Chair of Hydrological Modelling and Water Resources
Albert-Ludwigs-University Freiburg

Modelling Karst Groundwater Resources in Southwest China

Using the VarKarst model in a data scarce tropical environment

Master thesis submitted by
Laura-Marie Vecera

Examiner: Dr. Andreas Hartmann

Co-examiner: Prof. Dr. Stefan Hergarten

Scientific advisor: Fang Guo, Institute of Karst Geology (CAGS), Guilin

Freiburg im Breisgau, 11th of December 2017

Chair of Hydrological Modelling and Water Resources
Albert-Ludwigs-University Freiburg

Modelling Karst Groundwater Resources in Southwest China

Using the VarKarst model in a data scarce tropical environment

Master thesis submitted by
Laura-Marie Vecera

Examiner: Dr. Andreas Hartmann

Co-examiner: Prof. Dr. Stefan Hergarten

Scientific advisor: Fang Guo, Institute of Karst Geology (CAGS), Guilin

Freiburg im Breisgau, 11th of December 2017

Acknowledgements

I would like to thank all of the people who have made this thesis possible through their professional and personal support. Most importantly, my supervisor Dr. Andreas Hartmann, for his encouragement, patience and understanding as well as his advice and scientific input. His passion for the topic helped me to stay positive during all the lows of the writing of this thesis that might have occurred from time to time. Second, Mrs Fang Guo from the Institute of Karst Geology, without whom this thesis would not have been possible. Thank you for the co-operation, for supplying all the data and information on the study site and commitment in answering all my questions at all times. I would also like to thank Prof. Dr. Stefan Hergarten for co-referencing this thesis.

I also want to thank all of my friends and special people for bearing all my moods and doubts and getting me back on the track whenever I was about to get lost. Who supplied me with coffee, cookies and food to keep me going and were always there to share a laugh whenever work started to take its toll. Judith, thank you for your thorough proof-reading and your feedback.

Jesper, you have been a tremendous, priceless support throughout this whole time of this thesis. Thank you for everything!

And last but least I want to thank my mum. Thank you for always being there for me, be it day or night. This work is dedicated to you as you smoothed the way that helped me to get where I am today.

Contents

List of Figures i

List of Tables iii

List of Abbreviations & Symbols v

Abstract ix

Zusammenfassung xi

1 Introduction 1
- 1.1 Karst characteristics . . . 1
- 1.2 Karst in China and the Lingshui spring . . . 3
- 1.3 Exploration and management of karst groundwater resources . . . 6
- 1.4 Climate change impact studies and uncertainty . . . 10

2 Research questions and objectives 13

3 Study Site 15
- 3.1 Climate . . . 15
- 3.2 (Hydro-)Geology . . . 16

4 Methods 21
- 4.1 Data and quality assessment . . . 21
- 4.2 Rating curve calibration . . . 24
- 4.3 The VarKarst model . . . 26
- 4.4 Model calibration and evaluation . . . 32
  - 4.4.1 Model performance . . . 32
  - 4.4.2 Uncertainty analysis and parameter identifiability . . . 35
  - 4.4.3 Predictive capability . . . 36
- 4.5 Uncertainty under future climate scenarios . . . 37

5 Results 41
- 5.1 Rating curve calibration . . . 41
- 5.2 Model calibration and evaluation . . . 42
  - 5.2.1 Model performance . . . 42
  - 5.2.2 Uncertainty analysis and parameter identifiability . . . 45
  - 5.2.3 Predictive capability . . . 50
- 5.3 Uncertainty under future climate scenarios . . . 53

6 Discussion 57
- 6.1 Uncertainties from input/output data . . . 57
- 6.2 Model performance . . . 58
- 6.3 Uncertainties related to the rating curve . . . 61
- 6.4 Reliability of simulations . . . 62
- 6.5 Uncertainty under future climate scenarios . . . 66

7 Conclusion and Outlook 71

Bibliography 73

Appendix 81

A Discharge data 81

B Thornthwaite equation 82

C Parameter ranges of first calibration trial 83

D Groundwater stationarity - RNI 84

E Groundwater stationarity - RI 88

F Calibration results of EpiMax 92

G Parameter distributions of calibration trials (EpiMax) 93

H Uncorrected simulations of RNI 94

I Multi-objective calibration (Discharge - Water level) 95

List of Figures

Figure 1 General conceptual model of a karst aquifer 2
Figure 2 (a) Fenglin and (b) Fengcong landscapes of southwest China 4
Figure 3 Karst regions of southwest China 5
Figure 4 Conceptual depiction of the VarKarst model 10
Figure 5 Climate diagram for Nanning weather station (59431) 15
Figure 6 Location of the Lingshui spring and its catchment 16
Figure 7 Sketch of the karst system in the Wuming basin 17
Figure 8 (Hydro-)geological map of the Lingshui karst system 18
Figure 9 Map of the Lingshui spring pool 19
Figure 10 Water level duration curve of 2010, 2011 and 2016/2017 22
Figure 11 Filling of climate data gaps for daily (a) temperature and (b) precipitation sums 24
Figure 12 Plots of monthly (a) mean temperature and (b) sum of precipitation for 2010, 2011, 2016/2017 25
Figure 13 Comparison of PET data from 2010/2011: Fang Guo - Thornthwaite 26
Figure 14 Overview of methodology 31
Figure 15 Probabilistic climate scenarios adapted from Chen et al. (2011) . . 38
Figure 16 Conceptualization of the evaluation method established for comparing the model uncertainties under current conditions to those from scenarios under future conditions 39
Figure 17 Rating curve calculated from measured discharge data with least square fitted and modelled parameters. 41
Figure 18 Distribution of variable parameters resulting from RNI and RI for (a) the first calibration run and (b) the final calibration run 44
Figure 19 Comparison of modelled and observed values over the calibration time period for the final SCEM calibration of (a) RNI and (b) RI . . 45
Figure 20 Identifiability plots for RNI and RI 47
Figure 21 Distribution of model parameters (normalised by their ranges) for acceptable parameter sets of (a) RNI and (b) RI 48
Figure 22 Simulated water level from 50 selected parameter sets compared to observed water level over the calibration period (2010/2011) produced by (a) RNI and (b) RI. 49
Figure 23 Duration curve of observed and modelled water levels over the calibration time period (2010/2011) derived from (a) RNI and (b) RI . 50
Figure 24 Simulated water level from 50 selected parameter sets compared to observed water level over the validation period (2016/2017) produced by (a) RNI and (b) RI 51
Figure 25 Duration curve of observed and modelled water levels over the validation time period (2016/2017) derived from (a) RNI and (b) RI . . 52
Figure 26 Yearly (a) means of temperature and (b) sums of precipitation over the period of 1951-2016 calculated from the GHCN-D dataset . . . 53
Figure 27 Plot of results from analysis of uncertainties under current conditions and future scenarios 55
Figure D.1 Groundwater compartments(RNI - EpiMax - First calibration) . . . 84
Figure D.2 Groundwater compartments (RNI - SoiMaxEpiMax - First calibration) 85

Figure D.3 Groundwater compartments (RNI - EpiMax - Final calibration) . . 86
Figure D.4 Groundwater compartments (RNI - SoiMaxEpiMax - Final calibration) 87
Figure E.1 Groundwater compartments (RI - EpiMax - First calibration) 88
Figure E.2 Groundwater compartments (RI - SoiMaxEpiMax - First calibration) 89
Figure E.3 Groundwater compartments (RI - EpiMax - Final calibration) . . . 90
Figure E.4 Groundwater compartments (RI - SoiMaxEpiMax - Final calibration) 91
Figure G.1 Distribution of variable parameters resulting from RNI and RI for (a) the first calibration run and (b) the final calibration run (EpiMax) 93
Figure H.1 Modelled water level from 50 selected parameter sets from RNI before correction 94
Figure I.1 Comparison of modelled and observed values of water level and discharge over the calibration time period for a multi-objective calibration trial with RI 95

List of Tables

Table 1 Available data for the modelling study 21
Table 2 Description of parameters and parameter ranges used for final calibration trial (parameter ranges adapted from first calibration) . . . 33
Table 3 Values for climate change scenarios extracted from Chen et al. (2011) 38
Table 4 Results for parameters and objective functions for different calibration steps and rating curve approaches for SoiMaxEpiMax 43
Table 5 Statistics of uncertainty analysis by model efficiencies KGE_{rvar} (RNI) and KGE (RI) from acceptable model runs 46
Table 6 Percentage of exceedance probability points at which model runs for current and future conditions were significantly different according to Welch's t-test 54
Table A.1 Discharge data 81
Table C.1 Description of parameters and parameter ranges used for first calibration trial 83
Table F.1 Results for parameters and objective functions for different calibration steps and rating curve approaches for EpiMax 92

List of Abbreviations & Symbols

Abbreviations

a.s.l.	above sea level
AOGCMs	Atmosphere–Ocean General Circulation Models
BC	Before Christ
CMDC	China Meteorological Data Service Center
DEM	Digital Elevation Model
DJF	December, January, February (winter)
EpiMax	Epikarst Maximum (starting condition)
EPM	Equivalent Porous Medium Model
GCM	Global Climate Model
GHCN-D	Global Historical Climatological Network - Daily
GIS	Geographic Information System
GLUE	Generalised Likelihood Uncertainty Estimation
HSY	Hornberger - Spear - Young
JJA	June, July, August (summer)
KGE	Kling Gupta Efficiency
KGE_{rvar}	Kling Gupta Efficiency with β excluded
MC	Monte Carlo
MCAT	Monte Carlo Analysis Toolbox
MCMC	Monte Carlo Markov Chain
NA	Not Available
NSE	Nash Sutcliffe Efficiency
pc	Pearson's correlation coefficient
PET	Potential Evapotranspiration
RCM	Regional Climate Model
RI	Rating curve Included
RMSE	Root Mean Square Erorr

RNI	Rating curve Not Included	
SCE-UA	Shuffled Complex Evolution - University of Arizona	
SCEM-UA	Shuffled Complex Evolution Metropolis - University of Arizona	
SoiMaxEpiMax	Soil Maximum, Epikarst Maximum (starting conditions)	
SRES A1B	Special Report on Emissions Scenarios A1B	

Chemical formulae

CO_2	Carbon dioxide	
Cl^-	Chloride	
$HCO_3{}^-$	Hydrogencarbonate	
$NO_3{}^-$	Nitrate	
$SO_4{}^{2-}$	Sulfate	

Symbols

α	Measure of relative variability (KGE)	
α'	Normalized measure of relative variability (KGE)	
β	Measure of volume bias (KGE)	
γ	Equation parameter (Thornthwaite)	
μ_S, μ_O	Mean (simulated and observed)	
σ_S, σ_O	Standard deviation (simulated and observed)	
A	Recharge area	km^2
a, b, c	Calibration coefficients (Rating curve)	
a_{fsep}	Recharge separation variability constant	
a_{GW}	Groundwater variability constant	
a_{SE}	Soil/epikarst variability constant	
d	Sampling Distribution	
$E_{act,i}(t)$	Actual evapotranspiration	mm
$E_{pot}(t)$	Potential evapotranspiration	mm
ETp_{THORN}	Potential evapotranspiration after Thornthwaite	mm

$f_{C,i}$	Variable separation factor	
h	Water level	m
I	Warmth index	
$InSurfEpi_i(t-1)$	Redistributed excess water from epikarst to epikarst	mm
$InSurfSoi_i(t-1)$	Redistributed excess water from epikarst to soil	mm
K_C	Conduit storage coefficient	d
$K_{E,i}$	Variable epikarst storage coefficient	d
$K_{GW,i}$	Variable groundwater storage coefficient	d
$K_{max,E}$	Maximum epikarst storage coefficient	d
$K_{mean,E}$	Mean epikarst storage coefficient	d
L	Mean day length	h
N	Number of model compartments	
P	Precipitation	mm
Q	Discharge	$\mathrm{l \cdot s^{-1}}$
$Q_{Epi,i}(t)$	Flow from epikarst to groundwater	mm
$Q_{GW,i}(t)$	Groundwater contributions of the matrix	mm
$Q_{GW,N}(t)$	Groundwater contributions of the conduit system	mm
$Q_{spring}(t)$	Spring discharge	$\mathrm{l \cdot s^{-1}}$
r	Linear correlation coefficient (simulated and observed) (KGE)	
$R_{conc,i}(t)$	Concentrated recharge	mm
$R_{diff,i}(t)$	Diffuse recharge	mm
$R_{Epi,i}(t)$	Recharge to the epikarst	mm
T	Daily mean of temperature	°C
T_{mon}	Monthly mean of temperature	°C
$V_{E,i}$	Epikarst storage capacity	mm
$V_{Epi,i}(t)$	Actual amount of water in epikarst	mm
$V_{GW,i}(t)$	Actual amount of water in matrix storage	mm

$V_{GW,N}(t)$	Actual amount of water in conduit storage	mm
$V_{max,E}$	Maximum epikarst storage capacity	mm
$V_{max,S}$	Maximum soil storage capacity	mm
$V_{mean,E}$	Mean epikarst storage capacity	mm
$V_{mean,S}$	Mean soil storage capacity	mm
$V_{S,i}$	Soil storage capacity	mm
$V_{Soil,i}(t)$	Actual amount of water in soil	mm

Abstract

Karst groundwater resources play a key role in China. The Lingshui spring in the Southwest has until recently been a reliable water supply for 100,000 people but its quantity and quality are increasingly threatened by rapid land use changes as well as climate change. A model description of the hydrodynamics of the spring is needed to be able to make predictions on the future availability of the resource which could then serve as a scientific basis for groundwater management plans. The hydrological lumped parameter model VarKarst was chosen for this purpose as it is highly flexible due to its semi-distributed character. As VarKarst has never been tested in a tropical setting before, this study also seeks to explore the feasibility of using VarKarst in such a context. The main focus of this study was to test the model performance on water level data by finding an appropriate calibration for the model parameters. The scarce data availability for the study site was expected to introduce large uncertainties during the modelling process. These were evaluated with the help of a Monte Carlo sampling and a subsequent analysis of parameter identifiability and information content of water level data for parameter range confinement. It was also tested if the inclusion of a rating curve and thus further parameters in the calibration process leads to additional uncertainties. A statistical evaluation approach was employed to assess under which future climate scenarios the model can be expected to make reliable predictions. Generally, the VarKarst model performed well in the Lingshui spring catchment, especially considering its simulation of water level dynamics. Nevertheless, the low quantity of data available for calibration resulted in high parameter non-identifiability, particularly when the rating curve was included in the calibration. In spite of the reduced reliability of the model results, the predictive capability of the model was satisfactory enough to make it useful for future predictions. Uncertainty bands under current and future conditions often differed significantly, resulting in minimal overlap, so that the predictions can be regarded as reliable. Confidence in the model results might be significantly increased by further measurement campaigns in the catchment, e.g. by analysing hydrochemical parameters and recording groundwater levels at regular intervals. In spite of the large uncertainties caused by the scarcity of data, the model was successfully tested in the catchment and once further data are available, it could be a helpful and reliable tool for water management decisions in karst regions in the face of climate change.

Keywords: Southwest China, VarKarst, Lingshui Spring, Karst Modelling, Monte Carlo, SCEM-UA, Model Uncertainty, Rating Curve, Equfinality, Climate Change Impacts

Zusammenfassung

Karstgrundwasserressourcen spielen eine wichtige Rolle in China. Die Lingshui Quelle im Südwesten Chinas bot in der Vergangenheit eine zuverlässige Trinkwasserversorgung für 100.000 Menschen. Vor dem Hintergrund schnell fortschreitender Landnutzungsänderungen und dem Klimawandel wird der quantitative und qualitative Zustand dieser Quelle zunehmend gefährdet. Um Vorhersagen über die zukünftige Verfügbarkeit der Ressource machen zu können, ist die Beschreibung ihrer Hydrodynamik durch ein Modell unverzichtbar. Diese stellt außerdem eine wichtige wissenschaftliche Grundlage für die Formulierung von Grundwassermanagementstrategien für den Aquifer dar. Das semidistribuierte VarKarst Modell wurde aufgrund seiner großen Flexibilität für diesen Zweck ausgewählt. Des Weiteren sollte diese Studie zeigen, ob das Modell für einen tropischen Standort geeignet ist. Daher lag der Fokus darauf, die Modellgüte anhand einer Parameterkalibrierung für Wasserstandsdaten zu testen. Aufgrund der geringen Datenverfügbarkeit für den Standort konnten große Unsicherheiten in der Modellierung erwartet werden. Daher wurde ein Monte Carlo Sampling und eine anschließende Analyse der Bestimmbarkeit der Parameter und der Aussagekraft der verfügbaren Wasserstandszeitreihe durchgeführt. Außerdem wurde getestet, ob die Einbringung einer Wasserstands-Abfluss-Beziehung in die Kalibrierung aufgrund der zusätzlichen Parameter zu weiteren Unsicherheiten führt. Ein statistischer Ansatz wurde gewählt, um die Zuverlässigkeit der Modellvorhersagen unter verschiedenen Klimaprognosen zu testen. Generell waren die Modellergebnisse in Bezug auf die Wasserstandsdynamik der Lingshui Quelle zufriedenstellend. Trotzdem führte die geringe Datenverfügbarkeit, wie erwartet, zu einer verminderten Identifizierbarkeit angemessener Parameterwerte, vor allem wenn die Wasserstands-Abfluss-Beziehung in die Kalibrierung eingebracht wurde. Trotzdem wurde die Vorhersagefähigkeit des Modells als ausreichend eingestuft, um es für Zukunftsprojektionen zu verwenden. Die Unsicherheitsbänder der Simulationen unterschieden sich meist signifikant. Ihre Überlappung war so gering, dass die Zukunftsprojektionen des Modells als zuverlässig betrachtet werden können. Die Glaubwürdigkeit der Modellergebnisse könnte durch zusätzliche Messkampagnen, z.B. für hydrochemische Parameter und Grundwasserstände, deutlich verbessert werden. Trotz der großen Unsicherheiten durch die geringe Datenverfügbarkeit wurde das VarKarst Modell erfolgreich für die Lingshui Quelle eingesetzt. Sobald zusätzliche Daten vorhanden sind, könnte das Modelle eine hilfreiche und zuverlässige Grundlage für Entscheidungen bezüglich des Wasserressourcenmanagements in der Region darstellen.

Stichworte: Südwest-China, VarKarst, Lingshui Quelle, Karst Modellierung, Monte Carlo, SCEM-UA, Modellunsicherheit, Wasserstands-Abfluss-Beziehung, Equifinalität, Klimawandel

1 Introduction

All models are wrong but some are useful.

Box (1979)

1.1 Karst characteristics

Karst aquifers cover 7-12% of the Earth's continental area (Hartmann et al., 2014a). Characteristic for the hydrogeology of karst systems is their large-scale heterogeneity and anisotropy (Figure 1). This is mainly due to the different flow systems that evolve in carbonate rocks under the dissolution by CO_2 ranging from micropores (also referred to as primary or matrix porosity) over fissures (secondary porosity, 50 - 500 μm) to solutionally enlarged conduits (tertiary porosity) with apertures above 1 cm (White, 2002; Hartmann et al., 2014a; Ford and Williams, 2007). Ford and Williams (2007) describe the porosities as follows: Micropores develop during the genesis of carbonate rocks while fissure porosity is a result from rock folding and faulting. The gradual widening of these fissures through dissolution by contact with circulating water and groundwater leads to the formation of conduits and even caves. Terms for the description of porosities in karst systems are not used homogeneously in scientific literature. Hartmann et al. (2014a) report that the first two categories of micropores and fissures are often summarized under the term matrix system as their hydrological behaviour differs fundamentally from that of conduits. This is sensible, as Kiraly (1998) and Worthington (2003) report that in many karst aquifers the duality is the predominant characteristic. It has to be kept in mind, how diverse (hydro-)geological karst settings can be and that all three kinds of porosities can form the prevailing hydrodynamic character of the aquifer (Worthington, 2003). Figure 1 depicts a classical conceptual model of a karst aquifer based on the idea dual porosity. Recharge processes as well as the groundwater flow field can be divided into those driven by the matrix system and those with fast flow characteristics in the conduit system. Even the discharge at the outlet can be marked by duality as low permeability volumes result in diffuse seepage while concentrated discharge rises from the channel system. Worthington (2003) showed that the matrix system often drives the water storage of a karst aquifer, whereas dynamics of spring flow at its outlet are mainly controlled by conduits. The so called epikarst forms the soil-bedrock-interface and is often marked by fast dissolution processes due to the close contact to precipitation which forms highly distinctive fractures in this part (Ford and Williams, 2007). During intense precipitation events and flooding conditions, the hydraulic gradient of a karst aquifer can reverse so that it then takes up water from surrounding rivers (Ford and Williams, 2007).

The evolution of underground channel networks also leads to the phenomenon that catch-

ments of karst aquifers can often not be identified by classic topographic analysis as done for non-karstic groundwater basins (Goldscheider and Drew, 2007). The subsurface conduits develop independently from the surface topography and can carry water across topographical watersheds (Goldscheider and Drew, 2007). This often implies that catchment divides are also spatially and temporally changing (Goldscheider and Drew, 2007). In studies on management of karst aquifers, this can have a serious impact on the water balance.

Conduit-fed karst springs often discharge larger amounts of water compared to springs from non-karstic rocks, which makes them an easily exploitable resource (White, 2002). On the other hand, the heterogeneity and anisotropy caused by the different porosities result in large temporal discharge variability and make the aquifer particularly vulnerable to contamination (Sauter et al., 2006). With many stakeholders requiring a reliable water supply from karst aquifers and springs, questions like: "How much water can be used? Where is it coming from and what are the physical parameters characterizing the system?" (p.145, Ford and Williams (2007)) need to be answered. Thus, thorough system

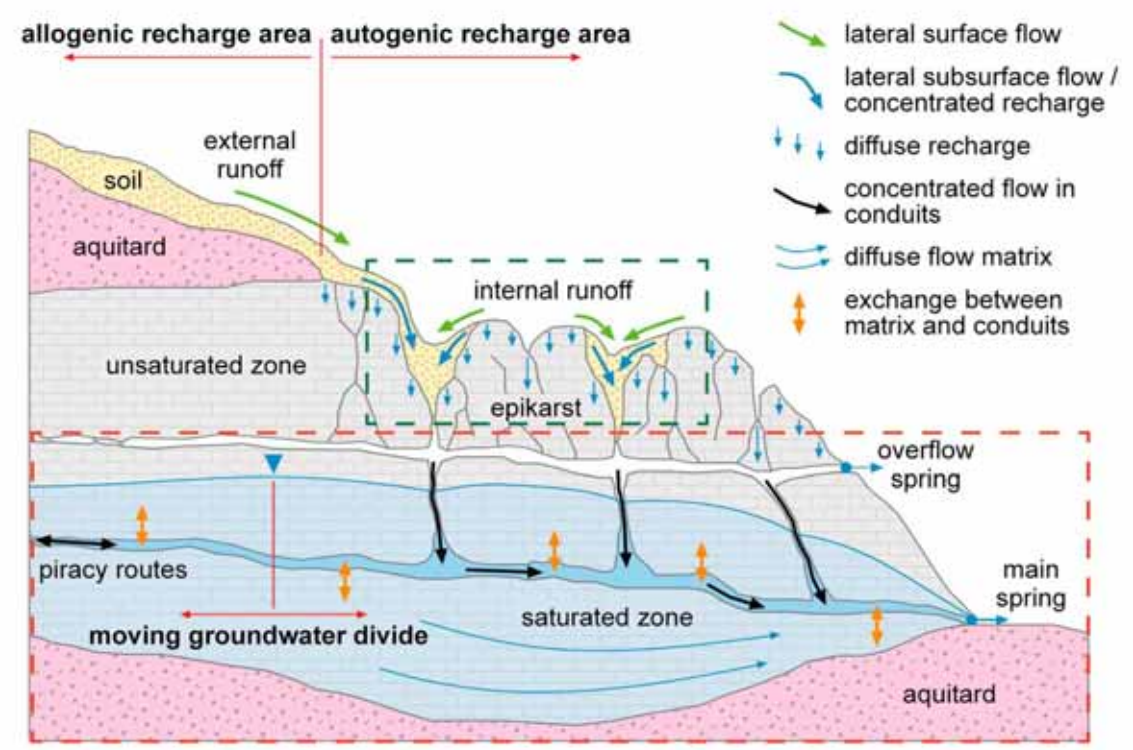


Figure 1: *Conceptual model of a karst aquifer. Above the surface, it is important to differentiate between allogenic and autogenic recharge areas as well as point infiltration (e.g. through sinkholes) and diffuse infiltration. The epikarst often forms the transition zone between surface and subsurface and is the zone of most intense karstification. Below, the distribution of matrix and conduit flowpaths (including caves and underground rivers) characterizes the karst system. Green dashed line: Soil/epikarst subsystem; red dashed line: groundwater subsystem. Retrieved from Hartmann et al. (2014a)*

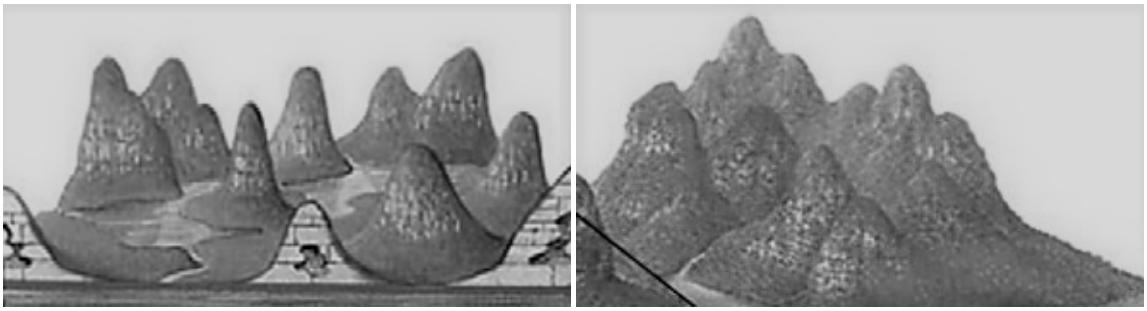
understanding becomes essential to develop management and protection plans to sustain the requirements of all stakeholders using the resource (Hartmann et al., 2014a).

1.2 Karst in China and the Lingshui spring

China's karst areas alone make up about 40% of the worldwide terrestrial karst. 1/7 of the country's area - about 1.3 million km^2 - is covered with openly perceivable karst (Sweeting, 1995). If covered and buried karst are added to this, the whole karst area of China adds up to 1/3 of the total Chinese territory (Sweeting, 1995). According to Guo et al. (2013), karst development in China is mainly based on pre-Triassic, old-phase, hard and compact carbonated rock. Covered karst develops under loose, unconsolidated sediment, while buried karst even occurs under non-soluble bed-rocks (Guo et al., 2013). Southwest China is probably one of the best known and scenic karst areas of the world. This is mainly because of the special landforms that have developed in the karst of this region. Sweeting (1995) lists several reasons for this development. First, the Chinese territory has been marked by intense tectonic activity, particularly Cenozoic uplifts. Secondly, the region is governed by tropical climate with high temperatures and large amounts and intensities of precipitation. This climate simulates a higher productivity of plants and therefore a higher CO_2 concentration in precipitation which is directly connected to higher carbonate solution rates. The main depositional period of main soluble rocks in southwest China occurred during the Caledonian and Variscan cycles Sweeting (1995). Therefore, rocks ages can be dated between the Devonian and Triassic. Sweeting (1995) reports that China's limestones usually have a porosity of less than 2% and a permeability of almost zero.

The best known landforms in southwest China, which are described in Sweeting (1995), can mainly be divided into two categories. Peak Forests (*Fenglin* 峰林) are isolated karst hills whose bases are at a similar level (Figure 2 (a)). They are usually divided by valleys which are often passed through by rivers. Another characteristic is the sparsely developed epikarst in peak forests (personal communication, Fang Guo (04.09.2017)). Guilin might be the Chinese city best known for its unique peak forests. Peak Clusters (*Fengcong* 峰丛), on the other hand, form groups of several peaks (cones and towers) which share a common basement (Figure 2 (b)). These do not necessarily have to be on the same surface level. In contrast to peak forests, they are often characterised by a well developed epikarst (personal communication, Fang Guo (04.09.2017)).

First records of karst spring usage for irrigation in China dates back to 453 BC in the Shanxi province (Ford and Williams, 2007). With China's growing economy and societal changes as well as its large population, karst water resources are increasingly put under stress (Western Develpoment Plan, 1999; Guo and Jiang (2011)). The Jinci spring in North China, fell dry within few years of exploitation (Hao et al., 2009). This example



(a) **(b)**

Figure 2: *(a) Fenglin and (b) Fengcong landscapes of southwest China. Images downloaded from: http://www.chinaspree.com/china-travel-guide/china-guizhou-guangxi-tours.html, retrieved 02.12.2017.*

shows the importance of a thorough understanding of karst systems used for water supply and additional resource management plans (Chalikakis et al., 2011; Li et al., 2012). As climate change can be expected to worsen this situation even more, Gu et al. (2017) see the question of how to deal with future climate change impacts on water resources availability as one of the key research requirements for China.

Southwestern China represents a subtropical region where a vast amount of area (about 500,000 km^2) is karstified and where around 100 million people depend on water supply from karst aquifers and springs (Guo and Jiang, 2011; Guo et al., 2013). All three types of karst (bare, covered and buried) occur in this region (Figure 3) (Zhang et al., 2013). Therefore, the loss or dry-up of a spring due to pollution or over-exploitation could be accompanied by serious supply problems for the local population.

The Lingshui spring system is situated in the Wuming basin, a county of Nanning city, the capital of the Guangxi region (Jiang and Guo, 2010). It supplies 120,000 people with its usually clean drinking water (Guo and Jiang, 2011). The spring pool is an important recreational area in the region and has been the training base for the national swimming team (Guo and Jiang, 2011; Guo et al., 2015, 2016). As the area is highly drained and produces low surface runoff, economic and social development was restricted in the past (Guo and Jiang, 2011). Nevertheless, over the last decades, the region has managed socioeconomic progress with intensified agriculture, mining activities, infrastructure development, a growing industrial base as well as urban expansion (Guo et al., 2010). This was inevitably accompanied by land use changes, e.g deforestation, urbanization and increasing eucalyptus cultivation as well as higher water consumption (Guo et al., 2016). Agricultural land use has changed from paddy land cultivation to dry land in form of economic forest planting which might also be a consequence of a reduced availability of water resources in the region (Guo et al., 2015). These developments and the lack of available surface water put the quantitative and qualitative status of the karst aquifer and its springs under serious threat, especially as many of these changes, e.g. the increase in mining activities, are highly water demanding. Guo et al. (2010) and Guo and Jiang

(2011) recognized in their study that developments in science and technology will be essential for promoting the sustainability and protection of karst environments in the region to:

- provide information about karst aquifer systems
- supply information on geological hazards
- map subsurface hydrology and geology to identify areas where productive water wells may be located as well as potential karst problems
- offer information for stakeholders
- supply solutions for environmental problems

Even though several policy acts have been passed to reduce threats to karst groundwater quality and quantity since the 1980's, due to different factors explained in Guo et al. (2010), groundbreaking changes in management practices could not be carried out so far. The current protection scheme in the region forbids drilling for pumping in the urban area and incorporates weekly water sampling for ions and bacteria as well as discharge measures on a small frequency (Jiang and Guo, 2010). Protection zones according to river boundaries and geology, strict control of underground water usage as well as higher discharge monitoring at monthly intervals are planned in the future (Jiang and Guo, 2010). The spring's karst system has been the object of several scientific studies, particularly

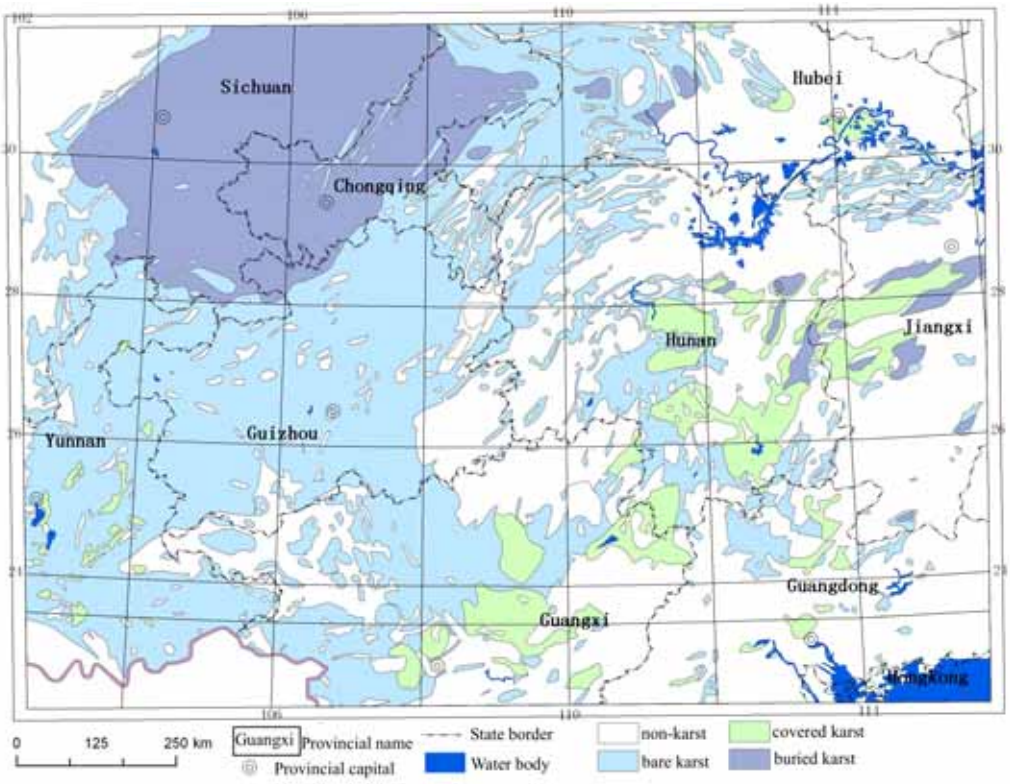


Figure 3: *Karst regions of southwest China. Retrieved from Guo and Jiang (2011).*

after a sewage accident in a starch factory polluted a spring in the system in 2010, cutting 4000 people from their water supply for over a month (Guo and Jiang, 2011). A hydrogeological survey was conducted in 2010 and quantity and quality of the Lingshui spring and other springs in the basin have been monitored between 2009 and 2012 in irregular intervals. In a study by Guo et al. (2015), measured results of these campaigns were compared to those of another campaign from the 1970's. Guo et al. (2015) tried to relate observable changes between these time periods to climate and land use change, concluding that the Lingshui spring discharge decreased and that the water temperature rose. Furthermore, the pollution with nitrate in the upper parts of the aquifer increased noticeably in recent years. Nevertheless, the karst system of the Lingshui spring has not been described by a numerical model so far. The concerning developments of the recent years make scientists, officials and locals in the area aware of the urge to manage the resource and formulate protection strategies (Guo and Jiang, 2011).

1.3 Exploration and management of karst groundwater resources

The exploration of karst groundwater resources often turns out to be a difficult task. The combination of lithology, structure, geomorphological history and climate produces highly individual karst systems (Ford and Williams, 2007). It is evident that this individuality makes generalizing approaches for system description, e.g. conceptual models, difficult (Hartmann et al., 2014a). Much research has been conducted in the last decades on methods to describe karst systems, assess available karst groundwater resources in these areas and assess their variability and quantity as well as quality (White, 2002). According to Goldscheider and Drew (2007), these include surface exploration and survey techniques (e.g. remote sensing or geophysical techniques (Chalikakis et al., 2011)), borehole analysis as well as natural or artificial tracer analysis (e.g. to delineate catchments and trace flow paths). All of these approaches are very important to understand the structure of the karst aquifer and its hydrological dynamics. Nevertheless, a description of the current status is not sufficient for water-resource management. To evaluate the impact of changing environmental conditions through e.g. pumping, land use or climate change on the water availability of an aquifer, predictions of future states become necessary (Scanlon et al., 2003). This requires the simplified translation of the functional characteristics of karst systems into process-based, deterministic models (Ford and Williams, 2007; Hartmann et al., 2014a). Hartmann et al. (2013a) even see these models as indispensable methods to sustainably manage present and future karst resources as they play such an important part in the world's drinking water supply.

The challenge in formulating models for complex karst systems lies in finding a model structure detailed enough to represent all important hydrological processes while at the same time keeping the number of parameters as low as possible to avoid overparametriz-

ation (referred to as parsimony) (Box, 1979; Kuczera and Mroczkowski, 1998; Wagener et al., 2001). The choices made in this context largely depend on the modelling task (Wagener et al., 2001). Distributed models attempt to incorporate the spatial and temporal dynamics of an aquifer to the highest degree possible. Reviews of these models are given in Sauter et al. (2006), Goldscheider and Drew (2007) and Ford and Williams (2007). All important physical processes in the catchment are modelled for every element on a 2D or 3D cell grid. The large number of parameters included in the physical descriptions restricts automatic calibration and thus relies heavily on data measured in the field or obtained from literature (Scanlon et al., 2003).

Classic equivalent porous medium models (EPM), are an example for distributed modelling approaches for karst systems (Hartmann et al., 2014a). They generalize the karst characteristics by assigning a value of hydraulic conductivity to a cell that represents the combined effects of matrix, fractures and conduits (Gondwe et al., 2011). Additionally, they are mostly based on the assumption of laminar flow in the pores that behaves according to Darcy's law (Hartmann et al., 2014a). These models are particularly useful at locations where a system of matrix and fissure porosity is dominating or where the cell volume can be chosen large enough to be representative of different conductivities (Teutsch, 1993; Scanlon et al., 2003). They were, for example, successfully tested in karst aquifers on the Yucatan Island (Mexico) (Gondwe et al., 2011), for the Edwards aquifer (USA) (Scanlon et al., 2003) and the Western Mountain Aquifer (Israel, Westbank) (Abusaada and Sauter, 2012).

Nevertheless, the basic assumption of flow according to Darcy's law is usually inappropriate in conduit-dominated aquifers (Sauter et al., 2006). Furthermore, extensive measuring campaigns, as they are needed to parametrize distributed models, are in most cases not feasible which is why lumped models have been introduced (see Rimmer and Hartmann (2012) for a detailed introduction). They summarize important structural units and their physical characteristics based on a conceptual model (as shown in Figure 1). The driving physical processes within and between these units are expressed by mostly (non-) linear functions whose parameters can be assessed by inverse calibration from output variables by using objective functions (e.g. Root Mean Square Error (RMSE), Nash-Sutcliffe-Efficiency (NSE), Kling Gupta Efficiency (KGE)) (Wagener et al., 2001; Gupta et al., 2009; Bennett et al., 2013).

Manual trial-and-error calibration of these model parameters is often work-intensive and reaches its limits for more complex models with parameter-interaction (Vrugt et al., 2003). Therefore, automatic calibration procedures have been developed that explore the space of parameter ranges similarly to manual calibration but with computational efficiency and objectivity (Boyle et al., 2000; Vrugt et al., 2003). These developed from algorithms with a focus on local optima to global search algorithms which explore the complex parameter space for a global optimum (Yapo et al., 1996). One of the most helpful

algorithms for highly complex calibration tasks available at this point is the Shuffled Complex Evolution Metropolis Algorithm developed at the University of Arizona (SCEM-UA) introduced by Vrugt et al. (2003), an improvement of the Shuffled Complex Evolution Algorithm (SCE-UA) (Duan et al., 1993). As many objective functions have a specialized focus, e.g. on high or low flows, multi-objective calibration procedures have lately found increasing attention (Efstratiadis and Koutsoyiannis, 2010). By weighting several objective functions in the calibration procedure, it becomes possible to take multiple aspects of the observed system response into account. In this context a use of additional multiresponse data (Kuczera and Mroczkowski, 1998), e.g. hydrochemical information, and so called soft-data has become increasingly popular (Seibert and McDonnell, 2002). Soft-data are usually based on the experimentalists qualitative knowledge of a catchment and include discontinuous proxies that can help to achieve higher realism in the modelling procedure and the model structure (Seibert and McDonnell, 2002; Choi and Beven, 2007; Winsemius et al., 2009).

In spite of this progress, even the best automated calibration procedure cannot detect an erroneous model structure or uncertainties in input/output data (Beven, 1993, 2016). Furthermore, the higher the number of parameters in a model, the more they are prone to loose identifiability as their interactions increase and they tend to compensate for physical processes not represented in the model (Kirchner, 2006). This often leads to the problem that differing parameter sets produce model outputs of similar quality according to the objective function(s) (Wilby, 2005). The often cited study of Beven (2006) was one of the first to introduce this effect as the concept of equifinality. Many studies have dealt with the problem of finding the appropriate degree of sophistication of lumped parameter models to represent all necessary processes and retain parsimony at the same time (Perrin et al., 2001, 2003; Hartmann et al., 2012b, 2013b; Chang et al., 2017). For example, Hartmann et al. (2013b) based their model selection procedure on discharge and hydrochemical time series as well as different discharge conditions to include as much information as possible in the identification process. Chang et al. (2017) established a multi-model framework for finding a parsimonious model with good performance by stepwise reduction of complexity to describe a karst system in southwest China.

A a first step to check the plausibility of parameter sets is to test the model's predictive capability outside the conditions used for calibration. This is often done by split-sample tests, where parts of the available input/output time series is set aside for validation (Bennett et al., 2013). In the best case, both time series inherit a large degree of the systems' variability in order to see whether the model is capable of representing different system states (Hartmann et al., 2014a). New approaches, like the ones introduced by Boyle et al. (2000) and Choi and Beven (2007), have further developed this split sampling by including a multi-period and multi-criteria approach and therefore considering the fact that hydrological systems are highly variable through time and that different states might need

an individual evaluation.

In the face of equifinality, claiming one set of parameter values as 'the best' for a model appears unrealistic (Beven, 2006). For this reason, the analysis and presentation of modelling uncertainties has become increasingly popular in the recent years. Wagener and Kollat (2007) stress that the communication of these uncertainties becomes more important when models are used as a basis for water resources management to give decision-makers the possibility to take them into account for the development of risk-based approaches. Renard et al. (2010) name the following uncertainties as essential in hydrological modelling:

- Input data
- Output data
- Structural uncertainty
- Parameter uncertainty

In order to assess parameter uncertainty, commonly used approaches today are often based on Baysian statistics (e.g. Generalised Likelihood Uncertainty Estimation (GLUE) or the Monte Carlo Analysis Toolbox (MCAT) (Beven and Binley, 1992; Wagener and Kollat, 2007)). Representation of uncertainty is given by Monte Carlo (MC) procedures that sample a large number of parameter sets from a uniform prior distribution and evaluate them according to the chosen objective function (Beven and Freer, 2001). By applying a threshold on the objective function, acceptable parameter sets can then be determined and a regional sensitivity analysis can be conducted in order to check the identifiability of parameters (see Section 4.4.2) (Hornberger and Spear, 1981). Hartmann et al. (2017) used this idea to assess the information content of different additional data, e.g. flow states and observation types, by considering in how far the usage of such data would be helpful in the confinement of parameter ranges.

A major disadvantage of lumped parameter models surely is the missing spatial variability in the physical processes, particularly when used in karstic environments. Thus, Hartmann et al. (2013a) introduced the semi-distributed VarKarst model. Instead of taking account of spatial variability by using a cell grid (like distributed models) the model reverts to pareto functions which attribute parameter values to a set number of compartments (Figure 4) (Moore, 2007).

Consequently, it is capable of accounting for variability of soil and epikarst depths as well as fractions of concentrated and diffuse recharge to groundwater and epikarst and groundwater hydrodynamics (Hartmann et al., 2013a). The VarKarst model has been successfully parametrized to be used in different catchments in Europe (e.g. Spain, Austria, Switzerland, England) as well as the Middle East (Israel, Palastine) (Hartmann et al., 2013c; Brenner et al., 2016). Most of the study sites were located in catchments with a clearly identifiable duality in porosity. Brenner et al. (2016) lately used VarKarst to study

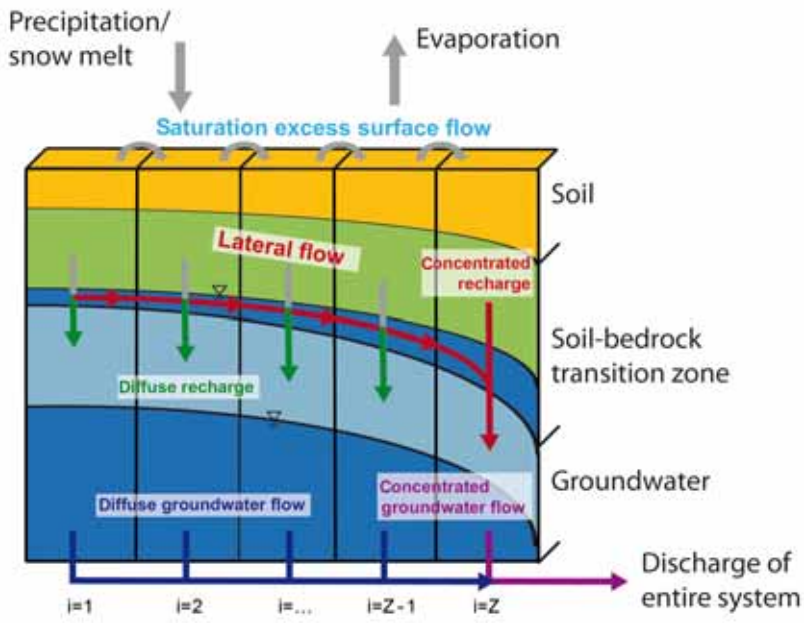


Figure 4: *Conceptual depiction of the VarKarst model. ´Retrieved from Hartmann et al. (2013a).*

the risk of groundwater flooding in a Chalk aquifer in England. They showed that the model structure is flexible enough to represent aquifers with a relatively high proportion of diffuse recharge through matrix and fissure porosity. VarKarst has also been adapted to incorporate local characteristics, e.g. by introducing a snow routine (Hartmann et al., 2012a) or hydrochemical information (Hartmann et al., 2013c) and by taking the variability of recharge areas into account (Hartmann et al., 2013a). Nevertheless, it has never been assessed before how well the VarKarst model works in (sub-)tropical karst systems and conditions as they are for example found in southwest China.

1.4 Climate change impact studies and uncertainty

Due to anthropogenic climate change, a global warming of air temperatures between 1.5°C and 2°C and changing precipitation patterns can be expected within this century (Stocker et al., 2013). Ren et al. (2012) report temperature increases of 0.08 ± 0.03°C during the period of 1906-2005 for the whole of China. Increases of temperature up to 2.5°C and precipitation declines of up to 20% until the end of the century can be expected in the country (Chen et al., 2011). In southwest China, a general tendency to increasing minimum daily temperatures has been reported by Lian et al. (2015). Furthermore, the region has generally become drier and is facing more extreme precipitation events resulting in increasing flood and drought frequencies, particularly during the summer monsoon period (Liu et al., 2014; Lian et al., 2015). This will significantly alter the water cycle in many regions and is a particular threat in karst dominated areas where surface water is

already scarce and many people depend on the water supply from groundwater aquifers (Lian et al., 2015; Huang et al., 2017). In 2010 the worst drought in 50 years caused large shortages of drinking water and losses in agricultural production (Zhang et al., 2012; Lian et al., 2015).

As mentioned before, future changes can only be anticipated with the help of models. Yet, these predictions inevitably come along with large uncertainties introduced by multiple steps of the modelling process. Kay et al. (2009) conducted a study which tried to differentiate the uncertainty sources in climate impact studies ('cascade of uncertainty', (p. 206, Henderson-Sellers (1993))) for flood frequencies in England. They considered the following error sources:

- Future greenhouse gas emissions
- Global climate model (GCM)
- Downscaling of the GCM
- Internal variability of the climate system
- Hydrological model structure
- Hydrological parameters

The first four aspects can be summarized as uncertainties from climate modelling, whereas the last two have to be assigned to the hydrological modelling procedure. It is evident that these uncertainties have to be communicated in order to make appropriate management decisions (Steinschneider et al., 2012). For instance, from a number of ensembles, scenarios with different probabilities can be provided to show the whole range of possible impacts but also give an assessment on how likely and under which conditions they can happen (Chen et al., 2011; Tian et al., 2015). Many studies show that awareness in the scientific community is rising towards the need of communication of modelling uncertainties in impact studies (Wilby, 2005; Wilby and Harris, 2006; Dobler et al., 2012; Steinschneider et al., 2012).

Different studies have focused on the attempt to assign uncertainties to the possible sources mentioned before. Poulin et al. (2011) compared parameter uncertainty to uncertainty resulting from model structure for future scenarios. Wilby and Harris (2006) even tried to implement a framework that can account for all uncertainties along the chain by using a weighting scheme for each step within a Monte Carlo set-up and Steinschneider et al. (2012) further improved this approach. Nevertheless, no study has tried to compare model uncertainties under current and future conditions driven by different climate scenarios before. This is important to find out under which climate scenarios uncertainties of hydrological model output differ largely enough from uncertainties under current conditions that predictions can be expected to be reliable. If uncertainties under a future climate scenario significantly overlap those under current climate, they cannot be trusted to formulate management plans on. Furthermore, none of the previous studies were conducted

in a karst setting with low data availability.

If a climate change impact study is to be conducted, several options for inferring future climate conditions for model input are available (Kay et al., 2009). The most commonly used method to obtain time series under future conditions is the delta approach (Kay et al., 2009; Chen et al., 2017). Anticipated absolute or percentage changes of temperature and precipitation, respectively, are simply applied to a measured time series and then used as a model input to obtain predictions under future conditions.

2 Research questions and objectives

In the face of expected climatic and land use changes in the Lingshui karst spring catchment, a thorough understanding of the karst system becomes increasingly necessary (Guo et al., 2015). To be able to make predictions on future states, the application of an appropriate modelling approach is reasonable (Hartmann et al., 2014a). The VarKarst model, developed by Hartmann et al. (2013a), has been tested in many different karst settings across Europe and the Middle East. However, no study has been conducted on its performance in subtropical karst settings before. Therefore, the aim of this study is to parametrize the VarKarst model for the Lingshui spring karst system. A thorough analysis of model performance, parameter identifiability, predictive capability and uncertainty will be carried out in order to evaluate whether the VarKarst model can appropriately reflect local conditions. This process will also be useful in order to identify driving processes of the Lingshui spring system. The adapted and parametrized model will then be used as a tool to predict future changes of water level under varying climate scenarios. Its reliability in a climate impact context will be tested by comparing model uncertainties under current conditions to those under future predictions. It is expected that model uncertainties of future predictions will only differ sufficiently from those under current conditions when extreme climate scenarios are applied.
The study of Guo et al. (2015) tried to rise awareness to threats of these highly important karst systems with regards to quantity and quality changes. The present work represents a first trial to catch up with these urgently needed scientific descriptions and starts a first formulation of the Linghsui spring system in terms of a model. Insights from this study will enhance the understanding of the Lingshui karst system and current knowledge about karst spring dynamics in the southwest China karst region. Furthermore, it will be a first step for advising local decision makers on management plans and point out necessary further data collection (Guo and Jiang, 2011).
The following research questions are going to be answered throughout this thesis:

- Does the hydrological VarKarst model work in the special hydrogeological karst setting of the Lingshui spring catchment in southwest China?
- How good is the parameter identifiability and how large is the model uncertainty when only a short time series is available for parameter calibration?
- Is it sensible to include the rating curve - needed to transfer modelled discharge output to water level - in the calibration procedure? Which impact will the inclusion have on parameter identifiability and model uncertainty?
- Which climatic changes would lead to uncertainties in future predictions that differ enough from those under current climate to be rated as meaningful or reliable?

3 Study Site

The Linghsui karst system and the Wuming basin in which it is embedded, have been subject to several field campaigns over the last decades during which basic information about the area (e.g. (hydro-)geological setting and land use) as well as the spring pool dynamics and quality were collected. The following descriptions are mainly based on studies published by Guo et al. (2015) and Guo et al. (2016).

3.1 Climate

To describe the general regional climate, daily data for precipitation as well as minimum and maximum temperatures for the whole available time series (1951-2016) were downloaded from the Global Historical Climatological Network (GHCN-D) dataset for Nanning airport weather station (59431, 22°37'48" N 108°13'1" E), which is located about 60 km away from the Lingshui spring pool (Figure 6) (NOAA, 2017). A climate diagram for the normal climate period (1981-2010) according to Walter and Lieth (1967) is displayed in Figure 5. Climate in the region is governed by the Southern tropical monsoon with an annual average precipitation of 1290 mm. In almost all summer months

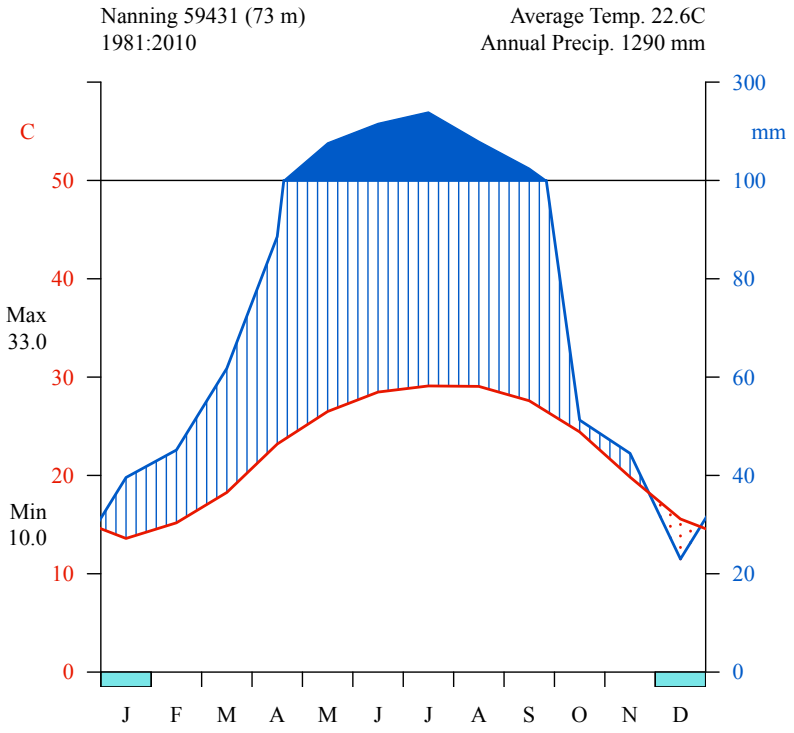


Figure 5: *Climate diagram for Nanning weather station (59431) over the normal weather period of 1981-2010. Monthly precipitation over 100 mm is scaled 15:1. Blue bars below the x-axis indicate months where frost can occur (absolute daily minimum temperatures below 0°C). The figure was created using the "iki-dataclim" package from the R software.*

between April and September, monthly precipitation sums exceed 100 mm. The maximum is reached between May and August where 63% of the yearly precipitation falls. Comparatively, low amounts of precipitation fall in the winter months between November and March (only 16.5% of the annual rainfall). In the climate diagram, December is even marked as a particularly dry month with arid conditions. The annual average temperature is 22.6°C. The highest temperatures occur over the summer months, particularly in July and August where the monthly mean lies above 25°C and daily maxima over 35°C can be reached. In January, on the contrary, the monthly mean temperature drops significantly below 15°C and daily minima can fall below 0°C. The combination of high temperatures and large amounts of precipitation lead to an average yearly potential evapotranspiration (PET) of 1287.4 mm (1957-2005) and a mean humidity of 78% (Guo et al., 2015).

3.2 (Hydro-)Geology

The Wuming basin, in which the Lingshui spring lies, includes all of Wuming County as well as parts of two other counties of Nanning city. Figure 6 shows details of the location of the study site[1]. Water from the basin feeds the largest river in the region, the Wuming river which is about 198 km long. Major contributaries are the Xijiang and Xiangshan rivers, emerging from surrounding non-karstic mountain ranges in the North and East. The basin contains 18 identified karst systems, which are divided by rivers and drained

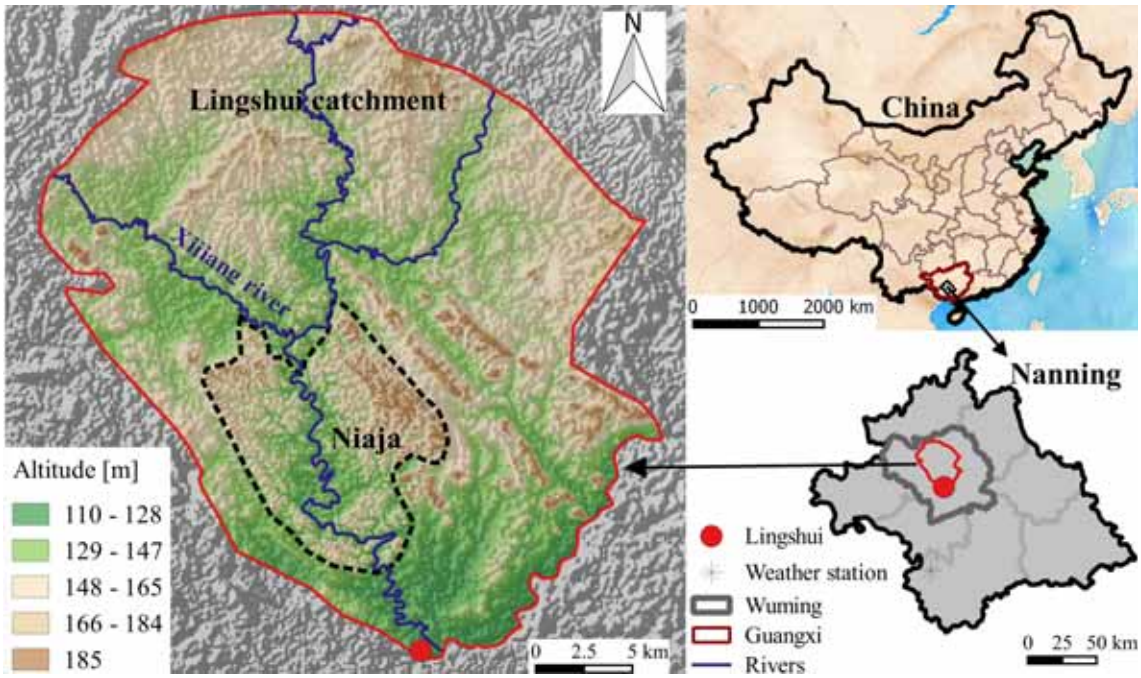


Figure 6: *Location of the Lingshui spring and its catchment.*

[1]DEM from USGS (2017); Data for Chinese administrative areas from GADM (2017); China Basemap from OpenStreetMap (2017); Polygons of Lingshui catchment and Niaja syncline digitized from maps in Guo et al. (2015) and Guo et al. (2016); River network from (Geofabrik, 2017)

by around 148 springs (personal communication, Fang Guo (22.11.2017)). The Lingshui spring system is the only one crossing a river by flowing underneath the river bed (Guo et al., 2015).

Along with the regional climate, the Wuming basin's (hydro-)geological setting is the main driving factor of the Lingshui karst system's hydrological dynamics (Guo et al., 2016). 70% of the area here are of karstic geology (Guo et al., 2015). Groundwater storage recharge is mainly autogenic by rainfall as only few obvious surface karst features like sinkholes can be found in the study area (Jiang and Guo, 2010). Therefore, it is assumed that the Lingshui spring has to come from a deep lying aquifer as a regional groundwater resource (Figure 7). Additionally, water is transferred to local karst aquifers by fractures in the bedrock and lateral recharge from rivers (Guo et al., 2015). During heavy rainfall events, the aquifer can become an estavelle which takes up water from the surrounding rivers (Guo et al., 2015). Figure 8 (adapted from Guo et al. (2016)) shows a geological map of the study region. The Lingshui spring lies in the downtown area of Wuming and forms the outlet of the Niaja syncline at an altitude of 96 m a.s.l.. The syncline is about 20 km long and exhibits the younger Middle Triassic geology of muddy limestones, which provide a mainly fissured karst aquifer (T_2). The Triassic to Upper Devonian carbonate depositions form peak forest at the edges of the syncline, the only bare karst that is found in the area (T_1b and T_1m). Siliceous and tuff aquitards, formed within Lower Triassic, Permian and Carboniferous strata, are present in the wings of the syncline (P_2). Otherwise, the hydrologically important covered karst occurs everywhere in the region where Quartenary (loam) clays overlay limestone ($Q_3{}^{al}$, $Q_4{}^{al}$). Covered karst aquifers are usually not as widely developed as bare karst aquifers and dominantly contain matrix flowpaths. According to Fischer et al. (2008), the dominating soil type in the region are acrisols, red soils typically found in the (sub-)tropics. In combination with overlaying soils of about 5 to 15 m (and in some parts up to 30 m), covered karst areas (and fissured aquifers of the Niaja syncline) build the groundwater aquifer feeding the Lingshui spring (Guo et al., 2015). Another proof for the dominance of matrix and fissure flowpaths in the system is the Xijiang river flowing through the Niaja syncline. In many highly fractured karst systems, precipitation is transported underground by sinkholes so

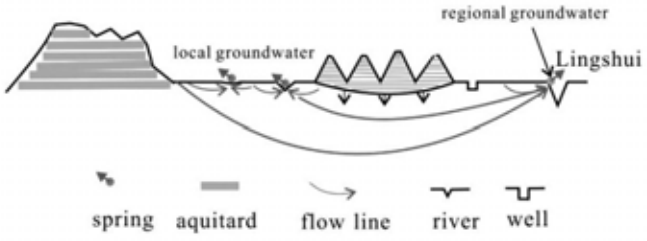


Figure 7: *Sketch of karst systems and springs in the Wuming basin. Retrieved from Guo et al. (2015).*

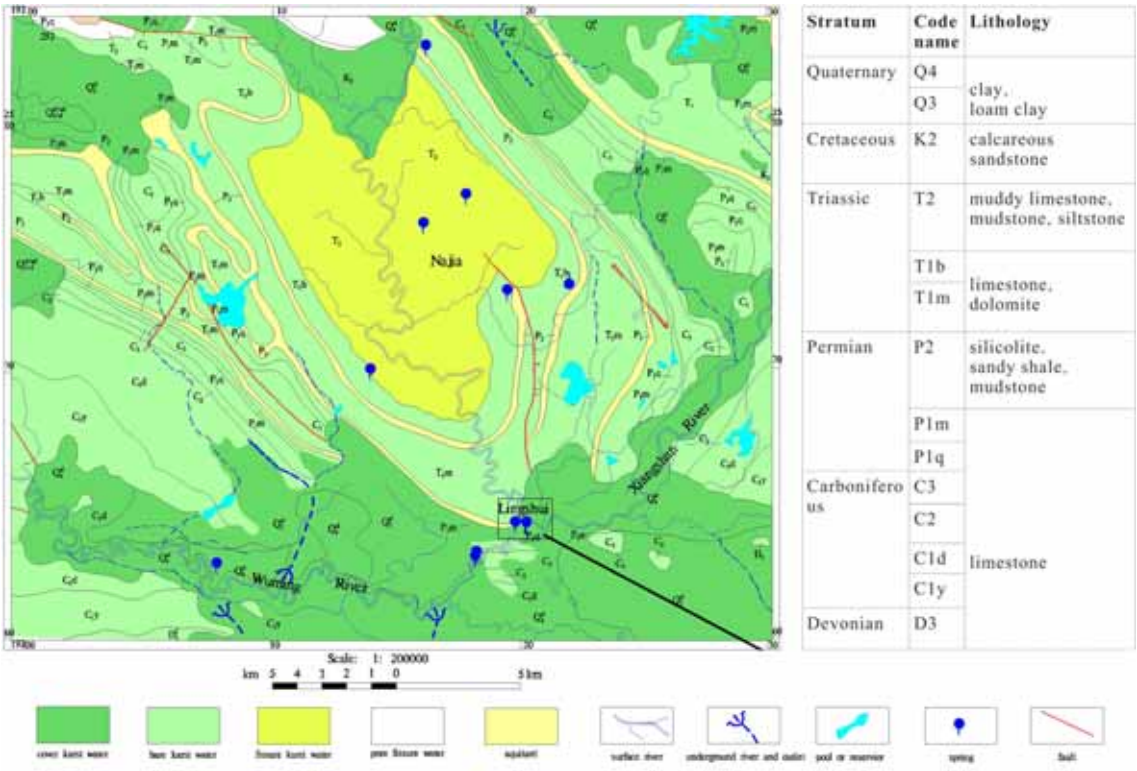


Figure 8: *(Hydro-)geological map of the study site. Adpated from Guo et al. (2016).*

fast, that surface runoff only occurs when all conduits are saturated with water. According to tracer tests, the covered karst in the area leads to comparatively long transit times of groundwater of about 17-23 $m \cdot d^{-1}$ (Guo et al., 2015). This could potentially result in mean residence times of up to 800 days over the whole Lingshui catchment. According to GIS-based[2] surface investigations as well as geophysical exploration, hydrochemistry, environmental isotope analysis and tracer tests, the catchment area of the Lingshui spring has been estimated to be 697 km^2 (Guo et al., 2015). The mean discharge of the Lingshui spring is 4300 $l \cdot s^{-1}$ (Guo et al., 2015). In their study on resilience of the Lingshui spring system to climate and land use changes, Guo et al. (2015) already found out that a general decline of discharge during the dry season of 50% took place between the 1970s and 2010s. Guo et al. (2016) have also investigated the qualitative changes of the Lingshui spring pool. Eight other springs from the karst system contribute to the pool's recharge so that it covers an area of 29300 m^2 (Figure 9). The water temperature of the spring pool of 24°C remains fairly stable over the year whereas the water level in the pool can range between 0.5 and 3 m, with peaks exceeding 3 m in the rainy season from May to August (Guo et al., 2016). Maxima of up to 7 m have been recorded after particularly strong rainfall events (personal communication, Fang Guo (04.09.2017)). An overflow dam was built near the outflow of the pool, 500 m before it drains to the Wuming river, to retain the scenic public area (Guo et al., 2016). In case of flooding, the river recharges the gorge

[2]Geographic Information Systems

with its polluted waters and changes the pools water quality in comparison to the springs (Guo et al., 2016).

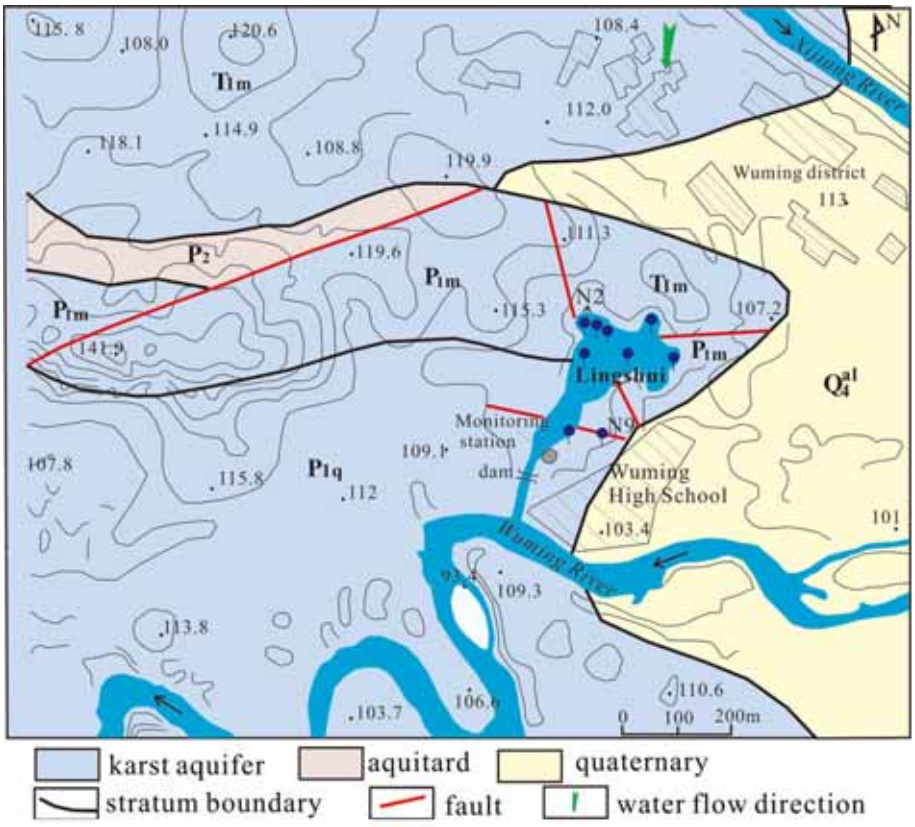


Figure 9: *Map of the Lingshui spring pool. Names of geologies are according to the nomenclature in Figure 8. Adapted from Guo and Jiang (2011).*

4 Methods

4.1 Data and quality assessment

The data on which this study is based have been kindly supplied by Fang Guo from the Institute of Karst Geology (CAGS) in Guilin. Only water level data were used in this study, as hydrochemical (Cl^-, $SO_4{}^{2-}$,HCO^{3-}, $NO_3{}^-$) and -physical parameters (electric conductivity and water temperature) were measured irregularly and were not available for both time series. Table 1 summarizes the available data. Regular measurements of the water level in the Lingshui spring pool were available for time periods of 01.01.2010 - 31.12.2011 (hourly, 2318 NAs (not availables)) and 18.04.2016 - 31.05.2017 (quarter-hourly, 151 NAs) measured by a Solinst water level meter. Figure 10 shows that the three years have very distinctive water level dynamics over the period 16.04 - 31.12. of a year. Where the duration curve of 2011 is marked by extensive lows (only about 5% of the data show values over 1 m), 2016 was much more variable with 50% over 0.75 m but also had much more variance in the low water level sector with the lowest values very close to 0. In 2010, the duration curve shows characteristics from both 2011 and 2016: It is more similar to 2016 for high water levels exceeding 1 m but is almost equal to 2011 for high exceedance probabilities. The maximum measured water level of all three time series reached 5.31 m on 22.04.2016 and the minimum occurred on the 05.12.2016 at 0.12 m. Yearly mean water levels were around 0.75 m.

Table 1: *Available data for the modelling study supplied by Fang Guo (Institute for Karst Geology, Guilin) and downloaded from GHCN-D. PET: Potential evapotranspiration.*

Parameter	Unit	Source	Time span	Res.
Water level	m	Fang Guo	01.01.2010 - 31.12.2011	hourly
			18.04.2016 - 31.05.2017	15 min
Precipitation	$mm \cdot d^{-1}$	GHCN-D	01.01.2010 - 31.12.2011	daily
		GHCN-D	26.08.2016 - 31.08.2016	daily
		GHCN-D	01.10.2016 - 16.10.2016	daily
		Fang Guo	18.04.2016 - 25.08.2016	30 min
			17.10.2016 - 31.05.2016	30 min
		Fang Guo	01.09.2016 - 30.09.2016	daily
PET	$mm \cdot d^{-1}$	Fang Guo	01.01.2010 - 31.12.2011	daily
Air temperature	°C	GHCN-D	01.01.2011 - 31.12.2011	daily
		GHCN-D	26.08.2016 - 31.08.2016	daily
		GHCN-D	01.10.2016 - 16.10.2016	daily
		Fang Guo	18.04.2016 - 25.08.2016	30 min
			17.10.2016 - 31.05.2016	30 min
		Fang Guo	01.09.2016 - 30.09.2016	daily

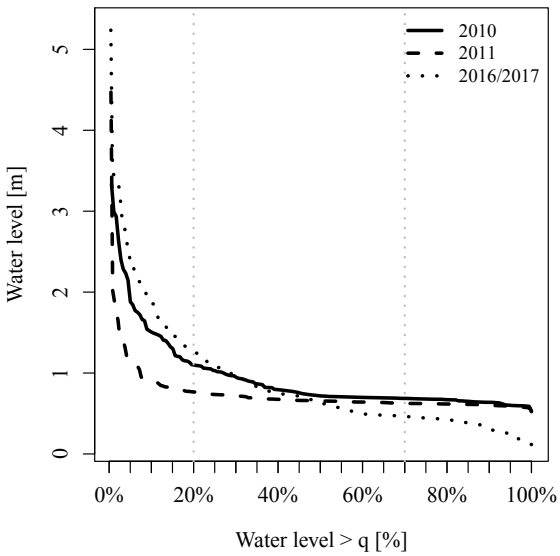


Figure 10: *Water level duration curve of 2010, 2011 and 2016/2017. The dotted grey lines differentiate very high (<20%) and low water levels (>70%)*

Discharge data for the Lingshui spring pool were obtained by a flow meter on a irregular basis (see appendix, Table A.1). Most data were measured in 2011 in an almost monthly interval. In 2010, 2016 and 2017 discharge was measured only occasionally. All measurements were conducted in average flow conditions so that extreme conditions (low/high flows) are not covered. Furthermore, information about the impact of backwater flow from the Wuming river during periods with very high flows is missing.

For the time period of 2010/2011 climate data from the GHCN-D weather station in Nanning, about 60 km away from the spring pool were the closest ones available. The data included daily means of temperature [°C] and daily sums of precipitation [mm]. Data for potential evapotranspiration [mm] for this time period have been provided by Fang Guo and were downloaded from the website of the China Meteorological Data Service Center (CMDC, 2017). For 2016/2017, climate data came from a HOBO weather station installed directly besides the Lingshui spring pool. Here, sums of precipitation and mean air temperature were measured in a half-hourly interval. For the HOBO time series, 2828 values are missing, mainly from 25.08.2016 to 17.10.2016. Data quality was ensured by checking for highly implausible values well out of expectable ranges. This occurred 6 times in the evapotranspiration data for 2010/2011 (values > 3000 $mm \cdot d^{-1}$) and 151 times for the water level data of 2016/2017 (values < 0 m). A thorough quality assessment of the rainfall and water level data turned out to be difficult due to the lack of further data from the region that could have served for plausibility checks.

As the Varkarst model conducts its calculations on daily time steps, both water level time series (2010/2011 and 2016/2017) and the HOBO climate data from 2016/2017 had to be aggregated. For water level, daily means were calculated from the (quarter-)hourly data, for precipitation daily sums and for temperature daily means. Before a day was declared to be NA, thresholds of a maximum allowed number of NAs per day were set. A maximum of 10% of the data missing was declared to be acceptable for a given time step. A thorough identification of NA-days like it has been conducted here is particularly important as input data for the model (precipitation and potential evapotranspiration) are expected to be continuous. Time gaps have to be filled by data from another station or interpolation methods. These additional inaccuracies should be avoided wherever possible, hence the

NA-threshold approach. After aggregation, a total of 112 days were missing of the water level data for 2010/2011 (about 15% of the whole time series) as well as 6 days of the evapotranspiration data. In 2016/2017 only 3 days were missing for water level and 63 days in the climate data (around 15%).

The short gaps in the 2010/2011 evapotranspiration time series were filled using the R na.spline-function. Gaps of precipitation and temperature in 2016 were filled by data provided by Fang Guo from a weather station in Nanning, 30 km away from the study site (personal communication, Fang Guo (25.09.2017)) and from the GHCN-D (see Table 1 and Figure 11). To keep this step simple, it was decided not to use a linear regression approach between the Lingshui weather station and the interpolation data provided from the aforementioned sources. Figure 11 shows the (a) temperature and (b) precipitation time series with gaps mainly occuring in the period of 25.08.2016 - 17.10.2016 and the data used for filling these gaps. As there were still NAs from the GHCN-D dataset for precipitation, they were simply set to zero. Data to fill the September gap were provided by Fang Guo from another weather station (light blue) whereas all other missing days were filled with data from GHCN-D.

In Figure 12 (a) monthly means of temperatures and (b) monthly sums of precipitation are depicted to compare the three years of climatic measurements (2010, 2011, and 2016/2017) and show whether any apparent differences did occur that could potentially have influenced water levels. In terms of their monthly temperatures, all three years were similar, even though 2011 had the coldest winter of all with a mean monthly temperature of under 10°C in January. On the mean, 2011 was also the driest year with 1252.9 mm of precipitation in spite of October being a particularly wet month with over 300 mm. This could be an explanation for the low water levels in 2011 shown in Figure 10. As can be identified in Figure 12 (b), the spring of 2010 was marked by a severe drought that occurred in the whole of southwest China which already started in September 2009 but was most severe between February and April 2010 (Zhang et al., 2012). The drought had negative effects on water availability in the area with impacts on agriculture and water supply (Zhang et al., 2012). August and October 2011 were marked by several high precipitation events that added up to a very high monthly sum. The most extreme precipitation event was also measured on 30.09.2011 with >100 mm of rain in one day. In spite of these variations, which are very characteristic for precipitation, the three time series can be regarded as fairly similar.

As the VarKarst model structure requires potential evapotranspiration data as an input, air temperature data from 2016/2017 had to be transformed using the Thornthwaite Equation (see Appendix B, Thornthwaite (1948)). To check for the reliability of evapotranspiration values calculated from air temperature by Thornthwaite, for 2010/2011 evapotranspiration data from Fang Guo were compared to those calculated by the Thornthwaite approach (Figure 13). It can be seen that lower evapotranspiration rates are highly underestimated

as the line of linear interpolation does significantly differ from the 1:1 line here. Higher evapotranspiration rates tend to be overestimated. This information has to be kept in mind for the interpretation of model results.

4.2 Rating curve calibration

To capture hydrodynamics of the Lingshui karst system, it is reasonable to use Lingshui spring pool for measurements, as it integrates the behaviour of all nine springs emerging

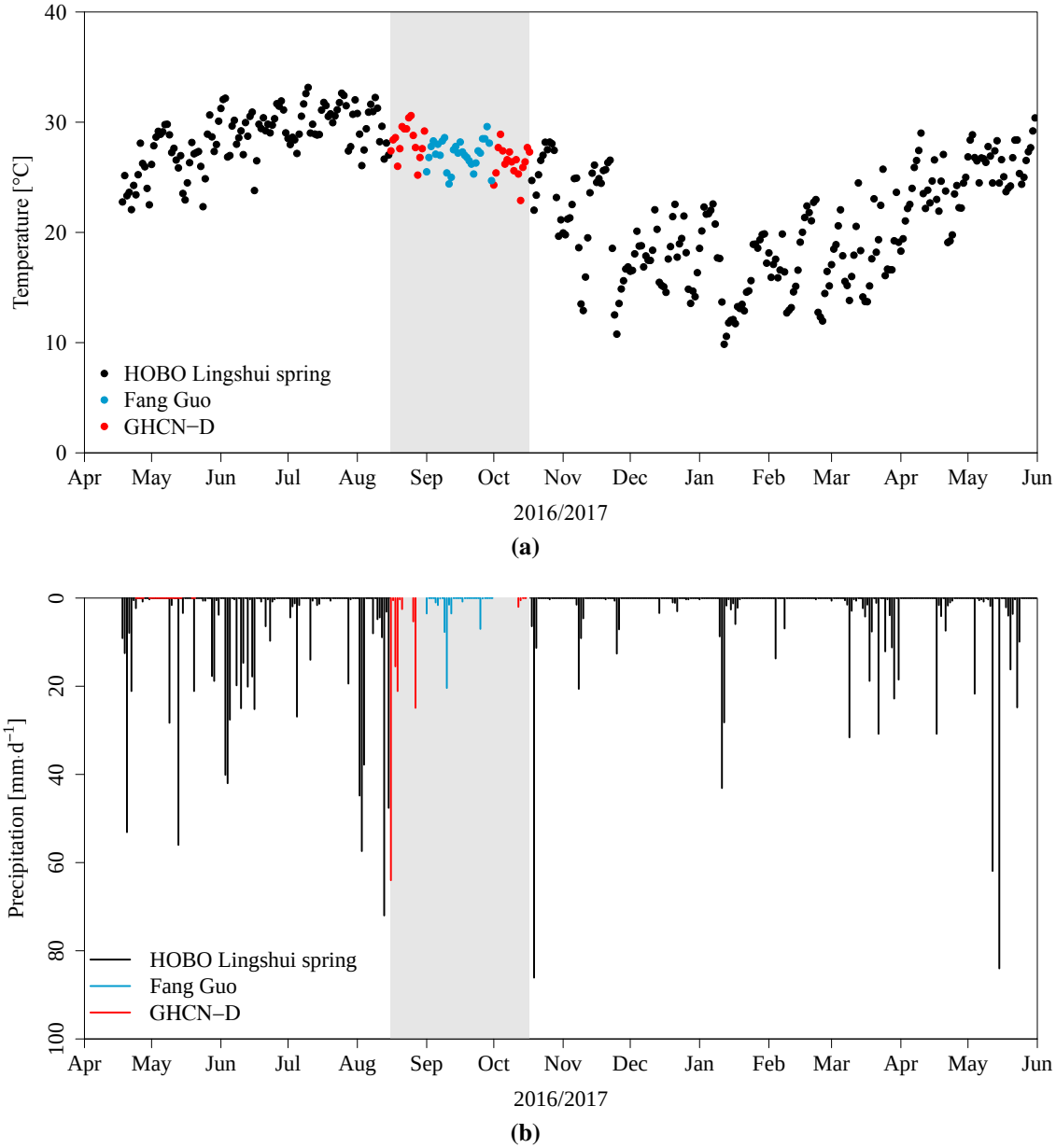


Figure 11: *Plot of values inserted into the gap of climate data (16.08.2016-17.10.2016) for (a) daily temperature and (b) precipitation sums. Interpolation values have been either provided by Fang Guo or were derived from GHCN-D (Station 59431) wherever still missing.*

from the system. Intensive discharge measurement campaigns, particularly in a large natural water body such as the Lingshui spring pool, are often difficult to realize. Water level, on the other hand, can be much easier recorded automatically, which is why a permanent monitoring station has been installed at the spring pool (Figure 9). As the VarKarst model simulates spring discharge dynamics, it is necessary to find a rating curve that transforms water level to discharge (or vice versa) to be able to compare simulations with measured data. The most commonly applied form of the rating curve is the power law in Equation 1 (Braca, 2008). It is particularly suitable in natural cross-sections without weirs where section and channel controls are unknown (Braca, 2008).

$$Q = c(h+a)^b \tag{1}$$

where: Q = Discharge [l·s^{-1}]
h = Water level [m]
a, b, c = Calibration coefficients [-]

The reverse can be accordingly calculated by converting Equation 1 into Equation 2.

$$h = (\frac{Q}{c})^{\frac{1}{b}} - a \tag{2}$$

21 discharge measurements were available for 2010, 2011 and 2016/2017 (11 measurements from 2011 alone) (Table A.1). Even though according to Braca (2008) this is a sufficiently large sample to formulate a rating curve, stationarity of the rating curve over time has to be assumed as conditions could have changed over the years, e.g. through

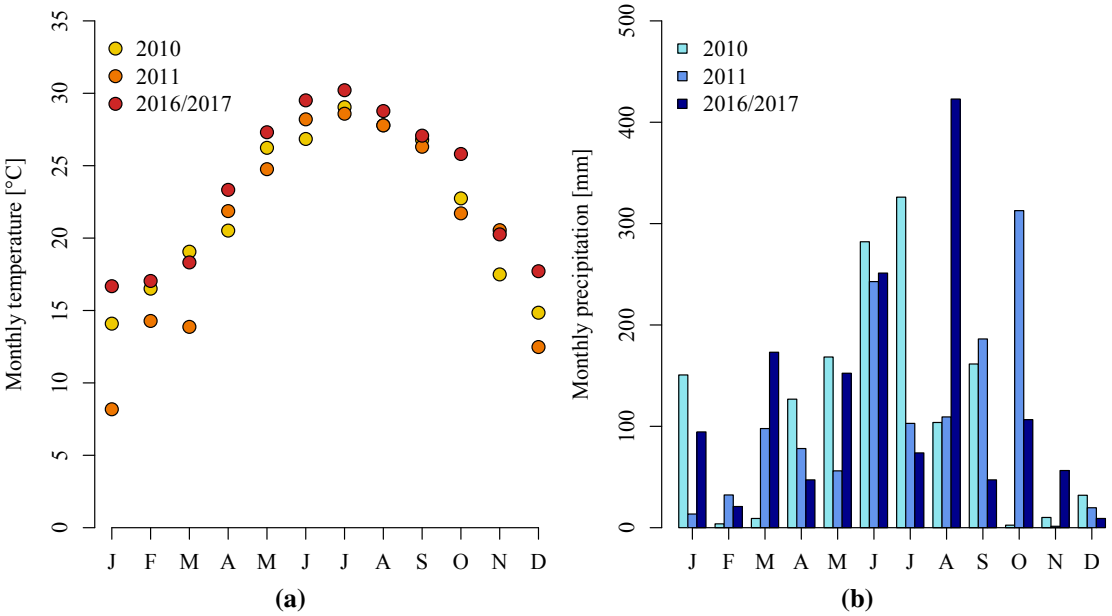


Figure 12: *Plots of monthly (a) mean temperature (b) sum of precipitation for 2010, 2011, 2016/2017 calculated after filling interpolation of NAs.*

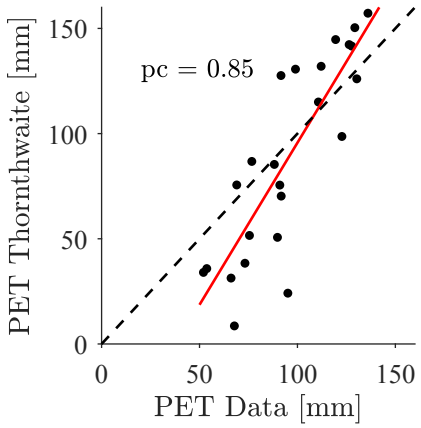


Figure 13: *Monthly sums of PET data provided by Fang Guo compared to monthly sums calculated with Thornthwaite. Solid blue line: least square linear fit of the data, dashed line: 1:1 line. Pearson's correlation coefficient = pc.*

sedimentation in the spring pool. The sample is not large enough to find rating curve formulations for both time periods separately.

Two approaches to include the discharge-water level transformation were considered in this study. One followed the classical approach of finding the rating curve formulation by fitting a linear model on log-scale with a least square approach (Braca, 2008). Based on the assumption that water level roughly follows discharge dynamics, the model calibration was conducted on water level dynamics and variability only, not taking water volume into account (see Section 4.4.1). Modelled discharge was then transformed into water level by using the fitted rating curve. This procedure will be referred to as RNI in the following (Rating curve Not Included) .

In a second approach it was tried to include the definition of appropriate rating curve parameters into the model calibration process. This approach will be called RI (Rating curve Included) hereafter. As this added another three parameters to be calibrated, it was particularly interesting to see whether uncertainties would increase with this approach due to equifinality (Beven, 2006).

4.3 The VarKarst model

In the following, the concepts behind the VarKarst model, as first introduced by Hartmann et al. (2013a), are going to be described. The VarKarst model was chosen for two reasons. First, for its low data input requirements in comparison to, for example, distributed models. Secondly, its semi-distributed character aims at taking the spatial variability of parameters driving the physical processes in a catchment into account. Even though the Lingshui spring catchment does not show the clear division between matrix and conduit porosity, VarKarst is expected to still be able to adapt to these conditions, as shown in Brenner et al. (2016). VarKarst incorporates variability of soil and epikarst depths,

fractions of concentrated and diffuse recharge to groundwater as well as epikarst and groundwater hydrodynamics. This flexibility is achieved by using one-parameter pareto functions as proposed and successfully tested in hydrological modelling by Moore (2007) and other applications of the VarKarst model. The pareto functions represent the heterogeneity of a karst catchment through model compartments without the data requirements of a distributed model, as their parameters can be identified by automated calibration. Due to the scarce availability of data for the Lingshui karst spring, a parsimonious model setup was chosen, which includes only basic characteristics of the karst system and does not incorporate chemical variables. By experience from former studies, the number N of model compartments was set to 15, e.g. Hartmann et al. (2013a), Hartmann et al. (2016) and Brenner et al. (2016).

Hereafter, the mathematical formulation of the VarKarst model is going to be presented, basically following the flow of water from precipitation over soil to epikarst, groundwater and discharge of the spring. This description is based on a previous one provided in Hartmann et al. (2013a).

The soil storage capacity of the compartments is determined by $V_{S,i}$ [mm] with Equation 3.

$$V_{S,i} = V_{max,S} \left(\frac{i}{N} \right)^{a_{SE}} \tag{3}$$

where: $V_{S,i}$ = Soil storage capacity [mm]
$V_{max,S}$ = Maximum soil storage capacity [mm]
N = Number of compartments
a_{SE} = Soil/epikarst variability constant [mm]

$V_{max,S}$ [mm] is approached by derivations in equation block 4 under the assumption that $V_{mean,S}$ [mm] represents the soil depth at compartment $i_{1/2}$ where the volumes on the left equal the volumes on the right.

$$\begin{aligned} \int_0^{i_{1/2}} V_{max,S} \left(\frac{x}{N} \right)^{a_{SE}} dx &= \frac{\int_0^N V_{max,S} \left(\frac{x}{N} \right)^{a_{SE}} dx}{2} \\ V_{mean,S} &= V_{max,S} \left(\frac{i_{1/2}}{N} \right)^{a_{SE}} \\ V_{max,S} &= V_{mean,S} \cdot 2^{\left(\frac{a_{SE}}{a_{SE}+1} \right)} \end{aligned} \tag{4}$$

where: $V_{mean,S}$ = Mean soil storage capacity [mm]

The epikarst storage capacity $V_{E,i}$ [mm] is found in the same manner with Equation 5 and

the derivations in Equation block 6.

$$V_{E,i} = V_{max,E} \left(\frac{i}{N} \right)^{a_{SE}} \tag{5}$$

where: $V_{E,i}$ = Epikarst storage capacity [mm]
$V_{max,E}$ = Maximum epikarst storage capacity [mm]

$$\begin{aligned} \int_0^{i_{1/2}} V_{max,E} \left(\frac{x}{N} \right)^{a_{SE}} dx &= \frac{\int_0^N V_{max,E} \left(\frac{x}{N} \right)^{a_{SE}} dx}{2} \\ V_{mean,E} &= V_{max,E} \left(\frac{i_{1/2}}{N} \right)^{a_{SE}} \\ V_{max,E} &= V_{mean,E} \cdot 2^{\left(\frac{a_{SE}}{a_{SE}+1} \right)} \end{aligned} \tag{6}$$

where: $V_{mean,E}$ = Mean epikarst storage capacity [mm]

The actual evapotranspiration $E_{act,i}$ [mm] from each compartment i in a time step t is determined by saturation of the soil and potential evapotranspiration E_{pot} [mm] (Equation 7). In comparison to all prior applications of the VarKarst model, a new runoff routine, introduced by Sarrazin et al. (2016) and soon to be published, was used in this study. It will be explained further below.

$$E_{act,i}(t) = E_{pot}(t) \frac{min[V_{Soil,i}(t) + P(t) + InSurfSoi_i(t-1), V_{S,i}]}{V_{S,i}} \tag{7}$$

where: $E_{act,i}(t)$ = Actual evapotranspiration [mm]
$E_{pot}(t)$ = Potential evapotranspiration [mm]
$V_{Soil,i}(t)$ = Actual amount of water in soil [mm]
$P(t)$ = Precipitation [mm]
$InSurfSoi_i(t-1)$ = Redistributed excess water from epikarst to soil [mm]

Wherever the input to a soil compartment exceeds its storage capacity $V_{S,i}$, recharge to the epikarst $R_{Epi,i}(t)$ [mm] is generated (Equation 8).

$$R_{Epi,i}(t) = max[V_{Soil,i}(t) + P(t) + InSurfSoi_i(t-1) - E_{act,i}(t) - V_{S,i}, 0] \tag{8}$$

where: $R_{Epi,i}(t)$ = Recharge to the epikarst [mm]

As in many karst areas, the epikarst can act as an important storage, outflow from the epikarst is controlled by variable storage coefficients for each compartment $K_{E,i}$ [d] defined

by Equation 9.

$$K_{E,i} = K_{max,E} \left(\frac{N - i + 1}{N} \right)^{a_{SE}} \tag{9}$$

where: $K_{E,i}$ = Variable epikarst storage coefficient [d]
$K_{max,E}$ = Maximum epikarst storage coefficient [d]

It is assumed, that a multiplication of the mean epikarst storage coefficient $K_{mean,E}$ [d] by the number of compartments N represents the area below the Pareto function whose variability is given by the distribution coefficient a_{SE}. $K_{max,E}$ can then be derived from $K_{mean,E}$ according to Equation block 10.

$$\begin{aligned} N \cdot K_{mean,E} &= \int_0^N K_{max,E} \left(\frac{x}{N} \right)^{a_{SE}} dx \\ K_{max,E} &= K_{mean,E}{}^{(a_{SE}+1)} \end{aligned} \tag{10}$$

where: $K_{mean,E}$ = Mean epikarst storage coefficient [d]

Flow from epikarst to groundwater can then be described by Equation 11 and for a time step t depends on the current volume of the epikarst storage, the recharge coming from the soil storage and the storage coefficient.

$$Q_{Epi,i}(t) = \frac{min[V_{Epi,i}(t) + R_{Epi,i}(t) + InSurfEpi_i(t-1), V_{E,i}]}{K_{E,i}} \cdot \Delta t \tag{11}$$

where: $Q_{Epi,i}(t)$ = Flow from epikarst to groundwater [mm]
$V_{Epi,i}(t)$ = Actual amount of water in epikarst [mm]
$InSurfEpi_i(t-1)$ = Redistributed excess water from epikarst to epikarst [mm]

The VarKarst model separates the slow diffuse recharge of groundwater (which varies between compartments i=1 to i=N-1) from concentrated recharge in conduits (compartment i=N). This variability is calculated by the separation factor $f_{c,i}$ [-] determined by a distribution coefficient a_{fsep} [-] (Equation 12).

$$f_{C,i} = \left(\frac{i}{N} \right)^{a_{fsep}} \tag{12}$$

where: $f_{C,i}$ = Variable separation factor [-]
a_{fsep} = Recharge separation variability constant [-]

Concentrated recharge $R_{conc,i}$ [mm] and diffuse recharge $R_{diff,i}$ [mm] are then calculated determined by Equation 13 and Equation 14.

$$R_{conc,i}(t) = f_{C,i} \cdot Q_{Epi,i}(t) \tag{13}$$

$$R_{diff,i}(t) = (1 - f_{C,i}) \cdot Q_{Epi,i}(t) \tag{14}$$

where: $R_{conc,i}(t)$ = Concentrated recharge [mm]
$R_{diff,i}(t)$ = Diffuse recharge [mm]

Groundwater contributions of the matrix system $Q_{GW,i}$ [mm] (compartments i=1...N-1) are controlled by variable groundwater storage coefficients $K_{GW,i}$ [d] (Equation 15).

$$K_{GW,i} = K_C \left(\frac{N - i + 1}{N} \right)^{-a_{GW}} \tag{15}$$

where: $K_{GW,i}$ = Variable groundwater storage coefficient [d]
K_C = Conduit storage coefficient [d]
a_{GW} = Groundwater variability constant [-]

and result in the discharge volume generated by each compartment (i=1...N-1) as described with Equation 16.

$$Q_{GW,i}(t) = \frac{V_{GW,i}(t) + R_{diff,i}(t)}{K_{GW,i}} \tag{16}$$

where: $Q_{GW,i}(t)$ = Groundwater contributions of the matrix [mm]
$V_{GW,i}(t)$ = Actual amount of water in matrix storage [mm]

The contribution of the conduit flow $Q_{GW,N}$ [mm] is calculated by Equation 17.

$$Q_{GW,N}(t) = \frac{min[V_{GW,N} + \sum_{i=1}^{N} R_{conc,i}(t), V_{crit,OF}]}{K_C} \tag{17}$$

where: $Q_{GW,N}(t)$ = Groundwater contributions of the conduit system [mm]
$V_{GW,N}(t)$ = Actual amount of water in conduit storage [mm]

A transformation of $Q_{GW,i}$ [mm] to spring discharge Q_{spring} [l·s^{-1}] is then achieved by taking the mean of $Q_{GW,i}$ over all compartments at each time step t and rescaling it by the recharge area A [km^2] (Equation 18).

$$Q_{spring}(t) = \frac{A}{N} \sum_{i=1}^{N} Q_{GW,i}(t) \tag{18}$$

where: $Q_{spring}(t)$ = Spring discharge [l·s^{-1}]
A = Recharge area [km^2]

As mentioned before, the runoff routine has been adapted by (Sarrazin et al., 2016). In

former versions of the model, water masses exceeding the storage capacity of the epikarst were only redistributed to the following compartment without taking its saturation into account and therefore causing errors in the mass balance. The new runoff routine incorporates the soil and epikarst storage capacity by redistributing saturation excess according to the saturation deficit of soil and epikarst of the following compartment. In practice, this means that all excess water from saturated compartments is summed up and redistributed to the next unsaturated compartment ($InSurfSoi_i$, $InSurfEpi_i$ in Equations 7, 8 and 11). If there is still excess water left once this compartment reaches saturation, again the next unsaturated compartment is filled up. This procedure is being continued until no excess water is left or the last compartment is reached. All water leftover after the saturation of the last compartment becomes surface runoff.

Based on this introduction to the VarKarst model, the next subsections will present the workflow used to parametrize the model. It largely follows steps based on Wagener et al. (2001). A mind map on the process is depicted in Figure 14.

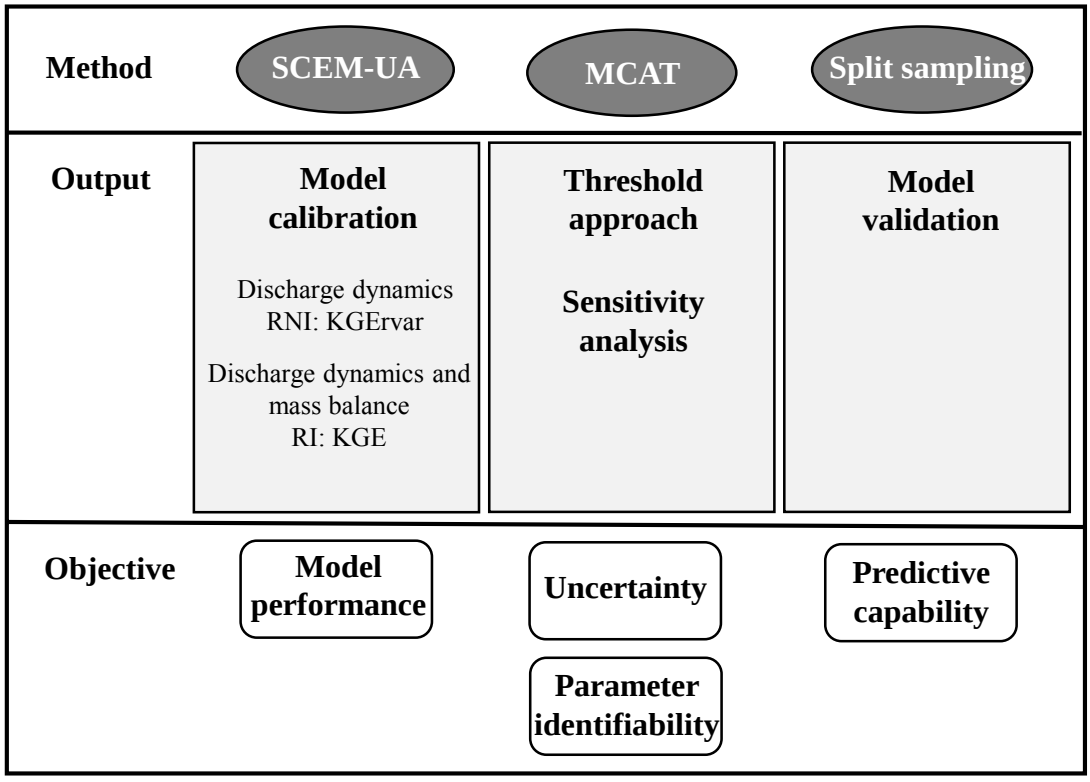


Figure 14: *Overview of the methodology used in this study.*

4.4 Model calibration and evaluation

4.4.1 Model performance

The general model performance was assessed by identification of an optimal parameter set with the SCEM-UA as introduced by Vrugt et al. (2003). This automated calibration procedure combines the strengths of the Metropolis algorithm (a Monte Carlo Markov Chain sampler), controlled random search, competitive evolution, and complex shuffling. It includes Monte Carlo Markov Chains (MCMC) (Metropolis-Hastings algorithm) to account for uncertainties in parameter calibration. In combination with the MCMC search, the SCEM-UA is used to constantly tune the prior distribution (usually chosen as a uniform distribution between the minimum and maximum expected range of a parameter) after each iteration. This gives the approach all the flexibility, efficiency and precision needed to ensure the exploration of the whole feasible parameter space and to find a realistic approximation of the stationary posterior distribution. Vrugt et al. (2003) see the algorithm as particularly useful in modelling studies with a high number of model parameters and therefore with high dimensional optimisation problems as is the case for the VarKarst model. Basic SCEM-UA settings were used in this study. The number of complexes was set to five and the number of samples to 1000. A maximum number of 3000 iterations was chosen to avoid an endless search for an optimal parameter set. Minimization of the objective function was chosen as the optimization method.
As this study uses a reverse modelling approach by fitting the model parameters to input and output variables, the choice of objective function on which the parameter calibration is evaluated is an important factor to obtain optimal modelling results. Gupta et al. (2009) have elaborated that the commonly used NSE consists of three elements with different contents of information which are often prone to interactions. Thus, they developed a decomposition of the NSE - the KGE - which separately accounts for components of the NSE in an orthogonal set up (Equation 19). This also gives the opportunity to exclude parts of the KGE if, depending on the research question, a focus on certain components is useful in model calibration.

$$\mathrm{KGE} = 1 - \sqrt{(r-1)^2 + (\alpha-1)^2 + (\beta-1)^2)} \quad \text{with } \alpha = \frac{\sigma_S}{\sigma_O} \text{ and } \beta = \frac{\mu_S}{\mu_O} \tag{19}$$

where:
r = Linear correlation coefficient (simulated and observed)
α = Measure of relative variability
β = Measure of volume bias (KGE)
μ_S, μ_O = Mean (simulated and observed)
σ_S, σ_O = Standard deviation (simulated and observed)

As the rating curve in the RNI approach would only be included after calibration, it was necessary to exclude the volume bias (β) from KGE and just calibrate modelled dis-

charge on water level dynamics and variability. This was done under the assumption that discharge dynamics of the spring pool could roughly be expected to linearly follow water level dynamics. Hartmann et al. (2013a) used the same approach in order to calibrate the VarKarst model separately on discharge dynamics and variability as well as variation of recharge area (KGE_{rvar}, Equation 20). Furthermore, the standard deviation α had to be normalized by the mean of the observed and modelled values in order to completely exclude the volume bias from the calculation of the KGE.

$$KGE_{rvar} = 1 - \sqrt{(r-1)^2 + (\alpha' - 1)^2)} \qquad \text{with } \alpha' = \frac{\sigma_S/\mu_S}{\sigma_O/\mu_O} \tag{20}$$

A basic version of the model was used to check model performance. The number of parameters was kept at its minimum of eight for RNI and eleven for RI, respectively, to reduce overparametrization and potential interactions between parameters. In a first run, the parametrization was conducted with parameter values and ranges by experience from another study using the VarKarst model in an English Chalk aquifer region (see appendix, Table C.1) (Brenner et al., 2016). They applied the VarKarst model in a wet climate and in a geological setting were diffuse recharge is dominant. The description of the study site of the Lingshui spring catchment allowed to conclude that a similar parametrization could be useful for the Lingshui spring system. Only parameter values of $V_{mean,S}$ and $V_{mean,E}$ were adapted to match information on local conditions with soil depths between 5 and 15 m and a poorly developed epikarst.

According to Mazzilli et al. (2012), the inclusion of a warm-up period is often highly recommended in order to allow storage compartments in a model to reach steady state

Table 2: *Description of parameters and parameter ranges used for final calibration trial (parameter ranges adapted from first calibration).*

Parameter	Description	Unit	Ranges	
			Lower	Upper
A	Recharge area	km^2	650	750
$V_{mean,S}$	Mean soil storage capacity	mm	5000	15000
$V_{mean,E}$	Mean epikarst storage capacity	mm	0	5000
a_{SE}	Soil/epikarst depth variability constant	-	0.1	100
$K_{mean,E}$	Epikarst mean storage coefficient	d	0.1	2.5
a_{fsep}	Recharge separation variability constant	-	0.1	100
K_C	Conduit storage coefficient	d	1	100
a_{GW}	Groundwater variability constant	-	1	100
a	Rating curve parameter	-	0	5
b	Rating curve parameter	-	0	5
c	Rating curve parameter	-	0	10000

and to reflect starting conditions at the beginning of the calibration period. They also mention that the warm-up period should ideally be extracted from a data set that was actually measured in the field. Due to the sparse availability of data, no information could be spared, so that a pseudo warm-up period had to be generated (referred to as pre-warm-up in Mazzilli et al. (2012)). For the warm-up period of the calibration time series (2010/2011), the 2010 part of the data were repeated three times. The calibration was then conducted on the 2010/2011 time series itself. For the validation period, the period between 18.04.2016 - 18.04.2017 was repeated three times for warm-up.

Several calibration trials were conducted in order to test whether the model was capable of producing realistic results at all. Any obvious erroneous behaviour, e.g. emptying of storage compartments against the assumption of stationarity (Nützmann and Moser, 2016) or large discrepancies between the model results and measured values, were analysed after this first step. Potential sources for unexpected behaviour (e.g. starting conditions, model structure, parameter ranges) were adapted. An appropriate parameter range confinement is highly important to ensure optimal sampling space for the MCAT (see Section 4.4.2) and to acquire a realistic impression of uncertainty. An insufficient confinement of the parameter space is inevitably linked to higher non-uniqueness of parameters and therefore equifinality (Beven, 2006). As some discharge measurements were available for the calibration period, different weighting schemes for measured discharge and water level with different combinations of KGE and KGE_{rvar} were tested, too.

Most former applications of the VarKarst model were conducted in the Mediterranean with a distinctive dry season with high evapotranspiration rates over the summer months. Therefore, it was concluded that the soil and epikarst would be completely dried out at the end of the dry season, so that their initial volume was set to zero (Hartmann et al., 2013a). In the tropics, on the other hand, with large amounts of rain falling during the monsoon season coinciding with the highest evapotranspiration rates, it can be expected that during the winter months upper parts of the aquifer retain their storage capacity. This has been proven by Hu et al. (2015) for a dolomite region in southwest China. Mean residence time in the well developed epikarst was over one year in this region. This is due to the missing duality of the system as is common in limestone aquifers. Guo et al. (2015) report, that the karst system of the Lingshui spring can be expected to be dominated by a fissure aquifer with few conduits and fractures. It therefore resembles the dolomite aquifer described by Hu et al. (2015) in its hydrogeological behaviour. The initial storage of the groundwater storage compartments was calculated by the linear relation between discharge and the storage of the groundwater reservoir (Equation 16). The first discharge value as obtained from the calculated rating curve was used as initial discharge. For the calibration period this was 3000 $l \cdot s^{-1}$ and for the validation period 1543 $l \cdot s^{-1}$.

Two plausible scenarios were tested for the starting conditions of the soil and epikarst storage in order to find out which would perform better in the calibration process. Con-

ducting such an additional sensitivity analysis was justified as a large impact of initial conditions on the model output could be expected due to the shortness of the available time series (Mazzilli et al., 2012).
The first scenario is mainly based on the study of Hu et al. (2015) which allows the assumption that by mid-winter (the time where the calibration time series starts), the soil compartment could have dried out completely. The epikarst on the other hand acts as an additional storage compartment and is still filled to its maximum. It will be called EpiMax (Epikarst Maximum) in the following.
The second scenario rather relies on actual hydrogeological conditions in the Lingshui catchment, where epikarst is only sparsely developed and soils with thicknesses of up to 30 m occur. Here, epikarst can be expected to only play a minor role, whereas soil might act as an additional storage compartment. Zhang et al. (2013) stress the importance of red soils as water storage in covered karst areas of south China. Both, soil and epikarst, were therefore set to maximum storage capacity for the start of the calibration period. This scenario will be called SoiMaxEpiMax (Soil Maximum, Epikarst Maximum) in the following. Of course, the soil could have dried out considerably by mid-winter but such nuances were not considered in the scope of this study.

In summary, the aim of the whole first calibration procedure by SCEM-UA was to identify final parameter ranges to be used in the model and to make a decision on the initial conditions to be used. The plausibility of parameter values and ranges was evaluated by taking existing publications and knowledge of the study site into account.

4.4.2 Uncertainty analysis and parameter identifiability

The short time series cannot contain the large temporal variability usually observed in the spring pool. Therefore, it will presumably introduce large uncertainties to the parameters and model output. Fundamental to the model evaluation approach that was used in this study was a Monte Carlo sampling of parameter sets to quantify uncertainty. This was done using MCAT, a sampling routine developed by Wagener and Kollat (2007), written in the programming language MATLAB®, which combines functionalities for Monte Carlo sampling and uncertainty analysis. As no assumptions about an error function could be made beforehand, the prior distribution was set uniform without updating, so that all parameter values got the same chance to be drawn in every run. It was decided that one million Monte Carlo runs were needed to evaluate parameter identifiability with more confidence and to have a higher chance to find parameter sets with satisfactory results.

The output of the MCAT allowed an interpretation from several perspectives. All of the further analyses were based on selected, ‘acceptable’ MC parameter sets which were identified by choosing only those which remained under a set threshold of the objective

function performance. For interpretation of parameter identifiability, a method similar to the regional sensitivity analysis of Hornberger, Spear and Young (HSY) (Hornberger and Spear, 1981) was used, which had been adapted for several similar studies before (Yapo et al., 1996; Choi and Beven, 2007; Moussu et al., 2011; Hartmann et al., 2013a; Brenner et al., 2016). In identifiability plots where the cumulative parameter distributions of the accepted parameter sets are plotted against the 1:1 line of a uniform distribution, parameters with a high identifiability should show an increase within certain value ranges. The narrowness of the cumulative density distribution can therefore be used to diagnose identifiability (Moussu et al., 2011).

Additionally, explanatory power of the water level based calibration was evaluated using an idea introduced by Hartmann et al. (2017). They calibrated the VarKarst model using different hydro-chemical data and information on flow states available for a spring in a study site in the southwest of Spain. Alternative simulation runs with the MCAT that produce large samples of parameter sets were then evaluated by different 'soft rules' (i.e. performance in calibration on discharge and the hydro-chemical parameters $SO_4{}^{2-}$ and $NO_3{}^{-}$). They then adopted the idea that when parameter ranges are normalized to values between 0 and 1, the confinement of distance between the 25^{th} and 75^{th} percentile due to the application of a threshold can give an indication on the informativeness of different flow states and hydro-chemical data for the specified karst system. As there was not as much additional information for the Lingshui spring available yet, the confinement of parameter ranges was simply evaluated by reducing the parameter set using the same threshold for KGE/KGE_{rvar} on water level.

In the last step, general uncertainty resulting from the modelling procedure was evaluated by using the restricted parameter sets, identified by the threshold procedure, for simulation. During the calibration time period, the distribution of model simulations could give a first impression on model performance under consideration of uncertainties. Ultimately, these should be diagnostic tools that help to learn about the underlying perceptual model in the case of the Lingshui karst spring - or at least show us where gaps in our knowledge are most severe and are most strongly affecting prediction uncertainty - and therefore help to guide field measurement campaigns (Wagener and Kollat, 2007).

4.4.3 Predictive capability

For validation of the parametrization of the model, it was decided to use a cross-validation approach, also called split sampling, where available data time series are divided in a calibration and validation period (Bennett et al., 2013). This was particularly important as the model is supposed to be used for further predictive analyses at the Lingshui study site in the future, e.g. climate change impact analyses, for which a high predictive capability

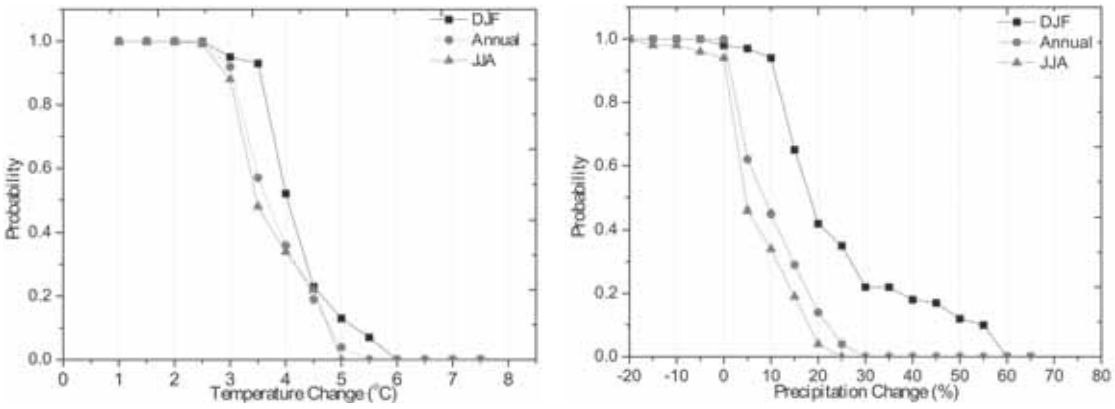
is required. For this purpose, it has to be determined whether the parametrized model is capable of making realistic predictions outside of the conditions of the calibration time series. 2010/2011 was chosen as calibration period and 2016/2017 for validation. One argument for this choice is that 2010/2011 were dry and normal years whereas 2016/2017, with 1455.3 mm of precipitation, were relatively wet in comparison. The model was therefore calibrated for a dry/normal period and tested in wet conditions.

4.5 Uncertainty under future climate scenarios

As the Lingshui spring system supplies about 100,000 people with drinking water, it is vital to estimate the impact of climate change on the quantity and seasonal dynamics of the spring. The question is, in how far a model calibration for which large uncertainties have to be expected due to the availability of only a short time series, is capable of making reliable predictions for future conditions. This becomes an even bigger issue considering the fact that most climate predictions are connected to significant uncertainties themselves.

To evaluate in how far uncertainties from the model under current conditions differ from those under future conditions using different possible scenarios, an approach following Chen et al. (2017) was used. They modified a baseline of observations under current conditions through climate change scenarios obtained from a perturbation of values extracted from probabilistic climate change modelling. By using this method, they accounted for the high uncertainty connected to climate change modelling that often even produces opposing results (Semadeni-Davies et al., 2008). The present study is based on results obtained by Chen et al. (2011). They calculated probabilistic projections for temperature and precipitation changes for China under the Special Report on Emissions Scenarios A1B (SRES A1B) using 28 coupled atmosphere–ocean general circulation models (AOGCMs) for the time periods 2011-2040 and 2070-2099. Expected temperature changes were presented as absolute changes whereas precipitation changes were reported as percentages of decrease/increase from the current scenario. The reference period for the changes was 1961-1990. Among the many climate projection studies for China and its Southwest, this study was particularly interesting for the purpose of the present one as it provides figures of probability distributions (Figure 15). Even though 2011-2040 represents the near-time period critical for policy and decision makers (Chen et al., 2011), according to Poulin et al. (2011) and Brenner et al. (2016), for 2070-2099, climatic changes from global warming are generally expected to exceed natural climate variability. Therefore, the latter time period was used for the climate projections of the present study.

The median as well as the confidence intervals (2.5% and 97.5%) were extracted from figures published in Chen et al. (2011), using the online digitizing tool WebPlotDigitizer (Rohatgi, 2017). Extracted values are listed in Table 3. A perturbation of three temperature scenarios with three precipitation scenarios led to nine possible climate change



(a) **(b)**

Figure 15: *Probabilistic winter (DJF, December, January, February), summer (JJA, June, July, August) and annual climate scenarios of (a) temperature and (b) precipitation change for the period of 2070-2099 under the A1B scenario for China. Adapted from Chen et al. (2011).*

scenarios that were applied to the validation time series by using the delta change approach (Kay et al., 2009; Chen et al., 2017). Depending on the climate scenario, the baseline of daily values of temperature and precipitation from the 2016/2017 time series were modified by the values in Table 3.

The method introduced in the following complies with the example of Kay et al. (2009) who compared the difference of uncertainties caused by natural variability under current and future conditions. For this study, it was adapted similar to the idea of Wilby (2005), who was interested in the impact of model uncertainties on future predictions. The concept behind this idea is best explained by the visualisation in Figure 16. Boxes 1) and 2) show duration curves of n model simulations under current (black) and future (coral pink) conditions. Note that the duration curves are only shown as lines for visual clarity. As an example, the sampling distribution (d) of the n model simulations is depicted as a small inset diagram for a random exceedance probability (dashed line). Such a sampling distribution can be calculated for each of the 409 exceedance probabilities from the validation time series. The shape of the sampling distribution represents the accumulation of modelled values at certain water levels for an exceedance probability (small inset

Table 3: *Median and confidence intervals (97.5% and 2.5%) for climate change scenarios of temperature change [°C] and precipitation change [%].Winter: DJF (December, January, February); summer: JJA (June, July, August). Digitized from Chen et al. (2011).*

Percentile	Temperature			Precipitation			Scenario
	Annual	DJF	JJA	Annual	DJF	JJA	
0.975	2.7	2.7	2.6	0.4	3.3	-8.2	H
0.5	3.7	4	3.5	8.6	18.3	4.6	M
0.025	5.2	5.8	4.9	26.9	58.7	21.9	L

diagram within boxes 1 and 2). It is basically a representation of uncertainty. Boxes 3) and 4) show the polygon which envelops all model simulations and therefore gives a more concise image of the overall model uncertainty. In box 5), the uncertainties are brought together to allow a comparison between current and future conditions. The aim of this analysis is to see in how far model uncertainties differ from another. If the difference is large enough, results from a climate prediction scenario can be perceived as significant. The novelty of this study is that it was decided to quantify this difference by a statistical approach, instead of making only qualitative statements. Welch's t-test appeared to be most appropriate for this purpose, as it is fairly indifferent to unequal variances of the samples and robust to non-normality in case of sufficiently large samples according to the central limit theorem (Welch, 1947). The number of accepted parameter sets from MCAT was thus set at 50, which, according to Ghasemi and Zahediasl (2012), is a sufficiently large sample to neglect normality tests. Welch's t-test evaluates the difference of two sample distributions by their means. In case of a significant difference, its p-value can give an impression on how strong the difference actually is. The overlap of the uncertainty bands gives an additional visualisation of the difference between uncertainties.

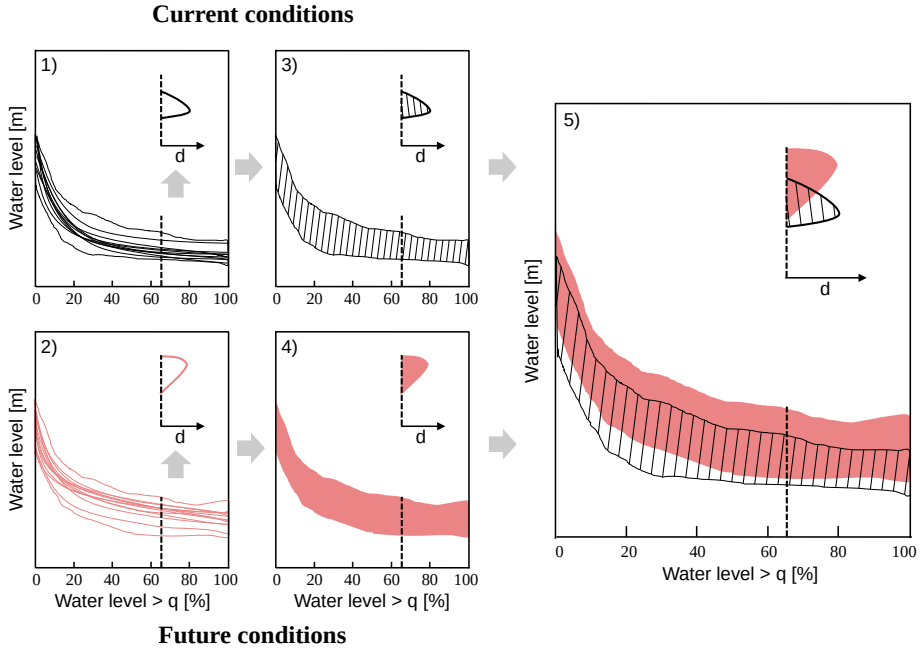


Figure 16: *Conceptualization of the evaluation method established for comparing the model uncertainties under current conditions (black) to those from scenarios under future conditions (coral pink). 1) and 2) represent the duration curve of n model simulations from which sampling distributions can be extracted. The small inset diagrams show the sampling distributions at an exemplary exceedance probability. 3) and 4) show the polygons which envelop the simulations runs. 5) shows the overlap of the uncertainty bands as well as the overlap of the sampling distributions. Welch's t-test was used to test the statistical difference between the sampling distributions.*

5 Results

5.1 Rating curve calibration

Parameters of the rating curve were obtained in two different ways: First, by a classical least square fitting and secondly by including the rating curve into the calibration procedure (RI). The results of these two methods are displayed in Figure 17. Black dots show the predicted discharge values calculated from measured water levels by the least square fitted rating curve formula. The predicted values lie well within the measured ones and the rating curve shows a satisfactory fit. In comparison, the rating curve with parameters identified by the SCEM-UA calibration procedure (Q modelled in Figure 17) lies noticeably under the measured values. A correction of the input water level by 0.5 m (see equation in Figure 17) moves the modelled curve upwards so that it fits with the measured data (Q modelled adjusted).

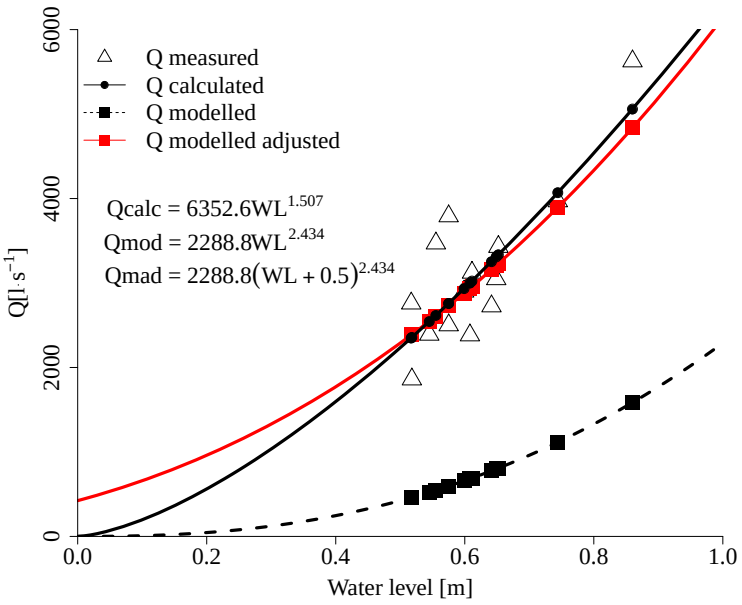


Figure 17: *Comparison of rating curves obtained from classical least square fitting on log-scale (Q calculated, Qcalc) and from including the rating curve into the model calibration procedure (Q modelled, Qmod). Measured water level - discharge data are marked by triangles. Calculated discharge data from measured water levels are marked by points and squares. The red line and points mark the transformation of Q modelled by raising the water level by a factor of 0.5 m (Qmad).*

5.2 Model calibration and evaluation

5.2.1 Model performance

In the following section, the results of the SCEM-UA calibration procedure are presented. As expected, the choice of parameter ranges had a large impact on the outcomes. Results of the first calibration trial with parameter ranges as described in Section 4.4.1 as well as those from the final calibration trial are demonstrated. The influence of the two chosen initial condition scenarios (EpiMax and SoiMaxEpiMax) was not as pronounced but still had a clear impact on the calibration results.

The most critical point in all calibration trials was the stationarity of groundwater compartments. Figures showing the development of groundwater storage in each compartment over the whole modelling time (warm-up + calibration) for all settings (RNI, RI, first and final parameter ranges as well as EpiMax and SoiMaxEpiMax) are presented in appendices D and E. Groundwater stationarity was mostly driven by the parameter ranges whereas a change of initial conditions evoked only small effect, if any at all. Nevertheless, it was decided to use SoiMaxEpiMax in the following analyses for mainly three reasons: Under the final calibration, compartment 14 of RNI still ran completely empty with EpiMax initial conditions (Figure D.3). Furthermore, it could also be observed that some soil compartments showed a tendency to fill up. Additionally, RNI and RI obtained a better KGE_{rvar} and KGE in the final calibration with the initial conditions of SoiMaxEpiMax compared to EpiMax. Thus, hereafter, only results of calibration runs with initial conditions set to SoiMaxEpiMax will be considered. Results from EpiMax calibration runs can be found in Appendices F and G.

Generally, it can be stated that no matter which setting was used, KGE_{rvar} and KGE for RNI and RI respectively had fairly high values (Table 4). All were above 0.7 except for the RNI run under the first trial parameter ranges, which only reached a KGE_{rvar} of 0.47. Nevertheless, other relevant aspects of model performance evaluation such as groundwater stationarity and parameter values differed largely among the calibrations and were therefore the basis on which a final decision on appropriate parameter ranges was based.

For the first calibration run, with parameter ranges mostly chosen according to Brenner et al. (2016), some parameter values identified by SCEM-UA did reach limits of the parameter ranges, indicating that wider ranges might be necessary in order to find optimal values (Table 4). This was particularly the case for a_{SE}, a_{GW} and a_{fsep} in the RNI modelling setup. Furthermore, a look at the dynamic part of the groundwater compartments showed non-stationarity (Figure D.2 and E.2). Generally, the following dynamics could be observed: The first compartments steadily ran empty over time whereas the middle ones constantly filled up. Only the last compartments ($\approx$ 10-15) showed recurring dynamics.

Table 4: *Results for parameters and objective functions for different calibration steps and rating curve approaches for SoiMaxEpiMax*

Parameter	Unit	RNI		RI	
		Calibration step			
		1st	Final	1st	Final
A	km^2	652.06	668.3	731.63	687.81
$V_{mean,S}$	mm	5011.7	5266.82	5317.97	12393.3
$V_{mean,E}$	mm	2044.84	4935.64	19.55	4665.66
$K_{mean,E}$	d	2.5	1.91	0.98	2.16
K_C	d	5.91	4.41	2.3	3.79
a_{SE}	-	5.98	40.13	5.22	52.61
a_{fsep}	-	4.87	25.49	0.89	29.64
a_{GW}	-	8.45	61.74	8.07	73.92
a	-	-	-	7599.13	2288.76
b	-	-	-	3.01	2.43
c	-	-	-	0.03	1.43
KGE$_{rvar}$	-	0.47	0.77	-	-
KGE	-	-	-	0.77	0.79

Based on these observations, parameter ranges of a_{SE}, a_{GW} and a_{fsep} were constantly widened in further calibration steps. This had to be done for both RNI and RI in order to keep up comparability. Only the final calibration step - presented in Table 4 - resulted in the intended stationarity for RNI and RI, particularly in the first compartments. Under the final calibration setting they remain completely stationary over time (Figure D.4 and E.4). All other compartments (except for 14 and 15) still show a clear tendency to filling up over time. Even so, referring to the explanations given in Section 6, this calibration was considered the best compromise possible.

The variable parameters clarify the difference in calibration results from the first to the final calibration of SoiMaxEpiMax (Figure 18 (a) and (b), respectively). The most apparent dissimilarity between the two calibrations is the extreme difference between the compartments. This effect can be noticed for all parameters, particularly for K_{GW}, the groundwater storage coefficient representing the mean residence time of the water in each compartment. The representation on log-scale was chosen to enable a visual differentiation between the first and the other compartments. Values for the first compartment reached up to $1.8 \cdot 10^{73}$ days for RNI and $3.3 \cdot 10^{87}$ days for RI. The number of days in groundwater then decreases rapidly, even though this decrease levels off towards the last compartments. In compartment 15, the mean residence time is at only 4.4 days (RNI) and 3.8 days (RI). These mean residence times result in a mean thickness of groundwater compartments over the calibration period between $7 \cdot 10^{66}$ km and 24.1 mm (RNI) (between $1.3 \cdot 10^{81}$ km and 21.2 mm for RI respectively). Even if the stark contrast between the compartments becomes most obvious in K_{GW}, the same behaviour can be noticed for the

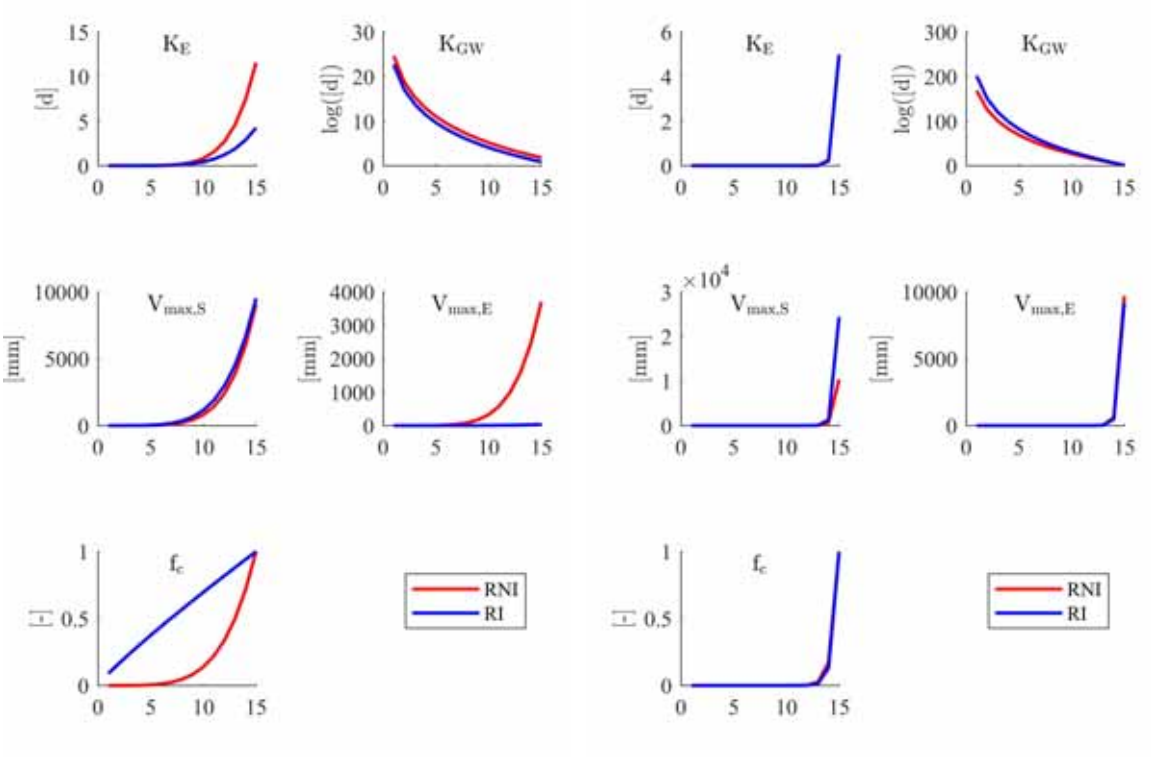


Figure 18: *Distribution of variable parameters resulting from RNI and RI for (a) the first calibration run and (b) the final calibration run.*

other parameters, too. For example, compartments 1-14 show minimal mean residence times of under a day in the epikarst (parameter K_E) in comparison to compartment 15 (about 5 days). The extremely low values of f_c in compartments 1-14, suggest a significant reduction of groundwater recharge through conduits in the model. Another noticeable outcome from the final calibration is that there is only little difference left between the distribution of variable parameters of RNI and RI. Knowing that such unrealistic values as observed for the first compartments of K_{GW} can occur in lumped parameter models (Gallagher and Doherty, 2007), modelling was still continued, particularly as an uncertainty analysis was included as part of the modelling process. Possible reasons for such missing realism of parameters will be given in Section 6.

Figure 19 shows the comparison of results for (a) modelled discharge of RNI and (b) modelled water level of RI versus observed water level over time. By visual evaluation, the discharge dynamics modelled by RNI with KGE_{rvar} follow the water level measurements well. All major peaks in water level are traced by the discharge dynamics, even though the discharge peaks are often not as pronounced as the measured water level peaks. Before the largest water level peak in autumn 2011, two small discharge peaks are modelled where no water level peaks were actually measured. Generally, the modelled discharge values (mean of 16599.4 $\text{l}\cdot\text{s}^{-1}$) highly exceed the mean of measured discharge during the

calibration period (3319.5 $l \cdot s^{-1}$).

RI is equally capable of reproducing the water level dynamics of the calibration period even though the modelled water level values almost always miss the complete height of the measured water level (Figure 19). Nevertheless, water levels modelled by RI were much more pronounced than those of RNI and could therefore capture the dynamics slightly better. The same periods in early summer 2010 and 2011 that caused trouble for RNI are missed by RI, too. This model setup also produces small peaks in late summer 2011 where none were actually measured. Recession periods are captured well, though.

5.2.2 Uncertainty analysis and parameter identifiability

In this section, the output of the MCAT with RNI and RI and the final parameter ranges are going to be presented. Initial conditions were set to SoiMaxEpiMax. The analysis of identifiability showed that with the high number of samples (1 Mio.) a clear threshold under which parameters sets could be considered 'acceptable' as in other studies (e.g. Hartmann et al. (2017)) was hard to choose. Therefore, a number of 50 acceptable parameter sets

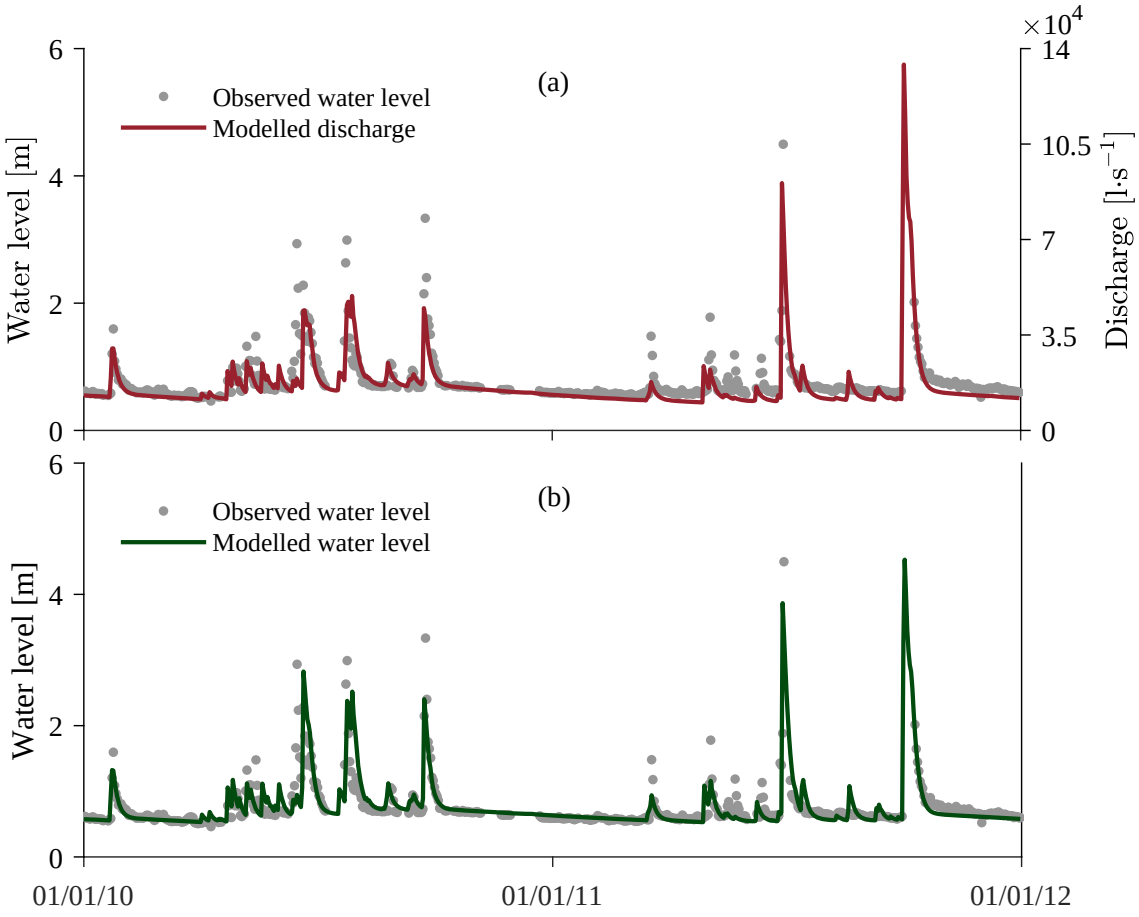


Figure 19: *Comparison of modelled and observed values over the calibration time period for the final SCEM calibration of (a) RNI and (b) RI.*

was selected instead (equalling 0.005% of the whole sample) under the assumption that they would adequately represent the modelling uncertainty and that the subsample would be sufficiently large for the Welch's test used in Section 5.3. For RNI, the threshold of 50 behaviourals was a KGE_{rvar} of 0.71 and for RI a KGE of 0.66 (Table 5, minima of calibration).

The identifiability plots for RNI in Figure 20 show that about half of the parameters are at least moderately identifiable (a_{SE}, $V_{mean,E}$, a_{fsep}, a_{GW} and K_C). For a_{SE}, a_{GW} and K_C, the cumulative distribution is particularly narrow, which shows that the values for these parameters are very similar among the acceptable parameter sets. Even though $V_{mean,E}$ and a_{fsep} also appear to clearly differ from the uniform distribution (black reference line), their cumulative distributions are more widespread. The reduction between the 25th and 75th percentile (boxplots in Figure 21 (a)), which is caused by the application of a threshold to water-level-based calibration results, confirms these observations. For a_{SE}, the reduction is most efficient with 81%. Clear decreases in distance between the percentiles are also noticeable for $V_{mean,E}$ and a_{GW}, leading to a reduction to 22.6% and 32% respectively. Contrary to the high identifiability indicated for K_C in Figure 20, the reduction between the percentiles is almost not noticeable. The comparatively high reduction for $V_{mean,S}$ of 25%, on the contrary, was not indicated as clearly by the identifiability plots. KGE_{rvar} excludes the volume bias and therefore the catchment area A from the objective function which is the reason for the missing identifiability of A.

Table 5: *Statistics of uncertainty analysis by model efficiencies KGE_{rvar} (RNI) and KGE (RI) for accepted parameter sets under the chosen threshold (0.005% = 50 accepted parameter sets). Cal: calibration, val: validation, min: minimum, std: standard deviation and max: maximum.*

	RNI (KGE_{rvar})		RI (KGE)	
	cal	val	cal	val
min	0.71	0.31	0.66	0.35
mean	0.72	0.4	0.67	0.45
std	0.01	0.05	0.01	0.05
max	0.74	0.58	0.72	0.61

The results for identifiability from the RI modelling strongly differ from those described for RNI. The blue lines in Figure 20 show that most parameters are not clearly identifiable. Only the cumulative distribution of K_C is as narrow as in RNI. Deviating from uniformity, a_{SE} lies at around 40, even though the tendency is by far not as clear as for RNI. Furthermore, the rating curve parameters a and b have clear tendencies towards certain parameter values. Again, the boxplots in Figure 21 (b) can serve to confirm these observations. The reduction here is highest with 56% for the rating curve parameter a. a_{SE}, K_C and b are reduced by 24%, 32% and 28%, respectively. The distance between the percentiles for c, on the contrary, was increased by 30% to 0.65. All in all it can be said that identifiability is clearly higher for parameters with RNI compared to RI, particularly for the original model parameters.

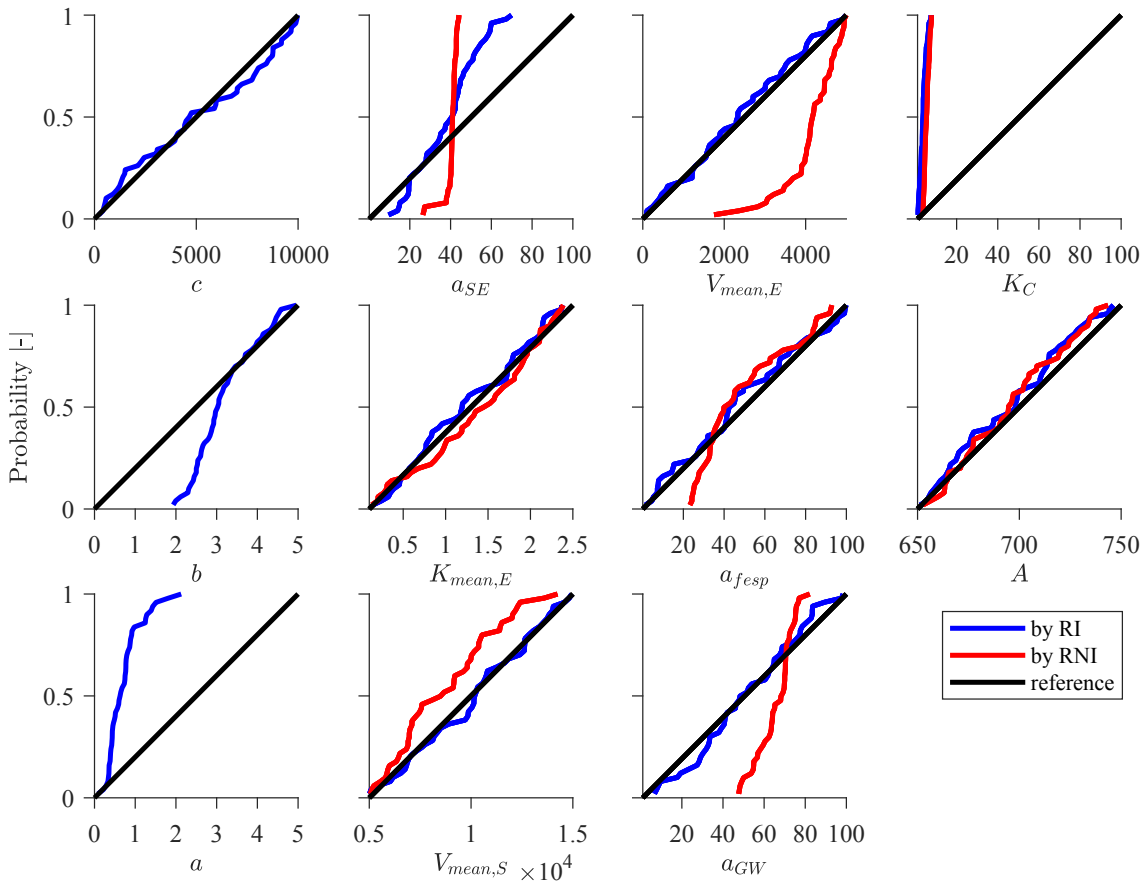


Figure 20: *Identifiability plots for RNI (red line) and RI (blue line) showing likelihood for parameter values for acceptable parameter sets compared to uniformity (black line).*

After identifying acceptable parameter sets from MCAT, these were now used for producing water level time series. For RNI, a transformation of the discharge model results in water levels using the least square fitted rating curve from Section 5.1 was necessary to allow a direct comparison to measured water level results. This did at first lead to modelled water levels shifted upwards (see appendix, Figure H.1), compared to observed water levels,which was overcome by a correction of 1.1 m to level the mean of modelled water levels (1.85 m) to the observed ones (0.75 m). Figure 22 shows the final modelling results for the 50 selected parameter sets over the calibration time period for (a) RNI and (b) RI in comparison to measured water level.

For RNI, the measured water level almost always lies within the uncertainty bounds of the simulations. Only in early summer 2011 do measured water levels exceed the modelled ones. In late summer 2011, a small peak was modelled as a result from a relatively large precipitation event which was not represented in measured water level data. A very efficient way to evaluate uncertainty and how well it captures measured values is the look at the duration curve of water levels (Figure 23). The generated output displays a slightly

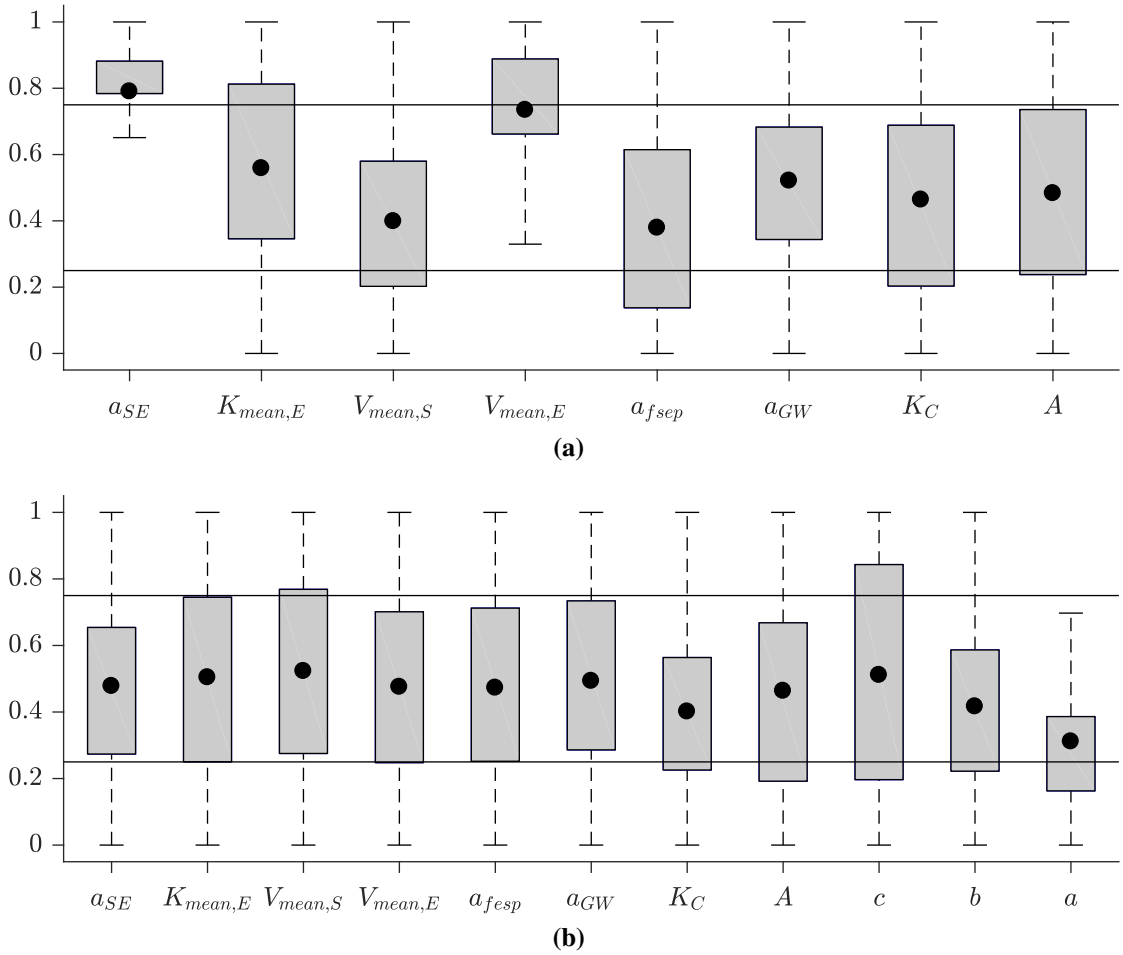


Figure 21: *Distribution of model parameters (normalised by their ranges) for acceptable parameter sets of (a) RNI and (b) RI. The black dot in the boxplots marks the mean. The two black lines identify the 25th and 75th percentiles.*

transparent black line for each simulation curve, so that areas with higher accumulation of curves appear to be black while single curves appear grey. For RNI, the measured duration curve overlaps with the dark areas of high density of the simulations. Only single parameter sets seem to be the cause for the width of the uncertainty band (Figure 23 (a)). Between an exceedance probability of 5% and 40% the uncertainty is largest and between 5% and 10% the modelled values tend to mostly overestimate measured data by almost 2 m (higher density of curves). Nevertheless, some simulations are still able to depict the actual dynamics, as the measured data still lie within the grey uncertainty band. As there is a gap in water level measurements during the highest precipitation event in October 2011 (Figure 22 (a)), modelled water levels in the very high flows do not have equivalent measured water levels and reach values of over 7 m.

For RI, the picture is quite different. As indicated by the low identifiability of parameters for the accepted parameter sets, the progression of the simulated time series is much more variable between different model runs (Figure 22 (b)). In the base flow period in winter 2010/2011 some model runs produce peaks whereas others clearly follow the

recession of the measured water levels. Modelled water level peaks without measured equivalent also occur in the recession of later summer 2011 and are more pronounced than for RNI. In comparison to RNI, RI simulations are noticeably less dispersed which indicates a generally reduced uncertainty. Furthermore, water level peaks in 2011 are simulated more adequately by RI, particularly for the events in 2011. The highest water

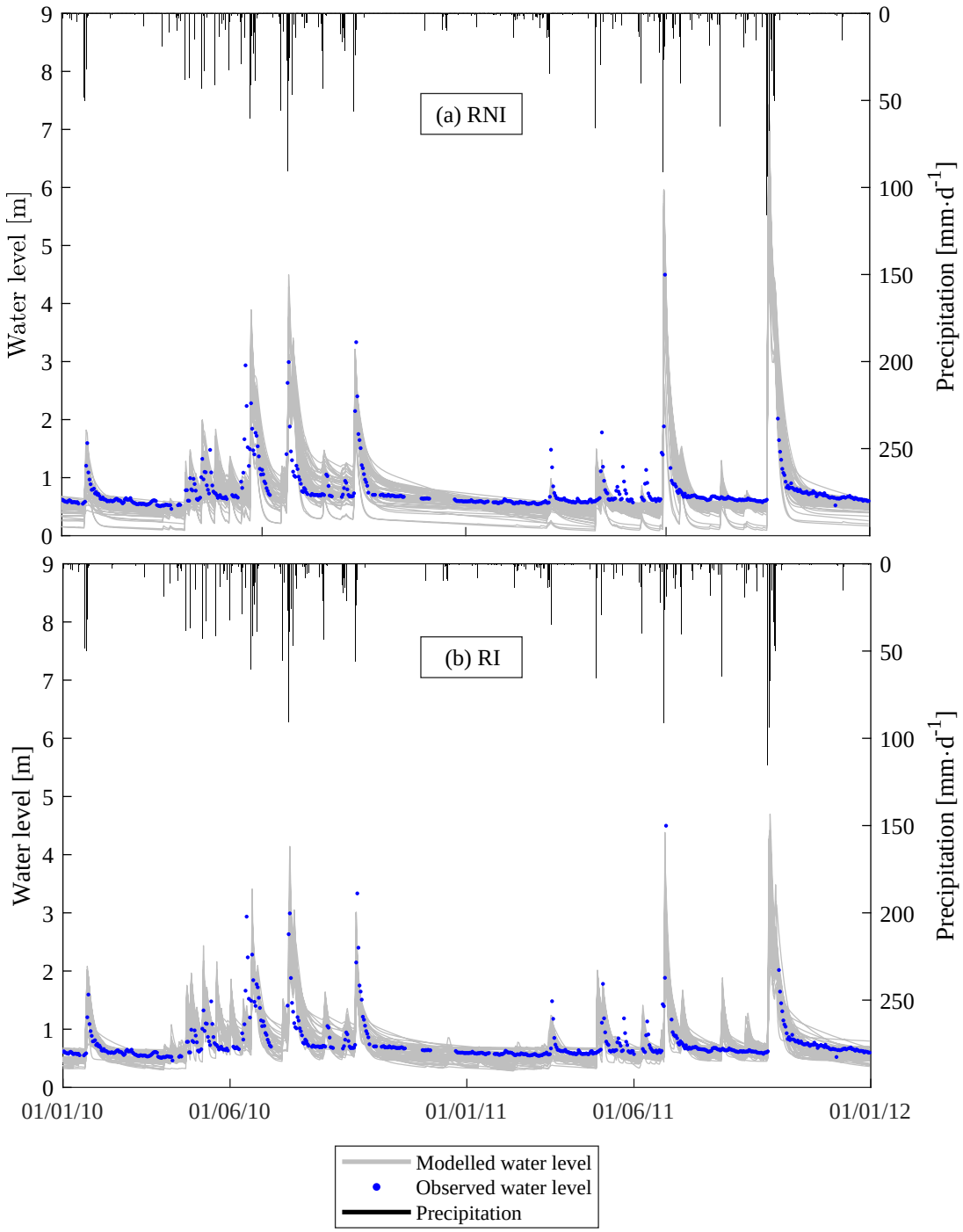
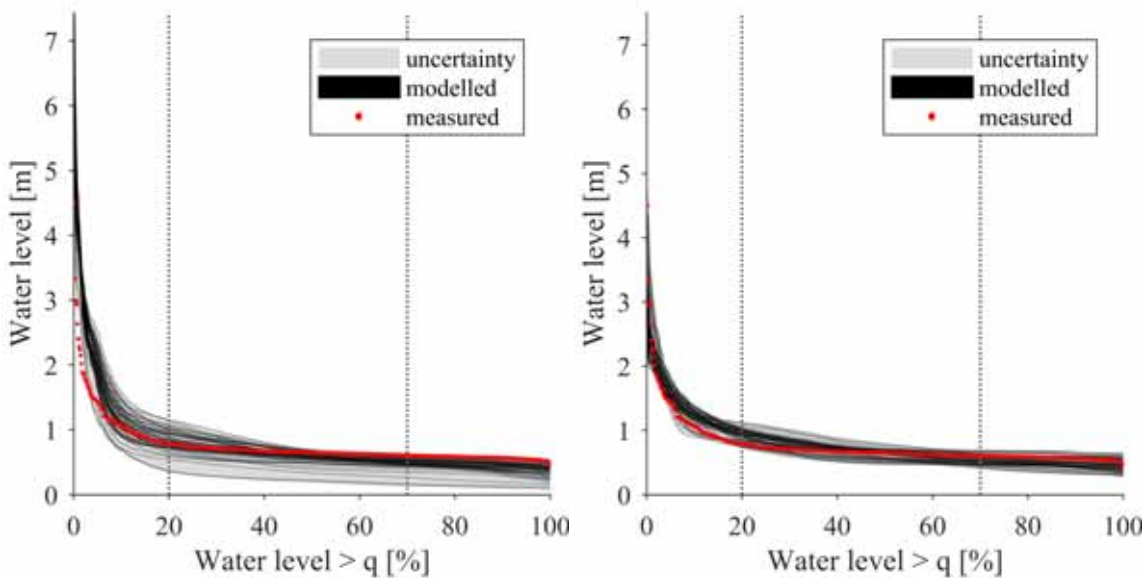


Figure 22: *Simulated water level from 50 selected parameter sets compared to observed water level over the calibration period (2010/2011) produced by (a) RNI and (b) RI.*



(a) **(b)**

Figure 23: *Duration curve of water levels over the calibration time period 2010/2011 for measured and modelled water levels derived from (a) RNI and (b) RI. The percentages on the x-axes indicate the probability of exceedance. A value of 5% therefore means that only 5% of the water level data are equal or higher than this value (high water levels, dotted black line). Low water levels > 70%.*

level peak modelled by RI is considerably lower than that of RNI. All these observations are confirmed by the water level duration curve of RI (Figure 23 (b)). The uncertainty band is much narrower and measured water levels overlap with the high density areas of the modelled curves. Only at frequencies between 5% and 30%, water levels tend to be overestimated slightly. The highest water level values are almost equal for simulations and measured values.

5.2.3 Predictive capability

To evaluate the predictive capability of the VarKarst model with the chosen initial conditions and 50 behaviourals, it was applied to a validation period for RNI and RI. Again, discharge values modelled by RNI had to be transformed by the rating curve calculated in Section 5.1 and corrected to the mean of measured values. Figure 24 (a) depicts the progression of the 50 corrected simulations from RNI over time with the measured water levels. Generally, the water level dynamics during the validation time period are captured fairly well by the simulations. Especially minor peaks are successfully reproduced. High peaks, on the contrary, are missed consistently by all simulations, particularly in (late) summer 2016. Furthermore, some model runs periodically fall under a water level of 0, which would mean a drying out of the Lingshui spring pool.

The duration curve of RNI in Figure 25 (a) summarizes these results. The black line marks the point where water levels fall under 0 m. The uncertainty band is remarkably large,

spanning almost 1.5 m at around 20%. Even so, the line of measured water levels is still close to the highest density areas of the simulations for most of the time. Contrary to the calibration time period, water levels are now mostly underestimated between exceedance probabilities of 5% and 30%. It has to be said that measured water levels always lie within the general uncertainty band except for some isolated events in the high water

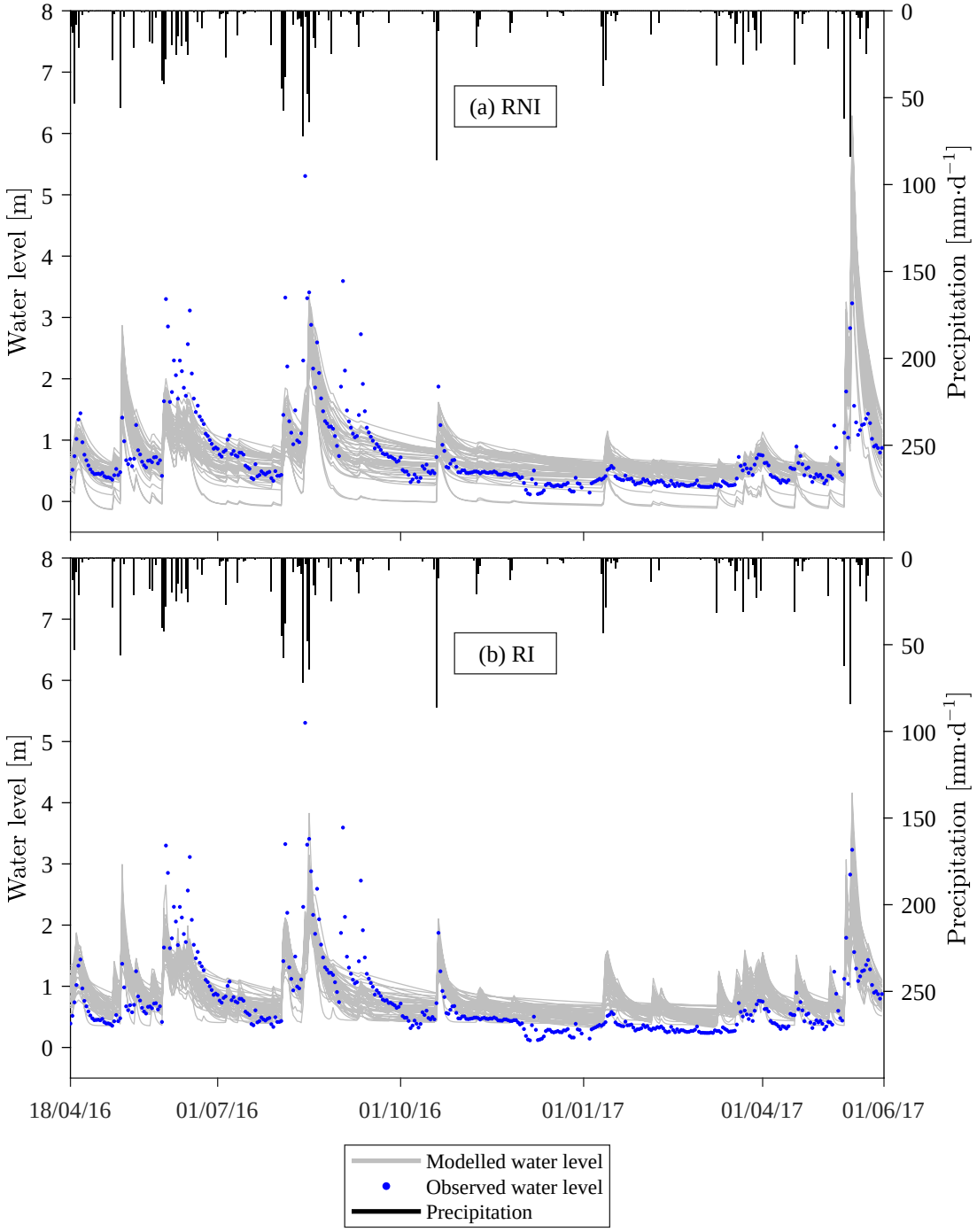


Figure 24: *Simulated water level from 50 selected parameter sets compared to measured water level over the validation period (2016/2017) produced by (a) RNI and (b) RI.*

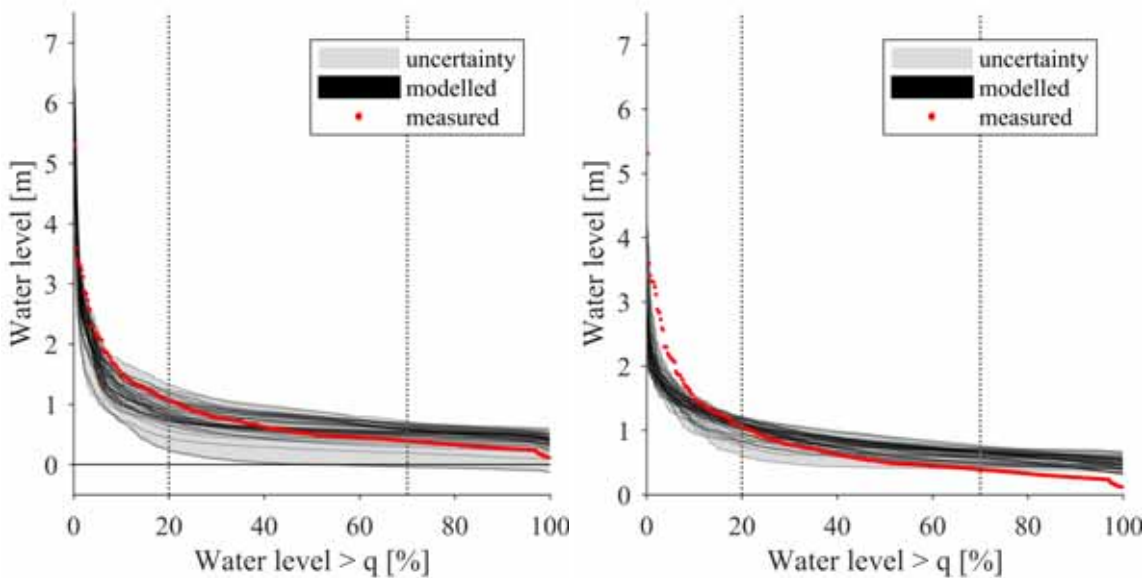


Figure 25: *Duration curve of water levels over the validation time period 2016/2017 for measured and modelled water levels derived from (a) RNI and (b) RI. The percentages on the x-axis indicate the probability of exceedance. A value of 20% therefore means that only 20% of the water level data are equal or higher than this value (high water levels, dotted black line). Low water levels > 70%.*

level sections over 2 m.

The predictive capability of the model can be evaluated by comparing the mean of the objective function values of the 50 selected parameter sets of the calibration period to that of the validation period. The results are presented in Table 5. For RNI, the mean KGE_{rvar} is reduced by 44%. While the best model run for RNI during the calibration period reached a KGE_{rvar} of 0.74, it was only 0.58 for validation.

For RI the picture is slightly different. As model runs spread less widely over the validation period (Figure 24 (b)), they miss the actual dynamics of water level more often. This happens during the peaks in summer 2016 as well as during the winter recession of 2016/2017, where all of the model runs overestimate the extremely low water levels. Furthermore, RI simulations produce pronounced peaks between January and May 2017 that do not appear so noticeably in the water level data. The duration curve of water levels confirms these observations (Figure 25 (b)). The uncertainty band is considerably narrower than for RNI, so that measured water level is not always captured by it. This is especially the case during very low flow conditions (exceeded by 80% and 100% of the values) and high flow conditions exceeded by only 10% to <1% of the values. Between the 10% and 70% exceedance probability, measured water levels are well contained within the uncertainty band.

A 33% reduction of mean KGE can be noticed between the calibration and validation period for RI (Table 5). RI simulations therefore perform better during the validation

period compared to RNI simulations. The performance of the best RNI run reduces the KGE from 0.72 to 0.61. This is a clear argument for using the selected parameter sets of RI for the prediction of future states, whose results are presented in the next subsection.

5.3 Uncertainty under future climate scenarios

To see how far actual trends for the main climatic variables could already be identified in the region, yearly means and sums for temperature and precipitation, respectively, were calculated for the time series 1951-2016 obtained for Nanning weather station from GHCN-D (Figure 26). A simple linear regression was used to give a general idea on tendencies. Overall, the temperature increases by 0.45°C over the whole time period at a rate of 0.007°C·yr^{-1}. For precipitation, the identified yearly decline is barely recognizable at 0.81 mm·yr^{-1}. Furthermore, there is a large variability of yearly precipitation sums. The driest year of the time series occurred in 1989 with a minimum of 827.0 mm whereas 2001 was the wettest year with 1987.5 mm of precipitation. The standard deviation around the mean of 1291.1 mm is 231.2 mm.

The PET results determined with the Thornthwaite equation from temperature values obtained from the delta change scenarios differed largely. The highly probable (97.5% probability) temperature rise of 2.7°C, results in a yearly PET of 2101.2 mm, an increase in PET of 41% compared to the validation time period. The median and low probability (2.5%) scenarios result in much higher PETs of 2452.7 mm and 3188.8 mm, which predict increases of 64.6% (114%, respectively) compared to the validation period.

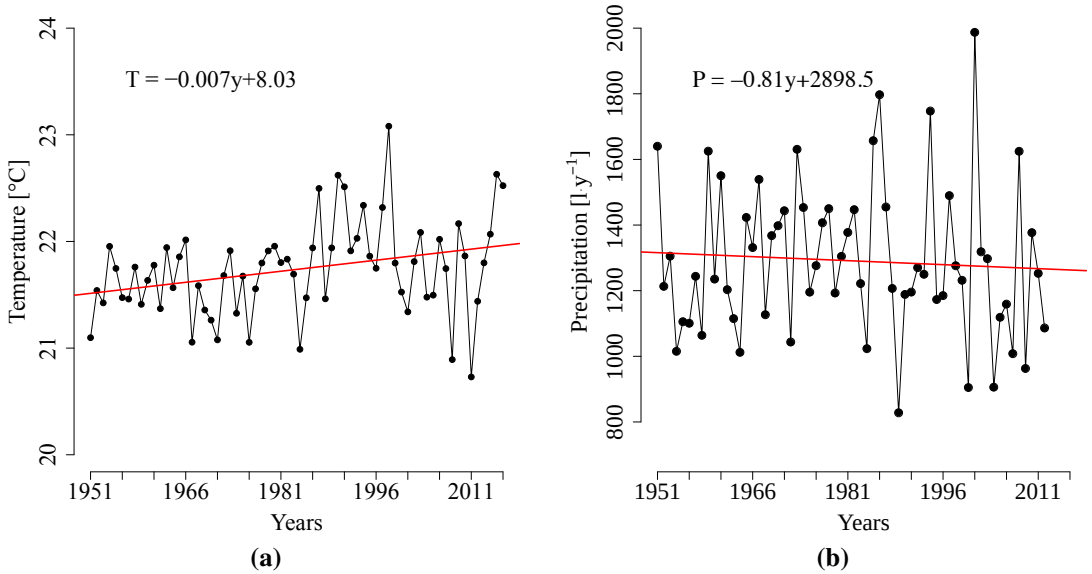


Figure 26: *Yearly means of temperature (a) and sums of precipitation (b) over the period of 1951-2016, calculated from the GHCN-D dataset. The red line is the linear regression through the yearly data for an impression of general tendencies.*

Figure 27 shows duration curves of model simulations for the validation times series under current conditions combined with nine different future conditions based on the nine climate scenarios described in Section 4.5. It was decided not to display the results of every single model run as the aim of this analysis was to compare general uncertainty instead of evaluating individual model run performance. The displayed uncertainty bands are the envelopes around all 50 simulations. The plots under each climate scenario show the p-value results of Welch's t-test. Notice that p-values often fell under the 0.01 significance level (as marked by a dashed line). Table 6 summarises the percentage of exceedance probability points where Welch's t-test identified a significant difference between current and future conditions. For many exceedance probability points, the differences between uncertainties of current and future conditions were highly significant. A layout was chosen in Figure 27 which allows to evaluate not only if a point was significant but also how strong the significance was. The highest degree of difference between current and future conditions was found for scenarios ETLPH, ETLPM and ETHPL. p-values for these scenarios get close to or fall below 10^{-10} at exceedance probabilities under 20%. For ETLPH (high evapotranspiration, low precipitation) and ETLPM (high evapotranspiration, median precipitation), the uncertainty band of future conditions is clearly shifted towards lower water level values whereas for ETHPL (low evapotranspiration, high precipitation) a general shift towards higher water values was observed. The least significant differences between current and future conditions could be identified for ETHPM, ETLPL and ETMPM, where the significance level was never or only partially undershot and uncertainty bands are largely identical. A clear trend that can be identified for all scenarios except for ETHPM and ETLPL is that the difference between simulations is not or only slightly significant during low water levels with high exceedance probabilities and then strongly increases towards water levels with exceedance probabilities < 50%. The significance levels level of again under exceedance probabilities of 5% for most scenarios. Another trend common to all scenarios is a general tendency (even if sometimes only minor) towards lower water levels. Only ETHLPL and ETMPL contrast this trend and show slight shifts towards higher water levels for future conditions.

Table 6: *Percentage of counts at which model runs for current and future conditions were significantly different (significance level 0.01) according to Welch's t-test. Total number of exceedance probability points is 409 (length of the validation time series).*

Scenario	Count [%]
ETHPH	38.39
ETHPM	0.00
ETHPL	97.07
ETMPH	92.42
ETMPM	5.62
ETMPL	37.65
ETLPH	100.00
ETLPM	95.35
ETLPL	0.00

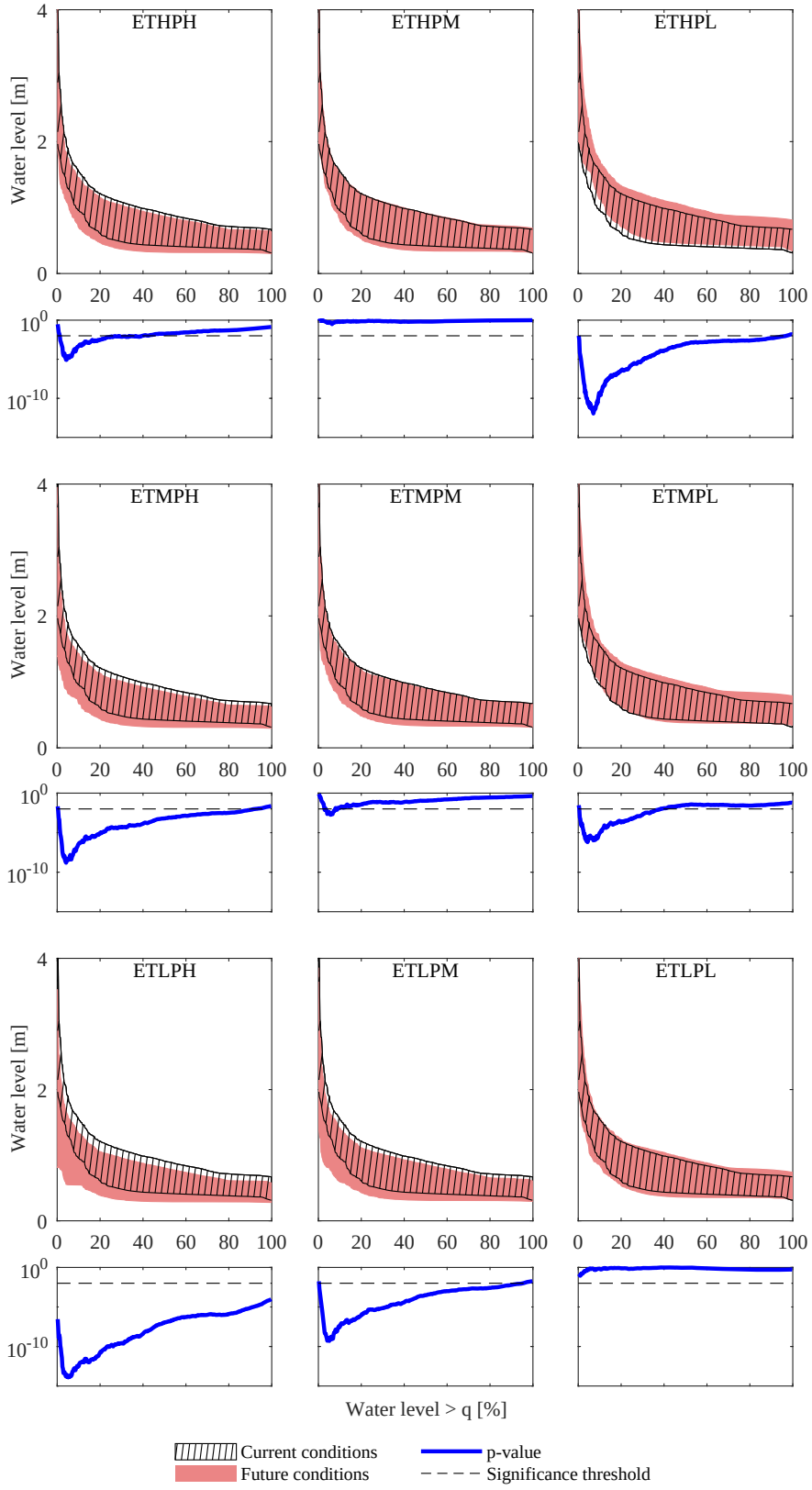


Figure 27: *Uncertainty bands of duration curves from acceptable parameter sets for current conditions and predictions for nine future climatic scenarios from Chen et al. (2011). Underneath each climate scenario plot, the distribution of p-values from Welch's t-test is mapped on the exceedance probabilities. The dashed line marks the 0.01 significance level. Notice that the x-axis is plotted on log-scale for better representation of small p-values.*

6 Discussion

6.1 Uncertainties from input/output data

Renard et al. (2010) stress the importance of quality and representativeness of input data for modelling to ensure reliability of model output. In his study on facets of uncertainty, Beven (2016) stresses how important it is to know about input and output data uncertainty. According to him, this can avoid Type II errors of rejecting a model with partly insufficient performance due to errors in the data source. Therefore, uncertainties in input data should be carefully considered when interpreting model results, particularly in a low data environment as it was given in the present study.

The climate data for 2010/2011 were retrieved from a weather station in Nanning, about 60 km away from the study site and for gap filling in 2016 from a weather station about 30 km away. These stations were used for precipitation assuming its representativity for the study site. However, precipitation patterns often occur with a high spatial variability (Chaubey et al., 1999; Kirchner, 2006) and the assumption could be highly erroneous as precipitation values could have significantly differed between the Lingshui spring basin and the weather station in Nanning. This is particularly the case as the subtropical climate leads to extreme precipitation events whose intensities can vary largely in space. An interpolation of multiple precipitation stations among the Wuming basin would have been necessary to take spatial variability into account and reduce the use of inaccurate precipitation data for modelling (Tabios and Salas, 1985).
The quality of the evapotranspiration data also has to be treated with caution. Data for 2010/2011 have been downloaded online and there was no reference data to how they were measured/calculated and no additional data set available to check their integrity. As shown in Figure 13, the comparison between the downloaded data and data calculated from air temperatures by Thronthwaite, the transformation of temperature data for 2016/2017 brings additional uncertainties. Low values are highly underestimated whereas higher values have a slight tendency of overestimation. Chen et al. (2005) compared the Thornthwaite approach to the more physically-based Penman-Monteith method and evapotranspiration pan measurements throughout China. They found out that Thornthwaite has a high tendency of underestimating PET in southwest China, particularly over the dry season in winter. This should be kept in mind as the climate scenarios show how highly sensitive VarKarst reacts to the PET input.

As the VarKarst model follows an inverse modelling approach, the quality of water level data used for calibration is just as important as that of input climate variables. As Guo et al. (2016) reported, water level in the Wuming spring pool is sometimes impacted by backwater flows from the Wuming river with accompanying sedimentation. This im-

pact cannot be corrected or described, as there are no records about when such events happened and which effect they could have had on the water level data. Hartmann et al. (2017) showed how important it can be to exclude disinformative time periods from the modelling to reduce uncertainty and improve parameter identifiability. They claim that as the model has to compensate for structural errors during such periods, a wider range of parameter combinations is needed which favours equifinality. To get a better impression on the influence of backwater flows on water level, it would be helpful to have a documentation of such events for future studies. Another aspect with regards to output data is the length of the water level time series of only about 3 years, which is a relatively short time period. Kay et al. (2009) claim that a time series needs to be sufficiently long in order to cover a wide range of flows (or water levels) and to be effective for calibration. Yapo et al. (1996) performed a sensitivity analysis on the optimal length of calibration data for a lumped rainfall-runoff model and found out that an 8-year period of data would be optimal. They also report that uncertainties can be decreased when rather wet years are included.

An additional downside to the data for the modelling purpose of this study are the gaps in the already relatively short time series. For climatic data, these had to be filled by information from other stations, which adds further uncertainty. Measuring devices for water level often failed during periods with extremely high peaks which resulted in missing values. These high water level values would be particularly important for recording the variability of the system (Hartmann et al., 2014a).

It is also problematic that no supplementary information was available on withdrawals from wells in the region. According to personal communication with Fang Guo (04.09.2017), usage of private wells has increased in the region in recent years without continuous documentation. This information would not only be useful to be included in the modelling process to get a correct record of the water balance but also for management purposes of the groundwater resources. In the worst case, overexploitation of an aquifer could lead to drying up of the spring, as has happened with the Jinci spring in North China (Hao et al., 2009). Another aspect which is not well documented yet for the Lingshui spring system, that could have an impact on the groundwater storage, is that according to Guo et al. (2015), the Lingshui spring system acts as an estavelle during backwater flow conditions following heavy rainfalls.

6.2 Model performance

In order to evaluate model performance, the main focus lies on the following aspects: Could all prerequisites basic for model performance (e.g. groundwater stationarity) be met? How do "optimal" parameter ranges and values obtained for the Lingshui catchment compare to other VarKarst studies? Do parameter values reflect the local conditions? How

did the model simulation perform? Which factors could have impacted the quality of the calibration results?

The most difficult prerequisite to meet in the SCEM-UA calibration was finding settings that would allow stationarity in the groundwater compartments. Even though the choice on initial conditions (EpiMax versus SoiEpiMax) was influenced by groundwater stationarity in one particular setting, these conditions were not the governing factor. Stationarity could be achieved in the first compartments whereas the middle ones still showed a clear tendency towards filling up. This was considered an unavoidable artefact caused by the short time series available for calibration that could not be overcome in all trials of different parameter ranges (personal communication, Andreas Hartmann (30.10.2017)). A significant filling effect of the groundwater compartments would also have led to underestimation of water levels at the beginning and overestimation at the end of the calibration time series. This was not observed.

Much more than on initial conditions did the choice of parameter ranges show a clear impact on groundwater storage behaviour. Parameter ranges for a_{SE}, a_{fsep} and a_{gw} had to be significantly increased in comparison to Brenner et al. (2016) in order to achieve stationarity. The need to change these parameters can be explained by the local hydrogeological conditions. An increase of a_{SE} results in a more extreme distribution of soil and epikarst storage size, meaning that they remain comparatively small in all compartments except for the last two. This also leads to a much smaller epikarst storage coefficient $K_{mean,E}$ and thus increased recharge rates. This is in accordance with the variable soil thickness found in the catchment (ranging from bare karst to up to 30 m of soil) that Guo et al. (2015) reported along with the sparsely developed epikarst as is typical for peak forest. The low distribution of recharge to the conduit compartment, as caused by a high a_{fsep}, lets all recharge water drain diffusively to the underlying groundwater storage compartments. Guo et al. (2015) and Guo et al. (2016) concurrently report of the dominance of matrix porosity in the covered karst and fissure porosity in the Niaja syncline, which favour diffuse transport of groundwater. Finally, the unusual values of a_{gw} result in extremely long mean residence times of groundwater in connection with extraordinary volumes of the storage compartments. Without a doubt, mean residence times in the range of 10^{73} and 10^{87} days cannot and do not realistically reflect real-world physical conditions. In spite of the missing realism of some of the values for K_{GW} and the extreme groundwater storage sizes, especially for the first few compartments, from a qualitative view, these findings are in high accordance with descriptions of the study site by Guo et al. (2015). They found relatively slow transport times of groundwater for the Lingshui catchment of 17 - 23 m·d^{-1} in a tracer experiment.

The large increase of parameter ranges needed for a_{SE}, a_{GW} and a_{fsep} and their high

parameter values are noteworthy. They much exceed the values found by Hartmann et al. (2013a,c, 2014b). This might be explained by the fact that all the other studies were conducted at study sites which showed a clear duality of porosity. However, it is remarkable that the values found in the present study also exceed parameter ranges found by Brenner et al. (2016) by such a large degree, as their study site of the English Chalk was comparable to the Lingshui catchment with regards to the dominance of fissures in groundwater recharge. In the present study, the best values obtained for a_{GW} cause highly unrealistic mean groundwater residence times and storage volumes. Similarly, the calibrated values of a_{SE} evoke very unrealistic distributions for the variable parameters K_E, $V_{max,S}$ and $V_{max,E}$ (Figure 18) with extremely low values over most of the compartments to very high values in the last two compartments. Gallagher and Doherty (2007) deal with realism of parameters in lumped parameter models and doubt whether their realism is even assessable. In the simplest setting, parameters are only a mean representation of real physical settings. In most lumped models, as is the case for VarKarst, the situation is even more complex and parameters might represent several processes at the same time. Therefore, the high values of a_{GW} and a_{fsep} in particular allow two possible conclusions: The model represents local hydrogeological conditions and recharge processes adequately. Large groundwater storages account for the baseflow component in the water level data which retains constant discharge in dry periods over winter. The second conclusion could be that the values actually reflect structural errors in the model and interactions between parameters (Gallagher and Doherty, 2007). Another indicator for this could be the low values of K_C of about 4 days in comparison to (Brenner et al., 2016) (about 40 days), which contradict the a_{fsep} and a_{GW} values. Water that gets to the conduits drains out of them really fast. This rather seems like a way of the model to counteract the large groundwater compartments in order to evoke fast-responding peaks in the water level dynamics. Not much information was available about the actual existence and importance of conduits in the Lingshui catchment. A fast component could also result from fractures in the bedrock as reported in Guo et al. (2016).

To give a final explanation for these results, additional data would be needed in order to conduct a multi-objective calibration approach (Gupta et al., 1998; Efstratiadis and Koutsoyiannis, 2010; Hartmann et al., 2014b; Brenner et al., 2016). This concept is going to be further elaborated in Section 6.4.

Considering the work of (Gallagher and Doherty, 2007), setting realistic parameter ranges to $V_{S,mean}$ and $V_{E,mean}$ might have prevented an effective search for more realistic parameter values. The catchment area A of 688 km^2 identified for RI almost completely coincides with the value reported by Guo et al. (2015).

Altogether, it can be said that the model results produced by optimal SCEM parameter sets can adequately represent the hydrodynamic behaviour of the Lingshui karst system, even

though quantitative realism of some parameters might be lacking. This is reflected by the high values of KGE_{rvar} and KGE as well as the performance of simulations compared to measured water levels. In spite of the obvious deficiencies, the VarKarst model was the best choice in this study, as it can take account of all different kinds of geological settings due to its semi-distributed character. This was important, to obtain a first understanding of the functioning of the aquifer and its predominant flow characteristics. It can serve as a basis for future measurement campaigns.
Nevertheless, as several studies showed (Hartmann et al., 2013b; Velázquez et al., 2013; Chang et al., 2017), it might as well be advisable to explore other possible model set-ups and approaches (e.g. equivalent porous medium) in order to ensure that an appropriate model structure is chosen for management purposes.

Some additional sources of uncertainty caused by the calibration procedure should be mentioned briefly at this point. First, Wilby (2005) report on how important the calibration time series can be for model identification as they compared interannual variability of the NSE for the same parameter set. Due to the scarce availability of measurements in the Lingshui spring catchment, there was only a very limited leeway possible on the calibration period. Secondly, according to Mazzilli et al. (2012) the choice of data for the warm-up period could have had a large impact on the initialisation bias and therefore uncertainty, too. Particularly, as both the calibration and validation time series were so short that no data could be spared for a warm-up period from real data. The chosen approach of repeating the first year of the time series three times can barely represent the actual initial conditions at the beginning of the calibration of the time series. Yet, as this is a commonly used method (e.g. Perrin et al. (2001); Mazzilli et al. (2012); Hartmann et al. (2017)), it was reasonable to be used here, too.

For interpretation of the SCEM results, it has to be considered that they only represent one parameter realisation of many possible ones and can only give a general orientation about the performance of the VarKarst model in the Linghsui spring system. It requires the uncertainty analysis to demonstrate, how identifiable parameters are and therefore how reliable predictions of the model can be.

6.3 Uncertainties related to the rating curve

Rating curve parameters found by least square fitting produce a good agreement with the measured data. The rating curve resulting from parameters from the calibration procedure, on the other hand, lies well beneath the least square fitted one. A simple adjustment can be obtained by rising the water level data by 0.5 m. Nevertheless, the question arises, where this difference comes from. Discharge values from the RI model run mostly lie well

above 10 000 $l \cdot s^{-1}$ and about 5 times higher than measured ones. This potentially requires quite different rating curve parameters to obtain the same values for the water level. Similarly, RNI water level results that were calculated by the least square fitted rating curve had to be corrected by a further 1.1 m downwards to be in accordance with actually measured water levels. Even though this is necessary and acceptable for the further modelling procedure, it adds another uncertainty source. All of this allows the following conclusion: Modelled discharge values are much higher than those actually observed in the Lingshui spring pool. The inclusion of the rating curve into model calibration allows to account for this fact and therefore incorporates all preceding uncertainties. A least square fitted rating curve on the other hand is bound to actually measured discharge data and therefore lacks the flexibility to account for model uncertainties. Of course it would have been useful to include discharge values in the calibration procedure, using a multi-objective calibration approach, with discharge values as well as water levels, to ensure that measured and modelled discharge values match. This was tested, even with different weighting schemes, and led to very unsatisfactory results with regards to water level dynamics (see appendix, Figure I.1). The results showed that the resolution of the available discharge data was not high enough for a multi-objective calibration approach. More frequent discharge measurements, particularly in extreme discharge conditions, can be expected to lead to a better reflection of discharge dynamics in simulations.

The translation of water level measurements into discharge (and vice versa) is a task that according to Braca (2008) requires a wide theoretical background. In this study, it has to be considered that this translation potentially adds another dimension to the uncertainties already implied in the data. Even though a rating curve could be identified that fits measured data well, its flaws have to be contemplated carefully. Of course, the aforementioned deficiencies of water level measurements have a direct impact on the rating curve. Additionally, as backwater flows happen in the spring pool, the assumption of stationary flow conditions that the formulation of the rating curve was based on may not be justified. According to Braca (2008), this non-stationarity would need to be incorporated in the rating curve by using individual formulas for different flow conditions. However, more discharge measurements would be necessary to clearly identify these conditions. Discharge measurements were mainly available for mean flow conditions. This is a problem often occurring in natural rivers and lakes that comes with the difficulty to conduct flow measurements during high flow conditions (Braca, 2008). If references are missing for high flows, extrapolated values become highly unreliable (Clarke, 1999).

6.4 Reliability of simulations

In this section, the reliability of simulations by the VarKarst model in the Lingshui catchment is going to be evaluated on the following questions: Which parameters show high

identifiability? How large was the informational value of the water level time series for parameter confinement? How did acceptable parameter sets act in prediction mode and during the validation period with regards to uncertainty? All of these questions are going to be answered by comparing results from RNI and RI in order to give an evaluation on whether the inclusion of a rating curve into the calibration procedure is sensible.

It has to be kept in mind that the choice of threshold for acceptable parameter sets could not be based on the basis of former studies. Even though similar ones have been conducted (e.g. Hartmann et al. (2017)), their background was different to this study. It was sensible to declare the 50 best MCAT runs as acceptable looking at the high number of MCAT runs necessary to capture model uncertainty adequately. Nevertheless, the impact of the choice of this threshold should not be neglected, as all further analyses of this study are going to be based on it.

With regards to parameter identifiability, large differences between RNI and RI could be observed. For RNI, mainly parameters which required widening of their ranges showed high identifiability. This proves that the widening of parameter ranges was necessary to find adequate values for those parameters. Whereas at least half of the parameters of RNI were identifiable, only the cumulative distribution of K_C was similarly narrow for RI. Instead, the rating curve parameters a and b became highly identifiable. This could be a confirmation for the aforementioned compensation of additional model parameters to ensure representation of hydrological dynamics. Furthermore, the increase of parameter ranges for selected parameter sets might be a hint that the model structure cannot appropriately represent the additional processes and thus compensates them through other model parameters (Hartmann et al., 2012a). Generally, it has to be acknowledged that parameter identifiability was very low in this study, even for RNI. This becomes obvious when compared to, for example, Hartmann et al. (2013a) and Brenner et al. (2016) where almost all parameters were highly identifiable. Yapo et al. (1996) list several possible reasons for missing parameter identifiability, of which the following could be interesting for evaluating the results of this study: Parameter interdependence, parameter stationarity, data noninformativeness and insensitivity. These are hard to differentiate but some factors of the model set-up and data used in this study might have favoured them, mainly, the high number of parameters used in the VarKarst model and the shortness of the calibration and validation time series. This could particularly be the case for RI. Hartmann et al. (2017) make a similar observation regarding the number of parameters in the VarKarst setup of their study. According to Kirchner (2006) every new parameter added to a model adds another dimension that could lead to over-parametrization and therefore parameter interaction, insensitivity and equifiniality (Beven, 2006). Kirchner (2006) states that these parameters potentially create much more flexibility than a calibration data set is capable

to resolve. Furthermore, he claims that a high number of parameters also prevents clear identification of structural model errors: 'By making it easier for models to get the right answer, overparameterization makes it harder to tell whether they are getting the right answer for the right reason' (p. 3, Kirchner (2006)). Jakeman and Hornberger (1993), Ye et al. (1997) and (Kuczera and Mroczkowski, 1998) claim that a model should consist of a maximum of 5-6 parameters if calibrated on discharge only (or on water level, as is the case here). In order to increase parameter identifiability, other proxies representing the system should be used in a multi-objective calibration. The missing model identifiability could potentially be overcome by including high-resolution chemical and isotopic data in the calibration process (Kirchner, 2006). Studies from Kuczera and Mroczkowski (1998); Hartmann et al. (2016, 2013c) and Rimmer and Hartmann (2014) proved that the inclusion of hydrochemical data, e.g. $SO_3{}^-$ or $SO_4{}^2$, is highly important to ensure higher identifiability of parameters. Hartmann et al. (2013a) even claim that hydrochemical information was crucial to find a reasonable parameter set in their study. Brenner et al. (2016) showed how groundwater level data and a corresponding weighting scheme can be incorporated in VarKarst. Worthington (2003) also demonstrated that both discharge and groundwater level measurements are important for identifying differences of flow in the matrix and fissures compared to conduits.

The value of additional data becomes even more obvious when the confinement of parameter ranges is considered. The reduction between the 25^{th} and 75^{th} percentiles in this study was quite low for most of the parameters which indicates that the informational content of water level data alone was not sufficient to clearly identify parameter values. Furthermore, a distinct reduction of parameter ranges was only achieved for parameters which were also highly identifiable. Hartmann et al. (2017) demonstrated that by including additional hydrochemical data ($NO_3{}^-$ and $SO_4{}^{2-}$) in the calibration process, informational value can be added and a significant confinement of parameter ranges achieved. Furthermore, they compared how the inclusion of information on saturated and unsaturated zone processes and the exclusion of disinformative periods in the discharge time series can help to confine parameter ranges. From their results it becomes clear that in comparison to the low confinement attained from a calibration on discharge (or water level in the present study) only, the inclusion of additional data highly increases the informational content for parameter range confinement. In the present study it could have been highly instructive to exclude disinformative periods caused by backwater flows from the Wuming river in the same manner as in Hartmann et al. (2017). They showed that the exclusion of disinformative flow states can result in a significant increase of confinement between the percentiles and a reduction of uncertainty in predictions. Another study by Hartmann et al. (2015) demonstrated that the confinement method can also be very useful when other soft rules can be used, e.g. bias between measured and simulated actual evapotranspiration or plausibility of parameters like $V_{mean,S}$ or $V_{mean,E}$. A verification of

a_{SE} by comparing the distribution of maximum soil depth $V_{S,i}$ with a map of soil depths in the Lingshui catchment could have been a soft rule to confine the parameter sample. As parameters show different responses to the additional information used, the confinement approach can also be used to identify which processes might drive certain parameters (Hartmann et al., 2015).

Despite of the restricted parameter identifiability, RI model runs generally lined up with measured values with smaller uncertainty. Particularly in the low exceedance probabilities, RNI model runs resulted in much higher uncertainties. This is an additional hint for equifinality in the RI model setup coming from the increased number of parameters introduced by the rating curve fitting with KGE. As more degrees of freedom are assigned to the model, it can compensate for structural model errors, so that parameter sets unexpectedly produce similarly good simulations with completely different parameter values (Wilby, 2005; Beven, 2006). This theory is supported by the fact that model runs from RI do show much more variable dynamics during the calibration time series. For RNI, on the contrary, model runs mainly follow the same patterns. In this context, the aforementioned methods to potentially increase identifiability and therefore confidence in the model and its parameters become essential in order to make sure that the modelled results are 'right for the right reasons' (Kirchner, 2006) and to avoid 'monsters' that make hydrological models right for the wrong reasons (Goswami and O'Connor, 2010). An investigation of the water balance showed no signs of 'misbehaviour' of the model. But the difficulties in finding stationarity in the groundwater compartments could hint at deficiencies of the model structure.

The validation proves the obvious drawbacks of the scarcity of data in the calibration. For both modelling approaches, the reduction of the mean of the objective function of all 50 selected parameter sets is noteworthy between the calibration and validation period. The large uncertainty band of RNI model runs completely captures measured water levels during the validation period. RI model runs, on the other hand, miss the measured water levels in the high and low flows exceedance probabilities. Nevertheless, RI based simulations lose less of their predictive capability in the validation period than RNI based ones. In the face of the high equifinality mentioned before, this result was unexpected. It could be a sign that the variability of the validation time period does not sufficiently differ from that of the calibration period and therefore the predictive capability of the model setups cannot be fully evaluated (Kirchner, 2006).

Surely, both modelling approaches have their flaws. RNI leads to much higher uncertainties compared to RI and its predictive capability is lower. In contrast, RI has a lower parameter identifiability and misses measured water levels during the validation period.

Therefore, no approach can clearly be rated as better. RI was mainly picked for climate predictions in the following because of its higher predictive capability. As the focus of this study was on describing the general behaviour of the Lingshui spring system, the fact that RI simulations missed some extremes of the validation time period was not a disqualifying factor for their usage in climate predictions. A final decision about the usefulness of the set-ups can only be made once sufficient data for parameter calibration and uncertainty analysis are available. It was justified to use the calibrated model for climate projections in spite of its comparatively poor predictive capability compared to other studies, as model uncertainties were communicated, which gives an honest picture on the model's performance (Blöschl and Montanari, 2009). Further arguments for the usage of the model in climate projections were the good results of the objective function (KGE) as well as the generally good representation of the water level dynamics.

6.5 Uncertainty under future climate scenarios

In the following, the newly implemented evaluation approach (see Section 4.5) for trustworthiness of climate impact studies based on models with high uncertainties is going to be considered thoroughly. In particular, potential limitations of the approach are going to be pointed out.

The identified tendencies in the climate time series from Nanning weather station go along well with observations of past climate change in the region. Ren et al. (2012) report temperature increases of 0.08 ± 0.03°C during the period of 1906-2005 for the whole of China but relatively weaker changes in the Southwest. This was confirmed by a study of Lian et al. (2015) who identified a yearly temperature increase of only 0.011°C in Guilin. The analysis of the current study showed, that in Nanning, about half of the years after 1980 show noticeably higher temperatures compared to the years before but the general trend of the time series was very low. No clear tendency was recognizable for precipitation over the period and natural variability appears to be very large. Accordingly, Ren et al. (2012) report of no significant precipitation changes between 1906 and 2005 for the whole of China. Guo et al. (2015) describe regular circulations of wet and dry periods every 13 to 16 years for the Wuming basin, which is in line with the high natural variability observed at Nanning weather station. This, of course, is a severe restriction on the simple delta approach used for the climate scenarios, as it does not account for any climatic variability, which could, according to Arnell (2003), lead to severe differences in model outputs. A general increase of interannual variability of temperatures and precipitation can also be identified in Figure 26. According to Praskievicz and Chang (2009), this could be another effect of climate change that potentially needs to be considered in management plans. It should be noted, that no proper statistical trend analysis was conducted for the

available data. The presentation of tendencies was rather intended to provide a general picture about climate variability and obvious tendencies. Nevertheless, the tendency towards rising temperatures shown in the dataset and the general climatic trends observed in (southwest) China published in other studies give good reason to conduct an analysis on reliability of predictions on future climate developments.

The enormous impact of increasing temperatures over China on hydrological systems, as predicted by Chen et al. (2011), becomes particularly obvious for PET. As the results of the calculations based on the Thornthwaite equation show, higher temperatures usually cause higher evapotranspiration, which can seriously affect the water balance in a region. For the time series of 2016/2017, the low probabilistic temperature scenario (+5.2°C) led to a PET increase of 114%. Guo et al. (2015) claim that groundwater depths of 7 to 50 m and the dominance of covered karst could potentially protect groundwater of the Lingshui system from evapotranspiration. Goldscheider and Drew (2007), on the other hand, declare that the presence of a soil cover, as is the case in most of the Lingshui spring catchment, significantly increases evapotranspiration and results in a much lower proportion of recharge. No study has been conducted on this specific topic at Lingshui yet, but would surely be interesting in order to formulate management plans in the face of climate change. Furthermore, it is questionable, whether in this specific climatic region temperature will actually be the driving force of PET as decreasing trends of PET in China were identified by Thomas (2000). They also found out that sunshine duration and relative humidity might have a much larger influence on PET in southwest China than temperature.

If we trust Welch's t-test and the chosen significance level to be reliable indicators, this study shows that predictions made by the VarKarst model for future scenarios can be declared trustworthy. The uncertainties of predictions under current and future climate differ significantly and over the whole range of exceedance probabilities for most of the examined climate scenarios. Particularly for high water levels of frequencies observed less than 10-20% of the time, Welch's t-test often indicated a very high significance well below the threshold of 0.01.

The impact of potential evapotranspiration becomes obvious in the climate prediction results. Only the 26.9% of precipitation increases of the low probability predictions (PL) outweigh the evapotranspiration losses caused by higher temperatures and do lead to higher water levels compared to current conditions. As mentioned before, the question remains, in how far PET will actually increase with rising temperatures. Nevertheless, the tendency towards a reduction of the water level also matches the results of Guo et al. (2015) who found a general decrease of discharge in the Lingshui spring over time. The results of the climate predictions are also interesting under the management aspect as they show that the highly different climate scenarios can lead to completely opposing trends

(generally lower water levels versus generally higher water levels) which would require different policy actions. In all these interpretations it has to be considered, though, that the RI model setup did already underestimate observed water levels under low exceedance probabilities during the validation time period (Figure 25). Therefore, under the assumption of stationarity, future predicted water levels can be expected to be underestimated for high water levels, too. Nevertheless, the goal to test an approach for evaluating the reliability of future predictions by considering the model uncertainties was successfully accomplished. If uncertainties of input and output data could be reduced and more extensive methodological approaches could be used to cut additional uncertainties along the way, this approach represents an interesting way for future climate impact studies to evaluate the effects of model uncertainty in future predictions.

Some additional uncertainty sources have to be mentioned. First of all, it was decided to use averaged climate projections for the whole of China instead of regionalized ones. Chen et al. (2011) also present probabilities for different warming levels and precipitation changes in a spatially differentiated manner. However, these were not helpful for the concept of our study, as they did not supply probabilistic distributions for different locations. In future work, spatial variability of climate change should be considered as results from Chen et al. (2011) show that e.g. the probabilities for an anticipated 4°C temperature change largely differ across the country with probabilities exceeding 60% in the northeast whereas in the south probabilities are much lower. Studies following this thesis could e.g. formulate more regionalized probabilistic climate scenarios for the area of the study site to overcome this issue. This could be achieved by using regional climate models (RCMs) to downscale the results obtained by AOGCMs as they are known to incorporate more realistic topographic forcings (Xuejie et al., 2001).

Another aspect neglected in this study that could be object to further research is the difference in projections between seasons. For instance, probabilities for precipitation increase equal to or over 10% for the 2070-2099 period were projected at 0.9 for winter and 0.34 for summer by Chen et al. (2011). The seasonal patterns also coincide with spatial patterns. For example, winter precipitation is projected to increase by over 20% with a high probability of >50% in northeast, north and northwest China. In summer, on the contrary, precipitation is likely to rise by 10% in the Southwest. As particularly scenarios containing at least one extreme (ETL or PL) caused most differences between uncertainties in current and future conditions in the present study, it can be expected that a differentiation between seasons would have supported an even clearer distinction between uncertainties.

Another assumption made in this study is the stationarity of parameter values over time under changing environmental conditions. Merz et al. (2011) clearly demonstrated in their study that this assumption is often quite unrealistic. They recalibrated parameters of a precipitation-runoff model with six consecutive five-year time periods of runoff. Clear

trends related to change of climatic conditions were distinguished for many parameters which clearly shows that the assumption of parameter stationarity is questionable, particularly in the face of climate change. Therefore, to carry out climate trend analyses, large time series are needed. If such data are not available, these deficiencies in climate impact modelling caused by the assumption of parameter stationarity have to be critically kept in mind when interpreting results.

The results of this study are in agreement with Poulin et al. (2011). In their study, which was conducted for a watershed in Canada, they showed that model uncertainty is usually more dominating in climate impact studies than model parameter uncertainty. They conclude that therefore models of different complexity should be used within such climate impact studies to be able to communicate uncertainties of future predictions related to model structure and potentially choose the model causing least uncertainties for management purposes.

7 Conclusion and Outlook

The aim of this study was to apply the VarKarst model for the Lingshui spring catchment in order to test its performance in the specific regional hydrogeological setting and its predictive capability under local conditions. Generally, it can be said that the model performance was satisfactory with regards to the objective functions and the representation of water level (dynamics) in the spring pool. Furthermore, the dominating hydrodynamical processes simulated by the model matched the general understanding of the aquifer. Nevertheless, some deficiencies were detected along the course of the study. Some of the optimal parameter values identified during the calibration procedure differed significantly from those of other studies conducted with VarKarst before (Hartmann et al., 2013a; Brenner et al., 2016). Even though this might be attributed to the predominating local conditions, it could be also be a hint for equifinality (Beven and Freer, 2001). The low identifiability of most parameters and low information content of the water level time series in terms of parameter range confinement further support this assumption (Hartmann et al., 2017). Furthermore, the calibrated models had a restricted predictive capability.

The most obvious reason for these deficiencies was the low quantity of data available for the catchment at the point of this study. In order to overcome the problems mentioned above, large additional measuring campaigns and longer time series are recommended. These should, for example, include regular hydrochemical measurements in the spring pool, documentation of backwater flow conditions, regular discharge measurements (ideally also during very high flow conditions) and observations of groundwater level throughout the catchment. These additional sets of information would be helpful in order to verify the appropriate functioning of the model and to increase identifiability of parameters (Kuczera and Mroczkowski, 1998; Efstratiadis and Koutsoyiannis, 2010; Renard et al., 2010; Hartmann et al., 2017). Innovative approaches of model evaluation, e.g. the use of soft rules in the calibration process, could further enhance the modelling process and give a clear record of parameter and model uncertainty (Seibert and McDonnell, 2002).

Additionally, alternative model structures, preferably with higher parsimony, should be investigated with the purpose of excluding possible structural model errors (Kirchner, 2006). As the epikarst does only play a minor role in the karst of the region, an exclusion of the epikarst routine from the VarKarst model could be a first step in order to reduce the number of parameters. Once more data are available, the testing of an EPM or a triple porosity approach, as has been used in other karst studies before (Scanlon et al., 2003; Cheng and Chen, 2005), could also be interesting alternatives to the VarKarst model for the Lingshui spring catchment with regards to the predominant local hydrogeology.

Presumably, the inclusion of a rating curve and therefore additional parameters into the modelling procedure did reduce parameter identifiability and possibly caused the model

to compensate for structural errors (Kirchner, 2006). Nevertheless, these assumptions could not be tested thoroughly due to the lack of additional information for the model calibration. To include a rating curve into the VarKarst model, additional discharge measurements as well as information on backwater flows would be especially needed to, for example, apply a multi-objective calibration approach.

In the projections based on future climate change scenarios, not only the extreme scenarios resulted in a signifiant difference between current and future uncertainties. According to the statistical test chosen for this study, almost all scenarios led to (more or less) significant differences and thus reliable predictions. As this was the first time this evaluation methodology for model uncertainties in climate impact studies was used, further studies are needed to investigate whether the high detection rate for significance is plausible. Other significance levels or even statistical tests could, for example, be evaluated with a simulated dataset. The application of the Thornthwaite equation seems inappropriate for the local climatic conditions, particularly when used for the prediction of PET changes (Thomas, 2000; Chen et al., 2005). In order to overcome this issue, other approaches to calculate PET should be used (Chen et al., 2005). Furthermore, more regionalised climate change predictions should be implemented in order to ensure that the predictions used for climate change studies are in accordance with actually observed trends.

Land use conditions in the Lingshui spring catchment have been changing rapidly over the last decades with large impacts on the aquifer and should therefore find consideration in future studies (Guo et al., 2015). It is advisable to investigate in how far these land use changes affect the local water cycle to allow more accurate predictions on future developments of water resources. The large increase in eucalyptus plantations, for example, can be expected to have a major effect on recharge patterns (Engel et al., 2005). Guo et al. (2015) even report of dried up springs in the region in correspondence with eucalyptus cultivation. In addition to this, the consideration of qualitative aspects would surely be equally important for local water resource management and should thus be included in the model.

In summary, the general aim of investigating the functioning of the VarKarst model in a data scarce tropical karst environment was accomplished. Even though some deficiencies, mostly related to data scarcity, became apparent, they were effectively communicated through an uncertainty analysis. A first impression of the driving processes of the Lingshui spring system could be collected and necessary steps for further measurement campaigns were identified that would eventually allow for a VarKarst centered model setup in the future that is reliable enough to serve as a basis for water management decisions in the Lingshui spring catchment, or more generally, in tropical karst environments.

Bibliography

M. Abusaada and M. Sauter. Studying the Flow Dynamics of a Karst Aquifer System with an Equivalent Porous Medium Model. *Groundwater*, 51(4):641–650, 2012.

N. W. Arnell. Climate change scenarios from a regional climate model: Estimating change in runoff in southern Africa. *Journal of Geophysical Research*, 108(D16), 2003.

N. D. Bennett, B. F. Croke, G. Guariso, J. H. Guillaume, S. H. Hamilton, A. J. Jakeman, S. Marsili-Libelli, L. T. Newham, J. P. Norton, C. Perrin, S. A. Pierce, B. Robson, R. Seppelt, A. A. Voinov, B. D. Fath, and V. Andreassian. Characterising performance of environmental models. *Environmental Modelling & Software*, 40:1–20, 2013.

K. Beven. Prophecy, reality and uncertainty in distributed hydrological modelling. *Advances in water resources*, 16(1):41–51, 1993.

K. Beven. A manifesto for the equifinality thesis. *Journal of Hydrology*, 320(1-2):18–36, 2006.

K. Beven. Facets of uncertainty: epistemic uncertainty, non-stationarity, likelihood, hypothesis testing, and communication. *Hydrological Sciences Journal*, 61(9):1652–1665, 2016.

K. Beven and A. Binley. The Future of Distributed Models: Model Calibration and Uncertainty Prediction. *Hydrological Processes*, 6:279–298, 1992.

K. Beven and J. Freer. Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. *Journal of hydrology*, 249(1):11–29, 2001.

G. Blöschl and A. Montanari. Climate change impacts—throwing the dice? *Hydrological Processes*, 24:374–381, 2009.

G. E. P. Box. Robustness in the Strategy of Scientific Model Building. In *Robustness in Statistics*, pages 201–236. Academic Press, Cambridge, Massachusetts, 1979.

D. P. Boyle, H. V. Gupta, and S. Sorooshian. Toward improved calibration of hydrologic models: Combining the strengths of manual and automatic methods. *Water Resources Research*, 36(12):3663–3674, 2000.

G. Braca. *Stage-discharge relationships in open channels: practices and problems*. Number 11 in FORALPS Technical Report. Università degli Studi di Trento, Dipartimento di Ingegneria Civile e Ambientale, Trento, 2008.

S. Brenner, G. Coxon, N. J. K. Howden, J. Freer, and A. Hartmann. A percentile approach to evaluate simulated groundwater levels and frequencies in a Chalk catchment in Southwest England. *Natural Hazards and Earth System Sciences Discussions*, 2016. in review.

K. Chalikakis, V. Plagnes, R. Guerin, R. Valois, and F. P. Bosch. Contribution of geophysical methods to karst-system exploration: an overview. *Hydrogeology Journal*, 19 (6):1169–1180, 2011.

Y. Chang, J. Wu, G. Jiang, and Z. Kang. Identification of the dominant hydrological process and appropriate model structure of a karst catchment through stepwise simplification of a complex conceptual model. *Journal of Hydrology*, 548:75–87, 2017.

I. Chaubey, C. T. Haan, J. M. Salisbury, and S. Grunwald. Quantifying model output uncertainty due to spatial variability of rainfall. *Journal of the American Water Resources Association*, 35(5):1113–1123, 1999.

D. Chen, G. Gao, C.-Y. Xu, J. Guo, and G. Ren. Comparison of the Thornthwaite method and pan data with the standard Penman-Monteith estimates of reference evapotranspiration in China. *Climate Research*, 28(2):123–132, 2005.

W. Chen, Z. Jiang, and L. Li. Probabilistic Projections of Climate Change over China under the SRES A1b Scenario Using 28 AOGCMs. *Journal of Climate*, 24(17):4741–4756, 2011.

Z. Chen, A. Hartmann, T. Wagener, and N. Goldscheider. Dynamics of water fluxes and storages in an Alpine karst catchment under current and potential future climate conditions. *Hydrology and Earth System Sciences Discussions*, 2017. in review.

J. M. Cheng and C. X. Chen. An integrated linear/non-linear flow model for the conduit-fissure-pore media in the karst triple void aquifer system. *Environmental Geology*, 47 (2):163–174, 2005.

H. T. Choi and K. Beven. Multi-period and multi-criteria model conditioning to reduce prediction uncertainty in an application of TOPMODEL within the GLUE framework. *Journal of Hydrology*, 332(3-4):316–336, 2007.

R. T. Clarke. Uncertainty in the estimation of mean annual flood due to rating-curve indefinition. *Journal of Hydrology*, 222(1):185–190, 1999.

CMDC. Evapotranspiration Data, 2017. http://data.cma.cn/site/index.html, supplied by Fang Guo.

C. Dobler, G. Bürger, and J. Stötter. Assessment of climate change impacts on flood hazard potential in the Alpine Lech watershed. *Journal of Hydrology*, 460-461:29–39, 2012.

Q. Y. Duan, V. K. Gupta, and S. Sorooshian. Shuffled complex evolution approach for effective and efficient global minimization. *Journal of optimization theory and applications*, 76(3):501–521, 1993.

A. Efstratiadis and D. Koutsoyiannis. One decade of multi-objective calibration approaches in hydrological modelling: a review. *Hydrological Sciences Journal*, 55(1): 58–78, 2010.

V. Engel, E. G. Jobbágy, M. Stieglitz, M. Williams, and R. B. Jackson. Hydrological consequences of Eucalyptus afforestation in the Argentine Pampas. *Water Resources Research*, 41(10), 2005.

G. Fischer, F. Nachtergaele, S. Prieler, H. van Velthuizen, L. Verelst, and D. Wiberg. Global Agro-ecological Zones Assessment for Agriculture (GAEZ 2008). Technical report, IIASA and FAO, Laxenburg and Rome, 2008. http://www.fao.org/soils-portal/soil-survey/soil-maps-and-databases/harmonized-world-soil-database-v12/en/.

D. Ford and P. Williams. *Karst Hydrogeology and Geomorphology*. Wiley, Chichester, UK, 2007.

GADM. Global Administrative Areas, 2017. http://gadm.org/, retrieved 30.09.2017.

M. R. Gallagher and J. Doherty. Parameter interdependence and uncertainty induced by lumping in a hydrologic model. *Water Resources Research*, 43(5), 2007.

Geofabrik. OpenStreetMap Data China, 2017. http://download.geofabrik.de/asia/china.html, retrieved 23.06.2017.

A. Ghasemi and S. Zahediasl. Normality Tests for Statistical Analysis: A Guide for Non-Statisticians. *International Journal of Endocrinology and Metabolism*, 10(2):486–489, 2012.

N. Goldscheider and D. Drew. *Methods in Karst Hydrogeology*. Number 26 in International Contributions to Hydrogeology. Taylor & Francis Group, London, 2007.

B. R. Gondwe, G. Merediz-Alonso, and P. Bauer-Gottwein. The influence of conceptual model uncertainty on management decisions for a groundwater-dependent ecosystem in karst. *Journal of Hydrology*, 400(1-2):24–40, 2011.

M. Goswami and K. M. O'Connor. A "monster" that made the SMAR conceptual model "right for the wrong reasons". *Hydrological Sciences Journal*, 55(6):913–927, 2010.

S. Gu, A. Jenkins, S.-J. Gao, Y. Lu, H. Li, Y. Li, R. C. Ferrier, M. Bailey, Y. Wang, Y. Zhang, X. Qi, L. Yu, L. Ding, T. Daniell, R. Williams, J. Hannaford, M. Acreman, S. Kirk, H. Liu, Z. Liu, L. Luo, D. Yan, X. Liu, F. Yu, D. Wang, B. Zhang, A. Ding, X. Xie, J. Liu, C. Ma, and A. Jobson. Ensuring water resource security in China; the need for advances in evidence-based policy to support sustainable management. *Environmental Science & Policy*, 75:65–69, 2017.

F. Guo and G. Jiang. Karst Groundwater Management through Science and Education. *Open Journal of Geology*, 1(3):45–50, 2011.

F. Guo, D. Yuan, and Z. Qin. Groundwater Contamination in Karst Areas of Southwestern China and Recommended Countermeasures. *Acta Carsologica*, 39(2):389–399, 2010.

F. Guo, G. Jiang, D. Yuan, and J. S. Polk. Evolution of major environmental geological problems in karst areas of Southwestern China. *Environmental Earth Sciences*, 69(7): 2427–2435, 2013.

F. Guo, G. Jiang, J. S. Polk, X. Huang, and S. Huang. Resilience of Groundwater Impacted by Land Use and Climate Change in a Karst Aquifer, South China. *Water Environment Research*, 87(11):1990–1998, 2015.

F. Guo, W. Wang, G. Jiang, and S. Huang. Distribution and stable isotopic compositions of organic carbon in surface sediments in hyporheic zone of karst springs. *Environmental Earth Sciences*, 75:850, 2016.

H. V. Gupta, S. Sorooshian, and P. O. Yapo. Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information. *Water Resources Research*, 34(4):751–763, 1998.

H. V. Gupta, H. Kling, K. K. Yilmaz, and G. F. Martinez. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, 377(1-2):80–91, 2009.

Y. Hao, Y. Zhu, Y. Zhao, W. Wang, X. Du, and T.-C. J. Yeh. The Role of Climate and Human Influences in the Dry-Up of the Jinci Springs, China. *Journal of the American Water Resources Association*, 45(5):1228–1237, 2009.

A. Hartmann, M. Kralik, F. Humer, J. Lange, and M. Weiler. Identification of a karst system's intrinsic hydrodynamic parameters: upscaling from single springs to the whole aquifer. *Environmental Earth Sciences*, 65(8):2377–2389, 2012a.

A. Hartmann, J. Lange, N. Mizyed, G. Smiatek, H. Kunstmann, and A. Vivó Aguado. A multi-model approach for improved simulations of future water availability at a large Eastern Mediterranean karst spring. *Journal of Hydrology*, 468-469:130–138, 2012b.

A. Hartmann, J. A. Barberá, J. Lange, B. Andreo, and M. Weiler. Progress in the hydrologic simulation of time variant recharge areas of karst systems – Exemplified at a karst spring in Southern Spain. *Advances in Water Resources*, 54:149–160, 2013a.

A. Hartmann, T. Wagener, A. Rimmer, J. Lange, H. Brielmann, and M. Weiler. Testing the realism of model structures to identify karst system processes using water quality and quantity signatures:. *Water Resources Research*, 49(6):3345–3358, 2013b.

A. Hartmann, M. Weiler, T. Wagener, J. Lange, M. Kralik, F. Humer, N. Mizyed, A. Rimmer, J. A. Barber, B. Andreo, C. Butscher, and P. Huggenberger. Process-based karst modelling to relate hydrodynamic and hydrochemical characteristics to system properties. *Hydrology and Earth System Sciences*, 17(8):3305–3321, 2013c.

A. Hartmann, N. Goldscheider, T. Wagener, J. Lange, and M. Weiler. Karst water resources in a changing world: Review of hydrological modeling approaches. *Reviews of Geophysics*, 52(3):218–242, 2014a.

A. Hartmann, M. Mudarra, B. Andreo, A. Marín, T. Wagener, and J. Lange. Modeling spatiotemporal impacts of hydroclimatic extremes on groundwater recharge at a Mediterranean karst aquifer. *Water Resources Research*, 50(8):6507–6521, 2014b.

A. Hartmann, T. Gleeson, R. Rosolem, F. Pianosi, Y. Wada, and T. Wagener. A large-scale simulation model to assess karstic groundwater recharge over Europe and the Mediterranean. *Geoscientific Model Development*, 8(6):1729–1746, 2015.

A. Hartmann, J. Kobler, M. Kralik, T. Dirnböck, F. Humer, and M. Weiler. Model-aided quantification of dissolved carbon and nitrogen release after windthrow disturbance in an Austrian karst system. *Biogeosciences*, 13(1):159–174, 2016.

A. Hartmann, J. A. Barberá, and B. Andreo. On the value of water quality data and informative flow states in karst modelling. *Hydrology and Earth System Sciences*, 21: 5971–5985, 2017.

A. Henderson-Sellers. An antipodean climate of uncertainty? *Climatic Change*, 25(3): 203–224, 1993.

G. M. Hornberger and R. C. Spear. An approach to the Preliminary Analysis of Environmental Systems. *Journal of Environmental Management*, 12:7–18, 1981.

K. Hu, H. Chen, Y. Nie, and K. Wang. Seasonal recharge and mean residence times of soil and epikarst water in a small karst catchment of southwest China. *Scientific Reports*, 5 (1), 2015.

L. Huang, G. Zeng, J. Liang, S. Hua, Y. Yuan, X. Li, H. Dong, J. Liu, S. Nie, and J. Liu. Combined Impacts of Land Use and Climate Change in the Modeling of Future Groundwater Vulnerability. *Journal of Hydrologic Engineering*, 22(7), 2017.

A. J. Jakeman and G. M. Hornberger. How Much Complexity Is Warranted in a Rainfall-Runoff Model? *Water Resources Research*, 29(8):2637–2649, 1993.

G. Jiang and F. Guo. Interpreting Source of Lingshui Spring by Hydrogeological, Chemical and Isotopic Methods. In F. Carrasco, editor, *Advances in Research in Karst Media*, Environmental Earth Sciences, pages 177–181. Springer, 2010.

A. L. Kay, H. N. Davies, V. A. Bell, and R. G. Jones. Comparison of uncertainty sources for climate change impacts: flood frequency in England. *Climatic Change*, 92(1-2): 41–63, 2009.

L. Kiraly. Modelling karst aquifers by the combined discrete channel and continuum approach. *Bulletin d'Hydrogéologie*, 16:77–98, 1998.

J. W. Kirchner. Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology. *Water Resources Research*, 42(3), 2006.

G. Kuczera and M. Mroczkowski. Assessment of hydrologic parameter uncertainty and the worth of multiresponse data. *Water Resources Research*, 34(6):1481–1489, 1998.

X. Li, L. Shu, L. Liu, D. Yin, and J. Wen. Sensitivity analysis of groundwater level in Jinci Spring Basin (China) based on artificial neural network modeling. *Hydrogeology Journal*, 20(4):727–738, 2012.

Y. Lian, G. J.-Y. You, K. Lin, Z. Jiang, C. Zhang, and X. Qin. Characteristics of climate change in southwest China karst region and their potential environmental impacts. *Environmental Earth Sciences*, 74(2):937–944, 2015.

M. Liu, X. Xu, A. Y. Sun, K. Wang, W. Liu, and X. Zhang. Is southwestern China experiencing more frequent precipitation extremes? *Environmental Research Letters*, 9(6), 2014.

N. Mazzilli, V. Guinot, and H. Jourde. Sensitivity analysis of conceptual model calibration to initialisation bias. Application to karst spring discharge models. *Advances in Water Resources*, 42:1–16, 2012.

R. Merz, J. Parajka, and G. Blöschl. Time stability of catchment model parameters: Implications for climate impact analyses. *Water Resources Research*, 47(2), 2011.

R. J. Moore. The PDM rainfall-runoff model. *Hydrology and Earth System Sciences Discussions*, 11(1):483–499, 2007.

F. Moussu, L. Oudin, V. Plagnes, A. Mangin, and H. Bendjoudi. A multi-objective calibration framework for rainfall–discharge models applied to karst systems. *Journal of Hydrology*, 400(3-4):364–376, 2011.

NOAA. Global Historical Climatological Network (GHCN-Daily), 2017. https://www.ncdc.noaa.gov/cdo-web/search, retrieved 02.11.2017.

G. Nützmann and H. Moser. *Elemente einer analytischen Hydrologie: Prozesse - Wechselwirkungen - Modelle*. Springer Spektrum, Wiesbaden, 2016.

OpenStreetMap. Map of China: Stamen Watercolor, 2017. http://maps.stamen.com/watercolor/, retrieved 30.09.2017.

C. Perrin, C. Michel, and V. Andréassian. Does a large number of parameters enhance model performance? Comparative assessment of common catchment model structures on 429 catchments. *Journal of Hydrology*, 242(3):275–301, 2001.

C. Perrin, C. Michel, and V. Andréassian. Improvement of a parsimonious model for streamflow simulation. *Journal of Hydrology*, 279(1-4):275–289, 2003.

A. Poulin, F. Brissette, R. Leconte, R. Arsenault, and J.-S. Malo. Uncertainty of hydrological modelling in climate change impact studies in a Canadian, snow-dominated river basin. *Journal of Hydrology*, 409(3-4):626–636, 2011.

S. Praskievicz and H. Chang. A review of hydrological modelling of basin-scale climate change and urban development impacts. *Progress in Physical Geography*, 33(5):650–671, 2009.

G. Ren, Y. Ding, Z. Zhao, J. Zheng, T. Wu, G. Tang, and Y. Xu. Recent progress in studies of climate change in China. *Advances in Atmospheric Sciences*, 29(5):958–977, 2012.

B. Renard, D. Kavetski, G. Kuczera, M. Thyer, and S. W. Franks. Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors. *Water Resources Research*, 46(5), 2010.

A. Rimmer and A. Hartmann. Simplified conceptual structures and analytical solutions for groundwater discharge using reservoir equations. In *Water Resources Management and Modeling*. InTech, 2012.

A. Rimmer and A. Hartmann. Optimal hydrograph separation filter to evaluate transport routines of hydrological models. *Journal of Hydrology*, 514:249–257, 2014.

A. Rohatgi. WebPlotDigitizer, 2017. https://automeris.io/WebPlotDigitizer/, last accessed 03.11.2017.

F. Sarrazin, A. Hartmann, F. Pianosi, and T. Wagener. To what extent do preferential flow pathways alter the impact of climate and land cover change on groundwater recharge? *AGU Fall Meeting Abstracts*, H31C-1374, 2016. http://adsabs.harvard.edu/abs/2016AGUFM.H31C1374S.

M. Sauter, T. Geyer, A. Kovács, and G. Teutsch. Modellierung der Hydraulik von Karstgrundwasserleitern-Eine Übersicht. *Grundwasser*, 11(3):143–156, 2006.

B. R. Scanlon, R. E. Mace, M. E. Barrett, and B. Smith. Can we simulate regional groundwater flow in a karst system using equivalent porous media models? Case study, Barton Springs Edwards aquifer, USA. *Journal of Hydrology*, 276(1-4):137–158, 2003.

J. Seibert and J. J. McDonnell. On the dialog between experimentalist and modeler in catchment hydrology: Use of soft data for multicriteria model calibration. *Water Resources Research*, 38(11):23–1–23–14, 2002.

A. Semadeni-Davies, C. Hernebring, G. Svensson, and L.-G. Gustafsson. The impacts of climate change and urbanisation on drainage in Helsingborg, Sweden: Combined sewer system. *Journal of Hydrology*, 350(1-2):100–113, 2008.

S. Steinschneider, A. Polebitski, C. Brown, and B. H. Letcher. Toward a statistical framework to quantify the uncertainties of hydrologic response under climate change. *Water Resources Research*, 48(11), 2012.

T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, B. Bex, and B. M. Midgley. *IPCC, 2013: climate change 2013: the physical science basis. Contribution of working group I to the fifth assessment report of the intergovernmental panel on climate change*. Cambridge University Press, 2013.

M. M. Sweeting. *Karst in China - Its Geomorphology and Environment*. Springer, 1995.

G. Q. Tabios and J. D. Salas. A comparative analysis of techniques for spatial interpolation of precipitation. *JAWRA Journal of the American Water Resources Association*, 21(3): 365–380, 1985.

G. Teutsch. An extended double-porosity concept as a practical modelling approach for a karstified terrain. *IAHS Publication*, 207:281–292, 1993.

A. Thomas. Spatial and Temporal Characteristics of Potential Evapotranspiration Trends over China. *International Journal of Climatology*, 20:381–396, 2000.

C. W. Thornthwaite. An Approach toward a Rational Classification of Climate. *Geographical Review*, 38(1):55–94, 1948.

D. Tian, Y. Guo, and W. Dong. Future changes and uncertainties in temperature and precipitation over China based on CMIP5 models. *Advances in Atmospheric Sciences*, 32(4):487–496, 2015.

USGS. Shuttle Radar Topography Mission - Void filled (3 arc-seconds), 2017. https://lta.cr.usgs.gov/SRTMVF, retrieved 30.09.2017.

J. A. Velázquez, J. Schmid, S. Ricard, M. J. Muerth, B. Gauvin St-Denis, M. Minville, D. Chaumont, D. Caya, R. Ludwig, and R. Turcotte. An ensemble approach to assess hydrological models' contribution to uncertainties in the analysis of climate change impact on water resources. *Hydrology and Earth System Sciences*, 17(2):565–578, 2013.

J. A. Vrugt, H. V. Gupta, W. Bouten, and S. Sorooshian. A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. *Water Resources Research*, 39(8), 2003.

T. Wagener and J. Kollat. Numerical and visual evaluation of hydrological and environmental models using the Monte Carlo analysis toolbox. *Environmental Modelling & Software*, 22(7):1021–1033, 2007.

T. Wagener, D. P. Boyle, M. J. Lees, H. S. Wheater, H. V. Gupta, and S. Sorooshian. A framework for development and application of hydrological models. *Hydrology and Earth System Sciences Discussions*, 5(1):13–26, 2001.

H. Walter and H. Lieth. *Klimadiagramm-Weltaltlas*. Gustav Fischer Verlag, Jena, 1967.

B. L. Welch. The generalization of "Student's" problem when several different population variances are involved. *Biometrika*, 34(1-2):28–35, 1947.

W. B. White. Karst hydrology: recent developments and open questions. *Engineering geology*, 65(2):85–105, 2002.

R. L. Wilby. Uncertainty in water resource model parameters used for climate change impact assessment. *Hydrological Processes*, 19(16):3201–3219, 2005.

R. L. Wilby and I. Harris. A framework for assessing uncertainties in climate change impacts: Low-flow scenarios for the River Thames, UK. *Water Resources Research*, 42(2), 2006.

H. C. Winsemius, B. Schaefli, A. Montanari, and H. H. G. Savenije. On the calibration of hydrological models in ungauged basins: A framework for integrating hard and soft hydrological information. *Water Resources Research*, 45(12), 2009.

S. R. Worthington. A comprehensive strategy for understanding flow in carbonate aquifer. *Speleogenesis and Evolution of Karst Aquifers*, 1(1):1–8, 2003.

G. Xuejie, Z. Zongci, D. Yihui, H. Ronghui, and F. Giorgi. Climate change due to greenhouse effects in China as simulated by a regional climate model. *Advances in Atmospheric Sciences*, 18(6):1224–1230, 2001.

P. O. Yapo, H. V. Gupta, and S. Sorooshian. Automatic calibration of conceptual rainfall-runoff models: sensitivity to calibration data. *Journal of Hydrology*, 181(1-4):23–48, 1996.

W. Ye, B. C. Bates, N. R. Viney, M. Sivapalan, and A. J. Jakeman. Performance of conceptual rainfall-runoff models in low-yielding ephemeral catchments. *Water Resources Research*, 33(1):153–166, 1997.

L. Zhang, J. Xiao, J. Li, K. Wang, L. Lei, and H. Guo. The 2010 spring drought reduced primary productivity in southwestern China. *Environmental Research Letters*, 7(4), 2012.

Z. Zhang, Z. Zhu, and Z. Wen. Problems and Solutions about Land Use of Different Karst Environments in Southern China. *Asian Agricultural Research*, 5(10):51–55;60, 2013.

Appendix

A Discharge data

Table A.1: *Discharge data measured in the Lingshui spring pool*

Date	Q [l·s^{-1}]	Water level [m]
29.03.10	1860.00	0.52
01.02.11	2500.73	0.58
01.03.11	3469.78	0.56
01.04.11	3123.58	0.61
02.05.11	3046.71	0.65
02.06.11	3937.73	NA
03.07.11	4247.08	NA
03.08.11	3428.92	0.65
03.09.11	2381.66	0.61
04.10.11	5143.70	NA
04.11.11	3968.97	0.74
05.12.11	2724.94	0.64
25.06.15	3547.30	NA
25.08.15	3384.70	NA
14.09.15	4283.80	NA
19.04.16	2758.90	0.52
29.06.16	5625.30	0.86
12.10.16	3791.30	0.58
16.01.17	2389.50	0.55
09.05.17	2900.00	0.60
28.08.17	4590.00	NA

B Thornthwaite equation

$$ETp_{THORN} = \left(\frac{0.533 \cdot L}{12}\right) \cdot \left(10 \cdot \frac{T}{I}\right)^{\alpha} \tag{21}$$

where: ETp_{THORN} = Potential Evapotranspiration after Thornthwaite [mm]
L = Mean day length [h]
T = Daily mean of temperature [°C]
γ = Equation parameter [-]

γ is being calculated through equation 22.

$$gamma = (6.75 \cdot 10^{-7})I^3 - (7.71 \cdot 10^{-5})I^2 + (1.792 \cdot 10^{-2})I + 0.49239 \tag{22}$$

where: I = Warmth index [-]

I [-] is the warmth index of monthly temperature means (equation 23).

$$I = \sum_{i=1}^{12} \left(\frac{T_{mon}}{5}\right)^{1.514} \tag{23}$$

where: T_{mon} = Monthly mean of temperature [°C]

C Parameter ranges of first calibration trial

Table C.1: *Description of parameters and parameter ranges used for first calibration trial (parameter ranges adapted from (Brenner et al., 2016)).*

Parameter	Description	Unit	Ranges	
			Lower	Upper
A	Recharge area	km^2	650	750
$V_{mean,S}$	Mean soil storage capacity	mm	5000	15000
$V_{mean,E}$	Mean epikarst storage capacity	mm	0	5000
a_{SE}	Soil/epikarst depth variability constant	-	0.1	6
$K_{mean,E}$	Epikarst mean storage coefficient	d	0.1	2.5
a_{fsep}	Recharge separation variability constant	-	0.1	5
K_C	Conduit storage coefficient	d	1	100
a_{GW}	Groundwater variability constant	-	1	10
a	Rating curve parameter	-	0	5
b	Rating curve parameter	-	0	5
c	Rating curve parameter	-	0	10000

D Groundwater stationarity - RNI

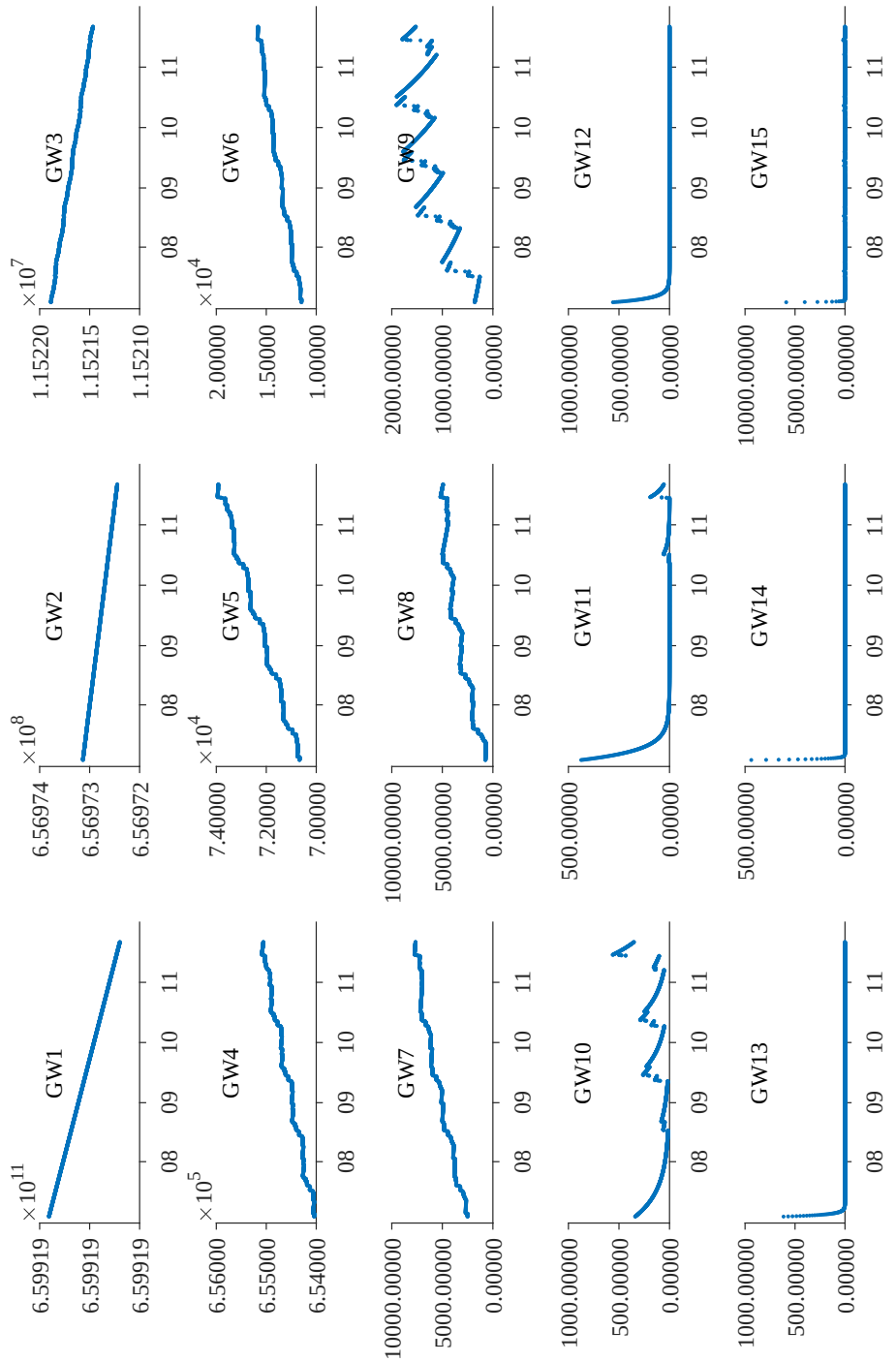


Figure D.1: *Groundwater compartments (RNI-EpiMax-First calibration).*

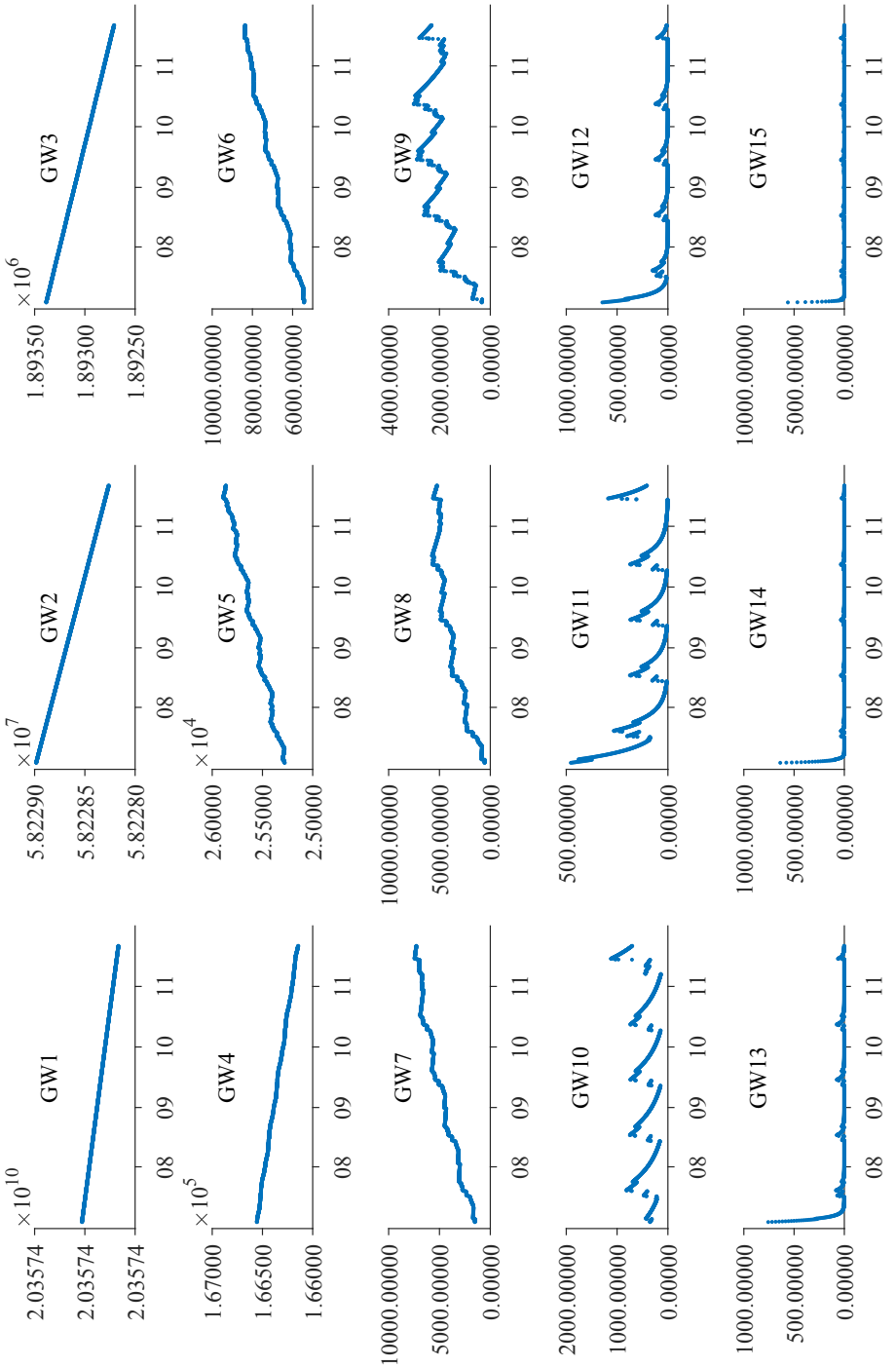


Figure D.2: *Groundwater compartments (RNI - SoiMaxEpiMax - First calibration).*

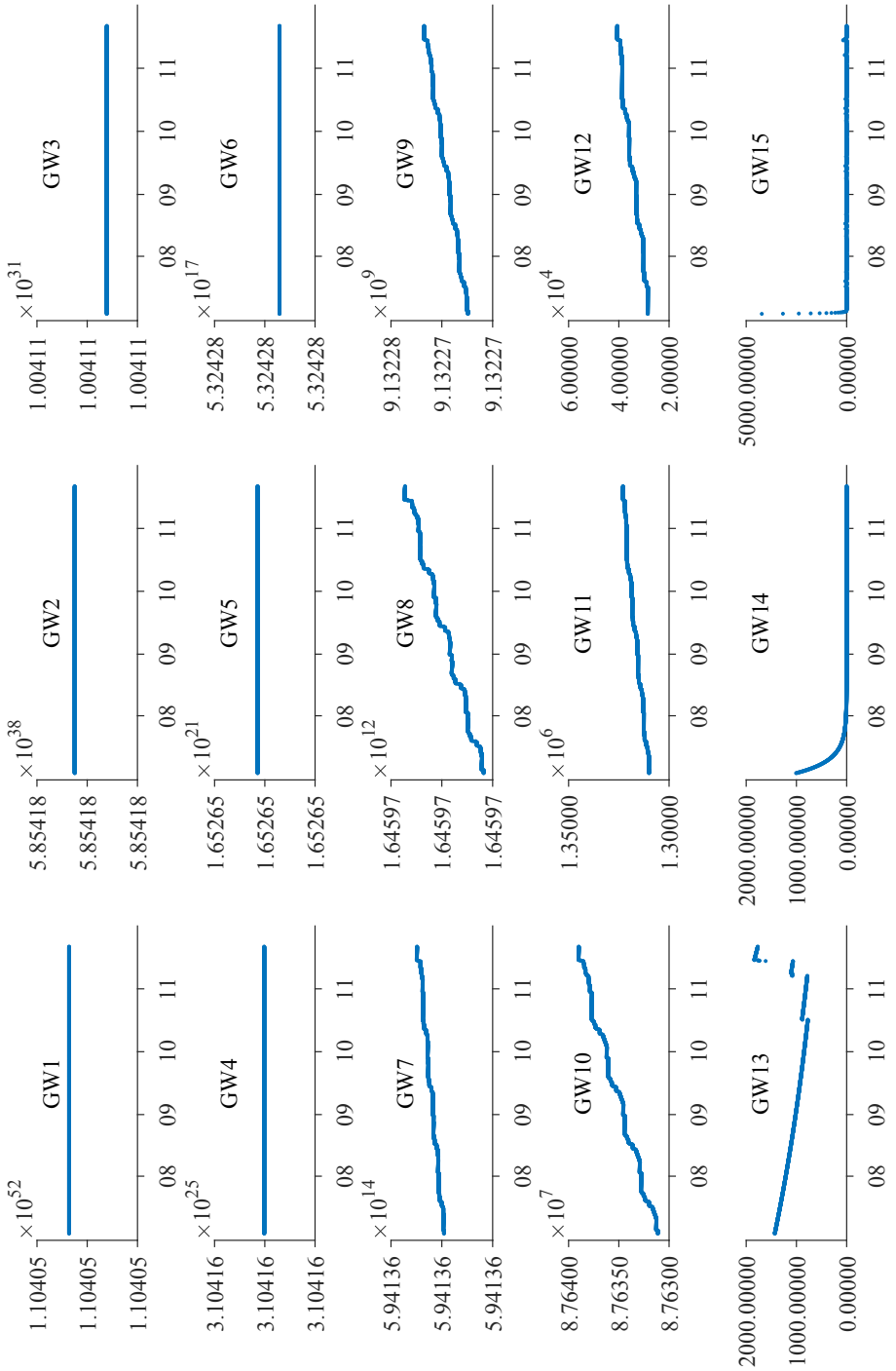


Figure D.3: *Groundwater compartments (RNI - EpiMax - Final calibration).*

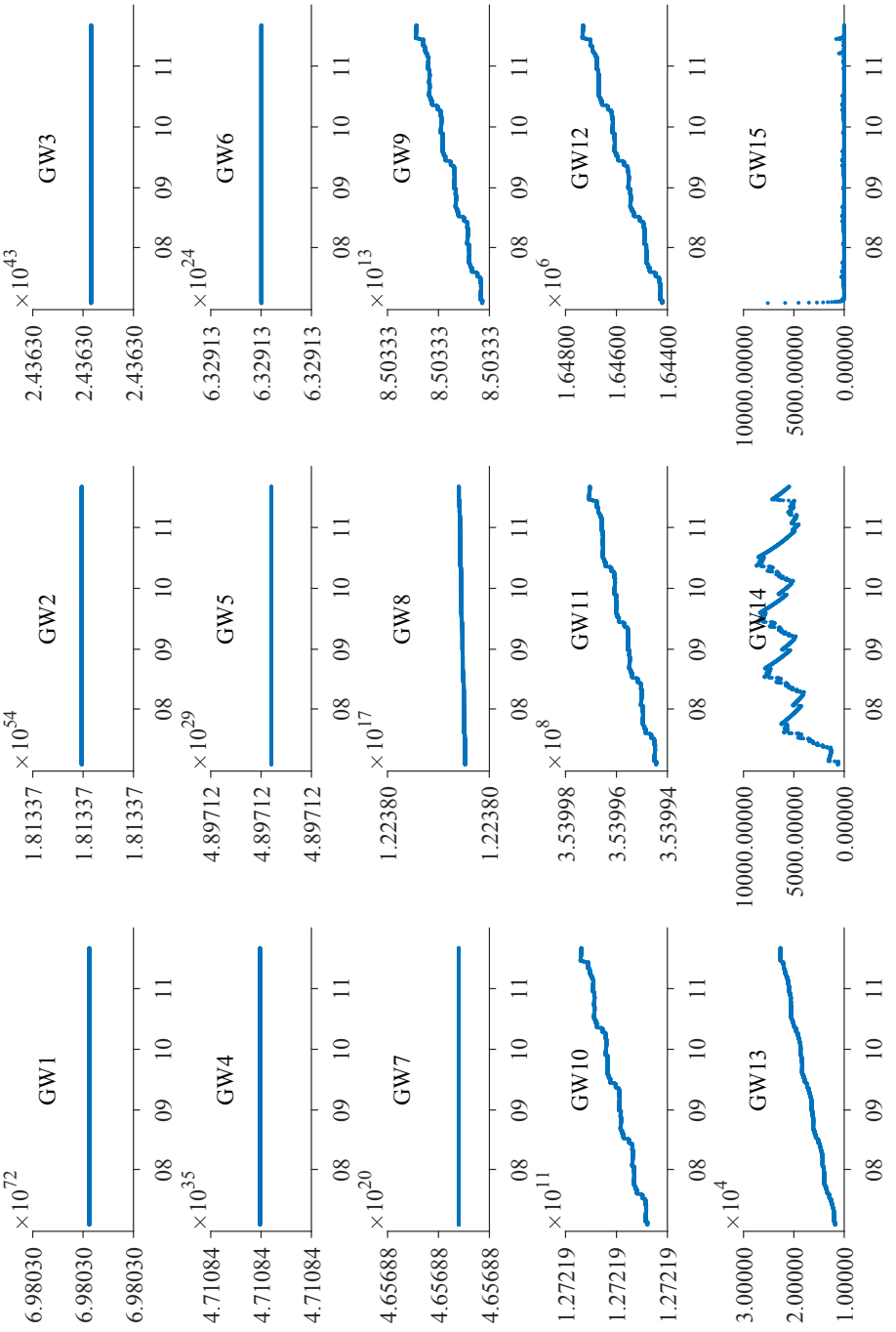


Figure D.4: *Groundwater compartments (RNI - SoiMaxEpiMax - Final calibration).*

E Groundwater stationarity - RI

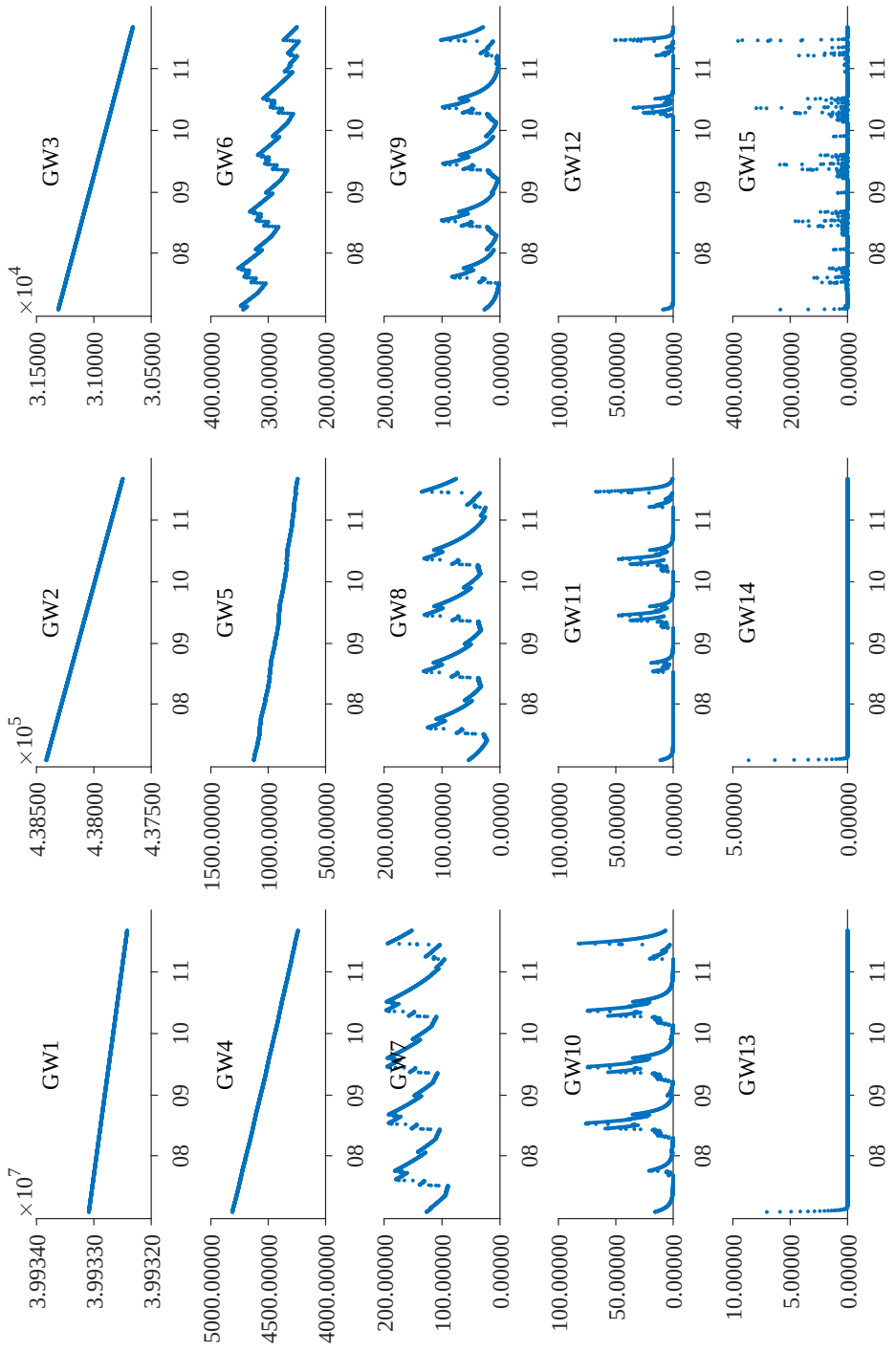


Figure E.1: *Groundwater compartments (RI - EpiMax - First calibration)*

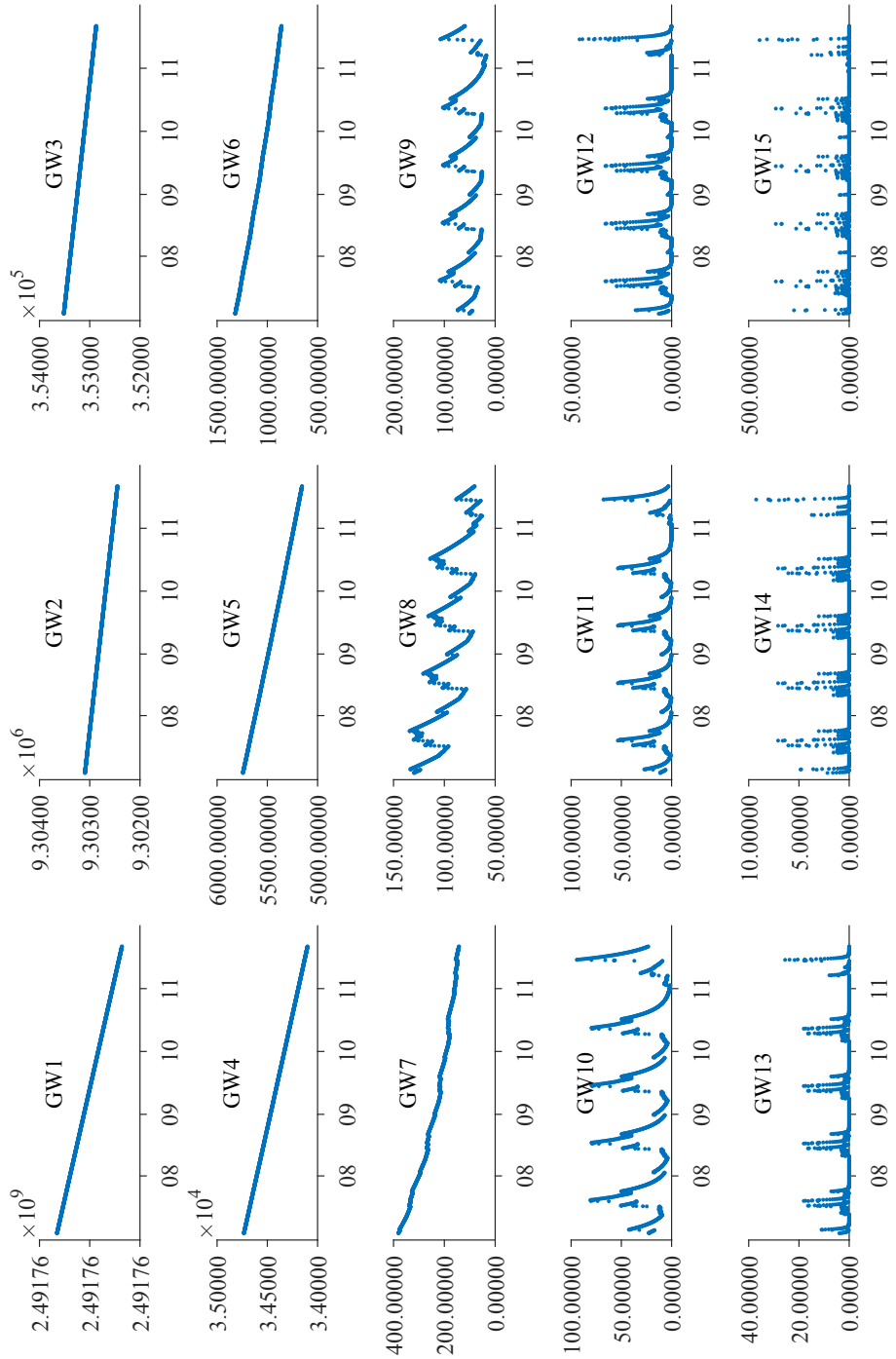


Figure E.2: *Groundwater compartments (RI - SoiMaxEpiMax - First calibration).*

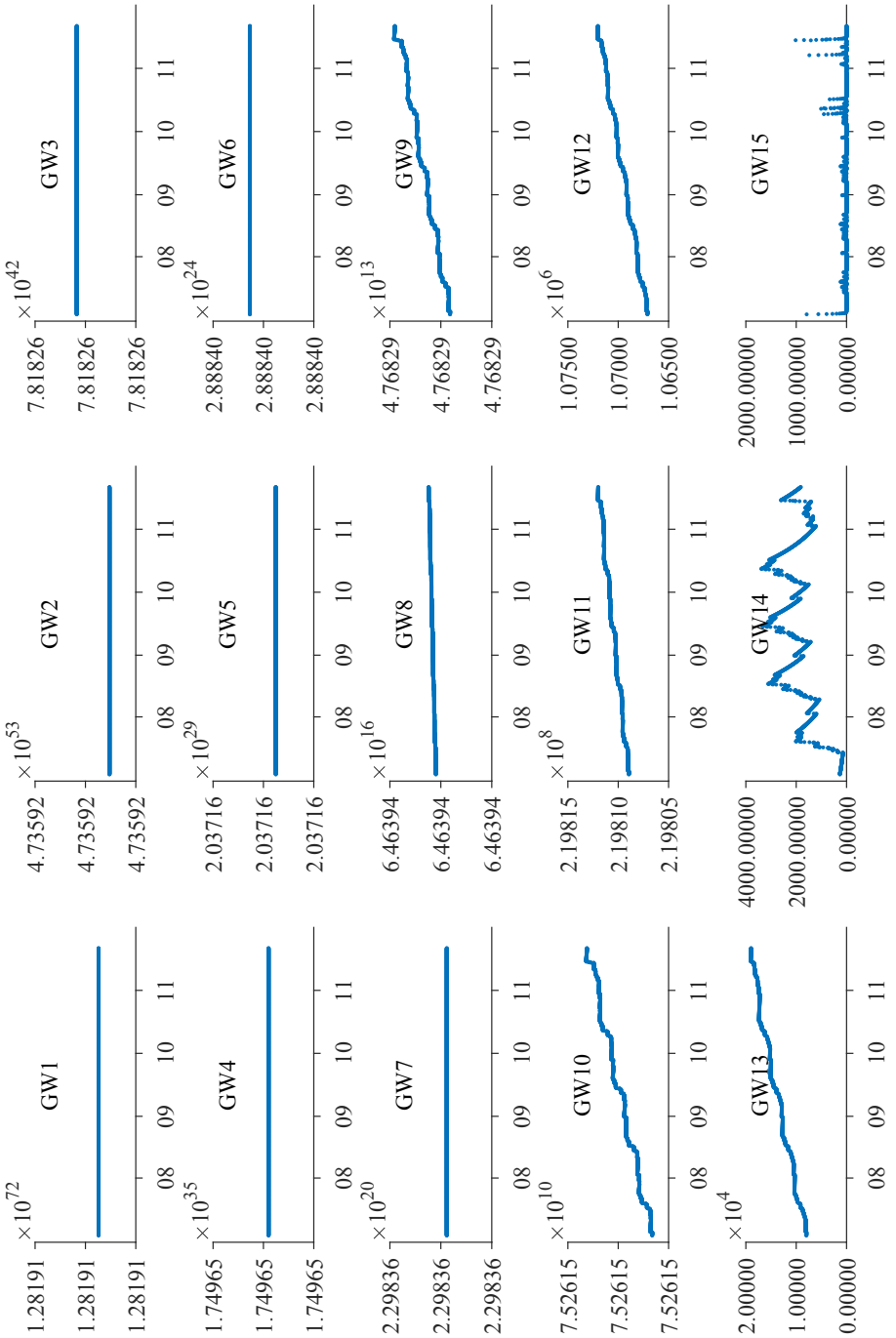


Figure E.3: *Groundwater compartments (RI - EpiMax - Final calibration).*

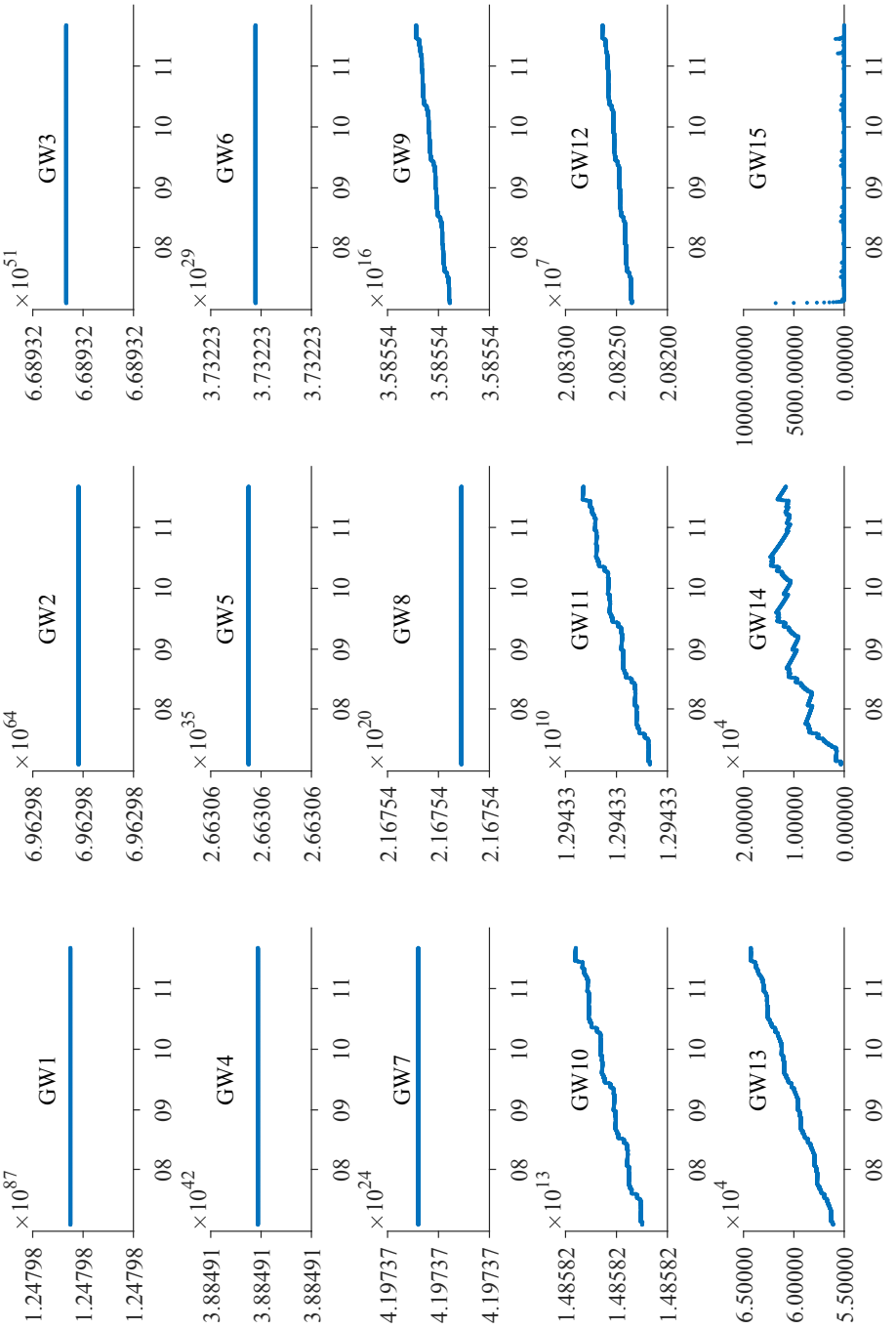


Figure E.4: *Groundwater compartments (RI - SoiMaxEpiMax - Final calibration).*

F Calibration results of EpiMax

Table F.1: *Results for parameters and objective functions for different calibration steps and rating curve approaches for EpiMax*

Parameter	Unit	RNI		RI	
		Calibration step			
		1st	Final	1st	Final
A	km^2	666.45	692.57	707.77	721.91
$V_{mean,S}$	mm	10117.74	14773.04	14290.63	5134.87
$V_{mean,E}$	mm	2342.52	2357.77	82.1	539.09
$K_{mean,E}$	d	0.11	0.73	0.23	0.15
K_C	d	3.17	4.02	2.41	3.42
a_{SE}	-	6	12.21	5.21	20.56
a_{fsep}	-	3.19	10.42	0.65	18.54
a_{GW}	-	9.97	44.1	6.51	61.23
a	-	-	-	0.49	0.79
b	-	-	-	3.95	3
c	-	-	-	1412.55	2711.94
KGE$_{rvar}$	-	0.72	0.76	-	-
KGE	-	-	-	0.77	0.78

G Parameter distributions of calibration trials (EpiMax)

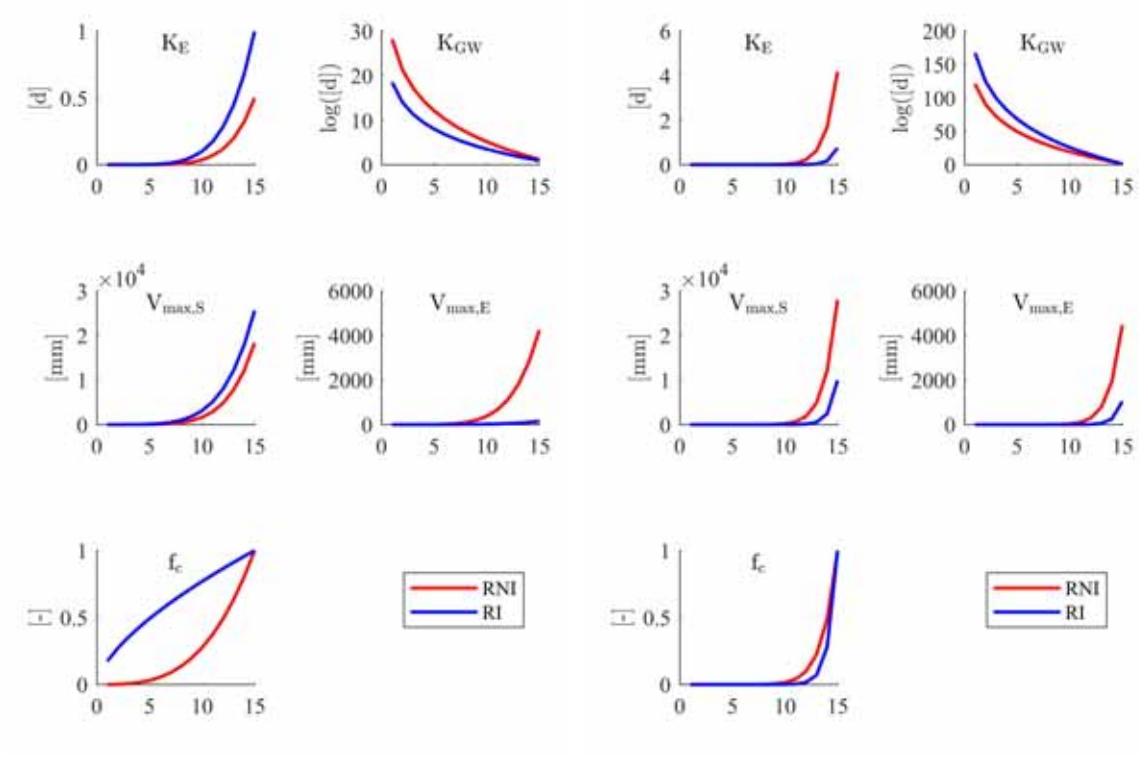


Figure G.1: *Distribution of variable parameters resulting from RNI and RI for (a) the first calibration run and (b) the final calibration run (EpiMax).*

H Uncorrected simulations of RNI

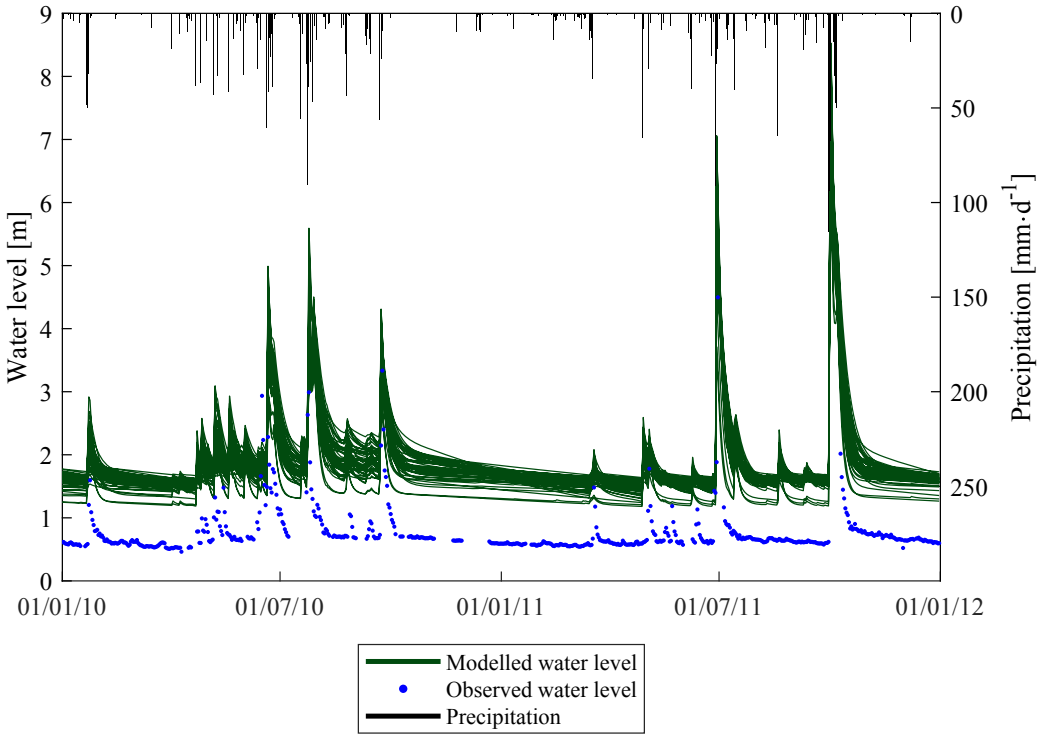


Figure H.1: *Modelled water level from 50 selected parameter sets compared to measured water level over the calibration period (2010 - 2011) produced by RNI and* ***before*** *correction.*

I Multi-objective calibration (Discharge - Water level)

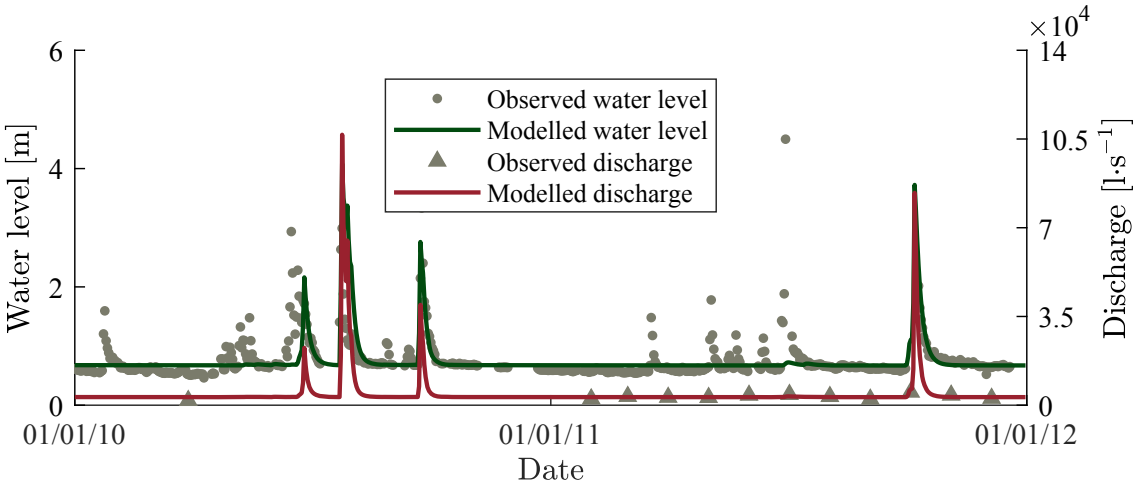


Figure I.1: *Comparison of modelled and observed values of water level and discharge over the calibration time period for a multi-objective calibration trial with RI (KGE weighting 0.5/0.5).*

Ehrenwörtliche Erklärung

Hiermit versichere ich, die vorliegende Arbeit selbstständig verfasst zu haben. Ich habe keine anderen als die angegebenen Quellen und Hilfsmittel benutzt und alle wörtlich oder sinngemäß aus anderen Werken übernommenen Inhalte als solche kenntlich gemacht. Die eingereichte Masterarbeit war oder ist weder vollständig noch in wesentlichen Teilen Gegenstandeines anderen Prüfungsverfahrens. Die elektronische Version der eingereichten Masterarbeit stimmt in Inhalt und Formatierung mit den auf Papier ausgeduckten Exemplaren überein.

Freiburg im Breisgau, 11. Dezember 2017 Laura-Marie Vecera