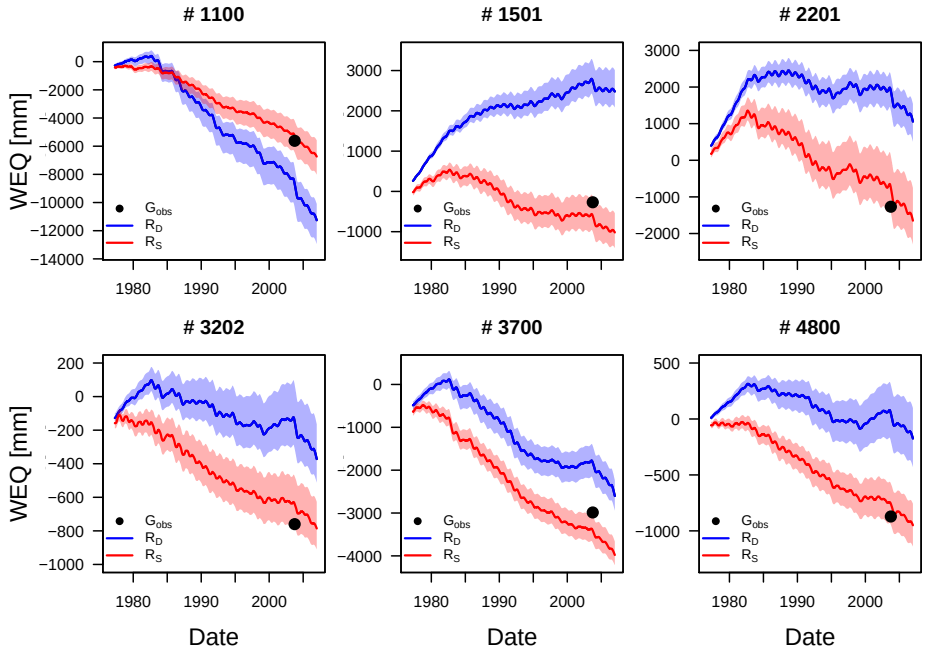
Chair of Hydrology

Albert-Ludwigs-University Freiburg i.Br.

Samuel Mayer – 3304294

Diagnostic Calibration of a Hydrological Model in Mountainous Watersheds



Master thesis under the supervision of Prof. Dr. Martin Weiler

Freiburg i.Br., May 15, 2018

Chair of Hydrology

Albert-Ludwigs-University Freiburg i.Br.

Samuel Mayer – 3304294

Diagnostic Calibration of a Hydrological Model in Mountainous Watersheds

Supervisor:
Prof. Dr. Martin Weiler

Co Supervisor:
Prof. Dr. Jan Seibert

Master thesis under the supervision of Prof. Dr. Martin Weiler
Freiburg i.Br., May 15, 2018

Acknowledgements

During the course of my thesis, a number of people helped me in very different ways. First and foremost I would especially like to thank Markus Weiler and Jan Seibert for the supervision of this work as well as for giving me the opportunity to write about a highly interesting subject. A special thanks goes to Irene Kohn who elaborately answered all my questions and provided me with the data and information required for this work. I am grateful to Jasper Vrugt and Florian Hartig for sharing their respective version of the DREAM algorithm. Last but not least I would like to express my gratitude to Kathrin, Janusch and Simon for their valuable comments and support as well as a great time.

Contents

List of Figures

List of Tables

Abstract

Runoff from snow and glacier melt can sustain streamflow during prolonged summer dry spells. The estimation of these runoff contributions to total catchment discharge, however, is a challenging task. One way to asses these contributions is to use glacio-hydrological modelling. The calibration of such models, however, often results in numerous parameter combinations with similar model efficiencies, which leads to uncertainties in model output. This issue can be addressed by incorporating multiple sets of observational data as well as different objective functions into model calibration in order to obtain hydrologically meaningful model parameters. In this study, the conceptual HBV-light model is calibrated to data of six partially glacierised headwater catchments of the Rhine river in the Swiss Alps. A time series for the calibration of glacier melt runoff is derived by hydrograph partitioning based on diurnal discharge amplitudes of total catchment runoff. Furthermore, snow melt runoff is calculated by using snow water equivalent data. The optimisation framework is mostly drawing upon likelihood based hydrological signatures, to enable model calibration rooted in Bayesian inference, using the DREAM optimisation algorithm. Modelling results are compared to an optimisation framework, not including time series of snow and glacier melt runoff as well as using residual based summary metrics as objective functions. The results reveal a systematic malfunction in the applied hydrograph partitioning. A comparison with data on glacier mass balance shows that runoff from glacier melt is strongly underestimated by the partitioned time series. For the employed hydrograph partitioning to be hydrologically meaningful, a number of enhancements are proposed. A comparison of the different optimisation frameworks reveals a better representation of overall hydrological catchment behaviour by optimising residual based summary metrics as compared to the employed hydrological signatures.

Keywords: hydrological modelling · diagnostic calibration · hydrograph partitioning · glacier runoff · hydrological signatures

Zusammenfassung

Schmelzwasser aus Schnee- und Gletscherschmelze kann im Sommer, während längerer Trockenperioden, den Gerinneabfluss aufrecht erhalten. Das Abschätzen dieser Abflussanteile am gesamten Gerinneabfluss stellt jedoch eine Herausforderung dar. Eine Möglichkeit dieser Abschätzung ist die Nutzung glazio-hydrologische Modellierung. Die Kalibrierung solcher Modelle resultiert allerdings oft in mehreren Parameterkombinationen mit vergleichbarer Modellgüte, was zu Unsicherheiten im Modelloutput führt. Diese Unsicherheit kann reduziert werden, indem mehrere Zeitreihen beobachteter Daten, als auch mehrere Zielfunktionen in der Modellkalibrierung kombiniert werden. In der vorliegenden Arbeit wird das konzeptionelle HBV-light Modell auf Daten aus sechs teilvergletscherten Kopfeinzugsgebieten des Rheins in den Schweizer Alpen kalibriert. Zur Kalibrierung des Gletscherabflussanteils kommt eine Ganglinienseparation, basierend auf täglichen Abflussamplituden zur Anwendung. Eine Zeitreihe des Abflussanteiles aus der Schneeschmelze wird aus Schnewasserequivalentsdaten abgeleitet. Die Modellkalibrierung basiert hauptsächlich auf hydrologischen Signaturen, um eine Optimierung im Rahmen Bayes'scher Inferenz zu ermöglichen, welche mit dem DREAM Optimierungsalgorithmus ausgeführt wird. Die Modellergebnisse werden mit einem Optmierungsansatz verglichen, welcher nicht auf Daten für Schnee- und Gletscherabfluss kalibriert und für welchen residuanbasierte Zielfunktionen zur Anwendung kommen. Die Ergebnisse zeigen systematische Fehler in der angewendeten Ganglinienseparation. Ein Vergleich mit Daten zur Gletschermassenbilanz der jeweiligen Einzugsgebiete zeigt eine Unterschätzung des Gletscherabflusses durch die angewendete Ganglinienseparation. Um der systematischen Abweichung entgegenzuwirken, werden mögliche Verbesserungen der angewendeten Ganglinienseparation vorgeschlagen. Der Vergleich der verschiedenen Optimierungsansätze zeigt, dass die Optimierung residuenbasierter Zielfunktionen zu Parameterkombinationen führt, die das hydrologische Verhalten der untersuchten Einzugsgebiete besser abbildet als die Optimierung der verwendeten hydrologischen Signaturen.

Stichworte: hydrologische Modellierung · diagnostische Kalibrierung · Ganglinienseparation · Gletscherabfluss · hydrologische Signaturen

1 Introduction

Snow and glacier melt in mountainous headwater catchments can sustain streamflow in larger rivers downstream during periods with no or negligible precipitation. This meltwater can be crucial for meeting water demands of downstream water users during prolonged dry spells (e.g. Kaser et al., 2010). Higher melting rates due to a warming climate can increase the water available from glacier melt during dry periods even though most glaciers are retreating. This so called glacier compensation can sustain streamflow from glacier melt at a high level until glacier mass eventually falls below a certain level (Frenierre & Mark, 2014; Huss et al., 2014). Furthermore, effects on shorter time scales can be observed in the streamflow of mountainous headwater catchments. Sub daily variations in energy available for melt processes lead to a diurnal amplitude in river discharge in catchments influenced by snow and glacier melt (e.g. Seibert et al., 2015; Hock, 2005; Jansson et al., 2003).

These long-term, seasonal and short-term effects of snow and especially glacier melt can influence downstream water users. Therefore, the potential impact of retreating glaciers on runoff generated from glacier melt on downstream river discharge necessitate a profound understanding and forecasting ability of glacier as well as snow melt driven runoff in headwater catchments.

Estimating runoff contributions of glacier melt to total streamflow remains a challenging task. Different approaches for estimating this contribution – including hydrochemical tracers and glacio-hydrological modelling – are reviewed by Frenierre and Mark (2014). Melt processes can be modelled with spatially distributed, physically based models solving the energy balance equation (e.g. Ragettli & Pellicciotti, 2012). A simpler, more computationally efficient method are conceptual models, mostly using empirical relationships between meteorological information (mainly air temperature) and melt processes. Due to the high correlation between air temperature and energy balance components, such conceptual temperature index models often closely match or even outperform fully distributed, physically based models on a daily time step (Hock, 2005, 2003; Rango & Martinec, 1995).

To fit observed data, however, conceptual hydrological precipitation-runoff models require parameter calibration (Beven, 2012). In the early days of hydrological

modelling this calibration was carried out manually. As a goodness of fit (GOF) measure mostly visual analysis of e.g. Unit Hydrograph parameters (e.g. Nash, 1959) or the scaling coefficient of the Rational Method was used (Chow et al., 1964; Linsley et al., 1949).

With increasing availability of computational power, an evolution towards more complex hydrological models can be observed. Many of these - mostly conceptual - models were designed to reproduce observed time series with a large number of model parameters (e.g. Fleming, 1975). The Stanford Watershed Model developed in the 1970s e.g. has more than 30 tunable parameters (Loague, 2010). With a high number of tunable model parameters, the parameter response surface becomes too complex to be examined manually. As a result, manual parameter calibration was replaced by automated calibration procedures (Clarke, 1973).

For automated calibration, a formulation of model efficiency (or malfunction) is needed as a benchmark. The objective functions commonly used in the hydrologic community are residual based metrics such as the Nash-Sutcliffe-Efficiency (NSE; Nash & Sutcliffe, 1970) or squared error functions such as the Root-Mean-Squared-Error (RMSE) - amongst others. Such residual based objective functions are often aiming at reproducing specific parts of the time series. The NSE, for example, emphasizes on matching peak flows rather than the whole time series (Beven, 2012; Krause et al., 2005). Furthermore, residual based performance criteria can be strongly influenced by timing errors as well as autocorrelation of the residuals (Beven, 2012). To avoid a calibration on one specific part of the hydrograph, the need for weighted combinations of objective functions has been stated (e.g. Jain & Sudheer, 2008; Legates & McCabe, 1999). Hereby the ideal model efficiency is the pareto optimum of the weighted objective functions used.

Arguably residual based objective functions aim at the mathematical reproduction of observed data, rather than being based on a hydrologically meaningful background. In recent years the need for a more "diagnostic" model optimisation based on hydrological realism has been stated (Gupta et al., 2009, 2008; Schaefli & Gupta, 2007). Yilmaz et al. (2008) - amongst others - introduced an optimisation scheme, based on so-called hydrological signature measures (subsequently referred to as signatures), rather than the observed time series itself. The theory behind this approach is to obtain sets of model parameters which are meaningful

in a hydrological sense as well as being able to trace structural deficits in model conceptualisation (Gupta et al., 2008). One of the most commonly used signatures is the Flow Duration Curve (FDC; Searcy, 1959) or parts thereof (e.g. Sadegh et al., 2016, 2015; Pokhrel et al., 2012; Blazkova & Beven, 2009b; Son & Sivapalan, 2007; Yadav et al., 2007). Signatures based on parts of the FDC are for example the bias in peak-flow and low-flow volume derived from the upper and lower parts of the FDC respectively as well as the bias in the FDC midsegment slope. To still capture the observed flow pattern in time, e.g. the bias in peak flow timing, can be incorporated (Yilmaz et al., 2008).

In order to reduce the usage of residual based summary metrics to quantify the GOF of signatures, several studies by Sadegh et al. (2015); Sadegh and Vrugt (2014); Vrugt and Sadegh (2013) suggest an optimisation framework, which is drawing upon a more formal statistical approach. Coupled with residual error functions, a maximum likelihood based approach is used by the aforementioned authors to asses model efficiency based on FDC signatures. Maximising the likelihood of a dataset, given a functional representation of the observed data, provides a couple of advantages over the approach of minimising residual errors. Likelihood based GOF measures are less likely to violate statistical assumptions of some optimisation frameworks, such as the independence of residual errors (e.g. Beven, 2012; Schoups & Vrugt, 2010a) or the assumption of constant variance (Pechlivanidis et al., 2011). However, maximum likelihood optimisation approaches can be strongly influenced by extreme values in the observational data (e.g. Nguyen & Welsch, 2010).

An issue in parameter optimisation independent from the used objective function is the so-called problem of equifinality. The term equifinality refers to the possibility of different parameter sets resulting in the same or similar model efficiency based on the objective function used and is a well-known issue in hydrology as well as other research fields (e.g. Beven & Binley, 2014; Beven & Freer, 2001; Hornberger & Spear, 1981; Spear & Hornberger, 1980; Hornberger & Spear, 1980).

To reduce equifinality issues and increase hydrological realism, hydrological models or their respective subroutines can be optimised by fitting more than one set of observational data. The usefulness of so-called multi-objective calibration approaches with respect to model realism has been recognised by a number of studies.

Kelleher et al. (2017) use measured Snow-Water-Equivalent (SWE) and groundwater table depth in addition to runoff data. Including calibration on snow covered area (SCA) as well as glacier mass balance increases the realism of snow and glacier subroutines in a study by Finger et al. (2015). Stahl et al. (2017) calibrate model output on runoff data as well as SWE, SCA and glacier mass balance. In studies by Larabi et al. (2018), He et al. (2018, 2015) and Shafii et al. (2017), hydrograph partitioning of discharge time series is applied. Hereby a stepwise calibration of model subroutines on parts of the hydrograph, theoretically represented by the respective subroutine, increased model performance and realism compared to a calibration on the total runoff time series only. The findings of Rakovec et al. (2016) and Shafii and Tolson (2015) suggest, that such pareto optimum based, multi-objective calibration approaches outperform the calibration on a single set of observational data. Seibert and McDonnell (2002) state that compromising on the efficiency of one objective function of e.g. streamflow is acceptable for achieving higher model realism through multi objective model calibration by including further sets of observational data in the optimisation approach.

To automatically optimise model performance on a chosen GOF criterion, many different algorithms are used. With Brute Force approaches lacking efficiency, more sophisticated model optimisation techniques arose. Initially so called "Hill Climbing Algorithms" were used for the optimisation of rainfall-runoff models. In such approaches the n-dimensional (with n corresponding to the number of parameters tuned) response surface is evaluated by following the parameter space in directions, where the chosen GOF criterion is improved. One popular example of such techniques is the Simplex scheme by Nelder and Mead (1965). Especially in high dimensional parameter spaces the results of such hill-climbing algorithms are highly dependent on the initial parameter values of the search, which can result in getting stuck in a local optimum instead of finding the global optimum (Beven, 2012).

To increase the probability of finding the global optimum for a specific GOF criterion, so called genetic algorithms, based on biological evolution (thus referred to as evolutionary algorithms as well), are used in hydrological modelling (e.g. Kuczera, 1997; Franchini & Galeati, 1997). To avoid terminating in local minima, the evolution of the initial parameter set in genetic algorithms can include e.g. larger

parameter jumps, to increase the sampled surface of the parameter space (Forrest, 1993).

The Shuffled-Complex-Evolution algorithm (SCE), widely used in hydrological modelling, combines the ideas from the Simplex algorithm with genetic algorithm techniques. Starting with multiple parameter sets, several Simplex hill-climbing searches are carried out. To further increase robustness in finding global optima the parameter sets are shuffled in between iterations (Duan et al., 1993, 1992). The SCE is considered to be an efficient and robust algorithm for hydrological optimisation problems (Beven, 2012; Kuczera, 1997).

The aforementioned algorithms are designed to result in one "best" set of model parameters given the GOF criterion (Chu et al., 2010). In contrast there are optimisation schemes based in Bayesian Inference. Given a prior distribution of parameter density, a GOF measure and an optimisation scheme, a probabilistic estimation (posterior density distribution) of model parameters is calculated. In hydrological modelling, a widely used optimisation scheme based on Bayesian Inference are so called Markov-Chain-Monte-Carlo (MCMC) methods. MCMC schemes originate from Monte Carlo (MC) sampling. MC sampling randomly draws parameter sets from predefined parameter distributions. The naming of MC methods refers to the gambling in Monte Carlo casinos, since finding acceptable parameter sets eventually relies on luck (Beven, 2012).

In contrast, MCMC schemes update the prior parameter distributions according to the model evaluations carried out – and thus the density – in a specific part of the parameter space. Drawing new parameter sets does therefore not solely rely on luck but on the density of the updated prior distribution. To update the prior distribution efficiently, a "smart" sampling scheme is required. Many of the MCMC schemes applied in parameter optimisation of hydrological models use the Metropolis-Hastings algorithm (Hastings, 1970; Metropolis et al., 1953). A popular example is the Shuffled-Complex-Evolution-Metropolis (SCEM) algorithm – an enhancement of the SCE algorithm – were the aforementioned Simplex scheme is replaced with the Metropolis-Hastings scheme (Vrugt et al., 2003). For applications of the SCEM scheme in hydrological paramter optimisation see e.g. Guo et al. (2013); McMillan and Clark (2009); Feyen et al. (2007).

As a further development of the SCEM scheme, Vrugt et al. (2009, 2008a) in-

troduced the DiffeRential-Evolution-Adaptive-Metropolis (DREAM) algorithm, which uses a different subsampling scheme within the individual Markov Chains. DREAM and especially the DREAM$_{ZS}$ enhancement by Laloy and Vrugt (2012) proved to be robust in targeting the true posterior distribution of model parameters in several toy problems. Furthermore, the DREAM$_{ZS}$ algorithm showed its efficiency in a very high dimensional hydrological parameter calibration problem with more than 200 parameters (Laloy & Vrugt, 2012).

To asses model uncertainty in a Bayesian optimisation framework – especially considering equifinality issues – multiple methods evolved. Based on the MC sensitivity analysis by Hornberger, Spear and Young (Young, 1983; Hornberger & Spear, 1981; Spear & Hornberger, 1980; Hornberger & Spear, 1980), the Generalised Likelihood Uncertainty Estimation (GLUE) methodology arose. GLUE recognises the need to systematically asses model uncertainty (Beven & Binley, 1992). Although GLUE is designed to asses data-, model-, and parameter-uncertainty, only parameter-uncertainty is discussed hereafter.

The GLUE methodology aims at providing an uncertainty estimation by accepting multiple model parameter sets. Acceptable parameter sets – often referred to as behavioural (e.g Beven, 2007) – are chosen by predefining a level of acceptance given a GOF measure. All parameter sets which satisfy the subjectively chosen level of acceptance, are used to predict model output. This procedure results in a density distribution of the model output at each modelling time step. Using the density distribution, model output uncertainty can be examined (Beven, 2012, 2007).

Being based on MC runs, the GLUE methodology is computationally demanding. In recent years Approximate Bayesian Computation (ABC) has been incorporated as uncertainty framework for hydrological models (Vrugt & Sadegh, 2013). ABC is less time consuming than GLUE since it can be incorporated in MCMC optimisation schemes and does not need additional MC runs (Sadegh & Vrugt, 2014). Yet ABC and GLUE share similar concepts and the results of both methodologies are similar in nature (Sadegh & Vrugt, 2013). Furthermore, with ABC incorporated into MCMC optimisation schemes, the used GOF measure does not necessarily need to be likelihood based (Sadegh et al., 2015; Sadegh & Vrugt, 2013).

2 Purpose and Scope

Diagnostic model calibration, as mentioned above, can increase model realism by using multiple objective functions, several time series of catchment behaviour or hydrologically meaningful signatures. Especially including multiple sets of observational data to calibrate model subroutines has proven to reduce issues associated with the calibration of hydrological models such as equifinality problems.
The present work is a follow up on the study carried out by Stahl et al. (2017), who modelled the relative discharge fractions of runoff from rainfall, snow melt and glacier ice melt in the Rhine river. The headwater catchments where the meltwater components originate from, are modelled with a semi distributed conceptual glacio-hydrological model (Stahl et al., 2017). The goal of the work at hand is to develop a diagnostic model optimisation framework for the above mentioned conceptual model. The framework is especially designed for mountainous headwater catchments influenced by snow and glacier melt processes. The optimisation approach to be developed aims at being diagnostic by implementing likelihood based hydrological signatures as well as multiple sets of observational data.
Stahl et al. (2017) did not incorporate observational data to calibrate runoff components originating from snow and glacier melt. The main goal within the scope of this work, is to test the applicability of snow and glacier melt time series to separately calibrate the respective model subroutines, in order to increase hydrological model realism as well as highlight structural model inadequacies.
Recent studies used seasonal partitioning of observed hydrographs to calibrate model subroutines representing snow and glacier melt. The respective model subroutines where calibrated against parts of the hydrograph representing runoff which originates from melt processes (Larabi et al., 2018; He et al., 2018; Shafii et al., 2017; He et al., 2015). However, to the authors best knowledge, high resolution (sub daily) discharge data has not yet been used for hydrograph partitioning. With diurnal discharge amplitudes mostly originating from melt processes in glacierised catchments (Seibert et al., 2015; Hock, 2005; Jansson et al., 2003) it is hypothesised that:

- (i) Incorporating a time series of glacier melt discharge (derived from diurnal discharge amplitudes) into the calibration process, increases model realism

when modelling glacierised headwater catchments.

Furthermore, using hydrological signatures as GOF measures allows model calibration and evaluation based on hydrological theory. Following the work on diagnostic model calibration by Shafii and Tolson (2015); Gupta et al. (2012); Clark et al. (2011); Gupta et al. (2008) it is assumed that:

- (ii) Diagnostic model optimisation, based on hydrological signatures, produces relatively stable parameter sets, which are able to capture overall hydrological catchment behaviour as well as the runoff time series.

These goals and hypothesis are tested by comparing the results of two optimisation frameworks. One framework is based solely on residual based summary metrics, while the second approach incorporates hydrological signatures, where applicable.

3 Material and Methods

3.1 HBV-light Model

The Hydrologiska Byråns Vattenbalansavdelnin Model (HBV) is a spatially semi-distributed conceptual precipitation-runoff model developed in the 1970s (Bergström, 1976). Since the models first implementation, several different versions have been published as well as additional subroutines added (e.g. Stahl et al., 2017; Seibert & Beven, 2009; Seibert, 1999, 1997). For a comprehensive description of the original HBV model structure see Seibert (1999); Lindström et al. (1997); Bergström (1995).

The HBV-light model (version 4.0.0.10), developed at the University of Zürich is used subsequently. Compared to the implementation by Seibert and Vis (2012), the used version contains a series of enhancements (Stahl et al., 2017). A visual representation of the HBV model structure is given in figure 1. Being a semi distributed model, the HBV-light structure allows for a distribution of parameters and forcing data (precipitation, air temperature and estimated long-term monthly potential evaporation) with elevation and vegetation zones and/or for different subcatchments (Seibert & Vis, 2012). In accordance with Stahl et al. (2017) elevation zones of 100 meters are implemented. The used HBV-light version includes a snow-routine (including snow redistribution), a glacier routine as well as a soil-moisture, a response and a routing routine.

With air temperaure above a certain threshold, precipitation enters the soil moisture routine where water can either evaporate or be stored. Out of the soil moisture storage water is subsequently transferred to the response routine, which can consist of two or three storages with their respective recession coefficients. Before water is released as streamflow, a triangular weighting function in the routing routine controls streamflow timing.

With air temperature below a certain threshold, precipitation is assumed to be solid and therefore enters the snow routine. In the snow routine snow water equivalent (SWE*; the addition of a * to a symbol or abbreviation subsequently indicates a model output component) as well as snow covered area (SCA*) are calculated. In non-glaciated elevation zones above 2500-2700 m. a.s.l. (catchment dependent)

snow is redistributed to lower elevation zones (above 1900 m a.s.l.) if a threshold of 500 mm in snow water equivalent is exceeded, in order to avoid the accumulation of so called snow towers (Freudiger et al., 2017). Snowmelt is calculated using a degree-day method, based on a seasonally varying melt factor (Stahl et al., 2008). Snow melt is delayed with elevation as well as on north facing slopes. On south facing slopes, in contrast, snow melt is accelerated (Konz & Seibert, 2010). Meltwater as well as precipitation can be held and refrozen in the snowpack, to allow for a more realistic representation of snow melt processes. Water release from the snowpack is simulated, if the maximum water holding capacity is reached (Seibert & Vis, 2012). The melt water from the snowpack enters the soil moisture routine and subsequently follows the same path through the model routines as non-solid precipitation (Stahl et al., 2017)

On a yearly basis, a small fraction of the accumulated snow is transferred into glacial ice. The glacier routine, calculating glacier mass balance, is based on the Δh parametrisation. The Δh parametrisation is an empirical approach designed to simulate glacier retreat, without needing computationally expensive models e.g. based on calculating energy balance fluxes (Huss et al., 2010). Glacier volume is calculated from an empirical relationship with glacier area. This empirical relation curve is available separately for small, medium and large glaciers. The method by Huss et al. (2010) was enhanced by Seibert et al. (2018) to include the simulation of phases with positive glacier mass balance, as for example occurred in the Swiss Alps during the 1970s. To achieve this enhancement, a lookup table is incorporated in the HBV-light model structure, in which the relation between glacier area and volume is stored for areas/volumes larger than the initial glacier conditions. This approach allows the glacier to advance as far as the initial glacier extent stored in the lookup table (Seibert et al., 2018).

Glacier ice melt, in the HBV model version used, is defined purely as ice melt, meaning firn and snow overlaying the glacier is not contributing to runoff from this subroutine (Stahl et al., 2017). Glacier ice melt is computed according to the procedure used in the snow routine. However, melt only takes place, if the snowpack on top of the glacier is already melted (Seibert et al., 2018). Glacier runoff is computed with a coefficient varying over the melt season to allow for an increasing outflow with the development of a pronounced drainage system, that develops

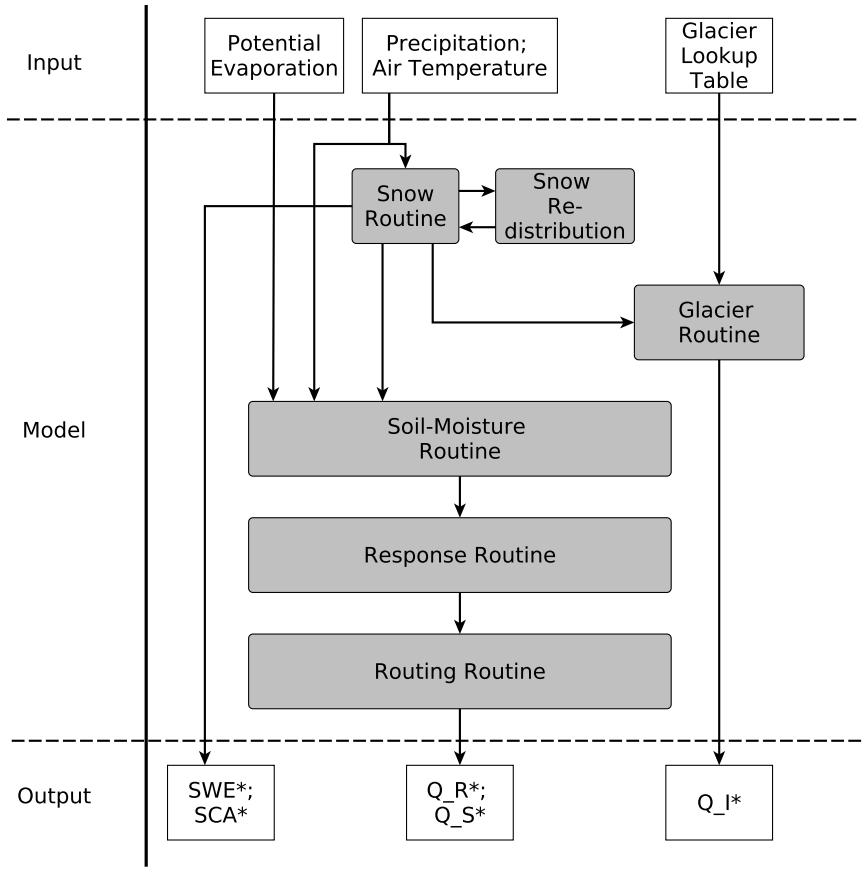


Figure 1: Structure of the HBV-light model. The dashed lines indicate model boundaries. Model subroutines are indicated in grey. The output components Q_R*, Q_S* and Q_I* denote runoff fraction from rainfall, snow melt and glacier melt respectively and in combination comprise total runoff.

within the glacier towards the end of the melt season (Stahl et al., 2008). The modelled discharge, driven by glacier ice melt (Q_I^*), is passed directly to streamflow, without entering any subsequent model subroutines. For a comprehensive description of the glacier subroutine see Seibert et al. (2018).

To allow for a tracking of the streamflow response of the modelled discharge components, driven by rainfall (Q_R^*) and snow melt (Q_S^*), the soil-moisture, response and routing routines are extended with mixing tanks. These mixing tanks are limited to 10 mm in storage in order to enable a tracking of the streamflow response of Q_R^* and Q_S^* on a daily basis. If unlimited, mixing tanks only show long term averages

of the respective streamflow fractions (Stahl et al., 2017). With the exception of refreezing rain in the snowpack, which is treated as Q_S^*, each discharge component is calculated on the basis of their respective origin. If e.g. snow melts on top of a glaciated area, the runoff is still assumed to be part of Q_S^* (Stahl et al., 2017). Table 1 shows the parameters of the HBV model, relevant for calibration in this work.

Table 1: HBV-light parameters which are calibrated in the work at hand. Parameters are ordered by model subroutine.

Parameter	Unit	Description
Snow Routine		
T_t	°C	Threshold temperature
CFMAX	mm °C$^{-1}\Delta t^{-1}$	Degree Δt factor
SFCF	–	Snowfall correction factor
CFR	–	Refreezing coefficient
CWH	–	Water holding capacity
Glacier Routine		
CFGlacier	–	Glacier correction factor
CFSlope	–	Slope correction factor
KGmin	t^{-1}	Minimum outflow coefficient
dKG	t^{-1}	Maximum minus minimum outflow coefficient
AG	mm^{-1}	Calibration parameter
KSI	–	Transition from snow in to glacier ice
Soil Moisture Routine		
FC	mm	Maximum soil moisture storage
LP	mm	Soil moisture threshold above which AET = PET
BETA	–	Shape parameter
Response Routine		
PERC	mmΔt^{-1}	Percolation threshold
Alpha	–	Non-linearity coefficient
K_1	Δt^{-1}	Recession coefficient 1
K_2	Δt^{-1}	Recession coefficient 2
Routing Routine		
MAXBAS	Δt	Length of triangular weighting function

3.2 Catchments and Data

For further analysis six meso-scale (21 – 165 km^2) catchments are used. All of these catchments are located in the Swiss Alps and are part of the Rhine watershed (see figure 2). The mountainous headwater catchments are all partly glacierised, with glacier coverage in 1973 ranging from 6% to 34%. The catchments do not contain large reservoirs or water transfers and thus can be regarded as fairly unregulated. An overview of catchment characteristics is given in table 2. For a more detailed catchment description see Stahl et al. (2017).

For all six catchments daily mean discharge data is available from 1973 to 2006, provided by the Swiss Federal Office for the Environment (FOEN). Additionally hourly discharge data for the period from 1974 to 2006 is available – also provided

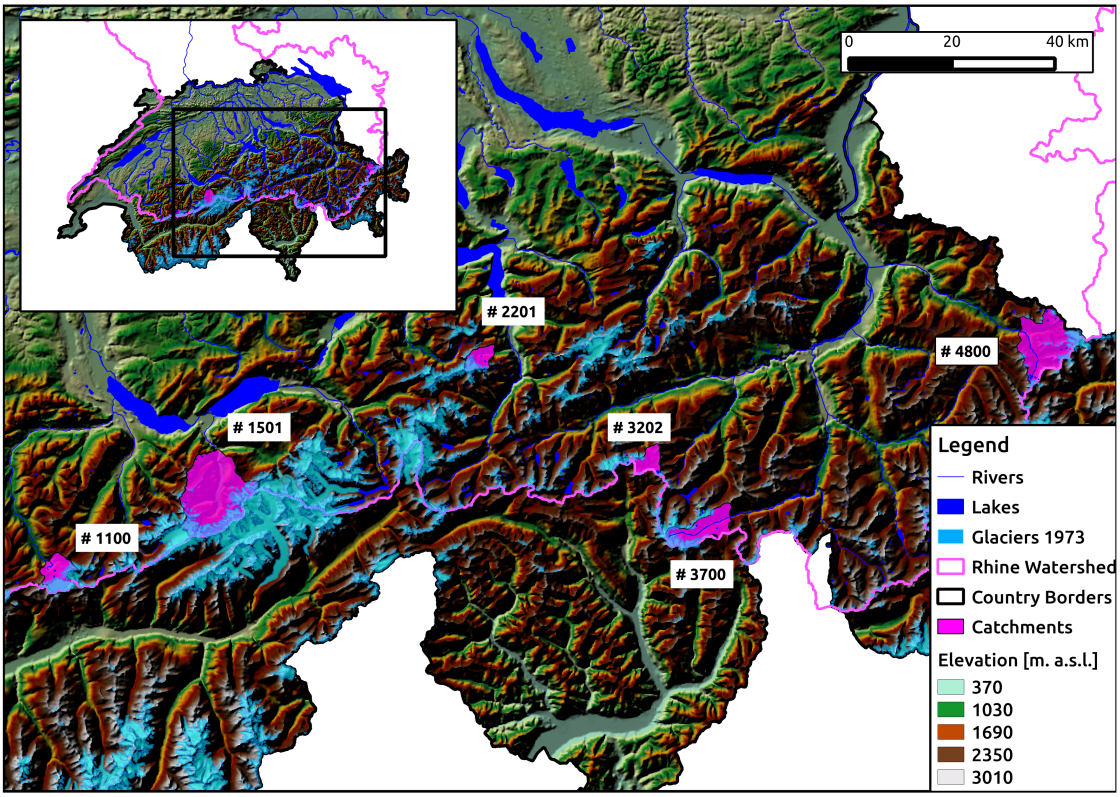


Figure 2: Location of the headwater catchments in the Swiss Alps. The catchment labelling corresponds to table 2. Water bodies are only displayed within the Rhine watershed. The maximum elevation is due to the spatial resolution (200 by 200 meters) of the digital elevation grid used.

Table 2: Catchment characteristics adapted from (Stahl et al., 2017). Ordered by catchment number #. "G" refers to glaciarised area.

#	Name	Gauge location	Elevation [m. a.s.l.]			Area [km^2]	G [%]	
			Min	Mean	Max		1973	2003
1100	Simme	Oberried/Lenk	1096	2336	3243	34.7	33.8	25.2
1501	Weisse Luetschine	Zweiluetschinen	650	2149	4146	164.9	18.5	16.0
2201	Alpbach	Erstfeld, Bodenberg	1022	2194	3192	20.7	26.9	22.1
3202	Somvixer Rhein	Somvix, Encardens	1489	2448	3159	21.8	6.1	3.2
3700	Hinterrhein	Hinterrhein	1587	2357	3387	53.9	17.8	8.8
4800	Landquart	Klosters, Auelti	1321	2339	3296	103.4	7.7	4.7

by the Swiss FOEN – which is the main reason, these particular catchments are chosen for further analysis.

Potential evaporation data is used, as calculated by Stahl et al. (2017) following the procedure of Oudin et al. (2005). As further meteorological input (precipitation; air temperature) the gridded HYRAS dataset provided by the German Weather Service (DWD), which covers the period from 1951 to 2006, is used (Frick et al., 2014; Rauthe et al., 2013). Stahl et al. (2017) found a systematic underestimation of precipitation sums in the HYRAS dataset and corrected the data accordingly.

Data on glacier extent are from different sources for the years 1973 and 2003. Data for September 1973 is derived from aerial photography (Maisch, 2000; Müller et al., 1976). Glacier extent in autumn 2003 is obtained from Landsat satellite imagery (Paul et al., 2011).

Maps of SWE and SCA are available from the Swiss Institute for Snow and Avalanche Research, a subunit of the Swiss Federal Institute for Forest, Snow and Landscape Research. The datasets cover the period from 1971 to 2012, comprising the months from November to May each winter. Data acquisition and processing is described by Jörg-Hess et al. (2014) and Jonas et al. (2009). Due to the highest density of observation stations in an elevation range between 2000 and 2500 m. a.s.l. and thus the highest accuracy in this particular elevation range, data of SWE and SCA is only used for this elevation band (Stahl et al., 2017).

3.3 Artificial Discharge Time Series

For the calibration of runoff from the snow and glacier subroutines of the HBV-light model, suitable discharge data is required. As a representation of runoff driven by snow melt, the SWE time series (see section 3.2) is used. The decrease in SWE (ΔSWE) for each day is computed to obtain an estimate of runoff originating from snow melt. To account for delay in snow melt runoff, the computed ΔSWE values are weighted on the subsequent days, using an Exponential Moving Average weighting function (Ulrich, 2013). This procedure results in a time series with daily artificial snow melt runoff (Q_S) between November and May for the period from 1973 to 2006.

The estimation of runoff driven by glacier ice melt is based on mean hourly discharge time series. With observable diurnal discharge amplitudes originating from diurnal variations in energy available for melt processes, these amplitudes can be linked to runoff driven by either snow or glacier ice melt (Seibert et al., 2015; Hock, 2005; Jansson et al., 2003). To obtain a time series representing glacier ice melt, the daily discharge minimum is computed. Subsequently the obtained minima are linearly interpolated. If the linear interpolation exceeds the observed streamflow (e.g. on declining limbs), the interpolation is equalled with the observed data following the procedure of Gustard and Demuth (2008) for baseflow separation. The area between the interpolated minima and the hourly runoff is numerically integrated in order to get daily amounts of runoff, that are driven by melt processes. To minimise the influence of the spring snow melt, values before Day of Year (DoY) 200 (19th of July in a non leap year) are discarded. DoY 200 is chosen in accordance with the findings of He et al. (2018) who found an increasing influence of glacier melt in late July for their specific catchment. In addition, high discharge values are removed from the time series since they are likely driven by large precipitation events rather than glacier ice melt (Seibert et al., 2015). Integrated amplitudes above a threshold of 20 mm are discarded as such high melt rates seem unrealistic (Hock, 2005). The resulting time series provides daily values of artificial glacier ice melt runoff (Q_I).

3.4 Objective Functions

For the optimisation of the HBV-light model a series of objective functions are employed. As a GOF measure for the different discharge components of total runoff (Q_{tot}), Q_S and Q_I, FDC based likelihood measures are used. Following the procedure reported in Sadegh et al. (2016), the FDC is calculated using the Weibull plotting position (Cunnane, 1978) with the following form:

$$e_i = \frac{1}{n}\left(R_i - \frac{1}{2}\right) \tag{1}$$

with e being the exceedance probability and R the rank of an observation i. Using equation 1, the FDCs are calculated for the discharge time series of Q_{tot}, Q_S and Q_I. For calculating a likelihood of modelled discharge, given the observed FDC, a derivable parametric expression of the observed FDC is required. Sadegh et al. (2016); Vrugt and Sadegh (2013) introduced a three parameter Van Genuchten (VG) model based on the water retention function by Van Genuchten (1980) to represent an observed FDC. The formulation in e-space is given as follows:

$$e_i^* = \left[1 + (a_{VG} \cdot y_i)^{b_{VG}}\right]^{-c_{VG}} \tag{2}$$

with calibration coefficients a_{VG} [d/mm], b_{VG} [-] and c_{VG} [-]. y denotes a discharge measurement and e_i^* the modelled exceedance probability (in contrast to the observed exceedance probability e_i).
Parameter calibration in e-space is not meaningful since the interval [0,1] is always matched by equation 2. Thus equation 2 has to be inverted for y:

$$y_i^* = \frac{1}{a_{VG}}\left[{e_i}^{(-1/c_{VG})} - 1\right]^{(1/b_{VG})} \tag{3}$$

with y_i^* being the calculated discharge. Given equation 3, the parameters a_{VG}, b_{VG} and c_{VG} are calibrated against the FDC of Q_{tot}, Q_S and Q_I, respectively. A SCE scheme is used as implemented in R (R Core Team, 2017) by Andrews (2012), minimising RMSE as GOF measure. A resulting single "best" set of parameters

is required for using a parametric expression of the FDC as GOF measure. Using a non probabilistic optimisation scheme such as the SCE achieves distinct values for the parameters a_{VG}, b_{VG} and c_{VG}. For further notations, the fitted parameters a_{VG}, b_{VG} and c_{VG} are referred to as θ.
For calculating the likelihood of a datapoint or dataset given a model – $l(data|model)$ – a probability density function (PDF) is required. With the FDC being a cumulative density function (CDF) the corresponding PDF is given as the first derivative of the CDF. For the e-space formulation of the VG model (equation 2) the first derivative with respect to y is given as follows:

$$e_i^{*'} = \left[b_{VG} \cdot c_{VG}(a_{VG} \cdot y_i)^{b_{VG}}((a_{VG} \cdot y_i)^{b_{VG}} + 1)^{(-c_{VG}-1)}\right] \cdot y_i^{-1} \tag{4}$$

Replacing y_i with modelled runoff Q_i one can calculate the likelihood of the individual data points $l(Q_i|e_i^{*'}[\theta])$. To obtain the log-likelihood (LL) for the entire modelled time series of a runoff component the sum of the logarithmic likelihoods is computed:

$$LL = \sum_{i=1}^{n} ln(Q_i|e_i^{*'}[\theta]) \tag{5}$$

with ln being the natural logarithm. Equation 5 now serves as an objective function for the modelled time series of total discharge and discharge driven by snow melt and ice melt (Q_{tot}^*, Q_S^* and Q_I^*), given the individually fitted parameter values θ for the discharge component specific FDC derived from optimising equation 3. GOF of Q_{tot}^* is calculated for the entire year. Due to data availability and structure, Q_S^* and Q_I^* are only calibrated during the period from November to May and DoY 200 to 365 respectively.
As an additional GOF measure for the calibration of Q_I^*, a minimum rejection sampling scheme can be incorporated in the Q_I^* objective function. This is achieved by rejecting model runs where $\sum_{i=1}^{n} Q_I^* < \sum_{i=1}^{n} Q_I$. This addition takes into account the limitation of the Q_I time series possibly underestimating the discharge from glacier ice melt (cf. Jansson et al., 2003). This rejection sampling is based in hydrological theory as compared to the ABC rejection sampling approach, where

rejecting parameter sets is based on statistical thresholds (Sadegh & Vrugt, 2014; Vrugt & Sadegh, 2013).

By optimising discharge to a FDC, information about the discharge timing is not taken into account. To not only calibrate against catchment behaviour in form of the FDC, an objective function for calibrating against peak flow timing is developed. To only calibrate on clearly identifiable peaks, a runoff measurement has to fulfil a series of conditions. The measurement has to be the highest runoff value in a 14 day interval, be higher or equal to the 25% quantile (Q25; exceedance probability) and the runoff has to decrease on 4 consecutive days after a peak. With peaks fulfilling these conditions, the Cross Correlation Function (CCF; Pearson r) is computed for the lags from $t-7$ to $t+7$. The lag with the highest correlation is assumed to be the shift in peak flow timing. The developed approach to account for peak flow timing is an adaption of a procedure introduced by Yilmaz et al. (2008)

As performance criterion to close the water balance during calibration, the bias in runoff ratio (biasRR) is used according to Yilmaz et al. (2008):

$$biasRR = \frac{\sum_{i=1}^{n}(Q^*_{tot_i} - Q_{tot_i})}{\sum_{i=1}^{n} Q_{tot_i}} \tag{6}$$

For SWE*, SCA* and Glacier mass balance (MB^*_G) the performance criteria are adapted from Stahl et al. (2017). SWE* is calibrated to SWE by using the mean absolute error (MAE) as objective function. For calibrating SCA*, the RMSE is utilised. Changes in MB^*_G are calibrated using an objective function, specifically designed by Stahl et al. (2017).

To achieve parameter sets near the pareto-optimal objective function value, given the GOF criteria mentioned above, normalisation of the objective functions is required for comparability. The following normalisation formula is used:

$$OF^* = \frac{N_{max} - N_{min}}{OF_{max} - OF_{min}} \cdot \Big(OF - OF_{min}\Big) + N_{min} \tag{7}$$

where OF^* and OF are scaled and unscaled objective function values respectively.

N_{min} and N_{max} are the minimum and maximum of the normalisation range and OF_{min} and OF_{max} the minimum and maximum of possible unscaled objective function values. Determining OF_{min} poses a problem especially for likelihood based objective function values, since these can theoretically be as low as -∞. This problem is bypassed by determining OF_{min} values for likelihood based objective functions from sensitivity analysis runs (described in section 3.5). Furthermore, normalisation with equation 7 induces a loss of information for the timing objective function. With the applied normalisation method, a distinction between positive and negative lags is not possible.
For model calibration the normalised objective functions are weighted as follows:

$$R_D = 0.15 \cdot R_{Q_{tot}} + 0.1 \cdot R_{Q_I} + 0.1 \cdot R_{Q_S} + 0.1 \cdot biasRR + 0.15 \cdot R_{timing} + 0.1 \cdot R_{SWE} + 0.1 \cdot R_{SCA} + 0.2 \cdot R_G \quad (8)$$

with the weighted combined R_D. With $R_{Q_{tot}}$, R_{Q_I} and R_{Q_S} being the FDC based GOF measures for Q_{tot}, Q_I and Q_S, respectively. R_{timing} is the bias in peak flow timing. R_{SWE} and R_{SCA} are the RMSE and MAE of modelled and observed SWE and SCA respectively. R_G is the objective function for calibrating MB^*_G.
A second optimisation scheme following the procedure of Stahl et al. (2017) is employed. Three distance based metrics are used to calibrate Q^*_{tot}, namely the Lindström measure (a variation of the NSE (Lindström et al., 1997)), the NSE on logarithmic discharge values as well as the NSE for the season from June to September. The optimisation of SWE*, SCA* and MB^*_G are carried out as described above. The model outputs Q^*_S and Q^*_I are not calibrated against time series data. The weighting of the objective functions is carried out as follows:

$$R_S = 0.2 \cdot R_{Q1} + 0.15 \cdot R_{Q2} + 0.15 \cdot R_{Q3} + 0.15 \cdot R_{SWE} + 0.1 \cdot R_{SCA} + 0.25 \cdot R_G \quad (9)$$

with R_S being the weighted objective function. R_{Q1} is the Lindström measure, R_{Q2} the NSE of the logarithmic flows and R_{Q3} the NSE of the flows from June to September. For a comprehensive description of the calibration approach summarised in equation 9 see Stahl et al. (2017).

3.5 Sensitivity Analysis

Sensitivity of the objective functions described above to parameter perturbations of the HBV-light model parameters is tested with a one at a time parameter alteration. Due to the number of parameters, a sensitivity analysis, which takes parameter interactions into account is not feasible.
During the sensitivity analysis one parameter is varied over the calibration range (see table 3) in 100 equidistant steps. The remaining parameters are kept at their respective default value (see table 3). To asses the meaningfulness of the objective functions used, the effect of varying each parameter on each objective function is evaluated. This assessment allows to test, whether an objective function targets the model subroutine it is designed for.
The sensitivity analysis is carried out with non normalised objective functions via visual inspection. This way the sensitivity analysis additionally allows to obtain theoretical maxima and minima of the likelihood based objective functions. The theoretical minima of the objective functions are used for normalisation, which is required for weighting.

3.6 Model Calibration

The calibration of the HBV-light parameters displayed in table 1 and 3 is carried out individually for the six catchments described in section 3.2. The years from 1973 to 1976 are used for model spin up. Parameter optimisation is carried out from 1976 to 2006. The Spin up and calibration periods are chosen mainly on data availability. Starting spin up in 1973 provides the advantage to be able to predefine initial glacier volume based on data. Furthermore, the HYRAS dataset is available until 2006. Thus, for the period from 1973 to 2006 data availability of all needed datasets is given. Model calibration is carried out separately for the optimisation framework developed in the work at hand (equation 8) and the framework following Stahl et al. (2017) (equation 9).
The DREAM$_{ZS}$ algorithm, as implemented by Hartig et al. (2017), is used for model optimisation. Uniform prior distributions within the predefined parameter calibration ranges (see table 3) are employed. The optimisation framework for the individual catchments consists of three Markov-Chains at 10000 function evalua-

Table 3: Valid ranges, default values and calibration ranges of the HBV-light parameters optimised in the work at hand. Parameters are ordered by model subroutine. Calibration ranges are adapted from Stahl et al. (2017).

Parameter	Valid Range	Default Value	Calibration Range
Snow Routine			
T_t	$-\infty; \infty$	0	-2; 3
CFMAX	$0; \infty$	3	4; 8
SFCF	$0; \infty$	1	0.8; 1.05
CFR	$0; \infty$	1	0.8; 1.05
CWH	$0; \infty$	0.05	0; 0.2
Glacier Routine			
CFGlacier	$0; \infty$	1	1; 2.5
CFSlope	$0; \infty$	1	1.2; 2.2
KGmin	0; 1	0.01	0.005; 0.2
dKG	0; 1	0.01	0.005; 0.2
AG	0; 1	0.1	0; 0.2
KSI	0; 1	0.002	0.001; 0.003
Soil Moisture Routine			
FC	$0; \infty$	200	50; 550
LP	0; 1	1	0.3; 1
BETA	$0; \infty$	1	1; 6
Response Routine			
PERC	$0; \infty$	1	0; 6
Alpha	$0; \infty$	0	0; 0.5
K_1	0; 1	0.1	0.01; 0.2
K_2	0; 1	0.05	0.00005; 0.15
Routing Routine			
MAXBAS	1; 100	1	1; 7

tions each. 10% of the function evaluations are discarded as burn-in of the Markov-Chains. Initial parameter values are randomly drawn from the prior distributions. During optimisation all parameters are updated jointly. Cross over probabilities of the parameter values are adapted during optimisation. This adaption favours large parameter jumps at the beginning of each chain, when the posterior distribution is close to the uniform prior distribution. With the evolution of each Markov-Chain, the probability of large jumps in the parameter space decreases but does not reach zero. The adaptive approach decreases the probability of the algorithm to get caught in a local optimum. This optimisation framework results in posterior parameter densities based on 27000 function evaluations for each catchment.

Compared to the original version of the DREAM$_{ZS}$ algorithm by Vrugt (2016); Laloy and Vrugt (2012), the implementation by Hartig et al. (2017) does not incorporate an automated breaking criterion based on convergence of the Markov Chains. Therefore the used implementation provides the advantage of not having to define a convergence criterion a priori, which makes it easier to estimate computational time needed. Furthermore, the posterior parameter densities are all based on the same sample size in order to increase comparability. Convergence of the Markov-Chains is checked after the optimisation is finished by using the method of Gelman and Rubin (1992b, 1992a).
For each catchment three individual optimisation runs are carried out. The weighted objective function R$_D$ (equation 8) is optimised with and without rejection sampling of runs where $\sum_{i=1}^{n} Q_I^* < \sum_{i=1}^{n} Q_I$. The weighted R$_S$ objective function (equation 9) is optimised without additional rejection sampling.

3.7 Model Evaluation

Due to computational limitations, an uncertainty analysis using GLUE is not feasible within the scope of this work. Neither is the implementation of ABC rejection sampling in the DREAM$_{ZS}$ algorithm. Instead, to keep computational time low, model uncertainty is evaluated by drawing 100 random parameter sets from the posterior parameter densities (derived from optimising R$_D$ and R_S in equation 8 and 9, respectively) for each catchment. The parameters are sampled from the region of highest posterior density $\pm$ 2.5% of the calibration range. In cases where this interval includes regions outside the calibration range, the range to draw parameters from is truncated to be strictly inside the calibration range. These 100 runs are subsequently referred to as MC runs. Technically speaking, these runs can not be classified as MC runs, however.
With the parameter sets from the MC runs, all objective functions (the weighted objective functions and the individual components thereof) are evaluated. This evaluation results in four sets of objective function values. Namely objective function values are derived for R$_D$ and R$_S$ with parameters from optimising R$_D$ and vice versa. Furthermore, two sets of modelled data with parameters from the standard (R$_S$; the term "standard" is used for the lack of more appropriate termi-

nology) and diagnostic (R_D) framework are obtained from the MC runs for each catchment. With these two data sets an uncertainty analysis is carried out by plotting the individual model output time series against the observed data.
The objective function values obtained from the MC runs are evaluated with a correlation analysis. On one hand, the aim of this analysis is to pinpoint individual objective functions, which can not be satisfied simultaneously, in order to detect malfunctioning model subroutines. On the other hand, this correlation analysis serves as a benchmark for the employed GOF measures for calibrating Q_I and Q_S. In addition, relations between objective function values and physiographic catchment characteristics are evaluated via correlation analysis.
Furthermore, the best model efficiencies from the calibration of R_S are compared to the efficiencies of model optimisation by Stahl et al. (2017) for the same catchments. Stahl et al. (2017) used a genetic algorithm (GAP) for the optimisation of R_S, as compared to the DREAM$_{ZS}$ algorithm used in the present work.

4 Results

4.1 Discharge Time Series and Flow Duration Curves

Figure 3 shows the daily median runoff regime of Q_{tot}, Q_S and Q_I for the calibration period from 1976 to 2006 for all six catchments. Q_{tot} shows a similar pattern for all catchments with low flows from late autumn to early spring. From late spring to early autumn, a significant increase in runoff can be observed. Q_S has a period with zero values from autumn to around DoY 100 (mid April). The increase of Q_S in spring shows a general agreement with the observed Q_{tot}. For the catchments #1100 and #1501, however, Q_S rises slightly earlier than Q_{tot}.
Q_I – like Q_S – shows no contribution to runoff during the winter months. From spring to autumn an increase of Q_I can be observed. With the exception of catchments # 1100 and # 3202, this increase occurs until around DoY 150. After DoY 150 a plateau can be observed before the discharge decreases at around DoY 240.

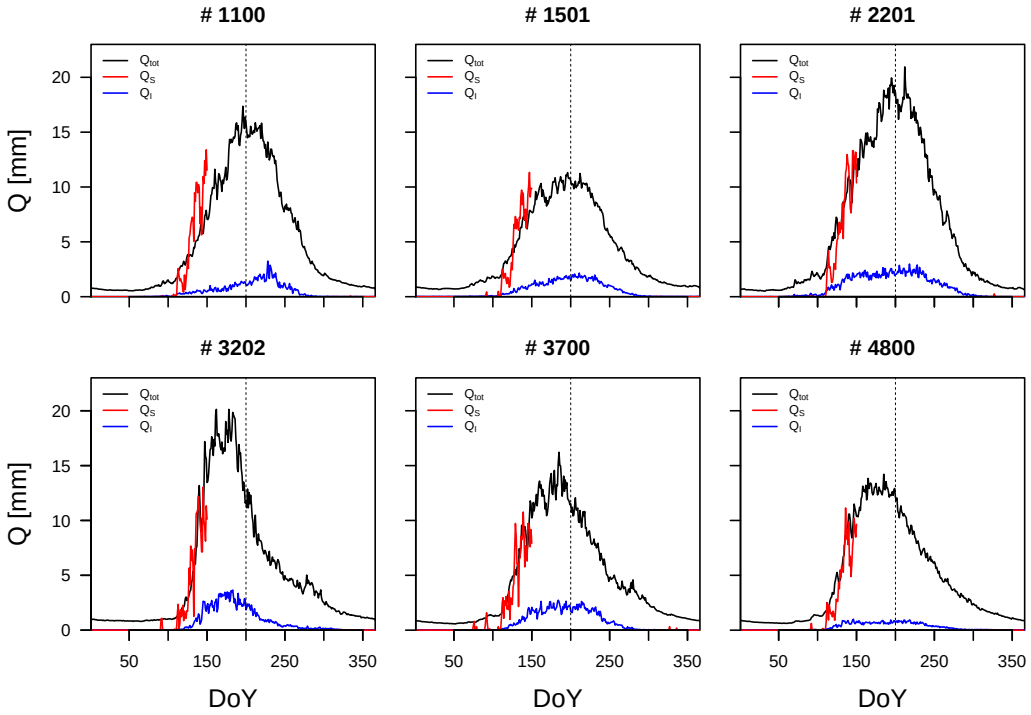


Figure 3: Longterm (30 years) daily median runoff values of Q_{tot},Q_S and Q_I for all catchments. The dashed line indicates DoY 200 wherefrom the blue time series is considered to represent Q_I. The Q_I values before DoY 200 are shown for completeness

The cutoff of the Q_I time series at DoY 200 (displayed in figure 3 as the dotted black line) is either at the peak of the discharge regime of Q_{tot} (catchments #1100, #1501 and #2201) or shortly thereafter (catchments #3202, #3700 and #4800). A similar pattern can be observed for Q_I. The only catchment where Q_I increases after DoY 200 is catchment #1100.

Figures 4, 5 and 6 show the FDCs of the time series of Q_{tot}, Q_S and Q_I, respectively. Furthermore, the fitted VG model for each FDC is displayed. The fitted VG models show a general trend of closely reproducing the low flow values ($e > 0.7$) of the FDC. Discharge values with $0.4 < e < 0.7$ are slightly overestimated by the VG model in most cases. In contrast, discharge values with $0.1 < e < 0.4$ are underestimated by the VG model. For the peak flow values the VG model shows an accurate fit of the observed discharge values in many cases. However, in some cases, the VG model overestimates the peak flows.

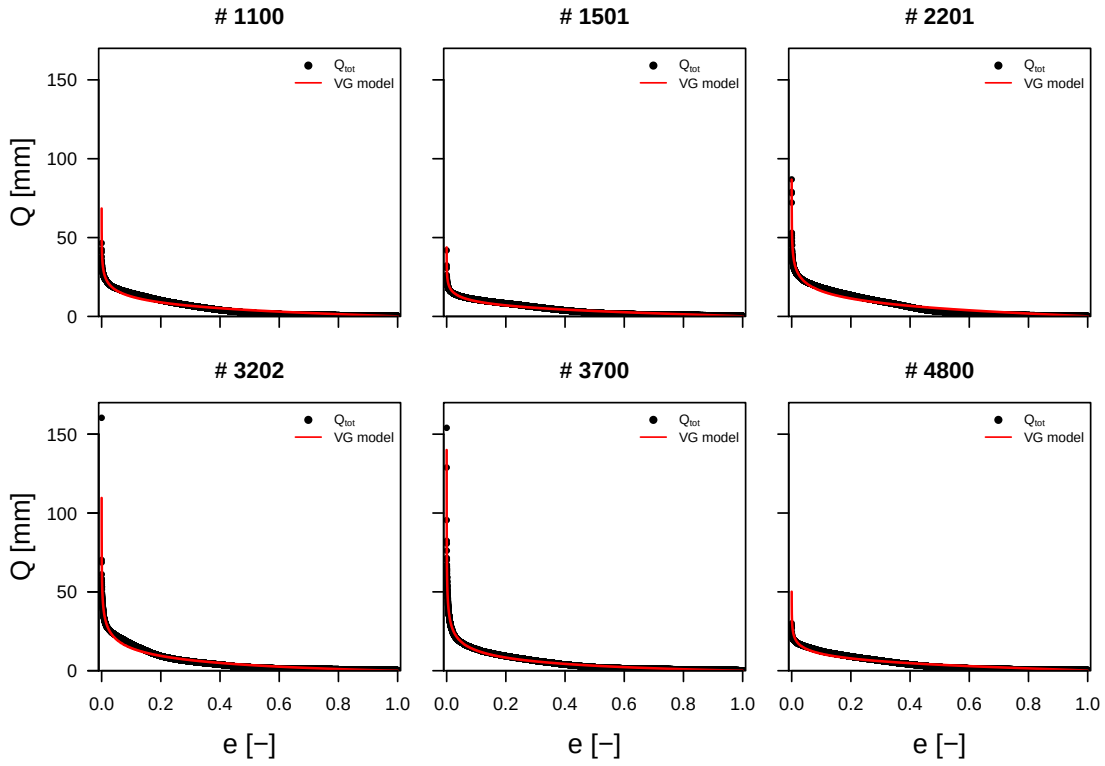


Figure 4: FDC of Q_{tot} for all six catchments (black dots). The fitted VG model for each FDC is displayed as the red line.

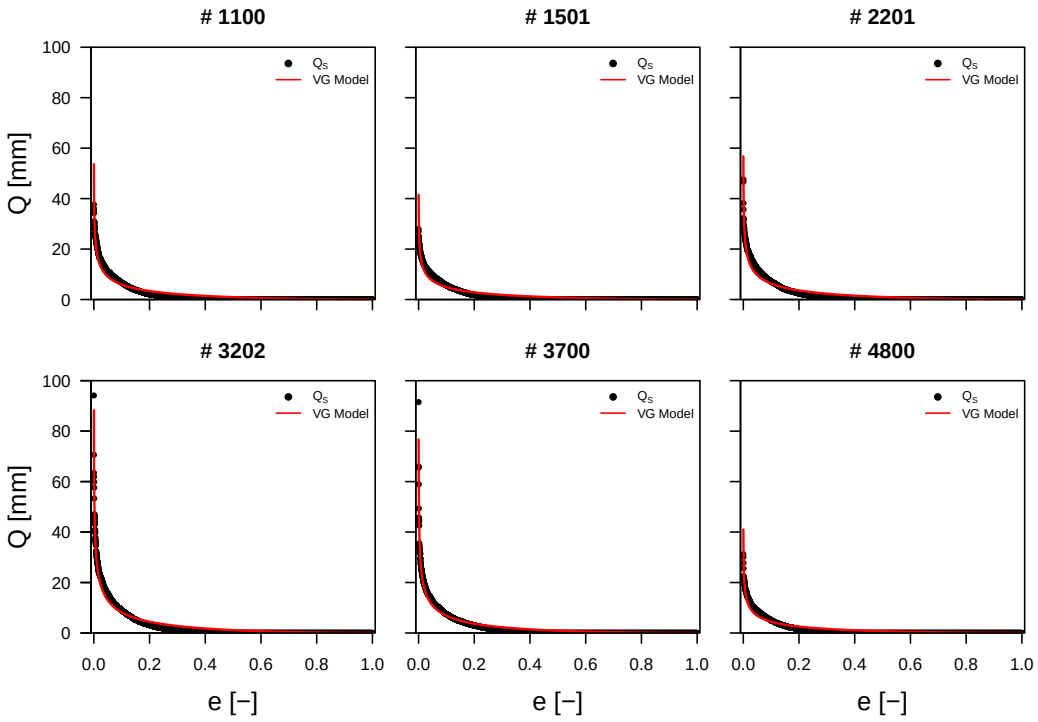


Figure 5: FDCs of Q_S for all six catchments (black dots). The fitted VG model for each FDC is displayed as the red line.

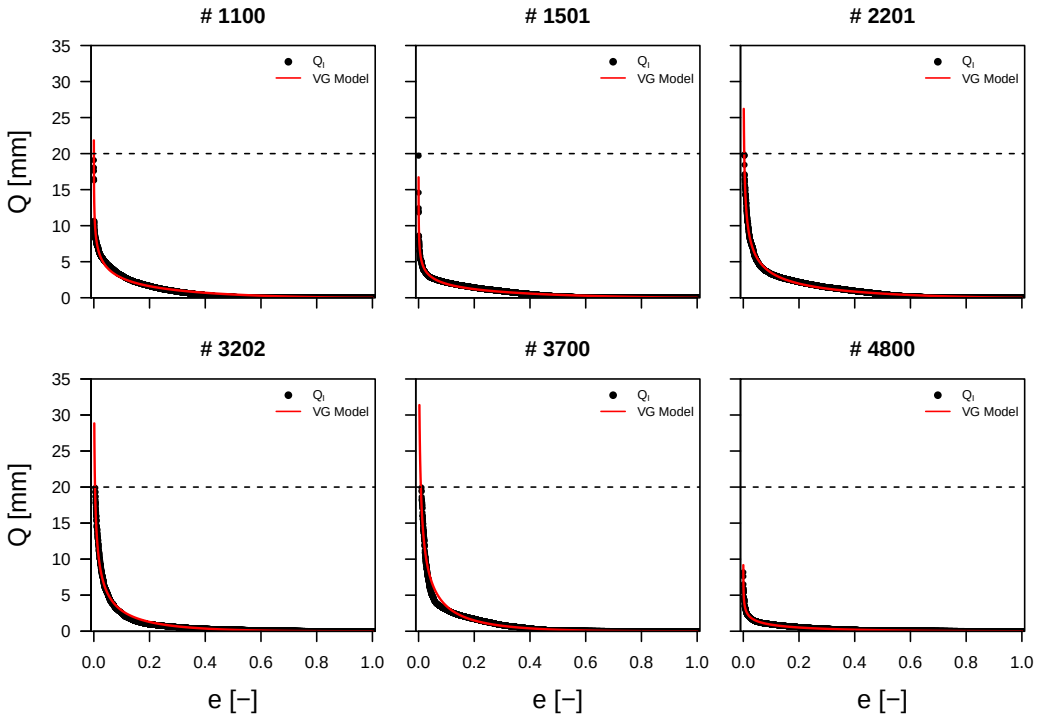


Figure 6: FDCs of Q_I for all six catchments (black dots). The fitted VG model for each FDC is displayed as the red line. The dashed black line indicates the cutoff value of 20 mm above which discharge values are discarded.

4.2 Sensitivity Analysis

Due to the number of catchments, parameters and objective functions, the visual results of the sensitivity analysis are shown in appendix B for catchment #1100 only. The sensitivity analysis in general yields similar results for all six catchments. Table 4 shows the objective functions analysed within the diagnostic and the standard framework and the parameters the respective objective function is sensitive to. Due to computational issues, parameters CWH and CFR could not be included in the sensitivity analysis. All objective functions except R_{timing} are sensitive to the parameter T_t. R_{Q_I} and R_{timing} are sensitive to most or all parameters of the targeted subroutine. For the other objective functions, such a clear conclusion can not be drawn.

In contrast, no objective function shows sensitivity towards the parameters AG, PERC, FC, LP and BETA – at least when parameter interactions are not taken into account. The diagnostic framework, in addition, is not targeted by the parameters Alpha and K2. The standard framework does not show sensitivity to the parameters KSI and CFGlacier.

Table 4: Objective functions and parameters the respective objective functions are sensitive to. For parameters in brackets the objective function is sensitive in at least four catchments. No brackets mean the objective function is sensitive to the respective parameter for all six catchments. Objective functions are ordered by calibration framework.

Objective function	Parameters
Diagnostic	
$R_{Q_{tot}}$	T_t; (K1); SFCF; KGmin
R_{Q_S}	KGmin; K1; T_t; (CFMAX)
R_{Q_I}	KSI; KGmin; dKG; T_t; CFMAX; CFGlacier; (CFSlope)
biasRR	T_t; (CFMAX); SFCF
R_{timing}	MAXBAS, K1
Diagnostic/Standard	
R_G	T_t; CFMAX; SFCF; CFGlacier; CFSlope
R_{SWE}	T_t; CFSlope
R_{SCA}	(T_t)
Standard	
R_1	(KGmin), (dKG), Alpha, K1, TT, CFMAX, (MAXBAS)
R_2	KGmin, dKG, K1, K2, TT, (CFMAX)
R_3	(KGmin), Alpha, K1, MAXBAS, TT, CFMAX, (SFCF), (CFSlope)
None	AG, PERC, FC, LP, BETA

4.3 Modelling Results and Evaluation

The posterior parameter distributions for catchment # 1100 are shown in figure 7 (standard framework) and figure 8 (diagnostic framework). The posterior densities of the other catchments are not shown (available upon request). The three Markov Chains resulting in these posterior densities do not converge for all parameters regardless of the catchment and optimisation framework. Figures 7 and 8 show unimodal posterior densities for some of the HBV-light parameters. Other parameters show bi- or multimodal posterior densities which in some cases are similar to the uniform prior distribution.

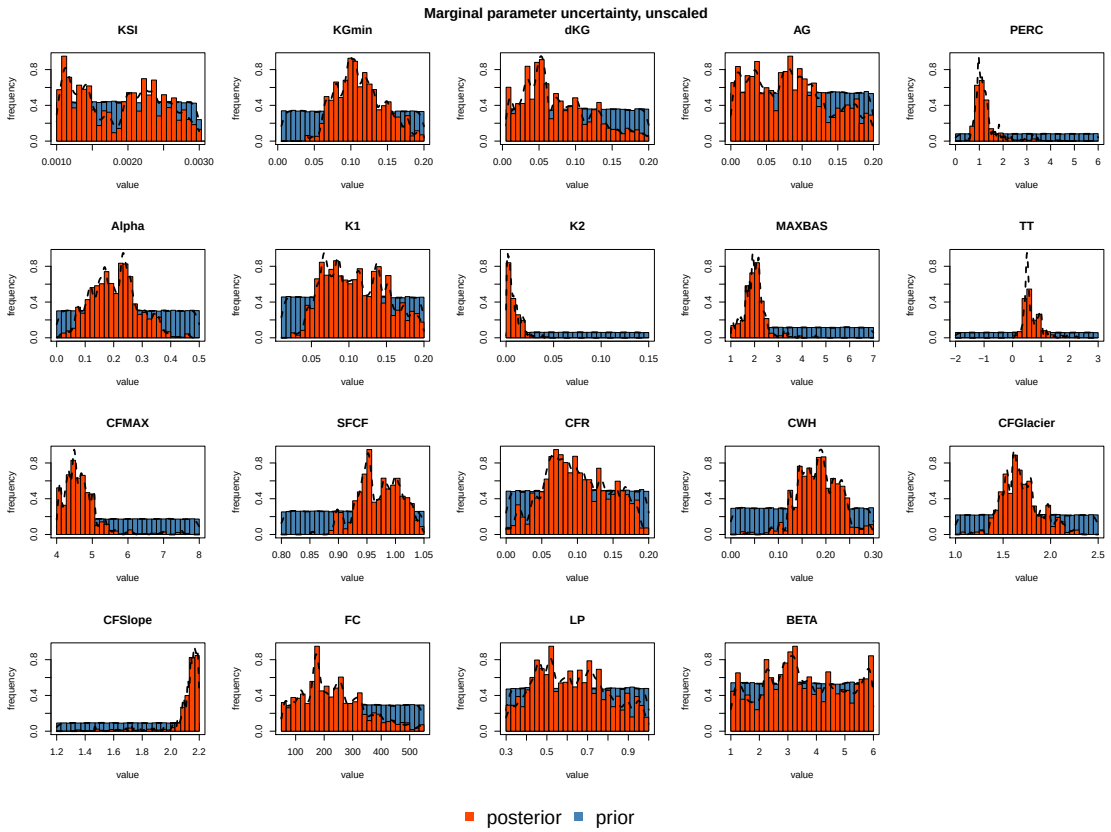


Figure 7: Posterior parameter distributions of catchment # 1100 for the standard calibration framework. The term marginal refers to the fact, that each parameter is only displayed in one dimensional parameter space (in contrast to the 19 dimensional parameter space of the optimisation problem).

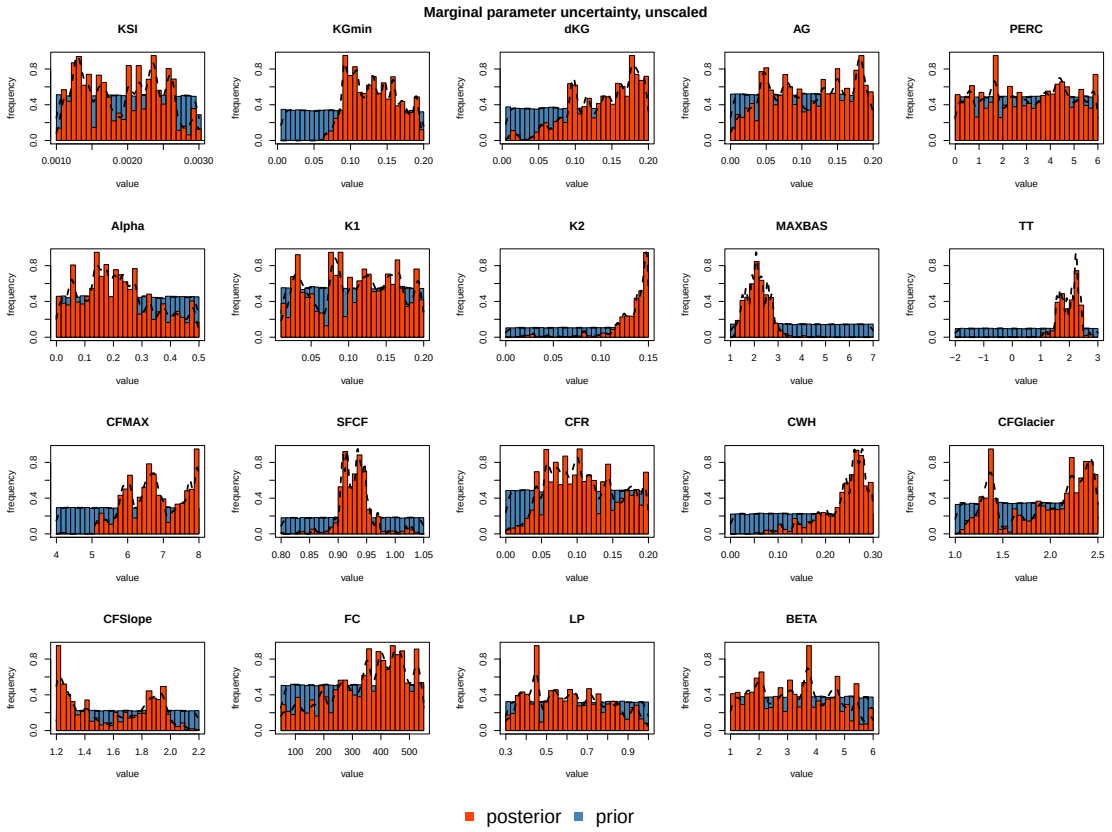


Figure 8: Posterior parameter distributions of catchment # 1100 for the diagnostic calibration framework. The term marginal refers to the fact, that each parameter is only displayed in one dimensional parameter space (in contrast to the 19 dimensional parameter space of the optimisation problem).

Figure 9 shows boxplots of the weighted model efficiencies obtained from 100 MC runs. Due to similar results for all catchments, combined boxplots are displayed. HBV-light evaluations with parameters from the standard optimisation scheme result in high model efficiencies for the weighted R_D objective function of the diagnostic framework. The median model efficiency with standard parameter sets is higher than the efficiency with parameters from the diagnostic framework for R_D. The ranges for R_D are similar for parameter sets from both optimisation frameworks.

Model efficiencies of the R_S objective function, using parameters from the diagnostic framework, are substantially lower than with parameter sets from the standard framework. Parameter sets from the diagnostic framework result in especially low values of the residual based distance metrics of R_{Q1}, R_{Q2} and R_{Q3} (results not

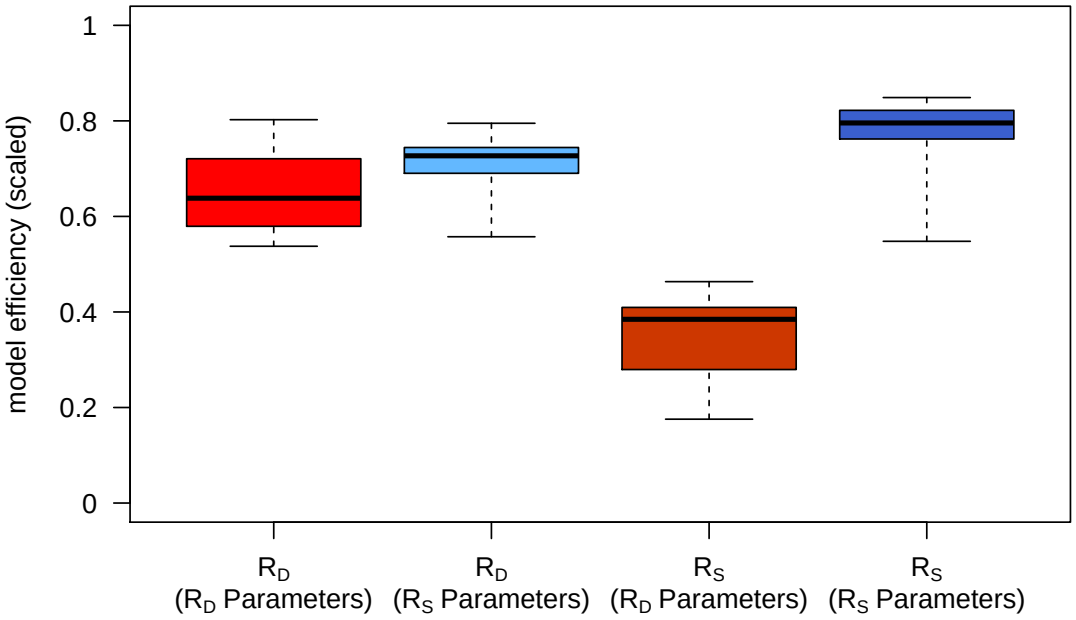


Figure 9: Box- and Whiskers plots of the weighted model efficiencies combined for all six catchments. Red colors indicate HBV-light evaluations with parameters from the diagnostic framework, while blue colors show evaluations with parameters from the standard framework. The bold black line indicates the median. The box includes the interquartile range and the whiskers include all objective function values received from the MC runs. An objective function value of 1 indicates a perfect fit.

shown), which results in low overall model efficiencies. Table 5 shows overall model efficiency of R_S from the optimisation with DREAM$_{ZS}$ compared to the results of Stahl et al. (2017). Very similar or slightly higher overall model efficiencies of the standard framework can be observed, when using the DREAM$_{ZS}$ algorithm compared to the GAP algorithm used by Stahl et al. (2017).

Table 5: Best overall model efficiencies from optimising R_S with the DREAM$_{ZS}$ algorithm compared to efficiencies from Stahl et al. (2017) using the GAP algorithm.

#	DREAM$_{ZS}$	GAP
1100	0.854	0.849
1501	0.847	0.858
2201	0.844	0.847
3202	0.853	0.789
3700	0.838	0.822
4800	0.863	0.829

Figure 10 shows the observed and modelled FDCs of Q_{tot} with their respective 95% confidence interval. Over a wide range of runoff values, the model runs from the standard framework show a close accordance with the observed data. However, the highest runoff peaks are systematically underestimated by model runs from the standard framework. In the model runs from the diagnostic framework, in contrast, a bias towards peak flow can be observed. Discharge values with low exceedance probabilities (but not the highest peak flows) are generally slightly overestimated by model runs from the diagnostic framework.

The modelled and observed FDCs of Q_S for the period from November to May are displayed in figure 11. The standard and the diagnostic framework show similar results for the modelled Q_S. The diagnostic framework shows a bias towards the highest runoff values for some catchments, however. While the observed values

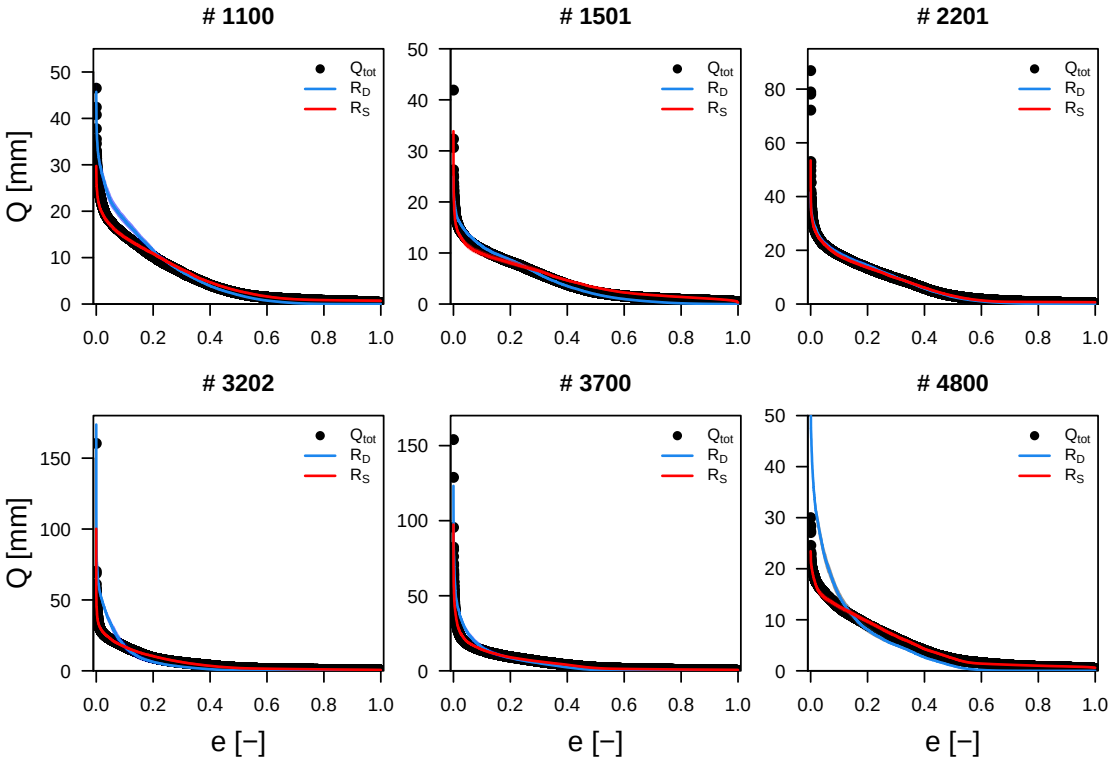


Figure 10: FDCs of Q_{tot} for all six catchments (black dots). The median of the modelled FDC from the R_D and R_S framework are displayed as the blue and red lines, respectively. The blue and red shaded areas are the 95% confidence interval obtained from the 100 MC runs.

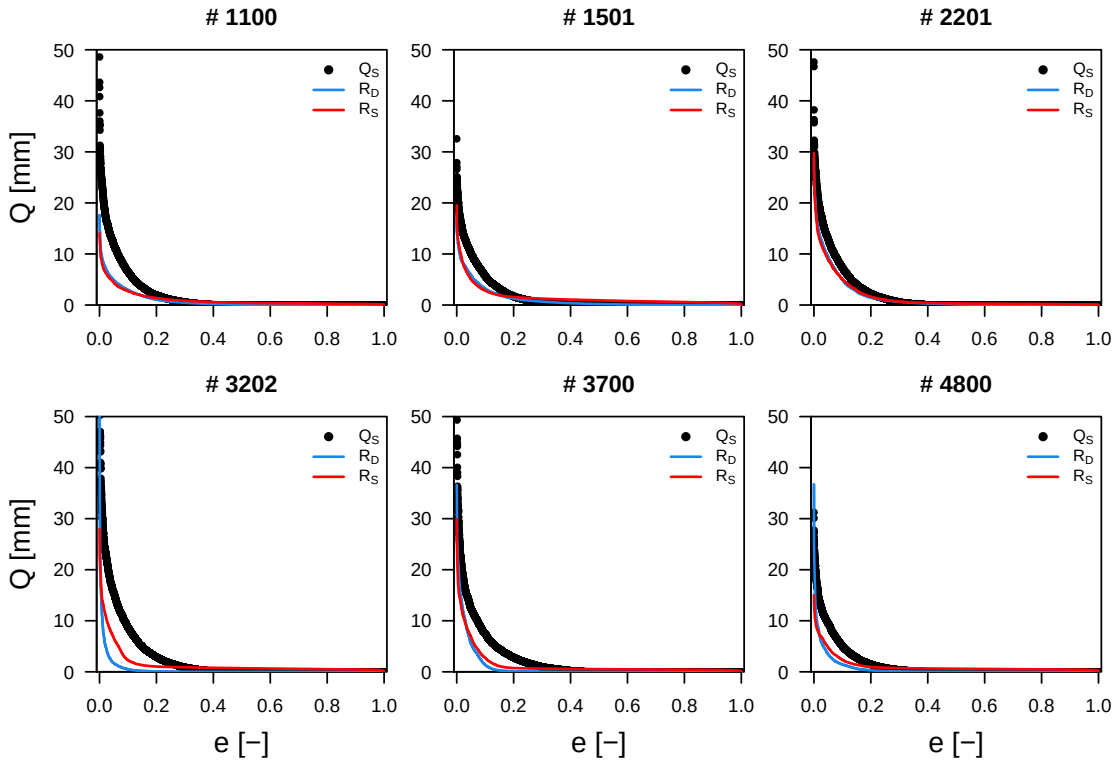


Figure 11: FDCs of Q_S for all six catchments (black dots). The median of the modelled FDC from the R_D and R_S framework are displayed in blue and red, respectively. The blue and red shaded areas are the 95% confidence interval obtained from the 100 MC runs. Only values for the period from November to May are shown.

mostly show zero runoff for exceedance probabilities above 0.4, the modelled runoff especially for the standard framework, in some cases, shows non zero runoff even for high exceedance probabilities (e.g. catchments # 1100, # 1501 and # 3202). Runoff values with an exceedance probability between 0.2 and around 0.01 are systematically underestimated by modelled data from both optimisation frameworks. Modelled and observed FDCs of Q_I are shown in figure 12 for DoY > 200. With the exception of catchment # 1100, the modelling results from the diagnostic framework show a closer fit to the observed data. For both optimisation frameworks a trend of underestimating the highest runoff values can be observed. This trend is, however, more pronounced for the standard framework. With the exception of catchment # 1100, no model run fulfils the condition $\sum_{i=1}^{n} Q_I^* > \sum_{i=1}^{n} Q_I$, resulting in all model runs being rejected (results not shown).

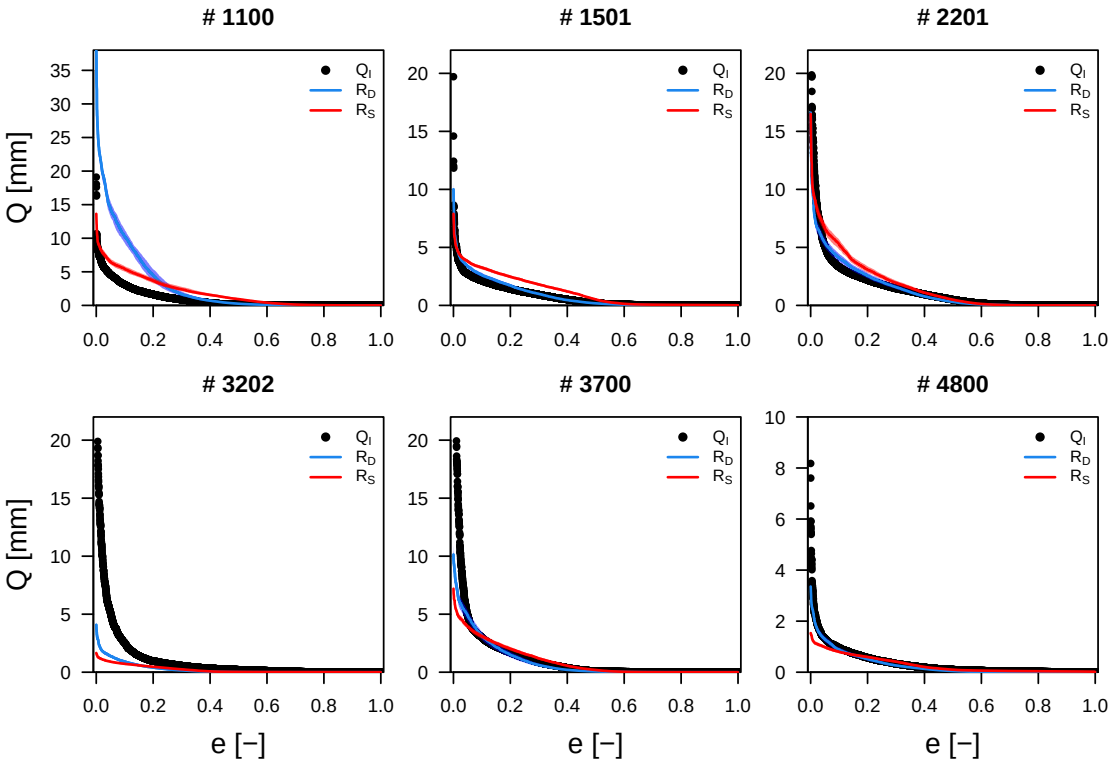


Figure 12: FDCs of Q_I for all six catchments (black dots). The median of the modelled FDC from the R_D and R_S framework are displayed in blue and red, respectively. The blue and red shaded areas are the 95% confidence interval obtained from the 100 MC runs. Only values for DoY $>$ 200 are displayed

For runoff values with low to medium exceedance probabilities (0.2 - 0.6), the standard framework results in higher runoff values than the diagnostic framework. For exceedance probabilities above around 0.6 both frameworks result in minimal or zero runoff. In particular for catchments # 3202 and # 4800, the standard framework underestimates runoff values with an exceedance probability below 0.2 (exact value catchment dependent).

Modelled and observed glacier mass balances for the calibration period from 1976 to 2006 are shown in figure 13. The standard framework shows a closer fit to the observed data in 2003 for all six catchments. The observed data point lies within the uncertainty range of the standard framework for all catchments except # 3700. Modelling results from the diagnostic framework do not capture the observed data. Furthermore, with the exception of catchment # 1100, the diagnostic framework systematically underestimates the loss in glacier mass balance. Note that for

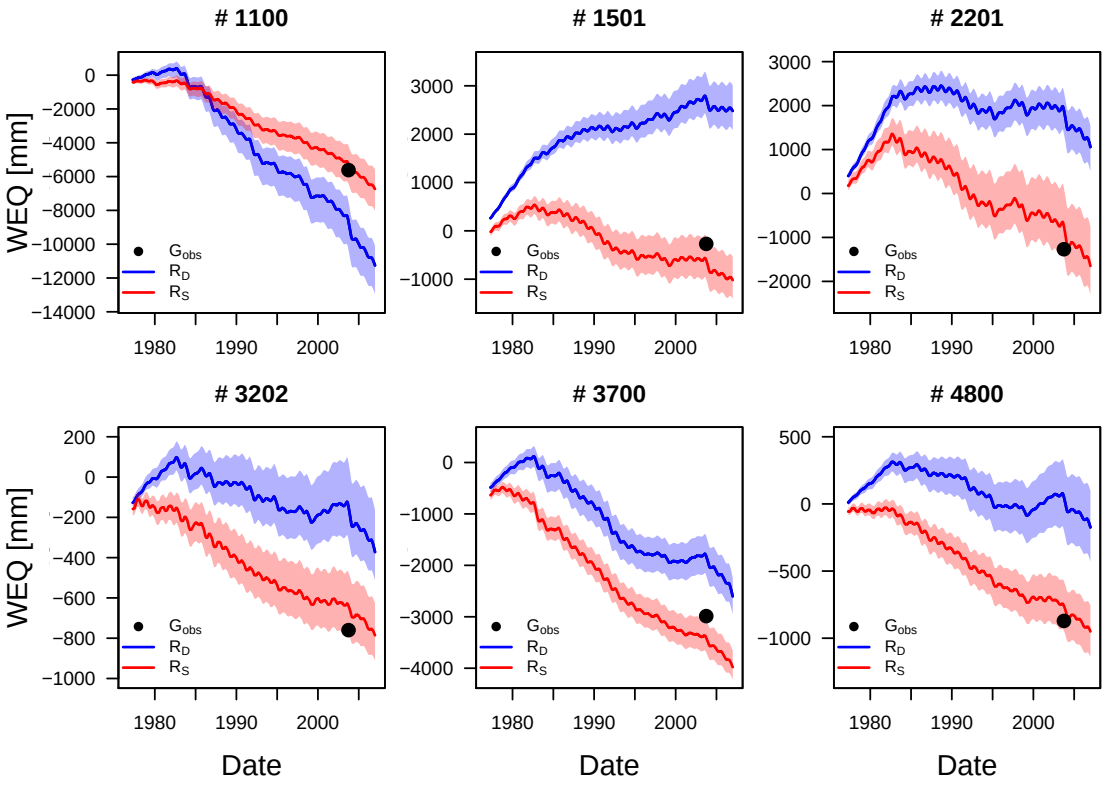


Figure 13: Observed glacier mass balance given in water equivalent (WEQ) relative to the initial conditions in 1973 for all six catchments (black dots). The median of the modelled WEQ from the R_D and R_S framework are displayed in blue and red, respectively. The blue and red shaded areas are the 95% confidence interval obtained from the 100 MC runs. The modelled data is smoothed using a 200 day moving average.

catchments # 1100 the diagnostic parameter sets also strongly overestimate the observed Q_I (see figure 12). For catchments # 1501 and # 2201 a gain of glacier mass is modelled by the diagnostic framework.

The model efficiencies of R_G and R_{Q_I} with parameter sets from the diagnostic framework show relations varying in their respective correlation coefficient (figure 14). The pearson correlation of the two model efficiencies ranges from -0.82 (catchments # 4800) to 0.98 (catchment # 1100). These pearson correlations in turn are correlated with the percentage of glacierised area in each catchment. For catchments with a high glacierised area, the correlation by trend is positive. For catchments with lower glacierised area, in contrast, the correlation is negative (see figure 15). Catchment # 1501 is diverging from this pattern.

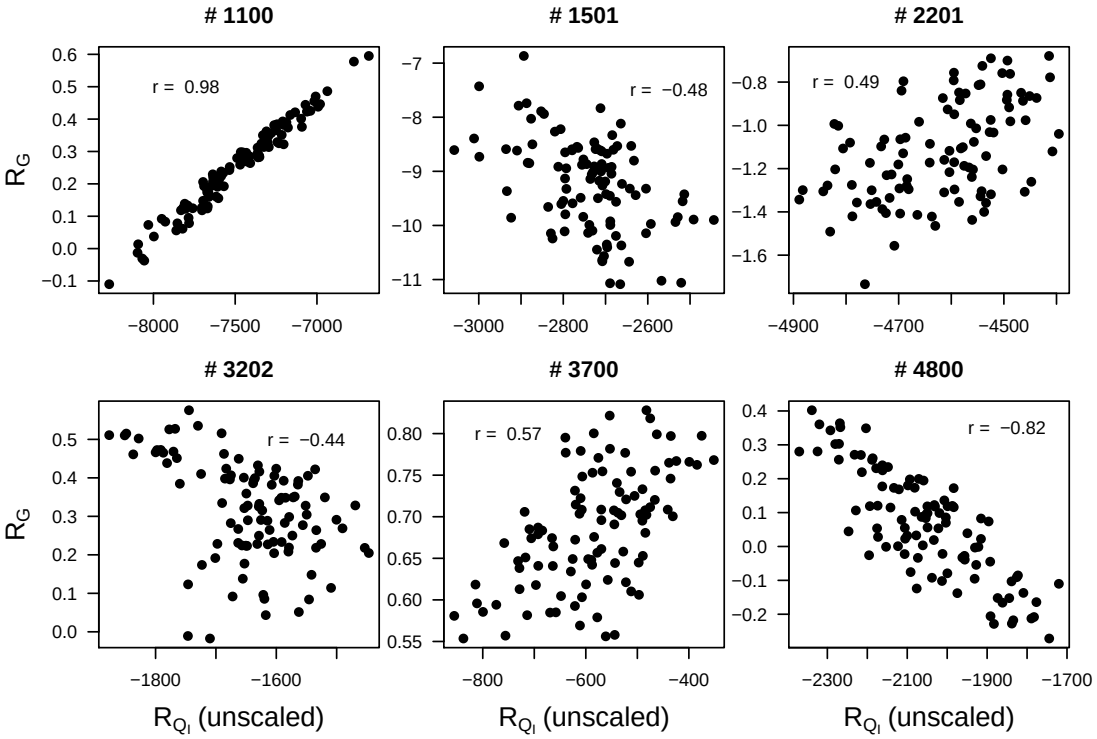


Figure 14: Scatterplots of the R_G and R_{Q_I} objective functions received from parameter sets from the diagnostic optimisation framework with the respective pearson correlation.

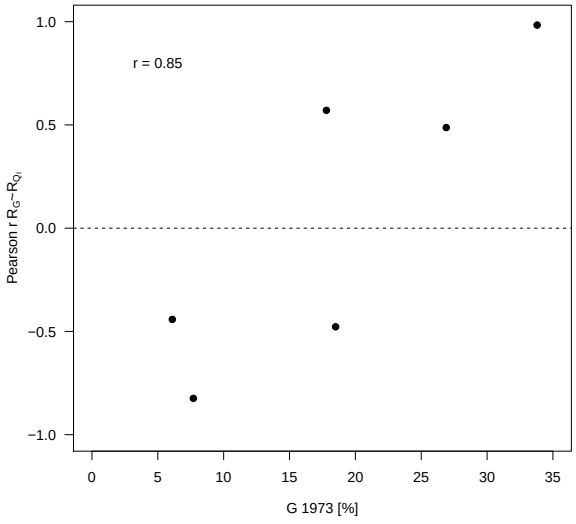


Figure 15: Pearson correlation between the correlation of R_G and R_{Q_I} and the galcier coverage in 1973 for all six catchments. The correlation is significant ($P < 0.05$).

An analysis of the relation between the objective functions $R_{Q_{tot}}$ and biasRR show a positive correlation for all six catchments. If Q^*_{tot} matches the observed FDC, the water balance is almost closed (results not shown).

Figure 16 shows scatterplots of the model efficiencies of the R_{SCA} and R_{SWE} objective functions obtained from the MC runs with parameter sets from the standard framework. With one exception, these two objective functions show a negative correlation. The changes in absolute model efficiencies especially of R_{SWE} are marginal, however. Results from the diagnostic framework show a similar pattern (data not shown).

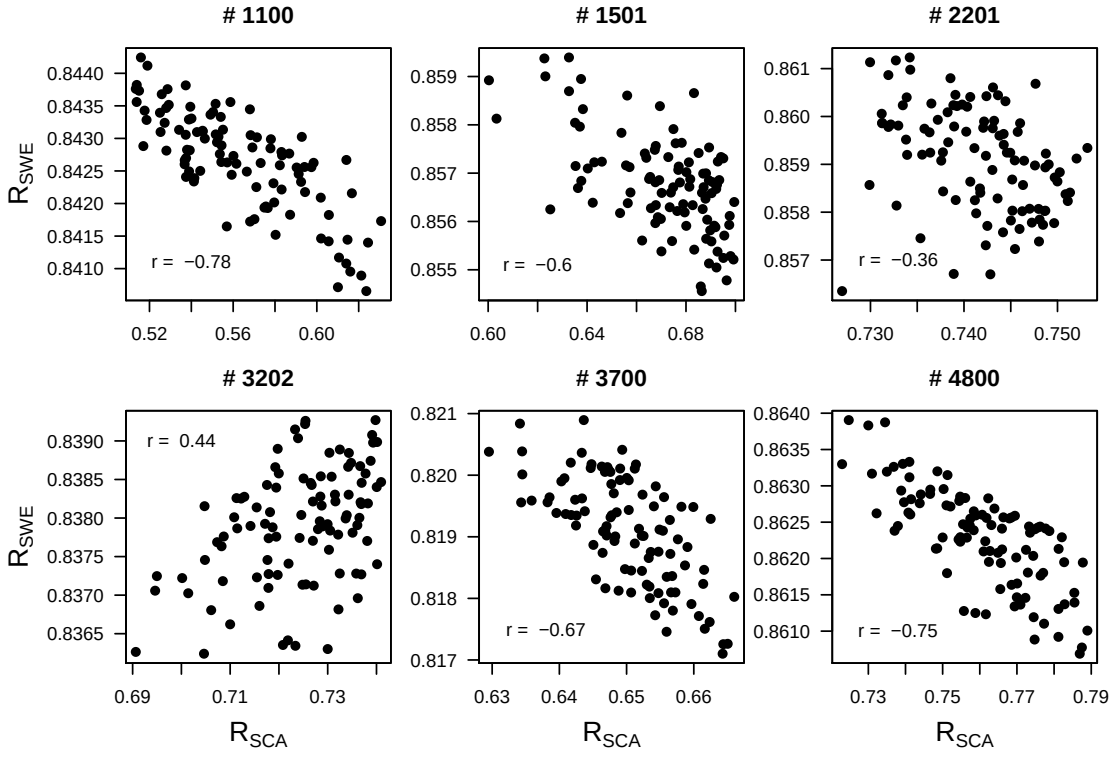


Figure 16: Scatterplots of the R_{SWE} and R_{SCA} objective functions received from parameter sets from the standard optimisation framework with the respective pearson correlation.

5 Discussion

The work at hand explored model calibration frameworks, based on different GOF measures as well as different time series of observational data. Time series of runoff from snow and glacier ice melt were calculated from SWE data and diurnal discharge amplitudes, respectively. The effect of calibrating snow and glacier melt runoff to these time series was evaluated and compared to results, where the respective model output was not directly calibrated against observational data. Furthermore, model efficiencies and interchangeability of parameter sets derived from optimising residual based summary metrics, as compared to hydrological signatures, were evaluated. Finally the effect of the used optimisation algorithm on overall model efficiency was compared to the results of Stahl et al. (2017).

5.1 Snow and Glacier Runoff

This section refers to results from the diagnostic optimisation framework. Where results from the standard framework are taken into account, it is explicitly stated. The time series of Q_S obtained from SWE data in the elevation between 2000 and 2500 m. a.s.l. shows some accordance with the increase of total runoff in spring. However, model output of Q_S^* does not capture the observed pattern. Most importantly, the Q_S time series does not cover the decline in SWE over the whole elevation range of the modelled catchments, which leads to a series of shortcomings in the representation of runoff generated by snow melt.
When comparing modelled and observed FDCs of Q_S and Q_S^*, model output overestimates low to medium flows. This overestimation most likely is not related to a malfunction in the HBV-light snow routine, but the lack of information on snow melt in altitudes below 2000 m. a.s.l.. Hence, snow melt processes in lower altitudes are not captured by the observational data, but modelled by HBV-light. In accordance with Stahl et al. (2017), who stated the need for SWE data in altitudes below 2000 m. a.s.l. for model calibration of SWE*, the calibration of Q_S would also benefit benefit from such data.
For snow melt runoff with low exceedance probabilities, Q_S^* systematically underestimates Q_S. This underestimation can most likely be attributed to a mismatch

in the conceptualisation of runoff delay in the Q_S time series compared to HBV-light. While Q_S^* requires time to pass the HBV-light subroutines, Q_S is delayed with an exponential moving average. An indication for an underestimated delay in the Q_S time series, however, is the finding that Q_S in some cases exceeds Q_{tot} during the spring snow melt.

To test the accuracy of the delay in the Q_S time series, artificial tracers could be applied during the built up of the snowpack in different elevations (Bales et al., 1993). In connection with SWE measurements, tracers could be used to investigate the time, snow melt requires to pass through a certain catchment. With measurements of SWE decrease in an area an artificial tracer was applied and tracer measurements at the streamflow gauge, travel times of snow melt runoff from different elevations in a catchment could be obtained. With these travel times, an elevation specific delay of Q_S could be obtained by applying elevation specific weighting functions (e.g. moving averages). However, such approaches are cost intensive as well as catchment specific.

Natural hydrochemical tracers for determining glacier runoff, as proposed by Frenierre and Mark (2014) are less suitable for the specific problem, since these tracers lack the information of origin within the catchment. Furthermore, snow and glacier melt are not necessarily sufficiently different in their composition of natural tracers – e.g. water stable isotopes in the water molecule – to be differentiated (Frenierre & Mark, 2014).

A comparison of both modelling frameworks reveals a systematic malfunction in the design of the time series of Q_I. This malfunction is especially evident when comparing modelling results of Q_I^* with the modelled and observed glacier mass balance. Furthermore, the applicability of Q_I as a representation of glacier runoff is dependant on the glacier coverage.

The following evaluations refer to the catchments #1100, #1501, #2201 and #3700, which had a glacier coverage of over 17.8% in 1973. The modelling results from the diagnostic framework mostly capture the FDCs of Q_I, the decline in glacier mass is underestimated, however. A notable exception is catchment # 1100, where this pattern is reversed. Modelling results from the standard framework, on the other hand, show a close fit to the observed decline in glacier mass balance, while overestimating medium to high Q_I values.

These findings indicate that the discharge amplitudes underestimate runoff originating from glacier ice melt, which is consistent with Jansson et al. (2003). When Q_I^* is calibrated to a time series which underestimates glacier runoff, HBV-light compensates by underestimating loss in glacier mass balance to achieve pareto optimal parameter sets, given the employed objective functions.
Given the aforementioned findings, the malfunction of the employed rejection sampling of parameter sets where $\sum_{i=1}^{n} Q_I^* < \sum_{i=1}^{n} Q_I$ is surprising. An explanation can be found in the design of the hydrograph partitioning. By solely discarding Q_I values above 20 mm, it is likely, that discharge events driven by non solid precipitation remain in the separated hydrograph. These discharge events likely result in an overestimation of Q_I peak flows which are mostly not captured by Q_I^*. In addition, the used maximum likelihood approach for optimising Q_I^* emphasizes the representation of high values (Nguyen & Welsch, 2010). These high flows, however, are underestimated by the glacier routine. This combination shapes the Q_I^* FDCs in such a manner, that peak flows as well as flows with exceedance probabilities above 0.4 are underestimated. Thus, for the employed rejection criteria to be meaningful, the hydrograph partitioning as well as the objective function should be redesigned.
With the aforementioned results, the positive correlation of R_G and R_{Q_I} in catchments with high glacier coverage is unexpected. The employed maximum likelihood approach is most likely causing this correlation. When peak flows are fitted closely, R_{Q_I} is higher, which leads to a slightly increased decline in glacier mass. When employing an objective function which does not emphasize peak flows, the correlation is most likely negative for all catchments analysed.
The negative correlation of R_G and R_{Q_I} in catchments with low glacier coverage, in contrast, results from an underestimation of peak flows of Q_I^*. This underestimation probably arises from rain and snow melt contributions to high discharge amplitudes. Especially catchment #3202 indicates that other runoff sources apart from bare glacier melt are included in some of the Q_I time series. This effect, however, seems to be less pronounced in catchments with higher glacier coverage, as indicated by the correlation with glacier coverage in 1973.
These findings suggest, DoY 200 being too early to exclude other melt components than bare glacier melt from diurnal discharge amplitudes, although late July

is identified as the beginning of increased glacier melt contribution to runoff in a semi arid catchment by He et al. (2018). The results of this work coincide with the findings of Koboltschnig and Schöner (2011), who identified mid to late August as the beginning of glacier ice melt being the main contribution from melt processes to streamflow in the Austrian Alps.

To facilitate the use of diurnal discharge amplitudes to derive a time series of glacier ice melt, further improvements in the employed hydrograph partitioning are necessary. Especially an exclusion of large rain events from the time series is needed, in order to obtain realistic high flow values. To avoid the time consuming task of manually excluding rain events, an automatic exclusion, based on temporal autocorrelation of the diurnal discharge amplitudes, could be employed. Measurements where the amplitude strongly exceeds the amplitudes on previous days, could be excluded from the time series. In combination with an analysis of precipitation time series, such an algorithm possibly is capable of identifying large rain events (cf. Seibert et al., 2015).

In combination with the data on glacier mass, a scaling method could be incorporated to increase Q_I to the point, where modelled glacier mass balance and Q_I^* are both able to satisfy the employed objective function. For catchments, where data on glacier mass balance is available, such an empirical scaling factor could add information on the timing of glacier melt runoff. Furthermore, the transferability of such a scaling factor to catchments without glacier mass balance data, but sub daily discharge measurements, could be tested. If such scaling factors are transferable, information on glacier melt runoff potentially could be derived solely from discharge measurements.

Periods dominated by discharge from snow and glacier melt are distinguishable by the lower and higher diurnal discharge amplitudes, respectively (Seibert et al., 2015). This distinction further enables the use of data on diurnal discharge amplitudes for hydrograph partitioning by dominant runoff source. Thus, methods of using air temperature gradients and precipitation data for determining runoff from snow and glacier melt (e.g. He et al., 2018, 2015) could possibly be enhanced to achieve more robust results.

5.2 Comparison of Model Efficiencies

Model parameters from the diagnostic framework are not able to adequately reproduce the objective functions of the standard framework. The other way around, the standard framework does capture the employed hydrological signatures of the diagnostic framework (excluding the FDCs for snow and glacier runoff, as discussed above).

Two main reasons can be identified for the malfunctioning of the diagnostic framework. First, although the FDC has successfully been used for model calibration purposes (e.g. Sadegh et al., 2016, 2015; Pokhrel et al., 2012; Blazkova & Beven, 2009b; Son & Sivapalan, 2007; Yadav et al., 2007), it does not provide information on runoff timing (Yilmaz et al., 2008). With the implemented maximum likelihood optimisation, high flows is given a high leverage during the optimisation, as discussed above. Furthermore, the employed objective function for streamflow timing also only targets high flows ($>$ Q25; exceedance probability). These two implementations lead to an emphasis on the representation of high flows by the diagnostic framework. Thus, especially the timing of low and medium flows is improvable, as can be seen by the inability of the diagnostic framework to result in acceptable model efficiency for the objective functions based on the NSE (R_{Q_1} – R_{Q_3}).

To overcome these shortcomings in the design of the used objective functions, multiple signatures, representing different parts of the FDC could be employed (Pfannerstill et al., 2014; Yilmaz et al., 2008). Furthermore, when optimising these signatures via maximum likelihood, a weighted optimisation, based on the leverage of a certain data point could be employed (Nguyen & Welsch, 2010), to reduce the bias towards extreme values inherent in maximum likelihood optimisation.

On the other hand, the calibration on these NSE based performance criteria does capture the runoff timing across the range of the runoff values as well as overall catchment behaviour represented by the FDC. The separate calibration on residual based metrics for total runoff, low flows as well as for the melt season (R_{Q_1} – R_{Q_3}) seems to be a robust choice of GOF measures for the analysed catchments. These findings coincide with McMillan and Clark (2009), who also found formal

likelihood measures being outperformed by residual based summary metrics for both streamflow timing as well as overall catchment behaviour.
The second reason for parameter sets from the diagnostic framework not resulting in high model performances for the standard framework lies in the combination of R_G and R_{Q_I}. While values for R_{Q_I} are reasonably high, R_G is not matched by the diagnostic framework. The standard framework, on the other hand, yields a good performance for R_G and does not need to fulfil a performance criteria for glacier melt runoff. Thus, when transferring parameter sets from the diagnostic framework, R_G is not matched, while no other objective function compensates this mismatch. When transferring parameters from the standard framework, R_G is matched, which compensates for low R_{Q_I} values in overall model performance.
Arguably, the standard framework as implemented by Stahl et al. (2017) can be considered diagnostic in terms of using multiple data sets on internal catchment behaviour (Kelleher et al., 2017; Shafii & Tolson, 2015; Gupta et al., 2012; Clark et al., 2011; Gupta et al., 2008). In addition, the multi objective calibration of the total runoff on three different performance criteria emphasizing different seasons and runoff processes goes in the direction of hydrograph partitioning for model calibration as done by Larabi et al. (2018); He et al. (2018, 2015); Shafii et al. (2017).
Given the stable runoff regime in the analysed catchments, however, this calibration could be improved by further increasing selective targeting of dominant runoff processes (e.g. by using diurnal discharge amplitudes as discussed above). With dominant runoff processes identified, it would also be possible to employ seasonally varying parameter sets emphasizing the source, runoff originates from.

5.3 Uncertainty and Correlation Analysis

The uncertainty analysis employed gives a first insight into model uncertainty for the analysed catchments. However, with the multimodal posterior parameter densities, the used sampling scheme does not provide an accurate uncertainty estimation. Due to computational time needed, a more sophisticated uncertainty estimation, using GLUE was not feasible.
Although the implementation of ABC into the optimisation framework was be-

yond the scope of this work, this uncertainty analysis would provide a couple of advantages over using GLUE. With ABC the uncertainty analysis relies on the same runs as the model calibration itself. Therefore, ABC can be considered as relatively efficient, regarding computational time. Furthermore, it allows using residual based summary metrics within a MCMC optimisation framework, relaxing the statistical assumption of one best set of parameters derivable by optimising a formal likelihood function (Sadegh & Vrugt, 2013).
Given the residual based metrics within the employed frameworks in this work, as well as the slightly better performance of the DREAM$_{ZS}$ algorithm in comparison to the algorithm used by Stahl et al. (2017), the implementation of ABC could be a step towards computationally efficient and statistically correct model optimisation of the HBV-light model.
Although improvable, the employed uncertainty analysis revealed two relations in the employed objective functions. Most notably, R$_{SWE}$ performs worse, if R$_{SCA}$ has higher values. The reason for this negative correlation could lie within the conceptualisation of the HBV-light snow routine. However, another explanation arises from the used objective functions. For optimising SCA, the MAE is used, which gives the same weight to all model residuals. The RMSE, applied for optimising SWE on the other hand, is more sensitive to outliers, which can produce biased model results, especially with non Gaussian error distributions (Chai & Draxler, 2014). Calibrating one model routine on two objective functions with slightly different focus possibly results in the aforementioned negative correlation. Secondly, the values of RQ_{tot} and biasRR show a positive correlation. This is little surprising, since the FDC passively contains information on cumulative runoff leaving the catchment. When calibrating on the entire FDC, including an objective function for closing the water balance seems to be arbitrary.

6 Conclusion

The presented work analysed the applicability of hydrograph partitioning to calibrate the snow and glacier subroutines of the HBV-light model. Diurnal discharge amplitudes were used to derive a time series of runoff from glacier ice melt. The obtained time series was found not to be coinciding with data on glacier mass balance. Due to the separation method, most parts of the time series systematically underestimate glacier runoff. Furthermore, rain events are not satisfactorily excluded from the time series, leading to an overestimation of peak flows. Hence, the used time series on glacier runoff does not increase model realism or parameter robustness.

For diurnal discharge amplitudes to be used for hydrograph partitioning within a diagnostic model optimisation framework, the employed partitioning method is to be enhanced with e.g. an empirical scaling approach (see section 5.1). In addition, rain events have to be excluded either manually or with an autocorrelation analysis as described in section 5.1, to obtain a realistic estimation of glacier runoff.

A time series representing runoff from snow melt was calculated from SWE data. Snow melt runoff was delayed by using an exponential moving average. However, this delay probably is underestimating travel times of snow melt runoff through the analysed catchments. To enable the use of SWE data as an estimate for snow melt runoff, an investigation of travel times e.g. with artificial tracers could be employed. Such an analysis could allow a height dependent delay in the calculation of snow melt runoff from SWE data. In addition, SWE data for the entire elevation range of the analysed catchments would be desirable.

The parameter sets from the diagnostic optimisation framework were not able to capture overall catchment behaviour as well as streamflow timing. Especially the maximum likelihood based optimisation approach leads to a strong emphasize on reproducing peak flows. This pattern is further intensified by the employed signature for streamflow timing.

Parameter sets from the optimisation of residual based metrics, in contrast, show an accordance with streamflow timing across a wide range of runoff values. Furthermore, these parameter sets capture overall catchment behaviour of total streamflow as represented by the FDC. These findings lead to the conclusion, that if

combined reasonably, summary metrics provide a solid basis for diagnostic model optimisation. As already mentioned by numerous studies, this is especially true, when calibration is carried out on multiple sets of observational data.

References

Andrews, F. (2012). hydromad: Hydrological model assessment and development [Computer software manual]. Retrieved from `http://hydromad.catchment.org/` (R package version 0.9.10)

Bales, R. C., Davis, R. E., & Williams, M. W. (1993). Tracer release in melting snow: diurnal and seasonal patterns. *Hydrological Processes*, *7*(4), 389–401.

Bergström, S. (1976). *Development and application of a conceptual runoff model for scandinavian catchments.* Norkoeping: SMHI RHO 7.

Bergström, S. (1995). The HBV model. In V. P. Singh (Ed.), *Computer models of watershed hydrology.* Colorado: Water Resources Publications.

Beven, K. J. (2007). *Environmental modelling: An uncertain future?* London: CRC Press.

Beven, K. J. (2012). *Rainfall-runoff modelling: the primer.* Chichester: John Wiley & Sons.

Beven, K. J., & Binley, A. (1992). The future of distributed models: model calibration and uncertainty prediction. *Hydrological Processes*, *6*(3), 279–298.

Beven, K. J., & Binley, A. (2014). GLUE: 20 years on. *Hydrological Processes*, *28*(24), 5897–5918.

Beven, K. J., & Freer, J. (2001). Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. *Journal of Hydrology*, *249*(1-4), 11–29.

Blazkova, S., & Beven, K. J. (2009b). A limits of acceptability approach to model evaluation and uncertainty estimation in flood frequency estimation by continuous simulation: Skalka catchment, Czech Republic. *Water Resources Research*, *45*(12), 1–12.

Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. *Geoscientific Model Development*, *7*(3), 1247–1250.

Chow, V. T., et al. (1964). *Handbook of applied hydrology.* New York: McGraw-Hill Inc.

Chu, W., Gao, X., & Sorooshian, S. (2010). Improving the shuffled complex

evolution scheme for optimization of complex nonlinear hydrological systems: Application to the calibration of the sacramento soil-moisture accounting model. *Water Resources Research*, *46*(9), 1–12.

Clark, M. P., Kavetski, D., & Fenicia, F. (2011). Pursuing the method of multiple working hypotheses for hydrological modeling. *Water Resources Research*, *47*(9), 1–16.

Clarke, R. (1973). A review of some mathematical models used in hydrology, with observations on their calibration and use. *Journal of Hydrology*, *19*(1), 1–20.

Cunnane, C. (1978). Unbiased plotting positions—a review. *Journal of Hydrology*, *37*(3-4), 205–222.

Duan, Q., Gupta, V. K., & Sorooshian, S. (1993). Shuffled complex evolution approach for effective and efficient global minimization. *Journal of Optimization Theory and Applications*, *76*(3), 501–521.

Duan, Q., Sorooshian, S., & Gupta, V. (1992). Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resources Research*, *28*(4), 1015–1031.

Feyen, L., Vrugt, J. A., Nualláin, B. Ó., van der Knijff, J., & De Roo, A. (2007). Parameter optimisation and uncertainty assessment for large-scale streamflow simulation with the lisflood model. *Journal of Hydrology*, *332*(3-4), 276–289.

Finger, D., Vis, M., Huss, M., & Seibert, J. (2015). The value of multiple data set calibration versus model complexity for improving the performance of hydrological models in mountain catchments. *Water Resources Research*, *51*(4), 1939–1958.

Fleming, G. (1975). *Computer simulation techniques in hydrology.* Elsevier, New York.

Forrest, S. (1993). Genetic algorithms: principles of natural selection applied to computation. *Science*, *261*(5123), 872–878.

Franchini, M., & Galeati, G. (1997). Comparing several genetic algorithm schemes for the calibration of conceptual rainfall-runoff models. *Hydrological Sciences Journal*, *42*(3), 357–379.

Frenierre, J. L., & Mark, B. G. (2014). A review of methods for estimating the

contribution of glacial meltwater to total watershed discharge. *Progress in Physical Geography: Earth and Environment*, *38*(2), 173-200.

Freudiger, D., Kohn, I., Seibert, J., Stahl, K., & Weiler, M. (2017). Snow redistribution for the hydrological modeling of alpine catchments. *Wiley Interdisciplinary Reviews: Water*, *4*(5), 1–16.

Frick, C., Steiner, H., Mazurkiewicz, A., Riediger, U., Rauthe, M., Reich, T., & Gratzki, A. (2014). Central european high-resolution gridded daily data sets (HYRAS): mean temperature and relative humidity. *Meteorologische Zeitschrift*, *23*(1), 15–32.

Gelman, A., & Rubin, D. B. (1992a). A single series from the Gibbs sampler provides a false sense of security. *Bayesian Statistics*, *4*, 625–631.

Gelman, A., & Rubin, D. B. (1992b). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*(4), 457–472.

Guo, J., Zhou, J., Song, L., Zou, Q., & Zeng, X. (2013). Uncertainty assessment and optimization of hydrological model with the shuffled complex evolution metropolis algorithm: an application to artificial neural network rainfall-runoff model. *Stochastic Environmental Research and Risk Assessment*, *27*(4), 985–1004.

Gupta, H. V., Clark, M. P., Vrugt, J. A., Abramowitz, G., & Ye, M. (2012). Towards a comprehensive assessment of model structural adequacy. *Water Resources Research*, *48*(8), 1–16.

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*(1-2), 80–91.

Gupta, H. V., Wagener, T., & Liu, Y. (2008). Reconciling theory with observations: elements of a diagnostic approach to model evaluation. *Hydrological Processes*, *22*(18), 3802–3813.

Gustard, A., & Demuth, S. (2008). *Manual on low-flow estimation and prediction – Operational hydrology report no. 50* (Vol. 1029). Vienna: World Meteorological Organization.

Hartig, F., Minunno, F., & Paul, S. (2017). BayesianTools: General-Purpose MCMC and SMC Samplers and Tools for Bayesian Statistics [Computer software manual]. Retrieved from `https://CRAN.R-project.org/package=`

`BayesianTools` (R package version 0.1.3)

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, *57*(1), 97–109.

He, Z. H., Tian, F. Q., Gupta, H. V., Hu, H. C., & Hu, H. P. (2015). Diagnostic calibration of a hydrological model in a mountain area by hydrograph partitioning. *Hydrology and Earth System Sciences*, *19*(4), 1807–1826.

He, Z. H., Vorogushyn, S., Unger-Shayesteh, K., Gafurov, A., Kalashnikova, O., Omorova, E., & Merz, B. (2018). The value of hydrograph partitioning curves for calibrating hydrological models in glacierized basins. *Water Resources Research*, *54*, 2336–2361.

Hock, R. (2003). Temperature index melt modelling in mountain areas. *Journal of Hydrology*, *282*(1-4), 104–115.

Hock, R. (2005). Glacier melt: a review of processes and their modelling. *Progress in Physical Geography*, *29*(3), 362–391.

Hornberger, G. M., & Spear, R. (1980). Eutrophication in Peel Inlet—I. the problem-defining behavior and a mathematical model for the phosphorus scenario. *Water Research*, *14*(1), 29–42.

Hornberger, G. M., & Spear, R. C. (1981). Approach to the preliminary analysis of environmental systems. *Journal of Environmental Management*, *12*(1), 7–18.

Huss, M., Jouvet, G., Farinotti, D., & Bauder, A. (2010). Future high-mountain hydrology: a new parameterization of glacier retreat. *Hydrology and Earth System Sciences*, *14*(5), 815–829.

Huss, M., Zemp, M., Joerg, P. C., & Salzmann, N. (2014). High uncertainty in 21st century runoff projections from glacierized basins. *Journal of Hydrology*, *510*, 35–48.

Jain, S. K., & Sudheer, K. (2008). Fitting of hydrologic models: a close look at the Nash–Sutcliffe index. *Journal of Hydrologic Engineering*, *13*(10), 981–986.

Jansson, P., Hock, R., & Schneider, T. (2003). The concept of glacier storage: a review. *Journal of Hydrology*, *282*(1), 116–129.

Jonas, T., Marty, C., & Magnusson, J. (2009). Estimating the snow water equivalent from snow depth measurements in the Swiss Alps. *Journal of Hydrology*, *378*(1-2), 161–167.

Jörg-Hess, S., Fundel, F., Jonas, T., & Zappa, M. (2014). Homogenisation of a gridded snow water equivalent climatology for alpine terrain: methodology and applications. *The Cryosphere*, *8*(2), 471–485.

Kaser, G., Großhauser, M., & Marzeion, B. (2010). Contribution potential of glaciers to water availability in different climate regimes. *Proceedings of the National Academy of Sciences*, *107*(47), 20223–20227.

Kelleher, C., McGlynn, B., & Wagener, T. (2017). Characterizing and reducing equifinality by constraining a distributed catchment model with regional signatures, local observations, and process understanding. *Hydrology and Earth System Sciences*, *21*(7), 3325–3352.

Koboltschnig, G., & Schöner, W. (2011). The relevance of glacier melt in the water cycle of the Alps: the example of Austria. *Hydrology and Earth System Sciences*, *15*(6), 2039–2048.

Konz, M., & Seibert, J. (2010). On the value of glacier mass balances for hydrological model calibration. *Journal of Hydrology*, *385*(1-4), 238–246.

Krause, P., Boyle, D., & Bäse, F. (2005). Comparison of different efficiency criteria for hydrological model assessment. *Advances in Geosciences*, *5*, 89–97.

Kuczera, G. (1997). Efficient subspace probabilistic parameter optimization for catchment models. *Water Resources Research*, *33*(1), 177–185.

Laloy, E., & Vrugt, J. A. (2012). High-dimensional posterior exploration of hydrologic models using multiple-try DREAM$_{(ZS)}$ and high-performance computing. *Water Resources Research*, *48*(1), 1–18.

Larabi, S., St-Hilaire, A., Chebana, F., & Latraverse, M. (2018). Multi-criteria process-based calibration using functional data analysis to improve hydrological model realism. *Water Resources Management*, *32*(1), 195–211.

Legates, D. R., & McCabe, G. J. (1999). Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation. *Water Resources Research*, *35*(1), 233–241.

Lindström, G., Johansson, B., Persson, M., Gardelin, M., & Bergström, S. (1997). Development and test of the distributed HBV-96 hydrological model. *Journal of Hydrology*, *201*(1-4), 272–288.

Linsley, R. K., Kohler, M. A., & Paulhus, J. L. (1949). *Applied hydrology.* New York: McGraw-Hill Inc.

Loague, K. (Ed.). (2010). *Benchmark papers in hydrology: Rainfall-runoff-modelling.* IAHS Press, Wallingford.

Maisch, M. (2000). *Die Gletscher der Schweizer Alpen: Gletscherhochstand 1850, aktuelle Vergletscherung, Gletscherschwund-Szenarien.* Vdf, Hochschulverlag AG an der ETH.

McMillan, H., & Clark, M. (2009). Rainfall-runoff model calibration using informal likelihood measures within a Markov chain Monte Carlo sampling scheme. *Water Resources Research*, *45*(4), 1–12.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, *21*(6), 1087–1092.

Müller, F., Caflish, T., & Müller, G. (1976). *Firn und Eis der Schweizer Alpen: Gletscherinventar.* Eidgenössische Technische Hochschule Zürich, Geographisches Institut.

Nash, J. (1959). Systematic determination of unit hydrograph parameters. *Journal of Geophysical Research*, *64*(1), 111–115.

Nash, J., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I — A discussion of principles. *Journal of Hydrology*, *10*(3), 282–290.

Nelder, J. A., & Mead, R. (1965). A simplex method for function minimization. *The Computer Journal*, *7*(4), 308–313.

Nguyen, T.-D., & Welsch, R. E. (2010). Outlier detection and robust covariance estimation using mathematical programming. *Advances in Data Analysis and Classification*, *4*(4), 301–334.

Oudin, L., Hervieu, F., Michel, C., Perrin, C., Andréassian, V., Anctil, F., & Loumagne, C. (2005). Which potential evapotranspiration input for a lumped rainfall–runoff model?: Part 2—towards a simple and efficient potential evapotranspiration model for rainfall–runoff modelling. *Journal of Hydrology*, *303*(1-4), 290–306.

Paul, F., Frey, H., & Le Bris, R. (2011). A new glacier inventory for the European Alps from Landsat TM scenes of 2003: challenges and results. *Annals of Glaciology*, *52*(59), 144–152.

Pechlivanidis, I., Jackson, B., McIntyre, N., & Wheater, H. (2011). Catchment

scale hydrological modelling: a review of model types, calibration approaches and uncertainty analysis methods in the context of recent developments in technology and applications. *Global NEST Journal*, *13*(3), 193–214.

Pfannerstill, M., Guse, B., & Fohrer, N. (2014). Smart low flow signature metrics for an improved overall performance evaluation of hydrological models. *Journal of Hydrology*, *510*, 447–458.

Pokhrel, P., Yilmaz, K. K., & Gupta, H. V. (2012). Multiple-criteria calibration of a distributed watershed model using spatial regularization and response signatures. *Journal of Hydrology*, *418*, 49–60.

R Core Team. (2017). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `https://www.R-project.org/`

Ragettli, S., & Pellicciotti, F. (2012). Calibration of a physically based, spatially distributed hydrological model in a glacierized basin: On the use of knowledge from glaciometeorological processes to constrain model parameters. *Water Resources Research*, *48*(3), 1–20.

Rakovec, O., Kumar, R., Attinger, S., & Samaniego, L. (2016). Improving the realism of hydrologic model functioning through multivariate parameter estimation. *Water Resources Research*, *52*(10), 7779–7792.

Rango, A., & Martinec, J. (1995). Revisiting the degree-day method for snowmelt computations. *Journal of the American Water Resources Association*, *31*(4), 657–669.

Rauthe, M., Steiner, H., Riediger, U., Mazurkiewicz, A., & Gratzki, A. (2013). A Central European precipitation climatology–Part I: Generation and validation of a high-resolution gridded daily data set (HYRAS). *Meteorologische Zeitschrift*, *22*(3), 235–256.

Sadegh, M., & Vrugt, J. (2013). Bridging the gap between GLUE and formal statistical approaches: approximate Bayesian computation. *Hydrology and Earth System Sciences*, *17*(12), 4831–4850.

Sadegh, M., Vrugt, J., Gupta, H., & Xu, C. (2016). The soil water characteristic as new class of closed-form parametric expressions for the flow duration curve. *Journal of Hydrology*, *535*, 438–456.

Sadegh, M., & Vrugt, J. A. (2014). Approximate Bayesian Computation using

Markov Chain Monte Carlo simulation: DREAM$_{(ABC)}$. *Water Resources Research*, *50*(8), 6767–6787.

Sadegh, M., Vrugt, J. A., Xu, C., & Volpi, E. (2015). The stationarity paradigm revisited: Hypothesis testing using diagnostics, summary metrics, and DREAM$_{(ABC)}$. *Water Resources Research*, *51*(11), 9207–9231.

Schaefli, B., & Gupta, H. V. (2007). Do Nash values have value? *Hydrological Processes*, *21*(15), 2075–2080.

Schoups, G., & Vrugt, J. A. (2010a). A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non-gaussian errors. *Water Resources Research*, *46*(10), 1–17.

Searcy, J. K. (1959). *Flow-duration curves*. Washington: US Government Printing Office.

Seibert, J. (1997). Estimation of Parameter Uncertainty in the HBV Model: Paper presented at the Nordic Hydrological Conference (Akureyri, Iceland-August 1996). *Hydrology Research*, *28*(4-5), 247–262.

Seibert, J. (1999). Regionalisation of parameters for a conceptual rainfall-runoff model. *Agricultural and Forest Meteorology*, *98*, 279–293.

Seibert, J., & Beven, K. J. (2009). Gauging the ungauged basin: how many discharge measurements are needed? *Hydrology and Earth System Sciences*, *13*(6), 883–892.

Seibert, J., Jenicek, M., Huss, M., & Ewen, T. (2015). Chapter 4 - snow and ice in the hydrosphere. In J. F. Shroder, W. Haeberli, & C. Whiteman (Eds.), *Snow and ice-related hazards, risks and disasters.* Boston: Academic Press.

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

Seibert, J., & Vis, M. (2012). Teaching hydrological modeling with a user-friendly catchment-runoff-model software package. *Hydrology and Earth System Sciences*, *16*(9), 3315–3325.

Seibert, J., Vis, M. J. P., Kohn, I., Weiler, M., & Stahl, K. (2018). Technical note: Representing glacier geometry changes in a semi-distributed hydrological model. *Hydrology and Earth System Sciences*, *22*(4), 2211–2224.

Shafii, M., Basu, N., Craig, J. R., Schiff, S. L., & Van Cappellen, P. (2017). A

diagnostic approach to constraining flow partitioning in hydrologic models using a multiobjective optimization framework. *Water Resources Research*, *53*(4), 3279–3301.

Shafii, M., & Tolson, B. A. (2015). Optimizing hydrological consistency by incorporating hydrological signatures into model calibration objectives. *Water Resources Research*, *51*(5), 3796–3814.

Son, K., & Sivapalan, M. (2007). Improving model structure and reducing parameter uncertainty in conceptual water balance models through the use of auxiliary data. *Water Resources Research*, *43*(1), 1–18.

Spear, R., & Hornberger, G. (1980). Eutrophication in Peel Inlet — II. identification of critical uncertainties via generalized sensitivity analysis. *Water Research*, *14*(1), 43–49.

Stahl, K., Moore, R., Shea, J., Hutchinson, D., & Cannon, A. (2008). Coupled modelling of glacier and streamflow response to future climate scenarios. *Water Resources Research*, *44*(2), 1–13.

Stahl, K., Weiler, M., Kohn, I., Freudiger, D., Seibert, J., Vis, M., Gerlinger, K., & Bohm, M. (2017). The snow and glacier melt components of streamflow of the river Rhine and its tributaries considering the influence of climate change. *CHR-KHR, Lelystad*.

Ulrich, J. (2013). TTR: Technical Trading Rules [Computer software manual]. Retrieved from `https://CRAN.R-project.org/package=TTR` (R package version 0.22.0)

Van Genuchten, M. T. (1980). A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. *Soil Science Society of America Journal*, *44*(5), 892–898.

Vrugt, J. A. (2016). Markov Chain Monte Carlo simulation using the DREAM software package: Theory, concepts, and MATLAB implementation. *Environmental Modelling & Software*, *75*, 273–316.

Vrugt, J. A., Gupta, H. V., Bouten, W., & Sorooshian, S. (2003). A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. *Water Resources Research*, *39*(8), 1–18.

Vrugt, J. A., & Sadegh, M. (2013). Toward diagnostic model calibration and

evaluation: Approximate Bayesian Computation. *Water Resources Research*, *49*(7), 4335–4345.

Vrugt, J. A., Ter Braak, C., Diks, C., Robinson, B. A., Hyman, J. M., & Higdon, D. (2009). Accelerating Markov Chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling. *International Journal of Nonlinear Sciences and Numerical Simulation*, *10*(3), 273–290.

Vrugt, J. A., Ter Braak, C. J., Clark, M. P., Hyman, J. M., & Robinson, B. A. (2008a). Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with Markov chain Monte Carlo simulation. *Water Resources Research*, *44*(12), 1–15.

Yadav, M., Wagener, T., & Gupta, H. (2007). Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins. *Advances in Water Resources*, *30*(8), 1756–1774.

Yilmaz, K. K., Gupta, H. V., & Wagener, T. (2008). A process-based diagnostic approach to model evaluation: Application to the NWS distributed hydrologic model. *Water Resources Research*, *44*(9), 1–18.

Young, P. (1983). The validity and credibility of models for badly defined systems. In M. B. Beck & G. van Straten (Eds.), *Uncertainty and forecasting of water quality.* Berlin: Springer.

Appendices

A

Tables A.1 and A.2 show symbols and abbreviations used in the text respectively.

Table A.1: Symbols used in the text with their units and descriptions. Symbols marked with a * in the text refer to a modelled version of the respective symbol and are not listed separately. OF is short for objective function.

Name/Description	Unit	Symbol
25 % quantile	%	Q25
bias in runoff ratio	–	biasRR
calibration coefficient	[d/mm]	a_{VG}
calibration coefficient	–	b_{VG}
calibration coefficient	–	c_{VG}
diagnostic OF	–	R_D
exceedance probability	[–]	e
fitted calibration coefficients	–	θ
glacier mass balance	[mm]	MB_G
glacier OF	–	R_G
glacier runoff	[mm]	Q_I
likelihood	–	l
log-likelihood	–	LL
objective function	–	OF
natural logarithm	–	ln
normalisation range	–	N
Pearson correlation	–	r
Rank	–	R
Runoff	[mm]	Q
runoff OFs	–	R_{Qtot}, R_{QI},R_{QS}, R_{Q1}, R_{Q2}, R_{Q3}
Snow melt runoff	[mm]	Q_S
standard OF	–	R_S
SCA OF	–	R_{SCA}
SWE OF	–	R_{SWE}
Time	[d]	t
Total runoff	[mm]	Q_{tot}
Discharge measurement	[mm]	y

Table A.2: Abbreviations used in the text with their units and descriptions. Abbreviations marked with a * in the text refer to a modelled version of the respective abbreviation and are not listed separately.

Name/Description	Unit	Abbreviation
Approximate Bayesian Computation	–	ABC
Cross Correlation Function	–	CCF
Cumulative density function	–	CDF
Day of Year	–	DoY
DiffeRential-Evolution-Adaptive-Metropolis	–	DREAM
Flow Duration Curve	–	FDC
Generalised Likelihood Uncertainty Estimation	–	GLUE
genetic algorithm	–	GAP
Goodness of fit	–	GOF
Hydrologiska Byråns Vattenbalansavdelnin	–	HBV
Markov-Chain-Monte-Carlo	–	MCMC
Mean absolute error	–	MAE
Monte-Carlo	–	MC
Nash-Sutcliffe-Efficiency	–	NSE
Objective function	–	OF
Probability density function	–	PDF
Root-Mean-Squared-Error	–	RMSE
Shuffled-Complex-Evolution	–	SCE
Shuffled-Complex-Evolution-Metropolis	–	SCEM
Snow-Covered-Area	[%]	SCA
Snow-Water-Equivalent	[mm]	SWE
Van Genuchten	–	VG

B

The results of the sensitivity analysis for catchment # 1100 are shown in figures B.1 to B.11.

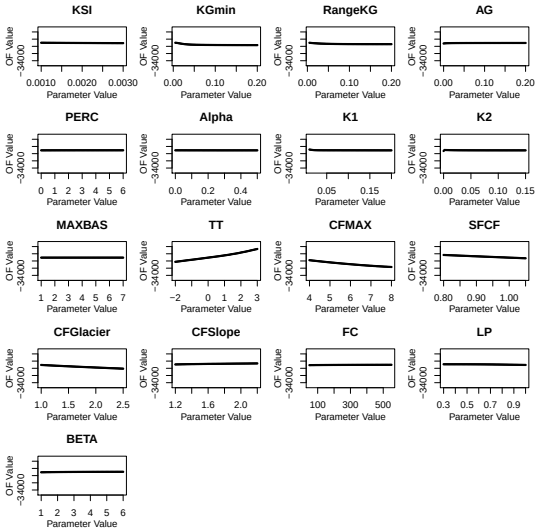


Figure B.1: Sensitivity of $R_{Q_{tot}}$ to one at a time parameter perturbations of selected HBV-light parameters.

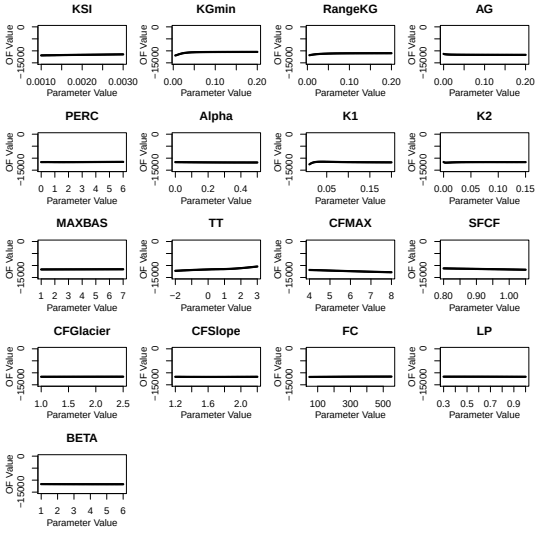


Figure B.2: Sensitivity of R_{Q_S} to one at a time parameter perturbations of selected HBV-light parameters.

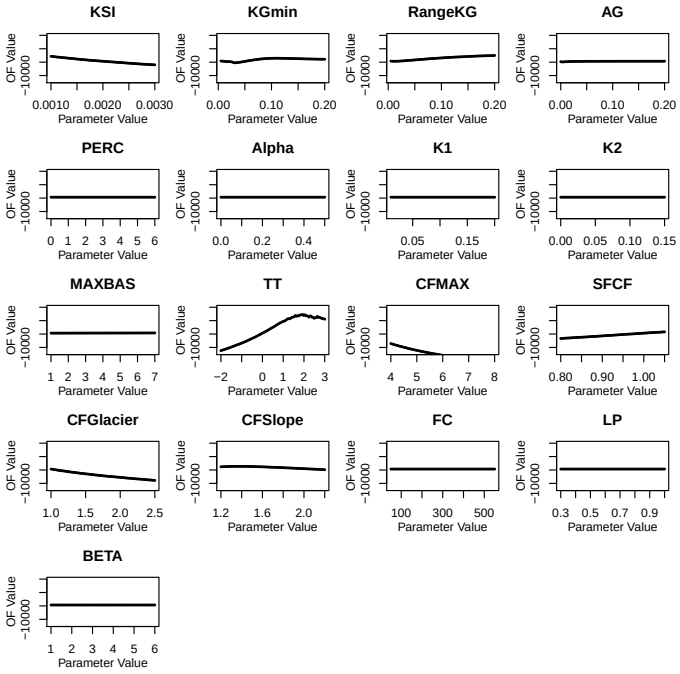


Figure B.3: Sensitivity of R_{Q_I} to one at a time parameter perturbations of selected HBV-light parameters.

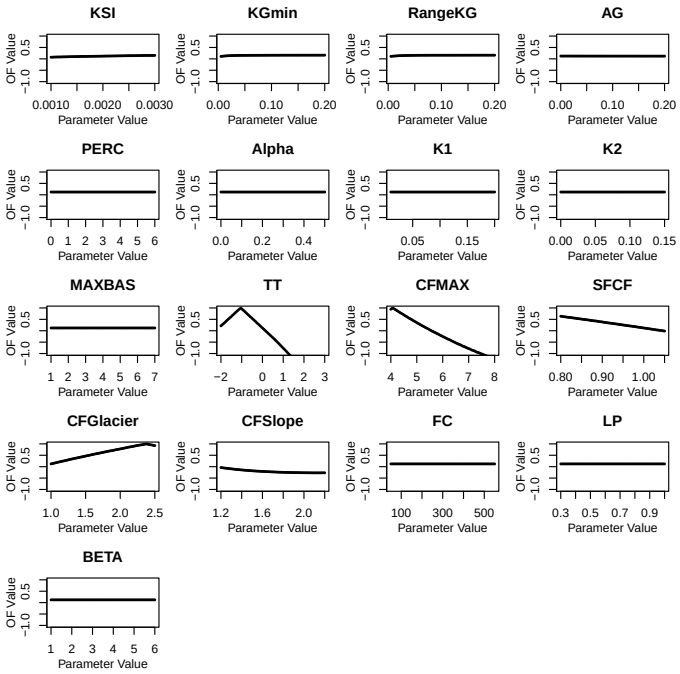


Figure B.4: Sensitivity of R_G to one at a time parameter perturbations of selected HBV-light parameters.

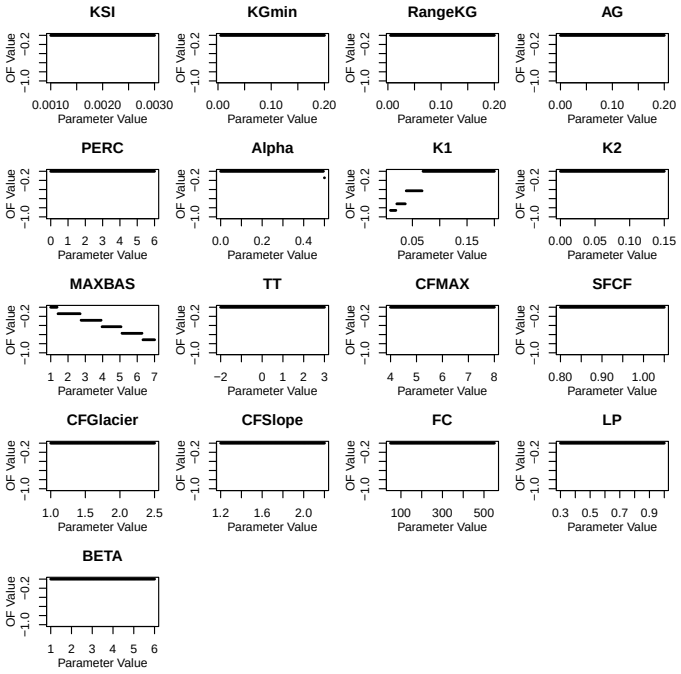


Figure B.5: Sensitivity of R_{timing} to one at a time parameter perturbations of selected HBV-light parameters.

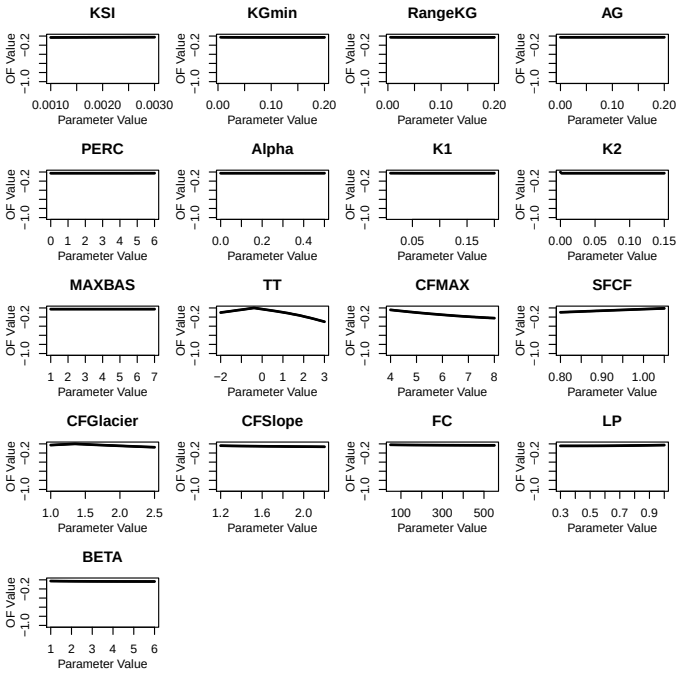


Figure B.6: Sensitivity of biasRR to one at a time parameter perturbations of selected HBV-light parameters.

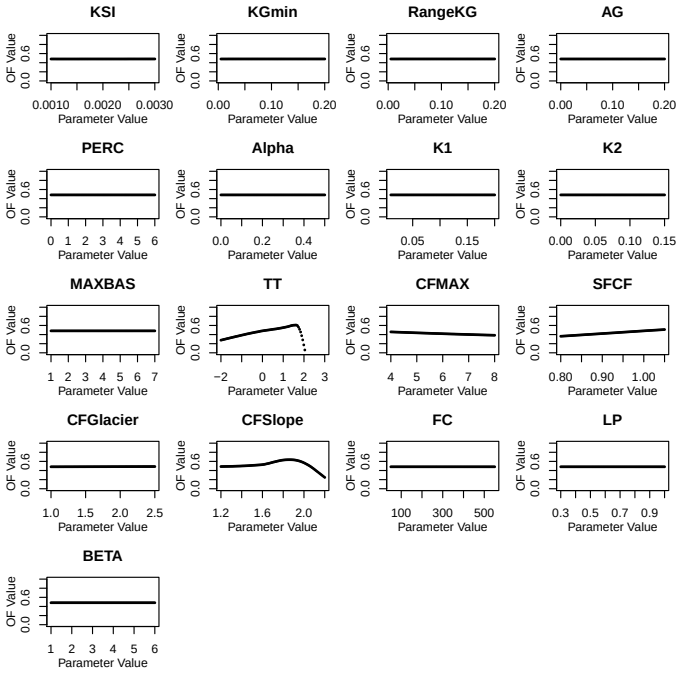


Figure B.7: Sensitivity of R_{SWE} to one at a time parameter perturbations of selected HBV-light parameters.

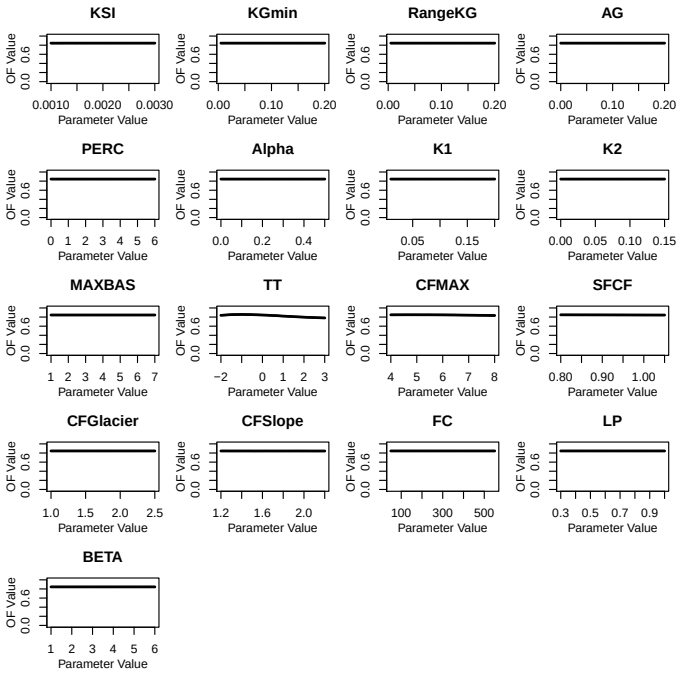


Figure B.8: Sensitivity of R_{SCA} to one at a time parameter perturbations of selected HBV-light parameters.

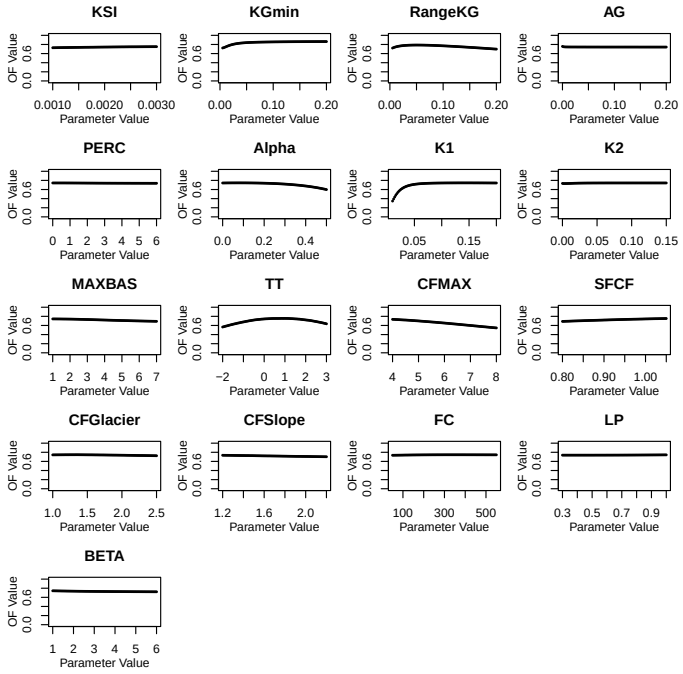


Figure B.9: Sensitivity of R_{Q_1} to one at a time parameter perturbations of selected HBV-light parameters.

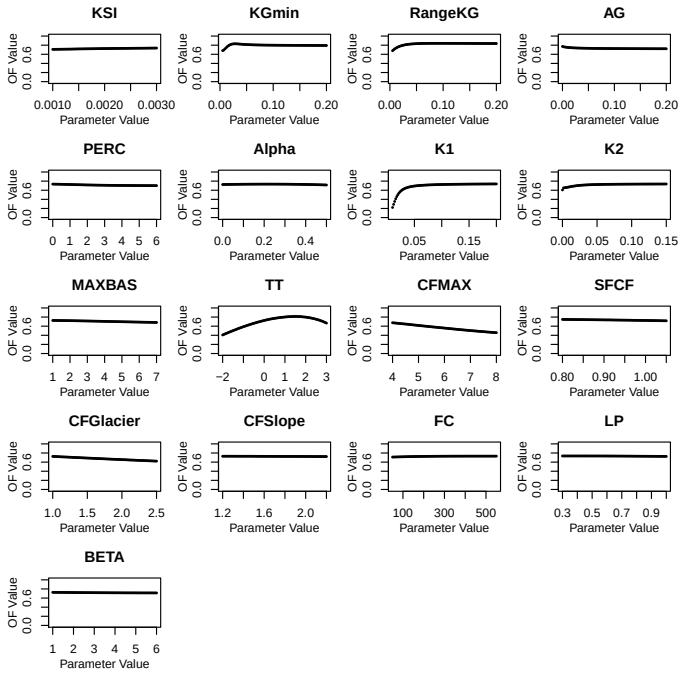


Figure B.10: Sensitivity of R_{Q_2} to one at a time parameter perturbations of selected HBV-light parameters.

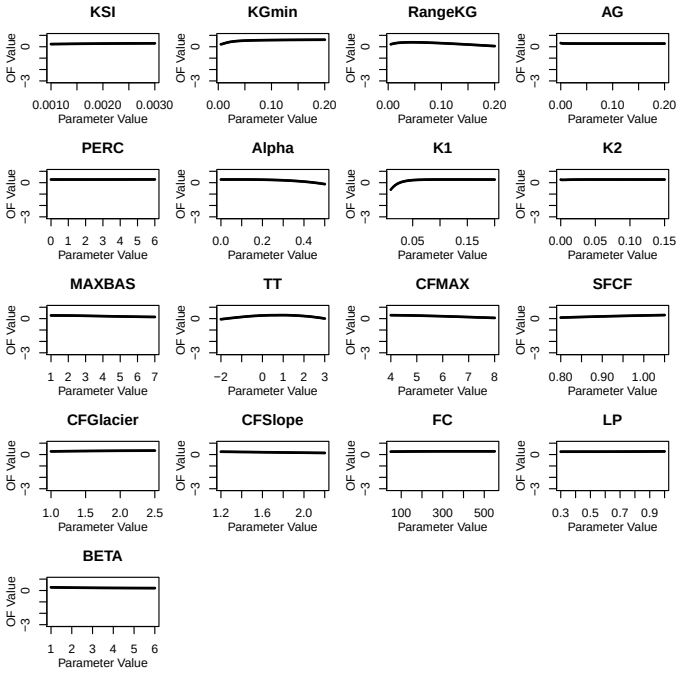


Figure B.11: Sensitivity of R_{Q_3} to one at a time parameter perturbations of selected HBV-light parameters.

Ehrenwörtliche Erklärung

Hiermit erkläre ich, Samuel Mayer (Matr. Nr. 3304294), dass die vorliegende Arbeit selbständig und nur unter Verwendung der angegebenen Hilfsmittel angefertigt wurde.

Ort, Datum

Unterschrift