Skill and value of global seasonal streamflow forecasts

by
Naze Candogan Yossef

Skill and value of global seasonal streamflow forecasts

De nauwkeurigheid en waarde van seizoensvoorspellingen van mondiale rivierafvoer

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de
Universiteit Utrecht
op gezag van de
rector magnificus, prof. dr. H.R.B.M. Kummeling,
ingevolge het besluit van het College voor Promoties
in het openbaar te verdedigen op

vrijdag 6 september 2024 des middags te 12.15 uur

door

Naze Candogan Yossef

geboren op 25 september 1970
te Ankara, Turkije

Promotor:
Prof. dr. M.F.P. Bierkens

Copromotor:
Dr. L.P.H. van Beek

Beoordelingscommissie:
Prof. dr. H. Cloke
Prof. dr. A. van Griensven
Prof. dr. H. Middelkoop
Prof. dr. ir. R. Uijlenhoet
Dr. M. Werner

Dit proefschrift werd (mede) mogelijk gemaakt met financiële steun van Deltares.

Copyright © 2024 Naze Candogan Yossef

Author information: ncandogan@hotmail.com

Cover design by Margot Stoete

ISBN 978-90-6266-696-6

USES 322

Printed in the Netherlands by Ipskamp.

Contents

Chapter 1
Introduction

1.1 Background

Global drivers such as climate change, population growth, rapid urbanization and land use changes may adversely affect the hydrological cycle, causing challenging water-related problems. The magnitude and frequency of hydrological extreme events, such as floods and droughts are rising; and so are the associated economic losses (WMO, 2021; Fig 1.1). Although there is evidence that the vulnerability to these extreme events has been decreasing due to successful risk management (Formetta and Feyen, 2019), the rise in flood and drought hazard under climate change (IPCC, 2012), together with increased exposure due to population and economic growth will likely lead to higher flood and drought risks in the future (Winsemius et al., 2015; Zhai et al., 2020).

Addressing these increased hazards and risks driven by global changes, calls for a global approach, where the interacting atmospheric, ocean and land as well as human systems are integrated in an Earth systems concept (Sood and Smakhtin, 2014; Harrigan et al., 2020). It also demands hydrological knowledge at the global scale, since climate and other human induced changes affect the terrestrial water cycle beyond the scale of a catchment (Bierkens et al., 2015; Eagleson, 1986). These needs have given rise to global hydrology as a growing research field, that collaborates closely with other scientific disciplines, such as atmospheric sciences and water resources management. Global hydrology is concerned with understanding the role of environmental change on the Earth's water cycle and focuses on the development of global-scale models to describe the terrestrial hydrology (Bierkens et al., 2015; Wood et al., 2011).

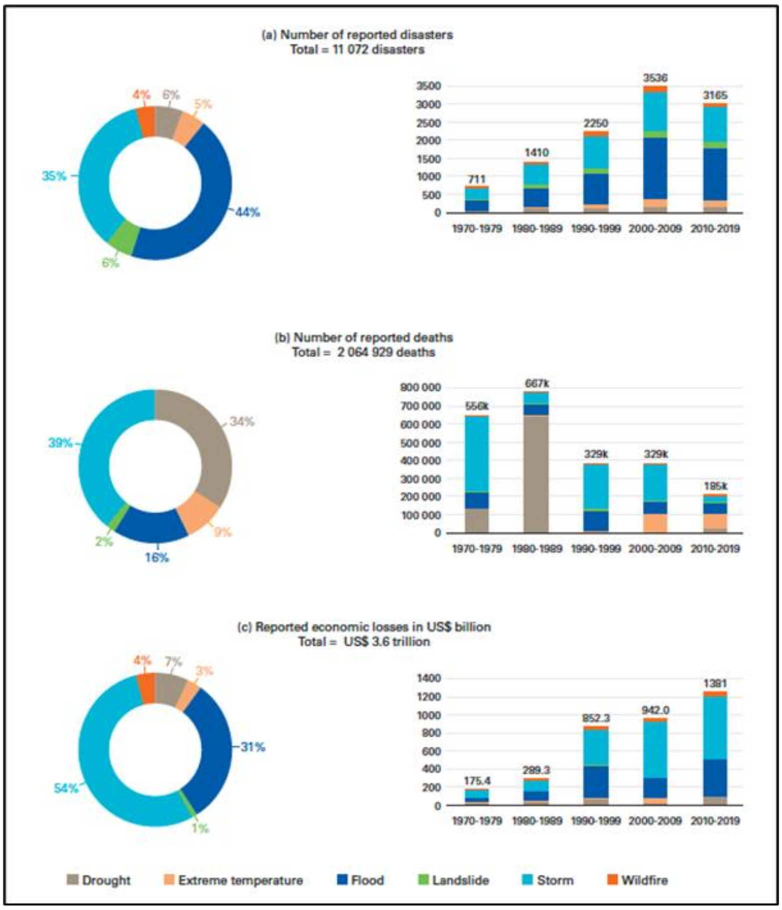


Figure 1.1 Number of (a) reported disasters, (b) number of death and (c) economic losses by hazard type by decade globally (WMO Atlas of Mortality and Economic Losses from Weather, Climate and Water Extremes (1970–2019)).

Global hydrological models (GHMs) that simulate land surface dynamics of the hydrological cycle on a global scale have developed rapidly over the past decades. GHMs are comparable to land surface models (LSMs), such as H-TESSEL (Pappenberger et al., 2011; Balsamo et al., 2009), ISBA-SGH (Decharme and Douville, 2006), MOSES (Gedney and Cox, 2003), NOAH (Ek et al., 2003), MATSIRO (Takata et al., 2003) and SWAP (Gusev and Nasonova, 2003), which were introduced in general circulation models (GCMs) to resolve the land component and provide realistic lower boundary conditions on temperature and moisture (Decharme and Douville, 2007). Although largely similar to LSMs, GHMs focus more on modeling runoff and streamflow, as well as a more comprehensive representation of the terrestrial hydrological processes. Examples are VIC (Wood et al., 1992; Nijssen et al., 2001), WaterGap (Alcamo et al., 2003; Döll et al., 2003), LaD (Milly and Schmakin, 2002), WBM (Vörösmarty et al., 2000; Fekete et al., 2002), MacroPDM (Arnell, 1999; 2004) and PCR-GLOBWB (Van Beek et al., 2011; Sutandudjaja et al., 2018). Most GHMs run at a daily time step and a spatial resolution of 30 arc min, corresponding to 50 km at the equator. The spatial resolution of GHMs is determined by computational resources as well as the available resolution of climate input data provided by global climate models. As higher computational power and finer resolution global input data become available, newer versions of GHMs are being developed which run on finer resolutions down to 5 arc min, such as WaterGAP, PCR-GLOBWB 2.0 and WBMplus (Bierkens et al., 2015; Sood and Smakhtin, 2015; Sivapalan, 2018). In addition to refinements of spatial resolution, newer versions of GHMs are being improved by features such as inclusion of reservoir operations, hydrodynamic routing and floodplain inundation (Bierkens et al., 2015).

GHMs and LSMs have been widely applied to estimate current and future continental runoff (Nijssen et al., 2001a; Fekete et al., 2002; Milly et al., 2005), to investigate the hydrological response to global warming (Arnell, 2004; Nijssen et al., 2001b; Milly et al., 2005), to study future projections of extremes in river discharge (Hirabayashi et al., 2008; Lehner et al., 2006), to assess freshwater availability (Alcamo et al., 2003; Islam et al., 2007; Oki et al., 2001; Vorosmarty et al., 2000; Van Beek et al., 2011; Wada et al., 2011), to map global flood hazard and risk (Pappenberger et al., 2012; Hirabayashi et al., 2013; Ward et al., 2013), to model global freshwater temperature (van Beek et al., 2012; Van Vliet et al., 2012) and to determine the contribution of terrestrial water stores to global sea level change (Wada et al., 2012; Pohkrel et al., 2013). While some of these applications focus on long-term trends, all rely on the ability of GHMs and LSMs to capture the relation between weather and hydrology, including the occurrence of extremes, such as floods and droughts, on shorter intervals, i.e., days to months.

1.2 Research problem

Studies on GHMs demonstrate the potential of their application for streamflow forecasting in large river basins. Streamflow forecasting with GHMs has been enabled in the past decade thanks to recent scientific and technological developments. These developments include advancement of global modelling capabilities both in meteorology and in land surface hydrology, enhanced collaboration between hydrological and meteorological communities, increased availability and quality of relevant data derived from ground observations and remote sensing by satellite and ground-based radars, as well as improvements in computing capabilities and resources (Emerton et al., 2016). Reliable and timely forecasts of water availability and scarcity are vital for mitigation of flood and drought hazards. While short- to mid-term forecasts inform immediate responses on the scales of hours to days, forecasts with lower temporal resolutions but longer lead times, e.g., months in advance, can be useful for increasing preparedness. Seasonal forecasts are beneficial not only in case of hydrological extreme events, but also during normal flow conditions, allowing several sectors to make more informed management decisions, in the fields of hydropower reservoirs, water supply, agriculture and navigation. The rationale behind operating global hydrological forecasting systems is that, as they are based on global meteorological datasets, they provide continuous and spatially consistent forecasts of streamflow. This may be valuable for regions where the spatial scale of hydrological extreme events goes beyond individual catchments or political borders as well as for the most vulnerable regions of the world where no local forecasting systems exist to alert the population. Still, where national scale forecasting systems exist, global forecasts provide an additional guidance at larger spatial scales (Harrigan et al., 2019). Disaster management organizations operating at global scale and international humanitarian aid agencies can benefit from global forecasts to prepare for appropriate response, and global water and energy markets can be informed about future availability of water and hydropower in different regions of the world. The economic rationale is that the provision of forecasts for basins across the globe does not require a large scale-up of resources. Rather than focusing on developing forecasting systems and issuing forecasts for individual basins in regions of scarce resources, it is more cost-effective to provide forecasts with global scale hydrological forecasting systems. Also, the economic benefit is evident for those countries who do have some existing capabilities, such as local hydrological models but are not able to produce hydrological forecasts, since they cannot afford to pay for access to, or processing of computationally expensive probabilistic and extended time scale meteorological forecast products (Emerton et al., 2018).

Despite their potential, GHMs have rarely been used for river flow forecasting up to now, mainly because appropriate routing of river discharge is not included, and forecasting systems are limited mainly to higher resolution national or regional domains. In the last decade, several seasonal hydrological forecasting systems have been developed for forecast applications and research purposes at the continental scale (De Roo et al., 2000; Bennett et al., 2016; Mo et al., 2014; Wood et al., 2002, 2005; Yuan et al., 2013). Continental scale operational systems include the European Flood Awareness System (EFAS) (Arnal et al., 2018), the European Hydrological Predictions for the Environment (E-HYPE) (Donnely et al., 2015), the Australian Government Bureau of Meteorology (BoM) Seasonal Streamflow Forecasts (BoM, 2018), and the National Hydrologic Ensemble Forecast Service (HEFS) for continental USA (Demargne et al., 2014; Emerton et al., 2016). Yet, currently only a few systems produce operational seasonal hydrological forecasts at the global scale, including the NASA Hydrological Forecast and Analysis System (NHyFAS) (Arsenault et al., 2020), the Global Flood Awareness System (GloFAS - Seasonal) (Emerton et al., 2018), the Global Flood Forecasting and Information System (GLOFFIS) (Emerton et al., 2016), and the Global Water Scarcity Information Service (GLOWASIS) (Weerts et al., 2013). In addition, a couple of studies on global scale forecast applications have been conducted for research purposes (Alfieri et al., 2013; Yuan et al., 2015), including three published studies by Candogan Yossef et al. (2012, 2013, 2017), which constitute the core of this thesis. These three studies together with an additional chapter of this thesis approach one main research problem from different angles. The main research problem with which this thesis is concerned may be stated as exploring the apparent but yet unmet potential of GHMs in operational seasonal forecasting applications on a global scale.

1.3 Research objectives

This thesis aims to assess the skill and value of seasonal streamflow forecasts produced by GHMs, as well as to investigate possible ways to improve the current skill and value. Here, the skill of a forecast is defined as its ability to capture the occurrence of an event; and the value as its usefulness in making informed decisions. In order to tackle the main research problem defined in the previous section, I have determined the following objectives which I address in the next four chapters.

1) to identify a methodology that can serve as a benchmark verification procedure for hydrological forecasting
2) to assess the prospect of using a GHM for forecasting hydrological extremes
3) to determine the relative contributions of meteorological forcing and initial hydrologic conditions to the skill of seasonal streamflow forecasts
4) to identify promising skill improvement methods
5) to assess the total skill of hydrological forecasts as affected by errors in model structure, in the estimation of initial hydrologic conditions as well as in the meteorological forcing obtained by numerical weather prediction
6) to shed light on the value of global scale seasonal streamflow forecasts for water management
7) to discuss possible ways to improve the value during various stages of the forecast chain

1.4 Context and outline

Figure 1.1 displays a conceptual schematization of the logical progression of my research and the organization of this thesis, with the research objectives addressed through Chapters 2 to 5. Chapter 2 presents the first study by Candogan Yossef et al. (2012). This study investigates the skill of the global hydrological model PC-Raster Global Water Balance (PCR-GLOBWB) in reproducing the occurrence of past extremes in the monthly discharges of 20 large rivers of the world. The motivation for this paper is to address the first two research objectives listed above: to present my evaluation of PCR-GLOBWB as an initial step in assessing the prospect of using a GHM for forecasting hydrological extremes, and to identify a methodology that can serve as a benchmark verification procedure for hydrological forecasting. This procedure uses methods and skill scores that were developed primarily for verification of meteorological forecasts. Global terrestrial hydrology is simulated for a historical period from 1958 until 2001, by forcing PCR-GLOBWB with a meteorological data set produced by combining ERA40 reanalysis (Uppala et al., 2005) and Climate Research Unit (CRU) data from the University of East Anglia (New et al., 2000). The use of a historical meteorological dataset implies that the hydrological forecasts are not affected by forecasting uncertainty in the forcing and the propagation thereof with increasing lead times. In this sense, the results presented here are indicative of the maximum skill that can currently be achieved by this and similar GHMs given the associated errors in forcing, discharge observations, model structure and parameterization. Monthly discharge observations from the Global Runoff Data Center (GRDC) reference dataset are used for verification. The skill of PCR-GLOBWB in reproducing hydrological extremes is assessed in three ways; a general verification of simulated hydrographs, assessment of the skill in reproducing significantly higher and lower flows than the monthly normals using skill scores for forecasts of ordinal categorical events, as well as the skill in reproducing flood and drought events using verification measures for forecasts of binary events, where floods and droughts are defined in terms of discharge values being higher or lower than discharges associated with a given return period.

This preliminary assessment in hindcast quantifies skill using a climate forcing based on observations and concludes that the prospects for seasonal forecasting with PCR-GLOBWB or comparable models are positive. However, this study does not include actual forecasts. Thus, the meteorological forcing errors due to uncertainty from numerical weather prediction models are not assessed. In an actual forecasting setup, the predictive skill of a hydrological forecasting system is affected by errors in model structure and parameterization, meteorological forcing (MF), and initial conditions (ICs), most importantly soil moisture, groundwater and snow. Skill of seasonal hydrological forecasts can thus be improved on the one hand by better prediction of future weather and on the other hand by better estimation of initial hydrologic states through assimilation of independent hydrological observations such as soil moisture and snow data from earth observation. The improvement in the overall predictability that may be attained depends on the relative importance of these two sources of uncertainty, MF and ICs, which varies considerably according to location, season and lead time (Bierkens and van den Hurk, 2007; Bierkens and van Beek, 2009; Shukla and Lettenmaier, 2011; Shukla et al., 2011). Therefore, determining the role of each factor is helpful in deciding which skill improvement methods are more promising (Paiva et al., 2012). The theoretical framework for quantifying the contributions of boundary forcing and initial conditions to predictability was developed in atmospheric sciences by Collins and Allen (2002). Wood and Lettenmaier (2008) and Wood et al. (2002, 2005) adopted this approach in hydrological forecasting. They presented an Ensemble

Streamflow Prediction (ESP) and reverse Ensemble Streamflow Prediction (revESP) approach and evaluated the relative roles of MF and ICs in seasonal hydrologic prediction in two western US basins. The ESP/revESP framework contrasts the forecast variance arising from a forecast ensemble based on perturbations of the initial states, and the forecast variance arising from an ensemble of meteorological forcing, to the internal, climatological variance.

The ESP/revESP approach is applied on the global scale to examine the relative contributions of ICs and MF to the skill of seasonal streamflow forecasts in my second study (Candogan Yossef et al., 2013), which is presented in Chapter 3 of this dissertation. As shown in Fig. 1.1, this study addresses research objectives 3 and 4; and investigates the roles of both sources of uncertainty in the skill of the global seasonal streamflow forecasting system Flood Early Warning System – World (FEWS-World), using the global hydrological model PCR-GLOBWB. Global monthly streamflow is simulated with lead times ranging from 1 to 6 months for a historical period of 30 years (1981–2010). The impact of both sources of uncertainty is analysed at 78 stations on large river basins across the globe by comparing the ESP and revESP forecast ensembles with retrospective model simulations driven by meteorological observations, and not with direct hydrological observations. In this way model errors are eliminated and predictability is related only to knowledge of ICs and the uncertainty in future MF. The results suggest that in some basins, and during certain seasons forecast skill may be improved by better estimation of initial hydrologic states through assimilation of snow, soil moisture or surface water data; whereas in others improvement of forecast skill depends on more accurate seasonal climate prediction. This analysis shows the relative contributions of ICs and MF to the potential skill of the forecasting system FEWS-World. However, in a real forecast mode, where both the ICs and the MF will be uncertain, the actual forecasting skill of the system should be assessed using probabilistic seasonal meteorological forecasts and comparing the ESP results to actual discharge observations, as I do in my third study.

Chapter 4 of this dissertation presents my third study (Candogan Yossef et al., 2017) which carries out a skill assessment in actual forecasting mode. In this study, I address research objective 5, as shown in Fig 1.1, and assess the total skill of hydrological forecasts as affected by errors in model structure, in the estimation of IC as well as in the actual MF obtained by numerical weather prediction. The skill of seasonal streamflow forecasts with the global hydrological forecasting system FEWS-World, which incorporates the global hydrological model PCR-GLOBWB, is investigated for 20 of the largest rivers in the world. The ability of FEWS-World to predict high and low flows, defined as discharges higher than the 75th and lower than the 25th percentiles for a given month respectively, is evaluated. The skill of streamflow forecasts using the seasonal climate forecast dataset S3 by the European Centre for Medium Range Weather Forecasts (ECMWF) is quantified as compared to the reference ESP forecasts using the Brier skill score (BSS), both for high and low flows. The analysis of the results in the context of prevailing hydroclimatic conditions suggest that the skill varies considerably according to location, season and lead time. The performance of the S3 forecast run is found to be generally close to that of the ESP run, with some basins where the ECMWF S3 forecast run performing significantly better than the ESP, during certain periods of the year and at certain lead times; but in fact, there are more cases where the forecast run performs worse than the ESP. The study concludes that in most cases, the apparent potential for improvement in seasonal hydrological forecasts by using climate predictions cannot be realized as yet until more accurate hydrological models and more skilful seasonal meteorological forecasts become available in the future.

Chapter 5 of this thesis investigates the value of global scale seasonal streamflow forecasts. I address the research objectives 6 and 7, as schematized in Fig 1.1, where skill contributes to value. With the skill varying considerably by region and season and showing a decrease with increasing lead time, in many cases, seasonal streamflow forecasts produced by these large-scale systems need to have a better skill before they can be adopted for water management applications (Samaniego et al., 2019). However, even with little added skill, the forecast may still be useful for end-users, allowing them to decide for themselves if they should take the risk of using the forecast information (Viel et al., 2016). The success of a hydrological forecasting system will ultimately be determined not by its skill but by the effect it has on decision-making for water management (Plummer et al., 2019). The current ability of seasonal streamflow forecasting systems to predict the right category of an event months ahead is potentially valuable for many water-related applications. The aim of this study, therefore, is to shed some further light on the *value* problem. For this purpose, I study the interaction between skill and value and discuss the possible ways to improve the value of seasonal streamflow forecasts on a global scale with an emphasis on flood and drought mitigation.

The final Chapter 6 of this thesis presents a synthesis of my conclusions and future recommendations.

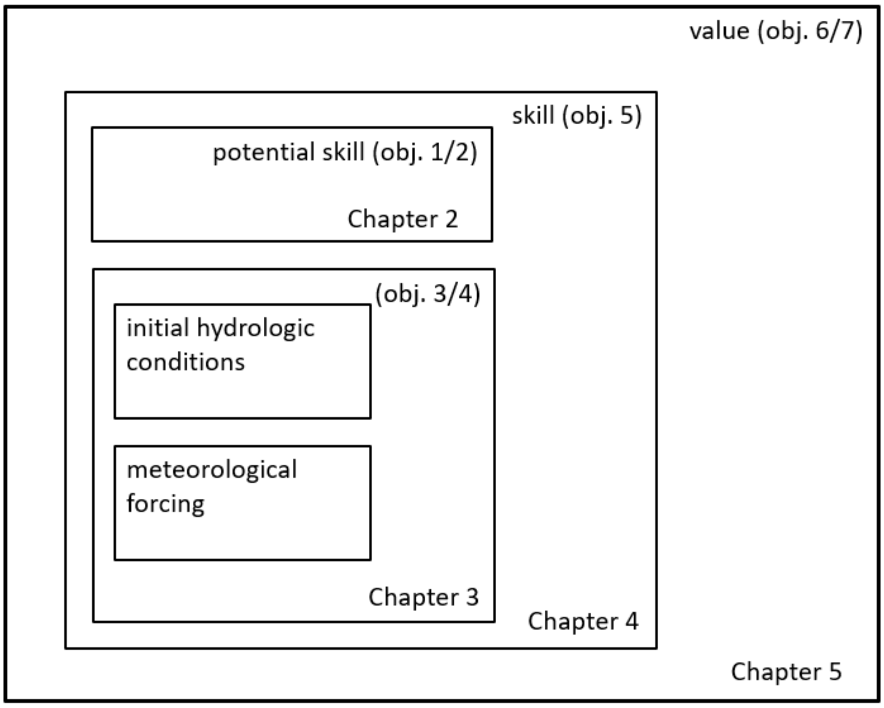


Figure 1.2 Conceptual schematization of the logical progression of research in this thesis on the skill and value of seasonal global hydrological forecasts.

Chapter 2
Assessment of the potential forecasting skill of a global hydrological model in reproducing the occurrence of monthly flow extremes

Abstract

As an initial step in assessing the prospect of using global hydrological models (GHMs) for hydrological forecasting, this study investigates the skill of the GHM PCR-GLOBWB in reproducing the occurrence of past extremes in monthly discharge on a global scale. Global terrestrial hydrology from 1958 until 2001 is simulated by forcing PCR-GLOBWB with daily meteorological data obtained by downscaling the CRU dataset to daily fields using the ERA-40 re-analysis. Simulated discharge values are compared with observed monthly streamflow records for a selection of 20 large river basins that represent all continents and a wide range of climatic zones.

We assess model skill in three ways all of which contribute different information on the potential forecasting skill of a GHM. First, the general skill of the model in reproducing hydrographs is evaluated. Second, model skill in reproducing significantly higher and lower flows than the monthly normals is assessed in terms of skill scores used for forecasts of categorical events. Third, model skill in reproducing flood and drought events is assessed by constructing binary contingency tables for floods and droughts for each basin. The skill is then compared to that of a simple estimation of discharge from the water balance (P-E).

The results show that the model has skill in all three types of assessments. After bias correction the model skill in simulating hydrographs is improved considerably. For most basins it is higher than that of the climatology. The skill is highest in reproducing monthly anomalies. The model also has skill in reproducing floods and droughts, with a markedly higher skill in floods. The model skill far exceeds that of the water balance estimate. We conclude that the prospect for using PCR-GLOBWB for monthly and seasonal forecasting of the occurrence of hydrological extremes is positive. We argue that this conclusion applies equally to other similar GHMs and LSMs, which may show sufficient skill to forecast the occurrence of monthly flow extremes.

2.1 Introduction

Global hydrological models (GHMs) that simulate land surface dynamics of the hydrological cycle on a global scale have developed rapidly over the past decades. GHMs are comparable to land surface models (LSMs), such as H-TESSEL (Pappenberger et al., 2011; Balsamo et al., 2009), ISBA-SGH (Decharme and Douville, 2006), MOSES (Gedney and Cox, 2003), NOAH (Ek et al., 2003), MATSIRO (Takata et al., 2003) and SWAP (Gusev and Nasonova, 2003), which were introduced in general circulation models (GCMs) to resolve the land component and provide realistic lower boundary conditions on temperature and moisture (Decharme and Douville, 2007). Although largely similar to LSMs, GHMs focus more on modelling runoff and streamflow, as well as a more comprehensive representation of the terrestrial hydrological processes. Examples are VIC (Wood et al., 1992), WaterGap (Döll et al., 2003), LaD (Milly and Schmakin, 2002), WBM (Fekete et al., 2002), and Macro-PDM (Arnell, 1999). GHMs and LSMs have been widely applied to estimate current and future continental runoff (Nijssen et al., 2001a; Fekete et al., 2002; Milly et al., 2005), to investigate the hydrological response to global warming (Arnell, 2004; Nijssen et al., 2001b; Milly et al., 2005), to study future projections of extremes in river discharge (Hirabayashi et al., 2008; Lehner et al., 2006) and to assess freshwater availability (Alcamo et al., 2003; Islam et al., 2007; Oki et al., 2001; Vörösmarty et al., 2000; Van Beek et al., 2011; Wada et al., 2011). Given the capability of GHMs to quantify streamflow, their relevance for integrated water resources management of large river basins has been recognized (Refsgaard, 2001). Reliable and timely forecasts of extremes in streamflow can help mitigate flood and drought risks and optimize water allocations to different sectors and sub-regions. The application of GHMs could be particularly promising for developing regions of the world where no effective flood and drought early warning systems are in place. However, up to now large-scale hydrological models have rarely been used for river flow forecasting, mainly because appropriate routing of river discharge is not included, and forecasting systems are limited to higher resolution national or regional domains (e.g., the European LISFLOOD system with a grid resolution of 5 × 5 km; De Roo et al., 2000).

In this paper we investigate the skill of the global hydrological model PCR-GLOBWB in reproducing the occurrence of past extremes in the monthly discharges of 20 large rivers of the world that represent all continents and a wide range of climatic zones. The motivation for the paper is twofold. The first objective is to present our evaluation of PCR-GLOBWB as an initial step in assessing the prospect of using a GHM for forecasting hydrological extremes. The second one is to identify a methodology that can serve as a benchmark verification procedure for hydrological forecasting. This procedure uses methods and skill scores that were developed primarily for verification of meteorological forecasts.

Global terrestrial hydrology is simulated for a historical period from 1958 until 2001, by forcing PCR-GLOBWB with a meteorological data set produced by combining ERA-40 reanalysis (Uppala et al., 2005) and Climate Research Unit (CRU) data from the University of East Anglia (New et al., 2000). The use of a historical meteorological dataset implies that the hydrological forecasts are not affected by forecasting uncertainty in the forcing and the propagation thereof with increasing lead times. In this sense, the results presented here are indicative of the maximum skill that can currently be achieved by this and similar GHMs given the associated errors in forcing, discharge observations, model structure and parameterization.

We assess the skill of PCR-GLOBWB in reproducing hydrological extremes in three ways. First, a general verification of simulated hydrographs is carried out. Second, model skill in reproducing significantly higher and lower flows than the monthly normals is assessed by constructing categorical contingency tables and applying skill scores used in meteorology for forecasts of ordinal categorical events. Third, model skill in reproducing flood and drought events is assessed by applying verification measures for forecasts of binary events, where floods and droughts are defined in terms of discharge values being higher or lower than discharges associated with a given return period. The model skill quantified in terms of these three sets of skill scores is then compared with the skill obtained by a simple estimation of discharge from the water balance (P-E) over each basin.

We use discharge observations from the Global Runoff Data Center (GRDC) reference dataset which contains monthly discharges for most basins. Consequently, the forecasting skill that we assess in this study is indicative for the potential skill that could be achieved in monthly and seasonal forecasting, rather than medium-range forecasting. Among other studies in which the discharge simulations of other GHMs and LSMs have been compared to discharge observations, the novelty of this work is to evaluate the ability of a GHM in reproducing the occurrence of anomalous flows and past flood and drought events with skill measures used in verification of meteorological forecasts, in the prospective context of operational hydrological forecasting.

The rest of this paper is set up as follows: Sect. 2 describes the GHM PCR-GLOBWB, the historical simulation, the meteorological forcing as well as the discharge data used for skill assessment. Section3 describes the assessment of skill in reproducing hydrographs, anomalous flows and floods and droughts. Results are presented and discussed in Sect. 4, followed by conclusions in the last section.

2.2 Historical simulation

2.2.1 Hydrological model

PCR-GLOBWB (PCRaster Global Water Balance) is a hydrological model that simulates the terrestrial part of the global water cycle (Van Beek and Bierkens, 2009; Bierkens and Van Beek, 2009). It is coded in the high-level computer language PCRaster for constructing environmental models (Wesseling et al., 1996). PCR-GLOBWB is fully distributed and operates on a regular grid with a cell size of 0.5×0.5^{O} (ca. 55 km squared at the Equator). Meteorological forcing is applied on a daily time step and assumed to be constant over the grid cell. Sub-grid variability is taken into account in the representation of short and tall vegetation, open water, different soil types, saturated area, surface runoff, interflow and groundwater discharge.

PCR-GLOBWB is a "leaky-bucket" type of model that calculates the water balance for every grid cell by tracking the transfer of water between the atmosphere and the cell, through stores within each cell, and laterally, as discharge, from one cell to the next. The model calculates the storages and fluxes of water, simulates the generation of runoff and its propagation as discharge through the river network. Precipitation falls either as snow or rain depending on atmospheric temperature. It can be intercepted by vegetation and added to the finite canopy storage, which is subject to open water evaporation. Snow is accumulated when the temperature is lower than 0^{O}C and melts when it is higher. Snow melt is added to rain and throughfall; it is stored in the available pore space in the snow cover, or reaches the top soil layer. Part of this water is transformed into surface runoff and the remainder infiltrates into the soil through two vertically stacked soil layers and an underlying groundwater layer. Water is exchanged between these layers following Darcy's law and the resulting soil moisture is subject to evapotranspiration. The remaining water contributes to lateral drainage as interflow from the soil layers or baseflow from the groundwater reservoir. The total drainage which consists of surface runoff, interflow and baseflow is routed through the drainage network of rivers, lakes and wetlands, based on DDM30 (Döll and Lehner, 2002), using the kinematic wave approach. An extensive description of PCR-GLOBWB can be found in Van Beek and Bierkens (2009) and Van Beek et al. (2011).

2.2.2 Meteorological dataset

The meteorological variables required to force PCR-GLOBWB are daily values of precipitation, evapotranspiration and temperature. In the absence of direct estimates of actual evapotranspiration, the model can be forced with values of potential evapotranspiration calculated from temperature, radiation, cloud cover, vapour pressure and wind speed.

In order to force PCR-GLOBWB with daily meteorological data at 0.5^{O} resolution, the monthly fields of the CRU TS 2.1 data set (New et al., 2000) have been downscaled to daily fields using ERA-40 reanalysis (Uppala et al., 2005). Precipitation fields are downscaled multiplicatively while an additive correction is used for temperature. Reference potential evapotranspiration is calculated first on a monthly basis, based on monthly cloud cover and vapour pressure deficit from CRU TS 2.1 as well as radiation and wind speed from CRU CLIM 1.0 (New et al., 2002). Reference evapotranspiration is converted to crop-specific potential evapotranspiration using crop factors derived following FAO guidelines. Finally, potential evapotranspiration is downscaled multiplicatively to daily values using ERA-40 temperature fields. The methodology used to calculate potential evaporation for the different land surfaces in PCR-GLOBWB and the downscaling of the meteorological data is described in detail by Van Beek (2008). The resulting meteorological data set is limited to the period from 1958 to 2001 for which ERA-40 data are available.

2.2.3 Simulated and observed discharge time series

The simulated discharge time series represent non-regulated flow. Twenty large river basins are selected for comparison of simulated and observed time series on the basis of three criteria. The first one is to represent all the continents, a wide range of climate zones and latitudes as well as a variety of precipitation regimes. The second criterion is the availability of observed monthly streamflow records for at least part of the period 1958–2001. The third criterion is to focus on developing regions which would benefit most from operational seasonal forecasting. Selected basins can be seen in Fig. 2.1 (Sperna Weiland et al., 2010). Basin characteristics and record length are presented in Table 1, adapted from Sperna Weiland et al. (2010).

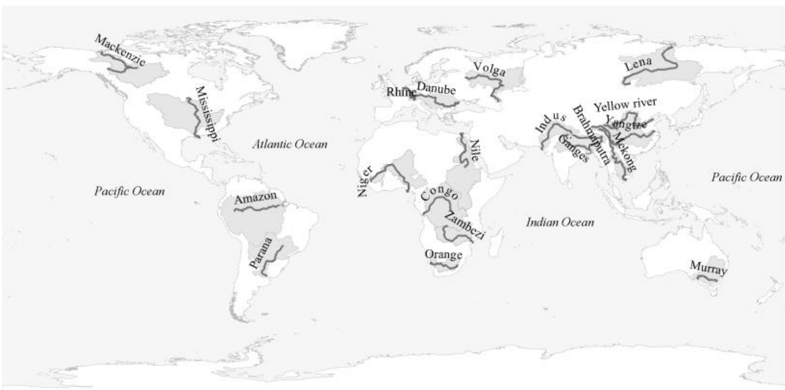
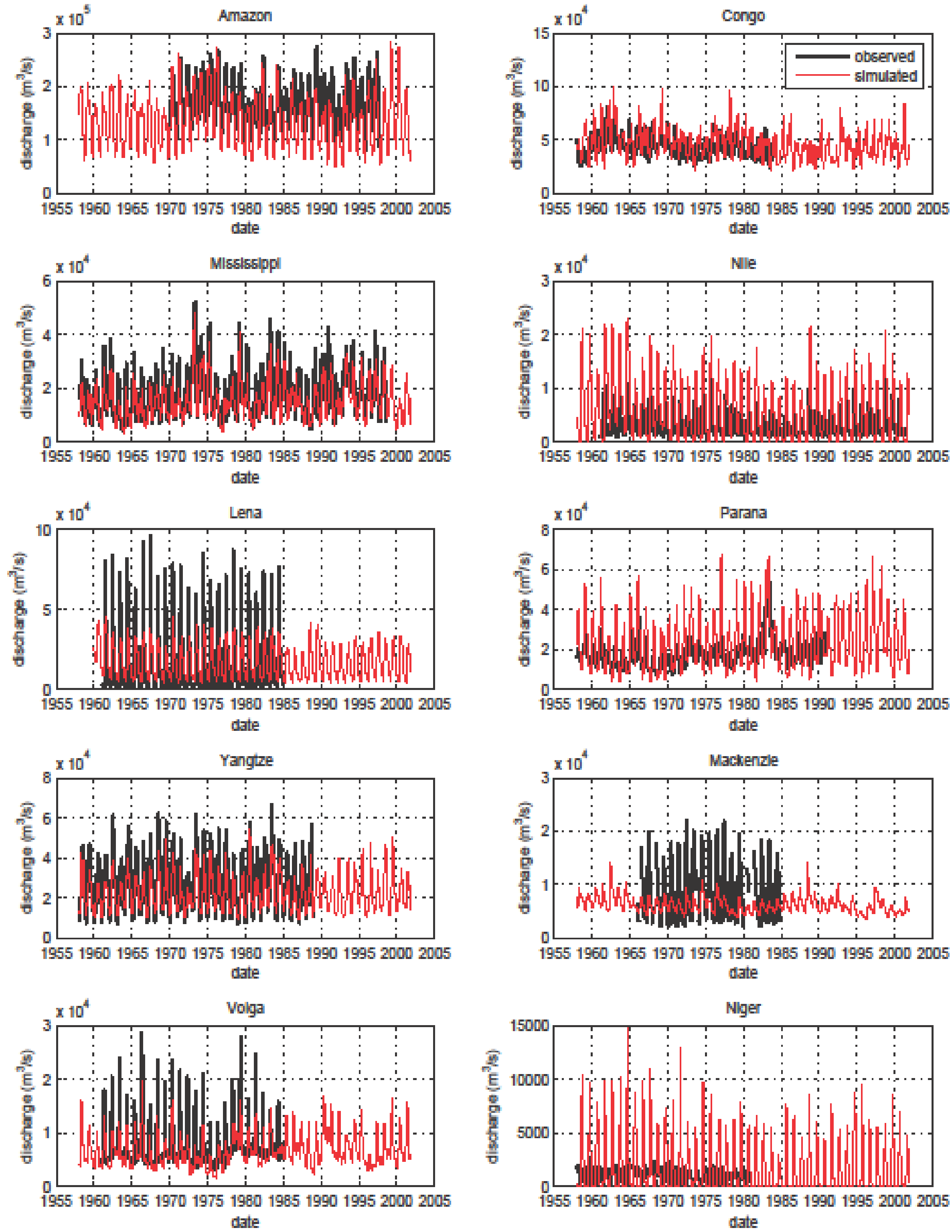


Figure 2.1 Selected basins.

The discharge data for most of the selected basins are obtained from the Global Runoff Data Center (GRDC, 2007). When GRDC data are not available, records from the Global River Discharge Database, RivDis 1.1 (Vörösmarty et al., 1998) are used. The period of record for the discharge values reported in the GRDC and RivDis databases varies widely from basin to basin (Table 2.1). Simulated daily discharges for the model grid cells corresponding to gauging stations are aggregated into monthly values, since this is the temporal resolution at which observed discharge data are available for validation. The simulated and observed discharge time series are used in the assessment of skill as described in the following section.

Table 2.1 Basins data.

Basin	Area (km^2)	Q avg (m^3/s)	Duration with available records
Amazon	6,915,000	190,000	28 years
Congo	3,680,000	41,800	26 years
Mississippi	2,981,076	12,743	40 years 9 months
Nile	3,400,000	2,830	40 years 7 months
Lena	2,500,000	17,000	24 years
Parana	2,582,672	18,000	33 years
Yangtze	1,800,000	31,900	31 years
MacKenzie	1,805,000	10,700	16 years 4 months
Volga	1,380,000	8,060	24 years
Niger	2,117,700	6,000	21 years 10 months
Murray	1,061,469	767	16 years
Orange River	973,000	365	20 years 3 months
Ganges	907,000	12,015	9 years
Indus	1,165,000	6,600	10 years 6 months
Danube	817,000	6,400	42 years 10 months
Yellow River	752,000	2,571	30 years
Brahmaputra	930,000	48,160	5 years 10 months
Rhine	65,638	2,200	29 years
Zambezi	1,390,000	3,400	4 years
Mekong	795,000	16,000	29 years 5 months


Amazon
Congo
observed
simulated
Mississippi
Nile
Lena
Parana
Yangtze
Mackenzie
Volga
Niger
discharge (m^3/s)
date
x 10^5
x 10^4

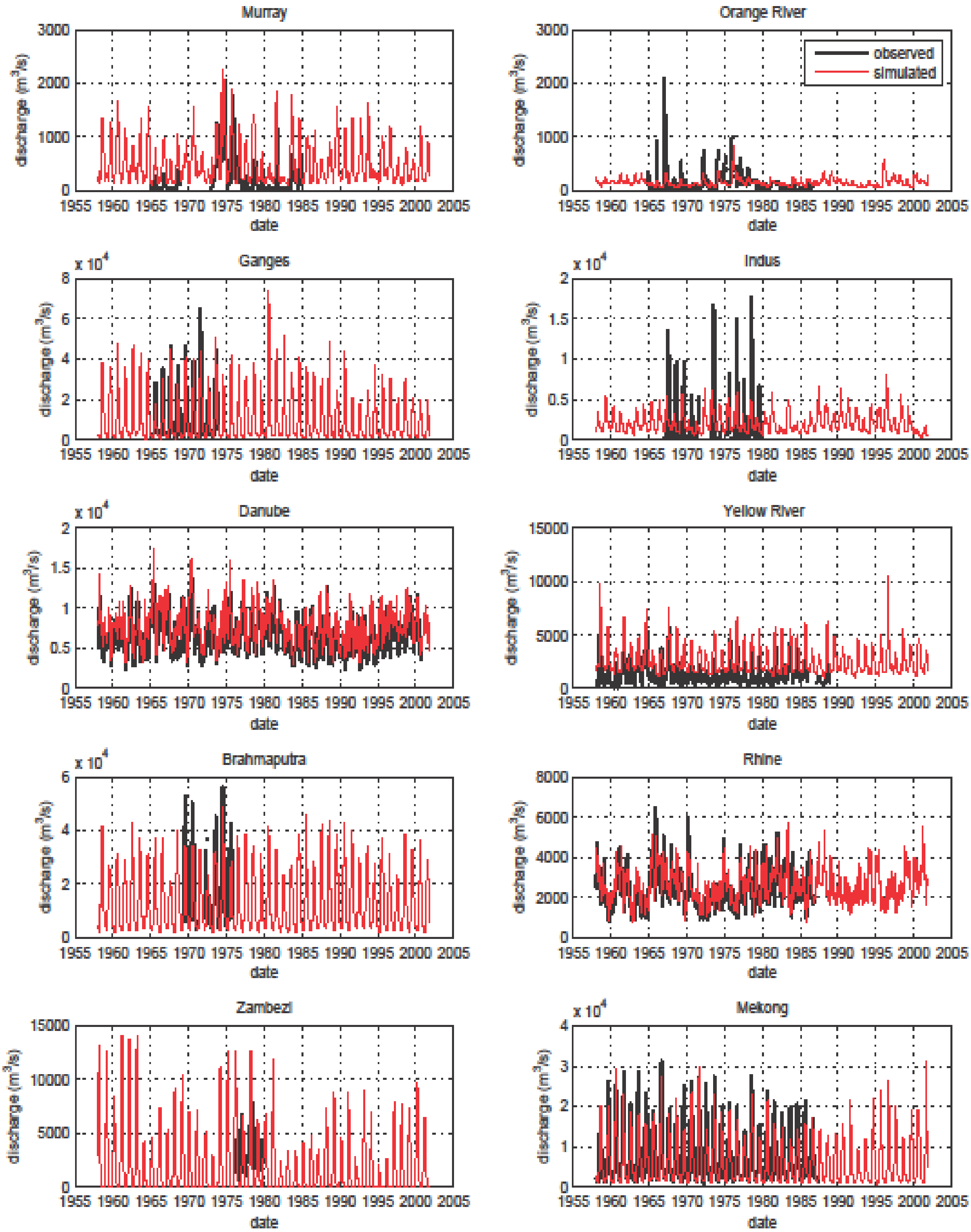


Figure 2.2 Discharge time series.

2.3 Skill assessment methodology

2.3.1 Measuring the skill in reproducing hydrographs

The performance of the model in hydrograph simulation is assessed in terms of verification measures used in forecasting of continuous variables, without applying thresholds. For this assessment, the most commonly applied statistical measure, mean squared error (MSE) is calculated for each river basin. In order to judge the predictive skill, the raw MSE scores are transferred into MSE Skill Scores, (MSESS). The MSESS provide a relative measure of the quality of the simulation compared to the mean climatology as a low skill alternative method of estimation. Here climatology refers to the long term mean of the available monthly discharge records for each of the 12 months of the year. The MSESS is defined as:

$$MSESS = 1 - \frac{MSE}{MSE_{climatology}}$$

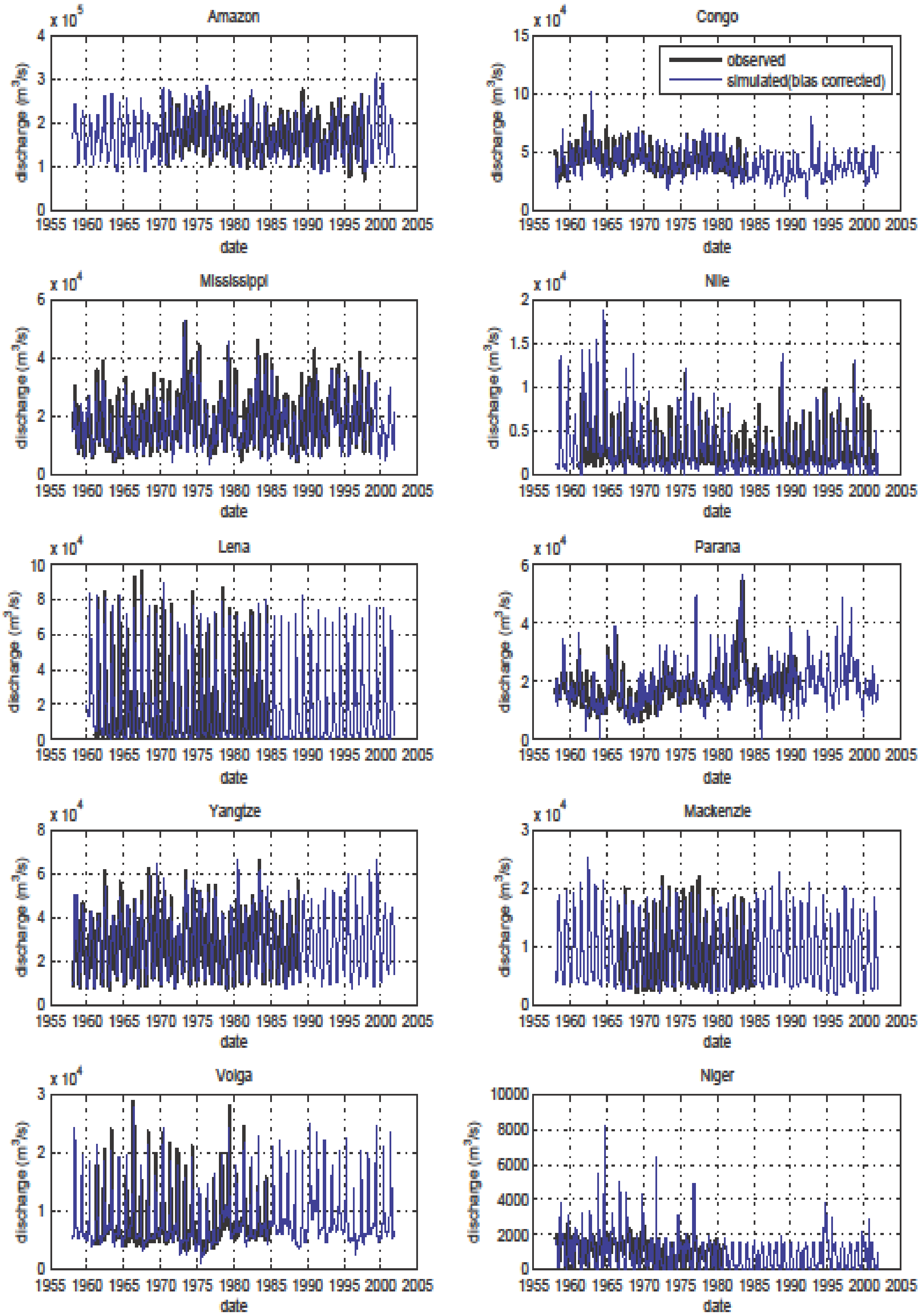
The range of values that MSESS can take is (-∞, 1); with the maximum value of 1 indicating perfect skill; a value of 0 indicating a model skill equivalent to the climatology; and a negative value implying that the model performs worse than the climatology.

Additionally, we use the coefficient of determination (R^2) and Nash and Sutcliffe's coefficient of efficiency (NS), which are often employed in the validation of hydrological models. These coefficients provide a measure of the model skill relative to the long-term mean, that is independent of the climatology. NS takes on the values (-∞, 1) and R2 (0, 1), with higher values indicating higher skill.

Bias due to errors in the meteorological forcing, discharge records, model parameters, or simplifying assumptions, can highly degrade the quality of the output of a hydrological model (Hashino et al., 2007). This is true for our simulations as well. We applied these skill measurement methods on both non-bias-corrected and bias-corrected simulation results. Verification with non-bias-corrected data presents a better reflection of potential shortcomings in the skill of the GHM and provides the opportunity to compare our simulations with the results of other studies which use non-bias-corrected data, such as the Water Model Intercomparison Project (WaterMIP), which quantifies and explains the differences in the results of five GHMs and six LSMs (Haddeland et al., 2011). Verification with bias-corrected data, on the other hand, is relevant for the assessment of forecasting skill, which is the ultimate purpose of this study. It provides an indication of the maximum skill that can be achieved when the systematic bias due to model errors or forcing is eliminated, as is generally the case in operational forecasting.

Table 2.2 Skill scores for reproducing hydrographs.

Basin	uncorrected			bias corrected		
	MSESS	R^2	NS	MSESS	R^2	NS
Amazon	-4.92	0.55	-0.13	-0.29	0.79	0.75
Congo	-3.83	0.27	-0.87	-0.35	0.64	0.48
Mississippi	0.40	0.77	0.68	0.72	0.85	0.85
Nile	-31.51	0.59	-4.35	-4.38	0.57	0.11
Lena	-7.81	0.62	0.52	0.40	0.97	0.97
Parana	-2.10	0.48	-1.70	0.48	0.65	0.54
Yangtze	-0.89	0.89	0.64	0.75	0.95	0.95
Mackenzie	-10.51	0.62	0.11	0.33	0.95	0.95
Volga	-0.81	0.58	0.51	0.50	0.86	0.86
Niger	-81.30	0.11	-18.62	-6.75	0.32	-0.85
Murray	-0.70	0.37	-0.45	0.32	0.48	0.42
Orange River	0.11	0.22	0.20	0.17	0.26	0.25
Ganges	0.33	0.90	0.90	0.47	0.92	0.92
Indus	-1.63	0.12	0.12	0.08	0.69	0.69
Danube	-0.04	0.68	0.38	0.50	0.76	0.70
Yellow River	-1.98	0.77	-0.49	0.57	0.79	0.78
Brahmaputra	-1.40	0.88	0.71	0.32	0.92	0.92
Rhine	0.57	0.72	0.65	0.74	0.79	0.79
Zambezi	-1.49	0.16	-1.13	0.24	0.38	0.35
Mekong	-0.61	0.85	0.82	0.13	0.90	0.90


Amazon
x 10^5
Congo
x 10^4
observed
simulated(bias corrected)
Mississippi
Nile
Lena
Parana
Yangtze
Mackenzie
Volga
Niger
discharge (m^3/s)
date

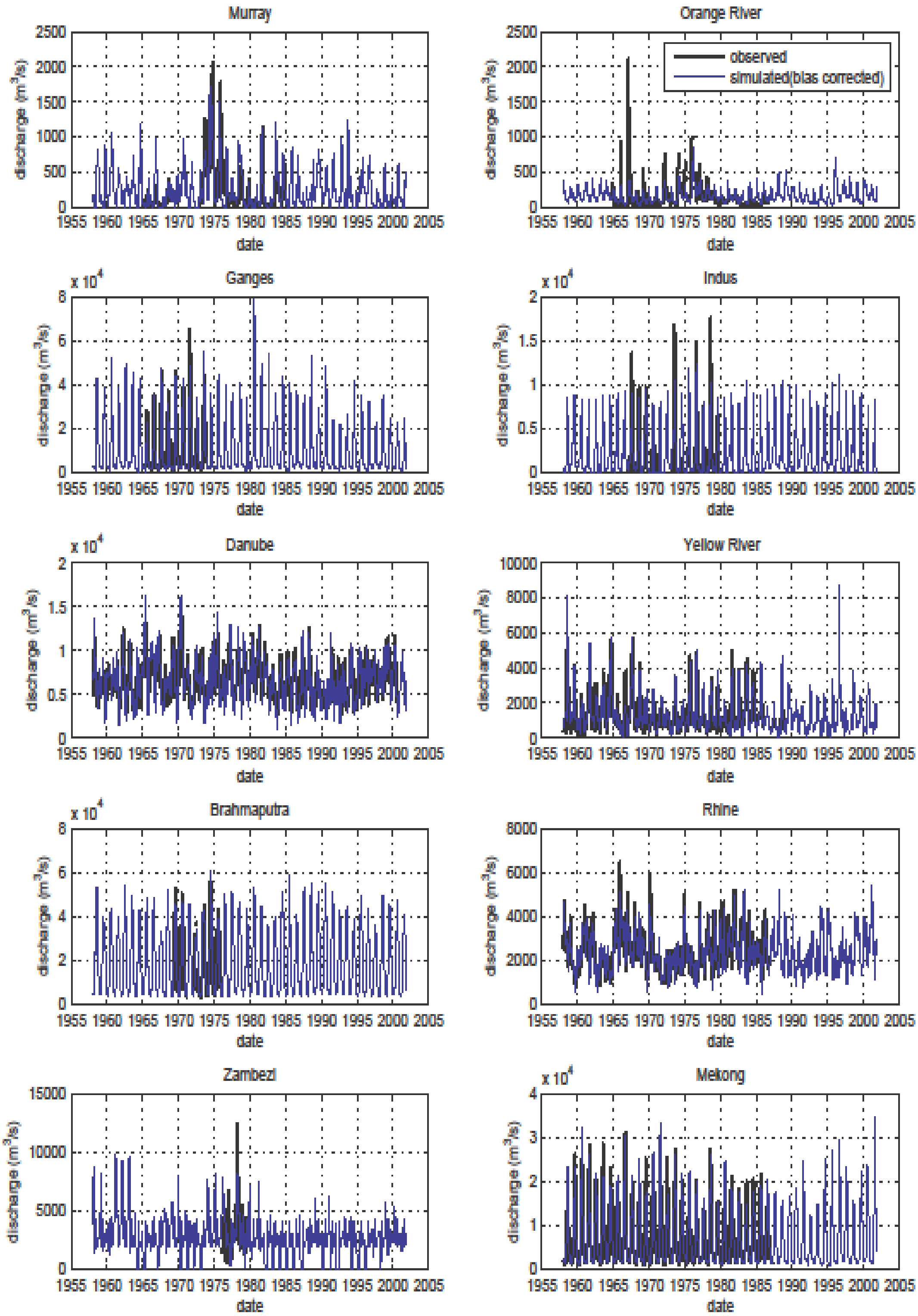
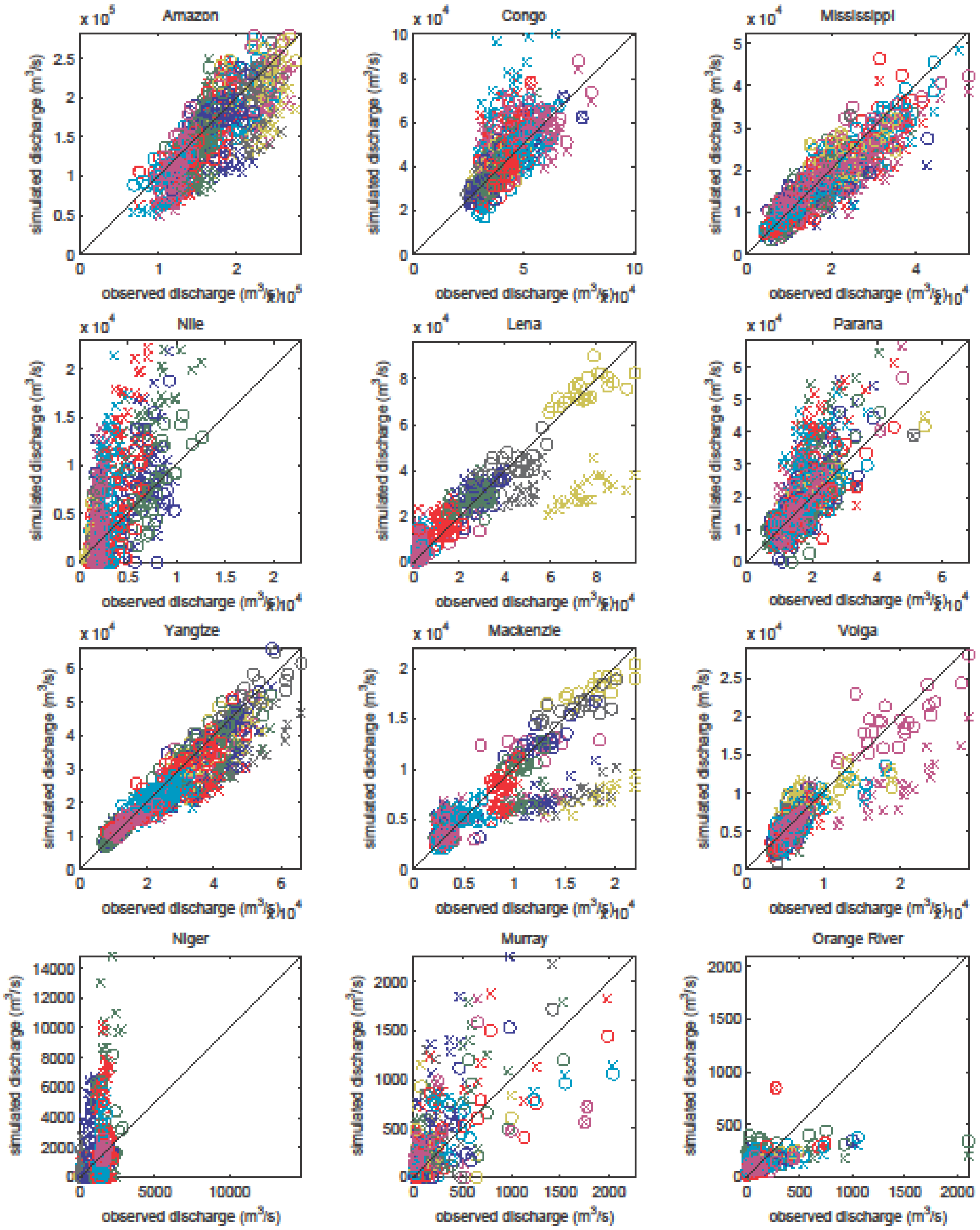


Figure 2.2 Bias-corrected discharge time series.


Amazon
Congo
Mississippi
Nile
Lena
Parana
Yangtze
Mackenzie
Volga
Niger
Murray
Orange River
simulated discharge (m³/s)
observed discharge (m³/s)
x 10⁵
x 10⁴

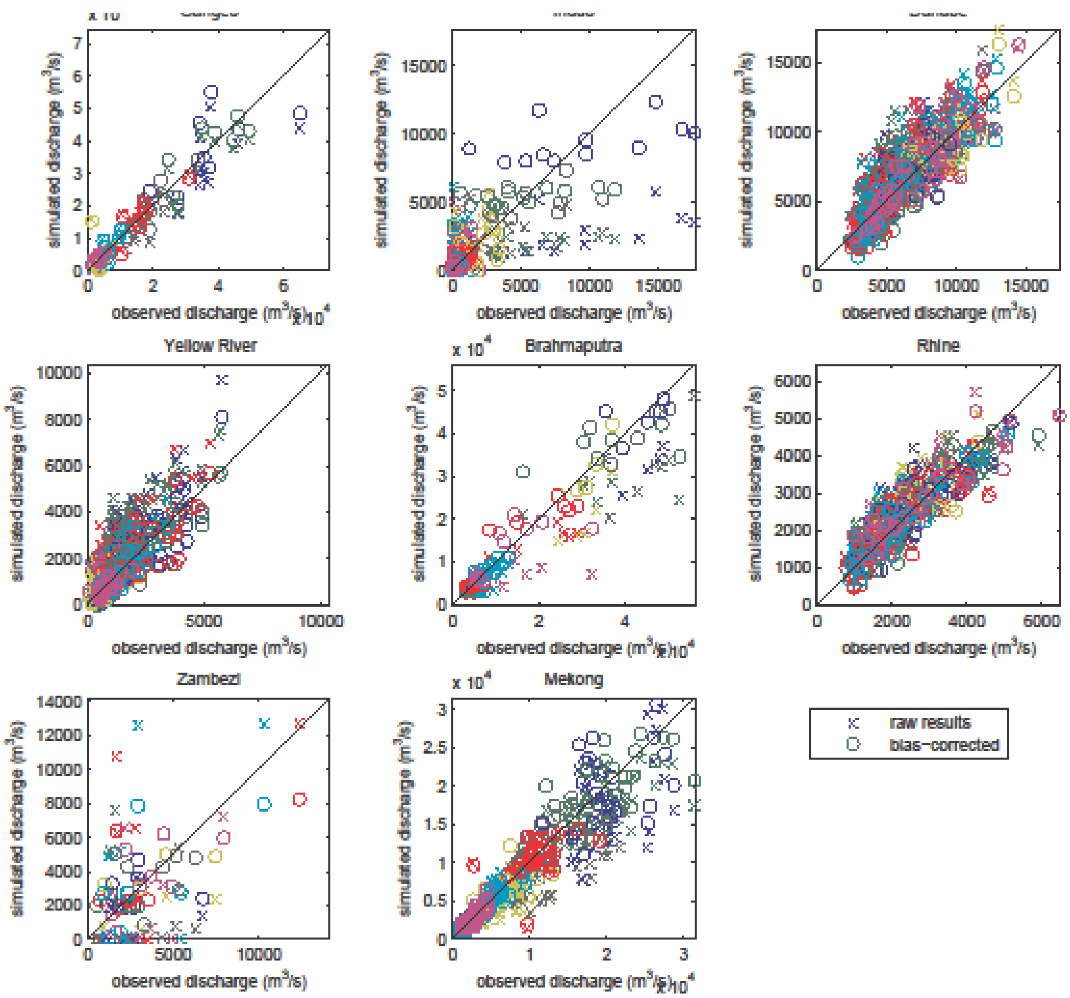


Figure 2.3 Reliability diagrams (different colors indicate different months of the year).

In this study a simple method of a posteriori bias-correction is carried out. It is true that an a priori correction by basin-specific calibration has a stronger physical basis than an a posteriori adjustment of the model output. On the other hand, given the time, data and computational capacity required for model calibration, a simple post-processing has the advantage of being far more straightforward and transparent. The post-processing method we employed is as follows: Bias is calculated for each pair of simulation and observation. Calculated biases are grouped into 12 months of the year, and a mean bias is calculated for each of these 12 months. Every discharge value is corrected for the mean bias calculated for the corresponding month of the year. The correction is done by simply subtracting the mean bias for the corresponding month from the simulated monthly discharge value.

2.3.2 Measuring the skill in reproducing anomalous flows

In order to analyse whether the model is capable of reproducing higher or lower flows than usual for a given month, the discharge time series are transformed into categorical events defined in terms of three categories of high, normal and low flow. The analysis is carried out for two different sets of categories. For the first set, high flow is defined as discharge values above the 75th percentile for the month in question; normal flow between the 75th and the 25th percentile; and low flow below the 25th percentile. For the second set, the 90th and the 10th percentiles are used. Thresholds are identified separately for simulated and observed discharge. This approach eliminates any systematic under- or overestimation in the simulations and thus removes the need for bias correction. The skill in simulating these three classes is assessed by constructing categorical contingency tables and applying skill scores for ordinal categorical events.

Here we use Gerrity Scores (GS) (Gerrity, 1992) which is a subset of the Gandin and Murphy (GM) family of equitable scores for deterministic categorical forecasts (Gandin and Murphy, 1992). The criterion of equitability is based on the principle that random forecasts or constant forecasts of the same single category receive a no-skill score (Murphy and Daan, 1985). GM scores use a scoring matrix which represents the reward or penalty accorded to each pair of simulation and observation on the contingency table. In contrast to other equitable scores such as the Heidke skill score (Heidke,1926), and Peirce's skill score (PSS) (Haansen and Kuipers, 1965), the GM family considers differences in relative sample probabilities of categories when appropriating a reward or penalty (Livezey, 2003). A correct forecast of a low probability category is rewarded more than that of a high probability category. Likewise, failure to forecast a rare event receives a lighter penalty than a common event.

GS and LEPSCAT scores (Potts et al., 1996) are the two subsets of the GM family that are appropriate for the specific case of ordinal categories, defined as ranges of a continuous variable such as discharge. In this study, GS are preferred since they are recommended by Livezey (2003) for ordinal categorical events, on the practical basis of being more convenient to use compared to LEPSCAT. GS provide higher penalties as the discrepancy between simulated and observed classes increase. For example, a forecast of low flow receives a heavier penalty when the observed flow is high, and a lighter one when the observed flow is normal.

This score takes on the maximum value of 1 for perfect skill, and the value of 0 for no-skill. The value of GS for a categorical forecast with K number of categories is given by Eq. (2.2):

$$GS = \sum_{i=1}^{K} \sum_{j=1}^{K} p_{ij}\, s_{ij}$$

where the relative sample frequency p_{ij} of each outcome on the K × K contingency table is multiplied by the corresponding scoring factor s_{ij} (i, j = 1, . . . , K) from a scoring matrix S with relative levels of rewards and penalties and summing the values. The elements s_{ij} of the scoring matrix S is given by Eq. (2.3):

$$S = \begin{bmatrix} s_{ii} & s_{ij} & \cdots & s_{iK} \\ s_{ji} & s_{jj} & \cdots & s_{jK} \\ \vdots & \vdots & \ddots & \vdots \\ s_{Ki} & s_{Kj} & \cdots & s_{KK} \end{bmatrix}$$

$$s_{ii} = b\left(\sum_{r=1}^{i-1} a_r^{-1} + \sum_{r=i}^{K-1} a_r\right)$$

$$s_{ij} = b\left(\sum_{r=1}^{i-1} a_r^{-1} - (j - i) + \sum_{r=j}^{K-1} a_r\right)$$

$$(1 \le i \le j \le K)$$
$$s_{ji} = s_{ij}$$

$$a_i = \frac{1 - \sum_{r=1}^{i} p_r}{\sum_{r=1}^{i} p_r}$$

$$p_r = \sum_{j=1}^{K} p_{rj}$$

$$b = \frac{1}{K - 1}$$

Table 2.3 Categorical contingency tables for 75th and 25th percentiles.

o: observed, s: simulated, L: low flow, N: normal flow, H: high flow.

Amazon

o \ s	L	N	H
L	53	27	4
N	35	96	37
H	1	32	51

Parana

o \ s	L	N	H
L	73	23	0
N	37	140	27
H	2	34	60

Murray

o \ s	L	N	H
L	30	14	4
N	29	46	21
H	4	18	26

Yellow River

o \ s	L	N	H
L	34	45	4
N	37	116	40
H	2	25	57

Congo

o \ s	L	N	H
L	24	40	8
N	16	101	51
H	1	14	57

Yangtze

o \ s	L	N	H
L	76	20	0
N	21	141	19
H	0	29	66

Orange River

o \ s	L	N	H
L	32	26	1
N	38	76	10
H	5	28	26

Brahmaputra

o \ s	L	N	H
L	6	6	0
N	9	29	7
H	2	7	4

Mississippi

o \ s	L	N	H
L	83	37	0
N	34	181	34
H	2	27	91

Mackenzie

o \ s	L	N	H
L	24	28	0
N	19	73	10
H	3	32	17

Ganges

o \ s	L	N	H
L	18	4	2
N	18	31	11
H	2	8	14

Rhine

o \ s	L	N	H
L	59	24	0
N	25	131	25
H	1	24	59

Nile

o \ s	L	N	H
L	61	49	10
N	57	133	57
H	11	48	61

Volga

o \ s	L	N	H
L	51	19	2
N	38	93	14
H	2	26	43

Indus

o \ s	L	N	H
L	12	11	4
N	25	32	14
H	2	11	15

Zambezi

o \ s	L	N	H
L	0	9	3
N	1	14	9
H	1	5	6

Lena

o \ s	L	N	H
L	26	39	6
N	14	103	28
H	2	29	41

Niger

o \ s	L	N	H
L	11	40	15
N	6	72	52
H	2	25	39

Danube

o \ s	L	N	H
L	92	35	3
N	34	182	38
H	2	38	90

Mekong

o \ s	L	N	H
L	41	36	7
N	24	119	43
H	7	27	49

Table 2.4 Categorical contingency tables for 90th and 10th percentiles.

o: observed, s: simulated, L: low flow, N: normal flow, H: high flow.

Amazon

o \ s	L	N	H
L	18	18	0
N	16	228	20
H	0	19	17

Parana

o \ s	L	N	H
L	21	15	0
N	25	291	8
H	0	12	24

Murray

o \ s	L	N	H
L	6	17	0
N	17	115	13
H	1	12	11

Yellow River

o \ s	L	N	H
L	9	27	0
N	17	253	18
H	0	15	21

Congo

o \ s	L	N	H
L	4	31	1
N	7	214	19
H	0	12	24

Yangtze

o \ s	L	N	H
L	20	16	0
N	17	277	7
H	0	13	22

Orange River

o \ s	L	N	H
L	13	11	0
N	15	174	5
H	1	15	8

Brahmaputra

o \ s	L	N	H
L	3	6	0
N	3	45	4
H	0	8	1

Mississippi

o \ s	L	N	H
L	25	23	0
N	18	360	15
H	0	15	33

Mackenzie

o \ s	L	N	H
L	4	19	0
N	10	149	2
H	0	12	10

Ganges

o \ s	L	N	H
L	6	5	1
N	10	70	4
H	0	6	6

Rhine

o \ s	L	N	H
L	18	16	0
N	17	252	9
H	0	13	23

Nile

o \ s	L	N	H
L	16	29	3
N	31	332	28
H	1	31	16

Volga

o \ s	L	N	H
L	12	10	0
N	24	213	5
H	0	13	11

Indus

o \ s	L	N	H
L	1	11	0
N	9	84	9
H	0	11	1

Zambezi

o \ s	L	N	H
L	0	0	0
N	0	35	13
H	0	0	0

Lena

o \ s	L	N	H
L	3	21	0
N	8	211	21
H	1	17	6

Niger

o \ s	L	N	H
L	0	22	2
N	4	181	29
H	1	14	9

Danube

o \ s	L	N	H
L	25	23	0
N	23	373	22
H	0	22	26

Mekong

o \ s	L	N	H
L	15	20	1
N	8	250	23
H	0	25	11

Table 2.5 Gerrity skill scores in reproducing anomalous flows for 75th and 25th percentiles; and 90th and 10th percentiles.

Basin	GS-75/25	GS-90/10	Basin	GS-75/25	GS-90/10
Amazon	0.47	0.43	Murray	0.33	0.27
Congo	0.40	0.34	Orange River	0.34	0.39
Mississippi	0.63	0.57	Ganges	0.47	0.42
Nile	0.32	0.26	Indus	0.21	0.01
Lena	0.35	0.13	Danube	0.60	0.48
Parana	0.58	0.58	Yellow River	0.39	0.36
Yangtze	0.67	0.56	Brahmaputra	0.25	0.16
Mackenzie	0.29	0.28	Rhine	0.61	0.54
Volga	0.53	0.45	Zambezi	0.07	n.a.
Niger	0.15	0.12	Mekong	0.39	0.31

2.3.3 Measuring the skill in reproducing floods and droughts

Floods and droughts are regarded as simple binary events defined as exceedances of threshold discharges. For some rivers a monthly time scale may seem to be too coarse to correctly predict flood sizes. However, when we limit ourselves to forecasting monthly flows in terms of binary events, these will certainly be indicative for increased probability of floods for large rivers. It can be seen in Appendix 2.A that at gauging station Lobith on the Rhine, throughout the years with available records during the period from 1815 to 2008, extreme daily discharges almost always coincide with large monthly discharges. When the annual maxima of daily discharge are plotted against the monthly mean discharge of the month in which this daily maximum occurred, resulting points cluster along a straight line (see Fig. 2.A1), with daily maxima higher than monthly mean values as would be expected. Moreover, Fig. 2.A2 shows that for most of the years, the month in which the annual maximum daily discharge occurred is also the month of maximum monthly flow or directly precedes or succeeds this month. Since the Rhine is the smallest of the 20 global rivers in this study, and given the fact that it has a rather complex regime, one can infer that the same assumption holds for other larger basins as well.

Table 2.6 Binary contingency tables for floods and droughts

o:observed, s:simulated

Amazon

Flood

o \ s	Yes	no
yes	4	5
no	5	322

Drought

o \ s	yes	no
yes	4	6
no	3	323

Parana

Flood

o \ s	yes	no
yes	11	6
no	7	372

Drought

o \ s	yes	no
yes	0	17
no	13	366

Congo

Flood

o \ s	Yes	no
yes	3	6
no	3	300

Drought

o \ s	yes	no
yes	0	5
no	10	297

Yangtze

Flood

o \ s	yes	no
yes	4	0
no	2	366

Drought

o \ s	yes	no
yes	5	5
no	2	360

Mississippi

Flood

o \ s	Yes	no
yes	7	3
no	3	476

Drought

o \ s	yes	no
yes	7	11
no	11	460

Mackenzie

Flood

o \ s	yes	no
yes	1	0
no	3	202

Drought

o \ s	yes	no
yes	0	4
no	7	195

Nile

Flood

o \ s	Yes	no
yes	5	6
no	8	468

Drought

o \ s	yes	no
yes	1	49
no	11	426

Volga

Flood

o \ s	yes	no
yes	2	1
no	3	282

Drought

o \ s	yes	no
yes	2	6
no	4	276

Lena

Flood

o \ s	Yes	no
yes	2	4
no	3	279

Drought

o \ s	yes	no
yes	0	1
no	5	282

Niger

Flood

o \ s	yes	no
yes	1	6
no	3	252

Drought

o \ s	yes	no
yes	0	31
no	6	225

Flood Drought

Murray

o \ s	**Yes**	**no**
yes	5	9
no	4	174

o \ s	**yes**	**no**
yes	2	25
no	2	163

Orange River

o \ s	**Yes**	**no**
yes	1	1
no	4	236

o \ s	**yes**	**no**
yes	4	13
no	12	213

Ganges

o \ s	**Yes**	**no**
yes	1	2
no	2	103

o \ s	**yes**	**no**
yes	0	3
no	2	103

Indus

o \ s	**Yes**	**no**
yes	1	2
no	1	122

o \ s	**yes**	**no**
yes	0	15
no	2	109

Danube

o \ s	**Yes**	**no**
yes	7	7
no	6	494

o \ s	**yes**	**no**
yes	4	7
no	6	497

Flood Drought

Yellow River

o \ s	**yes**	**no**
yes	6	6
no	3	345

o \ s	**yes**	**no**
yes	2	6
no	7	345

Brahmaputra

o \ s	**yes**	**no**
yes	1	0
no	0	69

o \ s	**yes**	**no**
yes	1	2
no	0	67

Rhine

o \ s	**yes**	**no**
yes	3	2
no	3	340

o \ s	**yes**	**no**
yes	5	9
no	6	328

Zambezi

o \ s	**yes**	**no**
yes	0	0
no	1	47

o \ s	**yes**	**no**
yes	0	0
no	1	47

Mekong

o \ s	**yes**	**no**
yes	2	3
no	4	344

o \ s	**yes**	**no**
yes	2	6
no	9	336

Decision thresholds for a basin may be defined using various hydrological and economic criteria. A comprehensive approach with verification over the full range of possible thresholds for each basin is beyond the scope of this study. Therefore, a single set of decision thresholds for floods and droughts common for all river basins is selected that can reasonably distinguish between the usual and extreme states of each basin. The flood and drought thresholds used in this study correspond to 5-yr return periods for each river. The discharges corresponding to the 5-yr flood and drought events have been derived using the Annual Maximum Series method.

The choice of 5-yr return periods for floods as well as droughts is made on the basis of two considerations. On the one hand, events with return periods of a few years do not reflect the long-term variability, and do not represent unusually extreme states of a river. On the other hand, the limited availability of discharge observations does not allow the estimation of rare events beyond a fraction of the record length. Five years in this case appears to be a practical return period for the assessment of model skill in reproducing both types of hydrological extremes observed in all basins, the record lengths for which are given in Table 2.1. For the two basins with the longest records, i.e., the Danube and the Mississippi, we repeat the analysis for return periods of ten years.

Similar to the approach used in the construction of categorical tables described in Sect. 2.3.2, for the construction of binary tables, the thresholds for observations and simulations are identified separately in order to decrease the effect of any systematic under- or over-estimation. The skill in simulating flood and drought events is assessed by constructing 2 × 2 contingency tables and applying binary skill scores. Binary contingency tables present the 2 × 2 possible combinations of simulated and observed event outcomes: hit, false alarm, miss and correct rejection.

Equitable skill scores used in the verification of binary forecasts are Heidke skill score (HSS) (Heidke, 1926), Peirce's skill score (PSS) (Haansen and Kuipers, 1965), Gilbert's skill score (GSS) (Schaefer, 1990) and odds ratio skill score (ORSS) (Stephenson, 2000). As stated in Sect. 3.2, the criterion of equitability is based on the principle that random forecasts or constant forecasts of the same single category receive a no-skill score (Murphy and Daan, 1985). Two of these four equitable scores, namely HSS and GSS, are markedly dependent on sample climate. Sample climate, defined as the sample estimate of the unconditional probability of occurrence of an event, is purely a characteristic of the observations with no direct relevance to skill assessment (Mason, 2003). Since dependence on sample climate makes a skill score unjustifiably sensitive to variations in observed climate and therefore unreliable, HSS and GSS are excluded in this study. The remaining two equitable scores PSS and ORSS are independent of the sample climate and recommended in several studies (McBride and Ebert, 2000;

Stephenson, 2000; Göber et al., 2004). However, ORSS is also excluded because the presence of zero in any cell of the contingency table renders this skill score inappropriate (Livezey, 2003). PSS is preferred to other scores in this study on the basis of these considerations.

The possible values of PSS are within the range (−1, 1) and its true zero-skill value is 0. Negative values imply less skill than a random prediction. The PSS for floods and droughts for each basin are calculated in terms of cell counts of the relevant contingency tables according to the formula:

$$PSS = \frac{a}{a+c} - \frac{b}{b+d}$$

where a, b, c and d represent the cell counts for each of the possible outcomes of hit, false alarm, miss and correct rejection respectively.

Table 2.7 Peirce's skill scores for floods and droughts.

Basin	**PSS-f**	**PSS-d**	**Basin**	**PSS-f**	**PSS-d**
Amazon	0.44	0.40	Murray	0.36	0.07
Congo	0.33	0.00	Orange River	0.50	0.24
Mississippi	0.70	0.39	Ganges	0.33	0.00
Nile	0.45	0.02	Indus	0.33	0.00
Lena	0.33	0.00	Danube	0.50	0.36
Parana	0.65	0.00	Yellow River	0.50	0.25
Yangtze	1.00	0.50	Brahmaputra	1.00	0.33
Mackenzie	1.00	0.00	Rhine	0.60	0.36
Volga	0.67	0.25	Zambezi	n.a.	n.a.
Niger	0.14	0.00	Mekong	0.40	0.25

2.3.4 Measuring added skill over a simple water balance estimate

In order to demonstrate the added value of running a complex hydrological model over a simple estimation of the water balance, the MSESS (non-bias corrected), GS and PSS are applied on an alternative set of monthly discharge values at the outlet of each basin. These discharge values are computed as follows: monthly actual evapotranspiration (E) is subtracted from the precipitation (P) on a monthly basis, then aggregated over the drainage network including downstream losses due to open water evaporation to obtain the instantaneous monthly discharge. This estimate of P-E incorporates the same information from the climatic forcing, but ignores hydrological information on stores and fluxes that lead to temporal and spatial redistribution. Skill comparison of model results with this estimate shows the added value of the routing and hydrology, while both suffer from the same poor climatological forcing.

2.4 Results and discussion

2.4.1 Skill in reproducing hydrographs

The results of the historical simulation and observed discharge time series for the selected rivers are presented in Fig.2.2 for visual inspection. The difference between the simulations and observations can be attributed to several errors such as those in the meteorological forcing, discharge records, model parameters, or simplifying assumptions. The possible model errors are discussed in depth in Van Beek and Bierkens (2009) and Van Beek et al. (2011).

Three groups of rivers present a large discrepancy between the simulations and observations. The first group is the Arctic rivers, such as the Lena and Mackenzie, and snow and glacier dominated rivers such as the Indus. Undercatch in the CRU snowfall amounts reported by Fiedler and Döll (2007) results in a large underestimation of the spring discharge after the start of snowmelt. The second group consists of those basins with heavy regulation and large amounts of withdrawal for irrigation and consumption, such as the Murray, Zambezi and Parana. The routing scheme in the current version of PCR-GLOBWB simulates natural discharge and does not include reservoir operations and withdrawals. Therefore, the simulated natural flow on these heavily regulated rivers is in disagreement with the measured discharge. Although it is one of the most heavily regulated rivers, the Nile does not show this discrepancy since measurements of natural flow upstream of the High Dam is available for comparison. The last group consists of rivers in the tropics, which show either overestimation as in Africa, or underestimation as in the Amazon. This is mostly attributable to the low station coverage over the tropics in the CRU dataset and to a lesser extent poor precipitation forecasts in ERA-40 (Troccoli and Kalberg, 2004).

The improvement in predictive skill due to the correction of bias can be seen on the discharge time series before and after the bias correction (Figs. 2.2 and 2.3), as well as the reliability diagrams (Fig. 2.4). It can be observed from these figures that bias correction highly improves the results. This improvement is documented quantitatively in Table 2, which shows the MSE skill scores for the selected basins, both before and after the bias correction. Table 2 shows that without a bias correction, the MSESS for the majority of basins are negative. The improvement in the MSESS due to the correction varies widely, but is quite high in general, yielding a skill higher than the climatology for most basins. The three basins where the highest skill is observed are the Yangtze, the Rhine and the Mississippi, with MSESS above 0.70. The model performs worse than the climatology in four basins. It is interesting to note that the three basins with the worst performance, namely the Niger, the Nile, and the Congo are all African rivers. The fourth basin with negative skill is the Amazon. The relatively low skill in the Amazon and other monsoon-dominated basins such as the Indus and the Mekong can be explained to a certain degree by the fact that for such basins the climatology is already a good estimate of the expected discharge, so that it is difficult to perform better than that. The relatively high values of R^2 and NS for these basins, which are also presented in Table 2, indicate that the model performance is not poor in monsoon-dominated basins, provided that it is evaluated using measures independent of the climatology.

2.4.2 Skill in reproducing anomalous flows

A complete summary of the joint distribution of categorical simulations and observations for the selected basins is presented in 3 × 3 contingency tables (Tables 2.3 and 2.4). These tables provide the basis for the calculation of the Gerrity Scores for each basin. As can be seen in Table 2.5, all the resulting values of GS are positive, indicating that the model has skill in reproducing categorical events. In general, GS values are higher for reproducing the 75th and the 25th percentile flows than for the 90th and the 10th, as the skill is expected to decrease for more extreme flow.

The same three rivers with the highest skill in simulating exact discharges, namely the Yangtze, the Rhine and the Mississippi, have again the highest scores for categorical events. The model performance in categorical simulations for the African rivers the Niger, the Nile, and the Congo is much better than in reproducing hydrographs. The lowest skill among all the basins is observed for another African river, the Zambezi, though still above the climatology. For the Amazon, where the skill in reproducing hydrographs is less than that of the climatology, we observe that the skill in reproducing anomalous flows is rather high compared to other basins. This shows that even in cases where the model simulations are biased and do not outperform the climatology in reproducing hydrographs, the skill in reproducing anomalous flows can be relatively high.

2.4.3 Skill in reproducing floods and droughts

The 2 × 2 contingency tables for flood and drought events for the selected basins can be seen in Table 6. The PSS calculated on the basis of these tables are presented in Table 2.7. The resulting PSS show that the skill obtained by binary forecasts of 5-yr floods and droughts is also higher than an unskilled forecasting system. The system has a markedly higher skill in forecasting floods compared to droughts. Model structure and process descriptions explain the difference in skill in reproducing floods and droughts. Floods are largely controlled by the rapid response of basins and thus react almost directly to the above-average rainfall of the forcing depending on the antecedent conditions. In contrast, droughts or low flows represent the response of the hydrological system to prolonged periods of below-average rainfall. As such, they are more sensitive to the uncertainty in model parameterization affecting processes such as the build-up of soil moisture deficit, the depletion of the groundwater system by baseflow and the regulation of discharge by reservoirs or changed withdrawal. With respect to baseflow, PCR-GLOBWB contains a conceptual model to describe the influence of lithology and drainage density. This model is parameterized using global datasets but not calibrated. As a consequence, it can resolve the general trend but not all local variations. Moreover, the simulated discharge in this study is the natural one and regulation and consumption are not considered. All in all, this makes droughts more sensitive to model uncertainty, all the more so as the rank order of these events can be less accurately assessed due to the relatively larger variability of this phenomenon.

There are no basins where the model has a negative skill score in reproducing either floods or droughts; but for seven basins, the PSS indicates no skill in reproducing droughts. This is because the PSS takes on the value of zero when the contingency table shows no hits. For some basins, the model demonstrates perfect skill in reproducing floods. This is a shortcoming of the skill score used. The score takes on the value of one in cases where there are either no misses or no false alarms. Yet, to be able to assign perfect skill, one would expect the number of both misses and false alarms to be zero.

The skill assessment in reproducing 5-yr events is not applicable to the Zambezi for which the available discharge record only covers four years (see Table 2.1). For this basin, PSS is undefined due to the absence of any observed event. The short length of the observed discharge records affects the assessment of skill negatively for the Brahmaputra (five years and ten months) and the Ganges (nine years; Table 2.1).

For the two basins with the longest records, i.e., the Danube and the Mississippi, we have repeated the analysis for return periods of ten years. The results, which are presented in Appendix C, show that for both basins, PSS in floods decrease when the return period increases, as expected. For the Mississippi, the PSS in reproducing 10-yr droughts is surprisingly slightly higher than in 5-yr droughts. For the Danube, the PSS in 10-yr droughts is zero since there are no hits on the contingency tables.

Notwithstanding the problems related to limited observation lengths, skill in reproducing flood and drought events is demonstrated.

2.4.4 Added skill over a simple water balance estimate

The added value of running a complex hydrological model over a simple estimation of the water balance is demonstrated by comparison of the skill scores MSESS (non-bias-corrected), GS and PSS for model simulated discharges and for the P-E estimate. Skill scores for both the model results and for the P-E estimate are presented in Appendix 2.B.

The results show that model skill by far exceeds that of the P-E estimate in all cases. Skill comparison of model results with this estimate shows the added value of the routing and hydrology, while both suffer from the same poor climatological forcing. In contrast, the monthly climatology of observed discharge performs better than the P-E estimate as it is more attuned to the actual climate, save for its anomalies, as well as the regulation.

2.5 Conclusions and recommendations

As an initial step in assessing the prospect of global hydrological forecasting, we tested the ability of a global hydrological model PCR-GLOBWB in reproducing the occurrence of past extremes in the monthly discharge of 20 large rivers of the world We assessed the model skill in three ways: first in simulating hydrographs, second in reproducing monthly anomalies and third in reproducing flood and drought events. The advantage of such a procedure is that it provides a more detailed assessment of forecasting skill and an insight into which types of forecasting are more promising.

Verification of non-bias-corrected hydrographs reflects model and forcing errors, thus providing the opportunity for improvement. In addition, it allows comparison with the results of other studies which use non-bias-corrected data. Eliminating the systematic bias due to model errors or forcing, on the other hand, provides an indication of the maximum skill that can be achieved in operational forecasting. Simulations with PCR-GLOBWB are biased for most basins, and the skill in reproducing hydrographs is lower than the observed climatology. The model skill improves significantly after a post-processing bias correction and surpasses the observed climatology in most basins.

Results of the analysis indicate that the skill obtained in reproducing monthly anomalies using non-bias-corrected data is higher than the climatology for all basins. The model also has skill in reproducing floods and droughts, with a markedly better performance in the case of floods. The model skill surpasses that of a simple water balance estimate in all cases. Although simulated hydrographs may be biased and do not always outperform the observed climatology even after bias correction, higher skills can be attained in forecasting the occurrence of monthly anomalies as well as floods. The prospects for operational forecasting of monthly hydrological extremes are thus positive. PCR-GLOBWB is similar to other GHMs in model structure and parameterization; and the forcing data is similar to those used in simulations with other GHMs and LSMs. The performance of PCR-GLOBWB in reproducing runoff is comparable to those of other GHMs (Sperna Weiland et al., 2010; Wada et al., 2008) and to LSMs (Sperna Weiland et al., 2011). Given these similarities we argue that our conclusion is valid for other comparable GHMs and LSMs as well.

This assessment in retrospect is a preliminary one and it shows a potential skill given the current GHM, with a meteorological forcing based on observations. The true skill should be assessed in forecasting mode using meteorological forecasts subject to uncertainty from numerical weather prediction (NWP) models.

2.6 Appendix 2.A: Correlation between annual maxima of daily and monthly discharges at gauging station Lobith on the Rhine

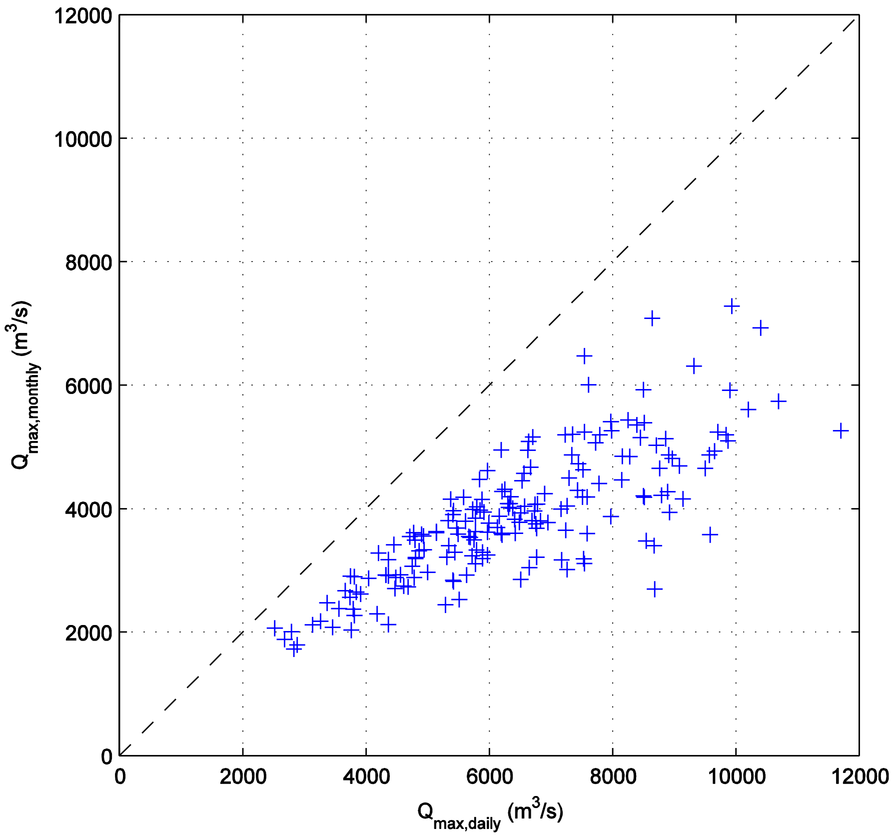


Figure 2.A.1 Annual maxima of daily discharge vs. corresponding monthly mean flows.

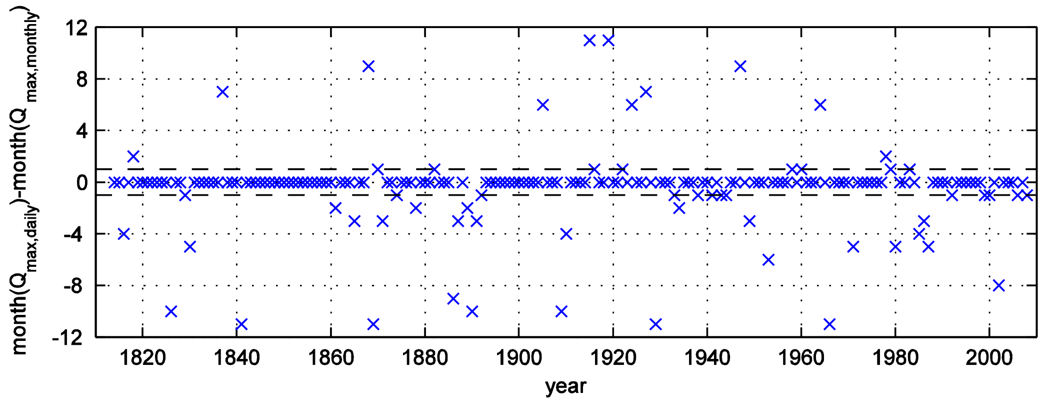


Figure 2.A.2 The difference between the month in which the annual maximum daily discharge occurred and the month of maximum monthly flow.

2.7 Appendix 2.B: Skill comparison between routed streamflow and estimated flow based on water balance

Table 2.B.1 Skill comparison of model results of routed streamflow and streamflow estimates based on P-E fields from the water balance.

Basin	MSESS		GS		PSSf		PSSd	
	model	estimate	model	estimate	model	estimate	model	estimate
Amazon	-4.92	-21.03	0.47	0.18	0.44	0.66	0.40	0.00
Congo	-3.83	-50.09	0.40	0.19	0.33	0.00	0.00	0.00
Mississippi	0.40	-6.69	0.63	0.11	0.70	0.00	0.39	0.14
Nile	-31.51	-75474.70	0.32	0.02	0.45	n.a.	0.02	0.00
Lena	-7.81	-13.21	0.35	0.02	0.33	n.a.	0.00	0.03
Parana	-2.10	-19.80	0.58	0.15	0.65	0.22	0.00	0.00
Yangtze	-0.89	-4.35	0.67	0.23	1.00	0.33	0.50	0.00
Mackenzie	-10.51	-12285.40	0.29	0.04	1.00	0.00	0.00	0.00
Volga	-0.81	-30.34	0.53	-0.01	0.67	0.00	0.25	0.03
Niger	-81.30	-696.49	0.15	0.05	0.14	0.00	0.00	0.03
Murray	-0.70	-13.63	0.33	0.04	0.36	0.00	0.07	0.00
Orange River	0.11	-2.58	0.34	0.08	0.50	0.00	0.24	0.01
Ganges	0.33	-14.04	0.47	0.06	0.33	n.a.	0.00	0.00
Indus	-1.63	-3.26	0.21	-0.03	0.33	0.00	0.00	0.00
Danube	-0.04	-15.17	0.60	0.13	0.50	0.00	0.36	0.02
Yellow River	-1.98	-32.76	0.39	0.11	0.50	0.33	0.25	0.01
Brahmaputra	-1.40	-2.25	0.25	0.12	1.00	n.a.	0.33	n.a.
Rhine	0.57	-2.40	0.61	0.35	0.60	1.00	0.36	0.00
Zambezi	-1.49	-17.34	0.07	0.04	n.a.	n.a.	n.a.	n.a.
Mekong	-0.61	-8.85	0.39	0.19	0.40	0.00	0.25	0.07

2.8 Appendix 2.C: Comparison of skill in reproducing 5-yr and 10-yr floods and droughts for the Mississipi and the Danube

Table 2.C.1 Binary contingency tables and PSS for the Mississippi

5-yr floods			10-yr floods			5-yr droughts			10-yr droughts		
o \ s	Yes	no	o \ s	yes	no	o \ s	yes	no	o \ s	yes	no
yes	7	3	yes	3	2	yes	7	11	yes	4	5
no	3	476	no	2	482	no	11	460	no	6	474
PSS= 0.70			PSS= 0.60			PSS= 0.39			PSS= 0.44		

Table 2C.1 Binary contingency tables and PSS for the Danube

5-yr floods			10-yr floods			5-yr droughts			10-yr droughts		
o \ s	Yes	no	o \ s	yes	no	o \ s	yes	no	o \ s	yes	no
yes	7	7	yes	3	4	yes	4	7	yes	0	5
no	6	494	no	4	503	no	6	497	no	5	504
PSS= 0.50			PSS= 0.43			PSS= 0.36			PSS= 0.00		

Chapter 3
Skill of a global seasonal streamflow forecasting system, relative roles of initial conditions and meteorological forcing

Abstract

We investigate the relative contributions of initial conditions (ICs) and meteorological forcing (MF) to the skill of the global seasonal streamflow forecasting system FEWS-World, using the global hydrological model PCRaster Global Water Balance. Potential improvement in forecasting skill through better climate prediction or by better estimation of ICs through data assimilation depends on the relative importance of these sources of uncertainty. We use the Ensemble Streamflow Prediction (ESP) and reverse ESP (revESP) procedure to explore the impact of both sources of uncertainty at 78 stations on large global basins for lead times up to 6 months. We compare the ESP and revESP forecast ensembles with retrospective model simulations driven by meteorological observations. For each location, we determine the critical lead time after which the importance of ICs is surpassed by that of MF. We analyse these results in the context of prevailing hydroclimatic conditions for larger basins. This analysis suggests that in some basins forecast skill may be improved by better estimation of initial hydrologic states through data assimilation; whereas in others skill improvement depends on better climate prediction. For arctic and snow fed rivers, forecasts of high flows may benefit from assimilation of snow and ice data. In some snow fed basins where the onset of melting is highly sensitive to temperature changes, forecast skill depends on better climate prediction. In monsoonal basins, the variability of the monsoon dominates forecasting skill, except for those where snow and ice contribute to streamflow. In large basins, initial surface water and groundwater states are important sources of skill.

3.1 Introduction

Forecasting of water availability and scarcity is a prerequisite for managing the risks and opportunities caused by the interannual variability of streamflow. Reliable seasonal streamflow forecasts are necessary to prepare for an appropriate response in disaster relief, management of hydropower reservoirs, water supply, agriculture, and navigation. Seasonal hydrological forecasting on a global scale could be valuable, especially for developing regions of the world, where effective hydrological forecasting systems are scarce. Furthermore, global seasonal forecasts may provide spatially consistent predictions of streamflow anomalies. These may provide information to disaster management organizations operating at global scale to prepare for response, as well as to the energy market about the regional availability of hydropower in the coming months.

Several studies demonstrated the capability of global hydrological models to predict streamflow, such as the WaterGap ((Alcamo et al., 2003; Döll et al., 2003), LaD (The Land Dynamics Model) (Milly and Schmakin, 2002), VIC (Variable Infiltration Capacity Model) (Nijssen et al.,2001), WBM (Water Balance Model) (Vörösmarty et al., 2000; Fekete et al., 2002), Macro-PDM (Probability Distributed Moisture Model) (Arnell, 1999, 2004), and PCRaster Global Water Balance (PCR-GLOBWB) (Sperna-Weiland et al., 2010; Van Beek et al., 2011). Candogan Yossef et al. (2012) assessed the skill of the global hydrological model PCR-GLOBWB in reproducing past discharge extremes in 20 large rivers of the world, as a first step toward developing and assessing a global seasonal hydrological forecasting system. This preliminary assessment in hindcast quantified skill using a meteorological forcing (MF) based on observations and concluded that the prospects for seasonal forecasting with PCR-GLOBWB or comparable models are positive. Note that this study did not include actual forecasts. Thus, the meteorological forcing errors due to uncertainty from numerical weather prediction models were not assessed.

In an actual forecasting setup, the predictive skill of a hydrological forecasting system is affected by errors in model structure and parameterization, MF, and initial conditions (ICs), most importantly soil moisture, groundwater and snow. Skill of seasonal hydrological forecasts can thus be improved on the one hand by better prediction of future climate and on the other hand by better estimation of initial hydrologic states through assimilation of independent hydrological observations such as soil moisture and snow data from earth observation. The improvement in the overall predictability that may be attained depends on the relative importance of these two sources of uncertainty, which varies considerably according to location, season and lead time (Bierkens and van den Hurk, 2007; Bierkens and van Beek, 2009; Shukla and Lettenmaier, 2011; Shukla et al., 2011). Therefore, determining the role of each factor is helpful in deciding which skill improvement methods are more promising (Paiva et al., 2012).

The theoretical framework for quantifying the contributions of boundary forcing and initial conditions to predictability was developed in atmospheric sciences by Collins and Allen (2002). Wood and Lettenmaier (2008) and Wood et al. (2002, 2005) adopted this approach in hydrological forecasting. They presented an Ensemble Streamflow

Prediction (ESP) and reverse Ensemble Streamflow Prediction (revESP) approach and evaluated the relative roles of MF and ICs in seasonal hydrologic prediction in two western US basins. The ESP/revESP framework contrasts the forecast variance arising from a forecast ensemble based on perturbations of the initial states, and the forecast variance arising from an ensemble of meteorological forcing, to the internal, climatological variance. Li et al. (2009) used a similar approach for the Ohio River basin and the southeastern US to investigate the varying roles of ICs and MF in seasonal hydrologic forecasting. The ESP/revESP approach was applied to seasonal forecasts in the US by Shukla and Lettenmaier (2011). Shukla et al. (2011) used the same approach to evaluate cumulative run-off and soil moisture forecasts on a global scale. Paiva et al. (2012) applied the ESP/revESP approach to the Amazon River basin.

In this study, we apply the ESP/revESP approach on a global scale to examine the relative contributions of ICs and MF to the skill of seasonal streamflow forecasts. We investigate the roles of both sources of uncertainty in the skill of the global seasonal streamflow forecasting system FEWS-World, using the global hydrological model PCR-GLOBWB. FEWS-World has been setup within the European Commission 7th Framework Programme project Global Water Scarcity Information Service (GLOWASIS). The assessment is based on the ESP/revESP procedure outlined by Wood and Lettenmaier (2008). We simulate global monthly streamflow with lead times ranging from 1 to 6 months for a historical period of 30 years (1981–2010). We analyze the impact of both sources of uncertainty at 78 stations on large river basins across the globe, for all the months of the year and for lead times up to 6 months. These 78 stations have previously been selected for analysis within the GLOWASIS project to represent different hydroclimatic conditions and all continents, but in the same time considering the availability of discharge data. In this study, we compare the ESP and revESP forecast ensembles with retrospective model simulations driven by meteorological observations, and not with direct hydrological observations. The advantage of comparing against simulations instead of observations is that in the former case model errors are eliminated and predictability is related only to knowledge of ICs and the uncertainty in future MF. Note that in previous work (Candogan Yossef et al., 2012) we investigated the model errors of the forecasting system.

The remaining part of this paper is set up as follows. Section 3.2 describes the global seasonal hydrological forecasting system, FEWS-World, the global hydrological model PCR-GLOBWB and the meteorological forcing data. Section 3.3 describes the hydrological simulations and the skill measures. Results are presented in section 3.4, followed by discussion in section 3.5 and conclusions in the last section.

3.2 Materials and methods

3.2.1 Global hydrological forecasting system FEWS-World

FEWS-World is a global hydrological forecasting system configured within the forecasting environment Delft-FEWS (Flood Early Warning System). Delft-FEWS is an open shell for managing, data handling and guiding of forecasting processes (Werner et al., 2013). It is used by a large number of operational forecasting centres and agencies around the world for various purposes such as forecasting hydrological storm surges, river flows, reservoir management and water quality. FEWS-World has been built as part of the GLOWASIS project. The FEWS-World system consists of a Master Controller, a Postgres database and 18 forecasting shells (i.e., computational cores) for efficient handling of ensemble forecasts and data processing. Within FEWS-World several workflows have been setup for running the global hydrological model PCR-GLOBWB sing the precipitation, temperature and potential evaporation fields from the ERA Interim-Global Precipitation Climatology Project (GPCP) data set (Balsamo et al., 2010). Further descriptions of the meteorological forcings are given in section 3.2.2.

PCR-GLOBWB simulates the terrestrial part of the global water cycle (van Beek et al., 2011; van Beek and Bierkens, 2009). It is coded in the high-level computer language PCRaster for constructing environmental models (Wesseling et al., 1996). The model is fully distributed and operates on a regular grid with a cell size of 0.5 x 0.5 ^{o} on a daily time step. Meteorological forcing is assumed to be constant over the grid cell. Subgrid variability of hydrological processes is taken into account in the representation of short and tall vegetation, open water, different soil types, saturated area, surface runoff, interflow and groundwater discharge.

PCR-GLOBWB calculates the water balance for every grid cell by tracking the transfer of water between the atmosphere and the cell, through stores within each cell, and laterally, as discharge, from one cell to the downstream neighbour. The model calculates the storages and fluxes of water, simulates the generation of runoff and its propagation as discharge through the river network. Precipitation falls either as snow or rain depending on atmospheric temperature. It can be intercepted by vegetation and added to the finite canopy storage, which is subject to open water evaporation. Snow is accumulated when the temperature is lower than 0 ^{o} C and melts when it is higher. Snow melt is added to rain and throughfall; it is either stored in the available pore space in the snow cover, or it infiltrates into the top soil layer. Part of this water is transformed into surface runoff and the remainder infiltrates into the soil through two vertically stacked soil layers and an underlying groundwater layer. Water is exchanged between these layers following Darcy's law and the resulting soil moisture is subject to evapotranspiration. The remaining water contributes to lateral drainage as interflow from the soil layers or base flow from the groundwater reservoir. The total drainage, consisting of surface runoff, interflow and base flow is routed through the drainage network of rivers, lakes, wetlands and reservoirs based on DDM30 (Döll and Lehner, 2002), using the kinematic wave approach. An extensive description of PCR-GLOBWB can be found in Van Beek and Bierkens (2009).

3.2.2 Meteorological forcing data

The meteorological variables required to force PCR-GLOBWB are daily values of precipitation, evapotranspiration and temperature. In the absence of direct estimates of actual evapotranspiration, the model can be forced with values of potential evapotranspiration calculated from temperature, radiation, cloud cover, vapor pressure and wind speed.

We force PCR-GLOBWB with the ERA Interim-GPCP data set (Balsamo et al., 2010). This is a global meteorological data set, which is a combination of the ERA Interim reanalysis (Dee et al., 2011) and GPCP monthly rainfall observations (Huffman and Bolvin, 2011; Huffman et al., 2009). ERA-Interim is the latest global atmospheric reanalysis produced by the European Centre for Medium-Range Weather Forecasts. It is an ''interim'' reanalysis initially started from year 1989; later extended back to the year 1979; and continues to be updated forward in time. ERA-Interim reanalysis was produced as a part of the next-generation extended reanalysis intended to replace ERA-40. The GPCP is part of the Global Energy and Water Cycle Experiment of the World Climate Research program. The GPCP provides global precipitation estimates by merging infrared and microwave satellite estimates with rain gauge data from more than 6000 stations.

Potential evaporation has been estimated from ERA-Interim as well. We estimated the monthly values of potential evaporation by application of the Penman-Monteith equation (Monteith, 1981; Penman, 1948) for a reference grass canopy, according to the Food and Agriculture Organization methodology (Allen et al., 1998). Reference potential evaporation is multiplied by a monthly crop factor to obtain land cover specific potential evaporation

3.2.3 Hydrological simulations

PCR-GLOBWB is run at a daily time step to produce ESP/revESP forecast ensembles as well as the control simulation. The model spin-up is carried out over the period 1979–1984 using ERA-Interim-GPCP data set. Subsequently, the hydrological states at the end of this 5 year spin-up are used as initial states for the 30 year historical simulation covering the historical period from 1981 to2010, with an extra 2 years spin-up period from 1979 to 1981.

The control run is a single simulation covering the whole 30 years period. Daily discharge values are aggregated into monthly totals. Monthly aggregation provides a more appropriate forecast at the seasonal scale and a proxy of the underlying distribution. Hydrologic states, as well as monthly discharge totals are saved at the end of each month. These states are used for both running the ESP forecasts and producing resampled ICs for each month of the year for the revESP forecasts.

The ESP and revESP forecast ensembles are produced with ESP and revESP workflows. In the ESP workflow, input ensembles of MF are created from the 32 year input data series (1979–2010). PCR-GLOBWB model runs are initialized on the first day of each month for the period 1981–2010 using the stored ICs. In the revESP workflow, stored ensembles of ICs for 12 x 30 = 360 months and the meteorological input data are used to start a PCR-GLOBWB run on the first day of each month for the period 1981–2010. This results in 360 ESP runs, each run containing 32 members and 360 revESP runs, each run containing 30 members. The runs are carried out in batch using the FEWS-World forecasting system. Each run continues for 6 months ahead and produces an ensemble of monthly discharge values for 6 lead times, from 1 to 6 months.

The skill of ESP forecasts comes from ICs and the ensemble spread is caused by MF uncertainty. RevESP on the other hand, uses an ensemble of ICs resampled from the hydrologic states for the same day of the year and the model is forced by observed MF for the given forecast period. RevESP skill results from the MF and the ensemble spread is caused by uncertainty in ICs.

3.2.4 Measures of skill

We assess the relative contribution of ICs and MF to forecast skill in 78 stations on large global basins. To quantify each contribution to predictability, we use a ratio of variances framework that compares the skill obtained from the ESP and revESP with the skill resulting when climatology is used as an ensemble prediction, which is equivalent to the climatological variance (Wood and Lettenmaier, 2008)

We calculate the mean squared error (MSE) values of the ESP and revESP as well as of the climatology for 78 global locations, for 12 months of the year and for 6 lead times. The MSE values for a given month and lead time are calculated according to the following formulas:

$$MSE[ESP] = \frac{1}{S}\sum_{s=1}^{S}\left[\frac{1}{M}\sum_{m=1}^{M}(h_{sm} - h_{ss})^2\right]$$

$$MSE[revESP] = \frac{1}{M}\sum_{m=1}^{M}\left[\frac{1}{S}\sum_{s=1}^{S}(h_{sm} - h_{mm})^2\right]$$

$$MSE[Cli] = \frac{1}{S}\sum_{s=1}^{S}\left[\frac{1}{S}\sum_{t=1}^{S}(h_{tt} - h_{ss})^2\right]$$

where, $h_{sm}(t)$ = streamflow for initial state s, meteorological forcing m
S = number of years in the historical period
M = number of ensemble members, i.e., number of meteorological forcing values

We then calculate the ratios of MSE of both ESP and revESP to the MSE of climatology for all locations, months and lead times.

When the ratio of the MSE of either forecast ensemble to the MSE of the climatology is equal to one, the forecast skill is equal to that of a climatological forecast. Ratios smaller than one indicate a forecast that is more skillful than the mere climatology; whereas ratios greater than one indicate less skill than the climatology. The ratio approaches zero for a perfect forecast.

When calculating the MSE ratios for a given month and a given lead time, we use the forecast ensembles that predict the total monthly discharge generated during that given month. In other words, we use the discharge ensembles resulting from the simulations which start at time t_0 and end at time t_n, with a lead time of n months, where t_0 is prior to the end of the given forecast month by n months. Thus, for the month of May and for 1 month lead time, n = 1, t_0 is the 1st of May and t_n is the 31st of May. For 2 months lead time, n = 2, t_0 is the 1st of April and t_n is again the 31st of May.

3.3 Results

3.3.1 Results of the hydrological simulations

The results of the 360 ESP runs and 360 revESP runs at each location, each one covering a period of 6 months, are combined in 12 discharge time series, 6 for ESP and 6 for revESP, each 360 months long. These time series represent the chained discharge forecasts at the 6 monthly lead times of interest. To demonstrate the typical behaviour of the ESP and revESP, the discharge time series in the Amazon and the Murray are presented in Figure 3.1. For better visual inspection the results of only the first 24 months period are shown. The ensemble results as well as the results of the control run are presented. Discharge time series for the ESP and revESP runs for all 78 locations can be seen in supporting information Figure 1 in the online supporting information data file 2013WR013487-fs01.doc.

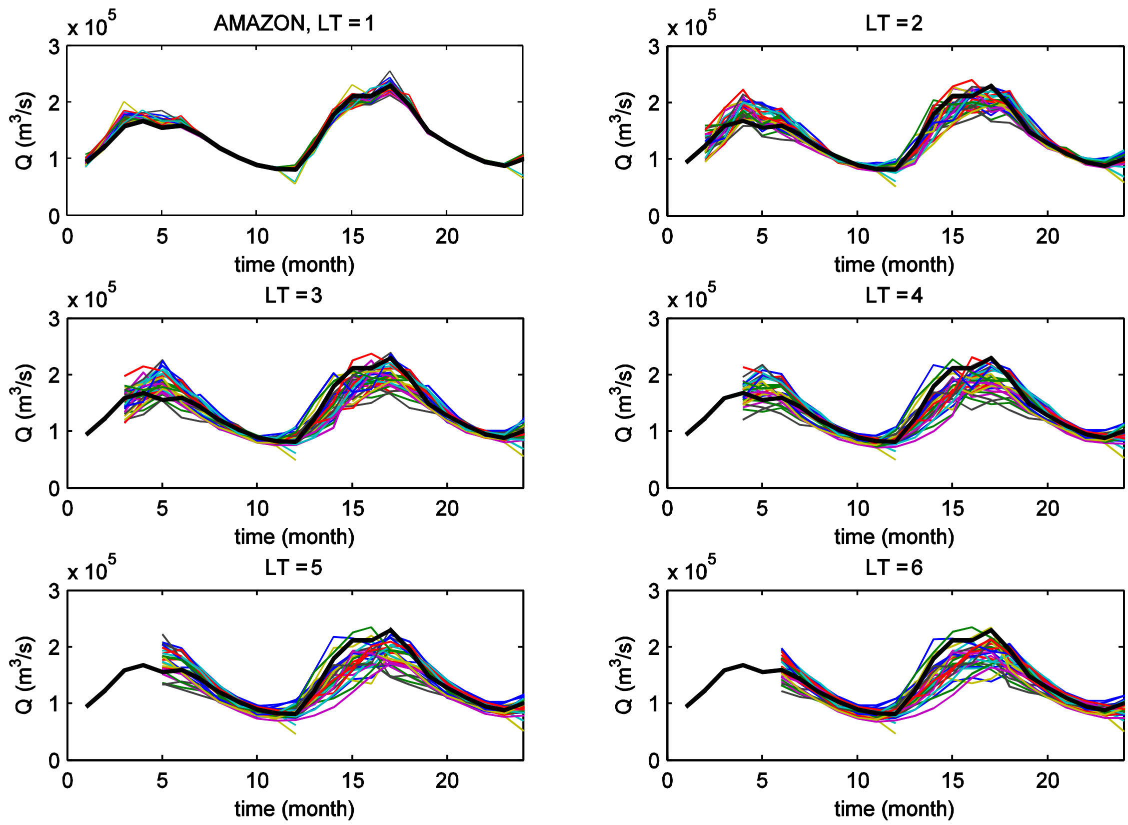


Figure 3.1a Chained discharge time series of ESP for the Amazon.

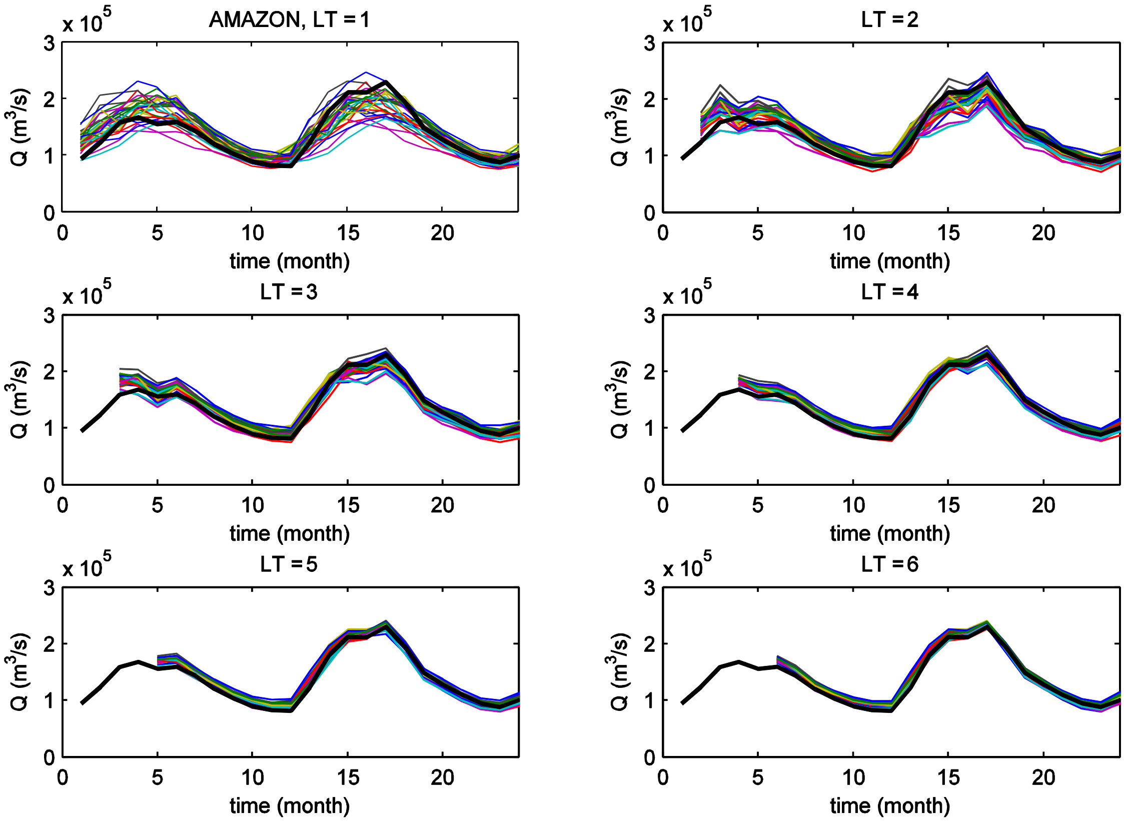


Figure 3.1b Chained discharge time series of reverse ESP for the Amazon.

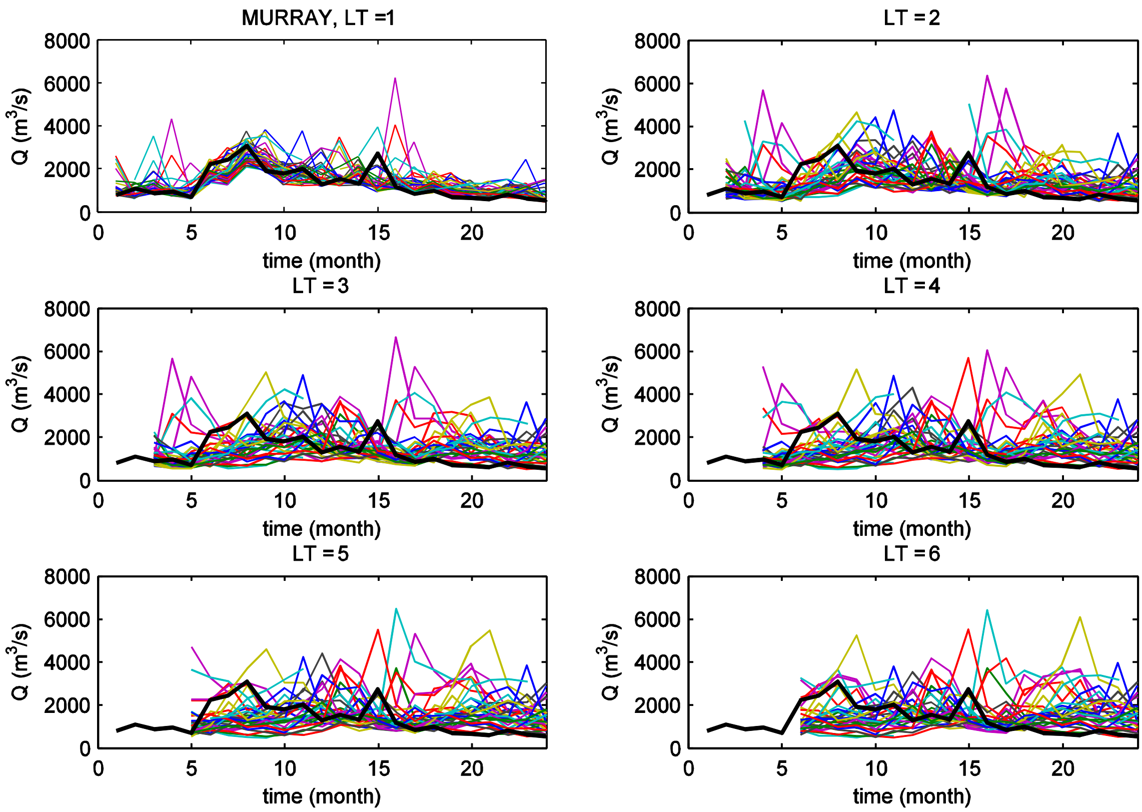


Figure 3.1c Chained discharge time series of ESP for the Murray.

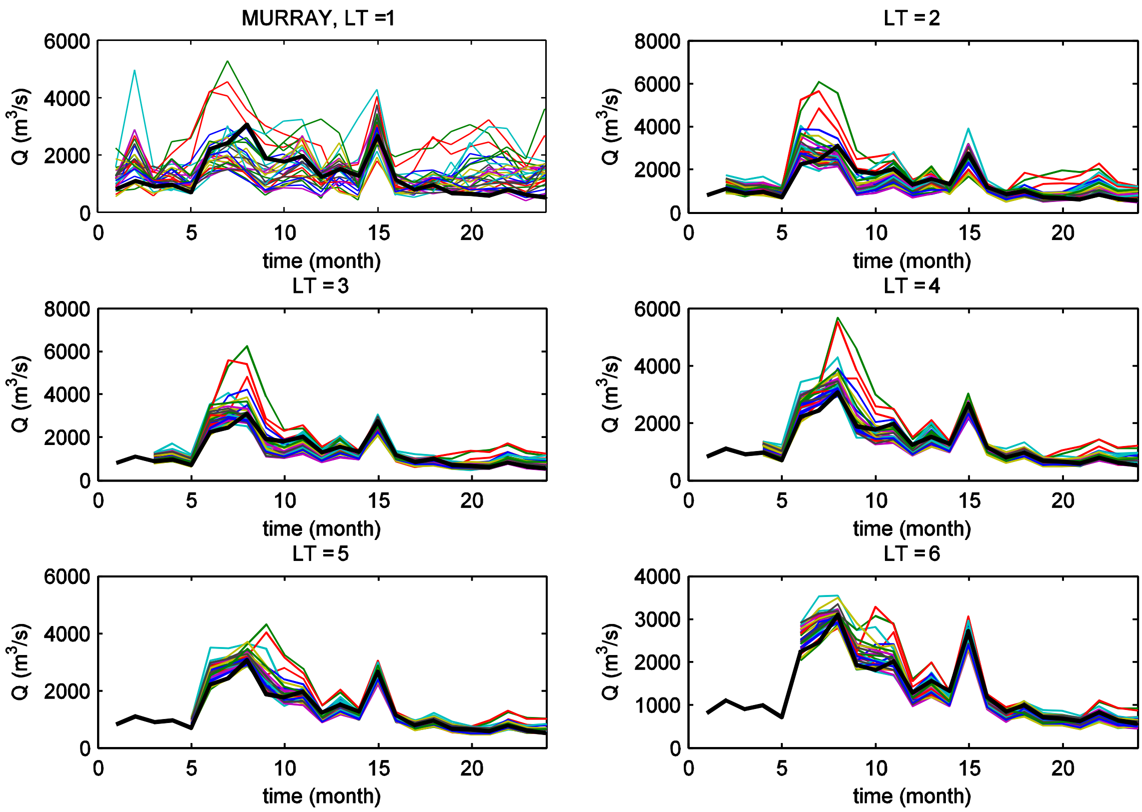


Figure 3.1d Chained discharge time series of reverse ESP for the Murray.

The ESP discharge time series in the Amazon (Figure 3.1a) show that the ensemble spread of the ESP simulations increases with increasing lead time as the relative importance of ICs diminish. For the revESP simulations at the same location (Figure 3.1b), the ensemble spread decreases with increasing lead time, converging toward the control run as the relative role of the MF becomes more dominant.

In the Murray basin the ESP ensemble spreads much more rapidly with increasing lead time (Figure 3.1c) than the discharge series in the Amazon, while the revESP ensemble spread collapses to the value of the control run very rapidly (Figure 3.1d). This means that the expected skill due to IC memory is much smaller in the Murray than in the Amazon, while the expected skill due to MF is much larger. The relative importance of ICs and MF is further discussed in section 3.4.

3.3.2 Skill for ESP and reverse ESP

Ratios of MSE of the ensemble simulations to the MSE of climatology for all 78 locations, 12 months of the year and 6 lead times are documented in the supporting information Table 1 that can be found online in the supporting information data file 2013WR013487-ts01.xls.

As explained in section 3.2.3, the MSE ratios for a given month are calculated using the results of the simulations starting at a prior time t_0, and end at the end of the considered month. Therefore, the month mentioned in the MSE tables denote the month for which the forecast is made, not the month at which the forecast starts.

3.3.3 Changes in skill with increasing lead time

The MSE ratios for the ESP and revESP are plotted on bar charts. As two contrasting examples, the skill charts for the Nile and Ob basins are presented in Figure 3.2. Skill charts for all 78 locations are presented in supporting information Figure 3.2 that can be found online in supporting information data file 2013WR013487-fs02.doc.

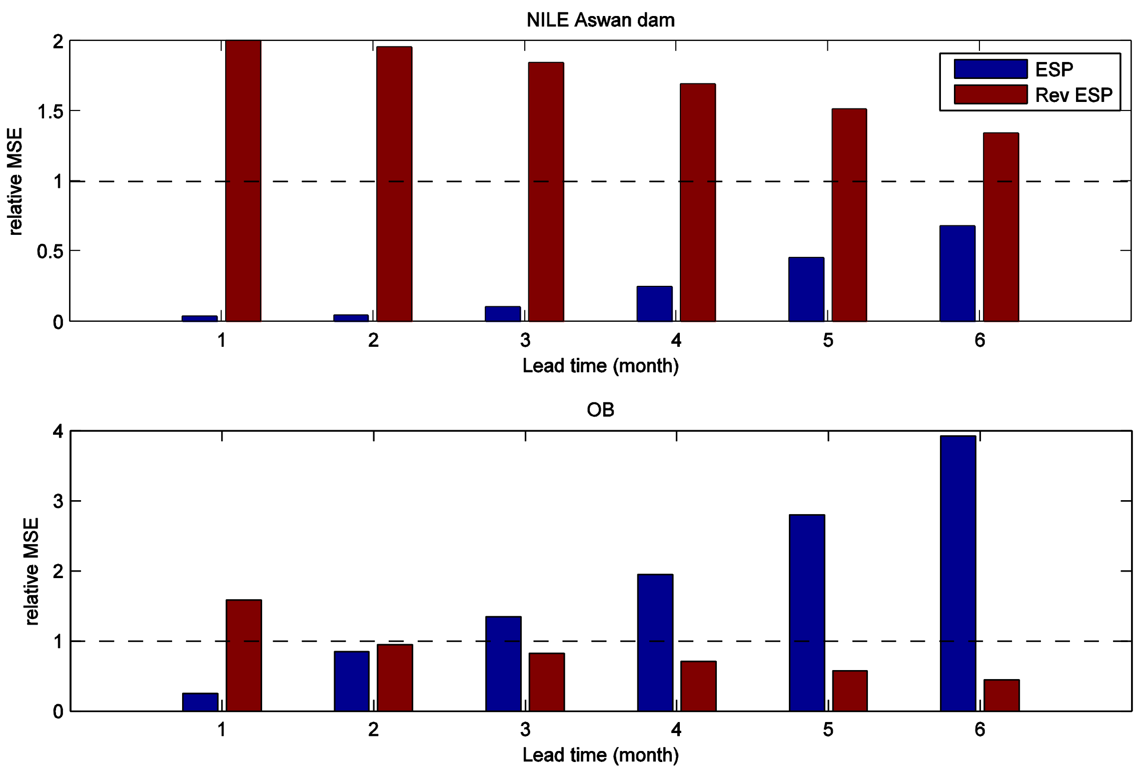


Figure 3.1d ESP and revESP skill charts for the Nile and the Ob.

The charts in Figure 3.2 demonstrate that the MSE ratios of the ESP increase with increasing lead time, while the MSE ratios of the revESP decrease. In the Nile basin, the MSE ratios of the ESP for all lead times are smaller than 1, indicating that the ESP is more skillful than climatology over the full lead time considered. In contrast, in the Ob the ESP is skillful only up to 2 months lead time. As the relative importance of the ICs diminishes with increasing lead time and the importance of MF becomes more dominant, the skill of the revESP increases, hence the MSE ratios decrease. It can be observed that for the Ob, the graphs of the ESP and revESP intersect each other at a point. This point denotes the time within the 6 months lead time range, after which the relative importance of MF surpasses that of ICs. In the Nile however, there is no such intersection point, since the ICs dominate throughout the 6 months range. The reasons for these differences are discussed at global scale in section 3.5.

3.3.4 Seasonal and geographical distribution of skill

Skill maps are prepared to present the seasonal and geographical distribution of skill. The maps indicate skill of the simulations for each month of the year at each location. Separate maps are prepared for the ESP and revESP runs, as well as for each lead time. MSE ratios for each basin are shown on each map for a given lead time and a given month of the year. Maps for the month of January for all lead times are presented in Figure 3.3. Maps for the ESP and revESP runs for all months of the year are presented in supporting information Figure 3 that can be found online in supporting information data file 2013WR013487-fs03.doc. The 78 stations are marked with a circle, coloured according to a skill scale from 0 to 1.5, 0 indicating perfect skill, 1 indicating skill equivalent to that of the climatology, and values above 1 indicating less skill than the climatology. The sizes of the circles reflect basin size.

The ESP maps for January and for all 6 lead times (Figure 3.3a) show in which basins and up to what lead times, forecasts for the given month may be skillful, with the knowledge of ICs only and no information besides past observed climatology on the future MF. The revESP maps (Figure 3.3b) show on the other hand, the forecast skill that may be attained for a given month, with perfect knowledge of future MF, but no information on ICs.

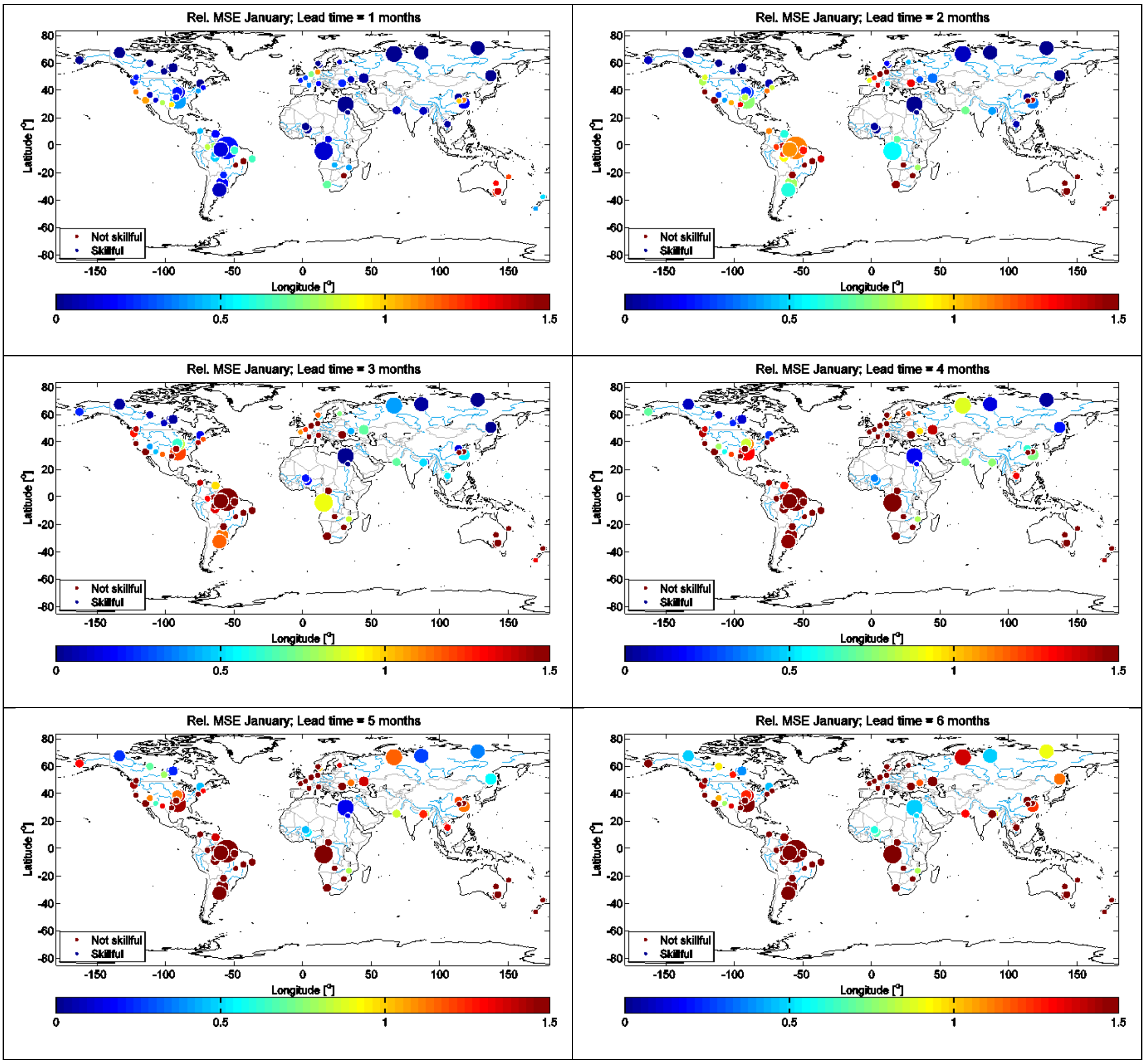


Figure 3.2a World maps indicating ESP skill scores in January

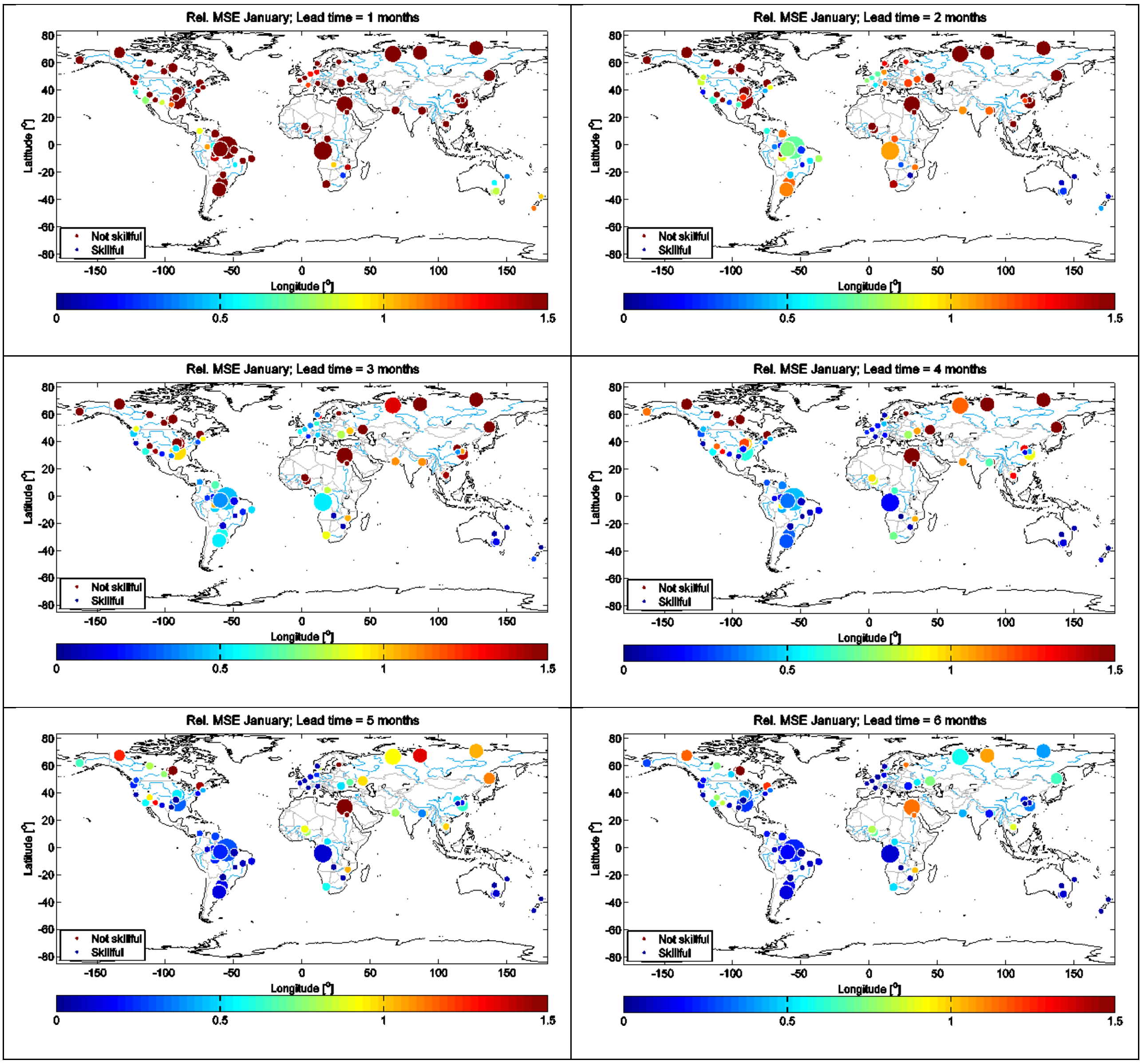
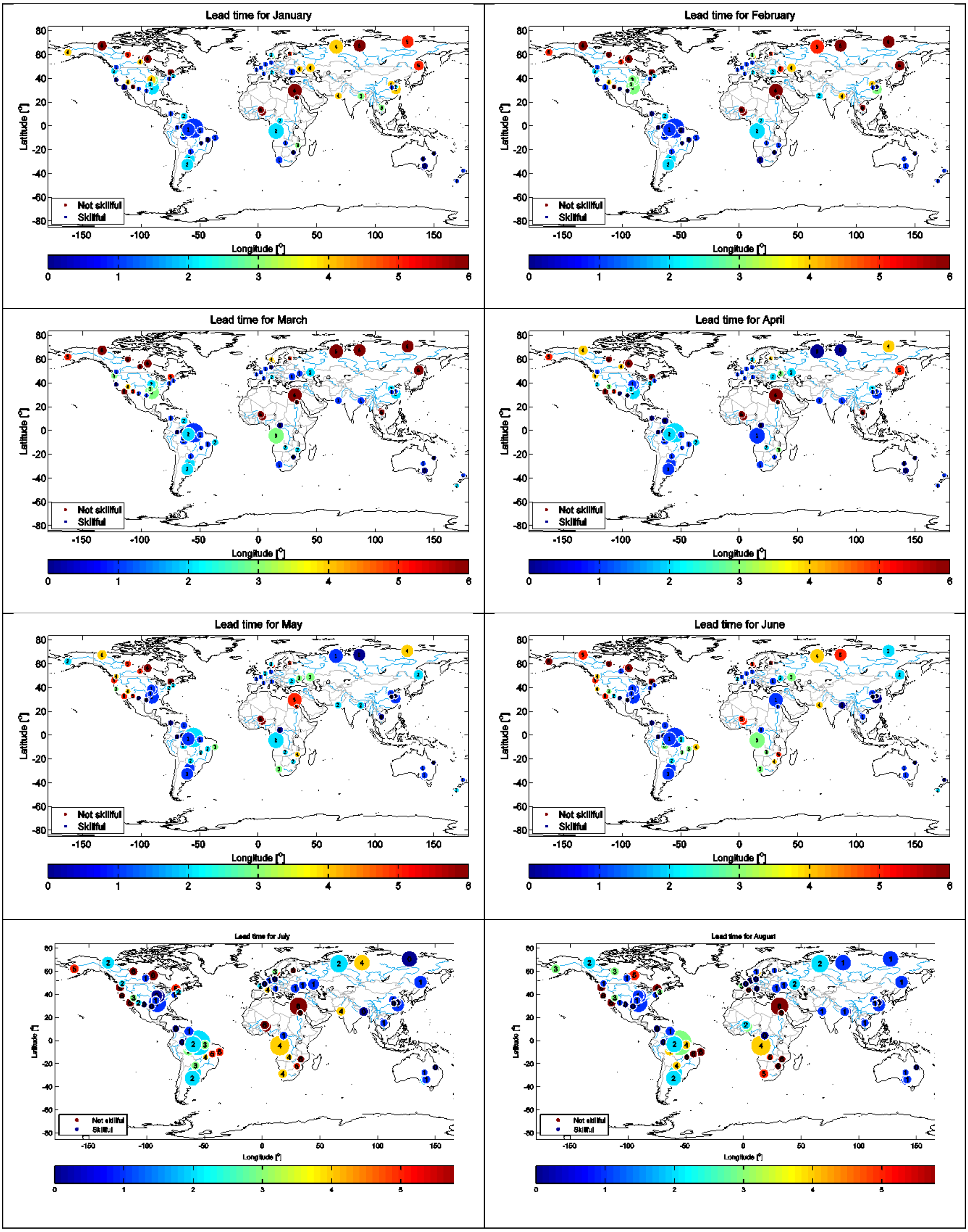


Figure 3.2b World maps indicating reverse ESP skill scores in January


Lead time for January
Lead time for February
Lead time for March
Lead time for April
Lead time for May
Lead time for June
Lead time for July
Lead time for August
Latitude [°]
Longitude [°]
Not skillful
Skillful
80
60
40
20
0
-20
-40
-60
-80
-150
-100
-50
0
50
100
150
0
1
2
3
4
5
6

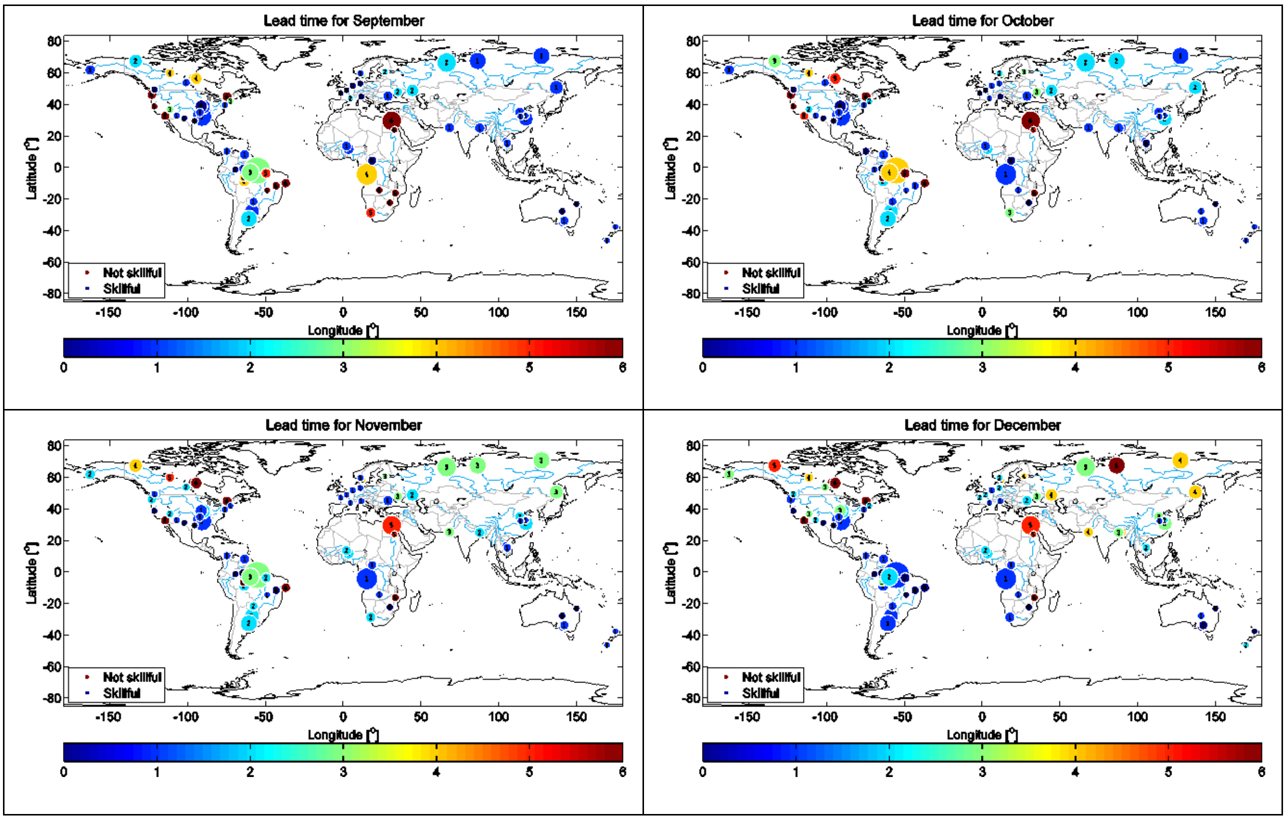


Figure 3.4 World maps indicating the Critical Lead Time (CLT) in months.

3.3.5 Critical lead times

Combining the skill measures of the ESP and revESP simulations, we calculate critical lead times (CLT) for each station and for each month of the year. CLT is defined as the lead time in months after which the importance of ICs is surpassed by that of MF. The CLT are plotted on 12 maps for each month and presented in Figure 3.4.

The maps show that there are some basins, where in certain months of the year, it is not likely to make skillful forecasts on any lead time, with only the knowledge of ICs and none on the future MF. These basins are marked with 0. There are also some basins where without any information on future MF, perfect knowledge of ICs allows forecasts for certain months up to 6 months ahead. It should be noted that CLT is only meaningful when there is skill in ESP and revESP. When both ESP and revESP are of poor skill, the resulting CLT may be high but this does not mean that forecasts are skillful. In most basins, this critical lead time changes significantly with the season. These issues are discussed in section 3.5.

3.4 Discussion of results

Our results show that the relative roles of ICs and MF in hydrological forecasting skill vary both seasonally and geographically. In this section, we discuss the results for several larger basins in the context of prevailing hydroclimatic conditions.

3.4.1 Tropical, monsoon-dominated basins

Results indicate that in the tropical, monsoon-dominated South American basins, such as the Amazon and the Parana, skill due to ICs is limited to 1–2 months lead time during most of the year. Hydrological forecasts beyond 1–2 months are dominated by MF. In the Amazon basin, forecasts for the spring months August until November have a higher CLT of 3–4 months. Hence knowledge of ICs contributes to the skill of forecasts, initiated 3–4 months ahead. This shows that better estimates of ICs during winter months (June, July, and August) may improve the spring forecasts in the Amazon. Our conclusion is in agreement with the conclusions of the study of Paiva et al. (2012) where they demonstrated that in the Amazon River uncertainty on ICs play an important role for hydrological predictability for up to 90 days. Paiva et al. (2012) conclude that ICs are more important especially during high flow conditions (March, April, and May) and recession (June, July, and August). Our results emphasize the importance of ICs during the recession period (June, July, and August), when the increased groundwater recharge plays an important role. Our findings, as well as the findings of Paiva et al. (2012) are in disagreement with Shukla et al. (2011), who found that MF uncertainty dominates the forecasts in the Amazon, even for shorter lead times. The reason for the disagreement may be the importance of routing in large basins such as the Amazon where long travel times are involved. Extensive floodplains store water and release it slowly, so that the outflow has already been in the river for a number of months. Therefore, the knowledge of surface water ICs is an important source of forecast skill. Shukla et al. (2011) studied cumulative runoff, without flow routing, whereas our study as well as the study by Paiva et al. (2012) include routing.

In the large rivers of the Indian subcontinent the relative roles of ICs and MF follow the seasonal patterns of the monsoon. In the Ganges and Brahmaputra, forecast skill is strongly dominated by MF during the monsoon season. At the onset of the monsoon, the ICs are always more or less the same, which reduces the impact of the ICs on model skill. Snowmelt coincides with the monsoon and the contribution of snowmelt to streamflow is very low compared to rainfall. The interannual variability of the monsoon is the determining factor in forecasts during the wet season. Maps in supporting information Figure 3 (online supporting information data file 2013WR013487-fs03.doc) indicate that in the Brahmaputra skill of the ESP is below the climatology from April to September even for lead times of 1 month; hence the forecasts are dominated by the MF for all lead times. In the period from October to March, the knowledge of ICs contributes to skill for lead times up to 6 months. The results for the Ganges show that during the monsoon season the ICs play a more important role than the Brahmaputra. Only June and July forecasts are dominated by the MF for all lead times. Skill derived from ICs is similar to the Brahmaputra for the rest of the year. The hydrographs for the Ganges and Brahmaputra are characterized by double peaks during the wet season. The second peak due to the reverse North-eastern monsoon in June is more pronounced in the Brahmaputra. Whereas the ICs prior to the first peak are generally the same every year, the ICs before the second peak are strongly conditioned by the strength of the monsoon in the first peak. Therefore, ICs are more important for the skill in the second peak. The decreased skill of the revESP for June to September can be seen on the maps in supporting information Figure 3 (online supporting information data file 2013WR013487-fs03.doc).

In the Indus, where snow and ice have a larger contribution to streamflow (Immerzeel et al., 2012), ICs play a relatively more important role. Snow at lower altitudes begins to melt in March to April, whereas glaciers and snow at higher altitudes begin to melt in June. Therefore, June and July forecasts in the Indus are dominated by the ICs for 4 months lead times. The effect of ICs in June on the flow after July is less however because the ice and snowpack at higher altitudes is more or less consistent from year to year. Flow is low during the winter months when most precipitation is stored as snow and ice. The effect of MF is relatively low from November to January, causing a higher CLT of up to 4 months. For the rest of the year forecasts beyond 1 month are dominated by MF.

Our results for the large rivers of China, such as the Yangtze and the Yellow River indicate that hydrologic forecasts are dominated by MF beyond 1 month lead time except for the low flow period in winter. High flows extend from May to October in the Yellow River and from April to September in the Yangtze. Summer flows are monsoon dominated for the most part and the onset of the monsoon coincides with the melting season of snowpack and glaciers, just as for the Indian basins. The variability in MF during these months dominates forecast skill. ICs do not contribute to skill of forecasts for the wet period longer than 1 or 2 months ahead. For the dry period from November through February, the relative importance of MF decreases and the CLT increases to 2, 3, or 4 months lead time.

The results for the Mekong are similar to the Yangtze and the Yellow River. The rainy months in the Mekong are May to October and highest discharges occur from July to October. During the wet season MF dominates the forecasts beyond 1 month lead time. ICs become more important during the dry months. Forecasts for the months February through April derive their skill from ICs of up to 6 months earlier.

3.4.2 Arctic basins

In the North American arctic rivers, such as the McKenzie, Yukon and Nelson, forecasts are dominated by ICs for lead times up to 6 months. The importance of ICs is due to the large memory of these arctic river systems. The temperature in the McKenzie is below freezing for about 300 days during the year. The rainy months are July to September. Discharge is lowest in March, and peaks from May to September. The Yukon basin is frozen for almost half of the year with little or no flow. Discharge peaks in May to June, following snowmelt and declines until November when the basin freezes again. Both snowpack and groundwater play an important role in these arctic rivers. These processes have a long memory, causing the large importance of ICs in forecast skill. The forecasts for the month of March in the McKenzie for instance have a CLT of 6 months, which reduces to 2 months by July. In the lake area the large memory introduced by the lakes continues through the summer months (June, July and August) as well.

The relative importance of ICs and MF in Asian arctic rivers follow seasonal patterns similar to the North American arctic rivers. The Ob, Yenisey, and Lena are ice bound from October to April. Forecasts for the colder months are dominated by ICs for lead times of 5–6 months. Following snowmelt, the discharge peaks in all three rivers in the beginning of June. The results show that the skill in the Yenisey basin derives from ICs for up to 6 months for March forecasts, whereas April forecasts are dominated by MF even for a lead time of 1 month. This sudden increase in the importance of MF can be attributed to the effect of temperature in April that determines the start of the melting season. Similar decrease in CLT occurs in the Ob from 6 months for March forecasts to 0 for April forecasts. During the summer, which is the rainy season, ICs play a less important role in these rivers. In the Lena a CLT of 4 months persist into the forecasts for April and May, which again decreases to 0 for July. The Lena basin has a more continental climate than the Ob and Yenisey, with a later onset of the snowmelt in June. This explains the time lag in Lena compared to Ob and Yenisey.

3.4.3 Temperate regions

European rivers display quite uniform results in the relative roles of ICs and MF in forecast skill. In general MF dominates beyond lead times of 1–2 months throughout the year. The forecasts for the months of July to October are dominated by MF even for 1 month lead time in Western Europe. In the Rhine basin, skill of the ESP forecasts for these months initiated 1 month ahead is below the climatology. The results agree with the fact that forecast skill in the Rhine is limited to a maximum of 2 weeks during the rain dominated season. Therefore, improvement in seasonal hydrological forecasts in these basins depends on better climate forecasts.

In the Danube and Volga, relative skill due to ICs and MF follows the same seasonal pattern as the Rhine but the role of ICs is relatively higher, especially in the Volga where skill due to ICs extends back up to 4 months lead time. Snowmelt dominates the flows in April and May more than it does in the Rhine and groundwater is important during the low flows in winter when most precipitation falls as snow, therefore reducing the importance of MF.

In the western United States Shukla and Lettenmaier (2011) showed that skill due to ICs is high during spring and summer months, mainly June. Our results for the Columbia and Colorado basins agree with the findings of this study. Discharge (80–90%) of the Colorado River originates from snowmelt and the rest from groundwater base flow and summer monsoon rain. Snowmelt begins in April, peaks in May to June, and finishes by July to August. ESP skill maps presented in the online supporting information data file 2013WR013487-fs03.doc demonstrate high skill for up to 6 months lead time in the Colorado basin in spring and summer. Our findings support the conclusions of Shukla and Lettenmaier (2011) that spring and summer forecasts for the western US regions could benefit from improvements in knowledge of ICs during winter and spring months.

For the eastern US, the results of Shukla and Lettenmaier (2011) show that ESP forecasts initialized from December to April are skillful only for 1–2 months lead times. Our results for the St. Lawrence River are in disagreement with these findings. Skill maps presented in supporting information Figure 3 (online supporting information data file 2013WR013487-fs03.doc) indicate that skill due to knowledge of ICs in this basin is higher than the climatology for up to 5–6 months. Peak flow in the St. Lawrence is due to spring and summer snowmelt accompanied by rain. Forecast skill in the spring therefore depends largely on the snowpack accumulated during the previous winter months. The disagreement between our results and those of Shukla and Lettenmaier (2011) is probably due to errors in one or both of the modelling applications in the estimation of snow accumulation. The presence of lakes implies that the routing in our study may also have an effect. The reasons for the different results may be clarified by a further comparative study involving both applications.

For the southeastern rivers such as Mississippi, Missouri and Arkansas our results also agree with those of Shukla and Lettenmaier (2011), showing that skill due to ICs diminishes after 1–2 months lead time. At Vicksburg station on the Mississippi peak flow in spring depends not only on the winter ICs of snowpack, but also on the MF which determines the timing of snowmelt and on rains in the Great Plains and the lower valley. Our results thus confirm the conclusion of Shukla and Lettenmaier (2011) that forecasts in the eastern US would benefit most from improvements in MF throughout the year.

3.4.4 Semi-arid regions

The results for the Australian basins such as the Murray, Darling, Fitzroy and Coopers Creek show that the relative importance of ICs is the lowest in this continent. Streamflow is not seasonal and highly sensitive to precipitation. Year-round rain is highly erratic with large interannual variability. Different precipitation patterns exist for different parts of the Murray basin. In summer, water from subtropical mountainous regions is lost on the way downstream due to high evaporation. Flow is not fed by groundwater and therefore knowledge of ICs does not contribute to the skill. MF dominates the forecast skill even for 1 month lead time throughout the year. Figure 4 shows that only for July forecasts in the Murray CLT is 2 months. However, the skill due to knowledge of ICs for 2 months lead time is below the climatology therefore the CLT is not a meaningful measure in this case. Improvement of hydrological forecasts for Australian basins therefore depends on better climate forecasts.

The interannual variability of rainfall is also high in the semiarid Orange River basin in Africa. The runoff coefficient in this basin is very sensitive to rainfall variability; therefore, the ICs do not contribute much to skill. Rainfall after a very dry period causes high unpredictability in the Zambezi as well. While skill of the ESP for November forecasts is high up to 6 months ahead, skill due to ICs is very low for February and March forecasts, which is the season when large floods occur. This is due to the peak inflows into the reservoir Cahora Bassa in February causing the reservoir to either spill or not spill, which in return depends very sensitively on the MF. The results for the Niger basin show the strongest seasonality in the relative importance of ICs and MF. Knowledge of ICs is the main source of skill for forecasts for the months of January to July initiated up to 6 months ahead. Groundwater outflow is the dominant factor for this season. For the rain dominated period from August to December on the other hand, the importance of ICs is limited to lead times of 1–2 months, after which MF becomes more important. In the Congo where there is no rain during the months of May, June and July, ICs play an important role in the forecasts for the months of July and August for up to 4 months lead time. From October to December MF dominate beyond 1 month, while for the rest of the year, ICs contribute to skill for 2–3 months lead time. The Nile stands out in Africa, with CLTs of 5–6 months. Since both stations on the Nile are downstream of the High Aswan Dam, skill of the ESP forecasts is very high throughout the year, assuming that the release strategy of the reservoir is known.

3.5 Conclusions

We investigated the relative contributions of ICs and MF to the forecasting skill of the global seasonal streamflow forecasting system FEWS-World. Potential improvement in forecasting skill through better climate prediction or by better estimation of initial conditions through data assimilation depends on the relative importance of these two sources of uncertainty. We explored the impact of both sources of forecast uncertainty at large river basins across the globe using the ESP/revESP procedure. Global monthly streamflow was simulated with lead times of 1–6 months for a historical period of 30 years (1981–2010). The ESP and revESP forecast ensembles were compared with retrospective model simulations driven by meteorological observations, thus model errors are eliminated and predictability is related only to knowledge of ICs and the uncertainty in future MF. We compared the variance of the ESP and revESP forecast ensembles to the climatological variance by calculating the ratios of the MSE of both ESP and revESP to the MSE of the climatology for 78 locations, for 12 months of the year and for 6 lead times. We also calculate for each basin and for each month of the year, the CLT after which the importance of ICs is surpassed by that of MF.

Skill maps for the ESP and revESP as well as the CLT values indicate that the contribution of ICs and MF to hydrological forecasting skill varies considerably according to location, season and lead time. We analysed these results in the context of prevailing hydroclimatic conditions for several larger basins. This analysis suggests that in some basins forecast skill may be improved by better estimation of initial hydrologic states through assimilation of snow, soil moisture or surface water data; whereas in others improvement of forecast skill depends on more accurate seasonal climate prediction. The conclusions can be summarized as follows:

1. For arctic rivers as well as for rivers fed by snow and ice from mountainous regions, such as the Volga and Colorado, forecasts of high flows during the melt season depend largely on the ice and snowpack, especially where these have a high interannual variability. These forecasts may thus benefit from assimilation of data on the snow and ice accumulated during the cold season.

2. In some snow fed basins such as the Yenisey and the Mississippi, the onset of ice and/or snowmelt and consequently the timing of peak flow are highly sensitive to temperature changes at the end of the cold season. The importance of ICs diminishes in these cases and improvement of forecast skill in these cases depends more heavily on better climate prediction.

3. In monsoonal basins, the interannual variability of the monsoon is the main factor determining the skill of hydrological forecasts for the wet period. In basins such as the Brahmaputra and the Yangtze where the onset of the thawing of snowpack and

glaciers coincides with the start of the monsoon season, forecasts of high flows are dominated by the MF and skill improvement depends on prediction of the monsoon. ICs play a more important role in basins like the Indus where snow and ice have a larger contribution to streamflow, especially when the ice and snowpack is variable from year to year. Better estimation of initial snow/ice states is likely to improve forecast skill during the wet season.

4. In large basins like the Amazon with extensive flood plains and large travel times of surface water, knowledge of ICs of surface water is an important source of skill for high flow forecasts on lead times of 2–3 months. The role of initial groundwater states also gains importance during the recession stage, when the groundwater discharge plays an important role.

The results of this study show the relative contributions of initial conditions and meteorological forcing to the potential skill of the global seasonal streamflow forecasting system FEWS-World. In an actual forecast, both the ICs and the MF will be uncertain. Therefore, as a next step, the actual forecasting skill of the system should be assessed in a real forecasting mode, using probabilistic seasonal meteorological forecasts and comparing the ESP results to actual discharge observations. Since model error cannot be excluded in actual forecasting mode, the resulting skill should be further improved by bias-correcting the meteorological input as well as by better estimation of ICs through data assimilation where it is indicated by this study to be potentially useful.

Chapter 4
Skill of a global forecasting system in seasonal ensemble streamflow prediction

Abstract

In this study we assess the skill of seasonal streamflow forecasts with the global hydrological forecasting system Flood Early Warning System (FEWS)-World, which has been set up within the European Commission 7th Framework Programme Project Global Water Scarcity Information Service (GLOWASIS). FEWS-World incorporates the distributed global hydrological model PCR-GLOBWB (PCRaster Global Water Balance). We produce ensemble forecasts of monthly discharges for 20 large rivers of the world, with lead times of up to 6 months, forcing the system with bias-corrected seasonal meteorological forecast ensembles from the European Centre for Medium-range Weather Forecasts (ECMWF) and with probabilistic meteorological ensembles obtained following the ESP procedure. Here, the ESP ensembles, which contain no actual information on weather, serve as a benchmark to assess the additional skill that may be obtained using ECMWF seasonal forecasts. We use the Brier skill score (BSS) to quantify the skill of the system in forecasting high and low flows, defined as discharges higher than the 75th and lower than the 25th percentiles for a given month, respectively. We determine the theoretical skill by comparing the results against model simulations and the actual skill in comparison to discharge observations. We calculate the ratios of actual-to-theoretical skill in order to quantify the percentage of the potential skill that is achieved. The results suggest that the performance of ECMWF S3 forecasts is close to that of the ESP forecasts. While better meteorological forecasts could potentially lead to an improvement in hydrological forecasts, this cannot be achieved yet using the ECMWF S3 dataset.

4.1 Introduction

Reliable seasonal streamflow forecasts potentially have many benefits including disaster relief, management of hydropower reservoirs, water supply, agriculture and navigation. Seasonal hydrological forecasting on a global scale could be especially valuable for developing regions, where effective hydrological forecasting systems are scarce. Furthermore, global seasonal forecasts provide spatially consistent predictions of streamflow anomalies. These may supply information to disaster management organizations operating at global scale to prepare for response as well as to the international water and energy markets about the regional availability of water and hydropower in the coming months.

Approaches to seasonal streamflow forecasting can be divided into two categories, empirical/statistical methods and numerical/dynamical methods. Empirical/statistical methods use statistical techniques (e.g., simple correlation, multiple regression, linear or quadratic discriminant analysis, canonical correlation analysis, and neural networks) to find statistically significant relationships between atmospheric/oceanic indicators and river flow on the basis of historical observations. While statistical forecasts are quite successful in some regions of the world and in some seasons, in many cases the available records are too short to accurately capture climatic variability. Moreover, forecasts derived from past climate do not include anthropogenic or other long-term changes in the climate, such as global warming, and statistical methods do not explain the underlying physical mechanisms. Although statistical methods are the more widely developed and reliable methods that are used for most current operational seasonal forecasts, dynamical modelling is thought to hold the greatest potential for future improvement in reliable seasonal streamflow forecasting (Zwiers and von Storch, 2004).

Dynamical model experiments involve the integration of general circulation models (GCMs), which model atmospheric, oceanic and land surface interactions and processes as a set of dynamic equations. Seasonal forecasting by GCMs is based on coupled ocean–atmospheric integrations, where both atmospheric and oceanic components of the Earth's system are taken into account. The main source of predictability for climate forecasting at seasonal scale is the long-term predictability of the oceanic circulation and its large impact on the global atmospheric circulation. The most important cause of seasonal climate variability is the ENSO (El Niño–Southern Oscillation) cycle, which is the large-scale fluctuation of ocean temperatures, rainfall, atmospheric circulation, vertical motion and air pressure centred over the tropical Pacific but affecting other ocean basins as well. Similarly, unusually warm or cold sea surface temperatures (SST) in other tropical oceans, the extent and thickness of snow cover and the amount of soil moisture can have a persistent influence on the atmospheric circulation (Persson and Grazzini, 2007). Due to the chaotic nature of the atmospheric–oceanic system, model runs made with small, random perturbations in the input data may produce a wide range of difference in the output. Therefore, GCMs are run multiple times with slightly different sets of initial conditions, producing a set of output data called an ensemble. The hydrological output from the land surface scheme of a GCM may be used as streamflow forecasts. Alternatively, the meteorological forecast ensemble by a GCM may be used as input to a hydrological model, which produces streamflow forecast ensembles, as we do in this research.

This paper investigates the skill of seasonal streamflow forecasts for 20 of the largest rivers in the world with the global hydrological forecasting system Flood Early Warning System (FEWS)-World, which has been set up within the European Commission 7th Framework Programme Project Global Water Scarcity Information Service (GLOWASIS). These 20 rivers have been selected for analysis to represent different hydroclimatic conditions and all continents. Selected basins can be seen in Fig. 4.1; gauging stations and basin characteristics are summarized in Table 4.1.

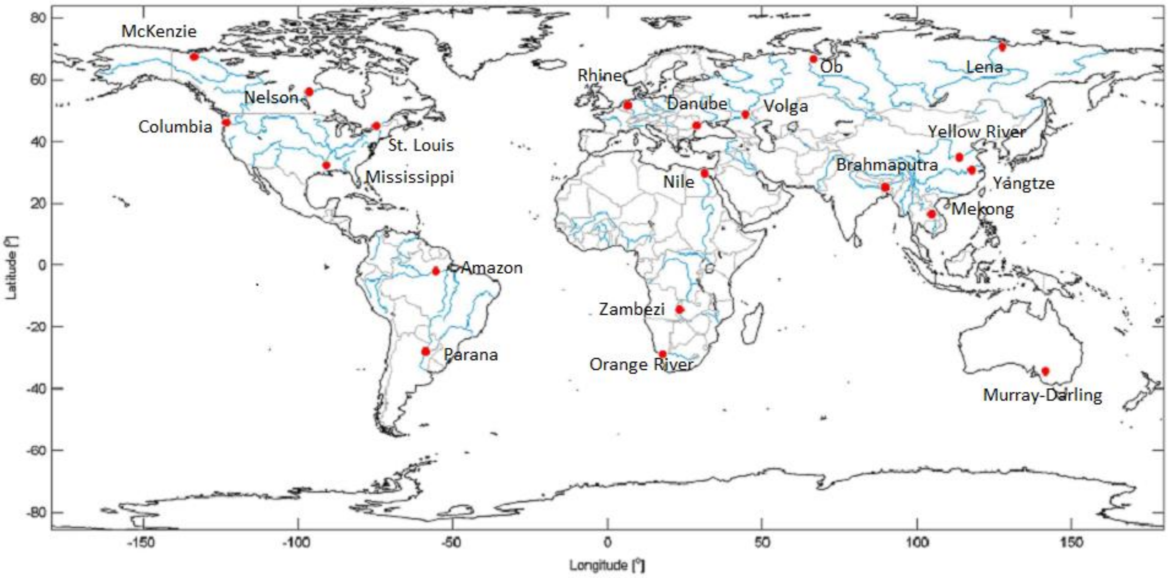


Figure 4.1 Selected basins.

FEWS-World incorporates the global hydrological model PCR-GLOBWB (PCRaster Global Water Balance). The capability of global hydrological models to predict streamflow was demonstrated previously by several studies such as the WaterGap (Alcamo et al., 2003; Döll et al., 2003), LaD (Milly and Schmakin, 2002), VIC (Nijssen et al., 2001), WBM (Vörösmarty et al., 2000; Fekete et al., 2002), Macro-PDM (Arnell, 1999, 2004) and PCR-GLOBWB (Sperna-Weiland et al., 2010; van Beek et al., 2011). Candogan Yossef et al. (2012) assessed the skill of the global hydrological model PCR-GLOBWB in reproducing past discharge extremes for 20 large rivers of the world, as a first step towards developing a global seasonal hydrological forecasting system and assessing its skill. The study quantified skill in deterministic hindcast mode, using the ERA-40 reanalyses by the European Centre for Medium-range Weather Forecasts (ECMWF). This preliminary assessment by Candogan Yossef et al. (2012) concluded that the prospects for seasonal forecasting with PCR-GLOBWB or comparable models are positive. Since actual probabilistic meteorological forecast ensembles were not used, the assessment did not include errors in the meteorological forcing.

Table 4.1 Basin characteristics and gauging stations (GRDC).

Basin	Gauging Station	Area (km^2)	Q_{avg} (m^3/s)
Amazon	Obidos	6,915,000	190,000
Parana	Corientes	2,583,000	18,000
Brahmaputra	Bahadurabad	930,000	48,160
Yangtze	Datong	1,800,000	31,900
Yellow River	Huayuankou	752,000	2,570
Mekong	Muhdahan	795,000	16,000
McKenzie	ArcticRedRiver	1,660,000	9,213
Nelson	Kettle Generating Station	1,060,000	3,447
Ob	Salekhard	2,950,000	12,680
Lena	Kyusur	2,430,000	17,000
Rhine	Rees	65,700	2,200
Danube	Ceatal Izmail	817,000	6,400
Volga	Volgograd Power Plant	1,360,000	8,115
Columbia	Beaver Army Terminal	665,400	6,670
St. Lawrence	Cornwall	774,000	7,367
Mississippi	Vicksburg	2,981,000	12,740
Murray-Darling	Lock 9 Upstream	991,000	257
Orange River	Vioolsdrif	866,500	259
Zambezi	Lukulu	206,530	776
Nile	El Ekhsase	2,900,000	1,251

However, in an actual forecasting setup, the predictive skill of a hydrological forecasting system is affected not only by errors in model structure and parameterization and initial conditions such as soil moisture, groundwater and snow, but also by meteorological forcing errors. Skill of seasonal hydrological forecasts can thus be improved by better meteorological forecasts on the one hand and by better estimation of initial hydrologic states through assimilation of independent hydrological observations on the other hand. The improvement in the overall predictability that may be attained depends on the relative importance of these two sources of uncertainty, which varies considerably among hydrological systems according to location, season and lead time (Bierkens and van den Hurk, 2007; Bierkens and van Beek, 2009; Shukla and Lettenmaier, 2011; Shukla et al., 2011; Yuan et al., 2015). Candogan Yossef et al. (2013) assessed the roles of initial conditions (ICs) and meteorological forcing (MF) in the skill of the global seasonal streamflow forecasting system FEWS-World, based on the ESP/revESP procedure outlined by Wood and Lettenmaier (2008). This study shows the potential for improvement in the skill of streamflow forecasts by a better estimation of IC or a more accurate MF input per region and per time of the year. The current paper aims to assess the total skill of hydrological forecasts, as affected by errors in model structure, in the estimation of IC as well as in the actual meteorological forecasts that are used to force the model.

The remaining part of this paper is set up as follows. Section 4.2 describes the global seasonal hydrological forecasting system, FEWS-World, the global hydrological model PCR-GLOBWB, the meteorological forcing data, the hydrological simulations and the skill assessment. Results are presented in Sect. 4.3, followed by discussion in Sect.4.4 and conclusions in the last section.

4.2 Materials and methods

4.2.1 Global hydrological forecasting system FEWS-World

FEWS-World is a global hydrological forecasting system configured within the forecasting environment Delft-FEWS. Delft-FEWS is an open shell for data handling, managing and guiding forecasting processes (Werner et al., 2013). It is used by a large number of operational forecasting centres and agencies around the world for various purposes such as forecasting hydrological storm surges, river flows, reservoir management and water quality. FEWS-World has been built as part of the GLOWASIS project. The FEWS-World system consists of a master controller, a Postgres database and 18 forecasting shells (i.e., computational cores) for efficient handling of ensemble forecasts and data processing. Within FEWS-World several workflows have been set up for running the global hydrological model PCR-GLOBWB using the precipitation, temperature and potential evaporation fields from the ERA-Interim/Land GPCP-corrected dataset (Balsamo et al., 2015). Further descriptions of the meteorological forcing datasets are given in Sect. 4.2.2.

PCR-GLOBWB simulates the terrestrial part of the global water cycle (van Beek et al., 2011; van Beek and Bierkens, 2009). It is coded in the high-level computer language PCRaster for constructing environmental models (Wesseling et al., 1996). The model is fully distributed and operates on a regular grid with a cell size of 0.5×0.5^{o} on a daily time step. Meteorological forcing is assumed to be constant over the grid cell. Sub-grid variability of hydrological processes is taken into account in the representation of short and tall vegetation, open water, different soil types, saturated area, surface runoff, interflow and groundwater discharge.

PCR-GLOBWB calculates the water balance for every grid cell by tracking the transfer of water between the atmosphere and the cell, through stores within each cell, and laterally as discharge from one cell to the downstream neighbour. The model calculates the storages and fluxes of water, and simulates the generation of runoff and its propagation as discharge through the river network. Precipitation falls either as snow or rain depending on atmospheric temperature. It can be intercepted by vegetation and added to the finite canopy storage, which is subject to open-water evaporation. Snow is accumulated when the temperature is lower than 0ºC and melts when it is higher. Snowmelt is added to rain and throughfall; it is either stored in the available pore space in the snow cover, or it infiltrates into the top soil layer. Part of this water is transformed into surface runoff and the remainder infiltrates into the soil through two vertically stacked soil layers and an underlying groundwater layer. Water is exchanged between these layers following Darcy's law and the resulting soil moisture is subject to evapotranspiration. The remaining water contributes to lateral drainage as interflow from the soil layers or baseflow from the groundwater reservoir. The total drainage, consisting of surface runoff, interflow and baseflow, is routed through the drainage network of rivers, lakes, wetlands and reservoirs, using the kinematic wave approach, based on the global drainage direction map DDM30, which describes the drainage directions of surface water with a spatial resolution of 300 longitude by 300 latitude (Döll and Lehner, 2002). An extensive description of PCR-GLOBWB can be found in van Beek and Bierkens (2009).

4.2.2 Meteorological forcing data

The meteorological variables required to force PCR-GLOBWB are daily values of precipitation, evapotranspiration and temperature. In the absence of direct estimates of actual evapotranspiration, the model can be forced with values of reference potential evapotranspiration, calculated from temperature, radiation, cloud cover, vapor pressure and wind speed.

We force PCR-GLOBWB with two different datasets. The first one is the ERA-Interim/Land dataset (Balsamo et al., 2015). This is a global meteorological dataset, which is a combination of the ERA-Interim reanalysis (Dee et al., 2011) and Global Precipitation Climatology Project (GPCP) monthly rainfall observations (Huffman and Bolvin, 2011; Huffman et al., 2009). ERA-Interim is a robust global atmospheric reanalysis produced by the ECMWF. It is an "interim" reanalysis initially started from the year 1989; later extended back to the year 1979, and continues to be updated forward in time. ERA-Interim reanalysis was produced as a part of the next-generation extended reanalysis intended to replace ERA-40. The GPCP is part of the Global Energy and Water Cycle Experiment (GEWEX) of the World Climate Research program (WCRP). The GPCP provides global precipitation estimates by merging infrared and microwave satellite estimates with rain gauge data from more than 6000 stations. Monthly values of potential evaporation have been estimated from ERA-Interim, using fields of temperature, radiation, cloud cover, vapor pressure and wind speed, by application of the Penman–Monteith equation (Monteith, 1981; Penman, 1948) for a reference grass canopy, according to the FAO methodology (Allen et al., 1998). Reference potential evaporation is multiplied by a monthly crop factor to obtain land cover specific potential evaporation in PCR-GLOBWB.

The second dataset that we use to force the model is the re-forecast ensemble of the System-3 (S3) seasonal forecast archives of the ECMWF covering the period 1981–2010. S3 seasonal forecasts are run in ensemble mode on a fully coupled ocean–atmosphere model. They are run on the first of every month as the initial date, integrated forward for 6 months. Verifications show that the skill of forecasts in regions and seasons known to have a teleconnection with the El Niño is much higher than during neutral conditions. ECMWF seasonal forecast system has been shown to be superior to statistical systems in forecasting the onset of El Niño or La Niña. But once an event has started statistical systems have comparable skill. The dynamical model is also better than the statistical models in forecasting the SST in the Atlantic Ocean and the Indian Ocean. In many parts of the tropics, where changes such as those associated with El Niño can have a large impact on global weather patterns, a substantial part of the year-to-year variation in seasonal-mean rainfall and temperature is predictable. In mid-latitudes, the level of predictability is lower and Europe, in particular, is a difficult area to predict. Seasonal forecasts start to show signs of systematic model errors after about 10 days into the forecast. The ECMWF does not introduce any artificial terms in the equations to reduce the drift. Rather, a daily bias correction based on quantile–quantile transformation is applied on each forecast. In order to account for drift, we applied a bias correction using datasets varying per forecast month. As a result, there are 12 bias correction datasets each with a length equal to a seasonal forecast. The bias correction dataset was provided by the ECMWF (Emanuel Dutra, personal communication, 2015) within the GLOWASIS project. Since November 2011 the seasonal forecast system S4 has become operational to replace S3 with the goal of improving those aspects, where S3 had problems. The improvements brought by S4 include, a next-generation ocean model, a higher spatial resolution, a larger ensemble size. The ensemble number of re-forecasts, which is relevant to our study, was increased from 11 to 15, and the forecasts integrated forward for 7, instead of 6 months. Though there are not many published references on S4 yet, initial studies indicate that there are some improvements in performance over S3, such as higher skill for ENSO forecasts. However, there are also certain aspects where the performance is worse. For instance, S4 suffers from a stronger bias in tropical Pacific SST than S3 (Molteni et al., 2011). Concerning the skill of re-forecast ensembles, an initial report by Norton and Rowlands (2011) compares the skill of 15-member S4 re-forecasts, to the 11-member S3 re-forecasts for the period 1981–2010; and concludes that there is no clear separation in skill between S3 and S4 on seasonal forecast timescales, from month 2 onwards. Therefore, taking into consideration that temperature and precipitation from the S3 re-forecast ensembles were bias corrected, we conclude that S3 is the preferred dataset for our study.

4.2.3 Streamflow forecast runs

PCR-GLOBWB is run at a daily time step to produce two sets of streamflow forecast ensembles, as well as the control simulation run. The first forecast run follows the ESP procedure using the ERA-Interim/Land dataset as basis for the meteorological input. The second forecast run uses actual ECMWF S3 seasonal forecasts as meteorological input.

Model spin-up is carried out over the period 1979–1984 using ERA-Interim/Land dataset. Subsequently, the hydrological states at the end of this 5-year spin-up are used as initial states for the control run. The control run started from these initial states with the ECMWF S3 seasonal forecasts for the period 1979–2010. Daily discharge values are aggregated into monthly totals. Monthly aggregation provides a more appropriate forecast at the seasonal scale and a proxy of the underlying distribution. Hydrologic states, as well as monthly discharge totals, are saved at the end of each month. These states are used as ICs for running the ESP as well as the ECMWF S3 seasonal forecasts.

The ESP forecast ensemble is produced with the ESP workflow within Delft-FEWS. Input ensembles of the meteorological forcing are created from the 32-year input data series (1979–2010). PCR-GLOBWB model runs are initialized on the first day of each month using the stored ICs. In order to avoid any further bias, we excluded the first 2 years and limited the subsequent analysis to the period 1981–2010. This results in 360 ESP runs, each run containing 31 members, excluding the year in question from the 32-year series. The ECMWF S3 streamflow forecast ensemble is produced by forcing the model with bias-corrected meteorological input dataset from the ECMWF S3 seasonal forecast archive, containing 11 ensemble members for each forecast and covering the period 1981–2010. (12 monthly forecasts over the 30-year period result in $30\times12 = 360$ runs, with 11 ensemble members for each run.) Both the ESP and ECMWF S3 runs are carried out in batch using the FEWS-World forecasting system. Each run spans 6 months and produces an ensemble of 11 monthly discharge values for six lead times.

4.2.4 Skill assessment

The Brier skill score (BSS) is commonly used for the skill assessment of meteorological probabilistic forecasts. In order to quantify the added skill obtained by using ECMWF S3 seasonal meteorological forecasts compared to the reference ESP forecast, we employ the BSS, calculated by Eq. (1):

$$BSS = 1 - \frac{BS_{forecast}}{BS_{ref}}$$

The BS values for a given month and lead time are given by Eq. (2):

$$BS = \frac{1}{N}\sum_{t=1}^{N}(p_t - o_t)^2$$

where N is the number of forecasting instances, p is the forecasted probability and o is the observed probability.

The range of the BSS is (−∞, 1) and the best value for a perfect forecast is 1. When the BSS is equal to 0, the forecast skill is equal to that of the reference forecast. Here, a skill of zero or less implies that the seasonal forecasts provide no additional information compared to the random generated climatology of the ESP forecast run. The range of the BS is (0, 1), 0 being the best value for a perfect forecast and 1 the worst.

Besides the BS and its associated skill score BSS, it is possible to use other verification metrics, such as the relative operating characteristic (ROC) score, or the continuous ranked probability skill score (CRPSS) for the skill assessment. We prefer to use the BS and BSS since we would like to assess the skill of our forecasting system in predicting a category of high, low or normal flow for the given month, rather than an exact discharge value, and BS is very suitable for this purpose. BS is the mean squared error of probabilistic forecasts for a given dichotomous event. A probability threshold is used to define the binary event to be observed and forecasted. The BS is a relevant verification metric for analysing the performance of a forecast system for specific categories, defined by a set of thresholds. It is preferred for being a proper score, i.e., being optimized for forecasts that correspond to the best judgment of the forecaster. It is also a highly compressed score; i.e., it directly accounts for forecast probabilities without necessitating a contingency table for each probability threshold (Bartholmes et al., 2009; Ferro, 2007).

In this study, we use two probability thresholds corresponding to the 25th and 75th percentiles for high and low flows, respectively. Values below the 25th percentile of a given month of the year are considered low flows and those above the 75th percentile are considered high flows. The thresholds are calculated separately for forecasted values and observed values. In other words, we classify a forecasted value as high flow if it exceeds the 75th percentile of all forecasted values for the same month of the year and low flow if it is below the 25th percentile. Similarly, an observed value is classified as high flow if it exceeds the 75th percentile of all observed values for the same month of the year and low flow if it is below the 25th percentile. This approach eliminates any systematic bias in the simulations compared to the observations. In this way, we are able to assess the skill in forecasting the occurrence of flows that are higher or lower than usual for a given month.

We calculate the BS and BSS values in 20 large global basins separately for the 12 months of the year and for all six lead times. When calculating the BS for a given month and a given lead time, we use the forecast ensembles that predict the total monthly discharge generated during that given month. In other words, we use the discharge ensembles resulting from the simulations that start at time t_0 and end at time t_n with a lead time of n months, where t_0 is prior to the end of the given forecast month by n months. Thus, for the month of May and for a 1-month lead time, n = 1, t_0 is1 May and t_n is 31 May. For a 2-month lead time, n = 2, t_0 is 1 April and t_n is again the 31 May.

For the ESP approach and the ECMWF S3 seasonal meteorological forecasts, we quantify the theoretical as well as the actual skill. To calculate the theoretical skill, we compare the ESP and ECMWF S3 streamflow forecast ensembles to the results of the control simulation; and for the actual skill we compare them to observed discharge records. The discharge records used are provided by the Global Runoff Data Centre (GRDC) and measured at stations located at the basin outlets. The meteorological datasets used in the calculation of scores are clarified in Table 4.2.

Table 4.2 Meteorological datasets used for calculating BS.

	theoretical (BS_{theo})		actual (BS_{act})	
	forecasted (p)	observed (o)	forecasted (p)	observed (o)
$BS_{forecast}$	ECMWF S3	ERA 40	ECMWF S3	GRDC
BS_{ref}	ESP	ERA 40	ESP	GRDC

4.3 Results

4.3.1 Skill scores

We present the results of the skill assessment in 20 score tables for 20 rivers (Tables S1–S20). The tables are presented in the Supplement. The first eight parts of each table show the BS values for the ECMWF S3 forecast as well as the BSS values, calculated for the four cases of actual and theoretical skill, for low and high flows, i.e., the 25th and the 75th percentiles. Tables present the scores for the 12 months of the year and for six lead times.

The tables are colour coded for easier visual inspection. Values are highlighted in blue where the accuracy of the ECMWF S3 forecasts is considerably higher than that of the ESP forecast, and in yellow where it is considerably lower. Since the best value for BS is 0, higher forecast accuracy corresponds to a lower BS. Where the difference between the BS values of the ECMWF S3 and ESP forecasts are larger or equal to 0.05, the value is highlighted in light blue or light yellow; where it is larger or equal to 0.1, it is highlighted in dark blue or dark yellow. The last two parts of each table show the ratios of the BS_{act} to BS_{theo} of both the ESP and ECMWF S3 forecasts, for the 12 months of the year and six lead times, for low and high flows, respectively.

4.3.2 Overview of the basins with added skill

We provide a global overview of the basins where added skill is obtained using ECMWF S3 meteorological forecast input compared to the ESP input. The locations of improved skill are presented on four world maps for the four cases of actual and theoretical skill, for low and high flows, i.e., the 25^{th} and the 75^{th} percentiles (Fig. 4.2). The maps indicate the number of months per year with skillful forecasts at each location, as well as the maximum lead time for which the skill is retained.

4.4 Discussion of results

In this section, we discuss the results for several larger basins in the context of prevailing hydroclimatic conditions.

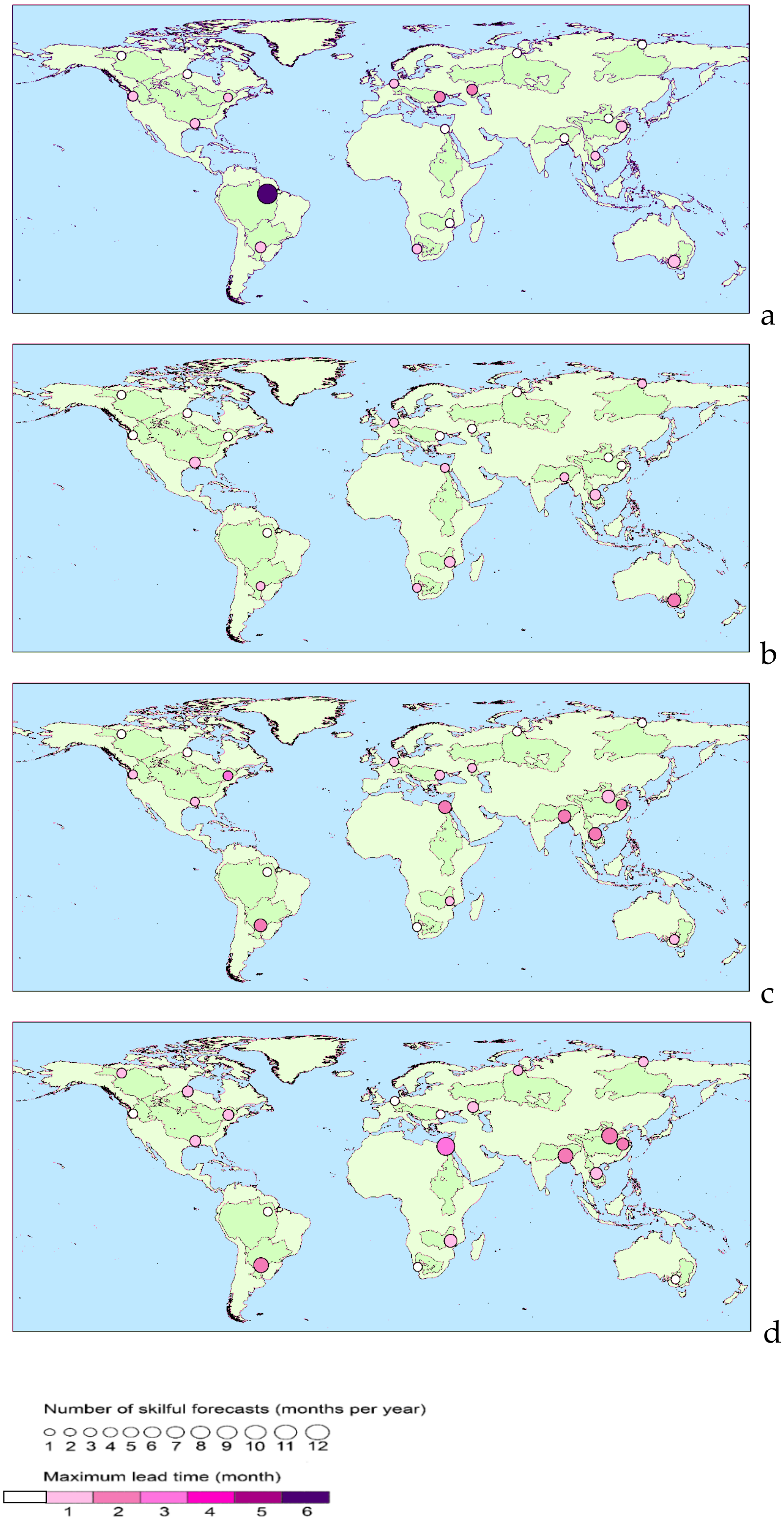


Figure 4.2 Global overview of basins with improved forecast skill.

Panels (a) theoretical skill in low flows, (b) theoretical skill in high flows, (c) actual skill in low flows, (d) actual skill in high flows.

4.4.1 Tropical, monsoon-dominated basins

As can be seen in Fig. 4.2a, results indicate that in the Amazon basin the theoretical skill of the ECMWF S3 forecasts is quite high for predicting lower flows than usual for the given month. In Table S1 for the Amazon, the colour-coded first part, which presents the BS_{theo} for low flow, shows that most of the values are coloured blue. This indicates that the accuracy of ECMWF S3 forecasts are significantly higher than the ESP forecasts; i.e., the difference between the BS values is higher than 0.05. For lead times of 1 and 2 months, the improvement is larger, as can be seen on the first two columns, which are coloured mostly dark blue, indicating a difference between BS values higher than 0.1.

The results for high flows are very different than those for low flows, as can be seen in Fig. 4.2b, as well as the third and fourth parts of Table S1. Most BS values of the ECMWF S3 are very close to the ESP, with only a few yellow highlighted values denoting a worse performance.

The results are also different for the actual skill as can be seen in Fig. 4.2c and d. Both for low and high flows (the fifth to eighth parts of the Table S1), the performance of the ECMWF S3 is either very close to the ESP or lower, as can be seen again by the yellow colour. The average ratio of BS_{act} to BS_{theo} of the ECMWF S3 forecasts over the year and the six lead times is 0.5 in forecasting low flows and 0.57 in high flows (the last two parts of Table S1). These ratios increase with increasing lead time, starting from 0.21 for low flows at a lead time of 1 month, and rising to 0.68 at a lead time of 6 months. There are considerable differences in the ratios between months as well.

Candogan Yossef et al. (2012) showed that hydrological forecasting skill in the Amazon basin is dominated by initial conditions for lead times of 1–2 months, and even up to 4 months for forecasting the discharge during the Southern Hemisphere spring, from August until November. Initial conditions are especially important during high-flow conditions (March, April and May) (Paiva et al., 2012) and the recession period (June, July, August), when the increased groundwater storage plays an important role. Moreover, in large basins such as the Amazon where long travel times are involved, the knowledge of surface water conditions several months ahead is an important source of forecast skill. Meteorological forcing starts to play a more important role beyond 1–2-month lead times throughout the rest of the year. The present study shows, however, that by using ECMWF S3 seasonal forecasts the biggest skill improvement over the ESP procedure can be attained at lead times of 1–2 months, but less at longer lead times when meteorological forcing plays a more important role. For lead times beyond 1–2 months an improvement in skill during most of the year still exists, but it should be noted that this improvement is observed only in the theoretical skill in forecasting low flows.

The results for the other tropical South American basin that we study, the Parana, show a somewhat similar pattern to the Amazon, in the sense that the theoretical skill of ECMWF S3 in forecasting low flows is higher than ESP in some cases, whereas for high flows it is mostly lower (see Table S2). In contrast, the actual skill of ECMWF S3 in forecasting both high and low flows in the Parana is quite different than that in the Amazon. The ratio of actual to theoretical skill of ECMWF S3 forecasts is much lower than that in the Amazon. Averaged over the months of the year and different lead times, it is 0.27 and 0.25 for low and high flows, respectively. Notwithstanding, comparing the actual skill of the ECMWF S3 forecasts to the ESP, we see several months and lead times where the actual skill is significantly improved by using ECMWF S3 forecasts, especially for forecasting high flows at longer lead times and during the first half of the year. For shorter lead times and for the second half of the year however, the actual performance of ECMWF S3 in forecasting high flows is significantly worse than ESP. In forecasting low flows, forecast accuracy is also mostly reduced by using ECMWF S3 forecasts.

Another monsoon-dominated tropical river, the Brahmaputra in the Indian sub-continent, shows a similar pattern to the Parana. In Table S3, we see again a significant improvement in the actual skill for forecasting high flows at longer lead times during the first half of the year. Just like the Parana, forecast accuracy is significantly lower at shorter lead times during the second half of the year. In contrast, the actual skill for forecasting low flows is significantly low at longer lead times, and high at a lead time of 1 month. In theoretical skill, the accuracy of ECMWF S3 re-forecasts in the Brahmaputra for both high and low flows is either very close to that of the ESP or lower. The ratio of the theoretical skill of ECMWF S3 to the actual skill varies considerably for high and low flows, as well as over the year and the range of lead times. The averages are 0.24 and 0.34 for low and high flows, respectively, ranging from as low as 0.2 for low flow forecasts in January to as high as 1.25 for high-flow forecasts in April. The BS values for April high flows at all lead times are higher for actual skill calculations where the forecasted discharges are compared to actual discharge records, than the theoretical skill where they are compared to model simulations. Indeed, it was shown by Candogan Yossef et al. (2012) that the ESP procedure performs worse than the unconditional climatological record of observed flow from April to September even for lead times of 1 month. The forecast skill in the Brahmaputra is strongly dominated by MF during the monsoon season for all lead times. During these months, at a lead time of 1 month, the ECMWF S3 performs significantly worse than the ESP, for the assessment of actual skill. This means the apparent potential for improvement in hydrological forecasts at short lead times by using ECMWF S3 seasonal meteorological forecasts cannot be realized at the moment.

In the two large rivers of China, the Yangtze and the Yellow River, there exists a potential for improving forecasts beyond 1-month lead time through better MF during the high-flow period (see Table S4 and S5). This period extends from May to October in the Yellow River and from April to September in the Yangtze (Candogan Yossef et al., 2012). Our results for the actual skill in forecasting high flows show that this opportunity may be partly realized in both rivers. The added skill of ECMWF S3 over ESP in forecasting higher than usual discharges during the high-flow periods at longer lead times may aid the estimation of increased probability of flooding at lead times of 4–6 months. Moreover, the actual skill of ECMWF S3 is also high in forecasting low flows at short lead times during some months of the highflow periods, especially for the Yellow River. This may help a better estimation of the probability of less than expected discharges during high-flow periods, at 1–2-month lead times. The actual skill of ECMWF S3 forecasts in the Yangtze captures on average 0.23 of the theoretical skill for low flows, and 0.25 for high flows. These numbers are 0.22 and 0.26 in the Yellow River for low and high flows, respectively. In both rivers, for both high and low flows, a significant pattern emerges in the ratios of actual to theoretical skill. The ratios are considerably higher during wet periods than during dry periods.

Similar to the Yellow River and the Yangtze, also in the Mekong basin forecast skill during the wet period from July to October is dominated by MF beyond 1-month lead time. However, the results for the Mekong are different from those for the Chinese basins. Added skill of ECMWF S3 over ESP in forecasting higher than usual discharges during the wet periods can be seen not at longer lead times, but only at a lead time of 1 month, as can be seen in Table S6. This may aid better estimation of flood probability at short notice. Beyond 1 month, the performance of ECMWF S3 forecasts are either worse or not significantly different than ESP. ECMWF S3 forecasts of lower than usual discharges during either the wet or dry periods perform worse than ESP at short lead times, but there are some months of improved skill at long lead times. The ratios of theoretical skill of ECMWF S3 forecasts to the actual skill in the Mekong are 0.37 and 0.60 for low and high flows, respectively. During the high-flow period from July to October, the actual skill in forecasting higher than usual discharges reaches more than 0.80 of the theoretical skill.

4.4.2 Arctic basins

In Arctic basins, snowpack, ice and groundwater processes have a long memory, causing the forecast skill to be dominated by ICs for lead times up to 6 months (Candogan Yossef et al., 2013). The North American Arctic rivers Mackenzie and Nelson, as well as the Asian Ob and Lena are ice bound for a significant part of the year and peak discharges follow snowmelt. The ESP forecasts already perform quite well in these Arctic rivers as would be expected for basin with such a large memory.

Tables S7–S10 show that the ECMWF S3 forecasts for these rivers are not significantly skillful when compared to the ESP. During May–June, which is the beginning of the high-flow season in Arctic rivers, one might expect some improvement in skill with ECMWF S3 forecasts over the ESP due to the temperature effect determining the onset of snowmelt. However, there is no significant increase in the performance of ECMWF S3 forecasts over the ESP forecasts, not even during the beginning of the high-flow season. ECMWF S3 forecasts perform very similar to ESP, and even worse in some cases. Especially the actual skill of ECMWF S3 forecasts in the Arctic basins in Asia is considerably low when compared to the ESP forecasts.

The ratios of actual skill to theoretical skill are not very low in the Arctic basins in general. Low ratios would be expected in areas where the model has large errors associated with snow and glaciers and consequent errors in the timing of peak discharges. In the river Ob for instance, where the discharge peaks in June, the actual skill reaches 0.60–0.70 of the theoretical skill, so it may be concluded that the timing of the model is well approximated.

4.4.3 Temperate regions

The ECMWF S3 forecasts in general do not perform significantly better than ESP in the temperate European basins, such as Rhine, Danube and Volga as can be seen in Tables S11–S13. There are some cases with improvement in the skill in forecasting flows lower than usual, especially in the theoretical skill. However, for high flows the ECMWF S3 forecasts perform worse than the ESP. In the Rhine basin, where improvement in forecast accuracy depends on better climate forecasts, using the ECMWF S3 forecasts does not provide an improvement over the ESP. In the Danube and the Volga, we see an improvement in the theoretical skill in forecasting low flows during winter months. In the Danube and especially the Volga basins snowmelt and groundwater processes play a bigger role than the Rhine. Low flows during winter months are actually dominated by the groundwater processes rather than the meteorological forcing. Nevertheless, this is where we see a consistent improvement in skill by using the ECMWF S3 forecasts. For high flows on the other hand, ECMWF S3 forecasts perform worse, both in their theoretical and actual skill.

The ratios of actual to theoretical skill are in general quite high for the European basins, but lower in temperate basins of North America. In the Columbia River forecasts are dominated by the ICs due to snow and the performance of ESP forecasts is already high. Using ECMWF S3 forecasts does not bring a significant improvement (see Table S14).

In the St. Lawrence River, peak flows are fed by spring and summer snowmelt accompanied by rain. Candogan Yossef et al. (2013) concluded that the forecasting skill in spring and summer months depends largely on the snowpack accumulated during the previous winter months, dominating seasonal forecasts up to 6 months ahead. These findings are in disagreement with the results of Shukla and Lettenmaier (2011), which show that ESP forecasts initialized from December to April are skillful only for 1–2-month lead times. As it was mentioned in Candogan Yossef et al. (2013), the disagreement is probably due to errors in one or both models in the estimation of snow accumulation. The results of the present study confirm the importance of ICs on the one hand. Table S15 shows that the theoretical skill of ECMWF S3 forecasts is considerably low compared to the ESP in the St. Lawrence, especially for forecasting higher flows than usual during the summer months. On the other hand, the actual skill of the ECMWF S3 forecasts in forecasting lower than usual summer flows is significantly high for 2, 3 and 4-month lead times. This finding supports the conclusion of Shukla and Lettenmaier (2011), which emphasizes the importance of MF beyond 1–2-month lead times. Additionally, the fact that the ratio of actual skill to theoretical skill in St. Lawrence is rather on the low side may be an indication of errors in our model in representing the snow processes.

For the southeastern US rivers, the results of Candogan Yossef et al. (2013) as well as those of Shukla and Lettenmaier (2011) show that skill due to ICs diminishes after 1–2- month lead time and that forecasts would benefit most from improvements in MF throughout the year. However, the results of the present study show that in general this potential improvement cannot be realized for the Mississippi by using ECMWF S3 forecasts. The performance of ECMWF S3 forecasts is similar to the ESP in most cases, as can be seen in Table S16, and it is lower than ESP in more case than it is higher, with no apparent pattern.

4.4.4 Semi-arid regions

Candogan Yossef et al. (2013) concluded that the relative importance of ICs is the lowest in the Murray–Darling basin and any improvement of hydrological forecasts depends on better climate forecasts. The results of the present study for this basin show that the theoretical skill of ECMWF S3 forecasts are significantly high in some cases, but lower in other cases, with no apparent pattern (see Table S17). The accuracy of ECMWF S3 forecasts in assessment of actual skill is lower than ESP in most cases. Also, the ratios of actual to theoretical skill are quite low in this basin for both high and low flows.

Similarly, in the semi-arid African basins of the Orange River and the Zambezi, where the knowledge of MF plays a very important role in the forecast skill, the performance of ECMWF S3 forecasts is worse compared to the ESP in most cases. Tables S18 and S19 show that the accuracy of ECMWF S3 is lower than ESP in these basins, particularly in actual skill. In contrast, in the Nile basin, the ICs dominate the forecast skill, resulting in high performance of ESP forecasts throughout the year assuming that the release strategy of the Aswan reservoir is known (Candogan Yossef et al., 2013). The results of the present study show that the theoretical skill of ECMWF S3 cannot surpass the already high performance of the ESP (see Table S20). Actually, forecasts with ECMWF S3 perform considerably worse. In actual skill however, the accuracy of the ESP forecasts in the Nile is very low due to the large effect of the reservoir operations. In fact, the ratio of actual to theoretical skill is the lowest by far in this basin. With such a low accuracy of ESP forecasts despite the dominance of ICs, comparison of the performance of ECMWF S3 to ESP is not very meaningful. Our results of actual skill in both high and low flows in the Nile appear to be very erratic indeed

4.5 Conclusions

We assessed the skill of seasonal streamflow forecasts with the global hydrological forecasting system FEWS-World, set up within the GLOWASIS project. Global hydrological model PCR-GLOBWB was run with the ESP procedure as well as with ECMWF S3 bias-corrected seasonal meteorological forecast ensembles. We produced ensemble forecasts of monthly discharges for 20 large rivers of the world, with lead times of up to 6 months. We quantified the skill of ECMWF S3 forecasts compared to the reference ESP forecasts using the BSS, both for high and low flows. We determined the theoretical skill by comparing the results against model simulations, as well as the actual skill by comparing against discharge observations. We also calculated the ratios of actual to theoretical skill.

We analysed these results in the context of prevailing hydroclimatic conditions. This analysis suggests that the skill varies considerably according to location, season and lead time. The conclusions can be summarized as follows:

– In general, the performance of the ECMWF S3 forecast run is close to that of the ESP forecast run.

– There are basins where the ECMWF S3 forecast run performs significantly better than the ESP, during certain periods of the year and at certain lead times.

– However, there are in fact more cases where the ECMWF S3 forecast run performs worse than the ESP.

– In most cases, the apparent potential for improvement in seasonal hydrological forecasts by using better meteorological forecasts cannot be realized as yet with the model PCR-GLOBWB and the ECMWF S3 re-forecast dataset.

– As more accurate global hydrological models and more skillful seasonal meteorological forecasts become available in the future, such as the most recent ECMWF system S4, further studies will be needed to assess the improvement in seasonal hydrological forecasts, as well as the effect of meteorological forecast quality vs. model errors on the hydrological forecasts.

The Supplement related to this article is available online at https://doi.org/10.5194/hess-21-4103-2017- supplement.

Chapter 5
Skill and value of global-scale seasonal hydrological forecasts, a perspective

5.1 Introduction

While the skill of global seasonal hydrological forecasting systems has been demonstrated in this thesis, their value for the society has not yet been examined systematically. In fact, definitions of skill and value differ significantly among studies. Existing studies have mostly investigated the value of seasonal hydrological forecasts for a specific sector in one basin or region, but the value of global seasonal hydrological forecasting systems for water management has not been discussed comprehensively. With this study, we aim to shed some further light on the value problem as a contribution to the improvement of the usefulness of global seasonal hydrological forecasting systems for water management. For this purpose, we present the current practice of global-scale seasonal hydrological forecasting in Section 5.2, and we study the interaction between skill and value in Section 5.3. In Section 5.4, we discuss the possible ways to improve the value of seasonal streamflow forecasts on a global scale during various stages of the forecast chain, i.e. forecast communication (5.4.1), forecast adoption (5.4.2), forecast use in decision making (5.4.3), feedback from forecast users to forecast providers (5.4.4); with an emphasis on flood and drought mitigation. Finally in Section 5.5, we present the conclusions of this chapter.

5.2 Global-scale seasonal hydrological forecasting

Seasonal hydrological forecasting is the attempt to provide useful information about hydrological variables such as streamflow, soil moisture etc. that can be expected in the coming months. Effectively communicated forecasts of water availability and scarcity months in advance have a potential for successful application in water related sectors. Reliable streamflow forecasts on monthly and seasonal time ranges are vital for mitigation of flood and drought hazards as they give disaster management agencies and humanitarian aid organizations the opportunity to prepare for an appropriate response. Seasonal forecasts are beneficial not only in case of hydrological extreme events, but also during normal flow conditions, allowing several sectors, e.g. water supply, hydropower production, agriculture and navigation, to make more informed management decisions. Seasonal hydrological forecasting on a global scale could be especially valuable for developing regions, where effective hydrological forecasting systems are scarce.

Global hydrological forecasting has been enabled in the past decade thanks to recent scientific and technological developments. These developments include advancement of global modelling capabilities both in meteorology and land surface hydrology, enhanced collaboration between hydrological and meteorological communities, increased availability and quality of relevant data derived from ground observations and remote sensing by satellite and ground-based radars, as well as improvements in computing capabilities and resources (Emerton et al., 2016). The rationale behind operating global hydrological forecasting systems is that, as they are based on global meteorological datasets, they provide continuous and spatially consistent forecasts of streamflow. This may be valuable for regions where the spatial scale of hydrological extreme events goes beyond individual catchments or political borders as well as for the most vulnerable regions of the world where no local forecasting systems exist to alert the population. Still, where national scale forecasting systems exist, global forecasts provide an additional guidance at larger spatial scales (Harrigan et al., 2019). Disaster management organizations operating at global scale and international humanitarian aid agencies can benefit from global forecasts to prepare for appropriate response, and global water and energy markets can be informed about future availability of water and hydropower in different regions of the world. The economic rationale is that the provision of forecasts for basins across the globe does not require a large scale-up of resources. Rather than focusing on developing forecasting systems and issuing forecasts for individual basins in regions of scarce resources, it is more cost-effective to provide forecasts with global scale hydrological forecasting systems. Also, the economic benefit is evident for those countries who do have some existing capabilities, such as local hydrological models but are not able to produce hydrological forecasts, since they cannot afford to pay for access to, or processing of computationally expensive probabilistic and extended time scale meteorological forecast products (Emerton et al., 2018).

Over the last decade, several seasonal hydrological forecasting systems have been developed for forecast applications and research purposes at the continental (Bennett et al., 2016; Mo et al., 2014; Wood et al., 2002, 2005; Yuan et al., 2013) and global scale (Yuan et al., 2015). Yet currently only a few systems produce operational seasonal hydrological forecasts on such large scales. Continental scale operational systems include the European Flood Awareness System (EFAS) (Arnal et al., 2018), the European Hydrological Predictions for the Environment (E-HYPE) (Donnely et al., 2015), the Australian Government Bureau of Meteorology (BoM) Seasonal Streamflow

Forecasts (BoM, 2018), and the National Hydrologic Ensemble Forecast Service (HEFS) for continental USA (Demargne et al., 2014; Emerton et al., 2016). Seasonal hydrological forecasting systems that are or have been operational at the global scale include the NASA Hydrological Forecast and Analysis System (NHyFAS) (Arsenault et al., 2020), the Global Flood Awareness System (GloFAS - Seasonal) (Emerton et al., 2018), the Global Flood Forecasting and Information System (GLOFFIS) (Emerton et al., 2016), and the Global Water Scarcity Information Service (GLOWASIS) (Weerts et al., 2013).

5.3 Skill and value of global seasonal streamflow forecasts

Seasonal streamflow forecasts produced by large scale hydrological forecasting systems are continually being verified, often as an automated quality check. Their predictive skill is being quantified using increasingly sophisticated methods and their value is being assessed for water management applications.

The most basic verification for streamflow forecasts consists of quantifying their skill in terms of how close they are to actual observations of river discharge. However, a key element in the evaluation of large-scale hydrological ensemble prediction systems is to determine whether the forecasts have added skill over climatology or another naïve forecast (Pappenberger et al., 2015). For this purpose, the skill of a probabilistic forecast can be assessed by comparing the relative closeness of both the forecast and a benchmark to the observations. Quantifying the added skill over a benchmark is important for forecasters to understand and improve the performance of forecast components, as well as for users to know how much better the forecast is compared to a lower-cost, second-best guess. In practice, several different benchmarks are being used for shorter range forecasts, but most seasonal hydrological forecasting systems use the ensemble streamflow prediction (ESP) as a benchmark (Samaniego et al., 2019). The ESP approach provides the forecast skill that can be achieved from the initial hydrological conditions, by using the climatology as meteorological forcing. For shorter lead times and in regions that have a long hydrological memory, the ESP can provide a highly skillful forecast because the impact of the initial hydrological conditions dominates the seasonal predictability (Wanders et al. 2019a). For longer lead times, the ESP tends to become close to hydroclimatology and the performance of ESP-based forecasts is comparable to that of dynamical forecasts based on general circulation models (GCMs). When the forecast skill of the GCM-based seasonal streamflow forecasts is benchmarked with respect to the skill of the ESP, it can be identified whether using actual meteorological forecasts provides an added skill over using historical meteorological observations.

The predictive skill of a hydrological forecasting system is affected not only by errors in model structure and parameterization and initial states such as soil moisture, groundwater and snow, but also by meteorological forcing errors. Skill of seasonal hydrological forecasts can be improved by including better meteorological forecasts, and/or by better estimation of initial hydrologic states through assimilation of independent hydrological observations. The improvement in the overall predictability that may be attained depends on the relative importance of these two sources of uncertainty, which varies considerably among hydrological systems according to location, season and lead time (Bierkens and van den Hurk, 2007; Bierkens and van Beek, 2009; Shukla and Lettenmaier, 2011; Shukla et al., 2011; Yuan et al., 2015) Here, the ESP ensembles, which contain no actual information on weather, serve as a benchmark to assess the added skill that may be obtained using actual seasonal meteorological forecasts. Verification metrics are calculated for the ESP benchmark as well as the GCM-based forecast and translated into skill scores which quantify this additional skill.

The skill assessment of ensemble probabilistic forecasts includes verification measures which describe the relationship between forecasts and observations based on their joint distribution. Various attributes of forecast quality are addressed by using different verification measures. These attributes include accuracy, sharpness, reliability, discrimination and overall performance (Arnal et al., 2018). Accuracy refers to the magnitude of the errors between the forecast ensemble mean and the observation, given by the mean absolute error (MAE). Sharpness is an attribute of the forecast only, which indicates the ability to predict forecast values with probabilities that differ from climatology. It is a measure of the spread of the ensemble members of a forecast, given by an interquantile range (IQR), i.e., the difference between the n^{th} and $(100\text{-}n)^{th}$ percentiles of the forecast distribution. Another skill attribute is reliability, which is the statistical consistency between forecast probabilities and observed frequencies. It is given by the probability integral transform (PIT) diagram, which is the cumulative distribution of the PIT values as a function of the PIT values, in which the PIT values measure where the observation falls relative to the percentiles of the forecast distribution. To determine the skill in terms of overall performance, several skill scores can be used in the verification of seasonal hydrological forecasts.

Rather than producing exact discharge amounts, a streamflow forecasting system may predict the category in which the expected discharge falls into, i.e., whether it is higher or lower than normal conditions. High and low flows are defined as discharges higher and lower than certain thresholds which are calculated for each given month, separately for forecasted values and observed values. This approach eliminates any systematic bias in the simulations compared to the observations. In this way, the added skill obtained by using actual seasonal meteorological forecasts compared to the reference ESP forecast in predicting the right category of high, normal or low flows for a given month can be assessed. In the case of categorical forecasts, the Brier skill score BSS, given by Eq. (5.1) is an appropriate metric that may be employed for the assessment of skill.

$$BSS = 1 - \frac{BS_{forecast}}{BS_{ref}}$$

The BSS quantifies skill on the basis of the ratio of the Brier score (BS) of the actual forecast over that of a chosen reference dataset, such as a naïve estimator:

The BS values for a given month and lead time are given by Eq. (5.2):

$$BS = \frac{1}{N}\sum_{t=1}^{N}(p_t - o_t)^2$$

where N is the number of forecasting instances, p is the forecasted probability and o is the observed probability of an event.

BS is the mean squared error of probabilistic forecasts for a given dichotomous event. A probability threshold is used to define the binary event to be observed and forecasted. The BS is one of the most commonly used skill scores for seasonal forecasting in meteorology and hydrology. It is a relevant verification metric for analysing the performance of a forecast system for specific categories, defined by a set of thresholds. It uses categorical forecast thresholds to determine the quality of the forecast compared to a reference simulation. It is preferred for being a proper score, i.e., being optimized for forecasts that correspond to the best judgment of the forecaster. It is also a highly compressed score; i.e., it directly accounts for forecast probabilities without necessitating a contingency table for each probability threshold (Bartholmes et al., 2009; Ferro, 2007). Besides the BS and the associated skill score BSS, it is possible to use other verification metrics, such as the relative operating characteristic (ROC) score, or the continuous ranked probability skill score (CRPSS) for the assessment of forecasting skill.

In chapter 3, we investigated the relative contributions of initial conditions (IC) and meteorological forecasts (MF) to the forecasting skill of the global seasonal streamflow forecasting system Flood Early Warning System-World (FEWS-World). Potential improvement in forecasting skill through better MF or by better estimation of IC through data assimilation depends on the relative importance of these two sources of uncertainty. The study explored the impact of both sources of forecast uncertainty at large river basins across the globe using the ESP/revESP procedure outlined by Wood and Lettenmaier (2008). The results suggested that in some basins, such as arctic rivers or very large rivers, forecast skill may be improved by better estimation of initial hydrologic states through assimilation of snow, soil moisture or surface water data; whereas in others, such as monsoonal rivers improvement of forecast skill depends on more accurate seasonal climate prediction. This analysis showed the relative contributions of IC and MF to the potential skill of the forecasting system FEWS-World. In an actual forecast, where both the IC and the MF will be uncertain, the actual forecasting skill of the system should be assessed in a real forecasting mode, using probabilistic seasonal meteorological forecasts and comparing the ESP results to actual discharge observations.

A skill assessment in real forecasting mode is presented in chapter 4, where we evaluated the ability of the FEWS-World seasonal forecasting system based on the global hydrological model PC-Raster Global Water Balance (PCR-GLOBWB) to predict high and low flows, defined as discharges higher than the 75th and lower than the 25th percentiles for a given month, respectively. The thresholds were calculated separately for forecasted values and observed values. This approach eliminates any systematic bias in the simulations compared to the observations. In this way, the skill in forecasting the occurrence of flows that are higher or lower than usual for a given month were assessed. This study quantified the skill of ECMWF S3 forecasts compared to the reference ESP forecasts using the BSS, both for high and low flows. The analysis of the results in the context of prevailing hydroclimatic conditions suggested that the skill varies considerably according to location, season and lead time. The performance of the ECMWF S3 forecast run was found to be generally close to that of the ESP forecast run. In some basins the ECMWF S3 forecast run performed significantly better than the ESP, during certain periods of the year and at certain lead times, for example in some tropical monsoon dominated basins. However, in fact there were more cases where the forecast run performs worse than the ESP, such as many of the semi-arid basins. The study concluded that in most cases, the apparent potential for improvement in seasonal hydrological forecasts by using climate predictions cannot be realized until more accurate hydrological models and more skillful seasonal meteorological forecasts become available in the future, such as the Seas6 forecasts that are now in production at ECWMF, with substantial improvements like standard variables of soil moisture and river flow.

Global hydrological models such as PCR-GLOBWB tend to show a long hydrological memory, which limits the impact of the dynamical forecast improvement. Wanders et al. (2019a) argue that hydrological models which respond rapidly to precipitation or temperature changes are more likely to benefit from accurate dynamical seasonal forecasts and thus show a stronger improvement in the BSS. A more accurate representation of the observed hydrology is therefore needed to benefit from GCM-based seasonal streamflow forecasts, especially in regions with shorter hydrological response times (Samaniego et al., 2019).

Besides FEWS-World used by Candogan Yossef et al. (2017), the skill of several other large-scale hydrological forecasting systems is being assessed by various research groups. However, only a few systems have been thoroughly validated in real-time, operational conditions. Moreover, in most cases, a comprehensive analysis of the relative importance of the main sources of hydrological forecast skill is lacking. Understanding the influence of these sources on skill is crucial to determine how the forecast performance can be improved, and the biases of models that can affect the quality of the forecasts in terms of reliability, sharpness and accuracy can be eliminated (Lavers et al., 2020).

What appears from the studies on continental and global hydrological forecasting systems is that forecast quality varies considerably by region and season. As would be expected, systems generally show a decrease in forecasting skill with increasing lead time. Similar to the skill assessment of the FEWS-World system, the evaluation of the operational global forecasting system GloFAS-Seasonal shows that in many regions and seasons, forecasts of both high and low flow events are more skillful than the climatology for 1 or 2 and in some cases up to 4 months ahead (Emerton et al., 2018). However, there are regions and seasons for which the GloFAS-Seasonal forecasts are less skillful than climatology. In these river basins it would still be more useful to use a long-term average climatology rather than GCM-based climate prediction. In many cases, seasonal streamflow forecasts produced by these large-scale systems do not yet have the skill necessary for their adoption for water management applications.

Samaniego et al., (2019) report that through consultations with stakeholder focus groups, including representatives of national government agencies, regional and local government authorities, international water and hydropower companies, agricultural sector, river basin authorities, consultancies, and academic sector, it was found that in general seasonal forecasts need to have a better skill before they can be used operationally. Lavers et al., (2020) refer to a gap between the low skill currently available in seasonal hydrometeorological forecasts and the high expectation from the user community for forecasts at such lead times. They also point out the lack of accuracy of seasonal forecasts at local scales due to local anthropogenic influences such as dams and reservoirs, which have large and potentially predictable impacts on streamflow. These impacts change the space and time dynamics of floods and droughts but are not usually taken into account in global models. It is recommended in the said study that forecasts may be improved by coupling hydrological models with reservoir management information such as data on regulated dam releases during drought periods, maximum storage capacity for flood retention or objective filling curves for seasonal reservoir operations. Still, incorporating human regulated systems and quantifying the impacts of human activities represent one grand challenge in large-scale hydrological modelling (Bierkens, 2015). Emerton et al., (2018) suggest that the use of river flow observations could lead to significant improvements in skill by adjusting the forecasts through model calibration using historical observations and assimilation of real-time data. However, this too remains a challenge due to the lack of openly available river flow data, especially in real-time. Another problem for large-scale hydrological forecasting, therefore, is the need for more observations, which is essential not only for providing initial hydrological states to force the models, but also for evaluation of the forecasts and continuous improvement of skill.

Murphy et al. (2001) argue that although the information given by large-scale seasonal streamflow forecasts may have limited utility for operational water management until forecast skill can be improved, users should be offered the opportunity to understand this information. This would allow users to decide for themselves whether to take the inherent risk of using the forecast information in their water management practice. Therefore, no matter what level of accuracy may actually be obtained, it is always important to provide users with streamflow forecasts. In fact, even if there is apparently little added skill, the forecast may still be useful and it may offer a large added value for decision-making (Viel et al., 2016).

Although the term added value is used in some studies, rather inaccurately, to denote improved forecast skill, forecast value signifies the benefit achieved by incorporating the forecast into the decision-making process within a certain field of application. Skill and quality are interchangeable terms which denote a measure of how similar the forecast is to the actual outcome, independent of how the forecast is used. Forecast value differs from forecast skill in that it is dependent on the forecast application (Anghileri et al., 2016 after Murphy, 1993). While decisions based on more accurate forecasts are expected to be more effective, better forecast skill does not necessarily translate into improved efficacy of the decisions based on the forecast.

According to Ritchie et al. (2004) a forecasting system may be considered useful if the forecast is statistically verified and offers a positive value of information. Benefits gained by using streamflow forecasts depend not only on their skill but also on the way they are presented, distributed and used. An effective application of a seasonal climate forecast is defined as the use of forecast information leading to a change in a decision that generates improved outcomes in the system where it is applied. Producing skillful predictions is a necessary but insufficient condition for this. The success of a hydrological forecasting system will ultimately be determined not by its skill but by the effect it has on decision-making for water management (Plummer et al., 2019). The current ability of seasonal streamflow forecasting systems to predict the right category of an event months ahead is potentially valuable for many water-related applications such as flood preparedness, drought-risk management, reservoir management, hydropower and navigation. It is not so straightforward however to convert the forecast skill into an added value for decision-making. Integrating new forecasting products into established decision-making chains is not an easy task (Arnal et al., 2018). A seemingly useful scientific finding does not simply translate into usable information that fits into any decision-making process (Soares and Dessai, 2016). Although the quality and usefulness of seasonal streamflow forecasts have been improving in the last decades, their usability for decision-making has lagged behind. The current gap between the usefulness and the usability of seasonal streamflow information is pointed out by White et al. (2017). It is important to recognize that the usefulness of seasonal forecasts depends strongly on user requirements. Actual realization of the potential added value by incorporating forecast information into decision-making, is an important outcome that may be achieved by collaboration of forecast producers and users throughout the forecast chain (Block, 2011; Plummer et al., 2019). Even though no single objective measure can quantify the value of a forecast for all users, cost-benefit studies may be carried out on individual applications to assess the value of forecasts for each user (Lavers et al., 2020; Bischiniotis et al., 2019.)

5.4 Improving the value of global seasonal streamflow forecasts

A typical streamflow forecast chain entails a sequence of actions, which starts with construction of forecasts and warnings, communication of forecasts to users, adoption of forecasts by users, decision-making processes to take appropriate response action and feedback from the users to forecasters. Suggestions in the literature to improve the value of seasonal streamflow forecasts through the steps of the forecast chain are discussed in the following sub-sections and summarized in Table 5.1 below.

Table 5.1: Suggestions to improve forecast value through the forecast chain

Step	Suggestion	Reference
Forecast construction	Tailoring forecast products to users' needs Aligning the timing of forecasts with the timing of users' decisions Information exchange between forecast providers and users	Soares and Dessai, 2016 Plummer et al., 2019 Lavers et al., 2020

Forecast communication	Timely transmission through correct channels Communication of forecast skill and uncertainty Assessment of users' perception of uncertainty Visualization tools for graphical representation Internet websites based on a user-centered design International and interdisciplinary cooperation	Plummer et al., 2019 Lavers et al., 2020 Samaniego et al., 2019 Ramos et al., 2010 Emerton et al., 2018
Forecast adoption	Development of local decision support systems Tailoring forecast information to needs of water managers Simulation modelling using retrospective forecasts Reducing the risks of forecast use through insurance hedging Documentation of existing uses and experiences with forecasts Development of knowledge portals for information sharing	Whateley et al., 2014
Forecast use in decision making	Training decision makers through workshops, training activities, games, simulation environments Development of decision support systems	Coelho and Costa, 2010
Feedback from users to providers	Increased communication within the forecast community Users taking an active role in forecast production Output from user-specific decision support systems	Soares and Dessai, 2016 Lavers et al., 2020

5.4.1 Forecast Construction

The first step in the forecast chain is forecast construction. This step includes running a hydrological model driven by meteorological forecasts and initial conditions determined by observations, interpretation of model outputs, defining appropriate threshold values and construction of forecasts and warnings according to these thresholds. Continued information exchange between scientific researchers, forecast providers, water managers and forecast users is crucial for the provision of timely and reliable forecasts (Lavers et al., 2020). Forecast providers should understand users' needs and tailor their present and future products to these needs. They should consider the timing of forecasts to align well with the timing of users' decisions (Plummer et al., 2019). The cooperation between forecast producers and users is important not only for providing better forecasts, but through all the steps in the forecast chain, as it increases accessibility of users to forecast information, contributes to informed decision-making and in return, provides users' feedback to forecasters so that they can improve the usefulness and usability of their forecasts and warnings (Lavers et al., 2020). Arnal et al. (2018) argue after Soares and Dessai (2016) that a key factor for improving the value of seasonal forecast information is collaboration between forecast producers and users. They mention two international projects, namely the Horizon 2020 IMPREX (IMproving PRedictions and management of hydrological EXtremes) project and the Hydrologic Ensemble Prediction EXperiment (HEPEX). The IMPREX project (van den Hurk et al., 2016) is a gathering where forecast providers and users have the opportunity to discuss the integration of seasonal information in decision-making in the management of floods and droughts over Europe. HEPEX is another international platform where forecasters and water management professionals come together for collaboration on the use of ensemble prediction in water management applications, that offers a perfect setting for discussing the usefulness and usability of seasonal streamflow forecasts for decision-making (Arnal et al, 2018).

5.4.2 Forecast Communication

Communication of forecasts is one of the vital steps in the forecast chain. Forecasts should be transmitted to decision makers through the correct channels in a timely manner. Forecast skill should always be communicated to users, so that they can decide for themselves how much confidence to attach to the forecast information.

Seasonal forecasting has become possible as a consequence of a shift from deterministic to probabilistic forecasting schemes (Murphy et al., 2000), producing large amounts of data, which necessitates providing the user with the essence of the information and enabling correct interpretation of the forecasts (Emerton et al., 2016; Samaniego et al., 2019). Therefore, communication of forecast uncertainty as well as forecast skill is crucial. This has resulted in an emphasis on the communication of probabilistic forecasts, with increased efforts from the forecasting community to understand how users perceive uncertainty information, to develop forecast visualization tools and products for graphical representation and to provide guidance and support in using ensemble forecasts in decision-making (Ramos et al., 2013).

Lavers et al. (2020) suggest that decision-making in several water management sectors is not fully developed to consider probabilistic scenarios. The user community includes not only specialists in the disciplines of hydrology and water management, but also individuals with no experience in using forecasts. Users may interpret ensemble streamflow forecasts in different ways and be reluctant to use them in their decision-making unless they have confidence in forecast skill and uncertainty. It is important therefore, to avoid uncertainty miscommunication by forecast providers, as well as misinterpretation by users. Efficient communication of forecasts requires quantification of uncertainty, as well as an assessment of users' perception of uncertainty (Ramos et al., 2010).

Communication of forecasts is a bigger challenge when seasonal streamflow forecasts are produced for the whole globe and need to be communicated to a large range of user groups in countries all over the world. An important medium for dissemination of global seasonal streamflow forecasts is internet websites. A good example is the website of the operational global-scale system GloFAS-seasonal (Emerton et al., 2018), which is based on a user-centred design, with user needs central to the web interface developed to promote simplicity, joy of use, and usability. Emerton et al. (2016) discuss the difficulties of incorporating global-scale forecasts into national warning chains, while respecting the single voice principle, which states that national services are the sole authoritative voice on hydro-meteorological warnings in their respective countries. They identify international and interdisciplinary cooperation as key in overcoming these difficulties and underline the importance of political agreement between upstream and downstream countries for the sharing of information.

5.4.3 Forecast Adoption

The next step in the forecast chain is adoption of forecasts by users. Adoption of forecast information is influenced by a complex array of factors including forecast skill, behavioural effects, financial and institutional constraints. A lack of forecast adoption is well documented, especially for seasonal time-scales (Pagano et al., 2002; Rayner et al., 2005; Yarnal et al. 2006; Lemos 2008; Block, 2011). Despite the potential benefits and wide public availability of seasonal streamflow forecasts, water managers are usually reluctant to incorporate the use of seasonal forecast information into their practice. Reasons for this reluctance include low forecast skill, excessive uncertainty, poor timing of forecast production and dissemination, risk averse and inflexible behaviour of water managers, individual accountability, preference for established practices, insufficient human and institutional capacities, financial constraints, political disincentives, and technical difficulties in incorporating information in the existing decision support systems (Kirchoff et al., 2013; Coelho and Costa, 2010; Block, 2011; Whateley et al., 2014; Crochemore et al., 2015).

In order to understand the adoption of seasonal hydro-climatic forecasts in water management, Whateley et al. (2014) applied the Diffusion of Innovations (DoI) framework (Rogers, 2003) to the water sector. DoI is a general tool to analyse the spread and adoption of innovations within a social system, by investigating the influence of five key characteristics of innovation on the rate of adoption, defined by Rogers (2003) as shown below in Table 5.2:

Table 5.2: Key characteristics of innovation by Rogers (2003)

1	Relative advantage	The amount of improvement a new innovation contributes to present conditions.
2	Complexity	The challenges associated with understanding and utilizing a new innovation.
3	Compatibility	The ease at which a new innovation can be incorporated into existing and future decision-making processes.
4	Trialability	The ability to experiment with a new innovation without having to fully commit to its adoption and implementation.
5	Observability	The ability to observe the implementation and use of a new innovation by an external source.

Application of the DoI framework to the use of seasonal hydro-climatic forecasts, in water management provides valuable insights into the issues related to the adoption of seasonal forecasts. Addressing the challenges faced in each of the key attributes of DOI may help increase forecast adoption by water managers, and therefore improve the value of seasonal hydro-climatic forecasts.

In terms of adoption of seasonal hydroclimatic forecasts, relative advantage appears to be the most challenging of the five key characteristics. Typically measured against climatology, relative advantage depends strongly on the forecast skill, as well as the nature of the system being managed. In most cases, the relative advantage of seasonal forecasts is not evident for water managers, and seasonal forecast information is rarely incorporated into their operations due to perceived low forecast skill. Whateley et al.

(2014) suggest that decision support systems (DSS) may present a higher relative advantage for a given skill level. The relative advantage of seasonal forecasts may increase in the future with expected climate change, increasing climatic variability and continuing human interventions to water resource systems, thus the existing historic climatological record becoming less useful in predicting future conditions.

Complexity is the next key characteristic influencing adoption and use of forecast information in water management. The challenges involved are the complexity of the forecast itself, i.e., the difficulty to understand or interpret, as well as the difficulty in obtaining it. Seasonal forecasts are typically presented in probabilistic terms, which makes it more difficult to incorporate into decisions than a simple deterministic prediction. Misinterpretation of forecast information as well as the tedious and time-consuming process of obtaining relevant forecast information stand out as the main problems related to complexity, which Whateley et al. (2014) suggest may be helped through the increased communication between forecast providers and users, as well as development of local, tailored DSS.

Compatibility of seasonal forecasts to water management appears to be one of the two most significant factors, besides relative advantage. It includes spatial, temporal and institutional compatibility, i.e., whether forecasts are geographically suitable for application to a specific system, temporally appropriate for operations, or consistent with institutional goals. Seasonal forecast information is often not compatible with water management operations. Spatial scales resolved in global climate models are often too coarse, which neglects local particularities, limits the predictive skill in hydrologic processes, and thereby the willingness of water managers to incorporate forecast information into the operation of local systems. The temporal compatibility of seasonal forecasts depends on the type of water-related system being managed, i.e., the temporal compatibility may be higher for the mitigation of droughts rather than floods. Due to their long temporal scales, seasonal forecasts are often perceived to be less useful than short-term forecasts, though they prove to be useful especially for planning purposes. Institutional compatibility presents many barriers to seasonal forecast adoption, including institutional rigidity, risk-aversion, financial and legal limitations. The issues associated with compatibility may be overcome by tailoring forecast information to the needs of water managers at specific locations.

Trialability appears to have a small effect on the use of seasonal hydro-climatic forecasts in water management, although the lack of trialability may be an important but overlooked aspect of seasonal forecast adoption. Given the extremely large negative consequences of an operational error in water management, any trial of a new innovation holds the risk of a major failure and societal impacts. Simulation modelling using retrospective forecasts may address this problem, but translating model results to practice is a challenge. Reducing the risks associated with forecast use for instance through employing insurance mechanisms to hedge the risk of low-probability events may help increase the adoption of seasonal hydroclimatic forecasts

The final key characteristic of the DoI framework applied to seasonal forecast adoption is observability. It would mean in this case, the ability for water managers to witness other users benefiting from the use of seasonal hydroclimatic forecasts. The documentation of existing uses and experiences with seasonal forecasts may provide insight into the value of forecasts in water management operations. The flow of information on the benefits of seasonal forecast use is limited however, because private hydroelectricity firms, which are often the leading agencies to adopt seasonal forecasts in their operations, tend to hold back from sharing their operational procedures. Knowledge portals may help address the challenges of observability as they improve communication between forecast providers and users and provide a medium for water managers to collaborate and share useful information.

5.4.4 Forecast use for decision making in water related sectors

Appropriate use of forecasts in decision making also presents room for improvement of seasonal forecast value. The decision-making processes for water management are extremely complex, as they involve many scientific, technical, financial, social and environmental issues; and decisions are expected to satisfy several possibly conflicting objectives (Crochemore et al., 2015). Decision makers thus face the challenge of arriving at a decision by making an assessment of forecasts and the uncertainty inherent to them. The assessment is dependent on the judgement of the decision maker, which means that it is inevitably subjective, and likely to be biased and inconsistent (Kreye et al., 2011). Individuals' biases and ignorance of uncertainty in the decision process may lead to misinterpretation of circumstances and erroneous decisions. Knowledge and experience on the part of the decision maker is crucial and may be achieved through cooperative engagement among forecast users, providers, researchers and stakeholders, in the context of workshops, training activities, games, simulation environments where they can experiment with forecasts and decisions. Another important factor is the development of decision support systems which produce relevant information for decision making. Decision makers need to be trained for assimilating the information produced by these decision support systems in order to maximize the benefits of using forecasts in their decisions (Coelho and Costa, 2010). With appropriate training and decision support tools, the potential for successful application of seasonal streamflow forecasts to decision making processes in several water-related sectors can be realized. These sectors include disaster management agencies and humanitarian organizations responding to floods and droughts, reservoir operators and hydro-power companies, water supply companies, irrigation agencies and the agricultural sector, the navigation sector, wildlife agencies and the tourism sector, and insurance companies. While many of these sectors use short- and medium-term forecasts as well, they may highly benefit from forecasts of expected water availability and scarcity on seasonal lead times, especially for longer term planning purposes. Provided they have sufficient skill, seasonal forecasts may present an opportunity to decision makers in water-related sectors to prepare for an appropriate response months in advance. Forecasts produced by global hydrological forecasting systems in particular could be useful to international aid organizations operating at the global scale.

5.4.5 Feedback from forecast users

The final step in the forecast chain is feedback from forecast users and decision makers to forecast providers and researchers. Increased communication within the forecast community may help improve the value of seasonal forecasts by giving users an opportunity to take a more active role in forecast production and provision. It allows users to feedback their needs and thus push for development of problem-driven science in forecasting. It also allows forecast producers and providers to understand users' needs and tailor their present and future products to these needs (Soares and Dessai, 2016). Output from user-specific application models for decision-making processes provides valuable feedback which is crucial in tailoring forecasts to specific applications (Lavers et al., 2020).

5.5 Seasonal streamflow forecasts in the management of floods and droughts

An essential application of seasonal streamflow forecasts is in the management of hydrologic extreme events such as droughts and floods. Flood and drought management aims to reduce the devastating socio-economic and environmental impacts of these disasters. The special report of the Intergovernmental Panel on Climate Change (IPCC, 2012), 'Managing the risks of extreme events and disasters to advance climate change adaptation', anticipates an increase in the frequency, duration and severity of hydrological extreme events globally due to climate change. Droughts are expected to increase in intensity and duration in arid regions, while extreme precipitation and floods events show a strong tendency to increase over large areas including wet tropical regions (Yuan et al., 2013; IPCC, 2013). In the same time a rise in the economic and social damages caused by these events is to be expected due to population growth, urbanization and economic development. The increasing risk associated with hydrological extreme events emphasize the importance of flood and drought mitigation for societies, and the necessity to predict the occurrence of these events in advance. Another implication of the changing climate and increasing climatic variability is that the past climate information is becoming less useful in predicting hydrological extreme events that may occur in the coming months. The assumption of a stationary climate, which is at the basis of the use of the historical climatological record to force hydrological models, is losing its validity; and the relative advantage of seasonal climate forecasts as opposed to climatology is increasing in those parts of the globe where seasonal climatic predictability is good (Viel et al., 2016). Climate change and the related increase in seasonal hydroclimatic uncertainty, as well as the rising complexity of socio-economic systems necessitate more adaptive strategies in decision making for better management of floods and droughts, a crucial part of which is making the best use of available seasonal forecast information.

One of the most useful measures for the mitigation of risks associated with hydroclimatic hazards is a well-functioning early warning system that provides reliable and timely information. This contributes economic benefits for developed countries, while it is essential for protection of lives and livelihoods in developing nations and vulnerable populations (Murphy et al., 2011). Seasonal forecasts buy the time needed to take vital actions well in advance of the anticipated hazard if they are integrated into efficient disaster management policies. This timing presents different opportunities for prevention, planning, preparedness and response (WMO, 2008). The value of a forecast depends on the degree to which it initiates effective mitigation measures. Specific response options to an expected flood or drought depend on the socio-economic conditions as well as the financial, technical and institutional capacity of the impacted regions. Often, a disaster situation overwhelms the capability of the affected society to cope and therefore an international response arises to provide assistance. Major international disaster response organizations such as the United Nations, the Red Cross/Red Crescent, international private voluntary organizations, and institutions from donor countries provide aid and assistance to national institutions in the affected countries.

Humanitarian organizations originally had a mandate to react to disasters only after they take place (de Perez et al., 2015). During the last couple of decades, the focus has shifted to understanding and mitigation of disaster risks, in addition to post-disaster response and restoration. The Sendai Framework that was adopted at the Third United Nations World Conference on Disaster Risk Reduction held in Sendai, Japan in 2015, introduced a number of innovations that emerged from the discussions (UNDRR, 2015). These include a shift to disaster risk management as opposed to disaster management and a strong emphasis on understanding and reduction of disaster risk as well as preparedness for effective response. Long term plans for response actions have become key components of disaster mitigation (Murphy et al., 2011). Increasing availability of forecasts and early warnings on several time scales has contributed to this development, allowing for a window of time in which to act before a potential extreme event occurs; and thereby reducing the adverse effects on the impacted society and environment. Disaster management agencies can take different types of preventative action after receiving forecasts at various lead-times, ranging from hours to months. These types of actions have different aims, costs and preparation needs (de Perez et al., 2015). Short lead-time forecasts are required for emergency response, but within this short period of time between the warning and the actual event, the response options are limited. Goods, services and information are often provided in the form of disaster relief in this initial phase of an emergency response.

In case of flood events disaster management agencies communicate the warning to local communities, and organise a response such as timely evacuation of the flood plain, relocation of valuables, distribution of water purification tablets, building shelters, pre-positioning relief items before roads get blocked, organizing civil defence and first aid procedures. Flood preparedness efforts include improvement of strategies for reservoir operation during high flows. In regulated basins the occurrence of flooding will be determined to a large extent by the operation of reservoirs upstream. In this case the mitigation action to prevent flooding may be the operation of the reservoir to release water prior to the onset of high flows into the reservoir. Longer lead time flood mitigation strategies include reinforcing infrastructure for flood defence, developing flood storage ponds, preparing for emergency response.

In case of drought warnings, emergency response includes arrangement of alternative water sources and allocation of available water resources based on priority use criteria. Aid organizations can initiate transport of food aid to affected regions. Longer lead time drought preparedness plans are aimed at extending the availability of water and reducing water demand. They include adjusting planting dates, planting drought-tolerant crops, selecting appropriate irrigation methods, applying restrictions of water use such as rationing programs, special water tariffs, and reduction of low-value uses. Water conservation measures that can be undertaken are recycling water, reduction of wastage, development of water allocation strategies among competing demands, examination of water pricing system, improved land-use practices, watershed management, rainwater/runoff harvesting, joint use of surface and groundwater, storing water in groundwater reservoirs, well improvement. Food policies can be adjusted to ensure adequate food stocks or regional trade linkages. With seasonal forecasts, aid organizations can act earlier to anticipate food-aid needs. This is important since international aid organizations require about four months to deliver food to an area after receiving confirmation of a need for aid.

Chances for reducing vulnerability to disasters are better when longer term forecasts are integrated into decision-making processes for flood and drought preparedness as they provide an extended time available for prevention, planning, preparedness and response in order to mitigate the risks of the anticipated hazard. Seasonal forecasts on the global scale are potentially valuable in two ways. Firstly, international disaster management and humanitarian aid agencies may benefit from them as they indicate regions of the world where the probability of drought or flood conditions is increased or reduced. Secondly, they may help improve the management of hydrological extremes in transboundary river basins as they offer a spatially consistent prediction of the future condition.

5.6 Conclusions and discussion

In this chapter, we discuss the skill and value of seasonal streamflow forecasts in general, and those produced by global scale hydrological models in particular. We argue that seasonal hydrological forecasting on a global scale could be especially valuable for transboundary river basins as well as for developing regions of the world, where no effective local hydrological forecasting systems exist. It has been established that the skill of forecasts produced by global hydrological forecasting systems varies considerably by region and season, showing a decrease in forecasting skill with increasing lead time, and that in many cases, seasonal streamflow forecasts produced by these large-scale systems need to have a better skill before they can be adopted for water management applications. However, even with little added skill, the forecast may still be useful for end-users, allowing them to decide if they should take the risk of using the forecast information. We conclude that the success of a hydrological forecasting system will ultimately be determined not by its skill but by the effect it has on decision-making for water management. The current ability of seasonal streamflow forecasting systems to predict the right category of an event months ahead is potentially valuable for many water-related applications.

A review of the literature to determine possible ways of improving the value of seasonal forecasts shows that the realization of the potential added value depends largely on the collaboration between forecast producers and users, during each link in the forecast chain, i.e., communication, adoption, use of forecasts in decision-making and feedback from users. Our conclusion is that appropriate communication of forecast uncertainty as well as forecast skill is the crucial first step in order to avoid misinterpretation by users. Faced with the challenges of communication when seasonal streamflow forecasts are produced for the whole globe and have to be disseminated to users in countries all over the world, we emphasize the need for international cooperation as well as the importance of internet websites based on a user-centred design as a medium for dissemination of global seasonal streamflow forecasts. We identify the reasons for the lack of adoption of seasonal forecasts by users as perceived low forecast skill, behavioural effects, technical, financial and institutional constraints. Our recommendations for increasing adoption by users are: developing decision support systems tailored to users' needs, reducing the risks of forecast use through insurance hedging mechanisms, as well as building knowledge portals to provide a medium for communication and information sharing. For successful application of seasonal forecasts in water-related sectors, we recognize the extreme complexity of the decision-making processes for water management, and the challenges of application of seasonal forecasts into these processes, given the uncertainty inherent to them. An important facilitator of forecast use appears to be the development of decision support systems and application models in various water-related sectors. Knowledge and experience turn out to be crucial in decision making, and attainable through collaboration among forecast users, providers, researchers and stakeholders; workshops, training activities, games, simulation environments where professionals can experiment with forecasts and decisions. Such communication opportunities allow decision makers and forecast users to provide valuable feedback to researchers and forecast providers, so that forecast products can be tailored to specific needs of users.

Appropriate use of seasonal forecasts is vital in the management of floods and droughts, in order to reduce the devastating socio-economic and environmental impacts of these disasters. The anticipated global increase in the frequency, duration and severity of hydrological extreme events due to climate change, as well as the rise in the economic and social damages caused by these events due to population growth, urbanization and economic development emphasize the importance of flood and drought mitigation and the necessity to predict the occurrence of these events in advance. The value of seasonal forecasts which provide the time needed for prevention, planning, preparedness and response, depends on their level of integration into efficient disaster management policies and on the degree to which they initiate effective mitigation measures. A shift of focus to disaster risk management instead of disaster management is arising as well as a strong emphasis on reduction of disaster risk and preparedness for effective response through long-term planning. There exist many different types of preventative action that may be taken by disaster management agencies in case of flood and drought events, in short and long-term, and the chances for reducing vulnerability to disasters are better when longer term forecasts are integrated into decision-making processes. We conclude that seasonal forecasts are useful because they provide a longer time for mitigation of the risks posed by the anticipated hazard, and that seasonal forecasting at the global scale has potential value for the operations of international disaster aid agencies, as well as for management of extreme hydrological events in transboundary river basins.

Chapter 6
Synthesis

This thesis consists of three studies on the skill of a global seasonal streamflow forecasting system and a discussion on the value of seasonal streamflow forecasts on the global scale. In this concluding chapter, the major findings of these studies are discussed in relation to the research objectives before general conclusions are drawn and recommendations presented. These objectives are iterated here:

1) to identify a methodology that can serve as a benchmark verification procedure for hydrological forecasting
2) to assess the prospect of using a global hydrological model GHM for forecasting hydrological extremes
3) to determine the relative contributions of meteorological forcing and initial hydrologic conditions to the skill of seasonal streamflow forecasts
4) to identify promising skill improvement methods
5) to assess the total skill of hydrological forecasts as affected by errors in model structure, in the estimation of initial hydrologic conditions as well as in the meteorological forcing obtained by numerical weather prediction
6) to shed light on the value of global scale seasonal streamflow forecasts for water management
7) to discuss possible ways to improve the value during various stages of the forecast chain

6.1 Assessment of the prospective skill of a GHM

The first study presented in Chapter 2 addresses the first 2 research objectives as can be seen in Fig 1.1, which displays a conceptual schematization of the logical progression of research in this thesis. This first study is an initial step in assessing the prospect of global hydrological forecasting, by testing the ability of the global hydrological model PC-Raster Global Water Balance (PCR-GLOBWB) in reproducing the occurrence of past extremes in the monthly discharge of 20 large rivers of the world. Model skill is assessed in three ways: first in simulating hydrographs, second in reproducing monthly anomalies and third in reproducing flood and drought events. This procedure provides a detailed assessment of forecasting skill and an insight into which types of forecasting are more promising. Verification of non-bias-corrected hydrographs that reflects model and forcing errors proves to be useful since it provides an opportunity for improvement and allows comparison with the results of other studies which use non-bias-corrected data. Eliminating the systematic bias due to model errors or forcing provides an indication of the maximum skill that can be achieved in operational forecasting. Simulations with PCR-GLOBWB are biased for most basins, and the skill in reproducing hydrographs is lower than the observed climatology. The model skill improves significantly after a post-processing bias correction and surpasses the skill of the observed climatology in most basins. Results of the analysis indicate that the skill obtained in reproducing monthly anomalies is higher than the climatology for all basins. The model also shows skill in reproducing floods and droughts, with a markedly better performance in the case of floods. The model skill surpasses that of a simple water balance estimate in all cases. Although simulated hydrographs may be biased and do not always outperform the observed climatology even after bias correction, higher skills can be attained in forecasting the occurrence of monthly anomalies as well as floods. I conclude that the use of PCR-GLOBWB for operational forecasting of monthly hydrological extremes is promising, and that the prospects for seasonal forecasting with PCR-GLOBWB are positive. Given the similarity of PCR-GLOBWB to other GHMs in model structure and parameterization, comparability of its performance in reproducing runoff to those of other GHMs and the forcing data being similar to those used in simulations with other GHMs and LSMs (land surface models), I argue that this conclusion is valid for other comparable GHMs and LSMs as well.

6.2 Assessment of the contribution of ICs and MF to forecast skill

In the previous assessment in Chapter 2, actual meteorological forecasts (MF) are not included but data from the observed climatology are used in order to concisely quantify the maximum attainable forecasting skill, also assuming initial conditions (ICs) are perfectly known. In this study presented in Chapter 3, research objectives 3 and 4 are assessed (see Fig 1.1). The relative contributions of initial conditions (ICs) and meteorological forcing (MF) to the forecasting skill of the global seasonal streamflow forecasting system Flood Early Warning System-World (FEWS-World) are investigated. Potential improvement in forecasting skill through better climate prediction or by better estimation of initial conditions through data assimilation depends on the relative importance of these two sources of uncertainty. The impact of both sources of forecast uncertainty is explored at large river basins across the globe using the ensemble streamflow prediction (ESP) and reverse ensemble streamflow prediction (revESP) procedure. Global monthly streamflow is simulated with lead times of 1–6 months for a historical period of 30 years (1981–2010). The ESP and revESP forecast ensembles are compared with retrospective model simulations driven by meteorological observations, thus eliminating model errors and relating predictability only to knowledge of ICs and the uncertainty in future MF. The variance of the ESP and revESP forecast ensembles are compared to the climatological variance by calculating the ratios of the mean squared error (MSE) of both ESP and revESP to the MSE of the climatology for 78 discharge stations on major global rivers, for 12 months of the year and for 6 lead times. Also, the critical lead time (CLT) after which the importance of ICs is surpassed by that of MF, is calculated for the corresponding basins and for each month of the year. Skill maps for the ESP and revESP as well as the CLT values indicate that the contribution of ICs and MF to hydrological forecasting skill varies considerably according to location, season and lead time. These results are analysed in the context of prevailing hydroclimatic conditions for several larger basins. This analysis suggests that in some basins forecast skill may be improved by better estimation of initial hydrologic states through assimilation of snow, soil moisture or surface water data. In others, improvement of forecast skill depends on more accurate seasonal climate prediction. General patterns are identified in the results based on hydro-climatological conditions, for arctic and snow-fed basins, monsoonal basins and for very large rivers. This shows that for arctic rivers as well as for rivers fed by snow and ice from mountainous regions, such as the Volga and Colorado, forecasts of high flows during the melt season depend largely on the ice and snowpack, especially where these have a high interannual variability. Thus, forecasts in these basins may benefit from assimilation of data on the snow and ice accumulated during the cold season. It is also recognized however, that in some snow-fed basins such as the Yenisey and the Mississippi, the onset of ice and/or snowmelt and consequently the timing of peak flow are highly sensitive to temperature changes at the end of the cold season, and consequently the importance of ICs diminishes in these cases. So, my conclusion is that improvement of forecast skill in these basins depends more strongly on better climate prediction. For monsoon-dominated rivers, it is observed that the interannual variability of the monsoon is the main factor determining the skill of hydrological forecasts for the wet period. In basins such as the Brahmaputra and the Yangtze where the onset of the thawing of snowpack and glaciers coincides with the start of the monsoon season, forecasts of high flows are dominated by the MF and skill improvement depends on prediction of the monsoon. ICs seem to play a more important role in basins like the Indus where snow and ice have a larger contribution to streamflow, especially when the ice and snowpack is variable from year to year. I conclude that better estimation of initial snow/ice states is likely to improve forecast skill during the wet season in such basins. Finally in large basins like the Amazon with extensive flood plains and large travel times of surface water, knowledge of ICs of surface water turns out to be an important source of skill for high flow forecasts on lead times of 2–3 months. In addition to surface water, the role of initial groundwater states also gains importance during the recession stage, when the groundwater discharge plays an important role.

6.3 Assessment of the skill of seasonal streamflow forecasts in actual forecasting mode

In the next study presented in Chapter 4, research objective 5 is addressed and the actual forecasting skill is assessed in real forecasting mode, where the ICs and the MF are uncertain (see Fig 1.1). To this end, I conduct a skill assessment of the FEWS-World system, which uses actual seasonal meteorological forecasts as input into the global hydrological model PCR-GLOBWB. The model is run with the ESP procedure as well as with European Centre for Middle Range Weather Forecasts (ECMWF) S3 bias-corrected seasonal meteorological forecast ensembles. Ensemble forecasts of monthly discharges are produced for 20 large rivers of the world, with lead times of up to 6 months. The skill of ECMWF S3 forecasts compared to the reference ESP forecasts are quantified using the Brier skill score (BSS), both for high and low flows. The theoretical skill is determined by comparing the results against model simulations, and the actual skill is determined by comparing against discharge observations. Also, the ratios of actual to theoretical skill are calculated. Analysis of these results in the context of prevailing hydroclimatic conditions suggested that the skill varies considerably according to location, season and lead time. It is determined that in general, the performance of the ECMWF S3 forecast run is close to that of the ESP forecast run. There are basins where the ECMWF S3 forecast run performs significantly better than the ESP, during certain periods of the year and at certain lead times. However, there are in fact more cases where the ECMWF S3 forecast run performs worse than the ESP. My conclusion is that in most cases, the apparent potential for improvement in seasonal hydrological forecasts by using better meteorological forecasts cannot be realized as yet with the model PCR-GLOBWB and the ECMWF S3 re-forecast dataset. I recommend that as more accurate global hydrological models and more skillful seasonal meteorological forecasts become available in the future, further studies would be needed to assess the improvement in seasonal hydrological forecasts, as well as the effect of meteorological forecast quality vs. model errors on the hydrological forecasts.

Although my findings in skill assessment studies indicate that the skill of forecasts produced by global hydrological forecasting systems varies considerably by region and season, showing a decrease in forecasting skill with increasing lead time, and that in many cases, seasonal streamflow forecasts produced by these large-scale systems need to have a better skill before they can be adopted for water management applications, I realize that even with little added skill, forecasts may still be useful for end-users, allowing them to decide if they should take the risk of using the forecast information. I recognize that the success of a hydrological forecasting system will ultimately be determined not only by its skill but also by its value to inform decision-making for water management.

6.4 Discussion of the skill and value of global seasonal streamflow forecasts

In Chapter 5, the last two research objectives 6 and 7 are addressed, as can be seen in Fig 1.1. The value of seasonal streamflow forecasts in general, and those produced by global scale hydrological models in particular are discussed. I argue that seasonal hydrological forecasting on a global scale could be especially valuable for transboundary river basins as well as for developing regions of the world, where no effective local hydrological forecasting systems exist. I conclude that the current ability of seasonal streamflow forecasting systems to predict the right category of an event months ahead is potentially valuable for many water-related applications. A review of the literature to determine possible ways of improving the value of seasonal forecasts shows that the realization of the potential added value depends largely on the collaboration between forecast producers and users, during each link in the forecast chain, i.e., forecast construction, communication, adoption, use of forecasts in decision-making and feedback from stakeholders and decision makers to forecasters and researchers. Appropriate communication of forecast uncertainty as well as forecast skill turns out to be a crucial step in order to avoid misinterpretation by users. Faced with the challenges of communication when seasonal streamflow forecasts are produced for the whole globe and have to be disseminated to users in countries all over the world, I emphasize the need for international cooperation as well as the importance of internet websites based on a user-centred design as a medium for dissemination of global seasonal streamflow forecasts. The reasons for the lack of adoption of seasonal forecasts by users are identified as perceived low forecast skill, behavioural effects, technical, financial and institutional constraints. My recommendations for increasing adoption by users are developing decision support systems tailored to users' needs, reducing the risks of forecast use through insurance mechanisms, as well as building knowledge portals to provide a medium for communication and information sharing. For a successful application of seasonal forecasts to decision making in water management, one needs to recognize the extreme complexity of the decision-making processes for water management, and the challenges of application of seasonal forecasts into these processes, given the uncertainty inherent to them. Knowledge and experience are identified to be crucial in decision making, and attainable through collaboration among forecast users, providers, researchers and stakeholders; workshops, training activities, games, simulation environments where professionals can experiment with forecasts and decisions. Another important facilitator of forecast use is found to be the development of decision support systems and application models in various water-related sectors.

As I look specifically into the case of forecast use in the mitigation of hydrological disasters, I conclude that appropriate use of seasonal forecasts is vital in the management of floods and droughts, in order to reduce the devastating socio-economic and environmental impacts of these disasters. I argue that the anticipated global increase in the frequency, duration and severity of hydrological extreme events due to climate change, as well as the rise in the economic and social damages caused by these events due to population growth, urbanization and economic development highlight the importance of flood and drought mitigation and the necessity to predict the occurrence of these events in advance. The value of seasonal forecasts which provide the time needed for prevention, planning, preparedness and response, is found to depend on their level of integration into efficient disaster management policies and on the degree to which they initiate effective mitigation measures. A shift of focus to disaster risk management instead of disaster management is arising as well as a strong emphasis on reduction of disaster risk and preparedness for effective response through long-term planning. With several existing types of preventative action that may be taken by disaster management agencies in case of flood and drought events, in short and long-term, I argue that the chances for reducing vulnerability to disasters are better when longer term forecasts are integrated into decision-making processes. I conclude that seasonal forecasts are vital as they provide the much-needed time for mitigation of the risks posed by the anticipated hazard, and that seasonal forecasting at the global scale has potential value for the operations of international disaster aid agencies, as well as for management of extreme hydrological events in transboundary river basins.

6.5 Conclusions and recommendations

Returning to the discussion of global hydrology presented at the very beginning of the thesis introduction, and given the aforementioned scientific, technical, and economic rationale behind operating global hydrological forecasting systems, I argue that in an era of global environmental change, seasonal streamflow forecasting on the global scale is not only relevant but also essential.

The conclusions reached through my studies may be summarized as:

1. Assessment of the prospective skill of a GHM

    a. The use of PCR-GLOBWB for operational forecasting of monthly hydrological extremes is promising, and the prospects for seasonal forecasting with PCR-GLOBWB are positive.

    b. Given the similarity of PCR-GLOBWB to other GHMs in model structure and parameterization, comparability of its performance in reproducing runoff to those of other GHMs and the forcing data being similar to those used in simulations with other GHMs and LSMs, the previous conclusion is valid for other comparable GHMs and LSMs as well.

2. Assessment of the contribution of ICs and MF to forecast skill

    a. In some basins forecast skill may be improved by better estimation of initial hydrologic states through assimilation of snow, soil moisture or surface water data. In others improvement of forecast skill depends on more accurate seasonal climate prediction.

    b. General patterns are identified in the results based on hydro-climatological conditions, for arctic and snow-fed basins, monsoonal basins and for very large rivers.

    c. In arctic rivers as well as in rivers fed by snow and ice from mountainous regions, such as the Volga and Colorado, forecasts of high flows during the melt season depend largely on the ice and snowpack, especially where these have a high interannual variability. Thus, forecasts in these basins may benefit from assimilation of data on the snow and ice accumulated during the cold season.

    d. In some snow-fed basins such as the Yenisey and the Mississippi, where the onset of ice and/or snowmelt and consequently the timing of peak flow are highly sensitive to temperature changes at the end of the cold season, the importance of ICs diminishes, and improvement of forecast skill in these basins depends more strongly on better climate prediction.

    e. In monsoon-dominated rivers, the interannual variability of the monsoon is the main factor determining the skill of hydrological forecasts for the wet period. In basins such as the Brahmaputra and the Yangtze where the onset of the thawing of snowpack and glaciers coincides with the start of the monsoon season, forecasts of high flows are dominated by the MF and skill improvement depends on prediction of the monsoon.

    f. ICs seem to play a more important role in some monsoonal basins like the Indus where snow and ice have a larger contribution to streamflow, especially when the ice and snowpack is variable from year to year. Better estimation of initial snow/ice states is likely to improve forecast skill during the wet season in such basins.

g. In large basins like the Amazon with extensive flood plains and large travel times of surface water, knowledge of ICs of surface water turns out to be an important source of skill for high flow forecasts on lead times of 2–3 months. In addition to surface water, the role of initial groundwater states also gains importance during the recession stage, when the groundwater discharge plays an important role.

3. Assessment of the actual skill of seasonal streamflow forecasts

   a. The skill varies considerably according to location, season and lead time, but in general, the performance of the ECMWF S3 forecast run is close to that of the reference ESP forecast run.

   b. While there are a few basins where the ECMWF S3 forecast run performs significantly better than the ESP, during certain periods of the year and at certain lead times, there are in fact more cases where the ECMWF S3 forecast run performs worse than the ESP

   c. In most cases, the apparent potential for improvement in seasonal hydrological forecasts by using better meteorological forecasts cannot be realized as yet with the model PCR-GLOBWB and the ECMWF S3 re-forecast dataset.

   d. As more accurate global hydrological models and more skillful seasonal meteorological forecasts become available in the future, further studies would be needed to assess the improvement in seasonal hydrological forecasts, as well as the effect of meteorological forecast quality vs. model errors on the hydrological forecasts.

4. Discussion of the skill and value of global seasonal streamflow forecasts

   a. Seasonal hydrological forecasting on a global scale could be especially valuable for transboundary river basins as well as for developing regions of the world, where no effective local hydrological forecasting systems exist.

   b. The current ability of seasonal streamflow forecasting systems to predict the right category of an event months ahead is potentially valuable for many water-related applications.

   c. The realization of the potential added value depends largely on the collaboration between forecast producers and users, during each link in the forecast chain, i.e., forecast construction, communication, adoption, use of forecasts in decision-making and feedback from stakeholders and decision makers to forecasters and researchers.

d. Forecast providers should understand users' needs and tailor their present and future products to these needs.

e. Appropriate communication of forecast uncertainty as well as forecast skill is crucial first in order to avoid misinterpretation by users.

f. The challenges of communication are greater when seasonal streamflow forecasts are produced for the whole globe and have to be disseminated to users in countries all over the world, which necessitates international cooperation as well as the use of internet websites based on a user-centred design.

g. The reasons for the lack of adoption of seasonal forecasts by users are perceived low forecast skill, behavioural effects, technical, financial and institutional constraints.

h. Adoption by users may be encouraged through developing decision support systems tailored to users' needs, reducing the risks of forecast use through insurance mechanisms, as well as building knowledge portals to provide a medium for communication and information sharing.

i. Given the extreme complexity of the decision-making processes for water management, and the inherent uncertainty in the application of seasonal forecast into these processes, knowledge and experience are crucial, and attainable through collaboration among forecast users, providers, researchers, and stakeholders.

j. Appropriate use of seasonal forecasts is vital in the management of floods and droughts, as they provide the much-needed time for prevention, planning, preparedness, and response.

k. The value of seasonal forecasts in reducing the devastating socio-economic and environmental impacts of these disasters depends on their level of integration into efficient disaster management policies and on the degree to which they initiate effective mitigation measures

l. Seasonal forecasting at the global scale has potential value for the operations of international disaster aid agencies, as well as for management of extreme hydrological events in transboundary river basins.

Synthesizing my conclusions, I maintain that global seasonal streamflow forecasting offers a great potential to fulfil a critical societal need, but there is a long way to go from its current state to reach this full potential. Continuing efforts by scientific research groups and the forecasting community at large would help reach this potential. I may summarize my main recommendations to improve the skill and value of global seasonal hydrological forecasts in the future as:

- improvements in model structure and parametrization
- inclusion of new and better model features such as reservoir operations and hydrodynamic routing
- refinement of spatial resolution of models
- enhancement of computing capabilities
- better prediction of future climate
- better estimation of initial hydrologic states through assimilation of higher quality data derived from ground observations and remote sensing by satellite and ground-based radars
- enhanced collaboration between hydrological and meteorological communities, as well as between forecast producers, users, researchers, and stakeholders
- enhanced international cooperation in data sharing and decision making
- developing internet websites based on a user-centred design as a medium for forecast dissemination
- developing decision support systems tailored to users' needs
- reducing the risks of forecast use through insurance hedging mechanisms
- building knowledge portals as a medium for communication and information sharing
- organizing workshops, training activities, games, simulation environments for professionals

References

Alcamo, J., P. Döll, T. Henrichs, F. Kaspar, B. Lehner, T. Rösch, and S. Siebert (2003), Development and testing of the WaterGAP 2 global model of water use and availability, Hydrol. Sci. J., 48, 317–337, doi:10.1623/hysj.48.3.317.45290.

Alfieri, L., P. Burek, E. Dutra, B. Krzeminski, D. Muraro, J. Thielen and F. Pappenberger (2013), GloFAS – global ensemble streamflow forecasting and flood early warning, Hydrol. Earth Syst. Sci., 17, 1161–1175, https://doi.org/10.5194/hess-17-1161-2013.

Allen, R. G., L. S. Pereira, D. Raes, and M. Smith (1998), Crop evapotranspiration, FAO Irrig. Drain. Pap. 56, Food and Agric. Organ., Rome.

Anghileri D., N. Voisin, A. Castelletti, F. Pianosi, B. Nijssen and D. P. Lettenmaier (2016), Value of long-term streamflow forecasts to reservoir operations for water supply in snow-dominated river catchments, Water Resour. Res. 52(6), doi: 10.1002/2015WR017864.

Arnal, L., H. I. Cloke, E. Stephens, F. Wetterhall, C. Prudhomme, J. Neumann, B. Krzeminski and F. Pappenberger (2018), Skilful seasonal forecasts of streamflow over Europe?: Hydrol. Earth Syst. Sci., 22, 2057–2072, https://doi.org/10.5194/hess-22-2057-2018.

Arnell, N. W. (1999), A simple water balance model for the simulation of streamflow over a large geographic domain, J. Hydrol., 27, 314–335.

Arnell, N. W. (2004), Climate change and global water resources: SRES emissions and socio-economic scenarios, Global Environ. Change, 14, 31–52, doi:10.1016/j.gloenvcha.2003.10.006.

Arsenault K. R., S. Shukla, A. Hazra, A. Getirana, A. McNally, S. V. Kumar, R. D. Koster, C. D. Peters-Lidard, B. F. Zaitchik, H. Badr, H. C. Jung, B. Narapusetty, M. Navari, S. Wang, D. M. Mocko, C. Funk, L. Harrison, G. J. Husak, A. Adoum, G. Galu, T. Magadzire, J. Roningen, M. Shaw, J. Eylander, K. Bergaoui, R. A. McDonnell and J. P. Verdin (2020), The NASA hydrological forecast system for food and water security applications, Bulletin of the American Meteorological Society. doi:10.1175/BAMS-D-19-0264.1

Balsamo, G., P. Viterbo, A. Beljaars, B. Van den Hurk, M. Hirschi, A.K. Betts, and K. Scipal (2009), A revised hydrology for the ECMWF model: Verification from field site to terrestrial water storage and impact in the integrated forecast system, J. Hydrometeorol., 10, 623–643.

Balsamo, G., S. Bousetta, P. Lopez, and L. Ferranti (2010), Evaluation of ERA-Interim and ERA-Interim-GPCP-rescaled precipitation over the U.S.A., ERA Rep. Ser., Tech. Rep. 5, ECMWF, Reading, U. K. Available at http://www.ecmwf.int/publications/library/do/references/list/782009

Balsamo, G., C. Albergel, A. Beljaars, S. Boussetta, E. Brun, H. Cloke, D. Dee, E. Dutra, J. Muñoz-Sabater, F. Pappenberger, P. de Rosnay, T. Stockdale and F. Vitart (2015), ERA-Interim/Land: a global land surface reanalysis data set, Hydrol. Earth Syst. Sci., 19, 389–407, https://doi.org/10.5194/hess-19- 389-2015.

Bartholmes, J. C., J. Thielen, M.H. Ramos and S. Gentilini (2009), The european flood alert system EFAS – Part 2: Statistical skill assessment of probabilistic and deterministic operational forecasts, Hydrol. Earth Syst. Sci., 13, 141–153, https://doi.org/10.5194/hess-13-141-2009.

Bierkens, M. F. P. (2015), Global hydrology 2015: State, trends, and directions, *Water Resour. Res.*, 51, 4923– 4947, doi:10.1002/2015WR017173.

Bierkens, M. F. P., and B. J. J. M. van den Hurk (2007), Groundwater convergence as a possible mechanism for multi-year persistence in rainfall, Geophys. Res. Lett., 34, LO2402, doi:10.1029/2006GL028396.

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.

Bischiniotis, K., B. van den Hurk, E. C. de Perez, T. Veldkamp, G. G. Nobre, J. Aerts (2019), Assessing time, cost and quality trade-offs in forecast-based action for floods, International Journal of Disaster Risk Reduction, 40 (2019) 101252.

Block, P. (2011), Tailoring seasonal climate forecasts for hydropower operations. Hydrol. Earth Syst. Sci., 15, 1355–1368,

www.hydrol-earth-syst-sci.net/15/1355/2011/ doi:10.5194/hess-15-1355-2011.

BoM, Bureau of Meteorology (2008), Forecast Verification - Issues, Methods and FAQ, Bureau of Meteorology, Melbourne.

http://www.bom.gov.au/bmrc/wefor/staff/eee/verif/verif_web_page.html.

Candogan Yossef, N., 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, 1–14, doi:10.5194/hess-16-1-2012.

Candogan Yossef, N., H. Winsemius, L. P. H. van Beek, A. Weerts 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, 4687–4699, doi:10.1002/wrcr.20350.

Candogan Yossef, N., L. P. H. van Beek, A. Weerts, H. Winsemius and M. F. P. Bierkens (2017), Skill of a global forecasting system in seasonal ensemble streamflow prediction, Hydrol. Earth Syst. Sci., 21, 4103–4114, doi:10.5194/hess-21-4103-2017.

Coelho, C. A. S., and S. M. S. Costa (2010), Challenges for integrating seasonal climate forecasts in user applications, Curr. Opin. Environ. Sustain., 2, 317–325, doi:10.1016/j.cosust.2010.09.002.

Collins, M., and M. R. Allen (2002), Assessing the relative roles of initial and boundary conditions in interannual to decadal climate predictability, J. Clim., 15, 3104–3109, doi:10.1175/1520-0442(2002)015.

Coughlan de Perez, E., B. van den Hurk, M. K. van Aalst, B. Jongman, T. Klose, and P. Suarez (2015), Forecast-based financing: an approach for catalyzing humanitarian action based on extreme weather and climate forecasts, Nat. Hazards Earth Syst. Sci., 15, 895–904, doi:10.5194/nhess-15-895-2015.

Crochemore, L., M. H. Ramos, F. Pappenberger, S. J. Andel, and A. Wood (2015), An experiment on risk-based decision-making in water management using monthly probabilistic forecasts, Bull. Amer. Meteor. Soc., doi:10.1175/BAMS-D-14- 00270.1

Decharme, B. and H. Douville (2006), Introduction of a sub-grid hydrology in the ISBA land surface model, Clim. Dynam., 26, 65–78.

Decharme, B. and H. Douville (2007), Global validation of the ISBA sub-grid hydrology, Clim. Dynam., 29, 21–37.

Dee, D. P., S. M. Uppala, A. J. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M. A. Balmaseda, G. Balsamo, P. Bauer, P. Bechtold, A. C. M. Beljaars, I. van de Berg, J. Biblot, N. Bormann, C. Delsol, R. Dragani, M. Fuentes, A. J. Greer, L. Haimberger, S. B. Healy, H. Hersbach, E. V. Holm, L. Isaksen, P. Kallberg, M. Kohler, M. Matricardi, A. P. McNally, B. M. Mong-Sanz, J. J. Morcette, B. K. Park, C. Peubey, P. de Rosnay, C. Tavolato, J. N. Thepaut and F. Vitart (2011), The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, Q. J. Roy. Meteorol. Soc., 137, 553–597, doi:10.1002/qj.828.

Demargne J., L. Wu, S. K. Regonda, J. D. Brown, H. Lee, M. He, D. J. Seo, R. Hartman, H. D. Herr, M. Fresch (2014), The science of NOAA's operational hydrologic ensemble forecast system, Bull. Am. Meteorol. Soc., 95:79–98.

De Roo, A. P. J., C. G. Wesseling and W. P. A. Van Deursen (2000), Physically based river basin modeling within a GIS: The LISFLOOD model, Hydrolog. Process., 14, 1981–1992.

Donnely C., J. C. M. Andersson, B. Arheimer (2015), Using flow signatures and catchment similarities to evaluate the EHYPE multi-basin model across Europe, Hydrol Sci J 2015, doi:10.1080/02626667.2015.1027710

Döll, P., and B. Lehner (2002), Validating of a new global 30-minute drainage direction map, J. Hydrol., 258, 214–231.

Döll, P., F. Kaspar, and B. Lehner (2003), A global hydrological model for deriving water availability indicators: Model tuning and validation, J. Hydrol., 270, 105–134.

Eagleson, P. (1986), The emergence of global-scale hydrology, Water Resour. Res., 22, 6S– 14S.

Ek, M. B., K. E. Mitchell, Y. Lin, P. Grunmann, E. Rogers, G. Gayno, V. Koren, and J. D. Tarpley (2003), Implementation of the upgraded Noah land-surface model in the NCEP operational mesoscale Eta model, J. Geophys. Res., 108, 8851, doi:10.1029/2002JD003296.

Emerton, R. E., E. M. Stephens, F. Pappenberger, T. C. Pagano, A. H. Weerts, A. W. Wood, P. Salamon, J. D. Brown, N. Hjerdt, C. Donnelly, C. A. Baugh and H. L. Cloke (2016), WIREs Water, doi: 10.1002/wat2.1137

Emerton, R. E, E. Zsoter, L. Arnal, H. L. Cloke, D. Muraro, C. Prudhomme, E. M. Stephens, P. Salamon and F. Pappenberger (2018), Developing a global operational seasonal hydro-meteorological forecasting system: GloFAS-Seasonal v1.0, Geosci. Model Dev., 11, 3327–3346, https://doi.org/10.5194/gmd-11-3327-2018.

Fekete, B. M., C. J. Vörösmarty, and W. Grabs (2002), High-resolution fields of global runoff combining observed river discharge and simulated water balances, Global Biogeochem. Cycles, 16, 1042, doi:10.1029/1999GB001254.

Ferro, C. A. T. (2007), Comparing probabilistic forecasting systems with the brier score, Weather Forecast., 22, 1089–1100, https://doi.org/10.1175/WAF1034.1.

Fiedler, K. and P. Döll (2007) Global modeling of continental water storage changes – sensitivity to different climate data sets, Adv. Geosci., 11, 63–68. Available at http://www.adv-geosci.net/11/63/2007/.

Gandin, L. S. and A. H. Murphy (1992), Equitable scores for categorical forecasts, Mon. Weather Rev., 120, 361–370.

Gedney, N. and P. M. Cox (2003), The sensitivity of global climate model simulations to the representation of soil moisture heterogeneity, J. Hydrometeorol., 4, 1265–1275.

Gerrity Jr., J. P. (1992), A note on Gandin and Murphy's equitable score, Mon. Weather Rev., 120, 2707–2712.

Gober, M., C. A. Wilson, S. F. Milton and D. B. Stephenson, D. B (2004), Fairplay in the verification of operational quantitative precipitation forecasts, J. Hydrol., 288, 225–236.

Gosling, S. N., N. W. Arnell (2016), A global assessment of the impact of climate change on water scarcity, Climatic Change 134, 371–385. https://doi.org/10.1007/s10584-013-0853-x

GRDC (2007), Major River Basins of the World, Global Runoff Data Centre, Federal Institute of Hydrology, D 56002, Koblenz, Germany.

Gusev, Y. M. and O. N. Nasonova (2003), The simulation of heat and water exchange in the boreal spruce forest by the land-surface model SWAP, J. Hydrol., 280, 162–191.

Haddeland, I., D. B. Clark, W. Franssen, F. Ludwig, F. Voß, N. W. Arnell, N. Bertrand, M. Best, S. Folwell, D. Gerten, S. Gomes, S. N. Gosling, S. Hagemann, N. Hanasaki, R. Harding, J. Heinke, P. Kabat, S. Koirala, T. Oki, J. Polcher, T. Stacke, P. Viterbo, G.P. Weedon and P. Yeh (2011) Multimodel Estimate of the Global Terrestrial Water Balance: Setup and First Results, J. Hydrometeor., 12, 869–884, doi:10.1175/2011JHM1324.1.

Hanasaki, N., S. Kanae, T. Oki, K. Masuda, K. Motoya, N. Shirakawa, Y. Shen, K. Tanaka (2008), An integrated model for the assessment of global water resources - Part 2: Applications and assessments, Hydrol. Earth Syst. Sci., 12, 1027–1037,

https://doi.org/10.5194/hess-12-1027-2008.

Hanasaki, N., S. Yoshikawa, Y. Pokhrel and S. Kanae (2018), A global hydrological simulation to specify the sources of water used by humans, Hydrol. Earth Syst. Sci., 22, 789–817, https://doi.org/10.5194/hess-22-789-2018.

Hanssen, A. W. and W. J. A. Kuipers (1965), On the relationship between the frequency of rain and various meteorological parameters, Koninklijk Nederlands Meteorologisch Institut, Mededelingen en Verhandelingen, 81, 2–15.

Hashino, T., A. A. Bradley and S. S. Schwartz (2007), Evaluation of bias-correction methods for ensemble streamflow volume forecasts, Hydrol. Earth Syst. Sci., 11, 939–950, doi:10.5194/hess-11-939-2007.

Heidke, P. (1926), Berechnung des Erfolges und der Gute der Windstarkevorhersagen im Sturmwarnungsdienst, Geografiska Annaler Stockholm, 8, 301–349.

Hirabayashi, Y., S. Kanae, S. Emori, T. Oki, and M. Kimoto (2008), Global projections of changing risks of floods and droughts in a changing climate, Hydrol. Sci. J., 53, 754–772.

Hirpa, F. A., P. Salomon, H. E. Beck, V. Lorini, L. Alfieri, E. Zsoter, S. J. Dadson (2018), Calibration of the Global Flood Awareness System (GloFAS) using daily streamflow data, J. Hydrol., 566, 595-606,

https://doi.org/10.1016/j.jhydrol.2018.09.052

Huffman, G. J., and D. T. Bolvin (2011), GPCP version 2.2 combined precipitation data set documentation.

Available at ftp://precip.gsfc.nasa.gov/pub/gpcp-v2.2/doc/V2.2_doc.pdf.

Huffman, G. J., R. F. Adler, D. T. Bolvin, and G. Gu (2009), Improving the global precipitation record: GPCP version 2.1, Geophys. Res. Lett., 36, L17808, doi:10:1029/2009GL040000.

Immerzeel, W. W., L. P. H. Beek, M. Konz, A. B. Shrestha, and M. F. P. Bierkens (2012), Hydrological response to climate change in a glacierized catchment in the Himalayas, Clim. Change, 110, 721–736.

IPCC (2013), Summary for Policymakers, in: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, United Kingdom and New York, NY, US.

Islam, M. S., T. Oki, S. Kanae, N. Hanasaki, Y. Agata and K. Yoshimura, K (2007), A grid-based assessment of global water scarcity including virtual water trading, Water Resour. Manage., 21, 19– 33.

Lavers, D. A., S. Harrigan, E. Andersson, D. S. Richardson, C. Prudhomme and F. Pappenberger (2019), A Vision for Improving Global Flood Forecasting, Environ. Res. Lett., 14, 121002, doi:10.1088/1748-9326/ab52b2.

Lavers, D. A., M. H. Ramos, L. Magnusson, I. Pechlivanidis, B. Klein, C. Prudhomme, L. Arnal, L. Crochemore, B. van den Hurk, A. H. Weerts, S. Harrigan, H. L. Cloke, D. S. Richardson and F. Pappenberger (2020), A Vision for Hydrological Prediction, Atmosphere 2020, 11, 237; doi:10.3390/atmos11030237.

Lehner, B., P. Döll, J. Alcamo, T. Henrichs and F. Kaspar (2006), Estimating the impact of global change on flood and drought risks in Europe: a continental integrated analysis, Climatic Change 75, 273–299.

Lemos, M. C. (2008), What influences innovation adoption by water managers? Climate information use in Brazil and the United States, J. Am. Water Resour. Assoc., 44(6), 1388–1396.

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, doi:10.1029/2008JD010969.

Liang, X., D. P. Lettenmaier, E. F. Wood, S. J. Burges (1994), A simple hydrologically based model of land surface water and energy fluxes for general circulation models, J. Geophys. Res., 99(D7), 14415–14428, doi:10.1029/94JD00483.

Livezey, R. E. (2003), Categorical events, in: Forecast Verification: A Practitioner's Guide in Atmospheric Science, edited by: Jollife, I. T. and Stephenson, D. B., Wiley, West Sussex, United Kingdom, 77–96.

Mason, I. B. (2003), Binary events, in: Forecast Verification: A Practitioner's Guide in Atmospheric Science, edited by: Jollife, I. T. and Stephenson, D. B., Wiley, West Sussex, United Kingdom, 37–73.

McBride, J. L. and E. E, Ebert (2000), Verification of quantitative precipitation forecasts from operational numerical weather prediction models over Australia, Weather Forecast., 15, 103–121.

Millner, A. (2009). What Is the True Value of Forecasts? Weather, Climate, and Society, 1(1), 22–37, http://www.jstor.org/stable/24906298.

Milly, P. C. D., and A. B. Schmakin (2002), Global modelling of land water and energy balances, Part I: The land dynamics (LaD) model, J. Hydrometeorol., 3, 283–299.

Milly, P. C. D., K. A, Dunne and A. V. Vecchia (2005), Global pattern of trends in streamflow and water availability in a changing climate, Nature, 438, 347–350, doi:10.1038/nature04312.

Mo, K. C., and D. P. Lettenmaier (2014), Hydrologic Prediction over the Conterminous United States Using the National Multi-Model Ensemble, J. Hydrometeorol., 15, 1457–1472, https://doi.org/10.1175/JHM-D-13-0197.1.

Molteni, F., T. Stockdale, M. Balmaseda, G. Balsamo, R. Buizza, L. Ferranti, L. Magnusson, K. Mogensen, T. Palmer and F. Vitart (2011), The new ECMWF seasonal forecast system (System 4), http://www.ecmwf.int/sites/default/files/elibrary/2011/11209-new-ecmwseasonal-forecast-system-system-4.pdf.

Monteith, J. L. (1981), Evaporation and surface temperature, Q. J. R. Meteorol. Soc., 107, 1–27.

Murphy, A. H. and H. Daan (1985), Forecast evaluation, in: Probability, Statistics and Decision Making in the Atmospheric Sciences, edited by: Murphy, A. H. and Katz, R. W., Westview Press, Boulder, Colorado, USA, 379–437.

Murphy, S. J., R. Washington, T. E. Downing, R. V. Martin, G. Ziervogel, A. Preston, M. Todd, R. Butterfield and J. Briden (2001), Seasonal Forecasting for Climate Hazards: Prospects and Responses, Natural Hazards 23: 171–196.

New, M., M. Hulme and P. Jones (2000), Representing twentieth-century space-time climate variability, Part 1: Development of a 1961–90 mean monthly terrestrial climatology, J. Climate, 12, 829–856.

New, M., D. Lister, M. Hulme and I. Makin (2002) A high-resolution data set of surface climate over global land areas, Climate Res., 21, 1–25.

Nijssen, B., G. M. O'Donnell, D. P. Lettenmaier, D. Lohmann, and E. F. Wood (2001a), Predicting the discharge of global rivers, J. Clim., 14, 3307–3323.

Nijssen, B., G. M. O'Donnell, A. F. Hamlet and D. P. Lettenmaier (2001b), Hydrologic sensitivity of global rivers to climate change, Clim. Change, 50, 143–175.

Norton, W. and D. Rowlands (2011), First impressions of seasonal forecasting system 4, available at: http://www.ecmwf.int/sites/default/files/elibrary/2011/ 14898-first-impressions-seasonal-forecasting-system-4.pdf.

Oki, T., Y. Agata, S. Kanae, T. Saruhashi, D. Yang and K. Musiake (2001), Global assessment of current water resources using total runoff integrating pathways, Hydrolog. Sci. J., 46, 983–995.

Pagano, T. C., H. C. Hartmann and S. Sorooshian (2002), Factors affecting seasonal forecast use in Arizona water management: A case study of the 1997–98 El Nino, Clim. Res., 21(3), 259–269.

Pahl-Wostl, C. (2007), Transitions towards adaptive management of water facing climate and global change, Water Resour. Manage., 21(1), 49–62.

Paiva, R. C. D., W. Collischonn, M. P. Bonnet, and L. G. Gonçalves (2012), On the sources of hydrological prediction uncertainty in the Amazon, Hydrol. Earth Syst. Sci., 9, 3739–3760, doi:10.5194/hessd-9-3739-2012.

Pappenberger, F., H. L. Cloke, G. Balsamo, T. Oki and T. Ngo-Duc (2011), Global routing of surface and subsurface runoff produced by the hydrological component of the ECMWF NWP system, Int. J. Climatol., 30, 2155–2174.

Pappenberger, F., M. H. Ramos, H. L. Cloke, F. Wetterhall, L. Alfieri, K. Bogner, A. Mueller, P. Salamon (2015), How do I know if my forecasts are better? Using benchmarks in Hydrological Ensemble Prediction, Journal of Hydrology (2015), doi:http://dx.doi.org/10.1016/j.jhydrol.2015.01.024.

Penman, H. L. (1948), Natural evaporation from open water, bare soil, and grass, Proc. R. Soc. London, Ser. A, 193(1032), 120–145, doi:10.1098/rspa.1948.0037.

Persson, A. and F. Grazzini (2007), User Guide to ECMWF forecast products, available at: http://old.ecmwf.int/products/forecasts/guide/ index.html.

Plummer, N., N. Tuteja, Q. J. Wang, E. Wang, D. Robertson, S. Zhou, A. Schepen, O. Alves, B. Timbal and K. Puri (2009), A Seasonal Water Availability Prediction Service: Opportunities and Challenges, 18th World IMACS / MODSIM Congress, Cairns, Australia 13-17 July 2009.

Potts, J. M., C. K. Folland, I. T. Jolliffe and D. Sexton (1996), Revised LEPS scores for assessing climate model simulations and long-range forecasts, J. Climate, 9, 34-53.

Ramos, M. H., T. Mathevet, J. Thielen, F. Pappenberger (2010), Communicating uncertainty in hydro-meteorological forecasts: mission impossible? Meteorol. Appl. 17: 223–235, doi:10.1002/met.202.

Ramos, M. H., S. J. van Andel and F. Pappenberger (2013), Do probabilistic forecasts lead to better decisions?, Hydrol. Earth Syst. Sci., 17, 2219–2232, doi:10.5194/hess-17-2219-2013.

Rayner, S., D. Lach and H. Ingram (2005), Weather forecasts are for wimps: Why water resource managers do not use climate forecasts, Clim. Change, 69(2), 197–227.

Refsgaard, J. C. (2001), Discussion of model validation in relation to the regional and global scale, in: Model Validation: Perspectives in Hydrological Science, edited by: Anderson, M. G. and Bates P. D., Wiley, West Sussex, United Kingdom, 461–483.

Ritchie, J. W., C. Zammit and D. Beal (2004), Can seasonal climate forecasting assist in catchment water management decision-making?: A case study of the Border Rivers catchment in Australia, Agr. Ecosyst. Environ., 104(3), 553–565.

Rogers, E. (2003), Diffusion of innovations, Free Press, New York.

Samaniego, L., S. Thober, N. Wanders, M. Pan, O. Rakovec, J. Sheffield, E. F. Wood, C. Prudhomme, G. Rees, H. Houghton-Carr, M. Fry, K. Smith, G. Watts, H. Hisdal, T.

Estrela, C. Buontempo, A. Marx and R. Kumar (2019), Hydrological forecasts and projections for improved decision-making in the water sector in Europe, Am. Meteo. Soc., BAMS: 2451-2471 doi:10.1175/BAMS-D-17-0274.1.

Schaefer, J. T. (1990), The critical success index as an indicator of forecasting skill, Weather Forecast., 5, 570–575.

Schumann, G., P. D. Bates, H. Apel, G. T. Aronica (2018), Global Flood Hazard Mapping, Modeling, and Forecasting, In Global Flood Hazard (eds G. J. - P. Schumann, P. D. Bates, H. Apel and G. T. Aronica), https://doi.org/10.1002/9781119217886.ch14

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, 3529–3538, doi:10.5194/hess-15-3529-2011.

Shukla, S., J. Sheffield, E. F. Wood, and D. P. Lettenmaier (2011), Relative contributions of initial hydrologic conditions and seasonal climate forecast skill to seasonal hydrologic prediction globally, Abstract H51N-05 presented at 2011 Fall Meeting, AGU, San Francisco, Calif., 5–9 Dec.

Soares, M. B. and S. Dessai (2016), Barriers and enablers to the use of seasonal climate forecasts amongst organisations in Europe, Climatic Change, 137, 89–103, https://doi.org/10.1007/s10584- 016-1671-8.

Sperna Weiland, F. C., L. P. H. van Beek, J. C. J. Kwadijk, and M. F. P. Bierkens (2010), The ability of a GCM-forced hydrological model to reproduce global discharge variability, Hydrol. Earth Syst. Sci., 14, 1595–1621, doi:10.5194/hess-14-1595-2010.

Sperna Weiland, F. C., L. P. H. van Beek, J. C. J. Kwadijk, and M. F. P. Bierkens (2011), On the suitability of GCM runoff fields for river discharge modeling; a case study using model output from HadGEM2 and ECHAM5, J. Hydrometeorol.,13, 140–154, doi:10.1175/JHM-D-10-05011.1.

Stephenson, D. B. (2000), Use of the "Odds Ratio" for diagnosing forecast skill, Weather Forecast., 15, 221–232.

Sutanudjaja, E. H., R. van Beek, N. Wanders, Y. Wada, J. H. C. Bosmans, N. Drost, R. J. van der Ent, I. E. M. de Graaf, J. M. Hoch, K. de Jong, D. Karssenberg, P. López, S. Peßenteiner, O. Schmitz, M. W. Straatsma, E. Vannametee, D. Wisser and M. F. P. Bierkens (2018), PCR-GLOBWB 2: a 5 arcmin global hydrological and water resources model. Geosci. Model Dev., 11, 2429–2453, https://doi.org/10.5194/gmd-11-2429-2018.

Troccoli, A. and P. Kallberg (2004), Precipitation correction in the ERA-40 reanalysis, ERA-40 Project Rep. Series 13, 6. ECMWF, Reading, United Kingdom.

UNDRR (2015), https://www.undrr.org/publication/sendai-framework-disaster-risk-reduction-2015-2030

Uppala, S. M., P. W. Kallberg, A. J. Simmons, U. Andrae, V. da Costa Bechtold, M. Fiorino, J.K. Gibson, J. Haseler, A. Hernandez, G. A. Kelly, X. Li, K. Onogi, S. Saarinen, N. Sokka, R.P. Allan, E. Andersson, K. Arpe, M.A. Balmaseda, A. C. M. Beljaars, L. van de Berg, J. Bidlot, N. Bormann, S. Caires, A. Dethof, M. Dragosavac, M. Fisher, M. Fuentes, S. Hagemann, E., Holm, B.J. Hoskins, L. Isaksen, P.A.E.M. Janssen, A. P. Mc Nally, J. F. Mahfouf, R. Jenne, J. J. Morcrette, N. A. Rayner, R. W. Saunders, P. Simon, A. Sterl, K.E. Trenberth, A. Untch, D. Vasiljevic, P. Viterbo and J. Woollen (2009), The ERA-40 re-analysis, Q. J. R.

Van Beek, L. P. H. (2008), Forcing PCR-GLOBWB with CRU meteorological data, available at: http://vanbeek.geo.uu.nl/suppinfo/ vanbeek2008.pdf.

Van Beek, L. P. H., and M. F. P. Bierkens (2009), The global hydrological model PCR-GLOBWB: Conceptualization, parameterization and verification, Report, Dep. of Phys. Geogr., Utrecht Univ., Utrecht, Netherlands.

Available at http://vanbeek.geo.uu.nl/suppinfo/vanbeekbierkens2009.pdf.

Van Beek, L. P. H., Y. Wada, and M. F. P. Bierkens (2011), Global monthly water stress: 1. Water balance and water availability, Water Resour. Res., 47, W07517, doi:10.1029/2010WR009791.

Viel, C., A. L. Beaulant, J. M. Soubeyroux and J. P. Céron (2016), How seasonal forecast could help a decision maker: an example of climate service for water resource management, Adv. Sci. Res., 13, 51–55, https://doi.org/10.5194/asr-13-51-2016.

Vörösmarty, C. J., B. Fekete and B. A, Tucker (1998) River Discharge Database, Version 1.1. Institute for the Study of Earth, Oceans, and Space, University of New Hampshire, Durham NH, USA.

Vörösmarty, C. J., P. Green, J. Salisbury, and R. B. Lammers (2000), Global water resources: Vulnerability from climate change and population growth, Science, 289, 284–288, doi:10.1126/science.289.5477.284.

Wada, Y., L. P. H. Van Beek, D. Viviroli, H.H. Durr, R. Weingartner, and M. F. P. Bierkens (2008), Water Stress over the Year: Quantitative Analysis of Seasonality and Severity on a Global Scale, MSc Thesis, University Utrecht, Utrecht, Netherlands, available at: http://igitur-archive.library. uu.nl/student-theses/2010-0308-200229/UUindex.html.

Wada, Y., L. P. H. Van Beek, and M. F. P. Bierkens (2011), Modelling global water stress of the recent past: on the relative importance of trends in water demand and climate variability, Hydrol. Earth Syst. Sci., 15(12), 3785-3805, https://doi.org/10.5194/hess-15-3785-2011.

Wada, Y., M. H. Lo, P. F. Yeh, J. T. Reager, J. S. Famiglietti, R. J. Wu and Y. H. Tseng (2016), Fate of water pumped from underground and contributions to sea-level rise, Nature Clim. Change 6, 777–780, https://doi.org/10.1038/nclimate3001

Wanders, N., S. Thober, R. Kumar, M. Pan, J. Sheffield, L. Samaniego, and E. F. Wood (2019b), Development and evaluation of a pan-European multi-model seasonal hydrological forecasting system, J. Hydrometeor., 20, 99–115, https://doi.org/10.1175/JHM -D-18-0040.1.

Wanders, N., M. T. H. van Vliet, Y. Wada, M. F. P. Bierkens, and L. P. H. van Beek (2019b), High-resolution global water temperature modeling, Water Resources Research, 55, 2760–2778, https://doi.org/10.1029/2018WR023250

Weerts, A., H. Winsemius, E. Dutra, J. Beckers, R. Brolsma, R. van Beek, F. Pappenberger, R. Westerhoff and M. Bierkens (2013), Seasonal Predictability of Water Scarcity at the Global Scale, Geophysical Research Abstracts Vol. 15, EGU2013-6060-1.

Werner, M., J. Schellekens, P. Gijsbers, M. van Dijk, O. van den Akker, and K. Heynert (2013), The Delft-FEWS flow forecasting system, Environ. Modell. Software, 40, 65–77, doi:10.1016/j.envsoft.2012.07.010.

Wesseling, C. G., D. Karssenberg, W. P. A. van Deursen, and P. A. Burrough (1996), Integrating dynamic environmental models in GIS: The development of a Dynamic Modeling language, Trans. GIS, 1, 40–48.

Whateley, S., R. N. Palmer, C. Brown (2015), Seasonal Hydroclimatic Forecasts as Innovations and the Challenges of Adoption by Water Managers, Journal of Water Resources Planning and Management, Volume 141 Issue 5, doi: 10.1061/(ASCE)WR.1943-5452.0000466.

White, C. J., H. Carlsen, A. W. Robertson, R. J. Klein, J. K. Lazo, A. Kumar, F. Vitart, E. Coughlan de Perez, A. J. Ray, V. Murray, S. Bharwani, D. MacLeod, R. James, L. Fleming, A. P. Morse, B. Eggen, R. Graham, E. Kjellström, E. Becker, K. V. Pegion, N. J. Holbrook, D. McEvoy, M. Depledge, S. Perkins Kirkpatrick, T. J. Brown, R. Street, L. Jones, T. Remenyi, I. Hodgson-Johnston, C. Buontempo, R. Lamb, H. Meinke, B. Arheimer and S. E. Zebiak (2017), Potential applications of subseasonal-to-seasonal (S2S) predictions, Meteorol. Appl., 24, 315–325, https://doi.org/10.1002/met.1654.

WMO (2008), Disaster Risk Reduction under Current and Changing Climate Conditions, Bulletin Vol 57 (2). https://public.wmo.int/en/bulletin/disaster-risk-reduction-under-current-and-changing-climate-conditions.

Wood, E. F., D. P. Lettenmaier and V. G. Zartarian (1992), A land-surface hydrology parameterization with sub-grid variability for general circulation models, J. Geophys. Res., 97, 2717–2728.

Wood, A. W., E. 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.

Wood, A. W., A. Kumar, and D. P. Lettenmaier (2005), A retrospective assessment of National Centers for Environmental Prediction climate model-based ensemble hydrologic forecasting in the western United States, J. Geophys. Res., 110, D04105, doi:10.1029/2004JD004508.

Wood, A. W., and D. P. Lettenmaier (2008), An ensemble approach for attribution of hydrologic prediction uncertainty, Geophys. Res. Lett., 35, L14401, doi:10.1029/2008GL034648.

Yarnal, B., A. L. Heasley, R. E. O'Connor, K. Dow and C. L. Jocoy (2006), The potential use of climate forecasts by community water system managers, Land Use Water Resour. Res., 6, 3.1–3.8.

Yuan, X., E. F. Wood, N. W. Chaney, J. Sheffield, J. Kam, M. Liang and K. Guan (2013), Probabilistic Seasonal Forecasting of African Drought by Dynamical Models, J. Hydrometeorol., 14, 1706– 1720, https://doi.org/10.1175/JHM-D-13-054.1.

Yuan, X., E. F. Wood and Z. Ma (2015a), A review on climate model-based seasonal hydrologic forecasting: physical understanding and system development, WIREs Water, 2, 523–536, https://doi.org/10.1002/wat2.1088.

Yuan, X., J. K. Roundy, E. F. Wood, J. Sheffield (2015b), Seasonal Forecasting of Global Hydrologic Extremes: System Development and Evaluation over GEWEX Basins, B. Am. Meteorol. Soc., 96, 1895–1912, https://doi.org/10.1175/BAMS-D-14-00003.1.

Zwiers, F. W. and H. von Storch (2004), On the role of statistics in climate research, Int. J. Climatol., 24, 665–680.

Summary

In our changing world, humans experience increasingly the negative consequences of floods and droughts. While short-term and mid-term hydrological forecasts can be vital for risk management on the event scale, seasonal forecasts are necessary to increase global preparedness and to inform local mitigation. Such forecasts with longer lead times extend over several months, and cover larger areas from continental to global scale. These global seasonal forecasts should eventually become operational, thus providing a constant and consistent feed of information. The quality of this information, however, needs to be assessed and this thesis explores the potential of global hydrological models in operational seasonal forecasting applications. The research aims to assess the skill and value of seasonal streamflow forecasts produced by global hydrological models, as well as to investigate possible ways to improve the current skill and value.

To assess the prospect of applying a global hydrological model for seasonal forecasting, global terrestrial hydrology is simulated with the model PCR-GLOBWB, which is very similar to other large scale hydrological models in its model structure and parameterization. As a first step, the model is forced with a meteorological data set based on historical observations and its skill in simulating the hydrology is analysed by adopting methods that were primarily developed for the verification of meteorological forecasts. The skill is assessed based on monthly discharges for twenty large river basins across the world. For these basins, PCR-GLOBWB cannot forecast the historical hydrographs adequately for all basins but the results indicate higher skills can be attained in forecasting the occurrence of monthly streamflow anomalies that are indicative of floods and droughts. The use of global hydrological models for operational forecasting of monthly hydrological extremes is found promising, and the prospects for seasonal forecasting with PCR-GLOBWB or other comparable models are assessed to be positive.

The simulated hydrological response depends on both the initial hydrological conditions and the meteorological forcing. Uncertainty in both inputs is important and is evaluated by comparing Ensemble Streamflow Prediction (ESP) and reverse ESP forecast ensembles with retrospective model simulations driven by meteorological observations. The results are analysed in the context of prevailing hydroclimatic conditions for larger rivers across the globe. The influence of the initial conditions and meteorological forcing on the hydrological forecasting skill is found to vary considerably according to location, season and lead time. As the meteorological

forcing deteriorates in quality with longer lead times, the initial conditions gain importance; without correct initial conditions, the relatively better quality of the meteorological forcing early in the forecasting period will be largely obscured. In particular snow and ice are important sources of water that contribute to the quality of long-term forecasts, but their influence varies depending on the nature of the river basin considered. For arctic and snow fed rivers, forecasts of high flows may benefit from assimilation of snow and ice data. In some snow fed basins where the onset of melting is highly sensitive to temperature changes, forecast skill depends on better climate prediction. Groundwater is a slow hydrological process that also strongly influences the skill. In very large rivers, initial surface water and groundwater states are important contributors to skill. In monsoonal basins, the variability of the monsoon dominates forecasting skill, except for those where snow and ice contribute to streamflow.

The total skill of a forecasting system is affected by the errors in the model structure and parameterization, in the initial conditions and in the meteorological forcing. When the total skill is assessed in actual forecasting mode, actual seasonal meteorological forecasts are used as input into the global hydrological model PCR-GLOBWB. The model is forced with S3 seasonal meteorological forecast ensembles from the European Centre for Medium-range Weather Forecasts (ECMWF) as well as with probabilistic meteorological ensembles obtained following the ESP procedure. Ensemble forecasts of monthly discharges for twenty large rivers of the world are produced with lead times of up to six months. The skill of ECMWF S3 forecasts compared to the reference ESP forecasts are quantified using the Brier skill score (BSS), both for high and low flows. Analysis of these results suggest that forecasting skill decreases with increasing lead time and that it varies considerably by region and season. The performance of ECMWF S3 forecasts is close to that of the ESP forecasts. In the current setup, the forecasting skill is limited and needs to be improved before forecasts can be adopted for water management applications. However, even with little added skill, forecasts may still be useful for end-users, allowing them to decide for themselves if they should take the risk of using the forecast information.

The success of a hydrological forecasting system will ultimately be determined not only by its skill but also by its value to inform decision-making for water management. This thesis concludes by presenting a study on the value of seasonal streamflow forecasts, where the interaction between skill and value is explored and possible ways to improve the value of seasonal hydrological forecasts on a global scale for water-related applications are discussed with an emphasis on flood and drought mitigation. The current ability of seasonal streamflow forecasting systems to predict the right category of an event months ahead is potentially valuable for many water-related applications. Seasonal hydrological forecasting on a global scale could be especially valuable for transboundary river basins as well as for developing regions of the world,

where no effective local hydrological forecasting systems exist. The realization of the potential added value depends largely on the collaboration between forecast producers and users, during each link in the forecast chain, i.e., construction of forecasts, their communication, their adoption, their use in decision-making and finally, during feedback from stakeholders and decision makers to forecasters and researchers.

Samenvatting

In onze veranderende wereld ondervindt de mens steeds meer de negatieve gevolgen van overstromingen en droogte. Hoewel hydrologische voorspellingen op korte en middellange termijn van vitaal belang kunnen zijn voor risicobeheer bij een gebeurtenis, is er behoefte aan seizoensvoorspellingen om de wereldwijde paraatheid te vergroten en lokale mitigatie form te geven. Zulke voorspellingen hebben langere aanlooptijden die zich uitstreken over meerdere maanden, en zij beslaan grotere gebieden, op continentale tot wereldschaal. Dergelijke wereldwijde seizoensvoorspellingen zouden uiteindelijk operationeel moeten worden, zodat een constante en consistente stroom van informatie kan worden geboden. De kwaliteit van deze informatie moet echter worden beoordeeld en dit proefschrift verkent het potentieel van wereldwijde hydrologische modellen in operationele seizoensvoorspellingen op wereldschaal. Het onderzoek heeft tot doel de nauwkeurigheid en de waarde van seizoensgebonden stroomvoorspellingen te beoordelen zoals die worden geproduceerd door wereldwijde hydrologische modellen, en om mogelijke manieren te onderzoeken om de huidige nauwkeurigheid en waarde hiervan te verbeteren.

Om de toepassing van een mondiale hydrologische model binnen seizoensvoorspellingen te beoordelen, is gebruik gemaakt van het mondiale hydrologische model PCR-GLOBWB, dat qua modelstructuur en parameterisatie sterk lijkt op andere grootschalige hydrologische modellen. Als eerste stap is dit hydrologische model toegepast met klimaatforceringen op basis van historische waarnemingen en is zijn nauwkeurigheid om de hydrologie te simuleren geanalyseerd met methoden die voornamelijk ontwikkeld zijn voor de verificatie van meteorologische voorspellingen. De nauwkeurigheid is beoordeeld op basis van de maandelijkse rivierafvoeren voor twintig grote stroomgebieden over de hele wereld. Voor deze stroomgebieden kan PCR-GLOBWB de historische afvoerreeksen niet volledig adequaat voorspellen, maar de resultaten geven aan dat hogere nauwkeurigheden verkregen kunnen worden bij het voorspellen van het optreden van maandelijkse anomalieën. Het gebruik van mondiale hydrologische modellen voor operationele voorspellingen van maandelijkse hydrologische extremen is veelbelovend, en de vooruitzichten voor seizoensvoorspellingen met PCR-GLOBWB of andere vergelijkbare modellen kan als positief worden beoordeeld.

De gesimuleerde hydrologische respons hangt af van zowel de beginvoorwaarden als de meteorologische forcering. Onzekerheid in beide invoeren is belangrijk en is geëvalueerd door Ensemble Streamflow Prediction (ESP)- en reverse-ESP-voorspellingsensembles te vergelijken met retrospectieve modelsimulaties op basis van meteorologische waarnemingen. De resultaten zijn geanalyseerd in relatie tot de heersende hydro-klimatologische omstandigheden voor grotere rivieren over de hele wereld. De invloed van de beginvoorwaarden en meteorologische forcering op de hydrologische voorspellingsnauwkeurigheid varieert aanzienlijk afhankelijk van de locatie, het seizoen en de aanlooptijd. Naarmate de meteorologische forcering in kwaliteit verslechtert met langere aanlooptijden, winnen de beginvoorwaarden aan belang; zonder de juiste beginvoorwaarden, zal de relatief betere kwaliteit van de meteorologische forcering vroeg in de voorspellingsperiode grotendeels teniet worden gedaan. Met name sneeuw en ijs zijn belangrijke waterbronnen die bijdragen aan de temporele variabiliteit, maar hun invloed varieert per stroomgebied. Voor arctische en voor met sneeuw gevoede rivieren kunnen voorspellingen van hoge afvoeren verbeteren door de assimilatie van sneeuw- en ijsgegevens. In sommige met sneeuw gevoede bekkens waar het begin van het smelten zeer gevoelig is voor temperatuurveranderingen, hangt de vaardigheid in het voorspellen af van betere klimaatvoorspellingen. Grondwater is een ander langzaam hydrologisch proces dat ook de nauwkeurigheid sterk beïnvloedt. In zeer grote rivieren zijn de initiële oppervlaktewater- en grondwatertoestanden belangrijke bronnen van nauwkeurigheid. In moessongebieden domineert de variabiliteit van de moesson de nauwkeurigheid bij het voorspellen, behalve in die gebieden waar sneeuw en ijs bijdragen aan de rivierafvoer.

De totale nauwkeurigheid wordt beïnvloed door de fouten in de modelstructuur en parameterisatie, door fouten in de beginvoorwaarden en door fouten in de meteorologische forcering. Bij de beoordeling van de totale vaardigheid in de werkelijke voorspellingsmodus, zijn actuele seizoensmeteorologische voorspellingen gebruikt als input voor het mondiale hydrologische model PCR-GLOBWB. Het model is geforceerd met S3-seizoensmeteorologische voorspellingsensembles van het European Centre for Medium-range Weather Forecasts (ECMWF). Daarnaast zijn probabilistische meteorologische ensembles verkregen volgens de ESP-procedure. Er zijn ensemblevoorspellingen van maandelijkse afvoeren gemaakt voor twintig grote rivieren in de wereld, met aanlooptijden tot wel zes maanden. De vaardigheid van de ECMWF S3-voorspellingen is vergeleken met de referentie-ESP-voorspellingen en is gekwantificeerd met behulp van de Brier-vaardigheidsscore (BSS), zowel voor hoge als lage afvoeren. Uit analyse van deze resultaten blijkt dat de voorspellingsvaardigheid afneemt met toenemende aanlooptijd en dat deze aanzienlijk varieert per regio en seizoen. De prestaties van de ECMWF S3-voorspellingen blijken vergelijkbaar te zijn met die van de ESP. In de huidige opzet is de voorspellende nauwkeurigheid dus beperkt en zal deze verbeterd moeten worden voordat zij toegepast kunnen worden voor toepassingen op het gebied van waterbeheer. Echter, zelfs zonder directe nauwkeurigheid, kan de voorspellingen nog steeds nuttig zijn voor eindgebruikers, waardoor ze zelf kunnen beslissen of ze het risico moeten nemen om de voorspellingsinformatie te gebruiken.

Het succes van een hydrologisch voorspellingssysteem zal uiteindelijk niet alleen worden bepaald door zijn nauwkeurigheid, maar ook door de waarde waarmee besluitvorming voor waterbeheer ondersteund kan worden. Tenslotte, presenteert dit proefschrift een onderzoek naar de waarde van seizoensgebonden afvoervoorspellingen, waarbij de interactie tussen nauwkeurigheid en waarde wordt onderzocht en mogelijke manieren worden besproken om de waarde van seizoensgebonden hydrologische voorspellingen op wereldschaal voor watergerelateerde toepassingen te verbeteren, met de nadruk op het matigen van de impact van overstromingen en droogte. Het huidige vermogen van hydrologische seizoensvoorspellingssystemen om maanden van tevoren de juiste categorie van een gebeurtenis te voorspellen, is potentieel waardevol voor veel watergerelateerde toepassingen. Seizoensgebonden hydrologische voorspellingen op mondiale schaal kunnen vooral waardevol zijn voor grensoverschrijdende stroomgebieden, maar ook voor ontwikkelingsgebieden, waar nog geen effectieve lokale hydrologische voorspellingssystemen bestaan. De potentiële toegevoegde waarde hangt grotendeels af van de wisselwerking tussen de producenten en de gebruikers van voorspellingen, tijdens elke schakel in de voorspellingsketen, dat will zeggen bij het opstellen van voorspellingen, bij de kennisoverdracht en adoptie, bij het gebruik van voorspellingen in het besluitvormingproces en bij het terugkoppelen van de wensen en ervaringen van eindgebruikers als belanghebbenden en besluitvormers aan de onderzoekers en de opstellers van voorspellingen.

Acknowledgements

I would like to thank all my colleagues, friends and family who have supported me in my PhD journey.

First and foremost, I thank my promotor Marc Bierkens and my direct supervisor Rens van Beek for their invaluable guidance, support, kindness and patience. I had felt from the start that I am really lucky to have both of you as my supervisors, but I could not have imagined how dear you would become to my heart through all these years that you have stood by me during the challenges I had, professional and personal.

I would like to express my gratitude to Prof. dr. Hannah Cloke, Prof. dr. Ann van Griensven, Prof. dr. Hans Middelkoop, Prof. dr. ir. Remko Uijlenhoet and Dr. Micha Werner for being part of my committee and reviewing my thesis.

I would also like to acknowledge the financial support and the wonderful work environment provided by Deltares. Thanks to all my colleagues and friends at Utrecht University as well as at Deltares, specially to my mentors Jaap Kwadijk, Eelco van Beek and Albrecht Weerts, and my room-mate Christiana Photiadou.

Thanks to all my friends and family who supported and encouraged me. Canım annem, canım babam, Nağmem, Selimim, Kerim G., Işık, Aydan, Yüc, Serkan, Nemana, Ayşe abla, teşekkürler.

And finally, dear Mohamed and Kerim, thank you for everything. This thesis is dedicated to you.

About the author

Naze Candogan Yossef was born in Ankara in 1970. She graduated as a civil engineer from the Middle East Technical University in 1995. She then studied at IHE Delft Institute for Water Education, where she obtained a master's degree in 1997 and also met her husband, Mohamed Yossef. She then came back to Ankara and started work at the Middle East Technical University, for the Med-Coast project for the international management of the coastal regions of the Mediterranean and the Black Sea. After marrying in 1998, she moved to Egypt and worked at the Cairo office of the Turkish contractor company STFA Enercom. In 1999, she travelled to Istanbul where she worked for the STFA Marine Construction Company.

In 2001, she relocated to the Netherlands with her newborn son Kerim Yossef to join her husband who had started his PhD study at TU Delft. Taking a parenting break from work, she had the opportunity to pursue her passion to study archaeology. She attended the international master's program Sciences in Archaeology at Leiden University. During the summer months, she joined the Alacahoyuk archaeological expedition led by Ankara University, where she worked as an archaeologist and hydraulic engineer for the excavation and environmental restoration of a 3200 years old Hittite dam. In 2006, she obtained her master's degree in Archaeology, with a thesis titled Hittite Water Works. Later she did an internship at Deltares, Hydrology division, where she took part in Lake Nasser Flood and Drought Control Project. She also worked for the Water Management division at Deltares as a flexible employee, and produced an analysis report on the water sector in Turkey.

In September 2008, she started her PhD research at Utrecht University under the supervision of Prof. Marc Bierkens and Dr Rens van Beek. Her research led to three publications in the journals Hydrology and Earth Systems Sciences and Water Resources Research. After a long break due to severe health problems, she resumed her studies at Utrecht as a visiting researcher and completed the final part of her thesis.