Hydrological impacts of climate change

Interpretation of uncertainties introduced by global models of climate and hydrology

Hydrologische effecten van klimaatverandering
Interpretatie van onzekerheden geïntroduceerd door globale modellen voor klimaat en hydrologie

(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. G.J. van der Zwaan,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op

maandag 19 december 2011 des middags te 12.45 uur

door

Frederiek Christianne Sperna Weiland

geboren op 8 november 1982 te Goes

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

Co-promotoren:
Dr. L.P.H. van Beek
Dr. J.C.J. Kwadijk

Hydrological impacts of climate change

Interpretation of uncertainties introduced by global models of climate and hydrology

Utrecht Studies in Earth Sciences

ISSN 2211-4335

Utrecht Studies in Earth Sciences 006

Hydrological impacts of climate change

Interpretation of uncertainties introduced by global models of climate and hydrology

Frederiek Sperna Weiland

Utrecht 2011

This work was funded by the Dutch Research Institute Deltares as part of its strategic research program.

ISBN 978-90-6266-287-6

Printed in the Netherlands by www.wps.nl, Zutphen.

Contents

List of figures

List of Tables

1 Introduction

1. 1 Context: Climate change and its hydrological impacts

Observed climate change

Meteorological observations show a global average temperature increase of 0.74°C over the last century. After 1970 global warming even accelerated with an average decadal temperature increase of 0.15°C. Most evident warming occurred over the mid- and low-latitudes and, although less obvious due to large year-to-year variation, strongest warming occurred over Asia and northern North-America. This global temperature increase is mainly the result of increasing greenhouse gas concentrations (IPCC, 2007a). The IPCC also reports high confidence on ongoing changes in a number of hydrological systems. Increased runoff and earlier spring peak discharges have been observed which can be related to increased snow and glacier melt. Furthermore, precipitation changes have become more pronounced over the last century. Increases are observed over the eastern parts of North and South America, northern Europe and northern and central Asia, whereas decreases occurred in the Sahel, the Mediterranean, southern Africa and parts of southern Asia. Globally it is likely that, especially after 1970, precipitation and temperature changes caused an increase in drought affected areas (IPCC, 2007a).

Observed hydrological extremes

Climate change may also influence the frequency of hydrological extremes, such as the disastrous flood and drougth events over the past year as for example the Indus floods in July 2010. During these floods one-fifth of Pakistan's total land area was inundated as a result of extremely heavy monsoon rains. About 20 million people were affected and almost 2000 people died (Singapore red cross, 2010). In January 2011 the worst Australian floods observed over the last fifty years inundated much of south and central Queensland. The floods were triggered by an unusual heavy monsoon, followed by much higher-than-normal rainfall afterwards. About 200.000 people were affected and at least 15 people died (CBC news, 2011). In April and May 2011 one of the largest and most damaging floods of the last century was recorded for the Mississippi. Two heavy precipitation storms passed over the Mississippi watersheds followed by large discharge

rises due to springtime snowmelt in May. So far, the estimated damage is more than 2 billion dollar (USA today, 2011).
At the time of writing, July 2011, West Africa is facing one of the worst droughts of the last decades. Currently, about 10 million people face starvation (Guardian, 2011). On the longterm droughts are likely to result in even larger economic losses than floods, due to among others the negative longterm impacts of malnutrition, such as decreased cognitive skills, less productive manual labor and loss of livestock (Loayza et al., 2009). Furthermore, especially in dry regions with weak governance, low rainfall rates may increase the chance on conflicts (World Bank, 2010).
To be prepared for possible increases in the frequency of hydrological extremes, adaptation strategies are sought for that minimize the damaging effects. Hydrological climate impact assessments can provide useful insights on the likelihood of change in occurrences and frequencies of such disastrous events.

Assessment of future climate change impacts
Large uncertainties exist in climate changes projected for the next century. Yet, there is confidence on future global temperature increase and consequently reductions of cold days and increases of extreme hot days are to be expected. Although uncertainties for future precipitation changes are larger than for temperature, an intensification of the global hydrological cycle is projected (Huntington, 2006; Arnell, 1999a). Several studies project increases in the intensity and frequency of extreme precipitation events (Trenberth et al., 2003; Allan and Soden, 2008; Meehl et al., 2000), these increases will particularly occur at high latitudes and in Monsoon influenced tropical regions (IPCC, 2007a).
Climate change will also affect global runoff regimes (e.g. annual average discharge cycles). Several hydrological impact assessments report on decreases in low flows for Southern Europe, parts of Africa and the Southwestern US, increases in runoff for Arctic and Monsoon influenced basins and a shift towards earlier peak discharges in northern regions due to an increased part of precipitation falling as rain instead of snow (Arnell, 2003; Nijssen et al., 2001; Vörösmarty et al., 2000; Alcamo and Henrichs, 2007). As a consequence, flood events may occur more often and increasing droughts may have devastating impacts on food security, freshwater ecosystems and economic development (World Bank, 2010).
So far, many studies investigated the hydrological consequences of climate change. Such assessments should preferably be based on an ensemble of general circulation models, also called global climate models (GCMs), due to the large deviations between GCMs (Murphy, 2004). Still, many applied water management studies focus on a few GCMs or the 'best performing' model only, herewith, ignoring the large uncertainty between GCMs. High investments are being made to implement adaptation measures which are defined based upon the outcomes of these studies. Yet, the outcomes could have been different if other GCMs would have been used and consequently the required adapta-

tion measures might be different from the implemented ones.
Moreover, in case multiple GCMs are used, future change is often simply quantified with the algorithmic mean of an ensemble of projected changes, as no common practice exists for the presentation of the uncertainties of projected changes to water managers. In addition, biases in GCM data are corrected with a variety of downscaling techniques, sometimes without proper investigation of their origin. Yet, these biases do provide important information on the quality of physical process representations within GCMs and possible guidance for model improvements and downscaling strategies.

1.2 The modeling chain of hydrological impact assessments

Within this thesis the inherent uncertainties of the different steps in the modeling chain of hydrological impact assessments are investigated in more detail. First, the usability of raw GCM data for hydrological impact studies is evaluated. Secondly, the global hydrological impacts of climate change are assessed from an ensemble of 12 GCMs. Finally, two methods for the interpretation of the uncertainties accompanying the projected changes are introduced. The remainder of this chapter discusses the uncertainties in the modeling chain of hydrological impact studies (Fig. 1.1) and at the end of this chapter thesis objectives and research questions are presented.

The modeling chain
The hydrological impact modeling chain exists of several steps, as displayed in Fig. 1.1. The driving force behind climate change are the increases in greenhouse gas emissions which are, for application in climate impact studies, pre-scribed by the IPCC SRES emission scenarios (IPCC, 2000). These emission scenarios are converted to atmospheric greenhouse gas concentrations with environmental models. The resulting greenhouse gas concentrations are used as boundary conditions for GCM simulations

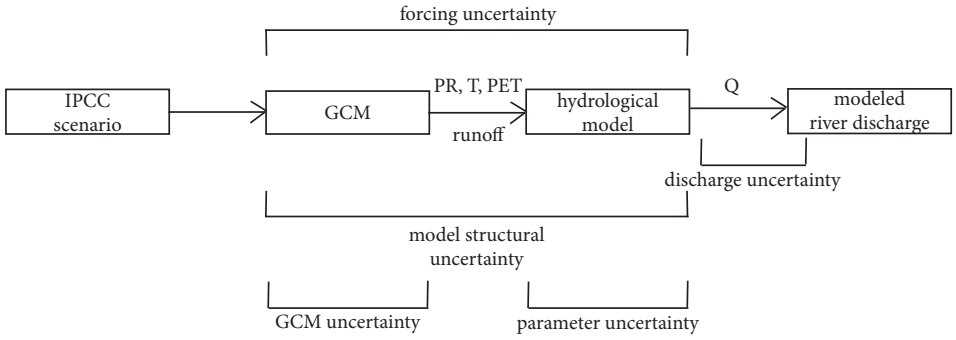


Figure 1.1 Schematization of the hydrological part of the modeling chain of climate impact assessments together with the uncertainties investigated in this thesis.

of the future climate. On their turn the, with GCMs calculated, future time series of meteorological variables can be employed as input for hydrological models. Based on this input a hydrological model produces discharge time-series which are used to assess future changes in river discharge and water availability.

IPCC emission scenarios

Climate change is induced by increasing greenhouse gas concentrations. Yet, future emissions of greenhouse gasses are highly depended on technological changes and socio-economic developments. Based on literature, the IPCC defined multiple scenarios which provide storylines of possible future development pathways. Focus is on population increases, economic growth, development of green technologies, convergence among regions, capacity building and social and cultural interactions (IPCC, 2000).
Within this thesis the IPCC SRES emission scenario A1B has been selected. This scenario is characterized by a rapid economic growth, an increasing population for the first half of the century and a rapid introduction of green and sustainable technologies. As can be seen from Fig. 1.2, the A1B scenario projects relatively large CO_2 emission increases, particularly for the first half of the century. Yet, for the period 2000 to 2006, observed emission increases were close to emissions projected by the worst case IPCC emission scenario (Global Carbon Project, 2008). This study focuses on a single emission scenario as differences between emission scenarios are often smaller than differences between GCMs (Arnell, 2003).

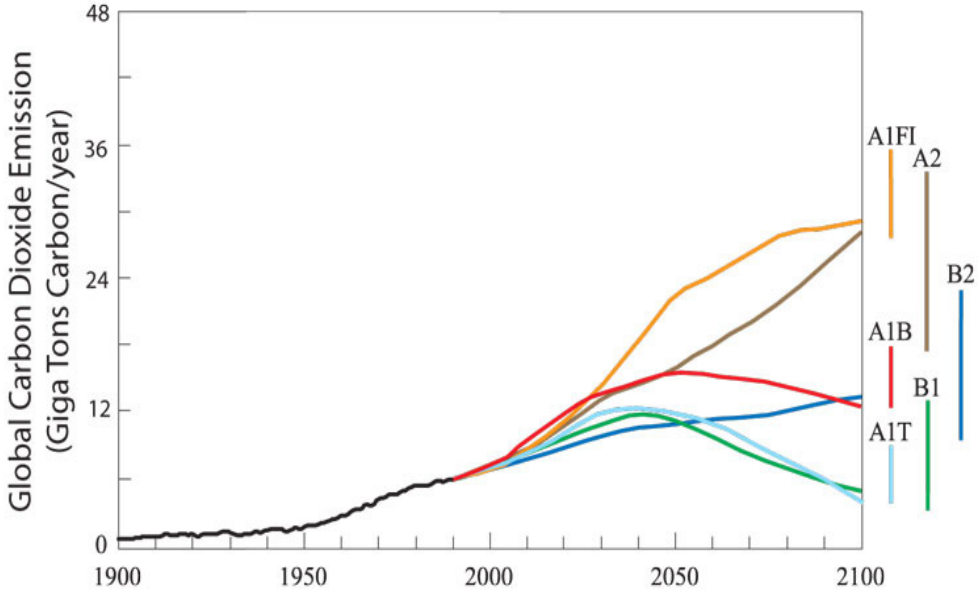


Figure 1.2 Global CO_2 emissions for the 21th century for different IPCC SRES emission scenarios. Vertical bars indicate uncertainty ranges of projected emissions for 2100 (U.S. Global Change Research Program 2008).

General Circulation Models

Nearly all impact assessments rely on datasets from GCMs. Within GCMs physical processes of the atmosphere, ocean and land surface are described by mathematical equations which solve energy and mass balances (Legates, 2000). All equations are unavoidable simplifications of real world processes. Within GCMs the world is discretized in grids and layers. Oceans and atmospheres are separately represented by multiple vertical layers of which the number ranges between 9 and 45 within the different GCMs. The GCM horizontal resolution is, compared to hydrological models, relatively coarse. In GCMs, climate change is induced by boundary conditions for aerosol and greenhouse gas concentrations which show an increase over time (Fig. 1.2). For the current climate these boundary conditions follow observed greenhouse gas and aerosol concentrations, for future periods these concentration increases are prescribed by the IPCC emission scenarios (IPCC, 2007b).

For this study 12 GCMs were selected (see Table 1.1) for which daily time series of the required meteorological variables could be retrieved from the data-portal of the program for climate model diagnosis and intercomparison (PCMDI).

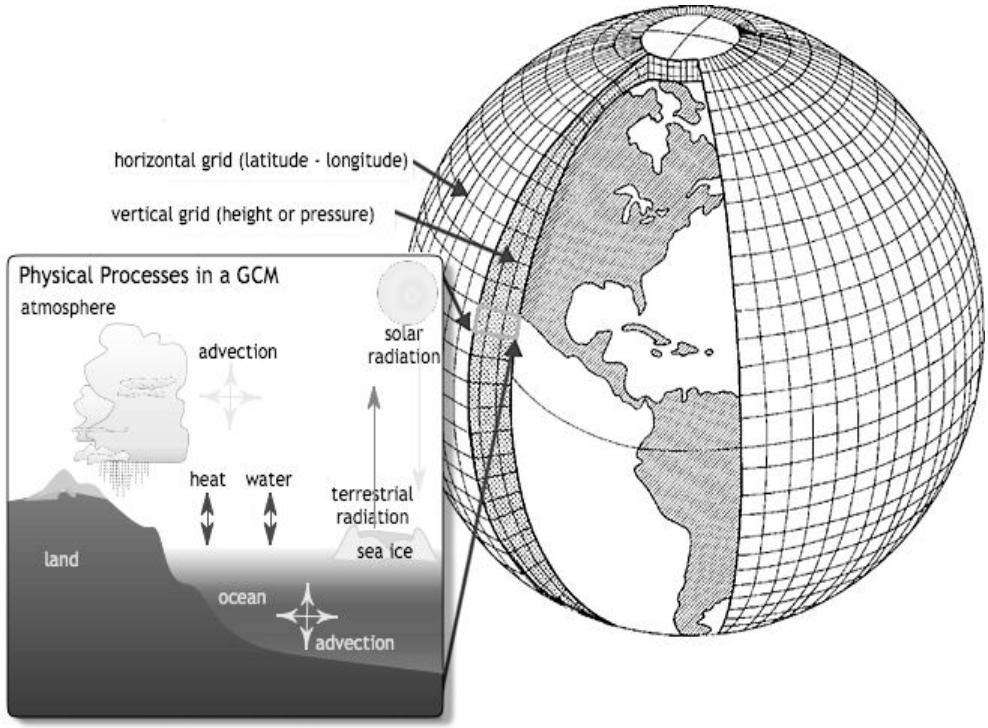


Figure 1.3 Schematic representation of GCM (ClimateChangeDispatch, 2011; Wikipedia, 2009).

Table 1.1 Overview of selected GCMs.

Model	*Institute*	*Country*	*Horizontal Resolution (degrees)**	*Vertical Resolution (nr atmospheric layers, nr ocean layers)**	*Acronym*
BCM2.0	Bjerknes Centre for Climate Research	Norway	2.8	(31 , 35)	BCCR
CGCM3.1	Canadian Centre for Climate Modelling and Analysis	Canada	3.75	(31 , 29)	CCCMA
CGCM2.3.2	Meteorological Research Institute	Japan	2.8	(30 , 23)	CGCM
CSIRO-Mk3.0	Commonwealth Scientific and Industrial Research Organization	Australia	1.875	(18 , 31)	CSIRO
ECHAM5	Max Planck Institute	Germany	1.5	(31 , 40)	ECHAM
ECHO-G	Freie Universität Berlin	Germany	3.75	(19 , 20)	ECHO
GFDL-CM 2.1	Geophysical Fluid Dynamics Centre	USA	1.0	(24 , 50)	GFDL
GISS-ER	Goddard institute for Space Studies	USA	4 * 5	(20 , 13)	GISS
IPSL-CM4	Institute Pierre Simon Laplace	France	2.5 * 3.75	(19 , 19)	IPSL
MIROC3.2	Center of Climate System Research	Japan	2.8	(20 , 43)	MIROC
CCSM3	National Center for Atmospheric Research	USA	1.4	(26 , 40)	NCAR
HADGEM1	Met Office's Hadley Centre for Climate Prediction	UK	1.25 * 1.875	(38 , 40)	HADGEM

* Parkinson et al. (2006)

GCM uncertainties

Different GCMs can produce quite varying and even contradictory results, especially for precipitation (Covey et al., 2003). The differences are amongst others a result of up-scaling of real world processes to the GCM resolution. Particularly small scale processes, as for example rain and cloud formation, cannot be represented well and are hard to parameterize (Pielke et al., 2009; Beven, 2011). In addition, GCMs differ in their schematization, resolution, process formulation and parameterization.

As a consequence, GCM derived changes should be interpreted with care. For example, the IPCC provides a projected global average mean temperature increase for the A1B scenario of 2.8°C and accompanies this change with an uncertainty range of 1.7°C to 4.4°C which is obtained from projections of all GCMs included in the IPCC AR4 assessment report (IPCC, 2007a). Over time, multiple studies concluded that for climate change assessments a large number of GCMs should be used. Such a large ensemble samples the widest possible range of modeling uncertainties and thereby provides a reliable specification of the spread of projected changes (Murphy et al., 2004).

The global hydrological model PCR-GLOBWB

The global hydrological model (GHM) PCR-GLOBWB has been developed at Utrecht University (Van Beek et al., 2011). PCR-GLOBWB follows in a long line of existing GHMs, as for example Macro-PDM (Arnell, 1999a), WBM (Vörösmarty et al., 1998b), VIC (Nijssen et al., 2001) and WGHM (Döll et al., 2003).

PCR-GLOBWB is able to reproduce the variability of global annual discharge cycles and changes therein well, which is the main focus of this thesis. And, in addition to the GHMs listed above, PCR-GLOBWB contains an advanced scheme for the subgrid parameterization of surface runoff, interflow and baseflow and an explicit routing scheme for surface water flow using the kinematic wave approximation that includes retention in, and evaporation loss from, wetlands, lakes and reservoirs. Similar to WBM

and WGHM PCR-GLOBWB has a grid resolution of 0.5 degrees. The model calculates river discharge at a daily time-step by routing the runoff fluxes which become available after the water balances for all individual cells are solved (see fig. 1.4). Extensive descriptions of the model can be found in chapters 2 - 6 and Van Beek et al. (2011).

Hydrological modeling within GCMs

Over the last decades it has been recognized that the hydrological cycle plays a major role within the global climate system. For instance, the hydrological cycle influences energy exchange between the land surface and atmosphere (Kite, 1998). Full inclusion of these energy exchanges requires the presence of land surface schemes (LSSs) in GCMs which define the partitioning of precipitation in evaporation, storage and runoff (van den Hurk et al., 2004; Pappenberger et al., 2009). As a consequence the representation of hydrological processes within GCMs has been improved over time and LSSs of GCMs can be as advanced as hydrological models (Clark and Gedney, 2008; Hagemann and Gates, 2003). One of the fluxes calculated by LSSs in GCMs is the cell specific runoff. River discharge derived from this GCM runoff flux may be of use to hydrological impact studies. It may even be possible to replace runoff derived with hydrological models with GCM runoff.

Hydrological model uncertainties

Model uncertainties in hydrological climate impact studies are not restricted to GCMs, a hydrological model also introduces uncertainties (Fig. 1.1). Within GHMs real world processes have to be represented on a model grid with a slightly higher resolution than that of GCMs (often 0.5 degrees). Yet, still a certain level of aggregation is required to implement small scale features as for example spatial varying vegetation patterns,

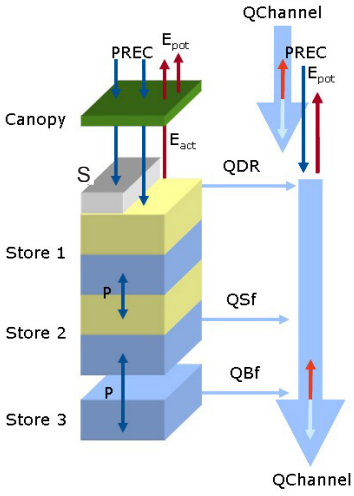


Figure 1.4 Schematic representation of PCR-GLOBWB (Van Beek et al., 2011).

soil characteristics and stream patterns. As a result, at least part of the model parameters does not represent measurable quantities and needs to be estimated (Gosling and Arnell, 2011; Beven and Binley, 1992; Vrugt et al., 2005). Parameter estimation can be accomplished by model calibration, a process where the difference between simulated model output and observations is minimized in order to obtain optimal parameter values (Vrugt et al., 2003). Yet, a full GHM calibration might be unfeasible because of the CPU demand, which will be especially high when a catchment specific calibration is considered. Moreover, estimated parameter values tend to compensate for simplifications within the model structure (Widén-Nilson, 2007; Beven, 1996), errors in discharge observations (McMillan et al., 2010; Vrugt et al., 2005) and, maybe even more important, uncertainty in meteorological model forcing (Fig. 1.1; Biemans et al., 2009; Fekete et al., 2004; Andréassian et al., 2001). As a result of forcing and parameter uncertainty, possible biases and uncertainties are introduced in the modeled discharge as well (Fig. 1.1). If modeled hydrological variability is biased for the current climate, as is often the case when using GCM data, future hydrological conditions are likely to be biased as well. This hampers the correct assessment of future hydrological changes.

1.3 Thesis objectives and research questions

This thesis focuses on the quantification of the global hydrological impacts of climate change and its accompanying uncertainties. The previous paragraphs and Fig. 1.1 provide an outline of the different uncertainties present in the modeling chain of hydrological climate impact studies. To investigate these uncertainties in more detail, two research objectives have been set:

To evaluate GCM datasets on their usability for hydrological impact studies

To assess trends and noise in projected changes in future water availability obtained from an ensemble of GCMs

To fulfill these objectives the following research questions will be answered:

Part 1: Uncertainties in hydrological impact studies

1. Is it feasible to search for optimal parameter estimates for the hydrological model PCR-GLOBWB when the model will be forced with multiple GCM datasets?
2. What is the skill of the hydrological model PCR-GLOBWB, forced with meteorological GCM datasets, in simulating hydrological variability in different regions of the world?
3. How well do runoff fields of GCMs reproduce observed hydrological variability and could these runoff fields make the use of a hydrological model redundant?

Part 2: Climate change impacts

4a. What hydrological impacts are to be expected for different regions of the world?
4b. What is the ratio between trend and noise for the ensemble of projected hydrological changes?
5. How can meaningful hydrological change information be extracted from an ensemble of GCMs in light of the large ensemble uncertainty?

1.4 Scope

The research presented in this thesis focuses on the interpretation and analysis of uncertainties in hydrological climate impact studies. This study does not aim to provide high resolution hydrological change projections for specific catchments. The global scale has been chosen to enable analysis of uncertainties between GCMs in different climate zones and in catchments of different size with different hydrological properties. Furthermore, for the future change assessment, no bias-correction or downscaling is applied to the GCM datasets. First of all, to ensure consistency between the different GCM variables as obtained form physical process modeling within GCMs. And second, to preserve changes in intensities and frequencies, which can particularly be expected for future precipitation (Trenberth et al., 2003).

1.5 Thesis outline

The research questions are answered in subsequent chapters. Each individual chapter has been published, submitted or will be submitted as an article to a peer reviewed journal. As a consequence, some chapters include repetitions of methods described in previous chapters. The different chapters are positioned in the hydrological impact modeling chain in Fig. 1.5.

In chapter 2 the modeled discharge uncertainty originating from parameter and forcing uncertainty is assessed. Based on this assessment we analyze whether it is feasible and realistic to calibrate a global hydrological model when the model is forced with multiple meteorological forcing or GCM datasets.
In chapter 3 meteorological GCM datasets are evaluated on their usability for hydrological impact studies. For this analysis the hydrological model PCR-GLOBWB is run with data from 12 different GCMs for the current climate. The results of these runs are compared with observed discharge data and a PCR-GLOWB model run forced with observed meteorological data.
In chapter 4 the usability of GCM-generated runoff for hydrological studies is investigated by evaluating derived annual average discharge quantities and annual hydrological

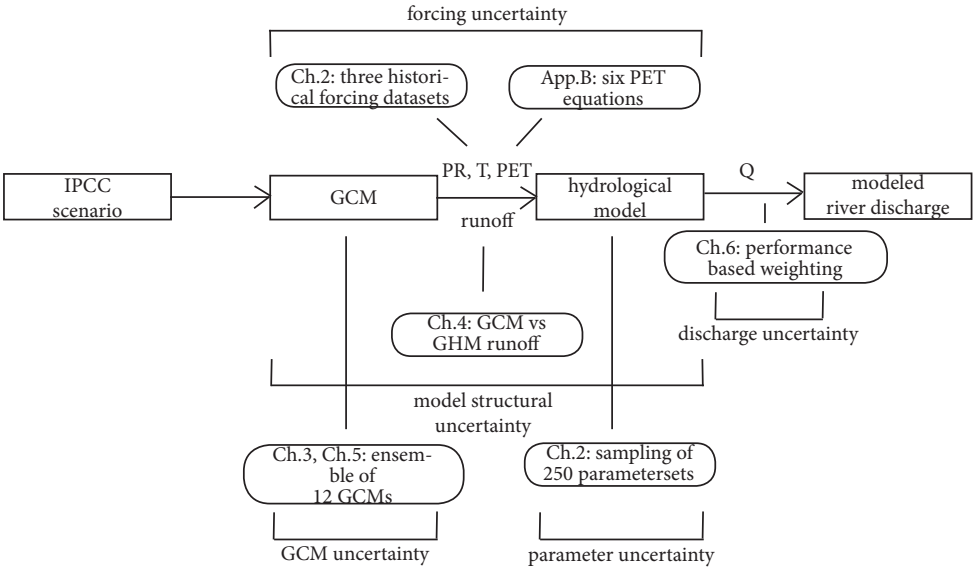


Figure 1.5 Schematization of the hydrological part of the modeling chain of climate impact assessments together with the uncertainties and processes investigated in this thesis.

cycles. This in order to assess whether the GCM runoff flux may even be useful in hydrological impact studies as a replacement of runoff calculated from GCM meteorological data with a hydrological model.

In chapter 5 the future global hydrological consequences of climate change are assessed from data of an ensemble of 12 GCMs. Changes in extreme discharges, inter-annual variability, annual average discharges and timing of the annual discharge cycle are given. In addition, the likelihood of these changes is quantified with the significance of changes projected by the individual GCMs and the consistency between GCMs.

In chapter 6 several performance based weighting and selection methods are compared. These methods can all be used for the derivation of a weighted ensemble mean change and its accompanying uncertainty.

Finally in chapter 7 at the conclusion of this thesis, all research questions are discussed based on the work presented in chapters 2 to 6. The presented results are placed in a broader perspective and related to ongoing research. At the end, several suggestions for future research are made.

2 The importance of parameter and forcing data uncertainty in global scale hydrologic modeling

This chapter is based on:
Sperna Weiland, F. C., J. A. Vrugt, L. P. H. van Beek, A. H. Weerts and M. F. P. Bierkens (2011), The importance of parameter and forcing data uncertainty in global scale hydrologic modeling, in prep.

Abstract

The global hydrological model (GHM) PCR-GLOBWB contains a large number of parameters whose default values were derived from readily available information. Experience with rainfall – runoff models suggests that their performance can be much enhanced if their parameters are calibrated against a historical record of discharge data. Yet, PCR-GLOBWB is computationally demanding, and distributed computing is required to be able to reliably calibrate the model parameter values. Perhaps more importantly, at the large spatial scale of PCR-GLOBWB the role of structural error, albeit significant, is poorly understood, and forcing data is highly uncertain. This study evaluates the influence of PCR-GLOBWB parameter and forcing data uncertainty on modeled discharge estimates. We force the model with three different rainfall data sets involving the CFSR reanalysis, the ERA-interim reanalysis, and a combined ERA-40 reanalysis and CRU data set and investigate the role of parameter uncertainty using Latin Hypercube sampling. Discharge data from five different catchments with contrasting properties are used to benchmark our modeling results.

The predictive ability of PCR-GLOBWB improves considerably by a simple search of the parameter space. Yet, it is typically difficult to find a single parameter combination that shows consistent performance for both the calibration and evaluation period and works well across the five different catchments.

Altogether our results demonstrate that parameter uncertainty constitutes only a relatively minor part of hydrograph prediction uncertainty. This is an important finding

and demonstrates that the apparent dichotomy between model predictions and data can currently not be resolved by increasing model complexity and resolving sub-grid processes. Instead, what is much-needed is an improved characterization of global rainfall amounts at spatial resolution of 0.5 degrees and smaller.

2.1 Introduction

Hydrologic models synthesize our knowledge of the way water flows through catchments, and they are widely used for flood forecasting, and investigation of water resources systems and climate change. These models use relatively simple mathematical equations to conceptualize and aggregate complex, spatially distributed, and highly interacting water, energy, and vegetation processes. As a result, most of the model parameters do not represent direct measurable quantities but can only be derived by calibration against a historical record of input-output data (Beven and Binley, 1992; Vrugt et al., 2005; Gosling and Arnell, 2010). In this process the parameters are adjusted in such a way that the difference between simulated model output and observations is minimized (Gupta et al., 1998; Vrugt et al., 2003).

In the past decade much progress has been made toward the development of efficient calibration strategies of hydrologic models. Despite these efforts, calibration parameters typically tend to compensate for structural inadequacies in the model (Beven, 1996; Vrugt et al., 2005; Widén-Nilson, 2007), errors in discharge observations (McMillan et al., 2010) and, perhaps more importantly, uncertainties in the meteorological model forcing (Andréassian et al., 2001; Fekete et al., 2004; Biemans et al., 2009; Vrugt et al., 2008). There is currently much debate going on in the hydrologic community on what constitutes an appropriate framework for uncertainty assessment. Such approach should appropriately consider model structural, forcing and calibration data error, and communicate probability distributions of parameter and model prediction uncertainty.

Much calibration analysis has focused on relatively simple hydrologic models that represent watersheds with area ranging between 100 – 10,000 km^2. Less attention has been paid to solving the parameter estimation problem for large-scale hydrologic models that attempt to simulate and forecast the largest river systems of the world. Not only do these models pose significant computational challenges, they are also prone to considerable structural error and forcing data error. Within global hydrological models (GHMs) real world processes have to be represented on a model grid with a very coarse resolution. Consequently, a high level of process aggregation is required, which unavoidably introduces structural inadequacies.

The influence of precipitation uncertainty on modeled global runoff fields has been analyzed with the GHM VIC (Wood et al., 1992) using six different precipitation data sets (Fekete et al., 2004). The uncertainty in model predicted runoff was comparable in size to the uncertainty in precipitation and especially large in semi-arid regions. Biemans et al. (2009) estimated the influence of precipitation uncertainty on modeled river discharge. They evaluated seven global gridded precipitation data sets for 294 basins using

the global vegetation and hydrology model LPJmL. The uncertainty in basin average discharge was found to be three times higher than basin average precipitation uncertainty. This illustrates the influence of the uncertainty introduced by a hydrological model and consequently the need for proper calibration.

Several contributions to the hydrologic literature can be found that have attempted to calibrate a global scale hydrologic model, without recourse to quantification of parameter uncertainty. An example includes the work of Fekete et al. (2002) that uses a rather simplistic calibration method to the WBMplus model. In this work, global modeled runoff fields were adjusted with a correction factor to match measured inter-station runoff estimated from discharge measured at neighboring stations. Another study of Alcamo et al. (2003) considered the calibration of the GHM WaterGAP model. They tuned the runoff coefficient of their model to improve the agreement between observed and simulated average discharge. Widén-Nilson et al. (2007) calibrated the global water balance model WASMOD-M using measurements of average discharge, hereby avoiding the disturbing influence of flow regulation. Basin specific parameter values were selected after running WASMOD-M with a sample of 1680 different parameter combinations. The global VIC model has been parameterized using the widely used Shuffled Complex Evolution (SCE) algorithm (Nijssen et al., 2001; Duan et al., 1992). After calibration both the annual average bias and the relative Root Mean Square Error of the long term average monthly discharge values were substantially reduced (respectively 62 to 37% and 29 to 10%). Most of these calibration studies report that parameter estimation is rather difficult and severely hampered by precipitation data errors and the lack of information about river regulation.

Both parameter and forcing data sensitivity have been investigated for the GHM Mac-PDM.09. Parameter sensitivity was analyzed by fourteen perturbed parameter runs in which the field capacity and the degree of variability in soil moisture capacity were varied and potential evaporation was calculated with either the Penman-Montheith or Priestley-Taylor equation (Gosling et al., 2010). In addition a previous version of Mac-PDM has been calibrated manually. Yet, in the first step of the calibration procedure the most adequate meteorological forcing data was selected (best fit to the observed discharge data) hereby excluding the influence of forcing data uncertainty on the final parameter estimates (Arnell, 1999). Recently, an extensive sensitivity analysis has been conducted for the Mac-PDM.09 model (Gosling et al., 2010). However, this analysis focused on forcing data uncertainly only, using an ensemble of 9 different scenarios from 21 different GCMs.

The goal of this study is to analyze the effect of forcing data and parameter uncertainty on global modeled river discharges. Latin Hypercube sampling (LHS) is used to sample the parameter space and provide estimates of hydrograph prediction uncertainty ranges. Streamflow data from five contrasting catchments, involving the Rhine, Amazon, MacKenzie, Mekong and Murray river systems, are used to benchmark and evaluate our results. Much research, focusing on the impact of parameter uncertainty on discharge

modeling, has appeared in the hydrologic literature in the past decades. Yet, most of this work has focused on relatively small river basins using relatively simple lumped watershed models. (Balin et al., 2010; Vrugt et al., 2003; Andréassian et al., 2001; Gupta et al., 1998; Sorooshian and Dracup, 1980; amongst many others). Ideally, we would have used a formal Bayesian paradigm (Schoups and Vrugt, 2010) and Markov Chain Monte Carlo (MCMC) simulation (Vrugt et al., 2003, 2008, 2009) to sample the parameter space. Yet, this methodology is rather computationally demanding, and our results show that LHS is sufficient to illustrate our main conclusions. Note that Laloy and Vrugt (2011) present an efficient distributed computing implementation of DREAM. We consider this formal approach in future work.

This chapter is organized as follows. Section 2 presents an overview of the meteorological data sets and river basins used, herein followed by a short description of PCR-GLOBWB and its most important calibration parameters. In section 3 we report the findings of our study and present hydrograph prediction uncertainty ranges for the different river basins and forcing data sets. Here, we are especially concerned with a comparison against the measured discharge values and evaluate whether model uncertainty decreases if we down sample the original LHS ensemble. Section 4 provides summary and conclusions and discusses future work.

2.2 Data and methods

2.2.1 Meteorological forcing

CRU monthly observations downscaled to daily values with ERA-40 reanalysis
The first forcing data set is made up of monthly data from CRU TS2.1 (Mitchell and Jones 2005, New et al., 2000). This data set has a spatial resolution of 0.5 degrees, and interpolates observed meteorological variables from constituent meteorological stations. Since our hydrologic model, PCR-GLOWWB requires daily values of temperature, rainfall and potential evaporation, temporal downscaling is necessary. To this end we utilize the ERA-40 reanalysis data set (Van Beek, 2008; Sperna Weiland et al., 2010). The ERA-40 data set is a reanalysis product developed at the European Centre for Medium-Range Weather Forecasts (ECMWF; Uppala et al., 2005) with a horizontal resolution of approximately 125 km and a 6-hourly temporal resolution. Reference potential evaporation was created the Penman-Monteith equation (Monteith, 1965). To supplement the CRU TS 2.1 time series, net incoming shortwave radiation was computed according to the FAO guidelines (Allen et al., 1996) using the CRU TS 2.1 cloud cover and a climatology of potential shortwave radiation while wind speed was taken from the long term average monthly climatology, CRU CLM 1.0 (New et al., 1999). The resulting daily forcing dataset follows the daily variability of the ERA-40 reanalysis and

complies with the monthly CRU TS2.1 and CRU CLM 1.0 with daily ERA-40 is a forcing data set with the monthly interpolated observed quantities of the CRU data sets at a 0.5 degrees resolution over the joint period 1958-2001. We opt for this combination as precipitation amounts tend to be biased in reanalysis products (Bosilovich et al., 2007; Troccoli and Kålberg, 2004). The resulting data set has been employed in previous applications of PCR-GLOBWB (Van Beek et al., 2011; Wada et al., 2010; Candogan et al., 2011; Sperna Weiland et al., 2010).

ERA-interim reanalysis

The second forcing data set used in our study is the ERA-interim reanalysis (Dee and Uppala, 2009). This data set supersedes the ERA-40 reanalysis, but includes several improvements to the numerical weather prediction system. Amongst others the horizontal resolution has been increased from T159 to T255, model physics have been improved, additional radiance information is used for bias-correction and improved data sources for wave height, radiance, ozone profiles are included. Yet, a strong relation exists between the ERA-40 and ERA-interim reanalysis data sets. Not only is the ERA-interim system an evolution of the existing ERA-40 system; Up until 2001 boundary forcing of the ERA-interim system have been taken from the ERA-40 system. For this analysis the ERA-interim timeseries extracted from the runs at 12:00 have been re-gridded to a regular 0.5 degrees resolution as required by the hydrological model. We consider the ERA-interim reanalysis product as a continuation of the CRU/ERA-40 data set within future retrospective studies with PCR-GLOBWB.

CFSR reanalysis

The last forcing data set is the CFSR reanalysis product which is developed as part of the Climate Forecast System (Saha et al., 2010) at the National Center for Environmental Predicition (NCEP). The CFSR data set has recently become available in 2010 and supersedes the previous NCEP reanalysis data sets which have been widely used. At this stage the CFSR data set spans the period of 1979 to present. The data has a spatial resolution of approximately 0.25 degrees around the equator to 0.5 degrees beyond the tropics (Higgins et al., 2010). In this study, 6-hourly precipitation, temperature, radiation, air pressure and windspeed were averaged to a daily time-step for the period 1991 to 2000. These daily values were subsequently interpolated to a regular 0.5 degrees grid and used in PCR-GLOBWB. Although previous research showed that CFSR potential evaporation calculated with the Penman-Monteith equation diverged somewhat from CRU estimates (Sperna Weiland et al., 2011a), the Penman-Monteith equation was also employed here for consistency. The CFSR reanalysis product will, in a future application of PCR-GLOBWB, be used for the statistical downscaling of climate model data.

Observed discharge data

This study focuses on five large river basins; the Amazon, Mackenzie, Murray, Mekong

and Rhine watersheds. These basins cover multiple different continents, and a wide range of climatic conditions and differ in total catchment area and degree of regulation. The basins were primarily selected because they have available an extensive record of discharge data at the catchment outlet for the entire period encapsulated by the three forcing data sets. The monthly streamflow observations were derived from the Global Runoff Data Center (GRDC, 2007) and Mekong River Commission. Fig. 2.1 provides a geographical overview of the different river systems along with their most important basin characteristics.

2.2.2 PCR-GLOBWB: The global hydrological model

The PCR-GLOBWB model used herein is a distributed hydrological model that is designed to resolve the global water balance at a spatial resolution of 0.5 degrees (Van Beek and Bierkens, 2009). We here provide a short summary of the most important components of the model. A detailed description of PCR-GLOBWB can be found in Van Beek et al. (2011) and is beyond the scope of this paper.

Each model cell of PCR-GLOBWB consists of two vertical soil layers and one underlying groundwater reservoir. The two soil layers are used to simulate vertical soil water flow and moisture dynamics in the vadose zone. Precipitation (forcing condition) in each grid cell is divided into rain or snow depending on the temperature of the air. Throughfall is computed from subgrid parameterization of the fraction of short and tall vegetation. The remaining water is stored on the vegetation (interception), and eventually returned back to the atmosphere. Rain water that reaches the soil surface either infiltrates or runs-off immediately. Evapotranspiration is calculated from potential evaporation (forcing condition) and the soil moisture status of the top part of the vadose zone. Water that is not taken up by plant roots is either transported to the river system by interflow or percolates to the groundwater store, and eventually ends up as base flow. Upward movement of water between the three different soil stores is possible, but amounts are typically negligibly small. Total runoff is in PCR-GLOWB computed for each individual grid cell as the sum of non-infiltrating melt water, throughfall, saturation excess surface runoff, interflow and base flow. The resulting runoff is accumulated and routed as river discharge along the drainage network using the kinematic wave approximation of the Saint-Venant equations.

The discharge estimates computed with PCR-GLOBWB in this way represent natural flow and do not take into consideration the effect of water use and regulation. We therefore post-process the simulated streamflow values by subtracting monthly water demand calculated using the model of Wada et al. (2010 and 2011). This model calculates global water demand at a spatial resolution of 0.5 degrees using information about industrial, domestic and agricultural water use. The amount of irrigation water subtracted from the surface water depends on PCR-GLOBWB modeled soil moisture conditions which influences the ratio between actual and potential evaporation. Irriga-

Catchment	*Area (km²)*	*Qavg (m³/s)*	*Gauge*
Amazon	6.915.000	190.000	Obidos
MacKenzie	1.805.000	10.700	Norman Wells
Mekong	2.981.076	12.743	Mukdahan
Murray	1.061.469	767	Wakool Junction
Rhine	65.683	2.200	Rees

Figure 2.1 World map (Miller cylindrical projection) with (top) locations of the different catchments used in this study, along with information (bottom) about catchment area, annual average discharge and location of gauge.

tion water is used to reach optimal growing conditions and to minimize the difference between actual and potential evaporation. The final amount of irrigation water subtracted from the surface water is limited by surface water availability.

2.2.3 Parameter uncertainty

PCR-GLOBWB: Calibration parameters

The PCR-GLOBWB model contains a relatively large number of parameters. Although the model was designed to be parameterized from existing global datasets, it remains difficult to assign all parameter values a-priori, and it is to be expected that model performance will increase substantially by calibration against a historical record of discharge data. For the purpose of this study, we selected ten calibration parameters. The ten individual parameters are listed in Table 2.1, including their upper and lower bounds. As we realize that the definition of the parameter sampling ranges highly impacts the modeled uncertainty ranges, the parameter value ranges were carefully derived with an extensive literature review (see table 2.1).

Obviously, PCR-GLOBWB has many more potential parameters that are subject to

considerable uncertainty. Yet, we included only these ten parameters to keep our uncertainty analysis manageable and computationally adequate. The calibration parameters were carefully selected, and evenly distributed among the most relevant hydrologic processes (e.g. saturation excess runoff, flow routing, baseflow release, snow melt / accumulation and evaporatranspiration). Preference was given to those parameters that are most conceptual and thus least measurable. What follows is a description of the ten parameters.

The first two model parameters, e.g. the global distributed values of the groundwater reservoir storage coefficient and soil permeability, directly determine the outflow from the linear groundwater reservoir. This outflow (Q_{bf}) is defined by a linear relation between the actual modeled reservoir storage (S_3) and the reservoir coefficient (J;T^{-1}):

$$Q_{bf} = S_3 J \tag{2.1}$$

The reservoir coefficient is defined by the Kraijenhoff van de Leur equation (Kraijenhof van de Leur, 1958; Sutanudjaja et al., 2011):

$$J = \frac{\pi^2 k_d}{4SL^2} \tag{2.2}$$

Where k_d is the aquifer transmissivity (L^2/T), S (-) is the aquifer porosity or storativity [m•m^{-1}] and L represents the length of the average drainage path [m]. Aquifer transmissivity (or soil permeability) varies spatially and cell specific values are derived from global lithology maps displayed in Fig. 2.2a (Dürr et al., 2005). Storage volume depends on soil porosity and cell specific values have been derived using the Holdridge life zones, see Fig. 2.2b (Gleeson et al., 2011; Leemans et al., 1990). Aquifer porosity largely originates from secondary porosity and is linked to lithology as well. Increased values have been assigned to silicic igneous rocks in tropical regions to account for deep weathering. The average length of the drainage path is defined as the inverse of the drainage density. Drainage density was computed per sub-catchment from the stream length and catchment area, as extracted from the VEMAP and Hydro1k datasets (Verdin and Greenlee, 1996) and corrected for scale effects using the climate-dependent information of Gregory (1976) and the Holdridge life zones of Leemans et al.(1990), before being aggregated to 0.5°. The Holridge life zones are used as an approximation for the quantification of drainage density and consequently storage volume.

In updating the reservoir coefficient, we took the spatial information of lithology and drainage density as constant and varied the values of aquifer transmissivity and porosity within the parameter sampling process by using multiplication factors for each individual lithology class (e.g. 7 lithology classes with one multiplier for the storage coefficient (SC_M) and one for permeability (PB_M)). By using these multiplication factors

Table 2.1 Parameters included in sensitivity analysis

Parameter	*definition*	*multiplier ranges*	*absolute value ranges*	*Default absolute value*	*unit*
	storage coefficient				-
SC_M1	multiplier	0.47 – 1.53	0.11 – 0.36	0.23	
SC_M2	„	0.20 – 2.0	0.01 – 0.1	0.05	-
SC_M3	„	0.33 – 2.0	0.01 – 0.06	0.03	-
SC_M4	„	0.20 – 2.0	0.01 - -0.1	0.05	-
SC_M5	„	0.50 – 3.0	0.02 – 0.12	0.04	-
SC_M6	„	0.20 – 10.0	0.001 – 0.05	0.005	-
SC_M7	„	0.25 – 2.25	0.01 – 0.09	0.04	-
PB_M1	permeability multiplier	0.1 - 10	0.25 – 25	2.50	m/day
PB_M2	„	0.1 – 10	0.1 – 10.00	1.00	m/day
PB_M3	„	0.1 – 10	0.01 – 1	0.10	m/day
PB_M4	„	0.1 – 10	0.1 -10	1.00	m/day
PB_M5	„	0.1 – 10	0.01 – 1	0.10	m/day
PB_M6	„	0.1 – 10	0.001 – 0.1	0.01	m/day
PB_M7	„	0.1 - 10	0.01 – 1	0.10	m/day
FC	matric suction at field capacity		1.0 – 3.0	1.0	m
FC_50	matric suction at which transpiration is halved		3.0 – 15.0	3.33	m
DDFsv	degree-day factor snow melt for areas with short vegetation		0.001 - 0.006	0.00239	m/°C/day
DDFt	degree-day factor snow melt for areas with tall vegetation		0.0005 - 0.004	0.0007	m/°C/day
DDFg	degree-day factor glacier melt		0.005 - 0.008	0.006	m/°C/day
TT	threshold temperature for freezing/thawing		0.0 – 2.0	0.0	°C
MAXFRAC	dimensionless shape factor (-) defining distribution of soil water storage within the cell	0.2 – 5.0		spatial varying	-
MN_M	multiplier of Manning coefficient	0.2 – 5.0		0.04 / 0.1	-

the original spatial distribution is maintained (McMillan et al., 2010) and the different lithology classes are treated independently. It should be noted that multiple lithology classes and Holdridge lifezones occur in a basin, yet in most basins not all lithology classes are present.

We influence the loss of soil moisture though evapotranspiration by calibrating the matric suction at which transpiration is halved (FC_50) and the matric suction at field capacity (FC), the latter also controls drainage.

Snow and glacier melt are influenced by varying the threshold temperature (TT) at

which melting occurs and the value or the degree-day factors (DDF) (Kuchment, 2004; Martinec, 1975). The degree-day factors are individually set for areas with short vegetation (DDFsv), tall vegetation (DDFtv) and glaciers (DDFg).
The parameter MAXFRAC is a measure of the sub-grid distribution of the soil water storage capacity as described by the Improved Arno Scheme (Hagemann and Gates, 2003). The value of MAXFRAC varies spatially as a function of soil properties and vegetation. Varying its value through a multiplier changes the dimensionless shape factor *b* (-), which relates this sub-grid distribution to the average soil water storage capacity which is constant throughout. In turn, adjusting *b* changes the fraction of saturated area given the actual soil water storage and the partitioning of net precipitation in direct runoff and infiltration. The infiltrated water is susceptible to evapotranspiration and infiltration delays runoff. Consequently, adjusting the amount of infiltrated water influences both the magnitude and shape of the discharge hydrograph (Gosling and Arnell, 2011; Sperna Weiland et al., 2011b). Cell specific values of b are calculated based on the distribution of maximum rooting depths, which have been derived from the 1×1 km distribution of vegetation types from GLCC (Hagemann, 2002).
River discharge is calculated with the kinematic wave approximation of the Saint-Venant Equations. In these equations river discharge depends in addition to the cross-section, slope and hydraulic radius of the river bed on the bed roughness defined by the Manning's roughness coefficient. Within the original PCR-GLOBWB model the floodplain manning coefficient is set to 0.1 and the channel coefficient to 0.04 (Chow et al., 1988), resulting in a map with spatial varying coefficient values. Within this analysis the global map is uniformly multiplied by sampled values of MN_M.

Latin hypercube sampling
In the absence of detailed prior information, the individual parameters were assumed to be uniformly distributed within the ranges listed in Table 2.1. A uniform distribution has shown to work well for exploratory sensitivity and uncertainty analysis (Haan et al., 1998). In a first step, 250 different parameter combinations are sampled by Latin Hypercube Sampling (LHS) using the ranges previously listed in Table 2.1 (Muleta and Nicklow, 2005). A total of 250 samples was deemed adequate for an initial assessment of parameter uncertainty. A larger sample likely will provide more robust results, yet further increases computational demands. After sampling the parameter space, PCR-GLOBWB is executed for each different parameter combination, resulting in 250 different time series of simulated discharge for each of the five different basins and three different forcing data sets. Steady state conditions at the start of each simulation were enforced by running PCR-GLOBWB multiple times for each different parameter combination using the climatology of each specific forcing data set until differences in basin average discharge for two consecutive runs was less than one percent.

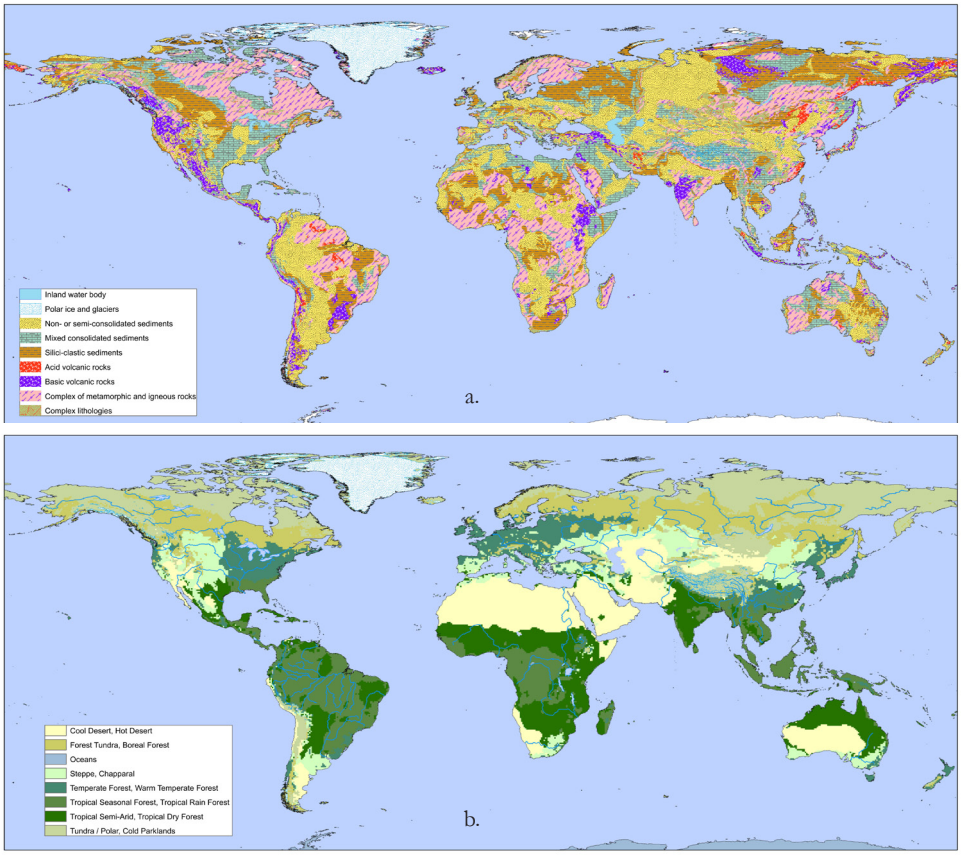


Figure 2.2 World maps (Miller cylindrical projection) that depict spatial variations in (A) lithology (Dürr et al., 2005), and (B) Holridge life zones (Gleeson et al., 2011).

2.2.4 Experiment design

Uncertainty analysis for calibration period (1991-1995)

The first analysis will be conducted for the period 1991 to 1995, which hereafter will be referred to as the calibration period. To benchmark the results, and assess the influence of meteorological forcing uncertainty on modeled discharge, PCR-GLOBWB is also individually run with default parameter values for all three different forcing data sets.

The results of the model simulations are evaluated by comparison against observed monthly discharge data. Three different diagnostics measures are used to quantify the correspondence between modeled, Q_{mod} and measured, Q_{obs} discharge values. This includes the Root Mean Square Error (RMSE; eq. 2.3a), the normalized root mean

squared error (eq. 2.3b), which has been obtained by dividing the RMSE through the mean observed monthly discharge, and the Nash-Sutcliffe (NS; eq. 2.4) coefficient.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Q_{mod,i} - Q_{obs,i})^2}{n}} \tag{2.3a}$$

$$RMSE_{normalized} = \frac{\sqrt{\frac{\sum_{i=1}^{n}(Q_{mod,i} - Q_{obs,i})^2}{n}}}{\overline{Q}_{obs}} \tag{2.3b}$$

$$NS = 1 - \frac{\sum_{i=1}^{n}(Q_{obs,i} - Q_{mod,i})^2}{\sum_{i=1}^{n}(Q_{obs,i} - \overline{Q}_{obs})^2} \tag{2.4}$$

where n denotes the total number of months (n = 60), and $\overline{Q}_{obs}$ is the average monthly-discharge.

The LHS approach utilized herein is anticipated to result in a large spread of the PCR-GLOBWB model predictions. However, our experience suggests that many of the ensemble members created this way will exhibit a rather poor predictive performance. Indeed, many of the randomly drawn parameter combinations will have difficulty to accurately represent the observed hydrologic signatures of the basins. Drawing inspiration from the GLUE method of Beven and coworkers (Beven and Binley, 1992; Beven, 2006), we adopt a subjective threshold to separate the ensemble in behavioral and non-behavioral solutions. We eliminate the worst 95% of the solutions of the LHS ensemble and focus our attention on the best 12 parameter combinations. We evaluate the performance of this 12-member ensemble using a monthly average uncertainty range, hereafter referred to as UR:

$$UR = \frac{1}{n}\sum_{i=1}^{N}(Q_{\min,i} - Q_{\max,i}) \tag{2.5}$$

where $Q_{\min,i}$ ($Q_{\max,i}$) represents the minimum (maximum) discharge of the i-th month. To demonstrate the advantages of this down sampling approach, we compare the value of UR obtained from the behavioral model realizations with its value separately derived

for the full ensemble of 250 streamflow predictions. This approach is repeated for each different forcing data set and catchment, and results are compared.

Uncertainty analysis for evaluation period (1996-2000)
To evaluate the consistency of the PCR-GLOBWB ensemble performance of the model during an independent evaluation period, please consider Table 2.2 to 2.4 and Figure 2.3 to 2.7 that summarize the performance of the behavioral parameter combinations for the period 1996 to 2000. For the evaluation period the monthly average discharge timeseries, obtained from the behavioral and best performing model realizations, are also evaluated with normalized RMSE and NS values, uncertainty ranges and hydrographs.

Uncertainty analysis for other hydrological model outputs
The discharge data used herein provides an integrated signature of catchment response. This data is of imminent importance to test and evaluate hydrologic models. Yet, even if PCR-GLOBWB is able to accurately describe the monthly discharge emanating from the basin, it remains to be seen whether the model can also accurately describe temporal dynamics of other hydrological processes in the watershed such as soil moisture conditions, snow accumulation and melt, local runoff generation and actual evapotranspiration. Ideally, the behavioral solutions identified so far, capture all relevant hydrologic processes, and parameter combinations of PCR-GLOBWB that fit the discharge data well and simultaneously also predict the other relevant fluxes and variables. Unfortunately, very limited measurement data exists of soil moisture dynamics and other hydrologic variables at large spatial scales. This makes it rather difficult to evaluate the suitability of PCR-GLOBWB for predicting the hydrologic cycle. Many advances have been made in recent years to improve characterization of hydrologic fluxes and variables at large spatial scales, but such data is not yet readily available. In the absence of such measurements, instead we evaluate the performance of PCR-GLOBWB by studying the reduction in simulation (prediction) uncertainty of these various variables from the behavioral solutions. This reduction is evaluated using the following simple measure:

$$Reduction = 100 \frac{UR_{behavioral}}{UR_{fullLHS}} \tag{2.6}$$

where $UR_{behavioral}$ denotes the prediction (simulation) spread pertaining to the down-sampled (behavioral) solutions and $UR_{fullLHS}$ signifies the original simulation uncertainty corresponding to the initial LHS sample of 250 parameter combinations.

Exploring possibilities for regionalization
For all rivers and all forcing data sets, PCR-GLOBWB has been run with the same

250 different parameter combinations. By comparing their individual RMSE values (Eq. 2.3a), for the five rivers and three different forcing data sets, we explore whether performance improvements are independent of catchment and climate characteristics. Such an independency would enable the estimation of one global parameter estimate applicable under multiple forcing data sets (Widén-Nilson et al., 2007). To assess this independency, we calculate average (linear) correlation coefficients, $r_{X,Y}$ between the 250 different RMSE values for different forcing data sets and catchments:

$$r_{X,Y} = \frac{\sum_{j=1}^{M} (RMSE_{X,j} - \overline{RMSE}_X)(RMSE_{Y,j} - \overline{RMSE}_Y)}{\sqrt{\sum_{j=1}^{M} (RMSE_{X,j} - \overline{RMSE}_X)^2 \sum_{j=1}^{m} (RMSE_{Y,j} - \overline{RMSE}_Y)^2}} \quad (2.7)$$

where the subscript X and Y denote two different basin-forcing combinations, j is the actual parameter combination, and M represents the total number of parameter estimates. Correlation coefficients close to 1 are desired and imply that performance is independent of catchment and climate characteristics. Values of $r_{X,Y}$ that deviate substantially from 1 indicate that the performance of the parameter combinations differs considerably between catchments and forcing data sets.

2.3. Results and discussion

2.3.1 Calibration period - full parameter and forcing uncertainty

Figures 2.3 to 2.7 present hydrograph prediction uncertainty ranges for the five different catchments considered in this study. Five different panels are used to summarize our findings. The light-blue region depicts the streamflow uncertainty ranges of the full LHS ensemble, whereas the dark-blue region denotes the results for the behavioral solutions obtained after selecting the behavioral parameter combinations of the original 250 samples. Each different forcing data set is color coded. To benchmark the results of the LHS sample, the top panel displays the results of PCR-GLOBWB using default values for the parameters and the three forcing datasets. The middle three panels illustrate the results for the three individual forcing data sets, whereas the bottom panel shows the results for a superimposed ensemble summarizing the results for all forcing data sets simultaneously. We now jointly discuss the results of Figs. 2.3 to 2.7.

For most rivers, the CFSR reanalysis forcing data set overestimates the observed monthly discharge data, irrespective of whether the default parameterization of PCR-GLOBWB is used or global sampling of the parameter space with LHS is employed. The hydrograph prediction uncertainty ranges of CFSR are also considerably wider than those computed with the other two forcing data sets. This is a result of the relatively high precipitation amounts in the CFSR datasets for all basins except the Amazon (Fig. 2.8). The overestimation is particularly large for the MacKenzie and Mekong.

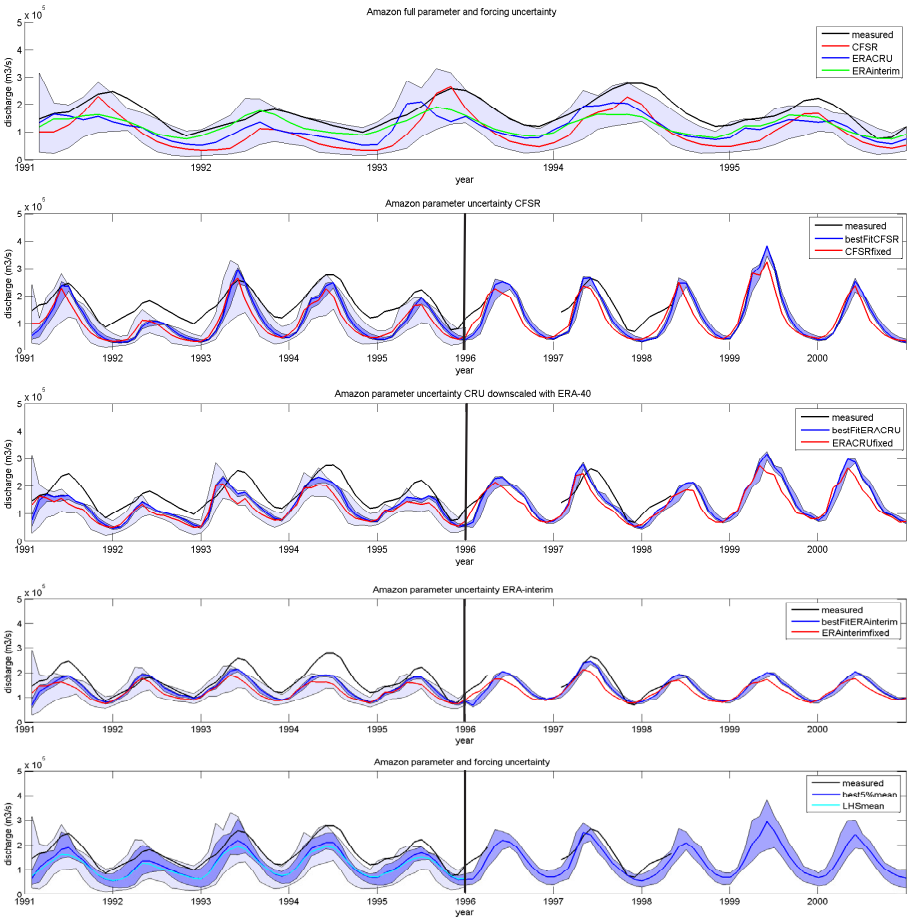


Figure 2.3 Hydrographs with monthly average discharge for the Amazon modeled with PCR-GLOBWB. In each individual panel, measured discharge data are shown in black, the light-blue area denotes the prediction uncertainty (difference between minimum and maximum discharge) obtained from the full LHS ensemble, and the dark-blue area represents the prediction uncertainty ranges derived from the behavioral parameter combinations. The top panel displays the results of the default PCR-GLOBWB parameterization forced with CFSR reanalysis data (red), the ERACRU data set (dark blue) and ERA-interim reanalysis data (green) and plots the LHS ensemble spread for all these different forcing data sets combined. The next three panels depict model realizations for the CFSR, ERACRU and ERA-interim data sets. The bottom panel displays the results for the full LHS ensemble and behavioral parameter combinations combined for all three forcing data sets. This signifies the combined effect of parameter and forcing uncertainty.

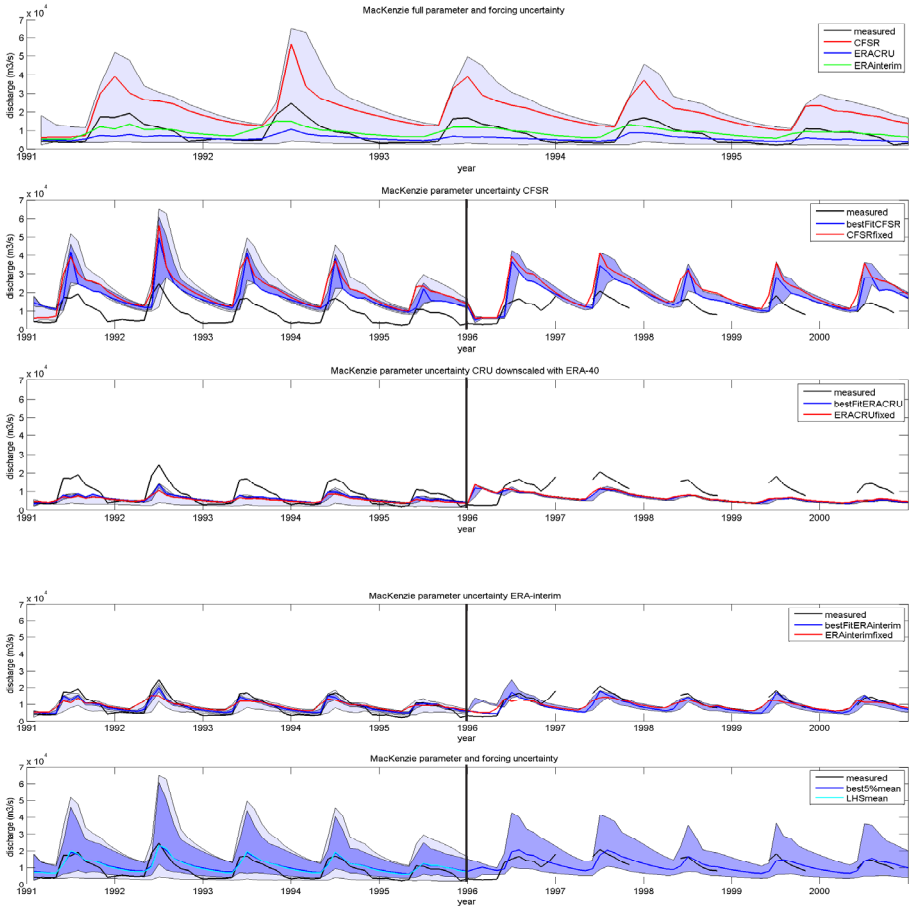


Figure 2.4 Measured (black line) and PCR-GLOBWB model predicted streamflow hydrographs for the MacKenzie basin. Please refer to the caption of figure 2.3 for further explanation.

For all rivers except the Amazon (where all forcing data sets underestimate the observed flow; consistent with the analysis of Widén-Nilson et al. (2007) and Döll et al. (2003)) the discharge prediction uncertainty ranges of the LHS ensemble for all three forcing data sets envelop the measured discharge data (see Fig. 2.3 to 2.7 a.). This is a rather encouraging result, and illustrates that PCR-GLOBWB is at least a reasonable description of the discharge dynamics of these river basins. The model is able to fit the monthly streamflow data reasonably well, irrespective of the forcing data used.

For the MacKenzie, Murray and Mekong (Figs. 2.4, 2.5 and 2.6) river systems there is a large variation in model predicted streamflow values among the different forcing data sets. A multitude of reasons can explain this apparent dichotomy, among which problems with precipitation undercatch due to snowfall (Fiedler and Döll, 2007) are perhaps most important and known to affect the ERA-CRU data set. These errors become most

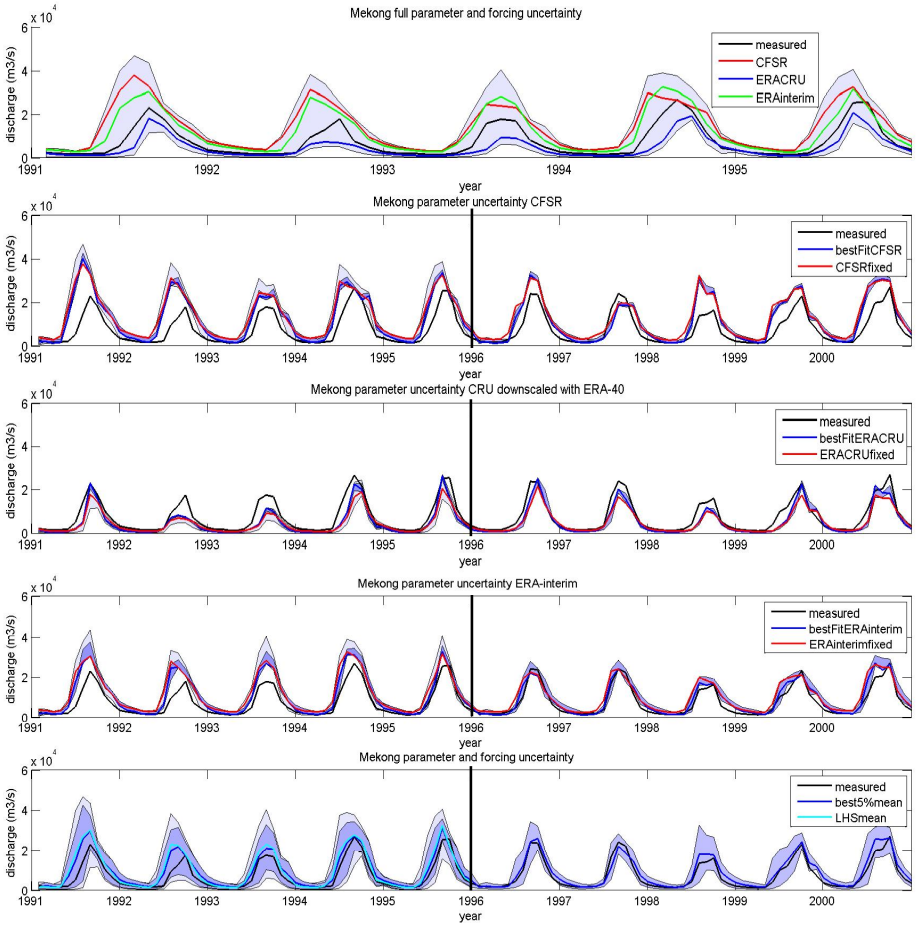


Figure 2.5 Measured (black line) and PCR-GLOBWB model predicted streamflow hydrographs for the Mekong river basin. Please refer to the caption of figure 2.3 for further explanation.

pronounced for high-latitude river systems such as the Mackenzie in Alaska that hence experiences the largest amounts of snowfall from all the selected basins.

For the Murray catchment, PCR-GLOBWB severely overestimates the actual monthly discharge when driven with the CFSR and ERACRU forcing data sets. This overestimation is not surprising, and was also reported earlier by other GHM studies for very dry basins (Gosling and Arnell, 2011 and references therein). Due to the large variations in rainfall between the three different forcing data sets, forcing data error constitutes the main source of uncertainty in monthly discharge modeling with PCR-GLOBWB. Apparently, parameter uncertainty represents only a relatively minor part of hydrograph prediction uncertainty. This is an important finding and demonstrates that the apparent mismatch between model predictions and data cannot be resolved by increasing model complexity and resolving sub-grid processes alone. Instead, what is much-needed is an

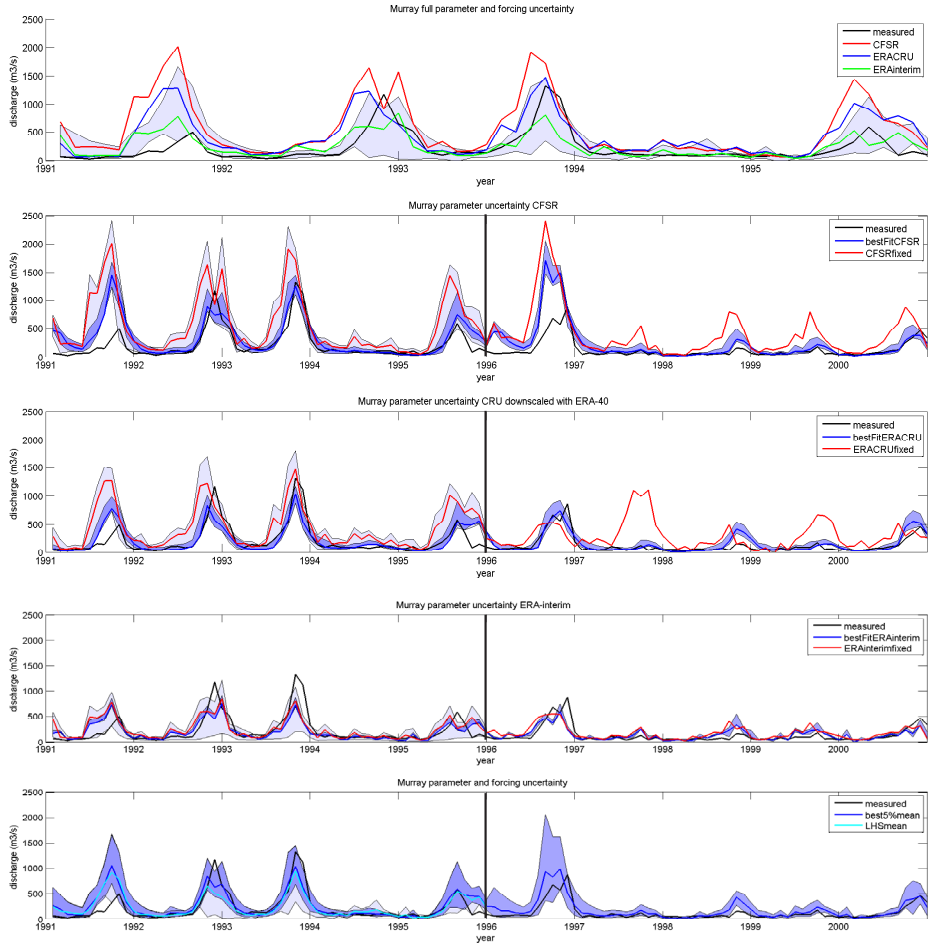


Figure 2.6 Measured (black line) and PCR-GLOBWB model predicted streamflow hydrographs for the Murray river basin. Please refer to the caption of figure 2.3 for further explanation.

improved characterization of global rainfall amounts at spatial resolution of 0.5 degrees and smaller.

For the Rhine (Fig. 2.7) river system an excellent resemblance is found for the ERA-interim and ERA-40/CRU data sets with the observed discharge data. Consequently, the modeled discharges are in high agreement with corresponding uncertainty ranges that appear rather narrow (Fig. 2.7 c and d), and generally envelop the observed stream-flow data.

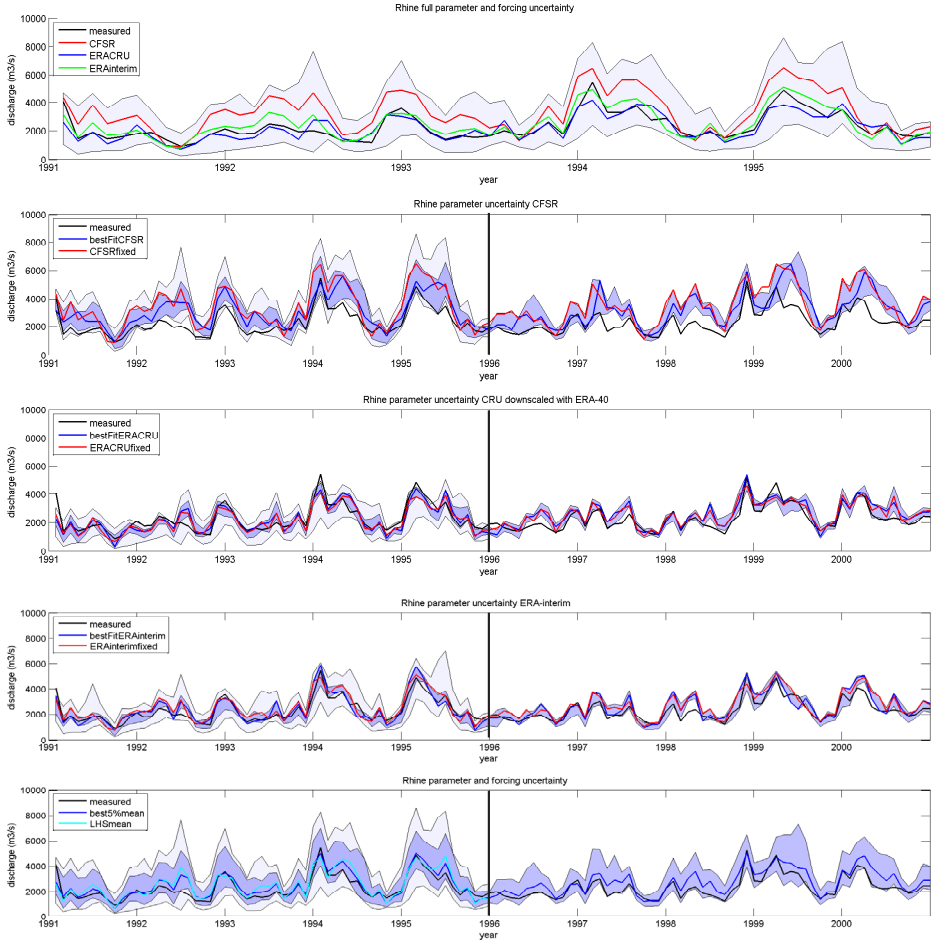


Figure 2.7 Measured (black line) and PCR-GLOBWB model predicted hydrographs for the Rhine river system. Please refer to the caption of figure 2.3 for further explanation.

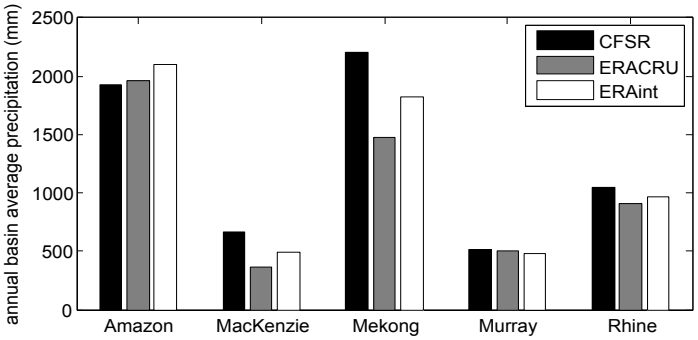


Figure 2.8 Annual basin average precipitation amounts (mm) for the CFSR reanalysis product (black), the combined ERA-40 / CRU dataset (grey) and the ERA-interim reanalysis product (white).

Table 2.2 Normalized root mean square errors (RMSE) of monthly modeled discharge time series for all forcing data sets and river systems considered herein. The various headings summarize the performance of the default model parameterization (*default*), best parameter combination (*best par*), mean prediction of the behavioral solutions (*behavioral*), and full LHS ensemble (*mean LHS*) respectively, for the calibration (*Calibration*: 1991 - 1995) and evaluation period (*Evaluation*: 1996 - 2000) respectively. The statistic "*Overall*" lists the average performance for all three different forcing data sets combined.

RMSE	***Calibration***				***Evaluation***		
Rhine	*default*	*best par*	*behavioral*	*mean LHS*	*default*	*best par*	*behavioral*
CFSR	0.56	0.43	0.46	0.53	0.60	0.50	0.47
ERACRU	0.21	0.21	0.24	0.29	0.23	0.23	0.22
ERAint	0.21	0.19	0.22	0.27	0.26	0.25	0.20
Overall				0.27			
MacKenzie							
CFSR	1.54	1.31	1.38	1.52	1.10	0.83	0.87
ERACRU	0.59	0.52	0.55	0.67	0.52	0.53	0.54
ERAint	0.43	0.28	0.29	0.35	0.27	0.24	0.33
Overall				0.43			
Murray							
CFSR	2.32	1.06	1.17	1.94	3.29	1.79	1.93
ERACRU	1.53	0.77	0.84	1.40	2.36	0.71	0.75
ERAint	0.93	0.87	0.87	0.90	0.99	0.88	0.89
Overall				1.25			
Amazon							
CFSR	0.42	0.39	0.40	0.41	0.34	0.36	0.37
ERACRU	0.37	0.29	0.31	0.34	0.29	0.26	0.27
ERAint	0.33	0.24	0.26	0.30	0.24	0.20	0.22
Overall				0.33			
Mekong							
CFSR	1.16	1.05	1.11	1.30	0.81	0.73	0.76
ERACRU	0.58	0.51	0.52	0.58	0.47	0.39	0.42
ERAint	0.83	0.67	0.76	0.93	0.46	0.32	0.36
Overall				1.16			

2.3.2 Reduced parameter uncertainty in calibration and evaluation period

Calibration period

Table 2.2 and 2.3 provide normalized RMSE and NS values of the monthly discharges obtained with the default model parameterization of PCR-GLOBWB and those values obtained from the LHS ensemble. Hydrographs of the best performing parameter combination for each individual forcing data set are depicted in dark blue in panels 2 to 4 of figures 2.3 to 2.7, respectively, whereas the mean discharge prediction of the full LHS ensemble is coded with the light blue line in panel 5. The most important findings are as follows.

Table 2.3 Nash-Sutcliffe (NS) coefficients of PCR-GLOBWB monthly discharge time series for all different forcing data sets and river catchments. The various headings summarize the performance of the default model parameterization (*default*), best parameter combination (*best par*), mean prediction of the behavioral solutions (*behavioral*), and full LHS ensemble (*mean LHS*) for the calibration (*Calibration*: 1991 - 1995) and evaluation period (*Evaluation*: 1996 - 2000) respectively. The statistic "Overall" lists the average performance for all three different forcing data sets combined.

NS	***Calibration***			***Evaluation***			
Rhine	*Default*	*best par*	*behavioral*	*meanLHS*	*default*	*best par*	*behavioral*
CFSR	-0.80	0.03	-0.02	-0.58	-1.59	-0.79	-0.56
ERACRU	0.76	0.74	0.72	0.53	0.62	0.62	0.65
ERAint	0.75	0.79	0.79	0.60	0.51	0.55	0.70
Overall				0.59			
MacKenzie							
CFSR	-4.61	-3.11	-3.28	-4.52	-7.88	-4.10	-4.61
ERACRU	0.17	0.35	0.29	-0.05	-0.99	-1.09	-1.18
ERAint	0.56	0.82	0.81	0.71	0.46	0.59	0.18
Overall				0.55			
Murray							
CFSR	-2.80	0.21	0.08	-1.66	-6.60	-1.28	-1.62
ERACRU	-0.66	0.58	0.54	-0.39	-2.92	0.65	0.60
ERAint	0.39	0.47	0.47	0.43	0.31	0.46	0.44
Overall				0.11			
Amazon							
CFSR	-0.98	-0.72	-0.74	-0.89	-0.06	-0.23	-0.27
ERACRU	-0.48	0.07	-0.02	-0.29	0.20	0.35	0.30
ERAint	-0.18	0.34	0.25	-0.01	0.48	0.61	0.55
Overall				-0.22			
Mekong							
CFSR	-0.41	-0.17	-0.29	-0.78	0.28	0.42	0.35
ERACRU	0.64	0.73	0.72	0.64	0.75	0.83	0.81
ERAint	0.28	0.41	0.52	0.08	0.76	0.88	0.86
Overall				0.43			

During the calibration period, the mean discharge prediction of the LHS ensemble generally outperforms the default parameterization of PCR-GLOBWB, particularly when the default parameterization shows a large bias from the observed discharge data. Yet, for several rivers for some of the forcing data sets, the default parameterization of PCR-GLOBWB results in lower RMSE values than the LHS mean. See for example the Rhine forced with ERA-CRU and ERA-interim and the Mackenzie forced with ERA-CRU (Table 2.2 and 2.3). For the Rhine the default parameterization of PCR-GLOBWB forced with the ERA-CRU data set performs even better than the best performing parameter combination from the LHS sample.

Although the set of behavioral parameter estimates was formed based on the RMSE

Table 2.4 Average spread of the PCR-GLOBWB simulated hydrograph prediction uncertainty ranges for the LHS ensemble ("*LHS ensemble*") and down-sampled behavioral solutions ("*Behavioral*") for all three forcing data sets individually, combined using the default parameterization ("*Forcing*") and joint LHS parameter and forcing data uncertainty ("*Pars + Forcing*").

range	***cal***		***eval***		***cal***		***eval***
Rhine	*LHS ensemble*	*Behavioral*	*Behavioral*	**Amazon**	*LHS ensemble*	*Behavioral*	*Behavioral*
CFSR	2665	1464	1434	*CFSR*	94102	23184	28292
ERACRU	1849	725	749	*ERACRU*	71001	14811	26252
ERAint	2195	739	812	*ERAint*	72315	14791	17761
pars + forcing	3560			*pars + forcing*	126940		
forcing	1270			*forcing*	44358		
MacKenzie	*LHS ensemble*	*Behavioral*	*Behavioral*	**Mekong**	*LHS ensemble*	*Behavioral*	*Behavioral*
CFSR	12750	6225	6130	*CFSR*	5337	2294	2148
ERACRU	3876	1080	1121	*ERACRU*	2982	1124	1702
ERAint	5387	1870	3705	*ERAint*	5766	3447	3794
pars + forcing	22806			*pars + forcing*	14346		
forcing	14418			*forcing*	9530		
Murray	*LHS ensemble*	*Behavioral*	*Behavioral*				
CFSR	490	190	138				
ERACRU	399	144	109				
ERAint	290	57	55				
pars + forcing	671						
forcing	332						

values of monthly discharge timeseries, for most forcing data sets for most rivers the mean monthly discharge timeseries calculated from the behavioral parameter combinations outperform the default parameter values according to the NS values as well. Overall the performance improvement is largest for the forcing data with the largest bias. Yet, this may be an effect of compensatory parameter estimates, where less realistic parameter values reduce the effect of the bias in the forcing data set (Beven, 1996; Widén-Nilson, 2007). Moreover, even after the relatively large performance improvement, the bias in discharge for the least performing data set remains highest. These results again suggest that forcing bias dominates model performance.

Evaluation period

To evaluate the consistency of our modeling results, we use a 5-year evaluation period from 1996 to 2000 to investigate the predictive ability of PCR-GLOBWB outside the calibration period. We focus on the behavioral solutions derived from the calibration period only. For most rivers, the mean ensemble discharge prediction of the behavioral

solutions and corresponding streamflow predictions of the best parameter combination outperform the results of the default parameterization. Exceptions are 1) the Rhine, where the default parameterization performs quite well (Van Beek et al., 2011) but note that the best parameter combination of the original LHS ensemble (and thus behavioral solutions) still obtains higher NS values when using the ERA-interim data set for this river system, 2) the MacKenzie forced with the ERA-CRU data set and 3) the Amazon catchment using the CFSR forcing data set. For both the Mackenzie and Amazon river systems the PCR-GLOBWB predicted monthly discharge simulations of the default model parameterization and best parameter combination show large deviations from the observed data. For the Amazon there is a consistent offset in the timing of the annual discharge cycle, which we posit is caused by an overestimation of the actual flow velocities in the floodplains. The floodplain areas are fixed in the version of PCR-GLOBWB used and hence it remains difficult to accurately represent the dynamics of storage and outflow and it is unlikely that this can be compensated for by increasing the prior sampling ranges of the Manning coefficient.

A rather discouraging result is that the best parameter combination found for each individual river system and forcing data set during the calibration period does not exhibit the best predictive performance during the evaluation period. Other solutions of the behavioral set demonstrate better performance, illustrating the difficulty of finding a single default parameterization that works well for each individual forcing data set and watershed. In the evaluation period, the RMSE (NS) value of the mean discharge prediction of the set of behavioral model realizations is often substantially lower (higher) than the respective statistics for the best parameter combination of the calibration period. Note as well that the PCR-GLOBWB prediction uncertainty ranges for the behavioral model realizations are comparable in size or slightly larger than those previously estimated for the calibration period (Table 2.4). These results are in line with previous studies of Krishnamurti et al. (1999), Ajami et al. (2006) and Vrugt and Robinson (2007) which led to the conclusion that it might be more productive to work with an ensemble of (behavioral) solutions and model predictions rather than a single best realization.

Table 2.4 also displays uncertainty ranges obtained by combining the behavioral model realizations for all three forcing data sets ("Behavioral pars"; Pars + Forcing). This uncertainty range is for almost all river systems at least twice as large as the hydrograph prediction spread corresponding to the behavioral model realizations for the individual forcing data sets. This indicates that bias and deviations in the forcing data sets highly impact the overall uncertainty and that differences between discharge predictions obtained from the different forcing data sets remain large, even when focus is only on the behavioral parameter estimates.

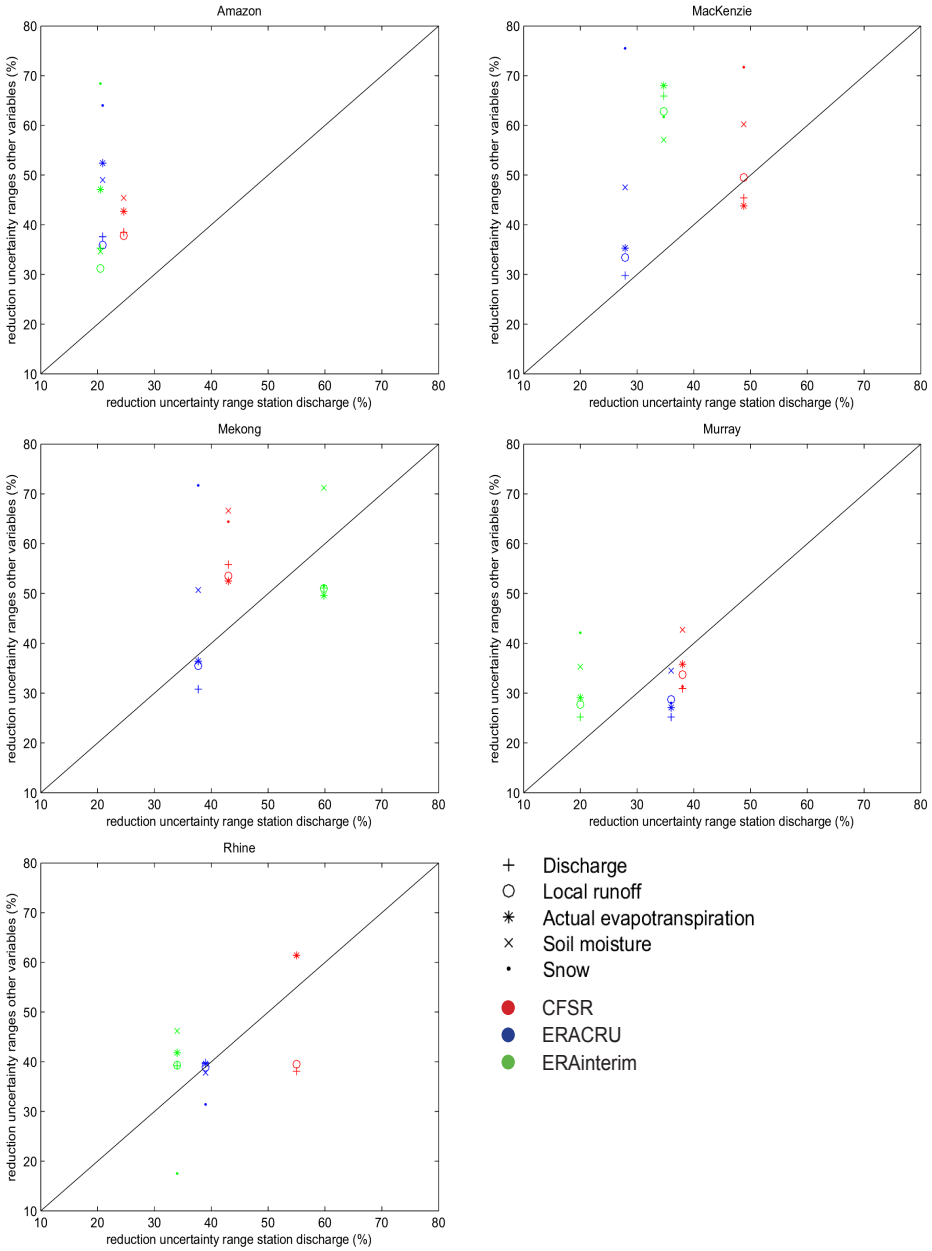


Figure 2.9 Reduction in uncertainty for the (A) Amazon, (B) MacKenzie, (C) Mekong, (D) Murray, and (E) Rhine river basin when down sampling the original full LHS ensemble to the behavioral parameter combinations. The "+", "o", "*", "x" and "·" symbols are used to denote reduction in uncertainty for discharge, local runoff, actual evapotranspiration, soil moisture and snow, respectively. Color coding is used to indicate different forcing data sets.

2.3.3 Reduction of uncertainty in other hydrological model outputs

The performance of the individual parameter combinations sampled with LHS has been quantified using the RMSE summary statistic of modeled and observed monthly discharge values. Ideally, the behavioral solutions should simultaneously also exhibit better predictive performance for other hydrological variables (e.g. soil moisture, local runoff, actual evapotranspiration and snow) not explicitly included in the calibration period. This would indicate that the performance of the hydrological model is overall improved when using only discharge data.

Figure 2.9 depicts the reduction in simulation uncertainty for the (A) Amazon, (B) MacKenzie, (C) Mekong, (D) Murray, and (E) Rhine system of each of the individual catchments when down-sampling the original LHS ensemble into the behavioral solutions considered herein. In general, there is an apparent correlation between the reduction in simulation uncertainty of the discharge of individual basins and the associated reduction in prediction spread of the other important hydrologic variables. The larger the average reduction in spread of the discharge simulations the more pronounced the reduction in uncertainty for snow (·), local runoff (o), actual evapotranspiration (*) and soil moisture (x). This reduction is especially large for the PCR-GLOBWB modeled state variable snow in the Mackenzie, Mekong and Amazon catchments, where snowfall determines the timing of the annual discharge cycle. A similar reduction of the uncertainty range was expected for the Rhine, but not visible in the plot. The runoff regime of the Rhine is influenced by both precipitation and snow. Consequently, multiple processes (e.g. snow melt, soil moisture storage and groundwater storage) influence the shape of the hydrograph. In the Rhine and Mekong catchments the reduction in uncertainty for modeled soil moisture appears rather large, most likely due to the temporal storage of water as soil moisture in the river deltas which directly influences the timing of the annual discharge cycle as well. Different combinations of parameter values may perform equally well due to compensating effects. As a result, the reduction in simulation uncertainty is smaller for most of the variables other than monthly discharge. And, although NS values are relatively high for the Rhine river system, the offset in the timing of the spring discharge indicates that some improvements to the model structure of PCR-GLOBWB are warranted. The behavioral solutions found herein, are likely compensating for these structural inadequacies, a subject of much interest in the hydrologic literature.

For the Murray system the reduction in uncertainty is smaller for the other hydrological variables than for monthly average discharge for both the ERA-CRU and CFSR forcing data sets. Performance of PCR-GLOBWB for the Murray, particularly in combination with the CFSR and ERA-CRU forcing data set, is poor as has also been observed for other GHM studies (Gosling and Arnell, 2011). As a consequence the better performing parameter combinations might be selected for the wrong reasons in this basin. Moreover, river discharge of the Murray is highly impacted by temporal changes in the storage in the reservoir due to management and river regulation. These human factors

Table 2.5 Lower diagonal values of average correlation coefficients between the 250 different RMSE values of PCR-GLOBWB simulated monthly discharge values of the five different basins and three forcing data sets. Bold values are used to indicate correlation coefficients larger than 0.7. High values are indicative for consistent performance across different catchments and forcing data sets; and favor the selection of a single parameter combination that exhibits good predictive performance for the different catchments, and hence can replace the current default PCR-GLOBWB parameterization.

		Amazon			MacKenzie			Mekong			Murray			Rhine		
		CFSR	ERACRU	ERAint	CFSR	ERACRU	ERAint	CFSR	ERACRU	ERAint	CFSR	ERACRU	ERAint	CFSR	ERACRU	ERAint
Amazon	CFSR															
	ERACRU	**0.91**														
	ERAint	**0.96**	**0.93**													
MacKenzie	CFSR	-0.39	-0.35	-0.30												
	ERACRU	0.44	0.41	0.39	-0.81											
	ERAint	0.23	0.20	0.23	-0.26	0.35										
Mekong	CFSR	-0.48	-0.45	-0.53	0.53	-0.70	-0.32									
	ERACRU	0.53	0.47	0.48	-0.79	**0.92**	0.37	-0.80								
	ERAint	-0.46	-0.44	-0.49	0.64	-0.72	-0.31	**0.95**	-0.84							
Murray	CFSR	-0.42	-0.37	-0.35	**0.87**	-0.94	-0.31	0.65	-0.91	0.69						
	ERACRU	-0.41	-0.37	-0.35	**0.82**	-0.94	-0.31	**0.70**	-0.91	**0.72**	**0.99**					
	ERAint	-0.06	0.13	-0.04	-0.03	0.18	0.16	-0.10	0.24	-0.14	-0.11	-0.14				
Rhine	CFSR	-0.33	-0.29	-0.30	0.67	-0.76	-0.27	0.53	-0.60	0.57	0.68	0.67	0.06			
	ERACRU	-0.02	0.00	-0.07	0.01	0.17	0.03	0.03	0.17	0.07	-0.17	-0.16	**0.37**	**0.13**		
	ERAint	-0.17	-0.15	-0.20	0.37	-0.28	-0.12	0.29	-0.20	0.35	0.22	0.21	0.23	0.64	**0.80**	

are not included in this version of PCR-GLOBWB (Van Beek et al., 2011).
The study of Anderton et al. (2002) illustrated that parameter optima can be positioned in different parts of the parameter space, depending on the selected response variable (e.g. soil moisture, phreatic surface level or discharge). And consequently, including multiple hydrological variables within a parameter estimation process, can allow for a rejection of those parameterizations for which the correct outlet discharge is obtained as a result of compensating errors in the internal model structure (Winsemius et al., 2006). Yet, for a global scale hydrologic model, data availability severely limits the possibilities of using such a multi-response calibration.

2.3.4 Possibilities for global parameter estimates or regionalization

In this section the dependency of the optimal parameter estimates on forcing data and catchment characteristics is investigated. In other words, we test whether it is feasible to find one parameter combination that works well for multiple different forcing data sets (Widén-Nilson et al., 2007). Of course, this would be highly desirable as it enables parameter transfer to ungauged basins and may reduce calibration efforts. We calculate the correlation of the RMSE values, obtained for the same 250 parameter combinations in the different basins for the different forcing data sets. The results of this analysis are presented in Table 2.5. Unfortunately the average correlation coefficient was rather low 0.09. This essentially indicates that it is rather difficult, if not impossible, to find a single realization of the parameter values that works well for different catchments and forcing data sets. The differences in parameter set performance are even larger when different forcing data sets are used.
First of all, this indicates that a proper GHM model calibration should be catchment specific, as the dominant hydrologic processes vary between individual catchments. In addition, all processes are schematized in a simplified form within the hydrological model and in each catchment optimal parameter values are compensating for other simplifications and possible biases in the model schematization. And although global hydrological models are consistently parameterized using existing global datasets, the quality of the underlying data is variable and the calibrated parameters will compensate for spatial varying errors for each catchment differently. Therefore, the current generation of global scale hydrologic models is simply not able to capture the dynamics across a range of catchments through a single realization of parameter values, whatever the skill and imagination of the modeler.
Secondly, model calibration calibration compensates for the substantial uncertainty in forcing datasets and therefore each time another forcing dataset is used the model should be recalibrated. Unfortunately, this poses a severe problem given the computational demands of global scale hydrologic models such as PCR-GLOBWB (Candogan et al., 2011; Wada et al., 2010; Sperna Weiland et al., 2010). This is rather unfortunate as large improvements in predictive ability are to be expected when optimizing the model parameters. Especially in climate impact studies, in which a model such as PCR-GLOB-

WB may be forced with meteorological data from an ensemble of GCMs for different time-slices (Vaze et al., 2010; Sperna Weiland et al, 2010) a global scale hydrologic model should preferably be used in its un-calibrated form where all parameter estimates are based on best available global data sets (Van Beek et al., 2011). Only if one single forcing data set is employed, as for example in a flood forecasting system (Candogan et al., 2011), it may be feasible to conduct an explicit model calibration.

2.4 Conclusions

The aim of this study was to explore the possibilities for the calibration of the global hydrological model PCR-GLOBWB by investigating the influence of both parameter and meteorological forcing uncertainty. From this analysis the following conclusions can be drawn:

- By using simple Latin Hypercube parameter sampling, parameter combinations can be obtained that outperform the default model parameterization.
- Parameter performance is not persistent over time. A set of behavioral parameter combinations should be sought for that account for uncertainties in meteorological forcing data, measurement data and the hydrological model and consequently perform well in both the calibration and evaluation period.
- Performance of behavioral parameters sets has limited to no persistence when being transferred to other rivers and or forcing data sets. Therefore, a catchment and forcing specific model calibration seems required and even than, the resulting optimal parameterization may not be robust under changing environmental conditions.
- Parameter uncertainty constitutes only a relatively minor part of hydrograph prediction uncertainty. This demonstrates that the apparent dichotomy between model predictions and data cannot be resolved by increasing model complexity and resolving sub-grid processes. Instead, what is much-needed is an improved characterization of global rainfall amounts at spatial resolution of 0.5 degrees and smaller.

Acknowledgements

The global discharge timeseries have been obtained from the Global Runoff Data Centre. Discharge data for the Mekong was provided by the Mekong River Commission. The CFSR reanalysis product used in this study originates from the Research Data Archive (RDA) which is maintained by the Computational and Information Systems Laboratory (CISL) at the National Center for Atmospheric Research (NCAR) and was downloaded and converted by Clement Tisseuil. The ERA-40 and ERA-interim data have been obtained from the ECMWF with help from Florian Pappenberger and the GLOWASIS project. Finally, we acknowledge Yoshi Wada for sharing the source code of his water demand model.

3 The ability of a GCM-forced hydrological model to reproduce global discharge variability

This chapter is based on:
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. Hydrology and Earth System Sciences 14, pp. 1595-1621

Abstract

Data from General Circulation Models (GCMs) are often used to investigate hydrological impacts of climate change. However GCM data are known to have large biases, especially for precipitation. In this study the usefulness of GCM data for hydrological studies, with focus on discharge variability and extremes, was tested by using bias-corrected daily climate data of the 20C3M control experiment from a selection of twelve GCMs as input to the global hydrological model PCR-GLOBWB. Results of these runs were compared with discharge observations of the GRDC and discharges calculated from model runs based on two meteorological datasets constructed from the observation-based CRU TS2.1 and ERA-40 reanalysis. In the first dataset the CRU TS 2.1 monthly time series were downscaled to daily time series using the ERA-40 dataset (*ERA6190*). This dataset served as a best guess of the past climate and was used to analyze the performance of PCR-GLOBWB. The second dataset was created from the ERA-40 time series bias-corrected with the CRU TS 2.1 dataset, using the same bias-correction method as applied to the GCM datasets (*ERACLM*). Through this dataset the influence of the bias-correction method was quantified. The bias-correction was limited to monthly mean values of precipitation, potential evaporation and temperature, as our focus was on the reproduction of inter and intra-annual variability (IAV).
After bias-correction the spread in discharge results of the GCM based runs decreased and results were similar to results of the ERA-40 based runs. Especially for rivers with

a strong seasonal pattern. Overall the bias-correction method resulted in a slight reduction of global runoff and the method performed less well in arid and mountainous regions. However, deviations between GCM results and GRDC statistics did decrease for $\overline{Q}$, Q90 and IAV. After bias-correction consistency amongst models was high for mean discharge and timing (Qpeak), but relatively low for IAV. This suggests that GCMs can be of use in global hydrological impact studies in which persistence is of less relevance (e.g. in case of flood rather than drought studies). Furthermore, the bias-correction influences mean discharges more than extremes, which has the positive consequence that changes in daily rainfall distribution and subsequent changes in discharge extremes will also be preserved when the bias-correction method is applied to future GCM datasets. However, it also shows that agreement between GCMs remains relatively small for discharge extremes.
Because of the large deviations between observed and simulated discharge, in which both errors in climate forcing, model structure and to a lesser extent observations are accumulated, it is advisable not to work with absolute discharge values for the derivation of future discharge projections, but rather calculate relative changes by dividing the absolute change by the absolute discharge calculated for the control experiment.

3.1 Introduction

Runoff regimes might change significantly due to climate change. Therefore strategies for water management are sought for that either mitigate the undesired effects of a changing climate or gain from the positive effects. The search for these strategies relies on reliable assessments of the effect of climate change on river discharge. Consequently, much research has been conducted investigating the hydrological response to climate change, both on local (Christensen and Lettenmaier, 2007; Prudhomme and Davies, 2008; Buytaert et al., 2009), regional (Lehner et al., 2006; Strzepek and Yates, 1997; Hagemann et al., 2009; Hurkmans et al., 2010) and global scale (Arnell, 1999a, 2003; Alcamo et al., 2000; Alcamo and Henrichs, 2002; Milly, 2005).
Future projections can be calculated using change factors; long-term average changes derived from climate model data which are applied to observed meteorological time series to obtain future meteorological time series as input for hydrological models (Alcamo et al., 2007; Nijssen et al., 2001; Vörösmarty et al., 2000). However, with this method changes in temporal variability are ignored. Therefore climate datasets from General Circulation Models (GCM) are often directly used as input for hydrological models when investigating possible changes. Unfortunately, different GCM datasets produce varying and even contradicting results (Varis et al., 2004). GCM results of the current climate control run (20C3M) do not always agree with the observed climate, deviations are especially apparent for precipitation (Covey et al., 2003; Perkins and Pitman, 2009). Often too many days with light rain are simulated and the frequency and amount of heavy rain events are underestimated (Dai, 2006). Because of the variance

amongst GCMs, many studies concluded that a multi-model ensemble of GCMs should be used to obtain a reliable impression of the spread of possible regional changes and the uncertainties accompanying these changes (Murphy et al., 2004; Boorman and Sefton, 1997; IPCC, 2007b). Furthermore, it has been widely recognized (Wood et al., 2004; Leander and Buishand, 2007; Fowler and Kilsby, 2007; Wilby et al., 1998) that precipitation data needs to be bias-corrected before they can be used.

Previous studies (Milly et al., 2005; Nohara et al., 2006) already investigated the correct reproduction of mean discharge and runoff regimes. Therefore we focus on the ability of a GCM-forced hydrological model to reproduce global discharge variability (extremes, seasonal and inter-annual variation), parameters relevant for water management. We realize that even if we obtain a correct reproduction of current discharge variability this is no guarantee that projected discharge variability is correct as well (Prudhomme and Davies, 2008). However, if discharge variability is biased for the current period, future projections of discharge variability are likely to be biased as well.

We restrict ourselves to a bias-correction of monthly mean values of GCM precipitation, temperature and potential evaporation (potential evaporation was derived using either Penman-Monteith or Blaney-Criddle, depending on data availability). Although a correction on monthly means does not guarantee that rainfall marginal distributions are well reproduced (Dai, 2006). No additional correction on GCM variability is applied for two reasons: First, because several climate change experiments have shown that GCM variability will change, especially for precipitation. Precipitation frequency and storm duration is likely to decrease, while intensity will increase resulting in heavier rain events (Trenberth et al., 2003; Allan and Soden, 2008; Meehl et al., 2000). Therefore, corrections based on past deviations between modeled and observed variability, although often applied (Leander and Buishand, 2007; Ines and Hansen, 2006; Wood et al., 2004), may not hold in future. Applying variability corrections might even mask important changes, since changes in rainfall distributions may have a much larger effect on the hydrological cycle than changes in mean precipitation (Allen and Ingram, 2002). Second, as discussed before, the very goal of this study is to examine the effect of GCM variability on modeled hydrological variability, thereby assessing the usability of GCM data for water management related climate effect studies. Replacing the differences in variability between GCMs to a single current observed variability (i.e. CRU) would rule out such an analysis.

Our evaluation of the ability of a GCM-forced hydrological model to reproduce global discharge variability consists of three steps. First, the performance of PCR-GLOBWB is assessed when forced directly with the CRU TS 2.1 monthly time series (New et al., 2000) which are downscaled to daily values with the ERA-40 reanalysis (Uppala et al., 2005) for the period 1961 to 1990. The CRU TS 2.1 time series are derived by interpolation of monthly meteorological observations while the ERA-40 reanalysis reflects the daily variations in the large-scale meteorological conditions over the globe. Thus, the down-scaled data set provides a best guess of the inter-annual and seasonal variability.

River discharge statistics resulting from the PCR-GLOBWB run driven with this forcing dataset are then compared with observed river discharge statistics from the GRDC dataset (GRDC, 2007). Although this comparison is subject to errors in the forcing dataset, evaluation of the model results from this run provides the best available information on the performance of PCR-GLOWB.
In a second step, the ERA-40 dataset is bias-corrected using the same method as used for the GCM datasets as a yearly month-by-month correction with the CRU TS2.1 monthly time series, similar to step 1, is not possible for GCM datasets; GCM time series only provide a realization of a given climate and do not represent the observed meteorology of a specific year. Rather, each monthly GCM field is updated in a manner that scales the 30-year monthly average GCM quantity to the 30-year monthly average CRU TS2.1 quantity. With this method the inter-annual variability remains similar to that of the original GCM or ERA-40 time series. Results of the hydrological model run driven with this forcing dataset are compared with results of the hydrological model run of step 1 and with observed GRDC data in order to quantify the influence of the bias-correction method as this should preserve the full temporal variability of the modeled meteorological product (ERA-40, GCMs).
In the last step PCR-GLOBWB is forced with the meteorological datasets of 12 GCMs, bias-corrected according to the method described in step 2. To quantify the spread between the results of the GCMs the coefficient of variation (CV) is calculated for a number of statistical quantities derived for the ensemble of GCMs. Combined with the findings on model performance (step 1) and loss of observed temporal variability (step 2), these results will show the usability of meteorological GCM datasets in hydrological impact studies with a focus on extremes and discharge variability.

3.2 Data and methods

3.2.1 Existing global hydrological models

Obviously, the global hydrological model (GHM) PCR-GLOBWB follows in a long line of existing GHMs. Without attempting to be complete, we refer to short reviews given by Arnell (1999a) and Döll et al. (2003) describing VIC (Nijssen et al., 2001), Macro-PDM (Arnell,1999), WBM (Vörösmarty et al., 1998a) and WGHM (Döll et al., 2003), four models frequently used in large-scale hydrological studies. Similar to PCR-GLOBWB, the last three models calculate for each time step the water balance of all individual grid cells. The grids of WBM and WGHM have a resolution of 0.5 degrees, corresponding to the finest resolution of most climate datasets available, within Macro-PDM grid cells can either be regular or catchment shaped. All three models contain at least one soil water layer and total runoff consists of a fast overland and a slow groundwater component. Size and partitioning of these fluxes depend on the degree of saturation of the soil water layer(s) that is calculated, either physically based or described by a statistical relation. The models all apply some form of routing to obtain realistic river

discharge. WGHM (Döll et al., 2003) is the sub- model of the global water use and availability model WaterGAP (Alcamo et al., 2003).
In addition to stand-alone hydrological models, global water balances have also been modeled by coupled vegetation water balance models and by land surface schemes (LSS) used in global climate models. Examples of global coupled vegetation water balance models are GEPIC (Liu et al., 2009) and LPJ (Gerten et al., 2004). GEPIC focuses on the calculation of crop yield and crop water productivity. It combines the extent of crop covered areas (with different properties for a variety of crop types), Hargreaves potential evaporation and soil water availability to calculate crop water use. LPJ is a global dynamic vegetation model that describes the interaction between the terrestrial biosphere and the water cycle. Runoff is one of the outputs of this vegetation model and exists of excess over field capacity from the upper two soil layers and percolation from the second soil layer. Runoff is translated to river discharges using a routing scheme. At this stage the global water balance and basin runoff is represented less reliable by most LSSs than by hydrological models (Gerten at al., 2004) and the resolution is too coarse for hydrological studies.
A notable exception is global VIC (Nijssen et al., 2001). VIC is designed to be a land surface scheme for climate models and solves both the energy and the water balance. At the global scale it currently has a resolution of 1 degrees (Sheffield et al., 2009), where vegetation, soil moisture and the application of precipitation are modeled using subgrid variability schemes. Runoff, existing of baseflow from the lower soil moisture store and fast response flow, is routed along a routing network using a convolution approach. VIC is one of the few LSSs that is frequently used in stand-alone hydrological studies (Hurkmans et al., 2010).
The main focus of this study is the reproduction of variability in runoff regimes. Therefore the hydrological model used should be designed to calculate this variability, including a good representation of hydrology. In addition to the models described above, PCR-GLOBWB contains an advanced scheme for the subgrid parameterization of surface runoff, interflow and baseflow and an explicit routing scheme for surface water flow using the kinematic wave approximation that includes retention in, and evaporation loss from, wetlands, lakes and reservoirs (Van Beek and Bierkens, 2009).

3.2.2 PCR-GLOBWB

PCR-GLOBWB is a global distributed hydrological model with a resolution of 0.5 degrees (Van Beek and Bierkens, 2009; Bierkens and van Beek, 2009). Each model cell consists of two vertical soil layers and one underlying groundwater reservoir. Sub-grid parameterization is used to represent short and tall vegetation, surface water and for calculation of saturated areas for surface runoff as well as interflow. Water enters the cell as precipitation and can be stored as canopy interception or snow. Snow melt or accumulation occurs depending on temperature. Melt water and throughfall are passed to the surface. Evapotranspiration is calculated from the potential evaporation and soil

moisture conditions. Exchange of water is possible between the soil and groundwater layers in both downward and upward direction depending on soil moisture status and groundwater storage. Total runoff consists of non-infiltrating melt water, saturation excess surface runoff, interflow and base flow. For each time step the water balance is computed per cell. Runoff is accumulated and transferred as river discharge along the drainage network using kinematic wave routing. The drainage network is taken from DDM30 (Döll and Lehner, 2002) and lakes, wetlands and large reservoirs are obtained from the GLWD dataset (Lehner and Döll, 2004). The model calculates natural flows. Water use is not considered and only recently, in a newer version of PCR-GLOBWB, reservoir management has been included.

3.2.3 Data

GCM data

In 1997 the IPCC developed a set of emission scenarios, representing possible future climate change and provided boundary conditions to be used in GCM runs. These scenarios are widely used in climate impact studies. In addition to boundary conditions for studies based on these scenarios, the IPCC also provided boundary conditions for a 20 century control experiment (20C3M). Climate modelling centers around the world conducted GCM runs with this data. The Program for Climate Model Diagnosis and Intercomparison (PCMDI) has collected the results and made these available through the PCMDI data portal (https://esg.llnl.gov:8443/index.jsp). The PCMDI provides results on a daily time step, whereas the IPCC data portal only provides derived monthly averages. Although it has been said that daily values are less reliable (Prudhomme et al., 2002), we prefer to use daily data, since they provide more information on extremes and climate variability.

We collected data from the GCMs for which complete model datasets are provided on a daily time step for both the 20C3M control experiment and the future scenario A1B, which we will analyze in future research. We selected the period 1961 to 1990, assuming that a thirty year period would be long enough to represent inter-annual variability, and all GCMs provide data for this period. For those GCMs with multiple runs available the first run was selected. We worked with a 30-year period of day values assuming this period is that long that potential biases in the selected GCM run are averaged out. Although the data portal does not provide all required parameters for the (state-of-the-art) Hadley centre climate models, HadGEM1 has been included. Data were retrieved from the CERA-gateway (http://cera-www.dkrz.de). Table 3.1 gives an overview of the selected models.

ERA-40 re-analysis data

Precipitation, temperature and evaporation were collected from the ERA-40 reanalysis as well. The ERA-40 dataset is obtained with a numerical weather prediction system

Table 3.1 Overview of selected GCMs.

Model	*Institute*	*Country*	*Acronym*
BCM2.0	Bjerknes Centre for Climate Research	Norway	BCCR
CGCM3.1	Canadian Centre for Climate Modelling and Analysis	Canada	CCCMA
CGCM2.3.2	Meteorological Research Institute	Japan	CGCM
CSIRO-Mk3.0	Commonwealth Scientific and Industrial Research Organization	Australia	CSIRO
ECHAM5	Max Planck Institute	Germany	ECHAM
ECHO-G	Freie Universität Berlin	Germany	ECHO
GFDL-CM 2.1	Geophysical Fluid Dynamics Centre	USA	GFDL
GISS-ER	Goddard institute for Space Studies	USA	GISS
IPSL-CM4	Institute Pierre Simon Laplace	France	IPSL
MIROC3.2	Center of Climate System Research	Japan	MIROC
CCSM3	National Center for Atmospheric Research	USA	NCAR
HADGEM1	Met Office's Hadley Centre for Climate Prediction	UK	HADGEM

that includes assimilation of meteorological observations. The prediction system reasonably represents the observed climate. It has a relative high horizontal resolution (ca. 125 km) and a vertical resolution of 60 levels (Uppala et al., 2005). Still precipitation is poorly approximated in ERA-40 data, particularly in the tropics (Troccoli and Kålberg, 2004). The ERA-40 datasets slightly overestimates precipitation globally. Overestimations are largest in the Congo basin, the southern slope of the Himalaya and the Andes region, while precipitation is underestimated in the Murray, Danube and Mississippi basins. In an attempt to eliminate the cold Arctic bias present in the ERA-15 dataset, a warm bias during winter is introduced in the Arctic region in ERA-40 dataset. Inter-annual variability is relatively high because the observational data included in the system vary over time (Hagemann et al., 2005). This variation increased, amongst others, the inter-annual variability of precipitation totals, which are relatively high after the first satellite data were included in the system for 1972.

CRU data

Precipitation, temperature and the available parameters required to derive Penman-Monteith reference potential evapotranspiration have been retrieved from the monthly CRU TS 2.1 timeseries (New et al., 2000). Since radiation and wind speed are not provided as monthly timeseries, the climatology from the CLIM 1.0 (New et al., 1999) was used. The CRU data is based on station data that is interpolated as a function of longitude, latitude and elevation above sea level using the thin-plate spline method. The CRU TS 2.1 timeseries have been evaluated in several studies. Unfortunately no benchmark meteorological dataset exists and evaluations are mainly based on differences between historical datasets. Available historical data sources for precipitation (FAO, GHCN and CRU) correspond well with each other (Beck et al., 2004). Uncertainties in precipitation amounts and seasonality are largest in dry regions (Fekete et al., 2004). In Arctic regions the CRU datasets underestimate precipitation amounts due to snow undercatch problems (Fiedler and Döll, 2007).

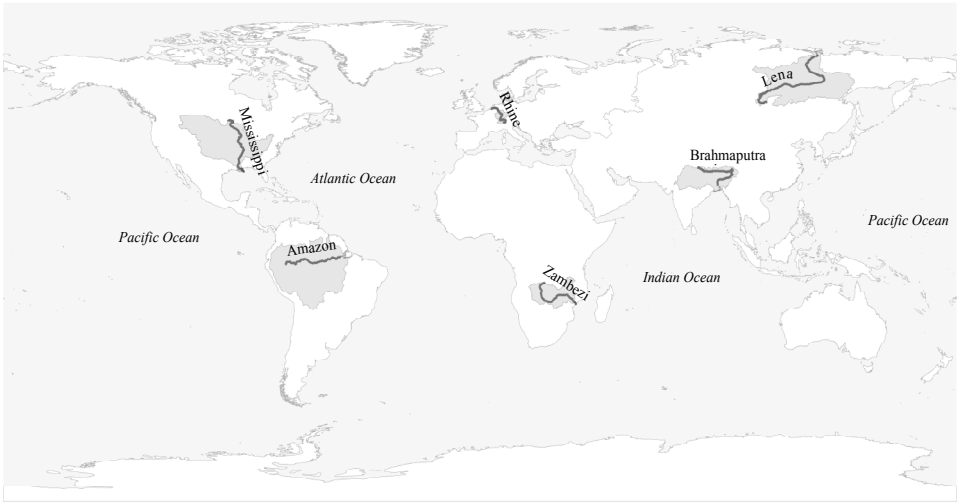
Hydrological data
We selected 19 large catchments to be evaluated in this study, covering a variety of climate zones, latitudes and continents. The selected catchments are shown in Fig. 3.1, together with basin characteristics. Discharge data are obtained from the Global Runoff Data Center (GRDC, 2007). For most rivers, daily data was available for at least part of the period of interest. For the Yangtze, Yellow River and Parana we retrieved monthly GRDC data and for the Indus we could only retrieve monthly data from the RivDis database (Vörösmarty et al., 1998b).

3.2.4 Derivation of potential evaporation

Evaporation is derived with the Penman-Monteith equation (Monteith, 1965). For those GCMs where the required atmospheric surface pressure data were not available, surface pressure was derived from the pressure at sea level using a global DEM. Air humidity fields, required to calculate the actual vapor pressure, could not be retrieved from the data portal for the complete period for some of the GCMs. Therefore we used a simplified method to calculate the actual vapor pressure from the minimum air temperature (Allen et al., 1998). For arid regions, the assumption that the air is saturated when the temperature is at its minimum, might not hold and therefore the minimum temperature will not equal the dew temperature. As suggested by Allen (1998) we subtracted 2 degrees from the minimum temperature in arid regions. The arid regions have been selected using the climate moisture indices of the WWDRII (UN, 2006). For those models where timeseries for other required parameters were missing or incomplete we used the Blaney-Criddle equation (Brouwer and Heibloem, 1986; Oudin et al., 2005) instead of Penman-Monteith. We realize this may have introduced additional noise between the model results (Kay and Davies, 2008). For several models we compared the potential evaporation we calculated using Penman-Monteith with evaporation calculated with the Blaney-Criddle equation. The full analysis has not been included in this article for brevity, however potential evaporation calculated with the Blaney-Criddle equation is relatively high for Europe, North-America and the north of Asia during summer and relatively low for Africa and Asia. Especially the overestimation in the Northern continents can result in deviations because, particularly at the beginning of summer, evaporation will not be limited by water availability and actual evaporation will be too high as well. However, hydrological model studies are forced to use what has been reported by the GCMs host institutes and evaporation is not provided for most models.

3.2.5 Bias correction of ERA-40 and GCM data with 30-year mean month values

For the bias-correction of the GCM and ERA-40 datasets, monthly scaling factors were calculated from the difference (temperature) or ratio (precipitation and evaportion) in 30-year average monthly means between the CRU TS 2.1 timeseries (New et al., 2000) and GCM/ERA-40 timerseries for the period 1961 to 1990. Applying the monthly cor-



Catchment	*Area* (km^2)	*Qavg* (m^3/s)	*Gauge*	*Catchment*	*Area* (km^2)	*Qavg* (m^3/s)	*Gauge*
Amazon	6.915.000	190.000	Obidos	Murray	1.061.469	767	Wakool Junction
Brahmaputra	930.000	48.160	Bahadurabad	Niger	2.117.700	6.000	Dire
Congo River	3.680.000	41.800	Kinshasa	Orange river	973.000	365	Aliwal North
Danube	817.000	6.400	Ceatal Izmail	Parana	2.582.672	18.000	Corientes
Ganges	907.000	12.015	Hardinge Bridge	Rhine	65.683	2.200	Rees
Indus	1.165.000	6.600	Kotri	Volga	1.800.000	31.900	Volgograd
Lena	2.500.000	17.000	Kusur	Yangtze	752.000	2.571	Datong
MacKenzie	1.805.000	10.700	Norman Wells	Yellow river	1.380.000	8.060	Huayuankou
Mekong	2.981.076	12.743	Mukdahan	Zambezi	1.390.000	3.400	Katom a Mulilo
Mississippi	795.000	16.000	Vicksburg				

Figure 3.1 Selected catchments.

rection factors also implies downscaling GCM data spatially to the higher spatial resolution of the CRU dataset, since correction factors are calculated for all individual cells of the CRU grid. In this section both the GCM and ERA-40 data will be referred to as modeled dataset (MOD), the bias-corrected ERA-40 dataset created in this section will be referred to as *ERACLM*.
For temperature an additive correction was used:

$$T_{corrected\_MOD} = T_{MOD} + (\overline{T}_{CRU} - \overline{T}_{MOD}) \tag{3.1}$$

Where T is the daily temperature (K) and $\overline{T}$ is the 30-year average monthly temperature. For evaporation (*ETP*) a multiplicative correction is used, to avoid the occurrence of negative evaporation.

$$PET_{corrected\_MOD} = PET_{MOD} \frac{\overline{PET}_{CRU}}{\overline{PET}_{MOD}} \tag{3.2}$$

Where PET is the daily potential evaporation (mm/day) and $\overline{PET}$ is the 30-year average monthly potential evaporation. For precipitation a similar multiplicative correction is used as for evaporation.

$$P_{corrected\_MOD} = P_{MOD} \frac{\overline{P}_{CRU}}{\overline{P}_{MOD}} \tag{3.3}$$

Where P is the daily precipitation (mm/day) and $\overline{P}$ is the 30-year average monthly precipitation. For some regions (North-Africa, Amazon and Himalaya) differences between the modeled monthly precipitation amount and CRU TS 2.1 monthly precipitation amounts and number of wet days can be very large. In these regions a simple multiplicative correction resulted in unrealistic precipitation peaks in the bias-corrected precipitation time series. Therefore the bias-correction of precipitation, Eq. (3.3), is extended with a minimum daily precipitation amount that has to be exceeded by the modeled total monthly precipitation amount before the multiplicative correction can be used (Van Beek, 2008). The threshold equals the monthly mean daily CRU precipitation amount:

$$P_{crit} = \frac{\overline{P}_{CRU}}{\overline{W}_{CRU}} \tag{3.4}$$

Where $\overline{W}_{CRU}$ is the 30-year average number of wet days for the specific month. In addition a threshold value of 10 is set for the maximum value of the multiplicative correction factor ($\overline{P}_{CRU} / \overline{P}_{MOD}$). If 1) the monthly precipitation sum of the GCM does not exceed the threshold P_{crit} or if 2) the multiplicative correction factor is higher than 10, the days with precipitation occurrence are calculated from a temperature limit below which a day becomes wet. With this method the number of wet days is increased to avoid large rain events on the few days with rain in the GCM timeseries. The temperature limit is defined by:

$$T_{critMOD} = T_{\min MOD} + (T_{\max MOD} - T_{\min MOD}) \cdot \frac{\overline{W}_{CRU}}{D} \tag{3.5}$$

Where T_{min} and T_{max} are the minimum and maximum temperature (K) of the given month and D is the total number of days in the specific month. With this formula the

number of modeled wet days per month is calculated (W_{MOD}) and the precipitation amount for these days equals:

$$P_{corrected\_MOD} = \frac{\overline{P}_{CRU}}{\overline{W}_{MOD}} \tag{3.6}$$

With this equation, precipitation is equally distributed over the wet days and the original temporal distribution of the modeled precipitation time series is lost. However, this correction method only had to be applied for a limited number of cells.

3.2.6 Downscaling of CRU monthly timeseries to daily timeseries with ERA-40

The CRU TS 2.1 monthly time series are downscaled to daily time series with the ERA-40 reanalysis dataset. The method used, resembles the bias-correction method described in Sect. 3.2.5. The only exception is that the 30-year average monthly precipitation, temperature, evaporation and number of wet days ($\overline{P}, \overline{T}, \overline{PET}$ and $\overline{W}_{CRU}$) should be replaced in all equations by the monthly mean values of the specific month in the specific year. Since for the ERA-40 dataset a year by year correction with the CRU TS 2.1 time series can be made. The resulting dataset will be referred to as *ERA6190* and is used in the first step of this study for the evaluation of the hydrological model.

3.2.7 Statistical analysis

Statistics

We calculated the thirty year average mean discharge, Q10 (low flow) and Q90 (high flow) percentile values for each catchment for results of all model runs and the observed discharge time series. From thirty year average quantities of the individual GCMs we calculated the GCM ensemble mean. When the mean modeled discharge deviates from the mean observed discharge, Q90 and Q10 values are likely to deviate as well. Therefore, instead of comparing the absolute Q10 and Q90 values, we compared the Q90 and Q10 values relative to the mean discharge (see Table 3.2). These quantities provide information on the shape of the discharge PDF, the positions of Q90 and Q10 are given relative to the mean discharge.

To compare the temporal behavior of the GCM and ERA-40 based model runs with observations, we also calculated the inter-annual discharge variability for the thirty year annual average discharges and the yearly month of regime peak occurrence (Table 3.2). Furthermore, the coefficient of variation (CV) of the ensemble of 12 GCM results is calculated for all statistics according to:

$$CV = \frac{\sqrt{\frac{1}{M}\sum_{m=1}^{12}(\overline{Q}_m - \overline{Q}_{ENS})^2}}{\overline{Q}_{ENS}} \quad (3.7)$$

Where M is the number of GCMs (12), m corresponds to one of the 12 GCMs, $\overline{Q}_m$ is the 30-year average discharge for the specific GCM and $\overline{Q}_{ENS}$ is the GCM ensemble mean 30-year average discharge; all discharges in m^3/s. In Eq. (3.7) $\overline{Q}$ is used as an example, however the same equation is used for relative *Q10*, relative *Q90*, *Qpeak* and IAV. The CV is a measure for the ensemble spread and indicates the resemblance of the PCR-GLOBWB runs driven with the different bias-corrected GCM datasets.

Regime curves
The hydrological regime curve consists of the 30-year average mean monthly discharges, calculated for all twelve months individually from modeled discharge timeseries at the station locations (Fig. 3.1). Regime curves have been calculated from: (1) discharge simulated with the *ERA6190* dataset, (2) discharge simulated with the *ERACLM* dataset, (3) discharges simulated for each individual bias-corrected GCM dataset, (4) the ensemble mean of the discharges calculated by these individual GCMs and for comparison (5) the observed GRDC discharges. In addition to the plots of the regime curves, we calculated Nash-Sutcliffe (NS) coefficients (Nash and Sutcliffe, 1970) from the 30-year average monthly values displayed in the regime curves.

Map comparison
PCR-GLOBWB calculates daily maps of routed accumulated runoff per cell. For all statistical variables (e.g. $\overline{Q}$, relative Q90 and Q10, Qpeak and IAV) the GCM ensemble mean results and the results of the two ERA-40 based runs are compared. Maps with the CV are derived from the ensemble of GCM results for all variables as well. These maps show a spatial distribution of the spread between the results of the bias-corrected GCM runs.

3.3 Results

3.3.1 PCR-GLOBWB validation using ERA6190

To assess its performance we ran the model PCR-GLOBWB for the period 1961 to 1990 with the monthly CRU TS 2.1 timeseries, downscaled to daily values using the ERA-40 re-analysis data. This forcing dataset (*ERA6190*) provides the best available guess of the current climate. Despite its coarse resolution PCR-GLOBWB can reproduce observed discharges reasonably well for most selected catchments. It should be noted that besides by the hydrological model structure, errors can also be introduced by the model input, in particular through the meteorological forcing.

Table 3.2 Descriptive statistics. Where Qday is a daily discharge value, *l* corresponds to the day number, *L* is the total number of days in the 30-year period, Q90 is the discharge exceeded at 10% of the days during the 30-year period, Q10 is the discharge exceeded at 90% of the days during the 30-year period, *h* corresponds to the year number, *H* is the total number of years (30), $Qyear_p$ is the mean annual discharge for the years 1961 to 1990, $\overline{Q}_{month}$ is the 30 year average mean discharge and is the 30-year average mean monthly discharge.

Parameter	*Description*	*Equation*	*Unit*
$\overline{Q}$	Thirty year average mean annual discharge	$\overline{Q} = \frac{1}{L}\sum_{l=1}^{L} Qday_l$	m^3/s
relative Q90	Discharge exceeded at 10% of the days during the thirty year period relative to the mean discharge	$Q90 / \overline{Q}$	-
relative Q10	Discharge exceeded at 90% of the days during the thirty year period relative to the mean discharge	$Q10 / \overline{Q}$	-
IAVrel	Inter annual variability derived from yearly average discharges	$IAV = \frac{\sqrt{\frac{1}{H}\sum_{h=1}^{30}(Qyear_h - \overline{Q})^2}}{\overline{Q}}$	-
Qpeak	Month of peak in annual discharge cycle	$Qpeak = \max(\overline{Q}_{month})$	month

* For the stations where no daily time-series were available (Indus, Parana, Yangtze and Yellow River) the observed Q90 and Q10 values could not be calculated and are therefore not included in the analysis.

Continental discahrges

In Table 3.3 the mean yearly continental discharges calculated with PCR-GLOBWB, are compared with continental discharges of previous studies calculated from both observations and model results (Wada et al., 2008). Continental discharges vary amongst studies. For Asia, North-America, South-America and Oceania the results of PCR-GLOBWB are within the range of continental discharges found in previous studies. Differences between the discharges modeled by PCR-GLOBWB and the average continental discharges of previous studies are small, globally the average difference is only 1%. For Europe PCR-GLOBWB slightly underestimates continental discharge. For Africa PCR-GLOBWB modeled discharge is on the high side, with PCR-GLOBWB overestimating the discharge for many African rivers. This is caused by the absence of water withdrawals in the model and limited river bed losses, like recharge to the groundwater and evaporation (which is particular relevant in the deltas of the Niger). In addition, the density of meteorological observations is relatively low for the African continent which limits the constraint of the CRU TS 2.1 and ERA-40 datasets (Fekete et al., 2004; Van Beek and Bierkens, 2009).

Discharge statistics

Figure 3.2a shows the annual mean modeled and observed discharges and Fig. 3.2b shows the modeled discharge as a percentage of the observed GRDC discharge. The annual mean modeled discharge shows little deviation from observed annual mean dis-

Table 3.3 Continental runoff based on data and model based estimates in km^3/year.

	Europe	*Asia*	*Africa*	*N.-America*	*S.-America*	*Oceania*	*Global*	*Time-period*
Data based estimates								
Baumgartner and Reichel (1975)	2564	12467	3409	5840	11039	2394	37713	-
Korzun et al. (1978)	2970	14100	4600	8180	12200	2510	44560	-
L'vovich (1979)	3110	13190	4225	5960	10380	1965	38830	-
Shiklomanov (1997)	2900	13508	4040	7770	12030	2400	42648	1921-1990
GRDC (2004)	3083	13848	3690	6294	11897	1722	40533	1961-1990
Average	2925	13423	3993	6809	11509	2198	40857	-
Model based estimates								
Fekete et al. (2000)	2772	13091	4517	5892	11715	1320	39319	-
Vörösmarty et al. (2000)	2770	13700	4520	5890	11700	714	39294	1961-1990
Nijssen et al. (2001)	-	-	3615	6223	10180	1712	36006	1980-1993
Oki et al. (2001)	2191	9385	3616	3824	8789	1680	29485	1987-1988
Döll et al. (2003)	2763	11234	3592	5540	11382	2239	36687	1961-1990
Widén-Nilsson et al. (2007)	3669	13611	3738	7009	9448	1129	38605	1961-1990
Average	2833	12204	3933	5730	10536	1466	36566	-
PCR-GLOBWB:								
ERA6190	2143	11461	5573	5249	11186	2633	36245	1961-1990
ERACLM	2159	10285	5223	4848	10174	2384	35075	1961-1990
GCM mean	2175	10572	5099	4803	10678	2371	35699	1961-1990

* This table with continental discharges from previous studies has been obtained from Wada et al. (2008)

charge for the Danube, Volga, Rhine, Mississippi and Lena. Deviations are relatively large for the drier basins; Murray, Niger, Zambezi and Orange. Modeled discharge for the MacKenzie is too low due to undercatch in the CRU snowfall amounts (Fiedler and Döll, 2007). Discharge is also relatively low for the Monsoon influenced Asiatic rivers; Brahmaputra, Indus, Mekong and Yangtze, while for the Ganges and Yellow River discharges are on the high side.

To investigate the possible influence of water use, we made the simplifying assumption that water demand equals water use. In reality water demand will, especially in drier regions, be higher than water use. Water demand is estimated on a grid of 0.5 degrees. It is the sum of estimated industrial, agricultural and domestic water demand (Wada et al., 2008). Calculated discharge is compared with the sum of observed discharge and water demand, which approximates natural discharge. Figure 3.2b shows that for the

Murray the deviations between observed and calculated discharge can partly be assigned to the lack of inclusion of water use. To a lesser extent this also applies to the Danube, Ganges, Yellow River and Rhine.
Besides mean discharge statistics, the positions of discharge extremes (e.g. the positions of Q90 and Q10 values relative to $\overline{Q}$) have been compared for those basins where daily GRDC timeseries were available, see Figs. 3.3 and 3.4 where the blue dots represent the results of the *ERA6190* run and black dots are GRDC values. In this analysis we will focus on the catchments with the largest deviations.

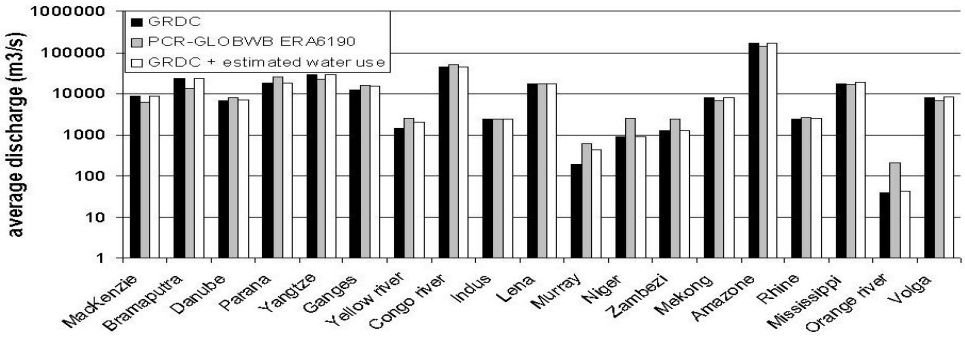


Figure 3.2a Annual mean observed discharge (GRDC), annual mean discharge calculated with PCR-GLOBWB forced with the ERA6190 dataset (PCR-GLOBWB ERA6190) and the sum of observed annual mean discharge and estimated water use (GRDC + estimated water use) as an approximation of natural flow.

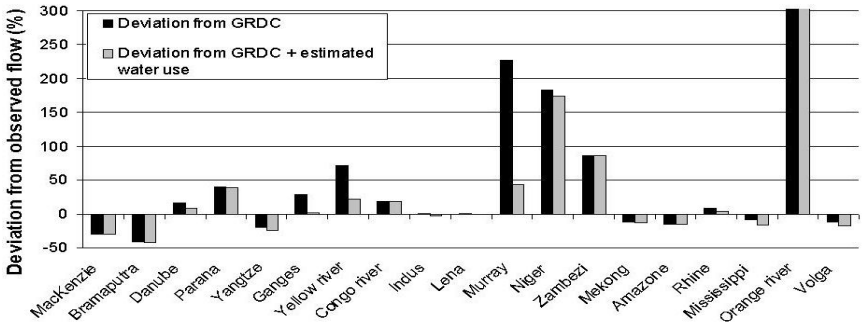


Figure 3.2b Percentage deviation of annual mean discharge (PCRGLOB-WB ERA6190) from 1) observed annual mean discharge (deviation from GRDC) and 2) the sum of observed annual mean discharge and estimated water use as an approximation of natural flow (Deviation from GRDC + estimated water use).

The relative position of the Q90 discharge is higher in the model results than in the GRDC observations. For the drier basins; Orange and Zambezi, the position of the Q10 discharge is too high as well as for the Niger, Orange and Murray, but too low for the Zambezi. For the Ganges the relative Q90 discharge is too low and for both the Lena and the MacKenzie, the relative Q90 is underestimated as well due to the relative low snow melt driven discharge rise in spring, yet for these rivers the position of Q10 flows are overestimated due to a higher baseflow during winter.

Deviations in relative inter-annual variability are small for European catchments (Fig. 3.5). IAV is underestimated in the drier catchments; Murray, Orange and Niger, in the Arctic catchments and in the Monsoon influenced basins; Brahmaputra, Indus, Ganges and Yellow River. For the remaining catchments differences are small.
Month of peak discharge occurrence is modeled well for the Arctic and most monsoon influenced basins (Fig. 3.6). Large deviations in timing are present for all dry catchments, because here discharge is low and relatively constant throughout the year.
For the Rhine, the discharge peak is modeled in April and observed in February due to a too late snow melt driven discharge peak. For both the Amazon and Congo the modeled regime shows a discharge rise in March and April which is not present in the observed discharges. For the Congo Zaitchik et al. (2010) found a similar shift in peak timing, which they assign to the complexity of the bimodal runoff regime and the poor quality of the GRDC gauge data for this river.

Regime curves
Finally, 30 year average modeled hydrological regimes are compared with observed regimes (GRDC) for all catchments, see the blue (*ERA6190*) and black (GRDC) lines in Fig. 3.7. The plots show that difference between the modeled and observed regime is especially large for the MacKenzie as mentioned before. Although for the Lena annual average modeled discharge is close to observed, the regime curve does show deviations. The observed snowmelt driven discharge peak is steeper than modeled and modeled baseflow is too high. Modeled discharge for the Murray, Niger and Zambezi is higher throughout the year. The regime curves of monsoon influenced rivers are modeled relatively well, except for the regime curve of the Indus, where the snow and glacier melt driven discharge rise is too early in spring and does not coincide with the monsoon influenced discharge peak, as in the observed regime (Immerzeel et al., 2010). For the Parana observed river flow is more constant through time, because of the presence of large reservoirs. The NS coefficients (Table 3.4) show that the modelled regime curves show large deviations, due to biases in absolute discharge quantities, for the arid catchments (Murray, Orange, Zambezi and Niger) and for the Yellow river. Furthermore, negative NS coefficients are obtained for the Parana and Congo where the shape of the modeled regime deviates from the observed regime. NS coefficients of more than 0.5 are obtained for the Brahmaputra, Ganges, Mekong and Yangtze and for the Mississippi and Lena.

3.3.2 Evaluation of the bias-correction method using the *ERACLM* dataset

The results of the *ERACLM* run are compared with the results of the *ERA6190* run and observed GRDC discharges in order to assess the influence of the bias-correction method applied to the GCM datasets.

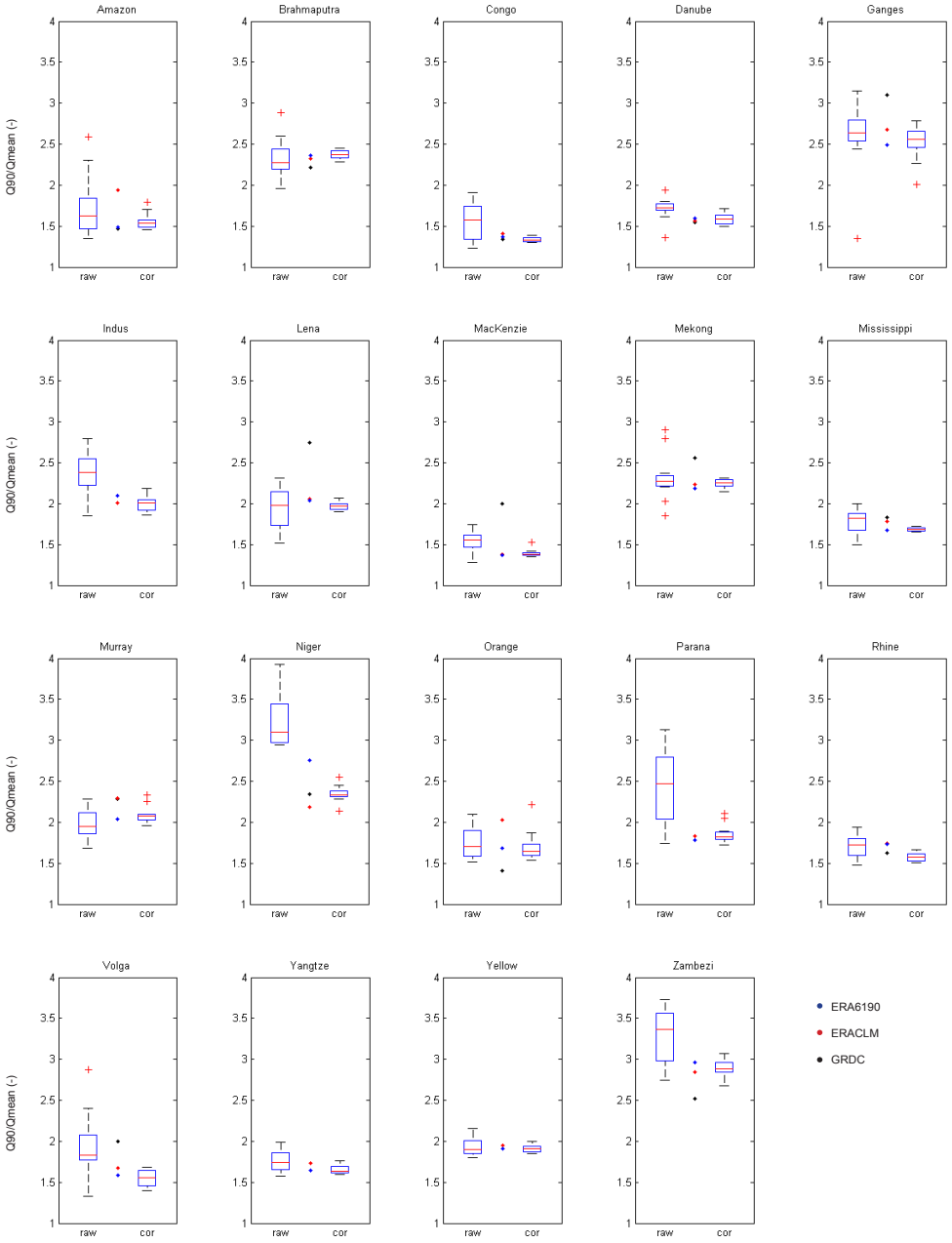


Figure 3.3 Thirty year average relative Q90 values ($Q90/\overline{Q}$ (-)) for the ERA6190 run (blue points), the ERACLM run (red points) and the GRDC observations (black points). Boxplots represent the values for the ensemble of non bias-corrected GCM runs (raw) and bias-corrected GCM runs (cor; boxes are drawn between the quartiles, whiskers extend to the most extreme data points and outliers are drawn individually as red crosses).

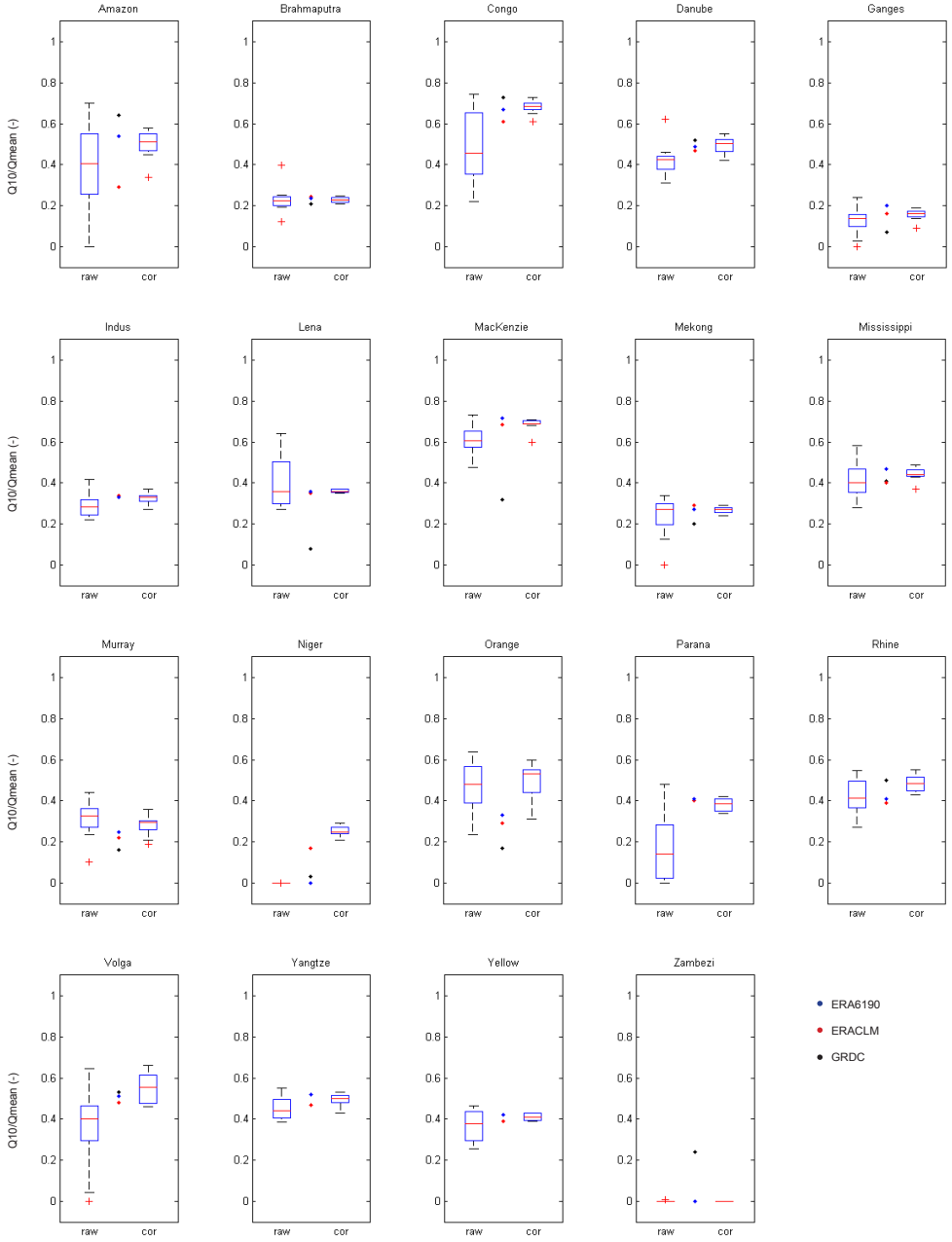


Figure 3.4 Thirty year average relative Q10 values ($Q10/\overline{Q}$ (-)) for the ERA6190 run (blue points), the ERACLM run (red points) and the GRDC observations (black points). Boxplots represent the values for the ensemble of non bias-corrected GCM runs (raw) and bias-corrected GCM runs (cor; boxes are drawn between the quartiles, whiskers extend to the most extreme data points and outliers are drawn individually as red crosses).

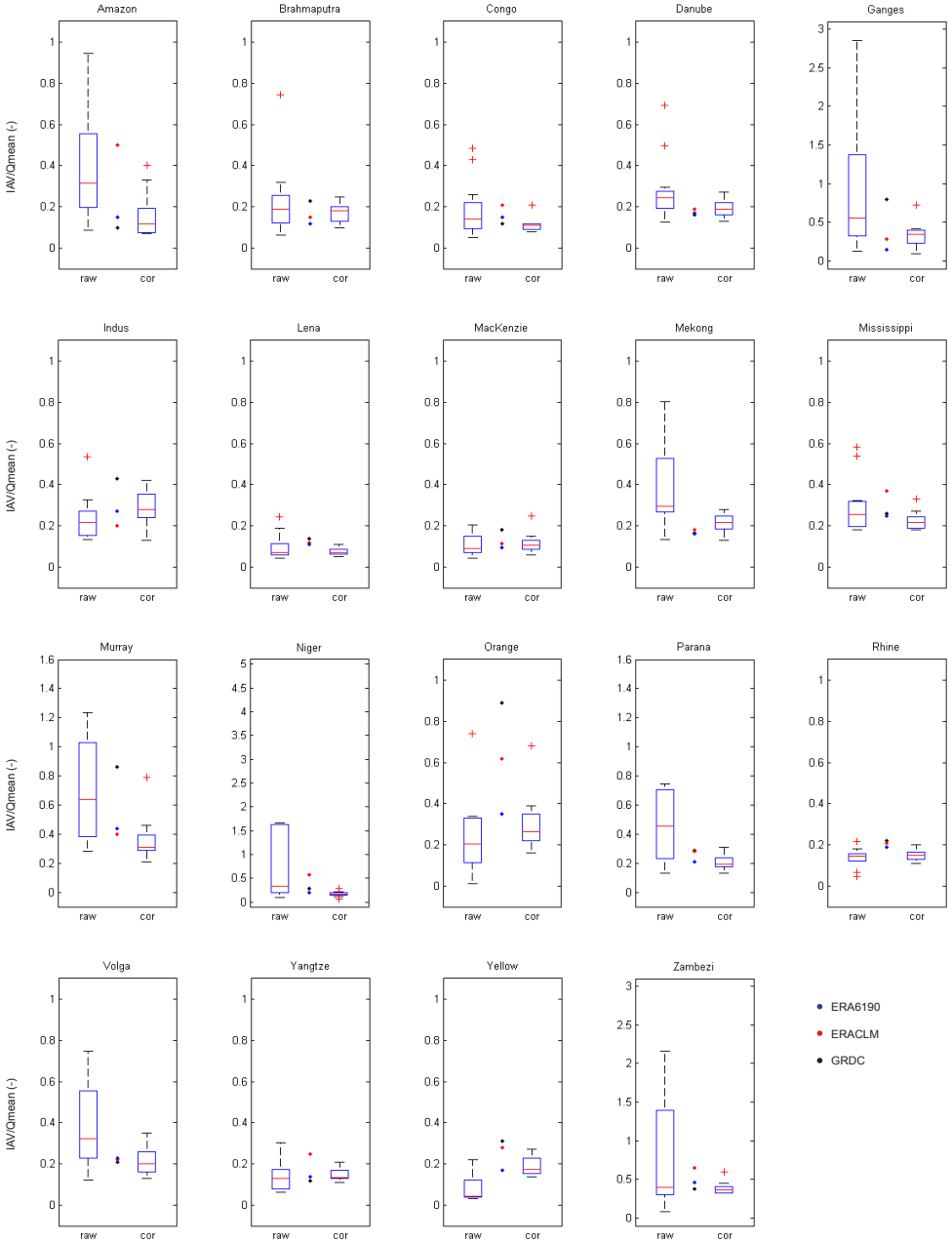


Figure 3.5 Thirty year average relative inter-annual variability values (IAV/$\overline{Q}$(-)) for the ERA6190 run (blue points), the ERACLM run (red points) and the GRDC observations (black points). Boxplots represent the values for the ensemble of non bias-corrected GCM runs (raw) and bias-corrected GCM runs (cor; boxes are drawn between the quartiles, whiskers extend to the most extreme data points and outliers are drawn individually as red crosses).

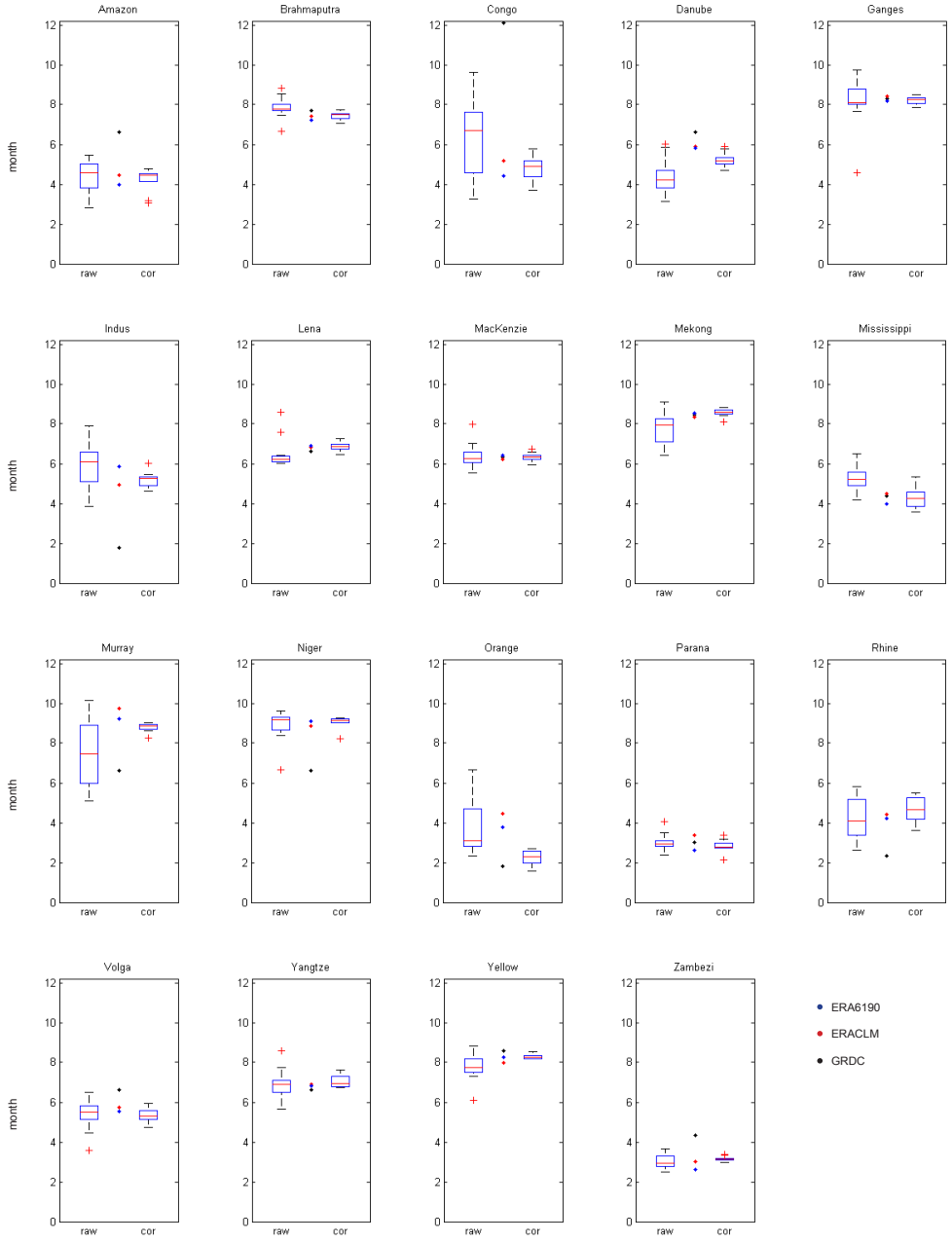


Figure 3.6 Thirty year average month of peak discharge occurence (Qpeak (month)) for the ERA6190 run (blue points), the ERACLM run (red points) and the GRDC observations (black points). Boxplots represent the values for the ensemble of non bias-corrected GCM runs (raw) and bias-corrected GCM runs (cor; boxes are drawn between the quartiles, whiskers extend to the most extreme data points and outliers are drawn individually as red crosses).

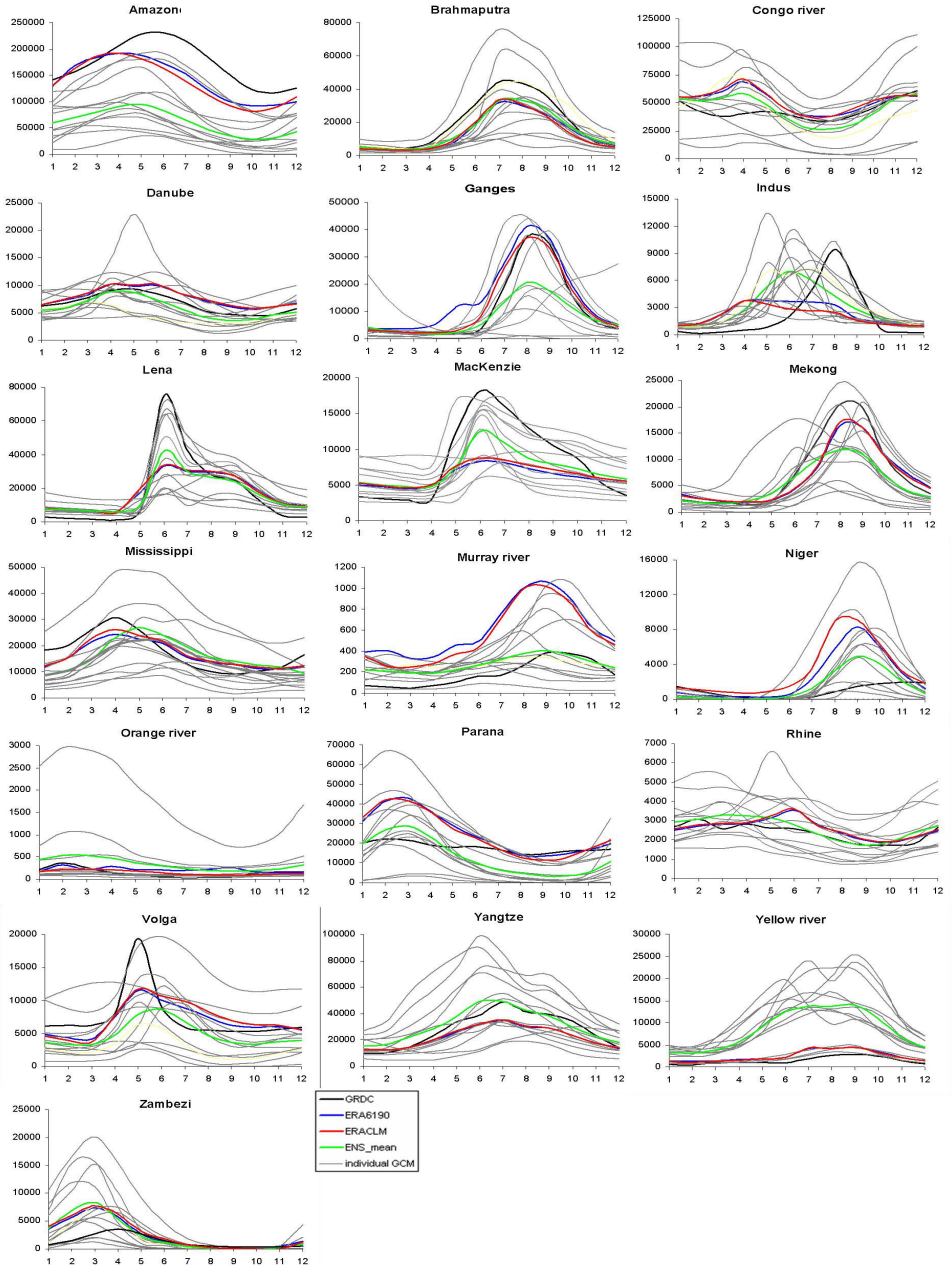


Figure 3.7 Hydrological regimes for all 19 catchments derived from discharges calculated for the period 1961-1990, with the average monthly discharge (m^3/s) on the y-axis and the month number on the x-axis. Black line is annual mean observed discharge (GRDC), grey lines are PCR-GLOBWB runs based on individual non bias-corrected GCM datasets, green line is ensemble mean result, blue line is derived from PCR-GLOBWB run forced with the ERA6190 dataset and red line is derived from PCR-GLOBWB run forced with the ERACLM dataset.

Continental discharges

The continental discharges of the *ERACLM* and the *ERA6190* run in Table 3.3 are comparable in size. Globally the difference is only 3%. The difference is especially small for Europe (0.07%). For all continents, except Europe, discharge is lower in the *ERACLM* run. The *ERACLM* run underestimates the global average modeled runoff obtained from previous studies by 4%.

Table 3.4 Nash-Sutcliffe coefficients for the regime curves (30-year average annual discharge cycles) of the ERACLM, ERA6190, ensmeble mean non bias-corrected (raw) and ensemble mean bias-corrected (cor) GCM runs relative to the GRDC regime.

NS	*raw*	*cor*	*ERA6190*	*ERACLM*
Amazon	-5.64	-0.08	0.43	0.50
Brahmaputra	-2.10	0.59	0.55	0.59
Congo river	-0.07	-2.01	-1.95	-1.48
Danube	0.87	0.33	0.38	0.50
Ganges	0.65	0.97	0.96	0.86
Indus	0.12	0.05	0.02	0.15
Lena	0.72	0.61	0.62	0.62
MacKenzie	0.58	0.27	0.30	0.21
Mekong	0.68	0.94	0.89	0.87
Mississippi	0.06	0.63	0.71	0.66
Murray	0.03	-8.34	-10.53	-13.26
Niger	-4.61	-14.98	-28.91	-16.78
Orange	-106.11	-7.86	-15.92	-17.75
Parana	-9,84	-23.61	-19.36	-19.44
Rhine	0.54	0.39	0.15	0.25
Volga	-0.05	0.50	0.43	0.50
Yangtze	0.87	0.68	0.66	0.63
Yellow river	-88.06	-0.70	-1.25	-1.27
Zambezi	-4.05	-2.17	-3.79	-2.86

Discharge statistics

Differences in results of the *ERACLM* and *ERA6190* are small, especially for mean discharge (see Fig. 3.8). Deviations of both the GCM runs and ERA-40 based runs from the observed GRDC discharge statistics are larger (Figs. 3.3–3.6). High and low-flow values are more extreme in the *ERACLM* run than in the *ERA6190* run; relative Q90 values are higher for 14 out of 19 basins and relative Q10 values are lower for 14 out of 19 basins. Differences between the runs in both relative Q10 and relative Q90 are largest for the Amazon and Niger.

Differences in relative IAV are more pronounced. For 16 out of 19 basins the *ERA6190* results have a lower IAV. The IAV of the *ERA6190* is prescribed by the CRU TS 2.1 timeseries, whereas the IAV of the *ERACLM* dataset is prescribed by the ERA-40 re-analysis dataset. The IAV of ERA-40 precipitation is known to be too high after 1972 due to gradual increasing inclusion of satellite observations in the system (Chen and Bosilovich, 2007; Hagemann et al., 2005; Troccoli and Kålberg, 2004), resulting in generally higher IAV values in the *ERACLM* than in the *ERA6190* run for 16 out of 19 basins. For 10 out of 19 basins (mainly located in monsoon and arctic regions) the

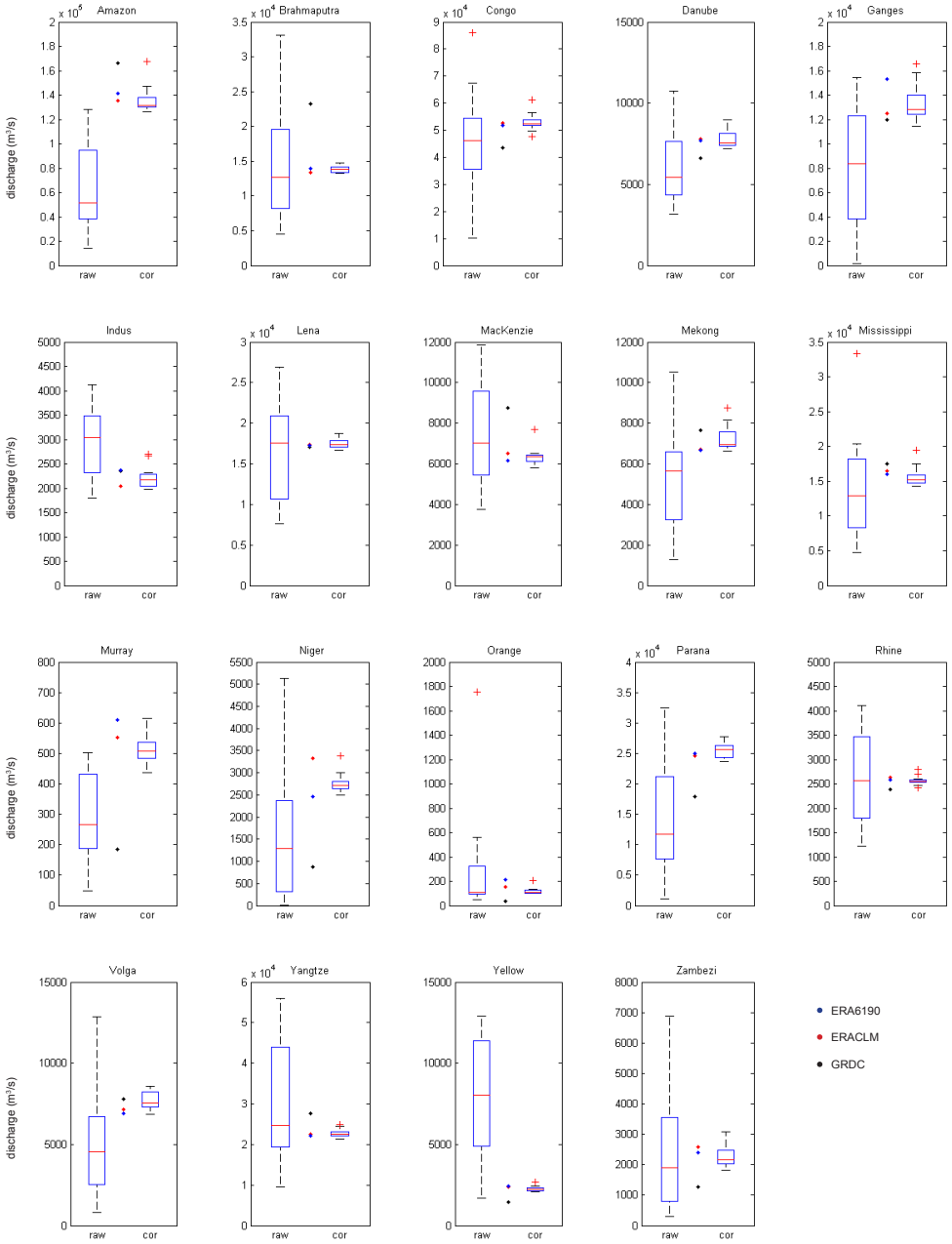


Figure 3.8 Thirty year average mean discharge (m^3/s) for the ERA6190 run (blue points), the ERACLM run (red points) and the GRDC observations (black points). Boxplots represent the values for the ensemble of non bias-corrected GCM runs (raw) and bias-corrected GCM runs (cor; boxes are drawn between the quartiles, whiskers extend to the most extreme data points and outliers are drawn individually as red crosses).

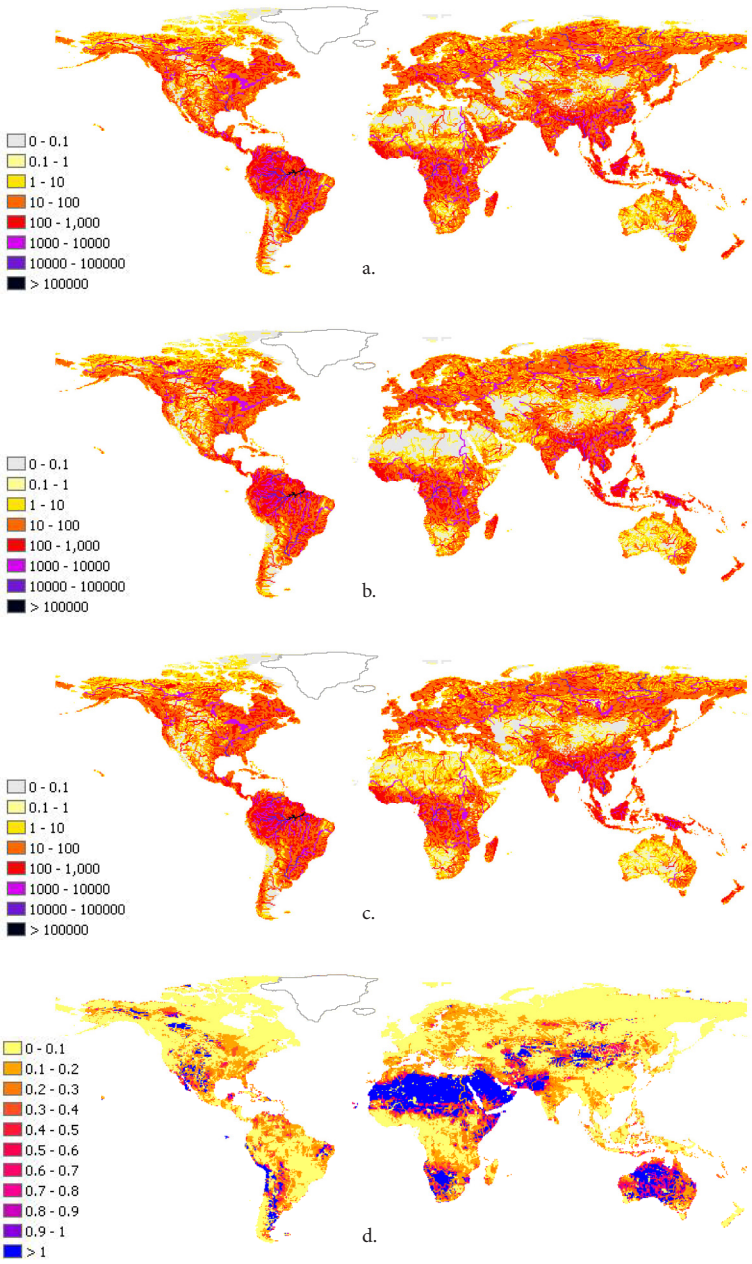


Figure 3.9 Thirty year average mean discharge (m^3/s) calculated from (a.) the PCR-GLOBWB run forced with the *ERA6190* dataset, (b.) the PCR-GLOBWB run forced with the *ERACLM* dataset, (c.) the ensemble mean discharge results of the 12 GCM based PCR-GLOBWB runs and (d.) the CV (-) of the thirty year average discharges of the ensemble of GCMS.

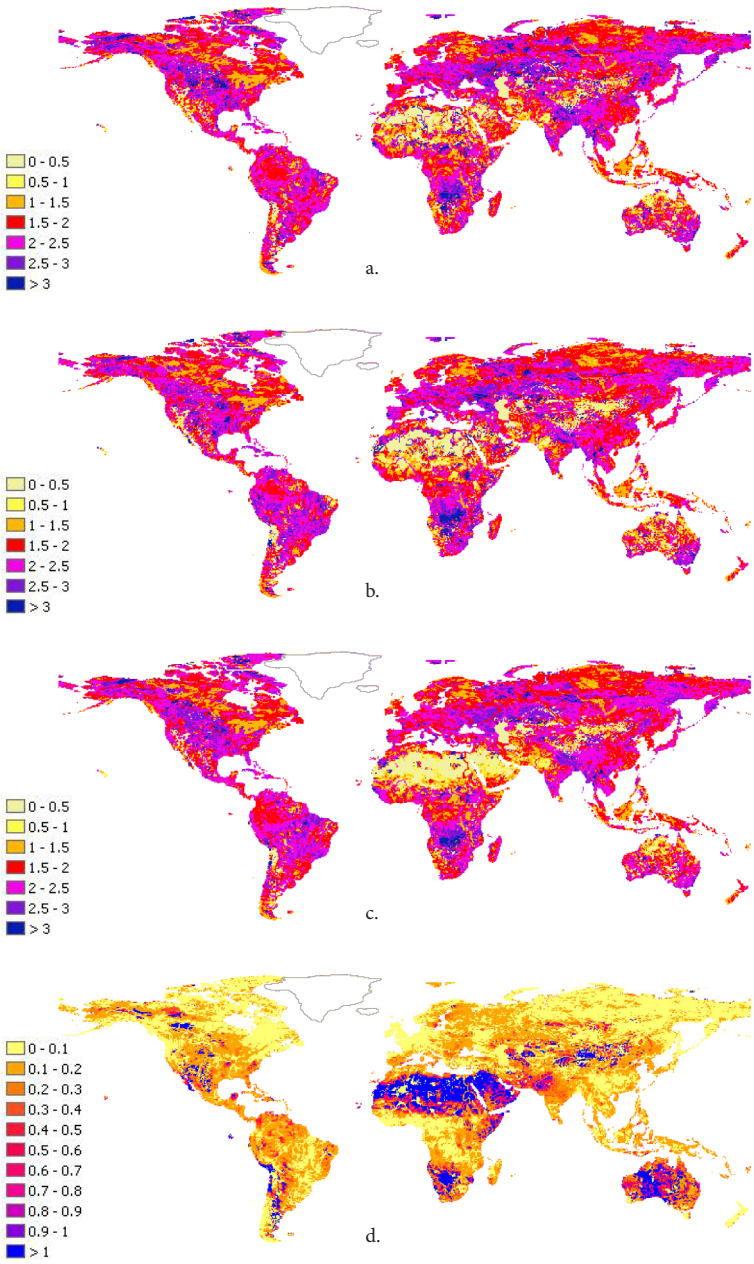


Figure 3.10 Thirty year average relative Q90 values: ($Q90/\overline{Q}$) calculated from (a.) the PCR-GLOBWB run forced with the *ERA6190* datatset, (b.) the PCR-GLOBWB run forced with the *ERACLM* dataset, (c.) the ensemble mean discharge results of the 12 GCM based PCR-GLOBWB runs and (d.) the CV (-) of the thirty year average Q90 of the ensemble of GCMs.

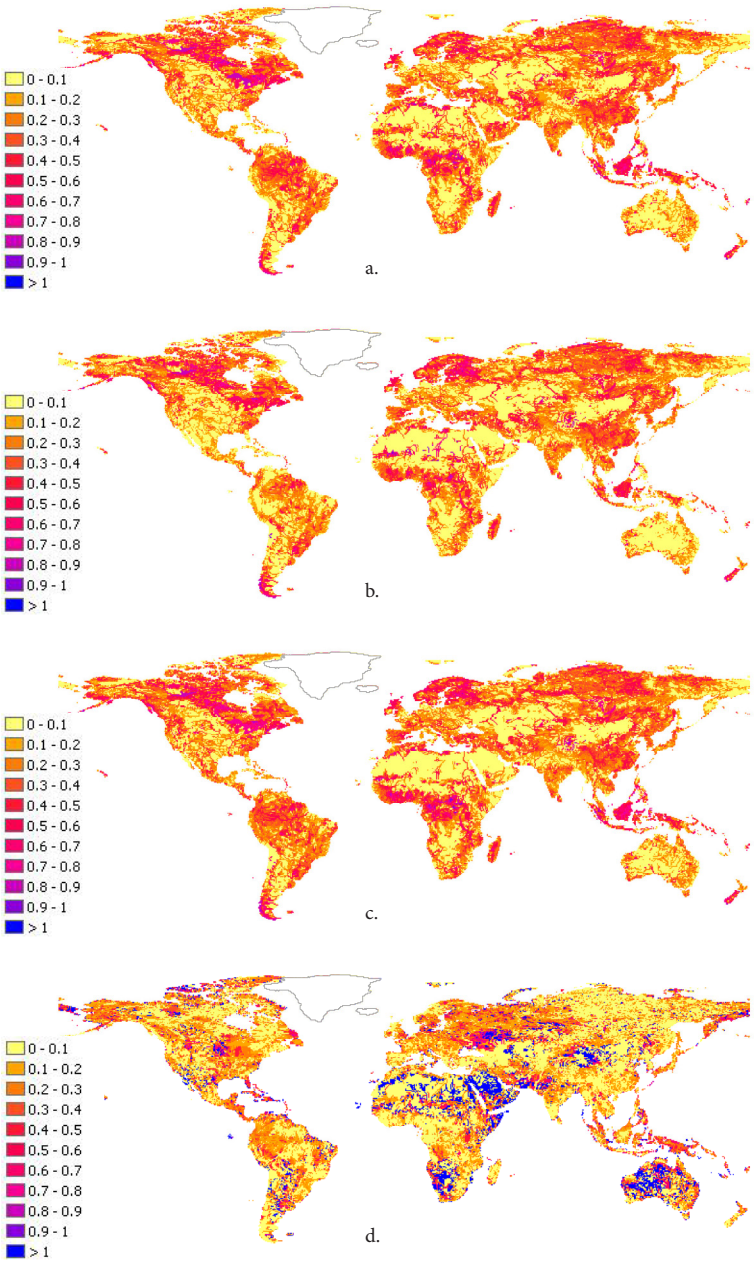


Figure 3.11 Thirty year average relative Q10 values: ($Q10/\overline{Q}$) calculated from (a.) the PCR-GLOBWB run forced with the *ERA6190* datatset, (b.) the PCR-GLOBWB run forced with the *ERACLM* dataset, (c.) the ensemble mean discharge results of the 12 GCM based PCR-GLOBWB runs and (d.) the CV (-) of the thirty year average Q10 of the ensemble of GCMs.

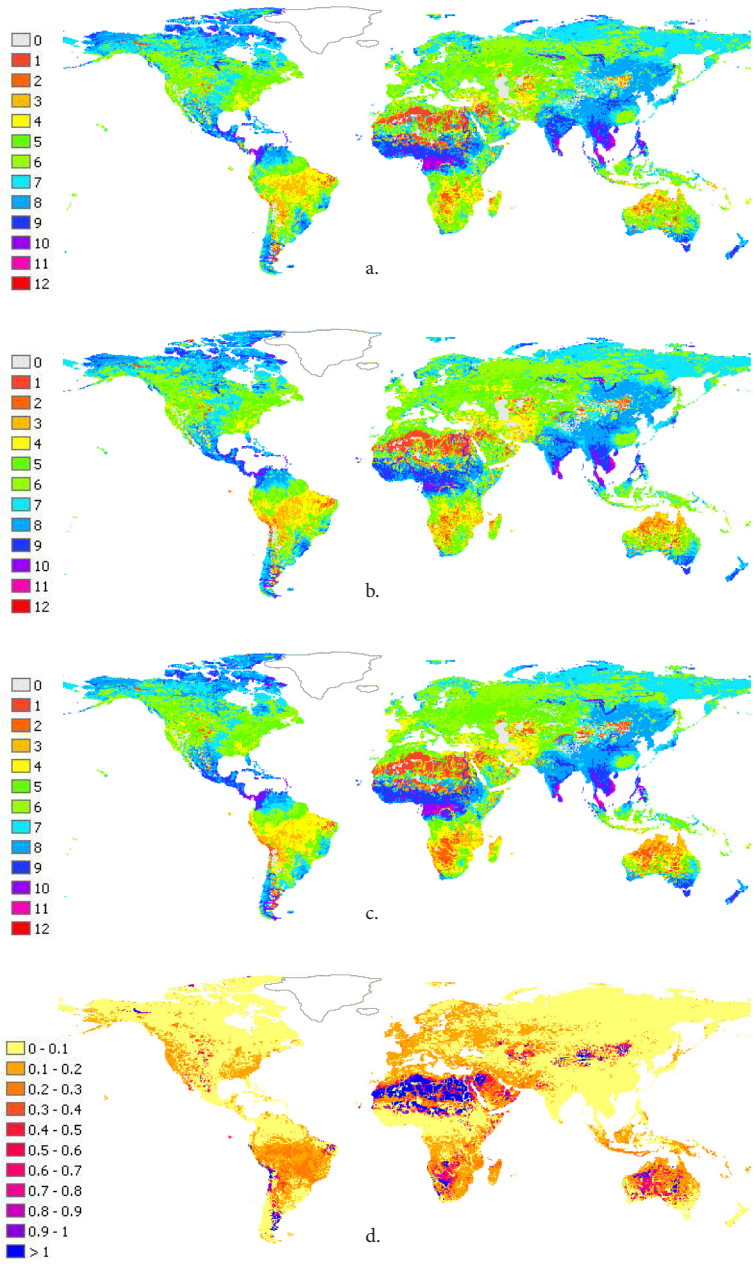


Figure 3.12 Thirty year average timing of peak discharge (month) calculated from (a.) the PCR-GLOBWB run forced with the *ERA6190* dataset, (b.) the PCR-GLOWBW run forced with the *ERACLM* dataset, (c.) the ensemble mean discharge results of the 12 GCM based PCR-GLOBWB runs and (d.) the CV (-) of the thirty year average timing of peak discharge of the ensemble of GCMs.

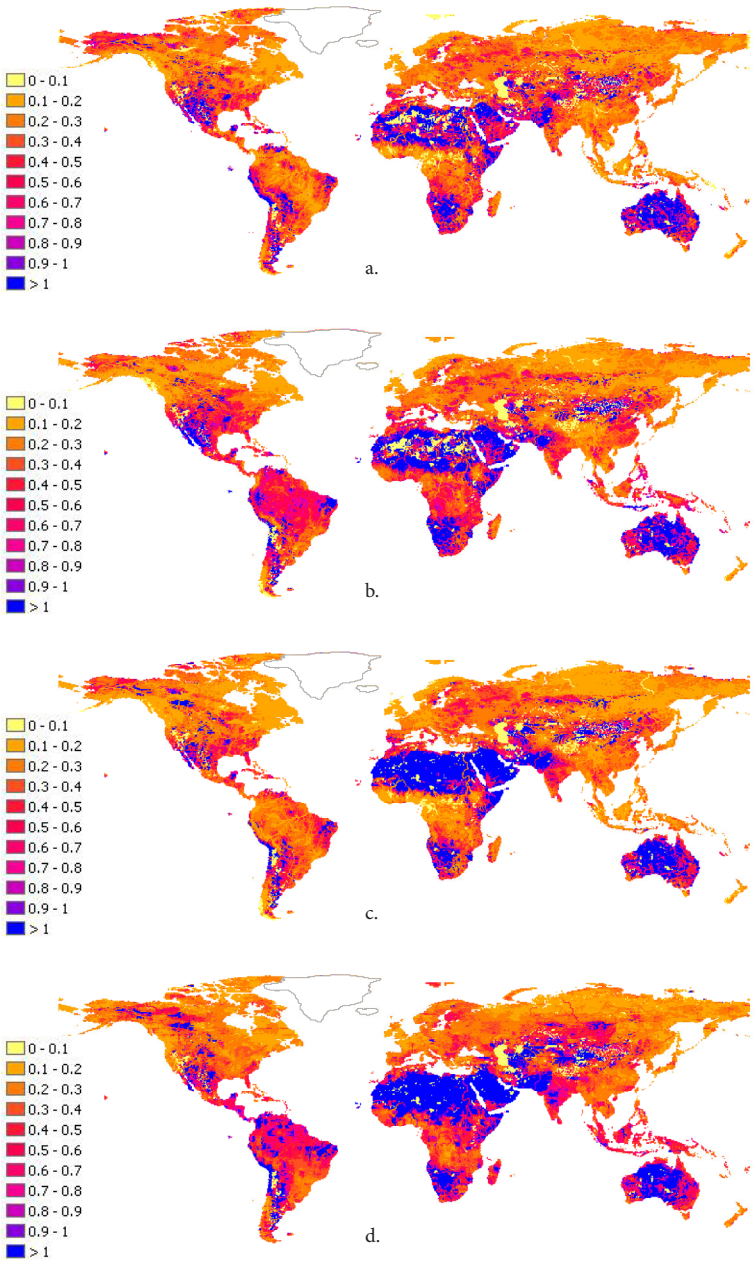


Figure 3.13 Thirty year average relative inter-annual variability: IAV/$\overline{Q}$ (-) calculated from (a.) the PCR-GLOBWB run forced with the *ERA6190* datatset, (b.) the PCR-GLOBWB run forced with the *ERACLM* dataset, (c.) the ensemble mean discharge results of the 12 GCM based PCR-GLOBWB runs and (d.) the CV (-) of the IAV of the ensemble of GCMs.

IAV of the *ERACLM* dataset is closer to the IAV of the GRDC than the *ERA6190* dataset. For these basins IAV is underestimated by both runs, yet the high ERA-40 IAV compensates for this underestimation.

Regime curves
Timing of peak discharge and regime curves are comparable for the two forcing datasets, see for comparison the blue (*ERA6190*) and red (*ERACLM*) lines in Fig. 3.7. Exceptions are (1) the Indus, where the *ERA6190* dataset results in a later timing of peak discharge that coincides with the monsoon instead of snow melt in spring, and (2) the Orange, where the average *ERA6190* regime is more constant throughout the year and the seasonality of the regime, with a low end of summer baseflow, is less pronounced.

Spatial statistics
The maps with $\overline{Q}$, relative Q90 and relative Q10 (Figs. 3.9, 3.10 and 3.11 upper two maps) show only a few differences between the *ERACLM* and *ERA6190* runs. Both the relative Q90 and $\overline{Q}$ are lower in western Australia for *ERACLM*. The south of Africa is relatively wet in the *ERA6190* run and relative Q90 values are higher in the *ERA6190* run for rivers and streams west of the Sahara. Mean discharge is slightly lower in the dry Western USA. Spatial patterns of timing of peak discharge are similar for the two runs (Fig. 3.12). Differences in IAV are also small (Fig. 3.13). Apparent are the relative high IAV values in the *ERACLM* run in Oman, Yemen and the Northern Sahel, where precipitation variability is influenced by the African Monsoon (Cook and Vizy, 2006). In dry regions the relative IAV values are the highest but of little importance as they are associated with mean discharge values that are close to zero.

3.3.3 Evaluation of discharge variability obtained with bias-corrected GCM datasets

Continental discharges
Differences in continental discharges derived from the *ERA6190*, *ERACLM* and ensemble of GCM runs are small (Table 3.3) as expected. The values of the GCM derived discharges are closer to the *ERACLM* discharges than to the *ERA6190*, globally the difference between *ERACLM* and the mean GCM discharge is less than 2%.

Discharge statistics
After bias-correction, deviations of the GCM ensemble discharge statistics from the GRDC observed discharge statistics have decreased for the majority of catchments (Fig. 3.8). The bias-correction decreased the deviation for $\overline{Q}$ for 11 out of 19 basins, for relative Q90 for 13 out of 19 basins and for the IAV for 12 out of 19 basins. However, for relative Q10 the deviation increased for 11 out of 19 basins and for Qpeak the deviation increased for 10 out of 19 basins.

The spread between the individual GCM results decreased. The CV of the ensemble of GCMs decreased for $\overline{Q}$ and Q10 for all catchments, for Qpeak and IAV for 18 out of 19 catchments and for relative Q90 for 12 out of 19 catchments. On average, after bias-correction, the CV of $\overline{Q}$ is lower than the CV of Q10 for 11 out of 19 basins and lower than the CV of relative Q90 for 14 out of 19 basins (Table 3.5). However, within the ensemble of bias-corrected GCM results, several GCMs still show deviations for $\overline{Q}$, especially for the Orange and Zambezi. The highest CV values are obtained for IAV. Low CV values and relatively small outliers of Q10, Q90 and $\overline{Q}$ are obtained for the Rhine, Lena, Yangtze and Brahmaputra.

Table 3.5 Ensemble coefficients of variation (CV; -) of the GCM results for the statistics of interest.

CV	Qmean		Q90		Q10		IAV		Qpeak	
	raw(*)	cor(*)	raw	cor	raw	cor	raw	cor	raw	cor
Amazon	0.56	0.08	0.21	0.13	0.50	0.12	0.83	0.72	0.17	0.12
Brahmaputra	0.54	0.04	0.10	0.05	0.26	0.05	0.71	0.29	0.06	0.02
Congo river	0.45	0.06	0.14	0.08	0.34	0.06	0.65	0.38	0.30	0.12
Danube	0.35	0.08	0.08	0.11	0.18	0.05	0.32	0.26	0.20	0.06
Ganges	0.59	0.11	0.16	0.10	0.51	0.11	0.88	0.56	0.15	0.2
Indus	0.24	0.10	0.11	0.08	0.19	0.06	0.46	0.37	0.17	0.07
Lena	0.34	0.03	0.12	0.04	0.30	0.04	0.55	0.24	0.11	0.03
MacKenzie	0.33	0.07	0.07	0.10	0.12	0.04	0.43	0.50	0.09	0.03
Mekong	0.51	0.08	0.12	0.08	0.39	0.06	0.68	0.21	0.11	0.02
Mississippi	0.53	0.09	0.08	0.10	0.19	0.05	1.16	0.28	0.12	0.11
Murray	0.46	0.09	0.09	0.13	0.27	0.13	0.92	0.49	0.21	0.02
Niger	0.95	0.08	0.47	0.07	3.32	0.14	0.96	0.38	0.09	0.03
Orange	1.45	0.23	0.1	0.37	0.35	0.10	1.08	0.80	0.34	0.15
Parana	0.65	0.05	0.19	0.09	0.91	0.07	0.97	0.27	0.14	0.10
Rhine	0.35	0.04	0.08	0.06	0.20	0.07	0.44	0.19	0.23	0.14
Volga	0.64	0.07	0.19	0.10	0.50	0.12	0.59	0.34	0.14	0.14
Yangtze	0.49	0.04	0.07	0.06	0.12	0.06	0.68	0.20	0.10	0.05
Yellow river	0.48	0.07	0.06	0.07	0.20	0.05	0.54	0.28	0.09	0.01
Zambezi	0.78	0.16	0.10	0.15	3.32	2.56	0.73	0.26	0.12	0.03

(*) raw refers to the statistics of the non bias-corrected GCM runs and **cor** refers to the statistics of the bias-corrected GCM runs

For the Lena and Brahmaputra the range between minimum and maximum values for IAV and timing of peak are also relatively small. This indicates the bias-correction reduced the spread between the GCMs best in these catchments. For the dry catchments; Orange and Zambezi, CV values and the range between minimum and maximum GCM results for Q90, Q10 and $\overline{Q}$ remain large. CV values for timing of peak discharge occurrence are small, especially for the Arctic rivers and the Monsoon influenced rivers Brahmaputra, Ganges, Mekong and Yellow river.

The results of the ensemble of bias-corrected GCMs are closer to the GRDC statistics than the results of the ERA-40 based runs for the majority of basins for $\overline{Q}$, Q10 and Qpeak (Figs. 3.3 – 3.6). For relative Q90 the results of the *ERACLM* run are closer to

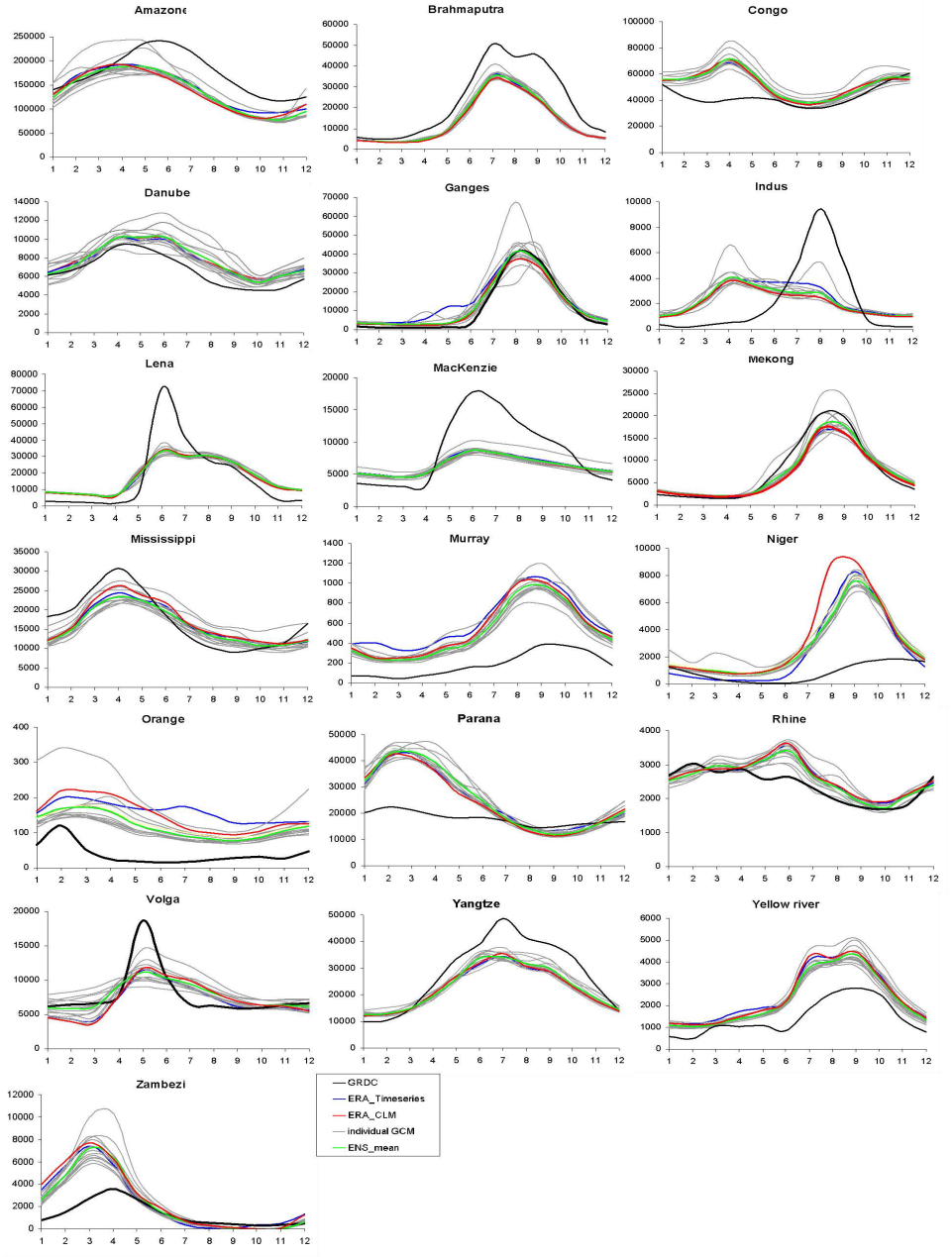


Figure 3.14 Hydrological regimes for all 19 catchments derived from discharges calculated for the period 1961-1990, with the average monthly discharge (m^3/s) on the y-axis and the month number on the x-axis. Black line is annual mean observed discharge (GRDC), grey lines are PCR-GLOBWB runs based on individual bias-corrected GCM datasets, green line is ensemble mean result, blue line is derived from PCR-GLOBWB run forced with the ERA6190 dataset and red line is derived from PCR-GLOBWB run forced with the ERACLM dataset.

the discharge observations for 11 out of 19 basins and the IAV of both ERA-40 based runs is closer to the GRDC values for more than half of the basins (*ERACLM* for 10 and *ERA6190* for 11 basins). On average the IAV is higher in discharge results of hydrological model runs based on the forcing datasets where the IAV is determined by the model generated meteorological datasets (e.g. the datasets obtained by bias-correction).

Regime curves

Overall the simple monthly bias-correction reduces the differences between the regimes derived from the GCM based runs (compare Fig. 3.7 and Fig. 3.14, where grey lines represent individual GCM runs and the green line is ensemble mean regime). After bias-correction, the resemblance of the GCM runs is large for rivers with a strong seasonal pattern, like the monsoon influenced and Arctic rivers. For other catchments that are influenced by both rainfall and snowmelt, like the Amazon, Rhine, Volga, Danube and Indus, differences in regime pattern are larger.

After bias-correction the NS coefficient of the regime curves decreased for 10 out of 19 basins; for several basins, such as the Indus, MacKenzie, Murray, Niger and Parana, the mean regime curves of the non bias-corrected GCMs is much closer to the GRDC regime than the mean regime curves of the bias-corrected GCMs. Still, the NS coefficient of the GCM ensemble mean is higher for 12 out of 19 basins than the NS coefficients derived for the regimes of both ERA-40 based runs.

For basins in the drier regions of the world, such as the Murray, Zambezi and Orange river, the GCM ensemble mean regime curve is slightly lower and closer to the GRDC regime than the regime curves of both ERA-40 runs. For the remaining basins the deviations of the GCM ensemble mean regime from the GRDC regime are comparable those of the *ERACLM* and *ERA6190* regimes.

Spatial statistics

Differences in ensemble mean spatial discharge patterns from both ERA-40 runs are small. $\overline{Q}$ is relatively high in the Sahara, resulting in lower relative Q90 values. In the Australian desert GCM mean discharge is relatively low and in the Southwestern USA the region with lower discharge values is slightly more extended. Ensemble uncertainties (CV) are large for arid African regions, the arid southwestern USA and for the desert region of Australia where the CV of precipitation of the ensemble of GCMs is large as well (Johnson and Sharma, 2009). High CV values are also obtained for mountainous regions, like the Andes ridge and parts of Asia, where the low GCM resolutions are unable to capture the high spatial heterogeneity (Viviroli et al., 2011). CV values are higher, and regions with high CV values are more extended, for relative Q90 and Q10 than for $\overline{Q}$. For all statistics the maps obtained from the ensemble of GCMs resemble the maps of the *ERACLM* run more than the map of the *ERA6190* does, except for the Northern Sahel.

3.4 Discussion

The analysis in this study was performed in three steps, i.e. to quantify the influence of both the hydrological model deficiencies and the bias-correction method before evaluating the discharge variability derived from bias-corrected GCM datasets. In this case PCR-GLOBWB was used as a means to assess the hydrological discharge variability obtained from meteorological GCM datasets, but the use of any other macro-scale hydrological model probably would have resulted only in minor differences. As assessed by the validation exercise in the first step of this study, the use of a hydrological model introduces uncertainty and bias in the discharge. This bias propagates in the results derived from the bias-corrected GCM datasets. Although the model errors found here are specific for PCR-GLOBWB, they are not unlike those found in other global model studies (e.g., Döll et al., 2003; Nijssen et al., 2001; Arnell, 1999). These errors can be reduced when the model is tuned or calibrated but this is only possible at the expense of a loss of its physical basis and the result is highly depended on data availability (Beven, 2001). Calibration would also amend for errors in the meteorological input, thus compromising the applicability of the model to different GCMs (see Chapter 2).

The performance assessment is hampered by the lack of a perfect representation of the historical meteorology. By downscaling the CRU TS 2.1 timeseries to daily values with the ERA-40 reanalysis we created a 'best guess'. However, biases present in both the CRU and ERA-40 datasets will also be included in this new dataset. Nonetheless, PCR-GLOBWB, forced with this dataset, reproduces global runoff quantities and regimes relatively well. For the European continent and the rivers Danube, Rhine and Volga mean discharges and inter-annual variability are relatively close to the GRDC observed values. Differences in total continental discharges from previous studies are small.

Globally the difference from the average of existing model studies is less than 1% (Table 3.3). Still, for the African continent as well as the Murray river, PCR-GLOBWB overestimates discharge. Regime curves are reproduced well for Monsoon influenced rivers but mean discharge and inter-annual variability are underestimated for most of the Asian monsoon rivers (Fig. 3.7).

Other than through model deficiencies, the biases between modeled and observed discharge statistics are caused by errors in 1) the observed discharges, 2) the ERA-40 dataset and 3) the CRU TS 2.1 dataset, which is known to be inaccurate for parts of Africa and suffers from undercatch in snow dominated areas (Fiedler and Döll, 2007). By using the CRU dataset as reference for the bias-correction, deviations present in the CRU set are introduced in the bias-corrected GCM datasets as well.

The influence of the bias-correction method is obtained form the difference in results between the *ERA6190* and the *ERACLM* runs. Differences between these two runs are small, overall relative IAV and Q90 values are higher in the *ERACLM* run (Fig. 3.3 and 3.5) and $\overline{Q}$ values are lower (Fig. 3.8). The IAV of ERA-40 precipitation is relatively high after 1972 when satellite observations are gradually included in the system

(Chen and Bosilovich, 2007). This high precipitation IAV compensates for the underestimation of observed IAV obtained with the *ERA6190* run.

The average annual continental runoff is lower for the *ERACLM* than for the *ERA6190* run, especially over Africa (Table 3.3). For the *ERACLM* dataset the bias-correction is not performed on a year-by-year basis, i.e. adjusting all monthly precipitation, temperature and evaporation values to the monthly observations from the CRU for the specific year. Instead, in the *ERACLM* (and the GCM) dataset the 30-year average annual mean precipitation, temperature and evaporation quantities are equalized to the 30-year average observed values of the CRU observations. The inter-annual variability for potential evaporation is relatively small compared to precipitation. Therefore, the correction of the 30-year mean potential evaporation results in similar increasing or decreasing effects for all 30 years. Yet, inter-annual variability for precipitation is larger and the multiplicative correction results in larger absolute changes for wet years. Herewith consistency between precipitation and evapotranspiration is lost over the wetter periods. Especially in wet years, with precipitation overestimations in the ERA-40 dataset, the correction results in relatively large precipitation decreases, while potential evaporation is only slightly adjusted. Here the *ERACLM* run will yield less effective precipitation with a consequent decrease in discharge.

Bias-correction decreases the deviations of GCM based model results from the observed discharges for $\overline{Q}$, relative Q90 and IAV for most basins, but increases deviations for Qpeak and relative Q10 for most basins. After bias-correction, continental discharges obtained from the ensemble of GCMs are close to the *ERACLM* results and lower than discharges modeled with the *ERA6190* dataset (Table 3.3). The results of the bias-corrected GCM runs are on average closer to the GRDC data for $\overline{Q}$, Q10 and Qpeak, while for Q90 the *ERA6190* run is closer to the GRDC values and for IAV both ERA-40 based runs are closer to the GRDC for a majority of the basins.

Application of the bias-correction method to the GCM datasets decreases the spread between the GCM ensemble members and lower CV values are obtained for the majority of basins for all discharge statistics. The largest ensemble CV values remain for drier catchments. Furthermore, bias-correction has less influence on extremes than on mean discharge; CV values are lower for $\overline{Q}$ than for Q10 and Q90 (Table 3.5). The CV of the inter-annual variability decreased for 17 out of 19 basins. Still, the CV values for IAV remain the highest of all statistics.

Bias-correction decreased the spread between the individual GCM derived regime curves. Especially, the resemblance of regimes for the monsoon influenced and Arctic rivers is large (Fig. 3.14). The NS coefficient decreased for 10 out of 19 basins after bias-correction. Yet, it can be questioned, whether this decrease is a result of better performance of the non bias-corrected datasets or a result of biases in the not corrected GCM datasets that compensate for inadequacies of the hydrological model (Fig. 3.7).

3.5 Conclusions

In order to evaluate the usability of bias-corrected GCM datasets in hydrological impact studies with a focus on changes in extremes and variability, we calculated hydrological regime curves and various discharge statistics for the period 1961 to 1990. We did this by simulating discharge with PCR-GLOBWB for a selection of 12 GCMs and comparison of the results with observed GRDC discharge data and two reference runs based on the ERA-40 and CRU TS 2.1 datasets.

As expected, the spread between the results of the different GCM based runs, decreases after bias-correction, especially for rivers with a strong seasonal pattern. Overall, the bias-correction decreased the deviation between GCM ensemble mean and GRDC observed statistics for $\overline{Q}$, Q90 and IAV. However, the bias-corrected datasets result in slightly lower mean annual discharge amounts and for arid and mountainous basins the method performs less well.

Observed discharge statistics (Q10, Q90 and $\overline{Q}$) are better approximated by the bias-corrected GCM datasets than inter-annual variability (IAV) and agreement amongst models is higher for mean discharge and timing than for IAV. This suggests that GCMs are mostly useful in global hydrological impact studies in which persistence is of less consequence (e.g. in case of flood studies rather than those on droughts).

Furthermore, the bias-correction influences mean discharges more than discharge extremes. This has the positive consequence that changes in daily rainfall distribution and subsequent changes in discharge extremes will also be preserved when the bias-correction method is applied to future GCM datasets. However it also shows that agreement between GCMs remains small for discharge extremes Finally, it can be concluded that, because of the large deviations between observed and simulated discharge in which both errors in climate forcing, model structure and to a lesser extent observations are accumulated, it is advisable not to work with absolute discharge values for the derivation of future discharge projections. Rather relative changes should be calculated by dividing the absolute change by the discharge calculated for the control experiment.

Acknowledgements

We acknowledge the GCM modeling groups, the Program for Climate Model Diagnosis and Intercomparison (PCMDI) and the WCRP's Working Group on Coupled Modelling (WGCM) for their roles in making available the WCRP CMIP3 multi-model dataset. Support of this dataset is provided by the Office of Science, US Department of Energy. We also want to acknowledge the Global Runoff Data Centre for providing the global discharge timeseries. We thank the two anonymous reviewers for their very useful comments.

4 On the suitability of GCM runoff fields for river discharge modeling; a case study using model output from HadGEM2 and ECHAM5

This chapter is based on:
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, Journal of Hydrometeorology, in press.

Abstract

The representation of hydrological processes in land surface schemes (LSSs) has recently been improved. In this study the usability of GCM runoff for river discharge modeling is evaluated by validating the mean, timing and amplitude of the modeled annual discharge cycles against observations.

River discharge was calculated for 6 large rivers, using runoff, precipitation and actual evaporation from the GCMs ECHAM5 and HadGEM2. Four methods were applied: 1) accumulation of GCM runoff, 2) routing of GCM runoff, 3) routing of GCM runoff combined with temporal storage of subsurface runoff and 4) off-line hydrological modelling with the global model PCR-GLOBWB using meteorological data from the GCMs as forcing.

The quality of discharge generated by all four methods is highly influenced by the quality of the GCM data. In small catchments, the methods which include runoff routing perform equally well, although off-line modelling with PRC-GLOBWB outperforms the other methods for ECHAM5 data. For larger catchments, routing introduces realistic travel times, decreased day-to-day variability and it reduces extremes. Complexity of the LSS of both GCMs is comparable to the complexity of the hydrological model. However, in HadGEM2 the absence of sub-grid variability for saturated hydraulic conductivity results in a large subsurface runoff flux and a low seasonal variability in the annual discharge cycle. The analysis of these two GCMs shows that when LSSs are

tuned to reproduce realistic water partitioning at the grid scale and a routing scheme is also included, discharge variability and change derived from GCM runoff could be as useful as changes derived from runoff obtained from off-line simulations using large-scale hydrological models.

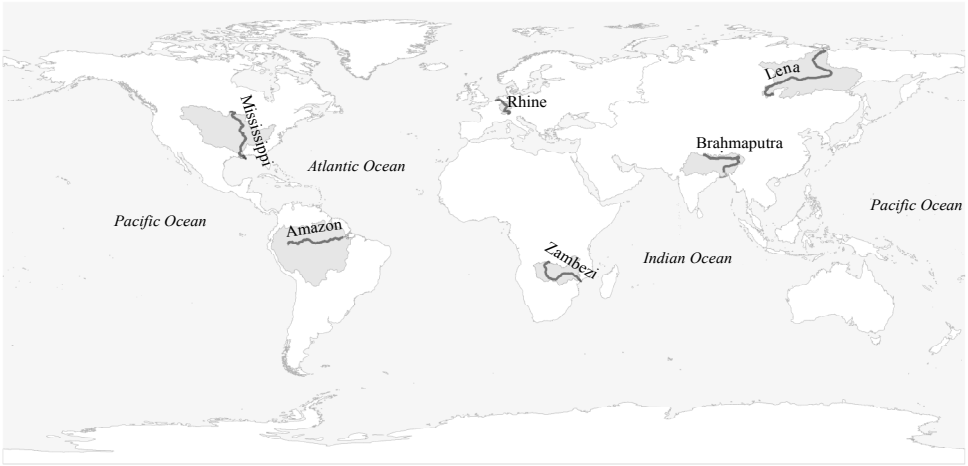
4.1 Introduction

Transport of water through rivers to oceans was previously often neglected in general circulation models (GCMs; Miller et al., 1993). In the last decade it has been recognized that land surface hydrology and river flow play an important role in the global climate system. For instance, the fresh water influx to oceans changes their salinity and consequently may affect ocean circulation and convection (Arora, 2001). Furthermore, the hydrological cycle influences feedback mechanisms between land surface and atmosphere (Kite, 1998). Full inclusion of these feedback mechanisms requires the presence of land surface schemes (LSSs). LSSs regulate the partitioning of precipitation into evaporation, storage and runoff and hereby determine the division of radiant energy in latent and sensible heat fluxes (Van den Hurk et al., 2004; Pappenberger et al., 2009). In addition to near-surface temperature, runoff is one of the few outputs of a LSS that can be validated with observed data. Therefore, the performance of LSSs, and of GCMs containing an LSS, is frequently evaluated by translating the grid cell runoff into river discharge and comparing the obtained discharges with observations (Graham and Jacob, 2000; Arora, 2001; Pappenberger et al., 2009). By now, resolutions of LSSs approach the resolutions of macro-scale hydrological models and the inclusion of hydrological processes can be as complete as in hydrological models (Clark and Gedney, 2008; Hagemann and Gates, 2003).

Therefore, in addition to using it as means of validation of LSSs, discharge obtained by routing GCM runoff may as well have a hydrological value. Moreoever, in large scale hydrological impact studies, GCM runoff may even be useful as a replacement for discharge obtained from hydrological modeling, as is currently common practice (Sperna Weiland et al., 2010; Alcamo et al., 2007; Fowler and Kilsby, 2007). When GCM runoff is used, a closed hydrological cycle is guaranteed and inconsistencies such as the double inclusion of hydrological processes in the modelling chain (for example storage of water in soil, snow and canopy and calculation of actual evaporation fluxes) can be avoided. At the same time computation times are reduced, which enables the analysis of larger ensembles of GCMs. Although the main purpose of a GCM is not the reproduction of realistic runoff fluxes, GCM runoff fields have been used directly in hydrological change studies before (e.g. Manabe et al., 2004). Other studies routed GCM runoff fields to obtain river discharge and to calculate future discharge changes (Arora and Boer, 2001; Milly et al., 2005; Nohara et al., 2006).

In this study we investigate the hydrological value of GCM runoff. This hydrological value is defined as the suitability of GCM-generated runoff fields for river discharge

modeling in hydrological impact studies. To this end, observed annual mean discharge, discharge extremes and timing and amplitude of the annual discharge cycle should be correctly reproduced using GCM runoff. In this study a first step of this assessment is made by comparing discharges derived from the runoff fields of two GCMs (ECHAM5 and HadGEM2) with observations. In this comparison we distinguish between correct implementation of local runoff generation mechanisms and the effect of the inclusion of a river routing scheme. To this end, discharge is generated using the following four methods: 1) daily accumulation of GCM runoff, 2) routing of GCM runoff, 3) routing of runoff in combination with attenuation of subsurface runoff in a groundwater reservoir and 4) off-line hydrological modelling with meteorological data from the two GCMs. In the analysis of the results, step-by-step comparisons of the different methods are made. The comparison of daily accumulated runoff with routed runoff illustrates the effect of runoff routing. The effect of additional storage in a groundwater reservoir is illustrated by comparison of direct routing of the complete runoff flux with the inclusion of a delay by temporal storage of the subsurface runoff before routing. The offline hydrological model is used to illustrate the influence of different land surface schemes.



Catchment	*Area (km²)*	*Qavg (m³/s)*	*Gauge*	*Catchment*	*Area (km²)*	*Qavg (m³/s)*	*Gauge*
Amazon	6.915.000	190.000	Obidos	Mississippi	2.981.076	12.743	Vicksburg
Brahmaputra	930.000	48.160	Bahadurabad	Rhine	65.683	2.200	Rees
Lena	2.500.000	17.000	Kusur	Zambezi	1.390.000	3.400	Katom a Mulilo

Figure 4.1 Selected catchments with total catchment area, long term average observed discharge and gauges used in analysis.

4.2 Methods and data

4.2.1 Hydrological data

This study focuses on six large river basins; the Amazon, Brahmaputra, Lena, Mississippi, Rhine and Zambezi. The selected rivers cover multiple continents, a wide range of climate conditions and they differ in total catchment area and degree of regulation. Fig. 4.1 shows the geographical location of the rivers together with some basin characteristics. Below follows a short description of the basins.

The Amazon originates in the Andes Mountains and flows through Brazil to its mouth at the Atlantic Ocean. Along its course, of about 6400 km, around 1000 tributaries join the river. The Amazon is at its narrowest (~ 1.5 km) at the Obidos discharge station where the river depth is 60 meters. Here the annual cycle has a typical high flow period from November to June. The river course is almost completely natural with limited water regulation and use.

The Lena is located in Russia and originates in the mountains near lake Baikal. It has a length of approximately 4500 km and ends with a large delta in the Arctic sea. Winter temperatures can be as low as -70°C and precipitation amounts are small (200 – 400 mm). The river is free from ice for 5 months a year and has a typical cycle, limited to zero flow (caused by freezing) during the winter and high flow during the summer. As for the Amazon, the river course is almost completely natural with limited water regulation and use.

The Rhine originates in the Swiss Alps. The river flows through Germany and drains through the Netherlands into the North Sea. It has a length of 1230 km. Many (Alpine) tributaries join the river in Switzerland and Germany. The annual cycle is characterized by a low winter discharge, a high discharge in spring originating from snowmelt and again a low end of summer discharge. Timing of the discharge cycle is influenced by both snow melt and rainfall. The river is highly regulated and embanked.

The Zambezi rises in northwestern Zambia and flows through Angola, Botswana and Mozambique to the Indian Ocean. The basin is partly located in the tropics. In the lower part, located in the valley, temperatures (~40°C) and consequently evapotranspiration, can become very high. Annual average basin precipitation is around 600 mm. Discharge is at its maximum in March-April and diminishes in October-November to less than 10 percent of the maximum flow. River (in)flow is regulated by several dams and in the populated areas water use influences river flow in the dry season.

The Mississippi rises at Lake Itasca, Minnesota. From there it flows south where it is joined by its major tributaries, the Missouri and Ohio rivers. The river ends at the Gulf of Mexico after a course of 3780 km. Temperature and precipitation amounts vary widely over the basin. In subarctic northern Minnesota water is stored as snow in winter (temperatures of -12°C). Annual rainfall amounts vary spatially from 750 - 1800 mm. The highest monthly discharge is observed in March and mainly originates from the Ohio river. The river is highly regulated with many levees, embankments and

Table 4.1 GCM parameters used.

Parameter	*Short name*	*Unit*
Precipitation	PR	$kg/m^2/s$
Actual evaporation	EVSPSBL	$kg/m^2/s$
Subsurface runoff	MRROSS	$kg/m^2/s$
Surface runoff	MRROS	$kg/m^2/s$
Total runoff	MRRO	$kg/m^2/s$
Temperature 2m	TAS	Kelvin

reservoirs for flood control and temporal water storage.
The Brahmaputra originates in the glaciated areas of the Himalaya in Tibet (China). The river has a length of 2900 km and traverses China, India, Bhutan and Bangladesh. In Bangladesh it merges with the Ganges and the two rivers form a delta. The climate of the basin is monsoon driven and has a wet season from June to September which accounts for 60–70% of the annual rainfall. The river course is still very natural and changes continuously over time, especially in the lower Brahmaputra delta (Britannica, 2011; Immerzeel et al., 2010).
Daily discharge time-series for measurement stations close to the basin outlets have been obtained from the Global Runoff Data Center (GRDC, 2007). The observed discharge data are used to validate modeled discharges derived with the methods described in 4.2.5.

4.2.2 GCM data

Daily global time-series of the GCMs ECHAM5 (Roeckner, 2006) and HadGEM2 (Johns, 2009) are obtained from the World Data Centre for Climate for the current climate experiment 20CM3 (period 1961 to 1990). The study is restricted to the GCMs ECHAM5 and HadGEM2 due to limited online availability of GCM runoff fields. Table 4.1 shows the GCM parameters used in this study (all parameters are cell-specific values).

4.2.3 Reference meteorological data

For historical comparison two reference precipitation datasets have been included. Two datasets were selected because the quality of precipitation data is always hampered by measurement errors and no true dataset exists (Fiedler and Döll, 2007).

GPCP
The GPCP precipitation dataset (Beck et al., 2004) is a global monthly gridded precipitation dataset for the period 1951 to 2000 with a resolution of 1.0 degrees. The gridded precipitation fields are derived from the Global Precipitation Climatology Centre precipitation database which contains validated station data. This precipitation dataset has only been used for the validation of the GCM precipitation since other variables, required to perform a hydrological model run, are not provided.

CRU monthly time-series downscaled to daily time-scale using ERA-40 re-analysis
In addition to precipitation, the second meteorological dataset contains all required variables (e.g. precipitation, temperature and reference crop potential evaporation) to perform a reference hydrological model run. This dataset is created by downscaling the CRU TS2.1 timeseries (New et al., 2000) to daily values with the ERA-40 re-analysis dataset (Uppala et al., 2005; see also: Van Beek, 2008; Sperna Weiland et al., 2010). For the calculation of potential evaporation, the required radiation and wind speed were not available as monthly time-series, therefore data from the CRU CLM 1.0 dataset have been used (New et al.,1999).

4.2.4 Land surface schemes of selected GCMs

ECHAM5
In this study ECHAM5 version 2.02a with a resolution of T63L31 (approximately 1.5 degrees) is used. ECHAM5's land surface scheme consists of four water reservoirs: snow intercepted at the canopy, snow stored at the surface, rain water intercepted by the canopy and a soil water compartment (Roeckner et al., 2003, Hagemann et al., 2006). Evapotranspiration is limited by water availability and stomatal resistance. Storage of rain and snow in the canopy is limited by the size of the interception reservoir. Snow leaves the canopy through slipping (wind dependent) and melting (temperature dependent). All water passing through these storage reservoirs, becomes available to the soil water reservoir. The soil water reservoir is a single bucket with sub-grid variability for the calculation of the maximum storage capacity. This storage capacity decreases according to a probability density function that defines, based on the water content, which fraction of the cell is saturated. All rainfall falling on this saturated fraction is transformed into runoff (Dumenil and Todini, 1992). Drainage (e.g. subsurface runoff) from the soil water reservoir depends on the percentage of the maximum storage capacity occupied. Water is released through a slow drainage process when storage is between 5% and 90% of the maximum storage capacity and through a fast drainage process when storage is between 90% and 100% of the maximum storage capacity.

HadGEM2
The LSS resolution of HadGEM2 is approximately 1.25 degrees latitude by 1.875 degrees longitude. The landsurface scheme is similar to MOSES1 (Cox et al., 1999; Clark and Gedney, 2008). It contains six water storage components: snow mass, canopy water and four soil layers of which thicknesses increase with depth. Evaporation is calculated using an extended Penman-Monteith equation including conductive heat transport through the soil. The evaporation flux consists of evaporation from the canopy store, transpiration from vegetation, bare soil evaporation and sublimation of snow. Snow melt and throughfall infiltrate into the soil and infiltration excess runoff occurs when the infiltration rate exceeds the saturated hydraulic conductivity. According to Clark and

Gedney (2008) this surface runoff flux is rare on the model's gridscale. The saturated hydraulic conductivity is uniform over a grid-cell and is most of the times not exceeded by the infiltration rate.
Vertical moisture fluxes between the soil layers are modeled using an approximation of the Richard's equation, where the soil hydraulic characteristics follow Clapp and Hornberger (1978). Drainage from the bottom soil layer forms the subsurface runoff flux. Drainage rates equal the constant conductivity of this bottom soil layer and the boundary condition is given by free drainage. The complexity of the hydrological schematization is comparable to that of distributed hydrological models. In addition to the hydrological part, the land surface scheme contains a vegetation module and a module that calculates soil thermodynamics. Both modules interact with the hydrological module through evapotranspiration rate, soil hydraulic conductivity and the freezing of soil water.

4.2.5 River discharge generation

We consider the following methods for the generation of river discharge from GCM-output.

Runoff accumulation – ACCU
In this method the cell-specific CM runoff flux (MRRO) is accumulated at a daily time-step. All water that becomes available in a catchment is, within the same time-step, transported along a drainage network to the catchment outlet. This results in daily discharge time-series for all individual grid cells. The cell-specific daily discharge amount is the sum of the runoff that becomes available in the cells' upstream catchment area in the specific time-step. Here the upstream area consists of cells upstream of the cell in question, which are connected to the same river branch of the drainage network. The drainage network is based on DDM30 and has a vertical resolution of 0.5 degrees (Döll and Lehner, 2002).

Routing of total runoff - MRRO
Within this method cell-specific runoff (MRRO) is accumulated for each individual time-step and, as an extension of the above method, routed along a drainage network using the kinematic wave equation. The drainage network is based on DDM30 and has a vertical resolution of 0.5 degrees (Döll and Lehner, 2002). It is extended with lakes, wetlands and reservoirs for which storage volumes have been obtained from the GLWD dataset (Lehner and Döll, 2004). Q-h relationships are used to calculate the discharge from reservoirs and lakes. No reservoir scheme has been implemented in this version of the model. Flow velocities are calculated using the kinematic wave equation (Chow et al., 1988) implemented in the PCRaster framework (Wesseling et al., 1996). A disadvantage of the kinematic wave equation is its simplicity. Complex dynamic processes, such as backwater effects, are not included. Furthermore, the method assumes that the

slope of the water surface is parallel to the slope of the bottom of the water course. This criterion is not always fulfilled in reality. Yet, important advantages of the application of the kinematic wave equation in this study are its speed, which enables running the model on a global scale, and its limited input data requirements, which is very useful for data-sparse regions of the world.
Sensitivity of the routing parameters has not been fully tested. Based on literature the value of the manning coefficient for the floodplains is set to 0.1 and the manning coefficient for channels is set to 0.04 (Chow et al., 1988). The length of a river segment is equal to either the length of the diagonal or the length of the grid cell, depending on the flow direction. To correct for the difference between modeled and real world river length a global average tortuosity multiplication coefficient of 1.3 has been implemented.
Initial conditions were obtained by running the model multiple times with a climatology calculated from the period 1961 to 1990. This climatology was derived from the CRU dataset which had been downscaled to daily values with the EA-40 re-analysis dataset. Steady state conditions were reached at the moment the global annual average catchment outflow differed less than 0.5% between two consecutive annual climatological runs.

Routing and attenuation of subsurface runoff - MRROSS
In this method the GCM subsurface runoff (MRROSS) is temporarily stored in a linear groundwater reservoir. Groundwater release from the reservoir is regulated with cell-specific reservoir constants which have been derived from a global lithology map (Dürr et al., 2005) following Van Beek and Bierkens (2009). After release, the groundwater flux is added to the GCM surface runoff flux (MRROS) and the combined flux is routed along the drainage network as described for the MRRO method. Initial groundwater reservoir storages, which is required for this method, has been obtained by back-to-back simulations with ten years of GCM subsurface and surface runoff. These runs were repeated until dynamic steady state conditions achieved for the soil and groundwater stores (e.g. less than 1 percent difference in catchment and reservoir outflow for consecutive ten year runs).

Hydrological modelling with PCR-GLOBWB - EVSPSBL
In the final method the global distributed hydrological model PCR-GLOBWB (Bierkens and Van Beek, 2009; Van Beek and Bierkens, 2009) has been used to derive river discharge from the actual evaporation, precipitation and temperature calculated by the GCMs. The model has a resolution of 0.5 degrees. Each cell consists of a canopy layer, a snow layer, two vertical soil layers and one underlying groundwater reservoir (i.e. the same reservoir as described in the MRROSS method). Sub-grid parameterization is used to represent surface water, short and tall vegetation and to calculate saturated areas for surface runoff as well as interflow. The amount of water entering the cell is prescribed

by the GCM precipitation flux and can be stored as canopy interception or snow. Snow accumulates at temperatures below 0°C. Snow melts at temperatures above 0°C. Melt water and throughfall are passed to the surface, where they either infiltrate into the soil or become surface runoff (Dumenil and Todini, 1992). Exchange of soil water is possible between the soil layers and groundwater reservoir in both up- and downward directions, depending on soil moisture status and groundwater storage. The GCM actual evaporation flux is subtracted from open water, bare soil and vegetation compartments depending on vegetated fractions of cells and crop factor values. For the cell fractions without open water, we apply the following rules: if the prescribed GCM actual evaporation flux exceeds the amount of water available on the canopy and in the bare soil, the remainder is subtracted from the first and second soil layer through transpiration. If the GCM actual evapotranspiration amount exceeds the actual water availability in the cell, the actual evapotranspiration is reduced to the total water availability.
Runoff consists of non-infiltrating melt water, saturation excess surface runoff, interflow and base flow. The model is run at a daily time-step. For each time-step the water balance is calculated for the individual cells and runoff is accumulated and routed along the drainage network as described for the MRRO method. An extensive description of the hydrological model can be found in Van Beek and Bierkens (2009).

Reference hydrological model run – CRU
As an additional reference for validation of the GCM derived annual discharge cycles, PCR-GLOBWB has been run with the combined CRU ERA-40 dataset (Sperna Weiland et al., 2010). In this reference run the model is forced with temperature, precipitation and potential evaporation calculated from the CRU data with the Penman-Monteith equation (Van Beek, 2008). Actual evapotranspiration is derived from potential evaporation and moisture conditions within the hydrological model.

4.3. Results

4.3.1 Effect of biases in different components of the hydrological modeling chain

Bias in precipitation
Figure 4.2 shows bar charts with the percentage deviation of ECHAM5, HadGEM2 and CRU mean annual basin average precipitation (P) from observed annual mean GPCP P. Figure 4.3 displays Taylor diagrams (Taylor, 2001) for comparison of the mean annual cycles of ECHAM5, HadGEM2 and CRU P with the observed mean annual cycle of GPCP P. The Taylor plots show, for all rivers except the Zambezi, that CRU P best resembles GPCP P. This indicates that both measurement-based historical datasets are comparable. And, since the GPCP dataset only consists of precipitation, the CRU dataset can best be used as reference forcing for the hydrological model.

HadGEM2 P overestimates observed GPCP P for all selected catchments and ECHAM5 overestimates GPCP P for all catchments except the Amazon and Zambezi. For the Amazon, both GCMs overestimate the amplitude of the annual cycle, but their correlation with observed P is above 0.9. For the Brahmaputra ECHAM5 shows a high correlation (>0.95) with observed P, yet this GCM overestimates the amplitude. HadGEM2 has a lower correlation (<0.8) and a comparable RMSD. The Taylor plot of the Lena is comparable to the Taylor plot of the Brahmaputra. For the Mississippi the correlation between HadGEM2 and observed P is low (≈0.4) and the RMSD is high. ECHAM5 performs somewhat better with lower RMSD values and higher correlation (>0.8). For the Rhine, correlation with observed P is low for both GCMs (<0.3). In addition, both GCMs overestimate the amplitude of the annual cycle. For the Zambezi correlation with observed P is high (>0.9) for both GCMs. Moreover, both GCMs have a smaller RMSD from GPCP P than the CRU P.

The effect of the precipitation bias on simulated discharge

Figure 4.4 shows bar charts with the annual average percentage bias of ECHAM5, HadGEM2 and CRU derived discharge (Q) from observed GRDC Q. Figure 4.5 displays Taylor plots for the annual average Q cycle obtained from GCM data with the different discharge generation methods. Here, observed GRDC Q has been used as a reference. In this section the effect of the P biases on discharge modeled with PCR-GLOBWB is estimated for the two GCMs (i.e. annual average PCR-GLOBWB modeled basin Q from HadGEM2 (Hp) and ECHAM5 (Ep) data). Under- and overestimations of ECHAM5 P are mirrored in modeled Q (plus and minus signs of P and Q differences are the same). However, especially underestimations are exaggerated in Q. For the Zambezi the underestimation of observed Q is a result of the high actual evaporation flux (e.g. ECHAM5 gives a runoff coefficient of only 0.11, while a runoff coefficient of more than 0.2 has been derived from observed data (Fig. 4.6)). This observation based

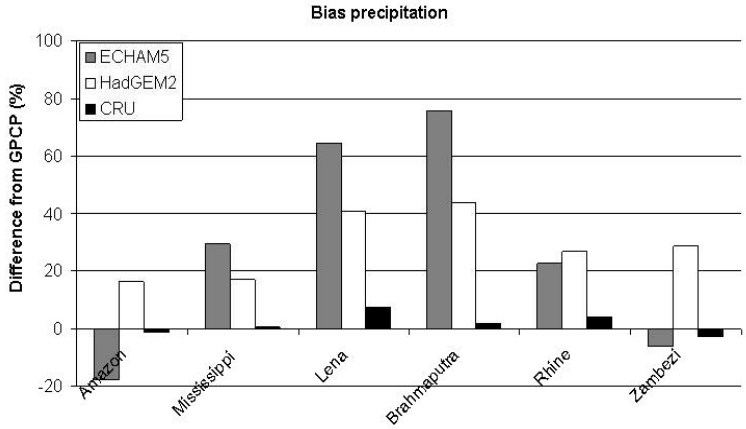


Figure 4.2 Deviation (%) of calculated monthly 30 year average GCM and CRU basin precipitation from 30 year average observed GPCP precipitation: $100 \cdot (\overline{P}_{GCM} - \overline{P}_{GPCP}) / \overline{P}_{GPCP}$.

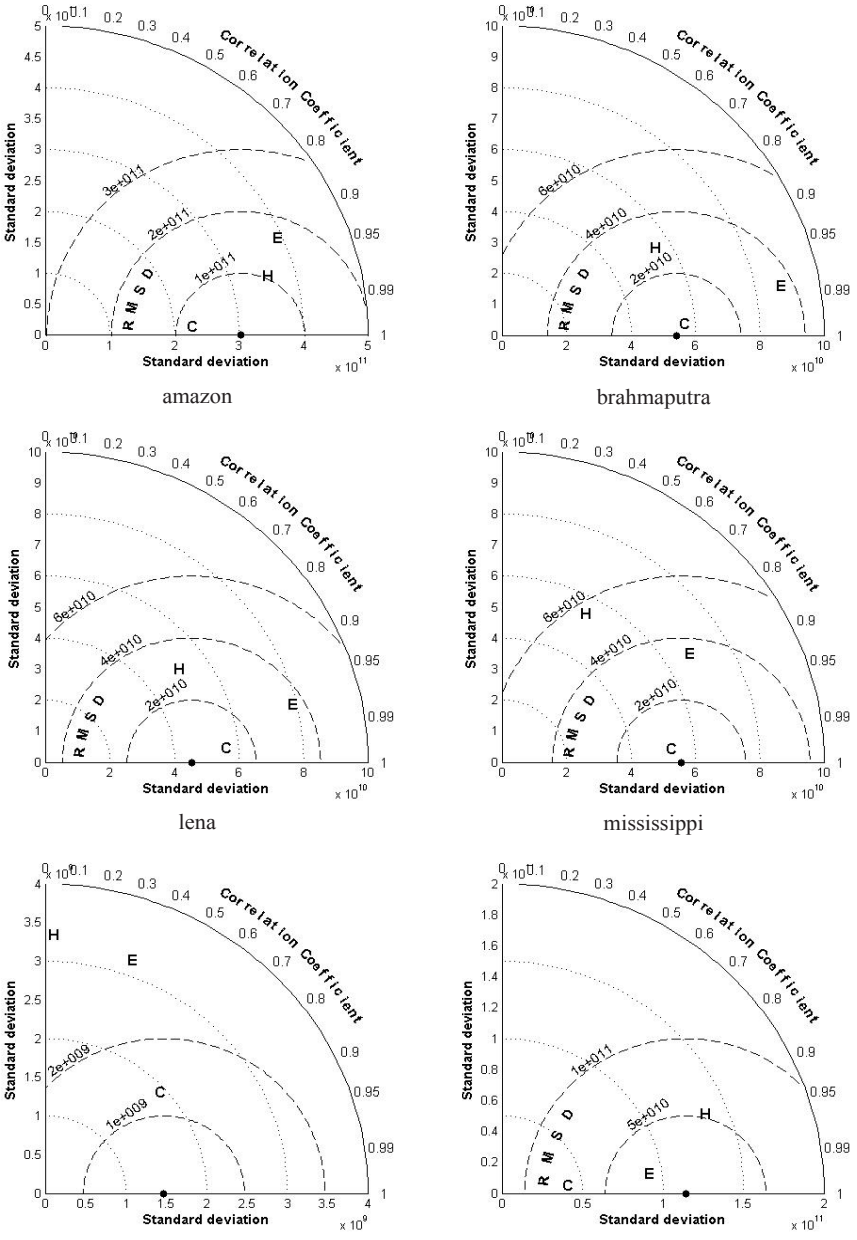


Figure 4.3 Taylor plots giving a comparison of the annual mean precipitation cycle of ECHAM5 (E), HadGEM2 (H) and CRU (C) with the observed annual mean precipitation cycle derived from GPCP precipitation (black dot). The standard deviations of the annual discharge cycle, representing the amplitude of the cycle, are displayed as the radial distance to the origin. The root mean square difference is displayed as the radial distance to the annual mean of the reference GPCP dataset. The correlation between the specific datasets and the reference GPCP dataset is given by the azimuthal positions.

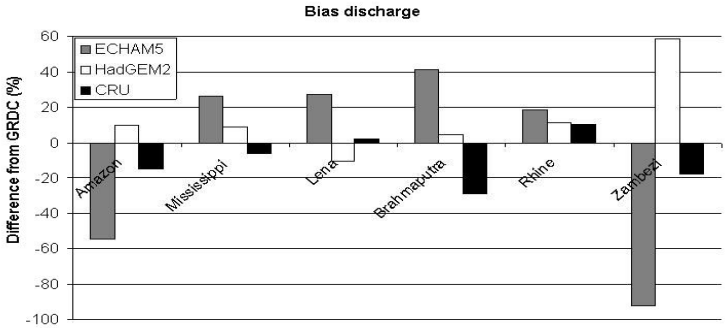


Figure 4.4 Deviation (%) of calculated monthly 30 year average modeled GCM and CRU discharge from 30 year average observed GRDC discharge: $100 \cdot (\overline{Q}_{GCM} - \overline{Q}_{GRDC}) / \overline{Q}_{GRDC}$.

runoff coefficient is calculated by dividing observed basin discharge (e.g. discharge measured at the most downstream gauge in the catchment) by basin total GPCP P (Fig. 4.6). Furthermore, for the Zambezi, the conversion of P to Q heavily diminishes the correlation with the observed annual cycle. Here, river discharge may be biased by water use, the model's overestimation of actual evapotranspiration and the presence of dams which are not schematized in detail in the routing model.

The correlation between the annual cycles of the two GCMs with the observed annual cycle is lower for Q than for P (0.6 vs 0.9) for the Amazon as well. This is most likely due to the complexity of the re-circulation of precipitation which is being converted to actual evapotranspiration and becomes available for precipitation again within the Amazon basin (McGuffie and Henderson-Sellers, 2005). For the Brahamaputra the Taylor plots are different for ECHAM5 and HadGEM2 P. Yet, the Taylor plots for modeled Q are comparable for the two GCMs, although different from the plot for CRU Q. The comparison of Taylor diagrams for P and Q for the Mississippi indicates a similar reduction of differences between the two GCMs as obtained for the Brahmaputra. Correlations with observed discharge are unexpectedly high for the Mississippi, as the river is highly regulated and this regulation is only included in the hydrological model in a limited form. Yet, the annual discharge cycle is influenced by the temporal storage of precipitation as snow in the mountainous regions. This temperature dependant temporal storage is likely to increase the correlation while transitioning from precipitation to discharge. For both GCMs for the Rhine, and for HadGEM2 for the Lena, the correlation with the observed Q cycle is larger than the correlation with the observed P cycle. This is also caused by temporal storage of precipitation as snow within the hydrological model, the improvement is largest for HadGEM2. Although the Brahmaputra basin has a significant snow component as well, the temperature impact on Brahmaputra discharge is small due to the strong influence of the Monsoon, which coincides with the melt season. The large overestimation of annual average precipitation in combination with a high runoff coefficient (Fig. 4.6) results in an overestimation of the annual average discharge in the Brahmaputra basin.

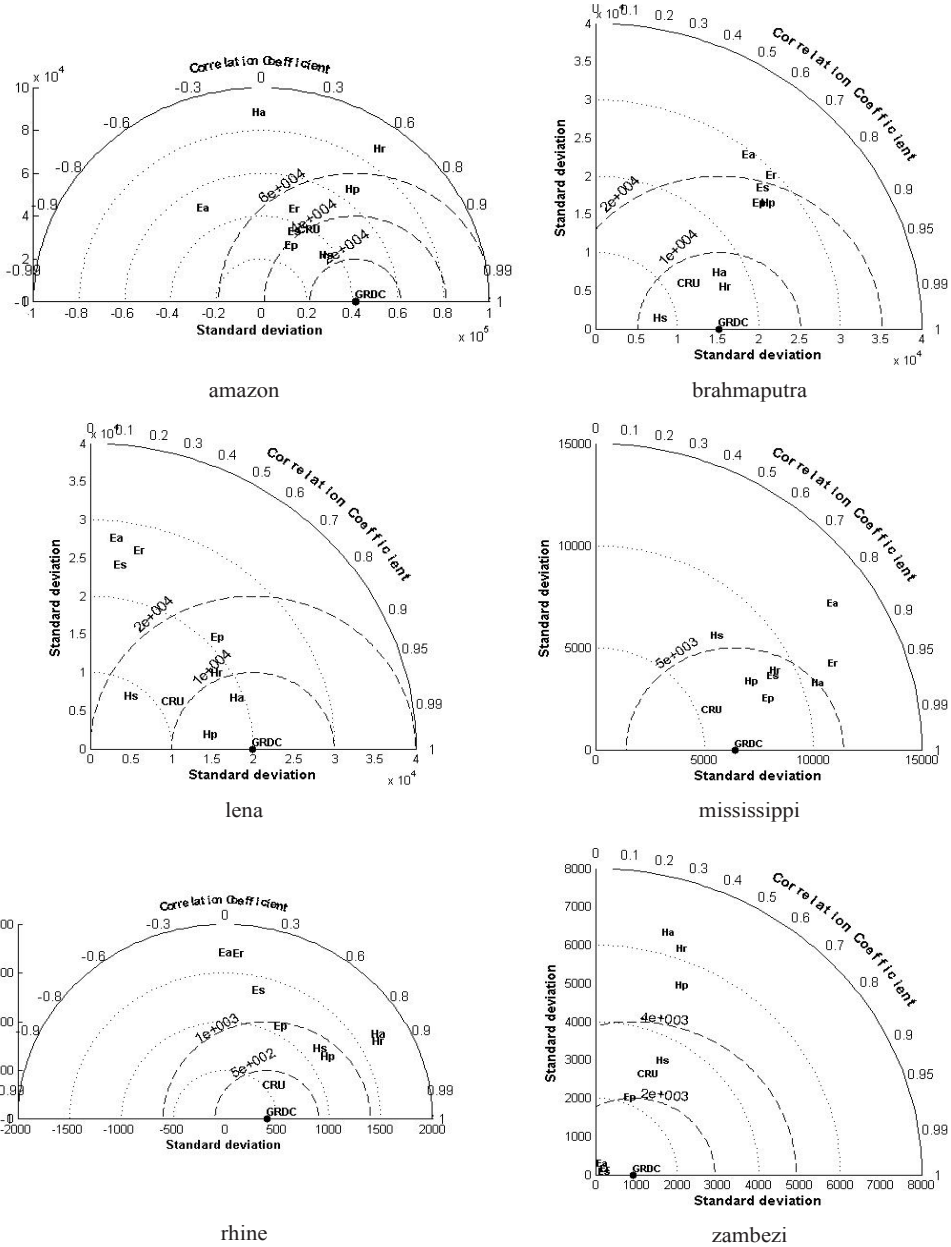


Figure 4.5 Taylor plots giving a comparison of the annual mean discharge cycle derived from ECHAM5 (E), HadGEM2 (H) and CRU (C) with the observed annual mean discharge cycle derived from GRDC discharge (black dot) for the ACCU (a), MRRO (r), MRROSS (s) methods and PCR-GLOBWB (p).

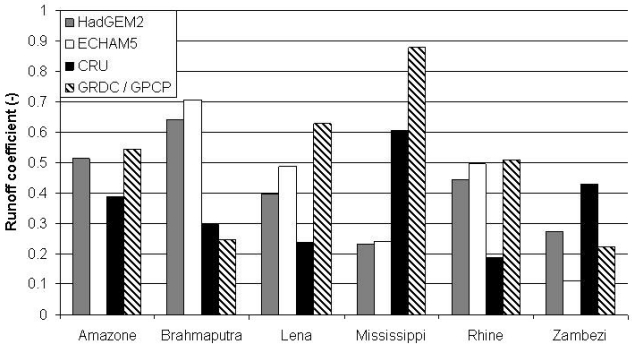


Figure 4.6 Runoff coefficients (-). For ECHAM5, HadGEM2 and the CRU dataset the RC is calculated by: $RC = 1 - (\overline{E} / \overline{P})$. As a reference a runoff coefficient derived from discharge observations (GRDC) and observed precipitation (GPCP) is included: $RC = \overline{Q} / \overline{P}$

The effect of kinematic wave routing

In this section we evaluate the importance of river routing for the reproduction of observed timing of the anuual discharge cycle. To this end the ACCU and MRRO method are compared (e.g. accumulated basin runoff from HadGEM2 (Ha) and ECHAM5 (Ea) and kinematic wave routed basin runoff from HadGEM2 (Hr) and ECHAM5 (Er) in Fig. 4.5). For ECHAM5 the correlation between the observed and modeled annual discharge cycle is highest for all rivers when kinematic wave routing of surface runoff is applied. The difference between routed and accumulated runoff is smallest for the Rhine where the, through routing introduced, discharge delay is smallest due to the small catchment size. The difference is also small for the Zambezi, where for all methods applied to ECHAM5 data, deviations from observed discharge are large. The largest differences between accumulated and kinematic wave routed runoff are obtained for the Amazon. Here simple accumulation of runoff results in a negative correlation with the observed annual discharge cycle, whereas routing introduces the required travel time which in turn results in a realistic annual cycle.

For the Lena for HadGEM2, kinematic wave routing does not result in an increase in the correlation with the observed annual cycle. The snow melt driven discharge rise in spring, which characterizes this basin, occurs too late in the year for the ACCU method and within the MRRO method this delay is actually increased. For HadGEM2 a large increase in the correlation (0 to 0.6) is only found for the Amazon. Yet, routing reduces RMSD values for both HadGEM2 and ECHAM5 runoff.

The effect of subsurface flow delay

To evaluate the effect of a delay of subsurface runoff by temporal storage in a linear reservoir, we compare the MRRO method (kinematic wave routing) and the MRROS method (kinematic routing and groundwater reservoir storage). In the Taylor-plots (Fig. 4.5) these model realizations are displayed as routing HadGEM2 (Hr) and ECHAM5

(Er) and routing and storage HadGEM2 (Hs) and ECHAM5 (Es). Overall, differences between the MRRO and MRROS methods applied to the ECHAM5 runoff fluxes are small, due to the relatively small percentage of subsurface runoff within the total runoff flux (Fig. 4.7a). For HadGEM2 the MRROSS method results in an underestimation of the amplitude of the annual hydrological cycle for all six basins. The subsurface flux is relatively large within HadGEM2 (Fig. 4.7b). As a consequence, the temporal storage of subsurface runoff in the groundwater reservoir is too large and it introduces an undesired delay in the annual Q cycle, which in turn results in a large baseflow that reduces seasonal discharge variation. Surface runoff is rare for HadGEM2 because saturated hydraulic conductivity is assumed to be uniform over a grid-cell (Clark and Gedney, 2008) and no - subgrid variable saturated area - concept is included. In most humid regions (Northern Hemisphere, Amazon and South-East Asia) the percentage subsurface runoff in HadGEM2 is more than 70% of total runoff. Therefore, for most basins, the MRRO method applied to HadGEM2 data, results in higher correlation with observed Q than the MRROS method. Recent attempts have been made to include subgrid saturation variability within HadGEM2 based on TOPMODEL principles, where soil moisture and runoff are related to topography (Clark and Gedney, 2008; Beven and Kirkby, 1979).

The effect of using different land surface hydrology schemes

In this section the influence of using different LSMs is tested. To this end the MRRO method (Er and Hr) is compared with PCR-GLOBWB runs forced with either ECHAM5 or HadGEM2 meteorological data (Ep and Hp). The land surface modelling within PCR-GLOBWB clearly outperforms ECHAM5 for the Lena and Rhine. Lower RMSD values are obtained and, particularly for the Lena, the correlation with the observed annual cycle is higher. Differences are smaller for the Missipippi and Amazon, but still the PCR-GLOBWB runs result in lower RMSD values and standard deviations closer to observed. Differences are small for the Zambezi and Brahamaputra.

PCR-GLOBWB outperforms the MRRO method applied to HadGEM2 for the Amazon, Rhine and Zambezi. Correlation coefficients with the observed annual cycle are comparable for the two methods, however direct routing of GCM runoff results in overestimation of the amplitude and larger absolute deviations from observed discharge values. Differences are small for the Lena and Mississippi and the LSS of HadGEM2 outperforms PCR-GLOBWB for the Brahmaputra.

Notable is the small difference between global patterns of the, with PCR-GLOBWB modeled, percentage subsurface runoff obtained from either ECHAM5 or HadGEM2 meteorological data (Fig. 4.7c and d). This suggests that the division of runoff in a surface and subsurface component is influenced mainly by the definition and parameterization of the LSS, or water balance model, and that the meteorological forcing has only a minor impact.

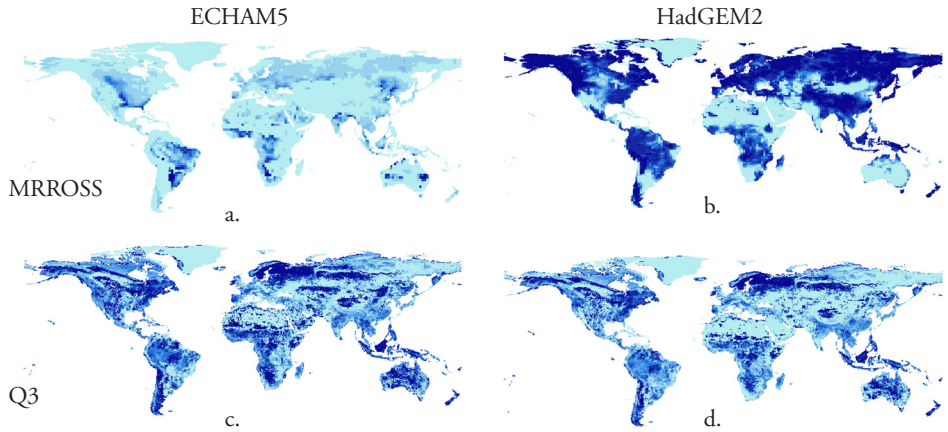


Figure 4.7 Percentage subsurface flux (MRROSS) of total runoff flux (MRRO) in ECHAM5 (a) and HadGEM2 (b). Percentage subsurface flux (Q3) of total runoff flux in PCR-GLOBWB forced with meteorological data of ECHAM5 (c) and HadGEM2 (d).

4.3.2 Auto-correlation for time-lags of several days

Figure 4.8 displays correlograms for the Amazon and Rhine at the most downstream discharge stations in the catchments (i.e. Obidos and Rees). The auto-correlation of the discharge of the Rhine decreases more quickly with an increasing time-lag than that of the Amazon. The smaller Rhine catchment has a higher sensitivity to variations in surface runoff, subsurface runoff and precipitation. Because of this sensitivity, the difference between methods is larger for the Rhine. The results of the PCR-GLOBWB runs, forced with meteorological data from either GCM, are comparable. This exemplifies again the importance of the parameterization of the partitioning of runoff in subsurface and surface runoff and the delaying influence of the groundwater reservoir and routing scheme. For the Amazon the correlation coefficient remains above 0.9 for time-lags of multiple days for most methods. The ACCU method for ECHAM5 is the only exception. For this method the correlation coefficient drops below 0.65 after three days. The auto-correlation is relatively small because the ECHAM5 runoff flux is accumulated over the entire catchment on a daily basis and because the cell-specific runoff itself has a larger day-to-day variability than the HadGEM2 flux due to a smaller subsurface component. This indicates again that the delays introduced by groundwater storage and channel travel time are the most important elements in explaining discharge autocorrelation.

4.3.3 Hydrographs

ECHAM5

The hydrographs of the Rhine and Amazon (Fig. 4.9) illustrate that the importance of routing on day-to-day variability depends on catchment size. Note that in fig. 4.9

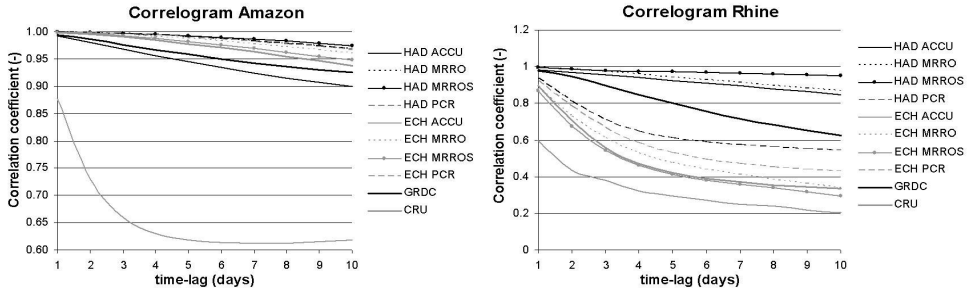


Figure 4.8 Correlograms for the Amazon (left) and Rhine (right) discharge.

hydrographs are shown for a random year. They can not be directly compared with one another as, because of their chaotic nature, a given year within a GCM is only a possible representation of a random year in that period. Therefore, the hydrographs only give an indication of performance for timing and the reproduction of extremes. For the Amazon, the hydrograph generated with the MRRO method from ECHAM5 runoff shows a realistic reduction in day-to-day variability. While with the ACCU method day-to-day variability is too high. Furthermore, when kinematic wave routing is applied, the peak discharge occurs later and a relatively constant baseflow is reached at the end of the summer. Even though ECHAM5 has a small baseflow component, the delay introduced by routing is so large that realistic hydrographs are obtained with the MRRO method. Although routing reduces the day-to-day variability for the Rhine as well, the variability remains larger than for the Amazon. For the Rhine, peak-flows of more than 20,000 m^3/s are calculated with both the MRRO and ACCU method, while the maximum observed discharge at Lobith is 12.600 m^3/s (Te Linde et al., 2010). In addition, discharge values below 1 m^3/s are obtained for low flow periods. This is clearly the result of an underestimation of the baseflow component, which in a smaller river is not compensated for by residence time in the drainage network. The hydrographs obtained with the MRRO and MRROSS method are comparable, mainly due to the small subsurface runoff flux. Notable is the small difference between the hydrographs of the MRROSS method and PCR-GLOBWB for the Rhine.

HadGEM2

The Rhine hydrographs obtained from HadGEM2 data with the MRRO and ACCU method show large resemblance to one another. Compared to ECHAM5, there is less day-to-day variability and the peak discharges are lower. This difference can mainly be explained by the large subsurface runoff flux in HadGEM2 for the Rhine catchment (93% of total runoff versus 36% of total runoff in ECHAM5) and by the small effect of routing within this relatively small basin. Due to the large baseflow component, the MRROSS method results in the most constant river flow throughout the year. This constant river flow best resembles the observed flow. We note however that the small variation in observed seasonal cycle is also a result of river regulation, which is in a limited form only included in the routing scheme. For the Amazon the MRROSS

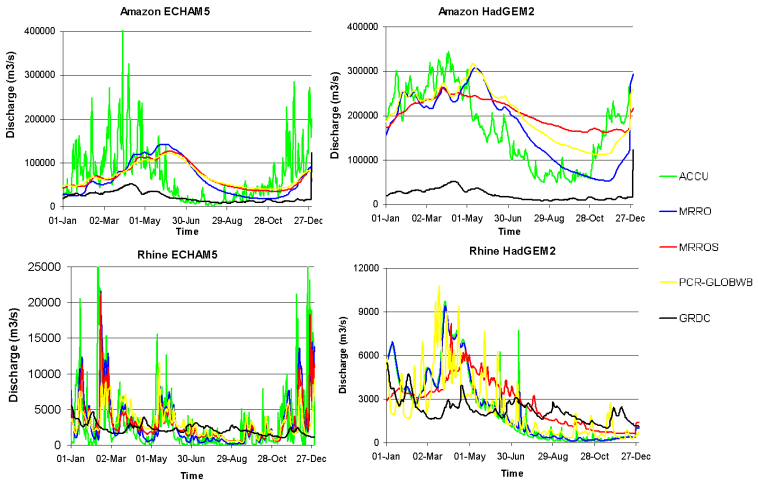


Figure 4.9 Hydrographs for the Amazon (top) and Rhine (bottom) for a random year for the different discharge generation methods applied to ECHAM5 (left) and HadGEM2 (right) data.

method results in the smallest seasonal variation with the highest summer baseflow, as was observed for the Rhine as well. The hydrographs obtained from the MRRO and ACCU method show the lowest baseflow and for these methods baseflow is closest to observed summer discharge. However, this is mainly due to the overestimation of observed discharge by HadGEM2 as a result of too much rainfall. The ACCU method applied to HadGEM2 data results in a lower day-to-day variability than the ACCU method applied to ECHAM5 data. Still, for HadGEM2, the ACCU method is the method with the highest day-to-day variability.

4.4 Discussion and conclusions

In this study we evaluated the usability of GCM runoff (from the GCMs ECHAM5 and HadGEM2) for hydrological impact studies. For this, we used four different discharge generation methods ranging from simple GCM runoff accumulation, to runoff routing, to advanced hydrological modelling. For all four methods, generated discharges show deviations from observed discharges due to biases in GCM runoff, precipitation and evaporation fluxes and the absence of water use in the models.

Routing and temporal storage in a groundwater reservoir, as included in the MRROSS and PCR-GLOBWB methods, introduce a delay (Oki and Sud, 1998). This delay reduces discharge peaks to realistic values and introduces constant baseflows.

Too high peak flows and day-to-day variability are obtained with the ACCU method, which lacks discharge routing. Differences between this method and the other methods are larger for the Amazon than for smaller catchments, such as the Rhine. However, we only conducted our analysis for relatively large catchments using GCM data and the value of GCM or RCM runoff should be assessed on smaller scales as well. For such

an analysis, additional runoff downscaling methods based on soil properties and topography (Beven and Kirkby, 1979) might be useful. Furthermore, the analysis is restricted to two GCMs. It would be interesting to extend the analysis when runoff fields become available for more GCMs.

The schematizations of the LSSs of ECHAM5 and HadGEM2 are comparable to the hydrological model PCR-GLOBWB, at least in detail and complexity of the hydrological processes included. However, the incoming precipitation is divided differently over the water balance components (e.g. surface runoff, subsurface runoff and actual evapotranspiration) due to a different parameterization and formulation of the subsurface hydrological processes. In addition, the resolution of PCR-GLOBWB is higher. In HadGEM2 the low resolution in combination with the lack of sub-grid parameterization results in a large subsurface runoff flux. This affects the annual discharge cycle and may influence the surface energy balance of the GCM as well. In ECHAM5 the low resolution is partly compensated for by sub-grid parameterization of surface runoff, yet here the fraction of groundwater runoff is probably too small.

The Taylor-plots show that PCRGLOB-WB outperforms the other runoff generation methods. For ECHAM5 this is the case for all six rivers, for HadGEM2 for the Mississippi, Lena and Rhine. Nevertheless, this study also illustrates that differences between PCR-GLOBWB and the MRRO and MRROSS method can be small. These two methods are promising because of their limited computational demand.

So far, land surface schemes are typically tuned to reproduce the surface energy balance correctly. However, the analysis of these two GCMs and comparison with results from a large-scale hydrological model, show that, if GCM runoff generation is additionally tuned with discharge observations and, even more importantly, a routing scheme is added, than discharge derived from GCM runoff can be as suitable as discharge derived from runoff calculated by an off-line hydrological model for large scale studies.

5 Global patterns of change in runoff regimes for 2100

This chapter is based on:
Sperna Weiland, F. C., L. P. H. van Beek, J. C. J. Kwadijk and M. F. P. Bierkens (2011), Global patterns of change in runoff regimes for 2100, Climatic Change, under review

Abstract

This study makes a thorough global assessment of effects of climate change on hydrological regimes and the accompanying uncertainties. Meteorological data from twelve GCMs (SRES scenario A1B, and control experiment 20C3M) are used to drive the global hydrological model PCR-GLOBWB. We reveal in which regions of the world changes in hydrology can be detected that are significant and consistent amongst the ensemble of GCMs. New compared to existing studies is: 1) the comparison of spatial patterns of regime changes and 2) the quantification of consistent significant change calculated relative to both the natural variability and the inter-model spread. The resulting consistency maps indicate in which regions likelihood of hydrological change is large.

Projections of different GCMs diverge widely. This underscores the need of using a multi-model ensemble. Despite discrepancies amongst models, consistent results are found: By 2100 the GCMs project consistent decrease in runoff for southern Europe, southern Australia, parts of Africa and southwestern South-America. Discharge decreases are large for most African rivers, the Murray and the Danube. While discharge of Monsoon influenced rivers slightly increases. In the Arctic regions runoff increases and a phase-shift towards earlier peaks is observed. Results are comparable to previous global studies, with a few exceptions. Globally we calculated an ensemble mean discharge increase of more than ten percent. The increase contradicts previously estimated decreases.

5.1 Introduction

Climate change will have significant effects on global runoff regimes and will affect water availability for agriculture and ecosystems as well. To anticipate on these changes, reliable assessments of the hydrological effects of climate change, including information on uncertainties are needed. Studies investigating hydrological effects of climate change on continental or global scales are often based on results from General Circulation Models (GCMs). However, especially for precipitation, GCMs produce quite varying and even contradictory results (Covey et al., 2003).

There have been quite a number of studies investigating hydrological consequences of climate change on a global scale (see table 5.1 for an overview). Overall the results of these studies project a decrease in runoff for southern Europe, north and south Africa, southwestern USA, Mexico and Brazil and an increase in discharge for Monsoon driven and Arctic rivers. Several studies (Alcamo and Henrichs, 2002; Arnell, 1999b; Arnell, 2003; Nijssen et al., 2001; Vörösmarty et al., 2000) used a change factor method instead of directly applying the climate model data for the future period. Within the change factor method observed precipitation, temperature or runoff fields are adjusted with a change factor derived from climate model data and those adjusted time-series are then used to derive future runoff and discharge changes. The method assumes that change is more reliable than absolute values. However, this only holds under the assumption of a constant model bias through time. Furthermore change in variability is ignored (Fowler et al., 2007). Although computationally more demanding than the change factor method, directly forcing a hydrological model with climate model data for current and future climate and calculation of differences in obtained discharges may give more reliable estimates of changes in variability and extremes.

Change in runoff can also directly be derived from runoff fields calculated by GCMs. Unfortunately such data is not accessible for most models and in most GCMs river routing is not included. To obtain information on changes in river regimes, additional routing of GCM runoff fields is needed (Arora and Boer, 2001; Milly et al., 2005; Nohara et al., 2006). Therefore river discharge is most often calculated with a hydrological model that includes a routing model, using either meteorological variables directly from GCMs (Aerts et al., 2006) or using observed meteorological time series perturbed with change factors derived from GCM results (Alcamo and Henrichs, 2002; Nijssen et al., 2001; Vörösmarty et al., 2000).

For a climate effect study it is possible to select datasets from multiple GCMs for multiple emission scenarios. Arnell (2003) showed that by 2050 there is little difference between the emission scenarios, i.e. correspondence between GCMs is weaker than between scenarios. This indicates that the choice of GCMs highly influences the calculated change and it has been concluded before that a multi-model ensemble of GCMs provides the most reliable impression of the spread and uncertainties of possible changes (Boorman and Sefton, 1997; IPCC, 2007b; Murphy et al., 2004). Arnell

Table 5.1 Literature overview of studies investigating hydrological effects of change on a global scale.

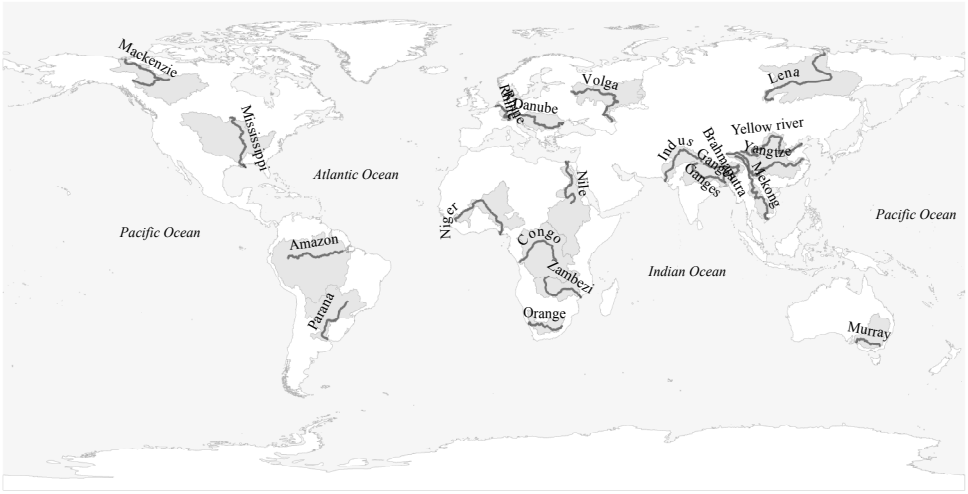
Study	*Climate model*	*Scenario*	*Runoff*	*Method*	*Horizon*	*Parameters*	*Significance / Consistency*	*Rivers / Regions*	*Results*
Aerts (2006)	ECBilt-CLIO-VECODE	A2	Hydrological model: STREAM including simplified routing scheme	Direct use of GCM meteo data	21st century	- Mean decadal change in discharge compared to discharge 1750-2000 - Inter decadal variability compared with natural variability	~ Comparison to past variability	Globe Amazon, Congo,Danube, Ganges, Lena,MekongMiss issippi, Murray, Nile, Odra, Rhine, Syr-Darya, Yukon, Volga,Volta	**Discharge increase:** Congo, Mekong, Ganges, Amazon, Rhine, Murray, Volga **Discharge decrease:** Nile, Danube, Mississippi **Seasonal shift:** Lena
Alcamo (2002)	ECHAM4 HadCM3	A2, B2	Hydrological model: WaterGAP including routing	Change factor	2020s, 2050s, 2080s	- Annual withdrawal-to-availability ratio - Consumption- to-Q90 ratio - Per capita wa-ter availability	Overlap between three parameters selected as indicators of climate change	Globe	**Severe water stress:** Southwester USA, central Mexico, northeast Brazil, West Coast Latin America, northern and southern Africa, Middle East
Arnell (1999b)	HadCM2 HadCM3	1% per year CO2 increase	Hydrological model, no routing	Change factor	2020s, 2050s, 2080s	- Average annual runoff - Water Stress	-	Globe 42 rivers	**Change in high flow:** North-America, east Asia, Ghana **Increasing water stress:** Mediterranean region, Middle-East, South- Africa, parts of south Asia **Seasonal shift:** Belarus
Arnell (2003, 2004)	HadCM3 CGCM2 CSIRO Mk2 ECHAM4 GFDL_R30_c CCSR/NIES2	A1, A2, B1, B2	Hydrological model, routing with monthly output	Change factor	2020s, 2050s, 2080s	Average annual runoff Drought runoff InterAnnual variability Flood runoff Annual cycle	Consistency among scenarios, compared to consistency among models	Globe	**Runoff increase:** High latitudes, east Africa, south and east Asia **Runoff decrease:** Southern and eastern Europe, western Russia, Middle East, Africa and much of North- and South-Africa
Arora and Boer (2001)	CGCM1	GHG+A based on IS92a	runoff from from GCM as input for routing model	Direct use of GCM runoff fields	2070-2100	Mean discharge, amplitude and phase, flood discharge, annual max discharge and sdv, flow duration curve	-	Globe 23 rivers	**Runoff decrease:** Africa, Amazon, Yangtze, Mekong. Global decrease 14% **Seasonal shift, decrease in amplitude:** Mid- and High latitude rivers

Table 5.1(continued)

Study	*Climate model*	*Scenario*	*Runoff*	*Method*	*Horizon*	*Parameters*	*Significance / Consistency*	*Rivers / Regions*	*Results*
Manabe (2004)	GFDL-GCM ensemble of 8 experiments	IS92a, CO2 quadrup-ling	Discharge derived from runoff fields GCMs	Direct use of GCM runoff fields	2050	Change in annual mean runoff Spatial pattern of change in seasonal soil moisture	-	Globe 42 rivers	**Discharge increase:** globally 7.3% by 2050, Arctic rivers, Brazil, Andes, northern India, Tibet, Indonesia, West-Africa, Amazon, Ganges, Brahmaputra **Discharge decrease:** Nile, Mekong **Soil moisture decrease:** North-America, Mediterranean Coast, northeast China, grasslands of Africa and southern and western regions of Australia
Milly (2005)	12 GCMs (best models from ensemble of 21 IPCC AR4 modds)	A1B	Runoff fields from GCMs as input for routing model	Direct use of GCM runoff fields	2041-2060	Annual mean change compared with past trends	Number of models showing positive changes minus number of models showing negative change	Globe	**Runoff decrease:** Southern Europe, Middle-East, mid-latitude western North-America, southern Africa **Runoff increase:** High latitude North-America, Eurasia, South-America, eastern equatorial Africa
Nohara (2006)	ensemble of 19 AOGCMs (part of IPCC AR4)	A1B	runoff fields from GCM as input for routing model	Direct use of GCM runoff fields	2100	Mean annual change in runoff River regimes	- Ensemble change relative to intermodel variability of the change signal - GCM deviations for control run	Globe, 24 major rivers	**Discharge increase:** Northern Hemisphere (Arctic rivers), southern to eastern Asia (Mekong, Ganges), central Africa **Discharge decrease:** Central America, southern Africa, Mediterranean region, southern North-America, Rhine, Danube **Seasonal shift:** Arctic and mid-latitude rivers

Table 5.1(continued)

Study	*Climate model*	*Scenario*	*Runoff*	*Method*	*Horizon*	*Parameters*	*Significance / Consistency*	*Rivers / Regions*	*Results*
Nijssen (2001)	HCCPR-CM2, HCCPR-MC3, ECHAM4, DOE-PCM3 (selected on resolution out of eight GCMs)	IS92a, A,B,C	Hydrological model: VIC including routing	Basin wide change factor	2025 2045	Annual hydrological cycle Change in water balance Seasonal change Moisture deficit periods Basin sensitivity	-	Amazon, Amur, MacKenzie, Xi, Mekong, Yellow River, Yenisei, Mississippi, Severnaya Dvina	**Discharge increase:** Arctic rivers **Discharge decrease:** mid-latitude and tropic basins **Seasonal shift:** Arctic rivers
Vörösmarty (2000)	HadCM2, CGCM1		Hydrological model: WBM including routing	Direct use of GCM meteo ata Change factor for discharge change	2025	Water stress Annual runoff	-	Globe	**Decreased water availability:** East Africa, southeast Asia, Mexico, Spain, parts of North- and South-America
This study	BCM2.0, CGCM3.1, CGCM2.3.2, CSIRO-Mk3.0, ECHAM5, ECHO-G, GFDL-CM2.1, GISS_ER, HADGEM1, IPSL-CM4, MIROC3.2medres, NCAR-CCSM3	A1B and A2	Hydrological model: PCRGLOB-WB including routing	Direct use of meteo data	2100	Mean, max and minimum annual runoff, annual cycle, inter-annual variability	- Significance compared to natural variability and ensemble uncertainty - Consistency amongst GCMs	Globe, 20 major rivers	**Discharge increase:** Arctic rivers **Discharge decrease:** Southern Australia, southern Africa, Mediterranean region, southwest South-America **Seasonal shift:** Sub-Arctic rivers



Catchment	*Area* (km^2)	*Qavg* (m^3/s)	*Gauge*	*Catchment*	*Area* (km^2)	*Qavg* (m^3/s)	*Gauge*
Amazon	6.915.000	190.000	Obidos	Murray	1.061.469	767	Wakool Junction
Brahmaputra	930.000	48.160	Bahadurabad	Niger	2.117.700	6.000	Dire
Congo River	3.680.000	41.800	Kinshasa	Orange river	973.000	365	Aliwal North
Danube	817.000	6.400	Ceatal Izmail	Parana	2.582.672	18.000	Corientes
Ganges	907.000	12.015	Hardinge Bridge	Rhine	65.683	2.200	Rees
Indus	1.165.000	6.600	Kotri	Volga	1.800.000	31.900	Volgograd
Lena	2.500.000	17.000	Kusur	Yangtze	752.000	2.571	Datong
MacKenzie	1.805.000	10.700	Norman Wells	Yellow river	1.380.000	8.060	Huayuankou
Mekong	2.981.076	12.743	Mukdahan	Zambezi	1.390.000	3.400	Katom a Mulilo
Mississippi	795.000	16.000	Vicksburg				

Figure 5.1 Selected catchments with total catchment area, average observed discharge (*Qavg*) and location of gauges for which statistics are calculated, annual cycles are given and comparisons are made.

(2003), Milly et al. (2005), Nohara et al. (2006) and Nijssen et al. (2001) used these multi-model ensembles.

All studies in table 5.1 indicated directions and amount of change for world regions or river basins, however quantification of the significance of these changes frequently played a minor role. In this study we will make a thorough assessment of the global hydrological effects of climate change by directly applying daily climate data from an ensemble of twelve GCMs for the IPCC SRES scenario A1B for the period 2081-2100 as input to the global hydrological model PCR-GLOBWB. In this hydrological model river discharge is calculated using an explicit routing scheme based on the kinematic wave equation, which also includes temporal storage in flood plains, lakes, wetlands and reservoirs. The relative changes between the current and future climate are analyzed instead of absolute changes, hereby reducing the influence of biases in the hydrological model and GCM data.

In addition to investigating annual mean changes in runoff fields and changes in river

regimes, as has also been done in previous studies, we will here focus on: 1) spatial patterns of change in the annual cycle looking at changes in timing of peak, 2) additional discharge statistics (e.g. maximum and minimum flow and interannual discharge variability), 3) significance of change, which is not only quantified by comparing change against the spread between models, as has been done in a few other studies before. Here we also test the significance of projected changes of individual models relative to their inter-annual variability. This will more clearly reveal regions with consistent and significant change in runoff statistics. And finally focus will be on 4) the consistency amongst model projections on the direction of change, hereby indicating on world maps in which regions the likelihood of hydrological changes is large.

5.2 Data and methods

The distributed global hydrological model PCR-GLOBWB (Van Beek and Bierkens 2009; Bierkens and van Beek 2009) was run on a daily time-step with meteorological time series from 12 GCMs for the 20C3M experiment for the period 1971-1990 and the SRES scenario A1B for the period 2081-2100. From the results we derived change fields of discharge regimes for which the significance of change and the consistency amongst GCMs was quantified. We selected 19 large catchments (Fig. 5.1) which cover a variety of climate zones, latitudes and continents. For these catchments changes in the mean annual cycle are quantified.

5.2.1 Hydrological model

PCR-GLOBWB is a global distributed hydrological model with a resolution of 0.5 degrees (Van Beek and Bierkens 2009). Each model cell consists of two vertical soil layers and one underlying groundwater reservoir. Sub-grid parameterization is used to represent fractions of short and tall vegetation, surface water and for calculation of saturated areas to quantify surface runoff and lateral outflow from the unsaturated zone. Water enters the cell as rainfall and can be stored as canopy interception or snow. Snow accumulation or melt depends on temperature (degree day method). Melt water and throughfall are passed to the surface. Evapotranspiration is calculated from potential evaporation and soil moisture conditions. Exchange of water is possible between the soil and groundwater layers. Runoff is made up of non-infiltrating melt and throughfall water, saturation excess surface runoff, interflow and base flow. For each time-step the water balance is computed per cell. Runoff is accumulated and routed as river discharge along the drainage network taken from DDM30 (Döll and Lehner, 2002) using the kinematic wave approximation of the Saint-Venant equation. Adaptations have been made to the network to improve the inclusion of storage in lakes, wetlands and large reservoirs. Hereto, a selection of substantial lakes and reservoirs (> 500 km^2) was obtained from the GLWD1 data set (Lehner and Döll, 2004). The resulting river discharge represents natural flow. Water and reservoir management, river regulation and other human

influences have not been included. Model parameterization is based on best available global datasets and so far the model has not been calibrated. More information about the model performance can be found in Appendix A.1 and in Van Beek et al. (2011). Because of some apparent deviations, mostly caused by biases in meteorological forcing and additionally by simplifications in model structure and related scale issues, we will focus on relative changes between current and future discharges instead of absolute values. To overcome initialization problems, initial states have been obtained for each GCM dataset individually. For the control climate experiment and the future scenario, PCR-GLOBWB was initialized with the first ten years of data starting with the initial states obtained from a 30 year run based on CRU TS2.1 monthly time series (New et al., 2000) downscaled to daily values using ERA-40 re-analysis data (Uppala et al., 2005). The end-states of the ten year during GCM runs are used as initial states for the 20 year GCM scenario runs

5.2.2 Climate data

Required model inputs are precipitation, temperature and reference potential evaporation. Temperature and rainfall data can directly be obtained from the GCMs. Reference potential evaporation is derived using a modification of the Penman-Monteith equation where missing air humidity fields are not required (Allen et al., 1998; Monteith, 1965). For those models where other required variables (e.g. radiation, air pressure, windspeed, minimum air temperature) were missing the simpler temperature based Blaney-Criddle equation was used (Brouwer and Heibloem, 1986; Oudin et al., 2005). We realize this may have introduced additional noise between the model results (Kay and Davies, 2008). Therefore, in Appendix A.2, an analysis of the influence of using either Blaney-Criddle or Penman-Monteith to calculate potential evaporation, on the modeled discharges and discharge changes is given. Within the hydrological model, crop specific potential evaporation is calculated based on global monthly crop factor maps. These

Table 5.2 Overview of selected GCMs.

Model	*Institute*	*Country*	*Horizontal Resolution (degrees)**	*Vertical Resolution (nr atmospheric layers, nr ocean layers)**	*Acronym*
BCM2.0	Bjerknes Centre for Climate Research	Norway	2.8	(31 , 35)	BCCR
CGCM3.1	Canadian Centre for Climate Modelling and Analysis	Canada	3.75	(31 , 29)	CCCMA
CGCM2.3.2	Meteorological Research Institute	Japan	2.8	(30 , 23)	CGCM
CSIRO-Mk3.0	Commonwealth Scientific and Industrial Research Organization	Australia	1.875	(18 , 31)	CSIRO
ECHAM5	Max Planck Institute	Germany	1.5	(31 , 40)	ECHAM
ECHO-G	Freie Universität Berlin	Germany	3.75	(19 , 20)	ECHO
GFDL-CM 2.1	Geophysical Fluid Dynamics Centre	USA	1.0	(24 , 50)	GFDL
GISS-ER	Goddard institute for Space Studies	USA	4 * 5	(20 , 13)	GISS
IPSL-CM4	Institute Pierre Simon Laplace	France	2.5 * 3.75	(19 , 19)	IPSL
MIROC3.2	Center of Climate System Research	Japan	2.8	(20 , 43)	MIROC
CCSM3	National Center for Atmospheric Research	USA	1.4	(26 , 40)	NCAR
HADGEM1	Met Office's Hadley Centre for Climate Prediction	UK	1.25 * 1.875	(38 , 40)	HADGEM

* Parkinson et al. (2006)

crop factor maps are derived from current land use (Van Beek, 2008). For the future runs possible changes in land use and growing season are neglected.
The Program for Climate Model Diagnosis and Intercomparison (PCMDI) collected model results from GCM runs based on the IPCC SRES scenarios and made the results available through the PCMDI data portal (https://esg.llnl.gov:8443/index.jsp). We selected the emission scenario A1B, which is positioned at the upper range of possible CO_2 emissions. This rather extreme scenario was selected because for the period 2000 to 2006 observed CO_2 emissions have been larger than estimated by models (Canadell et al., 2007; Global Carbon project, 2008) and because the signal to noise ratio is relatively clear for an extreme scenario, especially for a time horizon of 2100. Complete datasets, with the required variables available on a daily time-step for both the 20C3M control experiment (1971-1990) and the A1B emission scenario (2081-2100), could be retrieved for twelve GCMs (see table 5.2). Unfortunately the data availability restricted this analysis to these twelve GCMs, although a larger GCM ensemble would provide more uncertainty information. Furthermore a longer period would have been better for averaging out inter-decadal variability. However, for the future experiments data were only available for a 30 year period for some of the GCMs. Although the data portal does not provide all required variables for the Hadley centre climate models, HadGEM1 has been included for it is frequently used in climate change studies. HadGEM1 data has been retrieved from the CERA-gateway (http://cera-www.dkrz.de).
For a few GCMs multiple realizations were provided (five GCMs with two or more realizations). To avoid unequal influence of the different GCMs on the ensemble mean change and because the consistency amongst changes projected by multiple realization from a single GCM tends to be larger than the consistency of changes projected by an ensemble of multiple GCMs, we only included one run per GCM in our ensemble. In appendix A.3 a brief analysis of the consistency of the multiple realizations for the GCM with the highest number of realizations for both the 20C3M experiment and the A1B scenario (CGCM2.3.2) is given.

5.2.3. Statistical analysis

Statistics
To quantify the projected hydrological changes between the future and control experiments and the consistency of these changes, the statistics in table 5.3 were calculated for each GCM run. In the following sections we describe how changes in these statistical quantities are obtained from the multi-model ensemble and how the significance and consistency of these changes have been quantified.

Relative change
Discharge changes have been calculated relative to the baseline multi-model simulations. We did not look at absolute values, because the GCM precipitation and consequently

the derived discharges deviate from observed quantities for some of the catchments (see Appendix B.1). Relative changes have been calculated for each model individually, according to the following equation:

$$\Delta Q_{future} = (\overline{Q}_{future} - \overline{Q}_{past}) / \overline{Q}_{past} \quad (5.1)$$

Where $\overline{Q}$ can be one of the statistics in table 5.3, *past* refers to the 20C3M experiment and *future* refers to the A1B scenario. For the timing of peak discharges absolute changes were calculated. From the relative change fields per model (ΔQm) we calculated maps with the ensemble mean change (ΔQ) for the different statistics:

$$\overline{\Delta Q} = \frac{1}{12}\sum_{m=1}^{12} \Delta Q_m \quad (5.2)$$

We prefer to work with a non-weighted multi-model mean, since weights have to be derived from past performance and may not hold for future periods because of apparent small persistence in relative model skill (Reifen and Toumi, 2009). The multi- model ensemble, with equal weights assigned to each member, is likely to give good results and contains all the uncertainty information available. Furthermore, weighting on a limited number of indices of GCM performance may result in a misleading estimate of change, because the more complex picture of the relative merits of the individual GCMs is hidden (Gosling et al., 2011).

Table 6.3 Parameters included in analysis. Q_{mean_h}, Q_{min_h} and Q_{max_h} are respectively the mean, minimum and maximum daily discharge of year *h*. *h* is the year number and ranges from 1 to 20. $H = 20$, the total number of years. Q_{peak_h} is the number of the month in which discharge peak occurred in year *h*. $\overline{P}_{sum}$ is the twenty year average yearly precipitation sum and $\overline{AET}_{sum}$ is the twenty year average yearly actual evaporation sum.

Parameter	***Definition***	***Equation***
$\overline{Q}_{\min}$	20 year average annual minimum discharge	$\overline{Q}\min = \frac{1}{H}\sum_{h=1}^{20} Q_{\min_h}$
$\overline{Q}_{\max}$	20 year average annual maximum discharge	$\overline{Q}\max = \frac{1}{H}\sum_{h=1}^{20} Q_{\max_h}$
$\overline{Q}_{mean}$	20 year average annual mean discharge	$\overline{Q}mean = \frac{1}{H}\sum_{h=1}^{20} Q_{mean_h}$
$\overline{Q}_{peak}$	20 year mode of month in which yearly discharge peak occurs	$\overline{Q}_{peak} = \mathrm{mod}(Q_{peak_h})_{0,..,h}$
σ	Inter-annual variability in mean annual discharge	$\sigma = \sqrt{\frac{1}{H}\sum_{h=1}^{20}\left(Q_{mean_h} - \overline{Q}\right)^2}$
RC	20 year average runoff coefficient	$RC = \frac{\overline{P}_{sum} - \overline{AET}_{sum}}{\overline{P}_{sum}}$

Significance and consistency

Significance of change in the multi model ensemble mean is tested by comparing the ensemble mean discharge statistics calculated for the scenario A1B, with the ensemble mean discharge statistics obtained from the 20C3M experiment using a paired T-test. Each of the GCM runs is in this test an individual sample and significance is tested for a significance level of 95%. The test shows for which regions a reported change in discharge statistics is significant compared to the inter-model uncertainty. The T-statistics are calculated as:

$$x_d = \sum_{m=1}^{M} (Q_{GCM\_fut_m} - Q_{GCM\_past_m}) / M \tag{5.3}$$

$$S_d = sdv(Q_{GCM\_fut_m} - Q_{GCM\_past_m})_{m=1,..,M} \tag{5.4}$$

$$t = \frac{x_d}{s_d / \sqrt{M}} \tag{5.5}$$

Where m = model number, M = total number of models, $Q_{GCM\_fut_m}$ = average runoff result for the A1B scenario for one GCM, $Q_{GCM\_past_m}$ = average runoff results for the 20C3M experiment for one GCM. x_d and s_d are respectively the mean and standard deviation of the changes in mean runoff of the 12 GCMs.

In addition to the significance of the ensemble mean change *relative to the spread between GCMs*, we calculated for each GCM individually whether a change in statistics between 20th century climate and 2100 for the A1B scenario is significant *relative to its inter-annual variability*. This was done by applying the independent samples T-test for each GCM individually. However, an inter-annual autocorrelation is expected to exist in the yearly runoff time series, resulting in an effective decrease of the number of independent observations. This dependency was accounted for by calculating the effective sample size from the lagged correlation coefficient, ρ , according to Matalas and Langbein (1962):

$$\frac{1}{k_b^*} = \frac{1}{k} + \frac{2}{k^2} \sum_{h=1}^{k-1} (k-h)\rho_{h\Delta y} \tag{5.6}$$

Where Δy is the observation interval (= 1 year), k is the total number of observations (20) and $h\Delta y$ is the time lag for which the correlation coefficient is calculated. With this equation values of k_b^* (the effective sample size) were calculated for each model cell. Afterwards, independent two sample T-tests were conducted, calculating the significance of change per model using the effective sample size.

$$t = \frac{\overline{Q}_{fut} - \overline{Q}_{past}}{S_{fut\_past}\sqrt{\frac{1}{k_{fut}*} + \frac{1}{k_{past}*}}} \tag{5.7}$$

$$S_{fut\_past} = \sqrt{\frac{(k_{fut}*-1)S_{fut}{}^2 + (k_{past}*-1)S_{past}{}^2}{(k_{fut}*+k_{past}*-2)}} \tag{5.8}$$

Where S_{fut} and S_{past} are respectively the standard deviation of yearly average minimum, maximum or mean discharge for the A1B scenario and the 20C3M experiment. $\overline{Q}_{fut}$ and $\overline{Q}_{past}$ are the 20 year average discharge statistics for the A1B scenario and control experiment and $k_{fut}*$ and $k_{past}*$ are the effective degrees of freedom as calculated with Eq 5.6. To quantify the consistency in projected change between the twelve models, the number of models projecting significant change in the dominant direction (i.e. the direction of the mean change of the multi-model ensemble) was calculated for each individual model cell. The resulting consistency maps indicate for which regions of the world the models project consistent significant changes in discharge and where consequently likelihood of discharge changes is higher than in other regions.

Annual cycle

To illustrate the changes in monthly flows and possible seasonal shifts, mean annual cycles have been derived for each catchment. In the first step, mean annual cycles were derived over the twenty year model run period for each model individually for both the 20C3M experiment and the A1B scenario. The two resulting sets of twelve GCM derived annual cycles gave for each month long-term average distributions of GCM derived discharge from which for the 20C3M experiment and A1B scenario individually the mean, 10th- and 90th-percentile discharges per month were calculated. By doing so the plots of the resulting annual regimes do not only give information on the changes in mean annual cycle, but also on the spread in the annual cycles obtained from the ensemble of models.

5.3 Results

Global maps with monthly mean discharge and actual and potential evaporation derived from the daily results of the GCM based hydrological model runs (e.g. hydrological scenario data) can be downloaded from: http://public.deltares.nl/display/CAW/Global+hydrological+effects+of+climate+change

5.3.1 Global patterns of change

In figure 5.2 global maps with the multi model ensemble average relative change in mean, minimum and maximum annual discharge are shown. The mean discharge is the 20-year average annual mean daily discharge, minimum discharge is the average of the minimum daily discharge calculated for the twenty individual years and maximum is the average of the maximum daily discharge calculated for the twenty individual years. The regions where minimum, maximum and mean discharge increase and decrease are similar. Although regions with decreases are more extended for minimum discharge in the US and Eastern Europe and increases in maximum discharge are larger in Arctic and Sub-Arctic regions. Similar global patterns of change can be found in literature (Alcamo et al., 2007; Milly et al., 2005; Nohara et al., 2006). In addition, several studies (Alcamo and Henrichs, 2002; Alcamo et al., 2007; Arnell, 1999c; Vörösmarty et al., 2000) indicated large parts of the regions, for which we calculated discharge decreases, as areas currently experiencing water stress. According to these studies, water stress will increase for most of these areas, depending on the definition of the water use scenario.

Figure 5.3 shows ensemble average seasonal discharge changes. Seasonal changes in precipitation, temperature and actual evaporation were derived as well to explain discharge changes. However, for briefness, maps resulting from these calculations have not been included. Maximum discharge increases are projected for the Arctic and sub-Artic regions and for south-east Asia. These increases are related to an increase of precipitation in the JJA and SON seasons. Figure 5.3 shows that in North-Western Europe and the Eastern US winter runoff increases while summer runoff will decrease. This mirrors changes in precipitation distribution over the year, with wetter boreal winters and drier boreal summers. Areas around the Mediterranean Sea, the south-west of South-America, parts of south and north Africa and the south of Australia experience discharge decreases caused by large precipitation decreases. In south Africa this precipitation decrease is accompanied by an evaporation increase for the DJF and MAM season. The seasonal patterns of precipitation and evaporation of the multi model mean show that during the summer (JJA) the African monsoon reaches further north which results in rainfall and discharge increases in the Northern Sahel.
In figure 5.4 the significance of change relative to the ensemble spread is shown. Significant changes are only calculated for a few regions. The large spread in discharge changes calculated for the different GCMs results in a wide uncertainty range which renders many changes as insignificant. For change in annual mean discharge the area with significant change is largest. However the regions with significant change are highly comparable for mean, minimum and maximum discharge except for south-east Asia where the region with significant increases is larger for maximum discharge.
In figure 5.5 the globe is divided in arid and humid regions based on the climate moisture indices of the WWDRII (UN, 2006). We indicated whether arid (humid) regions are expected to become wetter (drier) according to our global ensemble mean projected

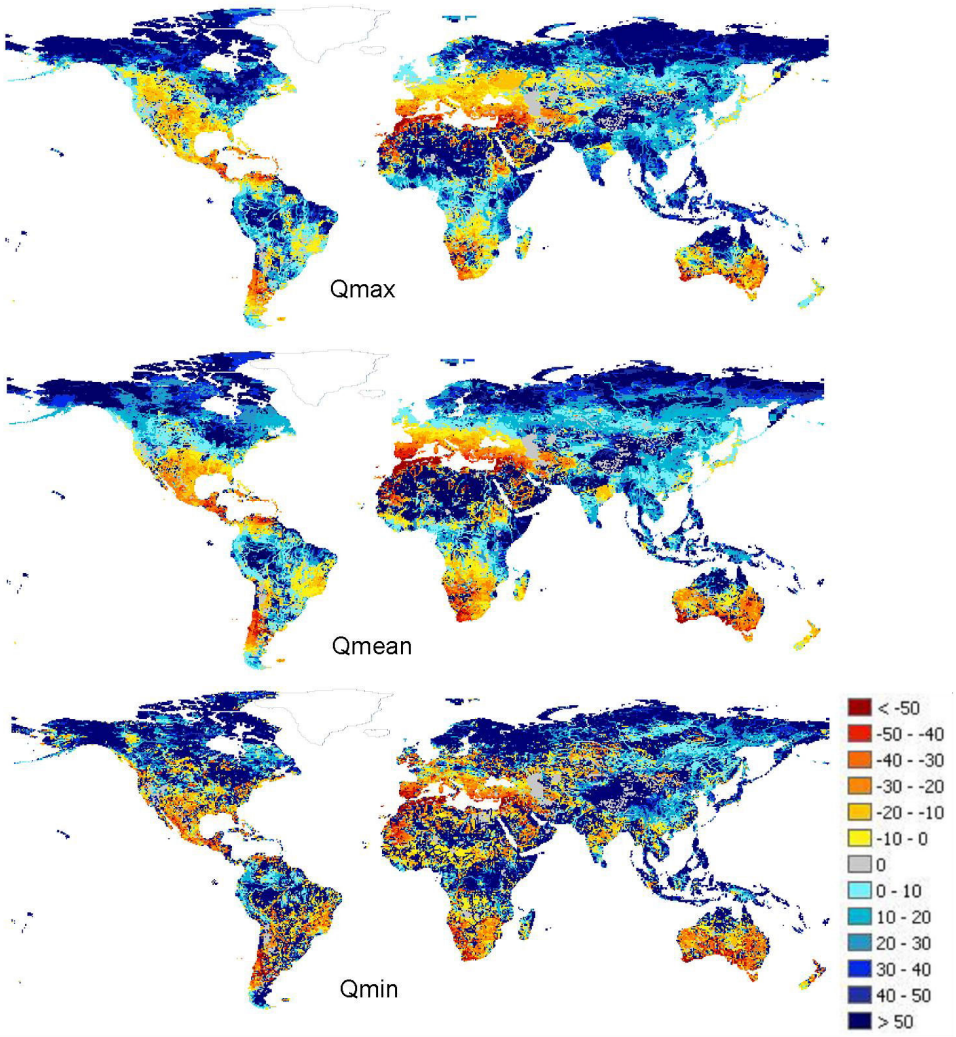


Figure 5.2 Maps showing the multi-model ensemble average percentage change (%) in the hydrological parameters annual maximum, minimum and mean discharge for the emission scenario A1B relative to the 20C3M control experiment.

changes (Fig. 5.2). For the arid regions; Southern Africa, the northern African coast, southern Australia, the southern US and Spain discharge decreases are projected. The more humid part of southern Europe will experience discharge decreases, for most other humid world regions (e.g. southeast Asia, Arctic and sub-Arctic regions, eastern US, the Amazon) discharge increases are projected. Current dry regions for which discharge increases are projected are northern Australia, parts of Asia, Russia and the central US. For northern Africa discharge increases are projected as well, however in absolute values these increases are negligible.

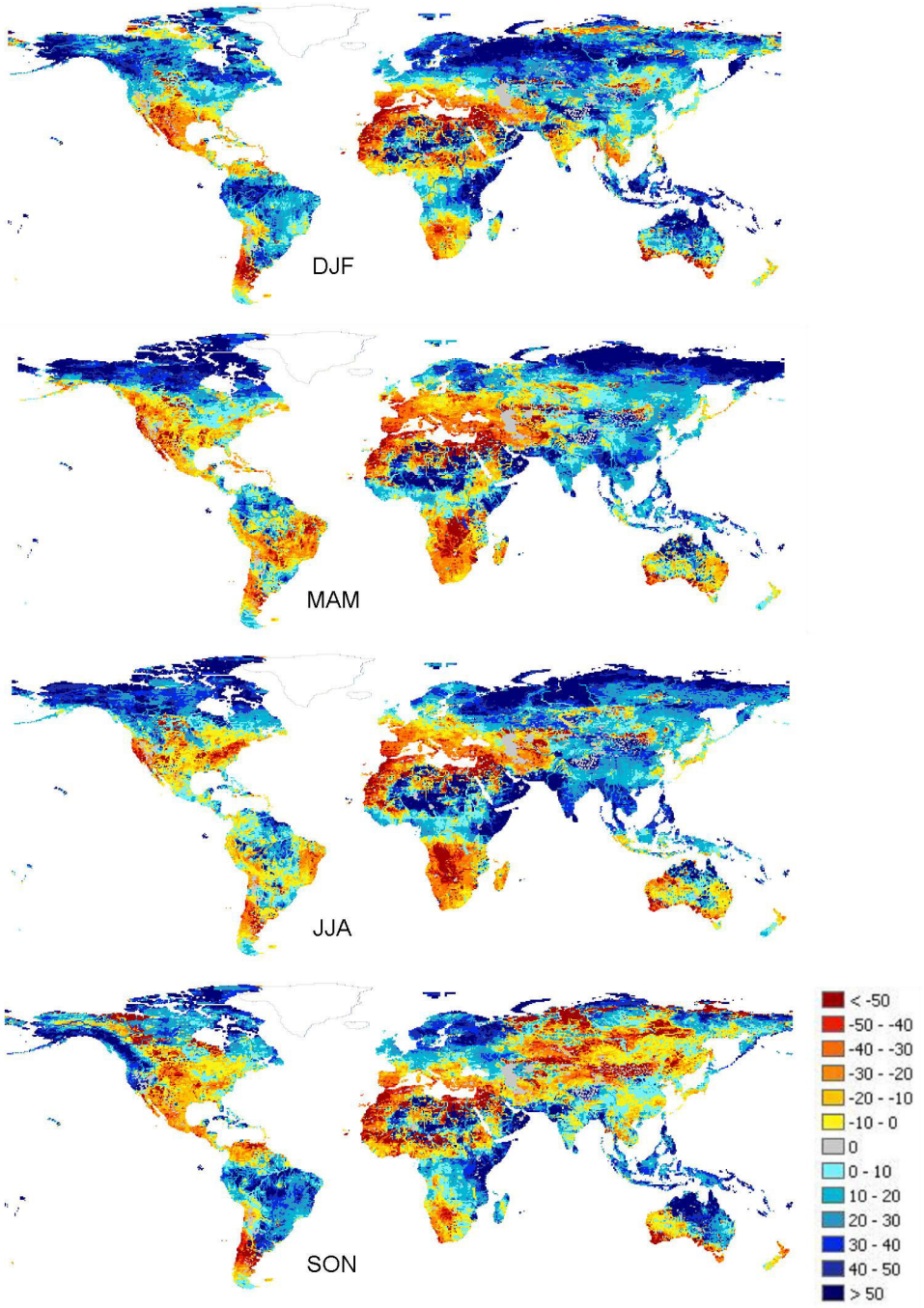


Figure 5.3 Multi-model ensemble average seasonal discharge changes for the scenario A1B as a percentage of the discharges calculated for the 20C3M control experiment. From top to bottom the seasons: December-January-February, March-April-May, June-July-August and September-October-November.

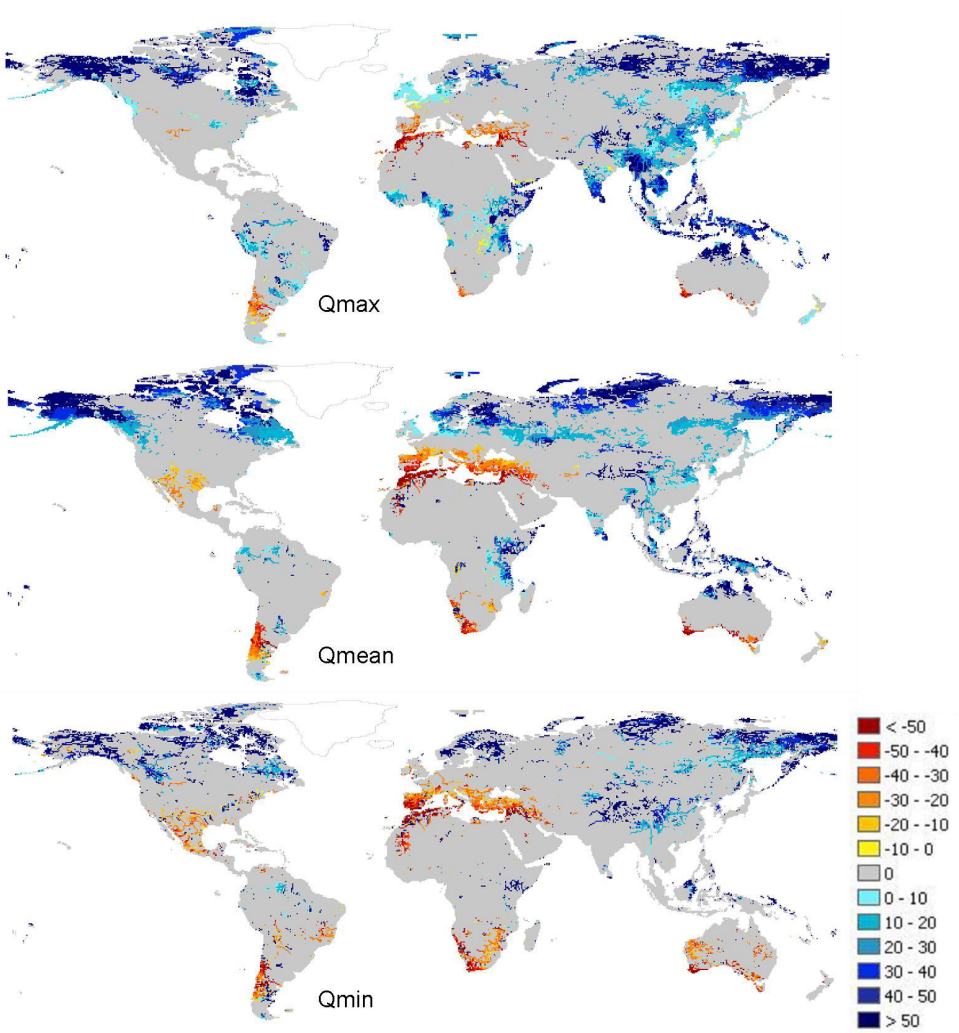


Figure 5.4 Regions with significant multi-model ensemble average change calculated relative to the GCM ensemble spread. Grey areas correspond to regions where calculated change is not significant, colored regions correspond to regions with significant change for a significance level of 95%, the values in these regions correspond to percentage change between the A1B scenario and 20C3M control experiment.

Besides change in runoff quantities, maps with shift in timing of peak discharge were calculated by taking the difference between the ensemble mode of the month of peak occurrence for the A1B scenario and the 20C3M experiment (fig. 5.6). For large parts of the world, shifts are less than a month. There is a shift backward in time for most of the sub-Arctic regions. This shift is caused by increased temperatures for the spring and summer season resulting in earlier snowmelt and more precipitation falling as rain. For parts of South-Asia a shift forward in peak timing of a half up to one month is calculated. This may result from a delay in the Monsoon rainfall that shifts from the

JJA to the SON season, caused by a later reversal of the meridional tropospheric temperature gradient (Ashfaq et al., 2009). However the plots of the annual cycles of other Monsoon influenced rivers do not show this shift. For most southern parts of the world changes are mixed. And, although shifts in timing are also displayed for deserts and tropic regions, they contain limited information since precipitation is relatively constant throughout the year in these regions and consequently the annual cycle has only a small amplitude.

5.3.2 Consistency on global patterns of change

GCM consistency maps for change in the different hydrological variables are given in figure 5.7 and 5.8. In these figures, significance is quantified for the individual GCMs relative to the GCM specific 20 year inter-annual variability. The maps show for each grid cell the number of models projecting significant change in the dominant direction. When significance is calculated relative to the ensemble spread the areas with significant change (Fig. 5.4) are smaller than when significance is obtained by calculating significance in changes projected by individual GCMs relative to their inter-annual variability.

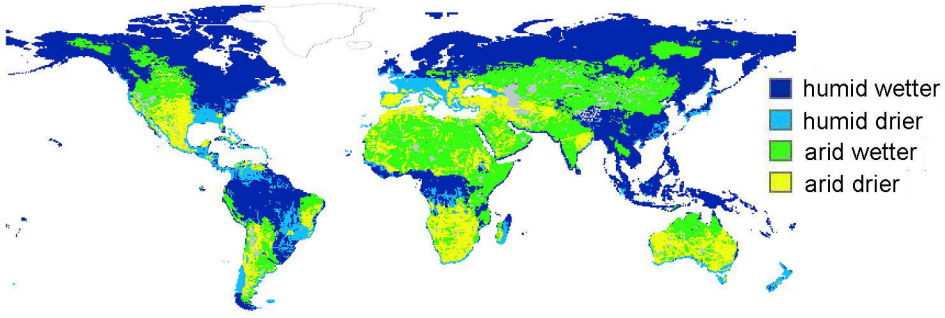


Figure 5.5 Change in aridity. The division in humid and arid regions is obtained from the WWRDII climate moisture indices (UN, 2006). The globe is divided in humid regions becoming wetter (dark blue), humid regions becoming drier (light blue), arid regions becoming wetter (green) and arid regions becoming drier (yellow) based on the ensemble average change calculated for the A1B scenario.

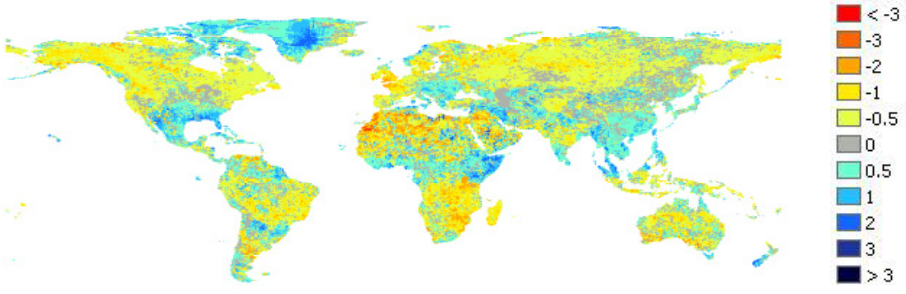


Figure 5.6 Map showing the number of months change in the timing of peak discharge occurrence calculated by taking the mode of the ensemble of timings calculated for the twelve individual GCMs for the scenario A1B relative to 20C3M control experiment.

By using this alternative analyses it is possible to denote regions with notable change, despite the uncertainty between models.
Consistent significant change amongst GCMs is especially large for increases in annual mean discharge in the Arctic regions and minimum discharge decreases in southern Australia, southern Europe, parts of Africa and the south-western coast of South-America. There is less agreement between the models on the changes in minimum, and especially maximum discharge, than on change in mean discharge. Consensus on seasonal shift of peak discharge (Fig. 5.8) is large for sub-Arctic regions where temperature rise causes an earlier snow melt driven discharge peak. For dry areas the timing of peak is difficult to assess due to low discharge values and small amplitudes, therefore models show little consensus on the direction of change in these areas.

5.3.3 Continental discharge changes

For each continent and each ocean the change in freshwater flowing into the oceans was calculated by summing the 20 year average mean accumulated runoff of rivers discharging into the oceans (Fig. 5.9). For all continents discharge to oceans increases according to the ensemble mean change. This confirms that there will be an intensification of the hydrological cycle (Huntington, 2006).
Discharge increases are smallest for Africa, Europe and South-America, as multiple GCMs also project discharge decreases for large parts of these continents. In Australia and Africa, despite the continental discharge increases, the effects of discharge decreases are large since they mainly occur in regions that are already arid at this stage (see in figure 5.5 the projected decreases in the arid regions of southern Africa and southern Australia including the Murray basin). Inflow to the oceans will increase for all oceans except the Mediterranean Sea. Inflow to the Mediterranean Sea originates from Southern Europe and Northern Africa, both regions with projected discharge decreases. Large discharge decreases for the Mediterranean region, up to 40%, have also been found by Sanchez-Gomez et al. (2009).
The spread in projected changes is smallest for Europe and South-America. Here discharge increases and decreases projected by the individual GCMs are small and the resulting ensemble mean projected change is close to zero. For the other continents ensemble mean change as well as the ensemble uncertainty is larger. For Australia and Asia a consistent discharge increase is projected and, although for Africa and North-America increases are projected as well, the ensemble mean change is smaller as some GCMs projected discharge decreases.
Globally we find an ensemble mean discharge increase of 11.0% by 2100. In contrast, Arnell (1999c) found a slight decrease for the HadCM2 ensemble (by 2080 an ensemble mean decrease of mean discharge from -0.4%), although three of the four individual ensemble members of his ensemble gave a discharge increase ranging between 0.6 and 1.0%. With HadCM3 data a decrease of -14.7% was projected. This illustrates the large differences amongst models. Vörösmarty (2000) found a global discharge decrease of

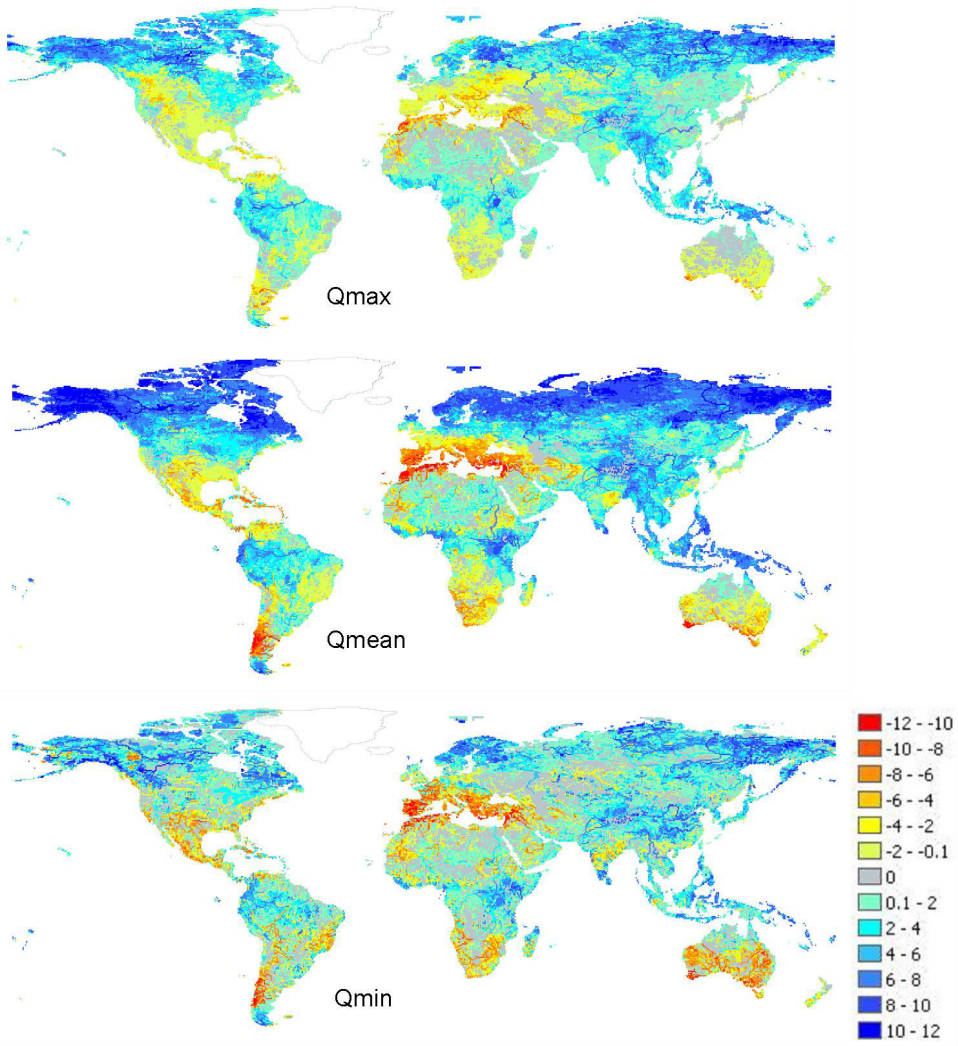


Figure 5.7 Maps showing the number of models projecting significant change (for a significance level of 95%) in the same direction as the ensemble mean direction of change (see section 2.4 for more information). From top to bottom the figure shows GCM consistencies for maximum, mean and minimum discharge. Negative values correspond to the number of models projecting discharge decrease, positive values correspond to the number of models projecting discharge increases, grey areas correspond to areas with no significant change.

-5.6% for their time horizon of 2025 and Arora and Boer (2001) found a larger decrease of -14% by the end of the 21st century. These differences might exist because they worked with the previous version of IPCC scenarios. However, more likely they are a result of the uncertainty between GCMs. Even for global average changes in temperature there is less variance amongst selected emission scenarios than amongst projections obtained from different GCMs. Depending on the selected scenario, the

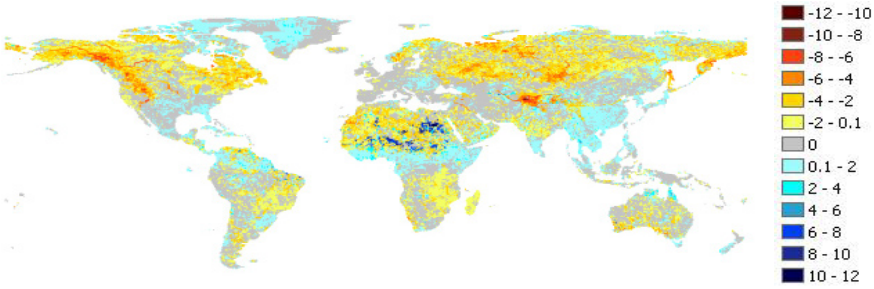


Figure 5.8 Maps showing the number of models projecting change in timing of the annual cycle consistent with the ensemble mean direction of change of timing. The negative numbers correspond to the number of models projecting advances in the annual cycle consistent with the ensemble mean advances, positive numbers correspond to consistent projected delays in the timing of regime, grey areas correspond to regions with zero change.

ensemble range of projected temperature ranges from 1.5 to 3 degrees Kelvin (K) or 2 to 4.5 degrees K. Arnell (1999c) calculated change for the 1% CO_2 increase scenario which has an average temperature increase of 2 degrees K. Arora and Boer (2001) used data for the IS92a scenario which projects an average increase of 3K. In this study we used data for the A1B IPCC SRES scenario which also projects an average increase of 3K (IPCC, 2007b). GCM ensemble uncertainty ranges for projected precipitation ranges are even larger than for temperature (see table 5.4).

5.3.4 Catchment results

Mean annual discharge cycles of the selected river basins are shown in figure 5.10 for the control experiment 20C3M, the A1B scenario and for discharge observations. Furthermore percentage changes in 20 year average minimum, maximum, mean discharge and runoff coefficient, absolute changes in timing of peak discharge, and changes in variability are shown in table 5.5, together with the significance level for which these change are significant.

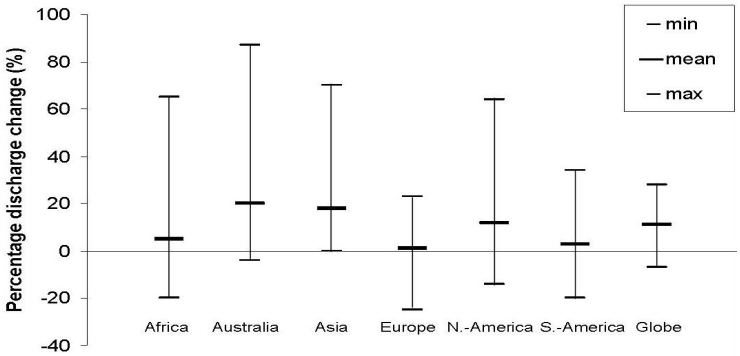
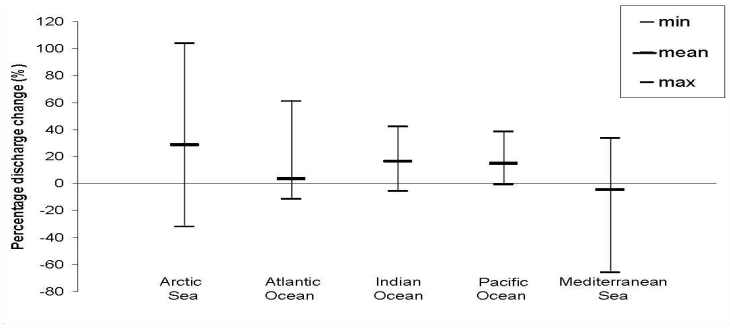


Figure 5.9 a. Continental discharge changes (%). Vertical bars represent range of changes covered by the entire ensembles of GCMs, large horizontal dashes represent ensemble mean change, small horizontal dashes represent minimum and maximum projected changes.



b.

Figure 5.9b Change in freshwater inflow to oceans (%), as figure 5.9a.

Variations between the individual GCMs are large and changes in discharge projected by individual GCMs are even contradictory for certain catchments. The projections show large discharge increases for the Yangtze, Mekong, Yellow river, Indus and Brahmaputra at high significance levels (80-95%). Furthermore, maximum discharge increases for these rivers, as well as for the Ganges, at a significance level of at least 90% (table 5.5). The changes are caused by an increase in Monsoon rainfall related to an increase of sea surface temperature which increases the moisture holding capacity of air above sea (Meehl and Arblaster, 2003). From the ensemble of GCMs an increase in precipitation is calculated for the JJA and SON seasons in this region. Nohara et al. (2006) also found an increase in discharge for the Mekong and Ganges. Arora and Boer (2001) found a slight increase in discharge for the Ganges, but a decrease in discharge for the Mekong and Yangtze. Differences are also found with the study of Nijssen et al. (2001). One of their main findings was the agreement of all models on a reduction of discharge for the Yellow River, while we found an increase of 19%. This may be a consequence of the fact that Nijssen et al. (2001) used a smaller ensemble, existing of GCMs different from our selection. Furthermore they worked with the previous set of emission scenarios (IS92a).

Table 5.4 Change in global temperature (K), precipitation (%) and discharge (%) for different emission scenarios.

Temperature	***Ensemble min***	***Ensemble mean***	***Ensemble max***	***Horizon***	***Source***
1% CO_2	1.5	2	3	2100	Andrews and Forster (2010)
IS92a	2	3	4.5	2100	IPCC (2007)
A1B	2	3	4.5	2100	IPCC (2007)
Precipitation					
1% CO_2	2	4	6	2100	Andrews and Forster (2010)
IS92a	1.5	4	6	2100	IPCC (2007)
A1B	1.5	4.5	7	2100	IPCC (2007)
Discharge					
1% CO_2	-14.7	-0.4	1	2080	Arnell (1999)
IS92a		-14		2100	Arora and Boer (2001)
A1B	-7	11	28	2100	This study

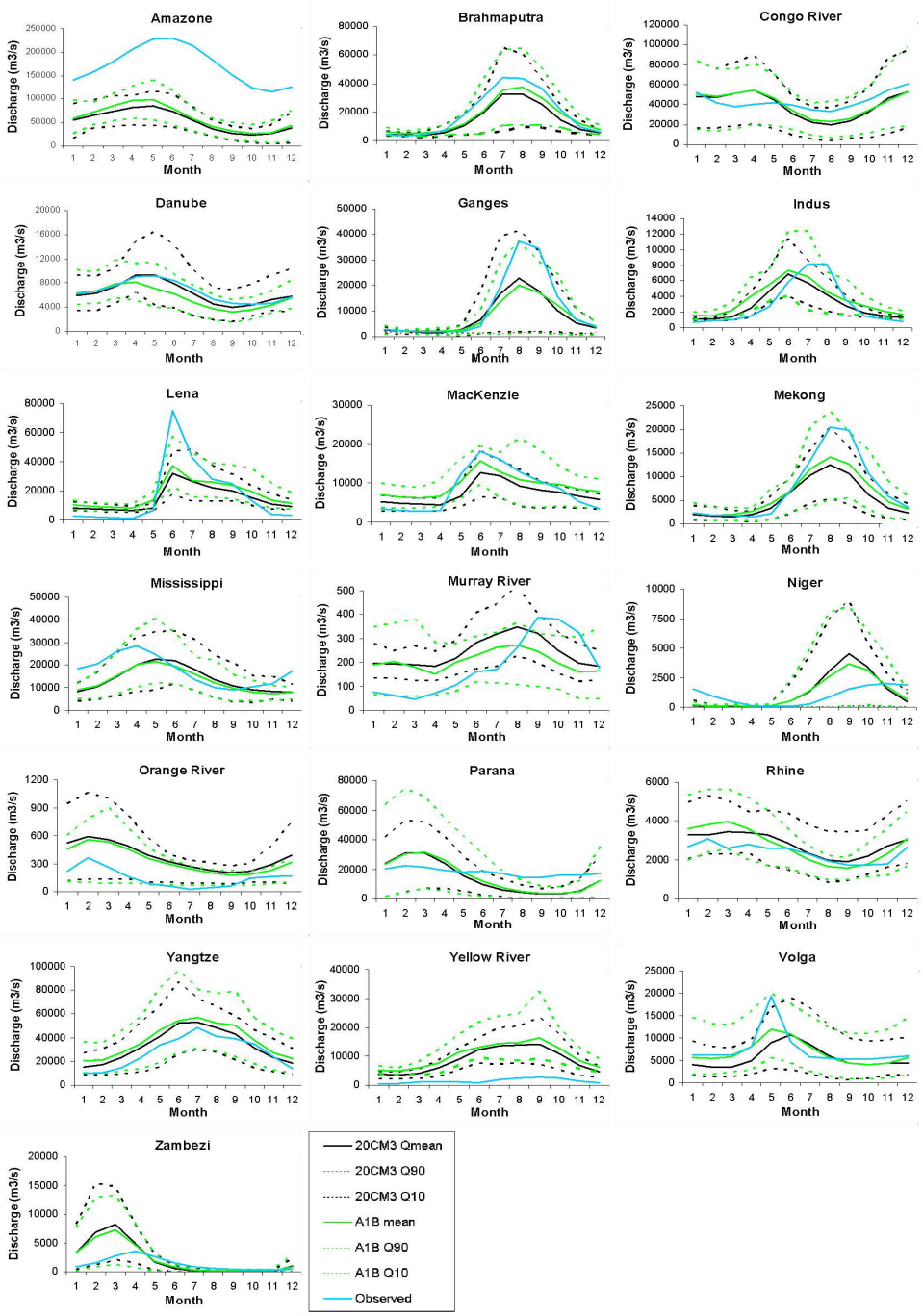


Figure 5.10 Modeled annual hydrological cycles for the 19 selected catchments, showing for each experiment the monthly 20-year average discharge and the monthly 10th and 90th percentile discharge derived from the discharges calculated from the 12 GCM datasets.

Table 5.5 Percentage change for the hydrological parameters of interest (see table 3). Calculated for the A1B experiment relative to the 20C3M experiment. When applicable the significance level (sig) for which change is significant is given as well.

	A1B		*A1B*		*A1B*		*A1B*		*A1B*		*A1B*	
	ΔQmean	**sig**	**ΔQmax**	**sig**	**ΔQmin**	**sig**	**ΔQpeak**	**sig**	**Δvar**	**sig**	**ΔRC**	**sig**
Amazon	13.2	80	21.7	95	-6.3		0.2	70	5.0	70	15.5	90
Bramaputra	14.3	80	34.7	95	10.2	70	-0.6	50	7.5	80	1.1	
Congo river	3.0		8.8	90	1.4		-0.3	70	2.1		7.1	50
Danube	-10.8	90	-15.7	80	-29.2	95	0.4	90	-1.6		-5.9	90
Ganges	-2.4		12.4	90	-28.2	70	0.4	60	3.1		-6.5	
Indus	25.0	95	28.8	95	23.4	95	-0.1		19.4	95	-5.4	95
Lena	21.4	95	47.3	95	10.9	95	-0.1		16.6	95	-5.1	50
MacKenzie	27.9	95	35.8	95	24.7	95	-0.5	95	7.1	70	-0.4	
Mekong	20.7	95	51.8	95	9.2	80	0.6	95	2.7		13.1	95
Mississippi	-4.5	50	4.6		-17.0	95	0.2		3.2	50	-10.2	90
Murray river	-14.4	90	-0.4		-23.1	90	-0.1		0.8		24.0	70
Niger	-52.5	80	19.0	80	-		-0.3		3.9		-9.9	50
Orange river	-10.4	80	5.3	50	-23.6	95	1.3	95	-4.8		4.9	
Parana	4.3		8.9	60	-27.6	90	0.3		-2.9		18.7	60
Rhine	-2.0		4.0	70	-30.6	95	-0.5	60	7.8	90	-0.8	
Volga	20.5	95	18.6	80	23.1	80	-0.1	95	13.8	95	-3.7	
Yangtze	14.1	95	23.1	95	19.1	95	0.2	70	3.7	70	0.2	
Yellow river	19.0	95	33.3	95	22.2	95	-0.1		9.6	95	-1.0	
Zambezi	-3.3		-5.9	60	-		-0.7	60	-1.9		13.9	

A decrease in mean discharge is projected for the African rivers; Zambezi, Orange and Niger. Furthermore the Zambezi shows a decrease of the 10-percentile of ensemble discharges towards no flow. Especially for the south of Africa, estimated precipitation decreases are large. The discharge decreases are in agreement with results of Arora and Boer (2001) who calculated a decrease of mean annual discharge for the warmer world and with results of Nohara et al. (2006), who found decreases for the African rivers. For the Orange river the large decrease in discharge results in a related ensemble mean decrease of inter-annual discharge variability.

For the Lena and MacKenzie a large discharge increase was estimated. This increase is related to earlier snowmelt due to higher temperatures and a calculated increase in precipitation for the SON season. The combination of both increases spring discharge for these basins. Nijssen et al. (2001) and Nohara et al. (2006) calculated an advance in peak for Arctic rivers, similar to our results. Arora and Boer (2001) also found an advances in phase for the high-latitude rivers and an increase in amplitude.

Discharge decreases are calculated for the Murray, the 10th percentile discharge even decreases to zero. Precipitation decreases and summer (JJA) evaporation increases are large for Australia. Our results for the Murray do not compare well with the results of Aerts et al. (2006). A pronounced difference was found here, they calculated a 43 % increase while we obtained a 14 % decrease for significance levels of 90 %. The difference might be caused by the difference in reference period used. This was 1750 to 2000 in their study. Within a longer reference period the interannual variability is likely to be larger and periods with relatively low discharge may have occurred before. Derived changes can therefore be smaller than the changes we derived from our 20-year reference period.

The discharge of the Danube significantly decreases at a level of 90%. Precipitation decreases are large in both the Danube and the Rhine basin in particular for the summer period. Annual average discharge decreases for the Rhine, yet there is an increase in maximum discharge in this basin. This is related to the calculated temperature increases, leading to earlier snowmelt and an increased amount of spring precipitation falling as rain instead of snow. Consequently, precipitation enters the river earlier in the year. The study of Nohara et al. (2006) showed a decrease in discharge for the river Danube and Rhine as well. However, again there is a difference with Aerts et al. (2006) who found little change in Danube discharge.

5.3.5 Change in catchment specific runoff coefficients

To quantify the relative change in water balance partitioning due to climate change, the change in runoff coefficients (RC) has been calculated for all individual catchments for the selected measurement stations (Fig. 5.1). For the calculation of the RC, twenty year average year sums of accumulated upstream precipitation and actual evaporation have been used to avoid the influence of storage changes in glaciers and soil water. Basins with an increase of RC of more than 10% are the Amazon, Parana, Murray, Zambezi, Mississippi and Mekong. For the Mekong this increase is caused by an increase in precipitation. The Parana also experiences an increase in precipitation together with a decrease of actual evaporation. For the Murray and Zambezi the decrease in actual evaporation is larger than the decrease in precipitation, also resulting in increasing RC. Except for the Niger, the African rivers all have an increasing runoff coefficient, indicating that the part of precipitation that evaporates decreases. The Niger is the only basin with a decrease in RC of more than 10%. This decrease is caused by small changes in precipitation and large evaporation increases. The runoff coefficient for the Danube and Rhine decrease, due to decreasing precipitation and increasing evaporation as a result of temperature increases.

5.4 Synthesis

Multiple studies investigating the hydrological consequences of climate change on a global scale have been conducted. In an attempt to make an as complete as possible assessment of the global hydrological effects of climate change we provided and overview of previous hydrological studies and presented our results in the context of the previous results. We used a, for hydrological studies, relatively large ensemble of GCMs existing of all the GCMs for which the PCMDI data portal provided the required daily time-series of meteorological variables which were needed as input to the hydrological model. We estimated changes in spatial and temporal discharge variability and calculated the ensemble consistency of the projected changes. In addition to previous studies we quantified significance of change relative to both the individual GCM inter-annual variability and the inter-model variability of the ensemble of GCMs. By using the alternative analyses of calculating significance relative to the GCMs inter-annual variability,

it is possible to denote regions with notable change, despite the uncertainty between models.

Result show that river discharge will increase for the Yangtze, Yellow river, Mekong, Ganges, Indus and Brahmaputra due to an increase in monsoon rainfall. As a result of earlier snowmelt and an increase of precipitation. the Lena and MacKenzie show an increase in spring discharge and a small shift in timing of peak. A decrease in both mean and extreme discharge is projected for the Orange, Niger, Murray and Danube. Comparable results have been found in previous studies especially when looking at global patterns of change, but differences exist both on catchment and continental scale.

The climate models do not always project consistent changes. Opposite directions of change are mainly projected for minimum and maximum discharge in regions with temperate climate. When calculated relative to the ensemble spread the significance is often small (Fig. 5.4). This is caused by the large range of projected discharge changes that is obtained when the results of the individual climate models are combined.

Changes in the downstream part of the river basins, especially in the main river courses, are often significant for higher significance levels than changes for grid cells located upstream in the catchment. This may be caused by accumulation of variations in climate patterns while moving downstream. It confirms the importance of discharge accumulation and the use of a routing scheme that, although biases are present for several catchments, allows for temporal storages in lakes and reservoirs and introduces realistic travel times which are especially relevant in larger catchments like the Amazon (Sperna Weiland et al., 2010). We expected the larger rivers basins, or basins with higher discharges, to have significant change for a higher significance level. However, this relation could not be found, probably because some of the larger rivers travel through multiple climate zones.

Significant changes were found for larger regions at higher significance levels for the individual GCMs when significance of change was calculated relative to the GCM specific 20 year natural variability. In addition to the information on the discharge change maps, the consistency maps (Fig. 5.7) indicate the agreement amongst models on the direction of significant change in relation to inter-annual variability and thereby give an indication of regions where discharge is likely to be affected by climate change. Such an analysis partly accounts for the influence of GCM model errors and may be the preferred change detection method for grid-based global assessment of discharge change.

According to the ensemble mean calculations, continental outflow to oceans will increase for all oceans except for the Mediterranean sea. The GCMs project a consistent decrease in runoff for southern Europe, South Australia, South Africa, parts of north Africa and the southwestern coast of South-America. There is also large consensus on discharge increase for the Arctic regions and the Northern Sahel. Besides these results, the following three findings are useful to hydrological climate effect studies in general: We found that the projected changes in our study show the largest differences with studies based on a small number of climate models. When using only small ensembles

the response may be biased through the influence of only one or two GCMs that deviate from the other models, while in larger ensemble these deviating GCMs will have less influence due to the averaging of multiple change projections. This underscores the value of using large ensembles. Secondly, from the differences with the study of Aerts et al. (2006) it can be concluded that choice of the reference period influences the change signal. Aerts used data for the period 1750 to 2000 as a reference for the change projections and to investigate the influence of interdecadal variability. When using a reference period of this length the influence of inter-annual variability is minimized, whereas in our twenty year period it is more likely that the average discharge is disturbed by effects like El Nino. Furthermore, in this study change is calculated between 2100 and the time-slice 1961-1990, which is likely to represent current climate conditions. Whereas Aerts et al. (2006) calculated change relative to the period 1750-2000 and changes will therefore either be relatively large or less extreme due to long-term variations in the climate that resemble future climate changes. And finally, our results are comparable to studies using the change factor method which, for computational reasons, might therefore be the preferable method to use.

5.5 Conclusions

In this paper we successfully revealed the regions of the world where significant and consistent hydrological changes are projected. By 2100 hydrological model runs, based on an ensemble of GCMs, project a consistent decrease in runoff for southern Europe, southern Australia, the south and north of Africa and southwestern South-America. Significant discharge decreases are also projected for most African rivers, for the Murray and for the Danube. Runoff increases are projected for sub-Arctic and Arctic regions and an advance in phase in the annual cycle is projected for the sub-Arctic regions. Overall, discharge of Monsoon influenced rivers slightly increases.

The results of this study are generally comparable to previous studies. Although, results of studies using only a small number of GCMs show relatively large differences from our study and the use of a multi-model ensemble is therefore preferable. We illustrated that by considering the consistency of change amongst models, i.e. in light of significance of projected change relative to natural variability, the regions with a high likelihood of changes in the annual cycle are more clearly revealed than when significance of change is calculated relative to inter-model variability.

Acknowledgements

We acknowledge the GCM modeling groups, the Program for Climate Model Diagnosis and Intercomparison (PCMDI) and the WCRP's Working Group on Coupled Modelling (WGCM) for their roles in making available the WCRP CMIP3 multi-model dataset. Support of this dataset is provided by the Office of Science, U.S. Department of Energy.

6 Extracting information from an ensemble of GCMs to reliably assess future global runoff change

This chapter is based on:
Sperna Weiland, F. C., L. P. H. van Beek, A. H. Weerts and M. F. P. Bierkens (2011), Extracting information from an ensemble of GCMs to reliably assess future global runoff change. Journal of Hydrology, doi:10.1016/j.jhydrol.2011.03.047, in press.

Abstract

Future runoff projections derived from different global climate models (GCMs) show large differences. Therefore, within this study, the information from multiple GCMs has been combined to better assess hydrological changes. For projections of precipitation and temperature the Reliability Ensemble Averaging (REA) method was introduced to calculate weighted average ensemble change with an accompanying uncertainty range. In this study the original REA method is compared with three other methods that calculate runoff change by selecting or weighting GCMs on their inter-model similarity for the current and future climate. All methods are applied to distributed runoff fields calculated with the hydrological model PCR-GLOBWB forced with meteorological data from an ensemble of 12 GCMs.
Differences between weighted and non-weighted average runoff changes for 2100 are small. Within a validation experiment, where GCM ensemble mean change is derived between the time slices 1961-1975 and 1976-1990, the non-weighted ensemble change resembled observed change best. Yet, both the weighted and non-weighted average change were too conservative. This underscores the importance of considering an uncertainty range alongside the ensemble average change. In this study the uncertainty range (or 95% confidence interval) is defined by four times the root mean square difference around the ensemble mean change. The uncertainty range derived with the non-weighted method is relatively wide, upper and lower uncertainty bounds show large

biases from observed change. However, the uncertainty range was reliably reduced by using only a selection of GCMs which show high inter-model similarity for the current and future climate.

6.1 Introduction

Global climate model (GCM) and regional climate model (RCM) projections are frequently used to quantify hydrological impacts of climate change (Christensen and Lettenmaier, 2007; Prudhomme and Davies, 2008; Arnell, 2003). Unfortunately large discrepancies exist between GCM datasets (Varis et al., 2004), especially for precipitation (Dai, 2006) the meteorological variable of utmost importance in hydrological processes. It has been recognized that the information of an ensemble of GCMs should be combined to obtain reliable projections (Murphy et al. 2004; Räisänen, 2007). Ideally, by using the average of an ensemble of GCMs, the individual model errors are cancelled out and the ensemble uncertainty decreases as increasingly more models are used. In reality, the models will not be completely independent as some biases are present in all models due to the use of similar parameterization, numerical methods and coarse grid resolutions. Therefore, biases will remain in the multi-model ensemble and the ensemble might not cover the full uncertainty range (Tebaldi and Knutti, 2007; Räisänen, 2007). Still multiple studies concluded that a multi-model mean tends to give better results than single model simulations (Giorgi and Mearns, 2002; Murphy et al., 2004; Boorman and Sefton, 1997; IPCC, 2007b; Sperna Weiland et al., 2010; Räisänen, 2007). Although often the non-weighted ensemble mean hydrological change is calculated, results of several studies showed that more reliable results are obtained by using projections of a cluster of better performing models (Smith and Chandler, 2010) or calculating a weighted ensemble average, where the individual GCM weights are derived from model performance and future ensemble convergence (Giorgi and Mearns, 2002; Min and Hense, 2007; Räisänen et al., 2009; Murphy et al., 2004). For the ensemble weighted averaging of future GCM temperature and precipitation projections, two reliability criteria are frequently used (Giorgi and Mearns, 2002; Tebaldi et al., 2004). The first criterion, model performance, is based on the ability of the GCM to reproduce characteristics of the current climate. The second criterion, the convergence criterion, is based on the convergence of future change projections of different GCMs. Using the convergence criterion in addition to the performance criterion is likely to result in a better weighting, since a correct reproduction of the current climate is no guarantee that the projected climate is correct as well (Räisänen, 2007; Tebaldi and Knutti, 2007; Reifen and Toumi, 2009; Prudhomme and Davies, 2008; Chiew et al., 2009). Räisänen et al. (2009) propose the idea of favoring models that show a high inter-model similarity in observable and future climate, since those models are most likely to perform well because of a realistic representation of the climatic system. Furthermore, down-weighting a model because its projected change is different from the weighted ensemble mean change involves the

risk of circular reasoning.
Deriving model weights from performance for precipitation and or temperature may not be the best solution in hydrological impact studies. Model weights for precipitation and temperature are assigned in individual procedures and in subsequent steps the different weights need to be translated into one reliability factor. It can be questioned whether GCM performance for temperature and precipitation are equally important and it is a subjective decision to define the ratio in which the weights should be combined (Tebaldi and Knutti, 2007). In addition, in a hydrological model, GCM precipitation, evaporation and temperature are used to derive runoff. The resulting runoff can be influenced by additional hydrological model errors which may overrule the importance of GCM performance for precipitation and temperature. Furthermore, the uncertainty in the derived runoff will differ from the uncertainty in precipitation and temperature. And finally, by weighting on runoff, weighting on parameters that have been used in the tuning process of the GCMs is avoided (Räisänen, 2007).
In the first step of this study, four GCM weighting and selection methods for ensemble averaging of runoff projections are evaluated. In a second step the better performing methods are used to calculate global runoff change by 2100. Global distributed runoff fields are calculated with the water balance model PCR-GLOBWB (Van Beek and Bierkens, 2009; Sperna Weiland et al., 2010) forced with meteorological data from an ensemble of 12 GCMs for both the current climate experiment (20C3M) and the future emission scenario (A1B). Reliability factors are calculated by comparison of GCM derived runoff fields with runoff fields calculated from the Global Runoff Data Centre (GRDC) observed discharges (GRDC, 2007, Fekete et al., 2002).
For the validation of the four methods a 30-year period of the 20C3M experiment (1961 to 1990) is divided into two time-slices of 15 years (1961-1975 and 1976-1990), see figure 6.1a. GCM ensemble average runoff changes are calculated between those time-slices and the changes resulting from the different methods are validated against observed runoff change and non-weighted ensemble average change. For the assessment of future global runoff projections, runoff change is calculated between two 20-year time-slices 1971 to 1990 (= 20C3M experiment) and 2081 to 2100 (= A1B scenario), see figure 6.1b. The following methods (listed in Table 6.1) are investigated: 1) the REA method as proposed by Giorgi and Mearns (2002), 2) the REA method extended with a reliability criteria for the reproduction of the observed trend (Räisänen,

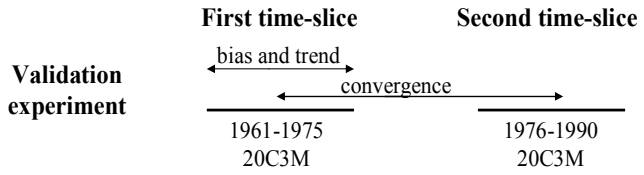


Figure 6.1a Schematic representation of the validation experiment.

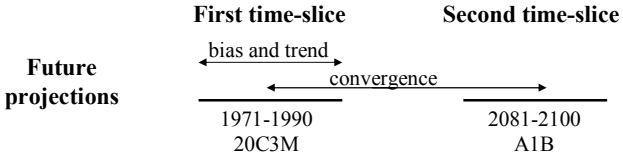


Figure 6.1b Schematic representation of the derivation of future change.

2007), 3) the REA method where convergence is replaced by inter-model similarity (Räisänen et al., 2009) and 4) a method selecting only those models showing a high inter-model similarity for the observed and future climate (Räisänen et al., 2009; Smith and Chandler, 2010).

Table 6.1 Overview of selecting and weighting methods.

Method	*Description*	*Reference*
Original REA: bias and convergence	The original REA method with reliability criteria for bias within the current climate and convergence / distance to ensemble average future change	Giorgi and Mearns, 2002
REA: bias, convergence and trend	The REA method with reliability criteria for bias within the current climate, future convergence / distance and an additional criterion for reproduction of past trend	Räisänen, 2007
REA: bias, inter-model similarity	The REA method with the reliability criterion for future convergence / distance replaced by future inter-model similarity	Räisänen et al., 2009
Selection of better performing models	A selection of models containing only those models having a small bias for the past climate and a high future inter-model similarity	Räisänen et al., 2009; Smith and Chandler, 2010

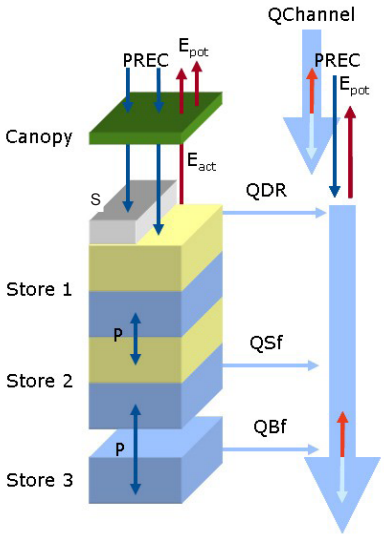


Figure 6.2 PCR-GLOBWB model outline.

6.2 Methodology

6.2.1 Global water balance model PCR-GLOBWB

PCR-GLOBWB is a global distributed hydrological model with a resolution of 0.5 degrees (Van Beek and Bierkens, 2009; Bierkens and van Beek, 2009; Sperna Weiland et al., 2010). A flow-chart of the model is given in Fig. 6.2. Each model cell consists of two vertical soil layers and one underlying groundwater reservoir. Sub-grid parameterization is used for the schematization of surface water, short and tall vegetation and for calculation of saturated areas for surface runoff as well as interflow. Water enters the cell as rainfall and can be stored as canopy interception or snow. Snow is accumulated when temperature is below 0°C and melts when temperature is higher. Melt water and throughfall are passed to the surface, where they either infiltrate in the soil or become surface runoff. Exchange of soil water is possible between the soil and groundwater layers in both downward and upward direction depending on soil moisture status and groundwater storage. Total runoff consists of non-infiltrating melt water, saturation excess surface runoff, interflow and base flow. For each daily time-step the water balance and resulting runoff is computed for all model cells. The model calculates natural runoff fluxes. Water use and management are not included.

Table 6.2 Overview of selected GCMs.

Model	*Institute*	*Country*	*Acronym*
BCM2.0	Bjerknes Centre for Climate Research	Norway	BCCR
CGCM3.1	Canadian Centre for Climate Modelling and Analysis	Canada	CCCMA
CGCM2.3.2	Meteorological Research Institute	Japan	CGCM
CSIRO-Mk3.0	Commonwealth Scientific and Industrial Research Organization	Australia	CSIRO
ECHAM5	Max Planck Institute	Germany	ECHAM
ECHO-G	Freie Universität Berlin	Germany	ECHO
GFDL-CM 2.1	Geophysical Fluid Dynamics Centre	USA	GFDL
GISS-ER	Goddard institute for Space Studies	USA	GISS
IPSL-CM4	Institute Pierre Simon Laplace	France	IPSL
MIROC3.2	Center of Climate System Research	Japan	MIROC
CCSM3	National Center for Atmospheric Research	USA	NCAR
HadGEM1	Met Office's Hadley Centre for Climate Prediction	UK	HADGEM

6.2.2 Data

GCM data

Data from a selection of 12 GCMs (listed in Table 6.2) were downloaded from the data portal of the Program for Climate Model Diagnosis and Intercomparison (PCMDI; https://esg.llnl.gov:8443/index.jsp) and from the CERA-gateway (http://cera-www.dkrz.de; data for HadGEM2). The selection of GCMs was based on data availability for the IPCC SRES emission scenario A1B and the control experiment 20C3M. In the

20C3M control experiment CO_2 forcing is prescribed by observed CO_2 concentration increases (IPCC, 2007b). For those GCMs with multiple runs available the first run was selected. Reference potential evaporation was calculated using either the equation of Penman-Monteith (Monteith, 1965; Sperna Weiland et al. 2010) or Blaney-Criddle (Brouwer and Heibloem, 1986) depending on data availability.

Observed runoff fields

Gridded GRDC runoff fields with a resolution of 0.5 degrees were derived from observed GRDC inter-station runoff fields provided by Fekete et al. (2002). These fields contain the area-average runoff becoming additionally available in the area between the selected GRDC stations (inter-station discharge):

$$\overline{R}_{oi} = \frac{\overline{Q}_{oi}}{A_{si}} \tag{6.1}$$

Where $\overline{Q}_{oi}$ is the mean annual observed inter-station discharge (m^3/year), A_{si} is the inter-station area (m^2) and $\overline{R}_{oi}$ is the resulting mean annual observed inter-station runoff (m/year).

In a first step, the inter-station runoff was redistributed uniformly over the inter-station area. Afterwards, the inter-station discharge was down-scaled to cell values. Since observational information on spatial runoff distribution was lacking, downscaling was based on the spatial runoff distribution obtained from a global distributed hydrological model. For this spatial down-scaling Fekete at al. (2002) used runoff fields calculated with the global water balance model WBM (Vörösmarty et al., 1998). Yet, in this study, runoff fields calculated with PCR-GLOBWB are used. Hereby avoiding the introduction of additional biases by the use of a water balance model different from the model used to derive runoff from GCM data.

To generate realistic spatial distributed runoff fields, PCR-GLOBWB was run with a historical meteorological dataset that consisted of precipitation, temperature, and reference crop potential evaporation. This dataset was created by downscaling the monthly CRU TS 2.1 (New et al., 2000) and CRU CLIM 1.0 datasets (New et al., 1999) to daily fields with the ERA40 reanalysis data (Uppala et al., 2005; see also: Van Beek, 2008; Sperna Weiland et al., 2010).

The runoff correction coefficient (ε_{SI}), used to equalize the PCR-GLOBWB total inter-station runoff to the observed inter-station runoff, was calculated from the ratio between average modelled ($\overline{R}_{mi}$) and observed ($\overline{R}_{oi}$) inter-station cell runoff:

$$\varepsilon_{SI} = \frac{\overline{R}_{oi}}{\overline{R}_{mi}} \tag{6.2}$$

The resulting corrected cell specific runoff (R_c) was derived by multiplying the modeled

cell specific runoff (R_m) with the runoff correction coefficient:

$$R_c = \varepsilon_{SI} R_m \tag{6.3}$$

6.2.3 GCM weighting and selection methods

REA based on bias and convergence for mean runoff

For each grid cell of the 0.5 degrees resolution runoff fields, the weighted ensemble average was calculated based on the Reliability Ensemble Averaging method proposed by Giorgi and Mearns (2002). Within this REA-method, performance based reliability factors are derived from current climate performance and future ensemble convergence for all individual GCMs and subsequently these reliability factors are used to calculate a weighted average. The performance and convergence criteria were applied to annual mean runoff fields where the cell specific normalized reliability factor for mean runoff, $R_{\bar{Q},m}$, of the *m*-th GCM is given by:

$$R_{\bar{Q},m} = \left[{R_{B,m}}^{c} \cdot {R_{D,m}}^{d} \right]^{1/(c \cdot d)} = \left\{ \left[\frac{\varepsilon_Q}{abs(B_{Q,m})} \right]^{c} \left[\frac{\varepsilon_Q}{abs(D_{Q,m})} \right]^{d} \right\}^{[1/(c \cdot d)]} \tag{6.4}$$

The overall GCM reliability factor, $R_{\bar{Q},m}$, is derived from the product of two independent reliability factors: $R_{B,m}$, which quantifies the models reliability based on the model bias for the current climate, and $R_{D,m}$, which is the normalized reliability factor which measures the GCM reliability in terms of the convergence of the GCM specific future runoff change with the weighted ensemble mean future change. $R_{B,m}$ is a function of $B_{Q,m}$, the bias of the *m*-th GCM derived runoff from the gridded observed runoff, the higher the bias the lower the reliability factor. $R_{D,m}$ is a function of $D_{Q,m}$, the distance of the runoff change calculated by the *m*-th GCM from the REA ensemble average change. High reliability factors are only obtained when both the bias for the current climate and the distance to ensemble weighted future change are small.

ε_Q is a measure of natural variability in either 15- (for the validation experiment) or 20-year (for the estimation of future change) average cell specific runoff. To calculate ε_Q, 15 or 20 year moving averages were derived from the time series of gridded runoff fields (section 6.2.2) after linearly detrending the data to remove trends in observed runoff change. ε_Q is equal to the difference between the minimum and maximum calculated moving average. When the bias ($B_{Q,m}$) or the distance to the weighted ensemble mean change ($D_{Q,m}$) is smaller than the natural variability, the model reliability factor for bias or distance is set to 1. The exponents *c* and *d* in Eq. 6.4 can be used to assign weights to the different criteria, however assignment of weights is mainly a subjective decision,

therefore the values of both c and d are set to 1.
The convergence or distance, $D_{Q,m}$, was calculated in an iterative process, where in a first step the distance of the individual GCM change, ΔQ_m, was calculated relative to the non-weighted ensemble mean change. From these distances reliability factors were derived and a first weighted average change could be calculated. In subsequent steps the deviations from the weighted average were used to derive new reliability factors until this process converged. The REA runoff change and the future REA runoff are defined by:

$$\widetilde{\Delta Q} = \frac{\sum_{m}^{M} R_m \Delta Q_m}{\sum_{m}^{M} R_m} \qquad (6.5) \qquad\qquad \widetilde{Q_{fut}} = \frac{\sum_{m}^{M} R_m Q_{fut,m}}{\sum_{m}^{M} R_m} \qquad (6.6)$$

Where $\widetilde{\Delta Q}$ corresponds to the ensemble average REA runoff change and $\widetilde{Q_{fut}}$ corresponds to the future ensemble REA runoff.

REA with an additional criteria for reproduction of observed trend
The ability of a GCM to reproduce the observed trend in climate change induced runoff change could be a good measure of its reliability as well (Räisänen, 2007). Only for a short period (1961 to 1990) both discharge observations and modelled runoff (20C3M experiment) were available to calculate modeled and observed trend in runoff change. Reliability factors for the bias of the GCM specific modeled runoff trend from the observed runoff trend have been derived for this period.

$$R_{BT,m} = \left[\frac{\varepsilon_Q}{abs(BT_{Q,m})} \right] \qquad (6.7)$$

Here $BT_{Q,m}$ is the bias of the modelled trend, of the mth GCM, from the trend obtained from discharge observations for the either the first 15-year period of the validation experiment (1961 to 1975) or the full historical period (1971 to 1990) for the future projections. The resulting overall normalized reliability factor is defined as:

$$R_m = R_{\overline{Q},B_m} \cdot R_{\overline{Q},D_m} \cdot R_{BT_m} \qquad (6.8)$$

REA with convergence replaced by future inter-model distance
Räisänen (2007) questioned the validity of down-weighting a model for which the projected change is different form the weighted ensemble mean, because this involves a risk of circular reasoning. He proposes to estimate future GCM performance from the

inter-model distances of the individual GCM changes, by doing so only a model that shows deviations from the majority of models is down-weighted. The total inter-model distance (*IMD*) of the *m*th GCM is given by the sum of absolute difference between its future derived mean runoff estimate (Q_m) and the mean future runoff estimates derived from all N other GCMs (Q_k).

$$IMD_m = \sum_{k=1}^{K} abs(Q_m - Q_k) \tag{6.9}$$

The reliability factor for this criteria is given by:

$$R_{IMD,m} = \left[\frac{\varepsilon_Q}{IMD_{Q,m}} \right] \tag{6.10}$$

Selection of models showing good performance for past and future

Performance based weighting is frequently criticized (Räisänen , 2007; Smith and Chandler, 2010). Therefore, we also evaluated a selection method proposed by Räisänen et al. (2009) where only those models having a large inter-model similarity in observed climate and a large inter-model similarity in simulated climate change are used. This method was tested by selecting the models with high reliability factors for both future inter-model distance and past performance. The definition of a threshold value above which models should be included in the selection, is a subjective decision. Therefore, two sets of models were created. The first set existed of the models with reliability factors (RFs) above or equal to the median RF value for both criteria. In the second set all models having RFs above or equal to the lower quartile RF value for both criteria were included.

6.2.4 Uncertainty bounds

In order to quantify the uncertainty of the projected runoff change, the uncertainty range was calculated from the root mean square difference (rmsd) of projected changes. For the REA method Giorgi and Mearns (2002) introduced the REA rmsd, $\tilde{\delta}_{\Delta Q}$, according to:

$$\tilde{\delta}_{\Delta Q} = \left[\frac{\sum_{m=1}^{M} R_m (\Delta Q_m - \widetilde{\Delta Q})^2}{\sum_{m=1}^{M} R_m} \right]^{1/2} \tag{6.11}$$

The lower and upper uncertainty bounds of the 95% prediction interval can be defined as:

$$\Delta Q_{+} = \widetilde{\Delta Q} + 2\tilde{\delta}_{\Delta Q} \qquad (6.12) \qquad \Delta Q_{-} = \widetilde{\Delta Q} - 2\tilde{\delta}_{\Delta Q} \qquad (6.13)$$

Here $\tilde{\delta}_{\Delta Q}$ can be replaced by the normal rmsd for the other methods.

6.3 Results

6.3.1 Validation of methods for the period 1961-1990

For the validation of the different methods the 30-year period (1961 to 1990) of the 20C3M experiment was divided into two time-slices of 15 years. Modeled change in runoff, calculated between the time-slices 1961-1975 and 1976-1990, was compared with change calculated from observed runoff fields. Ensemble average changes (Fig. 6.3b) are smaller than GRDC observed changes (Fig. 6.3a), although within the 20C3M experiment GCM greenhouse gas concentrations follow observed concentration increases. The fact that the ensemble average change is smaller than observed change, is a result of averaging of opposite changes, see for example the difference in change projected by individual GCMs for Southern Asia, Eastern North-America and Europe (Fig. 6.3c, d and e). The conservativeness of the ensemble mean change and the large variation between individual GCMs, which decreases the reliability of the non-weighted ensemble average change, indicate that an uncertainty range around the ensemble mean change should be considered.

On the right side of Fig. 6.3 projected changes for the validation period are given for all weighting and selection methods introduced in section 6.2.3. Changes calculated with the different methods show only slight deviations from the non-weighted average change. Yet, changes are less pronounced when only a selection of better performing models is used (Fig. 6.3i and 6.3j). The non-weighted ensemble projects a slightly larger area with runoff increases for the North-Western coast of South-America, while runoff decreases have been observed here (Fig. 6.3a). The non-weighted change also deviates from the other methods for Norway, this deviation is related to strong increases in westerly winds along the Scandinavian coast, present in only some of the GCMs (Räisänen, 2007). The resulting runoff increases are more pronounced in the alternative averaging methods.

The smallest bias from average observed change is obtained with the non-weighted ensemble mean (Table 6.3). However, differences in biases between methods are small compared to the absolute bias from observed change. This absolute bias is large because average modeled change deviates from observed runoff change, amongst others due to averaging of opposite changes. The original REA method and the REA method extended with the criteria for reproduction of observed trend show the largest biases from observed change. Yet, both methods show the smallest biases from observed absolute runoff.

Biases from observed runoff can be too high for all methods due to the absence of

water use in the model. The bias from observed runoff is largest for the non-weighted average and although the bias decreased with all selection and weighting methods, the selection of models with RF factors above the median showed smallest improvements. Table 6.4 shows the percentage of global map cells where the observed change lies within the uncertainty range. Unfortunately, the areas where observed change is covered by the uncertainty range are mainly regions with less pronounced observed changes and large uncertainties between models. However, some areas do correspond to regions with notable observed runoff decreases as for example Morocco, southeast Australia,

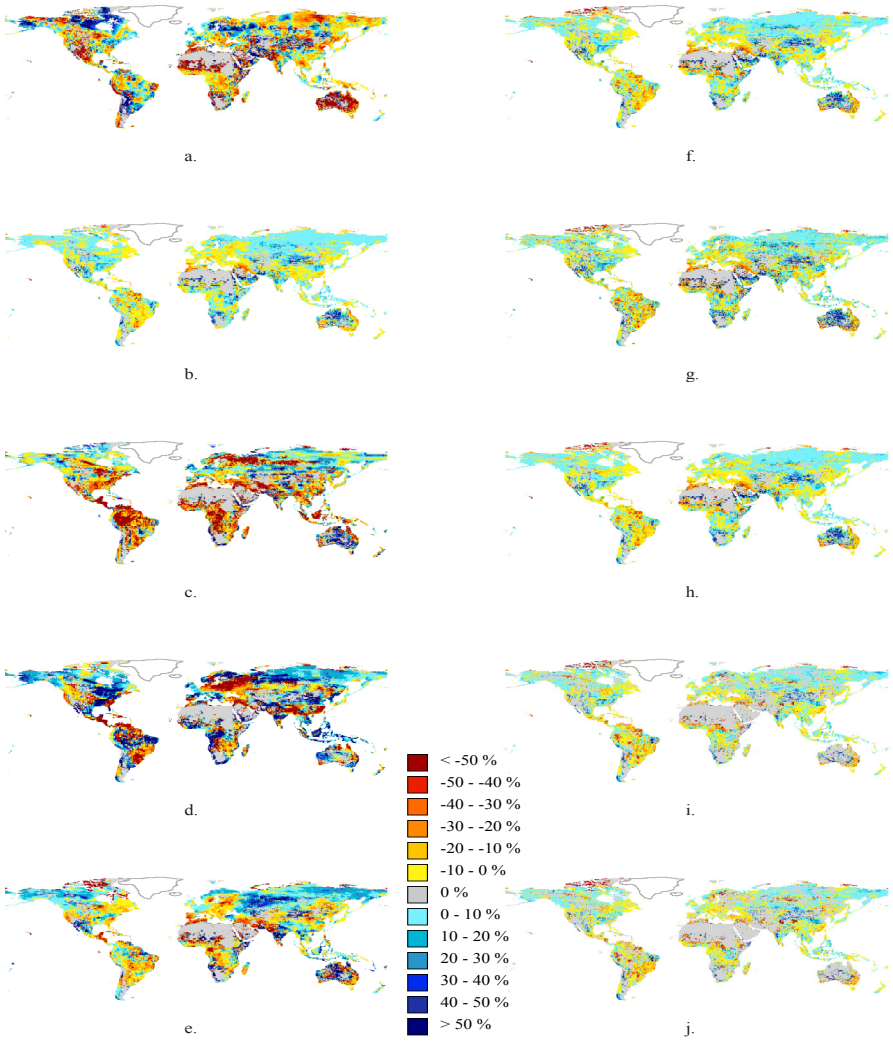


Figure 6.3 Percentage change in runoff for validation experiment as obtained from: observed (a), non-weighted (b), CSIRO-Mk3.0 (c), ECHAM5 (d), HadGEM2 (e), original REA method (f), REA with additional trend criterion (g), REA with future inter-model similarity criterion (h), selection of models performing above median (i) and lower quartile (j).

Table 6.3 Map average bias from observed runoff change between 1961-1975 and 1975-1990 and absolute runoff for the period 1961 to 1975 (%).

	Bias from observed change	*Bias from observed discharge*
non-weighted	108.8	58.7
REA: bias and convergence	111.0	40.9
REA: bias, convergence and trend	114.2	43.3
REA: bias, inter-model similarity	109.8	44.2
selection above median	110.1	53.3
bias, inter-model similarity, selection above lower quartile bias, inter-model similarity,	109.2	47.6

Table 6.4 Percentage of cells where observed change is covered by the uncertainty range (%).

Percentage of cells where observed change lies within uncertainty range	***(%)***
non-weighted	90.3
REA: bias and convergence	79.3
REA: bias, convergence and trend	80.9
REA: bias, inter-model similarity	78.7
bias, inter-model similarity, selection above median	70.4
bias, inter-model similarity, selection above lower quartile	80.7

California and parts of Southern Europe. The percentage coverage is highest for the non-weighted method, whereas selecting only those models that have a RF above or equal to the median RF resulted in the loweset coverage. Differences between the other methods are small. The inter-model uncertainty is often used as an estimator of the probability distribution of climate change (Räisänen, 2007), however in this study the GCM obtained uncertainty range underestimates the confidence interval (which should be 95% for a range of four times the rmsd) for all methods. In the application of the REA method by Giorgi and Mearns (2002) the full uncertainty range is defined by only 2 times the rmsd. Yet, in this study it is shown that when going from the percentage of cells covered by 2 times the rmsd (68% confidence interval) and the percentage covered by 4 times the rmsd (95% confidence interval) the coverage increases on average 10%. Therefore, and because the uncertainty range might be too narrow due to the informative prior and the dependence between GCMs, a larger uncertainty range of 4 times the rmsd was used as defined in equation 6.12 and 6.13. There is in general a pay-off between the width of the uncertainty range and the percentage of cells where observed change is covered. However, there is often also a pay-off between the width of the uncertainty range and the costs of the adaptation measures needed to respond to the full range of changes. Therefore attempts to reduce the uncertainty range realistically remain desirable.

The non-weighted ensemble average uncertainty range covers observed change better than the weighted ranges do. Yet, the percentage biases of the upper and lower uncertainty bounds from the observed change are largest for this method (Fig. 6.4). This indicates that the non-weighted uncertainty bounds are too wide. The uncertainty range is also wide for the REA method with an additional criteria for trend. Overall, the performance of this method is poor. The percentage of cells where observed change is covered by the uncertainty range is comparable to the other methods. Despite the wide

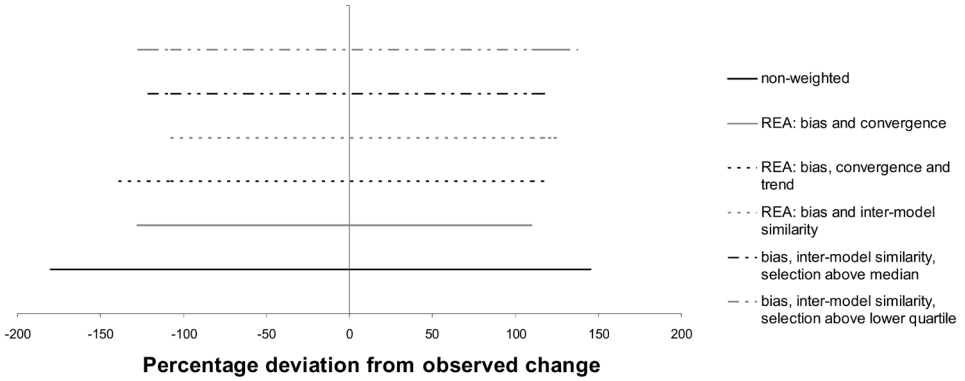


Figure 6.4 Percentage biases of upper and lower uncertainty bounds from observed change.

uncertainty range. In addition, the bias from observed mean change is large (Table 6.3). Weighting on trend was expected to have a larger influence than shown here. However, with the available datasets it was only possible to validate the weighting on bias in trend for the period 1961 to 1990. The trend is not very pronounced and highly influenced by inter-annual climate variability in such a short period (Räisänen, 2007). It is worth testing and validating this criterion with local observation datasets covering longer time-periods.

Combining the information in table 6.3, 6.4 and figure 6.4 illustrates that both the original REA method and the modified REA method (which weighs on inter-model distance instead of convergence) give good performance. The methods reduce the bias from observed runoff, the biases from observed runoff change are small and the percentages cells where observed change lies within the uncertainty range are comparable to the other methods, whereas the uncertainty ranges are narrower. The selection of the reliability criteria applied within the REA method is a subjective decision (Smith and Chandler, 2010) and will affect the weights assigned to the different GCMs (Räisänen, 2007). Therefore, the criteria used in the REA method and consequently the weighting of GCMs, should to our opinion preferably be based on the variable of interest (e.g. GCM performance for temperature, precipitation, water availability, discharge extremes, timing of regime or annual average runoff as in this study).

The selection of models performing above the lower quartile value for past performance and future inter-model distance gives reasonable results as well. The uncertainty range is slightly larger, coverage of observed change is comparable to the REA methods and the bias of the mean change from observed change is also small.

Fig. 6.5 shows the projected changes between the time-slices 1961 to 1975 and 1976 to 1990 for the non-weighted method and the best performing selection and weighting methods together with the upper and lower uncertainty bounds. Directions and amounts of change in runoff are comparable for the different methods, yet all three weighting and selection methods (Fig. 6.5b, c and d) have smaller uncertainty bounds.

The reduction of the uncertainty ranges originate from the outliers being either removed or assigned lower weights in the calculation of the rmsd. Again changes are less pronounced when calculated only from the selection of models performing above the lower quartile (Fig. 6.5d). For this method discharge increases (decreases) are present in the lower (upper) bounds in for example southern South-America, parts of the Northern Hemisphere, northern South-America and the European Mediterranean coast, indicating a higher confidence on the direction of change. Fig. 6.5b and 6.5c show little difference between change calculated with the REA method using either the future reliability criteria for future inter-model similarity or convergence.

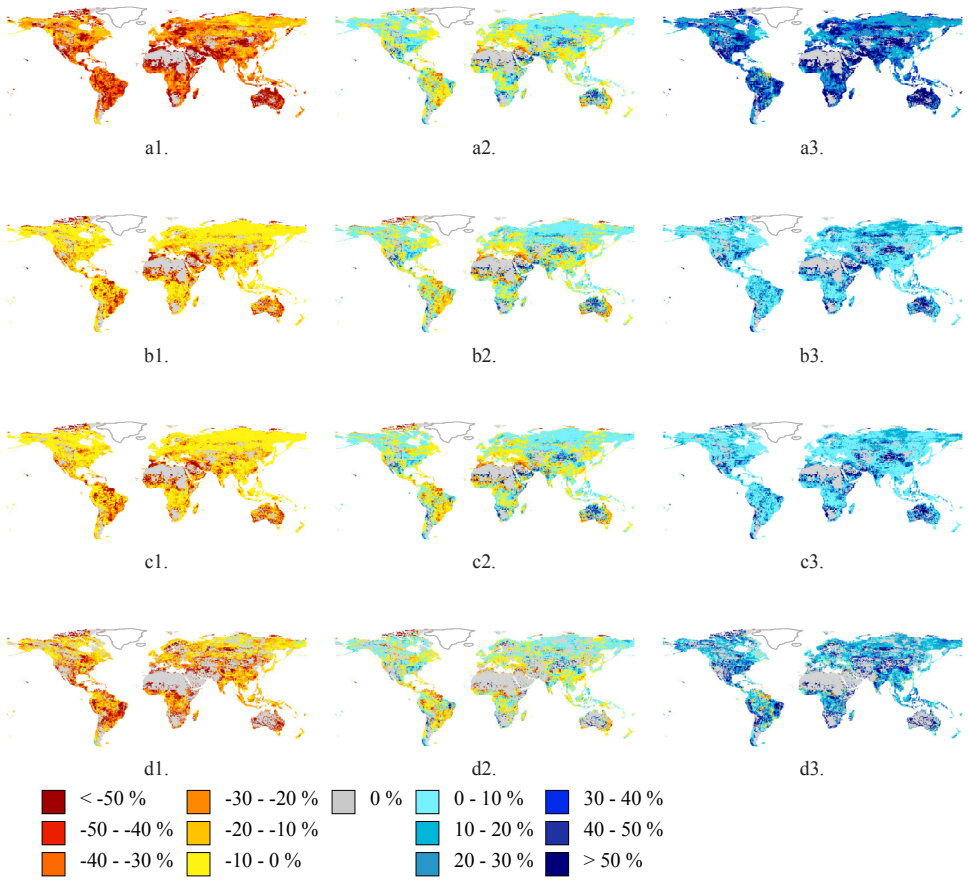


Figure 6.5 Average percentage change (middle) in runoff and lower (left) and upper (right) uncertainty bounds of change for the validation experiment: (a) non-weighted, (b) original REA method, (c) REA method with future inter-model similarity criterion, (d) selection of models performing above lower quartile.

6.3.2 Differences and consistency in runoff change projections for 2100

Figure 6.6 shows the future projected changes and uncertainty bounds derived from the non-weighted average change, and the three methods performing best in the validation experiment, being 1) the original REA method, 2) the REA average change using inter-model similarity as future criterion and 3) the change calculated from the selection of models which perform above the lower quartile RF value. As expected, projected changes for the next century are larger than changes projected for the validation period. Consequently, uncertainty ranges, especially for the non-weighted rmsd, are larger as well. Differences between the average non-weighted change and average changes derived with the other methods are small. Although individual GCMs project opposite directions of change by 2100. From Fig. 6.6 the following differences can be seen; on

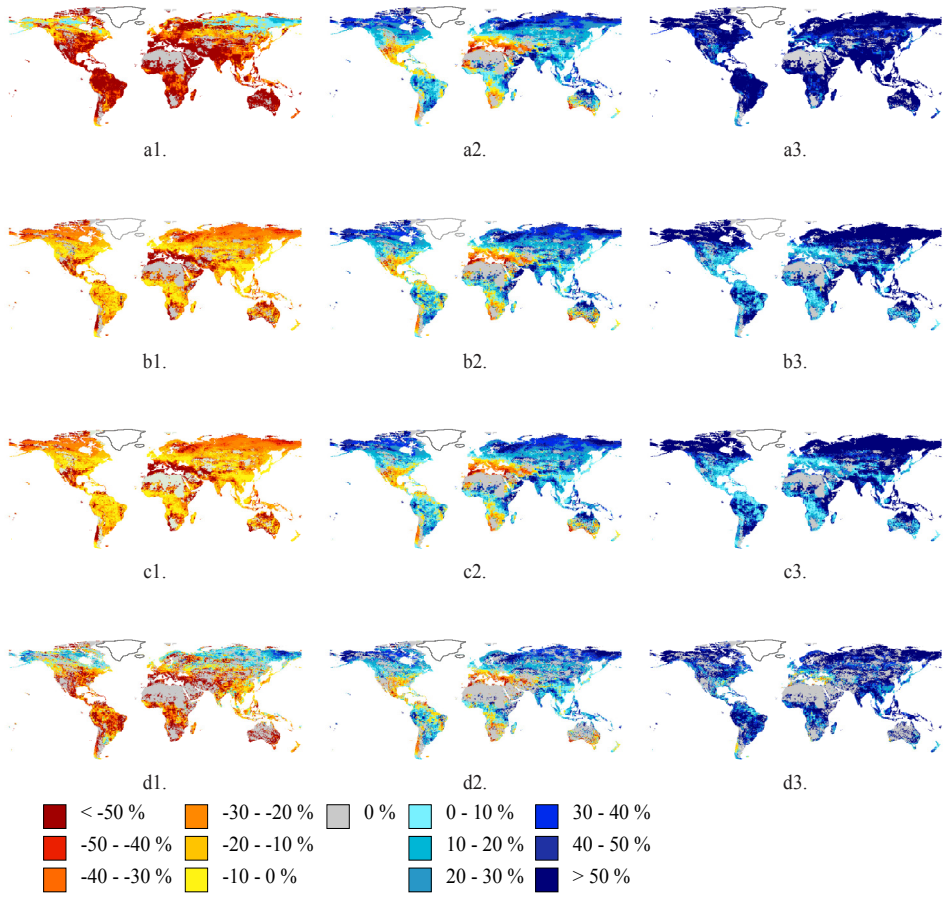


Figure 6.6 Average percentage change (middle) in runoff for 2100 together with lower (left) and upper (right) uncertainty bounds derived with: (a) non-weighted ensemble, (b) original REA method, (c) REA method with future inter-model similarity criterion, (d) selection of models performing above lower quartile

the eastern coast of Australia as well as in the Amazon, basin runoff increases for a larger region in the non-weighted average. Furthermore, there are larger regions with zero change calculated from the selection of better performing models. Unfortunately it is not possible to validate projected changes for 2100, however the validation experiment showed that uncertainty bounds deviated most from observed change for the non-weighted average and it can be expected this will hold for future periods. Especially because the REA method and the selection of better performing models reduce the uncertainty ranges for the validation period as well. For the selection of better performing models runoff increases are present in the lower uncertainty bounds in the Northern Hemisphere, and runoff decreases are visible in the upper uncertainty bounds in Turkey, parts of the European and African Mediterranean coast and the west coast of southern South-America. The future lower and upper bounds (respectively the left and right column in Fig. 6.6) of the uncertainty range derived with the REA-methods deviate from the other methods. The lower bounds indicate almost everywhere runoff decreases and the upper bounds show only runoff increases. Consequently these methods provide less confidence on the direction of change.
By 2100 the ensemble averages of all methods project runoff increases for the Northern Hemisphere, south-east Asia and the north-west coast of South-America. Furthermore, runoff is likely to decrease in southern Europe, southern Australia, the south-west coast of South-America and southern North-America.

6.4 Discussion

The global hydrological model PCR-GLOBWB was used to calculate runoff from meteorological data from GCMs, because runoff data is not available for most of the GCMs. The hydrological model may have introduced additional bias, thereby influencing the model weighting. However, since change in runoff is of interest in hydrological studies, weighting on modeled runoff is preferred above weighting on temperature and precipitation. Even if this runoff is biased due to the neglect of water use and water management in the model. As a consequence, high weights may be assigned to relatively poor (in terms of climate) performing GCMs due to interference of the hydrological model.
In this study it was assumed that the selected GCMs are independent, while in reality some process formulations and parameterizations are shared by all GCMs and output will be influenced by the relatively coarse resolution. Consequently, the calculated uncertainty ranges may still not cover the full range of possible changes (Tebaldi and Knutti, 2007). It can be questioned whether the root mean squared difference (rmsd) is a good measure to use for the derivation of the uncertainty range, since for the derivation of the rmsd from the ensemble of changes it is assumed that the changes follow a normal distribution. Yet, in our analysis other methods, as for example the derivation of the 95% prediction interval from the distributions of weighted runoff projections, showed

less effect of the use of the weighting and selection methods. Most likely because the ensemble of 12 changes is too small to reliably derive the 95% prediction interval.
The REA method shows similarity with the Bayesian model averaging approach (BMA; Tebaldi et al., 2004). However, by weighting model results against observations following fixed criteria an informative prior is used instead of the in BMA preferred non-informative distribution. This may result in a too narrow uncertainty range (Hargreaves, 2010). Although often recommended, BMA is less applicable in this study. First of all the limited number of GCMs does not allow for a thorough probabilistic interpretation of the uncertainty range. And secondly, the premise on the distribution shape of mean runoff, which is preferably known when working with a non-informative prior (Tebaldi et al., 2004), is difficult to assess and will differ per climate zone and catchment.
The study of Xu et al. (2010) provides an update of the REA method which was not considered in this study. As stated in their study the weighting on future model convergence, as included in the original REA method, can result in too narrow uncertainty bounds because highest weights are assigned to a cluster of GCMs which predict similar directions of change. Therefore Xu et al. (2010) abundant this criterion. They did not include any other critrerion for the future period. In this study we investigated a modification of the REA method, where the convergence criterion is replaced with a weighting of future model performance based on future inter-model similarity. In this method only a model showing deviations from the majority of models is down-weighted (Räisänen, 2007) hereby avoiding a too large reduction of the uncertainty range as well.

6.5 Conclusions

In this study four methods for performance based selection and weighted averaging of GCM results have been validated and the better performing methods have been used to derive future runoff change by 2100. The first method is the Reliability Ensemble Averaging (REA) method, which weights GCMs on their performance for the current climate and the convergence of the GCMs future change with the ensemble weighted average change. The other three methods are a performance based GCM selection method and two extensions of the REA method, with respectively 1) an additional criterion for reproduction of observed trend in runoff change and 2) replacement of the convergence criterion with future inter-model similarity. To obtain reliable assessments of global runoff change by 2100, all methods were applied to runoff fields calculated with a global hydrological model forced with meteorological data from an ensemble of 12 GCMs for both the twenty century control experiment (20C3M) and the IPCC SRES scenario A1B.
A validation experiment, conducted for the period 1961 to 1990 with data from the 20C3M experiment, showed that differences between weighted and non-weighted changes are small. Yet, all ensemble average changes were too conservative compared to changes obtained from observed runoff fields. Therefore it is recommend to use the

uncertainty range, here defined by four times the rmsd, alongside the weighted average change. Both the original REA method and the adjusted REA method (where the reliability criteria for future GCM convergence is replaced by future inter-model similarity) performed well although they provide less confidence on the direction of change for the future period than the other methods. Inclusion of the additional criterion for trend did not result in significant improvements. Restricting model averaging to a selection of models showing high inter-model similarity for the current and future climate gave good results, especially for the reduction of the uncertainty range.

Acknowledgements
We acknowledge the GCM modeling groups, the Program for Climate Model Diagnosis and Intercomparison (PCMDI) and the WCRP's Working Group on Coupled Modelling (WGCM) for their roles in making available the WCRP CMIP3 multi-model dataset. Support of this dataset is provided by the Office of Science, U.S. Department of Energy. We also acknowledge the Global Runoff Data Centre for providing the global discharge timeseries. And finally, we would like to thank Jaap Kwadijk for the ideas he shared with us at the beginning of this study.

7 Synthesis

7.1 Introduction

Greenhouse gas emissions have increased over the 20th century and there are strong indications that this has resulted in climate change. Temperature increased globally. Precipitation increased in North and South America, northern Europe and northern and central Asia and decreased over the Sahel, the Mediterranean, southern Africa and parts of southern Asia. Consequently, world wide river runoff regimes and water availability have been affected and they are likely to be altered even more under future changing climate conditions (IPCC, 2007a; Arnell, 2003; Nijssen et al. 2001; Vörösmarty et al. 2000; Alcamo and Henrichs, 2002).
To assess the hydrological impacts of climate change, data from global and regional climate models (RCMs and GCMs) have been used. However, absolute discharge quantities and projected changes, calculated from data of different GCMs, show large variation (Dai, 2006). Particularly for precipitation and consequently runoff. Many studies concluded that impact assessments should be based on a large ensemble of climate models. Such an ensemble samples the widest possible range of climate model uncertainties and therefore provides a reliable specification of the spread and uncertainties of projected changes (Murphy et al., 2004). Still, some change assessments, mainly in applied water management studies, are focused around one or a few GCMs. Moreover, no common practice exists for the presentation of uncertainties in the projected changes to watermanagers. And although multiple studies recognize the existence of uncertainties and biases in GCM data, the quality and usefulness of raw GCM data for hydrological applications is often not investigated before applying bias-correction or downscaling. Yet, these biases do provide important information on the deficiencies of physical process representations within GCMs and can therefore be used as guidance for model improvements and downscaling strategies.

In this thesis the different uncertainties present in the hydrological part of the modeling chain of climate impact studies (e.g. forcing uncertainty, model structural uncertainty, parameter uncertainty and resulting discharge uncertainty, see Fig. 7.1)

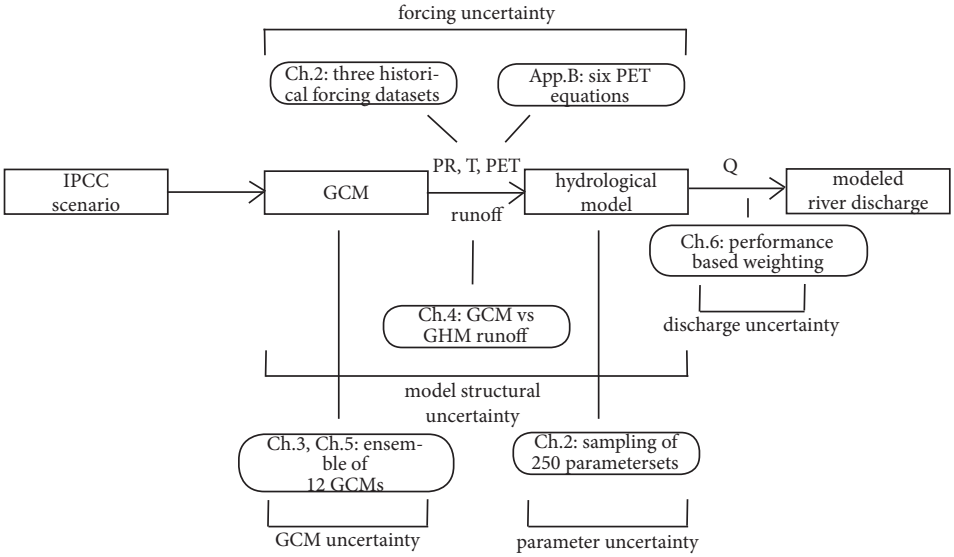


Figure 7.1 Schematization of the hydrological part of the modeling chain of climate impact assessments together with the uncertainties investigated in this thesis.

were investigated in more detail. The usability of raw GCM meteorological data and runoff for hydrological studies was assessed. And, in the final step, future hydrological changes were derived from data of an ensemble of 12 GCMs.

To address the different uncertainties present in the modeling chain of impact studies (Fig. 7.1) and to reliably derive future hydrological change, the following research objectives were formulated:

To evaluate GCM datasets on their usability for hydrological impact studies

To assess trends and noise in projected changes in future water availability obtained from an ensemble of GCMs

These objectives were divided in a number of research questions which have been answered in the preceding chapters. In this last chapter the answers to the questions are summarized and discussed in a wider perspective, giving suggestions and links to further research.

7.2 Hydrological model uncertainty analysis

Is it feasible to search for optimal parameter estimates for the hydrological model PCR-GLOBWB when the model will be forced with multiple GCM datasets?

In chapter 2 we analyzed the influence of meteorological input and parameter uncertainty on modeled discharge and explored the possibilities for the calibration of PCR-GLOBWB. The results of this analysis provide an answer to research question 1. Although some attempts for the calibration of GHMs already existed, most of these studies ignored uncertainty in meteorological forcing. Yet, biases in precipitation data are known to be large and have a major impact on hydrological model output. In this study, as an approximation of meteorological forcing uncertainty, PCR-GLOBWB was forced with three meteorological datasets: the CFSR re-analysis, the ERA-interim re-analysis and a combination of the ERA-40 re-analysis and CRU datasets. In order to assess parameter uncertainty, PCR-GLOBWB was run for each forcing dataset independently with 250 parameter combinations obtained by uniform latin hypercube sampling (LHS). Parameter values were sampled from pre-defined parameter ranges for 10 model parameters for the individual catchments. Multiple parameter estimates outperformed the default model parameterization which was obtained from available global datasets. For the forcing dataset with the largest bias, even with the best performing parameter combination from the LHS, bias in discharge remained highest. This indicates that parameter uncertainty constitutes only a relatively minor part of hydrograph prediction uncertainty. This is an important finding and demonstrates that the apparent dichotomy between model predictions and data cannot be resolved by increasing the complexity of hydrological models and resolving sub-grid processes. Instead, what is much-needed is an improved characterization of global rainfall amounts at spatial resolutions of 0.5 degrees and smaller.

Based on RMSE values of monthly discharge time-series, the 5 percent best performing parameter estimates (e.g behavioral model realizations) were selected and used for model simulations in an evaluation period. In the evaluation period, the parameter estimate performing best for the sampling period is outperformed by other parameter combinations included in the ensemble of behavioral model realizations. This illustrates the value of using an ensemble of good performing parameter combinations, instead of one optimal parameter estimate. An ensemble of model realizations is able to account for uncertainties in meteorological forcing, measurement data and the hydrological model and to transfer those uncertainties to discharge predictions. In addition to simply calculating the algorithmic mean of the ensemble of realizations (as was done in chapter 2), Bayesian model averaging (BMA) techniques could be applied. This would most likely result in even more reliable discharge estimates (Diks and Vrugt , 2010).
Although it is desirable to obtain one global optimal parameter estimate after model calibration, the analysis showed that performance of parameters sets has limited to

no persistence when being transferred to other rivers. Consequently, possibilities for regionalization are limited. Moreover, for the different forcing datasets different parameter sets performed best. These results indicate that a full model calibration of PCR-GLOBWB should be catchment and forcing specific. And, in case a hydrological model is forced with multiple non bias-corrected GCM datasets, it should be considered to work with an un-calibrated model.
As an alternative, a GHM calibration could also be conducted using input from multiple forcing datasets at the same time. The three datasets used in this analysis can be stochastically combined and the resulting meteorological time-series can be used as model input within the calibration process. Analysis of the results should clarify whether it is possible to obtain parameter sets which show reasonable performance for multiple meteorological forcing datasets. If so, the LH sampling presented in this thesis could be extended to an extensive automized uncertainty analysis, using for example the DREAM algorithm (Vrugt et al., 2009).
Finally, the optimal parameter estimates obtained by calibration, may perform less under changing climate conditions with different precipitation intensities and frequencies. By calibrating the model for different time-slices, with for example extreme dry and wet years, the stationarity in performance of different parameter estimates can be evaluated for future changing climate conditions (see the analysis of Vaze et al. (2010) for Australia).

7.3 Performance of a GCM-forced hydrological model

What is the skill of the hydrological model PCR-GLOBWB, forced with meteorological GCM datasets, in simulating hydrological variability in different regions of the world?

In chapter 3 we investigated the skill of the hydrological model PCR-GLOBWB in simulating hydrological variability when being forced with meteorological datasets from 12 different GCMs. Hereby, assessing the influence of biases in raw GCM data on river discharges obtained by hydrological modeling.
In a first analysis no bias-correction was applied to the GCM datasets. For comparison discharges were also calculated from CRU TS2.1 monthly time-series downscaled to daily values using the ERA-40 re-analysis product. In a second step the GCM data was bias-corrected by equalizing long-term monthly mean precipitation, potential evaporation and temperature to the long-term monthly mean quantities for these variables in the CRU TS2.1 dataset. As a reference, the same bias-correction was also applied to the ERA-40 re-analysis dataset.
The resulting discharge time-series modeled with the (bias-corrected and non bias-corrected) GCM datasets were compared to each other and with both observed discharge data and discharges modeled from the reference meteorological datasets. This was done in order to asses the influence of the applied bias-correction method on inter-annual variability, reproduction of discharge extremes, annual average discharge cycles and an-

nual mean discharge. The bias-correction method reduced the spread between the 12 GCMs. Long-term average discharges and timing of the peak in the annual discharge cycles were closer together after bias-correction. Yet, consistency between GCMs remained low for extreme discharges and inter-annual variability.
The large differences between discharge extremes derived from different GCM datasets, hampers the reliable assessment of changes in flood and drought occurrences. The simple bias-correction applied in this study does not overcome this problem. Nevertheless, we preferred this relatively simple bias-correction method over more advanced techniques. This, because for example downscaling by mapping of cumulative distribution functions (CDFs) of daily rainfall amounts (Haerter et al., 2010; Kallache, 2011) is likely to influence extremes in the GCM datasets for the current and future climate. Future changes in precipitation intensity and frequency, which impact flood and drought occurrences, also influence the shape of the CDF of daily rainfall values. As a consequence, in a downscaling procedure based on CDF mapping, future intensities and frequencies are often incorrectly mapped to current frequencies and intensities. In addition, by applying the simple bias-correction used in this thesis, differences in extremes for the ensemble of GCMs remain. This reduces the danger of underestimating the uncertainty present in future change projections of extremes.
Yet, the CDF downscaling techniques can be improved by considering a varying bias over time (Beven, 2011), or at least test the assumption of stationarity in bias for the current climate, as has been presented by Piani et al. (2010). Still, quality of bias-correction will remain dependent on the quality of the meteorological reference dataset used. Non bias-corrected data can be used as well if, instead of focusing on absolute discharge extremes, future change assessments focus on changes calculated relative to the discharge values for the current climate experiment (see chapter 5). The relative change is used in order to reduce the influence of biases in absolute discharge amounts. The main discrepancy introduced by using non bias-corrected data may result from the large number of drizzling days in GCM precipitation. As a consequence droughts will be modeled less frequent.

The analysis in chapter 3 illustrates the inadequacies of GCM datasets when being used for discharge modeling. In his review, Beven (2011) even questions the use of GCM data for hydrological climate change studies. Yet, as Beven (2011) also mentions, there are limited other tools which can be used to assess the possible hydrological consequences of climate change. There are two important strategies to improve the reliability of hydrological impact studies using GCM data. First of all, GCMs are continuously being improved and recently there is special focus on modeling of hydrological processes within GCMs by using improved land surface schemes (Van den Hurk et al, 2004; Pappenberger et al., 2009). In addition, several sources of errors within GCMs have been identified (Pielke et al., 2009; Beven, 2011). These are amongst others scale effects, process formulation of sub-grid rain, snow and cloud parameterization, inadequate heat

exchange between ocean and atmosphere and the influence of anthropogenic forcings other than green house gasses. By eliminating these deficiencies during future GCM developments, the suitability of GCM data for hydrological impact modeling could be improved.

7.4 Quality of GCM runoff

How well do runoff fields of GCMs reproduce observed hydrological variability and could these runoff fields make the use of a hydrological model redundant?

During the review process of two of our papers (chapter 3 and 5 in this thesis), we received the critical question why we used meteorological data from GCMs to derive runoff with our own hydrological model, instead of using GCM runoff directly. To answer this question the analysis in chapter 4 was conducted. Here we evaluated the quality of GCM runoff and its usability for hydrological impact studies. The analysis in this study was restricted to two GCMs (ECHAM5 and HadGEM2), as for most GCMs daily time series of gridded runoff are not yet downloadable through GCM databases (CERA-gateway, PCMDI data portal). Consequently, the results presented here only apply to these two GCMs.

Four different methods were employed to convert GCM runoff, precipitation and actual evaporation into river discharge: 1) simple accumulation of GCM runoff, 2) hydraulic routing of GCM runoff, 3) hydraulic routing of GCM runoff combined with temporal storage of subsurface runoff in a groundwater reservoir and 4) off-line hydrological modeling with the global model PCR-GLOBWB using meteorological data from the GCMs as forcing.

Comparison of the results with discharge observations revealed that the quality of the generated discharge was highly influenced by the quality of GCM data. In small catchments, like the Rhine, the methods which included runoff routing (2 and 3) performed equally well. However, off-line modeling with PRC-GLOBWB outperformed all other methods for ECHAM5. For HadGEM2, PCR-GLOBWB outperformed the other methods for the Mississippi, Lena and Rhine. For larger catchments, river routing was required to introduce realistic travel times, decrease day-to-day variability, reduce extremes and reproduce the long-term average annual discharge cycle correctly.

The complexity of the LSS of both GCMs is comparable to the complexity of large scale hydrological models. However, in HadGEM2 absence of sub-grid variability for generation of saturated surface runoff in combination with the coarser GCM resolution results in a large subsurface runoff flux and a low seasonal variability of the annual discharge cycle. In ECHAM5 on the other hand, infiltration and subsurface flow is limited. Consequently, the total runoff flux has a high day to day variability and a low baseflow.

At the current stage it is still useful to include a hydrological model in climate impact assessments. Not only does the routing result in realistic travel times and peak flows. The

separation of precipitation in surface and sub-surface runoff is also better represented in PCR-GLOBWB than in the LSS of the GCMs investigated. Moreover, baseflow can be modeled more realistically due to the delayed release of drainage from the groundwater reservoir. There is an increasing interaction between meteorologists and hydrologists which is amongst others established in the Global Energy and Water cycle EXperiment (GEWEX). This interaction results in the improvement of hydrological, or land surface processes included in LSS of GCMs (van den Hurk et al., 2004; Pappenberger et al., 2009). The hydrological outputs of GCMs can further be improved by focusing the tuning and verification of climate and weather models, in addition to performance for temperature and radiation, on outputs of interest to hydrologists.

7.5 Global patterns of change in runoff

What hydrological impacts are to be expected for different regions of the world and what is the ratio between trend and noise for the ensemble of projected hydrological changes?

In the analysis presented in chapter 3 the global hydrological model PCR-GLOBWB has been forced with meteorological datasets from an ensemble of 12 GCMs for the current climate experiment (20C3M) and future IPCC SRES scenario A1B. This exercise revealed the regions of the world where significant and consistent hydrological changes are most likely. New compared to existing studies (Arnell, 2003; Nijssen et al. 2001; Vörösmarty et al. 2000; Alcamo et al., 2007) was: 1) the comparison of spatial patterns of regime changes and 2) the quantification of consistent significant change calculated relative to both the natural variability of the individual GCMs and the inter-model spread using paired T-tests. Significance of projected changes was small when calculated relative to the inter-model spread due to the large differences between GCMs. In this study we showed that, when significance of change is calculated for each GCM individually relative to its own inter-annual variability, changes are significant for a larger number of rivers and the global areas with significant change are more extended. Using these significant changes of individual GCMs, the consistency of the ensemble of GCMs can be quantified by counting, for each model grid cell and each river, the number of GCMs projecting significant change in the dominant direction. The resulting consistency maps are a promising way forward to detect regions with high likelihood of change.

By 2100 the maps show a consistent projected decrease in runoff for southern Europe, southern Australia, the south and north of Africa and southwestern South-America. Significant discharge decreases were projected for most African rivers, for the Murray and for the Danube. Runoff increases were obtained for sub-Arctic and Arctic regions. The results also indicated that river discharges are expected to increase for the Yangtze, Yellow river, Mekong, Indus and Brahmaputra because of an increase in monsoon rainfall. Due to earlier snowmelt the Lena and MacKenzie showed an increase in spring

discharge and a small shift in timing of peak. In addition, freshwater outflow to the oceans will increase for all continents and oceans except the Mediterranean Sea.

7.6 Performance based weighting and selection of GCM projections

How can meaningful hydrological change information be extracted from an ensemble of GCMs in light of the large ensemble uncertainty?

In chapter 6 we explored several performance based weighting and selection methods to obtain a more reliable assessment of projected future runoff change from the ensemble of 12 GCMs which had already been used for the future discharge change assessment in chapter 5. In chapter 6 focus was on PCR-GLOBWB modeled runoff, instead of discharge changes. As modeled global runoff fields could be validated and weighted on a cell specific basis against reference runoff fields derived from GRDC discharge data (Fekete et al., 2002).

Four weighting and selection methods were explored. They were all based on a weighting method proposed by Giorgi and Mearns (2002) for the weighing of GCM precipitation and temperature, Reliability Ensemble Averaging (REA). The four methods are: 1) the original REA method, 2) the REA method extended with a reliability criteria for the reproduction of observed trend (Räisänen, 2007), 3) the REA method where future GCM convergence is replaced by future inter-model similarity (Räisänen et al., 2009) and 4) a method selecting only those models showing a high inter-model similarity for the observed and future climate (Räisänen et al., 2009; Smith and Chandler, 2010).

The differences between weighted and non-weighted average runoff changes for 2100 were small. Within a validation experiment, where GCM ensemble mean change was derived between the time-slices 1961 to 1975 and 1976 to 1990 and compared with observed change, the non-weighted ensemble change resembled observed change best. Yet, both the weighted and non-weighted average changes were too conservative. This underscored the importance of considering an uncertainty range around the (weighted) ensemble average change. In this chapter an uncertainty range (or 95% confidence interval) of four times the root mean square difference (RMSD) around the ensemble mean change was used. Here the cell specific RMSD was calculated form the ensemble of twelve projected changes. The uncertainty range derived with the non-weighted method was relatively wide, upper and lower uncertainty bounds showed large biases from observed change. However, the uncertainty range could reliably be reduced by using only a selection of GCMs which showed high inter-model similarity for the current and future climate. The weighting method in this chapter was presented on a global scale and not yet connected to a specific water management issue. A first application of the method, used to assess changes in high and low flows within the Lower Brahmaputra floodplain, can be found in Gain et al. (2011).

The analysis in chapter 6 revealed that changes projected by the (weighted) ensemble

mean can be relatively conservative due to averaging of opposite changes. We therefore urged the use of an accompanying uncertainty range. It was proposed to derive this uncertainty range from the weighted GCM ensemble root mean square difference. One critical comment that may be raised here is that the resulting uncertainty range is obtained from model uncertainty and does not represent the probability of real life climate turning out that way (Beven, 2011). Still, this probability distribution, which provides an indication of the uncertainty in projected changes, can be useful for adaptation strategies. By considering these uncertainty ranges, robust no regret adaptation strategies will be developed which can at least cope with the model uncertainty. Moreover, it is important to present these uncertainty ranges since the outcome of physically based models is often convincing evidence for political decisions and water managers are frequently looking for a quantification of the problem.

7.7 Main conclusions

Within this study the usefulness of GCM data for large scale hydrological impact assessments and the possibilities for incorporating the inherent uncertainties in the projected changes have been investigated. The main conclusions are summarized below:

- Parameter uncertainty constitutes only a relatively minor part of discharge prediction uncertainty. Therefore, deviations between hydrological model predictions and measured discharge data, cannot be resolved by increasing model complexity and resolving sub-grid processes alone. Instead, an improved characterization of global rainfall amounts at spatial resolution of 0.5 degrees and smaller is needed.
- Runoff calculated with a hydrological model outperforms GCM runoff when used for the reproduction of 1) discharge extremes and 2) river flow in large catchments. Runoff routing and correct representation of subsurface hydrological processes, as present in hydrological models, produce reliable flow velocities and consequently realistic discharge amounts.
- For the current climate, large differences between GCMs and large deviations from observed discharge were obtained. In addition, future hydrological projections vary widely between GCMs. Therefore, an ensemble of GCMs should be used.
- With the ensemble of 12 GCMs used in this study, consistent and significant future discharge changes have been retrieved. By 2100, runoff decreases are projected for southern Europe, southern Australia and the south and north of Africa. Runoff increases were obtained for sub-Arctic, Arctic and Monsoon regions.
- Performance based weighting or selection of GCMs results in reliable future hydrological changes. In addition, GCM performance can be used to derive reduced uncertainty ranges around the projected changes.

7.8 Outlook

Additional impact analysis using future hydrological change projections
The global hydrological change projections, created in this study, provide useful information for water resources assessments with a focus on multiple aspects. The projections can be used to explore the impact of the projected hydrological changes on:

- global water stress (Wada et al., 2011; Henrichs et al., 2002, Vorosmarty et al., 2000);
- connectivity of large fresh water ecosystems (Döll and Zhang, 2010);
- change in fresh water availability from mountains (Viviroli et al., 2007; Immerzeel et al., 2010);
- virtual water flows and possibilities to alter virtual water flows to inverse nega tive impacts of climate change (Hoekstra and Hung, 2005);
- planning of reservoir and damn developments (Döll et al., 2009; Van Soesbergen and Mulligan, 2010).

In addition, the future hydrological change projections could be stored in a freely accessible repository from where they can be used for quick water resources assessments in less developed countries with limited high resolution climate projections available. A first example of such and analysis can be found in Gain et al. (2011).

Assessing the influence of GCM data downscaling on future hydrological projections
In this thesis future hydrological change was calculated from non bias-corrected data. To minimize the influence of biases in GCM data, relative changes were calculated instead of absolute future discharge values. Yet, proper downscaling might improve or alter the hydrological projections. In ongoing research we will assess future hydrological changes and their impacts on the connectivity of fresh water ecosystems. To this end GCM data will be downscaled with the CFSR reanalysis product using the CDF-t method (Michelangeli et al., 2009). The resulting downscaled data will be used to force PCR-GLOBWB. In a first step, gridded daily potential evaporation time series have been calculated from the CFSR dataset with the Hargreaves equation. This equation was selected as most reliable in an application to CFSR dataset at a global scale (see for more information Appendix B). The resulting daily time-series will be used for the downscaling of daily potential evaporation time series derived from GCM data. A comparison between the in this thesis created future projections and the projections obtained from the PCR-GLOBWB run forced with downscaled GCM data should reveal the need for downscaling.

Assessing the likelihood of critical changes

Instead of presenting and confronting water managers with the probabilities of change for a given region, it would be interesting to investigate an alternative approach. In a first step, water managers should indicate the important upcoming problems in their region (e.g. change in flood frequencies, decreasing water availability) and they should conduct cost-effect analysis to assess to which extent increases or decreases can be handled by simple adaptation strategies. The resulting tipping points (= critical changes) could then be placed within the weighted GCM derived probability distribution of change, in order to indicate the future likelihood of their occurrences. In such an approach, apart from using the tipping points to indicate the expected time left to reach these critical thresholds (Kwadijk et al., 2010), the tipping points can be used to indicate the likelihood of the occurrence of critical changes and consequently the need to prepare for them. A similar GCM weighting approach (based on the REA method) has been introduced by Giorgi and Mearns (2003) for the calculation of threshold probabilities for precipitation changes.

The new generation IPCC scnenarios

In 2012/2013 the new generation of IPCC scenarios (AR5) becomes available (Moss et al., 2010). These scenarios provide projections for a shorter time horizon (up to 2035) as end-users are interested in climate scenarios that focus on the next two to three decades for the development of adaptation strategies. Within these short-term projections, a variety of adaptation measures and socio-economic scenarios can be included. The scenario development exists of a "parallel phase" in which climate and integrated assessment modelers work simultaneously to enable modeling of the influence of adaptation strategies on climate change and vice versa. In addition to the short-term projections, long-term projections (e.g. 2100 and beyond) have been developed. These are mainly important from a scientific point of view.

The new scenarios are divided in four forcing pathways which represent a diverse range of socio-economic and technological development scenarios. The scenarios are also called "representative concentration pathways" (RCPs) and have been selected based on extended literature reviews. The new emission scenarios are pre-scribed in more detail than previously, as more physical processes and land use data have been included in the emission models. In addition, most projections are available on a higher resolution. The question is whether the new scenarios will provide other insights on the global hydrological changes than the current IPCC AR4 SRES scenarios employed in this thesis.

Assessing the value of low resolution global hydrological projections on local scale

A local scale comparison between future hydrological projections obtained with the large scale hydrological model PCR-GLOBWB and projections obtained with higher resolution catchment-scale models can provide information on the possibilities for the calculation of future hydrological changes with a less demanding low resolution hy-

drological model. For such an analysis a comparison similar to the study of Gosling et al. (2011) for the large scale hydrological model Macro-PDM.09 would provide the necessary information. If projected changes calculated with the low resolution hydrological model are comparable to the high resolution model projections, the in this thesis created low resolution future projections are of great value for the definition of water management strategies in data and model sparse regions.

Extension of hydrological impact assessment with other factors

For a correct assessment of future water availability, the impact of climate change on hydrological conditions should be placed in light of other factors inducing hydrological changes, like population growth, land use change and water management (e.g. development of damns and reservoirs and sustainable water withdrawals and use). The impact of these factors may be comparable or larger than the impact of climate change (Beven, 2011; Vörösmarty et al., 2000; Alcamo et al. 2007; Arnell, 2004).

Appendix A

A.1 Model performance PCR-GLOBWB

To assess the performance of PCR-GLOBWB we ran the model for the period 1961 to 1990 with CRU TS 2.1 monthly time series (New et al. 2000) downscaled to daily values with the ERA-40 re-analysis data (Uppala et al., 2005). The monthly and annually average simulated discharges are compared with observed discharges for the individual catchments. For most of these catchments average monthly discharges are available from the Global Runoff Data Centre (GRDC, 2007). For the remaining catchments data from the Global River Discharge Database were used (Vörösmarty et al., 1998a).

A.1.1 Continental discharges

In Table A.1 the mean yearly continental discharges modeled with PCR-GLOBWB are compared with continental discharges of previous studies calculated from both observations and model results (Wada et al., 2008; Sperna Weiland et al., 2010). Continental discharges vary amongst studies and, as described in more detail below, the PCR-GLOBWB continental discharges are comparable to results from other global hydrological models. For Asia, North-America, South-America and Oceania the results of

Table A.1 Continental runoff based on data and model based estimates in km^3/year (Wada et al., 2008)

	Europe	Asia	Africa	N.-America	S.-America	Oceania	Global	Time Period
Data based estimates:								
Baumgartner and Reichel (1975)	2564	12467	3409	5840	11039	2394	37713	-
Korzun et al. (1978)	2970	14100	4600	8180	12200	2510	44560	-
L'vovich (1979)	3110	13190	4225	5960	10380	1965	38830	-
Shiklomanov (1997)	2900	13508	4040	7770	12030	2400	42648	1921-1990
GRDC (2004)	3083	13848	3690	6294	11897	1722	40533	1961-1990
Average	2925	13423	3993	6809	11509	2198	40857	-
Model based estimates:								
Fekete et al. (2000)	2772	13091	4517	5892	11715	1320	39319	-
Vörösmarty et al. (2000)	2770	13700	4520	5890	11700	714	39294	1961-1990
Nijssen et al. (2001)	-	-	3615	6223	10180	1712	36006	1980-1993
Oki et al. (2001)	2191	9385	3616	3824	8789	1680	29485	1987-1988
Döll et al. (2003)	2763	11234	3592	5540	11382	2239	36687	1961-1990
Widén-Nilsson et al. (2007)	3669	13611	3738	7009	9448	1129	38605	1961-1990
Average	2833	12204	3933	5730	10536	1466	36566	-
PCR-GLOBWB:	2143	11461	5573	5249	11186	2633	36245	1961-1990

*) This table with continental discharges from previous studies has been obtained from Wada et al. (2008).

PCR-GLOBWB are within the range of continental discharges from previous studies. Differences between the discharges modeled by PCR-GLOBWB and the average continental discharges of the other studies are small. Globally the average difference is only 1%. For Europe PCR-GLOBWB slightly underestimates continental discharge, whereas for Africa PCR-GLOBWB modeled discharge is relatively high. PCR-GLOBWB tends to overestimate discharge for most African rivers. This is due to the absence of water withdrawals in the model and limited river bed losses, like recharge to the groundwater and evaporation (which is particular relevant in the deltas of the Niger). In addition, and probably more important, the density of meteorological observations is relatively low in Africa which limits the quality of the CRU TS 2.1 and ERA-40 precipitation for this continent (Fekete et al., 2004; Van Beek and Bierkens, 2009).

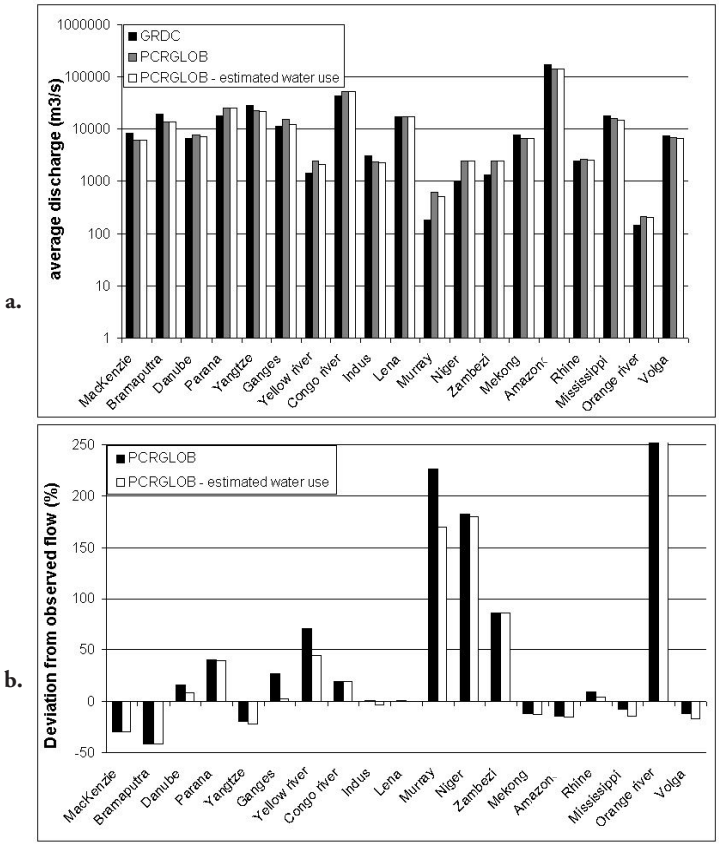


Figure A.1 a. Annual mean observed discharge (GRDC), annual mean discharge calculated with PCR-GLOBWB (forced with the CRU time-series downscaled to daily values with the ERA-40 dataset) and the annual mean discharge calculated with PCR-GLOBWB corrected with estimated of water use.
b. 1) Percentage deviation of annual mean discharge calculated with PCR-GLOBWB (forced with the CRU time-series downscaled to daily values with the ERA-40 dataset) from the GRDC observed mean annual discharge and 2) the percentage deviation of the annual mean discharge (calculated with PCR-GLOBWB corrected with an estimate of water use) from GRDC observed mean annual discharge.

A.1.2 Discharge statistics

Figure A.1a shows the annual mean modeled and observed catchment discharges and figure A.1b shows the modeled discharge as a percentage of the observed GRDC discharge. The annual mean modeled discharge shows little deviation from observed annual mean discharge for the Danube, Volga, Rhine, Mississippi and Lena. Deviations are relatively large for the drier basins; Murray, Niger, Zambezi and Orange where river discharge is highly influenced by water use and evaporation. For the MacKenzie modeled discharge is too low due to undercatch in the CRU snowfall amounts (Fiedler and Döll, 2007). Discharge is also relatively low for the Monsoon influenced Asiatic rivers Brahmaputra, Mekong and Yangtze, while for the Ganges and Yellow River discharges are relatively high.

To investigate the possible influence of water use on river discharge, we made the simplifying assumption that water demand minus the non-renewable groundwater abstractions equals water use from surface water (Wada et al., 2010a; Wada et al., 2010b). Although in reality water demand will, especially in drier regions, be higher than water use. The here described water use is estimated on a grid of 0.5 degrees and is the sum of estimated industrial, agricultural and domestic water. From this sum the amount of water which is abstracted from non-renewable groundwater sources is subtracted. According to Fig. A.1b the deviations between observed and calculated discharge for the Murray, Danube, Rhine, Ganges and Yellow river can partly be assigned to the lack of inclusion of water use.

In figure A.2 the observed monthly modelled discharges have been regressed on the simulated discharges for the reference period 1961 to 1990. For average monthly discharge values below 100 m^3/s simulated discharge show relatively large deviations from observed. Differences between simulated and observed discharge are larger than one

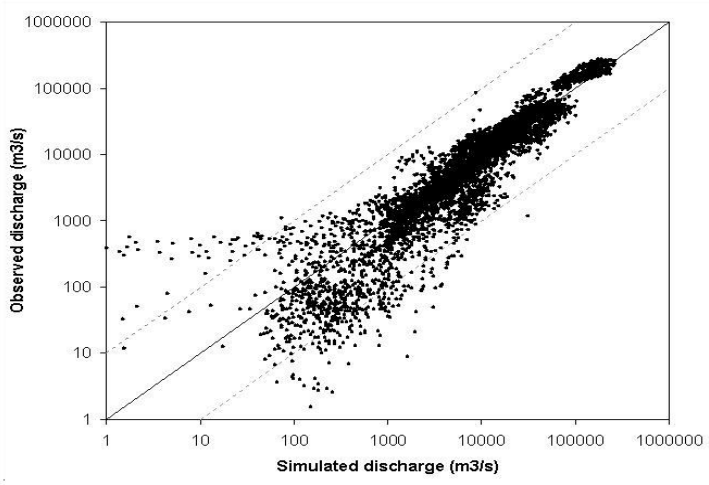


Figure A.2 Observed vs simulated monthly mean discharges (m^3/s) for the period 1961 to 1990 for the 19 selected rivers. The solid line represents the 1:1 slope, the dashed lines represent values deviating one order of magnitude.

order of magnitude and monthly discharges are both under- and overestimated by the hydrological model. In the range between 100 and 10000 m^3/s the model tends to slightly overestimate monthly discharge values, amongst others due to the lack of water use inclusion. However, the spread around the 1:1 line decreases with increasing monthly average discharge, which can be expected for a global hydrological model that cannot accurately simulate discharge for catchments smaller than 300 km due to its coarse resolution.

A.1.3 Annual discharge cycles

Twenty year average modeled annual discharge cycles are compared with observed cycles for the individual rivers (Fig. A.3). There are large differences in performance for the individual rivers. As stated above, meteorological forcing consists of CRU monthly data downscaled to daily data with ERA40, which is one of the best global climate time series there is to offer for the period considered. Despite this, there are large forcing errors causing biases in certain parts of the world. As expected, CRU (and also ERA40) has the highest quality in North America and Europe, where the density of rain gauges (as well as synoptic data assimilated into ERA40) is the largest. This can be seen by the good performance for e.g. the Mississippi, Danube and Rhine. Estimates of CRU (but also ERA40) are poor in the African tropics, which largely explains the larger bias for these rivers. Monsoon dominated rivers elsewhere perform reasonable well, expect for the Indus where a large glacial component adds to the discharge peak that is not accounted for (Immerzeel et al., 2010). The underestimation of the runoff peak of rivers with a large snow component (Lena, Mackenzie) is caused by the snow under catch that is prominent in CRU (Fiedler and Doll., 2007). River flow in the Parana is highly regulated with reservoirs and, although large reservoirs have been included in PCR-GLOBWB to enable temporal water storage, reservoir management and flow regulation are not implemented in the model. As a result the modeled annual cycle shows larger seasonal differences than observed. For the Rhine the discharge peak is modeled in April and observed in February due to a too late modeled snow melt driven discharge peak. For both the Amazon and Congo the modeled annual cycles show a discharge rise in respectively March and April which is not present in the observed discharges. For the Congo (Zaitchik et al., 2010) found a similar shift in peak timing which they assign to the complexity of the bimodal runoff regime and the poor quality of the GRDC gauge data for the Congo.

Annual cycles of modeled discharge minus estimated water use have been included in Fig. A.3 as well. The plots show that considering the influence of water use improves the resemblance between modeled and observed discharge considerably for the Danube, Ganges, Yellow river, Murray and Mississippi.

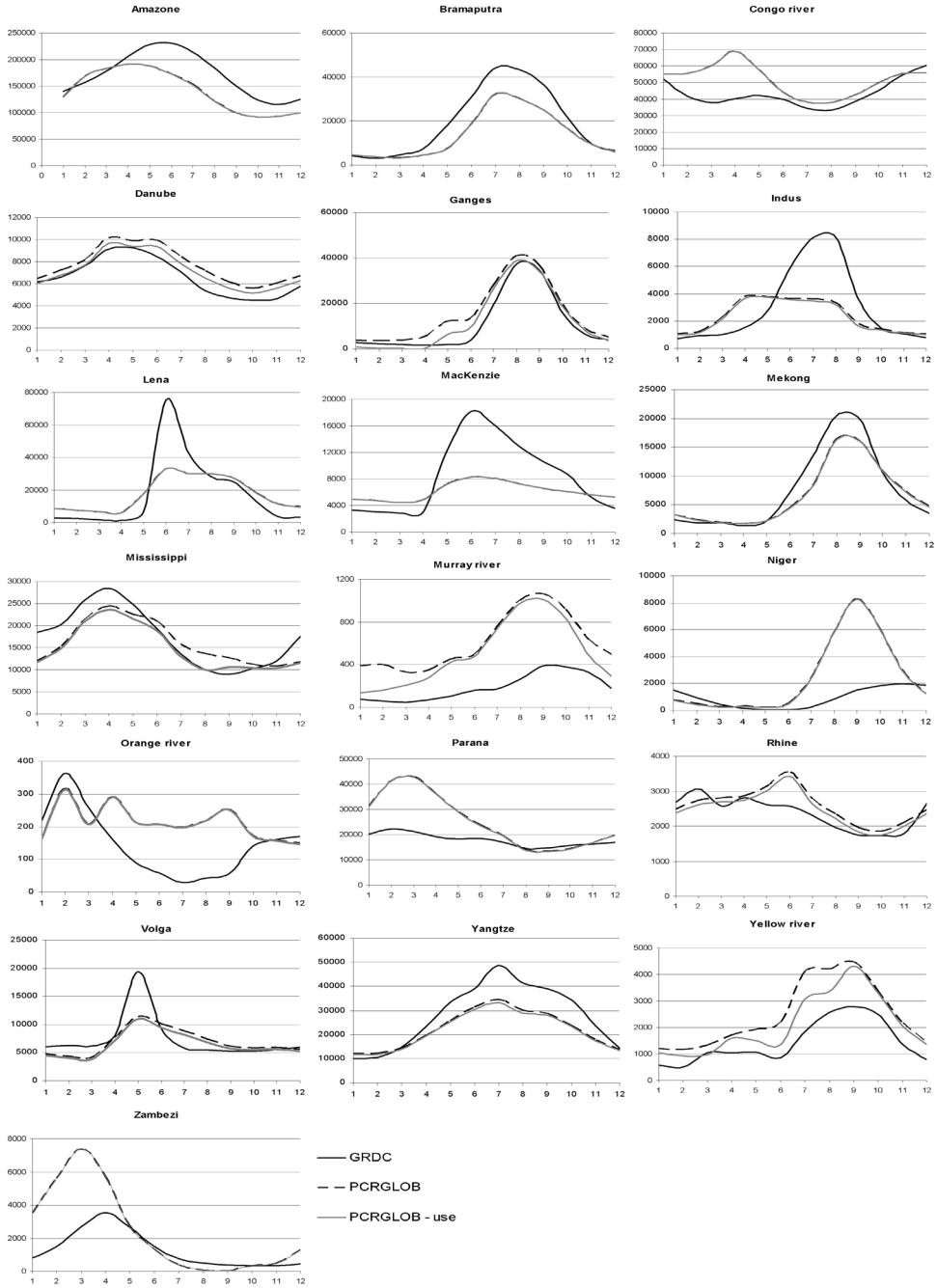


Figure A.3 Annual average modeled and observed discharge cycles (m^3/s) for 19 selected rivers. Showing for the period 1961-1990 the monthly 30-year average annual cycles derived from the PCR-GLOBWB run based on CRU downscaled data and annual cycles derived from observations. Additionally annual cycles of PCR-GLOBWB modeled discharge minus estimated water use have been constructed.

A.2 Penman-Monteith vs Blaney-Criddle

For most GCMs reference potential evaporation is calculated with a modification of the Penman-Monteith equation (Allen et al., 1998; Van Beek, 2008; Sperna Weiland at al., 2010). However, for some GCMs several input variables (e.g. wind speed, air pressure, radiation) required to calculate the Penman-Monteith evaporation were missing. For these GCMs the Blaney-Criddle equation is used (Brouwer and Heibloem, 1986; Oudin et al., 2005). The Blaney-Criddle equation is a simple temperature based potential evaporation estimator, whereas Penman-Monteith considers aerodynamics and radiation as well.

For several GCMs we compared potential evaporation calculated with the Penman-Monteith and the Blaney-Criddle equations and their resulting discharges. For brevity results are only shown for the CGCM2.3.2 model (Fig. A.4a and A.4b). For most GCMs potential evaporation calculated with the Penman-Monteith equation is high compared to Blaney-Criddle potential evaporation in Northern Australia, the Sahara, Southern Africa, the southwest US and Northern India and relatively low for Europe, the northen US, Canada, Russia, southeast Asia and the Amazon. However, only for specific periods, and in regions where evaporation limitation by soil moisture conditions is small, deviations in potential evaporation will introduce deviations in actual evaporation and runoff. Fig. A.4c shows the percentage difference in discharge calculated using either the Blaney-Criddle or the Penman-Monteith potential evaporation. Deviations are large for the northern regions of the northern Hemisphere, the Amazon basin, Europe and parts of southeast Asia where discharge calculated with Penman-Monteith potential evaporation is relatively high. The Penman-Monteith based discharge is relatively low in arid regions, the Indus basin and Himalayas. Unfortunately hydrological studies are restricted to the available GCM datasets and, since not all required Penman-Monteith variables are reported for all GCMs, the use of a simple temperature based equation like Blaney-Criddle can not be avoided. Still, for those GCMs where all variables were available we preferred to use the FAO recommended Penman-Monteith equation (Allen et al., 1998).

In this study not the absolute discharge quantities, but the changes in average discharge and discharge extremes were of interest. Therefore we analyzed the influence of using either Blaney-Criddle or Penman-Monteith potential evaporation as input to the hydrological model on the resulting discharge changes. Hereto discharge changes derived with the hydrological model forced with Blaney-Criddle potential evaporation are regressed on discharge changes derived with the hydrological model forced with Penman-Monteith potential evaporation (Fig. A.5). For this analysis we used data from the first realization of CGCM2.3.2 for the 20C3M experiment and A1B scenario. Overall, for 2100, the different potential evaporation equations result in similar directions of change. There are two exceptions. The first is the direction of change for maximum discharge in the MacKenzie, which is negative when using the Blaney-Criddle equation

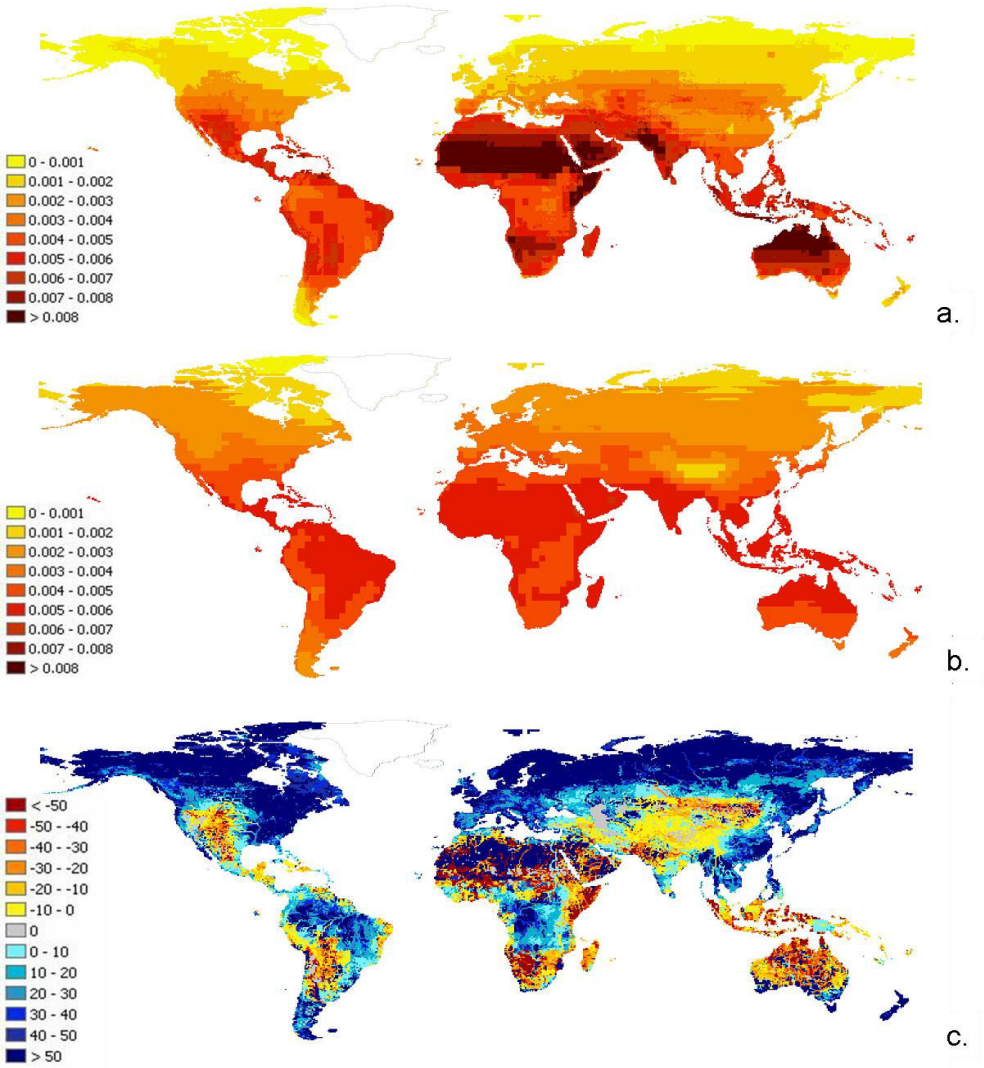


Figure A.4 Maps with reference potential evaporation (m/day) and resulting percentage discharge difference (%). **a.** twenty year average reference potential evaporation (m/day) calculated with the Penman-Monteith equation. **b.** twenty year average reference potential evaporation (m/day) calculated with the Blaney-Criddle equation and **c.** percentage difference (%) between twenty year average discharges calculated with PCR-GLOBWB from either Penman-Monteith or Blaney-Criddle.

and positive for the Penman-Monteith equation. This is a result of the large differences in absolute discharge quantities for the MacKenzie which tend to be twice as high for the Penman-Monteith equation. The second exception is the Ganges where minimum discharge decreases with the Penman-Monteith method, while it increases for the Blaney-Criddle method. For the remaining catchments directions of change in minimum, maximum and mean discharge are the same when using either two equations.

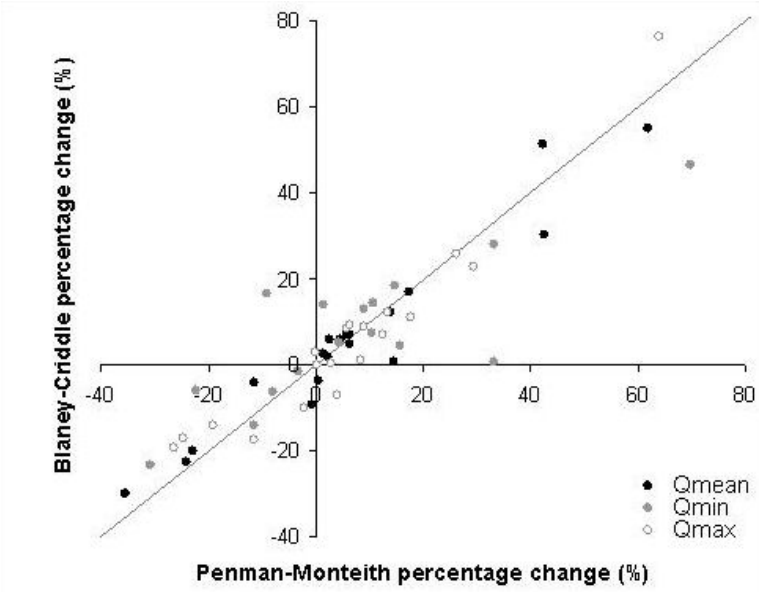


Figure A.5 Percentage change in discharge calculated using potential evaporation derived with Blaney-Criddle vs Penman-Monteith. Black dots represent projected change in average discharge, grey dot represent changes in high flows (Qmax), white dots represent changes in low flows (Qmin). The solid line represents the 1:1 slope.

In general the projected changes follow the 1:1 slope, although differences in magnitude of projected change exist.

A.3 Consistency of change for multiple realizations of one GCM

For CGCM2.3.2, the GCM with the highest number of realizations for both the 20C3M experiment and the A1B scenario, we calculated change in discharge by 2100 for all five available realizations. Boxplots of projected changes for the 19 catchments are shown in Fig. A.6a. Fig. A.6b shows boxplots of changes projected by the twelve individual GCMs included in our ensemble. The boxplots of the five realizations of CGCM2.3.2 cover much smaller ranges than the boxplots derived from the ensemble of GCMs. Furthermore, for 13 out of 19 catchments the direction of change is consistent for all five realizations of the CGCM2.3.2 model. In Fig. A.7a a global map with the number of CGCM2.3.2 realizations projecting change in the dominant direction (the direction of change projected by the majority of GCMs) is shown. For 55% of the globe all five realizations agree on the projected direction of change, for 81% of the globe at least four realizations agree on the direction of change and for the remaining 19% only three realizations are consistent.

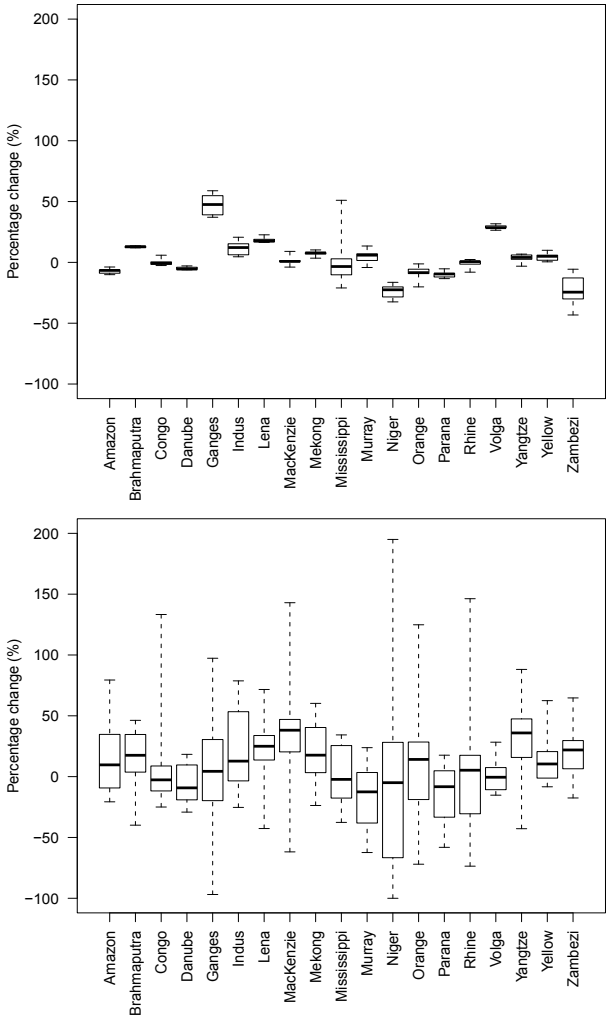


Figure A.6 a. Boxplots of changes projected by the five available realizations of the GCM CGCM 2.3.2 for the A1B scenario for 2100. Whiskers mark the maximum and minimum projected changes, boxes span the quartile range and horizontal dashes represent the median of projected changes. **b.** Same as a. but now for the changes projected by the ensemble of 12 GCMs.

From these results two conclusions can be drawn. Firstly, Fig. A.6 shows the inter-model uncertainty is much larger than the intra-model uncertainty, at least for the GCM data we have at our disposal. And secondly, for the majority of catchments the projected directions of change are consistent for the five realizations. This indicates that including different numbers of realizations for the individual GCMs in our ensemble would result in overweighting the direction of change projected by the GCMs with multiple realizations. Therefore we restricted ourselves to a single realization for each of the twelve GCMs included in the ensemble.

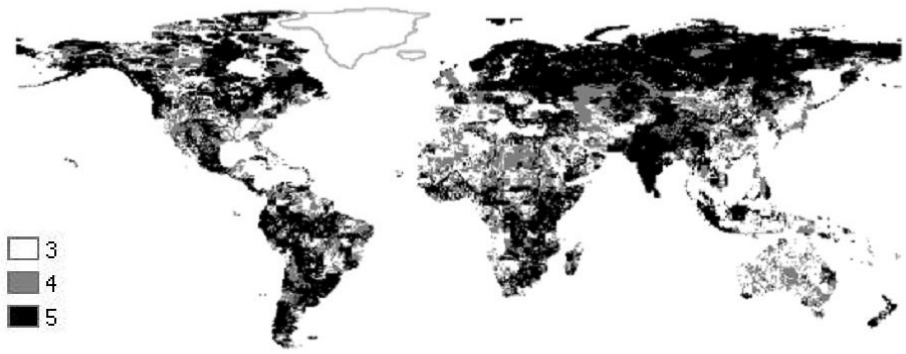


Figure A.7 Map showing the number of realizations of CGCM2.3.2 projecting mean change in the dominant direction. Black indicates regions where the five realizations project changes in the same direction, grey indicates regions where four realizations project similar directions of change and in the white regions only three realizations project the same direction of change. The dominant direction is the direction of change projected by the majority of models.

B Selecting the optimal method to calculate daily global reference potential evaporation from CFSR reanalysis data

This appendix is based on:
Sperna Weiland, F.C. , C. Tisseuil, H. H. Dürr, M. Vrac and L. P. H. van Beek (2011), Selecting the optimal mthod to calculate daily global reference potential evaporation from CFSR reanalysis data, Hydrology and Earth System Sciences Discussions 8, pp. 1-43, doi:10.5194/hessd-8-1-2011.

Abstract

Potential evaporation (PET) is one of the main inputs of hydrological models. Yet, there is limited consensus on which PET equation is most applicable in hydrological climate impact assessments. In this study six different methods to derive global scale reference PET time series from CFSR reanalysis data are compared: Penman-Monteith, Priestley-Taylor and original and modified versions of the Hargreaves and Blaney-Criddle method. The calculated PET time series are 1) evaluated against global monthly Penman-Monteith PET time series calculated from CRU data and 2) tested on their usability for modeling of global discharge cycles.
The lowest root mean squared differences and the least significant deviations (95% significance level) between monthly CFSR derived PET time series and CRU derived PET were obtained for the cell specific modified Blaney-Criddle equation. However, results show that the modified form is likely to be unstable under changing climate conditions and less reliable for the calculation of daily time series. Although often recommended, the Penman-Monteith equation did not outperform the other methods. In arid regions (e.g., Sahara, central Australia, US deserts), the equation resulted in relatively low PET values and, consequently, led to relatively high discharge values for dry basins (e.g., Orange, Murray and Zambezi). Furthermore, the Penman-Monteith equation has a high data demand and the equation is sensitive to input data inaccuracy. Therefore, we preferred the modified form of the Hargreaves equation, which globally gave refer-

ence PET values comparable to CRU derived values. Although it is a relative efficient empirical equation, like Blaney-Criddle, the equation considers multiple spatial varying meteorological variables and consequently performs well for different climate conditions. In the modified form of the Hargreaves equation the multiplication factor is uniformly increased from 0.0023 to 0.0031 to overcome the global underestimation of CRU derived PET obtained with the original equation. It should be noted that the bias in PET is not linearly transferred to actual evapotranspiration and runoff, due to limited soil moisture availability and the influence of precipitation.
The resulting gridded daily PET time series provide a new reference dataset that can be used for future hydrological impact assessments in further research or, more specifically, for the statistical downscaling of daily PET derived from raw GCM data.

B.1 Introduction

Climate change is likely to induce alterations in the hydrological cycle (IPCC, 2007b). To assess and quantify the possible changes, multiple hydrological impact studies have been conducted on the local, continental and global scale, the latter being of interest in this study. In addition to temperature and precipitation, evapotranspiration is one of the main components of the water balance at the land surface and required as input for hydrological models used in impact studies (Kay and Davies, 2008; Oudin et al., 2005). Whereas precipitation and temperature model input data are usually at hand, actual evapotranspiration (AET) is seldomly monitored. Furthermore, within hydro-climatic change studies, raw General Circulation Model (GCM) data and statistically or dynamically downscaled GCM data are frequently used to force global hydrological models (Sperna Weiland et al., 2011c). Yet, these datasets often lack AET data (PCMDI, 2010). Moreover, we prefer the calculation of potential evaporation (PET) from other GCM meteorological variables and the derivation of AET with a hydrological model, over using GCM AET directly. This because within GHMs, AET is calculated on a higher grid resolution and processes related to transpiration and soil moisture are schematized in more detail (Sperna Weiland et al., 2011b). In addition, AET of GCMs is often biased due to, amongst others, biases in precipitation and radiation (Mahanama and Koster, 2005; Elshamy et al., 2009). As a consequence, creating reference daily PET time-series that can be used as meteorological input for hydrological models is crucial in order to derive consistent AET, runoff and discharge (Oudin et al., 2005).
There is limited consensus on which PET equation is most applicable in global hydrological impact studies. Several studies illustrated that the selection of a method can actually determine the direction of projected change in future water availability (Boorman, 2010; Kingston et al., 2009; Arnell, 1999c). Note however, that the influence of biases and uncertainties in PET usually decreases while moving within the hydrological model chain from PET to AET to discharge, as water availability becomes limited (Vörösmarty et al., 1998a).

Generally, the Penman-Monteith equation is considered as the standard (Hargreaves et al., 2003; Droogers and Allen, 2002; Gavilán et al., 2006). This equation is preferred over simpler temperature based methods in climate impact studies as it includes the effect of changes in multiple atmospheric variables (Kay and Davies, 2008; Arnell, 1999c; Kingston et al., 2009). On the other hand, Penman-Monteith has a high input data requirement and, especially when input data is subject to inaccuracy, as is the case with reanalysis and GCM data, quality of the resulting PET might decrease (Oudin et al., 2005). Therefore the Hargreaves equation is often used as an alternative (Hargreaves and Samani, 1985; Hargreaves et al., 2003). Contrary to simpler empirical temperature based equations, like Blaney-Criddle (Blaney and Criddle, 1950), Hargreaves also considers the influence of humidity by an approximation using the diurnal temperature range. The equation is applicable in a variety of climatic conditions and no local calibration is needed. Globally, agreement between PET derived with Penman-Monteith and Hargreaves has been found to be reasonable (Droogers and Allen, 2002). Several studies tried to improve the Hargreaves equation by including aridity functions and wind data. However, the influence of these parameters on the quality of the calculated evaporation was limited (Hargreaves et al., 2003 and references therein).

There is an ongoing discussion on the reliability of the different methods, especially when working with reanalysis or GCM data. It could be questioned whether a physically based Penman-Monteith like formula with a high sensitivity to inaccuracy in input data would be more reliable than a more empirical equation with less input parameters and therefore smaller spread in uncertainties (Kingston et al., 2009). Or, whether the Priestley-Taylor equation, in which the aerodynamic term of the Penman-Monteith equation is replaced by an empirical multiplier, could better be used (Weiβ and Menzel, 2008; Lu et al., 2005). In addition, in order to reduce the computation time required for both calculation of PET and downscaling of the required input variables, the Blaney-Criddle equation might be useful as well. The Blaney-Criddle equation is an empirical temperature-based equation and it has given results comparable to the other PET methods (Oudin et al., 2005; Blaney and Criddle, 1950). Yet, Jensen (1966) showed that the climate dependency of the empirical Blaney-Criddle equation disables its application in multiple different climate zones. To overcome this problem, Ekström et al. (2007) presented three different methods to spatially bias-correct Blaney-Criddle PET. In this study we adopt their best performing method as one of the six methods we investigate. In addition, we investigate a modification of the more physically based empirical Hargreaves equation (Droogers and Allen, 2002), which might be less sensitive to climate conditions since a larger number of spatial varying meteorological variables are considered as input.

Seeing the variety of equations which all have their pros and cons, there is still a need to properly select the PET equation to be used within hydrological modeling studies, especially at continental to global scales. This selection will depend both on the meteorological dataset and the study area of interest. For the global analysis in this study, the

Climate Forecast System Reanalysis (CFSR) dataset is used (Saha et al., 2006; Saha et al., 2010) because it has a high spatial (~0.3°x~0.3° degrees) and temporal (6-hourly) resolution and it supersedes the frequently used earlier US NCEP/NCAR (National Center for Environmental Prediction / National Center for atmospheric research) reanalysis data (Kalnay et al., 1996). The dataset contains the required daily atmospheric fields to calculate and compare a range of PET equations and it can be used for the downscaling of raw GCM data to higher spatial resolution at a daily time-scale.

Daily input of PET equations increasingly becomes available and daily PET is often required as input for hydrological models. Therefore we decided to focus on calculation of daily PET time series using daily values of the required atmospheric variables instead of calculating monthly PET and downscaling this to daily values (for example based on temperature) as has frequently been done before (Sperna Weiland et al., 2010; Arnell, 2011), since for a correct temporal downscaling not only air temperature is important. The diurnal temperature range, vapor pressure, and incoming shortwave radiation influence PET as well and should preferably not be neglected in a temporal downscaling procedure. Yet, considering all these variables would result in complex empirical temporal downscaling relations.

This study is a preliminary step for the assessment of future global hydrological consequences of climate change. For this assessment the global hydrological model PCR-GLOBWB will be run with downscaled GCM data. A first analysis illustrated the value of directly downscaling PET derived from raw GCM data, based on the reference PET time series created in this study. The 'direct' downscaling approach was preferred over downscaling of the individual GCM input variables of the PET equation, since in independent procedures inconsistencies between the atmospheric input variables op the PET equations can be introduced (Piani et al., 2010). For future downscaling, PET will first be calculated from raw GCM data at the original GCM resolution. In a second step, the GCM derived PET will be downscaled with the here created reference CFSR PET time series, hereby generating bias-corrected GCM PET time series at the resolution of the hydrological model.

The main goal of this study is the construction of a global gridded dataset of reference PET at high spatial (0.5 degree) and temporal (daily) resolution from CFRS reanalysis data. Not only should the constructed PET dataset show high resemblance with the measurement based monthly PET time-series derived from the CRU datasets which are often considered as a standard (New et al., 2000; New et al., 1999, Droogers and Allen, 2002; IPCCb, 2007). In addition the dataset should be a reliable reference for the statistical downscaling of daily PET time series calculated from raw GCM data, which can be used as input for hydrological climate impact studies.

We will first compare six PET equations for the creation of daily PET time series; Penman-Monteith, Hargreaves, Priestley-Taylor, Blaney-Criddle and modifications of the Hargeaves and Blaney-Criddle equation. All equations are applied to the CFSR reanalysis dataset for the period 1979-2002 and the resulting PET time series are evaluated

against Penman-Monteith PET derived from the CRU datasets (New et al., 2000; New et al., 1999). In a second step, the transfer of bias in PET to modeled AET, runoff and discharge is assessed by inter-method comparison and comparison of modeled river discharge with discharge observations.

B.2 Data and methods

B.2.1 CFSR reanalysis data

The CFSR dataset is a reanalysis product which is developed as part of the Climate Forecast System (Saha et al., 2006; Saha et al., 2010) at the National Centers for Environmental Prediction (NCEP). The CFSR dataset became available in 2010 and supersedes the previous NCEP/NCAR reanalysis dataset which has been widely used in downscaling studies (e.g. Michelangeli et al., 2009; Maurer et al., 2010; Wilby et al., 2002). At this stage the CFSR dataset spans the period 1979 to present and has a resolution of approximately 0.25 degrees around the equator to 0.5 degrees beyond the tropics (Higgins et al., 2010). In this study, 6-hourly temperature, radiation, air pressure and wind data were averaged to a daily time-step for the period 1979-2002. These daily time series were then interpolated to a regular 0.5 degrees grid using bilinear interpolation in order to calculate PET at the grid resolution of the global water balance model PCR-GLOBWB.

B.2.2 CRU reference potential evaporation

For validation, reference historical PET time series were calculated from the CRU datasets with the FAO recommended Penman-Monteith equation (Monteith, 1965; Allen et al., 1998). Temperature, vapor pressure, diffusivity and net incoming radiation were retrieved from the CRU TS2.1 monthly time series (New et al., 2000). Cloud cover and wind speed were obtained from the monthly climatology, CRU CLIM 1.0 (New et al., 1999) because monthly CRU TS2.1 time series are not provided for these variables.

B.2.3 Potential evaporation equations

Within this study, daily PET time series derived from six different PET equations were compared. The equations ranged from the physically based Penman-Monteith (PM) equation, to the radiation- and temperature-based Hargreaves (HG) and Priestley-Taylor (PT) equations, to the simple temperature-based Blaney-Criddle (BC) equation and additional modified forms of the Hargreaves and Blaney-Criddle equations (Table B.1). The BC equation was applied in its original form (BCorig) and in a re-calibrated form (BCrecal) following Ekström et al. (2007). In this modified BC equation, the multiplicative and additive coefficients (e.g., 0.46 and 8) have been re-calibrated to cell-specific values (see the resulting coefficient values in Fig. B.1). This was done by linearly regressing the cell specific long-term average mean monthly CFSR temperature to the

CRU derived long-term average monthly PET for the complete period with overlapping data available for the two datasets (1979-2002). The slopes and intercepts of this linear regression exercise were used to calculate the coefficient values. For the empirical BC equation, which considers only limited meteorological variables, a cell specific re-calibration was preferred (this is also illustrated by the large spatial variation in bias between BC PET derived from CFSR data and reference PM PET derived from CRU data, as will be presented in the results section).

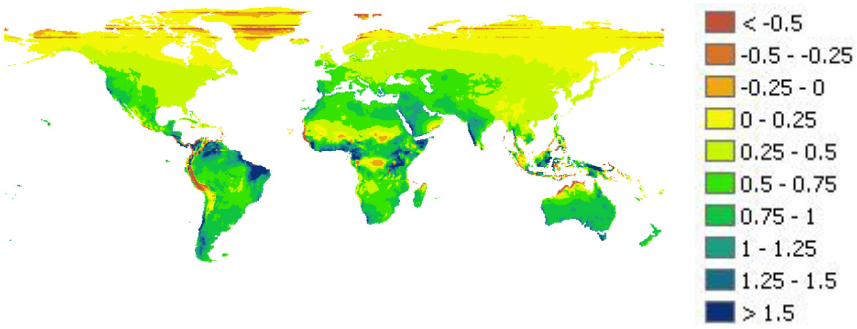
A simpler globally uniform modification was applied to the HG equation following Allen (1993) and Droogers and Allen (2002). The Hargreaves equation is an efficient empirical equation with low input data demand. Yet, the equation considers spatial variation in climate conditions in mode detail as it also includes the daily temperature range and spatial radiation pattern. PET derived from the CFSR dataset with the original HG (HGorig) equation underestimaties CRU PET with a small spatial variability, as will be shown in the results section. To increase the HG PET globally, we increased the multiplication factor for all grid cells uniformly from 0.0023 to 0.0031 by linear fitting long term average monthly HG PET against long term average monthly CRU PET (HGrecal). To this end, the multiplication factor was varied with intervals of 0.0001 until the lowest global average RMSD value was obtained for the monthly average PET time series.

Table B.1: Potential evaporation equations

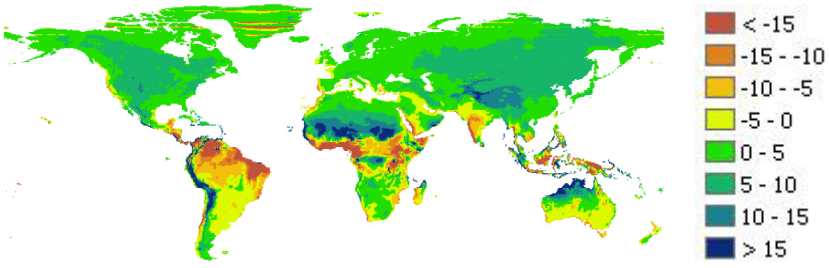
Method	*Acronym*	*Equation*	*Reference*
Penman-Monteith	PM	$PET = \frac{\Delta(R_n - G) + \rho_a c_p \frac{(e_s - e_a)}{r_a}}{\lambda_v \Delta + \gamma(1 + \frac{r_s}{r_a})}$	Monteith (1965)
Hargreaves	HGorig	$PET = 0.0023 \cdot R_a \cdot (\overline{T} + 17.8) \cdot TR^{0.50}$	Hargreaves and Samani (1985)
Modified Hargreaves	HGrecal	$PET = 0.0031 \cdot R_a \cdot (\overline{T} + 17.8) \cdot TR^{0.50}$	This study
Priestley-Taylor	PT	$PET = \alpha \frac{\Delta R_n}{\lambda_v (\Delta + \gamma)}$	Priestley and Taylor (1972)
Blaney-Criddle	BCorig	$PET = p(0.46\overline{T} + 8)$	Blaney and Criddle (1950)
recalibrated Blaney-Criddle	BCrecal	$PET = p(a\overline{T} + b)$	Ekström et al. (2007)

B.2.4 Global hydrological modelling

The global water balance was modelled with the global hydrological model PCR-GLOBWB. For a detailed description and validation of the model, see Van Beek et al. (2011), Van Beek and Bierkens (2009) and Sperna Weiland et al. (2010). Each model cell, with a resolution of 0.5 degrees, consists of two vertical soil layers and one underlying groundwater reservoir. Sub-grid parameterization is used for the schematization of



a.



b.

Figure B.1: Cell specific values of the coefficients in the re-calibrated Blaney-Criddle equation. The values in Fig. a replace the number 0.46 and the values in Fig. b replace the number 8 in the original Blaney-Criddle equation: $PET = p(0.46\overline{T} + 8)$

surface water, short and tall vegetation and for calculation of saturated areas for surface runoff as well as interflow. Water enters the cell as rainfall and can be stored as canopy interception or snow. Snow is accumulated when temperature is below 0°C and melts when temperature is higher. Melt water and throughfall are passed to the surface, where they either infiltrate in the soil or become surface runoff. Exchange of soil water is possible between the soil and groundwater layers in both up- and downward direction, depending on soil moisture status and groundwater storage. Total runoff consists of non-infiltrating melt water, saturation excess surface runoff, interflow and base flow. Within the hydrological model, AET is derived from PET time series. The total AET flux exists of plant transpiration and bare soil evaporation, which spatially depend on the presence of crop types and soil moisture conditions. Within the model in a first step,

reference PET is transferred to crop specific and bare soil PET values by multiplication with the minimum crop factor (for bare soil) or with the monthly climatology of crop factors (Van Beek, 2008) which have been projected on the 0.5 degrees hydrological model grid for the specific crop types:

$$ES_o = k_s PET \quad (1)$$

$$T_o = k_c PET \quad (2)$$

where PET is reference PET (m/day), k_s is the 'crop factor' used for bare soil, ES_0 is potential bare soil evaporation (m/day), k_c is the monthly crop factor and T_0 is potential crop specific transpiration (m/day).
Reduction of potential bare soil evaporation and potential plant transpiration to AET depends on soil moisture storage. For the saturated fraction of the soil no reduction occurs, except that the rate of potential evaporation can not exceed the saturated hydraulic conductivity. Potential bare soil evaporation of the un-saturated fraction is limited by the un-saturated hydraulic conductivity. Transpiration only occurs for the un-saturated fraction of the soil and depends on the total available soil moisture storage in the models soil layers (Van Beek, 2008).
For each daily time-step the water balance, and its resulting runoff and AET fluxes, are computed for all model cells. The cell specific runoff is accumulated and routed as river discharge along the drainage network taken from the global Drainage Direction Map (DDM30; Döll and Lehner, 2002) using the kinematic wave approximation of the Saint-Venant equation.

B.2.5 Statistical validation

The six PET time series derived from CFRS data were validated for the period 1979 to 2002 against monthly CRU based Penman-Monteith PET time series (CRUPM) and compared with each other, using six statistical quantities:

1) global maps of bias in long-term average annual means:

$$BIAS = \overline{PET}_{CFSR} - \overline{PET}_{CRU} \quad (3)$$

where $\overline{PET}_{CFSR}$ refers to annual average PET calculated from the CFSR dataset using one of the six equations (Table B.1) and $\overline{PET}_{CRU}$ refers to the annual average PET calculated from the CRU dataset with the Penman-Monteith equation.

2) global maps of areas with significant differences between CRU and CFSR derived PET, AET, runoff and discharge. These maps indicate for all six methods for all hydrological quantities of interest (PET, AET, local runoff and discharge) whether the CFSR derived values deviate significantly from the CRU derived values. The map comparison

is restricted to the deviations between results of PCR-GLOBWB runs forced with the full CFSR dataset (e.g. CFSR precipitation (PR), temperature (TAS) and PET calculate with one of the six equations) and the PCR-GLOBWB run forced with the full CRU dataset downscaled to daily values with the CFSR dataset. For the evaluation of differences in station discharge results of PCR-GLOBWB runs forced with CRU PR and TAS and CFSR PET have been included as well, this in order to exclude the influence of biases in CFSR PR from the analysis. Significance of differences between CRU and CFSR derived values have been quantified with the Welch's t-test for a significance level of 95%.

$$t = \frac{\overline{X}_{CRU} - \overline{X}_{CFSR}}{\sqrt{\frac{s_{CRU}^{\ 2}}{m_{CRU}} + \frac{s_{CFSR}^{\ 2}}{m_{CFSR}}}} \tag{4}$$

Where $\overline{X}_{CRU}$ is the long term annual average value calculated from the CRU dataset and $\overline{X}_{CFSR}$ is the average value calculated from the CFSR dataset for one of the six equations, S_{CRU} is the standard deviation of the 24 CRU derived annual values and S_{CFSR} is the standard deviation of the 24 CFSR derived annual values, m_{CFSR} and m_{CFSR} are the number (24) of annual average values for both datasets.

3) global maps of cell specific root mean squared differences (RMSD; m/day) of the monthly time series (Eq. 5) have been created. These maps give an indication of regional performance on the smallest time-scale at which validation data is provided:

$$RMSD = \sqrt{\frac{\sum_{i=1}^{N} (PET_{CRU_i} - PET_{CFSR_i})^2}{N}} \tag{5}$$

in this equation PET_{CFSR} refers to the monthly PET calculated from the CFSR data set, PET_{CRU} refers to the monthly PM-based PET calculated from the CRU dataset, i is the month number and N is the total number of months (N=288).
4) for each season individually, the mean (Eq. 6) of the global cell specific seasonal RMSD values (as in Eq. 5, except for monthly values being replaced by seasonal values) quantifies the overall performance of the PET methods. The standard deviation (Eq. 7) of the RMSD values quantifies the spatial variability in performance:

$$\overline{RMSD} = \frac{\sum_{g=1}^{G} RMSD_g}{G} \tag{6}$$

$$SDV_{RMSD} = \sqrt{\frac{\sum_{g=1}^{G}(RMSD_g - \overline{RMSD})^2}{G}} \quad (7)$$

where $\overline{RMSD}$ (m/day) is the mean of all cell specific seasonal RMSD values ($RMSD_g$), g is the grid cell number, G is the total number of grid cells and SDV_{RMSD} (m/day) is the standard deviation of all cell specific seasonal RMSD values.

5) global maps with long-term average and seasonal average PET, AET and runoff have been calculated to illustrate the differences between methods. Biases present in the PET time series are not linearly transferred to AET and runoff due to limiting soil moisture conditions. Therefore, the impact of the differences between PET methods on resulting AET and runoff fluxes obtained by hydrological modeling are also analyzed and mapped.

6) to quantify the variation between methods, cell specific values of the coefficient of variation (CV) have been calculated from the realizations for the six different PET equations:

$$CV = \frac{\sum_{g=1}^{G} \frac{\sqrt{\frac{1}{E}\sum_{e=1}^{E}(PET_{e,g} - \overline{PET}_g)^2}}{\overline{PET}_g}}{G} \quad (8)$$

Where $\overline{ETP}_g$ is the average of PET calculated with the 6 equations for the specific cell g. $PET_{e,g}$ is the PET calculated for the e-th equation for cell g, E is the total number of equations (6), G is the total number of grid cells.
Global average CV values have also been calculated for PET, AET, runoff and discharge and basin specific CV values have been calculated from annual average river discharge.

7) performance of the different methods for the reproduction of correct AET amounts is more explicitly evaluated by comparing long-term average modeled river discharge with discharge observations. To this end annual average discharge is calculated for a selection of 19 large river (Sperna Weiland et al., 2010). Within the bar-charts the following annual average discharges are shown: 1) Discharges derived from the full CFSR dataset for all six PET equations (for each individual river the group of bars on the left of the charts). 2) As an additional reference discharge calculated with an PCR-GLOBWB run forced with the full CRU forcing, e.g. CRU PR, TAS and PET downscaled to daily values based on the CFSR daily distributions shown in dark grey (hereafter called full CRU run; for the temporal downscaling procedure see Sperna Weiland et al. (2010)).

3) Discharge calculated with a PCR-GLOBWB run forced with a combination of CFSR derived PET and temporally downscaled CRU PR and TAS, in this run the influence of possible PR biases present in the CFSR dataset have been excluded. The results of these runs are shown by the group of bars on the right for each river. 4) Finally, for validation of all modeled discharge values, observed discharge obtained from the GRDC (GRDC, 2007) is modified by adding an estimation of water use (Wada et al., 2010; Sperna Weiland et al., 2010) the resulting corrected observed discharge is included in the charts in black.

B.3 Results

B.3.1 Global reference potential evaporation

Long-term average bias

The biases in long-term average CFSR annual PET from CRU PM PET show large differences for the six PET equations (Fig. B.2). Penman-Monteith PET derived from CFSR data (CFSRPM) underestimates Penman-Monteith PET derived from CRU data (CRUPM) in arid regions (e.g. the Sahara, Central Australia and the southwest of the US) and slightly overestimates CRUPM in southeast Asian Islands and parts of the Amazon basin (Fig. B.2a). The Priestley-Taylor equation (CFSRPT) highly overestimates CRUPM in the Amazon basin, Central Africa and Indonesia, whereas underestimations similar to those of the CFSRPM are present in the Sahara and parts of Australia (Fig. B.2d). The standard Hargreaves equation (CFSRHGorig) underestimates CRUPM globally (Fig. B.2b). By increasing the multiplication factor of the HG equation from 0.0023 to 0.0031, the lowest global average RMSD was obtained (Fig. B.2e). Similar increases of this coefficient have also been proposed by Droogers and Allen (2002) and Allen (1993). PET calculated with the Blaney-Criddle equation (CFSRBC) is too high for almost the entire world (Fig. B.2c). Overestimations are especially large in Central Africa and Central South-America. Yet, PET calculated with the re-calibrated Blaney-Criddle equation from the CFSR dataset (CFSRBCrecal) results in the highest similarity with CRUPM (Fig. B.2f). For illustrational purpose, global maps of absolute PET values for the different methods are shown in the supplementary information (Fig. S.1).

RMSD of monthly time series

The lowest monthly RMSD values are obtained for the BCrecal equation and the HGrecal equation (Fig. B.3e and B.3f). The pattern of RMSD values of PM and PT (Fig. B.3a and B.3d) are comparable, although the RMSD is slightly higher for CFSRPT over the Amazon basin and Central-Africa. The similarity of PET derived from the two equations is caused by the radiation term present in both equations. Overall, the global maps in Fig. B.3 show that performance of the PT, HGorig and BCorig equations is low, whereas the HGrecal equation and especially the BCcal equation perform well.

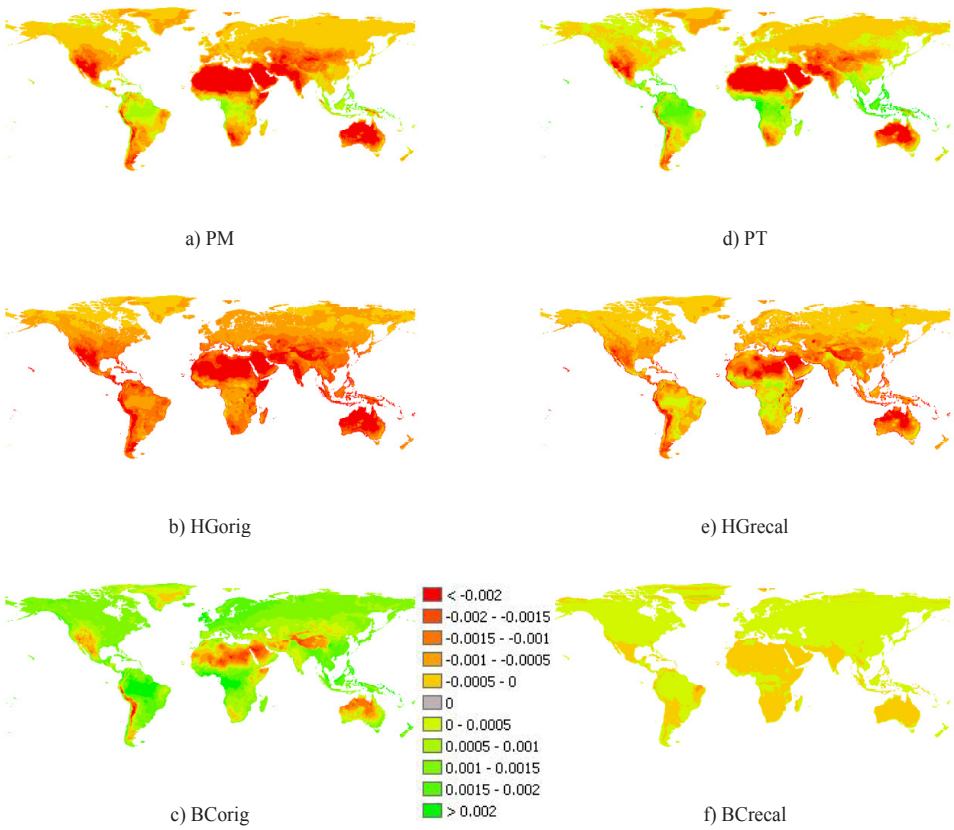


Figure B.2: Global maps with annual average bias of CFSR estimated daily reference potential evaporation (PET; m day-1) from annual average CRU Penman-Monteith reference PET. In the left column bias in PET obtained with the Penman-Monteith (PM), the standard Hargreaves (HGorig) and Blaney-Criddle (BCorig) method are displayed. In the right column bias obtained with Priestley-Taylor (PT), Hargreaves with increased multiplication factor (HGrecal) and the re-calibrated Blaney-Criddle equation (BCrecal) are displayed.

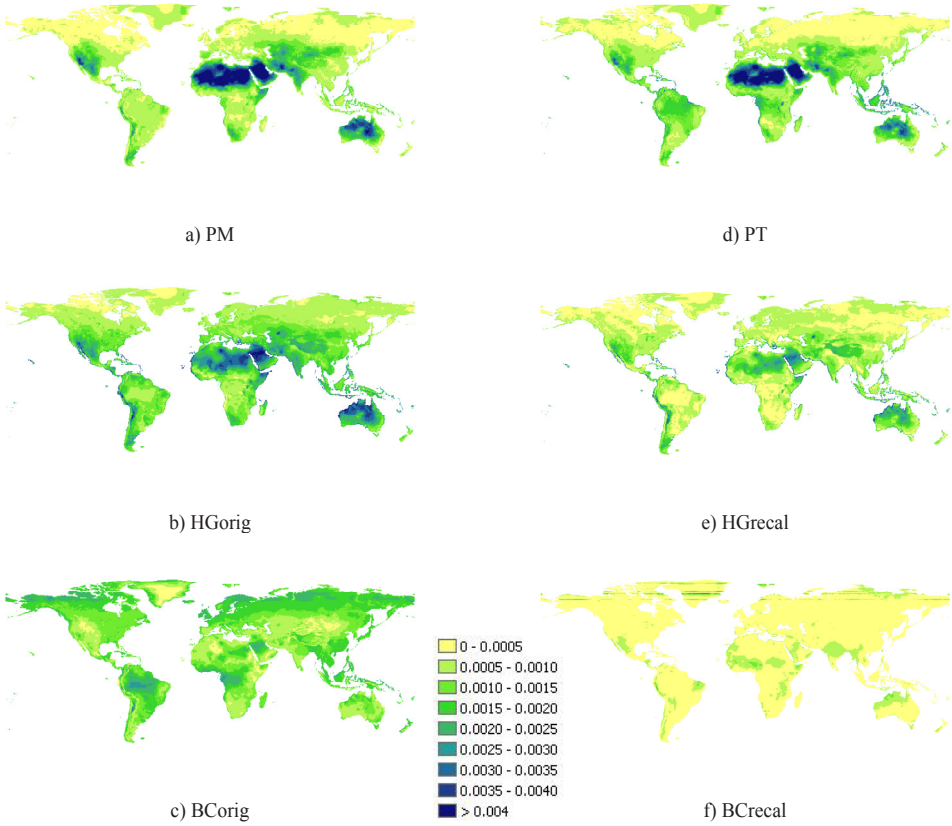


Figure B.3: Global maps with cell specific root mean square differences (RMSD) calculated between the CFSR derived monthly PET time series and the monthly PET timeseries derived from the CRU dataset with the Penman-Monteith equation. In the left column from top to bottom Penman-Monteith (PM), the original Hargreaves method (HGorig) and Blaney-Criddle equation (BCorig) and in the right column Priestley-Taylor (PT), Hargreaves with increased multiplication factor (HGrecal) and the re-calibrated Blaney-Criddle equation (BCrecal) are displayed.

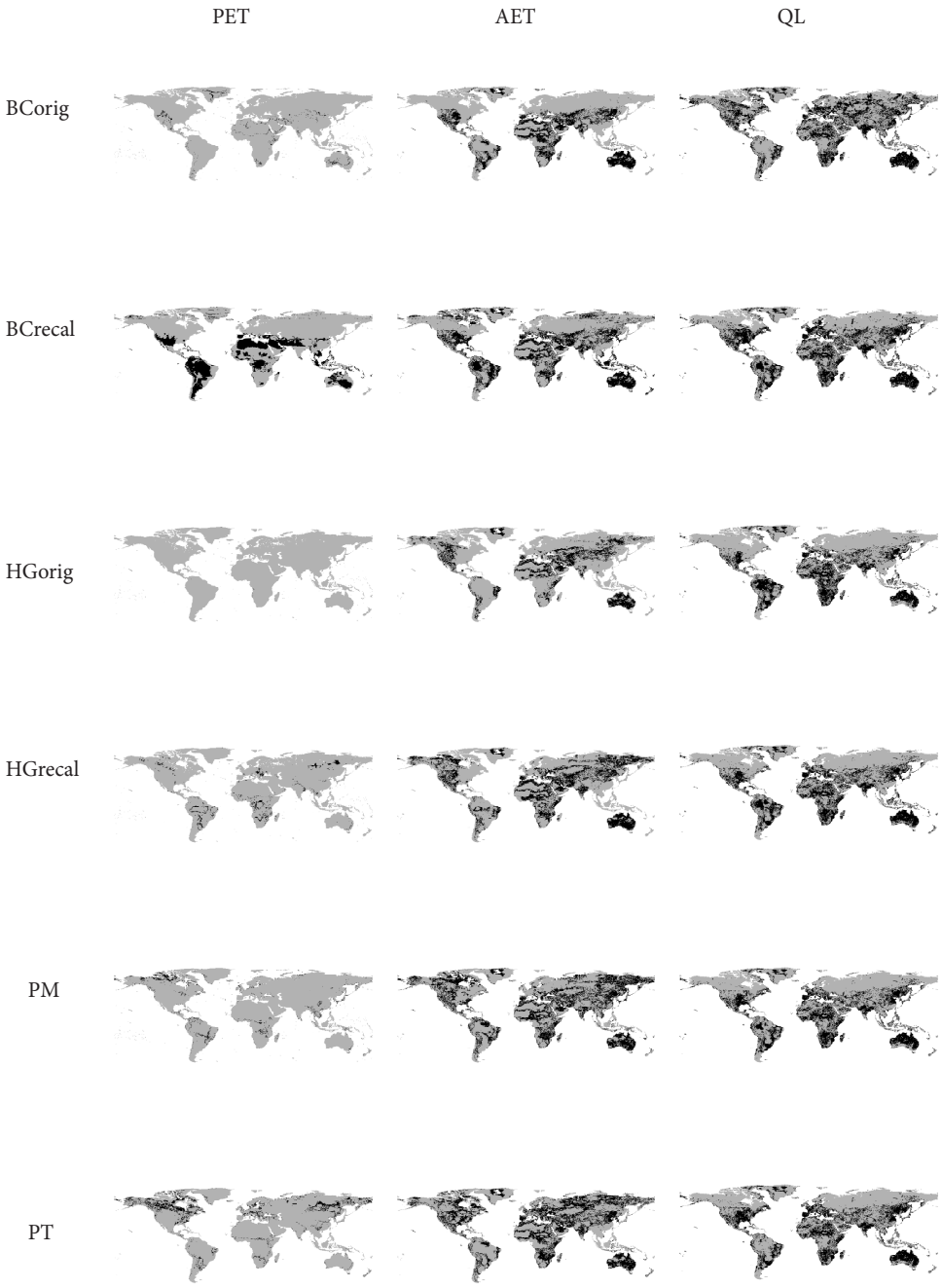


Figure B.4: Maps showing areas where CFSR derived PET, AET and local runoff (QL) significantly deviate from CRU derived values (in grey) and areas where annual average values are similar (in black) for all six PET equations.

Significance of deviations

In Fig. B.4 the significance of differences between the annual average PET, AET and local runoff derived from CFSR data (with any of the PET equations) and annual average values of the same variables derived from CRU data is indicated. Black areas correspond to regions where annual averages of CFSR and CRU PET derived values are equal. Large regions with CFSR PET values similar to CRU PET only occur for the BCcal method. The BCorig and HGorig equations obviously show least significant resemblance with CRU PET. While moving from PET to AET to local runoff (QL) the areas with similar CRU and CFSR derived annual average values increase in size. As difference between the different PET methods decrease due to limited soil moisture availability and the influence of PR on local runoff and discharge.

Global mean seasonal RMSD

The global mean and standard deviation of the cell specific RMSD values of seasonal PET, derived from the difference between CRUPM and the different CFSR PET time series (Eq. 5), have been calculated for the individual seasons to quantify the global seasonal performance and its spatial variability (Fig. B.5). For all four seasons the BCrecal equation gave the lowest global average RMSD values and the HGrecal equation performed second best (Fig. B.5). This confirms the performance improvement obtained by re-calibration of the two equations. Yet, due to the large spatial variability of the re-calibrated BC coefficient values (Fig. B.1), the stability of the equation under changing climate conditions is not guaranteed.

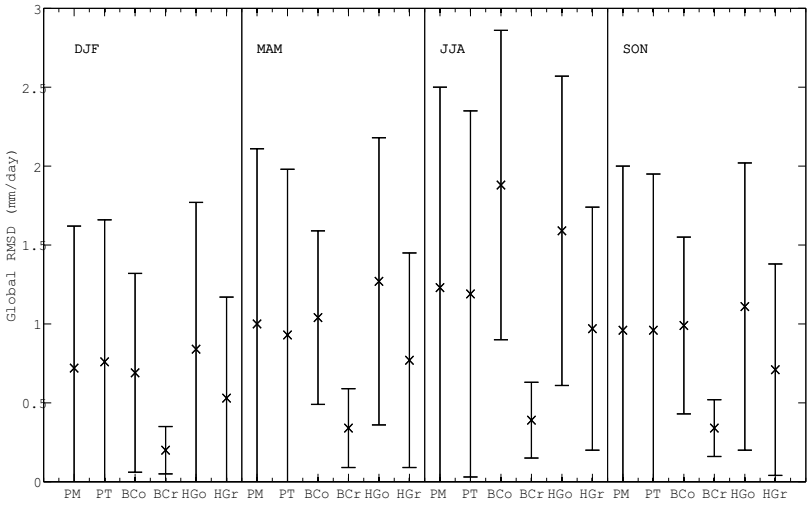


Figure B.5: : Error plots with mean (crosses) and standard deviation (horizontal bars) of cell specific RMSD values (m/day) calculated for the four seasons individually from the differences between the CFSR derived seasonal potential evaporation values and the CRU based seasonal potential evaporation. Abbreviations have been used for BCorig (BCo), BCrecal (BCr), HGorig (HGo) and HGrecal (HGr).

In addition, the daily BCrecal PET values span a relatively small range. The extreme daily values are modest compared to daily PET values derived with the other equations (for brevity, the full analysis has not been included, but as an example cumulative distribution functions (CDFs) of daily PET values are given for the MacKenzie, Amazon, Rhine and Zambezi river basins in Fig. B.6). This may be a result of the use of the equation on a daily time scale instead of the monthly time scale for which the equation was originally designed.

For the DJF, MAM and SON seasons, RMSD values of the PM, PT and BC methods are comparable. However, performance of the original BC equation is especially poor for the JJA season (Northern hemisphere boreal summer), the season in which evaporation has the largest influence on the water balance. For all seasons, except JJA, highest RMSD values are obtained with the HGorig equations.

The standard deviation of the cell specific RMSD values is an indication of the spatial variability in performance (Fig. B.5, error bars). The highest standard deviations are obtained from the PM and PT equation. The standard deviation of the HGorig equation is slightly lower, indicating a more constant performance in space. Increasing the multiplication factor in the HGorig equation to 0.0031 (HGrecal) did not only result in a lower global mean RMSD, but also decreased the spatial variability in performance. The highest spatially consistent performance was obtained for the BCrecal equation.

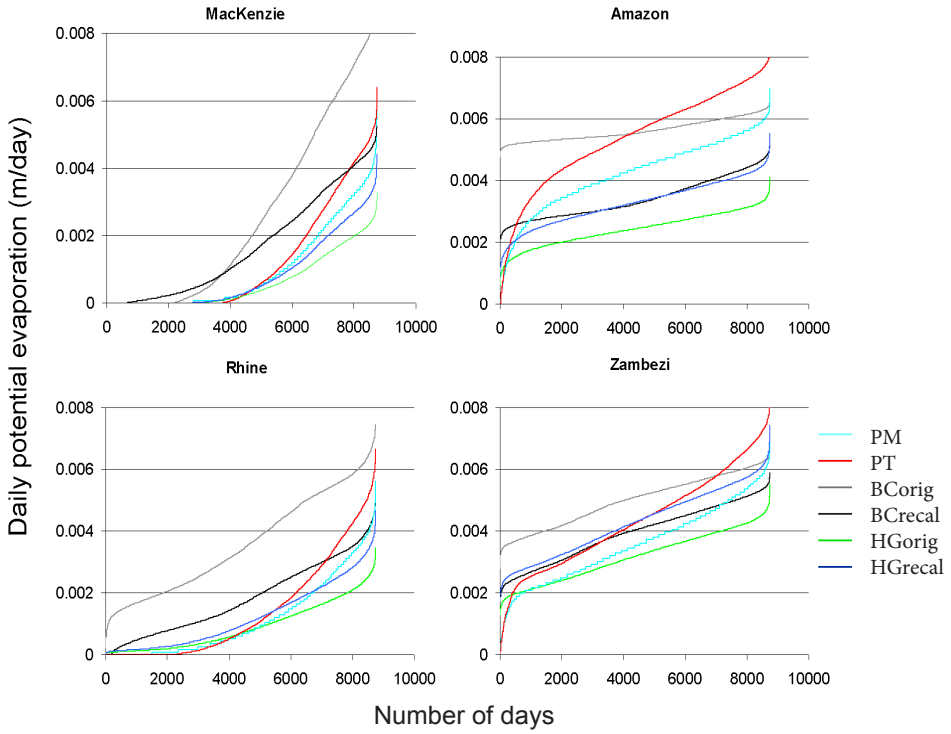


Figure B.6: CDFs of daily potential evaporation (m/day) for a selection of catchments; MacKenzie, Amazon, Rhine and Zambezi

B.3.2 Impact of different PET equations on actual evapotranspiration and runoff

For the evaluation of CFSR PET, global PET time series derived from CRU data with the PM equation, could be used as a reference. Unfortunately, there are no reference global gridded time series of AET available for the evaluation of the impact of the different PET equations on modeled discharge. Vörösmarty et al. (1998a) apply an approximation of observed AET by subtracting observed runoff from observed PR. Yet, as Vörösmarty et al. (1998a) already stated, this approximation is only valid in areas with little water regulation or abstractions and reliable PR and discharge measurements. We obtained negative AET values for a number of basins, as was also the case for several locations in their study, and concluded that the method was not reliable when being used in combination with our global datasets. Therefore, the CFSR derived AET and runoff maps are not explicitly validated. Here, only a comparison between methods is made. In section B.3.3 the bias in AET is evaluated by comparison of the resulting modeled discharge with observed GRDC discharge. Although illustrative, it should be noted that this comparison might be flawed by discharge measurement and hydrological model errors.

Variation between methods

Biases in CFSR PET from CRU PET are reduced while moving from PET to AET and runoff since AET is limited by soil moisture conditions. Consequently AET biases, are lower than biases in PET. Limitation by soil moisture deficits mainly occurs in arid regions (e.g., the Sahara, Central Australia and the South-Western US) or in the dry seasons. Global maps with annual average AET and runoff show the impact of deviations in PET on AET and runoff (Fig. B.7.1 and B.7.2). Globally the variability between the six different methods is smaller for AET than for PET, as can be seen from the cell specific values for the coefficient of variation (CV) obtained from the PET values calculated for the six different methods (Fig. B.8). The global cell average coefficient of variation (CV) for PET is 0.42, whereas for AET and runoff the CV values are respectively 0.25 and 0.27. High CV values for PET and AET are obtained for Northern regions and the Himalayas. Yet, CV values for runoff are low in these regions due to the relatively low absolute amount of AET related to the low air temperature and the large influence of PR. High CV values for PET are also present in the Sahara and central Australia as a result of the large underestimations of CRUPM PET in these regions by CFSRPT and CFSRPM PET. However, soil moisture is limited in these dry regions and AET amounts are comparably low for the different methods resulting in low CV values. Over South-East Asia, the Eastern US, parts of Europe, Russia and the Amazon and Congo basins, high CV values of 0.2 to 0.3 are obtained for both PET and AET. Here, except for South-East Asia, where there is a strong influence of the Monsoon precipitation, AET amounts are high and, the high CV values for AET are translated to high CV values for runoff. Only for those regions where both the CV of runoff generated by

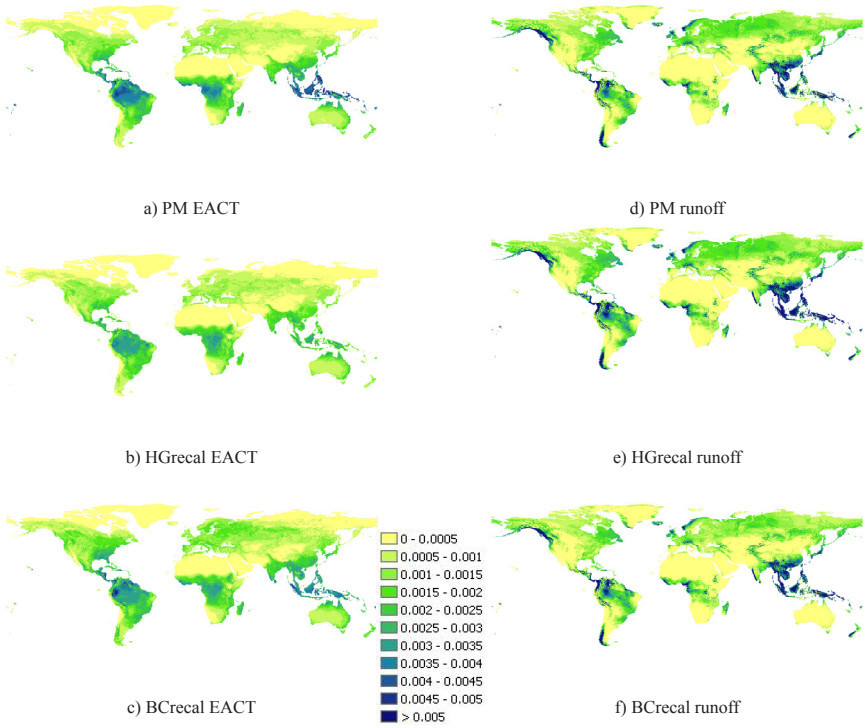


Figure B.7.1: Global maps with on the left annual average daily actual evapotranspiration (m day-1) and on the right annual average daily runoff (m/day). From top to bottom, Penman-Monteith (PM), Hargreaves with increased multiplication factor (HGrecal) and re-calibrated Blaney-Criddle (BCrecal).

the different PET is high and the absolute runoff amounts are of significant value, the selected PET method is likely to have a large impact on modeled runoff and discharge amounts.

Comparision of actual evapotranspiration

Absolute AET is relatively high for CFSRPM for the Amazon and Congo basins and the islands of southeast-Asia (Fig. B.7.1a). For these regions, even higher AET values are obtained with the PT and BCorig equations (Fig. B.7.2a and B.7.2c), resulting in low runoff values. AET calculated from both BCorig and BCrecal PET is high in Northern Europe and the Eastern US especially in the JJA season (Fig. B.7.1c and B.7.2c). This results in slightly lower runoff values for these regions (Fig. B.7.1f). The lowest AET values are derived from HGorig PET (Fig. B.7.2b) and even AET derived from HGrecal is relatively low. High similarity in AET values is found for HGrecal and BCrecal (fig. B.7.1b and B.7.1c). For illustrative purposes global maps of seasonal AET values are given in the supplementary information (Fig. S.2).

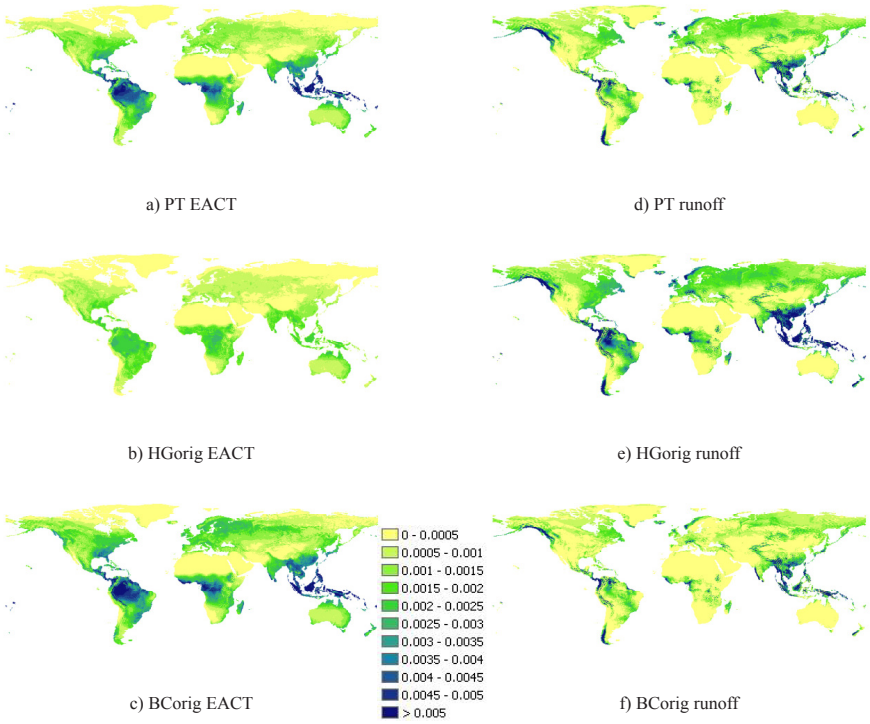


Figure B.7.2: Global maps with on the left annual average daily actual evapotranspiration (m/day) and on the right annual average daily runoff (m/day). From top to bottom, Priestley-Taylor (PT), the original Hargreaves equation (HGorig) and the original Blaney-Criddle equation (BCorig).

Comparison of runoff

Figures B.4, B.7.1 and B.7.2 show that differences in spatial runoff patterns are almost as small as the differences in AET patterns. This is a result of the fact that the runoff flux is influenced by both evapotranspiration and precipitation. Runoff is low for the PT and BCorig method (Fig. B.7.2d and B.7.2f). Although increasing the multiplication factor in the original HG equation to 0.0031 resulted in higher PET values, the difference in runoff derived from the two HG equations is still small (Fig. B.7.1e and B.7.2e). Global seasonal runoff maps for the different PET equation are provided as supplementary information (Fig. S.3).

B.3.3 Impact of different PET equations on discharge

The PET time series created in this study will be used in further research to downscale daily PET times-series derived from raw GCM data. The downscaled PET time-series can be employed for global hydrological impact assessments. We used the global hydrological model PCR-GLOBWB to evaluate the influence of the different PET equations on AET and river discharge. The model was forced with the daily CFSR PET time

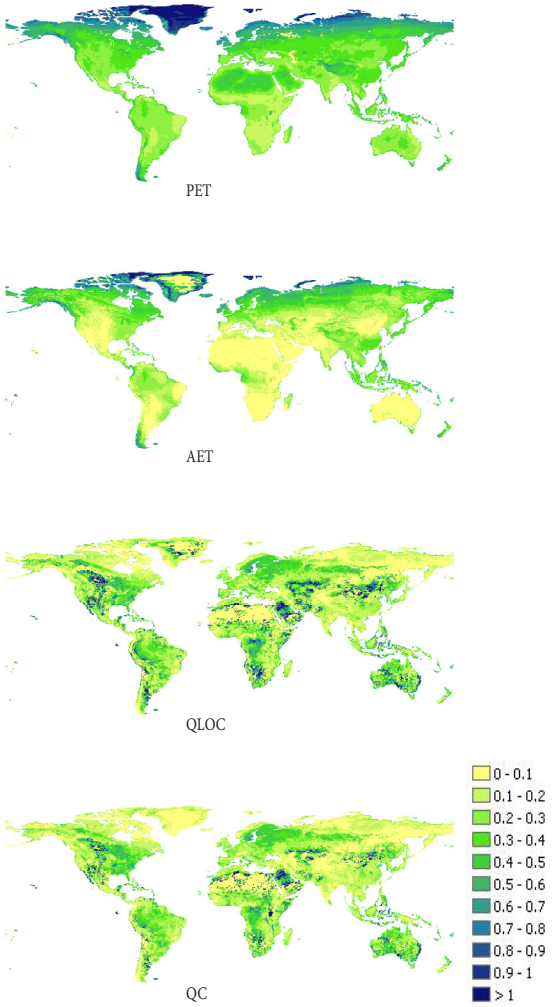


Figure B.8: Cell specific values of the coefficient of variation (CV; -) calculated from the six different potential evaporation methods for potential evaporation (PET), actual evapotranspiration (AET), local runoff (QLOC) and discharge (QC).

series, and both the CFSR and CRU PR and TAS, in subsequent steps. The latter was done to assess the influence of the bias in precipitation on modeled discharge as well.

Variation between methods

While being illustrative, the differences in runoff obtained from the six methods are hard to distinguish from the global runoff maps (Fig. B.7.1, B.7.2 and B.4). Therefore, basin specific discharge CV values, calculated from the discharges modeled with PCR-

Table B.2: Catchment specific coefficients of variation (CV) derived from long-term annual average modeled discharge for measurement stations closest to the catchment outlets, obtained with PET time series calculated with the six different potential evaporation equations.

Catchment	*CV*	*Catchment*	*CV*	*Catchment*	*CV*
Amazon	0.28	MacKenzie	0.24	Parana	0.25
Brahmaputra	0.10	Mekong	0.16	Rhine	0.17
Congo	0.34	Mississippi	0.31	Volga	0.28
Danube	0.19	Murray	0.26	Yangtze	0.19
Ganges	0.12	Niger	0.18	Yellow	0.17
Indus	0.05	Orange	0.29	Zambezi	0.30
Lena	0.10				

GLOBWB using the different CFSR PET time series as input, are listed in Table B.2 for 19 large rivers at measurement stations close to the catchment outlets. Basin discharge CV values are even lower than CV values for runoff, due to accumulation of processes along the river network. CV values of river discharge (Q) range between 0.05 and 0.34 and are on average 0.20. This indicates that the selection of a PET method is of minor relevance for modeled discharge (Oudin et al., 2005). The smallest variations in discharge between the different PET methods are found in the Monsoon influenced catchments where precipitation dominates discharge patterns. High CV values (0.26 – 0.30) are obtained for the Zambezi, Murray and Orange, basins in dry climate where PET has a large influence on resulting discharge. High values are also obtained for the

Amazon (0.28) and Congo (0.34). In these tropical basins the high variability between PET methods results in high variability in runoff and discharge as well, due to the humid climate. Contrary to the results of Oudin et al. (2005) this illustrates that for those basins with high CV values, which are unavoidable part of global scale studies, the selection of a PET equation does influence modeled discharge.

Deviations from observed discharge

Discharge calculated from the different PET equations and either CFSR or CRU PR and TAS are compared with observed discharge and discharge derived from CRU PET in Fig. B.9. The charts show that discharge derived from CFSR PET calculated with the BCorig equation is the lowest for all basins and underestimates corrected observed discharge for 6 out of 19 basins for both the runs forced with CFSR PR as well as for the runs forced with temporally downscaled CRU PR. Relatively low values are also obtained with the BCrecal and especially the PT equation. Discharges calculated from HGorig PET are highest for all basins.

In an overall comparison of the full CFSR runs with the corrected GRDC discharge and results of the full CRU run, BCorig performs the best. Discharge modeled from CFSR PET derived with this method has the lowest percentage bias from corrected GRDC discharge for 11 out of 19 basins and the lowest percentage bias from the full

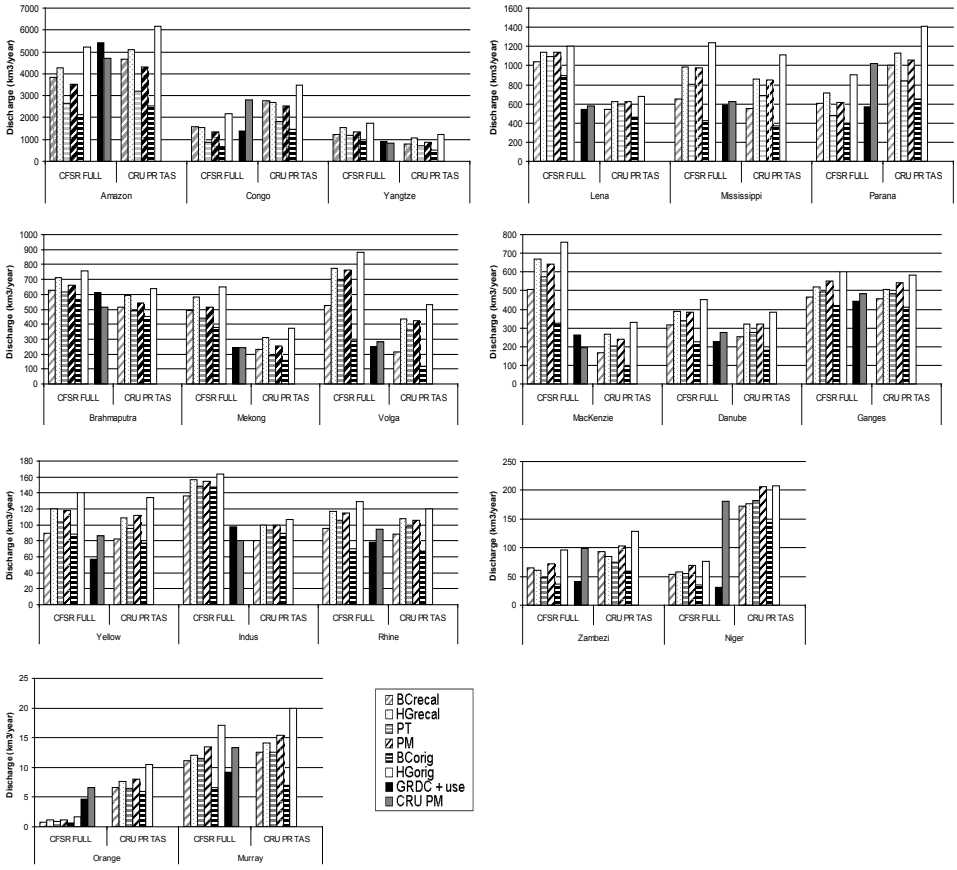


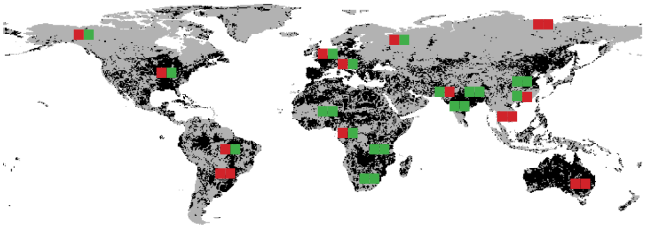
Figure B.9: Long-term average annual basin discharge (km^3/year) for 19 large river basins derived with PCRGLOB-WB forced with potential evaporation calculated from the CFSR dataset with the six different methods (group of bars on the right for each river). As references long-term average corrected observed GRDC basin discharge (black; periods do not completely overlap due to limited data availability), PCR-GLOBWB discharge modeled from the CRU dataset (dark grey) and discharges calculated from CRU precipitation, CRU temperature and CFSR PET (group of bars on the right) have been added.

CRU run for 5 out of 19 basins. The BCrecal method performs second best (lowest bias from corrected GRDC and full CRU run for 4 out of 19 basins). The HGorig method also shows good performance (best for 2 out of 19 in comparison with corrected GRDC and best for 5 out of 19 in comparison with full CRU run). However, as mentioned above, the HGorig discharge is the highest for all basins and the BCorig discharge is the lowest for all basins. These results suggest that these two methods perform the best for wrong reasons. The extreme values are likely to compensate for biases in observed discharge (McMillan et al., 2010; Vrugt et al., 2005) and biases in precipitation (Fekete et al., 2004; Biemans et al., 2009).

The influence of using either CFSR PR or CRU PR can be analyzed from Fig. B.9 by

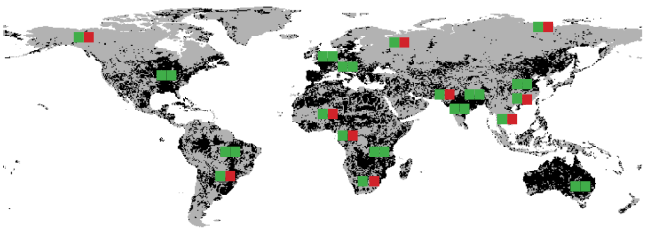
comparing the group of bars on the left and right side for each river. For example for the Amazon, Parana, Orange and Niger the use of CRU PR amounts results in much higher discharges, whereas for the Mekong, MacKenzie, Lena and Indus, CRU PR results in lower discharges. Differences between annual average discharge obtained from the two precipitation products are especially small for the Rhine, Murray and Yellow river.
Additional PCR-GLOBWB runs have been executed based on CFSR derived PET time series and the measurement based CRU PR and air TAS (bars on the right). This in order to minimize the influence of precipitation bias on modeled discharge and to distinguish the bias originating from PET from the bias originating from precipitation. The run based on CFSR HGrecal PET performs best for 11 out of 19 in comparison with the full CRU run, and best for 3 out of 19 in comparison with corrected GRDC discharge. In comparison with corrected GRDC discharge, the BCrecal (best for 7 basins) and BCorig (best for 8 basins) methods perform better. Yet, the BCorig method, which results in the lowest discharge values, mainly performs best for the dry basins (e.g. the Murray, Orange, Zambezi and Niger) where the hydrological model tends to underestimate AET and consequently overestimates discharge (Van Beek et al., 2011).
In section B.3.2 the impact of differences between ETP methods on resulting river discharge was quantified with CV values. Within the bar charts in Fig. B.9 the differences in runoff are displayed as well. Again it can be seen that for some basins the difference in discharge obtained from the different PET methods is large, see for example the Amazon and Congo, where due to the humid climate reduction from PET to AET is small. The difference is also large for the Mississippi which course travels through multiple climate zones (e.g. from sub-Arctic to semi-arid) and is therefore affected by differences in PET in at least part of the basin. Differences between methods are small for; the Niger and Orange due to limiting soil moisture conditions, the Lena because, as a result of the low temperatures, absolute AET amounts are low for this basins and the Indus where river discharge is highly influenced by precipitation.

Significance in deviations from CRU derived discharge
In Fig. B.10a and B.10b the significance of the differences between annual average (station) discharge (QC) derived from CRU and CFSR PET is indicated for PCR-GLOBWB runs forced with both full CFSR forcing and the forcing dataset existing of CFSR PET and CRU PR and TAS. Green squares correspond to similar annual average discharge derived from CFSR and CRU PET. Red squares indicate significant differences for a significance level of 95%.
Notable is the difference between QC derived from the full CFSR forcing and QC derived from CFSR PET and CRU PR and TAS amounts. When PCR-GLOBWB is forced with the observation based CRU PR and TAS and CFSR PET, significant differences from CRU PET derived values occur most often for the BCorig and HGorig equations. While for PCR-GLOBWB runs forced with the full CFSR dataset, the four

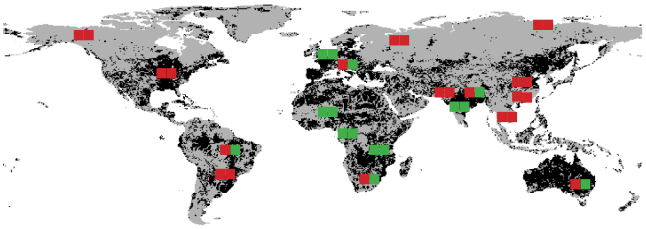


BC

significant basin difference
non-significant basin differnce
non-significant difference
significant difference

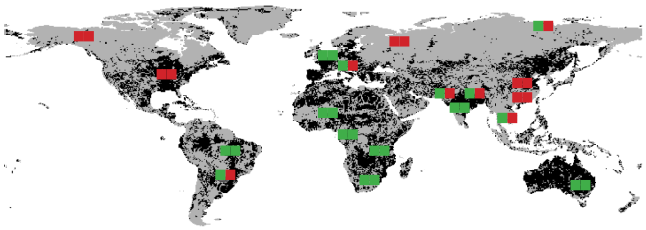


BCrecal



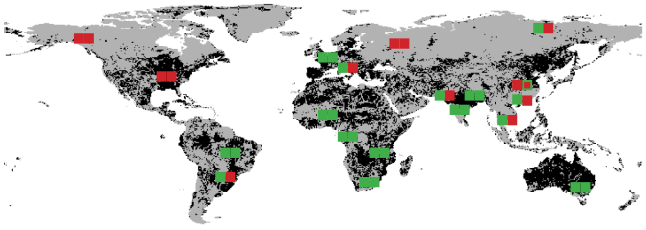
HGorig

Figure B.10a: Maps showing areas where CFSR derived (station) discharge (QC) significantly deviates from CRU derived values. Grey areas correspond to cells where CFSR derived QC significantly deviates from CRU derived QC, black cells indicate areas where annual average QC is similar. Squares indicate significance of difference in station discharge for the 19 major river basins in fig. B.9 (red is significant difference, green is similar annual average values). For each river, the left square corresponds to the PCR-GLOBWB run forced with CFSR PET and CRU PR and TAS and the right square corresponds to the runs with full CFSR forcing. Statistics are given for the original and re-calibrated Blaney-Criddle equation and the original Hargreaves equation.

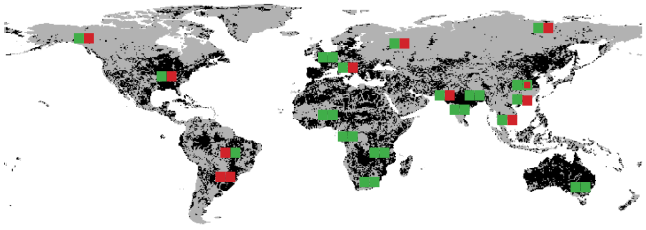


HGrecal

significant basin difference
non-significant basin differnce
non-significant difference
significant difference



PM



PT

Figure B.10b: Similar to B.10a, but now for the re-calibrated Hargreaves, Penman-Monteith and Priestley-Taylor equations.

other equations show lowest similarity with QC derived from CFSR PET. This indicates that the PET time series compensate for difference between CRU and CFSR PR. When evaluating the performance of the different PET equations for model runs forced with the observation based CRU PR and TAS, the BCrecal equation shows highest similarity with full CRU derived QC and the PT, PM and HGrecal equations also perform well.

B.4 Discussion and conclusions

In this study six different methods, to globally derive daily PET time series from CFSR reanalysis data, have been evaluated on 1) their resemblance with monthly PET time series calculated from the CRU datasets with the Penman-Monteith equation and 2) their impact on modeled runoff and river discharge and consequently usability for hydrological impact studies. The results of this study reveal that the selection of a PET method may be of minor influence on the resulting river flow modeled with a hydrological model. This can be seen from the transfer and reduction of the variability between PET methods throughout the hydrological modeling chain (see CV values Fig. B.8 and Table B.2) while moving from PET to AET to runoff (see also Kingston et al., 2009; Kay and Davies, 2008; Oudin et al., 2005). Only for those regions where 1) CV values calculated for AET and runoff obtained from the different PET equations are high and 2) absolute runoff amounts are of significant value (as for example the moderate Amazon, Congo and Mississippi regions) the selected PET method is likely to have a high impact on runoff and discharge amounts. As a consequence, selecting the most reliable PET method for future hydro-climatic impact studies remains important for specific regions or global and continental scale applications.
The PM equation is often recommended (Kay and Davies, 2008; Allen et al., 1998), whereas in this study it did not outperform the other methods, a similar conclusion was drawn by Oudin et al. (2005). In addition, there are other reasons why the PM equation is less suitable for application in climate impact studies. These are its high input data requirement and consequently long data-processing time. Furthermore, especially when working with reanalysis and bias-corrected GCM data, the sensitivity of the equation to input data accuracy will be of major relevance (Oudin et al., 2005). Therefore we do not promote the use of the PM equation for this global application.
The results of this study indicate that the re-calibrated Blaney-Criddle and Hargreaves equations applied to CFSR data in general outperform the other methods. Figure B.4, B.10a and B.10b show that the CFSR derived annual average AET, local runoff, (station) discharge and particularly PET show highest similarity with CRU derived values when using the re-calibrated Blaney-Criddle equation. And although these and other results suggest that PET is globally best calculated with the BCrecal equation, we here pose three critical remarks. Firstly, the original BC equation is developed for calculation of monthly instead of daily time series (Blaney and Criddle, 1950). Yet, in this study we violated this assumption and used the equation to calculate daily values, which are

required as input to the hydrological model PCR-GLOBWB. As a consequence the daily BCrecal PET spanned a relatively small range of daily PET values compared to the other methods (Fig. B.6). Secondly, discharge derived from BCrecal PET is too low compared to the other methods for most basins (second lowest after BCorig derived runoff or lowest for 12 out of 19 basins, Fig. B.9). Finally, the values of the Blaney-Criddle coefficients show a high spatial variability (Fig. B.1) due to the cell specific re-calibration which was required because of the large spatial variation in bias of CFSR BC PET from CRU PET. The sensitivity of the coefficients to the spatially varying climate conditions (Jensen, 1966) suggests that they will also be sensitive to future changing climate conditions.

By contrast, the Hargreaves equation, which could be globally uniform re-calibrated due to its small spatial variation in bias, performs well in multiple climate zones and is therefore also likely to perform well under changing climate conditions. It is also an efficient empirical equation, yet more spatial varying meteorological variables are considered than in the Blaney-Criddle equation, which increases its possibilities for spatial transfer. The large adjustment of the multiplication factor from 0.0023 to 0.0031 significantly improves the goodness-of-fit of the Hargreaves equation, similar increases were also suggested by Droogers and Allen (2002). The adjustment may directly result from the application of the equation to a daily time-step while the equation is said to perform best over longer time-steps (10 days – month; Hargreaves et al., 2003). From the discussion above we conclude that, for the calculation of daily PET time series from CFSR reanalysis data, the re-calibrated HG equation is the most reliable equation.

Acknowledgements

The global discharge time series have been obtained from the Global Runoff Data Centre. The CFSR reanalysis products used in this study are obtained from the Research Data Archive (RDA) which is maintained by the Computational and Information Systems Laboratory (CISL) at the National Center for Atmospheric Research (NCAR). NCAR is sponsored by the National Science Foundation (NSF). The original data are available from the RDA (http://dss.ucar.edu) in dataset number ds093.0. H.H. Dürr, was partly funded by the EU FP6 program "Carbo-North"(contract nr. 036993).

Supplemantary information

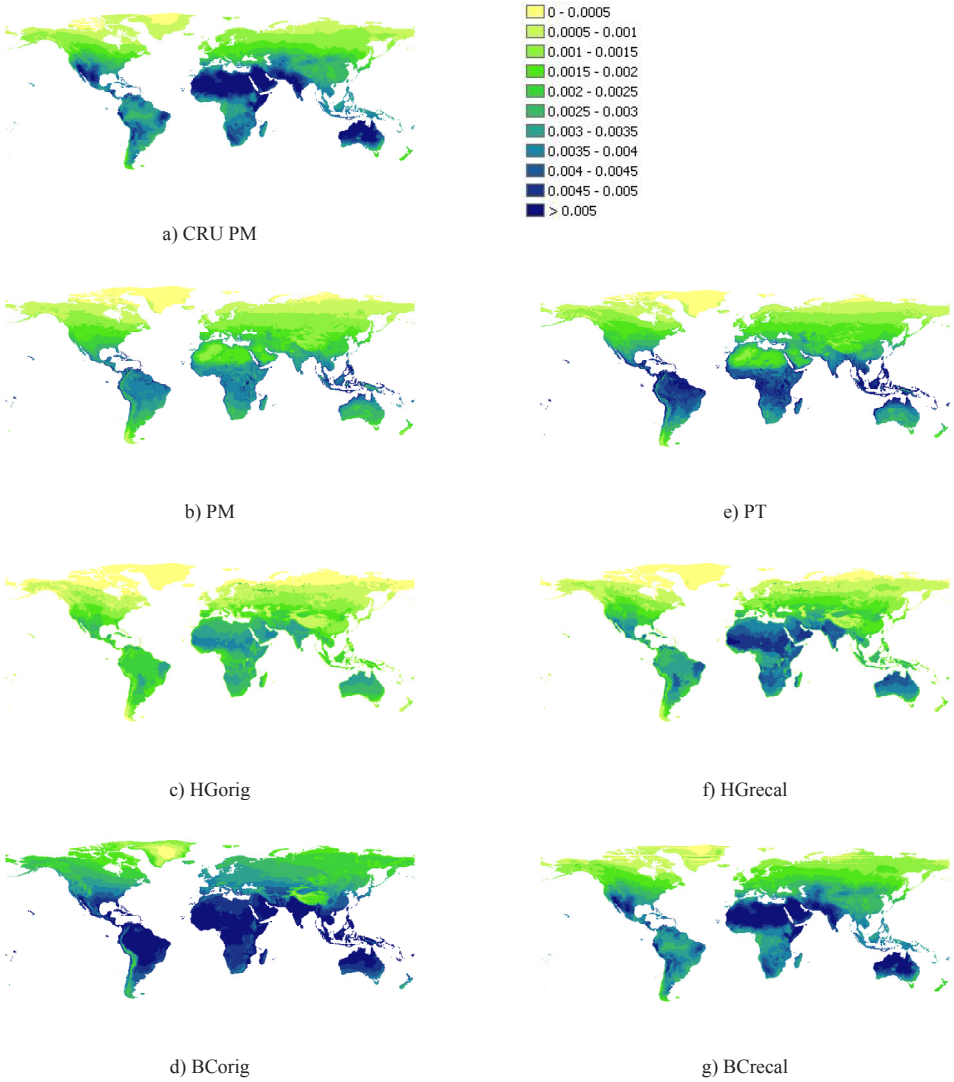


Figure S.1: Global maps with annual average reference potential evaporation (PET; m/day). On top annual average CRU Penman-Monteith reference PET. In the right column reference PET obtained with the Penman-Monteith (PM), the standard Hargreaves (HGorig) and Blaney-Criddle (BCorig) and tn the right column reference PET obtained with Priestley-Taylor (PT), Hargreaves with increased multiplication factor (HGrecal) and the re-calibrated Blaney-Criddle equation (BCrecal) are displayed.

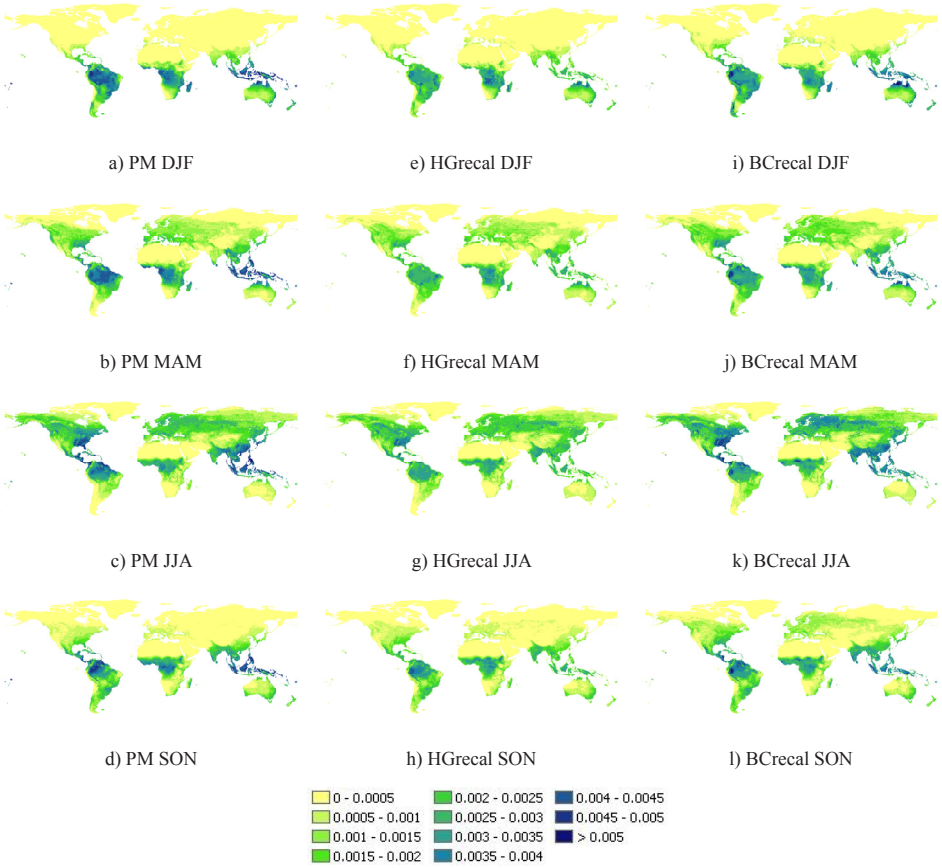


Figure S.2.1: Global maps with seasonal average daily actual evapotranspiration (m/day). From left to right Penman-Monteith (PM), Hargreaves with increased multiplication factor (HGrecal) and re-calibrated Blaney-Criddle (BCrecal) and from top to bottom the DJF, MAM, JJA and SON seasons.

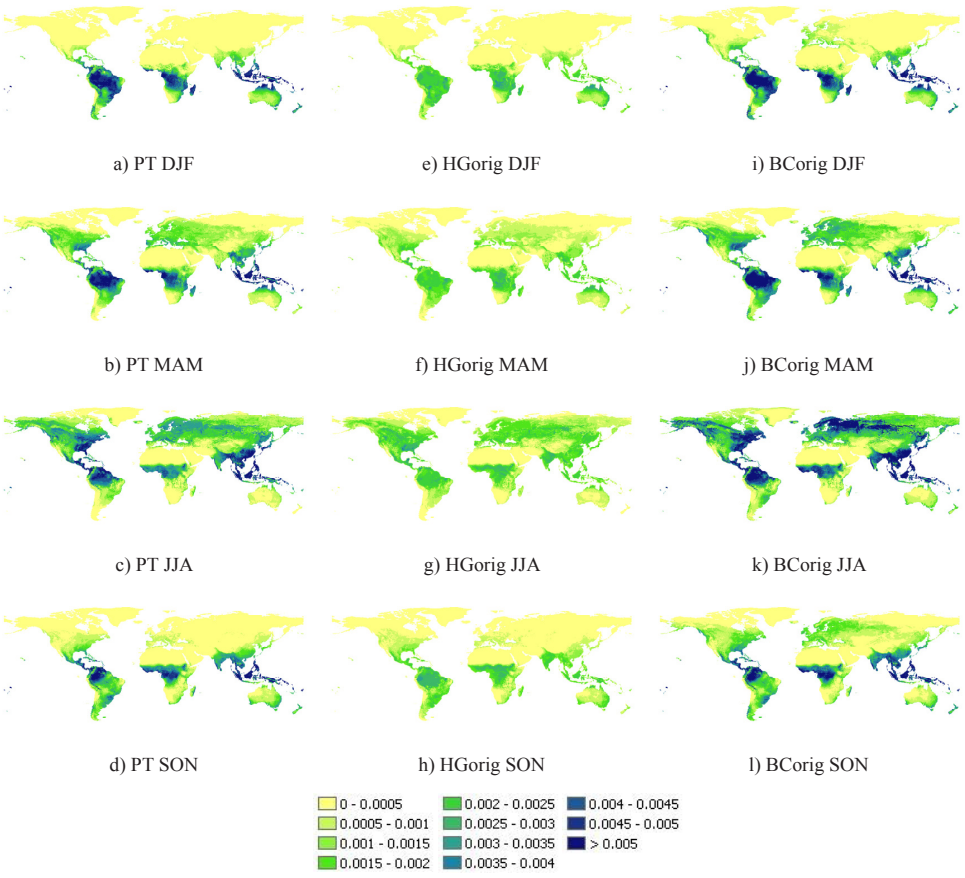


Figure S.2.2: Global maps with seasonal average daily actual evapotranspiration (m/day). From left to right Priestley-Taylor (PT), the original Hargreaves equation (HGorig) and the original Blaney-Criddle equation (BCorig) and from top to bottom the DJF, MAM, JJA and SON seasons.

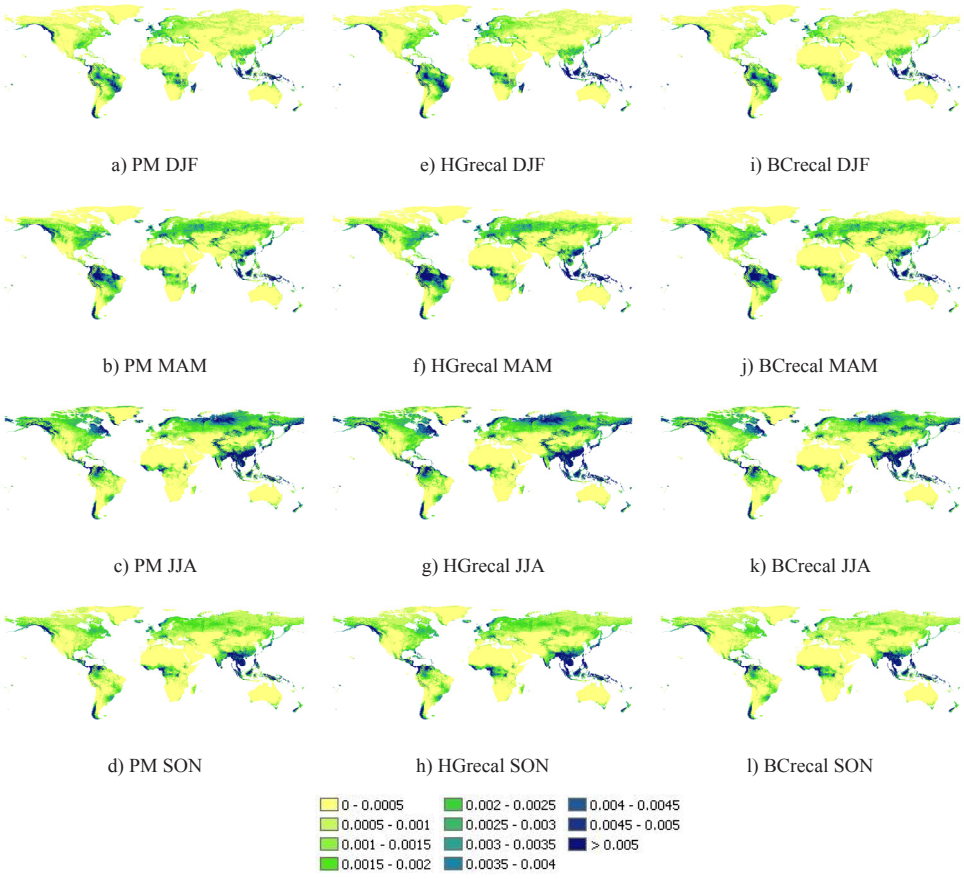


Figure S.3.1: Global maps with seasonal average daily cell specific runoff (m/day). From left to right Penman-Monteith (PM), Hargreaves with increased multiplication factor (HGrecal) and re-calibrated Blaney-Criddle (BCrecal) and from top to bottom the DJF, MAM, JJA and SON seasons.

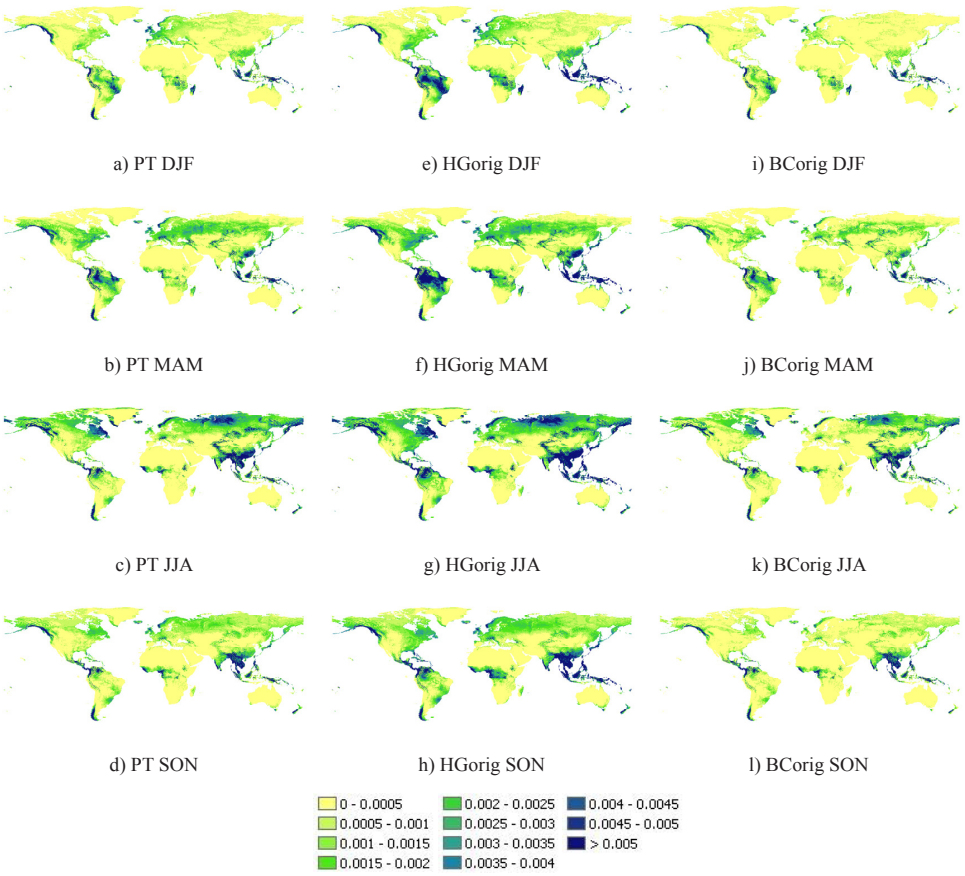


Figure S.3.2: Global maps with seasonal average daily cell specific runoff (m/day). From left to right Priestley-Taylor (PT), the original Hargreaves equation (HGorig) and the original Blaney-Criddle equation (BCorig) and from top to bottom the DJF, MAM, JJA and SON seasons.

References

Aerts, J., H. Renssen, P. J. Ward, H. de Moel, E. Odada, L. M. Bouwer & H. Goosse (2006), Sensitivity of global river discharges under Holocene and future climate conditions. Geophysical Research Letters 33(19), L19401, doi:10.1029/ 2006GL027493.

Ajami, N. K., Q. Duan, X. Gao & S. Sorooshian (2006), Mulitmodel combination techniques for analysis of hydrological simulations: Application to distributed model intercomparison project results. Journal of Hydrometeorology 7(4), pp. 755-768.

Alcamo, J. & T. Henrichs (2002), Critical regions: A model-based estimation of world water resources sensitive to global changes. Aquatic Sciences 64, pp. 352–263.

Alcamo, J., P. Döll, T. Henrichs, F. Kaspar, B. Lehner, T. Rösch & S. Siebert (2003), Development and testing of the WaterGAP 2 global model of water use and availability. Hydrological Sciences Journal 48(3), pp. 317–337.

Alcamo, J., M. Flörke & M. Märker (2007), Future long-term changes in global water resources driven by socio-economic and climatic changes. Hydrological Sciences Journal 52(2), pp. 247-275.

Alcamo, J., T. Henrichs & T. Rösch (2000), World water in 2025: Global modeling and scenario analysis for the world commission on water for the 21st century. Kassel World Water Series Report No. 2, Center for Environmental Systems Research, University of Kassel, Germany.

Allan, R. P. & B. J. Soden (2008), Atmospheric warming and the amplification of precipitation extremes. Science 321, pp. 1481–1484, doi:10.1126/science.1160787.

Allen, M. R. & W. J. Ingram (2002), Constraints on future changes in climate and the hydrological cycle. Nature 419, pp. 224-232, doi:10.1038/nature01092.

Allen, R. G., L. S. Pereira, D. Raes, D. & M. Smith (1998), Crop evapotranspiration: FAO Irrigation and drainage paper 56. FAO, Rome, Italy.

Allen R. G. (1993), Evaluation of a temperature difference method for computing grass reference evapotranspiration, Report submitted to the Water Resources Development and Man Service, Land and Water Development

Division. FAO, Rome. 49 p.
Anderton, S., J. Latron & F. Gallart (2002), Sensitivity analysis and multi-response, multi-criteria evaluation of a physically based distributed model. Hydrological Processes 16, pp. 333-353.
Andréassian, V., C. Perrin, C. Michel, I. Usart-Sanchez & J. Lavabre (2001), Impact of imperfect rainfall knowledge on the efficiency and the parameters of watershed models. Journal of Hydrology 250, pp. 206-223, doi:10.1016/S0022-1694(01)00437-1.
Arnell, N. W. (1999a), A simple water balance model for the simulation of streamflow over a large geographic domain. Journal of Hydrology 217, pp. 314–355.
Arnell, N. W. (1999b), Climate change and global water resources. Global Environmental Change 9, pp. 831-849.
Arnell, N. W. (1999c), The effect of climate change on hydrological regimes in Europe: a continental perspective. Global Environmental Change 9, pp. 5-23.
Arnell, N. W. (2003), Effects of IPCC SRES emissions scenarios on river runoff: a global perspective. Hydrology and Earth System Sciences 7, pp. 619–641, doi:10.5194/hess-7-619-2003.
Arnell, N. W. (2004), Climate change and global water resources: SRES emissions and socio-economic scenarios. Global Environmental Change 14, pp. 31-52, doi:10.1016/j.gloenvcha.2003.10.006.
Arnell, N. W. (2011), Uncertainty in the relationship between climate forcing and hydrological response in UK catchments. Hydrology and Earth System Sciences 15, pp. 897-912, doi:10.5194/hess-15-897-2011.
Arora, V. K. (2001), Streamflow simulations for continental-scale river basins in a global atmospheric general circulation model. Journal of Hydrology 24(7), pp. 775-591, doi:10.1016/S0309-1708(00)00078-6.
Arora, V. K., & G. J. Boer (2001), Effects of simulated climate change on the hydrology of major river basins. Journal of Geophysical Research, 106(4), pp. 3335-3348, doi:10.1029/2000JD900620.
Ashfaq, M., Y. Shi, W. Tung, R. J. Trapp, X. Gao, J. S. Pal & N. S. Diffenbaugh (2009), Suppression of south Asian summer monsoon precipitation in the 21st century. Geophysical Research Letters 36, L01704, doi:10.1029/2008GL036500.
Balin, D., H. Lee & M. Rode (2010), Are rainfall measurement errors likely to greatly impact on distributed complex hydrological modelling? Water Resources Research, doi:10.1029/2009WR007848.

Baumgartner, A. & E. Reichel (1975), The world water balance: Mean annual global, continental and maritime precipitation, evaporation and runoff. Elsevier, Amsterdam, 179 p.

Beck, C., J. Grieser, & B. Rudolf (2004), A new monthly precipitation climatology for the global land areas for the period 1951 to 2000. Climate Status Report 2004, German Weather Service, Offenbach, Germany, pp. 181–190.

Beven K. (1996), A discussion of distributed hydrological modelling. In Distributed Hydrological Modelling, Abbott MB, Refsgaard JC (eds), Kluwer Academic, pp. 255–278.

Beven, K. (2001), How far can we go in distributed hydrological modelling? Hydrology and Earth System Sciences 5, pp. 1–12, doi:10.5194/hess-5-1-2001.

Beven, K. (2006), A manifesto for the equifinality thesis. Journal of Hydrology 320, pp. 18–36, doi:10.1016/j.jhydrol.2005.07.007.

Beven, K. (2011), I believe in climate change but how precautionary do we need to be in planning for the future. Hydrological processes 25, pp. 1517-1520, doi: 10.1002/hyp.7939.

Beven, K. J. & A. M. Binley (1992), The future of distributed hydrological models: model calibration and uncertainty prediction. Hydrological Processes 6, pp. 279-298.

Beven, K. J. & M. J. Kirkby (1979), A physically based, variable contributing area model of basin hydrology. Hydrological Sciences Bulletin 24, pp. 43–69.

Biemans, H., R. W. A. Hutjes, P. Kabat, B. Strengers, D. Gerten & S. Rost (2009), Effects of precipitation uncertainty on discharge calculations for main river basins. Journal of Hydrometeorology 10(4), pp. 1011-1025, doi: 10.1175/2008JHM1067.1.

Bierkens, M. F. P. & L. P. H. van Beek (2009), Seasonal predictability of european discharge: NAO and hydrological response time. Journal of Hydrometeorology 10, pp. 953–968, doi:10.1175/2009JHM10341.

Blaney, H. F. & W. P. Criddle (1950), Determining water requirements in irrigated areas from climatological and irrigation data. USDA (SCS) TP-96, 48 p.

Boorman, H. (2010), Sensitivity analysis of 18 different potential evapotranspiration models to observed climatic change at German climate stations. Climatic Change 104 (3-4), pp. 729-753, doi: 10.1007/s10584-010-9869-7.

Boorman, D. B. & C.E.M. Sefton (1997), Recognizing the uncertainty in the quantification of the effects of climate change on hydrological response. Climatic Change 35(4), pp. 415–434, doi: 10.1023/A:1005372407881.

Bosilovich, M.G., J. Chen, F. R. Robertson & R. F. Adler (2007), Evaluation of

global precipitation in reanalyses, 21st Conference on Hydrology San Antonio, TX January 2007.

Britannica (2011), www.britannica.com. Retrieved June 15, 2011.

Brouwer, C. & M. Heibloem (1986), Irrigation water management: Irrigation water needs. FAO, Rome, Italy.

Buytaert, W., R. Célleri & L. Timbe (2009), Predicting climate change impacts on water resources in the tropical Andes: Effects of GCM uncertainty. Geophysical Research Letters 36, L07406, doi:10.1029/2008GL037048.

Canadell J. G., C. Le Quéré, M. R. Raupach, C. B. Field, E. T. Buitenhuis, P. Ciais, T. J. Conway, R. A. Houghton & G. Marland (2007), Contributions to accelerating atmospheric CO2 growth from economic activity, carbon intensity, and efficiency of natural sinks. Proceedings of the National Academy of Sciences, USA, 0702737104.

Candogan Yossef, N., L. P. H. Van Beek, J. C. J. Kwadijk & M. F. P. Bierkens (2011), Skill assessment of a global hydrological model in reproducing flow extremes. Hydrology and Earth System Sciences Discussions 8, 3469-3505, doi:10.5194/hessd-8-3469-2011.

CBC news (2011), The toll of Australia's floods. http://www.cbc.ca/news/world/story/2011/01/11/f-australia-queensland-floods-by-the-numbers.html, retrieved June 24, 2011.

Chen, J. & B.G. Bosilovich (2007), Hydrological variability and trends in global reanalyses. 19th Conference on Climate Variability and Change, San Antonio, Texas, January 2007, JP4.4.

Chiew, F. H. S., J. Teng, J. Vaze & D. G. C. Kirono (2009), Influence of global climate model selection on runoff impact assessment. Journal of Hydrology 379, pp. 172-180, doi:10.1016/j.jhydrol.2009.10.004.

Chow, V. T., D. R. Maidment & L. W. Mays (1988), Applied Hydrology. McGraw-Hill, New York, 572 p.

Christensen, N. S. & D. P. Lettenmaier (2007), A multimodel ensemble approach to assessment of climate change impacts on the hydrology and water resources of the Colorado River Basin. Hydrology and Earth System Sciences 11, pp. 1417–1434, doi:10.5194/hess-11-1417-2007.

Clap, R. B. & G. B. Hornberger (1978), Empirical equations for some soil hydraulic properties. Water Resources Research 14(4), pp. 601-604, doi: 10.1029/WR014i004p00601.

Clark, D. B. & N. Gedney (2008), Representing the effects of subgrid variability of soil moisture on runoff generation in a land surface model. Journal of Geophysical Research 113, D10111, doi:10.1029/2007JD008940.

ClimateChangeDispatch (2011), http://climatechangedispatch.com. retrieved June 24, 2011.

Cook, K. H. & E. K. Vizy (2006), Coupled model simulations of the West African monsoon system: Twentieth- and twenty-first- century simulations. Journal of Climate 19, pp. 3681–3703.

Covey, C., K. M. AchutaRao, U. Cubasch, P. Jones, S. J. Lambert, M. E. Mann, T. J. Phillips & K. E. Taylor (2003), An overview of results from the coupled model intercomparison project. Global Planetary Change 37, pp. 103–133, doi:10.1016/S0921-8181(02)00193-5.

Cox, P. M., R. A. Betts, C. B. Bunton, R. L. H. Essery, P. R. Rowntree & J. Smith (1999), The impact of new land surface physics on the GCM simulation of climate and climate sensitivity. Climate Dynamics 15, pp. 183-203.

Dai, A. (2006), Precipiation characteristics in eighteen coupled climate Models. Quarterly Journal of the American Meteorological Society 19, pp. 4605–4630.

Dee, D. P. & S. Uppala (2009), Variational bias correction of satellite radiance data in the ERA-Interim reanalysis. Quarterly Journal of the Royal Meteorological Society 135, pp. 1830-1841.

Diks, C. G. H. & J. A. Vrugt (2010), Comparison of point forecast accuracy of model averaging methods in hydrologic applications. Stochastic Environmental Research and Risk Assessment 24(6), pp. 809-821, doi:10.1007/s00477-010-0378-z.

Döll, P. & B. Lehner (2002), Validating of a new global 30-minute drainage direction map. Journal of Hydrology 258, pp. 214–231.

Döll, P. & J. Zhang (2010), Impact of climate change on freshwater ecosystems: a global-scale analysis of ecologically relevant river flow alterations. Hydrology and Earth System Sciences 14, pp. 783-799, doi:10.5194/hess-14-783-2010.

Döll, P., F. Kaspar & B. Lehner (2003), A global hydrological model for deriving water availability indicators: model tuning and validation. Journal of Hydrology 270, pp. 105–134, doi:10.1016/S0022-1694(02)00283-4.

Döll, P., K. Fiedler & J. Zhang (2009), Global-scale analysis of river flow alterations due to water withdrawals and reservoirs. Hydrology and Earth System Sciences 13, pp. 2413-2432, doi:10.5194/hess-13-2413-2009.

Droogers, P. & R.G. Allen (2002), Estimating reference evapotranspiration under inaccurate data conditions. Irrigation and drainage systems 16, pp. 33-45, doi: 10.1023/A:1015508322413.

Duan, Q., S. Sorooshian & V. Gupta (1992), Effective and efficient global opti-

mization for conceptual rainfall-runoff models. Water Resources Research 28(4), pp. 1015-1031.

Dumenil, L. & E. Todini (1992), A rainfall-runoff scheme for use in the Hamburg climate model. Advances in Theoretical Hydrology, A Tribute to James Dooge, J.P. O'Kane, Ed., Elsevier, Amsterdam, pp. 129–157.

Dürr, H. H., M. Meybeck & S. H. Dürr (2005), Lithologic composition of the Earth's continental surfaces derived from a new digital map emphasizing riverine material transfer. Global Biogeochemical Cycles 19, GB4S10, doi:10.1029/2005GB002515.

Ekström, M., P. D. Jones, H. J. Fowler, G. Lenderink, T. A. Buishand & D. Conway (2007), Regional climate model data used within the SWURVE project 1: projected changes in seasonal patterns and estimation of PET. Hydrology and Earth System Sciences 11 (3), pp. 1069–1083, doi:10.5194/hess-11-1069-2007.

Elshamy, M. E., I. A. Seierstad & A. Sorteberg (2009), Impacts of climate change on Blue Nile flows using bias-corrected GCM scenarios. Hydrology and Earth System Sciences 13, pp. 551-565, doi:10.5194/hess-13-551-2009.

Fekete, B. M., C. J. Vörösmarty & W. Grabs (2000), Global, composite runoff fields based on observed river discharge and simulated water balances. Technical Report 22, Global Runoff Data Centre, Koblenz, Germany.

Fekete, B. M., C. J. Vörösmarty & W. Grabs (2002), High resolution fields of global runoff combining observed river discharge and simulated water balances. Global Biogeochemical Cycles 16(3), pp. 1-6, doi:10.1029/1999GB001254.

Fekete, B. M., C. J. Vörösmarty, J. O. Roads & C. J. Willmott (2004), Uncertainties in precipitation and their impacts on runoff estimates. Journal of Climate 17, pp. 294-304, doi: 10.1175/1520-0442(2004)017.

Fiedler, K. & P. Döll (2007), Global modelling continental water storage changes – sensitivity to different climate data sets. Advances in Geosciences 11, pp. 63–68.

Fowler, H. J. & C. G. Kilsby (2007), Using regional climate model data to simulate historical and future ruver flows in northwest England. Climatic Change 80, pp. 337–367, doi: 10.1007/s10584-006-9117-3.

Fowler H. J., S. Blenkinsop & C. Tebaldi (2007), Review: Linking climate change modelling to impact studies: recent advances in downscaling techniques for hydrological modeling. International Journal of Climate 27, pp. 1547-1578, doi: 10.1002/joc.1556.

Gain, A. K., W. W. Immerzeel, F. C. Sperna Weiland & M. F. P. Bierkens (2011), Impact of climate change on the stream flow of the lower Brahmapu-

tra: trends in high and low flows based on discharge-weighted ensemble modeling. Hydrology and Earth System Sciences 15, pp. 1537-1545, doi:10.5194/hess-15-1537-2011.

Gavilán, P., I. J. Lorite, S. Tornero & J. Berengena (2006), Regional calibration of Hargreaves equation for estimating reference ET in a semiarid environment. Agricultural Water Management 81, pp. 257-281, doi:10.1016/j.ag.wat.2005.05.001.

Gerten, D., S. Schaphoff, U. Haberlandt, W. Lucht & S. Sitch (2004), Terrestrial vegetation and water balance – hydrological evaluation of a dynamic global vegetation model. Journal of Hydrology 286, pp. 249–270, doi:10.1016/j.jhydrol.2003.09.029.

Giorgi, F. & L. O. Mearns (2002), Calculation of average, uncertainty range, and reliability of regional climate changes from AOGCM simulations via the "Reliability Ensemble Averageing" (REA) method. Journal of Climate 15, pp. 1141-1158, doi: 10.1175/1520-0442(2002)015.

Giorgi, F. & L. O. Mearns (2003), Probability of regional climate change based on the Reliability Ensemble Averaging (REA) method. Geophysical Research Letters 30(12), 1629, doi: 10.1029/2003GL017130.

Gleeson, T., L. Smith, N. Moosdorf, J. Hartmann, H. H. Dürr, A. H. Manning, L. P. H. Van Beek & A. M. Jellinek (2011), Mapping permeability over the surface of the Earth. Geophysical Research Letters 38, L02401, doi:10.1029/2010GL045565.

Global Carbon Project Carbon budget and trends (2008), http://www.globalcarbonproject.org, retrieved 26 September 2008

Gosling, S. N. & N. W. Arnell (2011), Simulating current global river runoff with a global hydrological model: model revisions, validation and sensitivity analysis. Hydrological Processes 25, pp. 1129-1145, doi: 10.1002/hyp.7727.

Gosling, S.N, D. Bretherton, K. Haines & N. W. Arnell (2010), Global hydrology modelling and uncertainty: running multiple ensembles with a campus grid. Philosophical transactions of the royal society A 368, pp. 4005-4021, doi: 10.1098/rsta.2010.0164.

Gosling S. N., R. G. Taylor, N. W. Arnell & M. C. Todd (2011), A comparative analysis of projected impacts of climate change on river runoff from global and catchment-scale hydrological models. Hydrology and Earth System Sciences 15, pp. 279-294, doi:10.5194/hess-15-279-2011.

Graham, L. P. & D. Jacob (2000), Using large-scale hydrologic modelling to review runoff generation processes in GCM climate models. Meteorolo-

gische Zeitschrift, 9(1), pp. 49-57.
GRDC (2004), Long term mean annual freshwater surface water fluxes into the world oceans; comparisons of GRDC freshwater flux estimate with literature. http://grdc.bafg.de/servlet/is/7083.
GRDC (2007), Major river basins of the world. Global Runoff Data Centre, D – 56002, Federal Institute of Hydrology (BfG), Koblenz, Germany.
Gregory, K. J. (1976), Drainage networks and climate. Chapter 10 in: Geomorphology and climate (e.d. by E. Derbyshire). John Wiley, London.
Guardian (2011), Severe drought causes hunger for 10 million in west Africa. http://www.guardian.co.uk/environment/2010/jun/03/drought-hunger-west-africa, retrieved July 14, 2011.
Gupta, H.V., S. Sorooshian & P. O. Yapo (1998), Toward improved calibration of hydrological models: Multiple and noncommensurable measures of information. Water Resources Research 34, pp. 751-763.
Haan, C.T., D. E. Storm, T. Al-Issa, S. Prabhu, G. J. Sabbagh & D. R. Edwards (1998), Effect of parameter distribution on uncertainty analysis of hydrologic models. Transactions of the ASAE 41(1), pp. 65–70.
Haerter, J. O., S. Hagemann, C. Moseley & C. Piani (2010), Climate model bias correction and the role of timescales. Hydrology and Earth System Sciences 15, pp. 1065-1079, doi: 10.5194/hessd-7-7863-2010.
Hagemann, S. (2002), An improved land surface parameter data set for global and regional climate models, Max-Planck-Institute for Meteorology, Report 336, Hamburg.
Hagemann, S. & L. D. Gates (2003), Improving a subgrid runoff parameterization scheme for climate models by the use of high resolution data derived from satellite observations. Climate Dynamics 21, pp. 349-359, doi: 10.1107/s00382-003-0349-x.
Hagemann, S., K. Arpe & L. Bengtsson (2005), Validation of the hydrological cycle of ERA-40. Reports on earth system science, Max Planck Institute, Hamburg, Germany.
Hagemann, S., K. Arpe & E. Roeckner (2006), Evaluation of the hydrological cycle in the ECHAM5 model. Journal of Climate 19, pp. 3810-3827, doi: 10.1175/JCLI3831.1.
Hagemann, S., H. Göttel, D. Jacob, P. Lorenz & E. Roeckner (2009), Improved regional scale processes reflected in projected hydrological changes over large European catchments. Climate Dynamics 32, pp. 767–781, doi:10.1007/s00382-008-0403-9.
Hargreaves, G. H. & Z. A. Samani (1985), Reference crop evapotranspiration

from temperature. Applied engineering in agriculture 1(2), pp. 96-99.
Hargreaves, G. H., F. Asce & R. G. Allen (2003), History and evaluation of Hargreaves evapotranspiration equation. Journal of Irrigation and Drainage Engineering 129(1), pp. 53-63, .doi: 10.1061/(ASCE)0733-9437(2003)129:1(53).
Hargreaves, J. C. (2010), Skill and uncertainty in climate models. Wiley Interdisciplinary Reviews: Climate Change, doi:10.1002/wcc.58.
Henrichs, T., B. Lehner & J. Alcamo (2002), An Integrated Analysis of Changes in Water Stress in Europe. Integrated Assessment 3 (2), pp. 15-29.
Higgins, R. W., V. E. Kousky, V. B. S. Silva, E. Becker & P. Xie (2010), Intercomparison of daily precipitation statistics over the United States in observations and in NCEP reanalysis products. Journal of Climate 23, pp. 4637–4650, doi: 10.1175/2010JCLI3638.1.
Hoekstra, A. Y. & P. Q. Hung (2005), Globalisation of water resources: international virtual water flows in relation to crop trade. Global environmental change 15, pp. 45-56, doi:10.1016/j.gloenvcha.2004.06.004.
Huntington, T. G. (2006), Evidence for intensification of the global water cycle: Review and synthesis. Journal of Hydrology 319, pp. 83-95, doi: 10.1016/j.jhydrol.2005.07.003.
Hurkmans, R., W. Terink, R. Uijlenhoet, P. Torfs, D. Jacob & P. A. Troch (2010), Changes in streamflow dynamics in the Rhine basin under three high-resolution regional climate scenarios, Journal of Climate 23 (3), pp. 679–699, doi: 10.1175/2009JCLI3066.1.
Immerzeel, W. W., L. P. H. van Beek & M. F. P. Bierkens (2010), Climate change will affect the Asian water towers. Science 328 (5984), pp. 1382–1385, doi:10.1126/science.1183188.
Ines, A. V. M. & J. W. Hansen (2006), Bias correction of daily GCM rainfall for crop simulation studies. Agricultural and Forest Meteorology 138, pp. 44–53, doi:10.1016/j.agrformet.2006.03.009.
IPCC (2000), IPCC special reports Emission scenarios – summary for policy makers.
IPCC (2007a), Climate change 2007: Synthesis report – summary for policy makers.
IPCC (2007b), Climate Change 2007: The physical science basis, contribution of working group I to the fourth assessment report of the intergovernmental panel on climate change.
Jensen, M. E. (1966), Discussion of 'irrigation water requirements of lawns.'. Journal of Irrigation and Drainage Division 92, pp. 95-100.

Johns, T. C. (2009), Ensembles stream2 MetO-HC-HadGEM2AO SRA1B run1, daily values. World Data Center for Climate. CERA-DB "Ensembles2_HadGEM2_SRA1B_1_D".http://cera-www.dkrz.de/WDCC/ui/ compact.jsp?acronym= ENSEMBLES2 HADGEM2_SRA1B_1_D

Johnson, F. & A. Sharma (2009), Measurement of GCM skill in predicting variables relevant for hydroclimatological assessments. Journal of Climate 22 (16), pp. 4373–4382, doi:10.1175/2009JCLI2681.1.

Kallache, M., M. Vrac, P. Naveau & P.-A. Michelangeli (2011), Nonstationary probabilistic downscaling of extreme precipitation. Journal of Geophysical Research 116, D05113, doi:10.1029/2010JD014892.

Kalnay, E. & Coauthors (1996), The NCEP/NCAR 40-year reanalysis project, Bulletin of the American Meteorological Society 77, pp. 437-470.

Kay, A. L. & H. N. Davies (2008), Calculating potential evaporation from climate model data: A source of uncertainty for hydrological climate change impacts. Journal of Hydrology 358, pp. 221–239, doi: 10.1016/j.jhydrol.2008.06.005.

Kay, A. L., H. N. Davies, V. A. Bell, & R.G. Jones (2009), Comparison of uncertainty sources for climate change impacts: flood frequency in England. Climatic Change 92, pp. 41–63, doi: 10.1007/s10584-008-9471-4.

Kingston, D. G., M,. C. Todd, R. G. Taylor & J. R. Thompson (2009), Uncertainty in the estimation of potential evapotranspiration under climate change. Geophysical Research Letters 36, L20403, doi:10.1029/2009GL040267.

Kite, G. (1998), Land surface parameterizations of GCMs and macroscale models. Journal of the American Water Resources Association 34(6), pp. 1247-1254.

Korzun, V. I., A. A. Sokolov, M. I. Budyko, K. P. Voskresensky, G. P. Kalinin, A. A. Konoplyantsev, E. S. Korotkevich & M. I. L'vovitch (1978), World Water Balance and Water Resources of the Earth. UNESCO, 663 p.

Kraijenhoff van de Leur, D. (1958), A study of non-steady groundwater flow with special reference to a reservoir coefficient, De Ingenieur 70, pp. 87–94.

Krishnamurti, T. N., C. M. Kishtawal, T. LaRow, D. Bachiochi, Z. Zhang, C. E. Williford, S. Gadgil & S. Surendran (1999), Improved forecast skill of weather and seasonal climate forecasts from multimodel superensemble. Science 285, pp. 1548-1550.

Kuchment, L.S. (2004), The hydrological cycle and human impact on it. In Water Resources Management, [Eds. Arjen Y. Hoekstra, and Hubert H.G. Savenije], in Encyclopedia of Life Support Systems (EOLSS), Developed

under the Auspices of the UNESCO, Eolss Publishers, Oxford, UK, [http://www.eolss.net].

Kwadijk, J. C. J., M. Haasnoot, J. P. M. Mulder, M. M. C. Hoogvliet, A. B. M. Jeuken, R. A. A. van der Krogt, N. G. C. van Oostrom, H. A. Schelfhout, E. H. van Velzen, H. Waveren & M. J. M. de Wit (2010), Using adaptation tipping points to prepare for climate change and sea level rise: a case study in the Netherlands. Wiley Interdisciplinary Reviews: Climate Change 1 (5), pp. 729-740.

L'vovich, M. I (1979), World water resources and their future. American Geophysical Union, Washington DC, 415 p.

Laloy, E. & J. A. Vrugt (2011), High-dimensional posterior exploration using multiple-try MCMC simulation using $DREAM_{(ZS)}$ and high performance computing, Water Resources Research.

Leander, R. & A. T. Buishand (2007), Resampling of regional climate model output for the simulation of extreme river flows. Journal of Hydrology 332, pp. 487–496, doi:10.1016/j.jhydrol.2006.08.006.

Leemans, R. (1990), Global data sets collected and compiled by the Biosphere Project. Working Paper, IIASA-Laxenburg, Austria.

Legates, R. (2000), Climate models and the national assessment. George C. Marshall Institute, Washington, USA.

Lehner, B. & P. Döll (2004), Development and validation of a global database of lakes, reservoirs and wetlands. Journal of Hydrology 296, pp. 1–22, doi:10.1016/j.jhydrol.2004.03.028.

Lehner, B., P. Döll, J. Alcamo, T. Henrichs & F. Kaspar (2006), Estimating the impact of global change on flood and drought risks in Europe: a continental, integrated analysis. Climatic Change 75, pp. 273–299, doi: 10.1007/s10584-006-6338-4.

Liu, J., A. J. B. Zehnder & H. Yang (2009), Global consumptive water use for crop production: The importance of green water and virtual water. Water Resources Research 45, W05428, doi:10.1029/2007WR006051.

Loayza, N., E. Olaberría, J. Rigolini & L. Christiaensen (2009), Natural disasters and medium-term economic growth: the contrasting effects of different events on disaggregated output. World Bank, UN Assessment on the economics of disaster risk reduction.

Lu, J., G. Sun, S. G. McNulty & D. M. Amatya (2005), A comparison of six potential evapotranspiration methods for regional use in the southeastern united states. Journal of the American Water Resources Association 41 (3), pp. 621-633.

Mahanama, S. P. P. & R. D. Koster (2005), AGCM biases in evaporation regime: Impacts on soil moisture memory and land–atmosphere feedback. Journal of Hydrometeorology 6, pp. 656–669, doi: 10.1175/JHM446.1.

Manabe, S., P. C. D. Milly & R. D. Wetherald (2004), Simulated long-term changes in river discharge and soil moisture due to global warming. Hydrological Sciences Journal 49(4), pp. 625-642, doi: 10.1623/hysj.49.4.625.54429.

Martinec, J. (1975), Snowmelt – runoff model for stream flow forecasts. Nordic Hydrology 6(3), pp. 145–154, doi:10.2166/nh.1975.010.

Matalas N. C. & W. B. Langbein (1962), Information content of the mean. Journal of Geophysical Research 67(9), pp. 3441-3448.

Maurer, E. P., H. G. Hidalgo, T. Das, M. D. Dettinger & D. R. Cayan (2010), The utility of daily large-scale climate data in the assessment of climate change impacts on daily streamflow in California. Hydrology and Earth System Sciences 14, pp. 1125-1138, doi:10.5194/hess-14-1125-2010.

McGuffie, K. & A. Henderson-Sellers (2005), A climate modelling primer. John Wiley and Sons, 280 p.

McMillan, H., J. Freer, F. Pappenberger, T. Krueger & M. Clark (2010), Impacts of uncertain river flow data on rainfall-runoff model calibration and discharge predictions. Hydrological Processes 24, pp. 1270-1284, doi: 10.1002/hyp.7587.

Meehl, G. A. & J. M. Arblaster (2003), Mechanisms for projected future changes in south Asian monsoon precipitation. Climate Dynamics 21, pp. 659-675. doi: 10.1007/s00382-003-0343-3.

Meehl, G. A., F. Zwiers, J. Evans, T. Knutson, L. Mearns, & P. Whetton (2000), Trends in extreme weather and climate events: Issues related to modeling extremes in projections of future climate change. Bulletin of the American Meteorological Society 81(3), pp. 413–416.

Michelangeli, P.-A., M. Vrac, & H. Loukos (2009), Probabilistic downscaling approaches: application to wind cumulative distribution functions. Geophysical Research Letters 36, L11708, doi:10.1029/2009GL038401.

Miller, J. R., G. L. Russell & G. Caliri (1993), Continental-scale river flow in climate models. Journal of Climate 7, pp. 914-928.

Milly, P. C. D., K. A. Dunne & A.V. Vecchia (2005), Global pattern of trends in streamflow and water availability in a changing climate. Nature 438(17), pp. 347-350, doi:10.1038/nature04312.

Min, S. K. & A. Hense (2007), Hierarchical evaluation of IPCC AR4 coupled climate models with systematic consideration of model uncertainties. Climate Dynamics 29, 8530868, doi: 10.1007/s00382-007-0269-2.

Mitchell, T. D. & P. D. Jones (2005), An improved method of constructing a database of monthly climate observations and associated high-resolution grids. International Journal of Climatology 25, pp. 693-712.

Monteith, J. L. (1965), Evaporation and environment. Symposium Society for Experimental Biology 19, pp. 205–234.

Moss, R. H., J. A. Edmonds, K. A. Hibbard, M. R. Manning, S. K. Rose, D. P. van Vuuren, T. R. Carter, S. Emori, M. Kainuma, T. Kram, G. A. Meehl, J. F. B. Mitchell, N. Nakicenovic, K. Riahi, S. J. Smith, R. J. Stouffer, A. M. Thomson, J. P. Weyant & T. J. Wilbanks (2010), The next generation of scenarios for climate change research and assessment. Nature 463, pp. 747-756, doi:10.1038/nature08823.

Muleta, M. L. & J. W. Nicklow (2005), Sensitivity and uncertainty analysis coupled with automatic calibration for a distributed watershed model. Journal of Hydrology 306, pp. 127-145.

Murphy J. M., D. M. H. Sexton, D. N. Barnett, G.S. Jones, M.J. Webb, M. Collins & D.A. Stainforth (2004), Quantification of modelling uncertainties in a large ensemble of climate change simulations. Nature 430, pp. 768-772, doi:10.1038/nature02771.

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

New, M., M. Hulme & P. Jones (1999), Representing Twentieth-Century space-time climate variability. Part 1: Development of a 1961–90 mean monthly terrestrial climatology. Journal of Climate 12(3), pp. 829–856, doi: 10.1175/1520-0442(1999)012.

New, M., M. Hulme, M. & P. Jones (2000), Representing Twentieth-Century space–time climate variability. Part II: Development of 1901–96 monthly grids of terrestrial surface climate. Journal of Climate 13, pp. 2217–2238, doi: 10.1175/1520-0442(2000)013.

Nijssen, B., G. M. O'Donnel & D. P. Lettenmaier (2001), Predicting the discharge of global rivers. Journal of climate 14 (15), pp. 3307-3323, doi: 10.1175/1520-0442(2001)014.

Nohara, D., A. Kitoh, M. Hosaka & T. Oki (2006), Impact of climate change on river discharge projected by multimodel ensemble. Journal of Hydrometeorology 7, pp. 1076-1089.

Oki, T. & Y.C. Sud (1998), Design of total runoff integrating pathways (TRIP)—A global river channel network. Earth Interactions 2, pp. 1-36.

Oki, T., Y. Agata, S. Kanae, T. Saruhashi, D. Yang & K. Musiake (2001), Glob-

al assessment of current water resources using total runoff-integrating pathways. Hydrological Sciences Journal 46, pp. 983–996, doi: 10.1175/JHM531.1.

Oudin, L., F. Hervieu, C. Michel, C. Perrin, V. Andréassian, F. Anctil, & C. Loumagne (2005), Which potential evapotranspiration input for a lumped rainfall-runoff model? Part 2 – Towards a simple and efficient potential evapotranspiration model for rainfall-runoff modeling. Journal of Hydrology 303, pp. 290–306, doi: 10.1016/j.jhydrol.2004.08.026.

Pappenberger, F., H. L. Cloke, G. Balsamo, T. Ngo-Duc, & T. Oki (2009), Global runoff routing with the hydrological component of the ECMWF NWP system. International Journal of Climatology, doi: 10.1002/joc.2028.

Parkinson C. L., K. Y. Vinnikov & D. J. Cavalieri (2006), Evaluation of the simulation of the annual cycle of Arctic and Antarctic sea ice coverages by 11 major global climate models. Geophysical Research Letters 111, C07012, doi:10.1029/2005JC003408.

PCMDI (2010), Program for Climate Model Diagnosis and Intercomparison data portal, https://esg.llnl.gov:8443/index.jsp.

Perkins, S. E. & A. J. Pitman (2009), Do weak AR4 models bias projections of future climate changes over Australia? Climatic change 93, pp. 527–558, doi:10.1007/s10584-008-9502-1.

Piani, C., G. P. Weedon, M. Best, S. M. Gomes, P. Viterbo, S. Hagemann & J. O. Haerter (2010), Statistical bias correction of global simulated daily precipitation and temperature for the application of hydrological models. Journal of Hydrology 395 (3-4), pp. 199-215, doi: 10.1016/j.jhydrol.2010.10.024.

Pielke R., K. J. Beven, G. Brasseur, J. Calvert, M. Chahine, R. Dickerson, D. Entekhabi, E. Foufoula-Georgiou, H. Gupta, V. Gupta, W. Krajewski, E. P. Krider, W. K. M. Lau, J.J. McDonnell, W. Rossow, J. Schaake, J. Smith, S. Soroosh & E. F. Wood (2009), Climate change: the need to consider human forcings other than greenhouse gases. EOS, Transactions-American Geophysical Union 90(45), pp. 413.

Priestley, C. H. B. & R. J. Taylor (1972), On the assessment of surface heat flux and evaporation using large-scale parameters. Monthly Weather Review 100 (2), pp. 81–92.

Prudhomme, C. & H. Davies (2008), Assessing uncertainties in climate change impact analyses on the river flow regimes in the UK. Part 2: Future climate. Climatic Change 93, pp. 197–222, doi:10.1007/s10584-008-9461-6.

Prudhomme, C., N. Reynard & S. Crooks (2002), Downscaling of global climate models for flood frequency analysis: where are we now? Hydrological Pro-

cesses 16, pp. 1137–1150, doi:10.1002/hyp.1054.
Räisänen, J. (2007). How reliable are climate models? Tellus A 59, pp. 2–29. doi: 10.1111/j.1600-0870.2006.00211.x.
Räisänen, J., L. Ruokolainen & J. Ylhäisi (2009), Weighting of model results for improving best estimate of climate change. Climate Dynamics 35 (2-3), pp. 407-422, doi: 10.1007/s00382-009-0659-8
Reifen, C. & R. Toumi (2009), Climate projections: past performance no guarantee of future skill. Geophysical Research Letters 36, L13704, doi: 10.1029/2009GL038082.
Roeckner, E. (2006), Ensembles ECHAM5-MPI-OM SRESA1B run1, daily values. World Data Center for Climate. CERA-DB "Ensembles_MPEH5_ _SRA1B_1_D",http://cera-www.dkrz.de/WDCC/ui/Compact.jsp?acronym=ENSEMBLES_MPEH5_ SRA1B_1_D.
Roeckner, E. & Coauthors (2003), The atmospheric general circulation model ECHAM 5 Part 1. Max Planck Institut für Meteorologie, Hamburg, Germany.
Saha, S. & Coauthors (2006), The NCEP climate forecast system. Journal of Climate 19 (15), pp. 3483-3517, doi: 10.1175/JCLI3812.1.
Saha, S. & Coauthors (2010), The NCEP climate forecast system reanalysis. Bulletin of the American Meteorological Society, pp. 1015-1057.
Sanchez-Gomez, E., S. Somot & A. Mariotti (2009), Future change in the Mediterranean water budget projected by an ensemble of regional climate models. Geophysical Research Letters 36, L21401, doi: 10.1029/2009GL040120.
Schoups G. & J. A. Vrugt (2010), A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic and non-Gaussian errors. Water Resources Research 46, W10531, doi:10.1029/2009WR008933.
Sheffield, J., K. M. Andreadis, E. F. Wood & D. P. Lettenmaier (2009), Global and continental drought in the second half of the 20th century: severity-area-duration analysis and temporal variability of large-scale events. Journal of Climate 22(8), pp. 1962–1981, doi: 10.1175/2008JCLI2722.1.
Shiklomanov, I. A. (1997), Assessment of water resources and availability in the world. Comprehensive assessment of the freshwater resources of the world, Stockholm, Stockholm Environment Institute, 88 pp.
Singapore Red Cross (2010), Pakistan Floods:The Deluge of Disaster - Facts & Figures as of 15 September 2010. http://www.redcross.org.sg/news/pakistan-floods-the-deluge-of-disaster, retrieved October 18, 2010.
Smith, I. & E. Chandler (2010), Refining rainfall projections for the Murray Dar-

ling Basin of south-east Australia – the effect of sampling model results based on performance. Climatic Change 102, pp. 377-393, doi: 10.1007/s10584-009-9757-1.

Sorooshian, S. & J. A. Dracup (1980), Stochastic parameter estimation procedures for hydrologic rainfall-runoff models: correlated and heteroscedastic error cases. Water Resources Research 16, pp. 430-442.

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

Sperna Weiland, F.C., Tisseuil C., Dürr, H., Vrac, M. and Van Beek, L.P.H. (2011a), Selecting the optimal method to calculate daily global reference potential evaporation from CFSR reanalysis data. Hydrology and Earth System Sciences Discussions 8, pp. 7355-7398, doi:10.5194/hessd-8-7355-2011.

Sperna Weiland, F. C., L. P. H. van Beek, J. C. J. Kwadijk & M. F. P. Bierkens (2011b), On the suitability of GCM runoff fields for river discharge modeling; a case study using model output from HadGEM2 and ECHAM5. Journal of Hydrometeorology, in press.

Sperna Weiland, F. C., L. P. H. van Beek, J. C. J Kwadijk & M.F.P. Bierkens (2011c), Global pattern of change in runoff regimes for 2100. Climatic Change, under review.

Strzepek, K. M. & D. N. Yates (1997), Climate change impacts on the hydrologic resources of Europe: a simplified continental scale analysis. Climatic Change 36, pp. 79–92.

Sutanudjaja, E. H., L. P. H. Van Beek, S. M. de Jong, F. C. van Geer, & M. F. P. Bierkens (2011), Large-scale groundwater modeling using global datasets: a test case for the Rhine-Meuse basin. Hydrology and Earth System Sciences Discussions 8, pp. 2555-2608, doi:10.5194/hessd-8-2555-2011.

Taylor, K. E. (2001), Summarizing multiple aspects of model performance in a single diagram. Journal of Geophysical Research 106, pp. 7183-7192, doi:10.1029/ 2000JD900719.

Tebaldi, D. & R. Knutti (2007), The use of the multi-model ensemble in probabilistic climate projections. Philosophical Transactions of the Royal Society A. 365, pp. 2053-2075, doi:10.1098/rsta.2007.2076.

Tebaldi, C., R. L. Smith, D. Nychka & L. O. Mearns (2004), Quantifying uncertainty in projections of regional climate change: A Bayesian approach to the analysis of multimodel ensemble. Journal of Climate 18, pp. 1524-

1540, doi: 10.1.1.76.1308.

Te Linde, A. H., J. C. J. H. Aerts, A. M. R. Bakker & J. C. J. Kwadijk (2010), Simulating low probability peak discharges for the Rhine basin using resampled climate modeling data. Water Resources Research 46, W03512, doi:10.1029/2009WR007707.

Trenberth, K. E., A. Dai, R. M. Rasmussen & D. B. Parsons (2003), The changing character of precipitation. Bulletin of the American Meteorological Society 84, pp. 1205–1217, DOI: 10.1175/BAMS-84-9-1205.

Troccoli, A. & P. Kållberg (2004), Precipitation correction in the ERA-40 reanalysis. ERA-40 Project Rep. Series 13, ECMWF, Reading, UK, 6 p.

UN (2006), 2nd UN World Water Development Report: WWDRII data download page. http://wwdrii.sr.unh.edu/download.html.

USA today (2011), Mississippi flood damage could reach billions. http://www.usatoday.com/money/economy/2011-05-10-flood-impact_n.htm, retrieved June 24, 2011.

U.S. Global Change Research Program (2008), http://www.climas.arizona.edu/, retrieved June 24, 2011.

Uppala, S. M. & Coauthors (2005), The ERA-40 re-analysis. Quarterly Journal of the Royal Meteorological Society 131, pp. 2961–3012.

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

Van Beek, L. P. H. & M. F. P. Bierkens (2009), The global hydrological model PCR-GLOBWB: Conceptualization, parameterization and verification. Report Department of Physical Geography, Utrecht University, Utrecht, Netherlands, available at: http://vanbeek.geo.uu.nl/suppinfo/vanbeekbierkens2009.pdf.

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

Van den Hurk, B., M. Hirschi, C. Schär, G. Lenderink, E. van Meijgaard, A. van Ulden, B. Rockler, S. Hagemann, P. Graham, E. Kjellström & R. Jones (2004), Soil control on runoff response to climate change in regonal climate model simulations. Journal of Climate 18, pp. 3536-3551, doi: 10.1175/JCLI3471.1.

Van Soesbergen, A. J. J. & M. Mulligan (2010), Modelling climate change impacts on water resources of global dams. Journal of Hydrology Conference, Octobre 2010, San Diego.

Varis, O., T. Kajander & R. Lemmela (2004), Climate water: from climate models to water resources management and vice versa. Climatic Change 66, pp. 321–344, doi: 10.1023/B:CLIM.0000044622.42657.d4.

Vaze, J., D. Post, F. Chiew, J.-M. Perraud, N. Viney & J. Teng (2010), Climate non-stationarity - Validity of calibrated rainfall-runoff models for use in climate change studies. Journal of Hydrology 394 (3-4), pp. 447-457. doi: 10.1016/j.jhydrol.2010.09.017.

Verdin, K. L. & S. K. Greenlee (1996), Development of continental scale digital elevation models and extraction of hydrographic features. In: Proceedings, Third International Conference/Workshop on Integrating GIS and Environmental Modeling, Santa Fe, New Mexico, January 21-26.

Viviroli, D., H. H. Dürr, B. Messerli, M. Meybeck & R. Weingartner (2007), Mountains of the world – water towers for humanity: typology, mapping and global significance. Water Resources Research 43(7), W07447. doi:10.1029/2006WR005653.

Viviroli, D., D. R. Archer, W. Buytaert, H. J. Fowler, G. B. Greenwood, A. F. Hamlet, Y. Huang, G. Koboltschnig, M. I. Litaor, J. I. López-Moreno, S. Lorentz, B. Schädler, K. Schwaiger, M. Vuille, & R. Woods (2011), Climate change and mountain water resources: overview and recommendations for research, management and politics. Hydrology and Earth System Sciences 15, pp. 471-504, doi:10.5194/hess-15-471-2011.

Voisin, N., A. W. Wood & D. P. Lettenmaier (2008), Evaluation of precipitation products for global hydrological prediction, Journal of hydrometeorology 9, pp. 388-407, doi: 10.1175/2007JHM938.1.

Vörösmarty, C. J., C. A. Federer & A. L. Schloss (1998a), Potential evaporation functions compared on U.S. watersheds: Implications for global-scale water balance and terrestrial ecosystem modeling. Journal of Hydrology 207, pp. 147-169, doi:10.1016/S0022-1694(98)00109-7.

Vörösmarty, C. J., B. Fekete & B. A. Tucker (1998b), River Discharge Database, Version 1.1 (RivDIS v1.0 supplement). Available through the Institute for the Study of Earth, Oceans, and Space/University of New Hampshire, Durham NH, USA.

Vörösmarty, C. J., P. Green, J. Salisbury & R. B. Lammers (2000), Global water resources: Vulnerability from climate change and population growth. Science 289 (5477), pp. 284–288, doi: 10.1126/science.289.5477.284.

Vrugt, J. A. & B. A. Robinson (2007), Treatment of uncertainty using ensemble methods: Comparison of sequential data assimilation and Bayesian model averaging. Water Resources Research 43, W01411,

doi:10.1029/2005WR004838.
Vrugt, J. A., H. V. Gupta, W. Bouten & S. Sorooshian (2003), A shuffled complex evolution Metropolis algorithm for optimization and uncertainty assessment of hydrological model parameters. Water Resources Research 39 (8), 1201, doi:10.1029/2002WR001642.
Vrugt, J. A., C. G. H. Diks, H. V. Gupta, W. Bouten & J. M. Verstraten (2005), Improved treatment of uncertainty in hydrologic modeling: Combining the strengths of global optimization and data assimilation. Water Resources Research 41, W01017, doi:10.1029/2004WR003059.
Vrugt, J. A., C. J. F. ter Braak, M. P. Clark, J. M. Hyman & B. A. Robinson (2008), Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with Markov chain Monte Carlo simulation. Water Resources Research 44, W00B09, doi:10.1029/2007WR006720.
Vrugt, J. A., C. J. F. ter Braak, C. G. H. Diks, B. A. Robinson, J. M. Hyman & D. Higdon (2009), Accelerating Markov chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling. International Journal of Nonlinear Sciences and Numerical Simulation 10(3), pp. 273-290.
Wada, Y., L. P. H. van Beek, D. Viviroli, H. Dürr, R. Weingartner & 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, C. M. van Kempen, J. W. T. M. Reckman, S. Vasak & M. F. P. Bierkens (2010), Global depletion of groundwater resources. Geophysical Research Letters 37, L20402, doi:10.1029/2010GL044571.
Wada, Y., L. P. H. van Beek, D. Viviroli, H. H. Dürr, R. Weingartner & M. F. P. Bierkens (2011), Global monthly water stress: II. Water demand and severity of water stress. Water Resources Research 47, W07518, doi:10.1029/2010WR009792.
Weiβ, M. & L. Menzel (2008), A global comparison of four potential evapotranspiration equations and their relevance to stream flow modelling in semi-arid environments. Advances in Geosciences 18, pp. 15-23.
Wesseling, C.G., D. Karssenberg, W. P. A. van Deursen & P. A. Burrough (1996), Integrating dynamic environmental models in GIS: the development of a Dynamic Modelling language. Transactions in GIS 1, pp. 40-48.
Widén-Nilsson, E., S. Halldin & C.-Y. Xu (2007), Global water balance modelling with WASMOD-M: parameter estimation and regionalization. Journal

of Hydrology 340, pp. 105-118, doi:10.1016/j.jhydrol.2007.04.002.

Wilby, R. L., C. W. Dawson & E. M. Barrow (2002), SDSM - a decision support tool for the assessment of regional climate change impacts. Environmental Modelling Software 17, pp. 145-157, doi: 10.1016/S1364-8152(01)00060-3.

Wilby, R. L., O. J. Tomlinson & C. W. Dawson (2003), Multi-site simulation of precipitation by conditional resampling. Climate Research 23, pp. 183-194, doi:10.3354/cr023183.

Wilby, R. L., T. M. L. Wigley, D. Conway, P. D. Jones, B. C. Hewitson, J. Main & D. S. Wilks (1998), Statistical downscaling of general circulation model output: A comparison of methods. Water Resources Research 34(11), pp. 2995–3008, doi: 10.1029/98WR02577.

Winsemius, H. C., H. H. G. Savenije, A. M. J. Gerrits, E. A. Zapreeva & R. Klees (2006), Comparison of two model approaches in the Zambezi river basin with regard to model reliability and identifiability. Hydrology and Earth System Sciences 10, pp. 339–352, 10.5194/hess-10-339-2006.

Wood, E. F., D. P. Lettenmaier & V. G. Zartarian (1992), A landsurface hydrology parameterization with subgrid variability for general circulation models. Journal of Geophysical Research 97, D3, pp. 2717–2728.

Wood, A. W., L. R. Leung, V. Sridhar & D. P. Lettenmaier (2004), Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs. Climatic Change 62, pp. 189–216, doi: 10.1023/B:CLIM.0000013685.99609.9e.

World Bank (2010), Natural hazards, unnatural disasters; the economics of effective prevention. United Nations, 276 pp.

World data Center for Climate (2010), CERA-gateway. http://cera-www.dkrz.de/CERA/.

Xu, Y., X. Gao & F. Giorgi (2010), Upgrades to the reliability ensemble averaging method for producing probabilistic climate-change projections. Climate Research 41, pp. 61-81, doi: 10.3354/cr00836.

Zaitchik, B. F., M. Rodell & F. Olivera (2010), Evaluation of the global land data assimilation system using river discharge data and a source-to-sink routing scheme. Water Resources Research 46, W06507, doi:10.1029/2009WR007811.

Summary

Background

Multiple hydrological impact studies have assessed the hydrological consequences of climate change. Yet, especially in applied water management studies, often only one or a few GCMs are used and limited attention is given to the uncertainties present in future climate projections. Major investments for the implementation of adaptation measures are based on the outcomes of these overconfident and possibly biased assessments.
In this thesis an extensive analysis of the inherent uncertainties present in the modeling chain of hydrological impact studies is conducted. In addition, future global hydrological changes are derived and methods for the interpretation of the accompanying uncertainties are explored.

Uncertainty analysis

The first step of this study evaluates the influence of PCR-GLOBWB parameter and forcing data uncertainty on modeled discharge estimates. To this end PCR-GLOBWB was forced with three different forcing datasets and 250 different parameter combinations. The results show that, with simple Latin Hypercube sampling, parameter combinations can be obtained that outperform the default model parameterization. Yet, optimal parameter estimates highly depend on the forcing dataset used. This indicates that a hydrological model forced with often biased GCM datasets is preferably calibrated for each dataset independently. Otherwise, calibration attempts should be limited and parameter values could best be obtained from available global datasets, as has been done within this study.
In a second step the global hydrological model PCR-GLOBWB is forced with meteorological datasets from an ensemble 12 GCMs. The resulting annual average discharge cycles are compared with discharge observations. This comparison reveals large variations between GCMs and large deviations of GCM derived discharge from discharge observations. The analysis is extended with the evaluation of the usability of GCM runoff for hydrological studies. The results indicate that especially for large catchments full hydrological modeling outperforms the use of GCM-generated runoff. Runoff routing and correct representation of subsurface hydrological processes, as present in hydrological models, introduce realistic flow velocities and consequently a correct reproduction of discharge extremes.

Future hydrological impacts of climate change

To asses future runoff changes PCR-GLOBWB is forced with datasets from 12 GCMs for the IPCC A1B scenario. The change projections diverge widely. Yet, when discharge change is calculated relative to discharge amounts obtained for the 20th century control experiment for each individual GCM, ensemble consistent and significant changes can be found. By 2100 runoff decreases are projected for southern Europe, southern Australia and the south and north of Africa and runoff increases are projected for sub-Arctic and Monsoon influenced regions.

In a final step several performance based GCM weighting and selection methods have been investigated. These methods are used to derive weighted average future runoff changes from the ensemble of 12 GCMs and to quantify the accompanying weighted uncertainty ranges. The method which considered inter-model similarity, for both the current and future climate, performed best.

Nederlandse samenvatting

De aanleiding

De toenemende emissie van broeikasgassen heeft geleid tot veranderingen in het klimaat. Wereldwijd zijn temperatuurstijgingen waargenomen. De gebieden die ge-troffen worden door droogtes zijn in oppervlakte toegenomen en neerslag frequenties en intensiteit zijn verandert. Dit heeft gevolgen voor de globale waterbeschikbaarheid en afvoer regimes van rivieren.

De laatste decennia zijn vele hydrologische klimaat effect studies uitgevoerd om de mogelijke effecten van klimaatverandering op rivierafvoeren en waterbeschikbaarheid in kaart te brengen. Met name in toegepaste water management studies wordt echter vaak nog te weinig rekening gehouden met de onzekerheden die gepaard gaan met een dergelijke analyse. Er wordt hier regelmatig gebruik gemaakt van slechts één of een klein aantal klimaatmodellen. De resultaten worden vervolgens gepresenteerd zonder de bijbehorende onzekerheid te kwantificeren.

In deze studie wordt een uitgebreide analyse gemaakt van de onzekerheden die aanwezig zijn in de modelketen van hydrologische klimaat effect studies. Daarnaast worden toekomstige veranderingen in wereldwijde waterbeschikbaarheid berekend. Tot slot worden verschillende methoden voor de kwantificatie van de onzekerheden waarmee deze veranderingen gepaard gaan onderzocht.

Onzekerheidsanalyse

Gevoeligheidsanalyse PCR-GLOBWB

In een eerste stap is het globale hydrologische model PCR-GLOBWB gedraaid met achtereenvolgens drie verschillende meteorologische datasets om de gevoeligheid van het model voor onzekerheid in meteorologische model forcering in kaart te brengen. Vervolgens is het model gedraaid met 250 verschillende parameter sets (verkregen met behulp van latin hyper cube sampling) om zo in kaart te brengen wat de gevoeligheid van het model is voor parameteronzekerheid. Deze analyse toont aan dat de optimale parameter schattingen sterk afhankelijk zijn van de gebruikte meteorologische forcering. Dit geeft aan dat in klimaat effect studies, waarbij vaak gebruik wordt gemaakt van datasets van meerdere klimaatmodellen, het hydrologische model ofwel voor de datasets van ieder individueel klimaatmodel opnieuw gekalibreerd moet worden of dat parameter waarden verkregen moeten worden uit bestaande globale fysische datasets. Het laatste wordt toegepast in deze studie.

Bruikbaarheid meteorologische data klimaatmodellen
In een tweede stap zijn PCR-GLOBWB model runs uitgevoerd voor het huidige klimaat met meteorologische datasets van 12 globale klimaatmodellen (ook wel GCMs genoemd). Op basis van deze runs is voor iedere GCM apart de langjarige gemiddelde rivierafvoer cyclus bepaald en deze is vergeleken met afvoer metingen. De vergelijking laat zien dat er grote verschillen zijn tussen GCMs en dat afvoer berekend op basis van de GCM datasets grote afwijkingen vertonen van de waargenomen rivierafvoeren. Alhoewel met behulp van een eenvoudige bias-correctie de verschillen en afwijkingen sterk kunnen worden gereduceert, blijven de afwijkingen in afvoer extremen en interjaarlijkse variabiliteit ook na correctie aanwezig.

Bruikbaarheid cel specifieke afvoer klimaatmodellen
Bovenstaande analyse naar de bruikbaarheid van meteorologische GCM data is uitgebreid met de evaluatie van de bruikbaarheid van de cel specifieke GCM afvoer. Deze analyse laat zien dat, vooral voor rivieren groter dan de Rijn, volledige hydrologische modelering op basis van meteorologische GCM data een beter resultaat oplevert dan direct gebruik van GCM afvoer. Transport van afvoer langs een gedetailleerd rivier netwerk en een uigebreide weergave van ondergrondse hydrologische processen, zoals geïmplementeerd in de meeste hydrologische modellen, introduceren realistische stroomsnelheden en hiermee een correcte reproductie van waargenomen timing en hoeveelheid rivierafvoer.

Toekomstige hydrologische veranderingen

Hydrologische veranderingen
In het tweede deel van deze studie is PCR-GLOBWB geforceerd met meteorologische datasets van 12 verschillende GCMs voor de toekomst. De berekende veranderingen vertonen grote verschillen tussen de GCMs. Maar als de relatieve toekomstige veranderingen berekend met de individuele GCMs worden gebruikt, kunnen significante en consistente veranderingen in kaart worden verkregen. Rond 2100 lijkt rivier afvoer af te nemen in zuid Europa, zuid Australië en zuid en noord Afrika. Daarnaast zijn toenames berekend voor subarctische gebieden en gebieden waar rivierafvoer sterk beïnvloedt wordt door de Moesson, met name in Azië.

Weging hydrologische projecties op basis van GCM performance
In een laatste stap zijn verschillende methoden onderzocht waarbij GCMs worden gewogen of geselecteerd op basis van performance. Deze methoden zijn toegepast om een gewogen gemiddelde verandering in waterbeschikbaarheid te berekenen op basis van toekomstige hydrologische projecties berekend met de 12 GCMs. Daarnaast zijn de onderzochte methoden geschikt om onzekerheidsbanden rond de gewogen gemiddelde veranderingen te bepalen.

Dankwoord

Na ruim drie jaar hard werken zijn alle wetenschappelijke delen van dit proefschrift afgerond en rest alleen het dankwoord nog.

Allereerst wil ik Rens, Marc en Jaap heel erg bedanken. Jaap en Marc, bedankt voor het schrijven van het plan waarmee jullie geld binnen Deltares hebben gekregen zodat ik deze promotie heb kunnen uitvoeren. Marc bedankt voor je goede ideeen en je snelle aanpak als het ging om het uitzetten van de lijnen van een artikel of het oplossen van een probleem. De wijze waarop jij af en toe teksten van artikelen hebt weten aan te scherpen heeft veel geholpen en ik waardeer je bijdrage aan al die replies op kritisch review commentaren erg.

Rens, van jou heb ik heel veel geleerd. Alleen al omdat je manier van werken soms volledig anders is dan die van Marc, en ook dan die van mij. Jij kunt een vraagstuk zorgvuldig analyseren en doorzien tot in de kleinste details. Iets waarvoor bij mij soms het geduld ontbreekt, maar waarvan ik hoop dat ik het in de loop van de tijd een beetje overgenomen heb en ook later nog zal blijven doen. Je deur heeft de afgelopen jaren altijd voor me opengestaan. Waar dit in het begin vooral nodig was om me te helpen met mijn gevecht met python en praster hebben we later steeds meer over de inhoud en hydrologie gepraat en hebben je kritische reviews van de draften van mijn papers me veel geholpen. En niet te vergeten, bedankt dat je PCR-GLOBWB hebt gemaakt!

Jaap, bedankt dat je me überhaupt gevraagd hebt of ik dit promotie onderzoek wilde uitvoeren. Ik vind het heel leuk om met je samen te werken. Geniaal hoe uitgespoken jij je mening en ideeën af en toe kunt uiten. Bedankt ook dat je door de jaren heen fan bent gebleven van mijn onderzoek, ook al is het wel een beetje een "*academische exercitie geworden*" en had je het als kennismanager vaak al druk genoeg.

Albrecht bedankt dat je bij wilde springen toen Jaap me liet zitten. Hoewel het einde al in zicht kwam en ik redelijk mijn eigen gang ging ben ik blij dat er nog iemand op de gang rond liep die goed wist waarmee ik bezig was, iemand die af en toe een keertje extra binnenliep voor een praatje. Bedankt ook voor je bijdrage aan de twee papers die we dit laatste jaar nog hebben geschreven.

Dan nog een dankwoord aan de 'buitenlanders' waarmee ik tijdens mijn promotie heb mogen samenwerken. Jasper het was erg leuk om met je te werken. Je gedrevenheid en

enthousiasme zijn enorm en werken besmettelijk. Ook al had je het heel druk en hebben we je een beetje in een lastige positie gebracht, je hebt wel uitgebreid de tijd en moeite genomen om me te helpen en hard aan onze paper geschreven. Ik ben erg blij dat jij ook in Irvine was en heb goede herinneringen aan onze gesprekken in die tijd.
Clement, thanks to Hans Dürr our paths crossed and you came with a nice research question where we could apply our model and data. I hope we will soon publish our first paper and after that we should continue just as long as we finish the second. I start realizing how hard bias-correction can be.

Verder dank aan alle Utrechtenaren voor de gezelligheid bij de koffiemachine en de rondjes wandelen in de botanische tuinen. Arien, Edwin, Wiebe, Yoshi en alle anderen, ook al was ik er vaak maar een dag per week, het was altijd weer leuk om te horen waar jullie mee bezig waren.
Dank ook aan alle Deltarianen die interesse zijn blijven tonen in mij en mijn werk. OWB-ers, de gedrevenheid waarmee jullie aan projecten werken werkt aanstekelijk en daar heb ik zeker ook een deel van mijn drive aan te danken. HYD, bedankt voor een warm welkom, ook al was het af en toe wat lastig combineren met het afronden van dit proefschrift. Jaap en Ronald, bedankt voor het bouwen van FEWS-World en jullie support. Naze, it was great to have someone working with the same model and the same team. I am sure you will successfully defend your thesis in the near future as well. Simone, ik hoop dat we nog lang buren blijven op het werk en tijd overhouden om bij te kletsen tijdens een lunchwandeling. Marc, ik hoop dat we af en toe nog eens samen aan een project kunnen werken en bedankt dat je ooit een balletje voor me hebt opgegooid bij Jaap. Jan en Marjolijn bedankt voor jullie interesse en de vele literatuur updates.

Dan wil ik mijn vrienden bedanken voor alle leuke en gezellige afleiding. Ook hielp het dat ik met vele van jullie de struggles van een PhD kon delen. Het wordt hoog tijd om weer eens te onderzoeken of je op Brinta echt niet kunt sleeën (of anders varen). Monique bedankt voor de slaapplaats in Utrecht, de partijtjes squash en gezellige etentjes.

Lieve ouders, allereerst bedankt dat ik nog steeds een bed in Zeist heb en dat ik met enige (on)regelmaat aan mag komen waaien. Bedankt ook voor jullie steun, zorg, vertrouwen en (goed bedoelde, maar soms genegeerde) wijze raad in de afgelopen negenentwintig jaar.

Dan, als allerlaatste de allerliefste en leukste collega/partner. Bedankt dat je er altijd voor me bent Kymo. Ik ben blij dat je nog steeds geduldig kunt luisteren naar al mijn enthousiaste geratel over dit onderzoek en dat je hebt meegedacht als er weer een lastige situatie op mijn pad kwam. En… zonder jou waren de wereldkaarten nooit zo mooi geworden!

Curriculum Vitae

Frederiek Sperna Weiland was born on 8 November 1982 in Goes, The Netherlands. In 2001 she started her Bachelor of Science (BSc.) in Environmental Sciences at the Radboud University of Nijmegen. During this Bachelor she was triggered by several courses on water management and related topics and decided to continue her studies at Wageningen University with a Master of Science (MSc) in Hydrology and Quantitative Water Management. For her MSc thesis she visited Lancaster University for half a year and, supervised by Dr. F. Pappenberger, she performed an uncertainty analysis on a river flow model using distributed flow velocity data which she collected at the river Ribble. During her time in Lancaster she discovered the interesting and fun sides of scientific research. In 2006 she obtained her Master's degree and started working at Deltares (former Delft Hydraulics). After two years of working in the area of flood forecasting and related subjects she got the oppertunity to do a PhD while remaining employed at Deltares. She started this research in 2008 at the Physical Geography group at Utrecht University with Prof. Dr. Ir. M. Bierkens and Dr. R. van Beek. She investigated the uncertainties accompanying hydrological climate impact assessments. After finishing her PhD, she will remain working at Deltares, combining market oriented project work with scientifiic research.

Selected publications

Peer reviewed papers

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

Sperna Weiland, F. C., C. Tisseuil, H. H. Dürr, M. Vrac & L. P. H. van Beek (2011), Selecting the optimal method to calculate daily global reference potential evaporation from CFSR reanalysis data. Hydrology and Earth System Sciences Discussions 8, pp. 1-43, doi:10.5194/hessd-8-1-2011.

Sperna Weiland, F. C., L. P. H. van Beek, A. H. Weerts & M. F. P. Bierkens (2011), Extracting information from an ensemble of GCMs to reliably assess global future runoff change. Journal of Hydrology, doi:10.1016/j.jhydrol.2011.03.047.

Sperna Weiland, F. C., L. P. H. van Beek, J. C. J. Kwadijk & 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. Journal of Hydrometeorology, *in press.*

Sperna Weiland, F. C., L. P. H. van Beek, J. C. J. Kwadijk & M. F. P. Bierkens (2011), Global patterns of change in runoff regimes for 2100. Climatic Change, *under review.*

Sperna Weiland, F. C., J. A. Vrugt, L. P. H. van Beek, A. H. Weerts & M. F. P. Bierkens (2011), Exploring calibration strategies for the global hydrological model PCR-GLOBWB in light of forcing and parameter uncertainty, *submitted to Water Resources Research.*

Gain, A. K., W. W. Immerzeel, F. C. Sperna Weiland & M. F. P. Bierkens (2011), Impact of climate change on the stream flow of the lower Brahmaputra: trends in high and low flows based on discharge-weighted ensemble modelling. Hydrology and Earth System Sciences 15, pp. 1537-1545, doi:10.5194/hess-15-1537-2011.

Weerts, A. H., J. Schellekens & F. C. Sperna Weiland (2010), Real-Time geospatial data handling and forecasting: Examples from Delft-FEWS forecasting platform/system. IEEE Selected topics in applied earth observations and remote sensing 3 (3), pp. 386-394, doi: 10.1109/JSTARS.2010.2046882.

Conference abstracts

Sperna Weiland, F. C., L. P. H. van Beek, J. C. J. Kwadijk & M. F. P. Bierkens (2010), Interpreting global hydrological changes from climate model projections. Oral presentation, Conference of the Global Catchment Initiative, Bonn.

Sperna Weiland, F. C., L. P. H. van Beek, J. C. J. Kwadijk & M. F. P. Bierkens (2010), Translating multi-GCM derived future discharge projections into information suitable for water management. Oral presentation, Conference of the journal of hydrology, San Diego.

Sperna Weiland, F. C., L. P. H. van Beek, J. C. J. Kwadijk & M. F. P. Bierkens (2010), Global scale impacts of climate change on hydrological regimes. Poster presentation, EGU General Assembly 2010, Vienna.

Sperna Weiland, F. C., L. P. H. van Beek, J. C. J. Kwadijk & M. F. P. Bierkens (2009), Modelling hydrological effects of climate change on a global scale. Oral presentation, EGU General Assembly 2009, Vienna.