Assessing the impact of deforestation on observed runoff and an analysis on simulated runoff by WFLOW in the Amazon

Rogier de Wildt

June, 2022

Assessing the impact of deforestation on observed runoff and an analysis on simulated runoff by WFLOW in the Amazon

by R.S. de Wildt

to obtain the degree of Master of Science

at the Delft University of Technology,

to be defended publicly on Thursday June 15, 2022 at 15:00 AM.

Student number:	4154878	
Date:	June 8, 2022	
Thesis committee:	Dr. ir. R. Hut,	TU Delft, supervisor
	Dr. M. Hrachowitz,	TU Delft, chair
	Dr. S. R. de Roode,	TU Delft
	R. Hulsman,	Royal Haskoning DHV

Abstract

Recent reports show that deforestation in Latin-America has been severe over the last decades. Especially Brazil is subject to an alarming rate of forest loss, which will remain a factor in the coming decades. In addition to deforestation, there is an increasing amount of hazards like floods and mudslides. These hazards result in major damage to human life and nature. This research analyzes the relation between deforestation and river discharge for observed data, a simple conceptual model and complex model WFLOW for 30 catchments subject to deforestation in Brazil. These three studies use satellite data provided by Hansen et al. (2020) to determine annual cumulative deforestation per catchment. In the study with observed data the relation between deforestation and data from CAMELS-BR (Chagas et al., 2020) is researched by analyzing annual values of two discharge driven parameters: Runoff coefficient and recession coefficient. The second study consists of a simple self-made conceptual model. In this second study, the relation between recession coefficient α and annual deforestation is analyzed by calibrating the model per year for parameter α. Lastly, a study is conducted on the performance of WFLOW in catchments subject to deforestation. During this study, several input parameters are changed to observe the response of hydrological processes in WFLOW.

The study with observed data shows both increasing and decreasing trends in coefficients for several degrees of deforestation. Literature shows that landcover type after deforestation is a major factor in the interpretation of these results. However, a lack of quality annual landcover data prevents better research in the non-masked impact of deforestation on discharge. The results of the simple model study show no significant relation between deforestation and recession coefficient α. The simplicity of this self-made conceptual model is the weak and strong point of this sub-study. The simplicity makes the results less suitable for analyzing the exact impact of deforestation on discharge, but it is useful for observing general signals and is easily scalable to different catchments. The simulated discharge by WFLOW show a steep overestimation of discharge during peak flow in comparison to observed discharge by CAMELS-BR. Therefore it is not possible to analyze how WFLOW reacts to deforestation. Instead, an in-depth analysis on the cause of this poor performance is conducted by analyzing timeseries for hydrological factors like unsaturated zone depth. The results of this analysis indicate the overestimation of discharge is caused by a lack of outflow from soil layers. In addition, the difference between the Budyko framework of different data sets used in this research, show that uncertainty in quality of input data is a plausible factor on the output of WFLOW.

In conclusion it is observed that deforestation does not necessarily lead to higher runoff coefficients and recession coefficients for measured data in this study area. In addition, the conclusion of the simple model study is that no significant relation between deforestation and recession coefficient α is observed for this conceptual model. Finally, the performance of WFLOW is considered too poor to analyze the impact of deforestation on discharge. The root of this poor performance is considered to be a combination between lack of groundwater modelling and uncertainty on the quality of input-data.

Acknowledgements

This report is the final deliverable of my Master of Civil Engineering at Delft University of Technology. The research was carried out as a graduation internship at Royal Haskoning DHV.

First of all I want to thank my TU Delft committee members Rolf Hut, Markus Hrachowitz, Stephan de Roode for their guidance, discussions and valuable feedback during my research. Furthermore, I want to thank Rineke Hulsman from Royal Haskoning DHV for sharing her extensive knowledge and providing me the opportunity for this graduation internship. I also want to express my gratitude to Thijs van den Pol from Royal Haskoning DHV for his generous availability whenever I needed a sparring partner for Python related problems. Next, I want to thank Jerome Aerts for sharing his expert knowledge on WFLOW. I wish him all the best with the remainder of his PHD research.

Lastly, a special thanks goes to my friends, family and Rozemarijn for their support and much needed distractions during my research and COVID-19.

Rogier de Wildt
Delft, June 2022

List of symbols

α	Storage coefficient [-]
C_r	Runoff coefficient [-]
E_a	Actual evapotranspiration [mm]
E_p	Potential evapotranspiration [mm]
GEE	Google Earth Engine
$infcap$	Infiltration capacity [mm]
k	Recession coefficient [-]
KGE	Kling-Gupta Efficiency [-]
NSE	Nash-Sutcliffe Efficiency [-]
P_t	Daily precipitation [mm]
Q_t	Daily discharge $[m^3/s]$
TAS	Near-surface temperature [°C]

List of Figures

1 Annual forest area net change, by decade and region, 1990–2020 (Food & Organization, 2020) . . . 2
2 Data flow complete methodology . . . 8
3 Deforestation for 2000-2020. . . . 10
4 a) Concentration of deforestation over 20 years. The severity of deforestation is indicated by the darkness of the red pixels. b) Figure displaying the daily discharge (left ax) and cumulative deforestation (right ax) in time. c) Table showing how percentage of deforestation is calculated. . . . 11
5 Overview of the selected catchments . . . 12
6 Flowchart for selection of catchments . . . 13
7 Data flow: observed data study . . . 15
8 Data flow diagram for determining the correlation between runoff coefficient and deforestation . . . 16
9 Typical hydrograph showing recession of individual components of runoff (Jain, 2014) . . . 17
10 Time-series of natural logarithm of discharge with detection of discharge peaks (de Wildt, 2022) . . . 18
11 Data flow diagram for determining the correlation between recession coefficient and deforestation . . . 19
12 Data flow diagram for determining the impact of deforestation on a simple conceptual model . . . 19
13 Basic conceptual model with one storage bucket . . . 20
14 Data flow diagram for determining the impact of deforestation on a complex model . . . 22
15 Steps in creating the WFLOW parameterset Deltares (2021a) . . . 23
16 Overview of the different processes and fluxes in the WFLOW_sbm model (Deltares, 2021b) . . . 24
17 Complete WFLOW scheme (Deltares, 2021b) . . . 27
18 Relation between actual evapotranspiration (AET) and the cumulative percentage of deforestation per year with a individual plot for each catchment. Orange markers (green trendline) is obtained from GLEAM and blue markers (red trendline) is obtained from MGB. . . . 30
19 Budyko framework observed-data study . . . 31
20 Relation between the runoff coefficient C_r and the cumulative percentage of deforestation per year for all catchments. Each color represents a catchment. 32
21 Relation between the runoff coefficient C_r and the cumulative percentage of deforestation per year with a individual plot for each catchment . . . 33
22 Relation between the recession coefficient k and the cumulative percentage of deforestation per year for all catchments. Each color represents a catchment. . . . 34
23 Relation between the recession coefficient k and the cumulative percentage of deforestation for all catchments. . . . 35
24 Results "Trial and error" calibration for catchment "Bom Jardim" . . . 36

25 Discharge for optimized parameter α. NSE "Bom Jardim" = 0.79 and NSE "Cuiaba Paraguay" = 0.56. 37
26 Relation between yearly α and cumulative deforestation 38
27 Timeseries of observed and callibrated simulated discharge 39
28 Timeseries of discharge with different values of KsatHorFrac 40
29 Time series water balance 41
30 Budyko framework for CAMELS-BR and ERA5. The input data for CAMELS-BR (GLEAM) is orange. The input data for WFLOW (ERA5) is green. . . 42
31 Budyko framework for catchment "Bom Jardim" (top) and "Cuiaba Paraguay" (bottom). Input data for CAMELS-BR (GLEAM) is orange. Input data for WFLOW (ERA5) is green. 43
32 Timeseries of precipitation 44
33 Bar chart of cumulative yearly precipitation 45
34 Timeseries of mean daily temperature 46
35 Timeseries potential evapotranspiration 47
36 Timeseries standard deviation of daily potential evapotranspiration 47
37 Timeseries potential evapotranspiration and temperature from CAMELS-BR and input WFLOW 48
38 Timeseries canopy storage 49
39 Timeseries depth unsaturated zone and simulated overland flow 50
40 Timeseries depth unsaturated zone and precipitation 51
41 Timeseries depth unsaturated zone, infiltration and excess infiltration . . . 52
42 Timeseries depth saturated zone and discharge 53
43 Timeseries observed and simulated discharge with several values of parameter infiltration capacity 54
44 Timeseries observed and simulated discharge with several values of parameter soil depth 55
45 Relation between the runoff coefficient C_r and the percentage of deforestation per year for all catchments. Each color represents a catchment. 72
46 Relation between the runoff ratio C_r and the percentage of deforestation per year with a individual plot for each catchment 73
47 Relation between the recession coefficient k and the percentage of deforestation per year for all catchments. Each color represents a catchment. 74
48 Relation between the runoff ratio C_r and the percentage of deforestation per year for all catchments. 75
49 Timeseries observed discharge and simulated discharge for several values of infiltration capacity. The catchment is Bom Jardim. 77
50 Timeseries observed discharge and simulated discharge for several values of infiltration capacity. The catchment is Cuiaba Paraguay. 78
51 Timeseries observed discharge and simulated discharge for several values of soil depth. The catchment is Bom Jardim. 79
52 Timeseries observed discharge and simulated discharge for several values of soil depth. The catchment is Cuiaba Paraguay. 80
53 Overview of input and output for WFLOW this study. The defintion of the parameters for input and output are shown in Table 6 and Table 7. 81

List of Tables

1 Input sources of the observed data study 9
2 Input source of the model studies 9
3 Key characteristics of catchments in study area 14
4 Results of the calibration by objective function Kling-Gupta Efficiency (KGE) on parameter "KsatHorFrac". The table shows the value for KGE per value of "KsatHorfrac" for each catchment. The optimal KGE of a catchment is shown in green. 76
5 Results of the calibration by objective function Nash-Sutcliffe Efficiency (NSE) on parameter "KsatHorFrac". The table shows the value for NSE per value of "KsatHorfrac" for each catchment. The optimal NSE of a catchment is shown in green. 76
6 List containing all parameters in parameterset 82
7 List containing all output parameters 82
8 Description of each parameter changed in the model process. 83
9 Model logbook 84

Contents

List of Symbols **iii**

1 Introduction **1**

2 Problem Definition **6**

3 Methodology **7**
3.1 Pre-processing phase . . . 8
3.1.1 Data . . . 8
3.1.2 Study area . . . 11
3.2 Study with observed data . . . 14
3.2.1 Actual evapotranspiration . . . 15
3.2.2 Runoff Coefficient . . . 16
3.2.3 Recession coefficient . . . 16
3.3 Study with simple model . . . 19
3.4 Study with complex model . . . 21
3.4.1 Processes . . . 23
3.4.2 The model . . . 25

4 Results **29**
4.1 Study with observed data . . . 29
4.2 Study with simple model . . . 35
4.3 Study with complex model . . . 38

5 Discussion **56**
5.1 Study with observed data . . . 56
5.2 Study with simple model . . . 59
5.3 Study with complex model . . . 60

6 Conclusion **63**
6.1 Recapitulation of the research . . . 63
6.2 Answers to the sub-questions . . . 63
6.3 Main conclusion . . . 65
6.4 Limitations and recommendations . . . 65

Bibliography **71**

Appendices **72**

A Additional Results: Observed data study **72**
A.1 Runoff coefficient . . . 72
A.2 Recession coefficient . . . 74

B Calibration **76**

C Additional results: Complex model study **77**

C.1 Infiltration capacity 77
C.2 Soil depth 79

D Model parameters **81**

E Model logbook **83**

1

Introduction

November 2th 2021, the first major deal of the global climate conference is signed in Glasgow: Over a 100 world leaders have promised to end and reverse deforestation by 2030 (Rannard & Gillet, 2021). This statement is the result of numerous reports on the alarming rate of deforestation across the globe. Brazil is among the countries with the largest amount of deforestation in the last decades. However, deforestation is not the only disaster in Brazil that keeps occurring on a more frequent basis. The New York times reported the following headline on 28 December " Record floods stun Brazil's northeast, killing at least 20" (Milhorance, 2021). According to the newspapers, parts of Brazil are facing the worst floods in a century. But what causes this increase in disasters like floods? Researchers have been analyzing this for years. Often climate change is pointed at as the most obvious source. However, a link between deforestation and floods should not be disregarded.

This thesis will focus on the impact of deforestation in Brazil on river runoff. In particular the ability of the hydrologic model WFLOW to determine runoff in catchments subject to deforestation. First, an in-depth literature review is conducted to get and provide an overview of the existing knowledge on this subject. Second, a study is performed whereby measured data is used to analyze the relation between runoff and deforestation in the chosen study area. Third, a study is conducted on the impact of deforestation on a basic conceptual model. Lastly, river discharge (also called 'runoff') is simulated with the complex hydrological model WFLOW. The results of this complex model study are used to determine how well WFLOW is able to calculate runoff for deforested catchments in Brazil. In the next sub-chapter 'Background theory' results of the in-depth literature review are presented. This sub-chapter discusses literature related to deforestation and hydrologic modelling.

Background theory

This sub-chapter provides background theory for the research of this thesis. The sub-chapter discusses the relation between forests and the hydrological cycle, the impact of deforestation on discharge and the current representation of deforestation in models. The conclusions on these subjects are later used to interpret the results of the study with observed data and the model studies.

Deforestation

The literature review conducted by (Zhang, Yu, Zhao, & Qin, 2003) concludes that forests play an important role in the hydrological cycle. The temporal storage of water is one of the forests main abilities. This ability to store water occurs, among other things, through canopy interception, stem/root storage , shallow subsurface storage and deep surface storage (Breda, Huc, Granier, & Dreyer, 2006) (Farrick & Branfireun, 2014). In modelling these water storages are often called interception reservoir, the unsaturated zone, slow and fast responding reservoirs (Lindstrom, Johansson, Persson, & M. Gardelin,

1997). Different stores of water in models are connected by descriptions of hydrological processes such as infiltration or root water uptake (Lindstrom et al., 1997). Given enough knowledge of all relevant hydrological processes, river discharge can be predicted for a given atmospheric input (Beven, 1993). Jones et al. (2009) concluded that forest hydrology, the study of how water flows through forests, can aid in researching connections between water and forests. However, it must advance if it is to deal with current complex issues, including climate change, wildfires, changing patterns of development and ownership, and changing societal values (Jones et al., 2009).

Every five years the UN Food and Agriculture Organization (FAO) publishes a report on their assessment of global forest resources. The latest edition, 2020, states that the total amount of global forests in 2020 is 4.06 billion hectares, which is equivalent to 31% of total land area (Food & Organization, 2020). Brazil has the second largest amount of forest in the world with an area of 497 million hectares. Since 1990, the world has lost an estimated amount of 420 million hectares of forest through deforestation. Figure 1 shows the annual forest area net change by decade and region since 1990. It can be observed that the rate of deforestation in South America has been declining. However, the amount of deforestation in South America and especially Brazil remains an alarming issue.

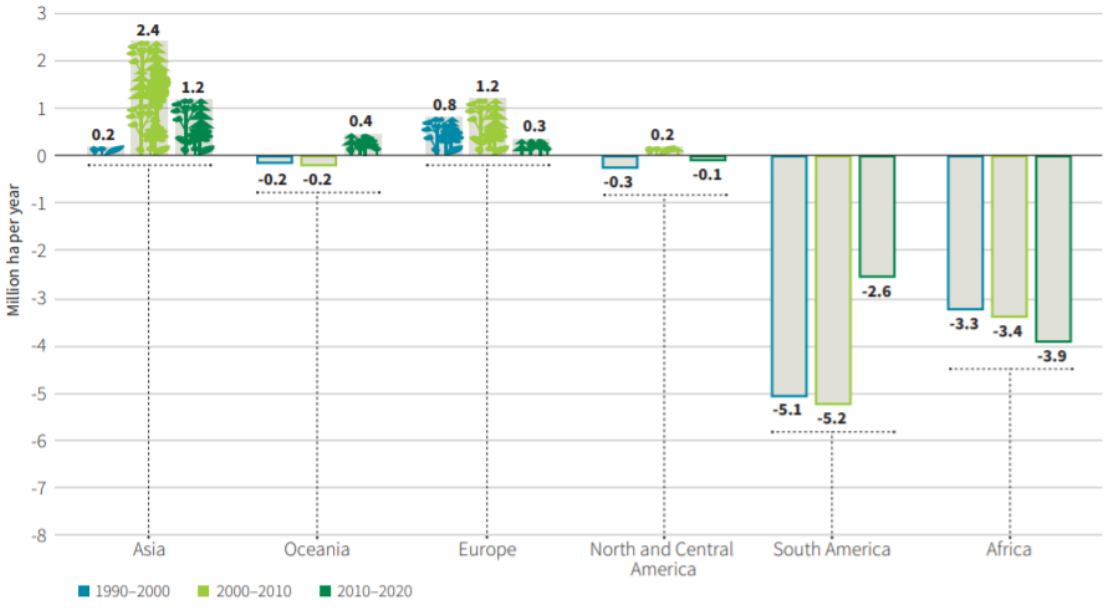

Figure 1: Annual forest area net change, by decade and region, 1990–2020 (Food & Organization, 2020)

The literature review conducted by A. Brown, Zhang, McMahon, Western, and Vertessy (2005a) shows that deforestation of catchments will have an impact on the hydrological processes in that catchment, which will be discussed in the next paragraph. Despite the recognized importance of the forest ecosystem, there is currently only one source with quantification of deforestation. Hansen et al. (2013) used earth observation satellite to map global forest loss (2.3 million square kilometers) and gain (0.8 million square kilometers) from 2000 to 2012 at a spatial resolution of 30 meters. Furthermore, Keenan et al. (2015)

show that the rate of deforestation is accelerating worldwide. Results of Guimberteau et al. (2017) emphasize the need to include the process of deforestation in climate change models. The main conclusion of Guimberteau et al. (2017) is that deforestation has the potential to mask the effects of climate change on surface hydrology. A mask means that the impact of deforestation on river runoff has an opposite effect as climate change. Thereby 'masking' the impact of deforestation on observed discharge.

The above conclusions indicate that deforestation will be an even more pressing topic in the coming years and must be taken into account when modelling rainfall-runoff relations for future scenarios.

Impact deforestation on discharge

There are several conclusions from research papers on the impacts of deforestation on hydrological processes like runoff. Siriwardena, Finlayson, and McMahon (2006) researched the impact of land-use change on catchment hydrology of large catchments. They observed that in the post-clearing period there was a 78% increase in runoff compared with the pre-clearing period. This is supported by (Mohammad & Adamn, 2010), (Zimmerman, Elsenbeer, & Moraes, 2006) and (Hrachowitz et al., 2020) who concluded that deforestation results in an increase of the rainfall-runoff ratio in a catchment. A second conclusion is the observation of Olang and Furst (2011) that deforestation results in a shorter travel time of precipitation. This means that after a precipitation event, the runoff-peak will occur earlier than in a situation without deforestation. In addition, these peaks are higher for catchments with deforestation (Olang & Furst, 2011),(Abdulkareem, Sulaiman, Pradhan, & Jamil, 2018), (Khaleghi, 2017). Fohrer, Haverkamp, Echkhardt, and Frede (2001) concluded that the waterstorage capacity decreases with deforestation, which can be observed by analyzing the runoff. This is supported by Hrachowitz et al. (2020) and Nijzink et al. (2016) who state that deforestation leads to a reduction of the water storage in the unsaturated soil.

The literature review conducted by Bruijnzeel (2004) showed that there are exceptions to these conclusions. The main conclusion of the studies in Bruijnzeel (2004) is that no significant variation in discharge is observed after deforestation. A straightforward explanation could be that this is caused by a period of less rainfall after a period of deforestation in a catchment (J. Fallas, personal communication)(Bruijnzeel, 2004). (S. Brown, 1990) observed that a rapid return to pre-disturbance levels of streamflow during forest regeneration after logging or clearing in the humid tropics may be expected in view of the generally vigorous growth of young tropical secondary vegetation. Additional evidence for a rapid recovery of forest water use after clearing and burning, comes from eastern Amazonia where the evapotranspiration of 2-year-old and 3.5-year-old secondary vegetation is similar to values obtained for mature forest in the same area (Klinge, Schmidt, & Folster, 2001) (Sommer, de Abreu Sa, Vielhauer, A. Carioca de Araujo, & Vlek, 2002). In addition, research shows that conversion from native forests to some invasive flora (Eucalyptus and oil palms), accelerates the return to original values of water yield due to their fast growing and high water consuming nature (Bruijnzeel, 2004). Factors like elevation and slope are also important for the effects of deforestation. No declines in annual streamflow totals have been reported following lowland tropical forest removal (Bruijnzeel, 2004). The scale of a study area is important for the observed effects. (Qian, 1983) was unable to detect any systematic changes in streamflow from catchments ranging in size from 7 to 727

km2 on the island of Hainan, southern China, despite a 30% reduction in tall forest cover over three decades. This is supported by (Dyhr-Nielsen, 1986) and (Wilk, Andersson, & Plermkamon, 2001). The prime cause of forest alteration in these researches was shifting cultivation.

The above paragraphs discussed the impact of deforestation on the amount of discharge. The general consensus is that deforestation results in an increase of discharge on both short and long timescales. However, some literature shows there are also examples that contradict this statement. In conclusion there does not seem to be an established global relation between deforestation and changes on downstream river discharge. Rather, the impact (if any) differs from one area to the next. Therefore it is important to look at the local situation and context when making claims on the impact of deforestation on river discharge

Models

Models are collections of hypotheses describing our understanding of water systems. This understanding is then translated to equations through physics and math. Finally, this set of equations is transformed into a computer model by using programming tools like Python. Rainfall-runoff models should be developed so that they represent the current consensus on system behaviour, whether this is informed by direct interpretation of published information with or without field and laboratory data, or whether further understanding has been extracted from these data by mathematical modelling (Brassington & Younger, 2010).

Rainfall-Runoff models are used among other things to predict flooding and therefore aid in minimizing the damages by these floods (Thielen, Bartholmes, Ramos, & Roo, 2009). In order to accurately predict floods in the future, it is important to know if land cover changes like deforestation affect the output of a model. A relatively new hydrological software model is WFLOW (Deltares, 2021c). It was designed by Deltares as a tool for modelling hydrological processes. WFLOW accounts for precipitation, interception, snow accumulation and melt, evapotranspiration, soil water, surface water and groundwater recharge in a fully distributed environment. There are currently 8 published open source papers that have used WFLOW for their research. This includes one paper about environmental change in relation to hydrological processes (Gebremicael, Mohamed, & der Zaag, 2019), but no paper about deforestation. More research on WFLOW in relation to deforestation is needed to gain insight in how well it predicts discharge in areas with deforestation.

From the above literature overview it is concluded that hydrologic models are important for research on hydrologic processes and provide a tool for flood predictions. Deforestation has an effect on these hydrologic processes, but not necessarily on the predicted discharge. Whether the WFLOW model needs to explicitly account for deforestation to provide accurate discharge predictions in areas with deforestation is an unanswered research question that will be addressed in this work.

Scale

Papers on land-cover change in models are often paired watershed studies. These studies look at two similar catchments where one study area is impacted by land-cover change and

the second study area without change is used as a reference. Zhang et al. (2003) shows that paired catchment studies have been widely used as a means of determining the magnitude of water yield changes resulting from changes in vegetation. The review paper written by A. Brown, Zhang, McMahon, Western, and Vertessy (2005b) argues that there is a lack of information available in the literature for making quantitative generalisations about the impacts of vegetation changes on flow behaviour. This is supported by Beven (1993) who states that all hydrological models on catchment-scale can be easily invalidated when used on different or even similar study areas.

Research on catchment scale is still very useful to compare different model structures, but generalization would help to advance the field of hydrological modelling (A. Brown et al., 2005b). Furthermore, Enemark, Peeters, Mallants, and Batelaan (2019) argues that there is a need for a systematic approach to conceptual model building where all aspects of conceptualization relevant to the study goals are covered. Research conducted with a larger amount of catchments is needed to achieve the goal of a general approach in rainfall-runoff modelling.

Summary

In summary, while the rate of deforestation is declining in Brazil, the amount of deforestation remains an alarming issue. Forests play an important role in the hydrological cycle, mainly due to their ability to store and evaporate water. How does that connect to deforestation? The first conclusion from the literature review is that deforestation results in larger runoff coefficients, higher discharge peaks and more frequent extremes. However, some research shows little to no impact on these factors at all. Hydrologic models like WFLOW are used as tool to determine quantities like river discharge for different scenarios. Currently deforestation is not implemented in WFLOW. If deforestation has an impact on the river discharge and if increasing deforestation leads to worse discharge predictions by WFLOW, than taking deforestation explicitly into account will be needed to have reliable river discharge prediction from WFLOW.

The above literature review shows that there is a need for more research on the representation of the impact of deforestation on rainfall-runoff models and in particular WFLOW. This thesis aims to provide a part of that need for research. More elaboration on the goal and research questions is provided in the next chapter: Problem Definition.

2

Problem Definition

The goal of this thesis is to research the impact of deforestation on river runoff generated by measurements, a simple model and a complex model (WFLOW). This goal and the literature review result in the following research questions:

Research questions

- **Main**: How adequate is deforestation represented in WFLOW?
- **Sub1**: In areas with (heavy) deforestation, is the observed discharge affected by deforestation with respect to: runoff coefficient, timing and magnitude of extremes?
- **Sub2**: What is the impact of deforestation on the simulated discharge and storage of a simple conceptual model?
- **Sub3**: In areas with (heavy) deforestation, is the calculated river discharge from WFLOW similar to observed river discharge with respect to: runoff coefficient, timing and magnitude of extremes?
- **Sub4**: If the answer to sub3 is 'no', what changes in model inputs (such as infiltration capacity, soil maps, etc.) can improve the calculated river discharge from WFLOW in regions with heavy deforestation?

The results of the sub-questions contribute to answering the main research question. The goal of the first sub-question is to research the impact of deforestation on measured discharge. Our hypotheses is that deforestation leads to higher runoff coefficients and steeper discharge peaks. The second sub-question acts as intermediate step between observed discharge and simulated discharge from a complex model like WFLOW.

Sub-question three analyzes how adequate WFLOW currently calculates the hydrologic response for catchments with deforestation. This current model uses the land-cover type measured in 2009 as input for the entire time scope 2000-2020. In other words, yearly deforestation is not taken into account. The utility of analyzing results for a model without input for deforestation has multiple arguments. Firstly, it is important to know if WFLOW needs input for yearly deforestation at all to calculate adequate results. After all a possible conclusion is that the current model is sufficient for deforested areas in Brazil. Secondly, the model with one value for forest-cover applies as a standard situation. Without a standard situation it is impossible to analyze if the model with yearly deforestation performs better than the model with one value for forest-cover.

The final sub-question researches the possibilities for improvements on calculated river discharge from WFLOW in catchments with deforestation. The necessity for these improvements depends on the outcome of sub-question 2. The main focus of improvement is the implementation of yearly deforestation in WFLOW. In theory, input for yearly deforestation will provide a better representation of reality. However, it is important to research if this also leads to a more accurate calculation of discharge. The method for answering these sub-questions is discussed in the next chapter Methodology.

3

Methodology

This chapter describes the method used to answer the research questions formed in section 2. The goal of this thesis is to analyze the representation of the impact of deforestation in WFLOW. The research questions are derived from the review of the literature, which is described in section 1. The method consists of four components:

- Pre-processing phase
- Observed-data study
- Simple model study
- Complex model study

An overview and flow diagram of the method is shown in Figure 2. First, a pre-processing phase is done as preparation for the three studies. In this phase aspects of the scope like study area and used data are discussed. The three studies are conducted after this pre-processing phase is concluded.

The literature describes several possible conclusions for the impact of deforestation on discharge. Therefore the first step in determining how adequate deforestation is represented in WFLOW, is to investigate the relation between deforestation and measured data. This is done in the observed-data study. The characteristics of this measured data are described in subsection 3.1 for 30 catchments in Brazil. The impact of deforestation in these catchments is researched by conducting an analysis on the time-series of discharge, precipitation and deforestation. As described in Chapter 1, deforestation can affect several components of discharge. The time-series of the runoff coefficient (subsubsection 3.2.2) is analyzed to research if deforestation leads to an increase in discharge per precipitation in mm. Furthermore, the recession curves of discharge peaks are analyzed to see the impact on the response time of a catchment.

There is a large step between the use of observed discharge and simulating discharge from a complex model like WFLOW. A basic conceptual model is build to create an extra step between those two studies. This simple model has a recession coefficient to represent the relation between water in storage and discharge. Relating the yearly value of this storage coefficient with deforestation, gives insight in the impact of deforestation on simulated discharge of a simple model.

The complex model study is the last step in determining how adequate deforestation is represented in WFLOW (subsection 3.4). The catchments are modelled in WFLOW to obtain time-series for discharge for each catchment. Two scenarios are conducted; One with a fixed value for forest land cover and one that tries to implement yearly values for forest land cover (and thus includes deforestation). The second scenario is replaced by an analysis on the weak points of WFLOW, when the performance of the first scenario is not deemed satisfactory.

Together the three steps form the tools to answer the research questions and ultimately determine how adequate the impact of deforestation is represented in WFLOW.

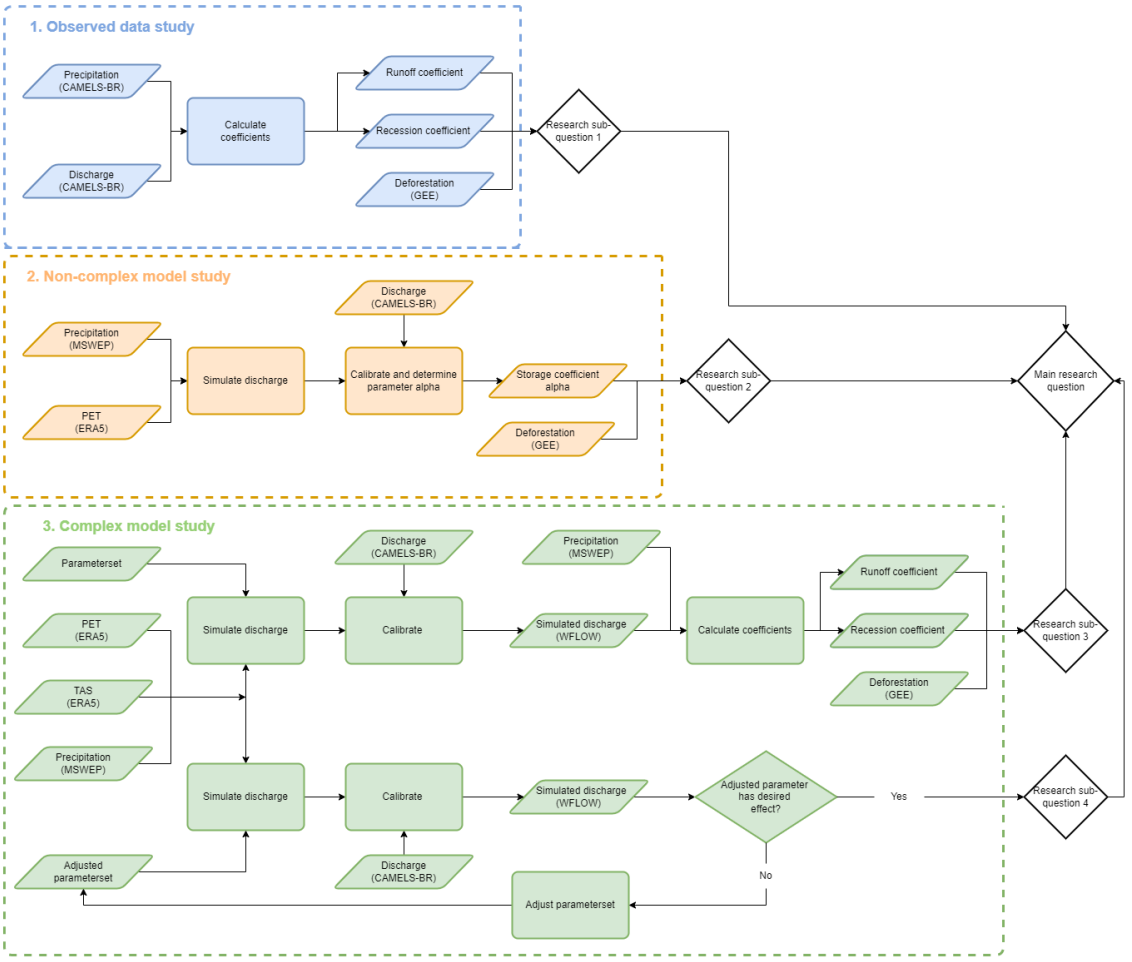

Figure 2: Data flow complete methodology

3.1 Pre-processing phase

The main part of the methodology is formed by three studies. A study with observed data, a simple model and a complex model study. Prior to these three studies, the study area and input for these studies are determined. This is done during the pre-processing phase. In the next paragraphs the chosen catchments and used data sets are described.

3.1.1 Data

Input for the studies is divided into forcing and metadata. Forcing describes input like precipitation and evaporation. Metadata is a term used to describes the characteristics of a catchment. The remainder of this sub-chapter elaborates on the properties of the different types of input and their sources.

Observed data study

Table 1 shows which sources are used for the study with observed data. The inputs in Table 1 are retrieved from CAMELS-BR, which is the abbreviation for Catchment Attributes and MEteorology for Large-sample Studies Brazil (Chagas et al., 2020). CAMELS-BR pro-

vides data for hundreds of catchments in the Amazon rain forest and Brazil. The data in CAMELS-BR is extensively screened en tested by the authors. One of their strong points is that uncertainties in precipitation and evapotranspiration are assessed with usage of several gridded products. Furthermore, CAMELS-BR consists of an large amount and variety of data, which adds to its versatility. More information about the screening and testing can be read in their paper (Chagas et al., 2020). The end result is a high quality data set perfectly suitable for a large-sample study like this thesis.

Table 1: Input sources of the observed data study

Input	Source		Unit	Timestep	Range
	Name	Reference			
Precipitation	CPC	(NOAA, 2019)	mm	daily	2000-2020
Evaporation	Gleam v3.3a	(Martens et al., 2017)	mm	daily	2000-2020
Discharge	ANA	(ANA, 2019)	m^3/s	daily	2020-2020
Area	ANA	(Do, Gudmundsson, Leonard, & Westra, 2018)	km ^{2}	-	-

Model studies

WFLOW is a distributed model, which means that the input for forcing also needs to be distributed. The CAMELS-BR data set does not provide distributed input, which makes it not suitable for the forcing of the model WFLOW. Therefore the MSWEP data set is used as input for precipitation in the model study. MSWEP provides 3-hourly global precipitation with a resolution of 0.1 degree. The data is available from 1979 to minus 3 hours of real time (Beck et al., 2019). Potential evapotranspiration and near-surface temperature are extracted from ERA5 (Hersbach et al., 2020). Input is not identical for the three studies, therefore a check is performed to see if the precipitation input for the studies are similar. This check is done to ensure that differences between results are not caused by dissimilarities in input. The result of this precipitation comparison is shown in subsection 4.3. There are catchments where some years differ significant between CAMELS-BR and MSWEP. However, these differences are not frequent and large. The conclusion of this check is that the differences between CAMELS-BR and MSWEP have no significant impact on the outcome of this research.

Table 2: Input source of the model studies

Input	Format	Source		Unit	Timestep	Range
		Name	Reference			
Precipitation	NetCDF	MSWEP	(Beck et al., 2019)	mm	daily	2000-2020
Potential evapotranspiration	NetCDF	MSWEP	(Beck et al., 2019)	mm	daily	2000-2020
Temperature	NetCDF	MSWEP	(Beck et al., 2019)	$°C$	daily	2000-2020
Shapefile	.shp	ANA	(ANA, 2019)	-	-	-
Digital elevation map	PCRaster	SRTM	(NASA, 2014)	-	-	-
Landcover map	PCRaster	Globcover2019	(ESA2010 & UCLouvain, 2010)	-	-	-
Soil map	PCRaster	HWSD	(FAO/IIASA/ISRIC/ISSCAS/JRC, 2013)	-	-	-
Leaf area index	Lookup table	Copernicus	(Copernicus, 2021)	-	-	-
Configuration file	.ini	Deltares	(Deltares, 2021a)	-	-	-

Deforestation

Deforestation is retrieved from the Global Forest Change map in Google Earth Engine (GEE) (Hansen et al., 2020).Trees are defined as vegetation taller than 5m in height and are expressed as a percentage per output grid cell. Forest Cover Loss is defined as a stand-

replacement disturbance, or a change from a forest to non-forest state, during the period 2000–2020 (Hansen et al., 2020). The data is expressed in pixels per year. The amount of pixels in a year displays the amount of deforestation in that year. The pixels are summed up per year (cumulative deforestation) to analyze the impact of deforestation over time. Figure 3 shows the forest cover loss of the part of Brazil including the study area of this thesis. The loss is represented as pixels over the entire 20 year period. Heavily deforested areas are indicated by a higher concentration of red pixels.

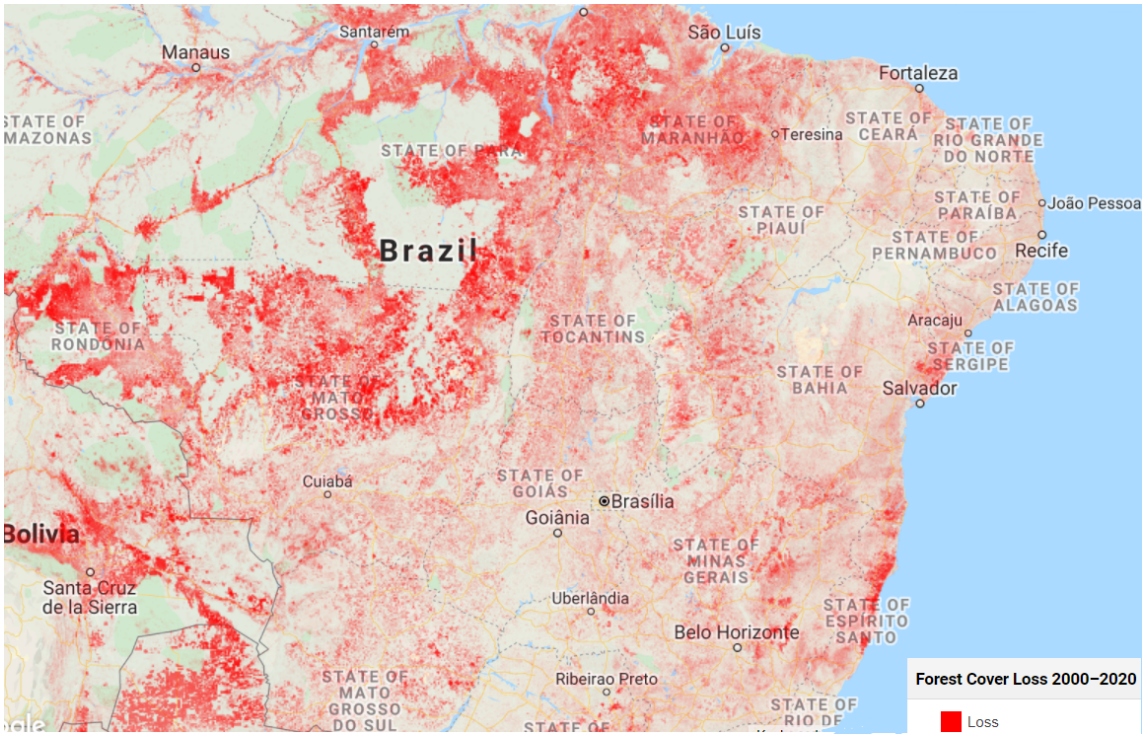

Figure 3: Deforestation for 2000-2020.

The definition deforestation can be interpreted in several ways. In this thesis deforestation is defined as the percentage of forest area loss in relation to the total land area of a catchment. This is illustrated in Figure 4. The top (a) of this figure shows the deforestation for two catchments in our study area. It is observed that catchments can have varying concentrations of deforestation. The center (b) displays a figure with the time series of daily discharge and cumulative deforestation. The cumulative deforestation for these two catchments is nearly equal, while having very different concentrations of deforestation. The bottom (c) of the figure shows a table with the total land area, the deforested area and percentage of deforestation per catchment. This table helps to illustrate our definition of deforestation. As shown in the table catchment 2 "Fazenda bandeira" has a much larger area than catchment 1 "Bom jardim". However the percentage of deforestation is almost equal. Hence, the percentage of deforestation is defined as the percentage of forest area loss in relation to the total land area of a catchment.

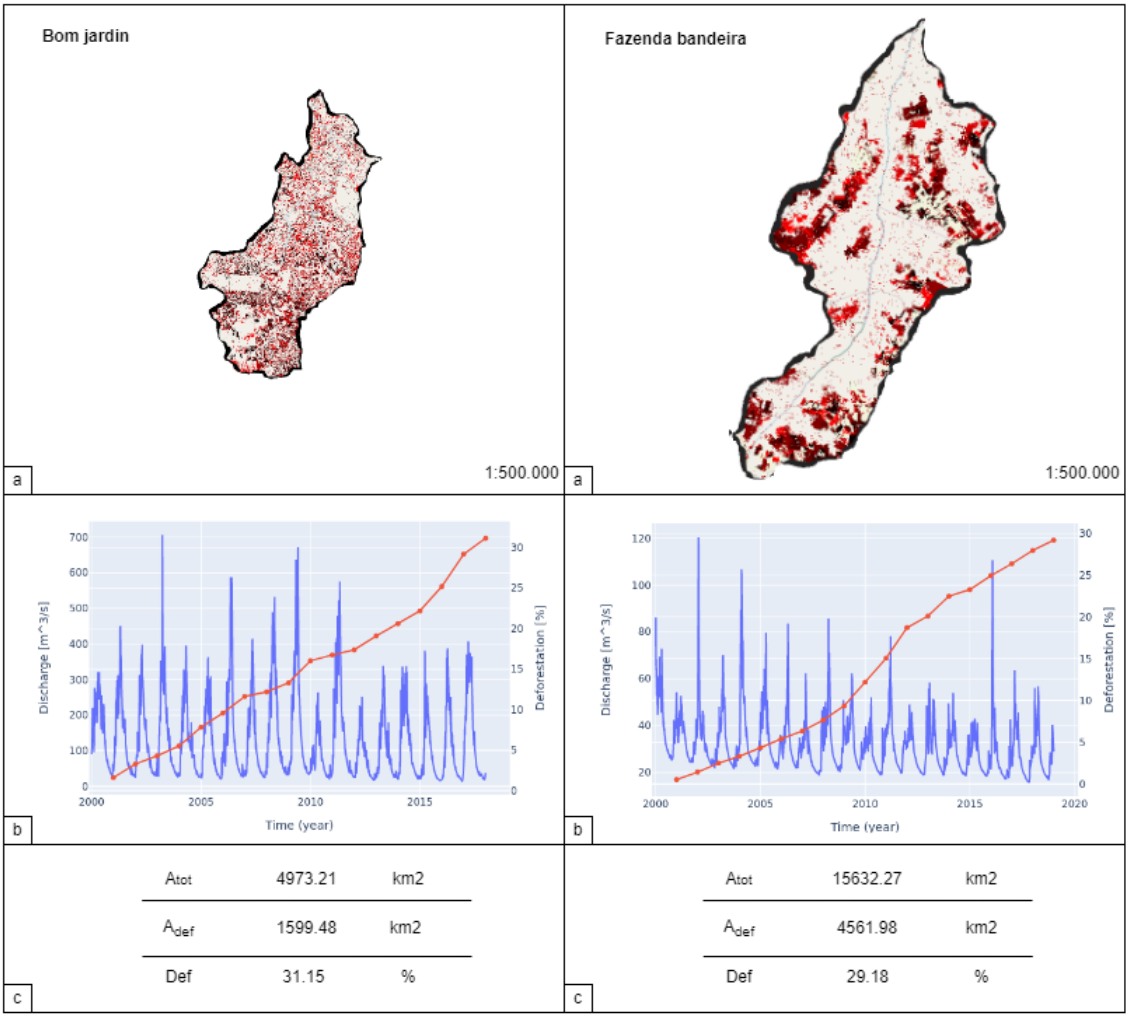

Figure 4: a) Concentration of deforestation over 20 years. The severity of deforestation is indicated by the darkness of the red pixels. b) Figure displaying the daily discharge (left ax) and cumulative deforestation (right ax) in time. c) Table showing how percentage of deforestation is calculated.

3.1.2 Study area

The study area is located within Brazil. The availability of high quality data and the magnitude of deforestation, make Brazil a suitable choice for this research. Within Brazil a total of 30 catchments are selected as study area. An overview of the selected catchments is shown in Figure 5.

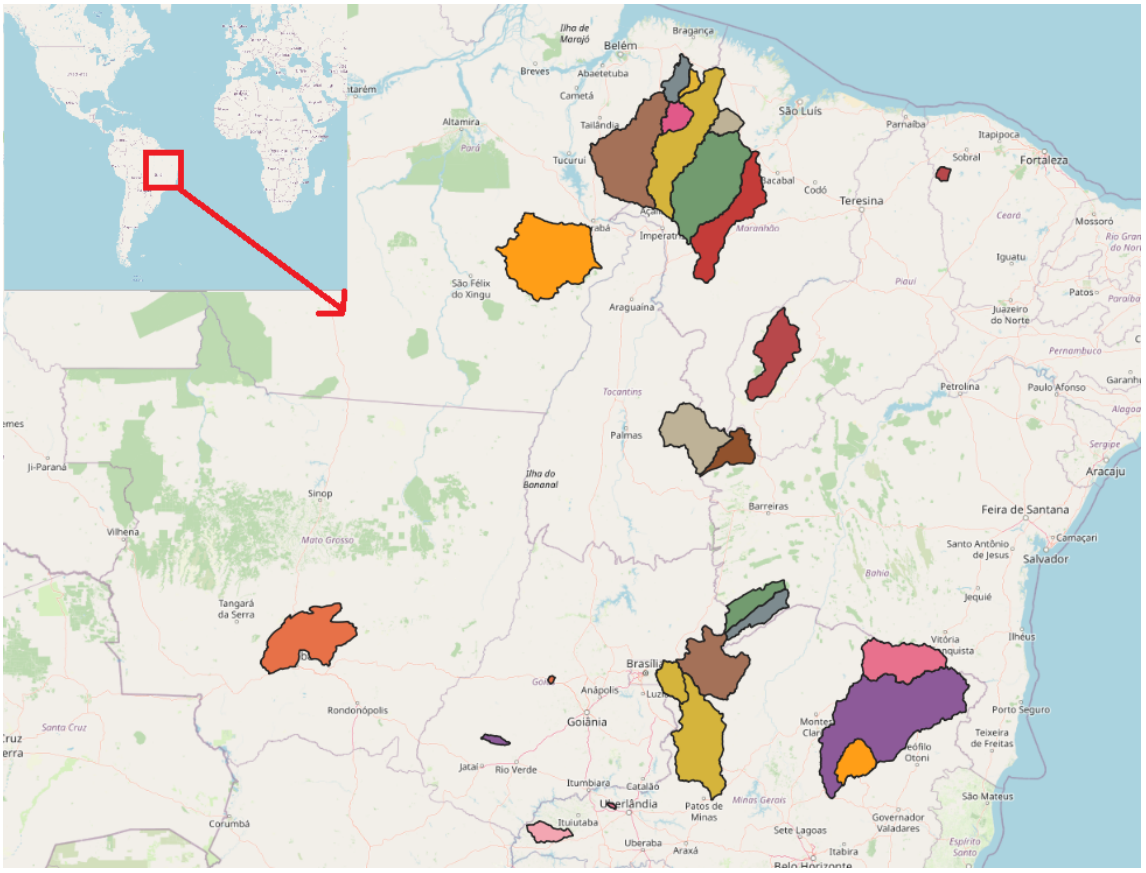

Figure 5: Overview of the selected catchments

The catchments of the study area are selected through five steps. These steps are shown in Figure 6. The outlined number in a step represents the amount of suitable catchments after that step.

The selection of the first step consists of all 897 catchments in CAMELS-BR. The data set offers a quality check by Agência Nacional de Águas (ANA), which is the Brazilian national water agency. The percentage of streamflow data (1990-2009) with quality control checks by ANA is used as a indicator for the second step. The criteria for selection of a catchment is that 95 % of streamflow data is checked on quality by ANA. This step reduces the amount of available catchments to 715. The deforestation data is obtained in pieces of 10x10 degree latitude and longitude from (Hansen et al., 2020). The area spanning the range latitude 0° north - 10° south and longitude 50° west - 60° west is chosen because of the degree of deforestation in that area. There are 281 catchments with quality control within the selected area.

The last two steps in the selection process require more explanation. Step 4 involves the occurrence of (heavy) deforestation. A catchment must be subject to deforestation in order to research the impact of deforestation on hydrological properties of that catchment. For this research catchments with at least 5% deforestation over a 20 year period are considered as study area. The final 30 catchments were selected based on the following criteria: Catchment character, degree of deforestation and the absence of abnormalities

in the data. Catchment character consists of area size and the location of a catchment. As described in subsubsection 3.1.1 deforestation is linked to the area of a catchment. Smaller catchments tend to be more sensitive for changes in land cover. A study area that consists of catchments with a wide range of area sizes is desired to research the impact of deforestation on a variety of basins. Catchments in the CAMELS-BR dataset can be part of a larger catchment. For example, one of the larger catchments consists of 12 smaller catchments. However, these smaller catchment are not 12 separate areas but have overlap instead. To illustrate, the second largest catchment in this subset of 12 is equal to the largest catchment minus one area. This is not desirable as it opposes the general character of this research. The overlap of catchments is taking into account during the final step of selection to avoid having almost identical catchments. The study area is also determined by the degree of deforestation. In step 4 a selection was made for catchments with at least 5% deforestation over a 20 year period. However, there is still a large variety of deforestation within that selection. Variety in degree of deforestation provides an opportunity to analyze the impact of several types of deforestation. For example, are there differences in results between catchments with slow and fast rates of deforestation? The previous selection criteria are related to the properties of a catchment. In addition, the quality of input is analyzed. The daily observed discharge is an important variable for this research. A number of catchments showed large gaps in the data of this daily observed discharge. Data sets with numerous large gaps, such as missing months every year, are not selected to preserve the quality of this thesis.

In summary, all steps in Figure 6 are used to select a wide variety of catchments which are subject to deforestation. The variety of the study area contributes to finding relations independent of catchment character.

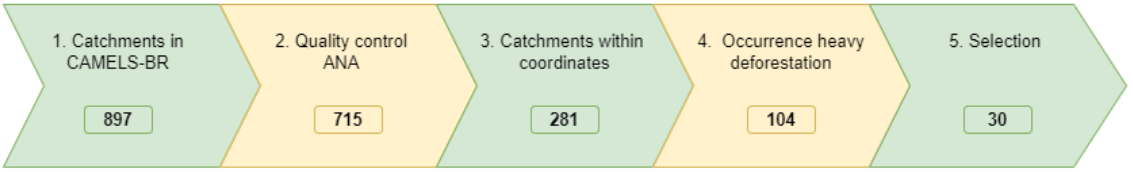

Figure 6: Flowchart for selection of catchments

Physical properties are an important factor in understanding the hydrology of a catchment. An overview of these psychical properties is shown below in Table 3. The values of these characteristics are obtained by CAMELS-BR through different sources. The name and coordinates of a catchment are provided by the National Water Agency Brazil (ANA) (ANA, 2019). The longitude and latitude of a gauge are also used in the WFLOW study as point where output river runoff is calculated. The mean elevation is determined through a Digital Elevation Map (DEM) from the HydroSHEDS database. The resolution of this DEM is 15 arc-second (Lehrer, Verdin, & Jarvis, 2008). The variety in mean elevation between the catchments adds to the diversity of the study area. Climate parameters like precipitation help to gain insight in the hydrology of a catchment. The annual mean precipitation is provided by the Climate Prediction Center (CPC). The soil largely consists of three types: sand, silt and clay. Knowing the structure of the soil is necessary to explain the behaviour of fluxes like infiltration and soil evaporation. CAMELS-BR obtains the percentage of each type of soil from (Batjes, Ribeiro, & van Oostrum, 2020). The

soil describes the underground properties of a catchment. The above ground properties are mainly determined by type of land cover. In CAMELS-BR these types consist of forest, crops, mosaic, shrubs and barren ESA2010 and UCLouvain (2010). The majority of dominant land cover in our study area is tropical forest or mosaic, where mosaic means that the land is covered by a mix of croplands and natural vegetation.

Detailed descriptions of the above and more metadata can be found on (Chagas et al., 2020).

Table 3: Key characteristics of catchments in study area

Catchment	Location			Annuel mean P [mm]	Soil			Dominant landcover
	°*Latitude*	°*Longitude*	Elevation [m]		Sand [%]	Silt [%]	Clay [%]	
Porto roncador	-13.55	-55.33	473	1690	59.12	13.63	27.25	Mosaic
Alto araguaia	-17.30	-53.21	776	1425	56.01	12.15	31.84	Mosaic
Fazenda alegria	-5.48	-49.22	270	1643	40.24	21.83	37.93	Tropical forest
Vila capoeira	-1.86	-47.05	79	2044	47.44	18.56	33.96	Tropical forest
Bom jardim	-1.54	-47.06	75	2078	47.31	18.69	34.00	Tropical forest
Badajos	-2.51	-47.76	145	1719	43.60	17.85	38.55	Tropical forest
Sete ilhas	-1.85	-46.70	82	2218	45.93	18.38	35.66	Tropical forest
Cafezal	-2.77	-46.80	129	1670	44.39	17.82	37.79	Tropical forest
Alto Bonito	-1.80	-46.31	144	1568	43.34	18.04	38.62	Tropical forest
Alto turi	-2.94	-45.66	72	1654	42.62	20.26	37.12	Tropical forest
Pindare-mirim	-3.66	-45.46	198	1315	44.02	16.90	39.08	Tropical forest
Grajau	-5.81	-46.14	308	1064	54.64	15.39	29.97	Mosaic
Aratoi grande	-3.77	-45.21	211	1321	50.46	17.19	32.35	Mosaic
Mendes	-5.70	-43.58	281	1152	57.10	13.83	29.07	Tropical forest
Fazenda bandeira	-7.39	-44.61	474	960	59.42	13.43	27.15	Shrub
Cristino castro	-8.79	-44.20	475	911	57.39	15.21	27.40	Mosaic
Porto da extrema	-17.03	-46.01	702	1219	44.98	18.06	36.96	Mosaic
Vila urucuia	-16.13	-45.74	756	1081	48.09	17.67	34.24	Mosaic
Correntina	-13.34	-44.63	794	1027	55.96	14.16	29.88	Mosaic
Gatos	-13.71	-44.63	790	1077	56.29	14.61	29.10	Shrub
Fazenda bom jardim	-10.99	-45.52	737	995	53.65	14.86	31.49	Shrub
Inhobim	-15.33	-40.93	867	735	46.86	14.80	38.28	Tropical forest
Fazenda jambeiro-grao mogol	-16.59	-42.91	927	946	47.99	16.04	35.97	Shrub
Barra do salinas	-16.61	-42.30	837	929	46.42	16.44	37.14	Shrub
Ponte alta	-17.28	-42.80	868	1172	41.43	15.89	42.68	Tropical forest
Jacinto	-16.13	-40.30	744	893	44.93	15.93	39.14	Tropical forest
Barra do rio preto	-18.69	-40.88	441	988	44.78	14.36	40.86	Mosaic
Ponte do pancas	-19.42	-40.68	300	994	44.77	13.65	41.58	Mosaic
Fazenda buriti do prata	-19.35	-49.18	686	1390	50.22	13.57	36.21	Mosaic
Cuiaba	-15.61	-56.10	375	1455	57.18	16.10	26.72	Mosaic

3.2 Study with observed data

The literature mentioned in chapter 1 has several conclusions on the impact of deforestation on rainfall-runoff models. Although the majority of literature shows a direct relation between deforestation and an increase in discharge, there is research that shows that deforestation does not necessarily results in an increase of discharge. A study on the CAMELS-BR data set is conducted to understand which conclusions are valid for the chosen study area. This study is referred to as the observed data study in the remainder of this thesis. The data flow for the observed data study is shown in Figure 7.

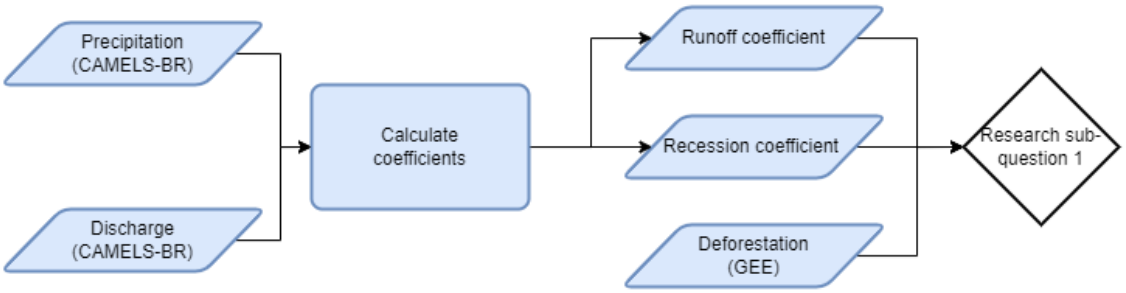

Figure 7: Data flow: observed data study

The programming language Python is used to analyze measured data like discharge. The code written for the observed data study is located on Github (de Wildt, 2022). Before analyzing the measured data (discharge, precipitation etc), the data obtained in subsection 3.1 has to be processed. The goal of this processing step is to create usable dataframes, which enhances the work-ability of the data. Daily rainfall and discharge measurements are summed to yearly values and missing data entries are filled by interpolation.

The second step is to process the deforestation data. As mentioned in subsubsection 3.1.1 the values of deforestation are measured in pixels per year. The amount of pixels represent the loss of forest in that year. A pixel is defined as a projection of 30 m^2 (Hansen et al., 2020). The amount of deforestation per year is then calculated in units of km2. In addition, the amount of deforestation is converted to the percentage of the total forest area of a catchment. The method is shown in Equation 1. The area of a catchment is determined with use of the metadata shown in Table 3.

$$\%F_i = \frac{F_i}{A_i} \tag{1}$$

F_i = Deforestation of catchment i [km^2], A_i = Total land area of catchment i [km^2].

The previous steps result in a basis for the remainder of the observed data study. This basis is used to analyze factors like yearly runoff coefficients and yearly discharge recession curve coefficients.

3.2.1 Actual evapotranspiration

Chapter 1 discussed conclusions from several research papers about the impact of deforestation on runoff. Some of these papers state that the impact of deforestation is masked by young forests or agriculture. Due to the unavailability of data for land cover type after deforestation, it is not possible to directly observe if this masking by other flora is occurring in this study area. However, it is possible to look at the relation between actual evapotranspiration (AET) and deforestation. Among others, Wang-Erlandsson, van der Ent, Gordon, and Savenije (2014) state that deforestation should lead to a decrease in AET. An absence of this decrease in AET can be used as indication for the presence of flora that is not detected by satellite images from Hansen et al. (2020) like young forests. The CAMELS-BR dataset offers two subsets for actual evapotranspiration; GLEAM (Martens et al., 2017) and MGB-SA (Siquera et al., 2018). Both of these sources use modelling to determine AET.

3.2.2 Runoff Coefficient

The literature mentioned in Chapter 1 shows that deforestation does not necessarily results in an increase of the runoff coefficient. The runoff coefficient is defined as a dimensionless coefficient relating the amount of runoff to the amount of precipitation received. This can be obtained by dividing the amount of precipitation in time by the amount of runoff in time. The equation is shown in Equation 2. Daily precipitation and discharge are used as input, these daily values are then summed to a yearly value to make it compatible with yearly values for deforestation. Therefore, the timescale of runoff coefficients is yearly. The data flow diagram can be observed in Figure 8.

$$Cr_{(t)} = \frac{\int_{t=to}^{t} Q_{(t)} dt}{\int_{t=to}^{t} P_{(t)} dt} \tag{2}$$

Cr_y = Runoff coefficient of period t [-], $Q_{(t)}$ = Cumulative discharge per year [m^3/s], $P_{(t)}$ = Cumulative precipitation per year [m^3/s].

The correlation between deforestation and the runoff coefficient is visualized by a scatter plot with C_r against the percentage of yearly deforestation. All catchments are included with yearly time steps. The runoff coefficient is set on the x-axis and the percentage of deforestation is set on the y-axis. The value of deforestation is presented as cumulative to represent the true amount of deforestation. In addition, the relation is analyzed per catchment to observe individual signals.

Finally the relation between the runoff coefficient and non-cumulative yearly deforestation is analyzed. Non-cumulative deforestation shows the short term impact of deforestation on river runoff. The combination of analysis on both cumulative and non-cumulative deforestation results in a more complete insight in the impact of deforestation on observed discharge.

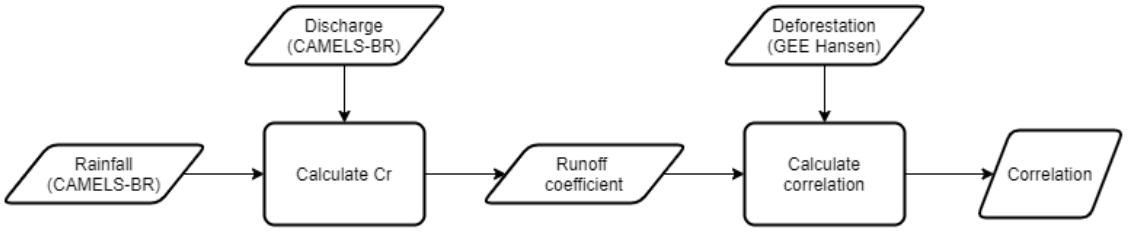

Figure 8: Data flow diagram for determining the correlation between runoff coefficient and deforestation

3.2.3 Recession coefficient

The literature mentioned in section 1 do not have a clear conclusion on the impact of deforestation on rainfall-runoff models. The majority of literature concludes that deforestation leads to higher recession coefficients. An analysis on the time series of discharge recession is conducted to validate if this general conclusion holds for our study area.

The recession curve for discharge is defined by (Jain, 2014) as "the specific part of discharge hydrograph after the crest (and the precipitation event), where streamflow diminishes dur-

ing a rainless or dry period is called the flow recession curve, at a specific river cross section of the basin. Hence, recession curves are the part of the hydrograph that are dominated by the release of water from natural storages". Figure 9 shows a typical discharge peak with different components of the recession curve.

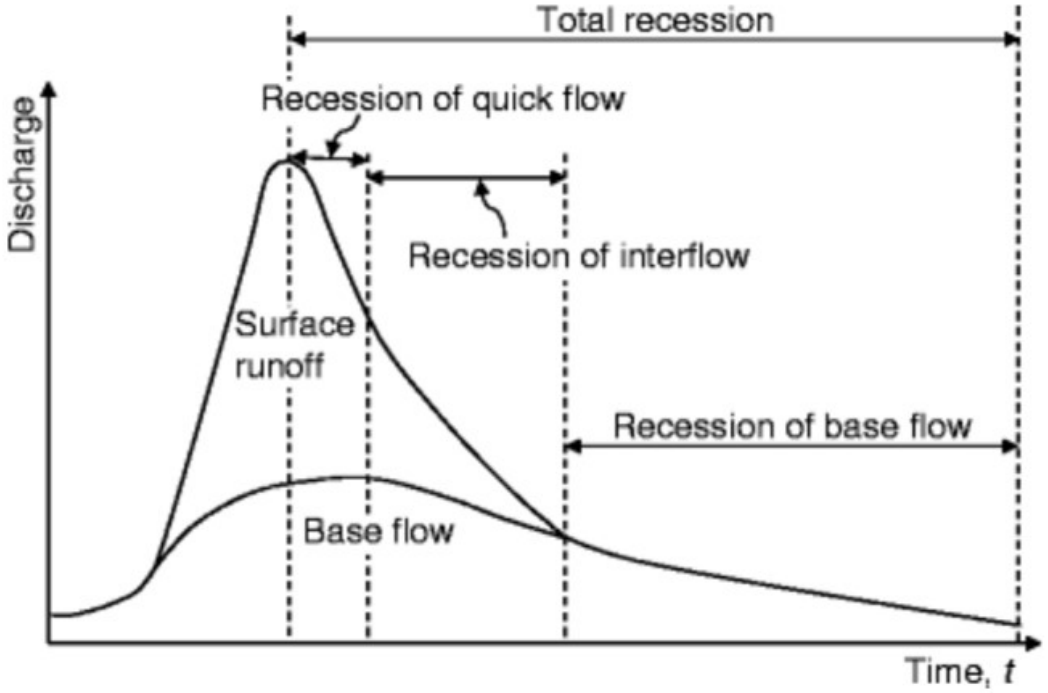

Figure 9: Typical hydrograph showing recession of individual components of runoff (Jain, 2014)

Equation 3 shows the relation between the flow at time t and t_0, where t_0 corresponds to the time t of the top of a discharge peak. In this equation the depletion factor k indicates the dominant type of flow at time t. Higher values of k results in faster draining of water from the basin, whereas a lower value of k represents slower drainage from the basin. Therefore the value of k is also an indication for the response of the different types of storages in the catchment.

$$Q_{(t)} = Q_0 * e^{-k(t-t_0)} \tag{3}$$

$$t - t_0 > 0$$

Q_t = discharge at time t [m^3/s], Q_0 = initial discharge at start of recession [m^3/s], k = depletion factor [-]

The time series of discharge for a single catchment consists of numerous peaks. These discharge peaks are caused by precipitation events. The recession curves of these discharge peaks must be defined. A recession curve has a start and an end. The top of a discharge

peak marks the start of a recession curve. An example of this peak detection is shown at Figure 10. Peaks are detected when they have a prominence of at least 0.2. The prominence of a peak measures how much a peak stands out from the surrounding baseline of the signal and is defined as the vertical distance between a peak and its lowest contour line. The value for prominence is determined by a trial and error approach. Very small peaks are considered noise and should therefore be disregarded. A balance is desired to cancel the noise without disregarding relevant peaks. This desire led to a prominence of 0.2 for this research.

The end of a recession curve is defined as time t where the minimum value for logarithmic discharge is reached between two peaks. It can occur that peaks are detected in close proximity of each other, which is also shown in Figure 10. The result of two peaks in close proximity is a short recession curve with a low amount of data. The reliability of a linear fit through a subset is partly dependent on the amount of data in that subset. A higher amount of data result in a more reliable linear fit. Recession curves with less than 20 days are removed, because they are considered as non-reliable.

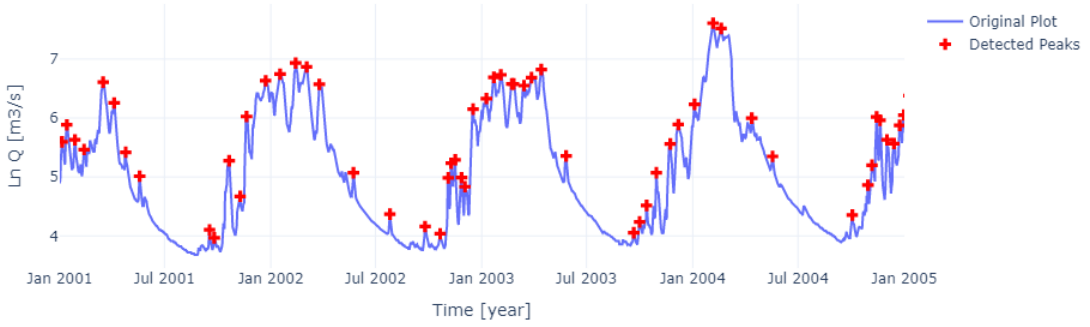

Figure 10: Time-series of natural logarithm of discharge with detection of discharge peaks (de Wildt, 2022)

The standard equation of a linear fit is used to rewrite Equation 3 into Equation 4 a linear equation. Recession coefficient k is calculated by applying a linear fit on the recession curves of the logarithmic discharge peaks, defined by Equation 4. As shown in Figure 10 a single year has multiple peaks of discharge, which also means that a single year has multiple recession coefficients k. The goal is to analyze the relation between k and deforestation. The unit for deforestation is km^2/year, which means the unit for k must be per year. This yearly k is determined by calculating the mean k for each year.

$$\ln Q_{(t)} = \ln Q_{(0)} - k * t \tag{4}$$

The results are used to analyze if a relation between discharge recession and deforestation is observed for our study area. The dataflow diagram can be observed in Figure 11.

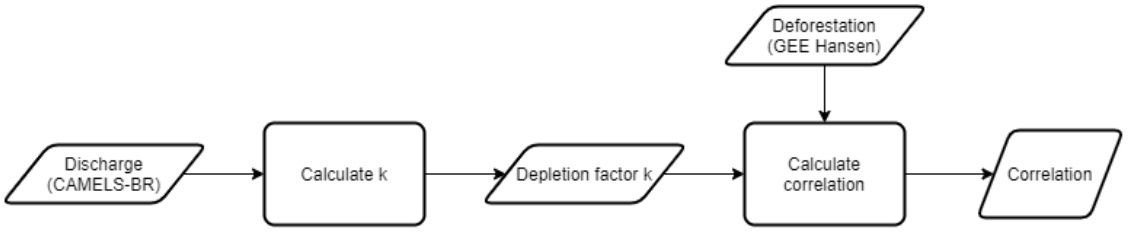

Figure 11: Data flow diagram for determining the correlation between recession coefficient and deforestation

3.3 Study with simple model

A basic conceptual model is build to accommodate a smoother transition from this measured discharge to WFLOW. The code written for the simple model study is located on Github (de Wildt, 2022).

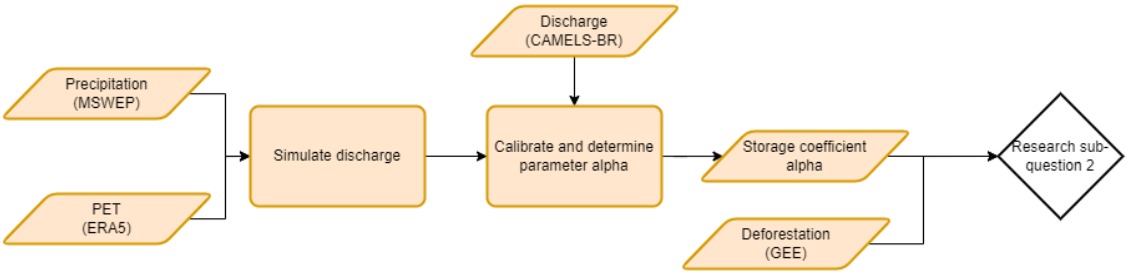

Figure 12: Data flow diagram for determining the impact of deforestation on a simple conceptual model

The model

A conceptual model is a hydrological model that uses simplified mathematical conceptualizations to describe a system with help of hydrological processes and interconnected storages.

Conceptual models differ in complexity ranging from very basic (one output and one input) to models with several storages and a variety of hydrological processes. In this research a simple model is build to see if the impact of deforestation is noticeable on such a basic level of complexity. Figure 13 shows the setup of the build conceptual model.

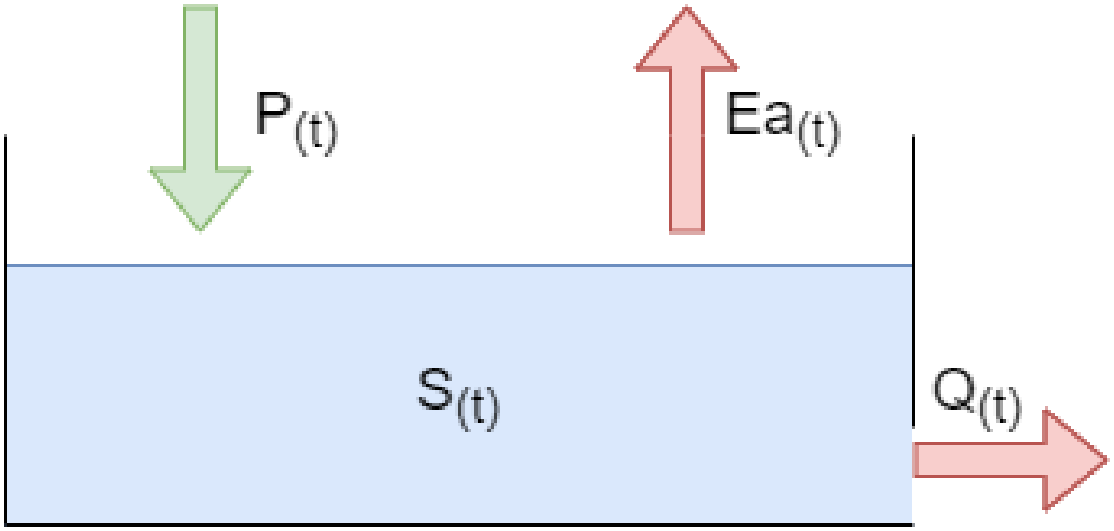

Figure 13: Basic conceptual model with one storage bucket

The hydrological concept from Figure 13 can be translated to a mathematical format shown in Equation 5. Storage ($S_{(t)}$) is determined by the amount of water currently in the system ($S_{(t-1)}$), precipitation ($P_{(t)}$), actual evapotranspiration ($E_{a(t)}$) and discharge ($Q_{(t)}$). Input is known for precipitation and potential evapotranspiration (PET) (Table 2). However, actual evapotranspiration (AET) is preferred over PET to enhance the quality of this simple model. Equation 6 is used to transform PET to AET. This equation describes that the actual evapotranspiration is dependent on the water available in the storage and the potential evapotranspiration. The minimum of these two parameters is used as evapotranspiration. This means that in this simple model PET can be equal to the actual evapotranspiration if there is enough water available in the storage.

The output of our model is discharge. Discharge is related to the amount of water in the system after precipitation and evapotranspiration. Equation 7 shows how discharge is determined by recession coefficient α. This coefficient is used to represent which part of the water in the system ends up in discharge. A larger value for α means more discharge and therefore less water in storage. The exact value of this coefficient is unknown and will later be used as parameter for calibration.

$$S_{(t)} = S_{(t-1)} + P_{(t)} - E_{a(t)} - Q_{(t)} \tag{5}$$

$S_{(t)}$ = Storage at time t [mm], $P_{(t)}$ = Precipitation at time t [mm], $E_{a(t)}$ = Actual evapotranspiration at time t [mm], $Q_{(t)}$ = Discharge at time t [mm]

$$E_{a(t)} = min(S_{(t)}, E_{p(t)}) \tag{6}$$

E_a = Actual evapotranspiration [mm], S = Storage [mm], E = Potential evapotranspiration [mm]

$$Q_{(t)} = \alpha \cdot S_{(t)} \tag{7}$$

$Q_{(t)}$ = *Discharge [mm],* α = *Storage coefficient [-],* S = *Storage [mm]*

Calibration

The calibration of a simple model can be done by hand with the so called "trial and error" method. However, this method can be less accurate and time consuming especially when the study area consist of multiple catchments. The Nash-Sutcliffe efficiency coefficient is used as objective function. The Nash–Sutcliffe efficiency is calculated as one minus the ratio of the error variance of the modeled time-series divided by the variance of the observed time-series, which is shown in Equation 8. In the situation of a perfect model with an estimation error variance equal to zero, the resulting Nash–Sutcliffe Efficiency equals 1 (NSE = 1). This also means that a model that produces an estimation error variance equal to the variance of the observed time series results in a Nash–Sutcliffe Efficiency of 0.0 (NSE = 0) (Nash & Sutcliffe, 1970).

$$NSE = 1 - \frac{\sum_{t=1}^{T}(Q_m^t - Q_o^t)^2}{\sum_{t=1}^{T}(Q_o^t - \overline{Q_o})^2} \tag{8}$$

NSE = Nash–Sutcliffe model efficiency coefficient [-], $\overline{Q_o}$ = *mean of observed discharges [*m^3/s*],* Q_m= *modeled discharge [*m^3/s*],* Q_o^t = *observed discharge at time t [*m^3/s*]*

Our goal is to analyze the impact of deforestation on model performance. In this basic model the only usable parameter to analyze this impact is the calibrated recession coefficient α. By calibrating each year individually we can determine the yearly α. The values of these yearly α can then be plotted against the yearly cumulative deforestation to analyze if deforestation results in a different division between storage and discharge.

3.4 Study with complex model

The observed data study used measured discharge to analyze the impact of deforestation on runoff and the simple model study is used to create a step between observed data and a complex model like WFLOW. The goal of this complex model study is to calculate discharge for catchments subject to deforestation. The dataflow illustrating the method to fulfill this goal is shown in Figure 14. The code written for the complex model study is located on Github (de Wildt, 2022).

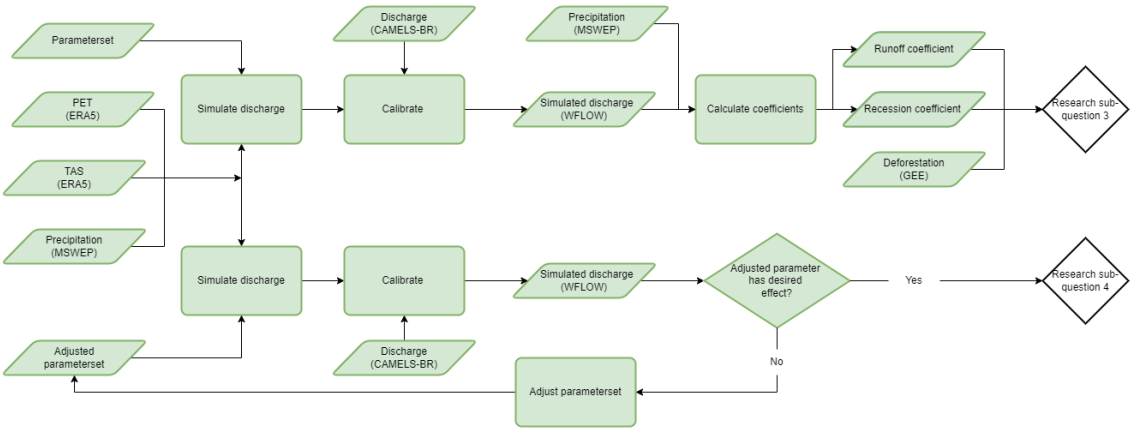

Figure 14: Data flow diagram for determining the impact of deforestation on a complex model

The model software used is WFLOW_sbm 2020.1.1. The virtual environment of eWaterCycle is used to run WFLOW (eWaterCycle, 2021). A more elaborate explanation of the eWaterCycle can be found on (Hut et al., 2021). WFLOW is a completely distributed modelling platform which maximizes the use of open earth observation data. Based on gridded topography, soil, land use and climate data, WFLOW calculates all hydrological fluxes at any given point in the model at a given time step. The data used as input is described in subsection 3.1.

As shown in Figure 14 WFLOW needs two types of input: forcing and a parameterset. Forcing input for WFLOW consists of precipitation, potential evapotranspiration and near-surface temperature. Forcing in WFLOW is generated through ESMValTool. This tool uses existing climate datasets like MSWEP to generate forcing for a determined scope. The time range of this scope is manually set in the model and the area is determined by providing a shapefile and DEM of a catchment. The sources for all input needed to create the forcing is shown in Table 2. More information about ESMValTool and its manual can be found on (eWaterCycle, 2021).

The second input for WFLOW is a parameterset. The parameterset describes the physical properties like the landcover or which type of soil is present in a catchment. This parameterset is build with two configuration script for each catchment, of which the steps are shown in Figure 15. The first configuration file (step 1) requires a DEM, landuse map and soil map as input. The sources for the maps used in this thesis can be found in Table 2. A

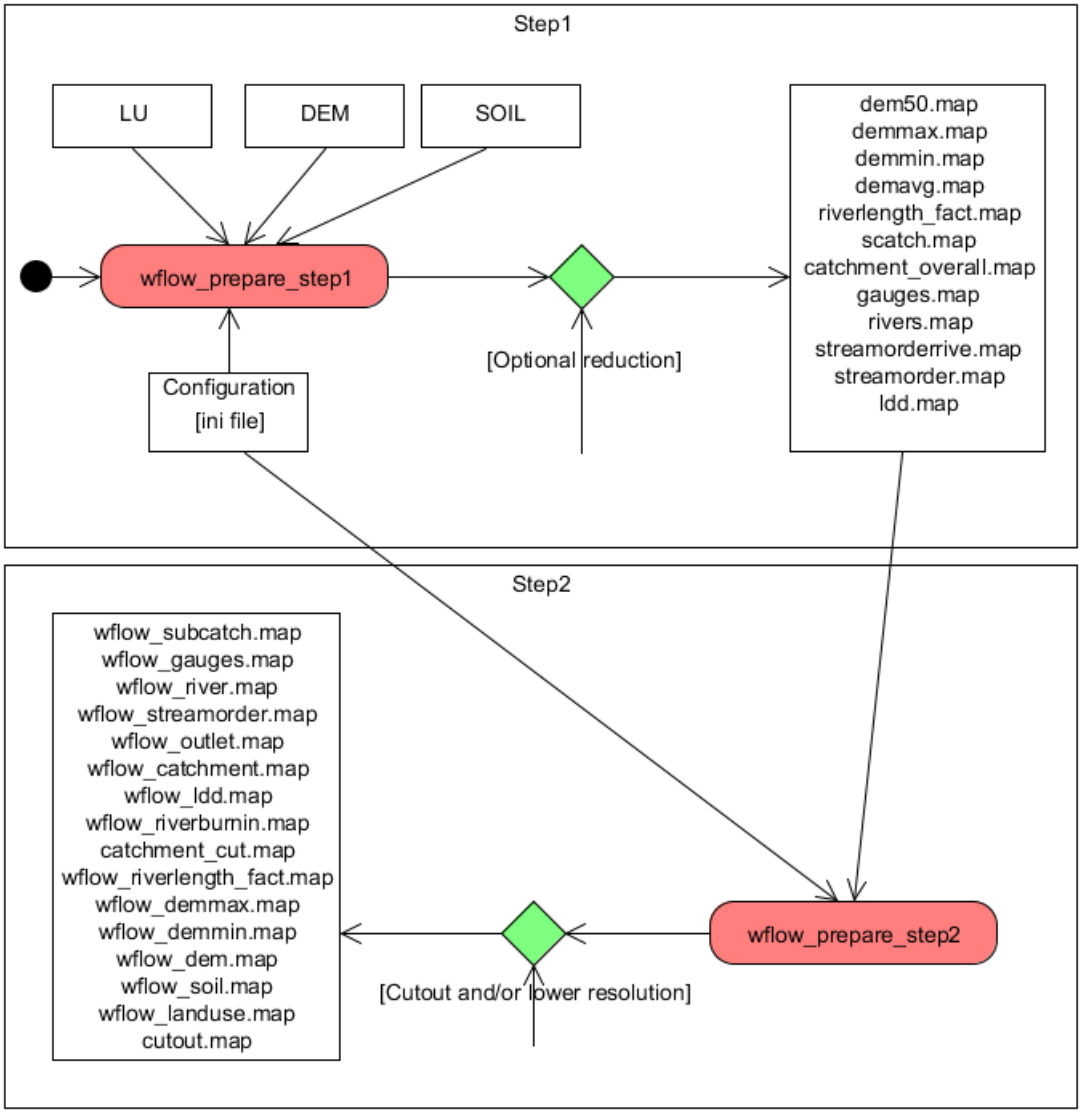

Figure 15: Steps in creating the WFLOW parameterset Deltares (2021a)

More information on the configuration scripts can be found on (Deltares, 2021a).

Together, the forcing and parameterset form the input of our model. The output of the model is discharge and several hydrological processes like excess infiltration. The simulated discharge will be used to determine runoff coefficient and recession coefficient. The method for determining these two coefficients is equal to the method described in subsection 3.2.

3.4.1 Processes

An overview of the different processes and fluxes implemented by Deltares in the WFLOW_sbm model is shown in Figure 16. The value of the fluxes and storages in Figure 16 are determined by a variety of hydrological and mathematical approaches. The chosen approach of

the major processes in relation to deforestation are discussed in this sub-chapter, because it is relevant in identifying suitable parameters for calibration.

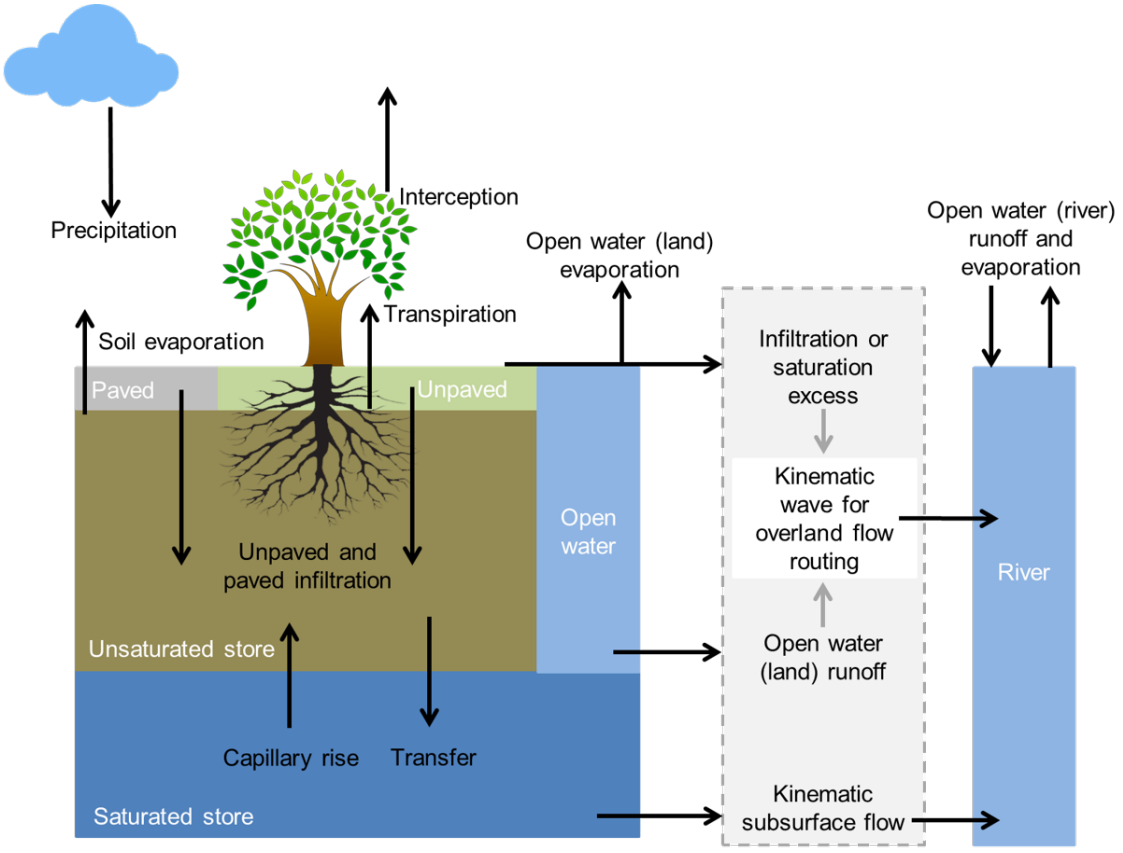

Figure 16: Overview of the different processes and fluxes in the WFLOW_sbm model (Deltares, 2021b)

The input for evaporation is assumed to be potential evaporation by Deltares. Potential evapotranspiration is defined as the amount of evaporation that would occur if a sufficient water and energy is available.

Rainfall interception by canopy is an important factor in research about the impact of deforestation. WFLOW_sbm makes use of the analytical Gash model. A detailed explanation on the Gash model is found in the documentary of WFLOW Deltares (2019).

The Gash parameters are determined by the implementation of Leaf Area Index (LAI) maps and specific leaf storage Sl. The Sl and C_{wood} are determined via a lookup table with different types of land cover. These tables can be observed in the documentation of WFLOW (Deltares, 2019) and use Globcover as source. The maximum storage of canopy is determined by Equation 9.

$$C_{max} = (Sl \cdot LAI) + C_{wood} \tag{9}$$

C_{max} = Maximum canopy storage [mm] , Sl= Specific leaf storage [mm] , LAI = Leaf area index [-], C_{wood} = Canopy storage woody part vegetation [mm]

Figure 16 illustrates the importance of the unsaturated zone in relation to vegetation. The change in storage of this zone is greatly dependent on the infiltration capacity of the soil. The infiltrating water is split in two parts, the part that falls on compacted areas and the part that falls on non-compacted areas. The water that can actual infiltrate is calculated by taking the minimum of the total maximum infiltration rate (compacted and non-compacted areas) and the remaining storage capacity (Deltares, 2019). If the infiltration capacity is smaller than the rate of infiltration, there will be an infiltration excess. This excess of water is then added to the overland flow.

The soil consists of the unsaturated and saturated zone. Both storages are important for this research and to identify possible parameters for calibration. The hydraulic conductivity k_{ver} [mm/day] is used to describe the flow of soil water. Soil parameter M [mm] determines the decrease of this vertical saturated conductivity with depth.

Limitations

The documentation of WFLOW (Deltares, 2019) discusses the limitations of the SBM concept for WFLOW. The WFLOW_sbm concept uses the kinematic wave approach for channel, overland and lateral subsurface flow, assuming that the topography controls water flow mostly. This assumption holds for steep terrain, but in less steep terrain the hydraulic gradient is likely not equal to the surface slope (subsurface flow), or pressure differences and inertial momentum cannot be neglected (channel and overland flow). In addition, while the kinematic wave equations are solved with a nonlinear scheme using Newton's method (Chow, Maidment, & Mays, 1988) , other model equations are solved through a simple explicit scheme. In summary the following limitations apply:

- Channel flow, and to a lesser degree overland flow, may be unrealistic in terrain that is not steep.
- The lateral movement of subsurface flow may be very wrong in terrain that is not steep.
- The simple numerical solution means that results from a daily timestep model may be different from those with an hourly timestep.

3.4.2 The model

The model plan describes which steps were taken during the model study with WFLOW. The model plan was designed on the basis of the Good Modelling Practice Handbook (van Waveren et al., 1999). The first step in modelling is to create a model journal. In this daily journal it is described what concrete actions were carried out, which choices were made and how reliable the end results are. The goal of this journal is to improve the reproducibility of the study. The model journal is shown in Appendix E

Set up model project

This step describes the boundaries of the project. Topics like the goal of the model, the different scenarios and used datasets are discussed. An overview of all input and output parameters is shown in Appendix D. A list containing definitions for all used parameters in the parametersets is included in Appendix D. Furthermore, a list describing the output parameters of this research is added to complete this overview.

Define the objective

The objective of the WFLOW study must be defined to ensure that this model is the right fit for the entire research.

- Study area: 30 catchments in Brazil. Their characteristics are shown in Table 3.
- Study period: 2000-2020.
- Used software: WFLOW_SBM 2020.1.1. and EwaterCycle
- Goal: To calculate/model daily discharge for all study areas during the study period.

Input

- Precipitation (Beck et al., 2019)
- Evaporation (Hersbach et al., 2020)
- Near-surface temperature (Hersbach et al., 2020)
- Landcover map (ESA2010 & UCLouvain, 2010)
- Digital elevation map (DEM) (NASA, 2014)
- Catchment delineation: Hydrosheds (Lehrer et al., 2008)
- Deforestation: Google Earth Engine deforestation map(Hansen et al., 2020)
- Soil map: Harmonized World Soil Database (HWSD) version 1.2 (FAO/IIASA/ISRIC/ISSCAS/JRC, 2013).
- Leaf area index (LAI): stored in directory of WFLOW (Copernicus, 2021)

Tools

The Dutch National supercomputer Snellius is used to run the model and calibration for our study area. This high performance computing system enables the user to produce results in a short time frame.

More information on Snellius is found at https://servicedesk.surfsara.nl/wiki/display/WIKI/Snellius.

Build model

The hydrological processes described in subsubsection 3.4.1 are used to build the model. Figure 17 shows the complete scheme for WFLOW. This scheme is a representation of how WFLOW deals internally with the hydrological processes and storages.

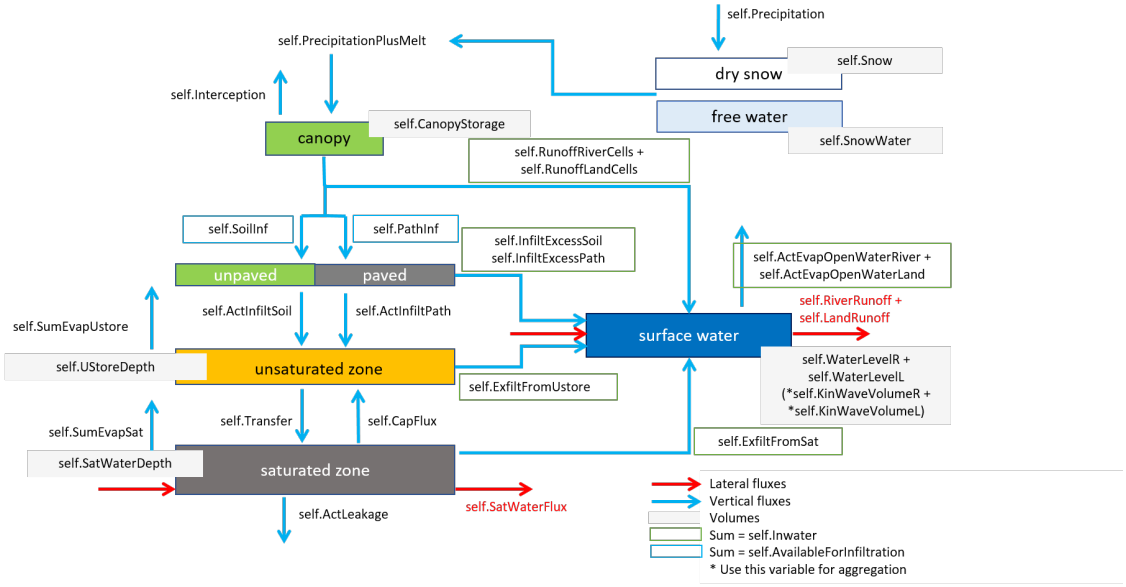

Figure 17: Complete WFLOW scheme (Deltares, 2021b)

Calibration

A model is calibrated to optimize the results. WFLOW offers several parameters for calibration:

- SoilThickness
- Soil parameter M
- Manning coefficient N
- KsatHorFrac

The definition of these calibration parameters is explained in Appendix D. This research will use KsatHorfrac as parameter for calibration. KsatHorfrac is a multiplication factor [-] applied to KsatVer for the horizontal saturated conductivity used for computing lateral subsurface flow. KsatVer is the vertical saturated conductivity [mm/day] at the surface. KsatHorfrac is used because it is the most suitable to optimize the signals observed in the default (run without calibration) output of this model. In addition, it is the most commonly used parameter for calibration in WFLOW. In general higher KsatHorFrac lead to smaller peaks and larger baseflow. The lateral ground flow becomes larger when KsatHorFrac becomes larger. More lateral groundflow means the ground will be less saturated and thus able to absorb more rain, resulting in lower discharge peaks. The lateral groundflow eventually contributes to the discharge of the river. Through this contribution, more lateral groundflow results in a larger baseflow.

This complex model uses two different objective functions for calibration. The Nash-Sutcliffe Efficiency (NSE) was explained earlier in subsection 3.3. The second objective function is the Kling-Gupta Efficiency (KGE). The KGE is derived from the constitutive components (mean bias, variability bias and correlation) of NSE (Gupta, Kling, Yilmaz, & Martinez, 2009). Like NSE, values typically range from 0 to 1 and larger values for KGE relate to a better model performance. Negative values are possible, but are regarded

as very poor performance. The general formula for Kling-Gupta efficiency is shown in Equation 10. The before mentioned constitutive components of KGE are represented by r, α and β in this equation. The variability and mean bias are determined through observed and simulated discharge, which is shown in Equation 11 and Equation 12.

$$KGE = 1 - \sqrt{(r-1)^2 + (\alpha-1)^2 + (\beta-1)^2} \tag{10}$$

KGE = Kling-Gupta Efficiency coefficient [-], r = linear correlation between observations and simulations [-], α= variability bias [-], β = mean bias [-]

$$\alpha = \frac{\sigma_{sim}}{\sigma_{obs}} \tag{11}$$

α = variability bias [-], σ_{sim} = standard deviation simulated discharge [m^3/s], σ_{obs}= standard deviation simulated discharge

$$\beta = \frac{\mu_{sim}}{\mu_{obs}} \tag{12}$$

β = mean bias [-], μ_{sim} = mean simulated discharge [m^3/s], μ_{obs}= mean simulated discharge

Output

The prime output from the complex model is discharge. This simulated discharge is used to analyze different aspects of the hydrologic response for each catchment. The time series of the runoff coefficient is used to observe the impact of deforestation on the percentage of rain that ends up in runoff. The method is identical to the method used in the study with observed discharge, which is explained in subsubsection 3.2.2. The time series of the recession curve coefficient is used to observe the impact of deforestation on the hydrologic response of the catchments. The method is identical to the method used in the CAMELS-BR study, which is explained in subsubsection 3.2.3.

In addition to discharge, the time series for several hydrological processes like overland flow can be extracted from WFLOW. These fluxes and storages are used to further analyze what is happening inside the model. A complete overview of of all input and output parameter is shown in

4

Results

This chapter shows the results of the analysis described in Chapter 3. The results are divided in three sub chapters similar to the three sub chapters described in Chapter 3.

4.1 Study with observed data

This sub-chapter shows the results of the study with observed data from CAMELS-BR. The measured discharge is used to determine the runoff coefficient and recession coefficient for discharge peaks.

Actual evapotranspiration

The relation between actual evapotranspiration and deforestation is used as indicator to determine if forests are replaced with different land cover types after deforestation. The method is described in subsubsection 3.2.1. Figure 18 shows the relation between actual evapotranspiration (AET) and the cumulative percentage of deforestation per year with a individual plot for each catchment. Two different data sources for AET are used; GLEAM and MGB. In Figure 18 the orange markers (green trendline) represent GLEAM and the blue markers (red trendline) respresent MGB. The majority of graphs within this figure show downward trends, indicating that the effects of deforestation on AET are not fully replaced by the evapotranspiration capabilities of land cover after deforestation. There are four catchments with upward trends, which means that the amount of actual evapotranspiration increases with deforestation for these areas. It is noticeable that these four catchment are subject to a relatively low percentage of deforestation. A possible explanation is that a catchment with less deforestation is more capable to replace the loss of evapotranspiration. Furthermore, these catchments wit low deforestation are relatively less impacted by effects of deforestation on evapotranspiration.

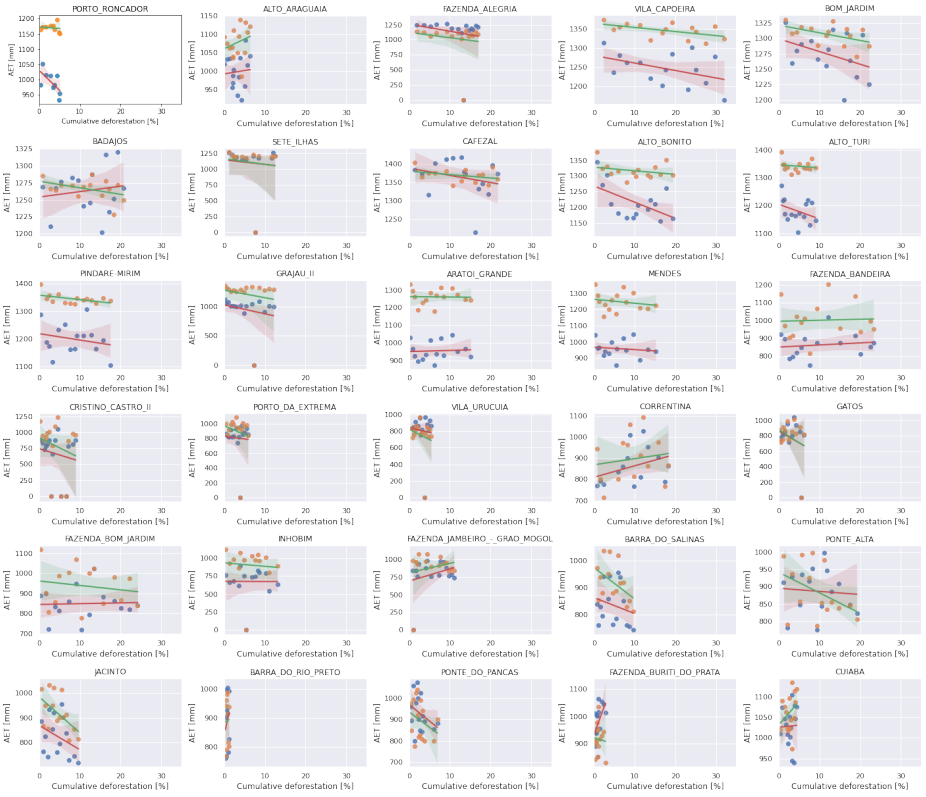

Figure 18: Relation between actual evapotranspiration (AET) and the cumulative percentage of deforestation per year with a individual plot for each catchment. Orange markers (green trendline) is obtained from GLEAM and blue markers (red trendline) is obtained from MGB.

Budyko framework

The Budyko framework describes the partitioning of the annual mean precipitation in evaporation and discharge. This partitioning is represented as a function between water supply and water demand. Figure 19 shows the Budyko framework for the observed-data study, for which the data is extracted from CAMELS-BR and its source is GLEAM (Martens et al., 2017). Each point in the framework represents a single year of a catchment in the study area. The color of a point matches with the degree of cumulative deforestation in that year.

The Amazon is a very humid region, which means values for PET/P are expected on the left side of the Budyko Framework. This is strengthened by Mianabadi, Coenders-Gerrits, Shirazi, and Ghahraman (2019) who showed that evergreen broadleaf forests (like the Amazon rainforest) have low PET/P values. Figure 19 shows the majority of points has PET/P < 1 as expected. The values of AET/P seem rather high for a humid area. High

values of AET/P indicate that a large part of precipitation is evaporated, resulting in an arid area.

Figure 19 shows several signals for deforestation. Firstly, it is observed that the majority of years with high degrees of deforestation are centered around the energy limit, which means that the actual evapotranspiration is similar to the potential evapotranspiration. Secondly, a cluster of high deforestation values is observed below 0.6 AET/P. The dominate presence of high deforestation values in the lower AET/P and PET/P areas of Figure 19 indicate that deforestation results in less potential and actual evapotranspiration. This observation does not match with conclusions from the majority of literature on Budyko Frameworks.

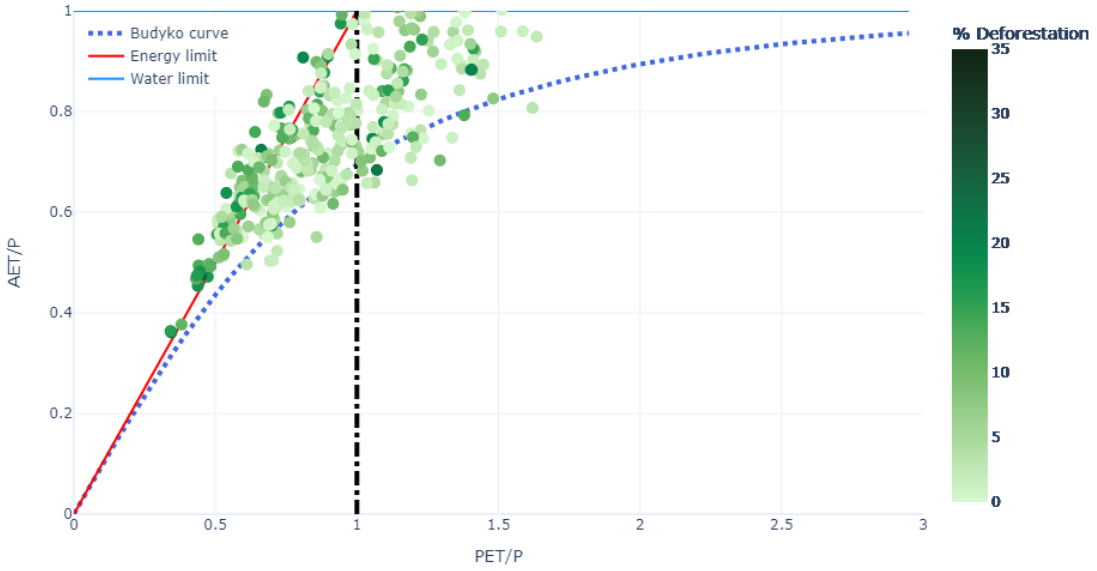

Figure 19: Budyko framework observed-data study

Runoff coefficient

The results for runoff coefficient are obtained with the method described in subsubsection 3.2.2. Figure 20 and Figure 21 show the relation between the runoff coefficient C_r and the cumulative percentage of deforestation. The cumulative percentage of deforestation in a certain year is defined as the total percentage deforestation of the previous years plus the deforestation in the corresponding year. Figure 20 shows the relation between runoff coefficient and cumulative deforestation for all catchments in one plot. This plot is used to analyze if a general relation is seen for the entire study area.

Several observations are extracted from Figure 20. Firstly, it is observed that the majority of the points are located between 0 - 0.4 C_r and 0 - 20 % deforestation. No clear correlation is found within this subset between C_r and deforestation. Secondly, the catchments with large cumulative deforestation ($> 30\%$) are considered as extreme. It is not possible to

determine a reliable correlation due to the low amount of data in this subset. However, it is noticed that this subset does not have any entries with a C_r smaller than 0.2. Thirdly, it is observed that high values for C_r (> 0.6) occur on the entire range of deforestation. In summary, Figure 20 does not show significant signals on the impact of deforestation on the runoff coefficient.

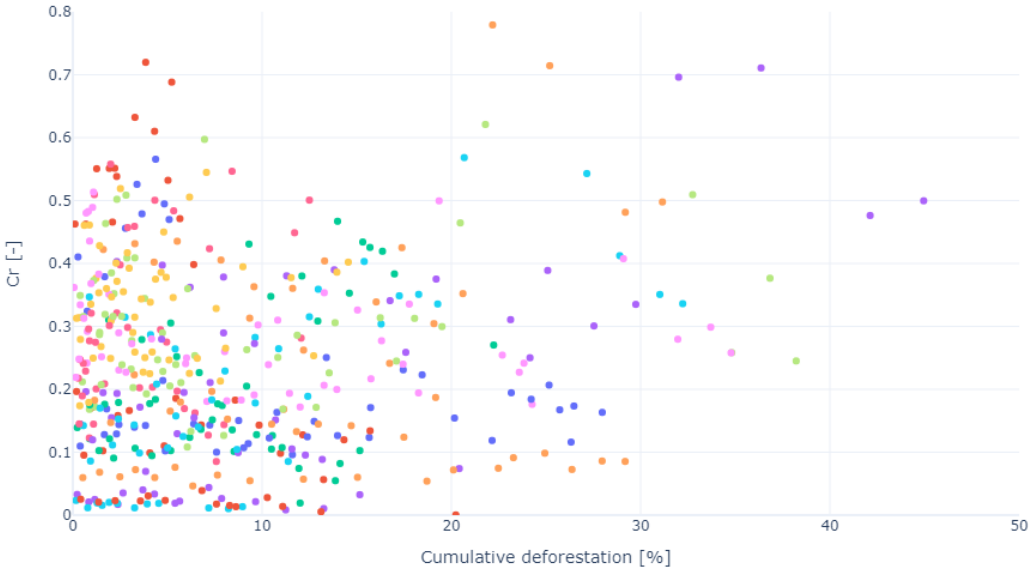

Figure 20: Relation between the runoff coefficient C_r and the cumulative percentage of deforestation per year for all catchments. Each color represents a catchment.

Figure 20 shows the relation between the runoff coefficient and cumulative deforestation for the entire study area. This relation is also determined per catchment to analyze if we observe different signals. Figure 21 shows the relation between deforestation and C_r per catchment and their 95% confidence interval. A larger interval means that the linear relation must be treated with more caution.

The graphs in Figure 21 show varying signals. Catchments like "Porto Roncador" and "Badajos" have an increasing trend. In these catchment the amount of rain that ends up in the river increases when the cumulative deforestation increases. Conversely, a downward trend is observed in catchments like "Porto da Extrema" and "Ponto da Pancas". Lastly, there are catchments like "Pindare-Mirim", that show minimal impact of deforestation on the runoff coefficient.

There seems to be a slight increasing trend among catchments with more than 25% cumulative deforestation in twenty years. There are 8 catchments with 25% or more cumulative deforestation, 6 of them show an increasing trend. However, some of these 6 catchments only slow a slight (almost negligible) upward trend. In addition, these 6 catchments also show the largest confidence interval indicating a less reliable result. In general it is observed that higher cumulative deforestation does not necessarily lead to higher C_r. This

is strengthened by increasing signals observed in catchments with less than 25% deforestation. For example, catchment "Porto Roncador" has a steep upward relation between deforestation and runoff coefficient, but total cumulative deforestation is less than 10 %. This indicates that increasing C_r also occurs in catchments with low deforestation.

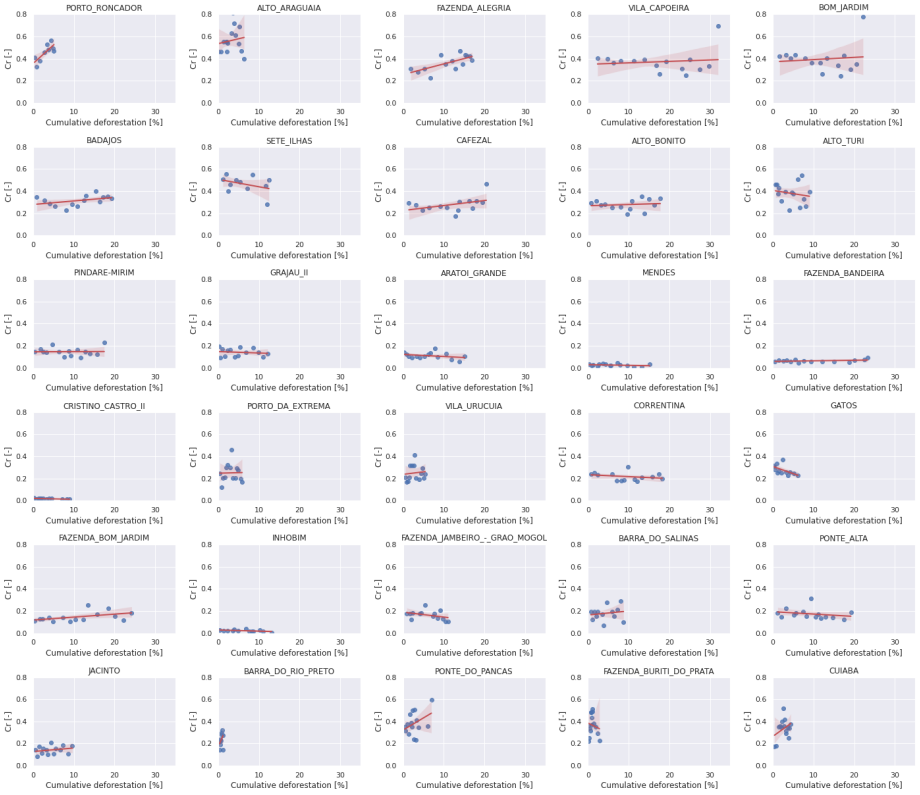

Figure 21: Relation between the runoff coefficient C_r and the cumulative percentage of deforestation per year with a individual plot for each catchment

Recession coefficient

The results in this sub-chapter are obtained with the method described in subsubsection 3.2.3. Figure 22 shows the relation between the recession coefficient k and the cumulative percentage of deforestation per year for all catchments. The recession coefficient is an indication for the rate of recession of a discharge peak. In other words, how fast discharge will decline from peak-flow (top discharge peak) to base flow. Higher values of k results in faster draining of water from the basin, whereas a lower value of k represents slower drainage from the basin.

Several observations are extracted from Figure 22. Firstly, it is observed that the majority of points are located between 0 - 0.05 k and 0 - 20 % deforestation. Within this subset no

correlation is observed between k and deforestation. Secondly, the subset with catchments having large cumulative deforestation ($> 30\%$) show very stable behaviour. It is noticeable that the values for k in this subset have no real outliers and remain below k = 0.04.

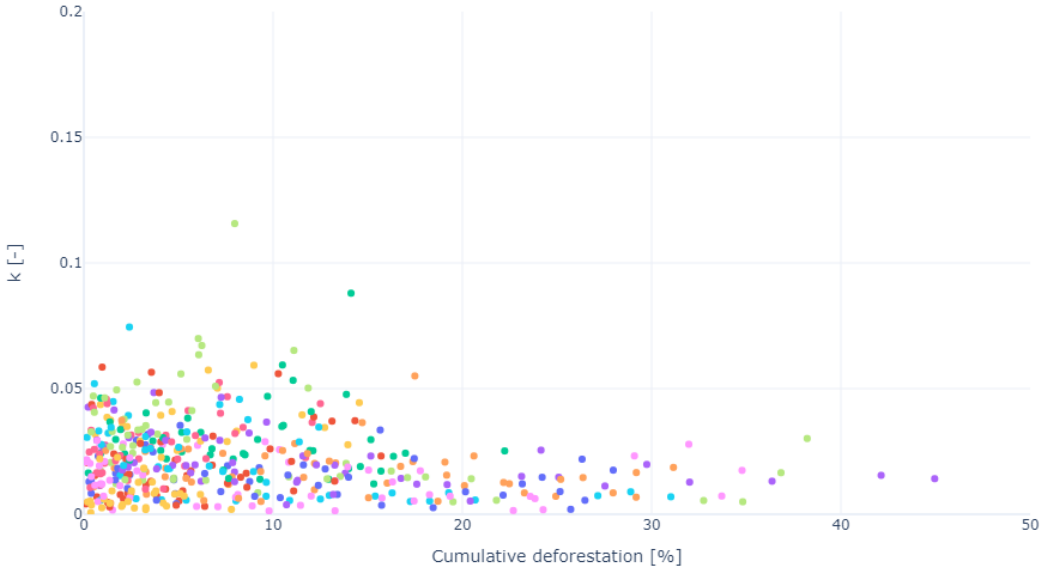

Figure 22: Relation between the recession coefficient k and the cumulative percentage of deforestation per year for all catchments. Each color represents a catchment.

Figure 22 showed the relation between the runoff coefficient and cumulative deforestation for the entire study area. This relation is also determined per catchment to analyze if we observe different signals. Figure 23 shows the relation between deforestation and k per catchment and the 95% confidence interval.

The graphs in Figure 23 show varying signals (upward, downward, stable). The majority of catchments show minimal impact of deforestation on the recession coefficient, resulting in a relatively flat line. This flat behaviour is seen in catchments like "Vila Capoeira" and "Bom Jardim". There are also catchments like "Vila Urucuia" and "Ponto de Pancas" with sharp increasing trends. An increasing trend indicates deforestation results in faster draining of the basin. Lastly, we observe several catchment with declining trends. Analyzing the different signals per catchments leads to the observation that there is no general relation between cumulative deforestation and the recession coefficient for this study area and this set of data.

In addition to the type of signal, it is also analyzed which signals are observed in a certain subset of catchments. Catchments with heavy deforestation ($> 25\%$) hardly show any up or downward trend. This does not mean that heavy deforestation automatically results in no impact on recession coefficients. Catchments with less deforestation have a more varying subset of signals. Especially in the subset of catchment with minor deforestation

(> 10%) steep upward and downward trends are observed. No general relation is observed for catchments with minor deforestation due to the spread of signals in this subset.

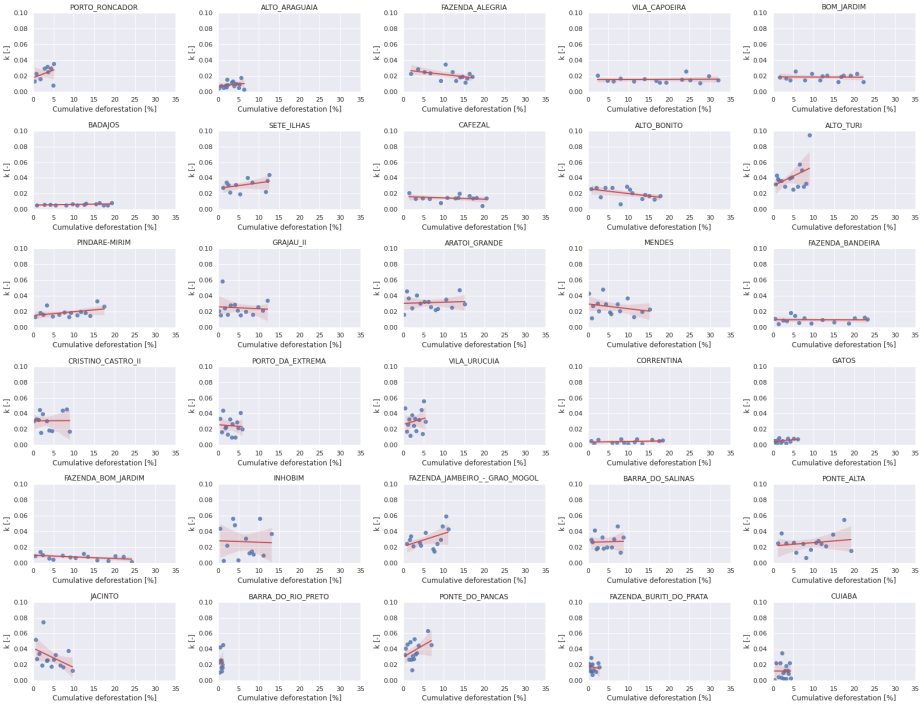

Figure 23: Relation between the recession coefficient k and the cumulative percentage of deforestation for all catchments.

In summary, there is no general relation observed between cumulative deforestation and the discharge coefficients for this study area and data. The variability of signals (upward, downward or stable) within coefficients is high for the entire range of deforestation. In addition, an analysis on the relation between yearly non-cumulative deforestation and the discharge coefficients (C_r and k) is conducted. The results of this additional analysis are shown in Appendix A. In short, these results show similar signals as cumulative deforestation in terms of trends (upward, downward or stable). No clear relation is observed between non-cumulative deforestation and the discharge coefficients for this study area and data.

4.2 Study with simple model

This sub-chapter shows the results of the model study obtained through the self-made "simple model". As described in subsection 3.3 the first method for calibration is "trial and error". This method was used on one catchment to test the simple model. The results of this manual calibration is shown in Figure 24. This figure shows the simulated discharge

with several values of parameter α. Parameter α represents the relation between water in storage and outflow (discharge). The entire volume of water in storage is turned into discharge when $\alpha = 1$. The graph with $\alpha = 1$ shows that this value does not perform well for our model. Figure 24 shows that decreasing α increases the performance of the model. A value of $\alpha = 1.5$ means the simulated output (discharge) is larger than the amount of water in storage. The only plausible cause for this situation is extra input of water (not precipitation), which is directly transferred to output. Trial and error calibration led to $\alpha_{obs} = 0.03$ for fitting the output of the model with the observed discharge from CAMELS-BR. Fitting the output of the model with the simulated discharge from WFLOW resulted in $\alpha_{sim} = 0.3$. As result of this trial and error method a large difference is observed between α_{obs} and α_{sim}. While being a relatively simple model, this difference in α indicates that the relation between storage and discharge is not optimal in WFLOW for our study area.

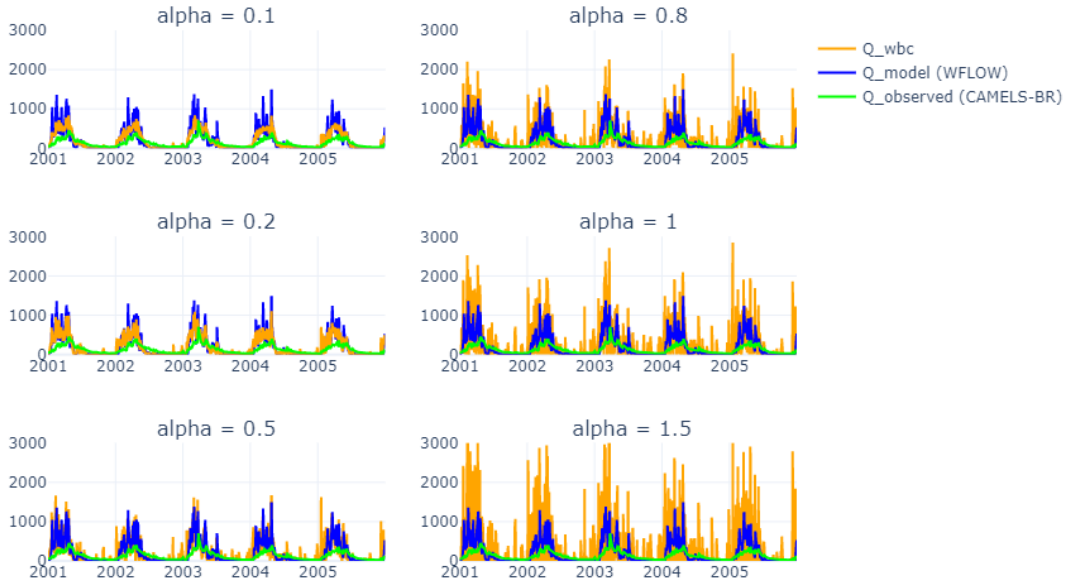

Figure 24: Results "Trial and error" calibration for catchment "Bom Jardim"

In addition to the "trial and error" method an objective calibration with Nash-Sutcliffe Efficiency (NSE) is conducted. The goal of this objective calibration is to find the optimal value α per catchment. The discharge calculated with the optimized α is shown in Figure 25. The Nash-Sutcliffe Efficiency for "Bom Jardim" is 0.79 and 0.56 for "Cuiaba Paraguay". These values are decent for a simple model like this and show that its performance is adequate. From Figure 25 it is observed that the model lacks performance during base flow. This behaviour is caused by the inability of the model to represent slow flows like groundwater. Consecutive days without precipitation lead to an empty storage and therefore absence of discharge.

From the optimized output in Figure 25 it is observed that α for "Bom Jardim is 0.0208

and 0.0253 for "Cuiaba Paraguay". The value for α is also determined per year to analyze the relation between α and deforestation. This relation is shown in Figure 26. From this figure it is observed that the value of α differs per year, but does not show a clear increasing or decreasing trend. A second notable observation is the zero values seen for several years in "Cuiaba Paraguay". A zero value means that the optimized objective function NSE is below 0.3 in that year. The value of 0.3 was chosen as threshold to exclude years with very poor performance.

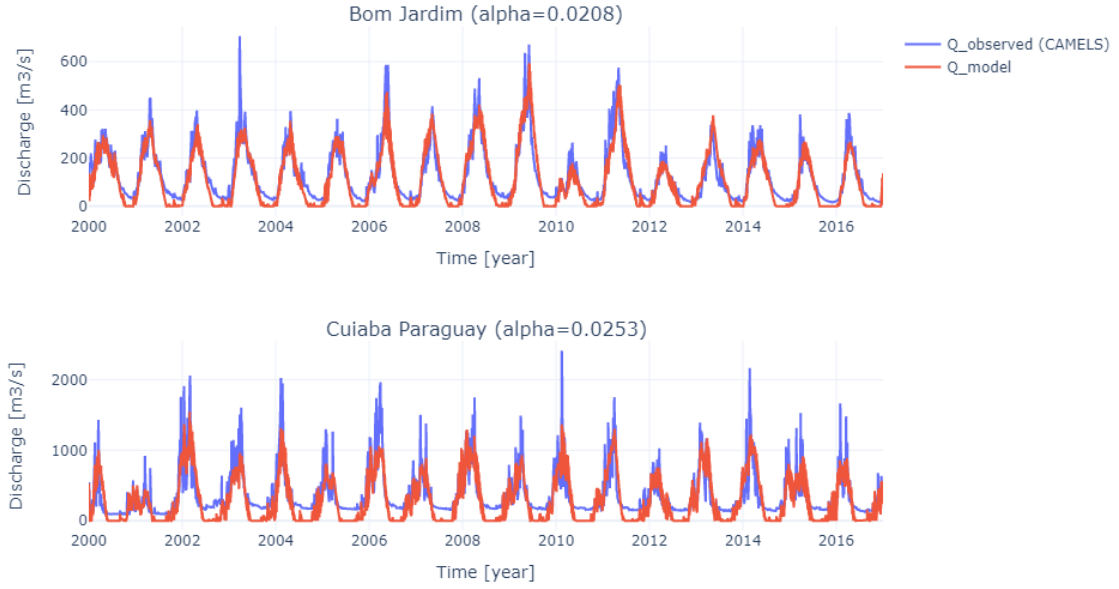

Figure 25: Discharge for optimized parameter α. NSE "Bom Jardim" = 0.79 and NSE "Cuiaba Paraguay" = 0.56.

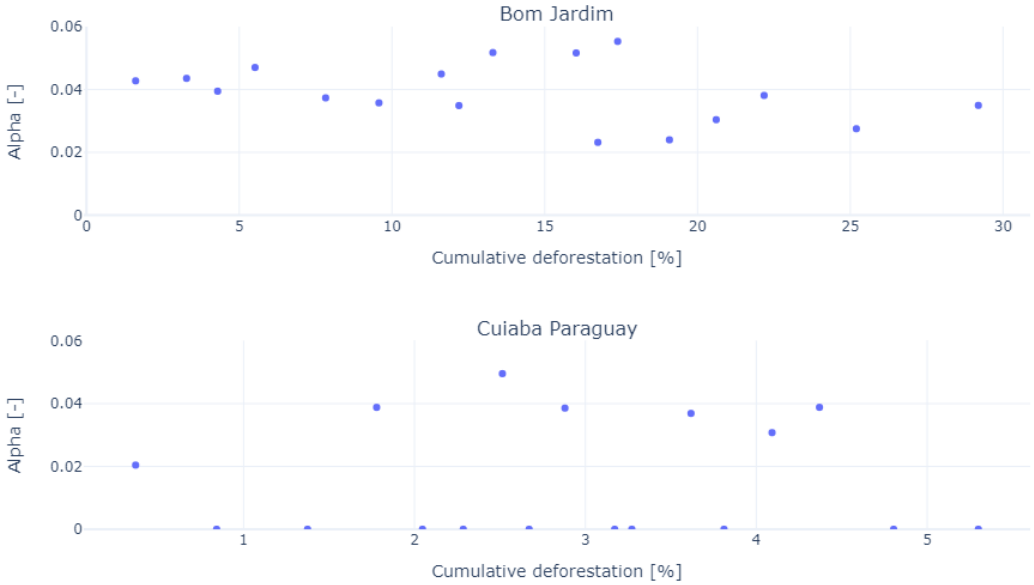

Figure 26: Relation between yearly α and cumulative deforestation

4.3 Study with complex model

This sub-chapter shows the results of the model study obtained through WFLOW. The main output of WFLOW is discharge, but a complete overview of output parameters is shown in Appendix D. Figure 27 shows the tim eseries of simulated discharge and observed discharge for the catchments "Bom Jardim" and "Cuiaba" after calibration. Other catchments in the study area show a similar performance as these two catchments. The KGE and NSE of theese catchments are included in Appendix C. It is observed through values of KGE and NSE that WFLOW does not perform well for the study area. This sub-chapter looks deeper into what is happening in the model and tries to answer why WFLOW does not perform as expected. The code for this analysis can be found on de Wildt (2022).

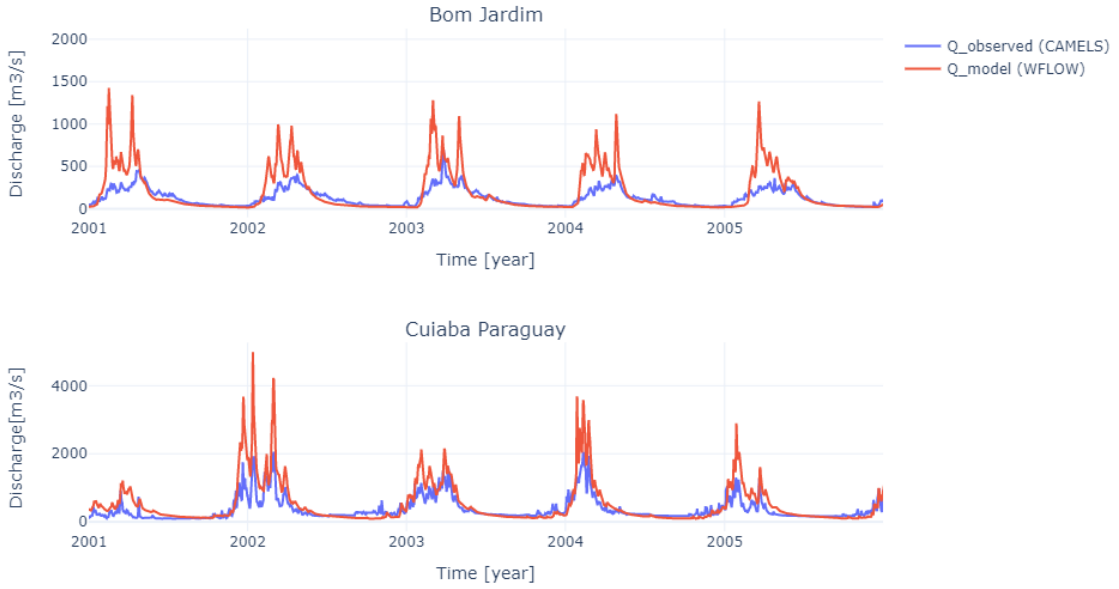

Figure 27: Timeseries of observed and callibrated simulated discharge

Initially, a period of 20 years was used to run the model. However, the output for discharge, fluxes and storages show similar behaviour in each year. Therefore the decision is made to only analyze the first five years to increase time efficiency. These five years are preceded by a so-called "warm-up" year to ensure the output is not affected by initial conditions. Furthermore, it is acknowledged that a period of five years does not fully represent the impact of long-term fluxes like groundwater flow. However the goal of this sub-chapter is to find an explanation for the errors observed in Figure 27. A time period of five years with daily output is deemed to be sufficient for this goal. The catchments "Bom Jardim" and "Cuiaba" are used due to their diverse characteristics, which is shown in Table 3. This diversity contributes to the objectivity of this research.

Calibration

The model is callibrated with parameter "KsatHorFrac". This parameter is defined as: "A multiplication factor [-] applied to KsatVer for the horizontal saturated conductivity used for computing lateral subsurface flow." (Deltares, 2021b). Figure 28 shows a selection of discharge output generated by WFLOW with different values for KsatHorFrac. In general higher KsatHorFrac lead to smaller peaks and larger baseflow. The lateral ground flow becomes larger when KsatHorFrac becomes larger. More lateral groundflow means the total depth of the soil will be less saturated and thus able to absorb more rain, resulting in lower discharge peaks. The lateral groundflow eventually contributes to the discharge of the river. Through this contribution, more lateral groundflow results in a larger baseflow. A complete overview of the calibration results is shown in Table 4. This table shows the value for the Kling Gupta Efficiency (KGE) per value of "KsatHorfrac" for each catchment. The optimal KGE of a catchment is shown in green.

With a KGE of 0.844, the best fit of all catchments is observed for Cuiaba Paraguay. With the exception of this catchment, none of the values of KsatHorFrac observed in Figure 28 and Appendix B result in decent discharge compared to the simple model.

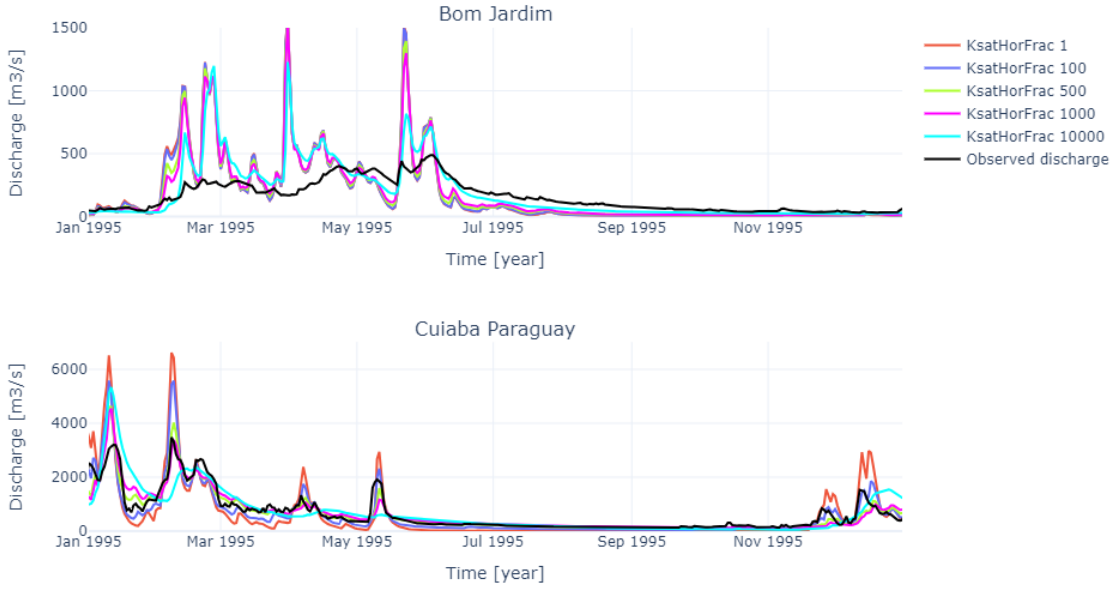

Figure 28: Timeseries of discharge with different values of KsatHorFrac

Water balance

In general, a check on the water balance is among the first steps in analyzing the performance of a model. The water balance is commonly described by Equation 13. This equation results from the definition of the water balance: The inflows to any water system or area is equal to its outflows plus change in storage during a time interval (Viessmann & Lewis, 1996). In a perfect situation the water balance of a model is closed, but some noise can be expected when using multiple sources of input. A closed water balance means that all water in the model is accounted for, which is shown in Equation 13.

$$\varepsilon = P - Q - ET - \Delta S \tag{13}$$

P = Precipitation [mm], Q = Discharge [mm], ET = Evapotranspiration [mm], ΔS = Change in storage [mm], ε = Error[mm].

Figure 29 shows the time series of the water balance. A value of zero means that the water balance is closed. Any other value indicates a surplus or deficit of the parameters in Equation 13. The values on the y-axis of Figure 29 are denoted in μ, which is equal to 10^{-6}. It is observed that the water balance remains close to zero for the entire time series and deviations are very small. This means the water balance of WFLOW is considered closed for these two catchments. In conclusion, not closing of the water balance is not the source for the differences between modelled and observed discharge in this research.

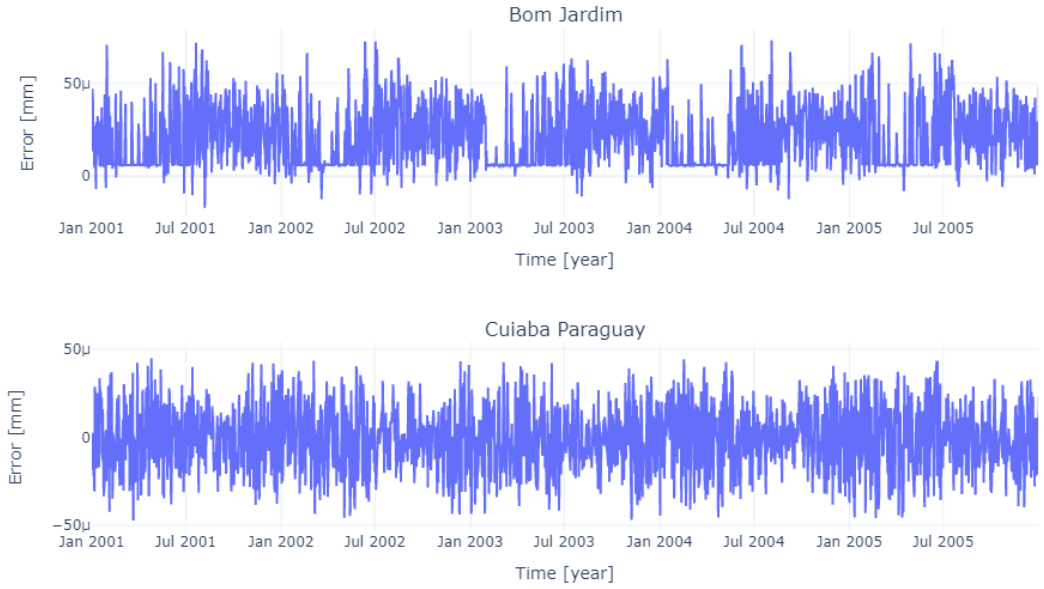

Figure 29: Time series water balance

Input WFLOW

Figure 30 shows the Budyko framework for CAMELS-BR (orange markers) and WFLOW (green markers). The data for WFLOW is extracted from ERA5 (Hersbach et al., 2020). Each point in the framework represents a single year of a catchment in the study area. The color of a point matches with the degree of cumulative deforestation in that year. As explained in subsection 4.1 values in the Budyko framework for humid areas are expected to be located on the left side with PET/P <1. Green markers in Figure 30 all have PET/P >1, which matches with arid areas. This is not compliant with literature about Budyko frameworks of humid areas like in this research. Figure 30 shows that data from CAMELS-BR (orange) is closer to expected values for this study area than WFLOW (green).

Figure 30 shows several signals for deforestation. First, it is observed that PET/P is larger than 1 for all points in the framework. This means that stream flow behaviour is not limited by energy, but by the amount of available water (precipitation). This is strengthened by the high values for AET/P, which indicates that a large part of precipitation is evaporated. Deforestation should result in higher values of PET/P and AET/P, which does not match with the observed values in Figure 30. A second signal is observed by analyzing the degree of deforestation in Figure 30. Points with low deforestation ($< 5\%$) have higher AET/P values in comparison with higher degrees of deforestation.

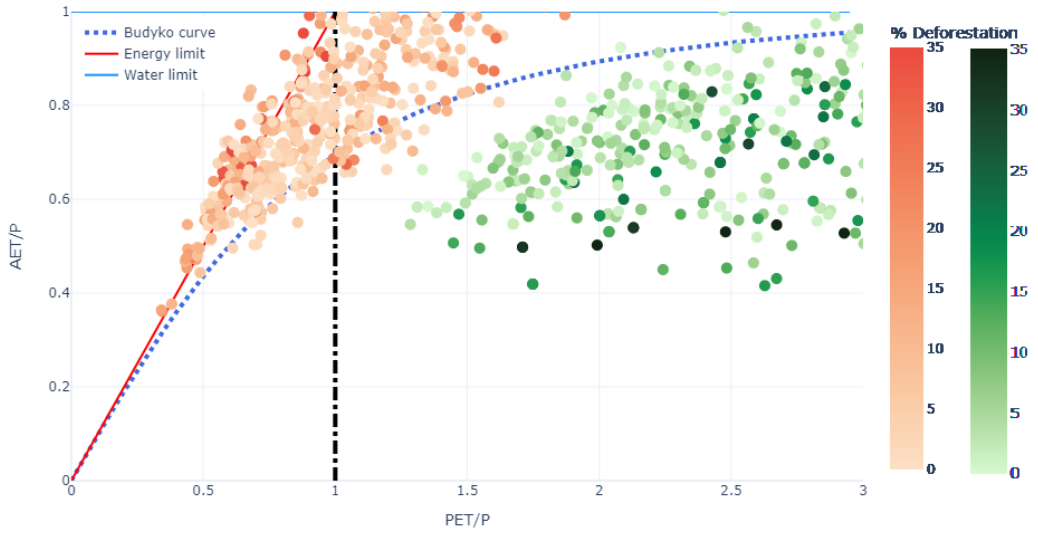

Figure 30: Budyko framework for CAMELS-BR and ERA5. The input data for CAMELS-BR (GLEAM) is orange. The input data for WFLOW (ERA5) is green.

Figure 31 shows the Budyko framework for catchment "Bom Jardim" (top) and "Cuiaba Paraguay" (bottom). The figures consist of data from CAMELS-BR (orange) and data used in WFLOW (green). The utility of Figure 31 is to observe signals for deforestation within a catchment. The spread of the markers for "Bom Jardim" does not show any signals for deforestation. An increase of AET/P and PET/P is expected with increasing deforestation, however this is not clearly shown in the data. A similar observation is made for "Cuiaba Paraguay". Markers with high deforestation have lower PET/P, which does not match with the hypothesis of increasing PET/P due to increasing deforestation.

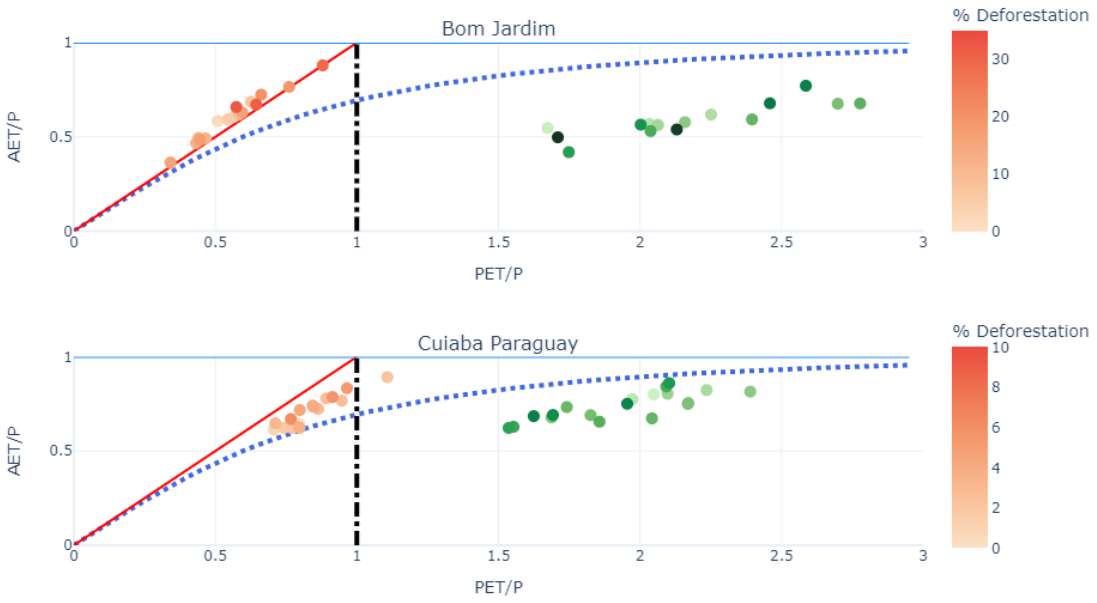

Figure 31: Budyko framework for catchment "Bom Jardim" (top) and "Cuiaba Paraguay" (bottom). Input data for CAMELS-BR (GLEAM) is orange. Input data for WFLOW (ERA5) is green.

The study with observed data and the complex model study are not able to use the same input, as was described in subsection 3.1. WFLOW requires distributed input for precipitation, potential evapotranspiration and near-surface temperature. Significant dissimilarities between the input of both studies are a plausible cause for dissimilarities in discharge of both studies. Figure 32 is a plot of the time series of precipitation for CAMELS-BR and WFLOW. It is observed that there are some dissimilarities between the data sets, but in general they are considered similar. The main observed difference is that some peaks from WFLOW are higher than CAMELS-BR. Figure 27 shows that the simulated discharge of WFLOW is higher during peak flow than the observed discharge. Incidental lower peak precipitation is not considered as a probable cause for higher discharge in WFLOW.

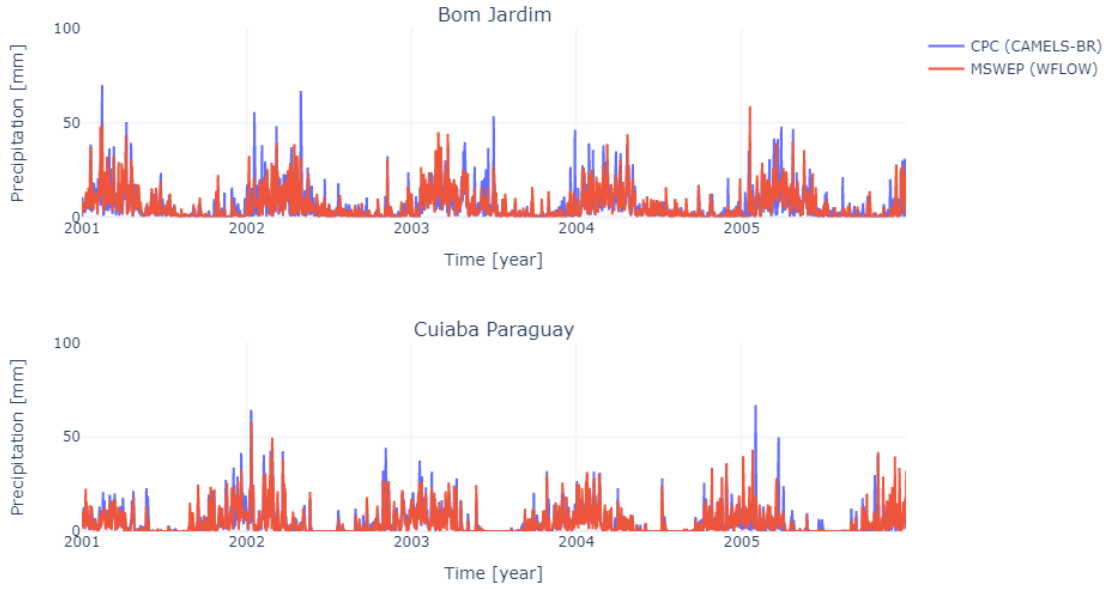

Figure 32: Timeseries of precipitation

An extra check is performed by analyzing the cumulative yearly precipitation for the entire time scope. The bar graph is shown in Figure 33. This bar graph shows that there are no significant differences between cumulative precipitation, with exception of the last three years in "Bom Jardim". The overestimation for discharge is already present during the years with similar cumulative precipitation. In conclusion, it is established that the input for precipitation is not a likely cause for the dissimilarities between modelled and observed discharge.

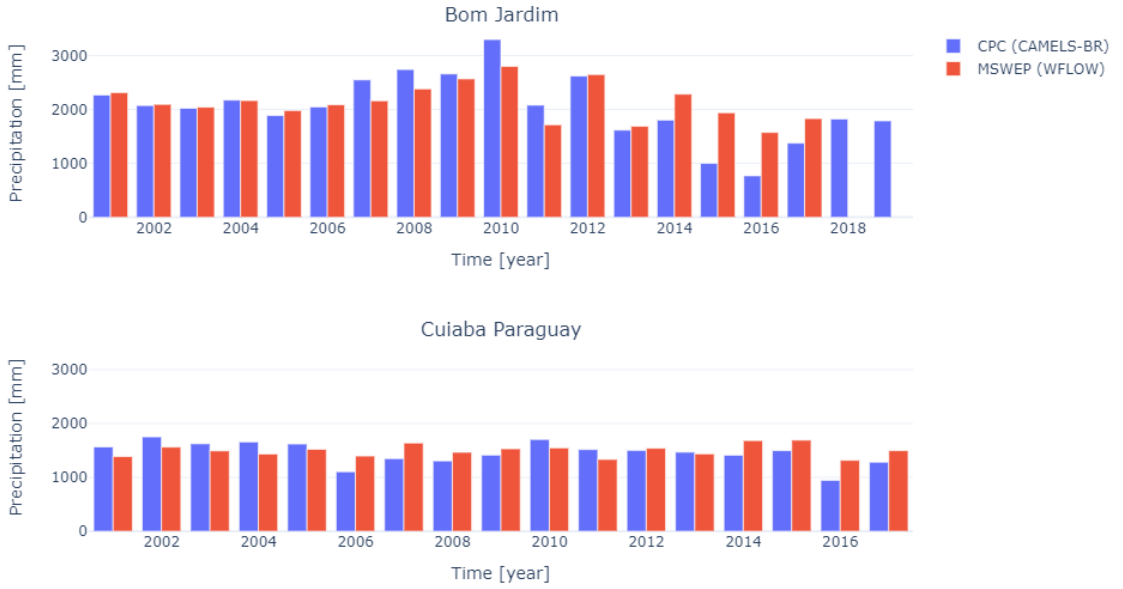

Figure 33: Bar chart of cumulative yearly precipitation

The near-surface temperature is used to determine fluxes like soil evaporation. Figure 34 shows the time series of the mean daily near-surface temperature from ERA5 used in WFLOW and obtained by CAMELS-BR. The temperature of both data sets show very similar behaviour. We can therefore conclude that it is not likely that input for temperature is causing the poor performance of the model.

Figure 34: Timeseries of mean daily temperature

An important factor of the water balance is the flux potential evapotranspiration (PET). From Equation 13 we can extract that a wrong estimation of PET must lead to a wrong estimation of discharge or storage. An underestimation of PET results in more water and an overestimation leads to less water in our system. The timing of PET is equally important to the reliability of a data set as the value of the estimation. For example, high temperatures during summer result in a larger PET than low temperatures during winter. Figure 35 shows time series of PET used by CAMELS-BR and WFLOW for both catchments. The first thing to notice is that the two series are not in sync with each other. This time lag observed between the series is alarming. If the input in WFLOW for potential evapotranspiration has a time lag in comparison with measured data, then it is a plausible explanation for the current performance of the model.

The two data sets with potential evapotranspiration are generated by a different model per study. The PET in CAMELS-BR is simulated by GLEAM and the PET used as input in WFLOW is simulated by MSWEP. The PET value per day for the entire catchment is determined by taking the mean value of all values (latitude , longitude) located in that catchment. Large daily deviations within a catchment can result in a mean value that does not match with evapotranspiration from CAMELS-BR. The daily standard deviation is calculated to rule out that the time lag is caused by mean values. This standard deviation is shown in Figure 36. The majority of standard deviation is less than 0.2 mm. It is observed in Figure 35 that the mean PET of MSWEP is largely located between 2 and 4 mm. Taking the 0.2 mm standard deviation into account, it is noticed that the maximum standard deviation per day is around 10% and for most days it is less. It can be argued that this percentage is not negligible, especially for research on deforestation. However, this standard deviation is not large enough to be the cause of the results shown in Figure 27.

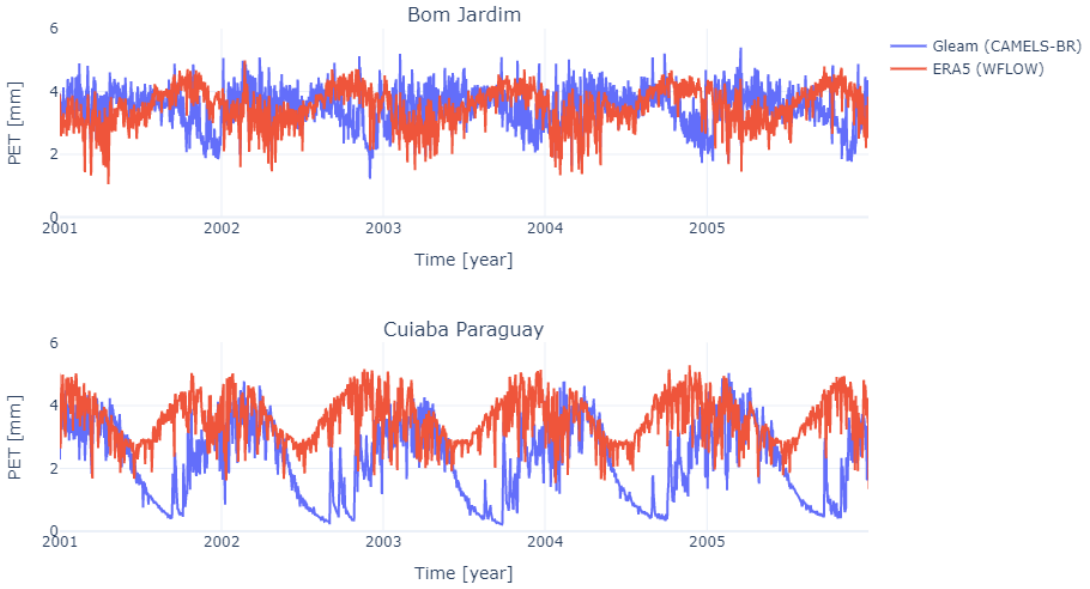

Figure 35: Timeseries potential evapotranspiration

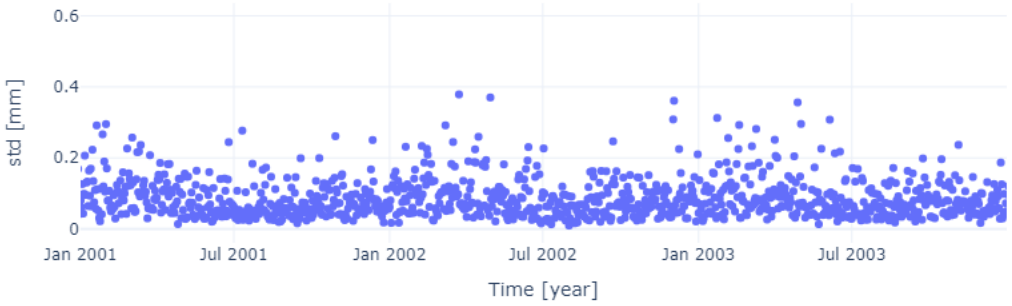

Figure 36: Timeseries standard deviation of daily potential evapotranspiration

There is a relation between evapotranspiration and temperature. An increase in temperature results in increased rate of evapotranspiration. A higher temperature means that there is more energy available to convert water to vapor, resulting in a higher amount of evaporation. In addition, the transpiration increases with higher temperature because plants open their stomata, thereby releasing more water vapor (NCCO, 2022). CAMELS-BR and MSWEP both use measurement for temperature. Potential evaporation is rarely equal to actual evaporation. However, the temperature in our study area is significant and there is plenty of water available. Therefore daily temperature should coincide with

measurements of daily PET in terms of timing.

Figure 37 shows time series of PET and temperature from CAMELS-BR and MSWEP. The potential evapotranspiration from MSWEP coincides with the temperature of both MSWEP and CAMELS-BR. Oddly enough the PET used by CAMELS-BR is not following the trend of the temperature used by CAMELS-BR. Due to the fact that the temperature of both data sets coincide with PET from MSWEP, it is concluded that the values for PET by MSWEP are correct. Therefore the potential evapotranspiration used as input in the model is not the cause of the poor model performance.

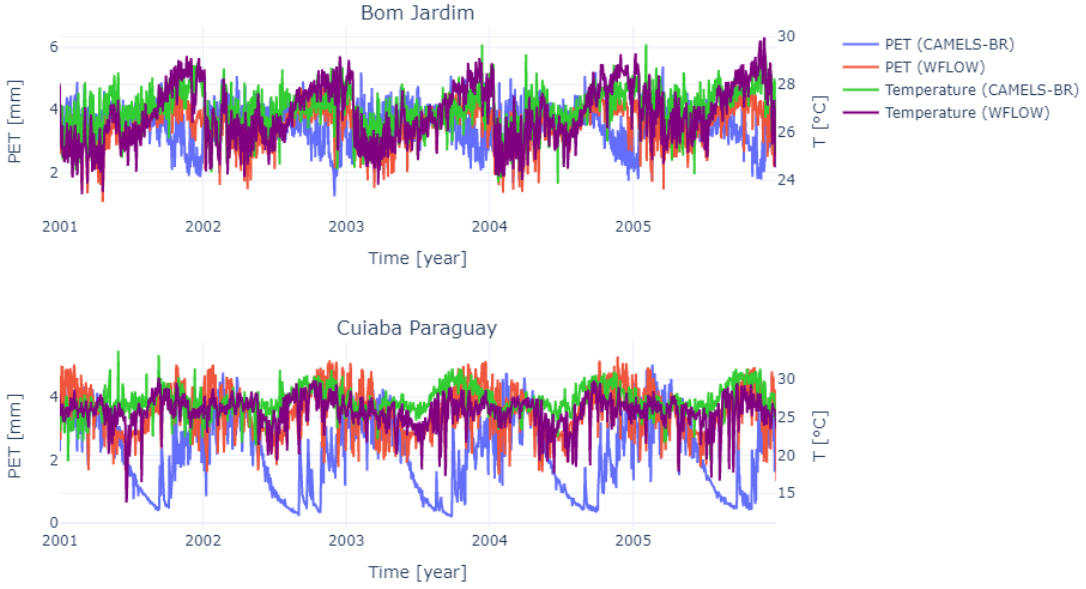

Figure 37: Timeseries potential evapotranspiration and temperature from CAMELS-BR and input WFLOW

Canopy storage

So far it is established that input for precipitation, potential evapotranspiration and near surface temperature are not causing the poor performance of the model. The next step is to analyze what is happening with hydrological parameters inside the model.

As explained in section 1 the canopy of a forest has several abilities. One of these abilities is the ability to store water, in particular rainfall. In WFLOW, the canopy storage is determined by several fluxes like evaporation. More information on the exact method WFLOW uses to calculate canopy storage is shown in the documentation for WFLOW by Deltares (2021b). Figure 38 shows the time series of canopy storage in mm. The value for canopy storage is defined as the available amount of storage that can be used to store precipitation in the present canopy. In other words, a canopy storage with a value of zero is equal to a full canopy storage. Figure 38 shows that the canopy storage in both catchments is always full, which can be explained by two different possibilities. The first possibility

is trivial; the canopy storage really is always full. This possibility is not unimaginable for an area like the Amazon with this amount precipitation. However, it seems unlikely that the daily storage is constantly full for a period of twenty years for 30 different catchments. This observation raises questions about how canopy storage is determined.

The second possibility is that the model does not take canopy into account. If the model indeed neglects canopy storage it could be a possible explanation for the deviation between observed and simulated discharge. However, canopy storage is defined during the setup of the model. This makes it difficult to check if the model really takes canopy storage into account or is somehow neglected in the model process.

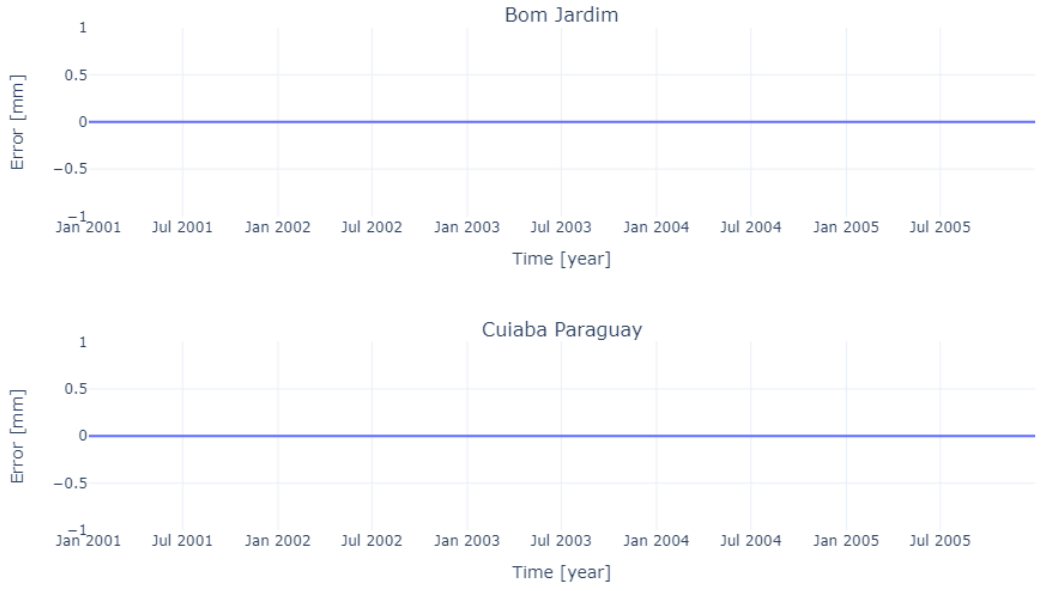

Figure 38: Timeseries canopy storage

Unsaturated zone

Figure 39 shows the relation between simulated discharge and the depth of the unsaturated zone simulated by the model. An unsaturated zone depth of zero means that the entire soil is saturated. Furthermore, an analysis on the relation between precipitation and unsaturated zone is conducted. This relation with precipitation is shown in Figure 39

Figure 39: Timeseries depth unsaturated zone and simulated overland flow

A clear signal obtained from Figure 39 is that discharge peaks occur when the depth of the unsaturated zone is close to zero. This signal makes sense as runoff reaches the river quicker when the ground is (almost) fully saturated. Runoff reaches the river quicker because there is less "damping" due to the soil. In addition, more rainfall ends up as overland flow when the ground is (almost) fully saturated due to less infiltration in comparison with an unsaturated soil. Figure 40 shows that the soil becomes more saturated (decreasing depth unsaturated zone) during precipitation events.

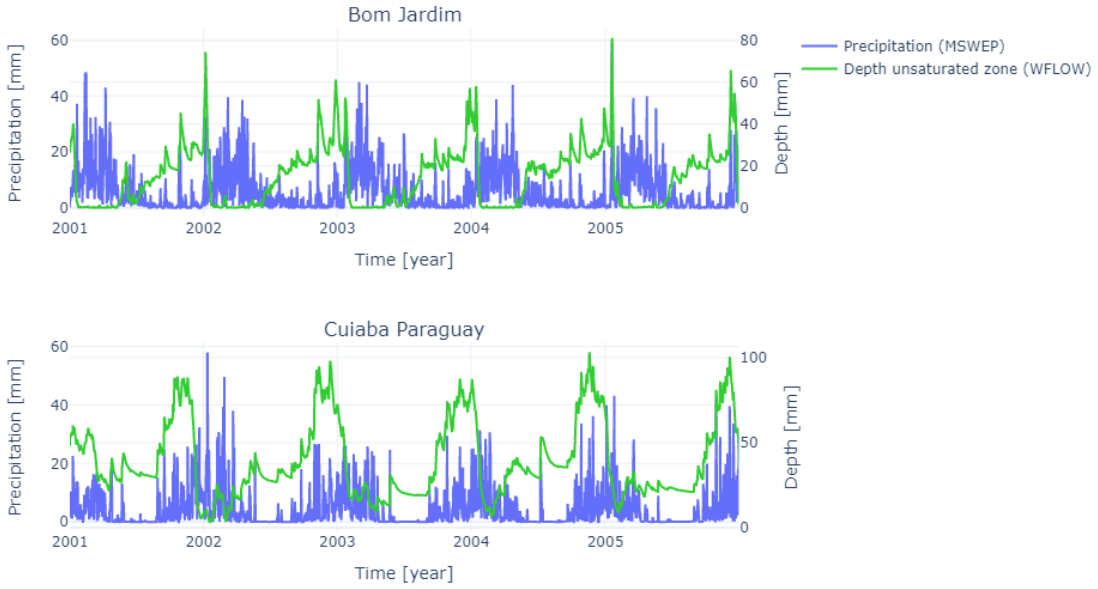

Figure 40: Timeseries depth unsaturated zone and precipitation

The above paragraph shows that the discharge peaks occur when the soil is saturated. Therefore, the focus of the next part of the analysis is on the unsaturated zone. Figure 41 shows the depth of the unsaturated zone, the daily infiltration and the daily excess infiltration. Excess infiltration occurs when the soil is saturated or the infiltration capacity is reached. Signals in Figure 41 are not the same for both catchments. In "Bom Jardim" the depth of the unsaturated zone becomes zero during precipitation events resulting in a large amount of excess infiltration. It is noticeable that the unsaturated zone fills in rapid pace. This can be caused by several reasons like excessive precipitation or a underestimation of the soil depth in WFLOW. In "Cuiaba Paraguay" excess infiltration starts occurring when the depth of the unsaturated zone starts decreasing. This is an indication that the excess infiltration is due to reaching the infiltration capacity of the soil. During heavy precipitation events the amount of rain is larger than the amount of water able to infiltrate the soil, which results in excess infiltration.

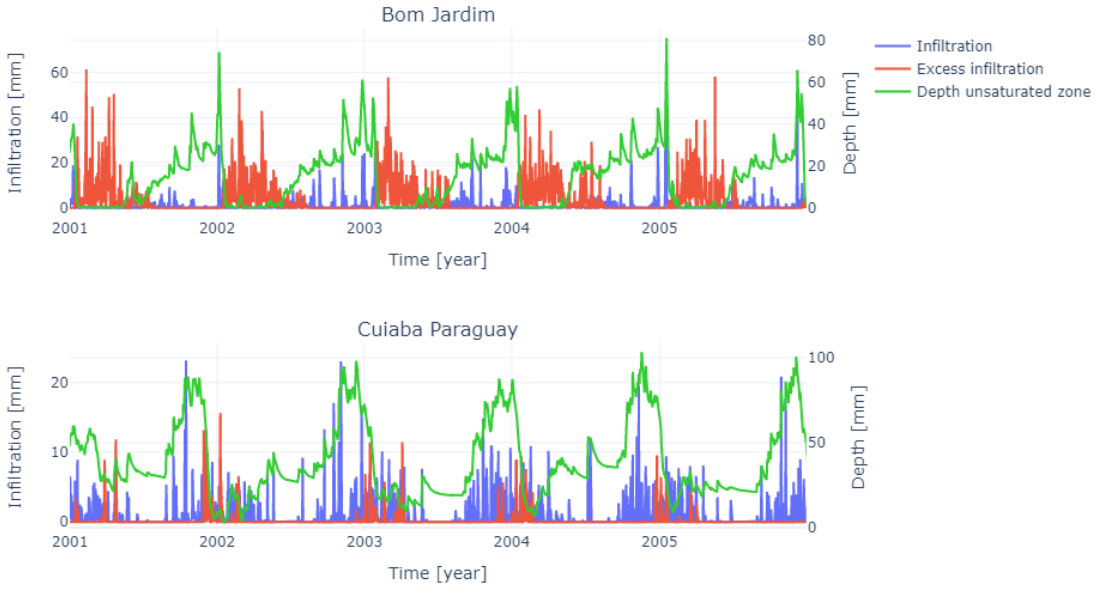

Figure 41: Timeseries depth unsaturated zone, infiltration and excess infiltration

It is observed that the overestimation of discharge by WFLOW is caused by excess infiltration. However, the origin of this excess infiltration is not quite clear. Before looking into infiltration capacity and soil depth as possible reasons for the overestimation of discharge, the saturated zone is analyzed.

Saturated zone

Figure 42 shows the depth of the saturated zone and discharge simulated by WFLOW. A zero value means that the soil is completely unsaturated.

Figure 42 shows the same signal for both catchments: Discharge peaks occur when the saturated zone depth is increasing. It is logical that the depth of the saturation zone increases during discharge peaks as this coincides with precipitation events. More precipitation leads to more water infiltrating the soil and therefore a more saturated soil. This is the same kind of behaviour that was earlier observed in Figure 39, where it is observed that discharge peaks occur when the unsaturated zone decreases.

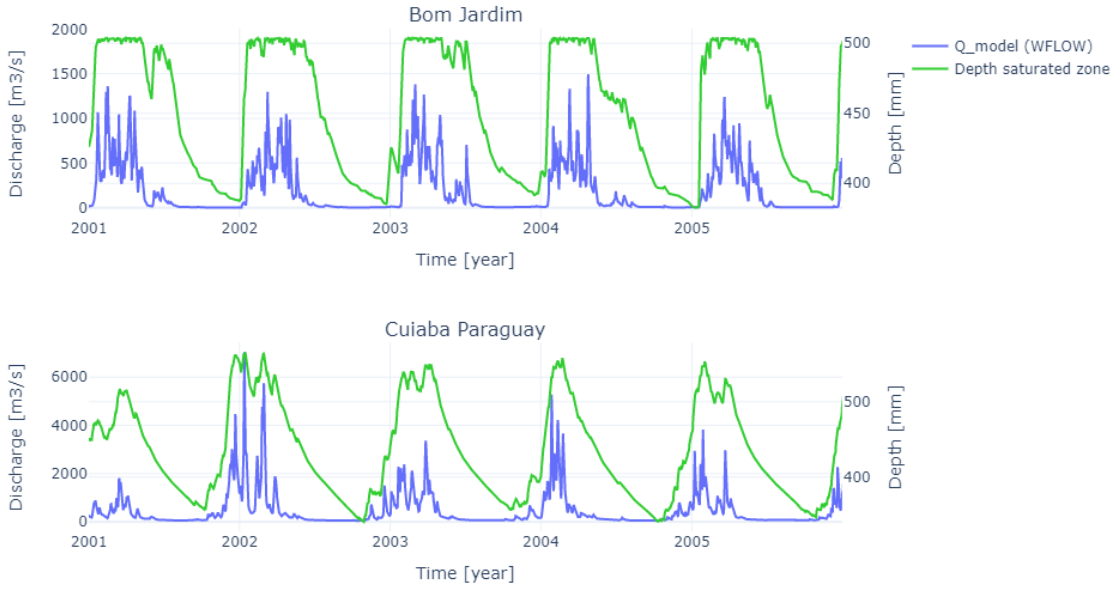

Figure 11: Saturated zone

Figure 42: Timeseries depth saturated zone and discharge

Changing infiltration capacity

The previous parts of the analysis on WFLOW indicate that excess infiltration is a large factor in the overestimation of discharge during peak flow. A logical follow-up is to research if the set value for maximum infiltration capacity is causing the excess infiltration. The default value for infiltration capacity for both catchments is 600 mm/day. Figure 43 shows the simulated discharge for several values of infiltration capacity. Individual graphs per value of infiltration capacity are shown in Appendix C for both catchments.

The first thing noticed in Figure 43 is that enlarging the infiltration capacity has no impact on the simulated discharge. The purple (infcap = 60.000 mm) and green line (infcap = 6000 mm) are exactly the same as the blue line (infcap = 600 mm). This means that the default infiltration capacity is sufficiently large. No excess infiltration is created due to reaching the maximum infiltration capacity. For smaller values of infiltration capacity (6 mm and 60 mm) an effect is observed on the base flow. During base flow, new minor discharge peaks are occurring. This can indicate two things: 1. Precipitation is unable to infiltrate, ergo the infiltration capacity is reached during precipitation events. This results in more excess infiltration and discharge peaks during base flow. 2. The infiltration capacity is not the problem during peak flows. Almost no differences are observed in peak flow with low values of infiltration capacity, indicating that precipitation hardly infiltrates at all during these peaks. This is caused by the saturated soil. It is shown in Figure 40 that the unsaturated zone fills rapidly during precipitation events, leaving no space for infiltration.

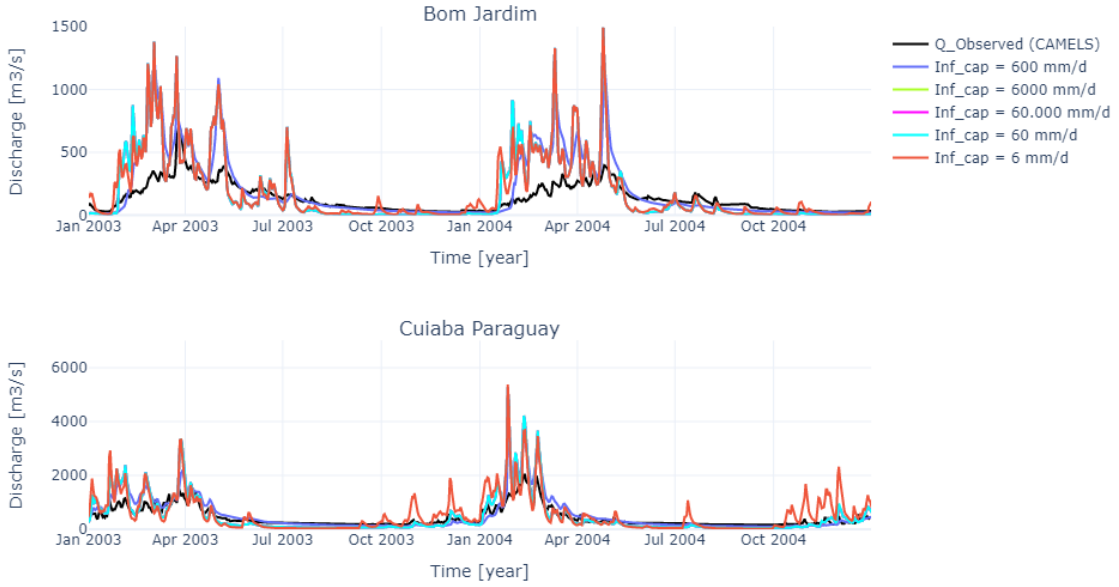

Figure 43: Timeseries observed and simulated discharge with several values of parameter infiltration capacity

Changing soil depth

The last part of this analysis on the performance of WFLOW is to change the soil depth. Soil depth is defined as the combined depths of the unsaturated and saturated zone. The default value for soil depth for both catchments is 2000 mm. Figure 44 shows the simulated discharge for several values of soil depth. Individual graphs per value of soil depth are shown in Appendix C for both catchments.

The first observation of Figure 44 is that multiplication factors of 10 or 0.1 have no impact on the simulated discharge. The green (soil depth = 20.000 mm) and cyan line (soil depth = 200 mm) are exactly the same as the blue line (soil depth = 2000 mm). This is a first indication that small alterations in soil depth are not effective in improving the performance of the model. A soil depth of 200.000 mm drastically improves the simulated discharge. A better representation of this signal is shown in Figure 51 and Figure 52 in Appendix C. A larger soil depth results in more available storage in the soil and is therefore more able to dampen peaks of discharge. Lastly, it is observed that simulated discharge for a soil depth of 20 mm is not affected by this soil depth reduction. A smaller soil depth would normally lead to faster saturation of the soil because the maximum storage is less. Therefore, resulting in a larger amount of excess infiltration. The absence of additional excess infiltration is noteworthy. This absence indicates a high rate of soil saturation, because the initial storage (determined by soil depth of 20mm and 20.000 mm) of both soil depths are filled in similar time frames.

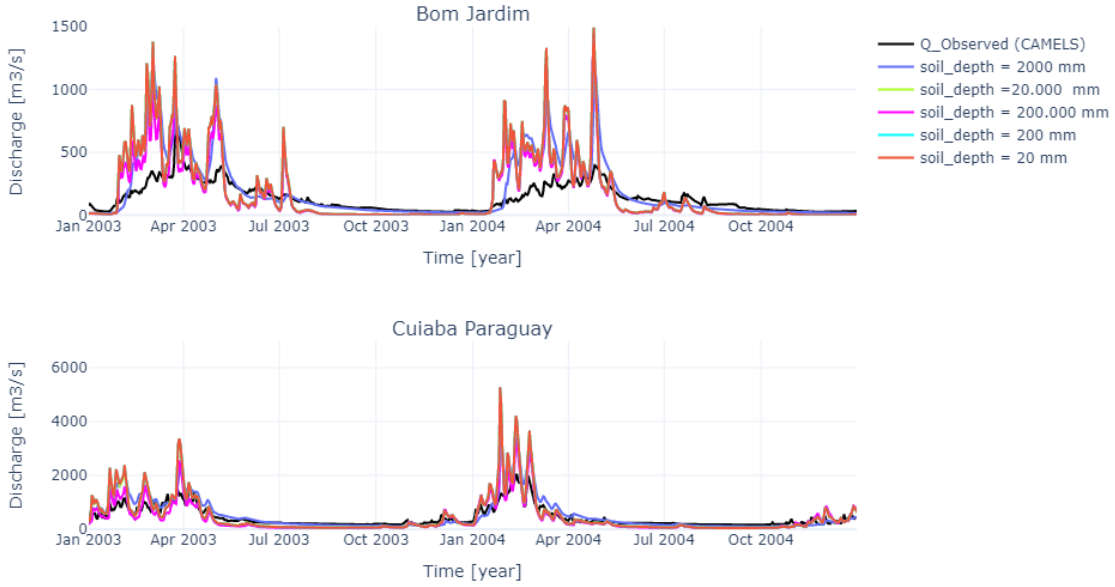

Figure 44: Timeseries observed and simulated discharge with several values of parameter soil depth

In summary, simulated discharge is not similar to observed discharge for the majority of catchments. A severe overestimation of discharge during peak flow is observed, while base flow is simulated adequately. Overestimation during peak flow is not caused by deviations between data sets of precipitation and temperature. However, large differences are observed between potential evapotranspiration of CAMELS-BR and input of WFLOW. It is observed that the soil saturates rapidly after a precipitation event, resulting in peak flow due to a (fully) saturated soil. An analysis on infiltration shows that excess infiltration has a large share in the overestimation of discharge during peak flow. Two possible explanations for this excess infiltration are poor estimations of the parameters infiltration capacity and soil depth. Different values for max infiltration capacity show that excess infiltration is not caused by lack of infiltration capacity. Multiplying the initial soil depth with factor 100 shows the largest impact on discharge. However, this dampening effect is temporary and is therefore not a viable solution.

5

Discussion

The results displayed in Chapter 4 raise several questions and topics for discussion. This chapter addresses these points of discussion and forms a basis for our conclusion on this research.

5.1 Study with observed data

The study with observed data from CAMELS-BR is performed to research the impact of deforestation on measured discharge. This impact is illustrated by two coefficients: runoff coefficient C_r and recession coefficient k. Firstly, the method for determining the coefficients and deforestation is discussed. Furthermore, we will critically look at our interpretation of the results. Lastly, we will reflect on this observed data study.

Method

The runoff coefficient quantifies how much of the precipitation ends up in the river as discharge. The method for calculating runoff coefficients uses yearly cumulative input for precipitation and discharge. In our study, a year is defined as the calendar year, ranging from January to December. However, a calendar year is not always equal to the hydrological year of a study area. For example, if the wet season ranges from November to February. Before making any conclusions, we must be certain that using the calendar year does not affect the outcome of our analysis. This is realized by looking at the precipitation charts for the catchments in our study area. The wet season of all catchments occur within the calendar year, which for example can be observed in Figure 32 for the catchments "Bom Jardim" and "Cuiaba Paraguay". Therefore the use of calendar years is justified and has no impact on the calculation of yearly runoff coefficients.

The recession coefficient of a discharge peak quantifies the hydrological response of a catchment. As described in subsubsection 3.2.3 our method detects peaks with a prominence larger than 0.2. Smaller peaks are regarded as noise and therefore not included in the calculation of yearly recession coefficients. In addition to these peaks with 0.2 prominence we also neglect recession curves shorter than 20 days. One could argue that by choosing to neglect these recession curves, our method is not objective anymore. Especially neglecting short recession curves is tricky as we are analyzing if deforestation leads to shorter recession curves. However, we need a threshold for the amount of entries in a curve to preserve the reliability of this research. Neglecting short recession curves could lead to false conclusions if these short curves are decisive for a catchment. However, with our method we detect multiple discharge peaks per year leading to multiple recession coefficients per year. The impact of neglecting very short curves is minimized because the yearly recession coefficient is determined by taking the mean of all recession coefficients in that year. A short recession curve also occurs when a precipitation event follows shortly after an earlier precipitation event. This type of short recession curve have no direct relation with deforestation and can therefore be neglected without jeopardising the quality of our research. In conclusion,

the value of 20 days is considered adequate to determine the impact of deforestation, while maintaining the quality of the calculated recession coefficients.

The entire analysis is conducted to relate deforestation to river runoff. Therefore we should not only look critically at the method for the two coefficients, but also at how deforestation is defined in our research. The method for determining yearly (cumulative) deforestation is described in subsubsection 3.1.1. Satellite data (Landsat) is used by Hansen et al. (2020) to create a global forest change map in Google Earth Engine. In this map, trees are defined as vegetation taller than 5 meters, probably due to the limitations of satellite images. This limitation brings two different points of discussion to our attention. Firstly, a forest does not solely consists of trees. A forest typically consists of trees, bushes and other vegetation. The dampening effects of bushes and other vegetation can not be neglected when analyzing rainfall-runoff relations. However, we can assume that in the process of deforestation not only trees are removed but also smaller vegetation is destroyed in the process. Therefore we conclude that looking solely at trees to determine deforestation is not an issue.

Secondly, there are also forests with smaller trees than 5 meter. The deforestation of these "low forests" are not detected by the satellite images, but research by several authors show that runoff dampening effects of "low forests" are significant (S. Brown, 1990) (Klinge et al., 2001) (Sommer et al., 2002). It is possible that catchments with a large area of "low forests" do not see the effects of deforestation of detectable forests. This effect of "low forests" is purely hypothetically as we do not have any data on the amount of "low forests" to validate this theory.

The method in subsubsection 3.1.1 also describes the definition of percentage deforestation used in our research. In this thesis deforestation is defined as the percentage of forest area loss in relation to the total land area of a catchment. One could argue that a better representation of deforestation is defined by taking the percentage of forest area loss in relation to the total forest area of that catchment. Two problems are identified by using forest area instead of land area. The first problem is caused by data availability. Information about the total forest area is retrieved from a landcover map (ESA2010 & UCLouvain, 2010). The data in this landcover map is obtained in 2009 and thereby represents the situation in 2009. However, the forest area in 2009 is not equal to the forest area in 2000 due to deforestation. We cannot use the forest area of 2009 because deforestation is a prime subject in our research. The second problem with using forest area instead of land area concerns the unequal presentation of deforestation it would cause between catchments. For example, logging 100 trees in a catchment with 100 trees will result in 100% deforestation when using forest area to determine percentage of deforestation. But logging 1 million trees in a catchment with 10 million trees will result in 10% deforestation. This method does not represent the true impact of deforestation on catchments. In addition, using forest area is likely to increase the chance of errors due to its dynamic properties. Forest area is dynamic due to deforestation and therefore has a different value each year. The measurements of deforestation obtained by satellite are reliable enough for this research. However, we can assume that these measurements deviate from the actual deforestation due to measuring errors. Using the dynamic forest area would mean that the effect of these measuring errors become significantly larger, especially when these errors are present in the first years of our scope. We can not remove the measuring errors by using the total land area of a catchment, but we can minimize the impact of these errors because the value of land area

is fixed.

Reflection on results

Chapter 1 showed that academics are divided on the impact of deforestation on runoff coefficients. A majority of researchers conclude that deforestation results in higher runoff coefficients. The research setup of this majority differs on numerous factors and ranges from models to studies with observed data. Our own hypothesis was that deforestation results in higher runoff coefficients and is therefore consistent with the majority of research. There is also a minority of researchers with an opposite conclusion: deforestation does not necessarily result in higher runoff coefficients. As described in Chapter 4, there are no clear signals that deforestation leads to higher runoff coefficients in our study area.

The research in section 1 showed that there are different conclusions on the relation between deforestation and the recession coefficient.(Olang & Furst, 2011),(Abdulkareem et al., 2018) and (Khaleghi, 2017) concluded that deforestation results in a shorter response time. The literature review (Bruijnzeel, 2004) shows that deforestation does not always result in an alternation of the response time. This is supported by (Klinge et al., 2001) and (Sommer et al., 2002) who found that the response time of forests in the Amazone make a fast recovery after deforestation. Figure 22 shows no significant relation between deforestation and the recession coefficient. This means that the conclusions from (Bruijnzeel, 2004), (Klinge et al., 2001) and (Sommer et al., 2002) apply for the study area of this thesis.

Chapter 4 shows that deforestation has no impact on the runoff and recession coefficient. The interpretation of this result depends on several factors. A major factor in the interpretation of our results is the absence of an analysis on what happens after deforestation. Deforestation by humans happens for a number of reasons. Firstly, logging of forests is happening for their resource of wood. In general, logging areas are replaced with new trees to be able use these areas again for logging in the future. We already discussed the damping ability of young forests. Literature shows that forests have the ability to completely regain their hydrological capacity in a period of 20 years. Secondly, logging of forests occurs to create area for other land uses like agriculture. Literature discussed in Chapter 1 mentions the effects of crops on runoff. Bruijnzeel (2004) observed that invasive flora like eucalyptus and oil palms are high in water consumption and fast growing, accelerating the return to pre-deforestation values for water yield. The above scenarios show that an impact of deforestation can be masked by the hydrological properties of young forests and agriculture. Thirdly, it is possible that forests are replaced with infrastructure and buildings resulting in less storage and infiltration possibilities. In general, replacing forests with infrastructure leads to more runoff. Lastly, there is non-forest land cover which can also be subject to change. Catchments in our study area do not fully consist of forests, other land cover types include grasslands, agriculture and barren plains. In our study we only look at land cover change by deforestation, but alternations of other land cover types possibly have equal impact on discharge. For example, a shift from barren plains to high-water yielding agriculture like palm oil have a dampening effect on peak flows (Costa, Botta, & Cardille, 2003). It is possible that shifts like these mask the impact of deforestation on river runoff. However, there is not much data available to observe if land cover change other than deforestation is occurring in our study area. The trend-lines in Figure 18 indicate that at least a portion of the AET loss caused by deforestation is

adsorbed. This adsorbing occurs either by new land cover after deforestation or by the transformation of areas with different land cover types.

The above arguments show that runoff is affected by whatever land cover replaces the forests after deforestation. The type of land cover after deforestation is not incorporated in our research. We still have to take it into account in the interpretation of our results. However, due to the unavailability of data we cannot analyze if our results are affected by land cover like oil palm plantations. In conclusion, the type of land cover after deforestation remains a factor of uncertainty.

5.2 Study with simple model

The simple model is used as step between observed discharge and discharge simulated by the complex model WFLOW. As shown in subsection 3.3 our simple model is a conceptual model with one bucket for storage. This means that our model is an oversimplified representation of reality and should be treated as such. The goal of our study with a simple model is to research if deforestation has an impact on the output of simple models with one parameter. In addition, it contributes to analyzing possible causes for the non-satisfactory behaviour of our WFLOW model.

The calibration is conducted with the objective function Nash-Sutcliffe Efficiency. It is difficult to pinpoint what a good threshold is for this objective function as it depends on the context of the research. Moriasi et al. (2007) state that in general a model performance can be judged as satisfactory when NSE is larger than 0.5. Figure 25 shows that the NSE of both catchments is above this threshold and therefore suffice to the general standards. However, it becomes more difficult for the scenario with yearly calibration. The threshold of the objective function is set to 0.3 for the yearly calibration to generate more values for parameter α. This manually chosen threshold is a good balance between creating sufficient output and preserving the quality of calibration. The most difficult part of simulating discharge is peak flow, due to the variability of precipitation during the wet season. Ergo, the wet season also has the most impact on the value of the objective function. The impact of a wet season becomes smaller when the time scope of the study is extended; years with bad simulations are balanced by years with good simulations. However, if you calibrate per year then it is not possible to balance a bad peak flow simulation with simulations from other years. Therefore it makes sense that values of the optimized objective function are lower if you calibrate years individually.

The question about the usability of simulation years with an objective function between 0.3 and 0.5 remains. We established that it is difficult to calibrate per year and the result will be lower values for the objective function. However, we need to calibrate per year to be able to analyze if there is a relation between parameter α and deforestation. It is not the goal of this basic model study to create a model to be used in different scenarios. The goal of this study is to observe if deforestation has impact on simple models with one parameter. This goal does not demand NSE values higher than 0.5. In conclusion, the use of a lower threshold in yearly calibration is justified for our study. A small footnote is that the goal of this study must be kept in mind when looking at the results.

Our study focused on the impact of deforestation on models with two different levels of complexity. One could wonder how you can analyze the impact of deforestation on a simple model without any parameter for land cover. It is true that there is no parameter for forests

in our model and therefore no direct connection with deforestation. However, this does not mean that the impact of deforestation can not be analyzed through an in-direct method. Parameter α illustrates the relation between water in storage and discharge. The storage "bucket" of our conceptual model represents multiple hydrological processes like canopy storage and groundwater. The type of land cover like tropical forest is important for these hydrological processes. This creates an indirect link between our storage and deforestation. By analyzing the trend of yearly α we can indirectly analyze if deforestation has impact on our simple model. A small footnote is that our storage "bucket" is a container concept for all hydrological processes active in a catchment. This means that α also includes the impact of parameters with less connection to deforestation like soil type. The impact of deforestation can be masked by the impact of other parameters. For example, an upward trend created by deforestation can be balanced by a downward trend created by the construction of a damn. This would result in a stable horizontal trend. Ergo, in our model we can not isolate the impact of deforestation.

In summary, using parameter α to analyze the impact of deforestation on a simple model is adequate for the goal of this study. A lower threshold for the objective function is a justified necessity to increase the quantity of output, while remaining a standard of quality. However, it must be noted that the conclusion of our analysis on simple models is not that deforestation has no impact on the relation between storage and discharge, but rather that we don't observe any relation for our model in our study area.

5.3 Study with complex model

The results described in Chapter 4 show several aspects of our model WFLOW. In summary, the simulated discharge does not match with the observed discharge and the performance of our model is considered poor. Good models with bad input will result in bad output. The input is tested to observe if the input for precipitation, actual and potential evapotranspiration is a probable cause for the poor output of the model. The figures in Figure 4.3 show that there are significant differences in data between sources (GLEAM, ERA5 etc.), with exception of precipitation and near-surface temperature. Uncertainty about data quality is a well-known struggle in hydrology, which is also observed in this thesis. This uncertainty about data quality is strengthened by comparing the two Budyko frameworks from the observed-data and complex model study. The Budyko frameworks Figure 19 and Figure 30 show two entirely different spreads. This difference in spread confirms there are significant deviations between our sources of input data, which leads to uncertainty about data used in this thesis. Furthermore, we observed that the expected values in a Budyko framework for humid catchments like our study area, have relatively low values for PET/P. Figure 30 does not match with this expectation of low PET/P values, which is a strong indication that the used input for the complex model study is not adequate. However, there are not many suitable data sets available as alternative for our complex model study. In conclusion, data uncertainty remains an important factor in interpreting the model output and is a plausible cause for the poor performance of WFLOW in this study area.

The performance of WFLOW in this study area is not sufficient to be able to analyze the impact of deforestation on WFLOW. After all, how can we analyze the impact of deforestation on WFLOW if a scenario without deforestation already results in poor performance in our study area? This would lead to false conclusions. Instead, we have analyzed what is

happening in our model and tried to find what causes the poor performance. Our analysis of the results indicates that WFLOW has trouble to correctly simulate groundwater flows. We observed that the unsaturated zone rapidly saturates after a precipitation event, which led to extensive excess infiltration and eventually higher discharge peaks. It makes sense that WFLOW has difficulties with the hydrological processes in the soil as WFLOW lacks groundwater modelling. The parameter "KsatHorFrac" is used to compensate for this gap in groundwater modelling. However, this parameter has no connection to a real physical parameter. There are no measurements of KsatHorfrac and therefore its values can not be validated. Still it is considered to be the best parameter for calibration in WFLOW. The results of our calibration are shown in Table 4. From this table we observe that the majority of optimized values for KsatHorFrac are at least 5000, which is considered very high. Normally, we would expect much lower values which is emphasized by the default value of 100. This raises questions on the reliability of these high values for KsatHorFrac. A KsatHorFrac of 5000 means that in saturated soils the conductivity of lateral ground flows is equal to 5000 times the vertical conductivity. These values are not realistic and should therefore be treated with caution. The optimal values for KGE are shown in green, in Table 4. We observe that the majority of these optimal values are negative, which relates to very poor model performance. However, (Knoben, Freer, & Woods, 2019) state that KGE values greater than -0.41 indicate that a model improves upon the mean flow benchmark, even if the model's KGE value is negative. The majority of our optimal KGE values indicate poor performance, even if we take these new limits for KGE into account.

Figure 43 and Figure 44 in subsection 4.3 show two things: 1. The infiltration capacity is not the reason for the performance of our model. 2. Soil depth can be increased to enhance the performance of our model. However, we only observe better performance with large multiplication factors like 100. Increasing the soil depth from 2 meter to 200 meter does not have a connection to the physical properties of our catchments. Therefore, we must conclude that a soil depth of 2 meter does not cause the overestimation of discharge during peak flows. If the infiltration and storage are not causing this overestimation, we must look at other possibilities. Our hypotheses is that the groundwater flows are not modelled correctly resulting in an underestimation of the outflow from the soil. In other words, the soil does not empty fast enough. This hypotheses is strengthened by signals observed in Figure 39 and Figure 42, where we observed that the unsaturated zone fills very quickly while emptying hardly occurs.

A different possibility is that the poor performance originates from factors above ground. River length and width are supplied by different map files. It is difficult to recreate the exact shape of a river in these map files. It is possible that this has an impact on the output of our model. For example, rivers often have floodplains which are used after heavy precipitation events. These floodplains act as temporary extensions of the river, creating more space to store the water from the precipitation event. The effect of these floodplains is the dampening of discharge peaks. Our simulated discharge shows an overestimation during peak flows. A possible explanation of this overestimation is the absence of floodplains in our model. River flow in WFLOW is determined through kinematic wave flow. The kinematic wave of channel does not only depend on the shape of a channel, but also on the roughness. For example, a river channel without plants or rifts has less resistance than a channel with weedy reaches. The roughness of a river channel is represented by the Manning coefficient. WFLOW uses lookup tables to determine the Manning value for

all channels (Deltares, 2021b). Determining the Manning coefficient of channels can be difficult, due to the absence of measured data. We have not calibrated on the Manning coefficient N. This means that we don't know if the used Manning coefficient in a channel matches with the true roughness of a channel. In theory an underestimation of the roughness can lead to larger peaks in discharge. Channels with low roughness, discharge water faster because the water experiences less resistance than channels with high resistance. Faster discharge leads to higher discharge peaks because the same amount of water must be discharged after a precipitation event. The above arguments show that the overestimation of discharge peaks is not necessarily caused by the absence of groundwater modelling. However, our results strongly indicate that the root of the poor model performance is in the representation of the hydrological soil processes.

6

Conclusion

In this section, the conclusion of this report is drawn. First, a brief recap of the research will be presented, followed by answering the sub-questions, answering the main research question, and lastly recommendations for future research.

6.1 Recapitulation of the research

The relevance of this research is proven by an increasing amount of deforestation and floods over the last decades. If hydrological models are used to predict the effects of precipitation events, then deforestation could be an important factor in the output of these predictions. The goal of this thesis is to research the impact of deforestation on river runoff generated by measurements, a simple model and a complex model (WFLOW), with the corresponding main research question:

"How adequate is deforestation represented in WFLOW?"

The representation of deforestation in WFLOW is determined through three studies. First, the relation between cumulative deforestation and measured discharge was analyzed during the study with observed data. The impact of discharge on observed data is quantified with the help of two coefficients: Runoff coefficient and the recession coefficient of discharge peaks. Then, a study with a simple model was conducted to reduce the step from observed data to a complex model like WFLOW. This simple model uses parameter α to represent the relation between storage and discharge, which is then set against the amount of cumulative deforestation. From this simple model a transition is made to the complex model WFLOW. With this model, discharge is simulated for a study area in Brazil subject to deforestation. The impact of deforestation simulated discharge would then again be broken down in runoff coefficient and recession coefficient to compare the output of observed data and complex model study. An analysis on the hydrologic processes of WFLOW is conducted when the performance of WFLOW is not satisfactory.

6.2 Answers to the sub-questions

Our main research question is divided into four sub-questions. This sub-chapter will provide answers to these sub-questions.

1. In areas with (heavy) deforestation, is the observed discharge affected by deforestation with respect to: runoff coefficient, timing and magnitude of extremes?

The answer for this sub-question requires to analyze the impact of deforestation on the runoff coefficient and the recession coefficient of discharge peaks. The general view among researchers is that deforestation leads to higher runoff coefficients. However, the figures in section 4 show that the percentage of cumulative deforestation has no relation with the trend of the runoff coefficient in time in our study area. Several signals are observed for catchments with low and high cumulative deforestation. Different catchments with

varying deforestation show both increasing and decreasing trends for C_r. Therefore, we conclude that deforestation does not necessarily lead to higher runoff coefficients in our study area.

The impact of deforestation on discharge peaks is quantified by the recession coefficient. The results observed and discussed in section 4 show that there is no general relation between deforestation and the recession coefficient in our study area. In conclusion, there are catchments that do show an upward trend for both coefficients against deforestation. However, we do not observe any causality between deforestation and the discharge coefficient for our study area.

2. What is the impact of deforestation on the simulated discharge and storage of a simple conceptual model?

The main goal of this sub-question was to research if a basic conceptual model is impacted by a dynamic land-cover change process like deforestation. The entire analysis of this study is build on one calibrated parameter due to the simplicity of this model. The recession parameter α represents the relation between water in storage and discharge, which is an ideal indicator for the impact of deforestation on the hydrological properties of a catchment. The results of the two catchments show no clear relation between deforestation and recession parameter α. A small footnote is that this study on a simple model is only conducted on two catchments. This must be taken into consideration when using our conclusion that deforestation has no impact on the simulated discharge of a simple model.

3. In areas with (heavy) deforestation, is the calculated river discharge from WFLOW similar to observed river discharge with respect to: runoff coefficient, timing and magnitude of extremes?

The answer to this sub-question is straightforward. The simulated river discharge from WFLOW is not similar to observed river discharge. section 4 shows that there is a severe overestimation of discharge during peak flows.

4. If the answer to sub3 is 'no', what changes in model inputs (such as infiltration capacity, soil maps, etc.) can improve the calculated river discharge from WFLOW in regions with heavy deforestation?

This sub-question is used to analyze the weak points of WFLOW in case the answer to sub-question 3 is 'no'. The first conclusion on this analysis is that a difference in input of precipitation, potential evapotranspiration and temperature is not the cause of the overestimation of discharge peaks. The results from subsection 4.3 do indicate that in our study area the problems occur in the groundwater flow modelling of WFLOW. Excess infiltration is identified as the main cause of the poor performance of WFLOW during peak flows. Logical explanations for this high excess infiltration is a lack of infiltration capacity or a lack of available storage in the soil. Changing the parameters "infiltration capacity" and "soil depth" did not solve the overestimation for discharge of our model. One exception was observed for a soil depth of 200 meters, which dampened the discharge during peak flows. However, this soil depth with a multiplication factor of 100 is not a realistic value and is therefore excluded as possible solution. In conclusion, the weak part of WFLOW is pinpointed to the area of groundwater modelling. The exact reason for the observed performance is not known, but there are strong indications that soil fluxes are not represented adequately in WFLOW.

6.3 Main conclusion

The answers of the four sub-questions provide the tools to form a final conclusion and determine how adequate deforestation is represented in WFLOW. The goal of this thesis was to research the relation between deforestation, observed discharge and simulated discharge. However, the simulated discharge by WFLOW shows a steep overestimation of discharge during peak flow, despite an extensive analysis on the hydrological processes within WFLOW. In conclusion, the performance of WFLOW is not adequate enough to determine the impact of deforestation on discharge in our study area. Though, no causality is observed between this performance and deforestation. Furthermore, it is difficult to quantify the impact of deforestation on a model if the observed discharge shows no clear relation with deforestation.

6.4 Limitations and recommendations

The results of our research lead to several conclusion and insights. Chapter 5 discussed some of the strong and weak points of our research. These points of interest lead to limitations on the interpretation of our results. Furthermore, we have recommendations on how to improve these weaker points.

Limitations

The limitations of this study show in which situations our method and results are applicable. Our study area consists of 30 catchments in Brazil. The hydrologic properties of a catchment are partly determined by characteristics like type of soil and landcover. The results of our study are only applicable to our study area due to the specific qualities of our catchments. Hydrological models differ in complexity and conceptualization of hydrologic processes. The output of a model is partially determined by these two factors. Our study is limited to the use of WFLOW. Using different models would possibly generate different results and is therefore noted as a limitation.

Input is a major factor in the output of a study. Our study is limited to the data sources described in subsection 3.1. Our results are generated with these datasets. Any conclusion on these results must therefore be limited to these sets of data. Different sources for data can lead to different results and should therefore always be validated upon usage. The input of deforestation is limited by the availability of data. Deforestation is obtained with satellite images. These satellite images provide yearly data from 2000 to 2020. This means our results can only be reproduced for this range of years.

The above limitations contribute to a better understanding of our study and prevent that our results are misinterpreted or used in studies with different characteristics.

Recommendations for future research

During our research we discovered multiple points of interest which can be used for future research. The first recommendations apply to the study with observed data. The observed relation between deforestation and runoff is considered notable as the majority of literature sees an opposite relation. However, most studies do not use the deforestation determined by Hansen et al. (2020) in their research. It is interesting to analyze the impact of deforestation on discharge in catchments with different properties like climate

and soil type. Figure 1 shows that Africa is subject to heavy deforestation. This could be an interesting study area for future research. The second recommendation regarding our study with observed data is the implementation of classes to categorize catchments. The categorization of catchments could lead to different insights and conclusions of our results. Possible categories could be: climate, dominant land cover and the mean slope of a catchment.

Our study with the basic model uses a conceptual model with one storage "bucket". It could be interesting to expand this model with more fluxes and storages to create a more complex model. The results of this "expanded" model can then be compared with the output of our basic model and WFLOW. The final recommendations apply to our study with WFLOW. The eWaterCycle platform offers several models next to WFLOW. We recommend too use our exact method to simulate discharge with one of these other models. This should be achievable due to the flexible nature of eWaterCycle.

Thirdly, we have a recommendation for future research on the representation of deforestation in WFLOW. Hydrological models typically don't have the ability to represent changing landcover like deforestation. A possible solution would be to run the model per year with different values for landcover. The first year uses the initial conditions of a catchment and stops running after one year. The second year would then need to implement the output of year 1 as initial conditions plus a different value for landcover. Implementing dynamic deforestation in WFLOW is only possible for catchments that already show adequate performance.

Fourthly, we identified that type of landcover after deforestation is an important factor in the uncertainty of our results. This factor was not included in this research due to data scarcity and time constraints. Including landcover type after deforestation presents a better opportunity for analyzing the impact of deforestation on discharge. We identified possible sources for annual landcover maps in Brazil:

- Copernicus Global Land Cover Layers: CGLS-LC100 Collection 3 (Buchhorn, Lesiv, Tsendbazar, Herold, & L. Bertels, 2020).
- ESA Climate Change Initiative - Land Cover (ESA, 2020).

The final recommendation involves expanding our analysis on why WFLOW is not performing well in Brazilian catchments. There are more parameters suitable for change, like we did with infiltration capacity and soil depth. Analyzing the effects of changing different parameters or even combinations, could give more insight in the sensitivity of WFLOW.

Recommendations for the field

One of the goals of this thesis was to explore the possibilities of WFLOW in catchments subject to deforestation. This goal was not achieved because the performance of our model is not considered adequate. Instead, we did an analysis on possible reasons for this performance and discovered that groundwater modelling can be considered a weak point in WFLOW. We would recommend that WFLOW is not used for research where groundwater flows are an important factor. In addition, we would critically look at the results for catchments with heavy deforestation.

References

Abdulkareem, J., Sulaiman, W., Pradhan, B., & Jamil, N. (2018). Relationship between design floods and land use land cover (lulc) changes in a tropical complex catchment. *Arabian journal of Geosciences*, 376. doi: 10.1007/s12517-018-3702-4

ANA. (2019). Retrieved from `http://www.snirh.gov.br/hidroweb/)`

Batjes, N., Ribeiro, E., & van Oostrum, A. (2020). Standardised soil profile data to support global mapping and modelling (wosis snapshot 2019). *Earth Syst. Sci. Data*, 299–320. doi: https://doi.org/10.5194/essd-12-299-2020

Beck, H., Wood, E., M.Pan, Fisher, C., Miralles, D., van Dijk, A., ... Adler, R. (2019). Mswep v2 global 3-hourly 0.1° precipitation: methodology and quantitative assessment. *Bulletin of the American Meteorological Society*, 473–500. doi: https://doi.org/10.1175/BAMS-D-17-0138.1

Beven, K. (1993). Prophecy, reality and uncertainty in distributed hydrological modelling. *Advances in Water Resources*, 41-51. doi: 10.1016/0309-1708(93)90028-E

Brassington, F., & Younger, P. (2010). A proposed framework for hydrogeological conceptual modelling. *Water and Environment*, 261-273. doi: 10.1111/j.1747-6593.2009.00173.x

Breda, N., Huc, R., Granier, A., & Dreyer, E. (2006). Temperate forest trees and stands under severe drought: A review of ecophysiological responses, adaptation processes and long-term consequences. *Annals of Forest Science*, 625 - 644. doi: 10.1051/forest:2006042

Brown, A., Zhang, L., McMahon, T., Western, A., & Vertessy, R. (2005a). A review of paired catchment studies for determining changes in water yield resulting from alterations in vegetation. *Journal of Hydrology*, 28 - 61. doi: 10.1016/j.jhydrol.2004.12.010

Brown, A., Zhang, L., McMahon, T., Western, A., & Vertessy, R. (2005b). A review of paired catchment studies for determining changes in water yield resulting from alterations in vegetation. *Journal of Hydrology*, 28-61. doi: 10.1016/j.jhydrol.2004.12.010

Brown, S. (1990). Tropical secondary forests. *Journal of tropical ecology*, 1 -32. doi: 10.1017/S0266467400003989

Bruijnzeel, L. (2004). Hydrological functions of tropical forests: Not seeing the soil for the trees? *Agriculture, Ecosystems and Environment*, 185 - 228. doi: 10.1016/j.agee.2004.01.015

Buchhorn, M., Lesiv, M., Tsendbazar, N., Herold, M., & L. Bertels, B. S. (2020). Copernicus global land cover layers—collection 2. *Remote sensing*. doi: doi:10.3390/rs12061044

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 Systems Science Data*, 2075–2096. doi: https://doi.org/10.5194/essd-12-2075-2020

Chow, V., Maidment, D., & Mays, L. (1988). *Applied hydrology*. McGraw-Hill Book Company.

Copernicus. (2021). Retrieved from `https://land.copernicus.eu/global/products/lai`

Costa, M., Botta, A., & Cardille, J. (2003). Effects of large-scale changes in land cover on the discharge of the tocantins river, southeastern amazonia. *Journal of Hydrology*, 206-217. doi: 10.1016/S0022-1694(03)00267-1

Deltares. (2019). Retrieved from `https://wflow.readthedocs.io/en/latest/wflow_sbm.html`

Deltares. (2021a). Retrieved from `https://wflow.readthedocs.io/en/latest/wflow_building.html`

Deltares. (2021b). Retrieved from `https://wflow.readthedocs.io/en/latest/wflow_sbm.html`

Deltares. (2021c). Retrieved from `https://www.deltares.nl/en/software/wflow-hydrology/`

de Wildt, R. (2022). Retrieved from `https://github.com/RogierWildt/Thesis_Msc.git`

Do, H., Gudmundsson, L., Leonard, M., & Westra, S. (2018). Camels-br: hydrometeorological time series and landscape attributes for 897 catchments in brazilthe global streamflow indices and metadata archive (gsim) – part 1: The production of a daily streamflow archive and metadata. *Earth Systems Science Data*, 765–785. doi: https://doi.org/10.5194/essd-10-765-2018

Dyhr-Nielsen, M. (1986). Hydrological effect of deforestation in the chao phraya basin in thailand. *International Symposium on Tropical Forest Hydrology and Application*, 15.

Enemark, T., Peeters, L., Mallants, D., & Batelaan, O. (2019). Hydrogeological conceptual model building and testing: A review. *Journal of Hydrology*, 310-329. doi: https://doi.org/10.1016/j.jhydrol.2018.12.007

ESA. (2020). Retrieved from `http://maps.elie.ucl.ac.be/CCI/viewer/index.php`

ESA2010, & UCLouvain. (2010). Retrieved from `http://due.esrin.esa.int/page_globcover.php`

eWaterCycle. (2021). Retrieved from `https://www.ewatercycle.org/`

FAO/IIASA/ISRIC/ISSCAS/JRC. (2013). *Harmonized world soil database (version 1.2).* Retrieved from `https://webarchive.iiasa.ac.at/Research/LUC/External-World-soil-database/HTML/`

Farrick, K. K., & Branfireun, B. A. (2014). Soil water storage, rainfall and runoff relationships in a tropical dry forest catchment. *Water Resources Research*, 625 - 644. doi: 10.1002/2014WR016045

Fohrer, N., Haverkamp, S., Echkhardt, K., & Frede, H. (2001). Hydrologic response to land use changes on the catchment scale. *Physics and Chemistry of the Earth, Part B: Hydrology, Oceans and Atmosphere*, 577 - 582. doi: https://doi.org/10.1016/S1464-1909(01)00052-1

Food, & Organization, A. (2020). Global forest resources assessment 2020 – key findings. doi: https://doi.org/10.4060/ca8753en

Gebremicael, T., Mohamed, Y., & der Zaag, P. V. (2019). tributing the hydrological impact of different land use types and their long-term dynamics through combining parsimonious hydrological modelling, alteration analysis and plsr analysis. *Science of the Total Environment*, 1155 - 1167. doi: 10.1016/j.scitotenv.2019.01.085

Guimberteau, M., Ciais, P., Boisier, J. P., Aguiar, A. P. D., Biemans, H., Deurwaerder, H. D., ... Verbeeck, H. (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*, 1455-1475. doi: 10.5194/hess-21-1455-2017

Gupta, H., Kling, H., Yilmaz, K., & Martinez, G. (2009). Decomposition of the mean squared error and nse performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, 81-91. doi: https://doi.org/10.1016/j.jhydrol.2009.08.003

Hansen, M., Potapov, P., Moore, R., Hancher, M., Turubanova, S., Tyukavina, A., ... Townshend, J. (2013). High-resolution global maps of 21st-century forest cover change. *Science*, 850-853. doi: 10.1126/science.1244693

Hansen, M., Potapov, P., Moore, R., Hancher, M., Turubanova, S., Tyukavina, A., ... Townshend, J. (2020). Retrieved from `https://glad.earthengine.app/view/global-forest-change#dl=1;old=off;bl=off;lon=20;lat=10;zoom=3;`

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horanyi, A., Munoz-Sabater, J., ... Abdalla, S. (2020). The era5 global reanalysis. *Quarterly journal of the royal meteorological society*, 1999-2049. doi: https://doi.org/10.1002/qj.3803

Hrachowitz, M., Stockinger, M., Coenders-Gerrits, M., van der Ent, R., Bogena, H., Lucke, A., & Stumpp, C. (2020). Deforestation reduces the vegetation-accessible water storage in the unsaturated soil and affects catchment travel time distributions and young water fractions. *HESS*. doi: https://doi.org/10.5194/hess-2020-293

Hut, R., Drost, N., van de Giesen, N., van Werkhoven, B., Abdollahi, B., Aerts, J., ... Weel, B. (2021). The ewatercycle platform for open and fair hydrological collaboration. *Geosci. Model Dev. Discuss. [preprint]*. doi: 10.5194/gmd-2021-344

Jain, M. (2014). Recession of discharge. *Encyclopedia of Snow, Ice and Glaciers*. doi: https://doi.org/10.1007/978-90-481-2642-2_437

Jones, J., Achterman, G., Augustine, L., Creed, I., Ffolliott, P., MacDonald, L., & Wemple, B. (2009). Hydrologic effects of a changing forested landscape - challenges for the hydrological sciences. *Hydrological Processes*, 2699-2704. doi: 10.1002/hyp.7404

Keenan, R. J., Reams, G. A., Achard, F., de Freitas, J., Grainger, A., & Lindquist, E. (2015). Dynamics of global forest area: Results from the fao global forest resources assessment 2015. *Forest ecology and management*, 9-20. doi: https://doi.org/10.1016/j.foreco.2015.06.014

Khaleghi, M. (2017). The influence of deforestation and anthropogenic activities on runoff generation. *Journal of Forest science*, 245 - 253. doi: 10.17221/130/2016-JFS

Klinge, R., Schmidt, J., & Folster, H. (2001). Simulation of water drainage of a rain forest and forest conversion plots using a soil water model. *Journal of Hydrology*, 82 - 95. doi: https://doi.org/10.1016/S0022-1694(01)00346-8

Knoben, W., Freer, J., & Woods, R. (2019). Technical note: Inherent benchmark or not? comparing nashsutcliffe and kling-gupta efficiency scores. *HESS*, 4323-4331. doi: 10.5194/hess-23-4323-2019

Lehrer, B., Verdin, K., & Jarvis, A. (2008). New global hydrography derived from spaceborne elevation data. *Advancing Earth and Space Science*, 93-94. doi: 10.1029/2008EO100001

Lindstrom, G., Johansson, B., Persson, M., & M. Gardelin, S. B. (1997). Development and test of the distributed hbv-96 hydrological model. *Journal of Hydrology*, 272-288. doi: 10.1016/S0022-1694(97)00041-3

Martens, B., Miralles, D., Lievens, H., van der Schalie, R., de Jeu, R., Fernández-Prieto, D., ... Verhoest, N. (2017). Gleam v3: satellite-based land evaporation and root-

zone soil moisture. *Geosci. Model Dev.*, 1903–1925. doi: https://doi.org/10.5194/gmd-10-1903-2017

Mianabadi, A., Coenders-Gerrits, M., Shirazi, P., & Ghahraman, B. (2019). A global budyko model to partition evaporation into interception and transpiration. *HESS*, 4983–5000. doi: https://doi.org/10.5194/hess-23-4983-2019

Milhorance, F. (2021). Record floods stun brazil's northeast, killing at least 20. *New York Times*. Retrieved from `https://www.nytimes.com/2021/12/28/world/americas/brazil-floods-climate-change.html`

Mohammad, A., & Adamn, M. (2010). The impact of vegetative cover type on runoff and soil erosion under different land uses. *Catena*, 97 - 103. doi: 10.1016/j.catena.2010.01.008

Moriasi, D., Arnold, J., van Lieuw, M., Binger, L., Harmel, R., & Veith, T. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *ASABE*, 885−900. doi: 10.13031/2013.23153

NASA. (2014). Retrieved from `https://www2.jpl.nasa.gov/srtm/`

Nash, J., & Sutcliffe, J. (1970). River flow forecasting through conceptual models part i — a discussion of principles. *Journal of Hydrology*, 282-290. doi: https://doi.org/10.1016/0022-1694(70)90255-6

NCCO. (2022). Retrieved from `https://legacy.climate.ncsu.edu/edu/,addendum={(accessed:"21.04.2022")}`

Nijzink, R., Hutton, C., Pechlivanidis, I., Capell, R., Arheimer, B., Freer, J., ... Hrachowitz, M. (2016). The evolution of root-zone moisture capacities after deforestation: a step towards hydrological predictions under change? *HESS*, 4775–4799. doi: https://doi.org/10.5194/hess-20-4775-2016

NOAA. (2019). Retrieved from `https://www.esrl.noaa.gov/psd/`

Olang, L., & Furst, J. (2011). Effects of land cover change on flood peak discharges and runoff volumes: Model estimates for the nyando river basin, kenya. *Hydrological processes*, 80-89. doi: 10.1002/hyp.7821

Qian, W. (1983). Effects of deforestation on flood characteristics with particular reference to hainan island, china. *Hydrology of Humid Tropical Regions: Aspects of Tropical Cyclones, Hydrological Effects of Agriculture and Forestry Practice*, 249 - 257.

Rannard, G., & Gillet, F. (2021). Cop26: World leaders promise to end deforestation by 2030. *BBC*. Retrieved from `https://www.bbc.com/news/science-environment-59088498`

Siquera, V., Paiva, R., Fleischmann, A., Fan, F., Ruhoff, A., Pontes, P., ... Collischonn, W. (2018). Toward continental hydrologic–hydrodynamic modeling in south america. *HESS*, 4815–4842. doi: https://doi.org/10.5194/hess-22-4815-2018

Siriwardena, L., Finlayson, B., & McMahon, T. A. (2006). The impact of land use change on catchment hydrology in large catchments: The comet river, central queensland, australia. *Journal of Hydrology*, 199 - 214. doi: 10.1016/j.jhydrol.2005.10.030

Sommer, R., de Abreu Sa, T., Vielhauer, K., A. Carioca de Araujo, H. F., & Vlek, P. (2002). Transpiration and canopy conductance of secondary vegetation in the eastern amazon. *Agricultural and Forest Meteorology*, 103 - 121. doi: https://doi.org/10.1016/S0168-1923(02)00044-8

Thielen, J., Bartholmes, J., Ramos, M., & Roo, A. D. (2009). The european flood alert system ĝ€" part 1: Concept and development. *Hydrology and Earth System Sciences*, 125-140. doi: 10.5194/hess-13-125-2009

van Waveren, R., Groot, S., Scholten, H., van Geer, F., Wosten, J., Koeze, R., & Noort, J. (1999). *Good modelling practice handbook*. STOWA 99-05.

Viessmann, W., & Lewis, G. (1996). *Introduction to hydrology (fourth ed.)*. HarperCollins College Publishers.

Wang-Erlandsson, L., van der Ent, R., Gordon, L., & Savenije, H. (2014). Contrasting roles of interception and transpiration in the hydrological cycle – part 1: Temporal characteristics over land. *Earth Syst. Dynam.,5*, 441–469. doi: https://doi.org/10.5194/esd-5-441-2014

Wilk, J., Andersson, L., & Plermkamon, V. (2001). Hydrological impacts of forest conversion to agriculture in a large river basin in northeast thailand. *Hydrological processes*, 2729-2748. doi: https://doi.org/10.1002/hyp.229

Zhang, Z., Yu, X., Zhao, Y., & Qin, Y. (2003). Advance in researches on the effect of forest on hydrological process. *Chinese Journal of Applied Ecology*, 113-116. doi: 10019332

Zimmerman, B., Elsenbeer, H., & Moraes, J. D. (2006). The influence of land-use changes on soil hydraulic properties: Implications for runoff generation. *Forest Ecology and Management*, 29 - 33. doi: 10.1016/j.foreco.2005.10.070

A

Additional Results: Observed data study

A.1 Runoff coefficient

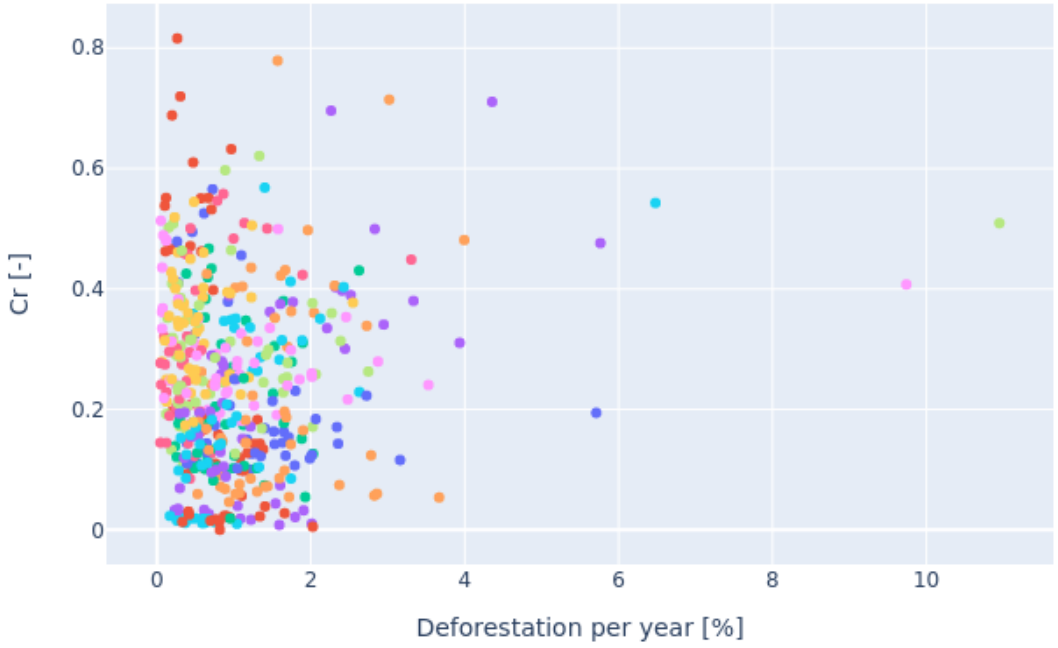

Figure 45: Relation between the runoff coefficient C_r and the percentage of deforestation per year for all catchments. Each color represents a catchment.

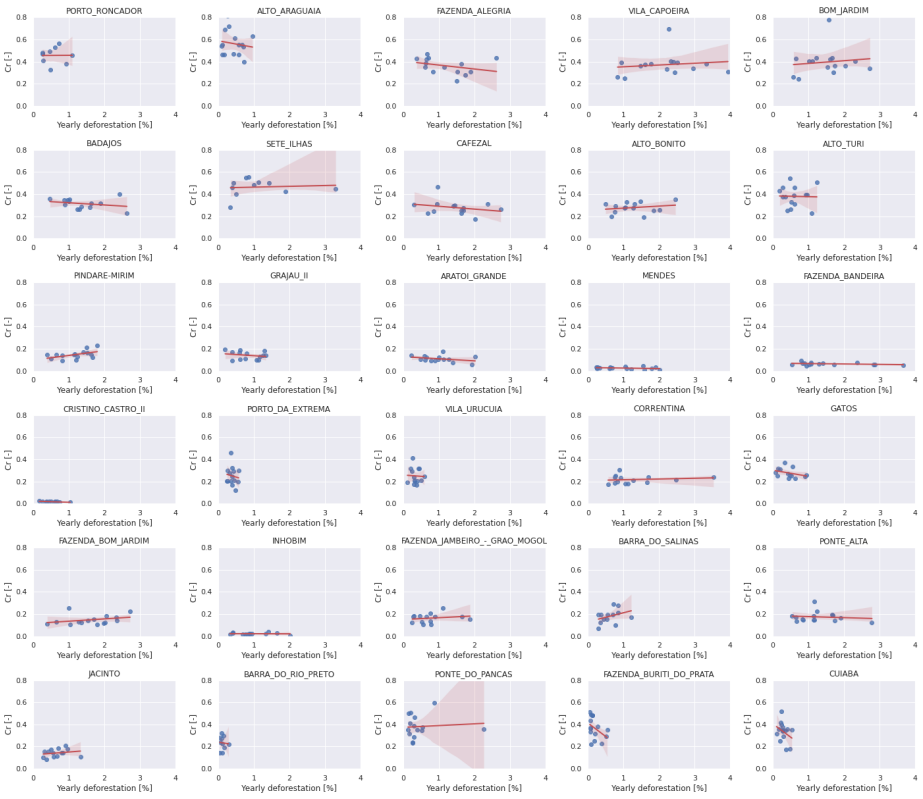

Figure 46: Relation between the runoff ratio C_r and the percentage of deforestation per year with a individual plot for each catchment .

A.2 Recession coefficient

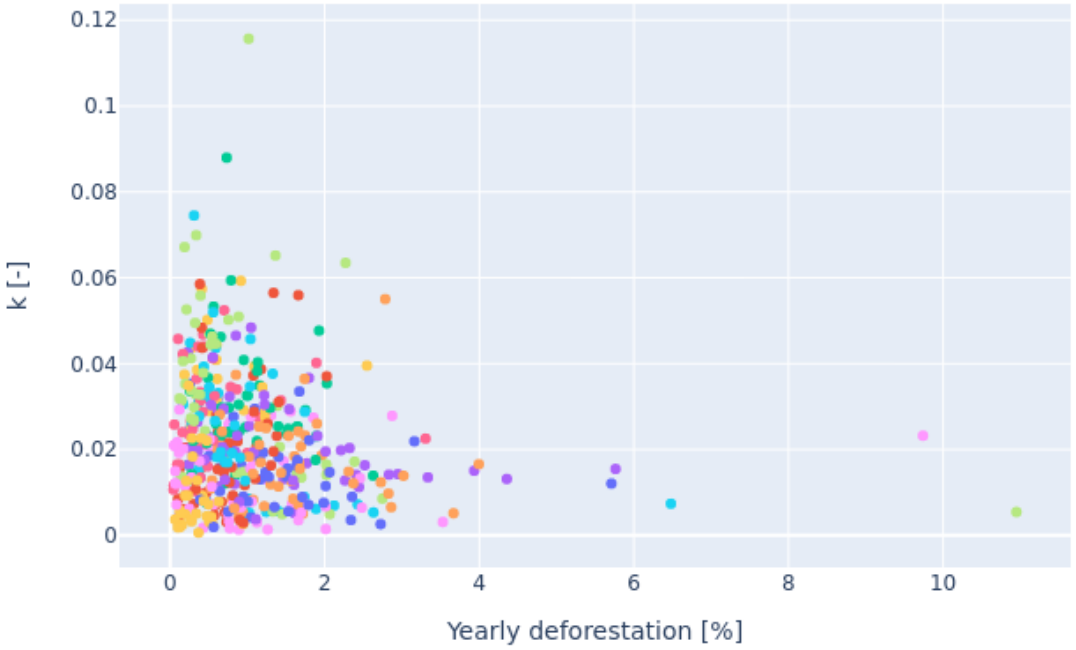

Figure 47: Relation between the recession coefficient k and the percentage of deforestation per year for all catchments. Each color represents a catchment.

Figure 48: Relation between the runoff ratio C_r and the percentage of deforestation per year for all catchments.

B

Calibration

Table 4: Results of the calibration by objective function Kling-Gupta Efficiency (KGE) on parameter "KsatHorFrac". The table shows the value for KGE per value of "KsatHorfrac" for each catchment. The optimal KGE of a catchment is shown in green.

Catchments	1	25	50	75	100	250	500	750	1000	5000	10.000	20.0000	50.0000
Bom Jardim	0.057	0.059	0.062	0.067	0.072	0.107	0.164	0.209	0.244	0.408	0.433	0.434	0.411
Badajos	-1.433	-1.433	-1.431	-1.427	-1.421	-1.378	-1.308	-1.251	-1.204	-0.960	-0.919	-0.924	-0.972
Sete Ilhas	0.525	0.527	0.530	0.533	0.538	0.561	0.594	0.616	0.633	0.653	0.642	0.635	0.626
Cafezal	-0.993	-0.990	-0.987	-0.983	-0.979	-0.954	-0.915	-0.883	-0.856	-0.733	-0.718	-0.721	-0.738
Alto Turi	0.240	0.241	0.242	0.244	0.245	0.257	0.276	0.293	0.307	0.381	0.389	0.387	0.379
Pindare Mirim	-1.989	-1.981	-1.976	-1.972	-1.969	-1.975	-2.001	-2.041	-2.071	-2.221	-0.252	-0.261	-2.256
Aratoi Grande	-1.955	-1.932	-1.922	-1.923	-1.929	-1.996	-2.100	-2.177	-2.233	-2.454	-2.484	-2.496	-2.501
Fazenda Bandeira	-6.814	-6.200	-5.777	-5.513	-5.335	-4.931	-4.848	-4.880	-4.930	-5.442	-5.754	-6.149	-6.831
Porto da Extrema	-0.524	-0.510	-0.485	-0.459	-0.433	-0.301	-0.167	-0.092	-0.048	-0.001	-0.033	-0.072	-0.138
Gatos	-10.859	-10.479	-10.062	-9.671	-9.324	-7.962	-6.932	-6.440	-6.160	-5.722	-6.008	-6.491	-7.360
Fazenda Bom Jardim	-10.864	-10.591	-10.339	-10.116	-9.918	-9.091	-8.378	-7.988	-7.735	-6.786	-6.634	-6.656	-6.975
Inhobim	-11.218	-11.021	-10.916	-10.897	-10.924	-11.384	-12.189	-12.775	-13.180	-14.260	-14.225	-14.127	-14.049
Ponte Alta	-0.373	-0.311	-0.259	-0.217	-0.184	-0.078	-0.046	-0.054	-0.069	-0.177	-0.207	-0.236	-0.287
Cuiaba	0.012	0.154	0.303	0.414	0.498	0.738	0.835	0.844	0.834	0.646	0.507	0.332	0.090

Table 5: Results of the calibration by objective function Nash-Sutcliffe Efficiency (NSE) on parameter "KsatHorFrac". The table shows the value for NSE per value of "KsatHorfrac" for each catchment. The optimal NSE of a catchment is shown in green.

Catchments	1	25	50	75	100	250	500	750	1000	5000	10.000	20.0000	50.0000
Bom Jardim	-2.29	-2.27	-2.25	-2.23	-2.06	-2.04	-1.78	-1.58	-1.43	-0.74	-0.63	-0.63	-0.73
Badajos	-30.48	-30.39	-30.25	-30.09	-29.92	-28.93	-27.59	-26.64	-25.88	-22.52	-22.16	-22.46	-23.48
Sete Ilhas	-0.32	-0.31	-0.30	-0.29	-0.27	-0.18	-0.05	0.04	0.12	0.35	0.36	0.33	0.26
Cafezal	-28.21	-28.12	-27.99	-27.85	-27.70	-26.81	-25.46	-24.33	-23.40	-18.44	-17.41	-17.18	-17.72
Alto Turi	-2.07	-2.06	-2.05	-2.04	-2.03	-1.94	-1.79	-1.66	-1.55	-0.88	-0.72	-0.68	-0.75
Pindare Mirim	-40.99	-40.45	-39.73	-39.03	-38.37	-35.36	-32.42	-30.71	-29.56	-26.16	-26.43	-27.47	-29.68
Aratoi Grande	-23.75	-22.67	-21.56	-20.66	-19.92	-17.35	-15.53	-14.75	-14.30	-14.14	-15.67	-17.09	-20.73
Fazenda Bandeira	-1507.23	-1308.46	-1177.05	-1094.62	-1036.12	-876.86	-802.01	-784.88	-785.94	-998.11	-1211.90	-1516.12	-2086.74
Porto da Extrema	-11.37	-11.06	-10.63	-10.22	-9.82	-8.03	-6.35	-5.40	-4.79	-3.05	-3.22	-3.77	-5.66
Gatos	-285.14	-272.92	-257.43	-245.04	-235.29	-200.55	-177.47	-167.12	-161.67	-166.02	-184.93	-212.91	-263.23
Fazenda Bom Jardim	-1335.78	-1296.92	-1260.34	-1230.48	-1205.69	-1109.55	-1032.72	-989.00	-962.70	-876.25	-885.18	-927.16	-1039.50
Inhobim	-428.67	-403.48	-380.14	-361.05	-345.19	-286.49	-244.66	-226.34	-216.83	-218.75	-243.89	-282.19	-352.87
Ponte Alta	-9.40	-8.36	-7.54	-6.90	-6.39	-4.63	-3.52	-3.08	-2.88	-3.12	-3.79	-4.76	-6.48
Cuiaba	-0.12	0.11	0.30	0.43	0.51	0.69	0.72	0.69	0.66	0.20	-0.17	-0.71	-.158

C

Additional results: Complex model study

C.1 Infiltration capacity

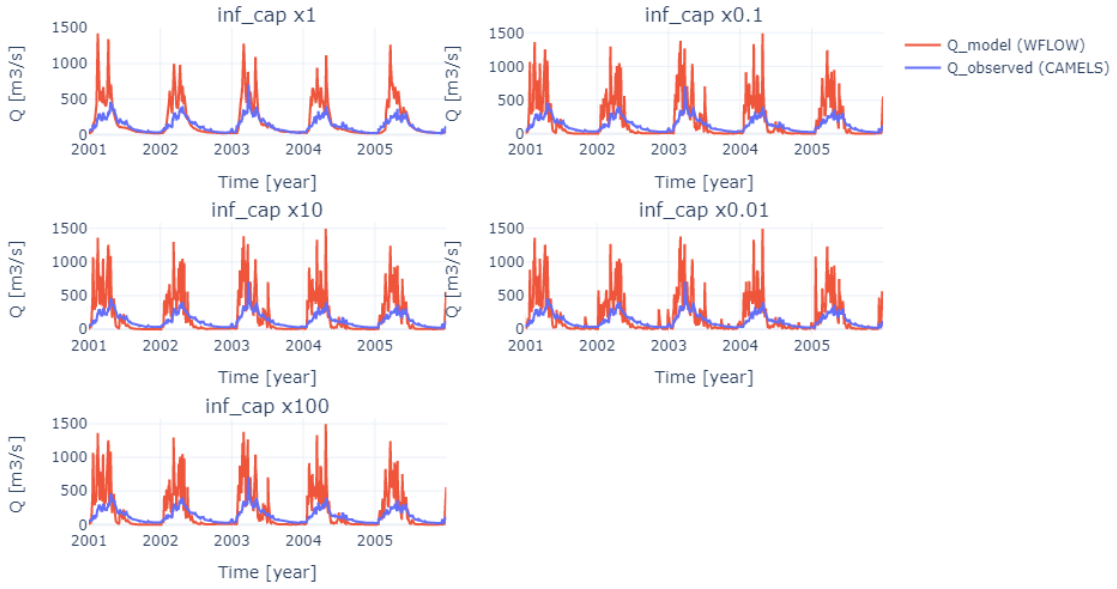

Figure 49: Timeseries observed discharge and simulated discharge for several values of infiltration capacity. The catchment is Bom Jardim.

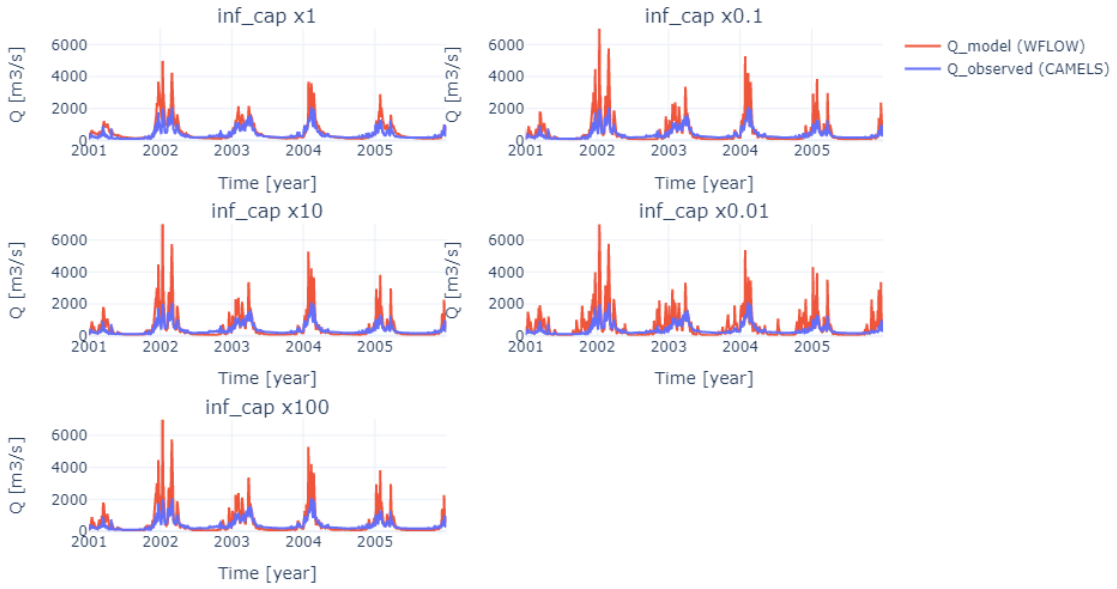

Figure 50: Timeseries observed discharge and simulated discharge for several values of infiltration capacity. The catchment is Cuiaba Paraguay.

C.2 Soil depth

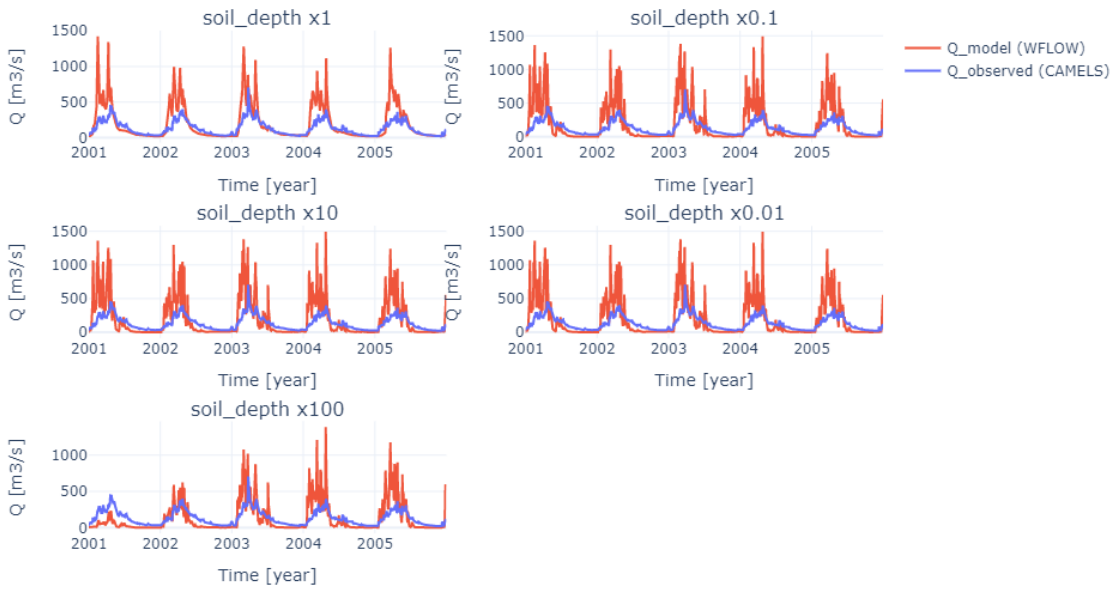

Figure 51: Timeseries observed discharge and simulated discharge for several values of soil depth. The catchment is Bom Jardim.

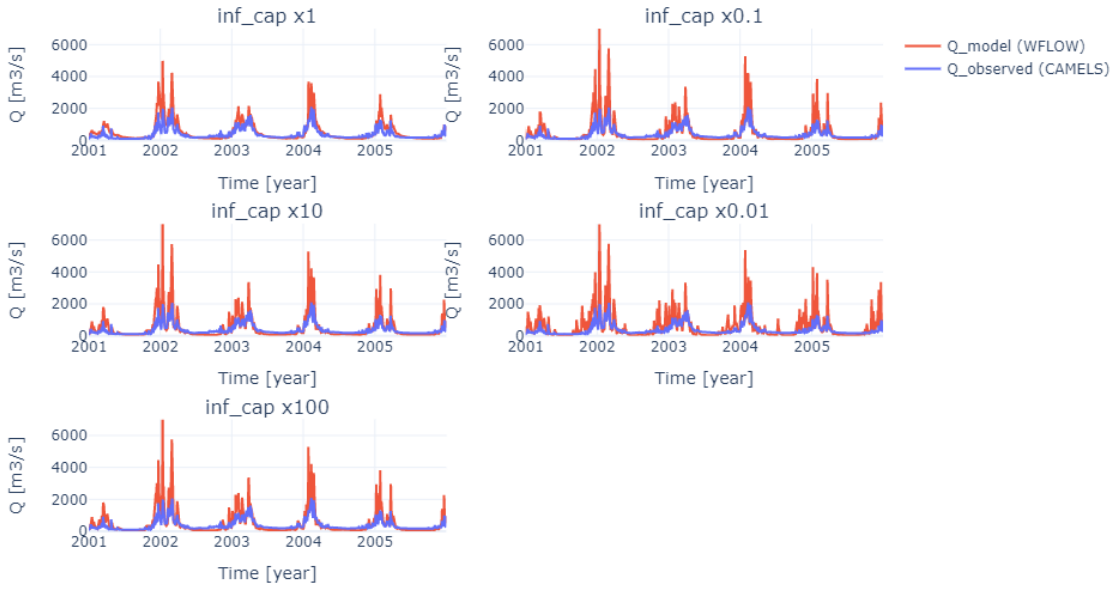

Figure 52: Timeseries observed discharge and simulated discharge for several values of soil depth. The catchment is Cuiaba Paraguay.

D

Model parameters

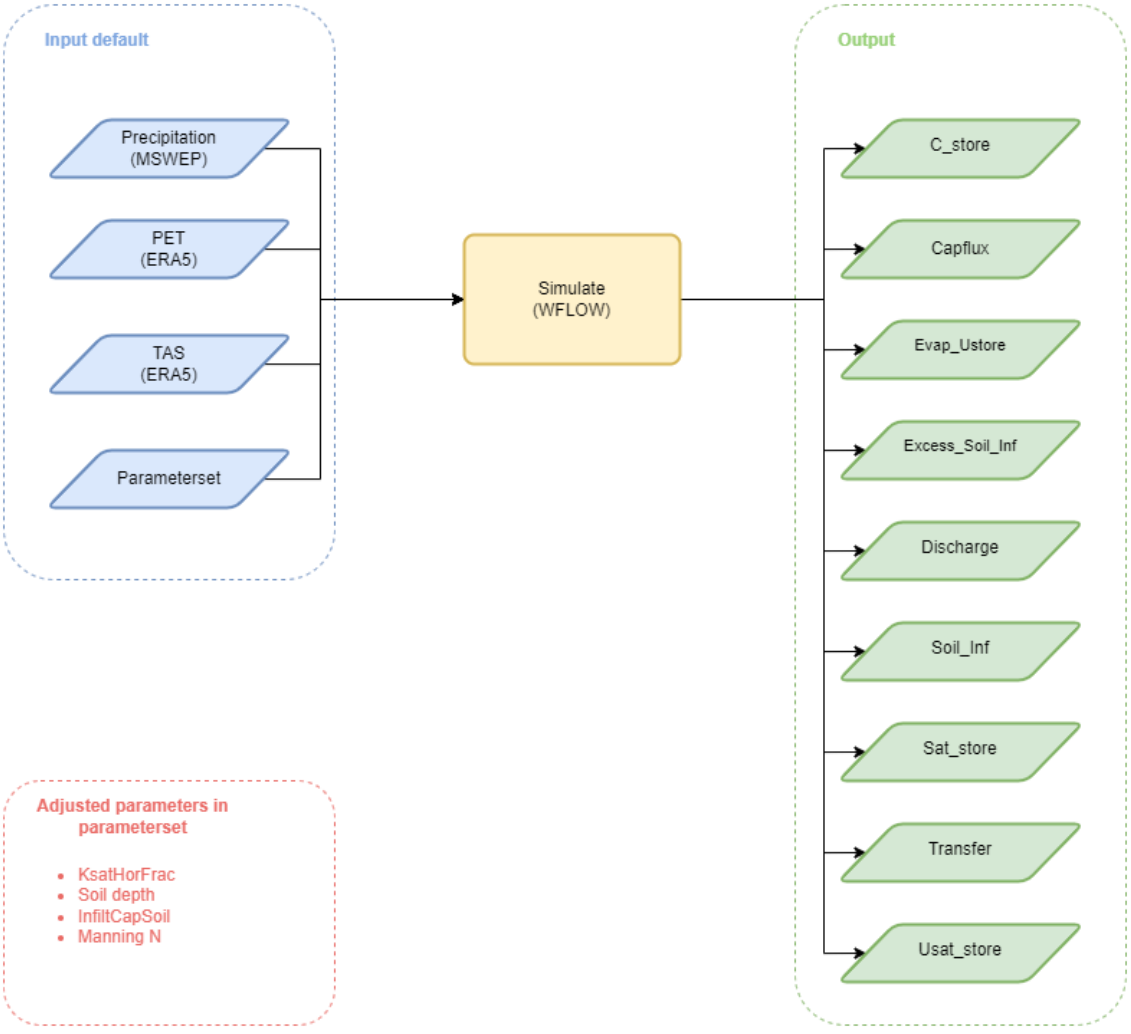

Figure 53: Overview of input and output for WFLOW this study. The defintion of the parameters for input and output are shown in Table 6 and Table 7.

List parameters in parameterset

Table 6: List containing all parameters in parameterset

Parameter	Definition
c	Brooks-Corey power coefficient [-]
Cf_soil	A reduction factor for the the maximum infiltration rate [-]
DEM	Digital elevation map
EoverR	Ratio of average wet canopy evaporation rate over average precipitation rate [-]
f	scaling parameter for s and r [mm-1]
LAI	Leaf area index
InfiltCapSoil	Infiltration capacityof the non-compacted soil fraction (unpaved area) of each gridcell [mm/d]
InfiltCapPath	Infiltration capacity of the compacted soil (or paved area) fraction of each gridcell [mm/d]
k	Extinction coefficient used for determining canopy gap fraction [-]
KsatHorFrac	Multiplication factor applied to KsatVer for horizontal saturated conductivity [-]
KsatVer	Vertical saturated conductivity of the store at the surface [mm/d]
M	Determines the decrease of vertical saturated conductivity with depth [mm]
MaxLeakage	Water lost to the model [mm/d]
N	Manning N parameter for the kinematic wave function for overland flow.
N_River	Manning N parameter for the kinematic wave function for river flow.
PathFrac	Fraction of compacted area per gridcell [-]
RootingDepth	Rooting depth of the vegetation [mm]
RiverSlope	Slope of river [mm/mm]
SI	Specific leaf storage
Slope	Slope of the area [mm/mm]
SoilDepth	Soil depth [mm]
θr	Residual water content [mm/mm]
θs	Water content at saturation [mm/mm]
WaterFrac	Fraction of water area per gridcell [-]
Wflow_gauges	Map with locations of gauges
Wflow_landuse	Map with landuse type per gridcell
Wflow_ldd	Map with local drainage direction
Wflow_river	Map with dimensions for rivers
Wflow_soil	Map with soil type per gridcell

List output parameters

Table 7: List containing all output parameters

Parameter	Definition
C_store	Canopy interception storage [mm]
Capflux	Capillary rise [mm/d]
Evap_Ustore	Soil evaporation from unsaturated zone [mm/d]
Excess_Soil_Inf	Excess from soil infiltration [mm/d]
Q	River discharge [mm/d]
Soil_Inf	Infiltration in soil [mm/d]
Sat_store	Depth of the saturated zone [mm]
Transfer	Flow between unsaturated and saturated zone [mm/d]
U_store	Depth of the unsaturated zone [mm]

E

Model logbook

The model logbook contains all records for the model runs in this thesis. Several parameters are changed during the modelling process. These parameters are described in Table 8. The actions and effects of changing parameters can be observed in the logbook of Table 9.

Table 8: Description of each parameter changed in the model process.

Parameter	Description	Default value
KsatHorFrac	A multiplication factor [-] applied to KsatVer for the horizontal saturated conductivity used for computing lateral subsurface flow.	100
ModelSnow	Enable or disable the modelling of snow	1
SoilInfRedu	The soil infiltration capacity can be adjusted in case the soil is frozen.	0
UsatLayer	The unsaturated layer can be split-up in different layers, by providing the thickness of three layers.	[100, 300, 800] mm
N_River	Manning N parameter for the kinematic wave function for overland and river flow. There are 8 different classes.	-
InfiltCapSoil	Infiltration capacity of the non-compacted soil fraction (unpaved area) of each gridcell.	600 mm/day
SoilDepth	Maximum soil depth	2000 mm
Scope	Time scope of the model run.	2000- 2020
MaxCanopy Storage	Canopy storage. Used in the Gash interception model.	mm
Whole_UST_Avail	the complete unsaturated storage is available for transpiration	0
Precipitation	Source for the input of precipitation in WFLOW	mm

Table 9: Model logbook

Date	Parameter	Action	Effect on output
10-02-2022	Default settings	-	Create base
10-02-2022	UsatLayer	Multiply depth all layer with factor 2	No visible effects
10-02-2022	UsatLayer	Multiply depth all layer with factor 2000	First two peakflow events are dampened, then returns to default peakflow.
11-02-2022	UsatLayer	Multiply depth all layer with factor 0.5	No visible effects
11-02-2022	UsatLayer	Multiply depth all layer with factor 0.01	No visible effects
11-02-2022	ModelSnow	Remove infiltration reduction for snow	No visible effects
11-02-2022	SoilInfRedu	Add infiltration reduction for frozen soil	Discharge increases overall
22-02-2022	KsatHorFrac	Change value to 1	Increases peakflow and decreases baseflow. Larger effect than 25
22-02-2022	KsatHorFrac	Change value to 25	Increases peakflow and decreases baseflow. Larger effect than 75
22-02-2022	KsatHorFrac	Change value to 75	Increases peakflow and decreases baseflow.
23-02-2022	KsatHorFrac	Change value to 250	Dampens peakflow and increases baseflow.
23-02-2022	KsatHorFrac	Change value to 500	Dampens peakflow and increases baseflow. Larger effect than 250
24-02-2022	KsatHorFrac	Change value to 750	Dampens peakflow and increases baseflow. Larger effect than 500
01-03-2022	KsatHorFrac	Change value to 1000	Dampens peakflow and increases baseflow. Larger effect than 750
01-03-2022	KsatHorFrac	Change value to 2000	Dampens peakflow and increases baseflow. Larger effect than 1000
02-03-2022	KsatHorFrac	Change value to 5000	Dampens peakflow and increases baseflow. Larger effect than 2000
02-03-2022	KsatHorFrac	Change value to 10.000	Dampens peakflow and increases baseflow. Larger effect than 5000
02-03-2022	KsatHorFrac	Change value to 20.000	Dampens peakflow and increases baseflow. Larger effect than 10.000
02-03-2022	KsatHorFrac	Change value to 50.000	Dampens peakflow and increases baseflow. Larger effect than 20.000
17-03-2022	Whole_UST_Avail	Set to 1; Complete unsaturated storage is available for transpiration	No visible effects
28-03-2022	N_River	Multiply all classes with factor 2	No visible effects
28-03-2022	N_River	Multiply depth all layer with factor 10	No visible effects
28-03-2022	N_River	Multiply depth all layer with factor 0.5	No visible effects
28-03-2022	N_River	Multiply depth all layer with factor 0.1	No visible effects
04-04-2022	InfiltCapSoil	Multiply with factor 10	No visible effects
04-04-2022	InfiltCapSoil	Multiply with factor 100	No visible effects
04-04-2022	InfiltCapSoil	Multiply with factor 1000	No visible effects
05-04-2022	InfiltCapSoil	Multiply with factor 0.1	No visible effects
05-04-2022	InfiltCapSoil	Multiply with factor 0.01	Extra discharge peaks during baseflow
06-04-2022	Soildepth	Multiply with factor 10	No visible effects
06-04-2022	Soildepth	Multiply with factor 100	First two peakflow events are dampened, then returns to default peakflow.
06-04-2022	Soildepth	Multiply with factor 0.1	No visible effects
11-04-2022	Soildepth	Multiply with factor 0.01	No visible effects
11-04-2022	MaxCanopy Storage	Multiply with factor 2	No visible effects
11-04-2022	MaxCanopy Storage	Multiply with factor 10	No visible effects
12-04-2022	InfiltCapSoil & Soildepth	Multiply both with factor 10	No visible effects
12-04-2022	InfiltCapSoil & Soildepth	Multiply both with factor 100	First two peakflow events are dampened, then returns to default peakflow.
12-04-2022	InfiltCapSoil & Soildepth	Multiply both with factor 0.1	No visible effects