Using Ensemble Streamflow Predictions for extreme discharge purposes in the river Rhine

Ivo Huiskes (s1221485)

27-10-2016

Deltares
Enabling Delta Life

UNIVERSITY OF TWENTE.

Front page picture: *Deltares building Delft, taken by Jeanne Dekkers Architectuur*

Using Ensemble Streamflow Predictions for extreme discharge purposes in the river Rhine

Final report

In partial fulfilment of the requirements for the degree of
Master of Science in Civil Engineering and Management

Author:
Ivo Huiskes BSc.
s1221485
i.huiskes@alumnus.utwente.nl
Master Civil Engineering and Management (M-CEM)
Track Water Engineering and Management

Graduation Committee:
Deltares, Hydrology department
Dr. ir. F.C. Sperna-Weiland

University of Twente, Department of Water Engineering and Management
Prof. Dr. J.C.J. Kwadijk
Dr. ir. M.J. Booij

Summary

Flood situations caused by high discharges in rivers have large societal impacts. Such as damaged properties and the potential loss of lives. Therefore flood protection measures are taken, which are based on the outcomes of extreme discharge distributions. The state of the art is in this field lies in the classical way to obtain annual maximum discharges from discharge series, plot them in one graph and fit a distributions through these annual maxima points. Subsequently discharges belonging to a specific return period can be deduced from this so called fitted extreme discharge distribution. In many river basins in the world relatively short observation records are present. With the classical approach this results in a low number of annual maxima and therefore a major uncertainty in the final extreme discharge distribution. Longer synthetic discharges series are required in order to make a more accurate estimation of extreme discharge distributions. This research creates long weather series which will be transformed into river discharges with a hydrological model thereafter. For the creation of this weather series two numerical weather products of the European Centre for Mid-range Weather Forecasts (ECMWF) are used. The first product is called EraClim which is a re-analysis for global weather in the period 1901 – 2010. The initial conditions of EraClim's deterministic re-analysis are 10 times slightly perturbed to obtain 10 so called weather ensemble members from 109 year long. The second product concerns GLOFAS which is available during the period 2003 – 2015 with every day weather forecasts for 15-days ahead. The initial conditions of GLOFAS are 51 times perturbed which results in 51 weather ensemble members. All weather ensemble members have an equal probability to occur. The objective of this research is as follows:

The objective is to investigate to which extent products provided by numerical weather and hydrological models for operational flow forecasting can be used for estimating high discharge events in rivers having relatively short return periods (< 50 year) and how do these estimates compare with the estimates derived from classical methods.

The river Rhine basin is used in this research and served as a test case. This because of the relative long discharge and meteorological records in this basin. This makes it a perfect river basin to compare the extreme discharge distributions based on the numerical weather products with.

First of all long weather series are constructed. For EraClim each of the 10 weather ensemble members is put in a subsequent order resulting in a 1090 year long weather series. For GLOFAS weather series are built with a high degree of chronology. This is done by leaving out the first five and last five days of each 15 day long weather ensemble forecast. The resulting segments consisting of day 6 to 10 are put in a chronological order per ensemble member. This yields 51 chronological GLOFAS weather series in the period between 2003 and 2015. Also a GLOFAS weather series is constructed which is based on a high degree of randomness (called GLOFAS synthetic). These synthetic GLOFAS weather series is constructed based on weather segments of 4 days long. This choice of this segment length is random.

Weather series are compared for the overlapping period 2003 – 2006, in which data for both EraClim, GLOFAS and the HYRAS weather reference is available. For EraClim each of the 10 weather ensemble members is used to compare with the HYRAS reference. For GLOFAS the ensemble members for the chronologic weather series are used for comparison to the HYRAS reference. On a basin-average scale, precipitation of both EraClim and GLOFAS ensemble members show roughly a high degree of correspondence with the HYRAS reference. However, more extreme precipitation is obtained for GLOFAS. EraClim shows also slightly higher precipitation than the HYRAS reference. Another issue which became clear is that both GLOFAS and EraClim produce less dry days than the

HYRAS reference. After analysing the precipitation input, discharges at Lobith are obtained by feeding the weather series of GLOFAS and EraClim into the hydrological HBV model. Results show that during the months May, June and July much more discharge is simulated by EraClim and GLOFAS than by the HYRAS reference. When flow duration curves (FDCs) of daily discharges are compared to the HYRAS reference FDC, the GLOFAS FDC seems to describe the HYRAS reference FDC better than the EraClim FDC does.

Subsequently the independency of all weather ensemble members is tested; how is each weather ensemble member correlated with all other weather ensemble members from the same dataset. For both EraClim and GLOFAS it turned out that this mutual ensemble member correlations are very low (0.006 and 0.16 respectively). Therefore each weather ensemble member of both EraClim as GLOFAS is supposed to be practically independent. Building infinite long independent weather series becomes possible in this way.

Extreme discharge distributions (EDDs) at Lobith are compared for different datasets. It was found that the EDD of the synthetic GLOFAS series underestimates the observed EDD and the HYRAS reference EDD significantly. This is caused by less extreme basin-average precipitation 10-days prior to the peak discharge. The 10-day precipitation prior to the peak discharge is more conform the HYRAS reference when using EraClim. EraClim perform therefore quite well concerning the extreme discharge distribution at Lobith. However, this good correspondence is because of unsatisfied reasons. The HYRAS reference shows only 13% of all annual discharge peaks in summer. This is much less compared to EraClim (26%) and GLOFAS (32%). When the Alpine sub-basins are assessed, it turned out that Alpine average temperatures are much lower than the HYRAS reference, causing more snow storage during winter. Compared to the HYRAS reference this larger snow volume starts to melt later for both EraClim and GLOFAS. Therefore more meltwater contributions are expected to the discharge peaks at Lobith in summer. Furthermore extreme discharge distributions in the Alpine sub-basins are sometimes very different for both EraClim and GLOFAS with respect to the HYRAS reference. This means per definition that no reliable extreme discharge distribution can be drafted for these sub-basins based on GLOFAS and EraClim.

Taking into account the final extreme discharge distributions at Lobith, the behaviour in rain-fed basins and in Alpine catchments, it is concluded that using EraClim or GLOFAS is not a suitable alternative to the classical way of estimating extreme discharge distributions. However when having a very few number of annual maxima, EraClim or GLOFAS can be used. Important hereby is the perception of making large over- or underestimations of extreme discharge distributions.

Preface

This thesis is the final part of my master Water Engineering and Management which I have studied at the University of Twente with great satisfaction. Last nine months I studied the opportunity to make long weather series for the river Rhine basin in order to estimate extreme high discharge events at Lobith. During this research at Deltares in Delft I really learned a lot. From more detailed programming to handling large and complex datasets. Especially this generic part of producing results was definitely something I really liked.

First of all I would like to thank Jaap Kwadijk for creating this thesis topic and helping me to formulate complex thing in a better way. Thanks also go to Martijn Booij for his sharp feedback and critical questions leading to better insights. Further, Frederiek Sperna-Weiland deserves some good words because her door was always open for providing me with good advice and answering a lot of questions as well. I have to thank the great presence of all my colleagues and fellow graduate students as well. I always could discuss things with them and talk about the not always easy-going graduation process. Finally, I would like to thank my family and friends for supporting me during the whole process.

Ivo Huiskes
Heerenveen, 27 October 2016

Abbreviations and definitions

Abbreviations

ADD	=	Average Relative Degree of Dryness
BfG	=	Bundesanstalt für Gewässkunde
EC	=	EraClim
EDD	=	Extreme discharge distribution
ESP	=	Ensemble Streamflow Prediction
EPS	=	Ensemble Prediction System
FDC	=	Flow duration curve
GLF	=	GLOFAS
GRD	=	GRADE (Generator of Rainfall And Discharge Extremes)
GRDC	=	Global Runoff Data Centre
Obs	=	(Discharge) observations
IQR	=	Inter Quartile Range
HBV	=	Hydrologiska byråns vattenbalansavdelning, a Swedish Hydrological model to transform precipitation and temperature input into discharges along a river system.
HYRAS weather	=	HYRAS weather reference
HYRAS discharge	=	HYRAS discharge reference based on HYRAS weather.
P	=	Precipitation [mm]
PDC	=	Precipitation duration curve
Q	=	Discharge [m^3/s]
T	=	Temperature [°C]

Definitions

Chronological GLOFAS series	=	GLOFAS weather series in which the dates of the ensemble members follow up each other in a chronological way. So a weather segment of 21 – 24 January is succeeded by a weather segment of 25 – 28 January of the same year
Ensemble member	=	When the initial conditions of a deterministic forecast are perturbed 51 times; also 51 ensemble members arise. Just one member of an ensemble.
HYRAS	=	Weather observations used as weather reference situation
Segment	=	Part of a weather ensemble member which is left when the first and last part of the ensemble member is removed.
Sub-basin	=	1 of the 7 sub-basin within the river Rhine basin such as the Moselle, Neckar and Main. Defined by Demirell et al. (2013)
Sub-catchments	=	Smaller subdivision of the sub-basins within the river Rhine basin. There are 134 sub-catchments covering the whole river Rhine basin.
Summer	=	The period between the 1st of May and the 31th of October
Synthetic GLOFAS series	=	Weather series in which the dates of the random chosen ensemble members follow up each other in a random sequence. So, a weather segment of 21 – 24 January 2008 from ensemble member number 16 can be succeeded by a weather segment of 9 – 12 January 2003 of ensemble member number 28. For constructing a month like January only January segments are used.

Weather	=	In this research defined as the combination of precipitation and temperature.
Winter	=	The period between the 1st November and 30th of April

Contents

Appendices

1. Introduction

1.1 Flood risk in Deltas

Rivers have a lot of functions which serve humanity in positive ways. Such as navigation, recreation, irrigation and drink water supplies. However, rivers can also be a potential threat. Discharge volumes can become so enormous that it will flood levees and people get negatively affected. De Moel et al. (2011) relate higher flood risk in deltaic areas to climate change, economic growth and an increasing population. Economic growth causes a higher risk for a certain area because damage, and thus the costs, is expected to be higher in an economic more developed zone. Also an increasing number of people will affect the risk because their properties and own lives are exposed to a possible flood. The IGBP (2016) states that 1 percent of the earth surface is a river delta, while 500 million people, 7% of the total world population, is living there. Therefore knowledge about extreme discharges and the frequency of these is important. This research will focus on these extreme discharge distributions and especially the ones which are derived with help of numerical weather models used for operational flow forecasting.

1.2 Flood risk in the Netherlands

The Netherlands is such a delta. The main river that enters the country is the river Rhine at Lobith. During the past century, three major flood events have occurred as shown in figure 1. Due to climate change, Bronstert (2003) expects higher flood events with higher frequencies. Because of these higher floods, measures should be taken to make sure that return periods of floods stay on an acceptable level. Although this acceptable level is a political issue, it should be based on a good quality of quantitative data.

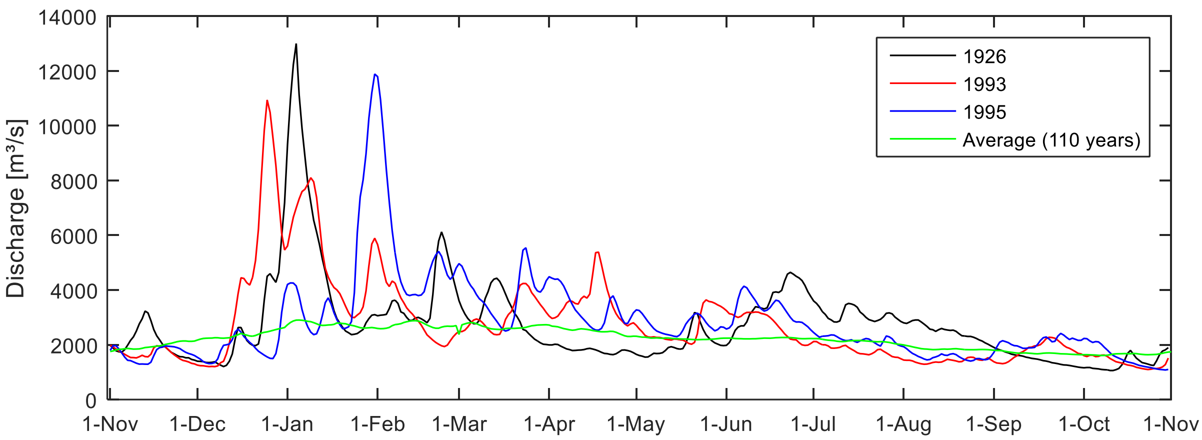


Figure 1 – Extreme flood events observed for the river Rhine at Lobith.

An extreme discharge analysis is thus required in order to investigate which discharge occurs with a specific return period. Van den Brink and Können (2009) mention that these extreme floods are generally determined on the basis of extrapolation methods. A problem with this approach is the lack of data series with a sufficient length. For the Dutch case the observed discharge series at Lobith are 110 years long. This results in 110 discharge maxima which forms the basis for the extrapolation process. It can be imagined that extrapolating based on this number of data points will cause a significant uncertainty in the final design discharge for a large return periods. The larger the return period the larger the uncertainty would be.

Ensemble Streamflow Predictions

Water management agencies are interested in early flood warnings in order to take preventive measures (Regimbeau et al., 2007). Therefore the European Centre for Medium-range Weather Forecasts (ECMWF) produces every day weather forecasts for the next 15 days (Cloke and Pappenberger, 2009; Demerit et al. 2007). Because of uncertainty reasons just one deterministic weather forecast is not enough to force a hydrological model. Therefore the numerical weather model called GLOFAS produces 51 different forecasts. According to NOAA (2006) this happens by perturbing the initial weather conditions of the deterministic forecast. Result are 51 weather forecasts which are called ensemble members. During the 15-day long forecast this ensemble form a bandwidth of 51 weather ensemble members around the deterministic weather forecast. An hydrological model is required to transform these meteorological weather ensemble members into discharges or so called ensemble streamflow predictions. As discussed before, these GLOFAS weather ensemble members are currently used for early flood warnings on the mid-term. However, the dataset of GLOFAS ensemble weather series contain lots of weather information. Possibly these ensemble members can be used in order to construct long weather series. These series can be used to research extreme discharge distributions.

1.3 State of the art

Extreme discharge distributions are assessed for many rivers around the world. Scientific work which assesses these extreme discharge distributions have in common that they all use the conventional method which basically consist of the following steps:

1) Selecting annual discharge maxima from the available dataset
2) Plotting these point into a graph
3) Fit a function to these points
4) Extrapolating to future return periods

Many of these classical examples are shown in literature, amongst others by Garba et al. (2013) and Willems et al. (2010). One can imagine however that discharge records are varying for different rivers. Sometimes these observations are very scarce. When just having 10 annual maximum discharges, uncertainties will grow immense when the extreme discharge distribution is interested in relative return periods (50 and 100 year return period is an example).

GRADE

To get a more certain estimation of an extreme discharge distribution, long synthetic time series can be constructed. Prior to a discharge series, weather series could be made to use as input into a hydrological model to produce river discharges. For the purpose of composing long weather series, a weather generation tool exist. This tool is called GRADE (Generator of Rainfall And Discharge Extremes) and generates daily weather data for the whole river Rhine catchment (Hegnauer et al., 2014). GRADE is able to generate infinite long weather series on a stochastic basis. However, this tool is only available for the river Rhine basin and needs calibration on their weather statistics in order to be useful. For the river Rhine this is not directly a problem, because more than sufficient data is available. For other river basins calibration can encounter problems.

1.4 Research gap

So far no research has been done to extreme discharge distributions based on long weather series using numerical ensemble weather forecasts. Since these GLOFAS forecasts are done on a daily scale with 51 ensemble members, a lot of weather data is available from this numerical weather product. Next to GLOFAS, the ECMWF has another relatively new numerical weather product.

EraClim is a numerical weather product which concerns the re-analysed weather in the whole last century. Basically it concerns 109 years of deterministic weather based on the fraction sea ice and the average daily temperature sea surface temperature. Perturbing the initial conditions of this deterministic re-analysis will result in 10 ensemble members lying around this deterministic re-analysis of weather. Because building long weather series with ensemble members has never done before, no standard method is available to prepare such series. Also no standard procedure exists to assess the quality of the extreme discharge distributions. Since relative long observation records are available for the river Rhine, this basin can function as a test case in order to investigate the potential of numerical weather products for extreme discharge estimations. It will be good to compare the observed and simulated extreme discharge distributions at Lobith to the ensemble-based ones in order to say something useful about both methods. It is interesting to investigate if extreme discharge distributions are different when using synthetic weather series from numerical weather products or using the conventional method which uses relatively less observed extreme discharges.

1.5 Objective and research questions

For the purpose of calculating design discharges in rivers, it is relevant to investigate to which extent it is possible to make use of products driven by the numerical weather models used for operational flow forecasting. Therefore the Master thesis will have the following objective:

The objective is to investigate to which extent products provided by numerical weather and hydrological models for operational flow forecasting can be used for estimating high discharge events in rivers having relatively short return periods (< 50 year) and how do these estimates compare with the estimates derived from classical methods.

In order to achieve the defined objective from the previous paragraph, main research questions are formulated as follows:

1) Which product seems most promising for the assessment of design discharges for short (< 50 year) return periods using Ensemble Streamflow Predictions?
2) Based on EraClim and GLOFAS weather ensemble members, in which way long and independent synthetic weather series can be composed to estimate the flood design level at a certain return period?
3) To what extent do extreme discharge distributions differ at Lobith when they are obtained by using extreme discharges based on numerical weather products compared to the observed extreme discharges?

1.6 Reading guide

The structure of this research is as follows. In chapter 2 an overview will be given of the study area of the river Rhine basin and all used datasets are discussed. This section also provides some more in depth information about what ensemble products exactly are. Chapter 3 firstly discusses how to build all weather series based on numerical weather products. Secondly the method to assess which numerical weather product (EraClim or GLOFAS) seems to have the most potential compared to the weather observations is discussed. Then the independence of mutual ensemble members is explained in more detail. Also a procedure is set up to compare extreme discharge distributions from different datasets at Lobith. Finally the analysis will be discussed in order to see how different Rhine sub basins perform concerning extreme discharge distributions. Next in chapter 4 the results of the performed method will be presented. This is followed by a discussion in chapter 5 and conclusions and recommendations in chapter 6.

2. Study area and data

This chapter describes the study area and all used datasets and models. Furthermore the essentials of Ensemble Streamflow Predictions are closer reviewed; how do they work and what do they produce.

2.1 Study area

The river Rhine flows from Switzerland, through Germany to the Netherlands and covers a total area of 185.000 km² (Hoffmann et al., 2007).

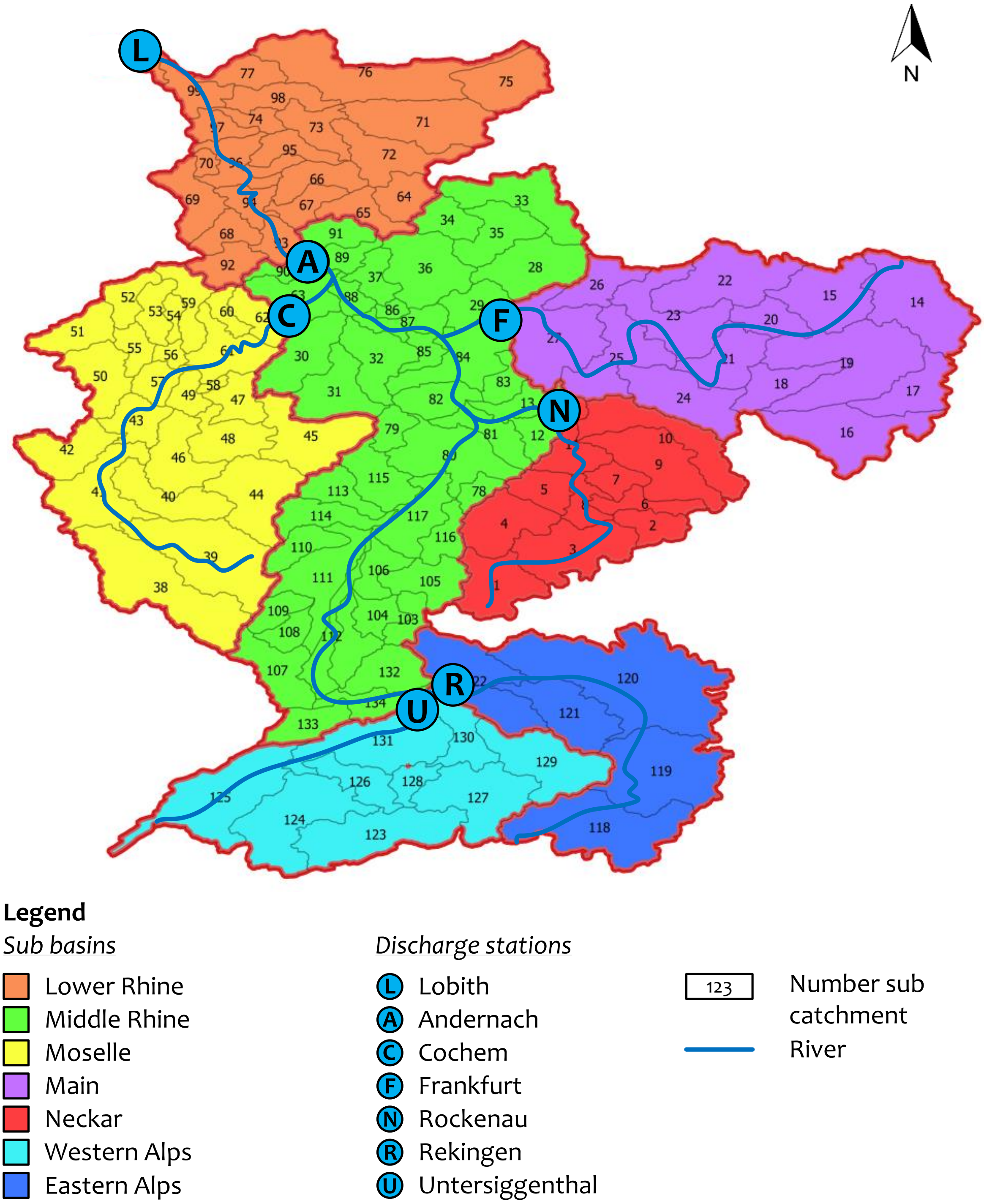


Figure 2 – River Rhine basin and their 7 sub-basins (coloured) according to Demirel et al. (2013).

The upstream part of Lobith (figure 2) is taken into account because this is the part of the basin which can actually influence discharges at Lobith. This part of the river Rhine basin can be divided in seven sub-basins according to Demirel et al. (2013). Herein both rainfall dominated rivers (Moselle, Main, Neckar) and meltwater dominated basins (Alpine regions) are captured. Furthermore two more central sub basins are taken into account to which the border basins will discharge (middle and lower Rhine). About half of the land use in the river Rhine area is agriculture, 32% consists of forest and only 9% is urban area (Tockner et al., 2009). The river Rhine has an average discharge of 2225 m^3/s at Lobith. Rain-fed basins peak mostly during winter while meltwater dominated rivers peak in summer. According to Middelkoop and Hasselen (1999) the summer discharge at Lobith consist for 70% of Alpine meltwater. This research uses the river Rhine basin specifically because of its long data records (see paragraph 2.4). Long series of weather observations and discharge observations are available. These long sets can function as a reliable reference.

2.2 Numerical weather products

Chapter 1 already discusses two numerical weather products very roughly. In this paragraph some more explanation will follow in order to understand the process of Ensemble Streamflow Predictions (ESP). This research uses two numerical weather products; GLOFAS and EraClim. They both produce meteorological ensemble members and are both provided by the ECMWF. However these two numerical weather products have a different nature.

GLOFAS

Meteorology tries to describe the state of the atmosphere, atmospheric phenomena and the atmospheric effects on the daily weather (Hogan, 2014). One can imagine that lots of uncertainties can possibly be involved. GLOFAS is a numerical weather product which is the result of a numerical weather model of the ECMWF. Every day GLOFAS produces one deterministic weather forecast for the next period of 15 days (Cloke and Pappenberger, 2009; Demerit et al. 2007). This deterministic forecast predicts precipitation and temperature parameters for every grid cell of 25 x 25 km on a worldwide scale (Alfieri et al., 2013). This deterministic forecast forms the basis for each of the ensemble members. The ensemble members are obtained by adding small disturbances in initial weather conditions of the deterministic meteorological forecast (Cloke and Pappenberger, 2009). This process can be repeated, which every time results in a new ensemble member (NOAA, 2006). It shows the spread around the deterministic meteorological forecast (Cloke and Pappenberger, 2009). GLOFAS creates on a daily basis 51 meteorological ensemble members with one deterministic reference member (Demmerit et al., 2007).

EraClim

EraClim (European Reanalysis of Global Climate observations) is a numerical weather product provided by the ECMWF. A global atmospheric reanalysis is done for the period 1901 – 2010 resulting in worldwide available day-to-day weather (Poli et al., 2013). The EraClim product is aimed to improve the observational weather records from 1901 to 2010 containing precipitation and temperature (ECMWF, 2015). Furthermore, the ECMWF (2015) wants to increase the data reliability by using ensemble members for the reanalysis of the weather data.

EraClim's deterministic atmospheric assimilation is forced by only two initial conditions (Dee, 2013). Only the daily average sea-surface temperature and the daily average sea ice fractions are required in each grid cell of 25 x 25 km. This may sound very attractive, but just a small deviation in these initial conditions will potentially lead to large deviations in the final assimilated weather. For that reason 10 ensemble members are created by perturbing the initial conditions of the deterministic assimilation (Poli et al., 2015)Not only one deterministic assimilation will be provided. This is done to help the model to realistic integrate the meteorological data through the century-long running period.

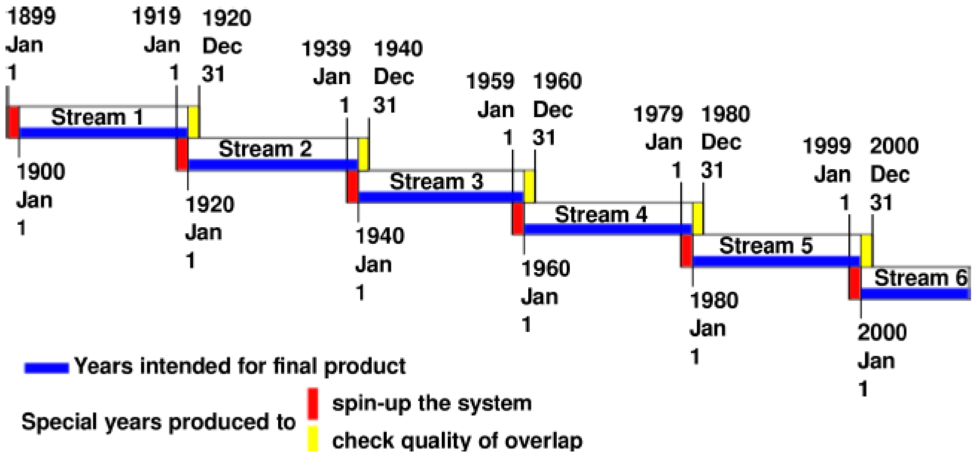


Figure 3 – Schematic layout of EraClims re-analysis production for the final EraClim product (Poli et al., 2015).

When a century long weather assimilation is executed, 6 runs (in figure 3 called 'streams') are performed of which each represents a segment of 20 years plus some extra overlap to correct for some bias in the deterministic weather assimilation (Poli et al. 2015). The red boxes are required for some spin up time to reach acceptable atmospheric circumstances for the deterministic weather assimilation.

2.3 From meteorological ensemble prediction to Ensemble Streamflow Predictions

In figure 4 the steps are described in order to process meteorological ensemble members into Ensemble Streamflow Predictions. Input are meteorological ensemble predictions in step 1. It can be seen that the meteorological ensemble members first have to be pre-processed in step 2. This contains an interpolation step to the same grid size as used in the catchment hydrology model (Cloke and Pappenberger, 2009). For all used data, this step is already done. When all 134 sub catchments are taken into account, the hydrology model calculates river discharges, or streamflow's, at important locations along the river system.

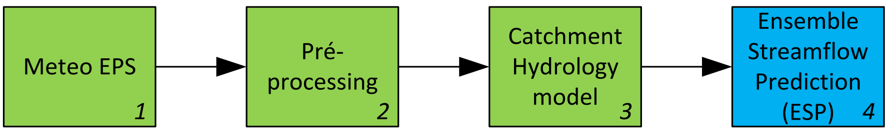


Figure 4 – Schematisation of the Ensemble Streamflow prediction process based on of meteorological EPS input. Adapted after Cloke and Pappenberger (2009).

Ensemble Streamflow Predictions can be visually schematised as shown in figure 5 (NOAA, 2015). The graph shows that the observed discharge series (black) are known until the current moment t = 0. From then on, different Streamflow Ensemble members will be created using the meteorological ensemble members. All these ensemble members have an equal probability of occurrence (Cloke and Pappenberger, 2009).

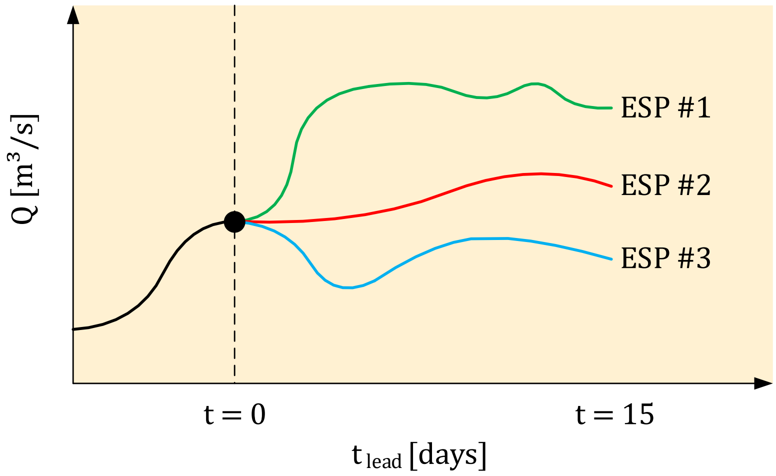


Figure 5 - Ensemble streamflow principle with 3 random ensemble members.

2.4 Datasets

The data used have been retrieved from different sources and have different data periods. Daily discharge observations at different locations along the River Rhine are acquired from the Global Runoff Data Centre (GRDC) provided by the Bundesanstalt für Gewässkunde (BfG). Two different historical and interpolated datasets for daily precipitation and temperature are used; HYRAS and E-OBS. The HYRAS weather dataset has a spatial resolution of 5 x 5 km based on more measurement stations than E-OBS which has a resolution of 25 x 25 km (Kjellström et al., 2015). Two numerical ECMWF weather products are used which both produce weather ensemble members; as discussed before 10 ensemble members for EraClim and 51 ensemble members for GLOFAS. Finally one discharge series is present for Lobith based on the weather generator of GRADE.

Table 1 – Used datasets (Q is discharge, P is precipitation and T is temperature). All data is on daily basis.

	Variable	Source	Location	Sub-basin	Data period
Observations	Q	GRDC, BfG	Lobith	Lower-Rhine	1901-2010
	Q	GRDC, BfG	Andernach	Middle-Rhine	1951-2006
	Q	GRDC, BfG	Cochem	Moselle	1951-2006
	Q	GRDC, BfG	Frankfurt	Main	1963-2006
	Q	GRDC, BfG	Rockenau	Neckar	1951-2006
	Q	GRDC, BfG	Rekingen	East-Alps	1951-2006
	Q	GRDC, BfG	Untersiggenthal	West-Alps	1951-2006
	P,T	HYRAS 2.0		134 sub-basins	1951-2006
	P,T	E-OBS		134 sub-basins	1951-2014
Numerical weather products	P,T	EraClim, ECMWF		134 sub-basins	1901-2010
	P,T	GLOFAS, ECMWF		134 sub-basins	2003-2015
Synthetic discharge series	Q	GRADE	Lobith	Lower-Rhine	5000 year

2.5 Hydrological model

As already mentioned in section 2.3, a hydrological model is required in step 3 of figure 4 in order to transform meteorological ensemble input into discharges. The HBV-96 model of Lindström et al. (1997) is suitable to transform precipitation and temperature data into discharges for the river Rhine basin. Based on figure 6 the model takes into account four submodules:

1) Snow and precipitation module
2) Soil module with infiltration, evaporation, percolation and capillary rise
3) Runoff module for both lower and upper zone
4) Routing module to obtain discharges along the river Rhine

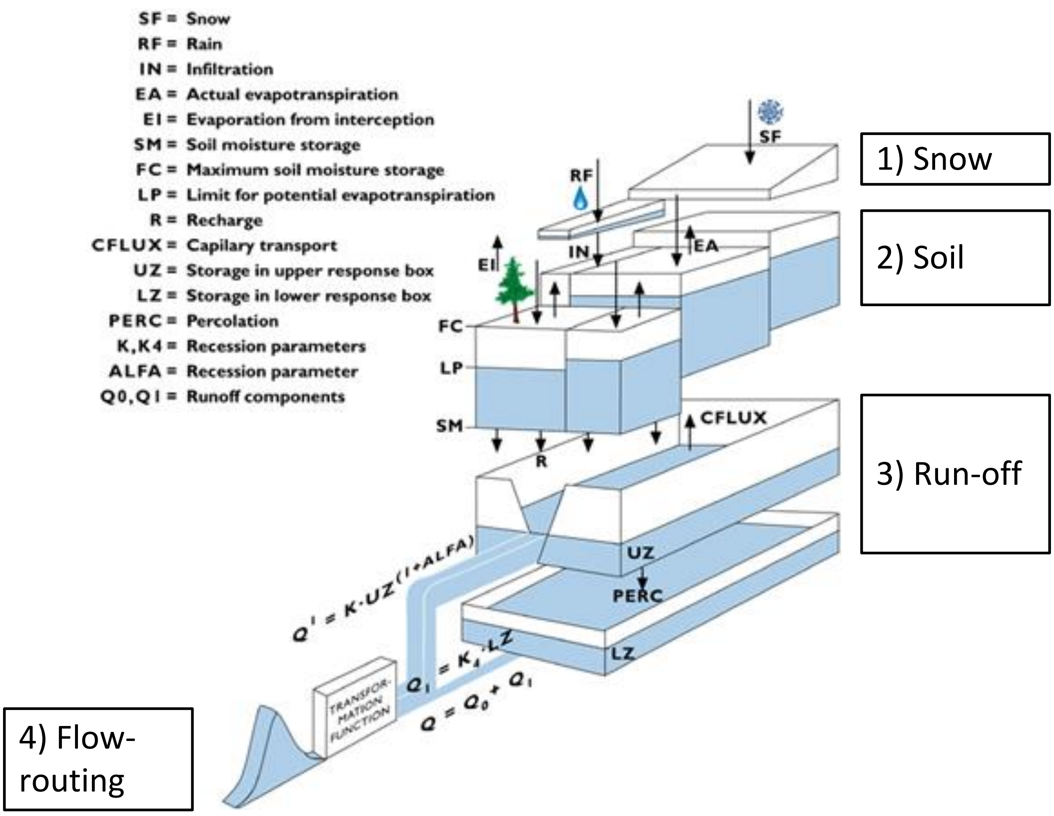


Figure 6 – HBV model schematisation for one of the 134 sub catchments (Lindström et al., 1997). Textual additions after Hegnauer et al. (2014).

The configuration of HBV takes all 134 sub catchments of the river Rhine basin into account since meteorological data (temperature and precipitation) are available for this number of sub-catchments. A calibrated HBV-model is present for the river Rhine basin. This calibration was performed by Hegnauer and Verseveld (2013) by grouping sub-catchments into 15 major sub basins. All these major sub basins are calibrated separately from up to downstream for the period between 1989 and 2006. The HYRAS dataset was used as weather input for this calibration. According to Steinrücke et al. (2012) HYMOG (Hydrologische Modellierungsgrundlagen im Rheingebiet) provided discharge data which has been used as reference for calibration.

3. Method

This chapter describes step by step which method will be followed in order to answer the research questions. Figure 7 shows the method outline with the corresponding sections. In case a more detailed method is required, an appendix will be concerned.

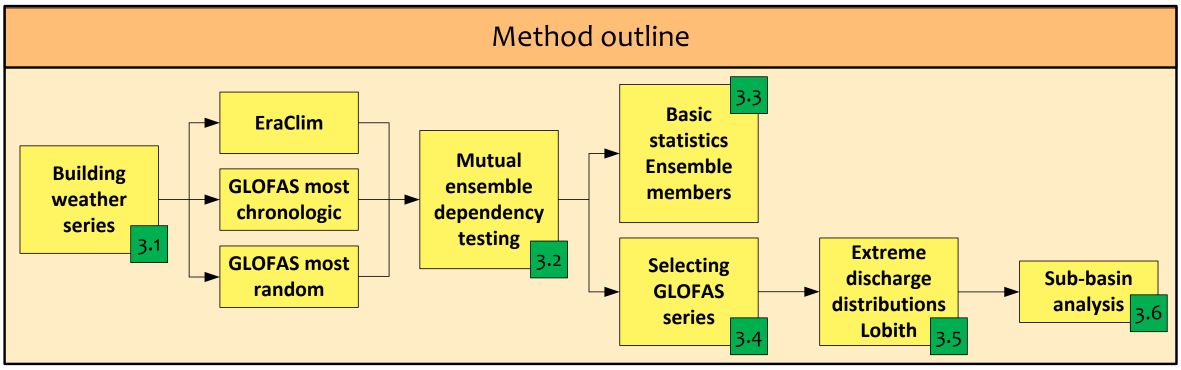


Figure 7 – Overview of steps taken in this research method. In green boxes the section designation is provided.

3.1 Building weather series

Firstly it is important to construct weather series from numerical weather products. Note that all construction steps below are only to build the weather series and say nothing about the quality or independence yet.

EraClim

The approach for one long EraClim weather series is kept quite simple. All 109 year long ensemble members are put in a subsequent order. Because 10 members are available this leads to a 1090 year long weather series with 10 possibly 'bad' connections. These possible bad connections are not further researched since this will only be 10 out of 398123 days. The applicability of this method depends of the dependency analysis on all 10 weather ensemble members. Only when EraClim weather ensemble members are independent enough, this approach is applicable.

GLOFAS

As described in section 2.2, ten EraClim ensemble members are available for the period between 1901 and 2010. These ensemble members are continuous during that whole period. This is not the case for the GLOFAS series that are only available during 2003 to 2015. Besides, GLOFAS has every day 51 possible ensemble members for the next fifteen calendar days. In order to build a long enough GLOFAS weather series, different GLOFAS ensemble member segments should be linked. An obstacle mentioned by Cloke and Pappenberger (2009) concerns the significant influence of the initial conditions in the first 5 days of each ensemble weather forecast. Therefore it is important to leave these first days out of the precipitation (and later discharge) analysis. On the other hand the last days of a 15-day long GLOFAS meteo ensemble forecast are not useful. Pappenberger (2015) mention a weather convergence to a climate average in the period from day 11 to day 15 of the GLOFAS forecasts. This is the reason for leaving these days out of the available dataset. Only days 6 to 10 are useful therefore, which is only 5 days of an actual ensemble member.

Construction of GLOFAS most chronological series

This method describes the construction of GLOFAS weather series based on the most chronological way possible. The series is called chronological because each calendar day follows up each other.

For this building process all 51 ensemble members are used for the whole period between 2003 and 2015. Hereby it is required to perform exactly the same process for precipitation as for temperature. In this way the corresponding temperatures with the precipitations stays consistent since precipitation is not totally independent of temperature according to CLIMAS (2016).

a) At the first of January of the year 2003 one of the 51 GLOFAS meteo- ensemble members is chosen (see upper ensemble in figure 8 below).
b) From this the days 6 up to 10 will be extracted. Subsequently they will be allocated to the dates from the 6th of January to the 10th of January (green boxes, figure 8).
c) Thereafter the same ensemble member is taken for a later moment in time, which is initially produced for the dates 6 up to the 20th of January. This meteo-ensemble member (which once again consist of 15 days) will be resized by cutting the first and last five days off (see second ensemble, figure 8).
d) After finishing this process, next the sixth day is allocated to the 11th of January.
e) In order to build a chronological weather series for each ensemble member, the steps a to d will be repeated. Finally a weather series is available between the 6th of January 2003 and the 31st of December 2015 for each ensemble member.
f) Figure 8 shows 5 days missing at the beginning of the final constructed weather series in in January. This is the only place that the 5 days of initial conditions are not left out. This is done to prevent date shifts in the weather series when they are linked as described in the next step.
g) All 51 ‘chronological’ GLOFAS weather series are linked in order to form a long series of 51 * 12 = 612 year.
h) This 612-year long meteo series is transformed into discharges using the calibrated HBV model for the River Rhine.

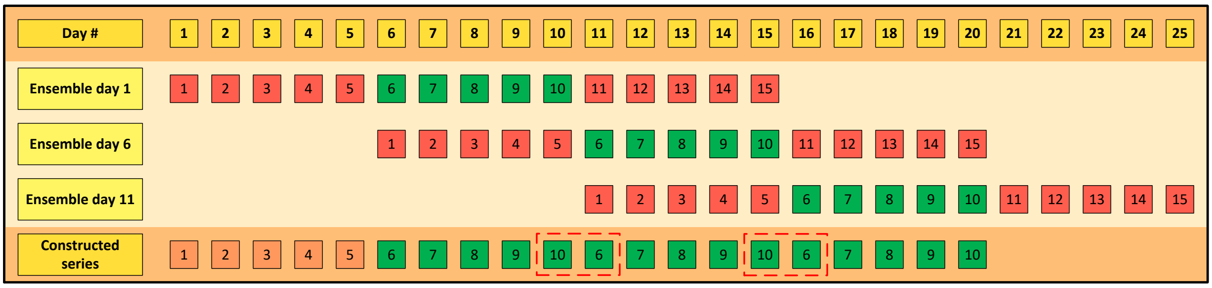


Figure 8 – Basic principle of the data building approach in chronologic sequence (non-random).

Construction of GLOFAS random synthetic series

A disadvantage of the ‘chronologic’ GLOFAS series as described above is that each ensemble member is only used once over the whole period between 2003 to 2015. In theory this obstructs the formation of infinite long weather series. Therefore it is interesting to investigate if it is possible to construct infinite long GLOFAS series. When it appears to be possible, longer weather series can be constructed leading to more annual discharge maxima (after processed by HBV). Possibly this leads to higher confidence in the extreme discharge distributions.

Synthetic GLOFAS weather series will be generated as described below. This approach is only applicable when it turns out that GLOFAS weather ensemble members are independent. First a selection process takes place on all the available GLOFAS data between 2003 and 2015. In this meteo-series weather segments of 4-days long will be used. This is a random choice. 1, 2, 3 or 5 days weather segment lengths are also possible.

1) Everyday 51 ensemble weather forecasts are available. Because 4-day long weather segments are chosen to use, these have to be selected from the 51 ensemble members for every new forecast day. This means that the ensemble weather forecast of the 1st of January delivers a 4-day weather segment from 6 to 9 January for each of the 51 ensemble members. The 2nd of January gives the 4-day long weather segment as forecasted for 7 to 10 January.
2) When a 4-day long precipitation segment only covers 4 January days, this segment is added to the January precipitation-matrix. The number corresponding with the ensemble member is saved together with the 4-day long segment. When a 4-day long precipitation segment covers both January and February days, it will not be used.
3) For each month such a matrix is built, so 12 large precipitation matrices are available as new resource to construct synthetic weather series from.
4) Step 1 to 3 are executed at the same time in order to find the corresponding temperature matrices. The layout will be in such a way that 4-day long precipitation exact fit with their corresponding temperatures.

These 12 matrices are the basis for the infinite long GLOFAS weather series. The only consistency which is taken into account to construct random synthetic (short: synthetic) GLOFAS series is a monthly based consistency. This means that only January weather is used in order to make a new synthetic January weather series. The synthetic GLOFAS weather series are constructed as follows;

1) First monthly synthetic GLOFAS series are constructed. For every month 8 random ensemble numbers between 1 and 51 are chosen (every year again).
2) For each of the 8 random chosen ensemble numbers, a random 4-day long segment will be taken from the precipitation matrix for the month of interest. All these segments will be linked to each other until a monthly series of 8 × 4 = 32 days is reached. Depending of the number of days in a certain month, the surplus of days will be cut off.
3) The corresponding temperature series is composed at the same time.
4) This process is repeated until the desired length of the synthetically constructed GLOFAS series is reached.

Figure 9 shows the linking of random chosen 4-day long segments for a period of 12 days. As can be seen the weather is originating from random ensemble members (red boxes) and different years (blue boxes). Because the EraClim (artificial) weather series will have a length of 1090 year, the synthetic GLOFAS series will be constructed for an equally long period. This is done because the performance with respect to each other can be more equitably be compared.

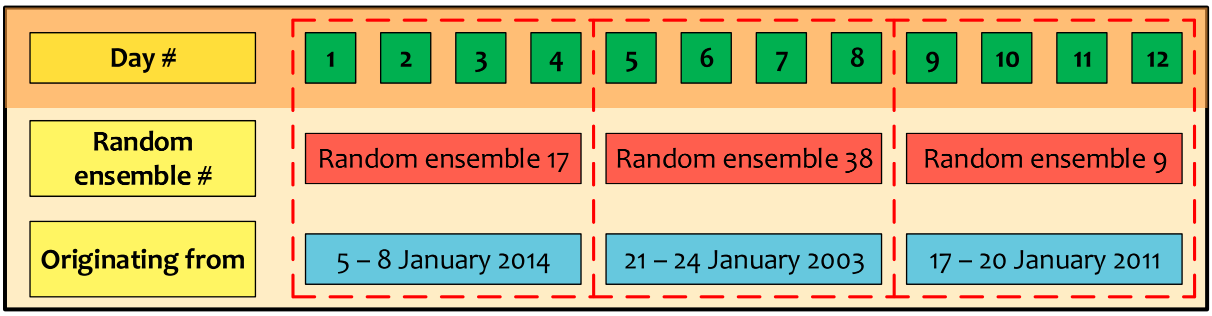


Figure 9 – Random possible construction example of synthetic GLOFAS precipitation series with a segment length of 4 days. In green, the final precipitation days of the artificial time series, in red the random chosen ensemble number, and in blue the original used timeslot.

3.2 Mutual ensemble member dependency

The chronological GLOFAS weather series as described in section 3.1 uses 51 different chronological GLOFAS discharge series from 2003 to 2010. Every discharge series is based on a weather series composed of one specific GLOFAS ensemble number. Question is how independent these 51 weather series actually are. When the resulting weather is not independent, it becomes hard to construct infinite long synthetic GLOFAS series in a later stage. Dependent weather ensemble members results in a final series where more segments of equal lengths are more less the same, so the weather series are less random. An independency analysis can be performed on the precipitation data or just on the final discharge series. Both analysis are executed, starting with a Pearson correlation analysis on the GLOFAS precipitation series in the period 2003 to 2015. The independency of each of the 51 precipitation ensemble member series is assessed by correlating each ensemble member to all other available precipitation ensemble members. In order to carry out a proper analysis, it is better to perform this precipitation correlation tests for each of the 134 Rhine sub catchments separately. In that case spatial differences in ensemble member correlation can be seen. Hereafter the same correlation tests are performed on the seven major sub basins as defined by Demirel et al. (2013). Figure 10 gives an overview of the independency test concerning the mutual ensemble member precipitations. A similar process is performed for the 10 109-year long ensembles of EraClim. The methodology for the discharge analysis is captured in appendix B.

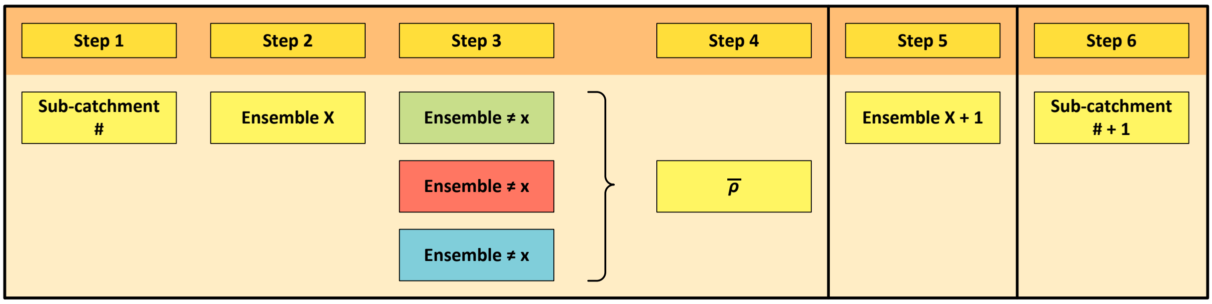


Figure 10 – Schematic overview of the mutual ensemble member dependency test.

3.3 Selecting GLOFAS time series

As described in section 3.1 two GLOFAS weather series are constructed; the so called chronological GLOFAS weather series and the most random GLOFAS synthetic series. In theory this random GLOFAS synthetic series can construct infinite long weather series. However, since calendar days are not used in a chronological order the GLOFAS synthetic series might perform worse than the chronological GLOFAS series. For this reason the chronological and the synthetic GLOFAS series are compared. Therefore a couple of basic statistics will be assessed for both of these datasets. Also some extreme discharge statistics are taken into account. The planned analysis are listed below.

- Monthly basin average precipitation and monthly basin average temperature
- Monthly average discharge at Lobith
- Time distribution plots for peak discharge to see in which months most peak discharges occur
- 10-day precipitation prior to the peak discharge plotted against peak discharge at Lobith. Regression lines should be more less the same
- Precipitation duration curves of 10-day precipitation prior to the peak discharge at Lobith
- Extreme discharge distributions

These results will be visually compared. When differences between chronological and synthetic GLOFAS weather series and discharge distributions are small it will be possible to construct infinitely long synthetic GLOFAS series.

3.4 Basic statistics EraClim and GLOFAS

In order to answer the first research question, this paragraph describes the basic statistics which will be performed on both precipitation and discharge when comparing ensemble products with the HYRAS reference.

Precipitation analysis

EraClim and GLOFAS input data provide meteorological input such as precipitation and temperature. Therefore it will be useful to perform some basic analysis to these meteo input since this will form the basis of all produced results. For extreme discharge events it is assumed that temperature has less influence on the discharge than precipitation does. So the 'weather analysis' will be limited to the precipitation part of the input datasets. For this precipitation analysis the precipitation data is used from each separate ensemble member. For EraClim this means 10 weather ensemble members of 109 years. For GLOFAS the chronological series of 51 ensemble members is used where each member is 12 year long. These ensemble precipitation data is compared to the HYRAS reference which contains observed precipitation and temperature. Why HYRAS is chosen as a reference set is described in appendix A. Goal of the precipitation analysis is to compare all ensemble members of each dataset, and compare them with the observed precipitation too. The basin averaged precipitation is determined on daily basis. This is done for the whole period between 2003 up to 2006, because for that period, both weather observations (HYRAS), EraClim and GLOFAS are available. Next statistics will be assessed and compared;

- Precipitation boxplots of each ensemble member, based on all daily precipitation
- Precipitation average and standard deviation for every ensemble member
- Maximum 10 day precipitation sums
- Monthly precipitation sums

Final step is to investigate how many dry days are modelled by EraClim and GLOFAS compared to the observations during 2003 and 2006. For every month the percentage of dry days ($P < 0.3$ mm/day according to MeteoGroup (2016)) will be assessed in all of the 134 sub-catchments. This percentage will be multiplied by the area-fraction of each sub catchment. Finally all these are summed to find an average degree of dryness of the whole Rhine system for a certain month. The degree of dryness will be assessed on all available meteorological data. This is important because EraClim will use all ensemble members, and GLOFAS will do this too by chance. The result of the dryness analysis will be visually compared.

Discharge analysis

After the analysis on the daily meteorological input data, an analysis of the discharge at Lobith is also important since the final interest lies in the field of extreme discharges. Precipitation and temperature input of the ensemble members will be processed by the HBV model in order to simulate the river Rhine discharges at Lobith. Just as with the precipitation analysis, EraClim uses all its 10 weather ensemble members to transform into discharges. For GLOFAS the 51 weather ensemble members are used to transform into discharges as mentioned in the so called 'chronological' series.

An analysis of the EraClim and GLOFAS ensemble members will be executed by comparing daily ensemble discharges with the reference discharge at Lobith. In fact there are two reference discharges at Lobith possible; the real measured discharge observations, or the HYRAS weather induced discharges. This last discharge reference uses the same HBV configuration that EraClim and GLOFAS, so the same bias is in the results. However, when using HYRAS only the period 2003 – 2006 is available to compare EraClim and GLOFAS with. When using the real discharge observations, the overlap period is much longer. Therefore the flow duration curves of HYRAS induced discharges and observations are compared. Based on a visual inspection is decided whether to use HYRAS of the real observations as a reference. From the discharge analysis the intention is not to find perfect fitting ensemble discharges to the reference discharges, but more to investigate how all ensemble discharges relate statistically to the observations.

Compared with GLOFAS, EraClim is a product which allows a larger weather variability because only once in the twenty years the ensemble magnitude is controlled (Poli et al., 2015). For GLOFAS every day the model starts is controlled because every day a new forecast is made where the initial conditions are observed. To see the impact it is fair to take a period in which EraClim, GLOFAS and reference (HYRAS or observations) discharges are all available. The following discharge statistics are assessed for ensemble products and reference;

- Average monthly discharge
- Discharge boxplot per ensemble member
- Flow duration curves per ensemble member

3.5 Estimation of extreme discharge distributions

Final interest lies in the extreme discharges with small return periods (5, 10, 20 and 50 years) at Lobith. The different used weather sets are at the basis of these plots. All available datasets will be used, meaning the discharge observations at Lobith, the simulated discharges based on the HYRAS reference, the 1090 years of synthetic GLOFAS series and the 1090 year long EraClim series. Last added dataset contains data of GRADE (see chapter 1 and section 2.4). This additional set is only used at Lobith as another extra reference.

To make sure that not two peak discharges from the same discharge wave are taken, the annual maximum discharges are determined. The hydrological year from the first of November to the 31th of October is taken therefore. Result is a highly variable number of annual discharge maxima per dataset. The annual discharge maxima are plotted against its Gumbel variate. This is because many literature in the field of conventional extreme discharge analysis takes into account the Gumbel distribution in their work (examples are given by: Garba et al. (2013) and Willems et al. (2010)). The results of these plots will be visually compared and discussed for all datasets. Thereafter the intention is to fit a Gumbel line through each cloud of annual maxima. Before doing this, a probability plot correlation coefficient (PPCC) test will be carried out to investigate if a Gumbel line can be fitted to the dataset of annual discharge maxima. An exact description of this test is captured in appendix C. When becomes clear that a Gumbel fit may be used, the approach of Shaw et al. (2005) will be followed in order to determine the Gumbel fit parameters (appendix C). In this way it becomes possible to estimate the discharges belonging to small return periods (5, 10, 20 and 50 years) at Lobith and compare them for the different datasets.

However the Gumbel fits will produce an uncertainty which should be taken into account. Shaw et al. (2005) provide a method for the calculation of the 95% confidence limits of each Gumbel fit which can be found in appendix C. For Lobith different plots will be drafted with the discharge confidence interval for small return periods for all assessed datasets. Note that these confidence limits are only based on the fitting uncertainty. Other uncertainties originating from the data or the HBV model are not taken into account. Despite this imperfection the rate of overlap is discussed with the HYRAS reference bandwidth and the one of the observed discharge.

It will be tried to explain the extreme discharge distributions using the next main analysis methods;

- 10-day precipitation prior to the peak discharge at Lobith will be plotted against the peak discharges since van Pelt et al (2015) mention that 10-day precipitation determines the peak discharge at Lobith to a significant extend.
- Plots of monthly relative frequencies of peak discharges.
- Precipitation duration curves will be plotted for each dataset summed over 10 days prior to the peak discharge.

3.6 Correctness extreme discharge distributions Lobith

In the previous section only an extreme discharge analysis is performed for Lobith. However discharge peaks at Lobith are composed of contributions of different sub-basins upstream from Lobith. There are rain-fed basins contributing to the peak discharge, but the regime in the Alpine basins is very different and is expected to contribute in another way.

Therefore it is interesting to analyse the behaviour in the sub basins of Demirel et al. (2013) concerning the extreme discharge statistics. Only the sub basins at the upstream borders of the river Rhine basin will be assessed, meaning that the lower and middle rhine basin are left out of the more detailed analysis. This will be done because the lower and middle Rhine basin receives discharge input from upstream basins. This makes it more difficult to find a direct relation between weather and peak discharge.

The same datasets will be used for the sub-basin analysis as used for Lobith. Only the GRADE dataset is not taken into account. Finally a division is made between rain-fed sub-basins and Alpine sub-basins. The following analysis will be performed. The ones marked with an asterisk are only performed in Alpine regions since temperature is expected to play an important role in those regions in order to force snowmelt.

- Extreme discharge distributions at discharge station at the lower part of the sub-basin.
- Precipitation prior to the peak discharge.
- Plots of monthly relative frequencies of peak discharges.
- Precipitation duration curves for both summer and winter periods
- Relative number of freezing days*
- Temperature distributions*

4. Results

The outline of this chapter is as follows. Firstly the ensemble member statistics of the ensemble products EraClim and GLOFAS will be compared with the HYRAS reference. Thereafter a dependency test is performed between ensemble members on both ensemble products. Then a choice is made for which GLOFAS weather series to use. Subsequently all extreme discharge distributions are plotted for Lobith and analysed. Finally the same analysis is performed on a smaller scale for some sub-basins in the river Rhine.

4.1 Basic statistics EraClim and GLOFAS

In this paragraph the results of the precipitation and discharge analysis of both EraClim and GLOFAS will be discussed.

Precipitation analysis

As mentioned in section 3.4 all precipitation data over the period 2003 up to 2006 is used for every sub catchment of the river Rhine in this analysis. For this timeslot EraClim, GLOFAS and HYRAS observations are all available. This precipitation analysis only uses the GLOFAS series with the chronological calendar days.

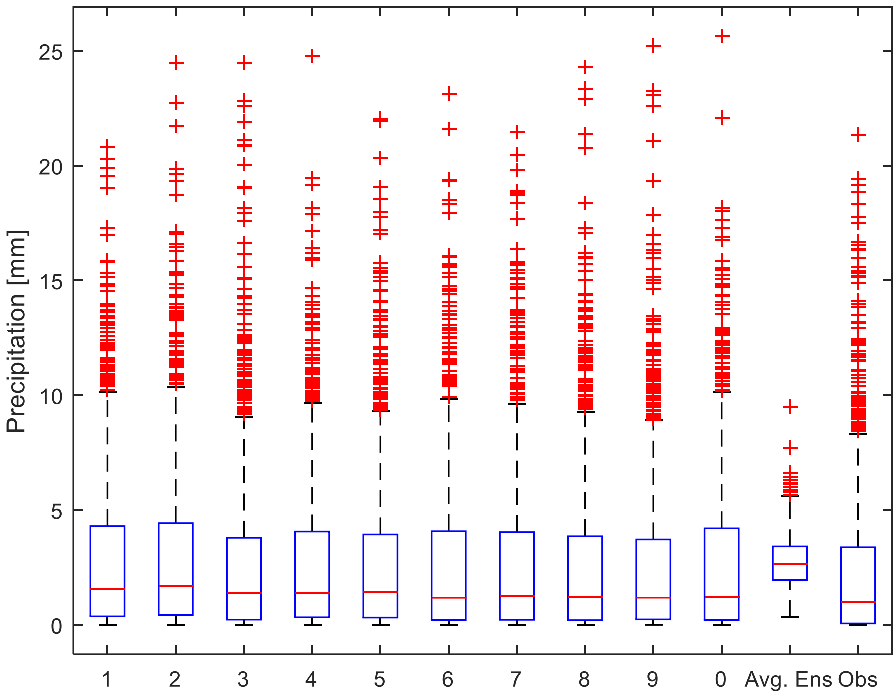


Figure 11 – EraClim (weighted) basin averaged precipitation against observed HYRAS (weighted) basin averaged precipitation over the period 2003 – 2006. Avg. Ens corresponds with the average precipitation of all EC ensembles over the whole period.

Figure 11 shows the (weighted) basin average EraClim precipitation for each of the 10 EraClim ensemble members (numbers 1 to 0). Also an ensemble average is shown which represents the daily ensemble average. The HYRAS reference case (Obs) is shown on the right. Looking to ensemble members zero to nine, the differences are not that large with respect to the HYRAS weather observations. All boxes seem to be equally large in terms of order of magnitude. This means that the lower and upper limit of all boxes (25 and 75 percent of the precipitation data) lies more less in the same interval. However the whiskers of the EraClim ensemble members are somewhat higher than those of the HYRAS observations. This means that somewhat higher precipitations are normal using EraClim. Further it looks like every ensemble member has a same proportion of outliers above the 1.5 times inter quartile range (IQR). There is a small variation visible in whisker length. However the EraClim whiskers are always larger than the observed HYRAS one. This points in the direction of more basin-average precipitation for more extreme weather. Reviewing the red outliers it can be seen that these extreme value distribution above the whisker distributions for HYRAS observations (Obs) and EraClim are more or less the same.

A notable issue concerns the first quartile differences of all ensemble members with the 'observed' boxplot. All precipitation ensembles have higher first quartile (25%) boxplots than the boxplot with the HYRAS weather observations. The average 25 percentile equals 0,35 mm for EraClim while for the HYRAS weather observations it is only 0,10 mm. This implies that there are less days with no or less precipitation using the EraClim ensemble members. Using a hydrological model, this might lead to a wetter state of the system because there would be more forcing of precipitation input. Last thing which is visible in figure 11 is that averaging over all ensemble members (Avg. Ens) will not give an appropriate result since the ensemble average boxplot differs quite a lot from the observed boxplot. In this case high and low precipitations of individual ensemble members are cancelling out each other. In appendix D all boxplot statistics are considered, including ensemble average and standard deviations. These averages and standard deviation do not differ that much with respect to the HYRAS weather observations.

For GLOFAS a similar precipitation analysis is performed. The boxplot with 51 precipitation ensemble members is shown in figure 12. Comparing the blue boxes of the ensemble members with the blue observed HYRAS precipitation, the GLOFAS dataset does not differ a lot from the EraClim precipitation data. Taking into account the red outliers however, much larger extreme precipitation is visible than the EraClim dataset provides. Especially when is mentioned that there are some outliers larger than 40 mm. These points are not plotted in the figure because of visibility considerations of the boxes, but they reach up to 60 mm. Part of the red extremes for GLOFAS precipitation is much more than observed. The maximum observed extreme precipitation equals 21 mm. For GLOFAS also extreme precipitation of 30, 40 and 60 mm show up. On the other hand also some weather ensemble members display lower extremes than observed, so the probability on having an outlier is supposed to be more less equal.

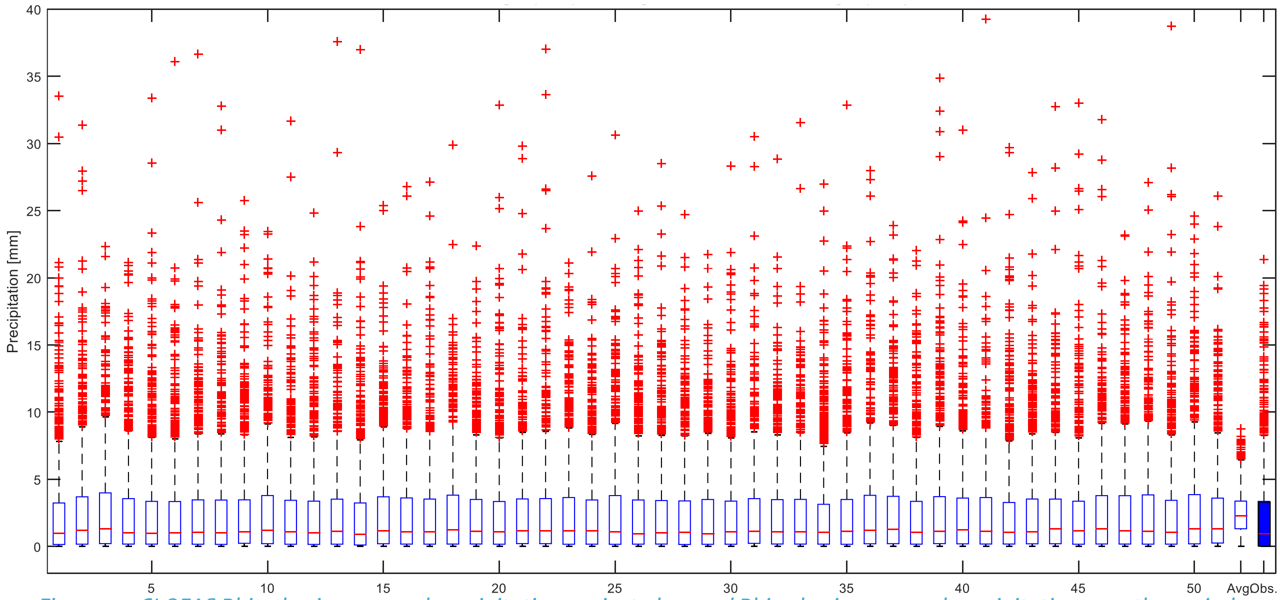


Figure 12 - GLOFAS Rhine basin averaged precipitation against observed Rhine basin averaged precipitation over the period 2003 – 2006. Avg. Ens corresponds with the average precipitation of all ensemble members over the whole period. The filled blue box corresponds with the 'observed' boxplot.

Figure 13 shows the basin-average monthly precipitation during the period of overlap. As can be seen, the average precipitation is consistently higher for both GLOFAS and EraClim compared to the HYRAS observed weather. These higher monthly precipitation sums can be obtained because EraClim boxplot whiskers (figure 11) are higher and GLOFAS produces higher extreme precipitation than the HYRAS reference. However, the period of overlap is too short (4 years only) too propose this as a very firm and solid conclusion. On the other hand not more overlap data is available, so it seems that GLOFAS and EraClim are generally wetter (more precipitation) than observed.

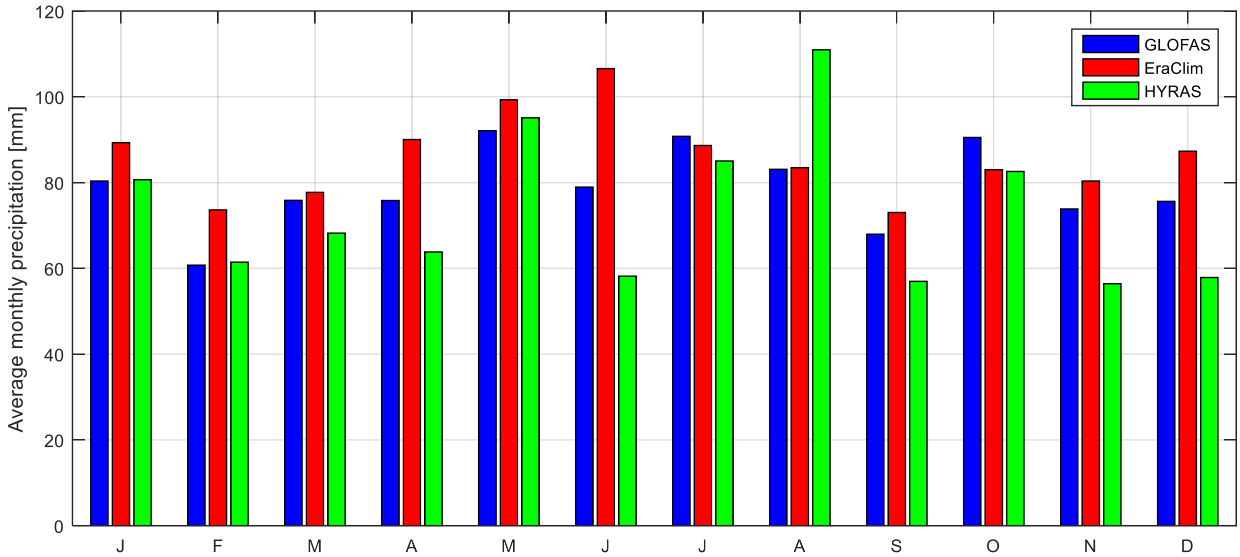


Figure 13 - Average monthly precipitation during the period 2003 – 2006 for three different weather sets.

It is already mentioned that both EraClim and GLOFAS ensemble members seem to produce less dry days (figure 11 and 12). Their lower boxplot limits are higher than the HYRAS reference during 2003-2006. Therefore figure 14 is drafted showing the average relative degree of dryness (ARDD) in the river Rhine system. To calculate the ARDD the percentage of dry days is assessed for every sub-catchment. Afterwards this is multiplied by the area fraction of the sub catchment. Finally all sub catchment contributions are summed. Based on EraClim for example, it becomes visible that the HYRAS observed ARDD lies consistently higher with an percentage of almost 25%, meaning less dry days for EraClim compared to the HYRAS observations. GLOFAS is performing more less in the same way, but shows a slightly higher dryness than EraClim. However, both EraClim and GLOFAS seems to have a significant lower dryness, meaning that the system will be more forced with precipitation. This will cause a wetter state of the system, leading to a higher base flow. Therefore less extreme precipitation is expected to cause peak discharges. Possible explanation for both 'wetter' products could be that they are built for usage in extreme wet periods. For such a case the number of dry days is of less importance than the wet ones.

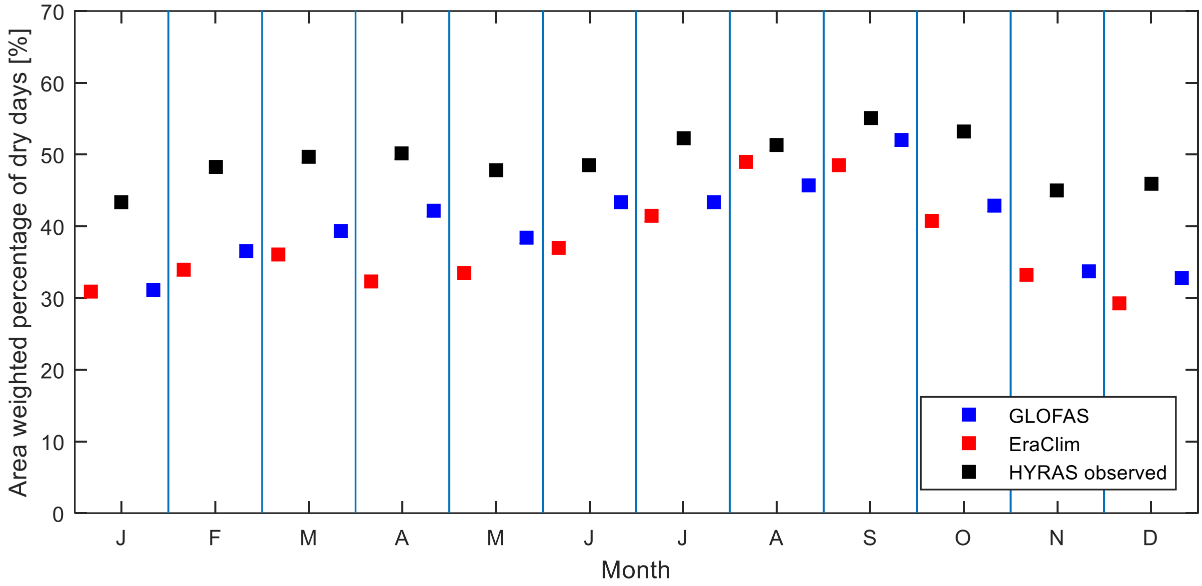


Figure 14 – Average percentage of dry days (P < 0,3 mm, (MeteoGroup, 2016)) during every month based on the entire datasets of GLOFAS, EraClim and HYRAS.

Discharge analysis

After the precipitation analysis, the discharges at Lobith are analysed for both ensemble products. For this purpose GLOFAS and EraClim ensemble weather is transformed into river discharges (then called ensemble streamflows) using the HBV model. These ensemble streamflows should be compared with a reference discharge set. From appendix D can be derived that the observed discharges at Lobith are taken in the period 2003 – 2010.

Before a detailed discharge analysis is performed, figure 15 is drafted. Herein the average monthly discharges of the HYRAS reference are compared to GLOFAS and EraClim. The full datasets are taken into account. It can be seen that during the months April up to July the discharges of GLOFAS and EraClim show large overestimations compared to the HYRAS reference. The rest of the year GLOFAS and EraClim merely show underestimations with respect to the HYRAS reference. The next sub-section goes deeper into the statistical discharge analysis for both EraClim and GLOFAS.

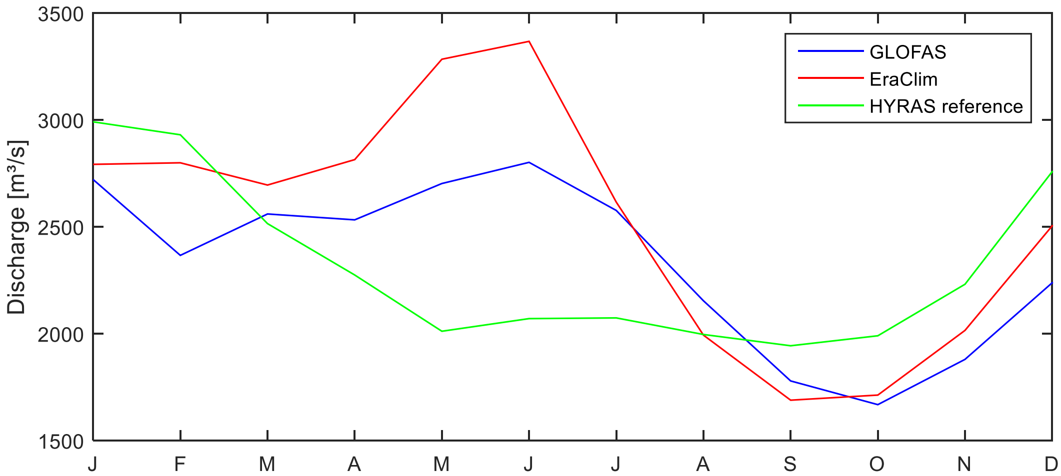


Figure 15 – Average discharge throughout the year for different datasets.

EraClim

Figure 16 shows 12 monthly EraClim boxplots of Lobith discharge. Every dataset consist of a maximum of 8 · 31 = 248 data points (all January observations during 2003-2010). For each ensemble member the whiskers display the 1.5 times the inter quartile ranges. Within this range the discharges are statistically not be considered as extreme. In general the observed blue discharge boxes are overlapping one or more ensemble member discharge boxes, but the 50^{th} percentiles (red line in the boxes) are highly variable as can be seen for two examples in figure 16. This might be caused by the nature of these EraClim ensemble members, which are built to describe a wide weather range. Due to this fact it is not strange that discharges are also varying a lot. Another noticeable point concerns the months May, June and July. They all show a quite large difference in the location of the ensemble member boxes (appendix D). The 75 percentile observed discharge in May is about 2370 m³/s. All 25 percentiles of the individual ensemble members are higher than this observed 75 percentile, so the EraClim ensemble members overestimate the discharges during these months. Further the observed upper whisker equals 3280 m³/s, while this is sometimes only the 50 percentile of some ensemble members. Only the months May, June and in a lesser extend July perform in a similar way.

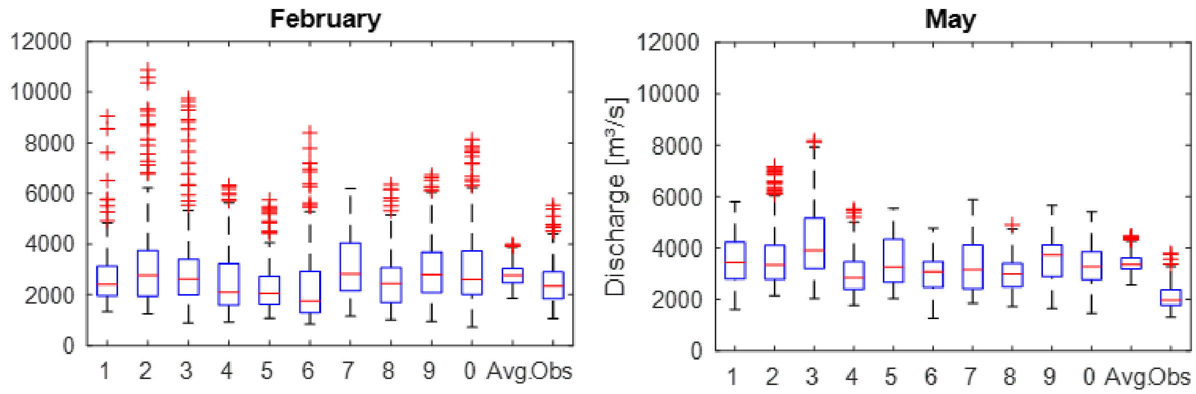


Figure 16 – EraClim based discharges at Lobith against observed Lobith discharge during the period 2003 -2010. Avg. Ens corresponds with the average precipitation of all ensemble members over the whole period. Only February and May are depicted. Appendix D contains the rest of the months.

GLOFAS

For the GLOFAS ensemble members similar plots can be drafted. These plots consist of 51 ensembles and are shown in appendix D. For almost all months the GLOFAS ensemble members show more less the similar results as the EraClim ensembles. Most of the observations (blue boxes) show a big overlap with the individual ensemble members. Meaning that half of the data points are in the same order of magnitude. There is also less variation visible between the 50^{th} percentile red lines. This last thing is much easier to reach with GLOFAS ensemble members than with the discharge ensemble members derived from EraClim data. GLOFAS weather ensemble members are expected to vary in a lesser extent than EraClim because the length of each GLOFAS ensemble member is only 15 days. For EraClim this is length is at max 20 years. So EraClim ensemble members reserve the right to have more freedom before new initial conditions (based on the weather observations) are added. Even as for the EraClim stream flows, it looks like the GLOFAS stream flows are overestimating the river discharges in the months May, June and July.

Flow duration curves

When looking at extreme discharges during the year, the moment of occurrence of an extreme discharge event is not relevant since only one annual maximum is taken. Rather the absolute value of the extreme discharge is of interest. Figure 17 shows flow duration curves (FDCs) during the period of investigation (2003 -2010) for both discharges of the EraClim ensemble members (10, red) and the GLOFAS ensemble members (51, blue). These are compared to the observed discharges at Lobith between 2003 and 2010. This figure displays a well visible difference in observed and ensemble distributions. Discharges ranging from 1200 m³/s to 14000 m³/s have a larger probability of occurrence for both the EraClim and GLOFAS ensemble streamflows. This is because of the majority of these ensemble curves lies above the observed discharge curve (black line) in the above called discharge interval. Only low and very high EraClim discharge events show a nice spread around the observed flow duration curve. When looking to the flow duration curves of the blue GLOFAS ensemble stream flows, a much better spread is visible around the observed black curve. This implies that GLOFAS stream flows might have more potential to describe a discharge distribution with more resemblance to the observed discharge distribution. Moreover, annual maximum discharges are not always that large; maxima around 3000 or 4000 m³/s will also occur. When using EraClim these discharges (3000 or 4000 m³/s) occur with a much larger frequency than observed.

According to this figure it might be concluded that EraClim ensemble streamflows have flow duration curves which are overestimating the observed flow duration curve. GLOFAS on the other hand overestimates the observed flow duration curve in a lesser extent. Some GLOFAS ensemble FDCs are even lower than the observed FDC. The position of GLOFAS FDCs with respect to the observed FDC is much better than EraClim. For that reason GLOFAS seems to have more potential than EraClim as an ensemble product. However more research is required in the field of extreme discharge distributions of long weather series for GLOFAS and EraClim.

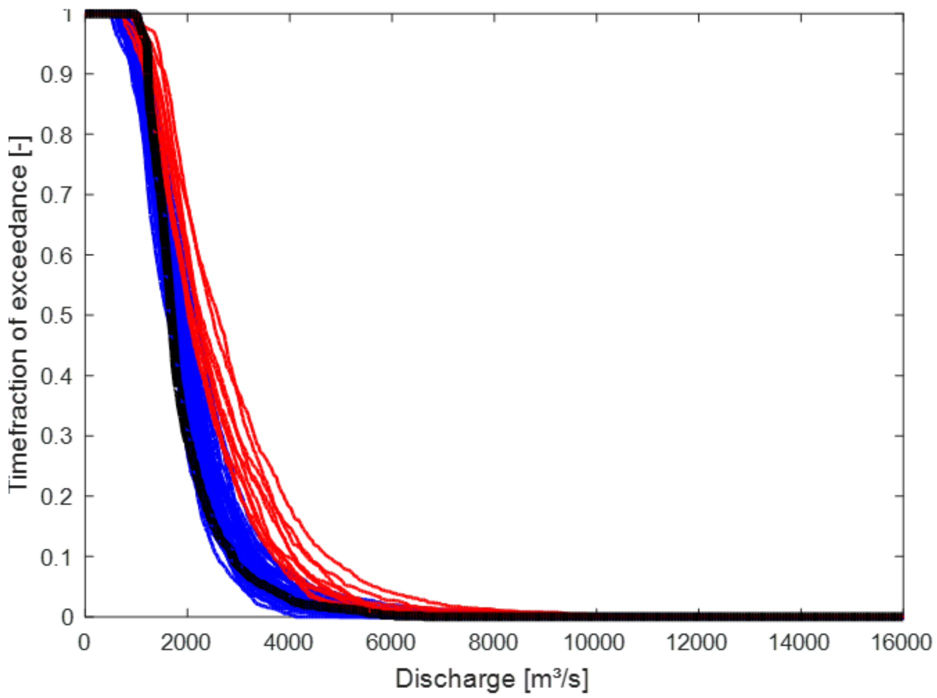


Figure 17 – Time fraction that a certain discharge is exceeded at Lobith for both observed (black), EraClim ensemble members (red) and GLOFAS ensemble member (blue) discharges. Plotting period is between 2003 and 2010.

4.2 Mutual ensemble member dependency

For each of the 51 chronological GLOFAS precipitation ensemble members the correlation is reviewed with each of the other ensemble members of the same sub basin. Figure 18 shows the resulting plot. As can be seen from this figure, the correlation distributions show high similarities for each of the weather ensemble members. The average correlation coefficient equals 0,16. Such a correlation is classified as very weak (Tilburg University, 2016). Even the somewhat higher correlations of 0,30 are supposed as weak. The average correlation of ensemble member 51 (the last one) in figure 18 shows a higher average correlation with all other ensemble precipitations. This is caused by the fact that one unperturbed ensemble member is present which describes the physical weather model most accurate. All other ensemble members are the perturbed derivations of this control ensemble (ECMWF, 2016).

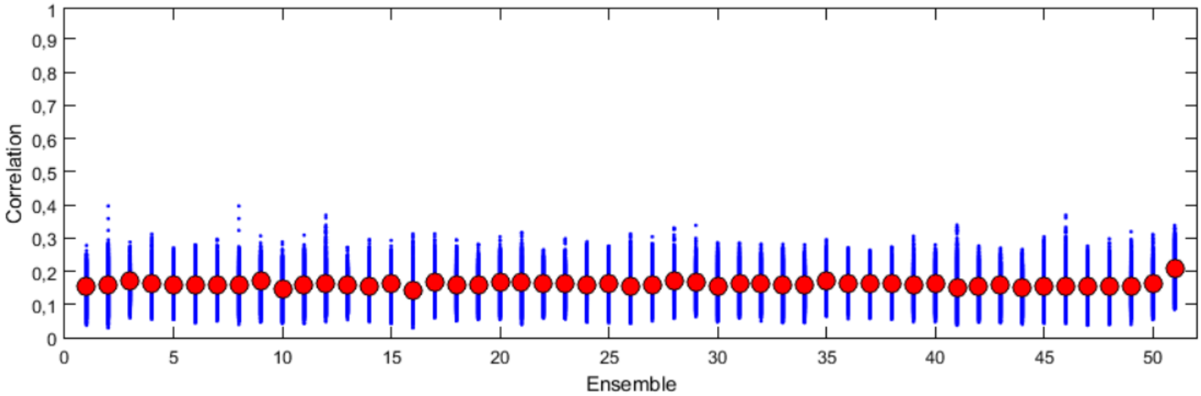


Figure 18 – Mutual correlations between a specific ensemble member with each of the other ensemble members. For every ensemble member x the blue dots display the correlations with the other ensemble members. The comparison is executed for each of the sub catchments 1 up to 134. So for ensemble member 1 a comparison is made with ensemble members 2 – 51. This is done for the catchments 1 – 134, so the first series consist of 6700 blue dots. The average of these dots is given by the red spot.

Also a closer look is given to the mutual ensemble member correlation per sub-catchment. Higher correlations are found in sub-catchments in the Alpine regions (see appendix F). Despite the correlations in mountain areas are higher (more less 0,20) one could still defend a very weak mutual precipitation correlation between GLOFAS ensemble members. The same test is also performed between all EraClim weather ensemble members. For EraClim even lower mutual ensemble member correlations are obtained of 0,006. These can be found in appendix F.

4.3 Selecting GLOFAS time series

Before an analysis can be performed using GLOFAS data it is important to know if it is better to use a GLOFAS series based on chronological calendar days (short: chronological series) or GLOFAS series with the highest level of synthetic randomness (short: synthetic series). For the synthetic series segments of 4-days are linked in order to construct a long enough time series. This number is randomly chosen. One could also choose to construct infinite long weather series with weather segments of just 1 day. For the chronological GLOFAS series weather segments from 6 – 10 January are adjacent to the one of 1 – 5 January. Using the whole available dataset, 11 years of chronological weather can be constructed. Hereby the weather consistency seems most plausible. Because 51 ensemble members are present, 612 annual peak discharges are modelled (51 ensemble members times 11 hydrological years which are put in sequence). Disadvantage of this chronological GLOFAS series is the lack of annual maxima for large return periods in an extreme discharge distribution. Therefore also synthetic GLOFAS series are constructed which are not fully chronological. Only weather consistency on a monthly basis is taken into account (see methodology, section 3.1). To investigate if rough weather characteristics of both chronological and synthetic GLOFAS series correspond to each other a statistical analysis is performed in order to map monthly average characteristics.

Precipitation

Figure 19 shows the average monthly precipitation for the Rhine catchment upstream from Lobith. Comparing the chronological GLOFAS series to the synthetic GLOFAS series, no significant differences are visible. Possible reason might be that synthetic GLOFAS series are drawn from the same ensemble segments than the chronological series, but only in a different order.

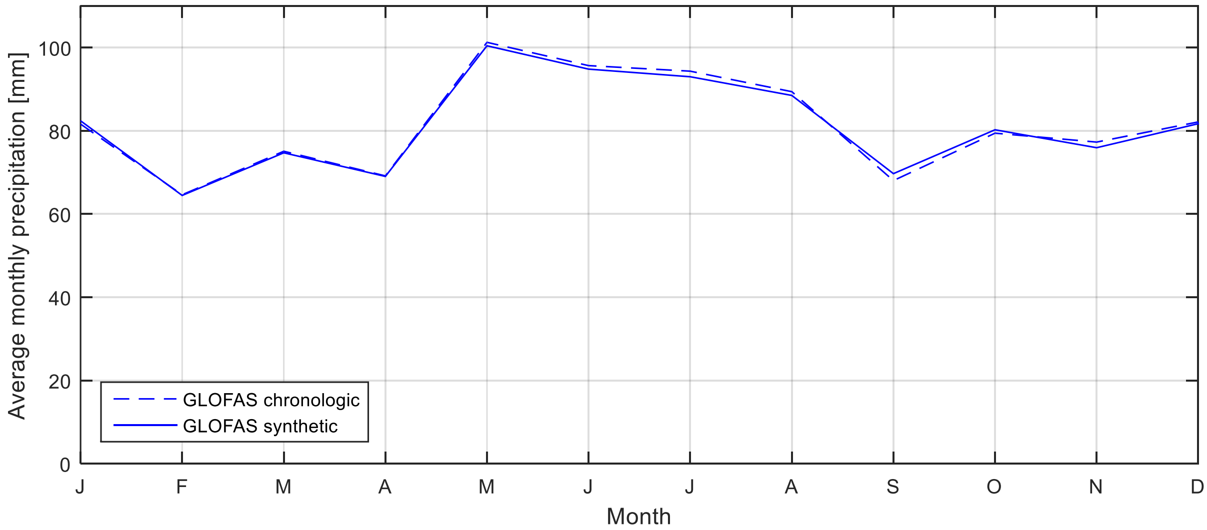


Figure 19 – Monthly averaged precipitation for the Rhine catchment upstream from Lobith for GLOFAS chronologic and GLOFAS synthetic series.

Temperature

Figure 20 only shows the input precipitations for HBV. However, a certain fraction of this precipitation will evaporate, using temperature as only forcing variable. Therefore the monthly average temperature for the area upstream from Lobith is shown in figure 20. Differences between chronological and synthetic GLOFAS series are negligible again. When there is still water available in the soil layers there is more potential to evaporate for HYRAS than for both GLOFAS sets. EraClim shows the lowest temperatures as can be seen from figure 20. The potential for higher discharges for GLOFAS series and EraClim series seems even more plausible now. Reviewing these results it can be stated that both EraClim and GLOFAS temperatures perform bad for data provided by ECMWF weather models.

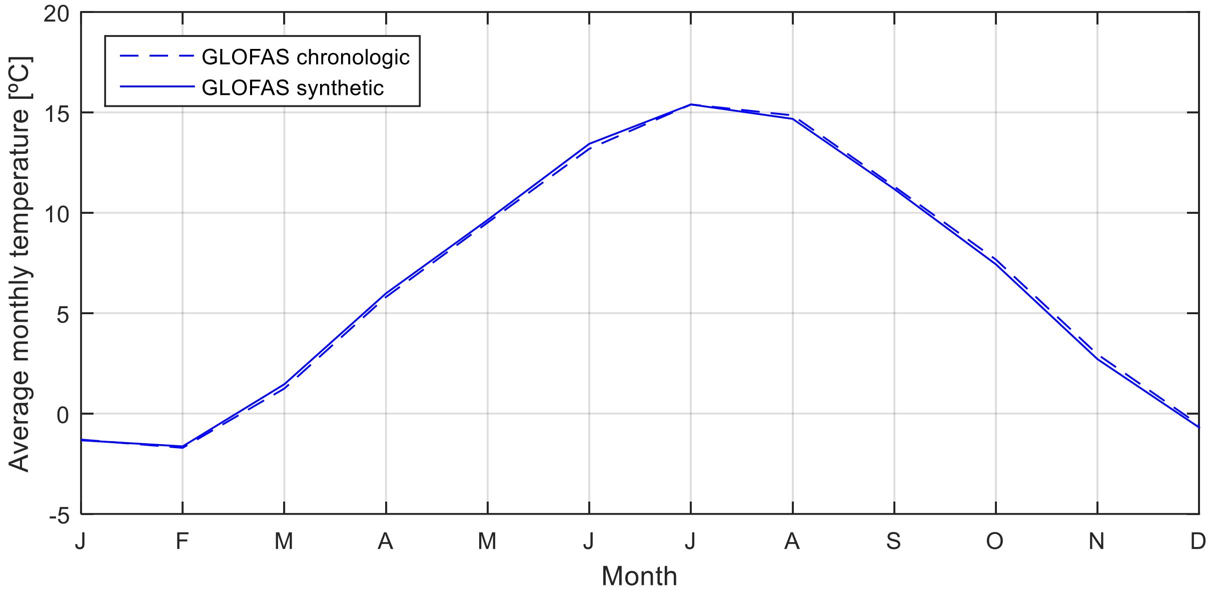


Figure 20 – Monthly average temperature for the Rhine basin upstream from Lobith for GLOFAS chronologic and GLOFAS synthetic.

Discharge

HBV is used as hydrological model to transform weather series into discharges at Lobith. Figure 21 shows the discharge distribution throughout the year. Again there is almost no difference between chronological or synthetic GLOFAS series. However, compared to the HYRAS modelled discharge reference, both EraClim and GLOFAS series show (much) more discharge.

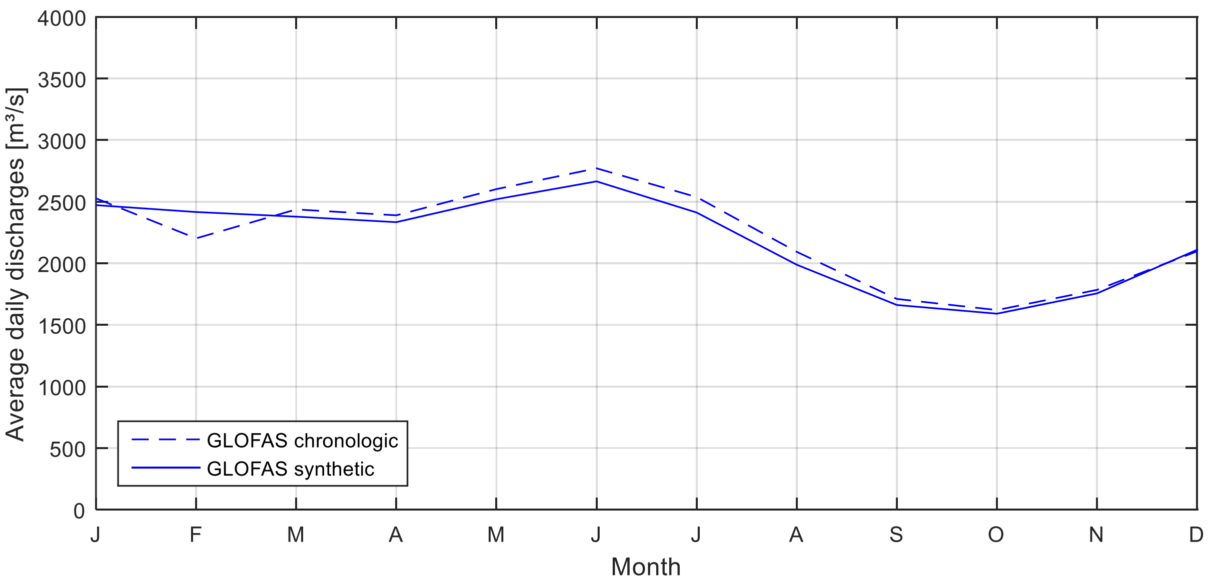


Figure 21 – Average monthly discharge for Lobith for GLOFAS chronologic and GLOFAS synthetic data series.

Discharge peak comparison

As can be seen before, rough basic statistics show a more less similar view for chronological and synthetic GLOFAS series. This does not directly mean that this behaviour is also applicable for the extreme discharge distributions. Therefore the peak frequency distribution is assessed in figure 22. There are slight differences between chronologic and synthetic GLOFAS series especially in the months April and July where the frequency bars are somewhat different for the synthetic GLOFAS series. However, in general the distributions from figure 22 show a more less similar pattern.

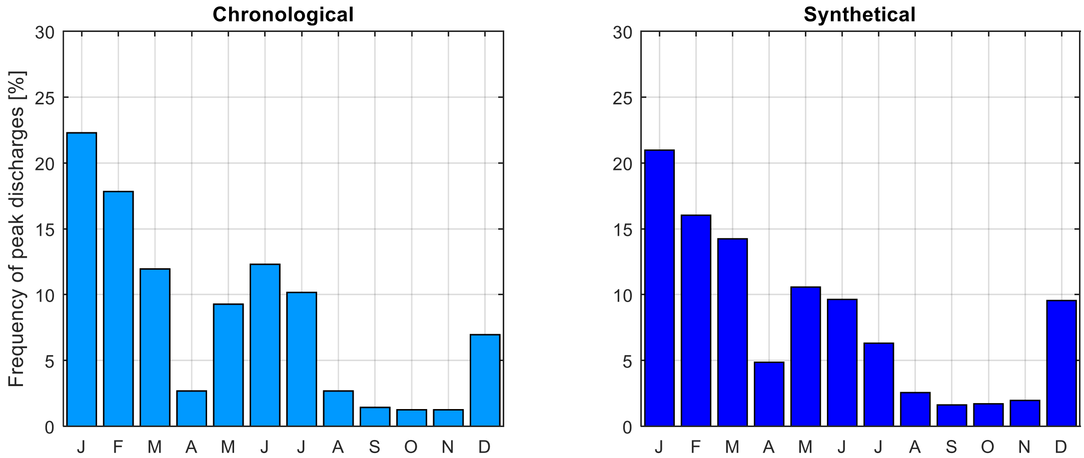


Figure 22 – Frequency plot according to the month wherein the peak discharge at Lobith occurs for chronological and synthetic GLOFAS series.

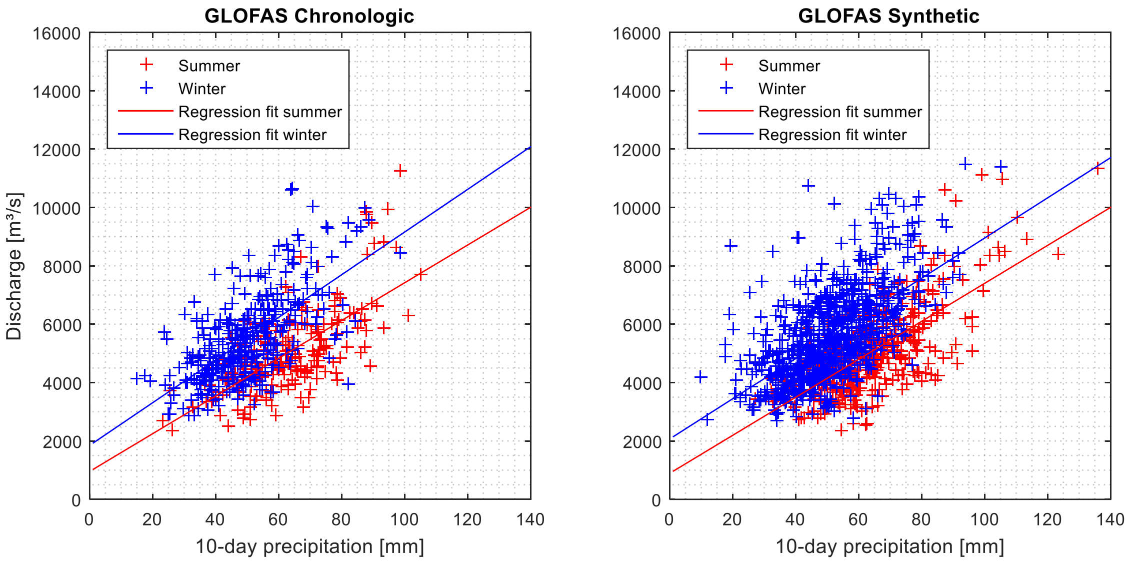


Figure 23 – 10 day precipitation prior to the discharge peak at Lobith plotted against modelled peak discharge for both chronological GLOFAS series as synthetic GLOFAS series.

Figure 23 shows a plot wherein peak discharge and the 10-day precipitation prior to the peak are plotted. For Lobith 10-day precipitation prior to the peak show highest correlations with the peak discharge. A distinction is made between winter and summer peaks and regression lines are fitted. Again, chronological and synthetic GLOFAS series does not seem very different. The y-intercepts and gradients are more less equal. Furthermore higher extreme precipitation is visible during summer months for both series 10 days in advance of the peak discharge. This corresponds with the analysis of the KNMI where precipitation intensities are higher in summer than during winters (KNMI, 2016).

For a complete view the precipitation duration curves are shown in figure 24. Comparing chronological and synthetic GLOFAS series shows no notable differences. Apparently the weather persistency in synthetic GLOFAS series is not that different from the chronological GLOFAS series because figure 24 shows comparative 10-day precipitation curves prior to the peak discharge. This can be explained by the design of the chronological GLOFAS series. Although they are put in a chronological day-to-day sequence, the chronological GLOFAS series still consist of 4-day long segments even like the synthetic GLOFAS series. Figure 25 shows this principle. The ensemble forecast from the 1st of January (blue box) is used to find weather data for 6, 7, 8 and the 9th of January. Four days later the forecast of the 5th of January (blue box) is used for weather data on the 10, 11, 12 and the 13th of January. The data learns that the overlapping parts of these two forecasts appear to be very different sometimes. This has to do with the perturbation of ensemble members; the forecast on the first of January gets another perturbation than the forecast on the fifth of January (despite the same ensemble number is used). This randomness is also included in the synthetic GLOFAS series. However, the ensemble member (1 – 51) is chosen here randomly and the 4-day long ensemble segments are used from a random moment of all January data (when constructing for the month January at least).

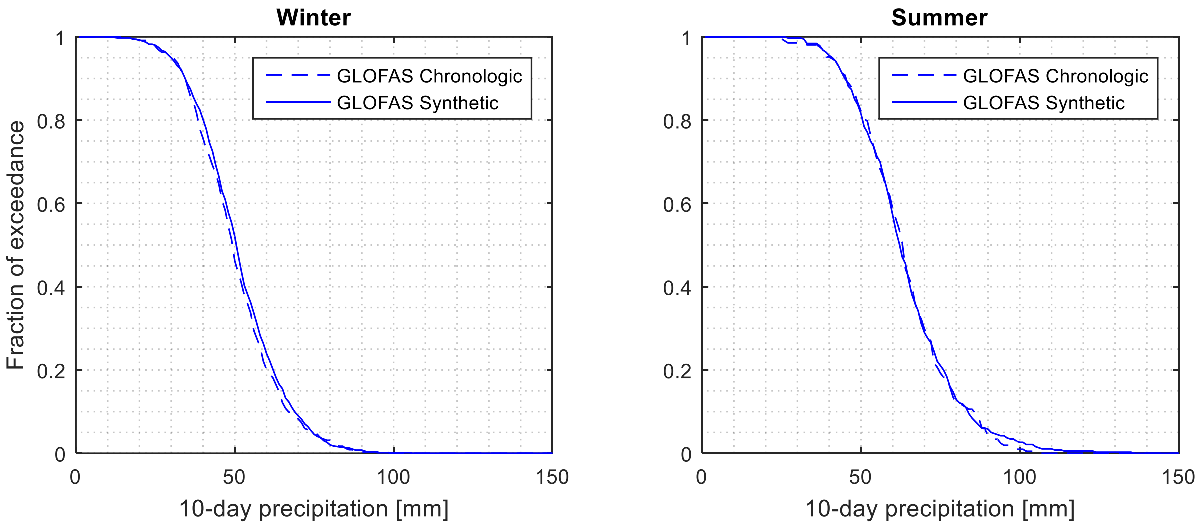


Figure 24 - Precipitation duration curves for 10-day precipitation sums prior to the discharge peaks at Lobith.

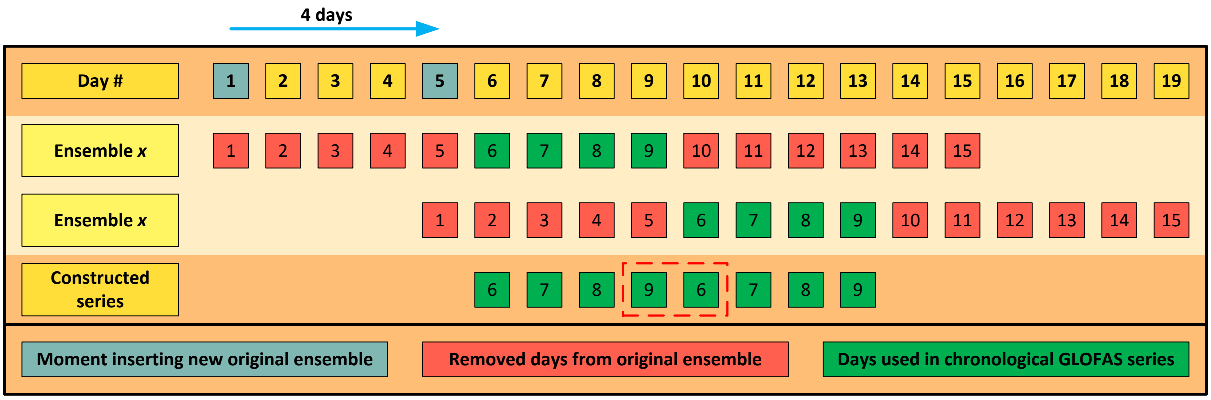


Figure 25 – Construction of chronological 4-day long GLOFAS weather series.

Because the more detailed precipitation statistics (figure 24) show minimal differences between chronological and synthetic GLOFAS series, the last step contains a comparison in the field of an extreme discharge distribution. Therefore figure 25 shows the annual maximum discharges at Lobith for 612 chronological and 1090 synthetic GLOFAS years plotted against their Gumbel variate. This difference is chosen, because EraClim also have 1090 data points. Despite this difference the global pattern of both datasets should be the same in order to favour the synthetic GLOFAS series. In line of what can be expected from the prior statistical analysis, the differences between chronological and synthetic GLOFAS series stay minimal, also when plotting a Gumbel distribution. Therefore only synthetic GLOFAS series are used from now on, because these have the potential to generate infinite long time series of.

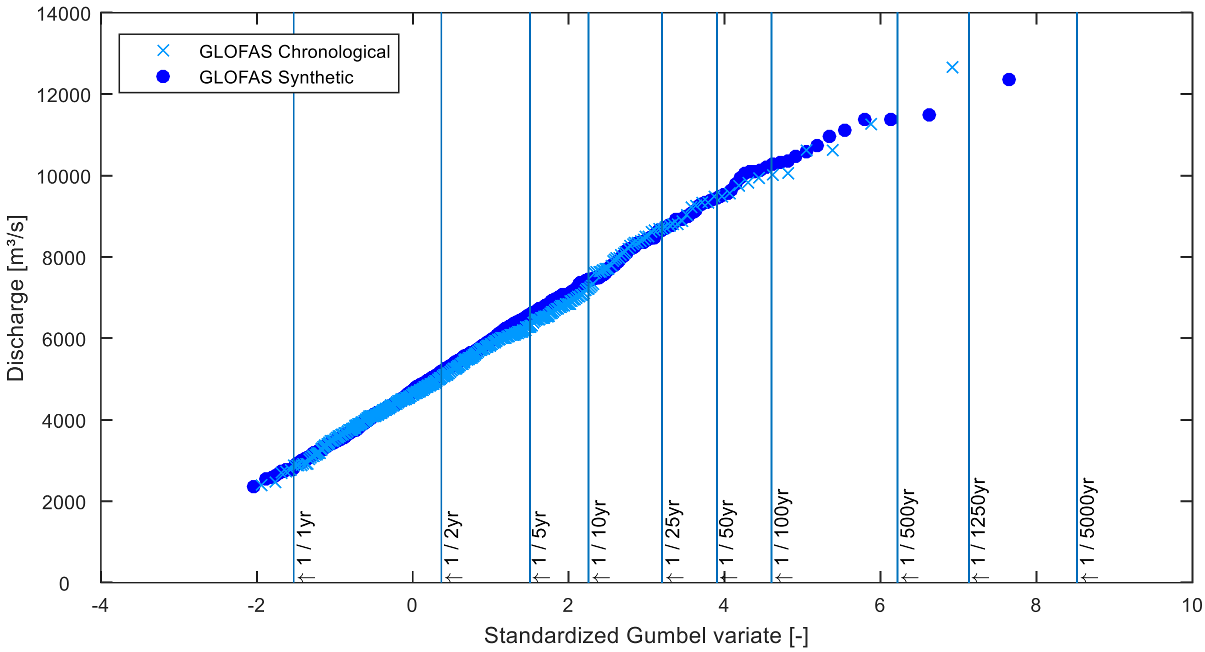


Figure 26 - Extreme discharge distribution at Lobith for both chronological (n = 612) and synthetic GLOFAS series (n = 1090).

4.4 Extreme discharge distributions Lobith

In order to formulate an answer to the third research question, the extreme discharge distribution at Lobith is of interest which is shown in figure 27. Before this plot is further analysed a probability plot correlation coefficient (PPCC) tests is applied to each of the point distributions. For Lobith it turned out that a Gumbel distribution can be used for each of the datasets. A Gumbel fit can be added to each dataset. The exact PPCC test is captured in appendix C. When the extreme discharge distributions in figure 27 are looked closer, it is visible that measured discharge observations at Lobith differ from the simulated HYRAS ones. Probably the model bias from HBV is a reason for this discrepancy.

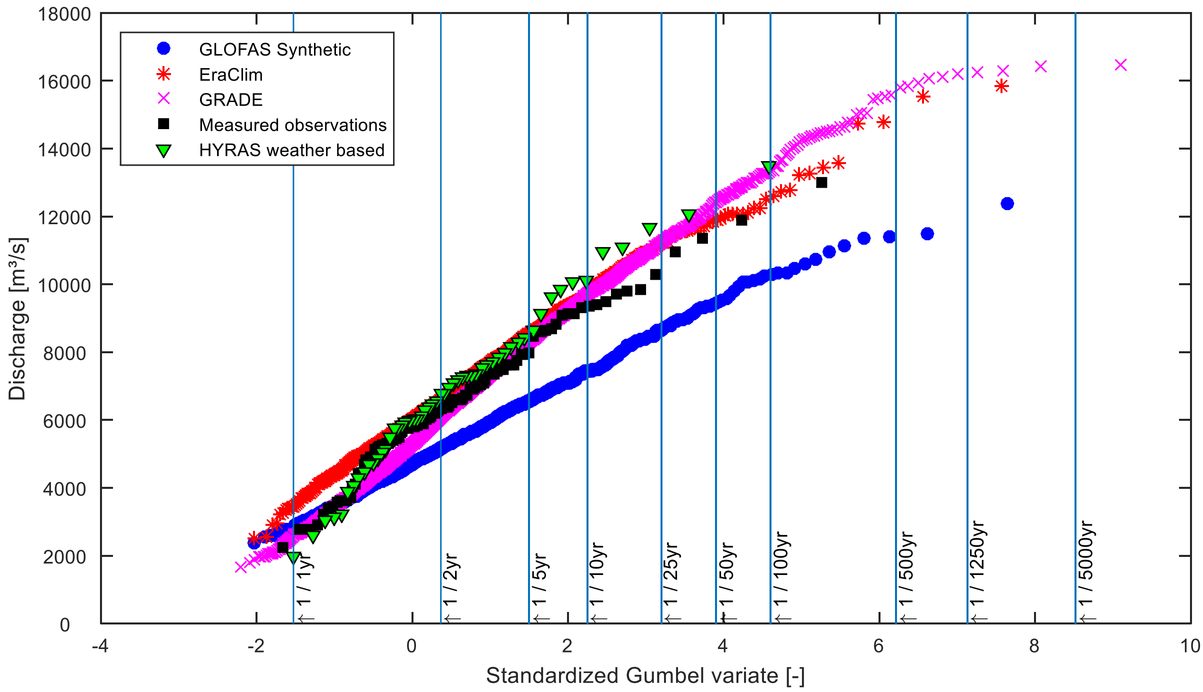


Figure 27 – Extreme discharge distribution at Lobith for more datasets; synthetic GLOFAS (n = 1090), EraClim (n = 1090), GRADE (n = 5000), observations (n = 109) and simulated HYRAS discharges (n = 55).

Comparing GLOFAS synthetic series (from now on named GLOFAS) to the observed discharges and simulated peak discharges of HYRAS, it gets visible that there are large underestimations for GLOFAS which already starts before a return period of 2 years is reached. EraClim on the other hand seems to produce extreme discharges which are more aligned to the observations and HYRAS simulation. For return periods larger than 10 years EraClim lies between observed Lobith peak discharges and HYRAS peak simulations. The 109 years real discharge observations is the most reliable set which is available. HYRAS however is also based on observations, but rather on temperature and precipitation input which is transformed into discharges using HBV. So both HYRAS and observations can be used as suitable reference situation, so the position of EraClim in between those two references is encouraging, also for larger return periods (>50 year).

GRADE has also been added to figure 27 in order to see how GLOFAS and EraClim perform compared to another 'synthetic' set. For the plotted series only the peak discharges were available. Also when using GRADE as a reference, GLOFAS produces very low peak discharges while EraClim overlaps quite well with GRADE for small return periods between 5 and 50 year.

Based on the Gumbel fits different datasets can be compared for different return periods. The red dots in figure 28 correspond with the Gumbel fit based peak discharge corresponding to a certain return period. These peak discharges of EraClim, GLOFAS and GRADE are compared to the average of the references (discharge observations and HYRAS). Results are shown below in table 2. As can be seen from this table, GLOFAS shows consistent lower peak discharges for every return period with more than 20 percent, when comparing it to the reference peak discharges (average of HYRAS peaks and observed peak discharges). With respect to the same reference both GRADE as EraClim are deviating less, anyway below 5 percent, which is assumed to be acceptable.

Table 2 – Relative difference (%) of design discharge for different return periods at Lobith with respect to the averaged design discharges of HYRAS and discharge observations.

Return period [y]	**5**	**10**	**25**	**50**
GLOFAS [%]	-20,6	-21,6	-22,6	-23,2
EraClim [%]	2,8	1,6	0,4	-0,4
GRADE [%]	-3,5	-1,9	-0,4	0,4

Since the interest lies in peak discharges for small return periods, figure 28 is drafted. The 95% confidence extrapolation intervals are provided by the black lines. An important note hereby is the absence of other errors. Errors in data, hydrological model, weather interpolation over the sub-catchments; it all can give uncertainties which are not taken into account here. For Grade figure 27 shows a flattening distribution for the highest return periods. From the PPCC test, officially a Gumbel fit may be used, but this fit is maybe not representative for the discharges belonging to the highest return periods. However, this research focusses on small (< 50 year) return periods.

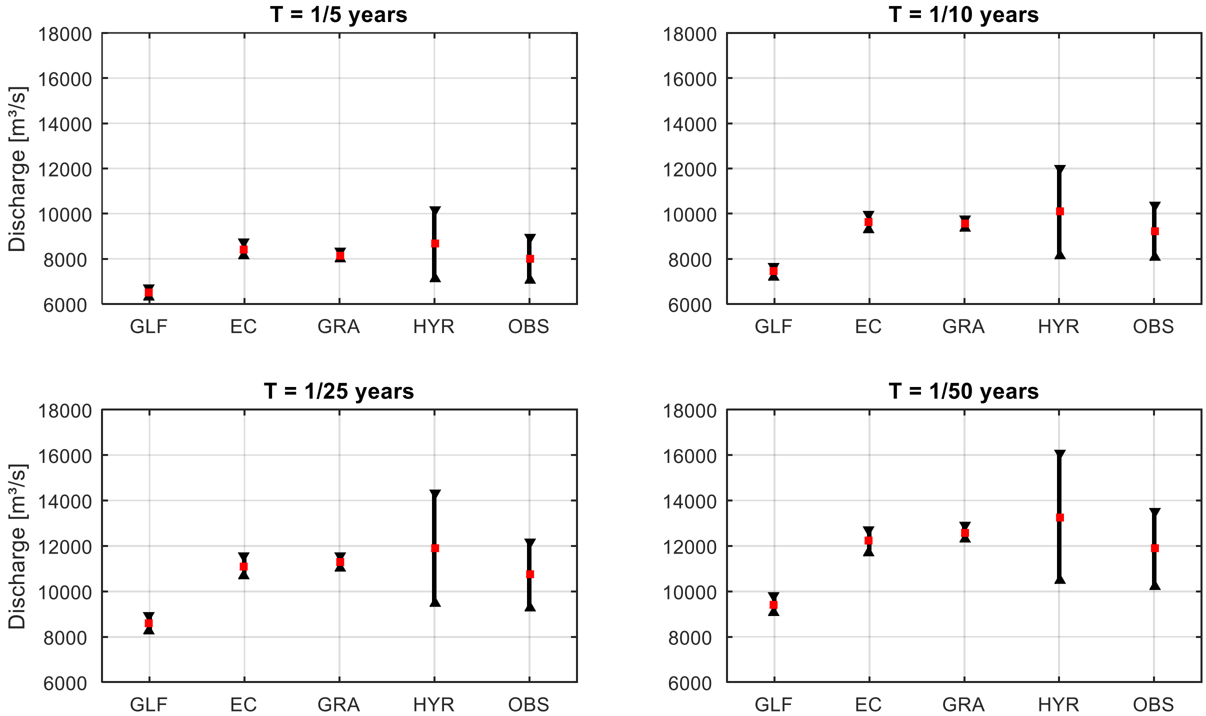


Figure 28 – Extreme discharge with 95% confidence limits for different datasets at Lobith based on Shaw et al. (2005). Plot is based on Gumbel fitted data. Red dot corresponds with fitted Gumbel discharge, black lines show the confidence limits. GLF = GLOFAS synthetic series (n = 1090), EC = EraClim series (n = 1090), GRA = GRADE weather generator series (n = 5000), HYR = HYRAS weather based discharges (n = 55), OBS = measured observations (n = 109).

As can be seen in figure 28, the 95% confident extrapolation limits for GLOFAS do not overlap the confidence limits for both references of HYRAS and measured observations at all. In consultation with the large deviations from table 2 it can be defended that GLOFAS does not describe a proper design discharge for small return periods. Larger return periods are not plotted in figure 28, but GLOFAS performs even worse there. EraClim on the other hand seems to perform more satisfactorily since their 95% confident extrapolation limits are fully positioned into the 95% confident extrapolation bandwidth of both HYRAS as the observed peak discharges. In addition the results of the EraClim Gumbel fitted line corresponds much better with the Gumbel averages of HYRAS and discharge observations (table 2). Based on this it cannot be rejected that EraClim is an unsuitable alternative to estimate design discharges for Lobith. For GLOFAS the opposite is applicable; extrapolation overlap bandwidth is less conveniently located with respect to the reference sets (HYRAS and observations). Furthermore GLOFAS Gumbel fit deviations are substantial.

Synthesis Lobith

It turns out that the synthetic GLOFAS series performs quite poor with respect to the observations and HYRAS simulated peak discharges. EraClim performs considerably better using the same reference. How these differences can occur is important. The explanation is expected to be in the field of precipitation, since these can be different with respect to the HYRAS weather reference. Firstly all peak discharges at Lobith are correlated with the summed n-day precipitation prior to the peak discharge. For Lobith it turned out that 10-day precipitation prior to the discharge peaks shows the highest correlation. In figure 29 these 10-day precipitation is plotted against the peak discharges. Also a distinction is made between summer and winter peaks.

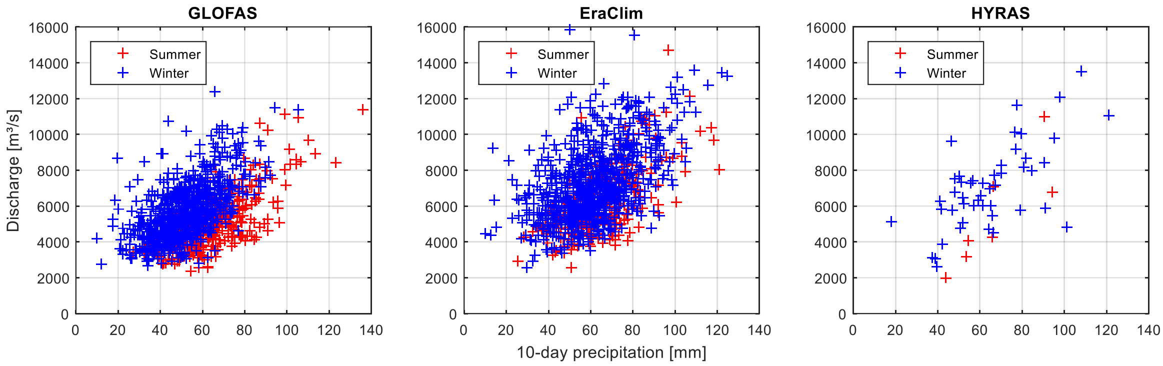


Figure 29 – 10-day precipitation prior to the peak discharge at Lobith. Total number of peaks; GLOFAS (n = 1090), EraClim (n = 1090), HYRAS (n = 55).

Roughly the same relationships between precipitation and discharge are visible in the scatterplots of figure 29. However it is also visible that GLOFAS produces relative less high precipitations during winter compared to HYRAS. EraClim seems to be more similar to the HYRAS reference. Figure 30 shows a relative frequency plot of the occurrence of discharge peaks. When using GLOFAS 32% of all discharge peaks are found during summer months. For EraClim this is less with 26%, while HYRAS shows only 13% of all peak discharges in summer (figure 30). These are remarkable differences and will be further evaluated in the next sub-section.

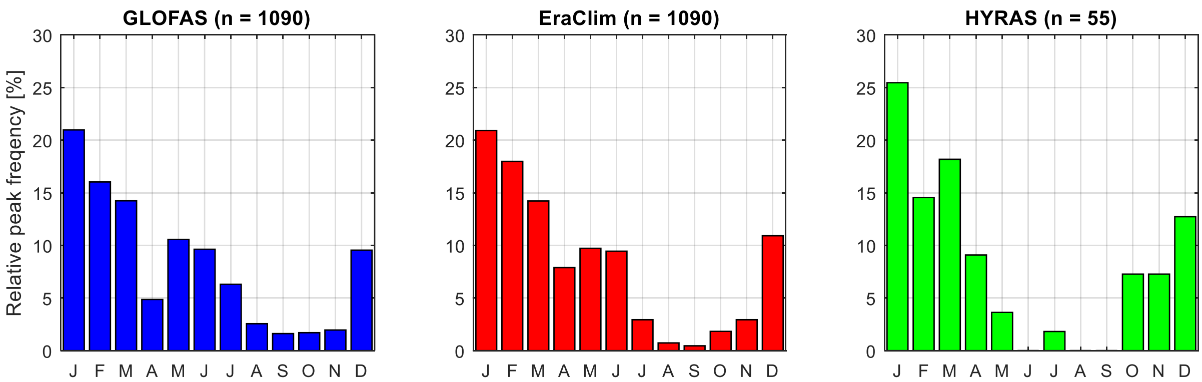


Figure 30 – Peak discharge frequency per month at Lobith for every dataset with n number of annual maximum discharges.

Because the real density of precipitation sums and discharges is not that easy to see, a precipitation duration curve is drafted in figure 31. The 10-day precipitation prior to the summer peaks shows a more less similar pattern for all three datasets. The precipitation prior to the winter discharges on the other hand is showing differences. It is clearly visible that there is much less basin average precipitation prior to the winter peaks when using GLOFAS. For GLOFAS the precipitation sums are larger than 50 mm in 50% of the cases. The HYRAS reference exceeds 63 mm of precipitation for 50% of the cases. So, GLOFAS receives in this case more than 20% less precipitation than obtained by weather observations. This relative difference is significant enough and will definitely affect the extreme discharge distribution in figure 27 and 28. The differences between the HYRAS reference and EraClim are much less. Only the higher HYRAS precipitation sums have a larger probability of occurrence than EraClim. This can be linked to the extreme discharge distribution, where one of the highest HYRAS extremes is positioned above the EraClim discharge extremes.

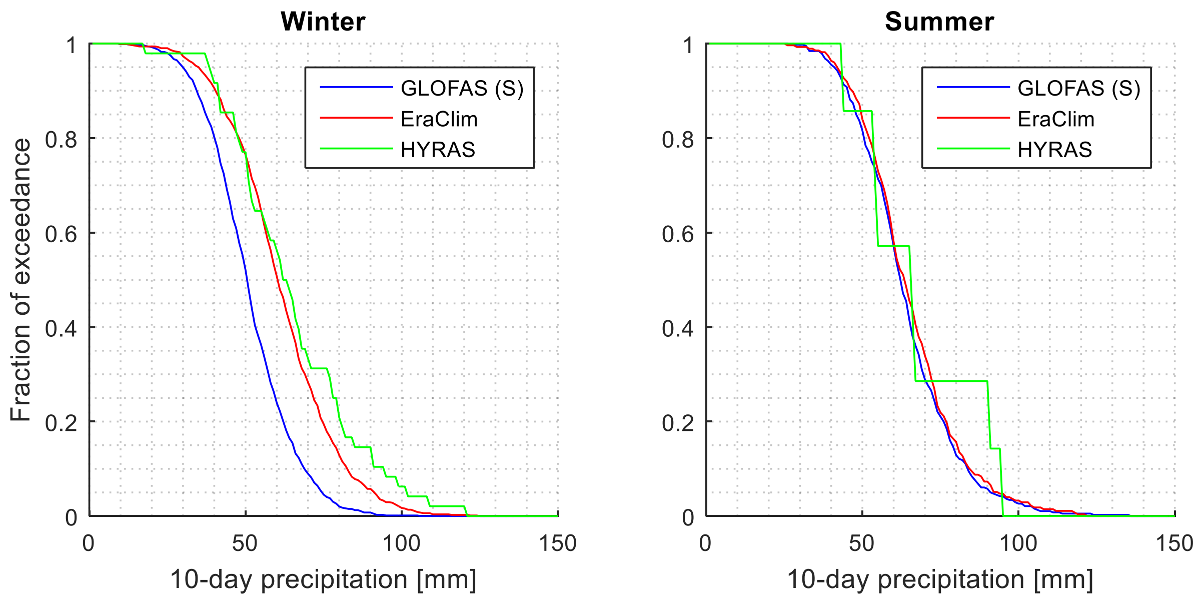


Figure 31 - Precipitation duration curves for different datasets at Lobith, 10-days prior to the peak discharge.

4.5 Upstream analysis

The previous paragraph explains why GLOFAS seems not useful for estimating design discharges at Lobith. EraClim seems to be better since its extreme discharge statistics match well with the HYRAS reference and measured observations. However, peak discharge frequencies at Lobith are obtained much more during summer when using both GLOFAS and EraClim compared to the HYRAS reference. Therefore the influence of the upstream system is assessed. The upstream basin can roughly be divided into two regimes; a rain-fed regime and a snow-melt regime. For rain-fed rivers the Mosel is found to be representative. For the Alps the West-Alpine basin will be discussed here. The discussion about the Neckar, Main and East-Alps is partly captured in the main-text here. Figures of these basins are captured in appendix G. At the end a summary will be provided in order to see how all investigated sub-basins perform.

Rain-fed rivers

A representative rain-fed river is the Moselle. This basin is used therefore to describe fundamentals in rain-fed rivers in the river Rhine basin. Figure 32 shows the extreme discharge distribution for the Moselle at Cochem. Both references; observed peak discharges and the ones of HYRAS show only small differences. When comparing GLOFAS to the HYRAS reference and the measured observations, a large underestimation is visible. The same applies when a comparison is made with the measured discharge peaks. The other rain-fed basins (appendix G) show also underestimations for GLOFAS but not as much as Cochem. However all these GLOFAS underestimations together can possibly accumulate to the large GLOFAS underestimation at Lobith. EraClim seems to produce a better estimation of extreme discharges since it approaches both HYRAS reference as measured observations quite well. In the other rain-fed rivers EraClim deviates also less from the HYRAS reference and measured observations. However, a small overestimation by EraClim is visible.

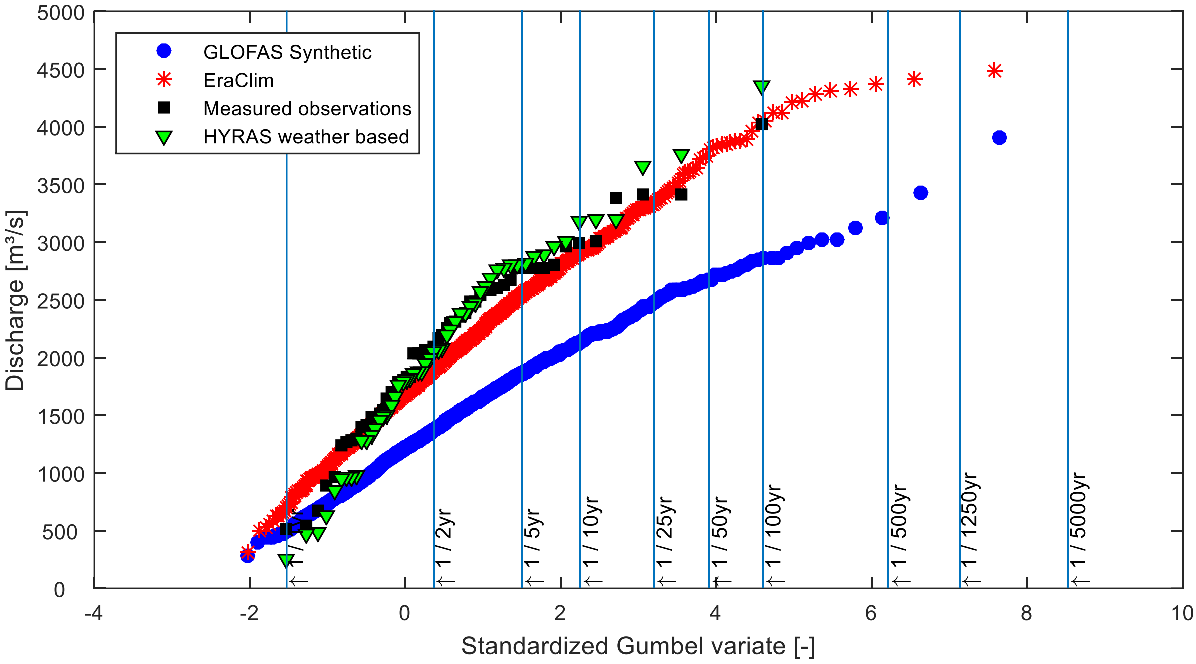


Figure 32 - Extreme discharge distribution at Cochem for more datasets; synthetic GLOFAS (n = 1090), EraClim (n = 1090), observations (n = 55) and simulated HYRAS discharges (n = 55).

The peak discharge distribution throughout the year (in which months do peak discharges occur) at Cochem is quite surprising when looking to figure 33. At Lobith there were relatively more summer peaks modelled for both GLOFAS and EraClim relative to the HYRAS reference. In the Moselle, but also in the other rain-fed rivers, this is not the case and peaks will mainly occur in winter (appendix G). At Cochem the HYRAS reference shows 10% summer peaks, while GLOFAS and EraClim show 8% and 4% respectively. Because this is characteristic for all rain fed basins, it becomes plausible that the summer peaks at Lobith are caused by the Alpine areas. This will be further analysed in the next section where the Alpine regions are analysed.

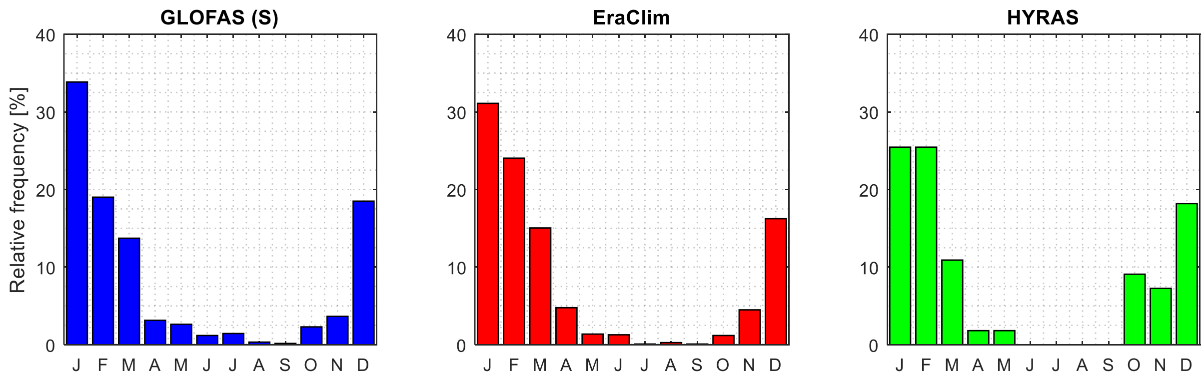


Figure 33 - Relative peak frequencies per month for multiple datasets at Cochem.

For Cochem the correlation between 4-day precipitation prior to peak discharge and the simulated peak discharges is the highest (figure 34). Regression lines are plotted for both winter and summer season. Focussing on the winter peaks (these are the majority) GLOFAS seems to produce relative more 4-day precipitation smaller than 60 mm compared to the HYRAS reference. However only

3.9% of the GLOFAS discharges at Cochem exceeds 2500 m³/s against 33% of the HYRAS reference. For EraClim a similar percentage (28 %) discharges above 2500m³/s is visible. However in the upper left quadrant is shown that EraClim produces more high (> 2500 m³/s) discharges than HYRAS while these EraClim peaks are preceded by less precipitation than in the HYRAS reference case. This might imply that EraClim has more wet soil conditions in the Moselle basin than the HYRAS reference. Also a higher regression line is visible for EraClim than for GLOFAS, so with the same amount of precipitation, more discharge is obtained. EraClim seems to be more wet than GLOFAS and the HYRAS reference, leading to higher discharge peaks. Note that this is only applicable during the period prior to the extreme discharge conditions.

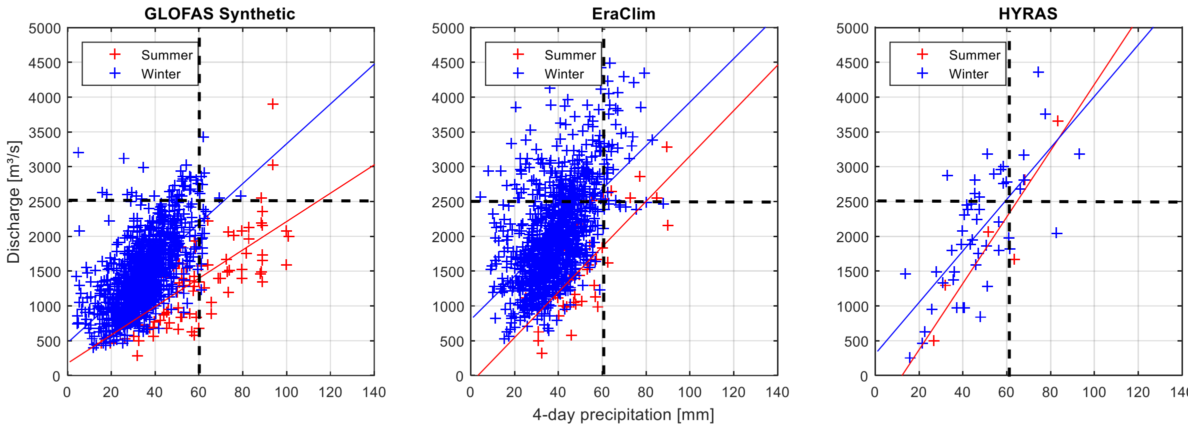


Figure 34 – Peak discharges plotted against 4-day precipitations prior to peak discharges at Cochem. Coloured lines represent seasonal regression fits.

To see a longer precipitation influence than only 4 days in the Moselle basin, figure 35 shows the 10 day precipitation sums prior to the peak discharge in a precipitation duration curve. Probably this is a better indicator to take also into account wetter soil conditions. The differences between the HYRAS reference and GLOFAS are quite large. GLOFAS produces much less precipitation prior to the peak discharge than for the HYRAS reference case. This view is obtained in all rain-fed sub basins (appendix G). This might be an explanation for GLOFAS where the extreme discharge distribution is also much lower than the HYRAS one. Large differences are visible too in 10-day precipitation distributions between the HYRAS reference and EraClim (figure 35). Lower extreme EraClim discharges compared to the HYRAS reference are expected therefore. However, this is not the case when looking to figure 32, where EraClim shows just a slightly lower extreme discharge distribution than the HYRAS reference. It can be stated that HYRAS only uses 48 winter peaks, so the confidence is not that high. Another possibility is that precipitation only does not fully explain the differences in extreme discharge distributions. When assessing the whole river Rhine basin for Lobith in the previous paragraph, it seems to be an indicator which explains differences in extreme discharge distributions well. On a smaller scale, looking to rain-fed sub basins it seems not to be the only explaining factor.

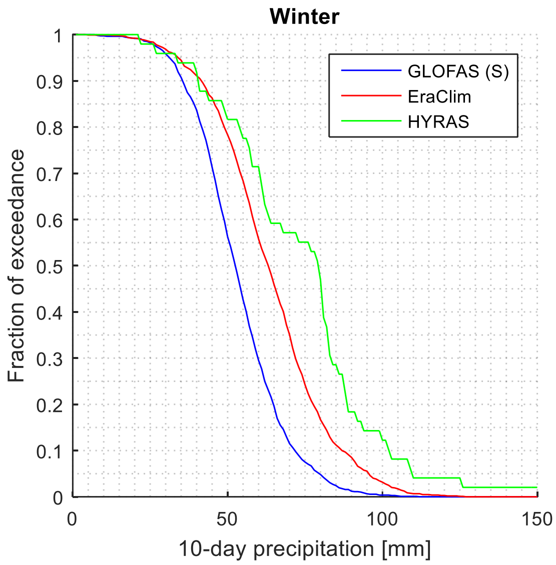


Figure 35 - Precipitation duration curves for different datasets at Cochem, 10 days prior to the peak discharge. Only winter curves are displayed since the majority of peak discharges occur during that period for all datasets.

The Main and Neckar show a similar view; higher extreme discharge distributions for EraClim than the HYRAS reference but lower 10-day EraClim precipitation prior to the peaks than the HYRAS reference. The precipitation deviations of GLOFAS and EraClim with respect to the HYRAS reference are smaller for the Main and Neckar than for the Moselle. Probably this causes less deviating extreme discharge distributions than the Moselle does. On the other hand, GLOFAS precipitations are still deviating more than EraClim with respect to the HYRAS reference. So also in the other rain-fed basins, the 10-day precipitation prior to the peak discharge is not the all-explaining indicator. For this reason it may be concluded that apparently other (smaller) hydrologic processes also play a role in the rest of the rain-fed sub basins.

Alpine areas

The extreme discharge distributions are quite different for the western- and eastern part of the Alpine region. This applies too for the frequency distribution for annual peaks over all months. Here only the Eastern Alps are discussed, with some links to the western Alps. Firstly the extreme discharge distribution for Rekingen is shown in figure 36. Something important what immediately stands out, is the discrepancy in references. HYRAS based discharges (green) and measured discharge observations (black) are deviating a lot. Probably the model parameterisation plays a role, so HYRAS is chosen to compare GLOFAS and EraClim with for this specific plot. This is done because HBV is used too to obtain GLOFAS or EraClim discharges at Rekingen and to transform HYRAS reference weather into HYRAS reference discharge at Rekingen.

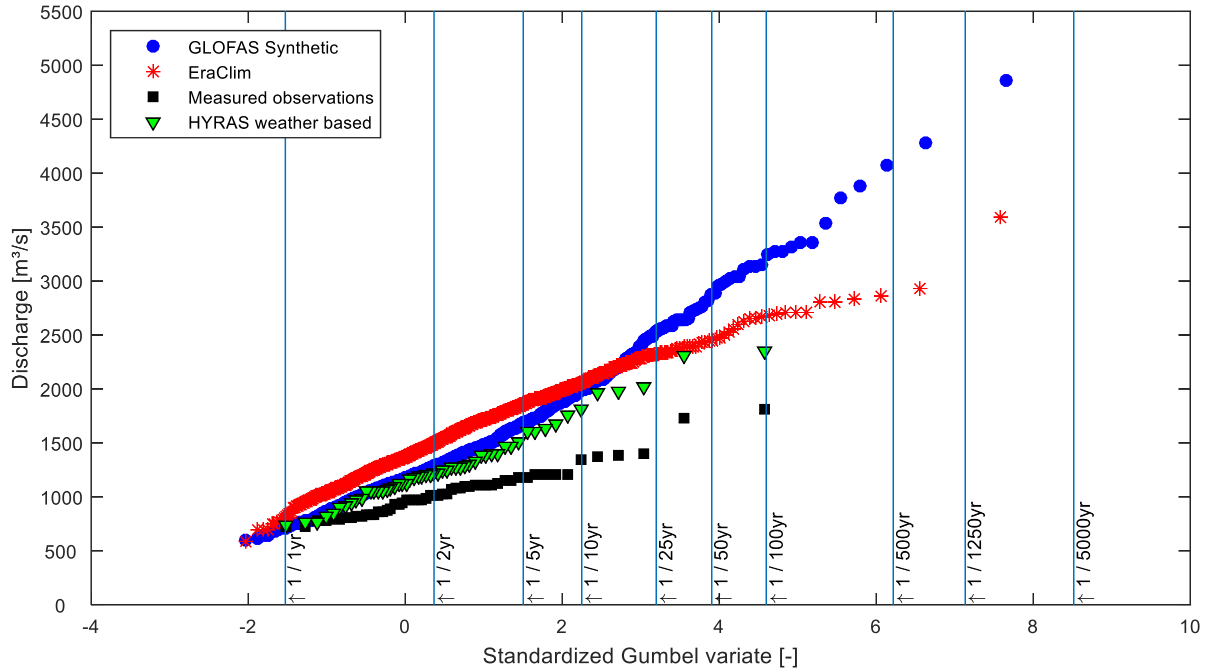


Figure 36 - Extreme discharge distribution at Rekingen for more datasets; synthetic GLOFAS (n = 1090), EraClim (n = 1090), observations (n = 55) and simulated HYRAS discharges (n = 55).

When figure 36 is considered it becomes clear that both GLOFAS and EraClim produce higher extreme discharges than the HYRAS reference (and measured observations). GLOFAS seems to be slightly higher than the HYRAS reference until a return period of 10 years. Thereafter the GLOFAS extreme discharges are increasing quickly. It is expected that temperature plays an important role in the Alpine regions in order to force the fraction of meltwater in the peak discharge at Rekingen. At first the relative monthly peak frequencies for Rekingen are plotted therefore in figure 37. From this it can be seen that GLOFAS and EraClim produce more peak discharge during summer (84% and 93% respectively) compared to the HYRAS reference (only 55% summer peaks). For the west Alpine area the majority of the peaks is also found during summer (appendix G). However, there the frequency distribution of GLOFAS and EraClim is more similar to the HYRAS reference. For both Alpine areas the period May, June and July show much higher peak frequencies than the HYRAS reference.

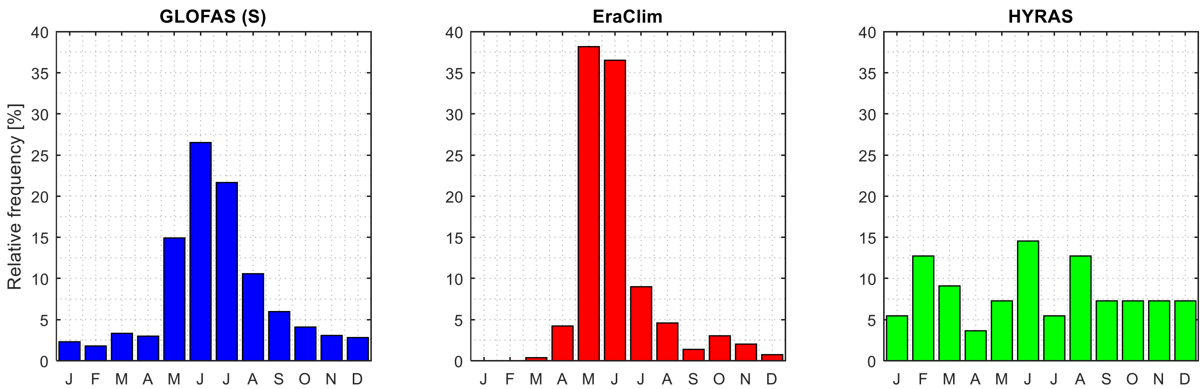


Figure 37 – Relative peak frequency per month for more datasets at Rekingen. GLOFAS (n = 1090), EraClim (n = 1090), observations (n = 55) and simulated HYRAS discharges (n = 55).

An explanation can be found in the number of frost days for the alpine basins. Figure 38 shows these frost-fractions in the eastern Alpine basin, but this figure is also representative for the western-alps. During the winter period more frost days are visible than the HYRAS temperatures which are used as a weather reference. This means that precipitation will be stored in the form of snow, and will not be discharged to the river system. When there is a no-frost day, the average temperature for GLOFAS is lower than the HYRAS observations. This means less snowmelt than the HYRAS reference, so during winter months even more snow stays stored for GLOFAS (figure 39). The temperatures for EraClim are even lower than those of GLOFAS, so the snow storage effects will be the largest for EraClim. This larger storage volume for GLOFAS and EraClim starts melting from April/May. This means that much more snowmelt will be added to the river system during the (beginning of the) summer period. A logical consequence is the higher peak frequency for GLOFAS and EraClim during summer months compared to the HYRAS reference. Because of the snow storage in both GLOFAS and EraClim during the winter period, the Alpine regions are expected to contribute barely to the winter peaks of Lobith. This means that winter peaks at Lobith are mainly caused by events in the rain-fed basins. The other way around, higher Alpine contributions to the summer peaks at Lobith are expected because of more Alpine snowmelt.

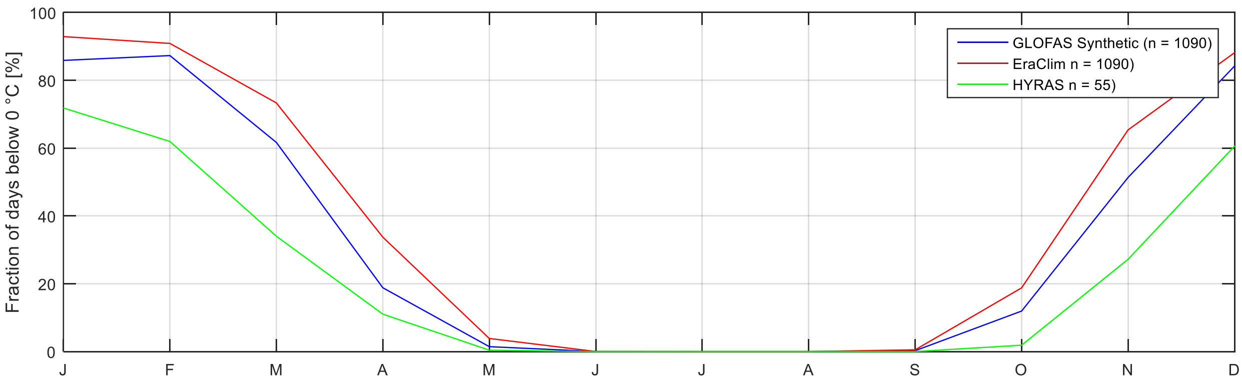


Figure 38 – Fraction of frost days with temperatures below 0 °C for different datasets.

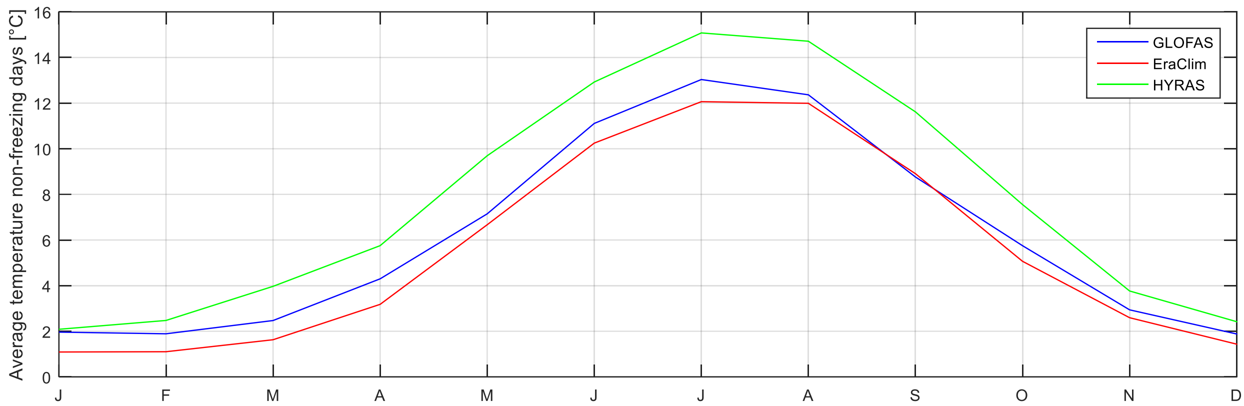


Figure 39 – Average temperature on non-freezing days in the eastern Alpine area.

For the Alpine regions the peaks are thus in summer, so it is important for this analysis to focus on the summer period therefore. Because the average summer temperature is larger than 0 °C the precipitation will be drained to the river system and has the potential to contribute to the peak discharge at Rekingen (or Untersiggenthal for the West-Alps). Visualisations of precipitation sums against peak discharges are for the Alpine regions less useful since n-day precipitation correlations with peak discharges are much less than for the rain-fed rivers. This probably to do with the snowmelt influences in the Alpine regions. An exact indicator for the snowmelt volumes should be meaningful to give an exact insight in the melting processes. However, these results are not published in the HBV-output interface. For that reason an exact ratio between precipitation and meltwater to each discharge peak cannot be determined. Since temperature seems to play an important role, the 5-day average temperature prior to the peak discharge is plotted against the peak discharges in figure 40A. 5-days are chosen because these show the largest correlations with the peak discharges. Figure 40A shows the 5-day average temperature prior to the discharge peaks for those who are modelled in May and June because of the high EraClim frequencies in these months. Because higher snow storage is expected for EraClim than GLOFAS, these two time series are compared in this case. Taking into account the curves in figure 40A EraClim shows higher average temperature prior to the peak discharge in May and June than GLOFAS. This means more meltwater will be available and forces the river system to produce higher extreme discharges than GLOFAS. This might partly explain the higher EraClim extremes captured in the extreme discharge distribution (figure 36), especially the part for return periods up to 10 years.

In contradiction to the rain-fed rivers, GLOFAS shows some higher extreme discharge distributions than EraClim in the Alpine basins (see also appendix G). This can be explained by the precipitation sums in figure 40B. Compared to the HYRAS observations, the GLOFAS precipitation prior to the peak discharges are always slightly higher. This is also visible in the extreme discharge distribution in figure 36 and is also applicable for the western Alps. When comparing EraClim to the HYRAS reference, EraClim shows less precipitation. However this gap is expected to be filled by more snowmelt for EraClim than HYRAS.

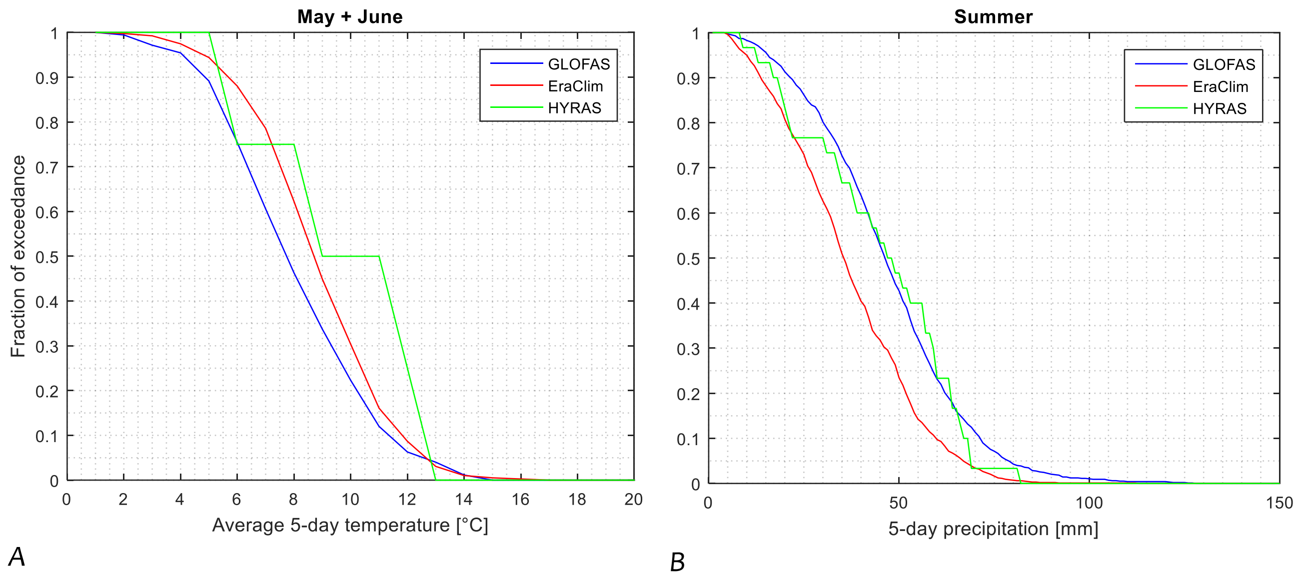


Figure 40 - A: temperature duration curve for EraClim peak months May and June. B: precipitation duration curve summer months (curves for May and June show a similar pattern).

For the Alpine regions it turned out to be difficult to find a complete, solid and firm quantitative explanation for the differences in extreme discharge distributions for different time series. Precipitation and snowmelt can both contribute to the river peak discharges. To draft a convincing analysis, information about melting volumes and their contributions to the peak discharges is required. Based on the average temperature distribution and the fraction of frost days during the year, it can be concluded that both EraClim as GLOFAS have much more snow storage in winter compared to the HYRAS reference. This makes the snowmelt influence during spring much more and affects the reliability of the extreme discharge plots at the Alpine basins. Moreover the large number of summer peaks in the Alpine areas and the contribution of meltwater, influences the Gumbel distribution at Lobith too.

This last reason makes it interesting for Lobith to see how different summer and winter Gumbel distributions are at Lobith. As discussed in the 'Lobith section', both GLOFAS and EraClim show much more peaks in summer months than observed and modelled by HYRAS. After this alpine analysis, the Alpine snowmelt seems to contribute considerably more to these Lobith summer peaks because the other rain-fed rivers do not show many summer peaks. When plotting summer and winter based Gumbel distributions at Lobith, differences are clearly visible in figure 41. Obviously summer peaks are lower than extreme winter discharges, but the significant high number of summer peaks will certainly have a reducing effect on the resulting Gumbel distribution for both GLOFAS as EraClim. For the HYRAS reference also summer peaks are obtained. Since the fraction HYRAS summer peaks is much lower than for GLOFAS and EraClim, the impact is considered less. Comparing the EraClim winter discharges to the HYRAS winter reference, EraClim performs quite similar. When the Alpine meltwater regime for GLOFAS and EraClim was more corresponding to the HYRAS reference, the final GLOFAS and EraClim extreme discharge distributions (figure 27) are expected to be somewhat higher than in the current case.

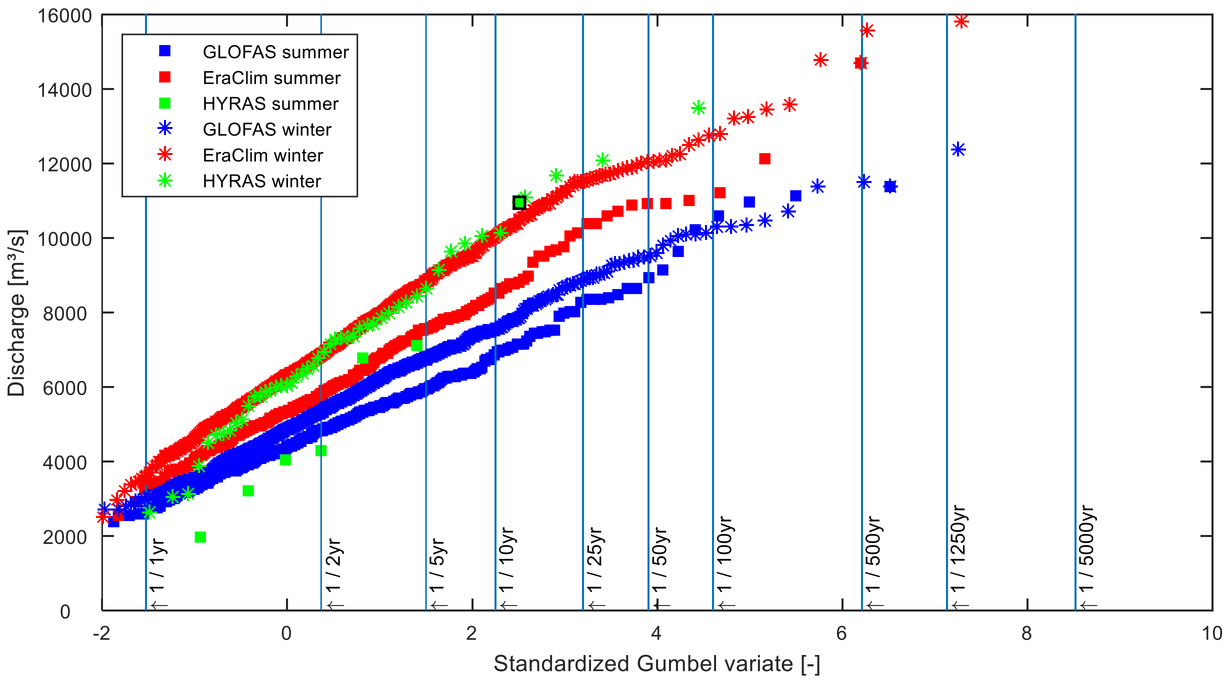


Figure 41 – Gumbel distribution at Lobith for GLOFAS and EraClim divided in summer and winter based annual maxima. Observed discharges are not plotted; these show more less a similar pattern than the HYRAS peaks. Also the fraction summer peaks in the observations is comparable.

Summary

Because not all graphs of all sub basins are shown and discussed in the previous section, a quantitative summary is provided in table 3. The following aspects are assessed in this table for both EraClim and GLOFAS

- Comparison extreme discharge distribution with the HYRAS reference
- Comparison of monthly discharge peak frequencies (in which months annual discharge peaks occur) with respect to the HYRAS reference
- Comparison of 10-day precipitation sums prior to the peak discharge with respect to the HYRAS reference

Table 3 indicates the rate of overestimation (++) or underestimation (- -) of a product with respect to the HYRAS reference. Thereby the colours indicate whether an over- or underestimation (red_ or a more less equal (green) result is obtained. When the over- or underestimation is just very small a yellow indicator is used.

Table 3– Summary of extreme discharge-weather analysis for different sub basins.

	GLOFAS			**EraClim**		
Distribution	Extreme discharge	Monthly peak discharge frequency	10-day precipi-tation	Extreme discharge	Monthly peak discharge frequency	10-day precipi-tation
Lobith	--	-	--	0	-	0
Cochem	--	0	--	+	0	-
Frankfurt	-	0	-	+	0	++
Rockenau	0-	0	-	+	0	-
Rekingen	0+	-	+*	+	-	-*
Untersiggenthal	0	0	-*	-	-	--*

** Alpine basin; meltwater is expected to play an important role. Precipitation distribution is not an encompassing indicator. Only used because of absence of melting water data.*

According to table 3 GLOFAS shows more extreme deviations with respect to the HYRAS reference than EraClim. The extreme discharge distribution of GLOFAS is underestimated for the rain-fed rivers and Lobith. This is explained by the lower 10-day precipitation sums in these basins prior to the peak discharge. Less precipitation leads to lower forcing of the system which reduces the extreme GLOFAS discharges. The GLOFAS extreme discharge distribution in the Alpine areas seems much better than the GLOFAS rain-fed rivers. However, it must be taken into account that snow starts to melt later compared to the HYRAS reference. This also affects the GLOFAS monthly peak discharge frequency. The 10 day GLOFAS precipitation is slightly underestimated for the Alpine regions. However, this indicator might not be appropriate since other processes such as snowmelt also play an important role. This is also applicable for EraClim since much less precipitation prior to the peak discharge is obtained while having higher extreme discharge distributions than the HYRAS reference.

In rain-fed rivers the extreme discharge distributions of EraClim show less extreme deviations than GLOFAS with respect to the HYRAS reference and the observed annual maxima. The 10-day precipitation prior to the peaks is much closer to the HYRAS reference when using EraClim. This partly explains the better performance of EraClim compared to GLOFAS with respect to the HYRAS reference extreme discharges. However on a sub-basin scale the precipitation sums prior to the peak discharge shows less convincing patterns than using the whole river Rhine basin for Lobith. Just as for GLOFAS, meltwater is expected to contribute more to the height of extreme discharges for the Alpine basins than using the HYRAS reference dataset. This because of the underestimation of precipitation sums prior to the peak discharges with respect to the HYRAS reference.

Based on table 3, EraClim seems to be a more suitable product than GLOFAS. This because of less extreme deviations in extreme discharge distributions and precipitation sums prior to the peak discharge with respect to the HYRAS reference. Due to lower temperatures the Alpine regions are expected to have more snow storage than referenced for both GLOFAS and EraClim. Therefore it is recommended to review the temperature generation produced by the ECMWF weather models, since this is direct input for HBV. For the GLOFAS precipitation the same is applicable. Extreme precipitation sums prior to the peak discharge are less than the HYRAS reference. The reasons for these differences should be further investigated.

5. Discussion

This chapter concerns a discussion regarding the research as conducted in this thesis. In section 5.1 the input data are discussed and in 5.2 and 5.3 the used methodology and the results respectively. Section 5.4 discusses the further applicability of this research.

5.1 Data

Daily discharge observations at Lobith are available from 1901 up to now. Annual maxima are selected based on these data, leading to a large uncertainty at higher return periods due to the low number of maxima for Lobith. Furthermore the reliability of the observed river discharges at Lobith can be doubted. Homogenisation corrections were already in the observed discharge data, however the degree in which homogenisations are implemented in this series is not fully certain. Result is different hydraulic conditions in the river system leading to different peak discharges at Lobith. With that more uncertainty is added to the results.

The input parameters for EraClim; the average sea surface temperature and the fraction of sea ice can cause large uncertainties. Just a small deviation in the deterministic re-analysis can cause unrealistic weather. This can be compensated by the ensemble members, but on the other hand it can have an amplifying effect causing even more unrealistic weather.

For GLOFAS there is a data availability of only 12 years during the period 2003 - 2015. The full length of these 12 years is used in the extreme discharge analysis. However the weather behaviour during these 12 years is relevant with respect to the long-term observed average. When those 12 years are much drier than the long-term average, also lower discharge distributions are expected. The GLOFAS average monthly discharge is higher than the HYRAS reference, so it can be stated that using weather from the period 2003 – 2015 is not representative. Results can be different when using longer time series. Also shorter time series than 'just' all available 12 years can be taken into account when these shorter period matches the long-term observed weather or discharge more appropriate. It is highly uncertain if the period 2003 – 2015 is representative enough for constructing GLOFAS time series with well-considered characteristics. The optimal data-period is not assessed in this research, but for further research this can certainly be recommended.

5.2 Method

For the synthetic GLOFAS series weather segments are linked in order to construct a weather series of 1090 years. Each segment is 4 days long, which is randomly chosen. This means that the synthetic GLOFAS weather series contains after 4 days a discontinuity in their weather series. To reduce this number of discontinuities another segment length of 5 days can be chosen. That is also the maximum segment length from each ensemble member since the first 5 and last 5 days are cut of the 15 day member length. However, the most appropriate segment length should be further investigated, and with them the discontinuities in the synthetic GLOFAS weather series. Since the construction process takes long this is not investigated in this research, but the method to investigate the quality of each weather series should not be that difficult. Synthetic GLOFAS weather series can be constructed from different segment lengths (1 day up to 5 days). The autocorrelation with a time lag of 1 day can be used to quantify the quality compared to the observed HYRAS weather series.

Another point can be associated with the total number of discontinuities between GLOFAS and EraClim. The synthetic GLOFAS series have every 4 days a discontinuity in their series, while EraClim has only 10 over a period of 1090 year. For GLOFAS this might cause weather patterns which is combined from totally contradictory weather. No attention is paid to a smooth transition to keep a high level of consistency in weather. In fact, this makes it possible to build a relative dry and cold august, with only one segment (of 4 days) with extreme high temperatures and precipitations for example. This weather transition is not reliable at all, and therefore the corresponding discharge neither.

5.3 Results

The extreme value distributions at Lobith seems to fit quite well for EraClim when only reviewing the absolute peak discharge. GLOFAS peaks too much during summer compared to the HYRAS reference, possibly caused by the meltwater contribution from the Alpine. However, to show this in a more convincing way a detailed analysis is required. Lag times such as defined by Torfs (2004) should be used to estimate the travel time from a certain discharge station to Lobith. Furthermore the flattening of a discharge peak should be considered. Both processes are far from trivial, but crucial to see the exact contribution of regional water volumes to the discharge peaks at Lobith.

During the result analysis prior to the peak discharges n-day precipitation sums are used. This turned out to be a good indicator to explain differences in extreme discharge distributions for Lobith when comparing GLOFAS and EraClim to the HYRAS reference. However, on a sub-basin scale, this indicator is not always very convincing. Where Lobith peak discharges (HYRAS reference) turned out to be most correlated with 10-day precipitation in advance, the peak discharges in sub-basins are best correlated between 3 and 5 days because of their smaller size. On sub-basin scale it turned out that these precipitation sums prior to the peak discharge do not explain the differences in sub basin extreme discharge distributions. To take into account saturation processes, 10-day precipitation is used to analyse differences between ensemble products and HYRAS reference in sub-basins. Since this analysis did not provide appropriate explanations for the differences in extreme discharge distributions, apparently other forcing factors play also a role in the development of peak discharges. Therefore more research is required in order to analyse which model variables are contributing to extreme discharges and in which proportion.

The first research question goes amongst others into the daily statistics for the GLOFAS, EraClim and the observed discharges. When comparing the flow duration curves for EraClim and GLOFAS to the observed discharges, GLOFAS shows a better correspondence. When assessing the extreme discharge statistics the view is totally the other way around. There EraClim shows much more similarities at Lobith than GLOFAS. Of course the data periods are different, but it is at least remarkable for GLOFAS that differences compared to the observed discharges are fluctuating that much when assessing daily or extreme discharges. This was not expected at the start of this research.

EraClim weather is basically produced by forcing a weather model by the average sea surface temperature and the fraction of sea ice. Since the sea ice fraction is decreasing and the average temperature rising due to climate change (KNMI, 2016) their might be some strange patterns in the used EraClim series. All 10 ensemble members are put simply in a subsequent order, but that means that the period 1901-2010 is put in a subsequent order too. Because the fraction sea ice was much more in 1901 than in 2010 and the average temperature was lower in 1901 than in 2010, strange weather jumps due to this climate change can be visible when two EraClim ensemble members are put in subsequent order.

For Lobith diagrams are plotted how design discharges are related to a certain return period for different datasets. The extrapolation uncertainty is taken into account based on an existing empirical method of Shaw et al. (2005). In fact the standard deviation of all annual maxima is divided by the root of the number of annual maxima. Subsequently this is multiplied with a variable depending on the return period. Since EraClim and GLOFAS consist of 1090 maxima, it may be doubted if it is fair to use the standard deviation of the whole dataset in order to define the extrapolation uncertainty. The number of annual maxima is much different, so a real fair uncertainty comparison cannot be done in this way. Possibly it is better to use a more local standard deviation based on a number of annual maxima around the return period of interest. The standard deviation of the x number of nearest neighbours can be used for example. However it needs more research to see if such an approach works properly. Another option is to search for another existing extrapolation uncertainty method and compare it to the method of Shaw (2005). Next to extrapolation uncertainties no other uncertainties are taken into account in this research because that was not in the scope of this research. Furthermore, it is very time consuming when assessing all sources of uncertainty in the extreme discharge distribution. However, it can be recommended to do so because a more fair comparison can be made between short observation records and long ensemble based datasets.

5.4 Applicability research outcomes

The general applicability of numerical weather models can be discussed. First of all artificial weather series of GLOFAS and EraClim are not in full correspondence with the sub-basin observed weather. In snowy mountain regions there is more frost than observed for both EraClim and GLOFAS and in rain-fed sub basins there is less extreme precipitation using GLOFAS. Also the system tends to a more wet state using EraClim or GLOFAS. Therefore additional research is required to find out why the weather model is producing this kind of results.

This research focusses on the extreme hydrological behaviour in the river Rhine basin. This makes the conclusions not worldwide applicable for ensemble products like GLOFAS or EraClim. Simply because the climatological circumstances cannot be assumed equal. Similar kind of assessments for both EraClim and GLOFAS are proposed in other climatological zones. However it can be proposed to treat final results with lots of care, just because it seems to be hard to model extreme discharge distributions in individual sub-basins appropriate enough with respect to the HYRAS reference case.

The EraClim product is aimed to improve the observational weather records from 1901 to 2010 containing precipitation and temperature (ECMWF, 2015). Even as for the GLOFAS product, EraClim shows much lower temperatures than observed. During some months this underestimation goes up to 5 degrees centigrade (basin averaged). It not know why this is happening, but this means that the EraClim weather assimilation should be researched and improved in this field.

6. Conclusions and recommendations

6.1 Conclusions

This research focussed on the estimation of extreme discharge distributions making use of products provided by numerical weather models used for operational flow forecasting. Hereto two numerical weather models are closer reviewed; GLOFAS and EraClim. GLOFAS creates 51 long weather ensemble forecasts every day for the next 15 days. This happens by perturbing the deterministic weather forecast. These GLOFAS data only cover the period 2003 – 2015. EraClim is available during the period 1901 – 2010 and contains 10 ensemble members covering a 109 year long period each. In this research the meteorological and discharge statistics of both weather products were compared. Subsequently a closer look was given to the construction of synthetic EraClim and in particular to GLOFAS weather series. Finally extreme discharge distributions are compared for both synthetic GLOFAS and EraClim series with respect to the HYRAS reference in the river Rhine basin. The objective was to investigate to which extent products provided by numerical weather and hydrological models for operational flow forecasting can be used for estimating high discharge events in rivers having relatively short return periods (< 50 year) and how these estimates compare with the estimates derived from classical methods.

Which product seems in first instance most promising for the assessment of design discharges for short (< 50 year) return periods using Ensemble Streamflow Predictions?

The overlapping period between 2003 – 2006 is compared for daily GLOFAS, EraClim and reference HYRAS precipitation at a basin-average scale. A boxplot precipitation analysis is performed from which it can be concluded that each of the GLOFAS and EraClim ensemble members produce overall about the same range of precipitation as for the reference case with measured HYRAS precipitation. Only difference is seen for GLOFAS which shows higher extreme precipitations than the HYRAS reference and EraClim. EraClim show slightly higher inter quartile ranges, so some more precipitation is expected there. For this reasons the monthly precipitation sums for both GLOFAS and EraClim are higher compared to the HYRAS reference. GLOFAS and EraClim show much less dry days compared to the HYRAS reference. HBV is used to convert the weather input into discharges along all river Rhine branches. The discharge boxplots for Lobith show significant higher discharges than observed (2003 – 2010) during May, June and July for both EraClim and GLOFAS. The flow duration curves (FDC's) for daily discharge are higher than observed for almost all GLOFAS ensemble members, just a few ensemble members show lower FDC's. The EraClim ensemble members show even higher discharges than the GLOFAS ensemble members. Especially EraClim discharges in the middle range (3000 – 6000 m^3/s) show large deviations from the observed discharge. Respecting this first rough analysis GLOFAS daily discharges are more similar to the observations than these of EraClim. Later on is found that the view for extreme discharge distributions is totally the other way around.

Based on EraClim and GLOFAS weather ensemble members, in which way synthetic weather series can be composed which are long and independent enough to estimate the flood design level at a certain return period?

Research is performed in order to see if (infinite) long time series may be constructed from products driven by numerical weather models. Because both EraClim as GLOFAS contain weather ensemble members induced by perturbations of the deterministic weather model, the independency of these mutual ensemble members is researched. For EraClim this results in negligible Pearson correlation coefficients for mutual ensemble members. For this reason EraClim series might be constructed by putting each of the 109 year long EraClim ensemble members in a subsequent order to get 1090 years of independent weather. The days 6 to 10 of an ensemble member are used to construct chronological weather series for all of the 51 ensemble members. These 51 chronological weather ensemble members are mutual tested for correlation. The average mutual Pearson correlation between all GLOFAS ensemble members is 0,16, which makes the GLOFAS ensemble members practically independent.

For the GLOFAS synthetic series with the largest degree of randomness, segment lengths of 4 days are used. This is a random choice since more research is required to the transitions between two GLOFAS segments. These 4-day long segments are put in a subsequent order. Hereby a random ensemble member is picked and a random period from the same month of interest. Finally statistics concerning synthetic GLOFAS series are compared to the chronological GLOFAS series. Daily precipitation, temperature and discharge turned out to be about the same throughout the year. For the extreme discharge statistics it is seen that discharge peaks occur roughly in the same months when using chronological or synthetic GLOFAS series. Also extreme precipitation prior to the peak and the extreme discharge distributions show much correspondence. Therefore synthetically series can be used in order to construct infinite long GLOFAS series.

To what extent do extreme discharge distributions differ at Lobith when they are obtained by using numerical weather products compared to observed weather?

The extreme discharge performance for GLOFAS and EraClim is qualitatively assessed by comparing the Gumbel fitted discharge distributions to the HYRAS reference. To provide an equitable comparison not only extreme discharge distributions are assessed at Lobith, but also the extrapolation uncertainty bands. Note that extrapolation uncertainty does not cover the total uncertainty. For small return periods at Lobith ($T \leq 50$ year) it turned out that for all assessed return periods extrapolation confidence of GLOFAS is below the extrapolation confidence intervals of both the observations as the HYRAS reference. This means that GLOFAS is not a suitable replacement for the conventional estimation of extreme discharges at Lobith. Initially EraClim shows a well corresponding extreme discharge distribution with both the HYRAS reference and the observations. This can be explained by the 10-day summed precipitation prior to the peak discharge which is corresponding well for EraClim to the HYRAS reference. For GLOFAS this correspondence is way less which results in an underestimation of the extreme discharge distribution at Lobith.

The EraClim extreme discharge distribution at Lobith fits well for the unsatisfactory reasons because 26% of all discharge peaks is found during summer while the HYRAS reference shows only 13% in summer. For GLOFAS the same is applicable with even more summer peaks (32%). Therefore the behaviour in all upstream sub-basins at the external border of the river Rhine system are analysed, because these basins do not have discharge input from upstream. For the rain-fed rivers (Moselle, Neckar, Main) it turned out that these do not show significant more summer peaks for GLOFAS and EraClim than the HYRAS reference. This means that rain-fed rivers are not expected to contribute in an extreme way to the difference in obtained summer peaks at Lobith.

Concerning the extreme discharge distributions in the rain-fed basins GLOFAS underestimates the HYRAS reference. On the other hand EraClim shows slightly higher extreme discharge distributions. Precipitation duration curves of the precipitation during 10 days prior to the peak discharge did not fully explain the differences in extreme discharge distributions on a sub-basin scale. Sometimes the differences between EraClim and the HYRAS reference are large while the differences in extreme discharge distributions are small.

In the Alpine areas temperature is an important factor. GLOFAS and EraClim temperature show more frost days in the winter period compared to the HYRAS reference. This results in more snow storage volumes during winter. In the first summer months this extra volume is expected to be added to the river system. This is clearly visible in the months May and June for GLOFAS and EraClim when the Alps are producing the most discharge peaks. This is also the period where Lobith shows more peak discharges for EraClim and GLOFAS than referenced by HYRAS. Therefore the Alps are pointed to be the cause of the summer peaks at Lobith. Because the large number of summer peaks for GLOFAS and EraClim are relatively lower than the average winter peaks, a reducing effect on the final discharge distributions at Lobith cannot be avoided. This affects the reliability of the Lobith plots significantly.

Concerning the extreme discharge distributions with respect to the HYRAS reference, the western Alps show a slightly higher extreme discharge distribution for GLOFAS and a slightly lower distribution for EraClim. In the eastern Alps GLOFAS shows also higher discharges than referenced for larger return periods. EraClim is slightly higher than the HYRAS reference, but lower than GLOFAS for the larger return periods. The combination snowmelt and (extreme) precipitation plays a role here. EraClim is expected to contribute more snowmelt than GLOFAS, but GLOFAS shows in the Alpine areas more extreme precipitation than EraClim prior to the peaks. This makes higher GLOFAS extreme discharges possible relative to EraClim.

All in all extreme discharge distributions of GLOFAS and EraClim differ in upstream sub-basins too much from the modelled extreme discharge based on weather observations. This means that both EraClim as GLOFAS cannot be recommended for extreme discharge analysis when they are used in the current way.

6.2 Recommendations

Based on the discussion and previous conclusions some recommendations can be to support further research.

First of all it can be recommended to assess the quality of transitions between two linked weather segments when constructing GLOFAS series. This research links random 4-day long segments in a subsequent order without checking the quality of the transitions. However, such an analysis is important since synthetic weather series have to reflect the same weather consistency than the weather observations. Furthermore it is important to look further to the best length of the segments which are used for the synthetic GLOFAS series.

GLOFAS shows lower extreme discharge distributions in the River Rhine basin compared to the HYRAS reference. Explanation lies partly in the lower extreme precipitation amounts prior to the peak discharge. This is interesting because these precipitation data are generated by ECMWF weather models. Therefore it should be assessed how extreme precipitation are formed in that model. This is also relevant for the GLOFAS and EraClim temperatures, since much more frost is obtained using these numerical weather products.

Another recommendation concerns the geographical location of the research area. The river Rhine is used because of its long set of discharge observations and measured weather (HYRAS). However, since GLOFAS and EraClim data are available worldwide other river basins (in other climate zone) should be assessed too in order to see if the same conclusions remain still valid.

Strictly it could not be concluded that GLOFAS or EraClim is a perfect alternative to estimate design discharges for Lobith. This does not mean that there is no potential at all. Companies and water management authorities could use both methods when extreme little observations are available. In case only five years observations are available, it is probably more accurate to use GLOFAS or EraClim in order to make an extreme discharge distribution. At the other hand these results should be handled with lots of care since some river Rhine sub-catchments also show very unreliable extreme discharge distributions for GLOFAS and EraClim. This could be the case when assessing another river basin.

7. References

Alfieri, L., Burek, P., Dutra, E., Krzeminski, B., Muraro, D., Thielen, J., & Pappenberger, F. (2013). Global ensemble streamflow forecasting and flood early warning. *Hydrology and Earth System Sciences*(17), 1161-1175.

BfG. (2010). *Global Runoff Data Centre, Koblenz*. Retrieved from Discharge data Rhine: http://www.bafg.de/GRDC

CLIMAS. (2016, September 25). *CLIMAS*. Retrieved from http://www.climas.arizona.edu/sw-climate/temperature-and-precipitation

Cloke, H. L., & Pappenberger, F. (2009). Ensemble flood forecasting: A review. *Journal of Hydrology*(735), 613-626.

de Moel, H., Aerts, J., & Koomen, E. (2011). Development of flood exposure in the Netherlands during the 20th century. *Global Environmental Change*(21), 620-627.

Demeritt, D., Cloke, H., Pappenberger, F., Thielen, J., Bartholmes, J., & Ramos, M. (2007). Ensemble predictions and perceptions of risk, uncertainty, and error in flood forecasting. *Environmental Hazards*(7), 115-127.

Demirel, M., Booij, M. J., & Hoekstra, A. Y. (2013). Identintification 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), 2742-2758.

ECMWF. (2015). *Era-Clim project website*. Retrieved from http://www.era-clim.eu/

ECMWF. (2016). *Medium-range forecasts*. Retrieved from ECMWF.int: http://www.ecmwf.int/en/forecasts/documentation-and-support/medium-range-forecasts

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

Garba, H., Ismail, A., & Tsoho, U. (2013). Fitting Probability Distribution Functions To Discharge Variability Of Kaduna River. *International Journal of Modern Engineering Research*(3), 2848-2852.

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

Hogan, C. M. (2016). *Encyclopedia of the Earth*. Retrieved from Encyclopedia of the Earth: http://www.eoearth.org/view/article/154587/

IGBP. (2016). *Future Earth*. Retrieved from http://www.futureearth.org/blog/2014-jun-26/deltas-risk

KNMI. (2014). *KNMI nieuws*. Retrieved 2016, from KNMI: https://knmi.nl/over-het-knmi/nieuws/2003-nederlands-zonrecord-warm-en-droog

KNMI. (2016). *Zeeijsbedekking in het noordpoolgebied*. Retrieved from KNMI kennis- en datacentrum: https://www.knmi.nl/kennis-en-datacentrum/achtergrond/zeeijsbedekking-in-het-noordpoolgebied-neemt-gestaag-af

KNMI. (2016). *Zware neerslag*. Retrieved from KNMI.nl: https://www.knmi.nl/kennis-en-datacentrum/uitleg/zware-neerslag

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

MeteoGroup. (2016). Retrieved from http://www.weer.nl/nieuws/detail/2009-09-12-droge-regen-of-natte-droogte/

Middelkoop, H., & Hasselen, C. (1999). *Twice a River, Rhine and Meuse in the Netharlands.* Arnhem: RIZA.

NOAA. (2006). *Ensemble Predictions Systems*. Retrieved from National Oceanic and Atmospheric Administration: http://www.wpc.ncep.noaa.gov/ensembletraining/

NOAA. (2015). *National Oceanic and Atmospheric Administration*. Retrieved from National Oceanic and Atmospheric Administration: http://www.erh.noaa.gov/mmefs/Help/EnsembleOverview/img0.html

Pappenberger. (2015). After 10 days the ensembles drift back to climate; mail conversation F.C. Sperna-Weiland and F. Pappenberger.

Poli, P., Hersbach, H., Tan, D., Dee, D., Thépaut, J. N., Simmons, A., . . . Fisher, M. (2015). *ERA report series: 14, the data assimilation system and initial performance evaluation of the ECMWF pilot reanalysis of the 20th-century assimilating surface observations only (ERA-20C).* Reading: ECMWF.

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

Rigimbeau, F. R., Habets, F., Martin, E., & Noilhan, J. (2007). Ensemble Streamflow forecasts over France. *ECMWF Newsletter 111*, 21-27.

Shaw, E. M. (2005). *Hydrology in Practice.* London.

Tockner, K., Uehlinger, U., & Robinson, C. (1999). *Rivers of Europe. .* Oxford: Accedemic Press, Elsevier.

Torfs, P. (2004). Black-box forecasting and uncertainty. Wageningen University.

University, Tilburg. (2016). *Colleges correlatie*. Retrieved from SPSS helpdesk: https://www.tilburguniversity.edu/nl/studenten/studie/colleges/spsshelpdesk/edesk/correlat/

Van den Brink, H. W., & Können, G. P. (2011). Estimating 10000-year return values from short time series. *International Journal of Climatology*, 115-126.

Vogel, R. M. (1956). The Probability Plot Correlation Coefficient Test for the Normal, Lognormal, and Gumbel Distributional Hypotheses. *Water resources research, 22*(4), 587-590.

Willems, P., Sonbol, M., Abdo, G., Mkhandi, S., Opere, A., Tadesse, L., . . . Al-Weshah, R. (2005). Regional flood frequency analysis in the Nile basin. Proceedings of the International Conference of UNESCO Flanders FIT FRIEND/Nile Project – “Towards a Better Cooperation”, Sharm El-Sheikh, Egypt: UNESCO.

A. Appendix A – HYRAS or EOBS reference

In chapter 3 the use of HYRAS weather input is preferred above the use of EOBS weather input for comparing HBV discharge results of EraClim and GLOFAS with. This choice is based on three assessors which are summed below

- Nash Sutcliffe (NS)
- Relative volume error (RVE)
- Qualitative consideration grid size.

Table 4 – Comparison EOBS and HYRAS relative to the real observed Lobith observations.

	NS	RVE
HYRAS 1951 – 2006	0,86	1,93%
EOBS 1951 – 2006	0,87	- 9.02%

As can be seen in table 4 the differences in NS-coefficient are not that large. However, the relative volume error for EOBS is much larger than for HYRAS. On that basis the daily discharge is better described by HYRAS.

A more qualitative assessment is based on the grid sizes of the two datasets. EOBS is available between 1951 and 2014 with a relative course horizontal resolution of 25 x 25 km (Kjellström et al., 2015). HYRAS on the other hand, has a more fine horizontal resolution of 5 km but is only available between 1950 and 2006 (Frick et al., 2014). In addition Rauthe et al. (2013) found that the (calibrated) HYRAS precipitation describes the observations much better than EOBS. They show a probability density function of the precipitation distribution at the measure stations. Herein EOBS overestimates precipitations between 0 and 1 mm and underestimates precipitations between 1 and 10 mm. To provide an accurate reference, the HYRAS dataset is taken therefore. This is also of importance when a closer view is given to some of the river Rhine sub basins. Accurate precipitation amounts become more and more important then.

B. Appendix B – Streamflow dependency

In section 3.2 the methodology describes how to test on mutual dependency on basis of precipitation ensemble members. Thereafter the discharge series comes into view. A correlation analysis on these series is expected to be useless because high correlations between streamflow ensemble members are assumed to be plausible. Covariance can be explained as the degree of two subsequent elements in two time series which are varying in the same direction (does the discharge increase or decrease for the next day in both series). When both series do so, a high covariance is found. This covariance is normalized to divide it by the product of both time series variances. GLOFAS streamflow hydrographs shown tend to have the same shape (mention: definitely different discharge values!). The assumption that these stream flows are highly correlated seems valid therefore. However, the correlation does not tell anything about the absolute differences between two streamflow ensemble members. When figure 42 is used as interpretation, it can be seen that those figures have more less the same shape. Covariance and thus correlation coefficients would be quite high. As can be seen from this fictive figure, absolute differences are certainly there. The difference between the mark of the red and black line is smaller than between the black and blue one. For that reason the stream flows originating from GLOFAS ensemble members will be tested on differences. A two tailed T-test (95% confidence limit) would be appropriate for that purpose since it concerns especially the significant difference between two normal distributed sets of data. Discharge is probably not normal but log-normal distributed (Bowers et al., 2012). To be able to apply a T-test, the discharges will be transformed to a normal distribution by taking the natural logarithm of each discharge value in the dataset. Every n days the difference is tested between one ensemble members and all others. Figure 43 shows in the red dashed box the red data planes which are taken into account for the first 5-day long T-test between two ensemble members (equation 1). Next the first 5 days of streamflow 1 will be compared to streamflow member 3. This process will continue until streamflow member 51. Subsequently days 2 up to 6 will be taken to perform a similar analysis but then for one day further (green planes). Result will be a large matrix of T-testing values in which every specific ensemble member is compared to all of the other 50 ensemble members. From this matrix a percentage will be derived how many T-tests display a significant difference.

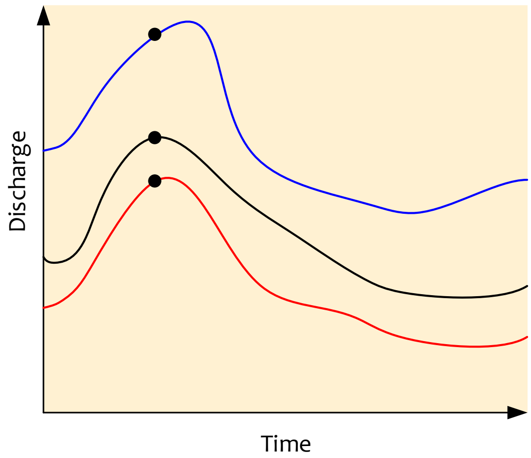


Figure 42 – Fictive discharge waves, with corresponding shapes

Because the discharge at Lobith is constructed by water originating from places which are located hundreds of kilometres upstream, time lag will be of influence. Therefore it would not be appropriate to only take a timespan of n is five days. Longer time intervals, up to 30 days will be assessed too. The T-test might obstruct a significant difference between two datasets when standard deviations are too large. For that reason a Fisher variance test will be performed too, to see if the variances deviate from each other. When they do so, two ensemble members are varying enough to say something useful about their differences. The variance test will be performed on the same intervals of n as the T-tests (equation 2).

$$T = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{(n-1)s_x^2 + (m-1)s_y^2}{n+m-2}} \cdot \sqrt{\frac{1}{n} + \frac{1}{m}}} \qquad (1)$$

$$F = \frac{s_x^2}{s_y^2} \qquad (2)$$

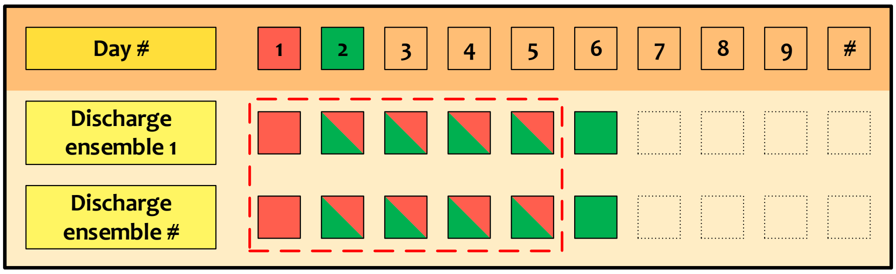


Figure 43 – T and F testing between mutual ensemble members on a n-day interval. As an example a 5-day interval is taken.

C. Appendix C – PPCC, Gumbel fitting and confidence

This appendix is an extension of section 3.5 of the methodology. In here it becomes clear how a PPCC test of Vogel (1986) is performed, how Gumbel lines are fitted to the annual discharge maxima and how the 95% confidence limits of this are determined using the method of Shaw et al. (2005).

PPCC-test

According to Vogel (1986) the correlation between standardized Gumbel variate and peak discharge can be expressed in equation 3. The expression describes the probability plot correlation coefficient (PPCC) r.

$$r = \frac{\operatorname{cov}\left(\ln[-\ln(Q_i)]\,,\ \ln\left[-\ln\left[\frac{i-0{,}44}{n+0{,}12}\right]\right]\right)}{\sqrt{\operatorname{Var}(\ln[-\ln(Q_i)])\cdot \operatorname{Var}(\ln[-\ln(Q_i)])}} \qquad (3)$$

For the different datasets from table 5 there is a different number of annual maxima which corresponds with a critical PPCC. When this critical PPCC is non-exceeded a Gumbel fit to the annual maxima of the dataset does not seem as an appropriate option.

Table 5 – Number of annual maxima and their critical correlations for each dataset. The asterisk indicates an interpolation of the simulation results of Vogel (1986).

Data	Annual maxima (N)	Critical correlations (95% C.L.)
HYRAS	55	0,940954*
Observations Lobith	110	0,979698*
EraClim	1090	0,996550*
GLOFAS Synthetic	1090	0,996550*
GRADE	5000	0,999030

* According to interpolation using simulation results of Vogel (1986)

Gumbel fit

When appears that all datasets are suitable for Gumbel fitting, the parameters alpha and beta corresponding to equation 5 and 6 are determined (Vogel, 1986; Shaw et al., 2005). In these parameters mu equals the average peak discharge of the observed annual series and sigma equals the standard deviation with respect to this average. Gamma is Euler's constant with a value of 0,5772.

$$F(X \le x) = \exp\left[-\exp\left\{-\frac{x-\alpha}{\beta}\right\}\right] \qquad (4)$$

$$\alpha = \mu - \gamma\beta \qquad (5)$$

$$\beta = \sigma\frac{\sqrt{6}}{\pi} \qquad (6)$$

Confidence interval extrapolation Gumbel fit

After fitting representative lines through the data points, it is important to investigate how the 95% confidence limits are situated. These are expected to be wider when using less data points and smaller for the EraClim and GLOFAS set which uses 1090 annual maxima. For the confidence limits the approach of Shaw et al. (2005) will be followed. The extrapolation confidence limits for the fitted data are described by equations 7 to 9.

$$SE(\hat{Q}) = \frac{S_Q}{\sqrt{N}} \cdot \sqrt{1 + 1{,}14\,K(T) + 1{,}10(K(T)^2)} \tag{7}$$

$$K(T) = \frac{\sqrt{6}}{\pi} \cdot \left(\gamma + \ln\ln\left[\frac{T(Q)}{T(Q) - 1}\right]\right) \tag{8}$$

$$Confidence\ extrapolation(\hat{Q})_{95\%} = \hat{Q} \pm 1{,}96 \cdot SE(\hat{Q}) \tag{9}$$

D. Appendix D – Basic analysis daily discharges and precipitation

Section 4.1 describes the results of both the precipitation as the discharge analysis for two different input products: EraClim and GLOFAS. This appendix describes the same analysis, but now performed on a more course scale; seasonal and annual with respect to the discharge. As mentioned in section 3.4 all precipitation data over the period 2003 up to 2010 is used for every sub catchment of the river Rhine in this analysis.

Precipitation analysis

In addition to the boxplots in section 4.1 the statistics are also shown for both EraClim and GLOFAS precipitation during 2003 – 2006. Table 6 shows the percentile statistics of each EraClim ensemble member.

Table 6 – Boxplot characteristics of ensemble precipitation using EraClim (area weighted average from each sub basin). (IQR = inter quartile range). In the last columns the average precipitation of a certain ensemble member is shown even as the standard deviation and the maximum 10-day precipitation sum.

$Group$	Q_1	Q_2	Q_3	$1,5 \cdot IQR$	P_{avg}	P_{std}	$P_{10:max}$
1	0,37	1,55	4,30	5,90	2,93	4,38	86,03
2	0,43	1,69	4,43	6,00	2,99	4,42	89,39
3	0,23	1,37	3,77	5,32	2,66	4,36	88,02
4	0,33	1,39	4,05	5,58	2,76	4,23	90,48
5	0,32	1,40	3,91	5,39	2,70	4,23	84,39
6	0,21	1,18	4,07	5,80	2,69	4,26	101,80
7	0,22	1,26	4,03	5,71	2,69	4,31	90,45
8	0,20	1,23	3,85	5,47	2,66	4,34	98,62
9	0,24	1,18	3,71	5,21	2,63	4,32	83,16
0	0,21	1,22	4,20	5,97	2,74	4,31	96,84
$Average\ ensembles$	1,95	2,66	3,42	2,20	2,75	4,32	43,49
$Observations$	0,06	0,98	3,36	4,96	2,32	4,64	88,30

The 10-day precipitation sum is assessed since this is assumed to be an important factor for large discharge events (Pelt et al., 2015)). The 10-day precipitations are shown in table 7 are more less in the same order of magnitude, when using a two times standard deviation interval for the 10 day long precipitation, each ensemble member maximum is in it.

Also for the GLOFAS precipitation ensemble members is investigated if their maximum 10-day precipitation is different from the observed precipitation in the research period. There is a bit more variation is visible. This will be addressed to the larger number of ensemble members comparing with EraClim.

Table 7 – Boxplot characteristics of ensemble precipitation using GLOFAS (area weighted average from each sub catchment). In the last columns the average precipitation of a certain ensemble member is shown even as the standard deviation and the maximum 10-day precipitation sum.

$Group$	Q_1	Q_2	Q_3	$1,5 \cdot IQR$	P_{avg}	P_{std}	$P_{10:max}$
1	0,15	0,98	3,24	4,63	2,36	4,51	87,98
2	0,17	1,21	3,70	5,29	2,60	4,94	80,31
3	0,21	1,31	4,00	5,68	2,64	4,77	95,43
4	0,18	1,02	3,56	5,08	2,49	4,59	82,00

5	0,18	0,97	3,36	4,77	2,47	4,93	80,83
6	0,18	1,01	3,35	4,75	2,36	4,59	84,37
7	0,17	1,07	3,47	4,96	2,42	4,60	82,48
8	0,16	1,03	3,47	4,96	2,43	4,84	68,53
9	0,24	1,11	3,47	4,86	2,46	4,63	74,88
10	0,21	1,19	3,79	5,38	2,60	4,78	74,57
11	0,17	1,10	3,44	4,91	2,40	4,35	96,86
12	0,15	1,01	3,38	4,83	2,39	4,95	111,49
13	0,17	1,12	3,53	5,04	2,39	4,50	83,49
14	0,11	0,91	3,24	4,70	2,31	4,56	78,59
15	0,20	1,16	3,70	5,24	2,55	4,55	76,13
16	0,18	1,09	3,62	5,16	2,43	4,67	96,57
17	0,17	1,10	3,53	5,04	2,45	4,41	87,98
18	0,18	1,23	3,82	5,46	2,49	4,50	84,44
19	0,18	1,13	3,50	4,98	2,39	4,30	69,04
20	0,17	1,09	3,35	4,77	2,43	4,64	85,78
21	0,21	1,16	3,54	4,99	2,41	4,33	102,90
22	0,21	1,16	3,57	5,04	2,50	4,76	78,92
23	0,17	1,18	3,65	5,22	2,48	4,57	89,64
24	0,17	1,17	3,47	4,94	2,41	4,39	63,62
25	0,16	1,11	3,79	5,44	2,57	4,73	92,20
26	0,17	0,93	3,46	4,94	2,54	5,23	95,56
27	0,18	1,03	3,42	4,86	2,44	4,53	77,85
28	0,15	1,04	3,39	4,87	2,43	4,69	81,88
29	0,16	0,95	3,50	5,01	2,38	4,30	73,11
30	0,17	1,10	3,38	4,82	2,43	4,61	82,29
31	0,23	1,14	3,57	5,01	2,46	4,70	69,81
32	0,18	1,10	3,49	4,96	2,40	4,41	73,33
33	0,18	1,10	3,51	5,00	2,45	4,75	75,85
34	0,18	1,04	3,14	4,45	2,39	4,59	89,66
35	0,16	1,13	3,51	5,02	2,48	4,52	85,81
36	0,20	1,22	3,81	5,41	2,62	4,77	71,90
37	0,21	1,29	3,73	5,29	2,65	4,78	89,17
38	0,17	1,06	3,36	4,78	2,42	4,44	75,00
39	0,21	1,12	3,71	5,25	2,71	5,34	124,88
40	0,20	1,23	3,60	5,11	2,59	4,64	83,59
41	0,18	1,12	3,65	5,21	2,41	4,39	71,91
42	0,18	1,07	3,28	4,65	2,49	4,80	90,04
43	0,20	1,10	3,54	5,01	2,52	4,69	93,41
44	0,20	1,32	3,53	4,99	2,52	4,53	64,97
45	0,17	1,16	3,37	4,81	2,41	4,71	81,20
46	0,19	1,33	3,79	5,40	2,64	5,36	86,88
47	0,18	1,16	3,78	5,41	2,62	4,62	86,80
48	0,17	1,15	3,84	5,51	2,61	4,73	94,55
49	0,17	1,07	3,44	4,91	2,45	4,80	87,06
50	0,20	1,33	3,86	5,50	2,72	5,11	98,57
0	0,26	1,30	3,60	5,02	2,54	4,63	85,13
$Avg.Ens.$	1,32	2,27	3,38	3,09	2,49	4,67	62,37
$Obs.$	0,06	0,98	3,36	4,96	2,32	4,64	88,30

Discharge analysis

Reference set

For the discharge analysis of ensemble members two possible references are available. When discharge observations at Lobith are chosen the period of overlap lies between 2003 and 2010 for EraClim, GLOFAS and observations. Another possibility is to use HYRAS weather. Then the HBV model is required to produce reference discharges, but only the period between 2003 and 2006 is in common. Because of the longer period of overlap the real discharge observations are preferred. The HYRAS based discharges and discharge observations are compared during the period 1951 – 2006 in order to see differences. Figure 44 shows two almost indistinguishable cumulative discharge distributions from both observed discharges and simulated HYRAS discharges. This means that observed discharges can be used just as well as HYRAS. For that reason the period between 2003 and 2010 is used in order to compare GLOFAS and EraClim to the discharge observations.

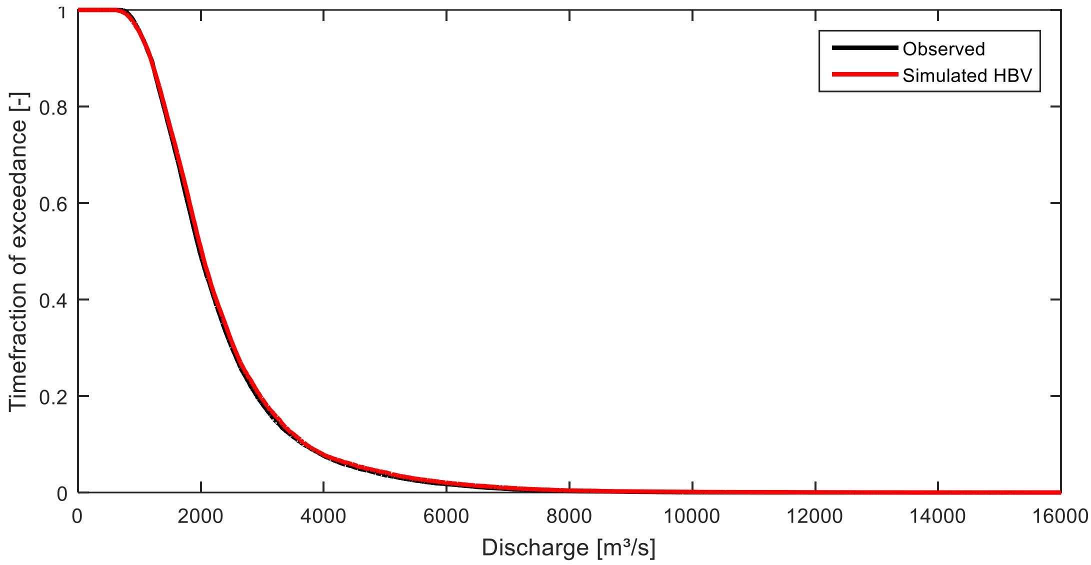
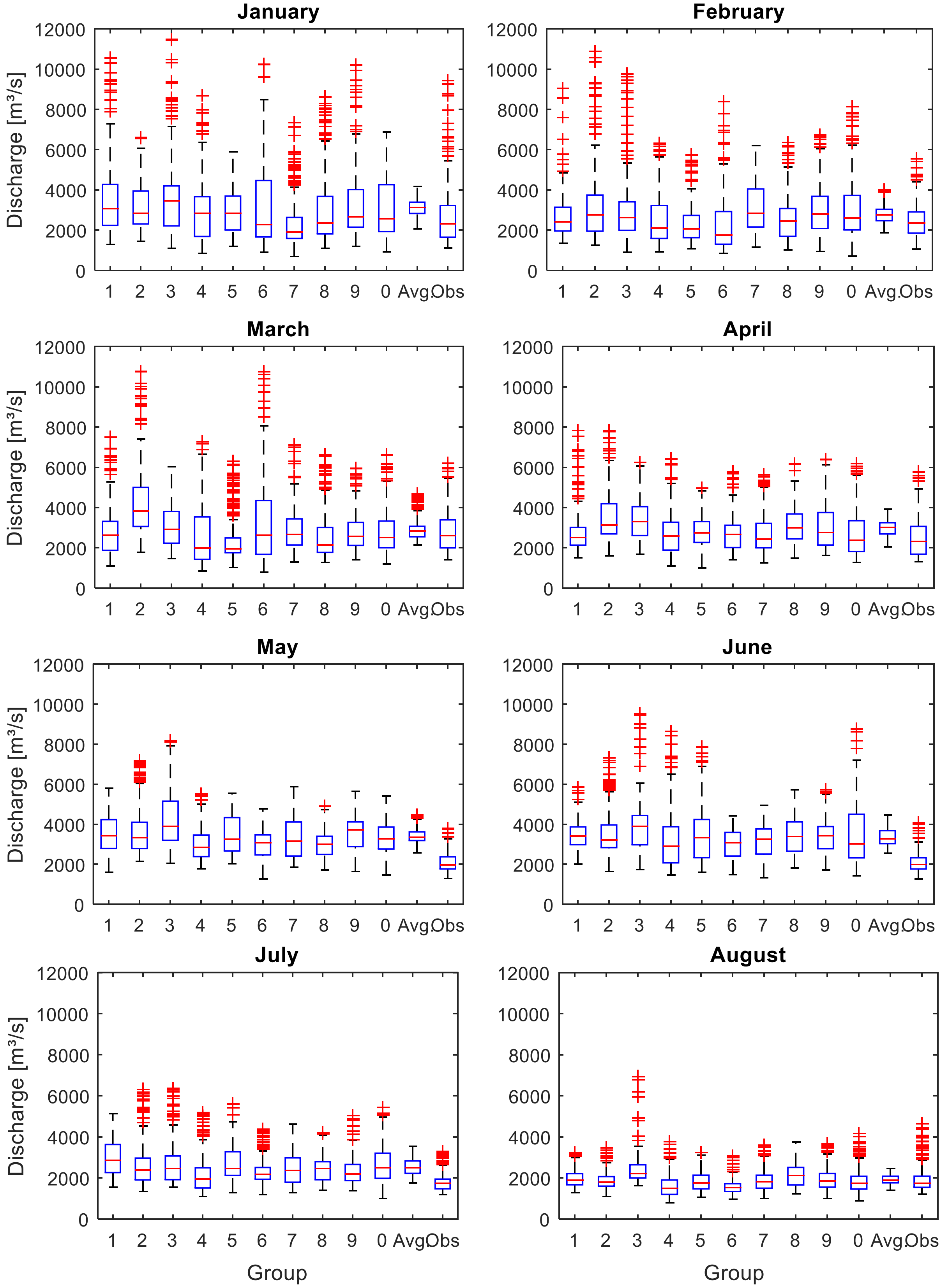


Figure 44 – Cumulative discharge distribution during 1951 -2006 for both observed discharges and discharges based on HYRAS weather.

Discharge boxplots EraClim



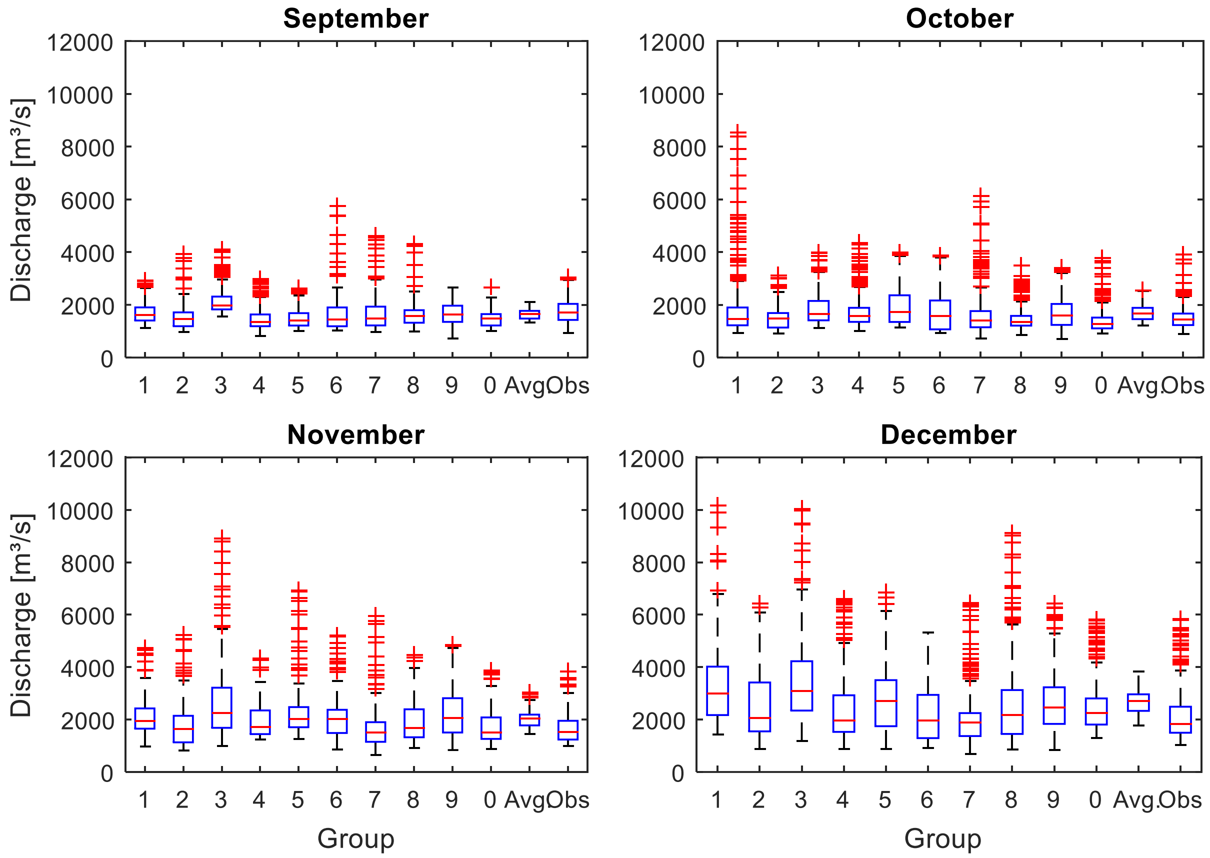


Figure 45 – EraClim based discharges at Lobith against observed Lobith discharge during the period 2003 -2010. Avg. Ens corresponds with the average precipitation of all ensemble members over the whole period.

Discharge boxplot GLOFAS

In the graphs below monthly boxplots are drafted using GLOFAS for the Lobith Discharge. The assessed period is still equal to 2003 – 2010. In the main report one can find the interpretation of these results in comparison with the EraClim boxplots.

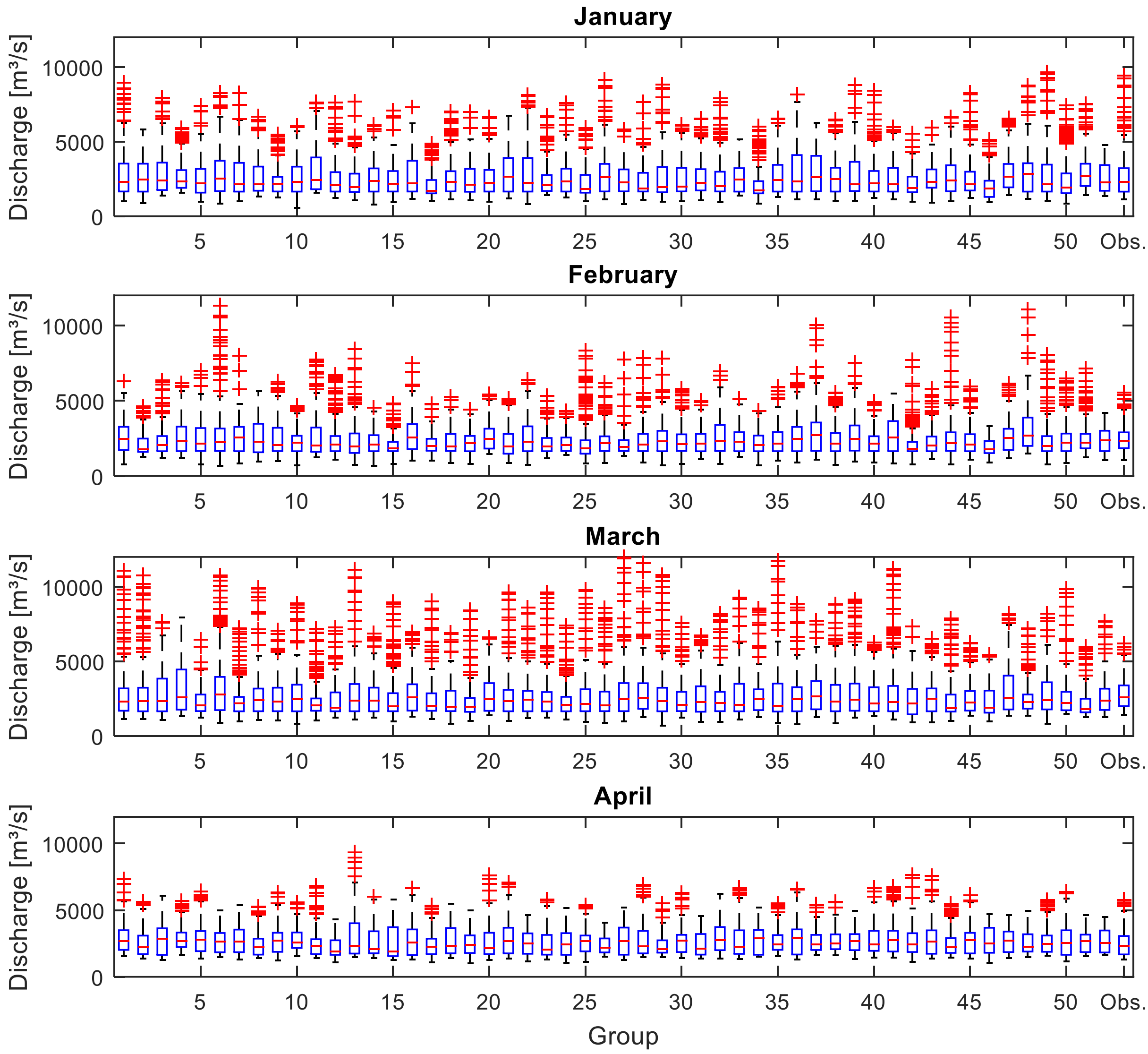


Figure 46A – Monthly GLOFAS based discharges at Lobith (2003 – 2010)

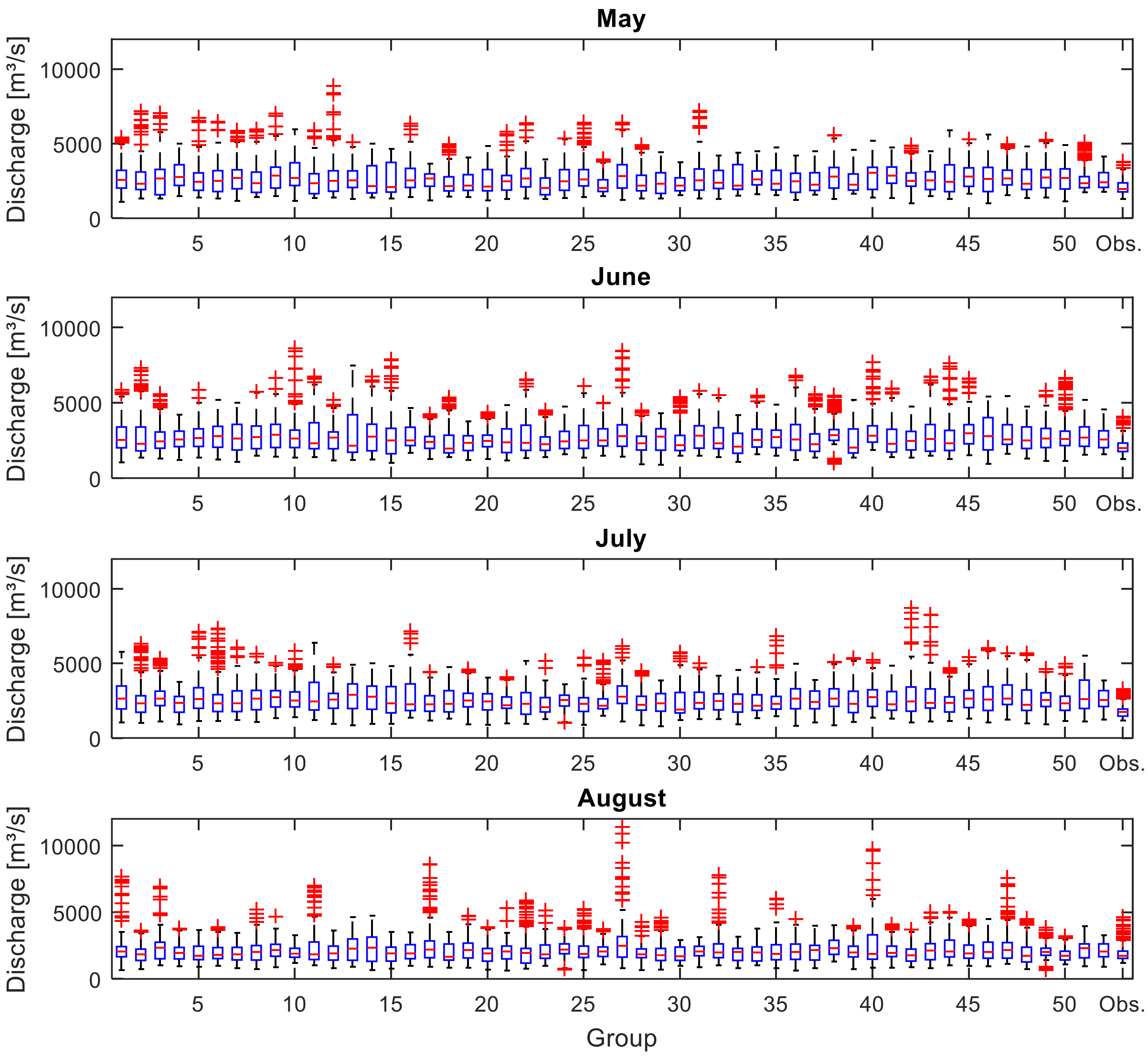


Figure 46B – Monthly GLOFAS based discharges at Lobith (2003 – 2010))

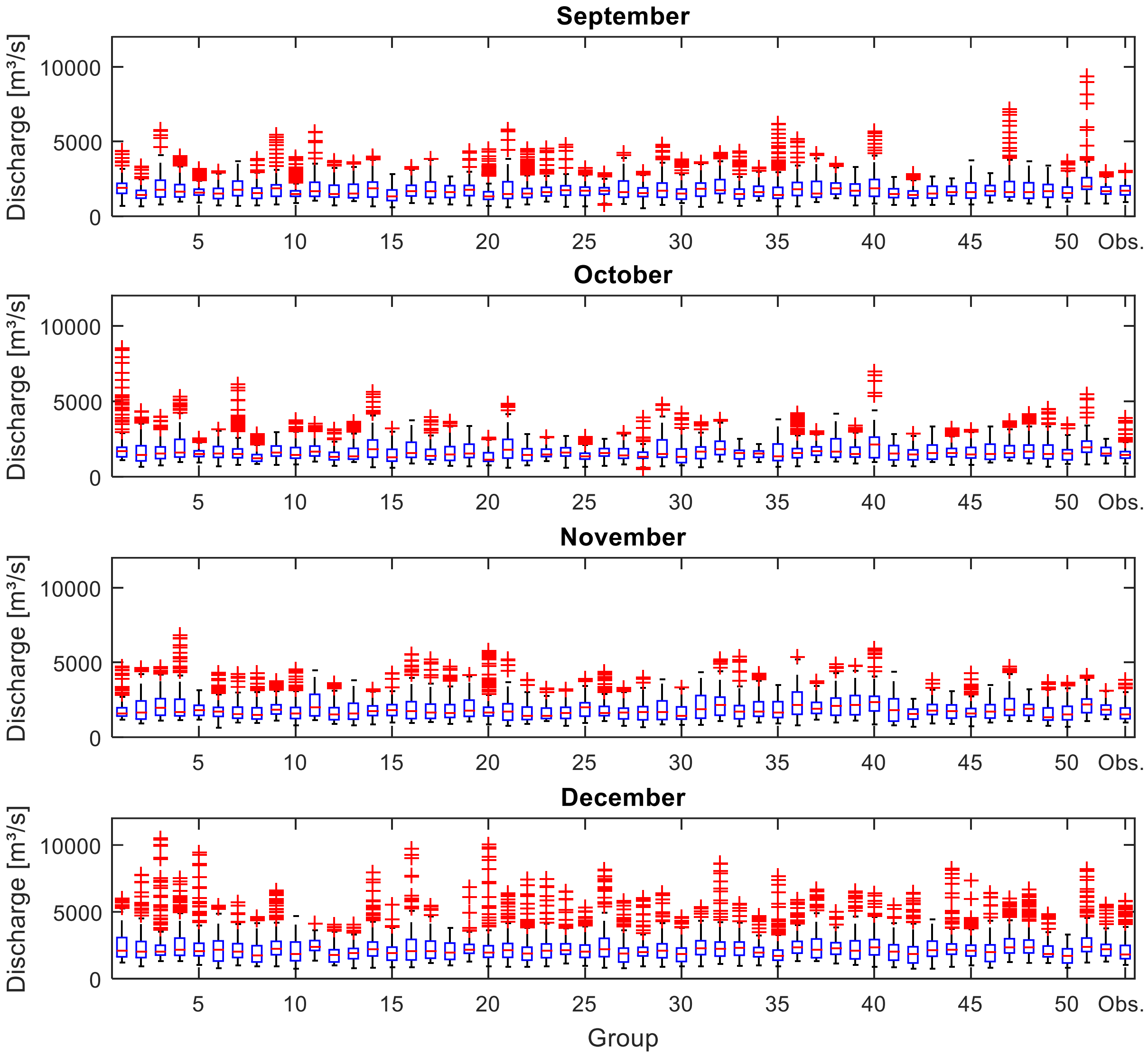


Figure 46C – Monthly GLOFAS based discharges at Lobith (2003 – 2010)

E. Appendix E – Streamflow dependency results

Using the chronological GLOFAS precipitation ensemble members in HBV gives 51 possible streamflow ensemble members for the river Rhine at Lobith for the period between 2003 and 2010. A histogram of all these ensemble members results in figure 47. As can be seen, a log-normal distribution is visible, just as can be expected by discharges (Bowers et al., 2012). The discharges are brought back to a normal distribution by simply taking the natural logarithm of each GLOFAS based streamflow in the discharge series.

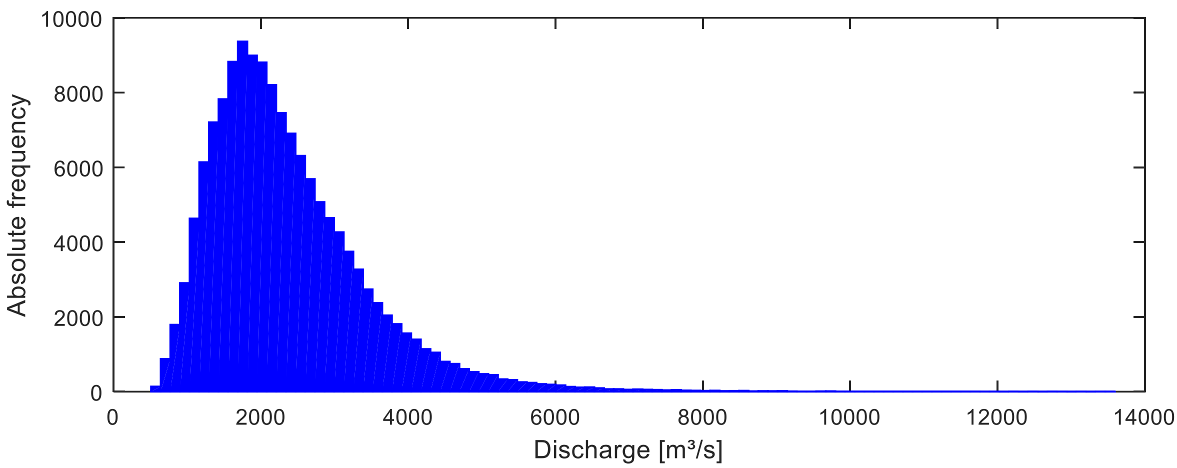


Figure 47 – GLOFAS based streamflow histogram (period 2003 – 2010) with input of all 51 ensemble members.

Subsequently a T-test is performed for the whole length of the time series with a sample size of $n = 10$ days. Every chronological time step of 10 days is statistical compared between mutual ensemble members. Finally a certain percentage of all T-test must be rejected. This means that two 5 day long samples of different ensemble members are statistically different (confidence limit of 95%). Results are shown in figure 48. On average 75,2% of the T-test turns out to be larger than the critical threshold, so 75,2% of the ensemble members deviate mutually. Nice to see is the stability of each streamflow member. Average deviations are not that large as can be seen in the plot.

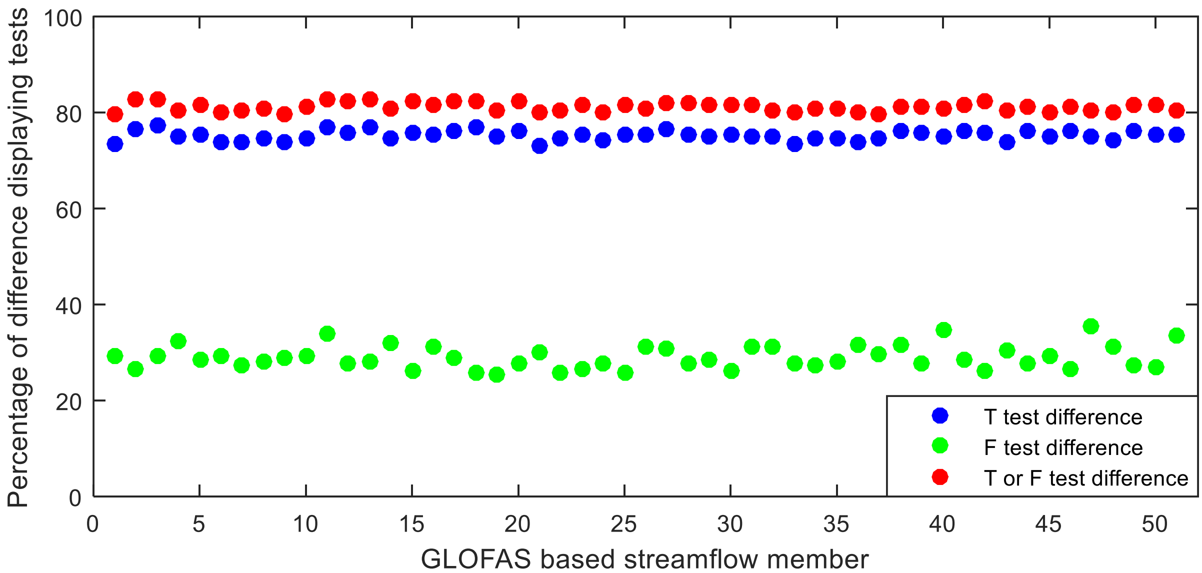


Figure 48 – Percentages of significant differences by testing every ensemble member x with each separate other ensemble members. Sample size in each test is 5 days in this figure (both for T and F test).

Now 75,2% difference is obtained between mutual ensemble members, they might be used to construct discharge time series later on. For now a hypothesis could be that it does not matter which initial precipitation ensemble member is inserted in a chronological ensemble series to obtain discharges after running these precipitation ensemble members with HBV. There will be differences in discharge output which is an important key issue for these series. Statistical differences will be present in different series when they would be build. This implies that different annual discharge statistics are present too, which means that by combining GLOFAS ensemble members, more 'independent' discharge series could be constructed for obtaining different discharge extremes. In fact streamflow ensemble members do not have to be fully different, but (large) variations is already enough. Therefore F- testing is performed in figure 48 too (green dots) to investigate the differences in both standard deviations. From this plot a mean of 29,9% can be derived for which both ensemble samples of 10 days differ concerning the standard deviations of both samples. In 81,1% of the comparisons one of both test (T or F) is positive. This is a slightly higher percentage than the T-test. However 81,1% is very valuable in terms of the degree of difference between ensemble members.

The tests as described above are executed for more sample sizes because of time lag issues. Water from upstream areas will take some time to arrive at Lobith. Figure 49 shows that there are not that many differences noticeable when increasing the sample size.

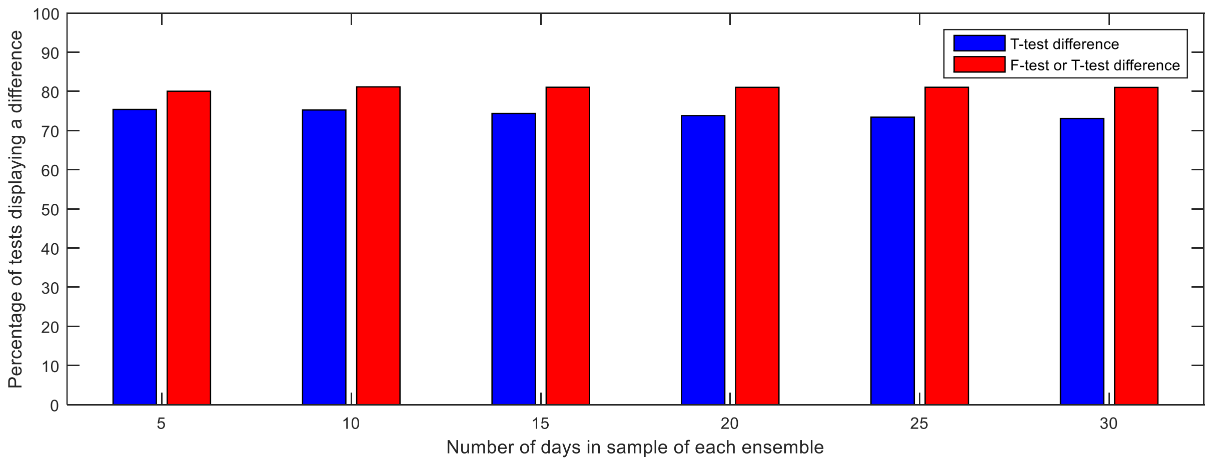


Figure 49 - Percentages of significant differences by testing every ensemble member x with each separate other ensemble members. Sample size varies from 5 days up to 30 days.

F. Appendix F – Ensemble member correlation

GLOFAS

Based on figure 50 some sub basins show a slightly higher correlation between ensemble members. Higher mutual ensemble member correlations are found in sub-catchments located in Alpine regions. Possible explanation for those higher correlations might be the large mountains. Clouds could possibly be get trapped against a ridge, so more of the 51 ensemble members tend to a similar kind of behaviour. This refers to a situation in which more ensemble members will increase precipitation on day $t + 1$ since these are forced by a complex ECMWF weather model. One might imagine that all (read: many) GLOFAS ensemble members will produce more precipitation for the next day when a cloud could not pass a mountain ridge. In case air pressures will change and the cloud could move over the ridge, less precipitation might be predicted by a majority of GLOFAS precipitation ensemble members. Despite the correlations in mountain areas are higher (more less 0,20) one could still defend a very weak mutual precipitation ensemble member correlation.

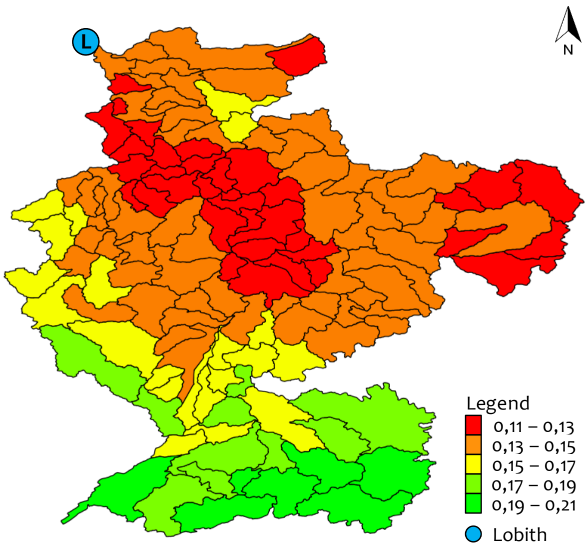


Figure 50 – Correlation distribution of each sub-catchment average GLOFAS precipitation ensemble members (2003-2010).

EraClim

This appendix shows the general and special mutual ensemble member correlation of all 10 EraClim ensemble members. The same method is followed in this appendix as described for GLOFAS in section 3.2 of the methodology.

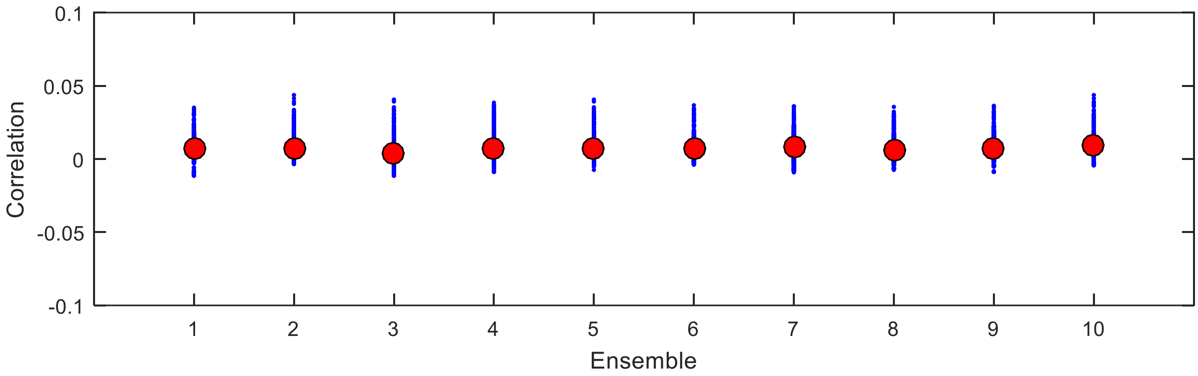


Figure 51 - Mutual correlations between a specific EraClim ensemble member with each of the other EraClim ensemble members. For every ensemble member x the blue dots display the correlations with the other ensemble members. The comparison is executed for each of the sub catchments 1 up to 134. So for ensemble member 1 a comparison is made with ensemble members 2 – 10. This is done for the catchments 1 – 134, so the first series consist of 1206 blue dots. The average of those dots is given by the red spot.

Figure 51 shows the mutual precipitation correlations between a certain ensemble member and all other ensemble members. So the precipitations of the first ensemble member are correlated with the precipitations from ensemble members 2, 3 etcetera. On average a very low (average) correlation is found which almost equals zero. These correlations are compared to GLOFAS (section 4.2) much lower. Probably this is caused by the fact that GLOFAS has many more 'correction points' in the series. After a couple of days the GLOFAS series are using 'fresh' input data which is generated just a few days later than the previous part of the series. EraClim only has such a point after 20 years (Poli et al., 2015). For EraClim a higher variability is present, which might influence the correlation.

Figure 52 shows the average mutual ensemble member correlation for each sub catchment of the River Rhine basin for the EraClim data. Most sub-catchments are performing more less the same with an average correlation of 0.006. At the right of the figure some sub catchments are visible with a strongly deviating character. These correlations are still very low, but the pattern shift is at least remarkable. Figure 53 is drafted in order to see the spatial spreading of these correlations in the river Rhine basin.

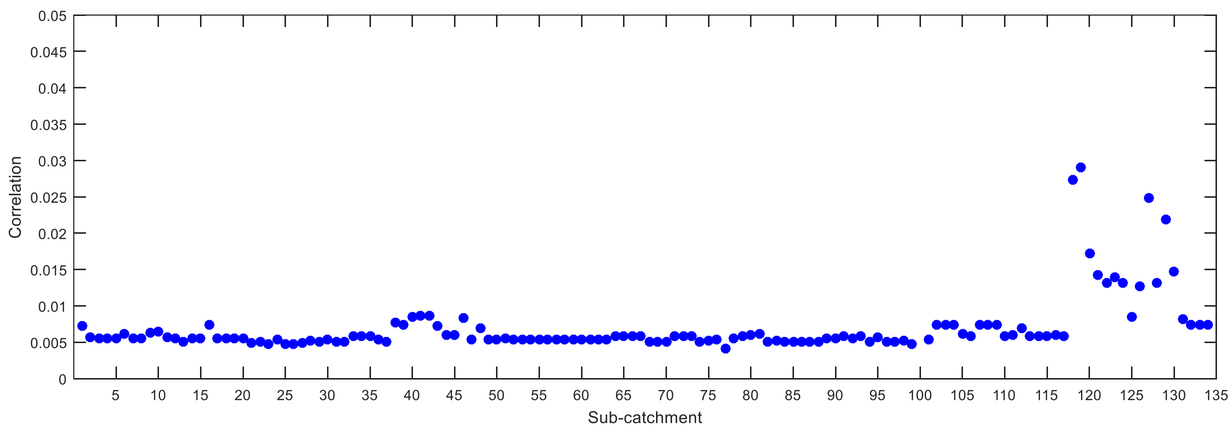


Figure 52 – Average correlations based on mutual comparison of EraClim precipitation ensemble members per sub-catchment

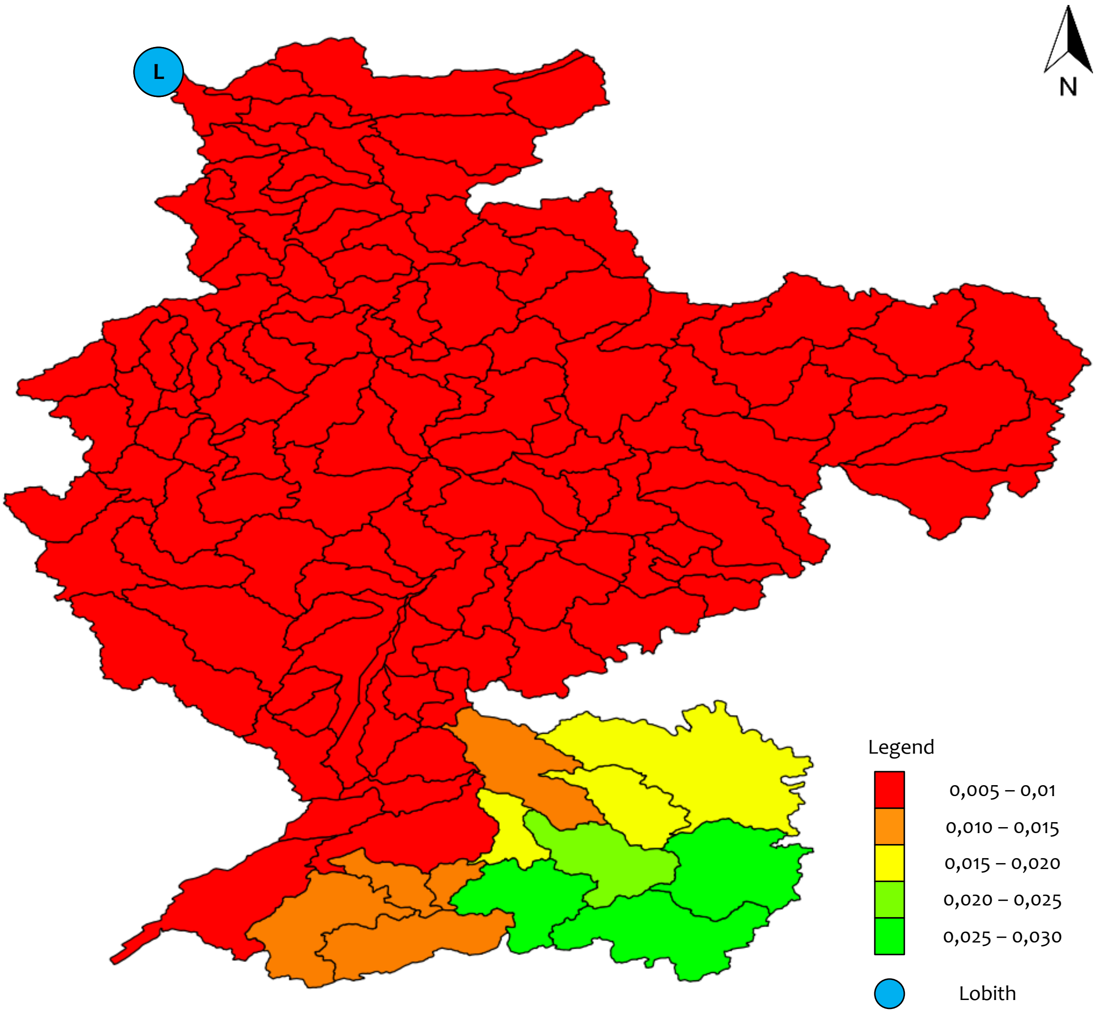


Figure 53 – Correlation distribution of each sub-catchment average EraClim precipitation ensemble members (2003-2010).

G. Appendix G – Closer view sub basin extreme discharge

In this appendix the extreme discharge statistics are captured for sub-basins which are not shown in the main-report. A separate analysis is not provided here, since most important 'general' patterns are described in section 4.6. The number of annual maxima used is as follows:

- GLOFAS (Synthetic), N = 1090
- EraClim, N = 1090
- Observations, N = 55 for all sub basins. Only Frankfurt has N = 43 annual maxima
- HYRAS N = 55

Extreme discharge distributions

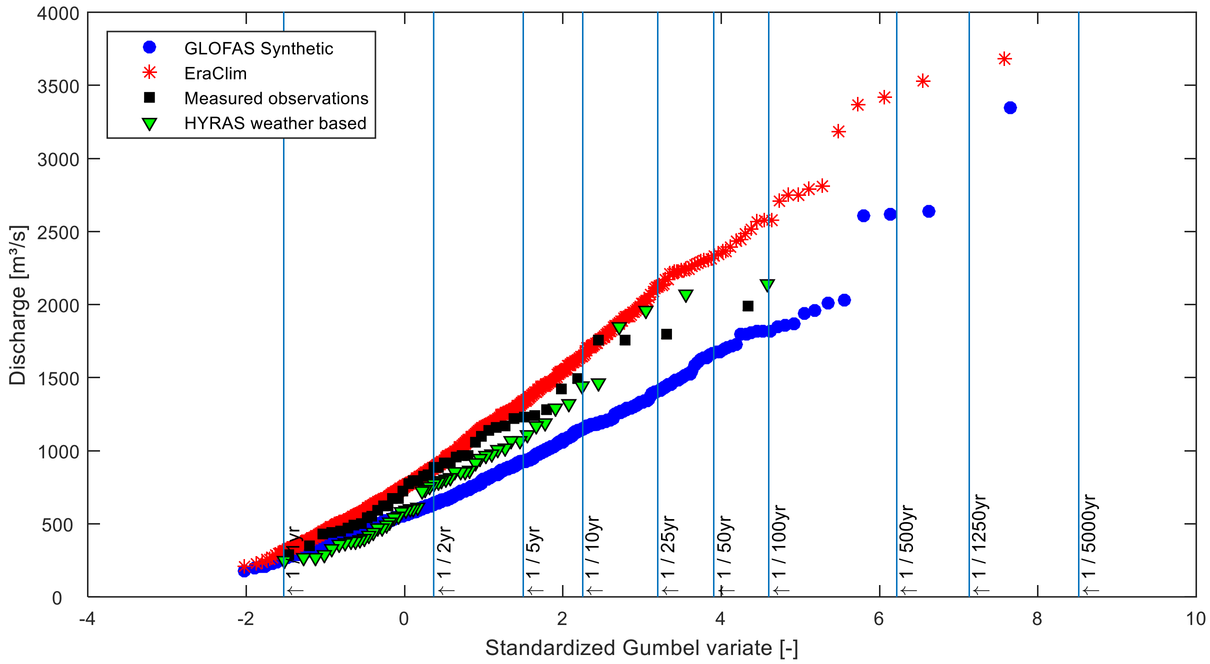


Figure 54 - Extreme discharge distribution at Frankfurt for more datasets.

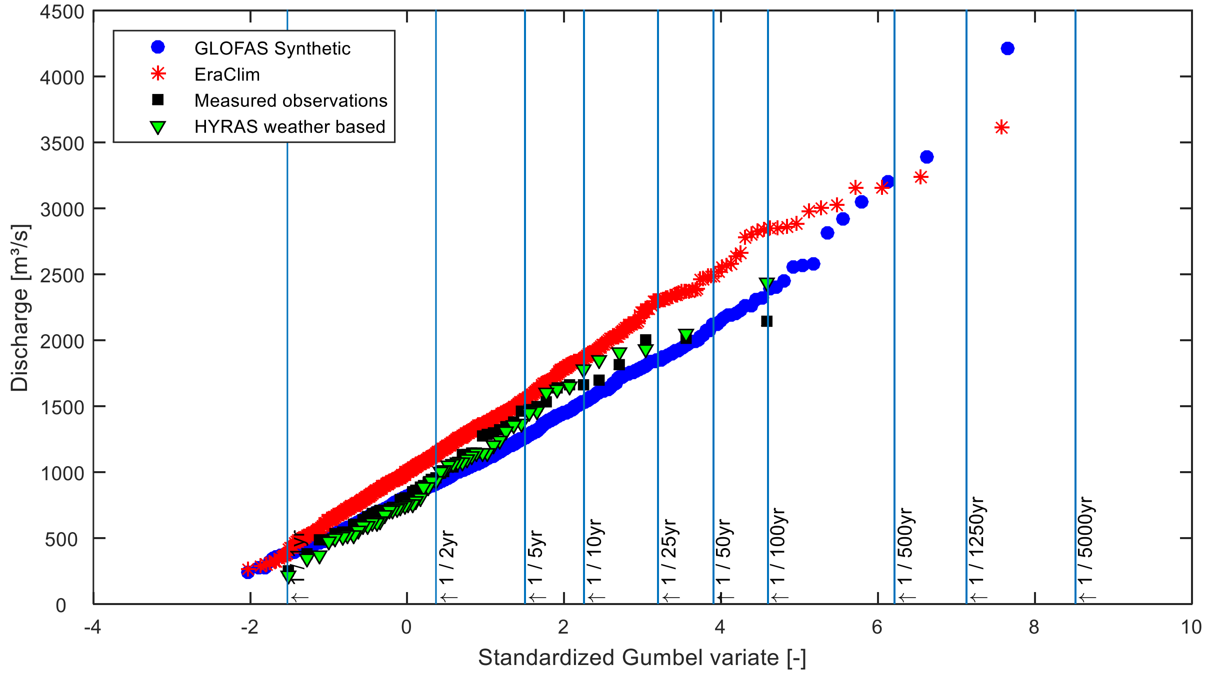


Figure 55 - Extreme discharge distribution at Rockenau for more datasets.

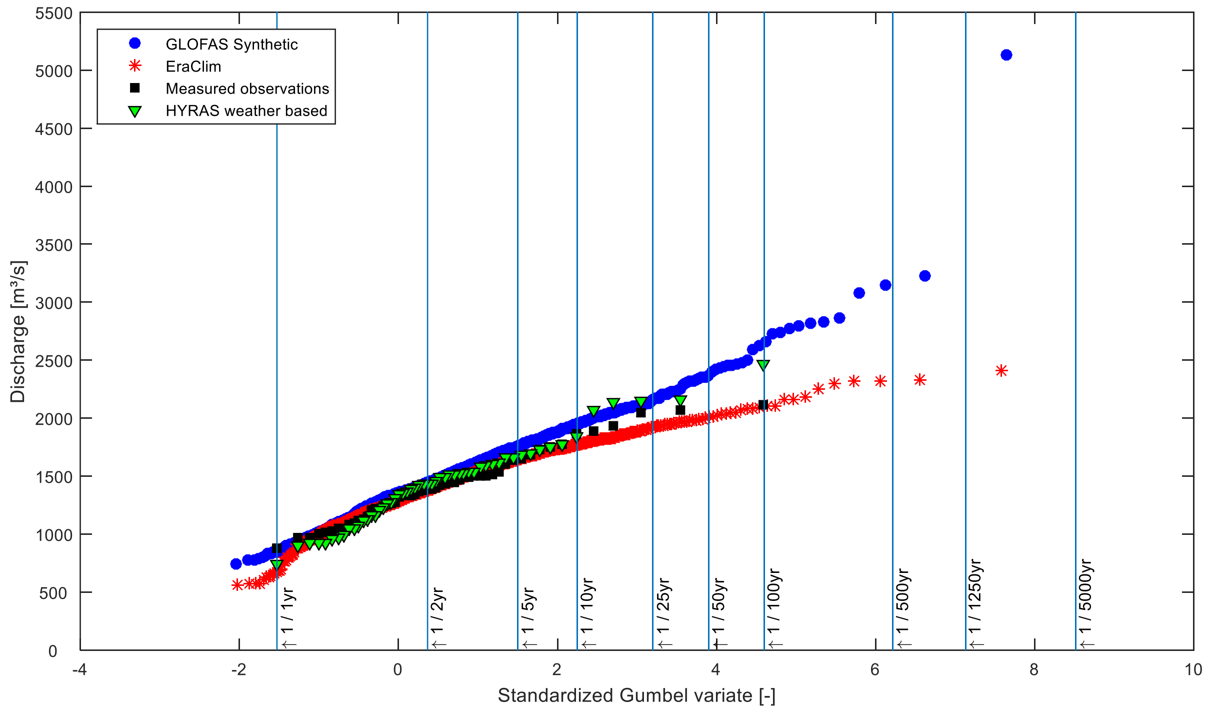


Figure 56 - Extreme discharge distribution at Untersiggenthal for more datasets.

Relative monthly peak frequencies

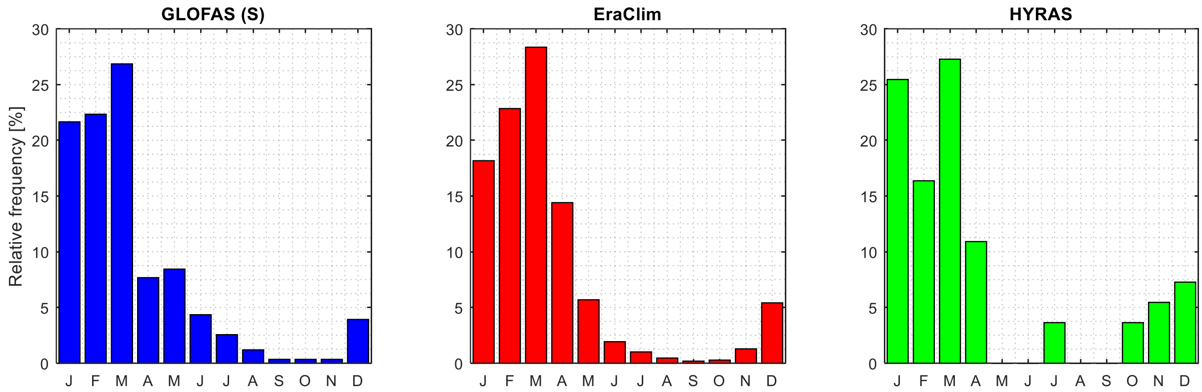


Figure 57 - Relative peak frequencies per month for multiple datasets at Frankfurt

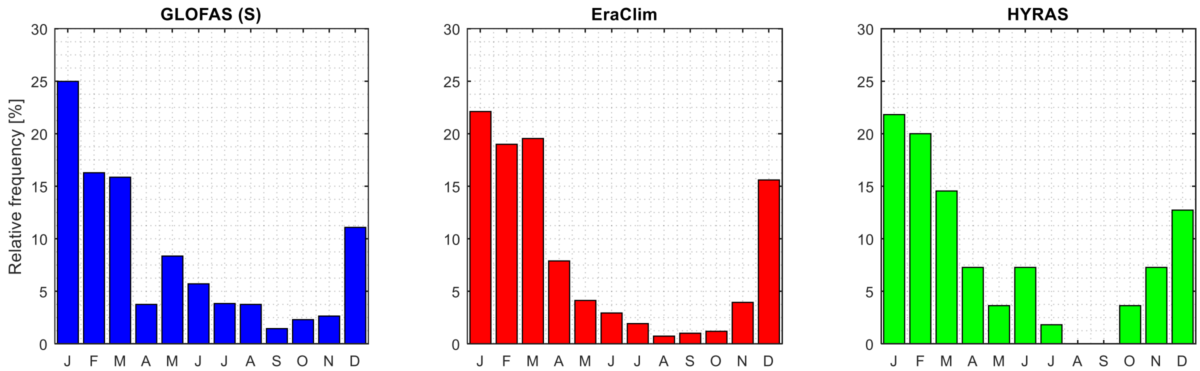


Figure 58 - Relative peak frequencies per month for multiple datasets at Rockenau.

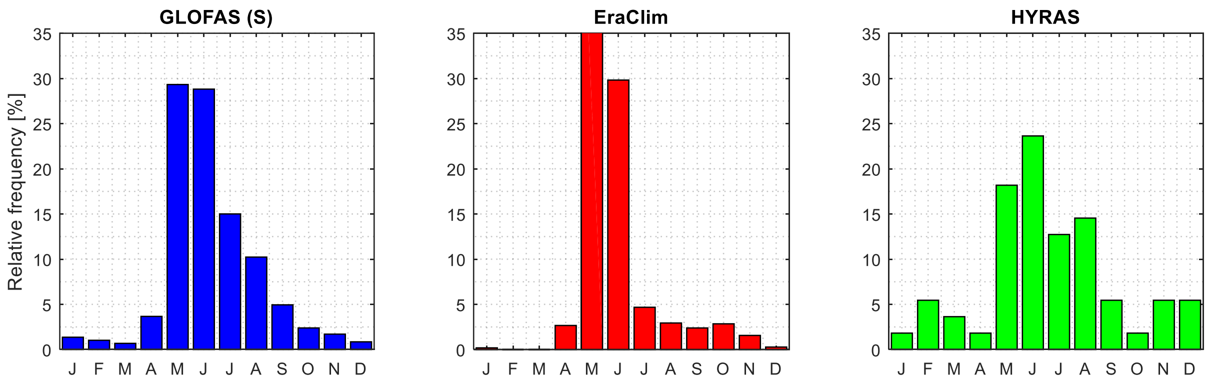


Figure 59 - Relative peak frequencies per month for multiple datasets at Untersiggenthal.

Precipitation prior to peak discharge

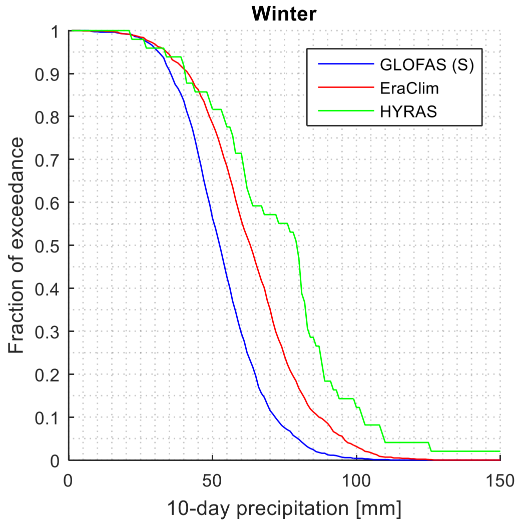


Figure 60 – Precipitation duration curve at Cochem in the Moselle basin.

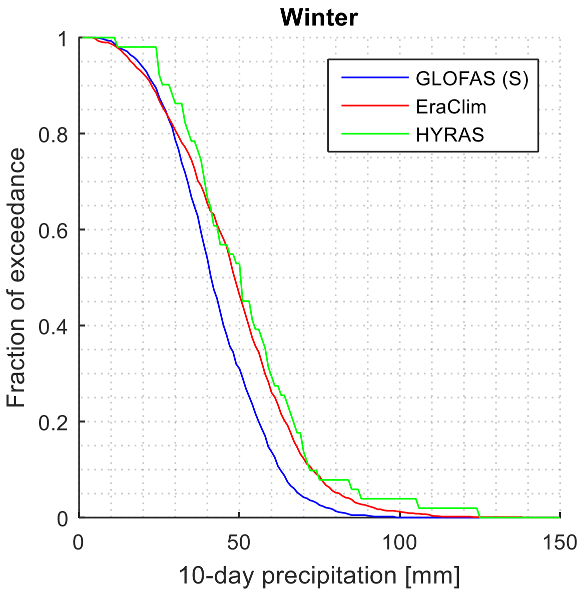


Figure 61 – Precipitation duration curve at Frankfurt in the Main basin.

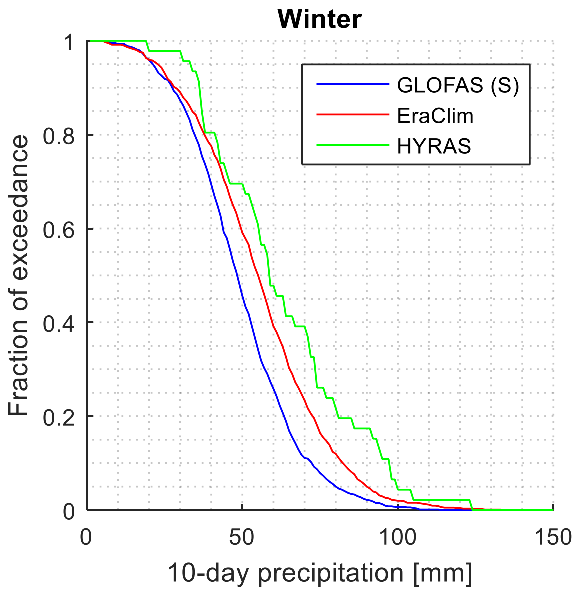


Figure 62 – Precipitation duration curve at Rockenau in the Neckar basin.

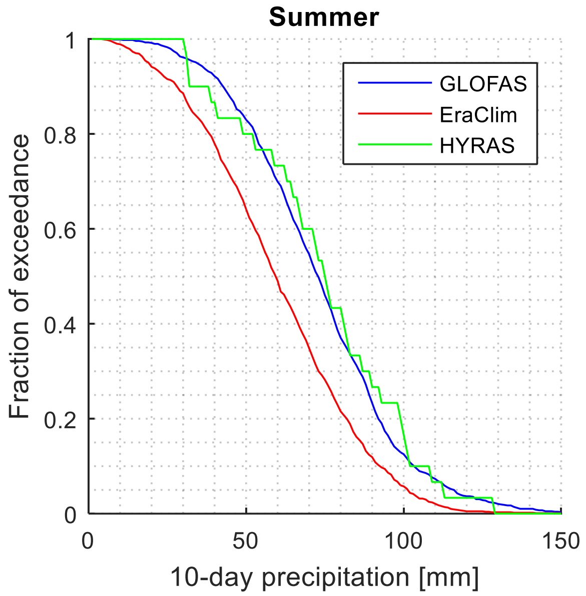


Figure 63 – Precipitation duration curve at Rekingen in the east-Alpine basin.

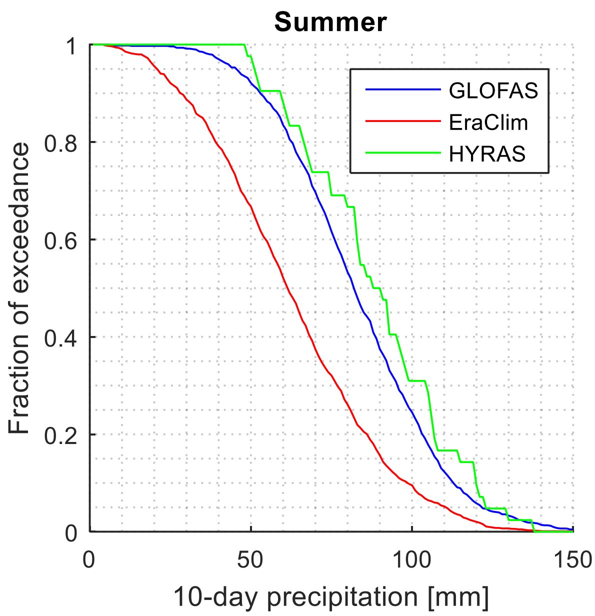


Figure 64 – Precipitation duration curve at Untersiggenthal in the west-Alpine basin.