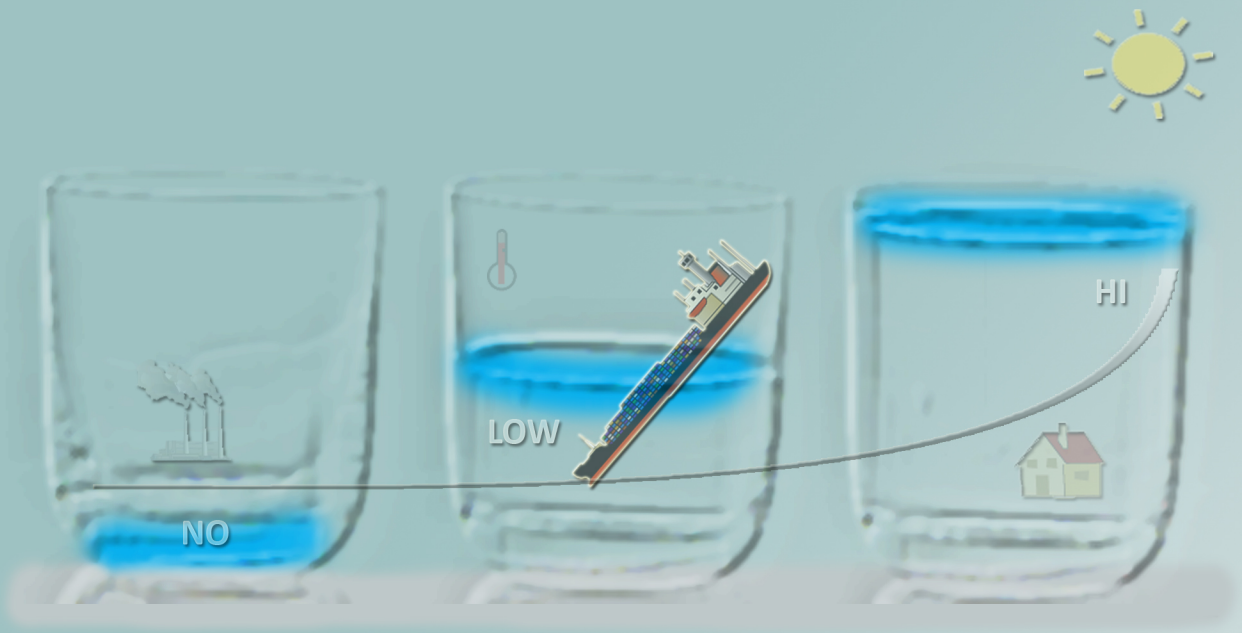
LOW FLOWS

MECHANISMS, FORECASTS and CLIMATE CHANGE IMPACTS



MEHMET CÜNEYD DEMİREL

LOW FLOWS

Mechanisms, forecasts and climate change impacts

Promotion committee:

prof.dr. G.P.M.R. Dewulf	University of Twente, chairman and secretary
prof.dr.ir. A.Y. Hoekstra	University of Twente, promotor
dr.ir. M.J. Booij	University of Twente, assistant promotor
prof.dr. J.C.J. Kwadijk	University of Twente
prof.dr. Z.B. Su	University of Twente
prof.dr.ir. R. Uijlenhoet	Wageningen University
prof.dr. T. Wagener	Bristol University

ISBN 978-90-365-35656

DOI: 10.3990/1.9789036535656

Typeset in LaTeX

Cover design: M.C. Demirel

Picture courtesy: Photographer Tom Buijse (Figure 1.1)

Printed by CPI - Koninklijke Wöhrmann, Zutphen, the Netherlands

LOW FLOWS

Mechanisms, forecasts and climate change impacts

PROEFSCHRIFT

ter verkrijging van
de graad van doctor aan de Universiteit Twente,
op gezag van de rector magnificus,
prof. dr. H. Brinksma,
volgens besluit van het College voor Promoties
in het openbaar te verdedigen
op vrijdag 20 februari 2014 om 16.45 uur

door

Mehmet Cüneyd Demirel
geboren op 03 jan 1980
te Konya, Türkije

Dit proefschrift is goedgekeurd door de promotoren:

prof.dr.ir. A.Y. Hoekstra (promotor)
dr.ir. M.J. Booij (assistant promotor)

Knowledge should mean a full grasp of knowledge:
Knowledge means to know yourself, heart and soul.
If you have failed to understand yourself,
Then all of your reading has missed its call.

What is the purpose of reading those books?
So that Man can know the All-Powerful.
If you have read, but failed to understand,
Then your efforts are just a barren toil.

The true meaning of the four holy books
Is found in the alphabet's first letter.
You talk about that first letter, preacher;
What is the meaning of that – could you tell?

Yunus Emre says to you, Pharisee,
Make the holy pilgrimage if need be
A hundred times – but if you ask me,
A visit to the heart is best of all.

Yunus Emre (1240 – 1321), *Anatolian poet*

Contents

Preface

High-impact low flows, yes, this is what my book is all about. It may seem controversial to do a research on low flows for a " wet" country protected by dykes and barriers like the Netherlands where contingency plans often focus on high flows. When I started this PhD project in August 2008, the topic of low flows in the River Rhine was a puzzle to me. The pieces of this puzzle were found during the trips to the water authorities in Bern, Zurich, Koblenz and Delft. The bicycle trip along the upper-Rhine from Köln to Dordrecht, with our master students (Erwin, Daniel, Wiebe), was instrumentally helpful to see (and to feel) the River Rhine in the summer period. Day by day, with ups and downs, the pieces of the puzzle were fit in a thesis context. I enjoyed the entire process and I am content with the final result. The reader will observe that the flow levels might be low but the societal impacts are high. The results were appreciated by several high impact scientific journals. Especially publishing in a journal so-called *Water Resources Research* was my obsession, my Nirvana mission, and it has happened. Without the help of many people around me, I could complete neither this mission nor this thesis.

First of all, the people who were with me in the project cockpit: Martijn Booij and Arjen Hoekstra. Both deserve my deep appreciation and thanks as being my co-pilots, advisers, mentors and friends. Their critical reviews, patience, support, motivation and wisdom (sense of humour usually after a beer too) enabled me to combine learning and enjoyment in this uplifting journey for my career. I continue the compliments with particular highlights for each member of the project cockpit. Arjen: our first philosophical talk was in Judith Janssen's PhD party evening. This was an important evening to know your life perspective and sincerity. You have provided me comfortable research conditions as long as I needed, and flexibility to overcome many bureaucratic pitfalls for a non EU student. Your critical views from your numerous experiences on writing helped me in making a clear story line for our scientific papers. You had a grip on the total project picture. Martijn: keeping on the metaphor of the project cockpit, you quickly assessed the engine–capacity and wing–width of our project plane which were the two key factors to recommend on reasonable flight altitudes and distances to reach the final destination safely. In other words, your wisdom and trust always motivated me. Your serious attention on every detail of the manuscripts improved

the quality of my work. Not surprisingly, none of our works have been rejected yet. This is also from the fact that I have followed The EU-Maastricht (Maartjenstrict) criteria properly. Further, I met many top–hydrologists in your network when we were together in a conference in India. My parents always mention about your family's warm approach to them during their first visit to the Netherlands.

I want to thank *Dr.ir. Cornellis Lely* foundation and University of Twente (UT) for funding this research project. The expert group of the project was very useful for the project. The members of this group were Eric Sprokkereef, Jules Beersma, Ron Passchier, Durk Klopstra and Marcel de Wit. Thank you for providing data, critically reviewing my research results and giving me the opportunity to meet other scientists in the Rhine domain. It has been a great and tragic loss that Marcel de Wit passed away in an early stage of my research.

The Water Engineering and Management group of the UT is a good place to do research. Besides, I have been given a considerable amount of budget to improve my skills by following courses and attending conferences and two summer schools. Everyone was always available for advice and I enjoyed the friendship of my colleagues in our water department. I would like to thank you all for that. Special thanks to my master and Erasmus students (Pieter, Jasper and Hakan), Kristine (pink teddycik), Ertuğ, Ramazan, Yusuf, Marcela, Tolga, Blanca, Mustafa, Gül, Judith, Freek, Mireia, Erika, Markus, Pieter van Oel, Matthieu, Hatem, Guoping, René, Kathelijne and Joep. Big scientists Bas and Jord, thanks for your help on LATEX. I would like to thank my "paranimfen" Mesfin Mekonnen and Suleyman Naqshband for their efforts during the period of my graduation. I would like to thank the secretaries of the WEM group, Joke, Anke and Brigitte for their help in administrative issues. I would like to thank the UT staff, Cecile Schouten, Bauke Visser, Sophie Vrielink-Witterick and Veronique Arnold. Their help in visa related issues and financial management of the project is indispensable. Futher, I would like to thank Isa Adıgüzel (and his lovely family) and my colleagues from Istanbul Technical University, Ercan Kahya, Mehmet Karaca, Azize Çırtlık, Mehmet Parlakyiğit and Mehmet Coşkun for their continuous support. I am very content that we have a good level of understanding since day one. I will remember the whole PhD period nicely and I hope you will remember me when you see (or feed) my six goldfishes (koi and shubunkin) in the pool in front of our faculty. The tree which I have planted in front of my house in 2010 will also remain as a witness to time.

A particular thank goes to my parents (Mehmet and Hediye), two brothers (Selman and Sami) whose belief in me through the pitfalls during the research phase was uplifting and rejuvenating.

Mehmet Cüneyd Demirel
Enschede, 28 November 2013

Summary

The water levels in rain-fed rivers can drop seriously after a prolonged dry period. Low flows may cause several problems for river users. Understanding low flows at different time scales to improve medium-range (10 day) and seasonal (90 day) low flow forecasts have both societal and scientific value as there is an increasing interest to account for low flow forecasts in decision support systems, e.g. how to operate river navigation and power plants during low flow periods to maximize benefits and minimize costs.

The objective of this thesis is to explore low flow mechanisms, develop forecast methods and assess climate change impacts on low flows. The Rhine basin is taken as case study area. To achieve the research objective, first, in **chapter 2** dominant low flow indicators are identified based on correlation analysis between low flow indicators and observed low flows at different lag times and temporal resolutions. The most important indicators in the Alpine sub-basins for medium-range low flow forecasts are potential evapotranspiration and lake levels. In the other sub-basins, groundwater and potential evapotranspiration correlate best to low flows. The most important indicators for seasonal low flow forecasts are potential evapotranspiration, lake levels and snow depths for the Alpine sub-basins, whereas in the other sub-basins the most important indicators are potential evapotranspiration and precipitation or groundwater. The identified low flow indicators and their appropriate temporal scales are then used to select appropriate models from available hydrological models and to develop data-driven models for medium-range and seasonal forecasts.

Based on the low flow indicators in **chapter 3**, the GR4J and HBV conceptual hydrological models are selected for evaluating the effect of major uncertainty sources, e.g. input, parameter and initial condition uncertainty, on medium-range low flow forecasts. Monte Carlo simulation is used to show the effect of different uncertainties on the skills of low flow forecasts and to derive the probabilities and ranges in forecasted low flows. The 10 day ensemble forecast results show that the daily observed low flows are captured by the 90% confidence interval for both models most of the time, whereas the GR4J model usually overestimates low flows and HBV is prone to underestimation of low flows. This is particularly the case if the parameter uncertainty is included into the forecasts. The total uncertainty in the GR4J model outputs is higher than in the HBV model outputs. Moreover, the parameter uncertainty has the largest effect on

the low flow forecasts and the input uncertainty the smallest. The forecast distribution based on 10 day low flow forecasts issued by the HBV model is the most reliable forecast distribution if only input uncertainty is considered. The number of correct low flow forecasts (hits) is about equal for the two models only if the input uncertainty is considered. The parameter uncertainty is the main reason reducing the number of hits. The number of false alarms is almost doubled with respect to the HBV model for the GR4J model when considering total uncertainty. The importance of parameter uncertainty for the quality of forecast is emphasized by all forecast quality measures used in this study. Overall, the output from two conceptual hydrological models is characterised by substantial uncertainty from model parameters. The parameter uncertainty effects mainly the reliability and the sharpness of the forecasts. This finding is new for medium-range low flow forecasts as the input uncertainty (mainly the rainfall forecast error) is generally assumed to be the most important uncertainty source in hydrological forecasting (Pappenberger *et al.*, 2005).

The selected two conceptual models, GR4J and HBV, are also used for seasonal low flow forecasts in **chapter 4**. In addition, two data-driven models are developed based on the low flow indicators. The effect of seasonal climate forecasts on low flow forecasts is assessed using the two conceptual hydrological models and two data-driven models, i.e. ANN-Ensemble and ANN-Indicator. These models are run using different seasonal meteorological forcings: **(1)** ensemble precipitation (P) and potential evapotranspiration (PET) forecasts, **(2)** ensemble P and climate mean PET, **(3)** climate mean P and ensemble PET, **(4)** climate mean P and PET, and **(5)** zero P and ensemble PET. The ensembles provided the forecast uncertainty range for the model inputs. The ranges, calculated for five cases, are compared, whereas the skill of low flow forecasts is evaluated for varying lead times up to 90 days. The results show that all models are prone to over-predict low flows using ensemble forcing and the largest range for 90 day low flow forecasts is found for GR4J using ensemble seasonal climate forecasts as input. The results of the comparison of forecast skills with varying lead times shows that the low flow forecasts using GR4J are less skilful than using the other three models. Further, the hit rate of ANN-Ensemble is higher than that of the other models for all lead times except for a 90 day lead time. The 90 day ahead low flow events in a very dry year, i.e. 2003, are correctly forecasted by ANN-I showing the skill of data-driven models for seasonal forecasting. Overall, the uncertainty arising from ensemble precipitation has a larger effect on seasonal low flow forecasts than uncertainties from ensemble potential evapotranspiration and model initial conditions.

In **chapter 5**, the impacts of climate change on the seasonality of low flows are assessed using the outputs of an ensemble of climate models to run a hydrological model. Three seasonality indices, namely seasonality ratio (SR), weighted mean occurrence day ($WMOD$) and weighted persistence (WP) were used to reflect the discharge regime, timing and variability in timing of low flow events, respectively. The analysis focuses

on the effects of the hydrological model and its inputs, the use of different GCMs and RCMs and the use of different emission scenarios. Significant differences are found between seasonality indices based on observed and simulated low flows with observed climate as input due to the uncertainty arising from hydrological model inputs and structure. Further, the weighted mean occurrence day and the weighted persistence in the two Alpine sub-basins show larger differences compared to the rain-dominated sub-basins. The comparison of the three seasonality indices based on observed inputs and simulated inputs reveals small differences in *SR* for all sub-basins except for the Moselle sub-basin. Large differences are found for the *WMOD* and *WP* indices, showing that these indices are very sensitive to uncertainties from the climate models. The comparison of the three seasonality indices using simulated inputs for the current climate and simulated inputs for the future climate resulted in the largest range for *WP* and the smallest range for *SR*. The *SR*s by 2063-2098 significantly decrease in all sub-basins, showing that a substantial change in the low flow regime in all sub-basins of the River Rhine is expected, whereas a regime shift from winter to summer low flows is likely to occur in the two Alpine sub-basins. Moreover, the *WMOD*s of low flows tend to be earlier in the year than for the current climate in all sub-basins except for the Middle Rhine and Lower Rhine sub-basins. By 2063-2098, the *WP*s slightly increase compared to current climate, showing that the predictability of low flow events increases as the variability in timing decreases. Overall, the comparison of the uncertainty sources evaluated in this study shows that different GCMs and RCMs have more influence on the timing of low flows (*WMOD*) than different emission scenarios. The influence of different GCMs and RCMs on *SR* is slightly larger than the influence of different emission scenarios on *SR*, whereas the influence of different GCMs and RCMs on *WMOD* is similar to the influence of different emission scenarios on *WMOD*.

In this thesis, different low flow mechanisms in the Rhine basin have been identified. These mechanisms are used to select two appropriate conceptual models for 10 day and 90 day ahead low flow forecasts. Moreover, the identified temporal scales of the dominant low flow mechanisms are used to develop two data-driven seasonal models. The effects of major uncertainty sources on low flow forecasts are assessed using Monte Carlo methods. Parameter uncertainty is found to have the largest effect on 10 day low flow forecasts, whereas ensemble seasonal precipitation forecasts have the largest effect on 90 day low flow forecasts. In the final step, we assessed the possible changes in seasonal low flow characteristics for the period 2063-2098 using three important indices reflecting the low flow regime, timing and persistence. Overall, the effect of the most important uncertainty sources on 10 day and 90 day ahead low flow forecasts and the impacts of climate change on the seasonality of low flows were evaluated. The identified lags and temporal resolutions are useful in creating operational medium-range and seasonal low flow forecast models for the River Rhine. The results can also be useful for other snow dominated and rain dominated catchments in the world. Further,

understanding the low flow mechanisms and subsequent storage responses should aid the selection of appropriate models and the choice of proper temporal scales. Critical catchment characteristics (e.g. the extent of aquifers) determine the applicability of the results for appropriate model selection. The exact results about the order of uncertainty and the type of models can be applicable for other river basins to evaluate low flow forecasts.

Chapter 1

Introduction

1.1 Rivers and low flows

Rivers have been very important to civilisations in all parts of the world since prehistoric times. An Earth without rivers is similar to a body without veins. Therefore, living along the rivers has always been preferred. The valley between the Tigris and the Euphrates rivers is blessed in many cultures. Many rivers, such as the Amazon, Guadiana, Nile, Danube, Meuse and Rhine, cross the borders of several countries. These rivers are used for drinking water supply, irrigation, industrial use, power production and freight shipment (De Wit *et al.*, 2007). They also provide an important habitat for wildlife and play an essential role in the ecology of rainforests and wetlands. Although most of the large rivers never dry up, the water levels can drop seriously after a prolonged dry period (i.e. hydrological drought). Similarly, large amounts of precipitation in a short time can cause flood events. Both floods and low flows are seasonal phenomena that may cause several problems to society.

Hydrological droughts are slowly developing events affecting a much larger area than floods. The economic loss during a hydrological drought period, causing low flows, is much higher than during floods (Pushpalatha *et al.*, 2011; Shukla *et al.*, 2012). Severe problems, e.g. water scarcity for drinking water supply and power production, hindrance to navigation and deterioration of water quality, have already been experienced during low flow events in the River Rhine in dry summers such as in 1976, 1985 and 2003. The River Rhine is selected as study area for this thesis (Figure 1.1). The reader will observe that this thesis is based on three main pillars: first, identification of dominant low flow mechanisms to improve forecast models; second, rigorously assessing uncertainty in 10 day and 90 day low flow forecasts; and third, evaluating the future trends in low flows due to climate change.

Figure 1.1: Low flow conditions in the River Rhine. Picture courtesy: Tom Buijse

1.2 Low flow mechanisms in the River Rhine

The river flow is the result of many interacting mechanisms acting at various spatial and temporal scales in a river basin. Different mechanisms are dominant at different scales during the transformation of precipitation to river flow due to the heterogeneity in water carrying media. The large Rhine basin, spanning from the Swiss Alps to the lowlands in the Netherlands, contains different land use types and aquifer systems. The average discharge downstream of the Alpine mountains is approximately 1000 m^3/s. It then increases up to 2,300 m^3/s at the Lobith gauging station after the German-Dutch border. The minimum observed discharge at this gauging station was 575 m^3/s in 1929. The contribution of the Alps to the total discharge can be more than 70% in summer, whereas it is only about 30% in winter (Middelkoop and Van Haselen, 1999). In the winter period, the precipitation is stored as snow and ice in the Alps until late spring. Due to the high evaporation and little melt-water input from the Alps, low flows typically occur in late summer or autumn (Nilson *et al.*, 2012).

The river discharge during a low flow period mainly originates from groundwater storage and the outflow follows a characteristic recession curve (Schneider, 2008). Basin characteristics such as geology, soil type, topography, vegetation, hydraulic conductivity and extent of the aquifer determine the magnitude and timing of groundwater discharge to streams (Hattermann *et al.*, 2004; Allen *et al.*, 2010; Burn *et al.*, 2008). Apart from that, the release from other large storages controlled by gravity, such as large lakes, snow storage and glaciers, can be important in sustaining low flows (Tallaksen and Van Lanen, 2004; Suweis *et al.*, 2010). Since low flows may occur in any season in the River Rhine, it is crucial to identify which dominant low flow mechanisms and temporal scales should be incorporated in a low flow model. This will be the first sub-

ject in this thesis. Selection of dominant temporal scales is important as it affects the required input data, the processes that can be well presented, and therefore, usefulness of the forecast results.

1.3 Ensemble low flow forecasts and uncertainty

To anticipate low flow events it is crucial that medium-range (10 day) and seasonal (90 day) low flow forecasts become available in addition to short-range (1 – 4 days) forecasts. The Dutch Water Service and the German Hydrology Institute forecast discharges with lead times varying from six hours to four days using different models including regression based models, unit hydrographs and flow routing schemes for the Rhine. Currently, various models that run with outputs from numerical weather forecast models at a daily or sub-daily time step are used for operational or semi-operational purposes by water agencies in the riparian countries (Cloke and Pappenberger, 2009). Hourly forecasts are important for flood warnings, whereas daily forecasts are used to forecast low flows. The low flow forecasts with longer lead times can be particularly beneficial for river navigation and power plants to maximize their gain. However, the reliability of weather forecasts is questionable for longer lead times than one week due to the uncertainties in the weather forecast results. Therefore, no single weather forecast is issued by climate institutes. Instead, an ensemble of forecasts, comprising 40 or 50 members, is issued using different model perturbations to incorporate uncertainty in the streamflow forecasts. This system is known as the ensemble streamflow prediction (ESP) system.
For assessing the effects of different sources of uncertainty on the model outputs, a systematic uncertainty analysis is usually carried out in hydrological modelling. A systematic uncertainty analysis consists of several steps such as, classification of uncertainty sources, importance assessment, quantification of uncertainty sources, uncertainty propagation through the model and, finally, communication of the uncertainty to the end users. After the identification of the main sources of uncertainty (Ewen *et al.*, 2006), these sources must be classified. There are many different approaches for source classification. For instance, Walker *et al.* (2003) classified uncertainty as originating from model context, input, model structure and parameters.
Other studies distinguished the uncertainty sources in observations, instruments and in the context of the problem, expert judgment and indicators (Janssen *et al.*, 2005; Van der Sluijs *et al.*, 2005; Warmink *et al.*, 2011). It has been commonly accepted that model inputs, parameters, initial conditions and structure are the major sources of uncertainty in conceptual hydrological models (Refsgaard *et al.*, 2006; Zappa *et al.*, 2011). Understanding the relative contributions of these sources to the total low flow forecast uncertainty and to the quality of forecasts can assist in the future development of ensemble forecasting systems for different lead times. This thesis focuses on two lead times, i.e. 10 and 90 days, as both medium-range and seasonal forecasts are important

in general for all river functions, river navigation in particular. Previous studies on the River Rhine focused either on only flood forecast models or on simulation models used for low flows, however no study to medium-range and seasonal forecasting of low flows is known to the author.

1.4 Climate change and low flows

Understanding low flow mechanisms and improved meteorological forecasts lead to better low flow forecasts. However, the long term trends in discharge series and inter-annual variations are also crucial, as long term changes in the Rhine discharge can have strong economical and societal impacts. For example, there is a growing concern that the occurrence of low flows will intensify due to climate change (Grabs *et al.*, 1997; Middelkoop *et al.*, 2001; Huang *et al.*, 2013). Intensified low flows will increase the cost of energy production and navigation in the River Rhine. The assessment of climate change impacts on hydrological catchment response is based on predicted meteorological variables like precipitation and temperature by climate models. Currently available climate change projections are mainly based on the outputs of general circulation models (GCMs) and, additionally, the outputs of regional climate models (RCMs) with a higher spatial resolution than GCMs. However, it is obvious that regional climate change projections based on these climate model outputs are highly uncertain due to unknown future greenhouse gas emissions and the simplified representation of processes in both RCMs and GCMs (Horton *et al.*, 2006). The impacts of climate change on the magnitude of simulated low flows have been previously assessed in different studies, e.g. the Rhineblick project (Görgen *et al.*, 2010). However, in this thesis, I focus not so much on the magnitude of low flows but on the seasonality properties of low flows (regime, timing and persistence of timing).

1.5 Research objectives and questions

The objective of this thesis is to explore low flow mechanisms, develop forecast methods and assess climate change impacts on low flows by identifying low flow indicators and their dominant temporal scales, analysing the effects of different uncertainty sources on low flow forecasts for different lead times, and comparing low flows for current and future climate conditions. For that, the following research questions are formulated:

Q1. What are the dominant low flow indicators and temporal scales in the Rhine Basin?
Q2. What is the effect of uncertainty originating from model inputs, parameters and initial conditions on 10 day ensemble low flow forecasts?
Q3. What is the effect of ensemble seasonal meteorological forecasts on the skill of seasonal ensemble low flow forecasts?
Q4. What is the impact of climate change on the seasonality of low flows in the Rhine basin?

1.6 Research approach and outline

The above mentioned four research questions are addressed in chapters 2, 3, 4 and 5 respectively. Figure 1.2 shows the setting of this thesis.
In chapter 2, the assessment of dominant low flow indicators will be based on a correlation analysis between low flow indicators and observed low flows at different lag times and different temporal resolutions of low flow indicators and observed low flows. These low flow indicators and their appropriate temporal resolutions will then be used in chapter 3 to select the appropriate models from available hydrological models and in chapter 4 to select two conceptual models and to develop two data-driven models. In chapter 3, the effect of major uncertainty sources, e.g. input, parameter and initial condition uncertainty, on the skill of 10 day low flow forecasts issued by two hydrological models are assessed using the forecast probabilities and ranges derived from ensemble forecast results. In chapter 4, the effect of seasonal meteorological forecasts on low flow forecasts is assessed using four hydrological models. The effects of different initial conditions and different model structures are also addressed. In chapter 5, the effect of different climate scenarios on low flows is evaluated based on a simulation approach using the outputs of an ensemble of climate models to drive a hydrological model. Three different seasonality indices are used to reflect the discharge regime, timing and variability in timing of low flow events respectively.

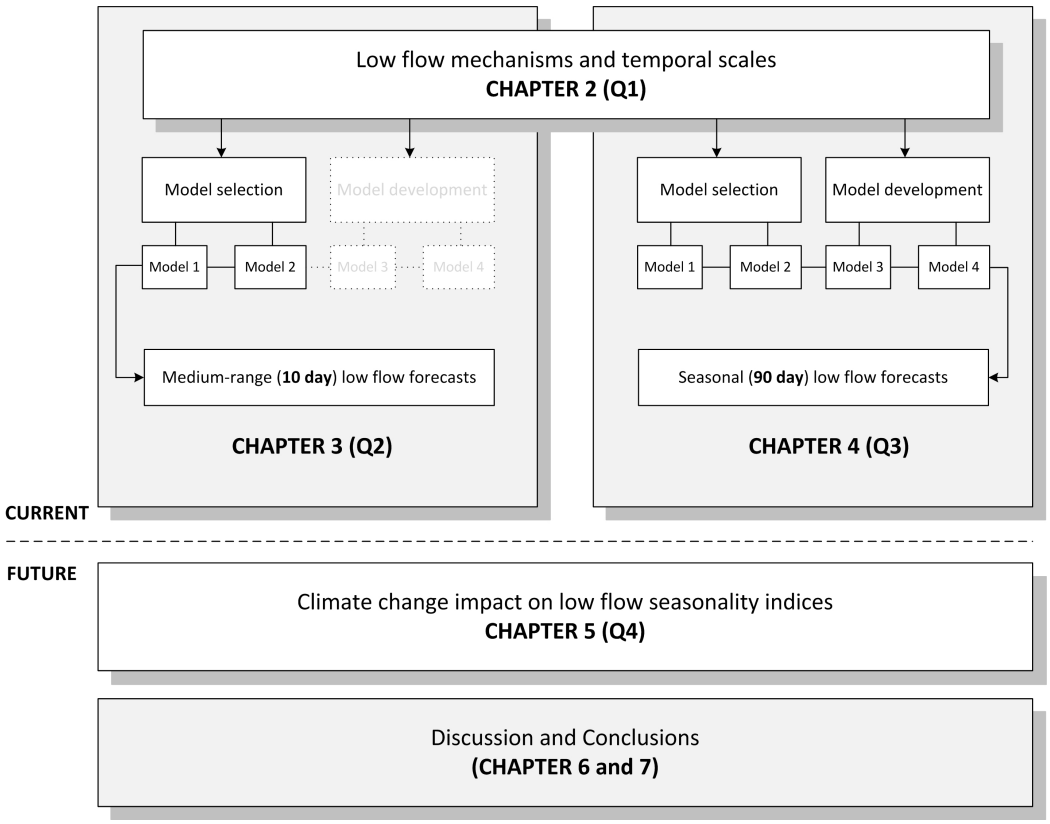


Figure 1.2: Setting of the thesis. Four research questions (Q1-Q4) are answered in chapters 2-5, and discussed and summarised in chapter 6 and 7 respectively.

This thesis focuses on the effects of the hydrological model and its inputs, the use of different GCMs and RCMs and the use of different emission scenarios. Finally, in chapters 6 and 7, the overall results from the preceding four chapters are discussed and the answers to the research questions are formulated.

Chapter 2

Identification of appropriate lags and temporal resolutions for low flow indicators

Abstract

The aim of this paper is to assess the relative importance of low flow indicators for the River Rhine and to identify their appropriate temporal lag and resolution. First, the Rhine basin is subdivided into seven sub-basins. By considering the dominant processes in the sub-basins, five low flow indicators were selected: precipitation, potential evapotranspiration, groundwater storage, snow storage and lake storage. Correlation analysis was then carried out to determine the relationship between observed low flows and pre-selected indicators with varying lags (days) and temporal resolutions (from one day to seven months). The results show that the most important low flow indicators in the Alpine sub-basins for forecasts with a lead time of 14 days are potential evapotranspiration with a large lag and temporal resolution, and lake levels with a small lag and temporal resolution. In the other sub-basins groundwater levels with a small lag and temporal resolution are important in addition to potential evapotranspiration with a large lag and temporal resolution. The picture is slightly different for forecasts with a lead time of 90 days. The snow storage in the Alpine sub-basins and the precipitation in the other sub-basins also become relevant for low flows. Consequently, the most important low flow indicators in the Alpine sub-basins for forecasts with a lead time of 90 days are potential evapotranspiration with a large lag and temporal resolution, lake levels with a small lag and temporal resolution and snow storage with a small lag and large temporal resolution. The resultant correlation maps provide appropriate lags and temporal resolutions for indicators to forecast low flows in the River Rhine with different lead times.

This chapter is based on the paper: Demirel MC, Booij MJ, Hoekstra AY, (2013). Identification of appropriate lags and temporal resolutions for low flow indicators in the River Rhine to forecast low flows with different lead times. *Hydrological Processes*, **27** (19): 2742-2758, doi: 10.1002/hyp.9402.

2.1 Introduction

Forecasting low flows several weeks or months in advance can benefit the management of water resources, river navigation and cooling water supply for the energy sector, particularly in Europe, where heavily industrialised cities are usually located along rivers. A two week forecast is often useful for the freight shipment sector whereas a longer lead time forecast like three month low flow forecast is a crucial reference for contingency plans of the energy sector. The water authorities can then make operational decisions on river traffic (e.g. maximum load allowance for ships), or decisions on reducing energy production because of a low cooling water supply.

Many different hydrological models exist which describe the transformation of rainfall to runoff at different spatial and temporal scales (Anderson *et al.*, 2004; Hannaford *et al.*, 2011; Hattermann *et al.*, 2004). Statistical models have been used to estimate low flows (Ouarda *et al.*, 2008) and conceptual models were applied to long-term low flow forecasting in France (Perrin *et al.*, 2002). To the best of our knowledge, none of the previous studies used conceptual and data-driven models to forecast low flows in the River Rhine. Selection of appropriate spatial and temporal scales is important as it affects the required input data, the processes that can be well presented, the scenarios that can be analysed, and usefulness of the resulting forecasts (Dumont *et al.*, 2008). However, the selection of a particular spatio-temporal scale in the model is usually not well reported, making the appropriateness of a chosen scale difficult to judge. Furthermore, modellers often have no clear criteria for selecting these scales for forecasting low flows.

This study focuses on identifying appropriate temporal scales (defined here as resolution) of dominant low flow processes (defined here as indicators). Identification of appropriate spatial scales is beyond the scope of this study. We present a framework for selecting the appropriate lag between the indicator and observed low flows and appropriate temporal resolutions of the indicators to include in the model or for selecting a suitable model for low flows. The lag provides information on the response time of the basin, including concentration time and travel time, while the temporal resolution gives information on the scale of the water volume entering or leaving the system.

Different processes can be dominant for different lead times (Haltas and Kavvas, 2011; Klemeš, 1983). There have been studies in which the hydrological processes leading to low flows and the relationship between low flows and drainage area were assessed (Burn *et al.*, 2008; Khaliq *et al.*, 2008; Ouarda *et al.*, 2008; Spence *et al.*, 2008). Most of these studies focused on low flows in Canadian rivers.

The river discharge during a low flow period mainly originates from groundwater storage and the outflow follows a characteristic recession curve (Schneider, 2008). Basin characteristics such as geology, soil type, topography, vegetation, hydraulic conductivity and extent of the aquifer determine the magnitude and timing of groundwater

discharge to streams (Allen *et al.*, 2010; Burn *et al.*, 2008; Hattermann *et al.*, 2004). Apart from that, the release from other large storages controlled by gravity, such as large lakes, snow storage and glaciers, can be important in sustaining low flows (Suweis *et al.*, 2010; Tallaksen and Van Lanen, 2004).

Low flows may occur in any season, mainly due to the lack of water input into a basin over a long period. This can be a dry period with a climatic water deficit (summer low flows) or a period with temperatures below zero, when the storage of precipitation is in the form of snow (winter low flows). Low flow is, therefore, defined as a seasonal phenomenon and an integral phase of the discharge cycle (Smakhtin, 2001; Warmink *et al.*, 2010).

Several studies have been carried out to analyse characteristics of low flows. Booij and De Wit (2010) analysed the relationship between the annual discharge deficit resulting in low flows and the annual minimum spatially-averaged precipitation at different temporal scales for the River Meuse in France and Belgium. Their study showed that the relationship becomes more significant at larger temporal scales. However, the central date of occurrence of spatially-averaged annual minimum precipitation did not show any relationship in time, while the annual discharge deficit mostly is observed in the period August-October. This is obviously because the annual cycle of evapotranspiration dominates the discharge deficit and the occurrence of low flows. Gudmundsson *et al.* (2011) analysed a pan-European dataset of 615 streamflow records, summarised as time series of annual streamflow percentiles. They revealed that under dry conditions the catchment response is more complex, as it depends on storage characteristics. Tallaksen *et al.* (2009) showed the importance of processes in modifying the drought signal in both time and space for the Pang catchment, UK. Their study disclosed that meteorological droughts frequently cover the whole catchment and last for a relatively short period (1-2 months). Moreover, hydrological droughts (e.g. groundwater drought) cover smaller areas and last longer (4-5 months) than meteorological droughts. Yue and Wang (2004) showed that low flows in Canadian rivers generally exhibit simple scaling behaviour and the drainage area alone explains most of the variability in the statistical properties of low flows.

In this study the main meteorological drivers in the River Rhine (precipitation and potential evapotranspiration) and the aforementioned storages (groundwater, lakes and snow) are defined as low flow indicators. It should be noted that, these relevant low flow indicators were not arbitrarily selected, but are based on previous reports (e.g. Belz and Frauenfelber-Kääb, 2007; Hurkmans *et al.*, 2008; Hurkmans *et al.*, 2010). We use the term 'indicator' rather than 'process' throughout the study, since not all the preselected indicators correspond to a hydrological process (e.g. lakes). The indicators usually act at different scales in the basin. While large storages are dominant at very large spatiotemporal scales, other indicators can be well described in small scales. However, these indicators can be modelled in an appropriate model scale.

Here 'appropriateness' is defined as a level between complex and simple for a model and its inputs. An appropriate model should then give adequate results through the use of appropriate input scales and a corresponding appropriate model scale. This is because the dominant processes are considered at their appropriate spatial and temporal scales. Consequently the appropriate model scale is estimated by integrating all input scales (Booij, 2003).

A model appropriateness procedure has been developed by Booij (2003). It has been cited in other studies as 'getting the right answers for the right reasons' (Kirchner, 2006). The procedure includes identification of dominant processes, appropriate scales and associated appropriate process formulations. In his study, Booij (2003) explored appropriate spatial scales for precipitation, elevation, soil and land use in a large river basin by using different relationships between scales and variable statistics and outputs for river basin modelling purposes. A framework was used to integrate the identified scales into an appropriate model scale of about 10 km with a corresponding temporal scale of 1 day. This result can drastically reduce the size of input data and model complexity. Booij (2002a) examined the effects of different spatial and temporal precipitation and hydrological model scales on extreme river flows. A spatial scale of 40 km was estimated to be appropriate for precipitation input into the model (Booij, 2002a). This agrees with the appropriate precipitation scale of about 20 km as assessed by Booij (2002b) for the same river basin.

To the best of our knowledge, none of the previous studies focused on quantifying the appropriate lag and temporal resolution of low flow indicators to forecast low flows. This is particularly important for modellers to develop a basin specific data-driven model or to select a model from existing hydrological models that is appropriate for low flow forecasting and for river managers. The importance of this research topic of identification of space-time patterns to improve large scale hydrological predictions by considering uncertainties was also emphasised in a recent special issue of a hydrological journal (Cloke and Hannah, 2011; Gudmundsson *et al.*, 2011; Hannaford *et al.*, 2011). Understanding dominant low flow indicators and their relationship to low flows in the River Rhine will help to improve the analysis of river behaviour during low flows and better identify the components of a low flow forecast model. This study is a first attempt that includes the most important low flow indicators and analyses their linkage to observed low flows for different temporal lags and resolutions.

In the following, the study area and data are described (Section 2.2). This is followed by the general structure of our lag and temporal resolution framework and details of the different basin averaged indices which form the basis of the correlation analysis (Section 2.3). Subsequently, the results are discussed (Section 2.4) and finally conclusions are drawn (Section 2.5).

2.2 Study area and data

2.2.1 Study area

The Rhine is the busiest inland waterway in Western Europe. It connects cities with heavy industry to the world market via Rotterdam harbour in the Netherlands. The river originates at the outlet of Lake Toma in the Piz Badus (2,928 m), Swiss Alps (Figure 2.1). It flows along a 1,233 km long course before discharging into the North Sea.

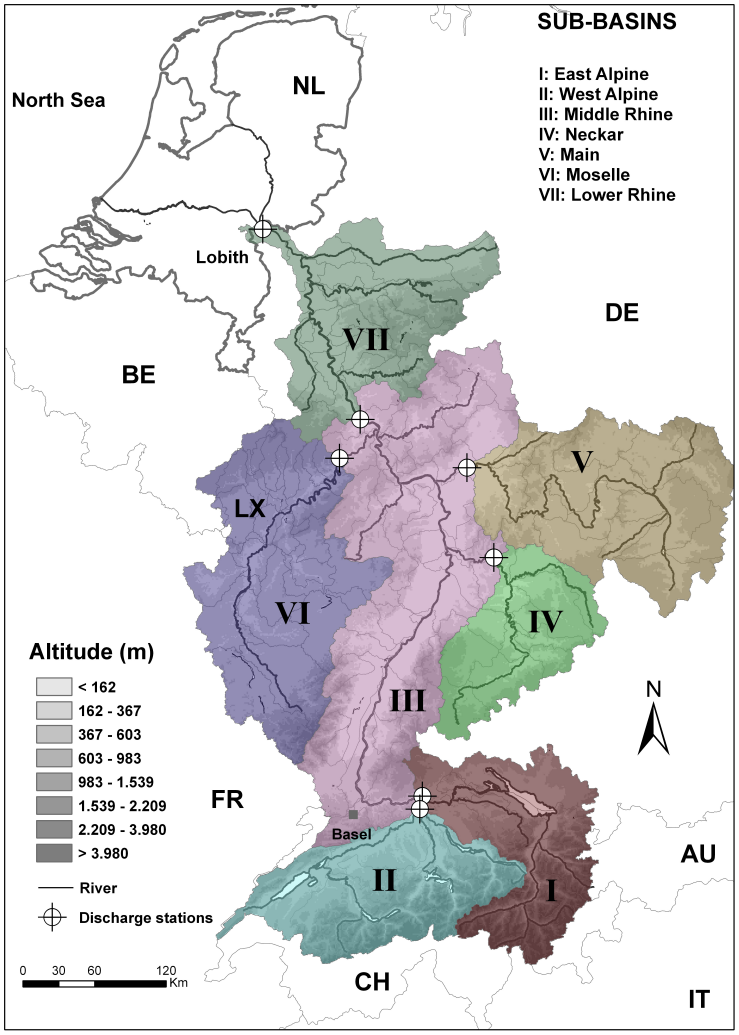


Figure 2.1: Schematisation of the seven major sub-basins of the River Rhine upstream of Lobith on the German-Dutch border.

The surface area of the Rhine basin is approximately 185,300 km^2, covering major parts of Switzerland, Germany, Luxembourg, France and the Netherlands, and has nearly 60 million inhabitants (Huisman *et al.*, 2000). Furthermore, more than 60 per cent of the Dutch fresh surface water comes from the Rhine (Middelkoop and Van Haselen, 1999).

The topography of the basin varies from 4,000 meters in the Alps to six meters below sea level in the Netherlands. The average discharge before Lake Constance, located in the East Alpine sub-basin (Figure 2.1), is approximately 1000 m^3/s. This is an indication of the Alps' contribution to the total discharge. It then increases to up to 2,300 m^3/s on average at the German-Dutch border (Lobith). The discharge regime of the Rhine at Basel is mainly dominated by melt-water from snow and around 150 Alpine glaciers, including those in the Gotthard massif (Figure 2.1). Here, the water is generally at its highest level in spring and early summer, when the snow melts (Figure 2.2). More than 70 per cent of the summer flow at Lobith (Figure 2.1) originates from the Alps, whereas only about 30 per cent of the winter flow comes from the Alps (Middelkoop and Van Haselen, 1999), because winter precipitation is stored as snow in this part of the basin until it melts in late spring (Figure 2.2).

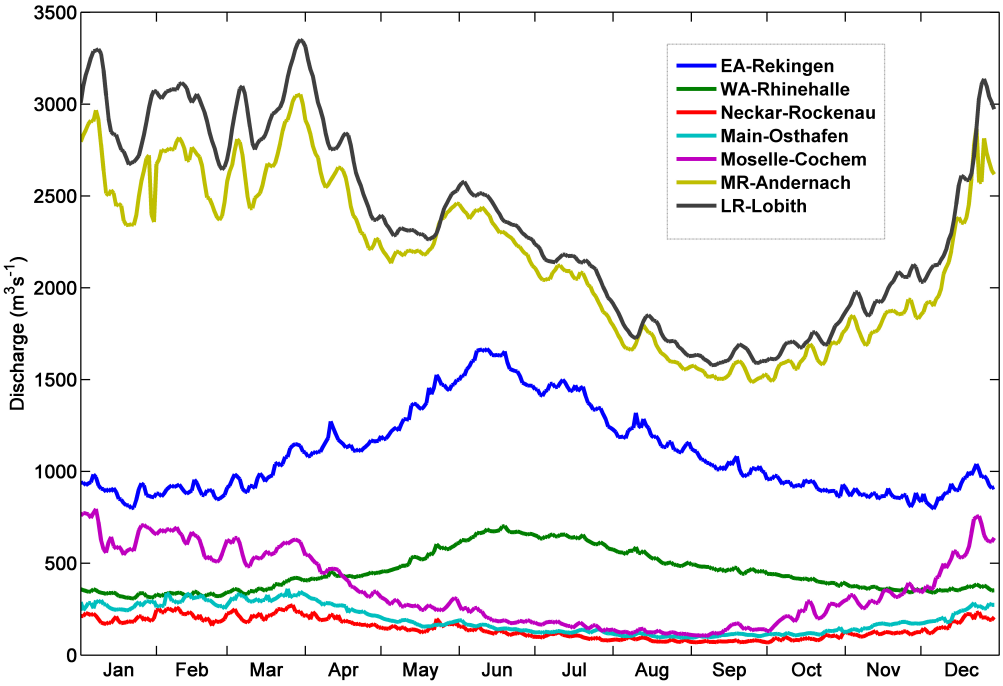


Figure 2.2: Long-term average discharge recorded at the outlets of the seven major sub-basins.

The tributaries Neckar, Main and Mosel join the Rhine in Germany. These tributaries carry vast volumes of water, mainly in winter, when there is intense rainfall and negligible evapotranspiration. Figure 2.2 shows that the snow-melt water and rain ensure that the River Rhine is navigable all year long. In addition, the difference between the minimum and the maximum flow in the Rhine is only a factor of 20, while it is a factor of around 150 in the neighbouring rainfed Meuse (Middelkoop and Van Haselen, 1999). The hydrology of the Rhine basin has been modelled with a conceptual model (HBV) using 134 catchments (Te Linde *et al.*, 2008; Te Linde *et al.*, 2010). As we are interested

in major storages in the basin and their relationship to low flows, the Rhine basin is analysed on two spatial scales that of 134 catchments and that seven sub-basins. The latter level of discretisation is chosen on the basis of basin characteristics such as topography and geology. The upper Rhine is divided into West Alpine (WA) and East Alpine (EA), to distinguish Lake Constance's impact on low flows in addition to that of snow storage in the Alps. The EA and WA sub-basins cover approximately 34,000 km^2, with a maximum altitude of 4,000 meters. There is great variability in altitude and subsequently in slopes, showing the heterogeneous topography in the WA and EA sub-basins (Table 2.1).

The glaciers cover about 400 km^2 of this mountainous area. The EA sub-basin stretches from the beginning of the river to Lake Constance (Alpenrhein), while the WA sub-basin covers the River Aare basin with an outlet in the Untersiggenthal before Basel (Hochrhein). The Untersiggenthal discharge station on the Aare covering 71 per cent of the total surface area of the WA sub-basin is chosen to represent WA sub-basin discharge. This leads to a clear picture of two Alpine sub-basins with totally independent discharge regimes. Moreover, estimating the annually-generated discharges become easier as it is the fraction of the independent discharge regimes to the sub-basins' surface areas (see Table 2.1). Note that the independent discharge is estimated by subtracting all inlet discharges from the outlet discharge.

Table 2.1: Spatial characteristics of the seven major sub-basins shown in Figure 2.1

Sub-basin	**Area** [km^2]	**Annually generated discharge** [mm]	**Altitude [m]**		
			Range	**Mean**	**Std. dev.**
East Alpine (I)	16051	890	143-3270	1250	761
West Alpine (II)	17679	1021	252-3980	967	603
Middle Rhine (III)	41473	344	67-1340	309	205
Neckar (IV)	12616	363	90-970	432	156
Main (V)	24833	244	83-939	344	115
Moselle (VI)	27262	410	59-1326	340	131
Lower Rhine (VII)	20174	273	5-779	237	150

In previous Rhine studies, Rheinfelden discharge station near Basel was used for hydrological modelling purpose (Renner *et al.*, 2009). The middle Rhine is divided into four parts: the Neckar, Main and Moselle tributaries and the remaining main channel between Basel and Koblenz, which is called the Middle Rhine (MR). The seventh sub-basin, the Lower Rhine (LR), starts after Koblenz, where the Main and the Moselle flow into the Rhine. Previous studies focusing on the Rhine basin used a similar subdivision into seven sub-basins (e.g. Belz and Frauenfelber-Kääb, 2007).

2.2.2 Discharge data and pre-selected low flow indicators

We use different hydrological variables to carry out the correlation analyses (see Table 2.2). Daily precipitation and potential evapotranspiration data as spatially averaged for 134 catchments were obtained from the German Federal Institute of Hydrology (BfG) in Koblenz (Germany). The outlet discharges for the EA (station *#6935054* at Rekingen), WA (station *#6935300* at Untersiggenthal), Neckar (station *#6335600* at Rockenau), Main (station *#6335304* at Frankfurt Osthafen) and Mosel (station *#6336050* at Cochem) and independent discharges for the MR and LR were used in the correlation assessment.

Table 2.2: Data characterisation/availability

Data	**Index**	**Spatial resolution**	**Number of stations/ sub-basins**	**Period**	**Temporal resolution**	**Source**
Discharge	Q	Point	7	1974-2008	Daily	GRDC in Koblenz
Precipitation	P	Sub-basins	134	1951-2006	Daily	BfG in Koblenz
Evapo-transpiration	PET	Sub-basins	134	1950-2006	Daily	BfG in Koblenz
Groundwater levels	G	Point	1402	1986-2009	Weekly, monthly	German states and BAFU.ch
Snow	S	Point	40	1978-2008	Daily, monthly	SLF.ch
Lake levels	L	Point	11	1978-2008	Daily	BAFU.ch

Daily lake level (L) and daily fresh snow depth data (S) are also included in the correlation analysis. The groundwater data (G) comprise time series of levels measured at 1,404 stations throughout the Rhine basin with different data lengths and temporal resolutions (Table 2.2). Also, the stations were not distributed evenly throughout the basin (Figure 2.3). For example, the groundwater levels in Bavaria in Germany are represented by a network of 661 wells, while only 99 wells were available in Switzerland. Therefore, pre-processing of groundwater level data was required. The data used are summarised in Table 2.2.

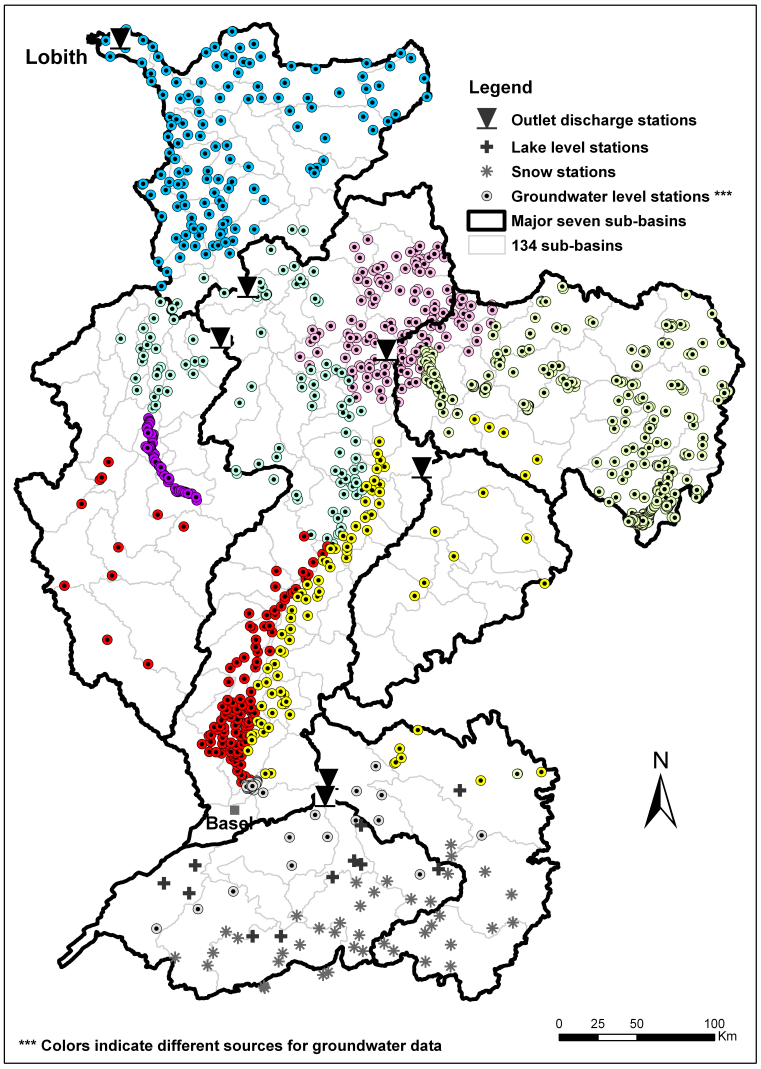


Figure 2.3: Location of discharge, lake level, snow depth and groundwater level stations.

2.3 Methodology

2.3.1 Overview

This study employs a five step framework to assess the relative importance of low flow indicators for the River Rhine and to identify their appropriate temporal lag and resolution (Figure 2.4). The low flow indicators were selected and analysed on the scale of seven sub-basins.

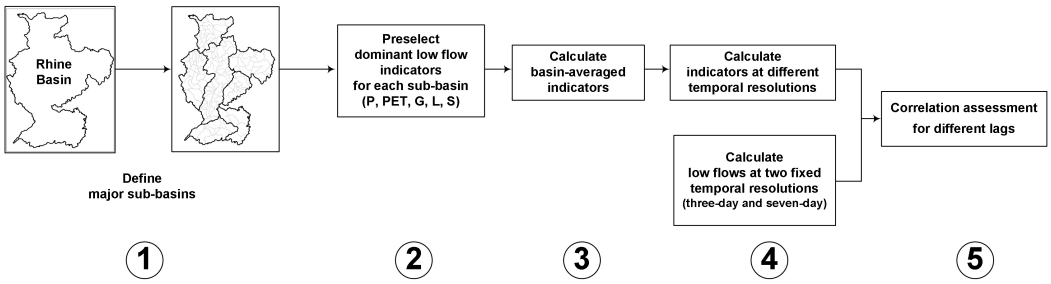


Figure 2.4: General framework for assessing the relative importance of low flow indicators for the Rhine and identifying their appropriate temporal lag and resolution.

We define seven sub-basins by spatially aggregating the 134 catchments according to similar hydrological characteristics and based on previous studies (e.g. Belz and Frauenfelber-Kääb, 2007; Hurkmans *et al.*, 2008). These sub-basins are the large tributaries except for the Middle and Lower Rhine sub-basins that are the remaining parts. Note that 134 catchments in the Rhine basin, shown in Figure 2.4, had already been identified and used in other Rhine studies for modelling purposes (e.g. Reggiani *et al.*, 2009; Renner *et al.*, 2009; Te Linde *et al.*, 2008). The scale of seven sub-basins is assumed to be sufficient to understand the pre-selected indicators and their relationship to low flows at the basin outlets. This raises the question of why the assessment of low flow indicators was not applied on the scale of 134 catchments. Identification of low flow indicators can be very complicated due to interacting processes on small scales, since the low flows are, in general, sustained by baseflow originating from a groundwater reservoir, which generally has a much greater spatial scale than the typical scale of each of the 134 catchments. Therefore the relationship between low flows and indicators on the scale of 134 catchments can lead to misinterpretation of existing storages in the Rhine basin.

In step 2 we selected precipitation (P), potential evapotranspiration (PET) and groundwater storage (G) as low flow indicators in all seven sub-basins. Additionally, snow storages (S) and lake storages (L) are considered as indicators, but only in the two upstream sub-basins (EA and WA) (Scherrer and Appenzeller, 2006; Tague and Grant, 2009; Verbunt *et al.*, 2003; Zappa and Kan, 2007).

In step 3, data at the point or catchment (134 catchments) scale were aggregated to the scale of seven sub-basins. This is done so as to have one basin-averaged time series for each indicator.

We used a classical standardisation method to avoid effects of spatial heterogeneity in the data. The standardised data have a zero mean and a standard deviation of one. The standardised series are obtained by subtracting the mean from each element and

dividing it by the standard deviation of the original series (Eq. (2.1)).

$$Z = \frac{X - \mu_x}{\sigma_x} \tag{2.1}$$

where Z is the standardised time series, X the original time series (observed data), μ_x the mean of the original time series and σ_x the standard deviation of the original time series.

In step 4, the temporal resolution of the indicators is varied between 1 day and 336 days (i.e. 1, 3, 7, 14, 21, 28, 56, 84, 112, 140, 168, 196, 224, 252, 280, 308 and 336 days). The lag between indicator and low flows is varied between 0 and 210 days (i.e. 0, 1, 3, 7, 14, 28, 42, 56, 70, 84, 98, 112, 126, 140, 154, 168, 182, 196 and 210 days). The maximum temporal resolution of 336 days and maximum lag of 210 days are assumed to be sufficient, as these ranges allow us to scan a long period of more than eighteen months. To the best of our knowledge, only large-scale atmospheric indicators can be significant over longer periods than those examined here.

Following the relevant reports by the principal authority for low water levels in the Netherlands i.e. Dutch National Coordinating Committee on Water Distribution (LCW), we selected the exceedence probability 75 per cent (Q75) as a threshold for the definition of low flows. Low flows at this threshold are still affecting the previously described river functions and the number of days with low flows is sufficient to calibrate a forecast model. Daily discharges observed at Lobith station equal to or below this threshold are used to construct the reference low flow series. The low flow occurrence days at Lobith are then used to construct low flow series for all seven major sub-basins, since our objective is to forecast low flows at Lobith. These seven Lobith-based low flow series are temporally aggregated on the scale of three and seven days for two lead times, namely 14 and 90 days respectively. The operational value of daily-averaged and three-day averaged low flow forecasts is about the same for the navigation and energy sectors. Moreover, forecasting three-day averaged flow is assumed to be reasonable with a lead time of 14 days (De Bruijn and Passchier, 2006). Similarly, the temporal resolution is increased to seven days when the lead time is 90 days. Correlation coefficients between low flows and indicators are estimated in step 5, to assess the relative importance of low flow indicators and to explore for which lag and temporal resolution the correlation is most significant.

2.3.2 Definition of storage indices

Daily standard groundwater storage index.

Groundwater levels from numerous stations in the Rhine basin were included in this study. The individual groundwater stations' measurements, shown in Figure 2.3, were

aggregated to the scale of seven sub-basins using standardised data (Eq. (2.1)). Therefore they are hereafter called standard groundwater storage indices.
It should be noted that the temporal resolution of the groundwater data is either weekly or monthly. For example, the groundwater data series from Bavaria are available with a monthly resolution, while the remaining part of the dataset has a weekly resolution. The differences in resolution are eliminated, firstly, by aggregating weekly into monthly data. This is done for every sub-basin where both weekly and monthly data are found. Particularly, in the west of the Main sub-basin there are catchments with weekly data from Hessen and monthly data from Bavaria (see Figure 2.3). After the aggregation of weekly data into monthly data the stations in each of the 134 catchments are standardised and then merged using arithmetic means, to estimate standard monthly storage index series. Disassembling these monthly storage index series to the daily scale is accomplished through linear interpolation. Since groundwater storage oscillations are generally very slow, a linear character is assumed for the data. This assumption was tested using a temporal correlogram. We examined the temporal variability of groundwater data in one of the seven major sub-basins. We selected nine stations in the western part of the Main sub-basin, where both weekly and monthly data are found. The temporal correlogram of stations with highly variable groundwater levels showed that they have long correlation lengths, varying from nine to 11 weeks. The appropriate temporal scale can be defined as 25 per cent of the correlation length, accepting a bias of 10 per cent (Booij, 2003). Therefore, a scale of two weeks was estimated as an appropriate temporal scale. For that reason the linear interpolation method was applied, assuming that the measurements were taken in the middle of each month. The standard lake level index and the standard snow storage index were estimated only for the EA and WA sub-basins.

Daily standard snow storage index.

Daily fresh snow depth data in the EA and WA sub-basins were used to estimate the standard snow storage index. Some of the small Alpine basins have more than one snow-monitoring station. Therefore firstly the data from the stations within the same sub-basin were standardised using zero mean and a standard deviation of one and secondly these stations were merged to the scale of 134 catchments using the arithmetic mean.

Daily standard lake storage index.

There are several large lakes in the EA and WA sub-basins. We selected Lake Constance as representative of the lake storage in the EA sub-basin. Daily lake levels from ten lakes in the WA sub-basin were used to calculate the daily lake storage index for the WA sub-basin. The data from these lakes were standardised using zero mean and a standard

deviation of one. The stations were then merged using the arithmetic mean. For the EA sub-basin, the correlations between the observed lake levels of Lake Constance and low flows for different lags and temporal resolutions were estimated directly.

2.3.3 Basin averaging of pre-selected low flow indicators from the scale of 134 catchments to that of seven sub-basins

The sub-basin averaged standard daily indices, i.e. P, PET, G and S, were aggregated to the scale of seven sub-basins using areal weighting. This refers to the fraction of a sub-basin area on the scale of 134 catchments relative to the sub-basin area on the scale of seven sub-basins. The basin-averaged standard groundwater storage index was estimated by using Eq. (2.2).

$$G_{index,j} = \sum g_i \frac{A_i}{A_j} \tag{2.2}$$

where i=1, 2..., 134 catchments, j=1, 2..., seven major sub-basins, g_i is the standardised daily groundwater level, A_i the area of each of the 134 catchments and A_j the total area of each of the seven major sub-basins. The basin-averaged standard snow storage index was estimated by using Eq. (2.3).

$$S_{index,j} = \sum s_i \frac{A_i}{A_j} \tag{2.3}$$

where j=1, 2 Alpine sub-basins, s_i is the standardised daily fresh snow height series observed in these sub-basins and A_j is the total area of the EA or WA sub-basin.

The basin-averaged standard lake storage index in the WA sub-basin was estimated by using Eq. (2.4).

$$L_{index,j} = \sum l_i \frac{A_i}{A_j} \tag{2.4}$$

where j= 2 indicating WA sub-basin, l_i is the standardised daily lake level series observed in the WA sub-basin and A_j is the total area of the WA sub-basin. For some of the 134 catchments no G, S or L data were available and these catchments were not included in the basin averaging. Finally, each pre-selected low flow indicator is represented by one series for each of the seven sub-basins. Most of these series are daily standard index series such as P, PET, G and S for the EA sub-basin, P, PET, G, S and L for the WA sub-basin and P, PET and G for the remaining five sub-basins.

2.3.4 Correlation assessment

In the last step of the framework, we used correlation analysis to screen potentially useful predictor-predictant relations for varying lags and temporal resolutions. The correlations were calculated for a lead time of 14 days and three-day temporally-averaged low flows and for a lead time of 90 days and seven-day temporally averaged low flows. The temporal resolution of the predictants (low flow indicators) varied between one day to 336 days, whereas the temporal resolution of the predictor (observed low flows) was either three day or seven-day moving average values (Figure 2.5). Note that the forecast lead times (i.e. 14 and 90 days) are added to the lag values in the correlation analysis.

Three different correlation coefficients between low flows and indicators are estimated. We are aware that the Pearson correlation coefficients are based on the assumption of stationary linear relationships between the low flow indicators and low flows (Steinschneider and Brown, 2011; Vicente-Serrano and López-Moreno, 2005). In practice, only slowly responding processes such as groundwater levels and lake and snow storages show linear behaviour (Wedgbrow *et al.*, 2002). Since all three correlation coefficients revealed similar results, only the Pearson correlation coefficients are presented.

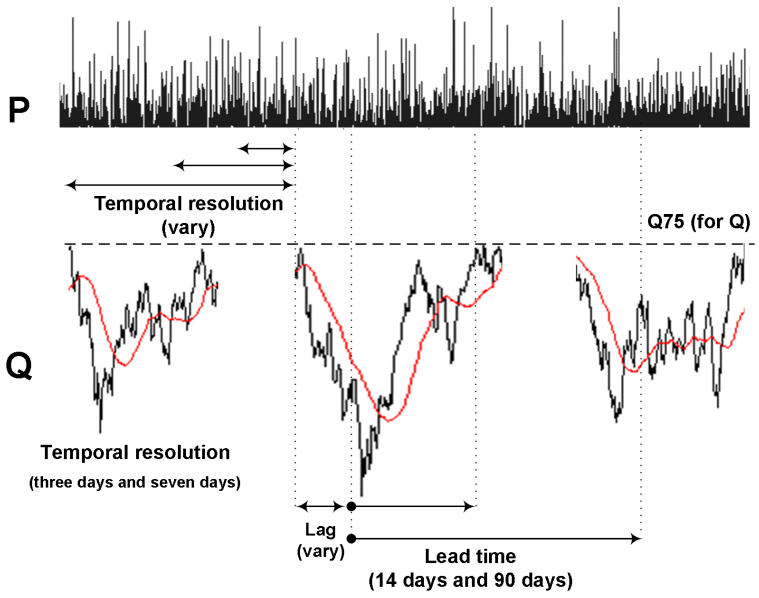


Figure 2.5: Conceptual diagram illustrating the correlation assessment of three-day and seven-day moving-averaged low flow data and precipitation data with varying temporal resolutions and lags for lead time of 14 and 90 days.

2.4 Results and discussion

2.4.1 Basin-averaged daily indicators

Figure 2.6 shows the basin-averaged low flow indicators in the East Alpine sub-basin for a three-year period. Because of the fact that the monitoring station at Lake Constance and the discharge station are close to each other, the fluctuations are very similar.

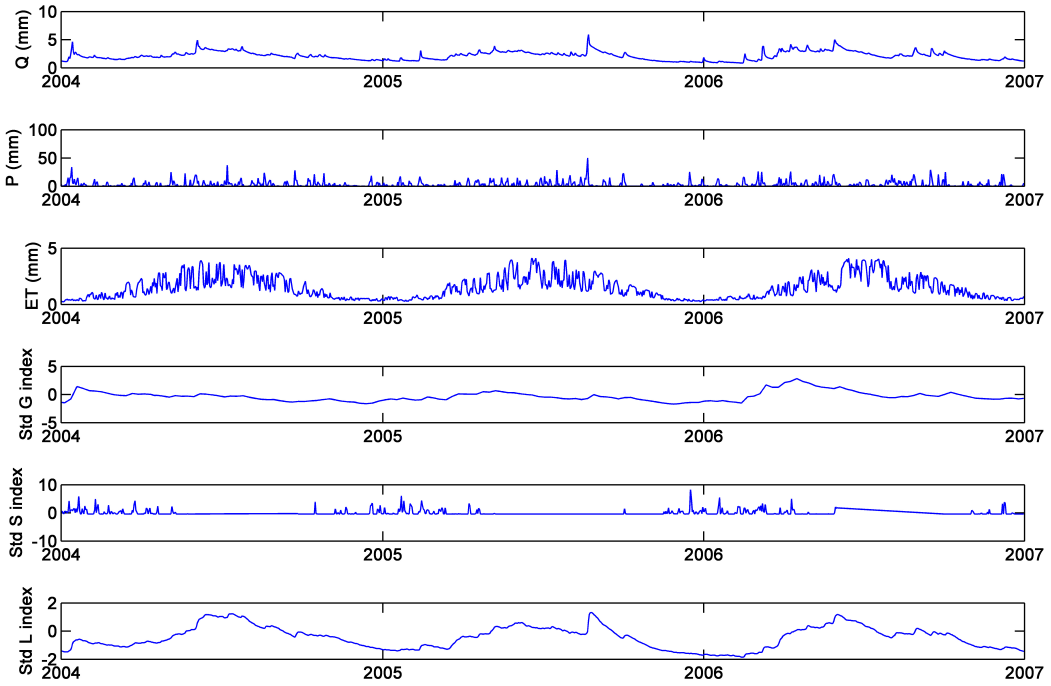


Figure 2.6: Observed daily time series of Q and pre-selected low flow indicators P, PET, G, S and L.

2.4.2 Correlation assessment

The correlation coefficients between low flows and indicators are presented by using colour maps. In these maps, the x axis shows the temporal resolution and the y axis the lag between indicators and low flows. The maximum ten percentile of the correlation coefficients in each colour map is indicated by two bars: a vertical bar for the range of lags and a horizontal bar for the range of temporal resolutions. These bars cross at the maximum correlation point, showing the appropriate lag and temporal resolution of an indicator. For clear visibility the crossing point is also highlighted by a circle. It is assumed that the ranges represent the uncertainty in the appropriate lag and temporal resolution. The correlation coefficients between PET and low flows are negative (i.e. blue colour in the maps). However, the absolute value of the maximum correlation for PET is used to compare the correlations. The most important features of these colour maps are discussed for each sub-basin, from upstream to downstream. Furthermore,

two major questions will be dealt with and discussed. (1) Why has one indicator higher correlations with low flows than other indicators? (2) Why are these lags and temporal resolutions appropriate?

Figure 2.7 shows the maximum correlation coefficients between indicators and low flows for all sub-basins. The ranking of the low flow indicators will determine the initial forecast model structures. It is obvious that the most important low flow indicators for forecasts with a lead time of 14 days in the Alpine sub-basins are potential evapotranspiration and lake levels, whereas in the other sub-basins in addition to potential evapotranspiration groundwater levels become important indicators (Figure 2.7a). For forecasts with a lead time of 90 days, potential evapotranspiration, lake levels and snow are the best predictors in the Alpine sub-basins. In the other sub-basins, as well as potential evapotranspiration precipitation or groundwater are important indicators for low flows.

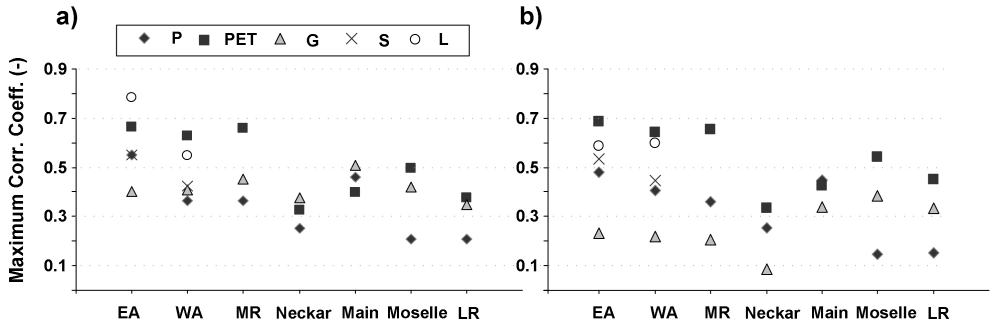


Figure 2.7: Maximum correlation coefficients between low flows and pre-selected indicators in the seven sub-basins: **a)** low flows with a temporal resolution of three days and a lead time of 14 days; **b)** low flows with a temporal resolution of seven days and a lead time of 90 days. Only for PET were absolute values of correlation coefficients used.

2.4.3 Appropriate lags and temporal resolutions

East Alpine

Figure 2.8 shows the Pearson correlation coefficients between observed low flows and pre-selected indicators with varying lags and temporal resolutions in the EA sub-basin. Of the five pre-selected low flow indicators, lake levels and potential evapotranspiration revealed the highest correlation coefficients for forecasts with a lead time of 14 days. In addition to these two snow storage is an important indicator for forecasts with a lead time of 90 days.

The proximity of Lake Constance to the discharge station at Rekingen must be the main reason for high correlations with low flows. The large lake storage in the EA sub-basin (55 km^3) can sustain river flows for several months. Because of the fact that the travel time between the outlet of Lake Constance and the discharge station is very short, small lags (zero days) and daily temporal resolution of lake levels are appropriate for low flow forecasts with lead times of 14 and 90 days. Therefore, the uncertainty around the appropriate lag and temporal resolution for the lake levels is very small.

The correlation maps mainly show two maximum correlation regions, as low flows have high seasonality. This can be clearly seen in the correlation figures for the lake levels. For example, Figure 2.8a shows the maximum correlation point at a lag of zero days and a temporal resolution of around one day for the lake levels, while Figure 2.8b shows the maximum correlation point at a lag of 210 days and a temporal resolution of around 140 days. Note that the lead times of 14 and 90 days were included in the lags shown in Figure 2.8a and b respectively. Therefore the maximum correlation point for lake levels in Figure 2.8b indicates that the lake levels in the preceding seven months with a temporal resolution of around five months are relevant for forecasts with a lead time of 90 days. This corresponds to a total lag of one year and can be explained by the annual hydrological cycle. Although this situation is justifiable from a mathematical point of view, such a large lag and temporal resolution are not physically meaningful and should be ignored. Consequently, small lags (zero days) and daily temporal resolution of the lake levels are appropriate for forecasts with lead times of 14 and 90 days, as changes in lake levels are seen directly observed in discharge levels.

Potential evapotranspiration is one of the main driving forces behind low flows, since it determines water loss to the atmosphere. If water storages sustaining river flows such as snow and groundwater are exhausted in a basin, with recurring water deficits, then it is very likely that low flows will be experienced in late summer. This is especially the case for the River Rhine, as most of the summer flow originates from the snow-dominated Alpine sub-basins.

Small lags (42 days) and large temporal resolutions (about nine months) for the *PET* index are appropriate for forecasts with lead times of 14 days. Moreover, small lags (zero days) and large temporal resolutions (around seven months) are appropriate for

forecasts with a lead time of 90 days. This shows that the *PET* averaged over the preceding summer and winter seasons is an important indicator for low flows at Lobith. However, the uncertainty around the appropriate lag and temporal resolution for *PET* is very large, due to the great climatic variability. The difference between the two lead times can be explained by the additional 76 days in the lag (Figure 2.8b).

The amount of snow-fall in the preceding winter period must also be a good indicator for the late summer low flows at Lobith. This is what we can see in Figure 2.8. The maximum correlation for the *S* index for forecasts with lead times of 14 days is found to be at a lag time of about four months and a temporal resolution of four months, while the maximum correlation for forecasts with lead times of 90 days is at a lag time of about one month and a temporal resolution of four months. Taking the forecast issue day as August 1, average fresh snow height during the December-March period is important for low flows at Lobith at the end of the summer. The differences found in the appropriate lag and temporal resolutions for the two lead times clearly confirm the effect of the additional lag of 76 days (Figure 2.8b).

Appropriate lags and temporal resolutions for other indicators are also given in the figures, as they will be tested in the forecasting phase. However, these indicators are less important and are not discussed here for reasons of brevity.

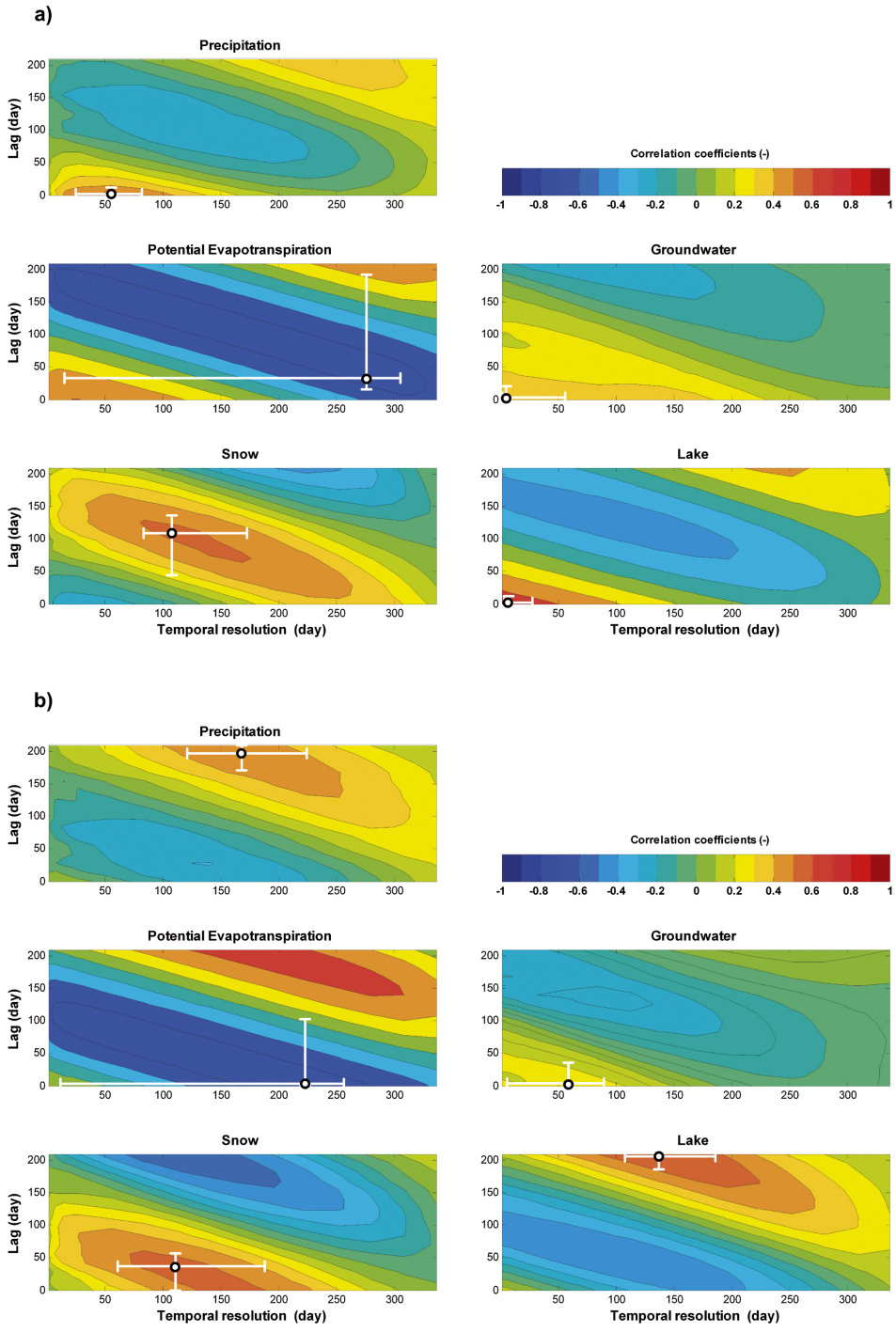


Figure 2.8: Cross correlation coefficients between low flows in the East Alpine sub-basin and pre-selected indicators as a function of lag and temporal resolution (days): **a)** low flows with a temporal resolution of three days and a lead time of 14 days; **b)** low flows with a temporal resolution of seven days and a lead time of 90 days.

West Alpine

The picture in the WA sub-basin is more or less the same as that in the EA sub-basin. Lake levels and potential evapotranspiration are the most important low flow indicators for forecasts with lead times of 14 and 90 days (Figure 2.9). However, the maximum correlation between lake levels and low flows in the WA sub-basin for forecasts with a lead time of 90 days is higher than that in EA. We assume that the increase in the maximum correlation for the L index in the WA sub-basin is not arbitrary.

The total lake storage in the WA sub-basin, comprised of ten lakes with varying storage capacities, is about 49 km^3. This is relatively less than the storage of Lake Constance (55 km^3). However, smaller lake storage should not be directly interpreted as an indication of shorter memory. Firstly, the annually-generated discharge in the WA sub-basin is higher than that in the EA sub-basin. Secondly, the ten lakes are distributed evenly through the WA sub-basin, whereas there is only one large lake (Lake Constance) in the EA sub-basin. The travel time from lake outlets to the discharge station in the WA sub-basin is longer than that for Lake Constance. Furthermore, the basin-averaged standard lake level index, an aggregated index series for the ten lakes, was used for the WA sub-basin. All this can cause an increase in the maximum correlation between lake levels and low flows for forecasts with a lead time of 90 days (Figure 2.9b).

The maximum correlations for L and PET have a similar magnitude and are significantly higher than for the other three indicators. Small lags (zero days) and daily temporal resolution for the L index and small lags (56 days) and large temporal resolutions (around seven months) for the PET index are appropriate for forecasts with lead times of 14 and 90 days. The large uncertainty around the appropriate lag and temporal resolution for the L index can be explained by the second high correlation region in the upper right part of the correlation map for forecasts with a lead time of 14 days (see Figure 2.9a). Therefore, the uncertainty range for the maximum correlation point of the L index should be from zero to 14 days on the y axis (lags) and from zero to one month on the x axis (temporal resolutions).

Another significant difference between the EA and WA sub-basins is that the snow storage in WA shows weaker correlations than that in EA for forecasts with lead times of 14 and 90 days. This may be because snow-melt water from the Alps is divided over several tributaries in the WA sub-basin, whereas in EA it is directly connected to the main river channel. The more immediate response of discharge to snow-melt in EA can explain the higher correlations between the S index and low flows in this sub-basin.

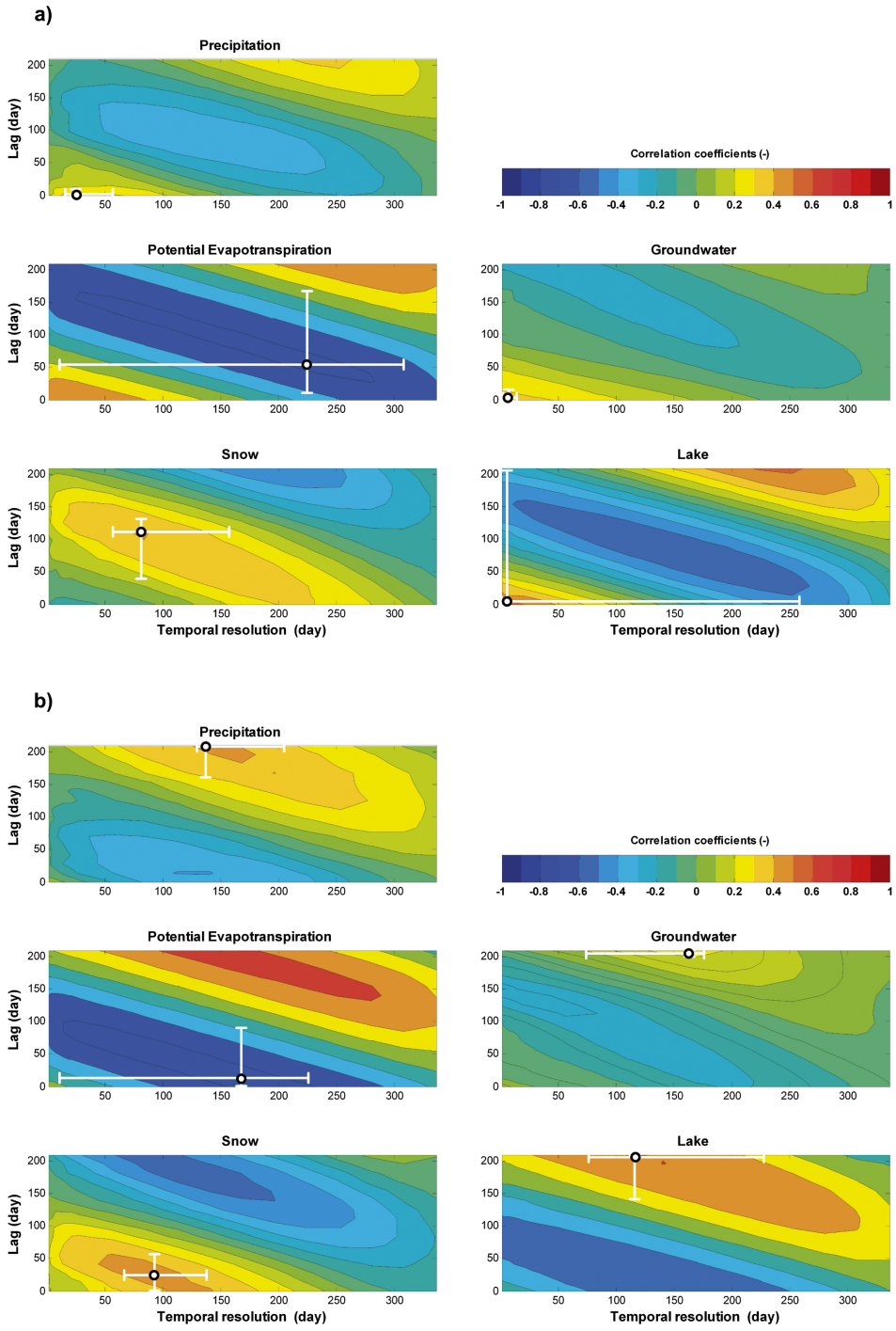


Figure 2.9: Cross correlation coefficients between low flows in the West Alpine sub-basin and pre-selected indicators as a function of lag and temporal resolution (days): **a)** low flows with a temporal resolution of three days and a lead time of 14 days; **b)** low flows with a temporal resolution of seven days and a lead time of 90 days.

Main

The general correlation patterns for the snow-dominated Alpine sub-basins and the downstream rainfed sub-basins are assumed to be different. This assumption is confirmed by comparing the correlation maps of the two groups of sub-basins. We only present the results for Main, as the correlation maps of the rainfed sub-basins are very similar (see Appendix). Figure 2.10 shows the correlations between low flows and indicators for the Main sub-basin. All pre-selected indicators have very similar maximum correlations for lead times of 14 and 90 days. The P index gives slightly higher correlations than the PET index for both lead times. Furthermore, the relative importance of the P index is also higher than that of the G index for forecasts with a lead time of 90 days, showing the importance of precipitation for this sub-basin.

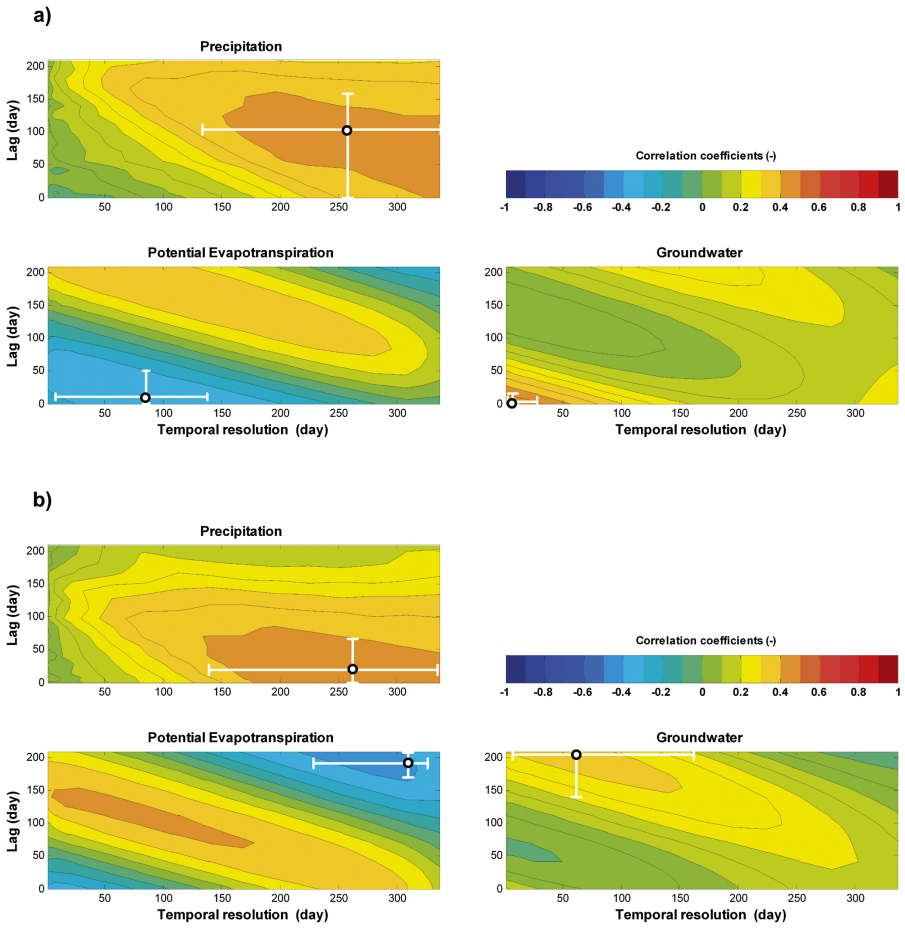


Figure 2.10: Cross correlation coefficients between low flows in the Main sub-basin and pre-selected indicators as a function of lag and temporal resolution (days): **a)** low flows with a temporal resolution of three days and a lead time of 14 days; **b)** low flows with a temporal resolution of seven days and a lead time of 90 days.

Large lags (around three months) and temporal resolutions (around eight months) are appropriate for the P index for forecasts with a lead time of 14 days. This result is in line with recent studies on the neighbouring rainfed Meuse basin, where low flows occur at the end of the summer and beginning of the autumn and depend on the amount of the preceding winter half year precipitation and also the preceding summer half year precipitation (Booij and De Wit, 2010; De Wit *et al.*, 2007). The uncertainty over the appropriate lag for the P index is high due to the uncertainty over rainfall events. The P index averaged over a period of five months or longer shows similar correlations with low flows in the Main sub-basin. Due to the rainfed characteristic of the sub-basin, low flow occurrence in the Main sub-basin is not as persistent as it is in the Alpine sub-basins. Therefore forecasting low flows can be more difficult than that for the sub-basins with high persistence.
Small lags (seven days) and large temporal resolutions (around three months) for the PET index are appropriate for forecasts with a lead time of 14 days. In other words, for a lead time of 14 days the potential evapotranspiration in the preceding 3 months, such as June, July and August, is important in issuing a low flow forecast in September. The appropriate lag and temporal resolution are assumed to be reasonable for such a medium scale rainfed sub-basin, where the summers can be dry. Small lags (zero days) and daily temporal resolution for the G index are appropriate for forecasts with a lead time of 14 days. The maximum correlation between groundwater and low flows is higher than for all other sub-basins (see Figure 2.10a), possibly due to the existence of large aquifers (Belz, 2010; Middelkoop and Van Haselen, 1999).
For forecasts with a lead time of 90 days, small lags (14 days) and large temporal resolutions (about eight months) are appropriate for the P index. Obviously, the appropriate lag for this index is shortened due to the additional lag of 76 days. For the PET index, the aforementioned effect of the annual hydrological cycle should be ignored and small lags (seven days) with large temporal resolutions (around three months) should be used as appropriate temporal scales for forecasts with a lead time of 90 days (see Figure 2.10b).

Middle Rhine

The MR and LR sub-basins have a mixed discharge regime as it originates from both Alpine and rainfed sub-basins. The daily-generated discharge series contain negative values, even after applying a lag to account for the travel time of the discharge wave. This is possibly due to damming, mining, storage changes and other anthropogenic effects (Belz, 2010; Harris, 1946; Hüffmeyer *et al.*, 2009). Moreover, given the large surface areas of these sub-basins the annual discharge generation is relatively low.
The MR sub-basin is located in the middle part of the Rhine basin, covering approximately 25 per cent of the total basin area. Therefore the PET index is an important low flow indicator for forecasts with lead times of 14 and 90 days. Similarly, we found

high correlations between the G index and low flows for forecasts with a lead time of 14 days (Figure 2.11). The high correlations for the PET index together with the low annual discharge generation rate show the significant role of PET in the water balance of the sub-basin.

Small lags (zero days) and large temporal scales (around two months) for the PET index are appropriate for forecasts with a lead time of 14 days. This means that the potential evapotranspiration amount in the preceding two months with a lag of zero days results in the highest correlation with low flows for a lead time of 14 days. On the other hand, small lags (zero days) and daily temporal resolution for the G index are appropriate for forecasts with lead times of 14 and 90 days. It can be concluded that the preceding daily groundwater levels are relevant for low flows in the MR sub-basin. The effect of the annual hydrological cycle on the correlations for the PET and G indices should be ignored and the aforementioned appropriate temporal scales for these low flow indicators should be used for forecasts with a lead time of 90 days. Although the groundwater levels do not change quickly, groundwater can be the only storage sustaining low flows in a dry period, in particular in rainfed sub-basins.

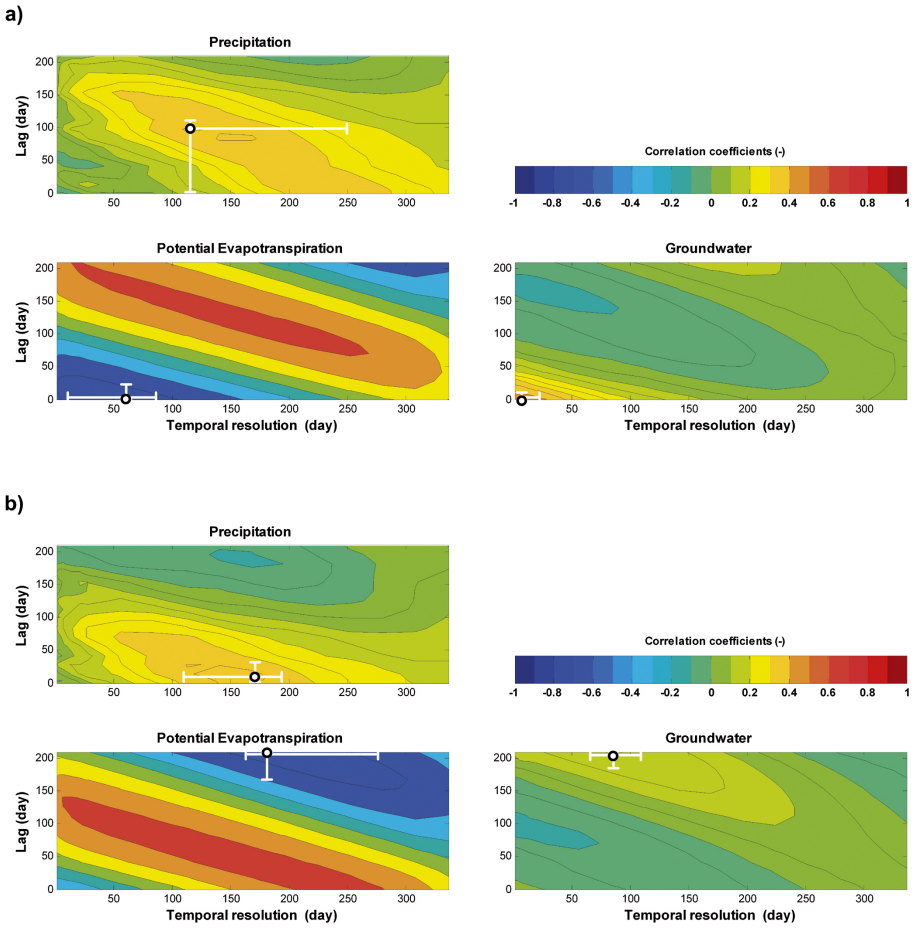


Figure 2.11: Cross correlation coefficients between low flows in the Middle Rhine sub-basin and pre-selected indicators as a function of lag and temporal resolution (days): **a)** low flows with a temporal resolution of three days and a lead time of 14 days; **b)** low flows with a temporal resolution of seven days and a lead time of 90 days.

Lower Rhine

Figure 2.12 shows the correlation coefficients between low flows and indicators for the LR sub-basin. The correlations are relatively low compared to upstream sub-basins. The *PET* and *G* indices are the most important low flow indicators here. The annually-generated discharge is about 273 mm. Large lags (around seven months) and temporal resolutions (around ten months) for the *PET* index are appropriate for forecasts with a lead time of 14 days, while small lags (zero days) and daily temporal resolution for the *G* index are appropriate for forecasts with lead times of 14 and 90 days. The annual hydrological cycle effect on the *G* index should be ignored (Figure 2.12b).

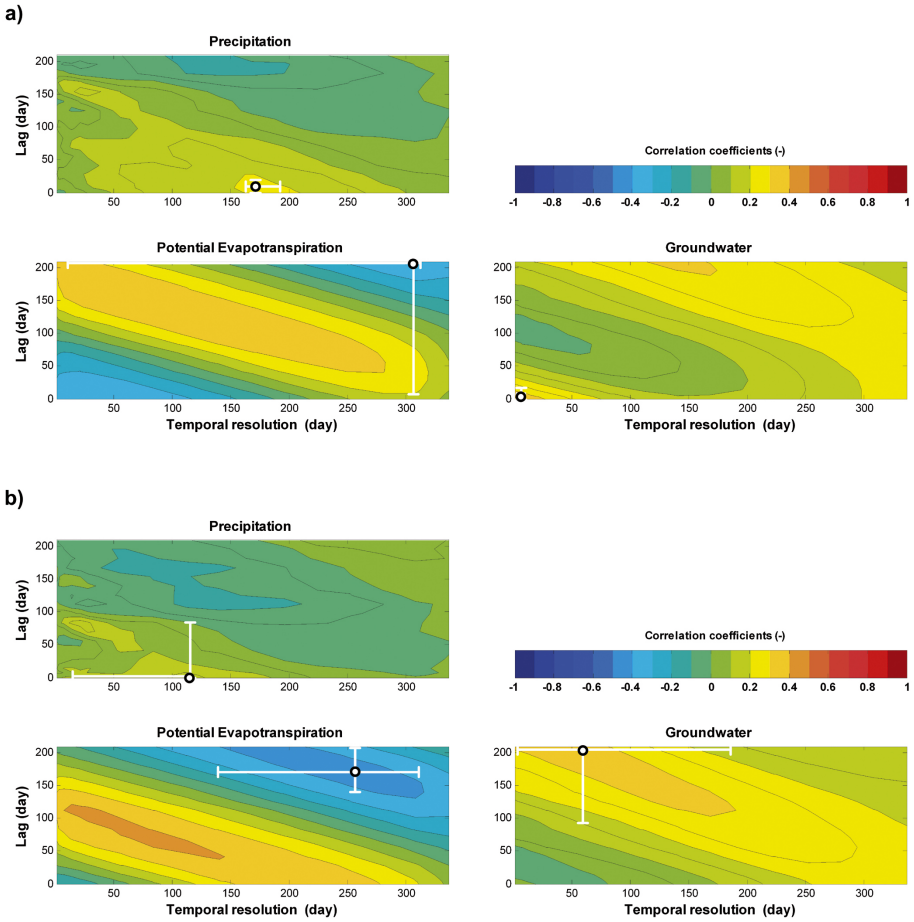


Figure 2.12: Cross correlation coefficients between low flows in the Lower Rhine sub-basin and pre-selected indicators as a function of lag and temporal resolution (days): **a)** low flows with a temporal resolution of three days and a lead time of 14 days; **b)** low flows with a temporal resolution of seven days and a lead time of 90 days.

Overall results agree with the general theory that the discharge response of a river basin is closely related to the preceding precipitation averages over several months. This is due to the fact that precipitation deficits over a long period can lead to a significant decrease in discharges (De Wit *et al.*, 2007; Vicente-Serrano and López-Moreno, 2005). Zaidman *et al.* (2001) found strong correlations between low flows and average precipitation deficits over the preceding two to four months in northwest Europe. They indicated that the catchment geology plays an important role in low flows. We also found strong correlations between groundwater storage and low flows in most sub-basins. We did not analyse the correlations between atmospheric and oceanic indicators, such as the North Atlantic Oscillation index and the El Niño Southern Oscillation index, and low flows. Recent work has suggested that these indices did not increase the ability to predict of summer discharge (low flows) and water temperature (Rutten *et al.*, 2008).

2.5 Conclusions

This study investigated the use of sub-basin averaged standard indices and correlation analysis to characterise low flows in the River Rhine basin for an ultimate goal of low flow forecasting. Correlation analysis is not new in low flow hydrology, as it has been applied in a number of river basins (Rutten *et al.*, 2008; Vicente-Serrano and López-Moreno, 2005; Wedgbrow *et al.*, 2002). However, to our knowledge this is the first study applying correlation analysis to low flows in the River Rhine.

This study presented a correlation analysis, to assess the relative importance of low flow indicators for the Rhine and to identify their appropriate lags and temporal resolutions. The most important indicators in the Alpine sub-basins for forecasts with a lead time of 14 days are potential evapotranspiration and lake levels. In the other sub-basins groundwater levels and potential evapotranspiration are relevant for low flows. Similarly, the most important indicators for forecasts with a lead time of 90 days are potential evapotranspiration, lake levels and snow depths for the Alpine sub-basins, whereas in the other sub-basins the most important indicators are potential evapotranspiration and precipitation or groundwater.

Overall, small lags and temporal resolutions are appropriate for lake levels and groundwater in the sub-basins for forecasts with lead times of 14 and 90 days, while large lags and temporal resolutions are appropriate for the P, PET and S indices. The uncertainty over the appropriate lags and temporal resolutions was estimated for each indicator as well. As a result we found a large uncertainty range around the maximum correlation points for most low flow indicators. In all sub-basins the largest uncertainties are found for the PET index. Lake levels and the G index show a small uncertainty range around the maximum correlation point.

The identified lags and temporal resolutions will be instrumentally useful in creating operational low flow forecast models for the River Rhine with lead times of 14 and 90 days. Additionally, understanding the low flow mechanisms and subsequent storage responses should aid the selection of appropriate models and the choice of proper temporal scales. Anticipating low flows would allow the making of more strategic decisions for river functions (e.g. navigation, cooling water supply) affected by low flows, as more low flows as well as extreme flood peaks are expected in the future (Hurkmans *et al.*, 2010). The framework presented in this study can be applied to all discharge regimes. Other methods, such as wavelet coherence analysis and chaotic correlation dimension methods, could also be used to identify dominant scales and the number of indicators in each sub-basin. Such methods are the basis for our on-going research.

2.6 Acknowledgements

We acknowledge the financial support of the Dr. Ir. Cornelis Lely Stichting (CLS), Project No. 20957310. Discharge data for the River Rhine were provided by the Global Runoff Data Centre (GRDC) in Koblenz (Germany). Areal precipitation and evapotranspiration data were supplied by the Federal Institute of Hydrology (BfG), Koblenz (Germany). REGNIE grid data were extracted from the archive of the Deutscher Wetterdienst (DWD: German Weather Service), Offenbach (Germany). Groundwater data for Bavaria were obtained from the Bayerisches Landesamt für Umwelt. Those for Hessen were provided by the Hessisches Landesamt für Umwelt und Geologie and for Baden-Württemberg by the Landesanstalt für Umwelt, Messungen und Naturschutz Baden-Württemberg. Groundwater data for the French part of the Moselle sub-basin were obtained from Portail national d'Accès aux Données sur les Eaux Souterraines (http://www.ades.eaufrance.fr/), those for North Rhine-Westphalia from the Landesamt für Natur, Umwelt und Verbraucherschutz Nordrhein-Westfalen, for Rhineland-Palatinate from the Landesamt für Umwelt, Wasserwirtschaft und Gewerbeaufsicht Rheinland-Pfalz and for the Saar sub-basin from the Wasser- und Schifffahrtsamt Saarbrücken. Groundwater and discharge data and lake level measurements for major lakes in Switzerland were provided by the Bundesamt für Umwelt (BAFU). The GIS base maps with delineated 134 catchments of the Rhine basin were provided by Eric Sprokkereef, the secretary general of the Rhine Commission (CHR). Finally, we thank Dr. Tobias Jonas from the WSL Institute for Snow and Avalanche Research, who provided fresh snow depth data and, made fruitful comments on the snow storage estimation procedure.

2.7 Appendix: Figures for the other rainfed sub-basins

Neckar

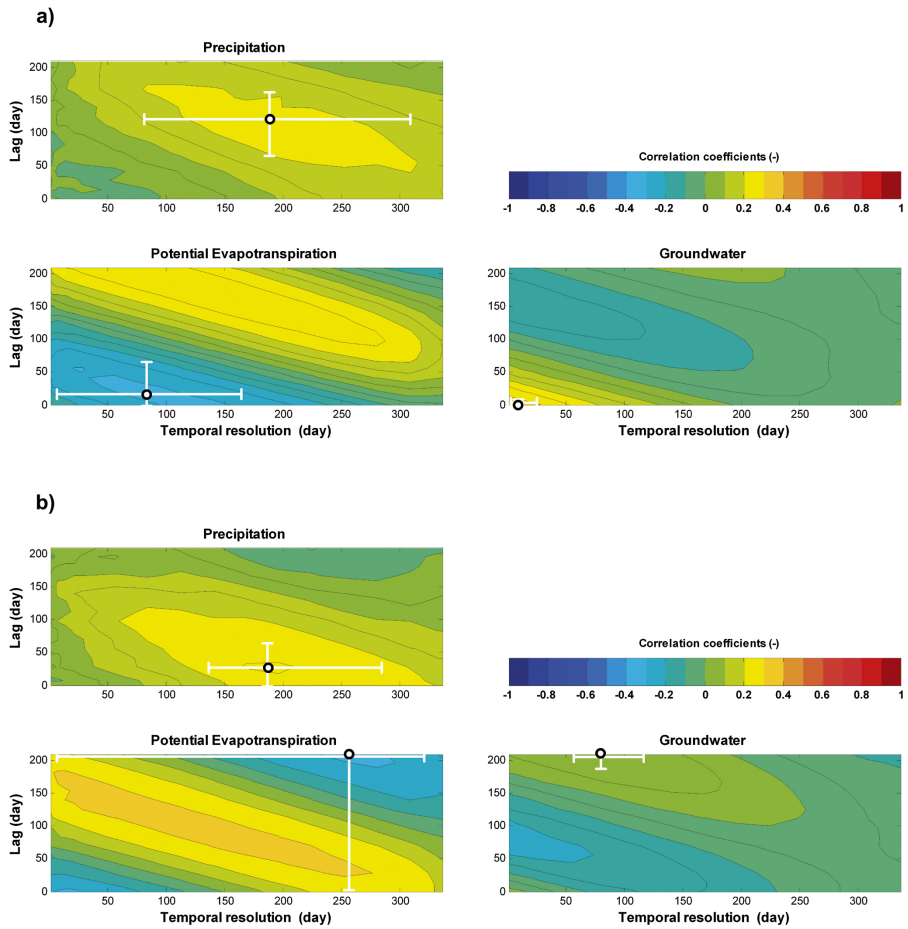


Figure 2.13: Cross correlation coefficients between low flows in the Neckar sub-basin and pre-selected indicators as a function of lag and temporal resolution (days): **a)** low flows with a temporal resolution of three days and a lead time of 14 days; **b)** low flows with a temporal resolution of seven days and a lead time of 90 days.

Moselle

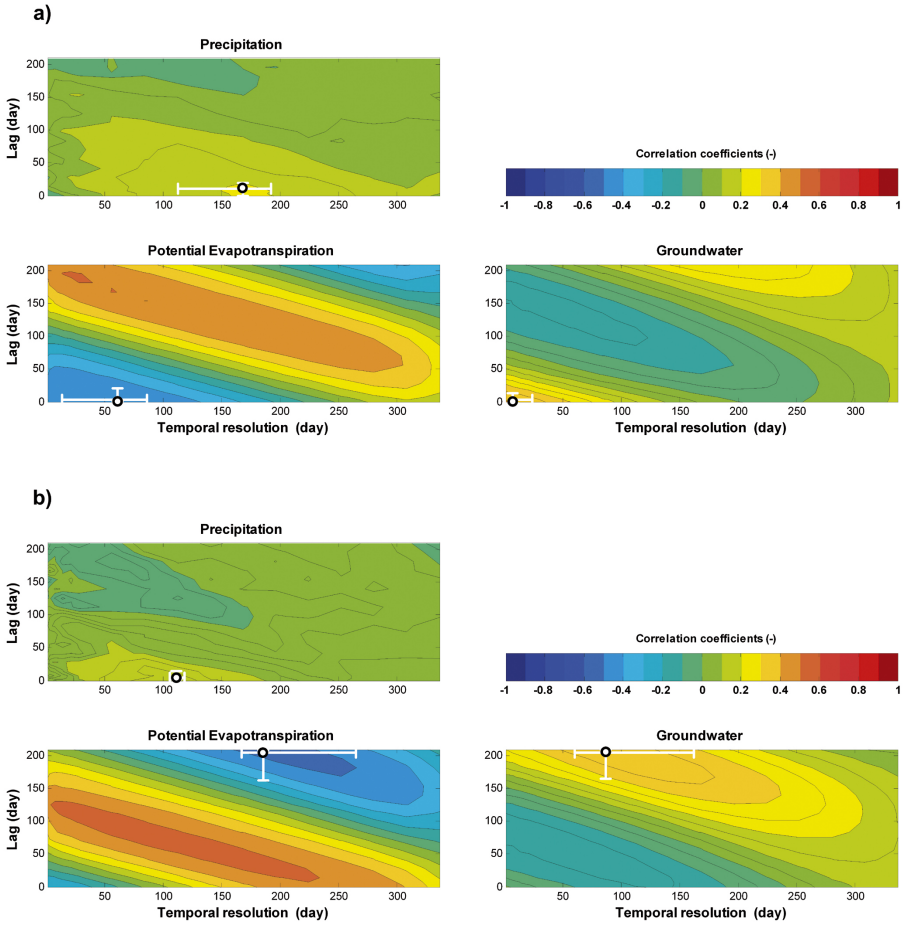


Figure 2.14: Cross correlation coefficients between low flows in the Moselle sub-basin and pre-selected indicators as a function of lag and temporal resolution (days): **a)** low flows with a temporal resolution of three days and a lead time of 14 days; **b)** low flows with a temporal resolution of seven days and a lead time of 90 days.

Chapter 3

Effect of different uncertainty sources on the skill of 10 day ensemble low flow forecasts

Abstract

This paper aims to investigate the effect of uncertainty originating from model inputs, parameters and initial conditions on 10 day ensemble low flow forecasts. Two hydrological models, GR4J and HBV, are applied to the Moselle River and performance in the calibration, validation and forecast periods, and the effect of different uncertainty sources on the quality of low flow forecasts are compared. The forecasts are generated by using meteorological ensemble forecasts as input to GR4J and HBV. The ensembles provided the uncertainty range for the model inputs. The Generalised Likelihood Uncertainty Estimation (GLUE) approach is used to estimate parameter uncertainty. The quality of the probabilistic low flow forecasts has been assessed by the relative confidence interval, reliability and hit/false alarm rates. The daily observed low flows are mostly captured by the 90% confidence interval for both models. However, GR4J usually overestimates low flows whereas HBV is prone to underestimate them, particularly when the parameter uncertainty is included in the forecasts. The total uncertainty in GR4J outputs is higher than in HBV. The forecasts issued by HBV incorporating input uncertainty resulted in the most reliable forecast distribution. The parameter uncertainty was the main reason reducing the number of hits. The number of false alarms in GR4J is twice the number of false alarms in HBV when considering all uncertainty sources. The results of this study showed that the parameter uncertainty has the largest effect whereas the input uncertainty had the smallest effect on the medium range low flow forecasts.

This chapter is based on the paper: Demirel MC, Booij MJ, Hoekstra AY, (2013). Effect of different uncertainty sources on the skill of 10 day ensemble low flow forecasts for two hydrological models. *Water Resources Research*, **49** (7): 4035-4053, doi:10.1002/wrcr.20294.

3.1 Introduction

Rainfed rivers in Western and Central Europe have a discharge regime with high-flows in winter and low-flows in late summer due to the temperate climate. The rivers, e.g. the River Rhine, are generally navigable throughout the year, a situation which has contributed to the region's industrial and trade development. The rivers are used for drinking water supply, irrigation, industrial use, power production, freight shipment and also fulfil ecological and recreational functions (De Wit *et al.*, 2007). Floods and low-flows are seasonal phenomena that may cause several problems to society. Since floods are eye-catching, quick and violent events risking human-life, contingency plans and water management bodies often focus on flood issues. In contrast, low flows are slowly developing events affecting a much larger area than floods. There is a growing concern that low flows will intensify due to climate change (Arnell, 1999; Grabs *et al.*, 1997; Hagemann *et al.*, 2008; Middelkoop *et al.*, 2001). Low flows in rivers may negatively affect the above mentioned river functions. Severe problems, e.g. water scarcity for drinking water supply and power production, hindrance to navigation and deterioration of water quality, have already been experienced during low flow events in the River Rhine in the dry summers such as in 1976, 1985 and 2003 indicating the importance of considering these events in addition to flood events.

To anticipate possible low flow events it is crucial that 10 day low flow forecasts become available in addition to short-range (1 – 4 days) forecasts. The forecasted low flow is commonly given as one value, even though it is an uncertain value. There is an increasing interest to account for uncertain information in decision support systems, e.g. how to operate river navigation and power plants during low flow periods to maximize the gain. One challenge is to develop systems that can use uncertain information (Engeland *et al.*, 2010). We are interested in forecasting low flows with a lead time of 10 days, and in presenting corresponding uncertainty to provide low flow information to major river users. This study focuses on assessing the uncertainty in ensemble 10 day low flow forecasts for two conceptual hydrological models.

Carrying out a systematic uncertainty analysis in hydrological modelling is an important field in hydrology, according to the numerous recent contributions in well-known journals (Cunha *et al.*, 2012; Rossa *et al.*, 2011; Salamon and Feyen, 2009; Tolson and Shoemaker, 2008). Uncertainty assessment has been one of the main goals of the Prediction in Ungauged Basins (PUB) initiative promoted by the International Association of Hydrological Sciences (Montanari, 2011). Similarly the Hydrological Ensemble Prediction Experiment (HEPEX), another international initiative, published a special issue on the results of the inter-comparison experiment for post-processing techniques for ensemble forecasts (Van Andel *et al.*, 2013). Systematic uncertainty analysis consists of several steps such as, classification, importance assessment, quantification, uncertainty propagation through model and, finally, communication of the uncertainty to the end

users. After the identification of the main sources of uncertainty (Ewen *et al.*, 2006), these sources must be classified. There are many different approaches for source classification. For instance, Walker *et al.* (2003) classified uncertainty as originating from model context, input, model structure and parameters. Other studies distinguished the uncertainty in observations, instruments and in the context of the problem, expert judgment and indicators (Janssen *et al.*, 2005; Van der Sluijs *et al.*, 2005; Warmink *et al.*, 2011). It has been commonly accepted that model inputs, parameters, initial conditions and structure are the major sources of uncertainty in conceptual hydrological models (Refsgaard *et al.*, 2006; Zappa *et al.*, 2011). We focus on these sources for further analysis. Quantification of the uncertainty sources is probably the most difficult step of the uncertainty analysis.

Uncertainty in forecasted input data, e.g. precipitation and temperature, is mainly from the assumptions and simplifications made when describing atmospheric processes in weather forecast models. In particular, future precipitation amounts are assumed to be very uncertain (Cunha *et al.*, 2012; Roulin, 2007). To quantify the uncertainty in the weather forecasts, an ensemble of lower resolution forecasts (ENS) has been developed by the European Medium Range Weather Forecasting Centre (ECMWF) and other national meteorological services (ECMWF, 2012). The system is operational since 1992, and a number of modifications have been implemented to its structure and grid resolution for improving the numerical weather predictions. In this system there are 50 different perturbed weather forecasts and an unperturbed control forecast. The 50 members, comprising an ensemble, are computed for a lead time of 15 days using perturbed initial conditions and model physics (Pappenberger *et al.*, 2005; Roulin and Vannitsem, 2005). Each member of an ensemble is assumed to be equally probable and provide useful information to address the uncertainty in future precipitation amounts (Roulin, 2007). However, in the context of flow forecasting it is important to assess the precipitation uncertainty in terms of the effect on runoff rather than in terms of comparing forecasted precipitation against observed precipitation (Arnaud *et al.*, 2011; Nester *et al.*, 2012). For example, Pappenberger *et al.* (2005) and more recently Pappenberger *et al.* (2011) used meteorological ensembles in hydrological models with different parameter settings to assess the uncertainty in flood forecasts. Similarly, Randrianasolo *et al.* (2010) coupled weather ensemble prediction system products from Météo-France with two hydrological models for forecasting discharges of 211 catchments in France for a lead time of two days.

Obviously there are other sources of uncertainty in low flow forecasts in addition to the model input (Meißner *et al.*, 2012; Zappa *et al.*, 2011). Hydrological models, whether using observed (Cunha *et al.*, 2012) or forecasted (Nester *et al.*, 2012) rainfall, are also limited by their capacity to represent the dominant processes in the river basin with appropriate spatial and temporal scales. Effective values of model parameters affected by local spatial heterogeneities and nonstationarities provide usually loose associations

with dominant processes (Lawal *et al.*, 1997; Pappenberger *et al.*, 2005; Stravs and Brilly, 2007). Therefore, the uncertainty due to model parameters will inevitably influence model outputs. There are a range of methods of quantifying model parameter errors including Monte Carlo simulations and analytical approaches (Montanari and Grossi, 2008). GLUE is a Monte Carlo based technique developed for calibration and estimation of uncertainty of predictive models using equifinality concept (Beven and Freer, 2001; Stedinger *et al.*, 2008; Viola *et al.*, 2009). Concerning the choice of the likelihood measure, Beven and Binley (1992) pointed out that many different likelihood measures in GLUE can be appropriate for a given application. Jin *et al.* (2010) compared different likelihood measures and the model uncertainty. They found that a less strict likelihood function, obviously leads to a wider confidence interval of the output uncertainty. Therefore, neither a too strict nor a too relaxed likelihood is appropriate for the GLUE assessment. The GLUE method has been widely used for flood forecasting (Pappenberger *et al.*, 2004; Pappenberger *et al.*, 2005) and for simulation of both high and low flows (Freer *et al.*, 1996; Tian *et al.*, 2013; Tolson and Shoemaker, 2008; Vázquez *et al.*, 2009; Viola *et al.*, 2009). In addition, the GLUE method is simple and relatively easy to implement. Therefore, GLUE is used in this study for model calibration and uncertainty analysis.

The drawbacks and advantages of the GLUE method have been enormously discussed in the hydrology literature (Beven *et al.*, 2007; Beven *et al.*, 2008; Beven and Young, 2003; Li *et al.*, 2010; Mantovan and Todini, 2006; Montanari, 2005; Stedinger *et al.*, 2008).

Beven *et al.* (2007) showed that if a correct formal likelihood was used in the GLUE method, the results would be identical with the formal Bayesian technique. Stedinger *et al.* (2008) also showed that GLUE can produce meaningful uncertainty and prediction intervals using a correct likelihood function. Beven (2006) argues whether the assumptions used in a formal Bayesian analysis are valid for any non-synthetic hydrologic system being modeled. In our ensemble low flow forecasting case study, performing a formal Bayesian method would require a very difficult process of deriving a correct description of the residual errors which are correlated in time and space. As a result, our study utilizes an elaborated low flow likelihood function within the GLUE method to assess parameter uncertainty.

Uncertainty in initial condition of state variables can have a significant effect on low flow forecasts. The summer forecasts, using a model initialised with an unsaturated soil state and with a saturated soil, will be very different. During prolonged dry periods the discharge largely originates from the release of groundwater storage (De Wit *et al.*, 2007; Tallaksen and Van Lanen, 2004). Therefore, uncertainty coming from the model initial conditions and groundwater storage in particular should be treated separately from the model parameter uncertainty, although model parameters and states are part of the model structure (Butts *et al.*, 2004). The uncertainties may be amplified when cascaded

through the hydrological model (Nester *et al.*, 2012). Komma *et al.* (2007) showed that small errors in rainfall may result in larger errors in model outputs. They showed that an uncertainty range of 70% in the precipitation ensemble increased to an uncertainty range of 200% in forecasted runoff with a lead time of 48 h. Although they related this to the nonlinearity of the catchment response, this amplification could also be caused by uncertainties from precipitation measurements and model parameters (Nester *et al.*, 2012). Komma *et al.* (2007) also showed that the assessment of precipitation uncertainty should be in terms of the effect on model outputs (herein forecasted low flows) instead of comparing only forecasted precipitation and observed precipitation.

Different methods i.e. particle filter (DeChant and Moradkhani, 2011), ensemble Kalman filter (Pasetto *et al.*, 2012), have been applied to asses initial condition uncertainty in the framework of ensemble streamflow prediction. The resulting update is similar to the static GLUE application (Pasetto *et al.*, 2012). These filters are examples of data assimilation techniques that are often used in short term flood forecasts (Liu *et al.*, 2012; Moradkhani *et al.*, 2012; Parrish *et al.*, 2012).

The aforementioned studies demonstrate the need for a systematic uncertainty analysis framework that isolates uncertainties due to various weather inputs, parameter estimation and initial conditions. Understanding the relative contributions of these sources to the total low flow forecast uncertainty and to the quality of forecasts can assist in the future development of ensemble forecasting systems. All studies mentioned constrain either to only flood forecast models or to simulation models used for low flows, but no similar application to low flow forecasts is known to the authors which also uses a sound model state updating procedure for assessing effect of initial condition uncertainty. There have been studies using ENS products for flood forecasting (Devineni *et al.*, 2008; Fundel and Zappa, 2011; Jaun and Ahrens, 2009; Muluye, 2011; Nester *et al.*, 2012; Pagano *et al.*, 2013; Renner *et al.*, 2009; Thirel *et al.*, 2010; Thirel *et al.*, 2008) and high resolution precipitation ensemble forecast of a regional climate model, i.e. COSMO-LEPS (Addor *et al.*, 2011; Zappa *et al.*, 2011) but no study is known to the authors which focuses on 10 day low flow forecasts. Only short-range low flow forecasts up to 4 days are issued by different water authorities for the entire Rhine basin (De Bruijn and Passchier, 2006). There have been different cross-border projects such as "Floods and low flow management in the Moselle and Saar Basin (FLOW MS)" focusing on climate impacts on low flows (Görgen *et al.*, 2010). However, low flow forecasts followed by a systematic uncertainty analysis do not exist for the Rhine basin and Moselle River in particular although there is a high demand (Meißner *et al.*, 2012; Rutten *et al.*, 2008) from different sectors (e.g. freight shipment, drinking water supply, and energy production).

The objectives of this study are to assess: (1) the uncertainty from ECMWF ensemble forecasted precipitation and potential evapotranspiration (2) the uncertainty from the parameters of two hydrological models by using the GLUE framework; (3) the uncer-

tainty due to the initial conditions; (4) the effect of these uncertainties on different low flow forecast quality and reliability measures.

Assessing the isolated three major sources of uncertainty, i.e. model input, parameters and initial conditions, in ensemble low flow forecasts by applying all steps from identification to communication of uncertainty, is an innovative way to understand and explain the effects of different uncertainties on the skill of low flow forecasts. In terms of model storage update, i.e. estimation of the model states at forecast issue day, a new method is proposed using observed discharge. This is superior compared to using only calibrated model run as there can be inevitable errors between simulated and observed discharge affecting the model initial condition. A further interesting aspect of this study is the use of a new hybrid low flow likelihood function for GLUE, which allows evaluation of low flows. Low flows in the Moselle River are investigated to allow the navigation and energy sectors to timely prepare for low flow conditions as they are the most important economic river functions (Li *et al.*, 2008; Rutten *et al.*, 2008; Svensson and Prudhomme, 2005). Since the River Rhine is a large scale river, the Moselle River is selected as a case study. We use an exceedence probability of 75% (Q75) as a threshold for the definition of low flows (Demirel *et al.*, 2013a). The number of days with low flows is sufficient to calibrate a forecast model, and low flows at this threshold are still affecting the important river functions. Several types of ensemble weather forecast products from the ENS dataset are incorporated in this study to prepare model inputs i.e. daily precipitation (P) and potential evapotranspiration (PET) for lead time of ten days from the ENS dataset. We address the model structure uncertainty by comparing two conceptual models with different complexities: the GR4J conceptual model with four parameters (Perrin *et al.*, 2003) and the HBV conceptual model with eight parameters (Lindström *et al.*, 1997). These models are assumed to represent dominant low flow indicators (predictors) with their appropriate temporal scales as identified by Demirel *et al.* (2013a). The GR4J model is a French conceptual hydrological model with a simple structure (Perrin *et al.*, 2003). With four parameters, it provides a minimum level of complexity. The HBV model has been calibrated and operationally used for the River Rhine (Renner *et al.*, 2009). Moreover, this model has been widely used in Rhine studies such as for real-time flow forecasts (Reggiani and Weerts, 2008a), climate impact assessment (Eberle, 2005; Hurkmans *et al.*, 2010; Te Linde *et al.*, 2010; Te Linde *et al.*, 2011) and for assessing uncertainties in flood forecasts due to ensemble weather forecasts by using a Bayesian postprocessor (Reggiani *et al.*, 2009; Reggiani and Weerts, 2008b).

The paper is organized as follows. In the next section, the study area and data are presented. The model structures and the uncertainty analysis method are described in section 3.3. The results are presented in section 3.4, and the conclusions are drawn in section 3.5.

3.2 Study area and data

3.2.1 Study area

The Moselle River has a surface area of approximately 27,262 km^2 and a length of 545 km. The source of the river is in the forested slopes of the Vosges massif and meanders before leaving France to form the border between Germany and Luxembourg for a short distance. The river enters Germany and flows past Trier to its confluence with the Rhine at Koblenz. Two major tributaries, the Sauer and Saar rivers flow into the Moselle before the Trier dam. There are other dams in the Moselle and Saar rivers whereas the Sauer river has a natural flow (Ackermann *et al.*, 2000; Belz *et al.*, 1999). Moreover, the river channels in the Moselle and Saar are mostly canalised for water management purposes and available for river navigation while the Sauer is not navigable (Behrmann-Godel and Eckmann, 2003). Annual generated discharge in the Moselle basin is about 410 mm (∼130 m^3/s). The measured discharge at Cochem station fluctuates between 14 m^3/s in dry summers and a maximum of 4000 m^3/s during winter floods. The altitude ranges from 59 to 1326 m with a mean altitude of 340 m (Demirel *et al.*, 2013a).

3.2.2 Data

Observed data

Observed daily precipitation (P) and potential evapotranspiration (PET) estimated with the Penman-Wendling equation (ATV-DVWK, 2002) were obtained from the German Federal Institute of Hydrology (BfG) in Koblenz (Germany). Both variables are spatially averaged, i.e. disaggregated over 26 Moselle sub-catchments.

The mean altitude of these sub-catchments has been also provided by BfG. The outlet discharge (Q) for the Moselle (station *#6336050* at Cochem) has been provided by the Global Runoff Data Centre (GRDC), Koblenz (Germany). The daily P, PET and Q data series span from 1951 to 2006 (Table 3.1).

Table 3.1: Observed data

Data	**Index**	**Spatial Resolution**	**Number of stations/sub-basins**	**Period**	**Time step**	**Source**
Discharge	Q	Point (Cochem)	1	1951-2006	24h	GRDC-Koblenz
Precipitation	P	Sub-catchments	26	1951-2006	24h	BfG-Koblenz
Evapotranspiration	PET	Sub-catchments	26	1951-2006	24h	BfG-Koblenz
Mean altitude	h	Sub-catchments	26	-	-	BfG-Koblenz

Meteorological ensemble forecast data

Both precipitation and other meteorological forecast data used in this study are originated from the ECMWF-ENS control and ensemble forecasts. These ensembles are computed for a lead time of 1 – 10 days using perturbed initial conditions and model physics (Table 3.2). A grid size of 0.25 degree (∼28km) is chosen to retrieve weather forecast products using the ECMWF Mars retrieval system. The PET forecasts are determined by the Penman-Wendling equation requiring only forecasted surface solar radiation and temperature at 2 meter data (ATV-DVWK, 2002). This is consistent with the observed PET estimation carried out by the Federal Institute of Hydrology in Koblenz, Germany. Both grid-based P and PET ensemble forecast data are first interpolated over 26 Moselle sub-catchments using areal weights. These sub-catchment averaged data are then aggregated to the Moselle basin level.

Table 3.2: Overview of the ensemble forecast data

Data	**Spatial Resolution**	**Ensemble size**	**Period**	**Time step**	**Lead time**	R^2 **(240h)**	**MAE (240h)**
Forecasted P	0.25 degree	50 + 1 control	2002-05	24h	240h	0,07	3,16 mm
Forecasted PET	0.25 degree	50 + 1 control	2002-05	24h	240h	0,77	0,59 mm

3.3 Methodology

3.3.1 Overview of the model structures

The two hydrological models (GR4J and HBV) are briefly described below. Figure 3.1 shows the simplified model structures.

GR4J

The GR4J conceptual model has a parsimonious structure with only four calibration parameters and has been frequently used over hundreds of catchments worldwide, with a broad range of climatic conditions from tropical to temperate and semi-arid catchments (Perrin *et al.* , 2003). The GR4J model requires only daily time series of precipitation (P) and potential evapotranspiration (PET) as input (Figure 3.1a). The four parameters in GR4J represent the maximum capacity of the production store ($X1$), the groundwater exchange coefficient ($X2$), the one day ahead capacity of the routing store ($X3$) and the time base of the unit hydrograph ($X4$). All four parameters are used to calibrate the model and estimate the parameter uncertainty (Table 3.3) based on Tian *et al.* (2013) and Thyer *et al.* (2009). The upper and lower limits are selected based on previous

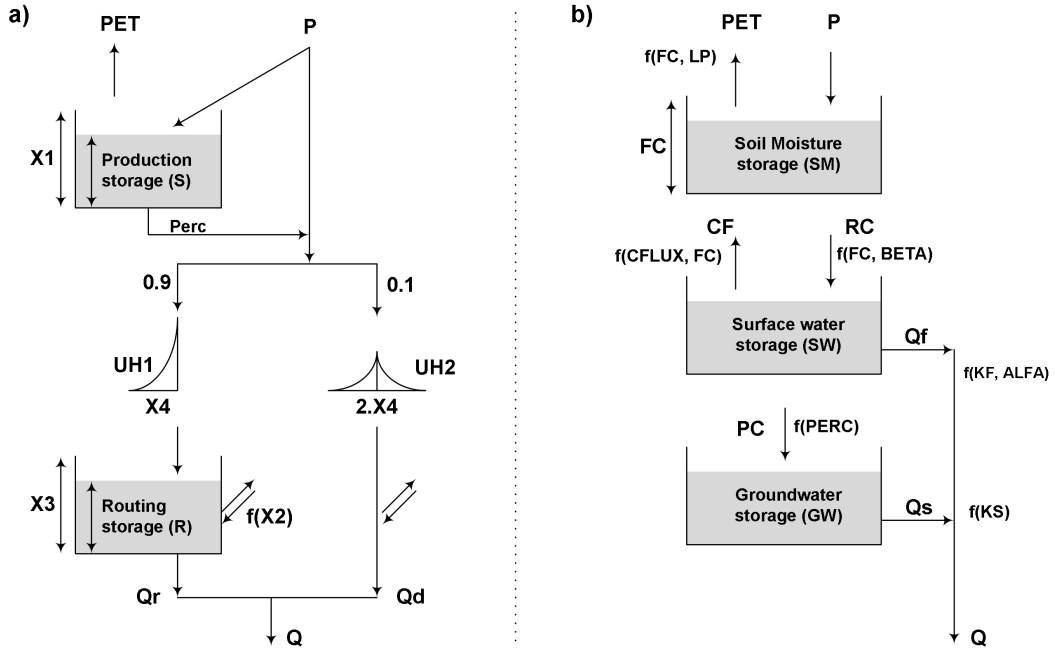


Figure 3.1: Schematic of **a)** GR4J model and **b)** HBV model

works (Booij, 2005; Eberle, 2005; Perrin *et al.*, 2003; Pushpalatha *et al.*, 2011; Tian *et al.*, 2013).

HBV

The HBV conceptual model was developed by the Swedish Meteorological and Hydrological Institute (SMHI) in the early 1970s (Lindström *et al.*, 1997). The HBV model consists of four subroutines: a precipitation and snow accumulation and melt routine, a soil moisture accounting routine and two runoff generation routines. The input data are daily P and PET. Since the Moselle basin is a rain-fed basin, the snow routine and daily temperature data are not used in this study (Figure 3.1b). The eight most important parameters in the HBV model (Table 3.3) are used to estimate the parameter uncertainty (Engeland *et al.*, 2010; Tian *et al.*, 2013; Van den Tillaart *et al.*, 2013).

Table 3.3: Parameters of the models used and their prior uncertainty ranges

Parameter	Unit	Range	Description
			GR4J model
X1	[mm]	10-2000	Capacity of the production store
X2	[mm]	-8 to +6	Groundwater exchange coefficient
X3	[mm]	10-500	One day ahead capacity of the routing store
X4	[d]	0-4	Time base of the unit hydrograph
			HBV model
FC	[mm]	100-800	Maximum soil moisture capacity
LP	[-]	0.1-1	Soil moisture threshold for reduction of evapotranspiration
BETA	[-]	1-6	Shape coefficient
CFLUX	[mm/d]	0.1-1	Maximum capillary flow from upper response box to soil moisture zone
ALFA	[-]	0.1-2	Measure for non-linearity of low flow in quick runoff reservoir
KF	[d^{-1}]	0.005-0.5	Recession coefficient for quick flow reservoir
KS	[d^{-1}]	0.0005-0.2	Recession coefficient for base flow reservoir
PERC	[mm/d]	0.01-6	Maximum flow from upper to lower response box

3.3.2 Calibration and validation

The GR4J and HBV models are calibrated using the GLUE method and historical Moselle low flows for the period from 01/01/1971 to 31/12/2001. The first forecast issue date is 01/01/2002 and the number of low flow events (i.e. 2762 days with low flows) is assumed to be long enough for calibrating a hydrological model (Perrin *et al.*, 2007). The validation period spans from 01/01/1951 to 31/12/1970. The definition of low flows, i.e. discharges below the Q75 threshold of $\sim$113 m^3/s, is based on previous work by Demirel *et al.* (2013a).

The Generalized Likelihood Uncertainty Estimation (GLUE) method (Beven and Binley, 1992) uses the "*equifinality*" concept rejecting only one optimal parameter set, instead, it uses many parameters sets that provide relatively equal performance (Beven and Freer, 2001). This method is developed as an extension of the Generalised Sensitivity Analysis (GSA) of Spear and Hornberger (1980) based on Bayesian Monte Carlo simulations. GLUE has been widely used for calibration of hydrological models since it is easy to implement and allows flexible definition of a likelihood function to evaluate the model outputs and to distinguish between behavioural (accepted) and non-behavioural (rejected) parameter sets (Freer *et al.*, 1996; Ratto *et al.*, 2007; Renard *et al.*, 2010; Shen *et al.*, 2012). Behavioural parameter sets are then those that provide predicted low flows that fall within the limits of acceptability with regard to a given likelihood measure (Zheng and Keller, 2007). It should be noted that the selection of the behavioural parameter sets is based on only the calibration period runs. In this study, the GLUE method, consisting of the three steps below, is applied for the selection of behavioural parameter sets. It is assumed that these parameter sets represent the uncertainty in model parameters.

Step 1: definition of a hybrid likelihood function for low flows

The most commonly used likelihood function in GLUE literature is the Nash-Sutcliffe (NS) coefficient (Beven and Freer, 2001; Nash and Sutcliffe, 1970; Shen *et al.*, 2012). However, other likelihood functions have been used for low flows (Pushpalatha *et al.*, 2012). In our study, we combined two low flow likelihood functions using subjectively selected weights. The new hybrid likelihood function (NS_{hybrid}) substantially improves the low flow forecasts as it combines NS based on only low flows (NS_a) and NS based on inverse discharge values (NS_b) (see equations (3.1)-(3.3)).

$$NS_a = 1 - \frac{\sum_{j=1}^{m} \left(Q_{sim}(j) - Q_{obs}(j)\right)^2}{\sum_{j=1}^{m} \left(Q_{obs}(j) - \bar{Q}_{obs}\right)^2} \tag{3.1}$$

where Q_{obs} and Q_{sim} are the observed and simulated values for the j-th observed low flow day (i.e. $Q_{obs} < Q_{75}$) and m is the total number of low flow days.

$$NS_b = 1 - \frac{\sum_{i=1}^{n} \left(1/(Q_{sim}(i) + \varepsilon) - 1/(Q_{obs}(i) + \varepsilon)\right)^2}{\sum_{i=1}^{n} \left(1/(Q_{obs}(i) + \varepsilon) - 1/(\bar{Q}_{obs} + \varepsilon)\right)^2} \tag{3.2}$$

where n is the total number of days (i.e. $m < n$), and ε is 1% of the mean observed discharge to avoid infinity during zero discharge days.

$$NS_{hybrid} = \alpha \left(\frac{NS_a}{2 - NS_a}\right) + \beta \left(\frac{NS_b}{2 - NS_b}\right) \tag{3.3}$$

where both NS_a and NS_b values range from -∞ to 1, with 1 indicating a perfect fit (Pushpalatha *et al.*, 2012). The weights α and β are selected as 0.3 and 0.7 respectively. These weights have been determined during calibration period. First component of our hybrid likelihood function is strictly developed for the low flows. Therefore, the resultant scores for this component can often be negative. Second component considers the inverse of all discharge values. The weights are determined for making the outcome values of our hybrid likelihood function positive for the calibration runs. In other words, the weights keep the balance between very strict and less strict likelihood functions since in cases with very strict low flow calibration i.e. high α values, both the GR4J and HBV models show results with very low likelihood values since the NS_a values are negative for both models.

Step 2: sampling parameter sets for two conceptual models

Previous model calibration and sensitivity analysis of the GR4J (Perrin *et al.*, 2003; Pushpalatha *et al.*, 2011) and HBV (Booij, 2005; Eberle, 2005; Tian *et al.*, 2013) in other rain-fed basins have allowed the prior uncertainty ranges of sensitive parameters to be assessed. These studies also indicated significant uncertainties for the sensitive parameters and emphasized the importance of inspecting the upper and lower parameter limits more in detail. Therefore, a sensitivity analysis is pursued using a large parameter space to select the most important parameters and their appropriate upper and lower limit values.

Independent uniform distributions for each effective parameter are chosen due to the lack of prior knowledge about the true distributions. The typical drawback of the GLUE method is the computational time caused by its random sampling strategy. Therefore, an improved sampling technique, i.e. Latin Hypercube Sampling (LHS), was used with the GLUE method (McKay *et al.*, 1979). Compared to a standard GLUE random sampling, LHS substantially reduces the computational burden for sampling and provide a 10-fold greater efficiency in parameter space coverage (Shen *et al.*, 2012). The sampling size should be large enough to ensure a sufficient calibration of the model. In this study, we generated 120,000 parameter sets for each conceptual model using LHS in the range of lower and upper limits given in Table 3.3. To our knowledge, this is the largest LHS sample size tested in low flow hydrology.

Step 3: threshold definition for behavioural model selection

The GR4J and HBV models are run for each of the 120,000 sets in the calibration. The output is evaluated against the observed daily discharge at Cochem station located at the outlet of the Moselle sub-basin using the NS_{hybrid} likelihood function to distinguish between behavioural parameter sets (accepted) and non-behavioural parameter sets (rejected). The parameter sets meeting the predefined threshold criterion (NS_{hybrid}>0.40) are accepted. Although the threshold value is a subjective decision (Jin *et al.*, 2010), we rigorously tested several thresholds based on low flow simulations and the size of the behavioural parameter sets for each model. The selected threshold resulted in two large behavioural parameter sets for parameter uncertainty analysis i.e. 9770×4 (GR4J) corresponding to ∼8% of the sample parameter set and 10909×8 (HBV) corresponding to ∼9% of the sample parameter set.

3.3.3 Model storage update procedure

Model storage updating is based on the observed discharge on the forecast issue day (Q_{obs}). This is a crucial step for medium range and seasonal low flow forecasts since the model initial state determines the model outputs (Wöhling *et al.*, 2006).

There are two storages in the GR4J model and three storages in the HBV model which are updated during low flow forecasts. The reader is referred to Perrin *et al.* (2003) and Lindström *et al.* (1997) for details of the process formulations in these models. A practical approach is used for both models. First, the two calibrated models are run with their best performing parameter sets and model states are analysed. This run is called the "*reference run*" in other recent works (Fundel and Zappa, 2011; Roulin, 2007; Roulin and Vannitsem, 2005). The empirical relations between the simulated discharge and the fast runoff for each model are used to divide the observed discharge between the fast and slow runoff components (equations (3.4) and (3.5)).

$$k_{GR4J} = \frac{Q_d}{Q_r + Q_d} \tag{3.4}$$

$$k_{HBV} = \frac{Q_f}{Q_f + Q_s} \tag{3.5}$$

The Q_r and Q_d in the GR4J model, and Q_f and Q_s in the HBV model are estimated using these fractions together with the observed discharge value at the forecast issue day. Subsequently, the routing storage (R) in the GR4J model is updated for a given value of the *X3* parameter using Eq. (3.6). Further, the surface water (SW) and groundwater (GW) storages in the HBV model can be updated for a given value of KF, $ALFA$ and KS parameters using equations (3.7) - (3.8).

$$Q_r = R\left\{1 - \left[1 + \left(\frac{R}{X3}\right)^4\right]^{-1/4}\right\} \tag{3.6}$$

$$SW = \left(\frac{Q_f}{KF}\right)^{(1/(1+ALFA))} \tag{3.7}$$

$$GW = \frac{Q_s}{KS} \tag{3.8}$$

The other two storages S (in GR4J) and SM (in HBV) are difficult to update using an empirical bottom-up approach. Instead, these two storages are updated using the calibrated model run until the forecast issue day (i.e. top-down approach). It is assumed that the two updated storages (S and SM) represent the reality due to the calibrated model run. However, we are aware that, there are inevitable uncertainties associated with this rough estimation of the initial conditions based on observed discharges and calibrated models' simulations.

3.3.4 Uncertainty sources and quantification

A robust assessment of uncertainties begins with identification of all sources (Refsgaard *et al.*, 2007). Obviously not all sources can be quantified. We used available classification schemes to identify and select the most important uncertainty sources in the two conceptual models and for low flow forecasting (Walker *et al.*, 2003; Warmink *et al.*, 2010). Three uncertainty sources and their quantification are described in the following sections. It should be noted that errors in model structure are also important for hydrological models (Götzinger and Bárdossy, 2008; Gupta *et al.*, 2012; Renard *et al.*, 2010; Tian *et al.*, 2013). In this study, the model structure uncertainty is addressed partly by comparing two different model structures.

Input uncertainty

A rainfall event after the forecast issue day can easily increase flows above low flow threshold in the Moselle River. Therefore, low flow forecasts are highly dependent on the quality of ENS weather forecasts. These forecasts are available at a spatial resolution of 0.25×0.25 degree for daily time steps. It has been reported that after several days these forecasts are highly uncertain due to the modelling limitations and complexity of the physical processes involved in the atmosphere (Fundel and Zappa, 2011; Reggiani *et al.*, 2009). In this study, the 51 ensemble members are used to quantify the uncertainty of future precipitation and potential evapotranspiration amounts. Obviously new uncertainties are introduced using the empirical PET formula (ATV-DVWK, 2002) and grid data interpolation over 26 Moselle sub-catchments.

Parameter uncertainty

The GLUE method is used for quantification of parameter uncertainties. This method rejects the idea of an optimal system representation and applies the equifinality concept accepting all forecasts using the behavioural parameter sets. This parameter ensemble then allows assessment of the output uncertainty arising from model parameters and partly from model structure (Pappenberger *et al.*, 2005). All four parameters of GR4J and eight parameters of the HBV are selected to estimate the parameter uncertainty.

Initial condition uncertainty

The importance of initial conditions for hydrologic forecasting is well established (Shukla and Lettenmaier, 2011; Wood and Lettenmaier, 2008). In most of the hydrologic modelling studies, initial conditions refer only to land surface states including soil moisture and snow cover (Li *et al.*, 2009). In this study, however, all model states, present in the conceptual models used, are included in the uncertainty analysis. This is from the fact that errors in estimated initial slow and fast runoff storages directly affect the low flow forecasts. We demonstrate a dynamic inverse-modelling approach based on observed discharge and uniformly distributed behavioural parameter sets on the forecast issue day for exploring the initial condition uncertainties and characterizing the relative importance of this uncertainty source for low flow forecasts. The *X3* parameter of GR4J and *KF*, *ALFA* and *KS* parameters of HBV are uncertain parameters that are directly linked to the model states (i.e. initial conditions). Other parameters are only assumed uncertain in the parameter uncertainty assessment.

3.3.5 Uncertainty propagation

The three sources of uncertainty described above are propagated through the GR4J and HBV models both separately and together. The latter case is executed to encapsulate the total uncertainty arising from all three sources together. In other words, this study employs the GLUE, an extended GSA method (Freer *et al.*, 1996; Ratto *et al.*, 2007), to apportion the output uncertainty of a model to different sources of uncertainty.

The 10 day low flow forecasts are issued every day for the test period from 1 January 2002 until 31 December 2005. The posterior distribution of the model outputs (e.g. confidence interval) is based on 10.000 Monte Carlo runs for each day (a total of 1461 days). The size of the Monte Carlo sample is assumed to be reasonable based on the number of behavioural parameter sets and on the

relevant literature (Blasone *et al.*, 2008; Franz and Hogue, 2011; Rossa *et al.*, 2011; Shen *et al.*, 2012).

For assessing the effect of the uncertainty in the forecasted input data, we run the models using randomly selected P and PET values from 51 members while the model parameters and model states are fixed according to the best performing calibrated parameter values (Table 3.4).

For assessing uncertainty in model parameters, we run the models using randomly selected behavioural parameter sets while the model inputs are fixed to ECMWF-ENS control forecast P and PET, and the model states are updated using the best performing calibrated parameter values.

For assessing the uncertainty in model states at forecast issue day, the routing storage (R) in the GR4J model is updated using randomly selected behavioural *X3* parameter values, and the surface water (SW) and groundwater (GW) storages in the HBV model are updated using randomly selected values of *KF*, *ALFA* and *KS* parameters from behavioural parameter sets for each of the 10.000 Monte Carlo runs for each day. The remaining model parameters are fixed to the best performing calibrated parameter values and the model inputs are fixed to ECMWF-ENS control forecast P and PET to evaluate the initial condition uncertainty (Table 3.4).

Table 3.4: Overview of the uncertainty propagation test scheme

Assessed Uncertainty	**Forecasted P and PET**		**Parameters**		**Initial Conditions**	
	1 Control	51 Ensemble	Calibrated set	GLUE set	Calibrated run	GLUE set ($X3$, KF, $ALFA$ and KS)
Deterministic	✓		✓		✓	
Input		✓	✓		✓	
Parameter	✓			✓	✓	
Initial Condition	✓		✓			✓
Total		✓		✓		✓

It should be noted that spatial and temporal consistency of the inputs are preserved to avoid nonphysical outcomes. For assessing the total uncertainty, we run the models using randomly selected model inputs, behavioural parameter sets and corresponding model states. For example, the storage S can never exceed the *X1* parameter value in the GR4J model (Perrin *et al.*, 2003). Similarly, we defined a "saturation rate" as the fraction of SM storage to the calibrated FC parameter.

$$SM\left(k,t\right) = FC\left(k\right) \times \frac{SM_{opt}\left(t\right)}{FC_{opt}} \tag{3.9}$$

For each k-th Monte Carlo run at each t-th forecast issue day, a new parameter set is randomly selected from the behavioural set and SM is calculated using Eq. (3.9). Therefore, the saturation rate is kept constant for a particular day

throughout the entire uncertainty propagation framework.

3.3.6 Uncertainty presentation

Uncertainty presentation allows the low flow forecasts to be monitored, thus helping to improve forecast quality by analysing uncertainties in the model outputs and allowing comparison of different models. Obviously, the added value of a low flow forecasts for decision makers depends on its uncertainty characteristics. We employed three forecast quality measures to analyse the results of the uncertainty quantification in 10 day low flow forecasts. These measures have been often used in meteorology (WMO, 2012) and flood hydrology (Renner *et al.*, 2009; Thirel *et al.*, 2008; Velázquez *et al.*, 2010). In WMO (2012), three properties of an accurate probabilistic forecast are defined as reliability, sharpness and resolution. In this study, three forecast quality measures have been rigorously selected to evaluate these three properties of the forecasts i.e. Reliability diagram - reliability, RCI - sharpness, and contingency table - resolution.

Relative confidence interval

The standard 90% confidence interval (90CI) was derived by ordering the 10.000 outputs on every day in the test period and then identifying the 5% and 95% percentiles (i.e. Q5 and Q95). The 90CI, observed discharge and 50% percentile (i.e. Q50 forecast median) are presented together. The relative confidence interval (RCI) is then calculated for only low flow days j using Eq. (3.10) to monitor the evolution of uncertainties with increasing lead time and to compare the effect of different uncertainty sources on the relative confidence interval.

$$RCI = \frac{1}{m} \times \sum_{j=1}^{m} \frac{Q95(j) - Q5(j)}{Q50(j)} \tag{3.10}$$

where m is the total number of low flow days.

Reliability diagram

The reliability diagram is an approach used to represent the performance of probabilistic forecasts of selected events, i.e. low flows (Bröcker and Smith, 2007). A reliability diagram shows the observed relative frequency as a function of forecast probability and the 1:1 diagonal represents the perfect reliability line (Olsson and Lindström, 2008; Velázquez *et al.*, 2010). In the present study, non-exceedence probabilities of 50%, 75%, 85%, 95%, and 99% are chosen as thresholds to catego-

rize the discharges from mean flows to extreme low flows. The forecast probability for each forecast day is estimated as the number of ensemble forecasts exceeding these thresholds divided by the total number of ensemble forecasts (i.e. 10.000 ensembles) in that forecast day. The forecasts are then divided into bins of probability categories; here, five bins (categories) are chosen 0 – 20%, 20% – 40%, 40% – 60%, 60% – 80% and 80% – 100%. The observed frequency for each day is estimated as 1 if the observed discharge exceeds the threshold, or 0, if not. The forecast probability and observed frequency can then be drawn.

Contingency table

We used contingency tables to assess the effect of uncertainty on the performance of low flow forecasts. Contingency tables, particularly used in flood warnings (Martina *et al.*, 2005), can be used to estimate the utility of hydrological forecasts and, in their simplest form, indicate the forecast models ability to anticipate correctly the occurrence or non occurrence of pre-selected events (i.e. Q75 low flows). The definitions of four cases are given in a two-by-two contingency table (Table 3.5).

Table 3.5: Contingency table for the assessment of threshold based forecasts

Low flow event (Q75)	**Observed**	**Not observed**
Forecasted	*hit*: the event forecast to occur and did occur	*false alarm*: event forecast to occur, but did not occur
Not forecasted	*miss*: the event forecast not to occur, but did occur	*correct negative*: event forecast not to occur and did not occur

The skill of a forecasting model can be represented on the basis of the hit rate and the false-alarm rate (Cloke and Pappenberger, 2009; Martina *et al.*, 2005). Both ratios can be easily calculated from the contingency table using equations (3.11) and (3.12).

$$hit\ rate = \frac{hits}{(hits + misses)} \tag{3.11}$$

$$false\ alarm\ rate = \frac{false\ alarms}{(correct\ negatives + false\ alarms)} \tag{3.12}$$

It should be noted that these ratios are also known as the probability of detection and the probability of false detection in other hydrological studies (Velázquez *et al.*, 2010). The hit and false-alarm rates, indicate respectively the proportion of events for which a correct warning was issued, and the proportion of non events for which a false warning was issued by the forecast model.

3.4 Results and discussion

3.4.1 Calibration and validation

The best performing parameter sets of the two models are shown in Table 3.6. The corresponding highest NS_{hybrid} values are 0.62 for the GR4J model and 0.56 for the HBV model. The GR4J model performs better than the HBV model on low flows based on only the best performing simulation in the calibration period. However, the HBV model performed better in the validation period. The highest NS_{hybrid} values did not change using another global optimisation technique i.e. a Genetic Algorithm (Velázquez *et al.*, 2010) showing that 120,000 parameter sets for each model are enough for calibrating the models. Considering the performance only in the low flow period (*i.e.* NS_a), the performance of the HBV model is better than the GR4J model in the calibration period. However, the drop in performance of the HBV model in the validation and the forecast periods is much larger than for the GR4J model. Such a drastic drop outside the calibration period is expected from a relatively complex model like HBV with 8 parameters since it has more degrees of freedom to adjust to the basin behaviour during the calibration period. This characteristic is somewhat concealed by the NS_{hybrid} results due to the subjective weights and due to the insensitive inverse performance index (*i.e.* NS_b). That is why all three performance indices have been presented in Table 3.6. The models are calibrated for a relatively wetter climate with ∼910 mm mean annual precipitation than for the validation period (∼890 mm) and forecast period (∼830 mm).

Table 3.6: Calibration, validation and forecast results

Parameter and likelihood	**1971-2001 Calibration**	**1951-1970 Validation**	**2002-05 Forecast**
	GR4J model		
X1 [mm]	649.2		
X2 [mm]	-1.3		
X3 [mm]	47.2		
X4 [d]	2.9		
NS_a	-0.68	-1.39	9.29
NS_b	0.99	0.76	0.99
NS_{hybrid}	**0.62**	**0.31**	**0.45**
	HBV model		
FC [mm]	239.6		
LP [-]	0.48		
BETA [-]	2.13		
CFLUX [mm/d]	0.13		
ALFA [-]	1.77		
KF [d^{-1}]	0.01		
KS [d^{-1}]	0.01		
PERC [mm/d]	0.76		
NS_a	-0.23	-1.13	-70.16
NS_b	0.92	0.9	0.99
NS_{hybrid}	**0.56**	**0.47**	**0.41**

The calibrated models are run for the test period (i.e. 2002-2005) to estimate the fraction of fast runoff to total runoff in the two models (Figure 3.2) and to update storages S and SM (i.e. top-down approach).
The exponential relation between the fraction values and the simulated discharge shows that the total discharge is dominated by flows from the fast runoff storage during high flows. These categorized values have been used to estimate fast and slow runoff storages in the GR4J and HBV models. The k_{GR4J} fraction is zero for low flows and 0.04 for high flows above 6 mm whereas the k_{HBV} fraction is about zero for low flows and 1 for high flows above 6 mm. Table 3.7 shows the empirical equations fitted to the simulation data presented in Figure 3.2.

Table 3.7: Empirical equations to divide observed discharge, (Q_{obs}), into fast and slow runoff

	Q_{obs} **category (mm)**		
	<=Q75	**>Q75 and <=6**	**>6**
k_{GR4J}	0	$-0.00007 \times {Q_{obs}}^3 + 0.001 \times {Q_{obs}}^2 + 0.0003 \times Q_{obs} + 0.004$	0.04
k_{HBV}	$4.9 \times {Q_{obs}}^{3.7} + 0.002$	$0.81 \times e^{(0.02 \times Q_{obs})} - 1.4 \times e^{(-1.9 \times Q_{obs})}$	1

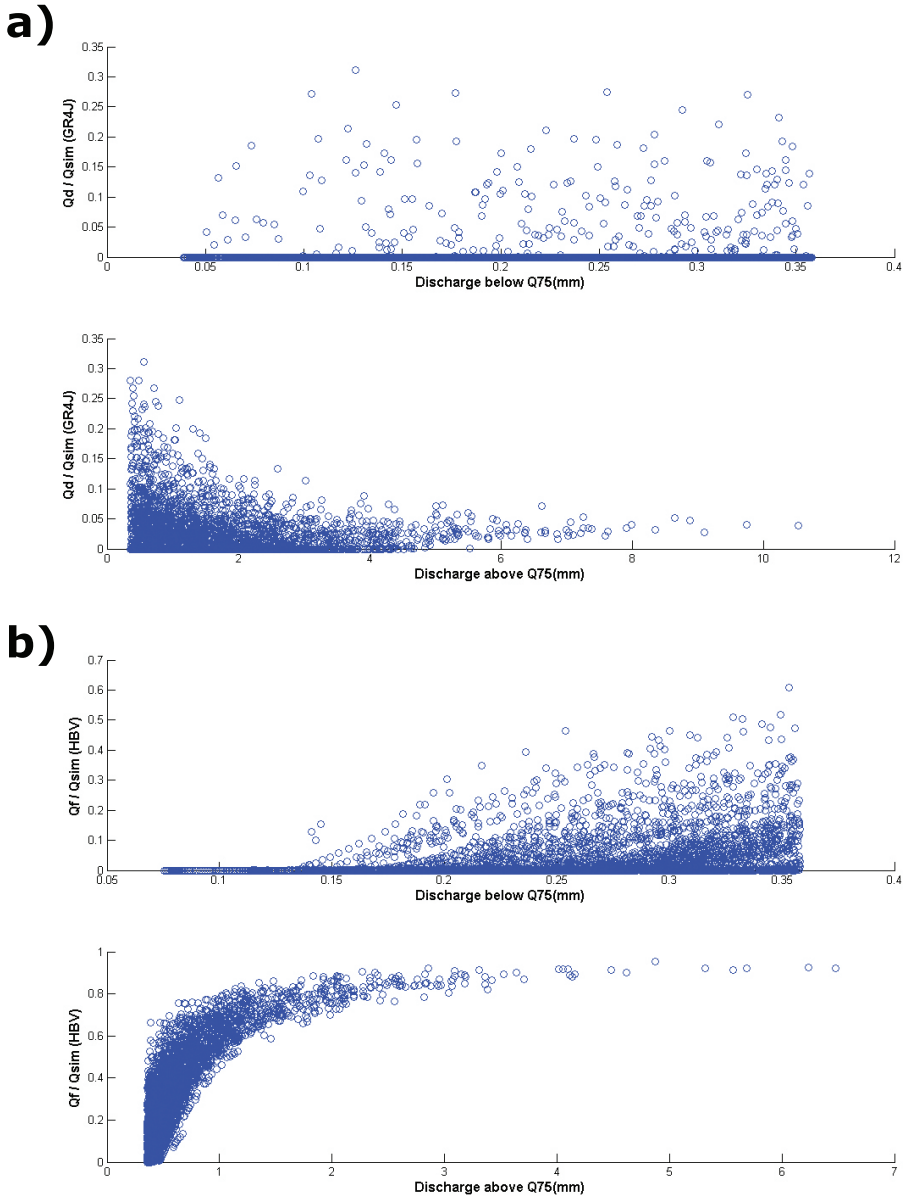


Figure 3.2: Fraction of fast runoff Q_f or Q_d to total runoff Q_{sim} as a function of simulated discharge, i.e. above and below Q75 for **a)** the GR4J and **b)** HBV models

3.4.2 Effect of uncertainty on confidence intervals of low flow forecasts

For the purpose of determining the extent to which different sources of uncertainty affects low flow forecasts for a lead time of 10 days, the degree of uncertainty in model outputs is expressed by a 90% confidence interval (90CI). The 90CI, the forecast median and the observed low flows for both models are shown in Figure 3.3. Daily discharge values (m^3/s) have been presented on a logarithmic y axis. There are significant differences between the two model results as 10 day ahead low flows are mostly overestimated by the GR4J model under uncertain conditions. As can be seen from Figure 3.3 the overestimation is more pronounced for the parameter uncertainty case than for other cases. First thing to be considered are the dependencies and interactions between groundwater storages and model parameters since the fraction of fast runoff to total runoff is about zero showing that the discharge, during low flows, is mainly produced in the groundwater storage (Figure 3.2). The more pronounced overestimation of GR4J compared for the underestimation of HBV may indicate that the slow responding groundwater storage of the HBV is less sensitive to different behavioural parameter sets. The more complex soil moisture and percolation components of the HBV model can also be a reason for the successful low flow forecasts of the HBV model under uncertain conditions. The systematic overestimation of forecasted precipitation is, therefore, well handled by the HBV model. However, it should be noted that the GR4J model is slightly better than HBV in deterministic 10 day forecasts (Table 3.6). Further, the low flows are usually underestimated by the HBV model, as shown in the last plot in Figure 3.3b.

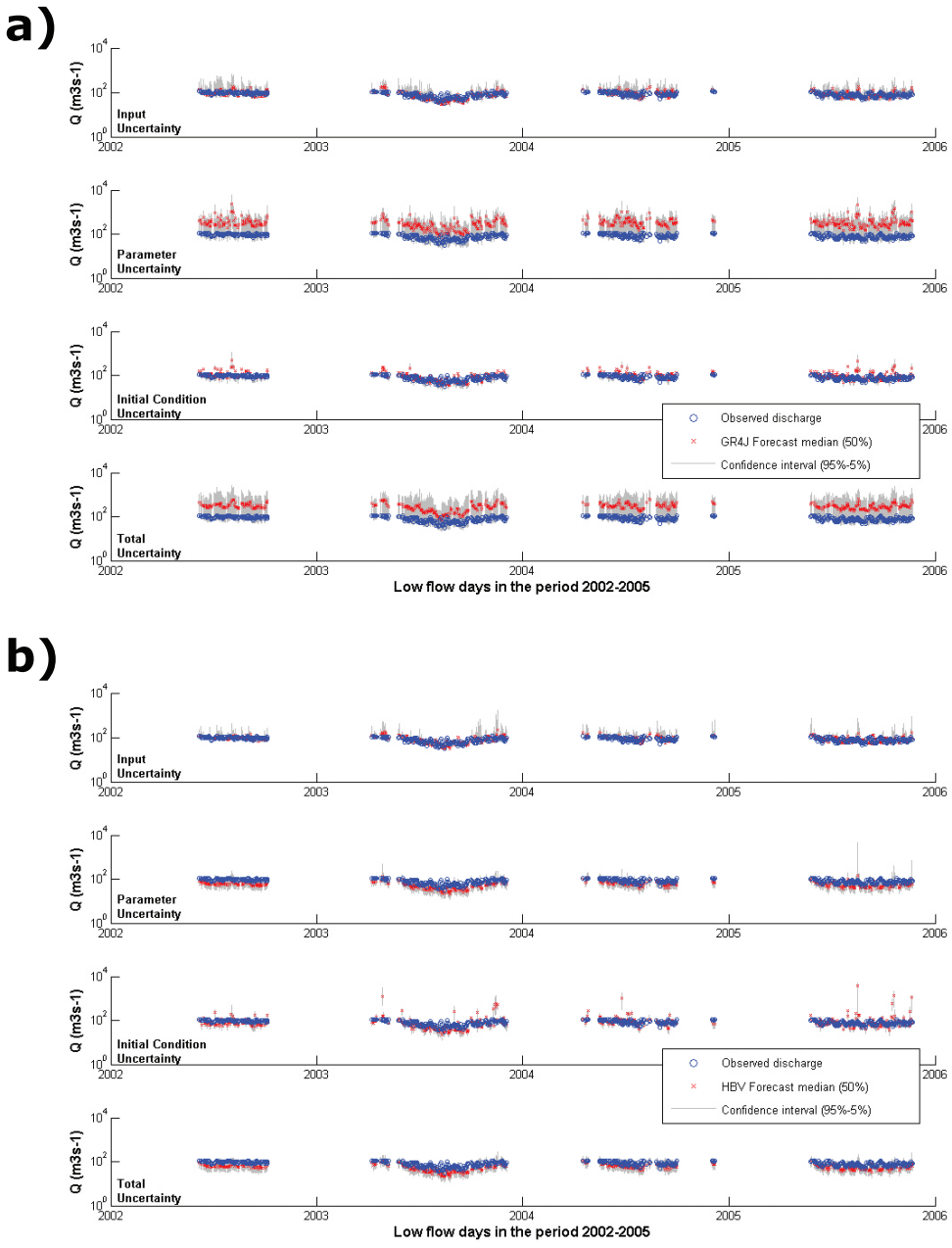


Figure 3.3: Different uncertainty sources and confidence intervals of low flow forecasts for a lead time of 10 days using **a)** the GR4J and **b)** HBV models

Surprisingly, there have been excessive rainfall forecasts for several days in summer months (e.g. in August 2005) causing high forecasted discharges (Figure 3.3). These days were carefully examined to determine if they significantly change the overall RCI results. However, they caused very minor effects on the RCI results that are based on the mean of the CI statistics from 567 low flow days. Thus the uncertainty in 10 day low flow forecasts is larger in the GR4J model compared to the HBV model. The GR4J model overestimates low flows for all sources of uncertainty and for parameter uncertainty in particular, whereas the HBV model tends to underestimate low flows.

Figure 3.4 compares two models and the effect of different uncertainty sources on the RCI of low flow forecasts with increasing lead time. From Figure 3.4, we can clearly see that the total uncertainty in the GR4J outputs is much higher than in the HBV outputs. This is similar to the results that we have seen in Figure 3.3. Comparing only 10 day forecasts issued by the two models, the RCI is ∼110% for the HBV model and ∼300% for the GR4J model (i.e. nearly tripled). One anticipated finding is that the RCI tends to increase with increasing lead time for both models and for all evaluated uncertainty cases. The increase of RCI for the initial condition uncertainty is slowest, showing that the initial condition uncertainty is less sensitive to increasing lead time. This is expected since our storage update procedure only depends on observed discharge and some of the model parameters. However, the uncertainty due to the model inputs (forecast P and PET) increases considerably with increasing lead time. This is from the fact that the error in the ensemble meteorological forecasts increases for longer lead times due to the atmospheric model limitations. It is interesting to note that the 10 day forecasts are even better using zero precipitation as model input i.e. NS_{hybrid} for the GR4J results is increased from 0.45 to 0.54 in the test period from 2002 to 2005.

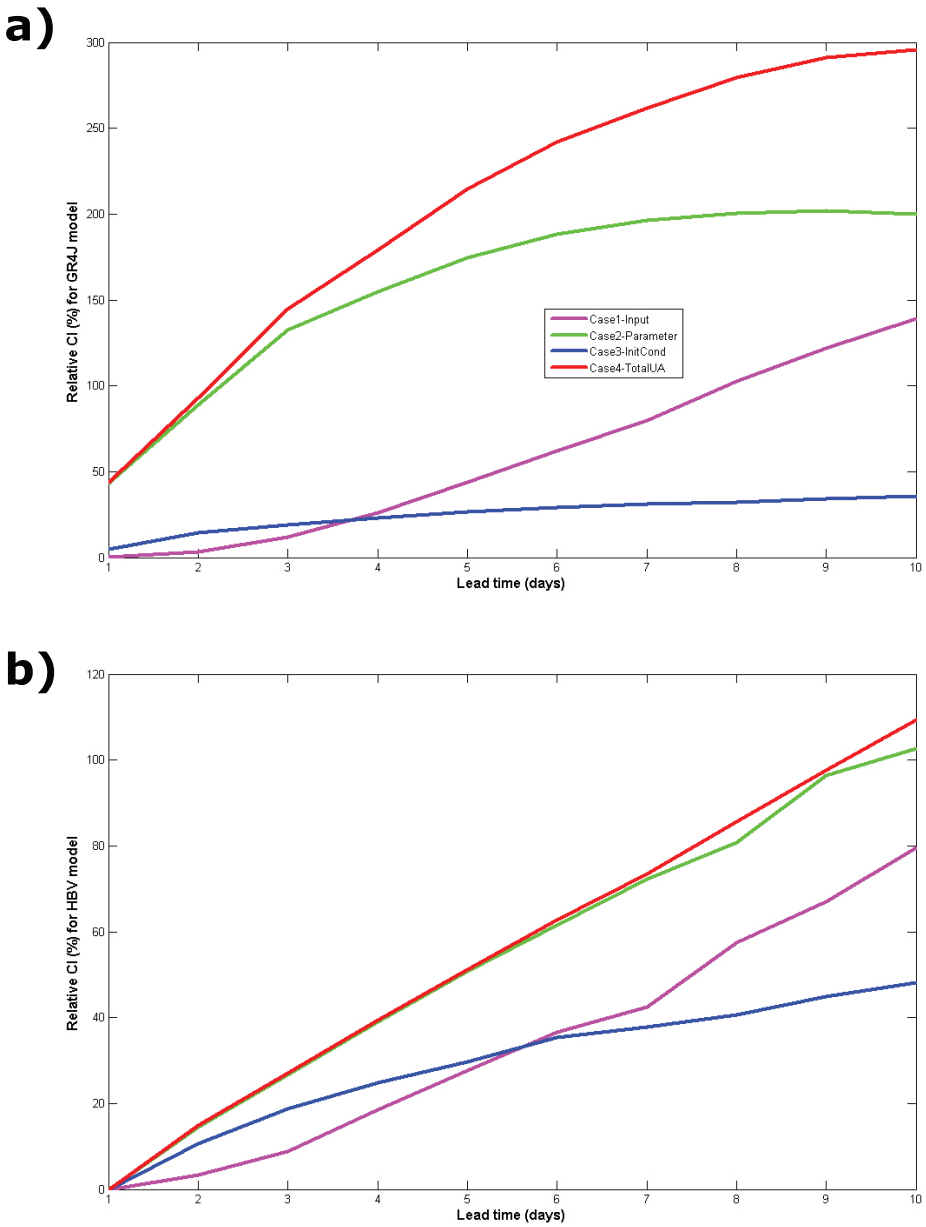


Figure 3.4: Effect of different uncertainty sources on relative confidence intervals of low flow forecasts with **a)** the GR4J and **b)** HBV, as a function of lead time

The total uncertainty for the GR4J model is sum of the three sources of uncertainty assessed in this study. Moreover, only half of the RCI comes from uncertain GR4J parameters. The most striking result to emerge from the RCI results is that the parameter uncertainty is dominating the total uncertainty in the HBV model outputs. It is somewhat surprising that nearly all uncertainty comes from the HBV parameters. Parameter interactions in the HBV model can be the main reason for the unexpected total uncertainty which is not the sum of the three sources of uncertainty. Interestingly, the results, as shown in Figure 3.4, indicate that input uncertainty is even smaller than initial condition uncertainty. On the one hand, this is not expected as ensemble meteorological forecasts are assumed to be one of the most important uncertainty sources in streamflow forecasts (Engeland *et al.*, 2010; Thirel *et al.*, 2008; Vrugt *et al.*, 2008; Zappa *et al.*, 2011). On the other hand, the large range of soil moisture related parameters randomly selected from the GLUE behavioural parameter set could enhance the impact of initial condition uncertainty compared to input uncertainty from only 51 P and PET forecast ensemble members. In other words, the dominating effect of parameter uncertainty, certainly determines the impact of the initial condition uncertainty due to the parameters used in the storage update procedure. Moreover, during low flow periods slow responding processes like groundwater are more dominant than precipitation as low flows usually occur after prolonged dry periods. This can explain the smaller effect of uncertainty from precipitation compared to the uncertainty from initial conditions. The total uncertainty in 10 day lead time low flow forecasts is not a linear sum of three uncertainty sources in the HBV model due to parameter interactions. This finding is in agreement with the findings of Zappa *et al.* (2011) who showed the full spread obtained from uncertainty superposition of three sources is growing non-linearly for a hydrological model (i.e. PREVAH) similar to HBV.

3.4.3 Effect of uncertainty on reliability of low flow forecasts

Figure 3.5 compares the reliability of 10 day ensemble forecasts of low flows for below Q75 and Q95 thresholds using the GR4J and HBV models. The figure exhibits the portion of observed data inside predefined forecast intervals. The reliability plots based on forecasts associated with different uncertainties show that GR4J and HBV over- or under-estimate middle forecast intervals but the narrowest (i.e. 0%–20%) and the 90% intervals are correctly estimated. From Figure 3.5 we see clearly that the ensemble Q75 forecasts issued by HBV including only the input uncertainty are the most reliable forecasts, and it is confirmed that GR4J provides too wide forecast intervals if all sources of uncertainty are

included. The plot for the evaluation of the Q75 forecasts using the HBV model in Figure 3.5 shows that the average overestimation for the total uncertainty is ∼25% inside the highest interval bin (i.e. 80%–100%). The Q75 low flow event will then occur in only ∼75% of the cases when it is forecasted to almost certainly happen, indicating that every fourth low flow warning will be a false alarm. Initial condition uncertainty has less effect than parameter uncertainty on the reliability of Q75 forecasts by the two models. For the Q95 low flows, all intervals except for the narrowest interval (i.e. 0%–20%) are overestimated by the HBV model. A comparison of the four subplots in Figure 3.5 reveals that the parameter uncertainty has a negative effect on the reliability of the forecasts. Moreover, the overestimation of the GR4J model and underestimation of the HBV model are also visible in Figure 3.5. Finally, the forecasts of extreme low flows (Q95) issued by the GR4J model are more reliable than the forecasts by the HBV model.

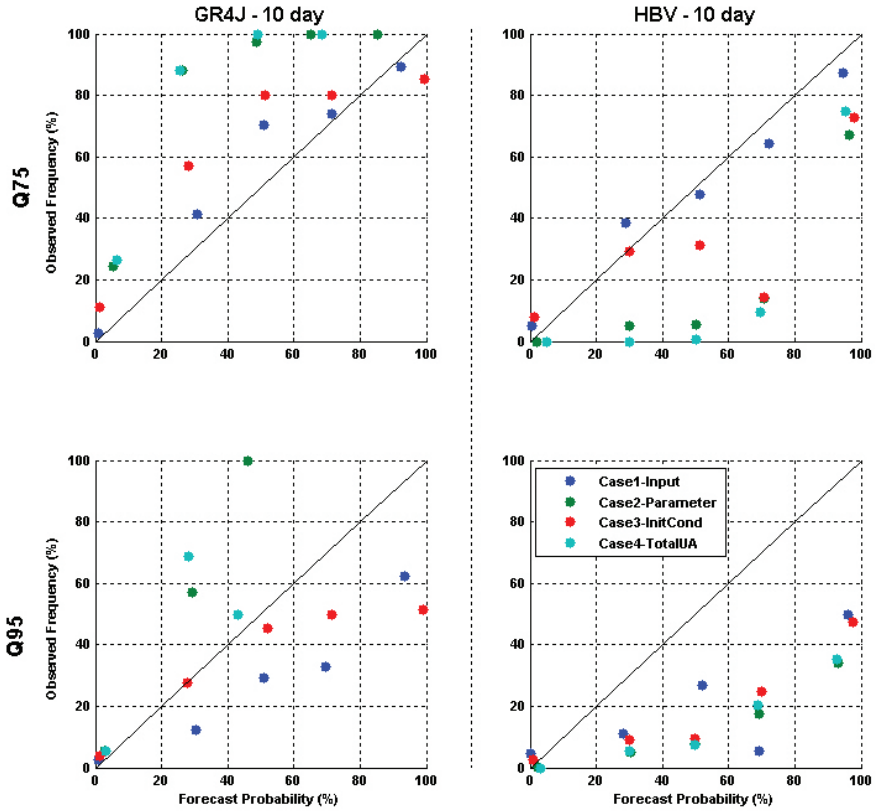


Figure 3.5: Effect of different uncertainty sources on reliability of 10 day low flow forecasts with GR4J and HBV as a function of two low flow thresholds (i.e. Q75 and Q95)

3.4.4 Effect of uncertainty on contingency table of low flow forecasts: hits and false alarms

From operational point of view, the main purpose of investigating uncertainty from 10 day ensemble low flow forecasts is to improve the forecasts (e.g. hits) and to reduce false alarms and missed targets in the low flow contingency measures. Figure 3.6 shows the comparison between the GR4J and HBV models, based on the number of hits, false alarms, misses and correct rejections for the pre-selected Q75 low flow events. It should be noted that a threshold probability of 0.5 is used to issue a low flow forecast alarm from 10.000 forecasts each day in the test period. Subsequently, a low flow event is assumed to occur if more than half of the 10.000 forecasts are low flows. Subsequently, the 10.000×1461 forecast ensemble matrix is transformed to a 1×1461 binary vector consists of zero (no low flow) and one (low flow) values by applying the aforementioned warning threshold of 0.5. This corresponds to a total of 1461 forecasts divided into four subplots in Figure 3.6.

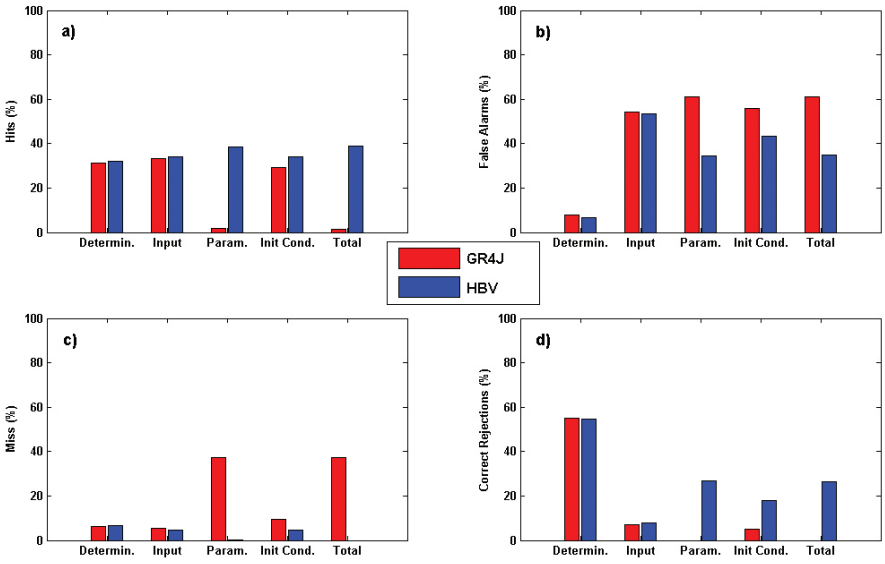


Figure 3.6: Effect of different uncertainty sources on contingency table. Percentage of **a)** hits **b)** false alarms **c)** misses **d)** correct rejections, to total warnings for 10 day ahead forecasts issued in the test period (i.e. 1461 warnings in the test period)

The threshold approach was not necessary for the deterministic run as we run the models only one time with calibrated parameters and control forecasts. The y axes of the subplots show the percentage of different contingency measures for each evaluated uncertainty source, i.e. input, parameter, initial condition and total. We are aware that the contingency table is very sensitive to the pre-selected

threshold and the number of forecasts (Devineni *et al.*, 2008). From Figure 3.6a, we can clearly see that the two models perform similar for the deterministic run whereas the number of hits declines for the GR4J model forecasts for the total uncertainty. This suggests that adding parameter uncertainty to the model certainly reduces the number of hits. This is what we have seen also in Figure 3.3a. In case of the HBV model, there is no drop in the number of hits indicating that most of the low flow events (a total of 567 events in the test period) are correctly forewarned by the ensemble forecasts. This is a significant success for a hydrological model calibrated for low flows. Moreover, non occurrence of low flow events was also correctly indicated by the HBV model. This could be inferred from the correct rejections plot in Figure 3.6d.

The most striking result to emerge from Figure 3.6b is that the percentage of false alarms is highest (∼50%) for the forecasts issued by HBV including only input uncertainty. This may seem contradictory with the results presented in Figure 3.5 as the same ensemble forecasts have been indicated as the most reliable forecasts. However, it should be noted that these two quality measures evaluate the forecasts from totally different aspects namely the reliability diagram for the reliability and the contingency table for the sharpness of the forecasts (WMO, 2012).

Figure 3.7 shows the utility of the low flow forecasts as a function of lead time using the hit rate and the false-alarm rate derived from contingency tables. What is surprising is that the hit rate of GR4J drops significantly from 0.9 to 0.1 by increasing the lead time whereas the hit rate of HBV is slightly increased from 0.9 to 1.

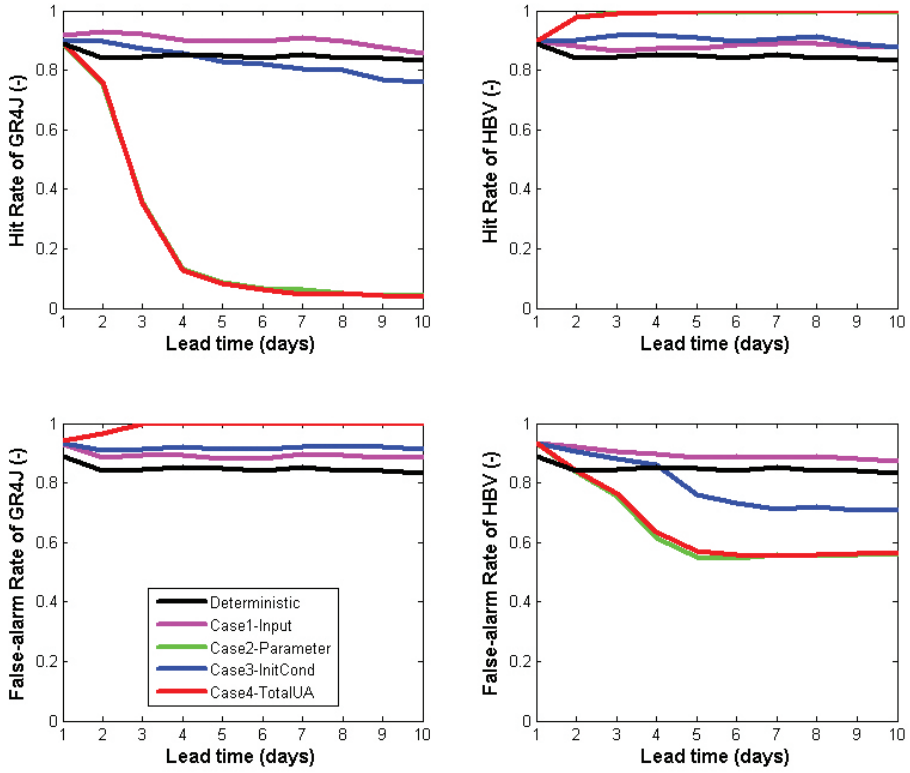


Figure 3.7: Effect of different uncertainty sources on hit/false-alarm rate of two models as a function of different lead times

Another important finding was that the hit rate and false alarm rate of the GR4J and HBV models do not vary significantly as a function of lead time for the deterministic forecasts. Moreover, the false-alarm rate of GR4J does not change considerably by increasing the lead time and for different uncertainty sources. The drop in false-alarm rate is higher for the HBV model. The importance of parameter uncertainty for both models can be clearly seen in Figure 3.7. The effect of the storage updating procedure and input uncertainty on both models outputs is much smaller. The findings of the current study (see Figure 3.7) are consistent with those of Zappa *et al.* (2011) who found slight decreases in the hit rates of low flow forecasts for a leadtime of 1 day and 5 days.

3.5 Conclusions

The performance of two hydrological models in the calibration, validation and forecast periods have been compared, and the effect of different uncertainty sources on the quality of 10 day low flow forecasts have been assessed. We applied a systematic uncertainty analysis to identify where the uncertainty comes from and to provide quantified model output uncertainty information to make a robust model comparison. An hybrid performance metric is used for evaluating low flow simulations whereas the quality of the probabilistic low flow forecasts has been assessed based on relative confidence intervals, reliability and hit/false alarm rates. Based on the results presented in this study we can draw the following conclusions.

- The 10 day ensemble forecast results show that the daily observed low flows are captured by the 90% confidence interval for both models most of the time. However, the GR4J model usually overestimates low flows whereas HBV is prone to underestimate low flows. This is particularly the case if the parameter uncertainty is included into the forecasts.
- The total uncertainty in the GR4J model outputs is higher than in the HBV model.
- The parameter uncertainty has the highest effect and the input uncertainty has the smallest effect on the low flow forecasts.
- A direct relation is found between the number of parameters and the parameter uncertainty according to the RCI results.
- The parameter uncertainty for 10 day low flow forecasts issued by the HBV model with eight parameters is almost half of the parameter uncertainty coming from the GR4J with four parameters. This is because the rainfall-runoff process resulting in low flows in the study area is better described by the HBV model.
- The forecast distribution based on 10 day low flow forecasts (i.e. Q75) issued by the HBV model was the most reliable forecast distribution if only input uncertainty is considered.
- The number of hits is about equal for two models only if the input uncertainty is considered. The parameter uncertainty was the main reason reducing the number of hits.
- The deterministic forecasts using the GR4J and HBV resulted in similar performance indices and also similar hit false alarm rates.

- The performance of the HBV model for correct rejections is remarkable indicating that the model is not only successful for low flows but also correctly indicates other flows above the Q75 threshold while being calibrated on low flows below Q75.
- The number of false alarms is almost doubled for the GR4J model considering total uncertainty. The importance of parameter uncertainty on the quality of forecast is emphasized by all forecast quality measures used in this study.

In essence, this paper has shown that the output from two conceptual hydrological models, calibrated for a medium sized $\sim$27.000 km^2 river basin, fed by raw ECMWF meteorological forecasts, is characterised by substantial uncertainty from model parameters. This source of uncertainty effects both the reliability and the sharpness of the forecasts. This finding is new for low flow forecasts as the significance of the rainfall prediction error is well known and documented for high flows (Pappenberger *et al.*, 2005). This study has taken a step in the direction of assessing major sources of uncertainties in medium range low flow forecasts in addition to flood forecasts for the Moselle River. However, further research has to be conducted on the effect of uncertainties on seasonal low flow forecasts using coarse seasonal weather products. Different types of models, especially data-driven models, may be considered to include in the uncertainty analysis framework for assessing model structure uncertainty explicitly.

3.6 Acknowledgements

We acknowledge the financial support of the Dr. Ir. Cornelis Lely Stichting (CLS), Project No. 20957310. The research is part of the programme of the Water Engineering and Management in the University of Twente and it supports the work of the UNESCO-IHP VII FRIEND-Water programme. Discharge data for the River Rhine were provided by the Global Runoff Data Centre (GRDC) in Koblenz (Germany). Areal precipitation and evapotranspiration data were supplied by the Federal Institute of Hydrology (BfG), Koblenz (Germany). REGNIE grid data were extracted from the archive of the Deutscher Wetterdienst (DWD: German Weather Service), Offenbach (Germany). ECMWF ENS data used in this study have been obtained from the ECMWF data server. We thank Dominique Lucas from ECMWF who kindly guided through the data retrieval process. The GIS base maps with delineated 134 catchments of the Rhine basin were provided by Eric Sprokkereef, the secretary general of the Rhine Commission (CHR). The GR4J and HBV model codes were provided by Tian Ye. We are grateful to the members of the Referat M2 - Mitarbeiter/innen group in BfG, Koblenz, in particular, Peter Krahe, Dennis Meißner, Bastian Klein, Robert Pinzinger, Silke Rademacher and Imke Lingemann for discussions on the value of low flow forecasts, biases in the models and statistical tests adapted to ensemble forecasts. The constructive review comments of Hamid Moradkhani (Associate Editor), Florian Pappenberger, Thomas E. Adams and one anonymous reviewer significantly improved this paper.

Chapter 4

The skill of seasonal ensemble low flow forecasts for four different hydrological models

Abstract

This paper investigates the skill of 90 day low flow forecasts using two conceptual hydrological models and two data-driven models based on Artificial Neural Networks (ANNs) for the Moselle River. One data-driven model, ANN-Indicator (ANN-I), requires historical inputs on precipitation (P), potential evapotranspiration (PET), groundwater (G) and observed discharge (Q), whereas the other data-driven model, ANN-Ensemble (ANN-E), and the two conceptual models, HBV and GR4J, use forecasted meteorological inputs (P and PET), whereby we employ ensemble seasonal meteorological forecasts. We compared low flow forecasts without any meteorological forecasts as input (ANN-I) and five different cases of seasonal meteorological forcing: (1) ensemble P and PET forecasts; (2) ensemble P forecasts and observed climate mean PET; (3) observed climate mean P and ensemble PET forecasts; (4) observed climate mean P and PET and (5) zero P and ensemble PET forecasts as input for the other three models (GR4J, HBV and ANN-E). The ensemble P and PET forecasts, each consisting of 40 members, reveal the forecast ranges due to the model inputs. The five cases are compared for a lead time of 90 days based on model output ranges, whereas the four models are compared based on their skill of low flow forecasts for varying lead times up to 90 days. Before forecasting, the hydrological models are calibrated and validated for a period of 30 and 20 years respectively. The smallest difference between calibration and validation performance is found for HBV, whereas the largest difference is found for ANN-E. From the results, it appears that all

models are prone to over-predict low flows using ensemble seasonal meteorological forcing. The largest range for 90 day low flow forecasts is found for the GR4J model when using ensemble seasonal meteorological forecasts as input. GR4J, HBV and ANN-E under-predicted 90 day ahead low flows in the very dry year 2003 without precipitation data, whereas ANN-I predicted the magnitude of the low flows better than the other three models. The results of the comparison of forecast skills with varying lead times show that GR4J is less skilful than ANN-E and HBV. Furthermore, the hit rate of ANN-E is higher than the two conceptual models for most lead times. The seasonal low flows are correctly forecasted by ANN-I, showing the ability of data-driven models in forecasting the magnitude of low flows. However, ANN-I is not successful in distinguishing between low flow events and non-low flow events. Overall, the uncertainty from ensemble P forecasts has a larger effect on seasonal low flow forecasts than the uncertainty from ensemble PET forecasts and initial model conditions.

This chapter is based on the paper: Demirel MC, Booij MJ, Hoekstra AY. (2013) The skill of seasonal ensemble low flow forecasts for four different hydrological models (*submitted*).

4.1 Introduction

Rivers in Western Europe usually experience low flows in late summer and high flows in winter. These two extreme discharge phenomena can lead to serious problems. For example, high flow events are quick and can put human life at risk, whereas streamflow droughts (i.e. low flows) develop slowly and can affect a large area. Consequently, the economic loss during low flow periods can be much bigger than during floods (Pushpalatha *et al.*, 2011; Shukla *et al.*, 2012). In the River Rhine, severe problems for freshwater supply, water quality, power production and river navigation were experienced during the dry summers of 1976, 1985 and 2003. Therefore, forecasting seasonal low flows (Coley and Waylen, 2006; Li *et al.*, 2008; Towler *et al.*, 2013) and understanding low flow indicators (Demirel *et al.*, 2013a; Fundel *et al.*, 2013; Saadat *et al.*, 2013; Vidal *et al.*, 2010; Wang *et al.*, 2011) have both societal and scientific value. The seasonal forecast of water flows is therefore listed as one of the priority topics in EU's Horizon 2020 research program (EU, 2013). Further, there is an increasing interest to incorporate seasonal flow forecasts in decision support systems for river navigation and power plant operation during low flow periods. We are interested in forecasting low flows with a lead time of 90 days, and in presenting the effect of ensemble meteorological forecasts for four hydrological models.

Generally, two approaches are used in seasonal hydrological forecasting. The first one is a statistical approach, making use of data-driven models based on relationships between river discharge and hydroclimatological indicators (Van Ogtrop *et al.*, 2011; Wang *et al.*, 2011). The second one is a dynamic approach running a hydrological model with forecasted climate input. The first approach is often preferred in regions where significant correlations between river discharge and climatic indicators exist, such as sea surface temperature anomalies (Chowdhury and Sharma, 2009), AMO - Atlantic Multi-decadal Oscillation (Ganguli and Reddy, 2013; Giuntoli *et al.*, 2013), PDO - Pacific Decadal Oscillation (Soukup *et al.*, 2009) and warm and cold phases of the ENSO - El Niño Southern Oscillation - index (Chiew *et al.*, 2003; Kalra *et al.*, 2013; Tootle and Piechota, 2004). Kahya and Dracup (1993) identified the lagged response of regional streamflow to the warm phase of ENSO in the south-eastern United States. In the Rhine basin, no teleconnections have been found between climatic indices, e.g. NAO and ENSO, and river discharges (Bierkens and van Beek, 2009; Rutten *et al.*, 2008). However, Demirel *et al.* (2013a) found significant correlations between hydrological low flow indicators (e.g. standard snow storage and groundwater indices) and observed low flows. They also identified appropriate lags and temporal resolutions of low flow indicators to build data-driven models.

The dynamic seasonal forecasting approach has long been explored (Fundel *et al.*, 2013; Gobena and Gan, 2010; Pokhrel *et al.*, 2013; Shukla *et al.*, 2013; Van Dijk *et al.*, 2013; Wang *et al.*, 2011), which has led to the development of the current ensemble streamflow prediction system (ESP) used by different national climate services like the National Weather Service in the United States. The seasonal hydrologic prediction systems are most popular in regions with a high risk of extreme discharge situations like hydrological droughts (Robertson *et al.*, 2013). Well-known examples are the NOAA Climate Prediction Centre's seasonal drought forecasting system (Shukla and Lettenmaier, 2011), the University of Washington's Surface Water Monitoring system (Wood and Lettenmaier, 2006), Princeton University's drought forecast system (Shukla *et al.*, 2012) and University of Utrecht's global monthly hydrological forecast system (Yossef *et al.*, 2012). These models provide indications about the hydrologic conditions and their evolution across the modelled domain using available weather ensemble inputs (Gobena and Gan, 2010; Yossef *et al.*, 2012). Many studies have investigated the seasonal predictability of low flows in different rivers such as the Thames and different other rivers in the UK (Bell *et al.*, 2013; Wedgbrow *et al.*, 2002; Wedgbrow *et al.*, 2005), the Shihmen and Tsengwen Rivers in Taiwan (Kuo *et al.*, 2010), the River Jhelum in Pakistan (Archer and Fowler, 2008), more than 200 rivers in France (Giuntoli *et al.*, 2013; Sauquet *et al.*, 2008), five semi-arid areas in South Western Queensland, Australia (Van Ogtrop *et al.*, 2011), four rivers including the Blue Nile in Africa (Dutra *et al.*, 2013), the Bogotá River in Colombia (Felipe and Nelson, 2009), the Ohio in the eastern US (Li *et al.*, 2009; Luo *et al.*, 2007; Wood *et al.*, 2002), the North Platte in Colorado, US (Soukup *et al.*, 2009), large rivers in the US (Schubert *et al.*, 2007; Shukla and Lettenmaier, 2011) and the Thur River in the north-eastern part of Switzerland (Fundel *et al.*, 2013). The common result of the above mentioned studies is that seasonal forecasts made with global and regional hydrological models are limited to lead times of 1-3 months (Shukla and Lettenmaier, 2011; Wood *et al.*, 2002) and these forecasting systems are all prone to large uncertainties as their forecast skills mainly depend on the knowledge of initial hydrologic conditions and weather information during the forecast period (Doblas-Reyes *et al.*, 2009; Li *et al.*, 2009; Shukla *et al.*, 2012; Yossef *et al.*, 2013). In a recent study, Yossef *et al.* (2013) used a global monthly hydrological model to analyse the relative contributions of initial conditions and meteorological forcing to the skill of seasonal streamflow forecasts. They included 78 stations in large basins in the world including the River Rhine for forecasts with lead times up to 6 months. They found that improvements in seasonal hydrological forecasts in the Rhine depend on better meteorological forecasts, which underlines the importance of meteorological forcing quality particularly for forecasts beyond

lead times of 1-2 months.
Most of the previous River Rhine studies use only one hydrological model, e.g. PREVAH (Fundel *et al.*, 2013) or PCR-GLOBWB (Yossef *et al.*, 2013), to assess the value of ensemble meteorological forcing, whereas in this study, we compare four hydrological models with different structures varying from data-driven to conceptual models. The objective of this study is to assess the effect of ensemble seasonal forecasted precipitation and potential evapotranspiration on low flow forecast quality and skill scores. By comparing four models with different model structures we address the use of model structure uncertainty. Moreover, the effect of initial model conditions is partly addressed using climate mean data in one of the cases.
The analysis complements recent efforts to analyse the effects of ensemble weather forecasts on low flow forecasts with a lead time of 10 days using two conceptual models (Demirel *et al.*, 2013b), by studying the effects of seasonal ensemble weather forecasts on 90 day low flow forecasts using not only conceptual models but also data-driven models.
The outline of the paper is as follows. The study area and data are presented in section 4.2. Section 4.3 describes the model structures, their calibration and validation set-ups and the methods employed to estimate the different attributes of the forecast quality. The results are presented in section 4.4 and discussed in section 4.5, and the conclusions are summarised in section 4.6.

4.2 Study area and data

4.2.1 Study area

The study area is the Moselle River basin, the largest sub-basin of the Rhine River basin. The Moselle River has a length of 545 km. The river basin has a surface area of approximately 27,262 km^2. The altitude in the basin varies from 59 to 1326 m, with a mean altitude of 340 m (Demirel *et al.*, 2013a). Approximately 410 mm ($\sim$130 m^3/s) discharge is annually generated in the Moselle basin (Demirel *et al.*, 2013b). The outlet discharge at Cochem varies from 14 m^3/s in dry summers to a maximum of 4000m^3/s during winter floods.

4.2.2 Data

Observed data

Observed daily data on precipitation (P) and potential evapotranspiration (PET) are obtained from the German Federal Institute of Hydrology (BfG) in Koblenz, Germany (Table 4.1). PET is estimated using the Penman-Wendling equation (ATV-DVWK, 2002) and both variables have been spatially averaged by BfG over 26 Moselle sub-basins using areal weights. Observed daily discharge (Q) data at Cochem (station *#6336050*) are provided by the Global Runoff Data Centre (GRDC), Koblenz. The daily observed data (P, PET and Q) are available for the period 1951-2006.

Table 4.1: Overview of observed data used

Variable	Name	Number of stations/sub-basins	Period	Time step (days)	Spatial resolution	Source
Q	Discharge	1	1951-2006	1	Point	GRDC
P	Precipitation	26	1951-2006	1	Basin average	BfG
PET	Evapotranspiration	26	1951-2006	1	Basin average	BfG
h	Mean altitude	26	-	-	Basin average	BfG

Ensemble seasonal meteorological forecast data

The ensemble seasonal meteorological forecast data, comprising 40 members, are obtained from the European Centre for Medium-Range Weather Forecasts (ECMWF) seasonal forecasting system (ECMWF, 2012). This dataset contains regular 0.25×0.25 degree latitude-longitude grids and each ensemble member is computed for a lead time of 184 days using perturbed initial conditions and model physics (Table 4.2). We estimated the PET forecasts using the Penman-Wendling equation requiring forecasted surface solar radiation, temperature at

2 meter above surface and altitude of the sub-basin (ATV-DVWK, 2002). The mean altitudes of the 26 sub-basins are provided by BfG in Koblenz, Germany. The PET estimation is consistent with the observed PET estimation carried out by BfG. The grid-based P and PET ensemble forecast data are firstly interpolated over 26 Moselle sub-basins using areal weights. These sub-basin averaged data are then aggregated to the Moselle basin level.

Table 4.2: Overview of ensemble seasonal meteorological forecast data

Data	**Spatial resolution**	**Ensemble size**	**Period**	**Time step (days)**	**Lead time (days)**
Forecasted P	0.25 x 0.25 degree	39 + 1 control	2002-2005	1	1-90
Forecasted PET	0.25 x 0.25 degree	39 + 1 control	2002-2005	1	1-90

4.3 Methodology

4.3.1 Overview of model structures and forecast scheme

The four hydrological models (GR4J, HBV, ANN-E and ANN-I), calibration and validation of the models and five cases with different combinations of ensemble meteorological forecast input and climate mean input are briefly described in sections below. Figure 4.1 shows the simplified model structures.

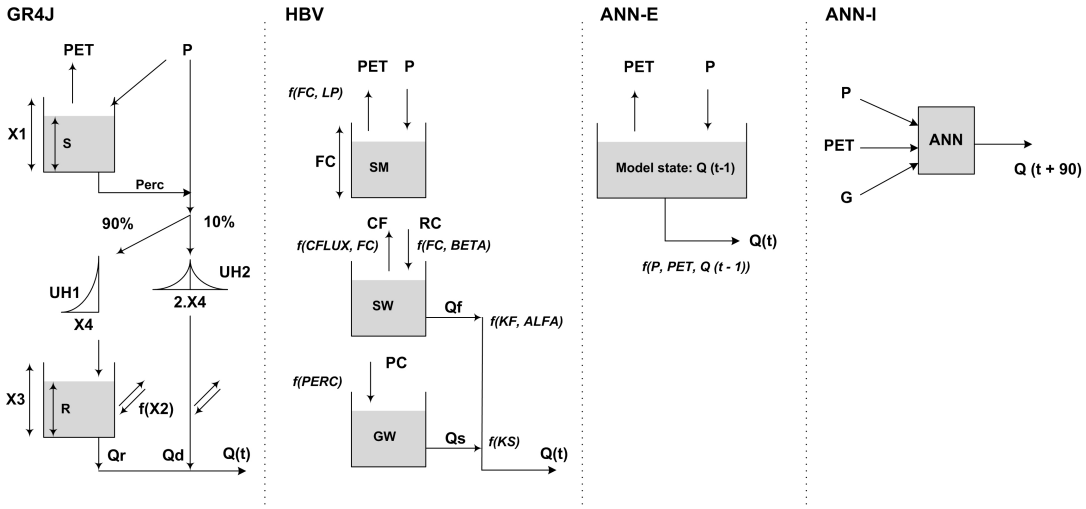


Figure 4.1: Schematisation of the four models. PET is potential evapotranspiration, P is precipitation, G is groundwater, Q is discharge and t is the time (day).

GR4J

The GR4J model (Génie Rural à 4 parametres Journalier) is used as it has a parsimonious structure with only four parameters. The model has been tested over hundreds of basins worldwide, with a broad range of climatic conditions from tropical to temperate and semi-arid basins (Perrin *et al.*, 2003). GR4J is a conceptual model and the required model inputs are daily time series of P and PET (Table 4.3). The four parameters in GR4J represent the maximum capacity of the production store ($X1$), the groundwater exchange coefficient ($X2$), the one day ahead capacity of the routing store ($X3$) and the time base of the unit hydrograph ($X4$). All four parameters (Figure 4.1a) are used to calibrate the model. The upper and lower limits of the parameters are selected based on previous works (Perrin *et al.*, 2003; Pushpalatha *et al.*, 2011; Tian *et al.*, 2013).

HBV

The HBV conceptual model (Hydrologiska Byråns Vattenbalansavdelning) was developed by the Swedish Meteorological and Hydrological Institute (SMHI) in the early 1970s (Lindström *et al.*, 1997). The HBV model consists of four subroutines: a precipitation and snow accumulation and melt routine, a soil moisture accounting routine and two runoff generation routines. The required input data are daily P and PET. The snow routine and daily temperature data are not used in this study as the Moselle basin is a rain-fed basin. Eight parameters (see Figure 4.1b) in the HBV model are calibrated (Engeland *et al.*, 2010; Tian *et al.*, 2013; Van den Tillaart *et al.*, 2013). The ranges of the eight parameters for calibration are selected based on previous works (Booij, 2005; Eberle, 2005; Tian *et al.*, 2013).

ANN-E and ANN-I

An Artificial Neural Network (ANN) is a data-driven model inspired by functional units (neurons) of the human brain (Elshorbagy *et al.*, 2010). A neural network is a universal approximator capable of learning the patterns and relation between outputs and inputs from historical data and applying it for extrapolation (Govindaraju and Rao, 2000). Three-layer feed-forward neural networks (FNNs) are the most widely preferred model architecture for prediction and forecasting of hydrological variables (Adamowski *et al.*, 2012; Kalra *et al.*, 2013; Shamseldin, 1997). Each of these three layers has an important role in processing the information. The first layer receives the inputs and multiplies them with a weight (adds a bias if necessary) before delivering them to each of the hidden neurons in the next layer (Gaume and Gosset, 1999). The weights determine the strength of

the connections. The number of nodes in this layer corresponds to the number of inputs. The second layer, the hidden layer, consists of an activation function (also known as transfer function) which non-linearly maps the inputs data to output target values. In other words, this layer is the learning element of the network which simulates the relationship between inputs and outputs of the model. The third layer, the output layer, gathers the processed data from the hidden layer and delivers the final output of the network.
A hidden neuron is the processing element with n inputs (x_1, x_2, x_3, ..., x_n), and one output y using Eq. (4.1).

$$y = f(x_1, x_2, x_3, \ldots, x_n) = logsig\left[\left(\sum_{i=1}^{n} x_i w_i\right) + b\right] \tag{4.1}$$

where w_i are the weights, b is the bias, and *logsig* is the logarithmic sigmoid activation function. Based on our preliminary tests, the *logsig* activation function is selected for this study. ANN model structures are determined based on the forecast objective. In this study, we used two different ANN model structures: ANN-Ensemble (ANN-E) and ANN-Indicator (ANN-I). The first model, i.e. ANN-E, requires daily P, PET and historical Q as input. Historical Q from the previous day is used to update the model states (Table 4.3). This is a one day memory which also exists in the conceptual models, i.e. GR4J and HBV (Figure 4.1). The ANN-E is assumed to be comparable with the conceptual models with similar model structures. The second model, ANN-I, uses historical Q to update initial model conditions and three low flow indicators, i.e. P, PET and G, as model input. The model uses historical data and does not require forecasted weather inputs. The appropriate lags and temporal resolutions of these indicators have been identified using the discharge data for the period of 1978-2006 in a previous study by Demirel *et al.* (2013a). The determination of the optimal number of hidden neurons in the second layer is an important issue in the development of ANN models. Three common approaches are ad hoc (also known as trial and error), global and stepwise (Kasiviswanathan *et al.*, 2013). We used a global approach (i.e. Genetic Algorithm) (De Vos and Rientjes, 2008) and tested the performance of the networks with one, two and three hidden neurons corresponding to a number of parameters (i.e. number of weights and biases) of 6, 11 and 16 respectively. Based on the parsimonious principle, testing ANNs only up to three hidden neurons is assumed to be enough as the number of parameters increases exponentially for every additional hidden neuron.

Table 4.3: Model descriptions. PET is potential evapotranspiration, P is precipitation, G is groundwater, Q is discharge and t is the time (day).

Model Type		Input	Temporal resolution of input	Lag between forecast issue day and final day of temporal averaging (days)	Model time step	Model lead time (days)
Conceptual	**Data-driven**					
GR4J		P: Ensemble PET: Ensemble Q: State update	Daily P Daily PET	P: 0 PET: 0 Q: 1	Daily	1 to 90
HBV		P: Ensemble PET: Ensemble Q: State update	Daily P Daily PET	P: 0 PET: 0 Q: 1	Daily	1 to 90
ANN-E		P: Ensemble PET: Ensemble Q: State update	Daily P Daily PET Daily Q	P: 0 PET: 0 Q: 1	Daily	1 to 90
	ANN-I	P: Observed PET: Observed G: Observed	110-day mean P 180-day mean PET 90-day mean G	P: 0 PET: 210 G: 210	Daily	90

Calibration and validation of models

A global optimisation method, i.e. Genetic Algorithm (GA) (de Vos and Rientjes, 2008), and historical Moselle low flows for the period from 1971-2001 are used to calibrate the models used in this study. The 30-year calibration period is carefully selected as the first low flow forecast is issued on 01/01/2002. For all GA simulations, we use 100 as population size, 5 as reproduction elite count size, 0.7 as cross over fraction, 2000 as maximum number of iterations and 5000 as the maximum number of function evaluations based on the studies by De Vos and Rientjes (2008) and Kasiviswanathan *et al.* (2013). The validation period spans from 1951-1970. The definition of low flows, i.e. discharges below the Q75 threshold of $\sim$ 113 m^3/s, is based on previous work by Demirel *et al.* (2013a). Prior parameter ranges and deterministic equations used for dynamic model state updates of the conceptual models based on observed discharges on the forecast issue day are based on the study by Demirel *et al.* (2013b). In this study, we use a hybrid Mean Absolute Error (MAE) based on only low flows (MAE_{low}) and inverse discharge values ($MAE_{inverse}$) as objective function (see Eq. (4.4)).

$$MAE_{low} = \frac{1}{m}\sum_{j=1}^{m}|Q_{sim}(j) - Q_{obs}(j)| \tag{4.2}$$

where Q_{obs} and Q_{sim} are the observed and simulated values for the j-th observed low flow day (i.e. $Q_{obs} < Q_{75}$) and m is the total number of low flow days.

$$MAE_{inverse} = \frac{1}{n}\sum_{i=1}^{n}|\frac{1}{Q_{sim}(i)+\varepsilon} - \frac{1}{Q_{obs}(i)+\varepsilon}| \tag{4.3}$$

where n is the total number of days (i.e. $m < n$), and ε is 1% of the mean observed discharge to avoid infinity during zero discharge days.

$$MAE_{hybrid} = MAE_{low} + MAE_{inverse} \tag{4.4}$$

Case description

In this study, four hydrological models are used for the seasonal forecasts. While only historical input is used for the ANN-I model, five ensemble meteorological forecast input cases for ANN-E, GR4J and HBV models are compared: (1) ensemble P and PET forecasts (2) ensemble P forecasts and observed climate mean PET (3) observed climate mean P and ensemble PET forecasts (4) observed climate mean P and PET (5) zero P and ensemble PET forecasts (Table 4.4).

Cases 1-4 are the different possible combinations of ensemble and climate mean meteorological forcing. Case 5 is analysed to determine to which extent the precipitation forecast in a very dry year (2003) is important for seasonal low flow forecasts.

Table 4.4: Details of the five input cases

Case	Precipitation	Potential evapotranspiration
1	Ensemble forecast	Ensemble forecast
2	Ensemble forecast	Climate mean
3	Climate mean	Ensemble forecast
4	Climate mean	Climate mean
5	Zero	Ensemble forecast

4.3.2 Forecast skill scores

Three probabilistic forecast skill scores (Brier Skill Score, reliability diagram, hit and false alarm rates) and one deterministic forecast skill score (Mean Forecast Score) are used to analyse the results of low flow forecasts with lead times of 1-90 days. Forecasts for each day in the test period (2002-2005) are used to estimate these scores. The Mean Forecast Score focusing on low flows is introduced in this study, whereas the other three scores have been often used in meteorology (WMO, 2012) and flood hydrology (Renner *et al.*, 2009; Thirel *et al.*, 2008; Velázquez *et al.*, 2010). For the three models, i.e. GR4J, HBV and ANN-E, the forecast probability for each forecast day is estimated as the ratio of the number of ensemble members exceeding the pre-selected thresholds (here Q75) and the total number of ensemble members (i.e. 40 members) for that forecast day. The ANN-I model issues single deterministic forecast; therefore, the probability for each forecast day is either zero or one.

Brier skill score

The Brier Skill Score (BSS) (Wilks, 1995) is often used in hydrology to evaluate the quality of probabilistic forecasts (Devineni *et al.*, 2008; Hartmann *et al.*, 2002; Jaun and Ahrens, 2009; Roulin, 2007; Towler *et al.*, 2013).

$$BSS = 1 - \frac{BS_{forecast}}{BS_{climatology}} \tag{4.5}$$

where the $BS_{forecast}$ is the Brier Score (BS) for the forecast, defined as:

$$BS = \frac{1}{N}\sum_{t=1}^{N}(F_t - O_t)^2 \tag{4.6}$$

where F_t refers to the forecast probability, O_t refers to the observed probability (O_t=1 if the observed flow is below the low flow threshold, 0 otherwise), and N is the sample size. $BS_{climatology}$ is the BS for climatology, which is also calculated from Eq. (4.6) for every year using climatological probabilities. BSS values range from minus infinity to 1 (perfect forecast). Negative values indicate that the forecast is less accurate than the climatology. BSS=0 indicates equal skill, and a BSS>0 indicates more skill compared to the climatology.

Reliability diagram

The reliability diagram is used to evaluate the performance of probabilistic forecasts of selected events, i.e. low flows. A reliability diagram represents the observed relative frequency as a function of forecasted probability and the 1:1 diagonal shows the perfect reliability line (Olsson and Lindström, 2008; Velázquez *et al.*, 2010). This comparison is important as reliability is one of the three properties of a hydrological forecast (WMO, 2012). A reliability diagram shows the portion of observed data inside pre-selected forecast intervals.
In this study, non-exceedence probabilities of 50%, 75%, 85%, 95%, and 99% are chosen as thresholds to categorize the discharges from mean flows to extreme low flows. The forecasted probabilities are then divided into bins of probability categories; here, five bins (categories) are chosen 0-20%, 20%-40%, 40%-60%, 60%-80% and 80%-100%. The observed frequency for each day is chosen to be 1 if the observed discharge non-exceeds the threshold, or 0, if not.

Hit and false alarm rates

We used hit and false alarm rates to assess the effect of ensembles on low flow forecasts for varying lead times. The hit and false alarm rates indicate respectively the proportion of events for which a correct warning was issued, and the proportion of non events for which a false warning was issued by the forecast model. These two simple rates can be easily calculated from contingency tables (Table 4.5) using Eq. (4.7) and (4.8). These scores are often used for evaluating flood forecasts (Martina *et al.*, 2006), however, they can also be used to estimate the utility of low flow forecasts as they indicate the models' ability to correctly forecast the occurrence or non-occurrence of pre-selected events (i.e. Q75 low flows). There are four cases in a contingency table as shown in Table 4.5.

Table 4.5: Contingency table for the assessment of Q75 forecasts

	Observed	**Not observed**
Forecasted	*hit*: the event forecasted to occur and did occur	*false alarm*: event forecasted to occur, but did not occur
Not forecasted	*miss*: the event forecasted not to occur, but did occur	*correct negative*: event forecasted not to occur and did not occur

$$hit\ rate = \frac{hits}{(hits + misses)} \tag{4.7}$$

$$false\ alarm\ rate = \frac{false\ alarms}{(correct\ negatives + false\ alarms)} \tag{4.8}$$

Mean forecast score

The Mean Forecast Score (MFS) is a new skill score which can be derived from either probabilistic or deterministic forecast probabilities. These probabilities are calculated only for the days that low flows occurred. Table 4.6 shows the low flow contingency table for calculating MFS. In this study we used a deterministic approach for calculating the observed frequency for all four models. However, a deterministic approach for calculating the forecast probability is used only for the ANN-I model. For the other three models, ensembles are used for estimating forecast probabilities.

Table 4.6: Low flow contingency table for the assessment of forecasts

		Observed low flows	
		Deterministic	**Probabilistic**
Forecasted low flows	**Deterministic**	*O_j=1 (Low flow observed)* *F_j=1 (Low flow forecasted if more than half of the ensemble members indicate low flows) otherwise 0*	*O_j= Observed frequency based on long term climate (e.g. 34/50 years indicates low flow for day j)* *F_j=1 or 0*
	Probabilistic	$O_j = 1$ *F_j=Forecast frequency based on 40 ensemble members (e.g. 23/40 members indicate low flows for day j)*	*O_j= Observed frequency based on long term climate* *F_j=Forecast frequency based on 40 ensemble members*

The score is calculated as below for only for deterministic observed low flows (left column in Table 4.6).

$$MFS = \frac{1}{m} \sum_{j=1}^{m} F_j \tag{4.9}$$

where F_j is the forecast probability for the j-th observed low flow day (i.e. $O_j \leq Q_{75}$) and m is the total number of low flow days. For instance, if 23 of the 40 ensemble forecast members indicate low flows for the j-th low flow day then F_j = 23/40. It should be noted that this score is not limited to low flows as it has a flexible forecast probability definition which can be adapted to any type of discharges. MFS values range from zero to 1 (perfect forecast).

4.4 Results

4.4.1 Calibration and validation

Table 4.7 shows the parameter ranges and the best performing parameter sets of the four models. The GR4J and HBV models have both well-defined model structures; therefore, their calibration was more straightforward than the calibration of the ANN models. Calibration of the ANN models was done in two steps. First, the number of hidden neurons was determined by testing the performance of the ANN-E model with one, two and three hidden neurons.
The daily P, PET and Q are used as three inputs for the tested ANN-E model with one, two and three hidden neurons due to the fact that these inputs are comparable with the inputs of the GR4J and HBV models. Figure 4.2a shows that the performance of the ANN-E models does not improve with additional hidden neurons. Based on the performance in the validation period, one hidden neuron is selected. Second, GR4J, HBV and ANN-I are calibrated. Based on the results of the first step, ANN-I with one hidden neuron is calibrated for its long term averaged inputs. The results of the four models used in this study are presented in Figure 4.2b.
The performances of GR4J and HBV are similar in the calibration period, whereas HBV performs better in the validation period (Figure 4.2b). This is not surprising, since HBV has a more sophisticated model structure than GR4J. The performance of ANN-E and ANN-I is similar in both calibration and validation periods.

Table 4.7: Parameter ranges and calibrated values of the pre-selected four models

Parameter	Unit	Range	Calibrated value	Description
			GR4J model	
X1	[mm]	10-2000	461.4	Capacity of the production store
X2	[mm]	-8 to +6	-0.3	Groundwater exchange coefficient
X3	[mm]	10-500	80.8	One day ahead capacity of the routing store
X4	[d]	0-4	2.2	Time base of the unit hydrograph
			HBV model	
FC	[mm]	200-800	285.1	Maximum soil moisture capacity
LP	[-]	0.1-1	0.7	Soil moisture threshold for reduction of evapotranspiration
BETA	[-]	1-6	2.2	Shape coefficient
CFLUX	[mm/d]	0.1-1	1.0	Maximum capillary flow from upper response box to soil moisture zone
ALFA	[-]	0.1-3	0.4	Measure for non-linearity of low flow in quick runoff reservoir
KF	[d^{-1}]	0.005-0.5	0.01	Recession coefficient for quick flow reservoir
KS	[d^{-1}]	0.0005-0.5	0.01	Recession coefficient for base flow reservoir
PERC	[mm/d]	0.3-7	0.6	Maximum flow from upper to lower response box
			ANN-E model	
W1	[-]	-10 to +10	-2.3	Weight of connection between 1st input node and hidden neuron
W2	[-]	-10 to +10	0.03	Weight of connection between 2nd input node and hidden neuron
W3	[-]	-10 to +10	-0.02	Weight of connection between 3rd input node and hidden neuron
W4	[-]	-10 to +10	3.7	Weight of connection between 4th input node and hidden neuron
B1	[-]	-10 to +10	0.02	Bias value in hidden layer
B2	[-]	-10 to +10	1.1	Bias value in output layer
			ANN-I model	
W1	[-]	-10 to +10	0.4	Weight of connection between 1st input node and hidden neuron
W2	[-]	-10 to +10	0.9	Weight of connection between 2nd input node and hidden neuron
W3	[-]	-10 to +10	0.9	Weight of connection between 3rd input node and hidden neuron
W4	[-]	-10 to +10	0.6	Weight of connection between 4th input node and hidden neuron
B1	[-]	-10 to +10	0.001	Bias value in hidden layer
B2	[-]	-10 to +10	0.3	Bias value in output layer

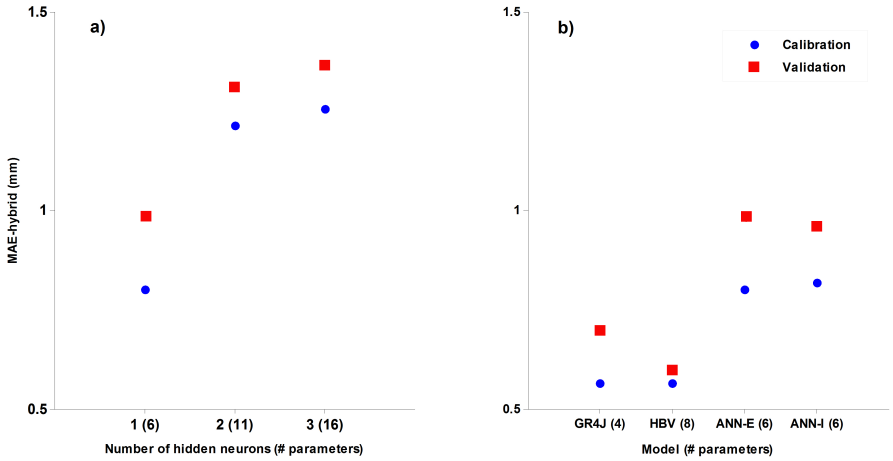


Figure 4.2: Calibration and validation results of **a)** the ANN-E model with one, two and three hidden neurons and **b)** the four models used in this study. The same calibration (1971-2001) and validation (1951-1970) periods are used for both plots.

4.4.2 Effect of ensembles on low flow forecasts for 90 day lead time

The effect of ensemble P and PET on GR4J, HBV and ANN-E is presented as a range bounded by the lowest and highest forecast values in Figure 4.3 and Figure 4.4. In these figures, there is no range for the ANN-I results as the model issues only one forecast using historical low flow indicators as input. The two years, i.e. 2002 and 2003, are carefully selected as they represent a relatively wet year and a very dry year respectively. Figure 4.3 shows that there are significant differences between the four model results. The 90 day ahead low flows in 2002 are mostly overpredicted by the ANN-E model, whereas GR4J, HBV and ANN-I models overpredict and underpredict low flows observed after August. The forecast results of ANN-I are considerably better than the results of the other three models. The overprediction of low flows is more pronounced for GR4J than for the other three models. The overprediction of low flows by ANN-E is mostly at the same level. This less sensitive behaviour of ANN-E to the forecasted ensemble inputs shows the effect of the logarithmic sigmoid transfer function on the results. Due to the nature of this algorithm, input is rescaled to a small interval [0, 1] and the gradient of the sigmoid function at large values approximates zero (Wang *et al.*, 2006). Further, ANN-E is also not sensitive to the initial model conditions updated on every forecast issue day. The less pronounced overprediction of low flows by HBV compared to GR4J may indicate that the slow responding groundwater storage in HBV is less sensitive to different forecasted ensemble P and PET inputs.

Figure 4.3: Range (shown as grey shade) of low flow forecasts in 2002 (the wettest year of the test period) for a lead time of 90 days using ensemble P and PET as input for GR4J, HBV and ANN-E models and using historical P, PET and G as input for the ANN-I model (case 1 - 2002).

The results for 2003 are slightly different than those for 2002. As can be seen from Figure 4.4 the number of low flow days has increased in the dry year, i.e. 2003, and the low flows between August and November are not captured by any of the 40-ensemble forecasts using ANN-E. Moreover, ANN-I performed better in 2002 than in 2003. The most striking result in Figure 4.4 is that the low flows observed in the period between April and June are not captured by any of the three models, i.e. GR4J, HBV and ANN-E. The 90 day low flows between October and November are better forecasted by GR4J and ANN-I than the other two models.

Figure 4.4: Range (shown as grey shade) of low flow forecasts in 2003 (the driest year of the test period) for a lead time of 90 days using ensemble P and PET as input for GR4J, HBV and ANN-E models and using historical P, PET and G as input for the ANN-I model (case 1 - 2003).

For the purpose of determining to which extent ensemble P and PET inputs and different initial conditions affect 90 day low flow forecasts, we run the models with different input combinations such as ensemble P or PET and climate mean P or PET and zero precipitation. Figure 4.5 shows the forecasts using ensemble P and climate mean PET as input for three models. The picture is very similar to Figure 4.4 as most of the observed low flows fall within the constructed forecast range by GR4J and HBV. The forecasts issued by GR4J are better than those issued by the other two models. However, the range of forecasts using GR4J is larger than for the other models showing the sensitivity of the model for different precipitation inputs. It is obvious that most of the range in all forecasts is caused by uncertainties originating from ensemble precipitation input. The results of the fourth model, ANN-I, are the same as in Figure 4.4 and therefore, they are not presented again in the remaining figures.

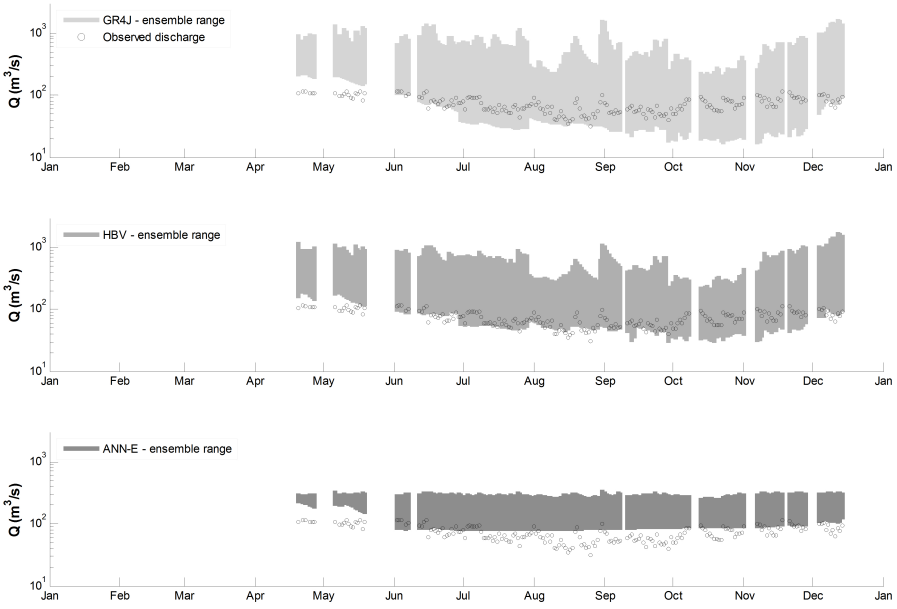


Figure 4.5: Range (shown as grey shade) of low flow forecasts in 2003 for a lead time of 90 days using ensemble P and climate mean PET as input for GR4J, HBV and ANN-E models (case 2).

Figure 4.6 shows the forecasts using climate mean P and ensemble PET as input for three models, i.e. GR4J, HBV and ANN-E. Interestingly, only GR4J could capture the 90 day low flows between July and November using climate mean P and ensemble PET showing the ability of the model to handle the excessive rainfall. None of the low flows were captured by HBV, whereas very few low flow events were captured by ANN-E (Figure 4.6).

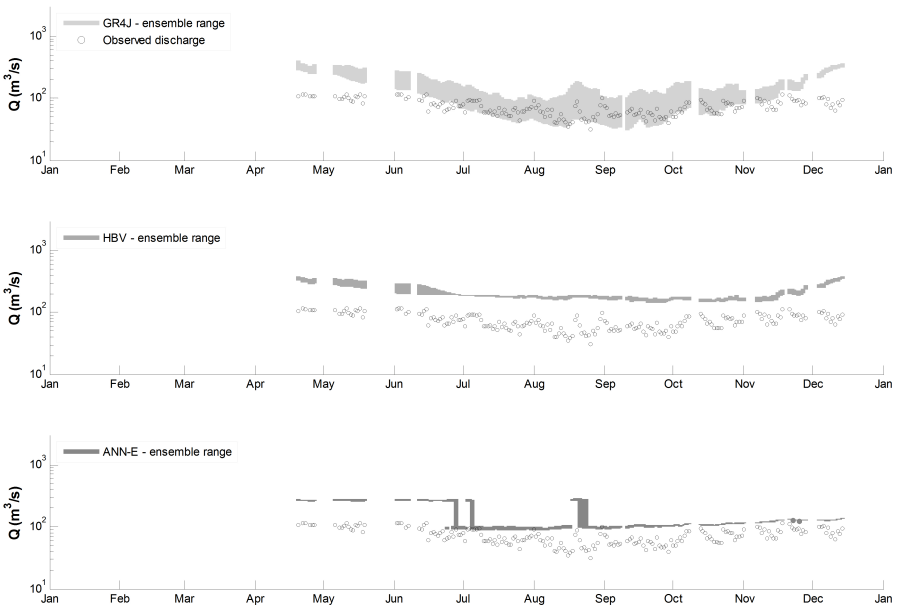


Figure 4.6: Range (shown as grey shade) of low flow forecasts in 2003 for a lead time of 90 days using climate mean P and ensemble PET as input for GR4J, HBV and ANN-E models (case 3).

Figure 4.7 shows the forecasts using climate mean P and PET as input for three models. The results are presented by point values without a range since only one deterministic forecast is issued. There are significant differences in the results of the three models. For instance, all 90 day ahead low flows in 2003 are over-predicted by HBV, whereas the over-prediction of low flows is less pronounced for ANN-E. Moreover, GR4J results are better than those for the other two models as GR4J can forecast some of the low flows accurately. It is remarkable that GR4J can forecast a very dry year accurately using the climate mean. The low values of the calibrated maximum soil moisture capacity and percolation parameters of HBV (FC and $PERC$) can be the main reason for over-prediction of all low flows as the interactions of parameters with climate mean P input can result in higher model outputs. The errors in the forecasts by GR4J and HBV models can also be caused by the uncertainty originating from the model initial conditions.

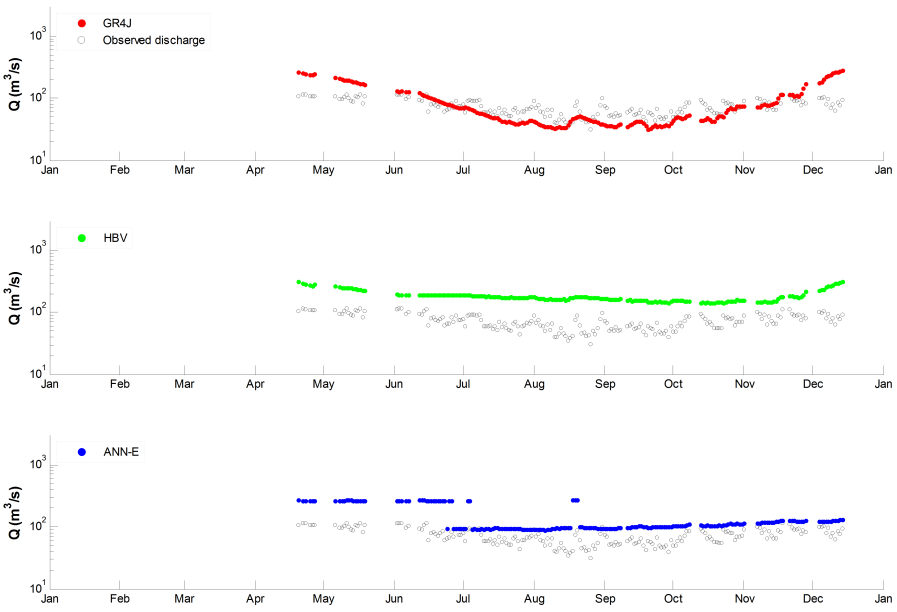


Figure 4.7: Low flow forecasts in 2003 for a lead time of 90 days using both climate mean P and PET as input for GR4J, HBV and ANN-E models (case 4).

Figure 4.8 shows the forecasts using zero P and ensemble PET as inputs for three models. Not surprisingly, both GR4J and HBV underpredicted most of the low flows when they are run without precipitation input. The results from Figure 4.8 confirm that the P input is very crucial for improving low flow forecasts although obviously less precipitation is usually observed in a low flow period compared to other periods. Interestingly, the results of ANN-E are much better than the other two conceptual models showing the ability of partly data-driven models for seasonal low flow forecasts.

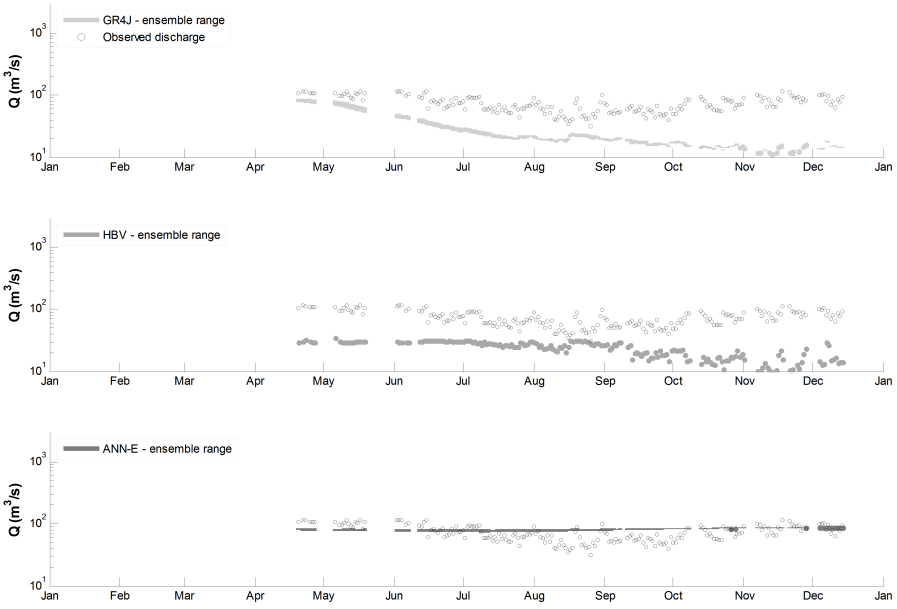


Figure 4.8: Range (shown as grey shade) of low flow forecasts in 2003 for a lead time of 90 days using zero P and ensemble PET as input for GR4J, HBV and ANN-E models (case 5).

4.4.3 Effect of ensembles on low flow forecast skill scores

Figure 4.9 compares the three models and the effect of ensemble P and PET on the skill of probabilistic low flow forecasts with varying lead times. In this figure, four different skill scores are used to present the results of probabilistic low flow forecasts issued by GR4J, HBV and ANN-E. From an operational point of view, the main purpose of investigating the effect of ensembles and model initial conditions on ensemble low flow forecasts with varying lead times is to improve the forecast skills (e.g. hits, reliability, BSS and MFS) and to reduce false alarms and misses. As anticipated, all scores decrease with increasing lead time. From Figure 4.9 we can clearly see that the results of GR4J show the lowest BSS, MFS and hit rate. The false alarm rate of forecasts using GR4J is also the lowest compared to those using other models. This indicates that the non-occurrence of low flow days is better forecasted by GR4J than the other models. It appears from the results that ANN-E and HBV show comparable skill in forecasting low flows up to a lead time of 90 days. It should be noted that the probabilistic skill scores for ANN-I were calculated only for a leadtime of 90 days and are not shown in Figure 4.9. The mean forecast score and hit rate are equal to one, confirming the good deterministic ANN-I forecast results in Figure 4.3 and Figure 4.4. However,

the ANN-I model is less skilful than climatology (i.e. BSS<0) for non-low flow events. Similarly, the false alarm rate of ANN-I is equal to one, showing that the model predicts only low flows and misses all non-low flow events. This is from the fact that ANN-I is solely developed for forecasting low flow days. In other words, only observed low flows and corresponding input data with appropriate lags and temporal resolution were introduced to the ANN-I model during calibration and validation.

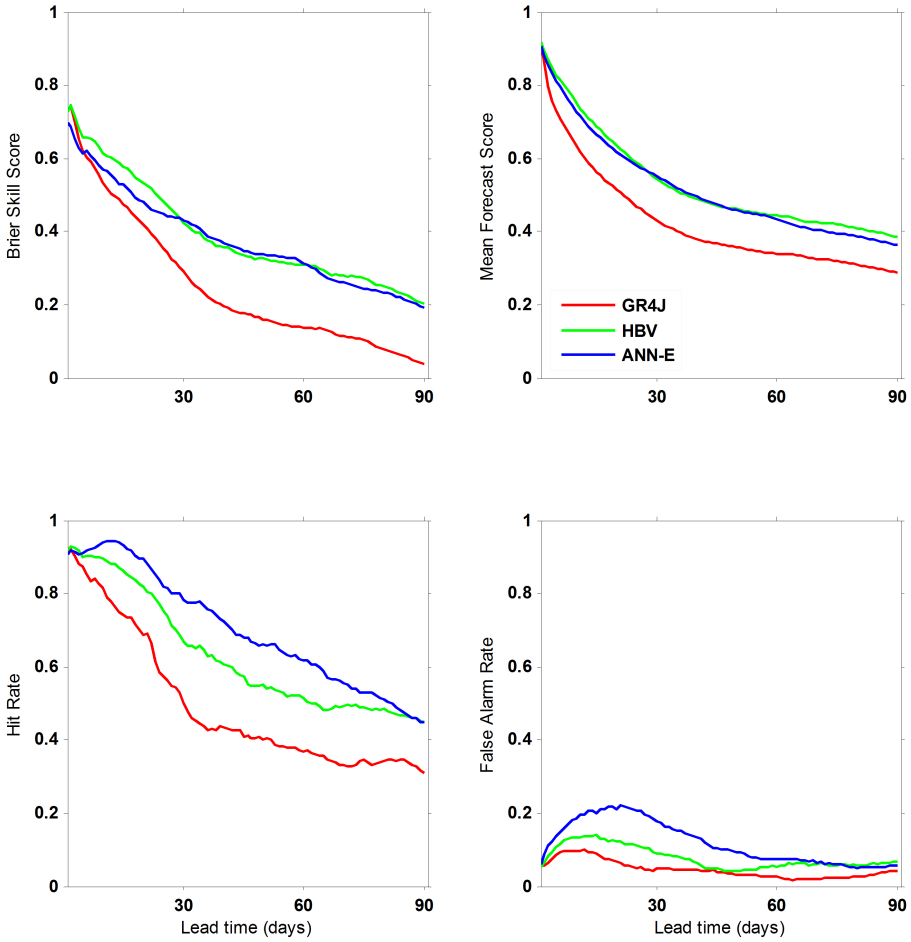


Figure 4.9: Skill scores for forecasting low flows at different lead times for three different hydrological models.

Figure 4.10 compares the reliability of probabilistic 90 day low flows forecasts below different thresholds (i.e. Q75, Q90 and Q95) using ensemble P and PET as input for three models. The figure shows that the Q75 and Q90 low flow forecasts issued by the HBV model are most reliable compared to the other models. Moreover, all three models under-predicts most of the forecast intervals. It appears from Figure 4.10c that very critical low flows (i.e. Q99) are under-predicted by GR4J model.

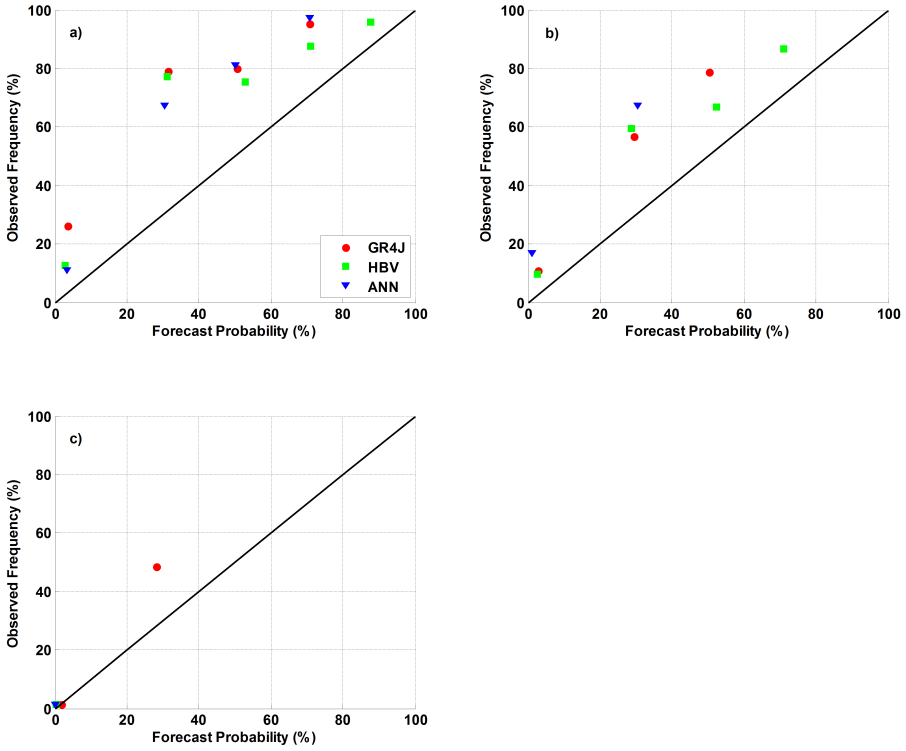


Figure 4.10: Reliability diagram for different low flow forecasts **a)** Low flows below Q75 threshold **b)** Low flows below Q90 threshold **c)** Low flows below Q99 threshold. The forecasts are issued for a lead time of 90 days for the test period 2002-2005 using ensemble P and PET as input for GR4J, HBV and ANN-E models.

4.5 Discussion

To evaluate the effects of seasonal meteorological forecasts on low flow forecasts, 40-member ensembles of ECMWF seasonal meteorological forecasts were used as input for four low flow forecast models. Different input combinations were compared to distinguish between the effects of ensemble P and PET and model initial conditions on 90 day low flow forecasts. The models could reasonably

forecast low flows when the amount of precipitation, i.e. ensemble P was introduced into the models. This result is in line with that of Shukla and Lettenmaier (2011) who found that seasonal meteorological forecasts have a greater influence than initial model conditions on the seasonal hydrological forecast skills. Two other related studies also showed that the effect of a large spread in ensemble seasonal meteorological forecasts is larger than the effect of initial conditions on hydrological forecasts with lead times longer than 1-2 months (Li *et al.*, 2009; Yossef *et al.*, 2013). The encouraging results of low flow forecasts using ensemble seasonal precipitation forecasts for the hydrological models confirm the utility of seasonal meteorological forcing for low flow forecasts. Not surprisingly, all models under-predicted low flows without rainfall information (zero P).

In this study, we also assessed the effects of ensemble P and PET on the skill scores of low flow forecasts with varying lead times up to 90 days. In general, the four skill scores show similar results. The most evident two patterns in these scores are that; first, the forecast skill drops sharply until a lead time of 30 days and second, the skill of probabilistic low flow forecasts issued by GR4J is lowest, whereas the skill of forecasts issued by ANN-E is highest compared to the other three models. Further, our study showed that data-driven models can be good alternatives to conceptual models for issuing seasonal low flow forecasts.

It is noteworthy to mention that the two data-driven models developed in this study, i.e. ANN-E and ANN-I, can be applied to other large river basins elsewhere in the world. Surprisingly, ANN-E and HBV showed similar skill on seasonal forecasts although we expected the two conceptual models, GR4J and HBV, would show similar results up to a lead time of 90 days. The skill score results of ANN-I may seem contradictory, but it shows that ANN-I is useless to predict whether low flow (as defined, below a threshold) will occur or not. Therefore, one of the other three models will be required. Though, if one of the other models predicts that low flow below a threshold will occur, ANN-I can be used to predict the magnitude of low flows, better than the other three models.

For that, the low flow indicators in the basin and their appropriate temporal scales should be firstly identified. Identification of low flow indicators in a small size basin can be cumbersome due to the slow responding storages and the large spatial extent of groundwater aquifers.

4.6 Conclusions

Four hydrological models have been compared regarding their performance in the calibration, validation and forecast periods, and the effect of seasonal meteorological forecasts on the skill of low flow forecasts has been assessed for varying lead times. We run the models with different input combinations, e.g. climate mean precipitation and ensemble potential evapotranspiration, to identify which input source led to the largest range in the forecasts. A new hybrid low flow objective function, comprising the mean absolute error of low flows and the mean absolute error of inverse discharges, is used for comparing low flow simulations, whereas the skill of the probabilistic seasonal low flow forecasts has been evaluated based on the ensemble forecast range, Brier Skill Score, reliability, hit/false alarm rates and Mean Forecast Score. The latter skill score (MFS) focusing on low flows is firstly introduced in this study. In general our results showed that;

- Based on the results of the calibration and validation, one hidden neuron was found to be enough for seasonal forecasts as additional hidden neurons in ANNs did not increase the simulation performance. Interestingly, the data-driven models, i.e. ANN-E and ANN-I performed similarly in the calibration and validation periods showing the utility of identified indicators in simulating low flows by ANN-I. The difference between calibration and validation performances was smallest for the HBV model, i.e. the most sophisticated model used in this study.

- Based on the results of the comparison of different model inputs, the largest range for 90 day low flow forecasts is found for the GR4J model when using ensemble seasonal meteorological forecasts as input. Moreover, the uncertainty arising from ensemble precipitation has a larger effect on seasonal low flow forecasts than the effects of ensemble potential evapotranspiration. All models are prone to over-predict low flows using ensemble seasonal meteorological forecasts. However, the precipitation forecasts in the forecast period are crucial for improving the low flow forecasts. As expected, all three models, i.e. GR4J, HBV and ANN-E underpredicted 90 day ahead low flows in 2003 without rainfall data.

- Based on the results of the comparison of forecast skills with varying lead times, the low flow forecasts using GR4J are less skilful than the other three models. However, the false alarm rate of GR4J is also the lowest indicating the ability of the model of forecasting non-occurrence of low flow days. The low flow forecasts issued by HBV model are most reliable compared to the other models. The ANN-I model can predict the magnitude of the low

flows better than the other three models. However, ANN-I is not successful in distinguishing between low flow events and non-low flow events for a lead time of 90 days. The hit rate of ANN-E is higher than that of the two conceptual models used in this study. Overall, the ANN-E and HBV models are the best performing two of the three models using ensemble P and PET.

4.7 Acknowledgements

We acknowledge the financial support of the Dr. Ir. Cornelis Lely Stichting (CLS), Project No. 20957310. The research is part of the programme of the Department of Water Engineering and Management at the University of Twente and it supports the work of the UNESCO-IHP VII FRIEND-Water programme. Discharge data for the River Rhine were provided by the Global Runoff Data Centre (GRDC) in Koblenz (Germany). Areal precipitation and evapotranspiration data were supplied by the Federal Institute of Hydrology (BfG), Koblenz (Germany). REGNIE grid data were extracted from the archive of the Deutscher Wetterdienst (DWD: German Weather Service), Offenbach (Germany). ECMWF ENS data used in this study have been obtained from the ECMWF seasonal forecasting system, i.e. Mars System 3. We thank Dominique Lucas from ECMWF who kindly guided us through the data retrieval process. The GIS base maps with delineated 134 basins of the Rhine basin were provided by Eric Sprokkereef, the secretary general of the Rhine Commission (CHR). The GR4J and HBV model codes were provided by Tian Ye. We are grateful to the members of the Referat M2 - Mitarbeiter/innen group at BfG, Koblenz, in particular Peter Krahe, Dennis Meißner, Bastian Klein, Robert Pinzinger, Silke Rademacher and Imke Lingemann, for discussions on the value of seasonal low flow forecasts.

Chapter 5

Impacts of climate change on the seasonality of low flows in 134 catchments in the River Rhine basin using an ensemble of bias-corrected regional climate simulations

Abstract

The impacts of climate change on the seasonality of low flows were analysed for 134 sub- catchments covering the River Rhine basin upstream of the Dutch-German border. Three seasonality indices for low flows were estimated, namely the seasonality ratio (*SR*), weighted mean occurrence day (*WMOD*) and weighted persistence (*WP*). These indices are related to the discharge regime, timing and variability in timing of low flow events respectively. The three indices were estimated from: (1) observed low flows; (2) simulated low flows by the HBV model using observed climate as input; (3) simulated low flows using simulated inputs from seven combinations of General Circulation Models (GCMs) and Regional Climate Models (RCMs) for the current climate (1964-2007); (4) simulated low flows using simulated inputs from seven combinations of GCMs and RCMs for the future climate (2063-2098) including three different greenhouse gas emission scenarios. These four cases were compared to assess the effects of the hydrological model, forcing by different climate models and different emission scenarios on the three indices.

Significant differences were found between cases 1 and 2. For instance, the HBV model is prone to overestimate *SR* and to underestimate *WP* and simulates very late *WMOD*s compared to the estimated *WMODs* using observed discharges. Comparing the results of cases 2 and 3, the smallest difference was found for

the *SR* index, whereas large differences were found for the *WMOD* and *WP* indices for the current climate. Finally, comparing the results of cases 3 and 4, we found that *SR* decreases substantially by 2063-2098 in all seven sub-basins of the River Rhine. The lower values of *SR* for the future climate indicate a shift from winter low flows ($SR > 1$) to summer low flows ($SR < 1$) in the two Alpine sub-basins. The *WMOD*s of low flows tend to be earlier than for the current climate in all sub-basins except for the Middle Rhine and Lower Rhine sub-basins. The *WP* values are slightly larger, showing that the predictability of low flow events increases as the variability in timing decreases for the future climate. From comparison of the error sources evaluated in this study, it is obvious that different RCMs and GCMs have a larger influence on the timing of low flows than different emission scenarios. Finally, this study complements recent analyses of an international project (Rhineblick) by analysing the seasonality aspects of low flows and extends the scope further to understand the effects of hydrological model errors and climate change on three important low flow seasonality properties: regime, timing and persistence.

This chapter is based on the paper: Demirel MC, Booij MJ, Hoekstra AY, (2013). Impacts of climate change on the seasonality of low flows in 134 catchments in the River Rhine basin using an ensemble of bias-corrected regional climate simulations, *Hydrology and Earth System Sciences*, **17** (10): 4241-4257, doi:10.5194/hess-17-4241-2013.

5.1 Introduction

The rivers in Western Europe have a seasonal discharge regime with high flows in winter and low flows in late summer. Many cities are located along these rivers like the River Rhine, as the rivers are used for drinking water supply and industrial use. The rivers are also used for irrigation, power production, freight shipment (Demirel *et al.*, 2010; Jonkeren *et al.*, 2013) and fulfil ecological and recreational functions (De Wit *et al.*, 2007). Floods and low flows in these rivers may cause several problems to society. Since floods are eye-catching, quick and violent events risking human-life, water authorities often focus on flood issues. In contrast, hydrological droughts, causing low flows, develop slowly and affect a much larger area than floods (Van Lanen *et al.*, 2013). Low flows in rivers may negatively affect all important river functions. Severe problems, e.g. water scarcity for drinking water supply and power production, hindrance to navigation and deterioration of water quality, have already been seen during low flow events in the River Rhine in dry summers such as in 1976, 1985 and 2003. Consequently, understanding low flows and its seasonal to inter-annual variation has both societal and scientific value as there is a growing concern that the occurrence of low flows will intensify due to climate change (Grabs *et al.*, 1997; Huang *et al.*, 2013; Middelkoop *et al.*, 2001) and reduced summer runoff contribution from Alpine glaciers (Huss, 2011). We are interested in evaluating the effects of climate change on the seasonality of low flows, and in presenting corresponding uncertainty to provide low flow seasonality information under different climate projections.

Assessing the impacts of climate change and associated uncertainties of the climate change projections is an important field in hydroclimatology (Arnell and Gosling, 2013; Bennett *et al.*, 2012; Chen *et al.*, 2011; Jung *et al.*, 2013; Minville *et al.*, 2008; Prudhomme and Davies, 2009; Taylor *et al.*, 2012). The assessment of the effect of climate change impacts on hydrological catchment response is based on predicted meteorological variables like precipitation and temperature by climate models. Currently available climate change projections are mainly based on the outputs of general circulation models (GCMs) and additionally the outputs of regional climate models (RCMs) with a higher spatial resolution than GCMs. However, it is obvious that regional climate change projections based on these climate model outputs are highly uncertain due to unknown future greenhouse gas emissions and the simplified representation of processes in both RCMs and GCMs (Graham *et al.*, 2007). Therefore, design practices will face new challenges which will require a better quantitative understanding of potential changes in seasonality of low flows complicated by several sources of uncertainty linked to climate change.

Many studies have investigated the impacts of climate change on hydrological regimes of different rivers such as the Nile River (Beyene *et al.*, 2010), the Columbia River in Canada (Schnorbus *et al.*, 2012), the Thames in the UK (Diaz-Nieto and Wilby, 2005; Wilby and Harris, 2006) and the River Rhine (Bosshard *et al.*, 2013; Lenderink *et al.*, 2007; Shabalova *et al.*, 2003). Most of the River Rhine studies focus on the snow processes in the Swiss Alps (Bormann, 2010; Horton *et al.*, 2006; Jasper *et al.*, 2004; Schaefli *et al.*, 2007). The River Rhine studies show that the projected temperature increase by GCMs strongly determines the temporal evolution of snowmelt and, accordingly, high flows in the catchments studied. Shabalova *et al.* (2003) showed a decrease of summer low flows and an increase of winter high flows in the River Rhine leading to an increased flood risk in the winter period. Jasper *et al.* (2004) used 17 combinations of GCMs and emission scenarios to assess the impact of climate change on runoff in two Swiss catchments. They found substantial reductions in snowpack and shortened duration of snow cover, resulting in time-shifted and reduced runoff peaks. The recent Rhineblick project (Görgen *et al.*, 2010) focused on climate change impacts on the magnitude of different discharge regimes, high flows in particular.

Several studies documented potential effects of climate change on low flows in the River Rhine (Huang *et al.*, 2013; Te Linde *et al.*, 2010) and on low flows in the Thames River (Diaz-Nieto and Wilby, 2005; Wilby and Harris, 2006). Huang *et al.* (2013) analysed the effects of three climate change projections on the length of the low flow period and on the 50-year return period of deficit volumes for the Rhine sub-catchments in Germany. Their study showed that low flow events are likely to occur more frequently by 2061-2100 in Western Germany (Huang *et al.*, 2013). Wilby and Harris (2006) assessed the effects of emission scenarios, GCMs, statistical downscaling methods, hydrological model structure and hydrological model parameters on simulating changes in low flows. Their study showed that GCMs and the downscaling method were the most important sources of uncertainty. Although GCMs are a very important source of uncertainty (Graham *et al.*, 2007; Prudhomme and Davies, 2009), the effects of uncertainty from RCMs should not be neglected (Horton *et al.*, 2006; Yimer and Andreja, 2013). The uncertainty due to the hydrological model used generally is relatively small compared to the uncertainty from emission scenarios and climate models (Prudhomme and Davies, 2009).

Most of the above mentioned studies focus on the effects of climate change uncertainty on river flow regimes. Earlier work exists for seasonality analysis of observed low flows (Laaha and Blöschl, 2006; Tongal *et al.*, 2013) and floods (Parajka *et al.*, 2009; Parajka *et al.*, 2010) to understand the hydrological processes in the studied catchments. However, only few studies analysed the impacts

of climate change on the seasonality of floods in Switzerland (Köplin *et al.*, 2013) and the seasonality of dam inflows in Korean rivers (Jung *et al.*, 2013). The first study by Köplin *et al.* (2013) assessed the changes in the seasonality of annual mean and annual maximum flows for a 22 year period for 189 catchments in Switzerland using circular statistics and an ensemble of climate scenarios. They assessed both changes in the mean occurrence date of floods as well as changes in the strength of the flood seasonality. The latter study by Jung *et al.* (2013) has investigated monthly dam inflow series and the standard deviation of these monthly series to reflect the seasonality of dam inflows using 39 climate simulations (13 GCMs with three emission scenarios) and three hydrologic models. They explicitly take into account the hydrological model uncertainty (Jung *et al.*, 2013). To our knowledge, so far no study has assessed the impacts of climate change, driven by state of the art climate scenarios, on the seasonality of low flows.

The objective of this study is to assess the effects of climate change on the seasonality of low flows in the River Rhine basin using different climate change projections. The effects of the hydrological model, the forcing by different combinations of GCMs and RCMs, and different emission scenarios on the seasonality of low flows are evaluated. The seasonality of a hydrological variable is often described in terms of mean value during fixed seasons (e.g. June, July, and August, or JJA) (Baldwin and Lall, 1999; Guo *et al.*, 2008). In this study, following the study of Laaha and Blöschl (2006), seasonality of low flows is described through the analysis of three indices namely the Seasonality Ratio (the ratio of summer low flow and winter low flow), the Weighted Mean Occurrence Day and the Weighted Persistence (measuring the variability in timing) of low flows. Daily observed low flow series from 101 sub-catchments and simulated low flow series from 134 sub-catchments are available and used to assess the effects of climate change on the three indices. This study complements the recent analyses of the Rhineblick project (Görgen *et al.*, 2010) by analysing the effects of climate change on three important low flow seasonality properties (regime, timing and persistence of timing) and extending the scope further to understand the effects of hydrological model errors and climate change on these seasonality properties: regime, timing and persistence.

The outline of the paper is as follows. The study area is introduced in section 5.2. The seasonality indices, the hydrological model and the data used in this study are described in section 5.3. The results are presented in section 5.4. The findings are discussed in section 5.5, and the conclusions are drawn in section 5.6.

5.2 Study area

The River Rhine basin is a major and densely populated river basin in Western Europe accommodating nearly 60 million inhabitants. The surface area of the basin is approximately 185,300 km^2 and the river flows along a 1,233 km course from the Alps to the North Sea. The topography of the basin is quite diverse varying from high Alpine mountains to flat lands in the downstream part. In addition to its importance as an inland water, the River Rhine serves as a vital freshwater resource for the Netherlands as well as for the other upstream countries such as Luxemburg, Germany and Switzerland (Middelkoop and Van Haselen, 1999). The average discharge downstream of the Alpine mountains is approximately 1000 m^3/s. It then increases up to 2,300 m^3/s at the Lobith gauging station after the German-Dutch border. The minimum observed discharge at this gauging station was 575 m^3/s in 1929. The contribution of the Alps to the total discharge can be more than 70% in summer, whereas it is only about 30% in winter (Middelkoop and Van Haselen, 1999). In the winter period, the precipitation is stored as snow and ice in the Alps until late spring. Due to the high evapotranspiration and little melt-water input from the Alps, low flows typically occur in late summer or autumn (Nilson *et al.*, 2012).

Figure 5.1 shows the River Rhine basin at two spatial scales, i.e. 134 sub-catchments and seven sub-basins. The hydrology of the River Rhine basin has already been modelled at a spatial scale of 134 sub-catchments (Eberle, 2005; Görgen *et al.*, 2010; Renner *et al.*, 2009; Te Linde *et al.*, 2008), whereas the indicators of low flow events have been assessed at an aggregated spatial scale of seven major sub-basins by Demirel *et al.* (2013a).

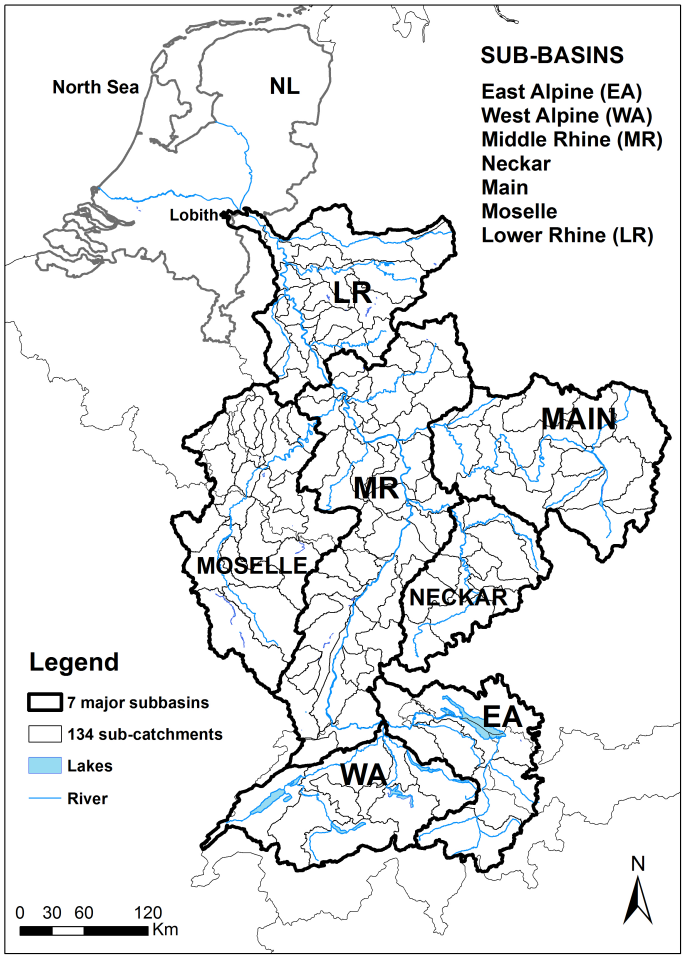


Figure 5.1: Schematisation of the 134 sub-catchments (spatial scale of HBV model) and seven major sub-basins of the River Rhine upstream of Lobith.

The spatial scales of 134 sub-catchments and seven sub-basins are used to present our results. The first spatial scale allows us to compare the differences in the three indices at a very detailed level, whereas the second spatial scale gives insight about the hydrological processes in the major tributaries of the River Rhine. The outlet discharges for the East Alpine (EA) (station #2143 at Rekingen), West Alpine (WA) (station #2016 at Aare-Brugg), Neckar (station #6335600 at Rockenau), Main (station #24088001 at Frankfurt), Moselle (station #6336050 at Cochem), Middle Rhine (MR) (station #6335070 at Andernach) and Lower Rhine (LR) (station #6435060 at Lobith) are used in the seasonality assessment. Although the MR and LR sub-basins have mixed discharge regimes originating from snow- and rainfall-dominated sub-catchments, they are also included in this study.

5.3 Methods and data

In this study, a simulation approach was used to assess the effects of climate change on the seasonality of low flows in the River Rhine. In this approach, observed inputs and simulated inputs from bias-corrected outputs of seven climate scenarios were used as forcing for the hydrological model. Observed low flows (case 1 in Table 5.1) and the outputs of the hydrological model (case 2, 3 and 4) were then used to estimate three seasonality indices as discussed below.

Table 5.1: Overview of the seasonality calculations.

Case number	**Number of calculations**	**Description of calculations**
1	1	The three indices are based on observed discharge series with varying lengths
2	1	The three indices are based on simulated discharge using observed climate for 1964-2007 as input
3	7	The three indices are based on simulated discharge using simulated climate for 1964-2007 as input
4	7	The three indices are based on simulated discharge using simulated climate for 2063-2098 including three emission scenarios as input

Cases 1 and 2 are compared to assess the effects of the hydrological model errors on the three seasonality indices. Secondly, we compare cases 2 and 3 to assess the effects of the meteorological forcing on the three indices. In the third and final comparison, cases 3 and 4 are used to assess the effects of different emission scenarios on the seasonality of low flows. We present the three indices at two spatial scales that are 134 sub-catchments and seven major sub-basins.

5.3.1 Seasonality indices

Laaha and Blöschl (2006) give an overview of seasonality indices and how they can be estimated based on discharge time series. Seasonality indices were estimated to describe different aspects of the discharge regime of a river. We used three seasonality indices described below as they focus on the differences in discharge regime, timing and variability in timing of the recurrent event (persistence).

Seasonality ratio

The Seasonality Ratio (SR) index reveals the low flow characteristics in summer and winter periods (Laaha and Blöschl, 2006). The definitions of a low flow threshold and the seasons are crucial for the SR results as the underlying hydrological processes for summer and winter low flows are different (Laaha and Blöschl, 2006; Tongal *et al.*, 2013). Following De Wit *et al.* (2007), we selected the period from November to April as winter half-year and the period from May to October as summer half-year season. The low flow series were then divided into winter and summer low flow series. We used the 75% exceedence probability (Q_{75}), as in Demirel *et al.* (2013a), as a threshold for defining summer low flow (Q_{75}) and winter low flow (Q_{75w}). The SR index is calculated as the ratio of Q_{75s} and Q_{75w} (Eq. 5.1) (Laaha and Blöschl, 2006).

$$SR = \frac{Q_{75s}}{Q_{75w}} \tag{5.1}$$

A value of SR greater than one indicates the presence of a winter low flow regime and a value smaller than one indicates the presence of a summer low flow regime.

Weighted mean occurrence day

The Weighted Mean Occurrence Day *(WMOD)* is an index similar to the seasonality index of Laaha and Blöschl (2006). For each sub-catchment, the days on which the discharge is below the Q_{75} threshold are transformed into Julian dates D_i, i.e. the day of the year ranging from 1 to 365 in regular years and 1 to 366 in leap years. The day number of each low flow event (D_i) is weighted by the inverse low flow value ($1/Q_i$) on the same day to address the severity of a low flow event as well as its occurrence day. The weighted mean occurrence day is estimated first in radians to represent the annual cycle correctly. Otherwise, a simple averaging of low flow occurrences in winter months, e.g. January and December, can lead to a large error in the results. The weighted mean of Cartesian coordinates x_θ and y_θ of a total number of low flow days i is defined as

$$x_\theta = \frac{\sum_i \frac{\cos(\frac{D_i \times 2\pi}{365})}{Q_i}}{\sum_i Q_i^{-1}} \tag{5.2}$$

$$y_\theta = \frac{\sum_i \frac{\sin(\frac{D_i \times 2\pi}{365})}{Q_i}}{\sum_i Q_i^{-1}} \tag{5.3}$$

The directional angle (θ) is then estimated by

$$\theta = arctan\left(\frac{y_\theta}{x_\theta}\right) \qquad 1^{st}\ and\ 4^{th}\ quadrants:\ x_\theta > 0 \tag{5.4}$$

$$\theta = arctan\left(\frac{y_\theta}{x_\theta}\right) + \pi \qquad 2^{nd}\ and\ 3^{rd}\ quadrants:\ x_\theta < 0 \tag{5.5}$$

The values of θ can vary from 0 to 2, where a zero value indicates 1^{st} of January, $\pi/2$ represents 1^{st} of April, π represents 1^{st} of July and $3\pi/2$ represents 1^{st} of October. The main advantage of using circular statistics is that it allows us to correctly average low flow occurrences in the winter half-year period. The *WMOD* is then obtained by back-transforming the weighted mean angle to a Julian date:

$$WMOD = \theta\,\frac{365}{2\pi} \tag{5.6}$$

Weighted persistence

The weighted persistence (*WP*) is calculated using the weighted mean of Cartesian coordinates x_θ and y_θ in Eq. (5.7).

$$WP = \sqrt{x_\theta^2 + y_\theta^2} \tag{5.7}$$

The dimensionless *WP* indicates the variability in timing of low flows, where a value of 1 indicates that low flow events occurred on exactly the same day of the year (high persistence) and a value of zero indicates that low flow events are uniformly distributed over the year (no persistence) (Laaha and Blöschl, 2006).

5.3.2 Hydrological model

The HBV-96 model (Hydrologiska Byråns Vattenbalansavdelning) is a semi distributed conceptual hydrological model which was developed by the Swedish Meteorological and Hydrological Institute (SMHI) in the early 1970s (Bergström, 1976; Lindström *et al.*, 1997). It consists of five subroutines for snow accumulation and melt, soil moisture accounting, fast runoff, groundwater response and river routing. It operates at a daily time step using precipitation (P) and potential evapotranspiration (PET) as inputs. The HBV model has been used in the field of operational forecasting and climate impact modelling in more than 50 countries around the world (Şorman *et al.*, 2009), in northwestern Europe in particular (Booij, 2005; Driessen *et al.*, 2010; Engeland *et al.*, 2010; Görgen *et al.*, 2010; Te Linde *et al.*, 2008; Wöhling *et al.*, 2006). Its good performance with a low number of parameters is the main advantage of the HBV model for large basins (Te Linde *et al.*, 2008). The HBV model has been applied to the River Rhine since 1997 by the Dutch Water authorities, i.e. Rijkwaterstaat Waterdienst (previously RIZA) and Deltares, and the German Federal Institute of Hydrology (BfG) in Koblenz. We use the HBV-96 model running at a daily time step and covering the area upstream of the Lobith gauging station comprising 134 sub-catchments. The HBV model was first calibrated by Eberle (2005) on the basis of expert knowledge at the BfG in Koblenz. The HBV model upstream of Maxau has been recalibrated again by Berglöv *et al.* (2009) at SMHI using a hybrid objective function (NS_{HBV} in Eq. 5.8) to improve low flow simulations. The calibration was carried out locally for 95 sub-catchments, and validated both locally and for the total river flow. Further, the calibration was mainly done using an automatic routine (Lindström *et al.*, 1997) for the period 01/11/2000 - 01/11/2007 and the period 01/11/1996 - 01/11/2000 was used for validation.

$$NS_{HBV} = 0.5 \cdot R^2 + 0.5 \cdot R^2_{\log} + 0.1 \cdot relaccdiff \tag{5.8}$$

Where R^2 is the efficiency criterion based on Nash and Sutcliffe (Nash and Sutcliffe, 1970), $R^2_{\log}$ is similar to R^2 but using the logarithmic discharge values giving more weight to low flows, and *relaccdiff* is the accumulated relative difference between the simulated and observed discharge (see Eq. (5.9), Berglöv *et al.* (2009)).

$$relaccdiff = \frac{\sum_i (Q_{sim,i} - Q_{obs,i})}{\sum_i Q_{obs,i}} \tag{5.9}$$

The HBV model has served as a robust platform for climate impact studies in the River Rhine basin (Görgen *et al.*, 2010; Nilson *et al.*, 2012; Te Linde *et al.*, 2010). The model simulations for the current and future climate were started on the 1^{st} of January 1961 and 2060 respectively. The first three years were used as a " warm-up" period and model simulation results for these periods were not used in the estimation of the seasonality indices.

5.3.3 Observed data

Daily observed discharge (Q_{obs}) data at the outlets of 101 of the 134 sub-catchments were provided by the Global Runoff Data Centre (GRDC) in Koblenz (Germany) and the Bundesamt für Umwelt (BAFU) in Bern (Switzerland). A complete set of daily P, T and PET data were obtained from Deltares (the Netherlands) and the German Federal Institute of Hydrology (BfG) in Koblenz. PET has been estimated with the Penman-Wendling equation (ATV-DVWK, 2002). All three climate variables were spatially averaged over each of the 134 sub-catchments.
The mean altitude of these sub-catchments has been provided by the International Commission for the Hydrology of the Rhine basin (CHR). The daily P, T and PET data series span from 1961 to 2007, whereas the length of the Q_{obs} data series varies from station to station.

5.3.4 Bias-corrected climate model outputs and transformation to catchment average

All seven regional climate model (RCM) outputs (Jacob, 2006) that were used in this study were provided by the Royal Netherlands Meteorological Institute (KNMI) and BfG in Koblenz. The grid-based RCM outputs have firstly been transferred into daily catchment averages over 134 sub-catchments of the River Rhine basin and then corrected for biases by Görgen *et al.* (2010) for the Rhineblick2050 project. The daily time series of areally-averaged PET estimated following the approach of Penman-Wendling (ATV-DVWK, 2002). This is consistent with the observed PET estimation carried out by the Federal Institute of Hydrology in Koblenz, Germany. The main characteristics of the pre-processed climate dataset, comprising an ensemble of bias-corrected outputs of scenarios based on four regional climate models (RCMs), four driving global climate models (GCMs) and three different emission scenarios (SRES), are shown in Table 5.2.
The three scenarios, i.e. A2, A1B and B1, are based on three different greenhouse gas emission scenarios as defined by the Intergovernmental Panel on Climate Change (IPCC) in the Special Report on Emissions Scenarios (Hurkmans *et*

Table 5.2: Climate data availability and seven climate scenarios (CSs).

ID	SRES	GCM	RCM	Bias correction	Common Period
CS 1	A1B	ECHAM5r3	RACMO	Eq. (5.10) and (5.11) (Görgen *et al.*, 2010)	1961-2007 (Current)
CS 2	A1B	ECHAM5r3	REMO		
CS 3	A1B	HADCM3Q16	HADRM3Q16		2060-2098 (Future)
CS 4	A1B	HADCM3Q3	HADRM3Q3		
CS 5	A1B	ECHAM5r1	REMO		
CS 6	A2	ECHAM5r1	REMO		
CS 7	B1	ECHAM5r1	REMO		

al., 2010; Nakićenović and Swart, 2000). The A2 scenario assumes a world with a continuously increasing population and very regionally oriented economic growth, whereas A1B indicates a globalized, very rapidly growing economy with fast introduction of new technologies that are balanced between fossil fuel intensive and sustainable and clean ones. The global population in the A1B scenario increases rapidly until the middle of 21^{st} the century and decreases thereafter. The third scenario, B1, assumes a globalized, rapidly growing population with changes in economic structure with an environmental emphasis and fast introduction of clean and efficient technologies.

Transferring the indicators of climate change from climate models to hydrological models is not a straightforward process due to the systematic errors in simulated meteorological variables, i.e. precipitation and temperature. For example, many RCMs exhibit a bias in the order of 25% for the amount of summer precipitation in the Alpine region (Graham *et al.*, 2007). Hydrological simulations using uncorrected inputs would be pointless for assessing impacts of climate change on low flow seasonality as summer precipitation amounts are crucial for low flows (Demirel *et al.*, 2013a). The biases from the RCM outputs for precipitation have been corrected by Görgen *et al.* (2010) using the following equation:

$$P_{cor} = a\, P_{RCM}^{b} \tag{5.10}$$

where P_{cor} (mm) is the bias-corrected precipitation, P_{RCM} (mm) is the precipitation from RCMs and, a and b are transformation coefficients which are determined separately for each of the 134 sub-catchments and for each of the 12 calendar months. The frequency distribution of the wet-day precipitation, i.e. location and shape, is not affected by this nonlinear bias-correction method (Eq. 5.10), whereas the frequency of wet days is corrected as in most RCMs the frequency of wet days is overestimated (Görgen *et al.*, 2010). The biases from the RCM outputs for temperature have been corrected by Görgen *et al.* (2010) using the following equation:

$$T_{cor} = \frac{\sigma_o}{\sigma_m}(T_{RCM} - \bar{T}_m) + \bar{T}_o \tag{5.11}$$

where T_{cor} (°C) is the bias-corrected temperature, σ_o (°C) is the standard deviation of the observed daily temperature, σ_m (°C) is the standard deviation of the daily RCM temperature, T_{RCM} (°C) is the RCM temperature, $\bar{T}_m$ (°C) is the long term mean of the RCM temperature and, $\bar{T}_o$ (°C) is the long term mean of the observed temperature series for each of the 134 sub-catchments. By using Eq. (5.11) the mean and standard deviation of the bias-corrected RCM temperature data are forced to be equal to those of the observed current climate data. The bias-corrections are described in detail in Görgen *et al.* (2010).

5.4 Results

5.4.1 Sensitivity of low flow seasonality to hydrological model

Figure 5.2 shows the three seasonality indices based on observed and simulated low flows for the common 101 catchments. These catchments are grouped into the seven major sub-basins as consistent with the previous low flow studies in the River Rhine (Demirel *et al.*, 2013a).
The results in Figure 5.2 reveal that there are significant differences between observed and simulated seasonality indices. The differences in the rain-dominated catchments are smaller than in the snow-dominated catchments. The differences in snow-dominated catchments can be partly explained by the effect of dam operations in the Alpine catchments. Obviously the dam effect is recorded in the observed discharge data, but dams are not incorporated in the hydrological model. Although HBV simulates overall low flows with an error of less than 5% in the simulation of the mean of minimum annual discharges (Eberle, 2005), dam operations can still affect the seasonality characteristics of the low flows (*e.g. WP*).
The results in Figure 5.2 are presented as a function of the mean catchment altitude. This altitude sorting (high to low altitude from left to right) is done within the seven major sub-basins since the mean catchment altitude is an important catchment characteristic for the discharge regime in the Rhine basin. A significant correlation ($r = \sim 0.7$, $p < 0.05$) between *SR* and catchment altitude is found in the 101 sub-catchments as catchments with a higher altitude tend to have winter low flows and higher *SR* values. Contrary to expectations, no significant correlations are found between *SR* and catchment altitude in the Main and Moselle sub-basins. Further, no significant relation is found between catchment altitude and the two other indices, *WMOD* and *WP*.

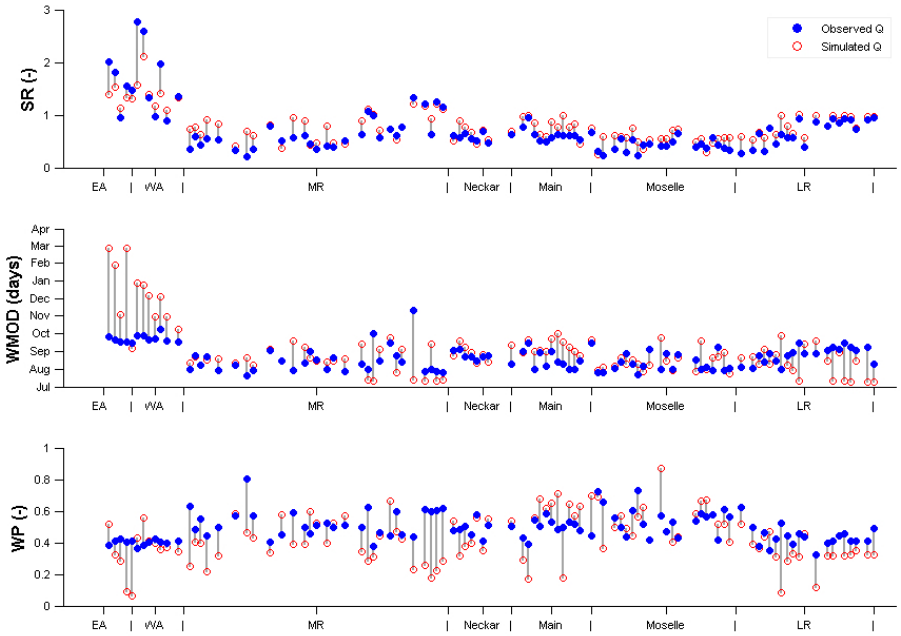


Figure 5.2: Three seasonality indices estimated from observed (case 1) and simulated (case 2) low flows in 101 catchments for the period 1964-2007. The grey line is used to connect observed and simulated indices for each catchment.

The weighted mean occurrence days (*WMOD*s) of simulated low flow events are too late for the EA and WA sub-basins. The *WMOD*s for observed low flows in these Alpine sub-basins are mostly around October, whereas the *WMOD*s for the simulated low flows considerably vary from October to March showing the uncertainty originating from the HBV model and its inputs (Figure 5.2). It should be noted that the effect of the varying lengths of observed discharge time series on the estimation of the *WMOD*s can be substantial for different catchments. This finding for the low flow simulation performance is consistent with that of Te Linde *et al.* (2008), who found variable performance of HBV on the low flow timing and significant errors in the duration of low flows. The weighted persistence (*WP*) of low flow events in the WA sub-basin is better simulated than in other sub-basins.

Figure 5.3 shows the three seasonality indices based on simulated low flows for the 134 catchments. From the *SR* and *WMOD* plots in Figure 5.3, it is apparent that the Alpine catchments have winter low flows, whereas other catchments have summer low flows. The *WMOD*s for the simulated winter low flows are mostly in January and February, whereas those for the simulated summer low flows are in September and October. Moreover, the *WP* in the rain-dominated catchments is generally higher than in the Alpine catchments. The dam operations in the Alpine

catchments in winter periods can marginally affect the *WP* as the dam operations are usually carried out in high flow periods for flood prevention (Bosshard *et al.*, 2013; Middelkoop and Van Haselen, 1999).

Table 5.3 compares the differences between the three seasonality indices based on observed and simulated low flows at the outlets of the seven sub-basins. It should be noted that the relative differences for *SR* and *WP* are presented as a percentage, whereas the difference for *WMOD* is equal to the difference in days at the outlet of the seven sub-basins.

Table 5.3: Differences between the three seasonality indices estimated from observed (case 1) and simulated (case 2) low flows at the outlets of the seven sub-basins in the River Rhine for the period 1964-2007.

	East Alpine	**West Alpine**	**Middle Rhine**	**Neckar**	**Main**	**Moselle**	**Lower Rhine**
SR (%)*	-11	-2	1	11	9	29	2
WMOD (days)**	-10	23	-83	33	5	54	-30
WP (%)*	-85	-17	-16	6	56	52	-34

** (Simulated index - Observed index) / Observed index*

*** Simulated WMOD - Observed WMOD*

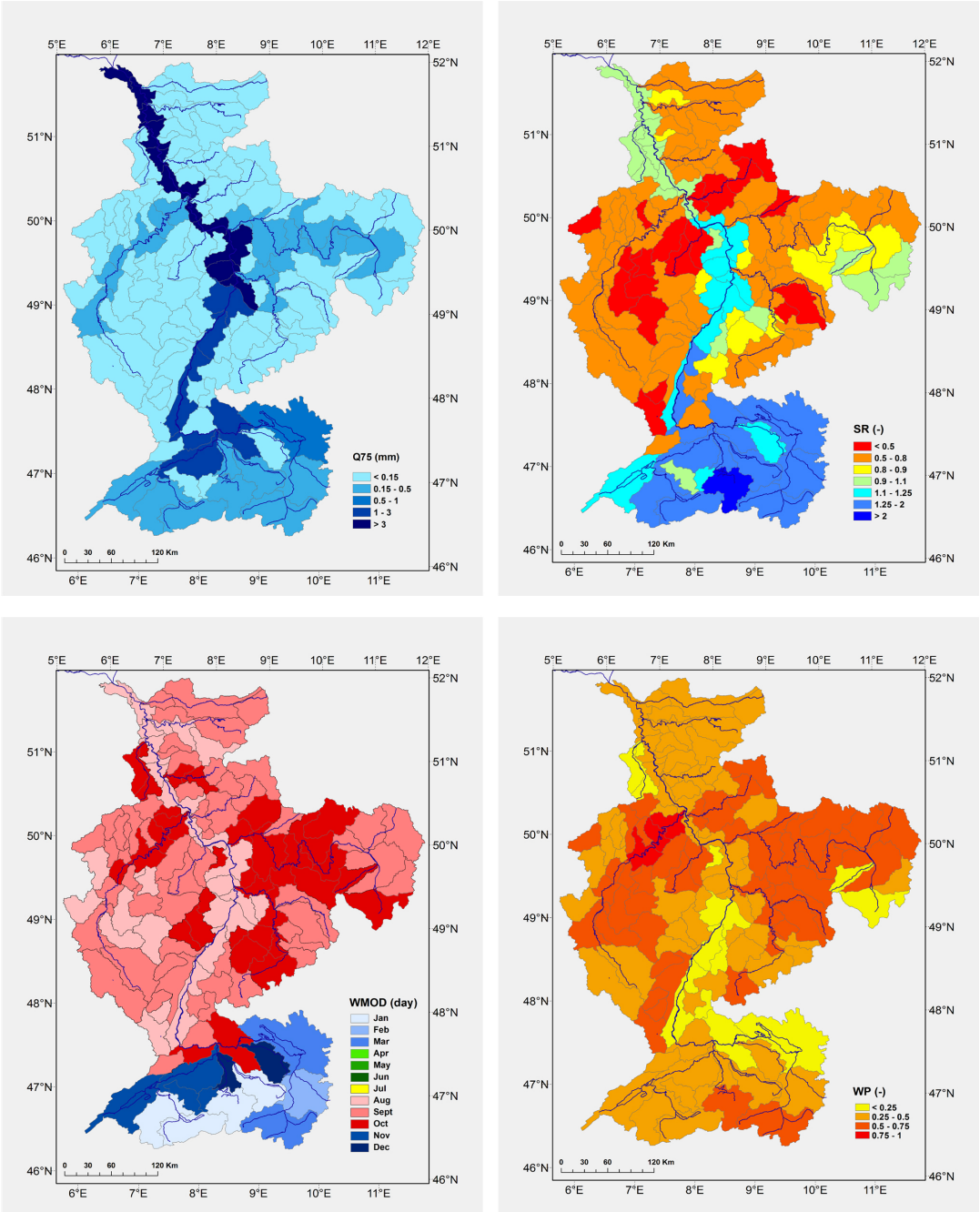


Figure 5.3: Low flow threshold (Q75 in mm/day and three seasonality indices (*SR*, *WMOD* and *WP*) estimated from simulated low flows using observed climate as model input in 134 sub-catchments for the period 1964-2007 (case 2).

No significant differences in *SR* were found between simulated and observed low flows in the WA, MR, Main and LR sub-basins, whereas the largest difference in *SR* was found in the Moselle sub-basin. The negative differences in *SR* were found only in the EA and WA sub-basins showing that the *SR* estimated from simulated low flows (case 2) is smaller than the *SR* estimated from observed low flows (case 1) at the outlet of the two Alpine sub-basins. It is obvious that the MR and LR sub-basins have mixed discharge regimes and, therefore, they are affected by the differences in the upstream sub-basins. For instance, the *WMOD* in the EA sub-basin, which is 10 days earlier than the *WMOD* estimated from observed low flows (case 1), resulted in 83 days earlier *WMOD* in the MR sub-basin. The effect is reduced to a 30 days earlier *WMOD* in the LR sub-basin after the inclusion of other tributaries with late *WMOD*s. The large differences in the *WP*s in all sub-basins except for the Neckar sub-basin show that the simulation of the distribution of low flow events in a year is a difficult task in hydrological modelling.

5.4.2 Sensitivity of low flow seasonality to meteorological forcing

The sensitivity of the three indices to different meteorological forcings is assessed at two spatial scales, i.e. 134 sub-catchments and seven major sub-basins. This is done for the current climate (1964-2007) using observed and simulated inputs for HBV. From the results in Table 5.4, we can see that the outputs of climate scenarios 3 and 4 result in smaller *SR*s than those simulated using observed climate as input for all sub-basins except the WA sub-basin for the current climate. The largest difference in *SR* is found for the Moselle sub-basin. The differences (mostly negative) for climate scenarios 3 and 4, both having boundary conditions from the HADCM3 GCM, are larger than the other five climate scenarios (except for the EA and WA sub-basins).

Table 5.4: Differences between the three seasonality indices estimated from simulated low flows using observed inputs for the reference period 1964-2007 (case 2) compared to the simulated low flows using simulated inputs from seven climate scenarios (CSs) for the same period (case 3)

Index	Climate Scenario	East Alpine	West Alpine	Middle Rhine	Neckar	Main	Moselle	Lower Rhine
SR (%)*	CS 1 (A1B)	6	13	6	5	-9	-5	7
	CS 2 (A1B)	9	19	12	23	6	8	12
	CS 3 (A1B)	-5	0	-13	-25	-19	*-33*	-12
	CS 4 (A1B)	-9	1	-15	-29	-13	*-31*	-13
	CS 5 (A1B)	8	20	6	18	14	-1	8
	CS 6 (A2)	10	23	10	21	16	-1	11
	CS 7 (B1)	6	19	4	13	11	-3	6
WMOD (days)**	CS 1 (A1B)	*45*	12	*90*	11	-24	*-67*	*75*
	CS 2 (A1B)	-11	14	*64*	-1	-1	-16	*56*
	CS 3 (A1B)	*72*	9	*56*	21	11	-16	*55*
	CS 4 (A1B)	*67*	-5	27	-25	-29	*-53*	14
	CS 5 (A1B)	-1	18	*81*	7	-17	-30	*72*
	CS 6 (A2)	*45*	*33*	*102*	1	19	25	*94*
	CS 7 (B1)	26	24	*87*	-9	0	*102*	*78*
WP (%)*	CS 1 (A1B)	*302*	4	23	*33*	*-62*	*-53*	-24
	CS 2 (A1B)	*57*	*-34*	-3	13	*-80*	*-72*	*-40*
	CS 3 (A1B)	*475*	*49*	*126*	*42*	12	-4	*106*
	CS 4 (A1B)	*390*	14	*37*	8	-20	*-42*	*64*
	CS 5 (A1B)	*232*	*-33*	14	10	*-63*	*-55*	7
	CS 6 (A2)	*325*	-4	23	-4	*-58*	*-84*	13
	CS 7 (B1)	*259*	-5	41	20	*-59*	*-75*	*32*

** (Based on simulated input - Based on observed input) / Based on observed input*

*** Based on simulated input - Based on observed input*

The differences in the *WMOD*s of low flows in the WA, Neckar and Main sub-basins are mostly less than 30 days, showing that the weighted mean occurrence day of low flows in these sub-basins is simulated well using the outputs of seven climate scenarios for the current climate. The picture is very different for the other sub-basins. For instance, the *WMOD*s based on simulated current climate as input in the HBV model in the EA, MR, Moselle and LR sub-basins are very different from the *WMOD*s simulated using observed climate. The differences vary from 1 day (by climate scenario 5) in the EA sub-basin to 102 days (by climate scenarios 6 and 7) in the MR and LR sub-basins respectively. Very large differences in the *WP*s in all seven sub-basins, in the EA sub-basin in particular, are simulated using the outputs of climate scenarios. All these differences are positive for the EA sub-basin, showing a substantially smaller variability in timing of low flow events (*WP*s), whereas all the differences are negative for the Moselle sub-basin, showing a larger variability in *WP*s. Since large differences are found in the *WP* index, we also present the detailed effects of seven climate scenarios on the weighted persistence in the 134 sub-catchments in Figure 5.4.

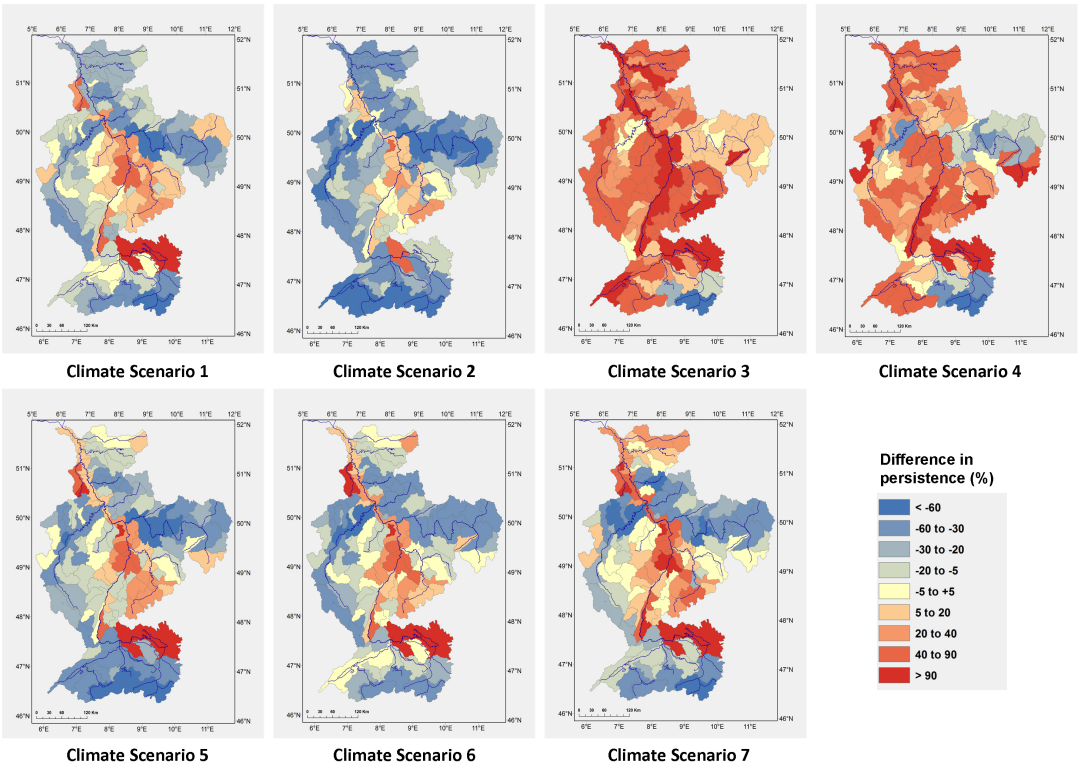


Figure 5.4: Relative differences (%)* between low flow persistence estimated from simulated low flows using simulated inputs from seven climate scenarios for the reference period 1964-2007 (case 3) and simulated low flows using observed inputs for the same period (case 2).

* *(Based on simulated input - Based on observed input) / Based on observed input*

There are large differences in the *WP*s using the outputs of climate scenarios. Climate scenarios 3 and 4 result in a higher *WP* than those simulated using observed climate as input. However, climate scenario 2 results in a lower *WP* than that simulated using observed climate as input. It should be noted that the *WP*s from climate scenarios 5, 6 and 7 are similar as the same version of ECHAM5 and REMO climate models with different emission scenarios are used in these climate scenarios. The significant differences in the climate scenarios can be partly explained by the inter-annual variability of monthly P and PET simulated by the climate scenarios over a year. We found large differences between cases 2 and 3 in the inter-annual variability of monthly P in winter months for all sub-basins, whereas large differences in the inter-annual variability of monthly PET in winter months were found only in rain-dominated sub-basins like in the Moselle sub-basin.

5.4.3 Sensitivity of low flow seasonality to changed climate

Figure 5.5 shows the differences in the three indices between the current and future climate. Here, the effects of the three emission scenarios (A1B, A2 and B1) on the sensitivity of the three indices are also evaluated.

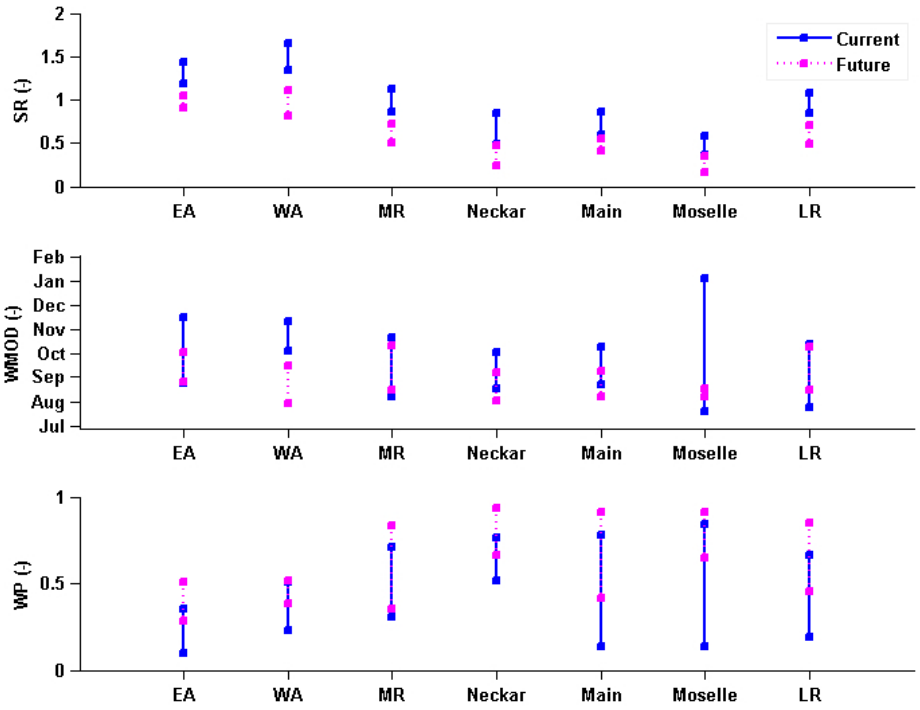


Figure 5.5: Range (shown as bar) of three seasonality indices in the seven sub-basins for the current climate (calculations for case 3) and future climate (calculations for case 4).

From the results in Figure 5.5, it is apparent that the range of *SRs* in all seven sub-basins for the future climate is not overlapping with those for the current climate. The uncertainty in *SRs* is considerably smaller than the uncertainty in the other two indices. Further, the *SR*s are always lower than for the current climate. The lower values of *SR* for the EA and WA sub-basins, for the latter in particular, indicate a substantial shift from winter low flows (*SR* >1) to summer low flows (*SR* <1) which is in line with other climate impact studies (Blenkinsop and Fowler, 2007; Bormann, 2010; Bosshard *et al.*, 2013; Huang *et al.*, 2013; Hurkmans *et al.*, 2010).

Comparing the results for the *WMOD*s, it appears that only the range of *WMOD*s in the WA sub-basin for the future climate is not overlapping with that for the current climate. The largest range of *WMOD*s for the current climate is found

in the Moselle sub-basin. Interesting is that low flows in most of the sub-basins tend to occur earlier by 2063-2098 based on the *WMOD* results in Figure 5.5. The uncertainty in the *WMOD*s varies from several weeks to five months in the sub-basins.

Large ranges are found for *WP* for all sub-basins except for the WA sub-basin using the inputs from seven climate scenarios, indicating that the *WP* index is highly uncertain. The distribution of precipitation over a year can affect the *WP* results significantly as the distribution of precipitation determines the variability in simulated discharges. A significant decrease in the variability in timing of low flows (*WP*s) in the EA sub-basin is found for the future climate. The existence of large lakes in the WA sub-basin can be a reason for a less sensitive *WP*. The most striking result from the *WP* plot in Figure 5.5 is that the weighted persistence is increased in all sub-basins for the future climate suggesting less variability in the timing of low flows. This finding is in line with the scientific consensus that climate change will likely increase the persistence of both high and low flows due to decreasing snowfall and earlier snowmelt, resulting in an earlier occurrence of snowmelt-induced peaks and drier summers (Horton*et al.*, 2006; Jung *et al.*, 2013). This means that the magnitude of extreme high and low flows will be amplified, whereas the timing of these extreme events is more predictable by 2063-2098.

Figure 5.6 shows the changes in the three indices for each climate scenario in the seven sub-basins. Substantial changes in the *SR* index are found, being more pronounced in the rain-dominated sub-basins than in the two Alpine sub-basins. Moreover, the *SR*s estimated from inputs by climate scenario 4 show the smallest change in all sub-basins except for the Main sub-basin, whereas climate scenario 5 shows the largest change in *SR*. Interestingly, the *SR*s estimated from the inputs by climate scenarios 2 and 5 are slightly different in all sub-basins although these two climate scenarios both use ECHAM5 (versions 1 and 3) as GCM and REMO as RCM. The difference in *SR* between these two climate scenarios with the same GCM, RCM and emission scenario can be explained by the different initial conditions used in their driving GCM (Görgen *et al.*, 2010).

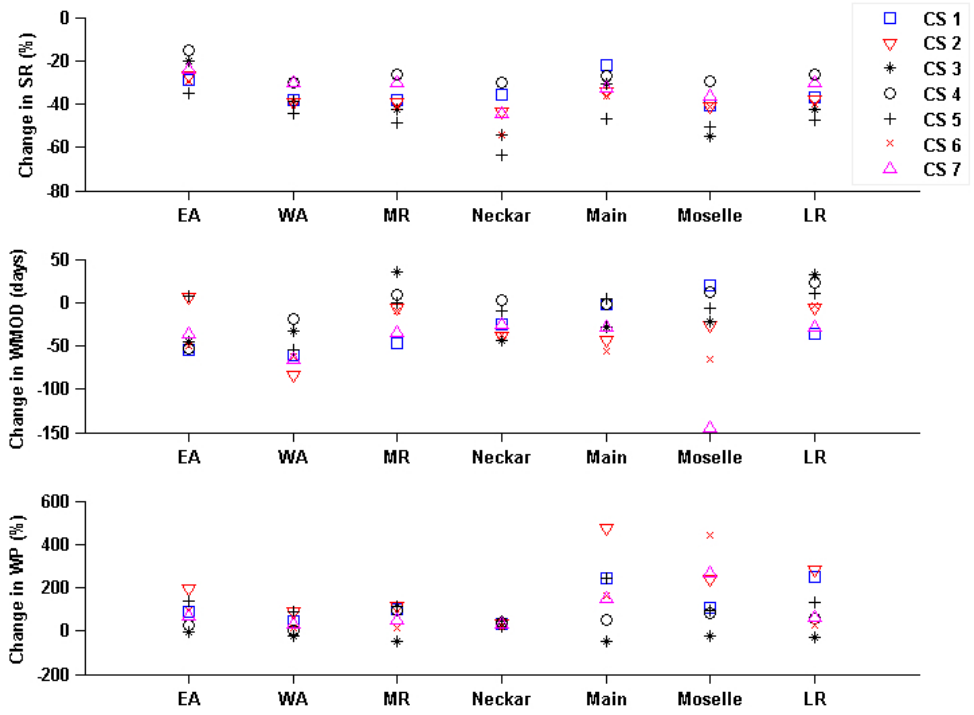


Figure 5.6: The relative changes (*) in *SR* and *WP* and the changes in *WMOD* (**) at the outlet of the seven sub-basins estimated from simulated low flows using simulated inputs for the future period 2063-2098 (case 4) compared to simulated low flows using simulated inputs for the reference period 1964-2007 (case 3) from seven climate scenarios (CSs).

* *(Based on simulated input for future climate - Based on simulated input for current climate) / Based on simulated input for current climate*

** *Based on simulated input for future climate - Based on simulated input for current climate*

From the results in Figure 5.6, it is apparent that climate change result in a negative change in *WMOD*s for the EA and WA sub-basins. Climate scenario 7 shows a very large change in *WMOD* for the Moselle sub-basin. The influence of climate scenario 2 on the change in the *WP* in the Main sub-basin and the influence of climate scenario 6 on the change in the *WP* in the Moselle sub-basin are both about 400%, suggesting much less variability in the timing of low flows in these sub-basins. Since large changes are found in the *WP* index for the future climate, we present Figure 5.7 to compare the effects of seven equally probable climate scenarios on the weighted persistence in the 134 sub-catchments. It is obvious from Figure 5.7 that the outputs of climate scenario 2 show the largest change in *WPs* in the 134 sub-catchments for the future climate, whereas climate scenario 3 shows the smallest change in the *WP*s.

It should be noted that the *WP*s from climate scenarios 5, 6 and 7 are significantly

different as different emission scenarios are used in these scenarios. The large changes in these climate scenarios for the future climate can be partly explained by the inter-annual variability of monthly P and PET simulated by the climate scenarios. We found large changes in the inter-annual variability of monthly P in all months in the Alpine sub-basins, whereas large changes are found mostly in summer months in the rain-dominated sub-basins. Further, large changes in the inter-annual variability of monthly PET were found in winter months in all sub-basins. Some of the Alpine catchments show significant increases in the low flow persistence which is consistent with the results of Huang *et al.* (2013) who reported less variability in the occurrence of low flows for the Alpine regions for all climate scenarios investigated.

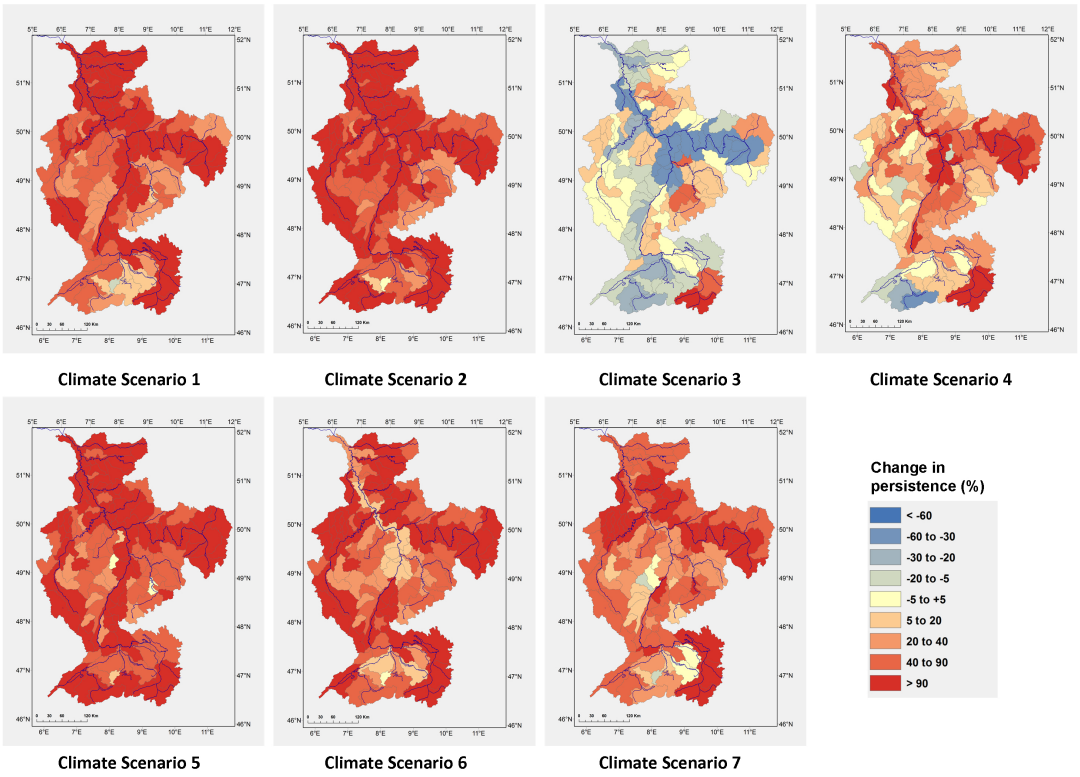


Figure 5.7: Relative change (%)* in low flow persistence in 134 sub-catchments based on simulated low flows using simulated inputs from seven climate scenarios for the future period 2063-2098 (case 4) compared to simulated low flows using simulated inputs for the reference period 1964-2007 (case 3).

* *(Future period - Current period) / Current period*

5.5 Discussion

For the River Rhine basin, a number of hydrological simulations were carried out using observed inputs and the outputs from an ensemble of seven climate scenarios. This was done to transfer the climate change signal from RCMs to a hydrological model and to evaluate the effects of climate change on the seasonality of low flows. The good low flow simulation performance of the hydrological model, i.e. an error of less than 5% in the simulation of the mean of minimum annual discharges (Eberle, 2005), was one of the reasons to select HBV for climate impact assessment. The difference between observed and simulated seasonality indices, and the change for the future climate, vary between the sub-basins. Moreover, the differences and changes also depend on the seasonality index considered. The dam operations, large lakes and the contribution of glacier storage are not explicitly incorporated in the HBV model structure (Berglöv *et al.*, 2009). However, all these factors are important for determining the seasonality characteristics of low flows and they can explain the significant differences between observed and simulated seasonality indices in the Rhine catchments and in the Alpine catchments in particular. This result is in line with that of Tallaksen and Van Lanen (2004), who found that the release from other large storages controlled by gravity, such as large lakes, snow storage and glaciers, can be important in sustaining low flows.

It appears from the results that the difference between observed and simulated indices is significantly larger compared to the change in the three indices between the current and future climate. This result is in line with that of Booij (2005) who found that the change with respect to the current climate conditions is like a systematic trend and much smaller than the uncertainty in modelling the extreme flow conditions.

The correlation coefficients between the three indices estimated from 134 catchments show that the seasonality ratio and weighted persistence indices are significantly negatively correlated. However, Figure 5.3 shows that the sub-catchments with lower seasonality ratio values (rainfed sub-catchments) show higher persistence. Similarly, the sub-catchments with higher seasonality ratio values (Alpine sub-catchments) experience low flow events in early winter months in the year compared to the downstream sub-catchments facing low flows in late summer. Therefore, the correlations are negative. It should be noted that the correlation coefficient between seasonality ratio and weighted persistence (i.e. -0.6) is higher than the correlation between seasonality ratio and weighted mean occurrence day (i.e. -0.4) and no significant correlation is found between weighted persistence and weighted mean occurrence day (i.e. 0.1). Regarding interrelations between

RCM outputs, as expected for time series resulting from stochastic processes in RCMs, no significant correlations were found (not shown). Moreover, in the IPCC special report on emission scenarios by Nakićenović and Swart (2000), it has been clearly stated that all A and B emission scenarios are equally valid with no assigned probabilities of occurrence.

The uncertainty originating from the RCMs, GCMs and emission scenarios is evaluated using the outputs from an ensemble of seven climate scenarios. If these seven climate scenarios are representative of climate change uncertainty, it appears from Figure 5.6 that the GCM and RCM uncertainty has the largest influence on weighted persistence. This result is in line with that of Prudhomme and Davies (2009) who found that the effect of emission scenario uncertainty was not larger than the effect of GCM uncertainty on the magnitude of changes in monthly summer flows. Further, the present findings seem to be consistent with other studies, which found that GCMs and RCMs were the most important sources of uncertainty in simulating climate change impacts on low flows (Wilby and Harris, 2006). Moreover, based on the ranges in average change in the three indices using simulated inputs from seven climate scenarios, shown in Figure 5.6, it appears that the influence of GCM and RCM uncertainty on seasonality ratio is slightly larger than the influence of emission scenario uncertainty on seasonality ratio, whereas the influence of GCM and RCM uncertainty on weighted mean occurrence day is similar to the influence of emission scenario on weighted mean occurrence day.

In this study, the errors induced by the hydrological model and observed inputs were not explicitly assessed as they are reported as less important than the uncertainty due to the climate predictions (Blenkinsop and Fowler, 2007; Muerth *et al.*, 2013). Further, the measurement errors in the observed discharges and the effect of different data lengths for the observed discharge series were implicitly addressed in this study. Nevertheless, it would be interesting to use a multi-model approach to assess model structural uncertainties and employing additional bias-correction techniques like quantile mapping (Teutschbein and Seibert, 2012; Gudmundsson *et al.*, 2012) to the outputs from different RCMs.

5.6 Conclusions

The results of this study about climate change impacts on the seasonality of low flows are based on a simulation approach using the outputs of an ensemble of climate models to drive a hydrological model. Three seasonality indices, namely the seasonality ratio (SR), weighted mean occurrence day ($WMOD$) and weighted persistence (WP), are used to reflect the discharge regime, timing and variability in timing of low flow events respectively. Our analysis focuses on the effects of the hydrological model and its inputs, the use of different GCMs and RCMs and the use of different emission scenarios. Sixteen model runs were considered. They are based on two periods, i.e. 1964-2007 and 2063-2098, four different GCMs, four different RCMs and three emission scenarios (A1B, A2 and B1). The 134 sub-catchments studied cover the entire River Rhine basin upstream of the Lobith gauging station at the Dutch-German border. They are representative of the different hydro-climatic regions and two distinct low flow regimes, winter and summer low flows, due to the Swiss Alps in the upstream part and rain-dominated catchments in the middle and downstream part of the basin. From the results presented in this study, we can draw the following conclusions.

- Significant differences have been found between seasonality indices based on observed low flows and simulated low flows with observed climate as input due to the uncertainty arising from hydrological model inputs and structure. The weighted mean occurrence day and the weighted persistence in the two Alpine sub-basins showed larger differences compared to the rain-dominated sub-basins.

- The comparison of the three seasonality indices based on observed inputs and simulated inputs reveals small differences in SR for all sub-basins except for the Moselle sub-basin. Large differences are found for the $WMOD$ and WP indices showing that these indices are very sensitive to uncertainties from the climate models.

- Based on the results of the comparison of the three seasonality indices using simulated inputs for the current climate and simulated inputs for the future climate, the largest range of change is found for WP, whereas the smallest range of change is found for SR. The SRs by 2063-2098 significantly decrease in all sub-basins, showing that a substantial change in the low flow regime in all sub-basins of the River Rhine is expected, whereas a regime shift from winter low flows to summer low flows is likely to occur in the two Alpine sub-basins. Further, the $WMOD$s of low flows tend to be earlier than for the current climate in all sub-basins except for the Middle Rhine and Lower

Rhine sub-basins. The *WP*s by 2063-2098 slightly increase, showing that the predictability of low flow events increases as the variability in timing decreases.

- From comparison of the uncertainty sources evaluated in this study, it is found that different RCMs and GCMs have a larger influence on the timing of low flows than different emission scenarios. The influence of different GCMs and RCMs on *SR* is slightly larger than the influence of different emission scenarios on SR, whereas the influence of different GCMs and RCMs on *WMOD* is similar to the influence of different emission scenarios on *WMOD*.

This study has evaluated the impacts of climate change on the seasonality of low flows in the River Rhine basin. A next step would be to assess the impacts of land use change on the seasonality of low flows and the relationship between groundwater seasonality and low flow seasonality. Furthermore, a detailed analysis of the climate change impacts on the return periods of extreme low flows is recommended.

5.7 Acknowledgements

We acknowledge the financial support of the Dr. Ir. Cornelis Lely Stichting (CLS), Project No. 20957310. The research is part of the programme of the Department of Water Engineering and Management at the University of Twente and it supports the work of the UNESCO-IHP VII FRIEND-Water programme. Discharge data for the River Rhine were provided by the Global Runoff Data Centre (GRDC) in Koblenz (Germany) and the Bundesamt für Umwelt (BAFU) in Bern (Switzerland). Catchment averaged precipitation, potential evapotranspiration data were supplied by the Federal Institute of Hydrology (BfG), Koblenz (Germany) and Deltares (the Netherlands). REGNIE grid data were extracted from the archive of the Deutscher Wetterdienst (DWD: German Weather Service), Offenbach (Germany). Bias-corrected forcing data were provided by Jules Beersma (KNMI) and Enno Nilson (BfG). The GIS base maps with delineated 134 sub-catchments of the River Rhine basin were provided by Eric Sprokkereef, the secretary general of the International Commission for the Hydrology of the Rhine basin (CHR). The stand-alone version of the HBV daily hydrological model environment was provided by Albrecht Weerts from Deltares (the Netherlands). The constructive review comments of Jan Seibert (Associate Editor), Renata Romanowicz and one anonymous reviewer significantly improved this paper.

Chapter 6

Discussion

6.1 Reflection on dominant low flow indicators and correlation analysis

In chapter 2, a framework was presented to assess the relative importance of pre-selected low flow indicators for the Rhine basin and to identify their appropriate lags and temporal resolutions. In this framework, the Rhine basin was divided into seven subbasins and correlations between observed low flows at the outlet of these subbasins and basin averaged low flow indicators were estimated. The low flow indicators were not arbitrarily selected, but based on previous studies (e.g. Belz and Frauenfelber-Kääb, 2007; Hurkmans *et al.*, 2010; Hurkmans *et al.*, 2008). I used the term 'indicator' rather than 'process' throughout the thesis, since not all the pre-selected indicators correspond to a hydrological process (e.g. lake level).

The lags and temporal resolutions of low flow indicators resulting in the best correlations were assumed to be appropriate temporal scales. This raises the question of validity of the presented evaluation results based on linear correlations between each indicator and the outlet discharge. The interactions between indicators are not considered in this thesis as these relations are difficult to assess. In practice, only slowly responding processes such as groundwater levels and lake and snow storages show linear behaviour (Wedgbrow *et al.*, 2002). Therefore, in chapter 2, not only linear correlations but three correlation coefficients measuring linear and non-linear relations between indicators and low flows were estimated. Since all three coefficients revealed similar results, only the Pearson correlation coefficients were presented. Finally, the schematization into seven sub-basins was assumed to be sufficient to understand the relationship between the pre-selected indicators and low flows at the basin outlets. Identification of low flow indicators at a higher spatial resolution than seven sub-basins would possibly be very

complicated due to interacting processes at small scales. Moreover, low flows are mostly sustained by base flows originating from groundwater aquifers having a much larger spatial scale than the typical scale of the 134 catchments.

6.2 Justification of selected models and model development

The model selection from available hydrological models was based on the identified dominant low flow indicators in chapter 2 and previous studies on the Rhine (Eberle, 2005; Te Linde *et al.*, 2008). The first conceptual model was GR4J with only four parameters, which was used since it has a parsimonious structure including a groundwater storage (routing store). This model has been tested over hundreds of catchments worldwide, for a broad range of climatic conditions from tropical to temperate and semi-arid catchments (Perrin *et al.*, 2003). Moreover, the model has been tested in numerous low flow studies in Europe (Pushpalatha *et al.*, 2012; Velázquez *et al.*, 2010; Velázquez *et al.*, 2011)

The second conceptual model was HBV, which has been used in the field of operational forecasting and climate impact modelling in more than 50 countries around the world (Şorman *et al.*, 2009), in North-Western Europe in particular (Görgen *et al.*, 2010; Driessen *et al.*, 2010; Engeland *et al.*, 2010; Te Linde *et al.*, 2008; Wöhling *et al.*, 2006; Booij, 2005). The good performance of the HBV model with a low number of parameters was the main advantage for application to large basins. In a previous study, this model was compared with a physically based model and showed better performance compared to the physically based model in the River Rhine (Te Linde *et al.*, 2008). The HBV model has been applied for the River Rhine since 1997 by the Dutch Water authorities, i.e. Rijkwaterstaat Waterdienst (previously RIZA) and Deltares, and the German Federal Institute of Hydrology (BfG).

The two conceptual models performed best in the Moselle subbasin. Therefore, this sub-basin was selected as a case study to apply a systematic uncertainty analysis. The uncertainty analysis required running the models 10.000 times for each forecast issue day within a Monte Carlo framework. A more detailed model structure, i.e. distributed and/ or physically-based, would make the uncertainty analysis infeasible. Therefore, a lumped model set-up was preferred in this thesis. The lags and temporal resolutions were identified not only for model selection but also for developing medium-range and seasonal forecasting models. In chapter 3, only the two selected conceptual models were used although data-driven models could also be developed. However, in chapter 4 two data-driven models, i.e. the

Artificial Neural Networks-Ensemble (ANN-Ensemble) and ANN-Indicator models, were developed. The ANN-Ensemble requires daily precipitation, potential evapotranspiration and historical discharge data as input. Historical discharge information from the previous day was used to update the model states. This is a one-day memory, which also exists in the conceptual models, i.e. GR4J and HBV. The ANN-Ensemble model is assumed to be comparable with conceptual models as it has a similar model structure. The ANN-Indicator model is a data-driven model incorporating only the three dominant low flow indicators for the Moselle subbasin, i.e. precipitation, potential evapotranspiration and groundwater. The model uses historical data and does not require forecasted meteorological data as input. The appropriate lags and temporal resolutions of these indicators have been taken from the results of chapter 2. The ANN-Ensemble and ANN-Indicator models with one hidden neuron were calibrated for seasonal low flow forecasts. The encouraging performance results of these models were presented in chapter 4.

6.3 Justification of model calibration and validation

All hydrological models require rigorous calibration for the specific purpose for which they are used, like for forecasting high flows or low flows. In this thesis, forecasting low flows was the main purpose of chapter 3 and 4. Therefore, a clear definition of low flows and related definitions of objective functions to calibrate the models were used. For that, I used Q75 as low flow threshold based on the exceedence probability of 75%, using the relevant reports by the principal authority for low flows in the Netherlands i.e. the Dutch National Coordinating Committee on Water Distribution (LCW). Low flows at this threshold are still affecting river navigation and the energy sector. Moreover, the number of days with low flows is sufficient to calibrate a forecast model.

The calibration period, i.e. 1971-2001, was selected as the first forecast issue date was 01/01/2002 and the number of low flow events (i.e. 567 days with low flows) in the calibration period was long enough for hydrological models (Perrin *et al.*, 2007). The validation period spans from 01/01/1951 to 31/12/1970.

In this thesis, I used two hybrid objective functions, each consisting of two single objective functions: *1)* hybrid Nash Sutcliffe coefficient (NS), and *2)* hybrid Mean Absolute Error (MAE). The first objective function was used in chapter 3 for 10 day low flow forecasts, whereas the second objective function was used in chapter 4 for 90 day low flow forecasts.

In both objective functions the common approach was that they focus on low flows by evaluating only low flows or inverse discharges. While the first objective

function estimates the Nash Sutcliffe coefficient, the second objective function uses the mean absolute difference between observed and simulated discharges. I could use only the NS of inverse discharges as objective function. However, it did not strongly influence the performance for low flow predictions. Similarly, I used a hybrid MAE to improve the ANN simulations. The two conceptual models, GR4J and HBV, were not influenced by a strict MAE based on only low flows as they have a sound physical description of the processes. However, the data-driven models, ANN-Ensemble and ANN-Indicator, meaningfully responded to input precipitation and potential evapotranspiration in non-low flow days only after incorporating a hybrid MAE function.

6.4 Assumptions and limitations in the uncertainty analysis framework

In chapter 3, I applied a systematic uncertainty analysis to low flow models to identify where uncertainties come from and to provide quantified model output uncertainty information for a robust model comparison. Therefore, I focused on three sources of uncertainty, i.e. model input, parameters and initial conditions, for further analysis. Quantification of the uncertainty sources is probably the most difficult step of an uncertainty analysis. Uncertainty in forecasted input data, e.g. precipitation and temperature, is mainly from the assumptions and simplifications made when describing atmospheric processes in weather forecast models. In particular, future precipitation amounts are assumed to be very uncertain (Roulin, 2007; Cunha *et al.*, 2012). To quantify the uncertainty in the weather forecasts, an ensemble of low resolution forecasts (ENS) has been generated by the European Medium Range Weather Forecasting Centre (ECMWF) and other national meteorological services (ECMWF, 2012).

To calibrate the two conceptual models and assess uncertainty originating from model parameters in chapter 3, we used the Generalised Likelihood Uncertainty Estimation (GLUE) method. For assessing uncertainty originating from model parameters and initial conditions, I used a Monte Carlo framework. For that, we generated 120,000 parameter sets for each conceptual model using Latin Hypercube Sampling (LHS). To my knowledge, this is the largest LHS sample size tested in low flow hydrology. Due to the lack of prior knowledge about the true distributions, independent uniform distributions were chosen for each parameter. Compared to a standard GLUE random sampling, LHS substantially reduces the computational burden for sampling and provide a 10-fold greater efficiency in parameter space coverage (Shen *et al.*, 2012). The sampling size used in my the-

sis was found to be large enough to ensure a sufficient calibration of the model as the highest NS_{hybrid} values did not change using another global optimisation technique, i.e. a Genetic Algorithm (Velázquez *et al.*, 2010).
The GR4J and HBV models are run for each of the 120,000 sets in the calibration. The output is evaluated against the observed daily discharge at Cochem station, which is located at the outlet of the Moselle sub-basin, using the NS_{hybrid} likelihood function to distinguish between behavioural parameter sets (accepted) and non-behavioural parameter sets (rejected). The parameter sets meeting the predefined threshold criterion (NS_{hybrid}>0.40) are accepted. Although the threshold value is a subjective decision (Jin *et al.*, 2010), I rigorously tested several thresholds based on low flow simulations and the size of the behavioural parameter sets for each model. The selected threshold resulted in two large behavioural parameter sets for parameter uncertainty analysis, namely 9770 × 4 (GR4J) corresponding to ∼ 8% of the sample parameter set and 10909 × 8 (HBV) corresponding to ∼ 9% of the sample parameter set.
The dynamic storage update was used in my thesis after having been tested on observed data. Compared to complex state updating procedures, the proposed new dynamic update procedure worked efficiently for assessing initial condition uncertainties using 10.000 runs each day in the test period 2002–2005.
As reported in chapter 3, the drawbacks and advantages of the GLUE method have been discussed in the hydrological literature. Concerning the choice of the likelihood measure, Beven and Binley (1992) pointed out that many different likelihood measures in GLUE can be appropriate for a given application. Jin *et al.* (2010) compared different likelihood measures and the model uncertainty. They found that a less strict likelihood function, obviously leads to a wider confidence interval of the output uncertainty. Therefore, neither a too strict nor a too relaxed likelihood is appropriate for the GLUE assessment. Stedinger *et al.* (2008) showed that GLUE can produce reasonable uncertainty and prediction intervals using a correct likelihood function. My thesis utilized an elaborated low flow likelihood function and a large parameter space within the GLUE method which resulted in meaningful uncertainty ranges for the GR4J and HBV models.

6.5 Applicability of methods

The framework presented in chapter 2 is applicable to all river basin types and all regime types (not only low flows) for identification of dominant indicators and their appropriate lags and temporal resolutions. For small river basins, subdivision of the river basin into sub-basins is not necessary. Therefore, the spatial scale of the assessment should be carefully selected to capture the signal of the

indicators. I used 3-day and 7-day window sizes for moving averaging of the low flows. This pre-selection of the window size should be determined based on the lead time and appropriate temporal scale accepted for the discharge forecast accuracy. In my case, the correlations were calculated for a lead time of 14 days and 3-day temporally-averaged low flows and for a lead time of 90 days and 7-day temporally averaged low flows.

The methodology presented in chapter 3 helped to quantify uncertainty originating from model inputs, parameters and initial conditions for two conceptual models. In principal, the steps followed in this methodology can be applied to other geographic locations as well. However, the two dynamic storage update procedures and subsequently developed empirical models for dividing observed discharges into fast and slow runoff components are model specific approaches and, therefore, they are limited to the GR4J and HBV models.

The methodology presented in chapter 4 helped in assessing the effect of ensemble seasonal forecasts on low flow forecasts. The selected five cases consisting of different input combinations were tested. A similar approach can be applied to other geographical areas and other regime types for evaluating the effect of model inputs on the forecasts. The objective function based on the mean absolute error (chapter 4) can be applied to all other low flow calibration problems, data-driven models in particular.

In this thesis, two data-driven models were developed in chapter 4. While one of these models, ANN-Ensemble, uses daily forecasted weather data, the ANN-Indicator model requires only historical data for 90 day ahead low flow forecasts. The results of ANN-Ensemble are comparable with those of GR4J and HBV. Despite the successful results of ANN-Indicator, there are still limitations to the applicability of this model:

1) The model is area dependent as its input and temporal scales were chosen for the Moselle sub-basin.

2) The model is limited to low flow forecasts as the model is calibrated and validated for observed low flows.

The methodology to develop ANN models for seasonal forecasts as described in chapter 4, can be generalized to any other river basin in the world. Particularly the ANN-Indicator type of model can be very useful for regions where seasonal climate forecast data are not available. The two low flow seasonality indices - weighted mean occurrence day and weighted persistence - are two indices similar to the seasonality indices of Laaha and Blöschl (2006). The new aspect added to these indices is the weighting factor based on inverse discharges. In other words, the day number of each low flow event was weighted by the inverse low flow value

at the same day to address the severity of a low flow event as well as its occurrence day. These weighted indices are applicable to all hydrological events.

6.6 Applicability of results

The identified lags and temporal resolutions in chapter 2 are useful in creating operational medium-range and seasonal low flow forecast models for the River Rhine. The results can be useful for other rain dominated catchments in the world. Further, understanding the low flow mechanisms and subsequent storage responses should aid the selection of appropriate models and the choice of proper temporal scales. Critical catchment characteristics (e.g. the extent of aquifers) determine the applicability of the results for appropriate model selection.
Most important uncertainty sources for low flow forecasts and future changes in low flows were quantified in my thesis. The assessment of low flow forecasts for lead times of 10 day and 90 day was limited to the Moselle River, whereas the impacts of climate change on the seasonality of low flows were assessed for the sub-catchments of the River Rhine. Anticipating low flows would facilitate strategic decisions regarding the various river functions affected by low flows (e.g. navigation, cooling water supply) since more frequent low flows as well as extreme flood peaks are expected in the future (Hurkmans *et al.*, 2010). The results of chapter 3 showed that parameter uncertainty had the largest effect whereas input uncertainty had the smallest effect on the medium range low flow forecasts. The order of uncertainty and the type of models can be applicable for other rainfed basins in the world. In chapter 4, we found that the uncertainty arising from ensemble precipitation had a larger effect on seasonal low flow forecasts than that from ensemble potential evapotranspiration and model initial conditions. A comparison of the uncertainty sources evaluated in chapter 5 showed that the uncertainty in RCM and GCM outputs has the largest influence on the seasonality of low flows in a future climate. These uncertainty sources should be treated carefully in low flow modelling practices and in the chain of decision making processes.

Chapter 7

Conclusions

The objective of this thesis was:

To explore low flow mechanisms, develop forecast methods and assess climate change impacts on low flows by identifying low flow indicators and their dominant temporal scales, analysing the effects of different uncertainty sources on low flow forecasts for different lead times, and comparing low flows for current and future climate conditions.

Four research questions were formulated to fulfil this objective. This concluding chapter starts with reflecting on the objective of this thesis and answering the research questions (Section 7.1). Subsequently, I provide recommendations for future work (Section 7.2).

7.1 Reflection on research objective and questions

I have identified important low flow mechanisms in the Rhine basin. Based on these mechanisms I have selected two conceptual models for 10 day and 90 day forecasts. Besides I have developed two data-driven models in which these mechanisms with appropriate lag and temporal resolutions were incorporated for 90 day ahead low flow forecasts. The effects of major uncertainty sources on low flow forecasts are assessed using Monte Carlo methods. Parameter uncertainty is found to have the largest effect on 10 day low flow forecasts, whereas ensemble seasonal precipitation forecasts have the largest effect on 90 day low flow forecasts. In the final step, I assessed the possible changes in seasonal low flow characteristics for the period 2063-2098 using important indices representing the low flow regime, timing and persistence.

Overall, the results of my thesis are assumed to be useful for hydrological modelers in general and decision makers in the River Rhine basin focusing on low flow events in particular.

Q1: *What are the dominant low flow indicators and temporal scales in the Rhine Basin?*

The analysis of pre-selected low flow indicators in chapter 2 has shown that precipitation, potential evapotranspiration, groundwater, lake levels and snow depth in the Alps are important indicators for low flow forecasting for lead times of 10 and 90 days. In the other sub-basins, groundwater and potential evapotranspiration correlate best to low flows. The correlation analysis also revealed the relative importance of these indicators for varying lags and temporal resolutions. Three correlation coefficients, focusing on both linear and non-linear relations, showed similar results. For seven different subbasins, correlations were assessed between low flow indicators and observed low flows at the outlet of the subbasins. The seven sub-basins were defined by spatially aggregating 134 catchments in the River Rhine basin according to similar hydrological characteristics and based on previous studies (e.g. Belz and Frauenfelber-Kääb, 2007; Hurkmans *et al.*, 2008). After selecting seven sub-basins and five low flow indicators, a classical standardisation method was applied to avoid effects of spatial heterogeneity in the data. The standardised low flow indicator data have a zero mean and a standard deviation of one. The correlations were estimated by varying the temporal resolutions of the indicators from one day to 336 days and the lag between the indicator and low flows from zero to 210 days.

The most important indicators in the Alpine sub-basins for 10 day low flow forecasts are potential evapotranspiration and lake levels. In the other subbasins, groundwater and potential evapotranspiration are most relevant for low flows. The most important indicators for 90 day low flow forecasts are potential evapotranspiration, lake levels and snow depths for the Alpine sub-basins, whereas in the other sub-basins the most important indicators are potential evapotranspiration and precipitation or groundwater. Overall, small lags and temporal resolutions are effective for lake levels and groundwater in sub-basins for medium-range and seasonal low flow forecasts, whereas large lags and temporal resolutions are important for precipitation, potential evapotranspiration and snow depth. The uncertainty in the appropriate lags and temporal resolutions was found to be large for most low flow indicators. In all sub-basins the largest uncertainties were found in potential evapotranspiration, whereas lake levels and

groundwater show small uncertainties. The indicators were used for model selection and model development in chapter 3 and 4.

Q2: *What is the effect of uncertainty originating from model inputs, parameters and initial conditions on 10 day ensemble low flow forecasts?*

To answer this question, I applied a systematic uncertainty analysis to 10 day ahead low flow forecasts in chapter 3, to provide quantified model output uncertainty information for a robust model comparison. A hybrid performance metric was used for evaluating low flow simulations whereas the quality of the probabilistic low flow forecasts was assessed based on relative confidence intervals, reliability and hit/false alarm rates.

The 10 day ensemble forecast results show that the daily observed low flows are captured by the 90% confidence interval for both models most of the time, whereas the GR4J model usually overestimates low flows and HBV is prone to underestimate low flows. This is particularly the case if the parameter uncertainty is included into the forecasts. The total uncertainty in the GR4J model outputs is higher than in the HBV model. Moreover, the parameter uncertainty has the largest effect and the input uncertainty has the smallest effect on the low flow forecasts. The parameter uncertainty for 10 day low flow forecasts issued by the HBV model with eight parameters was smaller than the parameter uncertainty coming from the GR4J with four parameters. This was because the rainfall-runoff process resulting in low flows in the study area is better described by the HBV model. The forecast distribution based on 10 day low flow forecasts (i.e. Q75) issued by the HBV model was the most reliable forecast distribution if only input uncertainty is considered. The number of hits is about equal for the two models only if the input uncertainty is considered. The parameter uncertainty was the main reason reducing the number of hits. The number of false alarms using the GR4J model was almost doubled, with respect to the HBV model, considering all uncertainty sources. The importance of parameter uncertainty on the quality of forecasts was emphasized by all forecast quality measures used in this study.

Overall, the systematic uncertainty analysis of different sources in chapter 3 has shown that the output from two conceptual hydrological models is characterised by substantial uncertainty from model parameters. The parameter uncertainty mainly affects the reliability and the sharpness of the forecasts. This finding is new for low flow forecasts; as the significance of the rainfall prediction error is well known and documented already for high flows (Pappenberger *et al.*, 2005).

Q3: *What is the effect of ensemble seasonal meteorological forecasts on the skill of seasonal ensemble low flow forecasts?*

The effect of ensemble seasonal climate forecasts, consisting of 40-members, on the skill of low flow forecasts has been assessed for varying lead times in chapter 4. For that purpose, four hydrological models - GR4J, HBV, ANN-Ensemble and ANN-Indicator - have been run using different seasonal meteorological forcings: (1) ensemble precipitation (P) and potential evapotranspiration (PET) forecasts, (2) ensemble P and climate mean PET (3) climate mean P and ensemble PET, (4) climate mean P and PET, and (5) zero P and ensemble PET. The ensembles provided the forecast uncertainty range for the model inputs. The ranges were compared for a lead time of 90 days, whereas the skill of low flow forecasts was evaluated for varying lead times up to 90 days. Based on the results of chapter 4 comparing the five cases above, all models were prone to over-predict low flows using ensemble forcing and the largest range for 90 day low flow forecasts was found for GR4J using ensemble seasonal climate forecasts as input. The results of the comparison of forecast skills with varying lead times showed that the low flow forecasts using GR4J were less skilful than using the other three models. Further, the hit rate of ANN-Ensemble was higher than that of the other models for all lead times except for a 90 day lead time. The 90 day ahead low flow days in a very dry year, i.e. 2003, are correctly forecasted by ANN-Indicator which illustrates the skill of data-driven models for seasonal forecasting. Overall, the uncertainty arising from ensemble P had a larger effect on seasonal low flow forecasts than uncertainties from ensemble PET and model initial conditions.

Q4: *What is the impact of climate change on the seasonality of low flows in the Rhine basin?*

The impacts of climate change on the seasonality of low flows were assessed in chapter 5 based on a simulation approach using the outputs of an ensemble of climate models to drive a hydrological model. Three seasonality indices, namely seasonality ratio (SR), weighted mean occurrence day ($WMOD$) and weighted persistence (WP) were used to reflect the discharge regime, timing and variability in timing of low flow events respectively. The impact analysis focused on the effects of the hydrological model and its inputs, the use of different GCMs and RCMs and the use of different emission scenarios. Sixteen experiments were considered. They were based on two periods, i.e. 1964-2007 and 2063-2098, four different GCMs, four different RCMs and three emission scenarios (A1B, A2

and B1). The 134 sub-catchments studied cover the entire River Rhine basin upstream of the Lobith gauging station at the Dutch-German border. These sub-catchments are representative of the different hydro-climatic regions and two distinct low flow regimes, winter and summer low flows, due to the Swiss Alps in the upstream part and rain-dominated catchments in the middle and downstream part of the basin respectively. In chapter 5, I found significant differences between seasonality indices based on observed low flows and simulated low flows with observed climate as input due to the uncertainty arising from hydrological model inputs and structure. Further, the weighted mean occurrence day and the weighted persistence in the two Alpine sub-basins showed larger differences compared to the rain-dominated sub-basins. The comparison of the three seasonality indices based on observed inputs and simulated inputs revealed small differences in *SR* for all sub-basins except for the Moselle sub-basin. Large differences were found for the *WMOD* and *WP* indices, showing that these indices were very sensitive to uncertainties from the climate models. The comparison of the three seasonality indices using simulated inputs for the current climate and simulated inputs for the future climate resulted in the largest range for *WP*, whereas the smallest range was found for *SR*. The *SR*s by 2063-2098 significantly decrease in all sub-basins, showing that a substantial change in the low flow regime in all sub-basins of the River Rhine is expected, whereas a regime shift from winter low flows to summer low flows is likely to occur in the two Alpine sub-basins. Moreover, the *WMOD*s of low flows tend to be earlier than for the current climate in all sub-basins except for the Middle and Lower Rhine sub-basins. The *WPs* by 2063-2098 slightly increase when compared to the current climate, showing that the predictability of low flow events increases as the variability in timing decreases. Overall, the comparison of the uncertainty sources evaluated in this chapter showed that differences in RCMs and GCMs have more influence on the timing of low flows (*WMOD*) than differences in emission scenarios. The influence of different GCMs and RCMs on *SR* was slightly larger than the influence of different emission scenarios on *SR*, whereas the influence of different GCMs and RCMs on *WMOD* was similar to the influence of different emission scenarios on *WMOD*.

7.2 Recommendations for further research

For further research on low flows, I recommend:

- to investigate the skill of different comprehensive techniques such as wavelet coherence analysis, chaotic correlation dimension analysis, canonical correlation analysis and principal component analysis for identifying the number of dominant processes and spatiotemporal scales in the river basin. In this thesis, correlation analysis was carried out to determine the relationship between observed low flows and pre-selected indicators with varying lag times and temporal resolutions. However, the recommended techniques can reveal new insights from the hydrological time series like the number of dominant processes.

- to assess the effect of parameter uncertainty on seasonal low flow forecasts. In this thesis, only the effect of ensemble seasonal climate forecasts (input uncertainty) on low flow forecasts was assessed.

- to assess uncertainty in forecasts for longer lead times such as 14 days and up to 6 months as the data are already available from the ECMWF Mars system,

- to use new climate data sets based on the new forthcoming IPCC Assessment Reports (e.g. AR5 working group I),

- to correct biases in meterological forcing data using climate statistics relevant for low flows.

Bibliography

Ackermann, T., D. P. Loucks, D. Schwanenberg, and M. Detering (2000), Real-time modeling for navigation and hydropower in the River Mosel, *J. Water Resour. Plan. Manage.-ASCE*, **126** (5), 298-303.

Adamowski, J., H. F. Chan, S. O. Prasher, B. Ozga-Zielinski, and A. Sliusarieva (2012), Comparison of multiple linear and nonlinear regression, autoregressive integrated moving average, artificial neural network, and wavelet artificial neural network methods for urban water demand forecasting in Montreal, Canada, *Water Resour. Res.*, **48** (1), W01528, doi: 10.1029/2010wr009945.

Addor, N., S. Jaun, F. Fundel, and M. Zappa (2011), An operational hydrological ensemble prediction system for the city of Zurich (Switzerland): skill, case studies and scenarios, *Hydrol. Earth Syst. Sci.*, **15** (7), 2327-2347.

Allen, D. M., P. H. Whitfield, and A. Werner (2010), Groundwater level responses in temperate mountainous terrain: regime classification, and linkages to climate and streamflow, *Hydrol. Processes*, **24** (23), 3392-3412.

Anderson, M. L., Z. Q. Chen, and M. L. Kavvas (2004), Modeling low flows on the Cosumnes River, *J. Hydrol. Eng.*, **9** (2), 126-134.

Archer, D. R., and H. J. Fowler (2008), Using meteorological data to forecast seasonal runoff on the River Jhelum, Pakistan, *J. Hydrol.*, **361** (1-2), 10-23.

Arnaud, P., J. Lavabre, C. Fouchier, S. Diss, and P. Javelle (2011), Sensitivity of hydrological models to uncertainty in rainfall input, *Hydrol. Sci. J.*, **56** (3), 397-410.

Arnell, N. W. (1999), The effect of climate change on hydrological regimes in Europe: a continental perspective, *Global Environ Chang*, **9** (1), 5-23.

Arnell, N. W., and S. N. Gosling (2013), The impacts of climate change on river flow regimes at the global scale, *J. Hydrol.*, **486**, 351-364.

ATV-DVWK (2002), Verdunstung in Bezug zu Landnutzung, Bewuchs und Boden, Merkblatt ATV-DVWK-M 504, Hennef.

Baldwin, C. K., and U. Lall (1999), Seasonality of streamflow: The upper Mississippi River, *Water Resour. Res.*, **35** (4), 1143-1154.

Behrmann-Godel, J., and R. Eckmann (2003), A preliminary telemetry study of the migration of silver European eel (Anguilla anguilla L.) in the River Mosel, Germany, *Ecology of Freshwater Fish*, **12** (3), 196-202.

Bell, V. A., H. N. Davies, A. L. Kay, T. J. Marsh, A. Brookshaw, and A. Jenkins (2013), Developing a large-scale water-balance approach to seasonal forecasting: application to the 2012 drought in Britain, *Hydrol. Processes*, doi: 10.1002/hyp.9863.

Belz, J. U. (2010), The flow regime of the River Rhine and its tributaries in the 20 th century- analysis, changes, trends, *Hydrologie und Wasserbewirtschaftung*, **54** (1), 4-17.

Belz, J. U., and R. Frauenfelber-Kääb (2007), *Das Abflussregime des Rheins und seiner Nebenflüsse im 20. Jahrhundert : Analyse, Veränderungen, Trends (in German)*, 390 pp., KHR / CHR, Lelystad.

Belz, J. U., N. Busch, H. Engel, and G. Gasber (1999), Comparative survey of measures on river training at Upper Rhine, Moselle and Saar and their effects on flood behaviour, *Vergleichende Darstellung der Ausbaumassnahmen an Oberrhein, Mosel und Saar und ihre Auswirkungen auf Hochwasser*, **43** (6), 283-292.

Bennett, K. E., A. T. Werner, and M. Schnorbus (2012), Uncertainties in Hydrologic and Climate Change Impact Analyses in Headwater Basins of British Columbia, *J. Clim.*, **25** (17), 5711-5730.

Berglöv, G., J. German, H. Gustavsson, U. Harbman, and B. Johansson (2009), Improvement HBV Model Rhine in FEWS, Final Report, SMHI Hydrology, Norrköping, Sweden.

Bergström, S. (1976), *Development and application of a conceptual runoff model for scandinavian catchments*. PhD Thesis. SMHI Reports RHO No. 7, Norrköping.

Beven, K. (2006), On undermining the science?, *Hydrol. Processes*, **20** (14), 3141-3146.

Beven, K., and A. Binley (1992), The Future of Distributed Models - Model Calibration and Uncertainty Prediction, *Hydrol. Processes*, **6** (3), 279-298.

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

Beven, K., and P. Young (2003), Comment on " Bayesian recursive parameter estimation for hydrologic models" by M. Thiemann, M. Trosset, H. Gupta, and S. Sorooshian, *Water Resour. Res.*, **39** (5), 1116.

Beven, K., P. Smith, and J. Freer (2007), Comment on " Hydrological forecasting uncertainty assessment: Incoherence of the GLUE methodology" by Pietro Mantovan and Ezio Todini, *J. Hydrol.*, **338** (3-4), 315-318.

Beven, K. J., P. J. Smith, and J. E. Freer (2008), So just why would a modeller choose to be incoherent?, *J. Hydrol.*, **354** (1-4), 15-32.

Beyene, T., D. P. Lettenmaier, and P. Kabat (2010), Hydrologic impacts of climate change on the Nile River Basin: implications of the 2007 IPCC scenarios, *Climatic Change*, **100** (3-4), 433-461.

Bierkens, M. F. P., and L. P. H. van Beek (2009), Seasonal Predictability of European Discharge: NAO and Hydrological Response Time, *J. Hydrometeorol.*, **10** (4), 953-968.

Blasone, R. S., J. A. Vrugt, H. Madsen, D. Rosbjerg, B. A. Robinson, and G. A. Zyvoloski (2008), Generalized likelihood uncertainty estimation (GLUE) using adaptive Markov chain Monte Carlo sampling, *Adv. Water Resour.*, **31** (4), 630-648.

Blenkinsop, S., and H. J. Fowler (2007), Changes in drought frequency, severity and duration for the British Isles projected by the PRUDENCE regional climate models, *J. Hydrol.*, **342** (1-2), 50-71.

Booij, M. J. (2002a), Extreme daily precipitation in Western Europe with climate change at appropriate spatial scales, *Int. J. Climatol.*, **22** (1), 69-85.

Booij, M. J. (2002b), Modelling the effects of spatial and temporal resolution of rainfall and basin model on extreme river discharge, *Hydrol. Sci. J.*, **47** (2), 307-320.

Booij, M. J. (2003), Determination and integration of appropriate spatial scales for river basin modelling, *Hydrol. Processes*, **17** (13), 2581-2598.

Booij, M. J. (2005), Impact of climate change on river flooding assessed with different spatial model resolutions, *J. Hydrol.*, **303** (1-4), 176-198.

Booij, M. J., and M. J. M. de Wit (2010), Extreme value statistics for annual minimum and trough-under-threshold precipitation at different spatio-temporal scales, *Hydrol. Sci. J.*, **55** (8), 1289-1301.

Bormann, H. (2010), Runoff Regime Changes in German Rivers Due to Climate Change, *Erdkunde*, **64** (3), 257-279.

Bosshard, T., M. Carambia, K. Görgen, S. Kotlarski, P. Krahe, M. Zappa, and C. Schar (2013), Quantifying uncertainty sources in an ensemble of hydrological climate-impact projections, *Water Resour. Res.*, **49** (3), 1523-1536.

Bröcker, J., and L. A. Smith (2007), Increasing the reliability of reliability diagrams, *Weather and forecasting*, **22** (3), 651-661.

Burn, D. H., J. M. Buttle, D. Caissie, G. MacCulloch, C. Spence, and K. Stahl (2008), The Processes, Patterns and Impacts of Low Flows Across Canada, *Can. Water Resour. J.*, **33** (2), 107-123.

Butts, M. B., J. T. Payne, M. Kristensen, and H. Madsen (2004), An evaluation of the impact of model structure on hydrological modelling uncertainty for streamflow simulation, *J. Hydrol.*, **298** (1-4), 242-266.

Chen, J., F. P. Brissette, A. Poulin, and R. Leconte (2011), Overall uncertainty study of the hydrological impacts of climate change for a Canadian watershed, *Water Resour. Res.*, **47** (12), W12509, doi: 10.1029/2011wr010602.

Chiew, F. H. S., S. L. Zhou, and T. A. McMahon (2003), Use of seasonal streamflow forecasts in water resources management, *J. Hydrol.*, **270** (1-2), 135-144.

Chowdhury, S., and A. Sharma (2009), Multisite seasonal forecast of arid river flows using a dynamic model combination approach, *Water Resour. Res.*, **45** (10), W10428, doi: 10.1029/2008wr007510.

Cloke, H. L., and F. Pappenberger (2009), Ensemble flood forecasting: A review, *J. Hydrol.*, **375** (3-4), 613-626.

Cloke, H. L., and D. M. Hannah (2011), Large-scale hydrology: advances in understanding processes, dynamics and models from beyond river basin to global scale, *Hydrol. Processes*, **25** (7), 991-995.

Coley, D. M., and P. R. Waylen (2006), Forecasting dry season streamflow on the Peace River at Arcadia, Florida, USA, *J. Am. Water Resour. Assoc.*, **42** (4), 851-862.

Cunha, L. K., P. V. Mandapaka, W. F. Krajewski, R. Mantilla, and A. A. Bradley (2012), Impact of radar-rainfall error structure on estimated flood magnitude across scales: An investigation based on a parsimonious distributed hydrological model, *Water Resour. Res.*, **48** (10), W10515, doi: 10.1029/2012wr012138.

De Bruijn, K. M., and R. Passchier (2006), Low-flow forecasts for the Rhine (Report number: Q3427), WL — Delft Hydraulics, Delft, the Netherlands.

De Vos, N. J., and T. H. M. Rientjes (2008), Multiobjective training of artificial neural networks for rainfall-runoff modeling, *Water Resour. Res.*, **44** (8), W08434, doi: 10.1029/2007wr006734.

De Wit, M. J. M., B. van den Hurk, P. M. M. Warmerdam, P. Torfs, E. Roulin, and W. P. A. van Deursen (2007), Impact of climate change on low-flows in the river Meuse, *Climatic Change*, **82** (3), 351-372.

DeChant, C. M., and H. Moradkhani (2011), Improving the characterization of initial condition for ensemble streamflow prediction using data assimilation, *Hydrol. Earth Syst. Sci.*, **15** (11), 3399-3410.

Demirel, E., J. van Ommeren, and P. Rietveld (2010), A matching model for the backhaul problem, *Transport Res B-Meth*, **44** (4), 549-561.

Demirel, M. C., M. J. Booij, and A. Y. Hoekstra (2013a), Identification of appropriate lags and temporal resolutions for low flow indicators in the River Rhine to forecast low flows with different lead times, *Hydrol. Processes*, **27** (19), 2742-2758.

Demirel, M. C., M. J. Booij, and A. Y. Hoekstra (2013b), Effect of different uncertainty sources on the skill of 10 day ensemble low flow forecasts for two hydrological models, *Water Resour. Res.*, **49** (7), 4035-4053.

Demuth, S. (2005), Low flows and droughts-A European prospective, IAHS Newsletter 82, March 2005, pp. 7-8.

Devineni, N., A. Sankarasubramanian, and S. Ghosh (2008), Multimodel ensembles of streamflow forecasts: Role of predictor state in developing optimal combinations, *Water Resour. Res.*, **44** (9), W09404, doi: 10.1029/2006wr005855.

Diaz-Nieto, J., and R. L. Wilby (2005), A comparison of statistical downscaling and climate change factor methods: impacts on low flows in the River Thames, United Kingdom, *Climatic Change*, **69** (2), 245-268.

Doblas-Reyes, F. J., A. Weisheimer, M. Déqué, N. Keenlyside, M. McVean, J. M. Murphy, P. Rogel, D. Smith, and T. N. Palmer (2009), Addressing model uncertainty in seasonal and annual dynamical ensemble forecasts, *Q. J. R. Meteorol. Soc.*, **135** (643), 1538-1559.

Driessen, T. L. A., R. T. W. L. Hurkmans, W. Terink, P. Hazenberg, P. J. J. F. Torfs, and R. Uijlenhoet (2010), The hydrological response of the Ourthe catchment to climate change as modelled by the HBV model, *Hydrol. Earth Syst. Sci.*, **14** (4), 651-665.

Dumont, E., E. J. Bakker, L. Bouwman, C. Kroeze, R. Leemans, and A. Stein (2008), A framework to identify appropriate spatial and temporal scales for modeling N flows from watersheds, *Ecological Modelling*, **212** (3-4), 256-272.

Dutra, E., F. Di Giuseppe, F. Wetterhall, and F. Pappenberger (2013), Seasonal forecasts of droughts in African basins using the Standardized Precipitation Index, *Hydrol. Earth Syst. Sci.*, **17** (6), 2359-2373.

Eberle, M. (2005), Hydrological Modelling in the River Rhine Basin Part III - Daily HBV Model for the Rhine Basin BfG-1451, Institute for Inland Water Management and Waste Water Treatment (RIZA) and Federal Institute of Hydrology (BfG) Koblenz, Germany

ECMWF (2012), Describing ECMWF's forecasts and forecasting system, in *ECMWF newsletter 133*, edited by B. Riddaway.

Elshorbagy, A., G. Corzo, S. Srinivasulu, and D. P. Solomatine (2010), Experimental investigation of the predictive capabilities of data driven modeling techniques in hydrology - Part 1: Concepts and methodology, *Hydrol. Earth Syst. Sci.*, **14** (10), 1931-1941.

Engeland, K., B. Renard, I. Steinsland, and S. Kolberg (2010), Evaluation of statistical models for forecast errors from the HBV model, *J. Hydrol.*, **384** (1-2), 142-155.

EU (2013), Horizon 2020 - Work Programme 2014-2015: Water 7_2015: Increasing confidence in seasonal-to-decadal predictions of the water cycle.

Ewen, J., G. O'Donnell, A. Burton, and E. O'Connell (2006), Errors and uncertainty in physically-based rainfall-runoff modelling of catchment change effects, *J. Hydrol.*, **330** (3-4), 641-650.

Felipe, P.-S., and O.-N. Nelson (2009), Forecasting of Monthly Streamflows Based on Artificial Neural Networks, *J. Hydrol. Eng.*, **14** (12), 1390-1395.

Franz, K. J., and T. S. Hogue (2011), Evaluating uncertainty estimates in hydrologic models: borrowing measures from the forecast verification community, *Hydrol. Earth Syst. Sci.*, **15** (11), 3367-3382.

Freer, J., K. J. Beven, and B. Ambroise (1996), Bayesian Estimation of Uncertainty in Runoff Prediction and the Value of Data: An Application of the GLUE Approach, *Water Resour. Res.*, **32** (7), 2161-2173.

Fundel, F., and M. Zappa (2011), Hydrological ensemble forecasting in mesoscale catchments: Sensitivity to initial conditions and value of reforecasts, *Water Resour. Res.*, **47** (9), W09520, doi: 10.1029/2010wr009996.

Fundel, F., S. Jörg-Hess, and M. Zappa (2013), Monthly hydrometeorological ensemble prediction of streamflow droughts and corresponding drought indices, *Hydrol. Earth Syst. Sci.*, **17** (1), 395-407.

Ganguli, P., and M. J. Reddy (2013), Ensemble prediction of regional droughts using climate inputs and SVM-copula approach, *Hydrol. Processes*, (accepted), doi: 10.1002/hyp.9966.

Gaume, E., and R. Gosset (1999), Over-parameterisation, a major obstacle to the use of artificial neural networks in hydrology?, *Hydrol. Earth Syst. Sci.*, **7** (5), 693-706.

Giuntoli, I., B. Renard, J. P. Vidal, and A. Bard (2013), Low flows in France and their relationship to large-scale climate indices, *J. Hydrol.*, **482**, 105-118.

Gobena, A. K., and T. Y. Gan (2010), Incorporation of seasonal climate forecasts in the ensemble streamflow prediction system, *J. Hydrol.*, **385** (1-4), 336-352.

Görgen, K., et al. (2010), Assessment of Climate Change Impacts on Discharge in the Rhine River Basin: Results of the RheinBlick 2050 Project, Lelystad, CHR, ISBN 978-90-70980-35-1, 211p., Available at: www.chr-khr.org/files/CHR_I-23.pdf, [Accessed: 24 September 2013].

Götzinger, J., and A. Bárdossy (2008), Generic error model for calibration and uncertainty estimation of hydrological models, *Water Resour. Res.*, **44**, W00B07, doi: 10.1029/2007wr006691.

Govindaraju, R. S., and A. R. Rao (2000), *Artificial Neural Networks in Hydrology*, 329 pp., Kluwer Academic Publishers Norwell, MA, USA.

Grabs, W., K. Daamen, and F. de Montmollin (1997), *Impact of Climate Change on Hydrological Regimes and Water Resources Management in the Rhine Basin (CHR-report I-16)*, CHR/KHR, Lelystad.

Graham, L., S. Hagemann, S. Jaun, and M. Beniston (2007), On interpreting hydrological change from regional climate models, *Climatic Change*, **81**, 97-122.

Gudmundsson, L., L. M. Tallaksen, and K. Stahl (2011), Spatial cross-correlation patterns of European low, mean and high flows, *Hydrol. Processes*, **25** (7), 1034-1045.

Gudmundsson, L., J. B. Bremnes, J. E. Haugen, and T. Engen-Skaugen (2012), Technical Note: Downscaling RCM precipitation to the station scale using statistical transformations; a comparison of methods, *Hydrol. Earth Syst. Sci.*, **16** (9), 3383-3390.

Guo, H., Q. Hu, and T. Jiang (2008), Annual and seasonal streamflow responses to climate and land-cover changes in the Poyang Lake basin, China, *J. Hydrol.*, **355** (1-4), 106-122.

Gupta, H. V., M. P. Clark, J. A. Vrugt, G. Abramowitz, and M. Ye (2012), Towards a comprehensive assessment of model structural adequacy, *Water Resour. Res.*, **48** (8), W08301, doi: 10.1029/2011wr011044.

Hagemann, S., H. Göttel, D. Jacob, P. Lorenz, and E. Roeckner (2008), Improved regional scale processes reflected in projected hydrological changes over large European catchments, *Climate Dynamics*, **32**, 767-781.

Haltas, I., and M. L. Kavvas (2011), Scale Invariance and Self-Similarity in Hydrologic Processes in Space and Time, *J. Hydrol. Eng.*, **16** (1), 51-63.

Hannaford, J., B. Lloyd-Hughes, C. Keef, S. Parry, and C. Prudhomme (2011), Examining the large-scale spatial coherence of European drought using regional indicators of precipitation and streamflow deficit, *Hydrol. Processes*, **25** (7), 1146-1162.

Harris, C. D. (1946), The Ruhr Coal-Mining District, *Geographical Review*, **36** (2), 194-221.

Hartmann, H. C., T. C. Pagano, S. Sorooshian, and R. Bales (2002), Confidence builders: Evaluating seasonal climate forecasts from user perspectives, *Bulletin of the American Meteorological Society*, **83** (5), 683-698.

Hattermann, F., V. Krysanova, F. Wechsung, and M. Wattenbach (2004), Integrating groundwater dynamics in regional hydrological modelling, *Environ Modell Softw*, **19** (11), 1039-1051.

Horton, P., B. Schaefli, A. Mezghani, B. Hingray, and A. Musy (2006), Assessment of climate-change impacts on alpine discharge regimes with climate model uncertainty, *Hydrol. Processes*, **20**, 2091-2109.

Huang, S. C., V. Krysanova, and F. F. Hattermann (2013), Projection of low flow conditions in Germany under climate change by combining three RCMs and a regional hydrological model, *Acta Geophysica*, **61** (1), 151-193.

Hüffmeyer, N., J. Klasmeier, and M. Matthies (2009), Geo-referenced modeling of zinc concentrations in the Ruhr river basin (Germany) using the model GREAT-ER, *Sci. Total Environ.*, **407** (7), 2296-2305.

Huisman, P., J. De Jong, and K. Wieriks (2000), Transboundary cooperation in shared river basins: experiences from the Rhine, Meuse and North Sea, *Water Policy*, **2** (1), 83-97.

Hurkmans, R., H. De Moel, J. Aerts, and P. A. Troch (2008), Water balance versus land surface model in the simulation of Rhine river discharges, *Water Resour. Res.*, **44** (1), W01418, doi: 10.1029/2007wr006168.

Hurkmans, R., W. Terink, R. Uijlenhoet, P. Torfs, D. Jacob, and P. A. Troch (2010), Changes in Streamflow Dynamics in the Rhine Basin under Three High-Resolution Regional Climate Scenarios, *J. Clim.*, **23** (3), 679-699.

Huss, M. (2011), Present and future contribution of glacier storage change to runoff from macroscale drainage basins in Europe, *Water Resour. Res.*, **47** (7), W07511, doi: 10.1029/2010WR010299.

Jacob, D. (2006), REMO climate of the 20th century run and A1B scenario run, UBA project, 0.088 degree resolution, 1h data, World Data Center for Climate. CERA-DB.

Janssen, P. H. M., A. C. Petersen, J. P. van der Sluijs, J. S. Risbey, J. R. Ravetz, and K. von Krauss (2005), A guidance for assessing and communicating uncertainties, *Water Sci. Technol.*, **52** (6), 125-134.

Jasper, K., P. Calanca, D. Gyalistras, and J. Fuhrer (2004), Differential impacts of climate change on the hydrology of two alpine river basins, *Clim. Res.*, **26** (2), 113-129.

Jaun, S., and B. Ahrens (2009), Evaluation of a probabilistic hydrometeorological forecast system, *Hydrol. Earth Syst. Sci.*, **13** (7), 1031-1043.

Jin, X., C.-Y. Xu, Q. Zhang, and V. P. Singh (2010), Parameter and modeling uncertainty simulated by GLUE and a formal Bayesian method for a conceptual hydrological model, *J. Hydrol.*, **383** (3-4), 147-155.

Jonkeren, O., P. Rietveld, J. Ommeren, and A. Te Linde (2013), Climate change and economic consequences for inland waterway transport in Europe, *Regional Environmental Change*, 1-13, doi: 10.1007/s10113-013-0441-7.

Jung, I. W., D. H. Bae, and B. J. Lee (2013), Possible change in Korean streamflow seasonality based on multi-model climate projections, *Hydrol. Processes*, **27** (7), 1033-1045.

Kahya, E., and J. A. Dracup (1993), U.S. streamflow patterns in relation to the El Niño/Southern Oscillation, *Water Resour. Res.*, **29** (8), 2491-2503.

Kalra, A., S. Ahmad, and A. Nayak (2013), Increasing streamflow forecast lead time for snowmelt-driven catchment based on large-scale climate patterns, *Adv. Water Resour.*, **53**, 150-162.

Kasiviswanathan, K. S., C. Raj, K. P. Sudheer, and I. Chaubey (2013), Constructing prediction interval for artificial neural network rainfall runoff models based on ensemble simulations, *J. Hydrol.*, **499**, 275-288.

Khaliq, M. N., T. B. M. J. Ouarda, P. Gachon, and L. Sushama (2008), Temporal evolution of low-flow regimes in Canadian rivers, *Water Resour. Res.*, **44** (8), W08436, doi: 10.1029/2007wr006132.

Kirchner, J. W. (2006), Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology, *Water Resour. Res.*, **42** (3), W03S04, doi: 10.1029/2005WR004362.

Klemeš, V. (1983), Conceptualization and scale in hydrology, *J. Hydrol.*, **65** (1-3), 1-23.

Komma, J., C. Reszler, G. Blöschl, and T. Haiden (2007), Ensemble prediction of floods catchment non-linearity and forecast probabilities, *Nat. Hazards Earth Syst. Sci.*, **7** (4), 431-444.

Köplin, N., B. Schädler, D. Viviroli, and R. Weingartner (2013), Seasonality and magnitude of floods in Switzerland under future climate change, *Hydrol. Processes*, (accepted), doi: 10.1002/hyp.9757.

Kuo, C.-C., T. Y. Gan, and P.-S. Yu (2010), Seasonal streamflow prediction by a combined climate-hydrologic system for river basins of Taiwan, *J. Hydrol.*, **387** (3-4), 292-303.

Laaha, G., and G. Blöschl (2006), Seasonality indices for regionalizing low flows, *Hydrol. Processes*, **20** (18), 3851-3878.

Lawal, S., W. Watt, and D. Watts (1997), A stochastic model of low flows, *Stoch. Hydrol. Hydraul.*, **11** (4), 303-321.

Lenderink, G., A. Buishand, and W. van Deursen (2007), Estimates of future discharges of the river Rhine using two scenario methodologies: direct versus delta approach, *Hydrol. Earth Syst. Sci.*, **11** (3), 1145.

Li, H., L. Luo, and E. F. Wood (2008), Seasonal hydrologic predictions of low-flow conditions over eastern USA during the 2007 drought, *Atmospheric Science Letters*, **9** (2), 61-66.

Li, H., L. Luo, E. F. Wood, and J. Schaake (2009), The role of initial conditions and forcing uncertainties in seasonal hydrologic forecasting, *J. Geophys. Res.*, **114** (D4), D04114, doi: 10.1029/2008jd010969.

Li , L., J. Xia, C.-Y. Xu, and V. P. Singh (2010), Evaluation of the subjective factors of the GLUE method and comparison with the formal Bayesian method in uncertainty assessment of hydrological models, *J. Hydrol.*, **390** (3-4), 210-221.

Lindström, G., B. Johansson, M. Persson, M. Gardelin, and S. Bergstrom (1997), Development and test of the distributed HBV-96 hydrological model, *J. Hydrol.*, **201** (1-4), 272-288.

Liu, Y., et al. (2012), Advancing data assimilation in operational hydrologic forecasting: progresses, challenges, and emerging opportunities, *Hydrol. Earth Syst. Sci.*, **16** (10), 3863-3887.

Luo, L., E. F. Wood, and M. Pan (2007), Bayesian merging of multiple climate model forecasts for seasonal hydrological predictions, *J. Geophys. Res.*, **112** (D10), D10102, doi: 10.1029/2006jd007655.

Mantovan, P., and E. Todini (2006), Hydrological forecasting uncertainty assessment: Incoherence of the GLUE methodology, *J. Hydrol.*, **330** (1-2), 368-381.

Martina, M. L. V., E. Todini, and A. Libralon (2006), A Bayesian decision approach to rainfall thresholds based flood warning, *Hydrol. Earth Syst. Sci.*, **10** (3), 413-426.

McKay, M. D., R. J. Beckman, and W. J. Conover (1979), A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code, *Technometrics*, **21** (2), 239-245.

Meißner, D., S. Gebauer, A. H. Schumann, M. Pahlow, and S. Rademacher (2012), Analysis of radar-based precipitation data products as input data to improve water-level forecasting for navigation on the River Rhine, *Hydrologie und Wasserbewirtschaftung*, **56** (1), 16-28.

Middelkoop, H., and C. O. G. Van Haselen (1999), Twice a river. Rhine and Meuse in The Netherlands. RIZA report 99.003 Arnhem: RIZA *Report*, 127 pp.

Middelkoop, H., K. Daamen, D. Gellens, W. Grabs, J. C. J. Kwadijk, H. Lang, B. Parmet, B. Schadler, J. Schulla, and K. Wilke (2001), Impact of climate change on hydrological regimes and water resources management in the Rhine Basin, *Climatic Change*, **49** (1-2), 105-128.

Minville, M., F. Brissette, and R. Leconte (2008), Uncertainty of the impact of climate change on the hydrology of a nordic watershed, *J. Hydrol.*, **358** (1-2), 70-83.

Montanari, A. (2005), Large sample behaviors of the generalized likelihood uncertainty estimation (GLUE) in assessing the uncertainty of rainfall-runoff simulations, *Water Resour. Res.*, **41** (8), W08406, doi: 10.1029/2004wr003826.

Montanari, A. (2011), Uncertainty of Hydrological Predictions, in *Treatise on Water Science*, edited by P. Wilderer, pp. 459-478, Academic Press, Oxford.

Montanari, A., and G. Grossi (2008), Estimating the uncertainty of hydrological forecasts: A statistical approach, *Water Resour. Res.*, **44**, W00B08, doi: 10.1029/2008wr006897

Moradkhani, H., C. M. DeChant, and S. Sorooshian (2012), Evolution of ensemble data assimilation for uncertainty quantification using the particle filter-Markov chain Monte Carlo method, *Water Resour. Res.*, **48** (12), W12520, doi: 10.1029/2012wr012144.

Muerth, M. J., B. Gauvin St-Denis, S. Ricard, J. A. Velázquez, J. Schmid, M. Minville, D. Caya, D. Chaumont, R. Ludwig, and R. Turcotte (2013), On the need for bias correction in regional climate scenarios to assess climate change impacts on river runoff, *Hydrol. Earth Syst. Sci.*, **17** (3), 1189-1204.

Muluye, G. Y. (2011), Implications of medium-range numerical weather model output in hydrologic applications: Assessment of skill and economic value, *J. Hydrol.*, **400** (3-4), 448-464.

Nakićenović, N., and R. Swart (2000), Special Report on Emissions Scenarios. Cambridge University Press, 612 pp., Cambridge, UK.

Nash, J. E., and J. V. Sutcliffe (1970), River flow forecasting through conceptual models part I – A discussion of principles, *J. Hydrol.*, **10** (3), 282-290.

Nester, T., J. Komma, A. Viglione, and G. Blöschl (2012), Flood forecast errors and ensemble spread - a case study, *Water Resour. Res.*, **48** (10), W10502, doi: 10.1029/2011wr011649.

Nilson, E., I. Lingemann, B. Klein, and P. Krahe (2012), Deliverable 1.4: Impact of Hydrological Change on Navigation Conditions, in *ECCONET - Effects of climate change on the inland waterway transport network*. Available at: http://ecconet.eu/deliverables/ECCONET_D1.4_final.pdf, [Accessed: 24 September 2013].

Olsson, J., and G. Lindström (2008), Evaluation and calibration of operational hydrological ensemble forecasts in Sweden, *J. Hydrol.*, **350** (1-2), 14-24.

Ouarda, T. B. M. J., C. Charron, and A. St-Hilaire (2008), Statistical models and the estimation of low flows, *Can. Water Resour. J.*, **33** (2), 195-206.

Pagano, T. C., D. L. Shrestha, Q. J. Wang, D. Robertson, and P. Hapuarachchi (2013), Ensemble dressing for hydrological applications, *Hydrol. Processes*, **27** (1), 106-116.

Pappenberger, F., J. Thielen, and M. Del Medico (2011), The impact of weather forecast improvements on large scale hydrology: analysing a decade of forecasts of the European Flood Alert System, *Hydrol. Processes*, **25** (7), 1091-1113.

Pappenberger, F., K. J. Beven, A. De Roo, J. Thielen, and B. Gouweleeuw (2004), Uncertainty analysis of the rainfall runoff model LisFlood within the Generalized Likelihood Uncertainty Estimation (GLUE), *International Journal of River Basin Management*, **2** (2), 123-133.

Pappenberger, F., K. J. Beven, N. M. Hunter, P. D. Bates, B. T. Gouweleeuw, J. Thielen, and A. P. J. de Roo (2005), Cascading model uncertainty from medium range weather forecasts (10 days) through a rainfall-runoff model to flood inundation predictions within the European Flood Forecasting System (EFFS), *Hydrol. Earth Syst. Sci.*, **9** (4), 381-393.

Parajka, J., S. Kohnová, R. Merz, J. Szolgay, K. Hlavčová, and G. Blöschl (2009), Comparative analysis of the seasonality of hydrological characteristics in Slovakia and Austria, *Hydrol. Sci. J.*, **54** (3), 456 - 473.

Parajka, J., et al. (2010), Seasonal characteristics of flood regimes across the Alpine-Carpathian range, *J. Hydrol.*, **394** (1-2), 78-89.

Parrish, M. A., H. Moradkhani, and C. M. DeChant (2012), Toward reduction of model uncertainty: Integration of Bayesian model averaging and data assimilation, *Water Resour. Res.*, **48** (3), W03519, doi: 10.1029/2011WR011116.

Pasetto, D., M. Camporese, and M. Putti (2012), Ensemble Kalman filter versus particle filter for a physically-based coupled surface-subsurface model, *Adv. Water Resour.*, **47**, 1-13.

Perrin, C., C. Michel, and V. Andréassian (2002), Long-term low flow forecasting for French rivers by continuous rainfall-runoff modelling, in *Meeting of the British Hydrological Society on Continuous River Flow Simulation. BHS Occasional Paper # 13*, pp. 21-29, Wallingford, UK, 5^{th} July 2001.

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

Perrin, C., L. Oudin, V. Andréassian, C. Rojas-Serna, C. Michel, and T. Mathevet (2007), Impact of limited streamflow data on the efficiency and the parameters of rainfall-runoff models, *Hydrol. Sci. J.*, **52** (1), 131-151.

Pokhrel, P., Q. J. Wang, and D. E. Robertson (2013), The value of model averaging and dynamical climate model predictions for improving statistical seasonal streamflow forecasts over Australia, *Water Resour. Res.*, (accepted), doi: 10.1002/wrcr.20449.

Prudhomme, C., and H. Davies (2009), Assessing uncertainties in climate change impact analyses on the river flow regimes in the UK. Part 2: future climate, *Climatic Change*, **93** (1-2), 197-222.

Pushpalatha, R., C. Perrin, N. L. Moine, and V. Andréassian (2012), A review of efficiency criteria suitable for evaluating low-flow simulations, *J. Hydrol.*, **420-421**, 171-182.

Pushpalatha, R., C. Perrin, N. L. Moine, T. Mathevet, and V. Andréassian (2011), A downward structural sensitivity analysis of hydrological models to improve low-flow simulation, *J. Hydrol.*, **411** (1-2), 66-76.

Randrianasolo, A., M. H. Ramos, G. Thirel, V. Andréassian, and E. Martin (2010), Comparing the scores of hydrological ensemble forecasts issued by two different hydrological models, *Atmospheric Science Letters*, **11** (2), 100-107.

Ratto, M., P. C. Young, R. Romanowicz, F. Pappenberger, A. Saltelli, and A. Pagano (2007), Uncertainty, sensitivity analysis and the role of data based mechanistic modeling in hydrology, *Hydrol. Earth Syst. Sci.*, **11** (4), 1249-1266.

Refsgaard, J. C., J. P. van der Sluijs, J. Brown, and P. van der Keur (2006), A framework for dealing with uncertainty due to model structure error, *Adv. Water Resour.*, **29** (11), 1586-1597.

Refsgaard, J. C., J. P. van der Sluijs, A. L. Højberg, and P. A. Vanrolleghem (2007), Uncertainty in the environmental modelling process-A framework and guidance, *Environ Modell Softw*, **22** (11), 1543-1556.

Reggiani, P., and A. H. Weerts (2008a), Probabilistic Quantitative Precipitation Forecast for Flood Prediction: An Application, *J. Hydrometeorol.*, **9**, 76-95.

Reggiani, P., and A. H. Weerts (2008b), A Bayesian approach to decision-making under uncertainty: An application to real-time forecasting in the river Rhine, *J. Hydrol.*, **356** (1-2), 56-69.

Reggiani, P., M. Renner, A. H. Weerts, and P. van Gelder (2009), Uncertainty assessment via Bayesian revision of ensemble streamflow predictions in the operational river Rhine forecasting system, *Water Resour. Res.*, **45** (2), W02428, doi: 10.1029/2007WR006758.

Renard, B., D. Kavetski, G. Kuczera, M. Thyer, and S. W. Franks (2010), Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors, *Water Resour. Res.*, **46** (5), W05521, doi: 10.1029/2009wr008328.

Renner, M., M. G. F. Werner, S. Rademacher, and E. Sprokkereef (2009), Verification of ensemble flow forecasts for the River Rhine, *J. Hydrol.*, **376** (3-4), 463-475.

Robertson, D. E., P. Pokhrel, and Q. J. Wang (2013), Improving statistical forecasts of seasonal streamflows using hydrological model output, *Hydrol. Earth Syst. Sci.*, **17** (2), 579-593.

Rossa, A., K. Liechti, M. Zappa, M. Bruen, U. Germann, G. Haase, C. Keil, and P. Krahe (2011), The COST 731 Action: A review on uncertainty propagation in advanced hydro-meteorological forecast systems, *Atmospheric Research*, **100** (2-3), 150-167.

Roulin, E. (2007), Skill and relative economic value of medium-range hydrological ensemble predictions, *Hydrol. Earth Syst. Sci.*, **11** (2), 725-737.

Roulin, E., and S. Vannitsem (2005), Skill of Medium-Range Hydrological Ensemble Predictions, *J. Hydrometeorol.*, **6** (5), 729-744.

Rutten, M., N. van de Giesen, M. Baptist, J. Icke, and W. Uijttewaal (2008), Seasonal forecast of cooling water problems in the River Rhine, *Hydrol. Processes*, **22** (7), 1037-1045.

Saadat, S., D. Khalili, A. Kamgar-Haghighi, and S. Zand-Parsa (2013), Investigation of spatio-temporal patterns of seasonal streamflow droughts in a semi-arid region, *Natural Hazards*, 1-24.

Salamon, P., and L. Feyen (2009), Assessing parameter, precipitation, and predictive uncertainty in a distributed hydrological model using sequential data assimilation with the particle filter, *J. Hydrol.*, **376** (3-4), 428-442.

Sauquet, E., J. Lerat, and C. Prudhomme (2008), La prévision hydrométéorologique á 3-6 mois. Etat des connaissances et applications, *La Houille Blanche* (6), 77-84.

Schaefli, B., B. Hingray, and A. Musy (2007), Climate change and hydropower production in the Swiss Alps: quantification of potential impacts and related modelling uncertainties, *Hydrol. Earth Syst. Sci.*, **11**, 1191-1205.

Scherrer, S. C., and C. Appenzeller (2006), Swiss Alpine snow pack variability: major patterns and links to local climate and large-scale flow, *Clim. Res.*, **32** (3), 187-199.

Schneider, J. (2008), Impacts of climate change on catchment storage, stream flow recession and summer low flow, 154 pp, Institut für Hydrologie Albert-Ludwigs-Universität Freiburg, Freiburg im Breisgau.

Schnorbus, M., A. Werner, and K. Bennett (2012), Impacts of climate change in three hydrologic regimes in British Columbia, Canada, *Hydrol. Processes*, (accepted), doi: 10.1002/hyp.9661.

Schubert, S., R. Koster, M. Hoerling, R. Seager, D. Lettenmaier, A. Kumar, and D. Gutzler (2007), Predicting Drought on Seasonal-to-Decadal Time Scales, *Bulletin of the American Meteorological Society*, **88** (10), 1625-1630.

Shabalova, M., W. van Deursen, and T. Buishand (2003), Assessing future discharge of the river Rhine using regional climate model integrations and a hydrological model, *Clim. Res.*, **23** (3), 233-246.

Shamseldin, A. Y. (1997), Application of a neural network technique to rainfall-runoff modelling, *J. Hydrol.*, **199** (3-4), 272-294.

Shen, Z. Y., L. Chen, and T. Chen (2012), Analysis of parameter uncertainty in hydrological and sediment modeling using GLUE method: a case study of SWAT model applied to Three Gorges Reservoir Region, China, *Hydrol. Earth Syst. Sci.*, **16** (1), 121-132.

Shukla, S., and D. P. Lettenmaier (2011), Seasonal hydrologic prediction in the United States: understanding the role of initial hydrologic conditions and seasonal climate forecast skill, *Hydrol. Earth Syst. Sci.*, **15** (11), 3529-3538.

Shukla, S., N. Voisin, and D. P. Lettenmaier (2012), Value of medium range weather forecasts in the improvement of seasonal hydrologic prediction skill, *Hydrol. Earth Syst. Sci.*, **16** (8), 2825-2838.

Shukla, S., J. Sheffield, E. F. Wood, and D. P. Lettenmaier (2013), On the sources of global land surface hydrologic predictability, *Hydrol. Earth Syst. Sci.*, **17** (7), 2781-2796.

Smakhtin, V. U. (2001), Low flow hydrology: a review, *J. Hydrol.*, **240** (3-4), 147-186.

Şorman, A. A., A. Şensoy, A. E. Tekeli, A. Ü. Şorman, and Z. Akyürek (2009), Modelling and forecasting snowmelt runoff process using the HBV model in the eastern part of Turkey, *Hydrol. Processes*, **23** (7), 1031-1040.

Soukup, T. L., O. A. Aziz, G. A. Tootle, T. C. Piechota, and S. S. Wulff (2009), Long lead-time streamflow forecasting of the North Platte River incorporating oceanic-atmospheric climate variability, *J. Hydrol.*, **368** (1-4), 131-142.

Spear, R. C., and G. M. Hornberger (1980), Eutrophication in peel inlet-II. Identification of critical uncertainties via generalized sensitivity analysis, *Water Research*, **14** (1), 43-49.

Spence, C., P. H. Whitfield, and T. B. M. J. Quarda (2008), Introduction to the special issue on low-flow prediction in ungauged basins (PUB) in Canada, *Can. Water Resour. J.*, **33** (2), 103-106.

Stedinger, J. R., R. M. Vogel, S. U. Lee, and R. Batchelder (2008), Appraisal of the generalized likelihood uncertainty estimation (GLUE) method, *Water Resour. Res.*, **44**, W00B06, doi: 10.1029/2008wr006822.

Steinschneider, S., and C. Brown (2011), Influences of North Atlantic climate variability on low-flows in the Connecticut River Basin, *J. Hydrol.*, **409** (1-2), 212-224.

Stravs, L., and M. Brilly (2007), Development of a low-flow forecasting model using the M5 machine learning method, *Hydrol. Sci. J.*, **52** (3), 466-477.

Suweis, S., E. Bertuzzo, G. Botter, A. Porporato, I. Rodriguez-Iturbe, and A. Rinaldo (2010), Impact of stochastic fluctuations in storage-discharge relations on streamflow distributions, *Water Resour. Res.*, **46** (3), W03517, doi: 10.1029/2009wr008038.

Svensson, C., and C. Prudhomme (2005), Prediction of British summer river flows using winter predictors, *Theoretical and Applied Climatology*, **82** (1), 1-15.

Tague, C., and G. E. Grant (2009), Groundwater dynamics mediate low-flow response to global warming in snow-dominated alpine regions, *Water Resour. Res.*, **45**, W07421, doi: 10.1029/2008WR007179.

Tallaksen, L., and H. A. J. Van Lanen (2004), *Hydrological Drought: Processes and Estimation Methods for Streamflow and Groundwater*, 579 pp., Elsevier, Amsterdam, The Netherlands.

Tallaksen, L. M., H. Hisdal, and H. A. J. Van Lanen (2009), Space-time modelling of catchment scale drought characteristics, *J. Hydrol.*, **375** (3-4), 363-372.

Taylor, I. H., E. Burke, L. McColl, P. Falloon, G. R. Harris, and D. McNeall (2013), The impact of climate mitigation on projections of future drought, *Hydrol. Earth Syst. Sci.*, **17** (6), 2339-2358.

Te Linde, A. H., J. Aerts, R. Hurkmans, and M. Eberle (2008), Comparing model performance of two rainfall-runoff models in the Rhine basin using different atmospheric forcing data sets, *Hydrol. Earth Syst. Sci.*, **12** (3), 943-957.

Te Linde, A. H., J. C. J. H. Aerts, A. M. R. Bakker, and J. C. J. Kwadijk (2010), Simulating low-probability peak discharges for the Rhine basin using resampled climate modeling data, *Water Resour. Res.*, **46** (3), W03512, doi: 10.1029/2009wr007707.

Te Linde, A. H., P. Bubeck, J. E. C. Dekkers, H. de Moel, and J. C. J. H. Aerts (2011), Future flood risk estimates along the river Rhine, *Nat. Hazards Earth Syst. Sci.*, **11** (2), 459-473.

Teutschbein, C., and J. Seibert (2012), Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods, *J. Hydrol.*, **456-457**, 12-29.

Thirel, G., F. Rousset-Regimbeau, E. Martin, and F. Habets (2008), On the Impact of Short-Range Meteorological Forecasts for Ensemble Streamflow Predictions, *J. Hydrometeorol.*, **9** (6), 1301-1317.

Thirel, G., F. Regimbeau, E. Martin, J. Noilhan, and F. Habets (2010), Short- and medium-range hydrological ensemble forecasts over France, *Atmospheric Science Letters*, **11**, 72-77.

Thyer, M., B. Renard, D. Kavetski, G. Kuczera, S. W. Franks, and S. Srikanthan (2009), Critical evaluation of parameter consistency and predictive uncertainty in hydrological modeling: A case study using Bayesian total error analysis, *Water Resour. Res.*, **45** (12), W00B14, doi: 10.1029/2008wr006825.

Tian, Y., M. Booij, and Y.-P. Xu (2013), Uncertainty in high and low flows due to model structure and parameter errors, *Stoch. Environ. Res. Risk Assess.*, 1-14, doi: 10.1007/s00477-013-0751-9.

Tolson, B. A., and C. A. Shoemaker (2008), Efficient prediction uncertainty approximation in the calibration of environmental simulation models, *Water Resour. Res.*, **44** (4), W04411, doi: 10.1029/2007wr005869.

Tongal, H., M. C. Demirel, and M. J. Booij (2013), Seasonality of low flows and dominant processes in the Rhine River, *Stoch. Environ. Res. Risk Assess.*, **27** (2), 489-503.

Tootle, G. A., and T. C. Piechota (2004), Suwannee River Long Range Streamflow Forecasts Based On Seasonal Climate Predictors, *JAWRA Journal of the American Water Resources Association*, **40** (2), 523-532.

Towler, E., M. Roberts, B. Rajagopalan, and R. S. Sojda (2013), Incorporating probabilistic seasonal climate forecasts into river management using a risk-based framework, *Water Resour. Res.*, **49** (8), 4997-5008.

Van Andel, S. J., A. Weerts, J. Schaake, and K. Bogner (2013), Post-processing hydrological ensemble predictions intercomparison experiment, *Hydrol. Processes*, **27** (1), 158-161.

Van den Tillaart, S. P. M., M. J. Booij, and M. S. Krol (2013), Impact of uncertainties in discharge determination on the parameter estimation and performance of a hydrological model, *Hydrology Research*, **44** (3), 454-466

Van der Sluijs, J., M. Craye, S. Funtowicz, P. Kloprogge, J. Ravetz, and J. Risbey (2005), Experiences with the NUSAP system for multidimensional uncertainty assessment, *Water science and technology: a journal of the International Association on Water Pollution Research*, **52** (6), 133.

Van Dijk, A. I. J. M., J. L. Peña-Arancibia, E. F. Wood, J. Sheffield, and H. E. Beck (2013), Global analysis of seasonal streamflow predictability using an ensemble prediction system and observations from 6192 small catchments worldwide, *Water Resour. Res.*, **49** (5), 2729-2746.

Van Lanen, H. A. J., N. Wanders, L. M. Tallaksen, and A. F. Van Loon (2013), Hydrological drought across the world: impact of climate and physical catchment structure, *Hydrol. Earth Syst. Sci.*, **17** (5), 1715-1732.

Van Ogtrop, F. F., R. W. Vervoort, G. Z. Heller, D. M. Stasinopoulos, and R. A. Rigby (2011), Long-range forecasting of intermittent streamflow, *Hydrol. Earth Syst. Sci.*, **15** (11), 3343-3354.

Vázquez, R. F., K. J. Beven, and J. Feyen (2009), GLUE Based Assessment on the Overall Predictions of a MIKE SHE Application, *Water Resour Manage*, **23**, 1325-1349.

Velázquez, J. A., F. Anctil, and C. Perrin (2010), Performance and reliability of multimodel hydrological ensemble simulations based on seventeen lumped models and a thousand catchments, *Hydrol. Earth Syst. Sci.*, **14** (11), 2303-2317.

Velázquez, J. A., F. Anctil, M. H. Ramos, and C. Perrin (2011), Can a multimodel approach improve hydrological ensemble forecasting? A study on 29 French catchments using 16 hydrological model structures, *Adv. Geosci.*, **29**, 33-42.

Verbunt, M., J. Gurtz, K. Jasper, H. Lang, P. Warmerdam, and M. Zappa (2003), The hydrological role of snow and glaciers in alpine river basins and their distributed modeling, *J. Hydrol.*, **282** (1-4), 36-55.

Vicente-Serrano, S. M., and J. I. López-Moreno (2005), Hydrological response to different time scales of climatological drought: an evaluation of the Standardized Precipitation Index in a mountainous Mediterranean basin, *Hydrol. Earth Syst. Sci.*, **9** (5), 523-533.

Vidal, J. P., E. Martin, L. Franchistéguy, F. Habets, J. M. Soubeyroux, M. Blanchard, and M. Baillon (2010), Multilevel and multiscale drought reanalysis over France with the Safran-Isba-Modcou hydrometeorological suite, *Hydrol. Earth Syst. Sci.*, **14** (3), 459-478.

Viola, F., L. V. Noto, M. Cannarozzo, and G. La Loggia (2009), Daily streamflow prediction with uncertainty in ephemeral catchments using the GLUE methodology, *Physics and Chemistry of the Earth, Parts A/B/C*, **34** (10-12), 701-706.

Vrugt, J. A., C. J. F. t. Braak, M. P. C. , J. M. Hyman, and B. A. Robinson (2008), Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with Markov chain Monte Carlo simulation, *Water Resour. Res.*, **44** (W00B09), doi: 10.1029/2007WR006720.

Walker, W. E., P. Harremos, J. Rotmans, J. P. van der Sluijs, M. B. A. van Asselt, P. Janssen, and M. P. K. von Krauss (2003), Defining Uncertainty: A Conceptual Basis for Uncertainty Management in Model-Based Decision Support, *Integrated Assessment*, **4** (1), 5-17.

Wang, E., Y. Zhang, J. Luo, F. H. S. Chiew, and Q. J. Wang (2011), Monthly and seasonal streamflow forecasts using rainfall-runoff modeling and historical weather data, *Water Resour. Res.*, **47** (5), W05516, doi: 10.1029/2010wr009922.

Wang, W., P. H. A. J. M. V. Gelder, J. K. Vrijling, and J. Ma (2006), Forecasting daily streamflow using hybrid ANN models, *J. Hydrol.*, **324** (1-4), 383-399.

Warmink, J. J., J. A. E. B. Janssen, M. J. Booij, and M. S. Krol (2010), Identification and classification of uncertainties in the application of environmental models, *Environ Modell Softw*, **25** (12), 1518-1527.

Warmink, J. J., H. Klis, M. J. Booij, and S. J. M. H. Hulscher (2011), Identification and Quantification of Uncertainties in a Hydrodynamic River Model Using Expert Opinions, *Water Resour. Manag.*, **25** (2), 601-622.

Wedgbrow, C. S., R. L. Wilby, and H. R. Fox (2005), Experimental seasonal forecasts of low summer flows in the River Thames, UK, using Expert Systems, *Clim. Res.*, **28** (2), 133-141.

Wedgbrow, C. S., R. L. Wilby, H. R. Fox, and G. O'Hare (2002), Prospects for seasonal forecasting of summer drought and low river flow anomalies in England and Wales, *Int. J. Climatol.*, **22** (2), 219-236.

Wilby, R. L., and I. Harris (2006), A framework for assessing uncertainties in climate change impacts: Low-flow scenarios for the River Thames, UK, *Water Resour. Res.*, **42** (2), W02419, doi: 10.1029/2005wr004065.

Wilks, D. S. (1995), *Statistical Methods in the Atmospheric Sciences*, Elsevier, New York.

WMO (2012), Forecastverification - issues, methods and faq. WWRP/WGNE Joint Working Group on Verification. Available at: www.cawcr.gov.au/projects/verification/, [Accessed: 24 September 2013].

Wöhling, T., F. Lennartz, and M. Zappa (2006), Technical Note: Updating procedure for flood forecasting with conceptual HBV-type models, *Hydrol. Earth Syst. Sci.*, **10** (6), 783-788.

Wood, A. W., and D. P. Lettenmaier (2006), A Test Bed for New Seasonal Hydrologic Forecasting Approaches in the Western United States, *Bulletin of the American Meteorological Society*, **87** (12), 1699-1712.

Wood, A. W., and D. P. Lettenmaier (2008), An ensemble approach for attribution of hydrologic prediction uncertainty, *Geophys. Res. Lett.*, **35** (14), L14401, doi: 10.1029/2008gl034648.

Wood, A. W., E. P. Maurer, A. Kumar, and D. P. Lettenmaier (2002), Long-range experimental hydrologic forecasting for the eastern United States, *J. Geophys. Res*, **107** (D20), 4429, doi: 10.1029/2001JD000659.

Yimer, E. G., and J. Andreja (2013), Downscaling technique uncertainty in assessing hydrological impact of climate change in the Upper Beles River Basin, Ethiopia, *Hydrology Research*, **44** (2), 377-398.

Yossef, N. C., L. P. H. van Beek, J. C. J. Kwadijk, and M. F. P. Bierkens (2012), Assessment of the potential forecasting skill of a global hydrological model in reproducing the occurrence of monthly flow extremes, *Hydrol. Earth Syst. Sci.*, **16** (11), 4233-4246.

Yossef, N. C., H. Winsemius, A. Weerts, R. van Beek, and M. F. P. Bierkens (2013), Skill of a global seasonal streamflow forecasting system, relative roles of initial conditions and meteorological forcing, *Water Resour. Res.*, **49** (8), 4687-4699.

Yue, S., and C. Y. Wang (2004), Scaling of Canadian low flows, *Stoch. Environ. Res. Risk Assess.*, **18** (5), 291-305.

Zaidman, M. D., H. G. Rees, and A. R. Young (2001), Spatio-temporal development of streamflow droughts in north-west Europe, *Hydrol. Earth Syst. Sci.*, **6** (4), 733-751.

Zappa, M., and C. Kan (2007), Extreme heat and runoff extremes in the Swiss Alps, *Natural Hazards and Earth System Sciences*, **7** (3), 375-389.

Zappa, M., S. Jaun, U. Germann, A. Walser, and F. Fundel (2011), Superposition of three sources of uncertainties in operational flood forecasting chains, *Atmospheric Research*, **100** (2-3), 246-262.

Zheng, Y., and A. A. Keller (2007), Uncertainty assessment in watershed-scale water quality modeling and management: 1. Framework and application of generalized likelihood uncertainty estimation (GLUE) approach, *Water Resour. Res.*, **43** (8), W08407, doi: 10.1029/2006wr005345.

List of symbols

Roman

a	transformation coefficient for precipitation	[1]
A_i	area of each of the 134 catchments	[L^2]
A_j	total area of each of the seven major sub-basins	[L^2]
b	bias	[-]
b	transformation coefficient for precipitation	[-]
$B1, B2$	bias value in hidden layer	[-]
BS	Brier Score	[-]
BSS	Brier Skill Score	[-]
$CFLUX$	maximum capillary flow from upper response box to soil moisture zone	[LT^{-1}]
D_i	day number of each low flow event	[T]
F_j	forecast probability for the j-th observed low flow day	[1]
F_t	forecast probability	[1]
FC	maximum soil moisture capacity	[L]
G	groundwater index	[L]
g_i	standardised daily groundwater level	[L]
GW	groundwater storage in HBV model	[L]
k	number of Monte Carlo run	[T]
k_{GR4J}	fraction of fast and slow runoff for GR4J model	[1]
k_{HBV}	fraction of fast and slow runoff for HBV model	[1]
KF	recession coefficient for quick flow reservoir	[T^{-1}]
KS	recession coefficient for base flow reservoir	[T^{-1}]
L	lake index	[L]
l_i	standardised daily lake level series observed in the WA sub-basin	[L]
LP	soil moisture threshold for reduction of evapotranspiration	[1]
MAE	Mean Absolute Error	[1]
MFS	Mean Forecast Score	[-]
N	number of observation	[-]

O_t	observed probability	[1]
P	precipitation index	[-]
P_{cor}	bias-corrected precipitation	[L]
P_{RCM}	precipitation from RCMs	[L]
$PERC$	maximum flow from upper to lower response box	[LT^{-1}]
PET	potential evapotranspiration index	[-]
Q	discharge	[L^3T^{-1}]
Q_{75s}	summer low flow	[LT^{-1}]
Q_{75w}	winter low flow	[LT^{-1}]
Q_d	fast runoff component in GR4J model	[LT^{-1}]
Q_f	fast runoff component in HBV model	[LT^{-1}]
Q_i	runoff	[LT^{-1}]
$Q_{obs}(i)$	observed discharge	[LT^{-1}]
$Q_{obs}(j)$	observed discharge for the j-th observed low flow day	[LT^{-1}]
Q_r	slow runoff component in GR4J model	[LT^{-1}]
Q_s	slow runoff component in HBV model	[LT^{-1}]
$Q_{sim}(i)$	simulated discharge	[LT^{-1}]
$Q_{sim}(j)$	simulated discharge for the j-th observed low flow day	[LT^{-1}]
$Q5$	5% discharge percentile	[LT^{-1}]
$Q50$	forecast median	[LT^{-1}]
$Q75$	75% discharge percentile as low flow threshold	[LT^{-1}]
$Q95$	95% discharge percentile	[LT^{-1}]
R^2	efficiency criterion based on Nash and Sutcliffe coefficient	[-]
$R_{\log}$	similar to R^2 but using the logarithmic discharge values giving more weight to low flows	[-]
RCI	Relative Confidence Interval	[L^3T^{-1}]
$relaccdiff$	accumulated relative difference between the simulated and observed discharge	[L^3T^{-1}]
S	snow index	[-]
s_i	the standardised daily fresh snow height series observed in these sub-basins	[-]
SM	soil moisture in HBV model	[L]
SM_{opt}	calibrated soil moisture in HBV model	[L]
SR	seasonality ratio	[1]
SW	surface water in HBV model	[L]

T	temperature	$[\theta]$
T_{cor}	bias-corrected temperature	$[\theta]$
T_m	long term mean of the RCM temperature	$[\theta]$
T_o	long term mean of the observed temperature	$[\theta]$
T_{RCM}	RCM temperature	$[\theta]$
w_i	weight of connection between input nodes and hidden neuron	[-]
$WMOD$	weighted mean occurrence day	[T]
WP	weighted persistence	[-]
$W1 \dots W3$	weight of connection between input nodes and hidden neuron	[-]
X	the mean of the original time series	[-]
x_θ	weighted mean of Cartesian coordinates	[-]
$x_1 \dots x_n$	input vectors to neural networks model	[-]
$X1$	capacity of the production store in GR4J model	[L]
$X2$	groundwater exchange coefficient in GR4J model	[L]
$X3$	one day ahead capacity of the routing store in GR4J model	[L]
$X4$	time base of the unit hydrograph in GR4J model	[T]
y_θ	weighted mean of Cartesian coordinates	[-]
Z	standardised time series	[-]

Greek

α	measure for non-linearity of low flow in quick runoff reservoir in HBV model	[1]
β	shape coefficient	[1]
μ_x	average	[-]
σ_m	standard deviation of the daily RCM temperature	[-]
σ_o	standard deviation of the observed daily temperature	[-]
σ_x	standard deviation of the original time series	[-]
θ	directional angle	$[\theta]$
ε	1% of the mean observed discharge	$[LT^{-1}]$

Acronyms and abbreviations

ANN-E	Artificial Neural Networks - Ensemble
ANN-I	Artificial Neural Networks - Indicator
BAFU	Federal Office for the Environment (Switzerland)
BFG	Federal Institute of Hydrology (Germany)
BSS	Brier Skill Score
CHR	International Rhine Commission
CLS	Dr.ir. Cornelis Lelystichting
COSMO-LEPS	Limited Area Ensemble Prediction System developed within COSMO consortium
DWD	German Weather Service
EA	East Alpine subbasin
ECHAM	ECMWF HAMburg AGCM (Germany)
ECMWF	European Centre for Medium-Range Weather Forecasts
ENS	Ensemble weather forecast
GCM	Global Climate Model
GIS	Geographic Information System
GLUE	Generalized Likelihood Uncertainty Estimation
GR4J	Génie Rural à 4 parametres Journalier
GRDC	Global Runoff Data Centre
HadCM3	Hadley Centre Coupled Model, version 3 (United Kingdom)
HadRM3	Hadley Centre Regional Model, version 3 (United Kingdom)
HBV	Hydrologiska Byråns Vattenbalansavdelning
IPCC	Intergovernmental Panel on Climate Change
KNMI	Royal Netherlands Meteorological Institute
LCW	Dutch National Coordinating Committee on Water Distribution
logsig	logarithmic sigmoid activation function
LR	Lower Rhine subbasin
MR	Middle Rhine subbasin
NS	Nash-Sutcliffe efficiency coefficient
PCR-GLOBWB	PCRaster Global Water Balance
PET	Potential Evapotranspiration
PREVAH	PREecipitation-Runoff-EVApotranspiration Hydrological Response Unit Model
RACMO	KNMI Regional Atmospheric Climate Model
RCM	Regional Climate Model
REMO	The Regional Model (Germany)
RIZA	Dutch Water Service (currently Rijkswaterstaat)
SMHI	Swedish Meteorological and Hydrological Institute
SRES	Emissions Scenarios
WA	West Alpine subbasin
WMO	World Meteorological Organization
WSL	Institute for Snow and Avalanche Research

List of publications

Journal papers

1. **Demirel MC**, Booij MJ, Hoekstra AY. 2013. Impacts of climate change on the seasonality of low flows in 134 catchments in the River Rhine basin using an ensemble of bias-corrected regional climate simulations. *Hydrology and Earth System Sciences* **17** (10): 4241-4257, doi: 10.5194/hess-17-4241-2013. (Impact Factor: 3.6, Ranking: 2/80 *Water Resources* journals listed in ISI Journal Citation Reports © 2013)

2. **Demirel MC**, Booij MJ, Hoekstra AY. 2013. Effect of different uncertainty sources on the skill of 10 day ensemble low flow forecasts for two hydrological models. *Water Resources Research* **49** (7): 4035–4053, doi: 10.1002/wrcr.20294. (Impact Factor: 3.1, Ranking: 3/80)

3. **Demirel MC**, Booij MJ, Hoekstra AY. 2013. Identification of appropriate lags and temporal resolutions for low flow indicators in the River Rhine to forecast low flows with different lead times. *Hydrological Processes* **27** (19): 2742–2758, doi: 10.1002/hyp.9402. (Impact Factor: 2.5, Ranking: 9/80)

4. Tongal H, **Demirel MC**, Booij MJ. 2013. Seasonality of low flows and dominant processes in the Rhine River. *Stochastic Environmental Research and Risk Assessment* **27**: 489–503. doi: 10.1007/s00477-012-0594-9. (Impact Factor: 1.9, Ranking: 20/80)

5. **Demirel MC**, Booij MJ, Kahya E. 2012. Validation of an ANN Flow Prediction Model Using a Multistation Cluster Analysis. *Journal of Hydrologic Engineering* **17**: 262–271. doi: 10.1061/(ASCE)HE.1943-5584.0000426. (Impact Factor: 1.4, Ranking: 36/80)

6. **Demirel MC**, Venancio A, Kahya E. 2009. Flow forecast by SWAT model and ANN in Pracana basin, Portugal. *Advances in Engineering Software* **40**: 467–473. doi: 10.1016/j.advengsoft.2008.08.002. (Impact Factor: 1.2, Ranking: 58/99 *Computer Science, Interdisciplinary Applications* journals listed in ISI Journal Citation Reports © 2013)

Discussion papers

1. **Demirel MC**, Kahya E, Rivera D. 2008. Discussion of " Clustering On Dissimilarity Representations for Detecting Mislabelled Seismic Signals at Nevado Del Ruiz Volcano" by Mauricio Orozco-Alzate, and Csar Germn Castellanos-Domnguez. *Earth Sciences Research Journal* **12**: 265–268

2. **Demirel MC**, Kahya E, Rivera D. 2009. Discussion of " Hydrologic Regionalization of Watersheds in Turkey" by Sabahattin Isik and Vijay P. Singh. *Journal of Hydrologic Engineering* **14**: 767

Book chapters

1. **Demirel MC**, Booij MJ. 2009. Identification of appropriate low flow forecast model for the Meuse River In *Hydroinformatics in hydrology, hydrogeology and water resources. Proc. Symposium JS.4 at the Joint IAHS & IAH Convention,* I.D. Cluckie YC, V. Babovic, L. Konikow, A. Mynett, S. Demuth and D.A. Savic ed.: Hyderabad, India; 296–303.

2. **Demirel MC**, Booij MJ. 2010. Identification of appropriate temporal scales of dominant low flow indicators in the Main River, Germany. In *Global Change : Facing Risks and Threats to Water Resources,* Servat E, Demuth S, Dezetter A, Daniell T (eds). IAHS Press; 538–543.

Conference Proceedings

1. **Demirel MC**, Booij MJ, Hoekstra AY. 2008. Seasonal and long-term prediction of low flows in the Rhine Basin In *NCR-days 2008,* Erdbrink CD, Os van AG eds). Dalfsen, the Netherlands; 54–55.(Full paper + poster presentation)

2. **Demirel MC**, Booij MJ. 2011a. Low flow forecasting with a lead time of 14 days for navigation and energy supply in the Rhine River. *Geophysical Research Abstracts* **13**: 8189. (Abstract + Poster presentation)

3. **Demirel MC**, Booij MJ. 2011b. Uncertainty analysis of a low flow model for the Rhine River. *Geophysical Research Abstracts* **13**: 8280. (Abstract + Poster presentation)

About the author

Mehmet Cüneyd Demirel (born and grown up in Konya-Turkey, 03-01-1980) received his civil engineering bachelor in 2001 from Selçuk University, Turkey. Right after graduation, he went to Guisveld, the Netherlands to work for a summer project organized by Staatsbosbeheer. In 2002, he moved to Istanbul to execute a Master study on hydrology. He received his first Master's degree from Istanbul Technical University in 2004. He also completed a European Joint Master program (Erasmus Mundus) on Water and Coastal Management in March 2007, which consisted of a 1.5 year study period spread over three different institutions, in Cádiz (Spain), Faro (Portugal) and Miami (USA). After this nomad Erasmus life, he started his PhD at the Water Engineering and Management (WEM) Department of the University of Twente (UT) in August 2008. During his PhD research, he supervised two master students from UT and one Erasmus PhD student from Isparta University, Turkey. Besides low flows in the River Rhine, described in this thesis, his interests lie in cooking, traveling, web design, dancing, horse riding, taking driving license exams at CBR-Enschede (!) and learning new languages.

He established the "Dutch Toastmasters" club for promoting Dutch speaking between UT international students. He organized lunch talks on water related topics in the WEM Department (2009-2013), participated in the Batavierenrace (2009 and 2012) and undertook a tour-de-Rhine from Köln to Dordrecht with his mountain bike. He also followed a horse riding course at the Hippocampus, Enschede until the thesis writing period (too risky ! ☺) and received his 2^{nd} driving license from CBR-Enschede, (puff, was a challenge for this Turco!).

Mehmet is currently a postdoc at Portland State University in Oregon, USA. He is working on an exciting project on the comprehensive assessment of climate change impact on the hydrology and water resources of the Columbia River Basin with quantification of uncertainty in all layers of hydro-climate modeling and projection.