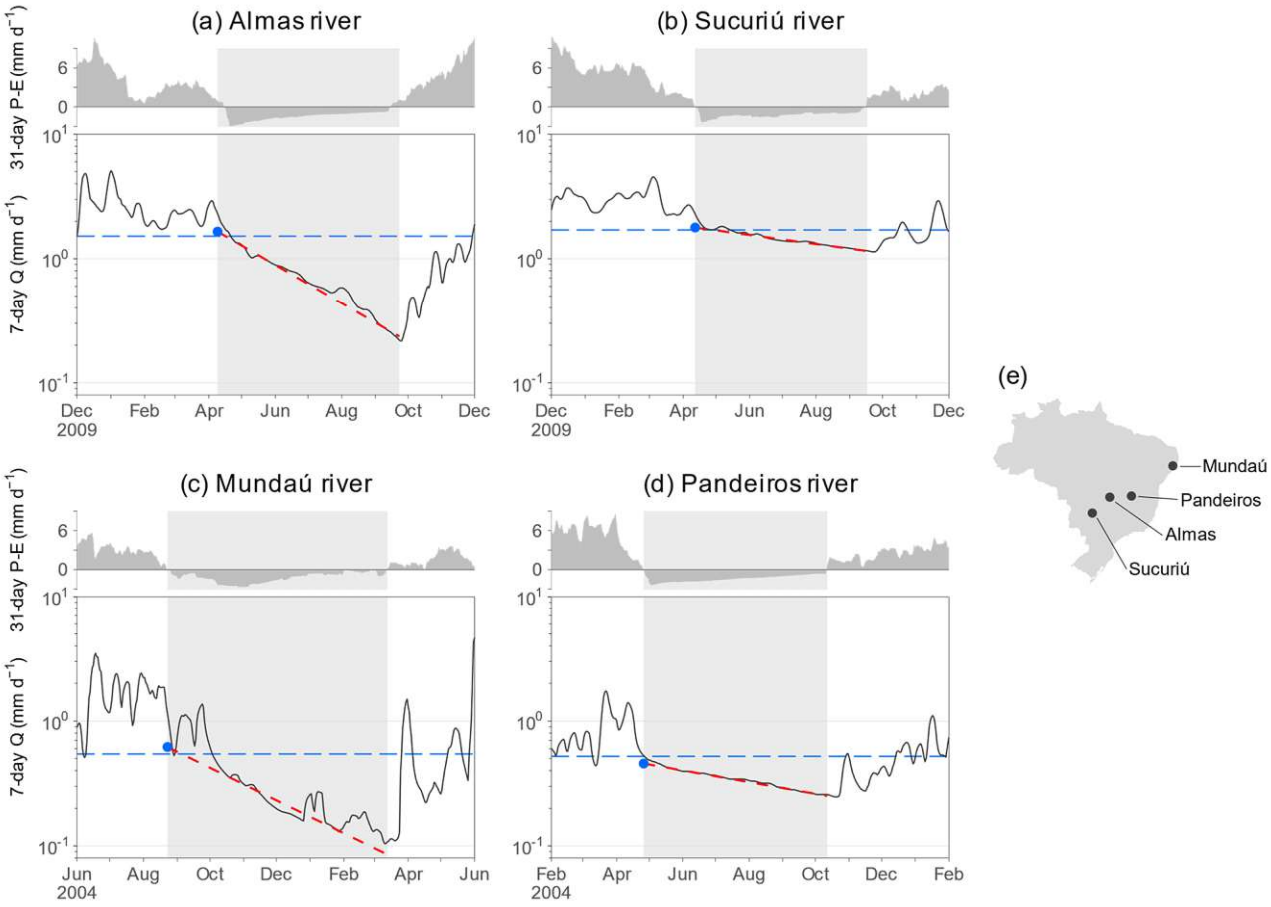

Figure 4.2. Low flow generation in two pairs of basins with similar climates but contrasting geologies. Time series of streamflow (Q) and rainfall (P) minus evaporation (E) in the: (a) Almas river (wet climate, low groundwater storage), 1,980 km^2, gauge code 20100000; (b) Sucuriú river (wet climate, high groundwater storage), 3,753 km^2, gauge 63001200; (c) Mundaú river (dry climate, low groundwater storage), 2,913 km^2, gauge 39740000; (d) Pandeiros river (dry climate, high groundwater storage), 3,263 km^2, gauge 44250000. (e) Location of the four basins. The blue dashed lines show the mean annual streamflow. The shaded rectangles indicate dry spells. The red dashed lines show the streamflow fit during the dry spell and the blue circle shows the estimated initial flow.

The third pair of basins (Fig. 4.3) illustrates how low flows are influenced by soil properties. The Iracema and Santo Cristo basins are close to each other and share similar wet climates (mean annual P – E of 2.18 and 2.47 mm d^{-1}, dry spell length of 65 and 71 days, respectively), geological formation (basalts), and comparable quick flow recessions (30 and 46 days to half flow). The main difference between the two basins is their soil properties. The Iracema basin features shallow, poorly-developed cambisols rich in clay content (56%), usually associated with low permeabilities (e.g., Gonçalves-Maduro et al., 2020; Reichert et al., 2020). In contrast, the Santo Cristo basin features ferralsols, a deep and well-developed soil, generally associated with high permeabilities (e.g., Reichert et al., 2020; Tomasella & Hodnett, 1997) despite the high clay content (56%). In the wet period, streamflow in the Iracema river is highly variable, dropping below 0.5 mm d^{-1} between storms (Fig. 4.3b, left panel) and surging above 50 mm d^{-1} during intense storms (Fig. 4.3b, right panel). Meanwhile, streamflow in the Santo Cristo river is less

variable, usually remaining above 1.0 mm d^{-1} between storms (Fig. 4.3a, left panel) and below 50 mm d^{-1} even during intense storms (Fig. 4.3a, right panel). When the dry spell begins, streamflow in the Iracema basin is lower (0.65 mm d^{-1} at the dry spell onset, 21% of the mean annual streamflow) than in the Santo Cristo basin (2.02 mm d^{-1}, 81% of the mean annual streamflow). It results in a minimum flow of 0.15 mm d^{-1} in the Iracema basin and 0.71 mm d^{-1} in the Santo Cristo basin. This phenomenon can be observed throughout the years in both basins. We infer that differences in the minimum flows are driven by streamflow magnitudes at the onset of dry spells, which in turn are driven by soil properties related to permeability. Soil controls the fraction of rainfall that infiltrates and recharges the groundwater stores before dry spells.

Table 4.2. Basin characteristics and hydrometeorological properties of the dry spells from Fig. 4.2-4.3. Q indicates streamflow and Q_{min} indicates the 7-day minimum annual flow.

Basin	Event Q_{min} (mm d^{-1})	Mean annual Q (mm d^{-1})	Event dry spell length (days)	Event flow recession (days to half flow)	Event initial flow (mm d^{-1})	Lithology	Soil
Fig. 4.2							
Almas	0.22	1.56	168	59	1.67	Metamorphic (schist, metagranite)	Deep, permeable ferralsols
Sucuriú	1.16	2.05	159	252	1.79	Sedimentary siliciclastic (sandstone)	Deep, permeable ferralsols and arenosols
Mundaú	0.10	0.91	203	71	0.63	Metamorphic (schist, orthogneiss)	Deep, permeable acrisols
Pandeiros	0.26	0.45	170	196	0.46	Sedimentary siliciclastic (sandstone, conglomerate)	Deep, permeable ferralsols and arenosols
Fig. 4.3							
St. Cristo	0.71	2.47	71	46	2.02	Volcanic (basalt)	Deep ferralsols
Iracema	0.15	2.18	65	30	0.65	Volcanic (basalt)	Shallow cambisols

The perceptual model derived from the examples above can be summarized in two stages, a wetting and a drying stage. In the wetting stage, mean annual rainfall minus evaporation and the soil permeability determine aquifer recharge. The groundwater recharge volume, in turn, is linked with the streamflow magnitude at the end of the wet season. In the drying stage, rainfall typically ceases and streamflow is sustained by groundwater storage. Geological properties govern groundwater release into the stream and, consequently, streamflow decreasing rates, as indicated by the flow recession. The combination of initial flow at the dry season onset, flow recession, and dry spell length leads to the minimum flows at the end of the dry period. Notably, this perceptual model works best in climates with considerable and univariate seasonality, like those found in most of Brazil, and negligible snowfall rates.

4.2.4 Diagnostic Model

We estimate the 7-day minimum annual flow magnitude (Q_{min}) with a diagnostic model and a predictive model. In the diagnostic model, we use observed streamflow data to estimate model parameters in order to evaluate the model's potential with parameter estimates as close to the true values as possible. In the predictive model, we use climatic and basin characteristics to estimate the model parameters.

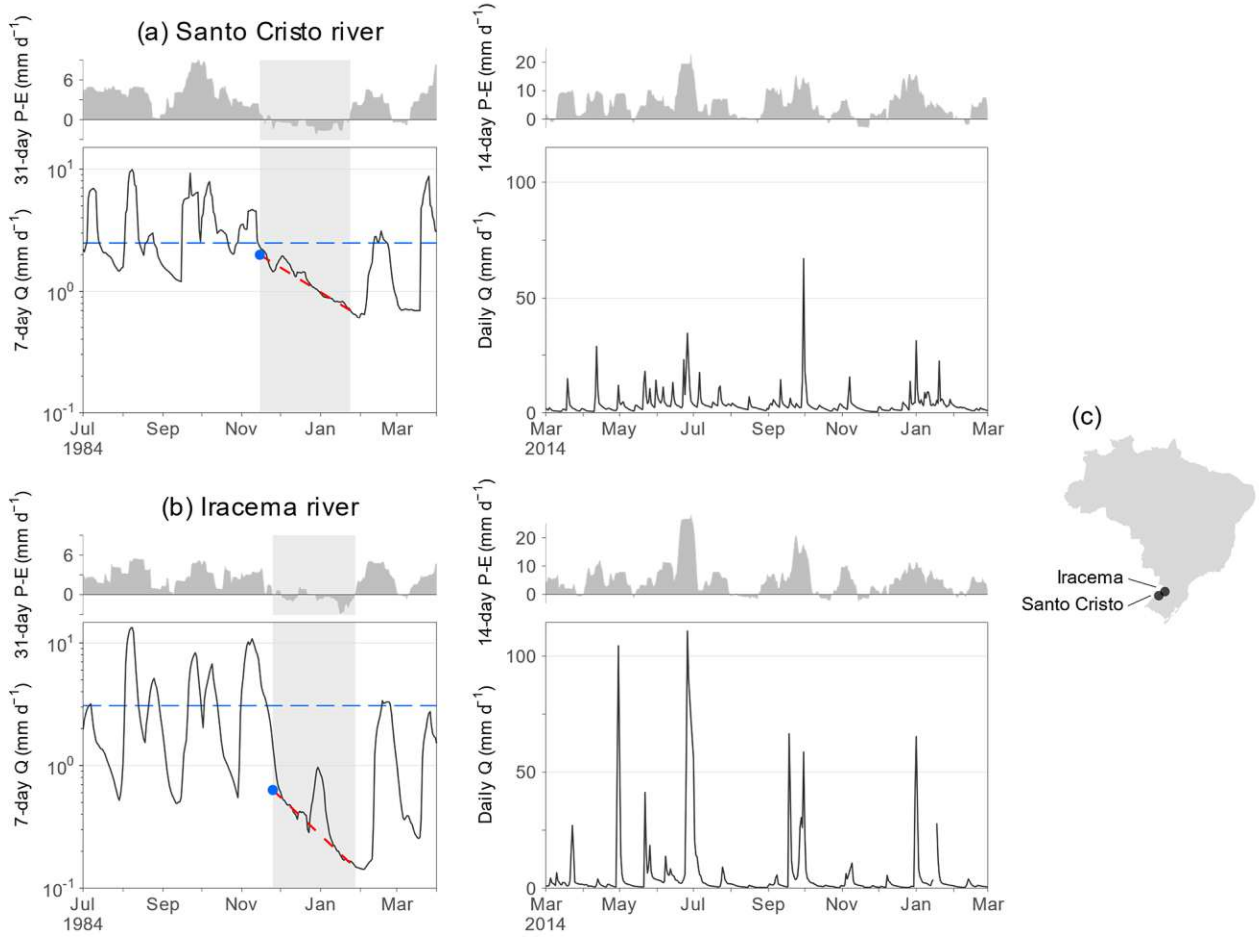

Figure 4.3. Low flow generation on a pair of basins with similar climate and geology but contrasting soils. Time series of streamflow (Q) and rainfall (P) minus evaporation (E) in the: (a) Santo Cristo river (high soil infiltration capacity), 346 km^2, gauge code 74750000; (b) Iracema river (low soil infiltration capacity), 300 km^2, gauge 74295000. Note the different Q and P – E scales between the left and right-hand-side panels. (c) Location map. The blue dashed lines show the mean annual streamflow. The shaded rectangles indicate dry spells. The red dashed lines show the streamflow fit during the dry spell and the blue circle shows the estimated initial flow.

Our model employs a simple exponential decay structure. It prioritizes interpretability while combining the two stages of low flow generation outlined in the perceptual model. The wetting stage determines the initial flow (Q_0) at the start of dry spells, while the drying stage determines the duration (t_{dry}) and rate (λ) of the exponential streamflow decrease.

In the diagnostic model, Q_{min} is estimated with:

$$\hat{Q}_{\mathrm{min},i,j} = Q_{0,i,j} \exp\left(-t_{\mathrm{dry},i,j} \,/\, \lambda_{i,j}\right) \tag{4.1}$$

where, for each basin i and year j, Q_0 (mm d^{-1}) is the streamflow at the start of the dry spell; t_{dry} (days) is the length of the longest dry spell of the year; and λ (days) is the dry-spell flow recession, denoting the streamflow decay constant in terms of how long it takes for streamflow to reduce to $1/e \approx 36.8\%$ of its current value.

In each basin, we define dry spells as continuous periods when the 31-day moving average P – E falls below half the 5-year moving average P – E. We chose the 31-day moving average as a trade-off between short (15 days) and long (91 days) time scales, although the conclusions are not sensitive to this choice (not shown). The threshold of half the 5-year moving average P – E was chosen to accommodate regions where the 31-day P – E rarely falls below zero (e.g., northwestern Amazonia) and where the mean annual P – E has been steadily decreasing in the past four decades (e.g., northeastern Brazil; Chagas et al., 2022). The conclusions are not sensitive to thresholds varying from 0.25 to 0.75 times the 3-year to 10-year moving average P – E (not shown). A dry spell ends when the 31-day P – E is above the threshold for at least seven consecutive days. We assume that the 7-day minimum flow coincides with the longest dry spell of the hydrological year (starting in the month after the highest monthly streamflow).

For each basin's dry spell, we estimate the diagnostic model parameters by fitting a regression line to the streamflow time series at the logarithmic scale. We first extract the streamflow values corresponding to the dry spell, computed as the streamflow in the dry spell period added by the flow wave propagation time. This adjustment accommodates large river basins (i.e., larger than 10^6 km^2) where streamflow can take more than ten days to reach the outlet. Then, we fit the streamflow time series with the regression:

$$\log\left(Q(t)_{i,j}\right) = a_{i,j} + b_{i,j}\, t \tag{4.2}$$

where, for each basin i and dry spell of year j, a is the regression intercept; b is the regression coefficient; and $Q(t)$ is streamflow at time t (days). Equation (4.2) is computed using a quantile regression (Koenker, 2005) with a percentile of 0.25, which captures the streamflow behavior of the recession during drier events. We arrive at the same conclusions with percentiles ranging from 0.1 to 0.75 (not shown). The initial flow (Q_0) and dry-spell flow recession (λ) of Equation (4.1) are computed with:

$$Q_{0,i,j} = \exp\left(a_{i,j}\right) \tag{4.3}$$

$$\lambda_{i,j} = -1/b_{i,j} \tag{4.4}$$

For an easier interpretation of the flow recession, we compute the half flow period ($Q_{1/2}$), that is, the number of days needed for streamflow to halve:

$$Q_{1/2} = \lambda \log(2) \tag{4.5}$$

Note that the recessions computed here deviate from traditional recession analysis (Tallaksen, 1995; Tashie et al., 2020). We allow for short and low-intensity rainfall events and consider only the lower portion of the recession, as the upper portion immediately after storms is automatically removed with our definition of dry spells. Nonetheless, the dry-spell recessions analyzed here are insensitive to rainfall events (e.g., Fig. 4.2-4.3) because the lower portion of the recession is the most stable (McMillan et al., 2022) and the quantile regression is robust to outliers (Koenker, 2005). Streamflow in the dry spells of some of the largest basins in Amazonia, however, result from flood runoff in different parts of the basin. Although the conclusions are insensitive to such basins, large low flow residuals might be observed in large Amazon basins.

4.2.5 Predictive Model

In the predictive model, we first separate the initial flow (Q_0) into:

$$Q_{0,i,j} = \beta_{i,j}\, Q_{\text{mean},i,j} \tag{4.6}$$

where, for each basin i and dry spell of year j, Q_{mean} is the mean annual streamflow; and β is the initial flow coefficient, a constant related to the fraction of the mean annual P – E that recharges the aquifer. We use Q_{mean} instead of mean annual P – E to avoid uncertainties from meteorological data sets (Alexander et al., 2020; Liu et al., 2016).

In the predictive model, we aim to estimate the typical Q_{min} of a basin. Thus, we estimate the median Q_{min} with:

$$\hat{Q}_{\text{min},i} = \hat{\beta}_i\, (P_{\text{mean},i} - E_{\text{mean},i}) \exp\left(-t_{\text{dry},i} / \hat{\lambda}_i\right) \tag{4.7}$$

where, for each river basin i, P_{mean} (mm d^{-1}) is the median mean annual rainfall; E_{mean} (mm d^{-1}) is the median mean annual evaporation; t_{dry} (days) is the median length of the longest dry spell of the year (days); $\hat{\lambda}$ (days) and $\hat{\beta}$ (unitless) are the median dry-spell flow recession and initial flow coefficient, respectively, as estimated with basin characteristics (Section 4.2.6).

4.2.6 Estimating the Flow Recession and Initial Flow Coefficient

We estimate the median dry-spell flow recession ($\hat{\lambda}$) of each basin using the median λ computed from streamflow data (Equations 4.2 and 4.4) as the dependent variable and basin characteristics (Table 4.1) as the independent variables. We use the basin's median value of each characteristic and transformed the categorical variables (e.g., geological class) into numeric variables (e.g., fraction of the basin with sedimentary siliciclastic rocks, fraction of the basin with plutonic intrusive rocks, etc.). The same reasoning applies to the initial flow coefficient ($\hat{\beta}$). We estimate $\hat{\lambda}$ and $\hat{\beta}$ using model-based recursive partitioning, a tree model in which each node has a fitted model (Hothorn & Zeileis, 2015; Zeileis et al., 2008).

The model-based recursive partitioning identifies groups of basins with similar relationships between the $\hat{\lambda}$ (or $\hat{\beta}$) and basin characteristics and, for each of these groups, estimates the dependent variable using linear regressions. It works in the following steps. First, it fits a multivariate ordinary least squares regression between the dependent variable (e.g., λ of each basin) and the independent variables (basin attributes) of the basins included in the present node. The first node contains all basins. Second, it divides the basins into two groups based on the classifying variables and evaluates the instability (i.e., sensitivity) of the regression coefficients. For example, when using the fraction of sedimentary siliciclastic rocks as a classifying variable, the method divides the basins into two groups considering a threshold (e.g., basins with more than 10% of sedimentary rocks), recompute the multivariate regressions for the two groups separately, and verify the differences in the regression coefficients among the two groups (using the outer product of gradients estimator). This process is repeated for many thresholds and for all classifying variables. In the third step, the model selects the classifying variable that leads to the highest sensitivity in the regression coefficients, if statistically significant ($p < 0.05$). Fourth, it finds the threshold of the selected classifying variable that optimizes the regressions (here, using residual

sum of squares). This optimal threshold is used as a split point to split the node into child nodes (i.e., divides the basins into groups in the tree model) and repeat these steps until no further partitioning is significant. After the tree is fully grown, we use the Bayesian information criterion in a post-pruning strategy to remove splits that do not significantly improve the model (e.g., Su et al., 2004), so that the model size can be reduced to facilitate its interpretation.

The flow recessions are estimated using model-based recursive partitioning with the basin fraction of each bedrock type as classifiers (to divide the nodes) and basin characteristics as estimators in each node. The reasoning is that groundwater moves with distinct mechanisms in each bedrock type, such as through pores in sedimentary rocks and fractures in metamorphic rocks. Thus, each bedrock type has a specific relationship between its physical attributes and groundwater release into the stream, which determines the flow recession. For example, in basalts, older formations can be linked with higher rock decomposition and deep groundwater systems (Tague & Grant, 2004), leading to slower flow recessions. This association might not hold for metamorphic rocks, as older formations can lead to longer metamorphism, higher degrees of compaction, recrystallization, and increased resistance to decomposition (Philpotts & Ague, 2009).

When including the basin characteristics as predictors in the tree model, we prioritize model parsimony for an easier interpretation. We only include in the model-based partitioning the basin characteristics with the highest explanatory power. We find the basin characteristics with the highest explanatory power by grouping basins with predominant bedrock types and, for each type, running a stepwise regression between the recession and the basin characteristics. The stepwise regression is performed with a combination of forward and backward variable selection using the Bayesian information criterion (Schwarz, 1978). As a result, we included only elevation, topographic slope, soil sand content, and saturated groundwater hydraulic conductivity as estimators in the tree model.

We estimate the median initial flow coefficient ($\hat{\beta}$) with slightly a different perspective. We use soil properties as classifiers (soil class, depth, sand content, topographic slope, land cover) and climatic indices as the node estimators (maximum annual 7-day rainfall, mean annual P – E, length of the longest dry spell of the year). The reasoning is that the model groups basins with similar permeabilities and, for each group, estimates $\hat{\beta}$ with climatic indices linked to stormflow and groundwater recharge. For example, regions with shallow soils and low permeabilities might be highly sensitive to intense rainfall, leading to higher stormflow and lower $\hat{\beta}$ than regions with deep permeable soils.

4.3 Results

4.3.1 Low Flow Predictions

The observed 7-day minimum annual streamflow (Q_{min}) in Brazil ranges from 0 to 2.7 mm d^{-1}, with a median of 0.37 mm d^{-1} (Fig. 4.4a). The lowest minimum flows (below 0.18 mm d^{-1}, the first quartile) occur in the semiarid northeast region, southwestern Amazonia, extreme south, and floodplains of the Araguaia river. Conversely, the highest minimum flows (above 0.61 mm d^{-1}, the third quartile) occur in northwestern Amazonia, the mountain ranges along the coast of southern and southeastern Brazil, and the highly productive aquifers of Botucatú, Parecis and Urucuia in central Brazil.

The diagnostic model, which uses streamflow data for parameter estimation, explains 96% of the spatial variance in observed Q_{min} (Fig. 4.4b, e). This means that the conceptual model has

great potential to estimate low flows and that streamflow in dry spells indeed follows a simple exponential decay.

The predictive model, which uses climatic and basin characteristics for parameter estimation, explains 61% of the spatial variance in observed Q_{min} (Fig. 4.4c, f). This model's explanatory power increases to 84% when considering non-specific streamflow units (i.e., $m^3\ s^{-1}$, not standardized by basin area). The large-scale spatial distribution of predicted Q_{min} closely matches that of observed Q_{min}, with the same regions exhibiting the lowest and highest values. The estimated Q_{min} of the example basins (Fig. 4.2-4.3) are close to the 1:1 line (Fig. 4.4f), indicating good estimates. The next step evaluates if these estimates are right for the right reasons by examining how the parameters estimated from climatic and basin characteristics led to these results.

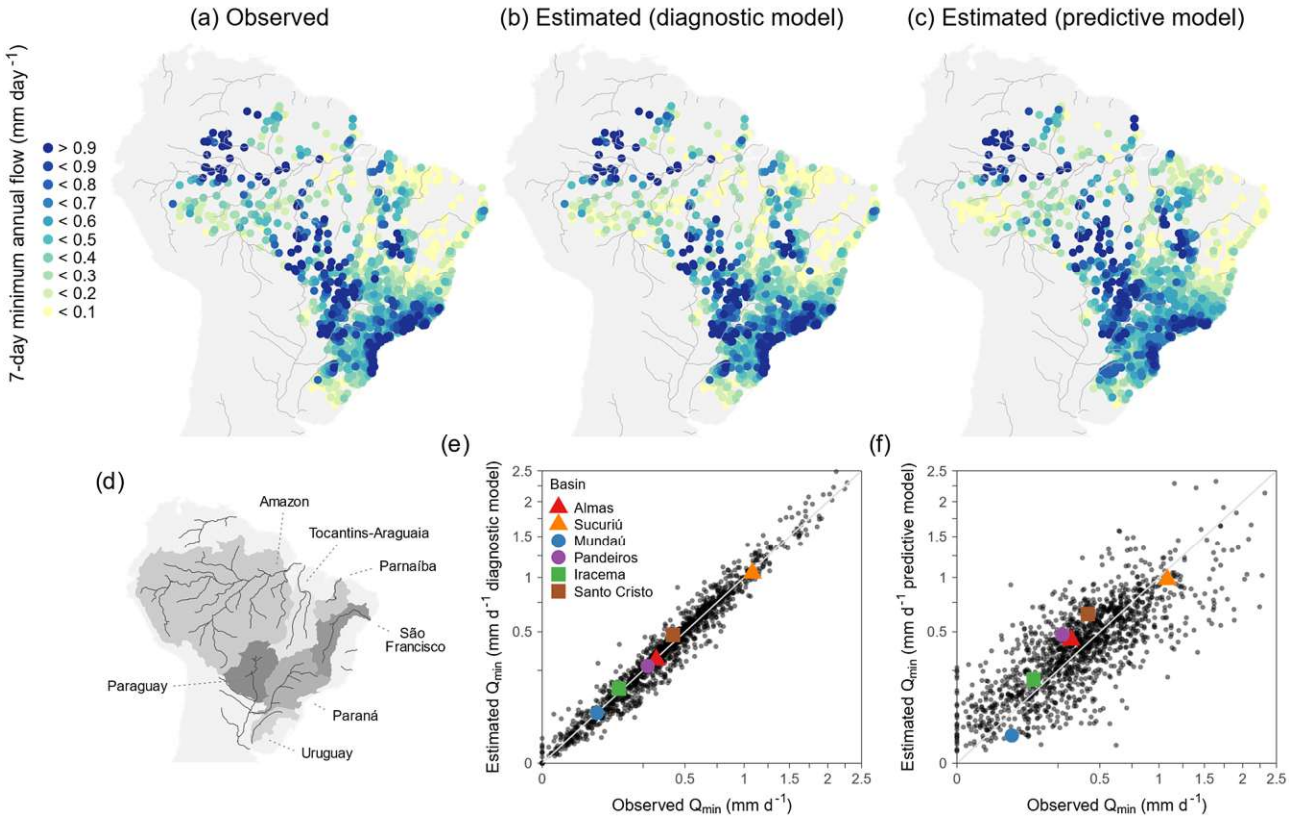

Figure 4.4. Observed and estimated low flow magnitudes of each river basin. (a) Observed median 7-day minimum annual flow (Q_{min}). (b) Estimated Q_{min} with the diagnostic model and (c) predictive model. (d) Location of the major Brazilian basins. Observed against estimated Q_{min} with the (e) diagnostic and (f) predictive models. Each point represents one river basin. The colored symbols indicate the example basins from Fig. 4.2-4.3.

4.3.2 Streamflow Recession

The observed dry-spell flow recession to half flow ranges from 11 to 1,307 days, with a median of 87 days (Fig. 4.5a). The quickest recessions (below 47 days to half flow, the first quartile) occur in the south, part of the northeastern semiarid, the floodplains of southern Amazonia (especially the Madeira basin) and the Araguaia river. In contrast, the slowest recessions (above

139 days to half flow, the third quartile) occur in the most productive aquifers composed of sandstones in central Brazil (Botucatú, Parecis, Urucuia) and the southeastern mountain ranges. Both regions with the quickest and slowest flow recessions are aligned with the lowest and highest Q_{min}.

The large-scale spatial pattern of the estimated flow recessions is similar to that of the observed recession (R^2 of 63%, Fig. 4.5b, c). Notably, the recession estimates of the example basins are accurate (Fig. 4.5c). These include the quick recessions in the southern basaltic basins (Santo Cristo and Iracema rivers), moderate recessions in basins with metamorphic and plutonic rocks (Almas and Mundaú rivers), and slow recessions in the sandstone basins (Sucuriú and Pandeiros rivers).

The model-based recursive partitioning divides the basins into seven geological groups (Fig. 4.5d), with three bedrock types driving most of the recession spatial variability: sedimentary siliciclastic, volcanic, and metamorphic or plutonic rocks.

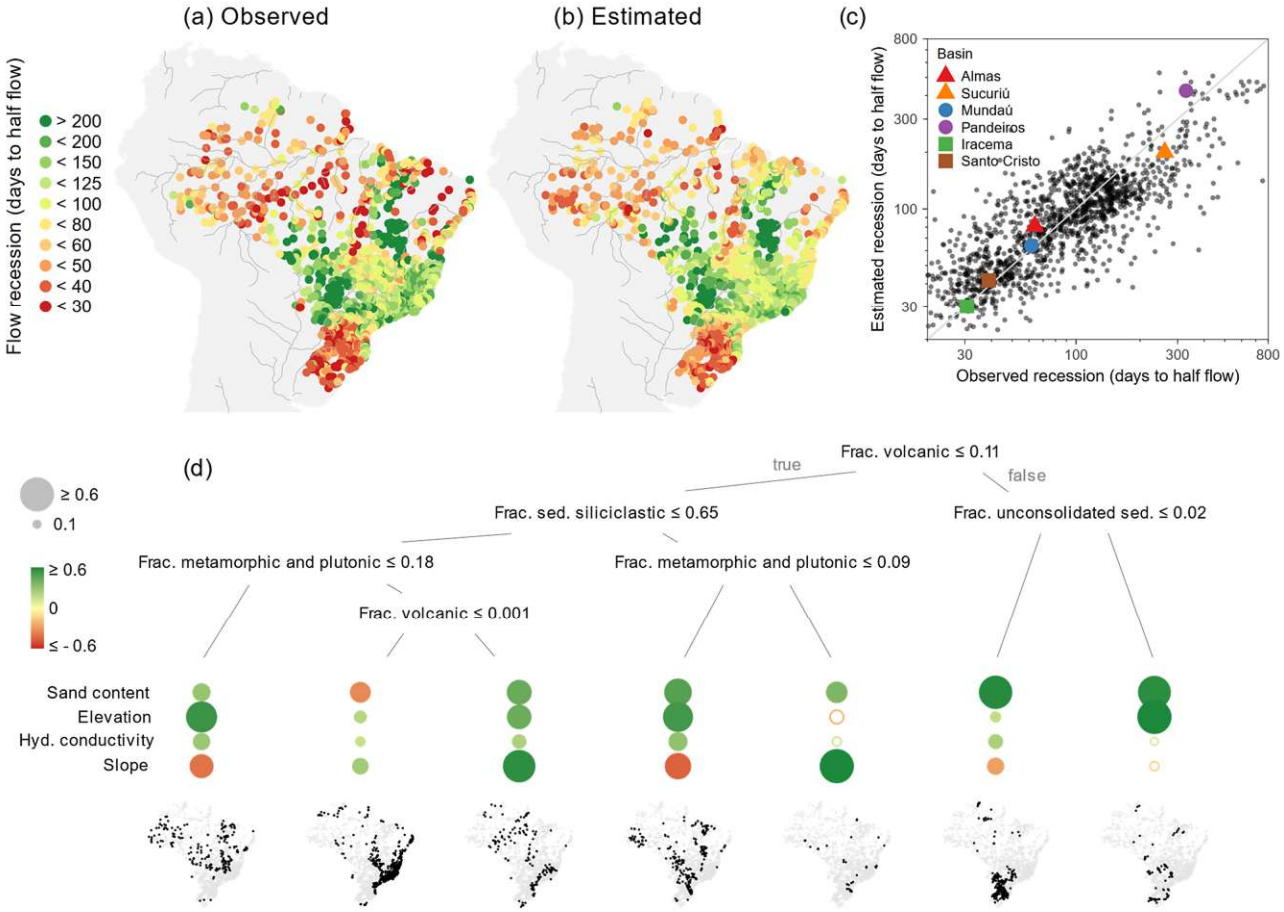

Figure 4.5. (a) Observed and (b) estimated median dry-spell flow recession of each basin. (c) Observed against estimated recessions. (d) Results of the model-based recursive partitioning. The classifiers are basin fractions of bedrock types. The color bars indicate the regression coefficients within each node using standardized values of the basin median sand content, elevation, groundwater hydraulic conductivity and topographic slope. Circle size represent the absolute value of the regression coefficients. Solid and open circles denote significant and insignificant coefficients respectively ($\alpha = 0.05$). The maps show each group's basins.

Basins dominated by sedimentary siliciclastic rocks (groups 4 and 5 in Fig. 4.5d) include the slowest flow recessions (median of 95 days to half flow, standard deviation of 132). They are located on sandstone aquifers (Botucatú, Parecis, and Urucuia; median of 310 days), claystones in southwestern Amazonia (Purus basin; median of 52 days), and shales in southern Brazil (median of 40 days). The flow recession in sedimentary siliciclastic rocks is highly related to soil sand content (Fig. 4.5d), a proxy for the bedrock material. Higher soil sand content indicates sandstones (slow recessions) and lower soil sand content indicates shales or claystones (quick recessions). The flow recession also positively correlates with elevation and have mixed responses to the topographic slope depending on the percentage of metamorphic and plutonic rocks.

Basins dominated by volcanic rocks (groups 6 and 7) exhibit the quickest flow recessions (median of 60 days to half flow, standard deviation of 68). They are found mainly in southern Brazil, underlain by basalts. The recession in these basins relates mainly to soil sand content (Fig. 4.5d). Our interpretation is that higher sand contents are linked to higher fractions of geological formations other than basalts, in particular the sandstones that underlay the basaltic formations in Brazil (Schobbenhaus et al., 2004).

Flow recessions of basins dominated by metamorphic and plutonic rocks (groups 2 and 3) are intermediate between those of sedimentary siliciclastic and volcanic rocks (median of 85 days to half flow, standard deviation of 54). This is expected because fractures in metamorphic and plutonic rocks are more numerous and connected than in basalts, although rock permeability is much lower than in sandstones (Neuman, 2005). In basins dominated by metamorphic and plutonic rocks, the recession is significantly slower in higher elevations and topographic slopes (Fig. 4.5d), interpreted as resulting from a higher number and connectivity of fractures. For example, the mountain ranges along the southeastern Brazilian coast have slow recessions (median of 139 days) due to exceptionally steep terrains, number and connectivity of fractures (Neves & Morales, 2007; Vieira & Gramani, 2015). In contrast, the central Brazilian highlands have quicker flow recessions (median of 79 days) due to landscapes dominated by plateaus with flat terrains (Salgado et al., 2019). Moreover, the recession model indicates a significant negative association between recession and soil sand content. This association contrasts with those for sedimentary siliciclastic and volcanic rocks, although the causes are unclear.

4.3.3 Initial Flow Coefficient

The observed initial flow coefficient (β) ranges from 0 to 2.6 (unitless), with a median of 0.79 (Fig. 4.6a). High values (above 0.90, the third quartile) are found in the central highlands (upper Paraná basin), central-western Brazil (western tributaries of the Paraná basin), Araguaia basin, and southern Amazonia. In contrast, low values (below 0.60, first quartile) occur in the south, southwestern Amazonia, and upper São Francisco basin.

The large-scale distributions of the estimated and observed initial flow coefficients are well aligned (R^2 of 54%, Fig. 4.6a-c). The initial flow coefficient model divides the basins into 16 groups (Fig. 4.6d), with no single predominant variable. The higher number of groups than in the recession model is attributed to a larger number of classifying variables, particularly more soil classes than rock types.

The groups captured by the initial flow coefficient model can be categorized based on soil depth, namely shallow soils (depth below 15.7 m) and deep soils (depth above 15.7 m). Shallow soil basins (groups 1 to 7) have the lowest initial flow coefficients (group median of 0.51), located mostly in southern Brazil. The initial flow coefficient is negatively associated with the 7-day maximum annual rainfall in five out of the seven groups, three of which have a significant

relationship (significance level $\alpha = 0.05$; Fig. 4.5d). This indicates that the soil saturates quickly during storms. The basins include those with (i) leptosols and regosols, characterized by the shallowest profiles in high mountains; (ii) acrisols and alisols, with considerable accumulation of clay in the subsoil; (iii) nitisols, with similar clay accumulation in the subsoil but to a lower degree; and (iv) ferralsols, characterized by highly weathered and well-drained soils, hence the weaker link between the initial flow coefficient and intense rainfall events.

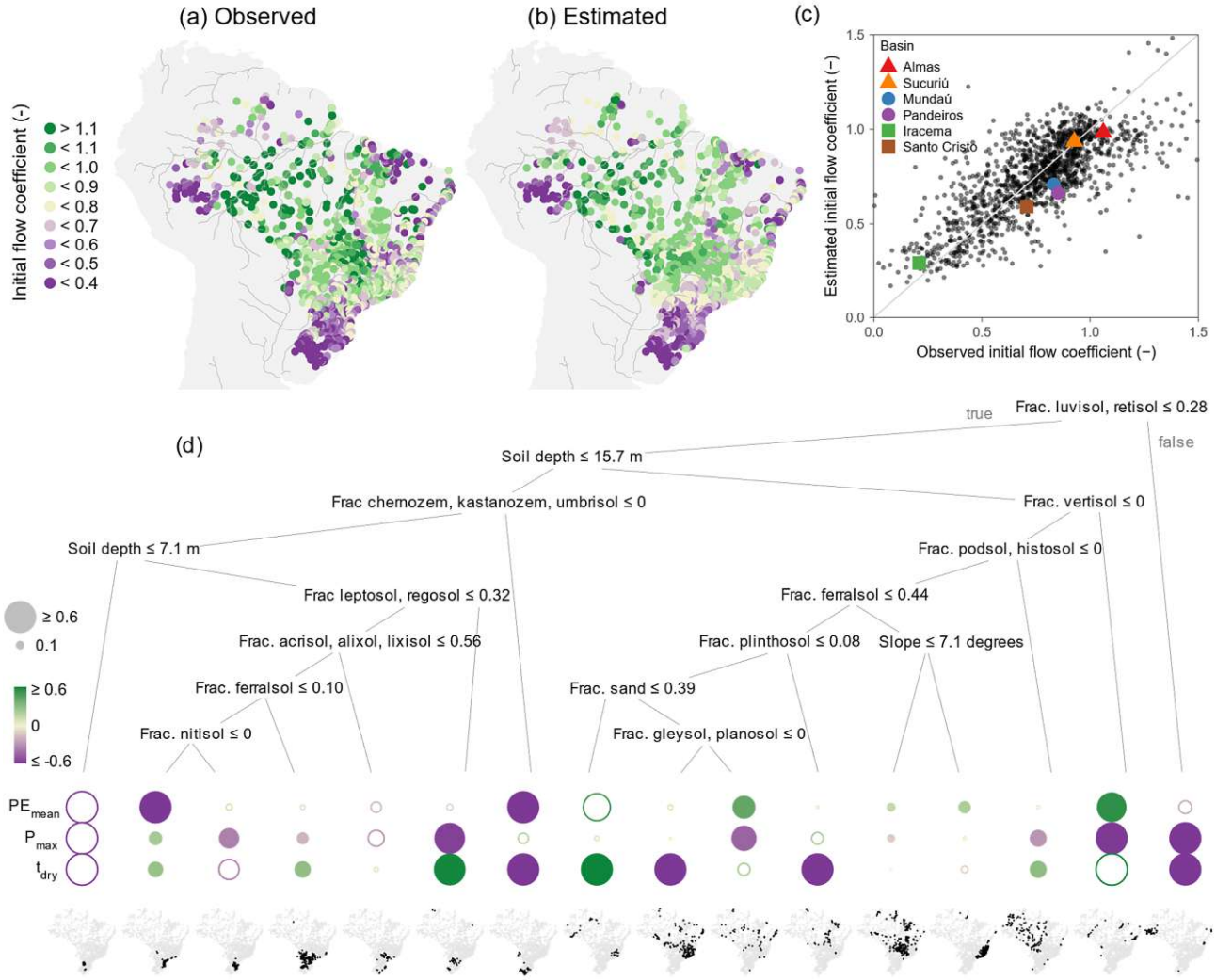

Figure 4.6. (a) Observed and (b) estimated median initial flow coefficient of each basin. (c) Observed against estimated median initial flow coefficient. (d) Results of the model-based recursive partitioning. The classifiers are basin fractions of soil classes, median soil depth, sand content, and topographic slope. The color bars indicate the regression coefficients within each node using standardized values of the basin mean annual rainfall minus evaporation (PE_{mean}), maximum annual 7-day rainfall (P_{max}), and dry spell length (t_{dry}). Circle size represent the absolute value of the regression coefficients. Solid and open circles denote significant and insignificant coefficients respectively ($\alpha = 0.05$). The maps show each group's basins.

Deep soil basins (groups 8 to 15) have higher initial flow coefficients (group median of 0.80) and are widespread in tropical Brazil. In most of these regions, the initial flow coefficient is not sensitive to the 7-day maximum annual rainfall magnitude (Fig. 4.5d) due to their high

weathering rates, infiltration and storage capacities. However, three out of the eight groups exhibit strong and significant negative associations ($\alpha = 0.05$) between the initial flow coefficient and intense rainfall events due to limited infiltration capacity. These include (i) gleysols, characterized by long periods of saturation such as lakes and wetlands; (ii) vertisols, which have low infiltration capacity when wet; and (iii) podzols, which often have low soil storage capacity (WRB, 2015).

4.3.4 Low Flow Generating Mechanisms in the Example Basins

The example basins are well estimated by the initial flow coefficient model (Fig. 4.6c). Together with the accurate flow recession estimates, it suggests that Q_{min} estimates are right for the right reasons. For instance, both the Almas and Pandeiros basins have moderate Q_{min}, although through different mechanisms. On the one hand, the Almas basin has a Q_{min} of 0.32 mm d^{-1} due to a high initial flow coefficient (1.06) linked with permeable deep soils, a wet climate (mean P - E of 1.56 mm d^{-1}, dry spell length of 168 days), but a moderate flow recession (64 days to half flow) linked with metamorphic rocks in flat terrain. It indicates that despite a large fraction of the wet-season P – E recharging the aquifer, it is quickly released into the stream in the long dry season. On the other hand, the Pandeiros basin has a Q_{min} of 0.27 mm d^{-1} due to a high initial flow coefficient (0.85) linked with permeable deep soils, a slow flow recession (329 days to half flow) due to highly permeable sandstones, but a drier climate (mean P – E of 0.45 mm d^{-1}, dry spell length of 194 days). It indicates that even though the aquifer has low recharge volumes due to a dry climate, it is largely mitigated by high storage capacity and slow release into the stream.

Similarly, both the Mundaú and Iracema basins have low Q_{min} for different reasons. The Mundaú basin has a Q_{min} of 0.07 mm d^{-1} due to a drier climate (mean annual P – E of 0.91 mm d^{-1}, dry spell length of 174 days), which is barely mitigated by the high initial flow coefficient (0.83) and moderate flow recession (62 days to half flow). In contrast, the Iracema basin's Q_{min} of 0.14 mm d^{-1} results from an exceptionally wet climate (mean annual P – E of 2.74 mm d^{-1}, dry spell length of 58 days) that does not compensate for one of the lowest initial flow coefficients (0.21) linked with shallow soils and quickest flow recessions (31 days for half flow) linked with basalts.

4.3.5 Scaling Properties of Low Flow Controls

To examine the impact of the spatial scale of analysis (domain size) on the controls of low flow spatial variability, we assess Q_{min} sensitivity to the spatial variability of the diagnostic model's parameters at scales ranging from the regional (smaller than 10^5 km^2) to the continental (2×10^7 km^2) (Fig. 4.7a-b). We achieve this by dividing the study area into blocks, starting at 80^2 blocks (first run) to one block (last run). Each block results from dividing the study area into non-overlapping geographical squares of equal size. At each block, we compute the spatial standard deviation of the Q_{min} controls (i.e., diagnostic model's parameters) of the basin outlets within that block. Then, at each basin, we estimate Q_{min} using the diagnostic model but with the spatial standard deviations added to each parameter. The change in estimated Q_{min} resulted from the added standard deviation, compared with the original Q_{min} estimate, indicates the sensitivity of the spatial variability of Q_{min} to the respective parameter.

The relative importance of the controls of low flow spatial variability changes considerably with the spatial scale of analysis (Fig. 4.7). At regional scales (smaller than 10^5 km^2), physiography controls most of the low flow spatial variability (Fig. 4.7b). Low flows are driven primari-

ly by geology (flow recession), followed by mean annual P and soil (initial flow coefficient) with similar relative importance, and mean annual E and dry spell length as the least important (Fig. 4.7a). At the 10^5 km^2 scale, a change in one spatial standard deviation in the flow recession changes Q_{min} by a median of 39%, followed by 19% and 21% for mean annual P and soil respectively, and 7% and 9% for mean annual E and dry spell length respectively. These values change proportionately at scales ranging from 10^3 to 10^6 km^2.

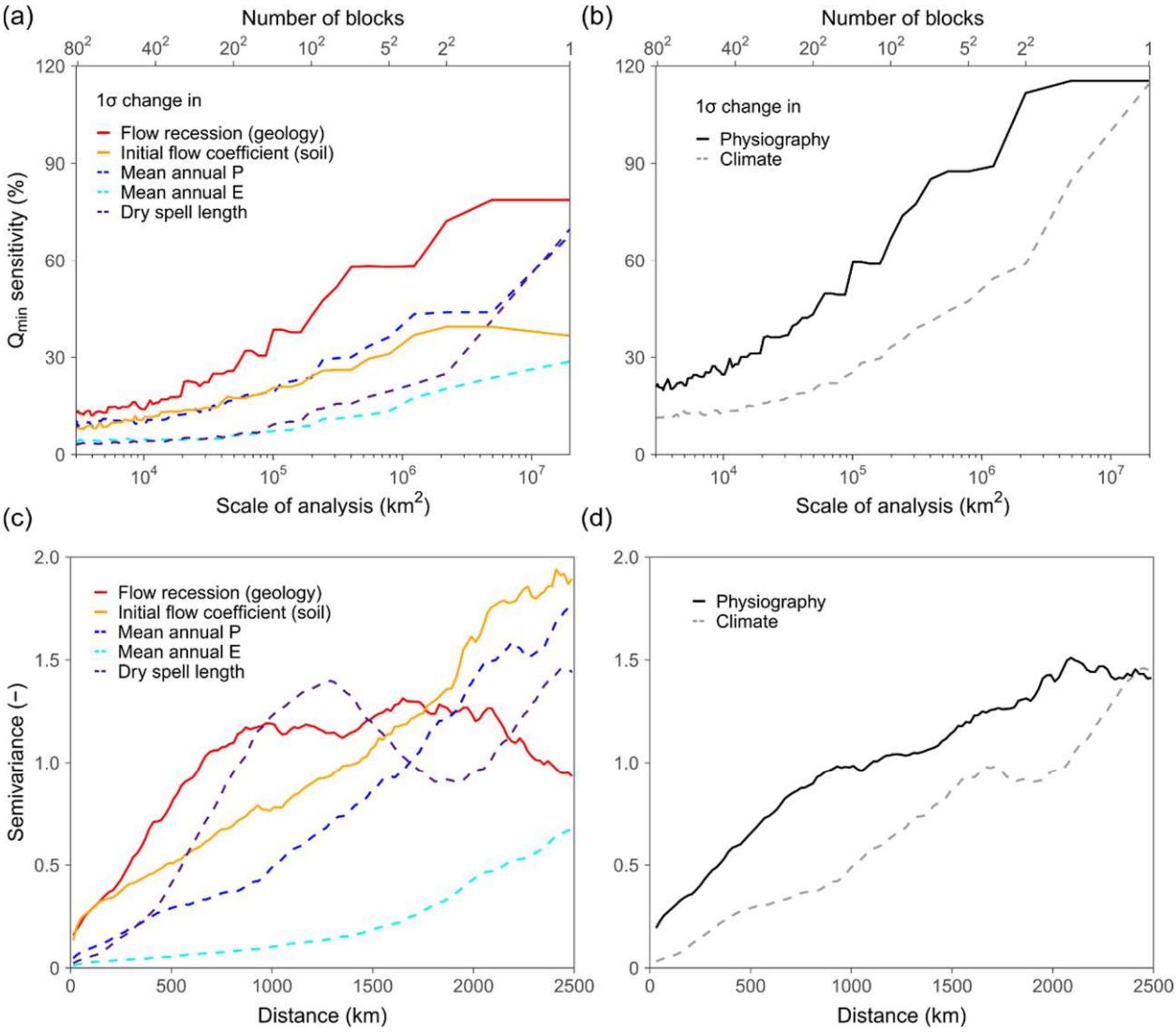

Figure 4.7. Scaling property of the controls of low flow spatial variability. (a) Sensitivity of the 7-day minimum annual flow (Q_{min}) to a change in one spatial standard deviation (σ) in each control as a function of the scale of analysis (domain size). The scale of analysis indicates the average block size and is inversely proportional to the number of blocks used to divide the study area. (b) Same as (a), but for the sum of the basin characteristics (geology and soil) and climatic variables (dry spell length and mean annual P – E). (c) Variogram indicating the spatial variance of the z-scaled Q_{min} controls as a function of distance. The semivariance indicates a change in one σ from the mean value of the respective control. (d) Median semivariance of the basin characteristics and climatic variables.

In contrast, at the continental scale (2×10^7 km^2), basin characteristics and climate hold similar importance for the Q_{min} spatial variability (Fig. 4.7b). The most important control is geology, in which a change in one spatial standard deviation changes Q_{min} by 79%. The most important climatic controls are mean annual P (Q_{min} sensitivity of 68%) and dry spell length (Q_{min} sensitivity of 70%). The least important controls are mean annual E (Q_{min} sensitivity of 29%) and soil (Q_{min} sensitivity of 37%).

The scaling property of low flow controls is related to a larger spatial variance of basin characteristics compared with climate, particularly at small distances. This is seen with variograms, which depict a variable's spatial variance as a function of distance (Fig. 4.7c-d). Up to distances of 1,000 km, the spatial variance of basin characteristics is at least twice that of climate (Fig. 4.7d). This pattern is also seen on maps (Fig. 4.1), where bedrock and soil types change abruptly over short distances while climate changes mainly over large distances. Above 1,000 km, the spatial variance of climate increases at a faster rate than that of basin characteristics. Their spatial variances match at about 2,500 km. This pattern is aligned with the Q_{min} sensitivities to its controls (Fig. 4.7b), where the relative importance of climate grows rapidly after $1{,}000^2$ km^2 scales, eventually matching that of basin characteristics.

4.4 Discussion

4.4.1 Low Flow Predictions

The low flow hydrology approach of this study allows us to uncover the black-box components of traditional low flow statistics by including even a simple conceptual model in the analysis. Our predicted Q_{min} explanatory power is on par with other large-scale studies (Salinas et al., 2013). However, the framework using conceptual models offers (i) opportunities for model improvement, such as including other components, model structure, or machine learning techniques for parameter estimation, and (ii) a better understanding of how climate, geology and soils control the low flow spatial variability.

Part of the unexplained variance in low flow estimation is due to uncertainty in the meteorological data. The water balance does not close in many regions, as the R^2 between the mean annual P – E and the mean annual streamflow is 0.62. By running the conceptual model with an ideal meteorological water balance, that is, using mean annual streamflow instead of mean annual P – E and assuming no water exchange between basins, the predictive model accounts for 73% of the observed variance in Q_{min}.

4.4.2 Geological Controls of Low Flows

Our findings indicate that geology is the most important control of the low flow spatial variability. Geology primarily determines how streamflow behaves in dry periods by regulating the dry-spell flow recession and, ultimately, a basin's potential to mitigate the climate seasonality.

Bedrock type is the main geological control on low flows at large scales. It can be considered a surrogate for groundwater pathways and their connectivity degrees, reflecting if groundwater moves mainly through pores (unconsolidated sediments), a combination of pores and fractures (siliciclastic sedimentary), fractures with high connectivity potential (metamorphic and plutonic rocks), fractures with generally low connectivity (volcanic rocks). Bedrock type can explain why, for instance, increased low flows are linked with the fraction of sandstones (Carlier et al., 2018; Floriancic et al., 2022) or colluvium (Price et al., 2011), both with high porosities

and water storage capacities. The effects of bedrock type on low flows are noticeable even with a coarse geological map such as the one available here (1:1 million scale; Schobbenhaus et al., 2004). Including additional geological data can improve low flow predictions even further, such as by specifying metamorphic rocks with bedding planes (e.g., formed by sedimentary rocks, with horizontal fractures) and without bedding planes (e.g., formed by igneous rocks, with irregular fracture directions).

Other geological and topographic properties are relevant for low flows by regulating the water storage and release of each bedrock type. The relevant properties are highly specific for the bedrock types and most noticeable at small scales. This specificity explains why, for example, low flows are linked with lithological age in basalts (Tague & Grant, 2004) and topographic slope in clay-rich sedimentary rocks (Sayama et al., 2011).

4.4.3 Soil Controls of Low Flows

Our results suggest that soil is half as influential as geology for low flow spatial variability. While geology dictates streamflow behavior in dry periods, soil governs streamflow response to rainfall and the fraction of rainfall that recharges the aquifer. This recharge subsequently impacts the streamflow magnitude at the onset of dry spells, as indicated by the initial flow coefficient. Such a relationship between the initial flow coefficient and the fraction of rainfall recharging the aquifer can also be assessed with the inverse process, stormflow generation. In fact, examining the 1,400 basins in Brazil, we found that the initial flow coefficient is inversely related to the ratio of maximum annual flow to median flow (correlation of 0.70) and positively related to the baseflow index (correlation of 0.78), calculated using the method of Institute of Hydrology (1980) with a five-day time window.

The most critical soil properties for low flows in our study are soil depth and class. Soil depth is related to soil storage capacity, while soil class is a proxy for unavailable large-scale data that drive several hydrologic processes in the tropics. Soil class can indicate the presence of macropores due to soil aggregates, defined by how individual particles of sand, silt, and clay are aggregated into larger particles. Macropores can drive soil infiltration capacity more than soil texture (Beven & Germann, 2013; Fatichi et al., 2020), particularly where aggregates dominate the soil profile such as Brazilian ferralsols (e.g., Madari et al., 2006; Silva et al., 2014; Valani et al., 2022). Soil class can also reveal how the hydraulic conductivity decreases along the soil profile (Bonell, 2004; Fan et al., 2019). For example, acrisols and nitisols have a high clay accumulation in the subsoil, often leading to reduced infiltration capacity and perched water tables (Bonell, 2004). Moreover, soil classes like gleysols and histosols point to regions with wetlands or lakes, which can increase the basin water storage capacity.

Land cover was not deemed a significant variable for low flow spatial variability. Although land cover can greatly impact low flows (e.g., Bruijnzeel, 2004), its relationship with soil properties is inconsistent across all cases. In the tropics, the effects of deforestation on low flows depend on the consequent soil degradation (soil compaction caused by agricultural management and machinery), specific vegetation types before and after deforestation, reduction of soil fauna, and deforestation timing and methods (Bruijnzeel, 2004; Chagas & Chaffe, 2018).

4.4.4 Climatic Controls of Low Flows

Climate is half as important as basin characteristics (geology and soil) in determining low flow spatial variability in Brazil. Mean annual P – E is the most important variable, while dry spell length is the least important. In our perspective, this happens because the influence of the mean

annual P – E on low flows is only moderately regulated by the soil, whereas the influence of dry spell length is highly attenuated by geology. The importance of the mean annual P – E is aligned with the predominance of the aridity index over other climatic indices for low flows (Addor et al., 2018; Beck et al., 2015), as both indices are strongly correlated (correlation of -0.81 for the 1,400 basins in Brazil).

The impact of climate on low flow spatial variability is most evident when comparing Brazil's semiarid and wet regions. In the semiarid (mean annual P – E below 1 mm d^{-1}), Q_{min} is usually low (below 0.2 mm d^{-1}). Reduced water availability leads to restricted groundwater recharge, restricting the geology's role in attenuating the climatic variability. Conversely, in wet regions (mean annual P – E above 1 mm d^{-1}), Q_{min} ranges from low (below 0.2 mm d^{-1}) to high (above 1 mm d^{-1}) as adequate water availability allows geology and soil to have a larger impact on low flows.

4.4.5 Scaling Properties of Low Flows

The dominant controls of low flow spatial variability depend on the spatial scale of analysis. At the continental scale, both basin characteristics and climate are equally important, as they exhibit high spatial variance. However, at subcontinental scales, basin characteristics become the primary control since their spatial variance is higher than that of climate. The scaling properties of low flows differ fundamentally from those of floods (Blöschl, 2022; Viglione et al., 2016). While flood scaling is linked with basin size in which different hydrological processes predominate (Blöschl, 2022), the low flow scaling found in this study is related to the variables' spatial variance captured by the model.

This multiscale spatial analysis reconciles the two hypotheses on whether climate or basin characteristics primarily control low flow spatial variability. On the one hand, large-scale studies emphasizing the control of climate (e.g., Beck et al., 2013, 2015; Addor et al., 2018; Peña-Arancibia et al., 2010; Apurv and Cai, 2020) recognize that climate explains half or less of the low flow spatial variability. These studies argue that the unexplained variability is associated with physiography. On the other hand, small-scale studies that show the predominant control of basin characteristics on low flows (e.g., Bloomfield et al., 2009; Carlier et al., 2018; Floriancic et al., 2022; Pfister et al., 2017; Price et al., 2011) complement large-scale studies by examining low flow generating mechanisms and identifying specific basin characteristics linked with low flows in each location.

4.5 Conclusions

In this study, we develop a new conceptual model to analyze how climate and basin characteristics control low flow spatial variability in 1,400 Brazilian river basins. Using comparative hydrology, we build a model that combines process-based hydrological knowledge with statistics, providing a basic model structure for other methods to build upon, such as machine learning techniques. This framework, referred to as regional low flow hydrology, allows for a more transparent examination of low flow generating mechanisms by addressing the place-specific and scale-dependent properties of low flows. Our key findings are summarized below.

(i) Basin characteristics primarily control the spatial variability of low flows. They are twice as important as climate, indicating that low flows are governed by the basin's ability to mitigate climatic variability through water storage.

(ii) Geological properties are the most important basin characteristics for the low flow spatial variability by determining streamflow recession rates during dry spells. The main control is bedrock type, understood as an indicator of groundwater pathways such as through highly permeable pores (e.g., sedimentary siliciclastic rocks, sandstones) or fractures with low connectivity (e.g., volcanic rocks, basalts). The water storage and release dynamics of each bedrock type are best predicted by soil sand content (a proxy for the bedrock material), elevation, and topographic slope (a proxy for fracture number and connectivity).

(iii) Soil is half as important as geology for the low flow spatial variability. It governs streamflow behavior during storms and the fraction of rainfall that recharges the aquifer. The most significant soil characteristics are soil depth, which relates to soil storage capacity, and soil class, an indicator of relevant properties for tropical hydrological processes such as the presence of macropores and the decrease in the hydraulic conductivity with soil depth (Bonell, 2004).

(iv) The impact of climate on the low flow spatial variability is mainly driven by mean annual rainfall minus evaporation, which is linked with the potential groundwater recharge in a basin. In contrast, the length of the longest dry spell of the year is the least influential component, as its effects on low flows are largely attenuated by geology and water storage dynamics.

(v) The dominant controls of low flow spatial variability depend on the spatial scale of analysis, that is, the study's domain size. Basin characteristics primarily control the variability up to the continental scale (10^7 km^2), where their relative importance matches that of climate. This phenomenon is linked to a higher spatial variance of basin characteristics at smaller scales and comparable variances at large scales.

The multiscale analysis proposed in this study offers a framework for connecting hydrological processes across small and large scales, a primary research topic of the IAHS Scientific Decade 2023-2032 (https://iahs.info/). Furthermore, the conceptual model enables improved low flow predictions under extrapolated conditions, such as those resulting from climate change. The integration of process-based hydrological knowledge and statistics presents a promising approach to understanding low flow generating mechanisms.

Chapter 5

Climate and land management accelerate the South American water cycle

This chapter presents the following publication in its original form. The Methods section is at the end of the chapter because of the requirements of *Nature Communications*.

Chagas, V. B. P., Chaffe, P. L. B., & Blöschl, G. (2022). Climate and land management accelerate the Brazilian water cycle. Nature Communications, 13(1), 5136. https://doi.org/10.1038/s41467-022-32580-x

Abstract

Increasing floods and droughts are raising concerns of an accelerating water cycle, however, the relative contributions to streamflow changes from climate and land management have not been assessed at the continental scale. We analyze streamflow data in major South American tropical river basins and show that water use and deforestation have amplified climate change effects on streamflow extremes over the past four decades. Drying (fewer floods and more droughts) is aligned with decreasing rainfall and increasing water use in agricultural zones and occurs in 42% of the study area. Acceleration (both more severe floods and droughts) is related to more extreme rainfall and deforestation and occurs in 29% of the study area, including southern Amazonia. The regionally accelerating water cycle may have adverse global impacts on carbon sequestration and food security.

5.1 Introduction

Floods and droughts cause more damage worldwide than any other natural hazard (Field et al., 2012; UNDRR, 2019) and their risks may be exacerbated by climate change and socio-economic activities (Field et al., 2012; Schwalm et al., 2017; Winsemius et al., 2016). Often an increase in floods is aligned with a decrease in droughts as a result of more abundant rainfall, and the opposite is the case as rainfall becomes scarcer (Gudmundsson et al., 2019, 2021). However, some models suggest a joint increase in the severity of floods and droughts (Allan et al., 2020; Field et al., 2012; Ward et al., 2020), a phenomenon referred to as acceleration of the terrestrial component of the water cycle. This acceleration could lead to large compound impacts (Zscheischler et al., 2020) on global food production (Brás et al., 2019; Lesk et al., 2016), ecosystem health (Brienen et al., 2015; Phillips et al., 2009), and infrastructure (Ward et al., 2020).

There are a number of processes that potentially cause an acceleration of the water cycle. In a warming climate, the moisture carrying capacity of the atmosphere is increased (Trenberth, 2011) enhancing extreme rainfall (Donat et al., 2016; Fischer & Knutti, 2016) which may increase streamflow during floods. Enhancement of rainfall seasonality (Chou et al., 2013) may decrease streamflow during hydrological droughts. Additionally, the global atmospheric and oceanic circulations are affected (Allan et al., 2020; Durack et al., 2012; Trenberth, 2011). Weaker meridional pressure gradients in a warmer climate may lead to the amplification of stationary waves causing more persistent rainfall and drought periods (Screen & Simmonds, 2014) and rapid shifts between these two regimes (He & Sheffield, 2020; Swain et al., 2018; Wang et al., 2017). Changes in monsoon patterns with increasing contrasts between land and sea surface temperature (Allan et al., 2020) can similarly increase floods and droughts. Land management can also accelerate the water cycle. Agricultural practices can reduce rainwater infiltration into the soil which increases overland flow and thus floods, and reduces groundwater recharge and thus low flows during droughts (Bruijnzeel, 2004). River engineering (Ward et al., 2020), urbanization (Zhang et al., 2018), and groundwater pumping (de Graaf et al., 2019) can have similar effects on streamflow. While there is some evidence for the acceleration of the water cycle over the ocean (Durack et al., 2012), there is little such evidence over land (Greve et al., 2014; Sharma et al., 2018) because of insufficient streamflow data and the confounding effects of the growing human interference in the terrestrial water cycle (Ward et al., 2020).

Here, we analyze a comprehensive hydrometeorological, land cover, and human water use data set in Brazil and show that water use and deforestation have amplified climate change effects on Brazilian streamflow extremes over the past four decades. This region encompasses some of the world's largest basins with mounting concerns of changing floods and droughts (Gudmundsson et al., 2021). Our analysis is based on daily streamflow observations from 886 hydrometric stations (Fig. B.1) for the period from 1980 to 2015. For each station, we compute annual time series of annual minimum 7-day streamflow as a measure of drought flows, mean daily streamflow as a measure of water availability, and annual maximum daily streamflow as a measure of flood flows. We quantify the trend magnitude of each time series (i.e., local trend) with the Theil-Sen slope estimator, the significance of each trend with the Mann-Kendall test, and obtain regional trends by spatial interpolation with ordinary kriging.

For each basin, we consider three climate drivers of streamflow change, computed from daily meteorological data from 1980 to 2015: (i) mean daily atmospheric water balance, computed as precipitation (P) minus evaporation (E, including transpiration from plants); (ii) annual minimum 90-day P – E to indicate droughts caused by seasonal variability; and (iii) annual maximum 14-day P – E because, as basin response times range from less than a day in small basins to a few months in large basins, the 14-day time scale is a compromise on which basins are most sensitive. Additionally, we consider two non-climatic drivers: (i) water use for irrigation and other purposes, and (ii) native vegetation cover. All variables are analyzed in units of mm d^{-1} so that they are independent of basin size, except for native vegetation which is analyzed in % of basin area. Trends in streamflow and their climate drivers are expressed in units of % per decade by dividing each trend by the long-term average value of the same time series. We analyze the links between streamflow changes and their drivers with panel regressions, which allows us to investigate the hydrological variability in space and time in a single framework. We set the regressions with fixed effects for location and use logarithmic-transformed variables. In addition, we explore the acceleration of the water cycle with bivariate frequency distributions with respect to flood and drought flows.

5.2 Results

5.2.1 Detection and attribution of streamflow trends

Our data show that streamflow changes have been widespread in Brazil (Fig. 5.1). Further diminishing low flows (i.e., increasing severity of hydrological droughts) can be found in southern Amazonia and central-eastern Brazil (Fig. 5.1a) while increasing flood flows can be found in Amazonia and in the southeast (Fig. 5.1b). Regional trends in drought flows range from -37 to +16% per decade and those in flood flows from -17 to +10% per decade. Local trends in drought flows at the hydrometric stations (Fig. B.2) range from -65 to +59% per decade and those in flood flows from -39 to +32% per decade. The average trends over the entire domain are -5% and -1% per decade for droughts and floods, respectively. Out of the 886 stations, 353 and 56 stations show significantly ($\alpha = 0.05$) decreasing and increasing drought flows, respectively, and the corresponding figures for floods are 104 and 51 stations.

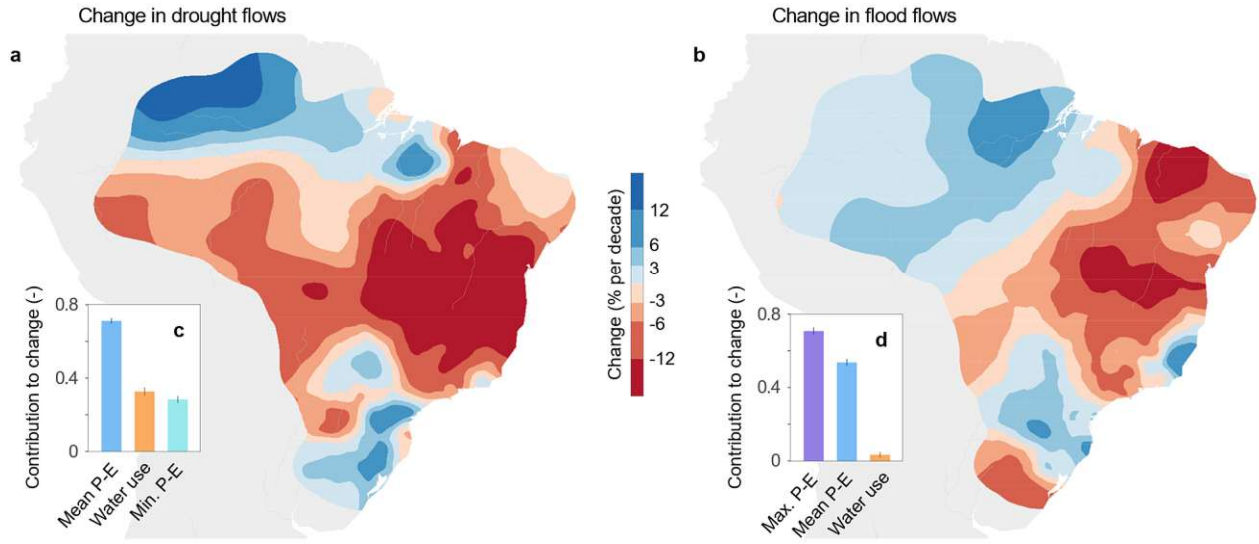

Figure 5.1. Observed streamflow trends and their drivers in Brazil (1980-2015). (a) Change in annual minimum 7-day streamflow (drought flows). (b) Change in annual maximum daily streamflow (flood flows). Blue and red indicate increasing and decreasing streamflow respectively (in % change relative to the long-term drought or flood flow, per decade). (c-d) Contributions to streamflow change in terms of coefficients of two panel regressions between streamflow (n = 25,682 for droughts and 27,299 for floods) and mean daily P – E (precipitation minus evaporation), annual minimum 90-day P – E, annual maximum 14-day P – E, and water use. A coefficient of 0.5 indicates that a 1% change in a particular driver leads on average to a 0.5% change in drought or flood flows. Error bars represent the standard error.

The regression analysis suggests that streamflow change is related to the combined effects of climate variability and increasing water use (Fig. 5.1c-d). Drought trends are driven primarily by changes in mean daily P – E, with substantial effects of water use and minimum P – E (Fig. 5.1c). Water use impacts are noticeable particularly in central-eastern Brazil, where decreases in drought flow and increases in water abstraction are the greatest (Fig. B.3-4). Flood changes are related to maximum P – E and mean daily P – E (Fig. 5.1d), indicating that the floods change in response to modified extreme precipitation and antecedent soil moisture conditions.

To interpret our results, we focus on four hotspots of change with distinct streamflow regimes, land management, and in the upstream areas of major South American basins with mounting environmental concerns such as the Amazon, São Francisco, Paraná, Uruguay and Iguaçu basins (Fig. B.2-B.4). In the southern Brazil and northern Amazonia hotspots, drought flows are aligned with increasing mean P – E and minimum P – E with little land management effect on streamflow (Fig. 5.2). Floods in the southern Brazil hotspot, a subtropical region, have increased in line with increasing maximum P – E and mean P – E. In the Highlands hotspot, a region with intensive agriculture, the reduction of drought flows is aligned with decreasing mean P – E and increasing water use but, from the year 2000 onward, drought flows have become dissociated from mean P – E with a rapid increase in water use (Fig. B.7). In the southern Amazonia hotspot, drought flows have decreased substantially, even though the climatic variables have barely changed, suggesting an effect of large-scale deforestation of the tropical rainforest.

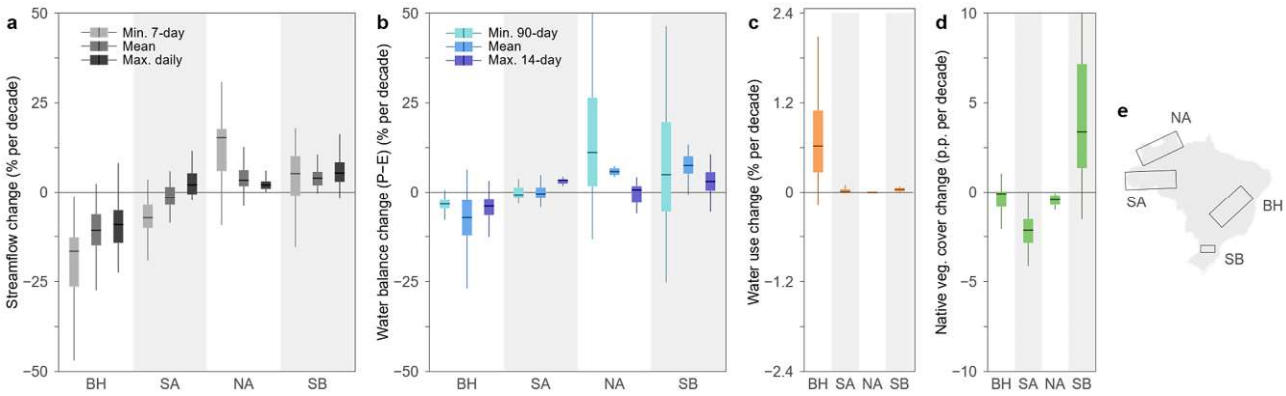

Figure 5.2. Streamflow trends and contributing drivers in four hotspots of change. (a) Streamflow trends, with light grey boxes indicating minimum 7-day flows (drought flows), medium grey indicating mean flows, and dark grey indicating maximum daily flows (flood flows). (b) Climatic trends, with light blue boxes indicating minimum 90-day precipitation minus evaporation (P – E), medium blue indicating mean P – E, and dark blue indicating maximum 14-day P - E. (c) Water use trends in % of the long-term mean daily streamflow per decade. (d) Native vegetation cover trends in percentage points (p.p.) of the total area per decade. The boxplots represent the spatial variability of the local trends within each hotspot. (e) The hotspot locations (Northern Amazonia – NA, Southern Amazonia – SA, Southern Brazil – SB, and Brazilian Highlands – BH). Boxplots show the median value, the first and third quartiles, and 1.5 times the interquartile range. Outliers are not shown.

5.2.2 Four quadrants of streamflow change

Changes in the extremes may not always be synchronized with changes in mean flows. For example, an increase in mean streamflow combined with an increase in the variance of streamflow could lead to increasing high flows but decreasing low flows. Here, we examine how both flow extremes have changed in a single analysis by classifying the trends into four quadrants (Fig. 5.3a). The northern Amazonia and southern Brazil hotspots show increases in flood and drought flows (wetting conditions), which implies that floods have become more frequent and droughts less frequent. The Brazilian Highlands show decreasing flood and drought flows (drying), and southern Amazonia increasing floods and decreasing drought flows (accelerating).

Even though the trends in flood and drought flows are highly correlated (Spearman correlation in the spatial variability of regional trends of 0.61, Fig. 5.3b), there is a tendency towards an accelerating and drying water cycle. A total of 29% of the study area has been accelerating (Fig. 5.3b), which is double the expected percentage of a standardized, bivariate normal distribution with a correlation of 0.61 (i.e., the correlation between drought and flood flow trends; Equation (5.5) in the Methods section). Moreover, 25% and 42% of the study area exhibit wetting and drying trends respectively, whereas 35% would be expected in a standardized, bivariate normal distribution. These figures are quite robust against estimation uncertainty (Fig. B.9).

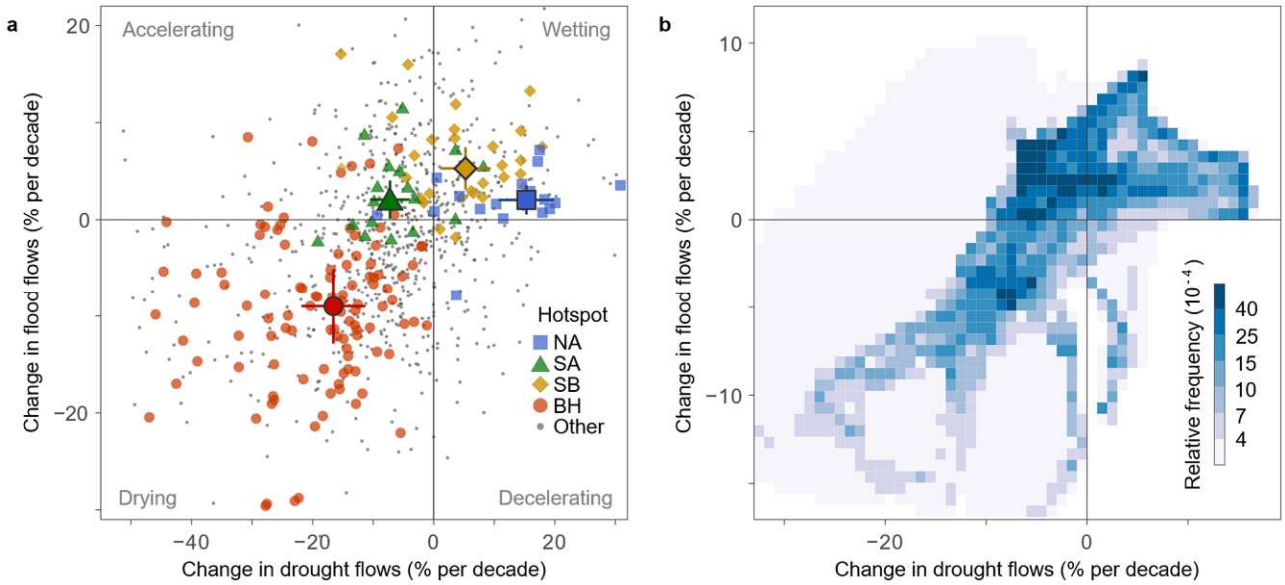

Figure 5.3. Classification of streamflow trends into accelerating, decelerating, wetting, and drying quadrants. (a) Symbols without borders indicate flood and drought flow trends of n = 886 stations. Hotspots (Northern Amazonia – NA, Southern Amazonia – SA, Southern Brazil – SB, and Brazilian Highlands – BH) are indicated by colors. Symbols with borders represent the median trend of each hotspot, and the error bars indicate the median temporal uncertainty of the trend estimates. (b) Classification of regional trends, with darker colors indicating higher areal fraction per bin.

In order to analyze the causes of the acceleration of the terrestrial component of the water cycle, we computed the average trend of each driver from the locations associated with the bins of the bivariate histogram of Fig. 5.3b (Fig. 5.4). Mean P – E trends are strongly positive (on average +5% per decade) and negative (on average -3% per decade) in the wetting and drying quadrants, respectively (Fig. 5.4a), while they are less important in the other quadrants. Increasing water use has amplified the decreasing trends in the drying quadrant (Fig. 5.4c). Virtually all areas where water use has increased by more than +0.5% of the long-term mean flow per decade are in the drying quadrant. The accelerating quadrant is dominated by two factors. The first is increasing trends in maximum P – E which has increased on average by +2% per decade. The second factor is decreasing native vegetation cover. A total of 60% of all areas where native vegetation cover has decreased by more than 2 percentage points per decade are in the accelerating quadrant.

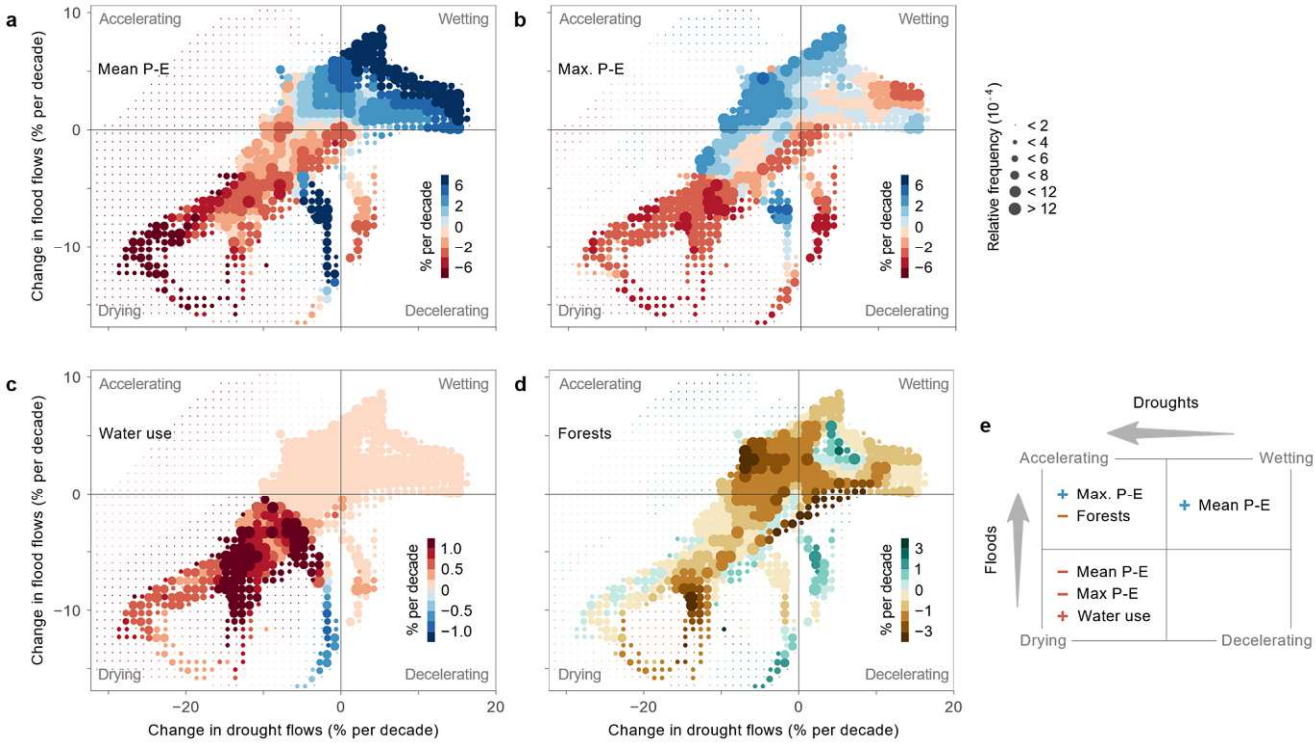

Figure 5.4. Trends of the drivers mapped on the quadrants of the accelerating, decelerating, wetting, and drying streamflow trends of Fig. 5.3b. (a) Mean daily precipitation minus evaporation (P – E). (b) Maximum annual 14-day P – E. (c) Water use in % of the long-term mean daily flow per decade. (d) Native vegetation cover in percentage points (p.p.) of the total area per decade. Larger circles indicate higher areal fractions of trends. Colors indicate the average trends of the drivers for each bin. (e) Schematic of the main drivers of streamflow changes.

5.3 Discussion

While in the past some of the drivers of streamflow change such as climate (Barichivich et al., 2018; Cavalcanti et al., 2015; Cunha et al., 2018; Gloor et al., 2013; Heerspink et al., 2020; Marengo & Espinoza, 2016; Skansi et al., 2013; Zilli et al., 2019) and land management (ANA, 2017, 2019b; Dias et al., 2016; Khanna et al., 2017; Matricardi et al., 2020; Rajão et al., 2020; Staal et al., 2018) have been analyzed individually in South America, here we are showing a clear, spatially coherent signal of streamflow changes that can be interpreted in terms of the compound effects of these drivers. Drying trends are the largest in central and northeastern Brazil (Fig. 5.5). One possible explanation for the change is the southward shift of the South American Convergence Zone (SACZ), a major source of precipitation, which has moved away from central Brazil (Zilli et al., 2019). The drying trends of Fig. 5.5 may also be related to a northward displacement of the Intertropical Convergence Zone (ITCZ), which has moved the equatorial precipitation band farther away from northeastern Brazil (Cunha et al., 2018). Even though the average temperature has been increasing in central and northeastern Brazil over the past four decades (Skansi et al., 2013), evaporation trends have been mostly not significant (Fig. B.5) possibly because of reduced precipitation water supply. An expansion of irrigated agricul-

ture from 15 to 70 thousand km^2 (i.e., by 367%) (ANA, 2017) from 1980 to 2015 has led to a rapid growth of water abstraction, which in 2017 constituted 68% of the total Brazilian water use (ANA, 2019b). Increases in crop productivity and water demands due to a drier climate have boosted agricultural water use even further (Dias et al., 2016). Water abstraction occurs mainly from May to September (ANA, 2017) during the dry season in most of central and eastern Brazil, which has caused a substantial reduction in drought flows.

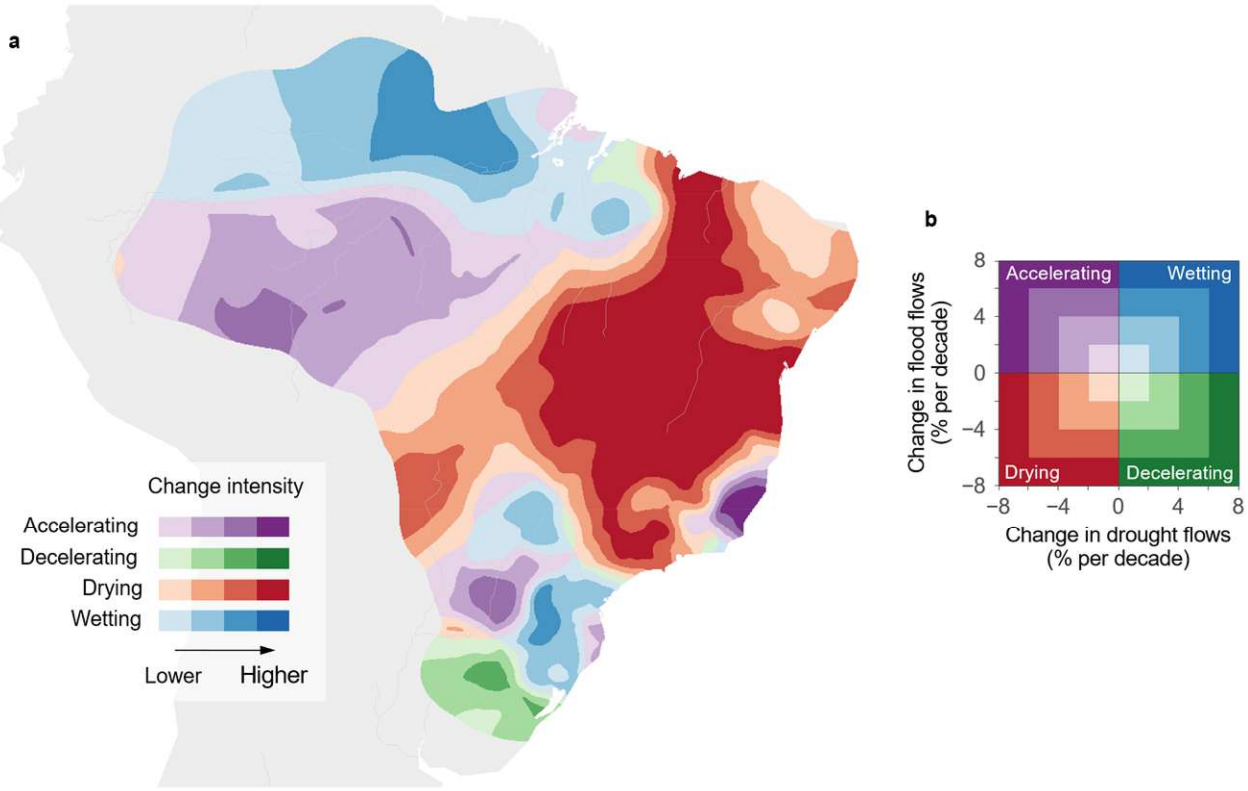

Figure 5.5. Spatial distribution of the accelerating, decelerating, wetting, and drying streamflow trends in Brazil. (a) The location of the four quadrants of regional streamflow trends, with darker colors indicating larger change intensities. (b) Explanation of the color code of (a). Accelerating water cycle has occurred in 29% of the region (2.7 million km^2); deceleration in 4% (0.4 million km^2); drying in 42% (3.9 million km^2); and wetting in 25% (2.4 million km^2).

The northward shift of the ITCZ that has contributed to the reduced precipitation in northeastern Brazil has also contributed to the wetting trends in northern Amazonia (Gloor et al., 2013; Marengo & Espinoza, 2016), even though average temperature (Skansi et al., 2013) and evaporation have increased. On the other hand, the wetting trends in southern Brazil might be associated with stronger effects of the El Niño-Southern Oscillation climate mode (Cavalcanti et al., 2015) and the strengthening and southwards shift of the SACZ (Zilli et al., 2019).

An acceleration of the terrestrial water cycle has occurred extensively in southern Amazonia (Fig. 5.5). The northward shift of the ITCZ is linked to an expansion of dry season length (Marengo & Espinoza, 2016) and warmer temperatures, which have increased evaporation particularly in southwestern Amazonia. Extreme wet-season precipitation and floods have increased as a result of the intensified ascending air masses of the Walker circulation since the 1990s (Barichivich et al., 2018). This intensification has been associated with warming trends of

sea surface temperatures (SST) in the North Atlantic and cooling trends of SST in the tropical Pacific (Barichivich et al., 2018).

Another factor contributing to the acceleration trend of streamflow is deforestation, i.e., the substitution of tropical native vegetation by croplands and pasture, which has caused widespread land degradation in southern Amazonia (Matricardi et al., 2020; Rajão et al., 2020). Land degradation is associated with reduced soil infiltration capacity through soil compaction by agricultural machinery or grazing, reduction of soil fauna, and continued exposure of bare soil (Bruijnzeel, 2004; Rogger et al., 2017). Consequently, surface runoff might increase and groundwater recharge decrease, thus both increasing floods and reducing the baseflow that maintains drought flows in the dry season (Bruijnzeel, 2004). This effect is particularly pronounced where streamflow is highly seasonal and drought flows depend on baseflow (Van Loon et al., 2014; Van Loon & Van Lanen, 2012), as in southern Amazonia. Additionally, deforestation may increase extreme precipitation by triggering convection due to warmer land surface temperatures and increased patchiness (Khanna et al., 2017). On the other hand, deforestation may increase dry season length due to reduced moisture recycling (Staal et al., 2018) thus further extending hydrological droughts. Annual streamflow to rainfall ratios in three small southern Amazonian basins cultivated with soy beans was found to be twice that of neighboring forested basins, flows in the dry season were lower and those in the wet season were higher (Hayhoe et al., 2011) similar to the present study. In contrast, analyses of streamflow in about 50 basins in Amazonia suggest that deforestation has increased low flows, likely because of decreasing transpiration, but without an effect on high flows (Heerspink et al., 2020; Levy et al., 2018), indicating that deforestation may potentially mask the effects of climate change on the water balance.

If the observed changes of extreme streamflow continue into the future in an unabated way, they will have substantial impacts in South America and on the global scale, some of which are already manifesting themselves. The impact will differ depending on where the region falls in the quadrant classification. In the Brazilian Highlands, for example, which lies in the drying quadrant, the drought flow of a 10-year return period has become a 1-year drought (90% CI 1, 2) over the past four decades (Fig. B.10). Such an increase in drought risk threatens agricultural productivity and global food security (Brás et al., 2019). During the 2012-2013 drought in Brazil and the U.S., global soybean prices soared to $550 per metric ton (Ash & Dohlman, 2013) and it is likely that these types of events will happen more frequently in the future. In wetting regions, flood hazards will increase and events such as the 2008 floods in São Paulo city, which caused damages on the order of 110 million USD (Haddad & Teixeira, 2015), may become more frequent. In the urban areas of southeastern South America, those changes may exacerbate the coastal hazards resulting from sea level rise (Bevacqua et al., 2020).

In regions with an accelerating water cycle, the situation of both droughts and floods will deteriorate if current changes continue. In Amazonia, where most of the hydropower potential is still untapped (Latrubesse et al., 2017), future reservoir construction will have to account for the increased flood risk as the average 100-year flood in 1980 in the region has now become a 25-year flood (90% CI 5, 190) (Fig. B.10). Enhanced flooding may increase tree mortality through the inundation of floodplain forests (Gloor et al., 2015), the “Achilles heel” of the Amazonian rainforest (Flores et al., 2017) and this process may be exacerbated by more intense droughts in the same region (Brienen et al., 2015). Reduced tree longevity could accelerate the transformation of Amazonia from a global carbon sink, currently sequestrating 0.4 Petagrams of carbon per year (25% of the terrestrial world total), to a global carbon source (Brienen et al., 2015; Phillips et al., 2009). Reduced vegetation health can further reduce moisture recycling, increase the duration of dry spells and extreme precipitation events, potentially leading to a tipping point of forest dieback (Staal et al., 2018).

Given the evidence for the acceleration of the terrestrial water cycle demonstrated here for Brazil and by global climate model projections (Field et al., 2012; Guimberteau et al., 2017), similar climate and land management changes can also occur in other regions. It would therefore be advisable to conduct observation-based mapping studies globally. The evidence for the acceleration found here also provides an opportunity for Earth System models to attribute the joint changes in floods and droughts to climate, deforestation and water use. In the face of still increasing carbon emissions and agricultural expansion, climate mitigation efforts need to go hand in hand with the adaptation of land management practices, in order to maintain food security and infrastructure safety through the compound risk management of floods and droughts.

5.4 Methods

5.4.1 Streamflow data

We used daily streamflow data from 886 hydrometric stations obtained from the CAMELS-BR dataset (Catchment Attributes and Meteorology for Large-sample Studies – Brazil) (Chagas et al., 2020). The hydrometric stations cover most of the largest river basins in tropical South America (Fig. B.1). The data have been collected following similar measurement protocols as the average of two daily staff gauge readings converted to streamflow through stage-discharge relationships. We selected the study period (1980 to 2015) as a trade-off between the number of stations and consistent record length since most hydrometric stations in the northern and western parts of Brazil were established in the second half of the 1980s. Only the stations that satisfied the following criteria were included in the study: (i) at least 25 years with less than 5% of data missing; (ii) with data starting before 1990; (iii) with data ending after 2005.

We removed hydrometric stations with typographic errors and unrealistically large discharges. Since we are interested in analyzing trends at a large scale, we also removed from the analysis the stations strongly affected by urban land cover (covering more than 10% of the basin area) or reservoirs (stations with a degree of regulation, i.e., the ratio of total reservoir storage capacity to total annual discharge, above 25%). The reservoirs considered were those from the CAMELS-BR data set, which is formed by a combination of data from the Global Reservoir and Dam database (GRanD) (Lehner et al., 2011), ANA's Dam Safety Report 2017 (ANA, 2018) and water bodies identified from Landsat satellite images (Pekel et al., 2016). The catchment boundaries were derived from the Global Streamflow Indices and Metadata Archive (GSIM) (Do et al., 2018; Gudmundsson et al., 2018). Catchment areas range from 11 km^2 to 5,120,000 km^2 with a median of 2,080 km^2.

5.4.2 Climate and land management data

We used daily precipitation time series from CHIRPS v2.0 (Climate Hazards Group InfraRed Precipitation with Station) (Funk et al., 2015) from 1981 to 2015. CHIRPS has a spatial resolution of 0.05° and includes data from rain gauges and satellite sensors. Our choice of using CHIRPS data was based on a comparison of precipitation trends from several gridded datasets (CHIRPS, MSWEP (Beck et al., 2019), PERSIANN (Ashouri et al., 2015), and CPC) with precipitation trends of ANA weather stations (ANA, 2019a). For the comparison, we estimated trends in the ANA dataset using daily precipitation data from 2315 weather stations with at least 25 years without missing values between 1980 and 2015. We interpolated the trends from the

weather stations with ordinary kriging and correlated the interpolation with the trends of the gridded datasets. The trends of the CHIRPS and MSWEP data had the highest correlations with those of the weather stations (Table B.1a) and the median trends of the CHIRPS data were closest to those of the weather stations (Table B.1b).

Daily evaporation time series (including transpiration from vegetation) were obtained from GLEAM v3.3a (Global Land Evaporation Amsterdam Model) (Martens et al., 2017; Miralles et al., 2011) from 1980 to 2015. GLEAM has a spatial resolution of 0.05º and is based on satellite soil moisture data and multiple meteorological products. We also conducted an alternative analysis using evaporation data from ERA5-Land (Copernicus Climate Change Service, 2019; Hersbach et al., 2020) and obtained similar results.

We used global land cover data from ESA/CCI Land Cover v2.0.7 (European Space Agency/Climate Change Initiative) from 1992 to 2015 with a 300-meter spatial resolution and annual temporal resolution. We merged the following land cover classes: forests, shrublands, grasslands, sparse vegetation, and wetland. For simplicity, we denote this merged class as native vegetation cover.

Consumptive water use (i.e., abstracted water that does not directly return to the river basin) was extracted from the ANA's Manual of Consumptive Water Use in Brazil (ANA, 2019b). The data set is composed of monthly water use estimates for each municipality in Brazil from 1931 to 2015, classified into six categories: (i) irrigation, mapped from satellite images and characterized using national agricultural censuses; (ii) livestock, mapped from national agricultural censuses; (iii) households, estimated from the number of people in each municipality; (iv) industry, estimated from the number of employees in each industrial category; (v) mining, estimated from annual production; and (vi) cooling water for thermal power plants, estimated from annual production. Evaporation from reservoirs is not included in the estimates of ANA, thus water use might be underestimated in some regions. We assumed water use in each municipality to be spatially homogeneous and converted the data to a 500-meter grid. Water use outside Brazil was not considered since the data were not available, but the main basins outside Brazil are in western Amazonia which has minor anthropogenic interventions.

5.4.3 Trend analysis

For each hydrometric station we computed annual time series from 1980 to 2015 for the following variables: (i) minimum 7-day streamflow (drought flows), as it is widely used in Brazilian water management and trend analysis worldwide (Blum et al., 2019; Chagas & Chaffe, 2018; Cigizoglu et al., 2005; Dudley et al., 2020; Ehsanzadeh & Adamowski, 2010; Fiala et al., 2010); (ii) mean daily streamflow (water availability); (iii) maximum daily streamflow (flood flows); (iv) minimum 90-day precipitation (P) minus evaporation (E, including transpiration from vegetation); (v) mean daily P – E; (vi) maximum 14-day P – E; (vii) native vegetation cover; (viii) consumptive water use. The meteorological, native vegetation and water use variables are computed considering the contributing basin area of their respective hydrometric stations. The annual time series are computed in units of mm d^{-1} so that their values are independent of basin size, except for native vegetation, which is computed in % of the basin area. Changing the time scales of maximum and minimum P – E did not modify the conclusions; the minimum 90-day P – E had correlations of at least 0.79 with other time scales ranging from 60 to 120 days; similarly, the maximum 14-day P – E had correlations of at least 0.77 with other time scales ranging from 7 to 30 days. Similarly, changing the minimum 7-day streamflow and maximum daily streamflow by the 5th and 95th flow percentiles yielded similar results, as the Spearman

correlations between their local trends are 0.94 and 0.72 respectively. In order to capture both the dry and wet seasons well, we used the water year from March to February for the minimum 7-day flow and minimum 90-day P – E; and from September to August for the other variables.

A linear trend magnitude in each annual time series (i.e., local trend) was estimated with the Theil-Sen slope estimator (Sen, 1968; Theil, 1950) (Fig. B.2-B.5). We evaluated the statistical significance of the trends with the Mann-Kendall test (Mann, 1945). We removed significant (at the 5% level) lag-1 autocorrelation by trend-free pre-whitening (Yue et al., 2002). We multiplied the trend magnitude by 10 to express it in terms of change per decade. The estimated local trend in each series was divided by the long-term average value of its own time series to transform it into units of % change per decade. For example, the lower Madeira river in southern Amazonia (gauge ID 15700000, latitude -5.8167, longitude -61.3019) has a drought flow trend of -0.00588 mm d^{-1} yr^{-1} and a long-term average drought flow of 0.4398 mm d^{-1}, which results in a trend of -13.4% per decade. There are two exceptions to this transformation: (i) native vegetation cover, for which no transformation was necessary because the data is already in % of the basin area, therefore its trends are expressed in percentage points per decade; and (ii) water use, which was instead divided by the long-term mean daily streamflow because it is a more relevant index to relate to water abstractions.

We estimated regional trends (Fig. 5.1, Fig. B.2-B.5) by spatially interpolating local trends with ordinary block kriging using the gstat R package (Gräler et al., 2016; Pebesma, 2004) and the best fit variogram models. The regional trends of the drivers are estimated by interpolating the local trends (which considers the contributing basin area of the hydrometric stations) so that it is consistent with the regional streamflow trends. The blocks are sized 4° by 4° (approximately 445 by 445 km at the equator), which allows for a robust analysis particularly in the Amazon (where gauge density is the lowest) with on average three gauges in each block. Interpolations using block sizes ranging from 1° by 1° to 6° by 6° yield similar results. The uncertainties of the estimated local trends (Fig. 5.3a and Fig. B.2-B.4) were evaluated with bootstrapping (Wilcox, 1998) (α = 0.34) and the uncertainties of regional trends with kriging standard deviations (i.e., kriging errors) (Fig. B.2-B.4). We checked the correlations between trends and catchment area to potentially account for its effect in the interpolation (Hodgkins et al., 2017), but the Spearman correlations were close to null.

To evaluate the possible inflated variance effects on the spatial interpolation due to the spatial correlation between stations, which could lead to an overestimation of regional trends, we repeated the trend interpolation using two subsets of randomly selected stations: (i) using only stations with distances larger than 0.5° from each other (Fig. B.6); and (ii) using only stations with distances larger than 1° from each other (Fig. B.6). The spatial patterns of the trends are similar to those using all stations, with Spearman correlations of at least 0.93 between them. Additionally, we computed the regional Mann-Kendall test (Helsel & Frans, 2006) for the trends in each hotspot of change. Changes in flood flows and drought flows are statistically significant for every hotspot ($p < 0.001$).

5.4.4 Trend attribution

We evaluated the potential causes of streamflow trends with panel regressions (Fig. 5.1c-d), similarly to previous studies (Bassiouni et al., 2016; Blum et al., 2020; Steinschneider et al., 2013). Panel regression includes time series data across multiple cross-sections (i.e., basins) in a single regression framework, allowing us to investigate the hydrological variability both in space and in time. We use fixed-effects (for location) regressions as we are mostly interested in analyzing the impacts of variables over time and as indicated by a significant ($p < 0.001$) Hausman

specification test (Hausman, 1978). We compute two panel regressions, one for drought flows and another for flood flows. The drought flows regression has the form:

$$\ln(Qmin_{i,t}) = \beta_1 \ln(Pm_{i,t}) + \beta_2 \ln(Pmin_{i,t}) + \beta_3 \ln(U_{i,t}) + \beta_4 Veg_{i,t} + \alpha_i + u_{i,t} \quad (5.1)$$

where $Qmin_{i,t}$ is the drought flow (i.e., minimum annual 7-day minimum flow) of basin i at year t; $Pm_{i,t}$ is the mean daily P – E for that basin and year; $Pmin_{i,t}$ is the minimum annual 90-day P - E; $U_{i,t}$ is the mean daily consumptive water use; $Veg_{i,t}$ is the percentage of native vegetation; β_0 to β_4 are the coefficients of the independent variables; α_i represents the intercept for basin i and $u_{i,t}$ represents the idiosyncratic error. The flood flows regression has the form:

$$\ln(Qmax_{i,t}) = \beta_1 \ln(Pm_{i,t}) + \beta_2 \ln(Pmax_{i,t}) + \beta_3 \ln(U_{i,t}) + \beta_4 Veg_{i,t} + \alpha_i + u_{i,t} \quad (5.2)$$

where $Qmax_{i,t}$ is the flood flow (i.e., maximum annual daily flow) of basin i at year t; and $Pmax_{i,t}$ is the maximum annual 14-day P – E. We use logarithms of mm d^{-1} units for all variables except native vegetation as it is already expressed in percent coverage. Therefore, the regression coefficients can be interpreted in relative terms. For example, a 1% change in maximum annual P – E would lead to a β_2% change in flood flows assuming that the remaining independent variables are unchanged. We computed the standardized errors of the regression coefficients with robust covariance matrix estimators (MacKinnon & White, 1985). The regression analysis was performed with the R packages plm (Croissant & Millo, 2008), sandwich (Zeileis, 2006; Zeileis et al., 2020) and lmtest.

The panel regressions were computed in two steps. First, we computed the regressions of Equations (5.1) and (5.2) for the years 1992 to 2015, which is the period covered by vegetation data. Both regressions had null and non-significant ($p > 0.01$) native vegetation coefficients. Thus, we removed the native vegetation terms and computed the regressions a second time including data from 1980 to 2015 (Fig. 5.1c-d). The regressions are robust to changes in the analysis period, with similar coefficients for the two time intervals analyzed (1992-2015 and 1980-2015).

We investigate the interannual variability of streamflow and its drivers in four hotspots with mounting environmental concerns (Fig. 5.2, Fig. B.2). The selected hotspots are located in the upstream areas of major South American basins with distinct streamflow regimes, land and water management. The Brazilian Highlands hotspot has widespread water-intensive crops with increasing drought and water scarcity issues (Bevacqua et al., 2021; Getirana et al., 2021), which covers the most arid regions upstream of the São Francisco and Paraná basins. The Southern Amazonia and Northern Amazonia hotspots have been under large-scale deforestation with potential hydrometeorological impacts (Khanna et al., 2017; Longo et al., 2020; Nobre et al., 2016), particularly in the south where land cover change is the highest (Hansen et al., 2013; Matricardi et al., 2020). The Southern Brazil hotspot has been under increasing flooding in recent decades (Cavalcanti, 2012; Chagas & Chaffe, 2018), which covers the upstream areas of the subtropical Uruguay and Iguaçu basins. We note that the results are robust to variations in hotspot sizes (by ± 20%) and orientations (by ± 20º).

Following the methodology of a previous study (Blöschl et al., 2019), for each hotspot we standardized the annual time series at the stations of each variable to zero mean and unit variance to make the time series comparable within hotspots (Fig. B.7), for example:

$$Q_{i,k}^{0} = \frac{Q_{i,k} - \mu_{Q_k}}{\sigma_{Q_k}} \tag{5.3}$$

where μ_{Q_k} is the mean and σ_{Q_k} is the standard deviation of the streamflow time series for station k, from which we estimated the long-term mean μ_{Qh} for each hotspot and the square root σ_{Qh} of the mean temporal variance. We compared the results between hotspots by denormalizing the series k of each hotspot h:

$$Q_{i,k}^{*} = \sigma_{Qh} Q_{i,k}^{0} + \mu_{Qh} \tag{5.4}$$

The hotspot time series (Fig. B.7) were smoothed using the LOESS method with a smoothing parameter of 0.5.

5.4.5 Quadrant classification

We examine how both flow extremes have changed in a single analysis using the quadrant classification. We classified the trends into four quadrants (Fig. 5.3): (i) wetting, when trends in drought flows and flood flows were positive; (ii) drying, when trends in drought flows and flood flows were negative; (iii) accelerating water cycle, when trends in drought flows were negative but those in flood flows positive; (iv) decelerating water cycle, when trends in drought flows were positive but those in flood flows negative. To analyze the role of the drivers in these changes we first computed a 2-dimensional histogram of regional trends in drought flows and flood flows (Fig. 5.3b). Then, we identified the spatial coordinates included in each 2-dimensional bin of the histogram. For each bin, we computed the average regional trends of the drivers at the associated spatial coordinates and plotted them along with their relative frequencies (Fig. 5.4 and Fig. B.8).

To determine the expected trend frequency in each quadrant, we considered the standardized, bivariate-normally distributed variables $\mathbf{Z_1}$ and $\mathbf{Z_2}$ evaluated with respect to regional trends in drought and flood flows. The quadrant probability can be evaluated (Cramér, 1999) as

$$\mathrm{P}(\mathbf{Z_1} \leq 0, \mathbf{Z_2} \leq 0) = \mathrm{P}(\mathbf{Z_1} \geq 0, \mathbf{Z_2} \geq 0) = \frac{1}{4} + \frac{\sin^{-1}(\rho)}{2\pi} \tag{5.5}$$

where ρ is the correlation of $\mathbf{Z_1}$ and $\mathbf{Z_2}$. A significant correlation between changes in drought and flood flows is expected as they are often consistent with each other and the entire flow distribution moves either upward or downward (Gudmundsson et al., 2019, 2021). Here, we set ρ to 0.61, corresponding to the spatial correlation between the regional trends in drought and flood flows found in the trend analysis in the present study (Fig. 5.3b). According to Equation (5.5), in a random set, 15% of the trends would be expected to fall in each of the accelerating and decelerating quadrants and 35% in each of the drying and wetting quadrants.

To demonstrate the robustness of the results, we examined the sensitivity of trend frequency in each quadrant as a function of the spatial uncertainty in regional trends (ie., the kriging errors) (Fig. B.9). Even if 30% of the locations with the highest average kriging errors were not considered in the analysis, the areal coverage of the accelerating quadrant would only change

from 29% to 24%, which is still well above that of a random sample (i.e., 15%). The drying quadrant becomes even more frequent and the wetting quadrant less frequent as the locations with the highest kriging errors are not considered.

5.4.6 Changes in the return period

In evaluating observed changes in the return periods of drought and flood flows (Fig. B.10), we followed a previous study (Blöschl et al., 2019) where the location parameter of the probability distribution is allowed to change with time. For compatibility of streamflows in catchments of different sizes, this analysis was made using streamflow per unit catchment area. The probability density function $f(x)$ of the annual maximum, x, was estimated for each station using a generalized extreme value distribution (GEV)

$$f(x|\mu_t,\sigma,\xi) = \frac{1}{\sigma}\left[1+\xi\left(\frac{x-\mu_t}{\sigma}\right)\right]^{-\left(\frac{1}{\xi}+1\right)} \exp\left\{-\left[1+\xi\left(\frac{x-\mu_t}{\sigma}\right)\right]\right\}^{-\frac{1}{\xi}} \tag{5.6}$$

where μ is the location, σ is the scale, and ξ is the shape parameter of the GEV distribution. The location parameter, μ_t, changes linearly with time t as

$$\mu_t = a + bt \tag{5.7}$$

The parameters a, b, σ, and ξ were estimated from the maximum flow series using Bayesian inference through a Markov chain Monte Carlo (MCMC) with the Differential Evolution Adaptive Metropolis (DREAM$_{(ZS)}$) (Luke et al., 2017; Vrugt, 2016). Non-informative uniform prior distributions were used for a, b, and σ, whereas a normal distribution consistent with the geophysical prior (Martins & Stedinger, 2000) was used for ξ. We drew 12,000 parameter samples from the posterior distributions, from which 12,000 100-year flood flows in 1980 were calculated for each station by inverting the cumulative distribution function of the GEV and using Equation (5.7) with t = 1,980. The changed return period of these 12,000 flood flows in 2015 were computed using the cumulative distribution function of the GEV and using Equation (5.7) with t = 2,015. Finally, the median of the 12,000 return periods was used as the 2015 return period of the 100-year flood flow in 1980.

In the case of low flows, Equation (5.6) was used after taking the negative of the original minimum 7-day flow series. The parameters a, b, σ, and ξ were estimated using the same MCMC algorithm (Luke et al., 2017; Vrugt, 2016) with non-informative priors for all parameters in this case. We drew 12,000 parameter samples from the posterior distributions, from which 12,000 10-year minimum 7-day flows in 1980 were calculated. The changed return period of these 12,000 drought flows in 2015 were computed using the cumulative distribution function of the GEV and using Equation (5.7) with t = 2,015. Finally, the median of the 12,000 return periods was used as the 2015 return period of the 10-year drought flow in 1980. Those stations for which the 5th and the 95th percentiles of the uncertainty distribution agreed in the sign of change are plotted as large points in Fig. B.10, whereas the remaining stations are plotted as smaller points to indicate the uncertainty involved in the estimation.

Chapter 6

Drought-rich periods were twice as frequent as flood-rich periods in Brazil from 1940 to 2020

This chapter presents the following manuscript under preparation:

Chagas, V. B. P., Chaffe, P. L. B., & Blöschl, G. (2023). Drought-rich periods were twice as frequent as flood-rich periods in Brazil from 1940 to 2020. Under review at Water Resources Research.

Abstract

Streamflow exhibits persistent variability over decades, with extreme events often clustered into anomalous periods. However, it is unclear if the magnitude and spatial extent of these variabilities are symmetric for droughts and floods. Here, we examine time clustering of extreme events by analyzing drought-rich and flood-rich periods in Brazil from 1940 to 2020. We explore data from 319 river gauges and 312 rainfall gauges and contrast streamflow time clustering with rainfall-rich and rainfall-poor periods, water abstractions, the Atlantic Multidecadal Oscillation (AMO) and the Pacific Decadal Oscillation (PDO). Drought- and flood-rich periods are detected by computing annual minima and maxima time series and using scan statistics to verify if events exceeding 5, 10, and 20-year return periods follow a Bernoulli process. Results show a dry period from the 1940s to the 1960s, a shift in the late 1970s with extensive flooding in the 1980s, and another reversal in the 2000s marked by record low flows. Overall, drought-rich periods are significant in 61.1% of the basins, 11.5 times higher than the false positive rate (5.3%) and 2.2 times higher than flood-rich periods (27.9% of the basins). These findings are consistent even when equal numbers of droughts and floods are ensured in each time series. The higher frequency of drought-rich periods is linked with a higher prevalence of rainfall-poor periods (43.6% of gauges) compared to rainfall-rich (29.2% of gauges), a sharp increase in water abstractions since the late 1990s, and a persistence of water storage deficits between years. While streamflow time clustering is commonly detected at scales as small as 6 years, rainfall time clustering is detected mostly at scales above 15 years. At large spatiotemporal scales, drought-rich periods are well-predicted by rainfall-poor periods, the AMO, PDO, and water abstractions (pseudo R^2 up to 0.83 depending on the region). In contrast, flood-rich periods are well-predicted only by rainfall-rich periods. These findings highlight the nonlinearity and asymmetry of drought and flood changes, especially at decadal and multidecadal scales.

6.1 Introduction

The magnitudes of droughts and floods have considerably changed worldwide in recent decades (Gudmundsson et al., 2021; Milly et al., 2015; Slater et al., 2021). These changes are often analyzed through linear trends spanning up to five decades, but as the analysis period increases, hydrometeorological time series start exhibiting decadal to multidecadal variabilities including abrupt shifts and temporal clustering (Blöschl et al., 2020; Cook et al., 2022; Kundzewicz et al., 2019; Markonis & Koutsoyiannis, 2016). Droughts and floods may be clustered into drought-rich and flood-rich periods, when they occur within a few decades, rather than sparsely distributed over the century. Understanding how drought- and flood-rich periods arise and their causes remain an unsolved problem in hydrology (Blöschl, Bierkens, et al., 2019), suggesting that we should go beyond linear trend analyses as they may underestimate other forms of nonstationarity (Hannaford et al., 2013).

The temporal persistence of hydrometeorological time series has been observed across a range of time scales, from decades to millennia. Evidence indicates that, in the 20th century, meteorological drought-rich periods followed cycles of 12-13 years in Europe (Ionita et al., 2012), multidecadal variability in Australia (Kiem & Franks, 2004) and the U.S. (McCabe et al., 2008; Seager, 2015), and cycles of 48 years in Thailand (Buckley et al., 2007). Similarly, flood-rich periods have been identified in Africa (Bola et al., 2022), Australia (Liu & Zhang, 2017), and Europe (Blöschl et al., 2020; Lun et al., 2020; Tarasova et al., 2023). However, precipitation persistence is significantly larger in the southern hemisphere than in the northern hemisphere (O'Connell et al., 2022). In South America, hydrometeorological anomalies have been detected at multidecadal scales (García & Mechoso, 2005; Marengo et al., 2018; Saurral et al., 2017) and centennial scales in reconstructed paleoclimatic records (Bernal et al., 2016; Flantua et al., 2016; Morales et al., 2020; Vuille et al., 2012). Most South American anomalies have been linked with sea surface temperature variabilities in the Pacific and Atlantic oceans (Cavalcanti, 2012; Flantua et al., 2016; Grimm, 2019; Lima & AghaKouchak, 2017; Marengo et al., 2018), especially long-term variabilities in the Atlantic Multidecadal Oscillation (AMO) and Pacific Decadal Oscillation (PDO) (Grimm, 2019).

It is unclear, however, how drought-rich and flood-rich periods are linked with one another and if their occurrence is symmetric in terms of frequency, duration and spatial extent. While research has simultaneously investigated linear trends in droughts and floods (e.g., Chagas et al., 2022a; Ekolu et al., 2022; Gudmundsson et al., 2021), the concurrent analysis of streamflow drought-rich and flood-rich periods and their causes has not been undertaken on a large-scale in Brazil, particularly over periods nearing a century. In this study, we address the following questions:

(i) How frequent were drought-rich and flood-rich periods in Brazil from 1940 to 2020?

(ii) What are their spatiotemporal characteristics?

(iii) How are drought-rich and flood-rich periods linked with rainfall-rich periods, rainfall-poor periods, the Atlantic Multidecadal Oscillation and the Pacific Decadal Oscillation?

We analyze observed daily streamflow data of 319 river basins and rainfall data of 312 gauges from 1940 to 2020. For each gauge, we compute the time series of minimum and maximum annual streamflow and detect drought-rich and flood-rich periods using scan statistics assuming a Bernoulli process (Lun et al., 2020). We analyze extreme events that exceed thresh-

olds of 5, 10, and 20-year return periods. The causes of drought-rich and flood-rich periods are investigated using quasi-Poisson regressions as a function of rainfall-rich periods, rainfall-poor periods, water use for irrigation and other purposes, and anomalies in the AMO and PDO.

6.2 Data and Methods

6.2.1 Data

We examine daily streamflow data from 319 river gauges (ANA, 2019; Chagas et al., 2020) and daily rainfall data from 312 rainfall gauges (ANA, 2019). Our analysis includes only gauges with at least 60 years of data from 1940 to 2020, ensuring that each year has at least 80% of the daily data available. We visually inspected the time series of each gauge, excluding from the analysis gauges with zeroes instead of missing data, orders of magnitude greater than expected, or abrupt changes tied to instrument changes or anthropogenic influence. Furthermore, we restrict our investigation to rainfall gauges located within 50 km of the analyzed river basin boundaries. This constraint ensures a closer match between the spatial distribution of rainfall and streamflow gauges.

The streamflow data are also contrasted with water use and sea surface temperature data. The consumptive water use data include water use for irrigation and other purposes that do not return to the basin (such as via evaporation or incorporation into products), spanning from 1940 to 2017 on an annual scale (ANA, 2019b). The sea surface temperature anomaly data from the Pacific Decadal Oscillation and Atlantic Multidecadal Oscillation cover from 1940 to 2020 at a monthly time scale (Huang et al., 2017).

6.2.2 Detection of flood-rich and drought-rich periods

We detect flood-rich and drought-rich periods using the scan statistics method of Lun et al. (2020). For each river gauge, we compute the time series of maximum annual flow (water year starting in September) and minimum annual 7-day flow magnitudes (water year starting in March). A flood event is defined by when the maximum annual flow of a river basin exceeds a threshold defined by a recurrence interval, here computed for three scenarios of 5, 10, and 20-year return periods (Fig. 6.1a shows 10-year thresholds). Similarly, streamflow drought events are defined by when the minimum annual 7-day flow falls below its corresponding return period thresholds (Fig. 6.1d). The thresholds are estimated with order statistics (Makkonen & Pajari, 2014) as it does not assume a distribution, is less prone to bias, and allows for an exact count of flood and drought events at each time series depending on its length and the return period. Given a median length of the annual time series of 79 years, the three return period scenarios lead to a total of 16, 8, and 4 flood and drought events respectively for each gauge analyzed. Furthermore, the scanning method does not allow for missing values in the annual time series, so we interpolated the missing data in some gauges using linear regression with the highest correlated gauge as a donor. On average, 8% and 13% of the annual streamflow and rainfall time series were reconstructed, which may impact particularly the spatial consistency of the results.

Flood-rich and drought-rich periods are defined by when the occurrence of floods and droughts at the annual time scale deviates from a reference condition of independently and identically distributed series, as modeled by a Bernoulli process (Lun et al., 2020). In other

words, these are periods when flood and drought occurrences are significantly clustered in time. Given an annual time series of length n, we calculate the number of flood events (or droughts) observed in each window of w years that runs from the start to the end of the series. We scan for and identify the window with the most events, marking it as a potential flood-rich or drought-rich period. The probability of occurrence of k or more events within a window, given that a events are observed over the entire time series, is denoted by $P(k \mid w,n,a)$. A larger number of k events within a window correlates with lower probabilities and, if below a significance level α, indicates a significantly anomalous flood-rich or drought-rich period. Thus, flood occurrence (or drought) in a river basin can be either sparsely distributed over the years without a significant time cluster, such as the example shown in Fig. 6.2a, or include a significant flood-rich period (or drought) such as in Fig. 6.2b, where 6 out of 8 floods (1940-2015) were clustered within a 13-year window from 1979 to 1992.

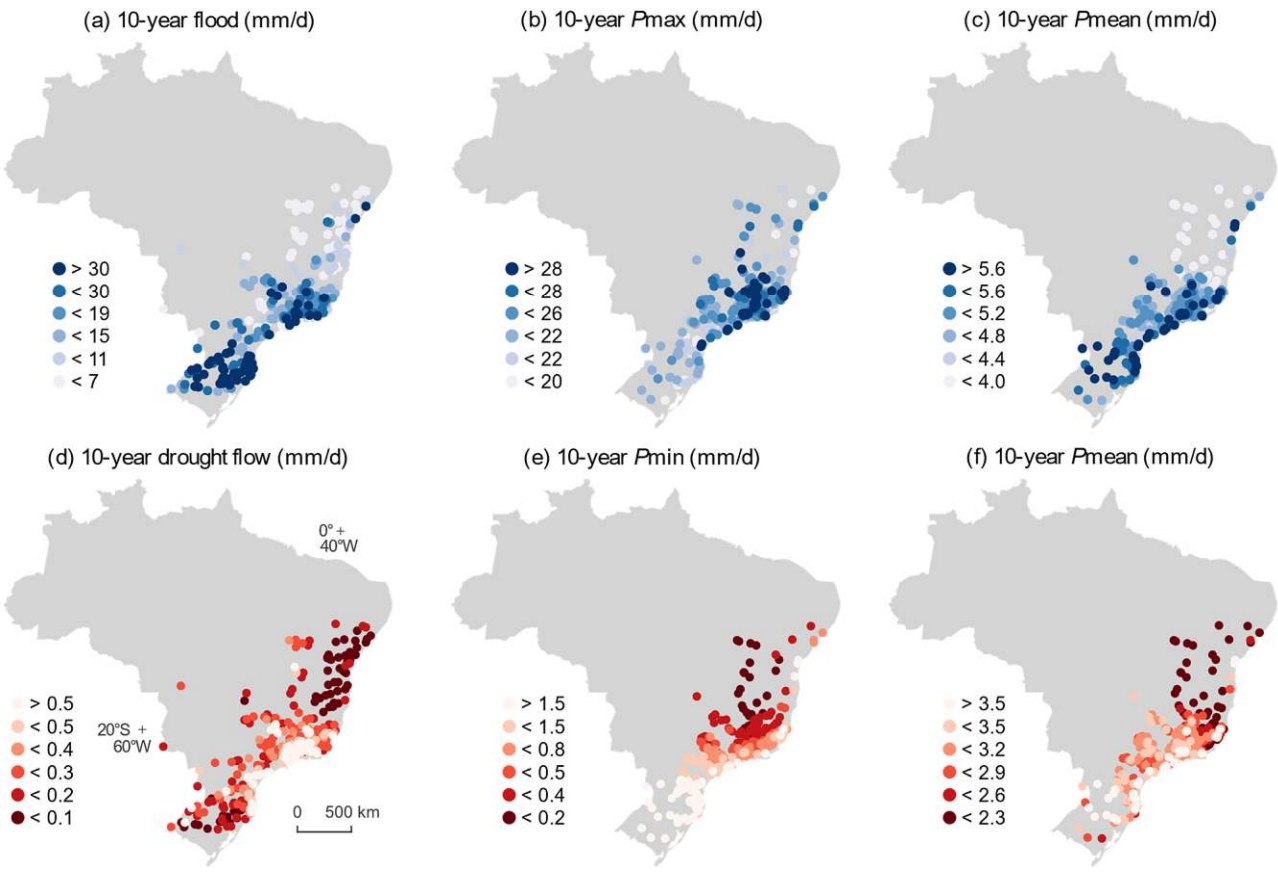

Figure 6.1. Exceedance thresholds of 10-year return period (a) floods, (b) P_{max} (maximum annual 14-day rainfall), and (c) P_{mean} (mean annual rainfall). Corresponding thresholds for when values drop below the 10-year (d) drought flows, (e) P_{min} (minimum annual 180-day rainfall), and (f) P_{mean}.

We consider different window sizes for the return periods of 5 years (w = 6, 9, 12 and 16 years), 10 years (w = 7, 13 and 21 years), and 20 years (w = 7 and 21 years). This is because the results vary considerably depending on the significance level (not shown). The number of significant flood-rich (or drought) periods relative to α is inflated for lower α, particularly those below 0.01. Thus, for comparable results between window sizes and return periods, we selected only window sizes with probabilities close to 0.05, even though they are never exactly 0.05 due

to the discrete nature of the method (see Table C.1 for probabilities of 10-year return period events).

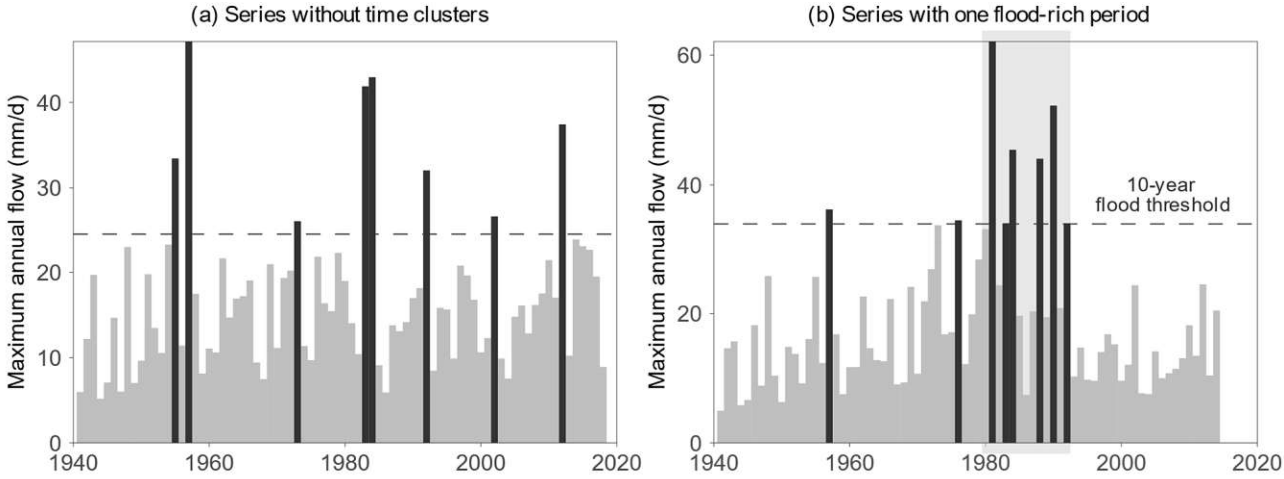

Figure 6.2. Example of a series (a) without significant time clusters (gauge ID 83690000) and (b) with a significant flood-rich period ($p < 0.01$) (gauge ID 83440000). The dashed line is the 10-year return period threshold, the black bars are floods above the threshold and the grey bars are flows below the threshold. The shaded area in (b) is the window when the flood-rich period was detected.

6.2.3 Determining the causes of flood-rich and drought-rich periods

We explore relationships between flood-rich and drought-rich periods with rainfall, water use, and sea surface temperature by contrasting their spatiotemporal patterns, grouping the gauges into regions, and applying quasi-Poisson regression models for the count of gauges with time clusters in each region. Initially, we investigate rainfall-rich and rainfall-poor periods at each rainfall gauge for various indices: P_{max}-rich periods (using maximum annual 14-day rainfall time series, Fig. 6.1b); P_{mean}-rich (abnormally wet mean annual rainfall, Fig. 6.1c); P_{min}-poor (minimum annual 180-day rainfall, Fig. 6.1e); and P_{mean}-poor (abnormally dry mean annual rainfall, Fig. 6.1f). Similar rainfall indices have been linked with trends in Brazil in the past four decades (Chagas et al., 2022a). Next, we group rainfall and river gauges into four regions with similar climates and biomes: South (temperate climate with low seasonality); Center (tropical monsoonal climate); Southeast (tropical coastal and mountainous climates); and Northeast (semiarid). Grouping into large-scale regions lets us focus on the climatic component of floods and droughts and limit local-scale impacts of artificial regulation and anthropogenically-induced changes. In each region, we count the number of gauges with significant flood-rich, drought-rich, rainfall-rich and rainfall-poor periods at each year. Finally, we use quasi-Poisson regressions for each region to model flood-rich counts as a function of P_{max}-rich and P_{mean}-rich counts. Drought-rich counts are modeled as a function of P_{min}-poor, P_{mean}-poor, and the average water use relative to the mean annual streamflow. Similar models are used to explore links between flood- and drought-rich counts, the Atlantic Multidecadal Oscillation and the Pacific Decadal Oscillation, the last two analyzed as long-term anomalies from a 121-month moving average.

Quasi-Poisson regressions between streamflow and rainfall are computed with a linear link between the dependent and independent variables, as count data are present on both sides of the equation and the water use data follows an exponential distribution. In contrast, the regressions between streamflow and sea surface temperature are analyzed with a natural logarithmic link. The quasi-Poisson model is appropriate for our data as it has distinct parameters for the mean and variance, in contrast with the Poisson model. We include an offset in the regressions to transform the count data into relative frequencies ranging from 0 to 100%. Standard errors are calculated using robust covariance matrix estimators (MacKinnon & White, 1985). The statistical significance of coefficients is analyzed with z-scores, computed as the ratio of the coefficient estimate to the standard error. The regressions' goodness of fit is evaluated using a pseudo R^2 of McFadden (1973), which cannot be interpreted as the variance explained by the model. We examined alternative regressions that include interactions among the terms, with results close to those presented here (not shown).

6.3 Results

6.3.1 Frequency of drought-rich and flood-rich periods

Between 1940 and 2020, basins with significant streamflow drought-rich periods were 2.2 times as frequent as those with flood-rich periods (Table 6.1; $\alpha = 0.531$). This ratio considers droughts and floods that exceed a 10-year return period threshold and a statistical significance in at least one window size, ensuring the same total number of droughts and floods are identified in each time series. Drought-rich periods were observed in 61.1% of the basins, 11.5 times higher than the expected value from the type I error of 5.31%. In contrast, flood-rich periods occurred in 27.9% of the basins, 5.3 times the expected value. This flood-rich frequency is similar to those found in Europe, which range from 2 to 4 times above the expected value (Lun et al., 2020; Merz et al., 2016). Overall, 72.1% of the basins had either a significant drought- or flood-rich period, with 16.9% experiencing both drought- and flood-rich periods, although mostly non-concurrent. Thus, over the past 80 years, the assumption of independently and identically distributed series, commonly used in statistical hydrology, is not met in most river basins.

Table 6.1. Frequency of gauges with significant drought-rich, flood-rich periods, P_{max}-rich (maximum annual 14-day rainfall), P_{mean}-rich (mean annual rainfall), P_{min}-poor (minimum annual 180-day rainfall), P_{mean}-poor periods from 1940 to 2020. The expected values represent the significance level (type I error) and the last column is the ratio of the observed frequency to what would be expected by chance. The values presented are for 10-year return period events.

Variable	Total n. of gauges	Frequency (%)	Expected (%)	Freq. / Expected (-)
Drought-rich	319	61.13	5.31	11.52
P_{min}-poor	312	31.73	5.31	5.97
P_{mean}-poor	312	21.47	5.30	4.05
Flood-rich	319	27.90	5.31	5.26
P_{max}-rich	312	12.82	5.29	2.42
P_{mean}-rich	312	20.51	5.30	3.87

Rainfall-rich and -poor periods have also been widely detected (Table 6.1). A total of 43.6% of the rainfall gauges recorded either a significant P_{min}-poor or P_{mean}-poor period. Both rainfall indices are associated with streamflow droughts as, in the tropics, they are commonly caused by dry spells in the season immediately before the drought (P_{min}-poor) or by abnormally low rainfall in the previous wet season resulting in reduced groundwater recharge volumes (P_{mean}-poor) (Van Loon & Van Lanen, 2012). In contrast, 29.2% of the gauges recorded either a significant P_{max}-rich or P_{mean}-rich period, both of which could lead to floods due to extreme rainfall events (P_{max}-rich) or abnormally wet years resulting in humid soils (P_{mean}-rich) (Chagas et al., 2022b).

The number of gauges with time clusters is notably related to the window size analyzed (that is, the time scale) and the return period threshold (Fig. 6.3). The frequency of time clusters decreases as the thresholds for droughts and floods get more restrictive, that is, higher return periods (Fig. 6.3). Drought-rich periods were detected in at least one window size in 73.4% of the basins considering 5-year droughts, 61.1% for 10-year droughts, and 49.8% for 20-year droughts. For floods, these values are 33.2%, 27.9%, and 17.2%. A similar pattern is observed for rainfall extremes. However, the relationship between the frequency of time clusters and window size varies considerably for streamflow and rainfall. While the frequency of drought- and flood-rich periods is stable across window sizes, the frequency of rainfall-rich and -poor periods is close to zero in smaller windows and nearly as high as streamflow clusters in larger windows. Streamflow time clusters are common at time scales as small as 6 years, but rainfall time clusters are mostly observed in time scales above 15 years. This suggests that streamflow extremes exhibit a stronger interannual dependency compared to rainfall extremes.

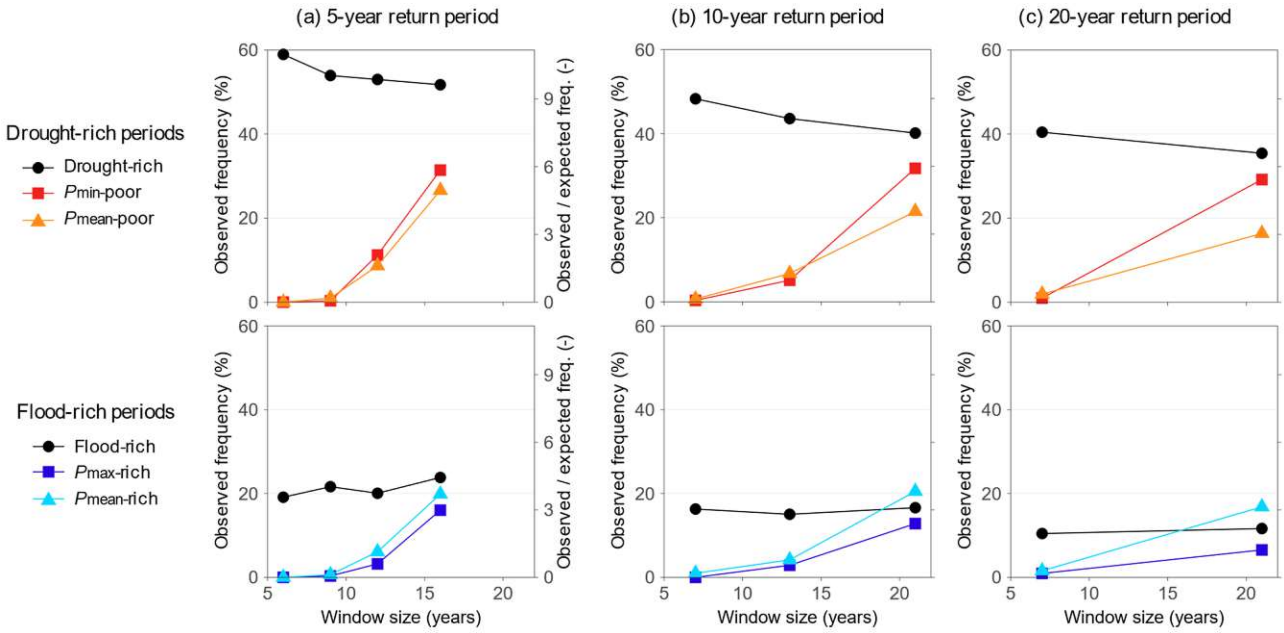

Figure 6.3. Frequency of gauges with significant drought-rich, flood-rich, rainfall-rich and rainfall-poor periods as a function of the running window size (that is, the time scale) and return period thresholds of 5, 10, and 20 years.

6.3.2 Spatiotemporal distribution of drought-rich and flood-rich periods

Drought-rich periods were observed across most of the study area, while flood-rich periods were observed mainly in the south, southeast, and center (Fig. 6.4a, 6.4d, 6.4g). The results are shown for 10-year return periods but are similar for 5 and 20-year return periods. On the other hand, rainfall time clustering shows a different pattern, with a clear spatial configuration only for 21-year windows. The northeast has the lowest frequency of rainfall time clusters. In the southeast and center, there is a higher frequency of P_{min}-poor periods compared with P_{max}-rich (Fig. 6.4h), contrasted by a lower frequency of P_{mean}-poor periods than P_{mean}-rich (Fig. 6.4i).

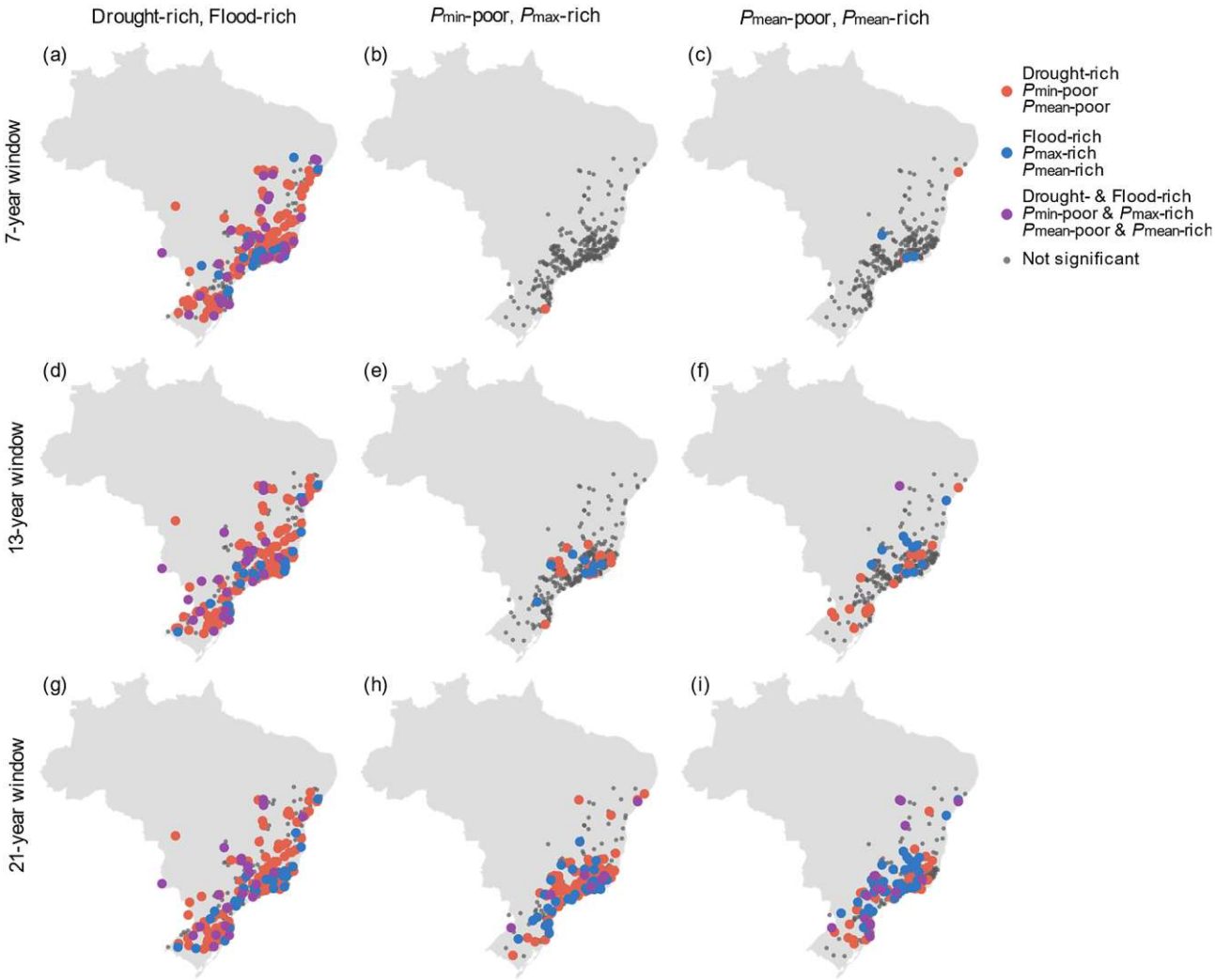

Figure 6.4. Spatial distribution of gauges with significant time clusters for (a-c) 7-year windows, (d-f) 13-year windows, and (g-i) 21-year windows considering 10-year return period thresholds.

Drought-rich and flood-rich periods have changed considerably since 1940, exhibiting three main cycles: a dry period in the initial decades; a shift in the late 1970s with frequent flooding; and another reversal in the 2000s marked by a predominance of droughts. The temporal evolution can be seen in Fig. 6.5, showing the spatial distribution of time clusters in each decade, and in Fig. 6.6, which displays the number of gauges with a significant time cluster in each year in four regions.

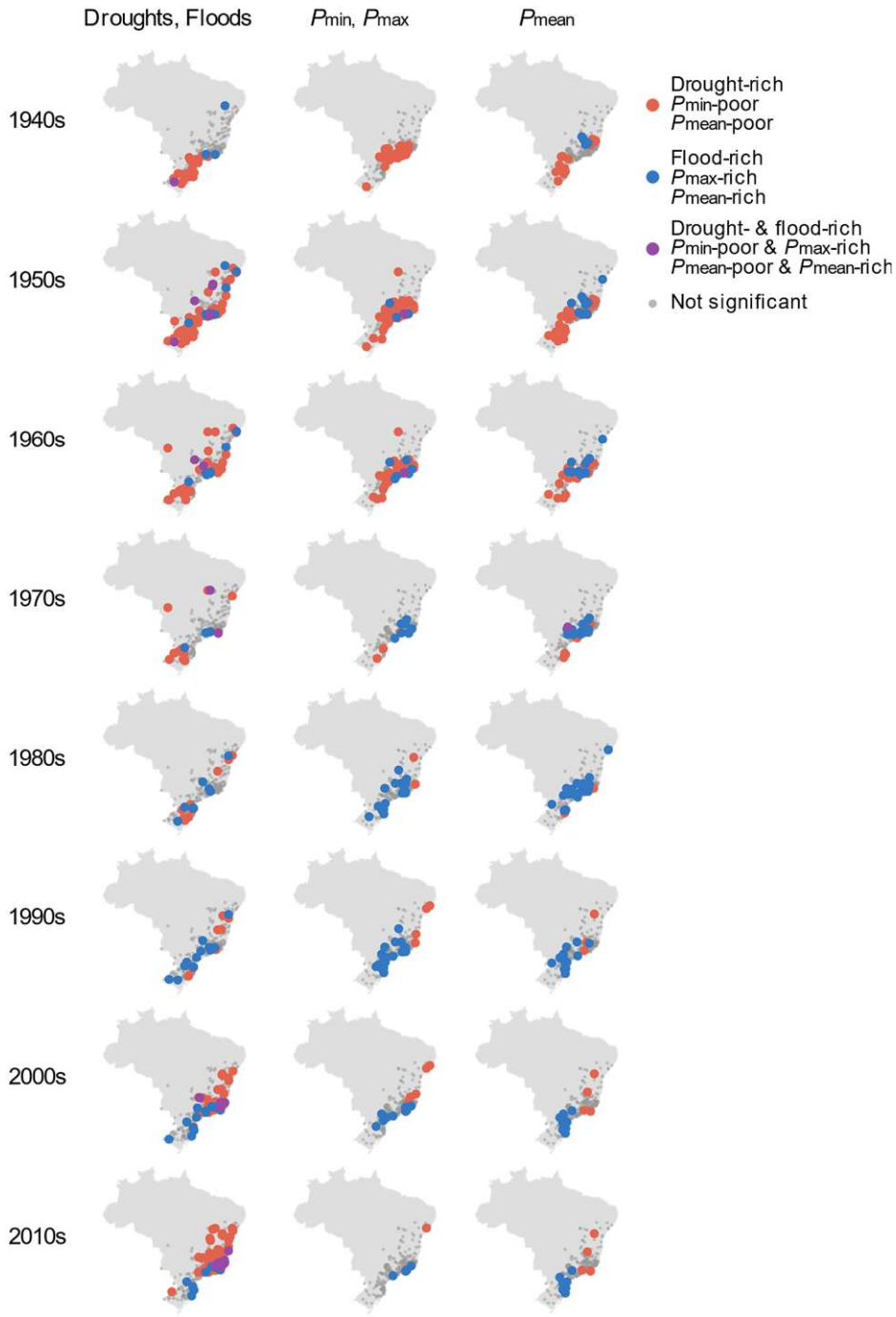

Figure 6.5. Spatial distribution of gauges with significant time clusters in each decade, considering 10-year return period thresholds and a significance in at least one window size. The gauges with significant clusters are plotted only in the decade that includes the central year of the cluster window.

The first dry period extends from the 1940s to the 1960s (Fig. 6.5), with a peak number of basins with significant drought-rich periods in the 1950s (Fig. 6.6). The drought-rich periods are typically aligned with rainfall-poor periods, more prevalent in the south (Fig. 6.6c) and less frequent in the southeast (Fig. 6.6a). However, most records of this period's socioeconomic

impacts are from the northeast, as it is the most arid and susceptible region, with consequences such as mass emigration and economic losses (Marengo et al., 2017; Ponce, 1995).

The dry period was reversed in the late 1970s and early 1980s, with an increased number of basins with flood-rich periods mainly in the south and central regions (Fig. 6.6c, 6.6d). The 1980s was the wettest period in a large part of Brazil since 1940, witnessing some of the largest floods recorded (Fleischmann et al., 2020). The shift from dry to wet period started with an increase in rainfall-rich periods, linked with a shift in the Pacific Decadal Oscillation in the late 1970s (Carvalho et al., 2011; Jacques-Coper & Garreaud, 2015), later followed by flood-rich periods (Fig. 6.5). This delay between the rainfall-rich and flood-rich periods suggests that basins may take several years to recharge the groundwater deficit accumulated from previous droughts.

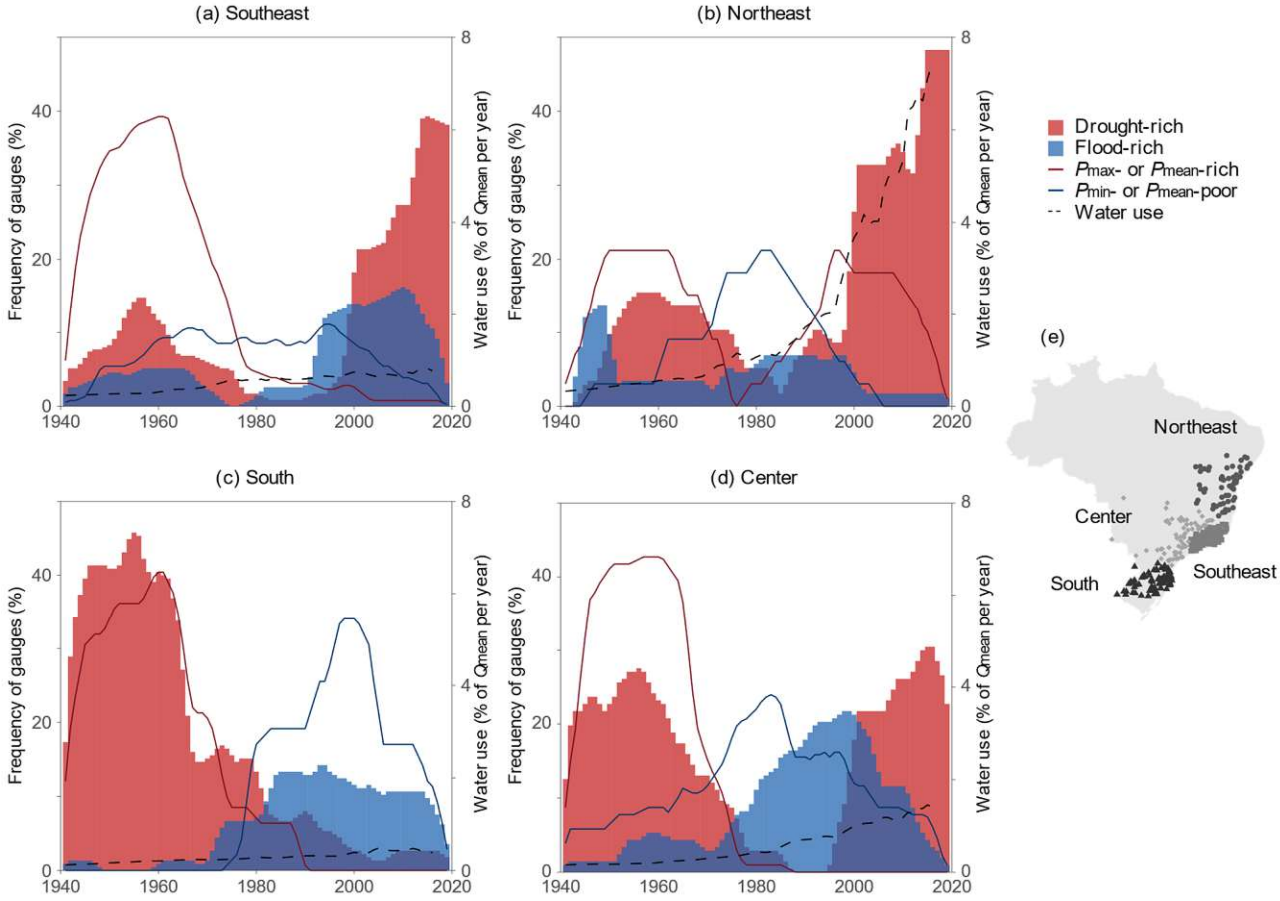

Figure 6.6. Time series of the frequency of gauges with significant drought-rich, flood-rich, rainfall-rich and rainfall-poor periods in (a) the southeast (number of river gauges n_q = 117, number of rainfall gauges n_p = 129), (b) northeast (n_q = 58, n_p = 33), (c) south (n_q = 75, n_p = 47), and center (n_q = 69, n_p = 103). Each year counts the number of gauges that were in a significant time cluster.

The second dry period started in the 2000s, with an increased frequency of basins with drought-rich periods in most regions except for the south (Fig. 6.6). There were more basins with significant drought-rich periods than in the first dry cycle before 1970, with frequencies up to 48% of the basins in the northeast region, 39% in the southeast, and 30% in the center (Fig. 6.6). Streamflow in the 2010s often reached record lows (Cuartas et al., 2022; Getirana et al., 2021),

resulting in severe water crises like those in São Paulo in 2014-2015 (Braga & Kelman, 2016) and partial disruption of the national hydropower generation (Tomasella et al., 2022). However, rainfall-poor periods were not as frequent as drought-rich periods after 2000. The higher frequency of drought-rich basins can be related to a combination of lower rainfall volumes (which are insufficient for a significant rainfall-poor period) and a rapid increase in water abstractions. This is especially evident in the central (Fig. 6.6d) and northeastern regions (Fig. 6.6b), which aligns with the linear trend analysis of Chagas et al. (2022a) for the period 1980-2015. Furthermore, a part of Brazil, close to the Doce river basin in the southeast, experienced an increase in both drought-rich and flood-rich periods simultaneously (Fig. 6.5), suggesting that the acceleration of the water cycle (Chagas et al., 2022a) may be perceptible even in the most extreme events.

6.3.3 Links with rainfall, water use, and sea surface temperature

Two out of the four regions analyzed had a good alignment between the time series of drought-rich periods, rainfall-poor periods and water use in terms of quasi-Poisson regressions (Fig. 6.7a). In the northeast, the frequency of basins with drought-rich periods is well predicted by the frequency of P_{mean}-poor periods and water use (pseudo R^2 of 0.83, $p < 0.01$), with increasing importance of water use in the last two decades. In the south, the impact of water use is not as significant, where the frequency of drought-rich periods mirrors P_{mean}-poor periods (pseudo R^2 of 0.93, $p < 0.01$). The higher importance of P_{mean}-poor over P_{min}-poor period suggests that drought clusters are primarily linked with decreases in groundwater recharge in the wet season than with rainfall deficits in the dry season. While the regressions are explored for 10-year return periods, the outcomes of 5 and 20-year return periods show good predictions for the same two regions. However, P_{min}-poor periods are the predominant variable for 20-year return periods, indicating that different mechanisms might be responsible for droughts of different magnitudes.

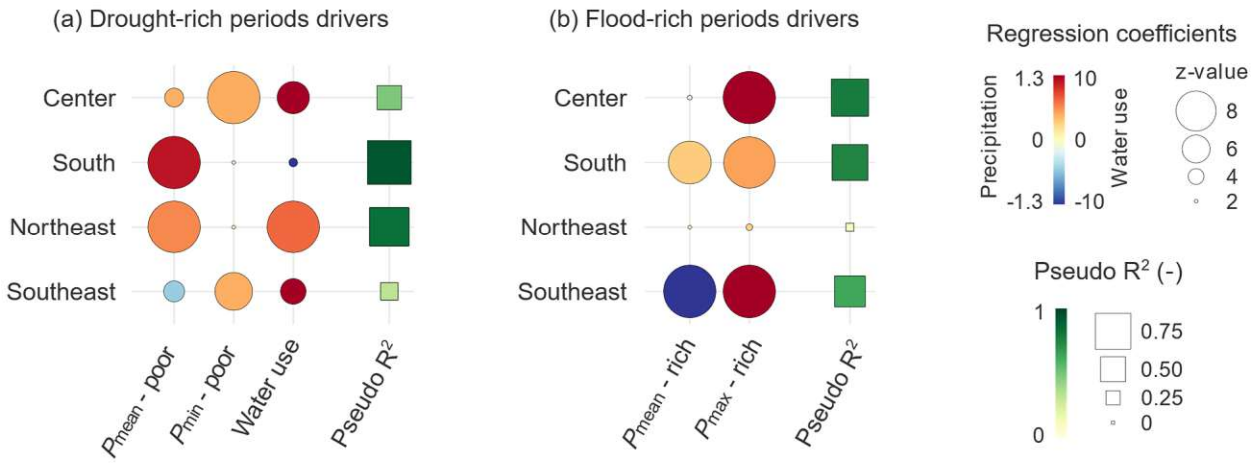

Figure 6.7. Quasi-Poisson regression coefficients for (a) the frequency of gauges with drought-rich periods in each region as the dependent variable and P_{mean}-poor, P_{min}-poor, and consumptive water use as independent variables. (b) Coefficients for the regressions between flood-rich periods, P_{mean}-rich and P_{max}-rich periods. The regressions are computed with a linear link between the variables. Darker green colors and larger squares indicate higher pseudo R^2.

Flood-rich periods are aligned with rainfall-rich periods in three out of four regions (Fig. 6.7b). Water use is not included in the equation, as it is not significant for flood changes in Brazil (Chagas et al., 2022a). Flood-rich periods are linked mostly with P_{max}-rich periods in the center (pseudo R^2 of 0.79, $p < 0.01$) and southeast (pseudo R^2 of 0.63, $p < 0.01$), and with both P_{mean}-rich and P_{max}-rich in the south (pseudo R^2 of 0.75, $p < 0.01$). The results are similar for 5-year return periods, but flood-rich periods are not as well explained for 20-year return periods (pseudo $R^2 < 0.65$ in all regions).

Drought-rich periods are significantly associated with 121-month anomalies in the Atlantic Multidecadal Oscillation (AMO) and Pacific Decadal Oscillation (PDO) in the center (pseudo R^2 of 0.65), northeast (pseudo R^2 of 0.76), and southeast regions (pseudo R^2 of 0.83) (Fig. 6.8a). An increase in drought-rich periods is exponentially related to warm AMO phases and cold PDO phases in these three regions (Fig. 6.8c). Water use is again highly influential in the northeast, with added importance in the southeast. The south region is not associated with either AMO or PDO, suggesting influences of other teleconnections such as the Southern Annular Mode (Morales et al., 2020). Conversely, flood-rich periods are not well-predicted by either AMO or PDO, suggesting that other short-term phenomena could better predict floods, such as the El Niño-Southern Oscillation (Kayano et al., 2013; Tedeschi et al., 2015) and Madden-Julian Oscillation (Grimm, 2019). The results are similar for 5 and 20-year return periods.

6.4 Discussion

The results suggest that the decadal clustering of streamflow extremes is considerably asymmetric for droughts and floods. Streamflow drought-rich periods are more frequent, last longer, and cover a larger spatial extent than flood-rich periods. This finding holds even after ensuring the exact same number of droughts and floods in every time series and accounting for varying return period thresholds. From our perspective, the asymmetry is linked with three main factors: a higher frequency of time clustering of dry spells than wet spells; a stronger carry-over effect of streamflow droughts than floods due to an interannual persistence of water storage deficits; and an amplification of streamflow droughts due to water abstractions.

A higher frequency of gauges with significant rainfall-poor periods is observed (43.6% of gauges, mostly in 1940-1970) compared to rainfall-rich periods (29.2% of gauges, mostly in 1980-2000). Even though the causes of this discrepancy are not entirely understood, a large persistence of dry spells might be associated with teleconnections (Mo & Berbery, 2011), soil-moisture and precipitation feedbacks (Seneviratne et al., 2010), and a larger spatial and temporal coverage of dry spells compared to wet spells (Cavalcanti, 2012). Dry spells often cover areas on the order of 10^6 km^2 over periods of weeks to years (Cavalcanti, 2012; Herrera-Estrada et al., 2017; Sheffield et al., 2009), whereas wet spells and flooding typically cover areas up to 10^4 km^2 and last from days to weeks (Berghuijs, Allen, et al., 2019; Brunner et al., 2020; Cavalcanti, 2012; Ratan & Venugopal, 2013).

Rainfall anomalies accumulate as basin water storage over months and years, resulting in a higher persistence of streamflow compared to rainfall time series (de Lavenne et al., 2022). This carry-over effect of water storage between years contributes to a higher frequency of basins with streamflow time clusters than rainfall time clusters, especially at short time scales. For instance, for 7-year windows and 10-year return period thresholds, drought-rich periods were observed at 48.3% of gauges even though rainfall-poor periods occurred in only 1.0% of the gauges. Generally, the persistence of low flows is even higher than high flows (e.g., David et al., 2022; McMillan, 2022), potentially contributing to a higher frequency of drought-rich than flood-rich periods,

as a sequence of short-lived meteorological droughts gets pooled into longer hydrological droughts (Bevacqua et al., 2021; Van Loon, 2015).

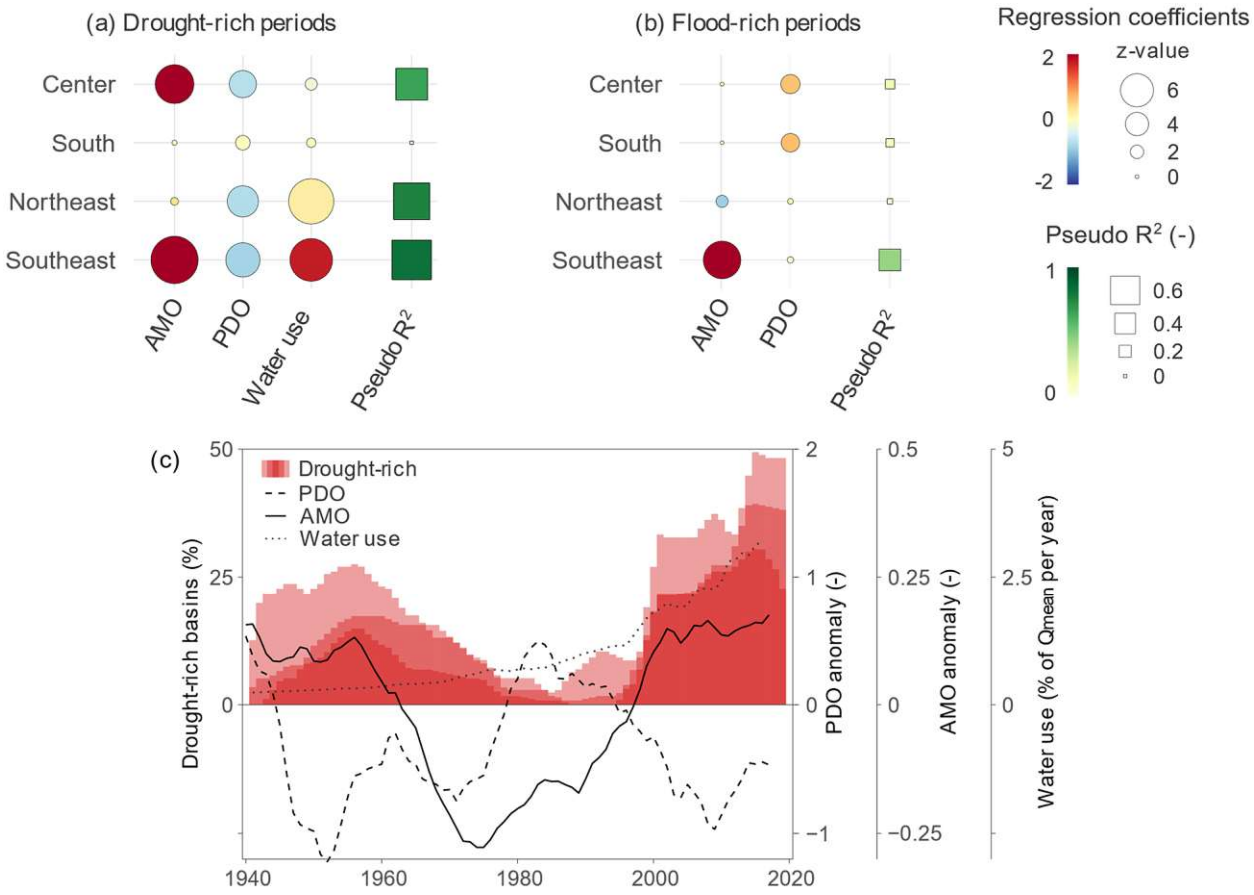

Figure 6.8. Quasi-Poisson regression coefficients for (a) the frequency of gauges with drought-rich periods in each region as the dependent variable and 121-month anomalies in the AMO (Atlantic Multidecadal Oscillation), PDO (Pacific Decadal Oscillation) and consumptive water use as independent variables. (b) Coefficients for the regressions between flood-rich periods, AMO, and PDO anomalies. The regressions are computed with a natural logarithmic link between variables. (c) Time series of AMO, PDO, consumptive water use, and the frequency of drought-rich basins at the regions of Southeast, Northeast, and Center.

Water abstractions contribute to an increase in streamflow droughts in many regions of the world (Custodio, 2002; Di Baldassarre et al., 2018), amplifying the frequency of drought-rich periods even further. In Brazil, 68% of the total water use in 2017 was for irrigation purposes (ANA, 2019c). Irrigated agriculture expanded from 15 to 70 thousand km^2 from 1980 to 2015 (ANA, 2019b), with an increase in crop productivity leading to even more irrigation. Water abstractions are particularly high in the dry season, which generally lasts from 4 to 6 months (Grimm, 2019), contributing to the lowering of groundwater tables and dry-season flows (Chagas et al., 2022a).

The connection between streamflow time clustering and long-term sea surface temperature anomalies suggests that streamflow extremes might often undergo nonlinear changes. For example, the Pacific Decadal Oscillation consists of quasi-periodic cycles of 20 years (d'Orgeville & Peltier, 2007), marked by extended periods of opposite phases and rapid transi-

tions between them (Newman et al., 2016), such as that of the late 1970s (Carvalho et al., 2011; Jacques-Coper & Garreaud, 2015). Thus, studies of drought and flood change would benefit from considering nonstationarities such as cycles and step changes. In this regard, the analysis of drought-rich and flood-rich periods (Blöschl, Bierkens, et al., 2019; Lun et al., 2020) show promising results and an opportunity for exploring the causes of such decadal variabilities.

6.5 Summary and conclusions

In this study, we investigate streamflow drought-rich and flood-rich periods (that is, interannual time clustering of drought and flood occurrence) in Brazil from 1940 to 2020. We looked at their association with rainfall-poor and rainfall-rich periods, water abstractions, the Atlantic Multidecadal Oscillation (AMO) and the Pacific Decadal Oscillation (PDO). Considering events that exceed 10-year return periods, we found that basins with significant (α = 0.053) drought-rich periods were 2.2 times as frequent as flood-rich periods, even after ensuring an equal number of droughts and floods for each time series. We detected drought-rich periods in 61.1% of the basins, 11.5 times higher than expected by chance. This shows how the most severe droughts of the century often cluster into periods of one to two decades, indicating that the assumption of independently and identically distributed series does not hold in most river basins. We believe the asymmetry between drought- and flood-rich periods in Brazil is linked to a higher frequency of rainfall-poor periods than rainfall-rich periods, an amplified effect of water abstractions on streamflow droughts, and a persistence of water storage deficits between years.

The large-scale patterns of drought-rich periods are well-predicted by rainfall-poor periods, water abstractions, the AMO and PDO. In contrast, flood-rich periods are well predicted only by rainfall-rich periods. In general, Brazil experienced a dry period from the 1940s to the 1960s, a wet period in the 1980s and 1990s with the largest floods on record (Fleischmann et al., 2020), and another dry period from the 2000s onward with record low flows. The current dry period has been linked with water abstractions (Chagas et al., 2022a) and to a partial disruption of the national hydropower generation (Tomasella et al., 2022). These multidecadal cycles are associated with 121-month anomalies in the AMO and PDO, characterized by extended periods of opposite phases with rapid transitions between them (Newman et al., 2016).

Our findings suggest that drought and flood change investigations can benefit from analyzing nonlinear changes, especially for periods nearing one century. The analysis of drought-rich and flood-rich periods (Blöschl, Bierkens, et al., 2019; Lun et al., 2020) shows promising results, particularly when considering asymmetries of decadal drought and flood variabilities.

Chapter 7

Summary and conclusions

The goal of this thesis was to investigate flood and drought flow generating mechanisms, change and decadal variability over the past century in Brazil. This study focuses on data-driven, large-scale approaches and comparisons between both hydrology extremes. Such investigation was only possible due to a new comprehensive data set, as presented in Chapter 2, which included information on streamflow, rainfall, evaporation, soil, geology, topography, land cover, and human interference.

7.1 Flood and drought mechanisms

In Chapter 3, we show that the timing of floods is more highly correlated with that of soil moisture peaks than with maximum annual rainfall peaks in 87% of Brazil. Two main patterns of flood generation were identified. In central, western, and northern Brazil, floods are modulated mainly by antecedent soil wetness and less so by event rainfall. Rainfall in the wet season tends to get stored, increasing soil moisture and rising groundwater tables until they peak at the end of the wet season with the annual floods. In contrast, in southern and southeastern Brazil, floods are driven mainly by the magnitude of rainfall events. Intense rainfall quickly saturates the soil and generates floods.

The patterns of flood mechanisms found here are consistent with previous studies in Brazil linking flooding with large-scale meteorological phenomena such as the El Niño (e.g., Cavalcanti, 2012; Lima et al., 2017; Marengo & Espinoza, 2016; Towner et al., 2021). However, we highlight that the mechanisms by which such phenomena cause floods are primarily through soil saturation from long rainfall events rather than by short extreme rainfall, particularly in central, western, and northern Brazil. Such mechanisms are similar to those of western Europe (Berghuijs, Harrigan, et al., 2019; Kemter et al., 2020) and central U.S. (Brunner et al., 2020; Stein et al., 2020). These findings link to a predominant role of soil water storage capacity in regulating flood occurrence. Regions where floods are driven by soil moisture have deeply weathered soils, commonly tens of meters deep (Hengl et al., 2017; Pelletier et al., 2016), with exceptional water storage capacities given the high rainfall magnitudes (Bonell, 2004). On the other hand, regions where floods are driven by event rainfall have either shallower soils or deep soils with a steep decrease in hydraulic conductivity with depth (Bonell, 2004; Hengl et al., 2017; Pelletier et al., 2016), both of which limit soil storage capacities.

In Chapter 4, we propose a new conceptual model and multiscale analysis to investigate the controls of drought flows in terms of climate and basin characteristics linked to water storage. The conceptual model is based on a simple exponential decay function, with a streamflow recession component associated with active water storage dynamics. Findings reveal a similar importance of basin characteristics and climate at continental scales (above 10^7 km^2), but basin characteristics are twice as important as climate at regional scales (below 10^6 km^2). The impact

of the scale of analysis is due to a higher spatial variance of basin characteristics compared with climate, especially over small distances. Such findings integrate the two contradicting hypotheses from previous studies linking low flows either with climate (e.g., Addor et al., 2018; Apurv & Cai, 2020; Beck et al., 2013, 2015) or basin characteristics (e.g., Bloomfield et al., 2009; Carlier et al., 2018; Floriancic et al., 2022; Pfister et al., 2017).

Among the climatic and basin attributes analyzed, we find that geology is the most important control of the spatial variability of drought flows, particularly bedrock type. The highest drought flows in Brazil are observed in basins underlain by siliciclastic sedimentary rocks, especially sandstones, linked with high porosities and fracturing degrees. The lowest drought flows are observed either in the semiarid (where streamflow is intermittent) or in basins underlain by volcanic rocks, particularly basalts, impermeable rocks where fractures are usually disconnected from each other. Intermediate drought flow magnitudes are observed in metamorphic and plutonic formations. The importance of these findings lies in a predominance of basin storage capacity for the spatial variability of drought flows, especially groundwater storage, which attenuates the climatic seasonality. From our perspective, a highest influence of climate on drought flow magnitudes is through wet-season rainfall magnitudes, which determine groundwater recharge potential, than through dry-season length, which is considerably attenuated by the stored water.

Water storage dynamics govern both floods and droughts in Brazil. The storage capacity relevant for floods is linked to soils (a more superficial component), whereas the storage capacity relevant for droughts is linked to geology (a deeper subsurface component). However, these storage components are not necessarily correlated, despite the joint coevolution of soils and bedrock over millennia (Troch et al., 2015). Example of an agreement between the edaphic and geological storage components is in central Brazil, where the high storage capacities dampen the occurrence of both floods and droughts. Similarly, southern Brazil has an agreement of low storage capacities in the edaphic and geological components, associated with increased flood magnitudes and reduced drought flows relative to the mean flow. In contrast, disagreements between the storage components can be found in the southeastern mountain ranges, where low soil storage capacity but high geological storage capacity links to increased floods (aligned with the timing of extreme rainfall) and increased drought flows relative to the mean flow. Southern Amazonia has the opposite condition, with reduced flood and drought flow magnitudes relative to the mean flow. Even though the role of water storage on streamflow generation was studied both on small scales (Harman, 2019; Staudinger et al., 2017) and large scales (Berghuijs, Hartmann, et al., 2016), here we argue that further understanding can come from contrasting flood and drought generating mechanisms.

7.2 Flood and drought change

In Chapter 5, regional linear trends reveal how Brazil has been drying (joint decreases in flood and drought flows; 42% of the study area), wetting (joint increases; 25% of the area), or accelerating (increasing flood flows but decreasing drought flows; 29% of the area) from 1980 to 2015. The predominant signal is of reducing drought flows. The central and northeastern regions were the most impacted, with regional decreases of up to 37% per decade, indicating that some areas have changed from a perennial to an intermittent flow regime. These changes are driven by decreasing mean annual rainfall minus evaporation, with significant influence of the minimum annual 90-day rainfall minus evaporation and increasing water abstractions. On the other hand, increases in flood flows were observed mainly in the south and Amazonia, with changes up to

16% per decade, driven by maximum and mean annual rainfall minus evaporation with similar contributions.

Observed joint changes with increasing flood and decreasing drought flow magnitudes were twice that of the expected value, revealing a consistent signal of an acceleration of the terrestrial component of the water cycle. Even though previous studies have observed an acceleration in Amazonia (Barichivich et al., 2018), here we provide evidence on multiple regions, a robust methodology, and examine its causes. Basins marked by acceleration are aligned with increasing maximum annual rainfall minus evaporation and extensive deforestation. The effects of deforestation, particularly in the tropics, depend on subsequent soil degradation (Bruijnzeel, 2004). It can reduce soil infiltration capacity, which increases runoff and decreases groundwater recharge (Bruijnzeel, 2004), in agreement with the acceleration found here.

Chapter 6 shows a bigger picture of the changes observed in Chapter 5. In general, Brazil had a dry period from 1940 to the late 1960s, an abrupt shift in the late 1970s with frequent flooding, followed by a reversal around 2000 with a steep increase in droughts linked to expanding water abstractions. More importantly, droughts are 2.2 times more likely to show decadal persistence (or time clustering) compared to floods. These drought-rich periods were observed in 61.1% of the basins, 11.5 times higher than the false positive rate of 5.3%, showing that most river basins do not follow the common assumption of independent and identically distributed streamflow time series. Thus, changes in drought and flood flows are considerably asymmetric, with more noticeable decadal persistence (Chapter 6) and change magnitudes (Chapter 5) in drought flows than in flood flows.

Both the linear trends from 1980-2015 and the decadal variabilities from 1940-2020 link drought flow changes more to the mean annual rainfall (recharging the groundwater) than to the minimum annual rainfall (that is, a proxy for dry spells). Thus, drought flow changes are generally in agreement with their generating mechanisms. Flood flow changes shown in Chapters 5 and 6 link to both mean annual rainfall and maximum annual rainfall, indicating that floods often change in both mechanisms identified in Chapter 3 (i.e., where flood depends on intense rainfall magnitudes or antecedent soil moisture).

7.3 Future directions

This thesis advances research on hydrological extremes in three main directions.

First, we provide evidence that the large-scale spatial variability of flood and drought flows is controlled mainly by basin characteristics rather than by climate. Floods and droughts are more dependent on the basin's capacity to attenuate the climatic variability through water storage dynamics than on rainfall magnitudes and timing. Compared with other studies (e.g., Addor et al., 2018; Apurv & Cai, 2020; Beck et al., 2013, 2015; Bloomfield et al., 2009; Carlier et al., 2018; Floriancic et al., 2022; Pfister et al., 2017), the novelty of this contribution comes from considering a multiscale analysis and building a process-based conceptual model that limits the impact of the subjective choice of basin attributes. We find that each streamflow extreme is impacted by basin characteristics in distinct ways that do not necessarily agree with one another, linked to distinct water storage compartments.

Second, we find a consistent signal of the acceleration of the water cycle on land by exploring joint increases in flood and drought flows in the past four decades in Brazil. These changes are driven by increases in the maximum annual rainfall, water demands and deforestation. Even though changes in flood or drought flows have been explored in the literature (e.g., Gudmundsson et al., 2021), a consistent signal of joint changes shows a new perspective on how hydrological extremes can change in the future.

Third, we show how the most severe streamflow extremes of the century often cluster within one to two decades, although at a rate twice as high for droughts compared to floods. More than half of the river basins explored had such persistence, suggesting that research on the long-term behavior of streamflow and rainfall can benefit from analyzing nonlinear changes and asymmetries between flood-rich and drought-rich periods.

The findings presented here suggest several strategies for Brazilian water resources management. For example, flood forecasting and prevention in the south and southeast can benefit from high accuracy detection of short and extreme rainfall events, as these regions are the most sensitive to intense rainfall events. Hydrological drought forecasting can benefit from considering rainfall deficits accumulated in multiple years and the geological configuration of the aquifer, as droughts show high interannual persistence and a particular dependence on groundwater stores. Future national water security can be enhanced by increasing the reliance on groundwater and adopting irrigation techniques that optimize water use. Additional contributions can come from land management measures focused on increasing basin water storage, such as introducing reforestation especially on the southeastern mountain ranges and in south Amazonia. As hydrological extremes remain a growing concern worldwide, this thesis' findings offer a foundation for creating better models and strategies to alleviate the societal impacts of floods and droughts.

References

Addor, N., Newman, A. J., Mizukami, N., & Clark, M. P. (2017). The CAMELS data set: catchment attributes and meteorology for large-sample studies. *Hydrology and Earth System Sciences*, *21*, 5293–5313. https://doi.org/10.5194/hess-21-5293-2017

Addor, N., Nearing, G., Prieto, C., Newman, A. J., Le Vine, N., & Clark, M. P. (2018). A Ranking of Hydrological Signatures Based on Their Predictability in Space. *Water Resources Research*, *54*(11), 8792–8812. https://doi.org/10.1029/2018WR022606

Addor, N., Do, H. X., Alvarez-Garreton, C., Coxon, G., Fowler, K., & Mendoza, P. A. (2019). Large-sample hydrology: recent progress, guidelines for new datasets and grand challenges. *Hydrological Sciences Journal*, 1–14. https://doi.org/10.1080/02626667.2019.1683182

Alexander, L. V., Bador, M., Roca, R., Contractor, S., Donat, M. G., & Nguyen, P. L. (2020). Intercomparison of annual precipitation indices and extremes over global land areas from in situ, space-based and reanalysis products. *Environmental Research Letters*, *15*(5), 055002. https://doi.org/10.1088/1748-9326/ab79e2

Alfieri, L., Lorini, V., Hirpa, F. A., Harrigan, S., Zsoter, E., Prudhomme, C., & Salamon, P. (2020). A global streamflow reanalysis for 1980–2018. *Journal of Hydrology X*, *6*, 100049. https://doi.org/10.1016/j.hydroa.2019.100049

Allan, R. P., Barlow, M., Byrne, M. P., Cherchi, A., Douville, H., Fowler, H. J., et al. (2020). Advances in understanding large-scale responses of the water cycle to climate change. *Annals of the New York Academy of Sciences*, *1472*(1), 49–75. https://doi.org/10.1111/nyas.14337

Allen, G. H., David, C. H., Andreadis, K. M., Hossain, F., & Famiglietti, J. S. (2018). Global Estimates of River Flow Wave Travel Times and Implications for Low-Latency Satellite Data. *Geophysical Research Letters*, *45*(15), 7551–7560. https://doi.org/10.1029/2018GL077914

Alvares, C. A., Stape, J. L., Sentelhas, P. C., Gonçalves, J. L. de M., & Sparovek, G. (2013). Köppen's climate classification map for Brazil. *Meteorologische Zeitschrift*, *22*(6), 711–728. https://doi.org/10.1127/0941-2948/2013/0507

Alvarez-Garreton, C., Mendoza, P. A., Boisier, J. P., Addor, N., Galleguillos, M., Zambrano-Bigiarini, M., et al. (2018). The CAMELS-CL dataset: catchment attributes and meteorology for large sample studies – Chile dataset. *Hydrology and Earth System Sciences*, *22*, 5817–5846. https://doi.org/10.5194/hess-22-5817-2018

ANA - Brazilian National Water Agency. (2017). *Atlas Irrigação: Uso da Água na Agricultura Irrigada* (p. 86). Brasília.

ANA - Brazilian National Water Agency. (2018). *Relatorio de Seguranca de Barragens 2017* (p. 84). Brasília.

ANA - Brazilian National Water Agency. (2019a). HIDROWEB. Retrieved June 15, 2019, from http://www.snirh.gov.br/hidroweb

ANA - Brazilian National Water Agency. (2019b). *Levantamento Da Agricultura Irrigada Por Pivôs Centrais No Brasil (1985-2017), 2a edição* (p. 49). Brasília.

ANA - Brazilian National Water Agency. (2019c). *Manual De Usos Consuntivos Da Água No Brasil* (p. 75). Brasília.

ANA - Brazilian National Water Agency. (2019d). *Relatório de Segurança de Barragens 2018* (p. 54). Brasília.

Apurv, T., & Cai, X. (2020). Drought Propagation in Contiguous U.S. Watersheds: A Process-Based Understanding of the Role of Climate and Watershed Properties. *Water Resources Research*, *56*(9), e2020WR027755. https://doi.org/10.1029/2020WR027755

Archfield, S. A., Hirsch, R. M., Viglione, A., & Blöschl, G. (2016). Fragmented patterns of flood change across the United States. *Geophysical Research Letters*, *43*(19), 10,232-10,239. https://doi.org/10.1002/2016GL070590

Arino, O., Perez, J. J. R., Kalogirou, V., Bontemps, S., Defourny, P., & Bogaert, E. V. (2012). *Global Land Cover Map for 2009 (GlobCover 2009)*. PANGAEA. Retrieved from https://doi.org/10.1594/PANGAEA.787668

Ash, M., & Dohlman, E. (2013). Oil crops outlook. In *Department of Agriculture Economic Research Service*.

Ashouri, H., Hsu, K.-L., Sorooshian, S., Braithwaite, D. K., Knapp, K. R., Cecil, L. D., et al. (2015). PERSIANN-CDR: Daily Precipitation Climate Data Record from Multisatellite Observations for Hydrological and Climate Studies. *Bulletin of the American Meteorological Society*, *96*(1), 69–83. https://doi.org/10.1175/BAMS-D-13-00068.1

Barichivich, J., Gloor, E., Peylin, P., Brienen, R. J. W., Schöngart, J., Espinoza, J. C., & Pattnayak, K. C. (2018). Recent intensification of Amazon flooding extremes driven by strengthened Walker circulation. *Science Advances*, *4*(9), eaat8785. https://doi.org/10.1126/sciadv.aat8785

Bartiko, D., Oliveira, D. Y., Bonumá, N. B., & Chaffe, P. L. B. (2019). Spatial and seasonal patterns of flood change across Brazil. *Hydrological Sciences Journal*, *64*(9), 1071–1079. https://doi.org/10.1080/02626667.2019.1619081

Bassiouni, M., Vogel, R. M., & Archfield, S. A. (2016). Panel regressions to estimate low-flow response to rainfall variability in ungaged basins. *Water Resources Research*, *52*(12), 9470–9494. https://doi.org/10.1002/2016wr018718

Bayliss, A. C., & Jones, R. C. (1993). *Peaks-over-threshold flood database: summary statistics and seasonality*. Wallingford: Institute of Hydrology.

Beck, H. E., van Dijk, A. I. J. M., Miralles, D. G., de Jeu, R. A. M., (Sampurno) Bruijnzeel, L. A., McVicar, T. R., & Schellekens, J. (2013). Global patterns in base flow index and recession based on streamflow observations from 3394 catchments. *Water Resources Research*, *49*(12), 7843–7863. https://doi.org/10.1002/2013WR013918

Beck, H. E., de Roo, A., & van Dijk, A. I. J. M. (2015). Global Maps of Streamflow Characteristics Based on Observations from Several Thousand Catchments. *Journal of Hydrometeorology*, *16*(4), 1478–1501. https://doi.org/10.1175/JHM-D-14-0155.1

Beck, H. E., van Dijk, A. I., De Roo, A., Miralles, D. G., McVicar, T. R., Schellekens, J., & Bruijnzeel, L. A. (2016). Global-scale regionalization of hydrologic model parameters. *Water Resources Research*, *52*(5), 3599–3622. https://doi.org/10.1002/2015WR018247

Beck, H. E., Van Dijk, A. I., De Roo, A., Dutra, E., Fink, G., Orth, R., & Schellekens, J. (2017). Global evaluation of runoff from 10 state-of-the-art hydrological models. *Hydrology and Earth System Sciences*, *21*(6), 2881–2903. https://doi.org/10.5194/hess-21-2881-2017

Beck, H. E., Vergopolan, N., Pan, M., Levizzani, V., van Dijk, A. I. J. M., Weedon, G. P., et al. (2017). Global-scale evaluation of 22 precipitation datasets using gauge observations and hy-

drological modeling. *Hydrology and Earth System Sciences*, *21*(12), 6201–6217. https://doi.org/10.5194/hess-21-6201-2017

Beck, H. E., Wood, E. F., Pan, M., Fisher, C. K., Miralles, D. G., van Dijk, A. I. J. M., et al. (2019). MSWEP V2 Global 3-Hourly 0.1° Precipitation: Methodology and Quantitative Assessment. *Bulletin of the American Meteorological Society*, *100*(3), 473–500. https://doi.org/10.1175/BAMS-D-17-0138.1

Beck, H. E., Pan, M., Miralles, D. G., Reichle, R. H., Dorigo, W. A., Hahn, S., et al. (2021). Evaluation of 18 satellite- and model-based soil moisture products using in situ measurements from 826 sensors. *Hydrology and Earth System Sciences*, *25*(1), 17–40. https://doi.org/10.5194/hess-25-17-2021

Berghuijs, W. R., & Woods, R. A. (2016). Correspondence: Space-time asymmetry undermines water yield assessment. *Nature Communications*, *7*(1), 11603. https://doi.org/10.1038/ncomms11603

Berghuijs, W. R., Sivapalan, M., Woods, R. A., & Savenije, H. H. G. (2014). Patterns of similarity of seasonal water balances: A window into streamflow variability over a range of time scales. *Water Resources Research*, *50*(7), 5638–5661. https://doi.org/10.1002/2014WR015692

Berghuijs, W. R., Woods, R. A., Hutton, C. J., & Sivapalan, M. (2016). Dominant flood generating mechanisms across the United States. *Geophysical Research Letters*, *43*(9), 4382–4390. https://doi.org/10.1002/2016GL068070

Berghuijs, W. R., Hartmann, A., & Woods, R. A. (2016). Streamflow sensitivity to water storage changes across Europe. *Geophysical Research Letters*, *43*(5), 1980–1987. https://doi.org/10.1002/2016GL067927

Berghuijs, W. R., Allen, S. T., Harrigan, S., & Kirchner, J. W. (2019). Growing Spatial Scales of Synchronous River Flooding in Europe. *Geophysical Research Letters*, *46*(3), 1423–1428. https://doi.org/10.1029/2018GL081883

Berghuijs, W. R., Harrigan, S., Molnar, P., Slater, L. J., & Kirchner, J. W. (2019). The Relative Importance of Different Flood-Generating Mechanisms Across Europe. *Water Resources Research*. https://doi.org/10.1029/2019WR024841

Bernal, J. P., Cruz, F. W., Stríkis, N. M., Wang, X., Deininger, M., Catunda, M. C. A., et al. (2016). High-resolution Holocene South American monsoon history recorded by a speleothem from Botuverá Cave, Brazil. *Earth and Planetary Science Letters*, *450*, 186–196. https://doi.org/10.1016/j.epsl.2016.06.008

Bevacqua, A. G., Chaffe, P. L. B., Chagas, V. B. P., & AghaKouchak, A. (2021). Spatial and temporal patterns of propagation from meteorological to hydrological droughts in Brazil. *Journal of Hydrology*, *603*, 126902. https://doi.org/10.1016/j.jhydrol.2021.126902

Bevacqua, E., Vousdoukas, M. I., Zappa, G., Hodges, K., Shepherd, T. G., Maraun, D., et al. (2020). More meteorological events that drive compound coastal flooding are projected under climate change. *Communications Earth & Environment*, *1*(1), 1–11. https://doi.org/10.1038/s43247-020-00044-z

Beven, K., & Germann, P. (2013). Macropores and water flow in soils revisited. *Water Resources Research*, *49*(6), 3071–3092. https://doi.org/10.1002/wrcr.20156

Bierkens, M. F. (2015). Global hydrology 2015: State, trends, and directions. *Water Resources Research*, *51*(7), 4923–4947. https://doi.org/10.1002/2015WR017173

Bloomfield, J. P., Allen, D. J., & Griffiths, K. J. (2009). Examining geological controls on baseflow index (BFI) using regression analysis: An illustration from the Thames Basin, UK. *Journal of Hydrology*, *373*(1), 164–176. https://doi.org/10.1016/j.jhydrol.2009.04.025

Blöschl, G. (2022). Flood generation: process patterns from the raindrop to the ocean. *Hydrology and Earth System Sciences Discussions*, 1–22. https://doi.org/10.5194/hess-2022-2

Blöschl, G., & Sivapalan, M. (1995). Scale issues in hydrological modelling: A review. *Hydrological Processes*, *9*(3–4), 251–290. https://doi.org/10.1002/hyp.3360090305

Blöschl, G., Sivapalan, M., Savenije, H., Wagener, T., & Viglione, A. (2013). *Runoff prediction in ungauged basins: synthesis across processes, places and scales*. Cambridge University Press.

Blöschl, G., Hall, J., Parajka, J., Perdigão, R. A. P., Merz, B., Arheimer, B., et al. (2017). Changing climate shifts timing of European floods. *Science*, *357*(6351), 588–590. https://doi.org/10.1126/science.aan2506

Blöschl, G., Hall, J., Viglione, A., Perdigão, R. A. P., Parajka, J., Merz, B., et al. (2019). Changing climate both increases and decreases European river floods. *Nature*, *573*, 108–111. https://doi.org/10.1038/s41586-019-1495-6

Blöschl, G., Bierkens, M. F. P., Chambel, A., Cudennec, C., Destouni, G., Fiori, A., et al. (2019). Twenty-three unsolved problems in hydrology (UPH) – a community perspective. *Hydrological Sciences Journal*, *64*(10), 1141–1158. https://doi.org/10.1080/02626667.2019.1620507

Blöschl, G., Kiss, A., Viglione, A., Barriendos, M., Böhm, O., Brázdil, R., et al. (2020). Current European flood-rich period exceptional compared with past 500 years. *Nature*, *583*(7817), 560–566. https://doi.org/10.1038/s41586-020-2478-3

Blum, A. G., Archfield, S. A., Hirsch, R. M., Vogel, R. M., Kiang, J. E., & Dudley, R. W. (2019). Updating estimates of low-streamflow statistics to account for possible trends. *Hydrological Sciences Journal*, *64*(12), 1404–1414. https://doi.org/10.1080/02626667.2019.1655148

Blum, A. G., Ferraro, P. J., Archfield, S. A., & Ryberg, K. R. (2020). Causal Effect of Impervious Cover on Annual Flood Magnitude for the United States. *Geophysical Research Letters*, *47*(5), e2019GL086480. https://doi.org/10.1029/2019GL086480

de Boer-Euser, T., McMillan, H. K., Hrachowitz, M., Winsemius, H. C., & Savenije, H. H. G. (2016). Influence of soil and climate on root zone storage capacity. *Water Resources Research*, *52*(3), 2009–2024. https://doi.org/10.1002/2015WR018115

Bola, G. B., Tshimanga, R. M., Neal, J., Trigg, M. A., Hawker, L., Lukanda, V. M., & Bates, P. (2022). Understanding flood seasonality and flood regime shift in the Congo River Basin. *Hydrological Sciences Journal*, *67*(10), 1496–1515. https://doi.org/10.1080/02626667.2022.2083966

Bonell, M. (2004). Runoff generation in tropical forests. In *Forests, water and people in the humid tropics: past, present and future hydrological research for integrated land and water management* (pp. 314–406). Cambridge University Press.

Börker, J., Hartmann, J., Amann, T., & Romero-Mujalli, G. (2018). Terrestrial Sediments of the Earth: Development of a Global Unconsolidated Sediments Map Database (GUM). *Geochemistry, Geophysics, Geosystems*, *19*(4), 997–1024. https://doi.org/10.1002/2017GC007273

Braga, B., & Kelman, J. (2016). Facing the challenge of extreme climate: the case of Metropolitan São Paulo. *Water Policy*, *18*(S2), 52–69. https://doi.org/10.2166/wp.2016.113

Brás, T. A., Jägermeyr, J., & Seixas, J. (2019). Exposure of the EU-28 food imports to extreme weather disasters in exporting countries. *Food Security*, *11*(6), 1373–1393. https://doi.org/10.1007/s12571-019-00975-2

Brienen, R. J. W., Phillips, O. L., Feldpausch, T. R., Gloor, E., Baker, T. R., Lloyd, J., et al. (2015). Long-term decline of the Amazon carbon sink. *Nature*, *519*(7543), 344–348. https://doi.org/10.1038/nature14283

Brocca, L., Ciabatta, L., Massari, C., Camici, S., & Tarpanelli, A. (2017). Soil Moisture for Hydrological Applications: Open Questions and New Opportunities. *Water*, *9*(2), 140. https://doi.org/10.3390/w9020140

Bruijnzeel, L. A. (2004). Hydrological functions of tropical forests: not seeing the soil for the trees? *Agriculture, Ecosystems & Environment*, *104*(1), 185–228. https://doi.org/10.1016/j.agee.2004.01.015

Brunner, M. I., Gilleland, E., Wood, A., Swain, D. L., & Clark, M. (2020). Spatial Dependence of Floods Shaped by Spatiotemporal Variations in Meteorological and Land-Surface Processes. *Geophysical Research Letters*, *47*(13), e2020GL088000. https://doi.org/10.1029/2020GL088000

Buckley, B. M., Palakit, K., Duangsathaporn, K., Sanguantham, P., & Prasomsin, P. (2007). Decadal scale droughts over northwestern Thailand over the past 448 years: links to the tropical Pacific and Indian Ocean sectors. *Climate Dynamics*, *29*(1), 63–71. https://doi.org/10.1007/s00382-007-0225-1

Budyko, M. I. (1974). *Climate and life* (Vol. 508). Academic press New York.

Carlier, C., Wirth, S. B., Cochand, F., Hunkeler, D., & Brunner, P. (2018). Geology controls streamflow dynamics. *Journal of Hydrology*, *566*, 756–769. https://doi.org/10.1016/j.jhydrol.2018.08.069

Carvalho, L. M. V., Jones, C., Silva, A. E., Liebmann, B., & Silva Dias, P. L. (2011). The South American Monsoon System and the 1970s climate transition. *International Journal of Climatology*, *31*(8), 1248–1256. https://doi.org/10.1002/joc.2147

Cassalho, F., Beskow, S., Rogério de Mello, C., Oliveira, L. F., & Sanchotene de Aguiar, M. (2019). Evaluation of Flood Timing and Regularity over Hydrological Regionalization in Southern Brazil. *Journal of Hydrologic Engineering*, *24*(8), 05019022. https://doi.org/10.1061/(ASCE)HE.1943-5584.0001815

Cavalcanti, I. F. A. (2012). Large scale and synoptic features associated with extreme precipitation over South America: A review and case studies for the first decade of the 21st century. *Atmospheric Research*, *118*, 27–40. https://doi.org/10.1016/j.atmosres.2012.06.012

Cavalcanti, I. F. A., Carril, A. F., Penalba, O. C., Grimm, A. M., Menéndez, C. G., Sanchez, E., et al. (2015). Precipitation extremes over La Plata Basin – Review and new results from observations and climate simulations. *Journal of Hydrology*, *523*, 211–230. https://doi.org/10.1016/j.jhydrol.2015.01.028

Chagas, V. B. P., & Chaffe, P. L. B. (2018). The Role of Land Cover in the Propagation of Rainfall Into Streamflow Trends. *Water Resources Research*, *54*(9), 5986–6004. https://doi.org/10.1029/2018WR022947

Chagas, V. B. P., Chaffe, P. L. B., Addor, N., Fan, F. M., Fleischmann, A. S., Paiva, R. C. D., & Siqueira, V. A. (2020a). CAMELS-BR: hydrometeorological time series and landscape attributes for 897 catchments in Brazil. *Earth System Science Data*, *12*(3), 2075–2096. https://doi.org/10.5194/essd-12-2075-2020

Chagas, V. B. P., Chaffe, P. L. B., Addor, N., Fan, F. M., Fleischmann, A. S., Paiva, R. C. D., & Siqueira, V. A. (2020b, July 30). CAMELS-BR: Hydrometeorological time series and landscape attributes for 897 catchments in Brazil - link to files. Zenodo. Retrieved from DOI: 10.5281/zenodo.3964745

Chagas, V. B. P., Chaffe, P. L. B., & Blöschl, G. (2022a). Climate and land management accelerate the Brazilian water cycle. *Nature Communications*, *13*(1), 5136. https://doi.org/10.1038/s41467-022-32580-x

Chagas, V. B. P., Chaffe, P. L. B., & Blöschl, G. (2022b). Process Controls on Flood Seasonality in Brazil. *Geophysical Research Letters*, *49*(5). https://doi.org/10.1029/2021GL096754

Chou, C., Chiang, J. C. H., Lan, C.-W., Chung, C.-H., Liao, Y.-C., & Lee, C.-J. (2013). Increase in the range between wet and dry season precipitation. *Nature Geoscience*, *6*(4), 263–267. https://doi.org/10.1038/ngeo1744

Cigizoglu, H. K., Bayazit, M., & Önöz, B. (2005). Trends in the Maximum, Mean, and Low Flows of Turkish Rivers. *Journal of Hydrometeorology*, *6*(3), 280–290. https://doi.org/10.1175/JHM412.1

Clausen, B., & Biggs, B. J. F. (2000). Flow variables for ecological studies in temperate streams: groupings based on covariance. *Journal of Hydrology*, *237*(3–4), 184–197. https://doi.org/10.1016/S0022-1694(00)00306-1

Collischonn, W., Tucci, C. E. M., & Clarke, R. T. (2001). Further evidence of changes in the hydrological regime of the River Paraguay: part of a wider phenomenon of climate change? *Journal of Hydrology*, *245*(1–4), 218–238. https://doi.org/10.1016/S0022-1694(01)00348-1

Cook, B. I., Smerdon, J. E., Cook, E. R., Williams, A. P., Anchukaitis, K. J., Mankin, J. S., et al. (2022). Megadroughts in the Common Era and the Anthropocene. *Nature Reviews Earth & Environment*, *3*(11), 741–757. https://doi.org/10.1038/s43017-022-00329-1

Copernicus Climate Change Service. (2019). ERA5-Land hourly data from 2001 to present. ECMWF. Retrieved from https://cds.climate.copernicus.eu/doi/10.24381/cds.e2161bac

Costa, M. H., Botta, A., & Cardille, J. A. (2003). Effects of large-scale changes in land cover on the discharge of the Tocantins River, Southeastern Amazonia. *Journal of Hydrology*, *283*(1–4), 206–217. https://doi.org/10.1016/S0022-1694(03)00267-1

Coxon, G., Addor, N., Bloomfield, J. P., Freer, J., Fry, M., Hannaford, J., et al. (2020). CAMELS-GB: hydrometeorological time series and landscape attributes for 671 catchments in Great Britain. *Earth System Science Data*, *12*(4), 2459–2483. https://doi.org/10.5194/essd-12-2459-2020

Cramér, H. (1999). *Mathematical methods of statistics* (Vol. 43). Princeton University Press.

Crochemore, L., Isberg, K., Pimentel, R., Pineda, L., Hasan, A., & Arheimer, B. (2019). Lessons learnt from checking the quality of openly accessible river flow data worldwide. *Hydrological Sciences Journal*, 1–13. https://doi.org/10.1080/02626667.2019.1659509

Croissant, Y., & Millo, G. (2008). Panel Data Econometrics in R : The plm Package. *Journal of Statistical Software*, *27*(2), 1–43. https://doi.org/10.18637/jss.v027.i02

Cuartas, L. A., Cunha, A. P. M. do A., Alves, J. A., Parra, L. M. P., Deusdará-Leal, K., Costa, L. C. O., et al. (2022). Recent Hydrological Droughts in Brazil and Their Impact on Hydropower Generation. *Water*, *14*(4), 601. https://doi.org/10.3390/w14040601

Cunha, A. P. M. A., Tomasella, J., Ribeiro-Neto, G. G., Brown, M., Garcia, S. R., Brito, S. B., & Carvalho, M. A. (2018). Changes in the spatial-temporal patterns of droughts in the Brazilian Northeast. *Atmospheric Science Letters*, *19*(10), e855. https://doi.org/10.1002/asl.855

Custodio, E. (2002). Aquifer overexploitation: what does it mean? *Hydrogeology Journal*, *10*(2), 254–277. https://doi.org/10.1007/s10040-002-0188-6

David, P. C., Chaffe, P. L. B., Chagas, V. B. P., Dal Molin, M., Oliveira, D. Y., Klein, A. H. F., & Fenicia, F. (2022). Correspondence Between Model Structures and Hydrological Signatures: A Large-Sample Case Study Using 508 Brazilian Catchments. *Water Resources Research*, *58*(3). https://doi.org/10.1029/2021WR030619

Di Baldassarre, G., Wanders, N., AghaKouchak, A., Kuil, L., Rangecroft, S., Veldkamp, T. I. E., et al. (2018). Water shortages worsened by reservoir effects. *Nature Sustainability*, *1*(11), 617–622. https://doi.org/10.1038/s41893-018-0159-0

Dias, L. C. P., Pimenta, F. M., Santos, A. B., Costa, M. H., & Ladle, R. J. (2016). Patterns of land use, extensification, and intensification of Brazilian agriculture. *Global Change Biology*, *22*(8), 2887–2903. https://doi.org/10.1111/gcb.13314

Diniz, J. A. O., Monteiro, A. B., Silva, R. de C. da, & Paula, T. L. F. de. (2014). *Mapa hidrogeológico do Brasil ao milionésimo: nota técnica*. CPRM.

Diniz, J. A. O., Silva, R. de C. da, & Paula, T. L. F. de. (2015). *Mapa hidrogeológico da Bacia do Prata ao milionésimo: nota técnica*. CPRM.

Do, H. X., Gudmundsson, L., Leonard, M., & Westra, S. (2018). The Global Streamflow Indices and Metadata Archive (GSIM) – Part 1: The production of a daily streamflow archive and metadata. *Earth System Science Data*, *10*(2), 765–785. https://doi.org/10.5194/essd-10-765-2018

Do, H. X., Westra, S., Leonard, M., & Gudmundsson, L. (2020). Global-Scale Prediction of Flood Timing Using Atmospheric Reanalysis. *Water Resources Research*, *56*(1), e2019WR024945. https://doi.org/10.1029/2019WR024945

Döll, P., Douville, H., Güntner, A., Müller Schmied, H., & Wada, Y. (2016). Modelling freshwater resources at the global scale: challenges and prospects. *Surveys in Geophysics*, *37*, 195–221. https://doi.org/10.1007/s10712-015-9343-1

Donat, M. G., Lowry, A. L., Alexander, L. V., O'Gorman, P. A., & Maher, N. (2016). More extreme precipitation in the world's dry and wet regions. *Nature Climate Change*, *6*(5), 508–513. https://doi.org/10.1038/nclimate2941

Dudley, R. W., Hirsch, R. M., Archfield, S. A., Blum, A. G., & Renard, B. (2020). Low streamflow trends at human-impacted and reference basins in the United States. *Journal of Hydrology*, *580*, 124254. https://doi.org/10.1016/j.jhydrol.2019.124254

Durack, P. J., Wijffels, S. E., & Matear, R. J. (2012). Ocean Salinities Reveal Strong Global Water Cycle Intensification During 1950 to 2000. *Science*, *336*(6080), 455–458. https://doi.org/10.1126/science.1212222

Durkee, J. D., Mote, T. L., & Shepherd, J. M. (2009). The Contribution of Mesoscale Convective Complexes to Rainfall across Subtropical South America. *Journal of Climate*, *22*(17), 4590–4605. https://doi.org/10.1175/2009JCLI2858.1

Ehret, U., Gupta, H. V., Sivapalan, M., Weijs, S. V., Schymanski, S. J., Blöschl, G., et al. (2014). Advancing catchment hydrology to deal with predictions under change. *Hydrology and Earth System Sciences*, *18*(2), 649–671. https://doi.org/10.5194/hess-18-649-2014

Ehsanzadeh, E., & Adamowski, K. (2010). Trends in timing of low stream flows in Canada: impact of autocorrelation and long-term persistence. *Hydrological Processes*, *24*(8), 970–980. https://doi.org/10.1002/hyp.7533

Ekolu, J., Dieppois, B., Sidibe, M., Eden, J. M., Tramblay, Y., Villarini, G., et al. (2022). Long-term variability in hydrological droughts and floods in sub-Saharan Africa: New perspectives from a 65-year daily streamflow dataset. *Journal of Hydrology*, *613*, 128359. https://doi.org/10.1016/j.jhydrol.2022.128359

Elsenbeer, H. (2001). Hydrologic flowpaths in tropical rainforest soilscapes—a review. *Hydrological Processes*, *15*(10), 1751–1759. https://doi.org/10.1002/hyp.237

ESA/CCI - European Space Agency - Climate Change Initiative. (2016). ESA Land cover maps (1992-2015). Retrieved December 21, 2020, from http://maps.elie.ucl.ac.be/CCI/viewer/download.php

Falkenmark, M., & Chapman, T. (1989). *Comparative hydrology: An ecological approach to land and water resources*. Unesco.

Fan, Y. (2015). Groundwater in the Earth's critical zone: Relevance to large-scale patterns and processes. *Water Resources Research*, *51*(5), 3052–3069. https://doi.org/10.1002/2015WR017037

Fan, Y., Li, H., & Miguez-Macho, G. (2013). Global Patterns of Groundwater Table Depth. *Science*, *339*(940), 1–5. https://doi.org/10.1126/science.1229881

Fan, Y., Clark, M., Lawrence, D. M., Swenson, S., Band, L. E., Brantley, S. L., et al. (2019). Hillslope Hydrology in Global Change Research and Earth System Modeling. *Water Resources Research*, *55*(2), 1737–1772. https://doi.org/10.1029/2018WR023903

Fatichi, S., Or, D., Walko, R., Vereecken, H., Young, M. H., Ghezzehei, T. A., et al. (2020). Soil structure is an important omission in Earth System Models. *Nature Communications*, *11*(1), 522. https://doi.org/10.1038/s41467-020-14411-z

Feng, X., Thompson, S. E., Woods, R., & Porporato, A. (2019). Quantifying Asynchronicity of Precipitation and Potential Evapotranspiration in Mediterranean Climates. *Geophysical Research Letters*, *46*(24), 14692–14701. https://doi.org/10.1029/2019GL085653

Fiala, T., Ouarda, T. B. M. J., & Hladný, J. (2010). Evolution of low flows in the Czech Republic. *Journal of Hydrology*, *393*(3), 206–218. https://doi.org/10.1016/j.jhydrol.2010.08.018

Field, C. B., Barros, V., Stocker, T. F., & Dahe, Q. (2012). *Managing the risks of extreme events and disasters to advance climate change adaptation: special report of the intergovernmental panel on climate change*. Cambridge University Press.

Fischer, E. M., & Knutti, R. (2016). Observed heavy precipitation increase confirms theory and early models. *Nature Climate Change*, *6*(11), 986–991. https://doi.org/10.1038/nclimate3110

Flantua, S. G. A., Hooghiemstra, H., Vuille, M., Behling, H., Carson, J. F., Gosling, W. D., et al. (2016). Climate variability and human impact in South America during the last 2000 years: synthesis and perspectives from pollen records. *Climate of the Past*, *12*(2), 483–523. https://doi.org/10.5194/cp-12-483-2016

Fleig, A. K., Tallaksen, L. M., Hisdal, H., & Demuth, S. (2006). A global evaluation of streamflow drought characteristics. *Hydrology and Earth System Sciences*, *5*, 535–552. https://doi.org/10.5194/hess-10-535-2006

Fleischmann, A. S., Siqueira, V. A., Wongchuig-Correa, S., Collischonn, W., & Paiva, R. C. D. D. (2020). The great 1983 floods in South American large rivers: a continental hydrological modelling approach. *Hydrological Sciences Journal*, *65*(8), 1358–1373. https://doi.org/10.1080/02626667.2020.1747622

Flores, B. M., Holmgren, M., Xu, C., van Nes, E. H., Jakovac, C. C., Mesquita, R. C. G., & Scheffer, M. (2017). Floodplains as an Achilles' heel of Amazonian forest resilience. *Proceed-*

ings of the National Academy of Sciences, *114*(17), 4442–4446. https://doi.org/10.1073/pnas.1617988114

Floriancic, M. G., Spies, D., van Meerveld, I. H. J., & Molnar, P. (2022). A multi-scale study of the dominant catchment characteristics impacting low-flow metrics. *Hydrological Processes*, *36*(1), e14462. https://doi.org/10.1002/hyp.14462

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., et al. (2015). The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Scientific Data*, *2*(1), 1–21. https://doi.org/10.1038/sdata.2015.66

García, N. O., & Mechoso, C. R. (2005). Variability in the discharge of South American rivers and in climate. *Hydrological Sciences Journal*, *50*(3), 5. https://doi.org/10.1623/hysj.50.3.459.65030

Getirana, A., Libonati, R., & Cataldi, M. (2021). Brazil is in water crisis — it needs a drought plan. *Nature*, *600*(7888), 218–220. https://doi.org/10.1038/d41586-021-03625-w

Ghiggi, G., Humphrey, V., Seneviratne, S. I., & Gudmundsson, L. (2019). GRUN: an observation-based global gridded runoff dataset from 1902 to 2014. *Earth System Science Data*, *11*(4), 1655–1674. https://doi.org/10.5194/essd-11-1655-2019

Gleeson, T., Moosdorf, N., Hartmann, J., & van Beek, L. P. H. (2014). A glimpse beneath earth's surface: GLobal HYdrogeology MaPS (GLHYMPS) of permeability and porosity. *Geophysical Research Letters*, *41*(11), 3891–3898. https://doi.org/10.1002/2014GL059856

Gloor, M., Brienen, R. J. W., Galbraith, D., Feldpausch, T. R., Schöngart, J., Guyot, J.-L., et al. (2013). Intensification of the Amazon hydrological cycle over the last two decades. *Geophysical Research Letters*, *40*(9), 1729–1733. https://doi.org/10.1002/grl.50377

Gloor, M., Barichivich, J., Ziv, G., Brienen, R., Schöngart, J., Peylin, P., et al. (2015). Recent Amazon climate as background for possible ongoing and future changes of Amazon humid forests. *Global Biogeochemical Cycles*, *29*(9), 1384–1399. https://doi.org/10.1002/2014GB005080

Gnann, S. J., McMillan, H. K., Woods, R. A., & Howden, N. J. K. (2021). Including Regional Knowledge Improves Baseflow Signature Predictions in Large Sample Hydrology. *Water Resources Research*, *57*(2), e2020WR028354. https://doi.org/10.1029/2020WR028354

Gómes Tapias, J., Schobbenhaus, C., & Montes Ramírez, N. E. (2019). Geological map of south America.

Gonçalves-Maduro, L., Armindo, R. A., Turek, M. E., & Wendroth, O. (2020). Soil water and fuel permeability of a Cambisol in southern Brazil and its spatial behavior: A case study. *Vadose Zone Journal*, *19*(1), e20035. https://doi.org/10.1002/vzj2.20035

de Graaf, I. E. M., Gleeson, T., (Rens) van Beek, L. P. H., Sutanudjaja, E. H., & Bierkens, M. F. P. (2019). Environmental flow limits to global groundwater pumping. *Nature*, *574*(7776), 90–94. https://doi.org/10.1038/s41586-019-1594-4

Gräler, B., Pebesma, E., & Heuvelink, G. (2016). Spatio-Temporal Interpolation using gstat. *The R Journal*, *8*(1), 204. https://doi.org/10.32614/RJ-2016-014

GRDC - Global Runoff Data Centre. (2019). GRDC. Retrieved December 24, 2019, from https://www.bafg.de/GRDC/EN/Home/homepage_node.html

Greve, P., Orlowsky, B., Mueller, B., Sheffield, J., Reichstein, M., & Seneviratne, S. I. (2014). Global assessment of trends in wetting and drying over land. *Nature Geoscience*, *7*(10), 716–721. https://doi.org/10.1038/ngeo2247

Grimm, A. M. (2011). Interannual climate variability in South America: impacts on seasonal precipitation, extreme events, and possible effects of climate change. *Stochastic Environmental Research and Risk Assessment*, *25*(4), 537–554. https://doi.org/10.1007/s00477-010-0420-1

Grimm, A. M. (2019). South American monsoon and its extremes. In *Tropical Extremes* (pp. 51–93). Elsevier.

Gudmundsson, L., Wagener, T., Tallaksen, L. M., & Engeland, K. (2012). Evaluation of nine large-scale hydrological models with respect to the seasonal runoff climatology in Europe. *Water Resources Research*, *48*(11). https://doi.org/10.1029/2011WR010911

Gudmundsson, L., Do, H. X., Leonard, M., & Westra, S. (2018). The Global Streamflow Indices and Metadata Archive (GSIM) – Part 2: Quality control, time-series indices and homogeneity assessment. *Earth System Science Data*, *10*(2), 787–804. https://doi.org/10.5194/essd-10-787-2018

Gudmundsson, L., Leonard, M., Do, H. X., Westra, S., & Seneviratne, S. I. (2019). Observed Trends in Global Indicators of Mean and Extreme Streamflow. *Geophysical Research Letters*, *46*(2), 756–766. https://doi.org/10.1029/2018GL079725

Gudmundsson, L., Boulange, J., Do, H. X., Gosling, S. N., Grillakis, M. G., Koutroulis, A. G., et al. (2021). Globally observed trends in mean and extreme river flow attributed to climate change. *Science*, *371*(6534), 1159–1162. https://doi.org/10.1126/science.aba3996

Guimberteau, M., Ciais, P., Ducharne, A., Boisier, J. P., Dutra Aguiar, A. P., Biemans, H., et al. (2017). Impacts of future deforestation and climate change on the hydrology of the Amazon Basin: a multi-model analysis with a new set of land-cover change scenarios. *Hydrology and Earth System Sciences*, *21*(3), 1455–1475. https://doi.org/10.5194/hess-21-1455-2017

Gupta, H. V., Perrin, C., Blöschl, G., Montanari, A., Kumar, R., Clark, M., & Andréassian, V. (2014). Large-sample hydrology: a need to balance depth with breadth. *Hydrology and Earth System Sciences*, *18*(2), 463–477. https://doi.org/10.5194/hess-18-463-2014

Haddad, E. A., & Teixeira, E. (2015). Economic impacts of natural disasters in megacities: The case of floods in São Paulo, Brazil. *Habitat International*, *45*, 106–113. https://doi.org/10.1016/j.habitatint.2014.06.023

Haddeland, I., Clark, D. B., Franssen, W., Ludwig, F., Voß, F., Arnell, N. W., et al. (2011). Multimodel estimate of the global terrestrial water balance: setup and first results. *Journal of Hydrometeorology*, *12*(5), 869–884. https://doi.org/10.1175/2011JHM1324.1

Hall, J., Arheimer, B., Borga, M., Brázdil, R., Claps, P., Kiss, A., et al. (2014). Understanding flood regime changes in Europe: a state-of-the-art assessment. *Hydrology and Earth System Sciences*, *18*(7), 2735–2772. https://doi.org/10.5194/hess-18-2735-2014

Hannaford, J., Buys, G., Stahl, K., & Tallaksen, L. M. (2013). The influence of decadal-scale variability on trends in long European streamflow records. *Hydrology and Earth System Sciences*, *17*(7), 2717–2733. https://doi.org/10.5194/hess-17-2717-2013

Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., et al. (2013). High-Resolution Global Maps of 21st-Century Forest Cover Change. *Science*, *342*(6160), 850–853. https://doi.org/10.1126/science.1244693

Harman, C. J. (2019). Age-Ranked Storage-Discharge Relations: A Unified Description of Spatially Lumped Flow and Water Age in Hydrologic Systems. *Water Resources Research*, *55*(8), 7143–7165. https://doi.org/10.1029/2017WR022304

Hartmann, J., & Moosdorf, N. (2012). The new global lithological map database GLiM: A representation of rock properties at the Earth surface. *Geochemistry, Geophysics, Geosystems*, *13*(12), 1–37. https://doi.org/10.1029/2012GC004370

Hausman, J. A. (1978). Specification Tests in Econometrics. *Econometrica*, *46*(6), 1251–1271. https://doi.org/10.2307/1913827

Hayhoe, S. J., Neill, C., Porder, S., Mchorney, R., Lefebvre, P., Coe, M. T., et al. (2011). Conversion to soy on the Amazonian agricultural frontier increases streamflow without affecting stormflow dynamics. *Global Change Biology*, *17*(5), 1821–1833. https://doi.org/10.1111/j.1365-2486.2011.02392.x

Haylock, M., & Nicholls, N. (2000). Trends in extreme rainfall indices for an updated high quality data set for Australia, 1910 – 1998. *International Journal of Climatology*, 9. https://doi.org/10.1002/1097-0088(20001115)20:13<1533::AID-JOC586>3.0.CO;2-J

He, X., & Sheffield, J. (2020). Lagged Compound Occurrence of Droughts and Pluvials Globally Over the Past Seven Decades. *Geophysical Research Letters*, *47*(14). https://doi.org/10.1029/2020GL087924

Heerspink, B. P., Kendall, A. D., Coe, M. T., & Hyndman, D. W. (2020). Trends in streamflow, evapotranspiration, and groundwater storage across the Amazon Basin linked to changing precipitation and land cover. *Journal of Hydrology: Regional Studies*, *32*, 100755. https://doi.org/10.1016/j.ejrh.2020.100755

Helsel, D. R., & Frans, L. M. (2006). Regional Kendall Test for Trend. *Environmental Science & Technology*, *40*(13), 4066–4073. https://doi.org/10.1021/es051650b

Hengl, T., Mendes de Jesus, J., Heuvelink, G. B. M., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., et al. (2017). SoilGrids250m: Global gridded soil information based on machine learning. *PLOS ONE*, *12*(2), e0169748. https://doi.org/10.1371/journal.pone.0169748

Herrera-Estrada, J. E., Satoh, Y., & Sheffield, J. (2017). Spatiotemporal dynamics of global drought. *Geophysical Research Letters*, *44*(5), 2254–2263. https://doi.org/10.1002/2016GL071768

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

Hirpa, F. A., Salamon, P., Beck, H. E., Lorini, V., Alfieri, L., Zsoter, E., & Dadson, S. J. (2018). Calibration of the Global Flood Awareness System (GloFAS) using daily streamflow data. *Journal of Hydrology*, *566*, 595–606. https://doi.org/10.1016/j.jhydrol.2018.09.052

Hodgkins, G. A., Whitfield, P. H., Burn, D. H., Hannaford, J., Renard, B., Stahl, K., et al. (2017). Climate-driven variability in the occurrence of major floods across North America and Europe. *Journal of Hydrology*, *552*, 704–717. https://doi.org/10.1016/j.jhydrol.2017.07.027

Hoekstra, A. Y., & Mekonnen, M. M. (2012). The water footprint of humanity. *Proceedings of the National Academy of Sciences*, *109*(9), 3232–3237. https://doi.org/10.1073/pnas.1109936109

Hothorn, T., & Zeileis, A. (2015). partykit: A modular toolkit for recursive partytioning in R. *The Journal of Machine Learning Research*, *16*(1), 3905–3909.

Huang, B., Thorne, P. W., Banzon, V. F., Boyer, T., Chepurin, G., Lawrimore, J. H., et al. (2017). NOAA Extended Reconstructed Sea Surface Temperature (ERSST), Version 5. *NOAA National Centers for Environmental Information*, *30*(8179–8205), 25.

Huscroft, J., Gleeson, T., Hartmann, J., & Börker, J. (2018). Compiling and Mapping Global Permeability of the Unconsolidated and Consolidated Earth: GLobal HYdrogeology MaPS 2.0 (GLHYMPS 2.0). *Geophysical Research Letters*, *45*(4), 1897–1904. https://doi.org/10.1002/2017GL075860

IBGE - Brazilian Institute of Geography and Statistics. (2007). *Censo Agropecuário 2006* (p. 141). Rio de Janeiro.

Institute of Hydrology. (1980). *Low Flow Studies: Report No 1*. Institute of Hydrology.

Ionita, M., Lohmann, G., Rimbu, N., Chelcea, S., & Dima, M. (2012). Interannual to decadal summer drought variability over Europe and its relationship to global sea surface temperature. *Climate Dynamics*, *38*(1), 363–377. https://doi.org/10.1007/s00382-011-1028-y

Jacques-Coper, M., & Garreaud, R. D. (2015). Characterization of the 1970s climate shift in South America. *International Journal of Climatology*, *35*(8), 2164–2179. https://doi.org/10.1002/joc.4120

Jammalamadaka, S. R., & Sengupta, A. (2001). *Topics in circular statistics* (Vol. 5). world scientific.

Junk, W. J., Piedade, M. T. F., Lourival, R., Wittmann, F., Kandus, P., Lacerda, L. D., et al. (2014). Brazilian wetlands: their definition, delineation, and classification for research, sustainable management, and protection. *Aquatic Conservation: Marine and Freshwater Ecosystems*, *24*(1), 5–22. https://doi.org/10.1002/aqc.2386

Kayano, M. T., Andreoli, R. V., & Ferreira de Souza, R. A. (2013). Relations between ENSO and the South Atlantic SST modes and their effects on the South American rainfall. *International Journal of Climatology*, *33*(8), 2008–2023. https://doi.org/10.1002/joc.3569

Kemter, M., Merz, B., Marwan, N., Vorogushyn, S., & Blöschl, G. (2020). Joint Trends in Flood Magnitudes and Spatial Extents Across Europe. *Geophysical Research Letters*, *47*(7), e2020GL087464. https://doi.org/10.1029/2020GL087464

Khanna, J., Medvigy, D., Fueglistaler, S., & Walko, R. (2017). Regional dry-season climate changes due to three decades of Amazonian deforestation. *Nature Climate Change*, *7*(3), 200–204. https://doi.org/10.1038/nclimate3226

Kidd, C., & Huffman, G. (2011). Global precipitation measurement. *Meteorological Applications*, *18*(3), 334–353. https://doi.org/10.1002/met.284

Kiem, A. S., & Franks, S. W. (2004). Multi-decadal variability of drought risk, eastern Australia. *Hydrological Processes*, *18*(11), 2039–2050. https://doi.org/10.1002/hyp.1460

Knoben, W. J. M., Woods, R. A., & Freer, J. E. (2018). A Quantitative Hydrological Climate Classification Evaluated With Independent Streamflow Data. *Water Resources Research*, *54*(7), 5088–5109. https://doi.org/10.1029/2018WR022913

Koenker, R. (2005). *Quantile regression* (Vol. 38). Cambridge University Press.

Kundzewicz, Z. W., Szwed, M., & Pińskwar, I. (2019). Climate Variability and Floods—A Global Review. *Water*, *11*(7), 1399. https://doi.org/10.3390/w11071399

Laaha, G., Demuth, S., Hisdal, H., Kroll, C. N., van Lanen, H. A., Nester, T., et al. (2013). Prediction of low flows in ungauged basins. In *Runoff Prediction in Ungauged Basins: Synthesis across Processes, Places and Scales* (pp. 163–188). Cambridge University Press.

Ladson, A., Brown, R., Neal, B., & Nathan, R. (2013). A standard approach to baseflow separation using the Lyne and Hollick filter. *Australian Journal of Water Resources*, *17*(1). https://doi.org/10.7158/W12-028.2013.17.1

Latrubesse, E. M., Arima, E. Y., Dunne, T., Park, E., Baker, V. R., d'Horta, F. M., et al. (2017). Damming the rivers of the Amazon basin. *Nature*, *546*(7658), 363–369. https://doi.org/10.1038/nature22333

de Lavenne, A., Andréassian, V., Crochemore, L., Lindström, G., & Arheimer, B. (2022). Quantifying multi-year hydrological memory with Catchment Forgetting Curves. *Hydrology and Earth System Sciences*, *26*(10), 2715–2732. https://doi.org/10.5194/hess-26-2715-2022

Lehner, B. (2012). *Derivation of watershed boundaries for GRDC gauging stations based on the HydroSHEDS drainage network*. Global Runoff Data Centre in the Federal Institute of Hydrology (BFG).

Lehner, B., Liermann, C. R., Revenga, C., Vörösmarty, C., Fekete, B., Crouzet, P., et al. (2011). High-resolution mapping of the world's reservoirs and dams for sustainable river-flow management. *Frontiers in Ecology and the Environment*, *9*(9), 494–502. https://doi.org/10.1890/100125

Leite, C. C., Costa, M. H., Soares-Filho, B. S., & de Barros Viana Hissa, L. (2012). Historical land use change and associated carbon emissions in Brazil from 1940 to 1995. *Global Biogeochemical Cycles*, *26*(2), 1–13. https://doi.org/10.1029/2011GB004133

Lesk, C., Rowhani, P., & Ramankutty, N. (2016). Influence of extreme weather disasters on global crop production. *Nature*, *529*(7584), 84–87. https://doi.org/10.1038/nature16467

Levy, M. C., Lopes, A. V., Cohn, A., Larsen, L. G., & Thompson, S. E. (2018). Land Use Change Increases Streamflow Across the Arc of Deforestation in Brazil. *Geophysical Research Letters*, *45*(8), 3520–3530. https://doi.org/10.1002/2017GL076526

Liebmann, B., Kiladis, G. N., Allured, D., Vera, C. S., Jones, C., Carvalho, L. M., et al. (2011). Mechanisms associated with large daily rainfall events in Northeast Brazil. *Journal of Climate*, *24*(2), 376–396. https://doi.org/10.1175/2010JCLI3457.1

Lima, C. H. R., & AghaKouchak, A. (2017). Droughts in Amazonia: Spatiotemporal Variability, Teleconnections, and Seasonal Predictions. *Water Resources Research*, *53*(12), 10824–10840. https://doi.org/10.1002/2016WR020086

Lima, C. H. R., AghaKouchak, A., & Lall, U. (2017). Classification of mechanisms, climatic context, areal scaling, and synchronization of floods: the hydroclimatology of floods in the Upper Paraná River basin, Brazil. *Earth System Dynamics*, *8*(4), 1071–1091. https://doi.org/10.5194/esd-8-1071-2017

Linke, S., Lehner, B., Ouellet Dallaire, C., Ariwi, J., Grill, G., Anand, M., et al. (2019). Global hydro-environmental sub-basin and river reach characteristics at high spatial resolution. *Scientific Data*, *6*(1), 1–15. https://doi.org/10.1038/s41597-019-0300-6

Liu, J., & Zhang, Y. (2017). Multi-temporal clustering of continental floods and associated atmospheric circulations. *Journal of Hydrology*, *555*, 744–759. https://doi.org/10.1016/j.jhydrol.2017.10.072

Liu, W., Wang, L., Zhou, J., Li, Y., Sun, F., Fu, G., et al. (2016). A worldwide evaluation of basin-scale evapotranspiration estimates against the water balance method. *Journal of Hydrology*, *538*, 82–95. https://doi.org/10.1016/j.jhydrol.2016.04.006

Longo, M., Saatchi, S., Keller, M., Bowman, K., Ferraz, A., Moorcroft, P. R., et al. (2020). Impacts of Degradation on Water, Energy, and Carbon Cycling of the Amazon Tropical Forests. *Journal of Geophysical Research: Biogeosciences*, *125*(8), e2020JG005677. https://doi.org/10.1029/2020jg005677

Luke, A., Vrugt, J. A., AghaKouchak, A., Matthew, R., & Sanders, B. F. (2017). Predicting nonstationary flood frequencies: Evidence supports an updated stationarity thesis in the United States. *Water Resources Research*, *53*(7), 5469–5494. https://doi.org/10.1002/2016WR019676

Lun, D., Fischer, S., Viglione, A., & Blöschl, G. (2020). Detecting Flood-Rich and Flood-Poor Periods in Annual Peak Discharges Across Europe. *Water Resources Research*, *56*(7), e2019WR026575. https://doi.org/10.1029/2019WR026575

MacKinnon, J. G., & White, H. (1985). Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *Journal of Econometrics*, *29*(3), 305–325. https://doi.org/10.1016/0304-4076(85)90158-7

Madari, B. E., Reeves, J. B., Machado, P. L. O. A., Guimarães, C. M., Torres, E., & McCarty, G. W. (2006). Mid- and near-infrared spectroscopic assessment of soil compositional parameters and structural indices in two Ferralsols. *Geoderma*, *136*(1), 245–259. https://doi.org/10.1016/j.geoderma.2006.03.026

Makkonen, L., & Pajari, M. (2014). Defining Sample Quantiles by the True Rank Probability. *Journal of Probability and Statistics*, *2014*, e326579. https://doi.org/10.1155/2014/326579

Mann, H. B. (1945). Nonparametric Tests Against Trend. *Econometrica*, *13*(3), 245. https://doi.org/10.2307/1907187

Mardia, K. V., & Jupp, P. E. (2009). *Directional statistics* (Vol. 494). John Wiley & Sons.

Marengo, J. A., & Espinoza, J. C. (2016). Extreme seasonal droughts and floods in Amazonia: causes, trends and impacts. *International Journal of Climatology*, *36*(3), 1033–1050. https://doi.org/10.1002/joc.4420

Marengo, J. A., Liebmann, B., Grimm, A. M., Misra, V., Silva Dias, P. L., Cavalcanti, I. F. A., et al. (2012). Recent developments on the South American monsoon system. *International Journal of Climatology*, *32*(1), 1–21. https://doi.org/10.1002/joc.2254

Marengo, J. A., Torres, R. R., & Alves, L. M. (2017). Drought in Northeast Brazil—past, present, and future. *Theoretical and Applied Climatology*, *129*(3–4), 1189–1200. https://doi.org/10.1007/s00704-016-1840-8

Marengo, J. A., Souza, C. M. J., Thonicke, K., Burton, C., Halladay, K., Betts, R. A., et al. (2018). Changes in Climate and Land Use Over the Amazon Region: Current and Future Variability and Trends. *Frontiers in Earth Science*, *6*. https://doi.org/10.3389/feart.2018.00228

Markonis, Y., & Koutsoyiannis, D. (2016). Scale-dependence of persistence in precipitation records. *Nature Climate Change*, *6*(4), 399–401. https://doi.org/10.1038/nclimate2894

Martens, B., Miralles, D. G., Lievens, H., van der Schalie, R., de Jeu, R. A. M., Fernández-Prieto, D., et al. (2017). GLEAM v3: satellite-based land evaporation and root-zone soil moisture. *Geoscientific Model Development*, *10*(5), 1903–1925. https://doi.org/10.5194/gmd-10-1903-2017

Martinez, J. A., & Dominguez, F. (2014). Sources of Atmospheric Moisture for the La Plata River Basin. *Journal of Climate*, *27*(17), 6737–6753. https://doi.org/10.1175/JCLI-D-14-00022.1

Martins, E. S., & Stedinger, J. R. (2000). Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data. *Water Resources Research*, *36*(3), 737–744. https://doi.org/10.1029/1999wr900330

Matricardi, E. A. T., Skole, D. L., Costa, O. B., Pedlowski, M. A., Samek, J. H., & Miguel, E. P. (2020). Long-term forest degradation surpasses deforestation in the Brazilian Amazon. *Science*, *369*(6509), 1378–1382. https://doi.org/10.1126/science.abb3021

McCabe, G. J., Betancourt, J. L., Gray, S. T., Palecki, M. A., & Hidalgo, H. G. (2008). Associations of multi-decadal sea-surface temperature variability with US drought. *Quaternary International*, *188*(1), 31–40. https://doi.org/10.1016/j.quaint.2007.07.001

McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior.

McMillan, H. (2022). A taxonomy of hydrological processes and watershed function. *Hydrological Processes*, *36*(3), e14537. https://doi.org/10.1002/hyp.14537

McMillan, H., Westerberg, I., & Branger, F. (2017). Five guidelines for selecting hydrological signatures. *Hydrological Processes*, *31*(26), 4757–4761. https://doi.org/10.1002/hyp.11300

McMillan, H. K., Gnann, S. J., & Araki, R. (2022). Large Scale Evaluation of Relationships Between Hydrologic Signatures and Processes. *Water Resources Research*, *58*(6), e2021WR031751. https://doi.org/10.1029/2021WR031751

Melo, D. de C. D., Scanlon, B. R., Zhang, Z., Wendland, E., & Yin, L. (2016). Reservoir storage and hydrologic responses to droughts in the Paraná River basin, south-eastern Brazil. *Hydrology and Earth System Sciences*, *20*(11), 4673–4688. https://doi.org/10.5194/hess-20-4673-2016

Merz, B., Vorogushyn, S., Uhlemann, S., Delgado, J., & Hundecha, Y. (2012). More efforts and scientific rigour are needed to attribute trends in flood time series. *Hydrology and Earth System Sciences*, *16*(5), 1379–1387. https://doi.org/10.5194/hess-16-1379-2012

Merz, B., Nguyen, V. D., & Vorogushyn, S. (2016). Temporal clustering of floods in Germany: Do flood-rich and flood-poor periods exist? *Journal of Hydrology*, *541*, 824–838. https://doi.org/10.1016/j.jhydrol.2016.07.041

Merz, R., & Blöschl, G. (2003). A process typology of regional floods. *Water Resources Research*, *39*(12). https://doi.org/10.1029/2002WR001952

Merz, R., & Blöschl, G. (2008a). Flood frequency hydrology: 1. Temporal, spatial, and causal expansion of information. *Water Resources Research*, *44*(8), 1–17. https://doi.org/10.1029/2007WR006744

Merz, R., & Blöschl, G. (2008b). Flood frequency hydrology: 2. Combining data evidence. *Water Resources Research*, *44*(8). https://doi.org/10.1029/2007WR006745

Milliman, J. D., Farnsworth, K. L., Jones, P. D., Xu, K. H., & Smith, L. C. (2008). Climatic and anthropogenic factors affecting river discharge to the global ocean, 1951–2000. *Global and Planetary Change*, *62*(3–4), 187–194. https://doi.org/10.1016/j.gloplacha.2008.03.001

Milly, P. C. D., Betancourt, J., Falkenmark, M., Hirsch, R. M., Kundzewicz, Z. W., Lettenmaier, D. P., et al. (2015). On Critiques of "Stationarity is Dead: Whither Water Management?" *Water Resources Research*, *51*(9), 7785–7789. https://doi.org/10.1002/2015WR017408

Miralles, D. G., Holmes, T. R. H., De Jeu, R. A. M., Gash, J. H., Meesters, A. G. C. A., & Dolman, A. J. (2011). Global land-surface evaporation estimated from satellite-based observations. *Hydrology and Earth System Sciences*, *15*(2), 453–469. https://doi.org/10.5194/hess-15-453-2011

Mo, K. C., & Berbery, E. H. (2011). Drought and Persistent Wet Spells over South America Based on Observations and the U.S. CLIVAR Drought Experiments. *Journal of Climate*, *24*(6), 1801–1820. https://doi.org/10.1175/2010JCLI3874.1

Montanari, A. (2007). What do we mean by 'uncertainty'? The need for a consistent wording about uncertainty assessment in hydrology. *Hydrological Processes*, *21*(6), 841–845. https://doi.org/10.1002/hyp.6623

Montanari, A., Young, G., Savenije, H. H. G., Hughes, D., Wagener, T., Ren, L. L., et al. (2013). "Panta Rhei—Everything Flows": Change in hydrology and society—The IAHS Scientific Decade 2013–2022. *Hydrological Sciences Journal*, *58*(6), 1256–1275. https://doi.org/10.1080/02626667.2013.809088

Morales, M. S., Cook, E. R., Barichivich, J., Christie, D. A., Villalba, R., LeQuesne, C., et al. (2020). Six hundred years of South American tree rings reveal an increase in severe hydroclimatic events since mid-20th century. *Proceedings of the National Academy of Sciences*, *117*(29), 16816–16823. https://doi.org/10.1073/pnas.2002411117

Müller Schmied, H., Eisner, S., Franz, D., Wattenbach, M., Portmann, F. T., Flörke, M., & Döll, P. (2014). Sensitivity of simulated global-scale freshwater fluxes and storages to input data, hydrological model structure, human water use and calibration. *Hydrology and Earth System Sciences*, *18*(9), 3511–3538. https://doi.org/10.5194/hess-18-3511-2014

Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., et al. (2021). ERA5-Land: a state-of-the-art global reanalysis dataset for land applications. *Earth System Science Data*, *13*(9), 4349–4383. https://doi.org/10.5194/essd-13-4349-2021

Nardi, F., Annis, A., Di Baldassarre, G., Vivoni, E. R., & Grimaldi, S. (2019). GFPLAIN250m, a global high-resolution dataset of Earth's floodplains. *Scientific Data*, *6*(1), 180309. https://doi.org/10.1038/sdata.2018.309

Neuman, S. P. (2005). Trends, prospects and challenges in quantifying flow and transport through fractured rocks. *Hydrogeology Journal*, *13*(1), 124–147. https://doi.org/10.1007/s10040-004-0397-2

Neves, M. A., & Morales, N. (2007). Well productivity controlling factors in crystalline terrains of southeastern Brazil. *Hydrogeology Journal*, *15*(3), 471–482. https://doi.org/10.1007/s10040-006-0112-6

Newman, A. J., Clark, M. P., Sampson, K., Wood, A., Hay, L. E., Bock, A., et al. (2015). Development of a large-sample watershed-scale hydrometeorological data set for the contiguous USA: data set characteristics and assessment of regional variability in hydrologic model performance. *Hydrology and Earth System Sciences*, *19*(1), 209–223. https://doi.org/10.5194/hess-19-209-2015

Newman, M., Alexander, M. A., Ault, T. R., Cobb, K. M., Deser, C., Di Lorenzo, E., et al. (2016). The Pacific Decadal Oscillation, Revisited. *Journal of Climate*, *29*(12), 4399–4427. https://doi.org/10.1175/JCLI-D-15-0508.1

NOAA. (2019a). CPC Global Temperature. Retrieved June 15, 2019, from https://www.esrl.noaa.gov/psd/

NOAA. (2019b). CPC Global Unified Precipitation. Retrieved June 15, 2019, from https://www.esrl.noaa.gov/psd/

Nobre, C. A., Sampaio, G., Borma, L. S., Castilla-Rubio, J. C., Silva, J. S., & Cardoso, M. (2016). Land-use and climate change risks in the Amazon and the need of a novel sustainable development paradigm. *Proceedings of the National Academy of Sciences*, *113*(39), 10759–10768. https://doi.org/10.1073/pnas.1605516113

O'Connell, E., O'Donnell, G., & Koutsoyiannis, D. (2022). On the spatial scale dependence of long-term persistence in global annual precipitation data and the Hurst Phenomenon. *Water Resources Research*, e2022WR033133.

Olden, J. D., & Poff, N. L. (2003). Redundancy and the choice of hydrologic indices for characterizing streamflow regimes. *River Research and Applications*, *19*(2), 101–121. https://doi.org/10.1002/rra.700

ONS - National Electrical System Operator. (2019). SIGEL - Sistema Geográfico de Informações do Sistema Elétrico. Retrieved December 10, 2019, from https://sigel.aneel.gov.br/

d'Orgeville, M., & Peltier, W. R. (2007). On the Pacific Decadal Oscillation and the Atlantic Multidecadal Oscillation: Might they be related? *Geophysical Research Letters*, *34*(23). https://doi.org/10.1029/2007GL031584

Paiva, R. C. D., Buarque, D. C., Collischonn, W., Bonnet, M.-P., Frappart, F., Calmant, S., & Bulhões Mendes, C. A. (2013). Large-scale hydrologic and hydrodynamic modeling of the Amazon River basin. *Water Resources Research*, *49*(3), 1226–1243. https://doi.org/10.1002/wrcr.20067

Parajka, J., Kohnová, S., Bálint, G., Barbuc, M., Borga, M., Claps, P., et al. (2010). Seasonal characteristics of flood regimes across the Alpine–Carpathian range. *Journal of Hydrology*, *394*(1), 78–89. https://doi.org/10.1016/j.jhydrol.2010.05.015

Pasquini, A. I., & Depetris, P. J. (2007). Discharge trends and flow dynamics of South American rivers draining the southern Atlantic seaboard: An overview. *Journal of Hydrology*, *333*(2–4), 385–399. https://doi.org/10.1016/j.jhydrol.2006.09.005

Pebesma, E. J. (2004). Multivariable geostatistics in S: the gstat package. *Computers & Geosciences*, *30*(7), 683–691. https://doi.org/10.1016/j.cageo.2004.03.012

Pekel, J.-F., Cottam, A., Gorelick, N., & Belward, A. S. (2016). High-resolution mapping of global surface water and its long-term changes. *Nature*, *540*(7633), 418–422. https://doi.org/10.1038/nature20584

Pelletier, J. D., Broxton, P. D., Hazenberg, P., Zeng, X., Troch, P. A., Niu, G., et al. (2016). A gridded global data set of soil, intact regolith, and sedimentary deposit thicknesses for regional and global land surface modeling. *Journal of Advances in Modeling Earth Systems*, *8*(1), 41–65. https://doi.org/10.1002/2015MS000526

Peña-Arancibia, J. L., van Dijk, A. I. J. M., Mulligan, M., & Bruijnzeel, L. A. (2010). The role of climatic and terrain attributes in estimating baseflow recession in tropical catchments. *Hydrology and Earth System Sciences*, *14*(11), 2193–2205. https://doi.org/10.5194/hess-14-2193-2010

Pfister, L., & Kirchner, J. W. (2017). Debates-Hypothesis testing in hydrology: Theory and practice. *Water Resources Research*, *53*(3), 1792–1798. https://doi.org/10.1002/2016WR020116

Pfister, L., Martínez-Carreras, N., Hissler, C., Klaus, J., Carrer, G. E., Stewart, M. K., & McDonnell, J. J. (2017). Bedrock geology controls on catchment storage, mixing, and release: A comparative analysis of 16 nested catchments. *Hydrological Processes*, *31*(10), 1828–1845. https://doi.org/10.1002/hyp.11134

Phillips, O. L., Aragão, L. E. O. C., Lewis, S. L., Fisher, J. B., Lloyd, J., López-González, G., et al. (2009). Drought Sensitivity of the Amazon Rainforest. *Science*, *323*(5919), 1344–1347. https://doi.org/10.1126/science.1164033

Philpotts, A. R., & Ague, J. J. (2009). *Principles of igneous and metamorphic petrology*. Cambridge University Press.

Poggio, L., de Sousa, L. M., Batjes, N. H., Heuvelink, G. B. M., Kempen, B., Ribeiro, E., & Rossiter, D. (2021). SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty. *SOIL*, *7*(1), 217–240. https://doi.org/10.5194/soil-7-217-2021

Ponce, V. M. (1995). Management of droughts and floods in the semiarid Brazilian Northeast—The case for conservation. *Journal of Soil and Water Conservation*, *50*(5), 422–431.

Price, K. (2011). Effects of watershed topography, soils, land use, and climate on baseflow hydrology in humid regions: A review. *Progress in Physical Geography: Earth and Environment*, *35*(4), 465–492. https://doi.org/10.1177/0309133311402714

Price, K., Jackson, C. R., Parker, A. J., Reitan, T., Dowd, J., & Cyterski, M. (2011). Effects of watershed land use and geomorphology on stream low flows during severe drought conditions in the southern Blue Ridge Mountains, Georgia and North Carolina, United States. *Water Resources Research*, *47*(2). https://doi.org/10.1029/2010WR009340

Raia, A., & Cavalcanti, I. F. A. (2008). The Life Cycle of the South American Monsoon System. *Journal of Climate*, *21*(23), 6227–6246. https://doi.org/10.1175/2008JCLI2249.1

Rajão, R., Soares-Filho, B., Nunes, F., Börner, J., Machado, L., Assis, D., et al. (2020). The rotten apples of Brazil's agribusiness. *Science*, *369*(6501), 246–248. https://doi.org/10.1126/science.aba6646

Ratan, R., & Venugopal, V. (2013). Wet and dry spell characteristics of global tropical rainfall. *Water Resources Research*, *49*(6), 3830–3841. https://doi.org/10.1002/wrcr.20275

Rayleigh, Lord. (1880). XII. On the resultant of a large number of vibrations of the same pitch and of arbitrary phase. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, *10*(60), 73–78. https://doi.org/10.1080/14786448008626893

Reichert, J. M., Albuquerque, J. A., Solano Peraza, J. E., & da Costa, A. (2020). Estimating water retention and availability in cultivated soils of southern Brazil. *Geoderma Regional*, *21*, e00277. https://doi.org/10.1016/j.geodrs.2020.e00277

Rogger, M., Agnoletti, M., Alaoui, A., Bathurst, J. C., Bodner, G., Borga, M., et al. (2017). Land use change impacts on floods at the catchment scale: Challenges and opportunities for future research. *Water Resources Research*, *53*(7), 5209–5219. https://doi.org/10.1002/2017WR020723

Rosbjerg, D., Bloschl, G., Burn, D., Castellarin, A., Croke, B., Di Baldassarre, G., et al. (2013). Prediction of floods in ungauged basins. In *Runoff prediction in ungauged basins: Synthesis across processes, places and scales* (pp. 189–225). Cambridge University Press.

Salgado, A. A. R., Assis, W. L., Magalhães Júnior, A. P., do Carmo, F. F., de Sordi, M. V., & de Oliveira, F. S. (2019). Semi-humid: The Landscape of Central Brazil. In *The Physical Geography of Brazil: Environment, Vegetation and Landscape* (pp. 93–117).

Salinas, J. L., Laaha, G., Rogger, M., Parajka, J., Viglione, A., Sivapalan, M., & Blöschl, G. (2013). Comparative assessment of predictions in ungauged basins - Part 2: Flood and low flow studies. *Hydrology and Earth System Sciences*, *17*(7), 2637–2652. https://doi.org/10.5194/hess-17-2637-2013

Salio, P., Nicolini, M., & Zipser, E. J. (2007). Mesoscale convective systems over southeastern South America and their relationship with the South American low-level jet. *Monthly Weather Review*, *135*(4), 1290–1309. https://doi.org/10.1175/MWR3305.1

Sankarasubramanian, A., Vogel, R. M., & Limbrunner, J. F. (2001). Climate elasticity of streamflow in the United States. *Water Resources Research*, *37*(6), 1771–1781. https://doi.org/10.1029/2000WR900330

dos Santos, H. G., Carvalho Junior, W. de, Dart, R. de O., Áglio, M. L. D., de Sousa, J. S., Pares, J. G., et al. (2011). O novo mapa de solos do Brasil: legenda atualizada.

Saurral, R. I., Camilloni, I. A., & Barros, V. R. (2017). Low-frequency variability and trends in centennial precipitation stations in southern South America. *International Journal of Climatology*, *37*(4), 1774–1793. https://doi.org/10.1002/joc.4810

Sawicz, K. A., Kelleher, C., Wagener, T., Troch, P., Sivapalan, M., & Carrillo, G. (2014). Characterizing hydrologic change through catchment classification. *Hydrology and Earth System Sciences*, *18*(1), 273–285. https://doi.org/10.5194/hess-18-273-2014

Sayama, T., McDonnell, J. J., Dhakal, A., & Sullivan, K. (2011). How much water can a watershed store? *Hydrological Processes*, *25*(25), 3899–3908. https://doi.org/10.1002/hyp.8288

Schewe, J., Heinke, J., Gerten, D., Haddeland, I., Arnell, N. W., Clark, D. B., et al. (2014). Multimodel assessment of water scarcity under climate change. *Proceedings of the National Academy of Sciences*, *111*(9), 3245–3250. https://doi.org/10.1073/pnas.1222460110

Schobbenhaus, C., Gonçalves, J., Santos, J., Abram, M., Leão Neto, R., Matos, G., et al. (2004). *Carta geológica do Brasil ao milionésimo: Sistema de Informações Geográficas*. CPRM - Serviço Geológico do Brasil, Brasília.

Schöngart, J., & Junk, W. J. (2007). Forecasting the flood-pulse in Central Amazonia by ENSO-indices. *Journal of Hydrology*, *335*(1–2), 124–132. https://doi.org/10.1016/j.jhydrol.2006.11.005

Schwalm, C. R., Anderegg, W. R. L., Michalak, A. M., Fisher, J. B., Biondi, F., Koch, G., et al. (2017). Global patterns of drought recovery. *Nature*, *548*(7666), 202–205. https://doi.org/10.1038/nature23021

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 461–464.

Screen, J. A., & Simmonds, I. (2014). Amplified mid-latitude planetary waves favour particular regional weather extremes. *Nature Climate Change*, *4*(8), 704–709. https://doi.org/10.1038/nclimate2271

Seager, R. (2015). Decadal Hydroclimate Variability Across the Americas. In C.-P. Chang, M. Ghil, M. Latif, & J. M. Wallace, *World Scientific Series on Asia-Pacific Weather and Climate* (Vol. 6, pp. 235–254). WORLD SCIENTIFIC. https://doi.org/10.1142/9789814579933_0015

Seager, R., Naik, N., Baethgen, W., Robertson, A., Kushnir, Y., Nakamura, J., & Jurburg, S. (2010). Tropical Oceanic Causes of Interannual to Multidecadal Precipitation Variability in Southeast South America over the Past Century. *Journal of Climate*, *23*(20), 5517–5539. https://doi.org/10.1175/2010JCLI3578.1

Sen, P. K. (1968). Estimates of the Regression Coefficient Based on Kendall's Tau. *Journal of the American Statistical Association*, *63*(324), 1379–1389. https://doi.org/10.1080/01621459.1968.10480934

Sena, J. A., Beser de Deus, L. A., Freitas, M. A. V., & Costa, L. (2012). Extreme Events of Droughts and Floods in Amazonia: 2005 and 2009. *Water Resources Management*, *26*(6), 1665–1676. https://doi.org/10.1007/s11269-012-9978-3

Seneviratne, S. I., Corti, T., Davin, E. L., Hirschi, M., Jaeger, E. B., Lehner, I., et al. (2010). Investigating soil moisture–climate interactions in a changing climate: A review. *Earth-Science Reviews*, *99*(3–4), 125–161. https://doi.org/10.1016/j.earscirev.2010.02.004

SGB - Geological Survey of Brazil. (2021). Mapa Integrado do Brasil ao Milionésimo. *SGB*.

Shangguan, W., Hengl, T., Mendes de Jesus, J., Yuan, H., & Dai, Y. (2017). Mapping the global depth to bedrock for land surface modeling. *Journal of Advances in Modeling Earth Systems*, *9*(1), 65–88. https://doi.org/10.1002/2016MS000686

Sharma, A., Wasko, C., & Lettenmaier, D. P. (2018). If Precipitation Extremes Are Increasing, Why Aren't Floods? *Water Resources Research*, *54*(11), 8545–8551. https://doi.org/10.1029/2018WR023749

Sheffield, J., Andreadis, K. M., Wood, E. F., & Lettenmaier, D. P. (2009). Global and Continental Drought in the Second Half of the Twentieth Century: Severity–Area–Duration Analysis and Temporal Variability of Large-Scale Events. *Journal of Climate*, *22*(8), 1962–1981. https://doi.org/10.1175/2008JCLI2722.1

Shen, X., Anagnostou, E. N., Mei, Y., & Hong, Y. (2017). A global distributed basin morphometric dataset. *Scientific Data*, *4*(1), 160124. https://doi.org/10.1038/sdata.2016.124

Silva, A. P. da, Babujia, L. C., Franchini, J. C., Ralisch, R., Hungria, M., & Guimarães, M. de F. (2014). Soil structure and its influence on microbial biomass in different soil and crop management systems. *Soil and Tillage Research*, *142*, 42–53. https://doi.org/10.1016/j.still.2014.04.006

Singh, R., Archfield, S. A., & Wagener, T. (2014). Identifying dominant controls on hydrologic parameter transfer from gauged to ungauged catchments–A comparative hydrology approach. *Journal of Hydrology*, *517*, 985–996. https://doi.org/10.1016/j.jhydrol.2014.06.030

Siqueira, V. A., Paiva, R. C. D., Fleischmann, A. S., Fan, F. M., Ruhoff, A. L., Pontes, P. R. M., et al. (2018). Toward continental hydrologic–hydrodynamic modeling in South America. *Hydrology and Earth System Sciences*, *22*(9), 4815–4842. https://doi.org/10.5194/hess-22-4815-2018

Sivapalan, M., Blöschl, G., Merz, R., & Gutknecht, D. (2005). Linking flood frequency to long-term water balance: Incorporating effects of seasonality. *Water Resources Research*, *41*(6). https://doi.org/10.1029/2004WR003439

Skansi, M. de los M., Brunet, M., Sigró, J., Aguilar, E., Arevalo Groening, J. A., Bentancur, O. J., et al. (2013). Warming and wetting signals emerging from analysis of changes in climate extreme indices over South America. *Global and Planetary Change*, *100*, 295–307. https://doi.org/10.1016/j.gloplacha.2012.11.004

Slater, L., Villarini, G., Archfield, S., Faulkner, D., Lamb, R., Khouakhi, A., & Yin, J. (2021). Global Changes in 20-Year, 50-Year, and 100-Year River Floods. *Geophysical Research Letters*, *48*(6). https://doi.org/10.1029/2020GL091824

Slater, L. J., Singer, M. B., & Kirchner, J. W. (2015). Hydrologic versus geomorphic drivers of trends in flood hazard. *Geophysical Research Letters*, *42*(2), 370–376. https://doi.org/10.1002/2014GL062482

Smakhtin, V. U. (2001). Low flow hydrology: a review. *Journal of Hydrology*, *240*(3–4), 147–186. https://doi.org/10.1016/S0022-1694(00)00340-1

Song, X.-P., Hansen, M. C., Stehman, S. V., Potapov, P. V., Tyukavina, A., Vermote, E. F., & Townshend, J. R. (2018). Global land change from 1982 to 2016. *Nature*, *560*(7720), 639–643. https://doi.org/10.1038/s41586-018-0411-9

Sophocleous, M. (2002). Interactions between groundwater and surface water: the state of the science. *Hydrogeology Journal*, *10*(1), 52–67. https://doi.org/10.1007/s10040-001-0170-8

Staal, A., Tuinenburg, O. A., Bosmans, J. H. C., Holmgren, M., van Nes, E. H., Scheffer, M., et al. (2018). Forest-rainfall cascades buffer against drought across the Amazon. *Nature Climate Change*, *8*(6), 539–543. https://doi.org/10.1038/s41558-018-0177-y

Staudinger, M., Stoelzle, M., Seeger, S., Seibert, J., Weiler, M., & Stahl, K. (2017). Catchment water storage variation with elevation. *Hydrological Processes*, *31*(11), 2000–2015. https://doi.org/10.1002/hyp.11158

Stedinger, J. R., Vogel, R. M., & Foufoula-Georgiou, E. (1993). Frequency Analysis of Extreme Events. In *Handbook of Hydrology* (pp. 697–754). McGraw-Hill,.

Stein, L., Pianosi, F., & Woods, R. (2020). Event-based classification for global study of river flood generating processes. *Hydrological Processes*, *34*(7), 1514–1529. https://doi.org/10.1002/hyp.13678

Steinschneider, S., Yang, Y.-C. E., & Brown, C. (2013). Panel regression techniques for identifying impacts of anthropogenic landscape change on hydrologic response. *Water Resources Research*, *49*(12), 7874–7886. https://doi.org/10.1002/2013WR013818

Stevaux, J. C., Martins, D. P., & Meurer, M. (2009). Changes in a large regulated tropical river: The Paraná River downstream from the Porto Primavera Dam, Brazil. *Geomorphology*, *113*(3–4), 230–238. https://doi.org/10.1016/j.geomorph.2009.03.015

Su, X., Wang, M., & Fan, J. (2004). Maximum likelihood regression trees. *Journal of Computational and Graphical Statistics*, *13*(3), 586–598. https://doi.org/10.1198/106186004X2165

Sun, Q., Miao, C., Duan, Q., Ashouri, H., Sorooshian, S., & Hsu, K. (2018). A Review of Global Precipitation Data Sets: Data Sources, Estimation, and Intercomparisons. *Reviews of Geophysics*, *56*(1), 79–107. https://doi.org/10.1002/2017RG000574

Swain, D. L., Langenbrunner, B., Neelin, J. D., & Hall, A. (2018). Increasing precipitation volatility in twenty-first-century California. *Nature Climate Change*, *8*(5), 427–433. https://doi.org/10.1038/s41558-018-0140-y

Tague, C., & Grant, G. E. (2004). A geological framework for interpreting the low-flow regimes of Cascade streams, Willamette River Basin, Oregon. *Water Resources Research*, *40*(4). https://doi.org/10.1029/2003WR002629

Tallaksen, L. M. (1995). A review of baseflow recession analysis. *Journal of Hydrology*, *165*(1), 349–370. https://doi.org/10.1016/0022-1694(94)02540-R

Tarasova, L., Merz, R., Kiss, A., Basso, S., Blöschl, G., Merz, B., et al. (2019). Causative classification of river flood events. *WIREs Water*, *6*(4), e1353. https://doi.org/10.1002/wat2.1353

Tarasova, L., Lun, D., Merz, R., Blöschl, G., Basso, S., Bertola, M., et al. (2023). Shifts in flood generation processes exacerbate regional flood anomalies in Europe. *Communications Earth & Environment*, *4*(1), 1–12. https://doi.org/10.1038/s43247-023-00714-8

Tashie, A., Pavelsky, T., & Band, L. E. (2020). An Empirical Reevaluation of Streamflow Recession Analysis at the Continental Scale. *Water Resources Research*, *56*(1). https://doi.org/10.1029/2019WR025448

Tedeschi, R. G., Cavalcanti, I. F. A., & Grimm, A. M. (2013). Influences of two types of ENSO on South American precipitation. *International Journal of Climatology*, *33*(6), 1382–1400. https://doi.org/10.1002/joc.3519

Tedeschi, R. G., Grimm, A. M., & Cavalcanti, I. F. A. (2015). Influence of Central and East ENSO on extreme events of precipitation in South America during austral spring and summer. *International Journal of Climatology*, *35*(8), 2045–2064. https://doi.org/10.1002/joc.4106

Theil, H. (1950). A Rank-Invariant Method of Linear and Polynomial Regression Analysis. *Proceedings of the Royal Netherlands Academy of Sciences*, *53*, 386–392.

Tomasella, J., & Hodnett, M. G. (1997). Estimating unsaturated hydraulic conductivity of Brazilian soils using soil-water retention data. *Soil Science*, *162*(10), 703–712. https://doi.org/10.1097/00010694-199710000-00003

Tomasella, J., Borma, L. S., Marengo, J. A., Rodriguez, D. A., Cuartas, L. A., A. Nobre, C., & Prado, M. C. R. (2011). The droughts of 1996-1997 and 2004-2005 in Amazonia: hydrological response in the river main-stem. *Hydrological Processes*, *25*(8), 1228–1242. https://doi.org/10.1002/hyp.7889

Tomasella, J., Cunha, A. P. M. A., Simões, P. A., & Zeri, M. (2022). Assessment of trends, variability and impacts of droughts across Brazil over the period 1980–2019. *Natural Hazards*. https://doi.org/10.1007/s11069-022-05759-0

Towner, J., Ficchí, A., Cloke, H. L., Bazo, J., Coughlan de Perez, E., & Stephens, E. M. (2021). Influence of ENSO and tropical Atlantic climate variability on flood characteristics in the Amazon basin. *Hydrology and Earth System Sciences*, *25*(7), 3875–3895. https://doi.org/10.5194/hess-25-3875-2021

Tramblay, Y., Villarini, G., Khalki, E. M. E., Gründemann, G., & Hughes, D. (2021). Evaluation of the drivers responsible for flooding in Africa. *Water Resources Research*, *n/a*(n/a), e2021WR029595. https://doi.org/10.1029/2021WR029595

Trenberth, K. (2011). Changes in precipitation with climate change. *Climate Research*, *47*(1), 123–138. https://doi.org/10.3354/cr00953

Troch, P. A., Lahmers, T., Meira, A., Mukherjee, R., Pedersen, J. W., Roy, T., & Valdés-Pineda, R. (2015). Catchment coevolution: A useful framework for improving predictions of hydrological change? *Water Resources Research*, *51*(7), 4903–4922. https://doi.org/10.1002/2015WR017032

UNDRR. (2019). *Global assessment report on disaster risk reduction : 2019*. Geneva: UN Office for Disaster Risk Reduction.

Valani, G. P., Martíni, A. F., Pezzopane, J. R. M., Bernardi, A. C. de C., & Cooper, M. (2022). Soil physical quality in the topsoil of integrated and non-integrated grazing systems in a Brazilian Ferralsol. *Soil and Tillage Research*, *220*, 105357. https://doi.org/10.1016/j.still.2022.105357

Van Lanen, H. A. J., Wanders, N., Tallaksen, L. M., & Van Loon, A. F. (2013). Hydrological drought across the world: impact of climate and physical catchment structure. *Hydrology and Earth System Sciences*, *17*(5), 1715–1732. https://doi.org/10.5194/hess-17-1715-2013

Van Loon, A. F. (2015). Hydrological drought explained. *WIREs Water*, *2*(4), 359–392. https://doi.org/10.1002/wat2.1085

Van Loon, A. F., & Van Lanen, H. A. J. (2012). A process-based typology of hydrological drought. *Hydrology and Earth System Sciences*, *16*(7), 1915–1946. https://doi.org/10.5194/hess-16-1915-2012

Van Loon, A. F., Tijdeman, E., Wanders, N., Van Lanen, H. a. J., Teuling, A. J., & Uijlenhoet, R. (2014). How climate seasonality modifies drought duration and deficit. *Journal of Geophysical Research: Atmospheres*, *119*(8), 4640–4656. https://doi.org/10.1002/2013jd020383

Veldkamp, T. I. E., Zhao, F., Ward, P. J., De Moel, H., Aerts, J. C., Schmied, H. M., et al. (2018). Human impact parameterizations in global hydrological models improve estimates of monthly

discharges and hydrological extremes: a multi-model validation study. *Environmental Research Letters*, *13*(5), 055008. https://doi.org/10.1088/1748-9326/aab96f

Vieira, B. C., & Gramani, M. F. (2015). Serra do Mar: the most "tormented" relief in Brazil. In *Landscapes and landforms of Brazil* (pp. 285–297). Springer.

Viglione, A., Merz, R., Salinas, J. L., & Blöschl, G. (2013). Flood frequency hydrology: 3. A Bayesian analysis. *Water Resources Research*, *49*(2), 675–692. https://doi.org/10.1029/2011WR010782

Viglione, A., Merz, B., Viet Dung, N., Parajka, J., Nester, T., & Blöschl, G. (2016). Attribution of regional flood changes based on scaling fingerprints. *Water Resources Research*, *52*(7), 5322–5340. https://doi.org/10.1002/2016WR019036

Villarini, G. (2016). On the seasonality of flooding across the continental United States. *Advances in Water Resources*, *87*, 80–91. https://doi.org/10.1016/j.advwatres.2015.11.009

de Vries, J. J., & Simmers, I. (2002). Groundwater recharge: an overview of processes and challenges. *Hydrogeology Journal*, *10*(1), 5–17. https://doi.org/10.1007/s10040-001-0171-7

Vrugt, J. A. (2016). Markov chain Monte Carlo simulation using the DREAM software package: Theory, concepts, and MATLAB implementation. *Environmental Modelling & Software*, *75*, 273–316. https://doi.org/10.1016/j.envsoft.2015.08.013

Vuille, M., Burns, S. J., Taylor, B. L., Cruz, F. W., Bird, B. W., Abbott, M. B., et al. (2012). A review of the South American monsoon history as recorded in stable isotopic proxies over the past two millennia. *Climate of the Past*, *8*(4), 1309–1321. https://doi.org/10.5194/cp-8-1309-2012

Wada, Y., Wisser, D., & Bierkens, M. F. P. (2014). Global modeling of withdrawal, allocation and consumptive use of surface water and groundwater resources. *Earth System Dynamics*, *5*(1), 15–40. https://doi.org/10.5194/esd-5-15-2014

Wagener, T., Gleeson, T., Coxon, G., Hartmann, A., Howden, N., Pianosi, F., et al. (2021). On doing hydrology with dragons: Realizing the value of perceptual models and knowledge accumulation. *WIREs Water*, *8*(6). https://doi.org/10.1002/wat2.1550

Wang, S.-Y. S., Yoon, J.-H., Becker, E., & Gillies, R. (2017). California from drought to deluge. *Nature Climate Change*, *7*(7), 465–468. https://doi.org/10.1038/nclimate3330

Wang-Erlandsson, L., Bastiaanssen, W. G. M., Gao, H., Jägermeyr, J., Senay, G. B., van Dijk, A. I. J. M., et al. (2016). Global root zone storage capacity from satellite-based evaporation. *Hydrology and Earth System Sciences*, *20*(4), 1459–1481. https://doi.org/10.5194/hess-20-1459-2016

Ward, P. J., de Ruiter, M. C., Mård, J., Schröter, K., Van Loon, A., Veldkamp, T., et al. (2020). The need to integrate flood and drought disaster risk reduction strategies. *Water Security*, *11*, 100070. https://doi.org/10.1016/j.wasec.2020.100070

Wasko, C., Nathan, R., & Peel, M. C. (2020). Changes in Antecedent Soil Moisture Modulate Flood Seasonality in a Changing Climate. *Water Resources Research*, *56*(3). https://doi.org/10.1029/2019WR026300

Werner, M. (2001). Shuttle radar topography mission (SRTM) mission overview. *Frequenz*, *55*(3–4), 75–79.

Westerberg, I. K., & McMillan, H. K. (2015). Uncertainty in hydrological signatures. *Hydrology and Earth System Sciences*, *19*(9), 3951–3968. https://doi.org/10.5194/hess-19-3951-2015

Western, A. W., Grayson, R. B., & Blöschl, G. (2002). Scaling of Soil Moisture: A Hydrologic Perspective. *Annual Review of Earth and Planetary Sciences*, *30*(1), 149–180. https://doi.org/10.1146/annurev.earth.30.091201.140434

Wilcox, R. R. (1998). A Note on the Theil-Sen Regression Estimator When the Regressor Is Random and the Error Term Is Heteroscedastic. *Biometrical Journal*, *40*(3), 261–268. https://doi.org/10.1002/(SICI)1521-4036(199807)40:3<261::AID-BIMJ261>3.0.CO;2-V

Winsemius, H. C., Aerts, J. C. J. H., van Beek, L. P. H., Bierkens, M. F. P., Bouwman, A., Jongman, B., et al. (2016). Global drivers of future river flood risk. *Nature Climate Change*, *6*(4), 381–385. https://doi.org/10.1038/nclimate2893

Wohl, E., Barros, A., Brunsell, N., Chappell, N. A., Coe, M., Giambelluca, T., et al. (2012). The hydrology of the humid tropics. *Nature Climate Change*, *2*(9), 655–662. https://doi.org/10.1038/nclimate1556

Woldemeskel, F., & Sharma, A. (2016). Should flood regimes change in a warming climate? The role of antecedent moisture conditions. *Geophysical Research Letters*, *43*(14), 7556–7563. https://doi.org/10.1002/2016GL069448

Wongchuig, S. C., de Paiva, R. C. D., Siqueira, V., & Collischonn, W. (2019). Hydrological reanalysis across the 20th century: A case study of the Amazon Basin. *Journal of Hydrology*, *570*, 755–773. https://doi.org/10.1016/j.jhydrol.2019.01.025

Wongchuig-Correa, S., Paiva, R. C. D. de, Espinoza, J. C., & Collischonn, W. (2017). Multi-decadal Hydrological Retrospective: Case study of Amazon floods and droughts. *Journal of Hydrology*, *549*, 667–684. https://doi.org/10.1016/j.jhydrol.2017.04.019

WRB, I. W. G. (2015). World Reference Base for Soil Resources 2014, *201*(5).

Xavier, A. C., King, C. W., & Scanlon, B. R. (2016). Daily gridded meteorological variables in Brazil (1980-2013). *International Journal of Climatology*, *36*(6), 2644–2659. https://doi.org/10.1002/joc.4518

Yadav, M., Wagener, T., & Gupta, H. (2007). Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins. *Advances in Water Resources*, *30*(8), 1756–1774. https://doi.org/10.1016/j.advwatres.2007.01.005

Yokoo, Y., & Sivapalan, M. (2011). Towards reconstruction of the flow duration curve: development of a conceptual framework with a physical basis. *Hydrology and Earth System Sciences*, *15*(9), 2805–2819. https://doi.org/10.5194/hess-15-2805-2011

Yue, S., Pilon, P., Phinney, B., & Cavadias, G. (2002). The influence of autocorrelation on the ability to detect trend in hydrological series. *Hydrological Processes*, *16*(9), 1807–1829. https://doi.org/10.1002/hyp.1095

Zeileis, A. (2006). Object-Oriented Computation of Sandwich Estimators. *Journal of Statistical Software*, *16*(9), 1–16. https://doi.org/10.18637/jss.v016.i09

Zeileis, A., Hothorn, T., & Hornik, K. (2008). Model-Based Recursive Partitioning. *Journal of Computational and Graphical Statistics*, *17*(2), 492–514. https://doi.org/10.1198/106186008X319331

Zeileis, A., Köll, S., & Graham, N. (2020). Various Versatile Variances: An Object-Oriented Implementation of Clustered Covariances in R. *Journal of Statistical Software*, *95*(1), 1–36. https://doi.org/10.18637/jss.v095.i01

Zhang, W., Villarini, G., Vecchi, G. A., & Smith, J. A. (2018). Urbanization exacerbated the rainfall and flooding caused by hurricane Harvey in Houston. *Nature*, *563*(7731), 384–388. https://doi.org/10.1038/s41586-018-0676-z

Zhao, F., Veldkamp, T. I., Frieler, K., Schewe, J., Ostberg, S., Willner, S., et al. (2017). The critical role of the routing scheme in simulating peak river discharge in global hydrological models. *Environmental Research Letters*, *12*(7), 075003. https://doi.org/10.1088/1748-9326/aa7250

Zilli, M. T., Carvalho, L. M. V., & Lintner, B. R. (2019). The poleward shift of South Atlantic Convergence Zone in recent decades. *Climate Dynamics*, *52*(5), 2545–2563. https://doi.org/10.1007/s00382-018-4277-1

Zscheischler, J., Westra, S., van den Hurk, B. J. J. M., Seneviratne, S. I., Ward, P. J., Pitman, A., et al. (2018). Future climate risk from compound events. *Nature Climate Change*, *8*(6), 469–477. https://doi.org/10.1038/s41558-018-0156-3

Zscheischler, J., Martius, O., Westra, S., Bevacqua, E., Raymond, C., Horton, R. M., et al. (2020). A typology of compound weather and climate events. *Nature Reviews Earth & Environment*, *1*(7), 333–347. https://doi.org/10.1038/s43017-020-0060-z

Appendix A

Supporting Information of Chapter 3 – Process controls on flood seasonality in Brazil

Table A.1. Mean correlations between the interannual variability of floods and maximum annual rainfall or maximum annual soil moisture according to their: (i) timing (Fig. 3.3); (ii) timing considering only the lowest 50% of floods (Fig. A.3); (iii) timing considering only the highest 50% of floods (Fig. A.3); and (iv) magnitudes (Fig. A.4). We considered only the spatially interpolated correlations. The timing variables are correlated using circular correlation and the magnitude variables using the Spearman rank correlation.

Variable	Mean correlation (floods and rainfall)	Mean correlation (floods and soil moisture)
Timing	0.25	0.37
Timing (lowest floods)	0.16	0.29
Timing (highest floods)	0.35	0.44
Magnitude	0.43	0.56

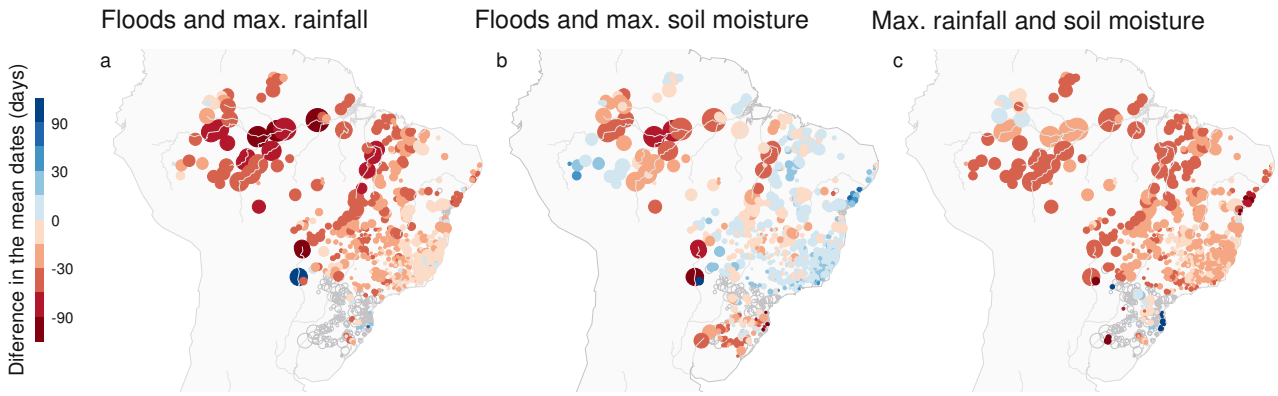

Figure A.1. Differences between the mean dates of floods and their drivers. (a) Floods and maximum annual rainfall, with red colors indicating floods that occur after maximum rainfall. (b) Floods and maximum annual soil moisture, with red colors showing floods that occur after maximum soil moisture. (c) Maximum annual rainfall and soil moisture, with red colors showing maximum rainfall that occurs before maximum soil moisture. Gray open circles are non-significant or non-unimodal seasonalities (α = 0.05). Circle size is proportional to the logarithm of the basin area.

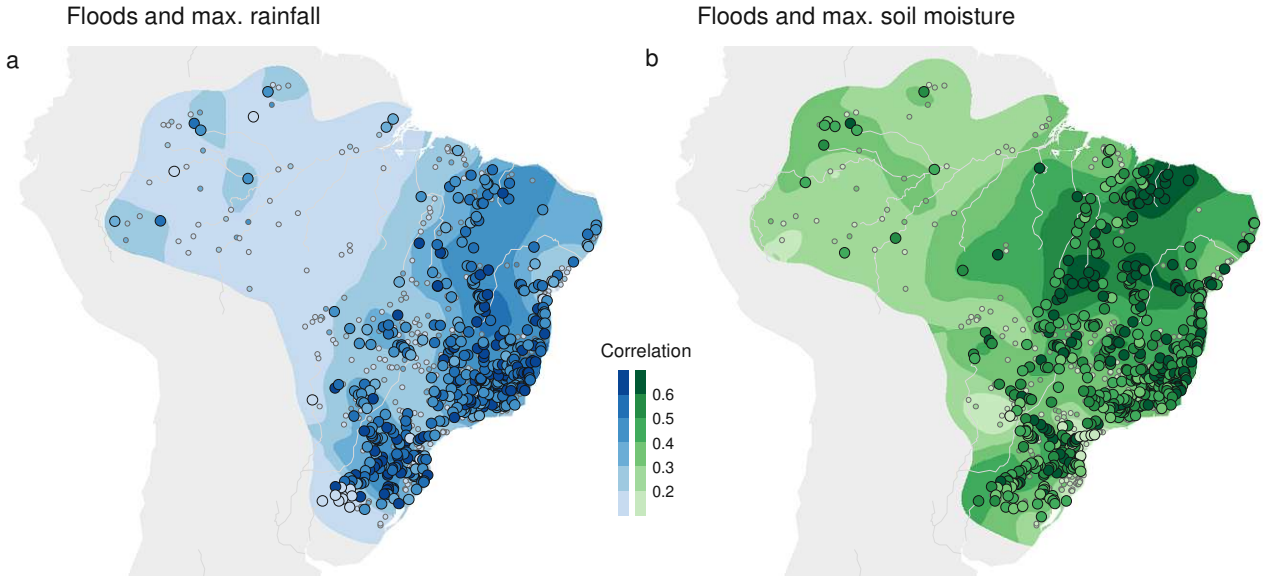

Figure A.2. Circular correlation between the interannual variability of the timing of (a) floods and maximum annual rainfall, (b) floods and maximum annual soil moisture. Large circles indicate basins with significant correlations (α = 0.05) and small circles show basins with non-significant correlations. The background colors are obtained with interpolation using block kriging.

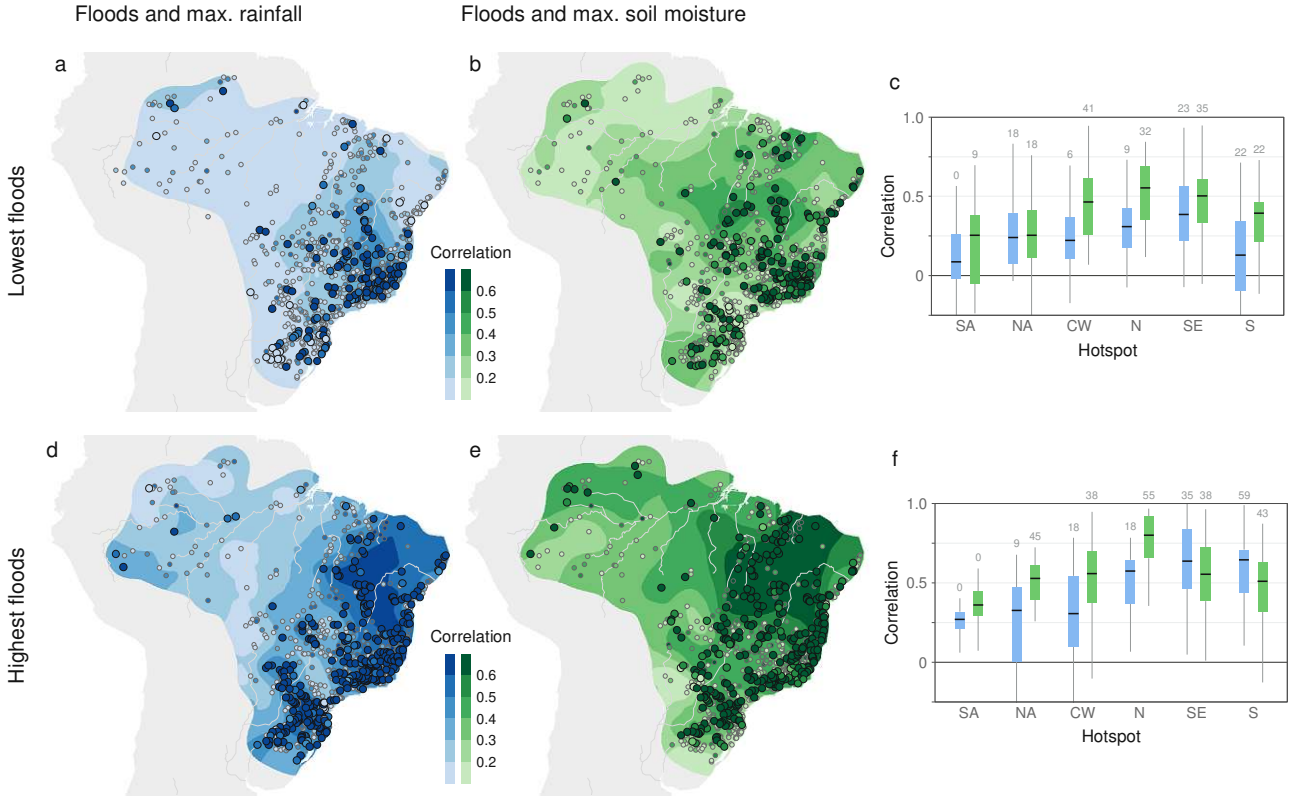

Figure A.3. Circular correlation between the interannual variability of the timing of floods and maximum annual rainfall (blues) or maximum annual soil moisture (greens). (a-c) Correlations including only the lowest 50% of the floods. (d-f) Correlations including only the highest 50% of the floods. Large circles indicate basins with significant correlations (α = 0.05) and small circles show basins with non-significant correlations. The background colors in (a, b, d, e) are obtained with interpolation using block kriging. The numbers above the boxplots indicate the percentage of basins with significant correlations (α = 0.05). The hotspots are Southern Amazonia (SA), Northern Amazonia (NA), Central-West (CW), North (N), Southeast (SE), and South (S).

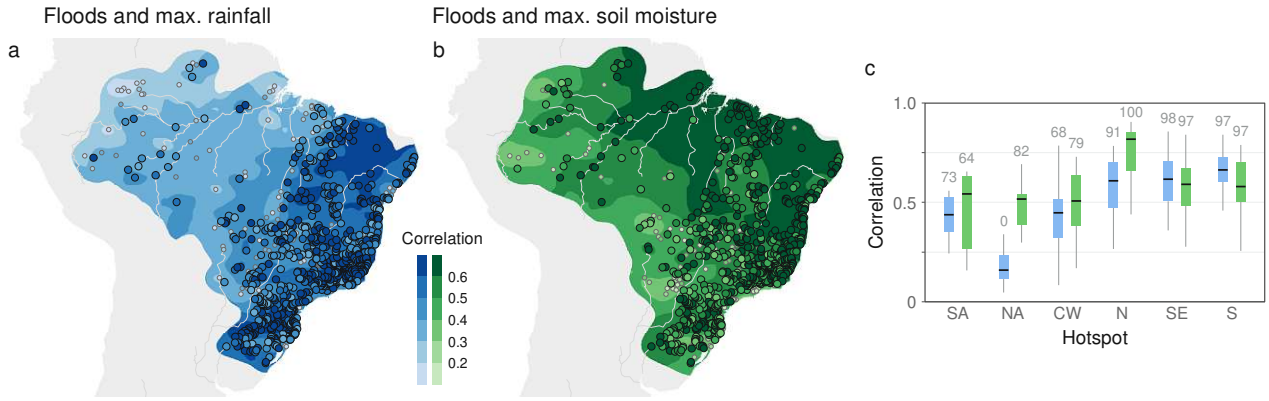

Figure A.4. Spearman correlation between the interannual variability of the magnitudes of (a) floods and maximum annual rainfall, (b) floods and maximum annual soil moisture. Large circles indicate basins with significant correlations ($\alpha = 0.05$) and small circles show basins with non-significant correlations. The background colors in (a) and (b) are obtained with interpolation using block kriging. (c) Spatial variability of the correlations with rainfall (blue boxes) and soil moisture (green boxes) over each hotspot. The numbers above the boxplots indicate the percentage of basins with significant correlations ($\alpha = 0.05$). The hotspots are Southern Amazonia (SA), Northern Amazonia (NA), Central-West (CW), North (N), Southeast (SE), and South (S).

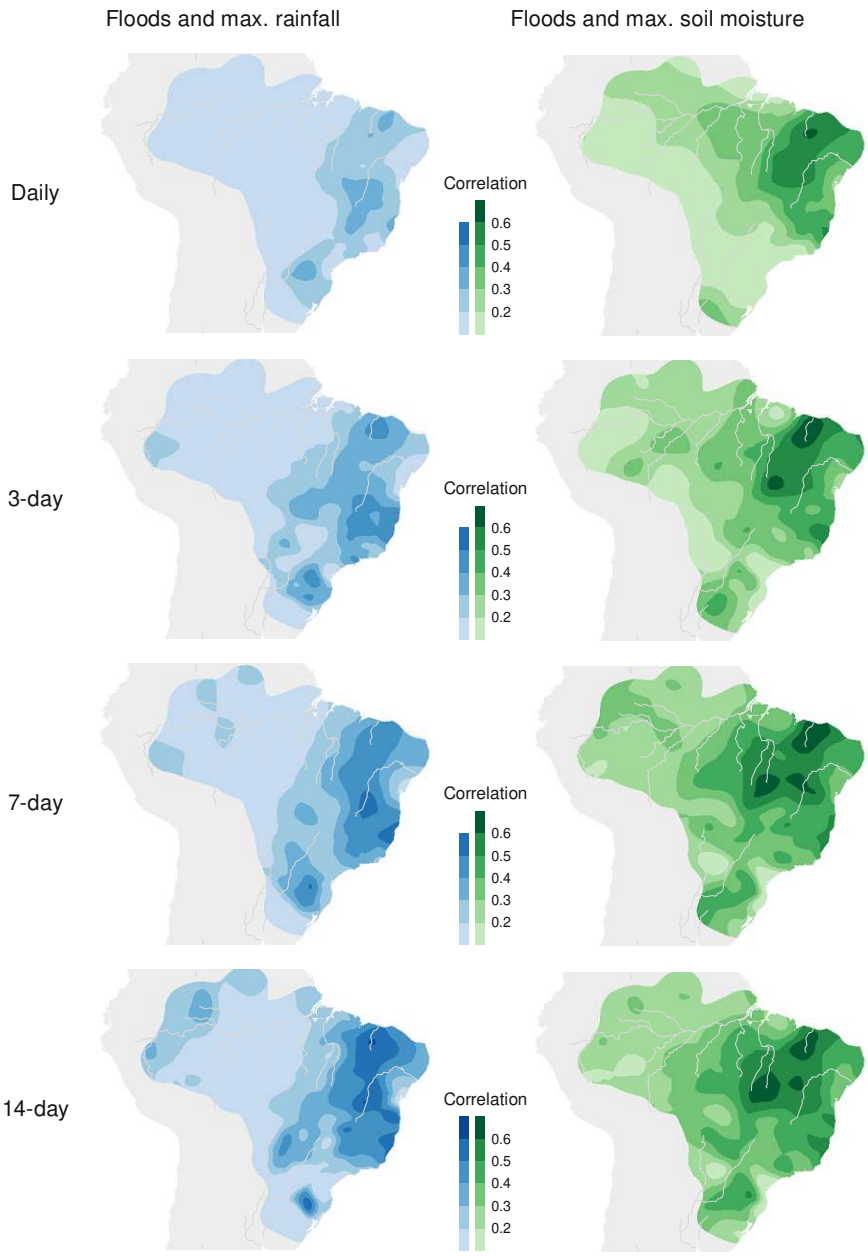

Figure A.5. Circular correlations between the interannual variability of the timing of (left column) floods and maximum annual rainfall, (right column) floods and maximum annual soil moisture. The rows indicate the length of the moving average applied to the time series of daily rainfall and soil moisture before computing the annual maxima. The correlations are obtained with interpolation using block kriging.

Appendix B

Supporting Information of Chapter 5 – Climate and land management accelerate the South American water cycle

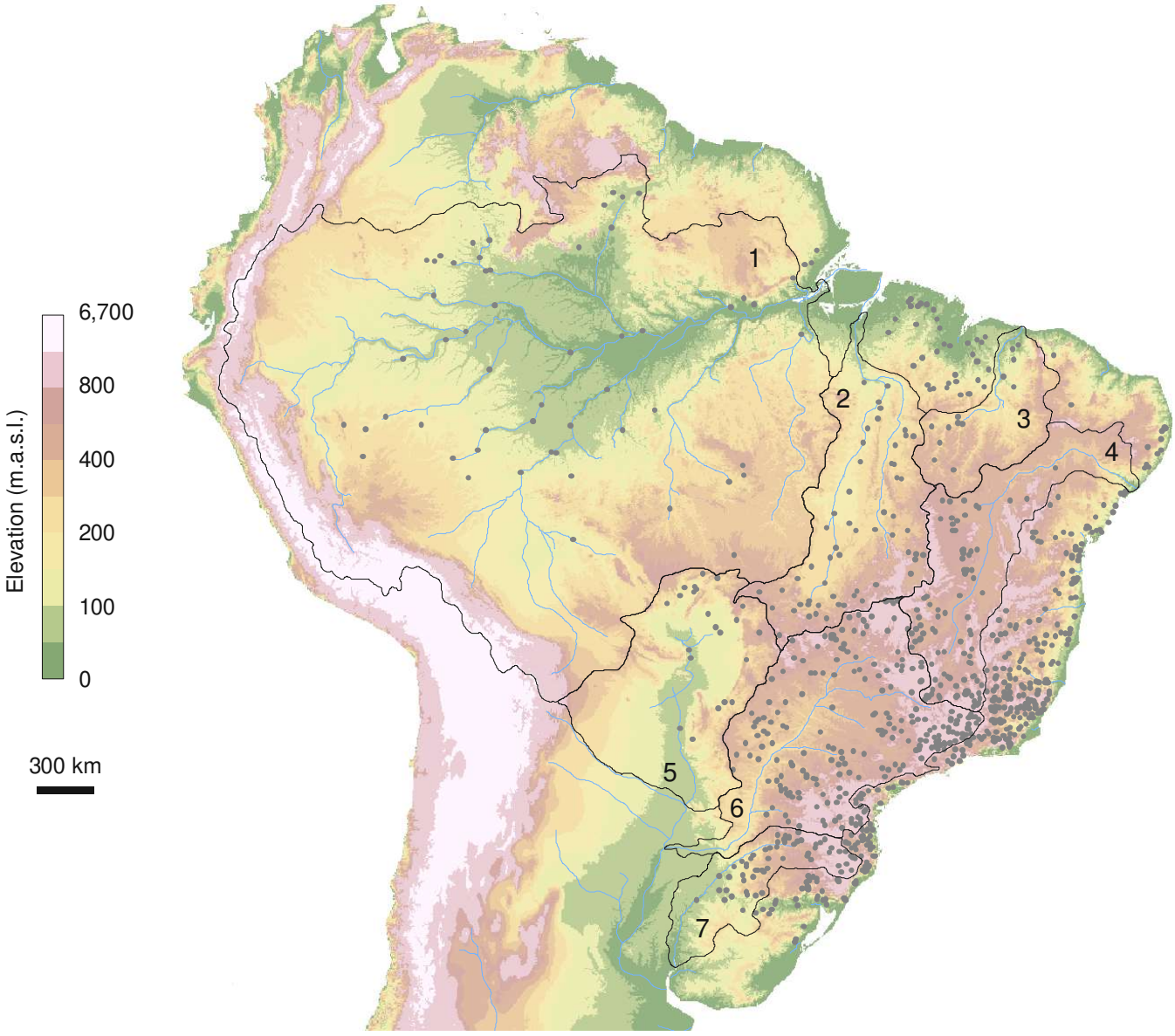

Figure B.1. Hydrometric stations and major Brazilian river basins. Points indicate the locations of the hydrometric stations analyzed. Black lines represent the watersheds of major Brazilian basins: 1, Amazon; 2, Tocantins-Araguaia; 3, Parnaíba; 4, São Francisco; 5, Paraguay; 6, Paraná; 7, Uruguay.

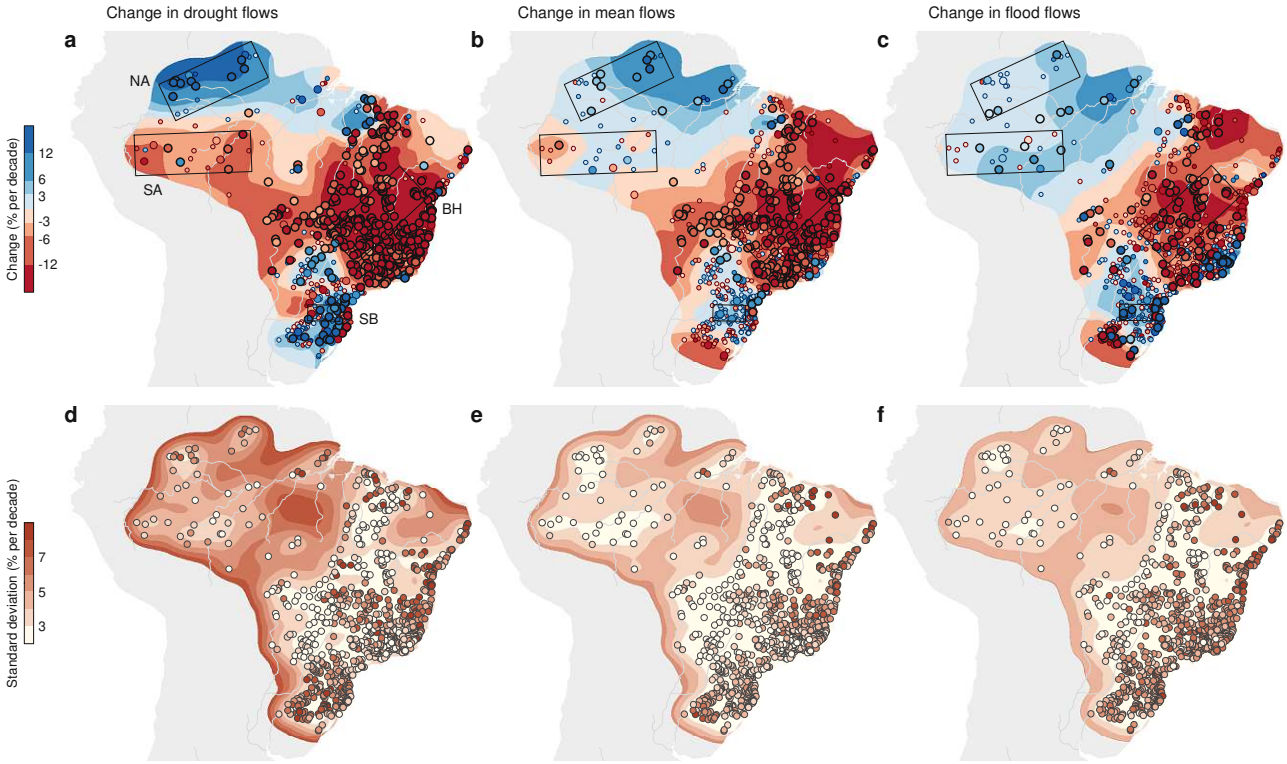

Figure B.2. Streamflow trends in Brazil (1980-2015). (a) Change in annual minimum 7-day streamflow (drought flows). (b) Change in mean daily streamflow. (c) Change in annual maximum daily streamflow (flood flows). The background pattern displays regional trends and the points display local trends (n = 886). Large circles with thick borders in (a-c) show significant trends for a significance level $\alpha = 0.10$, large circles with thin borders show significant trends for a significance level $\alpha = 0.05$, and small circles show non-significant trends. (d-f) Uncertainties of local trends (points) and regional trends obtained by kriging (background). Black rectangles indicate four hotspots of change: Northern Amazonia (NA); Southern Amazonia (SA); Brazilian Highlands (BH); and Southern Brazil (SB).

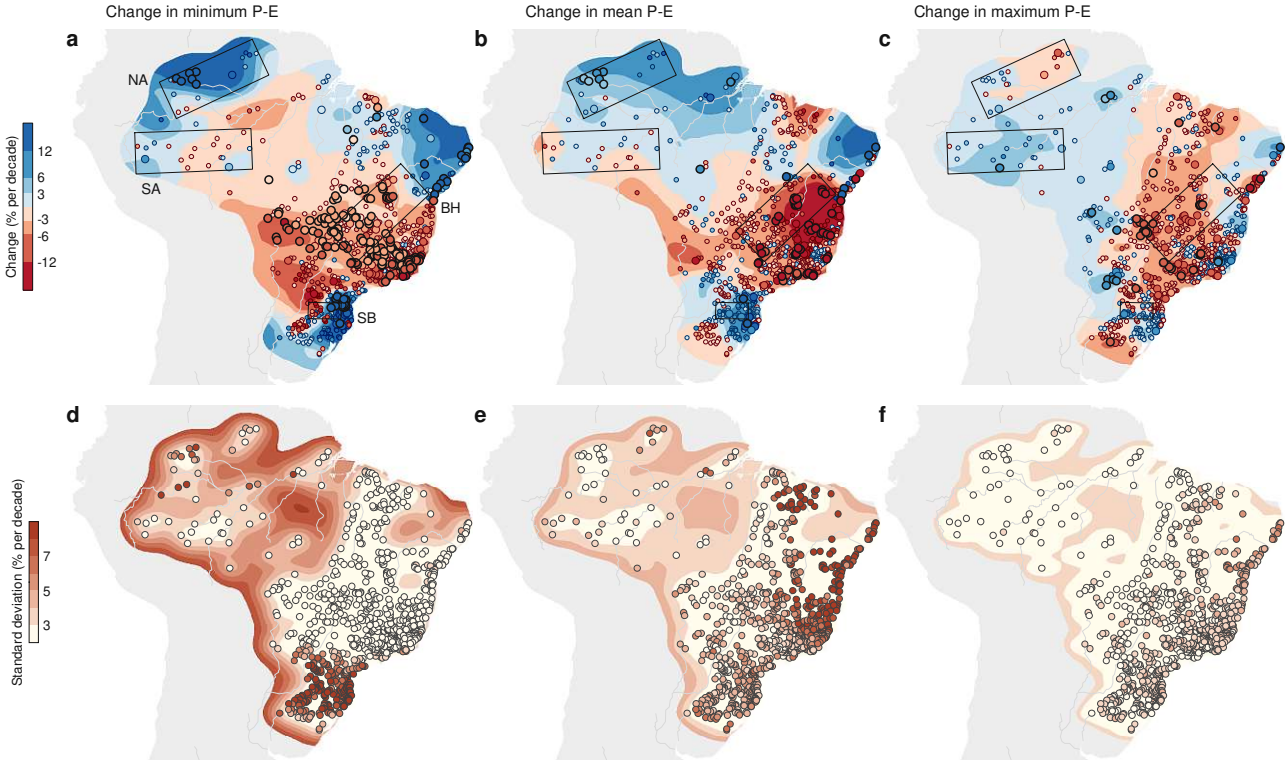

Figure B.3. Trends of climate drivers of streamflow changes in Brazil (1980-2015). (a) Change in annual minimum 90-day precipitation minus evaporation (P – E). (b) Change in mean daily P – E. (c) Change in annual maximum 14-day P – E. The background pattern displays regional trends (interpolated from local trends) and the points display local trends (which includes the contributing basin area). Large circles with thick borders in (a-c) show significant trends for a significance level $\alpha = 0.10$, large circles with thin borders show significant trends for a significance level $\alpha = 0.05$, and small circles show non-significant trends. (d-f) Uncertainties of local trends (points) and regional trends obtained by kriging (background). Black rectangles indicate four hotspots of change: Northern Amazonia (NA); Southern Amazonia (SA); Brazilian Highlands (BH); and Southern Brazil (SB).

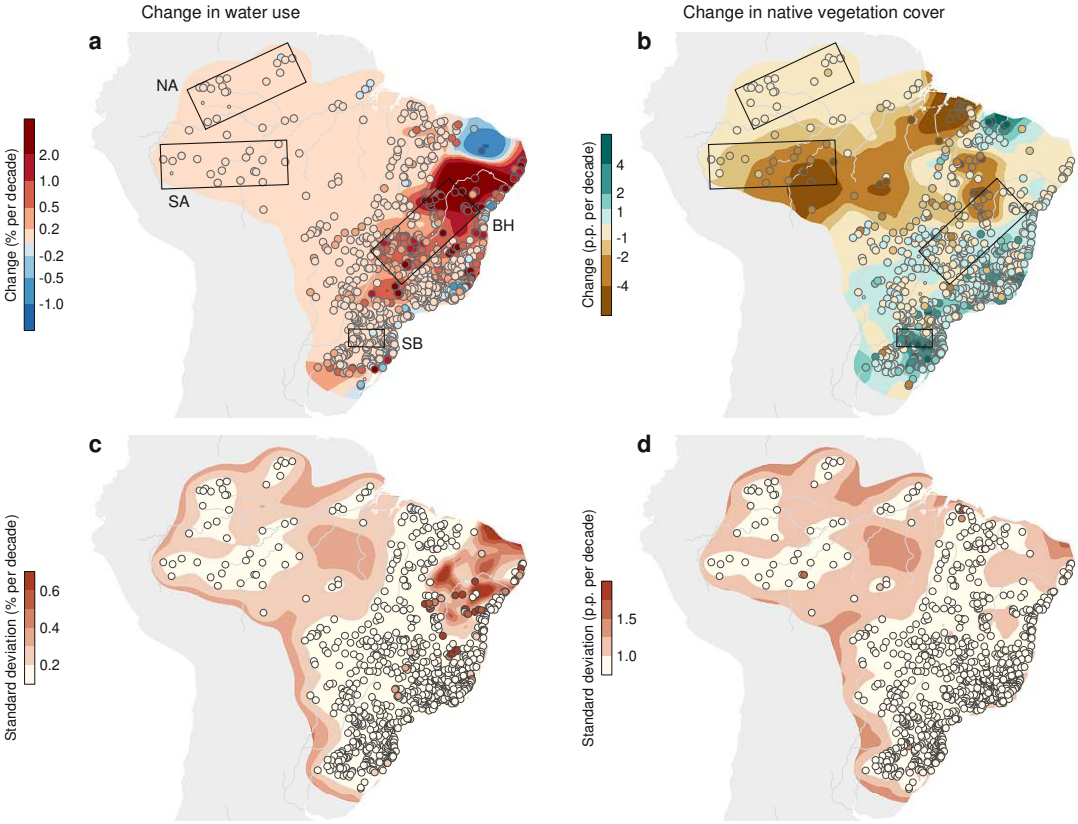

Figure B.4. Trends of land management drivers (water use and native vegetation cover) of streamflow changes in Brazil (1980-2015). (a) Change in water use in % of the long-term mean daily flow per decade. (b) Change in native vegetation cover in percentage points (p.p.) of basin area per decade. The background pattern displays regional trends (interpolated from local trends) and the points display local trends (which includes the contributing basin area). Large circles with thick borders in (a-b) show significant trends for a significance level $\alpha = 0.10$, large circles with thin borders show significant trends for a significance level $\alpha = 0.05$, and small circles show non-significant trends. (c-d) Uncertainties of local trends (points) and regional trends obtained by kriging (background). Black rectangles indicate four hotspots of change: Northern Amazonia (NA); Southern Amazonia (SA); Brazilian Highlands (BH); and Southern Brazil (SB).

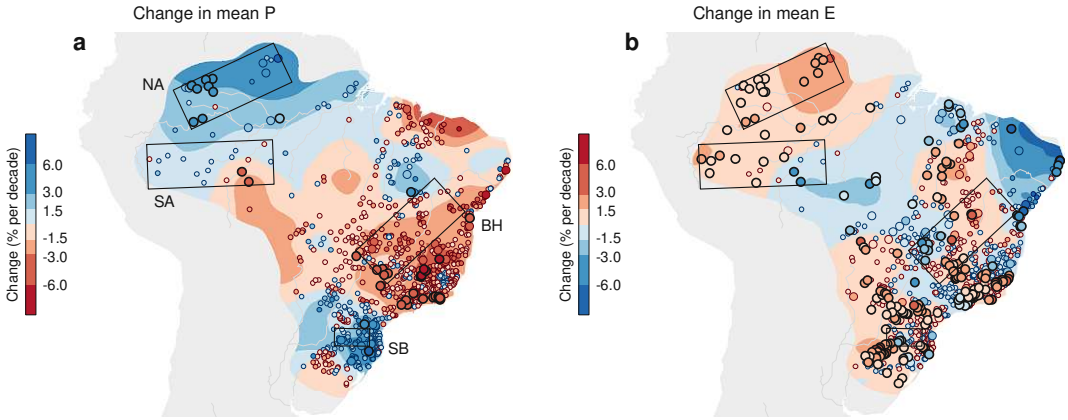

Figure B.5. Precipitation and evaporation changes in Brazil (1980-2015). (a) Change in mean daily precipitation (P). (b) Change in mean daily evaporation (E). The background pattern displays regional trends (interpolated from local trends) and the points display local trends (which includes the contributing basin area). Large circles with thick borders in show significant trends for a significance level $\alpha = 0.10$, large circles with thin borders show significant trends for a significance level $\alpha = 0.05$, and small circles show non-significant trends. Black rectangles indicate four hotspots of change: Northern Amazonia (NA); Southern Amazonia (SA); Brazilian Highlands (BH); and Southern Brazil (SB).

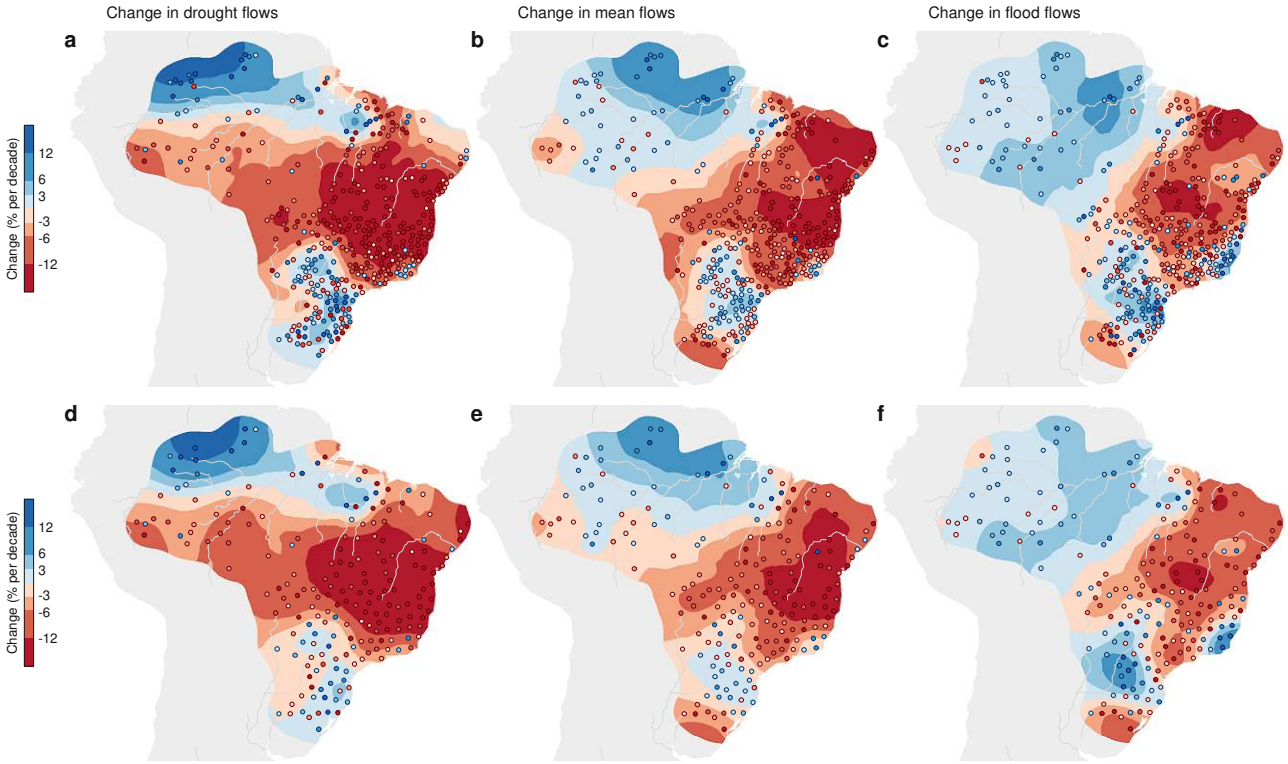

Figure B.6. Streamflow trends as in Fig. B.1 and Fig. B.2 but with fewer stations. (a-c) Change in drought flows, mean flows, and flood flows using only stations with distances of at least 0.5° between each other ($n = 388$). (d-f) Change in drought flows, mean flows, and flood flows using only stations with distances of at least 1° between each other ($n = 200$).

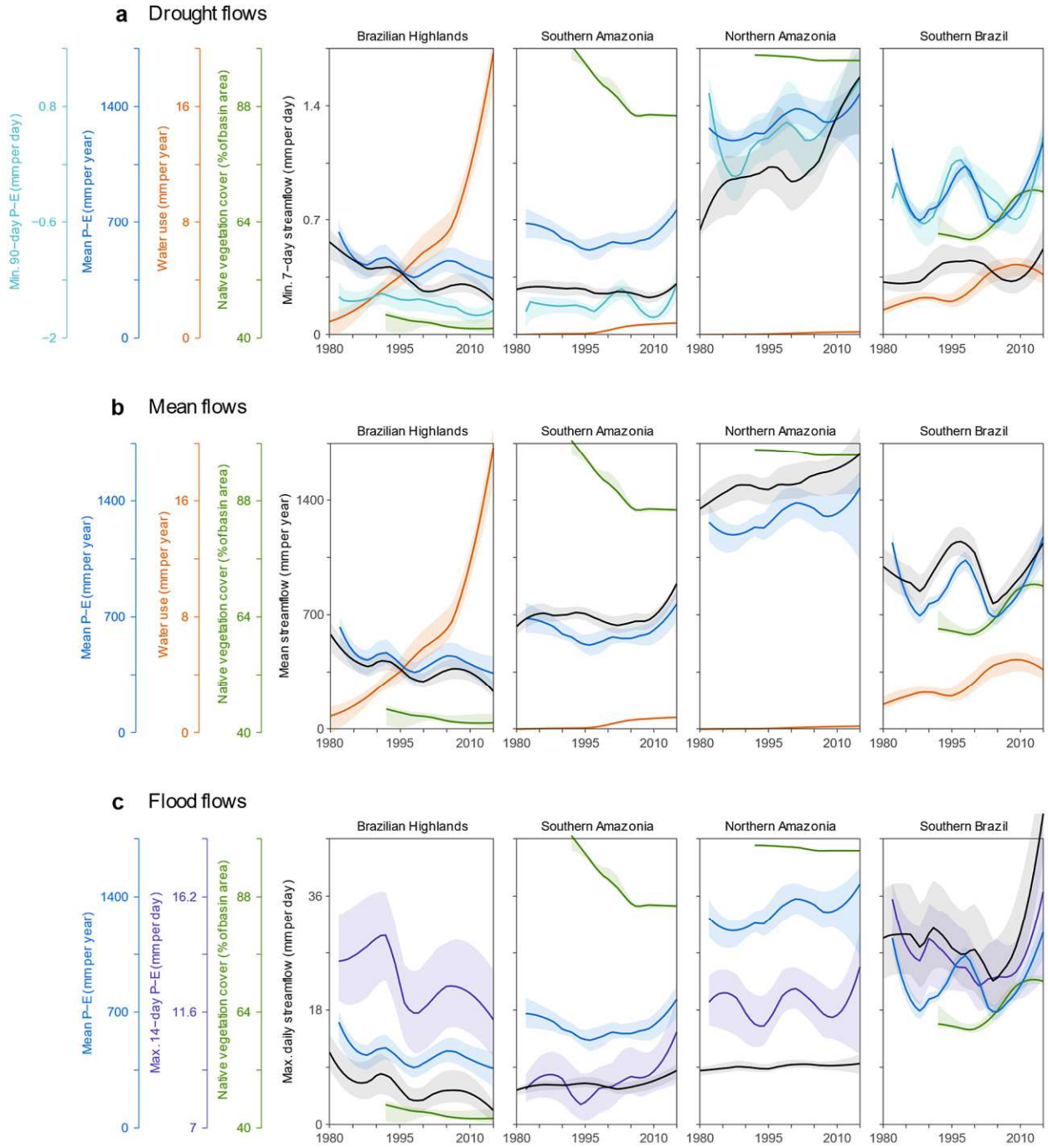

Figure B.7. Temporal evolution of streamflow and their drivers in four hotspots of change in Brazil. (a) Minimum 7-day streamflow (drought flows). (b) Mean daily streamflow (mean flows). (c) Annual maximum daily streamflow (flood flows). Black lines indicate drought flows in the upper panels, mean flows in the middle panels, and flood flows in the bottom panels. Light blue lines indicate minimum 90-day precipitation minus evaporation (P – E). Medium blue lines indicate mean P – E. Purple lines indicate maximum 14-day P – E. Orange lines indicate water use. Green lines indicate native vegetation cover. Lines represent the median and bands indicate the spatial variability within each hotspot (25th and 75th percentiles). The time series were smoothed with LOESS. The boxes represent the hotspots: Brazilian Highlands (n = 108); Southern Amazonia (n = 18); Northern Amazonia (n = 19); and Southern Brazil (n = 32).

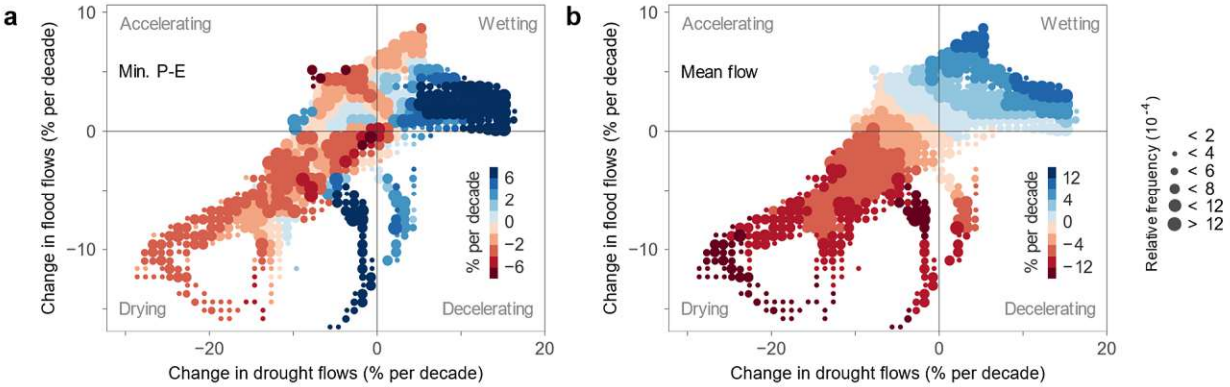

Figure B.8. Trend of water availability and mean flows mapped on the quadrants of the accelerating, decelerating, wetting, and drying streamflow trends. (a) Annual minimum 90-day precipitation minus evaporation (P – E). (b) Mean daily streamflow. Larger circles indicate higher areal fractions. Colors indicate the average trends of the drivers for each bin.

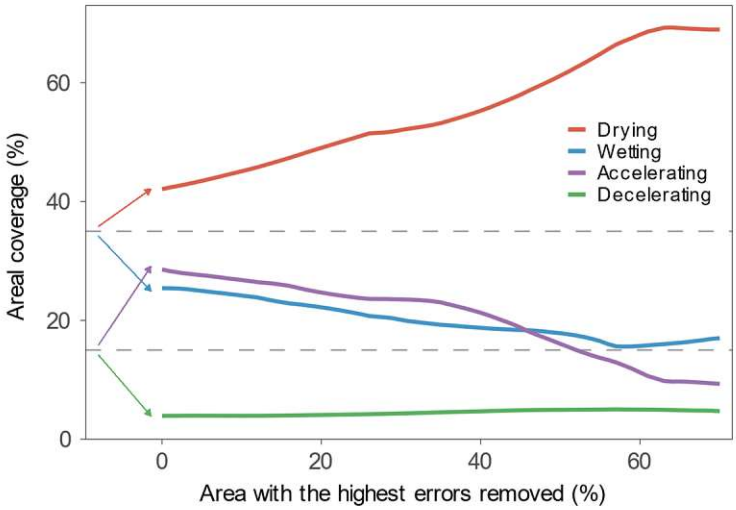

Figure B.9. Areal coverage of each quadrant as areas with the highest errors in the regional trend estimates (kriging errors) are removed. The dashed lines represent 15% (the expected area for the accelerating and decelerating quadrants for a bivariate distribution with $\rho = 0.61$ according to Equation (5.5) and 35% (the expected area for the wetting and drying trends). The arrows indicate differences from the expected values and the quadrant areal coverage when all areas are included in the quadrant classification.

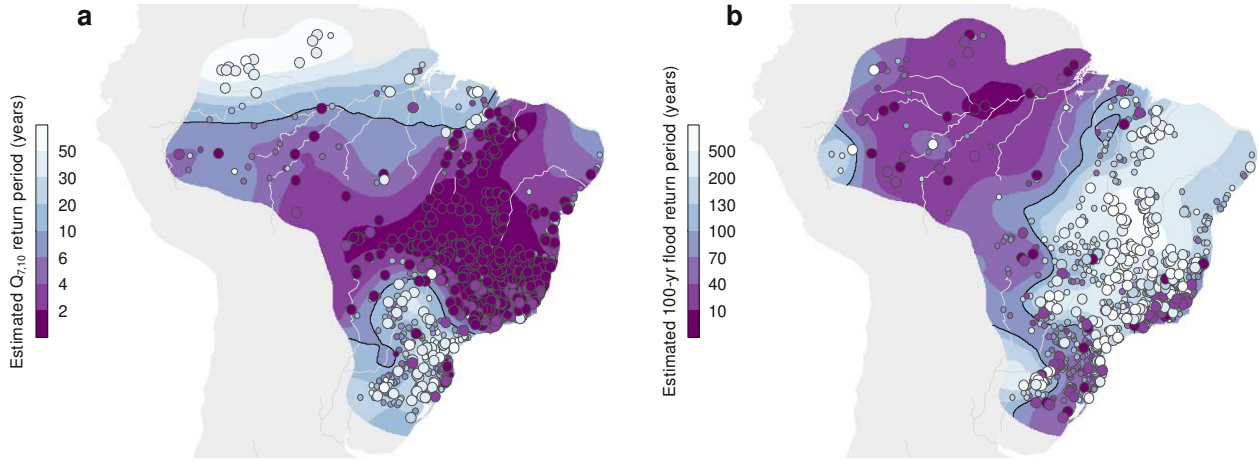

Figure B.10. Estimated return period in 2015 for the 1980 drought and flood flows. (a) Return period in 2015 for the 1980 10-year annual minimum 7-day flow ($Q_{7,10}$, drought flows). (b) Return period in 2015 for the 1980 100-year annual maximum daily streamflow (flood flows). Dark colors indicate lower return periods, representing increasing drought and flood risk. Light colors indicate higher return periods, representing decreasing risk. Points show local return periods (n = 886). Larger points indicate agreement of the 5th and the 95th percentiles of the uncertainty distribution in the sign of change. The black lines separate regions of increasing and decreasing risks.

Table B.1. Comparison of precipitation trends among datasets. (a) Spearman correlation coefficients between trends of the weather station series from the ANA dataset and trends in CHIRPS, MSWEP, PERSIANN and CPC datasets for: mean daily precipitation (Mean P); 14-day maximum precipitation (Maximum P); and 90-day minimum precipitation (Minimum P). (b) The median trend in each dataset for the same precipitation variables as in (a), in % per decade.

(a)

Variable	CHIRPS	MSWEP	PERSIANN	CPC
Mean P	0.62	0.67	0.29	0.08
Maximum P	0.61	0.71	0.22	0.61
Minimum P	0.49	0.48	-0.15	0.30

(b)

Variable	CHIRPS	MSWEP	PERSIANN	CPC	ANA
Mean P	-0.09	1.92	2.61	-5.41	-1.26
Maximum P	-0.13	1.95	2.55	-3.32	-0.51
Minimum P	-4.44	-2.78	-7.06	-5.59	-2.32

Appendix C

Supporting Information of Chapter 6 – Drought-rich periods were twice as frequent as flood-rich periods in Brazil from 1940 to 2020

Table C.1. Probabilities of occurrence $P(k \mid w,n,a)$ of k or more events within a window w sized from 5 to 15 years, in a time series of length n = 79 years (the median length for river gauges) and a = 8 total events over the 79 years (that is, events above a 10-year return period threshold).

k (events)	w = 5	w = 6	w = 7	w = 8	w = 9	w = 10	w = 11	w = 12	w = 13	w = 14	w = 15
1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
2	0.976	0.993	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
3	0.243	0.359	0.475	0.584	0.681	0.765	0.833	0.887	0.927	0.956	0.975
4	0.012	0.029	0.053	0.085	0.125	0.172	0.225	0.283	0.344	0.408	0.472
5	$1.8\ 10^{-04}$	$8.4\ 10^{-04}$	$2.4\ 10^{-03}$	0.005	0.010	0.016	0.026	0.038	0.053	0.072	0.094
6		$7.3\ 10^{-06}$	$4.2\ 10^{-05}$	$1.4\ 10^{-04}$	$3.6\ 10^{-04}$	$7.6\ 10^{-04}$	$1.5\ 10^{-03}$	$2.5\ 10^{-03}$	$4.2\ 10^{-03}$	$6.4\ 10^{-03}$	$9.5\ 10^{-03}$
7			$2.0\ 10^{-07}$	$1.4\ 10^{-06}$	$5.3\ 10^{-06}$	$1.5\ 10^{-05}$	$3.7\ 10^{-05}$	$8.0\ 10^{-05}$	$1.5\ 10^{-04}$	$2.8\ 10^{-04}$	$4.7\ 10^{-04}$
8				$2.8\ 10^{-09}$	$2.2\ 10^{-08}$	$9.7\ 10^{-08}$	$3.2\ 10^{-07}$	$8.7\ 10^{-07}$	$2.1\ 10^{-06}$	$4.4\ 10^{-06}$	$8.7\ 10^{-06}$