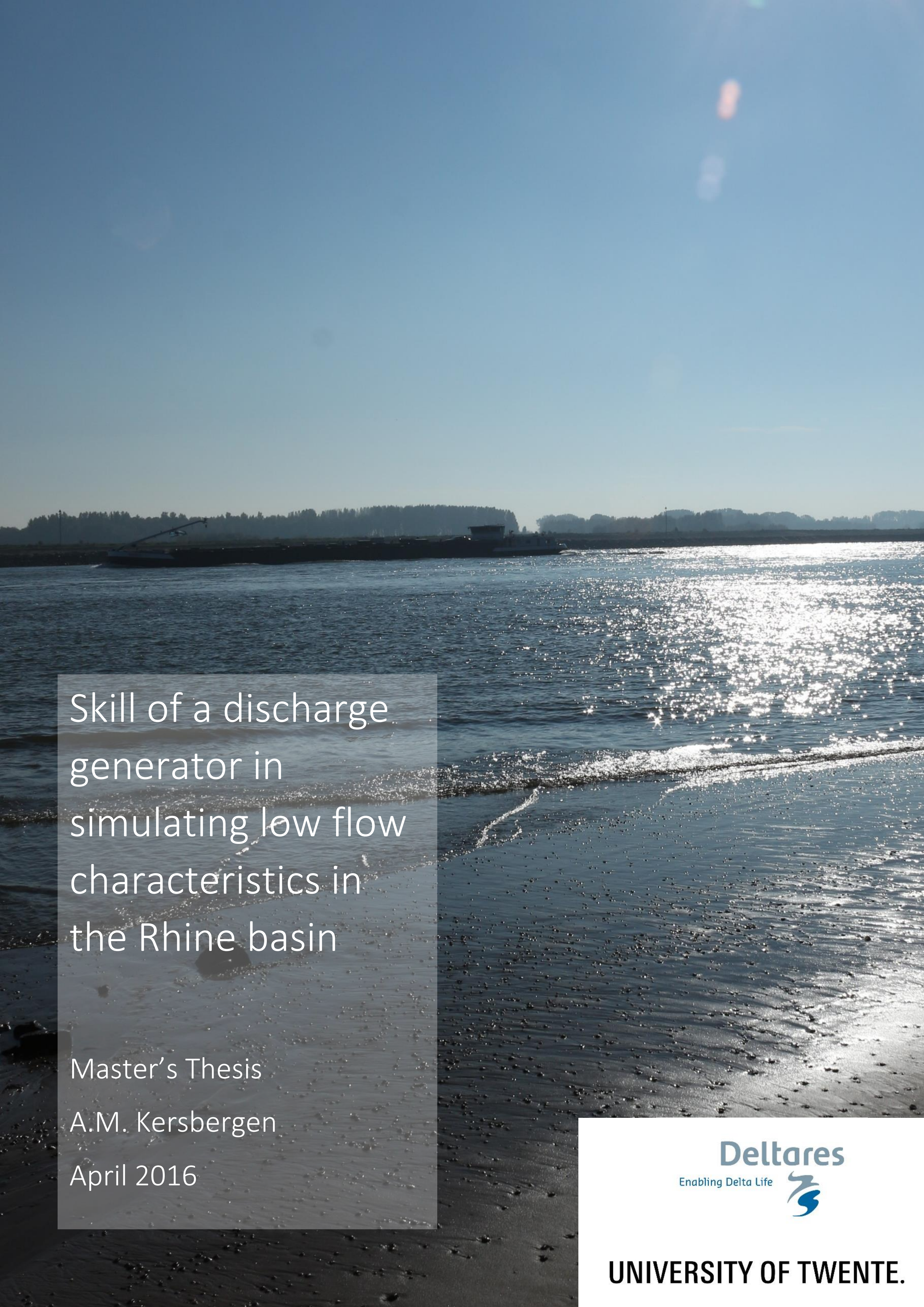


Skill of a discharge generator in simulating low flow characteristics in the Rhine basin

Master's Thesis

A.M. Kersbergen

April 2016

Deltares
Enabling Delta Life

UNIVERSITY OF TWENTE.

Skill of a discharge generator in simulating low flow characteristics in the Rhine basin

Master's Thesis Water Engineering and Management

University of Twente

Faculty of Engineering Technology

Civil Engineering and Management

Author:

Annemarleen Kersbergen BSc
a.m.kersbergen@alumnus.utwente.nl

Graduation Committee:

University of Twente, Department of Water Engineering and Management:

Prof. Dr. J.C.J. Kwadijk

Dr. ir. M.J. Booij

Deltares, Hydrology department:

Ir. M. Hegnauer, daily supervisor

Dr. ir. F. Sperna Weiland, daily supervisor

Enschede, April 2016

Cover photo: Ton Arbouw – Omroep Gelderland 30-10-2015 Laagwater op de Rijn

Summary

Low flows are important to consider in water management as low flows can have societal and economic impact: by e.g., navigation problems, lack of irrigation water for agriculture, salt intrusion, lack of cooling water and bloom of algae. Synthetic time series can be used for low flow frequency analysis and for gaining information about the development and characteristics of low flows. The Generator of Rainfall And Discharge Extremes (GRADE), consisting of a weather generator, a hydrological model and a hydraulic model, has given satisfactory results for simulating peak flows with large return periods in the Rhine basin, and it is expected that there is potential to apply the combination of models also for low flow analysis. In this research the skill of the hydrological model, the skill of the weather generator and the skill of the combination of weather generator and hydrological model for simulating low flows in the Rhine basin are evaluated and a first start is made with the improvement of skill of the hydrological model.

Low flows are defined as discharges under the monthly thresholds determined by the Dutch National Committee of Water allocation (LCW) and split into thresholds in the growing season and thresholds throughout the entire year. For seven discharge locations at the outflow of seven mayor sub-basins in the Rhine (Lobith (Lower Rhine), Andernach (Middle Rhine), Cochem (Moselle), Frankfurt (Main), Rockenau (Neckar), Rekingen (East Alpine sub-basin) and Untersiggenthal (West Alpine sub-basin)) analyses are conducted on the low flow discharges, the low flow events (duration and cumulative discharge deficit below the threshold), lake levels, snow covers, groundwater levels and the meteorology. After the evaluation of the skill has been decided to improve the performance of the hydrological model in part of the East Alpine sub-basin by recalibration. Five parameters, including the snowfall correction factor, have been selected and a Monte Carlo simulation has been performed for four sub-catchments.

The results from the evaluation of the skill show that discharges are mainly underestimated in the historical simulations by the hydrological model. This causes more low flow events and more severe events. In the Alps most underestimation takes place in the summer. The simulation of snow plays a role in this. Although a conceptual hydrological model is used, variations in processes like snow, lake levels and groundwater are captured well. The synthetic series of the weather generator simulates periods of dry weather, but less persistent dry periods (especially in the summer), which makes that less low flows occur and there is a decrease of extreme severities of low flow events compared to the observations, especially for events with the growing season thresholds. In the West Alpine sub-basin the snowfall from the weather generator is less than with observed weather, causing more low flows in summer. Comparing the synthetic simulations with the observations gives a good skill of the model for discharges at Rockenau, return periods of duration at Lobith and Andernach and the return periods of duration and severities in the growing season at the Alpine locations. This skill is however based on the compensation of two errors. The skill of the hydrological model for simulating low flow characteristics has been improved by the recalibration. There is less underestimation of flow and thus there are less false alarms. The performance on the other analyses has improved or stayed the same.

Synthetic weather series are a useful tool in low flow risk assessment, when both the weather generator and the hydrological model give acceptable results. In this study is shown that models made for simulating peak flows are not necessarily acceptable for low flows. By tracing the important processes in the model (in this case snow) and with focus on low flows, improvements in the skill are possible.

Preface

This thesis is the final part of my master Water Engineering and Management. I have studied the Rhine and its low flows for seven months. A year ago I saw a first description of the subject and I was immediately interested. It has not disappointed me; I enjoyed working on this thesis and I have learned a lot during the process. There are some people that I want to thank here for their contributions.

First of all I want to thank my supervisors Jaap Kwadijk and Martijn Booij, who were enthusiastically thinking along with me about this project from the start. Lots of questions were answered by Mark Hegnauer and Frederiek Sperna Weiland and especially lots of data have been gathered for me by them. Other colleagues at Deltares were very helpful as well, providing me with new ideas, and made my time at the Hydrology department very nice. Rest my thanks to the students with whom I shared an office, both in Enschede and Delft. I always could go for a coffee with you or talk about the not always easy graduation process. Special last thanks to my family and friends, who always had a lot of faith in me and cheered me up.

Annemarleen Kersbergen

Delft, 1 April 2016

Content

List of figures

List of tables

1. Introduction

Rivers have an important function in society; the water is used for drinking water, irrigation for agriculture, cooling water for the industry, navigation and recreation. While there is much attention for floods, with the risk of inundations and casualties, also low flows affect the river functions. Low flow can be defined in different ways, but it usually refers to the flow in a river during dry periods of the year (Smakhtin, 2001). Low flows can cause economic damages: discharge deficits can lead to unstable flood defences, salt intrusion, reduction of water quality (higher temperature of water and bloom of algae), agricultural losses and navigation problems (Bolwidt et al., 2006). To address the return periods of extreme events, outside of the available historical observations, synthetic time series can be used. Recently an instrument is made for the Rhine and Meuse river basins that generates synthetic discharge series based on synthetic weather series. It has been evaluated for floods. In this study the topic of interest is the skill of the synthetic series of simulating low flow characteristics in the Rhine.

Low flows

The occurrence of low flows depends on a couple of factors: the meteorology, the storages (groundwater, lakes and snow) and anthropogenic factors. In flood conditions there is usually a large amount of precipitation and the aquifers and the soils are saturated. There is a quick response of the basin to precipitation. Low flow conditions exist when there is a lack of precipitation, resulting in storages that are not recharged. The storages show a recession curve and only the base flow is adding to the stream flow (Smakhtin, 2001). Also melting snow and glaciers and discharge from lakes in the Alps are more important for low flow periods. For the Rhine about 70% from the total discharge comes from the Alps during summer, when the snow that accumulated in the winter months is melting (Middelkoop et al., 1999)). Human influences like groundwater abstractions, river abstractions and regulation of the river flow regime by dams and weirs are directly and indirectly influencing the low flows too. The processes in low flow periods are more complex than in high water periods and are more catchment specific (Gudmundsson et al., 2011).

Low flows in the Netherlands

The years 1949, 1959, 1976 (and 2003) are characteristic low flow years in the Netherlands (see Figure 1). The return periods of these discharge deficits (cumulative volume under the threshold of 1800 m^3/s) are determined (Beersma & Buishand, 2002). In the autumn of 2015 the most recent low flow period on the Rhine occurred. The discharge was below 1500 m^3/s for a long time, and also a large part of the time around 1000 m^3/s (LCW, 2015). This caused low water levels, and problems for navigation. Because the low flow period was outside of the growing season, it had no impact on agriculture. In press releases there were simple statistics made of the inter-event time between this event and a similar one. There is only limited information about the probabilities of low flows that occur or can develop in the future in the Netherlands. One of the reasons that there is less attention for low flows in rivers in the Netherlands is because there is no legal foundation as there exists for floods. There is not a certain low flow event for which there must be protection. The regulation by law for low flows is restricted in the Netherlands to setting priorities in the case of a scarcity of water. When the discharges from the main rivers Rhine and Meuse are under a threshold value the National Committee of Water allocation (LCW) advices on which measures to take. There is a priority list of functions to which the water can be allocated. When problems are foreseen, sluices and weirs in the Dutch delta are closed to provide more water depth for navigation and to prevent salt water intrusion (Bolwidt et al., 2006).

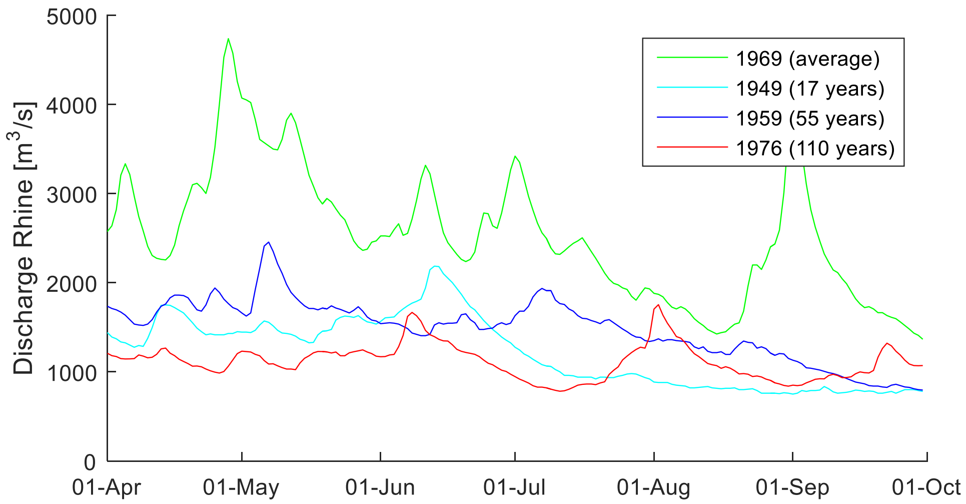


Figure 1 Summer discharges for an average year and three characteristic low flow years, with return periods based on joint probability of precipitation deficit and discharge (based on RIZA et al. (2005)).

Low flow frequency analysis, synthetic series and GRADE

Low flow frequency analysis is regularly used in studies on low flows. Beersma and Buishand (2002) took the discharge deficit (severity) under a fixed threshold as event to study. Others chose the annual minimum discharge (Tallaksen & Van Lanen, 2004), the mean annual minimum flow (over n days) (Du et al., 2015), the pits under thresholds (Önöz & Bayazit, 2002) or the duration of a continuous low flow event under a threshold (Sung & Chung, 2014). A univariate distribution, resampling of discharges, a multivariate distribution and severity-duration-frequency curves are used to connect return periods to low flow characteristics. A disadvantage of these methods is that there is no indication of how the low flow periods develop in time and throughout the basin. There is only a statistical extrapolation of observed discharge (deficits) at one location.

Using synthetic series can be a method to incorporate the spatial and temporal development of low flows and to assign probabilities of occurrences to low flow characteristics. For flood risk assessment in basins in Germany and Italy the influence of short synthetic rainfall events on the discharge is simulated (Brocca et al., 2013; Liersch & Volk, 2008). This is an example of a coupled weather generator and hydrological model. Only short time series are simulated. The benefit of this kind of database of rainfall events and hydrological responses is that no hydrological experience is needed when looking at the events and the impacts. Also for low flow periods synthetic series are used (for evaluating reservoir operation rules, the vulnerability of the water resources system or the return periods of low flow events), but then often the synthetic series is a discharge series (without the weather generator) (Bolgov & Korobkina, 2011; Borgomeo et al., 2015a; Borgomeo et al., 2015b; Salas et al., 2005).

In 2014 the model combination GRADE (a weather generator, a hydrological model and a hydraulic model) has been delivered to address return periods of flood waves (Hegnauer et al., 2014). It is an instrument to estimate design discharges and corresponding flood hydrographs by stochastic simulation of the weather and hydrological and hydrodynamic modelling. It is considered a useful and more realistic approach in addition to the existing high water procedures that exist in the Netherlands, because it is no ‘blind’ extrapolation of observed extremes (ENW, 2015). Also for low flows this can be a useful method. Using a synthetic series for the weather can make that there are lower discharges simulated than observed (something that cannot be achieved by resampling the discharges without extrapolating the current distributions).

HBV and low flows

Several hydrological models of river basins are used for low flow analyses, e.g. forecasting, studying the effects of climate change or effects of land use changes (De Wit et al., 2007; Demirel et al., 2015; Jörg-Hess et al., 2015; Nicolle et al., 2014; Querner et al., 1997; Te Linde et al., 2008). Several of these studies used the HBV model from the Swedish Meteorological and Hydrological Institute (SMHI), the model that is also used in GRADE. The previous mentioned studies conclude that HBV can be used for low flow estimation, but there are several points of attention particularly for low flows. The difficulty that HBV has with the distribution of low flow events in a year (Demirel et al., 2013c; Te Linde et al., 2008) is expected to influence the time series modelling, but the effect on the frequency modelling is uncertain. The lower complexity of HBV in comparison with other models is an advantage in this study, because large time series have to be simulated.

1.1. Objective and research questions

Because of the complexity of low flow mechanisms and the difficulties that studies with HBV to low flows have come across, a systematic evaluation of the model GRADE is necessary to see if the model (the combination of the weather generator, the hydrological model and the hydraulic model) gives realistic representations. Possible improvements can be suggested based on the results of the evaluation. With the results of this study the appropriateness of this model for estimating return periods of low flows is determined. Using a synthetic time series will improve the risk-based approach to low flows, which is already used for floods.

The objectives of this study are (1) to evaluate the skill of GRADE in simulating low flow events by validation on historical data and (2) to indicate whether this skill can be improved. The skill of GRADE can be divided into the skill of the weather generator and the skill of the HBV model (the SOBEK model is not evaluated in this study). The skill means both whether the model can be used for its purpose and whether the model can represent the observed conditions of the area. The results can be used to indicate if GRADE can be used to address the return periods of low flows in the Netherlands.

The following research questions are set up, to achieve the objectives.

1. What is the skill of HBV?
    1.1. What is the performance of HBV in simulating low flow events with observed meteorology?
    1.2. What is the performance of HBV in simulating lake levels, snow cover and groundwater levels?
2. What is the skill of the weather generator?
    2.1. What is the performance of weather generator in simulating weather conditions that cause low flow events?
    2.2. What is the performance of the weather generator in simulating weather conditions in correspondence with observed weather conditions?
3. What is the overall skill of GRADE in simulating low flow events and low flow event characteristics?
4. Which improvements within GRADE can be realised in simulating (characteristics of) low flow events and how large is the impact of the improvements?

1.2. Outline report

The report is structured as follows: in section 2 the study area of the Rhine is presented and the used data is discussed. Here also a small data analysis is done to see the characteristics of the low flows at Lobith and in the sub-basins. In section 3 the methodology is explained, with firstly the used definition of low flows as it is used in the study, secondly the evaluation methods of the performances of HBV and the weather generator and at last the methods of recalibration. In section 4 the results of the model evaluation are shown, and in section 5 the results of the recalibration of the East Alpine basins. The discussion follows in section 6 and in section 7 the conclusions and recommendations can be found.

2. Study area and data

In this section the study area and the data are described. In section 2.1 the Rhine basin is discussed with its sub-basins as used in this study. In section 2.2 the model GRADE is explained and in section 2.3 the data that is used to evaluate the model is presented.

2.1. Rhine basin and sub-basins

The Rhine is in European terms a medium sized river basin. The Rhine basin has an area of 185,000 km^2 (Middelkoop et al., 1999) and stretches out from the Alps in Switzerland to the delta in the Netherlands. In this study the part of the basin upstream of Lobith (160,000 km^2 (Demirel et al., 2013b)) is examined. At Lobith the Rhine flows into the Netherlands and a large part of the Dutch river and flood policy is based on the discharge at this location. Downstream of Lobith the Rhine bifurcates into the Waal, the Nederrijn and the IJssel and finally discharges in the North Sea.

The Rhine can be divided into three major hydrological areas with specific characteristics: the Alpine area, the German Middle Mountain area and the lowland area (Middelkoop et al., 2001). In this study a division of the Rhine basin into seven sub-basins is used (see Figure 2). These are the same basins as used before in the studies of Tongal et al. (2013) and Demirel et al. (2013b). The sub-basins consist of the three important tributaries (Main, Moselle and Neckar), two sides of the Alps (East and West, splitting the Alpenrhein and the Aare) and two sections of the Rhine (Middle and Lower Rhine). This division gives the required level of detail to see differences in performance within the Rhine basin.

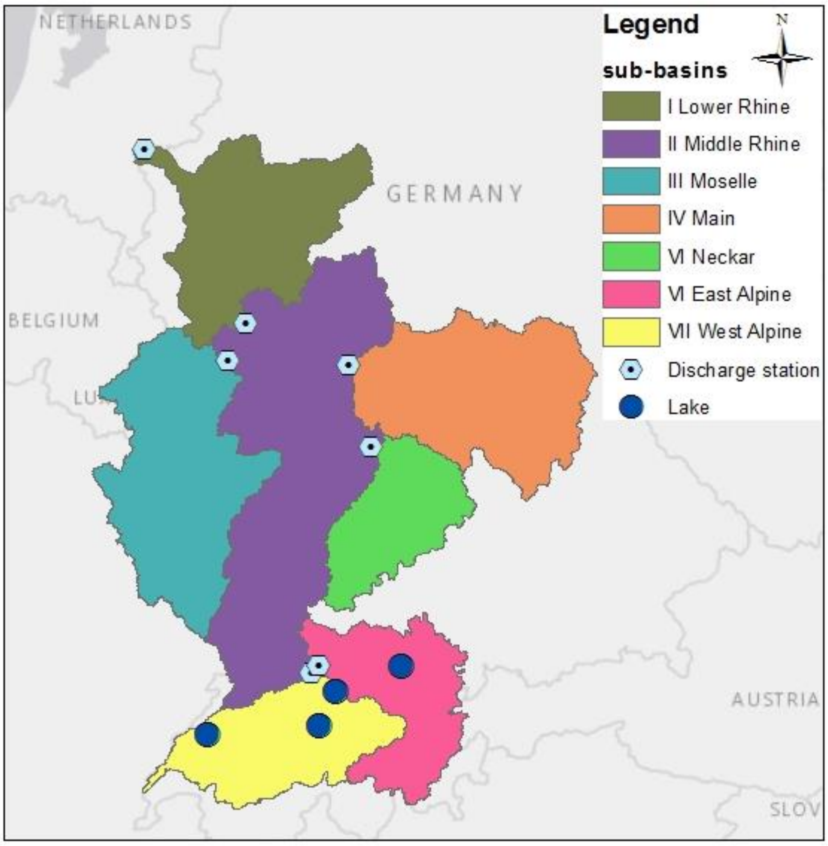


Figure 2 The seven sub-basins of the Rhine with the locations of the discharge stations and the major lakes.

2.1.1. (Low) flows in the Rhine basin

The seven sub-basins of the Rhine area have different regimes. The regimes are seen in Figure 3 where they are characterized as the median discharges per day at the outflow locations of the sub-basins. In the Alps more often precipitation falls as snow, due to the low temperatures at high altitudes. The snow builds up and stays in the mountains during the winter season. When temperatures rise, a lot of the snow melts and flows to the Rhine. In summer months the contribution of flow from the Alpine basins is the highest, but also in the winter the median discharge of the alpine basins is about the same as the other tributaries. The tributaries Moselle, Main and Neckar are rain fed rivers and have the largest discharges in winter and low discharges in summer and autumn, although the differences in median discharge in the Neckar basin between seasons are quite limited. In the Middle Rhine basin the discharges from the upstream sub-basins join. The discharge regime at Andernach is quite similar to that of Lobith. In the Lower Rhine basin the Ruhr, Lippe, Sieg and Erft discharge on the Rhine.

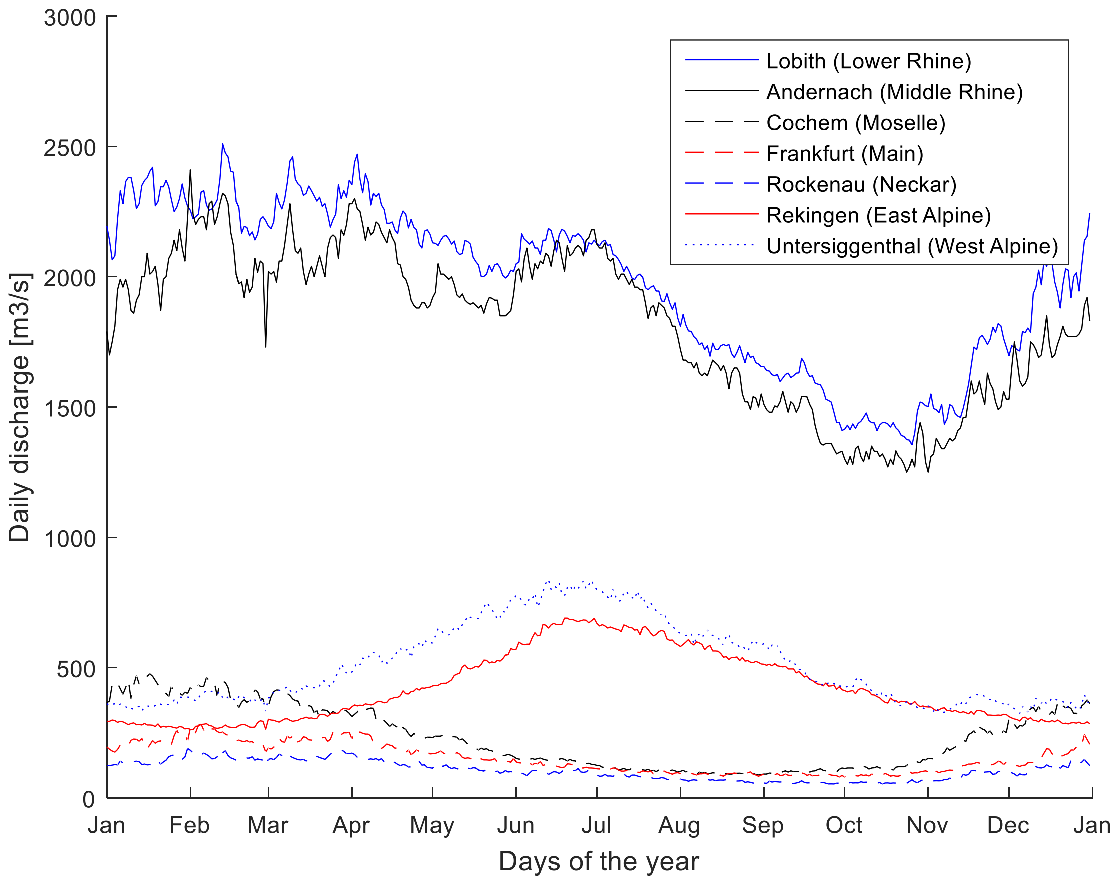


Figure 3 Discharge regimes in the sub-basins of the Rhine. Daily median discharges (Q50) throughout the year.

In Table 1 different low flow characteristics are calculated for the discharge stations of the sub-basins. The characteristics are derived from the whole time series, the time series with a moving average and the flow duration curves based on the whole time series. In a flow duration curve (FDC) the empirical cumulative frequency of discharges is plotted against the percentage of time that the discharge is equalled or exceeded (Tallaksen & Van Lanen, 2004). The values are calculated separately per discharge location.

Table 1 Low flow characteristics for the discharge locations of the sub-basins.

	Lobith (Lower Rhine)	Andernach (Middle Rhine)	Cochem (Moselle)	Frankfurt (Main)	Rockenau (Neckar)	Rekingen (East Alpine sub-basin)	Untersiggenthal (West Alpine sub-basin)
Area (sub-)basin (km^2)	160,087	139,913	27,262	24,833	12,616	16,051	17,678
Derived from time series							
Mean Daily Flow (mm/day)	1.20	1.26	1.00	0.66	0.94	2.38	2.74
Coefficient of variation (SD/Mean) (%)	52%	53%	107%	88%	98%	44%	45%
Absolute Minimum Flow (mm/day)	0.31	0.35	0.03	0.03	0.12	0.66	0.67
Derived from flow duration curve							
Q50 (mm/day)	1.06	1.11	0.64	0.47	0.69	2.17	2.51
Q75 (mm/day)	0.79	0.82	0.36	0.33	0.43	1.60	1.75
Q95 (mm/day)	0.54	0.54	0.20	0.21	0.25	1.05	1.19
Derived from time series with moving averages							
Minimum 7-day low flow (mm/day)	0.34	0.35	0.05	0.06	0.16	0.67	0.75
Minimum 30-day low flow (mm/day)	0.35	0.36	0.08	0.09	0.18	0.76	0.80

Lobith and Andernach have very similar values, as can also be seen before in the discharge regime throughout the year (Figure 3). Lobith has lower values, because the basins of the Ruhr, Lippe, Sieg and Erft add to the area but not much to the discharge. The low flow characteristics of the three tributaries Moselle, Main and Neckar have smaller values than those of the Alpine locations. The Moselle has on average the highest low flow discharge of the tributaries, and has a large variation in discharge. The flow in these tributaries can reduce to only several cubic meters per second and it can stay very low for at least 30 days. The Alpine areas have a more steady contribution to the discharge, with a lower coefficient of variation.

Anthropogenic factors play a significant role in the discharge of the Rhine. Humans always have had a preference to be close to rivers because of their water supply. Anthropogenic impacts in the catchments can affect the flows in both direct and indirect ways. Building a reservoir has a direct effect on the discharge. Especially in the Alps there are a lot of reservoirs built; they have a total storage capacity of 1.9 billion m^3 (about 56 mm) (Belz, 2007). The reservoirs, embankments and channelization in the Alps are mainly built for flood protection and hydropower. When the reservoirs were built, it gave problems to the navigation downstream from Basel. Engineers have put effort into designing a system of dams, wing dams and locks to make the Rhine navigable (Cioc, 2002).

The distances between the discharge locations, the course of the river and the anthropogenic measures determine together the travel time of the water from the discharge locations of the sub-basins to Lobith. In Table 2 these travel times in days can be found. These are average travel times, determined during low flow periods.

Table 2 Travel times in days from discharge locations to Lobith during low flow periods

	Travel time to Lobith [days]
Andernach	2
Cochem	3
Frankfurt	3
Rockenau	4
Rekingen	6
Untersiggenthal	6

2.2. GRADE

The project GRADE (Generator of Rainfall and Discharge Extremes in the Rhine and Meuse Basins) is initiated to develop an alternative to the common practice for determining design discharges for long return periods by extrapolating the measured time series of (yearly) maxima. It consists of three parts: a stochastic weather generator, a conceptual hydrological model and a hydraulic model. The inputs are the historical time series of daily precipitation and temperature per sub-catchment and the output is the discharge at Borgharen (Meuse) and Lobith (Rhine). In Figure 4 this is schematically shown. In this section all the different parts of GRADE are presented as a background of the origin of the model.

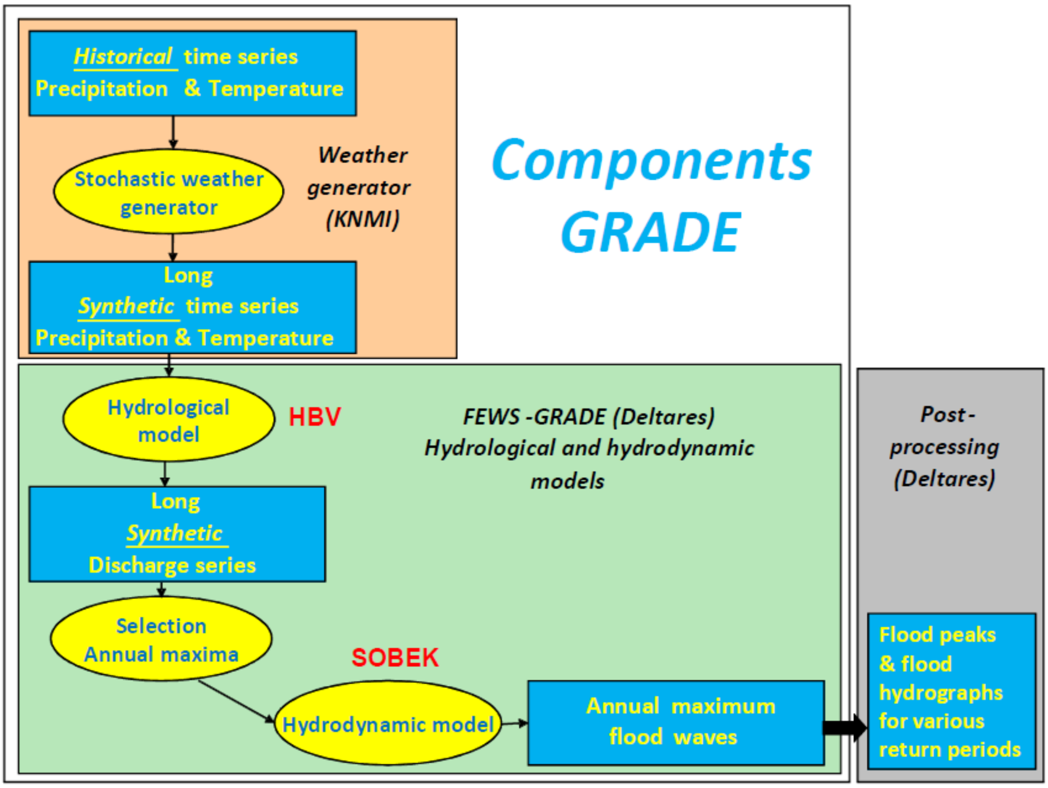


Figure 4 Components of GRADE (Hegnauer et al., 2014)

Weather Generator

The weather generator has been developed by KNMI and generates synthetic time series of daily precipitation and temperature distributed over the basin (using 134 sub-basins) by resampling historical data. The resampling takes place with the nearest-neighbour method. For the Rhine the nearest neighbours are selected as follows (Hegnauer et al., 2014):

The starting day is day *n* with a certain date. Within a window of 61 days around this date there is searched for 10 nearest neighbours (in terms of a weighted Euclidean distance) on the variables:

- Standardized daily temperature, averaged over 134 sub-basins,
- Standardized daily precipitation, averaged over 134 sub-basins,
- The fraction of sub-basins with daily rainfall > 0.3 mm.

Temperature is standardized by subtracting the calendar-day mean and dividing by the calendar-day standard deviation. Precipitation is standardized by dividing by the mean wet-day precipitation amount for that calendar day (with a threshold for wet days of 0.3 mm). Randomly one out of the ten days is selected (a decreasing kernel is used to give more weight to closest neighbours) and then the historical succeeding day is added to the series as day *n+1* (Schmeits et al., 2014). A schematic representation of this (with two variables and only 5 nearest neighbours is given in Figure 5.

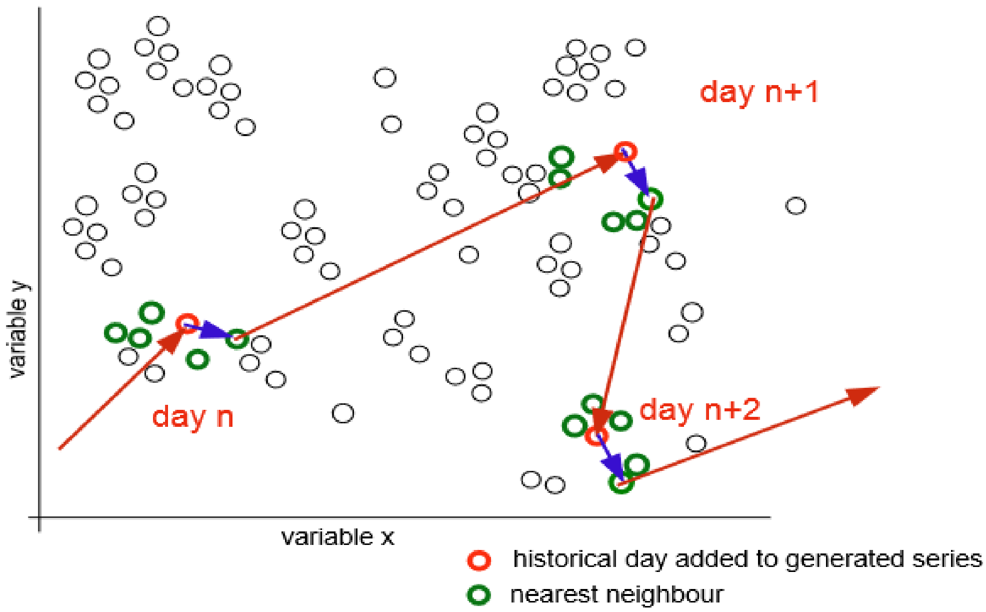


Figure 5 Schematic representation of the nearest-neighbour method, here with two variables. One of the k=5 states (green) which are closest to that of the last sampled day (red) is selected at random (blue arrow), using a decreasing kernel. Its historical successor (red arrow) provides the values for the new simulated day. (Leander & Buishand, 2004)

Hydrological model

For the Rhine basin upstream from Lobith the HBV-96 model is used to convert the precipitation and temperature data into discharges. The model (originally developed at the Swedish Meteorological and Hydrological Institute (SMHI) for runoff simulation and hydrological forecasting) consists of the following routines (Lindstrom et al., 1997), see also Figure 6:

- Precipitation and snow;
- Soil moisture and evapotranspiration;
- Runoff response; for the lower zone (base flow) and the upper zone
- Routing; by a simple version of the Muskingum method.

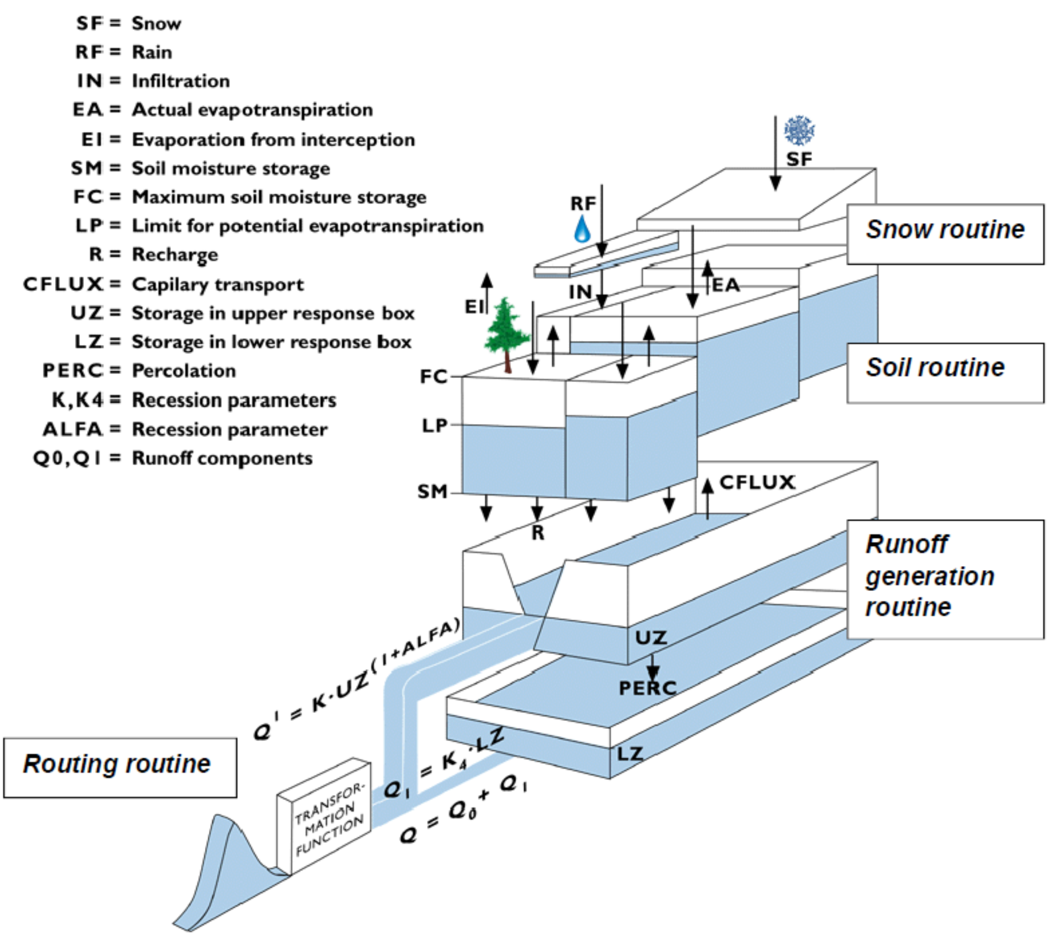


Figure 6 Schematic presentation of the HBV model for one sub-catchment (Hegnauer et al. (2014) after Lindstrom et al. (1997)).

From 1997 on the HBV model is applied in the Rhine basin. With the implementation of HBV in the forecasting system FEWS in 2005, the HBV model of the Rhine was updated and recalibrated. New meteorological data (the data available for forecasting) was used and the objective function (consisting of a weighting of the Nash Sutcliffe efficiency, the Nash Sutcliffe efficiency of logarithmically transformed flows and the Relative Volume Error) focussed on both high and low flows to make also low flow forecasting operational (Berglöv et al., 2009).

The HBV model used in GRADE is an adapted version of the version used in FEWS. For the purpose of GRADE four large lakes in Switzerland are added as sub-basins: Lake Constance, Lake Neuchâtel, Lake Lucerne and Lake Zürich (Hegnauer et al., 2014). Therefore the Rhine is now modelled in 148 sub-basins. The model is calibrated with (a slightly adapted version of) the Generalized Likelihood Uncertainty Estimation (GLUE) method and focus on high flow measures. During the calibration there is not one best set of parameters per sub-basin chosen, but all sets of parameters with a performance above a certain threshold are considered as good and equally likely: the “behavioural sets”. When moving to downstream basins to calibrate, a behavioural set for the more upstream basins is randomly selected next to a random parameter set for the downstream basin. In this way the uncertainty of the parameter sets is taken along downstream. When the whole basin is calibrated in this way, the different combinations of behavioural sets can be tested on, for example, the 1/10 year event. The range of values that is derived here, gives an range of uncertainty. Within GRADE the combinations with 5%, 25%, 50%, 75% and 95% value of the 1/10 year event are presented for fifteen mayor sub-basins of the Rhine (Hegnauer et al., 2014).

Hydraulic model

The hydraulic model in GRADE simulates the discharge wave more accurately for the stretch between Maxau and Pannerdensche Kop. Two models are available: One model with simulation of flooding and one model without simulation of flooding (Hegnauer & Becker, 2013). Retention areas and dike-overtopping are interesting aspects to study for floods. The lower sections of the tributaries Neckar, Main, Nahe, Lahn, Mosel, Sieg, Ruhr and Lippe are also modelled including structures with operation rules. Other tributaries are modelled as lateral inflows (Hegnauer & Becker, 2013). Dike overtopping and discharge waves in the time scale of days are not relevant for low flows. The structures with operation rules could be objects of interest, because they can be used during low flows to retain the water, but are not within the scope of this study.

Post processing

The extreme high discharges from the hydraulic model are post-processed. The annual maximum discharges are selected and ranked in increasing order, to determine the return periods. For return periods larger than 500 years the Weissman fit is used to reduce the effect of random fluctuations in the upper tail of the distribution and to extrapolate the series to return periods of 100,000 years (Hegnauer et al., 2014). Also the post-processing is not within the scope of this study on low flows.

2.3. Data

The data used for this study has different sources. The observed river discharges are from the data set of the Global Runoff Data Centre (2010) and the Federal Office for Environment (BAFU) in Switzerland. The observed lake levels, the snow covers and the groundwater levels are retrieved from information of respectively the BAFU, the Institute for Snow and Avalanche Research SLF and research of Demirel et al. (2013b). The observed meteorology is from the HYRAS 2.0 dataset (Deutscher Wetterdienst (Rauthe et al., 2013)), this is the same dataset as is used by the weather generator. Historical simulated output is simulated with the observed meteorology and the HBV with 50%-parameter set (GRADE reference). The GRADE 4,000 year synthetic simulations are created with the synthetic input of the weather generator and the HBV model with 50%-parameter set. Longer runs of GRADE are available for the discharge, but for this first analysis of low flows the first 4,000 years are adequate for the analyses. The data used for the precipitation and temperature analysis of the weather generator are the 2,000 year time series, the first half of what is used for generating the GRADE output. The metadata of the time series is presented in Table 3.

Table 3 Metadata, all the time series have daily time steps (Q= Discharge, P = Precipitation, T=Temperature).

	Variable	Location	Sub-basin	Start date	End date	Source
Observation	Q	Lobith	Lower Rhine	1-1-1901	31-12-2004	GRDC
	Q	Andernach	Middle Rhine	1-1-1931	31-12-2003	GRDC
	Q	Cochem	Moselle	1-1-1901	31-12-2003	GRDC
	Q	Frankfurt-Osthafen	Main	1-1-1964	31-12-2004	GRDC
	Q	Rockenau	Neckar	1-1-1951	31-12-2003	GRDC
	Q	Rekingen	East Alpine	1-1-1920	31-12-2003	GRDC
	Q	Untersiggenthal	West Alpine	1-1-1935	31-12-2003	GRDC
	Q	Domat/Ems		1-1-1978	31-12-2008	BAFU
	Q	Diepoldsau		1-1-1978	31-12-2008	BAFU
	Q	Neuhausen - Flurlingerbrücke		1-1-1978	31-12-2008	BAFU
	Lake level	4 lakes				BAFU
	Snow cover	Alpine basins		19-6-2002	3-10-2011	SLF
	Groundwater	7 sub-basins				Demirel et al. (2013b)
	P, T	134 sub-basins		1-1-1961	31-12-2007	HYRAS 2.0
Historical simulation	Q	7 discharge locations		3-1-1951	31-12-2006	GRADE reference
	Lake levels	4 lakes		3-1-1951	31-12-2006	GRADE reference
	Snow cover, ground water	148 sub-basins		3-1-1951	31-12-2006	GRADE reference
Synthetic simulation	Q	7 discharge locations				GRADE 4,000 y
	Lake levels	4 lakes				GRADE 4,000 y
	Snow cover, ground water	148 sub-basins				GRADE 4,000 y
	P,T,	134 sub-basins				Weather generator 2,000 y

3. Methodology

In this section the steps towards the objective of the study are explained in more detail. The first step is to select a valid definition of low flows, that gives useful information when it is used in the context of GRADE in the Netherlands (Section 3.1). The second step is to see how well the models simulate the low flow characteristics that follow from the low flow definition, so in section 3.2 several methods are explained to compare and value the historical and synthetic time series (related to research questions 1, 2 and 3). The third step is to improve GRADE for low flow simulations (related to research question 4), consisting of a recalibration of the HBV model, and check the skill of the recalibrated model after the improvement in the same way as in the second step. The methodology of the improvement is covered in section 3.3.

3.1. Definition of low flows

3.1.1. Low flows in the Rhine

In section 2 a general introduction is given on the low flow conditions of the Rhine and its sub-basins. However, these characteristics do not incorporate the relation between water demand and low flows. As is described in the book of Tallaksen and Van Lanen (2004) and in the study of Jörg-Hess et al. (2015), the threshold level method (introduced by Yevjevich (1967)) can be used to derive low flow events. An event is characterised by the duration and the severity (total deficit) as in Figure 7. The day that the discharge falls below the threshold is the start of a low flow event, and when it rises above the threshold again the event ends. This is an appropriate way to look at situations when low flow is causing damage to the society similar to what Sung and Chung (2014) did in their study. Threshold values have to be chosen which refer to a level when problems occur due to low flow. In the Netherlands these thresholds for the discharge at Lobith are defined by the National Committee of Water allocation (LCW) (see Table 4). When the discharge is below the threshold and it is expected to stay low for at least three days the committee determines which measures are needed to mitigate the damage (RIZA et al., 2005).

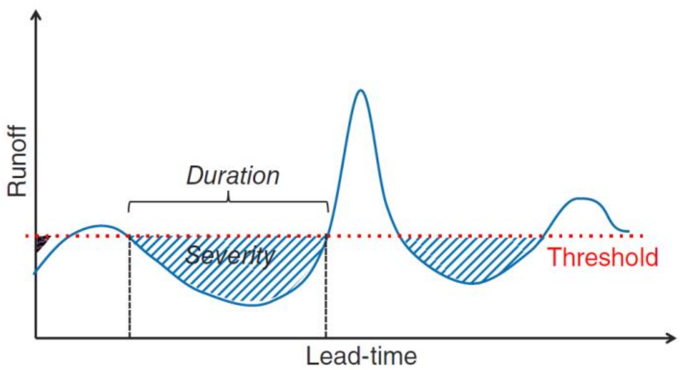


Figure 7 Low flow definition using a threshold to calculate duration and severity of the low flow event (Jörg-Hess et al., 2015).

Table 4 LCW thresholds for low flows in the Rhine at Lobith (RIZA et al., 2005).

Month	Low flow threshold Discharge at Lobith (m^3/s)
May	1400
June	1300
July	1200
August	1100
September- April	1000

The LCW thresholds are higher in the growing season (May-August) than in the rest of the year. In the growing season low flow can cause shortages in irrigation of crops, while in the rest of the year this is not the case and low flow mainly affects navigation, industries and dike stability and increases salt water intrusion. To separate these two aspects, the LWC thresholds are evaluated on different thresholds. For the growing season only the months May-August are evaluated and the LCW thresholds of these months are used. For navigation a threshold of 1000 m^3/s is used all year round.

The LCW thresholds are only appropriate for Lobith. To be able to indicate low flow events at the discharge stations of the sub-basins a conversion of the LCW thresholds is made. First the threshold discharges at Lobith are converted per month into exceedance percentages of discharges in that month, using the flow duration curve (explained in section 2.1). Then also per month a flow duration curve is made for the other discharge locations and the discharge corresponding with the exceedance percentage of the LCW threshold is selected. This results in the thresholds presented in Table 5, and graphically in Figure 8. The terms 'growing season threshold' and 'navigation threshold' refer to the LCW thresholds at Lobith. It does not mean that these values necessarily correspond with navigation or agriculture in the sub-basins itself.

Table 5 Threshold values for low flows. Thresholds at Lobith with corresponding exceedance percentages and threshold values at other discharge locations.

Navigation thresholds [m^3/s]								
	Lobith		**Andernach**	**Cochem**	**Frankfurt**	**Rockenau**	**Rekingen**	**Untersiggenthal**
Jan	1000	Q96	799	118	63	40	159	215
Feb	1000	Q97	823	124	83	44	153	220
Mar	1000	Q98	969	128	73	59	149	218
Apr	1000	Q99	986	79	75	47	200	286
May	1000	Q99	966	65	58	44	243	348
Jun	1000	Q99	903	46	25	34	327	376
Jul	1000	Q98	863	45	28	33	304	360
Aug	1000	Q96	855	45	36	31	297	359
Sep	1000	Q91	867	52	52	35	286	300
Oct	1000	Q87	858	69	63	37	248	260
Nov	1000	Q87	868	83	72	40	217	249
Dec	1000	Q93	843	108	77	42	180	228
Growing season thresholds [m^3/s]								
May	1400	Q89	1270	104	86	62	350	472
Jun	1300	Q97	1170	65	57	45	389	465
Jul	1200	Q96	1070	54	48	38	360	422
Aug	1100	Q92	966	52	46	34	323	400

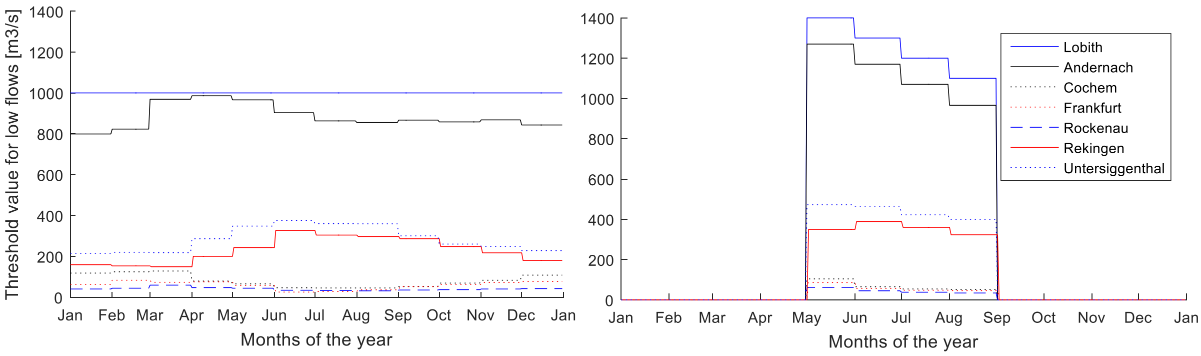


Figure 8 Converted thresholds for all discharge locations. On the left the navigation thresholds, on the right the growing season thresholds.

The thresholds define whether there is a low flow event. These low flow periods are pooled together when there are only small exceedances of the threshold within a period of low flow. The pooled events give more valuable information to the users: With a few days of higher discharges the ships (depending on the duration of their trip and the planning) may not be fully loaded again yet and the benefit that agriculture has with only a small period discharges above the low flow threshold is marginal. Therefore low flow events are pooled when the inter-event time is smaller than a week, as also used by Zelenhasic and Salvai (1987) and Woo and Tarhule (1994). The duration of the pooled event is the total duration from the start of the first event until the end of the last event. The pooled severity is the sum of the severities of the pooled events.

The inter-event volume is not taken into account; events are pooled based on only the inter-event time and when events are pooled only the volumes under the threshold are added up to the severity. In this way pooled events are longer events with a small severity. The longer duration represents the time that the discharge varies around the low flow threshold. The severity stays the cumulative discharge deficit under the threshold. When the inter-event volume would be taken as a criterion, there would be more events, but less events with a hit in the simulations. Therefore is chosen to only base the pooling on the inter-event time.

3.2. Evaluation of GRADE for low flows

Based on the previous explained definition of low flows, performance criteria and graphical tests are set up to compare the simulations with the observations. A distinction has been made between observed series/events, historical simulated series/events (based on historic input) and synthetic simulated series/events (based on the weather generator input). Comparing the observations with the historical simulations gives the skill of the HBV model, comparing the historical simulations with synthetic simulations gives the skill of the weather generator and comparing the observations with the synthetic simulations gives the skill of the combination of weather generator and hydrological model.

The historical simulations can be evaluated as time series because they cover the same period as the observations. But they will also be evaluated on statistics. The synthetic simulations can only be compared with the observations based on statistics; the weather input is different so the exact timing cannot be compared. The evaluation is done on different variables (discharge, snow cover, lake level, groundwater, precipitation and temperature) and on different locations (entire basin/Lobith, sub-basins/discharge locations, lakes).

3.2.1. Discharges

To correctly model the low flows the discharge value of low flows itself should agree, but also the persistence in this low value. For a first look in the skill of the models the discharges for the different discharge stations are compared with flow duration curves (see section 0). For low flows the lower end of the curve is of specific interest. In studies of Demirel et al. (2013a) and Görgen et al. (2010) on the Rhine, low flows were indicated as respectively Q75 and Q90. The low flow exceedances (indices) can be read from the FDC. Time series with a different length can also be compared.

In the flow duration curve there is no information about subsequent days of high or low flow. To say something about persistence in the time series the autocorrelation is used. A method used by Wilks (2006) is to take the lag-1 autocorrelation as measure for persistence in weather. Discharges in the Rhine usually have large persistence, so this measure would result in all values being very close to one. Therefore the correlation coefficients of the lags are summed for the lags where the correlation is still significant. This correlation length is a measure of the shape of the correlogram. The correlation length is calculated for the different discharge stations and the different time series. The observed and simulated time series are evaluated for the same time period. The limit for when the autocorrelation is still significant is calculated with equation 1 (Anderson, 1976) for lag k. With N is the length of the time series and r_i is the autocorrelation on lag i.

$$-1/N \pm 1.96\sqrt{\frac{1}{N} * (1 + 2 * \sum {r_i}^2)} \text{ with } i = 1..(k-1) \tag{1}$$

This results in a general comparison of observed and simulated discharges and simulated and GRADE discharges. It gives information about bias (structural under- or overestimation) and persistence (serial correlation).

3.2.2. Low flow events

Low flow events are defined as in section 3.1. The LCW thresholds are used to derive events in all three time series. These events are compared on duration and severity. The moment of occurrence (timing) of historical simulated events is compared with the observed events. For events at Lobith also the contribution of sub-basins to the low flows can be summarized.

Matching low flow events

For the evaluation of the performance of the HBV model on low flow events, the time series for the seven discharge locations for both observed and simulated are compared with the low flow threshold values and a list of observed events and a list of simulated events are made. With perfect correspondence all the observed events would be found in the simulations, no extra events are simulated and the duration and severity of the events are the same.

To check this, first the events from the observations and the simulations have to be matched (both pooled). When the start or end date of an event is within a window of 11 days around the start or end date of an event in the other time series, the events are matched. Also small observed events that fall totally within the timespan of a large simulated event are matched. When multiple events fall within the criteria, only the one with the largest duration is matched. The others are considered without match and are thus misses or false alarms. The matched events are hits (when an event is simulated that also is observed), misses are events that are observed but not simulated and false alarms are events that are simulated but not observed.

Now contingency tables can be made of the observed and simulated events. The hits, false alarms and the misses are the variables of interest and are used to show if the model is able to simulate the occurrence of events. The following indices can be calculated (Wilks, 2006):

$$Critical\ Score\ Index\ (CSI) = \frac{hits}{hits + false\ alarms + misses} \tag{2}$$

$$False\ Alarm\ Ratio\ (FAR) = \frac{false\ alarms}{\#\ simulated\ events} = \frac{false\ alarms}{hits + false\ alarms} \tag{3}$$

$$Hit\ Rate = \frac{hits}{\#\ observed\ events} = \frac{hits}{hits + misses} \tag{4}$$

All three measures can vary between 0 and 1. A CSI of 1 means a perfect match. The more false alarms and misses there are, the lower the CSI gets. The false alarm ratio gives the number of false alarms over the number of simulated events. The optimal value is 0. The hit rate gives the number of hits over the number of observed events. The optimal value is 1.

Next to the occurrence of events, the characteristics of events are of interest. Of the matched events (the hits) two measures are used to quantify the performance of simulation of duration and severity: the Mean Absolute Relative Error (MARE) and the coefficient of determination R^2.

Mean Absolute Relative Error (Staudinger et al., 2011)[1]

$$MARE = \frac{1}{n}\sum_{i=1}^{n}\frac{|O_i - S_i|}{O_i}$$

(5)

Coefficient of determination (Krause et al., 2005)

$$R^2 = \left(\frac{\sum_{i=1}^{n}(O_i - \bar{O}_i)(S_i - \bar{S}_i)}{\sqrt{\sum_{i=1}^{n}(O_i - \bar{O}_i)^2}\sqrt{\sum_{i=1}^{n}(S_i - \bar{S}_i)^2}}\right)^2$$

(6)

The MARE gives the absolute error of the simulates values over the observed value. It thus gives more weight to errors in small observed events than in large ones. This is a measure to quantify the error. To see if the error is due to a bias or is more random, the coefficient of determination is used. The coefficient of determination (also regarded as performance measure for discharges in Krause et al. (2005); Pushpalatha et al. (2012) and Crochemore et al. (2015)) indicates the linearity of the relation between observed and simulated. It gives a value between 0 and 1, where 1 is perfect linearity. Linearity does not mean perfect correspondence between observed and simulated properties, it can also indicate a bias.

Measures for the timing bias are the difference in start day and the difference in end day of the historical simulated events in comparison with the observed events.

Number of events

Without looking at matches or hits, the average number of events per (calendar) year with the growing season thresholds and the navigation thresholds is used to compare the occurrence of events in the three time series. For the navigation threshold events are evaluated for the whole year, the winter half year (with a start in October to March) and the summer half year (events with a start in April to September). With the distinction between seasons, information is obtained for low flows originating from different processes.

Return periods of low flow events

By plotting the return periods of severities and durations, the performances of the hit events and the number of events *N* per year are combined in one analysis. From the three time series (observed, historical simulated and synthetic simulated) the low flow events are derived. The characteristics severity and duration are sorted and assigned a probability based on their rank *r* and the Gringorten formula (Shaw et al., 2010):

$$P(X) = \frac{r - 0.44}{N + 0.12}$$

(7)

[1] O = observed value, S = simulated value, $\bar{O}$ = mean value, n = length of series

$$P_{year}(X) = P(X) * \frac{N}{t}$$

(8)

$$T(X) = \frac{1}{P_{year}(X)}$$

(9)

The return period T is calculated with the probability P and the average number of events per year (number of events N divided by length time period t). The return periods are plotted on logarithmic scale, to be able to examine them better visually.

Contribution of sub-basins during low flow events

It is good to place the results from the analyses above in perspective for the low flow events at Lobith by analysing which sub-basins contribute much to the low flow at Lobith and which do not. For evaluating the contribution of sub-basins the period of the low flow event at Lobith is taken, plus and minus six days (this is the average travel time from the Alps, flow can go both slower and faster). Travel times from the sub-basins to Lobith (see Table 2) are also taken into account. For every time frame the total amount of m^3 of discharge is calculated, at Lobith and at the other discharge locations. The ratio of the total discharge of the discharge location and the total discharge at Lobith is the contribution of the basin. The contributions of the basins Moselle, Main, Neckar, East Alpine and West Alpine are making up a large part of the discharge at Lobith. The remaining amount of discharge is then from the Middle and Lower Rhine basin. The distance between Andernach and Lobith is too short and the variation in the travel time between the two locations is too large, to make a separation between the Middle Rhine and Lower Rhine, especially when considering the daily time step in the data.

By comparing the relative contributions of the sub-basins in low flow events in the observed, historical simulated and synthetic simulated time series the performance of HBV and the weather generator can be tested again. In this way, it is evaluated whether the discharge during low flow events originates from the same basins as in the observations.

3.2.3. Lakes

The evaluation of lakes in this study focuses on the large lakes in Switzerland: Lake Constance, Lake Neuchâtel, Lake Lucerne and Lake Zurich. The lake levels from the HBV model compare well to the observations because they are derived from a volume/water level relation. Therefore the performance of lake levels from the simulations can be evaluated with the Nash-Sutcliffe Efficiency and the Mean Absolute Error (MAE). The MAE is used here, because the values itself can already be compared with the other lakes, so no relative error is needed. The MAE is evaluated for the entire year, and for the summer and winter half year.

Nash-Sutcliffe efficiency (Nash & Sutcliffe, 1970)

$$NSE = 1 - \frac{\sum_{i=1}^{n}(O_i - S_i)^2}{\sum_{i=1}^{n}(O_i - \bar{O}_i)^2} \tag{10}$$

With perfect correspondence the Nash-Sutcliffe efficiency is 1. A high NSE is also retrieved when the error in the simulations are smaller than the variation in the observations. The simulations are compared to a benchmark model; the mean and the variability of the observations (Shaw et al., 2010).

Mean Absolute Error (Crochemore et al., 2015)

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|O_i - S_i| \tag{11}$$

Lake levels from observations, historical simulations and synthetic simulations are evaluated with the same method as the discharges. In this case level duration curves are made and the lower end is evaluated.

3.2.4. Snow cover

Observed snow cover data is available from stations on several locations in the East Alpine and West Alpine basins. To compare the snow station data with HBV output, only the data from basins where also a measurement station is, are taken into account. These basins are listed in Table 6 and shown in Figure 9.

Table 6 HBV basin names with number of snow stations with available observation snow series.

	HBV basin name	Number of snow stations with observations		HBV basin name	Number of snow stations with observations
EA	Thur	2	**WA**	Limmat_Reuss	1
EA	Rhein1	9	**WA**	KleineEmme	1
EA	Rhein2	3	**WA**	Thunersee	7
			WA	Aare1	3
			WA	Sihlzuer	1
			WA	Lintwees	1
			WA	Lintmoll	2
			WA	Muotinge	2
			WA	Reusluze	1
			WA	Engebuoc	1
			WA	Reusseed	3

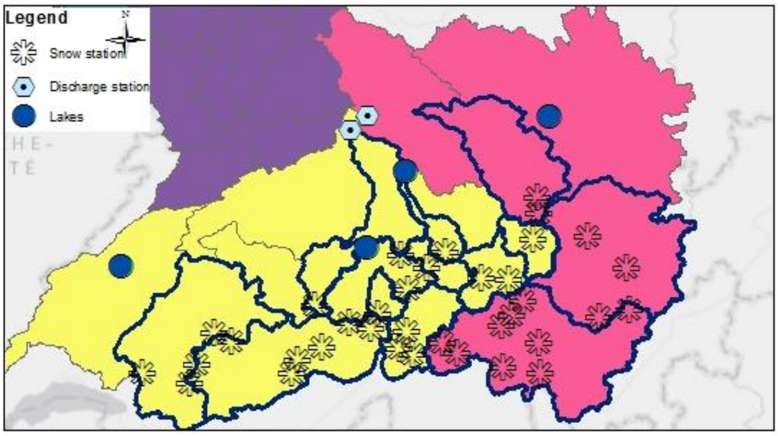


Figure 9 Snow stations in the Alps and HBV basins taken into account for snow (bold outlined).

In HBV the snow cover is the snow water equivalent and to compare it with the measured snow cover in the observations there should be more information on the composition of the snow cover. This information is not available for all time steps, and therefore an index is used rather than the raw snow cover data. The snow cover index is an index for the entire East Alpine or West Alpine basin in which the snow cover data is standardised. Standardisation is done in the same way as Demirel et al. (2013b) did: From every station or HBV basin the mean and the standard deviation are determined. Every time step is standardised by subtracting the mean and dividing by the standard deviation. The average per time step of all the snow stations or the HBV basins within the East Alpine or West Alpine sub-basin is the snow cover index.

The snow cover indices of observed and simulated time series are compared with a correlation test. The correlation is high when the timing of the snow cover is correctly simulated. To see the bias in timing also the mean number of days that the start, maximum and end of the snow cover are shown. The snow cover is considered to exist when the snow cover index is above a certain value. This value can be read from the snow index duration curve and is the number where the graph flattens out (to the value of zero snow or the cover in summer). This point is determined for the observations at the value that is exceeded 49% of the time and for the simulations 69% of the time (because of the difference in snow cover duration between observed and historical simulated snow cover). The hydrological year that is used for the snow cover runs from the 1st of August to the 31st of July. The correlation of the maximum cover per year shows how well the yearly variation in snow cover is captured by the model.

Possible differences occurring due to the weather generator are detected by the duration curves made for the yearly maximum snow cover and the duration of the snow cover. Because both series are HBV output they are not standardized. The average over the basins is taken. Because of the snow accumulation over the years in some HBV sub-catchments the value for which the snow cover starts and ends is determined in a different way. The minimum value is calculated for every year, and the snow cover starts and ends when it has a level of 20 mm above this yearly minimum, so the first small snow covers are neglected.

3.2.5. Groundwater

Groundwater is an important contribution to low flows (Smakhtin, 2001). Comparing groundwater levels from observations with a ground water related variable in HBV (the volume of the Lower Zone) gives an indication of the skill of HBV to model groundwater processes.

The observations of groundwater have been gathered in the study of Demirel et al. (2013b) from different sources and consisted of point information with different data lengths and temporal resolutions. The groundwater stations with observations are not evenly distributed over the Rhine basin. By standardization (subtracting the mean and dividing by the standard deviation per station), interpolation and averaging over the sub-basins a series of daily groundwater indices per basin has been derived (see Demirel et al. (2013b) for the complete pre-processing method). For the West Alpine the length of the time series is too short to compare with simulated results. Historical groundwater simulations are standardized in the same way as the snow data.

Correlation coefficients are used to compare groundwater indices from observations and simulations and also to see the relation of observed groundwater index and discharge and the relation of simulated groundwater index and discharge. This gives an indication whether the groundwater processes that are observed are also simulated.

Again duration curves are used to look at the influence of the weather generator. Here the groundwater content in mm (lower zone variable) is used.

3.2.6. Meteorology

The influence of the weather generator is evaluated through the output of the HBV model, but the input data itself (synthetic series of precipitation and temperature) is also evaluated and compared with the observed meteorological series.

Precipitation and temperature values are available per HBV sub-catchment and are averaged per day over the seven sub-basins and over the entire Rhine basin. In the previous studies on the weather generator principally the average precipitation of the entire basin is taken into account (Schmeits et al., 2014). The characteristics that are studied within the context of low flows are the annual 30 days minimum sum and the annual 120 days minimum sum of precipitation. This is done for the entire year (January –December), the summer half year (April – September) and the winter half year (October – March). With the calculations of the 30 and 120 days sum, the 30 or 120 days before the start of the year and seasons are also taken into account to calculate the sum of the past 120 days from the start day on. The annual values are sorted and with the Gringorten formula (see section 3.2.2) the probability and the return period are determined.

For temperature especially the days with frost in the Alpine basins are of interest, because of the snowfall. An average temperature will not exactly capture this characteristic. Therefore the temperature is evaluated with the number of days per year with days below zero degree Celsius. The same hydrological year as in the snow evaluation is chosen: from the 1st of August to the 31st of July.

3.3. Recalibration

To improve part of the model (research question 4) and indicate how recalibration could improve the skill for low flows part of the HBV model is recalibrated. This section gives the methodology of the recalibration. In section 3.3.1 the sensitivity analysis is explained. In section 3.3.2 the calibration procedure is can be found and section 3.3.3 covers the validation. In this section only the methods are explained, the specific information about the set-up of the recalibration can be found in section 5.2.

3.3.1. Sensitivity analysis

In previous studies of HBV models in the Rhine several calibration parameters were used, listed in Table 7. From these parameters several are selected to perform a sensitivity analysis on: Threshold temperature (Tt), melting factor (cfmax), snowfall correction factor (Sfcf), maximum soil moisture storage (fc), limit for potential evapotranspiration (lp), soil parameter (beta), high flow recession parameter at high flow HQ (khq), recession coefficient lower reservoir (k1) and the percolation (perc). These are the parameters for the snow melt and accumulation, soil moisture and hydrograph shape and are mentioned in the manual of HBV to use in the calibration of the model (SMHI, 2008).

Table 7 Calibration parameters for different studies of HBV models in the Rhine.

Parameter			**Te Linde et al. (2008)**	**Demirel et al. (2013a)**	**Hegnauer et al. (2014)**	**Berglöv et al. (2009)**	**Davids et al. (2015)**
Focus on			General	Low flow	Floods	General	Low flow
RFCF	Rainfall correction factor	-				X	
SFCF	Snowfall correction factor	-				X	
cfmax	Degree day factor melt	Mm/day/°C			X	X	
Tt	Temperature limit for snow/rain	°C			X	X	X
khq	Recession parameter at high flow level hq	1/day	X		X	X	X
K1	Recession coefficient lower reservoir	1/day		X		X	
perc	Percolation from upper to lower reservoir	Mm/day		X	X	X	X
beta	Soil parameter, controls contribution to the response function	-		X	X	X	X
maxbas	Time base of the triangular distribution of the transformation function	day				X	
fc	Maximum soil moisture content	mm	X	X	X		X
Alfa	Measure of non-linearity	-		X	X		
lp	Limit for potential evapotranspiration (fraction)	-		X	X		X
Cflux	Maximum capillary flow	Mm/day		X			X
Kf	Recession coefficient for quick flow reservoir	1/day		X			

First a sensitivity analysis is performed to see which parameters have the most influence on the discharges. This is done by a univariate sensitivity analysis where the parameters were changed from the current value to lower and higher values (10% and 20%) separately and the effect on the performance measure were evaluated. The parameters with the largest influence on the performance are chosen to incorporate in the recalibration, all the other parameters (also the ones not incorporated in the sensitivity analysis) are fixed on the value used in the current configuration of HBV in GRADE.

3.3.2. Calibration procedure

The HBV-catchments are calibrated one by one, so there are more discharge locations needed with time series observed and simulated discharge. The available time series is divided in two, to have a time series for the calibration and the other for validation. The calibration period is from 1-1-1978 to 31-12-1991. Recalibration is done with a Monte Carlo simulation. The HBV model is run with different parameter sets, five thousand for each individual basin. The sets are varied by selecting random values of the parameters from a uniform distribution between a minimum and a maximum value of the parameter. This setup is the same as used by (Hegnauer et al., 2014) for the calibration of HBV in GRADE.

In the analysis of the current model a lot of performance measures are used for the discharges, lake levels, snow covers and groundwater levels (see section 3.2). In the recalibration there are a many runs and discharges are evaluated at other discharge locations than in the previous analysis. During the recalibration more simple ways of calculating performance of the simulated discharge that HBV calculated based on the selected set of parameters compared with a time series of observed discharges are used: the Nash Sutcliffe Efficiency (NSE), the Nash Sutcliffe Efficiency for inverse flows, the Relative Volume Error (RVE) and the Relative Volume Error in Spring (March – July) are all used to indicate performance for each parameter set. The NSE, the NSE on inverse flows and the RVE are criteria that are also used in other calibrations of the Rhine (Berglöv et al., 2009; Davids et al., 2015; Demirel et al., 2013a; Hegnauer et al., 2014; Te Linde et al., 2008). The NSE is already explained in section 3.2.3. The NSE on inverse flows (NSE_i) does not show sensitivity to high-flow values and focusses on the lowest 20% (Pushpalatha et al., 2012). The RVE is comparable with the Mean Absolute Error (section 3.2.3) as it addresses the difference between simulated and observed discharges. The Relative Volume Error normalizes the error by dividing it by the observed discharges. It averages the error over the whole time series, but is favourable because of the high fluctuation in the observed discharges at the Alpine locations (short peaks of over- and underestimation are levelled out). The NSE has a tendency to emphasize the high flows and therefore is less relevant to this study. It is incorporated in the analysis to check the high flow performance and compare with other calibration studies of the Rhine.

<u>Relative Volume Error (Booij, 2005)</u>

$$RVE = \frac{\sum_{i=1}^{n}(S_i - O_i)}{\sum_{i=1}^{n} O_i} \tag{12}$$

<u>Nash-Sutcliffe efficiency on inverse flows (Pushpalatha et al., 2012)</u>

$$NSE_i = 1 - \frac{\sum_{i=1}^{n}\left(\frac{1}{O_i} - \frac{1}{S_i}\right)^2}{\sum_{i=1}^{n}\left(\frac{1}{O_i} - \overline{\frac{1}{O_i}}\right)^2} \tag{13}$$

All four performance measures are calculated for the five thousand runs. From all these runs the sets with the best performance are selected as behavioural sets (see also section 2.2). The best performance is determined with a bounded Pareto front. The Pareto front are the sets with values for the performance criteria that have such “optimal” set that none of the performance criteria can be improved without decreasing at least one of the other performance criteria (explained and also used in multi-objective optimization of a hydrological model by Guo et al. (2014)). Because in the Pareto front also very low values of (one of) the measures can occur, threshold values are used to bound the Pareto optimal functions. The NSE is only used to calculate the Pareto front, but not for bounding the values. Boundaries will be chosen such that there are at least 10 behavioural sets. An illustrative explanation of the bounded Pareto front is shown in Figure 10.

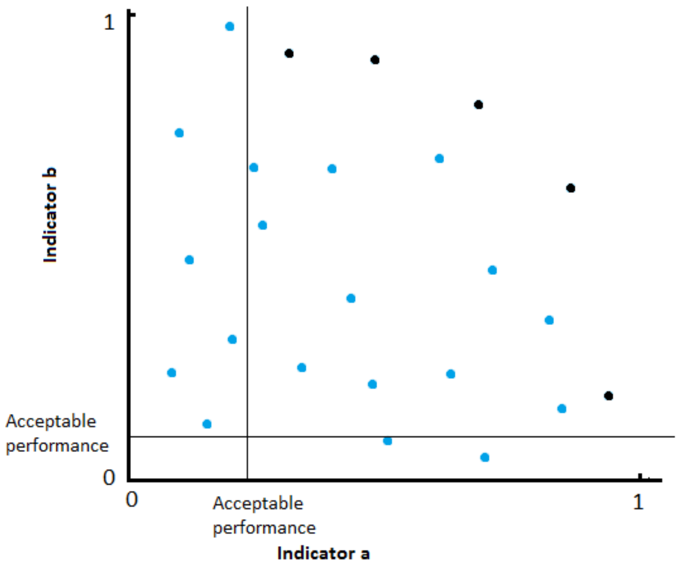


Figure 10 Illustration of the bounded Pareto front of two indicators. Here is 1 the optimal performance and 0 the worst performance. The acceptable levels of performance are indicated by lines. The black dots are the points on the bounded Pareto front.

When moving to a downstream basin, every run a behavioural set of parameters for the upstream basin is selected next to the random set of parameters. When reaching the outflow location of the larger sub-basin, the same analysis for low flow events as section 3.2.2 in is performed. The choice for an optimal parameter set is based on the 1 in 10 year event under the growing season threshold. From this analysis also the 5%, 50% and 95% values are evaluated as they represent the uncertainty.

3.3.3. Validation

The optimal parameter combination, as explained in section 3.3.2, is taken to validate flows on the validation time period of 1-1-1992 to 31-12-2003. The analysis for low flow events as presented in section 3.2 are performed to validate the time series and to compare with the original simulation. An interesting point of the validation from user perspective is whether the recalibrated model still performs well for floods. Therefore the yearly maximum values of discharge from the observations and the simulations are plotted against the probabilities expressed in standard Gumbel variates. The probabilities are calculated with equation 7.

4. Results skill of low flow

In this chapter the results of the evaluation of the HBV model and the weather generator from GRADE for low flows are presented. These analyses provide the answers to research questions 1,2 and 3. This chapter consists of two parts: first the results are shown per subject (section 4.1 till 4.6) and after that the interpretations of the results per sub-basin are discussed (section 4.7) which gives an overview of skill of HBV and the weather generator.

4.1. Discharges

In Figure 11 the flow duration curves for the discharges at the outlet stations of the sub-basins are presented. At Lobith, Andernach and Cochem simulations of Q50-Q100 give lower discharges than the observations, meaning general underestimation of the discharge by the simulations. At Frankfurt-Osthafen the simulations of Q65-Q100 are lower than the observations, the Q50-Q65 discharges are higher than the observations, so also here there is underestimation of the lowest discharges. In the Neckar differences are small, so there is no structural underestimation. At Rekingen the simulations are underestimating the observed discharge in the entire range of Q50-Q100. In the West Alpine sub-basin at Untersiggenthal there is both over- and underestimation: the Q90-Q100 discharges are overestimated in the simulations, and the Q50-Q85 flows are underestimated.

Differences in discharges in the FDC caused by the weather generator are only seen at Lobith, Andernach and Cochem in the Q65-Q100 where it gives higher discharges than the simulations (but lower than the observations). In the other basins differences between the historical simulation and the synthetic simulation in the flow duration curve are very small.

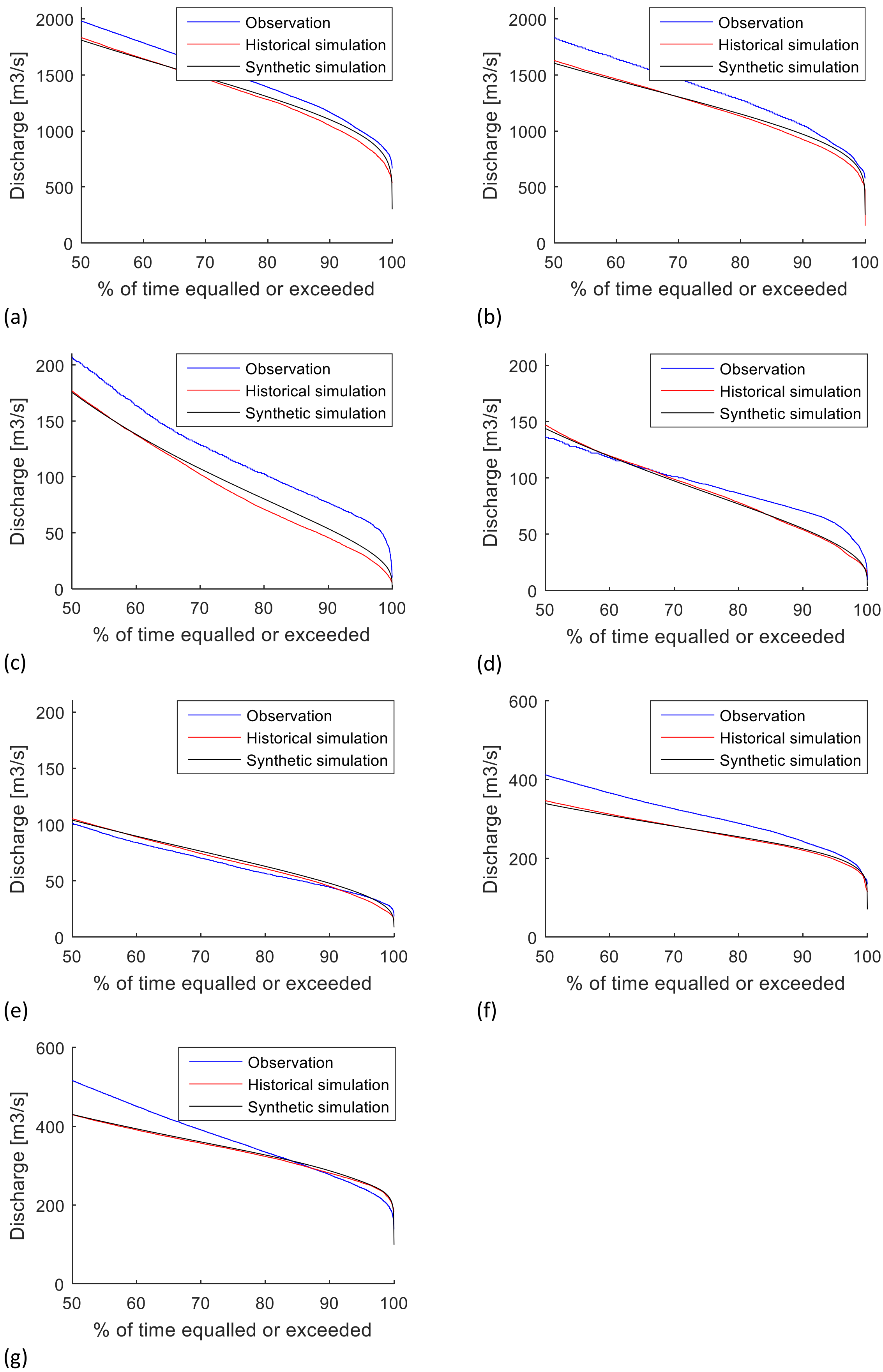


Figure 11 The lower end of the flow duration curves of the discharges for the seven sub-basins: a) Lobith (Lower Rhine), b) Andernach (Middle Rhine), c) Cochem (Moselle), d) Frankfurt-Osthafen (Main), e) Rockenau (Neckar), f) Rekingen (East Alpine) and g) Untersiggenthal (West Alpine).

From evaluating the autocorrelation length there are three interesting differences that can be seen in Table 8. At Frankfurt-Osthafen there is a larger autocorrelation length in the simulations than in the observations. So the discharges are more persistent in the simulations than in the observations. At the Alpine discharge locations there is the opposite effect: there is a large correlation length in the observations, because the resulting discharge is damped by lakes and originating from slowly and continually melting snow, and this is lower in the simulations.

The difference between the synthetic and historical simulations is small, but in the weather generator the autocorrelation length is slightly lower. This means that there are no large errors in the persistence of the discharge in the synthetic simulations caused by the weather generator, but that there is a little less persistence in the weather generator than in the observed weather.

Table 8 Correlation length (in days) for the discharge locations..

	Observation	**Historical simulation**	**Synthetic simulation**
Lobith	32	30	29
Andernach	31	29	28
Cochem	26	25	25
Frankfurt- Osthafen	30	37	38
Rockenau	20	18	18
Rekingen	45	33	32
Untersiggenthal	35	28	27

4.2. Low flow events

Matching low flow events

The observed and simulated events are matched with each other and the results from the contingency tables of the evaluated periods are shown in Table 9 and Table 10. The overall performance, indicated by the critical score index (CSI), is the highest for Lobith, Andernach and Rockenau and the lowest for the Alpine discharge locations.

The hit rate at Lobith for the navigation threshold events is high, there are only a few misses. Ninety percent of the observed events are also simulated. The False Alarm Ratio (FAR) is considerable; 40% of the simulated events are false alarms. For the events with growing season threshold the hit rate is lower and the FAR higher. The CSI is therefore lower. At Cochem, Frankfurt and Rockenau the hit rates for both thresholds are around 50%. Also the FAR is around this value, except for Rockenau, that has a smaller percentage of false alarms. At the Alpine locations there are many false alarms, so more low flow events in the simulations. The hit rates are high for the growing season, so there are a large number of events in the summer period in the simulations. The low hit rates for the navigation thresholds then can be explained by misses of events in winter (especially at Untersiggenthal).

Table 9 Results from the contingency table for events in time series with navigation thresholds.

Navigation threshold	Number of observed events	Hits	False alarms	Misses	CSI	Hit rate	False alarm ratio
Lobith	45	41	28	4	0.56	0.91	0.41
Andernach	47	40	42	7	0.45	0.85	0.51
Cochem	86	54	76	32	0.33	0.63	0.58
Frankfurt-O.	55	31	33	24	0.35	0.56	0.52
Rockenau	87	42	25	45	0.38	0.48	0.37
Rekingen	40	23	48	17	0.26	0.58	0.68
Untersiggenthal	77	20	64	57	0.14	0.26	0.76

Table 10 Results from the contingency table for events in time series with growing season thresholds.

Growing season threshold	Number of observed events	Hits	False alarms	Misses	CSI	Hit rate	False alarm ratio
Lobith	28	22	24	6	0.42	0.79	0.52
Andernach	26	20	24	6	0.40	0.77	0.55
Cochem	52	30	31	22	0.36	0.58	0.51
Frankfurt-O.	21	9	7	12	0.32	0.43	0.44
Rockenau	51	25	8	26	0.42	0.49	0.24
Rekingen	30	24	52	6	0.29	0.80	0.68
Untersiggenthal	46	33	86	13	0.25	0.72	0.72

In Table 11 and Table 12 the performance indicators of the hit events are presented for respectively the events with the navigation thresholds and the events with the growing season thresholds. At Lobith the Mean absolute Errors (MARE's) of the severity for both thresholds are higher than the MARE's of the durations. That means that the error in the duration, relative to the durations, is smaller than the error found in the severities, relative to the severities. The coefficients of determination R^2 for events at Lobith under the navigation thresholds are high, meaning that there is a quite clear relation between the characteristics of observed events and simulated events. For events at Lobith under the growing season thresholds there is a smaller MARE, and also a smaller R^2, pointing at a smaller error in the characteristics with less linearity in over- or underestimation. The timing bias in the events is that the events mainly start earlier and end later. This is in line with the overestimation of the durations.

Remarkable results from the other discharge locations are the large increase in MARE under the growing season thresholds, from Lobith to Andernach, while these locations have almost the same discharge regime. The MARE and R^2 at Frankfurt under the growing season threshold are very good, although the events seem to start on average very late in the simulations. This is only based on a small number of event hits. In Rekingen and Untersiggenthal there are very low R^2 values, meaning there is no correlation in the errors, there are under- and overestimations. In most of the locations the simulated events start earlier and end later. Very large MARE's can be found in the severities under the navigation threshold, at the locations Cochem, Frankfurt and Rockenau and in the severities under the growing season threshold at Rekingen. This is mainly due to hits of observed events with a very small severity with simulated events with a large severity. This can be quickly a factor 10 or higher in difference.

Table 11 Performance indicators of hit events in time series with navigation thresholds.

Navigation threshold	**Number of observed events**	**hits**	**Duration**		**Severity**		**Timing**	
			MARE [-]	R^2 [-]	MARE [-]	R^2 [-]	Mean # days start date later	Mean # days end date later
Lobith	45	41	1.57	0.77	18.07	0.74	-6.15	3.17
Andernach	47	40	2.46	0.42	24.58	0.61	-4.60	3.48
Cochem	86	54	6.26	0.63	80.35	0.57	-10.41	8.83
Frankfurt-O.	55	31	6.54	0.60	95.98	0.37	-10.06	10.61
Rockenau	87	42	3.75	0.40	43.01	0.64	3.07	4.67
Rekingen	40	23	2.14	0.024	12.82	0.0066	-5.09	-0.91
Untersiggenthal	77	20	3.20	0.0060	7.81	0.098	-5.85	0.40

Table 12 Performance indicators of hit events in time series with growing season thresholds.

Growing Season threshold	**Number of observed events**	**hits**	**Duration**		**Severity**		**Timing**	
			MARE [-]	R^2 [-]	MARE [-]	R^2 [-]	Mean # days start date later	Mean # days end date later
Lobith	28	22	0.95	0.53	6.85	0.67	-5.64	5.68
Andernach	26	20	3.50	0.65	18.69	0.35	-6.15	9.35
Cochem	52	30	6.05	0.29	21.03	0.48	-6.70	11.63
Frankfurt-O.	21	9	0.48	0.85	1.16	0.39	11.22	0.89
Rockenau	51	25	1.77	0.28	10.25	0.24	-0.12	1.16
Rekingen	30	24	5.14	0.29	52.96	0.0077	-7.46	18.29
Untersiggenthal	46	33	4.59	0.45	26.10	0.41	-9.79	9.21

Number of events

In Table 13 the average number of events per year can be found. At Lobith in the observations there is on average once every two years an event in the growing season, and once every 1.5 year an event with the navigation thresholds. In Andernach there are slightly more events than at Lobith. Almost the same number of events per year in the different seasons as at Lobith is seen at Andernach. Cochem and Rockenau have similar numbers of events with more events than at all other locations. Also Frankfurt has more events than Lobith, but the numbers of events in the growing season are comparable. At Rekingen the number of events is also similar to the number at Lobith, while at Untersiggenthal the number of events is overall higher than at Lobith.

Table 13 Average number of events per year in different time series for different thresholds and periods.

Events per year		**Navigation threshold**	**Growing season threshold**	**Navigation threshold summer**	**Navigation threshold winter**
				April-September	October-March
Lobith	Obs	0.85	0.53	0.26	0.58
	Sim historical	1.30	0.87	0.53	0.77
	Sim synthetic	1.24	0.78	0.42	0.82
Andernach	Obs	0.90	0.63	0.35	0.56
	Sim historical	1.58	1.08	0.69	0.88
	Sim synthetic	1.25	1.08	0.45	0.80
Cochem	Obs	1.72	1.04	0.70	1.02
	Sim historical	2.60	1.22	1.40	1.20
	Sim synthetic	2.23	1.16	1.05	1.19
Frankfurt	Obs	1.49	0.57	0.38	1.11
	Sim historical	1.73	0.43	0.59	1.14
	Sim synthetic	1.57	0.30	0.50	1.07
Rockenau	Obs	1.67	0.98	0.63	1.04
	Sim historical	1.29	0.63	0.48	0.81
	Sim synthetic	1.13	0.36	0.26	0.87
Rekingen	Obs	0.77	0.58	0.29	0.48
	Sim historical	1.37	1.46	1.08	0.29
	Sim synthetic	1.20	1.51	0.96	0.24
Untersiggenthal	Obs	1.48	0.88	0.58	0.90
	Sim historical	1.62	2.29	1.42	0.19
	Sim synthetic	1.68	2.55	1.45	0.23

At Lobith, Andernach and Cochem, the historical simulations give more events than the observations. In the Main the historical simulations this is also the case, except for events under the growing season threshold. At Rockenau all the historical simulated events occur less often than the observed events. At the Alpine discharge locations the number of winter events are remarkable. With the other thresholds the number of events is larger in the historical simulations, but in the winter events this is lower. This means that there are less low flows in winter in the historical simulations, and more in the summer. The number of events in the growing season (and summer) is much larger.

The influence of the synthetic weather is in most cases a small decrease from the historical simulations in the number of events per year, which brings the number closer to the observed number. At Cochem the decrease of events in summer is quite substantial. At Untersiggenthal the influence of the weather generator is an increase in the number of events per year in all cases.

Return periods of low flow events

When looking at the event properties (Figure 12) there are several things to see. At Lobith and Andernach the severities of events are more overestimated by the historical simulations than the durations. The synthetic series has less severe and shorter events than the historical simulations. The synthetic series is closer to the observations.

At Cochem also the historical simulations overestimate the event properties. In the synthetic simulations there is also overestimation, but slightly less than in the historical simulations. In the growing season events the difference is larger than in the navigation threshold events. At the influence of the weather generator is very strong in the synthetic simulations, and is causing underestimation of severity and duration compared to both the historical simulations as the observations. In the navigation threshold events at Frankfurt the synthetic and historical simulations lay closer together and the event properties are overestimated compared to the observations. For the Neckar there is a good representation of low flow severities and durations in both historical and synthetic simulations. At the Alpine discharge locations there is a large difference between performances of the simulations in both thresholds. With the navigation thresholds the performances are good, but in the growing season the historical simulations give large overestimations of severity and duration. The influence of the weather generator seems the same with both thresholds, and is rather small. At Rekingen and Untersiggenthal some synthetic simulated events reach the maximum duration of the growing season events, the duration of the growing season. These cause the horizontal lines in the graphs.

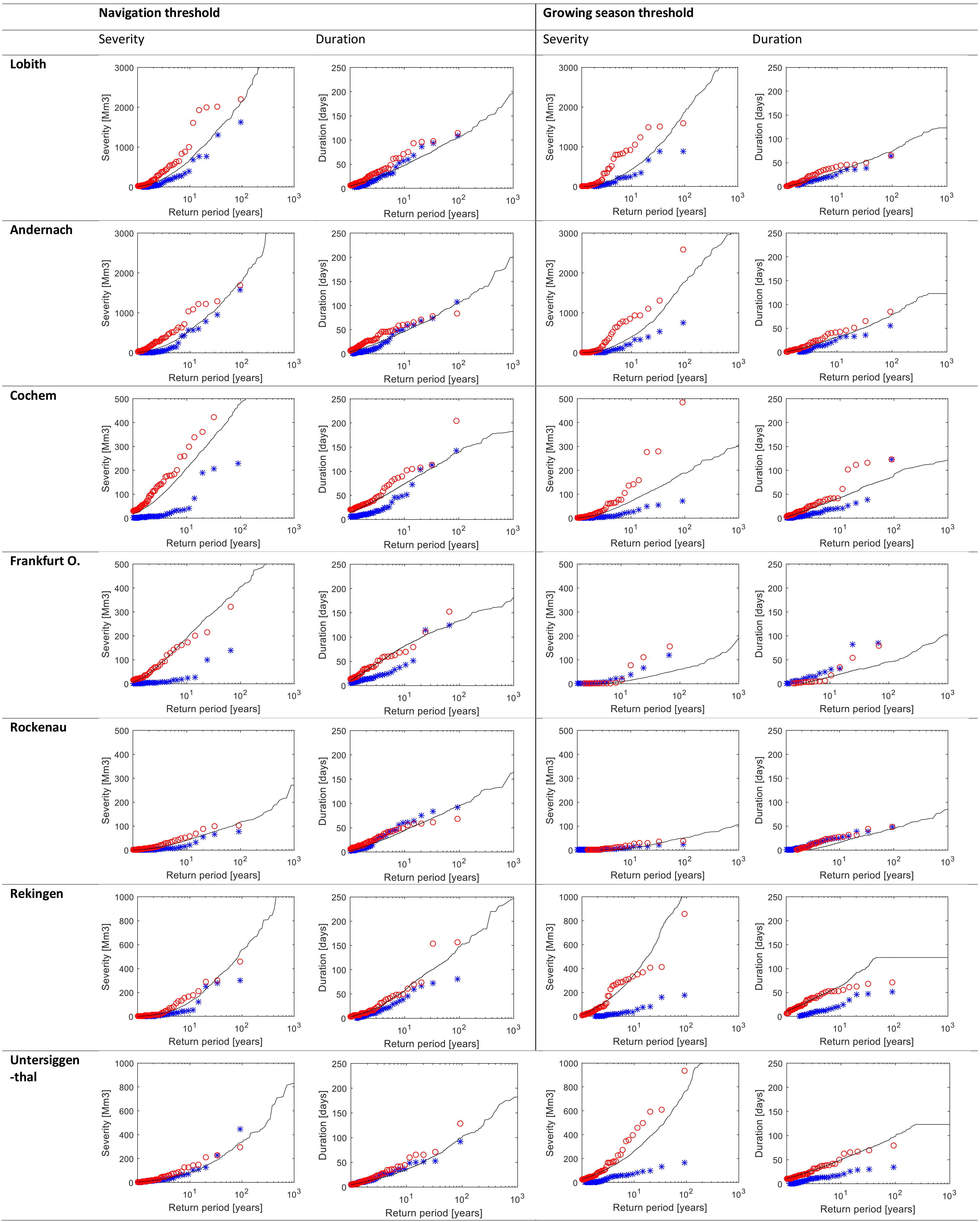


Figure 12 Return periods of event characteristics severity and duration. Blue=observations, red= simulations, black= synthetic simulations

Contributions of sub-basins during low flow events

In Table 14 the characteristics of relative contributions of the discharge from the sub-basins during low flow events at Lobith are presented. Per event the contributions of sub-basins vary, but by looking at the average contributions, it can be seen that the West Alpine sub-basin has the largest contribution, and also the East Alpine sub-basin adds more water to the Rhine discharge than the other tributaries. The Alpine sub-basins have the least variation in the contributions. The contributions of the Main, Neckar and Moselle are rather low, but the Moselle has some higher contributions. The Lower and Middle Rhine contributions together are also a considerable part of the discharge in low flow events. For the growing season events the contribution from the Alps is larger, and the from the Moselle, Middle and Lower Rhine basins smaller.

Table 14 Observed relative contributions of sub-basins to low flow events at Lobith, for two thresholds.

	Navigation				**Growing season**			
	Min	**Max**	**Std/ mean**	**Mean**	**Min**	**Max**	**Std/ mean**	**Mean**
Lower and Middle Rhine	0.10	0.35	29%	0.22	0.06	0.27	32%	0.15
Moselle	0.03	0.20	38%	0.10	0.03	0.13	33%	0.08
Main	0.03	0.11	27%	0.07	0.03	0.11	35%	0.07
Neckar	0.03	0.07	22%	0.05	0.03	0.09	29%	0.05
East Alpine sub-basin	0.16	0.35	20%	0.26	0.21	0.38	15%	0.29
West Alpine sub-basin	0.20	0.44	22%	0.29	0.27	0.45	12%	0.37

In Table 15 the average contributions are shown for different thresholds and also for the simulated time series. In general the picture is the same in the observations, the historical simulations and the synthetic simulations: the largest contributions come from the Alps, than the Lower and Middle Rhine and the contributions of the other tributaries are rather small. In the winter events the contribution from the Alps is lower and the contribution from the tributaries is larger than the summer and growing season events. The proportion of the contributions of the basins to low flow events seems to be represented well by both simulations. There are some differences between the historical simulations and the observations, mainly in the contributions of the Lower and Middle Rhine in summer. In the synthetic simulations these are even further off. In the Alpine areas the synthetic simulations make the contribution of the Alps go down, and the contribution of the Lower and Middle Rhine go up.

Table 15 Relative contributions of sub-basins to low flow events at Lobith, for different time series and thresholds.

		Lower + Middle Rhine	Moselle	Main	Neckar	East Alpine	West Alpine
Navigation	Obs	0.22	0.10	0.07	0.05	0.26	0.29
	Sim historical	0.24	0.08	0.06	0.05	0.26	0.31
	Sim synthetic	0.25	0.09	0.06	0.05	0.25	0.31
Growing Season	Obs	0.15	0.08	0.07	0.05	0.29	0.37
	Sim historical	0.23	0.06	0.07	0.05	0.25	0.33
	Sim synthetic	0.24	0.07	0.08	0.05	0.24	0.32
Navigation Summer	Obs	0.16	0.07	0.07	0.04	0.30	0.36
	Sim historical	0.22	0.05	0.06	0.05	0.27	0.35
	Sim synthetic	0.23	0.05	0.06	0.05	0.27	0.35
Navigation Winter	Obs	0.25	0.12	0.08	0.05	0.24	0.26
	Sim historical	0.26	0.10	0.06	0.05	0.24	0.29
	Sim synthetic	0.26	0.10	0.06	0.05	0.24	0.29

4.3. Lakes

The performance of historical lake level simulation is presented in Table 16 and measured by the Nash Sutcliffe Efficiency (NSE) and the Mean Absolute Error (MAE). Lake Neuchâtel and Lake Zurich are simulated best, illustrated by the highest NSE and the very small MAE. Lake Lucerne has a very low NSE so is simulated less well. The absolute errors in the simulations are the largest in Lake Constance. From the duration curves in Figure 13 the same conclusion can be drawn, and it is clear that the simulations are underestimating the lake levels at Lake Constance and Lake Lucerne. In summer months the errors there are larger than in winter months. When the lakes are more empty in summer in the simulations, there is less outflow. The influence of the weather generator on the lake levels is small when looking at the level duration curves of historical simulated and synthetic simulated.

Table 16 Performance indicators of historical lake level simulations.

Lake	NSE	MAE	MAE summer (April-September)	MAE winter
	[-]	[m]	[m]	[m]
1 Lake Constance	0.60	0.30	0.42	0.18
2 Lake Neuchâtel	0.85	0.060	0.062	0.057
3 Lake Lucerne	-0.16	0.17	0.20	0.15
4 Lake Zurich	0.79	0.039	0.039	0.038

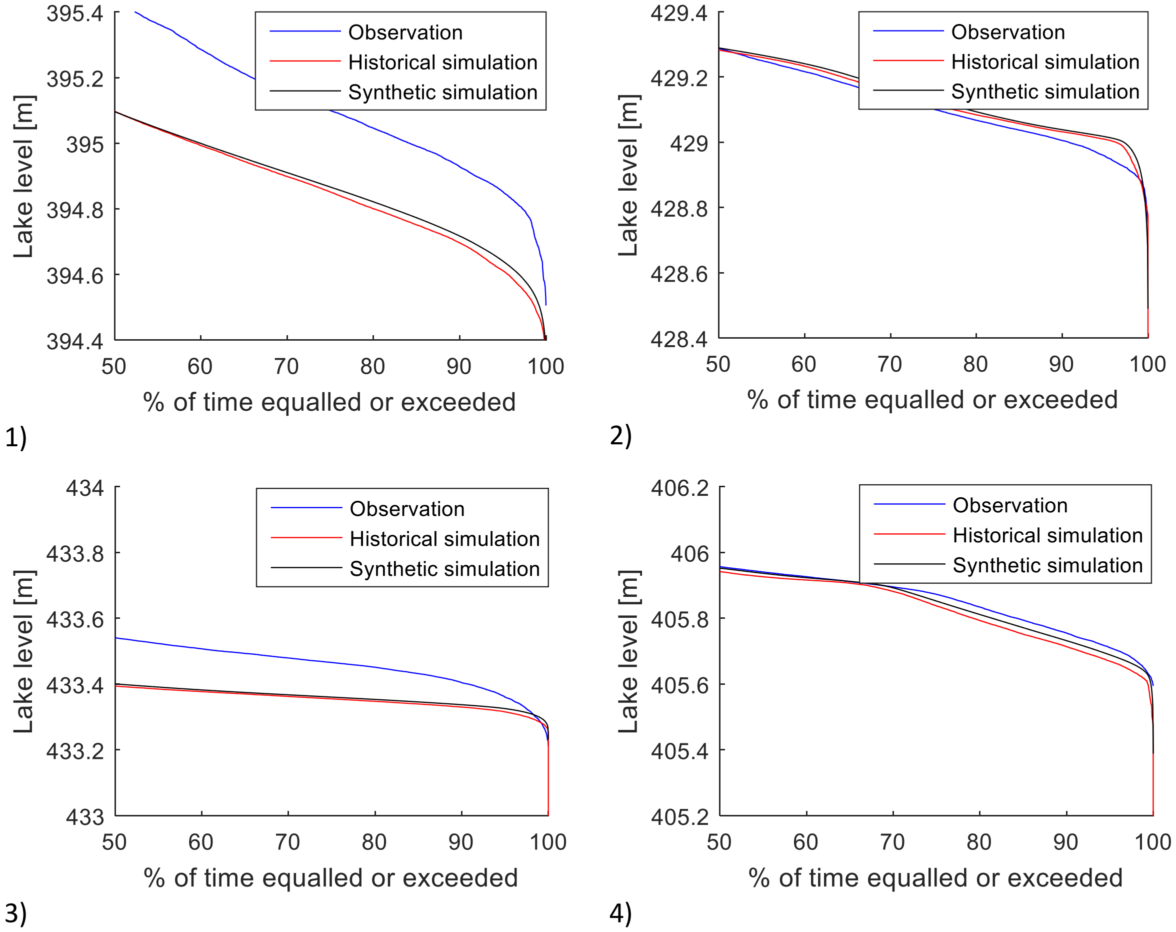


Figure 13 The level duration curves for the four Swiss lakes , 1=lake Constance, 2= lake Neuchâtel, 3= lake Lucerne, 4= lake Zurich.

4.4. Snow

Plotting both observed and simulated snow cover indices shows that it gives quite a good representation (Figure 14). The same thing is seen in the correlation between the series and the correlation between the yearly maximum snow cover indices (Table 17). In the West Alpine the representation is better than in the East Alpine, although both correlations are high. The snow covers in the simulations tend to start later, end later, and in total lasts longer. Also the maximum is reached later. In the East Alpine the end of the snow cover (the melt) has a larger timing error than the start. In the West Alpine the average error in timing of the start and the end is almost the same.

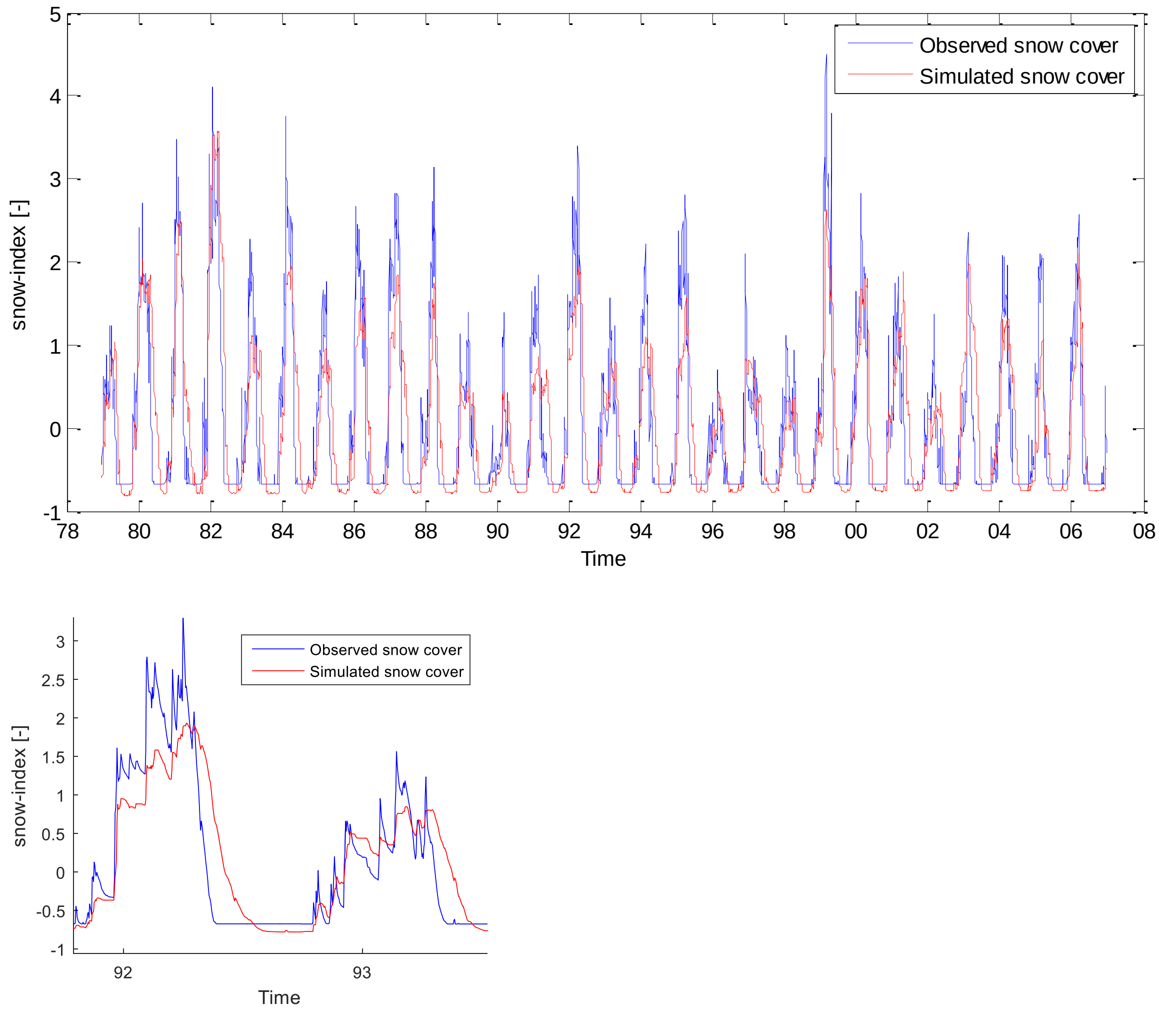


Figure 14 Snow cover index of observed and simulated series for the East Alpine basin. With a detail of the years 1992 and 1993.

Table 17 Performance indicators of snow cover simulation

	Correlation time series	Mean # days start day later	Mean # days end day later	Mean # days max snow cover later	Correlation maximum cover
East Alpine	0.83	24	46	28	0.88
West Alpine	0.94	32	31	16	0.93

In Figure 15 statistical properties of the snow covers in the series of the historical and synthetic simulations are shown with the unstandardized snow water contents in mm. There can be seen that the influence of the weather generator is larger in the West Alpine than in the East Alpine. The yearly maximum snow cover is underestimated, and also the durations are smaller.

In some of the HBV sub-catchments (Rhein2 in the East Alpine sub-basin and Reusseed and Lintmoll in the West Alpine sub-basin) the snow does not all melt so there is accumulation over the years. In the East Alpine this is on average 0.35 mm per year both in the historical simulations and in the synthetic simulations. In the West Alpine this is much more: 4.9 mm per year and 1.4 mm with the weather generator (5.6 m in 4000 years). The influence of the weather generator is thus that there is less snow accumulation over the years. The accumulation has probability to do with the snow parameters in the HBV model and the temperature on the elevations that is used in the model. Because the accumulation per year is very small, it is considered to have a very small effect on discharge simulations.

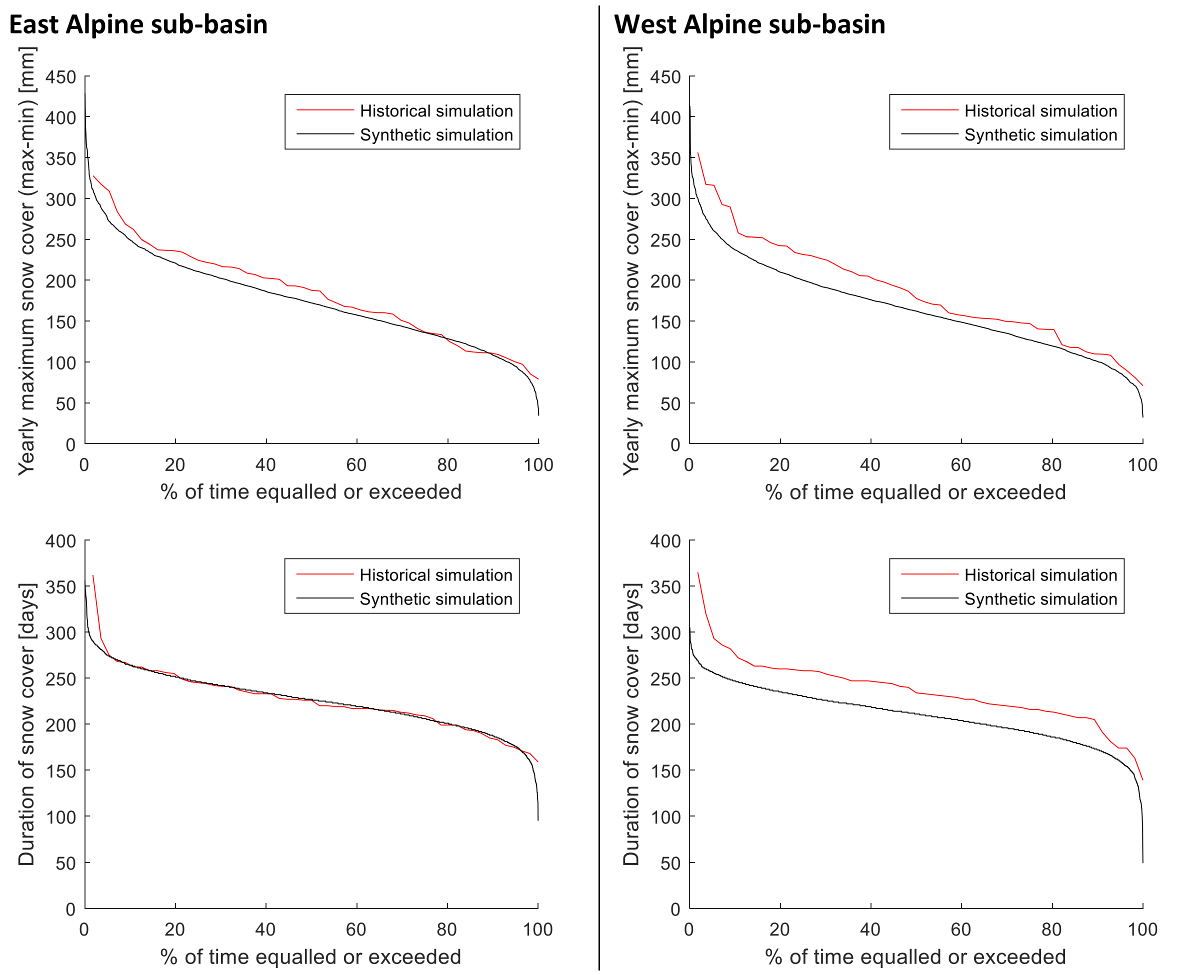


Figure 15 Duration curves for the yearly maximum depth and the duration of the snow cover.

4.5. Groundwater

The correlation between the groundwater index of observed groundwater levels and the index of simulated lower zone volumes gives an indication whether the simulated groundwater storage has the same pattern as the observed ground water levels. From the results (Table 18) appears that simulations are quite good, only in the East Alpine sub-basin the correlation between the two indices is low. It seems that the groundwater processes there are behaving less like the observations than in the other sub-basins.

Table 18 Correlation coefficients for groundwater indices of observed and simulated series. For the West Alpine basin not enough data is available to carry on this analysis.

	Correlation groundwater indices	**Correlation groundwater index with discharges observed**	**Correlation groundwater index with discharges simulated**
Lower Rhine	0.86	0.70	0.51
Middle Rhine	0.78	0.71	0.52
Moselle	0.72	0.62	0.71
Main	0.82	0.57	0.87
Neckar	0.88	0.43	0.40
East Alpine sub-basin	0.51	0.70	0.69

A few remarks can be made on correlations of groundwater indices and discharge: in the Lower and Middle Rhine these correlations are higher in the observations than in the simulations, in the Main the correlation is higher in the simulation. The correlations in observation and simulation in the Neckar and East Alpine agree very well. In Figure 16 an example from observed and historical simulated groundwater index in the Main is shown.

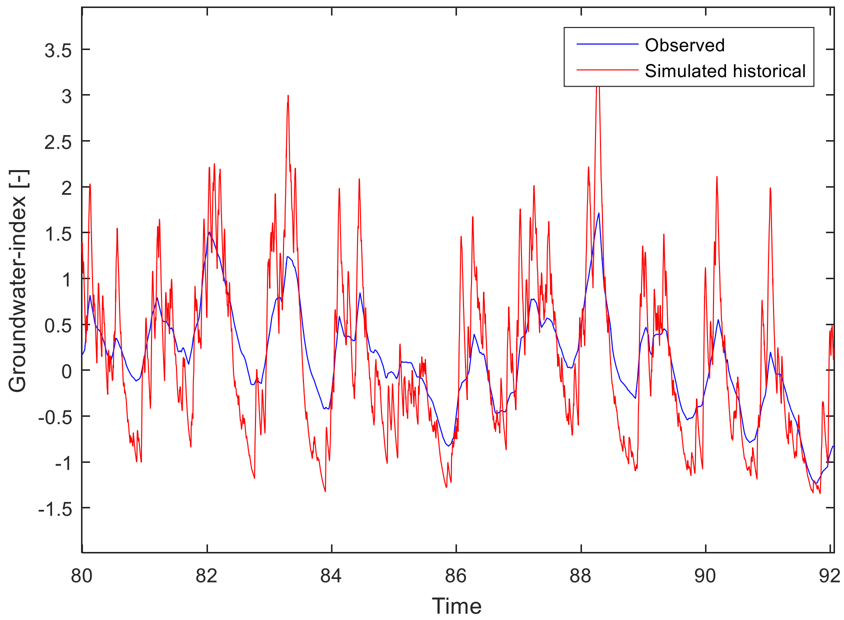


Figure 16 Observed and simulated groundwater indices from 1980 to 1992 in the Main. Correlation between these two series is high.

As can be seen in Figure 17, the influence of the weather generator is not causing large differences in the groundwater contents. The duration curves are almost the same. In the Middle Rhine and East

Alpine the large values are somewhat underestimated, and in the Lower Rhine the lower values are a bit overestimated.

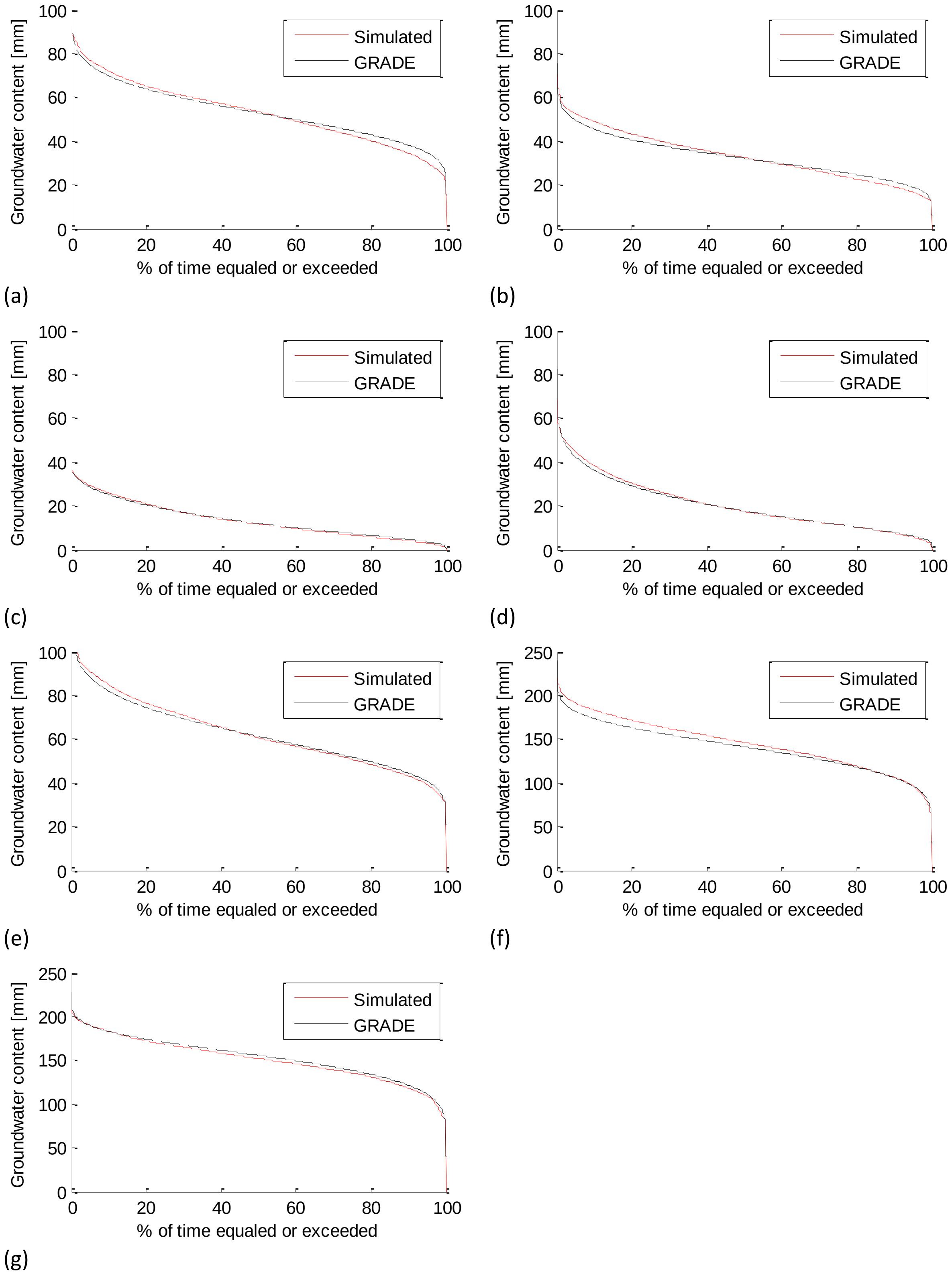


Figure 17 Duration curves for the ground water indices for the seven sub-basins : a) Lobith, b) Andernach, c) Moselle, d) Main, e) Neckar, f) East Alpine sub-basin and g) West Alpine sub-basin.

4.6. Meteorology

4.6.1. Precipitation

In Figure 18 the return periods are plotted for minimum annual cumulative precipitation sums for observed and weather generator precipitation, averaged over the whole Rhine basin. In the minima of 30 day and 120 day sums, there is some overestimation of the weather generator, especially in the summer. This is also seen in the separate basins.

In the separate basins the minima are overestimated in the Lower Rhine, Middle Rhine, Moselle and Main, especially in the summer. In the Neckar and the Alpine basins there is less overestimation of the lowest minima, only in summer some differences can be seen. Examples are shown in Figure 19, the rest of the graphs can be found in Appendix A.

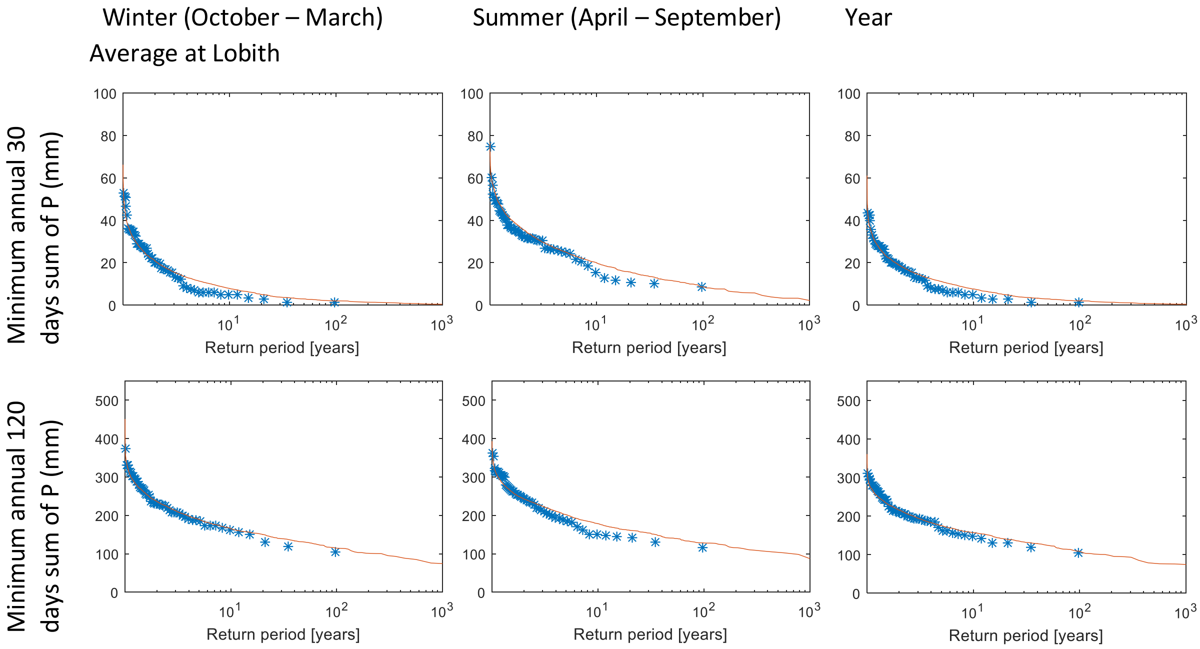


Figure 18 Probability of occurrence of annual precipitation sums for winter, summer and the entire year, averaged over the Rhine basin. Blue dots are the observed values, the red line indicates the values from the weather generator.

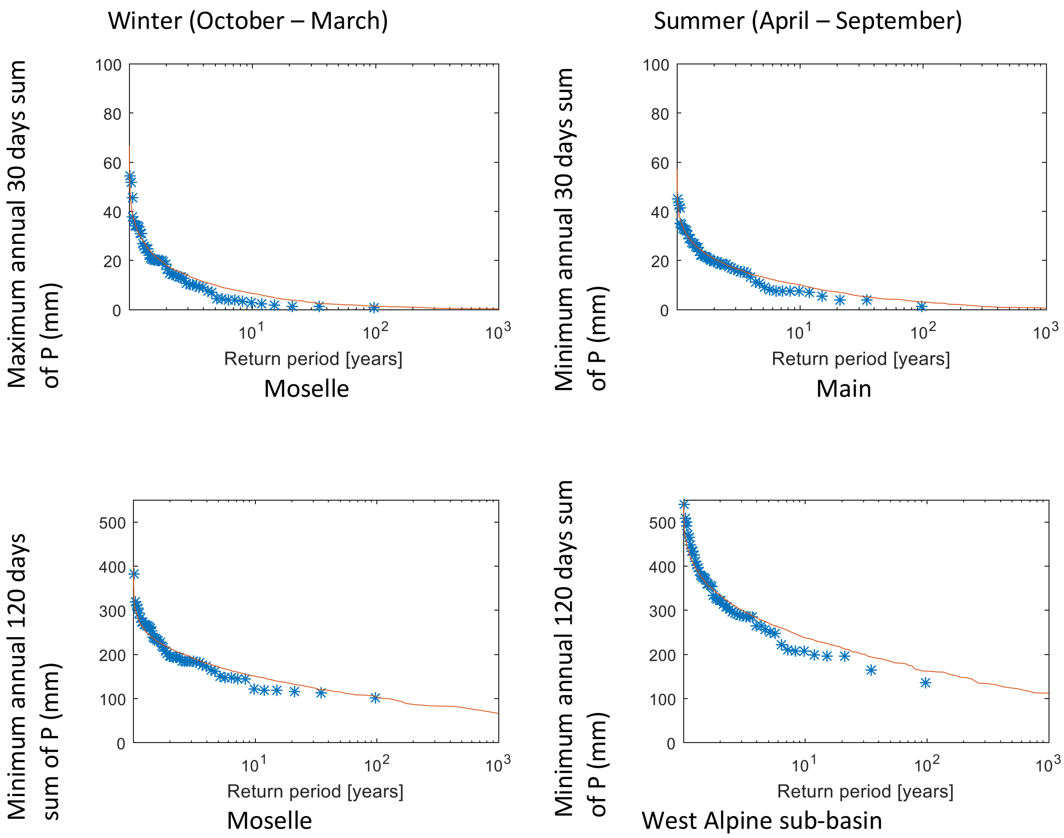


Figure 19 Examples of probabilities of occurrences of annual precipitation sums for different periods and different basins . Blue dots are the observed values, the red line indicates the values from the weather generator.

4.6.2. Temperature

In Figure 20 the probabilities of occurrence for numbers of days with temperatures below zero degrees Celsius are given. On these days precipitation will fall as snow and the snow will not melt. For both Alpine basins appears that there are slightly less days with temperatures below zero in the weather generator, but the difference between the synthetic series and the observations is very small.

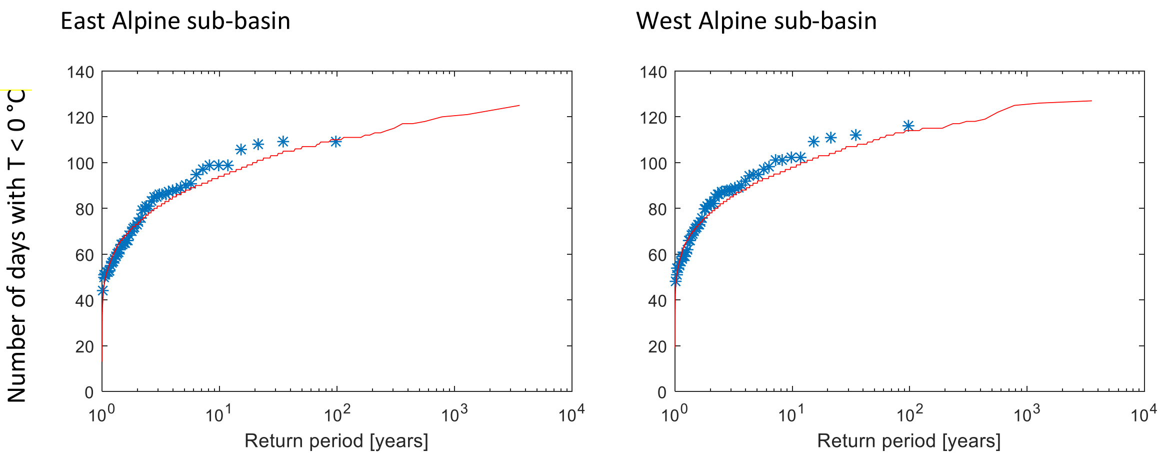


Figure 20 Probabilities of occurrence of annual number of days with temperatures below zero in the Alpine sub-basins. Blue dots are the observed values, the red line indicates the values from the weather generator.

4.7. Interpretation of results

In this section the results from the evaluations in the sections 4.1 to 4.6 are combined into a resulting analysis per sub-basin for the hydrological model and for the weather generator. The results are discussed from upstream to downstream. At the end of this section the strengths and limitations of both parts are given.

4.7.1. Hydrological model HBV

East Alpine sub-basin

At Rekingen the discharge of low flows is underestimated by the simulations, appearing from the flow duration curve (FDC) and the large number of false alarms in the events. The underestimation is larger in summer, resulting in a smaller number of missed events under the growing season thresholds (GS events) than under the navigation thresholds (N events), a large error in the return periods of GS events and a large mean absolute error (MAE) in summer in Lake Constance. Also the average contribution of the East Alpine sub-basin to low flow events in summer (GS and N) at Lobith is smaller in the simulations. There are both shorter and longer events, indicated by the low R^2 values of the durations and severities of the matched events. The shorter events result from the larger number of N events, which do not result in large errors in the return periods, and the longer events indicated by the large overestimation of return periods in the GS events. The underestimation of discharge in summer, especially in the first months, can be caused by the longer period it takes to melt and the later start of the melting process as appears from the snow cover data. This difference in snow simulation can also explain the smaller autocorrelation length that is found in the simulated discharge series compared to the observations.

West Alpine sub-basin

For the West Alpine discharge location Untersiggenthal applies that there is both overestimation and underestimation; this is already seen in the FDC. Because of the reasonable number of hits but many false alarms in the GS events, there is underestimation of discharge in the summer. In the N events there are only few hits and also many false alarms. Because the N events occur in both summer and winter, the false alarms are likely to derive from the underestimation of discharge in summer and the misses come from the overestimation of discharge in winter, also seen in the low R^2 of N events. This also results in the many more events in summer and less events in winter months and the lower average contribution of the West Alpine sub-basin to low flow events at Lobith under the growing season threshold and the larger contribution to events in winter under the navigation threshold. There is almost no error in the return periods of N event characteristics, because under- and overestimation balance each other out. In the severities of events under the growing season threshold there are large errors. Especially the severity is much overestimated. The error in the discharge is also visible in the autocorrelation length which is smaller in the simulations than in the observations. Part of the error in discharge can be explained by the error in lake levels of Lake Lucerne. The snow, although the correlation is high, starts later to build up and melts later and in shorter time. Too less snow would be an explanation for the overestimation of flow in winter (not much snow accumulation) and for the underestimation of summer flows (too less melt water).

Neckar

In the Neckar at Cochem there is a very accurate representation of the discharge when looking at the FDC and the autocorrelation length. The large number of misses and false alarms can be explained by having very short events, also seen in the short autocorrelation length. The well simulated discharge makes also the simulated return periods of event characteristics quite accurate. There is a slight overestimation of the severity (also seen in the Mean Absolute Relative Errors (MARE's)) and a slight underestimation of the duration. There is a high correlation of the groundwater indices and the groundwater indices have the same relation with the discharge. This is possibly the origin of the good discharge simulation.

Main

At Frankfurt the simulations underestimate the discharge, as can be seen in the FDC. Remarkable is the twice as long autocorrelation length in the simulations compared to the observations, which leads to more persistence and longer events. That this is especially the case in the winter months, can be seen in the large MARE's for N events and the overestimation of severity and duration for the return periods of N events. In the growing season there are not many events, and while there are quite some misses, the characteristics are simulated correctly. There is a good correlation between groundwater indices of observations and simulations, but the correlation of groundwater index and discharge is larger in the simulations than in the observations. This may be the reason of the larger persistence of discharge. Because the underestimation of flow is especially in the winter, and there are both in the observations as the simulations few low flow events, the lack of regulation by weirs is not considered to be a large part of the error.

Moselle

The underestimation of discharges at Cochem that is seen in the FDC, is leading to more low flow events. This causes a large number of false alarms, and twice as much N events in summer as in the observations. The severities and durations are overestimated, this can be seen in the high MARE's in the matched events and the large differences in return periods for the severities. Also the durations are overestimated, but the most extreme durations have very similar return periods in the simulations and observations. The groundwater-discharge relation is somewhat higher in the simulations, but the correlation between the two groundwater indices is good. Because there is underestimation of discharge taking place throughout the entire year, the lack of regulation by weirs in the hydrological model could be the explanation of the differences.

Andernach

Also discharges at Andernach are underestimated, this is probably because the tributary flows, except for the Neckar, are also mainly underestimated. The autocorrelation lag in the simulations is slightly smaller than in the observations, this could be because the inflow of the tributaries is also different from the observations, and in the Alps there is a large decrease in autocorrelation. From the high hit rate, high false alarm rate and the increased number of events can be concluded that the underestimation of flow results in more events, especially in summer. This underestimation in summer is also seen in Rekingen, Untersiggenthal and Cochem. The return periods of duration are represented quite good, but the discharge during low flow events is lower, so that the return periods of severities are largely overestimated.

Lobith

There are only a few differences between the discharge at Andernach and the discharge at Lobith. The underestimation of discharge in the FDC is slightly less. Perhaps the simulation of the small tributaries in the Lower Rhine basin is the cause, which could also explain the larger average contribution of the Lower and Middle Rhine basin to low flow events in summer at Lobith. The event properties at Andernach and Lobith in return periods are almost the same, but the number of events is lower, which can also be seen in the larger autocorrelation lag. The matching of events is rather good, and the overestimation of severity is smaller than in Andernach. Especially the number of events in the summer with the navigation thresholds is overestimated in the simulations.

4.7.2. Weather generator

East Alpine sub-basin

The weather generator in the East Alpine sub-basin performs very well. Looking at the precipitation of the weather generator there are less extreme 30 day dry periods in winter and more extreme 30 day dry periods in summer and with 120 day periods this is opposite. There is no difference in FDC and return periods of severity and duration of N events, while there are less events both in summer and winter. There are slightly more events under the growing season thresholds, and there is a large increase in duration in extreme events, showing thus both more and longer events. The combination temperature and precipitation does not seem to influence the snow simulations.

West Alpine sub-basin

The number of days with temperatures below zero degrees in the West Alpine sub-basin are a bit less in the synthetic weather, but the combination temperature and precipitation in winter causes a significant change in duration and maximum depth of snow cover. This causes more low flows because there is less melt water in summer. There are less extreme 120 day dry periods in the summer, but an increase of the number of events. There is a decrease of severities of these events seen in the return periods. The difference in rainfall makes that there are less severe events. In the FDC, the lake levels and the groundwater curves is no difference seen, so any differences average out over the years.

Neckar

In the Neckar sub-basin little difference is seen in the precipitation statistics. There is a slight increase of the 30 day minimum in summer, so less extreme dry periods. This is also seen in the number of events in summer, with both thresholds. The durations of the GS events seem to be a little underestimated. The weather generator causes no change in the FDC or the groundwater duration curve. Also the short autocorrelation length shows that the overall regime stays the same.

Main

In the Main sub-basin there is also a small difference in the summer dry periods in the weather generator compared to the observations. The less dry periods cause no difference in the FDC, the autocorrelation lag or the groundwater level duration curve. They do cause less low flow events, and a large error in the return periods of severity and duration of GS events. There is less low flow in the growing season, and in the winter there are probably more low discharges so that the return periods of the characteristics of N events are represented well.

Moselle

In the sub-basin of the Moselle is a more clear relation between the weather generator precipitation and the outcomes of the low flow analyses. For the Moselle the weather generator gives an increase of the extreme minimum sums of precipitation, especially in summer. This causes less low flows in summer, which is confirmed by the large decrease in N events in summer and the higher return periods of all event characteristics, especially the severity of GS events. The flow duration curve here shows also an increase of the lowest flows. There is no change in the autocorrelation lag and the groundwater level duration curve.

Andernach

The weather influences of the tributaries and the less dry periods in the Middle Rhine basin itself cause less N events, especially in the summer. Therefore the severity of GS events has decreased. The severity of N events and the duration of GS events have decreased a little. There is a slight increase of the lowest flows in the FDC, no change in the autocorrelation lag, and less variance in the groundwater levels.

Lobith

Also in the Lower Rhine basin there are less dry periods from the weather generator, so the effects from dry periods in the summer period in Andernach are strengthened; there are less low flow events in the summer and the largest difference in the return periods of the event characteristics is the decrease of severity of GS events. There are more low flow events, but still a decrease of severity. Also there is a light increase of the lowest discharges in the FDC, so there are no indications that there are more low flows caused by the weather generator. There is a little decrease in autocorrelation length and the variance in groundwater levels is less.

4.7.3. Performances

Hydrological model

The hydrological model causes mainly underestimations of flow. In the Alpine basins this is especially the case in summer, which can be related to errors in the snow simulation. At Frankfurt the underestimation is mainly in winter, which can be related to groundwater. In the tributaries Moselle, Main and Neckar regulation of flow by weirs can play a role in the underestimations, but this should not change much in the simulation of flow at Lobith, because the contributions from the tributaries are small. Therefore also the good performance of simulated discharge in the Neckar has little effect for Lobith. The underestimation of discharge at Lobith is considered to be caused mainly by the errors in the Alpine basins. The hit rate is high at Lobith and the probabilities of occurrence for the durations are simulated well. For the severities however, the performance is poor.

Weather generator

The weather generator has less dry 30 and 120 day periods than there are observed. This is especially the case in the Lower and Middle Rhine, the Moselle, the Main and the West Alpine sub-basin and more in summer than in winter. Therefore there are smaller numbers of low flows, and decreases of severities and durations. In the West Alpine sub-basin there is a significant difference in simulations of snow cover depth and duration. The decrease of snow causes more low flows in summer.

5. Results of recalibration East Alpine sub-basin

In this section the results of the recalibration of the East Alpine sub-basin are presented. This is the interpretation of research question 4. In section 5.1 the problem definition is given on which this recalibration is based. In section 5.2 a detailed set-up of the calibration is presented and in section 5.3 the calibration results are shown. Section 5.4 covers the validation of the recalibrated set.

5.1. Problem definition

The aim of the recalibration is to improve the modelling of low flows in the East Alpine basin. In Figure 13 in the results of the evaluation was shown that the modelling of Lake Constance has a low performance. This could be because of the inflow or because of the lake schematization itself. To check this, differences between observation and simulation of the upstream and downstream discharge locations are plotted, as seen in Figure 21. At the discharge locations Domat/Ems and Diepoldsau upstream of Lake Constance the flow is regulated (also seen in the large variability of errors), which makes it more difficult to calibrate, but still is seen that the error is the largest from February/March to July/August and mainly negative, directing to a systematic underestimation of the flow by the HBV model. The same underestimation is seen in the lake levels and the outflow of the lake at Neuhausen. The underestimation of flow in the spring can have its origin in underestimation of the precipitation measurements in the Alps (Davids et al., 2015). Precipitation that falls as snow, is more difficult to measure, as wind can cause large errors (Berglöv et al., 2009).

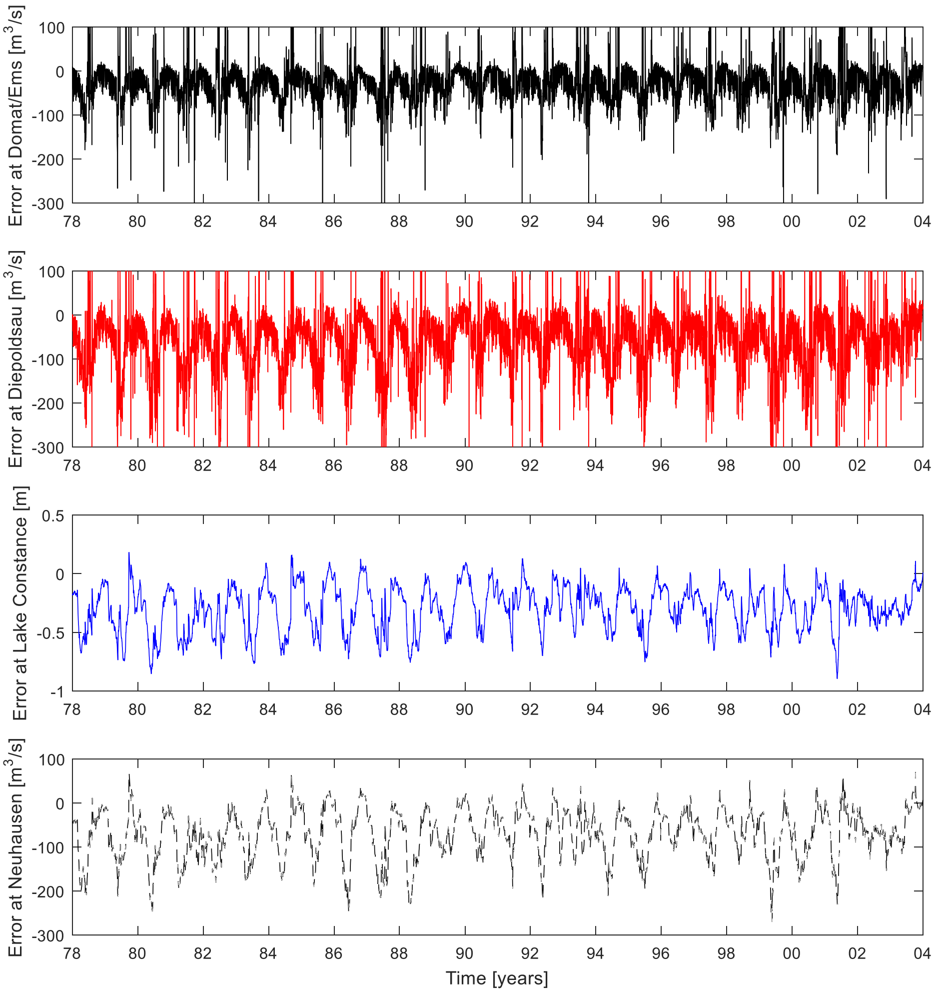


Figure 21 Graphical representation of the differences between the simulated and observed time series in the East Alpine sub-basin of discharges and lake levels at the locations Domat/Ems, Diepoldsau, Lake Constance and Neuhausen-Flurlingerbrücke (before recalibration).

In the recalibration is tried to reduce this error in spring and because it is seen in all upstream basins, focus is on the HBV sub-catchments Rhein1 and Rhein2 (upstream of Lake Constance), Rheineuh (Lake Constance) and Rhein3 (downstream of Lake Constance and with outflow location Rekingen). These are the basins in which the Alpenrhein flows.

5.2. Set up of the calibration

To determine which parameters are used in the calibration a sensitivity analysis is conducted. The sensitivity analysis is performed for catchment Rhein1, the assumption is that the other sub-catchments have similar characteristic behaviour because they are also in the Alpine area. The changes in parameters act differently with the four performance indicators as shown in Figure 22 (an increase in NSE/NSE_i is good, and an increase in RVE is bad, because they are all negative). In Table 19 the effects in the sensitivity analysis are summarized.

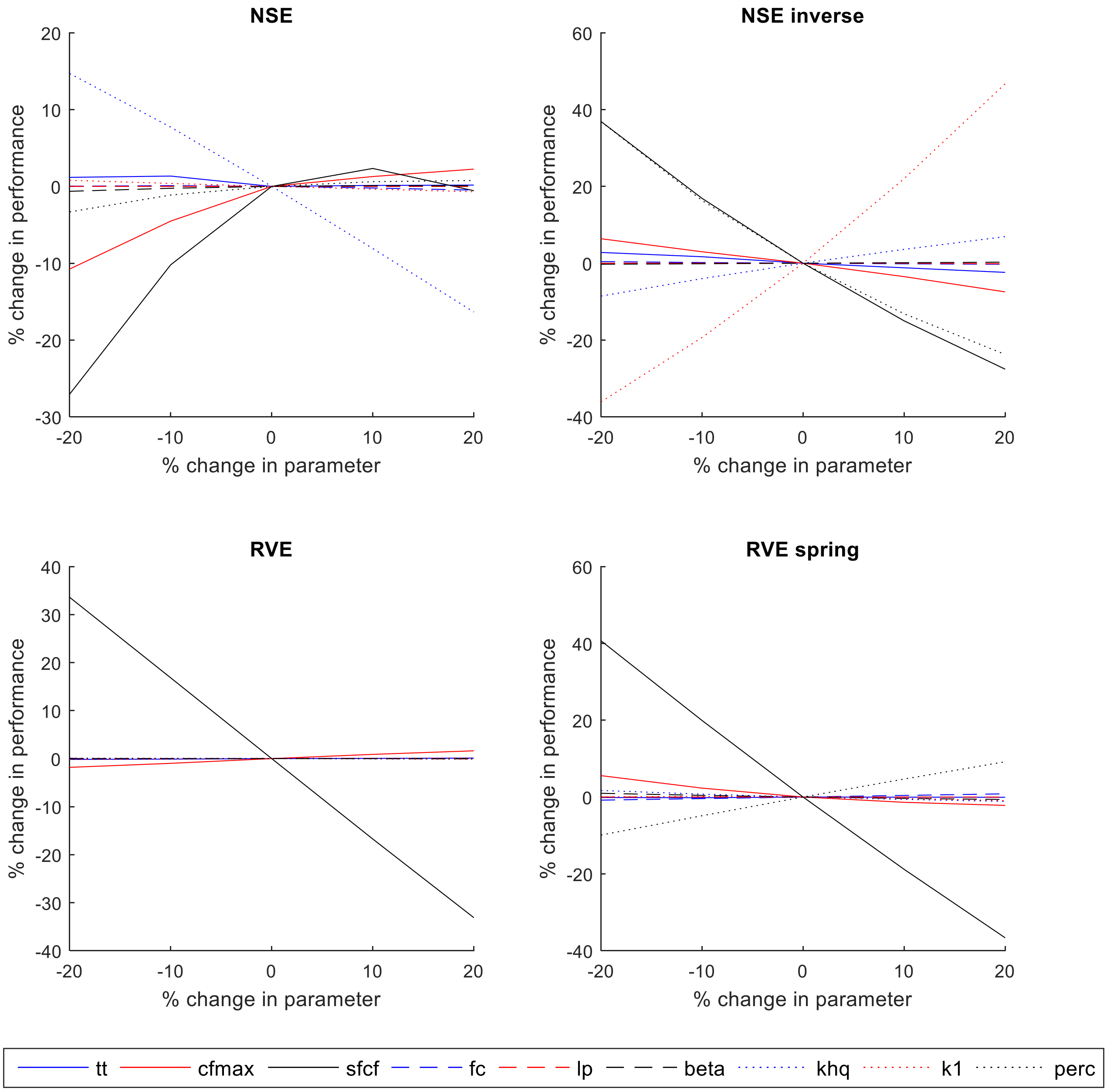


Figure 22 Relative changes of the four performance measures for relative changes of different parameters. Improvements are indicated by an increase of NSE or NSE inverse or a decrease of RVE or RVE spring.

Table 19 Positive and negative effects on the performance measures from the sensitivity analysis.

	--	-	+	++
NSE	Decrease cfmax Decrease Sfcf	Increase Khq Decrease perc	Decrease khq Increase cfmax	
NSE_i	Decrease k1 Increase Sfcf Increase perc	Decrease khq Increase cfmax	Decrease cfmax Increase khq	Increase k1 Decrease Sfcf Decrease perc
RVE	Decrease Sfcf	Increase cfmax	Decrease cfmax	Increase Sfcf
RVE_{spring}	Decrease Sfcf	Increase perc Decrease cfmax	Decrease perc	Increase Sfcf

Looking at the results of the sensitivity analysis the largest differences in the model are found for the parameters Cfmax, Sfcf, Khq, Perc and K1. These are the parameters that are calibrated. Some of them also have opposite effects on the different performance measures. Performing five thousand runs per HBV sub-catchment should cover for finding optimal parameter combinations. This is the same number as in Hegnauer et al. (2014) and Davids et al. (2015).

The parameters used in the recalibration together with their lower and upper bounds are presented in Table 20. The values for these boundaries are taken from other studies of HBV calibrations, and are considered realistic.

Table 20 Parameters used for the calibration of the Alpine basin with lower and upper boundary and references to studies with the same boundaries.

			Min	Max	
Cfmax	Degree day factor melt	Mm/day/°C	1.0	6.0	(Hegnauer & van Verseveld, 2013)
Sfcf	Snowfall correction factor	-	0.7	1.4	(Berglöv et al., 2009)
Khq	Recession parameter at high flow level hq	1/day	0.005	1.0	(Berglöv et al., 2009; Davids et al., 2015; Hegnauer & van Verseveld, 2013)
Perc	Percolation from upper to lower reservoir	Mm/day	0.01	5.5	(Berglöv et al., 2009; Demirel et al., 2013a; Hegnauer & van Verseveld, 2013)
K1	Recession coefficient lower reservoir	1/day	0.0005	0.2	(Demirel et al., 2013a)

5.3. Results of the calibration

After every calibration of the sub-catchment the Pareto front of the four performance measures NSE, NSE inverse, RVE and RVE spring, is determined. The boundary thresholds are shown in Table 21. When applying the boundaries of NSE_i>0.5, RVE<0.1 and RVE_{spring}<0.1 to the results of the basins Rhein 1 and Rhein 2, there are not enough behavioural sets because it is difficult to get a high performance at these locations with unnatural flow patterns (Hegnauer & van Verseveld, 2013).

Table 21 Thresholds of performance measures to bound the behavioural sets for the four HBV sub-catchments , and the resulting number of behavioural sets.

Performance measure	Threshold Rhein 1	Rhein 2	Rheineuh	Rhein 3
NSE	-	-	-	-
NSE_i	> - 0.5	> 0.5	> 0.5	> 0.5
RVE	< 0.25	< 0.15	< 0.1	< 0.1
RVE_{spring}	< 0.25	< 0.15	< 0.1	< 0.1
Final number of behavioural sets	11	12	22	63

The results on the severity for the growing season threshold are displayed for all 63 behavioural sets at Rekingen in Figure 23. The severity of low flow events in the growing season is, except for two sets, largely overestimated but lower than the original calibration. The 5%, 50% and 95% values based on the 1/10 year event are likely to have underestimations of the discharge. The set that is, based on visual inspection, closest to the observations is the dashed line in Figure 23. This set is chosen to be evaluated further. In Table 22 can be seen that this set does not give very different values for the performance measures as the 5%, 50% and 95% sets. In Table 23 the old and new parameter values can be found. Here can be seen that especially the snow factor has increased.

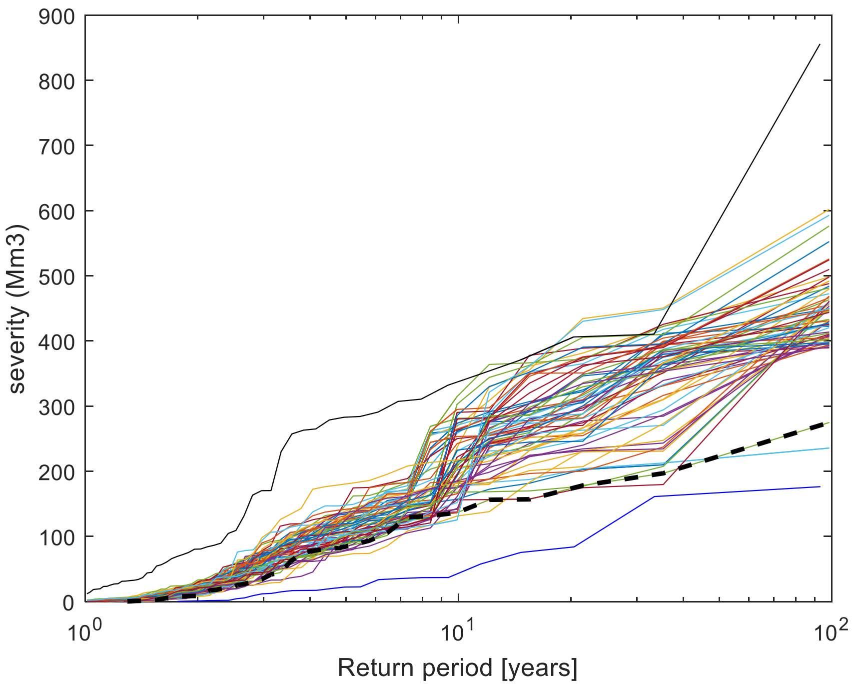


Figure 23 Return periods for the severity of low flow events in the growing season at Rekingen for the observations (blue), original historical simulation (black) and the 63 behavioural sets (different colours). The black dashed line is the visual inspection set.

Table 22 Performance measures for the 5%, 50% and 95% behavioural sets of the recalibration plus the set that gives the best result based on visual inspection (visual) based on the severity of the 1/10 year event in the growing season.

NSE	5%	50%	95%	visual	NSE_i	5%	50%	95%	visual
Rhein1	0.07	-0.64	-0.43	-0.64	**Rhein1**	0.14	0.13	0.04	0.13
Rhein2	0.60	0.52	0.49	0.63	**Rhein2**	0.10	0.07	0.24	0.34
Rheineuh	0.85	0.89	0.85	0.89	**Rheineuh**	0.62	0.75	0.79	0.78
Rhein3	0.88	0.92	0.77	0.92	**Rhein3**	0.86	0.88	0.83	0.91

RVE	5%	50%	95%	visual	RVE_{spring}	5%	50%	95%	visual
Rhein1	-0.10	-0.10	-0.11	-0.10	**Rhein1**	-0.08	-0.11	-0.06	-0.11
Rhein2	-0.11	-0.12	-0.12	-0.12	**Rhein2**	-0.08	-0.09	-0.09	-0.11
Rheineuh	-0.09	-0.10	-0.10	-0.10	**Rheineuh**	-0.12	-0.10	-0.08	-0.13
Rhein3	-0.02	-0.02	-0.02	-0.02	**Rhein3**	-0.04	-0.03	0.00	-0.04

Table 23 Parameter values of the recalibration and the original calibration.

		Rhein 1	Rhein 2	Rheineuh	Rhein 3
cfmax	org	2.737	2.737	1.427	3.597
	new	**2.050**	**3.717**	**4.370**	**1.899**
Sfcf	org	1.000	1.000	1.000	1.000
	new	**1.376**	**1.331**	**1.333**	**1.281**
Khq	org	0.164	0.164	0.696	0.674
	new	**0.788**	**0.058**	**0.366**	**0.051**
K1	org	0.010	0.010	0.010	0.010
	new	**0.002**	**0.075**	**0.138**	**0.007**
perc	org	2.172	2.172	1.264	5.075
	new	**3.478**	**4.349**	**2.564**	**2.130**

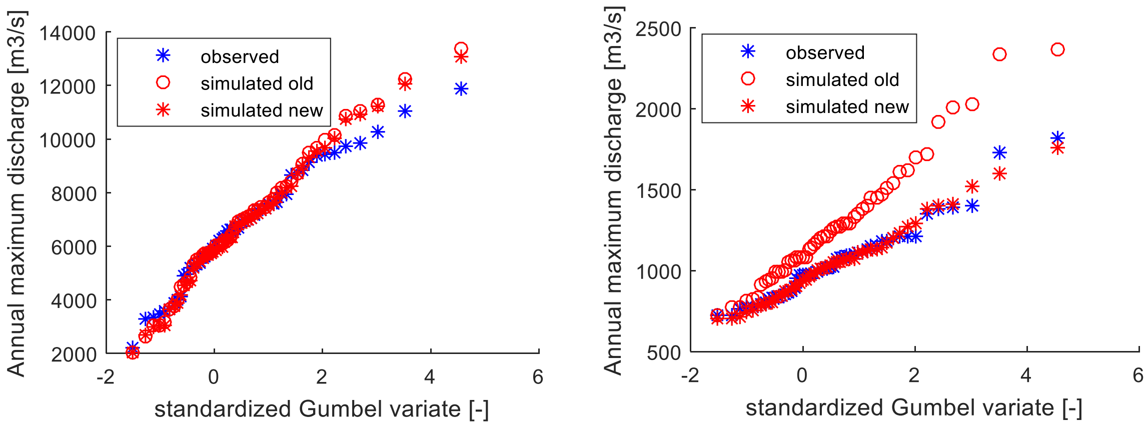
5.4. Validation

In Appendix B the tables and figures that show the validation results of the visual inspection set can be found.

With the recalibration there is less underestimation of discharge at Rekingen, and the autocorrelation length is closer to the observations, meaning that the persistence in discharge is also simulated better. This better simulation of the discharge pattern makes that there are more hits and less false alarms (also reflected in the lower number of events). The MARE's are smaller and the R^2's larger, so there are less errors in the matched events. Also the timing has improved. This leads to better simulation of lake levels and eventually the return periods. The snow simulation has improved, but not much, and the groundwater correlation is lower than before.

At Lobith the effects of the recalibration of the Alpenrhein are of course smaller, because of the contributions of the other sub-basins, but they are present. The underestimation of medium discharges is less, and the autocorrelation length has changed. The improvement in summer is clear, looking at the improved number of events under the growing season thresholds, the increased number of hits and the lower number of false alarms. But in the winter there are more events, and the return periods of severity under the navigation threshold are slightly worse. This is compensated with the improvement in summer.

Looking at the peak flows from the newly calibrated series the peak flows are better represented than in the old simulation (Figure 24). They are not overestimated anymore. From the 63 behavioural sets, there are 44 sets that give smaller yearly maximum values than the original simulations. There are four sets that give smaller values than the observations. That means that there are forty sets (the majority) with yearly maximum values with better performance than the original simulation.



Lobith **Rekingen**

Figure 24 Annual maximum discharges expressed as a function of standard Gumbel variate.

The good performance on high flows (giving lower peaks) can be caused by the groundwater parameters. In the original calibration these parameters were used to counteract on the missing snow. Now the snow is added, the other parameters do not have to correct for it anymore and can be set in a way that is more appropriate for the whole series of flow. Besides the better performance for low flows, the recalibrated set also lowered the annual peaks at Rekingen and therefore improved the performance for high discharges. The highest peaks at Lobith have decreased a little, bringing them closer to the observed values.

6. Discussion

The discussion is divided in four parts. In section 6.1 the value of the research is discussed, in 6.2 and 6.3 respectively the methodology and the results are evaluated and in section 6.4 this study is looked at in a broader perspective , linked to other research.

6.1. Value of the research

This research adds to the knowledge of the Rhine system and especially the low flows. Information about duration and severity under LCW thresholds with synthetic weather input has a large potential to give more information than characteristic years, and there is a possibility of GRADE to use it as replacement for using extreme value distributions for low flows as well as for floods. This is a completely new approach to low flow analysis in the Netherlands. In other river basins, such as the Thames (UK), the Volga (Russia) and the Poudre river (Colorado, USA) synthetic series are already used to assess return periods of low flow events or vulnerabilities of the water systems (Bolgov & Korobkina, 2011; Borgomeo et al., 2015a; Salas et al., 2005).

The benefit of GRADE for low flow frequency analysis in the Netherlands is that no distribution function has to be chosen and fitted. Although there are very advanced techniques for applying distribution functions, they are still very sensitive to certain outliers and they give no information about the hydrological and meteorological background of the low flows. With GRADE also information about the development of extreme low flow events can be derived.

6.2. Methods

In this research the calibrated model that is used in GRADE is chosen as starting point of the evaluation and the recalibration. A recalibration of the HBV model of the Rhine already existed, made by Davids et al. (2015). That hydrological model is however calibrated on an hourly time step base and only for the summer, while in this study only daily time steps are used and low flows are evaluated throughout the entire year.

The choices made in the low flow definition are quite important in the evaluation. Because the threshold level is a fixed value, the difference between a hit and a false alarm can be quite small, and it can be seen as subjective. In the synthetic series this is no different, but then cannot be checked whether it is a hit or a false alarm. The pooling procedure may be focussed more on the user's perspective when using it in practice. This depends on the information that the user requires.

In this study the return periods of duration and severity are evaluated separately. There is chosen not to incorporate the relation between these two and the joint probability in the evaluation. There are too less events to give a good representation. There is a correlation between the duration and the severity, but severities with an event of a certain duration can differ much. This depends on the discharge during the event. Also the joint probability is not of use to every user.

In the recalibration a lot of choices have been made that influence the results. The simulation of the lake level Bodensee had a very bad performance, and this is a large part of the East Alpine sub-basin. Therefore is chosen to try to improve the performance in this basin. There is also an error found in the West Alpine sun-basin (in the growing season), in the Moselle (in the summer) and in the Main (in the persistence of the discharge). Also the performance of these sub basins could benefit from a more in-depth analysis and recalibration. The East Alpine sub-basin is chosen because of reasons of time and curiosity about Lake Constance of the researchers. Also in the operational model often a deviation is seen between the observed lake levels and the simulated lake levels. Not all HBV sub-catchments are taken into account in the recalibration. So it could be that the current calibrated parameters balance the uncalibrated upstream catchments (Schugerm, Argegerm, Bregaust and Thur).

The fact that the snow correction factor seems so important in the calibration and leads to better performance than in the original GRADE model, also for the Nash-Sutcliffe Efficiency, leads to think that the assumption made in the reports of the GLUE analysis in GRADE (Hegnauer & van Verseveld, 2013; Winsemius et al., 2013) is open for discussion. They say that good results are established without precipitation correction factors and poorer results could be explained by other factors. But in the case of snow in the Alps, this can be invalid: snow is difficult to measure and it can differ a lot per location, also there is a negative relative volume error in the discharges as well as in Lake Constance. So applying a correction factor can be justified, and has been improved with a correction factor, resulting in more snow and more melt water. A maximum snowfall correction factor (Sfcf) of 1.4 is chosen because that was considered still realistic. From the recalibration results appears that all the behavioural sets have Sfcf values in the range 1.2 – 1.4. It could be that with a higher value for the upper bound more behavioural sets with good performance can be found. There is also still an underestimation of discharge in spring, so maybe even larger Sfcf would be more representative. It is advised to test this in next calibrations. Also satellite images can be used to estimate the snowfall correction factor, as is also done in the study of Shrestha et al. (2014).

6.3. Results

Because of the standardisation the snow indices and the groundwater indices can only be compared on the deviation from the mean value. Because of the small number of snow stations in the observations it is also questionable whether the simulations, that give a value for a larger area, should be compared at all with the station data, although they are both averaged for the large East and West Alpine sub-basins. Also the change in density of the snow pack, that is in the observed snow depth, cannot be taken into account in comparisons with the simulations.

Because of the standardisation it is also difficult to say if the snow simulation itself has improved in the recalibration, using the snowfall correction factor. This will only change the mean and the standard-deviation, but not the standardised index. What is remarkable of the snow simulation is the snow accumulation throughout the years, while the snow depth in the Alps show a decreasing trend from 1980 (Laternser & Schneebeli, 2003). After the recalibration the HBV model shows more snow accumulation throughout the years (0.94 mm/year) than in the original calibration, in spite of a better performance on discharge simulation.

In the weather generator only values averaged over the whole basin determine the resampling and are used to validate the extreme maximum values generated (Schmeits et al., 2014). For producing annual flood statistics this is good enough, and the weather generator also simulated dry periods in this way, so that low flows occur. When looking at the dry periods and the snow simulation of the West Alps, the representation of the weather is significantly different from the observations, resulting in differences in the flow. Although differences in low flow events are small, for better representation of low flows in the Rhine, the weather generator should be altered to incorporate (local) persistence of the weather. The weather generator with a four-month memory term in it (Beersma & Buishand, 2007) can be used as improvement, as also mentioned in GRADE report (Hegnauer et al., 2014). The persistent weather conditions then can be simulated better. For individual low flow events the spatial distribution is more important, so the validation of the weather generator my include different regions in the basin.

The comparison of the simulations with the observations is only valid when the measurements in the observed time series are reliable. In this case it is assumed that there are no large errors in the measurements of low flows. However, regulation of flow cannot be seen in the discharge series, and this can play an important role in low flows. As is seen in the results especially at Cochem regulation could play a role in the underestimation of flow by the simulations.

Comparing the observations with the synthetic simulations has its limitations. Because there are two factors in the synthetic simulations (the weather generator and the hydrological model) that balance out some effects. Because there are more low flows due to the hydrological model and less low flows due to the weather generator, the synthetic series can be similar to the observations in some characteristics. However, this similarity is based on two errors, so is not reliable to draw conclusions from.

6.4. Other research

De Wit et al. (2007) concluded in their research that a model with good performance for average and high discharges may not be adequate to simulate low flows. There are different processes occurring in the basin and contributing to the flow. Often HBV can only be tuned either for high flow processes or for low flow processes. Also in the recalibration of the HBV model for the Rhine by Davids et al. (2015) is advised to use two different models; one for high flows and one for low flows. The same is seen in this study; while the model performance is acceptable for peak flows, there are larger errors in the low flow simulations.

Discharge generators that are especially made for low flows do not have this problem of models for floods not being suitable for low flows. They are validated on low flow statistics. The benefit from using a weather generator and a hydrological model for synthetic series is that meteorology is resampled and has no other values than ever observed, but the discharges can be different from observed, without using a fitted distribution. This is the same method as Brocca et al. (2013) and Liersch and Volk (2008) used for flood risk assessment, but in GRADE a continuous time series is used. Because not only a database of possible events is collected, but also the probability of occurrence in a time period is estimated from it. In the Rhine basin also the timing of events within the basin are of importance.

The reality of climate change means that there is a need to go beyond methods that are based on historical data alone require synthetic discharge generators that can reproduce observed statistics and then can be adapted to generate discharges with other properties than the observed data (Borgomeo et al., 2015a). Making the current model applicable for low flow statistics simulation is the first start in assessing the implications of climate change on low flows. GRADE is already accepted for floods, and for floods also a synthetic series with climate change incorporated is already evaluated (Sperna Weiland et al., 2015). As De Wit et al. (2007) and Demirel et al. (2013c) found in their studies, climate change also affects low flows in the Rhine and Meuse basins.

7. Conclusions

In this research the skill of the existing model combination GRADE (Generator of Rainfall and Discharge Extremes) is evaluated for low flow events at the outflow locations of seven mayor sub-basins of the river Rhine. The low flow events are defined with threshold levels based on the thresholds of the Dutch National Committee of Water allocation and are divided into constant year-round thresholds and higher thresholds for only the growing season to take impact of low flows on agriculture into account. The skill of the HBV model, the skill of the weather generator and the skill of the combination of weather generator and HBV model are studied. Furthermore a part of the HBV model has been improved for the skill on low flows.

7.1. Skill of HBV

When looking at the skill of HBV in simulating low flows it is seen that there is usually an underestimation of the discharge in the lower spectrum. That causes more low flows than there are observed and larger severities of low flow events. The return periods of the durations are simulated quite well at Lobith, Andernach, Frankfurt (Main) and Rockenau (Neckar). In the Neckar the discharges are simulated well and the underestimation is less. In the West-Alpine sub-basin there is, next to underestimation in the summer, overestimation of discharges in the winter. This causes less low flow events. There is a large overestimation of the discharges in summer in the Alpine sub-basins. This is related to the snow simulation. The lake levels, snow covers and groundwater levels, that are simulated in the conceptual HBV model, correlate well with the observed levels when standardized. The variations through the years can be seen.

7.2. Skill of weather generator

Although the weather generator is not made to simulate dry periods and is not evaluated for it, it does generate dry periods. But these synthetic dry periods are not extreme enough to simulate the same number of low flows and the same durations and severities as with the historical weather series. The snow simulation is not adequate at all locations, while this is important for the flows in summer.

7.3. Skill of GRADE

Because of the increase in low flows caused by the HBV model and the decrease in low flows caused by the weather generator the synthetic simulations of low flows are closer to the observations than the historical simulations. The less dry periods compensate for the underestimation of discharge. But still there are large errors. Good performances are realised for the discharges and the return periods of severity and duration in the Neckar sub-basin at Rockenau, the return periods of severity and duration of events under the navigation thresholds at the Alpine discharge locations and the return periods of durations of events at Andernach and Lobith. The performances are however based on the compensation of two errors. So the model performs well for the wrong reasons and it is doubted if the information from the model is usable.

7.4. Improvement of the skill

With the recalibration of four East-Alpine sub-catchments of the HBV model is shown that the performance for low flows can be improved. After the recalibration the underestimation of discharges is less, and therefore the skill of the model to simulate low flow characteristics as duration and severity is better. Especially large improvements are gained in the summer and spring. The snow simulation played a large role in this, although it is not sure whether the simulation of snow itself has become better.

7.5. Recommendations and further research

Because of the described differences between the synthetic simulations and the observations of the low flow characteristics in the Rhine basin, it is not recommended to use GRADE in its current condition for estimating probabilities of low flows. The information about low flows that can be drawn from the synthetic series, is not representing the observations. But this study does show that there is a large potential to use an adapted version of GRADE for gaining insight in low flows in the Netherlands. Low flow policies can be connected to the probabilities of low flows and measures can be proposed based on evaluation of generated extreme low flow events.

The improved models should have a better skill on low flows. The hydrological model can be improved by e.g., a recalibration of the whole basin. The weather generator can be improved by adding a memory term to the algorithm to simulate more persistence in the dry weather.

Further research into the model can be useful for the improvements. For example satellite images can be used to estimate the snowfall correction factor. Also it could be interesting to look at the influence of hydraulic structures such as weirs in the low flow conditions.

When GRADE is adapted for low flows, it becomes more suitable to run this with synthetic series based on climate scenarios. This gives extra information about the change in low flows that can be expected in the future.

Now the analysis only includes the Rhine. For the Meuse the same analysis can be done. The Meuse is a rain-dominated river so problems in the snow simulation will not have a large impact. For the Meuse also LCW thresholds exist, these can be used for the analysis.

References

Anderson, O. D. (1976). *Time series analysis and forecasting : the Box-Jenkins approach*. London: Butterworth.

Beersma, J. J., & Buishand, A. (2002). Droog, droger, droogst. Bijdrage van het KNMI aan de eerste fase van de Droogtestudie Nederland *KNMI-publicatie 199-I*. De Bilt.

Beersma, J. J., & Buishand, T. A. (2007). Drought in the Netherlands - Regional frequency analysis versus time series simulation. *Journal of Hydrology, 347*(3-4), 332-346. doi: 10.1016/j.jhydrol.2007.09.042

Belz, J.U.; Brahmer, G.; Buiteveld, H.; Engel, H.; Grabher, R.; Hodel, H.; Krahe, P.; Lammersen, R.; Larina, M.; Memdel, H.-G.; Meuser, A.; Müller, G.; Plonka, B.; Pfister, L.; Van Vuuren, W. . (2007). Das Abflussregime des Rheins und seiner Nebenflüsse im 20. Jahrhundert - Analyse,Veränderungen, Trends. CHR report no. I-22. Lelystad.

Berglöv, G., German, J., Gustavsson, H., Herbman, U., & Johansson, B. (2009). Improvement HBV model Rhine in Fews Final report *Hydrology* (Vol. 112): SMHI, BfG.

Bolgov, M., & Korobkina, E. (2011). Low-flow risk assessment for water management. In G. Bloschl, K. Takeuchi, S. Jain, A. Farnleitner & A. Schumann (Eds.), *Risk in Water Resources Management* (Vol. 347, pp. 77-84). Wallingford: Int Assoc Hydrological Sciences.

Bolwidt, L., Schoor, M., van Hal, L., & Roukema, M. (2006). *Laagwater op de Rijn en de Maas*. Lelystad: Ministerie van Verkeer en Waterstaat, Rijkswaterstaat.

Booij, M. J. (2005). Impact of climate change on river flooding assessed with different spatial model resolutions. *Journal of Hydrology, 303*(1–4), 176-198. doi: 10.1016/j.jhydrol.2004.07.013

Borgomeo, E., Farmer, C. L., & Hall, J. W. (2015a). Numerical rivers: A synthetic streamflow generator for water resources vulnerability assessments. *Water Resources Research, 51*(7), 5382-5405. doi: 10.1002/2014wr016827

Borgomeo, E., Pflug, G., Hall, J. W., & Hochrainer-Stigler, S. (2015b). Assessing water resource system vulnerability to unprecedented hydrological drought using copulas to characterize drought duration and deficit. *Water Resources Research, 51*(11), 8927-8948. doi: 10.1002/2015wr017324

Brocca, L., Liersch, S., Melone, F., Moramarco, T., & Volk, M. (2013). Application of a model-based rainfall-runoff database as efficient tool for flood risk management. *Hydrology and Earth System Sciences, 17*(8), 3159-3169. doi: 10.5194/hess-17-3159-2013

Cioc, M. (2002). *The Rhine: an eco-biography, 1815-2000*. Seattle: University of Washington Press.

Crochemore, Louise, Perrin, Charles, Andréassian, Vazken, Ehret, Uwe, Seibert, Simon P., Grimaldi, Salvatore, . . . Paturel, Jean-Emmanuel. (2015). Comparing expert judgement and numerical criteria for hydrograph evaluation. *Hydrological Sciences Journal, 60*(3), 402-423. doi: 10.1080/02626667.2014.903331

Davids, F., Ten Velden, C., Eilander, D., & Van Verseveld, W. (2015). Kalibratie en aanpassingen HBV model voor de Rijn voor laagwater *Deltares report*. Delft, Netherlands.

De Wit, M. J. M., van den Hurk, B., Warmerdam, P. M. M., Torfs, P.J.J.F., Roulin, E., & van Deursen, W. P. A. (2007). Impact of climate change on low-flows in the river Meuse. *Climatic Change, 82*(3-4), 351-372. doi: 10.1007/s10584-006-9195-2

Demirel, M. C., Booij, M. J., & Hoekstra, A. Y. (2013a). Effect of different uncertainty sources on the skill of 10 day ensemble low flow forecasts for two hydrological models. *Water Resources Research, 49*(7), 4035-4053. doi: 10.1002/wrcr.20294

Demirel, M. C., Booij, M. J., & Hoekstra, A. Y. (2013b). Identification of appropriate lags and temporal resolutions for low flow indicators in the River Rhine to forecast low flows with different lead times. *Hydrological Processes, 27*(19), 2742-2758. doi: 10.1002/hyp.9402

Demirel, M. C., Booij, M. J., & Hoekstra, A. Y. (2013c). Impacts of climate change on the seasonality of low flows in 134 catchments in the River Rhine basin using an ensemble of bias-corrected regional climate simulations. *Hydrology and Earth System Sciences, 17*(10), 4241-4257. doi: 10.5194/hess-17-4241-2013

Demirel, M. C., Booij, M. J., & Hoekstra, A. Y. (2015). The skill of seasonal ensemble low-flow forecasts in the Moselle River for three different hydrological models. *Hydrology and Earth System Sciences, 19*(1), 275-291. doi: 10.5194/hess-19-275-2015

ENW. (2015). Advies GRADE en afvoerstatistiek *ENW-advies*. Utrecht: Expertise Netwerk Waterveiligheid.

Görgen, K., Beersma, J., Brahmer, G., Buiteveld, H., Carambia, M., de Keizer, O., . . . Volken, D. (2010). Assessment of Climate Change Impacts on Discharge in the Rhine River Basin: Results of the RheinBlick2050 Project *CHR report, I-23*. Lelystad.

GRDC. (2010). *Discharge data Rhine* Retrieved from: Global Runoff Data Centre, Federal Institute of Hydrology (BfG), Koblenz

Gudmundsson, L., Tallaksen, L. M., & Stahl, K. (2011). Spatial cross-correlation patterns of European low, mean and high flows. *Hydrological Processes, 25*(7), 1034-1045. doi: 10.1002/hyp.7807

Guo, Jun, Zhou, Jianzhong, Lu, Jiazheng, Zou, Qiang, Zhang, Huajie, & Bi, Sheng. (2014). Multi-objective optimization of empirical hydrological model for streamflow prediction. *Journal of Hydrology, 511*, 242-253. doi: 10.1016/j.jhydrol.2014.01.047

Hegnauer, M., & Becker, A. (2013). Technical Documentation GRADE part II Models Rhine *Deltares report*. Delft, Netherlands.

Hegnauer, M., Beersma, J.J., van den Boogaard, H.F.P., Buishand, T.A., & Passchier, R.H. (2014). Generator of Rainfall and Discharge Extremes (GRADE) for the Rhine and Meuse basins; Final report of GRADE 2.0 *Deltares report*. Delft, Netherlands: .

Hegnauer, M., & van Verseveld, W. (2013). Generalised Likelihood Uncertainty Estimation for the daily HBV model in the Rhine Basin, part B: Switzerland *Deltares report*. Delft, Netherlands.

Jörg-Hess, S., Kempf, S. B., Fundel, F., & Zappa, M. (2015). The benefit of climatological and calibrated reforecast data for simulating hydrological droughts in Switzerland. *Meteorological Applications, 22*(3), 444-458. doi: 10.1002/met.1474

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

Laternser, M., & Schneebeli, M. (2003). Long-term snow climate trends of the Swiss Alps (1931-99). *International Journal of Climatology, 23*(7), 733-750. doi: 10.1002/joc.912

LCW. (2015). Laagwaterbericht Nr. 2015-2 *LCW laagwaterberichten*: Landelijke Commissie Waterverdeling.

Leander, R., & Buishand, T. A. (2004). Rainfall generator for the Meuse basin - Development of a multi-site extension for the entire drainage area *KNMI-publicatie 196-III*. De Bilt.

Liersch, S., & Volk, M. (2008). *Application of a model-based rainfall-runoff database as efficient tool for flood risk management*. Paper presented at the International Congress on Environmental Modelling and Software, Barcelona, Catalonia.

Lindstrom, G., Johansson, B., Persson, M., Gardelin, M., & Bergstrom, S. (1997). Development and test of the distributed HBV-96 hydrological model. *Journal of Hydrology, 201*(1-4), 272-288. doi: 10.1016/s0022-1694(97)00041-3

Middelkoop, H., de Boo, M., Breukel, R. M. A., & van Haselen, C. O. G. (1999). Twice a river : Rhine and Meuse in the Netherlands *RIZA report; no. 99.003*. Arnhem.

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

Nicolle, P., Pushpalatha, R., Perrin, C., François, D., Thiéry, D., Mathevet, T., . . . Morice, E. (2014). Benchmarking hydrological models for low-flow simulation and forecasting on French catchments. *Hydrol. Earth Syst. Sci., 18*(8), 2829-2857. doi: 10.5194/hess-18-2829-2014

Önöz, B., & Bayazit, M. (2002). Troughs under threshold modeling of minimum flows in perennial streams. *Journal of Hydrology, 258*(1-4), 187-197. doi: 10.1016/s0022-1694(01)00562-5

Pushpalatha, R., Perrin, C., Le Moine, N., & Andreassian, V. (2012). A review of efficiency criteria suitable for evaluating low-flow simulations. *Journal of Hydrology, 420*, 171-182. doi: 10.1016/j.jhydrol.2011.11.055

Querner, E.P., Tallaksen, L.M., Kaspárek, L., & van Lanen, H. A. J. (1997). *Impact of land-use, climate change and groundwater abstraction on streamflow droughts using physically-based models.* Paper presented at the FRIEND '97 - Regional Hydrology: Concepts and Models for Sustainable Water Resource Management.

Rauthe, Monika, Steiner, Heiko, Riediger, Ulf, Mazurkiewicz, Alex, & Gratzki, Annegret. (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}. doi: 10.1127/0941-2948/2013/0436

RIZA, HKV, Arcadis, KIWA, Hovelynck, Korbee en, Klopstra, D., . . . Kroon, T. (2005). Aard, ernst en omvang van de droogte in Nederland: eindrapport: Droogtestudie Nederland *RIZA report; no. 2005.016*.

Salas, J. D., Fu, C. J., Cancelliere, A., Dustin, D., Bode, D., Pineda, A., & Vincent, E. (2005). Characterizing the severity and risk of drought in the Poudre River, Colorado. *Journal of Water Resources Planning and Management-Asce, 131*(5), 383-393. doi: 10.1061/(asce)0733-9496(2005)131:5(383)

Schmeits, M.J., Wolters, E.L.A., Beersma, J. J., & Buishand, T. A. (2014). Rainfall generator for the Rhine basin: Description of simulations using gridded precipitation datasets and uncertainty analysis *KNMI-publicatie 186-VII*. De Bilt.

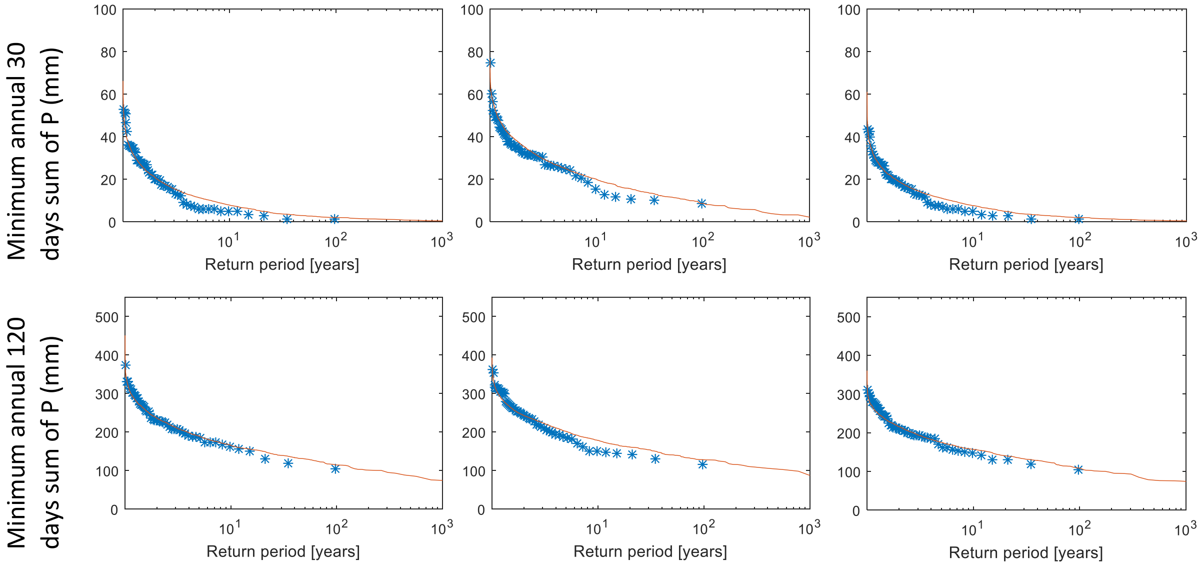
Shaw, E. M., Beven, K., Chappell, N., & Lamb, R. (2010). *Hydrology in practice 4th edition*. Abingdon, Oxon: Spon (E&F).
Shrestha, M., Wang, L., Koike, T., Tsutsui, H., Xue, Y., & Hirabayashi, Y. (2014). Correcting basin-scale snowfall in a mountainous basin using a distributed snowmelt model and remote-sensing data. *Hydrol. Earth Syst. Sci., 18*(2), 747-761. doi: 10.5194/hess-18-747-2014
Smakhtin, V. U. (2001). Low flow hydrology: a review. *Journal of Hydrology, 240*(3–4), 147-186. doi: 10.1016/S0022-1694(00)00340-1
SMHI. (2008). IHMS : Integrated Hydrological Modelling System (Manual version 4.5). Norrköping, Sweden: Swedish Meteorological and Hydrological Institute.
Sperna Weiland, F., Hegnauer, M., Bouaziz, L., & Beersma, J. J. (2015). Implications of the KNMI'14 climate scenarios for the discharge of the Rhine and Meuse *Deltares report*. Delft, Netherlands.
Staudinger, M., Stahl, K., Seibert, J., Clark, M. P., & Tallaksen, L. M. (2011). Comparison of hydrological model structures based on recession and low flow simulations. *Hydrology and Earth System Sciences, 15*(11), 3447-3459. doi: 10.5194/hess-15-3447-2011
Sung, J. H., & Chung, E. S. (2014). Development of streamflow drought severity-duration-frequency curves using the threshold level method. *Hydrology and Earth System Sciences, 18*(9), 3341-3351. doi: 10.5194/hess-18-3341-2014
Tallaksen, L. M., & Van Lanen, H. A. J. (2004). *Hydrological Drought: Processes and Estimation Methods for Streamflow and Groundwater*: Elsevier.
Te Linde, A. H., Aerts, J.C.J.H., Hurkmans, R.T.W.L., & Eberle, M. (2008). Comparing model performance of two rainfall-runoff models in the Rhine basin using different atmospheric forcing data sets. *Hydrology and Earth System Sciences, 12*(3), 943-957.
Tongal, H., Demirel, M. C., & Booij, M. J. (2013). Seasonality of low flows and dominant processes in the Rhine River. *Stochastic Environmental Research and Risk Assessment, 27*(2), 489-503. doi: 10.1007/s00477-012-0594-9
Wilks, D.S. (2006). *Statistical Methods in the Atmospheric Sciences - second edition*. USA: Elsevier.
Winsemius, H., van Verseveld, W., Weerts, A., & Hegnauer, M. (2013). Generalised Likelihood Uncertainty Estimation for the daily HBV model in the Rhine Basin, Part A: Germany *Deltares report*. Delft, Netherlands.
Woo, M. K., & Tarhule, A. (1994). Streamflow droughts of Northern Nigerian rivers. *Hydrological Sciences Journal-Journal Des Sciences Hydrologiques, 39*(1), 19-34.
Yevjevich, V. (1967). *An Objective Approach to Definitions and Investigations of Continental Hydrologic Droughts* (Vol. 23). Fort Collins, Colorado: Colorado State University.
Zelenhasic, E., & Salvai, A. (1987). A method of streamflow drought analysis. *Water Resources Research, 23*(1), 156-168. doi: 10.1029/WR023i001p00156

Appendix A – Return periods of annual minimum precipitation sums

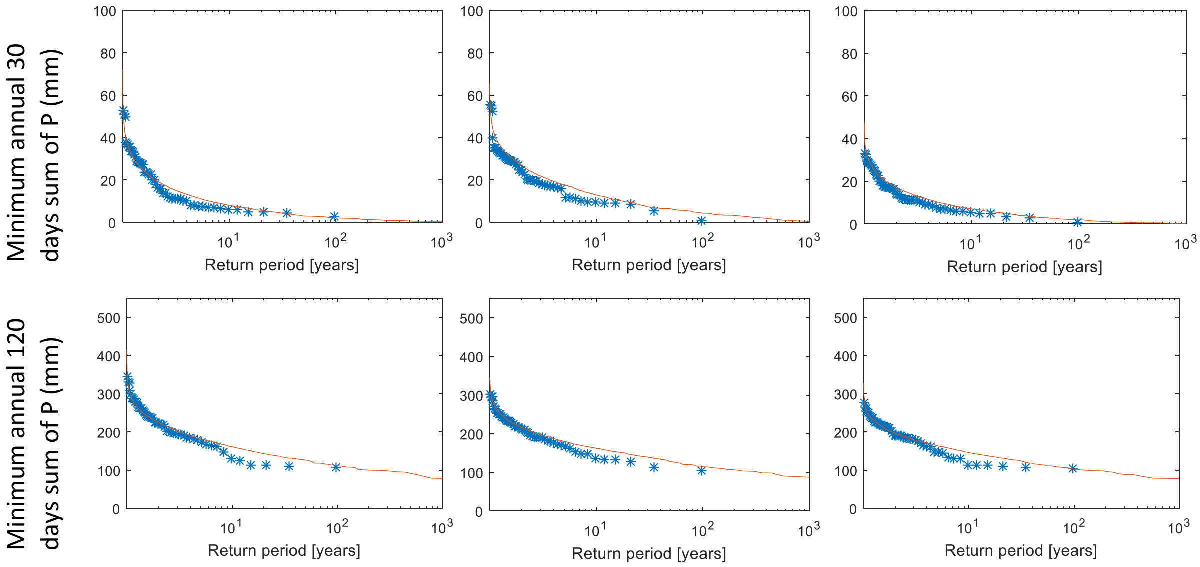
In this Appendix the graphs can be found for the return periods of precipitation sums in the different sub-basins of the Rhine (Figure 25). The observations are represented by blue dots and the synthetic weather series by a red line.

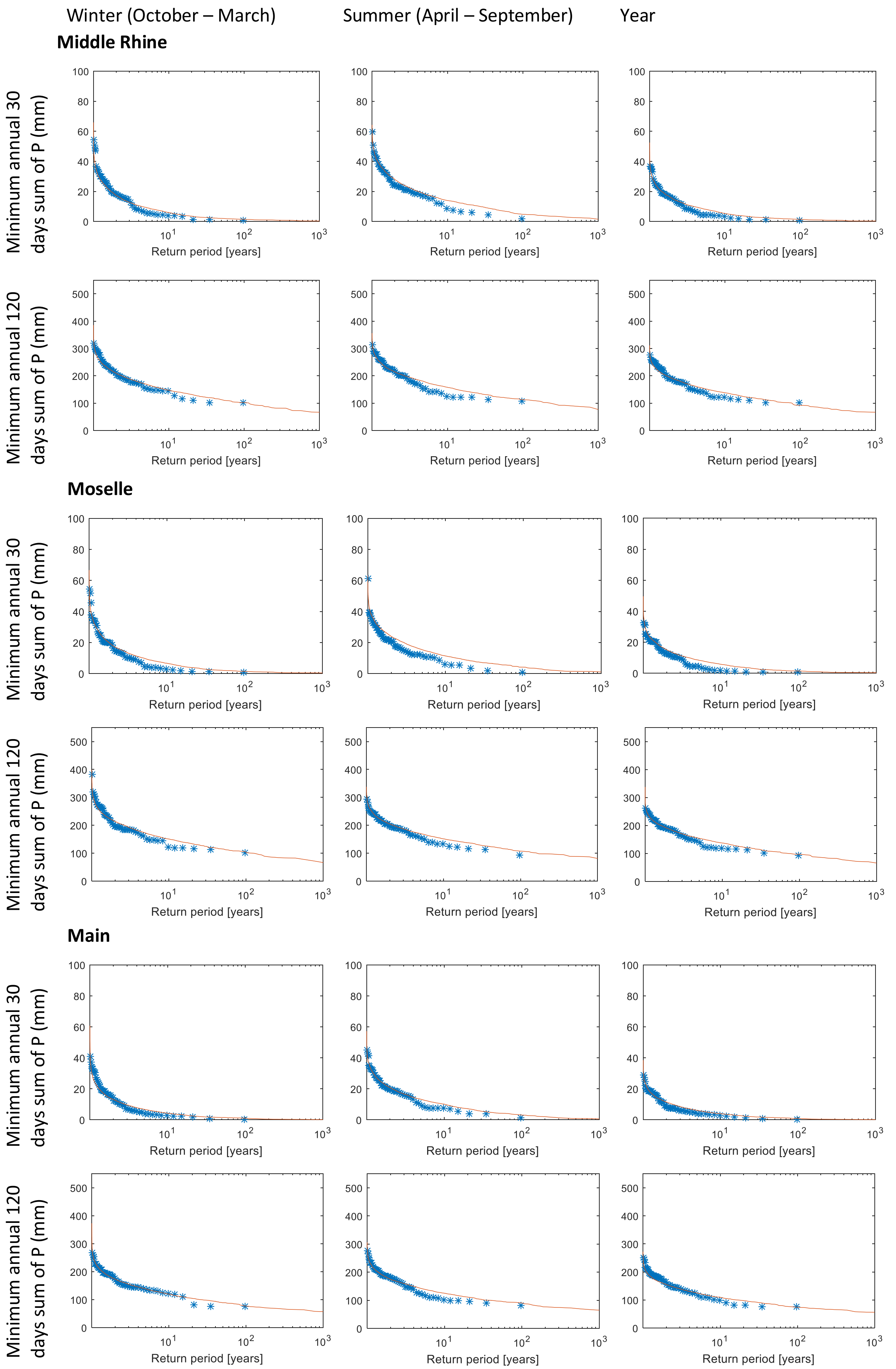
Winter (October – March) | Summer (April – September) | Year

Average at Lobith



Lower Rhine




Winter (October – March)
Summer (April – September)
Year
Middle Rhine
Moselle
Main
Minimum annual 30 days sum of P (mm)
Minimum annual 120 days sum of P (mm)
Return period [years]

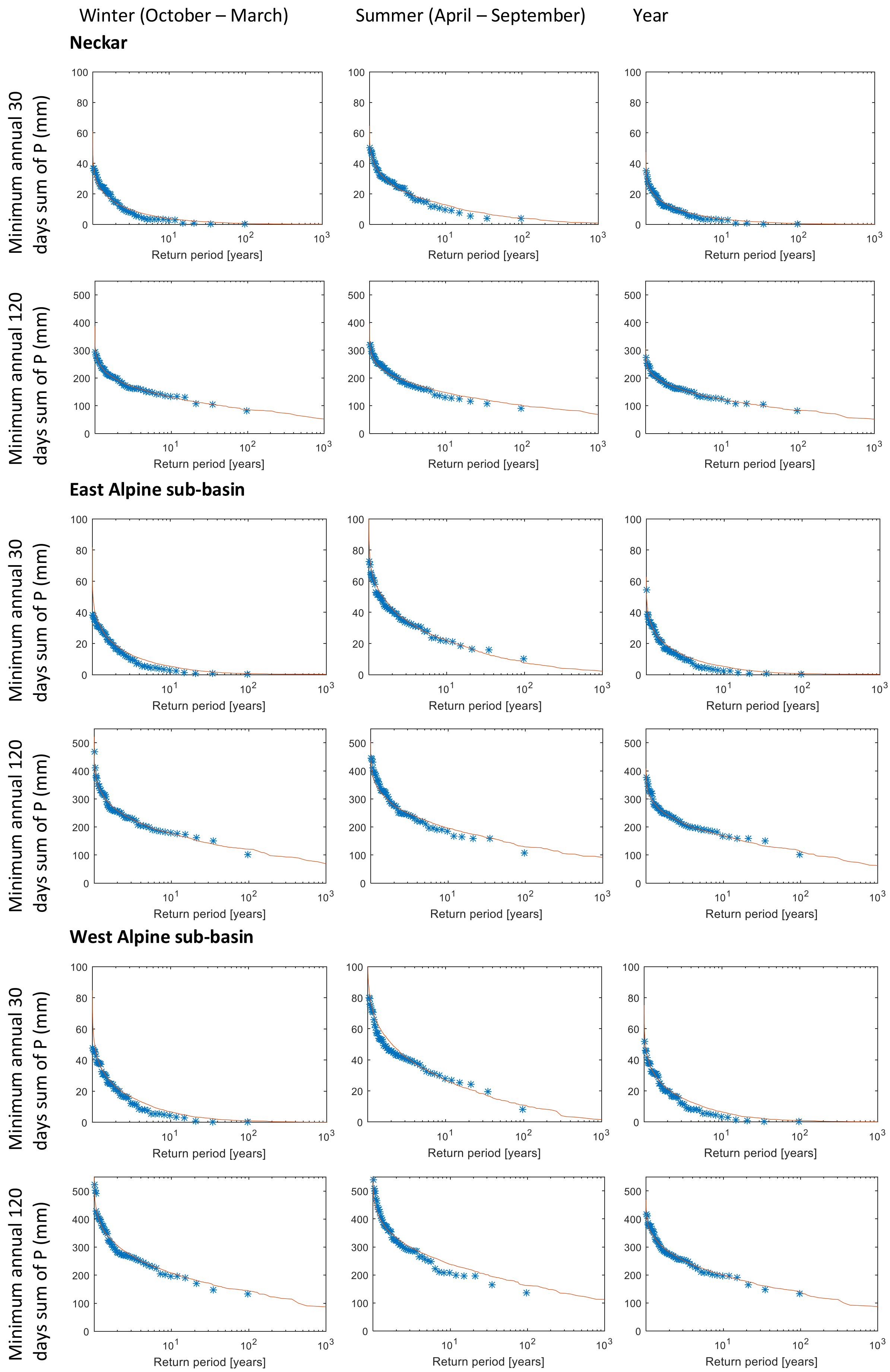


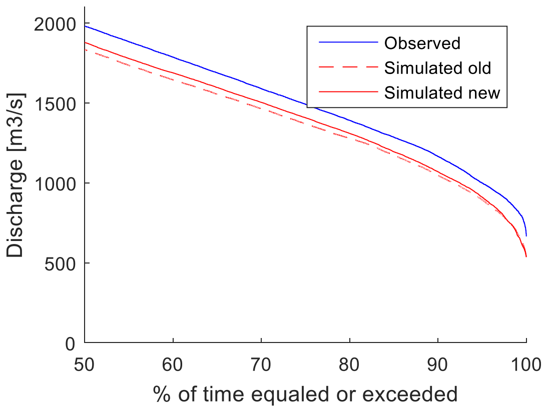
Figure 25 Probability of occurrence of annual precipitation sums for winter, summer and the entire year.

Appendix B – Validation of recalibration

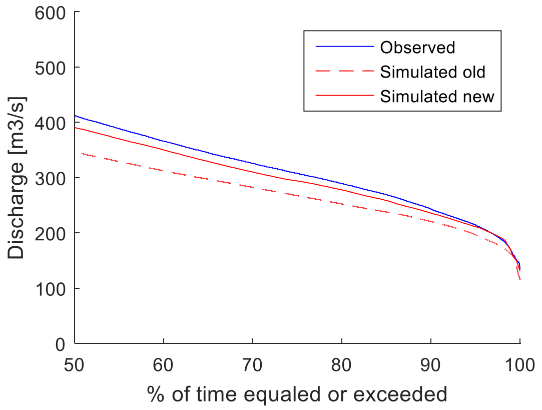
Here the figures and tables are shown for the low flow analyses presented in section 3.2 for the East Alpine basin (discharge location Rekingen) and the entire Rhine basin (discharge location Lobith).

Table 24 Performance measures for the original simulation (org), the recalibration (visual inspection set) (cal) and the validation (val)

	NSE			**NSEi**			**RVE**			**RVE spring**		
	Org	*Cal*	*Val*	*Org*	*Cal*	*Val*	*Org*	*Cal*	*Val*	*Org*	*Cal*	*Val*
Rhein 1 Domat/Ems	0.38	-0.64	-0.69	-2.09	0.13	0.14	-0.28	-0.10	-0.12	-0.29	-0.11	-0.10
Rhein2 Diepoldsau	0.57	0.63	0.57	-0.63	0.34	0.22	-0.26	-0.12	-0.12	-0.30	-0.11	-0.11
Rheineuh Neuhausen-Flurlingerbrücke	0.64	0.89	0.88	0.22	0.78	0.73	-0.19	-0.10	-0.090	-0.27	-0.13	-0.10
Rhein3 Rekingen	0.68	0.92	0.91	0.59	0.91	0.89	-0.11	-0.024	-0.042	-0.27	-0.044	-0.053
Rhine basin Lobith	0.93		0.95	0.90		0.92	-0.059		-0.043	-0.080		-0.053



Lobith



Rekingen

Figure 26 Flow duration curves of lower end of the discharges after the recalibration at Lobith and Rekingen

Table 25 Correlation length (in days) for the discharge locations

	Observation	**Historical simulation**	**Synthetic simulation**	**Recalibrated historical simulation**
Lobith	32	30	29	31
Rekingen	45	33	32	42

Table 26 Results from the contingency table for events in time series with navigation thresholds after recalibration.

Navigation threshold	**Hits**	**CSI**	**Hit rate**	**False alarm ratio**	**Duration**		**Severity**		**Timing**	
					MARE [-]	R^2 [-]	MARE [-]	R^2 [-]	Mean # days start date later	Mean # days end date later
Lobith (45)	41	0.56	0.91	0.41	1.57	0.77	18.07	0.74	-6.15	3.17
Lobith new	42	0.58	0.93	0.39	1.36	0.61	13.72	0.76	-3.45	2.14
Rekingen (40)	23	0.26	0.58	0.68	2.14	0.024	12.82	0.0066	-5.09	-0.91
Rekingen new	27	0.48	0.68	0.37	0.77	0.65	4.13	0.79	-0.074	0.33

Table 27 Results from the contingency table for events in time series with growing season thresholds after recalibration.

Growing season threshold	**Hits**	**CSI**	**Hit rate**	**False alarm ratio**	**Duration**		**Severity**		**Timing**	
					MARE [-]	R^2 [-]	MARE [-]	R^2 [-]	Mean # days start date later	Mean # days end date later
Lobith (28)	22	0.42	0.79	0.52	0.95	0.53	6.85	0.67	-5.64	5.68
Lobith val	23	0.56	0.82	0.36	0.64	0.66	4.58	0.72	-3.65	3.70
Rekingen (30)	24	0.29	0.80	0.68	5.14	0.29	52.96	0.0077	-7.46	18.29
Rekingen val	28	0.68	0.93	0.28	1.81	0.62	13.70	0.31	-1.43	6.14

Table 28 Average number of events per year in different time series for different thresholds and periods after recalibration.

Events per year		**Navigation threshold**	**Growing season threshold**	**Navigation threshold summer**	**Navigation threshold winter**
				April-September	October-March
Lobith	Obs	0.85	0.53	0.26	0.58
	Historical sim	1.30	0.87	0.53	0.77
	Historical sim new	1.30	0.68	0.51	0.80
Rekingen	Obs	0.77	0.58	0.29	0.48
	Historical sim	1.37	1.46	1.08	0.29
	Historical sim new	0.83	0.75	0.44	0.38

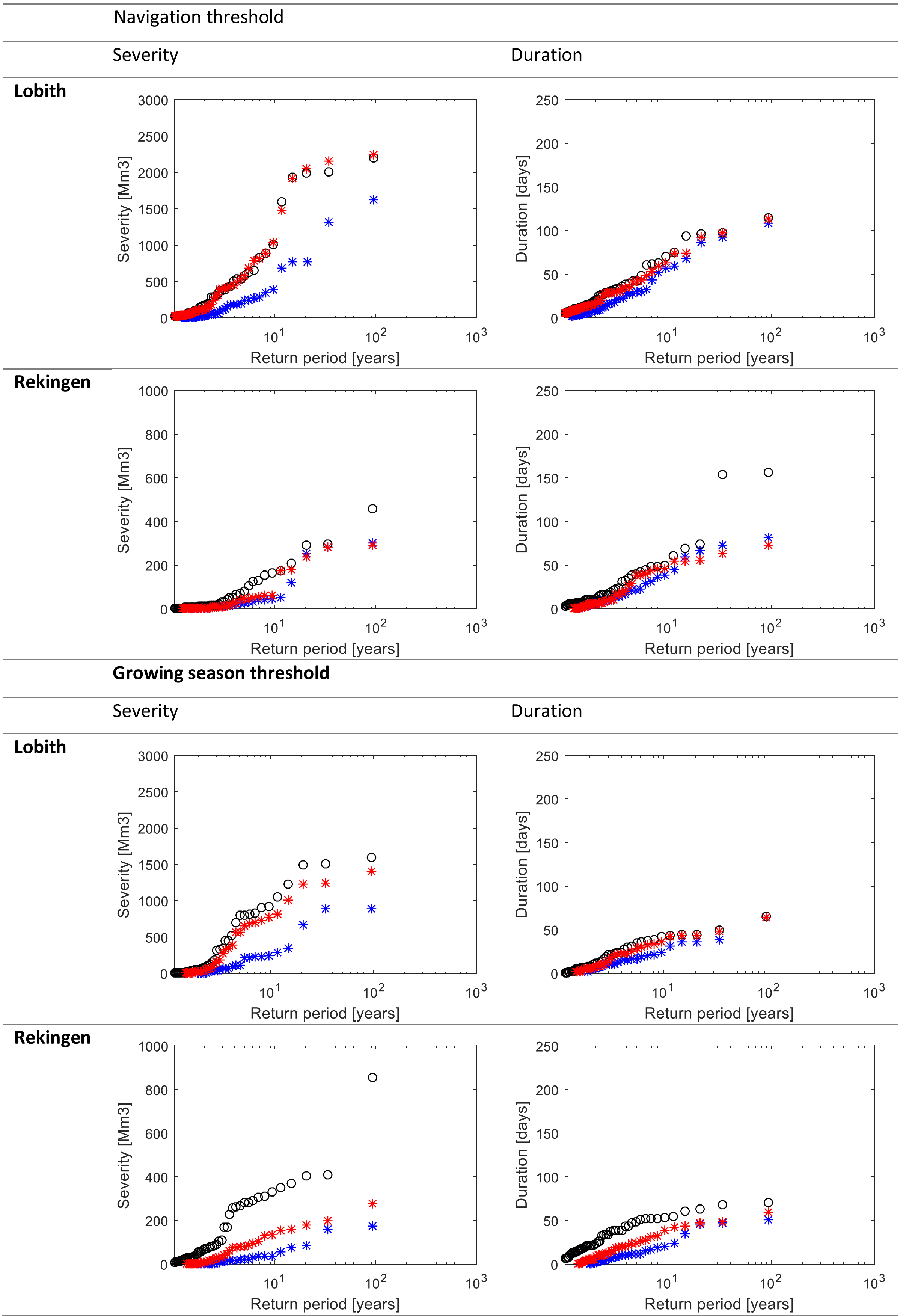


Figure 27 Return periods of event characteristics severity and duration after recalibration. Blue=observed, black= historical simulated old, red= historical simulated new.

Table 29 Performance indicators of lake level simulations after recalibration.

Lake	NSE	MAE	MAE summer (April-September)	MAE winter
Lake Constance	0.60	0.30	0.42	0.18
Lake Constance new	0.85	0.18	0.24	0.13

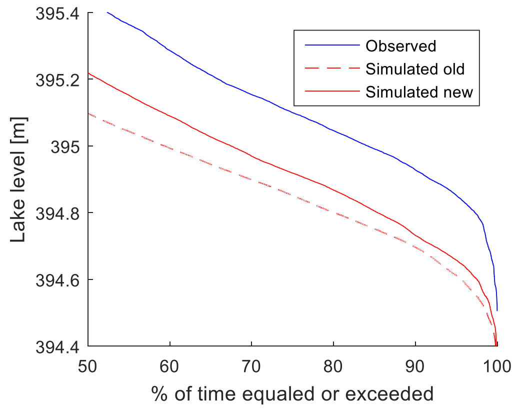


Figure 28 Level duration curve for lake levels in Lake Constance after recalibration.

Table 30 Performance indicators of snow cover simulation after recalibration

	Correlation time series	Mean # days start day later	Mean # days end day later	Mean # days max snow cover later	Correlation maximum cover
East Alpine	0.83	27	55	28	0.88
East Alpine new	0.85	25	54	26	0.89

Table 31 Correlation coefficients for groundwater indices of observed and simulated series after recalibration.

	Correlation groundwater indices	Correlation GW index with discharges observed	Correlation GW index with discharges simulated
East Alpine	0.51	0.70	0.69
East Alpine new	0.48	0.70	0.67