

Satellite data in rainfall-runoff models

Exploring new opportunities for semi-arid, data-scarce river basins

Hulsman, P.

DOI

[10.4233/uuid:bb898ff7-2982-4a8a-bee5-4a1f02d526e6](https://doi.org/10.4233/uuid:bb898ff7-2982-4a8a-bee5-4a1f02d526e6)

Publication date

2021

Document Version

Final published version

Citation (APA)

Hulsman, P. (2021). *Satellite data in rainfall-runoff models: Exploring new opportunities for semi-arid, data-scarce river basins*. [Dissertation (TU Delft), Delft University of Technology].
<https://doi.org/10.4233/uuid:bb898ff7-2982-4a8a-bee5-4a1f02d526e6>

Important note

To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

Satellite data in rainfall-runoff models

Exploring new opportunities for semi-arid,
data-scarce river basins



Satellite data in rainfall-runoff models

Exploring new opportunities for semi-arid,
data-scarce river basins

Satellite data in rainfall-runoff models

Exploring new opportunities for semi-arid,
data-scarce river basins

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology,
by the authority of the Rector Magnificus, Prof.dr.ir. T.H.J.J. van der Hagen,
chair of the Board for Doctorates,
to be defended publicly on
Monday 1 March 2021 at 12:30 o'clock

by

Petra HULSMAN

Master of Science in Civil Engineering, Delft University of Technology, the
Netherlands
born in Leiden, the Netherlands.

This dissertation has been approved by the promotor.

Composition of the doctoral committee:

Rector Magnificus, Dr. M. Hrachowitz, Prof.dr.ir. H.H.G. Savenije,	chairperson Delft University of Technology, promotor Delft University of Technology, promotor
--	---

Independent members:

Prof.dr.ir. R. Uijlenhoet,	Delft University of Technology
Prof.dr.ir. S.C. Steele-Dunne,	Delft University of Technology
Prof.dr.ir. P. van der Zaag,	IHE Delft, Delft University of Technology
Dr. F. Fenicia,	EAWAG, Switzerland

Other members:

Prof.dr.ir. N.C. van de Giesen,	Delft University of Technology
---------------------------------	--------------------------------



Keywords: hydrological modelling, poorly gauged, semi-arid regions, satellite data

Printed by: Ipskamp Printing

Front & Back: Petra Hulsman

Copyright © 2021 by P. Hulsman

ISBN 978-94-6421-226-6

An electronic version of this dissertation is available at

<http://repository.tudelft.nl/>.

*The eye is the lamp of the body.
If your eyes are healthy,
your whole body will be full of light.*

Matthew 6:22, New International Version

Preface

This PhD research was part of the research project "Enhancing Water, Food and Energy Security in the Lower Zambezi" (ZAMSECUR). The overall objective of this project was to improve the water management in the Lower Zambezi basin by enhancing knowledge on the water resources of its mostly ungauged tributaries in Zimbabwe, Zambia and Mozambique. This project gave me many interesting opportunities such that I enjoyed being part of it!

One of the great opportunities was the people I got to work with within this project. This included Henry Zimba from Zambia, who focused his PhD research on the evaporation in miombo woodlands in Zambia, and Hubert Samboko from Zimbabwe, who focused his PhD research on using drones to derive river geometry information and stream velocities which can be used to estimate the discharge. In addition, multiple students were involved in this project including Felix Chabala from Zambia, and Ivar Abas, Sylvia van Doorn, Geerten van der Zalm and Jan van Engelenhoven from The Netherlands. I enjoyed getting to know them and being involved (some more than others) in their research.

Another great opportunity was the traveling experience. Every year, I participated with the WaterNet symposium to communicate my results with people involved in the project which included people working at local water authorities or universities. It was also interesting to get in touch with (non-)scientists who work in the same or similar regions. This symposium was hosted by a different Sub-Saharan country each year and included a field trip on the last day during which I learned some very interesting facts. For example, did you know there is a plant in Namibia called the welwitschia which can get more than 2000 years old? Also, did you know the Victoria Falls was not always located at its current position? The huge gorges downstream of the falls indicate its past locations.

All in all, I am thankful for having been part of the ZAMSECUR project and am curious how it will proceed in the near future!

*Petra Hulsman
The Hague, October 2020*

Summary

Throughout the world, many people have been affected by water related issues in the past, some more extreme than others. In this context, hydrological models have often been used to gain more insight into the situation and to limit negative impacts as much as possible. There are many different types of hydrological models with each their strengths and weaknesses, but all models need a certain amount of reliable data. However, many river basins throughout the globe are poorly gauged which means there are only limited reliable ground observations available. That is why satellite observations provide many interesting opportunities to fill this gap of which many are not yet explored. Therefore the goal of this research was to answer the following main research question: *What is the added value of satellite-based observations for hydrological modelling in a semi-arid, data-scarce river basin?*

This research focused on the Luangwa River in Zambia which is a large tributary of the Zambezi River and has a basin area of 159,000 km². This river basin is poorly gauged, mostly unregulated and sparsely populated. In this semi-arid region, the mean annual precipitation (970 mm yr⁻¹) is lower than the potential evaporation (1555 mm yr⁻¹). During the dry season, the river meanders between sandy banks while during the wet season from November to May it can cover flood plains several kilometres wide. A process-based distributed hydrological model with sub-grid heterogeneity was developed in this research and modified step-wise when exploring the added value of different satellite observations for different aspects within hydrological modelling.

First, the added value of satellite-based river water level, i.e. altimetry, from 18 virtual stations combined with total water storage observations according to GRACE (Gravity Recovery and Climate Experiment) for model calibration was assessed by testing different calibration strategies. As a benchmark, feasible model parameter sets were identified using traditional model calibration with observed river discharge data. The results indicated the flows were modelled well with the benchmark calibration strategy. Assuming no discharge data is available, the discharge was reproduced best when calibrating with respect to GRACE and altimetry if the modelled discharge was converted to stream levels using the Strickler-Manning equation and river geometry information, hence cross-section and river gradient, extracted from Google Earth. The identification of feasible parameter sets improved further when using more accurate cross-section data and when increasing the number of virtual stations used.

Second, the added value of satellite-based evaporation and GRACE data to increase the understanding of hydrological processes through step-wise model structure improvement and model calibration was assessed. For this purpose, the benchmark rainfall-runoff model was adjusted iteratively. It was shown that the benchmark model calibrated with respect to discharge, reproduced the discharge, basin-

average evaporation and total water storage well, but poorly modelled the evaporation in wetland dominated areas and the spatial pattern of the evaporation and total water storage. By testing five alternative model structures, it was shown that all variables, hence discharge, evaporation and total water storage, were represented better when including the process of upwelling groundwater flow from a distributed groundwater reservoir and calibrating with respect to all three variables simultaneously. These changes improved the spatial and temporal variability of the individual variables with the exception of the temporal variation of the basin-average total water storage.

Third, processes underlying long-term total water storage variations observed with GRACE were identified through data analysis and model hypothesis testing. The benchmark model did not reproduce the observed long-term storage variations. However, this was improved by using different forcing data and by incorporating regional groundwater export. The variation in the long-term annual maximum total water storage improved by using different forcing data sources, while the annual minima improved by allowing groundwater to seep into a deep groundwater layer during wet conditions from where it leaks out of the basin.

In addition, analysing satellite observations provides new insights into the hydrological system which can be implemented in hydrological models. In general, people often adapt to dry conditions, depending on how they perceived them, by applying drought coping mechanisms affecting the hydrological system. Comparing satellite observations to local perceptions allows us to increase our understanding of factors influencing local perceptions. This was illustrated by analysing the drought of 2019 in the Zambezi River Basin using multiple satellite observations to determine whether it was, as locally perceived, indeed the most extreme over at least 20 years. It was shown this differed depending on the data variable, drought characteristic and location within the basin. Data analysis indicated that it depends on the drought characteristic, the hydrological variable considered, and on the location within the basin. On the one hand, the drought of 2019 resulted in the lowest basin-averaged annual rainfall over at least 27 years, most severe local rainfall deficit in the central and north-western part of the basin for at least 25 years, and lowest reservoir level since 1995. However, on the other hand, the spatially averaged rainfall deficit, as well as locally in the north of the basin, was more severe in 2002. Also in 2004, the spatially averaged total storage deficit, as well as locally in the central part of the basin, was more severe. Similarly, the reservoir water level-based drought severity was more extreme in 2015.

Overall, satellite-based observations have been used successfully to improve our understanding of the hydrological processes in the data-scarce Luangwa river basin, to improve the hydrological model structure and to allow for more reliable parameter identifications in the absence of reliable discharge data. This research focused on a selection of satellite-based observations and hydrological model applications. In other words, there remain many opportunities yet to be explored!

List of symbols

α	Pearson correlation coefficient [-]
β	Shape parameter [-]
γ	Exponent [-]
δ	Coefficient of variation [-]
Δt	Time step [d]
ϵ	Fraction of histogram intersection between K and L
μ	Mean
σ	Standard deviation
ϕ_{mod}	Modelled map [-]
ϕ_{obs}	Observed map [-]
a	Rating curve parameter
A	River cross-section area [m ²]
b	Rating curve parameter
B	River bed width [m]
C_e	Transpiration coefficient [-]
C_{max}	Maximum upwelling groundwater [mm d ⁻¹]
d	Water depth [m]
D_D	Drought duration [months]
D_E	Euclidian distance over multiple variables [-]
D_I	Drought intensity [month ⁻¹]
D_S	Drought severity [-]
E_i	Interception evaporation [mm d ⁻¹]
E_{NS}	Nash-Sutcliffe efficiency [-]
$E_{\text{NS},Q}$	Nash-Sutcliffe efficiency of the discharge time-series [-]
$E_{\text{NS},\log Q}$	Nash-Sutcliffe efficiency of logarithm of the discharge time-series [-]
$E_{\text{NS},\text{FDC}}$	Nash-Sutcliffe efficiency of the flow duration curve [-]
$E_{\text{NS},\log\text{FDC}}$	Nash-Sutcliffe efficiency of the logarithm of the flow duration curve [-]
$E_{\text{NS},\text{AC}}$	Nash-Sutcliffe efficiency of the autocorrelation function of the discharge time-series [-]
E_p	Potential evaporation [mm d ⁻¹]
$E_{R,\theta}$	Relative error of variable θ [-]
$E_{R,\text{RCdry}}$	Relative error of the mean seasonal runoff coefficient during the dry periods [-]
$E_{R,\text{RCwet}}$	Relative error of the mean seasonal runoff coefficient during the wet periods [-]
$E_{R,\text{RLD}}$	Relative error of the rising limb density of the hydrograph [-]
$E_{R,\text{WL}}$	Spearman-Rank correlation coefficient [-]
E_{SP}	Spatial efficiency metric [-]
E_t	Plant transpiration [mm d ⁻¹]
h	Water level [m]
h_0	Reference water level [m]
I_{max}	Interception capacity [mm]
i	Mean channel slope [-]
i_1, i_2	River bank slopes [-]

k	River roughness [$\text{m}^{1/3} \text{s}^{-1}$]
K	Observed histogram [-]
K_f	Reservoir time scales for fast runoff [d]
K_s	Reservoir time scales for slow runoff [d]
K_{sd}	Reservoir time scales for slow, deep runoff [d]
L	Modelled histogram [-]
P	Precipitation [mm d^{-1}]
P_e	Effective precipitation [mm d^{-1}]
p_{HRU}	Areal weights within a grid cell [-]
$Q_{f,\text{tot}}$	Fast runoff combining all relevant HRUs [mm d^{-1}]
Q_f	Fast runoff for each relevant HRU [mm d^{-1}]
Q_L	Groundwater loss [mm d^{-1}]
Q_m	Total runoff [mm d^{-1}]
Q_s	Slow runoff [mm d^{-1}]
Q_{sd}	Deep groundwater flow [mm d^{-1}]
Q_{ss}	Shallow groundwater flow [mm d^{-1}]
R	Hydraulic radius [m]
R_f	Drainage to fast runoff component [mm d^{-1}]
R_{fl}	Delayed fast runoff [mm d^{-1}]
$R_{\text{GW,tot}}$	Upwelling groundwater combining all relevant HRUs [mm d^{-1}]
R_{GW}	Upwelling groundwater for each relevant HRU [mm d^{-1}]
$R_{r,\text{tot}}$	Groundwater recharge combining all relevant HRUs [mm d^{-1}]
R_r	Groundwater recharge for each relevant HRU [mm d^{-1}]
R_s	Groundwater recharge into deeper groundwater reservoir [mm d^{-1}]
R_u	Infiltration into the unsaturated zone [mm d^{-1}]
S_f	Storage in fast reservoir [mm]
S_i	Storage in interception reservoir [mm]
S_s	Storage in groundwater/slow reservoir [mm]
$S_{s,\text{ref1}}$	Reference groundwater level [m]
$S_{s,\text{ref2}}$	Reference groundwater level [m]
S_{su}	Storage in upper groundwater reservoir [mm]
S_{sd}	Storage in deeper groundwater reservoir [mm]
S_u	Storage in unsaturated root zone [mm]
$S_{s,\text{ref}}$	Reference storage in the saturated zone [mm]
$S_{u,\text{max}}$	Maximum root zone storage capacity [mm]
T_{lag}	Time lag [d]
v	River velocity [m s^{-1}]
W	Splitter [-]
W_s	Groundwater splitter [-]

List of abbreviations

ADCP	Acoustic Doppler Current Profiler
CHIRPS	Climate Hazards Group InfraRed Precipitation with Station
CMRSET	CSIRO MODIS Reflectance Scaling EvapoTranspiration
CNES	Centre national d'études spatiales
CRU	Climatic Research Unit
CSIRO	Commonwealth Scientific and Industrial Research Organisation
CSR	Centre for Space Research
DAHITI	Database for Hydrological Time Series of Inland Waters
DWA	Department of Water Affairs in Zambia
EAPRS	Earth and Planetary Remote Sensing Lab
ECMWF	European Centre for Medium-Range Weather Forecasts
ERA5	5th generation ECMWF atmospheric reanalysis dataset
ESA	European Space Agency
FAO	Food and Agriculture Organization
GEOS	Goddard Earth Observing System Model
GFZ	GeoForschungsZentrum Potsdam
GLEAM	Global Land Evaporation Amsterdam Model
GMTED	Global Multi-resolution Terrain Elevation Data
GR4J	Modèle du Génie Rural à 4 paramètres Journalier
GRACE	Gravity Recovery and Climate Experiment
GRDC	Global Runoff Data Centre
HBV	Hydrologiska Byråns Vattenbalansavdelning model
HRU	Hydrological Response Unit
ISRO	Indian Space Research Organisation
IAHS	International Association of Hydrological Sciences
JPL	Jet Propulsion Laboratory
LAI	Leaf Area Index
LEGOS	Laboratoire d'Etudes en Géophysique et Océanographie Spatiales
MERRA	Modern-Era Retrospective analysis for Research and Applications
mHM	Mesoscale Hydrologic Model
MOD16	MODIS Global Evapotranspiration Project
MODIS	Moderate Resolution Imaging Spectroradiometer
NASA	National Aeronautics and Space Administration
NDII	Normalized Difference Infrared Index
NDVI	Normalized Difference Vegetation Index
NOAA	National Oceanic and Atmospheric Administration
PUB	Predictions in Ungauged Basins
RMSE	Root mean square error
SEBS	Surface Energy Balance System
SPI	Standardized Precipitation Index

SSEBop	Operational Simplified Surface Energy Balance
SPOT	Satellite Pour l'Observation de la Terre
SRTM	Shuttle Radar Topography Mission
SWAT	Soil & Water Assessment Tool
SWOT	Surface Water Ocean Topography
TRMM	Tropical Rainfall Measuring Mission
TSD	Total Storage Deficit
TSDI	Total Storage Deficit Index
VIC	Variable Infiltration Capacity model
WaPOR	Water Productivity Open Access Portal
WARMA	Water Resources Management Authority in Zambia
WLDI	Water-Level Deficit Index
ZMD	Zambia Meteorological Department

Contents

Preface	ix
Summary	xi
List of symbols	xiii
List of abbreviations	xv
1 Introduction	1
1.1 Hydrological modelling	2
1.2 Poorly gauged river basins	3
1.3 Opportunities with satellite observations.	3
1.4 Research objectives	4
2 Study area: Luangwa river basin	9
2.1 Site description	10
2.2 Floods & droughts	10
2.3 Ground observations.	11
3 Altimetry-based parameter set identification	13
3.1 Introduction	14
3.2 Site description	16
3.3 Data availability	16
3.4 Hydrological model development.	21
3.5 Results and discussion	32
3.6 Evaluation	46
4 Model structure development using satellite observations	49
4.1 Introduction	50
4.2 Site description	51
4.3 Data availability	51
4.4 Modelling approach	53
4.5 Model results	61
4.6 Discussion.	74
4.7 Evaluation.	75
5 Long-term total water storage variations in the Luangwa basin	77
5.1 Introduction	78
5.2 Site description	79
5.3 Data availability	81
5.4 Approach	81
5.5 Results	91

5.6 Discussion	106
5.7 Evaluation	109
6 Sub-Saharan drought of 2019 observed from space	111
6.1 Introduction	112
6.2 Site description	113
6.3 Data availability	114
6.4 Approach	115
6.5 Results	117
6.6 Discussion	124
6.7 Evaluation	125
7 Conclusions and outlook	127
A Appendix	135
B Appendix	145
C Appendix	157
References	169
Acknowledgements	195
Curriculum Vitæ	197
List of Publications	199

1

Introduction

*Do not insult a crocodile
while your feet are still in the water.*

African proverb

This chapter is based on:

Hulsman, P., Winsemius, H. C., Michailovsky, C. I., Savenije, H. H. G., and Hrachowitz, M.: Using altimetry observations combined with GRACE to select parameter sets of a hydrological model in a data-scarce region, *Hydrol. Earth Syst. Sci.*, 24, 3331–3359, doi: 10.5194/hess-24-3331-2020, 2020a.

Hulsman, P., Savenije, H. H. G., and Hrachowitz, M.: Learning from satellite observations: increased understanding of catchment processes through stepwise model improvement, *Hydrol. Earth Syst. Sci. Discuss.*, 2020, 1–26, doi: 10.5194/hess-2020-191, 2020b.

Hulsman, P., Savenije, H. H. G., and Hrachowitz, M.: Why are long-term storage variations observed but not modelled in the Luangwa basin?, *Water Resources Research*, in review.

Hulsman, P., Savenije, H. H. G., and Hrachowitz, M.: Zambezi River Basin: Drought of 2019, *Journal of Hydrology: Regional Studies*, submitted.

1.1. Hydrological modelling

Throughout the globe, water related issues such as droughts or floods have affected many people. According to “The International Disasters Database” at least 460 million people have been affected world-wide by riverine floods in 2010 – 2019 and 671 million people by droughts (EM-DAT and CRED, Accessed 2020). This number is expected to increase in the future with climatic and socio-economic changes (Di Baldassarre et al., 2017; Hallegatte et al., 2013; Markus et al., 2019). In this context, hydrological models have often been used to gain more insight into the situation and to limit negative impacts as much as possible. For instance, hydrological models have been important tools to predict floods (Kauffeldt et al., 2016; Price et al., 2012) and droughts (Hao et al., 2018; Trambauer et al., 2013), but also to assess the effect of climate (Jahandideh-Tehrani et al., 2019; Krysanova and Hattermann, 2017) and land-cover changes (Cuo, 2016; Nijzink et al., 2016a) on the water availability.

In the past, many different types of hydrological models have been developed. Some of the commonly used rainfall-runoff models are HBV (Bergström, 1992), SUPERFLEX (Fencia et al., 2011), FLEX-Topo (Gao et al., 2014a), mHM (Samaniego et al., 2010), SWAT (Arnold et al., 1998), VIC (Liang et al., 1994) and MIKE-SHE (Refsgaard et al., 2010) to name a few. These models can be classified based on for example the simplification strategy of the hydrological system (for example conceptual or physically-based), spatial representation (lumped, semi-distributed or distributed), model architecture (continuum or bucket-based) or model scaling strategy (bottom-up or top-down) (Hrachowitz and Clark, 2017; Pechlivanidis et al., 2011). In other words, there is an overabundance of hydrological models each with their strengths and weaknesses without knowing whether there is a “correct” model among them (Clark et al., 2011).

Within hydrological modelling, there are various challenges including 1) uncertainties in data, model structure and model parameters (Bourdin et al., 2012; Pechlivanidis et al., 2011), 2) equifinality problems as a result of a high number of parameters and too limited data to identify feasible parameter combinations representing the hydrological system well (Beven, 2006; Savenije, 2001), and 3) changing hydrological conditions as a result of climate changes or human actions such as land-use change or river regulations (Peel and Blöschl, 2011). To deal with these challenges, sufficient good quality ground observations are required. However, in many poorly gauged river basins there are no or only very limited ground observations available (see next section) such that these challenges become even more difficult to tackle resulting in more uncertain predictions.

1.2. Poorly gauged river basins

In many river basins world wide, there are only limited or no river discharge observations available such that they are poorly gauged or ungauged. According to the Global Runoff Data Centre (GRDC, https://www.bafg.de/GRDC/EN/Home/homepage_node.html), 9952 discharge stations exist throughout the globe of which merely 208 stations provide data after 2010 which are mostly located in the United States and Europe. In large parts of Africa and Asia on the other hand, there are no discharge observations available after 2000 in the GRDC database. Please note such online databases are not necessarily complete as local water authorities often have more recent discharge time-series. However, this does illustrate the large scale of poorly gauged river basins especially in Africa, Asia and some regions in South-America.

This challenge was the focus of the research initiative "Predictions in Ungauged Basins" (PUB) in 2003 – 2012, initiated by the International Association of Hydrological Sciences (IAHS). According to the PUB initiative, high uncertainties in model predictions were a result of incomplete understanding of hydrological processes, spatial-temporal heterogeneity of processes and unsuitable regionalization techniques (Hrachowitz et al., 2013). Therefore, the goal of this initiative was to improve hydrological model predictions in ungauged basins by improving existing models and by developing new models better reproducing the spatial-temporal variability of hydrological processes (Hrachowitz et al., 2013). During the PUB decade, advances were made in the fields of 1) data collection and processing, 2) modelling strategies, calibration techniques and uncertainty assessment, and 3) catchment classification methods. In the PUB synthesis book (Blöschl et al., 2013) all findings of the PUB decade on predicting runoff signatures are organized. Even though this PUB decade was completed in 2012, there remain many research opportunities unexplored in the context of PUB to obtain reliable and robust predictions in ungauged basins.

1.3. Opportunities with satellite observations

Remote sensing products offer increasingly new opportunities in hydrological modelling, especially in ungauged or poorly gauged basins, as they offer large-scale spatially distributed observations on a regular temporal scale (Lakshmi, 2004; Xu et al., 2014). For example, satellite data have been used to estimate precipitation, evaporation, soil moisture, snow cover, water level, land surface temperature, river width and total water storage (Jiang and Wang, 2019). In the context of hydrological modelling, previous studies used satellite products among others to 1) describe the basin characteristics such as landscape or land-cover (e.g. Fenicia et al., 2016; Kiptala et al., 2013; Savenije, 2010), 2) provide forcing data such as precipitation (Meier et al., 2011; Serrat-Capdevila et al., 2014; Winsemius et al., 2006a), 3) estimate parameters (Gao et al., 2014b; Wang-Erlandsson et al., 2016), or 4) calibrate models (e.g. Immerzeel and Droogers, 2008; Kunnath-Poovakka et al., 2016; Santhi et al., 2008). Despite the increased use of satellite observations for

hydrological modelling, many opportunities remain unexplored. For instance, the full potential of satellite-based river water level observations from multiple missions and at multiple locations for model calibration is yet to be exploited. In addition, only limited studies used satellite observations to calibrate hydrological models with respect to the temporal *and* spatial variability simultaneously (Dembélé et al., 2020; Rajib et al., 2018). Also, spatial information available in satellite data have rarely been used for step-wise model structure improvement (Fenicia et al., 2016; Roy et al., 2017).

1.4. Research objectives

This research mainly focused on the semi-arid Luangwa river in Zambia which is a large, poorly gauged tributary of the Zambezi in Sub-Saharan Africa. Accurate estimations of the water availability are important for water allocation planning especially during dry seasons (Hamududu and Ngoma, 2020), whereas reliable flow estimations are important for the management of the Cahora Bassa dam downstream (Gumindoga et al., 2016; Winsemius et al., 2008). However, the ability to estimate flows and water availability accurately is impeded by the limited ground observations available (see Chapter 2 for more information). As a result, the Luangwa river basin is an interesting study region to analyse the added value of satellite observations to increase our understanding of local hydrological processes and to improve hydrological model predictions despite the limited availability of ground observations similar to many other poorly gauged river basins in the world. Hence, the main research question was:

What is the added value of satellite-based observations for hydrological modelling in a semi-arid, data-scarce river basin?

Several new opportunities of satellite observations in hydrological modelling not yet fully explored in previous studies have been analysed in this research. For this purpose, a selection of satellite observations have been used to improve our understanding of local hydrological processes through step-wise model improvement and to calibrate hydrological models considering the temporal and spatial variability. As a result, the main research question was divided into the following sub-questions:

1. What is the combined information content of satellite-based river water level and total water storage observations to identify feasible parameter sets?

In the absence of reliable discharge data as commonly the case in poorly gauged regions, alternative calibration methods are required. In a previous study, it was shown hydrological models can be calibrated with respect to river water levels if uncertain or no rating curve information is available (Hulsman et al., 2018). In that study, Hulsman et al. (2018) developed a semi-distributed hydrological model for the Mara river basin in Kenya and used the Strickler-Manning equation to convert modelled discharges to river water levels calibrating the slope-roughness parameter. This study illustrated that with this calibration method river water levels can be

reproduced well and discharge-water level relations can be derived, even in basins with unreliable rating curve information.

Following up on that study, satellite-based river water level observations, i.e. altimetry, provide an interesting opportunity to calibrate hydrological models in poorly gauged basins as they are available world-wide, near real-time and at multiple locations along the river depending on the river size and location (Calmant et al., 2009; Musa et al., 2015; Schwatke et al., 2015). In previous studies, altimetry observations have been used to estimate discharge, calibrate/evaluate models and for data assimilation typically using altimetry data from only a few locations obtained from a single satellite mission (Getirana, 2010; Michailovsky et al., 2012, 2013; Sun et al., 2012; Tourian et al., 2017). However, previous studies did not use altimetry observations from multiple locations and satellite missions for model calibration comparing different calibration strategies. As altimetry observations only describe water level dynamics without any information on discharge amounts, total water storage estimates according to the Gravity Recovery and Climate Experiment (GRACE) were used to support model calibration to reduce the uncertainty in modelled discharge arising from the missing information on flow amounts (Bai et al., 2018; Rakovec et al., 2016). Therefore, the objective was to explore the information content of altimetry observations and total water storage data according to GRACE for the calibration of a hydrological model for the Luangwa basin comparing different strategies. This was discussed in Chapter 3.

2. What is the added value of satellite-based evaporation and total water storage data to increase the understanding of hydrological processes through step-wise model structure improvement and model calibration for a large river system in a semi-arid, data scarce region?

In the Luangwa Basin, there is a strong spatial heterogeneity in the landscape, land cover and precipitation resulting in spatially varying rainfall-runoff processes. Traditionally, hydrological models are developed and calibrated using discharge observations at the basin outlet only which can result in robust discharge predictions in small catchments (Daggupati et al., 2015; Fenicia et al., 2011). However, for large and heterogeneous river basins this could result in poor representations of the spatial-temporal variability of model internal fluxes and states (Clark et al., 2008; Garavaglia et al., 2017; Hrachowitz et al., 2014; Kirchner, 2006). In this context, satellite-based spatial-temporal observations provide interesting opportunities for model development and calibration to improve the overall representation of the hydrological system. Many previous studies have used satellite observations for model calibration considering either the temporal or spatial variability (e.g. Demirel et al., 2018; Kittel et al., 2018; Koch et al., 2016; Nijzink et al., 2018; Zink et al., 2018), whereas only limited studies combined both aspects in the calibration procedure (Dembélé et al., 2020; Rajib et al., 2018). In addition, many studies have used spatial aggregated observations such as discharge to improve the model structure (e.g. Fenicia et al., 2008; Hrachowitz et al., 2014; Kavetski and Fenicia, 2011), while only limited studies used spatial information for this purpose (e.g. Fenicia et al., 2016; Roy et al., 2017). Therefore, the objective was to explore the added value

of satellite-based evaporation and total water storage observations for model development and spatial-temporal model calibration. This was discussed in Chapter 4.

3. Why are long-term total water storage variations observed in the Luangwa basin with satellite data but not modelled?

In the Luangwa river basin, long-term fluctuations were observed in the total water storage similar to many river basin world-wide (Long et al., 2017; Scanlon et al., 2018). However, this was not reproduced by a standard conceptual model encapsulating our current understanding of the regional hydrological processes. Similarly, previous studies highlighted inconsistencies between the observed and modelled long-term total water storage fluctuations in different river basins using various rainfall-runoff models (Fowler et al., 2020; Scanlon et al., 2018; Winsemius et al., 2006b). While several studies focused on identifying differences between observed and modelled long-term storage variations (e.g. Jing et al., 2019; Leblanc et al., 2009; Scanlon et al., 2018), only limited studies attempted to modify a hydrological model to reproduce long-term storage variations (e.g. Grigg and Hughes, 2018). Fowler et al. (2020) even concluded that commonly used conceptual hydrological models cannot reproduce long-term storage variations as they lack long-term memory processes and hence should not be used for discharge predictions in for example drying climates. Therefore, the objective was to identify processes underlying the observed long-term storage variations in the Luangwa basin in a combined data analysis and model hypothesis testing approach. This was discussed in Chapter 5.

With these three sub-questions, several new opportunities were explored using satellite observations to improve our current understanding of local dominant hydrological processes, improve hydrological model structures, and improve the identification of feasible parameter sets in a data-scarce region. In addition, satellite-based data analysis results provide valuable new information to improve our understanding of the hydrological system which can then be implemented in hydrological models. For example, satellite observations can be used to analyse drought events in large regions with respect to the temporal and spatial variability (e.g. Bayissa et al., 2018; Hao and Singh, 2015; Mishra and Singh, 2010). In general, drought events can influence human activities with respect to drought coping mechanisms, depending on how they perceived specific drought events, which can affect the hydrological system (Haile et al., 2019; Iqbal et al., 2018; Van Loon et al., 2016). That is why it would be interesting to compare local perceptions with multiple satellite observations to increase our understanding of factors influencing local perceptions. As an illustration, multiple satellite observations were used to analyse droughts in the Zambezi River Basin with the following research question:

4. Was the drought of 2019 in the Zambezi River Basin the most extreme in at least 20 years according to multiple satellite observations as perceived by locals?

During the dry season of 2019, extreme low water levels were observed in the reservoir upstream of the Kariba hydro-power dam resulting in frequent power cuts of up to 18 hours for at least 3 months (Carlowicz, 2019; Matiashe, 2019;

Tshili, 2019). According to popular news media, locals perceived this drought as the worst in several decades (Brown, 2019; Edel, 2019). Previous studies illustrated discrepancies can occur between local perceptions and data observations for different reasons including economic issues as farmers who suffered crop losses were more inclined to perceive decreased rainfall even when that was not the case (Albertini et al., 2020; Di Baldassarre et al., 2017; Foguesatto et al., 2020). While previous studies compared local perceptions of drought events to rainfall observations (e.g. Giordano et al., 2013; Iqbal et al., 2018; Osgood et al., 2018; Ovuka and Lindqvist, 2000; Solano-Hernandez et al., 2020), they did not incorporate satellite-based total water storage and reservoir water level observations which provide additional information on drought events. Therefore, the objective was to analyse the drought of 2019 in the Zambezi River Basin upstream of the Kariba reservoir using satellite-based precipitation, total water storage and reservoir water level observations to determine whether it was indeed the most extreme drought in at least 20 years as perceived by locals. This was discussed in Chapter 6.

2

Study area: Luangwa river basin

*If you have escaped the jaws
of the crocodile while bathing in the river,
you will surely meet a leopard on the way.*

African proverb

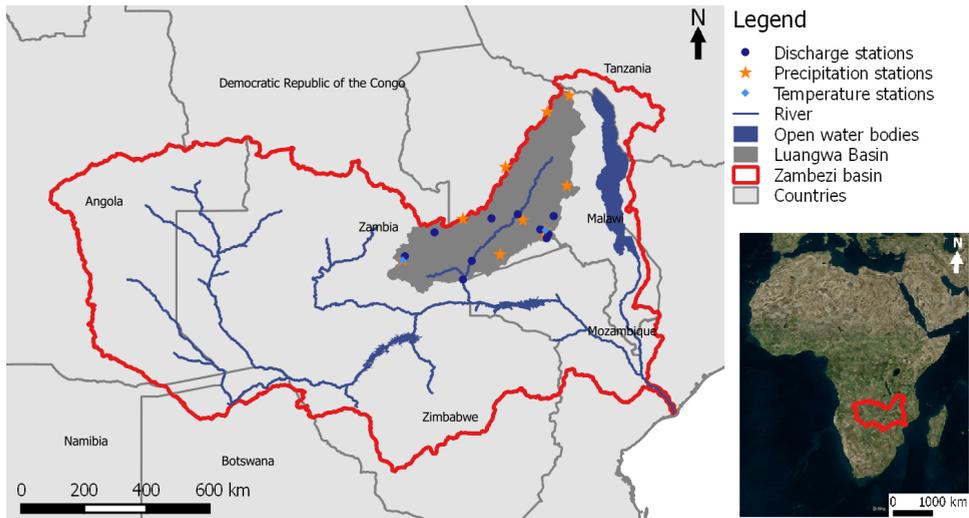


Figure 2.1: Map of the Luangwa River Basin in Zambia including available in-situ hydro-meteorological stations.

2.1. Site description

The Luangwa River in Zambia, a large tributary of the Zambezi River (Figure 2.1), has a basin area of 159,000 km² which is about 10% of the Zambezi River Basin. The Luangwa Basin is poorly gauged, mostly unregulated and sparsely populated with about 1.8 million inhabitants in 2005 (The World Bank, 2010). The mean annual precipitation is around 970 mm yr⁻¹ and potential evaporation around 1555 mm yr⁻¹ (The World Bank, 2010). The main land cover consists of broad-leaf deciduous forest (55%), shrub land (25%) and savanna grassland (16%) (GlobCover, 2009). The irrigated area in the basin is limited to about 180 km², i.e. roughly 0.1% of the basin area with an annual water abstraction of about 0.8 mm yr⁻¹ averaged over the entire basin which amounts to 0.6% of the annual basin water balance (The World Bank, 2010). The landscape varies between low lying flat areas along the river to large escarpments mostly in the North West of the basin and highlands with an elevation difference up to 1850 m. During the dry season, the river meanders between sandy banks while during the wet season from November to May it can cover flood plains several kilometres wide.

2.2. Floods & droughts

The Luangwa drains into the Zambezi downstream of the Kariba Dam and upstream of the Cahora Bassa Dam. The operation of both dams is crucial for hydropower production, and flood and drought protection, but is very difficult due to the lack of information from poorly gauged tributaries such as the Luangwa (SADC, 2008; Schleiss and Matos, 2016; The World Bank, 2010). As a result of high rainfall

variability and poor dam management, the local population suffered from severe floods and droughts (Beilfuss et al., 2001; Hanlon, 2001; SADC, 2008; Schumann et al., 2016; ZAMCOM et al., 2015). For example, severe droughts have occurred in 1995 and 2015 in multiple Sub-Saharan countries causing for instance reduced crop production (Libanda et al., 2019; The World Bank, 2010). Between 2000 and 2009, about 12.5 million people were affected by droughts in Mozambique, Zambia and Zimbabwe (ZAMCOM et al., 2015). Limited water availability is expected to become a more severe challenge due to population growth, increased water demand for instance for irrigation, climate change and poor drought management planning to cope with these changes (Kirchhoff and Bulkley, 2008; SADC, 2008; Schleiss and Matos, 2016). In addition, about 9.5 million people were affected by floods between 2000 and 2009 in Mozambique, Zambia and Zimbabwe (ZAMCOM et al., 2015). According to previous studies, high flood risks were a result of poor flood management planning, weak infrastructure, poor communication facilities and poor real-time information exchange for flood warning systems (Schumann et al., 2016; ZAMCOM et al., 2015).

To improve the flood and drought prediction in the Zambezi river basin, increased understanding of the local hydrological processes is very important. This research focused on the Luangwa river as it is a large and poorly gauged tributary flowing into the Zambezi upstream of the Cahora Bassa dam. Improving the flow prediction for the Luangwa river would contribute to a more accurate inflow prediction for the Cahora Bassa reservoir which is valuable to improve its management.

2.3. Ground observations

In the Luangwa River Basin, daily hydro-meteorological observations were available for the discharge, precipitation and temperature. In total, data was available for 11 gauge stations, 10 precipitation stations and 2 temperature stations. Their locations are visualised in Figure 2.1 and their characteristics are summarized in Table 2.1. Discharge data was obtained from the Global Runoff Data Centre (GRDC), the local Department of Water Affairs (DWA) and Water Resources Management Authority in Zambia (WARMA), precipitation data from the Zambia Meteorological Department (ZMD) and the National Oceanic Atmospheric Administration (NOAA) from where also temperature was obtained.

The discharge stations were located in the middle and lower part of the basin of which 3 stations were located in the main river. The in-situ discharge observations were available since 1948, but only limited data was available since 2002 resulting in a temporal coverage for the time period 2002 – 2016 between 0% and 55% (Table 2.1). One gauge station, the Luangwa Road Bridge gauging station, was located near the confluence with the Zambezi river and had the largest temporal coverage of 55% for the time period 2002 – 2016. In the following chapters, discharge data from this station was used for model calibration and/or validation. The precipitation and temperature stations were mostly located near the basin border resulting in only limited information throughout the basin. Their temporal coverage for the time period 2002 – 2016 varied between 0% and 51% (Table 2.1). Especially with respect to temperature only limited information was available after

2002.

For an improved understanding of the hydrological processes, it is crucial to capture the spatiotemporal heterogeneity in the precipitation and temperature well. Especially with respect to precipitation, very local rainfall occurs frequently in the Luangwa basin as observed by local inhabitants and confirmed through field visits. Such local rain events cannot be observed by merely 10 precipitation stations for an area of 159,000 km². In addition, in the context of flood and drought protection, it is more interesting to improve our understanding on the current situation rather than the historic situation. That is why it is interesting to explore the added value of satellite observations to improve our understanding of regional hydrological processes in the Luangwa basin.

Table 2.1: Summary ground-observations available in the Luangwa basin

	Number of stations	Duration	Time period	Temporal coverage in 2002 – 2016
Discharge	11	16 – 68 years	1948 – 2017	0 – 55%
Precipitation	10	5 – 42 years	1970 – 2011	0 – 51%
Temperature	2	30 – 44 years	1970 – 2014	0 – 1%

3

Altimetry-based parameter set identification

Smooth seas do not make skillful sailors.

African Proverb

This chapter investigated the potential of using remotely sensed river water level, i.e. altimetry observations, from multiple satellite missions to identify parameter sets for a hydrological model in the semi-arid Luangwa River Basin in Zambia. As a benchmark, feasible model parameter sets were identified using traditional model calibration with observed river discharge data. For the parameter identification using remote sensing, data from the Gravity Recovery and Climate Experiment (GRACE) were used in a first step to restrict the feasible parameter sets based on the seasonal fluctuations in total water storage. Next, three alternative ways of further restricting feasible model parameter sets using satellite altimetry time-series from 18 different locations along the river were compared. These three strategies compare altimetry observations to 1) modelled discharge by applying the Spearman Rank Correlation coefficient, and to modelled stream levels by converting modelled discharge using 2) rating curves whose parameters were treated as free model calibration parameters and 3) the Strickler-Manning equation to infer water levels directly from hydraulic properties of the river.

This chapter is based on: Hulsman, P., Winsemius, H. C., Michailovsky, C. I., Savenije, H. H. G., and Hrachowitz, M.: Using altimetry observations combined with GRACE to select parameter sets of a hydrological model in a data-scarce region, *Hydrol. Earth Syst. Sci.*, 24, 3331–3359, doi: 10.5194/hess-24-3331-2020, 2020a.

3.1. Introduction

Reliable models of water movement and distribution in terrestrial systems require sufficient good quality hydro-meteorological data throughout the modelling process. However, the development of robust models is challenged by the limited availability of ground measurements in the vast majority of river basins world-wide (Hrachowitz et al., 2013). Therefore, modellers increasingly resort to alternative data sources such as satellite data (Dembélé et al., 2020; Demirel et al., 2018; Lakshmi, 2004; Nijzink et al., 2018; Pechlivanidis and Arheimer, 2015; Rakovec et al., 2016; Sun et al., 2018a; Winsemius et al., 2008; Zink et al., 2018).

In the absence of directly observed river discharge data, various types of remotely sensed variables provide valuable information for the calibration and evaluation of hydrological models. These include, for instance, remotely sensed time series of river width (Sun et al., 2012, 2015a), flood extent (Montanari et al., 2009; Revilla-Romero et al., 2015), or river and lake water levels from altimetry (Garambois et al., 2017; Getirana, 2010; Getirana et al., 2009; Pereira-Cardenal et al., 2011; Sun et al., 2012; Velpuri et al., 2012).

Satellite altimetry observations provide estimates of the water level relative to a reference ellipsoid. For these observations, a radar signal is emitted from the satellite in the nadir direction and reflected back by the earth surface. The time difference between sending and receiving this signal is then used to estimate the distance between the satellite and the earth surface. As the position of the satellite is known at very high accuracy, this distance can then be used to infer the surface level relative to a reference ellipsoid (Calmant et al., 2009; Łyszkowicz and Bernatowicz, 2017). Satellite altimetry is sensed and recorded along the satellite's track. Altimetry based water levels can therefore only be observed where these tracks intersect with open-water surfaces; for rivers, these points are typically referred to as "virtual stations" (Birkett, 1998; de Oliveira Campos et al., 2001; Jiang et al., 2017; Schneider et al., 2017; Seyler et al., 2013). Depending on the satellite mission, the equatorial inter-track distance can vary between 75 km and 315 km, the along-track distance between 173 m and 374 m, and the temporal resolution between 10 days and 35 days (CNES, Accessed 2018; ESA, Accessed 2018; Schwatke et al., 2015; Łyszkowicz and Bernatowicz, 2017). Due to this rather coarse resolution, the application of remotely sensed altimetry data is at this moment limited to large lakes or rivers of more than approximately 200 m wide (Biancamaria et al., 2017; de Oliveira Campos et al., 2001; Getirana et al., 2009). Use of altimetry for hydrological models so far also remains rather rare due to the relatively low temporal resolution of the data, with applications typically limited to monthly or longer modelling time steps (Birkett, 1998).

In some previous studies, altimetry data were used to estimate river discharge at virtual stations in combination with routing models (Michailovsky and Bauer-Gottwein, 2014; Michailovsky et al., 2013) or stochastic models (Tourian et al., 2017). Other studies either directly related river altimetry to modelled discharge (Getirana and Peters-Lidard, 2013; Getirana et al., 2009; Leon et al., 2006; Paris et al., 2016) or they relied on rating curves developed with water level data from either in-situ measurements (Michailovsky et al., 2012; Papa et al., 2012; Tarpanelli

et al., 2013, 2017) or, alternatively, from altimetry data (Kouraev et al., 2004). In typical applications, radar altimetry data from one single or only a few virtual stations were used for model calibration, validation or data assimilation. These data were mostly obtained from a single satellite mission, either TOPEX/Poseidon or Envisat (Bauer-Gottwein et al., 2015; Fleischmann et al., 2018; Getirana, 2010; Liu et al., 2015; Michailovsky et al., 2013; Pedinotti et al., 2012; Sun et al., 2012). In previous studies, hydrological models have been calibrated or validated successfully with respect to (satellite based) river water levels for example by 1) applying the Spearman Rank Correlation coefficient (Jian et al., 2017; Seibert and Vis, 2016b), or by converting modelled discharge to stream levels using 2) rating curves whose parameters are free calibration parameters in the modelling process (Sikorska and Renard, 2017; Sun et al., 2012) or 3) the Strickler-Manning equation to directly estimate water levels over the hydraulic properties of the river (Hulsman et al., 2018; Liu et al., 2015).

In the Zambezi river basin, altimetry data has been used in previous studies for hydrological modelling (Michailovsky and Bauer-Gottwein, 2014; Michailovsky et al., 2012). These studies used the altimetry data from the Envisat satellite in an assimilation procedure to update states in a Muskingum routing scheme. Including the altimetry data improved the model performance, especially when the model initially performed poorly due to high model complexity or input data uncertainties.

Despite these recent advances in using river altimetry in hydrological studies, exploitation of its potential is still limited. Various previous studies have argued and provided evidence based on observed discharge data that, in a special case of multi-criteria calibration, the simultaneous model calibration to flow in multiple sub-basins of a river basin, can be beneficial for a more robust selection of parameter sets and thus for a more reliable representation of hydrological processes and their spatial patterns (e.g. Ajami et al., 2004; Clark et al., 2016; Hasan and Pradhanang, 2017; Hrachowitz and Clark, 2017; Santhi et al., 2008). Hence, there may be considerable value in simultaneously using altimetry data not only from one single satellite mission but in combining data from multiple missions, which has not yet been systematically explored. While promising calibration results using data from Envisat were found by Getirana (2010) in tropical and Liu et al. (2015) in snow-dominated regions, altimetry data from multiple sources has not yet been used to calibrate hydrological models in semi-arid regions.

As altimetry observations only describe water level dynamics, it does not provide direct information on the discharge amount. In an attempt to reduce the uncertainty in modelled discharge arising from the missing information on flow amounts, data from the Gravity Recovery and Climate Experiment (GRACE), which provides estimates of the total monthly water storage anomalies, were used to support model calibration. With GRACE, discharge can be constrained through improved simulation of the rainfall partitioning into runoff and evaporation as illustrated in previous studies (Bai et al., 2018; Rakovec et al., 2016).

Therefore, the overarching objective of this study is to explore the combined information content (cf. Beven, 2008) of river altimetry data from multiple satellite missions and GRACE observations to identify feasible parameter sets for the cal-

ibration of hydrological models of large river systems in a semi-arid, data scarce region. More specifically, in a step-wise approach we use GRACE observations together with altimetry data from multiple virtual stations to identify model parameters following three different strategies and we compare model performances to a traditional calibration approach based on in-situ observed river discharge. These three strategies compare altimetry observations to 1) modelled discharge by applying the Spearman Rank Correlation coefficient, and to modelled stream levels by converting modelled discharge using 2) rating curves whose parameters were treated as free model calibration parameters and 3) the Strickler-Manning equation to infer water levels directly from hydraulic properties of the river. These three strategies are tested on a distributed process-based rainfall-runoff model with sub-grid process heterogeneity for the Luangwa Basin. More specifically, we test the following research hypotheses: 1) the use of altimetry data combined with GRACE observations allows a meaningful selection of feasible model parameter sets to reproduce river discharge depending on the applied parameter identification strategy, and 2) the combined application of multiple virtual stations from multiple satellite missions improves the model's ability to reproduce observed hydrological dynamics.

3.2. Site description

The study area is the Luangwa River in Zambia, a tributary of the Zambezi River (Figure 3.1). Its 159,000 km² large basin area is poorly gauged, mostly unregulated and sparsely populated with about 1.8 million inhabitants in 2005 (The World Bank, 2010). The mean annual precipitation is around 970 mm yr⁻¹, whereas the potential evaporation is around 1555 mm yr⁻¹. The landscape varies between low lying flat areas along the river to large escarpments mostly in the North West of the basin and highlands with an elevation difference up to 1850 m (see Figure 3.1B and Section 3.4.2 for more information on the landscape classification). During the dry season, the river meanders between sandy banks while during the wet season from November to May it can cover flood plains several kilometres wide. See Chapter 2 for more detailed information on the Luangwa basin.

3.3. Data availability

In-situ discharge and water level observations

In the Luangwa basin, historical in-situ daily discharge and water level observations were available from the Zambian Water Resources Management Authority at the Great East Road Bridge gauging station, located at 30° 13' E and 14° 58' S (Figure 3.1) about 75 km upstream of the confluence with the Zambezi. In this study, all complete hydrological years of discharge data within the time period 2002 to 2016 were used; these are the years 2004, 2006 and 2008.

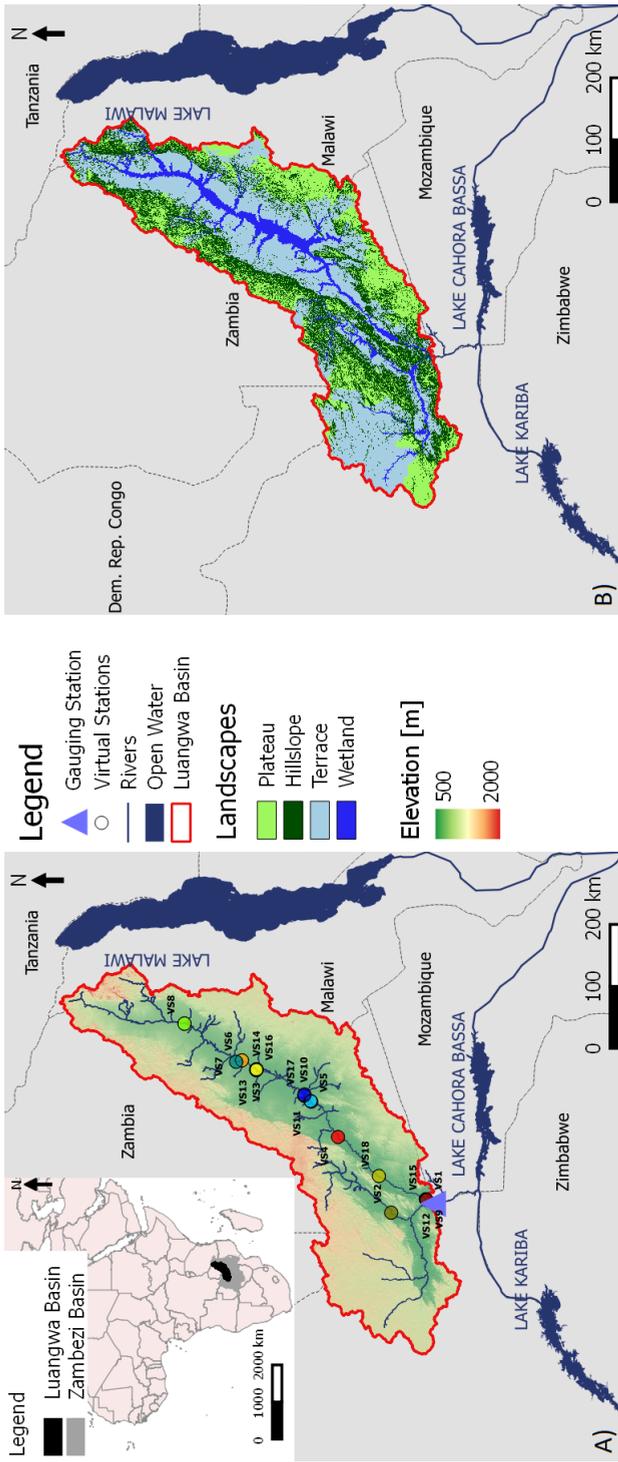


Figure 3.1: A) Elevation map of the Luangwa River Basin in Zambia including the Great East Road Bridges river gauging station and the locations of the 18 virtual stations (VS1 – VS18, the red dot is VS4) with altimetry data used in this study; their colours correspond to those in Figure 3.3. B) Map of the Luangwa River Basin with the main landscape types (see Section 3.4.2).

Gridded data products

Besides the above in-situ observations, gridded data products were used in this study for topographic description, model forcing (precipitation and temperature), and model parameter selection/calibration (total water storage anomalies), as shown in Table 3.1. The temperature data was used to estimate the potential evaporation according to the Hargreaves method (Hargreaves and Allen, 2003; Hargreaves and Samani, 1985).

Gravity Recovery and Climate Experiment (GRACE) observations describe the monthly total water storage anomalies which includes all terrestrial water stores present in the groundwater, soil moisture and surface water. Two identical satellites observe the variations in the Earth's gravity field to detect regional mass changes which are dominated by variations in the terrestrial water storage once atmospheric effects have been accounted for (Landerer and Swenson, 2012; Swenson, 2012). In this study, processed GRACE observations of Release 05 generated by CSR (Centre for Space Research), GFZ (GeoForschungsZentrum Potsdam) and JPL (Jet Propulsion Laboratory) were downloaded from the GRACE Tellus website (<https://grace.jpl.nasa.gov/>; the average of all three sources were used. The raw data were previously processed by CSR, GFZ and JPL to remove atmospheric mass changes using ECMWF (European Centre for Medium-Range Weather Forecasts) atmospheric pressure fields, systematic errors causing north-south-oriented stripes and high frequency noise using a 300 km wide Gaussian filter via spatial smoothing (Landerer and Swenson, 2012; Swenson and Wahr, 2006; Wahr et al., 1998). Processed GRACE observations describe terrestrial water storage anomalies in "equivalent water thickness" in [cm] relative to the 2004 – 2009 time-mean baseline. In other words, the water storage anomaly is the water storage minus the long-term mean (Landerer and Swenson, 2012).

All gridded information was rescaled to the model resolution of 0.1° . The temperature and GRACE data were rescaled by dividing each cell of the satellite product into multiple cells such that the model resolution is obtained, retaining the original value. The precipitation was rescaled by taking the average of all cells located within each model cell.

Table 3.1: Gridded data products used in this study

	Time period	Time resolution	Spatial resolution	Product name	Source
Digital elevation map	n/a	n/a	0.02°	GMTED	(Danielson and Gesch, 2011)
Precipitation	2002 – 2016	Daily	0.05°	CHIRPS	(Funk et al., 2014)
Temperature	2002 – 2016	Monthly	0.5°	CRU	(University of East Anglia Climatic Research Unit et al., 2017)
Total water storage	2002 – 2016	Monthly	1°	GRACE	(Landerer and Swenson, 2012; Swenson, 2012; Swenson and Wahr, 2006)

Altimetry data

The altimetry data used in this study was obtained from the following sources: the Database for Hydrological Time Series of Inland Waters (DAHITI; <https://dahiti.dgfi.tum.de/en/>) (Schwatke et al., 2015), HydroSat (<http://hydrosat.gis.uni-stuttgart.de/php/index.php>) (Tourian et al., 2013), Laboratoire d'Etudes en Géophysique et Océanographie Spatiales (LEGOS; <http://www.legos.obs-mip.fr/soa/hydrologie/hydroweb/>; see Appendix A), and the Earth and Planetary Remote Sensing Lab (EAPRS; <http://www.cse.dmu.ac.uk/EAPRS/>). In total, altimetry data was obtained for 18 virtual stations in the Luangwa basin (Figure 3.1A) for the time period 2002 – 2016 from the satellite missions Jason 1 – 3, Envisat and Saral (Table 3.2, Figure A.2 in Appendix A).

River geometry information

In the Luangwa Basin, very limited detailed in-situ information was available on the river geometry such as cross-section and slope. For that reason, this information was extracted from global high-resolution terrain data available on Google Earth as done successfully in previous studies for other purposes (Pandya et al., 2017; Zhou and Wang, 2015). This was done for each virtual station and the basin outlet. Google Earth only provides river geometry information above the river water level. As the Luangwa is a perennial river, parts of the cross-section remain submerged throughout the year and are thus unknown. To limit uncertainties arising from this issue, the cross-section geometry for each virtual station was extracted from the Google Earth image with the lowest water levels at each individual virtual station. The dates of these images in general fall in the dry season, with flows at the Great East Road Bridges gauging station on the respective days ranging from 1% to 4% relative to the maximum discharge (see Table A.3 in Appendix A for the dates of the satellite images and the associated flows at the Great East Road Bridges gauging station). The database underlying the global terrain images in Google Earth originate from multiple, merged data sources with varying spatial resolutions. For the Luangwa Basin these include the Shuttle Radar Topography Mission (SRTM) with a spatial resolution of 30 m, Landsat 8 with a spatial resolution of 15 m and the Satellite Pour l'Observation de la Terre 4/5 (SPOT) with a spatial resolution of 2.5 – 20 m (Drusch et al., 2012; Irons et al., 2012; Smith and Sandwell, 2003). In addition to Google Earth data, the submerged part of the channel cross-section was surveyed in the field on April 27th 2018 near the Great East Road Bridges river gauging station at the coordinates 30° 13' E and 15° 00' S (Abas, 2018) with an Acoustic Doppler Current Profiler (ADCP).

Table 3.2: Overview of the altimetry data in the Luangwa River Basin used in this study

Nr.	Longitude	Latitude	Time period	Nr. of data points	Source	Mission	Space Agency	Temporal resolution	Equatorial inter-track distance	Along-track distance	Literature
1	30.2823°	-14.8664°	2008 – 2016	246	DAHITI	Jason 2,3	NASA/CNES	10 days	315 km	294 m	(CNES, Accessed 2018; Schwatke et al., 2015)
2	30.0864°	-14.366°	2008 – 2015	92	DAHITI	Jason 2,3	NASA/CNES	10 days	315 km	294 m	(CNES, Accessed 2018; Schwatke et al., 2015)
3	32.1715°	-12.4123°	2008 – 2016	248	DAHITI	Jason 2,3	NASA/CNES	10 days	315 km	294 m	(CNES, Accessed 2018; Schwatke et al., 2015)
4	31.1868°	-13.5927°	2002 – 2016	104	DAHITI	Envisat, Saral	ESA (Envisat), ISRO/CNES (Saral)	35 days	80 km (Envisat), 75 km (Saral)	374 m (Envisat), 173 m (Saral)	(CNES, Accessed 2018; ESA, Accessed 2018; Schwatke et al., 2015)
5	31.6984°	-13.2039°	2002 – 2016	82	DAHITI	Envisat, Saral	ESA (Envisat), ISRO/CNES (Saral)	35 days	80 km (Envisat), 75 km (Saral)	374 m (Envisat), 173 m (Saral)	(CNES, Accessed 2018; ESA, Accessed 2018; Schwatke et al., 2015)
6	32.2998°	-12.2007°	2002 – 2016	100	DAHITI	Envisat, Saral	ESA (Envisat), ISRO/CNES (Saral)	35 days	80 km (Envisat), 75 km (Saral)	374 m (Envisat), 173 m (Saral)	(CNES, Accessed 2018; ESA, Accessed 2018; Schwatke et al., 2015)
7	32.2805°	-12.1157°	2002 – 2016	103	DAHITI	Envisat, Saral	ESA (Envisat), ISRO/CNES (Saral)	35 days	80 km (Envisat), 75 km (Saral)	374 m (Envisat), 173 m (Saral)	(CNES, Accessed 2018; ESA, Accessed 2018; Schwatke et al., 2015)
8	32.831°	-11.3674°	2002 – 2016	105	DAHITI	Envisat, Saral	ESA (Envisat), ISRO/CNES (Saral)	35 days	80 km (Envisat), 75 km (Saral)	374 m (Envisat), 173 m (Saral)	(CNES, Accessed 2018; ESA, Accessed 2018; Schwatke et al., 2015)
9	30.2704°	-14.8809°	2008 – 2015	247	Hydrosat	Jason 2	NASA/CNES	10 days	315 km	294 m	(Tourian et al., 2013, 2016)
10	31.78405°	-13.0995°	2002 – 2010	65	EARRS	Envisat	ESA	35 days	80 km	374 m	(ESA, Accessed 2018; Michailovsky et al., 2012)
11	31.71099°	-13.1943°	2002 – 2010	93	EARRS	Envisat	ESA	35 days	80 km	374 m	(ESA, Accessed 2018; Michailovsky et al., 2012)
12	30.2740°	-14.8763°	2008 – 2015	231	LEGOS	Jason 3	NASA/CNES	10 days	315 km	294 m	(CNES, Accessed 2018; Frappart et al., 2015)
13	32.15843°	-12.412°	2016 – 2016	28	LEGOS	Jason 3	NASA/CNES	10 days	315 km	294 m	(CNES, Accessed 2018; Frappart et al., 2015)
14	32.15989°	-12.4127°	2002 – 2009	137	LEGOS	Jason 1	NASA/CNES	10 days	315 km	294 m	(CNES, Accessed 2018; Frappart et al., 2015)
15	30.2740°	-14.8763°	2008 – 2016	271	LEGOS	Jason 2	NASA/CNES	10 days	315 km	294 m	(CNES, Accessed 2018; Frappart et al., 2015)
16	32.16056°	-12.4125°	2008 – 2016	283	LEGOS	Jason 2	NASA/CNES	10 days	315 km	294 m	(CNES, Accessed 2018; Frappart et al., 2015)
17	31.80001°	-13.0909°	2013 – 2016	35	LEGOS	Saral	ISRO/CNES	35 days	75 km	173 m	(CNES, Accessed 2018; Frappart et al., 2015)
18	30.61577°	-14.1852°	2013 – 2016	24	LEGOS	Saral	ISRO/CNES	35 days	75 km	173 m	(CNES, Accessed 2018; Frappart et al., 2015)

3.4. Hydrological model development

3.4.1. General approach

The potential of river altimetry for model calibration was tested with a process-based hydrological model for the Luangwa river basin. This model relied on distributed forcing allowing for spatially explicit distributed water storage calculations. The model was run on a daily time scale for the time period 2002 to 2016. To reach the objective of this study, the following distinct parameter identification strategies were compared in a stepwise approach: (1) traditional model calibration to observed river flow as benchmark; (2) identification of parameter sets reproducing the seasonal water storage anomalies based on GRACE data only; (3a) Altimetry Strategy 1: identification of parameter sets directly based on remotely sensed water levels combined with GRACE data; (3b) Altimetry Strategy 2: identification of parameter sets based on remotely sensed water levels by converting modelled discharges into water levels using calibrated rating curves combined with GRACE data; (3c) Altimetry Strategy 3: identification of parameter sets based on remotely sensed water levels by converting modelled discharges into water levels using the Strickler-Manning equation and including river geometry information (cross-section and gradient) extracted from Google Earth combined with GRACE data; (4a) Water level Strategy 1: identification of parameter sets based on daily river water level at the catchment outlet only using the Strickler-Manning equation and including river geometry information extracted from Google Earth combined with GRACE data; and (4b) Water level Strategy 2: identification of parameter sets based on daily river water level at the catchment outlet only using the Strickler-Manning equation and including river geometry information obtained from a detailed field survey with an Acoustic Doppler Current Profiler (ADCP) combined with GRACE data. Note that (1) is completely independent of (2) to (4) where no discharge data was used for the identification of parameter sets.

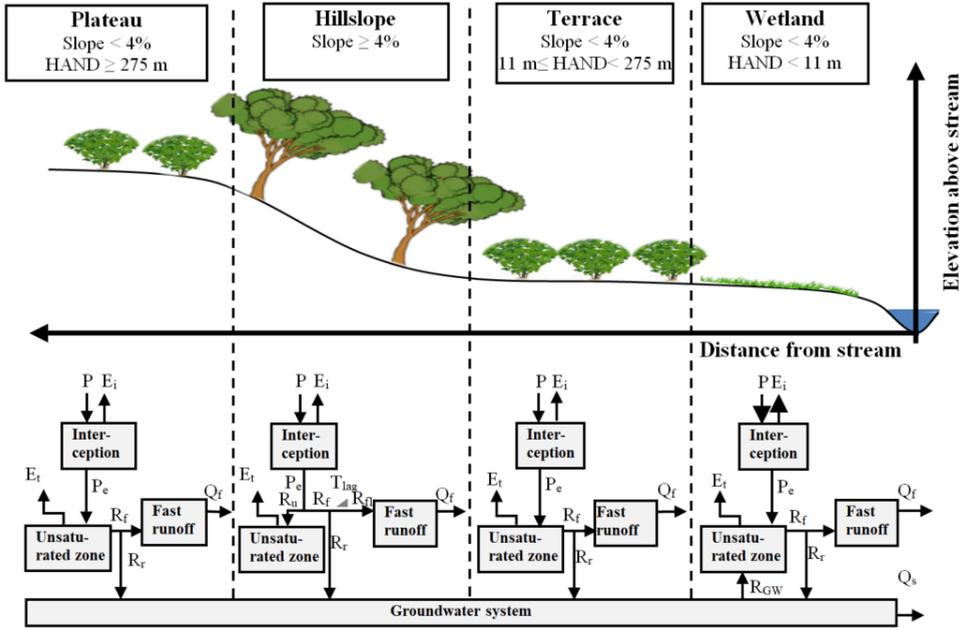


Figure 3.2: Sketch of the hydrological response units including the thresholds used in this analysis for the slope and HAND (Height Above Nearest Drainage) and including their corresponding model structures. This spatial sub-grid discretization was applied to each grid cell. Symbol explanation: precipitation (P), effective precipitation (P_e), interception evaporation (E_i), plant transpiration (E_t), infiltration into the unsaturated root zone (R_f), drainage to fast runoff component (R_{fi}), lag time (T_{lag}), groundwater recharge (R_r), upwelling groundwater flux (R_{GW}), fast runoff (Q_f), groundwater/slow runoff (Q_s).

3.4.2. Hydrological model structure

In this study, a process-based rainfall-runoff with distributed water accounting and sub-grid process heterogeneity was developed on daily time-scale (Ajami et al., 2004; Euser et al., 2015). The river basin was discretized into a grid with a spatial resolution of $10 \times 10 \text{ km}^2$. Each model grid cell was characterized by the same model structure and parameter sets but forced by spatially distributed, gridded input data (Table 3.1). Runoff was then calculated in parallel for each cell separately. Subsequently, a routing scheme was applied to estimate the aggregated flow in each grid cell at each time step.

Adopting the FLEX-Topo modelling concept (Savenije, 2010) and extending it to a gridded implementation, each grid cell was further discretised into functionally distinct hydrological response units (HRU) as demonstrated by Nijzink et al. (2016b). Each point within a grid cell was assigned to a response class based on its position in the landscape as defined by its local slope and "Height-above-the-nearest-drainage" (HAND; Gharari et al., 2011; Rennó et al., 2008). Similar to previous studies (e.g. Gao et al., 2016; Nijzink et al., 2016b), the response units plateau, hillslope, terrace and wetland were distinguished. Reflecting earlier work (e.g. Gharari et al., 2011), all locations with slope of $> 4\%$ were assumed to be hillslope. Locations with lower

slopes were then either defined as wetland ($HAND < 11$ m), terrace ($11 \text{ m} \leq HAND < 275$ m) or plateau ($HAND \geq 275$ m), see Figure 3.2. Following this classification wetlands make up 8%, terraces 41%, hillslopes 28% and plateaus 23% of the total Luangwa River Basin area as mapped in Figure 3.1B.

Each response class consisted of a series of storage components that are linked by fluxes. The flow generated from each grid cell at any given time step is then computed as the area-weighted flow from the individual response units plus a contribution from the common groundwater component which connects the response units (Figure 3.2). Finally, the outflow from each modelling cell was routed to downstream cells to obtain the accumulated flow in each grid cell at any given time step. For this purpose, the mean flow length of each model grid cell to the outlet was derived based on the flow direction extracted from the digital elevation model. The flow velocity, which was assumed to be constant in space and time, was calibrated. With this information on the flow path length and velocity, the accumulated flow in each grid cell was calculated at the end of each time step. The relevant model equations are given in Table 3.3. This concept was previously successfully applied in a wide range of environments (Fovet et al., 2015; Gao et al., 2014a; Gharari et al., 2014; Nijzink et al., 2016b; Prenner et al., 2018).

3.4.3. Parameter selection procedures

To evaluate the information content and thus the utility of altimetry data for the selection of feasible model parameter sets, a step-wise procedure as specified in detail below was applied (Table 3.4). Note that given data scarcity and the related issues of epistemic uncertainties (Beven and Westerberg, 2011; McMillan and Westerberg, 2015) and equifinality (Beven, 2006; Savenije, 2001) we did not aim to identify the “optimal” parameter set in what is frequently considered a traditional calibration approach. In most hydrological applications the available data have limited strength for rigorous model tests (Clark et al., 2015; Gupta et al., 2008; Jakeman and Hornberger, 1993). Thus, to reduce the risk of rejecting good parameters when they should have been accepted (Beven, 2010; Hrachowitz and Clark, 2017), we rather attempted to identify and discard the most implausible parameter sets (Freer et al., 1996) that violate our theoretical understanding of the system or that are inconsistent with the available data (Knutti, 2008). This allowed us to iteratively constrain the feasible parameter space and thus the uncertainty around the modelled hydrograph (Hrachowitz et al., 2014). To do so, a Monte-Carlo sampling strategy with uniform prior parameter distributions was applied to generate $5 \cdot 10^4$ model realizations. This random set of solutions was in the following steps used as baseline and iteratively constrained by identifying parameter sets that do not satisfy pre-specified criteria (see below), depending on the data type and source used.

Benchmark: Parameter selection based on observed discharge data

As benchmark, and following a traditional calibration procedure, the model was calibrated with observed daily discharge based on the Nash-Sutcliffe efficiency ($E_{NS,Q}$, Eq.3.1 in Table 3.5) using all complete hydrological years within the time period

Table 3.3: Equations applied in the hydrological model. **Fluxes** [mm d^{-1}]: precipitation (P), effective precipitation (P_e), potential evaporation (E_p), interception evaporation (E_i), plant transpiration (E_t), infiltration into the unsaturated zone (R_u), drainage to fast runoff component (R_f), delayed fast runoff (R_{fl}), groundwater recharge (R_r for each relevant HRU and $R_{r,tot}$ combining all relevant HRUs), upwelling groundwater (R_{GW} for each relevant HRU and $R_{GW,tot}$ combining all relevant HRUs), fast runoff (Q_f for each relevant HRU and $Q_{f,tot}$ combining all relevant HRUs), groundwater/slow runoff (Q_s), total runoff (Q_m). **Storages** [mm]: storage in interception reservoir (S_i), storage in unsaturated root zone (S_u), storage in groundwater/slow reservoir (S_s), storage in fast reservoir (S_f). **Parameters**: interception capacity (I_{max}) [mm], maximum upwelling groundwater (C_{max}) [mm d^{-1}], maximum root zone storage capacity ($S_{u,max}$) [mm], splitter (W) [-], shape parameter (β) [-], transpiration coefficient (C_e) [-], time lag (T_{lag}) [d], reservoir time scales [d] of fast (K_f) and slow (K_s) reservoirs, areal weights (p_{HRU}) [-], time step (Δt) [d]. Calibration parameters are shown in bold letters in the table below. The equations were applied to each hydrological response unit (HRU) unless indicated differently.

Reservoir system	Water balance equation	Process functions
Interception	$\frac{\Delta S_i}{\Delta t} = P - P_e - E_i$	$E_i = \min(E_p, \min(P, \frac{I_{max}}{\Delta t}))$ $P_e = P - E_i$
Unsaturated zone	Plateau/Terrace: $\frac{\Delta S_u}{\Delta t} = P_e - E_t - R_f$ Hillslope: $\frac{\Delta S_u}{\Delta t} = R_u - E_t$ Wetland: $\frac{\Delta S_u}{\Delta t} = P_e - E_t - R_f + R_{GW}$	$E_t = \min((E_p - E_i), \min(\frac{S_u}{\Delta t}, (E_p - E_i) \cdot \frac{S_u}{S_{u,max}} \cdot \frac{1}{C_e}))$ $R_{GW} = \min((1 - \frac{S_u}{S_{u,max}}) \cdot C_{max} \cdot \frac{S_s}{p_{HRU}})$ if $S_u + R_{GW} \cdot \Delta t > S_{u,max}$: $R_{GW} = \frac{S_{u,max} - S_u}{\Delta t}$ Hillslope: $R_u = (1 - C) \cdot P_e$ $C = 1 - (1 - \frac{S_u}{S_{u,max}})^\beta$
Fast runoff	$\frac{\Delta S_f}{\Delta t} = R_{fl} - Q_f$	$Q_f = \frac{S_f}{K_f}$ Plateau/Terrace/Wetland: $R_f = \frac{\max(0, S_u - S_{u,max})}{\Delta t}$ $R_{fl} = R_f$ Hillslope: $R_f = (1 - W) \cdot C \cdot P_e$ $R_{fl} = R_f * f(T_{lag})$
Groundwater	$\frac{\Delta S_s}{\Delta t} = R_{r,tot} - R_{GW,tot} - Q_s$	$R_r = W \cdot C \cdot P_e$ $R_{r,tot} = \sum_{HRU} (p_{HRU} \cdot R_r)$ $R_{GW,tot} = \sum_{HRU} (p_{HRU} \cdot R_{GW})$ $Q_s = \frac{S_s}{K_s}$
Total runoff	$Q_m = Q_s + Q_{f,tot}$	$Q_{f,tot} = \sum_{HRU} (p_{HRU} \cdot Q_f)$
Supporting literature	(Euser et al., 2015; Gao et al., 2014a; Gharari et al., 2014)	

2002 to 2016 (Nash and Sutcliffe, 1970); these are the years starting in the fall of 2004, 2006 and 2008.

To limit the solutions to relatively robust representations of the system while allowing for data and model uncertainty (e.g. Beven, 2006; Beven and Westerberg,

2011) only parameter sets that resulted in $E_{NS,Q} \geq 0.6$ were retained as feasible. The hydrological model consisted of 18 free calibration parameters (Table 3.4, Figure A.1 in Appendix A) whose uniform prior distributions are given in Table A.1 in Appendix A with associated parameter constraints as summarised in Table A.2.

Parameter selection based on the seasonal water storage (GRACE)

In a next step we assumed that discharge records in the Luangwa Basin were absent. The starting assumption thus had to be that all model realizations, i.e. all sampled parameter sets, were equally likely to allow feasible representations of the hydrological system. In a step-wise approach, confronting these realizations with different types of data, we sequentially identified and discarded solutions that were least likely to provide meaningful system representations, thereby gradually narrowing down the feasible parameter space.

We first identified and discarded solutions that were least likely to preserve the observed seasonal water storage (S_{tot}) fluctuations. To do so, the monthly modelled total water storage ($S_{tot,mod} = S_i + S_u + S_f + S_s$) relative to the 2004-2009 time-mean baseline in each grid cell was compared to water storage anomalies observed with GRACE where this same time-mean baseline was used (Fang et al., 2016; Footan et al., 2019; Khaki and Awange, 2019; Tang et al., 2017).

The model's skill to reproduce the seasonal water storage, i.e. S_{tot} , was assessed using the Nash-Sutcliffe efficiency $E_{NS,S_{tot}}$ (Eq.3.1). Note that $E_{NS,S_{tot,j}}$ was computed at first from the time series of S_{tot} in each grid cell j which were then averaged to obtain $E_{NS,S_{tot}}$. If no additional data were available, a hypothetical modeller relying on $E_{NS,S_{tot}}$ to calibrate a model, may choose only the solution with the highest $E_{NS,S_{tot}}$ or allow for some uncertainty. To mimic this traditional approach but to balance it with a sufficient number of feasible solutions to be kept for the subsequent steps we here identified and discarded the poorest performing 75% of all solutions in terms of $E_{NS,S_{tot}}$ as unfeasible for the subsequent modelling steps.

Parameter selection based on satellite altimetry data

Next, the remaining feasible parameter sets were used to evaluate their potential to reproduce time series of observed altimetry applying three distinct parameter selection and model evaluation strategies. Assuming again the situation of an ungauged basin (i.e. no time-series of river flow available), we kept for each strategy as feasible the respective 1% best performing parameter sets according to the specific performance metric associated to that strategy.

Altimetry Strategy 1: Direct comparison of altimetry data to modelled discharge

In the simplest approach, we directly used altimetry data to correlate observed water levels with modelled discharge based on the Spearman rank correlation coefficient $E_{R,WL}$ (Spearman, 1904) using Eq.3.2 (Table 3.5). This strategy, hereafter referred to with subscript WL, i.e. water level, requires the assumption that the relationship between water level and discharge is monotonic. The Spearman rank correlation was applied successfully in previous studies to calibrate a rainfall-runoff model to water level time series (Seibert and Vis, 2016b). As there

Table 3.4: Overview of the parameter identification strategies applied in this study

Strategy	Calibration data	Objective function	Parameter group	Calibration parameters	Comments	Q – h conversion	Benefits (+) & limitations (-)
Discharge (reference)	Discharge (at basin outlet)	$E_{NS,Q}$ (Eq.3.1)	Entire basin Plateau/Terrace Hillslope Wetland River profile	$K_{s,T}, C_e$ $I_{max}, S_{urmax}, K_f, W$ $I_{max}, S_{urmax}, K_f, W, \beta, T_{lag}$ $I_{max}, S_{urmax}, K_f, W, C_{max}$ Total: 18	<ul style="list-style-type: none"> Traditional model calibration 8 flow signatures combined 	-	-
Seasonal water storage	GRACE	$E_{NS,Stor}$ (Eq.3.1)	Entire basin Plateau/Terrace Hillslope Wetland River profile	$K_{s,T}, C_e$ $I_{max}, S_{urmax}, K_f, W$ $I_{max}, S_{urmax}, K_f, W, \beta, T_{lag}$ $I_{max}, S_{urmax}, K_f, W, C_{max}$ Total: 18	<ul style="list-style-type: none"> No discharge data used 	-	-
Altimetry Strategy 1	Altimetry (at 18 virtual stations) & GRACE	Altimetry: $D_{GR,VP}$ (Eq.3.2,3,4) GRACE: $E_{NS,Stor}$ (Eq.3.1)	Entire basin Plateau/Terrace Hillslope Wetland River profile	$K_{s,T}, C_e$ $I_{max}, S_{urmax}, K_f, W$ $I_{max}, S_{urmax}, K_f, W, \beta, T_{lag}$ $I_{max}, S_{urmax}, K_f, W, C_{max}$ Total: 18	<ul style="list-style-type: none"> No discharge data used 18 virtual stations combined Combined with GRACE 	-	+ No extra parameters or data needed + Assumption: monotonic relation between discharge and river water level – Focus on dynamics only
Altimetry Strategy 2	Altimetry (at 18 virtual stations) & GRACE	Altimetry: $D_{GR,RC}$ (Eq.3.1,3,4) GRACE: $E_{NS,Stor}$ (Eq.3.1)	Entire basin Plateau/Terrace Hillslope Wetland River profile	$K_{s,T}, C_e$ $I_{max}, S_{urmax}, K_f, W$ $I_{max}, S_{urmax}, K_f, W, \beta, T_{lag}$ $I_{max}, S_{urmax}, K_f, W, C_{max}$ $v, a_1, a_2, a_3, a_4, b_1, b_2, b_3, b_4$ Total: 26	<ul style="list-style-type: none"> No discharge data used 18 virtual stations combined Combined with GRACE 	Calibrated rating curve	+ No extra data needed – Two extra parameters per cross-section
Altimetry Strategy 3	Altimetry (at 18 virtual stations) & GRACE	Altimetry: $D_{GR,SM}$ (Eq.3.1,3,4) GRACE: $E_{NS,Stor}$ (Eq.3.1)	Entire basin Plateau/Terrace Hillslope Wetland River profile	$K_{s,T}, C_e$ $I_{max}, S_{urmax}, K_f, W$ $I_{max}, S_{urmax}, K_f, W, \beta, T_{lag}$ $I_{max}, S_{urmax}, K_f, W, C_{max}$ Total: 18	<ul style="list-style-type: none"> No discharge data used 18 virtual stations combined Combined with GRACE 	Strickler-Manning	+ Only 1 extra parameter needed – Cross-section data – Assumption: constant roughness in space & time
Water level Strategy 1	Water level (at basin outlet) & GRACE	Water level: $E_{NS,SM,GR}$ (Eq.3.1) GRACE: $E_{NS,Stor}$ (Eq.3.1)	Entire basin Plateau/Terrace Hillslope Wetland River profile	$K_{s,T}, C_e$ $I_{max}, S_{urmax}, K_f, W$ $I_{max}, S_{urmax}, K_f, W, \beta, T_{lag}$ $I_{max}, S_{urmax}, K_f, W, C_{max}$ v, R Total: 19	<ul style="list-style-type: none"> No discharge data used Combined with GRACE 	Strickler-Manning	+ Only 1 extra parameter needed – Cross-section data – Assumption: constant roughness in space & time
Water level Strategy 2	Water level (at basin outlet) & GRACE	Water level: $E_{NS,SM,ADCP}$ (Eq.3.1) GRACE: $E_{NS,Stor}$ (Eq.3.1)	Entire basin Plateau/Terrace Hillslope Wetland River profile	$K_{s,T}, C_e$ $I_{max}, S_{urmax}, K_f, W$ $I_{max}, S_{urmax}, K_f, W, \beta, T_{lag}$ $I_{max}, S_{urmax}, K_f, W, C_{max}$ v, R Total: 19	<ul style="list-style-type: none"> No discharge data used Combined with GRACE 	Strickler-Manning	+ Only 1 extra parameter needed – Cross-section data – Assumption: constant roughness in space & time

Table 3.5: Equations used to calculate the model performance

Name	Objective function	Symbol explanation
Nash-Sutcliffe efficiency	$E_{NS,\theta} = 1 - \frac{\sum_t (\theta_{\text{mod}}(t) - \theta_{\text{obs}}(t))^2}{\sum_t (\theta_{\text{obs}}(t) - \theta_{\text{obs}})^2} \quad (3.1)$	θ : variable
Spearman-Rank correlation coefficient	$E_{R,WL} = \frac{\text{COV}(r_{Q,\text{mod}}, r_{WL,\text{obs}})}{\sigma(r_{Q,\text{mod}}) \cdot \sigma(r_{WL,\text{obs}})} \quad (3.2)$	$r_{Q,\text{mod}}$: ranks of the modelled discharge $r_{WL,\text{obs}}$: ranks of the observed water levels
Relative error	$E_{R,\theta} = 1 - \frac{ \theta_{\text{mod}} - \theta_{\text{obs}} }{\theta_{\text{obs}}} \quad (3.3)$	θ : variable
Euclidian distance over multiple virtual stations	$D_{E,\beta,\gamma} = 1 - \sqrt{\sum_t w_i \cdot (1 - E_{\beta,\gamma})^2} \quad (3.4)$	w_i : relative weight of virtual station i β : model performance metric γ : parameter selection method
Euclidian distance over multiple signatures	$D_E = 1 - \sqrt{\frac{1}{N+M} \left(\sum_n (1 - E_{NS,\theta_n})^2 + \sum_m (1 - E_{R,\theta_m})^2 \right)} \quad (3.5)$	θ : signature n : signatures evaluated with Eq.3.1 with maximum N m : signatures evaluated with Eq.3.3 with maximum M

were multiple virtual stations with water level data available in this study, the $E_{R,WL}$ was computed at each location simultaneously. The individual values $E_{R,WL}$ were weighted based on the record length of the corresponding virtual stations and then combined into the Euclidean distance as aggregate metric $D_{E,R,WL}$ with Eq.3.4.

Altimetry Strategy 2: Rating curves In the second strategy, as successfully applied in previous studies (Getirana and Peters-Lidard, 2013; Jian et al., 2017), model parameters were selected based on the models' ability to reproduce water levels by converting the modelled discharge to water levels, assuming these two are related through a rating curve in the form of a power function (Rantz, 1982):

$$Q = a \cdot (h - h_0)^b \quad (3.6)$$

Where h is the water level, h_0 a reference water level, and a and b are two additional free calibration parameters, determining the shape of the function and lumping the combined influences of different river cross-section characteristics, such as geometry or roughness. Note, that here for each virtual station h_0 is the elevation that corresponds to the water level of the Google Earth image with the lowest flow available, corresponding to the assumption of no-flow at that time. This strategy is hereafter referred to as with subscript RC, i.e. rating curve. As river-cross sections vary in space, each of the 18 virtual stations would require an individual set of these parameters a and b . To limit the number of additional calibration parameters, we here classified the river-cross sections of the 18 virtual stations into 4 groups

(Figure 3.1A and Figure 3.3). For cross-sections within each class, i.e. geometrically similar, the same values for a and b were used, resulting in 4 sets of a and b and thus a total of 8 additional calibration parameters. The river cross-sections were extracted from global high-resolution terrain data available on Google Earth (see Section 3.3). The modelled river water levels were evaluated against the observed water levels at each virtual station using the Nash-Sutcliffe efficiency $E_{NS,RC}$ (equivalent to Eq.3.1 in Table 3.5), weighted based on the record length of the corresponding virtual stations and then combined into the Euclidean distance $D_{E,NS,RC}$ as an aggregated performance metric (Eq.3.4).

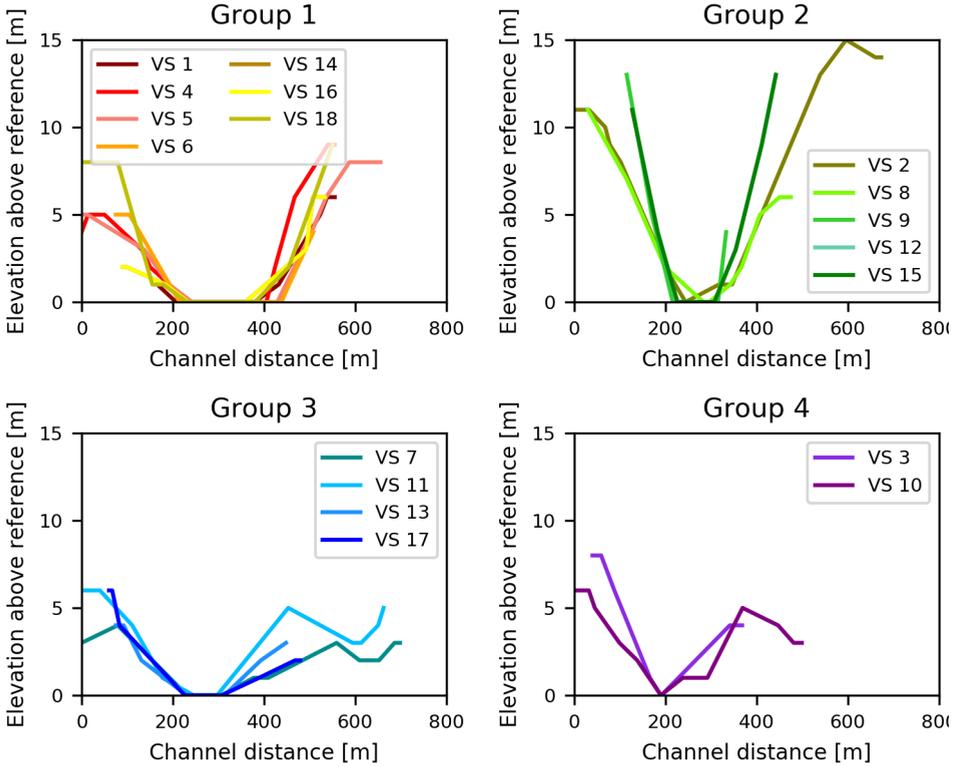


Figure 3.3: River profiles at 18 virtual stations (VS) divided into four groups. The reference level is equal to the lowest water level in the river profile for each location separately.

Altimetry Strategy 3: Strickler-Manning equation As a third strategy, we converted the modelled discharge to river water levels using the Strickler-Manning equation (Manning, 1891):

$$Q = k \cdot i^{1/2} \cdot A \cdot R^{2/3} \quad (3.7)$$

Where k is a roughness parameter, here treated as free calibration parameter and assumed constant for all virtual stations, i is the mean channel slope, extracted

here over a distance of 10 km, while A and R are the river cross-section area and hydraulic radius. Assuming trapezoidal cross-sections (see Figure 3.4 as illustrative example), A and R were calculated for each cross section according to:

$$A = B \cdot d + \frac{1}{2} \cdot d^2 \cdot (i_1 + i_2) \quad (3.8)$$

$$R = \frac{A}{B + d \cdot ((1 + i_1^2)^{1/2} + (1 + i_2^2)^{1/2})} \quad (3.9)$$

$$d = h - h_0 \quad (3.10)$$

Where B is the assumed river bed width, i_1 and i_2 are the river bank slopes, d the water depth, h the water level and h_0 the reference water level, here assumed to be the lowest observed river water level to limit the number of calibration parameters. In contrast to previous studies that use a similar approach but relied on locally observed river-cross sections (Hulsman et al., 2018; Liu et al., 2015; Michailovsky et al., 2012), here both, the river bed geometries (Figure 3.3) at and the channel slopes upstream of the 18 virtual stations were computed using high-resolution terrain data retrieved from Google Earth (see Section 3.3). Similar data sources were already used in previous studies to extract the river geometry (e.g. Gichamo et al., 2012; Michailovsky et al., 2012; Pramanik et al., 2010). The reader is referred to Table A.3 in Appendix A for the values of the variables for each virtual station. This strategy is hereafter referred to as with subscript SM, i.e. Strickler-Manning. Equivalent to above, the modelled river water levels were then evaluated against the observed water levels at each virtual station using the Nash-Sutcliffe efficiency $E_{NS,SM}$ (equivalent to Eq.3.1), weighted based on the record length of the corresponding virtual stations and then combined into the Euclidean distance $D_{E,NS,SM}$ as an aggregated performance metric (Eq.3.4).

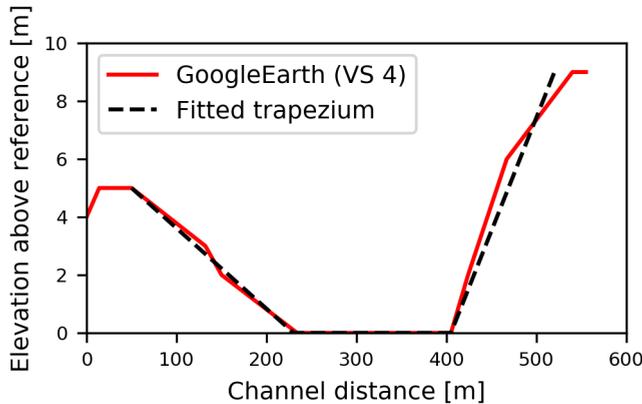


Figure 3.4: Example of approximating a trapezoidal cross-section (black) into the Google Earth based cross-section data (red) for virtual station "VS4" (see also Figure 3.1A and Figure 3.3). The reference level is equal to the lowest water level in the river profile.

Parameter selection based on daily river water level at the basin outlet

For the previous parameter identification strategy (Altimetry Strategy 3), river geometry information was extracted from high-resolution terrain data retrieved from Google Earth which have a low accuracy. Unfortunately, more accurate cross-section information from in-situ surveys was only available at the Great East Road Bridge gauging station, i.e. the basin outlet, where, in turn, no altimetry observations were available. That is why water level time series were used to illustrate the influence of the cross-section accuracy.

As shown in Figure 3.5, the Google Earth based above-water cross-section at the basin outlet corresponded in general well to the field survey considering that satellite images have limited spatial resolution. However, the in-situ measurement also illustrated the relevance of the submerged part of the channel cross-section at that location on the day the image was taken (June 2nd 2008).

Water level Strategy 1: River geometry information extracted from Google Earth First, cross-section information was extracted from global high-resolution terrain data available on Google Earth (subscript GE) and used with the Strickler-Manning equation (Eq.3.7) to convert the modelled discharge to water levels. This was combined with GRACE observations to restrict the parameter space in an equivalent way as in the previous section. The model performance with respect to river water levels was calculated with the Nash-Sutcliffe efficiency $E_{NS,SM,GE}$ (Eq.3.1).

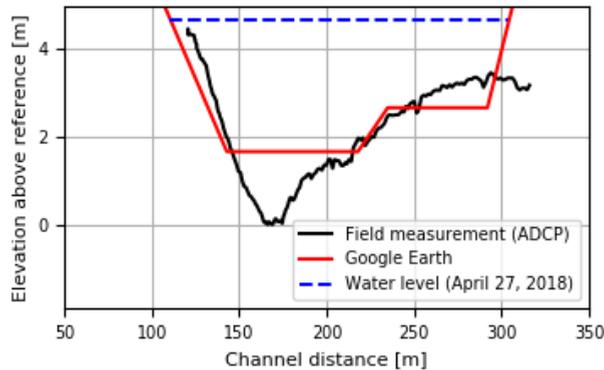


Figure 3.5: River cross-section at Luangwa Bridge obtained from Google Earth and detailed field survey including the river water level on June 2nd 2008. Field measurements were done with an Acoustic Doppler Current Profiler (ADCP) on April 27th 2018 at the coordinates 30° 13' E and 15° 00' S; the satellite image was taken on June 27th 2008. The reference level is equal to the lowest elevation level measured with the ADCP.

Table 3.6: Overview of flow signatures used in this study

Flow signature	Explanation	Function	Model performance equation
Q	Daily flow time series	-	$E_{NS,Q} = 1 - \frac{\sum_t (Q_{mod,t} - Q_{obs,t})^2}{\sum_t (Q_{obs,t} - \bar{Q}_{obs})^2}$
logQ	Logarithm of daily flow time series	-	$E_{NS,\log Q} = 1 - \frac{\sum_t (Q_{log,mod,t} - Q_{log,obs,t})^2}{\sum_t (Q_{log,obs,t} - \bar{Q}_{log,obs})^2}$
FDC	Flow duration curve	-	$E_{NS,FDC} = 1 - \frac{\sum_t (Q_{sort,mod,t} - Q_{sort,obs,t})^2}{\sum_t (Q_{sort,obs,t} - \bar{Q}_{sort,obs})^2}$
logFDC	Logarithm of flow duration curve	-	$E_{NS,\log FDC} = 1 - \frac{\sum_t (Q_{log,sort,mod,t} - Q_{log,sort,obs,t})^2}{\sum_t (Q_{log,sort,obs,t} - \bar{Q}_{log,sort,obs})^2}$
RCdry	Runoff coefficient during dry periods	$RC_{dry} = \frac{Q_{dry}}{P_{dry}}$	$E_{R,RCdry} = 1 - \frac{ RC_{dry,mod} - RC_{dry,obs} }{RC_{dry,obs}}$
RCwet	Runoff coefficient during wet periods	$RC_{wet} = \frac{Q_{wet}}{P_{wet}}$	$E_{R,RCwet} = 1 - \frac{ RC_{wet,mod} - RC_{wet,obs} }{RC_{wet,obs}}$
AC	Autocorrelation function	$AC_t = \frac{\sum_i (Q_i - \bar{Q})(Q_{i+t} - \bar{Q})}{\sum_i (Q_i - \bar{Q})^2}$	$E_{NS,AC} = 1 - \frac{\sum_t (AC_{mod,t} - AC_{obs,t})^2}{\sum_t (AC_{obs,t} - \bar{AC}_{obs})^2}$
RLD	Rising limb density	$RLD = \frac{N_{peak}}{T_r}$	$E_{R,RLD} = 1 - \frac{ RLD_{mod} - RLD_{obs} }{RLD_{obs}}$

Water level Strategy 2: River geometry information obtained from a detailed field survey Second, cross-section information obtained from a detailed field survey with an ADCP (subscript ADCP) was used with the Strickler-Manning equation (Eq.3.7) to convert the modelled discharge to water levels. This was combined with GRACE observations to restrict the parameter space in an equivalent way as in the previous section. The model performance with respect to river water levels was calculated with the Nash-Sutcliffe efficiency $E_{NS,SM,ADCP}$ (Eq.3.1).

3.4.4. Model evaluation

For each calibration strategy, the performance of all model realizations was evaluated post-calibration with respect to discharge using seven additional hydrological signatures (e.g. Euser et al., 2013; Sawicz et al., 2011) to assess the skill of the model to reproduce the overall response of the system and thus the robustness of the selected parameters (Hrachowitz et al., 2014). The signatures included the logarithm of the daily flow time series (hereafter referred to with the subscript logQ), the flow duration curve (FDC), its logarithm (logFDC), the mean seasonal runoff coefficient during dry periods (April – September; RCdry), the mean seasonal runoff coefficient during the wet periods (October – March; RCwet), the autocorrelation function of daily flow (AC) and the rising limb density of the hydrograph (RLD). An overview of these signatures can be found in Table 3.6, and more detailed explanations in Euser et al. (2013) and references therein. As performance measures for the model to reproduce the individual observed signatures the Nash-Sutcliffe efficiency ($E_{NS,\log Q}$, $E_{NS,FDC}$, $E_{NS,\log FDC}$, $E_{NS,AC}$; equivalent to Eq.3.1 in Table 3.5) and a metric based on the relative error ($E_{R,RCdry}$, $E_{R,RCwet}$, $E_{R,RLD}$; equivalent to Eq.3.3) were used (Euser et al., 2013). The signatures were combined, with equal weights, into one objective function, which was formulated based on the Euclidian distance D_E (Eq.3.5) so that a value of 1 indicates a “perfect” model (Schoups et al., 2005).

3.5. Results and discussion

3.5.1. Parameter selection and model performance

The complete set of all model realizations unsurprisingly results in a wide range of model solutions (Figure 3.6A), with $E_{NS,Q}$ ranging from -6.4 to 0.78 and with the combined performance metric of all signatures D_E ranging from -334 to 0.79 (Figure 3.7). With respect to the individual flow signatures, the model performance varied such that the largest range was found in $E_{NS,Q}$ and smallest in $E_{NS,AC}$ as visualised in Figure 3.7 and tabulated in Table A.4 in Appendix A. Although containing relatively good solutions, this full set of all realizations clearly also contained many parameter sets that considerably over- and/or underestimate flows.

Benchmark: Parameter selection based on observed discharge data

For the benchmark case, applying the traditional model calibration approach using discharge data, this parameter selection and calibration strategy results in a reasonable model performance, in which the seasonal but also the daily flow dynamics and magnitudes are in general well captured as shown in Figure 3.6B. For some years, a number of solutions overestimate flows in the wet season and underestimate flows during the dry season, when the river becomes a small meandering stream with almost annual morphological changes which is difficult to meaningfully observe. The best performing solution has a calibration objective function $E_{NS,Q,opt} = 0.78$ (5/95th percentiles of all feasible solutions $E_{NS,Q,5/95} = 0.61 - 0.75$; Figure 3.7 and Table 3.7). For the post-calibration evaluation of all retained solutions, it was observed that most signatures are well reproduced by the majority of solutions, except for the dry season runoff coefficient (RCdry; Figure 3.7 and Table A.4 in Appendix A). This resulted in aggregated model performances, combining all signatures, of $D_{E,5/95} = 0.55 - 0.76$ with the above identified best performing solution (i.e. $E_{NS,Q,opt}$) reaching a value of $D_{E,opt} = 0.60$.

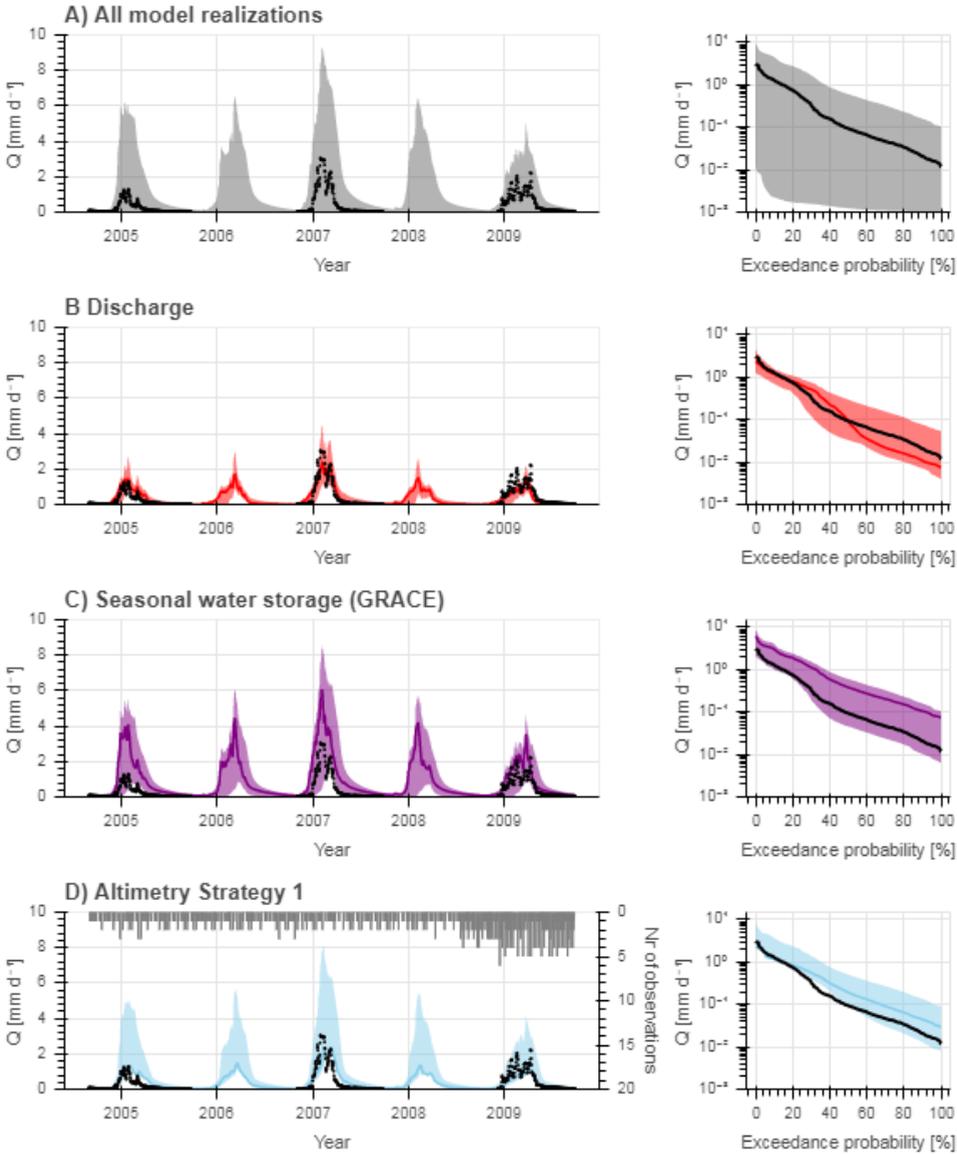


Figure 3.6: Range of model solutions. The left panel shows the hydrograph and the right panel the flow duration curve in logarithmic scale of the recorded (black) and modelled discharge: the line indicates the solution with the highest calibration objective function (E_{NS} or D_E) and the shaded area the envelope of the solutions retained as feasible. A) All model solutions included; solutions retained as feasible based on B) discharge (i.e. “traditional calibration”; $E_{NS,Q}$), C) GRACE ($E_{NS,Stot}$), and D) Altimetry Strategy 1 only ($D_{E,R,WL}$). The grey bars in the left subplot D indicate the number of altimetry observations available for each day.

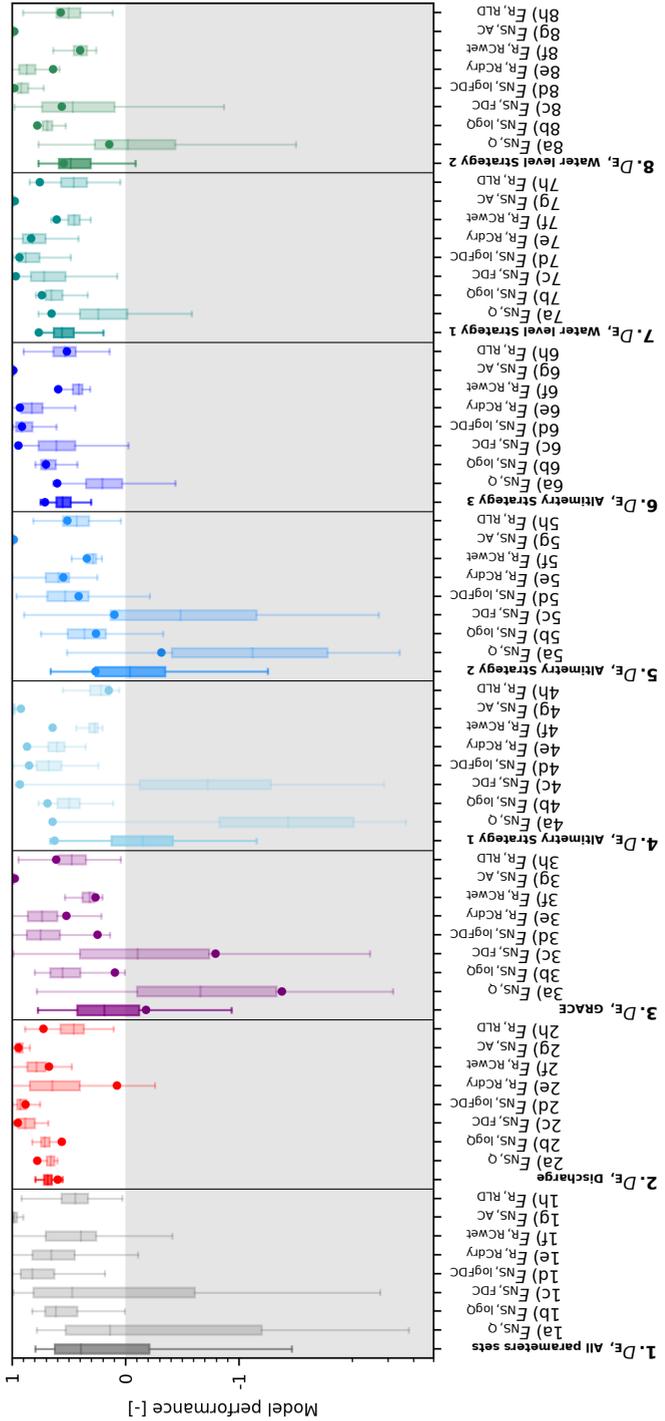


Figure 3.7: Comparison of different data sources to identify feasible parameter sets. Data sources applied: 1) All random parameters (no data), 2) Discharge, 3) GRACE, 4) Altimetry data combined with GRACE (Altimetry Strategy 1), 5) Altimetry data using the rating curves combined with GRACE (Altimetry Strategy 2), and 6) Altimetry data using the Strickler – Manning equation combined with GRACE (Altimetry Strategy 3), and 7) Daily river water level combined with GRACE using the Strickler – Manning equation and cross-section information retrieved from Google Earth (Water level Strategy 1), or 8) obtained from a detailed field survey with an Acoustic Doppler Current Profiler (ADCP, Water level Strategy 2). The boxplots visualise the spread in the overall model performance D_E with respect to discharge and the following individual signatures: a) daily discharge ($E_{NS,Q}$), b) its logarithm ($E_{NS,logQ}$), c) flow duration curve ($E_{NS,FDC}$), d) its logarithm ($E_{NS,logFDC}$), e) average runoff coefficient during the dry season ($E_{R,RCDry}$), f) average seasonal runoff coefficient during the wet season ($E_{R,RCDry}$), g) autocorrelation function ($E_{NS,AC}$), and h) rising limb density ($E_{R,RLD}$). The dots visualise the model performance when selecting the parameter set with the highest model efficiency according to each parameter identification strategy.

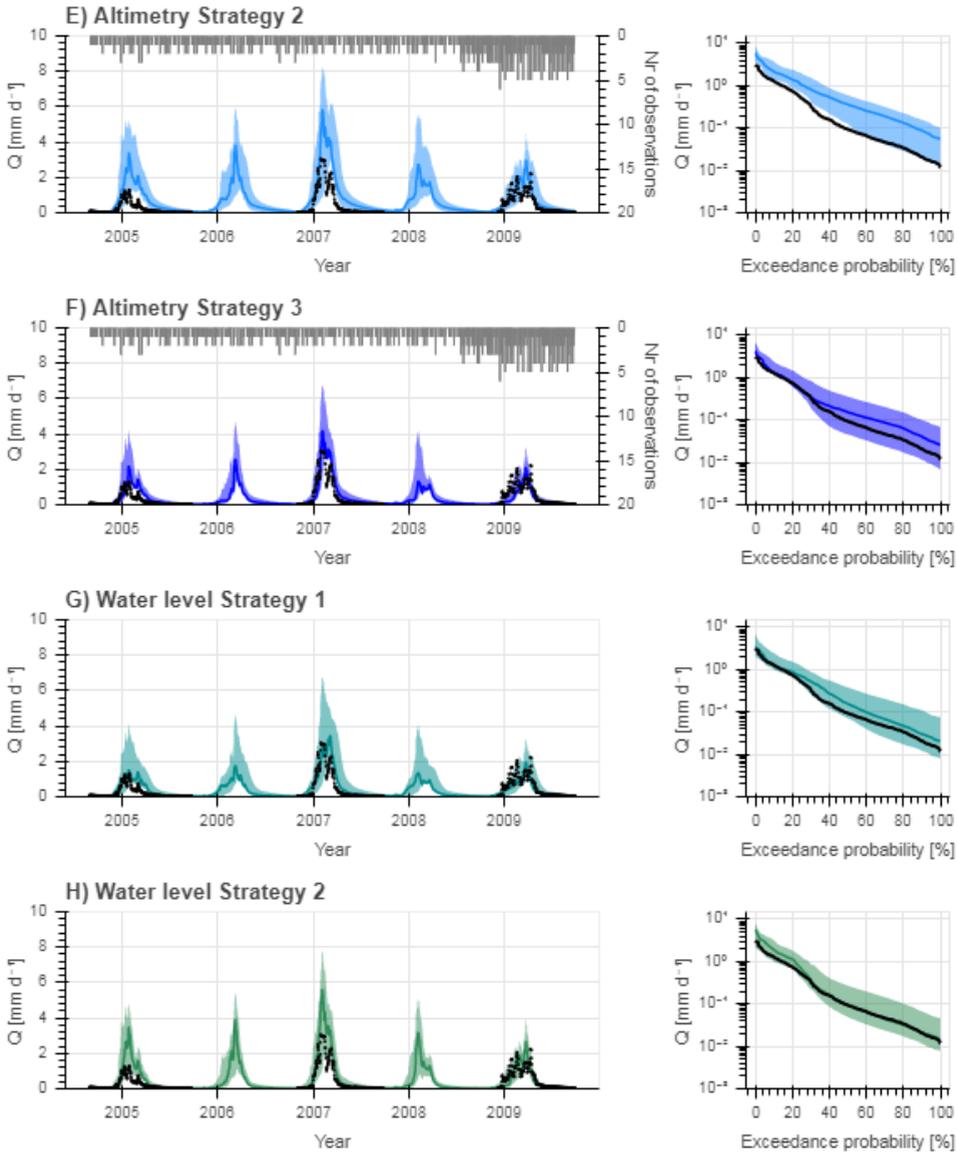


Figure 3.8: Range of model solutions. The left panel shows the hydrograph and the right panel the flow duration curve in logarithmic scale of the recorded (black) and modelled discharge: the line indicates the solution with the highest calibration objective function (E_{NS} or D_E) and the shaded area the envelope of the solutions retained as feasible. Solutions retained as feasible based on E) Altimetry Strategy 2 using rating curves for the discharge – water level conversion ($D_{E,NS,RC}$), F) Altimetry Strategy 3 using the Strickler-Manning equation for the discharge – water level conversion ($D_{E,NS,SM}$), and G) Daily in-situ water level using the Strickler Manning equation for the discharge – water level conversion with cross-section information retrieved from Google Earth (Water level strategy 1; $E_{NS,SM,GE}$) or H) obtained from a detailed field survey with an Acoustic Doppler Current Profiler (ADCP; Water level strategy 2; $E_{NS,SM,ADCP}$). The grey bars in the left subplots E and F indicate the number of altimetry observations available for each day.

Table 3.7: Summary of the model results: elimination of unfeasible parameter sets and detection of optimal parameter set according to each parameter identification strategy including the corresponding model performance range ($E_{NS,Q}$, D_E) indicating the model's skill to reproduce the discharge during the benchmark time period. For each strategy, the model efficiency for the optimal parameter set is summarised together with the corresponding performance metrics with respect to discharge ($E_{NS,Q,opt}$, $D_{E,opt}$). For all parameter sets retained as feasible, the maximum ($E_{NS,Q,max}$, $D_{E,max}$) and 5/95 percentiles ($E_{NS,Q,5/95}$, $D_{E,5/95}$) of all performance metrics with respect to discharge are summarised. Data sources used for the parameter set selection: 1) All parameter sets (no data), 2) Discharge, 3) GRACE, 4) Altimetry combined with GRACE (Altimetry Strategy 1), 5) Altimetry data using rating curves combined with GRACE (Altimetry Strategy 2), 6) Altimetry data using the Strickler – Manning equation combined with GRACE (Altimetry Strategy 3), and 7) Daily river water level combined with GRACE using the Strickler – Manning equation and cross-section information retrieved from Google Earth (Water level Strategy 1), or 8) obtained from a detailed field survey with an Acoustic Doppler Current Profiler (ADCP, Water level Strategy 2).

	Optimal parameter set		Feasible parameter sets	
	Model efficiency	$E_{NS,Q,opt}$ ($D_{E,opt}$)	$E_{NS,Q,max}$ ($E_{NS,Q,5/95}$)	$D_{E,max}$ ($D_{E,5/95}$)
1) All parameters sets	-	-	0.78 (-3.8 – 0.68)	0.79 (-1.4 – 0.71)
2) Discharge	$E_{NS,Q,opt} = 0.78$	0.78 (0.60)	0.78 (0.61 – 0.75)	0.79 (0.55 – 0.76)
3) Seasonal water storage (GRACE)	$E_{NS,Stot,opt} = 0.56$	-1.4 (-0.18)	0.78 (-2.3 – 0.38)	0.77 (-0.58 – 0.62)
4) Altimetry Strategy 1: Compare altimetry to discharge	$D_{E,R,WL,opt} = 0.76$	0.65 (0.63)	0.65 (-2.9 – 0.10)	0.66 (-0.83 – 0.50)
5) Altimetry Strategy 2: Rating curves	$D_{E,NS,RC,opt} = -0.50$	-0.31 (0.27)	0.51 (-2.6 – 0.25)	0.66 (-0.72 – 0.56)
6) Altimetry Strategy 3: Strickler-Manning equation	$D_{E,NS,SM,opt} = -1.4$	0.60 (0.71)	0.63 (-0.31 – 0.50)	0.75 (0.36 – 0.67)
7) Water level Strategy 1: satellite based cross-section	$E_{NS,SM,GE,opt} = -1.8$	0.65 (0.77)	0.77 (-0.48 – 0.60)	0.77 (0.28 – 0.70)
8) Water level Strategy 2: in-situ cross-section	$E_{NS,SM,ADCP,opt} = 0.79$	0.14 (0.55)	0.77 (-1.1 – 0.50)	0.77 (0.03 – 0.67)

Parameter selection based on the seasonal water storage (GRACE)

Starting from the set of all model realizations (Figures 3.6A and 3.7), and assuming no discharge observations are available, we identified and discarded parameter sets as unfeasible when they did not meet the previously defined criteria to reproduce the seasonal water storage ($E_{NS,Stot}$; see Section 3.4.3). The range of random model realizations with respect to the total water storage is visualised in Figure 3.9. The sub-set of solutions retained as feasible resulted in a significant reduction in the uncertainty around the modelled variables, which is illustrated by the narrower 5/95th percentiles of the solutions compared to the set of all realizations, as shown in Figure 3.6C. The feasible solutions with respect to the GRACE reached $E_{NS,Stot,opt} = 0.56$ ($E_{NS,Stot,5/95} = 0.45 – 0.52$) (Figure 3.7, Table 3.7). These parameter sets were then used to evaluate the model for the years 2004, 2006, 2008 used in the benchmark case. While the flow dynamics are captured relatively well, many of the retained solutions considerably overestimated flows across all seasons (Figure 3.6C) resulting in a decreased performance with respect to the individual flow signatures, only the dry runoff coefficient ($E_{R,RCdry}$) improved significantly compared to the

benchmark as shown in Table A.4 in Appendix A and Figure 3.7. The parameter set associated with the best performing model with respect to GRACE ($E_{NS,Stot,opt}$) resulted for the benchmark period in a $E_{NS,Q} = -1.4$ ($E_{NS,Q,5/95} = -2.3 - 0.38$) and the corresponding $D_{E,opt} = -0.18$ ($D_{E,5/95} = -0.58 - 0.62$) with respect to discharge (Figure 3.7, Table 3.7). As illustrated in Figure 3.7 and Figure 3.6C, many parameter sets that resulted in implausible representations of the seasonal signals were eliminated. However, as also indicated by the rather modest values of $E_{NS,Q}$ and D_E with respect to discharge, the data source used here obviously contained only limited information to avoid the over predictions of flow during all wet seasons. The sequence of applying first GRACE and then altimetry, or the reverse, did not affect the identification of feasible parameter sets when using altimetry data as shown in Figure A.8 in Appendix A. However, it did affect the selection of the “best” parameter set.

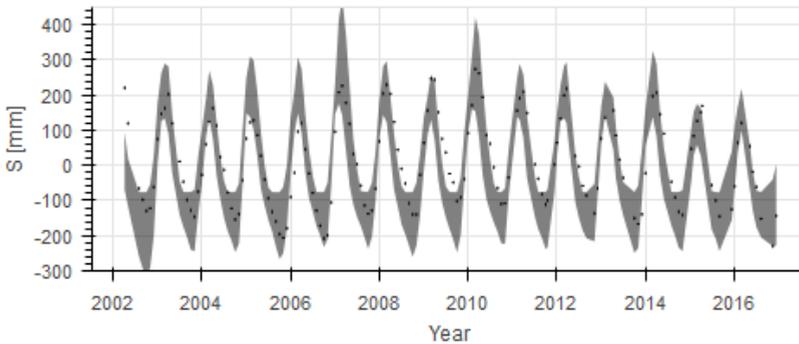


Figure 3.9: Range of random model realizations with respect to the total water storage (grey) including the observation according to GRACE (black)

Parameter selection based on satellite altimetry data

Altimetry Strategy 1: Directly compare altimetry data to modelled discharge The first approach, Altimetry Strategy 1, resulted in an overestimation of in particular intermediate and low flows as shown in Figure 3.6D. The feasible solutions reached an optimum of $D_{E,R,WL,opt} = 0.76$ ($D_{E,R,WL,5/95} = 0.74 - 0.75$) with respect to altimetry observations. Focusing on the model’s skill to reproduce the observed discharge using these feasible parameter sets for the benchmark period, the parameter set associated with the best performing model with respect to altimetry ($D_{E,R,WL,opt}$) resulted in a $E_{NS,Q} = 0.65$ ($E_{NS,Q,5/95} = -2.9 - 0.10$) and $D_E = 0.63$ ($D_{E,5/95} = -0.83 - 0.50$) with respect to discharge (Figure 3.7, Table 3.7). Hence, the parameter set with the highest model performance with respect to altimetry, did not perform best with respect to discharge as shown in Table 3.7 and Figure A.7 in Appendix A. While the optimum model performance with respect to discharge was similar to the benchmark, the very wide range in the 5/95th percentiles of the solutions indicated that this strategy has only limited potential to identify implausible parameter sets. This was also the case with respect to the individual flow signatures as shown in Figure 3.7 and Table A.4 in Appendix A.

Altimetry Strategy 2: Rating curves The second approach, Altimetry Strategy 2, also resulted in an overestimation of the flows (Figure 3.8E). The feasible solutions reached an optimum of $D_{E,NS,RC,opt} = -0.50$ ($D_{E,NS,RC,5/95} = -1.0 - -0.77$) with respect to altimetry observations. As example, Figure A.6A in Appendix A visualises the simulated and observed river water level at Virtual Station 4 (Figure 3.1) where the model significantly underestimated the stream levels. Focusing on the model's skill to reproduce the discharge using these parameter sets for the benchmark period, the parameter set associated with the best performing model with respect to altimetry ($D_{E,NS,RC,opt}$) resulted in $E_{NS,Q} = -0.31$ ($E_{NS,Q,5/95} = -2.6 - 0.25$) and $D_E = 0.27$ ($D_{E,5/95} = -0.72 - 0.56$) with respect to discharge (Figure 3.7, Table 3.7). Hence similar to Altimetry Strategy 1, the best parameter set with respect to altimetry, did not perform best with respect to discharge (Table 3.7 and Figure A.7 in Appendix A). The optimum model performance with respect to discharge was worse compared to the benchmark, and the wide range in the 5/95th percentiles of the solutions indicated this strategy poorly identified the feasible parameter sets. This was also the case with respect to the individual flow signatures as shown in Figure 3.7 and Table A.4 in Appendix A. Only the dry runoff coefficient ($E_{R,RCdry}$) improved significantly compared to the benchmark.

Altimetry Strategy 3: Strickler-Manning equation The third approach, Altimetry Strategy 3, resulted in improved flow predictions compared to the other two strategies using altimetry data (Figure 3.8F). Even though the feasible solutions exhibit a very poor ability to reproduce the altimetry data, with an optimum of $D_{E,NS,SM,opt} = -1.4$ ($D_{E,NS,SM,5/95} = -3.8 - -1.8$), the model's skill to reproduce the discharge for the benchmark period using these parameter sets, significantly increased compared to the two alternative strategies. As example, Figure A.6B in Appendix A visualises the simulated and observed river water level at Virtual Station 4 (Figure 3.1) where the model simulated the stream levels relatively well. The parameter set associated with the best performing model with respect to altimetry ($D_{E,NS,SM,opt}$) resulted in $E_{NS,Q} = 0.60$ ($E_{NS,Q,5/95} = -0.31 - 0.50$) and $D_E = 0.71$ ($D_{E,5/95} = 0.36 - 0.67$) with respect to discharge (Figure 3.7, Table 3.7). While the optimum model performance with respect to discharge was worse compared to the benchmark, the 5/95th percentiles of the solutions were significantly constrained by the removal of many implausible parameter sets. This was valid for the performance with respect to the individual flow signatures ($E_{NS,\theta}$ and $E_{R,\theta}$) and overall flow response (D_E) as shown in Figure 3.7 and Table A.4 in Appendix A. This indicated that, although the model performance with respect to altimetry observations was low, this strategy contains valuable information to considerably constrain the feasible solution space.

Parameter selection based on daily river water level at the basin outlet

Water level Strategy 1: River geometry information extracted from Google Earth The parameter identification strategy "Water level Strategy 1", using cross-section information extracted from Google Earth, resulted in a poor simulation of the river water level (Figure 3.10A) with an optimal objective function

value with respect to river water levels of $E_{NS,SM,GE,opt} = -1.8$ ($E_{NS,SM,GE,5/95} = -6.8 - -3.1$). Focusing on the model's skill to reproduce the discharge using these feasible parameter sets for the benchmark period, the parameter set associated with the best performing model with respect to river water levels ($E_{NS,SM,GE,opt}$) resulted in $E_{NS,Q} = 0.65$ ($E_{NS,Q,5/95} = -0.48 - 0.60$) and $D_E = 0.77$ ($D_{E,5/95} = 0.28 - 0.70$) with respect to discharge (Figure 3.7, Table 3.7). The model performance with respect to the remaining signatures as visualised in Figure 3.7 are tabulated in Table A.4 in Appendix A. As shown in Figure 3.8G, the discharge was overestimated in particular during intermediate and low flows.

Water level Strategy 2: River geometry information obtained from a detailed field survey The parameter identification strategy "Water level Strategy 2", using cross-section information obtained from a detailed field survey, resulted in improved river water level simulations (compare Figure 3.10A and B) with an optimal objective function value with respect to river water levels of $E_{NS,SM,ADCP,opt} = 0.79$ ($E_{NS,SM,ADCP,5/95} = 0.60 - 0.74$). The parameter set associated with the best performing model with respect to river water levels ($E_{NS,SM,ADCP,opt}$) resulted in $E_{NS,Q} = 0.14$ ($E_{NS,Q,5/95} = -1.1 - 0.50$) and in $D_E = 0.55$ ($D_{E,5/95} = 0.03 - 0.67$) with respect to discharge (Figure 3.7, Table 3.7); the model performance with respect to the remaining signatures as visualised in Figure 3.7 are tabulated in Table A.4 in Appendix A.

Compared to using river geometry information extracted from Google Earth (Water level Strategy 1), the overall model performance with respect to discharge did not increase since the parameter space was already restricted using GRACE data. However, the modelled flow duration curve during intermediate and low flows (compare Figure 3.8G with H) and rating curve (Figure 3.11) improved significantly when using more accurate geometry information obtained from a detailed field survey covering the cross-section that is submerged most of the year which is thus unlikely to be captured by satellite based observations. Note, that the in-situ cross-section information was limited to the submerged part during the time of measurement. The remaining part (water levels > 5 m) was extrapolated which is likely to explain the larger discrepancies during high flows visible in the flow duration curve (Figure 3.8H).

3.5.2. Number of virtual stations used for model calibration and evaluation

In this study, altimetry data was available at 18 virtual stations. However, would the model performance change if more or less virtual stations were used? To answer this question, n random stations were selected for model calibration; the remaining stations were used for cross-validation (Garavaglia et al., 2017; Gharari et al., 2013; Klemeš, 1986). This was repeated to cover all combinations of n stations and for $n = 1, 2 \dots 17$. When applying Strategy 3 using altimetry data with the Strickler-Manning equation, this analysis revealed that when increasing the number of calibration stations, the model calibration performance $D_{E,NS,SM}$ gradually decreased, but the ability to meaningfully reproduce the remaining observations

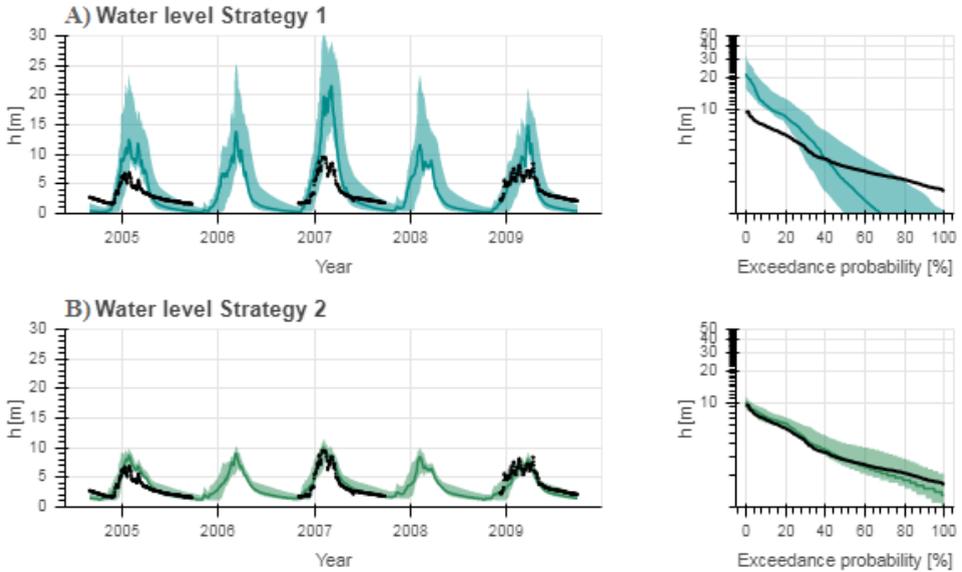


Figure 3.10: Range of model solutions. The left panel shows the hydrograph and the right panel the flow duration curve of the recorded (black) and modelled discharge: the line indicates the solution with the highest calibration objective function (E_{NS}) and the shaded area the envelope of the solutions retained as feasible. Solutions were retained as feasible based on daily water level time series at the basin outlet using the Strickler-Manning equation for the discharge – water level conversion; the cross-section was A) extracted from Google Earth (Water level Strategy 1), or B) obtained from a detailed field survey with an Acoustic Doppler Current Profiler (ADCP, Water level Strategy 2).

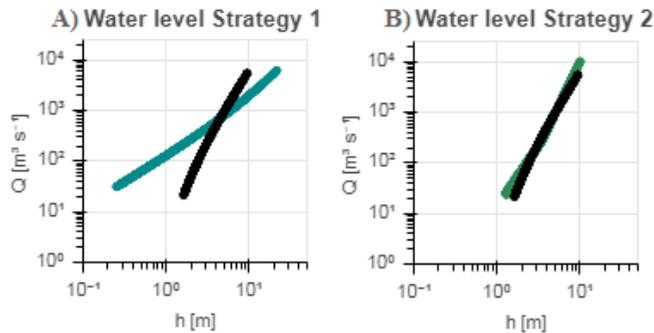


Figure 3.11: Discharge - water level graphs for the recorded (black) and modelled discharge and stream levels with the optimal model performance (E_{NS}) using the Strickler Manning equation for the discharge – stream level conversion with cross-section information A) extracted from Google Earth (Water level Strategy 1), or B) obtained from a detailed field survey with an Acoustic Doppler Current Profiler (ADCP, Water level Strategy 2).

which were not used for calibration increased significantly (Figure 3.12). Similar results were obtained for Strategies 1 and 2 (compare Figure 3.12 with Figures

A.3 and A.4 in Appendix A). Also the model performance with respect to discharge increased when using more virtual stations with an optimum at 7 – 15 stations depending on the calibration strategy (Figure A.5 in Appendix A). This provides evidence that in spite of reduced calibration performance, the simultaneous use of multiple virtual stations can contribute towards more plausible selections of model parameter sets and thus increase the model realism.

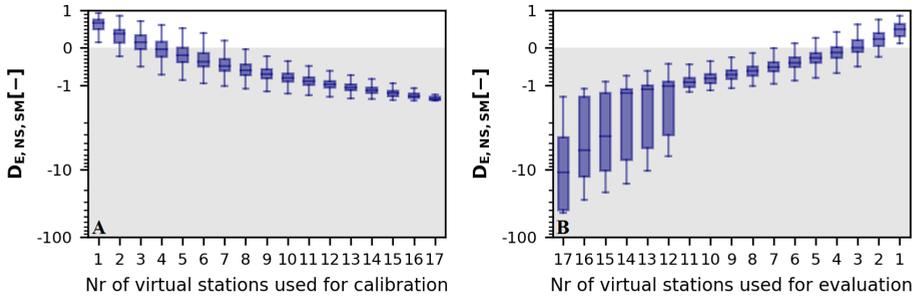


Figure 3.12: Influence of the number of virtual stations used for A) model calibration and B) evaluation on the model performance $D_{E,NS,SM}$ applying Altimetry Strategy 3.

3.5.3. Uncertainties and limitations

In the absence of discharge data for hydrological model calibration as commonly the case in poorly or ungauged regions, freely and globally available remotely sensed stream water levels could provide the opportunity to fill this gap as illustrated in this study, as well as in previous studies (e.g. Michailovsky and Bauer-Gottwein, 2014; Pereira-Cardenal et al., 2011; Sun et al., 2012). However, there are several limitations to the approach proposed in this study using altimetry for model calibration.

First, river altimetry data are prone to large uncertainties which increase for smaller river widths as a result of back-scatter effects of the surrounding topography (Biancamaria et al., 2017; Domeneghetti et al., 2015; Sulistioadi et al., 2015). Too small rivers could even be missed altogether. In this study, the Luangwa river becomes a small meandering stream in the dry season resulting in larger altimetry uncertainties. Unfortunately, this uncertainty could not be estimated for the virtual stations used in this study due to data limitations. However, in previous studies in the Zambezi Basin, the RMSE relative to in-situ stream levels ranged between 0.32 m and 0.72 m using Envisat (Michailovsky et al., 2012). Improving altimetry observations such that the uncertainties decrease would improve the identification of feasible parameter sets and simulation of stream levels and flow. However, comparison results between the three altimetry based calibration strategies are not expected to change since the same altimetry data were used. In other words, Altimetry Strategy 3 is still expected to perform best when decreasing the uncertainties in the altimetry observations. Second, large uncertainties in the forcing data (precipitation and temperature) with respect to the spatial-temporal variations

should not be ignored. This could compromise comparison results between modelled river water levels and altimetry within the basin since it has a low temporal resolution (10 or 35 days). Bias in the precipitation data affects storage calculations and hence the identification of feasible parameter sets based on GRACE (Le Coz and van de Giesen, 2020). This could explain why the flows were frequently overestimated when using GRACE only. In addition, precipitation bias could be compensated through calibration parameters introduced for the discharge – water level conversion. Therefore, such parameters should be constrained as much as possible. There are also data uncertainties in the cross-sections and river gradients extracted from high-resolution terrain data available on Google Earth due to its limited spatial resolution, but more importantly since no information is available below the water surface.

Further, GRACE observations are prone to uncertainties as a result of data (post-) processing including for example data smoothing (Blazquez et al., 2018b; Landerer and Swenson, 2012; Riegger et al., 2012) causing leakage between neighbouring cells of 1° (≈ 111 km) which are thus not completely independent from each other. Additionally, GRACE observations are more accurate for large areas. Depending on the applied processing scheme, the error is about 2 cm for basins with an area of around $63\,000$ km² (Landerer and Swenson, 2012; Vishwakarma et al., 2018). Also note that due to the coarse temporal resolution, monthly averaged GRACE observations are dominated by slow changing processes such as the groundwater and soil moisture system and seasonal variations reflected in all storage components. In addition, open water bodies or wetlands could affect GRACE observations if they are located in or near the basin, for example within a radius of about 300 km which is the distance often used for data smoothing. In this study, several open water bodies or wetlands were located ≤ 300 km of the Luangwa basin such as Lake Malawi, Kafue Flats, Cahora Bassa reservoir, Kariba reservoir, Bangweulu and Tanganyika. These open water bodies and wetlands had a limited impact on the GRACE observations due to limited fluctuations or different temporal variation as illustrated in Figure 3.13 for the Cahora Bassa reservoir. These uncertainties in the GRACE observations could influence the identification of plausible parameter sets. For example feasible parameter sets could be discarded incorrectly which could distort results obtained by calibrating with respect to altimetry and GRACE simultaneously. However, the comparison between the three altimetry based calibration strategies is not expected to change since the same GRACE data were used. In other words, Altimetry Strategy 3 is still expected to perform best when considering these uncertainties.

Uncertainties were not only introduced by the data, but also as a result of assumptions and simplifications. First, the reference level h_0 was assumed to be equal to the lowest river water level observed to limit the number of calibration parameters (Altimetry Strategy 2 and 3, Water level Strategy 1 and 2). However, uncertainties in the altimetry observations as explained previously influence h_0 estimates which results in a bias between the observed and simulated stream levels in Altimetry Strategies 2 and 3. Second, the roughness was assumed to be constant in time, over the entire cross-section and for all virtual stations throughout the

basin (Altimetry Strategy 3). However, this roughness can vary between 15 – 50 $\text{m}^{1/3} \text{s}^{-1}$ for natural rivers (Chow, 1959; Vatanchi and Maghrebi, 2019) changing the simulated stream levels between 42% – 75% in the Luangwa Basin with the low flows being the most sensitive. Third, all 18 virtual stations were grouped based on their cross-section similarity to limit the number of calibration parameters (Altimetry Strategy 2), but differences within each group remain such that the calibration parameters related to the rating curve varies slightly for each virtual station within a group. Fourth, the assumption of a constant flow velocity in space and time affects the timing of the simulated flow and stream levels influencing the comparison between model results and altimetry observations (all strategies).

Another limitation is the missing flow volume information when directly using (satellite based) river water levels for model calibration, using the Spearman Rank Correlations as model performance metric (Altimetry Strategy 1; Seibert and Vis, 2016a). This resulted here in an overestimation of intermediate and low flows due to the non-linear relation between stream levels and flows. In contrast, when converting the discharge to stream water levels, flow volume information was included at the cost of introducing additional calibration parameters (Altimetry Strategy 2 and 3), thereby increasing the degrees-of-freedom and thus the potential for parameter equifinality in the model (Beven, 2006; Sikorska and Renard, 2017; Sun et al., 2012).

Furthermore, it was assumed the Nash-Sutcliffe efficiency contained sufficient valuable information to describe the model performance with respect to river water level and total water storage when identifying feasible parameter sets. This performance measure is sensitive to the sample size, outliers, bias and time-offset (McCuen Richard et al., 2006). Unfortunately, simulated discharge and stream levels are prone to bias uncertainties as a result of spatiotemporal bias in the rainfall (Le Coz and van de Giesen, 2020). In addition, altimetry observations have a limited sample size for several virtual stations (Table 3.2) and are prone to bias due to uncertainties in the reference level h_0 as mentioned before. Moreover, a time-offset in the simulated flow can occur as a result of rainfall uncertainties. As comparison, the model performance with respect to altimetry only reached up to $D_{E,NS,SM} = -1.3$ for Altimetry Strategy 3, while it reached up to $E_{NS,SM,GE} = 0.61$ with respect to daily in-situ stream levels for Water level Strategy 1. Therefore, additional study is recommended to confirm this assumption and to assess which performance metric(s) would be most suitable. The model performance with respect to discharge was evaluated with respect to multiple hydrological signatures simultaneously (Table 3.6) to assess the model's skill to reproduce the internal dynamics of the system. Even though a few of these signatures have some overlapping information content (McMillan et al., 2017), each signature also contains at least some additional information not included in the other signatures. In general, the ambition is to represent a hydrological system as good as possible in a model which critically required that the model exhibits sufficient ability to simultaneously reproduce multiple flow signatures (Euser et al., 2013; Gupta et al., 2008; Hrachowitz et al., 2014).

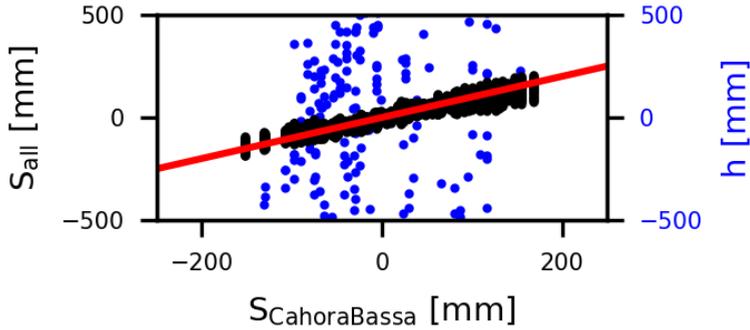


Figure 3.13: Temporal correlation of the GRACE observations for the cell in which the virtual station for Cahora Bassa is located (horizontal axis) and for A) all cells within an area surrounding the virtual station with a radius of 3 degree (GRACE area of influence, vertical axis, black), and B) the altimetry observation at Cahora Bassa (vertical axis, blue). The 1:1 line is visualised in red. The relatively strong temporal correlation between the GRACE cells could be a result of the strong seasonality in this area.

3.5.4. Comparison with previous studies

Previous studies have successfully used river altimetry data to calibrate and evaluate rainfall-runoff models using a few virtual stations (Getirana, 2010; Getirana et al., 2010; Liu et al., 2015; Sun et al., 2012). In these studies, the modelled discharge was converted to stream levels by means of a hydraulic model or empirical relations. Our results support several previous findings and added a number of new ones.

Similar to previous studies, the rainfall-runoff model reproduced river flow relatively well when calibrating on remotely sensed stream water levels preferably at several virtual stations simultaneously, but discharge based calibration results performed significantly better (Getirana, 2010). Thus, while river altimetry data cannot fully substitute discharge observations, they at least provide an alternative data source that holds information value where no reliable discharge data are available. In addition, our results suggest that in spite of the typically limited temporal resolution of altimetry observations, these data, when using multiple virtual stations simultaneously, provide enough information to select meaningful model parameter sets (Getirana, 2010; Seibert and Beven, 2009).

Strikingly, only limited studies combined altimetry with GRACE observations in the calibration procedure (Kittel et al., 2018). As altimetry observations only describe water level variations with no information on the flow amounts, GRACE provides additional valuable information to constrain the river discharge by improving the rainfall runoff partitioning as demonstrated in previous studies (Bai et al., 2018; Dembélé et al., 2020; Rakovec et al., 2016). Combining both data sources in the calibration procedure allowed for a more accurate identification of feasible parameter sets. The model performance range with respect to discharge improved from $D_{E,5/95} = -8.4 - 0.77$ when using only altimetry to $D_{E,5/95} = 0.19 - 0.75$ when combining GRACE and altimetry for Altimetry Strategy 3 (Figure A.8 in Appendix

A).

In contrast to previous studies, altimetry data originated from five different satellite missions rather than a single one. As a result, altimetry data was available at 18 locations for the time period 2002 to 2016. This gave the opportunity to analyse the effect of combining different numbers of stations for calibration and evaluation. This study illustrated that better predictions can be achieved when using more virtual stations for calibration. Furthermore, this study demonstrated that in particular the combination of altimetry with information on river geometry (cross section, gradient) proved beneficial for the selection of feasible parameter sets within relatively narrow bounds comparable to the benchmark using discharge. Using more accurate cross-section information obtained from a detailed field survey rather than Google Earth based estimates, improved the water level simulations, modelled rating curve and discharge simulations during intermediate and low flows significantly for which on-site cross-section data was available. That is why it is recommended to acquire accurate cross-section information on locations concurring with altimetry overpasses (not done in this study).

3.5.5. Opportunities for future studies

For future studies, it will be interesting to improve Altimetry Strategy 3 using additional data sources. For instance, the combination of altimetry observations with river width estimates derived from Landsat or Sentinel-1/2 (Hou et al., 2018; Pekel et al., 2016) may bear some potential as the combination of the two different hydraulic variables complements each other and increases the temporal sampling (Huang et al., 2018; Sichangi et al., 2016; Tarpanelli et al., 2017). For instance, during high flows river width estimates can be more accurate than altimetry observations especially when floodplains are inundated and small water level changes cause large river width changes. Alternatively, the altimetry observations used here could be combined with river surface water level slope estimates based on CryoSat observations which provide water level information at lower temporal resolution (every 369 days), but higher spatial resolution (equatorial inter-track distance of 7.5 km) (Jiang et al., 2017; Schneider et al., 2017). This allows for the estimation of the energy gradient based on stream levels as required in the Strickler-Manning equation, instead of the bed slope based on topography, which proved to be a good first estimate in absence of more reliable data. In addition, CryoSat observations are available annually such that there can be more overlap with altimetry observations in contrast to topography data. In addition, with the upcoming SWOT (Surface Water Ocean Topography) mission, more accurate altimetry observations should be available as well as river slope observations and width. The repeat cycle will be 21 days and across-track resolution between 10 m and 60 m increasing the number of observation points available within a specific area (Biancamaria et al., 2016; Langhorst et al., 2019; Oubanas et al., 2018). As a result, hydrological models can be calibrated with respect to river altimetry and width simultaneously at multiple locations even for small river basin improving the identification of plausible parameters sets and hence the model realism as illustrated in Section 3.5.2. It will also

be very valuable to improve cross-section estimates with respect to the submerged part of the cross-section as already explored in previous studies (Domeneghetti, 2016) or to use drone observations to obtain more accurate cross-section information and estimates of the river slope and roughness (Entwistle and Heritage, 2019). By improving the river profile description, the simulated stream levels become more accurate which is crucial when using this time series for model calibration. As illustrated with Water level Strategies 1 and 2, improving the cross-section resulted in a more accurate rating curve (Figure 3.11), stream level simulation (see Figure 3.10), and discharge simulation (Figure 3.8). Clearly, it will be interesting to analyze and disentangle different individual sources of uncertainty related to the discharge – water level conversion from the hydrological model in a more data rich region (Renard et al., 2010). Unfortunately, this was not possible in this study due to the scarcely available in-situ observations in the Luangwa. As concluded by Renard et al. (2010), reliable estimates of the data uncertainty are required to disaggregate multiple sources on uncertainty in rainfall-runoff modelling successfully.

3.6. Evaluation

This study investigated the potential value of river altimetry observations from multiple satellite missions to identify feasible parameters for a hydrological model of the semi-arid and poorly gauged Luangwa River Basin. A distributed process-based rainfall-runoff model with sub-grid process heterogeneity was developed on a daily timescale for the time period 2002 to 2016. Various parameter identification strategies were implemented step-wise to assess the potential of satellite altimetry data for model calibration. As a benchmark, when identifying parameter sets with the traditional model calibration strategy using discharge data, the model was able to simulate the flows relatively well ($E_{NS,Q} = 0.78$, $E_{NS,Q,5/95} = 0.61 - 0.75$). When assuming no discharge observations are available, the feasible parameter sets were restricted with GRACE data only resulting in an optimum of $E_{NS,Q} = -1.4$ ($E_{NS,Q,5/95} = -2.3 - 0.38$) with respect to discharge. Combining GRACE with altimetry data only from 18 virtual stations focusing on the water level dynamics resulted in frequently overestimated flows and poorly identified feasible parameter sets (Altimetry Strategy 1, $E_{NS,Q,5/95} = -2.9 - 0.10$). This was also the case when converting modelled discharge to water levels using rating curves (Altimetry Strategy 2, $E_{NS,Q,5/95} = -2.6 - 0.25$). The identification of the feasible parameter sets improved when including river geometry information, more specifically cross-section and river gradient extracted from Google Earth, in the discharge-water level conversion using the Strickler-Manning equation (Altimetry Strategy 3, $E_{NS,Q} = 0.60$, $E_{NS,Q,5/95} = -0.31 - 0.50$). Moreover, it was shown that more accurate cross-section data improved the water level simulations, modelled rating curve and discharge simulations during intermediate and low flows for which on-site cross-section information was available; the Nash-Sutcliffe efficiency with respect to river water levels increased from $E_{NS,SM,GE} = -1.8$ ($E_{NS,SM,GE,5/95} = -6.8 - -3.1$) using river geometry information extracted from Google Earth (Water level Strategy 1) to $E_{NS,SM,ADCP} = 0.79$ ($E_{NS,SM,ADCP,5/95} = 0.60 - 0.74$) using river geometry information obtained from a detailed field survey (Water level Strategy 2). The model performance also im-

proved when increasing the number of virtual stations used for parameter selection. Therefore, in the absence of reliable discharge data as commonly the case in poorly or ungauged basins, altimetry data from multiple virtual stations combined with GRACE observations have the potential to fill this gap if combined with river geometry estimates.

4

Model structure development using satellite observations

*A bridge is repaired only
when someone falls into the water.*

African Proverb

Satellite observations can provide valuable information for a better understanding of hydrological processes and thus serve as valuable tools for model structure development and improvement. While model calibration and evaluation has in recent years started to make increasing use of spatial, mostly remotely-sensed information, model structural development largely remains to rely on discharge observations at basin outlets only. Due to the ill-posed inverse nature and the related equifinality issues in the modelling process, this frequently results in poor representations of the spatiotemporal heterogeneity of system-internal processes, in particular for large river basins. In this chapter, it was shown that satellite-based evaporation and total water storage anomaly data are not only valuable for multi-criteria calibration, but can play an important role in improving our understanding of hydrological processes through diagnosing model deficiencies and step-wise model structural improvement.

This chapter is based on: Hulsman, P., Savenije, H. H. G., and Hrachowitz, M.: Learning from satellite observations: increased understanding of catchment processes through stepwise model improvement, Hydrol. Earth Syst. Sci. Discuss., 2020, 1–26, doi: 10.5194/hess-2020-191, 2020b.

4.1. Introduction

Traditionally, discharge observations at basin outlets are used for hydrological model development and calibration, which can be a robust strategy in small watersheds with relatively uniform characteristics such as topography and land cover, but not for larger, heterogeneous basins (Blöschl and Sivapalan, 1995; Daggupati et al., 2015). As a result, temporal dynamics of discharge may be well reproduced. This however, does not ensure that the spatial pattern and temporal dynamics of model internal storage and flux variables provide a meaningful representation of their real pattern and dynamics (Beven, 2006; Clark et al., 2008; Garavaglia et al., 2017; Gupta et al., 2008; Hrachowitz et al., 2014; Kirchner, 2006). Especially in large, poorly gauged basins this traditional model calibration and testing method is likely to result in a poor representation of spatial variability (Daggupati et al., 2015) due to equifinality and the related the boundary flux problem (Beven, 2006).

An increasing number of satellite-based observations have become available over the last decade, giving us insight into a wide range of hydrology-relevant variables, including precipitation, total water storage anomalies, evaporation, surface soil moisture or river width (Jiang and Wang, 2019; Xu et al., 2014). These data are increasingly used as model forcing or for parameter selection and model calibration (e.g. Li et al., 2015; Mazzoleni et al., 2019; Tang et al., 2019).

Many studies used a single satellite product in the calibration procedure, some of them additionally using discharge data, others not. For instance, hydrological models have been calibrated with respect to evaporation (e.g. Bouaziz et al., 2018; Immerzeel and Droogers, 2008; Odusanya et al., 2019; Vervoort et al., 2014; Winsemius et al., 2008), water storage anomalies from GRACE (Gravity Recovery and Climate Experiment; Werth et al., 2009), river width (Revilla-Romero et al., 2015; Sun et al., 2018a) or river altimetry (Getirana, 2010; Hulsman et al., 2020a; Michailovsky et al., 2013; Sun et al., 2015b).

Other studies simultaneously calibrated hydrological models with respect to multiple remotely-sensed variables, but only exploiting basin-average time series, without consideration for spatial pattern (e.g. Kittel et al., 2018; López López et al., 2017; Milzow et al., 2011; Nijzink et al., 2018). On the other hand, some studies exclusively calibrated models to spatial pattern of the observed variables (Demirel et al., 2018; Koch et al., 2016; Mendiguren et al., 2017; Stisen et al., 2011; Zink et al., 2018). As most satellite-based observations such as evaporation are not measured directly but are themselves a result of underlying models using satellite data as input (Xu et al., 2014), more focus has been recently placed on calibration to the relative spatial variability instead of using absolute magnitudes (Dembélé et al., 2020; Stisen et al., 2011; van Dijk and Renzullo, 2011).

To fully exploit the information content of satellite-based observations, simultaneous model calibration on both, temporal dynamics and spatial pattern of multiple variables has the potential to improve the representation of spatiotemporal variability and, linked to that, their underlying model internal processes and therefore the model realism (Herman et al., 2018; Rakovec et al., 2016; Rientjes et al., 2013). Strikingly, only a few studies so far used satellite-based observations to calibrate with respect to the temporal and spatial variation simultaneously (Dembélé et al.,

2020; Rajib et al., 2018).

In general, most studies that made use of remotely-sensed data for model applications have exclusively addressed the problem of parameter selection and thus model calibration. However, as models are always abstract and simplified representations of reality, every model structure needs to be understood as a hypothesis to be tested (Clark et al., 2011; Fenicia et al., 2011; Hrachowitz and Clark, 2017). Yet, most studies on model structural improvement have so far only relied on spatially aggregated variables (Fenicia et al., 2008; Hrachowitz et al., 2014; Kavetski and Fenicia, 2011; Nijzink et al., 2016b), while spatial data remain rarely used for that purpose (e.g. Fenicia et al., 2016; Roy et al., 2017).

The overall objective of this study is therefore to explore the simultaneous use of spatial pattern and temporal dynamics of satellite-based evaporation and total water storage observations for a step-wise structural improvement and calibration of hydrological models for a large river systems in a semi-arid, data scarce region. More specifically, we tested the research hypotheses that (1) spatial pattern and temporal dynamics in satellite-based evaporation and water storage anomaly data contain relevant information to diagnose and to iteratively improve on model structural deficiencies and that (2) these data, when simultaneously used with discharge data for calibration, do contain sufficient information for a more robust parameter selection.

4.2. Site description

The Luangwa River in Zambia is a large, mostly unregulated tributary of the Zambezi with a length of about 770 km (Figure 4.1). This poorly gauged river basin has an area of 159,000 km² which is mostly covered with deciduous forest, shrubs and savanna and where an elevation difference up to 1850 m can be found between the highlands and low lands along the river (Hulsman et al., 2020a; The World Bank, 2010). In this semi-arid basin, the mean annual evaporation (1555 mm yr⁻¹) exceeds the mean annual precipitation (970 mm yr⁻¹). See Chapter 2 for more detailed information on the Luangwa basin.

4.3. Data availability

In-situ discharge observations

Historical daily in-situ discharge data was available from the Zambian Water Resources Management Authority at the Great East Road Bridge gauging station, located at 30° 13' E and 14° 58' S (Figure 4.1), for the time period 2002 to 2016 yet containing considerable gaps resulting in a temporal coverage of 53%.

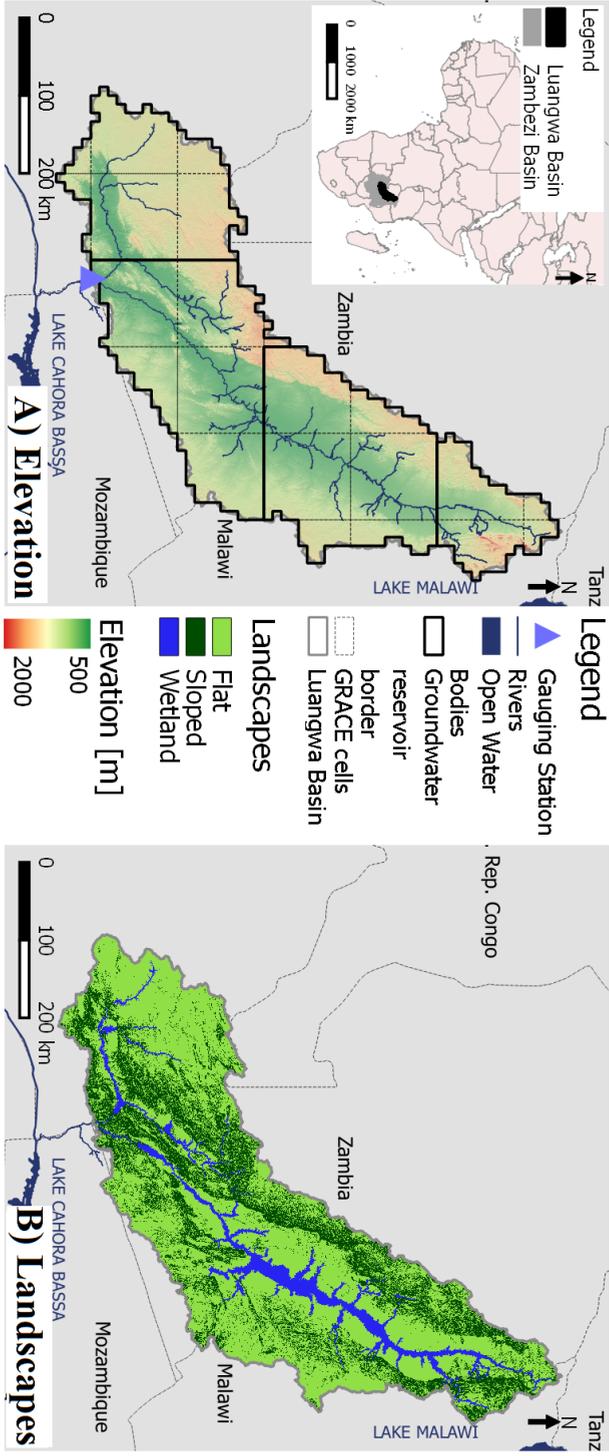


Figure 4.1: Map of the Luangwa River Basin in Zambia with A) the elevation, groundwater reservoir units at 0.1° resolution and 1° grid according to GRACE, and B) the main landscape types.

Spatially gridded observation

Spatially gridded data were used for a topography-based landscape classification into hydrological response units (HRU; Savenije, 2010), as model forcing (precipitation and temperature) and for parameter selection (evaporation and total water storage, see Table 4.1).

More specifically, topography was extracted from GMTED with a spatial resolution of 0.002° (Danielson and Gesch, 2011). Daily precipitation data was extracted from CHIRPS (Climate Hazards Group InfraRed Precipitation with Station) with a spatial resolution of 0.05° . Monthly temperature data extracted from CRU at a spatial resolution of 1° was used to estimate the potential evaporation applying the Hargreaves method (Hargreaves and Allen, 2003; Hargreaves and Samani, 1985). These monthly observations were interpolated to daily timescale using daily averaged in-situ temperature measured at two locations with the coordinates $28^\circ 30'$ E, $14^\circ 24'$ S and $32^\circ 35'$ E, $13^\circ 33'$ S. The satellite-based total evaporation data was extracted from WaPOR (Water Productivity Open-access portal; FAO, 2018) version 1.1 as it proved to perform well in African river basins (Weerasinghe et al., 2020). This product was available on 10-day temporal and 250 m spatial resolution. Satellite-based observations on the total water storage anomalies were extracted from the Gravity Recovery and Climate Experiment (GRACE). With two identical GRACE satellites, the variations in the Earth's gravity field were measured to detect regional mass changes which are dominated by variations in the terrestrial water storage after having accounted for atmospheric and oceanic effects (Landerer and Swenson, 2012; Swenson, 2012). In this study, the long term bias between the discharge, evaporation (WaPOR) and total water storage anomalies (GRACE) was corrected by multiplying the evaporation with a correction factor of 1.08 to close the long term water balance.

The gridded information provided for the precipitation, temperature and evaporation were rescaled to the model resolution of 0.1° . If the resolution of the satellite product was higher than 0.1° , then the mean of all cells located within each model cell was used. Otherwise, each cell of the satellite product was divided into multiple cells such that the model resolution is obtained, retaining the original value. In contrast, the modelled total water storage was rescaled to 1° , the resolution of the GRACE data set, by taking the mean.

4.4. Modelling approach

A previously developed and tested (Hulsman et al., 2020a) distributed, process-based hydrological model was implemented for the Luangwa Basin, see Section 4.4.1 for more information. This benchmark model (Model A) was calibrated with respect to discharge for the time period 2002 – 2012 and validated for the time period 2012 – 2016 with respect to discharge, evaporation and total water storage anomalies. Then, the model was calibrated with respect to all above variables, hence discharge, evaporation and total water storage anomalies simultaneously, for the time period 2002 – 2012 and validated with respect to the same variables for the time period 2012 – 2016. Model deficiencies were then diagnosed for this benchmark model (Model A) based on the results of both calibration strategies.

Table 4.1: Data used in this study

	Time period	Time resolution	Spatial resolution	Product name	Source
Digital elevation map	n/a	n/a	0.02°	GMTED	(Danielson and Gesch, 2011)
Precipitation	2002 – 2016	Daily	0.05°	CHIRPS	(Funk et al., 2014)
Temperature	2002 – 2016	Monthly	0.5°	CRU	(University of East Anglia Climatic Research Unit et al., 2017)
Evaporation	2009 – 2016	10-day	0.00223°	WaPOR	(FAO, 2018; FAO and IHE Delft, 2019)
Total water storage	2002 – 2016	Monthly	1°	GRACE	(Landerer and Swenson, 2012; Swenson, 2012; Swenson and Wahr, 2006)
Discharge (Luangwa Bridge gauging station)	2002 – 2016	Daily	n/a	n/a	WARMA

Next, model structure changes were applied creating Models B – D to improve the deficiencies found in Model A. These changes concerned the groundwater upwelling into the unsaturated zone as explained in Section 4.5.2. The same calibration and validation strategies as applied to Model A were applied to Models B – D. Model improvements were evaluated and further deficiencies were diagnosed for these models based on the calibration and validation results.

To improve the deficiencies diagnosed in Models B – D, further model structural changes, i.e. increased levels of spatial discretisation of the saturated zone as explained in Section 4.5.3, resulted in Models E and F. Similar to the previous models, the same calibration and validation strategies were applied to Models E and F, and model improvements and deficiencies were diagnosed based on the calibration and validation model performances.

The calculation of the model performance with respect to discharge, evaporation and total water storage are explained in Section 4.4.2. The calibration and validation procedures are described in Sections 4.4.3 and 4.4.4.

4.4.1. Hydrological models

Benchmark model (Model A)

This model is a process-based hydrological model developed in a previous study by Hulsman et al. (2020a) for the Luangwa basin. In this model, the water accounting was distributed by discretizing the basin and using spatially distributed forcing data while the same model structure and parameter set were used for the entire basin. Each 0.1° x 0.1° model cell was then further discretized into functionally distinct landscape classes, i.e. hydrological response units (HRU), inferred from topography (Figure 4.1B), but connected by a common groundwater component (Euser et al., 2015) following the FLEX-Topo modelling concept (Savenije, 2010) which was previously successfully applied in many different and climatically contrasting regions (Gao et al., 2014a, 2016; Gharari et al., 2014; Nijzink et al., 2016b). Here, the landscape was classified based on the local slope and “Height-above-the-nearest-drainage” (HAND; Rennó et al., 2008) into sloped areas (slope ≥ 4%), flat areas (slope < 4%, HAND ≥ 11 m) and wetlands (slope

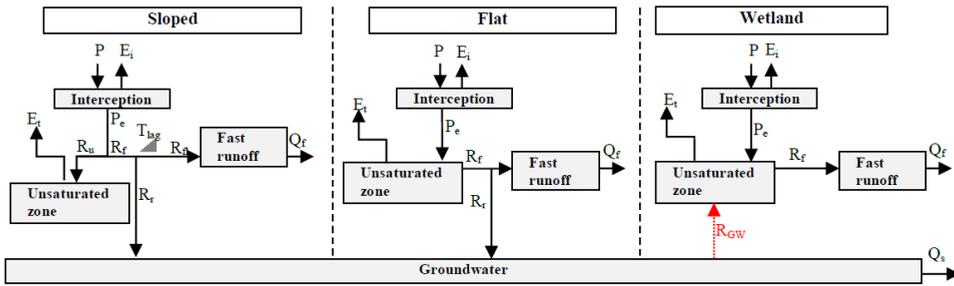


Figure 4.2: Schematisation of the model structure applied to each grid cell. Symbol explanation: precipitation (P), effective precipitation (P_e), interception evaporation (E_i), plant transpiration (E_t), infiltration into the unsaturated root zone (R_u), drainage to fast runoff component (R_f), delayed fast runoff (R_d), lag time (T_{lag}), groundwater recharge (R_r), upwelling groundwater flux (R_{GW}), fast runoff (Q_f), groundwater/slow runoff (Q_s).

< 4%, HAND < 11 m). For this purpose, the drainage network was derived from a digital elevation map extracted from GMTED (Section 4.3) using a flow accumulation map after having burned-in a river network map extracted from OpenStreetMap (<https://wiki.openstreetmap.org/wiki/Shapefiles>) to obtain an as accurate as possible drainage network as done successfully in previous studies (Seyler et al., 2009). According to this classification, the wetland areas covered 8% of the basin, flat areas 64% and sloped areas 28% (Figure 4.1).

The model consisted of different storage components schematised as reservoirs representing interception and unsaturated storage, as well as a slow responding reservoir, representing the groundwater and a fast responding reservoir (Figure 4.2). The water balance for each reservoir and the associated constitutive equations are summarised in Table 4.2. The individual model structures of each parallel HRU were very similar. Functional differences between HRUs were thus mostly accounted for by different parameter sets. To allow the use of partly overlapping prior parameter distributions while maintaining relationships between parameters of individual HRUs that are consistent with our physical understanding of the system and to limit equifinality, model process constraints (Gharari et al., 2014; Hrachowitz et al., 2014) were applied for several parameters (Table 4.3). For instance, in the Luangwa Basin, the sloped areas are dominated by dense vegetation, suggesting higher interception capacities and larger storage capacities in the unsaturated zone compared to the remaining part of the basin. In addition, for each HRU the model structure was adjusted where necessary to include processes unique to that area. For instance, water percolates and recharges the groundwater system in sloped and flat areas whereas in wetlands this was assumed to be negligible due to groundwater tables that are very shallow and thus close to the surface.

The runoff was first calculated for each individual grid cell. A simple routing scheme based on the flow direction and constant flow velocity as calibration parameter was applied to estimate the flow at the outlet. In total, this model consisted of 16 calibration parameters with uniform prior distributions and constraints as summarized in Table 4.3.

Table 4.2: Equations applied in the hydrological model. Fluxes [mm d⁻¹]: precipitation (P), effective precipitation (P_e), potential evaporation (E_p), interception evaporation (E_i), plant transpiration (E_t), infiltration into the unsaturated zone (R_u), drainage to fast runoff component (R_f), delayed fast runoff (R_{fl}), groundwater recharge (R_r for each relevant HRU and $R_{r,tot}$ combining all relevant HRUs), upwelling groundwater (R_{GW} for each relevant HRU and $R_{GW,tot}$ combining all relevant HRUs), fast runoff (Q_f for each relevant HRU and $Q_{f,tot}$ combining all relevant HRUs), groundwater/slow runoff (Q_s), total runoff (Q_m). Storages [mm]: storage in interception reservoir (S_i), storage in unsaturated root zone (S_u), storage in groundwater/slow reservoir (S_s), storage in fast reservoir (S_f). Parameters: interception capacity (I_{max}) [mm], maximum upwelling groundwater (C_{max}) [mm d⁻¹], maximum root zone storage capacity ($S_{u,max}$) [mm], reference storage in the saturated zone ($S_{s,ref}$) [mm], splitter (W) [-], shape parameter (β) [-], transpiration coefficient (c_e) [-], time lag (T_{lag}) [d], exponent (γ) [-], reservoir time scales [d] of fast (K_f) and slow (K_s) reservoirs, areal weights (p_{HRU}) [-], time step (Δt) [d]. Calibration parameters are shown in bold letters in the table below. The equations were applied to each hydrological response unit (HRU) unless indicated differently.

Reservoir system	Water balance equation	Process functions	
Interception	$\frac{\Delta S_i}{\Delta t} = P - P_e - E_i$ (4.1)	$E_i = \min(E_p, \min(P, \frac{I_{max}}{\Delta t}))$ (4.2)	
		$P_e = P - E_i$ (4.3)	
Unsaturated zone	Flat: $\frac{\Delta S_u}{\Delta t} = P_e - E_t - R_f$ (4.4)	$E_t = \min((E_p - E_i), \min(\frac{S_u}{\Delta t} \cdot (E_p - E_i) \cdot \frac{S_u}{S_{u,max}} \cdot \frac{1}{c_e}))$ (4.7)	
	Sloped: $\frac{\Delta S_u}{\Delta t} = R_u - E_t$ (4.5)	Model A: $R_{GW} = 0$ (4.8)	
	Wetland: $\frac{\Delta S_u}{\Delta t} = P_e - E_t - R_f + R_{GW}$ (4.6)	Model B: $R_{GW} = \min((1 - \frac{S_u}{S_{u,max}}) \cdot C_{max} \cdot \frac{S_s}{p_{HRU}})$ (4.9)	
		Model C,E,F: $R_{GW} = \min((\frac{\min(S_s, S_{s,ref})}{S_{s,ref}}) \cdot C_{max} \cdot \frac{S_s}{p_{HRU}})$ (4.10)	
		Model D: $R_{GW} = \min((\frac{\min(S_s, S_{s,ref})}{S_{s,ref}})^\gamma \cdot C_{max} \cdot \frac{S_s}{p_{HRU}})$ (4.11)	
		if $S_u + R_{GW} \cdot \Delta t > S_{u,max}$: $R_{GW} = \frac{S_{u,max} - S_u}{\Delta t}$ (4.12)	
		Sloped: $R_u = (1 - C) \cdot P_e$ (4.13)	
		$C = 1 - (1 - \frac{S_u}{S_{u,max}})^\beta$ (4.14)	
	Fast runoff	$\frac{\Delta S_f}{\Delta t} = R_{fl} - Q_f$ (4.15)	$Q_f = \frac{S_f}{K_f}$ (4.16)
			Flat/Wetland: $R_f = \frac{\max(0, S_u - S_{u,max})}{\Delta t}$ (4.17)
			$R_{fl} = R_f$ (4.18)
			Sloped: $R_f = (1 - W) \cdot C \cdot P_e$ (4.19)
		$R_{fl} = R_f * f(T_{lag})$ (4.20)	
Groundwater	$\frac{\Delta S_s}{\Delta t} = R_{r,tot} - R_{GW,tot} - Q_s$ (4.21)	$R_r = W \cdot C \cdot P_e$ (4.22)	
		$R_{r,tot} = \sum_{HRU} (p_{HRU} \cdot R_r)$ (4.23)	
		$R_{GW,tot} = \sum_{HRU} (p_{HRU} \cdot R_{GW})$ (4.24)	
		$Q_s = \frac{S_s}{K_s}$ (4.25)	
		$Q_{f,tot} = \sum_{HRU} (p_{HRU} \cdot Q_f)$ (4.27)	
Total runoff	$Q_m = Q_s + Q_{f,tot}$ (4.26)		
Supporting literature	(Euser et al., 2015; Gao et al., 2014a; Gharari et al., 2014)		

Table 4.3: Model parameter and ranges

Landscape class	Parameter	min	max	Unit	Constraint	Comment
Entire basin	K_s	50	200	d		
	C_e	0	1	-		
	$S_{s,ref}$	100	500	-		Only for Models C to F
Flat	I_{max}	0	5	mm d ⁻¹	$I_{max,sloped} > I_{max,flat}$	
	$S_{u,max}$	300	1000	mm		
	K_f	10	12	d		
	W	0.5	0.95	-		
Sloped	I_{max}	0	5	mm d ⁻¹		
	$S_{u,max}$	300	1000	mm	$S_{u,max,sloped} > S_{u,max,flat}$	
	β	0	2	-		
	T_{lag}	1	5	d		
	K_f	10	12	d		
Wetland	W	0.5	0.95	-	$W_{sloped} > W_{flat}$	
	I_{max}	0	5	mm d ⁻¹	$I_{max,wetland} < I_{max,sloped}$	
	$S_{u,max}$	10	500	mm	$S_{u,max,wetland} < S_{u,max,sloped}$	
	K_f	10	12	d		
	C_{max}	0.1	5	mm d ⁻¹		Only for Models B to F Only for Model D
	γ	0.01	0.5	-		
River profile	v	0.01	5.0	m s ⁻¹		

First model adaptation (Models B – D)

As first model adaption, groundwater upwelling (R_{GW}) was added in wetland areas (see Figure 4.2). This upwelling groundwater was made (1) a linear function of the water content in the unsaturated reservoir (Model B, Eq.4.9 in Table 4.2), (2) a linear function of the water content in the slow responding reservoir (Model C, Eq.4.10) and (3) a non-linear function of the water content in the slow responding reservoir (Model D, Eq.4.11). As a result, upwelling water from the saturated zone feeds the unsaturated zone, controlled by the water content in the unsaturated (Model B) or in the saturated zone (Models C and D), and thus increasing the water availability for transpiration from the unsaturated zone in wetland areas. Compared to the benchmark Model A, Model B introduces one additional calibration parameter, Model C two and Model D three (Tables 4.2 and 4.3). See Section 4.5.2 for more detailed information on the reasons for and processes behind these model adjustments.

Second model adaptation (Models E – F)

As second model adaptation, the spatial resolution of the slow responding reservoir was gradually increased from lumped (Models A – D) to semi-distributed (Model E) and fully distributed (Model F). In Model E, the slow responding reservoir was divided into four units as visualised in Figure 4.1A, whereas in Model F it was further discretized into a grid of 10 x 10 km², equivalent to the remaining parts of the model. For both alternative formulations, Models E and F respectively, the slow reservoir timescales K_s remained constant throughout the basin to limit the number of calibration parameters. For both Models E and F, groundwater upwelling was included according to Eq.4.10 (Table 4.2), hence using Model C as basis, introducing

two additional calibration parameters compared to the benchmark Model A (Tables 4.2 and 4.3). See Section 4.5.3 for more detailed information on the reasons for and processes behind these model adjustments.

4.4.2. Model performance metrics

Discharge

The model performance with respect to discharge was evaluated using eight distinct signatures simultaneously characterizing the observed discharge data (Euser et al., 2013; Hulsman et al., 2020a). The model performance measure was based either on the Nash-Sutcliffe efficiency ($E_{NS,\theta}$, Eq.4.29 in Table 4.4) or the relative error ($E_{R,\theta}$, Eq.4.30) depending on the individual signature. The resulting performance metrics for the eight signatures then included the Nash-Sutcliffe efficiencies of the daily discharge time series ($E_{NS,Q}$), its logarithm ($E_{NS,\log Q}$), the flow duration curve ($E_{NS,FDC}$), its logarithm ($E_{NS,\log FDC}$) and of the autocorrelation function of daily flows ($E_{NS,AC}$) and the relative errors of the mean seasonal runoff coefficient during dry and wet periods ($E_{R,RCdry}$, $E_{R,RCwet}$) and of the rising limb density of the hydrograph ($E_{R,RLD}$). All these signatures were combined into an overall performance metric based on the Euclidian distance to the “perfect” model ($D_{E,Qcal}$, Eq.4.32). In absence of more information and to obtain balanced solutions, all individual performance metrics were equally weighted in Eq.4.32. Here, a $D_{E,Qcal} = 1$ indicates a perfect fit.

The discharge data availability was very limited during the validation time period (2012 – 2016). As a result, hydrological years were not fully captured resulting in incomplete information on the hydrologic signatures such as rising limb density or auto correlation function. That is why the overall model performance ($D_{E,Qval}$) was calculated using the signatures $E_{NS,Q}$, $E_{NS,\log Q}$, $E_{NS,FDC}$ and $E_{NS,\log FDC}$ excluding $E_{R,RCdry}$, $E_{R,RCwet}$, $E_{R,RLD}$ and $E_{NS,AC}$. It is therefore important to note that $D_{E,Qcal}$ cannot be meaningfully compared with $D_{E,Qval}$. Instead, following the overall objective of the analysis, $D_{E,Qval}$ of the different alternative model hypothesis were compared to evaluate the differences between the models.

Evaporation and total water storage

The model performance was also evaluated with respect to both the temporal dynamics and the spatial pattern of evaporation and total storage, respectively. For this purpose, satellite-based evaporation data (WaPOR) was used on 10-day time scale, and total water storage anomaly data (GRACE) on monthly time scale.

Temporal variation To quantify the models’ skill to reproduce the temporal dynamics of evaporation and total water storage anomalies, the respective Nash-Sutcliffe efficiencies (Eq.4.29) were used as performance metrics. This performance metric was applied to assess the models’ skill to reproduce the basin-average time series of evaporation and total water storage anomalies, i.e. $E_{NS,Basin,E}$ and $E_{NS,Basin,S}$, respectively. Similarly, the models’ performance to mimic the dynamics of evaporation in all grid cells dominated by the wetland HRU was calculated with the Nash-Sutcliffe efficiency ($E_{NS,Wetland,E}$). Grid cells were considered as wetland

dominated if they were completely covered by wetlands, hence if $p_{\text{HRU}} = 1$ with p_{HRU} the areal weight of wetland areas within that cell. With respect to evaporation, the flux was normalised first with Eq.4.28 to emphasize temporal variations rather than absolute values in an attempt to reduce bias related errors in the observation:

$$E_{\text{normalised}} = \frac{E - E_{\text{min}}}{E_{\text{max}} - E_{\text{min}}} \quad (4.28)$$

Table 4.4: Equations used to calculate the model performance

Name	Objective function	Symbol explanation
Nash-Sutcliffe efficiency	$E_{\text{NS},\theta} = 1 - \frac{\sum_t (\theta_{\text{mod}}(t) - \theta_{\text{obs}}(t))^2}{\sum_t (\theta_{\text{obs}}(t) - \bar{\theta}_{\text{obs}})^2} \quad (4.29)$	θ : variable
Relative error	$E_{\text{R},\theta} = 1 - \frac{ \theta_{\text{mod}} - \theta_{\text{obs}} }{\theta_{\text{obs}}} \quad (4.30)$	θ : variable
Spatial efficiency metric	$E_{\text{SP}} = \frac{1}{t_{\text{max}}} \sum_t (1 - \sqrt{(\alpha - 1)^2 + (\delta - 1)^2 + (\epsilon - 1)^2}) \quad (4.31)$ <p>With:</p> $\alpha = \rho(\phi_{\text{obs}}, \phi_{\text{mod}})$ $\delta = \frac{\sigma_{\text{obs}}/\mu_{\text{obs}}}{\sigma_{\text{mod}}/\mu_{\text{mod}}}$ $\epsilon = \frac{\sum_{i=0}^{i=n} \min(K_i, L_i)}{\sum_{i=0}^{i=n} K_i}$	α : Pearson correlation coefficient $\phi_{\text{obs}}, \phi_{\text{mod}}$: observed/modelled map δ : coefficient of variation σ : standard deviation μ : mean ϵ : fraction of histogram intersection between K and L K : observed histogram L : modelled histogram $n = 100$ bins t time step within the dry season with maximum t_{max}
Euclidian distance over multiple signatures	$D_{\text{E}} = 1 - \sqrt{\frac{1}{N+M} \left(\sum_n (1 - E_{\text{NS},\theta_n})^2 + \sum_m (1 - E_{\text{R},\theta_m})^2 \right)} \quad (4.32)$	θ : signature n : signatures evaluated with Eq.4.29 with maximum N m : signatures evaluated with Eq.4.30 with maximum M
Euclidian distance over multiple variables	$D_{\text{E,ESQ}} = 1 - \sqrt{\frac{1}{n} \sum_n (1 - E_n)^2} \quad (4.33)$	E_n model performance metric of variable n

Spatial variation The model performance with respect to the spatial pattern of evaporation and total water storage anomalies was calculated with the spatial efficiency metrics $E_{SP,E}$ and $E_{SP,S}$ (Eq.4.31), respectively, which was successfully used in previous studies (Demirel et al., 2018; Koch et al., 2018). The spatial model performance was first calculated for each time step within the dry period which was in September/October and then averaged to obtain the overall model performance (E_{SP} , Eq.4.31). The spatial pattern was averaged over the dry season to minimize the effect of precipitation errors.

Multi-variable

The overall potential of the models to simultaneously reproduce the temporal dynamics as well as the spatial pattern of all observed variables, i.e. discharge, evaporation and total water storage anomalies, was tested with the overall model performance metric $D_{E,ESQ}$. This metric was the Euclidian distance (Eq.4.33) of the following individual metrics: the temporal variation of the basin-average evaporation ($E_{NS,Basin,E}$) and total water storage anomalies ($E_{NS,Basin,S}$), spatial pattern of the evaporation ($E_{SP,E}$) and total water storage anomalies ($E_{SP,S}$) as well as discharge ($D_{E,Q}$). See Table 4.5 for an overview of all model performance metrics used in this study.

Table 4.5: Overview of the applied model performance metrics

Data	Temporal dynamics /Spatial pattern	Performance metric	Symbol & equation nr.	Calibration/ validation
Discharge	Temporal dynamics	Euclidian distance over multiple signatures (combining $E_{NS,Q}$, $E_{NS,\log Q}$, $E_{NS,FDC}$, $E_{NS,\log FDC}$, $E_{NS,AC}$, $E_{R,RCdry}$, $E_{R,RCwet}$ and $E_{R,RLD}$)	$D_{E,Qcal}$ (Eq.4.32)	Calibration (2002 – 2012)
	Temporal dynamics	Euclidian distance over multiple signatures (combining $E_{NS,Q}$, $E_{NS,\log Q}$, $E_{NS,FDC}$ and $E_{NS,\log FDC}$)	$D_{E,Qval}$ (Eq.4.32)	Validation (2012 – 2016)
Evaporation	Temporal dynamics (basin-average)	Nash-Sutcliffe efficiency	$E_{NS,Basin,E}$ (Eq.4.29)	Validation (2012 – 2016)
	Temporal dynamics (wetland areas)	Nash-Sutcliffe efficiency	$E_{NS,Wetland,E}$ (Eq.4.29)	Validation (2012 – 2016)
	Spatial pattern	Spatial efficiency metric	$E_{SP,E}$ (Eq.4.31)	Validation (2012 – 2016)
Total water storage anomalies	Temporal dynamics (basin-average)	Nash-Sutcliffe efficiency	$E_{NS,Basin,S}$ (Eq.4.29)	Validation (2012 – 2016)
	Spatial pattern	Spatial efficiency metric	$E_{SP,S}$ (Eq.4.31)	Validation (2012 – 2016)
Multi-variable	Combination	Euclidian distance over multiple variables (combining $D_{E,Qcal}$, $E_{NS,Basin,E}$, $E_{SP,E}$, $E_{NS,Basin,S}$ and $E_{SP,S}$)	$D_{E,ESQcal}$ (Eq.4.33)	Calibration (2002 – 2012)
	Combination	Euclidian distance over multiple variables (combining $D_{E,Qval}$, $E_{NS,Basin,E}$, $E_{SP,E}$, $E_{NS,Basin,S}$ and $E_{SP,S}$)	$D_{E,ESQval}$ (Eq.4.33)	Validation (2012 – 2016)

4.4.3. Model calibration

In general, the model was calibrated by first running the model with $5 \cdot 10^4$ random parameter sets generated with a Monte-Carlo sampling strategy from uniform prior parameter distributions (Table 4.3). Then, the optimal and 5% best-performing parameter sets were selected according to the model performance metric as described in the previous section. The model was calibrated within the time period 2002 – 2012 with respect to 1) discharge ($D_{E,Qcal}$) and 2) all variables simultaneously ($D_{E,ESQcal}$).

4.4.4. Model validation

The model was validated with respect to discharge, evaporation and total water storage anomalies for the time period 2012 – 2016. During validation each variable was evaluated separately both temporally and spatially. This included the temporal variation of the basin-average evaporation ($E_{NS,Basin,E}$) and total water storage anomalies ($E_{NS,Basin,S}$), evaporation in wetland areas ($E_{NS,Wetland,E}$), spatial pattern of the evaporation ($E_{SP,E}$) and total water storage anomalies ($E_{SP,S}$) as well as discharge ($D_{E,Qval}$). In addition, the model was evaluated with respect to the overall performance ($D_{E,ESQval}$). This was done for the solutions from both calibration strategies.

4.5. Model results

4.5.1. Benchmark model (Model A)

Discharge based calibration

For the benchmark model (Model A), the model performance of all model realizations following the first calibration strategy, i.e. calibrating to discharge, resulted in an optimum $D_{E,Qcal,opt} = 0.76$ and $D_{E,Qval} = 0.37$ during validation (Table 4.6, Figure 4.3). As shown in Figure 4.4, the main features of the hydrological response were captured reasonably well. However, particularly in the validation period, low flows were somewhat underestimated. Note that in 2013, the observed high flows were probably underestimated due to failures in the recording which resulted in a truncated top in the hydrograph and flat top in the flow duration curve during the validation time period (Figure 4.4) and which affect the validated model performance values ($D_{E,Qval}$). The range in the calibrated model performance with respect to each discharge signature separately is visualised in Figure B.1 in Appendix B.

The basin-average evaporation ($E_{NS,Basin,E} = 0.54$) and total water storage anomalies ($E_{NS,Basin,S} = 0.74$) were in general also reproduced rather well (Figures B.3 and B.5 in Appendix B). In contrast, the model failed to mimic the evaporation dynamics in wetland dominated areas as it decreased rapidly to zero in the dry season in contrast to the observations ($E_{NS,Wetland,E} = 0.25$, Figure 4.5). Similarly, the spatial variability in evaporation ($E_{SP,E} = 0.17$) and water storage anomalies ($E_{SP,S} = -0.02$) were poorly captured as several areas were over- or underestimated (Figures 4.6 and 4.7). Note that in both figures the normalised evaporation and total water storage anomalies were plotted applying Eq.4.28 to emphasize relative spatial

differences rather than absolute values.

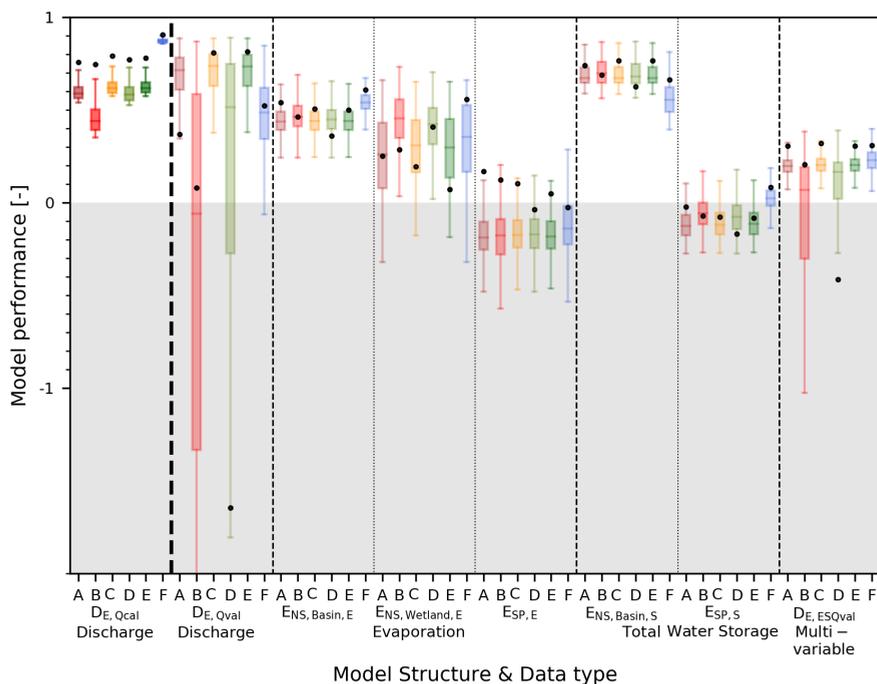


Figure 4.3: Model performance with respect to discharge, evaporation and storage for all models. The model is calibrated to discharge (darker boxplots in the first column) and validated to the discharge, evaporation and storage (lighter boxplots). The dots represent the model performance using the "optimal" parameter set and the boxplot the range of the best 5% solutions both according to discharge ($D_{E,Qcal}$). The following performance metrics were used: 1) discharge using the overall model performance metric ($D_{E,Qcal}$ for calibration and $D_{E,Qval}$ for validation), 2) evaporation temporally basin-average ($E_{NS,Basin,E}$), 3) evaporation temporally wetland areas only ($E_{NS,Wetland,E}$), 4) evaporation spatially ($E_{SP,E}$), 5) storage temporally basin-average ($E_{NS,Basin,S}$), 6) storage spatially ($E_{SP,S}$), and 7) the combination of evaporation, storage and discharge (combined metric $D_{E,ESQval}$).

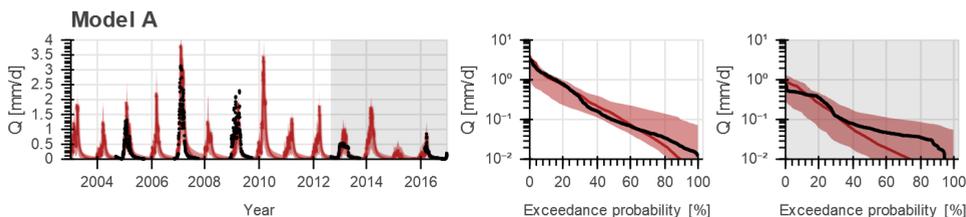


Figure 4.4: Range of model solutions for Model A. The left panel shows the hydrograph and the right panel the flow duration curve in logarithmic scale of the recorded (black) and modelled discharge: the line indicates the solution with the highest calibration objective function with respect to discharge ($D_{E,Qcal}$) and the shaded area the envelope of the solutions retained as feasible. The data in the white area were used for calibration and the grey shaded area for validation.

Table 4.6: Summary of model performance with respect to evaporation ($E_{NS,Basin,E}$, $E_{NS,Wetland,E}$ and $E_{SP,E}$), total water storage anomalies ($E_{NS,Basin,S}$ or $E_{SP,S}$), discharge ($D_{E,Qcal}$ and $D_{E,Qval}$) and all variables combined ($D_{E,ESQval}$): The parameter sets were selected based on discharge ($D_{E,Qcal}$).

Calibration (2002 – 2012)		Validation (2012 – 2016)						
	$D_{E,Qcal,opt}$ ($D_{E,Qcal,5/95}$)	$D_{E,Qval,opt}$ ($D_{E,Qval,5/95}$)	$E_{NS,Basin,E}$ ($E_{NS,Basin,E,5/95}$)	$E_{NS,Wetland,E}$ ($E_{NS,Wetland,E,5/95}$)	$E_{SP,E}$ ($E_{SP,E,5/95}$)	$E_{NS,Basin,S}$ ($E_{NS,Basin,S,5/95}$)	$E_{SP,S}$ ($E_{SP,S,5/95}$)	$D_{E,ESQval,opt}$ ($D_{E,ESQval,5/95}$)
A	0.76 (0.54 – 0.68)	0.37 (0.26 – 0.85)	0.54 (0.34 – 0.57)	0.25 (-0.14 – 0.58)	0.17 (-0.37 – 0.04)	0.74 (0.62 – 0.80)	-0.02 (-0.23 – 0.03)	0.30 (0.12 – 0.29)
B	0.75 (0.36 – 0.60)	0.08 (-3.9 – 0.78)	0.46 (0.34 – 0.63)	0.29 (0.09 – 0.65)	0.12 (-0.68 – 0.12)	0.69 (0.61 – 0.82)	-0.07 (-0.20 – 0.08)	0.21 (-1.3 – 0.27)
C	0.79 (0.58 – 0.70)	0.81 (0.27 – 0.85)	0.50 (0.34 – 0.58)	0.19 (-0.01 – 0.57)	0.10 (-0.39 – 0.06)	0.76 (0.62 – 0.81)	-0.08 (-0.23 – 0.04)	0.32 (0.12 – 0.30)
D	0.77 (0.53 – 0.68)	-1.7 (-2.4 – 0.84)	0.36 (0.33 – 0.60)	0.41 (0.11 – 0.62)	-0.04 (-0.57 – 0.10)	0.63 (0.61 – 0.82)	-0.17 (-0.22 – 0.06)	-0.41 (-0.72 – 0.28)
E	0.78 (0.58 – 0.70)	0.81 (0.27 – 0.85)	0.50 (0.34 – 0.58)	0.07 (-0.04 – 0.59)	0.05 (-0.39 – 0.05)	0.77 (0.62 – 0.81)	-0.08 (-0.23 – 0.04)	0.30 (0.12 – 0.29)
F	0.91 (0.86 – 0.89)	0.52 (0.12 – 0.74)	0.61 (0.45 – 0.63)	0.56 (-0.08 – 0.61)	-0.03 (-0.49 – 0.19)	0.66 (0.44 – 0.71)	0.08 (-0.07 – 0.13)	0.31 (0.12 – 0.34)

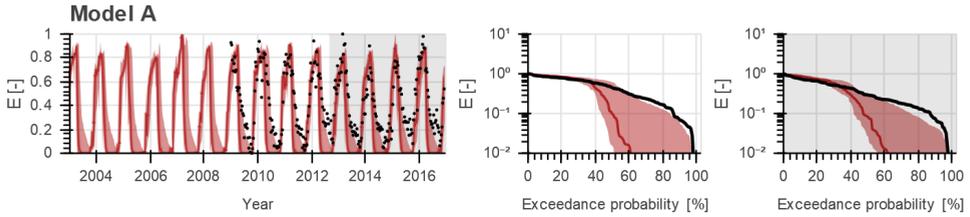


Figure 4.5: Range of model solutions for Models A to F. The left panel shows the time series and the right panel the duration curve in logarithmic scale of the recorded (black) and modelled normalised evaporation for wetland dominated areas: the line indicates the solution with the highest calibration objective function with respect to discharge ($D_{E,Q_{cal}}$) and the shaded area the envelope of the solutions retained as feasible. The data in the grey shaded area were used for validation.

4

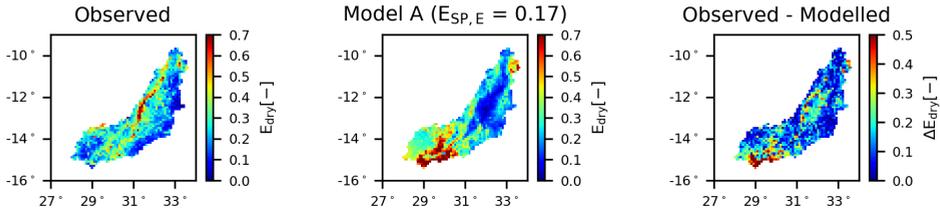


Figure 4.6: Spatial variability of the normalised total evaporation for Model A averaged over all days within the dry season. The left panel shows the observation according to WaPOR data, the middle panel the model result using the “optimal” parameter set with respect to discharge ($D_{E,Q_{cal}}$), and the right panel the difference between the observation and model.

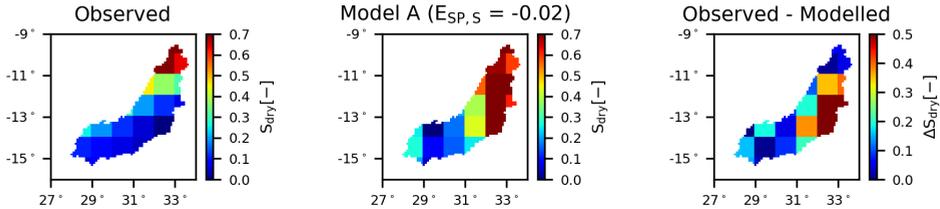


Figure 4.7: Spatial variability of the normalised total water storage anomalies for Model A averaged over all days within the dry season. The left panel shows the observation according to GRACE data, the middle panel the model result using the “optimal” parameter set with respect to discharge ($D_{E,Q_{cal}}$), and the right panel the difference between the observation and model.

Multi-variable calibration

Calibrating with respect to multiple variables simultaneously in the second calibration strategy, resulted in a reduced model skill to simultaneously reproduce all flow signatures in the validation period with $D_{E,Q_{val}} = 0.07$ (Table 4.7, Figures 4.8 and 4.9). Compared to the first calibration strategy, the simulated evaporation did not change significantly with respect to the temporal dynamics ($E_{NS,Wetland,E} = 0.27$, $E_{NS,Basin,E} = 0.57$) and spatial pattern ($E_{SP,E} = -0.18$). Evaporation from wetland dominated areas remained underestimated in the dry season (Figure 4.10) and large areas in the basin were still under- or overestimated (Figure 4.11). The reproduction of the total water storage anomalies decreased though, mostly with respect to the spatial pattern ($E_{SP,S} = -0.14$, Figure 4.12). On the other hand, when looking

at the 5/95th percentile range instead of the “optimal” parameter set, then an improvement was observed in the spatial pattern in evaporation ($E_{SP,E,5/95} = -0.10 - 0.22$) and in total water storage ($E_{SP,S,5/95} = -0.17 - 0.08$, compare Tables 4.6 and 4.7).

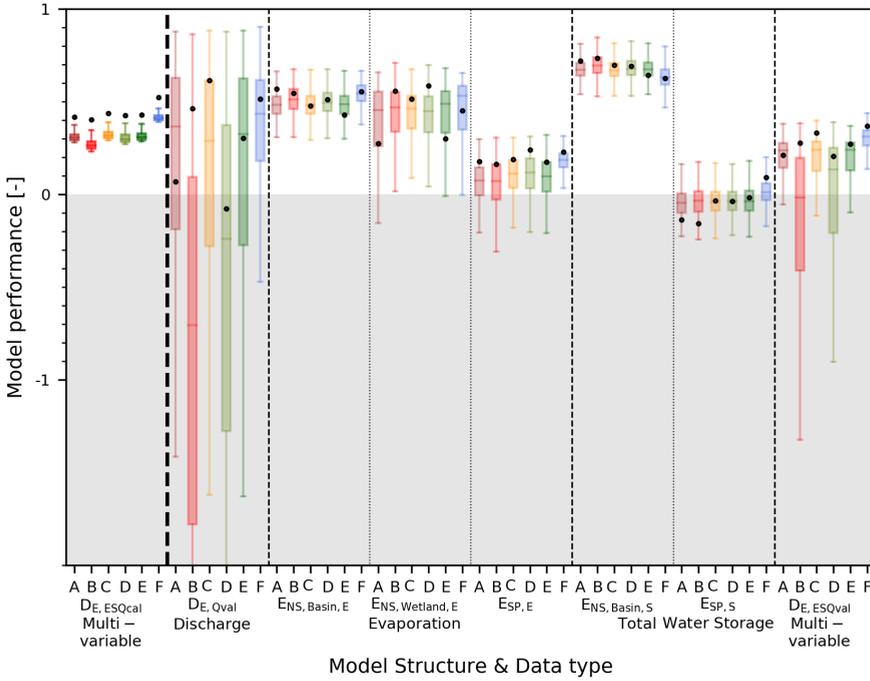


Figure 4.8: Model performance with respect to discharge, evaporation and storage for all models. The model is calibrated to all fluxes simultaneously ($D_{E,ESQcal}$, darker boxplots in the first column) and evaluated with respect to each flux individually (lighter boxplots). The dots represent the model performance using the “optimal” parameter set and the boxplot the range of the best 5% solutions both according to $D_{E,ESQcal}$. The following performance metrics were used: 1) discharge using the overall model performance metric ($D_{E,Qval}$), 2) evaporation temporally basin-average ($E_{NS,Basin,E}$), 3) evaporation temporally wetland areas only ($E_{NS,Wetland,E}$), 4) evaporation spatially ($E_{SP,E}$), 5) storage temporally basin-average ($E_{NS,Basin,S}$), 6) storage spatially ($E_{SP,S}$), and 7) the combination of evaporation, storage and discharge (combined metric $D_{E,ESQval}$).

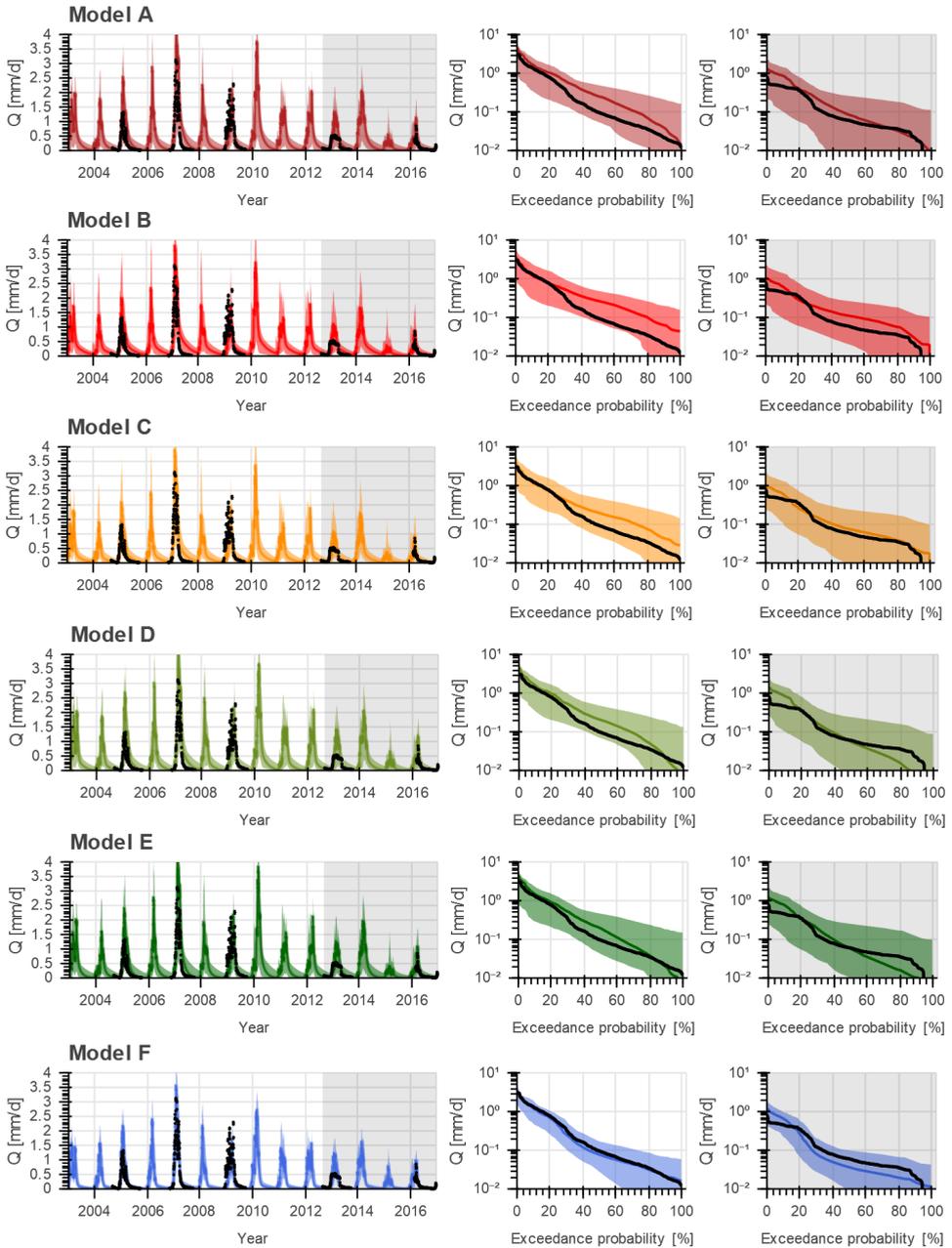


Figure 4.9: Range of model solutions for Models A to F. The left panel shows the hydrograph and the right panel the flow duration curve in logarithmic scale of the recorded (black) and modelled discharge: the line indicates the solution with the highest calibration objective function with respect to multiple variables ($D_{E,ESQ_{cal}}$) and the shaded area the envelope of the solutions retained as feasible. The data in the grey shaded area were used for validation.

Table 4.7: Summary of model performance with respect to evaporation ($E_{NS,Basin,E}$, $E_{NS,Wetland,E}$ and $E_{SP,E}$), total water storage anomalies ($E_{NS,Basin,S}$ or $E_{SP,S}$), discharge ($D_{E,Qcal}$ and $D_{E,Qval}$) and all variables combined ($D_{E,ESQval}$): Parameter sets selected based on multiple variables simultaneously ($D_{E,ESQcal}$).

	Calibration (2002 – 2012)			Validation (2012 – 2016)					
	$D_{E,ESQcal,opt}$ ($D_{E,ESQcal,5/95}$)	$D_{E,Qval,opt}$ ($D_{E,Qval,5/95}$)	$E_{NS,Basin,E}$ ($E_{NS,Basin,E,5/95}$)	$E_{NS,Wetland,E}$ ($E_{NS,Wetland,E,5/95}$)	$F_{SP,E}$ ($F_{SP,E,5/95}$)	$E_{NS,Basin,S}$ ($E_{NS,Basin,S,5/95}$)	$F_{SP,S}$ ($F_{SP,S,5/95}$)	$D_{E,ESQval,opt}$ ($D_{E,ESQval,5/95}$)	
A	0.42 (0.28 – 0.36)	0.07 (-1.4 – 0.80)	0.57 (0.37 – 0.60)	0.27 (-0.05 – 0.61)	0.18 (-0.10 – 0.22)	0.72 (0.60 – 0.77)	-0.14 (-0.17 – 0.08)	0.21 (-0.25 – 0.32)	
B	0.40 (0.23 – 0.33)	0.46 (-4.2 – 0.70)	0.55 (0.39 – 0.63)	0.56 (0.04 – 0.64)	0.16 (-0.14 – 0.25)	0.73 (0.61 – 0.79)	-0.16 (-0.17 – 0.09)	0.28 (-1.4 – 0.29)	
C	0.44 (0.29 – 0.37)	0.61 (-1.6 – 0.79)	0.48 (0.37 – 0.61)	0.51 (0.08 – 0.60)	0.19 (-0.07 – 0.25)	0.70 (0.60 – 0.77)	-0.03 (-0.16 – 0.09)	0.33 (-0.31 – 0.33)	
D	0.43 (0.27 – 0.36)	-0.08 (-3.5 – 0.75)	0.51 (0.38 – 0.62)	0.59 (0.06 – 0.61)	0.24 (-0.09 – 0.26)	0.69 (0.60 – 0.78)	-0.04 (-0.16 – 0.09)	0.21 (-1.1 – 0.32)	
E	0.43 (0.29 – 0.36)	0.30 (-1.6 – 0.79)	0.43 (0.38 – 0.61)	0.30 (0.03 – 0.61)	0.17 (-0.08 – 0.25)	0.64 (0.60 – 0.77)	-0.02 (-0.16 – 0.10)	0.27 (-0.31 – 0.32)	
F	0.52 (0.39 – 0.45)	0.51 (-0.24 – 0.81)	0.56 (0.45 – 0.63)	0.45 (0.01 – 0.63)	0.23 (0.08 – 0.27)	0.63 (0.53 – 0.73)	0.09 (-0.10 – 0.13)	0.37 (0.15 – 0.38)	

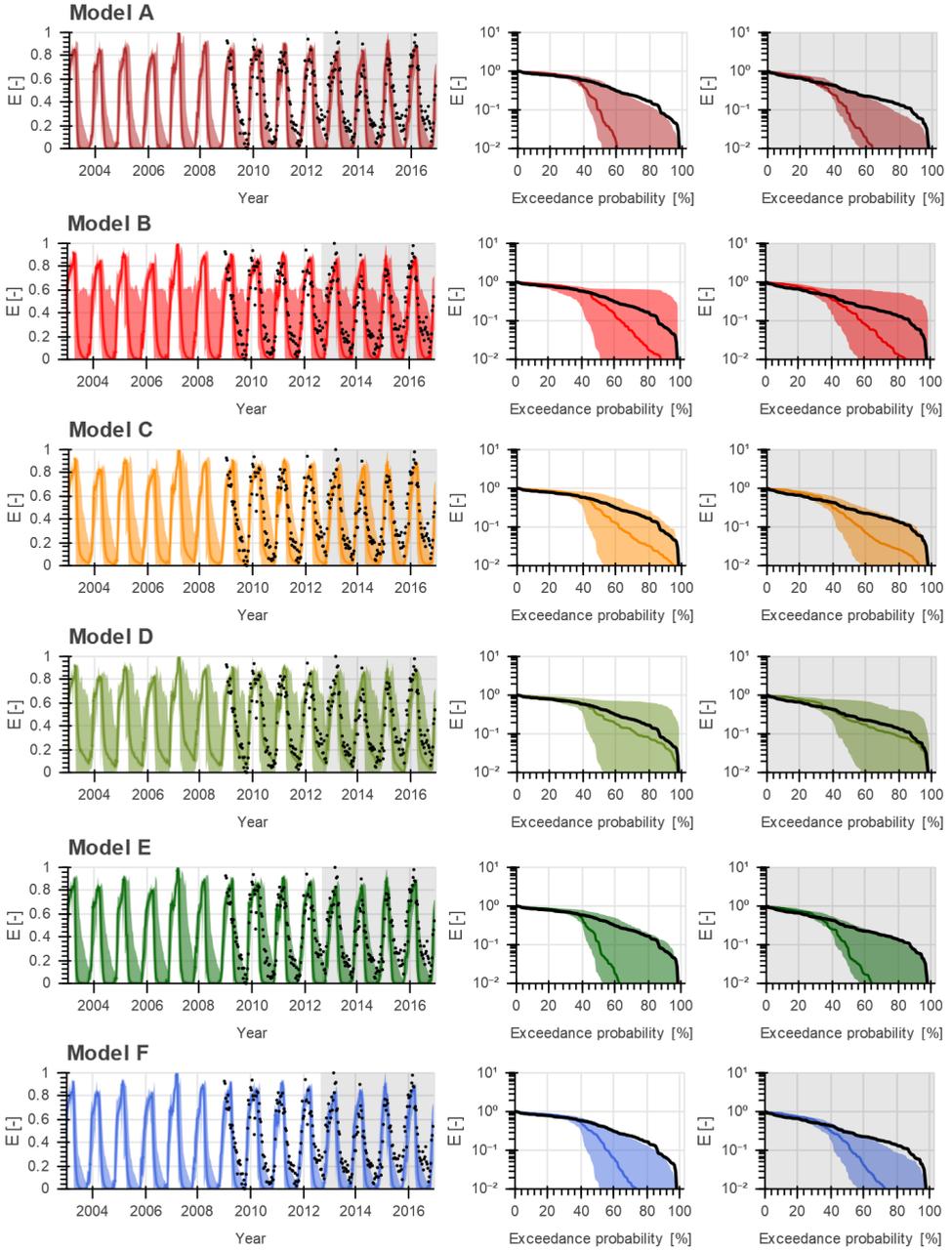


Figure 4.10: Range of model solutions for Models A to F. The left panel shows the time series and the right panel the duration curve in logarithmic scale of the recorded (black) and modelled normalised evaporation for wetland dominated areas: the line indicates the solution with the highest calibration objective function with respect to multiple variables ($D_{E,ESQ_{cal}}$) and the shaded area the envelope of the solutions retained as feasible. The data in the grey shaded area were used for validation.

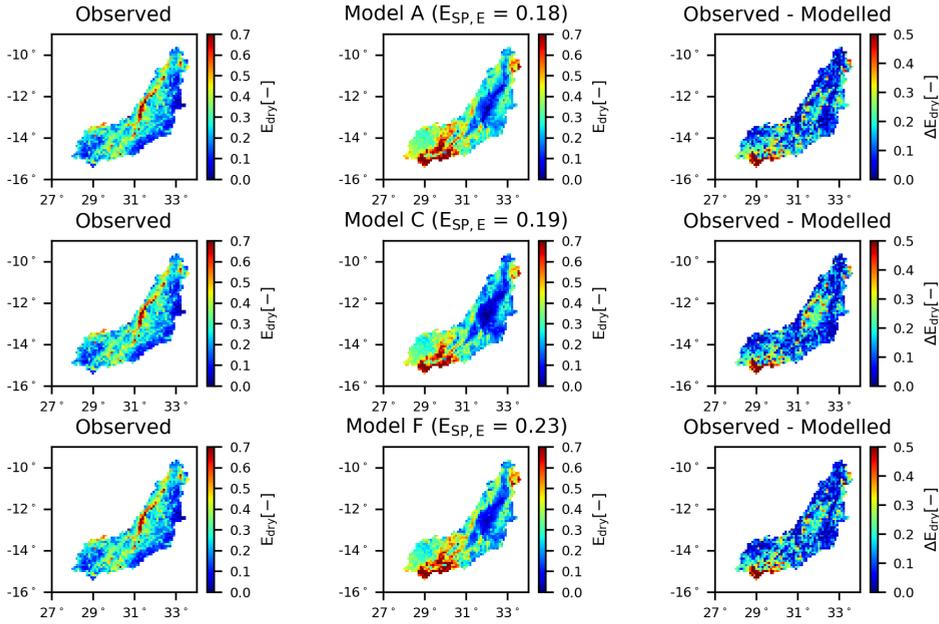


Figure 4.11: Spatial variability of the normalised total evaporation for Models A, C and F averaged over all days within the dry season. The left panel shows the observation according to WaPOR data, the middle panel the model result using the “optimal” parameter set with respect to multiple variables ($D_{E,ESQcal}$), and the right panel the difference between the observation and model.

Model deficiencies

Regardless of the calibration strategy, the benchmark model failed in particular to adequately reproduce evaporation dynamics in wetland dominated areas. During the dry seasons, the modelled evaporation decreased rapidly to zero in contrast to the observations (Figures 4.5 and 4.10). Partly as a consequence of that, the spatial pattern of evaporation was captured poorly as illustrated in Figures 4.6 and 4.11. Apart from the wetlands, the modelled average dry season evaporation was also extremely low in the centre of the basin which did not correspond with the satellite observations. At the same time, the evaporation was significantly overestimated in the southern part of the basin. Also the spatial pattern in total water storage anomalies were poorly represented since the model significantly overestimated storage anomalies in large parts of the basin (Figures 4.7 and 4.12).

4.5.2. Adding groundwater upwelling (Models B, C and D)

In the benchmark model (Model A), there was no groundwater upwelling into the wetlands and floodplains around the river channels, similar to many distributed conceptual hydrological models (Bieger et al., 2017; Samaniego et al., 2010). However, according to field and satellite-based observations, wetland areas remain moist at the end of the dry season while the remaining areas of the basin become very dry.

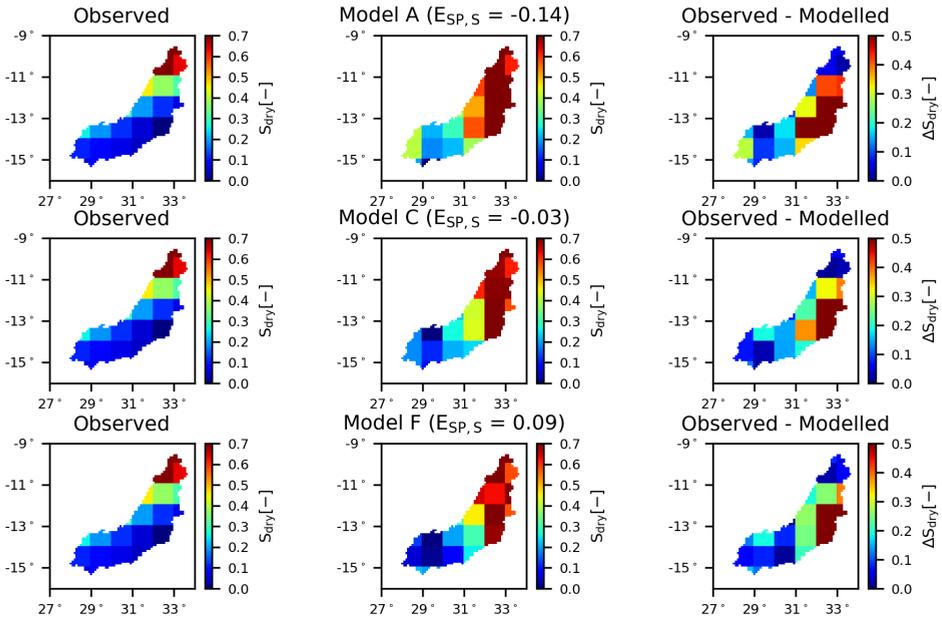


Figure 4.12: Spatial variability of the normalised total water storage anomalies for Models A, C and F averaged over all days within the dry season. The left panel shows the observation according to GRACE data, the middle panel the model result using the “optimal” parameter set with respect to multiple variables ($D_{E,ESQcal}$), and the right panel the difference between the observation and model.

Given the low elevation of these wetlands above rivers, it is plausible to assume that groundwater from higher parts of the catchment is pushed up into the unsaturated root zone of these wetlands. As a result, water deficits in the unsaturated zone are partly replenished by upwelling groundwater. It thereby can sustain relatively elevated levels of moisture, available for plant transpiration long into the dry season. To improve the representation of evaporation in the model, the process of upwelling groundwater (R_{GW}) was added to the model. In principle, it was assumed that the upwelling groundwater is regulated by the head difference between upland groundwater and the groundwater in the wetland. As this information was not available, due to the lack of continuous gradients in the type of model used (Hrachowitz and Clark, 2017), this was done in a simplified way. In three alternative formulations of this hypothesis, the upwelling groundwater was made (1) a linear function of the water content in the unsaturated reservoir (Model B, Eq.4.9), (2) a linear function of the water content in the slow responding reservoir (Model C, Eq.4.10) and (3) a non-linear function of the water content in the slow responding reservoir (Model D, Eq.4.11). In other words, in Model B the groundwater upwelling was driven by the water deficit in the unsaturated zone, hence the lower the water content in the unsaturated zone, the higher the groundwater upwelling. In Models C and D, the groundwater upwelling was driven by the water content in the slow responding reservoir, the groundwater system, such that the higher the water content in the

slow responding reservoir, the higher the groundwater upwelling. As a result of the non-linear relation between the groundwater upwelling and the water content in the slow responding reservoir in Model D, the groundwater upwelling increased the most under dry conditions and less under wet conditions. In Models B – D, the groundwater upwelling flowed into the unsaturated zone until it was saturated, hence until its maximum $S_{u,max}$ was reached (Eq.4.12). Model B required one additional calibration parameter, Model C two and Model D three (Tables 4.2 and 4.3).

Discharge based calibration

Following the first calibration strategy, the performances of Models B – D with respect to discharge did not improve significantly for the calibration period ($D_{E,Qcal} = 0.75 - 0.79$) compared to Model A, regardless of the model (Table 4.6, Figures 4.3 and B.2 in Appendix B). For the validation period, Models B and D experienced a pronounced reduction of their ability to adequately reproduce the discharge signatures with $D_{E,Qval} = 0.08$ and -1.7 , respectively, since the flows were mostly underestimated (Figure B.2 in Appendix B). On the other hand, Model C showed significant improvements with $D_{E,Qval} = 0.81$. With respect to the evaporation from wetland dominated areas, the largest improvements were found for Model D ($E_{NS,Wetland,E} = 0.41$) where the evaporation did not drop rapidly to zero anymore even though it was still significantly underestimated in the dry season (Figure B.4 in Appendix B). But this came at the cost of decreased simulations of all remaining variables (Table 4.6, Figure 4.3), hence the discharge, basin-average evaporation and total water storage and their spatial pattern (Figures B.2 – B.7 in Appendix B). For example Figure B.6 in Appendix B illustrates the poorly simulated temporally-averaged dry season evaporation for Model D which was higher in wetland areas (centre of the basin) compared to the surrounding areas which was not observed in the satellite based observations. For Models B and C, the model performances with respect to the remaining variables remained comparable to Model A or even decreased as can be seen in Table 4.6 and Figure 4.3. As a result, when considering all variables simultaneously, Model C performed the best with $D_{E,ESQval} = 0.32$.

Multi-variable calibration

Following the second calibration strategy, Model C experienced the largest increases compared to Model A in its ability to describe features of discharge with $D_{E,Qval} = 0.61$, while Model D decreased the most to $D_{E,Qval} = -0.08$ with the high flows being overestimated and low flows underestimated (Table 4.7, Figures 4.8 and 4.9). With this calibration strategy, large improvements were observed in the reproduction of the evaporation from wetland dominated areas for all three Models B – D, especially for Model D with $E_{NS,Wetland,E} = 0.59$ where the evaporation was simulated well even during the dry season as it did not decrease rapidly to zero in the dry season compared to Model A (Figure 4.10). For Models C and D, the spatial pattern in evaporation and total water storage anomalies improved, albeit moderately (Table 4.7) as large areas were still under- or overestimated (Figures B.10 and B.11 in Appendix B), whereas it decreased slightly for Model B. For all Models B – D, the basin-average temporal dynamics in evaporation and total water storage anomalies remained similar or decreased slightly (Table 4.7, Figures B.8 and B.9 in Appendix

B). Overall, when considering the model performance with respect to all variables simultaneously, Model C showed the highest performances with $D_{E,ESQval} = 0.33$.

Model deficiencies

According to the results, the representation of evaporation strongly benefitted from including upwelling groundwater as function of the water content in the slow responding reservoir (Eq.4.10, Model C) especially for the second calibration strategy. The incorporation of this flux resulted in increased levels of water supply to the unsaturated zone of wetlands to sustain higher levels of transpiration throughout the dry periods (Figure 4.10). But even though the evaporation increased during dry periods, it was still underestimated especially towards the end of the dry season due to too large groundwater upwelling depleting the slow responding reservoir. The major weakness of the model remained its very limited ability to represent the spatial pattern in evaporation as there were several local clusters of considerable mismatches, both over- and underestimating observed evaporation. This was clearly visible for example in the centre and southern part of the basin (Figure 4.11). Also the spatial pattern in the total water storage anomalies remained poorly represented, in spite of some improvements compared to Model A, as they were considerably overestimated in the northern parts of the basin (Figure 4.12). This could be a result of deficiencies in the hydrological models or in the satellite-based observations.

4

4.5.3. Discretizing the groundwater system (Models E and F)

In all above models, the groundwater layer was simulated as a single lumped reservoir assuming equal groundwater availability throughout the entire basin. As groundwater processes can occur on relatively large spatial scales, this assumption may be valid for small- or mesoscale catchments, but not necessarily for larger basins such as the Luangwa basin. This may partly be responsible for the deficiency of all above models to meaningfully reproduce the spatial pattern of the total water storage. Taking Model C as a basis for further model adaptations, two more alternative model hypothesis were formulated. In both models the slow responding reservoir, representing the groundwater, was spatially discretized. For Model E, the reservoir was split into four units with an area of 15,396 – 47,239 km² each containing four to six different GRACE cells (see Figure 4.1A). In contrast, Model F was formulated with a completely distributed slow reservoir at the resolution of the remaining parts of the model, i.e. 10 x 10 km². In Models E and F, the slow reservoir timescales K_s remained constant throughout the basin to limit the number of calibration parameters. Models E and F did not require additional calibration parameters. See Tables 4.2 and 4.3 for the corresponding model equations and calibration parameter ranges.

Discharge based calibration

Following the first calibration strategy, the calibrated and validated model performance with respect to discharge did not change significantly for Model E compared to Model C. For Model F on the other hand, the calibrated model performance increased to $D_{E,Qcal} = 0.91$ (Table 4.6, Figures 4.3 and B.2 in Appendix B), but during

validation it decreased to $D_{E,Qval} = 0.52$ compared to Model C as a result of overestimated high flows (Figure B.2 in Appendix B). In other words, the discharge simulation was only affected when applying a fully distributed groundwater system (Model F). Also the simulated dynamics of the evaporation improved for Model F, especially for wetland dominated areas ($E_{NS,Wetland,E} = 0.56$, Table 4.6) even though it remained significantly underestimated during the dry season (Figure B.4 in Appendix B). But for both models, no improvements in the spatial pattern of evaporation can be observed with $E_{SP,E} = 0.05$ and -0.03 for Models E and F, respectively. As shown in Figure B.6 in Appendix B, for Model E and F the temporally-averaged dry season evaporation was very low in the centre of the basin compared to the remaining part of the basin in contrast to the satellite-based observations. The spatial pattern of total water storage anomalies were at least slightly better mimicked by Model F with $E_{SP,S} = 0.08$ (Figure B.7 in Appendix B), which, in turn, came at the price of a poorer reproduction of the temporal dynamics of the basin-averaged total water storage anomalies ($E_{NS,Basin,S} = 0.66$, Figure B.5 in Appendix B).

Multi-variable calibration

Including multiple variables in the calibration process did not improve the representation of the hydrological response with respect to discharge for Models E and F compared to Model C with $D_{E,Qval} = 0.30$ and 0.51 , respectively (Table 4.7, Figures 4.8 and 4.9). For both models, the flows were underestimated during low flows and overestimated during high flows (Figure 4.9). Also the evaporation from wetland dominated areas did not improve for both models as it decreased rapidly in the dry season (Figure 4.10). On the other hand, the spatial pattern in the evaporation was slightly better mimicked for Model F ($E_{SP,E} = 0.23$), but still at low performance levels similar to Models A – D with large areas still being under- or overestimated (Figure B.10 in Appendix B). Slight improvements could be observed though for the representation of spatial pattern in total water storage in Models F ($E_{SP,S} = 0.09$, Figure B.11 in Appendix B), albeit modestly. Overall, when considering the model performance with respect to all variables simultaneously, Model F showed the highest performances with $D_{E,ESQval} = 0.37$.

Model deficiencies

Applying the second calibration strategy, Model F poorly reproduced the evaporation from wetlands (Figure 4.10) since the water availability for evaporation decreased rapidly in the dry season due to the limited water availability in the slow responding reservoir. This was a direct result of the limited connectivity in the distributed groundwater system within the basin and very likely points to the presence of contiguous groundwater systems extending beyond the modelling resolution that sustain dry season evaporation in wetlands. Strikingly, discretizing the groundwater basin only had limited effects on the spatial pattern in evaporation and total water storage anomalies. Despite their limited improvements, they remained poorly captured as several local clusters were over- and underestimated (Figures 4.11 and 4.12).

4.6. Discussion

As illustrated in the previous sections, satellite-based evaporation and storage anomaly data were used in an attempt to (1) iteratively improve a benchmark model structure and (2) identify parameter sets with which the model can simultaneously reproduce the temporal dynamics as well as the spatial pattern of multiple flux and storage variables.

The results suggested that among the tested models, Models C and F provided the overall best representation of the hydrological processes in the Luangwa basin, following the first and second calibration strategy respectively. The addition of upwelling groundwater alone (Model C) significantly improved the discharge simulations during validation regardless of the calibration strategy and the simulation of evaporation from wetland areas following the second calibration strategy. Discretizing the slow responding reservoir (Model F) reached reasonable overall performance levels, i.e. $D_{E,ESQval}$ when calibrating on discharge and its signatures only (Figure 4.3), with improved simulations of evaporation from wetland areas. But calibrating on multiple variables proved instrumental as it allowed to significantly improve the spatial pattern of the evaporation, while maintaining high levels for the other performance criteria (Figure 4.8). In general it could also be observed that a further discretization of the model lead to a better representation of the system especially with respect to the spatial pattern. Nevertheless, none of the tested models could adequately reproduce the observed spatial pattern in evaporation and total water storage anomalies which could be a result of model deficiencies or uncertainties in the satellite-based observations of the spatial pattern.

A potential reason for the models' problems to meaningfully describe the spatial pattern of the evaporation was in this study the use of the same parameters within a specific HRU in different model grid cells as also observed in previous studies (Stisen et al., 2018). As a result, the simulated spatial pattern was strongly influenced by the catchment classification method into distinct HRUs. In this study, the catchment was classified merely on the basis of topography into flat, sloped and wetland areas, whereas ecosystem diversity could also be considered as an additional layer in the classification. The poor representation of the spatial pattern in total water storage was also partly linked to that. Another likely reason is the absence of lateral exchange of sub-surface water between model grid cells in the tested models, as contiguous groundwater bodies of varying but unknown spatial scale will shape water transfer through the landscape in the real world which remain unaccounted for in the model.

In addition, each of the applied data sources have their own uncertainties and bias. These include uncertainties in observed discharge due to rating curve uncertainties (Domeneghetti et al., 2012; Tomkins, 2014; Westerberg et al., 2011) and limited data availability, in precipitation data, often as a result of poorly capturing mountainous regions or extreme events on small scales (Dinku et al., 2018; Hrachowitz and Weiler, 2011; Kimani et al., 2017; Le Coz and van de Giesen, 2020), in estimates of total water storage anomalies as a result of data (post-) processing including data smoothing using a radius of for example 300 km affecting the spatial variability on basin scale (Blazquez et al., 2018b; Landerer and Swenson, 2012) and

in evaporation data due to model, input data and parameter estimation uncertainties (Zhang et al., 2016). In general satellite products are a result of models that are prone to uncertainties related to the input data or model conceptualisation. In the ideal situation, the data would be validated with field measurements to assess the error magnitude. However, this was not possible due to data limitations.

The results in this study were sensitive to the choice of performance metrics with respect to the individual variables (discharge, evaporation and total water storage) and all variables combined. For instance the overall model performance measure $D_{E,ESQval}$ (Eq.4.33) was strongly influenced by the validated discharge model performance $D_{E,Qval}$ due to its large range and variation between models compared to the remaining variables where the range was smaller and similar for all models (Figure 4.8). As a result, the overall model performance measure might not reflect each variable equally well which affected the choice of best performing model.

Reflecting the results of previous studies, this study found that calibrating to multiple variables including the spatial pattern improved the simulation of the evaporation and storage with some trade-off in the discharge simulation depending on the model structure (Dembélé et al., 2020; Demirel et al., 2018; Herman et al., 2018; Rientjes et al., 2013; Stisen et al., 2011, 2018). But in contrast and additional to previous studies, this study also provided an example, illustrating that spatial data, here evaporation and total water storage, can contain relevant information to diagnose model deficiencies and to therefore enable step-wise model structural improvement. Previous studies have largely relied on discharge observations to improve model structures (Fencia et al., 2016; Hrachowitz et al., 2014) and only few studies used satellite data (Roy et al., 2017) even though it provides valuable information on the internal processes temporally and spatially which is not available with discharge data alone (Daggupati et al., 2015; Rakovec et al., 2016). Roy et al. (2017) observed that the simulated evaporation according to the spatially lumped model HYMOD (HYdrological MODel) rapidly dropped to zero in contrast to the satellite product GLEAM (Global Land Evaporation Amsterdam Model) in the Nyangores river basin in Kenya. They improved this simulated evaporation while maintaining good discharge performances by modifying the corresponding equation in HYMOD such that it was a function of the soil moisture. While here we focussed on upwelling groundwater and spatial discretization, a promising avenue for future studies may be to evaluate the incorporation of simple formulations of subsurface exchange fluxes between model grid cells. Similarly, a further discretization of HRUs into different land cover and ecosystem types may be worthwhile.

4.7. Evaluation

The objective of this study was to explore the added value of satellite-based evaporation and total water storage anomaly data to increase the understanding of hydrological processes through step-wise model structure improvement and model calibration for large river systems in a semi-arid, data scarce region. For this purpose, a distributed process-based hydrological model with sub-grid process heterogeneity for the Luangwa River basin was developed and iteratively adjusted. The results suggested that (1) the benchmark model (Model A) calibrated with respect

to discharge simulated the discharge well, and also the basin-average evaporation and total water storage anomalies, but poorly captured the evaporation for wetland dominated areas and the spatial pattern of evaporation and total water storage anomalies. (2) Testing five further alternative model structures (Models B – F), it was found that among the tested model hypotheses Model F, allowing for upwelling groundwater from a distributed representation of the groundwater reservoir and (3) simultaneously calibrating the model with respect to multiple variables, i.e. discharge, evaporation and total water storage anomalies, provided the best representation of all these variables with respect to their temporal dynamics and spatial pattern, except for the basin-average temporal dynamics in the total water storage anomalies. It was shown that satellite-based evaporation and total water storage anomaly data are not only valuable for multi-criteria calibration, but can play an important role in improving our understanding of hydrological processes through diagnosing model deficiencies and step-wise model structural improvement.

5

Long-term total water storage variations in the Luangwa basin

*A little rain each day
will fill the rivers to overflowing.*

African Proverb

In the Luangwa basin, long-term total water storage variations were observed with GRACE, but not reproduced by a standard conceptual hydrological model that encapsulates our current understanding of the dominant regional hydrological processes. The objective of this study was to identify potential processes underlying these low-frequency variations through combined data analysis and model hypothesis testing. First, we analysed the effect of data uncertainty to explain observed storage variations. Second, we evaluated the model's skill to reproduce the observed long-term storage variations using four different combinations of model forcing. Third, we formulated alternative model hypotheses to account for groundwater export as data analysis indicated this could potentially explain low-frequency storage variations.

This chapter is based on: Hulsman, P., Savenije, H. H. G., and Hrachowitz, M.: Why are long-term storage variations observed but not modelled in the Luangwa basin?, Water Resources Research, in review.

5.1. Introduction

Long-term and thus low-frequency total water storage variations have been observed in many regions world-wide (Long et al., 2017; Scanlon et al., 2018). This includes long-term storage variations in Australia during the Millennium Drought in 1997 – 2010 (e.g. Chen et al., 2016; Leblanc et al., 2009; Zhao et al., 2017b), in the United States (Boutt, 2017; Long et al., 2013), in the La Plata basin in South America (Chen et al., 2010), in China (Sun et al., 2018b; Zhang et al., 2015b) and in different African river basins (Awange et al., 2016; Bonsor et al., 2018; Werth et al., 2017).

However, many hydrological models cannot reproduce these observed long-term storage variations (Fowler et al., 2020; Scanlon et al., 2018; Winsemius et al., 2006b). As highlighted by previous studies, these observed long-term storage variations can be a result of climate variability, land-cover change, other human interventions or any combination thereof, while the inability of models to reproduce these variations can be a result of model structural deficiencies, poor parameterization, data errors, poor parameter values or any combination thereof (Fowler et al., 2018; Grigg and Hughes, 2018; Jing et al., 2019; Saft et al., 2016). For example, Bouaziz et al. (2020) showed that although a suite of different conceptual models could similarly well reproduce stream flow over almost two decades, they considerably varied in their skill to reproduce observed storage variations, which was attributed to deficiencies of different model architectures. With some exceptions (e.g. Bouaziz et al., 2018; Goswami et al., 2007; Hrachowitz et al., 2014; Le Moine et al., 2007; Perrin et al., 2003; Samaniego et al., 2011), processes that could potentially allow long-term memory effects, such as groundwater export, remain mostly unaccounted for in standard conceptual rainfall-runoff models (Bergström, 1992; Burnash et al., 1973; Euser et al., 2015; Fenicia et al., 2014; Liang et al., 1994; Willems, 2014). This leads to the situation that these models cannot capture long and slow processes dominating long-term storage variations, as convincingly demonstrated by Fowler et al. (2020). Their study, which focused on the Millennium Drought in Australia, illustrated that modelled annual minimum storage remained rather constant instead of showing a decreasing trend. The reason for this was that the modelled storage converged to or even reached zero towards the end of each dry season and hence could not decrease any further. Such an omission of processes that allow to account for long-term memory processes in rainfall-runoff models results in biased modelled discharge and impedes accurate estimations of water availability which is particularly crucial during extreme dry conditions (Saft et al., 2016).

In many river basins, detecting long-term storage variations and identifying their drivers is challenged by limited high-quality ground observations. That is why in this context satellite observations may play an important role. For example, satellite-based Gravity Recovery and Climate Experiment (GRACE) observations describe variations in the Earths' gravity field which can be used to detect regional mass changes that are dominated by variations in the terrestrial water storage after removing atmospheric effects. In other words, GRACE observations, which are available on monthly timescale, provide valuable information on total water

storage changes (Landerer and Swenson, 2012; Swenson, 2012). For example, GRACE observations have been used in the context of groundwater monitoring (Tangdamrongsab et al., 2018; Zhang et al., 2020), or drought analysis (Chao et al., 2016; Leblanc et al., 2009; van Dijk et al., 2013; Zhang et al., 2015a; Zhao et al., 2017a).

While several previous studies focused on identifying long-term storage variations in (satellite-based) observations, possible drivers for these variations, and differences between observations and model results (e.g. Fowler et al., 2020; Jing et al., 2019; Joodaki et al., 2014; Leblanc et al., 2009; Meng et al., 2019; Scanlon et al., 2018), only limited studies attempted to modify a hydrological model to allow for long-term storage variations. In one exception, Grigg and Hughes (2018) modified the GR4J rainfall-runoff model (Perrin et al., 2003) successfully to mimic long-term catchment memory effects. This was done by introducing a threshold in the storage reservoir such that percolation from this reservoir stopped when the storage was lower than the threshold while evaporation losses continued. Other studies improved the modelled long-term storage trends by assimilated total water storage observations according to GRACE into hydrological models (Khaki et al., 2018; Schumacher et al., 2018).

In this study, long-term storage variations were observed in the Luangwa river basin, but not reproduced by a standard implementation of a conceptual model. The objective of this study was to identify potential and so far overlooked processes underlying these low-frequency variations in a combined data analysis and model hypothesis testing approach. More specifically, we here tested the hypotheses that the degree to which a conceptual hydrological model can reproduce observed long-term, low-frequency water storage variations depends (1) on the choice of the forcing data source used as input to the model and (2) on the incorporation of processes allowing long-term memory effects in the model.

5.2. Site description

The Luangwa River is a 770 km long tributary of the Zambezi in Zambia which is mostly unregulated (see Figure 5.1). Its 159,000 km² large basin area is poorly gauged and mostly covered with deciduous forests, shrubs and savanna. The elevation varies up to 1850 m between the low lying areas around the river and the highlands. In this semi-arid area, there is a distinct wet season from October to April with heavy rains up to 100 mm month⁻¹. Nevertheless, the mean annual evaporation (1555 mm yr⁻¹) exceeds the mean annual precipitation (970 mm yr⁻¹) (Hulsman et al., 2020a; The World Bank, 2010). See Chapter 2 for more detailed information on the Luangwa basin.

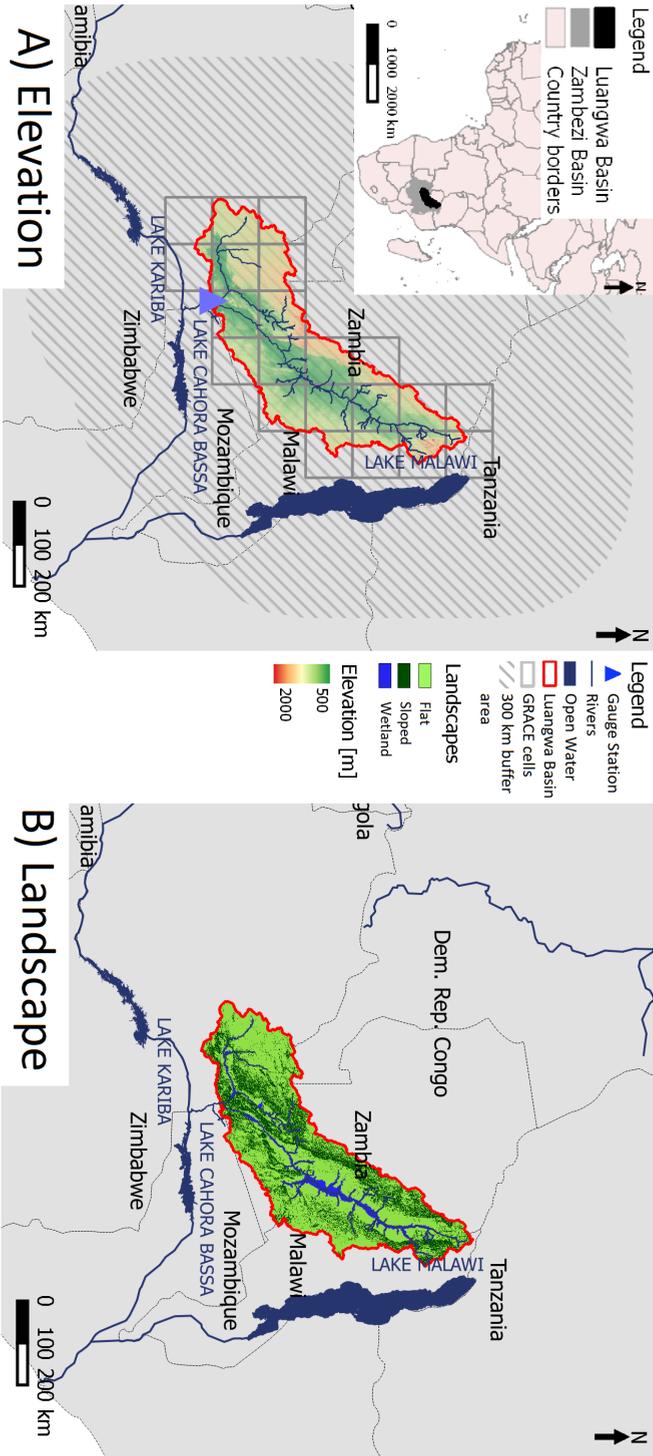


Figure 5.1: Map of the Luangwa River Basin in Zambia with A) the elevation, and B) the main landscape types.

5.3. Data availability

In this study, hydro-meteorological data as shown in Table 5.1 were used. This included two satellite-based precipitation products (CHIRPS and TRMM) and five actual evaporation products (WaPOR, SEBS, SSEBop, GLEAM and MOD16). Land-cover changes were assessed using the NDVI (Normalized Difference Vegetation Index) and LAI (Leaf Area Index). Temperature data according to CRU (Climatic Research Unit) was used to estimate the potential evaporation with the Hargreaves (Hargreaves and Allen, 2003; Hargreaves and Samani, 1985) and Thornthwaite (Maes et al., 2019) method.

Processed GRACE (Gravity Recovery and Climate Experiment) observations generated by CSR (Centre for Space Research), GFZ (GeoForschungsZentrum Potsdam) and JPL (Jet Propulsion Laboratory) were obtained from the GRACE Tellus website (<https://grace.jpl.nasa.gov/>). This study used the average of these three sources which previously processed the raw data to remove atmospheric mass changes, systematic errors and noise, and to subtract the 2004 – 2009 time-mean baseline (Landerer and Swenson, 2012; Swenson and Wahr, 2006; Wahr et al., 1998). As a result, total water storage anomalies were available in equivalent water thickness. Total water storage anomaly observations include all terrestrial water storage components, hence water stored in the surface water, soil moisture and groundwater.

Altimetry data was extracted from the DAHITI website (<https://dahiti.dgfi.tum.de/en/>) for the Cahora Bassa reservoir, Kariba reservoir and Lake Malawi (Schwatke et al., 2015). In-situ discharge data was used for the Great East Road Bridge gauging station at the basin outlet (30° 13' E and 14° 58' S) and was obtained from the Zambian Water Resources Management Authority (WARMA) for the time period 2002 to 2016 with a temporal coverage of 35%.

For the following data analysis, gridded observations were averaged for the entire basin, whereas for use in the distributed hydrological model, gridded observations were rescaled to the model resolution of 0.25° by (a) taking the mean of all cells located within a model cell if the resolution was smaller, or (b) dividing each cell into multiple cells if the resolution was larger. For the hydrological model, gridded observations were used for the topography to classify the landscape into hydrological response units (see Section 5.4.2), climate (precipitation and temperature) to force the model, and total water storage anomalies to calibrate/evaluate the model.

5.4. Approach

This study consisted of three steps. In the first step we analysed the effect of the choice of the data source used to explain observed total water storage variations to understand whether any of the data contain, in principle, sufficient information to at least broadly reflect the dynamics of storage variations. This was necessary to rule out that the model's inability to reproduce long-term storage variations is merely an artefact of unsuitable data. Thus, we investigated whether periods of high water storage anomalies roughly coincide with periods of high precipitation anomalies and/or low evaporation anomalies and vice versa. To do so, we con-

Table 5.1: Data used in this study

	Time period	Time resolution	Spatial resolution	Product n/ame	Long-term annual mean	Source
Digital elevation map Precipitation	n/a	n/a	0.02°	GMTED	n/a	(Danielson and Gesch, 2011)
	1998 – 2016	Daily	0.05°	CHIRPS	1127 mm yr ⁻¹	Version 2 (Funk et al., 2014)
	1998 – 2016	Daily	0.25°	TRMM	1029 mm yr ⁻¹	Version 3B42 (Huffman et al., 2014, 1995, 2007)
Evaporation	2009 – 2016	10-day	0.00223°	WaPOR	882 mm yr ⁻¹	Version 1.1 (FAO, 2018; FAO and IHE Delft, 2019)
	2002 – 2013	Monthly	0.05°	SEBS	657 mm yr ⁻¹	(Su, 2002)
	2003 – 2016	Monthly	0.01°	SSEBop	837 mm yr ⁻¹	Version 4 (Allen et al., 2007; Bastiaanssen et al., 1998; Senay et al., 2007)
	2003 – 2016	Monthly	0.25°	GLEAM	751 mm yr ⁻¹	Version 3.3b (Martens et al., 2017; Miralles et al., 2011)
Temperature	2002 – 2016	Monthly	500 m	MOD16	793 mm yr ⁻¹	MOD16A2 Version 6 (Running et al., 2017)
	2002 – 2016	Monthly	0.5°	CRU	22°	Time-series (TS) data version 4.01 (University of East Anglia Climatic Research Unit et al., 2017)
NDVI LAI	2002 – 2016	8 days	30 m	n/a	0.12	Derived from Landsat 7
	2002 – 2016	Monthly	0.05°	n/a	1.48	Version 5 (Claverie et al., 2014)
Total water storage	2002 – 2016	Monthly	1°	GRACE	8.8 mm	Pre-processed by CSR and GFZ (Version RL05.DSTVSCS1409), and JPL (Version RL05_1.DSTVSCS1411) https://grace.jpl.nasa.gov/ (Landerer and Swenson, 2012; Swenson, 2012; Swenson and Wahr, 2006)
Altimetry	2002 – 2016	10 or 35 days	n/a	DAHITI	n/a	https://dahiti.dgfi.tum.de/en/ (Schwatke et al., 2015)
Discharge (Luangwa Bridge gauging station)	2002 – 2016	Daily	n/a	n/a	138 mm yr ⁻¹	WARMA

trusted long-term estimates of variables such as precipitation, potential and actual evaporation from multiple data sources with the observed water storage variations. This allowed a preliminary assessment of which data sources are more consistent with the observed low-frequency storage variations than others. Based on that, we then analysed, in a second step, four different combinations of data sources, i.e. precipitation and potential evaporation, as input for a hydrological model and evaluated their respective effects to reproduce the observed long-term storage variations with the model. In a third step, we then iteratively formulated and tested several alternative model hypotheses, incorporating a model component, such as regional groundwater export, to account for long-memory effects.

In general, long-term total water storage variations are a result of changes in precipitation, evaporation, discharge or any combination thereof (Eq.5.1). While climate variability can cause long-term variations in precipitation and atmospheric water demand (i.e. potential evaporation), land-cover changes can affect the partitioning between evaporative fluxes and streamflow (Gallart and Llorens, 2003; Hrachowitz et al., 2020; Li et al., 2017; Nijzink et al., 2016b; Oguntunde et al., 2006; Saft et al.,

2016; Warburton et al., 2012). In addition, long-term storage variations can be a result of slow inter-basin groundwater exchange (Bouaziz et al., 2018; Nelson and Mayo, 2014; Pellicer-Martínez and Martínez-Paz, 2014).

$$\frac{dS}{dt} = P - E - Q \quad (5.1)$$

Where S is total water storage, P precipitation, E evaporation and Q discharge.

5.4.1. Data analysis

Long-term, basin-averaged satellite observations of the precipitation according to CHIRPS and TRMM, actual evaporation according to WaPOR, SEBS, SSEBop, GLEAM and MOD16, potential evaporation according to the Hargreaves (Hargreaves and Allen, 2003; Hargreaves and Samani, 1985) and Thornthwaite (Maes et al., 2019) methods, respectively, and land-cover based on the NDVI and LAI (Table 5.1) were contrasted with and compared to the water storage variations estimated by GRACE. For each of these data sources, the temporal variability was visualised on monthly and/or annual timescale.

To assess the potential role of regional groundwater import to or export from the basin, the long-term water balance was estimated using the average annual precipitation, evaporation and discharge from the different satellite products. Assuming negligible long-term storage changes and data uncertainties, surpluses or deficits in the long-term water balance, hence if $\bar{P} - \bar{E} - \bar{Q} \neq 0$, are then the result of groundwater import/export. In case of groundwater export, the average annual leaking flow can then be estimated according to (e.g. Bouaziz et al., 2018):

$$\bar{Q}_L = \bar{P} - \bar{E} - \bar{Q} \quad (5.2)$$

Where \bar{Q}_L is annual mean groundwater export [mm yr^{-1}], \bar{P} annual mean precipitation [mm yr^{-1}], \bar{E} annual mean evaporation [mm yr^{-1}] and \bar{Q} annual mean discharge [mm yr^{-1}].

5.4.2. Hydrological modelling

Benchmark model (Model A0)

The process-based distributed hydrological model used in this study for the Luangwa basin was step-wise developed and refined in previous studies (Hulsman et al., 2020a, b) following the FLEX-Topo modelling concept (Savenije, 2010). The benchmark model used in this study was similar to Model C in Chapter 4. Each $0.25^\circ \times 0.25^\circ$ model cell had the same model structure and parameter set, but was forced differently using spatially distributed forcing data with respect to the precipitation and potential evaporation (e.g. Euser et al., 2015). In addition, each cell was further discretized into functionally distinct landscape classes, i.e. hydrological response units (HRUs) based on the topography (Nijzink et al., 2016b). All HRUs

5

within a cell were connected through a common groundwater component (Figure 5.2A). This groundwater reservoir was also lumped over the entire basin assuming a homogeneous groundwater system (Hulsman et al., 2020b). The landscape was classified based on the local slope and “Height-above-the-nearest-drainage” (HAND; Rennó et al., 2008) into sloped areas (slope $\geq 4\%$), flat areas (slope $< 4\%$, HAND ≥ 11 m) and wetland areas (slope $< 4\%$, HAND < 11 m). As a result, 68% of the basin was classified as flat areas, 28% as sloped areas and 8% as wetlands (Figure 5.1B). This FLEX-Topo modelling concept was applied successfully in previous studies (Gao et al., 2014a; Gharari et al., 2014; Hulsman et al., 2020a). As illustrated in Figure 5.2A, the hydrological model consisted of multiple storage components representing the interception storage, unsaturated root-zone storage, as well as fast and slow responding storages. Each storage component was schematized as reservoir with corresponding water balance and constitutive equations as shown in Table 5.2. As the dominant processes and thus the associated model structures of the three individual HRUs were very similar to each other, the major differences between the HRUs were accounted for by different parameter values. Model process constraints were applied as shown in Table 5.3 to allow partly overlapping prior parameter distributions with relationships consistent with our physical understanding of the system (Gharari et al., 2014; Hrachowitz et al., 2014), and to limit equifinality (Beven, 2006). For example in the Luangwa basin, higher interception evaporation and larger root-zone storage capacities were expected in the densely vegetated, forest dominated sloped areas compared to the flat, grass- and shrub-land dominated areas and wetlands. Processes unique to a HRU were incorporated by adjusting the model structure where necessary. In sloped and flat areas for example, the groundwater system was recharged by downward infiltration whereas in wetlands this flow was assumed to be negligible due to shallow groundwater tables. Rather, water was assumed to be pushed upwards from the groundwater system into the unsaturated root-zone due to the groundwater head difference between the upland and wetland (Hulsman et al., 2020b). After having calculated the runoff for each grid cell, the total flow at the outlet was estimated by applying a simple routing scheme based on the flow distance to the outlet and a constant, calibrated flow velocity. This model consisted of 17 calibration parameters with uniform prior distributions and constraints as summarized in Table 5.3. In this benchmark model, the precipitation product CHIRPS was used and potential evaporation was calculated with the Hargreaves method (Table 5.4).

First model adaptation: Alternative forcing data (Models B0 – D0)

As first model adaptation, the forcing data was changed to assess the role of data uncertainty for the model’s ability to reproduce the observed long-term storage variations and to test whether some combinations of data sources allow model results to be more consistent with the observed storage variations than others. Starting with Model A0 as benchmark, different combinations of precipitation products, i.e. CHIRPS and TRMM, on the one hand and methods to estimate potential evaporation, i.e. Hargreaves and Thornthwaite, on the other hand were tested in Models B0 – D0 (Table 5.4).

Table 5.2: Equations applied in the hydrological model. Fluxes [mm d⁻¹]: precipitation (P), effective precipitation (P_e), potential evaporation (E_p), interception evaporation (E_i), plant transpiration (E_t), infiltration into the unsaturated zone (R_u), drainage to fast runoff component (R_f), delayed fast runoff (R_{fl}), groundwater recharge (R_r for each relevant HRU and $R_{r,tot}$ combining all relevant HRUs), upwelling groundwater (R_{GW} for each relevant HRU and $R_{GW,tot}$ combining all relevant HRUs), fast runoff (Q_f for each relevant HRU and $Q_{f,tot}$ combining all relevant HRUs), groundwater recharge into Deeper Groundwater reservoir (R_s), shallow groundwater flow (Q_{ss}), deep groundwater flow (Q_{sd}), groundwater loss (Q_L), total runoff (Q_m). Storages [mm]: storage in interception reservoir (S_i), storage in unsaturated root zone (S_u), storage in upper/deeper groundwater reservoir (S_{su} , S_{sd}), storage in fast reservoir (S_f). Calibration parameters (shown in bold): interception capacity (I_{max}) [mm], maximum upwelling groundwater (C_{max}) [mm d⁻¹], maximum root zone storage capacity ($S_{u,max}$) [mm], splitter (W) [-], shape parameter (β) [-], transpiration coefficient (c_e) [-], time lag (T_{lag}) [d], reservoir time scales [d] of fast (K_f) and slow (K_s , K_{sd}) reservoirs, reference groundwater level ($S_{s,ref1}$, $S_{s,ref2}$) [mm], groundwater splitter (W_s) [-]. Remaining parameters: areal weights for each grid cell (p_{HRU}) [-], time step (Δt) [d]. The equations were applied to each hydrological response unit (HRU) unless indicated differently.

Reservoir system	Water balance equation	Process functions	
Interception	$\frac{\Delta S_i}{\Delta t} = P - P_e - E_i \approx 0$ (5.3)	$E_i = \min(E_p, \min(P, \frac{I_{max}}{\Delta t}))$ $P_e = P - E_i$ (5.5)	
	Unsaturated zone	Flat: $\frac{\Delta S_u}{\Delta t} = P_e - E_t - R_f$ (5.6)	$E_t = \min((E_p - E_i), \min(\frac{S_u}{\Delta t}, (E_p - E_i) \cdot \frac{S_u}{S_{u,max}} \cdot \frac{1}{c_e}))$ $R_{GW} = \min((\frac{\min(S_{su}, S_{s,ref1})}{S_{s,ref1}}) \cdot C_{max} \cdot \frac{S_{su}}{p_{HRU}})$ (5.10)
Sloped: $\frac{\Delta S_u}{\Delta t} = R_u - E_t$ (5.7)		if $S_u + R_{GW} \cdot \Delta t > S_{u,max}$: $R_{GW} = \frac{S_{u,max} - S_u}{\Delta t}$ (5.11)	
Wetland: $\frac{\Delta S_u}{\Delta t} = P_e - E_t - R_f + R_{GW}$ (5.8)		Sloped: $R_u = (1 - C) \cdot P_e$ (5.12) $C = 1 - (1 - \frac{S_u}{S_{u,max}})^\beta$ (5.13)	
		Fast runoff	$\frac{\Delta S_f}{\Delta t} = R_{fl} - Q_f$ (5.14)
Upper Groundwater	$\frac{\Delta S_{su}}{\Delta t} = R_{r,tot} - R_{GW,tot} - Q_{ss}$ (5.21)		Flat/Wetland: $R_f = \frac{\max(0, S_u - S_{u,max})}{\Delta t}$ (5.18)
			Flat: $R_{fl} = (1 - W) \cdot R_f$ (5.19)
	$\frac{\Delta S_{su}}{\Delta t} = R_{r,tot} - R_{GW,tot} - Q_{ss} - Q_L$ (5.22)		Wetland: $R_{fl} = R_f$ (5.20)
		$R_r = W \cdot C \cdot P_e$ (5.24) $R_{r,tot} = \sum_{HRU} (p_{HRU} \cdot R_r)$ (5.25)	
Deeper Groundwater	$\frac{\Delta S_{su}}{\Delta t} = R_{r,tot} - R_{GW,tot} - Q_{ss} - R_s$ (5.23)	$R_{GW,tot} = \sum_{HRU} (p_{HRU} \cdot R_{GW})$ (5.26)	
		$Q_{ss} = \frac{\max(0, S_{su})}{K_s}$ (5.27)	
	$\frac{\Delta S_{sd}}{\Delta t} = R_s - Q_L$ (5.31)	$R_s = 0$ (5.28)	
		$R_s = \frac{W_s \cdot S_{su}}{\Delta t}$ (5.29)	
Total runoff	$Q_m = Q_{f,tot} + Q_{ss}$ (5.36) $Q_m = Q_{f,tot} + Q_{ss} + Q_{sd}$ (5.37)	$R_s = \frac{\min(S_{su}, \max(0, S_{su} - S_{s,ref2}))}{\Delta t}$ (5.30)	
		$Q_L = \text{const.}$ (5.33)	
	$Q_L = \frac{S_{sd}}{K_{sd}}$ (5.34) $Q_{sd} = \frac{S_{sd}}{K_{sd}}$ (5.35)		
		$Q_{f,tot} = \sum_{HRU} (p_{HRU} \cdot Q_f)$ (5.38)	

Table 5.3: Model parameters and prior distributions

Landscape class	Parameter	min	max	Unit	Constraint	Comment
Entire basin	C_e	0	1	-		All models
	K_s	90	110	d		All models
	$S_{s,ref1}$	1	50	mm		All models
	Q_L	0	0.5	mm d ⁻¹		Models A1, A2, A3
	K_{sd}	100	2500	d		Models A4, A5
	$S_{s,ref2}$	1	50	mm		Models A3, A4, A5
Flat	W_s	0	1	-		Model A2
	I_{max}	0	5	mm d ⁻¹	$I_{max,sloped} > I_{max,flat}$	All models
	$S_{u,max}$	10	800	mm		All models
	K_f	10	12	d		All models
Sloped	W	0.01	1	-		All models
	I_{max}	0	5	mm d ⁻¹		All models
	$S_{u,max}$	10	800	mm	$S_{u,max,sloped} > S_{u,max,flat}$	All models
	β	0	2	-		All models
	T_{lag}	1	5	d		All models
Wetland	K_f	10	12	d		All models
	W	0	1	-	$W_{sloped} > W_{flat}$	All models
	I_{max}	0	5	mm d ⁻¹	$I_{max,wetland} < I_{max,sloped}$	All models
	$S_{u,max}$	10	400	mm	$S_{u,max,wetland} < S_{u,max,sloped}$	All models
	K_f	10	12	d		All models
River profile	C_{max}	0.01	5	mm d ⁻¹		All models
	v	0.01	5	m s ⁻¹		All models

Table 5.4: Overview of model combinations

	Precipitation product	Potential evaporation method
Model A0	CHIRPS	Hargreaves
Model B0	CHIRPS	Thornthwaite
Model C0	TRMM	Hargreaves
Model D0	TRMM	Thornthwaite

Second model adaptation: Alternative model structure (Model A1 – A5)

As second model adaptation, the model structure was changed to test whether deep groundwater flow or inter-basin groundwater export/import was a relevant driver for the observed long-term storage variations. In this study, a distinction was made between shallow groundwater flow (Q_{ss}), deep groundwater flow (Q_{sd}) and groundwater loss (Q_L). While the shallow and deep groundwater flow reached the river, the groundwater loss (Q_L) leaked out of the Luangwa basin and potentially reached the Zambezi river further downstream. Based on benchmark Model A0, hence using CHIRPS for precipitation and the Hargreaves method to estimate potential evaporation, the model structure was modified to introduce long-term storage memory.

With Model A1, it was tested whether only groundwater export, hence groundwater leaking out of the Luangwa basin, was a dominant driver for the long-term storage variations. In this model, groundwater loss (Q_L) was introduced (Figures 5.2B and 5.3) which did not reach the river (Eq.5.36) and, in the spirit of model parsimony, was assumed to be constant, regardless of the water content in the Upper Groundwater reservoir to limit the number of calibration parameters as no

additional information was available. Thus, the Upper Groundwater reservoir (S_{su}) was formulated as a deficit store that can become negative. However, the shallow groundwater flow Q_{ss} only occurred when this storage was positive (if $S_{su} > 0$, Eq.5.27). Such a formulation allowed groundwater to keep on draining, and thus groundwater levels falling, even if discharge in the river ceased during dry periods (e.g. Bouaziz et al., 2018; Hrachowitz et al., 2014).

With Model A2, it was tested whether constant groundwater export from a second, Deeper Groundwater reservoir can explain the observed long-term storage variations. In this model, groundwater seeped from the Upper Groundwater reservoir into a Deeper Groundwater reservoir as fraction of the water content in the Upper Groundwater reservoir (R_s , Eq.5.29, Figures 5.2C and 5.3). From this Deeper Groundwater reservoir, constant groundwater loss (Q_L) leaked out of the basin similar to Model A1.

With Model A3, it was tested whether constant groundwater export from the Deeper Groundwater reservoir recharged only during wet seasons, was the main driver for long-term storage variations. In this model, groundwater only seeped into the Deeper Groundwater reservoir when the groundwater level in the Upper Groundwater reservoir exceeded a reference level ($S_{s,ref2}$, Eq.5.30, Figures 5.2D and 5.3). From there constant groundwater loss (Q_L) leaked out of the basin similar to Models A1 and A2.

With Model A4, it was tested whether variable groundwater export from the Deeper Groundwater reservoir recharged only during wet seasons, was the main driver for long-term storage variations. In this model, the groundwater loss (Q_L , Figures 5.2E and 5.3) was a function of the water content in the Deeper Groundwater reservoir (Eq.5.34). This groundwater loss (Q_L) did not reach the river similar to Models A1 – A3.

With Model A5, it was tested whether variable groundwater flow from the Deeper Groundwater reservoir recharged only during wet seasons, was the main driver for long-term storage variations. In this model, the groundwater drained from the Deeper reservoir into the river as Q_{sd} contributing to the total river flow (Eq.5.37, Figures 5.2F and 5.3). Hence, only in Model A5 deep groundwater reached the gauged river system whereas in Models A1 – A4 groundwater leaked out of the basin.

Figure 5.3 gives an overview of all alternative model hypotheses tested in this study. The relevant model equations are given in Table 5.2 and the corresponding prior parameter distributions in Table 5.3.

Third model adaptation: Alternative forcing data and model structure

As third model adaptation, the forcing and the model structure were changed simultaneously. For this purpose, the best performing model based on the results of the first model adaptation, i.e. changing the forcing data (Models A0 – D0) and the second model adaptation, i.e. changing the model structure (Models A0 – A5) were combined. For example, if Models D0 and A4 performed best, respectively, then the combined Model D4 using the forcing data applied in Model D0 and the model structure of Model A4 was tested. To ensure a robust representation of both, discharge and total water storage, the above model selection was based on the

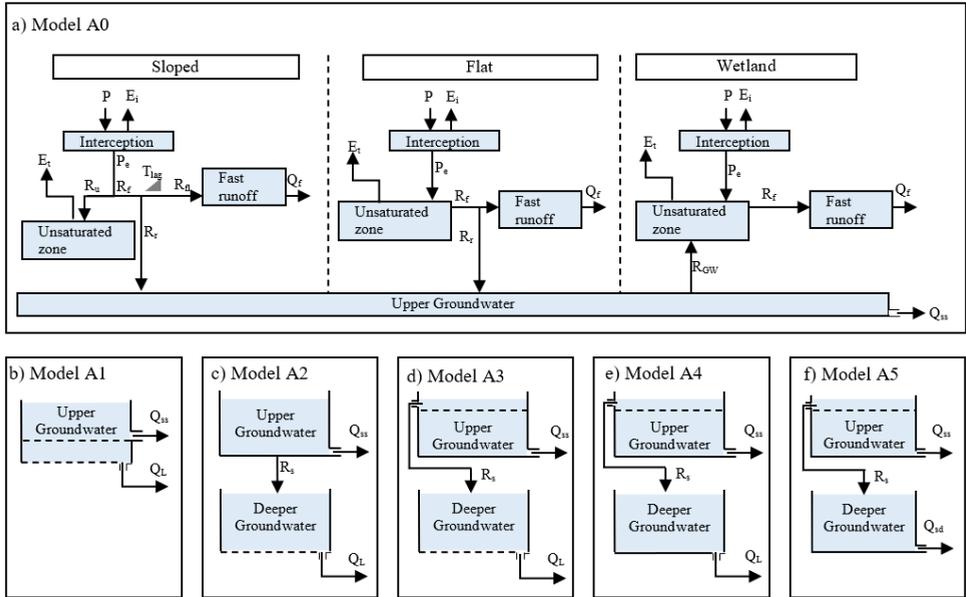


Figure 5.2: Schematisation of the model structure applied to each grid cell for Models A0 – A5. For Models A1 – A5 (B – F), only the groundwater module is shown for brevity and clarity of the presentation, as the rest of the model structure remained the same. Abbreviations: precipitation (P), effective precipitation (P_e), potential evaporation (E_p), interception evaporation (E_i), plant transpiration (E_t), infiltration into the unsaturated zone (R_u), drainage to fast runoff component (R_f), delayed fast runoff (R_{fl}), groundwater recharge (R_r), groundwater upwelling (R_{GW}), fast runoff (Q_f), groundwater recharge into Deeper Groundwater reservoir (R_s), shallow groundwater flow (Q_{ss}), groundwater loss (Q_L) and deep groundwater flow (Q_{sd}).

combined performance metrics for both variables. We explicitly acknowledge the possibility of this not being the combination that most reliably reflects real world processes. However, exhaustively testing all possible combinations goes beyond our computational capacity.

5.4.3. Model performance measures

The model performance was evaluated with respect to discharge and basin-average total water storage anomalies. With respect to discharge, eight hydrological signatures were evaluated simultaneously using the Nash-Sutcliffe efficiency ($E_{NS,\theta}$, Eq.5.39) or relative error ($E_{R,\theta}$, Eq.5.40), depending on the signature. The individual performance metrics included the Nash-Sutcliffe efficiency of the daily flow time-series ($E_{NS,Q}$) and its logarithm ($E_{NS,\log Q}$), of the flow duration curve ($E_{NS,FDC}$) and its logarithm ($E_{NS,\log FDC}$), and of the autocorrelation function of the daily flows ($E_{NS,AC}$). In addition the relative error of the mean seasonal runoff during dry and wet periods ($E_{R,RCdry}$, $E_{R,RCwet}$), and the rising limb density of the hydrograph ($E_{R,RLD}$) (Euser et al., 2013) were used. These signatures were combined, assuming equals weights, using the Euclidian distance ($D_{E,Q}$, Eq.5.41)

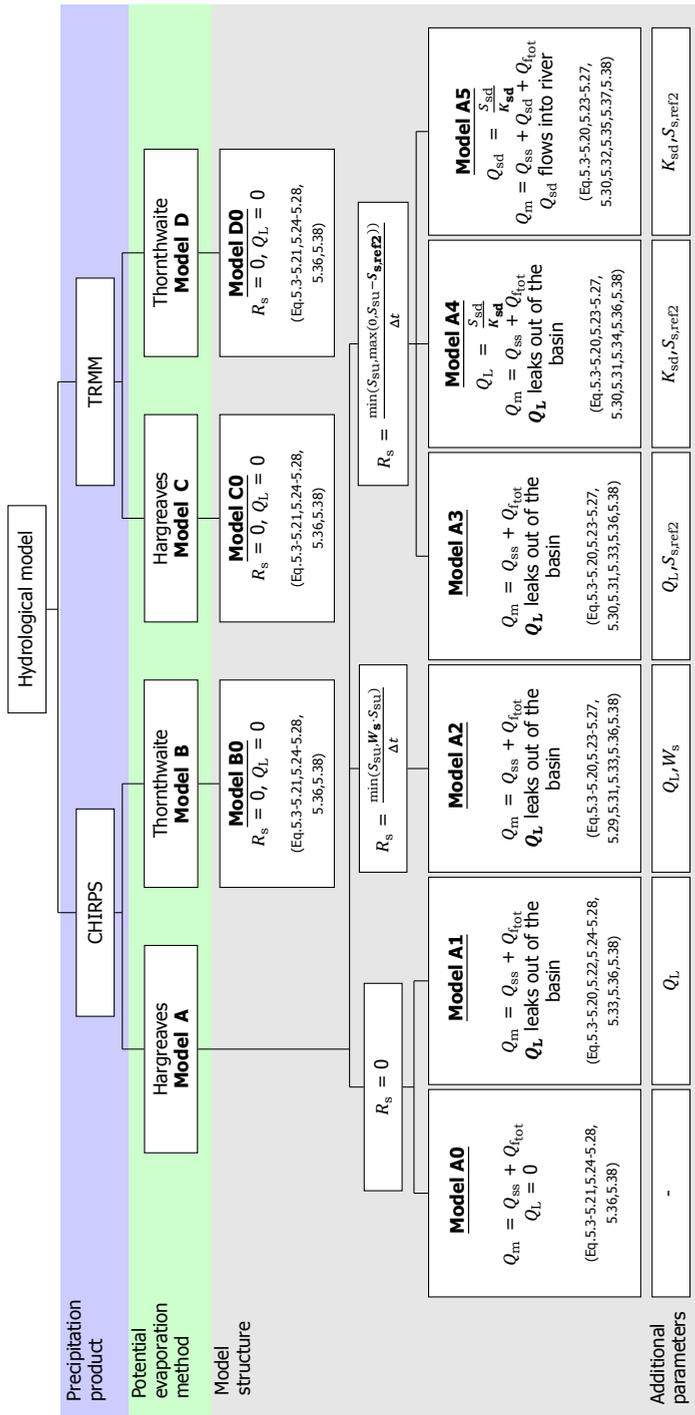


Figure 5.3: Overview hydrological models

with $D_{E,Q} = 1$ corresponding to the “perfect” model.

The model performance with respect to the basin-average total water storage anomalies was evaluated with the Euclidian distance ($D_{E,S}$, Eq.5.41) of the Nash-Sutcliffe efficiencies on monthly ($E_{NS,S,monthly}$) and annual ($E_{NS,S,annual}$) timescale. On annual timescale, the Nash-Sutcliffe efficiency was calculated for the annual minima and maxima separately which were then averaged to obtain $E_{NS,S,annual}$. The annual time-series were normalised by dividing it with the maximum range in the observed annual minima or maxima total water storage respectively. With this performance measure for the total water storage, more emphasis could be given to annual variations rather than to seasonal variations only.

The combined model performance with respect to discharge and total water storage anomalies ($D_{E,QS}$) was calculated with the Euclidian distance (Eq.5.41) using $D_{E,Q}$ for the discharge and $D_{E,S}$ for the total water storage. This performance measure was used to select the best performing models representing both the discharge and the total storage as good as possible.

Table 5.5: Overview of equations used to calculated the model performance

Name	Objective function	Symbol explanation
Nash-Sutcliffe efficiency	$E_{NS,\theta} = 1 - \frac{\sum_t (\theta_{mod}(t) - \theta_{obs}(t))^2}{\sum_t (\theta_{obs}(t) - \theta_{obs})^2}$	(5.39) θ : variable
Relative error	$E_{R,\theta} = 1 - \frac{ \theta_{mod} - \theta_{obs} }{\theta_{obs}}$	(5.40) θ : variable
Euclidian distance over multiple variables	$D_E = 1 - \sqrt{\frac{1}{N} \left(\sum_n (1 - E_n)^2 \right)}$	(5.41) E_n : model performance metric of variable n

5.4.4. Parameter selection procedure

The hydrological model was calibrated by running the model with 10^5 random parameter sets generated with a Monte-Carlo sampling strategy with uniform prior parameter distributions. Then, following two different strategies, the optimal parameter set was selected according to the model performance metrics as previously described with respect to 1) discharge ($D_{E,Q}$) and 2) discharge combined with total water storage ($D_{E,QS}$). The 5% best-performing parameter sets with respect to $D_{E,Q}$ or $D_{E,QS}$ were considered as feasible. The feasible parameter sets were used to evaluate the model performance with respect to discharge and total water storage anomalies individually and combined. The model was run for the time period 1995 – 2016 and calibrated/evaluated for the time period 2002 – 2016 using the first seven years as warm-up period. The entire time period (2002 – 2016) was

used to estimate the model performance with respect to discharge and total water storage to capture the long-term variability as good as possible.

In addition, the predictive strength of the benchmark Model A0 and the best performing model hypothesis (i.e. third model adaptation; Section 5.4.2) were compared by calibrating both models with respect to discharge and total water storage simultaneously ($D_{E, QS}$) for the time period 2002 – 2012, and post-calibration evaluating the models with respect to total water storage for the time period 2012 – 2016. Due to the limited data availability in 2012 – 2016, the model could not be evaluated with respect to discharge.

5.5. Results

5.5.1. Data analysis

Total water storage anomalies

In the Luangwa basin, the total water storage anomalies varied both seasonally and in the long-term (for example Figure 5.4A). The seasonal variation, hence the difference between the annual maximum and minimum, remained rather similar throughout the years (on average 225 mm). However, the annual minima, mean and maxima changed over the years indicating relatively dry conditions in the Luangwa basin for example during the 2005 – 2007 period and wetter conditions in the 2009 – 2011 period. The annual minima varied between -164 mm in 2016 and -67 mm in 2009, while the annual maxima varied between 75 mm in 2016 and 183 mm in 2010. Also the annual mean varied over the years between -46 mm in 2006 and 48 mm in 2010. This study focused on annual minima/maxima separately instead of the annual mean to distinguish processes dominant in wet seasons influencing the annual maxima and dry seasons affecting the annual minima.

One possibility is that these variations were a result of uncertainties in GRACE observations as the Luangwa basin is relatively small (150,000 km²) relative to the resolution of GRACE. Previous studies estimated errors in GRACE observations to be about 20 mm for areas of around 63,000 km² (Landerer and Swenson, 2012; Vishwakarma et al., 2018). But similar long-term variations were also observed for the entire Zambezi basin (Figure 5.4B), which is considerably larger (1,390,000 km²) and where the maximum variation (194 mm) was an order of magnitude larger than the average uncertainty error of 20 mm.

In addition, long-term variations in large open water bodies could influence the GRACE signal. In this study, multiple open water bodies were within a radius of 300 km of the Luangwa Basin (Figure 5.1B) which typically is the distance used for data smoothing when processing GRACE data (Blazquez et al., 2018a; Landerer and Swenson, 2012). The area of these open water bodies were 2% of the Luangwa basin for the Cahora Bassa reservoir, 4% for the Kariba reservoir and 20% for Lake Malawi. As no long-term variations were observed in the altimetry observations for the Cahora Bassa reservoir (Figure C.1 in Appendix C) and since this reservoir had a small area compared to the Luangwa basin, the effect of this reservoir was assumed to be negligible. For the Kariba reservoir (Figure 5.4C) and Lake Malawi (Figure 5.4F), long-term variations were observed in the altimetry data, but with

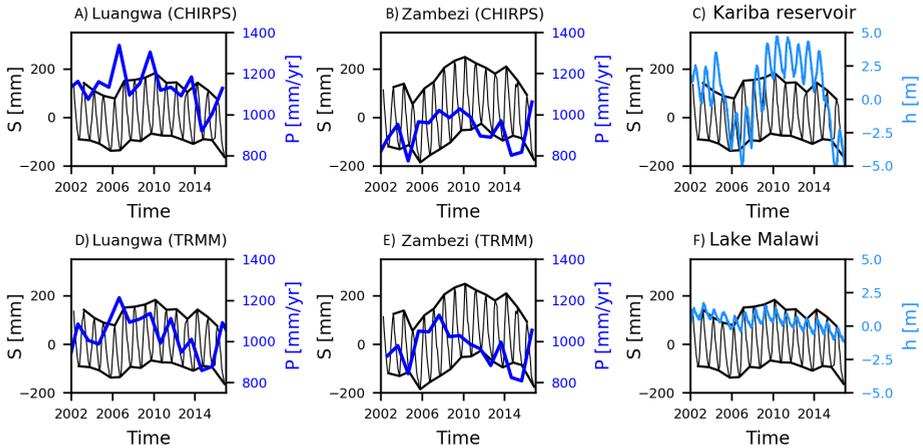


Figure 5.4: Basin-average total water storage (black) and annual rainfall (dark blue) according to CHIRPS (A and B) and TRMM (D and E) for the Luangwa (A and D) and Zambezi (B and E) river basin, or altimetry observations (light blue) at C) Kariba reservoir and F) Lake Malawi.

a low temporal correlation with the total water storage as shown in Figure C.2 in Appendix C. For the Zambezi basin where similar long-term storage variations were observed (Figure 5.4B), these three open water bodies covered together 2.7% of the basin. This was considered to be too small to have a significant effect. That is why it is plausible to assume that these long-term storage variations were not dominated by uncertainties in the GRACE observations.

Precipitation

Alternatively, long-term variations in the total water storage can be caused by changes in the precipitation. In the Luangwa basin, the annual observed precipitation volumes varied over the years, depending on the data source, from 920 mm to 1337 mm (CHIRPS) and from 858 mm to 1213 mm (TRMM), as shown in Figures 5.4A and D. In general, precipitation anomalies preceded storage variations by roughly 1 – 3 years. According to CHIRPS (Figure 5.4A), the rainfall volumes peaked in 2006 and 2009 with a significant decrease in 2008 – 2009 and 2014. While the increased rainfall volumes in 2006 and 2009 could explain the increased total water storage anomalies between 2008 and 2010, the significantly decreased rainfall volumes in 2008 – 2009 did not correspond to the long-term total water storage pattern. The correlation between the annual rainfall volumes according to CHIRPS and the annual maximum total water storage showed a $R^2 = 0.10$ without taking any time shift into account and reached up to $R^2 = 0.29$ with a two year time shift.

According to TRMM, the annual rainfall volumes decreased in 2004 – 2005 which could explain the decreased lower total water storage in 2006. This was followed by several wet years with a maximum rainfall volume of 1213 mm in 2006 which

could explain the increased total water storage starting in 2007. The annual rainfall volumes decreased significantly in 2014 – 2015 as low as 858 mm which corresponded to the decreased total water storage in 2016. The correlation between the annual rainfall volumes according to TRMM and the annual maximum total water storage reached $R^2 = 0.28$ without taking any time shift into account and reached up to $R^2 = 0.34$ with a two year time shift.

This difference between CHIRPS and TRMM illustrated the high sensitivity of the annual rainfall volumes to the underlying processing techniques (Cohen Liechti et al., 2012; Le Coz and van de Giesen, 2020; Mazzoleni et al., 2019; Thiemiig et al., 2012). Strikingly, for the entire Zambezi river basin the annual variability in the precipitation according to both CHIRPS and TRMM show a similar pattern compared to each other and to the storage variations. The annual rainfall volumes decreased in 2004 followed by low total water storages in 2006, after which both the rainfall and total water storage increased with a maximum in 2009 (CHIRPS), 2007 (TRMM) and 2010 (GRACE). These observations suggest that long-term variations in precipitation alone already contain considerable information to potentially explain much of the observed long-term storage variations.

Evaporation

The two different methods to estimate potential evaporation and its variations over the study time period, gave dramatically different results. While the Hargreaves method suggested a long-term mean annual $E_p = 1565 \text{ mm yr}^{-1}$ (Figure 5.5A), Thornthwaite estimated long-term mean $E_p = 1904 \text{ mm yr}^{-1}$ (Figure 5.5B). Major long-term variations in E_p were only observed for estimates based on the Thornthwaite method (Figure 5.5B), but with a different pattern compared to the total water storage resulting in low correlation coefficients when focusing on the annual mean variations (R^2). In contrast, no discernible long-term fluctuations were observed when applying the Hargreaves method ($R^2 = 0.03$). As the potential evaporation did change over the years according to the Thornthwaite method, it is possible this was one of the reasons why the modelled total water storage did not capture any long-term variations when using the Hargreaves method for the potential evaporation.

Analysis of the actual evaporation did not reveal any systematic long-term patterns that could clearly explain observed variations in the total water storage for most of the satellite products used in this study (Figure C.3 in Appendix C). In general, the magnitudes and long-term fluctuations varied for each satellite product as a result of different underlying assumptions and input data which could influence whether or not long-term fluctuations are visible. This resulted in a range of $R^2 = 0.02 - 0.17$ with respect to the annual minima for all satellite products used in this study except for SSEBop which showed the highest $R^2 = 0.37$ and where the evaporation increased between 2006 and 2010 similar to the storage (Figures 5.5C and C.3 in Appendix C). Note, that the observed annual minimum storage increase of 67 mm over three years (2006 – 2009), which in fact is an accumulated difference arising from the combined history of inputs and outputs over that period, can result among others from a mean deviation of only 0.06 mm d^{-1} in the evaporation, which

is by far within the uncertainty range of many satellite-based evaporation products (Long et al., 2014; Westerhoff, 2015). Hence, evaporation can potentially be one of the drivers for the observed long-term storage fluctuations, but additional in depth analyses is necessary to substantiate this hypothesis which was outside the scope of this study due to the limited ground observations available.

Overall, long-term variations in potential and actual evaporation, according to most satellite products used here, exhibited less direct correspondence with water storage variations, which was likely a consequence of the subtle spatially varying interactions between water supply and atmospheric water demand in this largely water limited environment. Thus, while actual evaporation is largely controlled by water supply in hillslope regions, it is to a higher degree dominated by variations in atmospheric water demand in wetland areas, where sufficient water supply is sustained by shallow groundwater throughout most of the year. On the basin average, these processes can, to some degree, cancel each other out and thus prevent the development of a clear long-term signal. Based on the above analysis it therefore remains difficult to meaningfully assess the uncertainty of the different analysed evaporation products.

5

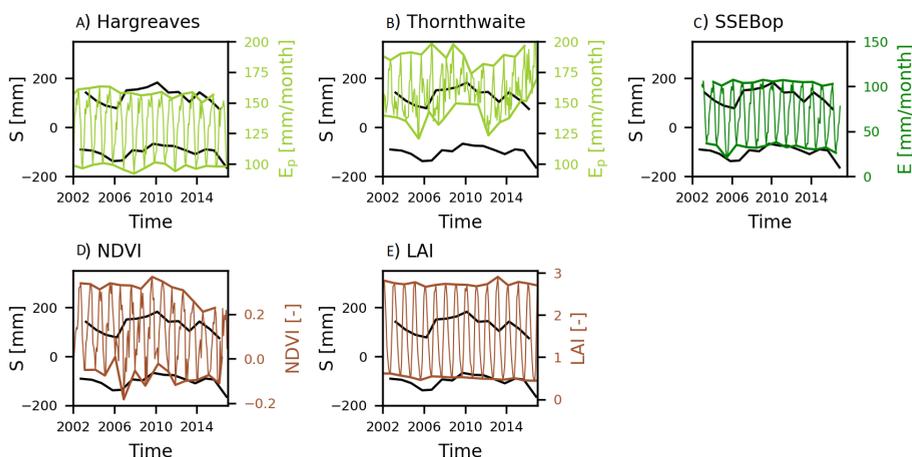


Figure 5.5: Basin-average total water storage (black) with respect to the annual minima/maxima combined with basin-average A) monthly potential evaporation according to Hargreaves (light green) and B) Thornthwaite (light green), C) monthly actual evaporation according to SSEBop (dark green), D) NDVI (brown), and E) LAI (brown) including the annual minima/maxima of the respective variables.

Land-cover

Affecting the magnitudes of transpiration, land-cover changes could also be one of the drivers for the observed annual storage variations. In the Luangwa basin, deforestation, forest recovery and agricultural expansion have occurred in the past (Handavu et al., 2019; Phiri et al., 2019a, b). However, inspections of time-series of LAI and NDVI (Figure 5.5) did not reveal any significant long-term variations directly corresponding with water storage variations over the 2002 – 2016 period.

While LAI did not exhibit any significant long-term variation, NDVI showed some fluctuations, including a considerable decrease after 2010, which, however, did not directly correspond with the observed water storage variations. This resulted in low correlations between the annual mean total water storage and LAI ($R^2 = 0.003$) and NDVI ($R^2 = 0.04$). It was therefore assumed that land use change did not play a major role for the observed long-term storage variations.

Water balance

Another potential reason for the observed long-term storage variations can be regional, inter-basin groundwater exchange. For example, groundwater may leak out of the Luangwa basin below the river, thus never contributing to the (river) flow at the basin outlet, and into the Zambezi river basin further downstream eventually draining into that river or potentially even directly into the sea. Given the available observations, this would result in a water balance surplus for the Luangwa basin. Depending on the rainfall and evaporation products used, the water balance surplus in the Luangwa basin for the study period ranged between 9 and 332 mm yr⁻¹ (Table 5.1). This suggested that even in the likely presence of data uncertainty, groundwater export may occur at least to some degree in the study region. Assuming an inter-basin export of $\overline{Q_L} = 332$ mm yr⁻¹, discharge would be considerably overestimated as compared to actual discharge observations (Figure 5.6). To remain within the ranges spanned by multiple analytical solutions for water partitioning in the Budyko space (dark grey area in Figure 5.6; Gerrits et al., 2009), groundwater export should not exceed $\overline{Q_L} = 143$ mm yr⁻¹, which corresponds to a mean daily flow of $\overline{Q_L} = 0.39$ mm d⁻¹ or 13% of the annual rainfall. Therefore, based on the water balance, a plausible estimate for groundwater export of $\overline{Q_L} = 0 - 0.39$ mm d⁻¹ is in the following assumed for the study basin.

5.5.2. Hydrological modelling

Benchmark model (Model A0)

Following the first strategy, i.e. calibrating with respect to discharge, the benchmark Model A0 captured the discharge well (Figures 5.7A and B) with an optimum model performance of $D_{E,Q,opt} = 0.85$ (Table 5.6, Figure 5.8A). The modelled flow dynamics such as the timings of the wet and dry season were broadly consistent with the observations (Figure 5.7A), but the high flows were slightly underestimated and low flows somewhat overestimated (Figure 5.7B). In contrast, and in spite of its general ability to reproduce discharge, the model could only poorly reproduce the time-series of monthly and annual total water storage anomalies with $D_{E,S} = -14$ (Table 5.6, Figure 5.8A). On monthly timescale, the general seasonal fluctuations were modelled well with respect to the timings of the wet and dry season (Figure C.6A in Appendix C). However, the annual maxima were significantly overestimated and the annual minima underestimated (Figure 5.7C). In addition, the modelled total water storage did not reflect any fluctuations in the annual minima in contrast to the observations (Figure 5.7E, $R^2 = 0.07$), whereas the modelled annual maxima varied throughout the years, but with a different pattern compared to the

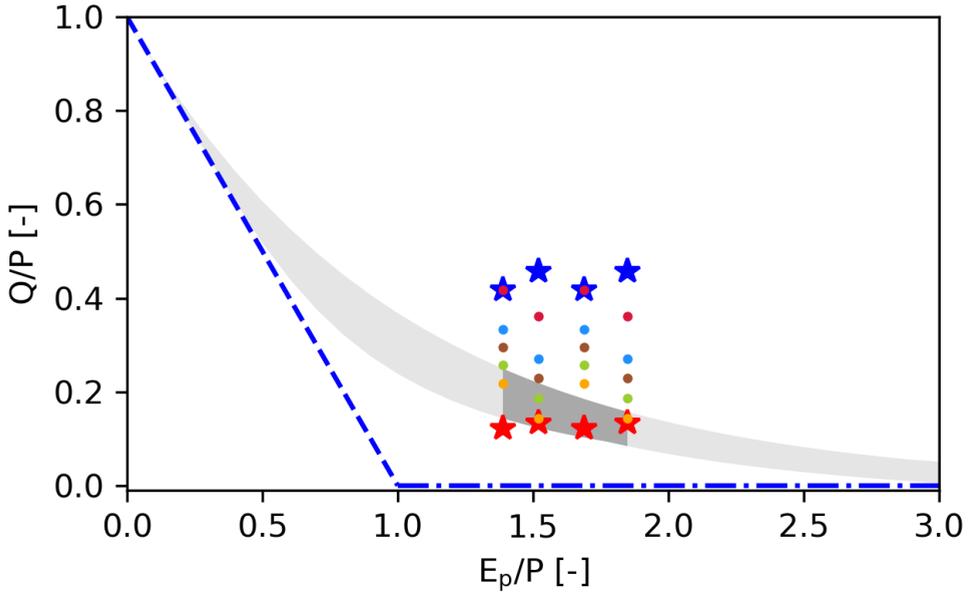


Figure 5.6: Runoff coefficient (Q/P) as a function of the dryness index (E_p/P) where Q is discharge, P precipitation, and E_p potential evaporation. The blue dashed line indicates the energy limit and the blue horizontal dash-dotted line the water limit. The grey area indicates envelope of analytical solutions according to Schreiber (1904), Ol'dekop (1911), Turc (1953), Pike (1964) and Budyko (1974). The dryness index was estimated using CHIRPS or TRMM for the precipitation and the Hargreaves method ($\bar{E}_p = 1565 \text{ mm yr}^{-1}$) or Thornthwaite ($\bar{E}_p = 1904 \text{ mm yr}^{-1}$) for the potential evaporation. The runoff coefficient was estimated with the same precipitation products and 1) recorded discharge without groundwater exchange (red stars), 2) estimated discharge including groundwater exchange ($\bar{Q} + \bar{Q}_L = \bar{P} - \bar{E}$, Eq.5.2) using the same precipitation products and SEBS (red dots), GLEAM (blue dots), MOD16 (brown dots), SSEBop (green dots) and WaPOR (orange dots) for the evaporation resulting in $\bar{Q}_L = 9 - 332 \text{ mm yr}^{-1}$ depending on the chosen satellite products, and 3) sum of recorded discharge and maximum groundwater export ($\bar{Q}_L = 332 \text{ mm yr}^{-1}$, blue stars). To remain within the Budyko space (dark grey area), the groundwater exchange should range between $\bar{Q}_L = -51 - 143 \text{ mm yr}^{-1}$ depending on the satellite products used. See Table 5.1 for the corresponding long-term values of the individual fluxes.

observations (Figure 5.7D, $R^2 = 0.20$). As a result, the overall model performance with respect to discharge and total water storage $D_{E,QS} = -9.6$ remained poor.

Following the second strategy, i.e. calibration with respect to discharge and total water storage simultaneously, the ability of the model to reproduce flow decreased significantly to $D_{E,Q} = -0.23$ (Table 5.6, Figure 5.8B). While the general flow dynamics were modelled well (Figure 5.9A), the flows were continuously overestimated (Figure 5.10A). In contrast, the modelled monthly and annual total water storage time-series improved ($D_{E,S} = -0.11$). The modelled total water storage mimicked the seasonal variations in the observation better (Figure C.7A in Appendix C), but with slight differences in the storage decrease during the dry seasons. The magnitudes of the annual maxima and minima corresponded better with the observations (Figure 5.10B) and the fluctuations in the annual maxima improved slightly (Figure

5.10C, $R^2 = 0.31$). However, the modelled storage did not reflect any fluctuations in the annual minima (Figure 5.10D, $R^2 = 0.06$). Hence, the overall model performance $D_{E,QS} = -0.17$ improved, but remained poor. Even when calibrating with respect to total water storage only, the annual minima did not reflect any fluctuations (Figure C.8 in Appendix C, $R^2 = 0.08$).

As a result, with this benchmark Model A0 the flows were modelled well as also the seasonal fluctuations in the total water storage. However, the long-term variations in the total water storage with respect to the annual maxima were poorly modelled and with respect to the annual minima completely missed.

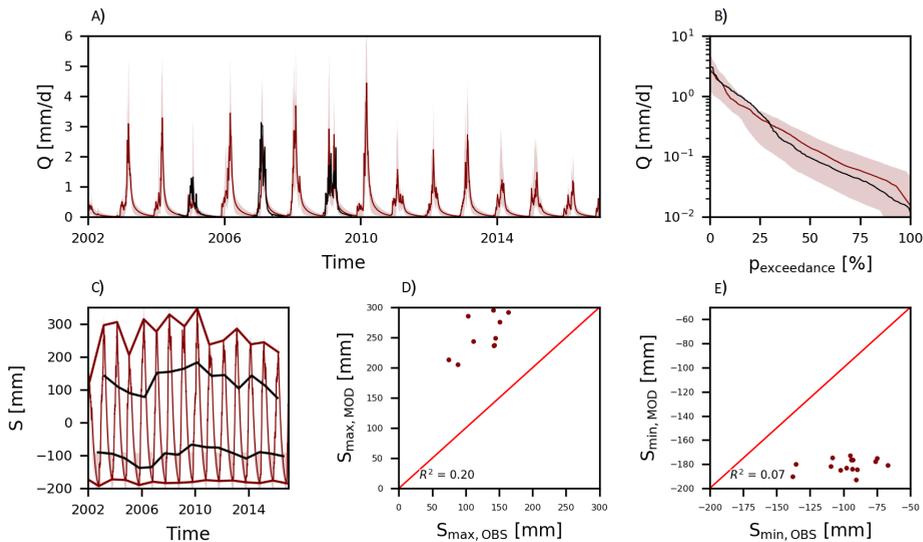


Figure 5.7: Range of model solutions for Model A0 for calibration strategy 1 with respect to A) hydrograph, B) flow duration curve in logarithmic scale, C) total water storage time-series, D) annual maximum total water storage, and E) annual minimum total water storage. In A) to C), the black line indicates the recorded data, the coloured line the solution with the highest calibration objective function with respect to discharge ($D_{E,Q}$) and the shaded area the envelope of the solutions retained as feasible. In D) and E), the recorded data are plotted on the horizontal axis and on the vertical axis the model solution with the highest calibration objective function with respect to discharge ($D_{E,Q}$). The red line indicates the 1:1 line.

First model adaptation: Alternative forcing data (Models B0 – D0)

Following the first calibration strategy, Models B0 – D0, using different combinations of input data sources, represented the discharge in general well with $D_{E,Q} = 0.85 - 0.92$ (Table 5.6, Figure 5.8A). All models reproduced the overall flow dynamics and magnitudes well (Figures C.4 and C.5A in Appendix C), especially Models C0 ($D_{E,Q} = 0.91$) and D0 ($D_{E,Q} = 0.92$). The monthly and annual total water storage remained poorly modelled for all models with $D_{E,S} = -3.4 - -0.48$ (Table 5.6, Figure 5.8A). On monthly timescale, the general seasonal fluctuations were modelled well

Table 5.6: Model performance with respect to discharge ($D_{E,Q}$), total water storage ($D_{E,S}$) and both combined ($D_{E,QS}$) including their 5/95% percentile ranges of the feasible parameter sets for Models A0 – D4.

	Strategy 1: Discharge calibration ($D_{E,Q}$)			Strategy 2: Multi-variable calibration ($D_{E,QS}$)		
	$D_{E,Q}$ ($D_{E,Q5/95\%}$)	$D_{E,S}$ ($D_{E,S5/95\%}$)	$D_{E,QS}$ ($D_{E,QS5/95\%}$)	$D_{E,Q}$ ($D_{E,Q5/95\%}$)	$D_{E,S}$ ($D_{E,S5/95\%}$)	$D_{E,QS}$ ($D_{E,QS5/95\%}$)
Model A0	0.85 (0.70 – 0.81)	-14 (-18 – -5.5)	-9.6 (-12 – -3.6)	-0.23 (-0.71 – -0.06)	-0.11 (-0.80 – -0.10)	-0.17 (-0.52 – -0.31)
Model B0	0.85 (0.72 – 0.81)	-3.4 (-9.2 – -1.7)	-2.1 (-6.2 – -0.94)	0.32 (-0.14 – 0.49)	0.00 (-0.65 – -0.09)	0.14 (-0.25 – 0.01)
Model C0	0.91 (0.80 – 0.88)	-0.85 (-4.5 – -0.34)	-0.31 (-2.9 – 0.05)	0.64 (0.26 – 0.72)	0.22 (-0.13 – 0.19)	0.39 (0.16 – 0.31)
Model D0	0.92 (0.84 – 0.90)	-0.48 (-2.2 – 0.21)	-0.05 (-1.3 – 0.43)	0.83 (0.56 – 0.88)	0.34 (0.09 – 0.28)	0.52 (0.34 – 0.46)
Model A1	0.84 (-0.13 – 0.71)	-15 (-15 – -0.87)	-11 (-10 – -0.51)	-0.20 (-1.1 – 0.07)	0.05 (-1.4 – -0.15)	-0.08 (-0.90 – -0.35)
Model A2	0.82 (-5.1 – 0.51)	-1066 (-813 – -3.4)	-753 (-575 – -3.3)	-0.24 (-11 – -1.0)	-0.47 (-7.5 – -0.68)	-0.36 (-7.6 – -3.3)
Model A3	0.87 (0.73 – 0.83)	-425 (-1133 – -11)	-300 (-801 – -7.2)	0.28 (-1.2 – 0.49)	-0.45 (-3.9 – -0.66)	-0.14 (-2.6 – -0.53)
Model A4	0.87 (0.73 – 0.83)	-9.8 (-27 – -3.6)	-6.7 (-19 – -2.3)	0.54 (-0.42 – 0.50)	0.16 (-0.64 – 0.11)	0.32 (-0.31 – 0.12)
Model A5	0.84 (0.68 – 0.79)	-13 (-18 – -5.3)	-9.0 (-12 – -3.5)	-0.31 (-0.72 – 0.03)	0.23 (-0.73 – 0.08)	-0.07 (-0.46 – -0.20)
Model D4	0.93 (0.85 – 0.91)	0.31 (-6.9 – -0.29)	0.51 (-4.6 – 0.49)	0.85 (0.61 – 0.89)	0.50 (0.11 – 0.37)	0.63 (0.35 – 0.53)

5

with slight differences mostly in the storage decrease during dry seasons (Figure C.6 in Appendix C). The magnitudes of the modelled annual minima corresponded well with the observation for all models, but the annual maxima were overestimated for Models B0 and C0, whereas this improved the most for Model D0 (Figure C.5B in Appendix C). In addition, the annual minimum storage did not exhibit any of the observed long-term variations in any of the models ($R^2 = 0.02 - 0.10$, Figure C.5C – D in Appendix C), whereas the fluctuations in the annual maxima improved the most for Model D0 ($R^2 = 0.35$). As a result, the overall model performance with respect to discharge and total water storage improved the most for Model D0 with $D_{E,QS} = -0.05$ (Table 5.6, Figure 5.8A) which remained poor.

Following the second calibration strategy, the modelled flow improved for all Models B0 – D0 to $D_{E,Q} = 0.32 - 0.83$ compared to the benchmark Model A0 (Table 5.6, Figure 5.8B). The general flow dynamics were represented well for all models (Figure 5.9), but the flow magnitudes were only captured well for Models C0 and D0 (Figure 5.10A). While Models A0 and B0 significantly overestimated the flows continuously, Model C0 only slightly overestimated the flows continuously and Model D0 only slightly underestimated the medium to low flows (Figure 5.10A). As a result, Model D0 had the highest model performance with respect to discharge with $D_{E,Q} = 0.83$ (Table 5.6, Figure 5.8B). Also the modelled monthly and annual total water storage improved for Models B0 – D0 with $D_{E,S} = 0.00 - 0.34$ compared to the benchmark Model A0 (Table 5.6, Figure 5.8B). On monthly timescale, the general seasonal variations were captured well for all models, but with slight differences in the storage decrease during dry seasons (Figure C.7 in Appendix C). The magnitudes of the annual minima and maxima corresponded well with the observations for all models (Figure 5.10B), whereas the fluctuations in the annual maxima only improved for Model D0 with $R^2 = 0.39$ (Figure 5.10C). On the other hand, the annual minima remained close to constant for all models ($R^2 = 0.00 - 0.03$; Figure 5.10D). The overall model performance with respect to discharge and

total water storage improved the most for Model D0 with $D_{E,QS} = 0.52$ (Table 5.6, Figure 5.8B).

As a result, the ability of the model to reproduce long-term variations of the total water storage during the wet seasons, i.e. the annual maxima, was considerably influenced by the choice of precipitation data source and the method to estimate potential evaporation. In contrast, the modelled dry season storage, i.e. annual minima, did not reflect the observed pattern for any combination of data sources but remained rather stable. Overall, the combination of TRMM with the Thornthwaite method (Model D0) here produced model results that were most consistent simultaneously with observed discharge and the observed total water storage variations. This suggests that the choice of data source can explain some of the inability of the model to reproduce long-term water storage variations.

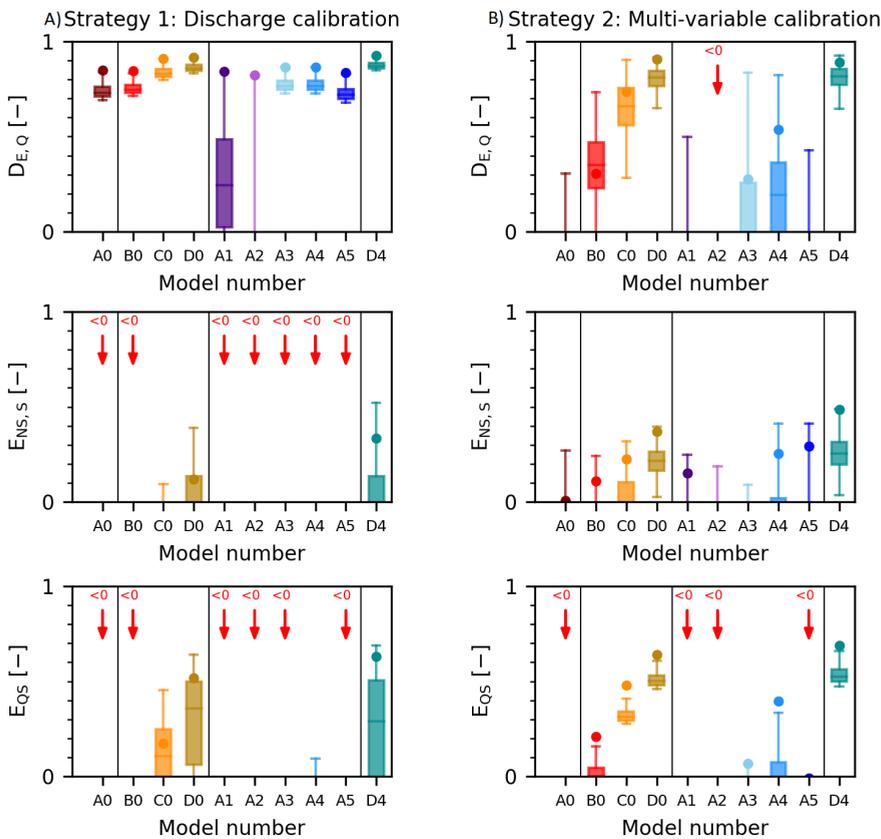


Figure 5.8: Model performance for Models A0 – D4 with respect to discharge ($D_{E,Q}$), total water storage anomalies ($D_{E,S}$) and both combined ($D_{E,QS}$). The model is calibrated with respect to A) discharge or B) both variables simultaneously. The dots represent the model performance using the “optimal” parameter set and the boxplot the range of the best 5% solutions according to $D_{E,Q}$ or $D_{E,QS}$. A red arrow was added if all solutions are below zero.

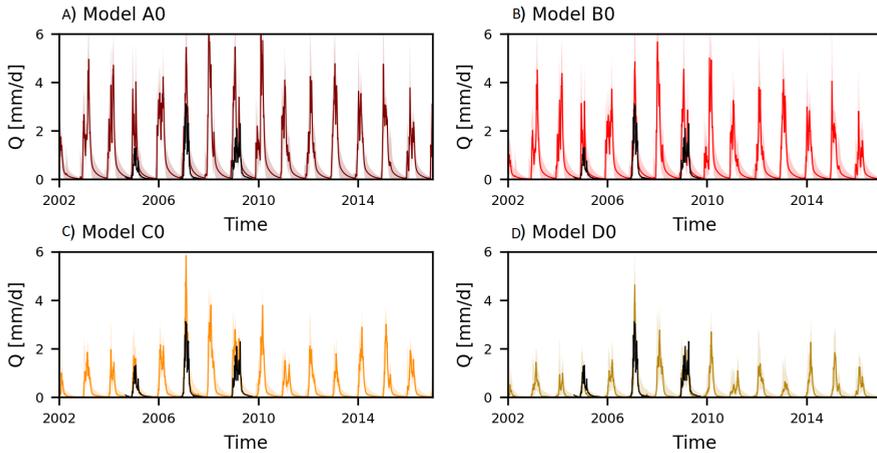


Figure 5.9: Range of model solutions for Models A0 – D0 for calibration strategy 2 with respect to discharge (hydrograph). The black line indicates the recorded data, the coloured line the solution with the highest calibration objective function with respect to discharge and total water storage ($D_{E,QS}$) and the shaded area the envelope of the solutions retained as feasible.

5

Second model adaptation: Alternative model structure (Model A1 – A5)

Following the first calibration strategy, all Models A1 – A5 reproduced the discharge well with $D_{E,Q} = 0.82 - 0.87$ (Table 5.6, Figure 5.8A). All models captured the general flow dynamics and magnitudes (Figures C.9 and C.10A in Appendix C). The monthly and annual total water storage time-series was modelled very poorly for all models ($D_{E,S} = -1066 - -9.8$, Table 5.6, Figure 5.8A). While Models A1 and A5 consistently over- or underestimated the storage with little resemblance in the fluctuations of the annual maxima ($R^2 = 0.19 - 0.22$) and minima ($R^2 = 0.08 - 0.16$), Models A2 and A3 substantially overestimated the long-term variations ($R^2 = 0.00 - 0.11$, Figures C.10 and C.11 in Appendix C). Also in Model A4, the storage was over- or underestimated, but the long-term variations improved with respect to the annual maxima ($R^2 = 0.56$) and minima ($R^2 = 0.27$). As a result, the overall model performance with respect to discharge and total water storage simultaneously improved the most for Model A4 with $D_{E,QS} = 0.32$ (Table 5.6, Figure 5.8A). Following the second calibration strategy, the modelled discharge improved considerably for Models A3 ($D_{E,Q} = 0.28$) and A4 ($D_{E,Q} = 0.54$) compared to the benchmark Model A0, but was poorly represented for the remaining models with $D_{E,Q} = -0.31 - -0.20$ (Table 5.6, Figure 5.8B). The general flow dynamics were reproduced well for Models A1 – A4 (Figure 5.11), albeit with slight differences in the timing of the wet season and dry season recession, whereas Model A5 poorly represented the recession during dry seasons. In addition, the flows were significantly over- or underestimated with Models A1 – A3 and A5 (Figure 5.12A), whereas Model A4 only slightly overestimated the high flows and underestimated the low flows. The monthly variations in the total water storage were captured well

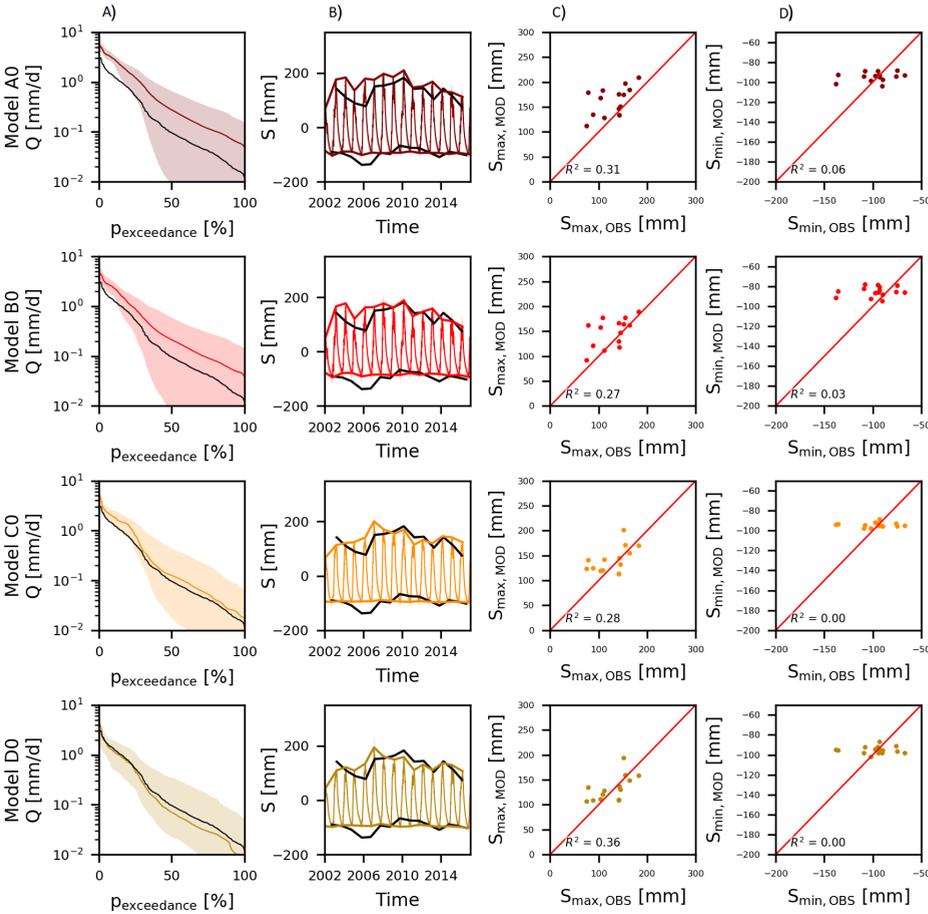


Figure 5.10: Range of model solutions for Models A0 – D0 for calibration strategy 2 with respect to A) flow duration curve in logarithmic scale, B) total water storage time-series, C) annual maximum total water storage, D) annual minimum total water storage. In A) – B), the black line indicates the recorded data, the coloured line the solution with the highest calibration objective function with respect to discharge and total water storage ($D_{E,QS}$) and the shaded area the envelope of the solutions retained as feasible. In C) – D), the recorded data are plotted on the horizontal axis and on the vertical axis the model solution with the highest calibration objective function with respect to discharge and total water storage ($D_{E,QS}$). The red line indicates the 1:1 line..

for all models with some differences in the storage decrease during dry seasons especially for Model A2 (Figure C.12 in Appendix C). While the magnitudes of the annual maxima and minima were captured well for all models (Figure 5.12B), the annual fluctuations improved the most for Model A5 with respect to the annual maxima ($R^2 = 0.51$, Figure 5.12C) and for Models A2 and A5 with respect to the annual minima ($R^2 = 0.23$, Figure 5.12D). When considering both the monthly and annual fluctuations and magnitudes, Models A4 ($D_{E,S} = 0.16$) and A5 ($D_{E,S} = 0.23$) improved the most (Table 5.6, Figure 5.8B).

As a result, the model's ability to reproduce the long-term total water storage variations during dry and wet seasons, i.e. annual minima and maxima, was significantly influenced by the model structure. The modelled annual and monthly total water storage improved the most for Models A4 and A5 (Table 5.6, Figure 5.8B) where a Deeper Groundwater reservoir was incorporated with groundwater loss/flow as function of the water content in the Deeper Groundwater reservoir. However, Model A5 only poorly captured the discharge ($D_{E,Q} = -0.31$, Figure 5.12A). Therefore, when considering the overall model performance with respect to discharge and total water storage simultaneously ($D_{E,QS}$), Model A4 performed the best with $D_{E,QS} = 0.32$ (Table 5.6, Figure 5.8B). This model captured the flows well as also the monthly and annual total water storage magnitudes and fluctuations, albeit with a slight overestimation of the annual minima and maxima in 2004 – 2006 (Figure 5.12B). These results indicated long-term storage fluctuations were most likely a result of groundwater loss from the Deeper Groundwater reservoir (Model A4).

5

Third model adaptation: Alternative forcing data and model structure

According to the first model adaptation (comparing Models A0 – D0), Model D0 performed the best using precipitation data according to TRMM and estimating the potential evaporation with the Thornthwaite method. According to the second model adaptation (comparing Models A0 – A5), Model A4 performed the best featuring a Deeper Groundwater reservoir which was only recharged during the wet season and from where groundwater leaked out of the basin (Figures 5.2 and 5.3). In this section, both models D0 and A4 were combined into Model D4 where we used TRMM as data source for precipitation, the Thornthwaite method to estimate potential evaporation and the model structure associated with Model A4.

Following the first calibration strategy, this model reproduced the discharge well (Figure C.13A in Appendix C) with $D_{E,Q} = 0.93$ which was better than all other alternative model hypotheses (Table 5.6, Figure 5.8A). Both, the general flow dynamics and magnitudes were captured well with this model (Figure C.13A, B in Appendix C). The monthly and annual total water storage improved significantly to $D_{E,S} = 0.31$ (Table 5.6, Figure 5.8A). The modelled monthly storage variations were broadly consistent with the observation (Figure C.14 in Appendix C), albeit with differences in the decrease during dry seasons and with high parameter uncertainty. The magnitudes of the annual minimum and maximum storage were modelled well for the time period 2010 – 2016, whereas before 2010 the storage was overestimated (Figure C.13C in Appendix C). Also the fluctuations in the annual maximum storage were modelled well with $R^2 = 0.48$ (Figure C.13D in Appendix C), but the annual minima were captured poorly ($R^2 = 0.19$, Figure C.13E in Appendix C). The overall model performance increased to $D_{E,QS} = 0.51$ which was better than all other alternative model hypotheses (Table 5.6, Figure 5.8A).

Following the second calibration strategy, the discharge was modelled well (Figure 5.13A), albeit with a slight decrease in the model performance ($D_{E,Q} = 0.85$) compared to the first calibration strategy (Table 5.6, Figure 5.8B). While the flow dynamics were captured well (Figure 5.13A), low flows were slightly underesti-

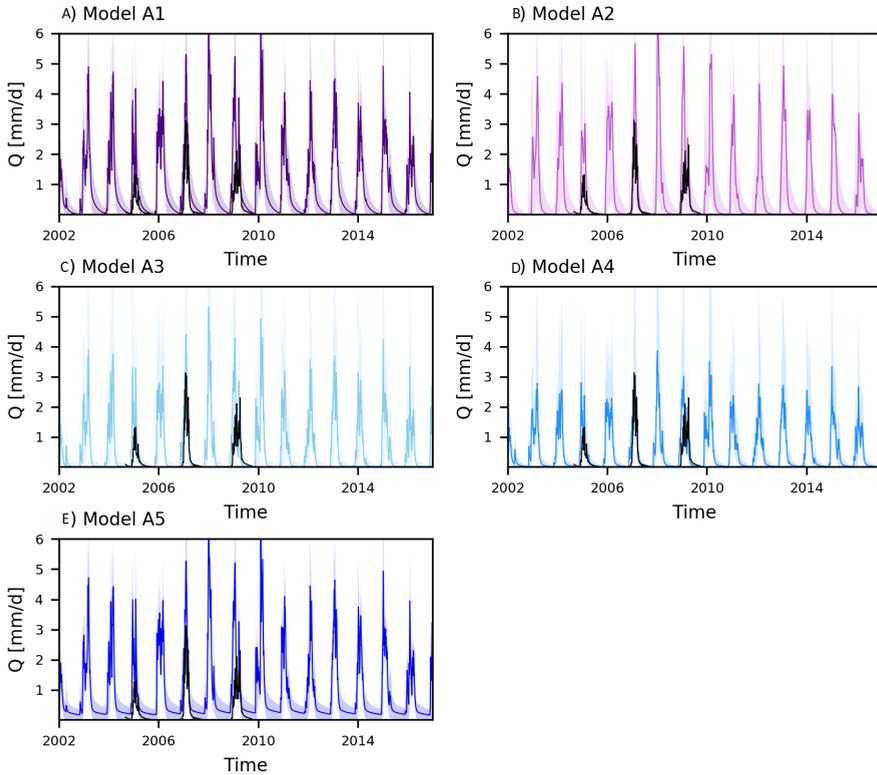


Figure 5.11: Range of model solutions for Models A1 – A5 for calibration strategy 2 with respect to discharge (hydrograph). The black line indicates the recorded data, the coloured line the solution with the highest calibration objective function with respect to discharge and total water storage ($D_{E,QS}$) and the shaded area the envelope of the solutions retained as feasible.

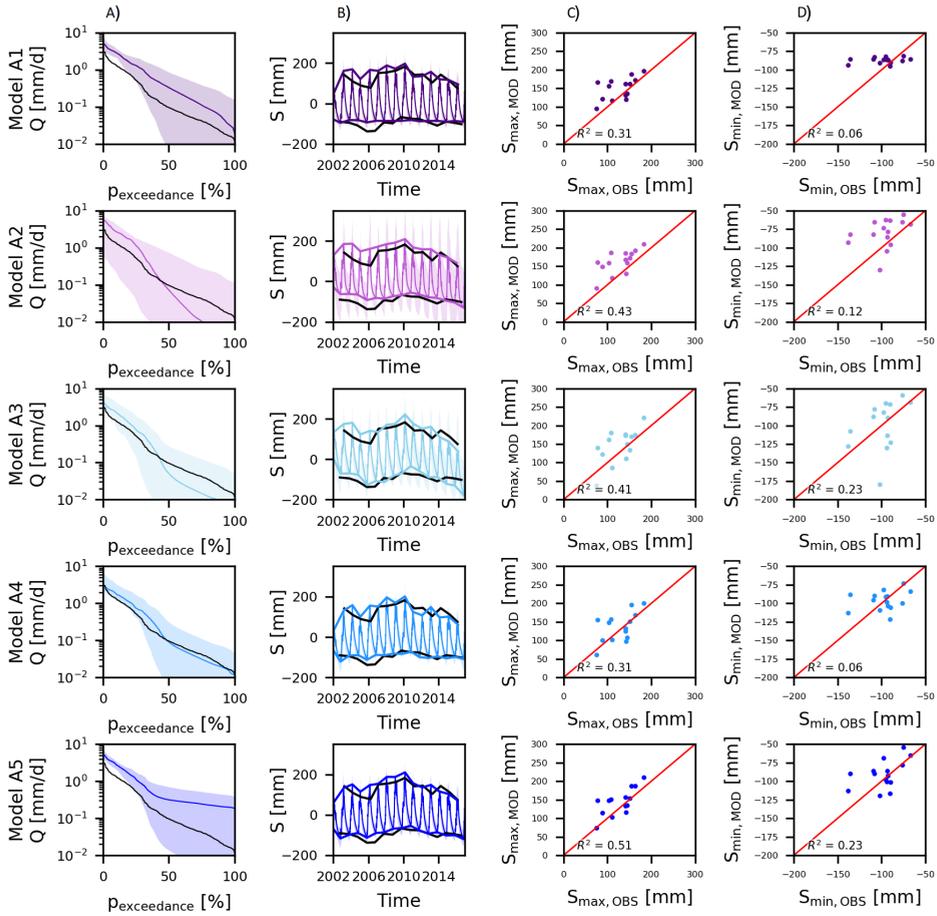


Figure 5.12: Range of model solutions for Models A1 – A5 for calibration strategy 2 with respect to A) flow duration curve in logarithmic scale, B) total water storage time-series, C) annual maximum total water storage, and D) annual minimum total water storage. In A) – B), the black line indicates the recorded data, the coloured line the solution with the highest calibration objective function with respect to discharge and total water storage ($D_{E,QS}$) and the shaded area the envelope of the solutions retained as feasible. In C) – D), the recorded data are plotted on the horizontal axis and on the vertical axis the model solution with the highest calibration objective function with respect to discharge and total water storage ($D_{E,QS}$). The red line indicates the 1:1 line.

mated (Figure 5.13B). The monthly and annual total water storage time-series improved considerably to $D_{E,S} = 0.50$ (Table 5.6, Figure 5.8B). With this model and this calibration strategy, the monthly variations were captured well (Figure C.15 in Appendix C), as also magnitudes and fluctuations in the annual maxima ($R^2 = 0.57$, Figure 5.13C,D) and minima ($R^2 = 0.41$, Figure 5.13C,E). The overall model performance increased to $D_{E,QS} = 0.63$ which was better than all other alternative

model hypotheses (Table 5.6, Figure 5.8B).

In a last step, the predictive strength of Model D4 was compared to that of the benchmark Model A0. For this purpose, both models were calibrated with respect to discharge and total water storage simultaneously (calibration strategy 2) for the time period 2002 – 2012, and post-calibration evaluated due to the lack of flow data only with respect to total water storage for the time period 2012 – 2016 (see Section 5.4.4). While the general flow dynamics were modelled well for both models (Figure C.16 in Appendix C), the magnitudes improved significantly for Model D4 as the flows were only slightly underestimated during medium flows (Figure 5.14A). Hence, the modelled flow improved from $D_{E,Q} = -0.13$ for Model A0 to $D_{E,Q} = 0.51$ for Model D4 (Table 5.7). Also the monthly and annual total water storage time-series improved for Model D4 to $D_{E,S} = 0.63$. On monthly timescale, Model D4 captured the seasonal variations better with considerable improvements in the storage decrease during dry seasons (Figure C.17 in Appendix C). While the magnitudes of the annual minima/maxima were captured well for both models (Figure 5.14B), long-term fluctuations improved for Model D4 with respect to the annual maxima ($R^2 = 0.57$, Figure 5.14C) and minima ($R^2 = 0.44$, Figure 5.14D) where R^2 corresponded to the calibration time-period 2002 – 2012 as merely four to five points were available for the evaluation time-period 2012 – 2016. With Model D4, the annual minimum and maximum storage increased before 2010 after which it decreased similar to the observations and in contrast to the benchmark Model A0. However, the annual minimum/maximum storage were frequently overestimated except in 2002 – 2004 when it was underestimated. During the evaluation time-period 2012 – 2016, the model performance with respect to the monthly and annual total water storage improved to $D_{E,S} = -1.0$ (Table 5.7) which remained negative due to the low model performance metrics with respect to the annual minima/maxima ($E_{NS,S,annual}$, Section 5.4.3). In this short time-period, the difference between the observed time-series and its mean was significantly lower compared to a longer time-period such as 2002 – 2012 resulting in a low denominator and hence a low Nash-Sutcliffe efficiency (Eq.5.39).

Table 5.7: Model performance with respect to total water storage and discharge ($D_{E,QS}$), and total water storage ($D_{E,S}$) including their 5/95% percentile ranges of the feasible parameter sets for Models A0 and D4 calibrated with respect to $D_{E,QS}$ for the time period 2002 – 2012.

	2002 – 2012		2012 – 2016	
	$D_{E,Q}$ ($D_{E,Q,5/95\%}$)	$D_{E,S}$ ($D_{E,S,5/95\%}$)	$D_{E,QS}$ ($D_{E,QS,5/95\%}$)	$D_{E,S}$ ($D_{E,S,5/95\%}$)
Model A0	-0.29 (-0.71 – -0.10)	-0.13 (-0.76 – -0.11)	-0.21 (-0.51 – -0.33)	-2.7 (-6.2 – -0.70)
Model D4	0.83 (0.62 – 0.89)	0.51 (0.08 – 0.37)	0.63 (0.33 – 0.53)	-1.0 (-3.3 – 0.43)

Overall the results suggest that the model's ability to simultaneously reproduce both the observed discharge and long-term and seasonal total water storage variations was considerably influenced by both, the choice of forcing data and model structure, respectively. Overall, the combination of TRMM data for precipitation, the Thornthwaite method for potential evaporation and the model structure associated

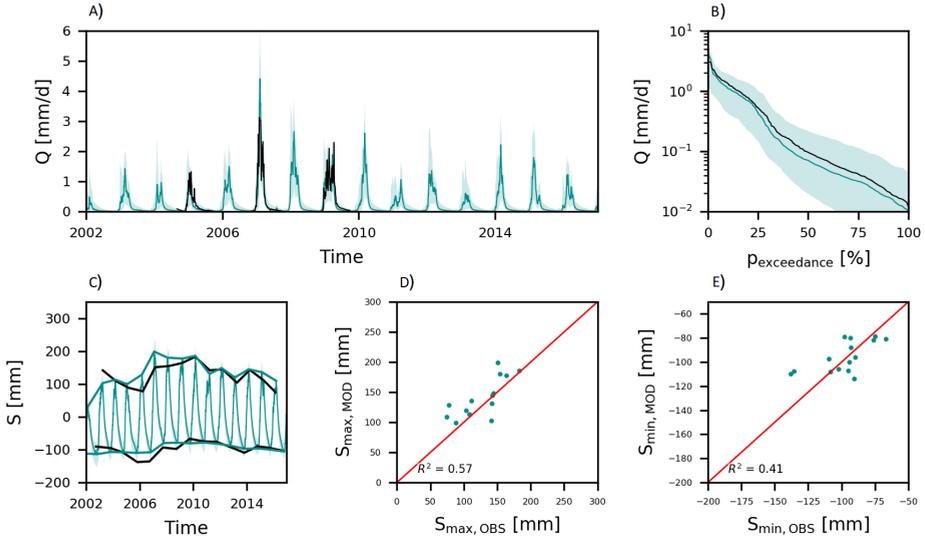


Figure 5.13: Range of model solutions for Model D4 for calibration strategy 2 with respect to A) hydrograph, B) flow duration curve in logarithmic scale, C) total water storage time-series, D) annual maximum total water storage, and E) annual minimum total water storage. In A) to C), the black line indicates the recorded data, the coloured line the solution with the highest calibration objective function with respect to discharge and total water storage ($D_{E,QS}$) and the shaded area the envelope of the solutions retained as feasible. In D) and E), the recorded data are plotted on the horizontal axis and on the vertical axis the model solution with the highest calibration objective function with respect to discharge and total water storage ($D_{E,QS}$). The red line indicates the 1:1 line.

with Model A4 here produced model results most consistent with the observed total water storage and discharge time-series. This Model D4 allowed for a better representation of the discharge and better prediction of the total water storage with respect to the seasonal and long-term fluctuations. The forcing data mostly controlled the model's ability to mimic annual storage maxima, whereas the annual storage minima improved the most when incorporating groundwater loss from the Deeper Groundwater reservoir (Model A4 and D4).

5.6. Discussion

In this study, we identified plausible drivers for the observed long-term total water storage variations in the Luangwa Basin. The results indicated modelled annual maximum storage fluctuations were to a large extent controlled by the choice of forcing data, whereas modelled annual minima were influenced by processes allowing long-term memory effects which were missing in the original benchmark Model A0. More specifically, the representation of monthly and annual total water storage fluctuations improved when using TRMM for the precipitation, the Thornthwaite method to estimate potential evaporation and incorporating groundwater loss from

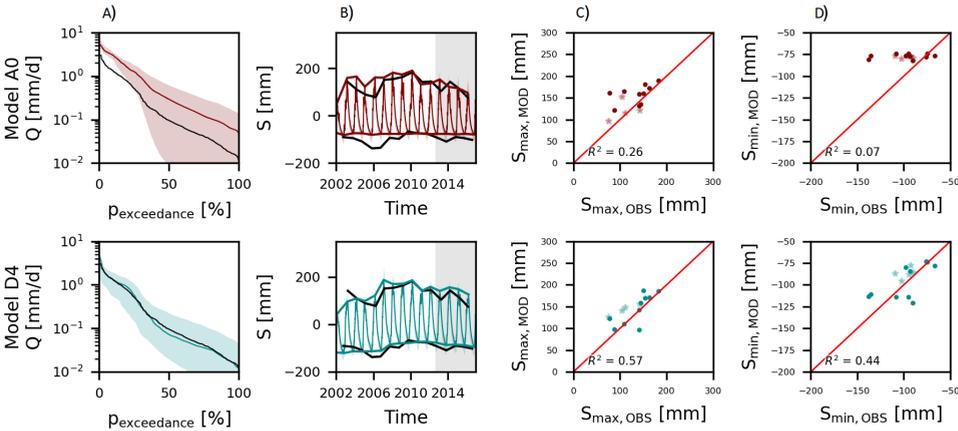


Figure 5.14: Range of model solutions for Models A0 and D4 for calibration strategy 2 with respect to A) flow duration curve in logarithmic scale, B) total water storage time-series, C) annual maximum total water storage, and D) annual minimum total water storage. In A) – B), the black line indicates the recorded data, the coloured line the solution with the highest calibration objective function with respect to discharge and total water storage ($D_{E,QS}$) and the shaded area the envelope of the solutions retained as feasible. The white area was used for calibration (2002 – 2012) and the grey area for evaluation (2012 – 2016). In C) – D), the recorded data are plotted on the horizontal axis and on the vertical axis the model solution with the highest calibration objective function with respect to discharge and total water storage ($D_{E,QS}$). The darker dots correspond to the 2002 – 2012 time-period and was used to calculate R^2 , whereas the lighter stars correspond to the 2012 – 2016 time-period. The red line indicates the 1:1 line.

a deeper groundwater layer (Model D4).

The results demonstrated that models that can adequately reproduce discharge do not necessarily reproduce storage well which was also observed by Bouaziz et al. (2020). In this study, the benchmark Model A0 reproduced the general dynamics and magnitudes of the discharge well but did not reproduce the observed storage magnitudes nor the long-term storage fluctuations. Incorporating the total water storage in the calibration procedure only improved the modelled storage magnitudes, but not the long-term fluctuations. While alternative forcing data sources improved the representation of the annual maximum storage fluctuations, the storage conditions during dry seasons, i.e. annual minima, remained poorly represented (Models A0 – D0) and only improved after modifying the model structure (Model D4). These results suggested that groundwater loss from the Luangwa basin played an important role to explain long-term annual storage variations. However, in many commonly used hydrological models such processes allowing long-term memory effects are missing (e.g. Bergström, 1992; Fenicia et al., 2014; Liang et al., 1994) resulting in biased predictions of discharge and storage which is especially crucial during extreme dry conditions (Fowler et al., 2020; Saft et al., 2016).

Furthermore, this study showed that processes allowing for long-term memory effects can be incorporated in conceptual hydrological models. In this study, several

model hypotheses were tested to assess which processes most likely dominated long-term memory effects in the Luangwa basin (Models A1 – A5). The results suggested long-term storage variations were a result of groundwater loss from a deeper groundwater layer which was only recharged during wet seasons (Model D4). With this model, the storage prediction substantially improved compared to the benchmark Model A0, yet remained at a modest level ($D_{E,S} < 0$, Table 5.7) most likely due to the chosen model performance metric and the limited number of data points for the evaluation when considering annual minima/maxima for the time-period 2012 – 2016 as explained in the previous section. In addition, these modifications also improved the modelled discharge time-series such that the general dynamics and magnitudes were represented better with Model D4 (Figure 5.13) compared to the benchmark Model A0 (Figure 5.7). Therefore, model hypothesis testing played a crucial role in improving the representation of real world processes to reproduce multiple variables simultaneously (Beven, 2018; Clark et al., 2011). Previous studies highlighted the inability of many conceptual models to reproduce long-term storage variations and attributed this to data errors, poor parameterization, model structural deficiencies or a combination thereof (Fowler et al., 2018; Jing et al., 2019; Saft et al., 2016; Scanlon et al., 2018; Winsemius et al., 2006b). Fowler et al. (2020) recently demonstrated that commonly used conceptual hydrological models cannot reproduce long-term storage variations as they lack long-term memory processes and hence should not be used for discharge predictions in for example drying climates. However, here we could show that following a careful data and model selection procedure, the representation of long-term storage variations in a conceptual model could be considerably improved. This further implies that although many typical implementations of hydrological models indeed cannot reproduce long-term storage changes, in particular with respect to annual fluctuations in dry season conditions, i.e. annual minima, as shown by Fowler et al. (2020) and here with Models A0 – D0, this inability is not an inherent property of conceptual models. Instead, our results provide some evidence that this inability can, at least to some degree, be overcome when adopting a systematic procedure to test alternative model hypotheses and thus to improve the representation of real world processes (here: Models A1 – A5).

For future studies, it will be interesting to explore the effects of evaporation on long-term storage fluctuations in a more detailed analysis. Our results suggest that long-term fluctuations in the potential evaporation can occur depending on the chosen estimation method (Hobbins et al., 2008; Huang et al., 2015; Roderick and Farquhar, 2005; Xu et al., 2018). It would therefore be interesting to look into alternative, potentially more accurate estimation methods. In addition, long-term fluctuations in the actual evaporation were observed depending on the satellite product due to the different underlying assumptions and input data (Bai et al., 2019; Feng et al., 2019; Goroshi et al., 2017; Wang et al., 2018). That is why, more in depth analyses on the effect of evaporation on long-term storage fluctuations is recommended which was outside the scope of this study due to the limited data availability.

5.7. Evaluation

In the Luangwa basin, long-term total water storage variations were observed with GRACE, but not reproduced by a previously developed process-based hydrological model that encapsulates our current understanding of the dominant regional hydrological processes. The objective of this study was to identify potential and so far overlooked processes underlying these low-frequency variations in a combined data analysis and model hypothesis testing approach. Overall, the results suggest that the initial model's inability to reproduce the observed low-frequency storage variations was a combined effect of the data source chosen to run the model and the missing representation of regional groundwater export. More specifically, it was shown that a different choice of the model input data source produced model results that are more consistent with observed fluctuations in long-term annual total water storage maxima. In contrast, the incorporation of a process representing regional groundwater export from a deep groundwater layer improved the model's ability to reproduce the observed variations in the annual minimum storage. The results highlighted the combined value of alternative data sources of multiple variables and iterative hypotheses testing to improve our understanding of hydrological processes, their quantitative description in models and eventually towards more reliable predictions of hydrological models.

6

Sub-Saharan drought of 2019 observed from space

Water doesn't stay in the sky forever.

African proverb

In the previous chapters, satellite observations were used to improve our current understanding of local dominant hydrological processes, improve hydrological model structures, and improve the identification of feasible parameter sets in a data-scarce region. In addition, satellite-based data analysis results provide valuable new information to improve our understanding of the hydrological system which can then be implemented in hydrological models. For example, satellite observations can be used to analyse drought events in large regions with respect to the temporal and spatial variability. In general, drought events can influence human activities with respect to drought coping mechanisms, depending on how they perceived specific drought events, which can affect the hydrological system. That is why it would be interesting to compare local perceptions with multiple satellite observations to increase our understanding of factors influencing local perceptions. As small illustration, the drought of 2019 in the Zambezi River Basin was analysed using multiple satellite observations to determine whether it was, as locally perceived, indeed the most extreme drought over at least 20 years. Data analysis indicates that it depended on the drought characteristic, the hydrological variable considered, and on the location within the basin.

This chapter is based on: Hulsman, P., Savenije, H. H. G., and Hrachowitz, M.: Zambezi River Basin: Drought of 2019, Journal of Hydrology: Regional Studies, submitted.

6.1. Introduction

During the dry season of 2019 in Sub-Saharan Africa, extreme low river water levels were observed which was especially visible in the Zambezi River at Victoria Falls and the reservoir upstream of the Kariba hydropower dam. Extreme low water levels were observed at the Kariba reservoir, down to 10% of usable water for hydro-power generation, which resulted in frequent power cuts of up to 18 hours per day for at least 3 months starting in November 2020 according to locals and news media (Carlowicz, 2019; Matiashe, 2019; Tshili, 2019). 250 km further upstream of the Kariba reservoir, the Victoria Falls which is known as one of the biggest waterfalls in the world, dried significantly reducing the 1.7 km wide falls to multiple small waterfalls (Childs, 2019; Henson, 2019).

As the Zambezi river basin is characterized by one distinct wet season, it exhibits high temporal but also spatial variability in water availability and demand such that the dry season demand frequently exceeds water availability, resulting in water stressed areas. In the past, severe droughts have occurred for example in 1995 and 2015 in multiple Sub-Saharan countries causing reduced crop production (Libanda et al., 2019; The World Bank, 2010). Between 2000 and 2009, about 12.5 million people were affected by droughts in Mozambique, Zambia and Zimbabwe (ZAMCOM et al., 2015). However, according to popular news media, locals perceived the drought in 2019 as the worst in several decades (Brown, 2019; Edel, 2019) or even a century (Carlowicz, 2019; Henson, 2019).

Drought is one of the most damaging natural disasters throughout the globe with widespread impacts on the society, economy, and ecology (Mishra and Singh, 2010; Van Loon, 2015). In general, droughts are classified among others into meteorological droughts related to precipitation deficits, hydrological droughts related to water deficits on the (sub-) surface and groundwater droughts related to decreased groundwater levels (Mishra and Singh, 2010; Van Loon et al., 2014; Wilhite and Glantz, 1985). Many previous studies have focused for example on 1) quantifying and predicting droughts in terms of for example intensity, duration, severity or spatial extent using drought indices (Bayissa et al., 2018; Hao and Singh, 2015; Hellwig and Stahl, 2018; Kumar et al., 2016; Naresh Kumar et al., 2009; Van Loon et al., 2017), 2) analysing the impact of drought on different sectors such as society, economy or ecology (Haile et al., 2019; He et al., 2019; Mishra and Singh, 2010; Stahl et al., 2016), and 3) analysing factors influencing droughts such as climate-change, human modifications or catchment characteristics (Firoz et al., 2017; Haile et al., 2019; Roodari et al., 2020; Van Loon and Laaha, 2015; Van Loon et al., 2016). Some of these studies focused on analysing droughts in Sub-Saharan regions such as the Zambezi river basin (e.g. Dutra et al., 2013; Thomas et al., 2014; Tirivarombo and Hughes, 2011).

In addition, previous studies illustrated discrepancies between people's perception of dry conditions and data analyses results (e.g. Foguesatto et al., 2020; Simelton et al., 2013; Solano-Hernandez et al., 2020). For instance, Foguesatto et al. (2020) showed multiple farmers in Africa and Asia perceived decreased rainfall amounts which was not reflected in meteorological records. They argued these discrepancies can often be a result of economic and psychological stress factors. In another

study, Taylor et al. (1988) showed farmers in the United States remembered the most recent drought and individual extreme drought events, but largely forget intermediate droughts. Similarly, Di Baldassarre et al. (2017) and Albertini et al. (2020) provided evidence that the impacts of droughts occurring after extreme flood events can be intensified as it is often not expected nor prepared for. In several Sub-Saharan countries, there was an extreme flood event due to the tropical cyclones Idai and Kenneth in March and April 2019 (United Nations Office for the Coordination of Humanitarian Affairs (OCHA), 2019) which thus may have impacted the local perception of the 2019 drought in the Zambezi River Basin.

While many previous studies compared local perceptions of drought events to rainfall observations (e.g. Giordano et al., 2013; Iqbal et al., 2018; Osgood et al., 2018; Ovuka and Lindqvist, 2000; Solano-Hernandez et al., 2020), these studies did not incorporate satellite-based total water storage and reservoir water level observations which provide additional information on drought events. In addition, the drought of 2019 in the Zambezi River Basin has not yet been fully analysed as it occurred recently. Therefore, the objective of this study was to analyse the drought of 2019 in the Zambezi River Basin upstream of the Kariba reservoir using multiple satellite observations to determine whether it was indeed the most extreme drought in at least 20 years as perceived by locals. More specifically, we tested the research hypothesis that the drought of 2019 was the longest, most intense and severe drought, basin-wide and locally, according to satellite-based precipitation, total water storage and reservoir water level observations since at least 20 years similar to the local perception.

6.2. Site description

The Zambezi river, the fourth longest river in Africa, is shared by the countries Zambia (42%), Angola (18%), Zimbabwe (16%), Mozambique (12%), Malawi (7.5%), Tanzania (2%), Botswana (1.5%) and Namibia (1%) (Figure 6.1; Kling et al., 2014). The river has a basin area of 1.37 million km², is about 2,660 km long and has an average discharge of 4,134 m³ s⁻¹ at the outlet in Mozambique (The World Bank, 2010). There is a distinct wet season from September – April and dry season from May – August. In this semi-arid basin, the potential evaporation (2000 mm yr⁻¹) exceeds the precipitation (1000 mm yr⁻¹) especially during dry seasons (Schleiss and Matos, 2016). Two large hydropower dams are located in the main Zambezi river which are the Cahora Bassa Dam (2075 MW) in Mozambique and further upstream the Kariba Dam (2130 MW) at the border of Zambia and Zimbabwe which is one of the main power sources for both countries according to the local power supply companies (Kesselring, 2017). The maximum water depth above the minimum operating level in the Kariba reservoir is 13 m according to the Zambezi River Authority (<http://www.zambeziriver.org/>).

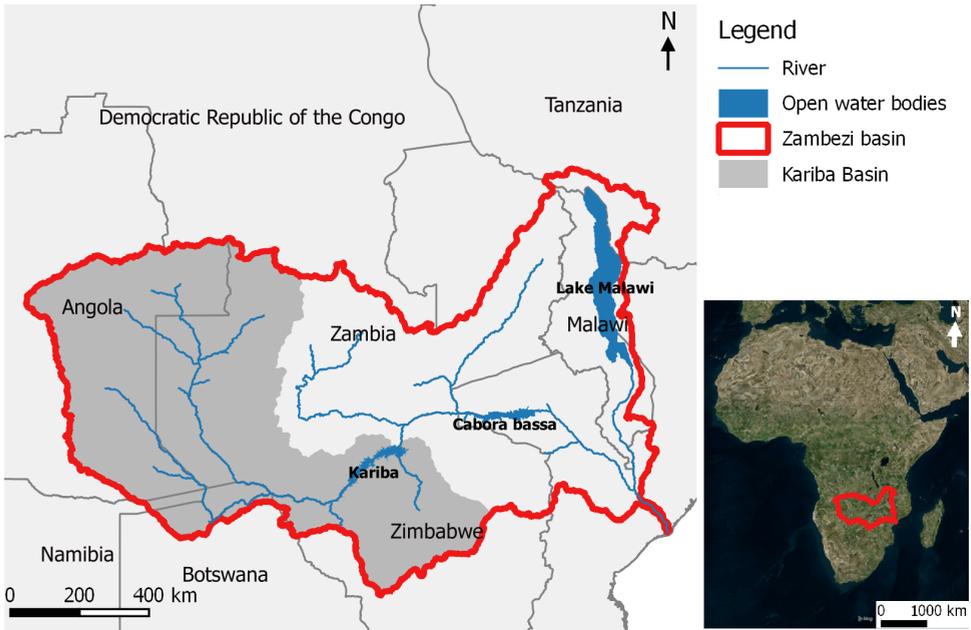


Figure 6.1: Map of the Zambezi river basin and the basin area upstream of the Kariba reservoir now called "Kariba basin".

6

6.3. Data availability

In this study, satellite observations were used to estimate precipitation, total water storage anomalies, actual and potential evaporation, and reservoir water levels (Table 6.1) as the available ground observations were very limited within the Zambezi river basin (e.g. Hulsman et al., 2020a). Briefly, monthly precipitation data was obtained from CHIRPS (Climate Hazards Group InfraRed Precipitation with Station data) and basin-averaged monthly actual evaporation from WaPOR (Water Productivity Open-access portal). Monthly satellite-based temperature data obtained from ERA5 (5th generation ECMWF atmospheric reanalysis dataset) and averaged over the Kariba reservoir were used to estimate the potential evaporation with the Hargreaves method (Hargreaves and Allen, 2003; Hargreaves and Samani, 1985) which was assumed to be equal to the total actual evaporation from this open water body.

Total water storage anomaly observations were obtained from the Gravity Recovery and Climate Experiment (GRACE). GRACE observations describe variations in the Earth's gravity field which are related to regional mass changes that are dominated by terrestrial water storage variations (Landerer and Swenson, 2012; Swenson, 2012). Readily available and pre-processed GRACE observations were generated by JPL (Jet Propulsion Laboratory) and downloaded from the GRACE Tellus website (<https://grace.jpl.nasa.gov/>). The JPL pre-processed the data to remove atmospheric mass changes, systematic errors and noise, and subtract the 2004 –

2009 time-mean baseline to obtain total water storage anomalies (Landerer and Swenson, 2012; Swenson and Wahr, 2006; Wahr et al., 1998).

Satellite-based lake water levels, i.e. altimetry observations, were extracted from the platform DAHITI (Database for Hydrological Time Series of Inland Waters) for the Kariba reservoir. Altimetry observations estimate the water level relative to a reference ellipsoid. The distance between the satellite and earth surface is estimated by sending a radar signal in nadir direction towards the earth and measuring the time difference between sending and receiving the reflected signal. With this distance and the known satellite position, the surface level relative to a reference ellipsoid is estimated (Calmant et al., 2009; Łyszkowicz and Bernatowicz, 2017). At the Kariba reservoir, the following satellite missions were used to create a time-series: Envisat, Jason 1 – 3, and Topex/Poseidon (Schwatke et al., 2015). Depending on the satellite mission, the temporal resolution is 10 and 35 days (CNES, Accessed 2018; ESA, Accessed 2018; Schwatke et al., 2015).

Table 6.1: Data used in this study

	Time period	Time resolution	Spatial resolution	Product name	Source/Reference
Precipitation	1992 – 2020	Daily	0.05°	CHIRPS	Version 2 (Funk et al., 2014)
Total water storage	2002 – 2020	Monthly	1°	GRACE	Pre-processed by JPL (Version RL05_1.DSTVSCS1411) https://grace.jpl.nasa.gov/ (Landerer and Swenson, 2012; Swenson, 2012; Swenson and Wahr, 2006)
Temperature	1992 – 2020	Monthly	0.25°	ERA5	5 th generation ECMWF atmospheric reanalysis dataset (Copernicus Climate Change Service (C3S), 2017)
Evaporation	2009 – 2020	Monthly	250 m	WaPOR	WaPOR V2 Level 1 (FAO, 2018; FAO and IHE Delft, 2019)
Altimetry	1992 – 2020	10 – 35 days	n/a	DAHITI	https://dahiti.dgfi.tum.de/en/ (Schwatke et al., 2015)

6.4. Approach

In this study, the temporal variability of remotely-sensed precipitation, total water storage, actual and potential evaporation as well as reservoir water levels at Kariba were compared with each other for the time period 1992 – 2020. As GRACE and WaPOR data were only available since 2002 and 2009 respectively, any comparison with these variables comprised the time-period 2002 – 2020 and 2009 – 2020, respectively. Annual values for the precipitation and evaporation were calculated from monthly observations considering hydrological years starting in August, rather than calendar years. As there were no discharge data available in the vicinity of the

Kariba reservoir after 2018 in the global discharge database GRDC (Global Runoff Data Centre), this variable was not included in this study.

The temporal variability of all individual variables were analysed and compared using basin-averaged time-series for the gridded observations considering the basin upstream of the Kariba hydropower dam (dark grey area in Figure 6.1). In addition, multiple drought indices were calculated using precipitation, GRACE and altimetry data as explained in Section 6.4.1. With these indices, the associated drought severities, durations and intensities were estimated and analysed as explained in Section 6.4.2. Then, the spatial variability of the drought severity and intensity was compared for the average, most severe and the drought of 2019 with respect to precipitation and GRACE for which gridded data was available.

6.4.1. Drought indices

Remotely-sensed precipitation data was used to estimate the Standardized Precipitation Index (SPI; McKee et al., 1993) which is typically used to quantify meteorological droughts. The SPI was calculated for each month using monthly precipitation time-series accumulated over the 12 preceding months to reflect short- and long-term effects. This time-series was fitted to the Gamma probability density function to compute the corresponding cumulative distribution function which was then transformed to a normal distribution function to estimate SPI values.

Total water storage observations according to GRACE were used to estimate the GRACE-based Total Storage Deficit Index (TSDI; Nie et al., 2018). This was calculated by first estimating the Total Storage Deficit (TSD) using Eq.6.1 to remove seasonal variations to allow comparisons of TSD between different seasons, and then standardizing TSD using Eq.6.2 to obtain TSDI (Nie et al., 2018):

$$TSD_{ij} = \frac{S_{ij} - S_{avg,j}}{S_{max,j} - S_{min,j}} \quad (6.1)$$

$$TSDI_{ij} = \frac{TSD_{ij} - \mu}{\sigma} \quad (6.2)$$

With S_{ij} [mm] total water storage in month j and year i , $S_{avg,j}$ [mm] long-term mean total water storage for month j , $S_{max,j}$ [mm] long-term maximum total water storage for month j , $S_{min,j}$ [mm] long-term minimum total water storage for month j , μ [-] mean of TSD [-] and σ [-] standard deviation of TSD.

Reservoir altimetry observations were used to estimate the corresponding drought index which was called Water-Level Deficit Index (WLDI) in this study. For this purpose, Eq.6.1 and 6.2 were used similar to the calculation of GRACE-based TSDI as reservoir water level and total water storage observations both already account for effects of preceding months. That is why the altimetry and GRACE time-series were not accumulated when calculating the corresponding drought indices as done when estimating SPI (Bloomfield and Marchant, 2013; Van Loon et al., 2017).

6.4.2. Drought characteristics

The three individual drought indices were used to characterise droughts in terms of drought severity, duration and intensity. The drought duration $D_{D,\theta}$ [months] for drought index θ was defined as the number of consecutive months with θ below zero, drought intensity $D_{I,\theta}$ [month⁻¹] as average θ over consecutive months with θ below zero, and drought severity $D_{S,\theta}$ [-] as the accumulated θ over consecutive months with θ below zero (Huang et al., 2016). This was applied for all drought indices used in this study such that $\theta = \text{SPI, TSDI or WLDI}$.

6.5. Results

6.5.1. Precipitation

The annual precipitation varied between 642 mm yr⁻¹ in 2019, and 1024 mm yr⁻¹ in 2008 (Figure 6.2A). In other words, in 2019 the annual rainfall was lowest since at least 27 years which was the duration of the available rainfall data. Considerably low rainfall amounts were also observed in 1995 (670 mm yr⁻¹), 2002 (728 mm yr⁻¹), 2005 (710 mm yr⁻¹) and 2015 (733 mm yr⁻¹). Consecutive wet years with above average rainfall amounts (871 mm yr⁻¹) were observed in 2006 – 2011 and 2017 – 2018, whereas in 1997 – 2001 the annual rainfall amounts were more than 800 mm yr⁻¹ (Figure 6.2A).

According to the precipitation-based drought index SPI, there have been multiple droughts in the Kariba basin (coloured areas in Figure 6.2B) of which five droughts are characterized by $\text{SPI}_{\min} < -1.5$ (Table 6.2) and can hence be classified as “severe meteorological drought” according to the drought categories as defined by McKee et al. (1993). These droughts started in 1994, 2002, 2005, 2015 and 2019. Their drought severities varied between $D_{S,\text{SPI}} = -11.6 - -29.5$, their durations between $D_{D,\text{SPI}} = 8 - 22$ months, and their intensities between $D_{I,\text{SPI}} = -1.0 - -1.5$ month⁻¹ (Table 6.2). Based on these results, the longest and most severe meteorological drought occurred in 1994 ($D_{D,\text{SPI}} = 22$ months, $D_{S,\text{SPI}} = -29.5$), whereas the most intense drought was in 2005 ($D_{I,\text{SPI}} = -1.5$ month⁻¹, Table 6.2). In contrast to the local perception, the meteorological drought of 2019 was “only” the most severe in 17 years since the drought of 2002 was more severe and longer (Table 6.2). Similarly, the 2005 drought was more intense than the 2019 drought. This largely contradicts the local perception that the drought of 2019 was the most extreme in more than 20 years.

6.5.2. Total water storage

The total water storage varied seasonally on average 271 mm, whereas in the long-term the annual mean storage varied between -115 mm and 143 mm (Figure 6.2C). In 2005 and 2019, the total water storage decreased to -217 mm and -215 mm, respectively, which were both dry years when considering the annual rainfall (710 mm yr⁻¹ and 642 mm yr⁻¹, respectively). Besides the low rainfall amounts, the decreased storage in 2019 is likely also a result of the above average evaporation in the basin (1030 mm yr⁻¹, Figure 6.3A). In contrast, the total water storage remained relatively high in 2015 despite the low rainfall (733 mm yr⁻¹) with an

Table 6.2: Overview of drought characteristics according to precipitation, GRACE and altimetry data including minimum Deficit Index (DI_{\min}), duration (D_D), intensity (D_I) and severity (D_S) for all droughts with $D_D > 6$ months. Bold numbers indicate droughts with $DI_{\min} \leq -1.5$. The data for the GRACE and water level based deficit in 2019 was incomplete as it continued after June 2020, the end of the time-series, since the corresponding deficit index remained below zero until the end of the time-series.

	Start of drought	End of drought	DI_{\min} [-]	D_D [months]	D_I [month ⁻¹]	D_S [-]	
Precipitation	Feb 1992	Dec 1992	-1.4	11	-0.97	-10.6	
	Apr 1994	Jan 1996	-1.6	22	-1.3	-29.5	
	Mar 1996	Nov 1996	-1.3	9	-1.1	-9.8	
	Apr 1997	Sep 1997	-0.3	6	-0.3	-1.6	
	Apr 1998	Nov 1998	-1.1	8	-0.9	-7.2	
	Apr 2000	Oct 2000	-0.7	7	-0.6	-4.4	
	Mar 2002	Nov 2003	-1.5	21	-1.0	-21.8	
	Apr 2005	Nov 2005	-1.5	8	-1.5	-11.6	
	Apr 2007	Sep 2007	-0.4	6	-0.4	-2.1	
	Apr 2012	Oct 2012	-0.9	7	-0.7	-5.1	
	Apr 2013	Oct 2013	-0.6	7	-0.5	-3.5	
	Mar 2015	Feb 2016	-1.5	12	-1.0	-12.5	
	Apr 2016	Nov 2016	-1.0	8	-0.7	-5.8	
	Mar 2019	Feb 2020	-1.7	12	-1.3	-15.7	
	GRACE	Apr 2002	Mar 2004	-1.4	24	-0.8	-16.8
		Aug 2004	Dec 2007	-1.8	41	-0.8	-32.9
Sep 2015		Dec 2016	-1.2	16	-0.6	-6.4	
Dec 2018		Jun 2020*	-1.8	19	-1.3	-23.7	
Altimetry	Sep 1992	May 1998	-2.0	69	-1.2	-79.6	
	May 2005	Jan 2008	-1.2	33	-0.7	-22.4	
	Apr 2015	Feb 2018	-1.6	35	-0.7	-25.3	
	Feb 2019	Jun 2020*	-1.7	17	-1.2	-20.1	

annual minimum storage of -80 mm (Figure 6.2A and C). This can plausibly be linked to the below average actual evaporation in the basin (902 mm yr^{-1} , Figure 6.3A) in that year, resulting in less water being released from the (sub-) surface and hence higher total water storage values.

According to the GRACE-based TSDI, water storage deficits occurred multiple times of which two exhibit $TSDI_{\min} < -1.5$ which started in 2004 and 2018 (Table 6.2 and Figure 6.2D), respectively. Their drought severities varied between $D_{S,TSDI} = -23.7 - -32.9$, their durations between $D_{D,TSDI} = 19 - 41$ months, and their intensities between $D_{I,TSDI} = -0.8 - -1.3 \text{ month}^{-1}$. In contrast to these droughts, the low rainfall in 2015 did not result in $TSDI < -1.5$ which corresponded to the observations in the total water storage (Figure 6.2D) as described above. Based on these results, the longest and most severe groundwater drought occurred in 2004 ($D_{D,TSDI} = 41$ months, $D_{S,TSDI} = -32.9$), whereas the most intense total storage related drought occurred indeed in 2019 ($D_{I,TSDI} = -1.3 \text{ month}^{-1}$, Table 6.2) which so far lasted for 19 months. Note, the TSDI remained below zero towards the end of the available time-series in June 2020 (Figure 6.2D) meaning the total storage deficit was still ongoing, thereby affecting the final duration, severity and intensity for this specific drought.

In contrast to the local perception, the GRACE-based drought of 2019 was not the most severe in several decades, since the drought of 2004 was more severe and

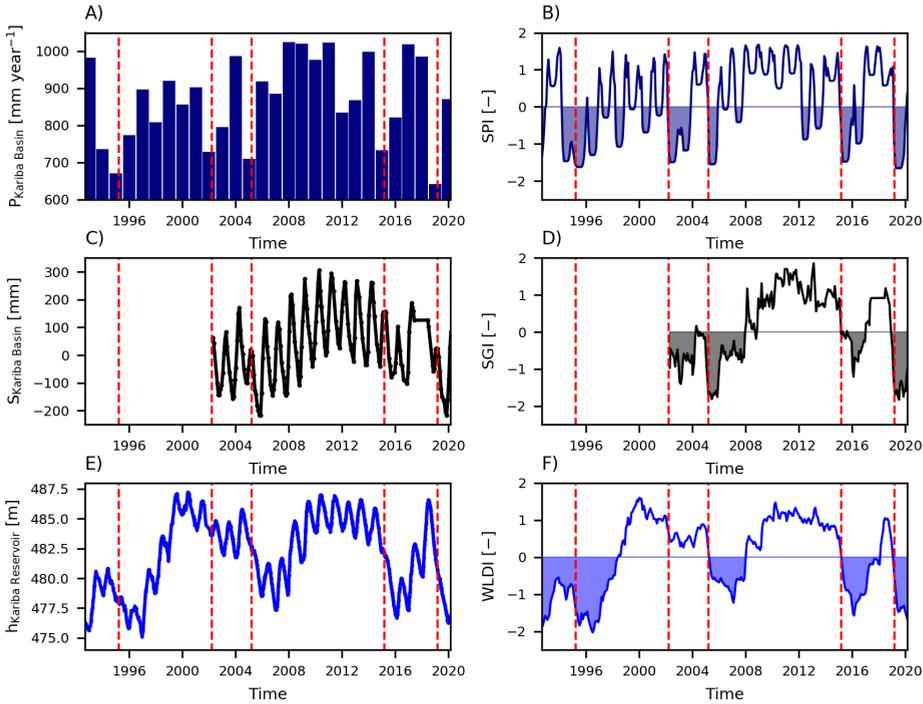


Figure 6.2: A) Basin-averaged annual precipitation, B) monthly SPI (Standardized Precipitation Index), C) basin-averaged monthly total water storage anomalies, D) monthly TSDI (Total Storage Deficit Index), E) altimetry observations at Kariba reservoir, and F) monthly WLDI (Water-Level Deficit Index). The red dashed lines mark the dates 15 March 1995, 2002, 2005, 2015 and 2019.

longer (Table 6.2). On the other hand, the drought of 2019 was the most intense according to GRACE data. However, as the available data did not cover the entire drought of 2019, its final duration and severity remain unknown. Thus, as of now there is insufficient data for a conclusive assessment of the local perception that the storage-related drought of 2019 was the most severe in more than 20 years as this depends on how this drought evolved after June 2020.

6.5.3. Reservoir water level

As illustrated in Figure 6.2E, the water level in Kariba reservoir changed both seasonally and inter-annually. While the seasonal variability was on average 2.8 m, the inter-annual variability ranged up to 10.2 m as the annual minima ranged between 475.1 m (1996) and 485.3 m (1999). The highest level was observed in 2000 (487.2 m) and the lowest in 1996 (475.1 m), whereas during the dry season of 2019, the reservoir level decreased to 476.2 m before increasing again with the new rains. In other words, the lowest reservoir level was observed in 1996 instead of 2019 also when considering uncertainties in altimetry observations which typically range between 4 – 36 cm for open water bodies according to previous studies (Crétau

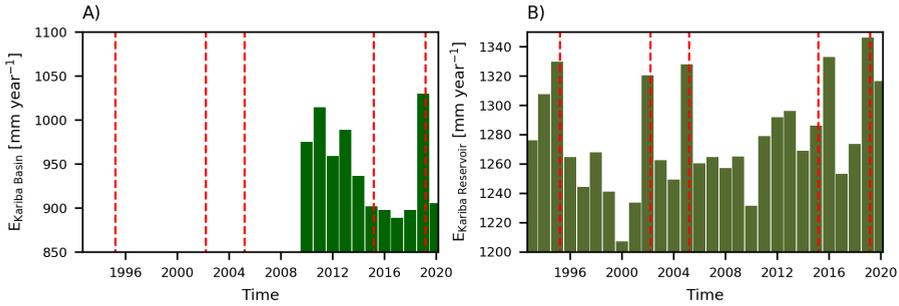


Figure 6.3: Annual total evaporation from A) Kariba basin and B) Kariba reservoir. The red dashed lines mark the dates 15 March 1995, 2002, 2005, 2015 and 2019.

and Birkett, 2006; Schwatke et al., 2015). The long-term variations in the altimetry often coincided with the annual precipitation amounts according to CHIRPS (Figure 6.2A). For example, the water levels decreased significantly in 1995, 2005, 2015 and 2019 when the annual rainfall was significantly low. Similarly, the water levels increased in 1997 – 1999, 2007 – 2011 and 2016 – 2018 during consecutively wet years (Figure 6.2A and E). In contrast, the annual rainfall was significantly lower in 2002 (728 mm yr⁻¹) which was not reflected as strong in the altimetry observations compared to the other dry years (Figure 6.2A and E). In general, high reservoir water levels despite low rainfall amounts can be due to decreased outflow as a consequence of reservoir operation considerations or decreased evaporation from the open water body. However, in 2002, the evaporation was considerably high (1320 mm yr⁻¹) compared to preceding and following years (Figure 6.3B). Therefore, the high altimetry observations in 2002, despite the low rainfall amount, were more likely a result of reservoir operation decisions.

According to the reservoir altimetry-based WLDI, significant water deficits continuing for more than a year started in 1992, 2005, 2015 and 2019 (coloured areas in Figure 6.2F). Their severities varied between $D_{S,WLDI} = -20.1 - -79.6$, their durations between $D_{D,WLDI} = 17 - 69$ months, and their intensities between $D_{I,WLDI} = -0.7 - -1.2$ month⁻¹ (Table 6.2). Based on these results, the longest, most severe and most intense deficit occurred between 1992 and 1998 ($D_{D,WLDI} = 69$ months, $D_{S,WLDI} = -79.6$, $D_{I,WLDI} = -1.2$ month⁻¹, Table 6.2). Note, the WLDI remained below zero towards the end of the time-series in June 2020 (Figure 6.2F) meaning the reservoir water deficit was still ongoing. Duration and severity of this specific drought as reported here are thus to be interpreted as lower bounds.

In contrast to the local perception, the altimetry-based drought of 2019 was not the most severe in the last decades, but only the most severe in four years since the drought of 2015 was more severe and longer (Table 6.2). However, as the available data did not cover the entire drought of 2019, its final duration and severity remain unknown. On the other hand, the drought of 2019 was the most intense since 1992 and resulted in the lowest reservoir water level since 1995 (475.8 m) which supports the local perception that the drought of 2019 was the most extreme in more than 20 years.

6.5.4. Spatial variability

In the previous sections, the basin-averaged drought characteristics were analysed using precipitation, GRACE and altimetry data. In this section, the spatial variability of the drought severity and intensity was compared for all droughts on average, the most severe drought and the 2019 drought with respect to precipitation and GRACE for which gridded data was available. The most severe basin-average precipitation deficit occurred in 1994 while the most severe total water storage deficit occurred in 2004 (Table 6.2).

Precipitation

The precipitation-based drought intensity averaged for all droughts in 2002 – 2020 did not change significantly in space with $D_{I,SPI} = -0.6 - -1.3 \text{ month}^{-1}$ (Figure 6.4A). In 1994, the drought intensity varied between $D_{I,SPI} = -0.5 - -1.6 \text{ month}^{-1}$ and also exhibited pronounced spatial contrasts with the lowest values in the central parts of the basin (Figure 6.4B), whereas in 2019, the intensity varied between $D_{I,SPI} = -0.1 - -1.7 \text{ month}^{-1}$ with the lowest values in the north-west (Figure 6.4C). Comparison between 1994 and 2019 further suggests that, in spite of comparable basin-average intensity, the 2019 drought was locally considerably more intense.

The drought severity averaged for all droughts in 2002 – 2020 varied between $D_{S,SPI} = -20.5 - -4.4$ in space with the lowest values in the northern part of the basin (Figure 6.5A). The most severe drought in 1994 showed more extreme spatial differences with $D_{S,SPI} = -3.6 - -63.3$ and with the lowest values in the central part of the basin (Figure 6.5B). In 2019, the drought severity ranged between $D_{S,SPI} = -0.3 - -27.2$ with the lowest values in the north-west of the basin (Figure 6.5C). Therefore, the drought was significantly more severe in 1994 compared to 2019 throughout large parts of the basin, whereas in other regions such as the north-western part of the basin the drought was more severe in 2019.

Overall, it depended on the location within the basin whether or not the drought of 2019 was the most severe or intense in at least 20 years as perceived by locals. On the one hand, the drought of 2019 was the most intense (Figure 6.6A) and severe (Figure 6.7A) for at least 20 years in the northern part of the basin which supports the local perception. On the other hand, in the southern parts of the basin the drought of 2019 was the most severe and intense in less than 10 years which contradicts the local perception.

Total water storage

The GRACE-based drought intensity averaged over all droughts in 2002 – 2020 ranged between $D_{I,TSDI} = -0.8 - -1.1 \text{ month}^{-1}$ with minor spatial differences (Figure 6.4D). In 2004, the intensity ranged between $D_{I,TSDI} = -0.7 - -1.3 \text{ month}^{-1}$ with the lowest values in the south-east (Figure 6.4E), whereas in 2019, it varied between $D_{I,TSDI} = -0.5 - -1.4 \text{ month}^{-1}$ with the lowest values in the north-west (Figure 6.4F). Therefore, the drought was more intense in 2004 (for example south-east of the basin) or 2019 (for example north-west of the basin) depending on the location within the basin.

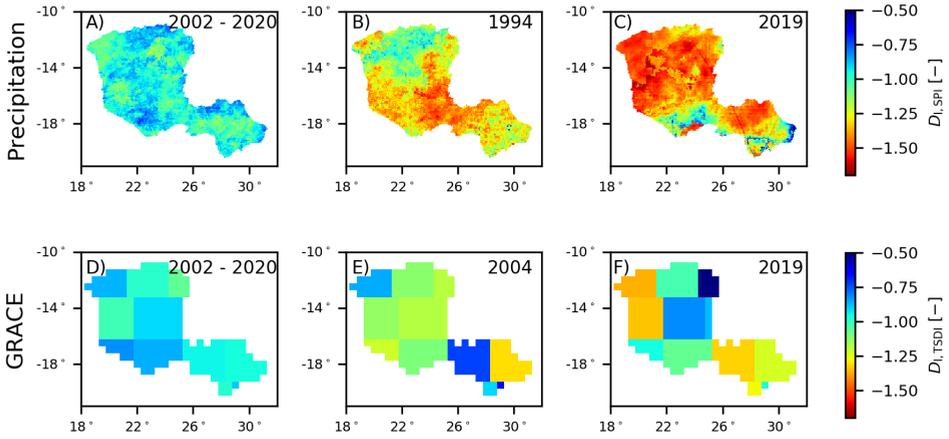


Figure 6.4: Spatial variability of the drought intensity according to the precipitation (A – C) and according to the total water storage (D – F). The maps were temporally averaged considering all droughts (A and D), the most severe drought according to precipitation (B) and GRACE (E), and drought of 2019 (C and F).

6

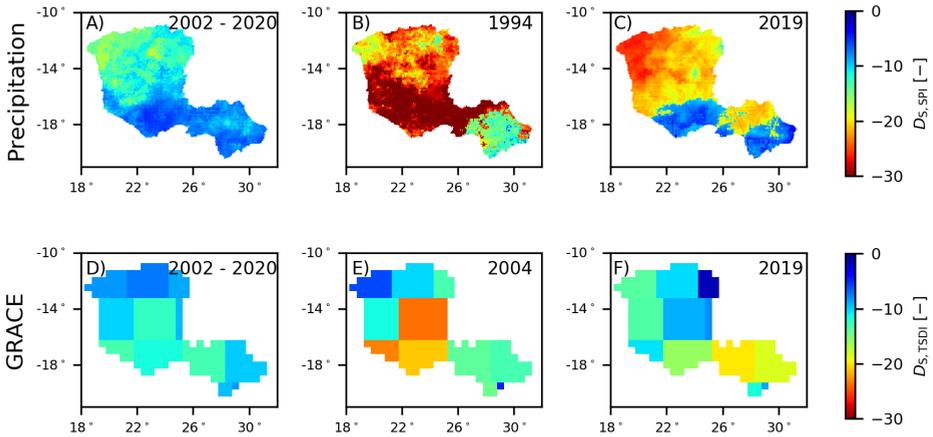


Figure 6.5: Spatial variability of the drought severity according to the precipitation (A – C) and according to the total water storage (D – F). The maps were temporally averaged considering all droughts (A and D), the most severe drought according to precipitation (B) and GRACE (E), and drought of 2019 (C and F).

The drought severity averaged over all droughts in 2002 – 2020 varied between $D_{S,TSDI} = -7.3 - -13.1$ in space with the lowest values in the central part of the basin (Figure 6.5D). In 2004, the severity ranged between $D_{S,TSDI} = -4.6 - -23.8$ with the

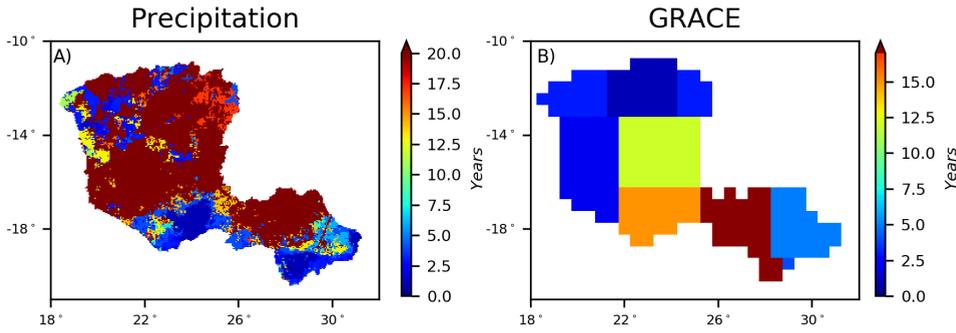


Figure 6.6: Spatial variability of the number of years between the drought of 2019 and the most recent drought before 2019 with more intense deficit according to A) precipitation or B) GRACE data. Grid cells where the drought of 2019 was the most intense observed with satellite data are coloured dark red.

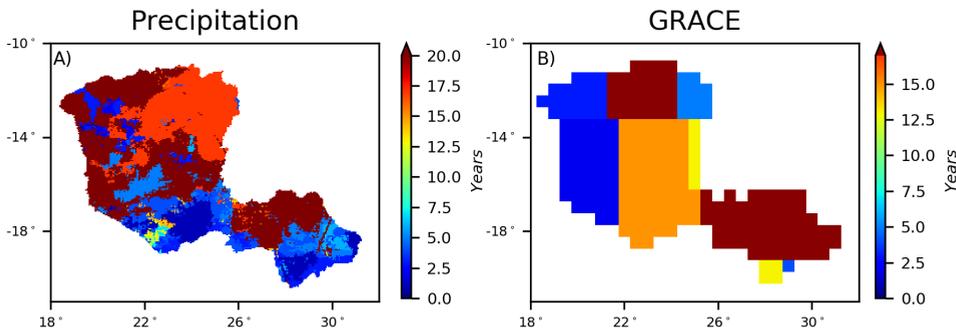


Figure 6.7: Spatial variability of the number of years between the drought of 2019 and the most recent drought before 2019 with more severe deficit according to A) precipitation or B) GRACE data. Grid cells where the drought of 2019 was the most severe observed with satellite data are coloured dark red.

lowest values in the central part of the basin (Figure 6.5E). In 2019, the severity ranged between $D_{S,TSDI} = -1.4 - -19.9$ with the lowest values in the south-east (Figure 6.5F). Therefore, depending on the location within the basin, as of June 2020, the drought was more severe in 2004 in the central parts of the basin and in 2019 in the south-east of the basin.

Overall, similar to precipitation-related drought, it here also depended on the location within the basin whether or not the drought of 2019 was the most severe and intense in at least 20 years as perceived by locals. On the one hand, in the western parts of the basin the drought of 2019 was the most severe and intense in less than 10 years (Figures 6.6B and 6.7B) which contradicts the local perception. On the other hand, the drought of 2019 was the most severe and intense for at least 17 years, the duration of the available GRACE data, in the south of the basin which at least does not contradict the local perception, but also does not support it with certainty. Insufficient data was available for this purpose as GRACE data was only available since 2002.

6.6. Discussion

While locals perceived the drought of 2019 as the most severe in at least 20 years, data indicated this differed depending on the drought characteristics, hydrological variable, and the location within the basin. On the one hand, the drought of 2019 resulted in the lowest basin-averaged annual rainfall since at least 27 years, most severe local rainfall deficit in the central and north-western part of the basin since at least 20 years, and lowest reservoir water level since 1995 which support the local perception that the drought of 2019 was the most extreme in more than 20 years. On the other hand, the rainfall deficit was more severe in 2002 with respect to the basin-average and locally in the north of the basin, reservoir water level deficit more severe in 2015, total storage deficit more severe in 2004 with respect to the basin-average and locally in the central part of the basin which contradict the local perception. However, the available data did not cover the entire water level and total storage deficit for 2019 as the TSDI and WLDI remained below zero towards the end of the available time-series in June 2020. Thus, the final duration and severity of the drought of 2019 remains unknown.

The water levels at Kariba reservoir not only depended on precipitation which largely dominates the inflow to the reservoir, but also on evaporation from the open water body and reservoir operation determining the amount of water leaving the reservoir. That is why low annual rainfall amounts do not necessarily result in low reservoir water levels. For example, reservoir operation is likely to explain the relatively high reservoir water levels in 2002 despite the very low rainfall and high evaporation in that year (Figures 6.2A, 6.2E and 6.3B).

The results in this study are sensitive to uncertainties in the satellite observations. Previous studies illustrated satellite-based precipitation observations were sensitive to bias (Kimani et al., 2017; Le Coz and van de Giesen, 2020). Uncertainties in total water storage anomalies are a result of data (post-) processing including data smoothing using a radius of for example 300 km affecting the spatial variability on basin scale (Blazquez et al., 2018b; Landerer and Swenson, 2012). Note, that GRACE observations were missing for several months among which July 2017 – May 2018 and August – September 2018 since the GRACE mission was ended in October 2017 and the mission GRACE-Follow-on (GRACE-FO) was launched in May 2018 (Kornfeld et al., 2019). These missing data resulted in an irregular temporal pattern in Figure 6.2C and D. Potential and actual evaporation data are prone to uncertainties related to the underlying estimation method and input data (Feng et al., 2019; Hobbins et al., 2008; Zhang et al., 2016). Altimetry uncertainties for open water bodies range between 4 – 36 cm depending on the lake size and climate conditions (Crétaux and Birkett, 2006; Schwatke et al., 2015). These uncertainties can affect long-term fluctuations patterns of these variables. Therefore, it is recommended to test these satellite observations against ground observations which was outside the scope of this study due to the limited in-situ data availability.

Previous studies have discussed various reasons for discrepancies between local perception of droughts and (satellite-based) observations. First, locals often perceive droughts through non-climatic factors such as reservoir water level or lack of electricity which is more clearly visible compared to rainfall or total water stor-

age amounts (Iqbal et al., 2018; Urquijo and De Stefano, 2016). Second, locals often remember most recent and extreme drought events, but forget intermediate droughts (Taylor et al., 1988). Third, local perceptions are often influenced by economic losses such as significant reductions in the expected crop production (Bola et al., 2014; Foguesatto et al., 2020; Meze-Hausken, 2004). Hence, farmers who experienced large crop failures often relate this to droughts even though the rainfall remained constant and also if they expanded their crop production in the past such that the water demand increased.

In contrast to previous studies, this study illustrated the local perception was partially supported and partially contradicted by satellite observations depending on the drought characteristic, hydrological variable and location within the basin. In other words, the simple statement “the drought of 2019 was worst in several decades” mentioned in news media (Brown, 2019; Edel, 2019) cannot be generalised. For example, locals in the west of the basin are more likely to consider the drought of 2019 as the most severe in at least 20 years if they mainly rely on rainwater (Figure 6.7A) instead of groundwater sources (Figure 6.7B). Note, monthly GRACE observations are dominated by slow processes which include changes in the groundwater system and seasonal variations in all storage components.

6.7. Evaluation

The objective of this study was to analyse the drought of 2019 in the Zambezi River Basin upstream of the Kariba reservoir to determine whether it was, as locally perceived, indeed the longest, most intense and severe drought, basin-wide and locally, according to remotely-sensed precipitation, total water storage and reservoir water level observations over at least 20 years. Data analysis indicates that it depends on the drought characteristic, the hydrological variable considered, and on the location within the basin. On the one hand, the drought of 2019 resulted in the lowest basin-averaged annual rainfall for at least 27 years, most severe local rainfall deficit in the central and north-western part of the basin for at least 25 years, and lowest reservoir water level since 1995 which supports the local perception that the drought of 2019 was indeed the most extreme over recent decades. However, on the other hand, the spatially averaged rainfall deficit, as well as locally in the north of the basin, was more severe in 2002. Also in 2004, the spatially averaged total storage deficit, as well as locally in the central part of the basin, was more severe. Similarly, the reservoir water level-based drought severity was more extreme in 2015. Therefore, it depends on the drought characteristic, the hydrological variable considered, and the location within the basin, whether the drought of 2019 was indeed the most extreme over at the last decades as perceived by local people.

7

Conclusions and outlook

*A genuine gift is like water from the tap;
when it flows out, it does not flow back*

African Proverb

This chapter is based on:

Hulsman, P., Winsemius, H. C., Michailovsky, C. I., Savenije, H. H. G., and Hrachowitz, M.: Using altimetry observations combined with GRACE to select parameter sets of a hydrological model in a data-scarce region, *Hydrol. Earth Syst. Sci.*, 24, 3331–3359, doi: 10.5194/hess-24-3331-2020, 2020a.

Hulsman, P., Savenije, H. H. G., and Hrachowitz, M.: Learning from satellite observations: increased understanding of catchment processes through stepwise model improvement, *Hydrol. Earth Syst. Sci. Discuss.*, 2020, 1–26, doi: 10.5194/hess-2020-191, 2020b.

Hulsman, P., Savenije, H. H. G., and Hrachowitz, M.: Why are long-term storage variations observed but not modelled in the Luangwa basin?, *Water Resources Research*, in review.

Hulsman, P., Savenije, H. H. G., and Hrachowitz, M.: Zambezi River Basin: Drought of 2019, *Journal of Hydrology: Regional Studies*, submitted.

The goal of this research was to assess the added value of satellite-based observations for hydrological modelling in a semi-arid, data-scarce river basin. This was done by focusing on a selection of satellite observations and hydrological model applications. More specifically, this research focused on exploring the added value of altimetry, evaporation and total water storage observations to improve our current understanding of local dominant hydrological processes, improve hydrological model structures, and improve the identification of feasible parameter sets in data-scarce regions.

In addition, satellite-based data analysis results provide valuable new information to improve our understanding of the hydrological system which can then be implemented in hydrological models. This was illustrated by analysing the drought of 2019 in the Zambezi River Basin using multiple satellite observations to determine whether it was the most extreme in at least 20 years as perceived by locals. As droughts influence human activities, depending on how they perceived specific drought events, and hence affect the hydrological system, satellite observations provide interesting opportunities to increase our understanding of factors influencing local perceptions.

Altimetry-based parameter set identification

Altimetry observations contain valuable information for model calibration in the absence of reliable discharge observations as explored in Chapter 3. Altimetry observations from 18 virtual stations obtained from multiple satellite missions were combined with GRACE data to identify feasible parameter sets comparing different calibration strategies. This was applied to a distributed process-based rainfall-runoff model with sub-grid process heterogeneity which was developed for the Luangwa river basin. With the benchmark calibration strategy using discharge observations as traditionally done, the model captured the flows relatively well reproducing the high flows, overestimating inter-mediate flows and underestimating low flows. When assuming no discharge observations are available and using GRACE data only to restrict the parameter space, the modelled discharge was continuously significantly overestimated. When combining GRACE with altimetry data by 1) focusing only on the water level dynamics or 2) converting modelled discharge to water levels using rating curves, the discharge was also continuously overestimated. On the other hand, the discharge was reproduced well when using GRACE and altimetry data combined with river geometry information for the discharge-water level conversion using the Strickler-Manning equation. With this parameter set identification strategy, the feasible parameter sets reproduced the discharge well similar to the benchmark. The identification of feasible parameter sets improved further when using more accurate cross-section data and when increasing the number of virtual stations used. Therefore, if there is no reliable discharge data available, altimetry observations from multiple virtual stations together with GRACE data can fill this gap when including river geometry information for the discharge-water level conversion.

Model structure development using satellite observations

Satellite-based evaporation and total water storage observations provide valuable spatial and temporal information which can be used to improve our understanding of hydrological processes through step-wise model development and to calibrate hydrological models as explored in Chapter 4. For this purpose, the benchmark distributed hydrological model for the Luangwa River basin was adjusted iteratively and during each step calibrated with respect to discharge only and multiple variables simultaneously. It was shown that the benchmark model calibrated with respect to discharge, reproduced the discharge, basin-average evaporation and total water storage well, but poorly modelled the evaporation in wetland dominated areas and the spatial pattern of the evaporation and total water storage. By testing five alternative model structures, it was shown that all variables, hence discharge, evaporation and total water storage, were represented better when including the process of upwelling groundwater flow from a distributed groundwater reservoir and calibrating with respect to all three variables simultaneously. These changes improved the spatial and temporal variability of the individual variables with the exception of the temporal variation of the basin-average total water storage. Therefore, satellite-based evaporation and total water storage observations have proven to be valuable not only for multi-criteria calibration, but also to diagnose model deficiencies, improve the model structure step-wise and hence improve our understanding of local hydrological processes.

Long-term total water storage variations in the Luangwa Basin

Long-term total water storage variations were observed with GRACE in the Luangwa basin, but were not reproduced by the previously developed hydrological model. Also in many other river basins, long-term storage variations were observed, but not reproduced by standard hydrological models. Therefore, processes underlying these variations were identified in Chapter 5 through a combined data analysis and model hypothesis testing approach. The results indicated the model did not reproduce long-term storage variations as a result of the chosen forcing data and the missing representation of regional groundwater export. The variation in the long-term annual maximum total water storage improved by using different forcing data. On the other hand, the fluctuations in the long-term annual minimum improved by allowing groundwater to seep into a deep groundwater layer during wet conditions from where it leaks out of the basin. Therefore, satellite observations were used successfully to identify discrepancies in the total water storage, identify potential drivers through satellite data analyses in a poorly gauged region, test the impact of alternative forcing data sources, improve our understanding of hydrological processes through iterative hypothesis testing and improve the quantitative description in the model to incorporate new processes well.

Sub-Saharan drought of 2019 from space

During the dry season in 2019, a drought that resulted in extremely low water levels at Kariba hydro-power dam, was perceived by locals as the most severe in several decades. To test the validity of this perception, Chapter 6 analysed the drought of 2019 in the Zambezi River Basin upstream of the Kariba reservoir to determine whether it was, as locally perceived, indeed the longest, most intense and severe drought, basin-wide and locally, according to remotely-sensed precipitation, total water storage and reservoir water level observations over at least 20 years. Data analysis indicates that it depends on the drought characteristic, the hydrological variable considered, and on the location within the basin. On the one hand, the drought of 2019 resulted in the lowest basin-averaged annual rainfall over at least 27 years, most severe local rainfall deficit in the central and north-western part of the basin for at least 25 years, and lowest reservoir level since 1995 which supports the local perception that the drought of 2019 was indeed the most extreme over recent decades. However, on the other hand, the spatially averaged rainfall deficit, as well as locally in the north of the basin, was more severe in 2002. Also in 2004, the spatially averaged total storage deficit, as well as locally in the central part of the basin, was more severe. Similarly, the reservoir water level-based drought severity was more extreme in 2015. Therefore, it depends on the drought characteristic, the hydrological variable considered, and the location within the basin, whether the drought of 2019 was indeed the most extreme over at the last decades as perceived by local people.

7

Outlook

This research highlighted the added value of a selection of satellite-based observations for a selection of hydrological model applications. However, there remain many opportunities yet to be explored as illustrated in the next sections.

Process understanding and model structure development

In this research, satellite observations were used to improve our understanding of hydrological processes through step-wise model improvement. For this purpose, first discrepancies between observations and model results were identified for system-internal fluxes and stores with respect to their temporal and spatial variability. Then, the model structure was adapted step-wise in an attempt to resolve the discrepancies under consideration. On the one hand, discrepancies can be a result of uncertainties in the (satellite-based) observation data, but on the other hand it can be due to uncertainties in the model input data, model structure, parameterization and/or parameter set (Fowler et al., 2018; Grigg and Hughes, 2018; Jing et al., 2019; Saft et al., 2016). Discrepancies can occur throughout the entire hydrological system including for example the discharge time-series at the basin outlet, runoff within the basin, evaporation or storage.

For example in this research, the benchmark model in Chapter 4 modelled the discharge well, but poorly captured the evaporation in wetland dominated areas as it decreased to zero in contrast to the satellite observations. Through model hy-

pothesis testing, it was illustrated this was a result of groundwater upwelling which was originally missing in the model similar to many other distributed conceptual hydrological models (Bieger et al., 2017; Samaniego et al., 2010). Another example is the long-term storage variations observed with GRACE which was originally not reproduced with the benchmark model in Chapter 5. Through data analysis and model hypothesis testing, it was shown these fluctuations were a result of the chosen forcing data source and the process of deep groundwater export which was originally assumed to be negligible.

Similar to these two examples, satellite-based evaporation and total water storage observations can be used to highlight other discrepancies in hydrological models to improve our understanding of local hydrological processes through model hypothesis testing. Some examples are: How does the evaporation in other hydrological response units such as forested areas change in time in the model and the observation? In which regions besides wetlands are roots connected to the groundwater system affecting the dry season evaporation (Fan et al., 2017)? Why are in the woodlands in the Luangwa basin new leaves growing on trees at the end of the dry season before the start of the wet season (Campbell, 1996; Zimba et al., 2020)? Is this reflected in the observed and modelled evaporation?

Also alternative variables such as satellite-based surface soil moisture, land surface or air temperature, snow cover (if there is snowfall), river width, land cover, NDVI (normalized difference vegetation index), NDII (normalized difference infrared index) or LAI (leaf area index) can be used for this purpose (Jiang and Wang, 2019; Lakshmi, 2004; Xu et al., 2014). Some examples are: When are floodplains or other areas in the basin flooded in the model and observations? Do such areas become flooded already with the first rains at the start of the wet season or after some time when the soil is saturated? Does the observed water content in the surface soil moisture correspond with the spatial and temporal rainfall pattern and model results? For all of the above mentioned questions, what could explain any anomalies?

Note, discrepancies can also be a result of uncertainties in the data used for model evaluation such that the model representation is correct rather than the observation. Therefore, it is important to eliminate this possibility through data validation if possible.

Catchment classification

Throughout this research, the Luangwa river basin was classified based on the topography into the hydrological response units (HRUs) sloped, flat and wetland dominated areas. The hydrological model structure and parameter sets then varied for each HRU considering the expected dominating hydrological processes as explained in Chapters 3, 4 and 5. However, additional or other properties besides topography can be used when discretizing the river basin depending on what controls the dominant hydrological processes. For example, the dominant runoff processes can be controlled by the topography, geology, land cover, land management, climate or any combination thereof (e.g. Fenicia et al., 2016; Gao et al., 2014a; Haghnegahdar et al., 2015; Knoben et al., 2018). It would be interesting

to explore the added value of additional (or other) variables for the discretization of the Luangwa basin. Previous studies illustrated the added value of incorporating more data sources when classifying the river basin into HRUs (Haghnegahdar et al., 2015; Haverkamp et al., 2002; Petrucci and Bonhomme, 2014). However, the more information is used, the larger the number of HRUs, the more calibration parameters. Therefore, it is important sufficient information is available to support the higher number of parameters to avoid equifinality (Beven, 2006; Hrachowitz and Clark, 2017).

Parameter set identification

Traditionally, hydrological models are calibrated with respect to discharge time-series at the basin outlet. However, an increasing number of studies have illustrated the added value of satellite observations in the calibration procedure (e.g. Michailovsky et al., 2013; Sun et al., 2018a; Werth et al., 2009; Winsemius et al., 2008). In this research, feasible parameter sets were identified using altimetry time-series at multiple virtual stations simultaneously, or using spatial-temporal information available in total water storage and evaporation data.

Building on Chapter 3, it would be interesting to combine altimetry observations with river width estimates derived from Landsat or Sentinel-1/2 (Hou et al., 2018; Pekel et al., 2016) or energy gradient estimates derived from CryoSat-based river water level observations (Jiang et al., 2017; Schneider et al., 2017) as discussed in Chapter 3 (Section 3.5.5). With the upcoming SWOT (Surface Water Ocean Topography) mission more accurate altimetry observations are expected at higher spatial resolution which can be used to further improve the parameter set identification also in small river basins (Biancamaria et al., 2016; Langhorst et al., 2019; Oubanas et al., 2018).

Building on Chapter 4, it would be interesting to explore the added value of additional/alternative satellite observations for spatial-temporal parameter set identification building on the findings of previous studies (e.g. Demirel et al., 2018; Mendiguren et al., 2017; Milzow et al., 2011; Nijzink et al., 2018; Zink et al., 2018). For example, satellite-based surface soil moisture, land surface or air temperature, snow cover (if there is snowfall), river width, NDVI (normalized difference vegetation index), NDII (normalized difference infrared index) or LAI (leaf area index) could provide valuable information (Jiang and Wang, 2019; Lakshmi, 2004; Xu et al., 2014).

It would also be interesting to look into different model performance metrics to exploit the available spatial and temporal information as much as possible. This was not done in this research as the focus was on assessing the added value of different variables for model calibration rather than comparing the impact of using different model performance metrics.

In conclusion, satellite observations have the potential to alleviate much of the data scarcity in many regions worldwide. Incorporating these observations in particular for step-wise hydrological model development enables more accurate representations of the hydrological system and hence more reliable model predictions in data scarce regions. This research illustrated several new hydrological applications of satellite observations and highlighted multiple additional opportunities in the previous sections. In the future, even more remotely-sensed hydrological information are expected to become available with higher spatial-temporal resolution and data quality such that the methods explored in this research can be further refined. Therefore, satellite observations provide us with abundant information which help us to overcome much of the limitations the hydrological community is currently facing with respect to data scarce regions.



Appendix

This appendix provides supplementary material for Chapter 3.

Background on the altimetry data from LEGOS

The altimetry data obtained from LEGOS come from the acquisitions of ENVISAT and Jason-2 radar altimetry missions on their nominal orbit (03/2002–10/2010 and 06/2008–10/2016 respectively). All the parameters necessary to estimate water levels (Crétaux et al., 2017) are contained in the Geophysical Data Records (GDR) made available by the space agencies. These data were obtained from Centre de Topographie des Océans et de l'Hydrosphère (CTOH – <http://ctoh.legos.obs-mip.fr>). Ranges used to derive altimeter heights are those processed using OCOG/Ice retracking algorithm (Wingham et al., 1986). Previous studies showed that altimeter heights derived using this retracking algorithm are more suitable for hydrological studies in terms of accuracy of water levels and availability of the data (Frappart et al., 2006; Santos da Silva et al., 2010; Sulistioadi et al., 2015) among the commonly available retracked data present in the GDRs. The Multi-mission Altimetry Processing Software (MAPS) was used to visualize and process the altimetry data in order to obtain the virtual stations (VS) at the cross-sections between the altimeter ground tracks and the rivers (Frappart et al., 2015; Normandin et al., 2018). Data processing is composed of three main steps: (i) a coarse delineation of the VS using Google Earth; (ii) a refined selection of the valid altimetry data based on visual inspection; and (iii) the computation of the time series of water level. The altimetry-based water level is computed for each cycle using the median of the selected altimetry heights, along with their respective deviation (i.e., mean absolute deviation). This process is repeated each cycle to construct the water level time series at the virtual stations; see Frappart et al. (2015) and Normandin et al. (2018) for more details.

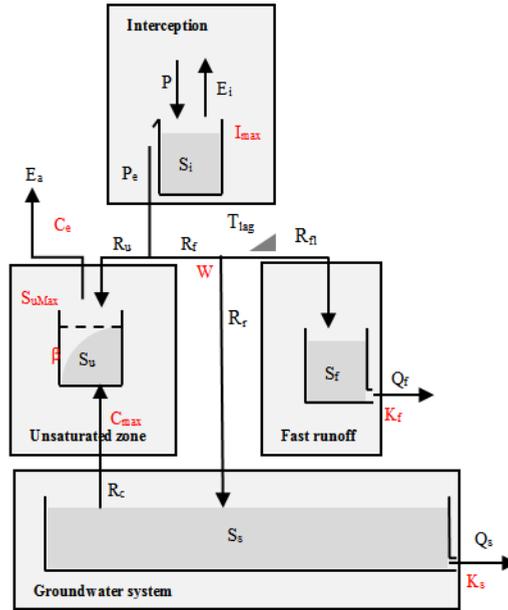


Figure A.1: Model structure. Parameters are marked in red, storages and fluxed in black. Symbol explanation: Fluxes [mm d⁻¹]: precipitation (P), effective precipitation (P_e), potential evaporation (E_p), interception evaporation (E_i), plant transpiration (E_t), infiltration into the unsaturated zone (R_u), drainage to fast runoff component (R_r), delayed fast runoff (R_n), groundwater recharge (R_r for each relevant HRU and $R_{r,tot}$ combining all relevant HRUs), upwelling groundwater (R_{GW} for each relevant HRU and $R_{GW,tot}$ combining all relevant HRUs), fast runoff (Q_f for each relevant HRU and $Q_{f,tot}$ combining all relevant HRUs), groundwater/slow runoff (Q_s), total runoff (Q_m). Storages [mm]: storage in interception reservoir (S_i), storage in unsaturated root zone (S_u), storage in groundwater/slow reservoir (S_s), storage in fast reservoir (S_f). Parameters: interception capacity (I_{max}) [mm], maximum upwelling groundwater (C_{max}) [mm d⁻¹], maximum root zone storage capacity ($S_{u,max}$) [mm], splitter (W) [-], shape parameter (β) [-], transpiration coefficient (c_e) [-], time lag (T_{lag}) [d], reservoir time scales [d] of fast (K_f) and slow (K_s) reservoirs.

Table A.1: Model parameter values and ranges. See Figure A.1 for the parameter explanation and Table A.2 for the parameter constrains applied during the random parameter generation.

Landscape class	Parameter	min	max	Unit
Entire catchment	K_s	100	100	d
	C_e	0.5	0.5	-
Plateau/Terrace	I_{\max}	0	2	mm d ⁻¹
	$S_{u,\max}$	200	2000	mm
	K_f	10	12	d
	W	0.1	0.5	-
Hillslope	I_{\max}	0	2	mm d ⁻¹
	$S_{u,\max}$	200	2000	mm
	β	0	2	-
	T_{lag}	1	5	d
	K_f	10	12	d
Wetland	W	0.1	0.5	-
	I_{\max}	0	2	mm d ⁻¹
	$S_{u,\max}$	200	2000	mm
	K_f	10	12	d
	W	0.1	0.5	-
River profile	C_{\max}	0.1	2	mm d ⁻¹
	v	0.01	5.0	m s ⁻¹
	k	5	45	m ^{$\frac{1}{3}$} s ⁻¹
	a	0.1	800	m ³ s ⁻¹
	b	1	3	-

Table A.2: Parameter constrains. See Figure A.1 for the parameter explanation.

Parameter	Constrain
Maximum root zone storage capacity	$S_{u,\max,\text{hillslope}} > S_{u,\max,\text{plateau/terrace}}$ $S_{u,\max,\text{hillslope}} > S_{u,\max,\text{wetland}}$
Maximum interception	$I_{\max,\text{hillslope}} > I_{\max,\text{plateau/terrace}}$ $I_{\max,\text{hillslope}} > I_{\max,\text{wetland}}$
Splitter for groundwater percolation	$W_{\text{hillslope}} > W_{\text{plateau/terrace}}$

Table A.3: Characteristics of the virtual stations in the Luangwa River basin for which remotely sensed river water levels are available: station number, coordinates (X , Y), river slope (i), river width (B), river bank slopes (i_1 and i_2), upstream catchment area, acquisition date of the image in Google Earth used to extract the river geometry information, and discharge at Luangwa Bridge gauge station (basin outlet; absolute values and relative to the maximum discharge); in the absence of discharge data on the acquisition dates, the long-term mean daily values for the entire time period available were used.

VS	X	Y	i [-]	B [m]	i_1 [-]	i_2 [-]	A [m ²]	Acquisition date	Q_{absolute} [m ³ s ⁻¹] Q_{relative} [%]
1	30.2823°	-14.8664°	0.00049	324	36	29	10211995771	13-9-2010	68 (1%)
2	30.0864°	-14.366°	0.00062	7	17	83	14859805930	13-10-2013	65 (1%)
3	32.1715°	-12.4123°	0.00019	3	19	42	44337218380	17-12-2013	211 (4%)
4	31.1868°	-13.5927°	0.00020	129	42	8	87227195673	5-6-2013	160 (3%)
5	31.6984°	-13.2039°	0.00020	185	31	20	78090945429	20-9-2013	60 (1%)
6	32.2998°	-12.2007°	0.00039	170	30	17	40935244516	13-6-2013	146 (3%)
7	32.2805°	-12.1157°	0.00030	78	38	77	40747298483	13-6-2013	146 (3%)
8	32.831°	-11.3674°	0.00031	10	48	21	21066101487	26-9-2013	97 (2%)
9	30.2704°	-14.8809°	0.00017	99	8	5	102140213550	14-11-2009	30 (1%)
10	31.78405°	-13.0995°	0.00029	100	26	20	77559639645	26-7-2013	89 (2%)
11	31.71099°	-13.1943°	0.00020	54	34	30	78051272962	20-9-2013	60 (1%)
12	30.2740°	-14.8763°	0.00017	82	8	15	102135928406	14-11-2009	30 (1%)
13	32.15843°	-12.412°	0.00019	87	43	30	44340963341	17-12-2013	211 (4%)
14	32.15989°	-12.4127°	0.00019	128	83	19	44339840479	13-6-2013	146 (3%)
15	30.2740°	-14.8763°	0.00017	82	8	15	102139379771	13-6-2013	146 (3%)
16	32.16056°	-12.4125°	0.00019	128	83	19	44339840479	17-12-2013	211 (4%)
17	31.80001°	-13.0909°	0.00029	86	21	83	77553414963	13-6-2013	146 (3%)
18	30.61577°	-14.1852°	0.00051	227	24	20	96231647197	20-9-2014	60 (1%)
Outlet	30.21491°	-14.96678°	0.00037	149	8.62	10.10	154325857000	26-7-2016	89 (2%)

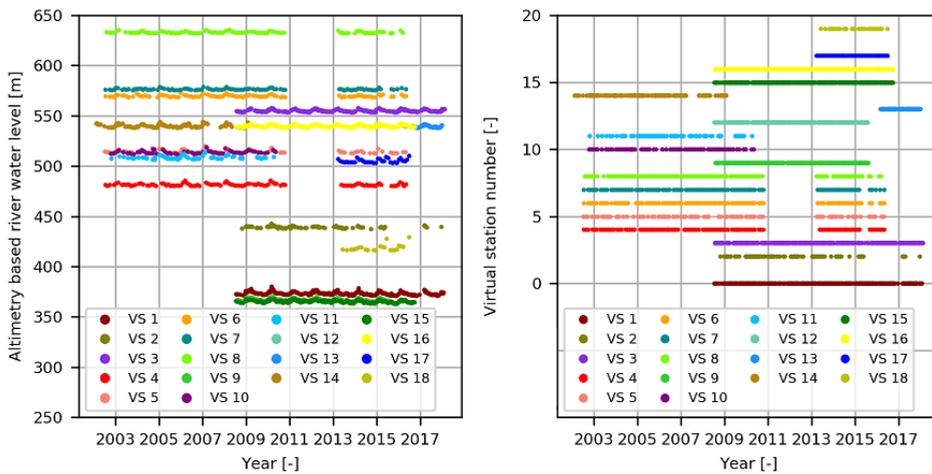


Figure A.2: Visualisation of the altimetry time series relative to a reference ellipsoid (left) and altimetry data availability (right) for all virtual stations used in this study. The colours for the individual stations correspond with those in Figure 3.1 and 3.3.

A

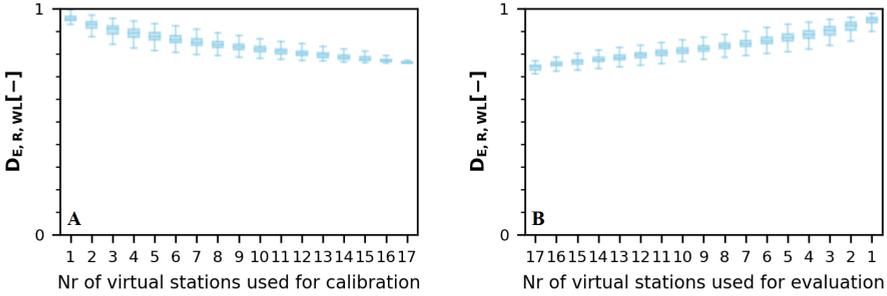


Figure A.3: Influence of the number of virtual stations used for A) model calibration and B) evaluation on the model performance $D_{E,R,WL}$ applying Altimetry Strategy 1.

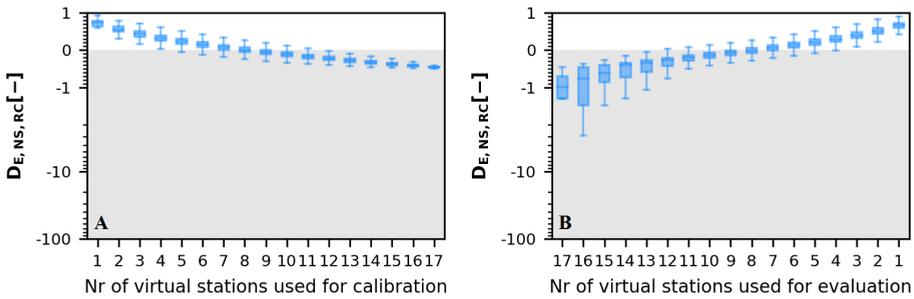


Figure A.4: Influence of the number of virtual stations used for A) model calibration and B) evaluation on the model performance $D_{E,NS,RC}$ applying Altimetry Strategy 2.

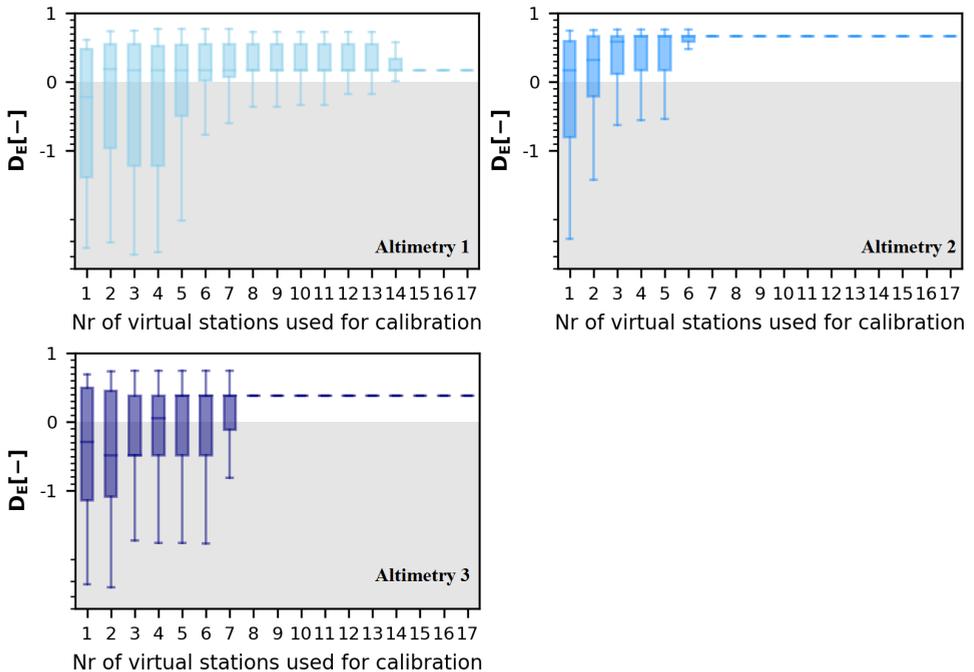


Figure A.5: Influence of the number of virtual stations used for model calibration on the model performance D_E with respect to discharge

Table A.4: Model performance with respect to each flow signature separately ($E_{NS,Q}$, $E_{NS,\log Q}$, $E_{NS,FDC}$, $E_{NS,\log FDC}$, $E_{R,RCDdry}$, $E_{R,RCDdry}$, $E_{R,RCDwet}$, $E_{NS,AC}$, $E_{R,RLD}$) for each parameter identification strategy.

	$E_{NS,Q}$	$E_{NS,\log Q}$	$E_{NS,FDC}$	$E_{NS,\log FDC}$	$E_{R,RCDdry}$	$E_{R,RCDwet}$	$E_{NS,AC}$	$E_{R,RLD}$
Discharge	0.78	0.56	0.95	0.88	0.08	0.67	0.94	0.73
Seasonal water storage	(0.68 - 0.76)	(0.61 - 0.78)	(0.81 - 0.97)	(0.80 - 0.97)	(0.11 - 0.97)	(0.61 - 0.84)	(0.89 - 0.97)	(0.29 - 0.75)
Altimetry Strategy 1	-1.38	0.09	-0.80	0.25	0.52	0.27	0.98	0.61
Altimetry Strategy 2	(-2.2 - -0.05)	(0.11 - 0.73)	(-1.6 - 0.40)	(0.27 - 0.93)	(0.45 - 0.97)	(0.24 - 0.36)	(0.97 - 0.99)	(0.14 - 0.79)
Altimetry Strategy 3	0.65	0.69	0.93	0.85	0.87	0.64	0.92	0.15
Water level Strategy 1	(-2.9 - 0.08)	(0.20 - 0.69)	(-2.0 - 0.68)	(0.37 - 0.89)	(0.42 - 0.81)	(0.22 - 0.44)	(0.97 - 0.99)	(0.10 - 0.46)
Water level Strategy 2	-0.31	0.26	0.10	0.41	0.55	0.34	0.99	0.51
	(-2.5 - 0.27)	(-0.12 - 0.66)	(-1.9 - 0.73)	(0.05 - 0.88)	(0.38 - 0.89)	(0.23 - 0.46)	(0.97 - 0.99)	(0.11 - 0.70)
	0.61	0.70	0.95	0.91	0.93	0.60	0.99	0.52
	(-0.30 - 0.50)	(0.45 - 0.78)	(0.16 - 0.88)	(0.65 - 0.99)	(0.54 - 0.98)	(0.34 - 0.54)	(0.98 - 0.99)	(0.29 - 0.79)
	0.65	0.74	0.97	0.94	0.83	0.64	0.98	0.76
	(-0.49 - 0.60)	(0.37 - 0.75)	(0.13 - 0.95)	(0.593 - 0.98)	(0.52 - 0.98)	(0.34 - 0.60)	(0.98 - 0.99)	(0.12 - 0.75)
	0.14	0.78	0.57	0.98	0.64	0.40	0.98	0.57
	(-1.2 - 0.50)	(0.53 - 0.77)	(-0.62 - 0.91)	(0.73 - 0.98)	(0.66 - 0.99)	(0.28 - 0.55)	(0.98 - 0.99)	(0.26 - 0.77)

A

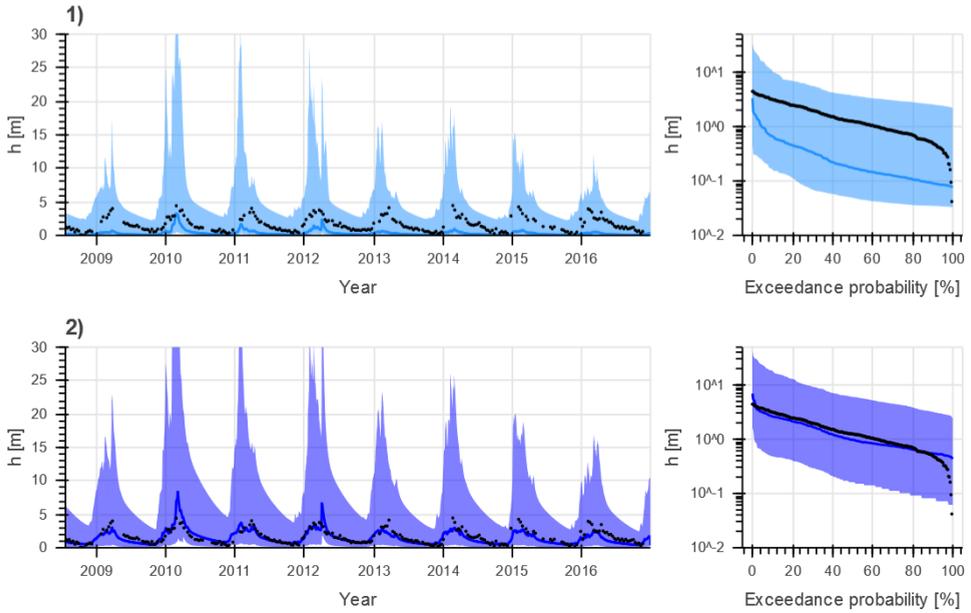


Figure A.6: Range of model solutions for Virtual Station 4 (see Figure 3.1 for its location). The left panel shows the time series and the right panel the exceedance probability graph of the recorded (black) and modelled water level: the line indicates the solution with the highest calibration objective function and the shaded area the envelope of the solutions retained as feasible. Solutions retained as feasible based on altimetry observations using all virtual stations within the basin and A) calibrated rating curves for the discharge – water level conversion (Altimetry Strategy 2) or B) the Strickler-Manning equation with cross-section information retrieved from Google Earth (Altimetry Strategy 3).

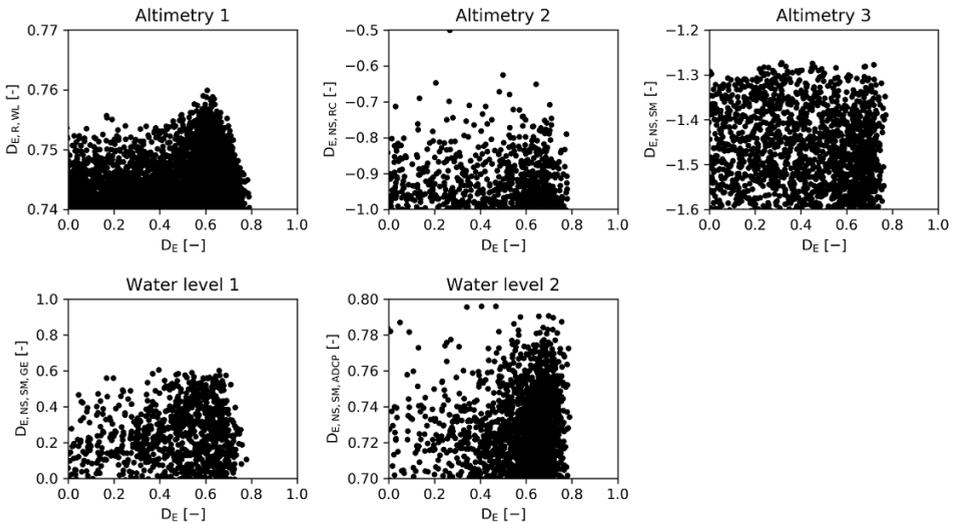


Figure A.7: Model performance with respect to discharge (horizontal axes) vs. model performance with respect to (satellite based) river water level (vertical axes) for each calibration strategy

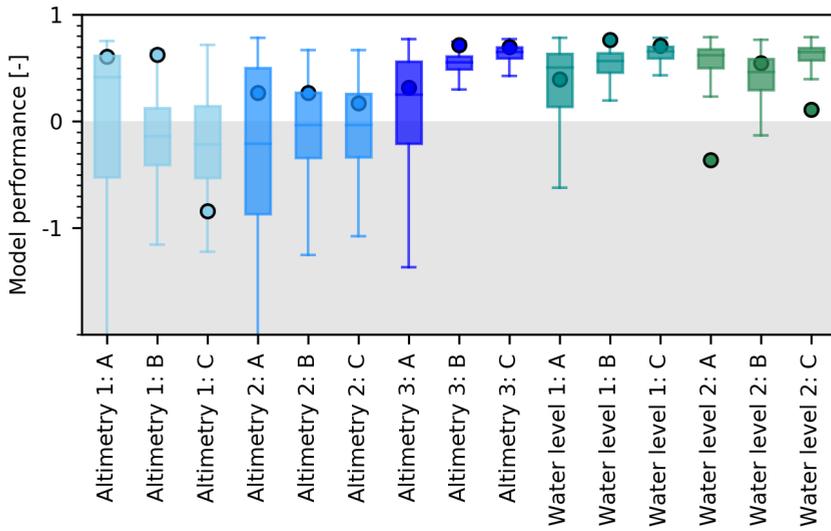


Figure A.8: Model performance with respect to discharge for each calibration strategy. Parameter sets were selected based on A) (satellite based) river water level only, B) first GRACE, then (satellite based) river water level, and C) first (satellite based) river water level, then GRACE

B

Appendix

This appendix provides supplementary material for Chapter 4.

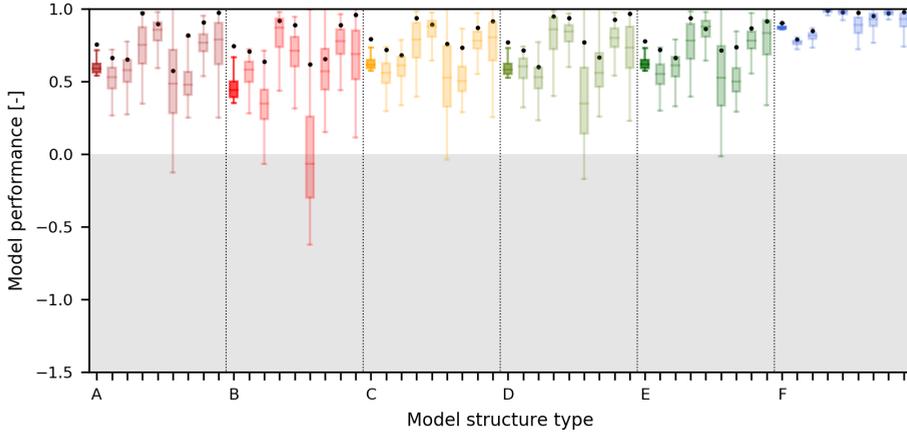


Figure B.1: Calibrated model performance of all models with respect to discharge (2002 – 2012). The boxplots visualise the spread of the best 5% solutions according to $D_{E,Q_{cal}}$ in the overall model performance $D_{E,Q_{cal}}$ and the following individual signatures: 1) daily discharge ($E_{NS,Q}$), 2) its logarithm ($E_{NS,\log Q}$), 3) flow duration curve ($E_{NS,FDC}$), 4) its logarithm ($E_{NS,\log FDC}$), 5) average runoff coefficient during the dry season ($E_{R,RC_{dry}}$), 6) average seasonal runoff coefficient during the wet season ($E_{R,RC_{wet}}$), 7) autocorrelation function ($E_{NS,AC}$), and 8) rising limb density ($E_{R,RLD}$). The dots visualise the model performance with the “optimal” parameter set using the overall model performance metric ($D_{E,Q_{calopt}}$).

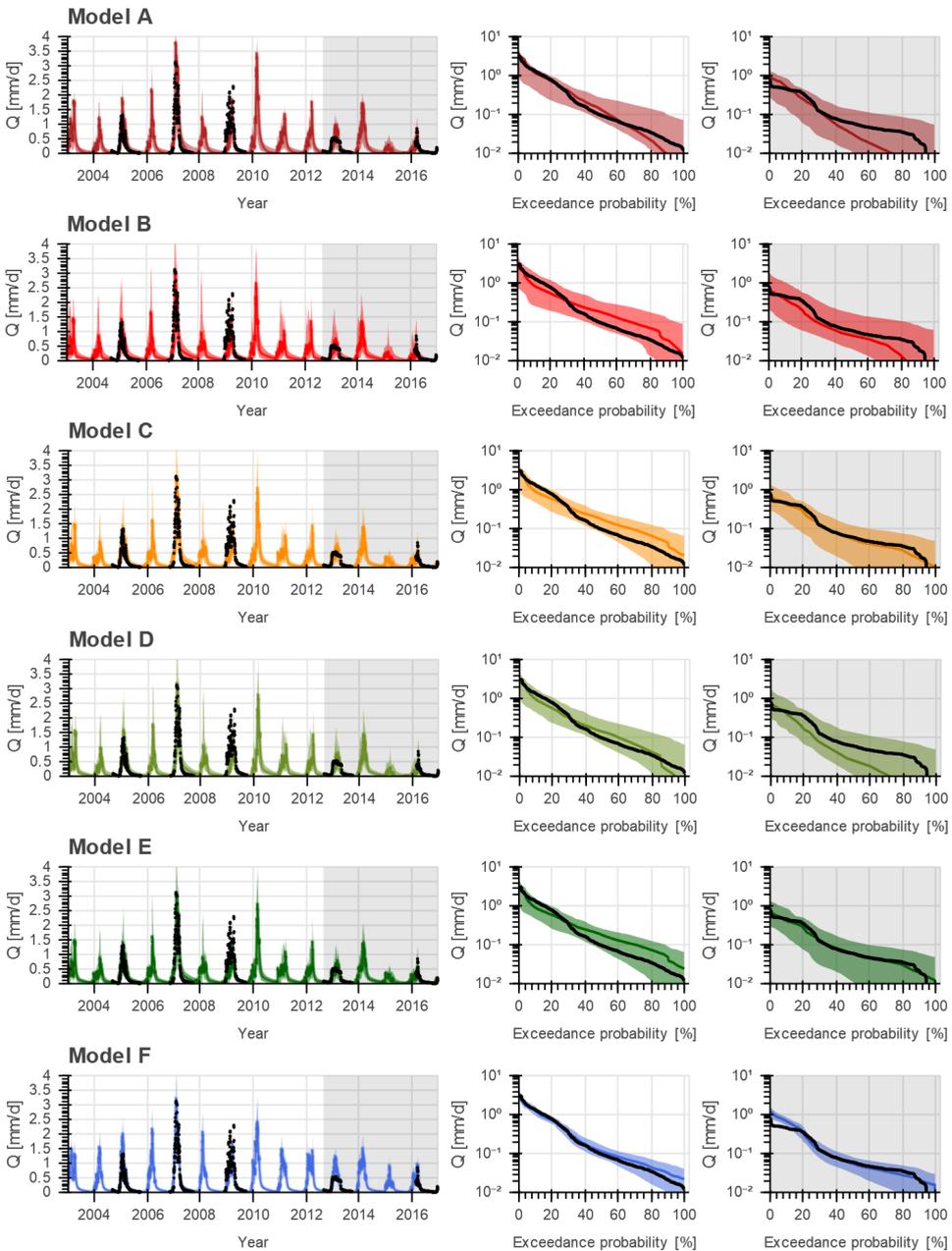


Figure B.2: Range of model solutions for Models A to F. The left panel shows the hydrograph and the right panel the flow duration curve in logarithmic scale of the recorded (black) and modelled discharge: the line indicates the solution with the highest calibration objective function with respect to discharge ($D_{E,Q_{cal}}$) and the shaded area the envelope of the solutions retained as feasible. The data in the white area were used for calibration and the grey shaded area for validation.

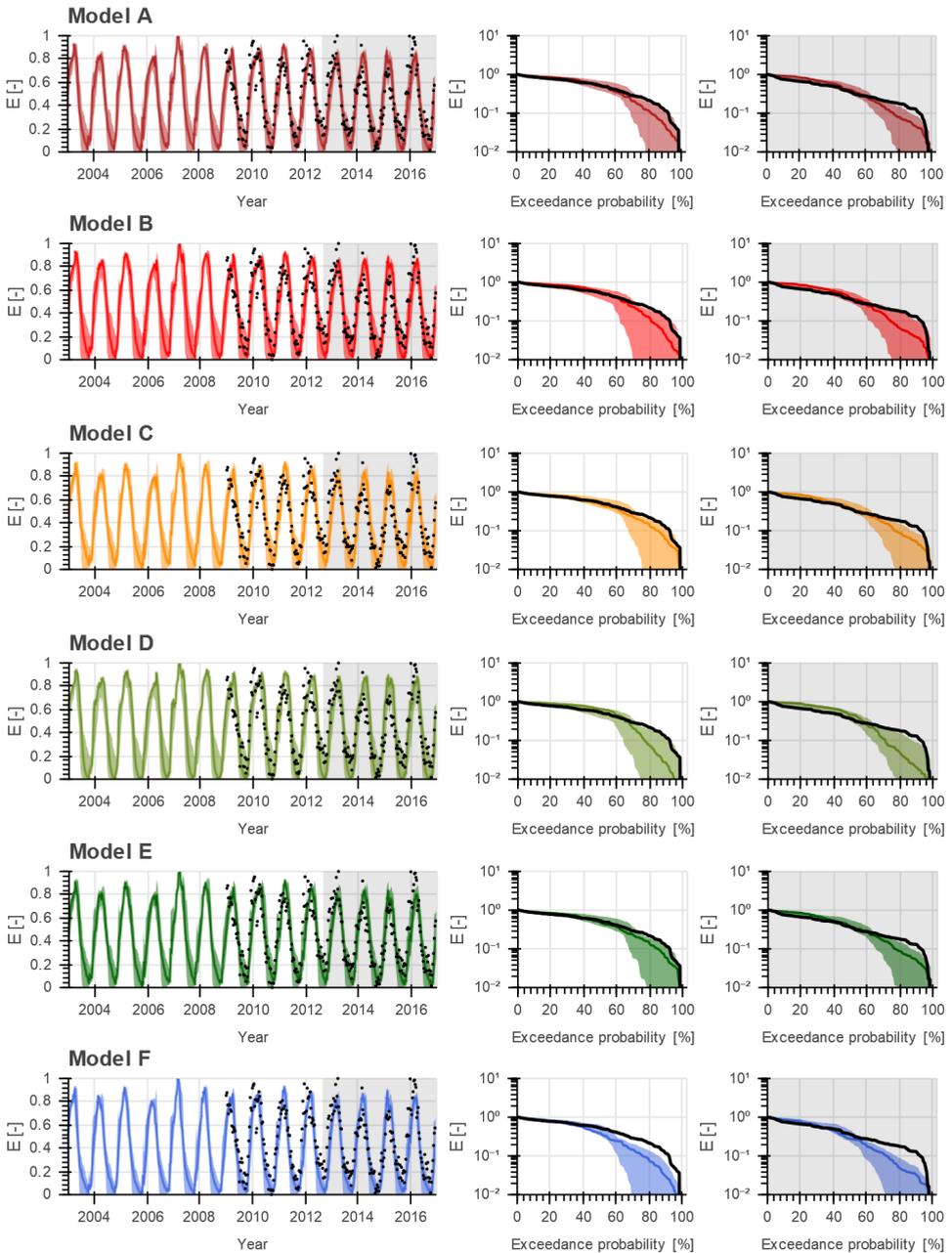


Figure B.3: Range of model solutions for Models A to F. The left panel shows the time series and the right panel the duration curve in logarithmic scale of the recorded (black) and modelled normalised basin average evaporation: the line indicates the solution with the highest calibration objective function with respect to discharge ($D_{E,Q_{cal}}$) and the shaded area the envelope of the solutions retained as feasible. The data in the grey shaded area were used for validation.

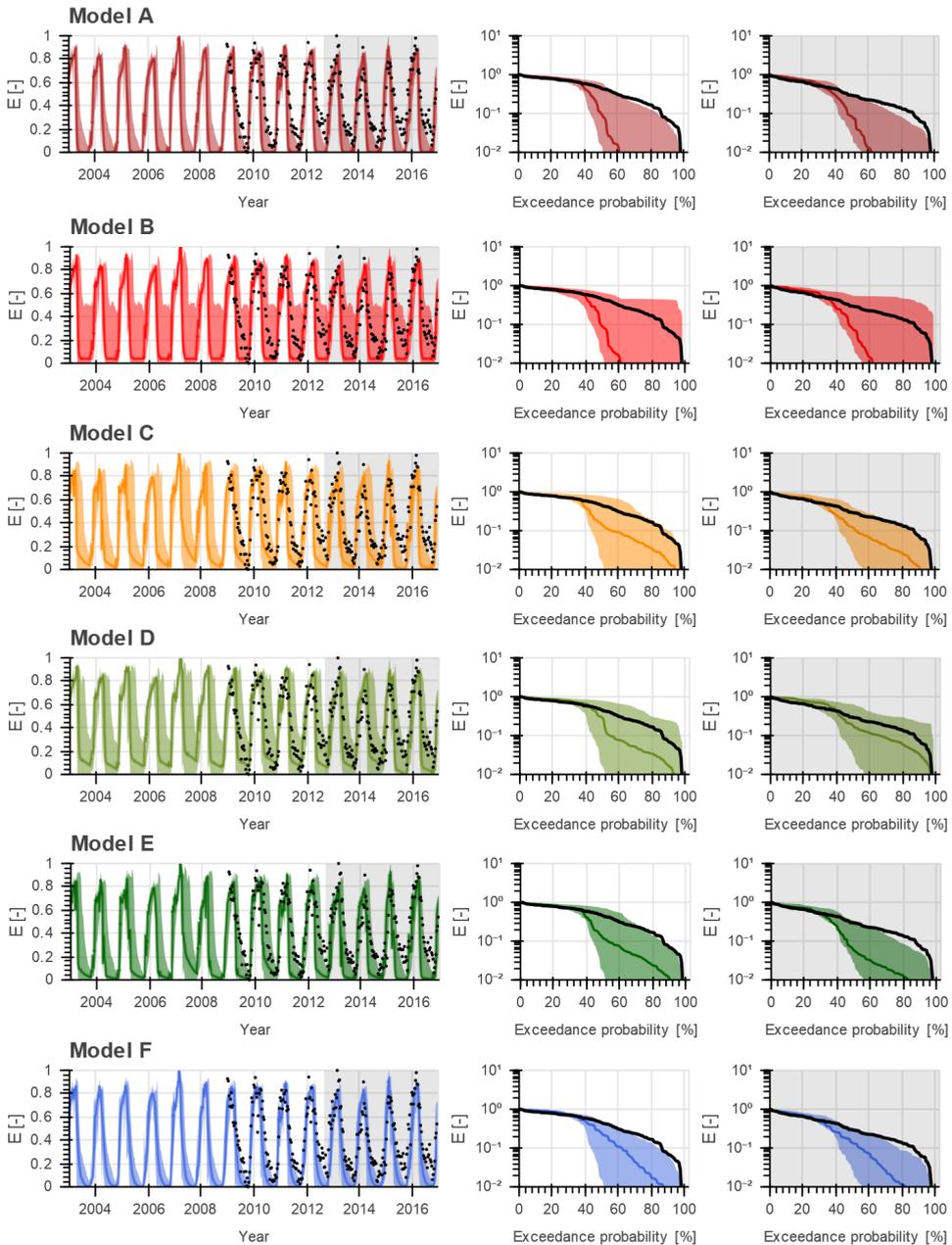


Figure B.4: Range of model solutions for Models A to F. The left panel shows the time series and the right panel the duration curve in logarithmic scale of the recorded (black) and modelled normalized evaporation for wetland dominated areas: the line indicates the solution with the highest calibration objective function with respect to discharge ($D_{E,Q_{\text{cal}}}$) and the shaded area the envelope of the solutions retained as feasible. The data in the grey shaded area were used for validation.

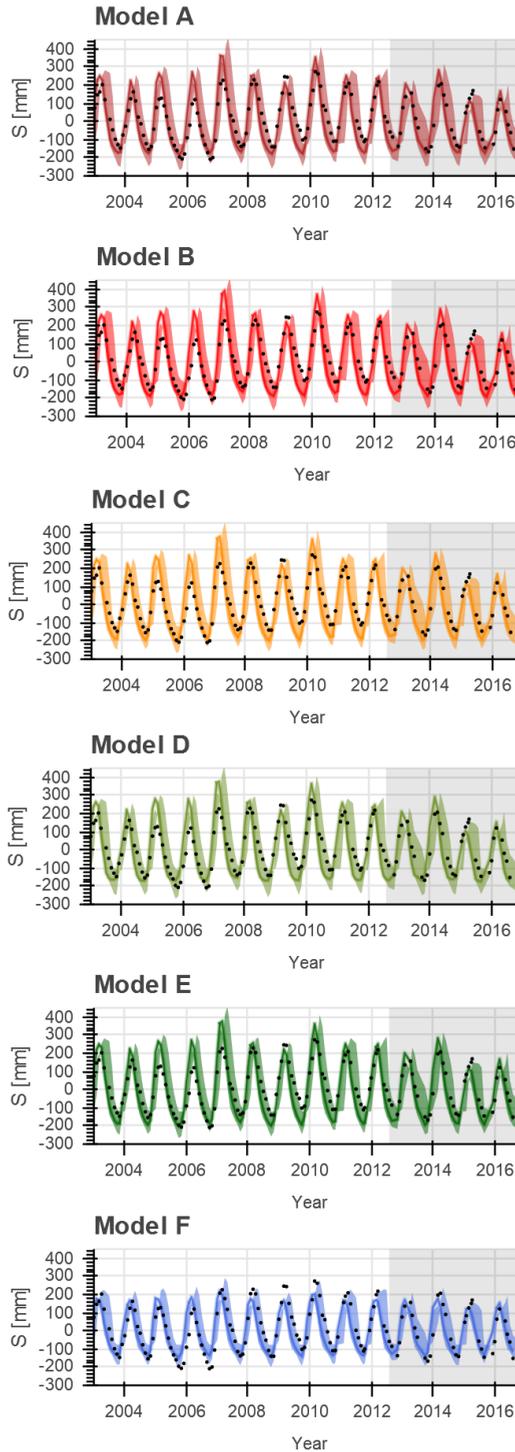


Figure B.5: Range of model solutions for Models A to F. Each panel shows the time series of the recorded (black) and modelled basin average total water storage: the line indicates the solution with the highest calibration objective function with respect to discharge ($D_{E,Qcal}$) and the shaded area of the solutions retained as feasible. The data in the grey shaded area were used for validation.

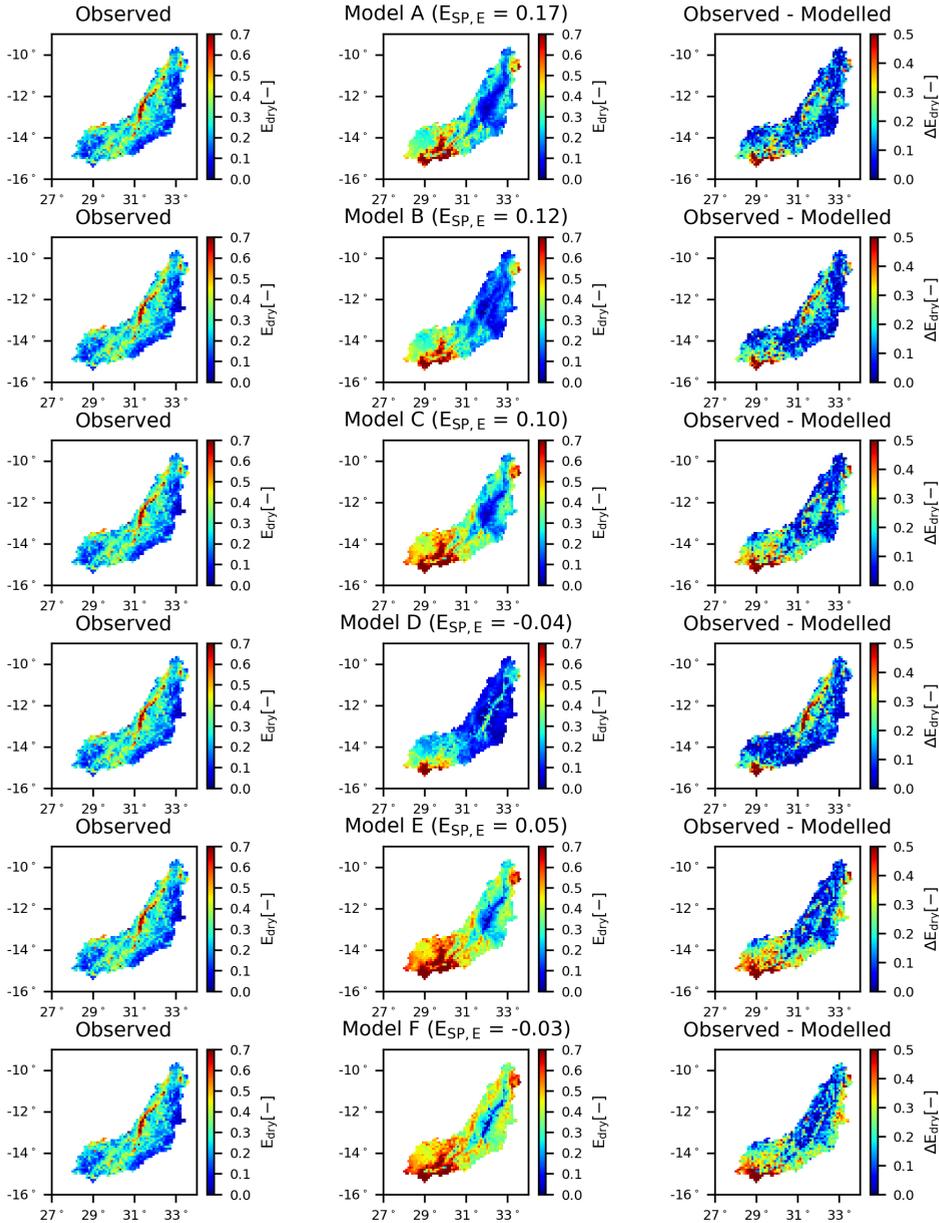


Figure B.6: Spatial variability of the normalised total evaporation for Models A to F averaged over all dry seasons. The left panel shows the observation according to WaPOR data; the middle panel the model result using the “optimal” parameter set with respect to discharge ($D_{E,Qcal}$); and the right panel the difference between the observation and model.

B

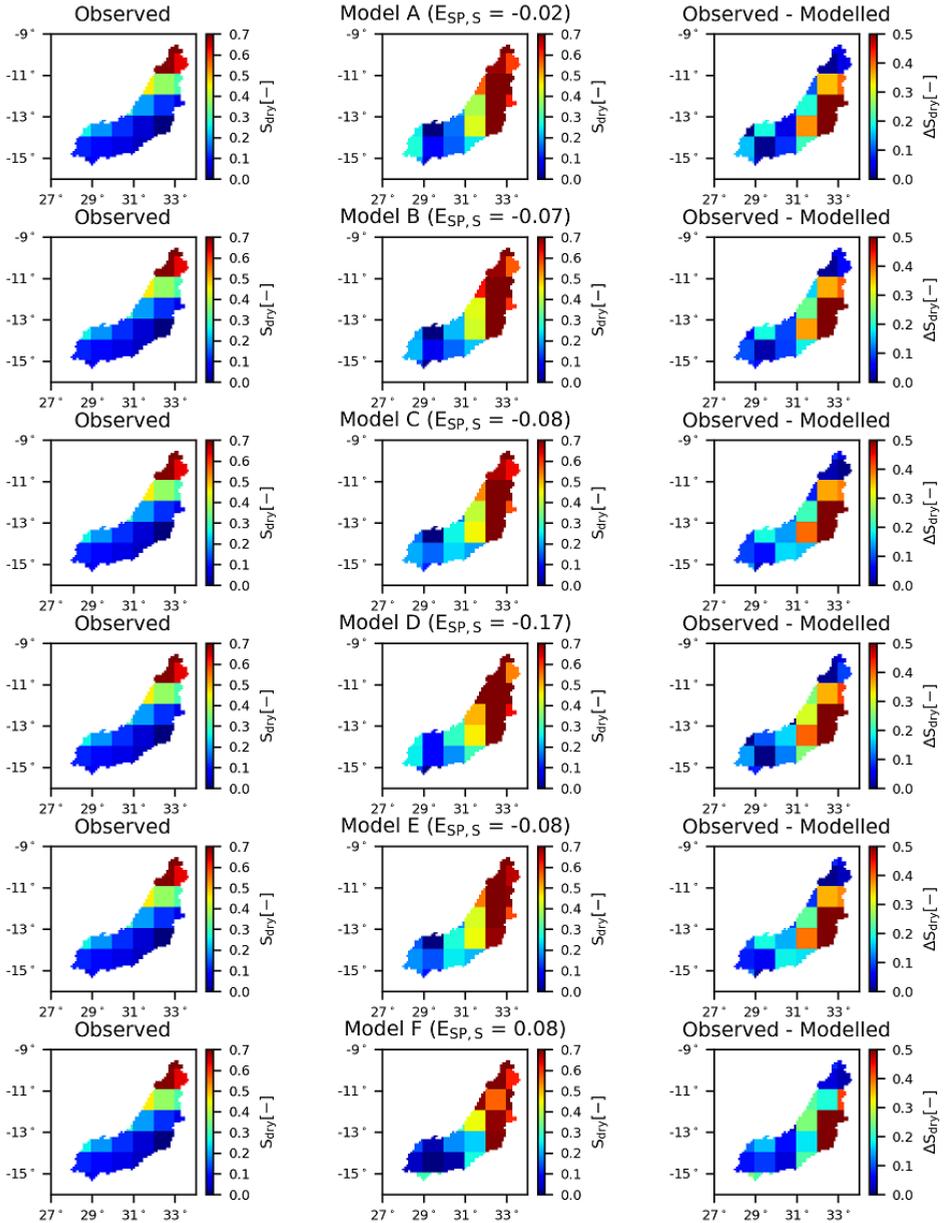


Figure B.7: Spatial variability of the normalised total water storage for Models A to F averaged over all dry seasons. The left panel shows the observation according to GRACE data; the middle panel the model result using the “optimal” parameter set with respect to discharge ($D_{E,Q_{cal}}$); and the right panel the difference between the observation and model.

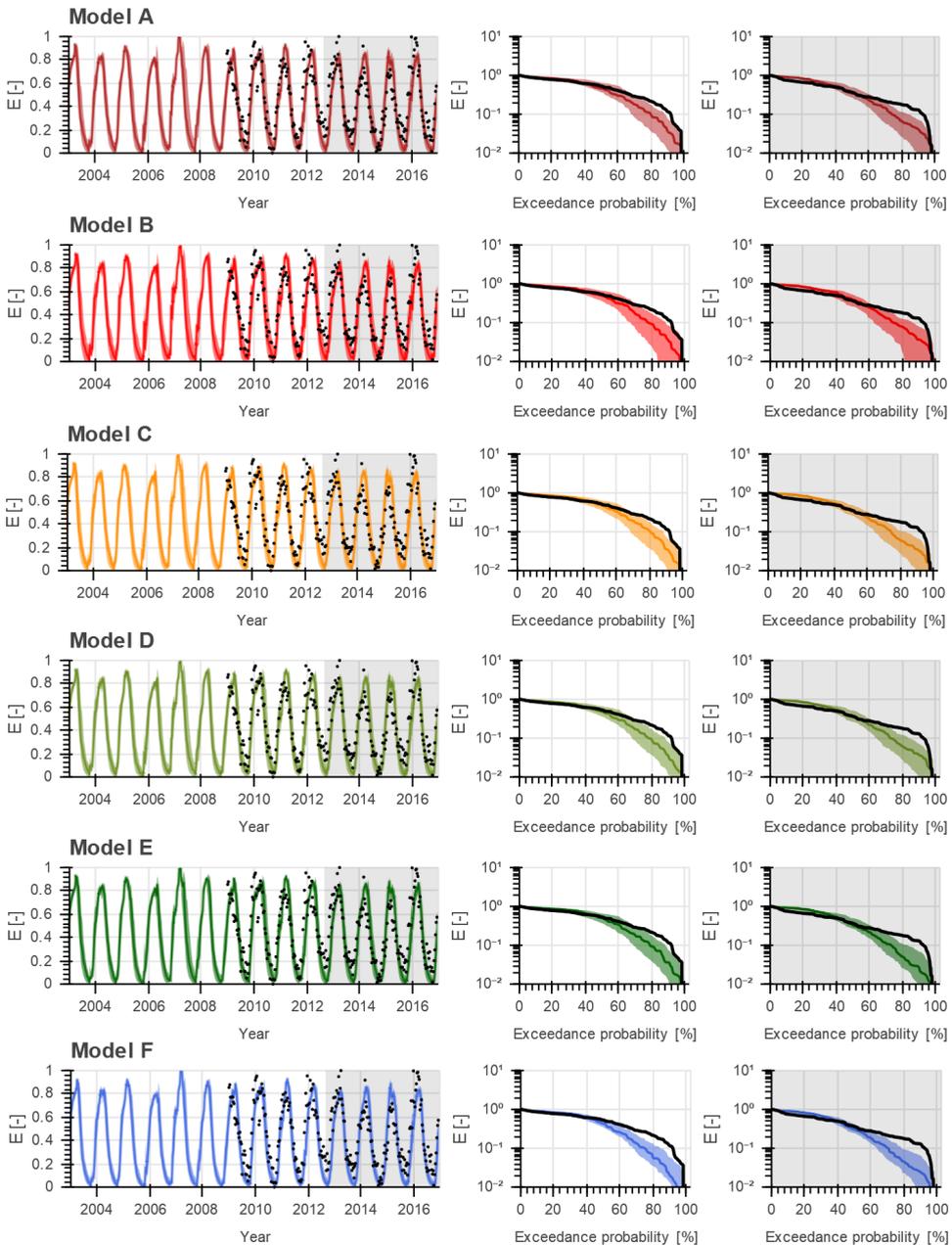


Figure B.8: Range of model solutions for Models A to F. The left panel shows the time series and the right panel the duration curve in logarithmic scale of the recorded (black) and modelled normalised basin average evaporation: the line indicates the solution with the highest calibration objective function with respect to multiple variables ($D_{E,ESQ}$) and the shaded area the envelope of the solutions retained as feasible. The data in the grey shaded area were used for validation.

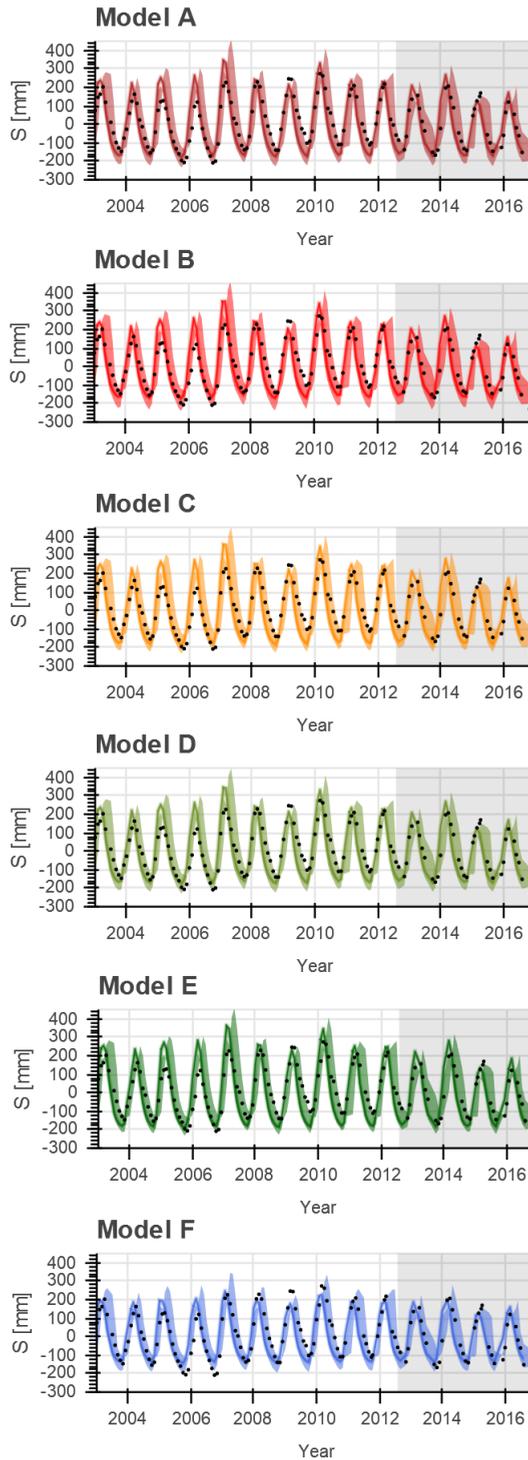


Figure B.9: Range of model solutions for Models A to F. Each panel shows the time series of the recorded (black) and modelled basin average total water storage: the line indicates the solution with the highest calibration objective function with respect to multiple variables ($D_{E,ESQ}$) and the shaded area the envelope of the solutions retained as feasible. The data in the grey shaded area were used for validation

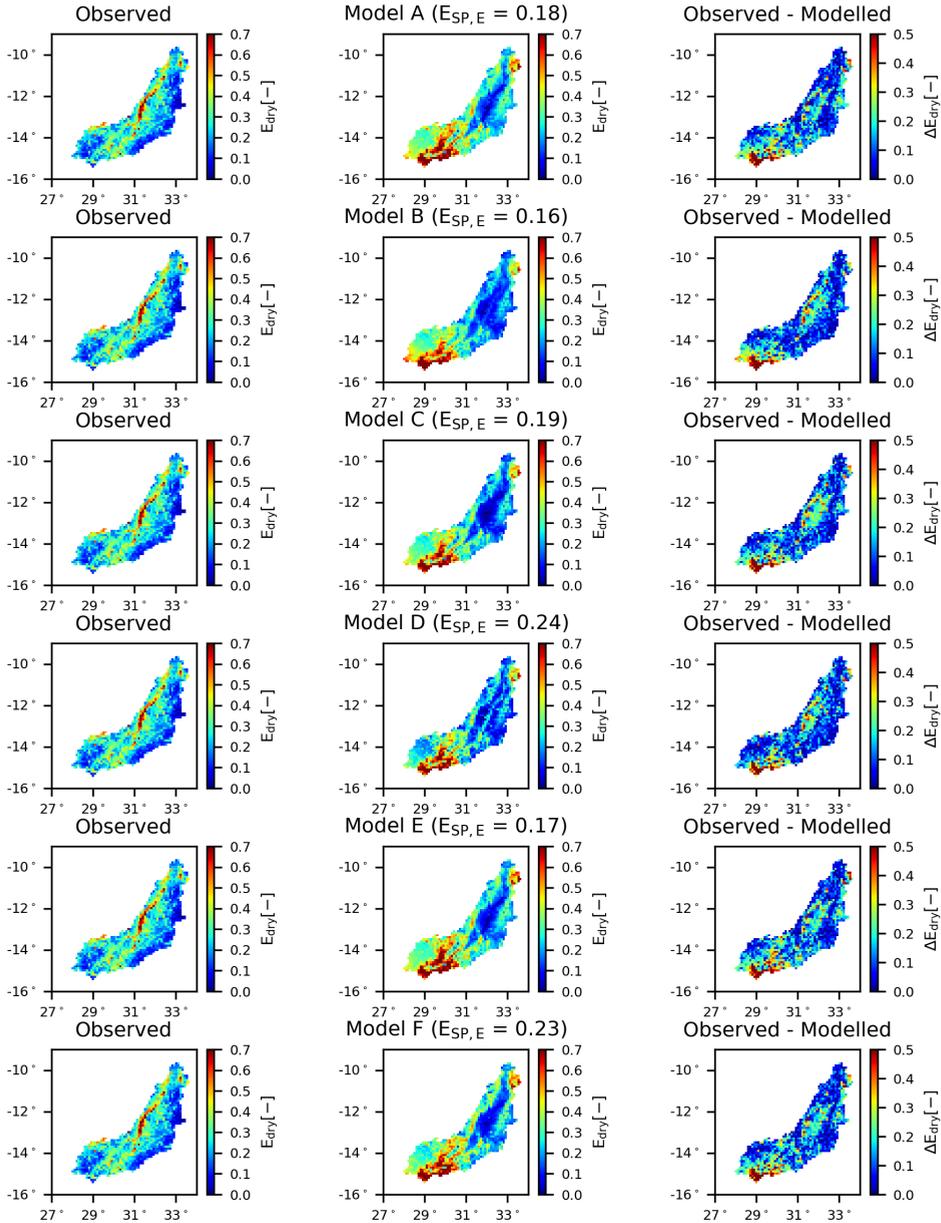


Figure B.10: Spatial variability of the normalised total evaporation for Models A to F averaged over all dry seasons. The left panel shows the observation according to WaPOR data; the middle panel the model result using the "optimal" parameter set with respect to multiple variables ($D_{E,ESQ}$); and the right panel the difference between the observation and model.

B

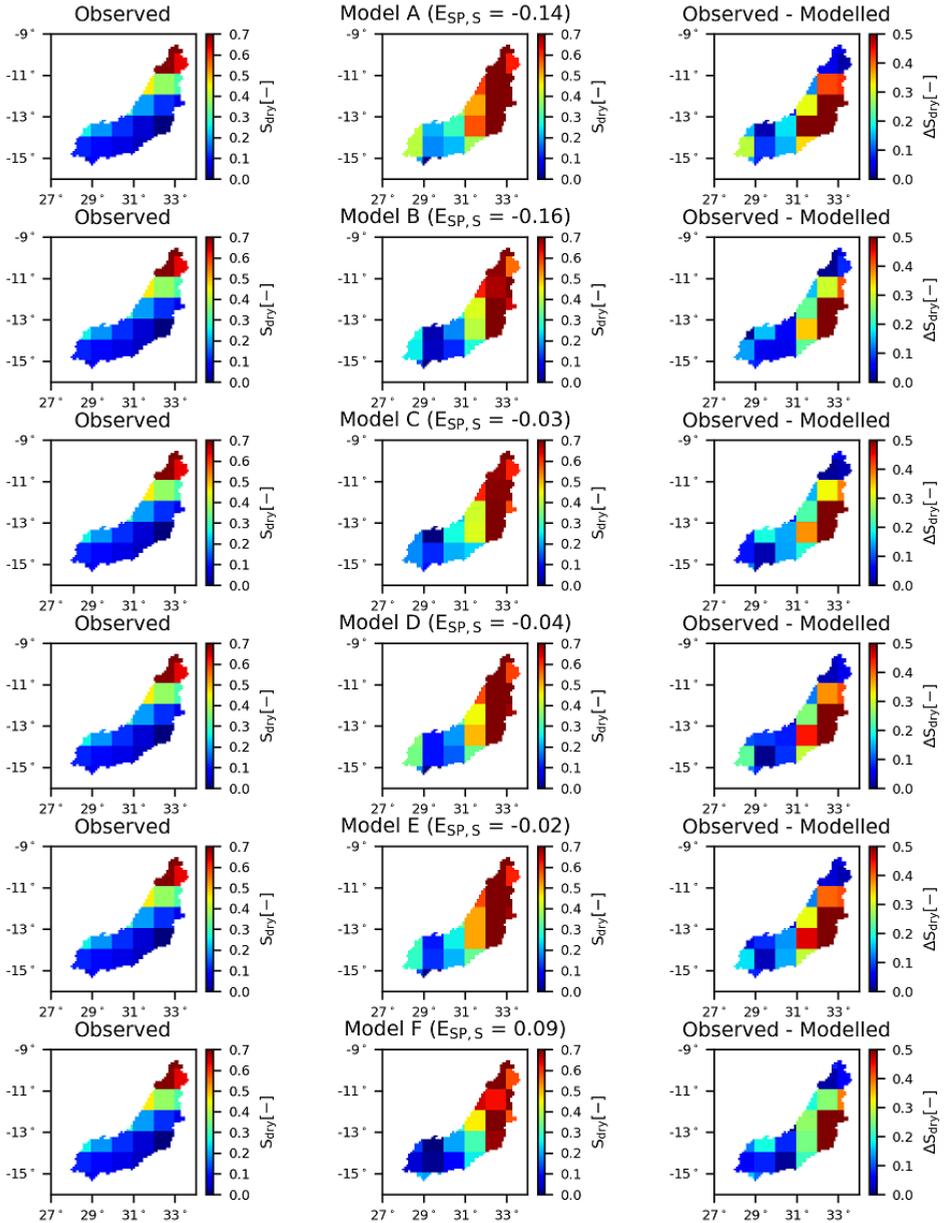


Figure B.11: Spatial variability of the normalised total water storage for Models A to F averaged over all dry seasons. The left panel shows the observation according to GRACE data; the middle panel the model result using the “optimal” parameter set with respect to multiple variables ($D_{E,ESQ}$); and the right panel the difference between the observation and model.

C

Appendix

This appendix provides supplementary material for Chapter 5.

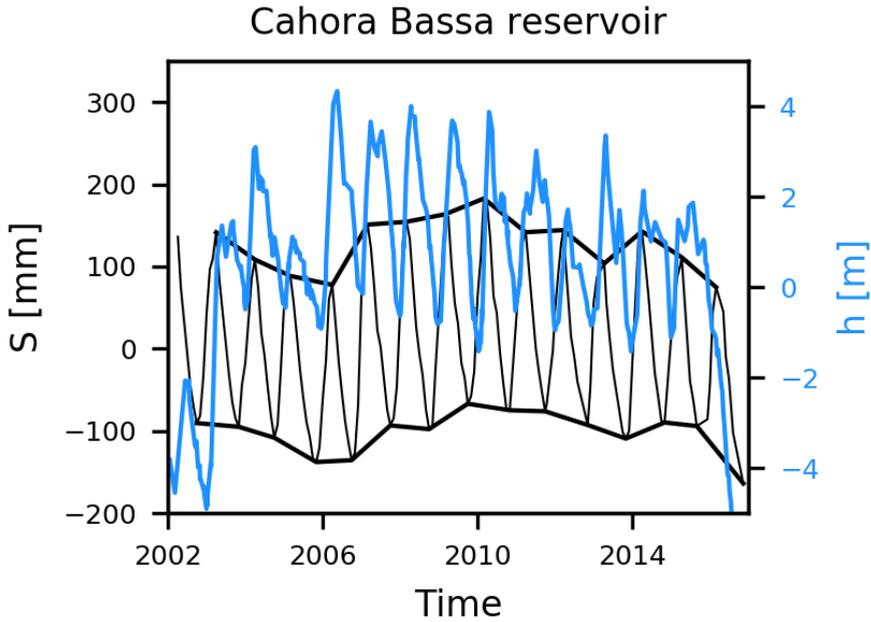


Figure C.1: Altimetry observations (blue) at Cahora Bassa reservoir and annual maximum/ minimum basin-average total water storage for the Luangwa basin (black).

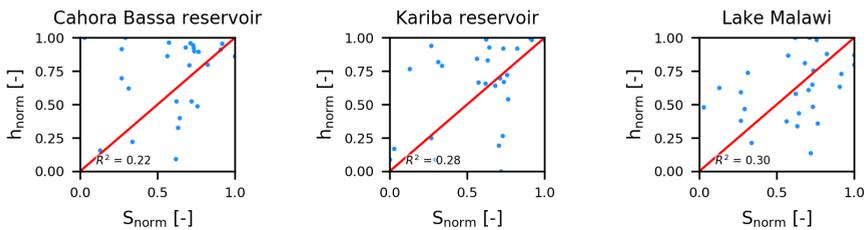
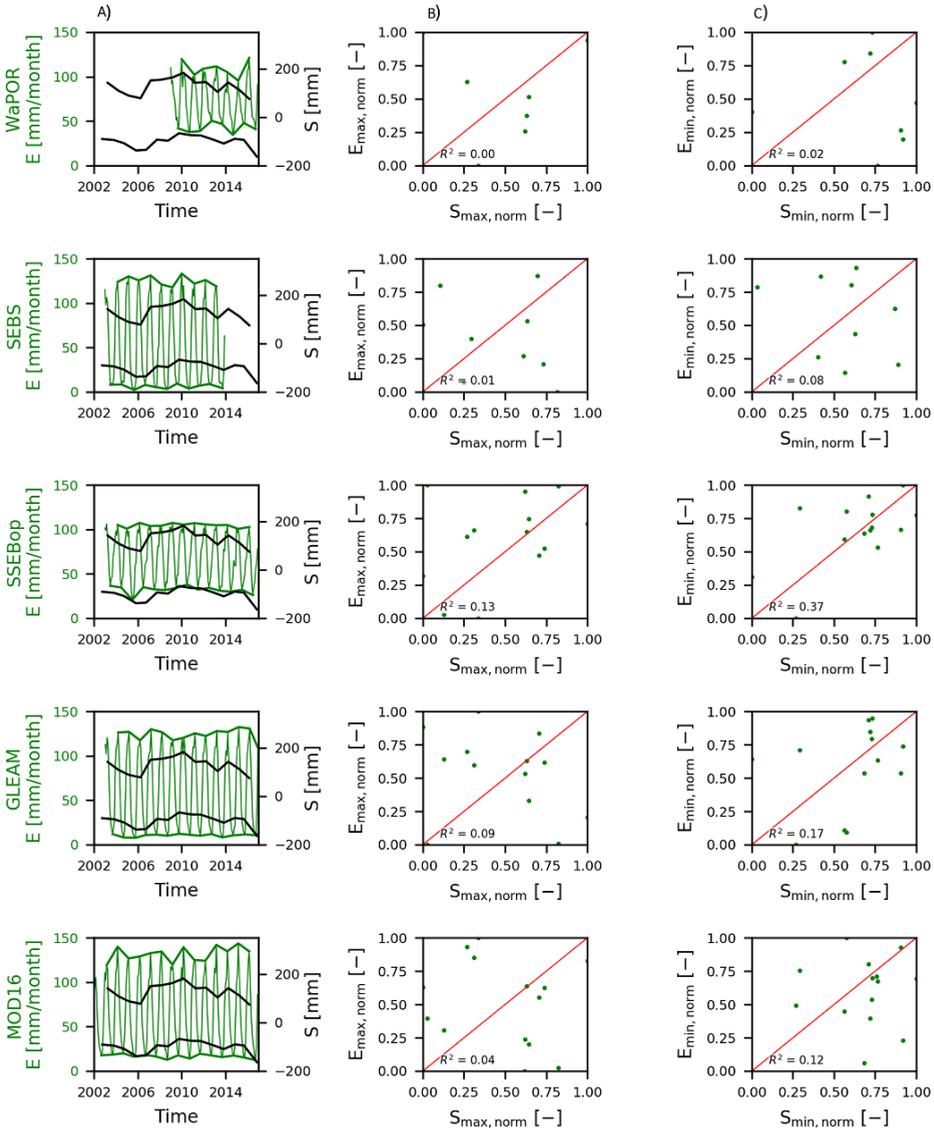


Figure C.2: Normalized annual maximum and minimum altimetry observations in large open water bodies near the Luangwa river basin and basin-average total water storage for the Luangwa basin. For the normalization, the following equation was applied: $X_{\text{norm}} = \frac{X - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}}$.



C

Figure C.3: A) Basin average total water storage (black) and monthly actual evaporation (green) according to WaPOR, SEBS, SSEBop, GLEAM and MOD16 for the Luangwa Basin. Annual maximum (B) and minimum (C) normalised total water storage (horizontal axis) and evaporation (vertical axis). For the normalization, the following equation was applied: $X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}}$.

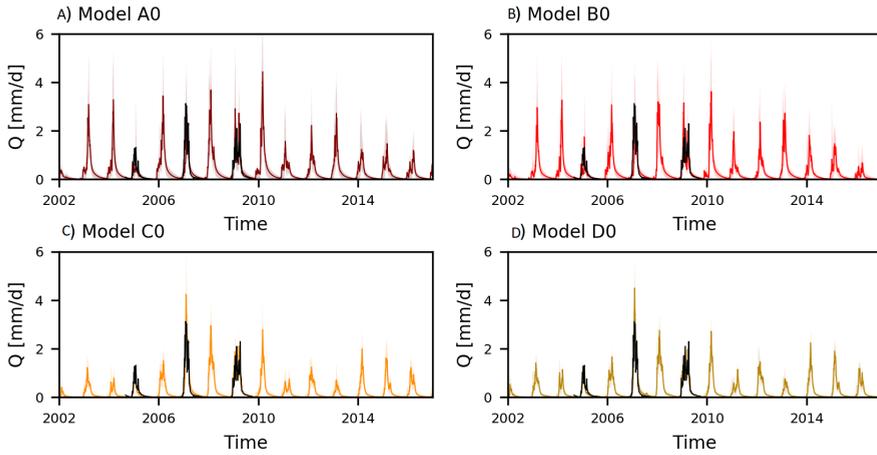


Figure C.4: Range of model solutions for Models A0 – D0 for calibration strategy 1 with respect to discharge (hydrograph). The black line indicates the recorded data, the coloured line the solution with the highest calibration objective function with respect to discharge ($D_{E,Q}$) and the shaded area the envelope of the solutions retained as feasible.

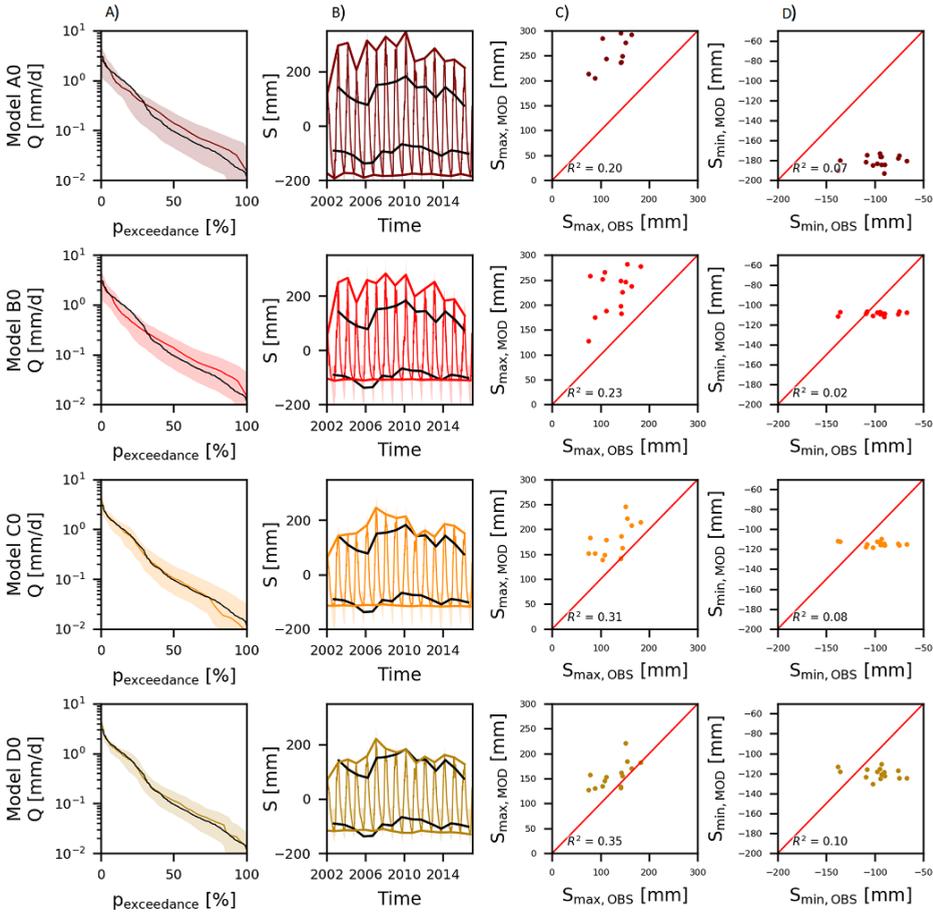


Figure C.5: Range of model solutions for Models A0 – D0 for calibration strategy 1 with respect to A) flow duration curve in logarithmic scale, B) total water storage time series, C) annual maximum total water storage, and D) annual minimum total water storage. In A) – B), the black line indicates the recorded data, the coloured line the solution with the highest calibration objective function with respect to discharge ($D_{E,Q}$) and the shaded area the envelope of the solutions retained as feasible. In C) – D), the recorded data are plotted on the horizontal axis and on the vertical axis the model solution with the highest calibration objective function with respect to discharge ($D_{E,Q}$). The red line indicates the 1:1 line.

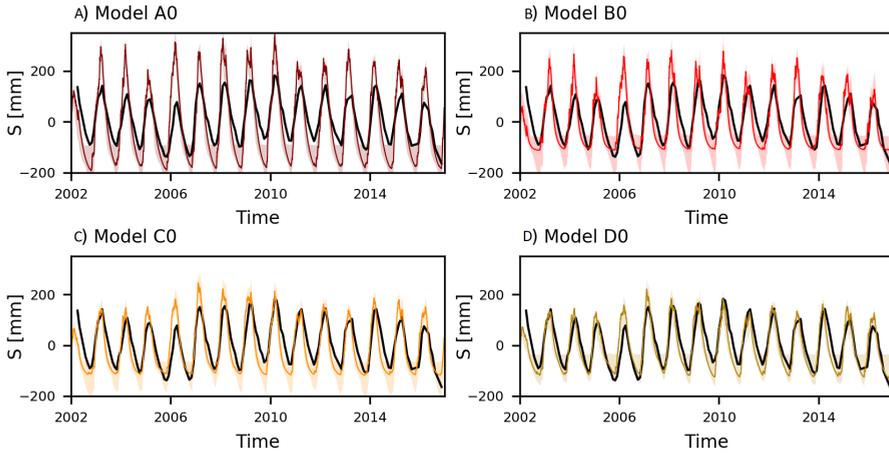


Figure C.6: Range of model solutions for Models A0 – D0 for calibration strategy 1 with respect to total water storage anomalies. The black line indicates the recorded data, the coloured line the solution with the highest calibration objective function with respect to discharge ($D_{E,Q}$) and the shaded area the envelope of the solutions retained as feasible.

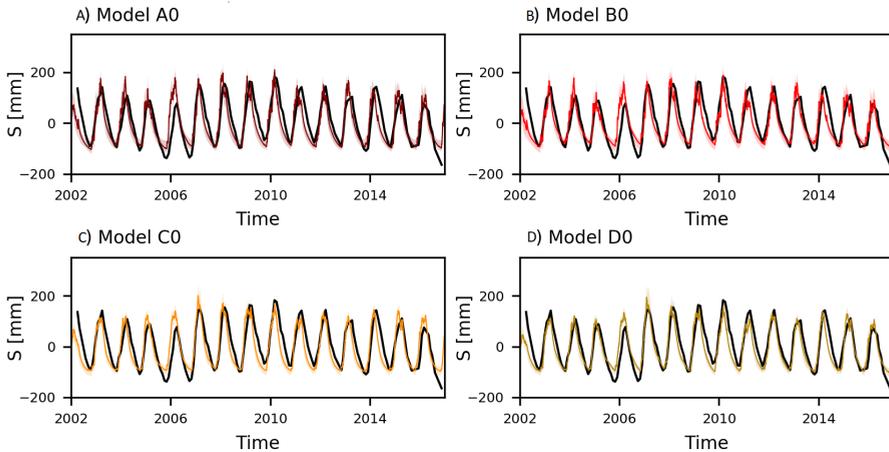


Figure C.7: Range of model solutions for Models A0 – D0 for calibration strategy 2 with respect to total water storage anomalies. The black line indicates the recorded data, the coloured line the solution with the highest calibration objective function with respect to discharge and total water storage ($D_{E,QS}$) and the shaded area the envelope of the solutions retained as feasible.

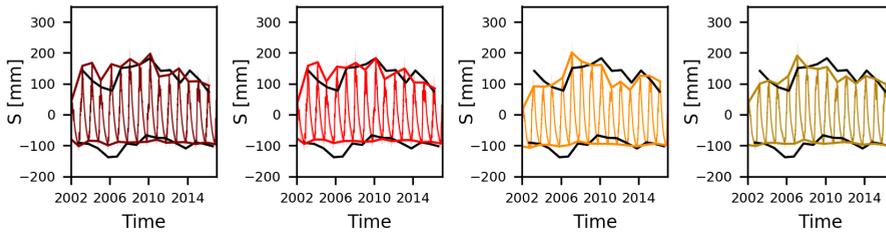


Figure C.8: Range of model solutions for Models A0 – D0 with respect to total water storage anomalies. The black line indicates the recorded data, the coloured line the solution with the highest calibration objective function with respect to discharge and total water storage ($D_{E,S}$) and the shaded area the envelope of the solutions retained as feasible.

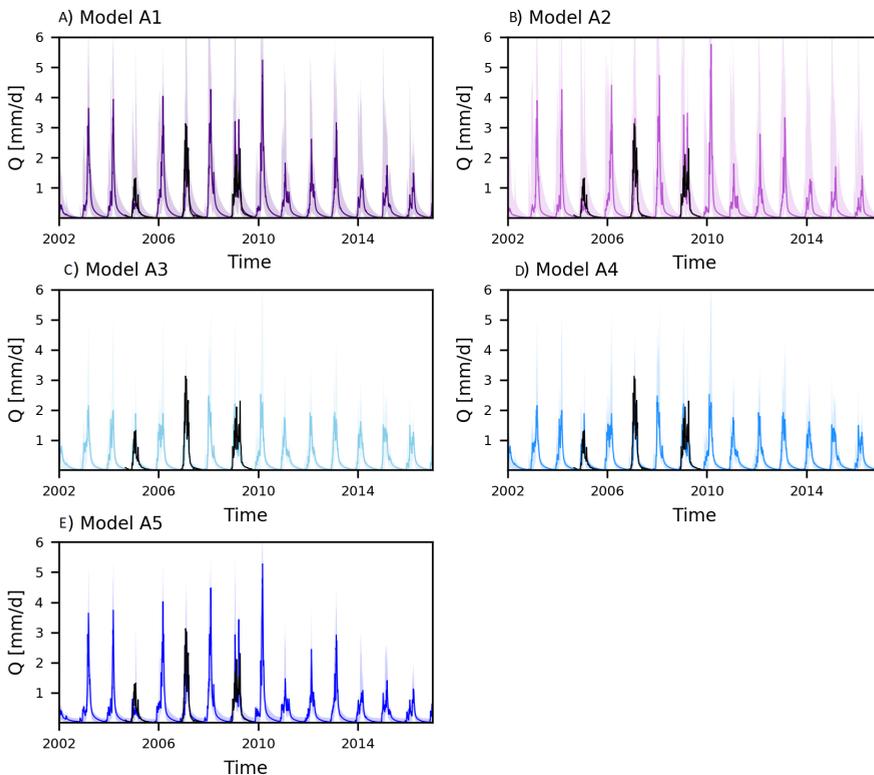


Figure C.9: Range of model solutions for Models A1 – A5 for calibration strategy 1 with respect to discharge (hydrograph). The black line indicates the recorded data, the coloured line the solution with the highest calibration objective function with respect to discharge ($D_{E,Q}$) and the shaded area the envelope of the solutions retained as feasible.

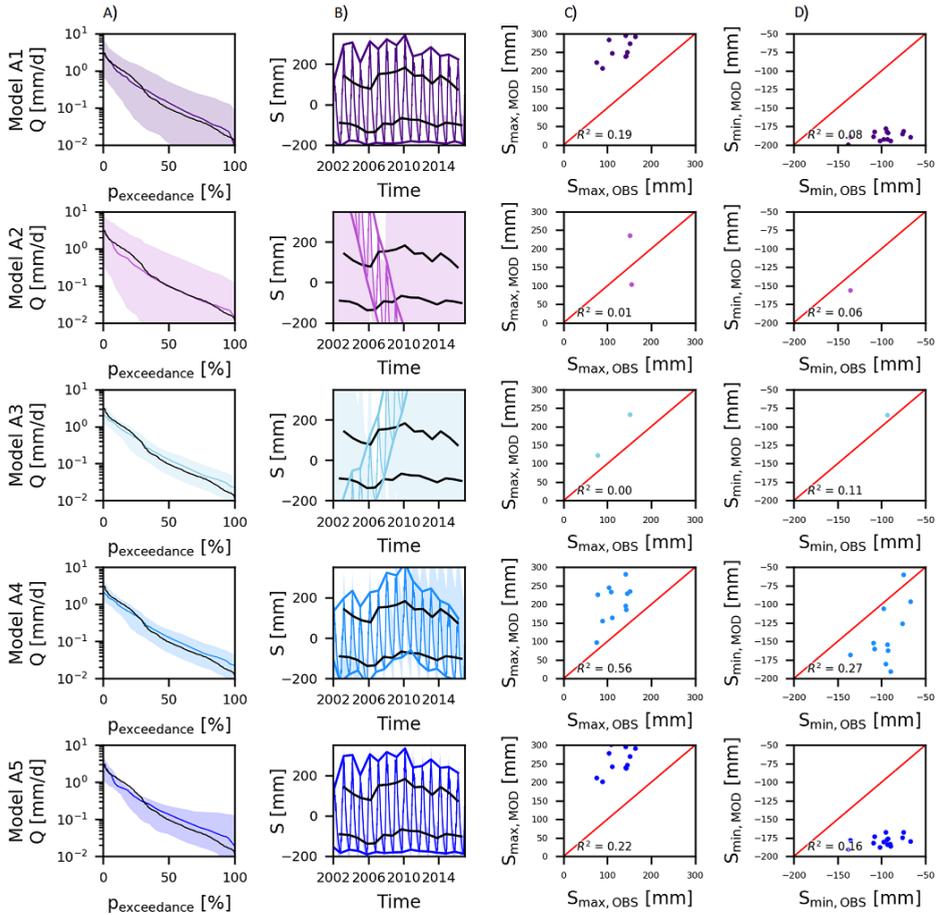
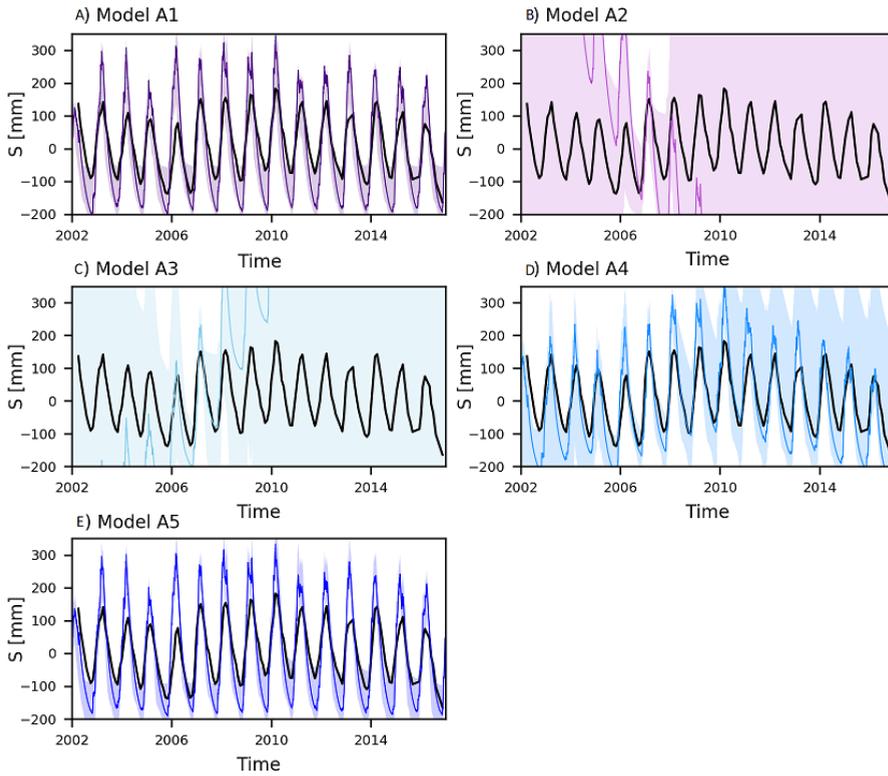


Figure C.10: Range of model solutions for Models A1 – A5 for calibration strategy 1 with respect to A) flow duration curve in logarithmic scale, B) total water storage time series, C) annual maximum total water storage, and D) annual minimum total water storage. In A) – B), the black line indicates the recorded data, the coloured line the solution with the highest calibration objective function with respect to discharge ($D_{E,Q}$) and the shaded area the envelope of the solutions retained as feasible. In C) – D), the recorded data are plotted on the horizontal axis and on the vertical axis the model solution with the highest calibration objective function with respect to discharge ($D_{E,Q}$). The red line indicates the 1:1 line.



C

Figure C.11: Range of model solutions for Models A1 – A5 for calibration strategy 1 with respect to total water storage anomalies. The black line indicates the recorded data, the coloured line the solution with the highest calibration objective function with respect to discharge ($D_{E,Q}$) and the shaded area the envelope of the solutions retained as feasible.

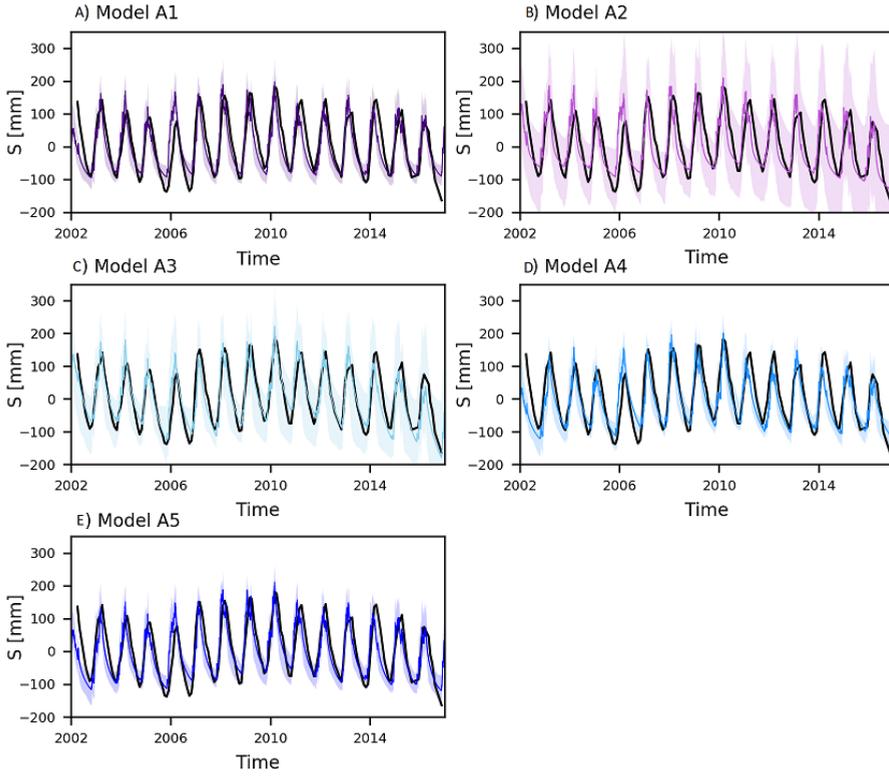


Figure C.12: Range of model solutions for Models A1 – A5 for calibration strategy 2 with respect to total water storage anomalies. The black line indicates the recorded data, the coloured line the solution with the highest calibration objective function with respect to discharge and total water storage ($D_{E, QS}$) and the shaded area the envelope of the solutions retained as feasible.

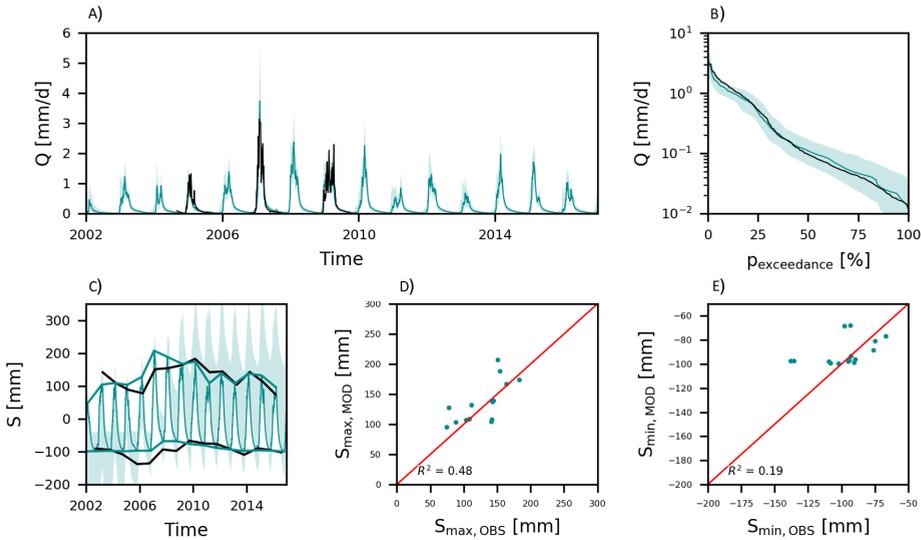


Figure C.13: Range of model solutions for Model D4 for calibration strategy 1 with respect to A) hydrograph, B) flow duration curve in logarithmic scale, C) total water storage time series, D) annual maximum total water storage, and E) annual minimum total water storage. In A) to C), the black line indicates the recorded data, the coloured line the solution with the highest calibration objective function with respect to discharge ($D_{E,Q}$) and the shaded area the envelope of the solutions retained as feasible. In D) and E), the recorded data are plotted on the horizontal axis and on the vertical axis the model solution with the highest calibration objective function with respect to discharge ($D_{E,Q}$). The red line indicates the 1:1 line.

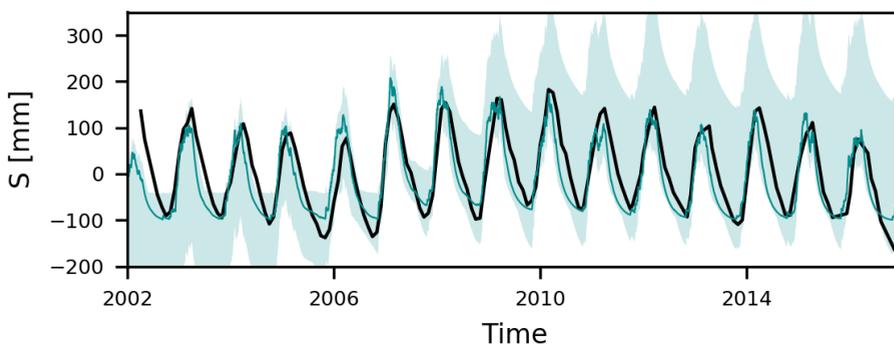


Figure C.14: Range of model solutions for Model D4 for calibration strategy 1 with respect to total water storage anomalies. The black line indicates the recorded data, the coloured line the solution with the highest calibration objective function with respect to discharge ($D_{E,Q}$) and the shaded area the envelope of the solutions retained as feasible.

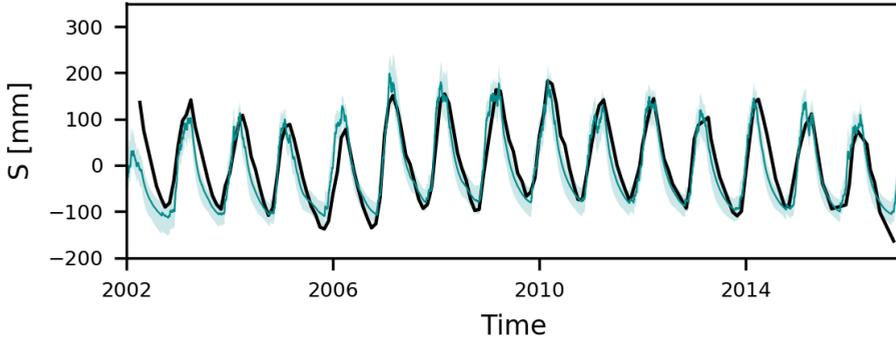


Figure C.15: Range of model solutions for Model D4 for calibration strategy 2 with respect to total water storage anomalies. The black line indicates the recorded data, the coloured line the solution with the highest calibration objective function with respect to discharge and total water storage ($D_{E,QS}$) and the shaded area the envelope of the solutions retained as feasible.

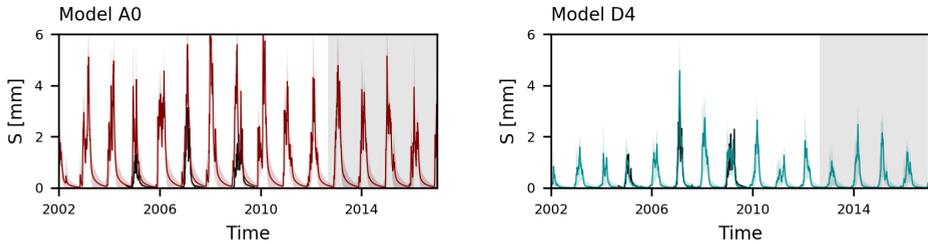


Figure C.16: Range of model solutions for Models A0 and D4 with respect to the hydrograph. The black line indicates the recorded data, the coloured line the solution with the highest calibration objective function with respect to discharge and total water storage ($D_{E,QS}$) for the time-period 2002 – 2012 (white area) and the shaded area the envelope of the solutions retained as feasible.

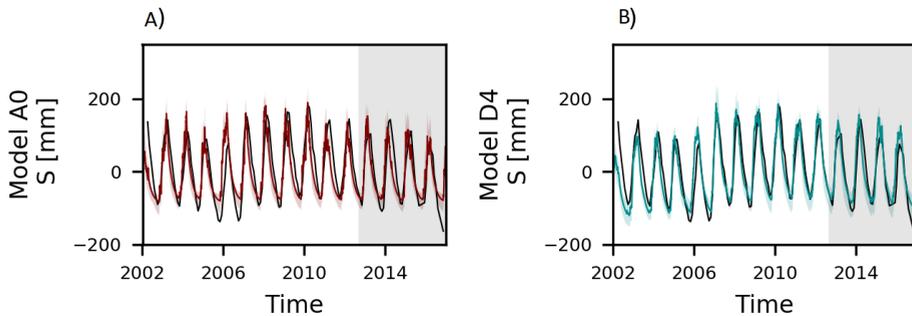


Figure C.17: Range of model solutions for Models A0 and D4 with respect to total water storage anomalies. The black line indicates the recorded data, the coloured line the solution with the highest calibration objective function with respect to discharge and total water storage ($D_{E,QS}$) for the time-period 2002 – 2012 (white area) and the shaded area the envelope of the solutions retained as feasible.

References

- Abas, I.: Remote river rating in Zambia: A case study in the Luangwa river basin, Thesis, Delft University of Technology, <http://resolver.tudelft.nl/uuid:22495ef4-7876-4bc5-bcf4-f639b097dcfb>, 2018.
- Ajami, N., Gupta, H., Wagener, T., and Sorooshian, S.: Calibration of a semi-distributed hydrologic model for streamflow estimation along a river system, *Journal of Hydrology*, 298, 112–135, doi: 10.1016/j.jhydrol.2004.03.033, 2004.
- Albertini, C., Mazzoleni, M., Totaro, V., Iacobellis, V., and Di Baldassarre, G.: Socio-Hydrological Modelling: The Influence of Reservoir Management and Societal Responses on Flood Impacts, *Water*, 12, doi: 10.3390/w12051384, 2020.
- Allen, R., Tasumi, M., and Trezza, R.: Satellite-Based Energy Balance for Mapping Evapotranspiration with Internalized Calibration (METRIC)—Model, *Journal of Irrigation and Drainage Engineering*, 133, 380–394, doi: 10.1061/(ASCE)0733-9437(2007)133:4(380), 2007.
- Arnold, J., Srinivasan, R., Muttiah, R., and Williams, J. R.: Large Area Hydrologic Modeling and Assessment Part 1: Model Development, *Journal of The American Water Resources Association*, 34, 73–89, doi: 10.1111/j.1752-1688.1998.tb05961.x, 1998.
- Awange, J. L., Khandu, Schumacher, M., Forootan, E., and Heck, B.: Exploring hydro-meteorological drought patterns over the Greater Horn of Africa (1979–2014) using remote sensing and reanalysis products, *Advances in Water Resources*, 94, 45–59, doi: 10.1016/j.advwatres.2016.04.005, 2016.
- Bai, M., Shen, B., Song, X., Mo, S., Huang, L., and Quan, Q.: Multi-Temporal Variabilities of Evapotranspiration Rates and Their Associations with Climate Change and Vegetation Greening in the Gan River Basin, China, *Water*, 11, 2568, doi: 10.3390/w11122568, 2019.
- Bai, P., Liu, X., and Liu, C.: Improving hydrological simulations by incorporating GRACE data for model calibration, *Journal of Hydrology*, 557, 291–304, doi: 10.1016/j.jhydrol.2017.12.025, 2018.
- Bastiaanssen, W. G. M., Menenti, M., Feddes, R. A., and Holtslag, A. A. M.: A remote sensing surface energy balance algorithm for land (SEBAL). 1. Formulation, *Journal of Hydrology*, 212–213, 198–212, doi: 10.1016/S0022-1694(98)00253-4, 1998.
- Bauer-Gottwein, P., Jensen, I. H., Guzinski, R., Bredtoft, G. K. T., Hansen, S., and Michailovsky, C. I.: Operational river discharge forecasting in poorly gauged basins: The Kavango River basin case study, *Hydrology and Earth System Sciences*, 19, 1469–1485, doi: 10.5194/hess-19-1469-2015, 2015.
- Bayissa, Y., Maskey, S., Tadesse, T., Van Andel, S. J., Moges, S., Van Griensven, A., and Solomatine, D.: Comparison of the Performance of Six Drought Indices in Characterizing Historical Drought for the Upper Blue Nile Basin, Ethiopia, *Geosciences*, 8, doi: 10.3390/geosciences8030081, 2018.
- Beilfuss, R., Moore, D., Bento, C., and Dutton, P.: Patterns of Hydrological Change in the Zambezi Delta, Mozambique, 2001.
- Bergström, S.: The HBV model – its structure and applications, 1992.
- Beven, K.: On doing better hydrological science, *Hydrological Processes*, 22, 3549–3553, doi: 10.1002/hyp.7108, 2008.

- Beven, K. J.: A manifesto for the equifinality thesis, *Journal of Hydrology*, 320, 18–36, doi: 10.1016/j.jhydrol.2005.07.007, 2006.
- Beven, K. J.: Preferential flows and travel time distributions: defining adequate hypothesis tests for hydrological process models, *Hydrological Processes*, 24, 1537–1547, doi: 10.1002/hyp.7718, 2010.
- Beven, K. J.: On hypothesis testing in hydrology: Why falsification of models is still a really good idea, *WIREs Water*, 5, e1278, doi: 10.1002/wat2.1278, 2018.
- Beven, K. J. and Westerberg, I.: On red herrings and real herrings: disinformation and information in hydrological inference, *Hydrological Processes*, 25, 1676–1680, doi: 10.1002/hyp.7963, 2011.
- Biancamaria, S., Lettenmaier, D. P., and Pavelsky, T. M.: The SWOT Mission and Its Capabilities for Land Hydrology, *Surveys in Geophysics*, 37, 307–337, doi: 10.1007/s10712-015-9346-y, 2016.
- Biancamaria, S., Frappart, F., Leleu, A. S., Marieu, V., Blumstein, D., Desjonquères, J.-D., Boy, F., Sottolichio, A., and Valle-Levinson, A.: Satellite radar altimetry water elevations performance over a 200m wide river: Evaluation over the Garonne River, *Advances in Space Research*, 59, 128–146, doi: 10.1016/j.asr.2016.10.008, 2017.
- Bieger, K., Arnold, J. G., Rathjens, H., White, M. J., Bosch, D. D., Allen, P. M., Volk, M., and Srinivasan, R.: Introduction to SWAT+, A Completely Restructured Version of the Soil and Water Assessment Tool, *JAWRA Journal of the American Water Resources Association*, 53, 115–130, doi: 10.1111/1752-1688.12482, 2017.
- Birkett, C. M.: Contribution of the TOPEX NASA Radar Altimeter to the global monitoring of large rivers and wetlands, *Water Resources Research*, 34, 1223–1239, doi: 10.1029/98WR00124, 1998.
- Blazquez, A., Meyssignac, B., Lemoine, J. M., Berthier, E., Ribes, A., and Cazenave, A.: Exploring the uncertainty in GRACE estimates of the mass redistributions at the Earth surface: implications for the global water and sea level budgets, *Geophysical Journal International*, 215, 415–430, doi: 10.1093/gji/ggy293, 2018a.
- Blazquez, A., Meyssignac, B., Lemoine, J. M., Berthier, E., Ribes, A., and Cazenave, A.: Exploring the uncertainty in GRACE estimates of the mass redistributions at the Earth surface: implications for the global water and sea level budgets, *Geophysical Journal International*, 215, 415–430, doi: 10.1093/gji/ggy293, 2018b.
- Bloomfield, J. P. and Marchant, B. P.: Analysis of groundwater drought building on the standardised precipitation index approach, *Hydrol. Earth Syst. Sci.*, 17, 4769–4787, doi: 10.5194/hess-17-4769-2013, 2013.
- Blöschl, G. and Sivapalan, M.: Scale issues in hydrological modelling: A review, *Hydrological Processes*, 9, 251–290, doi: 10.1002/hyp.3360090305, 1995.
- Blöschl, G., Sivapalan, M., Wagener, T., Viglione, A., and Savenije, H.: *Runoff Prediction in Ungauged Basins: Synthesis across Processes, Places and Scales*, Cambridge University Press, <https://books.google.co.zm/books?id=8Y0YAAAAQBAJ>, 2013.
- Bola, G., Mabiza, C., Goldin, J., Kujinga, K., Nhapi, I., Makurira, H., and Mashauri, D.: Coping with droughts and floods: A Case study of Kanyemba, Mbire District, Zimbabwe, *Physics and Chemistry of the Earth, Parts A/B/C*, 67-69, 180–186, doi: 10.1016/j.pce.2013.09.019, 2014.
- Bonsor, H. C., Shamsudduha, M., Marchant, B. P., MacDonald, A. M., and Taylor, R. G.: Seasonal and Decadal Groundwater Changes in African Sedimentary Aquifers Estimated Using GRACE Products and LSMs, *Remote Sensing*, 10, 904, doi: 10.3390/rs10060904, 2018.
- Bouaziz, L. J. E., Weerts, A., Schellekens, J., Sprokkereef, E., Stam, J., Savenije, H., and

- Hrachowitz, M.: Redressing the balance: quantifying net intercatchment groundwater flows, *Hydrol. Earth Syst. Sci.*, 22, 6415–6434, doi: 10.5194/hess-22-6415-2018, 2018.
- Bouaziz, L. J. E., Steele-Dunne, S. C., Schellekens, J., Weerts, A. H., Stam, J., Sprokkereef, E., Winsemius, H. H. C., Savenije, H. H. G., and Hrachowitz, M.: Improved Understanding of the Link Between Catchment-Scale Vegetation Accessible Storage and Satellite-Derived Soil Water Index, *Water Resources Research*, 56, e2019WR026365, doi: 10.1029/2019WR026365, 2020.
- Bourdin, D. R., Fleming, S. W., and Stull, R. B.: Streamflow Modelling: A Primer on Applications, Approaches and Challenges, *Atmosphere-Ocean*, 50, 507–536, doi: 10.1080/07055900.2012.734276, 2012.
- Boutt, D. F.: Assessing hydrogeologic controls on dynamic groundwater storage using long-term instrumental records of water table levels, *Hydrological Processes*, 31, 1479–1497, doi: 10.1002/hyp.11119, 2017.
- Brown, W.: Worst southern African drought in decades kills wildlife and slows Victoria Falls, <https://www.thetimes.co.uk/article/worst-southern-african-drought-in-decades-kills-wildlife-and-slows-victoria-falls-5081m82dx>, 2019.
- Budyko, M. I.: *Climate and Life*, Academic Press, New York, 1974.
- Burnash, R. J. C., Ferral, R. L., and McGuire, R. A.: A generalized streamflow simulation system: conceptual modeling for digital computers, 1973.
- Calmant, S., Seyler, F., and Cretaux, J.: Monitoring Continental Surface Waters by Satellite Altimetry, vol. 29, *Surveys in Geophysics*, doi: 10.1007/s10712-008-9051-1, 2009.
- Campbell, B.: The Miombo in transition: woodlands and welfare in Africa, Center for International Forestry Research (CIFOR), Malaysia, doi: 10.17528/cifor/000465, 1996.
- Carlowicz, M.: Water Levels Keep Falling at Lake Kariba, <https://earthobservatory.nasa.gov/images/146068/water-levels-keep-falling-at-lake-kariba>, 2019.
- Chao, N., Wang, Z., Jiang, W., and Chao, D.: A quantitative approach for hydrological drought characterization in southwestern China using GRACE, *Hydrogeology Journal*, 24, 893–903, doi: 10.1007/s10040-015-1362-y, 2016.
- Chen, J. L., Wilson, C. R., Tapley, B. D., Longuevergne, L., Yang, Z. L., and Scanlon, B. R.: Recent La Plata basin drought conditions observed by satellite gravimetry, *Journal of Geophysical Research: Atmospheres*, 115, D22 108, doi: 10.1029/2010JD014689, 2010.
- Chen, J. L., Wilson, C. R., Tapley, B. D., Scanlon, B., and Güntner, A.: Long-term groundwater storage change in Victoria, Australia from satellite gravity and in situ observations, *Global and Planetary Change*, 139, 56–65, doi: 10.1016/j.gloplacha.2016.01.002, 2016.
- Childs, J. W.: Water Levels at World's Largest Man-Made Reservoir Drop to Lowest Since 1996, <https://weather.com/news/news/2019-12-11-lowest-water-levels-kariba-dam>, 2019.
- Chow, V. T.: *Open-channel hydraulics*, McGraw-Hill, New York, 1959.
- Clark, M. P., Rupp, D. E., Woods, R. A., Zheng, X., Ibbitt, R. P., Slater, A. G., Schmidt, J., and Uddstrom, M. J.: Hydrological data assimilation with the ensemble Kalman filter: Use of streamflow observations to update states in a distributed hydrological model, *Advances in Water Resources*, 31, 1309–1324, doi: 10.1016/j.advwatres.2008.06.005, 2008.
- Clark, M. P., Kavetski, D., and Fenicia, F.: Pursuing the method of multiple working hypotheses for hydrological modeling, *Water Resources Research*, 47, W09 301, doi: 10.1029/2010WR009827, 2011.
- Clark, M. P., Nijssen, B., Lundquist, J. D., Kavetski, D., Rupp, D. E., Woods, R. A., Freer, J. E., Gutmann, E. D., Wood, A. W., Gochis, D. J., Rasmussen, R. M., Tarboton, D. G., Mahat,

- V., Flerchinger, G. N., and Marks, D. G.: A unified approach for process-based hydrologic modeling: 2. Model implementation and case studies, *Water Resources Research*, 51, 2515–2542, doi: 10.1002/2015WR017200, 2015.
- Clark, M. P., Schaeffli, B., Schymanski, S. J., Samaniego, L., Luce, C. H., Jackson, B. M., Freer, J. E., Arnold, J. R., Moore, R. D., Istanbuluoglu, E., and Ceola, S.: Improving the theoretical underpinnings of process-based hydrologic models, *Water Resources Research*, 52, 2350–2365, doi: 10.1002/2015WR017910, 2016.
- Claverie, M., Vermote, E., and NOAA CDR Program: doi: doi:10.7289/V5M043BX, 2014.
- CNES: AVISO+ Satellite Altimetry Data, www.avisio.altimetry.fr, Accessed 2018.
- Cohen Liechti, T., Matos, J. P., Boillat, J. L., and Schleiss, A. J.: Comparison and evaluation of satellite derived precipitation products for hydrological modeling of the Zambezi River Basin, *Hydrol. Earth Syst. Sci.*, 16, 489–500, doi: 10.5194/hess-16-489-2012, 2012.
- Copernicus Climate Change Service (C3S): ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate. Copernicus Climate Change Service Climate Data Store (CDS), <https://cds.climate.copernicus.eu/cdsapp#!/home>, 2017.
- Crétau, J.-F. and Birkett, C.: Lake studies from satellite radar altimetry, *Comptes Rendus Geoscience*, 338, 1098–1112, doi: 10.1016/j.crte.2006.08.002, 2006.
- Crétau, J.-F., Nielsen, K., Frappart, F., Papa, F., Calmant, S., and Benveniste, J.: Hydrological applications of satellite altimetry: rivers, lakes, man-made reservoirs, inundated areas, p. 459–504, CRC Press, Boca Raton, FL, USA, 2017.
- Cuo, L.: Land Use/Cover Change Impacts on Hydrology in Large River Basins, pp. 103–134, *Geophysical Monograph Series (Book 221)*, American Geophysical Union and John Wiley & Sons, Washington, D.C., doi: 10.1002/9781118971772.ch6, 2016.
- Daggupati, P., Yen, H., White, M. J., Srinivasan, R., Arnold, J. G., Keitzer, C. S., and Sowa, S. P.: Impact of model development, calibration and validation decisions on hydrological simulations in West Lake Erie Basin, *Hydrological Processes*, 29, 5307–5320, doi: 10.1002/hyp.10536, 2015.
- Danielson, J. J. and Gesch, D. B.: Global multi-resolution terrain elevation data 2010 (GMTED2010), Report 2011-1073, U.S. Geological Survey, doi: 10.3133/ofr20111073, 2011.
- de Oliveira Campos, I., Mercier, F., Maheu, C., Cochonneau, G., Kosuth, P., Blitzkow, D., and Cazenave, A.: Temporal variations of river basin waters from Topex/Poseidon satellite altimetry. Application to the Amazon basin, *Comptes Rendus de l'Académie des Sciences - Series IIA - Earth and Planetary Science*, 333, 633–643, doi: 10.1016/S1251-8050(01)01688-3, 2001.
- Dembélé, M., Hrachowitz, M., Savenije, H. H. G., Mariéthoz, G., and Schaeffli, B.: Improving the Predictive Skill of a Distributed Hydrological Model by Calibration on Spatial Patterns With Multiple Satellite Data Sets, *Water Resources Research*, 56, e2019WR026085, doi: 10.1029/2019WR026085, 2020.
- Demirel, M. C., Mai, J., Mendiguren, G., Koch, J., Samaniego, L., and Stisen, S.: Combining satellite data and appropriate objective functions for improved spatial pattern performance of a distributed hydrologic model, *Hydrology and Earth System Sciences*, 22, 1299–1315, doi: 10.5194/hess-22-1299-2018, 2018.
- Di Baldassarre, G., Martinez, F., Kalantari, Z., and Viglione, A.: Drought and flood in the Anthropocene: feedback mechanisms in reservoir operation, *Earth Syst. Dynam.*, 8, 225–233, doi: 10.5194/esd-8-225-2017, 2017.
- Dinku, T., Funk, C., Peterson, P., Maidment, R., Tadesse, T., Gadain, H., and Ceccato, P.: Validation of the CHIRPS satellite rainfall estimates over eastern Africa, *Quarterly Journal of the Royal Meteorological Society*, 144, 292–312, doi: 10.1002/qj.3244, 2018.

- Domeneghetti, A.: On the use of SRTM and altimetry data for flood modeling in data-sparse regions, *Water Resources Research*, 52, 2901–2918, doi: 10.1002/2015WR017967, 2016.
- Domeneghetti, A., Castellarin, A., and Brath, A.: Assessing rating-curve uncertainty and its effects on hydraulic model calibration, *Hydrology and Earth System Sciences*, 16, 1191–1202, doi: 10.5194/hess-16-1191-2012, 2012.
- Domeneghetti, A., Castellarin, A., Tarpanelli, A., and Moramarco, T.: Investigating the uncertainty of satellite altimetry products for hydrodynamic modelling, *Hydrological Processes*, 29, 4908–4918, doi: 10.1002/hyp.10507, 2015.
- Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., and Bargellini, P.: Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services, *Remote Sensing of Environment*, 120, 25–36, doi: 10.1016/j.rse.2011.11.026, 2012.
- Dutra, E., Di Giuseppe, F., Wetterhall, F., and Pappenberger, F.: Seasonal forecasts of droughts in African basins using the Standardized Precipitation Index, *Hydrol. Earth Syst. Sci.*, 17, 2359–2373, doi: 10.5194/hess-17-2359-2013, 2013.
- Edel, D.: Millions Affected by Victoria Falls Severe Drought, *Intelligent Living*, <https://www.intelligentliving.co/millions-affected-by-victoria-falls-severe-drought/>, 2019.
- EM-DAT and CRED: The International Disaster Database, www.emdat.be, Accessed 2020.
- Entwistle, N. S. and Heritage, G. L.: Small unmanned aerial model accuracy for photogrammetrical fluvial bathymetric survey, *Journal of Applied Remote Sensing*, 13, 1–19, doi: 10.1117/1.JRS.13.014523, 2019.
- ESA: Satellite Missions Database, <https://directory.eoportal.org/web/eoportal/satellite-missions>, Accessed 2018.
- Euser, T., Winsemius, H. C., Hrachowitz, M., Fenicia, F., Uhlenbrook, S., and Savenije, H. H. G.: A framework to assess the realism of model structures using hydrological signatures, *Hydrology and Earth System Sciences*, 17, 1893–1912, doi: 10.5194/hess-17-1893-2013, 2013.
- Euser, T., Hrachowitz, M., Winsemius, H. C., and Savenije, H. H. G.: The effect of forcing and landscape distribution on performance and consistency of model structures, *Hydrological Processes*, 29, 3727–3743, doi: 10.1002/hyp.10445, 2015.
- Fan, Y., Miguez-Macho, G., Jobbágy, E. G., Jackson, R. B., and Otero-Casal, C.: Hydrologic regulation of plant rooting depth, *Proceedings of the National Academy of Sciences*, 114, 10572, doi: 10.1073/pnas.1712381114, 2017.
- Fang, K., Shen, C., Fisher, J. B., and Niu, J.: Improving Budyko curve-based estimates of long-term water partitioning using hydrologic signatures from GRACE, *Water Resources Research*, 52, 5537–5554, doi: 10.1002/2016WR018748, 2016.
- FAO: WaPOR Database Methodology: Level 1. Remote Sensing for Water Productivity Technical Report: Methodology Series, 2018.
- FAO and IHE Delft: WaPOR quality assessment. Technical report on the data quality of the WaPOR FAO database version 1.0, Report, <http://www.fao.org/3/ca4895en/CA4895EN.pdf>, 2019.
- Feng, T., Su, T., Zhi, R., Tu, G., and Ji, F.: Assessment of actual evapotranspiration variability over global land derived from seven reanalysis datasets, *International Journal of Climatology*, 39, 2919–2932, doi: 10.1002/joc.5992, 2019.
- Fenicia, F., McDonnell, J. J., and Savenije, H. H. G.: Learning from model improvement: On the contribution of complementary data to process understanding, *Water Resources Research*, 44, doi: 10.1029/2007WR006386, 2008.
- Fenicia, F., Kavetski, D., and Savenije, H. H. G.: Elements of a flexible approach for concep-

- tual hydrological modeling: 1. Motivation and theoretical development, *Water Resources Research*, 47, doi: 10.1029/2010WR010174, 2011.
- Fenicia, F., Kavetski, D., Savenije, H. H. G., Clark, M. P., Schoups, G., Pfister, L., and Freer, J.: Catchment properties, function, and conceptual model representation: is there a correspondence?, *Hydrological Processes*, 28, 2451–2467, doi: 10.1002/hyp.9726, 2014.
- Fenicia, F., Kavetski, D., Savenije, H. H. G., and Pfister, L.: From spatially variable streamflow to distributed hydrological models: Analysis of key modeling decisions, *Water Resources Research*, 52, 954–989, doi: 10.1002/2015WR017398, 2016.
- Firoz, A. B. M., Nauditt, A., Fink, M., and Ribbe, L.: Quantifying human impacts on hydrological drought using a combined modelling approach in a tropical river basin in Central Vietnam, *Hydrol. Earth Syst. Sci. Discuss.*, 2017, 1–33, doi: 10.5194/hess-2017-86, 2017.
- Fleischmann, A., Siqueira, V., Paris, A., Collischonn, W., Paiva, R., Pontes, P., Crétaux, J. F., Bergé-Nguyen, M., Biancamaria, S., Gosset, M., Calmant, S., and Tanimoun, B.: Modelling hydrologic and hydrodynamic processes in basins with large semi-arid wetlands, *Journal of Hydrology*, 561, 943–959, doi: 10.1016/j.jhydrol.2018.04.041, 2018.
- Foguesatto, C. R., Artuzo, F. D., Talamini, E., and Machado, J. A. D.: Understanding the divergences between farmer's perception and meteorological records regarding climate change: a review, *Environment, Development and Sustainability*, 22, 1–16, doi: 10.1007/s10668-018-0193-0, 2020.
- Forootan, E., Khaki, M., Schumacher, M., Wulfmeyer, V., Mehrnegar, N., van Dijk, A. I. J. M., Brocca, L., Farzaneh, S., Akinluyi, F., Ramillien, G., Shum, C. K., Awange, J., and Mostafaie, A.: Understanding the global hydrological droughts of 2003–2016 and their relationships with teleconnections, *Science of the Total Environment*, 650, 2587–2604, doi: 10.1016/j.scitotenv.2018.09.231, 2019.
- Fovet, O., Ruiz, L., Hrachowitz, M., Fauchoux, M., and Gascuel-Odoux, C.: Hydrological hysteresis and its value for assessing process consistency in catchment conceptual models, *Hydrology and Earth System Sciences*, 19, 105–123, doi: 10.5194/hess-19-105-2015, 2015.
- Fowler, K., Coxon, G., Freer, J., Peel, M., Wagener, T., Western, A., Woods, R., and Zhang, L.: Simulating Runoff Under Changing Climatic Conditions: A Framework for Model Improvement, *Water Resources Research*, 54, 9812–9832, doi: 10.1029/2018WR023989, 2018.
- Fowler, K., Knoben, W., Peel, M., Peterson, T., Ryu, D., Saft, M., Seo, K.-W., and Western, A.: Many Commonly Used Rainfall-Runoff Models Lack Long, Slow Dynamics: Implications for Runoff Projections, *Water Resources Research*, 56, e2019WR025286, doi: 10.1029/2019WR025286, 2020.
- Frappart, F., Calmant, S., Cauhopé, M., Seyler, F., and Cazenave, A.: Preliminary results of ENVISAT RA-2-derived water levels validation over the Amazon basin, *Remote Sensing of Environment*, 100, 252–264, doi: 10.1016/j.rse.2005.10.027, 2006.
- Frappart, F., Papa, F., Marieu, V., Malbeteau, Y., Jordy, F., Calmant, S., Durand, F., and Bala, S.: Preliminary Assessment of SARAL/AltiKa Observations over the Ganges-Brahmaputra and Irrawaddy Rivers, *Marine Geodesy*, 38, 568–580, doi: 10.1080/01490419.2014.990591, 2015.
- Freer, J., Beven, K., and Ambrose, B.: Bayesian Estimation of Uncertainty in Runoff Prediction and the Value of Data: An Application of the GLUE Approach, *Water Resources Research*, 32, 2161–2173, doi: 10.1029/95WR03723, 1996.
- Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Rowland, J., Romero, B., Husak, G., Michaelsen, J., and Verdin, A.: A quasi-global precipitation time series for drought monitoring: U.S. Geological Survey, Data Series 832, Report, U.S. Geological Survey, doi:

- 10.3133/ds832, 2014.
- Gallart, F. and Llorens, P.: Catchment Management under Environmental Change: Impact of Land Cover Change on Water Resources, *Water International*, 28, 334–340, doi: 10.1080/02508060308691707, 2003.
- Gao, H., Hrachowitz, M., Fenicia, F., Gharari, S., and Savenije, H. H. G.: Testing the realism of a topography-driven model (FLEX-Topo) in the nested catchments of the Upper Heihe, China, *Hydrol. Earth Syst. Sci.*, 18, 1895–1915, doi: 10.5194/hess-18-1895-2014, 2014a.
- Gao, H., Hrachowitz, M., Schymanski, S. J., Fenicia, F., Sriwongsitanon, N., and Savenije, H. H. G.: Climate controls how ecosystems size the root zone storage capacity at catchment scale, *Geophysical Research Letters*, 41, 2014GL061668, doi: 10.1002/2014GL061668, 2014b.
- Gao, H., Hrachowitz, M., Sriwongsitanon, N., Fenicia, F., Gharari, S., and Savenije, H. H. G.: Accounting for the influence of vegetation and landscape improves model transferability in a tropical savannah region, *Water Resources Research*, 52, 7999–8022, doi: 10.1002/2016WR019574, 2016.
- Garambois, P.-A., Calmant, S., Roux, H., Paris, A., Monnier, J., Finaud-Guyot, P., Samine Montazem, A., and Santos da Silva, J.: Hydraulic visibility: Using satellite altimetry to parameterize a hydraulic model of an ungauged reach of a braided river, *Hydrological Processes*, 31, 756–767, doi: 10.1002/hyp.11033, 2017.
- Garavaglia, F., Le Lay, M., Gottardi, F., Garçon, R., Gailhard, J., Paquet, E., and Mathevet, T.: Impact of model structure on flow simulation and hydrological realism: from a lumped to a semi-distributed approach, *Hydrol. Earth Syst. Sci.*, 21, 3937–3952, doi: 10.5194/hess-21-3937-2017, 2017.
- Gerrits, A. M. J., Savenije, H. H. G., Veling, E. J. M., and Pfister, L.: Analytical derivation of the Budyko curve based on rainfall characteristics and a simple evaporation model, *Water Resources Research*, 45, W04403, doi: 10.1029/2008WR007308, 2009.
- Getirana, A. C. V.: Integrating spatial altimetry data into the automatic calibration of hydrological models, *Journal of Hydrology*, 387, 244–255, doi: 10.1016/j.jhydrol.2010.04.013, 2010.
- Getirana, A. C. V. and Peters-Lidard, C.: Estimating water discharge from large radar altimetry datasets, *Hydrol. Earth Syst. Sci.*, 17, 923–933, doi: 10.5194/hess-17-923-2013, 2013.
- Getirana, A. C. V., Bonnet, M.-P., Calmant, S., Roux, E., Rotunno Filho, O. C., and Mansur, W. J.: Hydrological monitoring of poorly gauged basins based on rainfall-runoff modeling and spatial altimetry, *Journal of Hydrology*, 379, 205–219, doi: 10.1016/j.jhydrol.2009.09.049, 2009.
- Getirana, A. C. V., Bonnet, M. P., Rotunno Filho, O. C., Collischonn, W., Guyot, J. L., Seyler, F., and Mansur, W. J.: Hydrological modelling and water balance of the Negro River basin: evaluation based on in situ and spatial altimetry data, *Hydrological Processes*, 24, 3219–3236, doi: 10.1002/hyp.7747, 2010.
- Gharari, S., Hrachowitz, M., Fenicia, F., and Savenije, H. H. G.: Hydrological landscape classification: investigating the performance of HAND based landscape classifications in a central European meso-scale catchment, *Hydrol. Earth Syst. Sci.*, 15, 3275–3291, 2011.
- Gharari, S., Hrachowitz, M., Fenicia, F., and Savenije, H. H. G.: An approach to identify time consistent model parameters: sub-period calibration, *Hydrol. Earth Syst. Sci.*, 17, 149–161, doi: 10.5194/hess-17-149-2013, 2013.
- Gharari, S., Hrachowitz, M., Fenicia, F., Gao, H., and Savenije, H. H. G.: Using expert knowledge to increase realism in environmental system models can dramatically reduce the need for calibration, *Hydrol. Earth Syst. Sci.*, 18, 4839–4859, doi: 10.5194/hess-18-4839-2014, 2014.

- Gichamo, T. Z., Popescu, I., Jonoski, A., and Solomatine, D.: River cross-section extraction from the ASTER global DEM for flood modeling, *Environmental Modelling and Software*, 31, 37–46, doi: 10.1016/j.envsoft.2011.12.003, 2012.
- Giordano, R., Preziosi, E., and Romano, E.: Integration of local and scientific knowledge to support drought impact monitoring: some hints from an Italian case study, *Natural Hazards*, 69, 523–544, doi: 10.1007/s11069-013-0724-9, 2013.
- GlobCover: http://due.esrin.esa.int/page_globcover.php, 2009.
- Goroshi, S., Pradhan, R., Singh, R. P., Singh, K. K., and Parihar, J. S.: Trend analysis of evapotranspiration over India: Observed from long-term satellite measurements, *Journal of Earth System Science*, 126, 113, doi: 10.1007/s12040-017-0891-2, 2017.
- Goswami, M., O'Connor, K. M., and Bhattarai, K. P.: Development of regionalisation procedures using a multi-model approach for flow simulation in an ungauged catchment, *Journal of Hydrology*, 333, 517–531, doi: 10.1016/j.jhydrol.2006.09.018, 2007.
- Grigg, A. H. and Hughes, J. D.: Nonstationarity driven by multidecadal change in catchment groundwater storage: A test of modifications to a common rainfall–run-off model, *Hydrological Processes*, 32, 3675–3688, doi: 10.1002/hyp.13282, 2018.
- Gumindoga, W., Makurira, H., Phiri, M., and Nhapi, I.: Estimating runoff from ungauged catchments for reservoir water balance in the lower middle Zambezi basin, *Water SA*, 42, 641–649, doi: 10.4314/wsa.v42i4.15, 2016.
- Gupta, H. V., Wagener, T., and Liu, Y.: Reconciling theory with observations: elements of a diagnostic approach to model evaluation, *Hydrological Processes*, 22, 3802–3813, doi: 10.1002/hyp.6989, 2008.
- Haghnegahdar, A., Tolson, B. A., Craig, J. R., and Paya, K. T.: Assessing the performance of a semi-distributed hydrological model under various watershed discretization schemes, *Hydrological Processes*, 29, 4018–4031, doi: 10.1002/hyp.10550, 2015.
- Haile, G. G., Tang, Q., Sun, S., Huang, Z., Zhang, X., and Liu, X.: Droughts in East Africa: Causes, impacts and resilience, *Earth-Science Reviews*, 193, 146–161, doi: 10.1016/j.earscirev.2019.04.015, 2019.
- Hallegette, S., Green, C., Nicholls, R. J., and Corfee-Morlot, J.: Future flood losses in major coastal cities, *Nature Climate Change*, 3, 802–806, doi: 10.1038/nclimate1979, 2013.
- Hamududu, B. H. and Ngoma, H.: Impacts of climate change on water resources availability in Zambia: implications for irrigation development, *Environment, Development and Sustainability*, 22, 2817–2838, doi: 10.1007/s10668-019-00320-9, 2020.
- Handavu, F., Chirwa, P. W. C., and Syampungani, S.: Socio-economic factors influencing land-use and land-cover changes in the miombo woodlands of the Copperbelt province in Zambia, *Forest Policy and Economics*, 100, 75–94, doi: 10.1016/j.forpol.2018.10.010, 2019.
- Hanlon, J.: Floods displace thousands in Mozambique, <https://www.theguardian.com/world/2001/mar/28/mozambique.unitednations>, 2001.
- Hao, Z. and Singh, V. P.: Drought characterization from a multivariate perspective: A review, *Journal of Hydrology*, 527, 668–678, doi: 10.1016/j.jhydrol.2015.05.031, 2015.
- Hao, Z., Singh, V. P., and Xia, Y.: Seasonal Drought Prediction: Advances, Challenges, and Future Prospects, *Reviews of Geophysics*, 56, 108–141, doi: 10.1002/2016RG000549, 2018.
- Hargreaves, G. H. and Allen, R. G.: History and evaluation of hargreaves evapotranspiration equation, *Journal of Irrigation and Drainage Engineering*, 129, 53–63, doi: 10.1061/(ASCE)0733-9437(2003)129:1(53), 2003.
- Hargreaves, G. H. and Samani, Z. A.: Reference Crop Evapotranspiration from Temperature,

- Applied Engineering in Agriculture, 1, 96–99, doi: 10.13031/2013.26773, 1985.
- Hasan, M. A. and Pradhanang, S. M.: Estimation of flow regime for a spatially varied Himalayan watershed using improved multi-site calibration of the Soil and Water Assessment Tool (SWAT) model, *Environmental Earth Sciences*, 76, 787, doi: 10.1007/s12665-017-7134-3, 2017.
- Haverkamp, S., Srinivasan, R., Frede, H. G., and Santhi, C.: SUBWATERSHED SPATIAL ANALYSIS TOOL: DISCRETIZATION OF A DISTRIBUTED HYDROLOGIC MODEL BY STATISTICAL CRITERIA1, *JAWRA Journal of the American Water Resources Association*, 38, 1723–1733, doi: 10.1111/j.1752-1688.2002.tb04377.x, 2002.
- He, X., Estes, L., Konar, M., Tian, D., Anghileri, D., Baylis, K., Evans, T. P., and Sheffield, J.: Integrated approaches to understanding and reducing drought impact on food security across scales, *Current Opinion in Environmental Sustainability*, 40, 43–54, doi: 10.1016/j.cosust.2019.09.006, 2019.
- Hellwig, J. and Stahl, K.: An assessment of trends and potential future changes in groundwater-baseflow drought based on catchment response times, *Hydrol. Earth Syst. Sci.*, 22, 6209–6224, doi: 10.5194/hess-22-6209-2018, 2018.
- Henson, B.: Drought, Heat, and Victoria Falls: A Climate Story with a Twist, https://www.wunderground.com/cat6/Drought-Victoria-Falls-Climate-Story-Twist?cm_ven=cat6-widget, 2019.
- Herman, M. R., Nejadhashemi, A. P., Abouali, M., Hernandez-Suarez, J. S., Daneshvar, F., Zhang, Z., Anderson, M. C., Sadeghi, A. M., Hain, C. R., and Sharifi, A.: Evaluating the role of evapotranspiration remote sensing data in improving hydrological modeling predictability, *Journal of Hydrology*, 556, 39–49, doi: 10.1016/j.jhydrol.2017.11.009, 2018.
- Hobbins, M. T., Dai, A., Roderick, M. L., and Farquhar, G. D.: Revisiting the parameterization of potential evaporation as a driver of long-term water balance trends, *Geophysical Research Letters*, 35, L12 403, doi: 10.1029/2008GL033840, 2008.
- Hou, J., van Dijk, A. I. J. M., Renzullo, L. J., and Vertessy, R. A.: Using modelled discharge to develop satellite-based river gauging: a case study for the Amazon Basin, *Hydrol. Earth Syst. Sci.*, 22, 6435–6448, doi: 10.5194/hess-22-6435-2018, 2018.
- Hrachowitz, M. and Clark, M. P.: HESS Opinions: The complementary merits of competing modelling philosophies in hydrology, *Hydrol. Earth Syst. Sci.*, 21, 3953–3973, doi: 10.5194/hess-21-3953-2017, 2017.
- Hrachowitz, M. and Weiler, M.: Uncertainty of Precipitation Estimates Caused by Sparse Gauging Networks in a Small, Mountainous Watershed, *Journal of Hydrologic Engineering*, 16, 460–471, doi: 10.1061/(ASCE)HE.1943-5584.0000331, 2011.
- Hrachowitz, M., Savenije, H. H. G., Blöschl, G., McDonnell, J. J., Sivapalan, M., Pomeroy, J. W., Arheimer, B., Blume, T., Clark, M. P., Ehret, U., Fenicia, F., Freer, J. E., Gelfan, A., Gupta, H. V., Hughes, D. A., Hut, R. W., Montanari, A., Pande, S., Tetzlaff, D., Troch, P. A., Uhlenbrook, S., Wagener, T., Winsemius, H. C., Woods, R. A., Zehe, E., and Cudennec, C.: A decade of Predictions in Ungauged Basins (PUB)—a review, *Hydrological Sciences Journal*, 58, 1198–1255, doi: 10.1080/02626667.2013.803183, 2013.
- Hrachowitz, M., Fovet, O., Ruiz, L., Euser, T., Gharari, S., Nijzink, R., Freer, J., Savenije, H. H. G., and Gascuel-Oudou, C.: Process consistency in models: The importance of system signatures, expert knowledge, and process complexity, *Water Resources Research*, 50, 7445–7469, doi: 10.1002/2014WR015484, 2014.
- Hrachowitz, M., Stockinger, M., Coenders-Gerrits, M., van der Ent, R., Bogaen, H., Lücke, A., and Stumpp, C.: Deforestation reduces the vegetation-accessible water storage in the unsaturated soil and affects catchment travel time distributions and young water fractions, *Hydrol. Earth Syst. Sci. Discuss.*, 2020, 1–43, doi: 10.5194/hess-2020-293, 2020.

- Huang, H., Han, Y., Cao, M., Song, J., Xiao, H., and Cheng, W.: Spatiotemporal Characteristics of Evapotranspiration Paradox and Impact Factors in China in the Period of 1960–2013, *Advances in Meteorology*, 2015, 519–207, doi: 10.1155/2015/519207, 2015.
- Huang, Q., Long, D., Du, M., Zeng, C., Qiao, G., Li, X., Hou, A., and Hong, Y.: Discharge estimation in high-mountain regions with improved methods using multisource remote sensing: A case study of the Upper Brahmaputra River, *Remote Sensing of Environment*, 219, 115–134, doi: 10.1016/j.rse.2018.10.008, 2018.
- Huang, S., Huang, Q., Chang, J., and Leng, G.: Linkages between hydrological drought, climate indices and human activities: a case study in the Columbia River basin, *International Journal of Climatology*, 36, 280–290, doi: 10.1002/joc.4344, 2016.
- Huffman, G., Stocker, E., Bolvin, D., and Nelkin, E.: doi: 10.5067/TRMM/TMPA/MONTH/7, 2014.
- Huffman, G. J., Adler, R. F., Rudolf, B., Schneider, U., and Keehn, P. R.: Global Precipitation Estimates Based on a Technique for Combining Satellite-Based Estimates, Rain Gauge Analysis, and NWP Model Precipitation Information, *Journal of Climate*, 8, 1284–1295, doi: 10.1175/1520-0442(1995)008<1284:GPEBOA>2.0.CO;2, 1995.
- Huffman, G. J., Bolvin, D. T., Nelkin, E. J., Wolff, D. B., Adler, R. F., Gu, G., Hong, Y., Bowman, K. P., and Stocker, E. F.: The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-Global, Multiyear, Combined-Sensor Precipitation Estimates at Fine Scales, *Journal of Hydrometeorology*, 8, 38–55, doi: 10.1175/JHM560.1, 2007.
- Hulsman, P., Bogaard, T. A., and Savenije, H. H. G.: Rainfall-runoff modelling using river-stage time series in the absence of reliable discharge information: a case study in the semi-arid Mara River basin, *Hydrol. Earth Syst. Sci.*, 22, 5081–5095, doi: 10.5194/hess-22-5081-2018, 2018.
- Hulsman, P., Winsemius, H. C., Michailovsky, C. I., Savenije, H. H. G., and Hrachowitz, M.: Using altimetry observations combined with GRACE to select parameter sets of a hydrological model in a data-scarce region, *Hydrol. Earth Syst. Sci.*, 24, 3331–3359, doi: 10.5194/hess-24-3331-2020, 2020a.
- Hulsman, P., Savenije, H. H. G., and Hrachowitz, M.: Learning from satellite observations: increased understanding of catchment processes through stepwise model improvement, *Hydrol. Earth Syst. Sci. Discuss.*, 2020, 1–26, doi: 10.5194/hess-2020-191, 2020b.
- Hulsman, P., Savenije, H. H. G., and Hrachowitz, M.: Why are long-term storage variations observed but not modelled in the Luangwa basin?, *Water Resources Research*, in review.
- Hulsman, P., Savenije, H. H. G., and Hrachowitz, M.: Zambezi River Basin: Drought of 2019, *Journal of Hydrology: Regional Studies*, submitted.
- Immerzeel, W. W. and Droogers, P.: Calibration of a distributed hydrological model based on satellite evapotranspiration, *Journal of Hydrology*, 349, 411–424, doi: 10.1016/j.jhydrol.2007.11.017, 2008.
- Iqbal, M. W., Donjadedee, S., Kwanyuen, B., and Liu, S.-y.: Farmers' perceptions of and adaptations to drought in Herat Province, Afghanistan, *Journal of Mountain Science*, 15, 1741–1756, doi: 10.1007/s11629-017-4750-z, 2018.
- Irons, J. R., Dwyer, J. L., and Barsi, J. A.: The next Landsat satellite: The Landsat Data Continuity Mission, *Remote Sensing of Environment*, 122, 11–21, doi: 10.1016/j.rse.2011.08.026, 2012.
- Jahandideh-Tehrani, M., Zhang, H., Helfer, F., and Yu, Y.: Review of climate change impacts on predicted river streamflow in tropical rivers, *Environmental Monitoring and Assessment*, 191, 752, doi: 10.1007/s10661-019-7841-1, 2019.
- Jakeman, A. J. and Hornberger, G. M.: How much complexity is warranted in a rainfall-runoff model?, *Water Resources Research*, 29, 2637–2649, doi: 10.1029/93WR00877, 1993.

- Jian, J., Ryu, D., Costelloe, J. F., and Su, C.-H.: Towards hydrological model calibration using river level measurements, *Journal of Hydrology: Regional Studies*, 10, 95–109, doi: 10.1016/j.ejrh.2016.12.085, 2017.
- Jiang, D. and Wang, K.: The Role of Satellite-Based Remote Sensing in Improving Simulated Streamflow: A Review, *Water*, 11, doi: 10.3390/w11081615, 2019.
- Jiang, L., Schneider, R., Andersen, O. B., and Bauer-Gottwein, P.: CryoSat-2 altimetry applications over rivers and lakes, *Water (Switzerland)*, 9, doi: 10.3390/w9030211, 2017.
- Jing, W., Yao, L., Zhao, X., Zhang, P., Liu, Y., Xia, X., Song, J., Yang, J., Li, Y., and Zhou, C.: Understanding Terrestrial Water Storage Declining Trends in the Yellow River Basin, *Journal of Geophysical Research: Atmospheres*, 124, 12 963–12 984, doi: 10.1029/2019JD031432, 2019.
- Joodaki, G., Wahr, J., and Swenson, S.: Estimating the human contribution to groundwater depletion in the Middle East, from GRACE data, land surface models, and well observations, *Water Resources Research*, 50, 2679–2692, doi: 10.1002/2013WR014633, 2014.
- Kauffeldt, A., Wetterhall, F., Pappenberger, F., Salamon, P., and Thielen, J.: Technical review of large-scale hydrological models for implementation in operational flood forecasting schemes on continental level, *Environmental Modelling & Software*, 75, 68–76, doi: 10.1016/j.envsoft.2015.09.009, 2016.
- Kavetski, D. and Fenicia, F.: Elements of a flexible approach for conceptual hydrological modeling: 2. Application and experimental insights, *Water Resources Research*, 47, doi: 10.1029/2011WR010748, 2011.
- Kesselring, R.: The electricity crisis in Zambia: Blackouts and social stratification in new mining towns, *Energy Research and Social Science*, 30, 94–102, doi: 10.1016/j.erss.2017.06.015, 2017.
- Khaki, M. and Awange, J.: The application of multi-mission satellite data assimilation for studying water storage changes over South America, *Science of the Total Environment*, 647, 1557–1572, doi: 10.1016/j.scitotenv.2018.08.079, 2019.
- Khaki, M., Forootan, E., Kuhn, M., Awange, J., van Dijk, A. I. J. M., Schumacher, M., and Sharifi, M. A.: Determining water storage depletion within Iran by assimilating GRACE data into the W3RA hydrological model, *Advances in Water Resources*, 114, 1–18, doi: 10.1016/j.advwatres.2018.02.008, 2018.
- Kimani, W. M., Hoedjes, C. B. J., and Su, Z.: An Assessment of Satellite-Derived Rainfall Products Relative to Ground Observations over East Africa, *Remote Sensing*, 9, doi: 10.3390/rs9050430, 2017.
- Kiptala, J. K., Mohamed, Y., Mul, M. L., Cheema, M. J. M., and Van der Zaag, P.: Land use and land cover classification using phenological variability from MODIS vegetation in the Upper Pangani River Basin, Eastern Africa, *Physics and Chemistry of the Earth, Parts A/B/C*, 66, 112–122, doi: 10.1016/j.pce.2013.08.002, 2013.
- Kirchhoff, C. J. and Bulkley, J. W.: Sustainable Water Management in the Zambezi River Basin, *The Journal of the International Institute*, 2008.
- Kirchner, J. W.: Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology, *Water Resources Research*, 42, doi: 10.1029/2005WR004362, 2006.
- Kittel, C. M. M., Nielsen, K., Tøttrup, C., and Bauer-Gottwein, P.: Informing a hydrological model of the Ogooué with multi-mission remote sensing data, *Hydrol. Earth Syst. Sci.*, 22, 1453–1472, doi: 10.5194/hess-22-1453-2018, 2018.
- Klemeš, V.: Operational testing of hydrological simulation models, *Hydrological Sciences Journal*, 31, 13–24, doi: 10.1080/02626668609491024, 1986.
- Kling, H., Stanzel, P., and Preishuber, M.: Impact modelling of water resources development

- and climate scenarios on Zambezi River discharge, *Journal of Hydrology: Regional Studies*, 1, 17–43, doi: 10.1016/j.ejrh.2014.05.002, 2014.
- Knoben, W. J. M., Woods, R. A., and Freer, J. E.: A Quantitative Hydrological Climate Classification Evaluated With Independent Streamflow Data, *Water Resources Research*, 54, 5088–5109, doi: 10.1029/2018WR022913, 2018.
- Knutti, R.: Should we believe model predictions of future climate change?, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 366, 4647–4664, doi: 10.1098/rsta.2008.0169, 2008.
- Koch, J., Siemann, A., Stisen, S., and Sheffield, J.: Spatial validation of large-scale land surface models against monthly land surface temperature patterns using innovative performance metrics, *Journal of Geophysical Research: Atmospheres*, 121, 5430–5452, doi: 10.1002/2015JD024482, 2016.
- Koch, J., Demirel, M. C., and Stisen, S.: The SPATial Efficiency metric (SPAEF): Multiple-component evaluation of spatial patterns for optimization of hydrological models, *Geoscientific Model Development*, 11, 1873–1886, doi: 10.5194/gmd-11-1873-2018, 2018.
- Kornfeld, R. P., Arnold, B. W., Gross, M. A., Dahya, N. T., Klipstein, W. M., Gath, P. F., and Bettadpur, S.: GRACE-FO: The Gravity Recovery and Climate Experiment Follow-On Mission, *Journal of Spacecraft and Rockets*, 56, 931–951, doi: 10.2514/1.A34326, 2019.
- Kouraev, A. V., Zakharova, E. A., Samain, O., Mognard, N. M., and Cazenave, A.: Ob' river discharge from TOPEX/Poseidon satellite altimetry (1992–2002), *Remote Sensing of Environment*, 93, 238–245, doi: 10.1016/j.rse.2004.07.007, 2004.
- Krysanova, V. and Hattermann, F. F.: Intercomparison of climate change impacts in 12 large river basins: overview of methods and summary of results, *Climatic Change*, 141, 363–379, doi: 10.1007/s10584-017-1919-y, 2017.
- Kumar, R., Musuza, J. L., Van Loon, A. F., Teuling, A. J., Barthel, R., Ten Broek, J., Mai, J., Samaniego, L., and Attinger, S.: Multiscale evaluation of the Standardized Precipitation Index as a groundwater drought indicator, *Hydrol. Earth Syst. Sci.*, 20, 1117–1131, doi: 10.5194/hess-20-1117-2016, 2016.
- Kunnath-Poovakka, A., Ryu, D., Renzullo, L. J., and George, B.: The efficacy of calibrating hydrologic model using remotely sensed evapotranspiration and soil moisture for streamflow prediction, *Journal of Hydrology*, 535, 509–524, doi: 10.1016/j.jhydrol.2016.02.018, 2016.
- Lakshmi, V.: The role of satellite remote sensing in the Prediction of Ungauged Basins, *Hydrological Processes*, 18, 1029–1034, doi: 10.1002/hyp.5520, 2004.
- Landerer, F. and Swenson, S.: Accuracy of scaled GRACE terrestrial water storage estimates, *Water Resources Research*, 48, W04 531, doi: 10.1029/2011WR011453, 2012.
- Langhorst, T., Pavelsky, T. M., Frasson, R. P. d. M., Wei, R., Domeneghetti, A., Altenau, E. H., Durand, M. T., Minear, J. T., Wegmann, K. W., and Fuller, M. R.: Anticipated Improvements to River Surface Elevation Profiles From the Surface Water and Ocean Topography Mission, *Frontiers in Earth Science*, 7, 102, 2019.
- Le Coz, C. and van de Giesen, N.: Comparison of Rainfall Products over Sub-Saharan Africa, *Journal of Hydrometeorology*, 21, 553–596, doi: 10.1175/JHM-D-18-0256.1, 10.1175/JHM-D-18-0256.1, 2020.
- Le Moine, N., Andréassian, V., Perrin, C., and Michel, C.: How can rainfall-runoff models handle intercatchment groundwater flows? Theoretical study based on 1040 French catchments, *Water Resources Research*, 43, W06 428, doi: 10.1029/2006WR005608, 2007.
- Leblanc, M. J., Tregoning, P., Ramillien, G., Tweed, S. O., and Fakes, A.: Basin-scale, integrated observations of the early 21st century multiyear drought in southeast Australia, *Water Resources Research*, 45, W04 408, doi: 10.1029/2008WR007333, 2009.

- Leon, J. G., Calmant, S., Seyler, F., Bonnet, M. P., Cauhopé, M., Frappart, F., Filizola, N., and Fraizy, P.: Rating curves and estimation of average water depth at the upper Negro River based on satellite altimeter data and modeled discharges, *Journal of Hydrology*, 328, 481–496, doi: 10.1016/j.jhydrol.2005.12.006, 2006.
- Li, C., Wu, P. T., Li, X. L., Zhou, T. W., Sun, S. K., Wang, Y. B., Luan, X. B., and Yu, X.: Spatial and temporal evolution of climatic factors and its impacts on potential evapotranspiration in Loess Plateau of Northern Shaanxi, China, *Science of The Total Environment*, 589, 165–172, doi: 10.1016/j.scitotenv.2017.02.122, 2017.
- Li, Z., Yang, D., Gao, B., Jiao, Y., Hong, Y., and Xu, T.: Multiscale hydrologic applications of the latest satellite precipitation products in the Yangtze river basin using a distributed hydrologic model, *Journal of Hydrometeorology*, 16, 407–426, doi: 10.1175/JHM-D-14-0105.1, 2015.
- Liang, X., Lettenmaier, D. P., Wood, E. F., and Burges, S. J.: A simple hydrologically based model of land surface water and energy fluxes for general circulation models, *Journal of Geophysical Research: Atmospheres*, 99, 14 415–14 428, doi: 10.1029/94JD00483, 1994.
- Libanda, B., Zheng, M., and Ngonga, C.: Spatial and temporal patterns of drought in Zambia, *Journal of Arid Land*, 11, 180–191, doi: 10.1007/s40333-019-0053-2, 10.1007/s40333-019-0053-2, 2019.
- Liu, G., Schwartz, F. W., Tseng, K. H., and Shum, C. K.: Discharge and water-depth estimates for ungauged rivers: Combining hydrologic, hydraulic, and inverse modeling with stage and water-area measurements from satellites, *Water Resources Research*, 51, 6017–6035, doi: 10.1002/2015WR016971, 2015.
- Long, D., Scanlon, B. R., Longuevergne, L., Sun, A. Y., Fernando, D. N., and Save, H.: GRACE satellite monitoring of large depletion in water storage in response to the 2011 drought in Texas, *Geophysical Research Letters*, 40, 3395–3401, doi: 10.1002/grl.50655, 2013.
- Long, D., Longuevergne, L., and Scanlon, B. R.: Uncertainty in evapotranspiration from land surface modeling, remote sensing, and GRACE satellites, *Water Resources Research*, 50, 1131–1151, doi: 10.1002/2013WR014581, 2014.
- Long, D., Pan, Y., Zhou, J., Chen, Y., Hou, X., Hong, Y., Scanlon, B. R., and Longuevergne, L.: Global analysis of spatiotemporal variability in merged total water storage changes using multiple GRACE products and global hydrological models, *Remote Sensing of Environment*, 192, 198–216, doi: 10.1016/j.rse.2017.02.011, 2017.
- López López, P., Sutanudjaja, E. H., Schellekens, J., Sterk, G., and Bierkens, M. F. P.: Calibration of a large-scale hydrological model using satellite-based soil moisture and evapotranspiration products, *Hydrol. Earth Syst. Sci.*, 21, 3125–3144, doi: 10.5194/hess-21-3125-2017, 2017.
- Maes, W. H., Gentine, P., Verhoest, N. E. C., and Miralles, D. G.: Potential evaporation at eddy-covariance sites across the globe, *Hydrol. Earth Syst. Sci.*, 23, 925–948, doi: 10.5194/hess-23-925-2019, 2019.
- Manning, R.: On the flow of water in open channels and pipes, *Transactions of the Institution of Civil Engineers of Ireland*, 20, 161–207, 1891.
- Markus, M., Cai, X., and Srivier, R.: Extreme Floods and Droughts under Future Climate Scenarios, *Water*, 11, doi: 10.3390/w11081720, 2019.
- Martens, B., Miralles, D. G., Lievens, H., van der Schalie, R., de Jeu, R. A. M., Fernández-Prieto, D., Beck, H. E., Dorigo, W. A., and Verhoest, N. E. C.: GLEAM v3: satellite-based land evaporation and root-zone soil moisture, *Geosci. Model Dev.*, 10, 1903–1925, doi: 10.5194/gmd-10-1903-2017, 2017.
- Matiashe, F. S.: Victoria Falls, one of the natural wonders of the world, is running dry due to climate change, <https://qz.com/africa/1765688/photos-of-victoria->

- falls-running-dry-in-zimbabwe-and-zambia/, 2019.
- Mazzoleni, M., Brandimarte, L., and Amaranto, A.: Evaluating precipitation datasets for large-scale distributed hydrological modelling, *Journal of Hydrology*, 578, 124076, doi: 10.1016/j.jhydrol.2019.124076, 2019.
- McCuen Richard, H., Knight, Z., and Cutter, A. G.: Evaluation of the Nash–Sutcliffe Efficiency Index, *Journal of Hydrologic Engineering*, 11, 597–602, doi: 10.1061/(ASCE)1084-0699(2006)11:6(597), 2006.
- McKee, T. B., Doesken, N. J., and Kleist, J.: THE RELATIONSHIP OF DROUGHT FREQUENCY AND DURATION TO TIME SCALES, in: Eighth Conference on Applied Climatology, Anaheim, California, 1993.
- McMillan, H., Westerberg, I., and Branger, F.: Five guidelines for selecting hydrological signatures, *Hydrological Processes*, 31, 4757–4761, doi: 10.1002/hyp.11300, 2017.
- McMillan, H. K. and Westerberg, I. K.: Rating curve estimation under epistemic uncertainty, *Hydrological Processes*, 29, 1873–1882, doi: 10.1002/hyp.10419, 2015.
- Meier, P., Frömelt, A., and Kinzelbach, W.: Hydrological real-time modelling in the Zambezi river basin using satellite-based soil moisture and rainfall data, *Hydrol. Earth Syst. Sci.*, 15, 999–1008, doi: 10.5194/hess-15-999-2011, 2011.
- Mendiguren, G., Koch, J., and Stisen, S.: Spatial pattern evaluation of a calibrated national hydrological model – a remote-sensing-based diagnostic approach, *Hydrol. Earth Syst. Sci.*, 21, 5987–6005, doi: 10.5194/hess-21-5987-2017, 2017.
- Meng, F., Su, F., Li, Y., and Tong, K.: Changes in Terrestrial Water Storage During 2003–2014 and Possible Causes in Tibetan Plateau, *Journal of Geophysical Research: Atmospheres*, 124, 2909–2931, doi: 10.1029/2018JD029552, 2019.
- Meze-Hausken, E.: Contrasting climate variability and meteorological drought with perceived drought and climate change in northern Ethiopia, *Climate Research - CLIMATE RES*, 27, 19–31, doi: 10.3354/cr027019, 2004.
- Michailovsky, C. I. and Bauer-Gottwein, P.: Operational reservoir inflow forecasting with radar altimetry: the Zambezi case study, *Hydrol. Earth Syst. Sci.*, 18, 997–1007, doi: 10.5194/hess-18-997-2014, 2014.
- Michailovsky, C. I., McEnnis, S., Berry, P. A. M., Smith, R., and Bauer-Gottwein, P.: River monitoring from satellite radar altimetry in the Zambezi River basin, *Hydrol. Earth Syst. Sci.*, 16, 2181–2192, doi: 10.5194/hess-16-2181-2012, 2012.
- Michailovsky, C. I., Milzow, C., and Bauer-Gottwein, P.: Assimilation of radar altimetry to a routing model of the Brahmaputra River, *Water Resources Research*, 49, 4807–4816, doi: 10.1002/wrcr.20345, 2013.
- Milzow, C., Krogh, P. E., and Bauer-Gottwein, P.: Combining satellite radar altimetry, SAR surface soil moisture and GRACE total storage changes for hydrological model calibration in a large poorly gauged catchment, *Hydrol. Earth Syst. Sci.*, 15, 1729–1743, doi: 10.5194/hess-15-1729-2011, 2011.
- Miralles, D. G., Holmes, T. R. H., De Jeu, R. A. M., Gash, J. H., Meesters, A. G. C. A., and Dolman, A. J.: Global land-surface evaporation estimated from satellite-based observations, *Hydrol. Earth Syst. Sci.*, 15, 453–469, doi: 10.5194/hess-15-453-2011, 2011.
- Mishra, A. K. and Singh, V. P.: A review of drought concepts, *Journal of Hydrology*, 391, 202–216, doi: 10.1016/j.jhydrol.2010.07.012, 2010.
- Montanari, M., Hostache, R., Matgen, P., Schumann, G., Pfister, L., and Hoffmann, L.: Calibration and sequential updating of a coupled hydrologic-hydraulic model using remote sensing-derived water stages, *Hydrol. Earth Syst. Sci.*, 13, 367–380, doi: 10.5194/hess-13-367-2009, 2009.
- Musa, Z. N., Popescu, I., and Mynett, A.: A review of applications of satellite SAR, optical,

- altimetry and DEM data for surface water modelling, mapping and parameter estimation, *Hydrol. Earth Syst. Sci.*, 19, 3755–3769, doi: 10.5194/hess-19-3755-2015, 2015.
- Naresh Kumar, M., Murthy, C. S., Sessa Sai, M. V. R., and Roy, P. S.: On the use of Standardized Precipitation Index (SPI) for drought intensity assessment, *Meteorological Applications*, 16, 381–389, doi: 10.1002/met.136, 2009.
- Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models part I — A discussion of principles, *Journal of Hydrology*, 10, 282–290, doi: 10.1016/0022-1694(70)90255-6, 1970.
- Nelson, S. T. and Mayo, A. L.: The role of interbasin groundwater transfers in geologically complex terranes, demonstrated by the Great Basin in the western United States, *Hydrogeology Journal*, 22, 807–828, doi: 10.1007/s10040-014-1104-6, 2014.
- Nie, N., Zhang, W., Chen, H., and Guo, H.: A Global Hydrological Drought Index Dataset Based on Gravity Recovery and Climate Experiment (GRACE) Data, *Water Resources Management*, 32, 1275–1290, doi: 10.1007/s11269-017-1869-1, 2018.
- Nijzink, R., Hutton, C., Pechlivanidis, I., Capell, R., Arheimer, B., Freer, J., Han, D., Wagener, T., McGuire, K., Savenije, H., and Hrachowitz, M.: The evolution of root-zone moisture capacities after deforestation: a step towards hydrological predictions under change?, *Hydrol. Earth Syst. Sci.*, 20, 4775–4799, doi: 10.5194/hess-20-4775-2016, 2016a.
- Nijzink, R. C., Samaniego, L., Mai, J., Kumar, R., Thober, S., Zink, M., Schäfer, D., Savenije, H. H. G., and Hrachowitz, M.: The importance of topography-controlled sub-grid process heterogeneity and semi-quantitative prior constraints in distributed hydrological models, *Hydrol. Earth Syst. Sci.*, 20, 1151–1176, doi: 10.5194/hess-20-1151-2016, 2016b.
- Nijzink, R. C., Almeida, S., Pechlivanidis, I. G., Capell, R., Gustafssons, D., Arheimer, B., Parajka, J., Freer, J., Han, D., Wagener, T., van Nooijen, R. R. P., Savenije, H. H. G., and Hrachowitz, M.: Constraining Conceptual Hydrological Models With Multiple Information Sources, *Water Resources Research*, 54, 8332–8362, doi: 10.1029/2017WR021895, 2018.
- Normandin, C., Frappart, F., Telly Diepkilé, A., Marieu, V., Mougine, E., Blarel, F., Lubac, B., Braquet, N., and Ba, A.: Evolution of the Performances of Radar Altimetry Missions from ERS-2 to Sentinel-3A over the Inner Niger Delta, *Remote Sensing*, 10, 833, doi: 10.3390/rs10060833, 2018.
- Oduşanya, A. E., Mehdi, B., Schürz, C., Oke, A. O., Awokola, O. S., Awomeso, J. A., Adejuwon, J. O., and Schulz, K.: Multi-site calibration and validation of SWAT with satellite-based evapotranspiration in a data-sparse catchment in southwestern Nigeria, *Hydrol. Earth Syst. Sci.*, 23, 1113–1144, doi: 10.5194/hess-23-1113-2019, 2019.
- Oguntunde, P. G., Friesen, J., van de Giesen, N., and Savenije, H. H. G.: Hydroclimatology of the Volta River Basin in West Africa: Trends and variability from 1901 to 2002, *Physics and Chemistry of the Earth, Parts A/B/C*, 31, 1180–1188, doi: 10.1016/j.pce.2006.02.062, 2006.
- Ol'dekop, E. M.: On evaporation from the surface of river basins, *Transactions on Meteorological Observations*, 4, 1911.
- Osgood, D., Powell, B., Diro, R., Farah, C., Enenkel, M., Brown, M. E., Husak, G., Blakeley, S. L., Hoffman, L., and McCarty, J. L.: Farmer Perception, Recollection, and Remote Sensing in Weather Index Insurance: An Ethiopia Case Study, *Remote Sensing*, 10, doi: 10.3390/rs10121887, 2018.
- Oubanas, H., Gejadze, I., Malaterre, P. O., Durand, M., Wei, R., Frasson, R. P. M., and Domeneghetti, A.: Discharge Estimation in Ungauged Basins Through Variational Data Assimilation: The Potential of the SWOT Mission, *Water Resources Research*, 54, 2405–2423, doi: 10.1002/2017WR021735, 2018.
- Ovuka, M. and Lindqvist, S.: Rainfall variability in murang'a district, kenya: meteorological

- data and farmers' perception, *Geografiska Annaler: Series A, Physical Geography*, 82, 107–119, doi: 10.1111/j.0435-3676.2000.00116.x, 2000.
- Pandya, U., Patel, A., and Patel, D.: RIVER CROSS SECTION DELINEATION FROM THE GOOGLE EARTH FOR DEVELOPMENT OF 1D HEC-RAS MODEL-A CASE OF SABARMATI RIVER, GUJARAT, INDIA, in: International Conference on Hydraulics, Water Resources & Coastal Engineering, Ahmedabad, India, 2017.
- Papa, F., Bala, S. K., Pandey, R. K., Durand, F., Gopalakrishna, V. V., Rahman, A., and Rossow, W. B.: Ganga-Brahmaputra river discharge from Jason-2 radar altimetry: An update to the long-term satellite-derived estimates of continental freshwater forcing flux into the Bay of Bengal, *Journal of Geophysical Research: Oceans*, 117, doi: 10.1029/2012JC008158, 2012.
- Paris, A., Dias de Paiva, R., Santos da Silva, J., Medeiros Moreira, D., Calmant, S., Garambois, P. A., Collischonn, W., Bonnet, M. P., and Seyler, F.: Stage-discharge rating curves based on satellite altimetry and modeled discharge in the Amazon basin, *Water Resources Research*, 52, 3787–3814, doi: 10.1002/2014WR016618, 2016.
- Pechlivanidis, I. G. and Arheimer, B.: Large-scale hydrological modelling by using modified PUB recommendations: The India-HYPE case, *Hydrology and Earth System Sciences*, 19, 4559–4579, doi: 10.5194/hess-19-4559-2015, 2015.
- Pechlivanidis, I. G., Jackson, B. M., McIntyre, N. R., and Wheater, H. S.: Catchment Scale Hydrological Modelling: A Review Of Model Types, Calibration Approaches And Uncertainty Analysis Methods In The Context Of Recent Developments In Technology And Applications, *Global NEST Journal*, 13, 193–214, doi: 10.30955/gnj.000778, 2011.
- Pedinotti, V., Boone, A., Decharme, B., Crétau, J. F., Mognard, N., Panthou, G., Papa, F., and Tanimoun, B. A.: Evaluation of the ISBA-TRIP continental hydrologic system over the Niger basin using in situ and satellite derived datasets, *Hydrology and Earth System Sciences*, 16, 1745–1773, doi: 10.5194/hess-16-1745-2012, 2012.
- Peel, M. C. and Blöschl, G.: Hydrological modelling in a changing world, *Progress in Physical Geography*, 35, 249–261, 2011.
- Pekel, J.-F., Cottam, A., Gorelick, N., and Belward, A. S.: High-resolution mapping of global surface water and its long-term changes, *Nature*, 540, 418–422, doi: 10.1038/nature20584, 2016.
- Pellicer-Martínez, F. and Martínez-Paz, J. M.: Assessment of interbasin groundwater flows between catchments using a semi-distributed water balance model, *Journal of Hydrology*, 519, 1848–1858, doi: 10.1016/j.jhydrol.2014.09.067, 2014.
- Pereira-Cardenal, S. J., Riegels, N. D., Berry, P. A. M., Smith, R. G., Yakovlev, A., Siegfried, T. U., and Bauer-Gottwein, P.: Real-time remote sensing driven river basin modeling using radar altimetry, *Hydrol. Earth Syst. Sci.*, 15, 241–254, doi: 10.5194/hess-15-241-2011, 2011.
- Perrin, C., Michel, C., and Andréassian, V.: Improvement of a parsimonious model for streamflow simulation, *Journal of Hydrology*, 279, 275–289, doi: 10.1016/S0022-1694(03)00225-7, 2003.
- Petrucci, G. and Bonhomme, C.: The dilemma of spatial representation for urban hydrology semi-distributed modelling: Trade-offs among complexity, calibration and geographical data, *Journal of Hydrology*, 517, 997–1007, doi: 10.1016/j.jhydrol.2014.06.019, 2014.
- Phiri, D., Morgenroth, J., and Xu, C.: Long-term land cover change in Zambia: An assessment of driving factors, *Science of The Total Environment*, 697, 134–206, doi: 10.1016/j.scitotenv.2019.134206, 2019a.
- Phiri, D., Morgenroth, J., and Xu, C.: Four decades of land cover and forest connectivity study in Zambia—An object-based image analysis approach, *International Journal of Ap-*

- plied Earth Observation and Geoinformation, 79, 97–109, doi: 10.1016/j.jag.2019.03.001, 2019b.
- Pike, J. G.: The estimation of annual run-off from meteorological data in a tropical climate, *Journal of Hydrology*, 2, 116–123, doi: 10.1016/0022-1694(64)90022-8, 1964.
- Pramanik, N., Panda, R. K., and Sen, D.: One Dimensional Hydrodynamic Modeling of River Flow Using DEM Extracted River Cross-sections, *Water Resources Management*, 24, 835–852, doi: 10.1007/s11269-009-9474-6, 2010.
- Prenner, D., Kaitna, R., Mostbauer, K., and Hrachowitz, M.: The Value of Using Multiple Hydrometeorological Variables to Predict Temporal Debris Flow Susceptibility in an Alpine Environment, *Water Resources Research*, 54, 6822–6843, doi: 10.1029/2018WR022985, 2018.
- Price, D., Hudson, K., Boyce, G., Schellekens, J., Moore, R. J., Clark, P., Harrison, T., Connolly, E., and Pilling, C.: Operational use of a grid-based model for flood forecasting, *Proceedings of the Institution of Civil Engineers - Water Management*, 165, 65–77, doi: 10.1680/wama.2012.165.2.65, 2012.
- Rajib, A., Evenson, G. R., Golden, H. E., and Lane, C. R.: Hydrologic model predictability improves with spatially explicit calibration using remotely sensed evapotranspiration and biophysical parameters, *Journal of Hydrology*, 567, 668–683, doi: 10.1016/j.jhydrol.2018.10.024, 2018.
- Rakovec, O., Kumar, R., Attinger, S., and Samaniego, L.: Improving the realism of hydrologic model functioning through multivariate parameter estimation, *Water Resources Research*, 52, 7779–7792, doi: 10.1002/2016WR019430, 2016.
- Rantz, S. E.: Measurement and computation of streamflow: Volume 2, Computation of Discharge, Report 2175, doi: 10.3133/wsp2175, 1982.
- Refsgaard, J. C., Storm, B., and Clausen, T.: Système Hydrologique Européen (SHE): review and perspectives after 30 years development in distributed physically-based hydrological modelling, *Hydrology Research*, 41, 355–377, doi: 10.2166/nh.2010.009, 2010.
- Renard, B., Kavetski, D., Kuczera, G., Thyer, M., and Franks, S. W.: Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors, *Water Resources Research*, 46, doi: 10.1029/2009WR008328, 2010.
- Rennó, C. D., Nobre, A. D., Cuartas, L. A., Soares, J. V., Hodnett, M. G., Tomasella, J., and Waterloo, M. J.: HAND, a new terrain descriptor using SRTM-DEM: Mapping terra-firme rainforest environments in Amazonia, *Remote Sensing of Environment*, 112, 3469–3481, doi: 10.1016/j.rse.2008.03.018, 2008.
- Revilla-Romero, B., Beck, H. E., Burek, P., Salamon, P., de Roo, A., and Thielen, J.: Filling the gaps: Calibrating a rainfall-runoff model using satellite-derived surface water extent, *Remote Sensing of Environment*, 171, 118–131, doi: 10.1016/j.rse.2015.10.022, 2015.
- Riegger, J., Tourian, M. J., Devaraju, B., and Sneeuw, N.: Analysis of grace uncertainties by hydrological and hydro-meteorological observations, *Journal of Geodynamics*, 59–60, 16–27, doi: 10.1016/j.jog.2012.02.001, 2012.
- Rientjes, T. H. M., Muthuwatta, L. P., Bos, M. G., Booij, M. J., and Bhatti, H. A.: Multi-variable calibration of a semi-distributed hydrological model using streamflow data and satellite-based evapotranspiration, *Journal of Hydrology*, 505, 276–290, doi: 10.1016/j.jhydrol.2013.10.006, 2013.
- Roderick, M. L. and Farquhar, G. D.: Changes in New Zealand pan evaporation since the 1970s, *International Journal of Climatology*, 25, 2031–2039, doi: 10.1002/joc.1262, 2005.
- Roodari, A., Hrachowitz, M., Hassanpour, F., and Yaghoobzadeh, M.: Signatures of human intervention – or not? Downstream intensification of hydrological drought along a large Central Asian River: the individual roles of climate variability and land use change, *Hydrol.*

- Earth Syst. Sci. Discuss., 2020, 1–40, doi: 10.5194/hess-2020-369, 2020.
- Roy, T., Gupta, H. V., Serrat-Capdevila, A., and Valdes, J. B.: Using satellite-based evapotranspiration estimates to improve the structure of a simple conceptual rainfall-runoff model, *Hydrol. Earth Syst. Sci.*, 21, 879–896, doi: 10.5194/hess-21-879-2017, 2017.
- Running, S., Mu, Q., and Zhao, M.: doi: 10.5067/MODIS/MOD16A2.006, 2017.
- SADC: Integrated Water Resources Management Strategy and Implementation Plan for the Zambezi River Basin, Euroconsult Mott MacDonald, 2008.
- Saft, M., Peel, M. C., Western, A. W., Perraud, J.-M., and Zhang, L.: Bias in streamflow projections due to climate-induced shifts in catchment response, *Geophysical Research Letters*, 43, 1574–1581, doi: 10.1002/2015GL067326, 2016.
- Samaniego, L., Kumar, R., and Attinger, S.: Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale, *Water Resources Research*, 46, doi: 10.1029/2008WR007327, 2010.
- Samaniego, L., Kumar, R., and Jackisch, C.: Predictions in a data-sparse region using a regionalized grid-based hydrologic model driven by remotely sensed data, *Hydrology Research*, 42, 338–355, doi: 10.2166/nh.2011.156, 2011.
- Santhi, C., Kannan, N., Arnold, J. G., and Di Luzio, M.: Spatial Calibration and Temporal Validation of Flow for Regional Scale Hydrologic Modeling, *JAWRA Journal of the American Water Resources Association*, 44, 829–846, doi: 10.1111/j.1752-1688.2008.00207.x, 2008.
- Santos da Silva, J., Calmant, S., Seyler, F., Rotunno Filho, O. C., Cochonneau, G., and Mansur, W. J.: Water levels in the Amazon basin derived from the ERS 2 and ENVISAT radar altimetry missions, *Remote Sensing of Environment*, 114, 2160–2181, doi: 10.1016/j.rse.2010.04.020, 2010.
- Savenije, H.: Topography driven conceptual modelling (FLEX-Topo), *Hydrol. Earth Syst. Sci.*, 14, 2681–2692, doi: 10.5194/hess-14-2681-2010, 2010.
- Savenije, H. H. G.: Equifinality, a blessing in disguise?, *Hydrological Processes*, 15, 2835–2838, doi: 10.1002/hyp.494, 2001.
- Sawicz, K., Wagener, T., Sivapalan, M., Troch, P. A., and Carrillo, G.: Catchment classification: empirical analysis of hydrologic similarity based on catchment function in the eastern USA, *Hydrol. Earth Syst. Sci.*, 15, 2895–2911, doi: 10.5194/hess-15-2895-2011, 2011.
- Scanlon, B. R., Zhang, Z., Save, H., Sun, A. Y., Müller Schmied, H., van Beek, L. P. H., Wiese, D. N., Wada, Y., Long, D., Reedy, R. C., Longuevergne, L., Döll, P., and Bierkens, M. F. P.: Global models underestimate large decadal declining and rising water storage trends relative to GRACE satellite data, *Proceedings of the National Academy of Sciences*, 115, E1080, doi: 10.1073/pnas.1704665115, 2018.
- Schleiss, A. J. and Matos, J. P.: Chapter 98: Zambezi River Basin, McGraw-Hill Education - Europe, United States, 2016.
- Schneider, R., Godiksen, P. N., Villadsen, H., Madsen, H., and Bauer-Gottwein, P.: Application of CryoSat-2 altimetry data for river analysis and modelling, *Hydrol. Earth Syst. Sci.*, 21, 751–764, doi: 10.5194/hess-21-751-2017, 2017.
- Schoups, G., Lee Addams, C., and Gorelick, S. M.: Multi-objective calibration of a surface water-groundwater flow model in an irrigated agricultural region: Yaqui Valley, Sonora, Mexico, *Hydrol. Earth Syst. Sci.*, 9, 549–568, doi: 10.5194/hess-9-549-2005, 2005.
- Schreiber, P.: Über die Beziehungen zwischen dem Niederschlag und der Wasserführung der Flüsse in Mitteleuropa, *Z. Meteorol*, 21, 441–452, 1904.
- Schumacher, M., Forootan, E., van Dijk, A. I. J. M., Müller Schmied, H., Crosbie, R. S., Kusche, J., and Döll, P.: Improving drought simulations within the Murray-Darling Basin by combined calibration/assimilation of GRACE data into the WaterGAP Global Hydrology Model,

- Remote Sensing of Environment, 204, 212–228, doi: 10.1016/j.rse.2017.10.029, 2018.
- Schumann, G., Kirschbaum, D., Anderson, E., and Rashid, K.: Role of Earth Observation Data in Disaster Response and Recovery: From Science to Capacity Building, pp. 119–146, Springer International Publishing, Seattle, USA, doi: 10.1007/978-3-319-33438-7_5, 2016.
- Schwatke, C., Dettmering, D., Bosch, W., and Seitz, F.: DAHITI – an innovative approach for estimating water level time series over inland waters using multi-mission satellite altimetry, *Hydrol. Earth Syst. Sci.*, 19, 4345–4364, doi: 10.5194/hess-19-4345-2015, 2015.
- Seibert, J. and Beven, K. J.: Gauging the ungauged basin: how many discharge measurements are needed?, *Hydrol. Earth Syst. Sci.*, 13, 883–892, doi: 10.5194/hess-13-883-2009, 2009.
- Seibert, J. and Vis, M. J. P.: How informative are stream level observations in different geographic regions?, *Hydrological Processes*, 30, 2498–2508, doi: 10.1002/hyp.10887, 2016a.
- Seibert, J. and Vis, M. J. P.: How informative are stream level observations in different geographic regions?, *Hydrological Processes*, 30, 2498–2508, doi: 10.1002/hyp.10887, 2016b.
- Senay, G. B., Budde, M., Verdin, J. P., and Melesse, A. M.: A Coupled Remote Sensing and Simplified Surface Energy Balance Approach to Estimate Actual Evapotranspiration from Irrigated Fields, *Sensors (Basel, Switzerland)*, 7, 979–1000, doi: 10.3390/s7060979, 2007.
- Serrat-Capdevila, A., Valdes, J. B., and Stakhiv, E. Z.: Water Management Applications for Satellite Precipitation Products: Synthesis and Recommendations, *JAWRA Journal of the American Water Resources Association*, 50, 509–525, doi: 10.1111/jawr.12140, 2014.
- Seyler, F., Muller, F., Cochonneau, G., Guimarães, L., and Guyot, J. L.: Watershed delineation for the Amazon sub-basin system using GTOPO30 DEM and a drainage network extracted from JERS SAR images, *Hydrological Processes*, 23, 3173–3185, doi: 10.1002/hyp.7397, 2009.
- Seyler, F., Calmant, S., Silva, J. S. d., Moreira, D. M., Mercier, F., and Shum, C. K.: From TOPEX/Poseidon to Jason-2/OSTM in the Amazon basin, *Advances in Space Research*, 51, 1542–1550, doi: 10.1016/j.asr.2012.11.002, 2013.
- Sichangi, A. W., Wang, L., Yang, K., Chen, D., Wang, Z., Li, X., Zhou, J., Liu, W., and Kuria, D.: Estimating continental river basin discharges using multiple remote sensing data sets, *Remote Sensing of Environment*, 179, 36–53, doi: 10.1016/j.rse.2016.03.019, 2016.
- Sikorska, A. E. and Renard, B.: Calibrating a hydrological model in stage space to account for rating curve uncertainties: general framework and key challenges, *Advances in Water Resources*, 105, 51–66, doi: 10.1016/j.advwatres.2017.04.011, 2017.
- Simelton, E., Quinn, C. H., Batisani, N., Dougill, A. J., Dyer, J. C., Fraser, E. D. G., Mkwambisi, D., Sallu, S., and Stringer, L. C.: Is rainfall really changing? Farmers' perceptions, meteorological data, and policy implications, *Climate and Development*, 5, 123–138, doi: 10.1080/17565529.2012.751893, 2013.
- Smith, B. and Sandwell, D.: Accuracy and resolution of shuttle radar topography mission data, *Geophysical Research Letters*, 30, doi: 10.1029/2002GL016643, 2003.
- Solano-Hernandez, A., Bruzzone, O., Groot, J., Laborda, L., Martínez, A., Tiftonell, P., and Easdale, M. H.: Convergence between satellite information and farmers' perception of drought in rangelands of North-West Patagonia, Argentina, *Land Use Policy*, 97, 104726, doi: <https://doi.org/10.1016/j.landusepol.2020.104726>, 2020.
- Spearman, C.: The proof and measurement of association between two things, *The American Journal of Psychology*, 15, 72–101, 1904.
- Stahl, K., Kohn, I., Blauhut, V., Urquijo, J., De Stefano, L., Acácio, V., Dias, S., Stagge, J. H.,

- Tallaksen, L. M., Kampragou, E., Van Loon, A. F., Barker, L. J., Melsen, L. A., Bifulco, C., Musolino, D., de Carli, A., Massarutto, A., Assimacopoulos, D., and Van Lanen, H. A. J.: Impacts of European drought events: insights from an international database of text-based reports, *Nat. Hazards Earth Syst. Sci.*, 16, 801–819, doi: 10.5194/nhess-16-801-2016, 2016.
- Stisen, S., McCabe, M. F., Refsgaard, J. C., Lerer, S., and Butts, M. B.: Model parameter analysis using remotely sensed pattern information in a multi-constraint framework, *Journal of Hydrology*, 409, 337–349, doi: 10.1016/j.jhydrol.2011.08.030, 2011.
- Stisen, S., Koch, J., Sonnenborg, T. O., Refsgaard, J. C., Bircher, S., Ringgaard, R., and Jensen, K. H.: Moving beyond run-off calibration—Multivariable optimization of a surface–subsurface–atmosphere model, *Hydrological Processes*, 32, 2654–2668, doi: 10.1002/hyp.13177, 2018.
- Su, Z.: The Surface Energy Balance System (SEBS) for estimation of turbulent heat fluxes, *Hydrol. Earth Syst. Sci.*, 6, 85–100, doi: 10.5194/hess-6-85-2002, 2002.
- Sulistioadi, Y. B., Tseng, K. H., Shum, C. K., Hidayat, H., Sumaryono, M., Suhardiman, A., Setiawan, F., and Sunarso, S.: Satellite radar altimetry for monitoring small rivers and lakes in Indonesia, *Hydrol. Earth Syst. Sci.*, 19, 341–359, doi: 10.5194/hess-19-341-2015, 2015.
- Sun, W., Ishidaira, H., and Bastola, S.: Calibration of hydrological models in ungauged basins based on satellite radar altimetry observations of river water level, *Hydrological Processes*, 26, 3524–3537, doi: 10.1002/hyp.8429, 2012.
- Sun, W., Ishidaira, H., Bastola, S., and Yu, J.: Estimating daily time series of streamflow using hydrological model calibrated based on satellite observations of river water surface width: Toward real world applications, *Environmental Research*, 139, 36–45, doi: 10.1016/j.envres.2015.01.002, 2015a.
- Sun, W., Song, H., Cheng, T., and Yu, J.: Calibration of hydrological models using TOPEX/Poseidon radar altimetry observations, *Proc. IAHS*, 368, 3–8, doi: 10.5194/piahs-368-3-2015, 2015b.
- Sun, W., Fan, J., Wang, G., Ishidaira, H., Bastola, S., Yu, J., Fu, Y. H., Kiem, A. S., Zuo, D., and Xu, Z.: Calibrating a hydrological model in a regional river of the Qinghai–Tibet plateau using river water width determined from high spatial resolution satellite images, *Remote Sensing of Environment*, 214, 100–114, doi: 10.1016/j.rse.2018.05.020, 2018a.
- Sun, Z., Zhu, X., Pan, Y., Zhang, J., and Liu, X.: Drought evaluation using the GRACE terrestrial water storage deficit over the Yangtze River Basin, China, *Science of The Total Environment*, 634, 727–738, doi: 10.1016/j.scitotenv.2018.03.292, 2018b.
- Swenson, S.: GRACE monthly land water mass grids NETCDF RELEASE 5.0, doi: 10.5067/TELND-NC005, 2012.
- Swenson, S. and Wahr, J.: Post-processing removal of correlated errors in GRACE data, *Geophys. Res. Lett.*, 33, L08402, doi: 10.1029/2005GL025285, 2006.
- Tang, X., Zhang, J., Gao, C., Ruben, G. B., and Wang, G.: Assessing the uncertainties of four precipitation products for SWAT modeling in Mekong River Basin, *Remote Sensing*, 11, doi: 10.3390/rs11030304, 2019.
- Tang, Y., Hooshiyar, M., Zhu, T., Ringler, C., Sun, A. Y., Long, D., and Wang, D.: Reconstructing annual groundwater storage changes in a large-scale irrigation region using GRACE data and Budyko model, *Journal of Hydrology*, 551, 397–406, doi: 10.1016/j.jhydrol.2017.06.021, 2017.
- Tangdamrongsub, N., Han, S.-C., Tian, S., Müller Schmied, H., Sutanudjaja, E. H., Ran, J., and Feng, W.: Evaluation of Groundwater Storage Variations Estimated from GRACE Data Assimilation and State-of-the-Art Land Surface Models in Australia and the North China

- Plain, *Remote Sensing*, 10, 483, doi: 10.3390/rs10030483, 2018.
- Tarpanelli, A., Barbetta, S., Brocca, L., and Moramarco, T.: River discharge estimation by using altimetry data and simplified flood routing modeling, *Remote Sensing*, 5, 4145–4162, doi: 10.3390/rs5094145, 2013.
- Tarpanelli, A., Amarnath, G., Brocca, L., Massari, C., and Moramarco, T.: Discharge estimation and forecasting by MODIS and altimetry data in Niger-Benue River, *Remote Sensing of Environment*, 195, 96–106, doi: 10.1016/j.rse.2017.04.015, 2017.
- Taylor, J. G., Stewart, T. R., and Downton, M.: Perceptions of Drought in the Ogallala Aquifer Region, *Environment and Behavior*, 20, 150–175, doi: 10.1177/0013916588202002, 1988.
- The World Bank: The Zambezi River Basin: A Multi-Sector Investment Opportunities Analysis, 2010.
- Thiemig, V., Rojas, R., Zambrano-Bigiarini, M., Levizzani, V., and De Roo, A.: Validation of satellite-based precipitation products over sparsely Gauged African River basins, *Journal of Hydrometeorology*, 13, 1760–1783, doi: 10.1175/JHM-D-12-032.1, 2012.
- Thomas, A. C., Reager, J. T., Famiglietti, J. S., and Rodell, M.: A GRACE-based water storage deficit approach for hydrological drought characterization, *Geophysical Research Letters*, 41, 1537–1545, doi: 10.1002/2014GL059323, 2014.
- Tirivarombo, S. and Hughes, D. A.: Regional droughts and food security relationships in the Zambezi River Basin, *Physics and Chemistry of the Earth, Parts A/B/C*, 36, 977–983, doi: 10.1016/j.pce.2011.08.003, 2011.
- Tomkins, K. M.: Uncertainty in streamflow rating curves: Methods, controls and consequences, *Hydrological Processes*, 28, 464–481, doi: 10.1002/hyp.9567, 2014.
- Tourian, M. J., Sneeuw, N., and Bárdossy, A.: A quantile function approach to discharge estimation from satellite altimetry (ENVISAT), *Water Resources Research*, 49, 4174–4186, doi: 10.1002/wrcr.20348, 2013.
- Tourian, M. J., Tarpanelli, A., Elmi, O., Qin, T., Brocca, L., Moramarco, T., and Sneeuw, N.: Spatiotemporal densification of river water level time series by multimission satellite altimetry, *Water Resources Research*, 52, 1140–1159, doi: 10.1002/2015WR017654, 2016.
- Tourian, M. J., Schwatke, C., and Sneeuw, N.: River discharge estimation at daily resolution from satellite altimetry over an entire river basin, *Journal of Hydrology*, 546, 230–247, doi: 10.1016/j.jhydrol.2017.01.009, 2017.
- Trambauer, P., Maskey, S., Winsemius, H., Werner, M., and Uhlenbrook, S.: A review of continental scale hydrological models and their suitability for drought forecasting in (sub-Saharan) Africa, *Physics and Chemistry of the Earth, Parts A/B/C*, 66, 16–26, doi: 10.1016/j.pce.2013.07.003, 2013.
- Tshili, N.: Zesa reverts to Stage 2 load shedding, <https://www.chronicle.co.zw/zesa-reverts-to-stage-2-load-shedding/>, 2019.
- Turc, L.: Le bilan d'eau des sols: relations entre les précipitations, l'évaporation et l'écoulement, 1953.
- United Nations Office for the Coordination of Humanitarian Affairs (OCHA): Southern Africa: Cyclones Idai and Kenneth Snapshot, as of 10 July 2019, <https://reliefweb.int/report/mozambique/southern-africa-cyclones-idai-and-kenneth-snapshot-10-july-2019>, 2019.
- University of East Anglia Climatic Research Unit, Harris, I., and Jones, P.: CRU TS4.01: Climatic Research Unit (CRU) Time-Series (TS) version 4.01 of high-resolution gridded data of month-by-month variation in climate (Jan. 1901- Dec. 2016), Centre for Environmental Data Analysis, doi: doi:10.5285/58a8802721c94c66ae45c3baa4d814d0, 2017.
- Urquijo, J. and De Stefano, L.: Perception of Drought and Local Responses by Farmers: A Perspective from the Jucar River Basin, Spain, *Water Resources Management*, 30, 577–

- 591, doi: 10.1007/s11269-015-1178-5, 2016.
- van Dijk, A. I. J. M. and Renzullo, L. J.: Water resource monitoring systems and the role of satellite observations, *Hydrol. Earth Syst. Sci.*, 15, 39–55, doi: 10.5194/hess-15-39-2011, 2011.
- van Dijk, A. I. J. M., Beck, H. E., Crosbie, R. S., de Jeu, R. A. M., Liu, Y. Y., Podger, G. M., Timbal, B., and Viney, N. R.: The Millennium Drought in southeast Australia (2001–2009): Natural and human causes and implications for water resources, ecosystems, economy, and society, *Water Resources Research*, 49, 1040–1057, doi: 10.1002/wrcr.20123, 2013.
- Van Loon, A. F.: Hydrological drought explained, *WIREs Water*, 2, 359–392, doi: 10.1002/wat2.1085, 2015.
- Van Loon, A. F. and Laaha, G.: Hydrological drought severity explained by climate and catchment characteristics, *Journal of Hydrology*, 526, 3–14, doi: 10.1016/j.jhydrol.2014.10.059, 2015.
- Van Loon, A. F., Tiedeman, E., Wanders, N., Van Lanen, H. A. J., Teuling, A. J., and Uijlenhoet, R.: How climate seasonality modifies drought duration and deficit, *Journal of Geophysical Research: Atmospheres*, 119, 4640–4656, doi: 10.1002/2013JD020383, 2014.
- Van Loon, A. F., Stahl, K., Di Baldassarre, G., Clark, J., Rangelcroft, S., Wanders, N., Gleeson, T., Van Dijk, A. I. J. M., Tallaksen, L. M., Hannaford, J., Uijlenhoet, R., Teuling, A. J., Hannah, D. M., Sheffield, J., Svoboda, M., Verbeiren, B., Wagener, T., and Van Lanen, H. A. J.: Drought in a human-modified world: reframing drought definitions, understanding, and analysis approaches, *Hydrol. Earth Syst. Sci.*, 20, 3631–3650, doi: 10.5194/hess-20-3631-2016, 2016.
- Van Loon, A. F., Kumar, R., and Mishra, V.: Testing the use of standardised indices and GRACE satellite data to estimate the European 2015 groundwater drought in near-real time, *Hydrol. Earth Syst. Sci.*, 21, 1947–1971, doi: 10.5194/hess-21-1947-2017, 2017.
- Vatanchi, S. M. and Maghrebi, M. F.: Uncertainty in Rating-Curves Due to Manning Roughness Coefficient, *Water Resources Management*, 33, 5153–5167, doi: 10.1007/s11269-019-02421-6, 2019.
- Velpuri, N. M., Senay, G. B., and Asante, K. O.: A multi-source satellite data approach for modelling Lake Turkana water level: calibration and validation using satellite altimetry data, *Hydrol. Earth Syst. Sci.*, 16, 1–18, doi: 10.5194/hess-16-1-2012, 2012.
- Vervoort, R. W., Miechels, S. F., van Ogtrop, F. F., and Guillaume, J. H. A.: Remotely sensed evapotranspiration to calibrate a lumped conceptual model: Pitfalls and opportunities, *Journal of Hydrology*, 519, Part D, 3223–3236, doi: 10.1016/j.jhydrol.2014.10.034, 2014.
- Vishwakarma, D. B., Devaraju, B., and Sneeuw, N.: What Is the Spatial Resolution of grace Satellite Products for Hydrology?, *Remote Sensing*, 10, 852, doi: 10.3390/rs10060852, 2018.
- Wahr, J., Molenaar, M., and Bryan, F.: Time variability of the Earth's gravity field: Hydrological and oceanic effects and their possible detection using GRACE, *Journal of Geophysical Research: Solid Earth*, 103, 30 205–30 229, doi: 10.1029/98JB02844, 1998.
- Wang, W., Li, J., Yu, Z., Ding, Y., Xing, W., and Lu, W.: Satellite retrieval of actual evapotranspiration in the Tibetan Plateau: Components partitioning, multidecadal trends and dominated factors identifying, *Journal of Hydrology*, 559, 471–485, doi: 10.1016/j.jhydrol.2018.02.065, 2018.
- Wang-Erlandsson, L., Bastiaanssen, W. G. M., Gao, H., Jägermeyr, J., Senay, G. B., van Dijk, A. I. J. M., Guerschman, J. P., Keys, P. W., Gordon, L. J., and Savenije, H. H. G.: Global root zone storage capacity from satellite-based evaporation, *Hydrol. Earth Syst. Sci.*, 20, 1459–1481, doi: 10.5194/hess-20-1459-2016, 2016.
- Warburton, M. L., Schulze, R. E., and Jewitt, G. P. W.: Hydrological impacts of land use change

- in three diverse South African catchments, *Journal of Hydrology*, 414–415, 118–135, doi: 10.1016/j.jhydrol.2011.10.028, 2012.
- Weerasinghe, I., Bastiaanssen, W., Mul, M., Jia, L., and van Griensven, A.: Can we trust remote sensing evapotranspiration products over Africa?, *Hydrol. Earth Syst. Sci.*, 24, 1565–1586, doi: 10.5194/hess-24-1565-2020, 2020.
- Werth, S., Güntner, A., Petrovic, S., and Schmidt, R.: Integration of GRACE mass variations into a global hydrological model, *Earth and Planetary Science Letters*, 277, 166–173, doi: 10.1016/j.epsl.2008.10.021, 2009.
- Werth, S., White, D., and Bliss, D. W.: GRACE Detected Rise of Groundwater in the Sahelian Niger River Basin, *Journal of Geophysical Research: Solid Earth*, 122, 10 459–10 477, doi: 10.1002/2017JB014845, 2017.
- Westerberg, I., Guerrero, J. L., Seibert, J., Beven, K. J., and Halldin, S.: Stage-discharge uncertainty derived with a non-stationary rating curve in the Choluteca River, Honduras, *Hydrological Processes*, 25, 603–613, doi: 10.1002/hyp.7848, 2011.
- Westerhoff, R. S.: Using uncertainty of Penman and Penman–Monteith methods in combined satellite and ground-based evapotranspiration estimates, *Remote Sensing of Environment*, 169, 102–112, doi: 10.1016/j.rse.2015.07.021, 2015.
- Wilhite, D. A. and Glantz, M. H.: Understanding the Drought Phenomenon: The Role of Definitions, *Water International*, 10, 111–120, doi: 10.1080/02508068508686328, 1985.
- Willems, P.: Parsimonious rainfall–runoff model construction supported by time series processing and validation of hydrological extremes – Part 1: Step-wise model-structure identification and calibration approach, *Journal of Hydrology*, 510, 578–590, doi: 10.1016/j.jhydrol.2014.01.017, 2014.
- Wingham, D., Rapley, C., and H D, G.: New Techniques in Satellite Altimeter Tracking Systems, in: *IGARSS 86 Symposium*, Zurich, 1986.
- Winsemius, H. C., Savenije, H. H. G., Gerrits, A. M. J., Zapreeva, E. A., and Klees, R.: Comparison of two model approaches in the Zambezi river basin with regard to model reliability and identifiability, *Hydrol. Earth Syst. Sci.*, 10, 339–352, doi: 10.5194/hess-10-339-2006, 2006a.
- Winsemius, H. C., Savenije, H. H. G., van de Giesen, N. C., van den Hurk, B. J. J. M., Zapreeva, E. A., and Klees, R.: Assessment of Gravity Recovery and Climate Experiment (GRACE) temporal signature over the upper Zambezi, *Water Resources Research*, 42, W12 201, doi: 10.1029/2006WR005192, 2006b.
- Winsemius, H. C., Savenije, H. H. G., and Bastiaanssen, W. G. M.: Constraining model parameters on remotely sensed evaporation: justification for distribution in ungauged basins?, *Hydrol. Earth Syst. Sci.*, 12, 1403–1413, doi: 10.5194/hess-12-1403-2008, 2008.
- Xu, S., Yu, Z., Yang, C., Ji, X., and Zhang, K.: Trends in evapotranspiration and their responses to climate change and vegetation greening over the upper reaches of the Yellow River Basin, *Agricultural and Forest Meteorology*, 263, 118–129, doi: 10.1016/j.agrformet.2018.08.010, 2018.
- Xu, X., Li, J., and Tolson, B.: Progress in integrating remote sensing data and hydrologic modeling, *Progress in Physical Geography*, 38, 464–498, doi: 10.1177/0309133314536583, 2014.
- ZAMCOM, SADC, and SARDC: *Zambezi Environment Outlook 2015*, 2015.
- Zhang, D., Zhang, Q., Werner, A. D., and Liu, X.: GRACE-Based Hydrological Drought Evaluation of the Yangtze River Basin, China, *Journal of Hydrometeorology*, 17, 811–828, doi: 10.1175/JHM-D-15-0084.1, 2015a.
- Zhang, J., Liu, K., and Wang, M.: Seasonal and Interannual Variations in China’s Groundwater Based on GRACE Data and Multisource Hydrological Models, *Remote Sensing*, 12, 845, doi:

- 10.3390/rs12050845, 2020.
- Zhang, K., Kimball, J. S., and Running, S. W.: A review of remote sensing based actual evapotranspiration estimation, *Wiley Interdisciplinary Reviews: Water*, 3, 834–853, doi: 10.1002/wat2.1168, 2016.
- Zhang, Z., Chao, B. F., Chen, J., and Wilson, C. R.: Terrestrial water storage anomalies of Yangtze River Basin droughts observed by GRACE and connections with ENSO, *Global and Planetary Change*, 126, 35–45, doi: 10.1016/j.gloplacha.2015.01.002, 2015b.
- Zhao, M., A. G., Velicogna, I., and Kimball, J. S.: Satellite Observations of Regional Drought Severity in the Continental United States Using GRACE-Based Terrestrial Water Storage Changes, *Journal of Climate*, 30, 6297–6308, doi: 10.1175/JCLI-D-16-0458.1, 2017a.
- Zhao, M., A. G., Velicogna, I., and Kimball, J. S.: A Global Gridded Dataset of GRACE Drought Severity Index for 2002–14: Comparison with PDSI and SPEI and a Case Study of the Australia Millennium Drought, *Journal of Hydrometeorology*, 18, 2117–2129, doi: 10.1175/JHM-D-16-0182.1, 2017b.
- Zhou, X. and Wang, H.: Application of Google Earth in Modern River Sedimentology Research, *Journal of Geoscience and Environment Protection*, 3, 1–8, doi: 10.4236/gep.2015.38001, 2015.
- Zimba, H., Coenders-Gerrits, M., Kawawa, B., Savenije, H., Nyambe, I., and Winsemius, H.: Variations in Canopy Cover and Its Relationship with Canopy Water and Temperature in the Miombo Woodland Based on Satellite Data, *Hydrology*, 7, doi: 10.3390/hydrology7030058, 2020.
- Zink, M., Mai, J., Cuntz, M., and Samaniego, L.: Conditioning a Hydrologic Model Using Patterns of Remotely Sensed Land Surface Temperature, *Water Resources Research*, 54, 2976–2998, doi: 10.1002/2017WR021346, 2018.
- Łyszkowicz, A. B. and Bernatowicz, A.: Current state of art of satellite altimetry, *Geodesy and Cartography*, 66, 259–270, doi: 10.1515/geocart-2017-0016, 2017.

Acknowledgements

This research was supported by the TU Delft | Global Initiative, a program of the Delft University of Technology to boost Science and Technology for Global Development, and NWO-WOTRO, a Dutch Organization for Scientific Research which aims to carry out research in support of the Sustainable Development Goals. The ZAM-SECUR project was a collaboration of Delft University of Technology (TU-Delft), University of Zimbabwe (UZ), University of Zambia (UNZA) and WaterNet. A summary of the research activities within this project can be found on <https://zamsecur.wixsite.com/home>.

This study would not have been possible without the help of those who provided us with the data. Local hydro-meteorological data was provided by WARMA (Water Resources Management Authority in Zambia), ZMD (Zambia Meteorological Department), GRDC (Global Runoff Data Centre) and NOAA (National Oceanic and Atmospheric Administration). Remotely sensed river water levels were obtained from DAHITI (Database for Hydrological Time Series of Inland Waters), HydroSat, EAPRS (Earth and Planetary Remote Sensing Laboratory) and LEGOS (Laboratoire d'Etudes en Géophysique et Océanographie Spatiales).

Curriculum Vitæ

Petra Hulsman

23-08-1990 Born in Leiden, Netherlands.

Education

1995–2009 Deutsche Evangelische Oberschule, in Cairo, Egypt

2009–2012 BSc Civil Engineering
Delft University of Technology, Netherlands

2012–2015 MSc Civil Engineering
Delft University of Technology, Netherlands &
National University of Singapore, Singapore

2021 PhD. Watermanagement
Thesis: Satellite data in rainfall-runoff models: Exploring
new opportunities for semi-arid, data-scarce river
basins
Promotor: Dr. M. Hrachowitz and Prof. dr. ir. H.H.G. Savenije

Experience

2015 Junior researcher
Delft University of Technology, Netherlands

2016 Water Allocation Plan Modelling Assistant
Delft University of Technology, Netherlands

List of Publications

P. Hulsman, H.H.G. Savenije, and M. Hrachowitz: *Zambezi River Basin: Drought of 2019*, Journal of Hydrology: Regional Studies, submitted.

P. Hulsman, H.H.G. Savenije, and M. Hrachowitz: *Why are long-term storage variations observed but not modelled in the Luangwa basin?*, Water Resources Research, in review.

P. Hulsman, H.H.G. Savenije, and M. Hrachowitz: *Learning from satellite observations: increased understanding of catchment processes through stepwise model improvement*, Hydrology and Earth System Sciences Discussions, doi: 10.5194/hess-2020-191, 2020.

P. Hulsman, H.C. Winsemius, C.I. Michailovsky, H.H.G. Savenije and M. Hrachowitz: *Using altimetry observations combined with GRACE to select parameter sets of a hydrological model in a data-scarce region*, Hydrology and Earth System Sciences, 24, 3331–3359, doi: 10.5194/hess-24-3331-2020, 2020.

P. Hulsman, T.A. Bogaard and H.H.G. Savenije: *Rainfall-runoff modelling using river-stage time series in the absence of reliable discharge information: a case study in the semi-arid Mara River basin*, Hydrology and Earth System Sciences, 22, 5081–5095, doi: 10.5194/hess-22-5081-2018, 2018.

