University of Zurich[UZH]

Spatial changes, temporal dynamics, and their influences on hydrological signals in an urban stream: A case study of the Wuhle, Berlin.

GEO 511 Master's Thesis

Author
Anna Czerniejewska
16-718-686

Supervised by
Prof. Dr. Jan Seibert
Prof. Dr. Dörthe Tetzlaff (doerthe.tetzlaff@igb-berlin.de)

Faculty representative
Dr. Ilja van Meerveld

31.01.2024
Department of Geography, University of Zurich

Content

Figures

Tables

Symbols and abbreviations

Symbol	Explanation	Unit
Catchments		
W4.0	Discharge Station at the Wuhle, lower catchment	
W4.3	Discharge Station at the Wuhle, middle catchment	
W4.4	Discharge Station at the Neue Wuhle, upper catchment	
HG	Discharge Station at the Hellensdorfer Graben, upper catchment	
HBV-Model		
TT	Threshold temperature	°C
CFMAX	Degree-day factor	mm $\Delta t^{-1}°C^{-1}$
SP	Seasonal variability in degree-Δt factor	-
SFCF	Snowfall correction factor	-
CFR	Refreezing coefficient	-
CWH	Water holding capacity	-
FC	Maximum storage in soil box	mm
LP	Threshold for reduction of evaporation	-
Beta	Shape coefficient	-
K1	Storage / Recession coefficient	Δt^{-1}
K2	Storage / Recession coefficient	Δt^{-1}
Alpha	Non-linearity coefficient	-
PERC	Maximal flow from upper to lower box	mm Δt^{-1}
MAXBAS	Routing, length of weighting function	Δt
Cet	Potential evaporation correction factor	$°C^{-1}$
PCALT	Increase of precipitation with elevation	% $(100\ m)^{-1}$
TCALT	Decrease of temperature with elevation	°C $(100\ m)^{-1}$
Pelev	Elevation of precipitation data in the input file	m
Telev	Elevation of temperature data in the input file	m
SM	Soil moisture storage	
SUZ	Soil upper zone storage	
SLZ	Soil lower zone storage	
Modelling process		
GAP	Genetic Algorithm and Powell optimization	
KGE	Kling-Gupta efficiency	
NPE	Non-parametric efficiency	
R_{eff}	Nash-Sutcliff efficiency	
r_s	Spearman rank correlation coefficient	
VE	Volumetric efficiency / Volume error	
Data		
AET	Actual evapotranspiration	mm Δt^{-1}
E_{act}	Actual evaporation	mm Δt^{-1}
E_{pot}	Potential evaporation	mm Δt^{-1}
GMC	Gravimetric soil moisture content	%
GMWL	Global meteorological water line	
Lc-excess	Line conditioned excess	
LMWL	Local meteorological water line	
PET	Potential evapotranspiration	mm Δt^{-1}
VMC	Volumetric soil moisture content	%

Symbol	Explanation	Unit
Other		
EU	European Union	
Eq.	Equation	
IGB	the Leibniz Institute of Freshwater Ecology and Inland Fisheries	
UHI	Urban heat island effect	

Acknowledgements

This great opportunity of experiencing new research areas and visiting another city for my field work portion of this thesis was made possible by the cooperation between the University of Zurich and the Humboldt-University Berlin. Thank you Dörthe Tetzlaff and Jan Seibert for making this project possible, for letting me have an insight into the academic work in both research groups and helping me structure this thesis. I hope you could enjoy your time off during your respective sabbaticals even when I had to disturb your well-earned time away, because my thesis overlapped with it.

Thank you all from the IGB for your support and inclusion in the lunch breaks and the evening program in Berlin. Special thanks to Gregorio López Moreira Mazacotte and Maria Warter for taking me with them during the field trip in autumn and the interesting conversations about hydrology and academic work. Thank you, Ann-Marie Ring and David Dubbert, for introducing me to the field work in isotopic hydrology and supporting me in the lab. And thank you Jessica Landgraf for brainstorming with me and helping me with the last step of processing my samples when I had to leave Berlin.

I thank Marc Vis and Franziska Clerc for showing me tips and trick in modelling with HBV and explaining to me the details of the model I had difficulty to understand. Thank you for showing me how to access HBV from R which sped up my calibration runs by excluding initialisation errors and the need to start each calibration individually. And thank you for the discussion about how to best analyse and visualise the results I got out of the modelling part.

I thank Ilja van Meerveld for supporting me during the absence of both of my supervisors, for reminding me to take deep breaths and solve one problem at a time. And thank you for asking from time to time how my thesis and me are doing.

I especially want to thank Bogdan, Dorota, and Jared for reading through my thesis, giving me feedback, and helping me improve it. I thank Bogdan for his love, for being such a positive force in my life and believing in me, especially in moments where I doubted myself. I thank my parents, Artur and Dorota, for supporting me during my years in university for enabling to find my voice and go my way in life. And I thank my brother Olaf, to whom I can always turn to for discussions about life and working through stressful situations.

Abstract

Urbanisation affects the environment and hydrology. It influences temperature and precipitation locally, an effect named urban heat island. With a warming climate, implementing new and maintaining existing cooling strategies in urban areas is crucial for a viable city environment. Urban green spaces provide ecosystem services including physical and mental well-being to its citizens, ecological diversity, cooling, and flood mitigation. Urban streams are heavily influenced by anthropogenic activity and have complex interactions with urban water systems' technical and natural components. For better understanding of urban catchments stable isotopes (Deuterium (D) and oxygen (^{18}O)) can be used as natural tracers. In recent years, more studies examined ecohydrological partitioning at spatially distributed sites or between different vegetation covers. One part of this thesis focused on grassland coupled with dynamics in shallow soil. Another method to gain insight into urban catchments is to model the discharge. Hydrological modelling is dependent on catchment information and available observation data. For the simulation hydrological processes, complex models are not always more accurate. A second part of this thesis was modelling an urban stream with a conceptual rainfall-runoff model, HBV. The overarching question in this thesis was how the hydrological signals change throughout an urban catchment. The research to this thesis was conducted in Berlin in the Wuhle catchment.

Autumn 2023 was dry and slightly warmer compared to the long-term trend. In open water measurements clear distinctions between flowing water and water taken from connected ponds were visible. Water and soil samples show evaporative fractioning. Differences in volumetric water contents and isotopic composition in depth could already be detected between the upper 5 cm and the samples from 5-15 cm in shallow soil. Site specific differences are present. A trend of isotopic differences linked to soil composition was stronger than tree presence and nearby degree of impervious surfaces.

In the modelling process, simulating peaks during and after precipitation events, was challenging due to the presence of direct runoff from sealed surfaces. Calibrated parameter values show statistical differences between the agricultural subcatchment and the whole semi-urban catchment. All modelled (sub)catchments had calibrated similar short discharge delays. Significant differences were modelled in threshold temperatures, soil moisture storage and percolation.

Signal changes in upper and lower catchment could be measured by stable isotopes and modelled by a simple conceptual model. The exact cause of those changes could not be derived from the available information in this thesis. Further research on environmental influences is needed. For better discharge simulations more trials within the existing model structure are needed. Alternatively, an expansion to model's direct runoff from sealed surfaces could improve the simulated discharge.

***Keywords**: urban hydrology, stable isotopes, natural tracer, urban green spaces, HBV, rainfall-runoff modelling, conceptual model*

1. Introduction

1.1. On urbanisation and urban hydrology

Over half of the world's population lives in urban areas; by 2050, 86 percent will presumably become urban (United Nations, 2019). In the European Union (EU), about 75 percent of the population lives in urban areas which is likely to grow in the future (European Commission, 2017). It is important to note that the term urbanisation is broad and includes transformation of rural areas into urban environments. Further, it encompasses socio-economic processes such as people moving from rural areas into cities and changes in social dynamics that transform urban and rural areas' demographic and social structure (United Nations, 2019). Urban area and city are terms that are often used interchangeably. Additionally, the definition of a city is specific to each country. In this thesis the exact definition was not relevant, as there was only one study location, and the thesis focussed on the physical aspects of urbanisation. The later described fieldwork was carried out in Berlin, Germany.

Urban areas have a different climate compared to its surrounding area. Urbanisation influences temperatures locally, with cities being warmer than their peripheral areas, known as the urban heat island effect (UHI) (Bowler et al., 2010; IPCC, 2021). Changes in surface albedo, thermal conductivity and specific heat capacity of materials, building geometry, and anthropogenic heat create changes in solar radiation absorption and reflection, causing UHI (Bowler et al., 2010; Souch & Grimmond, 2006). The most significant of those is the removal of vegetation, which is replaced by buildings or sealed surfaces. Vegetation reduces the latent heat flux, the energy needed for evaporation of water, and increases sensible heat flux, energy that causes the change of temperature within mass (Oswald et al., 2023). Urbanization also affects precipitation with increased mean and heavy precipitation over and/or downwind of cities (IPCC, 2021). In addition to the extent of the UHI effect, the review of Oswald et al. (2023) summarises the following factors to moderate precipitation in cities: Changes to aerodynamic roughness, increased wind speed and the extent of aerosol and pollutant loading. Combined with impervious areas, it results in more surface runoff with less lag time, leading to higher runoff intensities (IPCC, 2021; Shuster et al., 2005). Green spaces can mitigate pluvial floods by increasing catchment capacity for infiltrating, storing and releasing water (Golden & Hoghooghi, 2018). They are beneficial to cities and provide multiple ecosystem services. Green spaces have a cooling effect through evapotranspiration and lead to better air quality by reduced exposure to air pollutants, reduces noise and excessive heat, which affect the physical and mental well-being of citizens (Bowler et al., 2010; WHO, 2016; Pukowiec-Kurda, 2022).

Urban hydrology is very complex. Urban areas interact, influence and are affected by streams and, generally, the hydrological cycle. In their paper, Albert (2021) estimated that one third of the global freshwater passes through environments modified by humans for agricultural and industrial uses or urban infrastructure. In densely populated areas in Eurasia, more than 50 percent of rivers are diverted for human activities. The tight entanglement of urban water systems' technical and natural components adds to the complexity of interactions (Gessner et al., 2014). When precipitation falls, it is intercepted, evaporated, infiltrated, or runs off. Urban areas have reduced green spaces, which are prone to be compacted soils. Moreover, large parts are impervious areas with surface drainage which can guide runoff directly to the stream network or via sewers (Oswald et al., 2023). A study by Haase (2009) in Leipzig found an increased runoff and a decline in actual evapotranspiration rates. The increased amount of water leads to a need for a subsurface drainage system and wastewater management (Haase, 2009). This water is not necessarily disconnected from the natural water cycle as the subsurface drainage and sewer network can still interact with groundwater through leaky pipes (Gessner et al., 2014; Oswald et al., 2023). Gessner et al. (2014) mentioned that water in the sewer

network not only infiltrates the aquifer, but the hydrostatic pressure in aquifers can lead to water infiltration into the sewer system. Additionally, water may be imported or exported across natural catchment boundaries (Oswald et al., 2023). Stable isotopes can be used as natural tracers to get more insight into the movement and origin of water.

1.2. Stable isotopes as tracers

To understand the hydrological cycle in urban areas, stable isotopes of hydrogen (^{2}H or deuterium (D)) and oxygen (^{18}O) can be used, as isotopic compositions change throughout the water cycle. When measuring the isotopic composition of water, it is expressed as delta values as compared to the international Vienna Mean Standard Ocean Water (VMSOW):

$$\delta\,[‰] = \left(\frac{R_{sample}}{R_{VSMOW}} - 1\right) \times 1000 \qquad \text{(Eq. 1)}$$

R is the atom ration D/^{1}H and ^{18}O/^{16}O, respectively (Craig, 1961).
Positive δ-values signify an enrichment of heavy isotopes relative to the standard; negative δ-values indicate depletion (Gat, 1996). A correlation between D and ^{18}O was discovered and defined as the Global Meteorological Water Line (GMWL):

$$\delta\mathrm{D} = 8\,\delta^{18}O + 10\,[‰] \qquad \text{(Eq. 2)}$$

(Craig, 1961).
Local differences from the GMWL can occur and result in a Local Meteoric Water Line (LMWL) (Gat, 1996), where 10 [‰] is replaced by the local calculated deuterium excess value (Dansgaard, 2012). The LMWL can be calculated by amount-weighted least square regression from precipitation isotopes (Hughes & Crawford, 2012):

$$LWML: \qquad \delta D = a \times \delta^{18}O + b \qquad \text{(Eq. 3)}$$

where *a* is the slope and *b* is the intercept of the weighted isotopic composition (Ring et al., 2023).
After evaporation, water vapour is more depleted of heavy isotopes, while the remaining water is more enriched (assuming equilibrium fractionation). However, in natural open systems, where the air is unsaturated, lighter water vapour molecules D are more likely to evaporate than the heavier water vapour molecules ^{18}O (Sprenger et al., 2017). This process leads to distinct differences in isotopic compositions, and when plotted on an “evaporation line”, it diverges from the global or local water line (Dansgaard, 2012). These isotopic signals are also present in soil samples at different depths (Sprenger et al., 2017).

Isotope hydrology is a relatively new area of research (Ehleringer et al., 2016). Isotope-based studies need extensive monitoring data and field studies, which are difficult to maintain by local authorities or researchers (Kuhlemann et al., 2020). Nevertheless, in recent years, isotope hydrology has started to grow. Studies used isotopes to improve our understanding of water partitioning in urban soils and vegetation, to comprehend how urbanisation affects the age distribution and travel times of runoff, to investigate stormwater control measures and to assess their effects on event contributions (see in Kuhlemann et al., 2022; Marx et al., 2021).

This thesis focuses on Berlin, where recent isotope-based studies have significantly improved the understanding of urban water cycle. Particularly insightful for future scenarios were studies carried out including the dry summers of 2018, 2019 and 2020 (Kuhlemann et al., 2022). Isotope data could reveal the dominant water source of the river Panke. Marx et al. (2021) found that groundwater

accounted for 75 percent and urban storm drainage for 25 percent of the annual runoff in the Panke's upper catchment. In the lower catchment, the effluent discharge of a water treatment plant accounted for 80 percent; groundwater and urban storm runoff each accounting for 10 percent of the remaining runoff (Marx et al., 2021). The distinction of water origin becomes complicated within an urban area, where stormwater drainage systems and urban streams are tightly connected and respond quickly to precipitation events. Especially, since the isotopic contrast between natural and engineered runoff is limited due to the similar isotopic composition. Combining isotope with hydrogeochemical data made it possible to overcome this limitation and distinguish between the two water sources (Kuhlemann et al., 2020). Furthermore, Kuhlemann (2021) used soil water isotopes to examine ecohydrological partitioning at different vegetation covers. Trees showed a high evapotranspiration loss with dry soils, whereas urban grassland had high soil evaporation losses, allowed for more percolation, and had higher moisture levels (Kuhlemann et al., 2021). Another study also showed shallow soils to have rapid water turnover and to be dominated by young water; this signal decreased with more depth (Marx et al., 2022). Overall, the study concluded that urban trees and grassland will require irrigation to preserve urban green spaces whereas shrubs may be more resilient (Kuhlemann et al., 2021). In part of this thesis, grassland covered, shallow soil was looked at throughout an urban catchment to add to the ongoing study of isotopic research in Berlin.

1.3. Importance of hydrological modelling

A Model is a simplified representation of reality. With mathematical formulas, it describes variables and their relationships. Generally in hydrology, models usually have one of two goals: 1) to help understand a process, formulate, and test a hypothesis of the hydrological water cycle, and 2) to create predictions outside of the range of observations that can be tested (Solomatine, 2011). The hydrological water cycle is very heterogeneous and has a wide range of spatial and temporal scales. Thus, no model can be assumed universal (Rosbjerg & Madsen, 2005). Different criteria can classify models. One would be to name them after their final purpose or usage, such as rainfall-runoff model, river and reservoir model, or groundwater model (Rosbjerg & Madsen, 2005). Another description of models can be, how they represent the physical process in the model structure, such as deterministic vs. stochastic properties. Further, they are characterised by how the data inputs and model parameters are used in a function of space and time, for example, lumped vs. distributed (Rosbjerg & Madsen, 2005; Solomatine, 2011). Moreover, there is a distinction between empirical, conceptual, and physically based models (Solomatine, 2011). Models can be simple or rather complex, and one is not worse than the other. A conceptual-based model can be suited for a larger area with unknown processes. The parameters represent an overall average, and variables should be seen more as a model-specific index than the actual value. If a smaller scale area is investigated, where parameters can be measured or even manipulated and have a small variability, a physically based model can be more feasible (Bergström, 1991).

During the modelling process to quantify the results, the data is split into two: one part is used to train the model and the other to validate the model and the performance. There are many objective functions to calibrate a model depending on the focus of the simulation, for example, Nash-Sutcliff efficiency (Nash & Sutcliffe, 1970), Kling-Gupta model efficiency (Gupta et al., 2009), or a non-parametric efficiency (Pool et al., 2018). The objective functions give a general idea of how good the performance is but are rather vague without context. By using upper and lower benchmarks, a more accurate model performance evaluation can be made (Seibert, 2001). In the end, the result is one parameter set that performed the best, but often, different parameter sets can lead to similarly good results. This concept is called equifinality (Beven, 2012). To avoid this problem an ensemble mean of the an specified number of “best” parameter sets can be used as a final simulation output. If the validation performs significantly worse than the calibration, it could be a problem of over-

parameterisation, where there are too many degrees of freedom for the information contained in the observed data (Bergström, 1991).

Modelling for and in urban areas is very important since it affects a large and growing population. Additionally, foreseeing predicted changes due to climate change (IPCC, 2021), hydrological modelling helps with urban planning and mitigating risks (Beven, 2012). To be prepared for a higher frequency of floods, droughts, or rapid events relies on the ability to predict them at different time scales to prepare protection measures (Brunner et al., 2021). Rainfall-runoff models require accurate rainfall measurements with high levels of temporal and spatial precision (Fletcher et al., 2013). With a clearer understanding of urban runoff compositions in the future, challenges stated in the paper of Fletcher et al. (2013) such as the change of pervious to impervious areas and the resulting surface and subsurface flow, researchers can attempt to model it. For this thesis, the question is if a simple model can represent the runoff of an urban catchment well enough so that differences in the rural and urban (sub)catchments are recognisable.

1.4. Research question and hypothesis

This thesis focuses on the Wuhle catchment in Berlin, Germany. It is composed of two parts. One investigates the stable isotope dynamics in the shallow soil and open water in the catchment and the other focuses on the HBV Model and how well it performs in an urban environment such as the Wuhle catchment. The following research questions were guiding this thesis:

- How do hydrological signals change throughout an urban catchment, and what does influence them?
    - What is the spatial variability of isotopes in the upper soil and open water?
    - How do short-term hydrological dynamics of discharge differ from the upper more rural catchment to the lower urban catchment?
    - What could explain the differences in the upper and lower catchment?

Hypothesis:

- The isotope signals of the upper catchment of the Wuhle differ from the signals in the lower catchment. In the upper catchment, the composition of the isotopes changes more with depth as evaporation is higher than in more urbanised area.
- The HBV model does perform better in the upper catchment than the whole catchment. The performance of simulated runoff in a city environment is dependent on the imperviousness of the area.

Objective:

- Identify and quantify the spatial variability of isotopes in the upper soil (<15cm) and open water during autumn in the Wuhle catchment.
- Test the applicability of the HBV model in an urban area with the runoff in the Wuhle catchment. Identify input uncertainties and their implication for the simulated data.

2. Study area

2.1. Berlin

Berlin is in the northeast of Germany (Figure 1a), with an area of 890 km^2 and a population of close to 3.8 million (AFSBB, 2021a; AFSBB, 2021b). Berlin's flat topography was formed during the Pleistocene glaciation with an altitude from around 25 – 120 m.a.s.l. The surface consists mainly of Quaternary deposits of unconsolidated sediments (Stackebrandt & Manhenke, 2010). The catchment of the Wuhle originates from the Barnim plateau, which is made of subglacial till and flows into the Warsaw-Berlin Glacial Spillway (Kuhlemann et al., 2022; Limberg et al., 2007). Under the city lies a lower saline aquifer and an upper freshwater aquifer of Tertiary to Holocene age. The average thickness of the freshwater aquifer is around 150 m, for the Wuhle catchment the average is around 100 m (Limberg & Thierbach, 1997). The groundwater heads on the plateaus are > 10 m below the surface, in the glacial valley they are higher at around 4 m below the surface (Limberg et al., 2007).

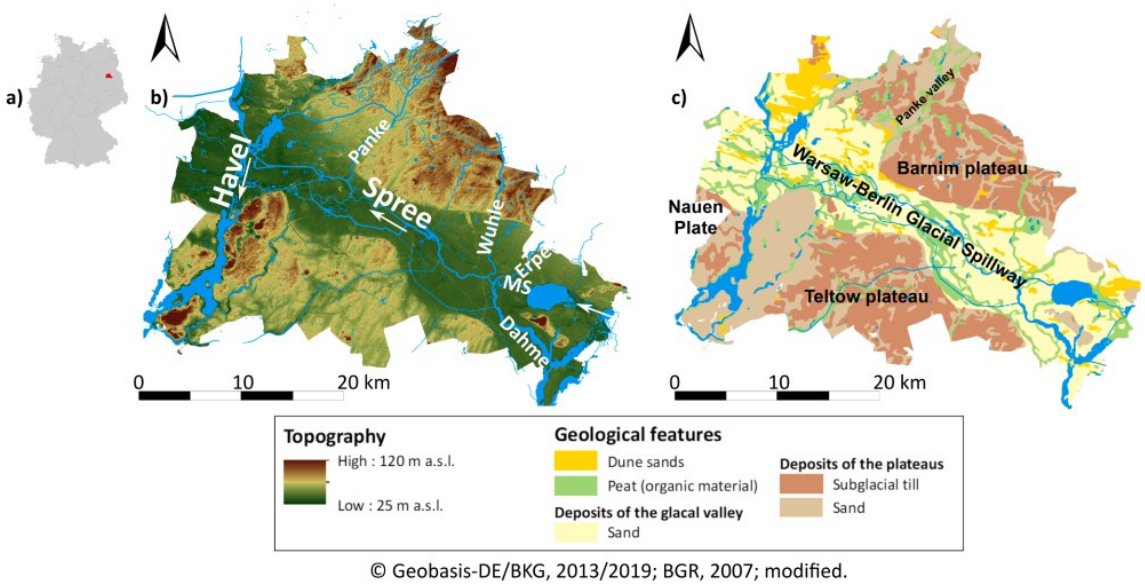


Figure 1: a) Location of Berlin (red) in Germany; b) Topographic Structure with surface waters (blue); c) Geological outline and surface waters (blue). Figure from Kuhlemann, et al. (2022).

The city has a lot of green and blue spaces including a forest (17.7 %), public green spaces (12 %), water bodies (6.6 %) and agricultural areas (4 %) (SenUMVK, 2018). Around 60 percent of Berlin is covered by buildings and roads (SenUMVK, 2018), and the average percentage of sealed surfaces is at 33.4 % (SenStadtWoh, 2017).

Climate data can be retrieved from the German weather service (Deutscher Wetterdienst [DWD]). The long-term mean annual rainfall in Berlin (1981-2010) ranges from 525 to 591 mm (DWD, 2023). Long-term mean temperature (1981-2010) ranges from 9.3 to 10 °C (DWD, 2023). In the context of climate change the linear trend of change since 1881 in Berlin Brandenburg was +1.4 °C for temperature, which is the lowest value together with Schleswig-Holstein in Germany, and for rainfall +2.8 mm (DWD, 2020).

2.2. Wuhle catchment

2.2.1. Catchment characteristics

The Wuhle lies on the east side of Berlin (see Figure 1) and is part of the Elbe river basin. The area of

the Wuhle varies between documents. In a report from the senate department for city development the catchment area of the Wuhle is 144 km^2 (SenStU, 2013a) or if divided into the river Wuhle and Alte Wuhle the areas of the catchments are 101 km^2 and 55.7 km^2 respectively (SenSt, 2004). This means that the catchment borders are not clearly defined. The slope of the stream is around 1,9 ‰ (SenStU, 2013a). The catchment is highly urbanised and is connected to stormwater drains (Kuhlemann et al., 2020). The riverbed is strongly modified (SenStU, 2013a). Depending on the source 11 – 14 percent of the catchment are sealed (Kuhlemann et al., 2020; SenSt, 2004). Because of that, precipitation events greatly impact the stream runoff (SenStU, 2013a).

There are multiple streams congregating into the Wuhle. The Wuhle starts in the north from Ahrensfelde on the Barnim Plateau, runs southwards and flows into the Spree (SenStU, 2013b). From the west side of the stream, the Neue Wuhle and the Biesdorf-Marzahner Grenzgraben join the Wuhle. The Neue Wuhle is an artificial stream that was built for the wastewater treatment plant Falkenberg, which was shut down in 2003. In the years 2006-2008 a redesign of the Neue Wuhle took place, which was necessary due to the massively reduced discharge. The Wuhlgraben and the Hellersdorfer Graben flow from the east into the Wuhle (SenStU, 2013b). The Hellersdorfer Graben is also an artificial stream, since there are no indications of a natural origin (SenUMVK, 2021a). There are three still water bodies connected to the Wuhle which are the Wuhlesee, Wuhleblase and Wuhleteich. The Wuhlesee is directly connected with the Wuhle but is part of a tributary. The lake was created to lower the groundwater level in the surrounding area and to trap sediments (SenUMVK, 2021a). In the GIS vector data of the water network in Berlin Brandenburg, additional water bodies are recorded (Landesamt für Umwelt Brandenburg, 2021). In the west lies the Biersdorfer Baggersee and in the east the Kaulsdorfer Baggersee, which consists of three water bodies. All the lakes in the Wuhle catchment cover less than 1 % of the whole catchment area. From the Kaulsdorfer Baggersee, a stream called Eichwaldgraben is recorded, which is not measured nor is it mentioned in any other papers. For an overview of the stream components, see Figure 2 or appendix 9.1, Table 11.

The Wuhle is a heavily urbanised stream and major anthropogenic changes are recorded since the 19th century (SenStU, 2013a). In 1860 a canal system was implemented which caused higher discharge and flooding. Consequently, the stream was artificially deepened and straightened. The upper part of the Wuhle was drained in 1908 for the construction of a cemetery. In 1916 the waterworks Kaulsdorfer Busch was put into service. As mentioned above, the sewage treatment plant Falkenberg was built in 1984 which was decommissioned in 2003. The Neue Wuhle was constructed and used to drain clear water from the treatment plant into the stream network (SenStU, 2013a).

The European Water Framework Directive, accepted in December 2000, defines one environmental objective as good chemical and ecological status for surface water, as long as the stream is not categorised as artificial or significantly changed (SenUMVK, 2021a). As part of the implementation the riverbed structures in the Elbe River basin were categorised from a scale of 1 (unchanged, near-nature) to 7 (entirely change, excessively damaged) (SenSt, 2004). The Wuhle is categorised as 5.3 (strongly changed and significantly damaged). One planned measure is to improve the treatment of collected rainwater before it is released into the stream (SenUMVK, 2022). Other measures are the construction of a berm for retention of runoff peaks, secondary floodplains and re-routing so that the Wuhle can develop dynamically. The measurements are planned to be implemented until 2027 (SenUMVK, 2021a).

2.2.2. Division into subcatchments

GIS data from the geoportal Brandenburg was used for the calculation and modelling of the subcatchments in combination with the Urban Atlas. The data was modified based on the catchment

areas from Berlin and Brandenburg (European Union, 2018; Landesamt für Umwelt Brandenburg, 2023; López Moreira Mazacotte, 2024) which calculates the Wuhle catchment to a size of 109 km^2. The discrepancy of catchment size reflects the complicated hydrology in urban areas. Furthermore, the length of the Wuhle and the tributary streams differ in the vector data (Landesamt für Umwelt Brandenburg, 2021) which is not thematised in this thesis.

*Table 1: Summary of the (sub)catchment division for hydrological modelling in HBV. Station numbers and station names are extracted from open water measuring stations of the water portal Berlin (SenUMVK, 2021b). In the column Information, it is noted if a station provided water level measurements (w) and/or discharge data (q). In subcatchment number, the subcatchments used to calculate the PET for modelling are noted. In W4.4, the subcatchment is market with * as subcatchment 5 was excluded. For a breakdown of different definitions of W4.4 see appendix 9.2, Table 12. A breakdown of area characteristics in subcatchment resolution can be found in appendix 9.1.*

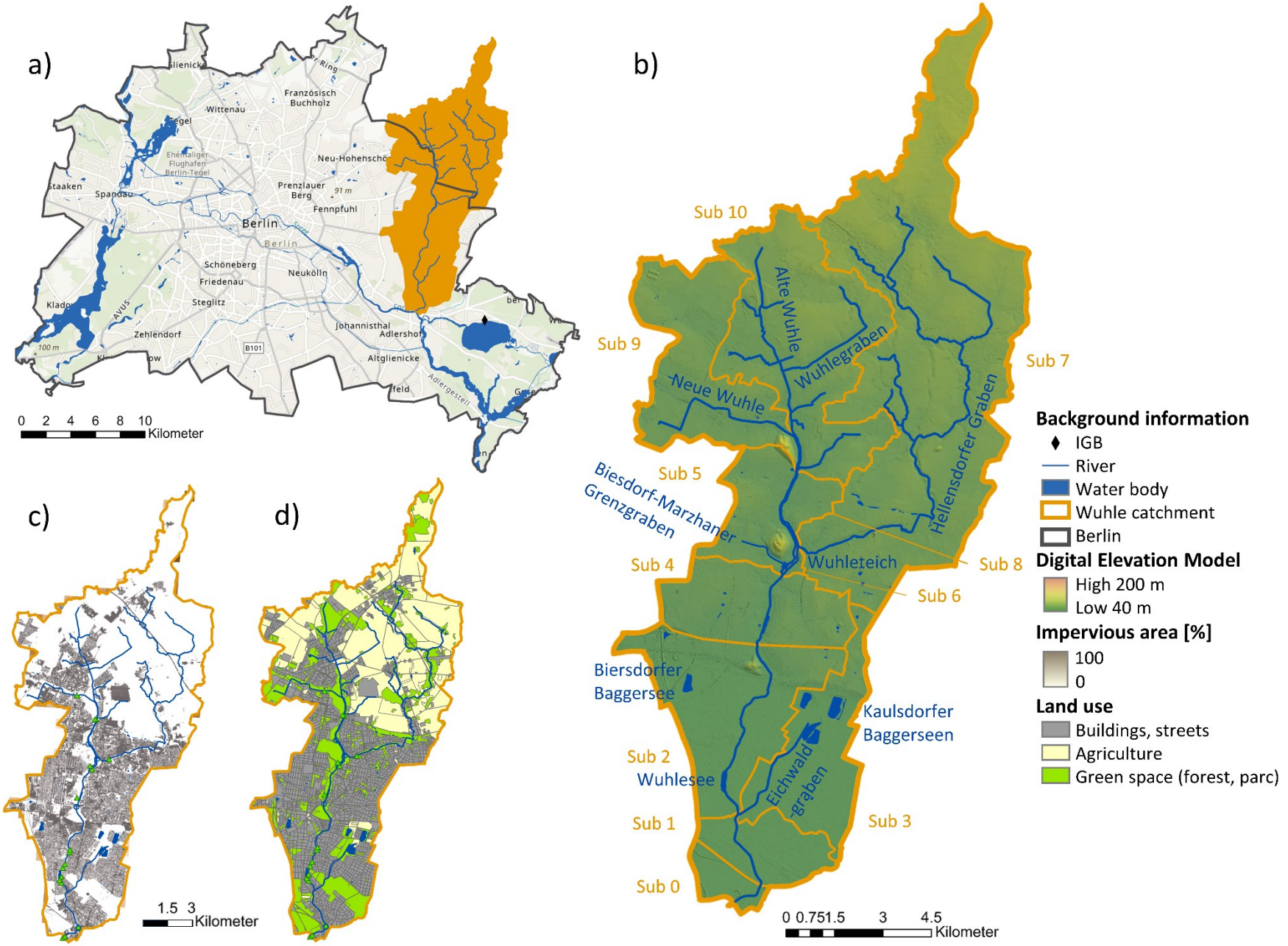
	HG	**W4.0**	**W4.3**	**W4.4**
Administrative information				
Number	5864801	5865300	5864800	5863000
Station name	Am Kienberg	Am Bahndamm	Wuhletal	Eisenacher Strasse
Water body	Hellersdorfer Graben	Wuhle	Wuhle	(Alte) Wuhle
Average discharge [m^3 s^{-1}]	0.022	0.286	0.115	-
Available information	w, q	w, q	w, q	w
Included subcatchment	7	0-10	4-10	10*
Area characteristics				
Area [km^2]	40.4	109.09	82.44	16.75
Area of Wuhle [%]	37.2	100	76	15.4
Ø Impervious [%]	14.2	23.7	20.4	16.4
Ø Sand [%]	76	81.7	78.7	76.2
Ø Clay [%]	7	5.4	6	6.5
Total tree [%]	15.7	25.6	19.3	21.6
Broadleaf tree [%]	12.4	20.4	16.1	15.6
Coniferous tree [%]	3.3	5.1	3.1	6

The Wuhle catchment was divided into subcatchments in two separate occasions (see Figure 5), once to transform the discharge data into specific discharge and once for the potential evapotranspiration (PET) calculations, which were modified. For the discharge data, the whole area was modelled with the data from the measuring station “Am Bahndamm” (W4.0, see Table 1). For two other stations (W4.3, HG) the Wuhle catchment was split into subcachments listed in Table 1. Station W4.4 did not have any discharge data, which is why the data did not have to be transformed according to the (sub)catchment area. The area above the measuring station was assumed to be the catchment area for the respective subcatchment for the specific discharge. Area information from the GIS data was used, as most of the measuring stations are situated close to the defined borders. It was prepared by the Leibniz Institute of Freshwater Ecology and Inland Fisheries (IGB) (López Moreira Mazacotte, 2024). For the PET calculations, the same areas were used. For W4.4 some additional assumptions were made, as the Neue Wuhle and Alte Wuhle cross riverbeds without confluencing and switch side geographically through an atrifical pipe. The station itself would lie in the subcatchment 5. However, in the complicated urban cycle it is not certain how much of the area influences the station directly. Subcatchment 10 is the only one taken into account for the PET calculations for W4.4 as it is the particular area that certainly does influence the measuring station and no further uncertainties are introduced to the calculations. See the map of the subcatchment devision in Figure 5b and Table 2 for a clear understanding of site groupig and subcatchment enviromental characteristics.

Table 2: Environmental characteristics by individual sampling spots. Site name is listed next to sample name. Each separated column is sorted from smallest to biggest value.

Sample/Site		Total tree [%]	Sample/Site		Broadleaf tree [%]	Sample/Site		Coniferous tree [%]	Sample/Site		Sand [%]	Sample/Site		Clay [%]	Sample/Site		Impervious area [%]
16	z	15.00	16	z	15.00	16	z	0.00	09	g	74.02	06b	e	2.49	06b	e	33.57
08	g	21.44	08	g	21.44	08	g	0.00	13	h	76.08	06c	e	2.55	06c	e	31.91
09	g	23.65	09	g	23.60	15	j	0.00	14	h	76.08	06	d	2.62	06	d	30.32
00	a	28.43	00	a	26.94	09	g	0.05	16	z	79.79	05b	c	2.66	05b	c	30.00
15	j	29.29	15	j	29.29	02	b	0.21	08	g	81.08	04	c	2.92	04	c	32.92
01	a	33.83	01	a	32.09	13	h	0.36	15	j	85.16	05	c	2.92	05	c	30.35
02	b	35.55	02	b	35.34	14	h	0.36	07	f	85.46	03	c	3.00	03	c	20.37
13	h	40.19	04	c	38.36	07	f	0.87	00	a	87.21	00	a	3.70	00	a	52.01
04	c	41.62	13	h	39.84	00	a	1.49	01	a	87.21	01	a	3.70	01	a	46.44
14	h	43.13	05	c	42.30	01	a	1.74	02	b	89.49	07	f	3.76	07	f	4.23
05	c	44.81	14	h	42.77	03	c	1.79	04	c	91.37	15	h	4.03	15	h	18.80
07	f	46.47	05b	c	44.13	05	c	2.51	05	c	91.37	02	b	4.22	02	b	50.11
05b	c	48.11	06b	e	44.64	04	c	3.26	03	c	91.44	08	g	4.70	08	g	28.23
06	d	49.69	07	f	45.59	06	d	3.47	06c	e	92.11	09	g	6.79	09	g	9.92
06c	e	49.85	06c	e	45.97	06c	e	3.88	06	d	92.35	16	z	7.41	16	z	24.63
06b	e	51.17	06	d	46.23	05b	c	3.98	06b	e	92.81	13	h	8.01	13	h	43.84
03	c	51.76	03	c	49.97	06b	d	6.53	05b	c	93.51	14	h	8.01	14	h	21.09

Figure 2: Overview of the Wuhle catchment. a) Location in comparison with Berlin, b) digital elevation model including all the streams described in (SenStU, 2013b) and from GIS Data (Landesamt für Umwelt Brandenburg, 2021) where Sub n *is subcatchment number n, e.g. Sub 1 is Subcatchment 1, c) impervious area (European Union, 2020), d) land use throughout the Wuhle catchment (European Union, 2018). Additional maps can be found in appendix 9.4, Figure 18.*



3. Data and methods

3.1. Public data

Daily air temperature from 2 m above the ground and daily precipitation data are available from the climate and data centre (DWD, 2022a; DWD, 2022b). The station Berlin-Marzahn (Station ID: 420) is located outside of the catchment in the west. It is the closest measuring station to the catchment area and is used as the reference point (Fig. 3). As the UHI temperature elevation is more pronounced in the city centre, no additional measures were taken to correct the temperature values (Menberg et al., 2013).

The senate department of the environment, mobility, consumer, and climate protection (SenUMVK, 2021b) provides mean daily discharge data and groundwater level data. Multiple stations along the Wuhle were selected to model the discharge for the catchment area and evaluate the results. Table 1 shows the open water measuring stations used for this thesis. Discharge and water level data was downloaded from water portal Berlin (SenUMVK, 2021b).

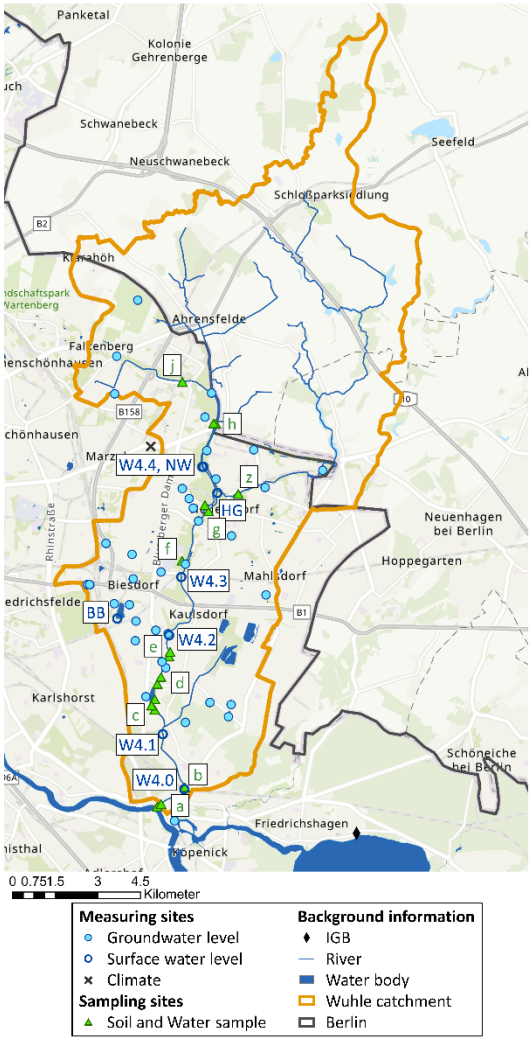


Figure 3: Location of the measuring stations used for temperature and precipitation (climate, black cross) from the climate and data centre (DWD, 2022a; DWD, 2022b), groundwater and surface water level (blue circle) from the water portal Berlin (SenUMVK, 2021b) along the Wuhle and in autumn 2022 sampled soil/water sample sites (green triangle). The location of IGB in Friedrichshagen is marked with a black diamond.

Vector data of the stream network and the polygon data for subcatchments was downloaded from the Geoportal of Brandenburg (Landesamt für Umwelt Brandenburg, 2021, 2023). Raster data of impervious density (European Union, 2020), were retrieved from Copernicus. The land cover data was cut out with a mask of the PET subcatchments (see Figure 2b) and in a circle of 250 m around each sampling point. The tree percentage was calculated by raster with trees divided by the total number of rasters within the circle polygon. The sand and clay information were available in a 500 m raster resolution (see appendix 9.3). The mean value of the circle buffer was calculated and used as the representative soil composition for each sampling point. The mean value of the raster data was additionally extracted for each subcatchment. The summary of the environmental characteristics can be seen in Table 1 for the modelled (sub)catchments and Table 2 for the individual sampling sites.

3.2. Field work

Field work was carried out during Autumn 2022. The soil samples were collected close to the river (25.10, 8 sites and 08.11, 10 sites) and open water samples were taken (25.10). The sites are situated close to the stream bed on grass covered soil. Some sites had trees in near proximity, but none were shaded as the leaves had already fallen. The sites were distributed along the catchment to get a good

overview of the stream when flowing into the city, following sampled sites from the paper by López Moreira Mazacotte et al. (2024).

Water samples were filtered on site (0.2 µm) and put in 1.5 mL glass vials (LLG Labware). They were transported in a thermally isolated box and stored in the refrigerator until laboratory analysis. At the site the soil was first cleaned, and all surface vegetation removed. The temperature and volumetric soil moisture content (VMC) was measured with a handheld probe ML3 Sensor from Delta-T Devices with an accuracy of 3 % (Delta-T, 2023; Marx et al., 2022). Each reading was carried out 3 times in a one metre radius around the soil sampling spot. The soil samples were collected with a HDPE deposit sampler (250 cm^3; Umwelt-Geräte-Technik, Müncheberg, Germany) at 0-5 cm, 5-10 cm, and 10-15 cm depths, measuring the depth with a measuring tape from soil surface to sampling point. The soil samples were extracted and put into bags (WEBAbag, Silver Range, Stand-Up Pouches, Weber Packaging GmbH, Güglingen, Germany) and sealed on site.

In the laboratory, the water samples were analysed by cavity ring-down spectroscopy with a L2130-I isotopic water analyser (Picarro, Inc., Santa Clara, CA, USA). Four lab standards were used for linear correction and standards of the International Atomic Energy Agency (IAEA) for calibration. Results were expressed in δ notation with Vienna Standard Mean Ocean Water (VSMOW) (Kuhlemann et al., 2020). Mean analytical precision was 0.02 ‰ standard deviation (SD) for $\delta^{18}O$ and 0.11 ‰ SD for δD. The isotope compositions of the soil samples were determined using the direct equilibration method, as introduced by Wassenaar et al. (2008). In the laboratory, additional bags were filled with 10 mL of two liquid lab standards, including duplicates, that were used to calibrate the sensors before and after measurements. The field bags were filled with dry synthetic air, welded, and a silicon septum was added. After ~48 h equilibrium was reached in the headspace of the bags. The vapour phase of D and ^{18}O were measured using the Picarro L2130-i by inserting a needle attached to a tube into the bags through the silicon septum to avoid air contamination from outside. Criteria for plateau detection during analysis was SD H^2O < 100 ppm. The vapour was measured for 8-10 minutes. Analytical precision was mean SD of 0.33 ‰ for $\delta^{18}O$ and 0.7-0.85 ‰ for δD. After quality-checking and averaging multiple analyses for each sample, the results were expressed in δ-notation with Vienna Standard Mean Ocean Water (VSMOW). After that, the soil was oven-dried at 105 °C for 24 h. Samples weighted twice before vapour analyse and after drying to determine their gravimetric soil moisture content (GMC) which will be later transformed into VMC. For all isotope samples, line conditioned excess (lc-excess) was calculated (Landwehr & Coplen, 2006).

Because of time restrictions, not many data points could be sampled during the field campaign. The isotopic samples were grouped geographically for general statements, the letter “i” was replaced by “j” for better readability (see Table 3). Enviromental characteristics were determined for each site individually (Table 2). The numbering of the sites is not continuous, because the sampling in October started at the outlet of the catchment (sample WR0) and the field day in November started in the upper catchment (sample WR15).

Table 3: Table with sampling sites and group label (a-j, z), water body information following sample type (w= water, s=soil) and sampling date. The letter" i" was replaced by "j" for better readability. The sampling date is not always applicable to all samples at a site. If the sampling date is relevant, the tables differentiate between measurements taken in October and November.

Site	Sample WR	Water body	Information	Sampling Date
a	0, 1	Wuhle	w, s	25.10.2022, 08.11.2022
b	2	Wuhle	w, s	25.10.2022, 08.11.2022
c	3, 4, 5, 5b	Wuhle	w, s	25.10.2022, 08.11.2022
d	6	Wuhle	w, s	25.10.2022
e	6c, 6b	Wuhle	w, s	08.11.2022
f	7, 7b	Karpfenteich	w, s	25.10.2022, 08.11.2022
g	8, 9	Neue Wuhle, Wuhleteich	w, s	25.10.2022, 08.11.2022
h	13, 14	Neue Wuhle	w, s	08.11.2022
j	15	Neue Wuhle	w, s	08.11.2022
z	16	Hellersdorfer Graben	w, s	08.11.2022

3.3. Other data

A local meteoric water line (LMWL) was calculated from rain samples at the IGB (see location Figure 3) in Berlin-Friedrichshagen and from a location in Berlin-Steglitz. An LMWL for Berlin was calculated using both datasets. The data was generated during ongoing research by Ring et al. (see location and experiment set up in: Ring et al., 2023). Measurements from 01.06.2022 until 17.11.2022 were used to cover a few of months before the sampling period of this thesis. The LMWL was calculated by amount-weighted least square regression (Hughes & Crawford, 2012). The line conditioned excess (lc-excess) as described by Landwehr & Coplen (2006) was calculated to better understand the local evaporative effects.

$$Ic - excess: \qquad \delta D - a \times \delta^{18}O - b \qquad \text{(Eq. 4)}$$

where *a* is the slope and *b* is the intercept of the weighted isotopic composition.
Percentage of sand and clay in soil were modified and provided by the IGB (López Moreira Mazacotte et al., 2024). PET was calculated with the HEC-HMS Model (USACE Hydrologic Engineering Center) using the Penman Monteith Method (Penman, 1948):

$$\lambda ET = \frac{\Delta(R_n - \mathrm{G}) + \rho_a c_p \frac{(e_s - e_a)}{r_a}}{\Delta + \gamma\left(1 + \frac{r_{sf}}{r_a}\right)} \qquad \text{(Eq. 5)}$$

Where *Rn* is the net radiation at the crop surface, *G* is the soil heat flux, ρ_a is the mean air density at constant pressure, c_p, is the specific heat of air, e_s is the saturation vapour pressure, e_a is the actual vapour pressure, and the difference is the vapour pressure deficit. Δ is the slope of the saturation vapour pressure temperature relationship, and *γ* is the psychrometric constant, and r_{sf} and r_a are the (bulk) surface and aerodynamic resistances (not to be confused with the Spearman rank correlation coefficient, r_s), respectively (USACE Hydrologic Engineering Center). The data and calculation were done by the IGB for each individual subcatchment (López Moreira Mazacotte et al., 2024). For this thesis the PET was aggregated to a daily value. The original input for the HBV model is potential evaporation (E_{pot}) data, which was not available nor pre-calculated by other researchers. The model does not explicitly model transpiration. The soil routine in the model (described below) will adapt and is not so sensitive to the exact E_{pot} or PET values used.

The GMC was transformed into VMC to make it comparable with depth. For this the following formula was used:

$$GMC: \qquad \theta g = \frac{soil_{wet} - soil_{dry}}{soil_{dry}} \qquad \text{(Eq. 6)}$$

$$VMC: \qquad \theta v = \theta g \times \left(\frac{soil_{Dry} \times bulk\ density}{water\ density}\right) \qquad \text{(Eq. 7)}$$

Bulk density is the weight of dry soil divided by the volume of the soil sampler (250cm^3) and *water density* assumed to be 1. The result is then given in percent (%).

3.4. Statistical analysis

For the analysis of possible differences and correlations in the data, R version 4.3.1 (2023.06.1) was used (R Core Team, 2023). The data was tested for normality using Shapiro–Wilk (Shapiro & Wilk, 1965). In the case of normally distributed values, the simple t-statistic was performed (Student, 1908) to test for significant differences. For non-parametric data or in the case of skewed data, the Kruskal-Wallis-Test was used to compare ranks for more than two groups (Kruskal & Wallis, 1952; Ring et al., 2023). As a post-hoc analysis the Dunn's-test was used (Dunn, 1964) as seen in the paper by Sprenger et al. (2017) and recommended by the website of the university of Zurich for statistical method consulting (*Kruskal-Wallis-Test*, 2023).

Packages used in R were "hydroGOF" (Zambrano-Bigiarini, 2020) to calculate the goodness of fit functions, with addition of the NPE function from Pool et al (2018). Multiple packages included in "tidyverse" were used, such as "readxl", "luberdate" and "ggplot2" for data manipulation and visualization (Wickham et al., 2019), as well as "RColorBrewer" to select the colours for the different sites plotted in the dual isotope figure (Neuwirth, 2014). Some of the coding was aided by Chat-GPT for problem-solving in R.

4. Modelling with HBV

4.1. Model description

The HBV Model introduced by Bergström in 1976 is a conceptual bucket-type and semi-distributed rainfall-runoff model (Bergström, 1976). The model was originally design for Scandinavian catchments, which can also be derived from its name after the Hydrologiska Byråns Vattenavdelning unit at the Swedish Meteorological and Hydrological Institute. The HBV and has been continuously developed and is a widely used application for runoff modelling (Bergström, 1976, 1992), in research such as quantifying land-cover changes, modelling ungauged catchment, detecting climate change impacts, and evaluating model calibration and uncertainties (Seibert & Bergström, 2022). In this thesis, 'HBV light 1.2' Model is used (Seibert & Vis, 2012).

The advantage of conceptual models, especially in education, is that the processes are transparent and therefore comprehensible but the requirements for input data is more moderate than for physically based models. The consequence of the model summarising the processes to a catchment scale is that the resulting parameters are more an average value and should be interpreted as an index rather than a true value (Bergström, 1991). The runoff simulations are calculated using time series of precipitation (P), temperature (T), observed runoff (Q_{obs}) and long-term E_{pot}. The data is provided in daily time steps, except E_{pot} which can be long-term daily or monthly values. Semi-distributed for HBV means, that catchments can be divided into several vegetation and elevation zones (Seibert & Vis, 2012). As the catchment of the Wuhle is urban and very flat, and no vegetation change is expected due to elevation, no subdivision of the catchment was defined. The (sub)catchments W4.0, W4.3, W4.4 and HG were independently run to compare the resulting parameters with each other.

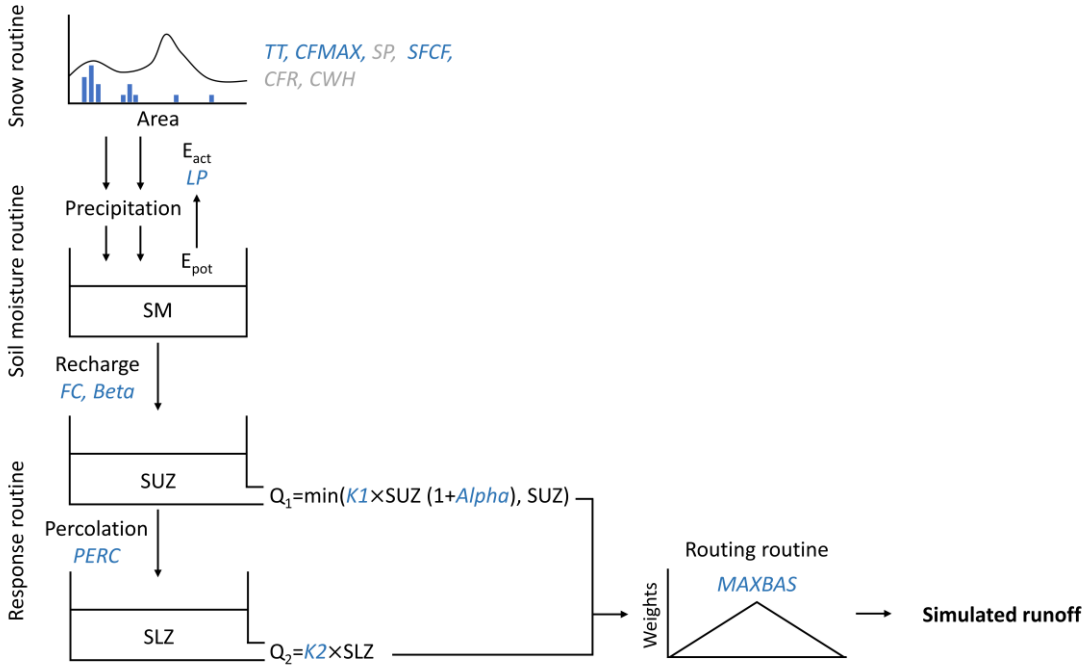


Figure 4: Structure of the HBV model. Adapted from Mendez, M. (2016) and HBV-light help section. Parameter shown in light grey have set values, parameters in light blue were calibrated. Explanation of parameter abbreviation can be found in "Symbols and abbreviations" or Table 5.

The model incorporates different routines which are the *snow routine, glacier routine*, *soil moisture routine*, *response routine*, and the *routing routine*. The model structure can be altered with model variants, which depends on how the routines are computed. In this thesis, the basic standard model is used without an additional threshold parameter and runoff component in the storage of the upper

zone (see Table 4) (Seibert & Vis, 2012). Further, the glacier routine is not needed for the Wuhle catchment. In the following paragraph the most important aspects of the model for this thesis are discussed. In this subchapter the parameters are written in cursive and bold for easier readability. A more detailed description of the model and different variants can be found in the help section of the HBV-light model and the reference cited there (e.g.: Bergström, 1992; Seibert, 1997, 1999; Seibert & Vis, 2012).

In the *snow routine*, precipitation and temperature data is used to calculate if the precipitation could accumulate as snow, taking melting, and refreezing processes into account. The output data of this routine include snowpack and snowmelt. The model distinguishes rain from snow when temperatures are below the *threshold temperature* (***TT***, °C) which is multiplied by a *snowfall correction factor* (***SFCF***, -). Because snowfall is elevation-dependent two catchment parameters ***PCALT*** (% (100 m)$^{-1}$, percent increase of precipitation per 100 m elevation) and ***TCALT*** (°C (100 m)$^{-1}$, decrease of temperature per 100 m elevation) are used to adjust temperature and precipitation input. Snow melt is calculated with the day-degree factor ***CFMAX*** (mm Δt^{-1} °C^{-1}) (Eq. 8). Additional parameters are used to calculate the amount of water refreezing in melting snow, which are influenced by the water holding capacity of snow (***CWH***, -) and the refreezing coefficient (***CFR***, -). Snowpacks are important especially for spring as they attribute to runoff delays.

$$melt = \boldsymbol{CFMAX} \times (T(t) - \boldsymbol{TT}) \quad \text{(Eq. 8)}$$

In the *soil moisture routine,* the output data of the *snow routine* (snowmelt, *P* (as rain) and E_{pot}, (mm Δt^{-1})) are used to calculate water storage in the soil moisture box (*SM*, mm), and water losses through actual evaporation (E_{act}) and groundwater recharge. ***LP*** (-) is a soil moisture value above which E_{act} reaches E_{pot}. The water storage in *SM* and recharge is dependent on the content of *SM* and its maximum storage value (***FC***, mm) (Eq. 9). ***Beta*** (-) determines the relative contribution to runoff from rain or snowmelt. E_{act} from the *SM* is calculated as equal to E_{pot} if *SM*/***FC*** is above ***LP***. If *SM*/***FC*** is below ***LP*** a linear reduction is used (Eq. 10).

$$\frac{recharge}{P(t)} = \left(\frac{SM(t)}{\boldsymbol{FC}}\right)^{\boldsymbol{Beta}} \quad \text{(Eq. 9)}$$

$$E_{act} = E_{pot} \times min\left(\frac{SM(t)}{\boldsymbol{FC} \times \boldsymbol{LP}}, 1\right) \quad \text{(Eq. 10)}$$

In the *response routine*, E_{pot} and the groundwater recharge output from the previous routine are used to calculate the runoff and groundwater level. ***PERC*** (mm t^{-1}) defines the maximum percolation rate from the upper (soil upper zone, *SUZ*) to the lower (*SLZ*, soil lower zone) groundwater box. Larger values for ***PERC*** increase contribution of the lower soil box. In the upper groundwater box (*SUZ*, mm) a non-linear runoff is assumed. From this box, the smaller of two water amounts are used as output: 1) the whole content of *SUZ* or 2) part of *SUZ* calculated as a product of the recession coefficient (***K1***, Δt^{-1}) and the *SUZ* by the power of (1+ non-linearity coefficient ***Alpha*** (-)) (Eq.11). For the lower groundwater box (*SLZ*, mm) a simple linear reservoir is assumed where the runoff *Q(t)* at the time *t* is proportional to the water storage *S(t)*. A certain proportion of *SLZ*, calculated with the second recession coefficient (***K2***, Δt^{-1}), contributes to the discharge (Eq. 12). The resulting runoff is the sum of the calculated *Q1* and *Q2*.

$$Q1(t) = \min(\boldsymbol{K1} \times SUZ(t)^{(1+\boldsymbol{Alpha})}, SUZ(t)) \quad \text{(Eq. 11)}$$

$$Q2(t) = \boldsymbol{K2} \times SLZ(t) \quad \text{(Eq. 12)}$$

Lastly, in the *routing routine*, the calculated runoff is transformed with a triangular weighting function (***MAXBAS***, Δt). This accounts for the delay between the generated runoff and the time it needs to reach the catchment outlet. A list with all the parameters used and their ranges can be seen in chapter 4.3.1, Table 5.

4.2. Model initialization

4.2.1. Input data

The information about precipitation, temperature and discharge (ptq-data file) was compiled from daily values from the climate and data centre (DWD, 2022a; DWD, 2022b) and daily discharge or water level values from the water portal (SenUMVK, 2021b). The discharge data was transformed from observed discharge ($m^3\ s^{-1}$) into specific discharge per day (mm t^{-1}) as follows:

$$Q_{mm/d} = Q_{m^s/s} \times 86400 \times \frac{1000}{A} \quad \text{(Eq. 13)}$$

where *A* is the respective (sub)catchment of the area above the measuring station and *86'400* is the number of seconds in a day. The water level values were not transformed as the later calculations used ranks and operate in the ordinal scale.

Because of data gaps between 1993 and 2007 the precipitation and temperature data of the year 2008 was duplicated to achieve three years for the warm-up period. The mean temperature in 2008 (10.7 °C) lies slightly above in the long term mean of 9.3 – 10 °C in Berlin (DWD, 2023), fits into the standard deviation of the following years of 2009 (10 °C), 2010 (8.7 °C), and 2011 (10.7 °C). The accumulated rainfall of 612.6 mm lies above the average mean of 525 – 591 mm in Berlin (DWD, 2023), but it is more aligned with the total rainfall amount in the year 2009 (629.6 mm), 2010 (641.6 mm) and 2011 (688.9 mm) (DWD, 2023).

The required E_{pot} was replaced by long-term PET data (López Moreira Mazacotte, 2024) which was modified from 10-year bidaily values (2011 to 2021) for each subcatchment (0-10) into summarised daily mean values for each modelled (sub)catchment (W4.0, W4.3, W4.4, HG) defined in this thesis.

4.2.2. Model and catchment settings

As mentioned above the standard basic model without the use of an additional runoff in the upper storage box (UZL and K0) was used (see help section, HBV light). The efficiency for specified season was set to the hydrological year from 1st of October to 30th of September. The warm-up period was set to three years before the start of the calibration period (see Figure 5).

Because of the flat topography of the Wuhle catchment, simply one elevation and vegetation zone were used and the settings for PCALT and TCALT have been left at their default values (see Table 5) (Bergström, 1992). The elevation of precipitation and temperature were set to the elevation of the weather station (61 m.a.s.l.) and the mean elevation of the catchment was derived from the GIS data and set at 42 m (GeoBasis-DE, 2019). Lakes in the catchment were neglected as there are mostly artificial ponds disconnected from the stream network and cover less than 1 % of the whole catchment area.

4.3. Model calibration

The calibration of the model is the process of parameter adjustment until simulated runoff is sufficiently similar to the observed runoff. To estimate the difference between observed and simulated discharge, an objective function is used (Solomatine, 2011). To later be able to validate the model on independent data, only part of the data time span was used for calibration. The model calibration was completed in two rounds. First, a pre-calibration to search for feasible parameter ranges with the Monte-Carlo approach and get an insight into parameter uncertainty was executed (Seibert & Vis,

2012). In this approach randomly chosen parameters are combined with other random parameters to build a parameter set (Harrison, 2010). In the HBV-light model, the 100 best performing parameter sets according to the chosen objective functions were than saved and evaluated.

Second, after parameter ranges were defined (see Table 5) a Genetic Algorithm and Powell optimization (GAP) was used to calibrate the HBV model (Seibert, 2000)(see Table 4). With this method, optimised parameter sets are found by starting with an initial parameter population and the value of the chosen objective function (see chapter 4.3.2) is evaluated. Parameter sets with good model performances are more likely to further evolve than parameter sets with bad model performances. The first step generates two parameter sets randomly. Then the parameters can be evolved from the previous parameter sets, a random value between the two parameter sets or a random value within the limits for the parameter mutation. In the second step, the parameter sets are fine-tuned using Powell's quadratically convergent method (as described in (Press William H., 2002; Seibert & Vis, 2012). Even though, theoretically this should lead to the best parameter set it can also result in equifinality (Beven, 2012). In the end the 100 best performing parameter sets were saved.

Table 4: GAP settings used for model calibration in HBV.

GAP calibration settings	Number
Number of Runs	5000
Number of Powell Runs	1000
Number of Parameter Sets	50
Number of Populations	1
Number of Calibrations	100

4.3.1. Parameter ranges

The parameter ranges and fixed parameter values used in the calibration are listed in Table 5. For twelve parameters, a range was defined. To minimize the calculation effort, the following rather unsensitive parameters CFR, CWH were fixed to their default value of 0.05 and 0.1 respectively. For the degree-day factor CFMAX no seasonality is expected so the parameter SP is also kept at the default value of 0.

For the other parameters, ranges were chosen after experimenting with multiple Monte-Carlo runs to gauge feasible ranges, described in the previous chapter on model calibration. In a paper by Seibert (2000) feasible ranges are proposed for the HBV model to calibrate with the GAP. A basic overview was created for each (sub)catchment by producing dotty plots. Ranges were established after defining acceptable objective function values by a visual examination of the clusters in the dotty plots (final resulting dotty plots of the 100 best parameter sets can be found in appendix 9.7-9.8 and 9.11.1-9.11.2). If point clusters were all at the range boundaries, the ranges were broadened. There were a few exceptions for that. The maximum storage in the soil box FC in the soil moisture routine was high especially in the subcatchment HG. In the case of this thesis the value was fixed with the range suitable for the other (sub)catchments (W4.0, W4.3) to avoid even larger variability. With the pre-calibration approach, the recession coefficient in the response routine K2 could have been higher than K1. This would logically not make sense in the model; therefore, the maximal range was kept at 0.4 for K2. For the non-linearity parameter Alpha, the range was limited from 0 to 1 based on previous experience in the Hydrology and Climate group (H2K) at the University in Zurich, as no improvement was expected as soon as 1+Alpha reached values larger than 2 (Schwarzenbach, 2022).

Table 5: Model parameter ranges used for the GAP optimisation for calibration in all (sub)catchments. The valid range was extracted exactly from the help section in HBV-light.

Parameter	Description	Unit	Range/Value	Valid range
Snow Moisture Routine				
TT	Threshold temperature	°C	[-2;4]	(-inf, inf)
CFMAX	Degree-day factor	mm t^{-1} $°C^{-1}$	[0.5;6]	[0, inf)
SP	Seasonal variability in degree-Δt factor	-	0	[-1, 1]
SFCF	Snowfall correction factor	-	[0;3]	[0, inf)
CFR	Refreezing coefficient	-	0.05	[0, inf)
CWH	Water holding capacity	-	0.1	[0, inf)
Soil Routine				
FC	Maximum storage in soil box	mm	[500;1500]	[0, inf)
LP	Threshold for reduction of evaporation	-	[0;1]	[0,1]
Beta	Shape coefficient	-	[1;6]	(0, inf)
Response Routine				
K1	Storage / Recession coefficient	Δt^{-1}	[0.01;0.5]	[0, 1)
K2	Storage / Recession coefficient	Δt^{-1}	[0.00005;0.4]	[0, 1)
Alpha	Non-linearity coefficient	mm	[0;1]	[0, inf)
PERC	Maximal flow from upper to lower box	mm t^{-1}	[0;15]	[0, inf)
MAXBAS	Routing, length of weighting function	Δt	[1;4]	[0, 100]
Other				
Cet	Potential evaporation correction factor	$°C^{-1}$	[0;0.3]	[0, 1]
PCALT	Increase of precipitation with elevation	% $(100m)^{-1}$	10	(-inf, inf)
TCALT	Decrease of temperature with elevation	°C $(100m)^{-1}$	0.6	(-inf, inf)
Pelev	Elevation of precipitation data in the input file	m	61	(-inf, inf)
Telev	Elevation of temperature data in the input file	m	61	(-inf, inf)

4.3.2. Objective functions

To compare the aggregated differences between observed and simulated discharge, objective functions are used during calibration (Solomatine, 2011). In this thesis, the main objective function used was the Non-Parametric Efficiency (NPE) introduced by Pool et al. (2018). For ordinal data from the subcatchment W4.4, where just water level data was available, the Spearman Rank Correlation Coefficient r_s was used for calibration (Spearman, 1904). For the later discussion the Nash-Sutcliff efficiency NSE or R_{eff} (as it is labelled in HBV light)(Nash & Sutcliffe, 1970) and the Volumetric Efficiency or Volume Error VE (Criss Robert E. & E., 2008) are mentioned. For an overview of the objective functions, see Table 6. In the paper by Nonik et al. (2021) threshold values for good performances are summarised for r_s as 0.8, for R_{eff} as 0.75 and for VE as 0.15. For a more specific evaluation of the model performance benchmark were used (see chapter 4.4.1).

The NPE was used rather than the Kling-Gupta efficiency (KGE) (Gupta et al., 2009). The KGE assumes data linearity, normality, and the absence of outliers. However, discharge time series and model simulations are often highly skewed (Pool et al., 2018). The NPE therefore, does reformulate the variability and correlation term of the Kling-Gupta model efficiency (KGE) to a non -parametric form.
The NPE is based on the normalized flow-duration curve (α), the Spearman rank correlation coefficient (r_s) and the mean discharge (ß) (Pool et al., 2018) which are listed in Table 6. For α absolute error was

computed between all ranked simulated and observed discharge values, where I(k) and J(k) are the time steps when the *k*th largest flow occurs within the simulated and observed time series respectively (Pool et al., 2018).

Table 6: Equation of the objective functions used in this thesis. Table inspired by Nonik et al. (2021). $R_{obs,j}$ *and* $R_{sim,j}$ *are ranks of observed and simulated data in the time step* j *respectively.* $\overline{Qobs}$ *and* $\overline{Qsim}$ *are the average of observed and simulated streamflow respectively and* n *the number of time steps in the period of simulation (Nonki et al., 2021). For* α *absolute error was computed between all ranked simulated and observed discharge values, where* I(k) *and* J(k) *are the time steps when the* k*th largest flow occurs within the simulated and observed time series, respectively (Pool et al., 2018).*

Goodness of Fit function	Calculation	Part of hydrograph	Value for 'perfect' fit
r_s	$\frac{\sum_{j=1}^{n}\left(R_{obs,j}-\overline{R_{obs}}\right)\left(R_{sim,j}-\overline{R_{sim}}\right)}{\sqrt{\left(\sum_{j=1}^{n}\left(R_{obs,j}-\overline{R_{obs}}\right)^2\right)\left(\sum_{j=1}^{n}\left(R_{sim,j}-\overline{R_{sim}}\right)^2\right)}}$	Timing	1
NPE	$1-\sqrt{(\alpha-1)^2+(ß-1)^2+(r_s-1)^2}$	Entire hydrograph	1
α	$1-\frac{1}{2}\sum_{j=1}^{n}\left\lvert\frac{Q_{sim}(I(j))}{n\overline{Q_{sim}}}-\frac{Q_{obs}(J(j))}{n\overline{Q_{obs}}}\right\rvert$	-	-
ß	$\overline{Q_{sim}}/\overline{Q_{obs}}$	-	-
R_{eff} (NSE)	$1-\frac{\sum_{j=1}^{n}(Q_{obs,j}-Q_{sim,j})^2}{\sum_{j=1}^{n}(Q_{obs,j}-\overline{Q_{obs}})^2}$	Focus on high flow	1
VE	$1-\frac{\left\lvert\sum_{j=1}^{n}Q_{sim,j}-Q_{obs,j}\right\rvert}{\sum_{j=1}^{n}(Q_{obs,j})}$	Water Balance	0

4.4. Model validation

For model validation the split sample method was used (Klemeš, 1986), where part of the available data is not used for the calibration to then be simulated with the found parameter values and the resulting objective function compared. Ideally, the data is split into two equal parts, but as the available data was not long enough (5-10 years, based on experience of the H2K), the data for the (sub)catchments was split into two simulations: Simulation A used the first 70 percent of the data for calibration and the last 30 percent for validation, simulation B used the last 70 percent for calibration and the first 30 percent for validation (see Figure 5). The model is deemed acceptable for further use (e.g. in decision making processes) only if the two results from simulation A and B are similar and the error in both validation runs acceptable (Klemeš, 1986). The chosen time periods also align with the differential split sample test for most of the (sub)catchments, except for HG. In this method, the data would be split into dry and wet climate scenarios (Klemeš, 1986). In HG the data split is in Autumn 2018, but 2019 was also a dry summer which affected the hydrological cycle (Kuhlemann et al., 2020). In a paper by Motavita (2019) they found that the length of calibration period is less relevant for model accuracy than the hydrological conditions. Further, they also showed that calibrating on longer periods of dry conditions leads to better results for the whole timespan than calibrating wet periods to predict later dry periods (Motavita et al., 2019).

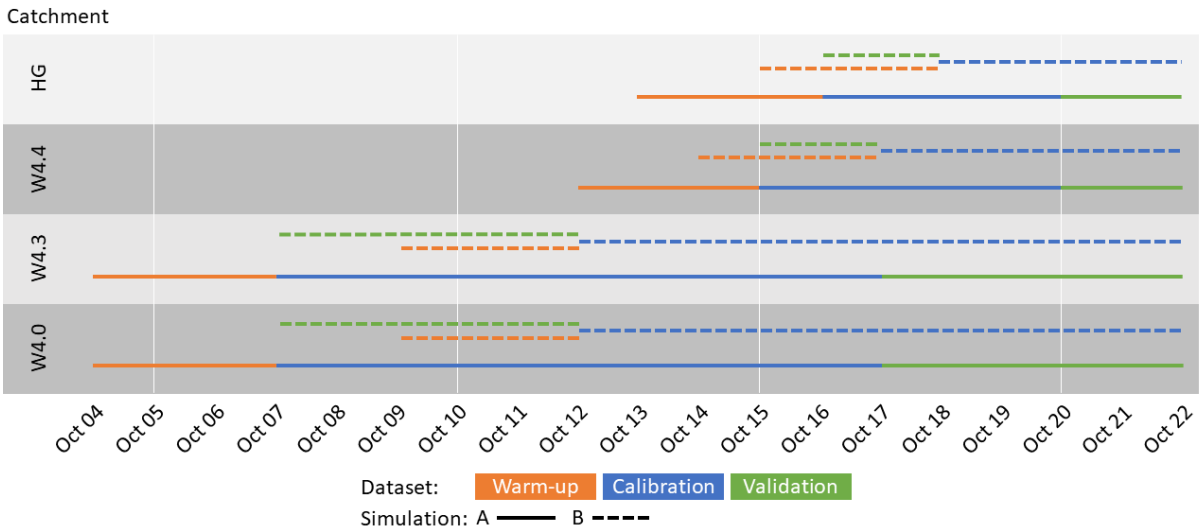


Figure 5: Time periods used for warm-up, calibration, and validation of the catchment W4.0 and subcatchments W4.4, W4.3 and HG. In simulation B the warm-up period and validation period data overlapped, which is not an issue as the warm-up period is not used for the calibration itself.

4.4.1. Upper and lower benchmark

Benchmarks allow to estimate the values derived from objective functions and put them in the context of an upper and lower benchmark. As the perfect value of objective functions (e.g. 1 for NPE or 0 for VE, see Table 6) can never be achieved. The proposed approach by Seibert et al. (2018) suggest using a bucket-type model, such as the HBV model, to compare any model to. For the upper benchmark, the catchment is calibrated with the best possible simulation. For the lower benchmark the model is run with random parameters within feasible ranges.

In this thesis the HBV-light model was already used for general calibration, therefore the benchmarks were defined as follows: The result of the calibration and the validation was calculated as the ensemble mean hydrograph of the 100 best performing parameter sets for the NPE. The upper benchmark was defined as the best calibration from the GAP optimisation, regardless of the parameters used, as an "ideal" value. To evaluate the model and how good the value of the objective functions was, a lower benchmark following the method described in the paper by van Meerveld et al. (2017) was introduced. The model was run with the Monte-Carlo calibration for each (sub)catchment with 1'000 runs using randomly chosen parameters within the same pre-defined ranges as the GAP-calibration (see Table 5). The ensemble mean was then selected as representative simulation and the resulting objective functions were chosen as a lower benchmark.

The upper and lower benchmarks were calculated only for the calibration period, as the validation period usually does perform worse. With this, a higher goal is set to achieve, because the model is calibrated particularly for the used data. The upper benchmarks (u.b.) and lower benchmark (l.b.) were then set into relation with the validation of the ensemble mean for each (sub)catchment to quantify the relative performance of the respective objective function $R_{x,rel}$ (Seibert et al., 2018):

$$R_{rel} = \frac{R_{(sub)catchment} - R_{l.b.}}{R_{u.b.} - R_{l.b.}} \quad \text{(Eq. 14)}$$

5. Results

5.1. Antecedent conditions

The fieldwork of this thesis was conducted in 2022. The mean temperature during that year was 11.4 °C which lies slightly above the mean temperature in Berlin (9.3 – 10 °C (DWD, 2023)). The mean temperature for autumn 2022 (September – November) was 11.06° C which is 1 °C warmer than the long-term average from 1993-2022 measured by the station Berlin-Marzahn. The precipitation during 2022 totalled 429.2 mm, which was 129 mm (23 %) lower than the long-term average (DWD, 2023). The total precipitation during autumn 2022 was 101.6 mm which is also lower mean total precipitation of 137 mm (1993-2022, Station Berlin-Marzahn: (DWD, 2023)). The biggest precipitation events before the sampling period occurred on the 18.10.2022 with 13.1 mm of precipitation. Between the two field days it rained a total of 4 mm (Berlin-Marzahn Station).

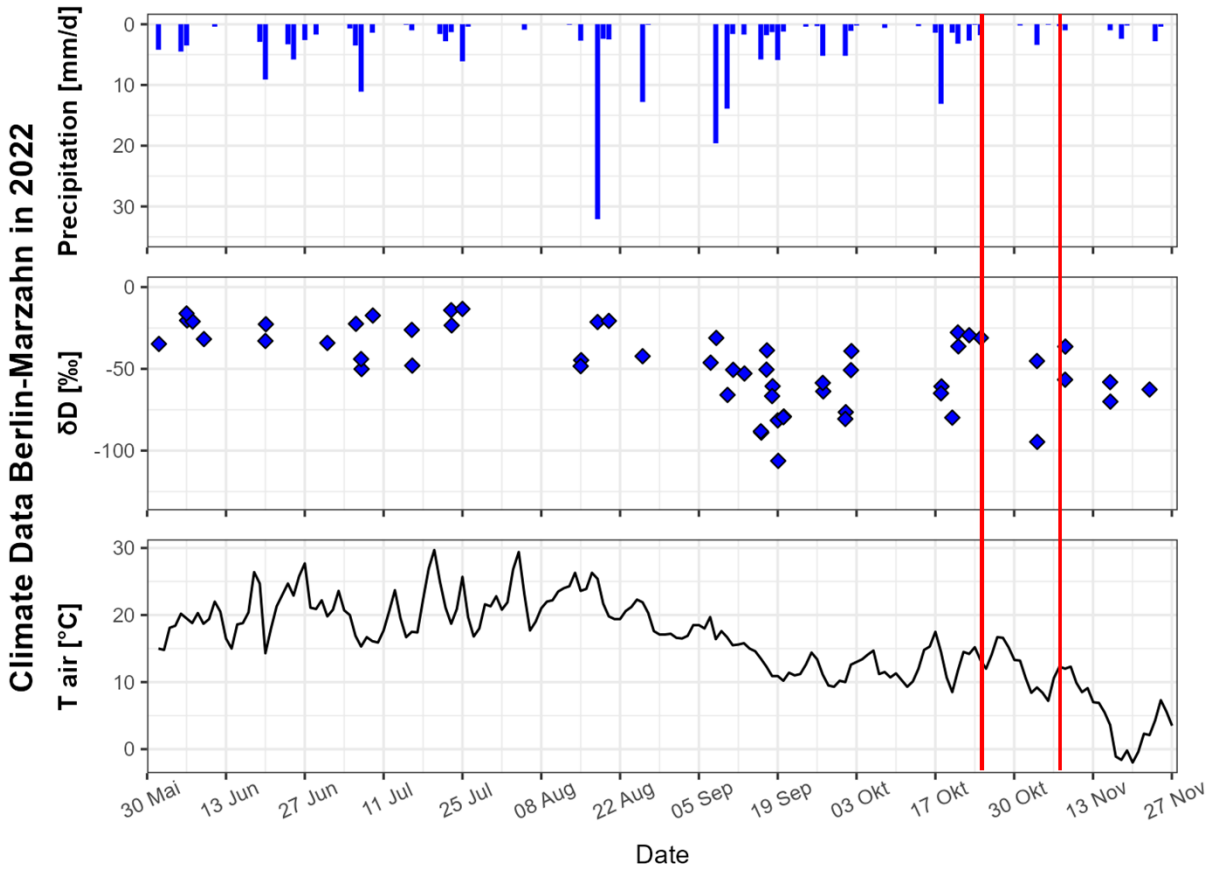


Figure 6: Antecendent conditions in the Wuhle catchment starting half a year before the samples were taken during field work in autumn 2022. Precipitation data was downloaded from the station Berlin-Marzahn (ID-number: 420) and isotope data was measured at the study the location at IGB in Berlin-Friedrichshagen (Ring et al., 2023). The two red lines mark the two field days.

Between September and the second sampling day (01.09 – 08.11.2022), 20 samples of isotopes were taken in Berlin-Friedrichshagen. Figure 6 shows that the δD varied highly. The isotope ratios ranged from -14.9 to -3.7 for $\delta^{18}O$ and from -106.3 to -25.4 ‰ δD (see Table 7). The isotopic signature of δD had more variable and more depleted values than in the summer months before (see Figure 6). The LMWL was calculated by a weighting of the isotopic data from Berlin-Friedrichshagen and Berlin-Steglitz, which is labelled with *Berlin* in the following plots. For the LWML calculation the slope from Friedrichshagen was 7.784 and the intercept 5.954 and for Steglitz 7.325 and 6.901. The resulting LWML for Berlin looks as follows:

$$LWML_{Berlin}: \quad \delta D = 7.810 \pm 0.26 \times \delta^{18}O + 8.00 \pm 2.38 \; (R^2 = 0.97) \qquad \text{(Eq. 15)}$$

The resulting lc-excess is therefore as follows:

$$Ic - excess_{Berlin}: \quad \delta D - 7.81 \times \delta^{18}O - 8 \qquad \text{(Eq. 16)}$$

Table 7: Weighted $\delta^{18}O$, δD and lc-excess values for precipitation in Berlin, from September until the second sampling day (08.11.2022). For reference the $\delta^{18}O$, δD and lc-excess values of Berlin-Friedrichshagen and Berlin-Steglitz are listed as well.

Precipitation	**‰**	**min**	**mean**	**max**	**SD**
Berlin	δD	-97.22	-60.15	-26.54	21.63
	$\delta^{18}O$	-13.71	-8.76	-3.94	2.83
	Lc-excess Berlin	-7.52	0.25	7.14	3.63
Friedrichshagen	δD	-106.27	-59.07	-25.35	22.94
	$\delta^{18}O$	-14.94	-8.61	-3.74	3.02
	Lc-excess IGB	-7.39	0.23	7.21	3.74
Steglitz	δD	-88.17	-61.80	-27.73	19.33
	$\delta^{18}O$	-12.48	-8.99	-4.15	2.51
	Lc-excess Steglitz	-11.04	-4.47	1.33	3.43

The discharge in the Wuhle catchment is very small. In W4.0 the average discharge (2009-2021) was 0.28 $m^3 s^{-1}$ and during autumn it was around 0.197 $m^3 s^{-1}$ and 0.209 $m^3 s^{-1}$ for September and October, respectively. The mean discharge for 2022 was 0.145 $m^3 s^{-1}$, for September and October it was 0.098 $m^3 s^{-1}$ and 0.078 $m^3 s^{-1}$, respectively. In W4.3 the average discharge (2005-2021) was 0.122 $m^3 s^{-1}$ and for September and October 0.075 $m^3 s^{-1}$ and 0.079 $m^3 s^{-1}$, respectively. In 2022 the average discharge was 0.024 $m^3 s^{-1}$, in September 0.012 $m^3 s^{-1}$ and October 0.004 $m^3 s^{-1}$. In HG, the measurements started later, the mean average discharge (2018-2021) was 0.023 $m^3 s^{-1}$ and in autumn it was 0.013 $m^3 s^{-1}$ and 0.017 $m^3 s^{-1}$ for September and October, respectively. For 2022 the average discharge was 0.012 $m^3 s^{-1}$, for September 0.014 $m^3 s^{-1}$ and for October 0.009 $m^3 s^{-1}$.

5.2. Isotope dynamics

5.2.1. Overall dynamics in the Wuhle catchment

Figure 7 shows the isotopic signatures plotted in a dual isotope space. The measurements do not show clear deviation from the LMWL. Precipitation is the most spread along the MWLs. The two most depleted values are measurements from the beginning of November. The surface water isotopes generally lie below the GMWL and the LMWL of Berlin. When tested statistically, surface water signatures deviate from the LMWL of Berlin and the GMWL ($p < 0.05$). The most depleted values are the open water measurements from site a, part of c (WR4 and WR5) and d. The noticeable datapoints, which are clearly beneath the LMWL, are the sites at WR3, WR9, and WR14 which are all larger water bodies in the stream (see Table 8 and appendix 9.1, Table 11). Although the data is limited, if just the flowing water of the stream is considered, no deviation from the LMWL is detectable. If a linear regression is calculated for only the water sampled from the ponds or all of the water samples and are compared to the LMWL, the data does deviate (simple t-test, $p < 0.001$). Figure 19 in appendix 9.4 shows a close-up of the water data with the corresponding catchment position. Since not enough data is available, this must be treated as a trend. The soil water measurements deviate significantly from both MWLs ($p < 0.001$). The data points cluster beneath the LMWLs indicating evaporation processes. A more detailed description of the results considering catchment position can be found in the following subchapter 5.2.2.

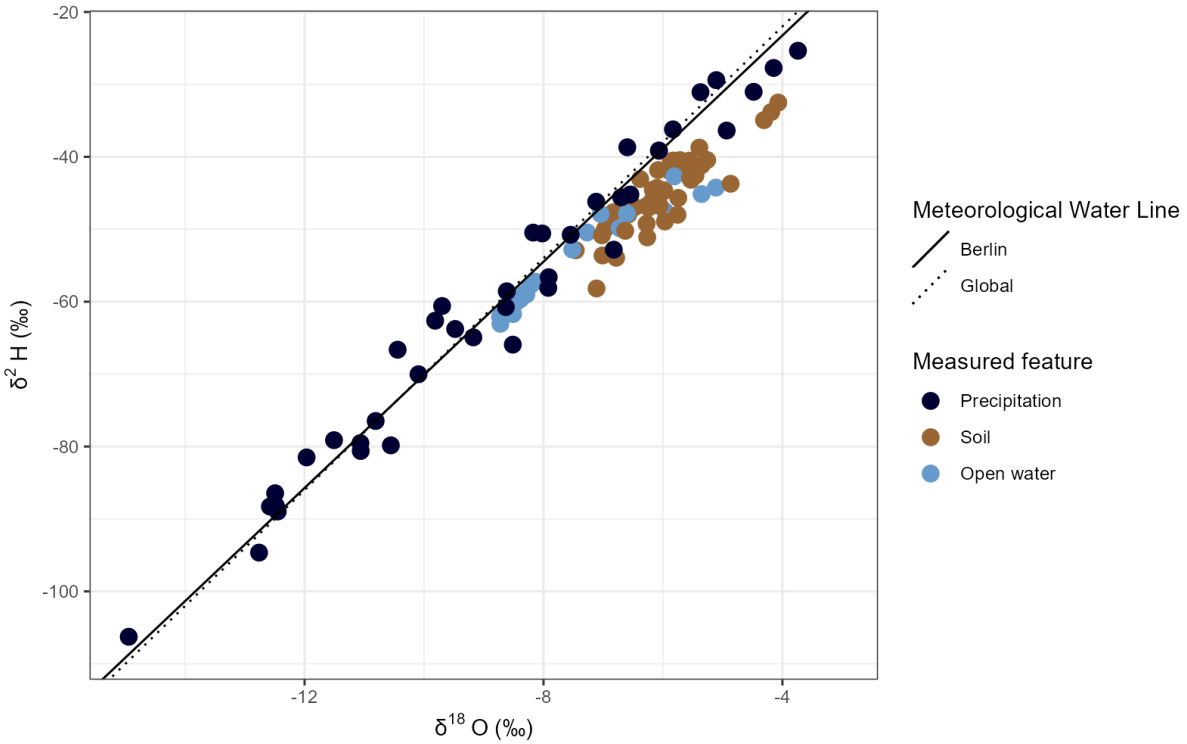


Figure 7: Overview of water signature of ^{18}O and D along the Wuhle in autumn 2022 for precipitation, soil water (labelled as "Soil") and open water.

5.2.2. Spatial aspect of volumetric water content and soil water isotope dynamics

The volumetric water content in soil changes with depth even in the upper 15 cm. Although the samples are from shallow soil, for the description of the data the samples from 0-5 cm are described as upper, the 5-10 cm as middle and the ones from 10-15 cm depth as lower sample or section. In Table 8, the soil measurements are separately written down for October and November, but all the site groups are averaged. For statistical analysis the measurements are tested against the grouped sites and environmental characteristics. In Figure 8, the volumetric water content (VMC, %) is shown. The handheld VMC measurements are combined with the calculated VMC from the upper soil samples. The soil samples for October and November are shown in different colours. The VMC data is not normally distributed which is why the Kruskal-Wallis-Test was used to test for significant differences. The Kruskal-Wallis-Test identified significant differences between the sample depths ($p < 0.0001$), the Dunn-Test showed that this difference is specifically between the upper samples compared to the lower samples (both with $p < 0.0001$). No significant difference was observed between the middle and lowest samples. No environmental characteristics showed an influence on the difference between the measured VMC. A simple t-test was done, and a significant difference was detected between the soil samples and the LMWL. This is also true between all the depths. No t-test was done for the individual sites as there is not enough data to make a sensible linear regression.

The mean difference of VMC content at the surface is ±2 % compared to the upper soil samples (Table 8). At site a, b, e, f, g and j the surface measurements show a higher VMC at the surface than the upper soil samples. Furthermore, there are interesting outliers, which are not visible in the data in Table 8. The samples WR2 (in November), WR3 and WR7b at site b, c and f were the most noticeable differences in the resulting surface and upper soil sample VMC. The surface measurements were over 4.5 % wetter than the upper soil sample.

In Figure 8 the upper point represented as 0 cm depth is the mean soil moisture content, calculated from the handheld probe measurements and the upper of the soil sample (0-5 cm). This value is usually higher than the calculated VMC of the middle and lower soil sampled (5-10 cm and 10-15 cm). Sites a, c, d and z have a wetter surface, a drier middle section and the lowest soil section has again a higher moisture content compared to the sample from the middle. This is also true for site b in October and for site g in November. For the averaged values, the sites a, c, d and z show a difference in VMC of 1.3-6.1 %. The biggest difference can be found between the first 5 cm and the middle section of the soil. Site a, c, d, e, f, g, and h are around 6% wetter at the upper section compared to the middle section. Site b, j and z are 10 % wetter in the first 5cm. Site h, sample WR14 is an outlier with less than 1 % difference in VMC. At site e, f, h, g (in October) and j, the VMC gets lower with depth. The difference in VMC is around 1.5 %, at site g (in October) and site j the difference was even higher with over 4 % drier middle soil samples than the lower soil samples. Sample WR14 (site h) has a nearly stable VMC throughout the 15 cm of soil measured depth.

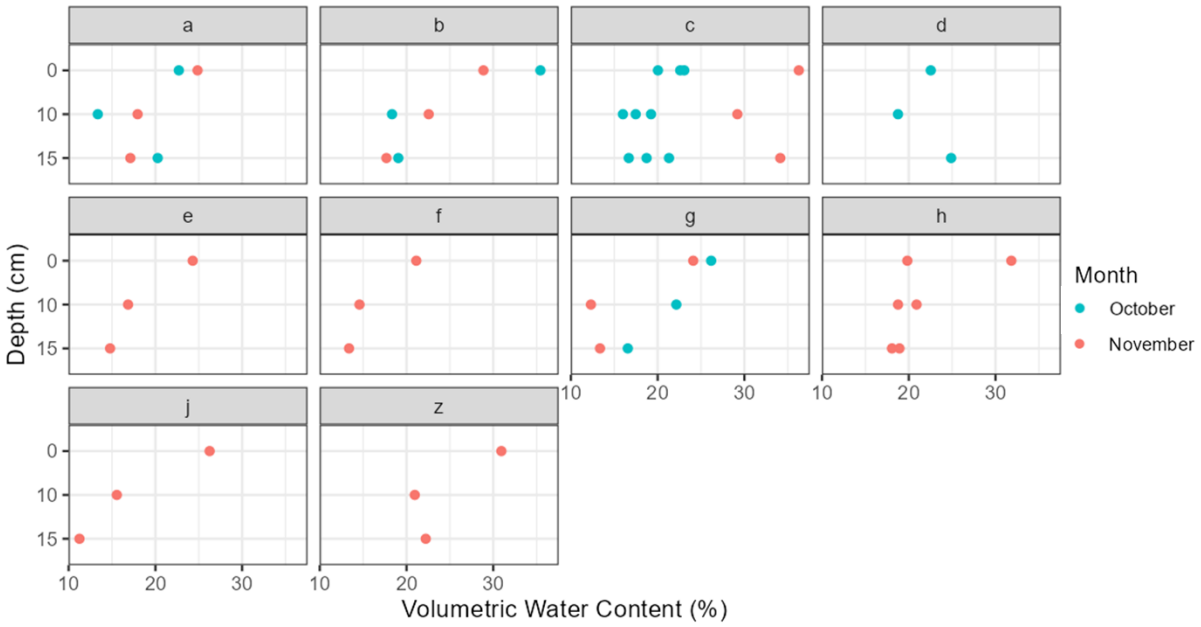


Figure 8: VMC per grouped sample in site (a-h, j and z) at each sampling depths (0-5, 5-10, 10-15 cm).

Table 8: Isotopic ratios, $\delta^{18}O$, δD and lc-excess, and VMC (%) values at the sampling sites for water bodies, surface grassland and soil samples at specific depth. For water samples the sample number WR is specified in the brackets. "Surface" are the handheld VMC measurements. For location orientation see Figure 3.

Site	Sample (n)	Soil/Water	D (‰)		^{18}O (‰)		Lc-excess		VMC (%)	
		Depth (cm)	Oct	Nov	Oct	Nov	Oct	Nov	Oct	Nov
a	2	water (0)	-58.89	-47.49	-8.31	-5.97	-2.03	-8.88		
	1	water (1)		-59.91		-8.44		-2.03		
	3	surface							22.70	24.63
	1	0-5	-46.70	-46.81	-6.57	-6.28	-3.38	-5.79	20.36	22.84
	1	5-10	-52.90	-44.54	-7.46	-6.17	-2.63	-4.35	13.36	17.95
	1	10-15	-46.06	-42.18	-6.18	-5.74	-5.77	-5.38	20.26	17.11
b	1	water	-59.64	-60.05	-8.39	-8.46	-2.14	-1.97		
	3	surface							35.43	28.87
	1	0-5	-47.09	-50.03	-6.69	-6.97	-2.87	-3.56	30.05	26.76
	1	5-10	-47.96	-47.10	-6.58	-6.64	-4.58	-3.24	18.33	22.56
	1	10-15	-48.84	-47.65	-6.92	-6.83	-2.82	-2.29	19.05	17.68
c	1	water (3)	-44.25	-45.14	-5.11	-5.35	-12.31	-11.33		
	1	water (4)	-59.01	-59.49	-8.29	-8.38	-2.28	-2.03		
	1	water (5)	-60.01		-8.47		-1.84			
	9(3)	surface							21.90	36.28
	3(1)	0-5	-40.11	-40.78	-5.92	-5.77	-1.88	-3.75	25.56	31.37
	3(1)	5-10	-42.30	-40.45	-6.11	-5.72	-2.56	-3.81	17.57	29.19
	3(1)	10-15	-42.17	-38.73	-5.82	-5.39	-4.69	-4.63	18.9	34.14
d	1	water	-60.01		-8.47		-1.84			
	3	surface							22.53	
	1	0-5	-34.94		-4.31		-9.28		24.16	
	1	5-10	-33.82		-4.19		-9.12		18.76	
	1	10-15	-32.51		-4.07		-8.72		24.88	
e	1	water	-52.79	-59.38	-7.52	-8.40	-2.07	-1.77		
	3	surface								24.30
	1	0-5		-58.18		-7.11		-10.63		22.38
	1	5-10		-46.93		-6.47		-4.41		16.85
	1	10-15		-46.95		-6.62		-3.26		14.78
f	1	water	-50.44	-59.46	-7.27	-8.39	-1.67	-1.91		
	3	surface								21.13
	1	0-5	-38.78	-45.67	-4.62	-5.74	-10.71	-8.82		16.58
	1	5-10	-32.73	-42.58	-3.70	-5.46	-11.85	-7.97		14.58
	1	10-15	-29.96	-42.27	-3.47	-5.58	-10.83	-6.69		13.37
g	1	water (8)	-47.90	-62.11	-7.04	-8.73	-0.89	-1.95		
	1	water (9)	-47.82	-49.86	-6.60	-6.73		-5.32		
	3	surface							26.17	24.10
	1	0-5	-44.61	-48.93	-5.98	-5.97	-5.91	-10.32	19.74	25.70
	1	5-10	-43.17	-44.79	-5.53	-5.99	-7.97	-5.98	22.15	12.31
	1	10-15	-40.44	-47.00	-5.26	-6.24	-7.36	-6.26	16.56	13.35
h	1	water (13)		-61.64		-8.53		-3.03		
	1	water (14)		-42.70		-5.81		-5.32		
	6	surface								29.04
	2	0-5		-52.53		-6.52		-9.59		29.64
	2	5-10		-48.48		-6.34		-6.94		18.22
	2	10-15		-48.84		-6.61		-5.18		14.65
j	1	water		-61.71		-8.51		-3.24		
	3	surface								19.83
	1	0-5		-48.00		-5.76		-11.04		18.98
	1	5-10		-49.23		-6.27		-8.23		18.76
	1	10-15		-53.62		-7.01		-6.85		18.93
z	1	water		-57.24		-8.14		-1.69		
	3	surface								30.93
	1	0-5		-43.72		-4.87		-13.71		31.98
	1	5-10		-42.85		-5.55		-7.54		20.95
	1	10-15		-41.15		-5.35		-7.35		22.22

5.2.3. Spatial aspect of soil and open water isotope composition

The dual-isotope plot shows the results of isotopic composition with $\delta^{18}O$ on the x-axis and δD on the y-axis. The catchment position is indicated by the colours and the depth of the sample is visualised by the shape (see Figure 9). All per site averaged samples are below the LMWL. Some soil measurements of a site were spread out others were more clustered. All calculated lc-excess values were negative for soil water. Water samples from flowing stream sites have less negative values than water samples from small ponds or disconnected artificial ponds. The isotopic data is normally distributed, and the variances are equal. Because the sample size is small and the assumptions of normality could be violated, the Kruskal-Wallis-Test is used to detect differences. Further the t-test only allows two levels in data being tested, which is not the case for the environmental factors that are considered.

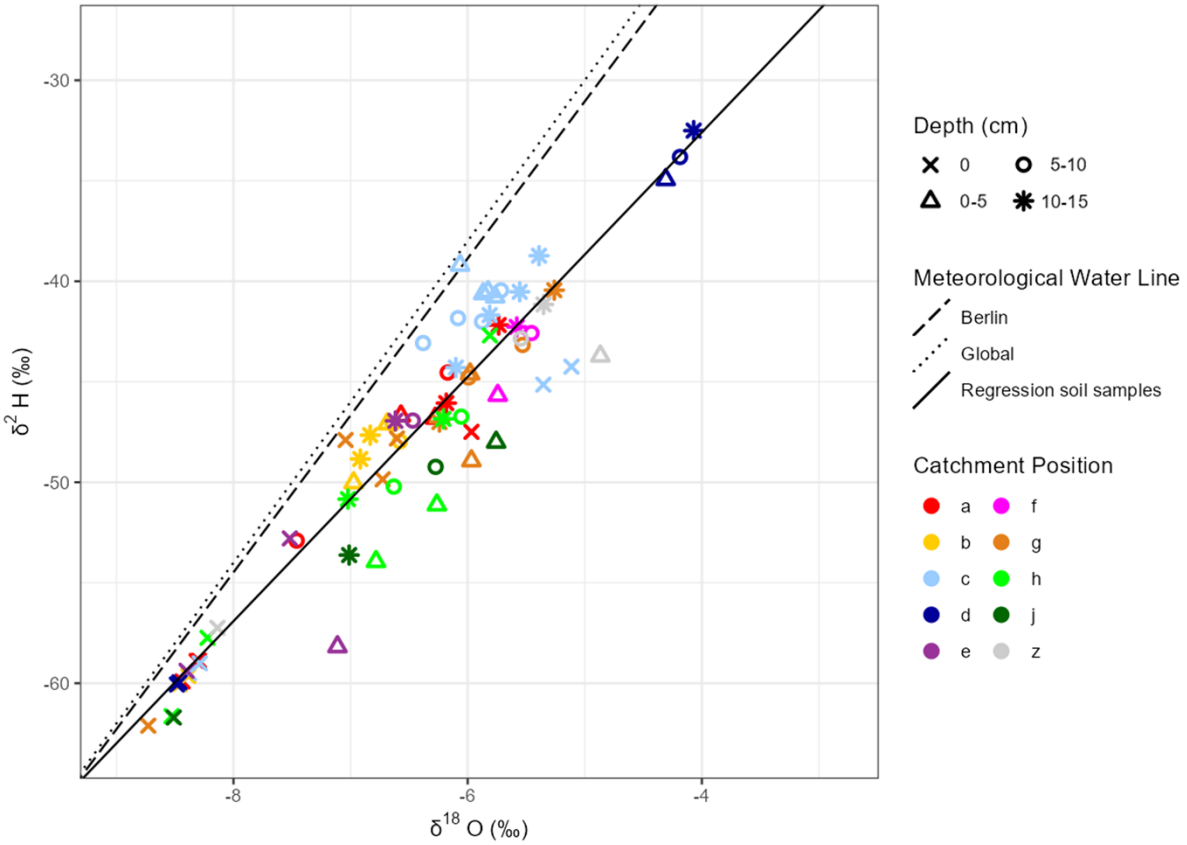


Figure 9: The Isotope composition including the Global Meteoric Water Line (GMWL; Craig, 1961) (dotted), Local Meteoric Water Line (LMWL; Hughes and Crawford, 2012) (dashed). Depth 0 cm signifying water samples (black cross). The solid line shows the regression of all the soil samples.

Not many of the differences are statistically significant, meaning a Kruskal-Wallis-Test p-value of <0.05 and a post-hoc Dunn-Test with an adjusted p-value of < 0.05. Differences detected between the subcatchments (as categorised in the modelling part) in δD are as following: W4.4 (including site h and j) is significantly different from the other catchments. Regarding $\delta^{18}O$, the subcatchment W4.4 (site h and j) is significantly different from HG (site z) and W4.3 (site f and g).

5.2.3.1. Water site specific results

Most of the water samples are clustered. They are depleted and close to the LMWL apart from the soil samples. Less depleted, but following the LMWL, are samples from site e, f and g from October. Next to them are six samples that are clearly departing from the LMWL along an evaporation line. These samples are from site a, g, h and c. The samples are from ponds or artificial ponds which potentially

are disconnected from the stream, where the water has a slower transit time. Noticeable is site a (Figure 9; red) as it was taken just at the outlet of the stream before it flows into the Spree. For the different water sites a significant difference was found for streams that were flowing and streams that passed through a pond or an artificial pond. The Dunn-Test returned significant differences for δD and $\delta^{18}O$ both with a value of $p < 0.01$. The simple t-test revealed no significant difference from the flowing water samples to the LMWL, but it did return significant difference of the pond samples to the LMWL. No environmental characteristics or other influences were tested.

5.2.3.2. Soil site specific results

When examining Figure 9 the measurements for site a are off set from the LMWL but spread along the regression line of the soil samples. The upper samples (0-5 cm) are close to each other. Comparing the middle samples (5-10 cm) the one from October is a lot more depleted than the sample from November (-52.9 ‰ δD, -7.46 ‰ $\delta^{18}O$ and -44.54 ‰ δD, -6.17 ‰ $\delta^{18}O$ respectively; see Table 8). The lowest soil sample (10-15 cm) is again more enriched.

Site b is clustered in all sampling depths ranging from -50.03 ‰ δD, -6.97 ‰ $\delta^{18}O$ to -47.09 ‰ δD, - 6.64 ‰ $\delta^{18}O$.

Site c has the most samples taken in proximity. The upper samples are clustered, the deeper samples start to spread out more, get more enriched in $\delta^{18}O$ and have a stronger offset from LWML. The middle section is the most depleted in δD. The lc-excess is less negative in the upper sample and gets more negative in the lower samples (-1.65 ‰ to -2.79 ‰).

Site d is also more clustered and one of the most enriched samples with values around -34 ‰ δD, - 4 ‰$\delta^{18}O$ and -8.4 ‰ lc-excess.

Site e has a much more depleted upper soil sample than the deeper samples (-58.18 ‰ δD, -7.11 ‰ $\delta^{18}O$ compared to -46.9 ‰ δD, -6.5 ‰ $\delta^{18}O$). The lc-excess is also more negative for the upper section, -10.68 ‰, as compared to the deeper samples, -4.30 ‰ and -3.19 ‰ respectively.

Site f is clustered with the upper sample being the most depleted. Only the sample WR7b was used for the statistical analysis, the removed soil sample WR7 was taken from considerably more sandy soil. The measurement was noticeably more enriched, a clear outlier, which is why it was removed from the results, as it is not comparable with the other measurements.

Site g is quite spread out again. The two upper soil samples have the same $\delta^{18}O$ value (-5.97 ‰) but the sample from November is more depleted regarding δD (-47.9 ‰ compared to -48.93 ‰). Aside from the upper soil sample from November, the other samples are close to the LMWL. The deepest sample from October was the most enriched and the most depleted was the lower and upper soil sample from November.

Site h has a big signal variability. Sample WR13 does not have a big signal change with depth in δD but is clearly more depleted with depth in $\delta^{18}O$. The lc-excess value gets less negative with depth (-10.06 ‰ to -4.02 ‰). Sample WR14 has a very depleted upper sample compared to the deeper samples, which are both approximately the same (-53.94 ‰ δD, -6.78 ‰ $\delta^{18}O$ compared to around -46.8 ‰ δD, -6.1 ‰ $\delta^{18}O$). The lc-excess is less negative with depth (-10.77 ‰ to -6.88 ‰).

Site J is removed from the LMWL and spread out with the surface sample being the most enriched and getting more depleted. The lc-excess changes with depth and gets less negative (-8.93 ‰ to -6.20 ‰).

Site z has a more depleted $\delta^{18}O$ value in the middle upper sample, but it is simultaneously more enriched in δD. The middle sample is more depleted than the deepest soil sample, which follow roughly the trajectory of the LMWL but with an offset. The lc-excess changes with depth and gets less negative (-13.24 ‰ to -7.00 ‰).

5.2.3.3. Environmental characteristics site specific

For the evaluation of the environmental characteristics of the sites, raster data in 250 m radius around the sampling points was considered (see Table 2). The results for environmental characteristics showed

the same results for total tree percentage and broad leaf trees. Coniferous trees are less abundant. Some significant differences were detected simply because of their absence. If those statistics were not confirmed in another comparison (e.g. between the individual samples), the results of differences between the sites based on coniferous tree abundance were neglected. If results are similar between the broadleaf and coniferous trees or between clay and sand composition, they are referred to as tree and soil composition.

The strongest difference, which is seen in δD, $\delta^{18}O$ and lc-excess, is between site b and d ($p < 0.01$). Lc-excess showed no additional differences between the samples across the environmental compositions. Compared to the percentage of pixels with trees, the Dunn-test shows significant differences in δD values between sample WR6 (site d) with WR2 (site b), WR13 (site h), and WR15 (site j). Further differences were found in WR3 (site c) with WR2 (site b), WR13 (site h) and WR15 (site j). Samples WR6 and WR3 have over 45 % of trees, whereas samples WR2 and WR13 have below 40 % and WR15 below 30 % (see Table 2). For $\delta^{18}O$ values, significant differences are found for sample WR6 (site d) with WR2(site b) and WR6c (site e; 46 %) and between WR2 (site b) and WR16 (site z; 15 %).

Regarding the soil composition δD, the resulting differences were the same as in tree composition with the addition of WR4 (site c) with WR13 and WR14 (site h). Here sample WR3, WR6 and WR4 have soil composed of above 91 % sand and below 5 % clay, while WR13, WR14 have 77% sand and 8% clay, WR2 and WR15 have 89 % and 85 % sand respectively and around 4 % clay. For $\delta^{18}O$ the same samples as in tree composition are different, with WR6 being different from WR2 and WR6c (92 % sand, and 2.5 % clay). And WR2 is different from WR16 (site z; 79.8 % sand and 7.4 % clay).

The impervious area for WR6 is 30 % and WR3 20.4 % compared to the samples with significant differences in δD with WR2 (50 %), WR13 (43.8 %), WR15 (18.8 %). For $\delta^{18}O$ the differences are between WR6 with WR2 and WR6c (31.9 %), and WR2 with WR16 (24.6 %).

5.2.3.4. Depth specific results

For an understandable visualisation of the isotopic composition with depth, a heatmap is used to illustrate the dynamics (see Figure 10). The simple t-test was used to check the difference of the depth to the LMWL and linear regression of the samples from 5-10 cm and from 10-15 cm depth show that they are not similar. No significant differences were detected in the three depth sections with the Kruskal-Wallis-Test on their own, except for samples in November between 0-5 cm and 10-15 cm. In the VMC result, significant differences between the upper section compared to the middle or lower section were detected. Therefore, the isotopic composition was examined for two depths excluding one sampling depth, to check if the same is true for isotopic signatures (e.g. 0-5 cm samples were compared to 10-15 cm samples, excluding the samples from 5-10 cm depth). No significant differences are found when the whole sampling data is used. From visual inspection of the heatmap, there is a noticeable difference between the sites a, b and c and the other grouped samples. This is why another test was conducted without the sites a, b and c. Between those pairings, significant differences resulted from the Kruskal-Wallis-Test. The Dunn-Test was not conducted, because the factor of interest is depth which is present in all sites.

For the upper section samples (0 - 5 cm) combined with the middle section samples (5 - 10 cm), there are no differences in the δD or $\delta^{18}O$ values considering depth. However, there are significant differences in δD regarding soil composition. In $\delta^{18}O$ values differences between tree and soil characteristics are significant. When lc-excess is investigated, no environmental differences can be detected but there are differences regarding the sampled depth.

For the middle section samples (5 - 10 cm) and the lowest section samples (10 - 15 cm), there were no differences for lc-excess. Significant difference is detected regarding δD and $\delta^{18}O$ with all environmental factors investigated (tree and soil composition, as well as impervious surfaces).
For the comparison of the upper and lower section samples (0 - 5 cm and 10 - 15 cm) to environmental characteristics δD has no significant differences. Regarding $\delta^{18}O$ significant differences can be found in tree and soil composition. There are significant differences in lc-excess.

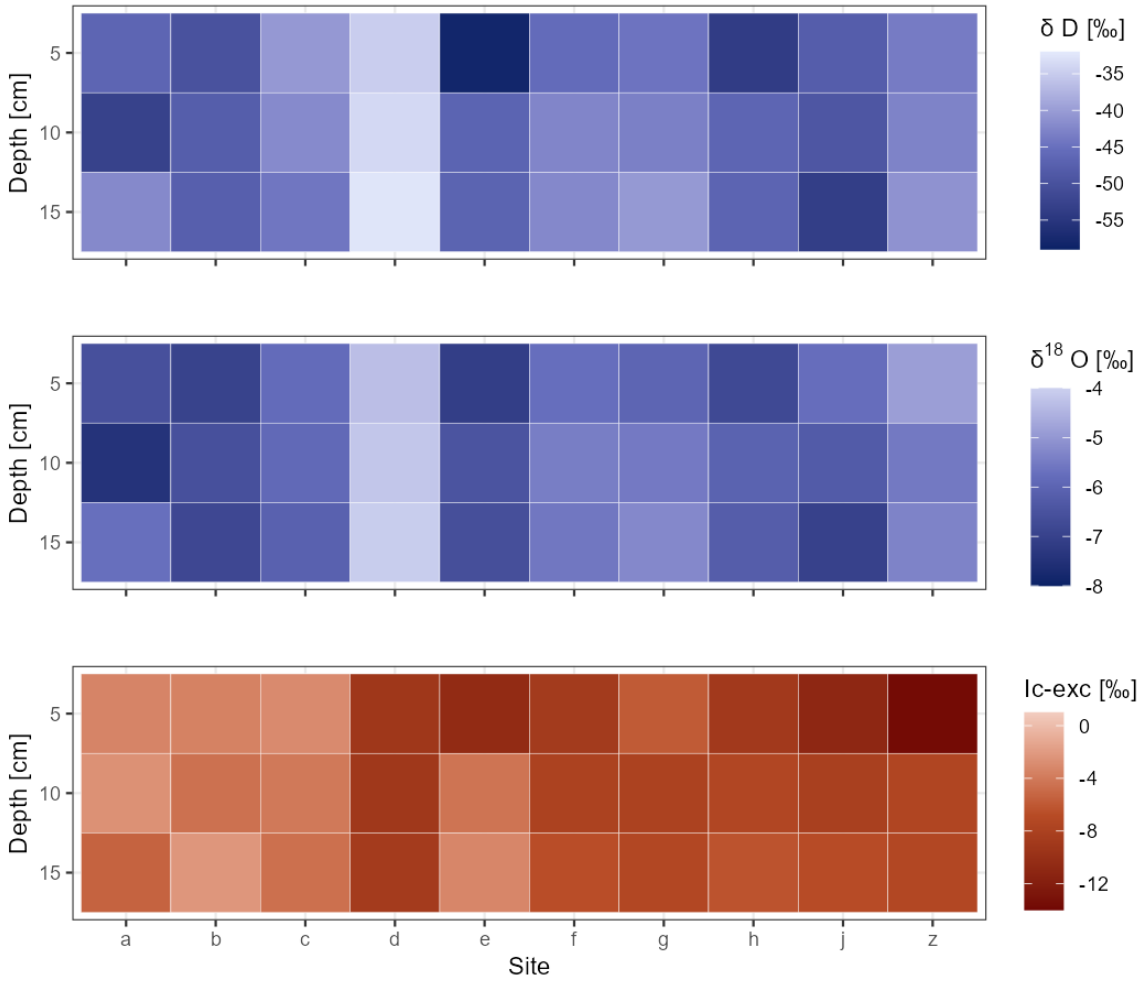


Figure 10: Heatmap of the sampled sites for each depth for δD, $\delta^{18}O$ and calculated lc-excess (in red). Site a located at the outlet and site g the last sampling group after the confluence of Alte Wuhle, Neue Wuhle and Hellersdorfer Graben. Site h and j are located close to the Neue Wuhle and site z close to the Hellersdorfer Graben.

5.3. Modelling

Because of equifinality, the term "best" parameter set is put in quotations. For individual parameters, all the returned values from the 100 best performing parameter sets are analysed and compared. For an overall assessment of the calibration and validation success, model performance is evaluated with an ensemble mean. Although the resulting parameter values can only be viewed as indexes rather than absolute values (Bergström, 1991), the comparison between the parameter values are quantifiable.

5.3.1. Calibrated parameter

5.3.1.1. Discharge data

To take a closer look at the differences between the catchments all the parameters are plotted in boxplots for the individual parameters (see Figure 11 and Figure 12) and as dotty plots (see appendix 9.7 and 9.8). In simulation A, the most defined parameters for all the catchments are FC, LP, PERC and

especially K2 and MAXBAS. For catchment W4.0 big ranges can be found for parameters TT, CFMAX, K1 and Alpha with skewed data for TT, CFMAX and Alpha (see appendix 9.7). W4.3 has more defined parameters except for K1. The medians in W4.3 are more centred compared to the other two catchments. Catchment HG has noticeable broad ranges in the snow routine parameters TT, CFMAX and SFCF, with the median value close to the box boarders (between the first and third quartile).

Because the resulting parameters are not normally distributed, and there are more than two groups to be compared, the Kruskal-Wallis-Test (Kruskal & Wallis, 1952) was done with a post-hoc analysis using the Dunn- Test (Dunn, 1964). The tests indicated significant differences for all resulting parameters (see appendix 9.5, Table 13). For simulation A, all the catchments have significant differences for the parameters FC, LP, PERC, K2, MAXBAS and Cet (p-value at least >0.01). The parameter Beta has no differences between the (sub)catchments. For the remaining parameters, only one catchment is significantly different from the others. For W4.0 it is SFCF and K1, for W4.3 it is TT and for HG it is CFMAX and Alpha.

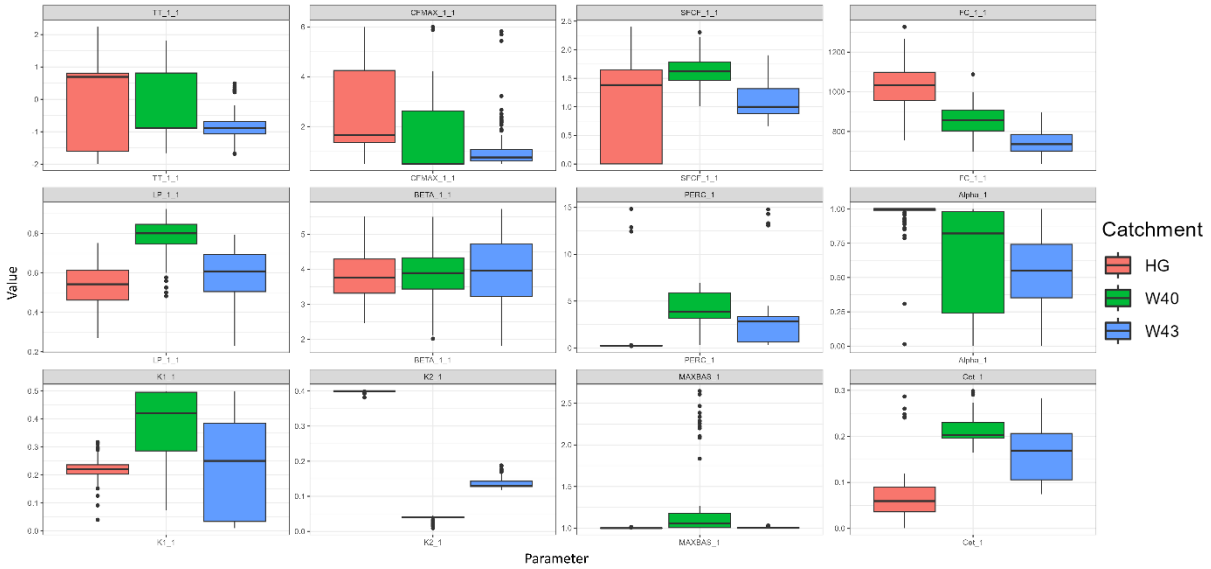


Figure 11: Boxplots of parameters for the discharge data calibration of simulation A for catchment HG (rose), W4.0 (green) and W4.3(blue).

For simulation B, the parameters PERC, K2 and MAXBAS are well defined for all the catchments. W4.0 and W4.3 have a well-defined Cet parameter, whereas HG has a well-defined TT parameter. However, it is HG that has wide ranges in SFCF, Alpha and Cet, with skewed data indicated by the median close to the third quartile border.

All the catchments have significant differences in the parameters FC, LP, PERC, K1, K2, MAXBAS and Cet. Alpha is also significantly different for all catchments, but it has to be mentioned that the p-value between W4.0 and W4.3 is only 0.03, whereas all the other p-values are below 0.005. Furthermore, W4.3 has a significant difference in the calibrated parameter for Beta, whereas HG has significant differences in TT, CFMAX and SFCF.

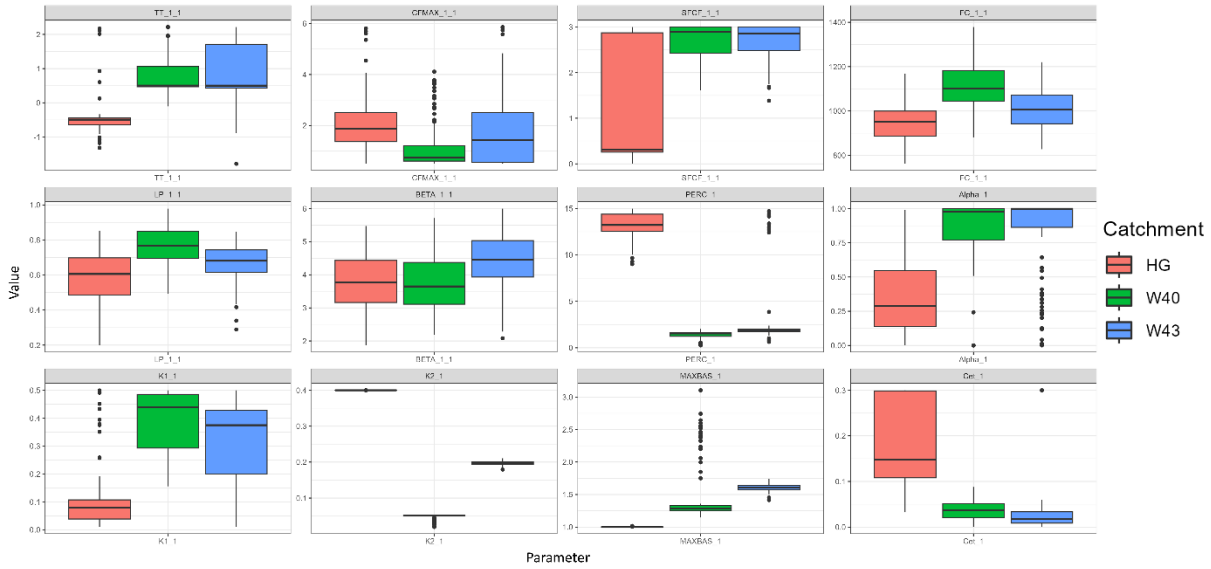


Figure 12: Boxplots of parameters for the discharge data calibration of simulation B for catchment HG (rose), W4.0 (green) and W4.3(blue).

5.3.1.2. Water level data

The boxplots of the calibrated parameters for the water level data are bigger compared to the discharge data (see Figure 13 and Figure 14). For simulation A, the widest range of possible parameter values in all (sub)catchments can be found in parameter Alpha (see Figure 13). Regarding the individual (sub)catchments, the biggest ranges for catchment W4.0 are the parameters CFMAX, SFCF, Beta and K1. For W4.3, only parameter Alpha has a slightly bigger boxplot. For W4.4, parameter TT, SFCF and K1 have big ranges, and for HG the biggest range can be found in the parameters TT and Alpha. The most well-defined parameter for all (sub)catchments is K2. For the whole catchment W4.0, the range of the parameter value FC is very small. In W4.3, most of the parameters are well defined with the above noted exception of parameter Alpha. Subcatchment W4.4 has a small, calibrated parameter range for FC, Beta, MAXBAS and Cet. In subcatchment HG, the parameter SFCF is well-defined as well as LP, Beta, MAXBAS and Cet.

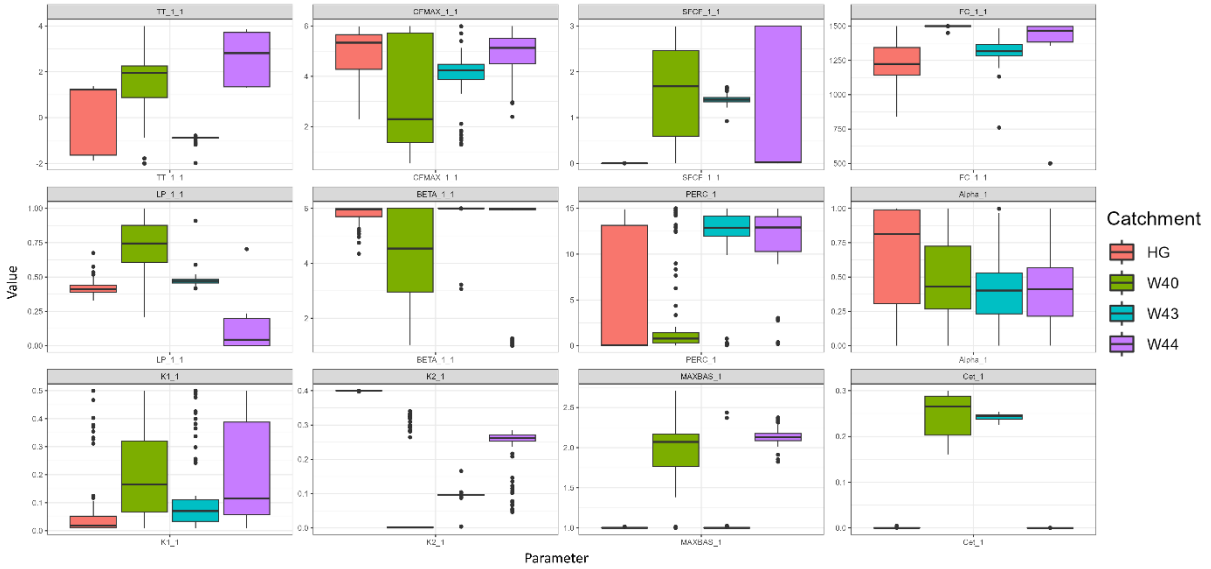


Figure 13: Boxplots of parameters for the water level data calibration of simulation A for catchment HG (rose), W4.0 (green), W4.3(blue), and W4.4 (purple).

Statistically, only the parameter LP is significantly different in all (sub)catchments (p-value <0.001, see appendix 9.6, Table 14). Catchment W4.0 is significantly different from the other catchments regarding the parameter values of TT and FC. For the parameter CFMAX, W4.0 is significantly different from HG and W4.4 but not from W4.3. Additionally, catchment W4.0 is different from subcatchment W4.3 regarding Beta. W4.0 has a clearly defined PERC parameter which is different from W4.3 and W4.4. For the parameter MAXBAS, W4.0 is similar to W4.4 but different from subcatchment HG and W4.3.The parameter TT and FC are similar between the catchment W4.3 and HG. In the calibration of Beta solely, catchment W4.3 is significantly different from all the other catchments. For CFMAX, W4.3 is different from W4.4 and HG, but not from W4.0. For K1, W4.3 is different from the other catchments. Subcatchment W4.4 is significantly different from the other catchments regarding the parameter values of TT and FC. For CFMAX, W4.4 is different from W4.0 and W4.3 but not from HG. Catchment W4.4 is significantly different in the calibrated parameter ranges SFCF, LP, K2 and Cet from the other catchments (p-value < 0.05 for SFCF, otherwise p < 0.0001).

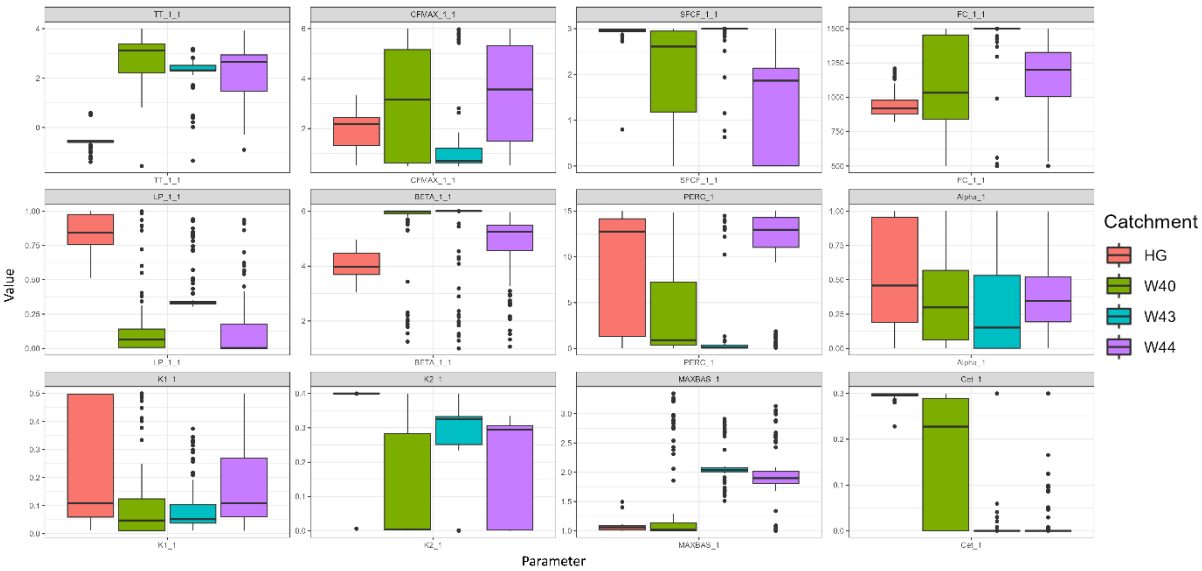


Figure 14: Boxplots of parameters for the water level data calibration of simulation B for catchment HG (rose), W4.0 (green), W4.3(blue), and W4.4 (purple).

The calibrated parameter ranges in SFCF, LP and Cet are significantly different in the subcatchment HG from the other subcatchments. For CFMAX, HG is different from W4.0 and W4.3 but not from W4.4. HG has a huge range for the parameter PERC and is significantly different to W4.3 but similar to the other (sub)catchments. For the parameter Alpha, only HG has significant differences in varying degrees (p-values of < 0.05, < 0.01 and < 0.0001 for W4.0, W4.4 and W4.3 respectively). For K1, HG is more significantly different from the other catchments, than W4.3 (p-value of < 0.05 and < 0.01, whereas in HG all p-values are < 0.0001). For the parameter MAXBAS, subcatchment HG and W4.3 are similar to each other but different from W4.0 and W4.4 which are in contrary similar to each other.

For simulation B (see Figure 14), W4.0 has also a big boxplot for the parameters CFMAX, SFCF, FC, K2, and Cet. In W4.3, only Alpha has a big boxplot. In W4.4, big ranges were calibrated for the parameters CFMAX, SFCF, FC and K2. Additionally, HG has big range for PERC, and K1. A well-defined parameter for all (sub)catchments is MAXBAS. W4.0 has a well-defined Beta value. W4.3 is well defined in nearly all parameters except for the above-mentioned Alpha. W4.4 is the least defined catchment with only a clear MAXBAS and Cet value. For HG the parameters TT, K2 and Cet are well-defined. All catchments have significant differences in parameters SFCF and Beta (for Beta the difference between W4.0 and W4.3 is lower with p-value of <0.05, see appendix 9.6, Table 14).

For the parameter TT, there is a significant difference between W4.0 and W4.3 (p-value <0.01). Regarding parameter Alpha, W4.0 does differ from HG and W4.4 but not from W4.3. As for Cet in the case of W4.0, it does significantly differ from the other catchments. In subcatchment W4.3, parameter CFMAX is significantly different from the other catchments (p-value <0.001). W4.3 differs from all the other catchments in LP, FC, PERC and K2. For the parameter Alpha only, W4.3 and HG have a significant difference (p-value of < 0.0001). The parameter MAXBAS does differ in W4.3 from all catchments. W4.4 is significantly different from the other catchments regarding CFMAX (p-value of <0.01). The parameter MAXBAS does differ in W4.4 from all catchments. For the parameter TT, the catchment HG is clearly different. HG is significantly different from the other catchments regarding CFMAX, except from W4.0. HG differs from all the other catchments in LP, FC, PERC and K2. As for Cet in the case of HG, it does significantly differ from the other catchments.

5.3.1.3. Comparison of discharge and water level data

The discharge data and water level data are calibrated separately and with different objective functions. The resulting parameter ranges are very different. For simulation A in the catchment W4.0, only LP is statistically similar. In W4.3, none of the parameters match and for HG, TT is similar. For simulation B in catchment W4.0, FC and PERC are similar. Further, there is a difference in Cet, although not as strong as with the other parameters (level of p-value <0.05, whereas the others are <0.0001). In W4.3, the parameter CFMAX is statistically similar. HG has slightly more similar parameters with TT, CFMAX and FC. Because of this, the calibration and validation are not comparable between the two data types. The results of water level data are not further described or discussed in this thesis. The figures and tables of the benchmark results for water level data can be found in the appendix 9.11.3.

5.3.2. Model uncertainty

5.3.2.1. Calibration of simulation A

The dotty plots display the relation of the parameter value and the performance with the objective function NPE. For 1'000 calibrated dotty plots with the Monte-Carlo approach, all the catchments show the same pattern. Most of the calibrated parameters show a wide range of performance from -4 to maximum of 0.0.776. Because the best performances are hard to read, the dotty plots were confined into a NPE range of 0 to 0-0.8 NPE. Most of the parameters are not well defined. W4.0 has a slight clustering along a linear line for the parameter LP, where higher LP values have better performance values. Further, there is a skewed plot for K2 with the highest value between 0.01 and 0.1 Δt^{-1}. W4.3 has a similar clustering for LP as W4.0. This similarity is not visible in parameter K2, as it has some bad performing values around 0-0.5 but reaches a plateau that has the same uncertainty as the other parameters in W4.3. The dotty plots for HG also have uncertain parameters but with smaller ranges. Here, just the parameter K2 shows a clear positive trend with a higher K2 resulting in a higher NPE performance.

Next, the dotty plots with the calculated 100 best GAP calibrations (see appendix 9.7, Figure 20, Figure 21 and Figure 22) are consulted. First, the catchment W4.0 (the whole catchment) is discussed (see appendix 9.7, Figure 20). The NPE ranges between 0.795 and 8.300. It is clear that the parameter TT and CFMAX are defined into two groups, one of them with a higher NPE value (around -1.0 °C and around 0 mm t^{-1} °C^{-1}, respectively). This is visible in the parameters FC, LP and Beta as two lines are visible in the dotty plots which indicate uncertain parameter values but at a set NPE performance. There are some outliers that are at the top of those lines. The dotty plot of SFCF looks similar to FC. Both have a clear line-cluster where a range of parameter values perform similarly. These clusters are influenced by the parameter TT and CFMAX. K2, MAXBAS and Cet each have a clear cluster of similarly performing parameter values and some outliers which have a higher performance. PERC, Alpha and K1 all have one column of high performing calibrations (between 0 and 1 mm Δt^{-1}, 0 (-) and around 1.3 Δt^{-}

^{1} respectively) and then a plateau of a broader parameter range with lower performing calibrations. There is a hint of negative trend in SFCF, FC and K2 for the outliers where, the higher the value of the parameter gets, the lower is the performance due to the outliers.

As for Catchment W4.3 (see appendix 9.7, Figure 21), The NPE range is between 0.780 and 0.810. The calibrated TT value is more distributed with a big range of parameter values getting good performance values. There are two performance lines visible in LP and Beta with a negative trend of higher values performing better but a plateau at the top. SFCF, FC, Alpha and Cet are not well-defined parameters with a visible cluster around the NPE value of 0.800 for all parameters. PERC has a visible cluster at the minimum valid value of 0 mm Δt^{-1} and another cluster with lower performance around 3 to 4 mm Δt^{-1}. K1 has also a visible line cluster around the NPE value of 0.800 but all the higher performing calibrated values are around 0 Δt^{-1} which is the minimum valid value for K1. K2 has a cluster around 0.13 Δt^{-1} with outliers of higher value linearly improving performance. MAXBAS has a lot of values around 1 Δt with higher values following the NPE performance of 0.800.

Regarding the dotty plots for catchment HG (see appendix 9.7, Figure 22), the NPE range is between 0.750 and 0.780. There is also a clear clustering with TT, CFMAX and SFCF in two groups, but both have high and low performance values. FC seems to be evenly distributed. PERC, Alpha, and K2 cluster at the edge of the calibration ranges (0 mm t^{-1}, 1 (-) and 0.4 Δt^{-1} respectively), which for PERC is the lowest minimal valid value. This can also be seen in MAXBAS but less extreme, value 1 Δt being the minimum valid value. LP and Beta both have a positive trend with higher values also performing better. Cet does cluster to the valid minimum value (0 $°C^{-1}$) and shows a negative trend with higher Cet values getting lower performance values.

5.3.2.2. Calibration of simulation B

The NPE performances for simulation B are all lower and the trends are less curved (see appendix 9.8, Figure 23). For W4.0 the NPE ranges are from 0.55 to 0.615. TT has two clear clusters at around 3.4 °C and 2.1 °C that have an NPE value of 0.55 and 0.57 respectively. The third and fourth group are more scattered, but one is around the NPE value of 0.595 with 1 °C and the other cluster would be around 3.2 °C, but between the two there is some more noise. This results in horizontal cluster lines in CFMAX, SFCF, FC, LP, Alpha and K1. The parameters Beta and MAXBAS also have a division in their points with the clearer highest NPE value around 6 for Beta and 1 Δt for MAXBAS which is also the minimum valid range for this parameter. PERC has two clusters between the values 12 and 15 mm Δt^{-1} and at the higher NPE value the values switch to the minimum valid range around 1 mm Δt^{-1}. K2 has a defined values for the respective NPE values at around 0.28, 3.2 and 0 Δt^{-1}. Cet has one cluster column around the value 0 $°C^{-1}$ and then two cluster lines with the most values around 0.3 $°C^{-1}$.

For the catchment W4.3 (see appendix 9.8, Figure 24) the NPE ranges are 0.78 to 0.86 with the higher performance with TT around 0.5 °C and the lower performance from 2.2 to 3.2 °C. The parameters that have a higher NPE value are more scattered, the lower NPE value parameter is more clustered for all calibrated parameters. CFMAX has values on either side of the range for the lower performance but most of the higher NPE parameter are around 0 mm Δt^{-1} $°C^{-1}$. SFCF, FC, LP and Beta have a scattered range with better NPE values, while lower NPE values are clustered at one end of the set range for calibration. PERC has most of the calibrated parameters around 0 mm Δt^{-1} and some points cluster with NPE of 0.78 around 13 to 15 mm Δt^{-1}. Alpha and K1 are divided by the NPE values but do scatter with no clear values. K2 has the best performance with 0 Δt^{-1} and lower performance with values from 0.2 to 0.4 Δt^{-1}. MAXBAS has one cluster with lower performance around 1.5 to 2.2 Δt and the values scattered around 2.6 to 2.9 Δt has a better performance. The parameter Cet has column cluster at 0 $°C^{-1}$.

For Catchment HG (see appendix 9.8, Figure 25) the NPE values range from 0.56 to 0.64. TT has a wide range for the lowest NPE values, for NPE 0.6 its clusters around 1.5 °C and for the highest NPE values around 0.64 TT clusters around 2.9 °C. CFMAX, FC, Alpha and K1 have very uncertain parameter values that are grouped in three line-clusters. SFCF has three distinct groups all with different parameter values with the highest NPE value from 1.5 to 2.5. LP, Beta, PERC and MAXBAS have some outlier points in lines but also freely scattered that are difficult to group but clear clusters around the highest NPE values. The points cluster around 0 for LP, between 5 and 6 for Beta, between 12 and 15 mm Δt^{-1} for PERC and around 1.9 Δt for MAXBAS. K1 has a column cluster at 0 Δt^{-1} and the higher performing cluster around 0.3 Δt^{-1}. Cet has some scattered outliers but most of the points cluster at 0 $°C^{-1}$.

5.3.3. Model performance

Simulation A used the data chronologically, first the warm-up period, then the calibration and lastly the validation. Simulation B used the latest data as calibration data set. The earlier data is used as both, the warm-up and validation data. As the warm-up data is not used for calibrating the model, this is not an issue. To evaluate the performance, the biggest discharge events were filtered for the validation periods of each catchment and simulation. For simulation A, all the catchments have the same observed discharge event on the 21.02.2022 and 22.02.2022. For simulation B and catchment W4.0 and W4.3, the observed discharge event was on the 21.01.2008 and for catchment HG the events on the 24. And 26.07.2017. The catchments were plotted for the whole hydrological year the event was taking place in.

5.3.3.1. "Best" Validation

Because the individual boxes of the model cannot be plotted for an ensemble mean, the "best" result from the parameter set is described first (see appendix 9.9, Table 15 and Figure 26-Figure 30). For this, the best validation results are chosen, regardless of how good the calibration was compared to the other 100 calibration results. All calibrations in all catchments perform better in simulation B for the calibrated NPE value, but the validation NPE are higher in the simulation A for catchment W4.0 and HG (see appendix 9.9, Table 15). For all catchments, the discharge simulation is mostly regulated by the soil box and the lower groundwater box. The upper groundwater box is rarely activated. This leads to an overestimation of the baseflow und underestimation of the peaks in all catchments. When water has accumulated over a longer precipitation period, the peaks are overestimated and because the lower groundwater box empties slowly, it leads to an overestimation of the baseflow for an extended period. The PET is bigger than the actual evapotranspiration (AET) in all simulations and time periods looked at during summer. HG as the smallest catchment has the quickest response in the lower groundwater box and therefore has less overestimations over a long period of time.

5.3.3.2. Validation Ensemble Mean

The ensemble mean takes all the 100 best results from the GAP optimization and calculates the mean discharge from all those simulations. NA values from observed discharge data were omitted. The resulting objective functions for the validation period can be seen in Table 9. Because the model was calibrated on NPE, those values are all positive. For simulation A, the hydrological year from 1.10.2021 to 30.09.2022 is selected to showcase how the simulation fits the observed discharge as visible in Figure 15. The simulation for W4.0 misses the peaks which is why the R_{eff} is negative. The timing of the simulated peaks follows the observed data, although some lower peaks are not plotted. The biggest volume error is quite big, as the simulated baseflow has a lag to get back to the level of the observed baseflow. W4.3 has a better R_{eff} value although the peaks are sometimes also overestimated. The timing is a little worse than with W4.0 but the water balance (VE) is better. For HG, the R_{eff} is a little lower than W4.3 but still positive. The timing of the simulation in HG is the best of the three and so is the water balance.

Table 9: Objective function values for discharge data. Shown is the validation period per catchment and simulation. The negative values are marked in orange.

Catchment	W4.0		W4.3		HG	
Simulation	A	B	A	B	A	B
NPE	0.684	0.704	0.622	0.625	0.709	0.193
R_{eff}	-0.071	-2.314	0.399	-0.599	0.289	-0.833
r_s	0.726	0.725	0.629	0.738	0.768	0.706
VE	0.542	0.596	0.396	0.387	0.259	-0.284

For simulation B, the hydrological year from 1.10.2008 to 30.09.2009 is selected to showcase how the simulation fits the observed discharge as visible in Figure 16. The NPE values of W4.0 and W4.3 are better, but the peaks are missed more frequently. This leads to a greater VE value for W4.0 but it is smaller for W4.3. The timing of the peaks is nearly the same for W4.0 and a lot better for W4.3. For HG, the simulation looks fitting, but the NPE value is a lot lower than for all other catchment in both simulations. R_{eff} and VE values are negative. Generally, the timing of simulation B is worse than for simulation A.

The objective functions solely for the showcased hydrological years can be found in appendix 9.109.10, Table 16.

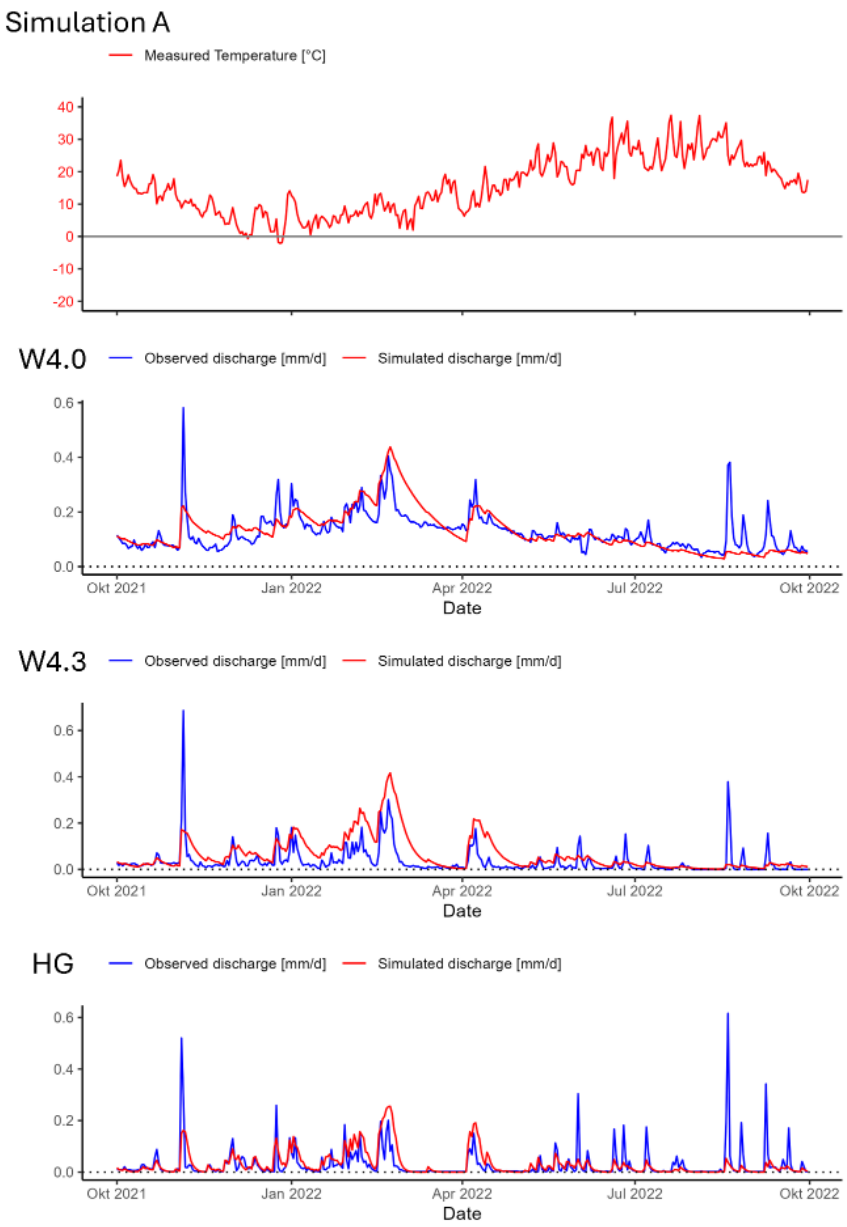


Figure 15: Example of simulation results during the hydrological year of 2021 to 2022 for all (sub)catchment in simulation A. The simulated discharge (red) is calculated from the ensemble mean of the validation period. The observed discharge (blue) shows the true value. The corresponding objective function values can be found in appendix 9.109.10, Table 16.

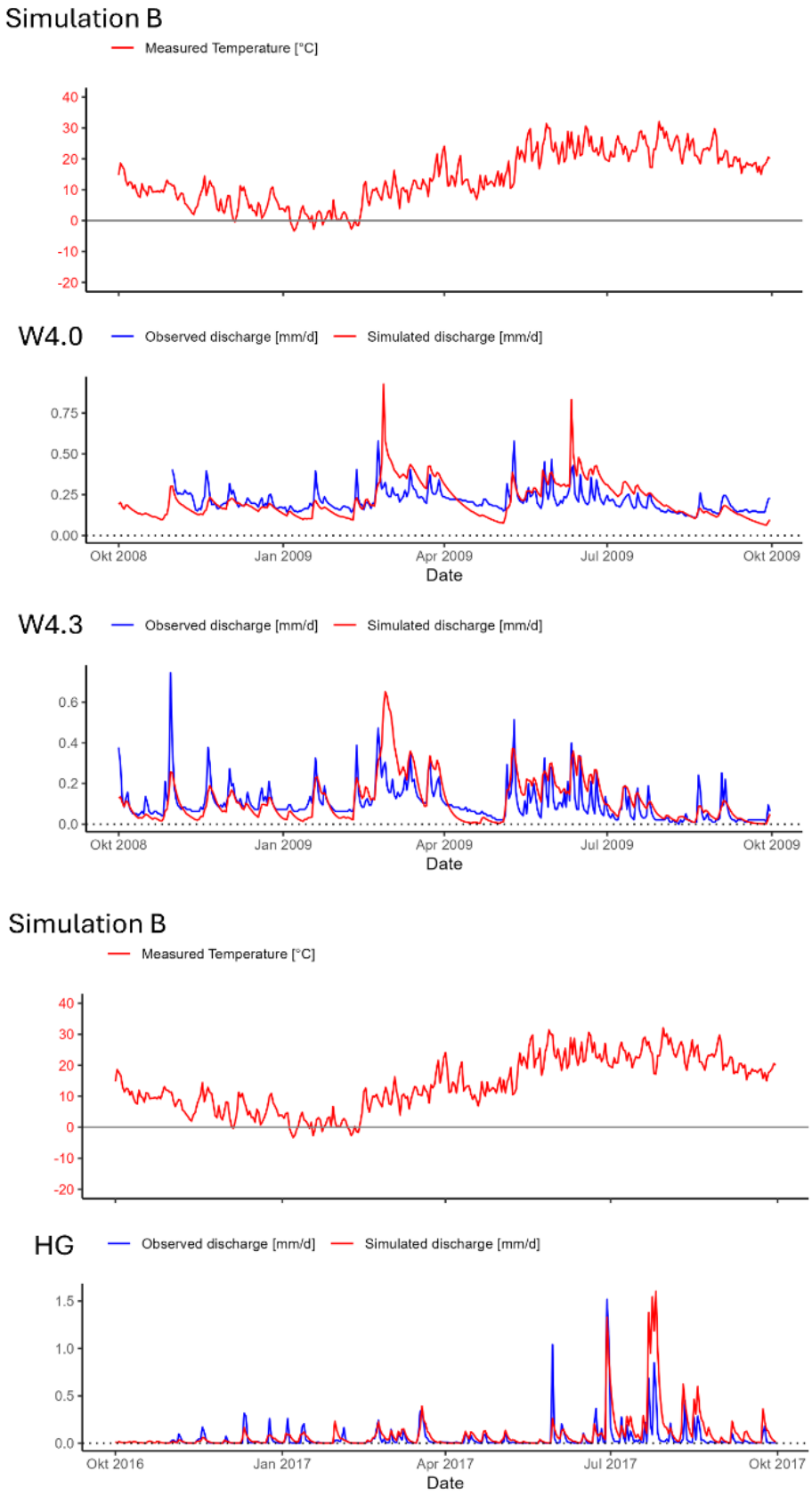


Figure 16: Example of simulation results during the hydrological year of 2008 to 2009 for W4.0 and W4.3 catchment in simulation B (upper plot) and the year from 2016 to 2017 for catchment HG (lower plot). The simulated discharge (red) is calculated from the ensemble mean of the validation period. The observed discharge (blue) shows the true value.

5.3.3.3. Benchmarks

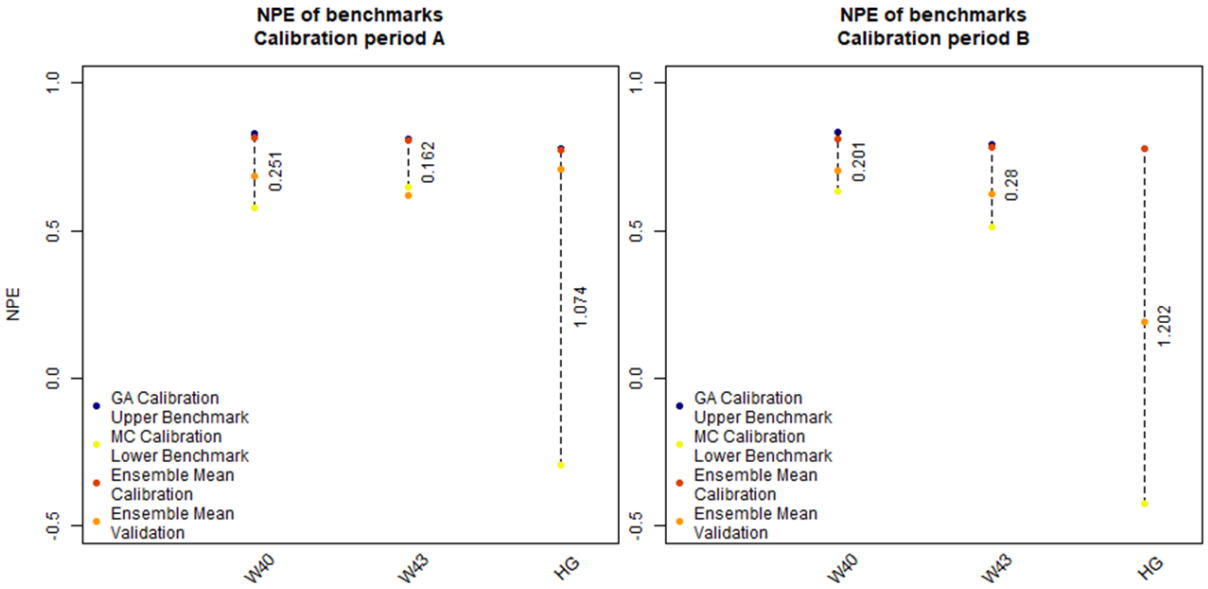


Figure 17: Benchmarks and results of calibration period for discharge data for each catchment with the NPE as the objective function. Simulation A is on the left, simulation B is on the right. The numbers describe the difference between lower benchmark, calculated with the Monte Carlo approach (yellow) and the upper benchmark calibrated with the GAP approach (blue) the best calibration result and the ensemble mean (red).

For a better evaluation of the validation results, the upper and lower benchmark are plotted against the ensemble mean calibration and validation (see Figure 17 and Table 10). It is clear from Figure 17 that the calibration for the catchment HG is the most difficult to model, as the difference between the upper and lower benchmark is enormous. W4.0 and W4.3 are easier to calibrate. The only case, where the validation is visibly bad is in W4.3 for simulation A, as it is lower than the lower benchmark. The best validation can be seen for HG in simulation A. A comparison of the other performances is easier with the relative objective function in Table 10. The higher the relative objective function value, the better is the overall validation. Therefore, simulation A is better than simulation B for W4.0 and HG. For W4.3 the opposite is the case.

Table 10: Summary of objective function NPE values for all catchments calculated with the discharge data for the simulation A and B. The upper benchmark is calculated with the GAP optimisation, the lower benchmark with the Monte-Carlo approach. The ensemble mean is calculated once for the calibration and once for validation period. The relative objective function describes the relation between the difference of validation ensemble mean and the lower benchmark to the difference of upper and lower benchmark. Negative values are marked in orange.

Catchment	**W4.0**		**W4.3**		**HG**	
Simulation	**A**	**B**	**A**	**B**	**A**	**B**
Upper benchmark	0.829	0.836	0.811	0.79	0.78	0.778
Ensemble mean lower benchmark	0.578	0.635	0.648	0.511	-0.294	-0.423
GAP ensemble mean calibration	0.816	0.812	0.808	0.783	0.775	0.777
GAP ensemble mean validation	0.684	0.704	0.622	0.625	0.709	0.193
Relative Objective Function	0.422	0.343	-0.160	0.409	0.934	0.513

6. Discussion

6.1. Field work planning and methods

During the available time for field work, a limited number of samples could be taken on two field trips. Few sites were visited twice. Especially the upper catchment has a limited number of measurements. In addition, the soil samples were only taken at one point for each sample in three depths. Taking more than one sample per sampling point and averaging the results would reduce the uncertainties introduced by human error and device measurement error. After the first field trip, the bags were left sealed for one night, and only in the following morning, were they filled with dry air. As an equilibrium is assumed after 48 hours, this could have caused an imbalance as some air might have escaped while the equilibrium was not reached. No recognizable differences caused by this imprecise handling of the samples were statistically detected for samples that were taken in both October and November.

The above-mentioned reasons make the resulting measurements pool limited. Therefore, the displayed results and discussed trends can merely be indications of processes that could take place. As the data is solely from autumn, seasonal signals may be present in the data. In papers about isotopic composition in Berlin, the comparison of soil samples with depth, were calculated with geometric means for plot and depth replicates (Kuhlemann et al., 2021). In this thesis, this was neglected due to limited data. If data points were compared a simple mean was used.

In the statistical analysis, the data is normally distributed. Nonetheless, the Kruskal-Wallis-Test was used. Even though a t-test would be more robust since it only compares two levels, it was not enough to analyse the available data. Because the sample size is small, validated normality of the data could be wrong in comparison to a larger data set. A multicriterial analysis with extensive research of environmental factors, such as radiation, shading or irrigation was outside of the scope of this thesis and the limited time in Berlin. If the available data showed no significant differences between grouped data and the known environmental factors, they could still contribute but not be as important in distinguishing sites from each other as other factors. A cross-correlation with Spearman rank correlation coefficient was not feasible as more data points are needed and for this study exclusively soil samples under grassland were sampled. For a cross-correlation independent data under tree or shrub would have been necessary, but that was already excessively studied in Berlin (Kuhlemann et al., 2022; Marx et al., 2022; Ring et al., 2023).

6.2. Signal changes throughout the Wuhle catchment

The autumn of 2022 was drier than the long-term mean. Isotopic composition from summer to autumn became more variable and negative, which was also seen in other studies done in Berlin (Kuhlemann et al., 2020; Ring et al., 2023). Due to the limitation of data from autumn particularly, the seasonal effect on the water and soil isotopes cannot be shown. Just a small amount of rain fell between the two sampling days, but samples taken in November are more depleted than in October in the first 5 cm and in site g in all sampling depth. Precipitation is closer to the samples taken from the flowing stream than to standing water and soil samples. A comparison between October and November was not possible, as just a few samples were taken on both field trips and because the uncertainty of the improper handling of the probes in October. The negative lc-excess is indicative for evaporative isotopic fractioning compared to precipitation as written by Ring et al. (2023). This was the case for all samples taken. Shifts between samples not following the LMWL show non-equilibrium fractioning through evaporation, which can be also seen visually in the regression line (Figure 9). There seems to be a difference between the lower catchment and the upper catchment regarding lc-excess, as shown in the depth specific results. However, this test does reduce the samples tested from 78 samples to 27 samples, meaning it merely is a trend and can be used as hypothesis for further studies.

6.2.1. Water content

The VMC is higher on the surface and in the upper 5 cm of the sampled soil than in the deeper soil. Interestingly, some of the samples show drier middle sections than lower sections, which is an unexpected dynamic in the shallow soil. Although no significant difference was detected between the sites, some have a steeper slope of wetness (sites h, sample WR13 and site j) and others are more homogeneous (site h, sample WR14). No connection to external influences was found to explain the differences between the sites. Possible explanations mentioned by Marx et al. (2022) are local conditions which seem to be almost equally important as differences in vegetation cover: radiation, humidity, irrigation, subsurface drainage, or run-on from impervious surfaces. What was not considered during sampling was noting the distances from trees and the amount of litter from trees that would interrupt evaporation processes. Another important factor could be compacted soil due to foot traffic outside of the designated path, which makes infiltration into the first few centimetres of soil slower. In the future, it would have been interesting to compare the data with more soil moisture monitoring data from the Berlin Senate (Pflanzenschutzamt, 2020) that is available as daily measurements since mid-2023.

6.2.2. Spatial variability of isotopes

The flowing stream isotope measurements were clustered and separated from the pond open water and soil isotope measurements. There are significant differences between stream water and open data. Not enough data is available to make a meaningful statement about the difference between artificial ponds and parts of the Wuhle that flow into smaller ponds which are still connected to the stream network. Nor is the information available, where effluent discharge is coming into the Wuhle, but it would have been an interesting point to consider. Kuhlemann et al. (2022) mention urban storm drains highly impacting the stream during precipitation events. The locations of those are not available, but merely a map with the amount of water collected by area (Gerstenberg, 2019). The amount of data did not allow to analyse the connection between the shallow soil and the Wuhle stream, which does not indicate noticeable divergence as found in the study by Kuhlemann (2022). The differences found between the catchments categorised in (sub)catchments by the modelled areas are likely due to the small number of samples. The catchment W4.4 has just one sample spot, as well as HG. And the other two catchment would have all the sampling spots included in them, which were not used in the calculations. For this, a new stream would have to be sampled and modelled, for example the Eichwaldgraben, which was not possible. The grouping as part of catchments is not well thought out, as the subcatchments W4.4 and HG are part of W4.3 and W4.4. Yet, in statistical tests, they were split from the bigger catchments. However, the overarching question in this thesis is if the upper, less urban samples differ from the lower more urban samples. The difference between W4.4 and HG in δD and $\delta^{18}O$ are valid, as they belong to different tributaries. Furthermore, the reasoning to divide the Wuhle into smaller subcatchment for the PET data was that local processes on the surface happen presumably more vertically instead of horizontally. Therefore, soil samples of W4.3 do not influence the soil samples from the catchment W4.0.

6.2.3. Differences in sites and subcatchments

The decision of taking a circle radius of 250 m around the sampling point was based on the raster size of the soil data in GIS of 500 m. The alternative to categorize the sites by the (sub)catchments, as used in the PET data, would have been too coarse as percolation processes seem to be very site specific (Marx et al., 2022). The difficulty in the characterisation lies in how big the circle around the sampling sites should be drawn. In this thesis, all of the investigated environmental characteristics were based on an equal distance from the sampling point. However, soil composition might influence the soil more locally. A proper soil texture analysis could have been interesting, although it would defeat the purpose

of finding catchment dynamics with simple methods. Impervious area could be considered with a larger distance to gauge if there is potentially run-on from sealed surfaces during precipitation events. This question cannot be properly analysed and answered with the limited amount of data for this thesis. The approach of testing differences between the isotope values and site characteristics in the catchment provided some insight into the possible influences. All found differences were mainly caused by the isotopic composition and not by the major differences in environmental characteristics, as the found differences were the same in all considered factors.

The biggest difference between samples was found in WR2 and WR6 which have around 10 % difference in tree abundance. WR2 and WR16 have a difference of over 15 %. The smallest reported difference in δD is between WR6 and the site WR6c 0.7 % and in $\delta^{18}O$ with only 6 %. This would suggest that at least for these two cases, tree cover is not the most influential factor. Regarding soil composition, WR13 and WR14 seem to be denser than WR6 or WR3 with 14 % less sand and 3 % more clay. Water cannot percolate through the soil as easy. The same is true for WR16 compared to WR2. And D is more affected by evaporation than ^{18}O which explains the bigger D depletion but a more enriched ^{18}O. WR2 and WR15 seem more similar to WR3 as well as WR6c compared to WR6. Here, other factors influence the sites as the differences in soil composition are small. The impervious area has again quite a gap between the WR6 and WR3 samples and the significantly different samples of WR2, WR13 and WR15. This might be the factor influencing the sample difference between WR2 and the samples WR6c and WR16. With more sealed surfaces in WR2 by 20-25 %.

Subcatchment 2 (site c, d, e) differs the most from other sites. This is due to site d, which is enriched in both δD and $\delta^{18}O$ compared to the other sites, and site c which is more depleted in $\delta^{18}O$. Both of those sample groups are around the Wuhlebecken. The soil samples in site c are more depleted in $\delta^{18}O$ but a little more enriched in δD than the open water samples of the Wuhlebecken. Overall, it seems that this site is more connected to the rainwater and was less influenced by evaporation processes during the sampling period than other sites. This could be due to run-on from nearby impervious areas. An explanation why site d is removed from the cluster of the other soil samples is not detectable, since the environmental characteristics are defined on a subcatchment level. From the pictures taken during sampling the only noteworthy feature would be more shrub surrounding the area opposed to trees that were more generally present in other sampling sites. Here a more detailed notation about distances to as well as sizes of vegetation would have been interesting.

No significant differences were detected between the sampling depth separately. This could have two meanings. First, the sample size is too small to see outliers or second, there are no outliers. In a visual inspection of the dual isotope plot, focusing on the depths individually, it seems that site a, b and c have a unique signature in the upper sampled section. This is especially visible when examining the lc-excess. Statistical tests can only be used as trends because the sampling pool is significantly smaller, when excluding the lower catchment. Stronger isotopic fractioning is found in the upper catchment with depth, as seen in the results of the depth specific results regarding lc-excess. No difference is found between samples from 5-10 cm and 10-15 cm, but strong differences were found between 0-5 cm compared to 5-10cm and 10-15 cm (both $p < 0.01$). If the whole catchment is examined no statistically significant differences are found in the lc-excess between the depths. However, if a comparison between two depths examines the δD and $\delta^{18}O$ values and focuses on the environmental factors over the whole catchment some of the found differences are stronger ($p < 0.01$) compared to the other results ($p < 0.05$). This is the case for samples with 0-5 cm and 5-10 cm depth in site c and h for δD and site b and d $\delta^{18}O$. For samples from 5-10 cm and 10-15 cm depth site d and site j show stronger differences. Lastly, for samples from 0-5 cm and 10-15 cm depth site h has stronger differences to site d and c in δD and regarding $\delta^{18}O$ site b and site z.

However, the hypothesis that differences between the sites regarding depth would be pronounced in all sites, especially between the samples of 0-5 cm and 10-15 cm were not proven to be true.

Kuhlemann et al. (2022) found that grassland has fast infiltration rates despite high evaporation losses in exposed shallow soil. From the results of this sites from the upper catchment (d, e, f, h, j and z) show very negative lc-excesses (< -9 ‰) indicating high evaporation losses, even in autumn. Most of the lc-excess get less negative with depth indicating that the water is mixed with older soil water, which corresponds with soil being generally wetter during autumn (Kuhlemann et al., 2022).

6.3. Modelled hydrological dynamics

6.3.1. Model settings

The HBV model was developed for rural ungagged catchments and was not used on urban areas yet. Because simple models are sometimes better to model processes than models that focus too much on physical accuracy, this conceptual model was tested on an urban stream flowing into the city. Although the PET values were calculated by the HEC-HMS Model and therefore the modelling process was not based exclusively on conceptual modelling, it was just the tool used for the calculations. Furthermore, simpler equations could be used to estimate PET, including e.g. simple seasonal or annual sine curves regardless of the variation in weather (Beven, 2012; Federer et al., 1996).

The HBV model can distinguish between various vegetation elevation units which in the flat Berlin was not applicable. Testing all different variants of the model would have been outside of this thesis' scope. It would have been interesting to compare the results of the standard basic model with the one groundwater box variant. The disadvantage of the HBV model in an urban setting prior to modelling seems to be the lack of an outflow option in the soil routine that would indicate impervious areas and direct runoff during precipitation.

Since the size of the Wuhle catchment is variable between GIS data and official documents, it introduces uncertainties about the actual precipitation and the specific discharge for the subcatchments. Another reason for uncertainties is that precipitation over urban areas is highly spatial. Therefore, a second or multiple meteorological station to compare and smooth out the amount of rain over the area would have made a more stable data set. Another variable input is temperature. Even though the catchment of the Wuhle is not as affected by the UHI as the city centre (Menberg et al., 2013; Vulova et al., 2020), the denser urban areas would experience some higher temperatures which can deviate from the measuring station Berlin-Marzahn and influence AET. Because the PET data was divided in subcatchments, for the HBV model it had to be recalculated into one value for each day of the year. The long term mean PET was calculated from the years 2011 to 2021 which includes the heatwave of 2015, and droughts of 2018 and 2019. Because PET is not used as yearly input but as a mean, this does inflate the PET for all modelled years. Here, it might have been preferable to calculate the median of all years to see if the difference of the resulting values is negligible.

The small lakes and artificial ponds were neglected in the modelling process, as they cover less than 1 % of the whole catchment area. Lakes signify storage and evaporation. Something that could have been tested out would be to treat the impervious area as a lake corrected by some factor to indicate direct evaporation from sealed surfaces. The problem is that lakes would retain water whereas impervious area immediately responds to precipitation events.

6.3.2. Calibration process

To set the parameter ranges a lot of pre-runs were done to gauge which settings would be reasonable. As a result, a feasible range as proposed by Seibert (2000) was used as an initial setting. Since an urban

setting is different than a rural setting, this might have been a limiting starting point. Ranges were then made bigger to see if parameter that clustered on the maximal range border would scatter at some point. This led to some unusual high values for some ranges such as K2. This parameter was limited because of the logic in the model structure, that would not make sense to have a higher value in K2 than in K1. In other areas, previous experiences lead to a limitation of the ranges as in Alpha, which perhaps could have been set higher. Calibration takes a lot of calculation efforts, which is why at some point the trial calibration had to be stopped. The overall choice of parameter ranges was based on a visual inspection of the resulting dotty plots from pre-runs rather than on a parameter sensitivity analysis, which would have given an idea of (un)certainty of the parameters. As the pre-calibration was done with a reduced number of runs to save on computing resources, and the fixed ranges did not show many clear signals, the calibration process proved to be difficult. In retrospect, greater attention to this step would probably result in the calibrated ranges being defined more accurately.

Because snowfall does not play a big role in city environment, the refreezing coefficient and water holding capacity was set to their default values, especially because snow in cities is artificially influenced and melts quickly if it falls at all due to higher temperatures from UHI. For the same reason, the seasonality of the day degree factor is set to a default value. The parameters of the snow moisture routine are deducted quickly from the dotty plots. The large ranges in TT were caused by two different ranges in the catchment. HG showed a preferable range from 0-4 °C whereas W4.0 and W4.3 did perform well in the range of -2 - +2 °C. To avoid bias, the range was set wider, but equal for all (sub)catchments. The CFMAX parameter has a smaller range than the proposed ranges by Seibert (2000), which can be explained by the study contrasting sites of urban Berlin and two rural catchments in Sweden. The biggest challenge in the soil routine was the maximum storage in soil box, FC. The pre-calibrated values are rather unusual, and until now solely in Brazil an even higher FC value was found (around 25’000). Otherwise an expected FC is around 500 as seen in the paper by Seibert (2000). As mentioned in the HBV-light help section, the FC is a model parameter and not necessarily equal to measured values of ‘field capacity’, which is why the high values are unusual but not impossible. The percolation parameter has also quite a large range, which can be explained by the fact that the standard basic model variant does not have a direct runoff and consequently all the precipitation is relocated to the groundwater box. MAXBAS parameter ranges are rather big considering that the catchment responds quickly to rainfall.

The objective function of NPE was chosen. The R_{eff} has a focus on high flow. A calibration with this parameter would probably simulate the peaks better but end up with too high base flow and unbalanced water balance. The NPE is based on the objective function KGE both of which use three aspects of mean, variability and dynamics (Pool et al., 2018). The KGE uses a bias between simulated and observed mean discharge, the bias between simulated and observed standard deviation and the Pearson correlation. The NPE as a non-parametric variant replaces the mean discharge with the flow duration curve and the Pearson correlation is replaced by the Spearman rank correlation coefficient. As Pool et al. (2018) mention in their paper, the NPE produces better results than KGE in their study except when evaluating the magnitude and timing of high flows. After the evaluation of the hydrograph, a second calibration with the KGE would have been interesting to compare, to see if it would be a better fit but it is outside of this thesis’ scope. As it was planned to compare water level data, which was calibrated on the Spearman rank correlation coefficient, to the discharge data it did make more sense to calibrate on the non-parametric efficiency. Calibrating on multiple objective functions was dismissed, because the NPE is already taking into account three criteria and adding another objective function would have weakened the calibration.

6.3.3. Validation process

Two simulations were run (A and B) because the dataset of HG and W4.4 were too short to have 50 % split for calibration and validation to compare. Literature showed that calibrating on longer periods of dry conditions leads to better results for the whole timespan than calibrating wet periods to predict later dry periods (Motavita et al., 2019). This would lead to the expectation that simulation B results in better validation values, which is true for W4.0 and W4.3. HG performed noticeably worse in simulation B than A. However, in subcatchment HG (and W4.4) both simulations have the drought periods included in their calibration dataset. In simulation B the drought is in the middle of the calibration period and in simulation A it is the beginning of the calibration period. The posing question now is if the catchment could not recover after 2018. This would make simulation B calibrating on extraordinary data which cannot reflect pre-drought conditions. Regarding the R_{eff} value, all catchments are performing better in simulation A than B. The timing of the simulation reflected in r_s is similar for W4.0, better for W4.3 in simulation B and better for HG in simulation A. The water balance aspect for W4.0 is especially bad, but it has to be assumed that the direct runoff from sealed surfaces is the problem. This is also reflected in the R_{eff} value which is noticeably worse for W4.0 compared to W4.3 or HG in the same simulation.

In literature, the split-sample method is used to evaluate if a parameter set qualifies for further use, which was also criticised by Arsenault et al. (2018) for sacrificing information contained in the rest of the time-series. The split sample method has three premises, the first one being that this is a method for models whose outputs are intended for use outside of hydrology, meaning that the focus is that the results are sound but not to contribute to the understanding of the hydrological processes. The second is that the criteria for model performance should be correctness of hydrological variables generated by the model and not the structural adequacy, and the third is that the basis for judging a model performance is the comparison with the available observed data (Klemeš, 1986). As the focus of this thesis is the difference between the (sub)catchments based on their parameters, the validity of the calibration cannot be gauged with the split sample method. It does indicate if the simulated discharge is robust in both scenarios and can be used in further decision making. The “best” parameter set was only used to investigate the modelled boxes behaviour but not as a final result. To simulate the discharge, as described by Seibert and van Meerveld (2016) an ensemble mean was used to avoid bias in response of the problem of equifinality. The resulting NPEs were acceptable and similar for W4.0 and W4.3 in both simulations. The discrepancy between simulation A and B were too big in the calibration for HG under the criteria of the split-sample method and should be rejected.

To generate context on how good the validation performance is, upper and lower benchmarks are used. According to the relative objective function, the best performing ensemble mean are simulation A for W4.0 and HG and simulation B for W4.3. Since the overall performance of simulation A in W4.3 is better, it has been chosen as the simulation to compare parameter differences, even though the main objective function of NPE is slightly better in simulation B. This means that for catchment W4.3 a calibration with the Monte-Carlo approach with ensemble mean is equally valid as the chosen parameters for W4.3. To fulfil the final goal of comparing catchment parameters and extract some information about the differences between the whole catchment and subcatchments for all the (sub)catchments, simulation A is used for discussion.

The resulting parameter values are discussed as index and in relation to the discharge present in each (sub)catchment rather than absolute values, as mentioned in the results (Bergström, 1991). The assumption was, that the visible calculated ranges would have an order of parameter values such as HG-W4.3-W4.0 or reversed. This would conclude that HG has different parameters as a more

agricultural and rural subcatchment, HG is a part of W4.3 and has the bigger influence on W4.3 than W4.0 which represents the whole catchment.

The snow routine has a wide range parameter SFCF for HG, but the mean value is in the upper part of the boxplot, and a lot of values are clustered at 0, which can also lead to high value in NPE results (see appendix 9.7). Because of this wide range, statistically only W4.0 is different from the other two catchments. Therefore, the urbanised area has a higher snowfall correction factor, than more rural areas presumably because of the difference in impervious areas (see Table 1). For the other two parameters in the snow routine, W4.3 has a smaller range and therefore more defined values than W4.0 and HG. The TT and CFMAX are split into two clusters in W4.0 and HG. Regarding TT, only W4.3 is significantly different. The higher-ranking TT values in NPE for W.40 are below 0 and for HG it is evenly distributed. For the threshold temperature, below which precipitation is considered snow, it is lower in the urban shaped areas. This can be explained by the UHI effect with concrete structures radiating energy accumulated throughout the day and thus increasing the overall temperatures of surrounding air as well as surfaces. Following this logic, it is also clear that CFMAX is higher and statistically different in HG than in the urban areas. This means that between HG and W4.3 snow melt is quicker in W4.3. In some cases of calibration CFMAX is 0 mm $\Delta t^{-1}°C^{-1}$ in W4.0, which means that TT can be ignored in calculations for snowmelt (see Eq. 8). Presumably, no snow is modelled to fall in these simulations. If the CFMAX is in the upper range and according to the dotty plot (see appendix 9.7) combined with a higher TT, the snowmelt is very high. This duality in W4.0 could indicate the difficulty of the mix in urban and rural areas in one catchment. Nevertheless, the subcatchments with urban parts have higher melt than the more agricultural subcatchment.

For the soil moisture routine all parameters are distinguishable, except for parameter Beta. HG has the biggest maximum soil moisture storage (FC), which would be logical. Curiously, the W4.0 has a supposedly larger FC value than W4.3. However, LP value is also higher for W4.0, which reduces the E_{act} for W4.0 compared to W4.3 as it is multiplied by FC and then used for division with the soil moisture box content (see Eq. 10). This would suggest that evapotranspiration is higher in the less urbanised HG and simultaneously, that less water is lost through evapotranspiration in urban W4.0 which in turn means more water does infiltrate or runoff directly.

For the response routine, the parameter Alpha shows that the discharge for W4.0 and W4.3 has mathematically a broader bell-shape than HG (Eq. 11). K1 and Alpha are only used if the result of $K1 \times SUZ(t)^{(1+Alpha)}$ is smaller than SUZ(t) (see Eq. 11). From the output data of the HBV, it cannot be determined when solely SUZ(t) or when the calculation with K1 and Alpha is used. Theoretically, if the same amount of water would be put into the catchments, HG would have a bigger discharge from SUZ than W4.0 and W4.3. Whichever part of the equation would generate outflow from the upper soil box, but there is rarely enough precipitation to produce a simulation of Q1 in most cases (see appendix 9.7). K1 is a parameter with a big range in W4.0 and W4.3. It is visible in the dotty plot for W4.3, that the best NPE values are achieved with low K1 (see appendix 9.7). This explains the bigger activity of the SUZ in the simulations in W4.3 (see appendix 9.7). The discharge is only slowly retreating to the usual baseflow in W4.0. As the peaks are all generated with Q2, the water level is gradually declining. Therefore, the SUZ should be more active for W4.0 for the discharge peaks to be simulated more precisely. The problem is that those peaks are probably not caused by subsurface flow but by additional discharge from direct run-off from impervious areas. This runoff is not delayed by the soil moisture box and upper and lower boxes but is directly transformed form precipitation to discharge, which is why it cannot be simulated by HBV. For PERC, all the catchments are different from each other. In this simulation, PERC for HG is very small. A high PERC value makes the final discharge less flashy and the return to baseflow slower, a low PERC value has higher peaks und returns quicker to the baseflow. The

more urbanised the whole catchment is, the flashier are the observed discharge data. Therefore, the distinction of HG from the other subcatchments is not easily comprehensible. The smaller the PERC value becomes compared to the K1 value, the more the discharge Q1 of the upper zone will approach that of a singular reservoir (Eq. 11) (Bergström, 1976). In this simulation, HG has very low percolation from the upper to the lower groundwater box (on average between 0.2 and 0.3 mm Δt^{-1}), which would suggest a higher activity of the upper soil box. However, the total discharge in HG is so small, that only the lower soil box is active. Just one calibrated PERC value is small enough for the upper soil box to be active, which is the one in W4.3 (see Figure 15 and Figure 16). K2 is set very high for subcatchment HG compared to the other catchments, which again causes lower baseflow. A lower K2 produces a higher baseflow. For W4.0, this value is quite low compared to the other subcatchments, which probably what improves the water balance as the direct runoff cannot be modelled. In turn the dynamics of the peaks becomes less accurate, and the simulated discharge can merely follow the general behaviour of the observed discharge with delays in returning to the baseflow.

MAXBAS is low as expected from literature, as the Wuhle catchment response quickly to precipitation events (Kuhlemann et al., 2022). Cet is the potential evaporation correction factor, which has the highest value for W4.0. This could be due to the fact that HG was the only subcatchment PET was calculated for, the other values had to be transformed.

To achieve a better calibration, the parameter Alpha should be tested on a wider range. With the insights gained from this thesis, MAXBAS would be forced into a range of 0;1 even if higher values could achieve good simulations, to reduce calibration effort. The same would be true for FC. As it is unusually high, it could be rewarding to fix it to a certain value, if the model structure is not adapted. Another idea that would go beyond the scope of this thesis is to experiment with different vegetation units and use them to indicate impervious areas with very low percolation rates to force the upper soil box to be more active. One proposition to make the HBV model structure work for urbanised areas is to add a box for direct runoff, maybe as a percentage of precipitation that directly contributes to discharge according to the percentage of impervious area in the city.

6.3.4. Water level and discharge

The goal of modelling both water level data and discharge data were to compare the resulting parameter ranges. If the calibrated results were similar, it would have required one more measuring station to analyse. W4.4 is situated at the Neue Wuhle before its confluences with the Hellersdorfer Graben. This would have provided one additional subcatchment to evaluate the difference of rural and urban streams, as the two separate subcatchments HG and W4.4 do not mathematically influence each other. Because all parameters were not well-defined and water level data had been calibrated with the Spearman rank correlation coefficient due to the ordinal scale, the ranges of parameters were very wide and the dynamics between the (sub)catchments did not reflect the calibrated ranges from the discharge data. This is why the results of water level were no longer discussed after the initial parameter ranges results.

6.4. Comparison of isotope and modelled results

In this thesis, the isotopic measurements and the modelled discharge of the Wuhle are difficult to combine mathematically, as the HBV model is not working with isotopic data, nor was it developed to be. Because of the limited data (just one site in subcatchment HG), no trends could be uncovered in trying to combine the results of the two approaches in this thesis. The VMC showed a very wet 0-5 cm of shallow soil throughout the catchment, which would disagree with the modelled PERC value of 0-7 mm Δt^{-1} for W4.0. The same is true for HG which has a high PERC value compared to measured precipitation. However, as seen in the dotty plots (see appendix 9.7), the lower percolation values

performed better during calibration in all (sub)catchments. In sites where the upper 5 cm of soil samples are noticeable more depleted and closer to the LMWL than the others, it might be an indication for run-on from sealed surfaces, as they are closer to the precipitation signatures. This would mean that sites a, b and c could indicate additional dynamics which are not properly modelled in HBV. However, with this amount of data it can be just confirmation bias. In the water isotopic composition, no additional information can be extracted regarding the modelled (sub)catchments, as there are too few to translate it back to the modelling efforts.

6.5. Uncertainties and shortcomings

This project was born out of a collaboration between the university Berlin and Zurich, which presented a great opportunity to travel to another city and try out new research methods that were otherwise not available. It also ended up congregating two research approaches that are not easy to combine. This made this thesis a combination of two very interesting topics and methods to look at a catchment, but also resulted in available time being split between the topics.

6.5.1. Catchment information

The resolution of environmental characteristics of environmental factors was on a subcatchment level, that was provided by the IGB. More refined methods could be used to distinguish the samples from each other. However, additional assumptions would be made which introduce uncertainties. The soil, tree, and imperviousness of an area around the sampling spot could be defined as a 10 m circle around it, or smaller. Some sampling spots were close together, but the sites were in different subcatchments (site h in subcatchment 9 and site j in subcatchment 5). The Wuhle and Neue Wuhle swap sides, which was already discussed in the chapter before. It is not as consequential for this thesis but an important fact to keep in mind while evaluating data in that catchment. A significant number of the sampling points were close to footpaths. The compaction of soil or additional moisture sources from animal waste could also influence the measurement. This is part of sample taking in an urban setting, the study site is bound to be closely integrated in everyday life and there are no undisturbed areas. However, some sampling sites were merely a meter from the footpath, which makes external influences more likely. In a bigger sampling pool, more detailed notes on the location of the samples, such as distance to footpath, might be worth to investigate. A shortcoming regarding the environmental context in this study is that other potential significant environmental influences were not taken into account. Slope and aspect were excluded because Berlin is smooth overall, and all the sites were taken in flat areas. Elevated areas are missing in this region. The Wuhle catchment has three smaller hills, which would have been interesting to include. One aspect of the catchment that was outside of the scope of this analysis is the depth of unconfined groundwater level. Data is available in the water portal of the Senate in Berlin (SenUMVK, 2021b) and was shortly investigated during this research process. However, to reduce the range of topics in this thesis, the analysis was abandoned. Irrigation of parks was excluded as factor, since the samples were taken after the vegetation period.

A noteworthy uncertainty is also the water paths in an urbanised area, where water is potentially disconnected and redirected by pipes. The only information found was the mention of street names where rainwater retention basins are located (SenUMVK, 2021a). If maps of underground pipes in sealed surfaces exist, which would show the inflow of additional runoff into the stream, it would be a great indicator where additional challenges should be excepted. Or it could confirm if the outlets are in the lower catchment as suspected. Another possibility is that some of the runoff could be redirected into another small stream or the Spree directly. With more time in the catchment, a more thorough inspection of the stream would give an indication of where additional input is occurring.

6.5.2. Sampling methods

As mentioned in the discussion, taking multiple samples from one site would have made the measurements more robust and detected trends more meaningful. Further, possible outliers would be averaged out. Both field trips were very long and ended at dusk. Therefore, more samples per site would have resulted in less samples taken throughout the Wuhle. For this, a third sampling day would have been an advantage, which was not possible in the short time period available. All the lab processes were a new experience, which can introduce small errors to the process. One point of discussion was to let the dry bags slowly cool down in another oven as completely dried out and warm soil does absorb some water from the surrounding air. But given the restricted time available and that it was not mentioned in other studies (e.g. Kuhlemann et al., 2021) this step was skipped. To test the consequences of the samples of the first field work trip, which were not filled up with dry air on the same day, site z was sampled twice in proximity. Three bags of each depth were filled with dry synthetic air in the evening of the same day they were sampled, and the other tree bags were filled the following morning. Unfortunately, the second test sample in site z (WR16x) was measured without measuring the two lab standards for calibration before and after, which made this test unusable. The sample is not included in the analysis of the isotopic composition.

6.5.3. Modelling methods

The parameter range definition introduces an uncertainty to model calibration. Especially the high FC values were difficult to handle as they are unusual, even though from the model logic perspective, in a valid range. For the input of E_{pot} data PET data was used. As the model is not as sensitive to this switch and other correction factors can mitigate this change, it is not considered to be very influential. PET values were provided by the IGB on a subcatchment level. The aggregation of those values introduced a level of uncertainty, as they might have been more accurate if calculated directly for the right areas. Another parameter that was unusual high is K2 which could have been confined into a narrower range. Because the model is used in an unusual setting from its initial purpose, the range was considered acceptable. A better approach could have been a sensitivity analysis of the parameters, even though the dotty plots did not suggest any of the parameters to be more well-defined than the others. During the modelling process with the HBV-light model, the basic standard model version with UZL and K0 parameter was used yet replaced by a version without the additional runoff component. This was due to recent experiences in the research group that showed only slightly better results with the UZL and K0 components, but with higher computational effort. With more time, also other model types could have been tested out, as for example the distributed SUZ model.

6.5.4. Statistical methods

It would have been interesting to see if the evaporation lines of each site differ from each other. But a regression analysis is not feasible for this thesis as the sample size for the sites ranges from 3 to 12. Further, the grouped depth samples do not go over 14 samples. With more samples per site and for each depth, a linear regression would have been reasonable to investigate.

The Kruskal-Wallis-Test was used on independently sampled, normally distributed data with more than two levels within the tested variables. The normally distributed data could be a coincidence and it is possible that the data is not normally distributed in a bigger sampling pool. This is why the Kruskal-Wallis-Test was used, even though it is designed for not normally distributed data. Although it is a less robust test to use, the decision is supported by other papers using the Kruskal-Wallis-Test in the same field of research (Ring et al., 2023; Sprenger et al., 2017). In some studies, the Wilcoxon-test is used as a post-hoc test to detect pairwise differences between variables. In this thesis, the Wilcoxon-test is replaced with the Dunn-Bonferroni-Test, as it is listed as a post-hoc analysis option on the statistical

methods consulting website of the university of Zurich (*Kruskal-Wallis-Test*, 2023). Furthermore, the paper by Sprenger et al. (2017), also used the Dunn-Test as a post-hoc analysis tool.

The NPE as the objective function was chosen, as the KGE assumes data linearity and normality and the absence of outliers as described by Pool et al. (2018). For a non-parametric alternative of the standard deviation used in KGE, the flow duration curve is used, as it is an indicator of flow variability across all flow magnitudes and flashiness of the hydrograph or baseflow can be linked to specific segments of the flow-duration curve (Pool et al., 2018). The Spearman rank correlation coefficient is a more robust characterization of correlation than the Pearson correlation. In another study by Pool et al. (2021) KGE is described as focusing on the magnitude and timing of high flows and NPE achieving a more balanced evaluation of broad range of hydrograph aspects. The flashiness of the Wuhle catchment discharge is difficult to calibrate and other objective functions, focusing more on the peak such as the KGE or even R_{eff}, could achieve better alignment in that regard. However, because the model structure is limited to simulate such discharge, a good calibration of the baseflow is already an achievement. The VE would be probably off in W4.0 either way, because of the direct runoff from the sealed areas.

To calculate the objective function of the simulated ensemble, mean results (see Table 9) had to be calculated outside of the HBV-light software. The observed discharge data of the measuring station in HG is adapted, as in the beginning of the simulation period discharge data is missing. Therefore, the length of the data was shortened. However, some measured discharge data points are missing ($n < 10$) during the investigated time period. This was a problem and had to be worked around for the Spearman rank correlation coefficient to be properly calculated in R, which was done with a case-wise deletion of NA values.

7. Conclusion

This study investigated isotopic composition through the Wuhle catchment in shallow soil (< 15 cm) and modelled three different (sub)catchments. The question posed at the beginning of this thesis was how hydrological signals change throughout the Wuhle catchment and what could influence it. The results of the two research approaches are not comparable or combinable. But both contribute to a conclusion for the Wuhle catchment.

Although a small number of samples was taken, some trends could be determined. The isotopic water composition was significantly different in all bigger water bodies than in the stream, due to evaporative fractioning in non-equilibrium conditions. The stream, although reacting quickly to precipitation events, showed also slow runoff as some main stream measurement showed pond similar compositions. The VMC showed a significantly wetter first 5 cm of the soil compared to the 5 to 15 cm of the shallow soil sampled, which can be an indication of slow percolation. The hypothesis of the upper catchment showing different signals from the lower catchment is strengthened, as the lc-excess is significantly different in the upper and middle catchment sites than the lower catchment. The hypothesis cannot be confirmed, mainly because not enough data was sampled from the upper catchment. The environmental characteristics present in upper catchment sites with significant differences confirmed, can be found in the δD and $\delta^{18}O$ values. Detectable influences are most often soil composition, followed by tree composition and rarely high percentage of impervious surfaces. The second hypothesis of significant changes with depth could not be proven for the whole catchment, as no significant differences with depth could be detected. However, indications of differences with depth are found in the upper and middle catchment, which are preliminary results mainly due to lack of samples. The sampled soil data revealed differences in depth values between the first 5 cm and the lower samples in the upper catchment.

The modelling efforts were overall unsuccessful. In comparison with the lower benchmark in some of the simulations, a relatively good objective function value could be achieved. Because of the flashy nature of heavily reshaped and constructed streams and potential direct runoff from impervious areas, the HBV model did not have the structure to model the discharge adequately. This is evident in the resulting imbalance of the simulated discharge volume. The more agricultural shaped HG subcatchment in one of the calibrated simulations had the best results but on the other hand the second simulation was very low performing. Although, the accurate representation of the (sub)catchments was not as successful, some parameter differences between the catchments are visible despite the unfavourable model structure. Quick reaction to precipitation events was modelled for all catchments and is backed by findings in literature (Kuhlemann et al., 2022). The UHI effect might be an explanation for uncertain snow routine in the whole catchment, whereas the smaller rural subcatchment HG was more defined in the temperature dependent routines. The upper catchment showed larger percolation parameters in context of the incoming water compared to the lower catchments. The hypothesis of the model performing better in the upper catchment than in the lower catchment could be confirmed, but not if the influence is linked to the percentage of the impervious area.

Further research with more environmental influences on a site-specific base is needed. Site c is noticeably closer to the LWML than the other sites and site d is more enriched. More samplings and observations are necessary to determine if there are outliers or if possibly run-on from sealed surfaces influenced the sites. Regarding the HBV model, further research is necessary to see if the model outcome could be improved within the existing structure. Another possibility for research would be the expansion of the model to see if the performance of the simulated runoff could be improved with an option for direct runoff for sealed surfaces.

8. Bibliography

Albert, J. S., Destouni, G., Duke-Sylvester, S. M., Magurran, A. E., Oberdorff, T., Reis, R. E., Winemiller, K. O., & Ripple, W. J. (2021). Scientists' warning to humanity on the freshwater biodiversity crisis. *Ambio*, *50*(1), 85-94. https://doi.org/10.1007/s13280-020-01318-8

Amt für Statistik Berlin-Brandenburg [AFSBB]. (2021a). *Einwohnerbestand Berlin*. Retrieved 01.10.2023 from https://www.statistik-berlin-brandenburg.de/kommunalstatistik/einwohnerbestand-berlin

Amt für Statistik Berlin-Brandenburg [AFSBB]. (2021b). *Flächennutzung*. Retrieved 01.10.2023 from https://www.statistik-berlin-brandenburg.de/flaechennutzung

Arsenault, R., Brissette, F., & Martel, J.-L. (2018). The hazards of split-sample validation in hydrological model calibration. *Journal of Hydrology*, *566*, 346-362. https://doi.org/10.1016/j.jhydrol.2018.09.027

Bergström, S. (1976). *Development and Application of a Conceptual Runoff Model for Scandinavian Catchments*. Swedish Meteorological and Hydrological Institute.

Bergström, S. (1991). Principles and Confidence in Hydrological Modelling. *Hydrology Research*, *22*(2), 123-136. https://doi.org/10.2166/nh.1991.0009

Bergström, S. (1992). *The HBV Model: Its Structure and Applications* (SMHI Reports Hydrology, Issue. S. Norrköping.

Beven, K. J. (2012). *Rainfall-Runoff Modelling: The Primer, Second Edition.* Wiley-Blackwell. https://doi.org/10.1002/9781119951001

Bowler, D. E., Buyung-Ali, L., Knight, T. M., & Pullin, A. S. (2010). Urban greening to cool towns and cities: A systematic review of the empirical evidence. *Landscape and Urban Planning*, *97*(3), 147-155. https://doi.org/10.1016/j.landurbplan.2010.05.006

Brunner, M. I., Slater, L., Tallaksen, L. M., & Clark, M. (2021). Challenges in modeling and predicting floods and droughts: A review. *WIREs Water*, *8*(3). https://doi.org/10.1002/wat2.1520

Commission, E., Joint Research Centre, Jacobs-Crisioni, C., Perpiña Castillo, C., Kompil, M., Ronchi, S., Aurambout, J., Pinto Nunes Nogueira Diogo, V., Batista e Silva, F., Vandecasteele, I., Baranzelli, C., Auteri, D., Rosina, K., Serpieri, C., Pontarollo, N., Vizcaino, M., Ribeiro Barranco, R., Marín Herrera, M., Lavalle, C., & Kavalov, B. (2017). *European territorial trends – Facts and prospects for cities and regions*. Publications Office. https://doi.org/doi/10.2760/148283

Craig, H. (1961). Isotopic Variations in Meteoric Waters. *Science*, *133*, 2. https://doi.org/10.1126/science.133.3465.1702

Criss Robert E., & E., W. W. (2008). Do Nash values have value? Discussion and alternate proposals. *Hydrological Processes*, *22*(14), 2723-2725. https://doi.org/https://doi.org/10.1002/hyp.7072

Dansgaard, W. (2012). Stable isotopes in precipitation. *Tellus A: Dynamic Meteorology and Oceanography*, *16*(4). https://doi.org/10.3402/tellusa.v16i4.8993

Delta-T. (2023). *ML3 ThetaProbe*. Retrieved 16.02.2023 from https://delta-t.co.uk/wp-content/uploads/2017/02/ML3_user_manual_version_2.0.pdf

Deutscher Wetterdienst [DWD]. (2020). *Klimastatusbericht Deutschland Jahr 2019*. Geschäftsbereich Klima und Umwelt. Retrieved 01.10.2023 from www.dwd.de/DE/derdwd/bibliothek/fachpublikationen/selbstverlag/selbstverlag_node.html, https://www.dwd.de/DE/leistungen/klimastatusbericht/klimastatusbericht.html

Deutscher Wetterdienst [DWD]. (2022a). *Daily station observations of mean temperature at 2 m above ground in °C for Germany*. https://cdc.dwd.de/portal/

Deutscher Wetterdienst [DWD]. (2022b). *Daily station observations precipitation height in mm for Germany*. https://cdc.dwd.de/portal/

Deutscher Wetterdienst [DWD]. (2023). *Klimatabelle Niederschlagshöhe und Lufttemperatur,, Berlin Brandenburg, Berlin-Dahlem, Berlin-Tempelhof*. 17.02.2023. https://www.dwd.de/DE/leistungen/kvo/berlin_brandenburg.html

Dunn, O. J. (1964). Multiple Comparisons Using Rank Sums. *Technometrics*, *6*(3), 241-252. https://doi.org/10.1080/00401706.1964.10490181

Ehleringer, J. R., Barnette, J. E., Jameel, Y., Tipple, B. J., & Bowen, G. J. (2016). Urban water – a new frontier in isotope hydrology†. *Isotopes in Environmental and Health Studies*, *52*(4-5), 477-486. https://doi.org/10.1080/10256016.2016.1171217

Europe, W. R. O. f. (2016). *Urban green spaces and health*.

European Union, Coperincus Land Monitoring Service,. (2018). *Urban Atlas LCLU 2018* [Vector]. https://land.copernicus.eu/local/urban-atlas/urban-atlas-2018

European Union, Copernicus Land Monitoring Service information,. (2020). *Imperviousness Density 2018 (raster 100 m), Europe, 3-yearly*. https://doi.org/10.2909/3bf542bd-eebd-4d73-b53c-a0243f2ed862

Federer, C. A., Vörösmarty, C., & Fekete, B. (1996). Intercomparison of Methods for Calculating Potential Evaporation in Regional and Global Water Balance Models. *Water Resources Research*, *32*(7), 2315-2321. https://doi.org/10.1029/96wr00801

Fletcher, T. D., Andrieu, H., & Hamel, P. (2013). Understanding, management and modelling of urban hydrology and its consequences for receiving waters: A state of the art. *Advances in Water Resources*, *51*, 261-279. https://doi.org/10.1016/j.advwatres.2012.09.001

Gat, J. R. (1996). Oxygen and Hydrogen Isotopes in the Hydrologic Cycle. *Annual Review of Earth and Planetary Sciences*, *24*(1), 38.

Geodaten der Deutschen Landesvermessung (GeoBasis-DE). (2019). *Digitale Geodaten*. Retrieved 01.10.2023 from https://isk.geobasis-bb.de/mapproxy/dgm/service/wms?SERVICE=WMS&REQUEST=GetCapabilities

Gerstenberg, J. (2019). *02.13.1 Oberflächenabfluss aus Niederschlägen*. Senatsverwaltung für Stadtentwicklung und Wohnen, III D Geodateninfrastruktur, Umweltatlas; Bundesanstalt für Gewässerkunde Retrieved 01.10.2023 from https://www.berlin.de/umweltatlas/wasser/wasserhaushalt/2017/karten/artikel.994317.php

Gessner, M. O., Hinkelmann, R., Nützmann, G., Jekel, M., Singer, G., Lewandowski, J., Nehls, T., & Barjenbruch, M. (2014). Urban water interfaces. *Journal of Hydrology*, *514*, 226-232. https://doi.org/10.1016/j.jhydrol.2014.04.021

Golden, H. E., & Hoghooghi, N. (2018). Green infrastructure and its catchment-scale effects: an emerging science. *WIREs Water*, *5*(1), 1254. https://doi.org/10.1002/wat2.1254

Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, *377*(1), 80-91. https://doi.org/https://doi.org/10.1016/j.jhydrol.2009.08.003

Haase, D. (2009). Effects of urbanisation on the water balance – A long-term trajectory. *Environmental Impact Assessment Review*, *29*(4), 211-219. https://doi.org/10.1016/j.eiar.2009.01.002

Harrison, R. L. (2010). Introduction to Monte Carlo Simulation. *AIP Conference Proceedings*, *1204*(1), 17-21. https://doi.org/10.1063/1.3295638

Hughes, C. E., & Crawford, J. (2012). A new precipitation weighted method for determining the meteoric water line for hydrological applications demonstrated using Australian and global GNIP data. *Journal of Hydrology*, *464-465*, 344-351. https://doi.org/10.1016/j.jhydrol.2012.07.029

Intergovernmental Panel on Climate Change [IPCC]. (2021). *Summary for Policymakers*.

Klemeš, V. (1986). Operational testing of hydrological simulation models. *Hydrological Sciences Journal*, *31*(1), 13-24. https://doi.org/10.1080/02626668609491024

Kruskal-Wallis-Test. (2023). Retrieved 01.10.2023 from https://www.methodenberatung.uzh.ch/de/datenanalyse_spss/unterschiede/zentral/kruskal.html#3.3._Post-hoc-Tests

Kruskal, W. H., & Wallis, W. A. (1952). Use of Ranks in One-Criterion Variance Analysis. *Journal of the American Statistical Association*, *47*(260), 583-621. https://doi.org/10.1080/01621459.1952.10483441

Kuhlemann, L.-M., Tetzlaff, D., Marx, C., & Soulsby, C. (2022). The imprint of hydroclimate, urbanization and catchment connectivity on the stable isotope dynamics of a large river in Berlin, Germany. *Journal of Hydrology*, *613*. https://doi.org/10.1016/j.jhydrol.2022.128335

Kuhlemann, L.-M., Tetzlaff, D., Smith, A., Kleinschmit, B., & Soulsby, C. (2021). Using soil water isotopes to infer the influence of contrasting urban green space on ecohydrological partitioning.

Hydrology and Earth System Sciences, *25*(2), 927-943. https://doi.org/10.5194/hess-25-927-2021

Kuhlemann, L. M., Tetzlaff, D., & Soulsby, C. (2020). Urban water systems under climate stress: An isotopic perspective from Berlin, Germany. *Hydrological Processes*, *34*(18), 3758-3776. https://doi.org/10.1002/hyp.13850

Landesamt für Umwelt Brandenburg. (2021). *Gewässernetz im Land Brandenburg*. https://geoportal.brandenburg.de/

Landesamt für Umwelt Brandenburg. (2023). *Oberirdische Einzugsgebiete im Land Brandenburg*. https://geoportal.brandenburg.de/

Landwehr, J., & Coplen, T. B. (2006). Line-conditioned excess: a new method for characterizing stable hydrogen and oxygen isotope ratios in hydrologic systems. In (pp. 132-135).

Limberg, A., Darkow, P., Faensen-Thiebes, A., Fritz-Taute, B., Günther, M., Hähnel, K., Hörmann, U., Jahn, D., Köhler, A., Krüger, E., May, S., Naumann, J., & Wagner, M. (2007). Grundwasser in Berlin, Vorkommen·Nutzung·Schutz·Gefährdung. In U. u. V. Senatsverwaltung für Gesundheit (Ed.): Berlin.

Limberg, A., & Thierbach, J. (1997). Gliederung der Grundwasserleiter in Berlin. *Brandenburgische Geowissenschaftliche Beiträge*, *4*(2), 21-26.

López Moreira Mazacotte, G. (2024). *Integrating multiple data sources and modelling to disentangle the complex spatio-temporal hydrological dynamics of urbanised streams under drought stress.* In preparation.

Marx, C., Tetzlaff, D., Hinkelmann, R., & Soulsby, C. (2021). Isotope hydrology and water sources in a heavily urbanized stream. *Hydrological Processes*, *35*(10). https://doi.org/10.1002/hyp.14377

Marx, C., Tetzlaff, D., Hinkelmann, R., & Soulsby, C. (2022). Seasonal variations in soil–plant interactions in contrasting urban green spaces: Insights from water stable isotopes. *Journal of Hydrology*, *612*. https://doi.org/10.1016/j.jhydrol.2022.127998

Menberg, K., Bayer, P., Zosseder, K., Rumohr, S., & Blum, P. (2013). Subsurface urban heat islands in German cities. *Sci Total Environ*, *442*, 123-133. https://doi.org/10.1016/j.scitotenv.2012.10.043

Mendez, M., & Calvo-Valverde, L. (2016). Development of the HBV-TEC Hydrological Model. *Procedia Engineering*, *154*, 1116-1123. https://doi.org/10.1016/j.proeng.2016.07.521

Motavita, D. F., Chow, R., Guthke, A., & Nowak, W. (2019). The comprehensive differential split-sample test: A stress-test for hydrological model robustness under climate variability. *Journal of Hydrology*, *573*, 501-515. https://doi.org/10.1016/j.jhydrol.2019.03.054

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I — A discussion of principles. *Journal of Hydrology*, *10*(3), 282-290. https://doi.org/https://doi.org/10.1016/0022-1694(70)90255-6

Neuwirth, E. (2014). *RColorBrewer: ColorBrewer Palettes. R package version 1.1-2.*

Nonki, R. M., Lenouo, A., Tshimanga, R. M., Donfack, F. C., & Tchawoua, C. (2021). Performance assessment and uncertainty prediction of a daily time-step HBV-Light rainfall-runoff model for the Upper Benue River Basin, Northern Cameroon. *Journal of Hydrology: Regional Studies*, *36*. https://doi.org/10.1016/j.ejrh.2021.100849

Oswald, C. J., Kelleher, C., Ledford, S. H., Hopkins, K. G., Sytsma, A., Tetzlaff, D., Toran, L., & Voter, C. (2023). Integrating urban water fluxes and moving beyond impervious surface cover: A review. *Journal of Hydrology*, *618*. https://doi.org/10.1016/j.jhydrol.2023.129188

Penman, H. L. (1948). Natural evaporation from open water, hare soil and grass. *Proc R Soc Lond A Math Phys Sci*, *193*(1032), 120-145. https://doi.org/10.1098/rspa.1948.0037

Pflanzenschutzamt Berlin. (2020). *Messungen der Bodenfeuchte*. Retrieved 01.10.2023 from https://www.berlin.de/pflanzenschutzamt/stadtgruen/beratung/messungen-der-bodenfeuchte/

Pool, S., Vis, M., & Seibert, J. (2018). Evaluating model performance: towards a non-parametric variant of the Kling-Gupta efficiency. *Hydrological Sciences Journal*, *63*(13-14), 1941-1953. https://doi.org/10.1080/02626667.2018.1552002

Pool, S., Vis, M., & Seibert, J. (2021). Regionalization for Ungauged Catchments — Lessons Learned From a Comparative Large-Sample Study. *Water Resources Research*, *57*(10). https://doi.org/10.1029/2021wr030437

Press William H., T. S. A., Vetterling William T, Flannery Brian P. . (2002). *Numerical Recipes: The Art of Scientific Computing* (S. Sahraei, Ed.). Cambridge University Press.

Pukowiec-Kurda, K. (2022). The urban ecosystem services index as a new indicator for sustainable urban planning and human well-being in cities. *Ecological Indicators*, *144*. https://doi.org/10.1016/j.ecolind.2022.109532

R Core Team. (2023). *R: A language and environment for statistical computing*. In (Version 4.3.1) R Foundation for Statistical Computing. http://www.R-project.org/

Ring, A.-M., Tetzlaff, D., Dubbert, M., Dubbert, D., & Soulsby, C. (2023). High-resolution in situ stable isotope measurements reveal contrasting atmospheric vapour dynamics above different urban vegetation. *Hydrological Processes*, *37*(9), e14989. https://doi.org/https://doi.org/10.1002/hyp.14989

Rosbjerg, D., & Madsen, H. (2005). Concepts of Hydrologic Modeling. In *Encyclopedia of Hydrological Sciences*. https://doi.org/https://doi.org/10.1002/0470848944.hsa009

Schwarzenbach, F. (2022). *Value of Citizen Science Data and Limited Other Information for Hydrological Model Calibration*

Seibert, J. (1997). Estimation of Parameter Uncertainty in the HBV Model. *Nordic Hydrology*, *28*(4/5), 16.

Seibert, J. (1999). Regionalisation of parameters for a conceptual rainfall-runoff model. *Agricultural and Forest Meteorology*, *98-99*, 279-293. https://doi.org/https://doi.org/10.1016/S0168-1923(99)00105-7

Seibert, J. (2000). Multi-criteria calibration of a conceptual runoff model using a genetic algorithm. *Hydrology and Earch System Sciences*, *4*(2), 215-224.

Seibert, J. (2001). On the need for benchmarks in hydrological modelling. *Hydrological Processes*, *15*(6), 1063-1064. https://doi.org/10.1002/hyp.446

Seibert, J., & Bergström, S. (2022). A retrospective on hydrological catchment modelling based on half a century with the HBV model. *Hydrology and Earth System Sciences*, *26*(5), 1371-1388. https://doi.org/10.5194/hess-26-1371-2022

Seibert, J., & van Meerveld, H. J. I. (2016). Hydrological change modeling: Challenges and opportunities. *Hydrological Processes*, *30*(26), 4966-4971. https://doi.org/10.1002/hyp.10999

Seibert, J., & Vis, M. J. P. (2012). Teaching hydrological modeling with a user-friendly catchment-runoff-model software package. *Hydrol. Earth Syst. Sci.*, *16*(9), 3315-3325. https://doi.org/10.5194/hess-16-3315-2012

Seibert, J., Vis, M. J. P., Lewis, E., & van Meerveld, H. J. (2018). Upper and lower benchmarks in hydrological modelling. *Hydrological Processes*, *32*(8), 1120-1125. https://doi.org/10.1002/hyp.11476

Senatsverwaltung für Stadtentwicklung [SenSt]. (2004). *Dokumentation der Umsetzung der EG-Wasserrahmenrichtlinie in Berlin (Länderbericht)*. https://www.berlin.de/sen/uvk/_assets/umwelt/wasser-und-geologie/europaeische-wasserrahmenrichtlinie/wrrl-laenderbericht.pdf

Senatsverwaltung für Stadtentwicklung und Wohnen [SenStadtWoh]. (2017). *01.12 Versiegelung*. Berlin Umweltatlas. https://www.berlin.de/umweltatlas/_assets/boden/versiegelung/de-texte/kd102.pdf

Senatsverwaltung für Stadtentwicklung und Wohnen [SenStadtWoh]. (2018). *02.09 Management of Rain and Waste*

Water (Edition 2018). Berlin Enviromental Atlas. https://www.berlin.de/umweltatlas/_assets/wasser/regen-und-abwasser/en-texte/ekd209.pdf

Senatsverwaltung für Umwelt, Mobilität, Verbraucher- und Klimaschutz [SenUMVK],. (2018). *Öffentliche Grünflächen in Berlin, Diagramm: Prozentualer Anteil Flächen*. Retrieved 01.10.2023 from

https://www.berlin.de/sen/uvk/_assets/natur-gruen/stadtgruen/daten-und-fakten/ausw_5.pdf
Senatsverwaltung für Umwelt, Mobilität, Verbraucher- und Klimaschutz [SenUMVK],. (2021a). *Ergänzender Länderbericht Berlins zur Aktualisierung des Bewirtschaftungsplans und des Massnahmenprogramms der Flussgebietsgemeinschaft Elbe für den Zeitraum 2022 bis 2027*.
Senatsverwaltung für Umwelt, Mobilität, Verbraucher- und Klimaschutz [SenUMVK],. (2021b). *Wasserportal Gewässerkundliche Messdaten*. Retrieved 17.02.2023 from https://wasserportal.berlin.de/stationen_start.php
Senatsverwaltung für Umwelt, Mobilität, Verbraucher- und Klimaschutz [SenUMVK],. (2022). *Masterplan Wasser Berlin - 1. Bericht*. https://www.berlin.de/sen/uvk/_assets/umwelt/wasser-und-geologie/masterplan-wasser/masterplan-wasser-berlin.pdf
Sentsverwaltung für Stadtentwicklung und Umwelt [SenStU]. (2013a). Ökologische Entwicklung der Wuhle, Informationsheft zur europäischen Wasserrahmenrichtlinie (WRRL). In (pp. 41). Berlin.
Sentsverwaltung für Stadtentwicklung und Umwelt [SenStU]. (2013b). *Technisch-integrative Zusammenfassung der vorbereitenden Maßnahmenplanung im Einzugsgebiet der Wuhle*. https://www.berlin.de/sen/uvk/_assets/umwelt/wasser-und-geologie/europaeische-wasserrahmenrichtlinie/wuhle-technisch-integrative_zusamenfassung.pdf?ts=1699574415
Shapiro, S. S., & Wilk, M. B. (1965). An Analysis of Variance Test for Normality (Complete Samples). *Biometrika*, *52*(3/4), 591-611. https://doi.org/10.2307/2333709
Shuster, W. D., Bonta, J., Thurston, H., Warnemuende, E., & Smith, D. R. (2005). Impacts of impervious surface on watershed hydrology: A review. *Urban Water Journal*, *2*(4), 263-275. https://doi.org/10.1080/15730620500386529
Solomatine, D. P. (2011). 2.16 – Hydrological Modeling.
Souch, C., & Grimmond, S. (2006). Applied climatology: urban climate. *Progress in Physical Geography*, *30*, 270 - 279.
Spearman, C. (1904). The Proof and Measurement of Association between Two Things. *The American Journal of Psychology*, *15*(1), 72-101. https://doi.org/10.2307/1412159
Sprenger, M., Tetzlaff, D., & Soulsby, C. (2017). Soil water stable isotopes reveal evaporation dynamics at the soil–plant–atmosphere interface of the critical zone. *Hydrology and Earth System Sciences*, *21*(7), 3839-3858. https://doi.org/10.5194/hess-21-3839-2017
Stackebrandt, W., & Manhenke, V. (2010). Geologie und Geopotenziale in Brandenburg. In *Atlas zur Geologie von Brandenburg* (Vol. 4. aktualisierte Auflage). Landesamt für Bergbau, Geologie und Rostoffe Brandenburg [LBGR].
Student. (1908). The Probable Error of a Mean. *Biometrika*, *6*(1), 1-25. https://doi.org/10.2307/2331554
United Nations. (2019). *World Urbanization Prospects - The 2018 Revision*.
USACE Hydrologic Engineering Center. (n.d.). *Penman Monteith Method - Basic Concepts and Equations*. Retrieved 01.10.2023 from https://www.hec.usace.army.mil/confluence/hmsdocs/hmstrm/evaporation-and-transpiration/penman-monteith-method
van Meerveld, H. J. I., Vis, M. J. P., & Seibert, J. (2017). Information content of stream level class data for hydrological model calibration. *Hydrology and Earth System Sciences*, *21*(9), 4895-4905. https://doi.org/10.5194/hess-21-4895-2017
Vulova, S., Meier, F., Fenner, D., Nouri, H., & Kleinschmit, B. (2020). Summer Nights in Berlin, Germany: Modeling Air Temperature Spatially With Remote Sensing, Crowdsourced Weather Data, and Machine Learning. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *13*, 5074-5087. https://doi.org/10.1109/JSTARS.2020.3019696
Wassenaar, L. I., Hendry, M. J., Chostner, V. L., & Lis, G. P. (2008). High Resolution Pore Water δ2H and δ18O Measurements by H2O(liquid)-H2O(vapor) Equilibration Laser Spectroscopy. *Enviromental Science Technology*, *42*(24), 6. https://doi.org/10.1021/es802065s
Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., & François, R., et al. (2019). Welcome to the Tidyverse. . *Journal of Open Source Software*, *4*(43). https://doi.org/https://doi.org/10.21105/joss.01686

Zambrano-Bigiarini, M. (2020). *hydroGOF: Goodness-of-fit functions for comparison of simulated and observed hydrological time series. R package version 0.4-0.* Retrieved 01.10.2023 from https://doi.org/10.5281/zenodo.839854

9. Appendix

9.1. Overview of the Wuhle stream

Table 11: Overview of the Wuhle stream, its tributaries and connected water bodies. Table inspired by Kuhlemann et al. (2018). Total catchment size and catchment size within Berlin including percentage of sealed surfaces from (SenStadtWoh, 2018; SenUMVK, 2021a; SenStU, 2013b). If no data was available length was substituted by GIS data from (Landesamt für Umwelt Brandenburg, 2021).

Name	**Catchment area (catchment in Berlin) [km²]**	**Length of the stream (km)**	**Type of water body**	**Type of ecology**
Wuhle	101 (57)	16.6	stream	organically shaped stream/river
Biesdorf-Marzahner Grenzgraben		1.25	tributary	organically shaped stream
Hellersdorfer Graben	31.2 (6.0)	14.4	tributary	Artificial
Alte Wuhle	55.7 (5.2)	-	Stream	organically shaped stream
Neue Wuhle	-	8	tributary	Artificial
Wuhlegraben		2.4		organically shaped stream
Eichwaldgraben		3	unknown	unknown
Wuhlesee			Tributary, still water body	Artificial
Wuhleblase			Still water body	
Wuhleteich			Still water body	
Biersdorfer Baggersee			still water body, disconnected	Artificial
Kaulsdorfer Baggerseen			Still water body, disconnected	Artificial

9.2. Comparison subcatchment W4.4 in different combinations

Table 12: Table with different definitions of subcatchment W4.4. The assumption in the last column is that in subcatchment 5 all the characteristics are evenly distributed which is unlikely.

	W4.4		
Administrative information			
Number	5863000		
Station name	Eisenacher Strasse		
Water body	(Alte) Wuhle		
Available information	w		
Division Subcatchments			
Subcatchment Number	10	5, 10	Half of 5, 10
Area [km^2]	16.75	21.42	19.08
Area of Wuhle [%]	15.4	19.6	17.5
Area characteristics [%]			
Ø Impervious	16.4	21.4	19.18
Ø Sand	76.2	77.7	77.1
Ø Clay	6.6	6	6.2
Total area with trees	21.6	22.8	22.3
Area with broadleaf trees	15.6	17.9	16.9
Area with Coniferous trees	6	4.8	3.3

9.3. Wuhle catchment soil composition

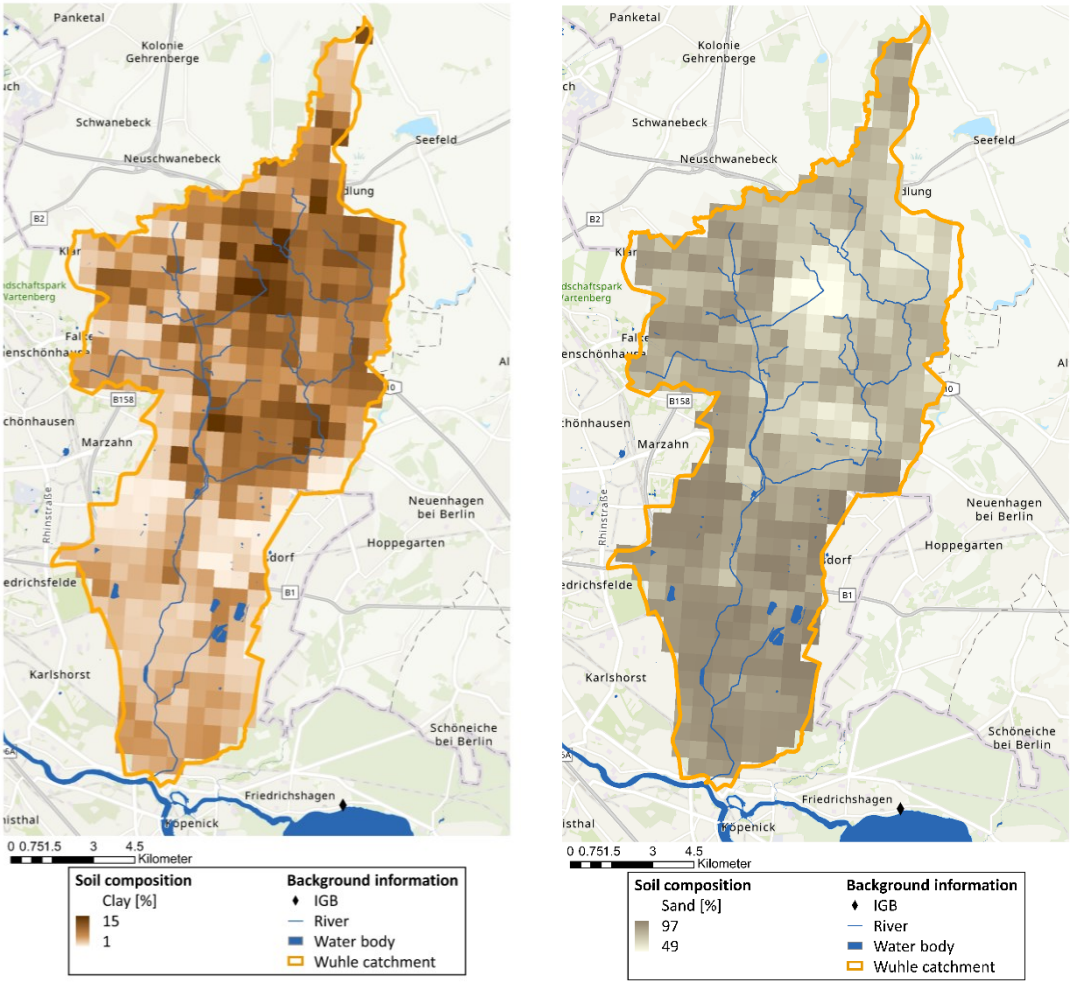


Figure 18: Percentage of clay in soil composition (left) and percentage of sand in soil composition (right) in the Wuhle catchment.

9.4. Isotopic compositions of the open water measurements in the Wuhle catchment

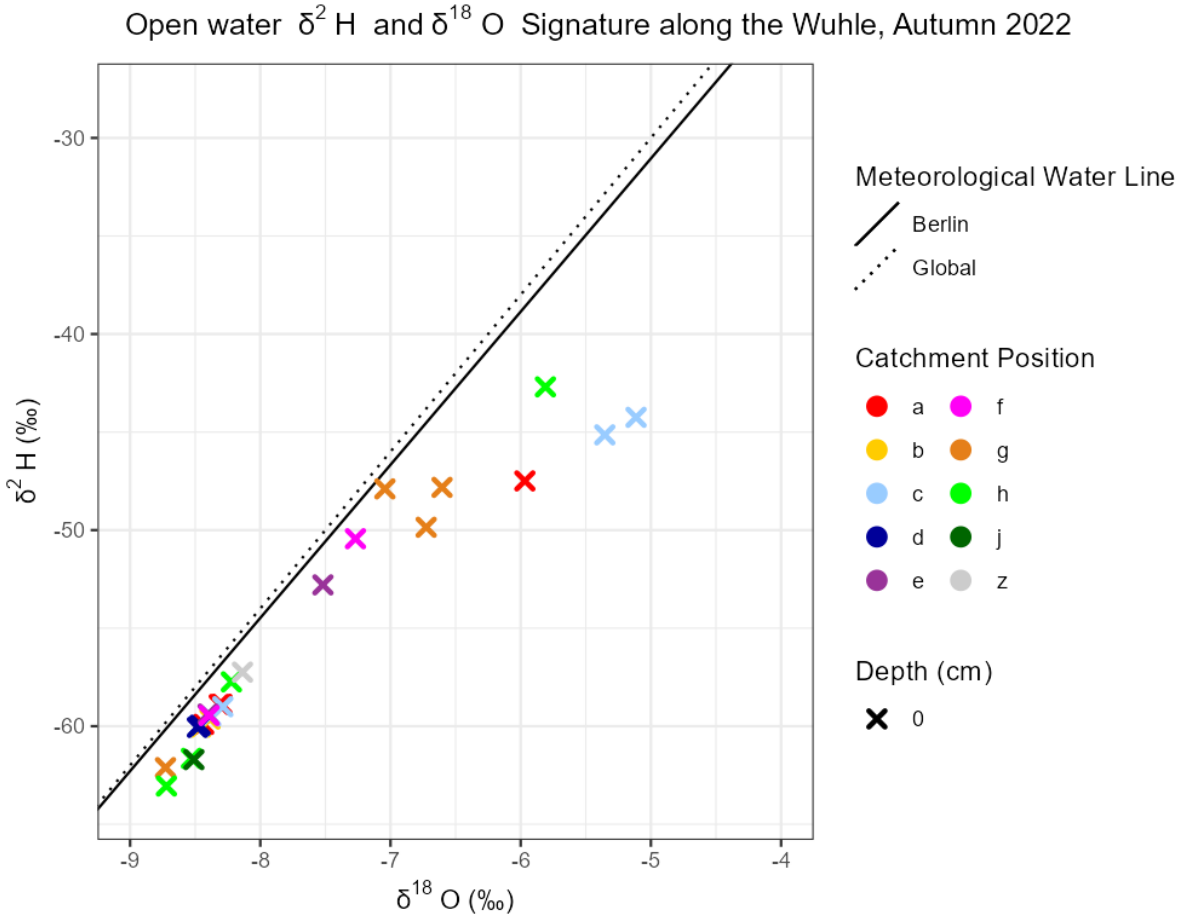


Figure 19: Close-up of the dual-isotope plot of surface water isotope measurements with catchment positions.

9.5. Dunn-Test for discharge data

*Table 13: Discharge data results of the Dunn-Test for significant differences between the parameters for each catchment in simulation A (white) and simulation B (grey). Adjusted p-value are given as: ns = not significant or > 0.05, * < 0.05, ** < 0.01, *** < 0.001 and **** < 0.0001*

	TT_1_1		
Catchment	HG	W40	W43
HG	-	****	****
W40	ns	-	ns
W43	**	****	-

	CFMAX_1_1		
Catchment	HG	W40	W43
HG	-	****	****
W40	****	-	ns
W43	****	ns	-

	SFCF_1_1		
Catchment	HG	W40	W43
HG	-	****	****
W40	****	-	ns
W43	ns	****	-

	FC_1_1		
Catchment	HG	W40	W43
HG	-	****	****
W40	****	-	****
W43	****	****	-

	LP_1_1		
Catchment	HG	W40	W43
HG	-	****	**
W40	****	-	****
W43	**	****	-

	BETA_1_1		
Catchment	HG	W40	W43
HG	-	ns	****
W40	ns	-	****
W43	ns	ns	-

	PERC_1		
Catchment	HG	W40	W43
HG	-	****	****
W40	****	-	***
W43	****	***	-

	Alpha_1		
Catchment	HG	W40	W43
HG	-	****	****
W40	****	-	*
W43	****	ns	-

	K1_1		
Catchment	HG	W40	W43
HG	-	****	****
W40	****	-	**
W43	ns	****	-

	K2_1		
Catchment	HG	W40	W43
HG	-	****	****
W40	****	-	****
W43	****	****	-

	MAXBAS		
Catchment	HG	W40	W43
HG	-	****	****
W40	****	-	****
W43	***	****	-

	Cet_1		
Catchment	HG	W40	W43
HG	-	****	****
W40	****	-	***
W43	****	***	-

9.6. Dunn-Test for water level data

*Table 14: Water level data calibration. Results of the Dunn-Test for significant differences between the parameters for each catchment in simulation A (white) and simulation B (grey). Adjusted p-value are given as: ns = not significant or > 0.05, * < 0.05, ** < 0.01, *** < 0.001 and **** < 0.0001*

TT_1_1				
Catchment	HG	W40	W43	W44
HG	-	****	****	****
W40	****	-	**	ns
W43	ns	****	-	ns
W44	****	****	****	-

CFMAX_1_1				
Catchment	HG	W40	W43	W44
HG	-	ns	***	**
W40	****	-	****	ns
W43	****	ns	-	****
W44	ns	****	****	-

SFCF_1_1				
Catchment	HG	W40	W43	W44
HG	-	****	****	****
W40	****	-	****	***
W43	****	ns	-	****
W44	****	*	*	-

FC_1_1				
Catchment	HG	W40	W43	W44
HG	-	*	****	***
W40	****	-	****	ns
W43	ns	****	-	****
W44	****	****	**	-

LP_1_1				
Catchment	HG	W40	W43	W44
HG	-	****	****	****
W40	****	-	****	ns
W43	***	****	-	****
W44	****	****	****	-

BETA_1_1				
Catchment	HG	W40	W43	W44
HG	-	****	****	***
W40	ns	-	*	****
W43	****	****	-	****
W44	ns	**	****	-

PERC_1_1				
Catchment	HG	W40	W43	W44
HG	-	****	****	ns
W40	ns	-	****	****
W43	****	****	-	****
W44	****	****	ns	-

Alpha_1_1				
Catchment	HG	W40	W43	W44
HG	-	ns	****	ns
W40	*	-	ns	ns
W43	****	ns	-	ns
W44	**	ns	ns	-

K1_1				
Catchment	HG	W40	W43	W44
HG	-	****	****	ns
W40	****	-	ns	***
W43	****	*	-	ns
W44	****	ns	**	-

K2_1				
Catchment	HG	W40	W43	W44
HG	-	****	****	****
W40	****	-	***	ns
W43	****	ns	-	**
W44	****	****	****	-

MAXBAS_1				
Catchment	HG	W40	W43	W44
HG	-	ns	****	****
W40	****	-	****	****
W43	ns	****	-	**
W44	****	ns	****	-

Cet_1				
Catchment	HG	W40	W43	W44
HG	-	****	****	****
W40	****	-	****	****
W43	****	ns	-	ns
W44	****	****	****	-

F

9.7. Dotty plots for simulation A with discharge data

Catchment W4.0, simulation A, discharge data

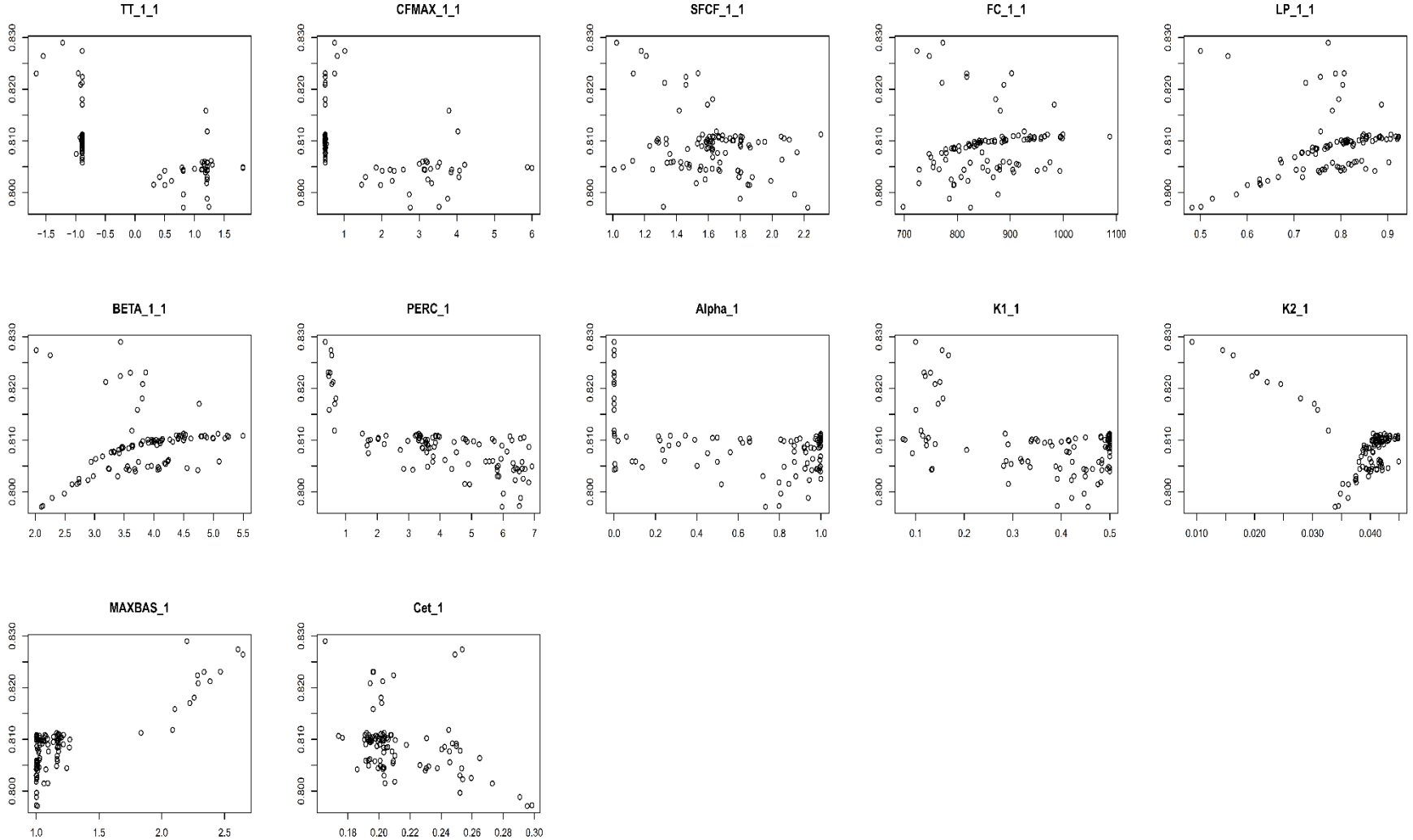


Figure 20: Dotty plot for catchment W4.0 for simulation A for the 12 calibrated parameters with discharge data. The title of the plot shows the parameter of the x-axis, the y-axis is the objective function NPE.

Catchment W4.3, simulation A, discharge data

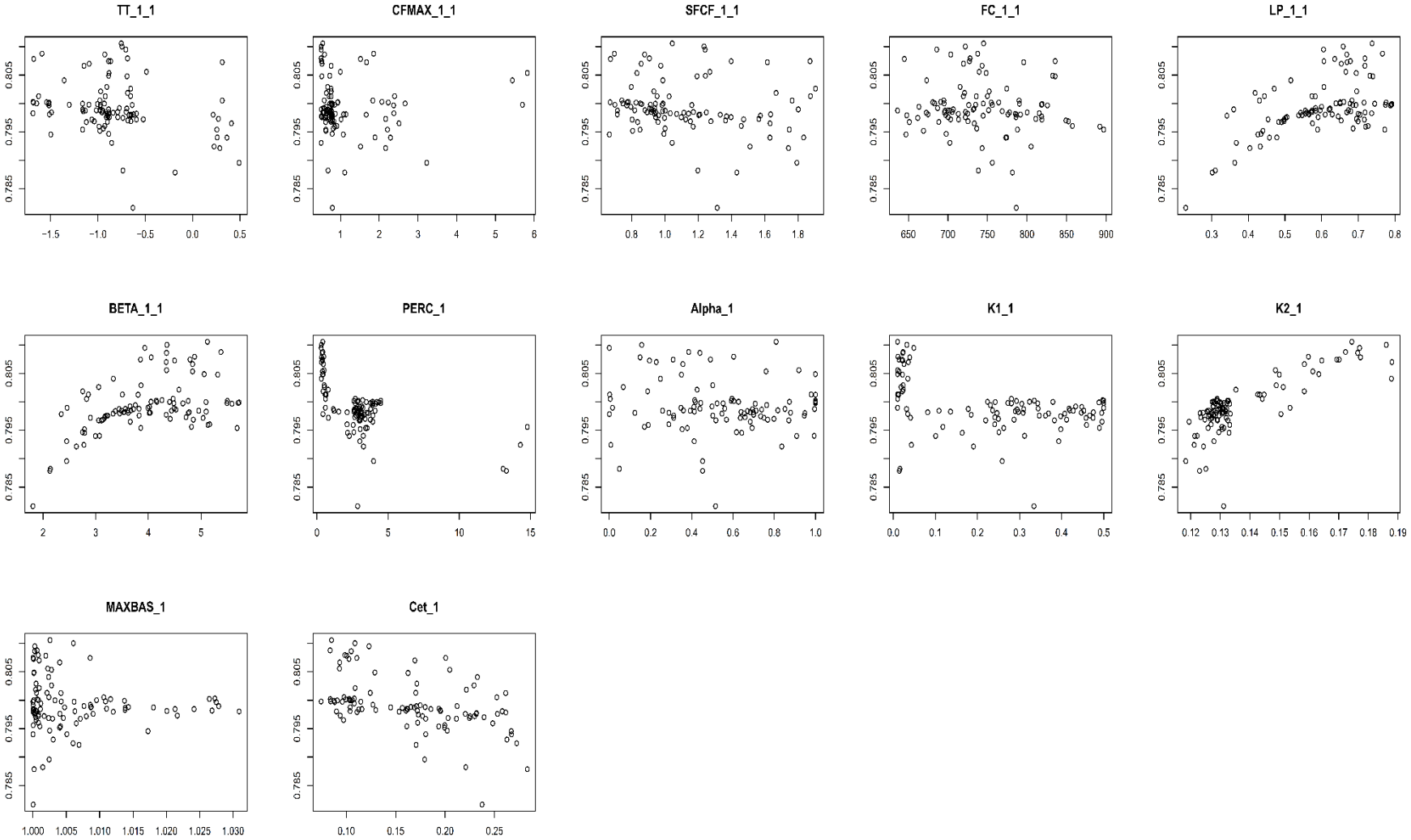


Figure 21: Dotty plot for catchment W4.3 for simulation A for the 12 calibrated parameters with discharge data. The title of the plot shows the parameter of the x-axis, the y-axis is the objective function NPE.

Catchment HG, simulation A, discharge data

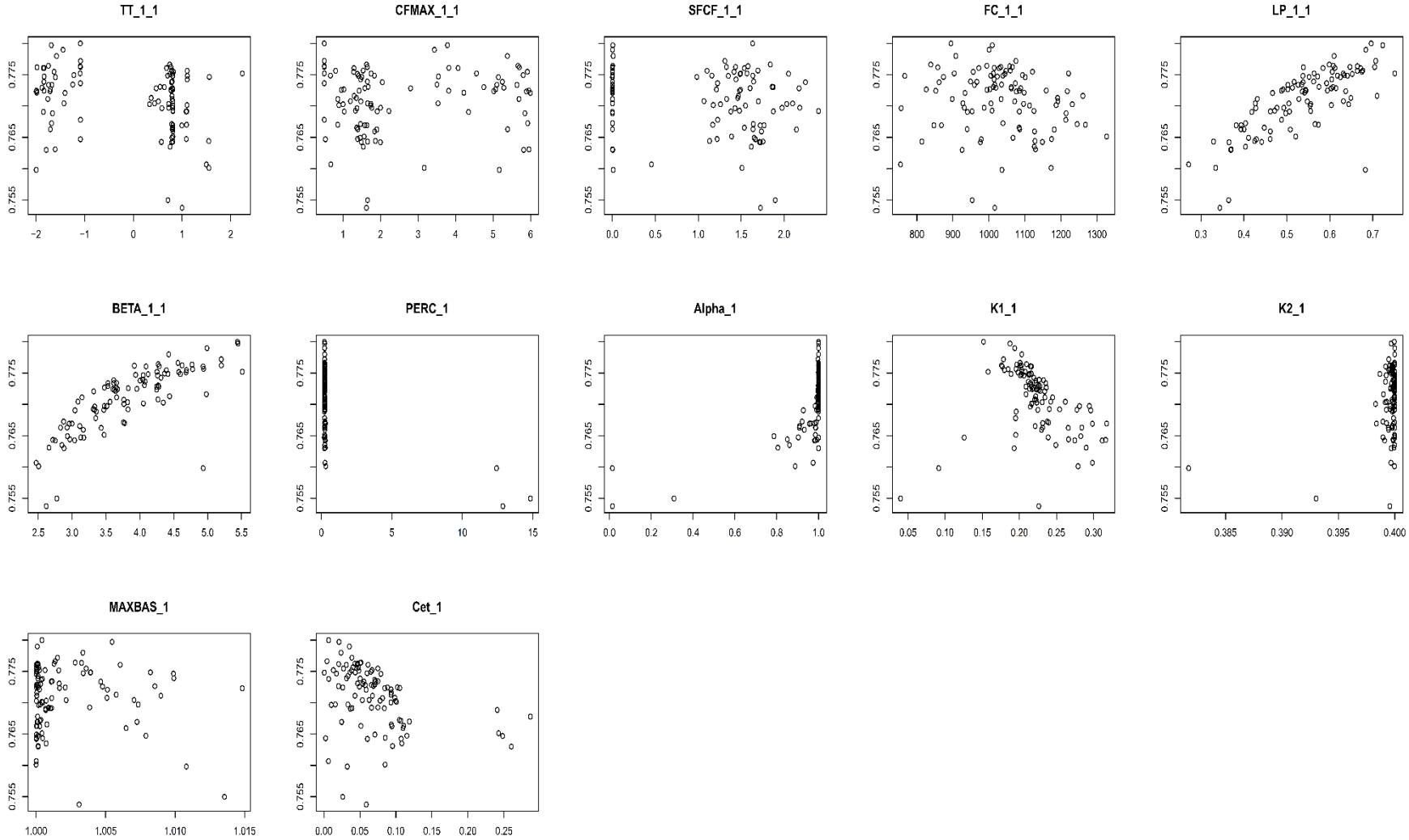


Figure 22: Dotty plot for catchment HG for simulation A for the 12 calibrated parameters with discharge data. The title of the plot shows the parameter of the x-axis, the y-axis is the objective function NPE.

9.8. Dotty plots for simulation B with discharge data

Catchment W4.0, simulation B, discharge data

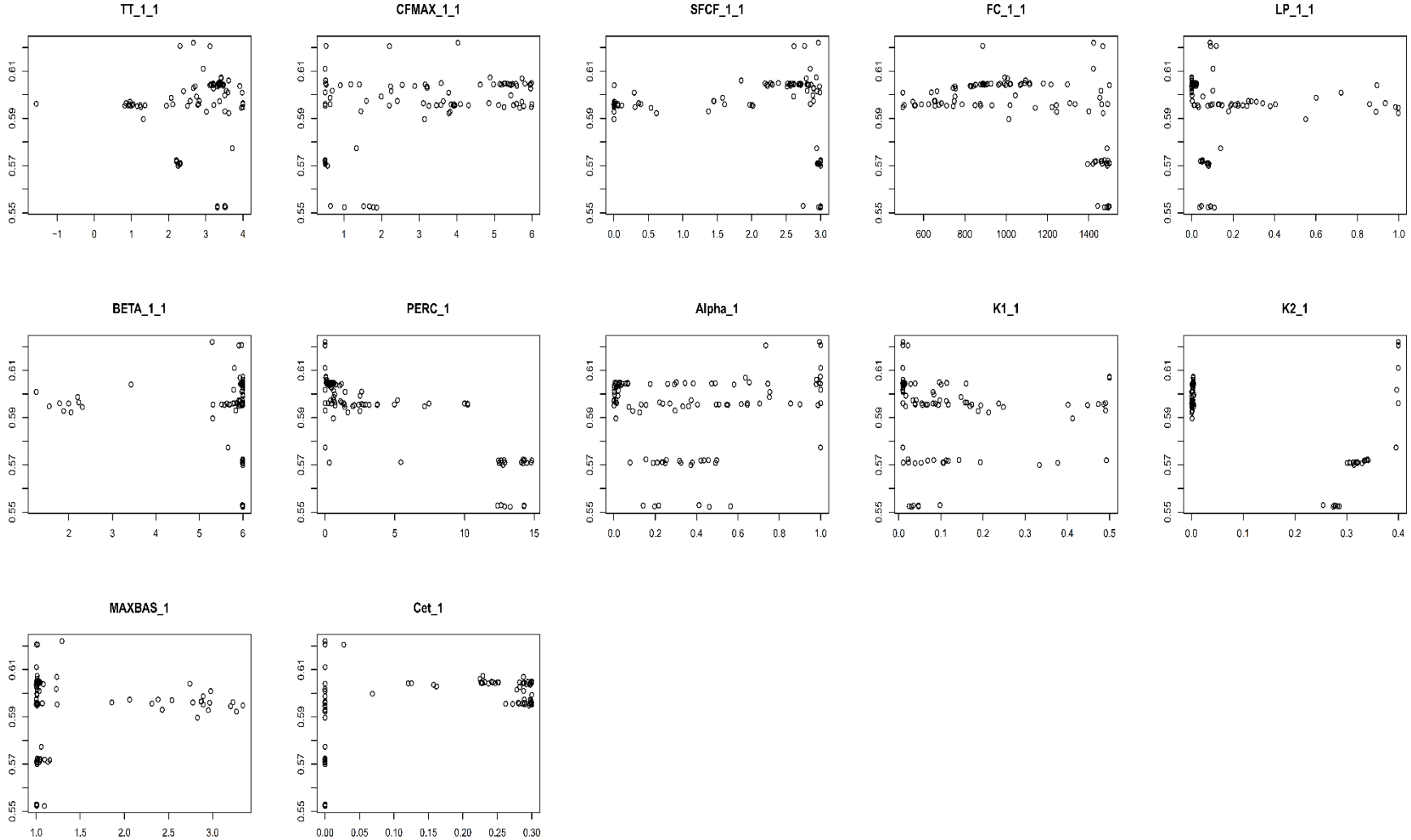


Figure 23: Dotty plot for catchment W4.0 for simulation B for the 12 calibrated parameters with discharge data. The title of the plot shows the parameter of the x-axis, the y-axis is the objective function NPE.

Catchment W4.3, simulation B, discharge data

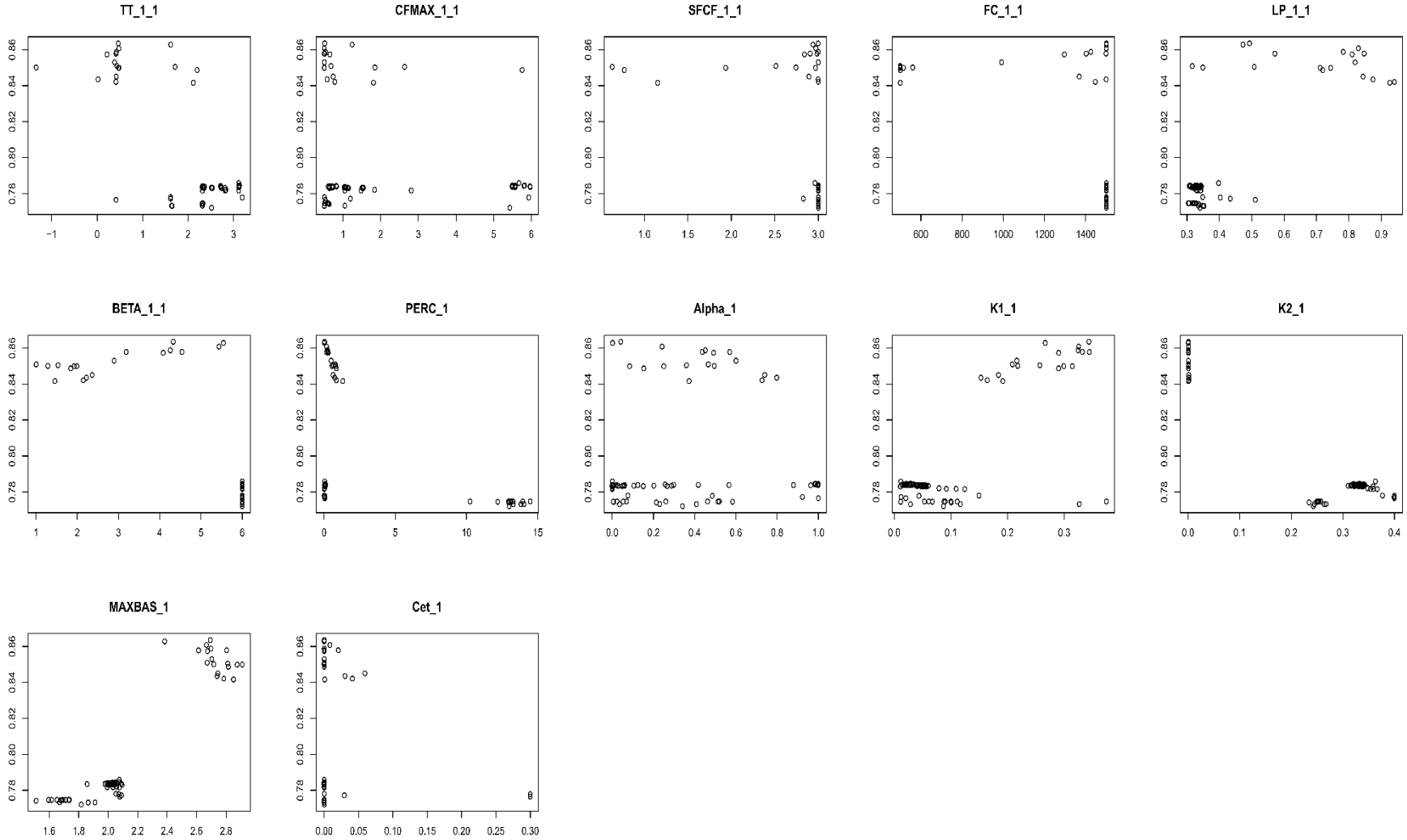


Figure 24: Dotty plot for catchment W4.3 for simulation B for the 12 calibrated parameters with discharge data. The title of the plot shows the parameter of the x-axis, the y-axis is the objective function NPE.

Catchment HG, simulation B, discharge data

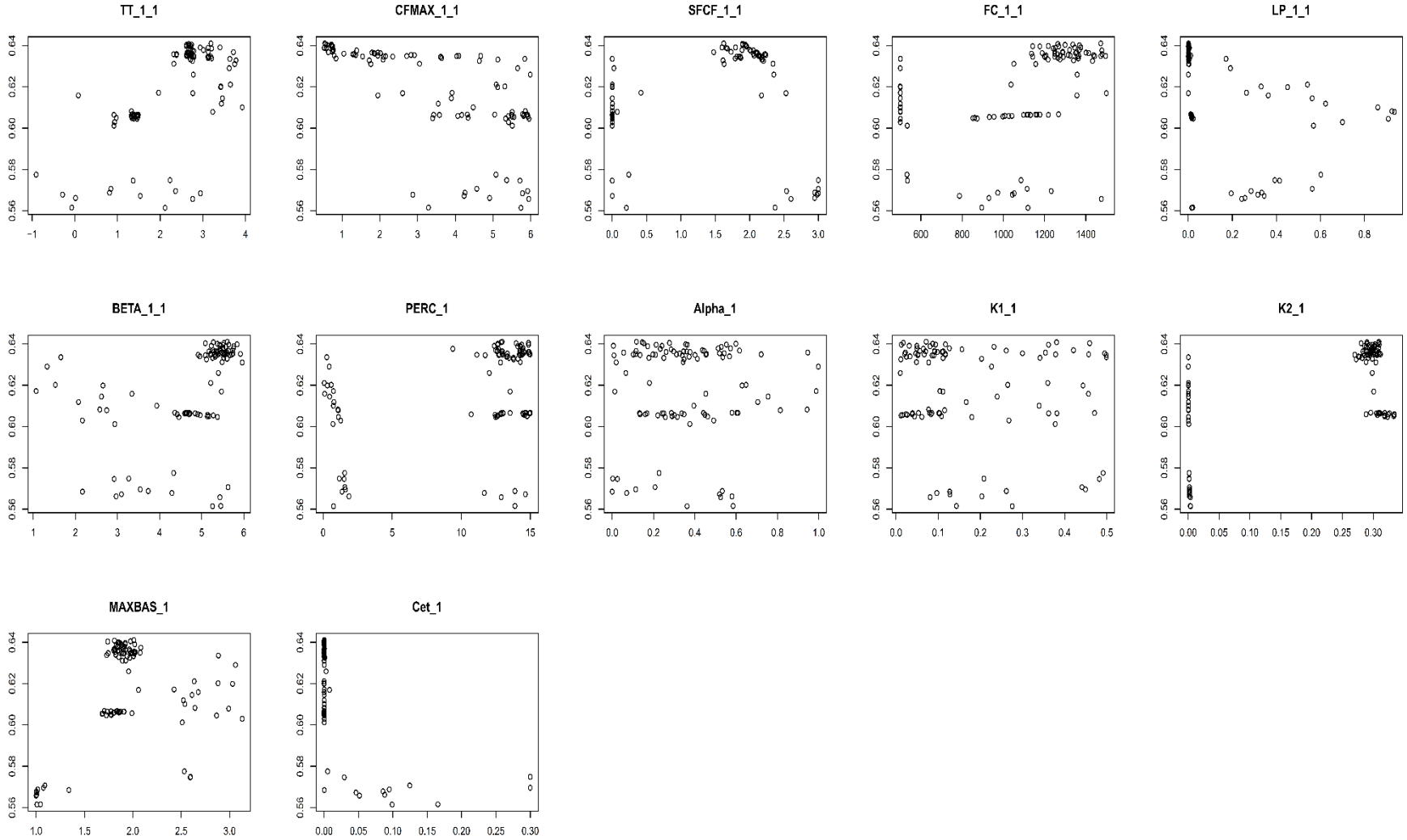


Figure 25: Dotty plot for catchment HG for simulation B for the 12 calibrated parameters with discharge data. The title of the plot shows the parameter of the x-axis, the y-axis is the objective function NPE.

9.9. “Best” calibrations per simulation and catchment for discharge data

Table 15: Parameter set per simulation and catchment with the highest NPE value for validation with the corresponding NPE value of the calibration.

Simulation	Catchment	Best validation parameter set	NPE Validation	NPE Calibration
A				
	W4.0	60	0.723	0.797
	W4.3	93	0.645	0.811
	HG	1	0.747	0.754
B				
	W4.0	29	0.711	0.804
	W4.3	98	0.682	0.781
	HG	83	0.555	0.774

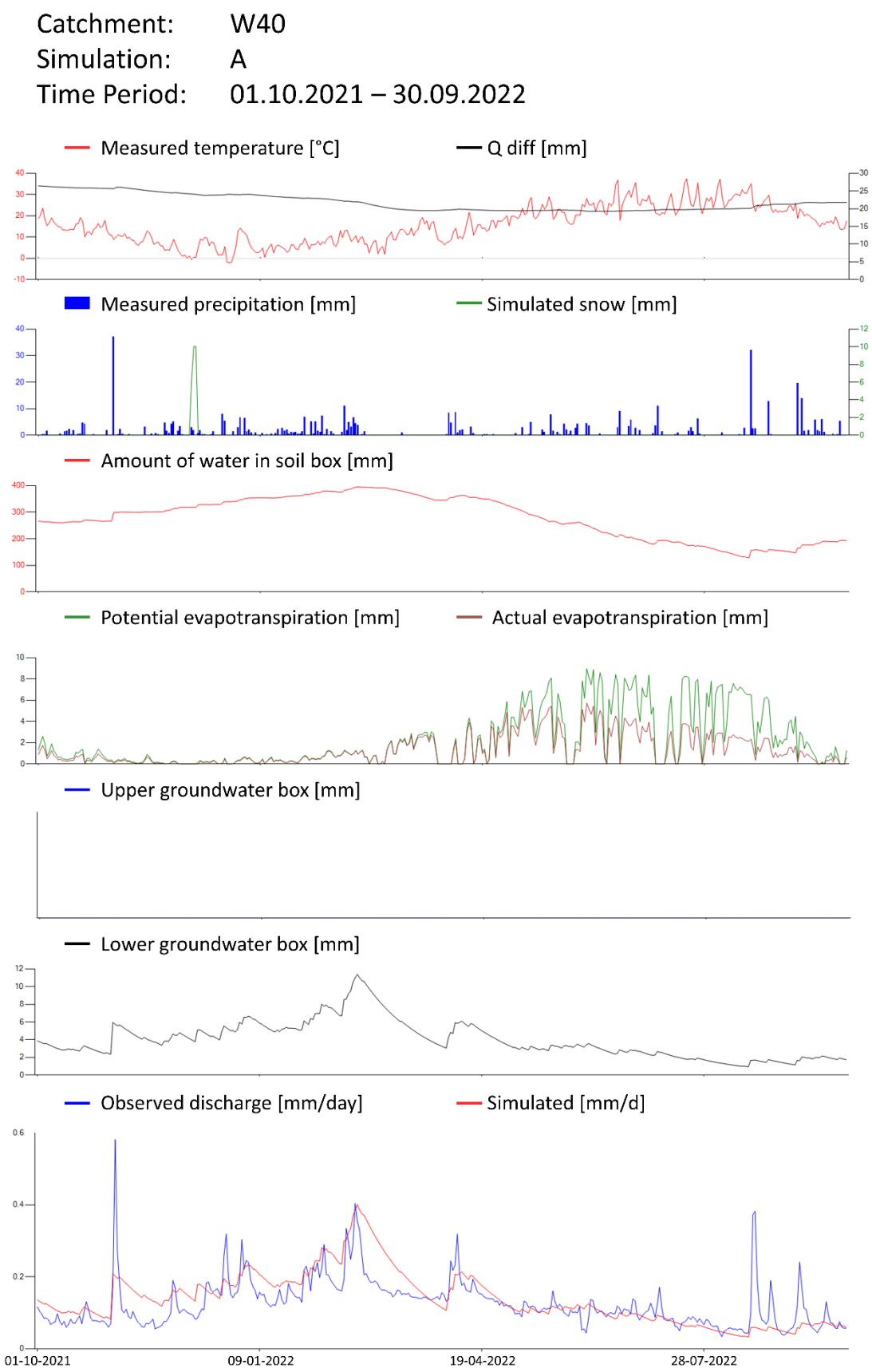


Figure 26: Composed HBV-light screenshot of all output simulated for one hydrological year (01.10.2021-30.09.2022) for simulation A, catchment W4.0.

N

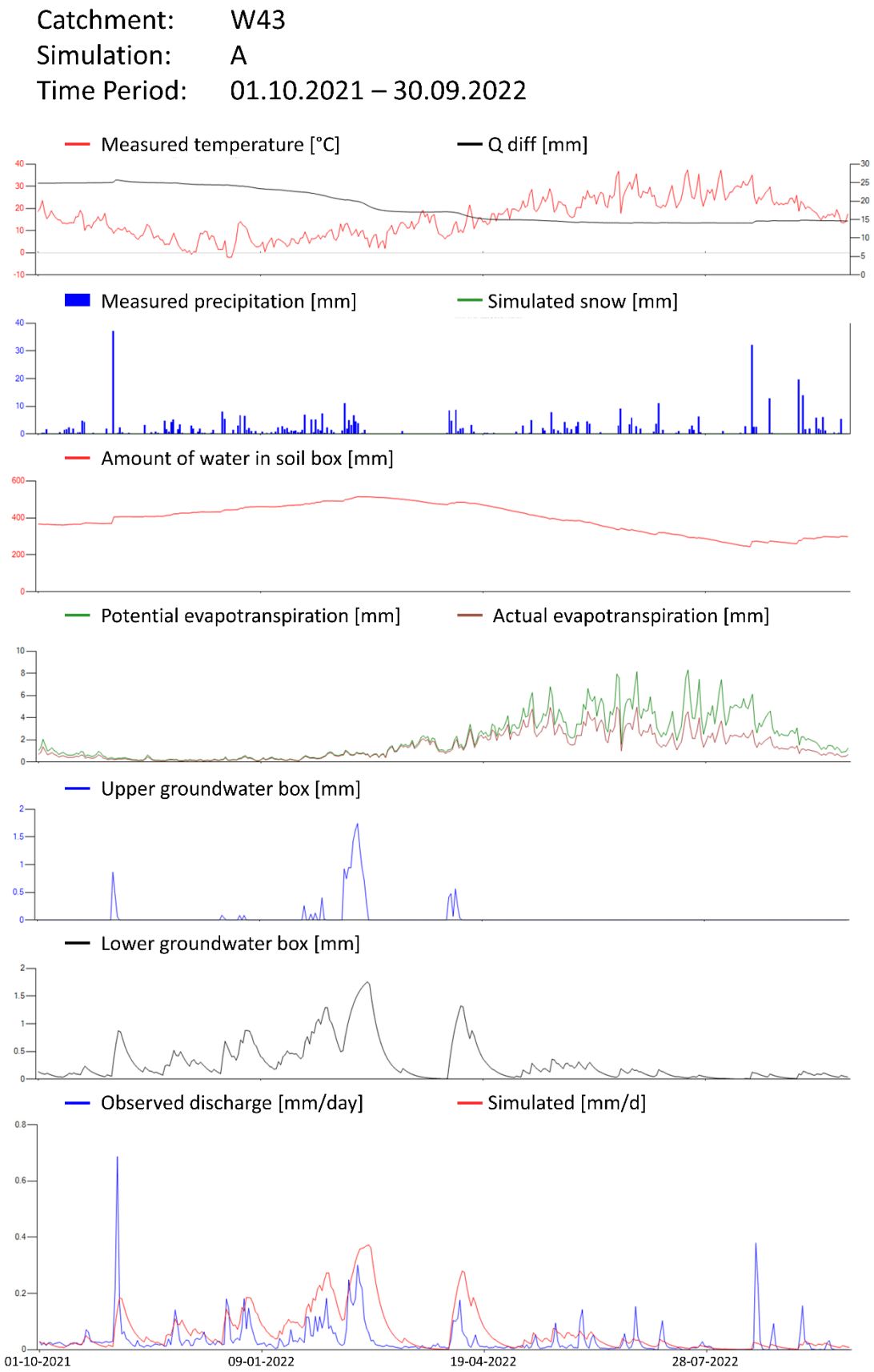


Figure 27: Composed HBV-light screenshot of all output simulated for one hydrological year (01.10.2021-30.09.2022) for simulation A, subcatchment W4.3.

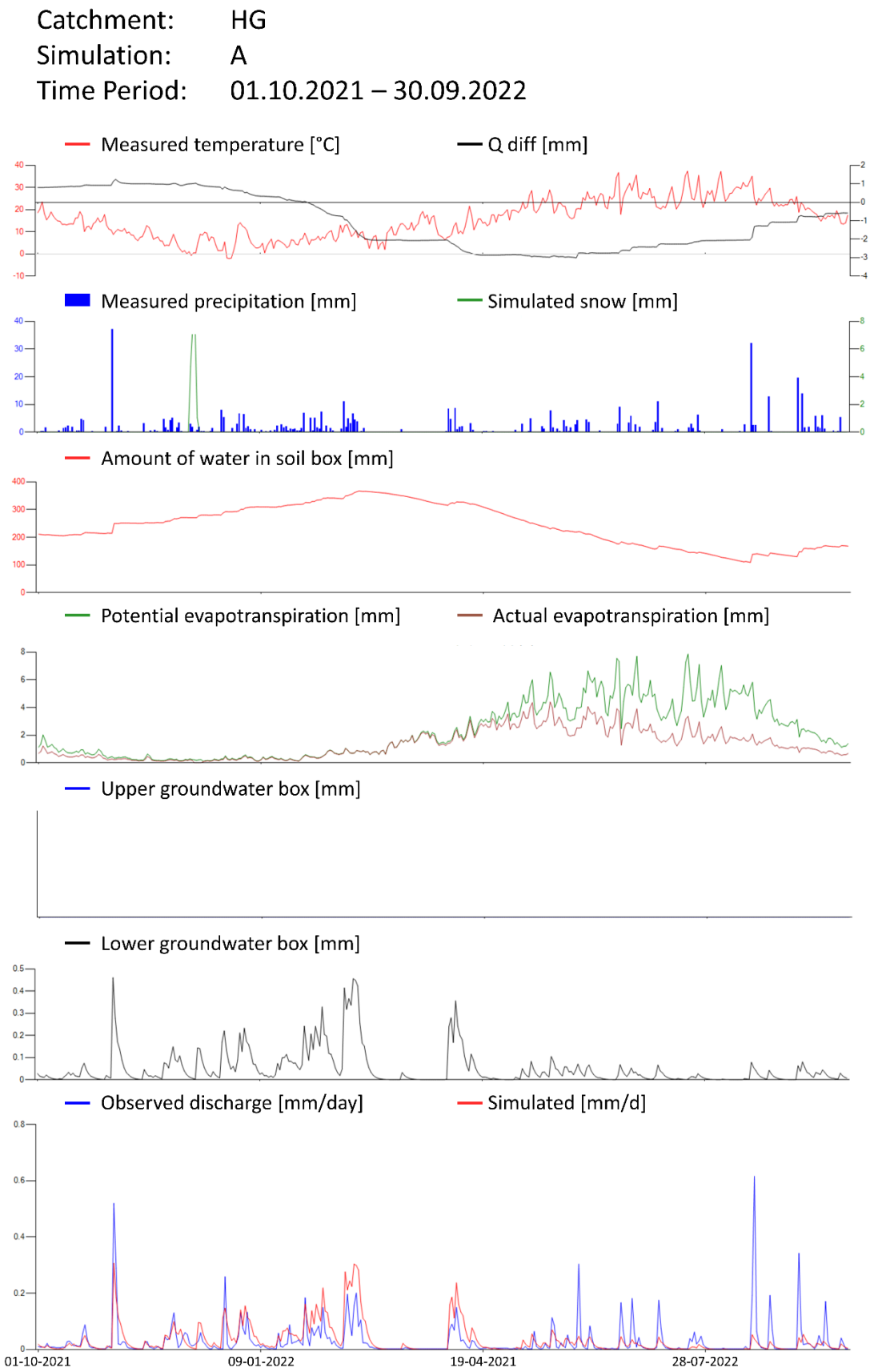


Figure 28: Composed HBV-light screenshot of all output simulated for one hydrological year (01.10.2021-30.09.2022) for simulation A, subcatchment HG.

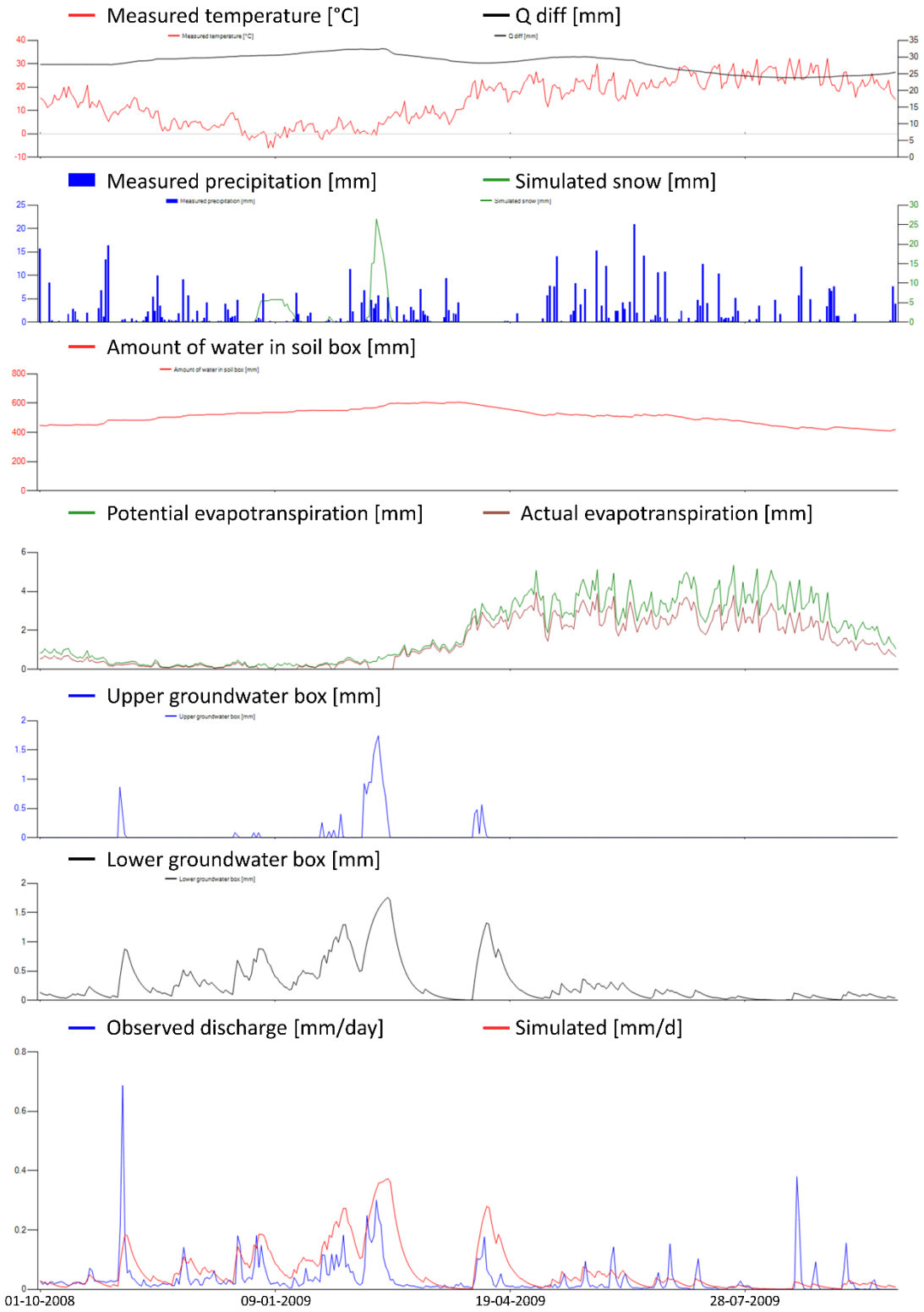


Figure 29: Composed HBV-light screenshot of all output simulated for one hydrological year (01.10.2008-30.09.2009) for simulation A, catchment W4.0.

Q

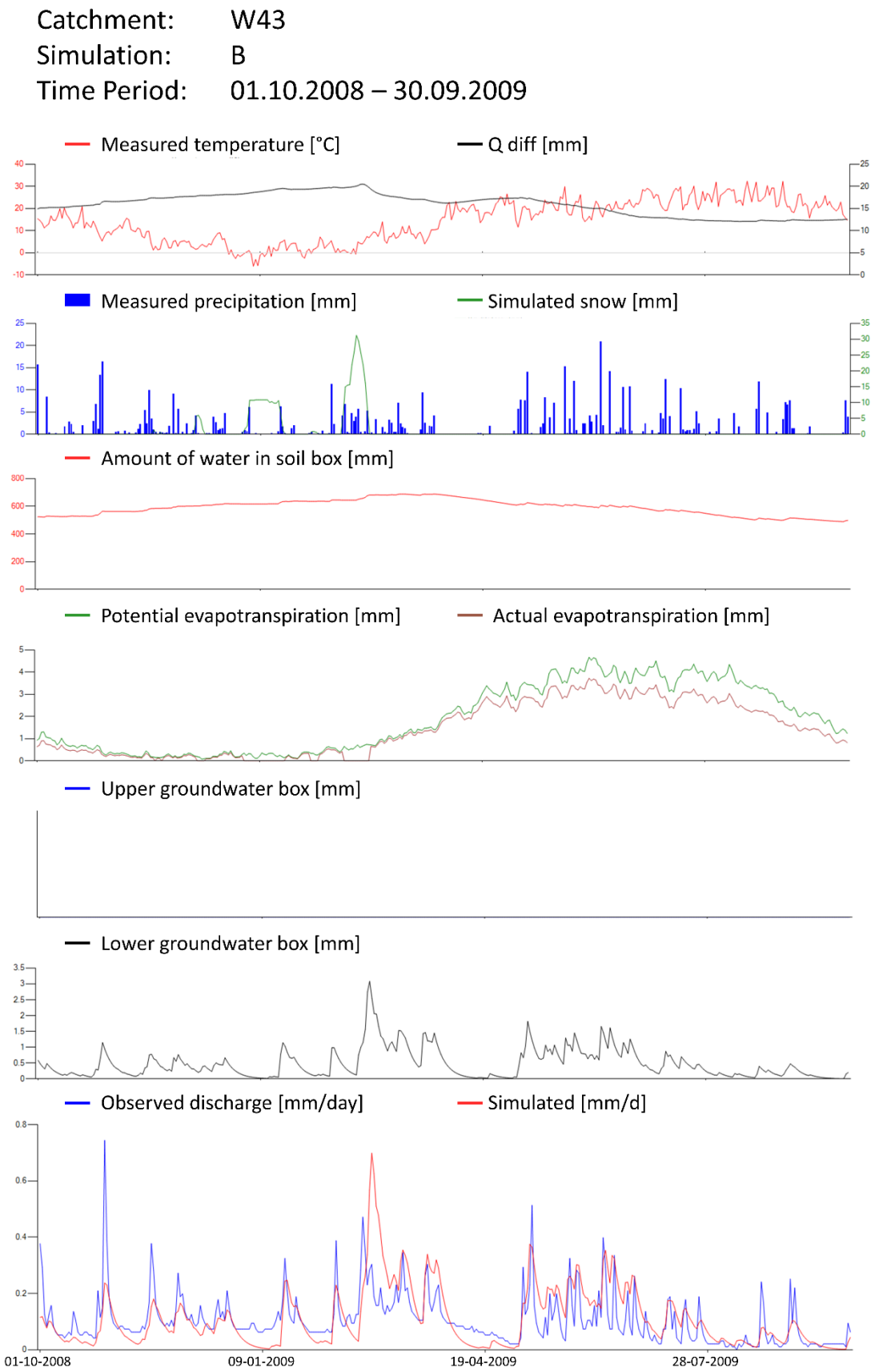
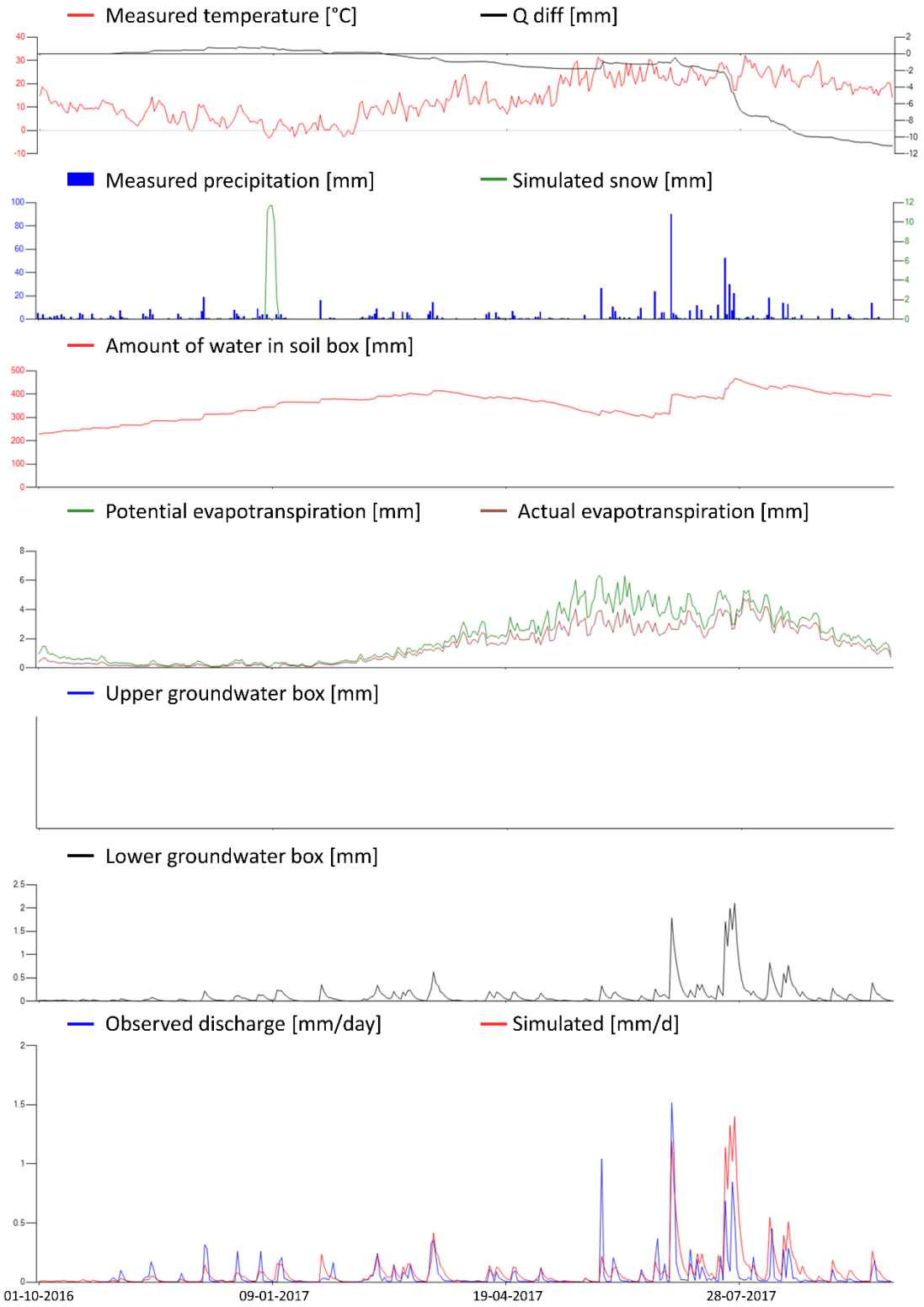


Figure 30: Composed HBV-light screenshot of all output simulated for one hydrological year (01.10.2008-30.09.2009) for simulation A, subcatchment W4.3.

Catchment: HG
Simulation: B
Time Period: 01.10.2016 – 30.09.2017



(DWD, 2018; DWD, 2021; DWD, 2005)

9.10. Objective function values for the hydrological year 2021-2022

Table 16: Objective function values for the hydrological year 2021 to 2022 for the validation period simulation A for all catchments. Simulation B represents the hydrological year of 2008 to 2009 for W4.0 and W4.3 and 2016 to 2017 for HG. Marked in orange are the negative values.

Catchment	**W4.0**		**W4.3**		**HG**	
Simulation	**A**	**B**	**A**	**B**	**A**	**B**
NPE	0.771	0.704	0.102	0.625	0.709	0.193
KGE	0.678	-0.262	0.030	-	0.392	-0.038
Reff	0.321	-2.314	0.399	-0.599	0.289	-0.833
r_s	0.776	0.725	0.629	0.738	0.768	0.706
VE	0.725	0.596	0.396	0.387	0.259	-0.284

9.11. Further results water level

9.11.1. Dotty plots for simulation A with water level data

Catchment W4.0, simulation A, water level

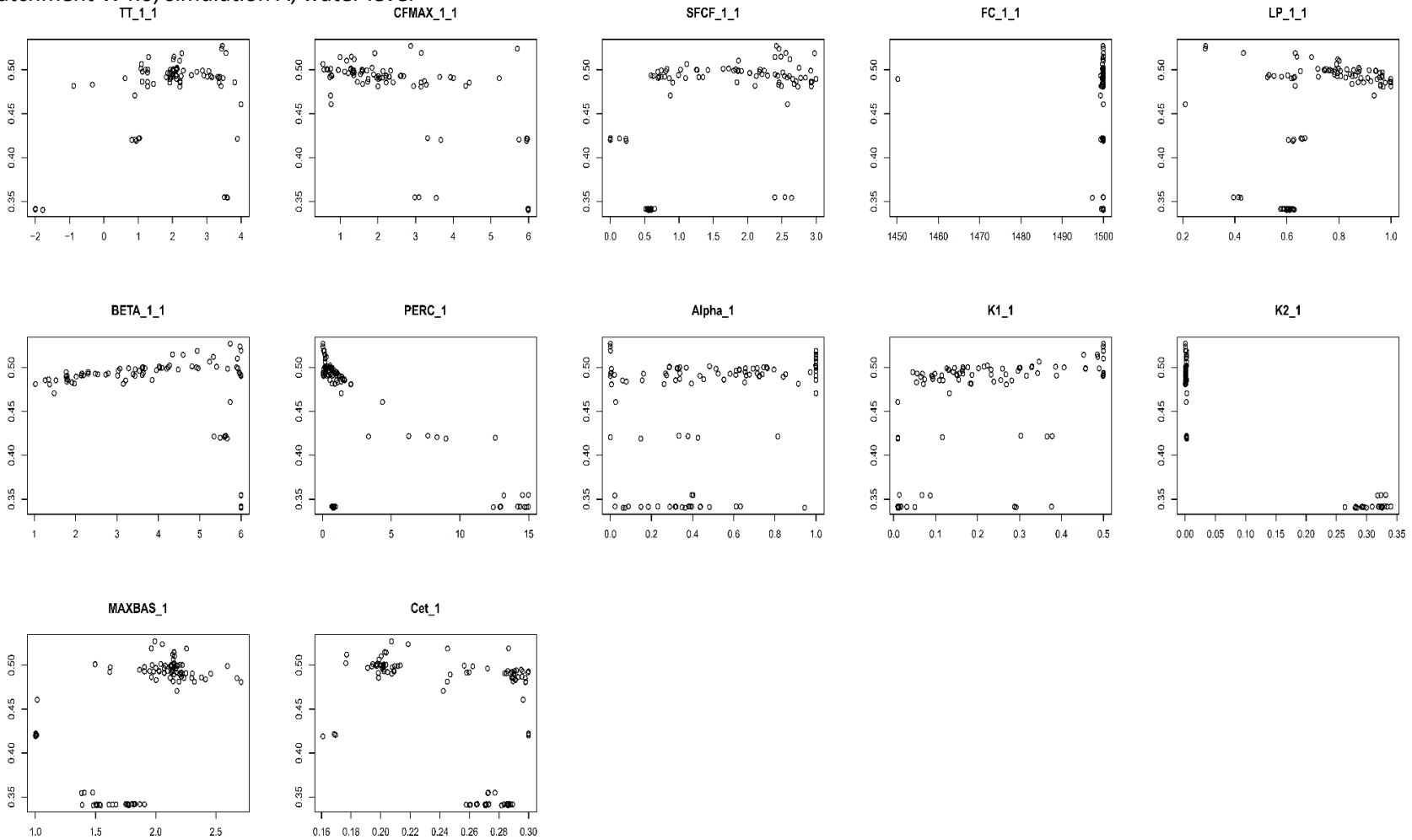


Figure 31: Dotty plot for catchment W4.0 for simulation A for the 12 calibrated parameters with water level data. The title of the plot shows the parameter of the x-axis, the y-axis is the objective function r_s.

Catchment W4.3, simulation A, water level

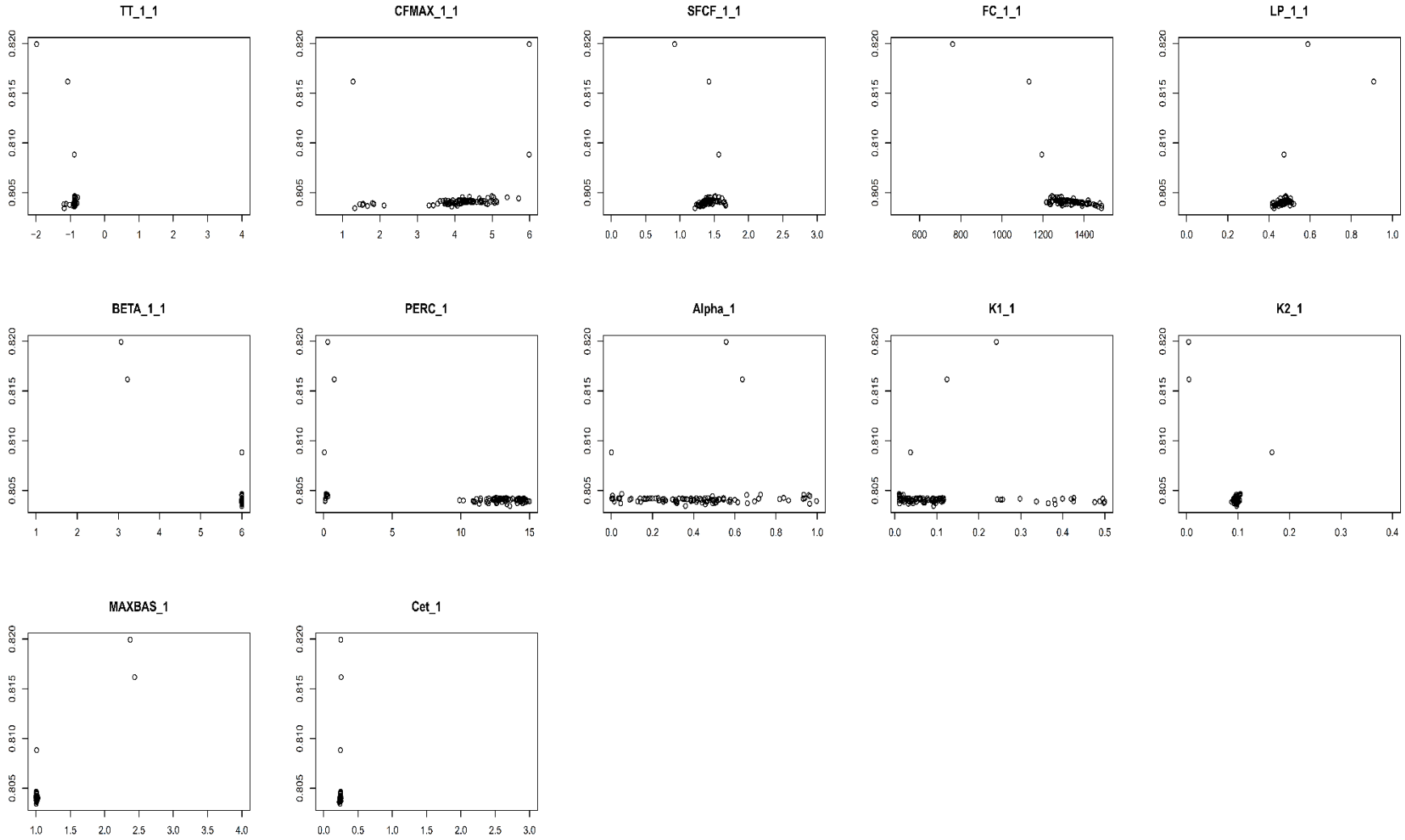


Figure 32: Dotty plot for catchment W4.3 for simulation A for the 12 calibrated parameters with water level data. The title of the plot shows the parameter of the x-axis, the y-axis is the objective function r_s.

Catchment W4.4, simulation A, water level

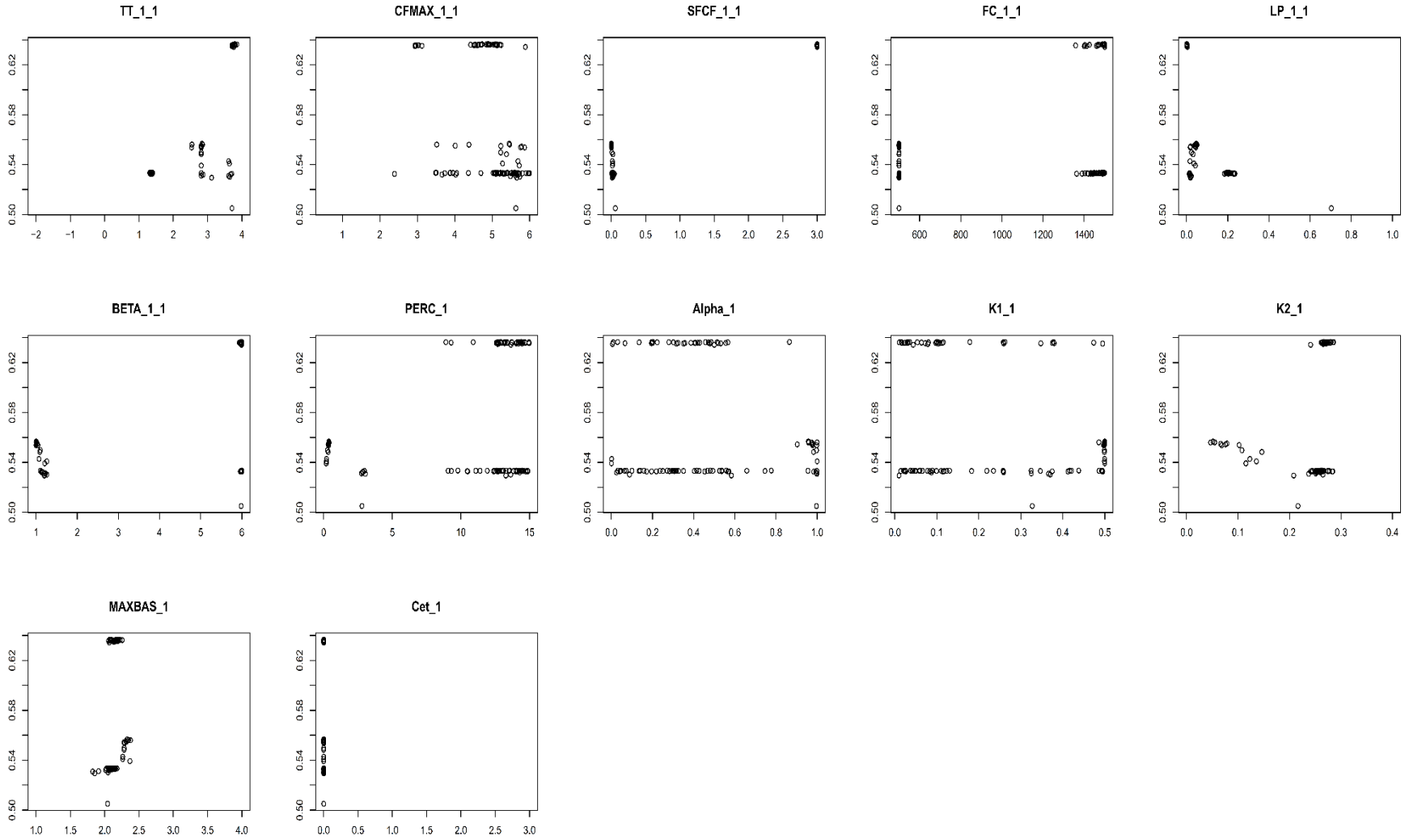


Figure 33: Dotty plot for catchment W4.4 for simulation A for the 12 calibrated parameters with water level data. The title of the plot shows the parameter of the x-axis, the y-axis is the objective function r_s.

Catchment HG, simulation A, water level

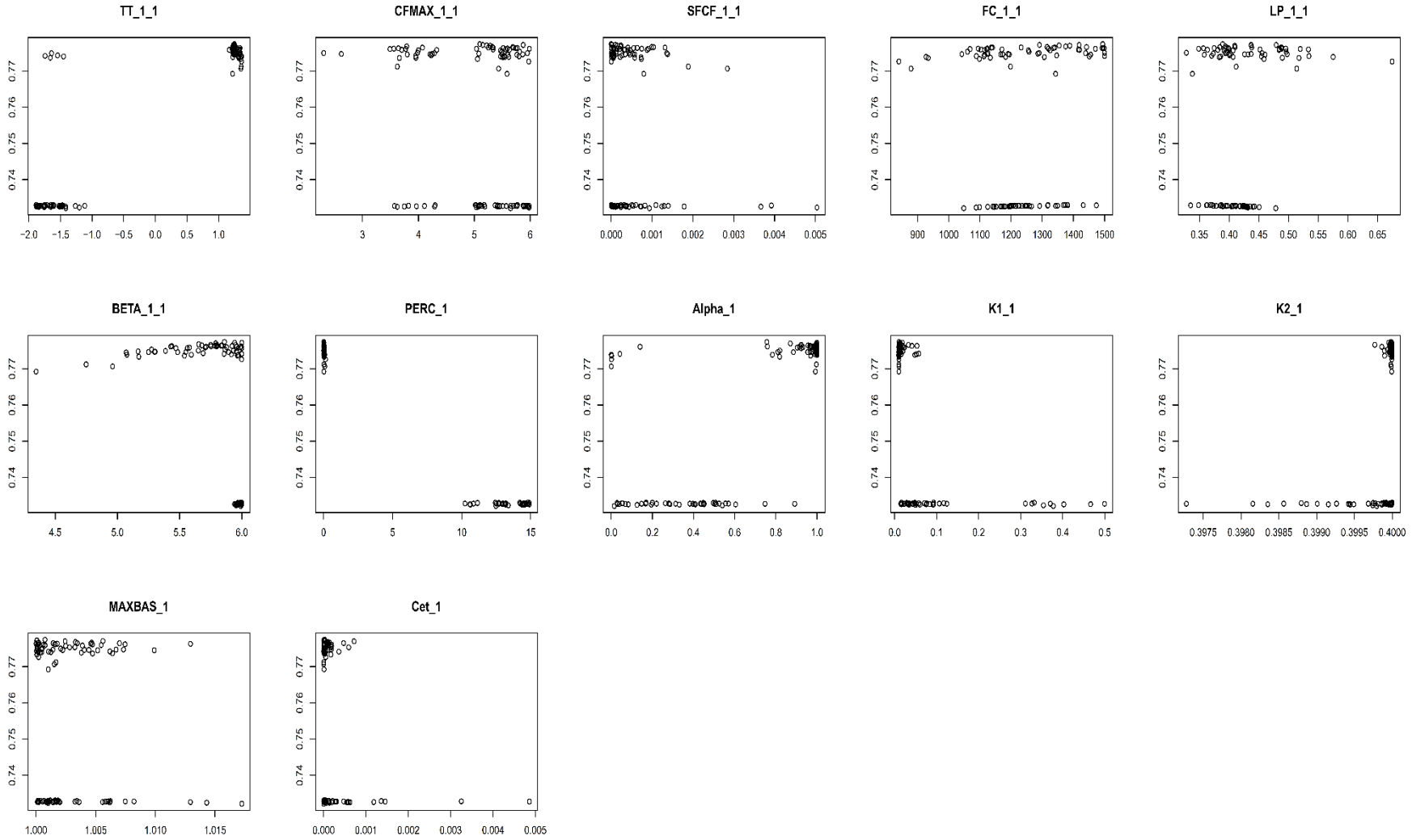


Figure 34: Dotty plot for catchment HG for simulation A for the 12 calibrated parameters with water level data. The title of the plot shows the parameter of the x-axis, the y-axis is the objective function r_s.

9.11.2. Dotty plots for simulation B with water level data

Catchment W4.0, simulation B, water level

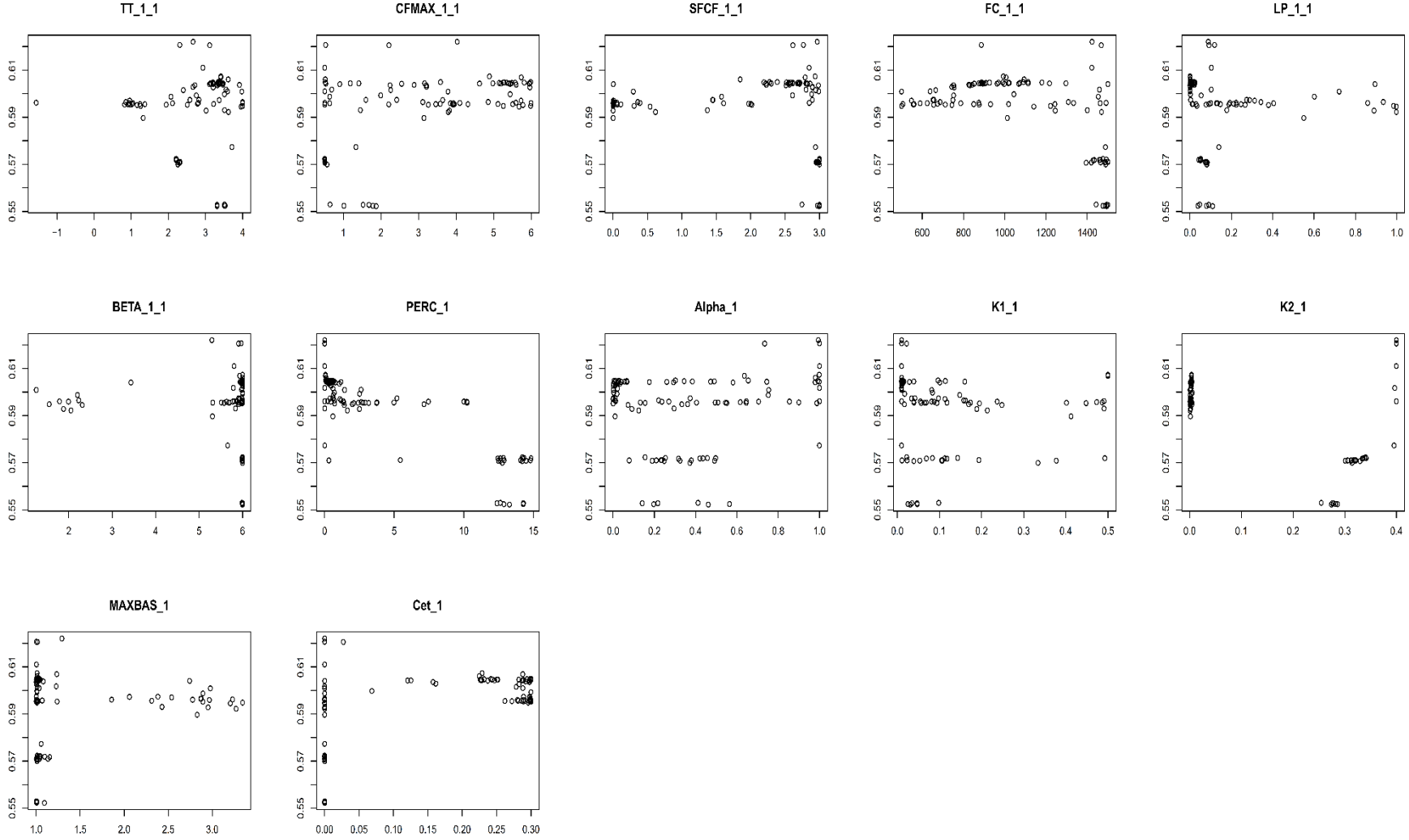


Figure 35: Dotty plot for catchment W4.0 for simulation B for the 12 calibrated parameters with water level data. The title of the plot shows the parameter of the x-axis, the y-axis is the objective function r_s.

Y

Catchment W4.3, simulation B, water level

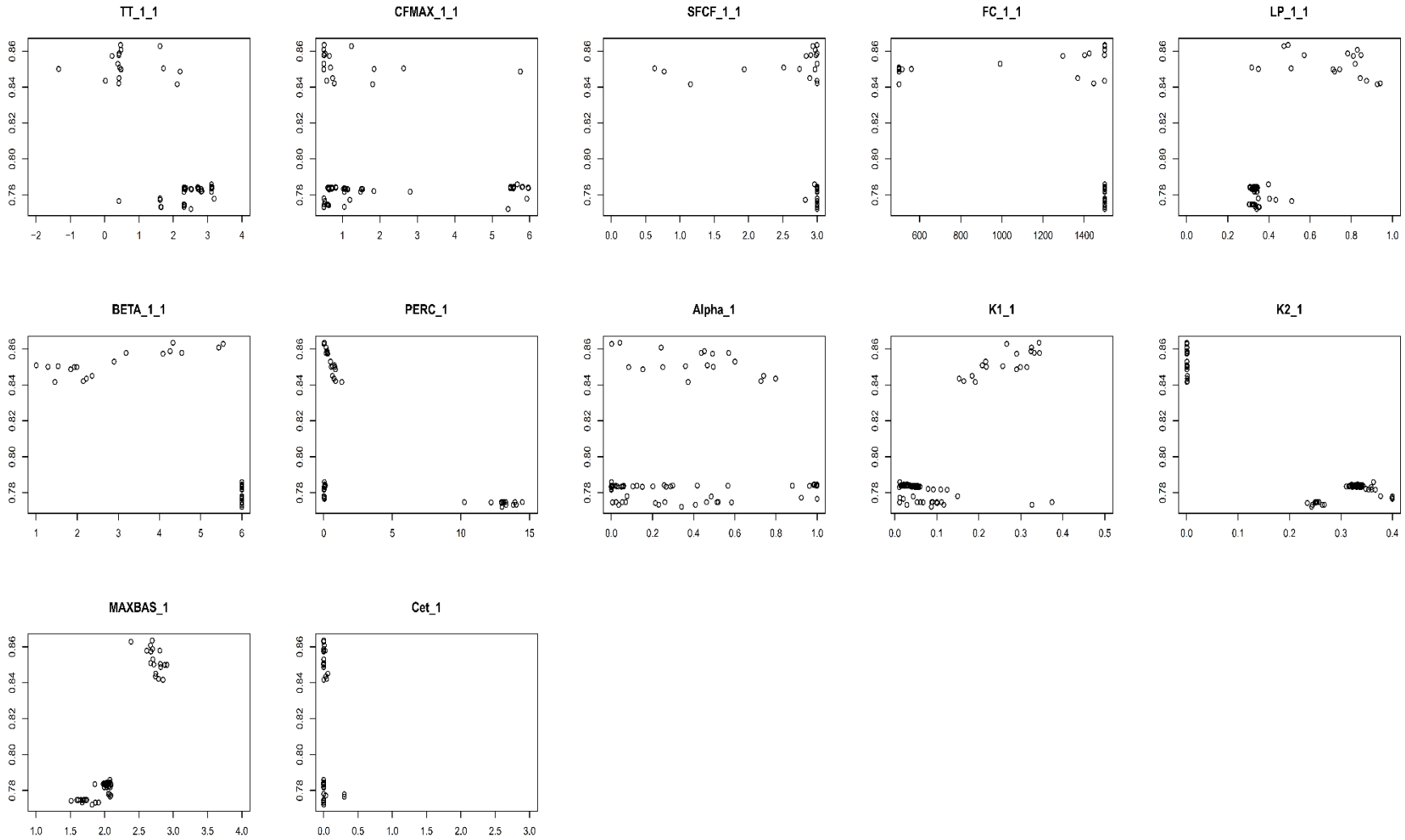


Figure 36: Dotty plot for catchment W4.3 for simulation B for the 12 calibrated parameters with water level data. The title of the plot shows the parameter of the x-axis, the y-axis is the objective function r_s.

Catchment W4.4, simulation B, water level

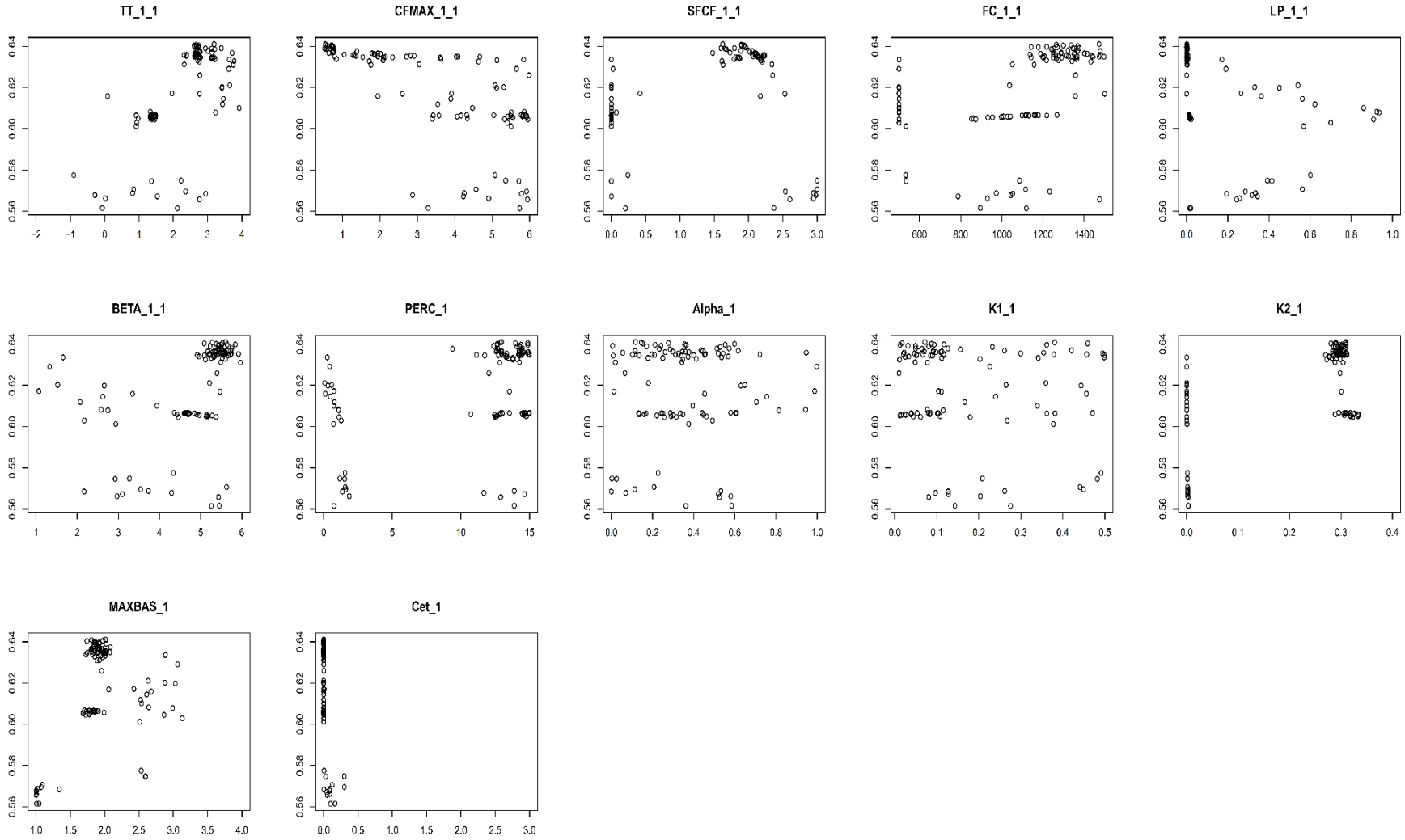


Figure 37: Dotty plot for catchment W4.4 for simulation B for the 12 calibrated parameters with water level data. The title of the plot shows the parameter of the x-axis, the y-axis is the objective function r_s.

Catchment HG, simulation B, water level

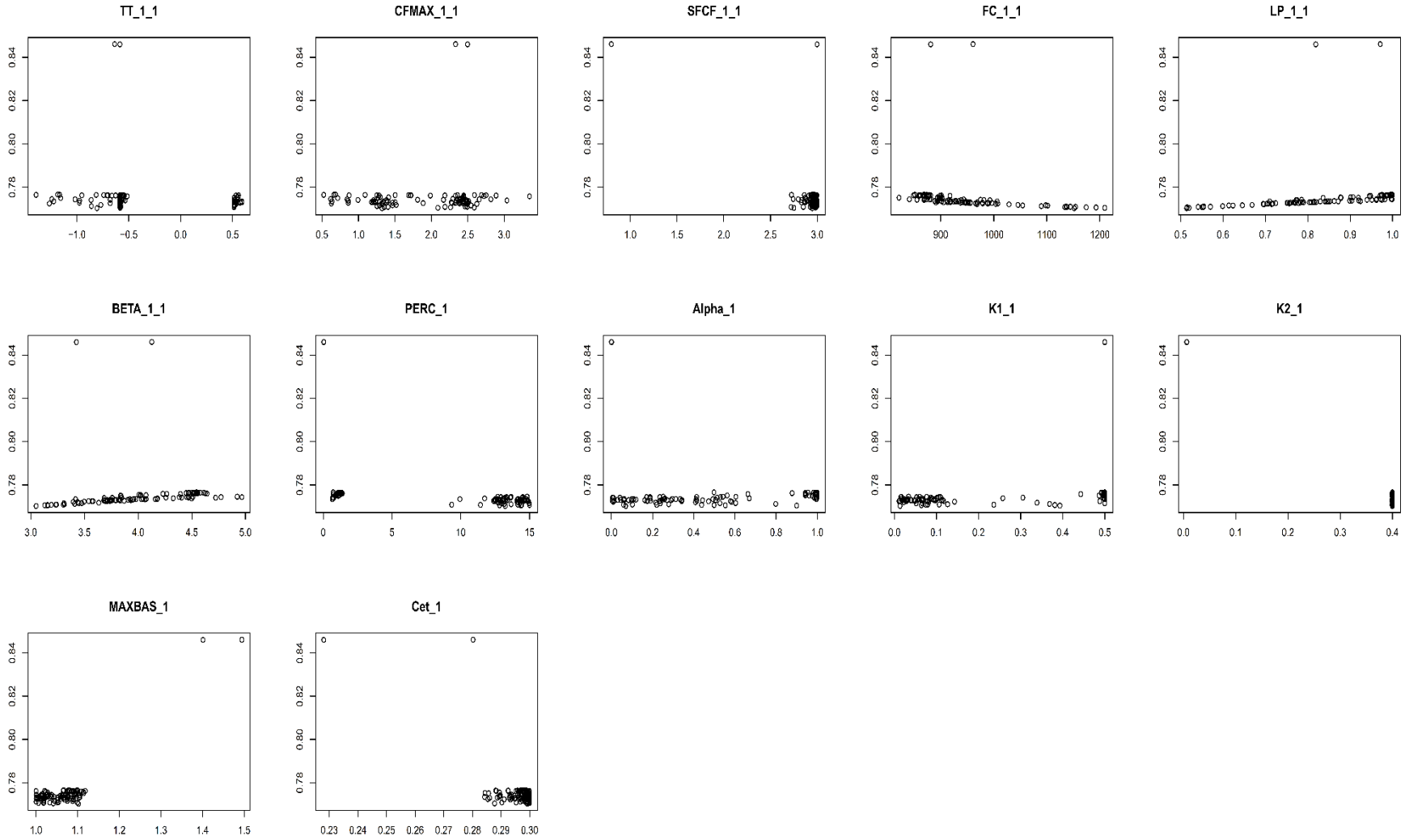


Figure 38: Dotty plot for catchment HG for simulation B for the 12 calibrated parameters with water level data. The title of the plot shows the parameter of the x-axis, the y-axis is the objective function r_s.

9.11.3. Water level benchmarks

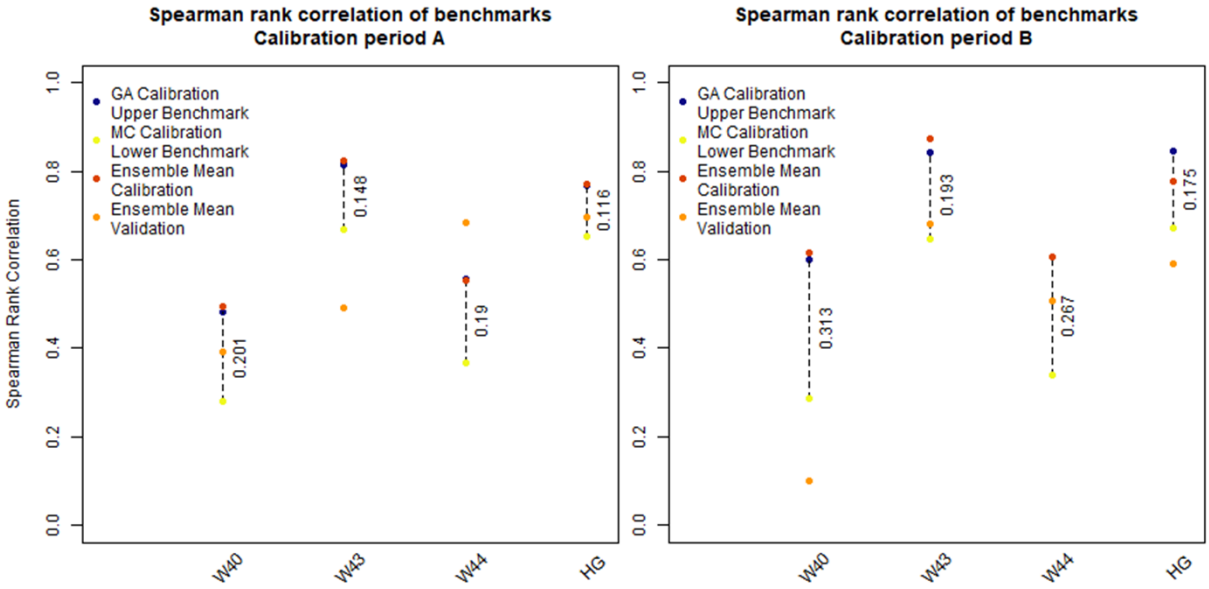


Figure 39: Benchmark and result of calibration period for water level data for each catchment with r_s as objective function. The number describes the difference between lower benchmark, calculated with the Monte-Carlo approach (yellow) and the upper benchmark calibrated with the GAP approach (blue) the best calibration result and the ensemble mean (red).

9.11.4. Model performance of model calibrated with water level data

Corresponding values of Figure 39 in summarized in a table.

Table 17: Water level data summary of objective function r_s values for all catchments for the simulation A and B. The upper benchmark is calculated with the GAP optimisation, the lower benchmark with the Monte Carlo approach. The ensemble mean is calculated once for the calibration and once for validation period. The relative objective function describes the relation between the difference of validation ensemble mean and the lower benchmark to the difference of upper and lower benchmark.

Catchment	**W40**		**W43**		**W4.4**		**HG**	
Simulation	**A**	**B**	**A**	**B**	**A**	**B**	**A**	**B**
Upper benchmark	0.481	0.601	0.816	0.841	0.556	0.608	0.769	0.846
Ensemble mean lower benchmark	0.281	0.288	0.669	0.648	0.367	0.340	0.653	0.671
GAP ensemble mean calibration	0.494	0.616	0.824	0.873	0.553	0.608	0.770	0.777
GAP ensemble mean validation	0.392	0.100	0.491	0.680	0.683	0.508	0.696	0.593
Relative Objective Function	0.555	-0.573	-1.211	0.166	1.672	0.627	0.371	-0.446

10. Personal declaration

I hereby declare that the submitted thesis is the result of my own, independent work. All external sources are explicitly acknowledged in the thesis.

24.01.2024,
Anna Czerniejewska