

## Significant uncertainty in global scale hydrological modeling from precipitation data errors



Frederiek C. Sperna Weiland <sup>a,b,\*</sup>, Jasper A. Vrugt <sup>c,d,e</sup>, Rens (L.) P.H. van Beek <sup>b</sup>, Albrecht H. Weerts <sup>a,f</sup>, Marc F.P. Bierkens <sup>b,g</sup>

<sup>a</sup> Deltas, P.O. Box 177, 2600 MH Delft, The Netherlands

<sup>b</sup> Department of Physical Geography, Utrecht University, P.O. Box 80115, 3508 TC Utrecht, The Netherlands

<sup>c</sup> Department of Civil and Environmental Engineering, University of California, Irvine, 4130 Engineering Gateway, Irvine, CA 92697-2175, USA

<sup>d</sup> Department of Earth System Science, University of California, Irvine, USA

<sup>e</sup> Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, Amsterdam, The Netherlands

<sup>f</sup> Environmental Sciences, Wageningen University and Research Centre, P.O. Box 47, 7600 AA Wageningen, The Netherlands

<sup>g</sup> Deltas, P.O. Box 85467, 3508 AL Utrecht, The Netherlands

### ARTICLE INFO

#### Article history:

Received 8 March 2015

Received in revised form 31 July 2015

Accepted 28 August 2015

Available online 4 September 2015

This manuscript was handled by Konstantine P. Georgakakos, Editor-in-Chief, with the assistance of Matthew McCabe, Associate Editor

#### Keywords:

Global hydrological modeling

Forcing uncertainty

Model calibration

Parameter uncertainty

### SUMMARY

In the past decades significant progress has been made in the fitting of hydrologic models to data. Most of this work has focused on simple, CPU-efficient, lumped hydrologic models using discharge, water table depth, soil moisture, or tracer data from relatively small river basins. In this paper, we focus on large-scale hydrologic modeling and analyze the effect of parameter and rainfall data uncertainty on simulated discharge dynamics with the global hydrologic model PCR-GLOBWB. We use three rainfall data products; the CFSR reanalysis, the ERA-Interim reanalysis, and a combined ERA-40 reanalysis and CRU dataset. Parameter uncertainty is derived from Latin Hypercube Sampling (LHS) using monthly discharge data from five of the largest river systems in the world. Our results demonstrate that the default parameterization of PCR-GLOBWB, derived from global datasets, can be improved by calibrating the model against monthly discharge observations. Yet, it is difficult to find a single parameterization of PCR-GLOBWB that works well for all of the five river basins considered herein and shows consistent performance during both the calibration and evaluation period. Still there may be possibilities for regionalization based on catchment similarities. Our simulations illustrate that parameter uncertainty constitutes only a minor part of predictive uncertainty. Thus, the apparent dichotomy between simulations of global-scale hydrologic behavior and actual data cannot be resolved by simply increasing the model complexity of PCR-GLOBWB and resolving sub-grid processes. Instead, it would be more productive to improve the characterization of global rainfall amounts at spatial resolutions of 0.5° and smaller.

© 2015 Elsevier B.V. All rights reserved.

### 1. Introduction

Hydrological models synthesize our knowledge of the rainfall–storage–runoff transformation. These models are used widely for flood forecasting, and investigation of water resources systems and climate change, and use relatively simple mathematical equations to conceptualize and aggregate the complex myriad of spatially distributed and highly interrelated water, energy and vegetation processes in a watershed. As a result, most of the model parameters in hydrologic models do not represent direct measurable quantities but can only be derived indirectly by calibration

against a historical record of input–output data (Beven and Binley, 1992; Vrugt et al., 2005; Gosling and Arnell, 2011). In this process the parameters are adjusted in such a way that the difference between the simulated model output and observations is minimized (Gupta et al., 1998; Vrugt et al., 2003).

In the past decades much progress has been made toward the development of efficient calibration strategies for hydrological models and the treatment and quantification of uncertainty. Most of this work has used relatively simple lumped and semi-distributed hydrological models that represent watersheds with area ranging between 100 and 10,000 km<sup>2</sup> (Sorooshian and Dracup, 1980; Gupta et al., 1998; Andréassian et al., 2001; Vrugt et al., 2003; Muleta and Nicklow, 2005; Balin et al., 2010; McMillan et al., 2010; Vaze et al., 2010, amongst many others). Less attention has been paid to calibration of global hydrological

\* Corresponding author at: P.O. Box 177, 2600 MH Delft, The Netherlands.  
Tel.: +31 88 335 8230.

E-mail address: [frederiek.sperna@deltas.nl](mailto:frederiek.sperna@deltas.nl) (F.C. Sperna Weiland).

models (GHMs) that attempt to simulate (predict) terrestrial-scale soil moisture, recharge, surface runoff, groundwater table, and discharge dynamics. Some notable exceptions include the recent work of [Troy et al. \(2008\)](#), [Gosling and Arnell \(2011\)](#), [Nasonova et al. \(2011\)](#) and [Pappenberger et al. \(2011\)](#). Not only do GHMs pose significant computational challenges, they also require a wealth of input data to accurately characterize global scale variations in land-use, soil type, elevation, climate conditions, and groundwater table depths (amongst others). Yet, all these data exhibit a large spatial variability and high degree of uncertainty which compromises, sometimes severely, the predictive capability of GHMs ([Beven and Cloke, 2012](#); [Duan et al., 2006](#); [Teuling et al., 2009](#)).

The lack of high-quality and high-resolution input and forcing data, and considerable CPU-requirements of GHMs, necessitates the use of a very coarse grid resolution to resolve global scale hydrologic fluxes and state variables ([Haddeland et al., 2011](#)). This introduces a very high level of process aggregation, which unavoidably introduces significant structural errors and requires appropriate sub-grid parameterization ([Beven and Cloke, 2012](#)). Moreover, the (discharge) datasets available for model evaluation are limited and their accuracy and reliability varies considerably over the world ([Renard et al., 2010](#)). Consequently appropriate parameterizations will not be spatially uniform and can only be tested and optimized locally ([Beven and Cloke, 2012](#); [McMillan et al., 2010](#)).

Several contributions can be found in the hydrological literature that have investigated the role of parameter and forcing data uncertainty in GHMs. For instance, [Fekete et al. \(2004\)](#) analyzed the influence of precipitation data uncertainty on simulated global runoff with the UHN global water balance model. The uncertainty in simulated runoff was of similar size as the uncertainty in the precipitation and especially large in semi-arid regions. A similar study by [Biemans et al. \(2009\)](#) used the global vegetation and hydrology model LPJmL to evaluate seven precipitation datasets for discharge simulation of 294 basins. The uncertainty in simulated discharge was found to be about three times higher than the uncertainty in basin average precipitation. These findings make a strong case for hydrological model calibration using the meteorological dataset selected for the final model application. [Pappenberger et al. \(2011\)](#) concluded that the quality of meteorological data has improved considerably in the past decade, which hence should improve our ability to simulate the hydrology of large river basins.

Recent studies by Gosling and coworkers have investigated the sensitivity of the Mac-PDM.09 GHM to parameter and forcing (precipitation) data. The study of [Gosling et al. \(2010\)](#) used fourteen different model simulations to determine the sensitivity of the model output to variations in the field capacity and variability of the soil moisture capacity. The second study, published in [Gosling and Arnell \(2011\)](#) used an ensemble of 9 different scenarios from 21 different GCMs to analyze the impact of forcing data uncertainty. More recently, [Nasonova et al. \(2011\)](#) investigated the effect of different forcing datasets on the SWAT simulated water balance. Results demonstrate that the simulated surface runoff strongly depends on the precipitation dataset being used. This finding is perhaps not surprising, but highlights the need for accurate forcing data and information on river regulation in global hydrologic modeling.

Several other recent studies have focused attention on the effect of model selection in global hydrologic modeling. For instance, [Haddeland et al. \(2011\)](#) combined several Land Surface Models (LSMs) and GHMs in the WATCH project to generate a multi-model ensemble of the global water cycle. The ensemble of simulations exhibited a large spread, even though the constituent models resolved similar processes, but with differing parameter values. [Gudmundsson et al. \(2011\)](#) also conducted a multi-model comparison in the context of the WATCH project and demonstrated that

the ensemble spread was particularly large during low flow events, but the ensemble mean reliably estimated mean and extreme flows.

Thus far, we have focused our attention on contributions whose main goal was to illustrate the effect of parameter, model, or forcing data uncertainty in global hydrologic modeling, without recourse to parameter estimation. Several authors have focused on global scale parameter estimation. For instance, [Fekete et al. \(2002\)](#) used a correction factor in the WBMplus model to match discharge data from neighboring stations. [Troy et al. \(2008\)](#) calibrated their global scale hydrological model at only 2% of the grid cells, and used the remaining cells to explore the potential for regionalization of the parameters and to assess sub-basin variability. Another study by [Döll et al. \(2003\)](#) considered the calibration of the GHM WaterGAP model. This work demonstrated that careful tuning of the runoff coefficient significantly improved the agreement between the observed and simulated discharge data. [Widén-Nilson et al. \(2007\)](#) calibrated the global water balance model WASMOD-M using measurements of average areal discharge, thereby avoiding problems of flow regulation. Basin specific values of the WASMOD-M model parameters were selected from a sample of 1680 different parameter combinations. [Wood et al. \(1992\)](#) calibrated the global VIC model using the well-known Shuffled Complex Evolution (SCE) algorithm ([Duan et al., 1992](#); [Nijssen et al., 2001](#)). Calibration reduced the annual average bias and the relative Root Mean Square Error (RMSE) of the monthly discharge values from 62 to 37% and 29 to 10%, respectively.

Altogether, published studies demonstrate that calibration of GHMs is difficult, and hampered by (1) a lack of quality and high-resolution input data to accurately characterize surface and subsurface properties, (2) significant uncertainty in the forcing data, (3) high computational demands, and (4) limited availability of reliable discharge observations. The present study will show that rainfall uncertainty constitutes the largest source of error in global scale hydrologic modeling, while parameter uncertainty explains only a minor source of streamflow simulation uncertainty. Our analysis is based on a single model, and unlike previous studies jointly considers the effect of parameter and rainfall data uncertainty in modeling discharge dynamics of some of the largest rivers in the world. We also investigate the merits of calibration of PCR-GLOBWB for each meteorological forcing dataset individually.

This paper is organized as follows. Section 2 presents an overview of the different meteorological datasets and river basins used herein. This is followed by a short description of PCR-GLOBWB and its most important calibration parameters, and a brief explanation of Latin Hypercube Sampling (LHS) used to quantify parameter uncertainty. In Section 3 we report some of the main findings of our study and present the simulated hydrograph uncertainty ranges for each different river basin and forcing dataset. Here, we are especially concerned with a comparison of the simulated discharge dynamics with their observed counterparts, and investigate whether the simulation (prediction) uncertainty of PCR-GLOBWB decreases by down sampling of the original behavioral parameter sets. Section 4 summarizes our main conclusions and provides an outlook for future work.

## 2. Data and methods

### 2.1. Meteorological forcing

Three different meteorological forcing datasets are used in this paper. This includes: (1) a combination of the dataset of the Climate Research Unit of the University of East Anglia

(CRU – Mitchell and Jones, 2005) and the European Centre for Medium-Range Weather Forecasts (ECMWF) re-analysis dataset ERA-40 (Uppala et al., 2005), (2) the ECMWF re-analysis dataset ERA-Interim (Dee and Uppala, 2009) and (3) the Climate Forecast System Reanalysis of the National Center for Environmental Prediction (NCEP CFSR – Saha et al., 2010). For practical purposes, we restrict our attention to the period from 1991 to 2000 for which we have available a continuous record of all three forcing time series and corresponding discharge observations. A traditional split sampling test was used to divide the dataset in a calibration (1991–1995) and evaluation period (1996–2000).

### 2.1.1. CRU monthly observations downscaled to daily values with ERA-40 reanalysis

The first forcing dataset is made up of monthly data from CRU TS2.1 (Mitchell and Jones, 2005; New et al., 2000). This dataset has a spatial resolution of 0.5°, and interpolates observed meteorological variables from constituent meteorological stations. The PCR-GLOWB model requires daily values of temperature, rainfall, and potential evaporation, which necessitates temporal downscaling. To this end we used the ERA-40 reanalysis dataset (Van Beek, 2008; Sperna Weiland et al., 2010), a product of the ECMWF with a horizontal resolution of approximately 125 km and temporal resolution of 6 h (Uppala et al., 2005). Reference potential evaporation was calculated using the Penman–Monteith equation (Monteith, 1965). To supplement the CRU TS 2.1 time series, net incoming shortwave radiation was computed according to the UN Food and Agriculture Organization (FAO) guidelines (Allen et al., 1996) using cloud cover data from CRU TS 2.1 and the climatology of potential shortwave radiation with wind speed data from the long term average monthly climatology, CRU CLM 1.0 (New et al., 1999). The resulting daily forcing dataset follows the diurnal variability of the ERA-40 reanalysis product and complies with the monthly CRU TS2.1 statistics from 1958 to 2001. We purposely combined the CRU datasets with the ERA-40 reanalysis to remove, as much as possible, common biases in precipitation amounts in reanalysis products (Troccoli and Källberg, 2004; Bosilovich et al., 2007). The resulting dataset has been used in previous applications of PCR-GLOWB (Sperna Weiland et al., 2010; Wada et al., 2010; Candogan Yossef et al., 2011; Van Beek et al., 2011).

### 2.1.2. ERA-Interim reanalysis

The second forcing dataset used herein constitutes the ERA-Interim reanalysis (Dee and Uppala, 2009). This dataset supersedes the ERA-40 reanalysis, and includes several improvements to the numerical weather prediction system. The horizontal resolution has been increased from T159 to T255, the model physics have been improved, radiance information is used for bias-correction and better data sources are utilized for wave height, radiance, and ozone profiles. Nevertheless, a strong correlation exists between the ERA-40 and ERA-Interim reanalysis datasets. Not only is the ERA-Interim system an evolution of the existing ERA-40 system, up to 2001 the boundary forcing of the ERA-Interim system has been taken from the ERA-40 system.

For the present analysis the ERA-Interim time-series of precipitation, temperature and data for the Penman–Monteith equation have been extracted from the ERA-Interim runs at 12:00 AM. The meteorological fields have been re-gridded to a regular 0.5° to satisfy the resolution of PCR-GLOWB used herein. The ERA-Interim reanalysis product is considered a continuation of the CRU/ERA-40 dataset for upcoming retrospective studies with PCR-GLOWB.

### 2.1.3. CFSR reanalysis

The third and last forcing dataset is the CFSR reanalysis product which is developed as part of the Climate Forecast System (Saha et al., 2010). The CFSR dataset has only become available in 2010

and supersedes the previous NCEP reanalysis datasets which have been used widely in many studies. At this stage the CFSR dataset spans the period of 1979 to present. The data has a spatial resolution of approximately 0.25° around the equator to 0.5° beyond the tropics (Higgins et al., 2010). In this study, 6-hourly estimates of precipitation, temperature, radiation, air pressure and windspeed were averaged to a daily time-step for the period 1991–2000. These daily values were subsequently interpolated to a regular 0.5° grid and used in PCR-GLOWB. Although previous studies have shown that Penman–Monteith potential evaporation estimates from CFSR diverge somewhat from their CRU derived counterparts (Sperna Weiland et al., 2011a), we nevertheless use this equation for consistency with the other forcing datasets.

## 2.2. Observed discharge data

This study focuses on five of the largest river basins of the world, including the Amazon, Mackenzie, Murray, Mekong and Rhine watersheds. These basins cover different continents and a wide range of climatic conditions, and differ in total catchment area and degree of regulation. We purposely selected these basins because of the availability of an extensive record of discharge observations at the catchment outlet for the period of 1991–2000 encapsulated by the three forcing datasets. The monthly streamflow observations were derived from the Global Runoff Data Center (GRDC, 2007) and the Mekong River Commission. These streamflow time-series were directly used without corrections for river regulations which may have affected the model parameterization. Fig. 1 provides a geographical overview of the different river systems along with their most important basin characteristics and the PCR-GLOWB run times. The degree of regulation is taken from Nilsson et al. (2005) and is quantified by the river's storage capacity as percentage of total discharge.

## 2.3. PCR-GLOWB: the global hydrological model

The PCR-GLOWB model used herein is a distributed hydrological model that is designed to solve the global water balance at a spatial resolution of 0.5° (see Fig. 2; Van Beek and Bierkens, 2009). We now provide a short summary of the most important components of the model. A detailed description of PCR-GLOWB can be found in Van Beek et al. (2011) and Sperna Weiland et al. (2010), and is beyond the scope of the present paper.

Each model cell of PCR-GLOWB consists of two vertical soil layers and one underlying groundwater reservoir. The two soil layers are used to simulate soil water flow and moisture dynamics of the vadose zone. Precipitation (forcing condition) in each grid cell is divided into rain or snow depending on the actual air temperature. Throughfall is computed from the sub-grid parameterization of the fraction of short and tall vegetation. The remaining water is intercepted by the vegetation, and eventually released back to the atmosphere. Rain water that falls on the soil surface either infiltrates or runs-off immediately. Evapotranspiration is calculated from potential evaporation (forcing condition) and soil moisture status of the upper part of the vadose zone. Soil moisture that is not lost to evaporation or taken up by plant roots for transpiration is either immediately transported to the river system by interflow or percolates to the groundwater store, possibly ending up as base flow. Upward flow of water in the vadose zone is physically possible, but numerical results demonstrate that these fluxes are often negligibly small. PCR-GLOWB computes total runoff 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. Attenuation of discharge by lakes



Catchment	Area (km <sup>2</sup> )	$Q_{avg}$ (m <sup>3</sup> /s)	Gauge	Degree of regulation (%)	Run time – 1 year (seconds)	Average run time – (minutes)
Amazon	7.050.000	190.000	Obidos	3	17	13
Mackenzie	1.805.000	10.700	Norman Wells	12	15	11
Mekong	2.981.100	12.743	Mukdahan	3	10	9
Murray	1.061.500	767	Wakool Junction	67	9	7
Rhine	170.000	2.200	Rees	5	8	6

**Fig. 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, location of gauge, degree of river regulation (Nilsson et al., 2005) and run time of PCR-GLOBWB for a yearly simulation and the full simulation with one parameter combination derived from Latin Hypercube sampling.

and reservoirs was accounted for by means of the weir formula which relates storage to outflow (Van Beek and Bierkens, 2009). No reservoir management was considered.

Thus far, we have not considered the effects of water use. We therefore post-process the PCR-GLOBWB simulated streamflow values by subtracting monthly water demands derived from each individual parameter combination using the model of Wada et al. (2010, 2011). This model calculates global water demands at a spatial resolution of 0.5° using information about industrial, domestic and agricultural water use. The amount of irrigation subtracted from the surface water depends on PCR-GLOBWB simulated soil moisture which influences the ratio between actual and potential evaporation. Irrigation water is used to maintain optimal growing conditions for crops 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.

#### 2.4. Parameter uncertainty

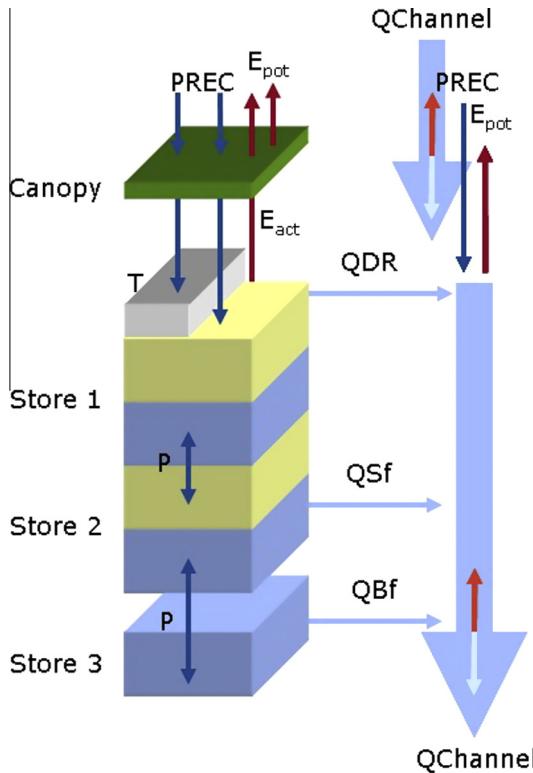
##### 2.4.1. PCR-GLOBWB: calibration parameters

The PCR-GLOBWB model contains a large number of parameters whose values are difficult to observe directly in the field, or at the scale of interest (see Table 1). Although the model was designed to be parameterized from existing global datasets, it remains difficult to estimate all parameters *a priori*. Indeed, it is to be expected that the predictive capability of PCR-GLOBWB will be enhanced if at least some of the parameters are estimated by calibration against discharge observations.

**2.4.1.1. Model parameters.** Table 1 lists all parameters of PCR-GLOBWB. The table contains a total of 29 free model parameters. For some parameters we explicitly distinguish between shallow, (1) and deep, (2) soil compartments, and short, "(s)" and tall, "(t)" vegetation types. The table does not include parameters or variables used for the dimensioning of the model compartments and river network. These values can be directly obtained from topographical datasets, as for example the global drainage network DDM-30 (Döll and Lehner, 2002) and the Global Lake and Wetland dataset (GLWD – Lehner and Döll, 2004).

Based on expert judgment a total of fifteen parameters were selected for the sensitivity analysis. This selection was made, in large part, based on the degree of conceptualization of each individual parameter, and/or their importance in the simulated water balance with PCR-GLOBWB. The following parameter groups have been excluded from the sensitivity analysis: (1) vegetation and root fraction values, (2) the snowfall correction factor and the varying water holding capacity of the snow cover, and (3) the fractional area where percolation to groundwater store is impeded. With respect to (1) values were directly obtained from global vegetation datasets (GLCC 2 – Olson, 1994) and linked to the dimensions of the soil layers (model compartments); (2) omission restricts the degrees of freedom in snow modeling, and (3) values were derived from the observed occurrence of permafrost and soils with impeded percolation (Brown et al., 2000). What follows is a detailed description of the fifteen selected parameters.

The first two model parameters, e.g. the global distributed values of the groundwater reservoir coefficient and storativity directly



**Fig. 2.** Model concept of PCR-GLOBWB (van Beek and Bierkens, 2009). The left-hand side represents the vertical structure for the soil hydrology representing the canopy, soil column (stores 1 and 2), and the groundwater reservoir (store 3). Precipitation (PREC) falls as rain if air temperature ( $T$ ) is above  $0^{\circ}\text{C}$  and as snow otherwise. Snow accumulates on the surface, and melt is temperature controlled. Potential evapotranspiration ( $E_{\text{pot}}$ ) is broken down into canopy transpiration and bare soil evaporation, which are reduced to an actual rate ( $E_{\text{act}}$ ) on the basis of the moisture content of the soil. Vertical transport in the soil column arises from percolation or capillary rise ( $P$ ). Drainage from the soil column to the river network occurs via direct runoff, interflow or subsurface stormflow, and base flow (QDR, Qsf, and Qbf, respectively). Drainage accumulates as discharge (QChannel) along the drainage network and is subject to a direct gain or loss depending on the precipitation and potential evaporation acting on the freshwater surface.

determine the outflow from the linear groundwater reservoir. This outflow,  $Q_{\text{bf}}$  [ $\text{L}^3 \text{T}^{-1}$ ] is defined by a linear relation between the actual modeled storage,  $S_3$  [ $\text{L}^3$ ] and the reservoir coefficient,  $J$  [ $\text{T}^{-1}$ ]:

$$Q_{\text{bf}} = S_3 J. \quad (1)$$

The reservoir coefficient is calculated using the Kraijenhoff van de Leur equation (Kraijenhoff van de Leur, 1958; Sutanudjaja et al., 2011)

$$J = \frac{\pi^2 k_d}{4SP^2}, \quad (2)$$

where  $k_d$  denotes the aquifer transmissivity [ $\text{L}^2 \text{T}^{-1}$ ],  $S$  [–] is the aquifer porosity or storativity [ $\text{L} \text{L}^{-1}$ ], and  $P$  represents the length of the average drainage path [L].

Aquifer transmissivity (or soil permeability) varies spatially and cell specific values are derived from the global lithological maps of Dürr et al. (2005; illustrated in Fig. 3a) and transmissivity estimates of Gleeson et al. (2011; illustrated in Fig. 3b). Aquifer porosity largely originates from secondary porosity and is linked to lithology as well. Note that larger porosity values have been assigned to silicic igneous rocks in tropical regions to account for deep weathering (Gleeson et al., 2014). The average length of the drainage path is defined as half the inverse of the drainage density (Dürr et al., 2005). Drainage density was extracted from the VEMAP and Hydro1k datasets (Verdin and Greenlee, 1996) and computed for

each sub-catchment from the stream length and catchment area. Scale effects were accounted for using the climate-dependent information of Gregory (1976) before aggregating to  $0.5^{\circ}$ . The Holridge life zones of Leemans (1990) were used as proxy for climatic variation, which influences drainage density and thus specific yield.

To update the reservoir coefficient, we used spatial lithology information and assumed the drainage density to be constant and varied the values of aquifer transmissivity and porosity by parameter estimation 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 the permeability (PB\_M)). The use of multiplication factors ensures that the different lithology classes are treated independently, so that the original spatial distribution is honored (McMillan et al., 2010).

The loss of soil moisture through evapotranspiration is dependent on the matric suction at which transpiration is reduced with 50% (FC\_50) and the matric suction at field capacity (FC), which also controls drainage. Snow and glacier melt are determined by the degree-day factors (DDF) (Kuchment, 2004; Martinec, 1975; Seibert, 1999), which are specified for areas with short vegetation (DDFsv), tall vegetation (DDFtv), and glaciers (DDFg). Preliminary sensitivity analysis for the selected basins has demonstrated that the simulation results of PRC-GLOBWB are mainly sensitive to changes in DDFsv.

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. We vary MAXFRAC using a single multiplier, which affects the dimensionless shape factor  $b$  [–] that relates the sub-grid distribution to the average soil water storage capacity, which is assumed constant throughout. Adjusting  $b$  in turn changes the partitioning of net precipitation in direct runoff and infiltration, and the fraction of the saturated area given the actual soil water storage. The infiltrated water delays runoff and is susceptible to evapotranspiration. Consequently, adjusting the amount of infiltrated water influences both the magnitude and shape of the hydrograph (Gosling and Arnell, 2011; Sperna Weiland et al., 2011b). Cell-specific values of MAXFRAC are calculated based on the distribution of the maximum rooting depth (Canadell et al., 1996).

Soil water content and moisture flow are strongly controlled by the saturated hydraulic conductivity ( $K_{\text{sat}}$ ), the air-entry value ( $PSI_A$ ) and the pore-size distribution (BCH), which jointly determine soil evaporation, transpiration, percolation, and capillary rise. The equation of Clapp and Hornberger (1978) is used to describe the relationship between the volumetric moisture content and suction, whereas Campbell (1974) is used to describe the hydraulic conductivity as function of soil moisture content.

The unsaturated hydraulic conductivity is calculated according to Eq. (3) which applies to both soil layers, 1 and 2:

$$K_{\text{unsat}}(S_e) = K_{\text{sat}} S_e^{2\beta+3}, \quad (3)$$

where  $\beta$  [–] is defined per soil type,  $K_{\text{sat}}$  [ $\text{LT}^{-1}$ ] signifies the saturated hydraulic conductivity, and  $S_e$  denotes the degree of saturation [ $\text{L}^3 \text{L}^{-3}$ ]:

$$S_e = \frac{(\theta - \theta_r)}{(\theta_s - \theta_r)}, \quad (4)$$

in which  $\theta$  (theta) denotes the average soil water content of layer 1 or 2 and  $\theta_s$  and  $\theta_r$  signify the saturated and residual soil moisture content, respectively.

River discharge is calculated with the kinematic wave approximation of the Saint-Venant Equations, in which channel flow depends on the bed roughness defined by Manning's roughness coefficient, and the cross-section, slope and hydraulic radius of

**Table 1**

PCR-GLOBWB model parameters.

Parameter	Definition	Original value	Unit	Multiplier ranges	Absolute value ranges	Reference
<i>Deep groundwater: lithological units simplified from Dürr et al. (2005), aggregated to 0.5°</i>						
<b>SC_M1</b>	<b>Storage coefficient multiplier</b>	<b>0.23</b>	–	0.47–1.53	0.11–0.36	Morris and Johnson (1967)
<b>SC_M2</b>	"	<b>0.05</b>	–	0.20–2.0	0.01–0.1	and Rasmussen (1963)
<b>SC_M3</b>	"	<b>0.03</b>	–	0.33–2.0	0.01–0.06	
<b>SC_M4</b>	"	<b>0.05</b>	–	0.20–2.0	0.01–0.1	
<b>SC_M5</b>	"	<b>0.04</b>	–	0.50–3.0	0.02–0.12	
<b>SC_M6</b>	"	<b>0.005</b>	–	0.20–10.0	0.001–0.05	
<b>SC_M7</b>	"	<b>0.04</b>	–	0.25–2.25	0.01–0.09	
<b>PB_M1</b>	<b>Permeability multiplier</b>	<b>2.50</b>	<b>m/day</b>	0.1–10	0.25–25	Domenico and Schwartz (1990), Lambe and
<b>PB_M2</b>	"	<b>1.00</b>	<b>m/day</b>	0.1–10	0.1–10.00	Whitman (1969) and
<b>PB_M3</b>	"	<b>0.10</b>	<b>m/day</b>	0.1–10	0.01–1	Rasmussen (1963)
<b>PB_M4</b>	"	<b>1.00</b>	<b>m/day</b>	0.1–10	0.1–10	
<b>PB_M5</b>	"	<b>0.10</b>	<b>m/day</b>	0.1–10	0.01–1	
<b>PB_M6</b>	"	<b>0.01</b>	<b>m/day</b>	0.1–10	0.001–0.1	
<b>PB_M7</b>	"	<b>0.10</b>	<b>m/day</b>	0.1–10	0.01–1	
<i>Soil layers: Digitized Soil Map of the World (FAO, 2003), aggregated at 0.5°</i>						
<b>psi_FC</b>	<b>Matric suction at field capacity</b>	<b>1</b>	<b>m</b>	Uniform	1.0–3.0	Seneviratne et al. (2010)
<b>psi_FC50</b>	<b>Matric suction at which transpiration is halved</b>	<b>3.33</b>	<b>m</b>	Uniform	3.0–15.0	and Dingman (1994)
THETASAT1 <sup>a</sup>	Saturated moisture content	0.371–0.837	$m^3/m^3$			
THETASAT2 <sup>a</sup>	Saturated moisture content	0.364–0.836	$m^3/m^3$			
K51	Saturated hydraulic conductivity	0.038–6.781	$m/day$			
KS2	Saturated hydraulic conductivity	<b>0.054–5.830</b>	<b>m/day</b>	0.2–5.0	~0.001–29.15	Clapp and Hornberger (1978)
psi_A1	Air entry value	<b>0.015–0.841</b>	<b>m</b>	0.2–2	~0.003–1.7	
psi_A2	Air entry value	<b>0.016–0.512</b>	<b>m</b>	0.2–2	~0.003–1.0	
BCH1	Pore size distribution	<b>0.255–11.039</b>	–	0.2–2	~0.05–22	
BCH2	Pore size distribution	<b>0.687–11.302</b>	–	0.2–2	~0.14–22	
P2_IMP	Fractional area with impeded drainage to the groundwater	0.0–1.0	–			FAO (2003) and Brown et al. (1997)
<i>Land cover: GLCC2 (USGS EROS Data Center) aggregated at 0.5°</i>						
SMAX_S	Interception storage	0–0.0009	<i>m</i>			
SMAX_T	Interception storage	0–0.0020	<i>m</i>			GLCC2 – Olson (1994) and Hagemann et al. (1999)
KC_S	Crop factors short vegetation	0.20–1.2	–			
KC_T	Crop factors tall vegetation	0.26–1.4	–			
KC_WATSTACK	Composite crop factor for channels and wetlands or lakes	Cell specific	–			
KC_MIN	Minimum bare soil crop factor	0.2	–			
VEGFRAC	Vegetation cover	0.0–1.0	–			
RFRAC	Root fraction per soil layer	0.0–1.0	–			
<b>MAXFRAC</b>	<b>Dimensionless shape factor (–) defining distribution of soil water storage within the cell (Improved Arno Scheme) based on the ratio of short and tall vegetation types (max and min rooting depths)</b>	<b>Spatially varying at subgrid level</b>	–	0.2–5.0	Cell specific	Canadell et al. (1996) and Hagemann and Gates (2003)
<i>Snow</i>						
<b>DDF_S<sup>b</sup></b>	<b>Degree-day factor snow melt for areas with short vegetation</b>	<b>0.00239</b>	<b>m°C/day</b>	Uniform	0.001–0.006	Martinec (1975) and Kuchment (2004)
DDF_T <sup>b</sup>	Degree-day factor snow melt for areas with tall vegetation	0.0007	$m^{\circ}C/day$	Uniform		
DDFg	Degree-day factor glacier melt	0.006	$m^{\circ}C/day$	Uniform		
TT <sup>c</sup>	Threshold temperature for freezing/thawing	0.0	°C	Uniform		Dingman (1994) and Bergström (1976)
CFR	Refreezing coefficient	0.05	–	Uniform		
LAMBDA_T	Lapse rate	–0.0065	°C/m	Uniform		
SFCF	Snowfall correction factor	1	–	Uniform		
CWH	Water holding capacity snow cover	0.1	–	Uniform		Bergström (1976)
<i>River routing</i>						
<b>MN_M</b>	<b>Multiplier of Manning coefficient</b>	<b>1</b> <i>(manning = 0.04)</i>	–	0.2–5.0	0.1–0.04	Chow et al. (1988)
QBETA	Kinematic wave parameter	0.6	–			

N.B. All parameters included in sensitivity analysis are formatted in Italics and all parameters included in uncertainty analysis are formatted bold.

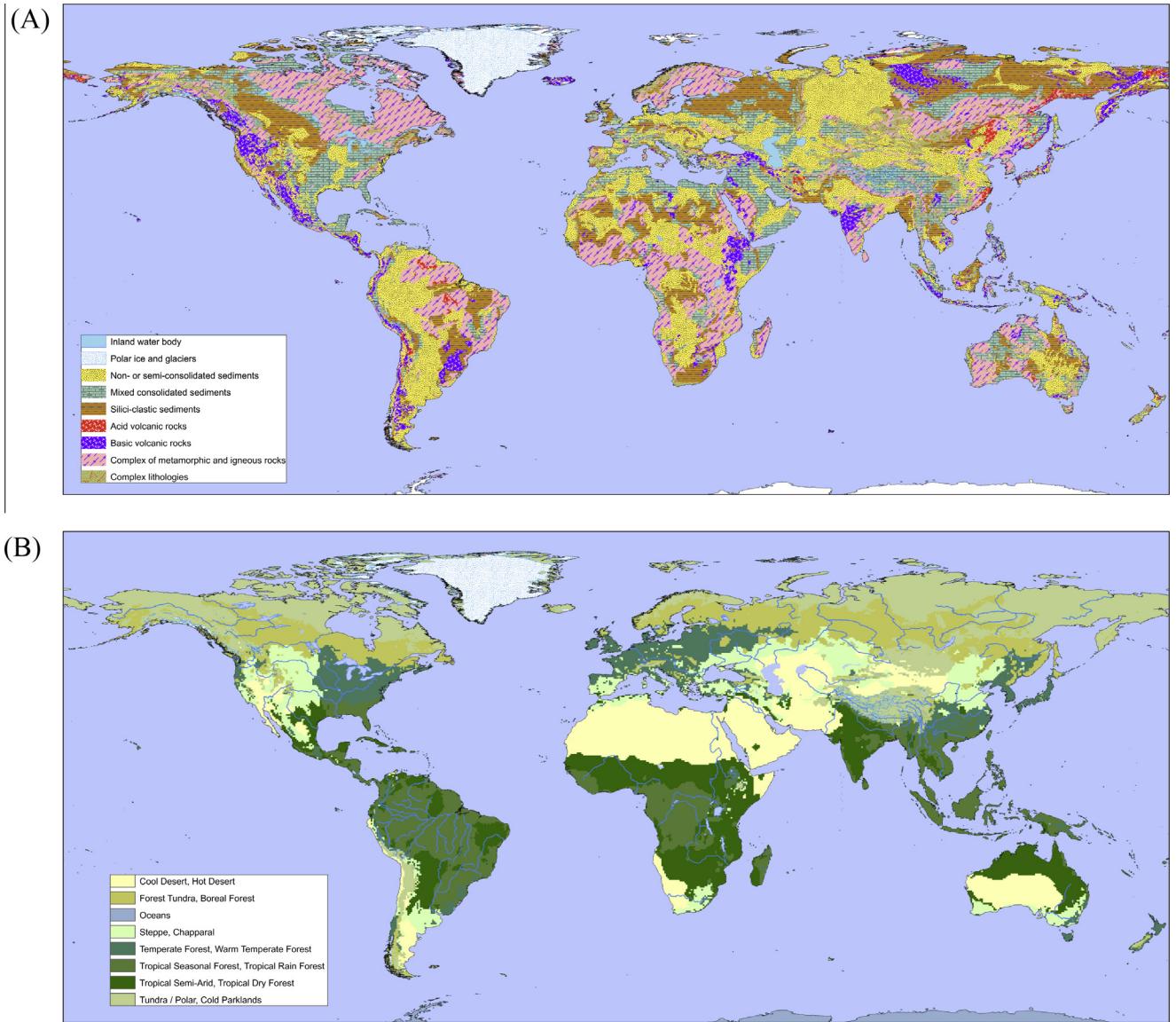
<sup>a</sup> Within the model the soil is divided over a deep (2) and undepth (1) layer. The numbers 1 and 2 in the parameter names refer to the two soil layers.<sup>b</sup> Within the model differentiation in parameterization is made in area with short (s) and tall (t) vegetation.<sup>c</sup> As the default values of the threshold temperature for freezing/thawing is 0 °C, the ±20% range was fixed to –0.2 to 0.2 °C.

the river bed. In all our PCR-GLOBWB calculations reported herein, the default value of 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 spatially varying coefficients. This map is uniformly multiplied with the parameter MN\_M.

**2.4.1.2. Sensitivity analysis.** To identify the most sensitive parameters, a one-at-a-time (OAT) local sensitivity analysis was performed in which each parameter was perturbed individually

using a two-sided interval of 20% from its default value, while keeping the other parameters at their original values. We purposely used two-sided intervals to derive stable parameter sensitivity values. This resulted in a total of 30 different PCR-GLOBWB model runs.

For the two different discharge simulations (+20% and –20%), the long-term annual average,  $\bar{Q}_{-20\%}$  and  $\bar{Q}_{+20\%}$ , were calculated. The sensitivity was now calculated as follows:



**Fig. 3.** World maps depicting spatial variations in (A) lithology (after Dürr et al., 2005), and (B) Holridge life zones (after Leemans, 1990).

$$\Delta Q = 100 \frac{|\bar{Q}_{-20\%} - \bar{Q}_{+20\%}|}{\bar{Q}_{\text{orig}}}, \quad (5)$$

where  $\bar{Q}_{\text{orig}}$  denotes the mean of the original (unperturbed) discharge simulation using default values of the parameters. If a parameter is highly sensitivity, then a small perturbation from its default value would result in a relatively large change in the PCR-GLOBWB simulated discharge time series.

#### 2.4.2. Parameters included in uncertainty analysis

Based on the results of the OAT sensitivity analysis, only those parameters were selected for the full uncertainty analysis with PCR-GLOBWB that demonstrated at least a 1% change in mean annual discharge simulation (averaged over all five basins). These twelve sensitive parameters are listed in bold in Table 1. We also included their upper and lower bounds, which were determined from an extensive literature research. Note that the results of the sensitivity analysis depend somewhat on the chosen parameter sampling ranges.

#### 2.4.3. Latin hypercube sampling

In the absence of detailed prior information, the individual parameters were assumed to be uniformly distributed within the ranges summarized in Table 1. A uniform distribution has shown to work well for exploratory sensitivity and uncertainty analysis (Haan et al., 1998). In a first step, 1000 different parameter combinations were randomly drawn from their uniform prior ranges using LHS (Muleta and Nicklow, 2005). This ensemble was assumed adequate for an initial assessment of parameter uncertainty. A larger sample will probably provide more robust results, yet further increases the already significant computational costs of the analysis. After sampling the parameter space, PCR-GLOBWB is executed for each different parameter combination, resulting in 1000 different time series of simulated discharge dynamics for each of the five different basins and three different forcing datasets. Steady state initial conditions were enforced by running PCR-GLOBWB multiple times for each different parameter combination using the climatology of each specific forcing dataset until differences in basin average discharge and basin average groundwater recharge between two consecutive runs were less than one percent.

## 2.5. Experiment design

### 2.5.1. Uncertainty analysis for calibration period (1991–1995)

The first analysis was conducted for the period 1991–1995, which we hereafter will conveniently refer to as the calibration period. To benchmark our results, and assess the influence of the forcing data uncertainty on simulated discharge dynamics, PCR-GLOBWB was also run individually for all three forcing datasets with default values for the parameters.

The results of the model simulations are evaluated by comparison against observed monthly discharge data. Three simple summary metrics are used to quantify the similarity of the simulated,  $Q_{\text{mod}}$  and measured,  $Q_{\text{obs}}$  discharge time series. This includes the RMSE (Eq. (6a)), the normalized RMSE (NRMSE; Eq. (6b)) and Nash–Sutcliffe (NS; Eq. (7)) coefficient:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (Q_{\text{mod},i} - Q_{\text{obs},i})^2}, \quad (6a)$$

$$\text{NRMSE} = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n (Q_{\text{mod},i} - Q_{\text{obs},i})^2}}{\bar{Q}_{\text{obs}}}, \quad (6b)$$

$$\text{NS} = 1 - \frac{\sum_{i=1}^n (Q_{\text{obs},i} - Q_{\text{mod},i})^2}{\sum_{i=1}^n (Q_{\text{obs},i} - \bar{Q}_{\text{obs}})^2}, \quad (7)$$

where  $n$  denotes the number of months ( $n = 60$ ), and  $\bar{Q}_{\text{obs}}$  signifies the monthly average observed discharge value.

The LHS approach used herein exhaustively samples the parameter space, and should induce a significant simulation spread if appropriate prior ranges are used. This type of random sampling will likely create many parameter combinations that rather poorly simulate the observed hydrology of the basins. To eliminate these non-behavioral solutions from the ensemble of 1000 samples, we draw inspiration from the Generalized Likelihood Uncertainty Esti-

mation (GLUE) method of Beven and coworkers (Beven and Binley, 1992; Beven, 2006), and introduce a subjective threshold to refocus the thrust of the analysis on behavioral samples that at least reasonably mimic the observed discharge dynamics. We therefore eliminate the worst 95% of the 1000 samples and focus our attention on the 50 best parameter combinations with lowest NRMSE. We evaluate the performance of this behavioral ensemble using a monthly average normalized uncertainty range, conveniently also referred to as  $UR [-]$ :

$$UR = \frac{1}{n} \sum_{i=1}^n \frac{(Q_{\max,i} - Q_{\min,i})}{\bar{Q}_{\text{obs},i}}, \quad (8)$$

where  $Q_{\min,i}$  and  $Q_{\max,i}$  denote the minimum and maximum discharge of the  $i$ th month, respectively. We purposely divide in Eq. (8) with the observed average monthly discharge to enable a comparison of the calibration and evaluation results of the different river systems.

To illustrate the advantages of down sampling, we compare  $UR$  of the behavioral solutions against its value derived for the original ensemble of 1000 streamflow simulations. This approach was repeated for each forcing dataset and catchment.

### 2.5.2. Uncertainty analysis for evaluation period (1996–2000)

We evaluate the performance of PCR-GLOBWB for the ensemble of the 50 behavioral parameter solutions for an independent evaluation period (1996–2000). The results of this analysis are summarized in Tables 2–4 and Figs. 6–10. The discharge simulations of the behavioral and best parameter values of the evaluation period are quantified using the NRMSE and NS metrics, and width of the uncertainty ranges and hydrographs.

### 2.5.3. Uncertainty analysis for other hydrological model outputs

The results of this study show that the performance of PCR-GLOBWB to accurately describe the observed monthly discharge

**Table 2**

Normalized Root Mean Square Errors (NRMSE, see Eq. (6b)) of monthly modeled discharge time series for all forcing datasets and river systems considered herein. The various headings summarize the performance of the default model parameterization (*default*), optimal parameter combination (*optimal*), parameter sets with lowest RMSE averaged over all forcings (*global optimal*), mean simulation of the behavioral solutions (*behavioral*), and full LHS ensemble (*mean LHS*) respectively, for the calibration (1991–1995) and evaluation period (1996–2000) respectively. The statistic “*Overall*” lists the average performance for all three different forcing datasets combined.

NRMSE	Calibration					Evaluation			
	<i>default</i>	<i>optimal</i>	<i>global optimal</i>	<i>behavioral</i>	<i>mean LHS</i>	<i>default</i>	<i>optimal</i>	“ <i>global</i> ” <i>optimal</i>	<i>behavioral</i>
<b>Rhine</b>									
CFSR	0.56	0.40	0.62	0.45	0.56	0.60	0.34	0.46	0.32
ERACRU	0.21	0.16	0.23	0.21	0.26	0.23	0.18	0.19	0.26
ERAint	0.21	0.19	0.25	0.21	0.22	0.26	0.22	0.17	0.31
<i>Overall</i>					0.26				
<b>Mackenzie</b>									
CFSR	1.54	1.09	1.17	1.15	1.45	1.10	0.54	0.63	0.59
ERACRU	0.59	0.37	0.87	0.42	0.63	0.52	0.22	0.65	0.29
ERAint	0.43	0.34	0.54	0.36	0.39	0.27	0.19	0.41	0.23
<i>Overall</i>					0.38				
<b>Murray</b>									
CFSR	2.32	0.91	2.72	1.11	2.13	3.28	1.50	3.47	1.71
ERACRU	1.53	0.80	1.20	0.88	1.57	2.36	0.69	1.45	0.71
ERAint	0.93	0.80	1.06	0.86	0.92	0.99	0.73	1.22	0.67
<i>Overall</i>					1.40				
<b>Amazon</b>									
CFSR	0.42	0.35	0.49	0.39	0.41	0.34	0.16	0.25	0.18
ERACRU	0.37	0.26	0.41	0.29	0.34	0.29	0.13	0.20	0.16
ERAint	0.33	0.23	0.37	0.25	0.30	0.24	0.12	0.18	0.13
<i>Overall</i>					0.32				
<b>Mekong</b>									
CFSR	1.16	0.91	1.09	1.05	1.30	0.86	0.63	0.78	0.71
ERACRU	0.58	0.33	0.60	0.41	0.56	0.47	0.22	0.50	0.56
ERAint	0.83	0.61	0.75	0.71	0.94	0.46	0.37	0.41	0.55
<i>Overall</i>					0.60				

**Table 3**

Nash–Sutcliffe (NS) coefficients of PCR-GLOWB simulated discharge time series for all different forcing datasets and river catchments. The various headings summarize the performance of the default model parameterization (*default*), optimal parameter combination (*optimal*), lowest NRMSE average over all forcings (*global optimal*), mean simulation of the behavioral solutions (*behavioral*), and original LHS ensemble (*mean LHS*) for the calibration (1991–1995) and evaluation period (1996–2000) respectively. The statistic “*Overall*” lists the average performance for all three different forcing datasets combined.

NS	Calibration					Evaluation			
	<i>default</i>	<i>optimal</i>	<i>global optimal</i>	<i>behavioral</i>	<i>mean LHS</i>	<i>default</i>	<i>optimal</i>	“ <i>global</i> ” <i>optimal</i>	<i>behavioral</i>
<b>Rhine</b>									
CFSR	−0.80	0.11	−1.09	0.00	−0.79	−1.59	0.19	−0.53	0.25
ERACRU	0.76	0.86	0.69	0.83	0.62	0.62	0.77	0.73	0.53
ERAint	0.75	0.80	0.66	0.83	0.73	0.51	0.66	0.78	0.32
<i>Overall</i>					0.61				
<b>Mackenzie</b>									
CFSR	−4.61	−1.84	−2.28	−2.03	−3.99	−7.88	−2.60	−3.90	−3.30
ERACRU	0.17	0.67	−0.79	0.60	0.05	−0.99	0.41	−4.18	−0.01
ERAint	0.56	0.72	0.29	0.71	0.64	0.46	0.55	−1.12	0.32
<i>Overall</i>					0.66				
<b>Murray</b>									
CFSR	−2.80	0.41	−4.23	0.26	−2.2	−6.60	−0.58	−7.50	−1.01
ERACRU	−0.66	0.55	−0.02	0.53	−0.74	−2.92	0.66	−0.49	0.65
ERAint	0.39	0.55	0.21	0.52	0.41	0.31	0.62	−0.06	0.68
<i>Overall</i>					0.40				
<b>Amazon</b>									
CFSR	−0.98	−0.38	−1.67	−0.54	−0.88	−0.06	0.29	−0.88	0.00
ERACRU	−0.48	0.26	−0.86	0.13	−0.28	0.20	0.52	−0.15	0.29
ERAint	−0.18	0.23	−0.54	0.34	0.01	0.48	0.58	0.07	0.52
<i>Overall</i>					−0.22				
<b>Mekong</b>									
CFSR	−0.41	0.13	−0.25	−0.13	−0.79	0.28	0.56	0.32	0.43
ERACRU	0.64	0.88	0.62	0.83	0.69	0.75	0.94	0.72	0.65
ERAint	0.28	0.61	0.41	0.49	0.07	0.76	0.85	0.81	0.66
<i>Overall</i>					0.62				

**Table 4**

Average spread of the PCR-GLOWB simulated hydrograph uncertainty ranges for the original LHS ensemble (*LHS ensemble*), behavioral solutions (*Behavioral*) for all three forcing datasets individually, combined using the default parameterization (*forcing*) and joint LHS parameter and forcing data uncertainty (*pars + forcing*).

Range	<i>Cal</i>		<i>Eval</i>		<i>Cal</i>		<i>Eval</i>	
	<i>LHS ensemble</i>	<i>Behavioral</i>	<i>Behavioral</i>	<i>Behavioral</i>	<i>LHS ensemble</i>	<i>Behavioral</i>	<i>Behavioral</i>	<i>Behavioral</i>
<b>Rhine</b>								
CFSR	1.29	0.67	1.12	CFSR	0.65	0.34	0.44	
ERACRU	1.11	0.48	0.76	ERACRU	0.53	0.23	0.49	
ERAint	1.23	0.50	0.77	ERAint	0.52	0.19	0.37	
<i>pars + forcing</i>	1.74	0.98	1.12	<i>pars + forcing</i>	0.84	0.58	0.78	
<i>forcing</i>	0.56			<i>Forcing</i>	0.26			
<b>Mackenzie</b>								
CFSR	1.58	0.61	0.43	CFSR	1.22	0.74	1.65	
ERACRU	1.03	0.40	0.37	ERACRU	0.69	0.32	0.86	
ERAint	1.32	0.33	0.36	ERAint	1.22	0.78	1.13	
<i>pars + forcing</i>	3.03	1.51	1.11	<i>pars + forcing</i>	2.24	1.60	1.65	
<i>forcing</i>	1.72			<i>forcing</i>	1.27			
<b>Murray</b>								
CFSR	4.96	1.56	1.91					
ERACRU	4.32	1.12	1.77					
ERAint	3.72	0.86	1.58					
<i>pars + forcing</i>	5.74	1.87	2.87					
<i>forcing</i>	1.46							

emanating from the basin outlet can be improved, yet it remains to be seen whether this is caused by an overall water balance improvement, i.e. storage in snow, soil, groundwater and losses due to evaporation and human influences. From the full set of 1000 realizations the 50 behavioral members have been selected based upon their agreement with discharge observations (5% lowest NRMSE values). Ideally, the behavioral solutions identified in this way, capture all relevant hydrological processes. They should not only fit discharge data well, but simultaneously show a similar increase in performance on simulating the temporal dynamics of

the soil moisture dynamics, snow accumulation/melt, local runoff generation and actual evapotranspiration. We quantify this with the reduction,  $\delta$ , in simulation uncertainty of the behavioral solutions relative to the uncertainty in the full LHS:

$$\delta = 100 \frac{UR_b}{UR_{LHS}}, \quad (9)$$

where  $UR_b$  [–] denotes the simulation uncertainty pertaining to the 50 behavioral solutions and  $UR_{LHS}$  quantifies the spread of the original ensemble of 1000 samples see Eq. (8).

### 2.5.4. Exploring possibilities for regionalization

For all watersheds and forcing datasets, PCR-GLOWBW was executed with the same 1000 different parameter combinations. By comparing their individual NRMSE values (Eq. (6b)), we explore whether the improvements in simulation performance are dependent on catchment and climate characteristics. If these improvements are catchment and/or forcing dataset independent, this would possibly indicate the presence of one global parameter set that works well for the different basins and rainfall scenarios (Widén-Nilson et al., 2007). To assess this independency, we calculate average (linear) correlation coefficients,  $r_{X,Y}$  between the 1000 different NRMSE values for different forcing datasets and catchments:

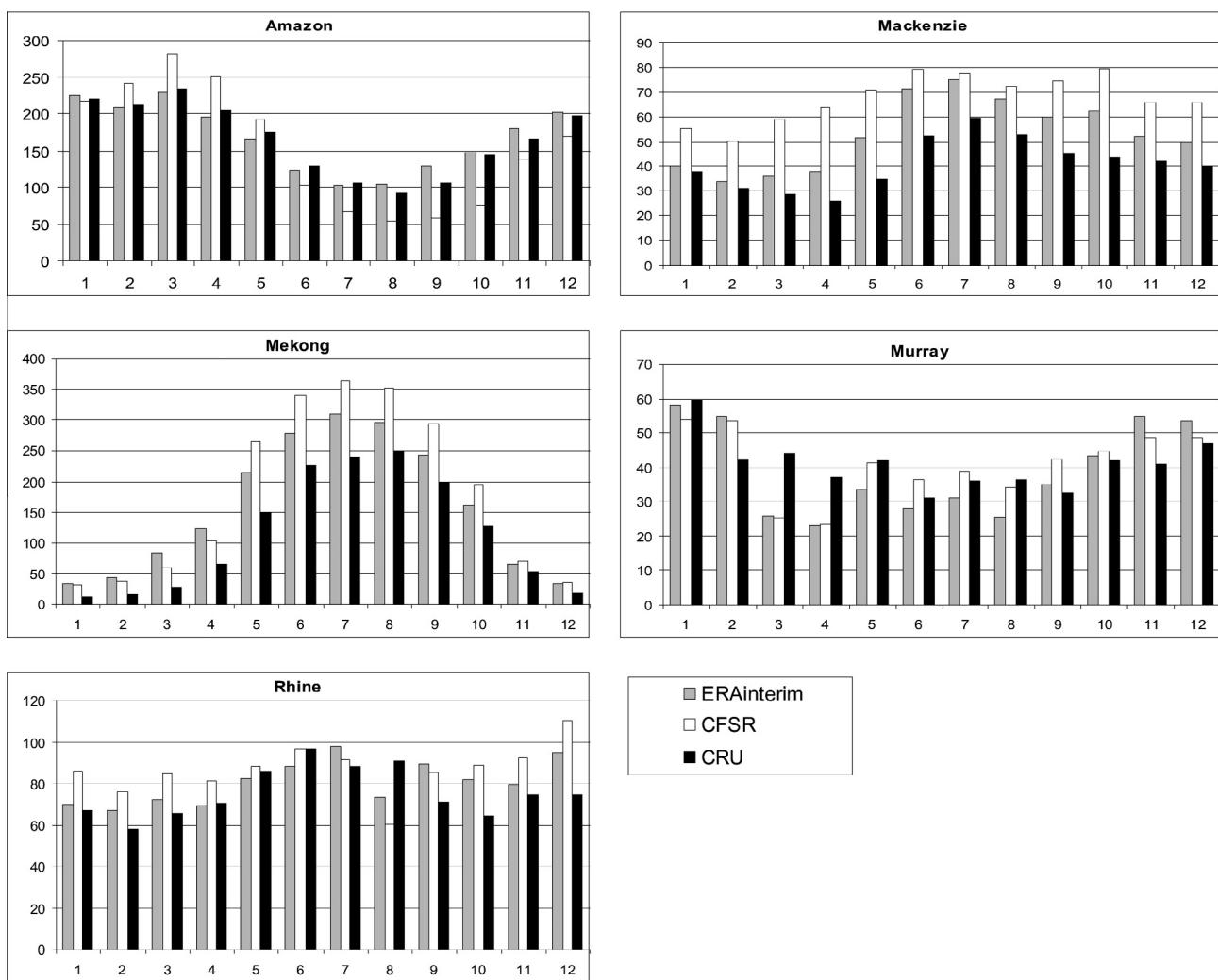
$$r_{X,Y} = \frac{\sum_{j=1}^M (\text{NRMSE}_{X,j} - \bar{\text{NRMSE}}_X)(\text{NRMSE}_{Y,j} - \bar{\text{NRMSE}}_Y)}{\sqrt{\sum_{j=1}^M (\text{NRMSE}_{X,j} - \bar{\text{NRMSE}}_X)^2 \sum_{j=1}^M (\text{NRMSE}_{Y,j} - \bar{\text{NRMSE}}_Y)^2}}, \quad (10)$$

where the subscript X and Y denote different basin and forcing combinations,  $j$  is the actual parameter combination, and  $M$  represents the size of the parameter ensemble. Correlation coefficients close to 1 are preferred and imply that model performance is independent of catchment and climate characteristics.

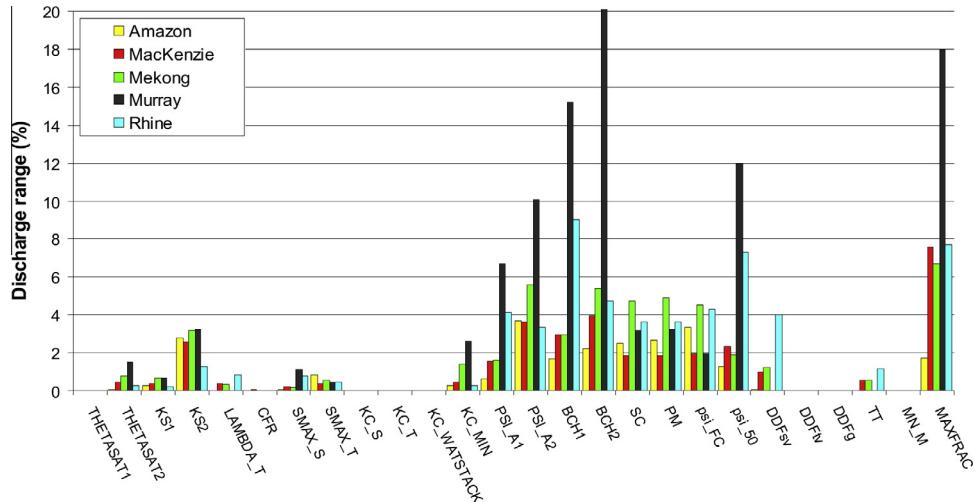
## 3. Results and discussion

### 3.1. Comparison of meteorological forcing datasets

To compare the three forcing datasets, monthly climatologies (long-term basin average monthly precipitation amounts) have been derived for the period 1991–2000. The climatologies are displayed in Fig. 4. In general the climatologies are comparable for the three different forcing datasets, with monthly differences ranging between 5% and 35% where the highest values are found for the CFSR dataset. Patterns are particularly similar for those basins with a strong seasonal variation such as the Amazon and Mekong. The CFSR reanalysis clearly assigns the highest precipitation amounts to the Amazon, Mekong and Mackenzie river basins, particularly during the wet seasons. This over-estimation, also found by Wang et al. (2010), may result from a bias in the updated reanalysis states (Zhang et al., 2012). The differences between the three meteorological datasets appear smallest for the Rhine (with a maximum difference of 30 mm/month) and Murray watersheds (maximum difference of 18 mm/month). This finding is perhaps not surprising, and an immediate effect of the relatively high data quality and measurement density in these basins. Forcing uncertainty decreases with increasing number of sampling points available for optimization and data-assimilation. Relatively low



**Fig. 4.** Long-term monthly basin average precipitation amounts (mm/month) for the selected river basins and three forcing datasets: ERA-Interim (gray), CFSR (white) and the combination of ERA-40 and CRU (black).



**Fig. 5.** Change in mean simulated discharge value (%) as function of a 20% (two-sided intervals) change in the individual parameters.

precipitation amounts are found in the ERA-CRU dataset for the Mackenzie and Mekong river systems. For the Mackenzie this can in part be explained by undercatch in snow measurements (Fiedler and Döll, 2007).

### 3.2. Results parameter sensitivity analysis

The results of the parameter sensitivity analysis are shown in Fig. 5. Quite interestingly, parameter sensitivity of PCR-GLOBWB is comparable for the different basins. High sensitivities are found for the pore size distribution, air entry value and the distribution of soil water storage. Changes in soil parameterization induce the largest changes in the Murray basin (up to 20%), the most arid basin with relatively thin soil layers and parameters that determine the freezing process only affect discharge dynamics of the Mackenzie, Rhine and Mekong where snow accumulation constitutes a major component of the hydrological cycle. Based on the sensitivity analysis with fifteen parameters, only those parameters were selected for the full uncertainty analysis with PCR-GLOBWB that demonstrated at least a 1% variation in the annual mean discharge (averaged over all basins). These twelve parameters are marked in bold in Table 1.

### 3.3. Calibration period – full parameter and forcing uncertainty

Figs. 6–10 present the simulated hydrograph uncertainty ranges for the five different catchments considered in this study. Six different panels are used to summarize our findings. The light-blue region depicts the streamflow uncertainty ranges of the original LHS ensemble, whereas the dark-blue region denotes the results for the behavioral solutions. Each different forcing dataset 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. In the second panel the discharge time-series obtained with the optimal parameter set are displayed (thick lines), the thin lines represent the overall best discharge simulation with lowest average NRMSE of all three forcing datasets (also referred to as the meteorological “global” optimal parameter combination). The third, fourth and fifth panel illustrate the results for the three individual forcing datasets, and finally the bottom panel displays the simulations of a superimposed ensemble that summarizes the joint results for all forcing datasets combined. We now discuss the results of Figs. 6–10.

For most rivers, the CFSR reanalysis forcing dataset consistently overestimates the observed monthly discharge data, irrespective of

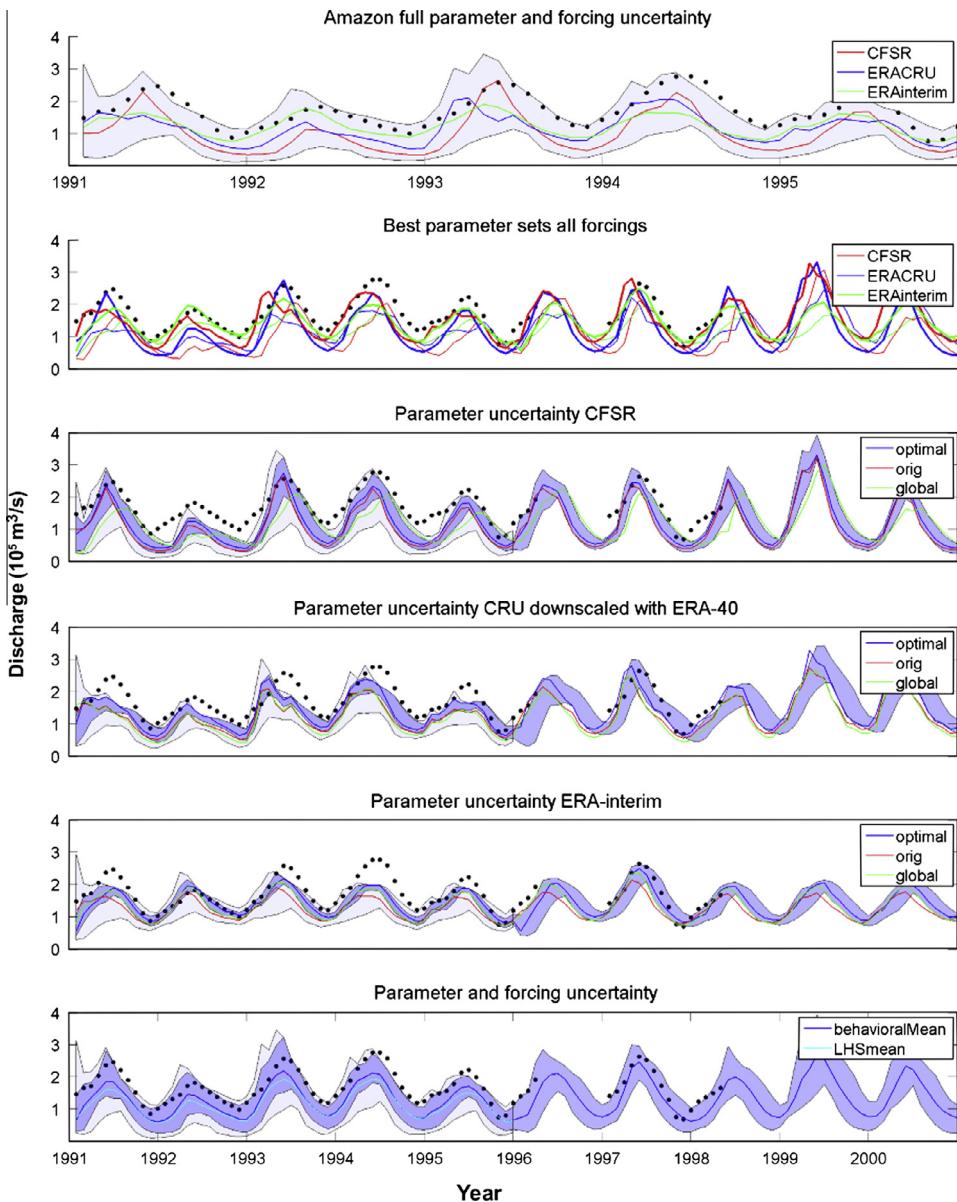
whether the default parameterization of PCR-GLOBWB is used or the parameters are varied using LHS. The simulated hydrograph uncertainty ranges of the full LHS for the CFSR dataset are significantly larger (for all basins at a significance level of 0.01) than those derived for the other two forcing datasets. This difference is due to the larger precipitation amounts of the CFSR dataset (Wang et al., 2010; Zhang et al., 2012), which is apparent for all basins, except the Amazon (Fig. 4). This overestimation is particularly large for the Mackenzie and Mekong.

For all catchments, except the Amazon (where all forcing datasets underestimate the observed flow; consistent with the analysis of Döll et al. (2003) and Widén-Nilsson et al. (2007)), the discharge simulation uncertainty ranges envelop the measured discharge data (see Figs. 6–10a). This is an encouraging result, and demonstrates the ability of PCR-GLOBWB to accurately describe large scale monthly discharge dynamics. The model is able to fit the monthly streamflow data reasonably well, irrespective of the actual forcing data that is being used.

For the Mackenzie, Murray and Mekong (Figs. 7–9) river systems considerable variation exists in the simulated streamflow values among the different forcing datasets. A number of reasons explain this apparent dichotomy. For instance, for the Mackenzie problems with precipitation under-catch due to snowfall (Fiedler and Döll, 2007) are perhaps most important and known to affect the ERA-CRU dataset. On the other hand, the timing and quantity of low flow events is simulated relatively well for the Mackenzie. This coincides with below zero temperatures and temporal storage of precipitation as snow or ice. For the Murray catchment, PCR-GLOBWB severely overestimates the actual monthly discharge when driven with the CFSR forcing dataset (up to four times too high). This overestimation has been reported in earlier GHM studies (Gosling and Arnell, 2011).

Thus far it is clear that forcing data error constitutes the main source of uncertainty in simulated monthly discharges. Apparently, parameter uncertainty represents only a subordinate part of the simulated hydrograph uncertainty. This is an important finding and demonstrates that the apparent mismatch between model simulations and data cannot be resolved by increasing model complexity and resolving sub-grid processes. Instead, it would be most productive to improve the characterization of global rainfall amounts at spatial resolutions of 0.5° and smaller.

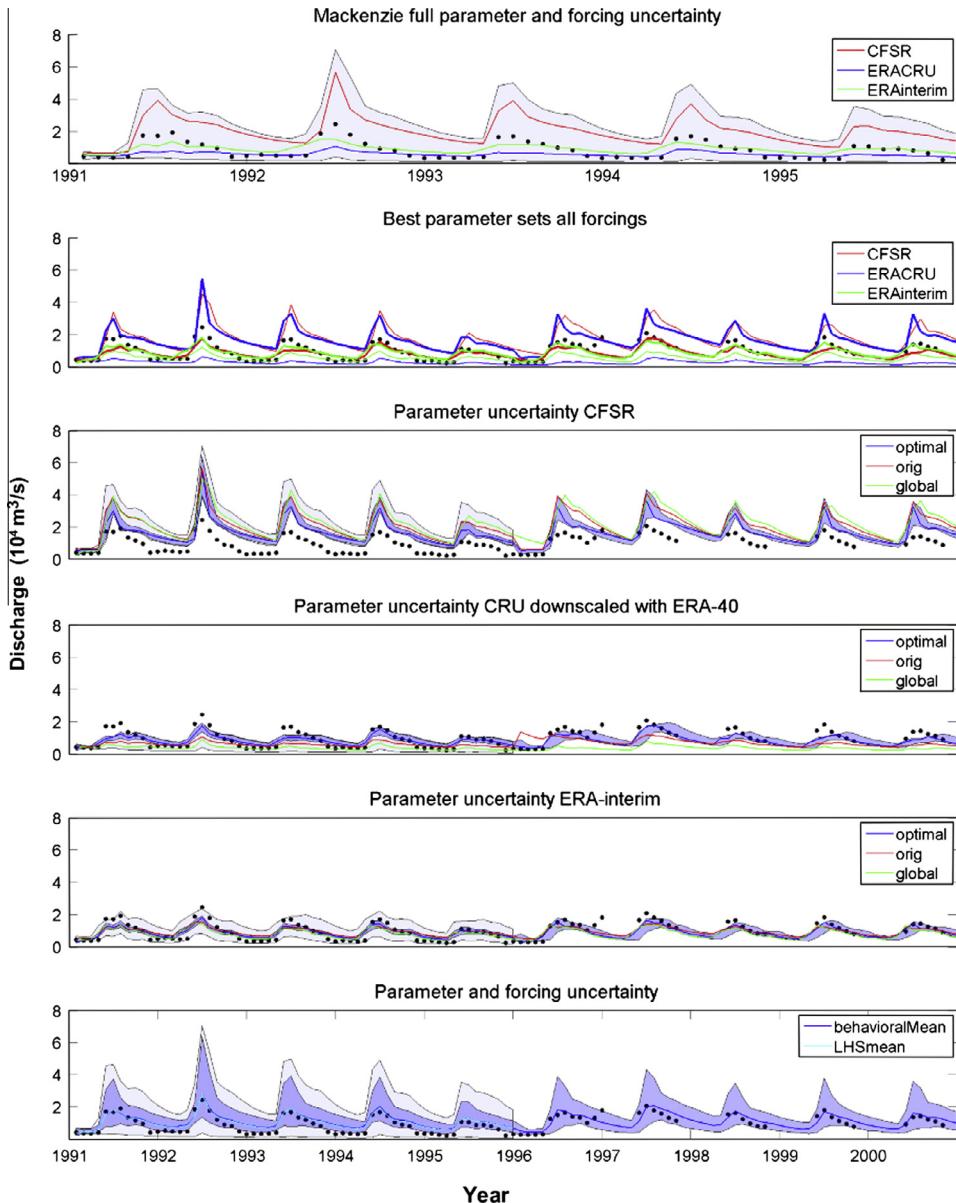
For the Rhine (Fig. 10) river system the different forcing datasets are in excellent agreement, especially the ERA-Interim and ERA-CRU time series for which the best performing parameter



**Fig. 6.** Hydrographs with monthly average discharge for the Amazon river basin modeled with PCR-GLOBWB. In each individual panel, measured discharge data (GRDC) are shown as black dots, the light-blue area denotes the simulation uncertainty (difference between minimum and maximum discharge) obtained from the original LHS ensemble, and the dark-blue area represents the simulated 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 ERA-CRU dataset (dark blue) and ERA-Interim reanalysis data (green) and plots the LHS ensemble spread for all these different forcing datasets combined for the calibration period (1991–1995). Panels 2–6 contain both the calibration and validation period (1991–2000) as indicated on the x-axis of the last panel. In the second panel the discharge time-series obtained with the optimal parameter set are displayed (thick lines), the thin lines represent the discharge simulations obtained with the parameters giving the lowest overall NRMSE for all three forcing datasets for each individual basin (the meteorological “global” optimal parameter combination). The next three panels depict model realizations for the CFSR, ERA-CRU and ERA-Interim datasets. The bottom panel displays the results for the original LHS ensemble and combined behavioral solutions of all three forcing datasets. This signifies the combined effect of parameter and forcing uncertainty. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

sets result in NS values of 0.80 and 0.86. Consequently, the simulated discharge dynamics closely correspond with each other and exhibit relatively narrow uncertainty ranges (Fig. 10c and d) which generally encompass the observed streamflow data. The excellent availability of soil data and detailed information about the river network as well as the dense system of meteorological stations used to derive the forcing datasets enables the model to accurately mimic the observed discharge dynamics. Indeed, Fig. 4 displays a strong agreement between the ERA-40 and ERA-Interim datasets for the Rhine (differences range between 1 and 20 mm/month).

To illuminate the dependency of the optimized parameter values on the forcing dataset used, the second panels in Figs. 6–10 compare the discharge simulations derived from the optimal parameter sets for each of the individual forcing time series (thick lines) with the global best simulation (lowest averaged NRMSE) for all three different forcing datasets (thin lines). The best simulations of each individual forcing data match the observed data much better than the global best simulation. This is further demonstrated in Tables 2 and 3 that list the NRMSE and NS values for the *global* optimal parameterization. The performance of the forcing specific parameter values is, at least during the calibration period, always



**Fig. 7.** Measured (GRDC, black dots) and PCR-GLOBWB simulated streamflow hydrographs for the Mackenzie basin. Please refer to the caption of Fig. 6 for further explanation.

notably higher than the global best parameter set (see also Tables 2 and 3).

#### 3.4. The effect of down sampling

##### 3.4.1. Calibration period

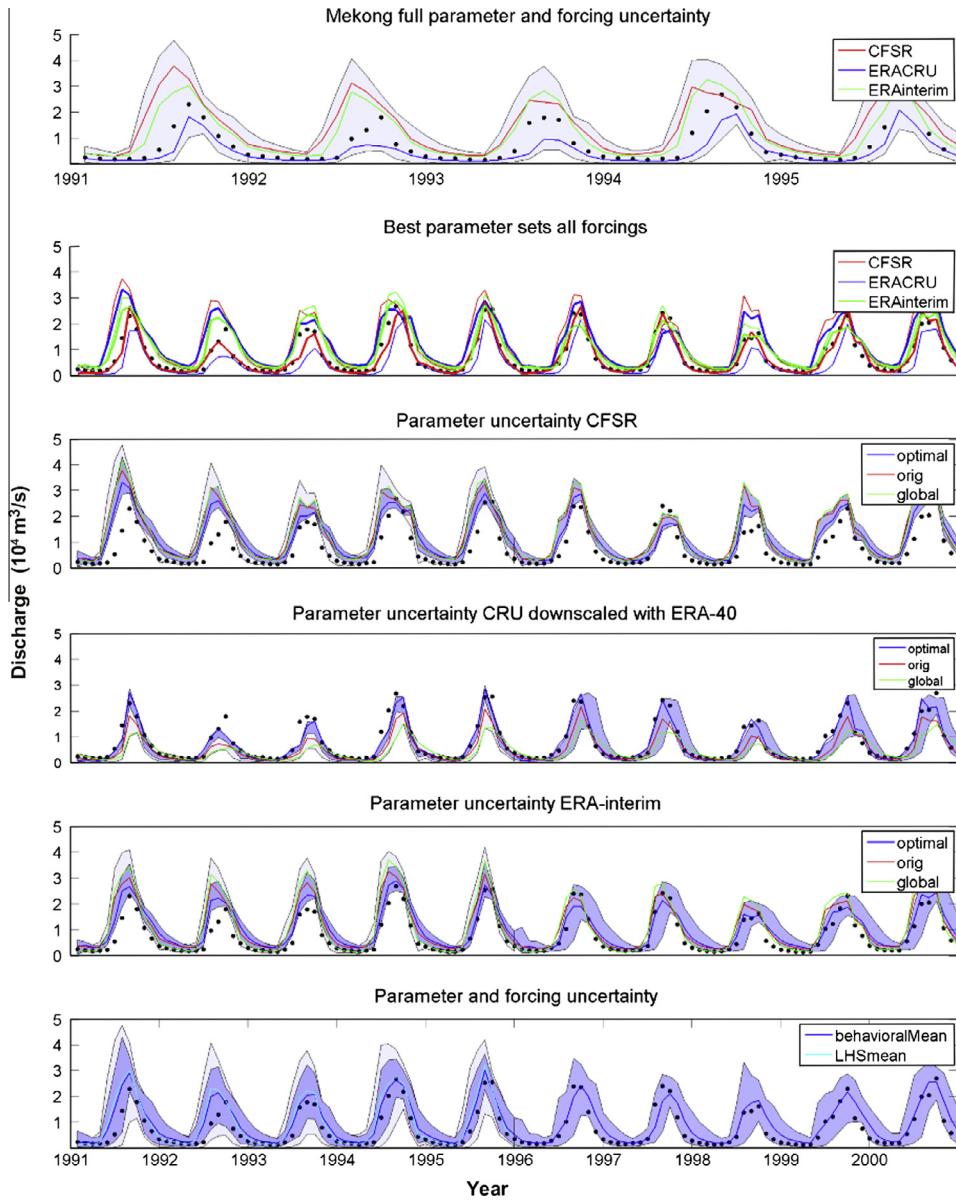
The mean discharge simulation of the LHS ensemble generally outperforms the default parameterization of PCR-GLOBWB during the calibration period (see Tables 2 and 3). This is most evident for those forcing – river basin combinations in which the default parameterization demonstrates significant bias. Yet, for several catchments and forcing datasets the default parameterization of PCR-GLOBWB gives lower NRMSE values than the LHS mean. This is most obvious for (a) the Rhine river forced with the ERA-CRU and ERA-Interim data, (b) the Mackenzie with ERA-CRU data, (c) the Murray with ERA-CRU data, and (d) the Mekong with CFSR and ERA-CRU datasets.

The fifty behavioral solutions with lowest NRMSE also result in higher NS values than the default parameterization of PCR-

GLOBWB. In general, the performance improvement is largest for the forcing data with the largest bias. Yet, this likely results in biased parameter estimates that compensate for systematic errors in the precipitation measurements (Beven, 1996; Widén-Nilsson, 2007). Despite the noticeable improvement in simulation performance of PCR-GLOBWB due to parameter tuning, the model bias remains large, especially for the CFSR dataset. This again reiterates the importance of accurate rainfall estimates.

##### 3.4.2. Evaluation period

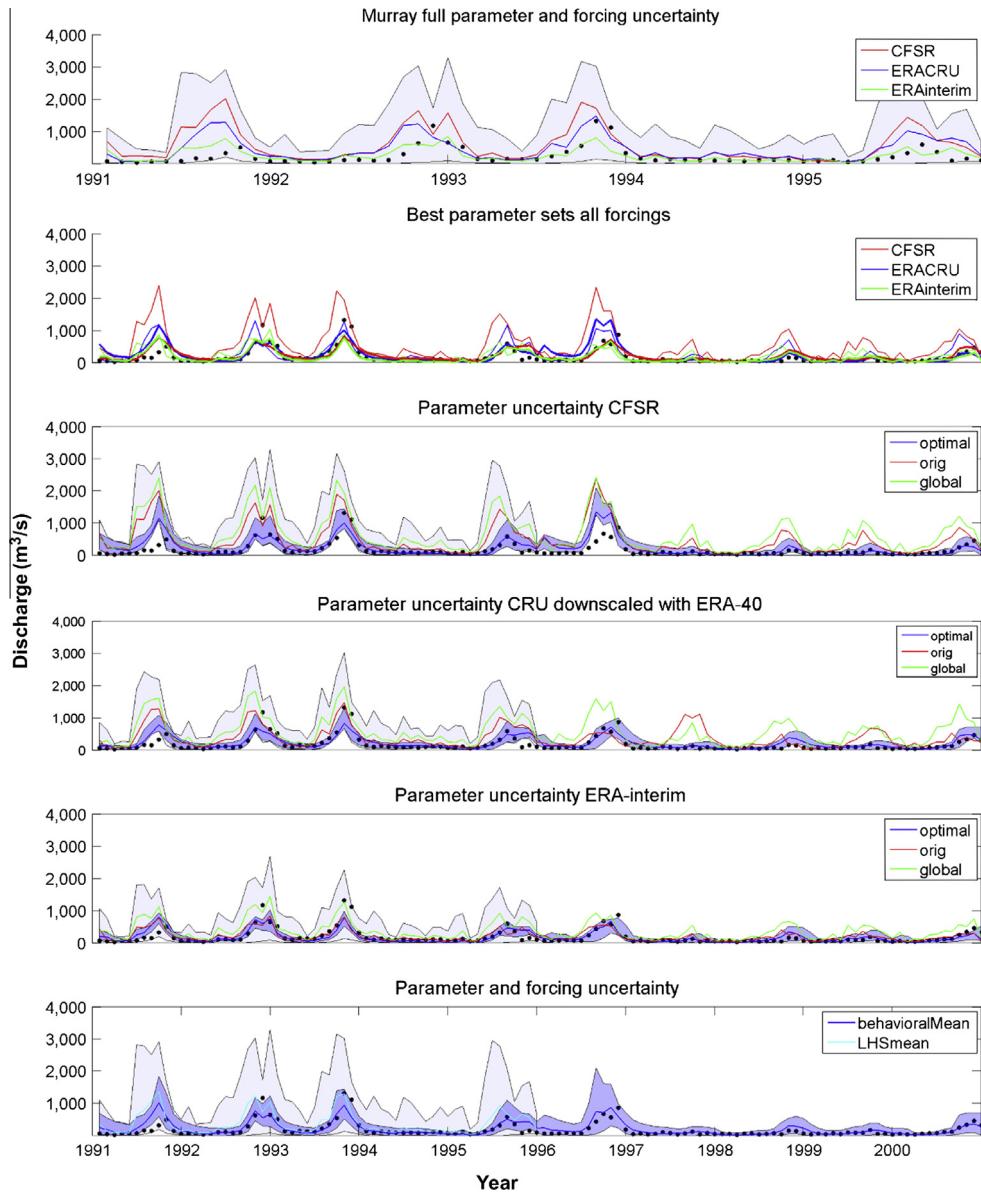
To evaluate the consistency of the calibration results, we use a 5-year evaluation period from 1996 to 2000 to investigate the simulation accuracy of PCR-GLOBWB outside the parameter estimation period. For simplicity, we limit our attention to the behavioral solutions of the calibration data period. From the NS and RMSE values it appears that for most river systems, the mean ensemble discharge simulation of the behavioral solutions and that of the optimal parameter combinations better fit the observed discharge dynamics than the default parameterization, with the



**Fig. 8.** Measured (GRDC, black line) and PCR-GLOBWB simulated hydrographs for the Mekong river basin. Please refer to the caption of Fig. 6 for further explanation.

exception of the Rhine watershed. For this catchment, the default parameterization of PCR-GLOBWB performs quite well when forced with the ERA-CRU dataset (Van Beek et al., 2011). Note that the optimal parameter combinations derived from the calibration period also receive higher NS values than the mean discharge simulation of the behavioral parameter solutions. Exceptions include (1) the Rhine forced with the CFSR dataset and (2) the Murray catchment forced with the ERA-Interim reanalysis. For both the Mackenzie and Amazon river basins the PCR-GLOBWB simulated discharge time series show large deviations from the observed data (with at worst an NS value of –2.6 for the optimal parameter set for the CFSR dataset in the Mackenzie for the evaluation period). For the Amazon there is a consistent offset in the timing of the annual discharge cycle, which we attribute to an overestimation of the actual flow velocities in the floodplains, most likely caused by an underestimation of the roughness coefficient. This model-data mismatch cannot be resolved by simply increasing the prior sampling range of the Manning coefficient. Instead it would be more productive to further refine the dynamics of water storage and outflow in the floodplains.

**Table 4** lists the average spread of the uncertainty ranges derived from the simulations of the LHS ensemble (the shaded blue regions in Figs. 6–10). This range is calculated by averaging the difference between the maximum and minimum simulated discharge values at a monthly time step. We present both the ranges of the original ensemble of 1000 samples and those of the behavioral solutions. One would expect the discharge simulation uncertainty of the behavioral solutions to be smaller than that of the original ensemble. This is because the behavioral solutions are constrained by the observed discharge data. Indeed, the PCR-GLOBWB simulation uncertainty of the behavioral solutions is smaller than that of the original ensemble of 1000 parameter samples. This is true for all catchments and the reduction is significant at a significance level of 0.01 for all basins for all forcing datasets. The uncertainty ranges of the behavioral models during the evaluation period are comparable in size (Mackenzie, Amazon, Murray) or slightly larger than those obtained for the calibration period. These results are consistent with findings from other studies (Krishnamurti et al., 1999; Ajami et al., 2006; Vrugt and Robinson, 2007; Gudmundsson et al., 2011; Haddeland et al., 2011) and favor the



**Fig. 9.** Measured (GRDC, black line) and PCR-GLOWB simulated streamflow hydrographs for the Murray river basin. Please refer to the caption of Fig. 6 for further explanation.

use of an ensemble of solutions rather than a single deterministic simulation for uncertainty quantification and prediction.

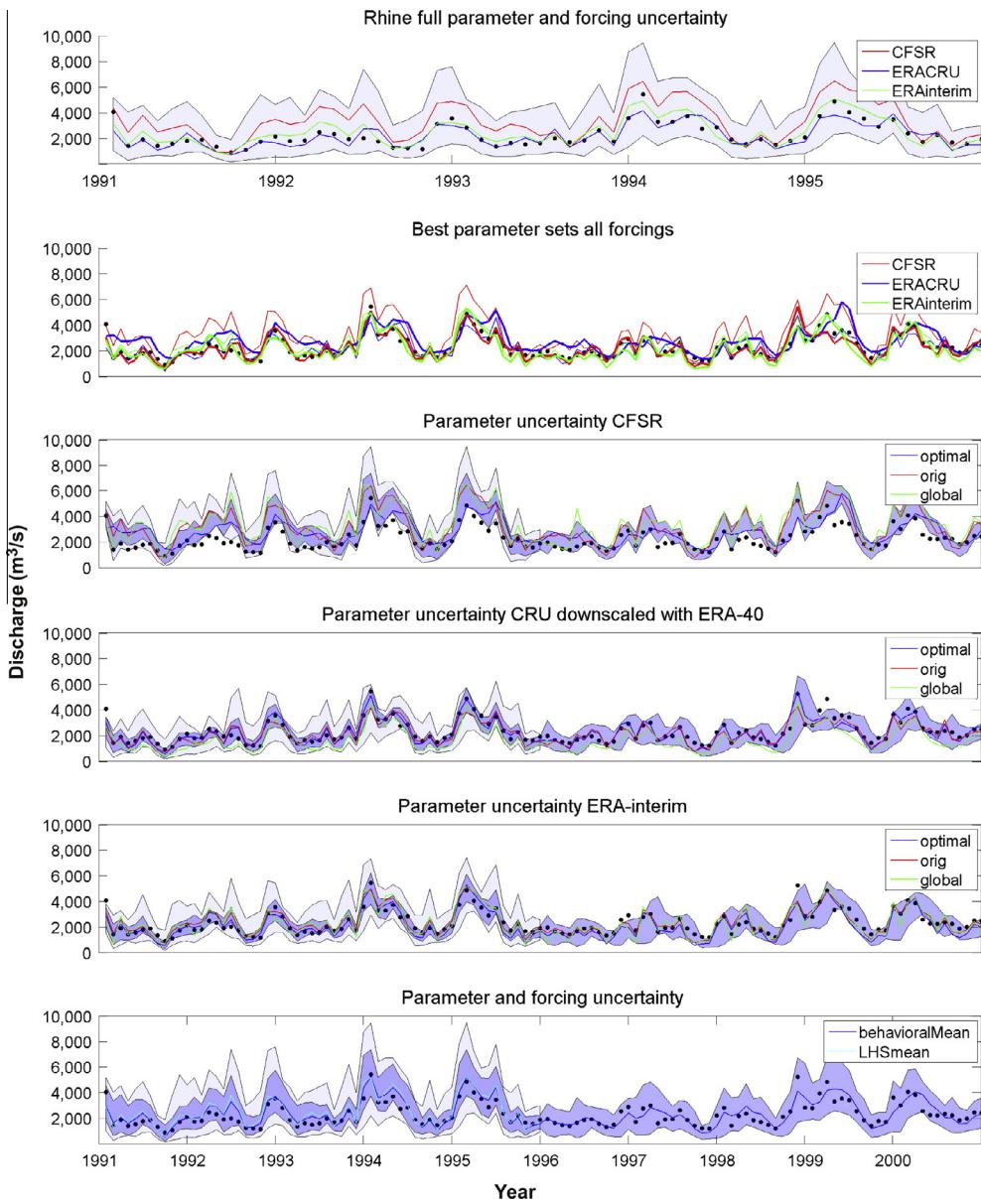
Table 4 also lists the discharge simulation uncertainty ranges derived from the combined behavioral models for all three forcing datasets (*Behavioral; pars + forcing*). For most catchments the simulation uncertainty is at least twice as large as the simulated hydrograph spread corresponding to the behavioral model realizations for the individual forcing datasets. This indicates that bias and differences among the forcing datasets highly impact the overall uncertainty and that differences between discharge simulations obtained from the different forcing datasets remain large, even when the focus is on the behavioral solutions.

An expected, but rather discouraging result is that the optimal parameter combination found for the calibration period of each river system and forcing dataset does not possess the best performance during the evaluation period. This again illuminates the difficulty of finding a single default parameterization that exhibits consistent performance across the different watersheds and forcing datasets. The NRMSE (NS) values of the optimal discharge sim-

ulation for the evaluation period is substantially higher (lower) than the respective statistics for the calibration period for the Rhine, Murray and Mackenzie (see Tables 2 and 3). A more consistent performance during both periods might be found if the length of the calibration period is increased (if possible) to better account for long-term climate variability or if a more advanced calibration strategy is used. This is beyond the scope of the current paper and will be explored in future work.

### 3.5. Reduction of uncertainty in other hydrological model outputs

Unfortunately, at large spatial scales, limited data is available to benchmark the performance of GHMs. Although some advances have been made in our ability to monitor and measure large-scale hydrological fluxes and state variables, such data is not yet readily and widely available. We therefore assume that the reduction of the spread of the ensemble, while moving from the full 1000 member ensemble to the 50 behavioral members, is an indicator of improved performance for other hydrological variables



**Fig. 10.** Measured (GRDC, black line) and PCR-GLOBWB simulated hydrographs for the Rhine river basin. Please refer to the caption of Fig. 6 for further explanation.

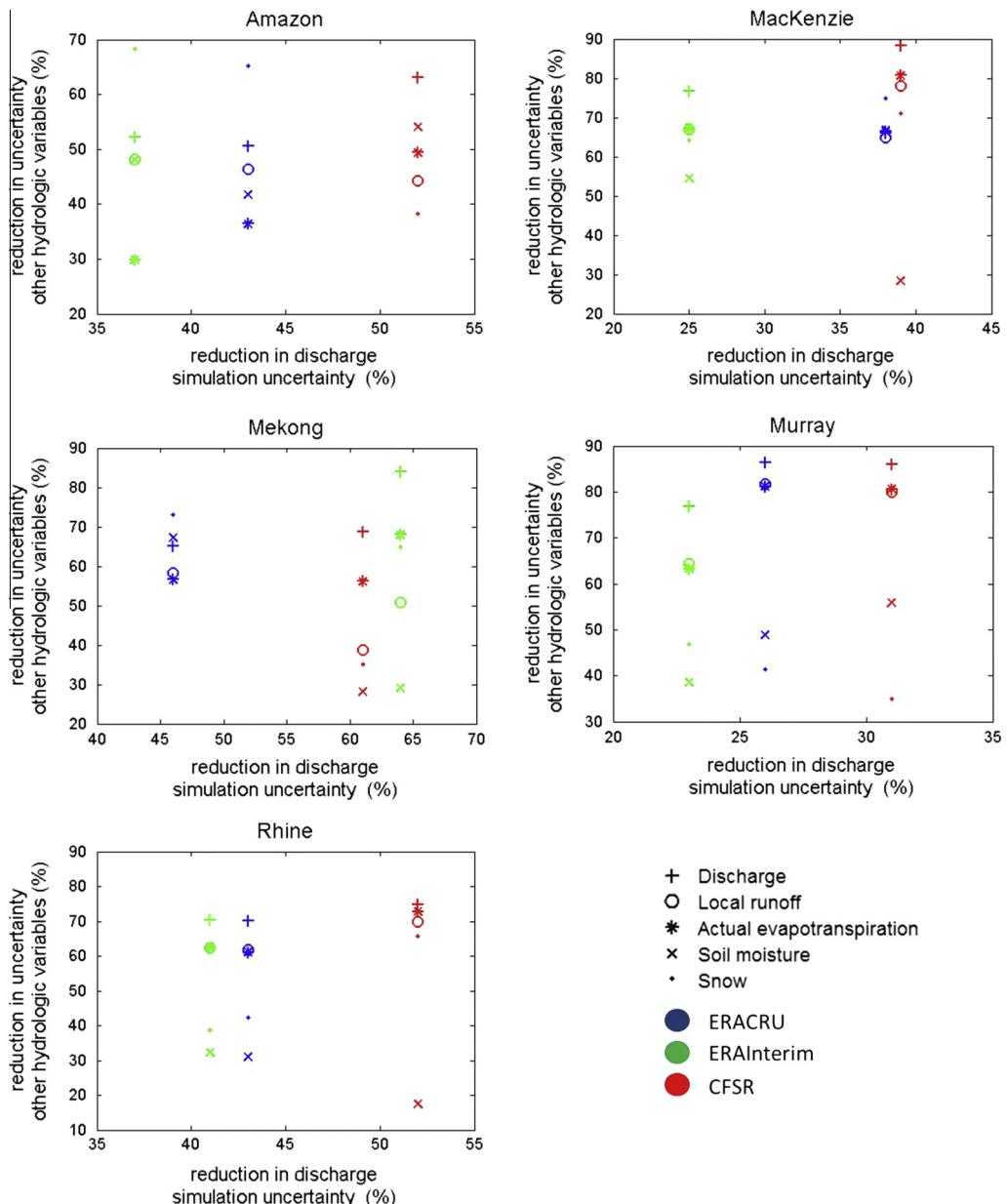
under consideration (e.g. soil moisture, local runoff, actual evapotranspiration and snow). A reduction in uncertainty for all variables would be a desirable finding and demonstrate model robustness and reliability.

Fig. 11 plots the reduction in simulation uncertainty for the behavioral solutions considered herein. Each hydrological variable is coded with a different symbol. The behavioral solutions substantially reduce the simulation uncertainty of the other hydrological variables. The larger the reduction in the width of the simulated discharge dynamics, the smaller the simulation spread of the other hydrological variables. This reduction in simulation uncertainty is most pronounced for the basin average local runoff which determines the river discharge at the basin outlet ( $x$ -axis), and hence this variable exhibits a high correlation with river discharge. The reduction in simulation uncertainty is most notable for the modeled snow dynamics of the Mackenzie watershed and the evapotranspiration amounts of the Murray river basin. These results are somewhat to be expected. The timing of the annual discharge cycle of the Mackenzie strongly depends on the simulated snow dynamics,

whereas the discharge regime of the Murray basin is strongly influenced by evapotranspiration.

The spread in simulated evapotranspiration amounts has been notably reduced for the Rhine river system, which heavily impacts discharge dynamics during the summer season. For this watershed we also expected a notable reduction in the uncertainty of the modeled snow dynamics, as snow fall and melt in the Alps strongly influence the discharge regime. But this reduction is not found. This may be the consequence of the runoff regime of the Rhine that is strongly controlled by antecedent rainfall as well. Although the NS values are relatively high for the Rhine (0.77 for ERACRU and 0.66 for ERAint for the optimal parameter set in the evaluation period), the onset of the spring discharge is wrongly estimated, pointing to structural deficiencies in PCR-GLOBWB.

Although the behavioral solutions reduce the uncertainty of the simulated discharge dynamics of the different river systems, the modeled soil moisture dynamics appears relatively unaffected. The simulation uncertainty remains rather large. This is most obvious for the arid Murray basin where soil moisture constitutes only



**Fig. 11.** Reduction in simulation uncertainty (see Eq. (9)) for the Amazon, Mackenzie, Mekong, Murray, and Rhine river basin derived from down sampling the original LHS ensemble. 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 datasets. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

a relatively minor part of the water balance, irrespective of the forcing dataset being used. Note that the performance of PCR-GLOBWB for the Murray basin is poor, particularly for the CFSR and ERA-CRU forcing dataset. Similar findings have been reported in other GHM studies (Gosling and Arnell, 2011). Moreover, reservoir management and river regulation strongly impact the discharge dynamics of the Murray. These human factors were excluded in the version of PCR-GLOBWB used in this study (Van Beek et al., 2011). The behavioral solutions for this basin are compensating for these structural deficiencies, a subject of much interest in the hydrological literature.

Quite interestingly, the behavioral solutions of the Mackenzie and Murray river basins provide a relatively large spread of the simulated discharge dynamics, but much smaller variation of the other basin average hydrological variables. Apparently, the cell

specific hydrologic fluxes and states of the behavioral ensemble are in much better agreement than the modeled river discharge time series.

The study of Anderton et al. (2002) illustrated that optima can be found in different parts of the parameter space, depending on the hydrological variable of interest (e.g. soil moisture, phreatic surface level or discharge). This makes it rather difficult to find a single parameterization that adequately fits river discharge, soil moisture dynamics, and groundwater tables (Winsemius et al., 2006). Finger et al. (2011) showed that the performance of a distributed hydrological model can be enhanced by including spatial datasets during calibration, for example snow depth or coverage. Unfortunately, data availability severely limits the possibilities of a multi-response calibration for global hydrological models. Fortunately, continued advances in measurement technologies should

**Table 5a**

Upper diagonal values of average correlation coefficients between the 1000 different NRMSE values of PCR-GLOWB simulated monthly discharge values of the five different basins and three forcing datasets. Bold values are used to indicate positive correlations that are significant at a significance level of 0.01. High values are indicative for consistent performance across different catchments and forcing datasets; and favor the selection of a single parameter combination that exhibits good performance for the different catchments, and hence can replace the current default PCR-GLOWB parameterization.

		Amazon			Mackenzie			Mekong			Murray			Rhine		
		CFSR	ERA CRU	ERA int	CFSR	ERA CRU	ERA int	CFSR	ERA CRU	ERA int	CFSR	ERA CRU	ERA int	CFSR	ERA CRU	ERA int
Amazon	CFSR	–	<b>0.88</b>	<b>0.86</b>	–0.51	<b>0.55</b>	<b>0.09</b>	–0.53	<b>0.58</b>	–0.56	–0.52	–0.51	–0.38	–0.50	–0.09	–0.40
	ERACRU	–	<b>0.92</b>	–0.36	<b>0.38</b>	<b>0.10</b>	–0.57	<b>0.38</b>	–0.54	–0.34	–0.35	–0.28	–0.36	–0.15	–0.31	
	ERAint	–	–0.17	<b>0.22</b>	0.03	–0.60	<b>0.25</b>	–0.54	–0.18	–0.20	–0.20	–0.22	–0.19	–0.21	–	–
Mackenzie	CFSR	–	–	–0.96	–0.03	<b>0.43</b>	–0.81	<b>0.56</b>	<b>0.86</b>	<b>0.81</b>	<b>0.58</b>	<b>0.91</b>	0.01	<b>0.70</b>	–	–
	ERACRU	–	–	–	<b>0.25</b>	–0.48	<b>0.85</b>	–0.60	–0.84	–0.79	–0.47	–0.84	<b>0.09</b>	–0.59	–	–
	ERAint	–	–	–	–0.03	<b>0.17</b>	–0.02	0.03	0.07	–0.45	<b>0.15</b>	<b>0.59</b>	<b>0.44</b>	–	–	–
Mekong	CFSR	–	–	–	–	–0.58	<b>0.97</b>	<b>0.49</b>	<b>0.54</b>	<b>0.46</b>	<b>0.53</b>	<b>0.31</b>	<b>0.49</b>	–	–	–
	ERACRU	–	–	–	–	–0.73	–0.93	–0.92	–0.58	–0.81	–0.01	–0.59	–	–	–	–
	ERAint	–	–	–	–	–0.66	<b>0.70</b>	<b>0.57</b>	<b>0.65</b>	<b>0.30</b>	<b>0.59</b>	–	–	–	–	–
Murray	CFSR	–	–	–	–	–	–	<b>0.99</b>	<b>0.75</b>	<b>0.89</b>	<b>0.19</b>	<b>0.76</b>	–	–	–	–
	ERACRU	–	–	–	–	–	–	–	<b>0.80</b>	<b>0.86</b>	<b>0.27</b>	<b>0.78</b>	–	–	–	–
	ERAint	–	–	–	–	–	–	–	–	<b>0.70</b>	<b>0.66</b>	<b>0.86</b>	–	–	–	–
Rhine	CFSR	–	–	–	–	–	–	–	–	–	–	–	<b>0.16</b>	<b>0.83</b>	–	–
	ERACRU	–	–	–	–	–	–	–	–	–	–	–	–	<b>0.64</b>	–	–
	ERAint	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–

open new avenues for calibration and evaluation of large-scale hydrological models.

### 3.6. Possibilities for global parameter estimates or regionalization

In this section we investigate, the dependency of the optimal parameter estimates on forcing data and catchment characteristics. In other words, we test whether it is feasible to locate a single parameter combination that receives consistent performance for multiple different forcing datasets and river basins (Widén-Nilsson et al., 2007). This would be desirable, and opens up possibilities to estimate the parameters of PCR-GLOWB from basic soil and catchment properties. This would reduce the need for calibration and facilitate prediction in ungauged basins. To provide insights into the transferability of the parameter estimates from one catchment to the next, we calculate the correlation coefficients and their significance, between the NRMSE values of the 1000 different samples of the LHS ensemble derived for each different basin and forcing dataset. The results of this analysis are summarized in Table 5a. Overall significant correlations (marked bold) are found, with the CFSR dataset as main exception, here coefficient values appear to be rather low, with the exception of the Murray and Rhine river basins. This again demonstrates that it is rather difficult, if not impossible to find a single realization of the parameter values that receives consistent performance for each of the basins and forcing datasets considered herein. The likelihood of finding such parameterization further deteriorates if different forcing datasets are being used. In a second step we investigated the possibilities for regionalization by calculating the same correlation coefficients of the NRMSE for basins with similar characteristics. We therefore select (1) the Rhine and Danube, two Alpine European catchments in moderate climate zones, and (2) the Mackenzie and Yukon, two arctic catchments in Northern Canada and Alaska. The correlation coefficients for the different forcing datasets are shown in Tables 5b and c. For the Rhine and Danube the correlation coefficients are relatively high and significant at a level of 0.01, especially for the CFSR (0.92) and ERA-Interim (0.94) forcing datasets. For the Mackenzie and Yukon basins, the correlation is somewhat lower, nonetheless, a significant correlation of 0.92 is obtained for the ERA-CRU dataset.

The results presented thus far highlight several important findings. In the first place, it seems impossible to find a single param-

**Table 5b and c**

Equal to Table 5a, but now for catchments with similar characteristics: Rhine and Danube (b) and Mackenzie and Yukon (c).

		Danube		
		CFSR	ERACRU	ERAint
Rhine	CFSR	<b>0.92</b>	<b>0.72</b>	<b>0.80</b>
	ERACRU	<b>0.20</b>	<b>0.69</b>	<b>0.80</b>
	ERAint	<b>0.79</b>	<b>0.92</b>	<b>0.94</b>
		Yukon		
		CFSR	ERACRU	ERAint
Mackenzie	CFSR	<b>0.62</b>	–0.94	<b>0.40</b>
	ERACRU	–0.57	<b>0.92</b>	–0.33
	ERAint	<b>0.13</b>	–0.01	<b>0.31</b>

eterization that works well across the different watersheds and forcing datasets. This necessitates a catchment specific GHM model calibration, which in hindsight is perhaps not a surprising result as the dominant hydrological processes vary considerably between individual catchments. Yet, there seems to be a possibility for regionalization when dividing catchments in groups with similar basin characteristics, as has been shown here for the Danube and Rhine river basins. But even then one should realize that the optimal parameter values derived for each catchment are compensating for model structural errors and input data measurement errors. This is common in virtually all watershed modeling studies and severely diminishes our ability to correlate the parameter values to basic soil and catchment properties. Note that the quality of the model input data varies considerably between different catchments, and hence the parameter compensation effect differs for each watershed. Therefore regionalization methods should properly be tested.

In the second place, our results demonstrate that the chosen forcing dataset strongly determines the optimal parameter values. Each individual forcing dataset has its own optimal parameter estimates. This result is perhaps not surprising given the relatively large differences among the different forcing datasets. Yet it further complicates parameter estimation, given the computational demands of global scale hydrological models such as PCR-GLOWB and the fact that the model will be applied using different operational, historical and climate model datasets (Candogan et al., 2015).

(Yossef et al., 2011; Wada et al., 2010; Sperna Weiland et al., 2010). Moreover, in climate impact studies, in which a model such as PCR-GLOBWB may be forced with meteorological data from an ensemble of GCMs for different time-slices (Vaze et al., 2010; Sperna Weiland et al., 2010) the model should preferably be used with a default parameterization derived from available global datasets (Van Beek et al., 2011). Only if one single forcing dataset is employed, as for example in a flow forecasting system (Candogan Yossef et al., 2011, 2013), it may be feasible to estimate the parameters from discharge and/or other data. In addition the current study used one single hydrological model, focusing on parameter and forcing uncertainty and ignoring hydrological model uncertainty, whereas several studies demonstrated that different models may respond differently to (climate) changes in forcing data (Vansteenkiste et al., 2012; Ficklin and Barnhart, 2014). The current study should be repeated with other GHMs to test whether the statement that forcing uncertainty is larger than parameter uncertainty is in general valid.

Finally, we would like to note that residual-based calibration or sampling approaches as used herein have inherent weaknesses that make it very difficult to detect model structural errors (Gupta et al., 2008; Vrugt and Sadegh, 2013). A diagnostic approach with compelling summary metrics rooted in hydrologic theory will significantly increase our chances to pinpoint reasons for model malfunction.

#### 4. Conclusions

The goal of this study was to explore the possibilities for the calibration of PCR-GLOBWB, a global hydrological model. Our analysis specifically analyzed the influence of parameter and meteorological forcing data uncertainty on simulation uncertainty. From this analysis the following conclusions can be drawn:

- Simple random search of the parameter space led to parameter estimates that exhibit better performance than the default parameterization for nearly all basins and forcing combinations.
- The model did not demonstrate consistent performance during the calibration and evaluation period. The parameter values are likely compensating for structural model deficiencies and biases in the forcing data. Consistency may be improved by implementing a more advanced calibration method or by increasing the length of the calibration period, as it may currently be too short to excite all components of model behavior and properly consider climate variability. Unfortunately, this significantly increases the computational requirements of the analysis. And even then, the resulting optimal parameterization may not be robust under changing environmental conditions.
- The behavioral parameter solutions derived from one watershed exhibit poor performance for other watersheds. A catchment and forcing specific model calibration appears most productive to optimize the predictive abilities of PCR-GLOBWB, although there may be possibilities for regionalization based on catchment similarities.
- From the current analysis we conclude that parameter uncertainty constitutes only a relatively minor part of hydrograph simulation uncertainty. This demonstrates that the apparent dichotomy between model simulations 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 resolutions of 0.5° and smaller. Physically-based hydrological model output along with stream flow observations can help to diagnose where biases in precipitation are largest.

- In light of the limited availability of high quality meteorological, river discharge, surface and subsurface data one might argue that the complexity of PCR-GLOBWB is not warranted by the available discharge observations. The use of a simple lumped, parsimonious, and CPU-efficient hydrologic model allows use of advanced Bayesian sampling approaches that could potentially lead to a better simulation performance. Yet, most of these models simulate only the discharge at the watershed outlet or interior point, and thus cannot be used to simulate other hydrologic variables of great importance in global modeling studies.

#### Acknowledgements

The global discharge time-series 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 EU FP7 GLOWASIS project. Finally, we acknowledge Yoshi Wada for sharing the source code of the water demand model.

#### References

- Ajami, N.K., Duan, Q., Gao, X., Sorooshian, S., 2006. Multmodel combination techniques for analysis of hydrological simulations: application to distributed model intercomparison project results. *J. Hydrometeorol.* 7 (4), 755–768.
- Allen, R.G., Smith, M., Pruitt, W.O., Pereira, L.S., 1996. Modification of the FAO crop coefficient approach. In: Camp, C.R., Sadler, E.J., Yoder, R.E. (Eds.), *Evapotranspiration and Irrigation Scheduling*. Proceedings of the International Conference, November 3–6, San Antonio, TX, pp. 124–132.
- Anderton, S., Latron, L., Gallart, F., 2002. Sensitivity analysis and multi-response, multi-criteria evaluation of a physically based distributed model. *Hydrolog. Process.* 16, 333–353.
- Andréassian, V., Perrin, C., Michel, C., Usart-Sánchez, I., Lavabre, J., 2001. Impact of imperfect rainfall knowledge on the efficiency and the parameters of watershed models. *J. Hydrol.* 250, 206–223.
- Balin, D., Lee, H., Rode, M., 2010. Are rainfall measurement errors likely to greatly impact on distributed complex hydrological modeling. *Water Resour. Res.* <http://dx.doi.org/10.1029/2009WR007848>.
- Bergström, S., 1976. The HBV model. In: Singh, V.P. (Ed.), *Computer Models of Watershed Hydrology*. Water Resources Publications, Colorado, pp. 443–476.
- Beven, K.J., Binley, A.M., 1992. The future of distributed hydrological models: model calibration and uncertainty prediction. *Hydrolog. Process.* 6, 279–298.
- Beven, K., 1996. A discussion of distributed hydrological modelling. In: Abbott, M.B., Refsgaard, J.C. (Eds.), *Distributed Hydrological Modelling*. Kluwer Academic, pp. 255–278.
- Beven, K., 2006. A manifesto for the equifinality thesis. *J. Hydrol.* 320, 18–36. <http://dx.doi.org/10.1016/j.jhydrol.2005.07.007>.
- Beven, K., Cloke, H., 2012. Comment on “Hyperresolution global land surface modeling: meeting a grand challenge for monitoring Earth's terrestrial water” by Eric F. Wood et al. *Water Resour. Res.* 48, W01801. <http://dx.doi.org/10.1029/2011WR010982>.
- Biemans, H., Hutjes, R.W.A., Kabat, P., Strengers, B., Gerten, D., Rost, S., 2009. Effects of precipitation uncertainty on discharge calculations for main river basins. *J. Hydrometeorol.* 10 (4), 1011–1025. <http://dx.doi.org/10.1175/2008JHM1067.1>.
- Bosilovich, M.G., Chen, J., Robertson, F.R., Adler, R.F., 2007. Evaluation of global precipitation in reanalyses. In: 21st Conference on Hydrology, San Antonio, TX, January 2007.
- Brown, J., Hinkel, K.M., Nelson, F.E., 2000. The circumpolar active layer monitoring (CALM) program: research designs and initial results. *Polar Geogr.* 24 (3), 165–258.
- Campbell, G.S., 1974. A simple method for determining unsaturated conductivity from moisture retention data. *Soil Sci.* 117, 311–314.
- Canadell, J., Jackson, R.B., Ehleringer, J.R., Mooney, H.A., Sala, O.E., Schulze, E.-D., 1996. Maximum rooting depth of vegetation types at the global scale. *Oecologia* 108, 583–595.
- Candogan Yossef, N., van Beek, L.P.H., Kwadijk, J.C.J., Bierkens, M.F.P., 2011. Skill assessment of a global hydrological model in reproducing flow extremes. *Hydrolog. Earth Syst. Sci.*
- Candogan Yossef, N., Winsemius, H.C., Weerts, A.H., Van Beek, L.P.H., Bierkens, M.F.P., 2013. Skill of a global seasonal hydrologic forecasting system: relative roles

- of initial conditions and meteorological forcing. *Water Resour. Res.* 49 (8), 4687–4699.
- Chow, V.T., Maidment, D.R., Mays, L.W., 1988. *Applied Hydrology*. McGraw-Hill, New York, 572 pp.
- Clapp, R.B., Hornberger, G.M., 1978. Empirical equations for some soil hydraulic properties. *Water Resour. Res.* 14 (4), 601–604. <http://dx.doi.org/10.1029/WR014004p00601>.
- Dee, D.P., Uppala, S., 2009. Variational bias correction of satellite radiance data in the ERA-Interim reanalysis. *Quart. J. R. Meteorol. Soc.* 135, 1830–1841.
- Dingman, S.L., 1994. *Physical Hydrology*, second ed. Prentice Hall, New York, 575 pp.
- Döll, P., Lehner, B., 2002. Validation of a new global 30-min drainage direction map. *J. Hydrol.* 258 (1–4), 214–231.
- Döll, P., Kaspar, F., Lehner, B., 2003. A global hydrological model for deriving water availability indicators: model tuning and validation. *J. Hydrol.* 270, 105–134. [http://dx.doi.org/10.1016/S0022-1694\(02\)00283-4](http://dx.doi.org/10.1016/S0022-1694(02)00283-4).
- Domenico, P.A., Schwartz, F.W., 1990. *Physical and Chemical Hydrogeology*. John Wiley & Sons, New York.
- Dürr, H.H., Meybeck, M., Dürr, S.H., 2005. Lithologic composition of the Earth's continental surfaces derived from a new digital map emphasizing riverine material transfer. *Global Biogeochem. Cycles* 19, GB4S10. <http://dx.doi.org/10.1029/2005GB002515>.
- Duan, Q., Sorooshian, S., Gupta, V., 1992. Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resour. Res.* 28 (4), 1015–1031.
- Duan, Q., Schaake, J., Andreassian, V., Franks, S., Goteti, G., Gupta, H.V., Gusev, Y.M., Habets, F., Hall, A., Hay, L., Hogue, T., Huang, M., Leavesley, G., Liang, X., Nasonova, O.N., Noilhan, J., Oudin, L., Sorooshian, S., Wagener, T., Wood, E.F., 2006. Model Parameter Estimation Experiment (MOPEX): an overview of science strategy and major results from the second and third workshops. *J. Hydrol.* 320 (1–2), 3–17. <http://dx.doi.org/10.1016/j.jhydrol.2005.07.031>.
- Fekete, B.M., Vörösmarty, C.J., Roads, J.O., Willmott, C.J., 2004. Uncertainties in precipitation and their impacts on runoff estimates. *J. Climate* 17, 294–304.
- Fekete, B.M., Vörösmarty, C.J., Grabs, W., 2002. High-resolution fields of global runoff combining observed river discharge and simulated water balances. *Global Biogeochem. Cycles* 16, 1042. <http://dx.doi.org/10.1029/1999GB001254>.
- Ficklin, D.L., Barnhart, B.L., 2014. SWAT hydrologic model parameter uncertainty and its implications for hydroclimatic projections in snowmelt-dependent watersheds. *J. Hydrol.* 519, 2081–2090. <http://dx.doi.org/10.1016/j.jhydrol.2014.09.082>.
- Fiedler, K., Döll, P., 2007. Global modelling continental water storage changes – sensitivity to different climate datasets. *Adv. Geosci.* 11, 63–68. <<http://www.adv-geosci.net/11/63/2007/>>.
- Finger, D., Pellicciotti, F., Konz, M., Rimkus, S., Burlando, P., 2011. The value of glacier mass balance, satellite snow cover images, and hourly discharge for improving the performance of a physically based distributed hydrological model. *Water Resour. Res.* 47, W07519. <http://dx.doi.org/10.1029/2010WR009824>.
- Gleeson, T., Moosdorf, N., Hartmann, J., van Beek, L.P.H., 2014. A glimpse beneath earth's surface: GLobal HYdrogeology MaPs (GLHYMPS) of permeability and porosity. *Geophys. Res. Lett.* 41, 3891–3898. <http://dx.doi.org/10.1002/2014GL059856>.
- Gleeson, T., Smith, L., Moosdorf, N., Hartmann, J., Dürr, H.H., Manning, A.H., van Beek, L.P.H., Jellinek, A.M., 2011. Mapping permeability over the surface of the Earth. *Geophys. Res. Lett.* 38, L02401. <http://dx.doi.org/10.1029/2010GL045565>.
- Gosling, S.N., Bretherton, D., Haines, K., Arnell, N.W., 2010. Global hydrology modelling and uncertainty: running multiple ensembles with a campus grid. *Philos. Trans. R. Soc. A* 368, 4005–4021. <http://dx.doi.org/10.1098/rsta.2010.0164>.
- Gosling, S.N., Arnell, N.W., 2011. Simulating current global river runoff with a global hydrological model: model revisions, validation and sensitivity analysis. *Hydrolog. Process.* 25, 1129–1145.
- Gregory, K.J., 1976. Drainage networks and climate. In: Derbyshire, E. (Ed.), *Geomorphology and Climate*. John Wiley, London.
- GRDC, 2007. Major River Basins of the World/Global Runoff Data Centre, D-56002, Federal Institute of Hydrology (BfG), Koblenz, Germany.
- Gudmundsson, L., Tallaksen, L.M., Stahl, K., Clark, D.B., Hagemann, S., Bertrand, N., Gerten, D., Heinke, J., Hanasaki, N., Voss, F., Koirala, S., 2011. Comparing large-scale hydrological model simulations to observed runoff percentiles in Europe. *J. Hydrometeorol.* 13, 604–620. <http://dx.doi.org/10.1175/JHM-D-11-083.1>.
- Gupta, H.V., Sorooshian, S., Yapo, P.O., 1998. Toward improved calibration of hydrological models: multiple and noncommensurable measures of information. *Water Resour. Res.* 34, 751–763.
- Gupta, H.V., Wagener, T., Liu, Y., 2008. Reconciling theory with observations: elements of a diagnostic approach to model evaluation. *Hydrolog. Process.* 22 (18), 3802–3813. <http://dx.doi.org/10.1002/hyp.6989>.
- Haan, C.T., Storm, D.E., Al-Issa, T., Prabhu, S., Sabbagh, G.J., Edwards, D.R., 1998. Effect of parameter distribution on uncertainty analysis of hydrologic models. *Trans. ASAE* 41 (1), 65–70.
- Haddeland, I. et al., 2011. Multimodel estimate of the global terrestrial water balance: setup and first results. *J. Hydrometeorol.* 12, 869–884.
- Hagemann, S., Botzet, M., Dümenil, L., Machenhauer, B., 1999. Derivation of global GCM boundary conditions from 1 km land use satellite data Max-Planck-Institute for Meteorology, Report 289, Hamburg, Germany. <<http://www.mpimet.mpg.de/deutsch/Sonst/Reports/HTMLReports/289/>>.
- Hagemann, S., Gates, L. Dümenil, 2003. Improving a subgrid runoff parameterization scheme for climate models by the use of high resolution data derived from satellite observations. *Clim. Dyn.* 21, 349–359.
- Higgins, R.W., Kousky, V.E., Silva, V.B.S., Becker, E., Xie, P., 2010. Intercomparison of daily precipitation statistics over the United States in observations and in NCEP reanalysis products. *J. Climate* 23, 4637–4650. <http://dx.doi.org/10.1175/2010JCLI3638.1>.
- Kraijenhoff van de Leur, D., 1958. A study of non-steady groundwater flow with special reference to a reservoir coefficient. *De Ingenieur* 70, 87–94.
- Krishnamurti, T.N., Kishtawal, C.M., LaRow, T., Bachiochi, D., Zhang, Z., Williford, C. E., Gadgil, S., Surendran, S., 1999. Improved forecast skill of weather and seasonal climate forecasts from multimodel superensemble. *Science* 285, 1548–1550.
- Kuchment, L.S., 2004. The hydrological cycle and human impact on it. In: Hoekstra, Arjen Y., Savenije, Hubert, H.G. (Eds.), *Water Resources Management*. In: Encyclopedia of Life Support Systems (EOLSS). Developed under the Auspices of the UNESCO. Eolss Publishers, Oxford, UK. <<http://www.eolss.net>>.
- Lambe, T.S., Whitman, R., 1969. *Soil Mechanics*. John Wiley & Sons, Inc., New York.
- Leemans, R., 1990. Global Data Sets Collected and Compiled by the Biosphere Project, Working Paper, IIASA-Laxenburg, Austria.
- Martinec, J., 1975. Snowmelt – runoff model for stream flow forecasts. *Nordic Hydrol.* 6 (3), 145–154. <http://dx.doi.org/10.2166/nh.1975.010>.
- McMillan, H., Freer, J., Pappenberger, F., Krueger, T., Clark, M., 2010. Impacts of uncertain river flow data on rainfall-runoff model calibration and discharge predictions. *Hydrolog. Process.* 24, 1270–1284.
- Mitchell, T.D., Jones, P.D., 2005. An improved method of constructing a database of monthly climate observations and associated high-resolution grids. *Int. J. Climatol.* 25, 693–712.
- Monteith, J.L., 1965. Evaporation and environment. *Symp. Soc. Exp. Biol.* 19, 205–234.
- Morris, D.A., Johnson, A.I., 1967. Summary of Hydrologic and Physical Properties of Rock and Soil Materials as Analyzed by the Hydrologic Laboratory of the U.S. Geological Survey, U.S. Geol. Surv. Water-Supply Paper 1839-D, 42p.
- Muleta, M.L., Nicklow, J.W., 2005. Sensitivity and uncertainty analysis coupled with automatic calibration for a distributed watershed model. *J. Hydrol.* 306, 127–145.
- Nasonova, O.N., Gusev, Y.M., Kovalev, Y.E., 2011. Impact of uncertainties in meteorological forcing data and land surface parameters on global estimates of terrestrial water balance components. *Hydrolog. Process.* 25 (7), 1074–1090. <http://dx.doi.org/10.1002/hyp.7651>.
- New, M., Hulme, M., Jones, P., 1999. Representing twentieth-century space-time climate variability. Part 1: Development of a 1961–90 mean monthly terrestrial climatology. *J. Climate* 12, 829–856.
- New, M., Hulme, M., Jones, P., 2000. Representing twentieth century space-time climate variability. Part 2: Development of 1901–96 monthly grids of terrestrial surface climate. *J. Climate* 13, 2217–2238.
- Nilsson, B., Reidy, C.A., Dynesius, M., Revenga, C., 2005. Fragmentation and flow regulation of the world's largest river systems. *Science* 308 (5720), 405–408. <http://dx.doi.org/10.1126/science.1107887>.
- Nijssen, B., O'Donnell, G.M., Lettenmaier, D.P., Lohmann, D., Wood, E.F., 2001. Predicting the discharge of global rivers. *J. Climate* 14, 3307–3323.
- Olson, J.S., 1994. Global Ecosystem Framework-definitions: USGS EROS Data Center Internal Report, Sioux Falls, SD, 37 p.
- Pappenberger, F., Thielen, J., Del Medico, M., 2011. The impact of weather forecast improvements on large scale hydrology: analysing a decade of forecasts of the European Flood Alert System. *Hydrolog. Process.* 25 (7), 1091–1113. <http://dx.doi.org/10.1002/hyp.7772>.
- Rasmussen, W.C., 1963. Permeability and Storage of Heterogeneous Aquifers in the United States. International Union of Geodesy and Geophysics, August 1963, Berkeley.
- Renard, B., Kavetski, D., Keczer, G., Thyer, M., Franks, S.W., 2010. Understanding predictive uncertainty in hydrologic modeling: the challenge of identifying input and structural errors. *Water Resour. Res.* 46, 5. <http://dx.doi.org/10.1029/2009WR008328>.
- Saha, S., Moorthi, S., Pan, H.-L., Wu, X., Wang, J., Nadiga, S., Tripp, P., Kistler, R., Woollen, J., Behringer, D., Liu, H., Stokes, D., Grumbine, R., Gayno, G., Wang, J., Hou, Y.-T., Chuang, H.-Y., Juang, H.-M.H., Sela, J., Iredell, M., Treadon, R., Kleist, D., van Delst, P., Keyser, D., Derber, J., Ek, M., Meng, J., Wei, H., Yang, R., Lord, S., van den Dool, H., Kumar, A., Wang, W., Long, C., Chelliah, M., Xue, Y., Huang, B., Schemm, J.-K., Ebisuzaki, W., Lin, R., Xie, P., Chen, M., Zhou, S., Higgins, W., Zou, C.-Z., Liu, Q., Chen, Y., Han, Y., Cucurull, L., Reynolds, R.W., Rutledge, G., Goldberg, M., 2010. The NCEP climate forecast system reanalysis. *Bull. Am. Meteorol. Soc.*, 1015–1057.
- Seibert, J., 1999. Regionalisation of parameters for a conceptual rainfall-runoff model. *Agric. Forest Meteorol.* 98–99, 279–293.
- Seneviratne, S.I., Corti, T., Davin, E.L., Hirschi, M., Jaeger, E.B., Lehner, I., Orlowsky, B., Teuling, A.J., 2010. Investigating soil moisture–climate interactions in a changing climate: a review. *Earth-Sci. Rev.* 99 (3–4), 125–161. <http://dx.doi.org/10.1016/j.earscirev.2010.02.004>.
- Sorooshian, S., Dracup, J.A., 1980. Stochastic parameter estimation procedures for hydrologic rainfall–runoff models: correlated and heteroscedastic error cases. *Water Resour. Res.* 16, 430–442.
- Sperna Weiland, F.C., Van Beek, L.P.H., Kwadijk, J.C.J., Bierkens, M.F.P., 2010. The ability of a GCM-forced hydrological model to reproduce global discharge variability. *Hydrolog. Earth Syst. Sci.* 14, 1595–1621. <http://dx.doi.org/10.5194/hess-14-1595-2010>.
- Sperna Weiland, F.C., Tisseuil, C., Dürr, H.H., Vrac, M., van Beek, L.P.H., 2011a. Selecting the optimal method to calculate daily global reference potential

- evaporation from CFSR reanalysis data. *Hydrol. Earth Syst. Sci. Discuss.* 8, 7355–7398. <http://dx.doi.org/10.5194/hessd-8-7355-2011>.
- Sperna Weiland, F.C., van Beek, L.P.H., Kwadijk, J.C.J., Bierkens, M.F.P., 2011b. On the usability of GCM runoff fields for river discharge modeling: a case study using model output from HadGEM2 and ECHAM5. *J. Hydrometeorol.* <http://dx.doi.org/10.1175/JHM-D-10-05011.1>.
- Sutanudjaja, E.H., van Beek, L.P.H., de Jong, S.M., van Geer, F.C., Bierkens, M.F.P., 2011. Large-scale groundwater modeling using global datasets: a test case for the Rhine-Meuse basin. *Hydrol. Earth Syst. Sci. Discuss.* 8, 2555–2608. <http://dx.doi.org/10.5194/hessd-8-2555-2011>.
- Teuling, A.J., Uijlenhoet, R., Van den Hurk, B., Seneviratne, S.I., 2009. Parameter sensitivity in LSMs: an analysis using stochastic soil moisture models and ELDAS soil parameters. *J. Hydrometeorol.* 10 (3), 751–763. <http://dx.doi.org/10.1175/2008JHM1033.1>.
- Troppocci, A., Källberg, P.W., 2004. Precipitation Correction in the ERA-40 Reanalysis. ERA-40 Project Rep. Series 13, ECMWF, Reading, UK, 6 pp.
- Troy, T.J., Wood, E.G., Sheffield, J., 2008. An efficient calibration method for continental-scale land surface modeling. *Water Resour. Res.* <http://dx.doi.org/10.1029/2007WR006513>.
- Uppala, S.M., Källberg, P.W., Simmons, A.J., Andrae, U., da Costa Bechtold, V., Fiorino, M., Gibson, J.K., Haseler, J., Hernandez, A., Kelly, G.A., Li, X., Onogi, K., Saarinen, S., Sokka, N., Allan, R.P., Andersson, E., Arpe, K., Balmaseda, M.A., Beljaars, A.C.M., van de Berg, L., Bidlot, J., Bormann, N., Caires, S., Chevallier, F., Dethof, A., Dragosavac, M., Fisher, M., Fuentes, M., Hagemann, S., Hólm, E., Hoskins, B.J., Isaksen, L., Janssen, P.A.E.M., Jenne, R., McNally, A.P., Mahfouf, J.-F., Morcrette, J.-J., Rayner, N.A., Saunders, R.W., Simon, P., Sterl, A., Trenberth, K.E., Untch, A., Vasiljevic, D., Viterbo, P., Woollen, J., 2005. The ERA-40 re-analysis. *Quart. J. R. Meteorol. Soc.* 131, 2961–3012.
- Van Beek, L.P.H., 2008. Forcing PCR-GLOBWB with CRU Meteorological Data. Utrecht University, Utrecht, Netherlands. <<http://vanbeek.geo.uu.nl/supinfo/vanbeek2008.pdf>>.
- Van Beek, L.P.H., Bierkens, M.F.P., 2009. The Global Hydrological Model PCR-GLOBWB: Conceptualization, Parameterization and Verification. Report Department of Physical Geography, Utrecht University, Utrecht, Netherlands. <<http://vanbeek.geo.uu.nl/supinfo/vanbeekbierkens2009.pdf>>.
- Van Beek, L.P.H., Wada, Y., Bierkens, M.F.P., 2011. Global monthly water stress: I. Water balance and water availability. *Water Resour. Res.* 47 (7). <http://dx.doi.org/10.1029/2010WR009791>.
- Vansteenkiste, T., Tavakoli, M., Ntegeka, V., Willem, P., De Smedt, F., Batelaan, O., 2012. Climate change impact on river flows and catchment hydrology: a comparison of two spatially distributed models. *Hydrol. Process.* 27 (25). <http://dx.doi.org/10.1002/hyp.9480>.
- Vaze, J., Post, D.A., Chiew, F.H.S., Perraud, J.-M., Viney, N.R., Teng, J., 2010. Climate non-stationarity – validity of calibrated rainfall-runoff models for use in climate change studies. *J. Hydrol.* 394, 447–457. <http://dx.doi.org/10.1016/j.jhydrol.2010.09.018>.
- Verdin, K.L., Greenlee, S.K., 1996. Development of continental scale digital elevation models and extraction of hydrographic features. In: Proceedings of the Third International Conference/Workshop on Integrating GIS and Environmental Modeling, Santa Fe, New Mexico, January 21–26.
- Vrugt, J.A., Robinson, B.A., 2007. Treatment of uncertainty using ensemble methods: comparison of sequential data assimilation and Bayesian model averaging. *Water Resour. Res.* 43, W01411. <http://dx.doi.org/10.1029/2005WR004838>.
- Vrugt, J.A., Sadegh, M., 2013. Toward diagnostic model calibration and evaluation: approximate Bayesian computation. *Water Resour. Res.* 49. <http://dx.doi.org/10.1002/wrcr.20354>.
- Vrugt, J.A., Gupta, H.V., Bouten, W., Sorooshian, S., 2003. A shuffled complex evolution Metropolis algorithm for optimization and uncertainty assessment of hydrological model parameters. *Water Resour. Res.* 39, 1201.
- Vrugt, J.A., Diks, C.G.H., Gupta, H.V., Bouten, W., Verstraten, J.M., 2005. Improved treatment of uncertainty in hydrologic modeling: combining the strengths of global optimization and data assimilation. *Water Resour. Res.* 41, W01017. <http://dx.doi.org/10.1029/2004WR003059>.
- Wada, Y., Van Beek, L.P.H., Van Kempen, C.M., Reckman, J.W.T.M., Vasak, S., Bierkens, M.F.P., 2010. Global depletion of groundwater resources. *Geophys. Res. Lett.* 37, L20402. <http://dx.doi.org/10.1029/2010GL044571>.
- Wada, Y., van Beek, L.P.H., Viviroli, D., Dürr, H.H., Weingartner, R., Bierkens, M.F.P., 2011. Global monthly water stress: II. Water demand and severity of water stress. *Water Resour. Res.* 47, W07518. <http://dx.doi.org/10.1029/2010WR009792>.
- Wang, W., Xie, P., Yoo, S.-H., Xue, Y., Kumar, A., Wu, X., 2010. An assessment of the surface climate in the NCEP climate forecast system reanalysis. *Clim. Dyn.* <http://dx.doi.org/10.1007/s00382-010-0935-7>.
- Widén-Nilson, E., Halldin, S., Xu, C.-Y., 2007. Global water balance modelling with WASMOD-M: parameter estimation and regionalization. *J. Hydrol.* 340, 105–118.
- Winsemius, H.C., Savenije, H.H.G., Gerrits, A.M.J., Zapreeva, E.A., Klees, R., 2006. Comparison of two model approaches in the Zambezi river basin with regard to model reliability and identifiability. *Hydrol. Earth Syst. Sci.* 10, 339–352.
- Wood, E.F., Lettenmaier, D.P., Zartarian, V.G., 1992. A landsurface hydrology parameterization with subgrid variability for general circulation models. *J. Geophys. Res.* 97 (D3), 2717–2728.
- Zhang, L., Kumar, A., Wang, W., 2012. Influence of changes in observations on precipitation: a case study for the Climate Forecast System Reanalysis (CFSR). *J. Geophys. Res.* 117, D08105. <http://dx.doi.org/10.1029/2011JD017347>.