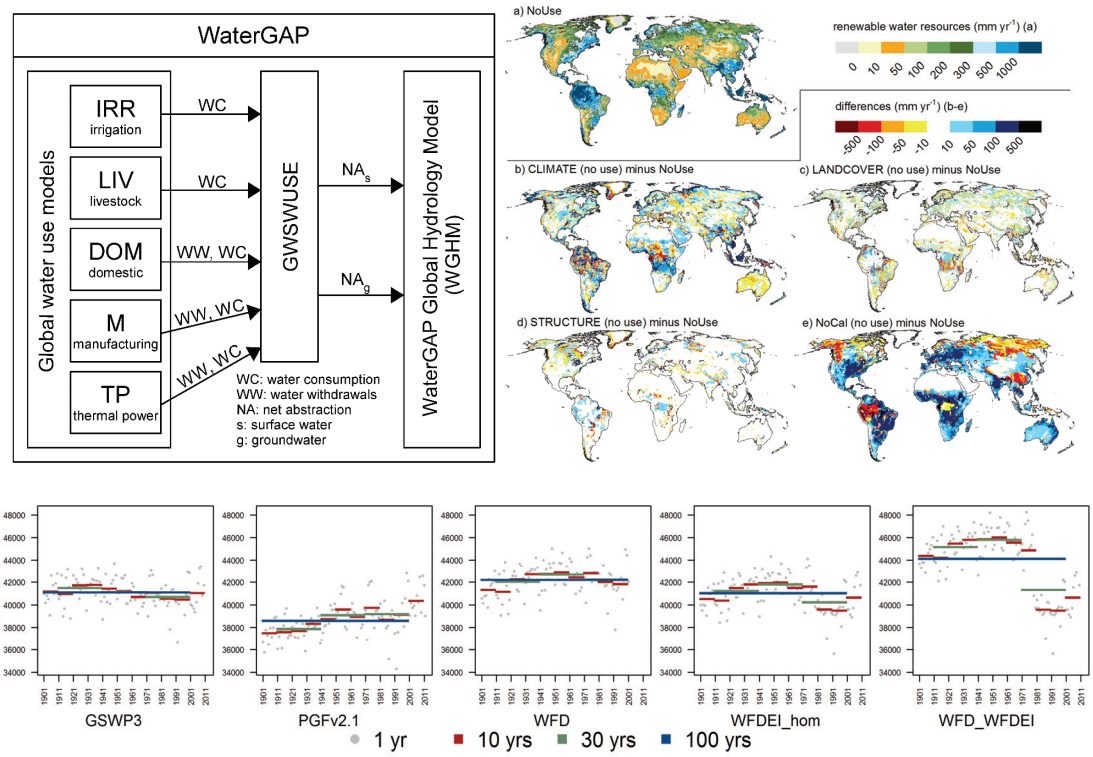
16

Evaluation, modification and application of a global hydrological model



Hannes Müller Schmied

2017

Frankfurt Hydrology Paper

GOETHE UNIVERSITÄT FRANKFURT AM MAIN

Frankfurt Hydrology Papers:

01 **A Digital Global Map of Irrigated Areas - An Update for Asia**

02 **Global-Scale Modeling of Nitrogen Balances at the Soil Surface**

03 **Global-Scale Estimation of Diffuse Groundwater Recharge**

04 **A Digital Global Map of Artificially Drained Agricultural Areas**

05 **Irrigation in Africa, Europe and Latin America - Update of the Digital Global Map of Irrigation Areas to Version 4**

06 **Global data set of monthly growing areas of 26 irrigated crops**

07 **The Global Crop Water Model (GCWM): Documentation and first results for irrigated crops**

08 **Towards mapping the extent of irrigation in the last century: time series of irrigated area per country**

09 **Global estimation of monthly irrigated and rainfed crop areas on a 5 arc-minute grid**

10 **Entwicklung der transdisziplinären Methode „Akteursbasierte Modellierung“ und ihre Anwendung im Problemfeld der mobilen, organischen Fremdstoffe**

11 **Expert-based Bayesian Network modeling for environmental management**

12 **Anthropogenic river flow alterations and their impacts on freshwater ecosystems in China**

13 **Design, implementation and evaluation of a participatory strategy development – A regional case study in the problem field of renewable electricity generation**

14 **Global-scale modelling and quantification of indicators for assessing transboundary aquifers. A contribution to the GEF Transboundary Waters Assessment Programme (TWAP)**

15 **Evaluating alternative water sources and their use for small-holder agriculture from a systemic perspective. A focus on water reuse and rainwater harvesting in Namibia**

16 **Evaluation, modification and application of a global hydrological model**

Institute of Physical Geography, Goethe University Frankfurt
Altenhöferallee 1, 60438 Frankfurt, Germany
Phone +49 (0)69 798 40155, Fax +49 (0)69 798 40170
http://www.uni-frankfurt.de/45217668/dl

Please cite as:
Müller Schmied, H. (2017): Evaluation, modification and application of a global hydrological model. *Frankfurt Hydrology Paper 16*, Institute of Physical Geography, Goethe University Frankfurt, Frankfurt am Main, Germany.

EVALUATION, MODIFICATION AND APPLICATION OF A GLOBAL HYDROLOGICAL MODEL

Dissertation
Zur Erlangung des Doktorgrades
der Naturwissenschaften

vorgelegt beim Fachbereich 11 Geowissenschaften/Geographie
der Johann Wolfgang Goethe-Universität
in Frankfurt am Main

von
Hannes Müller Schmied
aus Erfurt

Frankfurt am Main 2016
(D 30)

vom Fachbereich Geowissenschaften/Geographie der
Johann Wolfgang Goethe - Universität als Dissertation angenommen.

Dekan: Prof. Dr. Peter Lindner

1. Gutachter: Prof. Dr. Petra Döll
Institut für Physische Geographie,
Johann Wolfgang Goethe-Universität

2. Gutachter: Prof. Dr. Andreas Güntner
Deutsches GeoForschungsZentrum GFZ Potsdam
und
Institut für Erd- und Umweltwissenschaften,
Universität Potsdam

Datum der Disputation: 08. Februar 2017

Das Wasser ist ein freundliches Element für den,
der damit bekannt ist und es zu behandeln weiß.

Johann Wolfgang von Goethe

Table of contents

Abstract

The estimation of water balance components as well as water-related indicators on the land surface by means of global hydrological models have evolved in recent decades. Results of such models are frequently used in global- and continental-scale assessments of the current and future state of the terrestrial water cycle and provide a valuable data basis, e.g., for the Intergovernmental Panel on Climate Change. The Water – Global Assessment and Prognosis (WaterGAP) model is one of the state-of-the-art models in that field and has been in development and application for around 20 years. The evaluation, modification and application of WaterGAP is the subject of this thesis. In particular, the sensitivity of climate input data on radiation calculation and simulated water fluxes and storages is evaluated in the first part. Effects of model modification such as updated spatial input datasets, improved process representation or an alternative calibration scheme are the focus of the second part. Finally, three applications of WaterGAP give insight into the capabilities of that model, namely an estimate of global and continental water balance components, an assessment of groundwater depletion and the impact of climate change on river flow regimes. Model experiments, which are described in six journal papers as well as the appendices, were used as the basis for answering the total of 13 research questions. One of the major foci was to quantify the sensitivity of simulated water fluxes and storages to alternative climate input data. It was found that the handling of precipitation undercatch leads to the greatest difference in water balance components, especially in those areas where WaterGAP is not calibrated due to a lack of river discharge observations. The modifications of WaterGAP in the last few decades has led in general to an improved simulation of monthly river discharge, but process representation in semi-arid and arid regions still requires improvements. With the most current model version, WaterGAP 2.2b, and for the time period 1971–2000, river discharge to the oceans and inland sinks is estimated to be 40 000 $km^3\ yr^{-1}$, whereas actual evapotranspiration is simulated as 70 500 $km^3\ yr^{-1}$. Future research needs for WaterGAP in particular but also for the global hydrological model community in general are defined, promoting a community-driven effort for a robust assessment of the continental water cycle.

Abstract

Globale hydrologische Modelle gewannen in den letzten Jahrzehnten für die Berechnung von Wasserhaushaltskomponenten und wasserbezogenen Indikatoren auf der Landoberfläche an Bedeutung. Ergebnisse dieser Modelle werden häufig für globale und auf Kontinente bezogene Bewertungen des derzeitigen und projektierten Zustandes des terrestrischen Wasserkreislaufes verwendet und bilden eine wertvolle Datenbasis z.B. im Rahmen des Intergovernmental Panel on Climate Change. Das Water – Global Assessment and Prognosis (WaterGAP) Modell ist eines der weitverbreitetsten und aktuellsten Modelle in diesem Zusammenhang, und in Entwicklung und Anwendung seit etwa 20 Jahren. Die Evaluation, Entwicklung und Anwendung von WaterGAP ist Gegenstand dieser Dissertationsschrift. Im ersten Teil der Arbeit wird die Auswirkung der Unsicherheit von klimatischen Antriebsdaten und der Strahlungsberechnung auf berechnete Wasserflüsse und -speicher untersucht. Aktualisierte räumlich verteilte Eingangsdaten, verbesserte Prozessbeschreibungen sowie ein alternativer Kalibrieransatz sind Schwerpunkt des zweiten Teils, der Modellentwicklung. Im dritten Teil der Arbeit werden drei Anwendungen von WaterGAP vorgestellt, die einen Einblick in die Möglichkeiten des Modells geben. Konkret werden Abschätzungen der globalen und kontinentalen Wasserbilanz-komponenten, der Grundwasserzehrung sowie des Einflusses des Klimawandels auf Durchflussregime thematisiert. Die Basis der Arbeit bilden Modellexperimente, die in sechs international publizierten Artikeln sowie dem Anhang beschrieben sind, und die insgesamt 13 Forschungsfragen beantworten. Ein wesentlicher Schwerpunkt der Arbeit bildet die Beurteilung der Sensitivität von berechneten Wasserflüssen und -speichern auf verschiedene klimatische Antriebsdaten. Es wurde festgestellt, dass die hauptsächliche Unsicherheit auf der unterschiedlichen Berücksichtigung der Niederschlagskorrektur beruht und diese Unsicherheit insbesondere in den Gebieten groß ist, in dem keine Kalibrierung auf gemessenem Durchfluss möglich ist. Die Entwicklung von WaterGAP in den letzten Jahrzehnten führte zu einer verbesserten Simulation von monatlichen Durchflüssen. Allerdings wurden große Defizite in semi-ariden und ariden Gebieten festgestellt. Unter Verwendung der derzeit aktuellsten Modellversion WaterGAP 2.2b und des Zeitraums 1971 bis 2000 wurde der Durchfluss in die Ozeane und in Inlandsenken mit 40000 km^3 a^{-1} quantifiziert, die aktuelle Evapotranspiration mit 70500 km^3 a^{-1}. Einige Forschungsideen und Anwendungsmöglichkeiten für WaterGAP, aber auch für globale hydrologische Modelle im Allgemeinen, werden definiert, die in gemeinschaftlicher Arbeit zu einer robusten Quantifizierung des terrestrischen Wasserkreislaufs führen können.

1 Introduction

The evaluation, modification and application of the Water – Global Assessment and Prognosis (WaterGAP) is the subject of this thesis. In the following, the background of the thesis is introduced with some general insights into the global water cycle (Chapter 1.1) and human impacts on it (Chapter 1.2), as well as an overview of approaches to quantifying (parts of) the global water cycle (Chapter 1.3). More specific, global hydrological models and their uncertainties (Chapter 1.3.4) and the WaterGAP model used for this thesis (Chapter 1.4) are described. The objectives and research questions are formulated in Chapter 1.5. Chapters 2 to 7 consist of six journal papers that, in conjunction with the appendices, form the basis on which to answer the research questions in Chapter 8. The thesis is summarized in Chapter 9, which also presents an outlook on future research possibilities.

1.1 The global water cycle

Water moves in a global cycle as precipitation, evapotranspiration and runoff (discharge) between storages (e.g., glaciers, soil, surface water, groundwater, atmosphere, oceans) and plays a vital role in many regards. Water is the origin of life, forms landscapes and is a medium for diluted and solid materials. Aquatic ecosystems are fully dependent on the presence or absence of water (Dudgeon et al., 2006), and nearly all other ecosystems have a strong interaction with the resource of water (e.g., Austin et al., 2004; Bernacchi and VanLoocke, 2015). Not only do humans need water, e.g., for drinking, washing and cleaning; humans also consume it indirectly as virtual water (Hoekstra et al., 2011) that is included in foods, e.g., corn and meat, and other commodities, e.g., cotton and leather. On the other hand, water is needed to cool thermal power plants and to produce goods in industry. Water is a resource for joy, recreation and tourism but – especially if only small amounts are available – can also influence conflicts between countries or authorities as described in the case of the current situation in Syria by Kelley et al. (2015). But there are a number of other stressors, e.g., to freshwater as well as river biodiversity (Vörösmarty et al., 2010).

1.2 Human impacts on the (continental) water cycle

Humans have altered the water cycle in many ways, which are often interrelated. Changes in land use, especially intensification of agriculture, results in modified surface runoff and river discharge, e.g., when forest is replaced by agriculture (Foley et al., 2005; Piao et al., 2007). Closely related to this is the increasing area that is irrigated (Siebert et al., 2015) and the amount of water withdrawn to satisfy plant needs (de Marsily and Abarca-del-Rio, 2016). In some regions, where groundwater aquifers are overused, groundwater depletion

can occur with negative impacts for surrounding (groundwater-dependent) ecosystems and higher costs for pumping water. Such an additional pumping of water (especially for non-renewable fossil groundwater resources) can lead to sea level rise, even though the proportion related to other causes of sea level rise is disputed (Wada et al., 2016a). Some components of the interrelation of land use change and intensification include salinization; erosion (Smith et al., 2016); and, due to intensified fertilization and washout of soils, the water quality of rivers (Heathwaite, 2010). Alterations of the water cycle have also taken place due to the creation of dams leading to artificial reservoirs and regulated lakes. Lehner et al. (2011) created the Global Reservoir and Dam database (GRanD), which contains 6862 dams having a total storage capacity of 6200 km^3, resulting in 7.6% of worldwide rivers (with more than 1 m^3s^{-1} discharge) being affected by more than 2% of their annual flow. For the future, a large number of new dams are planned, which according to Zarfl et al. (2014) will reduce the number of remaining free-flowing rivers by 20%. Dam operation especially influences discharge seasonality as well as – due to the additional evaporation of the artificial water body – mean annual river flow (e.g., Döll et al., 2009), and reservoir water impoundment influences global sea level rise (Chao et al., 2008). Climate change effects on hydrology are the subject of a range of studies. As it is hard to assess climate change effects based on observations only, mainly due to a lack of long-term observations, e.g., of glaciers (Fernández and Mark, 2016), modeling approaches are often used, especially when focusing on climate extremes (Easterling et al., 2000). The effect of human impact on changed precipitation patterns is shown in the recent study of Schaller et al. (2016) for the 2014 flood event in southern England. Other studies use a multi-model and multi-forcing approach to assess projected future changes in water resources (e.g., Schewe et al., 2014).

1.3 Quantification of the (continental) water cycle components

1.3.1 Early assessments

The quantification of water balance components plays a vital role in assessments of historical and future water scarcity (Gosling and Arnell, 2013; Hertel, 2015; Hoekstra et al., 2012; Vörösmarty et al., 2010) and is also important in the water–food–energy nexus (Ringler et al., 2013). Due to its importance both for storages (e.g., to satisfy the demand from reservoir users) and river discharge (e.g., for management purposes), there is a long history of attempts to quantify water balance components. In the early work of Baumgartner and Reichel (1968, their Table 2), the authors compare assessments from the 19th and 20th century, which show a large range of uncertainty, e.g., for discharge from land areas (17 000–56 000 $km^3\ yr^{-1}$) and precipitation over land areas (78 000–122 000 $km^3\ yr^{-1}$). The authors themselves assess precipitation over land at 100 000 $km^3\ yr^{-1}$, evaporation from land at 65 000 $km^3\ yr^{-1}$ and discharge from land at 35 000 $km^3\ yr^{-1}$. In principle they

followed two approaches: i) they used measured data for precipitation to create a precipitation chart, used discharge data from Lvovitch (1964) as well as Marcinek (1964) and calculated evaporation using the Thornthwaite method, and ii) they flipped the water balance equation (which is also called the Birkner–Oppokov equation according to Lvovitch (1964)) to obtain the evaporation term as residual. Using both methods, Baumgartner and Reichel (1968) could estimate the uncertainty of the evaporation term. Even though the data basis is not comparable to recent available sources, those numbers have frequently been used as a reference in recent studies (e.g., Clark et al., 2015; Müller Schmied et al., 2014; Pan et al., 2012; Vinukollu et al., 2011).

1.3.2 Measurement-based approaches

Measuring components of the water cycle has historically been done for precipitation and other meteorological variables as well as for river discharge. The common station-based measurements of, e.g., precipitation have passed the times of the highest number of stations being globally available (Schneider et al., 2014), mainly for financial reasons. Traditional precipitation measurements are subject to a range of uncertainty, e.g., wind drift and undercatch errors (McMillan et al., 2012). Therefore, some general correction schemes exist (Adam and Lettenmaier, 2003; Hirabayashi et al., 2008; Willmott and Johnson, 2005); however, they cannot reduce the error completely (Döll et al., 2016). Due to its value for, e.g., management purposes, river discharge has been measured for centuries using different techniques. Here, measurement uncertainties are also high (McMillan et al., 2012); even for high-quality data (as in the example of the United Kingdom), errors can range up to 80% (Coxon et al., 2015). A number of discharge station data, but by no means all of them, are reported to databases, e.g., hosted by the Global Runoff Data Centre (GRDC). In addition to the decreasing number of stations in operation, sharing of data is limited, e.g., in water-scarce transboundary basins for political reasons, which hampers data-driven global analysis of water balance components (Fekete et al., 2015; Hannah et al., 2011). Even though a number of satellite missions can observe (parts of) the water cycle components (Famiglietti et al., 2015), in situ data are often needed for calibration (Fekete et al., 2015). Precipitation and river discharge are the two components which can be measured with reasonable accuracy.

The “loss” term of the water cycle, the evapotranspiration, is more complicated to measure. Whereas the potential evapotranspiration (PET, i.e., the amount that evaporates when water availability is unlimited) is assessed frequently and with relatively high spatial coverage using evaporation pans (Abtew et al., 2011; Jovanovic et al., 2008; Matsoukas et al., 2011; McVicar et al., 2007, 2012; Sanchez-Lorenzo et al., 2014; Yang and Yang, 2012), measurements of actual evapotranspiration (which takes water limitation into account and is thus of special interest for global assessments) are rather limited. One reasonable way

to determine actual evapotranspiration is by using weighted lysimeters (soil columns which are weighted and thus allow measuring water balance components), but those experiments are expensive and focused on special conditions (e.g., soil properties and type of vegetation) and management purposes (Allen et al., 2011; Liu et al., 2002; Rana and Katerji, 2000; Verstraeten et al., 2008). To enhance the communication of results and as a first step of data sharing, the Lysimeter Research Group (http://www.lysimeter.at) was established. In addition to some other methods like Bowen ratio and scintillometers, the so-called eddy flux towers are another resource of point-scale measurements of the actual evapotranspiration (Verstraeten et al., 2008). Due to its close relationship with carbon fluxes, a large number of such towers (844, thereof 567 currently active in June 2016) were installed in the past, and data are accessible via the FluxNet network (http://fluxnet.ornl.gov/). Jung et al. (2010) used data from those towers for a machine learning algorithm and combined it with remote sensing data to produce an observation-based data product of actual evapotranspiration which is frequently used as benchmark product (e.g., Mueller et al., 2013).

A relatively new branch of quantifying water balance components is the usage of the isotope signature of water to derive their source. For example, Schlesinger and Jasechko (2014) reviewed studies that distinguish the proportion of transpiration and evaporation due to the circumstance of the isotope $\delta^{18}O$ being enriched by evaporation but not by transpiration. Many other recent studies show the potential of this method to quantify, e.g., the groundwater volume, its age or circumstances of recharge (Evaristo et al., 2015; Gleeson et al., 2015; Jasechko, 2016; Jasechko and Taylor, 2015); to partition transpiration and evaporation (Kool et al., 2014); and to determine the age of river discharge (Jasechko et al., 2016) or water balances of catchments (Schulte et al., 2011).

1.3.3 Satellite-based approaches

In the era of satellite missions, several attempts were undertaken to quantify water balance components with the help of remote sensing (Rast et al., 2014; Rodell et al., 2015; Tang et al., 2009; Zhang et al., 2016). Whereas, e.g., surface water level estimation from space is in operation and provide reasonable accuracy (Durand et al., 2016; Ričko et al., 2012), research is still needed to accurately observe soil moisture by satellites (Al-Yaari et al., 2014; Dorigo et al., 2012). Several products of actual evapotranspiration are available, e.g., based on the Moderate Resolution Imaging Spectroradiometer (MODIS) satellite (Mu et al., 2011a, 2011b; Ruhoff et al., 2013) or other spacecrafts (Miralles et al., 2011), but these are not solely observations as, e.g., a PET model is included. For precipitation, radar observations from space were developed in the past (Neeck et al., 2005) and are incorporated, e.g., into the Tropical Rainfall Measuring Mission (TRMM) product (Tang et al., 2009). The Gravity Recovery And Climate Experiment (GRACE) mission provides monthly

information on total mass change with a spatial resolution of a few hundred kilometers (Tapley et al., 2004). With GRACE post-processing applied to achieve anomalies within the water compartment (Wahr et al., 1998), these variations are useful for hydrologic studies, e.g., to improve hydrological models (Eicker et al., 2014; Forman et al., 2012; Lo et al., 2010; Werth et al., 2009). In general there is strong potential in satellite-based observations of the hydrologic water cycle, and even more missions are under way (see details in Simmons et al., 2016).

1.3.4 (Global) hydrological models and their uncertainties

Models that simulate water balance components at the global scale as a basis for such applications can be classified into global hydrological models (GHMs); land surface models (LSMs); Earth system models (ESMs); and, to some degree, dynamic global vegetation models (DGVMs). DGVMs are specialized at representing the vegetation dynamics as well as structural and physiological response of plants to, e.g., changing CO_2 conditions and are also a subject to hydrological applications, e.g., the model Lund–Potsdam–Jena managed Land (LPJmL; Gerten, 2013) or effects of climate change on agriculture (Rosenzweig et al., 2014). Recent developments, the state of the art and research needs are discussed in detail in Bierkens (2015), Döll et al. (2016), Pokhrel et al. (2016) and Sood and Smakhtin (2015) and are thus touched upon only briefly here. In general, GHMs are designed to assess water balance components (e.g., renewable water resources) from a hydrologic perspective, incorporating storage equations and subsequent flows between those storages. They usually calculate the vertical water balance (e.g., canopy, snow and soil storages) and contain a river routing scheme directly within the model. This allows for comparison and calibration, e.g., in the case of the GHM WaterGAP (Döll et al., 2003) γ to observed river discharge. Traditionally, actual evapotranspiration is calculated using the PET approach, and energy balance is often not closed (not as the water balance). Much effort has been expended in the past for GHMs to include human impacts on the hydrological cycle, e.g., reservoir operation (Döll et al., 2009; Hanasaki et al., 2006), water use schemes including the source of water withdrawal (Döll et al., 2012; Flörke et al., 2012; Wada et al., 2014) and many others (see the review of Pokhrel et al., 2016, for details).

LSMs were developed as the land component of global circulation models (GCMs) preserving both the water and the energy balance. In most cases, LSMs calculate runoff, not the accumulated river discharge. The latter is subsequently calculated using a river routing model like Total Runoff Integrating Pathways (TRIP; Decharme et al., 2012; Oki and Sud, 1998). Even though integration of human impact on the hydrological cycle is in development (Pokhrel et al., 2012), there is room for improvements (Pokhrel et al., 2016).

Models that combine the physical, chemical and biological components of the Earth system (i.e., land, atmosphere and ocean) in a fully coupled mode are called ESMs. They integrate

LSMs as land components; GCMs as the boundary for atmospheric circulation; and ocean models for the exchange of water and energy (as well as other fluxes) between land, atmosphere and ocean. Due to their integrative character, ESMs are designed to simulate the historical and future status of the Earth system with frequent use of observations in data assimilation schemes (Simmons et al., 2016). For example, Wada et al. (2016b) used an ESM to quantify the pathways of pumped groundwater and to better understand sea level rise due to groundwater depletion, and they concluded that the existing studies (that are not done with ESMs) tend to overestimate that contribution. As for LSMs, the human impact on the water cycle is an area in which improvements can be made in ESMs (Pokhrel et al., 2016).

As discharge observations are not available at every river, model approaches can utilize those measurements to close the gap to ungauged regions (e.g., Fekete et al., 2002; Gudmundsson and Seneviratne, 2016). A big benefit of using such global-scale models is that they provide spatially distributed model outputs and are thus of value for assessing the water cycle both temporally and spatially. According to the brief overview here (and especially the review of Pokhrel et al., 2016), human impacts on the hydrological cycle are currently best represented in GHMs. In consequence (and following Oki and Kanae, 2006, as they state that neglecting such impacts weaken water resources assessment), they can be seen currently as the first choice to quantify the world's water resources, especially the amount of renewable freshwater resources. Nevertheless, modeling approaches are subject to a number of uncertainties, which thus results in a large range of water resources estimates (e.g., as presented in Haddeland et al., 2011).

As summarized comprehensively by Döll et al. (2016), quantification of human water use and uncertain climate input data are two of the major challenges in historical water resources assessment. Both relate to scarcity of observed data (or, in the case of human water use, of statistical data with reasonable spatial resolution) and differ geographically. The uncertainty of climate input data has consequently been the object of study in the past (Biemans et al., 2009; Voisin et al., 2008; Wisser et al., 2010); nevertheless, as more and more plausible climate input data evolves and the correction of measured precipitation is still handled differently (or completely neglected) by the data producers, the uncertain precipitation data can be seen as a major driver of differences in water resources assessments, including the modeling of, e.g., irrigation water demand. But other climate variables like downward solar radiation are also subject to uncertainties, as is shown, e.g., for the different WATCH Forcing Data products (Weedon et al., 2011, 2014). Additional variables like wind speed and water vapor content have their own significance (and uncertainty) in model results, albeit not in the dimension of precipitation (Haddeland et al., 2012). Therefore, a consistent evaluation of the sensitivity of water balance components to climate input data is beneficial.

When the source code of a model is modified – e.g., due to the inclusion of new/enhanced process descriptions, spatial input data or parameters – it is often required to alter the model structure. There are a number of studies that show effects of different model structures on simulated outputs as well as frameworks to assess those effects (Butts et al., 2004; Montanari and Di Baldassarre, 2013; Refsgaard et al., 2006). However, a simple comparison of model results (and/or parameters) due to modified code and/or inclusion of new data sets also enables insights on the effect of changes and is often incorporated as model experiments (Lo et al., 2010; Nijssen et al., 2001; Oudin et al., 2005).

Due to the simplified representation of physical or empirical relationships between model components (e.g., the outflow of a water storage) in hydrological models, parameters are needed which mimic the natural process to be represented. In most cases, those parameters cannot be measured, and optimal values can be set in a calibration procedure, e.g., by optimizing objective functions with the assistance of efficiency metrics (Beven, 2001). However, in a classical way and in contrast to most catchment models this kind of calibration is seldom done in GHMs, with WaterGAP being one exception (Döll et al., 2003). This GHM is currently used as the basis for GRACE data assimilation and multi-parameter calibration approaches (Eicker et al., 2014; Schumacher et al., 2016a). However, and especially if more than one parameter is to be calibrated, equifinality problems (more than one parameter set leads to successful objective function) can occur (Beven, 2006). As parameter uncertainty (and sensitivity) is discussed in earlier work with WaterGAP (Kaspar, 2003; Schumacher et al., 2016b; Werth and Güntner, 2010), this topic is touched upon only briefly in this thesis.

1.4 Water – Global Assessment and Prognosis

This section gives a brief overview of the Water – Global Assessment and Prognosis (WaterGAP) model. A thorough description of the model and its calibration approach is given in the appendix of Müller Schmied et al. (2014).

In the late 1990s, the global freshwater and water use model WaterGAP was developed at the Center for Environmental Systems Research (CESR, University of Kassel, Germany) under supervision of Joseph Alcamo and Petra Döll. The major objective of developing this new tool was to assess water resources including human water use for historical and scenario conditions. To allow for a robust estimation of renewable freshwater resources in particular, a basin-wide calibration routine was developed that aims to fit simulated discharge to long-term annual observed river discharge, which even now is unique in global-scale hydrological modeling.

WaterGAP calculates water storages and fluxes for global land area (except Antarctica) at a 0.5° x 0.5° spatial resolution. Temporal resolution of the calculations is one day, whereas

model outputs are analyzed predominantly at monthly time steps. WaterGAP (Fig. 1a) consists of the water use models, which are split into five water use sectors (irrigation, domestic, cooling of thermal power plants, manufacturing and livestock). In version 2.1h (see Table 1), the GroundWater Surface Water USE (GWSWUSE) sub model was introduced, which calculates the net abstractions from surface water and those from groundwater. Within the WaterGAP Global Hydrology Model (WGHM, Fig. 1b), the vertical and horizontal water balance is calculated for each of the roughly 70000 grid cells, and river discharge is transported through the drainage network DDM30 (Döll and Lehner, 2002). All calculations of vertical water balance are in units of water column (calculating canopy, snow and soil storage), whereas flow rates are used for calculating the flows in the horizontal water balance (e.g., from the groundwater, surface water bodies and river storage).

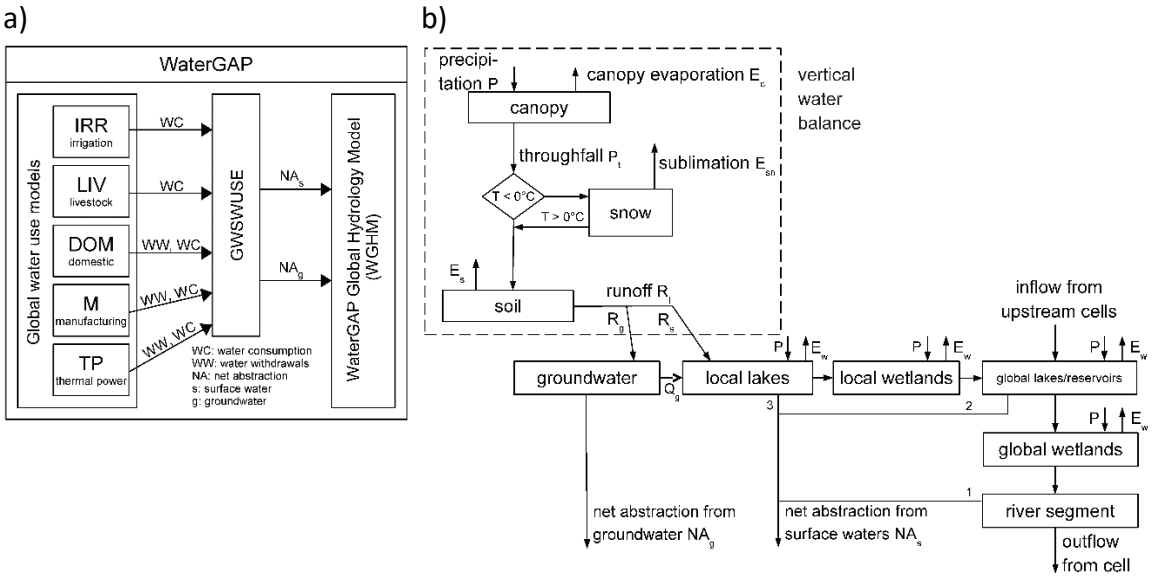


Figure 1: Scheme of the global freshwater model Water – Global Assessment and Prognosis (WaterGAP) with the water use models, the GroundWater Surface Water USE (GWSWUSE) submodel and WaterGAP Global Hydrological Model (WGHM) (a) and scheme of WGHM where boxes represent water storages and arrows represent related flows (b). Both figures are taken from Müller Schmied et al. (2014), their Fig. 1 and A1.

The irrigation model as well as WGHM is driven by meteorological input data with a daily or monthly time step and using precipitation, temperature, shortwave and longwave downward radiation (or cloud cover) as minimum required variables. Optionally, wind speed, specific humidity and air pressure can be read in, e.g., for more complex PET approaches. Other spatially distributed input data (e.g., soil texture, land cover) are used to parameterize the model equations and build the basis for calculation of storages (e.g., extent of lake and wetlands and corresponding storage capacities). Since 2003, WaterGAP has also been developed at the Goethe University Frankfurt. Parts of model development, and especially the integration of GRACE total water storage for calibration (and data assimilation), were/are done at GeoForschungsZentrum (GFZ) Potsdam as well as the University of Bonn.

Since the initial development of WaterGAP, many efforts have been made in order to implement new hydrological processes as well as spatially distributed input data to improve the model and its assessment of water resources. Starting from the first version after WaterGAP version 2.1d (Alcamo et al., 2003; Döll et al., 2003), Table 1 lists the modifications of the model versions compared to 2.1d. This table aims to provide an overview of the history of model developments as well as to indicate the role of the PhD candidate in model development, represented in percentage numbers of contribution. Note that this overview is not exhaustive as intermediate model versions were developed at CESR and GFZ which are not indicated with a separate version number. In a joint effort by CESR and the Goethe University Frankfurt, WaterGAP 2.2 was developed, which includes the main features of both developing lines, providing the basis for future joint model improvements.

Table 1: List of modifications for the specific WaterGAP versions (V) for water use models as well as the WaterGAP Global Hydrology Model (WGHM) and indication of the role of the PhD candidate, represented as percentage of contribution for the specific modification in brackets.

V	Modifications, references and proportion of PhD candidate's contribution (before version 2.2: 0%)	Reference R / application A
2.1e	Water use - Distinguishing water for cooling thermal power plants and manufacturing water use - Updates for irrigated areas for Oceania and Africa - New map of livestock types with improved spatial resolution WGHM - Update of land cover information from IMAGE 2.1 to IMAGE 2.2	R: internal documentation A: Schulze and Döll, 2004; Schuol et al., 2008
2.1f	Water use - Update of water use modules and accompanying input data WGHM - Dynamic allocation of a neighboring cell with the highest water storage in rivers/lakes for water abstractions - 64 additional reservoirs - Snow routine now based on 100 subgrids to improve snow accumulation and snow melt dynamics - Integration of a reduction factor for evaporation for lakes and wetlands that improves model simulation when water storage is at a low level - Enhanced number of calibration stations to 1235 stations	R: Döll and Fiedler, 2008; Hunger and Döll, 2008; Schulze and Döll, 2004 A: Chen et al., 2016; Döll, 2009; Fiedler and Döll, 2007; Güntner et al., 2007; Polcher et al., 2011; Ramillien et al., 2015; Richts et al., 2011
2.1g	WGHM - New precipitation correction approach - Improved groundwater recharge algorithm for semi-arid and arid regions - New reservoir management algorithm after Hanasaki et al. (2006) - Reservoirs and lakes were updated to a (preliminary) version of the GRanD database (Lehner et al., 2011) and quality-controlled	R: Döll et al., 2009 A: Döll et al., 2009; Döll and Zhang, 2010

V	Modifications, references and proportion of PhD candidate's contribution (before version 2.2: 0%)	Reference R / application A
2.1h	Water use - Source of water withdrawals (groundwater or surface water) is distinguished, and in WGHM not all water withdrawals are taken from surface waters any longer - Newly introduced sub model GroundWater Surface Water USE (GWSWUSE) computes net abstractions from surface water as well as from groundwater, including time-constant sectoral fractions of groundwater use and fraction of return flows to groundwater and surface water	R: Döll et al., 2012 A: Döll et al., 2012; Huang et al., 2012
2.2	Water use - Updates for sectoral water use (Flörke et al., 2012) (0%) - GWSWUSE with adaptations (fraction of return flows recharges groundwater now between values of 0.2 and 0.95; irrigation efficiency for groundwater-fed irrigation assumed to be globally 0.7) (0%) WGHM - Possibility to read in daily meteorological input (0%) - New land cover input map based on MODIS satellite (30%) - New attributes for land cover classes (albedo (0%), emissivity (100%)) - Refined leaf area index (LAI) calculation (no longer based not on PET) (0%) - Growing period starts when temperature exceeds 8°C and precipitation sum exceeds 40 mm (0%) - Variable flow velocity resulting in better representation of day-to-day characteristics (Schulze et al., 2005; Verzano et al., 2012) (0%) - Calculation of permafrost by using the frost number (Aus Der Beek and Teichert, 2008) (0%) - Second land-ocean mask possible (transferring of input data 20%) - Introduction of several new PET methods (Weiß and Menzel, 2008) (0%) - Optional use of Kc values for FAO-56 calculation of PET (0%) - Several new model outputs, e.g., daily water storages (100%) - Possibility to read in MODIS LAI and MODIS albedo (100%) - Calibration now with 1319 discharge stations (90%) - Efforts to close water balance, e.g., through assigning the Caspian Sea as ocean (was modeled as lake before) and neglecting those grid cells from calculation (100%) - Predecessor of final 2.2 is adapted to run in a calibration/data assimilation framework, e.g., by reading in a parameter matrix (30%)	R: Müller Schmied et al., 2014 A: Berger et al., 2013; Boulay et al., 2015; Buma et al., 2016; Döll et al., 2014b; Eicker et al., 2014; Knieper and Pahl-Wostl, 2016; Müller Schmied et al., 2014; Riedel and Döll, 2016; Schumacher et al., 2016b; Siebert et al., 2015

V	Modifications, references and proportion of PhD candidate's contribution (before version 2.2: 0%)	Reference R / application A
2.2a	WGHM - Groundwater recharge below surface water bodies is enabled in semi-arid and arid regions (80%) - Dynamic land area fractions as consequence of dynamic surface water extents (80%) - Precipitation input on surface water bodies is now also multiplied with the evaporation reduction factor (as evaporation) to keep water balance consistent (60%) - Modified routing approach where water is routed through the storages dependent upon the fraction of surface water bodies; otherwise water is routed directly into the river (70%) - Initial development of river reach evaporation (not used in standard version) (90%) - Integration of depression storage evaporation (not used in standard version) (10%) - Possibility to read in radiation components from reanalysis (not included in standard version) (100%) - Possibility to include output of HYOGA2 glacier model to represent glacier dynamics (not included in standard version) (90%) - New output options, e.g., possibility to save grid-cell-specific text files with daily states and fluxes (100%)	R: Döll et al., 2014b A: Döll et al., 2014a; Khandu et al., 2016; Müller Schmied et al., 2016a
2.2 (ISIMIP2a)	Water use - For irrigation, 70% of consumptive water use is applied (deficit irrigation) (0%) WGHM - Update of reservoir information, including year when reservoir began operation (30%) - Including reservoir operation years (100%) * - Parallel calibration routine on 23 nodes (100%)	R: Müller Schmied et al., 2016c A: Müller Schmied et al., 2016b, 2016c

V	Modifications, references and proportion of PhD candidate's contribution (before version 2.2: 0%)	Reference R / application A
2.2b	Water use - Deficit irrigation with 70% of optimal irrigation was applied in grid cells, which were selected based on Döll et al. (2014a) and have 1) groundwater depletion of > 5 mm yr^{-1} over 1989–2009 and 2) a >5% fraction of mean annual irrigation water withdrawals in total water withdrawals over 1989–2009 (10%) WGHM - New total water capacity input based on Batjes (2012) (100%) ** - For global lakes and reservoirs (where the water balance is calculated in the outflow cell), water demand of all riparian cells is included in the water balance of the outflow cell and thus can be satisfied by global lake or reservoir storage (0%) - All water storage equations in horizontal water balance are solved analytically (except for local lakes). Those equations now include net abstractions from surface water or groundwater. As a consequence, sequence of net abstractions has been changed to 1) global lakes or reservoirs, 2) rivers, 3) local lakes (0%) - Net cell runoff is strictly the difference between the outflow of a cell and inflow from upstream cells at the end of a time step (0%) - Area correction factor (CFA) is included in water balance of lakes and wetlands (as it was before 2.2) (80%) - In 2.2, local and global lake storage could vary between the maximum storage $S_{ma}x$ and zero. In 2.2b (as in versions before 2.2), local and global lake storage can drop to –Smax as described in Hunger and Döll (2008). The area reduction factor (corresponding to the evaporation reduction factor in Hunger and Döll (2008), their eq. 1) has been changed accordingly (denominator: 2 x S_{max}). If lake storage S equals S_{max}, the rediction factor is 1; if S equals $–S_{max}$, the reduction factor is 0 (0%) - Modified calibration routine: an uncertainty of 10% of long-term average river discharge is allowed (following Coxon et al., 2015), meaning that calibration runs in four steps: 1) test if γ alone is enough to calibrate to ±1% of observed value; 2) test if γ alone is enough to calibrate when 10% uncertainty of observed values are allowed; 3) adapt observed value by 10%, and test if γ plus CFA are sufficient for calibration; 4) add station correction factor (CFS) if all other steps were not successful, and set CFS values to 1 if between 0.98 and 1.02 (90%) *** - All model parameters which are potentially used for the calibration/data assimilation integration (including also multiplicators) are now read from a text file in Javascript Object Notation (JSON) format (10%) - Regional changes based on Döll et al. (2014a): 1) for Mississippi Embayment Regional Aquifer, groundwater recharge was overestimated, and thus the fraction of runoff from land recharging groundwater was reduced from 80–90% to 10% in these cells; 2) groundwater depletion in the North China Plain was overestimated by a factor of 4, and thus runoff coefficient γ was reduced from 3– 5 to 1 in this area; 3) all wetlands in Bangladesh were removed since diffuse groundwater recharge was unrealistically low (30%) - Reducing water balance error to a global sum of $<1*10^{-4}$ km^3 yr^{-1} (40%)	R: this dissertation

* Effect is described in Appendix A1,** effect is described in Appendix A2,*** effect is described in Appendix A3

1.5 Objectives and thesis outline

As outlined in Chapters 1.1 to 1.4, there are still challenges left when quantifying water balance components at the global scale. The focus of this thesis is to evaluate, improve and apply the modeling of water resources using the Water Global Assessment and Prognosis (WaterGAP) model. The dissertation is structured in three research topics that reflect the thesis title and cover 13 research questions (RQs). Within the first part, divided into two subfields, WaterGAP components are evaluated against independent observation data, and model outputs driven with alternative climate input data are analyzed (Chapter 1.5.1). The second part (which includes the Appendices A1–3) addresses the effects of model modifications on freshwater fluxes and storages as well as on model parameters (Chapter 1.5.2). Three WaterGAP applications are the focus of the third part of this thesis (Chapter 1.5.3).

1.5.1 Model evaluation

The evaluation of WaterGAP is split into two parts. The first part compares the simulated and observed radiation, while the second part focuses on the sensitivity of freshwater fluxes and storages to alternative climate input data.

1.5.1.1 Evaluation of net radiation and its components with station-based observations

Net radiation is the most important variable for many potential evapotranspiration (PET) equations and therefore requires a plausible calculation. As PET provides the upper limit of simulated actual evapotranspiration, an evaluation regarding the uncertainties related to obtaining radiation information from climate input data as well as modelling upward components with WaterGAP is needed. Within paper 1, the following research questions will be answered:

RQ 1 **How is the performance of simulated upward radiation components of WaterGAP compared to observations?**

State-of-the-art climate input data provide downward radiation components, so that upward radiation has to be simulated by WaterGAP. This is done by including land cover-dependent attributes like albedo and emissivity that determines in conjunction with climate input variables the upward radiation components. It is of interest to evaluate the performance of upward radiation components with observations as well as with the ERA-Interim reanalysis, where substantial effort were has been put into closing the energy balance and constraining the radiation components as much as possible with observations.

RQ 2 **Is it beneficial to substitute standard climate input data and calculation of upward radiation components by ERA-Interim reanalysis?**

Reanalysis approaches comprise modeling approaches (e.g., of a meteorological forecast model) with data assimilation techniques that allow for the integration of various observations (satellites, radiosondes, station measurements). Hence, reanalysis are currently the main resource for creating climate input data that are used in GHMs like WaterGAP. Traditionally, such climate input data provide downward radiation components for consistent land-cover-type calculations within the GHM. However, a characteristic of most GHMs is that no energy balance is computed (in contrast to the water balance that is closed). Reanalysis has a closed energy balance, and upward radiation fluxes that are augmented by observations make the usage of upward fluxes from reanalyses interesting. Here, it is to be analyzed whether it is beneficial to include all radiation fluxes from ERA-Interim reanalysis in WaterGAP. The question is answered by comparing radiation fluxes to station observations and to WaterGAP estimates as well as discussing the consequences for PET.

RQ 3 **What is the likely uncertainty of net radiation and its components?**

In order to enhance knowledge about the uncertainty of simulated radiation components, the mean absolute error efficiency metric from paper 1 is used to quantify the likely uncertainty of the simulations from the observations.

1.5.1.2 Evaluation of the sensitivity of simulated water balance components to climate input data

The main spatiotemporal driver of WaterGAP is the climate input data, and uncertainties of these forcing data can be large (see Chapter 1). It is thus of great interest to evaluate how the uncertainty of climate input data translates to model outputs. The sensitivity of simulated water balance components to climate input data is studied in papers 2, 3 and 4. Furthermore, it is evaluated how the calibration approach of WaterGAP affects this kind of uncertainty.

RQ 4 **What is the sensitivity of simulated water balance components, including human water use, to state-of-the-art climate input?**

When WaterGAP is forced with alternative climate input data, substantial differences occur in water flows and storages. Three papers (2–4) cover the sensitivity of simulated water balance components to climate input data. Whereas paper 2 covers two alternative forcings in relation to other kinds of uncertainty, papers 3 and 4 evaluate the effect of four state-of-the-art climate

input data plus one homogenized data set. The effect of climate input data on quantifying human water use is the focus of paper 4.

RQ 5 How does the WaterGAP calibration approach affect the uncertainty of climate input data?

The effect of calibration on simulated water flows and storages is demonstrated in paper 2. In paper 3, and for human water use in paper 4, the effect of WaterGAP calibration against long-term average annual river discharge to the climate-input-data-induced uncertainty is investigated.

1.5.2 Model modification

Models always contain simplified representations of (natural) processes, but, e.g., the availability of new input data sets allows for the improvement of processes or integration of new ones. However, in most cases the model complexity increases and new parameters might be introduced, which both can lead to a higher degree of uncertainty. Therefore, a structured assessment of those modifications is crucial. During the PhD, several modifications have been made to the WaterGAP model in order to improve the spatio-temporal database and the representation of processes (see Table 1). With a focus on papers 2 and 5 as well as on Appendices A1–3, some effects of those modifications on model behavior, parameters and water flows and storages are highlighted.

RQ 6 What is the effect of WaterGAP development on water flows and storages between version 2.1d and version 2.2?

One of the main steps in WaterGAP development was the creation of model version 2.2, which contains numerous new features and data sets and is based on two previous model versions from the two WaterGAP groups. In a model experiment described in paper 2, the effect of one model structure that is comparable to the initial WaterGAP structure (version 2.1d) as well as version 2.2 is investigated in terms of effect of hydrological model structure on simulated freshwater fluxes and storages.

RQ 7 What is the impact of an alternative land cover map on simulated water flows and storages?

In WaterGAP, land cover input data are important for deriving, e.g., total soil water capacity and attributes for radiation calculation. Whereas the historical source of land cover information is map information, a newly derived satellite-based land cover map was introduced in WaterGAP 2.2. Within this question and as part of paper 2, it is of interest to assess the resulting differences in simulated water fluxes and storages due to this modification.

RQ 8 How does the integration of reservoir commissioning years affect the model results?

The representation of human water use and reservoir management is a feature of many GHMs. Since generic reservoir management algorithms were developed, WaterGAP includes reservoirs (and regulated lakes) in the model scheme, although commissioning years were not considered. In model version WaterGAP 2.2 (ISIMIP2a), reservoir commissioning years were included to improve the simulated discharge time series. The effect of considering the year in which dam operation started is assessed in Appendix A1 by comparing model outputs with and without commissioning years.

RQ 9 What is the effect of an updated spatial input map of total available soil water capacity?

An essential part of the WaterGAP model scheme (and of many other GHMs) is the characterization of the soil layer(s), as this determines the amount of maximum stored water in the soil as well as the amount of water that flows out of this storage compartment either as runoff or groundwater recharge or is evapotranspired. WaterGAP includes an observation-based map of total available soil water capacity for the first meter and scales that with land-cover-dependent rooting depth to obtain the maximum soil water storage. However, the soil databases greatly improved both with respect to the number of soil profiles included as well as to the soil physical characteristics. Within Appendix A2, an updated version of the total available soil water capacity applied in WaterGAP 2.2b is described, and its effects on simulated model output are assessed.

RQ 10 How does river discharge measurement uncertainty influence calibration parameters and simulated freshwater fluxes?

To date, WaterGAP is the only GHM which includes a basin-specific calibration scheme to adjust simulated long-term annual river discharge to observed ones. However, that scheme does not consider any uncertainties in discharge observations, which are shown to be significant even for high quality measurements. Within WaterGAP 2.2b, the calibration scheme was modified by assuming a potential 10% uncertainty in river discharge observations. This model modification is described in Appendix A3, together with a thorough discussion of the effect on calibration parameters as well as model outputs.

1.5.3 Model application

WaterGAP was developed as a tool to assess the freshwater fluxes and storages for both historical and future scenario conditions. The model outputs are included in a wide range of applications (Table 1) in the past. Three examples of WaterGAP application are presented in the third part of the thesis to elucidate the potential of that modeling framework.

RQ 11 What is the current best estimate of global and continental freshwater fluxes and storages with WaterGAP?

Within papers 2-4, as well as using results of the most recent model version 2.2b (Appendix A3), water balance estimates (by component) are presented. Within this research question, the attempt is to present the current best estimate of freshwater fluxes and storages at the global and continental scale by using WaterGAP 2.2b along with a discussion of other values reported in the literature.

RQ 12 How can groundwater depletion at the global scale be assessed using multiple sources of observations jointly with hydrological modeling?

Many regions worldwide where water demand for agriculture is satisfied by groundwater resources and withdrawal exceeds groundwater recharge, suffer from groundwater depletion. Hydrological modeling can assist in monitoring groundwater depletion especially in regions where observations, e.g., from wells, are sparse. Within paper 5, groundwater depletion is investigated at the global scale by using the WaterGAP model, GRACE satellite data of total water storage changes and information from observations. Both the methodology and the assessment are presented in this paper.

RQ 13 What is the impact of climate change on river flow regimes?

One of the aims of developing the WaterGAP model was to quantify effects of global (climate) change on freshwater resources. Based on a model experiment with bias-corrected climate input data from two general circulation models (GCMs) and two socio-economic development scenarios as well as newly developed indicators, an assessment of river flow regime changes with climate change is given in paper 6.

The following chapters include the contributing papers, which are referred to as paper 1 to 6 in Table 2.

Table 2: List of papers (chapters) and appendices that are included in this dissertation together with research questions (RQs) to which they mainly refer.

Paper # / chapter	Citation	RQ
1/2	Müller Schmied, H., Müller, R., Sanchez-Lorenzo, A., Ahrens, B. and Wild, M.: Evaluation of radiation components in a global freshwater model with station-based observations, Water, 8(10), 450, doi:10.3390/w8100450, 2016.	1, 2, 3
2/3	Müller Schmied, H., Eisner, S., Franz, D., Wattenbach, M., Portmann, F. T., Flörke, M. and Döll, P.: Sensitivity of simulated global-scale freshwater fluxes and storages to input data, hydrological model structure, human water use and calibration, Hydrol. Earth Syst. Sci., 18, 3511–3538, doi:10.5194/hess-18-3511-2014, 2014.	4, 5, 6, 7, 11
3/4	Müller Schmied, H., Adam, L., Eisner, S., Fink, G., Flörke, M., Kim, H., Oki, T., Portmann, F. T., Reinecke, R., Riedel, C., Song, Q., Zhang, J. and Döll, P.: Variations of global and continental water balance components as impacted by climate forcing uncertainty and human water use, Hydrol. Earth Syst. Sci., 20, 2877-2898, doi:10.5194/hess-20-2877-2016, 2016.	4, 5, 11
4/5	Müller Schmied, H., Adam, L., Eisner, S., Fink, G., Flörke, M., Kim, H., Oki, T., Portmann, F. T., Reinecke, R., Riedel, C., Song, Q., Zhang, J. and Döll, P.: Impact of climate forcing uncertainty and human water use on global and continental water balance components, Proc. Int. Assoc. Hydrol. Sci., 374(7), 53–62, doi:10.5194/piahs-374-53-2016, 2016.	4, 5, 11
5/6	Döll, P., Müller Schmied, H., Schuh, C., Portmann, F. T. and Eicker, A.: Global-scale assessment of groundwater depletion and related groundwater abstractions: Combining hydrological modeling with information from well observations and GRACE satellites, Water Resour. Res., 50(7), 5698–5720, doi:10.1002/2014WR015595, 2014.	12
6/7	Döll, P. and Müller Schmied, H.: How is the impact of climate change on river flow regimes related to the impact on mean annual runoff? A global-scale analysis, Environ. Res. Lett., 7(1), 014037, doi:10.1088/1748-9326/7/1/014037, 2012.	13
A1	Integrating reservoir commissioning years	8
A2	Adapting input map for total soil water capacity	9
A3	A modified calibration approach to account for uncertainty in river discharge measurements	10

2 Evaluation of radiation components in a global freshwater model with station-based observations

Full citation: Müller Schmied, H., Müller, R., Sanchez-Lorenzo, A., Ahrens, B. and Wild, M.: Evaluation of radiation components in a global freshwater model with station-based observations, Water, 8(10), 450, doi:10.3390/w8100450, 2016.

Declaration of portion of own work

H. Müller Schmied designed the initial analysis setup after discussions with P. Döll. Data preparation, model code adaptation, model runs, analyses, tables, figures, and text were done by H. Müller Schmied. R. Müller, B. Ahrens, A. Sanchez-Lorenzo and M. Wild gave general suggestions regarding radiation comparison. Station data (GEBA) were provided by M. Wild. All co-authors helped revise the manuscript. In overall percentage, H. Müller Schmied contributed around 90% of this manuscript.

Hannes Müller Schmied

water

MDPI

Article

Evaluation of Radiation Components in a Global Freshwater Model with Station-Based Observations

Hannes Müller Schmied [1,2,*], Richard Müller [3], Arturo Sanchez-Lorenzo [4], Bodo Ahrens [5] and Martin Wild [6]

1 Institute of Physical Geography, Goethe-University, Frankfurt D-60438, Germany
2 Senckenberg Biodiversity and Climate Research Centre (BiK-F), Frankfurt D-60325, Germany
3 Center of Research and Development, German Meteorological Service, Offenbach D-63067, Germany; richard.mueller@dwd.de
4 Pyrenean Institute of Ecology, Spanish National Research Council, Zaragoza ES-50059, Spain; arturo.sanchez@ipe.csic.es
5 Institute for Atmospheric and Environmental Sciences, Goethe-University, Frankfurt D-60438, Germany; bodo.ahrens@iau.uni-frankfurt.de
6 Institute for Atmospheric and Climate Science, ETH Zurich, Zurich CH-8092, Switzerland; martin.wild@env.ethz.ch
* Correspondence: hannes.mueller.schmied@em.uni-frankfurt.de; Tel.: +49-069-7984-0216

Academic Editor: Karl-Erich Lindenschmidt
Received: 3 July 2016; Accepted: 11 October 2016; Published: 14 October 2016

Abstract: In many hydrological models, the amount of evapotranspired water is calculated using the potential evapotranspiration (PET) approach. The main driver of several PET approaches is net radiation, whose downward components are usually obtained from meteorological input data, whereas the upward components are calculated by the model itself. Thus, uncertainties can be large due to both the input data and model assumptions. In this study, we compare the radiation components of the WaterGAP Global Hydrology Model, driven by two meteorological input datasets and two radiation setups from ERA-Interim reanalysis. We assess the performance with respect to monthly observations provided by the Baseline Surface Radiation Network (BSRN) and the Global Energy Balance Archive (GEBA). The assessment is done for the global land area and specifically for energy/water limited regions. The results indicate that there is no optimal radiation input throughout the model variants, but standard meteorological input datasets perform better than those directly obtained by ERA-Interim reanalysis for the key variable net radiation. The low number of observations for some radiation components, as well as the scale mismatch between station observations and 0.5° × 0.5° grid cell size, limits the assessment.

Keywords: global hydrological modeling; radiation; validation with station observations; BSRN; GEBA; uncertainty

1. Introduction

The estimation of the Earth's surface radiation components are of high interest in climate science (e.g., for global radiation budget [1–3]) and as a driver for evaporation of water (e.g., [4]). Consequently, global hydrological models (GHMs), which are designed to simulate water fluxes and storages on terrestrial land surface, incorporate radiation information in their calculation of potential evapotranspiration (PET, maximum amount of evapotranspiration if no water limitation occurs) and subsequently actual evapotranspiration (AET, taking into account possible water availability limitations) [5,6].

Until the early 2000s, meteorological input data (hereafter referred to as meteorological forcings) for GHMs solely contained indirect information about radiation (e.g., sunshine duration or cloud

cover). Nowadays, meteorological input data for GHMs are based on reanalysis data, which are either driven by observations including satellite data (as in the case of radiation [7]) or sometimes supplemented (corrected) with ground or satellite based observations. A number of such forcings are now available that include downward radiation information, e.g., WATCH Forcing Data methodology applied to ERA-Interim (WFDEI, [8]) and Princeton meteorological forcing (PGFv2, [9]). However, these forcings differ from each other due to different underlying reanalysis and correction approaches, which is reflected in uncertain estimates of the water balance components (e.g., [10,11]).

Although precipitation is one of the major drivers of uncertainty (e.g., [12]), uncertainties in (net) radiation also affect model outputs significantly. Döll et al. [13] varied simulated net radiation within the global freshwater model WaterGAP by 20% resulting in a 4% to 7% difference for global scale AET and a 6% to 10% difference for global scale river discharge, which consequently alters the spatial pattern of renewable water resources (see their Table 2 and Figure 1). In the study of Nasonova et al. [11], variations in global averaged shortwave (longwave) fluxes among their datasets are estimated to be about 14 (13) Wm^{-2} with subsequent influences to water fluxes. The CMIP5 models vary globally by 19 Wm^{-2} for both downward fluxes with a much higher variation on the global land area (41 Wm^{-2} for shortwave and 33 Wm^{-2} for longwave downward fluxes) [3]. Consequently, net radiation over land varies among the 43 CMIP5 models by 29 Wm^{-2} which is translated into a latent heat flux (evaporation) variation of 14 Wm^{-2} [3].

However, modifying simulated radiation with a fixed percentage or comparing different products are not appropriate to assess the quality (in terms of the agreement with in-situ observations) of simulated radiation components. There are a number of studies evaluating spatially distributed radiation datasets with station-based observations. For the first time, Wild et al. [14] used surface observations to assess shortwave downward radiation from reanalyses, and later also the longwave downward radiation from reanalyses and climate models [15]. Troy and Wood [16] compared seven globally available radiation products with 32 stations from the World Radiation Data Centre (WRDC) archive over northern Eurasia and found differences of 20 Wm^{-2} for net radiation. Furthermore, NASA/GEWEX Surface Radiation Budget (SRB) data was assessed to be the closest dataset compared to the observations. In other recent studies, data products were evaluated using station observations (e.g., [17–21]) but were mostly focused on regional scales. Overall, to assess the global-scale coverage of a GHM, mainly two radiation-relevant station datasets are available to the scientific community: the Baseline Surface Radiation Network (BSRN, [22]) and the Global Energy Balance Archive (GEBA, [23]).

Heinemann and Kerschgens [24] investigated the representativeness of station measurements in a heterogeneous landscape in northern Germany to simulated surface energy fluxes with a 250 m and 1 km resolution. They concluded that large uncertainties exist which are also dependent on the aggregation method. Furthermore, a strong dependence on land cover types was found. In addition, Horlacher et al. [25] noted reasonable differences between two meteorological stations (8.5 km apart) especially in the upward radiation fluxes that are mainly land-cover-dependent. In a study of Hakuba et al. [26], an error assessment for the representativeness of mean monthly point measurements compared to grid cell satellite-based datasets for solar radiation showed a relatively small mean error of about 3 to 4 Wm^{-2}. These issues indicate the problem of representativeness when comparing point measurements with grid cell boxes. Nevertheless, and due to the given heterogeneous (and not omnipresent) data coverage of radiation observations, such a comparison can still be of value for assessments of coarse scale gridded products [26].

To the authors' knowledge, there is a lack of studies investigating the reliability of radiation components used in GHMs, and especially their calculation of upward radiation fluxes. In most GHMs, upward (and net) radiation fluxes are calculated according to land surface parameters as land-cover-dependent albedo or emissivity (e.g., [27]). This approach is common and consistent as some other land-cover-dependent characteristics are used, e.g., for modeling snow dynamics and rooting depth [27]. However, using radiation-relevant characteristics for calculating upward components is subject to uncertainties. Therefore, it is of interest to evaluate the general performance

of state-of-the-art meteorological forcings in terms of radiation fluxes and if such an internal model calculation can be substituted by radiation components, e.g., from reanalysis.

Following the approach of Budyko [28], the Earth's land surface (represented here by $0.5^\circ \times 0.5^\circ$ grid cells) can be hydrologically characterized, dividing PET by precipitation (P). If this quotient is less than one, the region can be characterized as energy-limited, otherwise it is water-limited. From the hydrological perspective, changes in net radiation (which is the major driver of PET in most approaches) are interesting, especially in energy-limited areas as, for example, increasing amounts of net radiation are translated into higher amounts of PET and AET and subsequent less available water resources. Therefore, the optimal calculation of radiation components plays a vital role in GHMs.

The overall objective of this study is to evaluate radiation components of two state-of-the-art meteorological forcings and from ERA-Interim reanalysis by comparing the downward components as well as the net components with station-based observations of the BSRN and GEBA datasets. Furthermore, the upward and net radiation components as calculated by WaterGAP (except for the ERA-Interim variant where net radiation is used) are evaluated. The general aim is to find the most plausible approach to obtain net radiation for PET simulation. Within this, we pursue four objectives:

1. To assess the spatial differences of global-scale net radiation as well its components from selected meteorological forcings.

Using two meteorological forcings (WFDEI, and PGFv2) and two ERA-Interim reanalysis variants, we evaluate spatial patterns of the simulated (or provided ones in the case of the ERA-Interim variant where net radiation is taken as input data) radiation components.

2. To identify optimal downward radiation input from meteorological forcings.

The downward radiation components of the two meteorological forcings, as well as ERA-Interim reanalysis, are assessed with respect to their performance to station observations of BSRN and GEBA data on a monthly time step.

3. To evaluate the calculation of upward radiation fluxes within WaterGAP and net radiation components.

For the two meteorological forcings as well as the reanalysis variant where only downward radiation is taken into account, the simulated upward radiation that is estimated by WaterGAP is evaluated by comparing it to observations and the reanalysis variant that provides upward radiation components. In addition, net radiation fluxes are assessed.

4. To test if improvements are achieved when using net radiation from the ERA-Interim reanalysis directly.

Net radiation is the most important variable for PET simulation. Therefore, the performance of net radiation and consequences for PET is evaluated in order to investigate if ERA-Interim net radiation could be used for future assessments within global hydrological models.

In Section 2, the data sources, model and efficiency criteria are described. The results are presented and discussed in Section 3 both for the global scale, but also if there are spatial differences between geographic regions and energy- and water-limited areas. Concluding remarks are given in Section 4.

2. Data and Methods

In this section, we firstly describe the model experiment and the meteorological forcings used in this study, followed by the WaterGAP Global Hydrology Model (WGHM). The two in-situ radiation databases, as well as the efficiency metrics, are presented thereafter.

2.1. Experimental Setup

Four different radiation data sets were used within WaterGAP, version 2.2a in order to answer the research questions. Even though the model version is not modified, the term model variant is

used to distinguish the meteorological forcings with a five-letter abbreviation (WFDEI, PGFv2, ERAID, and ERAIN, see Table 1). The WFDEI variant (see Section 2.1.2), based on ERA-Interim reanalysis, represents the current standard meteorological forcing and thus standard radiation input for WaterGAP (see STANDARD model variant in [27]). The PGFv2 forcing (see Section 2.1.3) is used to include an alternative forcing which is very frequently applied in global scale modeling and is thus evaluated here as well. Directly interpolated shortwave downward (S↓) and longwave downward radiation (L↓) data from ERA-Interim reanalysis were taken for variant ERAID, which differs from WFDEI due to the interpolation, elevation and bias correction method. Finally, substituting the calculation of upward radiation components by the corresponding fluxes from the ERA-Interim reanalysis is done in model variant ERAIN. For both ERAID and ERAIN, WFDEI was used for P and temperature (T). For ERAIN, the choice of P and T is irrelevant as radiation components are not calculated within WaterGAP. P and T are important for all other variants, especially for the snow dynamics, which have an influence on albedo and thus shortwave upward radiation (S↑), whereas T (at 2 m height) is used to calculate longwave upward radiation (L↑). We decided to use WFDEI T and P for ERAID as: (1) a simple interpolation of T to the 0.5° grid without considering the environmental lapse rate would lead to uncertainties, especially in mountainous areas [29]; (2) a bias correction of P with observations is required as there is still a (slight) wet bias in ERA-Interim when compared to P observations [7]; and (3) in case P and T for ERAID is the same as for WFDEI, the emissivity value can be evaluated for longwave upward radiation. In the following sections, the meteorological forcings and their sources used for the intercomparison are described in more detail.

Table 1. Model variant names for the experiment and source for climate variables. Abbreviations: S↓: shortwave downward radiation, S↑: shortwave upward radiation, L↓: longwave downward radiation; L↑: longwave upward radiation; P: precipitation; T: temperature; WaterGAP: Water—Global Assessment and Prognosis; WFDEI: WATCH Forcing Data methodology applied to ERA-Interim reanalysis; PGFv2: Princeton Global Meteorological Forcing Dataset version 2. Name for downward ERA-Interim radiation components is ERAID, for the variant where also upward ERA-Interim radiation is used, ERAIN.

Name	S↓	S↑	L↓	L↑	P, T
WFDEI	WFDEI	WaterGAP	WFDEI	WaterGAP	WFDEI
ERAID	ERA-Interim	WaterGAP	ERA-Interim	WaterGAP	WFDEI
ERAIN	ERA-Interim	ERA-Interim	ERA-Interim	ERA-Interim	WFDEI
PGFv2	PGFv2	WaterGAP	PGFv2	WaterGAP	PGFv2

2.1.1. ERA-Interim

The European Centre for Medium-Range Weather Forecast (ECMWF) Re-Analysis Interim (ERA-Interim) is the third generation of global atmospheric reanalysis from ECMWF and spans from 1979 until recent time. A reanalysis includes a meteorological forecasting model which uses all available observation data (e.g., station measurements, radiosonde—profiles or satellite data) to initialize the next forecast step. This is done cyclically a few times per day. The model behind ERA-Interim is called Cy31r1, an atmospheric model and data assimilation system that contains the three components: atmosphere, land surface and ocean waves. The land surface model of ERA-Interim is TESSEL (Tiled ECMWF Scheme for Surface Exchanges over Land). Within the data assimilation scheme and using surface (and cloud) albedo and emissivity values, the radiation fluxes are generated in an integrative way using a radiation transfer scheme, with the main aim to preserve the energy balance (details in [30]). Compared to ERA-40 the energy balance on land surface is improved [7]. Spatial resolution is horizontal ~80 km (spectral T255 grid) and vertical with 60 layers [7]. Reanalyses are widely applied in climate monitoring and analysis [31,32].

For this study, S↓ and L↓ from ERA-Interim (model variant ERAID), as well as S↑ and L↑ (model variant ERAIN) were used. ERA-Interim data were interpolated bi-linearly to 0.5° resolution (without bias and altitude correction) and aggregated to daily values.

2.1.2. WATCH Forcing Data Methodology Applied to ERA-Interim Reanalysis

The basis of this meteorological forcing is the ERA-Interim reanalysis [7]. Weedon et al. [8] prepared the output of this reanalysis within the Integrated Project Water and Global Change (WATCH) for global hydrological models by applying an interpolation and bias-correction scheme which is explained briefly here. ERA-Interim T was obtained from the lowest atmospheric model level (10 m) and interpolated to the 0.5° resolution using an elevation based environmental lapse rate. Furthermore, T is corrected to Climatic Research Unit Time-Series [33] (CRU TS) version 3.1 (1979–2009) respectively version 3.21 (2010–2012) average T as well as average diurnal T range [8,29]. The handling of P is described by Weedon et al. [29]. This variable is adjusted to CRU TS3.1 number of wet days and monthly totals using either CRU TS3.1/TS3.21 or data from the Global Precipitation Climatology Centre (GPCC) v5/v6. For 1979–2009, the adjustment is based on GPCC v5, and from 2010 to 2012 based on CRU TS3.21.

L↓ from ERA-Interim reanalysis data was elevation corrected after interpolation to 0.5° resolution using a fixed relative humidity as well as changes in T, surface pressure and specific humidity. Weedon et al. [8] found no necessity for monthly bias correction. S↓ is not elevation corrected after interpolation. In contrast to the previous version of WATCH Forcing Data (WFD) [29], S↓ was adjusted for interannual (but not seasonal) variations in aerosol loading (using CRU TS3.1/3.21 average cloud cover) which results in higher mean monthly values compared to the WFD forcing [8]. Furthermore, the aerosol distribution was changed in the ERA-Interim reanalysis so that WFDEI has higher values of S↓ in northern Africa (~40 Wm^{-2} for the year 2000) and lower values of ~30 Wm^{-2} in northern South America (same year) [7]. WFDEI climate input is already available for the spatial and temporal resolution of WaterGAP. T and P from the WFDEI [8] were used in all variants except for PGFv2 to keep consistency within that meteorological forcing (Table 1).

2.1.3. Princeton Global Meteorological Forcing Dataset

The Princeton Global Meteorological Forcing Dataset, version 2 (PGFv2, http://hydrology.princeton.edu/data.pgf.php) is an updated version of the 60 year-forcing (1948–2008) described by Sheffield et al. [9] and is available between 1901 and 2012. This dataset blends reanalysis data (NCEP-NCAR) with station and satellite observations. Radiation (S↓, L↓) is adjusted for systematic biases at a monthly scale to a product from the University of Maryland (by Rachel Pinker) developed within the NASA MEaSUREs project. S↓ trends are corrected using CRU TS 3.21 cloud cover and for L↓, the year-to-year variation of NCEP-NCAR reanalysis is retained. T is bias corrected by shifting to monthly CRU TS 3.21. P is bias corrected using CRU TS 3.21 and not undercatch corrected (in contrast to the previous version described by Sheffield et al. [9]). All information on this PGFv2 version was provided by personal communication with J. Sheffield, 2015. PGFv2 climate input is already available for the spatial and temporal resolution of WaterGAP.

2.2. WaterGAP Global Hydrology Model (WGHM)

The WaterGAP Global Hydrological Model (WGHM, [34]) belongs to the Water—Global Assessment and Prognosis (WaterGAP) model [35] and calculates freshwater fluxes and storages on the global land surface (except Antarctica). WaterGAP was developed to assess water availability and water scarcity and was applied in a range of studies with historic meteorological forcings [10,27,36–39] and climate change [40–43] scenarios.

The spatial resolution of the model is 0.5° × 0.5° (55 × 55 km at the equator) and calculations are performed on a daily time step, whereas output is analyzed on a monthly scale. A thorough description of the model and its components for version WaterGAP 2.2 can be found in the appendix

of Müller Schmied et al. [27]. The version used here is named WaterGAP 2.2a and is described in Döll et al. [38]. The only relevant modification of this model version (compared to the description in Müller Schmied et al. [27]) used in this study is the possibility to read in net radiation components, e.g., from ERA-Interim reanalysis. In this case, WaterGAP does not influence radiation calculation and acts simply as a tool to calculate PET as well as to provide the output file format that is then used for the comparison with station observations.

In general, S↓ and L↓ are provided by the meteorological forcings and have the unit in Wm^{-2}. Net shortwave radiation S_{net} (Wm^{-2}) for all model variants except ERAIN (Table 1) is calculated as:

$$S_{net} = S\downarrow (1 - \alpha_{LC}) , \quad (1)$$

where α_{LC} is the albedo (-) based on land cover type LC ([27], their Table A2). Albedo values for WaterGAP are taken from assumptions of the IMAGE model [44]. In the case of a reasonable snow cover, the albedo value is varying dynamically in WaterGAP to represent the influence of snow cover dynamics on radiation balance [27]. S_{net} from ERAIN is directly used for the assessment.

Upward shortwave radiation S↑ (Wm^{-2}) is calculated as:

$$S\uparrow = S\downarrow - S_{net} . \quad (2)$$

Upward longwave radiation L↑ (Wm^{-2}) is calculated as:

$$L\uparrow = \varepsilon_{LC}\sigma T^4 , \quad (3)$$

where ε_{LC} is the emissivity (-) based on land cover type ([27], their Table A2), σ is the Stefan–Boltzmann constant (5.67×10^{-8} ($Wm^{-2}{\cdot}K^{-4}$)) and T is temperature (K). Emissivity values are taken from Wilber et al. [45] who assessed the emissivity from different materials (e.g., minerals) in laboratory experiments and then upscaled it to a land cover classification scheme using the predominant material composition of the land cover class.

Net longwave radiation L_{net} (Wm^{-2}) for all model variants except ERAIN is calculated as:

$$L_{net} = L\downarrow - L\uparrow . \quad (4)$$

L_{net} from ERN is directly used for the assessment.

Finally, net radiation R_{net} (Wm^{-2}) is calculated as:

$$R_{net} = S_{net} + L_{net}. \quad (5)$$

2.3. Radiation Validation Data

2.3.1. Baseline Surface Radiation Network (BSRN)

The Baseline Surface Radiation Network (BSRN) was an initiative from the World Climate Research Programme (WCRP) Radiative Fluxes Working Group and is now incorporated into the WCRP Global Energy and Water Cycle Experiment (GEWEX). The aim of BSRN is to collect and provide data from a high qualitative radiometric network. The number of stations is relatively low (Table 2) but shortwave and longwave surface radiation fluxes of "best possible quality currently available" are provided [22] (p. 6). BSRN data were downloaded via the data portal http://pangaea.de/ in 01/2015 as monthly means (calculated by the data warehouse tool) of all available radiation flux variables (if available). In cases where more than one sensor reported data values, we obtained the mean value. Some errors in the database (e.g., interchanged variables) were found, reported and are already being updated. Forty-three BSRN stations (Figure 1a) are located within the land-ocean mask of WaterGAP, providing data for at least 6 months within the analysis time (1992–2012) and could thus be used for this study (see Table 2 for a summary of number of stations/months).

Table 2. Number of stations and months of the radiation data used within this study. Stations were only considered if at least 6 months of data were available.

Variable	# Stations	# Stations (Calculated)	# Months	# Months (Calculated)
		BSRN		
S↓	43		5099	
S↑	16		2317	
S_{net}		16		2289
L↓	43		4872	
L↑	16		2210	
L_{net}		16		2205
R_{net}		16		2171
		GEBA		
S↓	1061		171,169	
S↑	43		3067	
S_{net}		40		2791
L↓	42		2933	
L↑	14		839	
L_{net}		13		692
R_{net}	142		11,525	

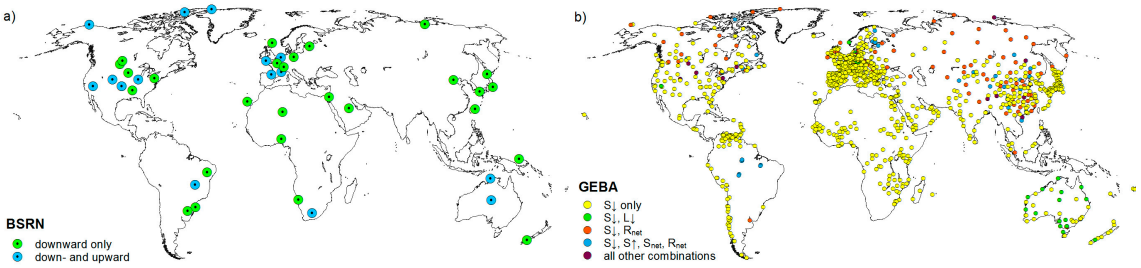


Figure 1. Location of the reference stations used for evaluating the radiation components: Baseline Surface Radiation Network (BSRN), (**a**) and Global Energy Balance Archive (GEBA); (**b**) A specific color for GEBA stations was assigned when the radiation variable(s) were available for at least ten stations, otherwise joined to "all other combinations".

2.3.2. Global Energy Balance Archive (GEBA)

In order to extend the geographical coverage of the evaluation sites, the Global Energy Balance Archive (GEBA) is used in addition to BSRN. GEBA contains long-time averaged surface radiation fluxes of more than 1500 stations worldwide. The data are quality controlled and coded with quality flags [23]. Station data were used if at least 6 monthly data values were available and quality flags indicated no problematic data (quality control procedure 1–4 have values > 4). Twenty GEBA stations were excluded as they provide monthly aggregates of BSRN stations. Figure 1b shows the location of the stations. The data were downloaded via http://www.geba.ethz.ch/ in August 2014. The coordinate was rounded to the next decimal point in case a GEBA station was located at the 0.5° grid cell border.

2.3.3. Calculation of Net Radiation and Its Components

Based on the monthly averages of the measurements, S_{net} (L_{net}) radiation was calculated as S↓ minus S↑ (L↓ minus L↑) and R_{net} as the sum of S_{net} and L_{net}.

2.4. Efficiency Metrics

Efficiency metrics are used to quantify the goodness-of-fit between observations and simulations. There are numerous metrics available (each with benefits and limitations) and have been subject to

reviews in the past e.g., [46,47]. Möller [48] analyzed WaterGAP model outputs with BSRN station data (from the four model variants, Table 1) by using ten efficiency metrics. She found that the Nash–Sutcliffe Efficiency and the Mean Absolute Error E_{MAE} are the best metrics to reflect the characteristics of the modeled radiation vs. the simulated ones. The advantages of the Kling–Gupta Efficiency E_{KG}, which builds on the Nash–Sutcliffe Efficiency, are the possibility to split this metric into the three components and to determine the dominant component [49]. In this study, E_{KG} and its components, as well as E_{MAE}, are used to assess the simulations with station observations.

2.4.1. Kling–Gupta Efficiency

Gupta et al. [49] and, later on, Kling et al. [50] decomposed the very popular Nash–Sutcliffe Efficiency metric into three components to allow diagnostic insights into the model performance and created the Kling–Gupta Efficiency metric by computing the Euclidian distance of the components from the ideal point (E_{KG}, Equation (6) in its version 2012 which is used here).

$$E_{KG} = 1 - \sqrt{(E_{KGr} - 1)^2 + (E_{KGbeta} - 1)^2 + (E_{KGgamma} - 1)^2} \tag{6}$$

where E_{KGr} is the correlation coefficient between simulated and observed values [-] and can act as indicator for the timing, E_{KGbeta} is the bias ratio (Equation (7)) [-] and can act as indicator if biases of the mean values occur and $E_{KGgamma}$ is the variability ratio (Equation (8)) [-] which can act as indicator for the variability of simulated (S) and observed (O) values.

$$E_{KGbeta} = \frac{\mu_S}{\mu_O}, \tag{7}$$

$$E_{KGgamma} = \frac{CV_S}{CV_O} = \frac{\sigma_S/\mu_S}{\sigma_O/\mu_O}, \tag{8}$$

where μ is the mean value (e.g., in Wm^{-2}), σ is the standard deviation of the value (e.g., in Wm^{-2}), CV is the coefficient of variation [-]. The optima of E_{KG} and its components are one. The lowest component determines the E_{KG} value. Furthermore, Gupta et al. [49] provided a methodology to assess the relative contribution g_i of the three components as:

$$g_i = \frac{G_i}{\sum_{j=1}^{3} G_j}, \tag{9}$$

with

$$G_1 = (E_{KGr} - 1)^2, \tag{10}$$

$$G_2 = (E_{KGgamma} - 1)^2, \tag{11}$$

$$G_3 = (E_{KGbeta} - 1)^2. \tag{12}$$

2.4.2. Mean Absolute Error

An absolute measure of model efficiency is the Mean Absolute Error E_{MAE}. It is calculated as the absolute difference between observed and simulated (monthly) data, which are then averaged over the number of observations (Equation (13)). A great advantage of this absolute efficiency criterion is the resulting number in the unit of measurement, which can then be used, for example, to assess absolute errors among the radiation components.

$$E_{MAE} = \frac{1}{n}\sum_{i=1}^{n} |O_i - S_i| \tag{13}$$

where O_i is the observed (monthly) value for time step i and S_i is the modeled (monthly) value of the variable.

2.4.3. Sum of Ranks

According to Möller [48], the model variants were ranked for the efficiency metrics, giving the value "1" for the model variant with the highest performance (maximum E_{KG} or minimum E_{MAE}, respectively) and incrementing values from the highest to lowest performing model variant. This is done for each station and then aggregated over the dataset. Finally, the model variant with the lowest rank can be seen as the best model variant amongst the data set and radiation component considered here.

3. Results and Discussion

In this section, we firstly compare the radiation components with observations from BSRN and GEBA datasets (Section 3.1). Then, we assess global averages and spatial differences of net radiation and its components (Section 3.2). Finally, we discuss the optimal radiation input and possible improvements in Section 3.3.

3.1. Comparison to Station Measurements

Monthly grid cell values generated from the different data sources (see Table 1) are compared with BSRN and GEBA ground measurements. The comparison results are evaluated and discussed using the efficiency metrics defined in Section 2.4.

Results are displayed as boxplots for the E_{KG} criterion (Figure 2) as well as single components of E_{KG} (Figures 3–5) and for E_{MAE} (Figure 6). The BSRN dataset is separated into one where all 43 stations are included (green color in the figures) and one with only the 16 stations where both components (upward, downward) are measured (blue color in the figures). Remarkable differences exist between the GEBA and BSRN comparison results for all radiation components (Figures 2 and 6). For the downward components and both efficiency criteria, the simulations fit best to the subset of 16 BSRN stations, followed by all BSRN stations. Generally, lower performance is achieved for the GEBA stations, which is expected due to the larger data uncertainty [51]. However, due to the larger spatial coverage of this data set (at least for some components), integration of GEBA data allows a more robust assessment.

Analyzing Figures 2 and 6, downward radiation components are represented with high performance by all model variants, indicated by the overall high value for E_{KG} as well as (compared to net radiation) low E_{MAE} of around 10 Wm^{-2} (slightly higher for S↓, lower for L↓). Except for PGFv2, S↓ has a mean bias (E_{KGbeta}) greater than one which relates to an overestimation compared to measurements and L↓ has a (slight) mean bias below one (Figure 3). This is in line with previous assessments [2,3]. L↑, as modeled by WaterGAP (all except ERAIN), has a lower performance (except E_{MAE} and GEBA) than with ERAIN. Two possible reasons are that: (a) surface temperature as the main factor for L↑ calculation is strongly constrained by observations; or (b) emissivity values are more realistic in ERAIN.

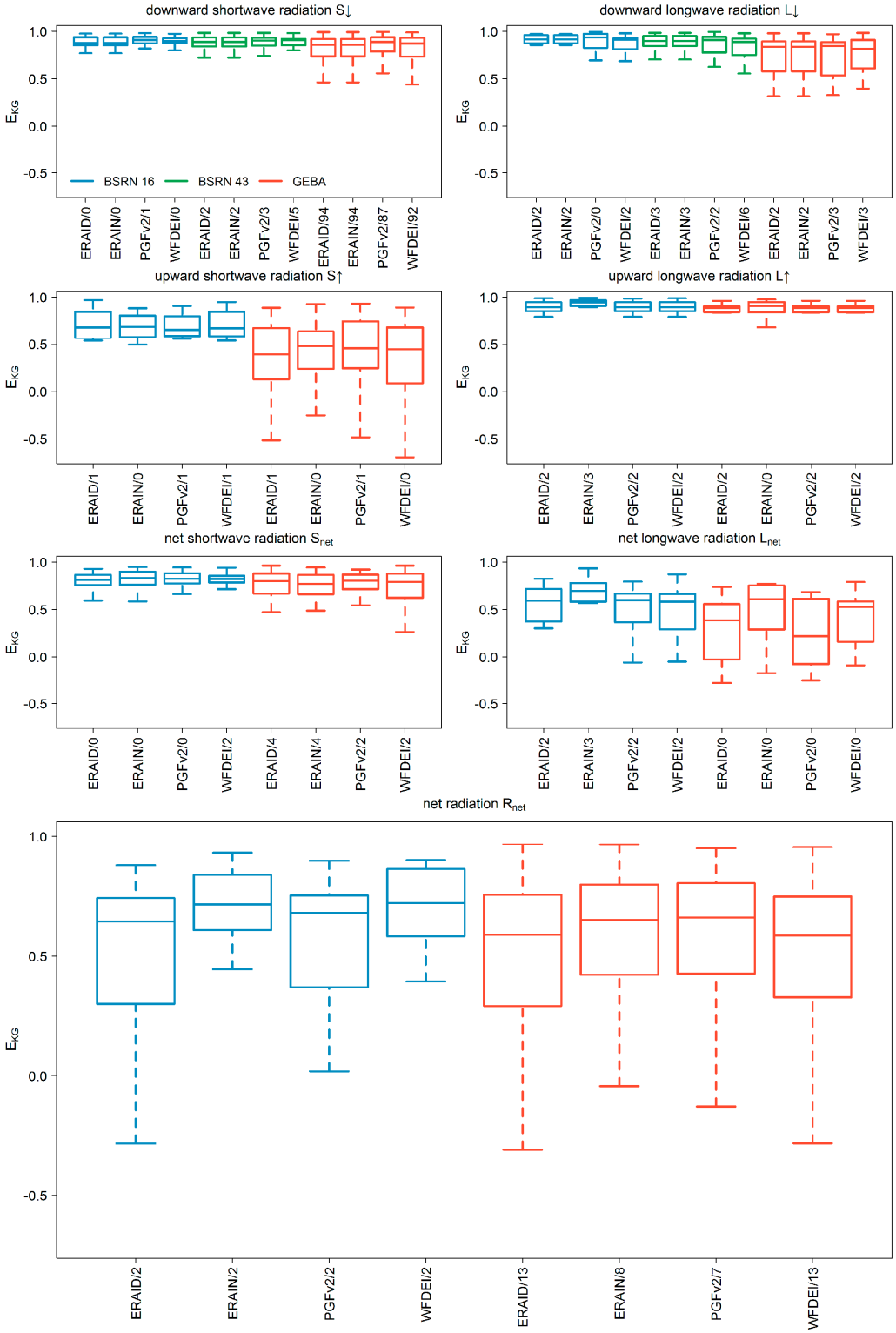


Figure 2. Validation of monthly radiation components with measurements of 16 BSRN stations (where all components are measured, blue color), 43 BSRN stations where only downward components were measured (green color) for the time period 1992–2012, as well as GEBA stations (red color) for the time period 1979–2012 using Kling–Gupta Efficiency E_{KG} (-). Numbers after the slash indicate the number of stations that are outside of the 1.5 × inter quartile range and thus not included in the boxplot. Significant differences of metric distribution (two sample Kolmogorov–Smirnov test, $p < 0.05$) are present for GEBA stations and shortwave downward radiation for all combinations except ERAID-ERAIN. For single components of E_{KG} , see Figures 3–5.

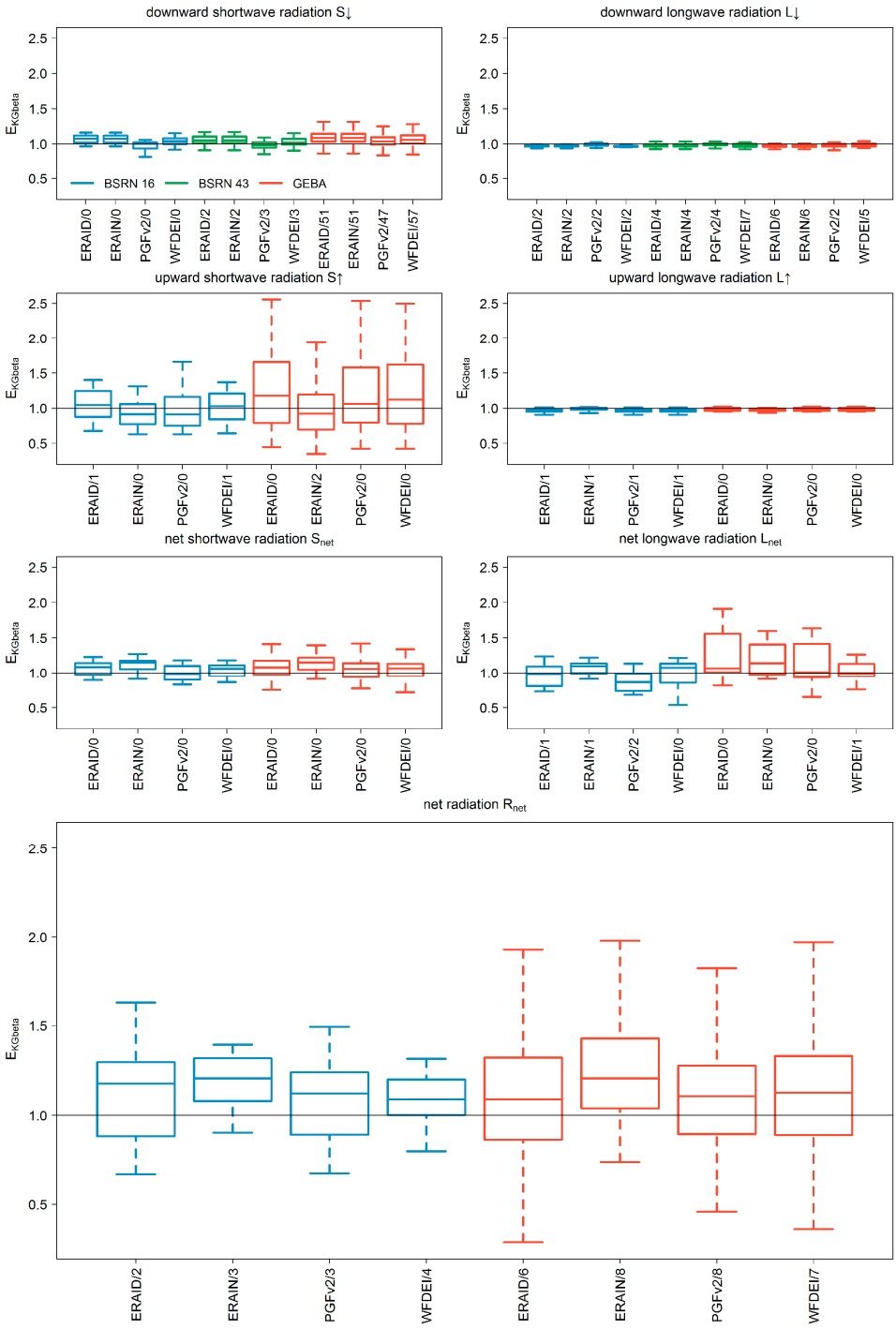


Figure 3. Validation of monthly radiation components with measurements of 16 BSRN stations (where all components are measured, blue color), 43 BSRN stations where only downward components were measured (green color) for the time period 1992–2012, as well as GEBA stations (red color) for the time period 1979–2012 using Kling–Gupta Efficiency component E_{KGbeta} (-) as indicator for mean bias. Numbers after the slash indicate the number of stations that are outside of the 1.5 × inter quartile range and thus not included in the boxplot. Significant differences of metric distribution (two sample Kolmogorov–Smirnov test, $p < 0.05$) are present for BSRN stations and the downward components for all combinations with PGFv2 and for the net shortwave as well as net longwave radiation for ERAIN-PGFv2. Within GEBA stations, significant differences occur for shortwave downward radiation between ERAIN/ERAID and PGFv2/WFDEI and also PGFv2-WFDEI; for net shortwave radiation between ERAIN-WFDEI and for net radiation between ERAID-ERAIN, ERAIN-PGFv2 and ERAIN-WFDEI.

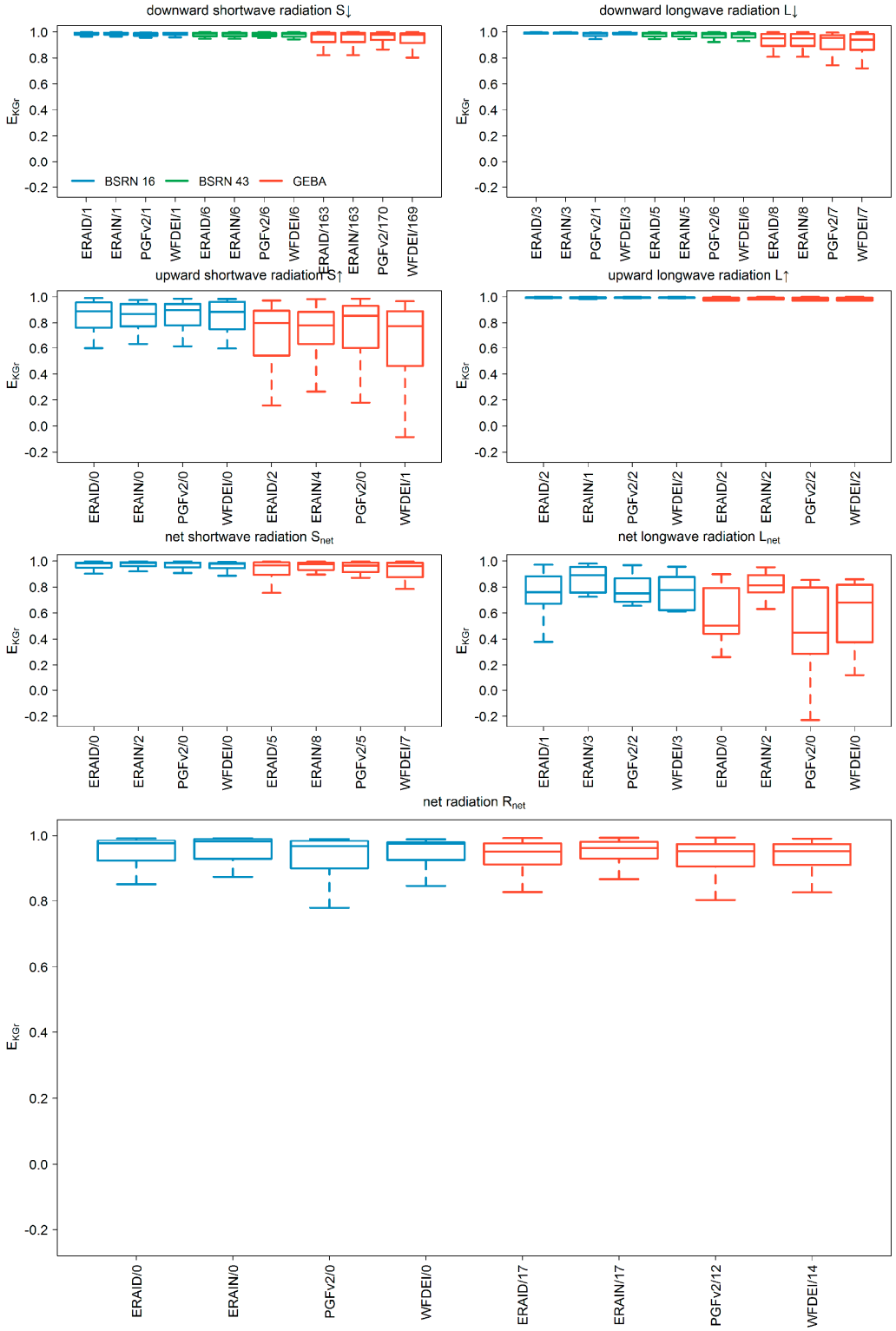


Figure 4. Validation of monthly radiation components with measurements of 16 BSRN stations (where all components are measured, blue color), 43 BSRN stations where only downward components were measured (green color) for the time period 1992–2012, as well as GEBA stations (red color) for the time period 1979–2012 using Kling–Gupta Efficiency component E_{KGr} (-) as indicator for correlation, thus timing. Numbers after the slash indicate the number of stations that are outside of the 1.5 × inter quartile range and thus not included in the boxplot. Significant differences of metric distribution (two sample Kolmogorov–Smirnov test, $p < 0.05$) are present for GEBA stations and shortwave downward radiation for ERAID-WFDEI and ERAIN-WFDEI.

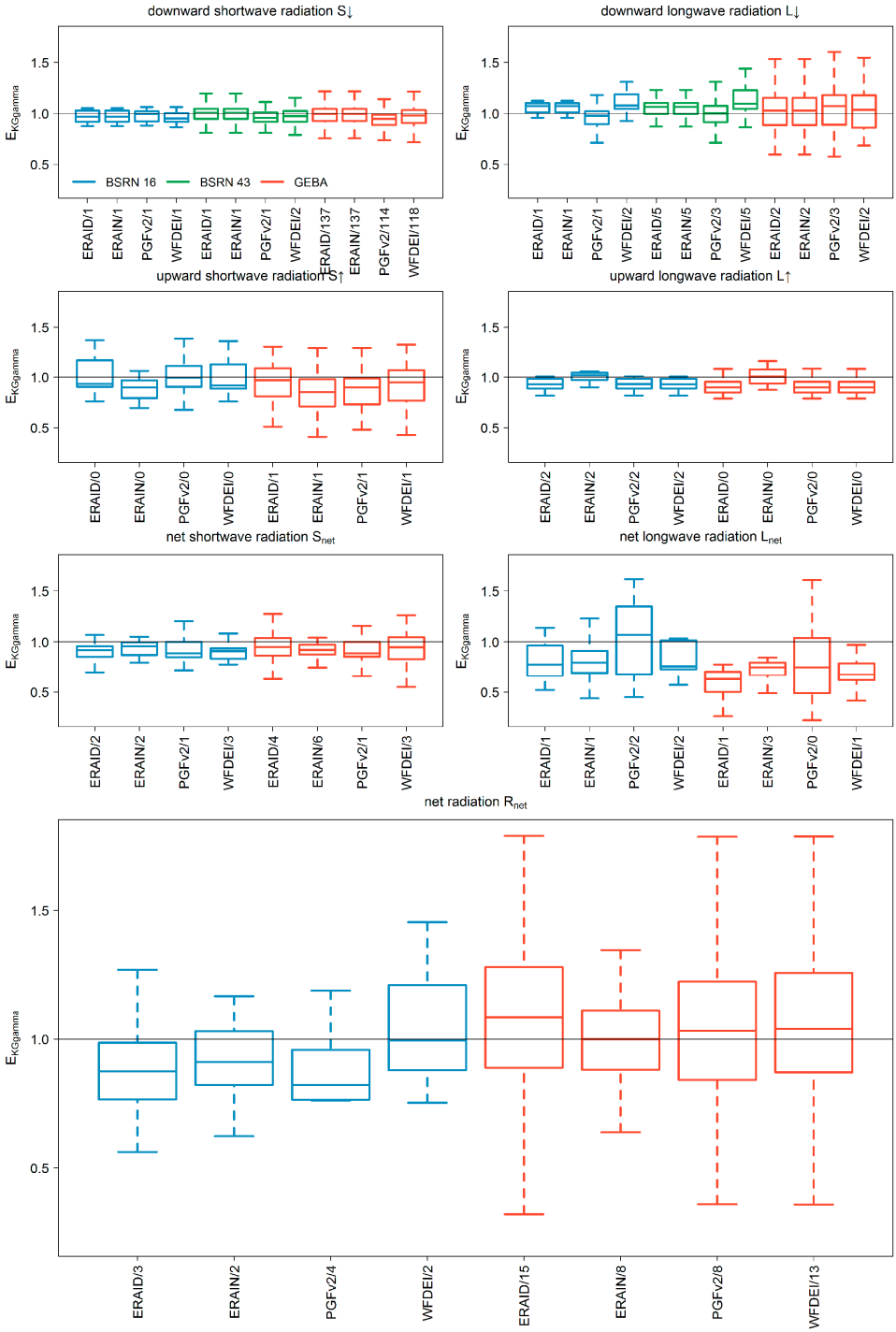


Figure 5. Validation of monthly radiation components with measurements of 16 BSRN stations (where all components are measured, blue color), 43 BSRN stations where only downward components were measured (green color) for the time period 1992–2012, as well as GEBA stations (red color) for the time period 1979–2012 using Kling–Gupta Efficiency component $E_{KGgamma}$ (-) as indicator for variability. Numbers after the slash indicate the number of stations that are outside of the 1.5 × inter quartile range and thus not included in the boxplot. Significant differences of metric distribution (two sample Kolmogorov–Smirnov test, $p < 0.05$) are present for BSRN stations and longwave downward radiation for all combinations with PGFv2, for longwave upward radiation for ERAID-ERAIN, ERAIN-PGFv2 and ERAIN-WFDEI as well as for net radiation between PGFv2 and WFDEI. For the GEBA stations, significant differences occur for shortwave downward radiation among all combinations (except ERAID-ERAIN), and for longwave upward as well as for net radiation between ERAID-ERAIN, ERAIN-PGFV2 and ERAIN-WFDEI.

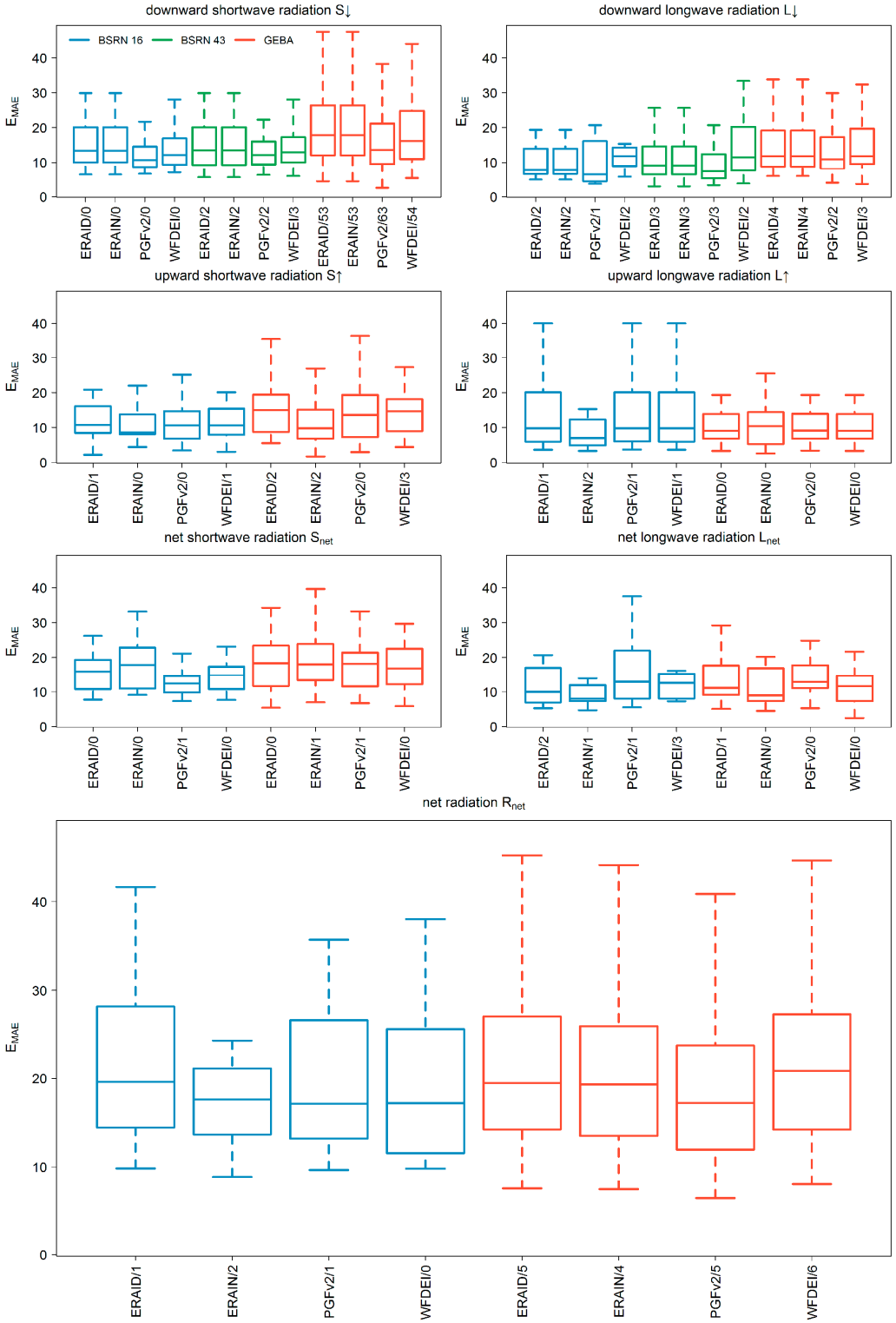


Figure 6. Validation of monthly radiation components with measurements of 16 BSRN stations (where all components are measured, blue color), 43 BSRN stations where only downward components were measured (and green color) for the time period 1992–2012, as well as GEBA stations (red color) for the time period 1979–2012 using Mean Absolute Error E_{MAE} (Wm^{-2}). Numbers after the slash indicate the number of stations that are outside of the 1.5 × inter quartile range and thus not included in the boxplot.

Simulated S↑ is modeled with low agreement in comparison to observed values. The timing of S↑ (indicated by E_{KGr}, Figure 4) is similar among the several forcings. The indicator of variability ($E_{KGgamma}$) is underestimated more strongly by ERAIN compared to the simulated ones (Figure 5) but most differences among the forcings occur in mean bias (E_{KGbeta}, Figure 3). For GEBA stations and simulated S↑, the mean bias E_{KGbeta} is greater than one. Besides the non-representative albedo and emissivity values of the surface beneath the measurement surface, obviously, different albedo values are responsible for different mean biases between simulations and reanalysis-based S↑ (compare ERAID and ERAIN for GEBA dataset of Figure 3).

There is no general best meteorological forcing for S_{net}. Due to the small footprint of upward radiation measurement, we did not expect better agreements. For example, if the area below the measurement consists of grassland but the majority of the 0.5° grid cell contains forest, the agreement is expected to be low due to the strongly differing albedo. For the BSRN stations, PGFv2 has the highest performance whilst for the GEBA stations WFDEI (and ERAID for E_{MAE}) ranks first (Table 3). The variability of net shortwave radiation is underestimated by all forcings, but to a lower degree by ERAIN (Figure 5). Differences in modeling L↑ are small (values close to optimum for E_{KGr} and E_{KGbeta}) to moderate ($E_{KGgamma}$). Interestingly, for L↓, variability is likely to be overestimated by all model variants (except PGFv2 and BSRN) and the variability of L↑ is at the same time underestimated by the WaterGAP simulations. All model variants have a low performance for L_{net} which is a result of differences in timing and variability (Figures 4 and 5). ERAIN ranks first (for E_{MAE} and GEBA together with WFDEI). However, the most important variable for calculating PET is R_{net}. For the BSRN stations, the model variant WFDEI has the highest performance to observations; for GEBA stations, the PGFv2 variant ranks first in both cases followed by ERAIN. In terms of mean bias, R_{net} is overestimated by the model variants, independently of the meteorological forcing (Figure 3). This is in agreement with Wild et al. [3], as they estimated a much lower value for the global land area (see Section 3.2, although Antarctica was included in their assessment). We assume that the overestimation of mean S↓ (Figure 3) is one reason for the high value of global land R_{net} obtained in this study. The measure of variability (Figure 5) (and to a lesser extent timing, Figure 4) for R_{net} varies by reference dataset. Nevertheless, the mean absolute biases for R_{net} are with 15 to 20 Wm^{-2} being the highest deviation from all the components throughout the forcings.

Table 3. Rank sums divided by number of stations for the radiation components of BSRN and GEBA stations and the efficiency metrics E_{KG} and E_{MAE} . The lowest rank is shown in bold font, indicating the best match to observations. Model variants are abbreviated with ERD (stands for ERAID), ERN (ERAIN), PGF (PGFv2), WFD (WFDEI) and net radiation components with a subscript "n".

	BSRN								**GEBA**							
	E_{KG}				E_{MAE}				E_{KG}				E_{MAE}			
	ERD	**ERN**	**PGF**	**WFD**	**ERD**	**ERN**	**PGF**	**WFD**	**ERD**	**ERN**	**PGF**	**WFD**	**ERD**	**ERN**	**PGF**	**WFD**
S↓	2.8	2.8	2.6	**1.8**	2.8	2.8	2.3	**2.1**	2.8	2.8	**2.0**	2.4	2.9	2.9	**2.0**	2.3
S↑	**2.2**	2.5	3.0	2.3	2.3	2.7	3.0	**2.0**	3.0	2.3	**1.9**	2.9	3.0	**2.0**	**2.0**	3.0
S_n	3.0	2.3	2.5	**2.2**	3.0	3.2	**1.3**	2.5	**2.3**	2.6	2.6	2.5	**2.4**	2.6	2.6	**2.4**
L↓	**2.4**	**2.4**	**2.4**	2.9	2.4	2.4	**1.8**	3.4	**2.5**	**2.5**	2.6	**2.5**	2.6	2.6	**2.2**	2.6
L↑	2.6	**2.0**	2.8	2.6	2.6	**2.0**	2.8	2.6	2.5	**2.1**	2.8	2.5	**2.1**	2.5	3.4	**2.1**
L_n	2.8	**1.5**	3.2	2.5	2.2	**1.7**	3.5	2.7	3.0	**1.5**	3.0	2.5	2.8	**2.1**	3.0	**2.1**
R_n	3.2	2.2	2.7	**2.0**	3.0	2.8	2.2	**2.0**	2.8	2.4	**2.2**	2.6	2.5	2.5	**2.3**	2.7

Analyzing the relative contribution of the E_{KG} components to the overall value (Figure 7), it can be stated that the correlation (thus the timing) contributes to the least extent (except S↑) for all radiation components. For the shortwave radiation components (and R_{net}), the mean bias is dominating, whereas variability is dominating for the longwave components.

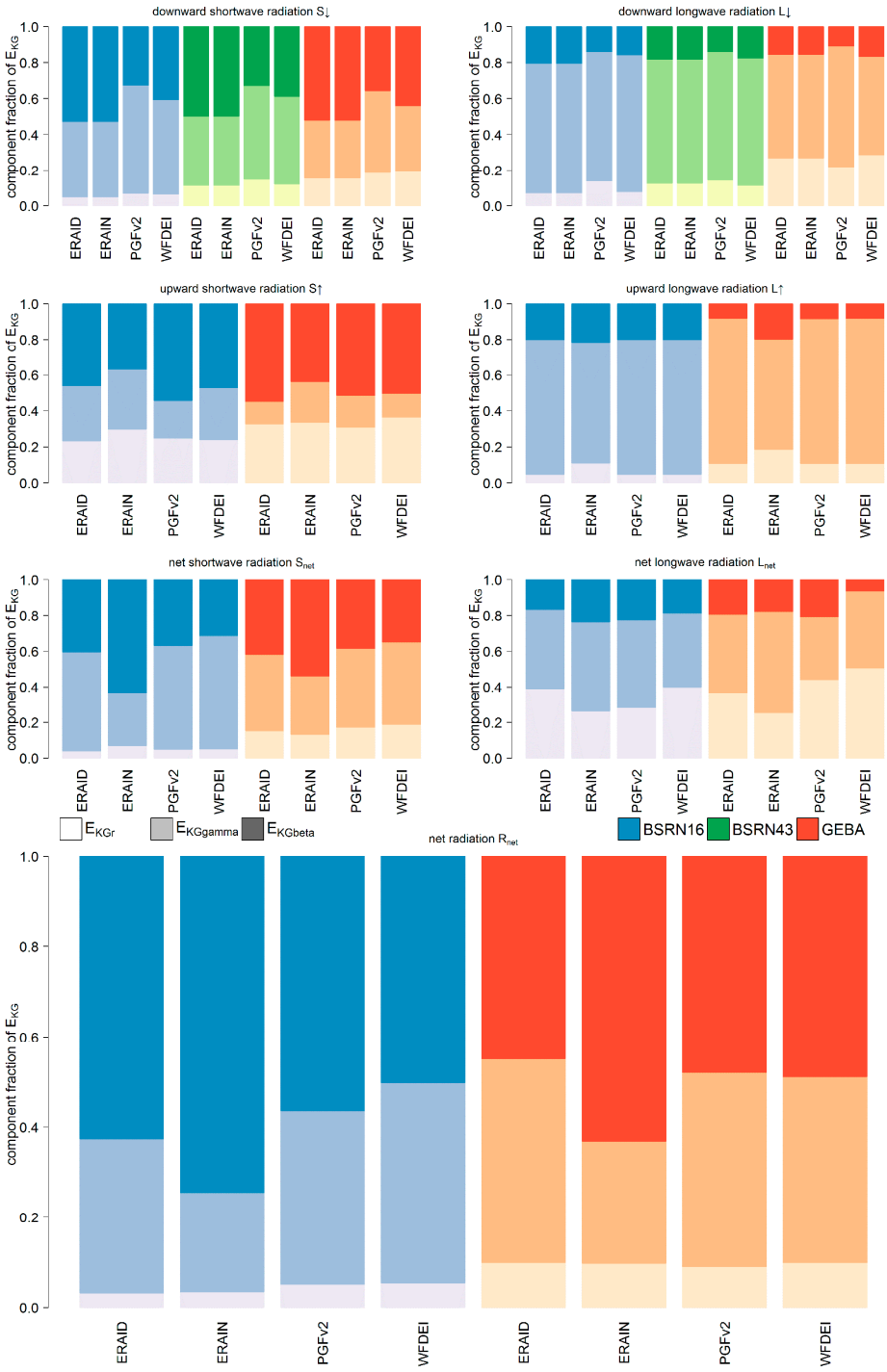


Figure 7. The average fraction of E_{KG} component (-) showing the different contribution of correlation, variability and mean bias to the overall efficiency criteria.

3.2. Intercomparison of Global-Scale Net Radiation and Its Components

Figure 8 shows simulated net radiation (bottom row) and its components from WFDEI and differences to the other model variants described in Table 1. It is worth noting that in northern Africa and Somalia, the Saudi-Arabian peninsula, central and Western Australia, as well as the Amazon and Mexico, R_{net} of ERAIN is significantly lower than the current standard meteorological forcing of WFDEI (brown to red colors in Figure 8, bottom line). Despite large areas of lower R_{net} values, there are numerous grid cells with drastically higher values for ERAIN compared to WFDEI

(maximum difference in the Himalayan Mountains is 196 Wm^{-2}). For long term (1979–2012) global land (except Antarctica) averages of net radiation, ERAIN has the highest values (83.0 Wm^{-2}), compared to (slightly) lower values of the other forcings (PGFv2: 82.9 Wm^{-2}, ERAID 82.0 Wm^{-2}, WFDEI 81.0 Wm^{-2}). Other estimates (that include Antarctica) are lower (e.g., 70 Wm^{-2}, [3]).

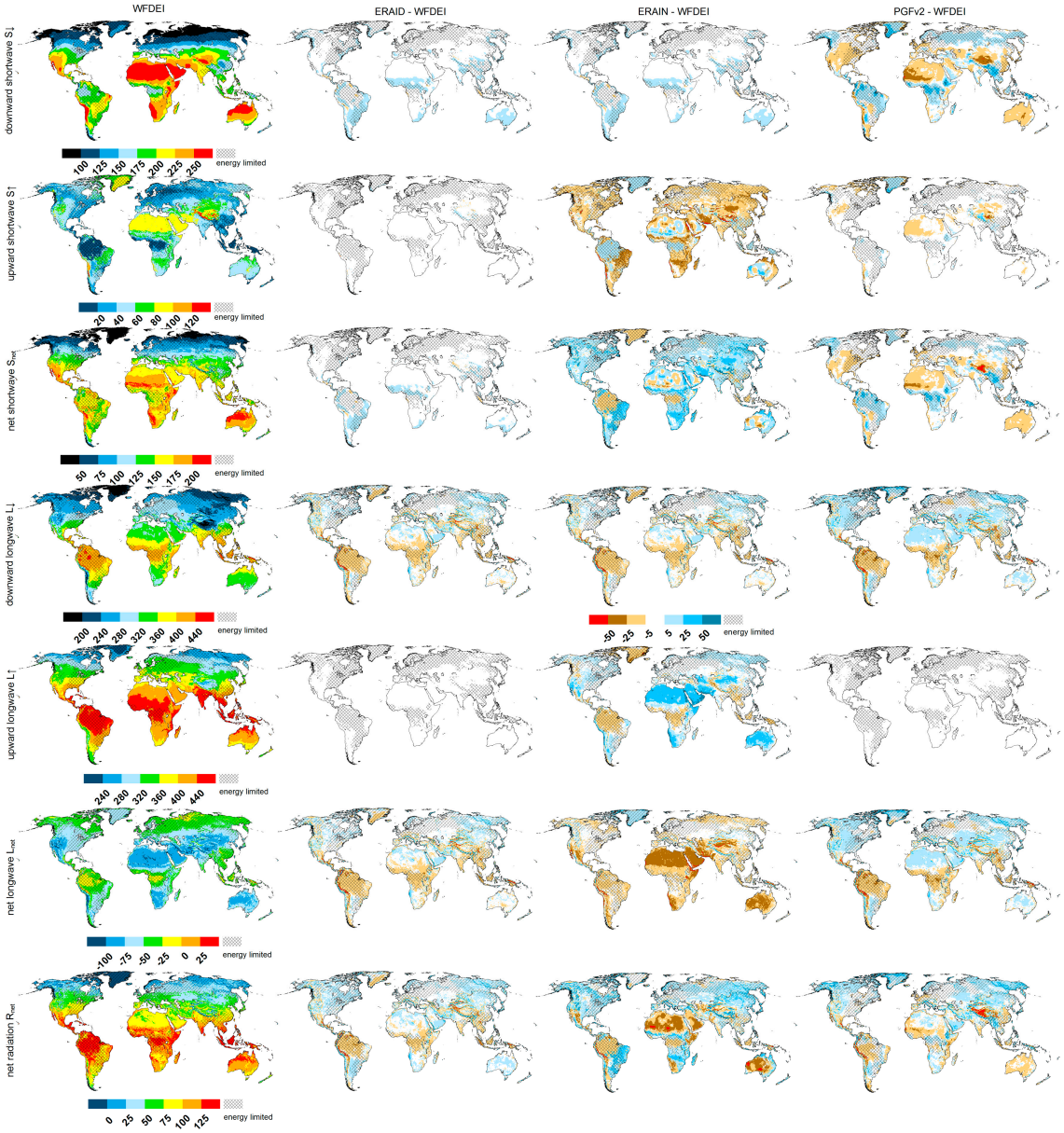


Figure 8. Average radiation components for 1980–2009 for WFDEI (1st column) and differences for ERAID-WFDEI (2nd), ERAIN-WFDEI (3rd) and PGFv2-WFDEI (last column). All units in Wm^{-2}. Cross-hatched areas are energy limited areas after the Budyko approach, taken from the STANDARD model run of [27].

Comparing WFDEI and ERAID (which differ in terms of the adjustments of Weedon et al. [8]), 40.2% of the global land area are within ± 5 Wm^{-2} difference for R_{net}, which shows a higher agreement when compared to the other forcings (PGFv2 35.1%, ERAIN 19.1%). Whereas higher (lower) R_{net} values for ERAID are calculated in 36.2% (23.7%) of land area, a similar pattern, i.e., the tendency to calculate higher R_{net} can be found for PGFv2 (37.2%, 27.8%) and ERAIN (48.9%, 32.0%). When separating the energy-limited and water-limited regions (Section 1), the forcing leads to pronounced

differences. For ERAID and energy-limited (water-limited) regions, 34.3% (45.3%) of land area is within ± 5 Wm^{-2} difference, with similar values for PGFv2 (38.3%, 32.3%) but are in much lower agreement with ERAIN (18.8%, 19.3%).

R_{net} is larger than 5 Wm^{-2} in the percentage of land area for energy-limited/water-limited regions and ERAID (40.3%/32.6%) as well as PGFv2 (43.3%/31.9%). For ERAIN, 62.0% of energy-limited regions have R_{net} values that are higher than 5 Wm^{-2} compared to WFDEI, whereas water-limited regions have, in the predominantly land area (43.2%), R_{net} values that are lower than 5 Wm^{-2} (see also the spatial differences in Figure 8). Hence, by using ERAID or PGFv2 instead of WFDEI, around 1/3 (and for ERAIN around 1/2) of energy-limited regions could profit from an increasing amount of energy for evapotranspiration. Taken into account the different fraction of energy-limited/water-limited regions, on 18.7% to 28.8% of the global land area, the forcings ERAID, PGFv2 and ERAIN are calculated to have higher R_{net} (compared to WFDEI) which would likely increase evapotranspiration. In contrast, on 8.6% to 11.8% on the global scale, R_{net} decreases with the alternative forcing.

Absolute differences for the radiation components from WFDEI are also shown in Figure 8. Differences in S↓ for WFDEI compared to ERAID and ERAIN are low, but compared to PGFv2, WFDEI has 5 to 25 Wm^{-2} less S↓ in the US, northern Africa, Australia and the mountain regions in central Asia. In tropical regions, and to a lesser extent in northern Eurasia, S↓ is higher in WFD (Figure 8, first row). S↑ differences of WFDEI and ERAID are (except for small regions) within ± 5 Wm^{-2}, which is expected due to the smaller absolute values. As the land cover class dependent albedo values are comparable, WFDEI and PGFv2 differ with more than ± 5 Wm^{-2} in areas where S↓ is differing, but to a much lesser extent. Most differences (between ± 50 Wm^{-2}) occur between WFDEI and ERAIN, which is interesting as S↓ is comparable. This could be related to a higher surface albedo of ERAIN that induces higher S↑. Other factors, e.g., clouds, might also contribute to the higher upward radiation, but no hints for an overestimation of cloudiness have been found by comparisons with BSRN stations in Europe and Africa [18]. S_{net} is very similar between WFDEI and ERAID, but differs from ERAIN (in general with lower values except the tropics) and PGFv2 (in general with higher values except for the tropics).

The pattern of differences between WFDEI to the other model variants for L↓ is similar in most regions; in the tropics, WFDEI has less L↓, whilst in most other regions, differences are within ± 5 Wm^{-2} or, in the case of PGFv2, higher (up to 25 Wm^{-2}) in most other regions. Differences > 5 Wm^{-2} can be found for L↑ only between WFDEI and ERAIN. Most of the dry regions have higher values in WFDEI (up to 50 Wm^{-2} for large areas), whereas the tropics (and Greenland) have lower values up to 25 Wm^{-2}. The difference pattern for L_{net} is diverse. ERAIN has, in general, smaller values for L_{net}, especially in dry regions; ERAID has lower values in the tropics and PGFv2 has lower values in the tropics but higher values for most other areas. When comparing the differences of WFDEI to ERAIN and ERAID, the effect of modeling outgoing radiation within WaterGAP is visible. Based on an additional model run, where T from ERA-Interim (instead from WFDEI) was used for ERAID, we can summarize that T affects model results in only a minor manner and thus, differences in emissivity values (or other, ERA-Interim related characteristics) determine the difference between ERAIN and the WaterGAP calculations. The same results occur when T and P from WFDEI are used in a PGFv2 model run. Whereas the differences in the Amazon are comparable, ERAIN shows large differences in Africa, the Saudi-Arabian peninsula as well as Australia. For some regions, the calculation of outgoing radiation components in WaterGAP (ERAID, PGFv2) leads to changing signs of the differences compared to those obtained from ERAIN.

Summarizing the main findings of the inter-comparison of the four model variants, it can be stated that ERAIN has the highest net radiation at the global land scale and the standard forcing WFDEI the lowest. In energy-limited regions, the alternative forcings (especially ERAIN) have higher net radiation which has an influence on the evapotranspiration calculation. Especially for the downward components and PGFv2 as well as for the upward components and ERAIN, differences among the model variants are visible.

3.3. Discussion of Optimal Radiation Input

The downward components of all model variants show a high performance compared to the observed radiation apparent in Figure 2 (top row) but may still have systematic biases in terms of over and underestimation (Figure 3). According to the rank sums in Table 3 and S↓, highest performances are simulated when WaterGAP is forced by WFDEI, followed by PGFv2 and ERAID/ERAIN variants. The preparation and adjustments of WFDEI [8,29] to radiation fluxes from ERA-Interim reanalysis (see Section 2.1.2) lead to improved results for S↓, independently of the station dataset and efficiency criterion (Table 3). Interestingly, the E_{KG} components, except $E_{KGgamma}$ (which is the dominant component (Figure 7) and thus influencing overall E_{KG} value at most), perform worse than the directly obtained S↓ from ERA-Interim (Tables 4 and 5). For L↓ (only elevation corrected), performance is worse for BSRN stations (except $E_{KGgamma}$ which is again the dominant component, Figure 7) but of the same quality for GEBA stations (Table 3). As the results of WFDEI are similar or better than ERAID (except for L↓ and BSRN), the thorough interpolation scheme of Weedon et al. [8] improved the quality of the meteorological forcing. The PGFv2 model variant ranks second for S↓ and BSRN and first for GEBA stations, whereas for L↓ the PGFv2 model variant is at the first position for both observation datasets (except for E_{KG} and GEBA data) and can, therefore, be seen as a valuable alternative for downward radiation.

Table 4. Rank sums divided by number of stations for the radiation components, BSRN stations and the efficiency metrics E_{KGr}, E_{KGbeta} and $E_{KGgamma}$. The lowest rank is shown in bold font, indicating the best match to observations. Model variants are abbreviated with ERD (stands for ERAID), ERN (ERAIN), PGF (PGFv2), WFD (WFDEI) and net radiation components with a subscript "n".

	E_{KGr}				E_{KGbeta}				$E_{KGgamma}$			
	ERD	**ERN**	**PGF**	**WFD**	**ERD**	**ERN**	**PGF**	**WFD**	**ERD**	**ERN**	**PGF**	**WFD**
S↓	**2.1**	**2.1**	2.6	3.2	**2.2**	**2.2**	2.6	2.9	2.6	2.6	2.5	**2.3**
S↑	**2.4**	2.6	**2.4**	2.7	2.5	2.8	**2.2**	2.5	2.6	2.6	2.7	**2.1**
S_n	2.2	**2.1**	2.2	3.5	2.3	**1.8**	2.9	2.9	2.3	3.1	2.6	**1.9**
L↓	**1.8**	**1.8**	3.1	3.2	2.4	2.4	3.1	**2.0**	2.7	2.7	2.5	**2.2**
L↑	2.5	**2.2**	2.9	2.5	2.5	3.2	**1.8**	2.5	**2.3**	2.9	2.4	**2.3**
L_n	2.8	**1.4**	2.6	3.2	2.7	2.8	**2.0**	2.6	2.6	2.8	**2.1**	2.6
R_n	2.4	**1.6**	3.1	2.9	**1.9**	2.5	2.4	3.2	2.3	3.2	**2.0**	2.4

Table 5. Rank sums divided by number of stations for the radiation components GEBA stations and the efficiency metrics E_{KGr}, E_{KGbeta} and $E_{KGgamma}$. The lowest rank is shown in bold font, indicating the best match to observations. Model variants are abbreviated with ERD (stands for ERAID), ERN (ERAIN), PGF (PGFv2), WFD (WFDEI) and net radiation components with a subscript "n".

	E_{KGr}				E_{KGbeta}				$E_{KGgamma}$			
	ERD	**ERN**	**PGF**	**WFD**	**ERD**	**ERN**	**PGF**	**WFD**	**ERD**	**ERN**	**PGF**	**WFD**
S↓	2.3	2.3	**2.2**	3.2	**2.1**	**2.1**	3.0	2.9	2.6	2.6	**2.4**	**2.4**
S↑	2.7	2.3	**1.8**	3.2	**2.0**	3.0	2.9	2.1	2.8	**2.1**	2.5	2.6
S_n	2.4	2.1	**2.0**	3.4	2.6	**2.3**	2.4	2.8	2.7	2.7	2.4	**2.2**
L↓	**2.1**	**2.1**	2.8	3.0	2.4	2.4	**2.9**	2.2	2.6	2.6	**2.4**	2.5
L↑	2.4	**2.2**	3.0	2.4	**2.7**	2.6	2.0	**2.7**	2.4	3.1	**2.1**	2.4
L_n	2.6	**1.2**	3.4	2.8	**2.2**	2.4	2.6	2.8	**1.8**	3.1	2.2	2.8
R_n	2.7	**1.8**	2.7	2.9	2.5	**2.4**	2.7	2.4	**2.3**	3.0	2.5	**2.3**

WaterGAP simulations of S↑ have a higher performance than those from ERAIN for the BSRN stations and a lower performance for WFDEI and ERAID for the GEBA stations consistently among the two efficiency criteria (Tables 3 and 4). The calculation of S↑ in WaterGAP and underlying assumptions (e.g., albedo) are of the same quality as those from ERAIN for the grid cells where measurement data are available, even though the E_{KG} values of S↑ are drastically lower compared to S↓. Due to the small footprint of upward radiation measurements (and thus the questionable representativeness for

the 0.5° grid cell) as well as the low number of stations, this result may also be driven by chance. The E_{KG} of L↑ is close to the optimum with highest performance with ERAIN for BSRN stations and for E_{KG} also for GEBA stations (Table 3), whereas E_{KGbeta} and E_{KGr} for all variants are close to optimum and the indicator of variability $E_{KGgamma}$ (which is dominant in E_{KG} value) differs between the model variants (Figures 3–5 and 7). For BSRN and GEBA stations, the WaterGAP calculation of L↑ variability $E_{KGgamma}$ has a higher performance compared to observations than those from ERAIN (Tables 4 and 5). Even though, overall, the E_{KG} value has a slightly weaker performance than those from ERAIN, the approach of WaterGAP using a land cover class based emissivity value to determine L↑ can be seen as plausible. However, here also, the small number of stations and months available for upward radiation components data (Table 2), as well as the small footprint for upward radiation measurement, limits the assessment. However, currently, the emissivity value is kept constant within WaterGAP which is a problem especially when land surface is covered by snow which influences emissivity values [52–54]. It should, therefore, be tested if a snow-cover-dependent emissivity value leads to improvements for WaterGAP calculation of L↑.

The evaluated L_{net} performance depends largely on the efficiency metric applied. Whereas E_{KG} values perform low for all model variants (Figure 2), values for E_{MAE} are, compared to the other radiation fluxes, relatively low. The subtraction of two large values (L↓, L↑) leads to low absolute values (e.g., compared to S_{net}, Figure 8) and thus, absolute errors can be lower. On the other hand, small uncertainties in L↓ and L↑ are translated into large uncertainties for L_{net} which is visible, e.g., for E_{KGbeta} (Figure 3).

For calculating PET, R_{net} is the most important variable. The relative efficiency criterion E_{KG} (Figure 2) is relatively low, independently of the model variant and the medians of absolute error E_{MAE} are high with around 20 Wm^{-2}. The rank sums for R_{net} do not perform highest when substituting the upward radiation calculation of WaterGAP with those values obtained directly by ERA-Interim (Table 3). ERAIN has a higher performance than ERAID but for each efficiency criteria and station database, either WFDEI or PGFv2 perform better than ERAIN (Table 3). Snow dynamics are modeled within WaterGAP and albedo changes when snow cover is present [27]. Therefore, the use of ERAIN may induce inconsistent representation of land cover dependent parameters as albedo is also modeled by ERAIN. This, and the small benefit of using ERAIN, has to be considered when deciding about the optimal input.

WaterGAP uses a very simple scheme to calculate the energy balance at the surface. It uses only land-cover dependent albedo and emissivity values. The radiation transfer scheme that is integrated in ERA-Interim is highly complex, taken into account different bandwidths of radiation transfer, cloud (and atmosphere) properties and is designed to close the energy balance both at the surface and top level of the atmosphere [7]. Even so, in principle, ERA-Interim also considers surface albedo and emissivity and is therewith comparable in a very general sense, even though the processes are far more physically constrained.

All model variants have a positive mean bias for R_{net} (Figure 3), which indicates that this variable is overestimated by about 20% compared to measurements. Consequently, E_{KGbeta} is the dominant component of the E_{KG} value (Figure 7). Obviously, the higher E_{KGbeta} of S↓ propagates to S_{net} and R_{net} (Figure 7) as this is described in the literature [2,3]. For the 16 BSRN stations, the current standard forcing of WaterGAP (WFDEI) provides the highest performance for both efficiency criteria (Table 3). For the 142 GEBA stations with net radiation data, the best match to simulations can be reached when using PGFv2 forcing (Table 3).

One of the foci of this study was to assess whether differences in results occur in energy- and water-limited areas. However, the discrepancy of the distribution of stations in energy limited vs. water limited regions (on average, only 28% of GEBA stations and 38% of BSRN stations are in water-limited grid cells) hinders a robust assessment of the performance in those distinct regions. Rank sums aggregated for BSRN and energy/water limited regions (not shown) indicate only slight deviations from those in Table 3.

The scale mismatch between (small scale) station observations and 0.5 ° grid cells results in uncertain comparison results. The median E_{MAE} for S↓ from this study is about three times higher (10 to 15 Wm^{-2}, Figure 6) than those of Hakuba et al. [26]. Obviously, due to the heterogeneity of land surface characteristics, the representativeness of downward and upward radiation fluxes differs fundamentally. However, monthly averaging can smear short-term (and small scale) atmospheric effects, which barley applies for land surface heterogeneities.

PET was calculated for the four model variants using the Priestley-Taylor approach [55] (Figure 9). In general, the difference patterns are similar to those of R_{net} (Figure 8). Less PET, in comparison with WFDEI, occurs for all other model variants mainly in the tropics and for ERAIN in the very dry regions of Northern Africa and Australia. On a global scale, WFDEI has highest value for PET and the overall range among the model variants is about 24 mm·$year^{-1}$ (lowest value for ERAIN). The highest absolute range (100 mm·$year^{-1}$) among the models occurs in water limited grid cells, where WFDEI and ERAID have the same values and ERAIN the lowest. For the energy limited regions, WFDEI has the second lowest value (ERAIN has the highest) while the models vary by about 70 mm·$year^{-1}$ (Table 6). As AET could be only increased in energy limited regions, this implies that WaterGAP would simulate higher AET when using ERAIN as radiation input.

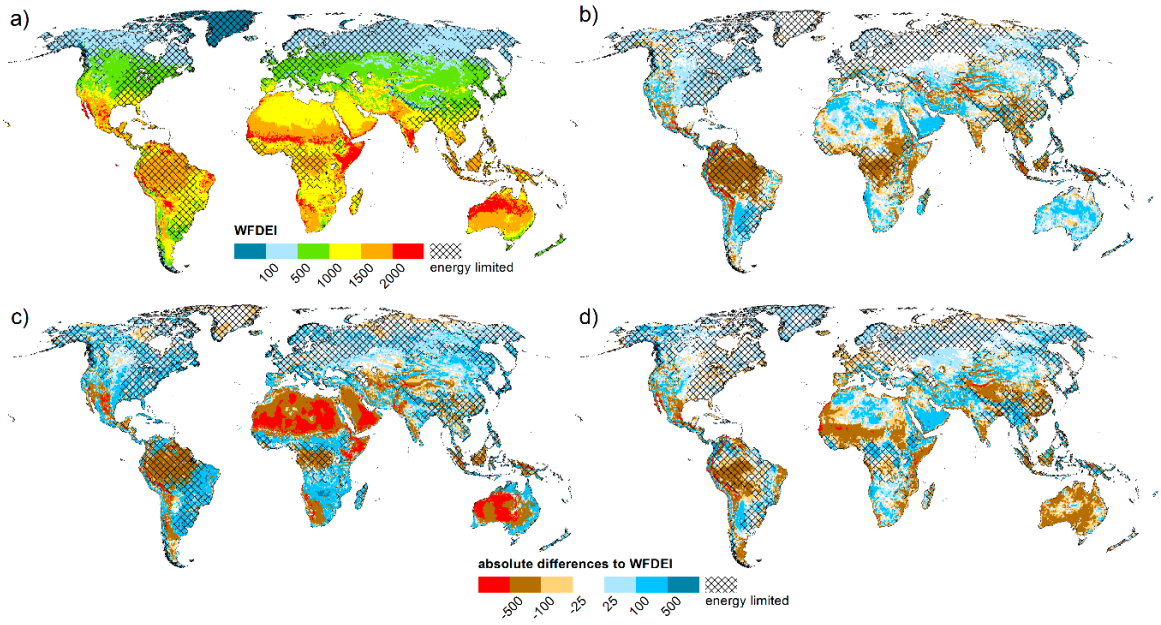


Figure 9. Long term (1980–2009) average of potential evapotranspiration as calculated by the Priestley–Taylor approach for WFDEI (**a**); and differences for ERAID-WFDEI (**b**); ERAIN-WFDEI (**c**); and PGFv2-WFDEI (**d**) (all units in mm·$year^{-1}$).

Table 6. Simulated long term (1980–2009) average PET (mm·$year^{-1}$) for the model variants and global scale as well as for energy- and water-limited regions. Model variants are abbreviated with ERD (stands for ERAID), ERN (ERAIN), PGF (PGFv2), WFD (WFDEI).

Region	ERD	ERN	PGF	WFD
Global	1113	1092	1103	1116
Energy-limited	864	933	894	872
Water-limited	1336	1235	1292	1336

4. Conclusions

In this study, we assessed the performance of radiation components in experiments with the global freshwater model WaterGAP on a 0.5° × 0.5° grid scale against station observations from the

BSRN and GEBA databases. The model was driven with two state-of-the-art meteorological forcings (WFDEI and PGFv2) as well as with the interpolated downward radiation data from ERA-Interim reanalysis (ERAID) and finally a radiation dataset, including both downward and upward components from ERA-Interim (ERAIN). We used two efficiency metrics (one absolute, and one relative) for the comparison analysis. In addition, we assessed spatial differences among the radiation components as well as their effect on PET. The results can be summarized as follows:

- For global land averages, R_{net} differs only slightly among the model variants (~2 Wm^{-2}). However, regionally large differences of R_{net}, S↑ and L_{net} were found, especially in comparison with ERA-Interim reanalysis (Figure 8).
- R_{net} values of WaterGAP as forced by WFDEI is within ±5 Wm^{-2} agreement to the ERA-Interim reanalysis for 19.1% of the global land area, with similar numbers for energy-limited and water-limited regions.
- In 62.0% of energy-limited regions (relates to 28.8% of the global land area), R_{net} of the full radiation dataset from ERA-Interim is higher by more than 5 Wm^{-2} and has, therefore, the highest potential to increase the simulated evapotranspiration as compared to the other forcings.
- The downward radiation components of ERA-Interim show less or similar agreement to station observations compared to those from meteorological forcings (Table 3). The interpolation and correction approach of Weedon et al. [8] improves both downward radiation components (Table 3). However, for all model variants, a systematical overestimation of S↓ and R_{net} was found when comparing to observation data.
- The performance of S↑ radiation of ERA-Interim lies between the meteorological forcings WFDEI and PGFv2. For BSRN stations, the model variant where ERA-Interim downward components are used has a higher performance than the variant where also ERA-Interim upward components are used, whereas the opposite is true for GEBA stations. ERA-Interim values for L↑ are superior compared to those of WaterGAP except for the absolute error measure derived at the GEBA stations (Table 3).
- Best results for R_{net} are found for current standard forcing (WFDEI, BSRN) or alternative forcing (PGFv2, GEBA) (Table 3), but median absolute errors are around 20 Wm^{-2} (comparable to the study of Troy and Wood [16]) and are mainly due to a higher mean value independent on the model variant (Figures 3 and 4).
- Global values for PET vary only by 25 mm·$year^{-1}$ and the highest values are achieved with WFDEI forcing. However, in energy limited regions, where a change in PET directly influences AET, WFDEI has the second lowest value while ERAIN is 60 mm·$year^{-1}$ higher.

Some limitations were found:

- The relatively small number of some radiation measurements (e.g., 16 stations with upward flux measurements for BSRN) limits the overall assessment and hinders a robust assessment of the performance of the model variants, especially when separating into energy-limited/water-limited areas.
- In contrast to the downward components, which show a reasonable representativeness of station measurements at the grid cell level [26], upward measurements have a small footprint and thus a high uncertainty in terms of representativeness for the grid cell.

The results imply a certain impact of radiation forcing and modeling on the simulation of PET, especially for energy-limited regions, which then affects the modeling of available water resources. However, calculation of PET can be done with many approaches which all have their own uncertainties. To quantify this, in particular, a study on the impact of PET (including different calculation approaches) is needed, followed by a quantification of the consequences for AET and freshwater resources. The results of this study can help to improve the understanding of net radiation estimates as

an important driver of simulated evapotranspiration in GHMs. Even though this assessment is model-specific, GHMs with a comparable approach (i.e., land cover dependent albedo and emissivity) can benefit from this analysis as two of the most frequently used meteorological forcings and one reanalysis for two radiation setups were evaluated.

Acknowledgments: We thank Petra Döll for initial discussions according to the study design and Melanie Seeger (Möller before marriage) for her input regarding the efficiency criteria. We greatly appreciate the suggestions of three anonymous reviewers that helped to improve the manuscript significantly.

Author Contributions: Hannes Müller Schmied designed the study, performed the research, prepared the data and ran the model, analyzed the model output, created figures and tables and wrote the paper. All authors contributed to discussions and manuscript revision.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Trenberth, K.E.; Fasullo, J.T.; Kiehl, J. Earth's global energy budget. *Bull. Am. Meteorol. Soc.* **2009**, *90*, 311–323. [CrossRef]
2. Wild, M.; Folini, D.; Schär, C.; Loeb, N.; Dutton, E.G.; König-Langlo, G. The global energy balance from a surface perspective. *Clim. Dyn.* **2013**, *40*, 3107–3134. [CrossRef]
3. Wild, M.; Folini, D.; Hakuba, M.Z.; Schär, C.; Seneviratne, S.I.; Kato, S.; Rutan, D.; Ammann, C.; Wood, E.F.; König-Langlo, G.; et al. The energy balance over land and oceans: An assessment based on direct observations and CMIP5 climate models. *Clim. Dyn.* **2015**, *44*, 3393–3429. [CrossRef]
4. Wild, M.; Liepert, B. The Earth radiation balance as driver of the global hydrological cycle. *Environ. Res. Lett.* **2010**, *5*. [CrossRef]
5. Sood, A.; Smakhtin, V. Global hydrological models: A review. *Hydrol. Sci. J.* **2014**, *60*, 549–565. [CrossRef]
6. Bierkens, M.F.P. Global hydrology 2015: State, trends, and directions. *Water Resour. Res.* **2015**, *51*, 4923–4947. [CrossRef]
7. Dee, D.P.; Uppala, S.M.; Simmons, A.J.; Berrisford, P.; Poli, P.; Kobayashi, S.; Andrae, U.; Balmaseda, M.A.; Balsamo, G.; Bauer, P.; et al. The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Q. J. R. Meteorol. Soc.* **2011**, *137*, 553–597. [CrossRef]
8. Weedon, G.P.; Balsamo, G.; Bellouin, N.; Gomes, S.; Best, M.J.; Viterbo, P. The WFDEI meteorological forcing data set: WATCH Forcing Data methodology applied to ERA-Interim reanalysis data. *Water Resour. Res.* **2014**, *50*, 7505–7514. [CrossRef]
9. Sheffield, J.; Goteti, G.; Wood, E.F. Development of a 50-year high-resolution global dataset of meteorological forcings for land surface modeling. *J. Clim.* **2006**, *19*, 3088–3111. [CrossRef]
10. Müller Schmied, H.; Adam, L.; Eisner, S.; Fink, G.; Flörke, M.; Kim, H.; Oki, T.; Portmann, F.T.; Reinecke, R.; Riedel, C.; et al. Variations of global and continental water balance components as impacted by climate forcing uncertainty and human water use. *Hydrol. Earth Syst. Sci.* **2016**, *20*, 2877–2898. [CrossRef]
11. Nasonova, O.N.; Gusev, Y.M.; Kovalev, Y.E. Impact of uncertainties in meteorological forcing data and land surface parameters on global estimates of terrestrial water balance components. *Hydrol. Process.* **2011**, *25*, 1074–1090. [CrossRef]
12. Biemans, H.; Hutjes, R.W.A.; Kabat, P.; Strengers, B.J.; Gerten, D.; Rost, S. Effects of precipitation uncertainty on discharge calculations for main river basins. *J. Hydrometeorol.* **2009**, *10*, 1011–1025. [CrossRef]
13. Döll, P.; Douville, H.; Güntner, A.; Müller Schmied, H.; Wada, Y. Modelling freshwater resources at the global scale: Challenges and prospects. *Surv. Geophys.* **2016**, *37*, 195–221. [CrossRef]
14. Wild, M.; Ohmura, A.; Gilgen, H.; Morcrette, J.-J. The distribution of solar energy at the Earth's sufrace as calculated in the ECMWF re-analysis. *Geophys. Res. Lett.* **1998**, *25*, 4373–4376. [CrossRef]
15. Wild, M.; Ohmura, A.; Gilgen, H.; Morcrette, J.-J.; Slingo, A. Evaluation of downward longwave radiation in General Circulation Models. *J. Clim.* **2001**, *14*, 3227–3239. [CrossRef]
16. Troy, T.J.; Wood, E.F. Comparison and evaluation of gridded radiation products across northern Eurasia. *Environ. Res. Lett.* **2009**, *4*. [CrossRef]
17. Jin, Y.; Randerson, J.T.; Goulden, M.L. Continental-scale net radiation and evapotranspiration estimated using MODIS satellite observations. *Remote Sens. Environ.* **2011**, *115*, 2302–2319. [CrossRef]

18. Posselt, R.; Mueller, R.W.; Stöckli, R.; Trentmann, J. Remote sensing of solar surface radiation for climate monitoring—The CM-SAF retrieval in international comparison. *Remote Sens. Environ.* **2012**, *118*, 186–198. [CrossRef]
19. Gómez, I.; Caselles, V.; Estrela, M. Seasonal Characterization of Solar Radiation Estimates Obtained from a MSG-SEVIRI-Derived Dataset and a RAMS-Based Operational Forecasting System over the Western Mediterranean Coast. *Remote Sens.* **2016**, *8*. [CrossRef]
20. Inamdar, A.; Guillevic, P. Net surface shortwave radiation from GOES imagery—Product evaluation using ground-based measurements from SURFRAD. *Remote Sens.* **2015**, *7*, 10788–10814. [CrossRef]
21. Boilley, A.; Wald, L. Comparison between meteorological re-analyses from ERA-Interim and MERRA and measurements of daily solar irradiation at surface. *Renew. Energy* **2015**, *75*, 135–143. [CrossRef]
22. König-Langlo, G.; Sieger, R.; Schmithüsen, H.; Bücker, A.; Richter, F.; Dutton, E.G. *Baseline Surface Radiation Network (BSRN) Update of the Technical Plan for BSRN Data Management*; WCRP Report 24/2013; World Meteorological Organization (WMO): Geneva, Switzerland, 2013.
23. Gilgen, H.; Ohmura, A. The Global Energy Balance Archive. *Bull. Am. Meteorol. Soc.* **1999**, *80*, 831–850. [CrossRef]
24. Heinemann, G.; Kerschgens, M. Simulation of surface energy fluxes using high-resolution non-hydrostatic simulations and comparisons with measurements for the LITFASS-2003 experiment. *Bound.-Layer Meteorol.* **2006**, *121*, 195–220. [CrossRef]
25. Horlacher, V.; Osborne, S.; Price, J.D. Comparison of two closely located meteorological measurement sites and consequences for their areal representativity. *Bound.-Layer Meteorol.* **2012**, *142*, 469–493. [CrossRef]
26. Hakuba, M.Z.; Folini, D.; Sanchez-Lorenzo, A.; Wild, M. Spatial representativeness of ground-based solar radiation measurements-Extension to the full Meteosat disk. *J. Geophys. Res. Atmos.* **2014**, *119*, 11760–11771. [CrossRef]
27. Müller Schmied, H.; Eisner, S.; Franz, D.; Wattenbach, M.; Portmann, F.T.; Flörke, M.; Döll, P. Sensitivity of simulated global-scale freshwater fluxes and storages to input data, hydrological model structure, human water use and calibration. *Hydrol. Earth Syst. Sci.* **2014**, *18*, 3511–3538. [CrossRef]
28. Budyko, M.I. *Climate and Life*; Academic Press: New York, NY, USA, 1974.
29. Weedon, G.P.; Gomes, S.; Viterbo, P.; Shuttleworth, W.J.; Blyth, E.; Österle, H.; Adam, J.C.; Bellouin, N.; Boucher, O.; Best, M. Creation of the WATCH Forcing Data and its use to assess global and regional reference crop evaporation over land during the twentieth century. *J. Hydrometeorol.* **2011**, *12*, 823–848. [CrossRef]
30. *ECMWF IFS Documentation—Cy40r1 Part IV: Physical Processes*; European Centre for Medium-Range Weather Forecasts: Reading, UK, 2014.
31. Uppala, S.M.M.; KÅllberg, P.W.; Simmons, A.J.; Andrae, U.; Bechtold, V.D.C.; Fiorino, M.; Gibson, J.K.; Haseler, J.; Hernandez, A.; Kelly, G.A.; et al. The ERA-40 re-analysis. *Q. J. R. Meteorol. Soc.* **2005**, *131*, 2961–3012. [CrossRef]
32. Stocker, T.F.; Qin, D.; Plattner, G.-K.; Tignor, M.; Allen, S.K.; Boschung, J.; Nauels, A.; Xia, Y.; Bex, V.; Midgley, P.M. (Eds.) *IPCC Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*; Cambridge University Press: Cambridge, UK; New York, NY, USA, 2013.
33. Harris, I.; Jones, P.D.; Osborn, T.J.; Lister, D.H. Updated high-resolution grids of monthly climatic observations—The CRU TS3.10 Dataset. *Int. J. Climatol.* **2014**, *34*, 623–642. [CrossRef]
34. Döll, P.; Kaspar, F.; Lehner, B. A global hydrological model for deriving water availability indicators: Model tuning and validation. *J. Hydrol.* **2003**, *270*, 105–134. [CrossRef]
35. Alcamo, J.; Döll, P.; Henrichs, T.; Kaspar, F.; Lehner, B.; Rösch, T.; Siebert, S. Development and testing of the WaterGAP 2 global model of water use and availability. *Hydrol. Sci. J.* **2003**, *48*, 317–337. [CrossRef]
36. Döll, P.; Fiedler, K.; Zhang, J. Global-scale analysis of river flow alterations due to water withdrawals and reservoirs. *Hydrol. Earth Syst. Sci.* **2009**, *13*, 2413–2432. [CrossRef]
37. Döll, P.; Fiedler, K. Global-scale modeling of groundwater recharge. *Hydrol. Earth Syst. Sci.* **2008**, *12*, 863–885. [CrossRef]
38. Döll, P.; Müller Schmied, H.; Schuh, C.; Portmann, F.T.; Eicker, A. Global-scale assessment of groundwater depletion and related groundwater abstractions: Combining hydrological modeling with information from well observations and GRACE satellites. *Water Resour. Res.* **2014**, *50*, 5698–5720. [CrossRef]

39. Haddeland, I.; Clark, D.B.; Franssen, W.; Ludwig, F.; Voß, F.; Arnell, N.W.; Bertrand, N.; Best, M.; Folwell, S.; Gerten, D.; et al. Multimodel estimate of the global terrestrial water balance: Setup and first results. *J. Hydrometeorol.* **2011**, *12*, 869–884. [CrossRef]
40. Schewe, J.; Heinke, J.; Gerten, D.; Haddeland, I.; Arnell, N.W.; Clark, D.B.; Dankers, R.; Eisner, S.; Fekete, B.M.; Colón-González, F.J.; et al. Multimodel assessment of water scarcity under climate change. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 3245–3250. [CrossRef] [PubMed]
41. Hagemann, S.; Chen, C.; Clark, D.B.; Folwell, S.; Gosling, S.N.; Haddeland, I.; Hanasaki, N.; Heinke, J.; Ludwig, F.; Voss, F.; et al. Climate change impact on available water resources obtained using multiple global climate and hydrology models. *Earth Syst. Dyn.* **2013**, *4*, 129–144. [CrossRef]
42. Döll, P.; Müller Schmied, H. How is the impact of climate change on river flow regimes related to the impact on mean annual runoff? A global-scale analysis. *Environ. Res. Lett.* **2012**, *7*. [CrossRef]
43. Portmann, F.T.; Döll, P.; Eisner, S.; Flörke, M. Impact of climate change on renewable groundwater resources: Assessing the benefits of avoided greenhouse gas emissions using selected CMIP5 climate projections. *Environ. Res. Lett.* **2013**, *8*. [CrossRef]
44. Alcamo, J.; Leemans, R.; Kreileman, E. *Global Change Scenarios of the 21st Century—Results from the IMAGE 2.1 Model*; Pergamon: Oxford, UK, 1998.
45. Wilber, A.C.; Kratz, D.P.; Gupta, S.K. *Surface Emissivity Maps for Use in Satellite Retrievals of Longwave Radiation*; NASA/TP-1999-209362; Langley Research Center: Hampton, VA, USA, 1999; 35p.
46. Legates, D.R.; McCabe, G.J. Evaluating the use of "goodness-of-fit" Measures in hydrologic and hydroclimatic model validation. *Water Resour. Res.* **1999**, *35*, 233–241. [CrossRef]
47. Krause, P.; Boyle, D.P.; Bäse, F. Comparison of different efficiency criteria for hydrological model assessment. *Adv. Geosci.* **2005**, *5*, 89–97. [CrossRef]
48. Möller, M.I. Vergleich von Modellierten Globalskaligen Rasterdaten der Strahlung Mit Stationsbasierten Messwerten: Methodenentwicklung und Analyse. Master's Thesis, Goethe-University Frankfurt, Frankfurt, Germany, 16 July 2015.
49. Gupta, H.V.; Kling, H.; Yilmaz, K.K.; Martinez, G.F. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *J. Hydrol.* **2009**, *377*, 80–91. [CrossRef]
50. Kling, H.; Fuchs, M.; Paulin, M. Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. *J. Hydrol.* **2012**, *424–425*, 264–277. [CrossRef]
51. Raschke, E.; Kinne, S.; Stackhouse, P.W. *GEWEX Radiative Flux Assessment (RFA) Volume 1: Assessment. A Project of the World Climate Research Programme Global Energy and Water Cycle Experiment (GEWEX) Radiation Panel*; WCRP Report 19/2012; World Meteorological Organization (WMO): Geneva, Switzerland, 2012.
52. Russak, V.; Niklus, I. Longwave radiation at the earth's surface in Estonia. *Proc. Estonian Acad. Sci.* **2015**, *64*, 480. [CrossRef]
53. Hori, M.; Aoki, T.; Tanikawa, T.; Motoyoshi, H.; Hachikubo, A.; Sugiura, K.; Yasunari, T.J.; Eide, H.; Storvold, R.; Nakajima, Y.; et al. In-situ measured spectral directional emissivity of snow and ice in the 8–14 μm atmospheric window. *Remote Sens. Environ.* **2006**, *100*, 486–502. [CrossRef]
54. Warren, S.G. Optical properties of snow. *Rev. Geophys.* **1982**, *20*, 67–89. [CrossRef]
55. Priestley, C.H.B.; Taylor, R.J. Assessment of surface heat flux and evaporation using large-scale parameters. *Mon. Weather Rev.* **1972**, *100*, 81–92. [CrossRef]

3 Sensitivity of simulated global-scale freshwater fluxes and storages to input data, hydrological model structure, human water use and calibration

Full citation: Müller Schmied, H., Eisner, S., Franz, D., Wattenbach, M., Portmann, F. T., Flörke, M. and Döll, P.: Sensitivity of simulated global-scale freshwater fluxes and storages to input data, hydrological model structure, human water use and calibration, Hydrol. Earth Syst. Sci., 18, 3511–3538, doi:10.5194/hess-18-3511-2014, 2014.

Declaration of portion of own work

H. Müller Schmied has shared with S. Eisner model code development. The model experiment setup was framed mainly between H. Müller Schmied and P. Döll. All model runs (except water use, which was done by S. Eisner and F. T. Portmann), analyses, figures and tables were done by H. Müller Schmied. The draft of the manuscript was written by H. Müller Schmied (except for small parts, e.g., description of water use by F. T. Portmann and the regionalization approach by S. Eisner) and revised by the co-authors. In overall percentage, H. Müller Schmied contributed around 80% of this manuscript.

Hannes Müller Schmied

Hydrol. Earth Syst. Sci., 18, 3511–3538, 2014
www.hydrol-earth-syst-sci.net/18/3511/2014/
doi:10.5194/hess-18-3511-2014

Hydrology and Earth System Sciences
Open Access

Sensitivity of simulated global-scale freshwater fluxes and storages to input data, hydrological model structure, human water use and calibration

H. Müller Schmied[1], S. Eisner[2], D. Franz[3,*], M. Wattenbach[3], F. T. Portmann[1,4,5], M. Flörke[2], and P. Döll[1]

[1]Institute of Physical Geography, Goethe-University Frankfurt, Altenhöferallee 1, 60438 Frankfurt am Main, Germany
[2]Center for Environmental Systems Research (CESR), University of Kassel, Wilhelmshöher Allee 47, 34117 Kassel, Germany
[3]GFZ German Research Centre for Geosciences, Section 5.4 Hydrology, Telegrafenberg, 14473 Potsdam, Germany
[4]Biodiversity and Climate Research Centre (LOEWE BiK-F), Senckenberganlage 25, 60325 Frankfurt am Main, Germany
[5]Senckenberg Research Institute and Natural History Museum, Senckenberganlage 25, 60325 Frankfurt am Main, Germany
[*]now at: Inorganic and Isotope Geochemistry, GFZ Potsdam, Potsdam, Germany

Correspondence to: H. Müller Schmied (hannes.mueller.schmied@em.uni-frankfurt.de)

Received: 8 January 2014 – Published in Hydrol. Earth Syst. Sci. Discuss.: 7 February 2014
Revised: 9 June 2014 – Accepted: 23 July 2014 – Published: 10 September 2014

Abstract. Global-scale assessments of freshwater fluxes and storages by hydrological models under historic climate conditions are subject to a variety of uncertainties. Using the global hydrological model WaterGAP (Water – Global Assessment and Prognosis) 2.2, we investigated the sensitivity of simulated freshwater fluxes and water storage variations to five major sources of uncertainty: climate forcing, land cover input, model structure/refinements, consideration of human water use and calibration (or no calibration) against observed mean river discharge. In a modeling experiment, five variants of the standard version of WaterGAP 2.2 were generated that differed from the standard version only regarding the investigated source of uncertainty. The basin-specific calibration approach for WaterGAP was found to have the largest effect on grid cell fluxes as well as on global AET (actual evapotranspiration) and discharge into oceans for the period 1971–2000. Regarding grid cell fluxes, climate forcing ranks second before land cover input. Global water storage trends are most sensitive to model refinements (mainly modeling of groundwater depletion) and consideration of human water use. The best fit to observed time series of monthly river discharge or discharge seasonality is obtained with the standard WaterGAP 2.2 model version which is calibrated and driven by daily reanalysis-based WFD/WFDEI (combination of Watch Forcing Data based on ERA40 and Watch Forcing Data based on ERA-Interim) climate data. Discharge computed by a calibrated model version using monthly CRU TS (Climate Research Unit time-series) 3.2 and GPCC (Global Precipitation Climatology Center) v6 climate input reduced the fit to observed discharge for most stations. Taking into account uncertainties of climate and land cover data, global 1971–2000 discharge into oceans and inland sinks ranges between 40 000 and 42 000 $km^3\,yr^{-1}$. Global actual evapotranspiration, with 70 000 $km^3\,yr^{-1}$, is rather unaffected by climate and land cover uncertainties. Human water use reduced river discharge by 1000 $km^3\,yr^{-1}$, such that global renewable water resources are estimated to range between 41 000 and 43 000 $km^3\,yr^{-1}$. The climate data sets WFD (available until 2001) and WFDEI (starting in 1979) were found to be inconsistent with respect to shortwave radiation data, resulting in strongly different actual evapotranspiration. Global assessments of freshwater fluxes and storages would therefore benefit from the development of a global data set of consistent daily climate forcing from 1900 to present.

Published by Copernicus Publications on behalf of the European Geosciences Union.

1 Introduction

The quantification of global-scale freshwater fluxes, in particular river discharge, is essential to assess availability and scarcity of water resources for humans and the environment for both present (Hoekstra et al., 2012; Oki and Kanae, 2006; Prudhomme et al., 2014) and scenario conditions (Döll and Müller Schmied, 2012; Masaki et al., 2014; Schewe et al., 2014). Further examples are the estimation of amounts and spatial distribution of precipitation (Harris et al., 2014; Schneider et al., 2014) and evapotranspiration (Jasechko et al., 2013; Jung et al., 2010; Sterling et al., 2012). As groundwater plays an important role for humans, e.g., for irrigation purposes, related fluxes such as groundwater recharge (Döll and Fiedler, 2008; Koirala et al., 2014; Portmann et al., 2013) or, as consequence of an overexploitation of groundwater resources, groundwater depletion (Döll et al., 2014b; Wada et al., 2010) are the focus of modeling activities.

These examples show attempts to quantify global-scale freshwater fluxes as well as water storages; however, the methodologies used differ largely. Interpolation of in situ measurements works well with a dense monitoring network. Hence, precipitation products (e.g., GPCC – Global Precipitation Climatology Center; Schneider et al., 2014; CRU – Climate Research Unit; Harris et al., 2014) are often based solely on interpolation of station data. In combination with other data sources like remote sensing, even less dense point measurements are used, e.g., to quantify global evapotranspiration (Jung et al., 2010). In particular, remote sensing is used to derive spatiotemporal input data for evapotranspiration schemes (Miralles et al., 2011; Vinukollu et al., 2011; Wang and Liang, 2008) or to assess total continental water storage variations (Schmidt et al., 2006). Spatiotemporal patterns of consistent multiple fluxes and storages can be obtained using land surface models (LSMs) and global hydrological models (GHMs). LSMs, which have evolved as "land components" of global circulation models (GCMs), usually have a high temporal resolution and solve the energy balance (Haddeland et al., 2011). However, they do have limitations, especially in runoff routing and with regard to human alterations of the water cycle (even though there are exceptions, e.g., Pokhrel et al., 2012). GHMs are explicitly designed to assess the state of freshwater resources and to address water-related problems like floods and droughts (Corzo Perez et al., 2011; Prudhomme et al., 2011) and human impacts on freshwater resources. In the last 20 years, a number of GHMs have been developed using different conceptual approaches; e.g., VIC (Variable Infiltration Capacity; Nijssen et al., 2001), WBM (Water Balance Model; Vörösmarty et al., 1998), Mac-PDM (Macro Probability Distribution Model; Gosling and Arnell, 2011), WASMOD-M (Water And Snow balance Modeling system – Macro scale; Widén-Nilsson et al., 2007), H08 (Hanasaki et al., 2008), WaterGAP (Water – Global Assessment and Prognosis; Alcamo et al., 2003; Döll et al., 2003) and PCR-GLOBWB (PCRaster GLOBal Water Balance; Sperna Weiland et al., 2010).

Freshwater fluxes and storages simulated by GHMs and LSMs are subject to several sources of uncertainty: spatially distributed input data (e.g., climate forcing, water use, and land cover), model structure (or modeling approach) and model parameters. In addition, epistemic uncertainty due to a lack of knowledge and understanding of processes is of particular importance at the global scale (see discussion in Beven and Cloke, 2012 and Wood et al., 2011, 2012).

Uncertainties due to the choice of climate forcing were the focus of few studies. For example, Guo et al. (2006) showed the large sensitivity of soil moisture simulated by 11 LSMs to different climate forcing data sets (especially to precipitation and radiation). They concluded that this uncertainty associated with land surface hydrology is as large as the variation among the LSMs. Biemans et al. (2009) evaluated seven global precipitation products for 294 river basins worldwide and quantified an average uncertainty of 30 % per basin. They studied the dynamic global vegetation and hydrology model LPJmL (Lund–Potsdam–Jena managed Land; Bondeau et al., 2007) with these precipitation forcings. As consequence of the large spread of precipitation input, discharge uncertainty was quantified with about 90 %. Even though climate forcing is of such importance, only few studies are available which study the uncertainty in a global hydrological model setup.

Uncertainties in terms of model structure are related to the design of the model, i.e., the processes considered and their representation by conceptual approaches. To consider this kind of uncertainty, Butts et al. (2004), Clark et al. (2008), Refsgaard et al. (2006) and Song et al. (2011) developed approaches to diagnose different structures of hydrological models and related uncertainties. The model intercomparison efforts WATCH WaterMIP (Water and Global Change Water Model Intercomparison Project; Haddeland et al., 2011) and ISI-MIP (Inter-Sectoral Impact-MIP; Warszawski et al., 2014) have shown the effects of different model structures (Gudmundsson et al., 2012a, b; Hagemann et al., 2013; Van Loon et al., 2012; Prudhomme et al., 2014; Schewe et al., 2014) even though this was not explored systematically. For example, values for global annual evapotranspiration between 60 000 and 85 000 $km^3\,yr^{-1}$ were reported in the WATCH WaterMIP study (Haddeland et al., 2011). In such multimodel studies, many completely different models are participating, which makes it very difficult to identify the reasons for different model behavior. A sensitivity study using basically the same model but with a refined model structure can therefore be of benefit (e.g., Thompson et al., 2013).

Model parameters are used to represent system dynamics in solvable equations, in particular when the hydrological process cannot be described physically. These parameters are generally not measurable and, hence, are a source of uncertainty that can influence model results to varying degrees. Within the GCM community, the perturbed physics ensemble

approach has been used to assess this kind of uncertainty in a structured way (Collins et al., 2006; Rowlands et al., 2012). Global-scale hydrology applications also assess parameter uncertainty. For example, Gosling and Arnell (2011) used seven sets of parameter perturbations for two model parameters of the GHM Mac-PDM.09 (Macro-scale – Probability-Distributed Mositure model.09). For the GHM WaterGAP, Kaspar (2003) investigated the impact of uncertainty of 38 model parameters on simulated river discharge by conducting various model runs with a sampling of parameter values within specific ranges. He found that major uncertainties are related to evapotranspiration parameters and land cover specific attributes. Schumacher et al. (2014) confirmed the sensitivity of model output (here: monthly total water storage) to radiation calculation and related parameters in WaterGAP which, together with a river roughness coefficient and precipitation, dominate uncertainty in many of the 33 investigated river basins. Groundwater-related parameters and soil parameters were found to be important for the timing and variation of total water storages in WaterGAP (Werth and Güntner, 2010).

Model parameters can be adjusted by calibration, such that model output matches an observed set of data. Whereas basin-scale hydrological models are routinely calibrated against observed river discharge (e.g., Beven, 2001), this is only seldom the case for GHMs. Widén-Nilsson et al. (2007) used different model parameter sets within the GHM WASMOD-M to define optimal parameter values on river-basin scale. WaterGAP is calibrated against observed river discharge in a basin-specific manner by varying one soil parameter (and up to two correction factors) (Döll et al., 2003; Hunger and Döll, 2008).

In this paper, we analyze and quantify the uncertainty in simulated global-scale freshwater water fluxes and storages due to (1) spatially distributed input data, and (2) model structure and modeling approach, using the most recent version of the GHM WaterGAP 2.2. Previous studies (Kaspar, 2003; Schumacher et al., 2014; Werth and Güntner, 2010) have already investigated both parameter sensitivity and uncertainty for WaterGAP. To assess recently available climate forcing and land cover input data as well as significant modifications of WaterGAP model structure during the last decade were major motivations of this experiment.

In particular, we will answer the following research questions:

1. How sensitive are freshwater fluxes and water storages to spatially distributed input data (climate forcing, land cover)?
2. What are the benefits of WaterGAP model structure refinements implemented during the last decade?
3. How does the modeling approach (calibration procedure, consideration of human water use) affect freshwater fluxes and water storages?

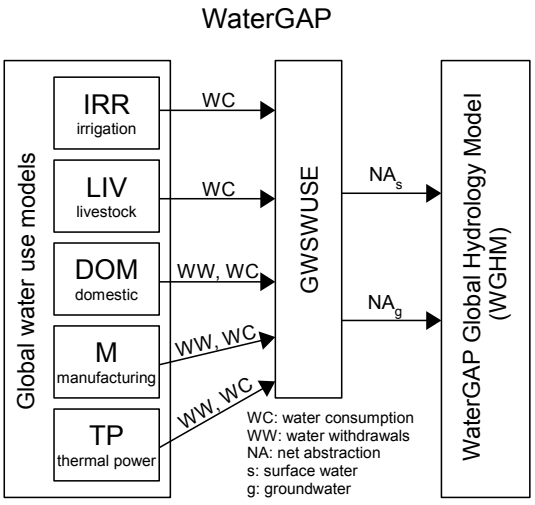


Figure 1. Schematic of WaterGAP 2.2. The output of five water use models is translated into net abstractions from groundwater NA_g and surface water NA_s by the submodel GWSWUSE, which allows computing the impact of human water use on water flows and storages by WGHM. For details see Döll et al. (2012).

4. Which type of uncertainty is dominant for specific fluxes and variations of total water storage?

After an initial description of WaterGAP 2.2 (for details see the Appendix), the experimental setup is explained (Sect. 2). In Sect. 3, the results are described; focusing on the effect of the different model variants on global freshwater fluxes and water storages as well as spatial patterns. In Sect. 4, we discuss the results with regard to the research questions. The paper ends with a summary and conclusions (Sect. 5).

2 Methods and experiment setup

2.1 Description of WaterGAP 2.2

The global hydrology and water use model WaterGAP (Fig. 1) consists of two major parts: the water use models for five different sectors (Appendix C) and the WaterGAP Global Hydrology Model (WGHM, Fig. A1). The submodel GroundWater Surface Water USE (GWSWUSE) (Appendix D) is used to distinguish water use from groundwater and surface water sources and computes net abstractions from both sources which are an input to WGHM (Fig. 1). Using a number of water storage equations (change of storage over time equals to inflow minus outflows; Appendix A), WGHM calculates daily water flows and storages at a spatial resolution of 0.5° by 0.5° (55 km by 55 km at the Equator) for the whole land area of Earth except Antarctica (66 896 cells). WaterGAP 2.2 is calibrated against mean annual river discharge at 1319 gauging stations, and the adjusted calibration factor is regionalized to grid cells outside the calibration basins (Appendix B).

Since the initial publication of WaterGAP 2.1d (Döll et al., 2003), major changes were made to keep the model up to date. For example, algorithms of reservoir operation were included (Döll et al., 2009), groundwater recharge was optimized by distinguishing semiarid/arid regions from humid regions (Döll and Fiedler, 2008), a variable flow velocity algorithm was included (Verzano et al., 2012) and the source of abstracted water was considered (Döll et al., 2012).

2.2 Experiment setup

Six WaterGAP model variants (Table 1) were designed as follows. The standard version of WaterGAP 2.2 (STANDARD) was modified regarding only one aspect, including either alternative climate forcing (CLIMATE), land cover input (LANDCOVER) or model structure (STRUCTURE). Each model variant was independently calibrated. Variant NoCal is an uncalibrated simulation with the standard version of WaterGAP 2.2 to study the impact of the calibration approach. Variant NoUse reflects naturalized water flows and storages without the impact of human water use, and thus also renewable water resources.

In addition, for assessing the effect of uncertainties on renewable water resources, variants CLIMATE, LANDCOVER, STRUCTURE and NoCal are also run without considering any water abstractions. The simulation period is 1901–2009 but model results are evaluated only for 1971–2000.

2.2.1 Climate input

Climate forcing data for global-scale hydrological models are a major source of uncertainty for two main reasons: (1) they are subject to measurement errors which were not corrected in the original input data and (2) they are subject to interpolation errors due to low spatial and temporal monitoring network density and/or because (temporal) data gaps have to be filled. To analyze the sensitivity of calibration and simulated freshwater fluxes to different climate forcing data sets, two climate forcings were used to force both WGHM and the Global Irrigation Model GIM (Döll and Siebert, 2002) (Appendix C).

In variant STANDARD, the daily WATCH Forcing Data methodology applied to ERA-40 (40-year European Centre for Medium-Range Weather Forecasts (ECMWF) Re-Analysis) data (WFD) (Weedon et al., 2011) for the years 1901–1978 (the years 1901–1957 are based on reordered reanalysis data) and the WATCH Forcing Data methodology applied to ERA-Interim reanalysis data (WFDEI) (Weedon et al., 2014) for the years 1979–2009 was chosen. Switching the climate input data set in 1979 leads to inconsistencies in terms of AET (actual evapotranspiration; higher in WFDEI) and therefore affects the storages until a new equilibrium is reached (see Sect. 3.1). WFD and WFDEI monthly sums/means are bias-corrected with other data sources (temperature bias correction, shortwave radiation adjustment using cloud cover and adjustment of number of wet days to CRU TS 2.1 for WFD and to CRU TS 3.1 for WFDEI as well as adjustment of monthly precipitation sum to GPCC v4 for WFD and GPCC v5 for WFDEI and snowfall undercatch corrected after Adam and Lettenmaier, 2003). To calculate net shortwave radiation, the incoming shortwave radiation is reflected by literature-based land cover specific albedo values (see Table A2). Literature-based-emissivity values for all land cover classes (Wilber et al., 1999) and the Stefan–Boltzmann equation are used to calculate outgoing longwave radiation. The difference to incoming longwave radiation represents net longwave radiation. Net radiation is the sum of both components.

In variant CLIMATE, the monthly data set CRU TS (time-series) 3.2 (Harris et al., 2014) was used but monthly precipitation totals were replaced by the latest GPCC v6 precipitation monitoring product (Schneider et al., 2014) because it includes more observation stations. Monthly means are disaggregated to daily values within WaterGAP (Döll et al., 2003). Neither CRU nor GPCC precipitation is corrected for observational errors, e.g., wind-induced-precipitation undercatch. Thus, Döll and Fiedler (2008) included the catch ratios of Adam and Lettenmaier (2003) and used the empirical function of Legates (1987) to correct especially snow undercatch by dividing snow and liquid precipitation using a temperature-based approach. The correction of precipitation measurement bias leads to an average increase of 8.7 % compared to the original product. On 37.5 % of the land area (except Greenland and Antarctica), the increase of precipitation is larger than 10 %. Differences of mean values from both data sets (CRU/GPCC and WFD/WFDEI) occur due to the slightly different precipitation correction approach and the GPCC version used for scaling monthly sums. Monthly precipitation is equally distributed to the number of wet days provided by the CRU 3.2 data set; the distribution of wet days within a month is modeled as a two-state, first-order Markov chain (Döll et al., 2003). Cloudiness fraction was used to calculate incoming shortwave radiation as well as outgoing longwave radiation after Shuttleworth (1993), see also Döll et al. (2003).

2.2.2 Land cover input data

The distribution of land cover classes and associated attributes are affecting simulated fluxes in terms of radiation energy balance (albedo and emissivity), snow dynamics (degree-day factor D_F), available soil water capacity (rooting depth) and interception capacity (L) (for details see Appendix A). To estimate the effect of different, homogeneous-source land cover data, two input maps were used (Fig. 2). Attributes and model parameters associated to land cover classes were derived from literature or previous model versions (Table A2) and left equal in both variants.

Table 1. Overview of the model variants.

Name	Characteristic	Description
STANDARD	Standard WaterGAP 2.2 model version	MODIS land cover for the year 2004. WATCH Forcing Data as daily climate input. For 1901–1978 WFD is used, for 1979–2009 WFDEI. Calibration against mean annual river discharge, including regionalization of the calibration parameter to grid cells outside calibration basins. Consideration of human water use.
CLIMATE	Alternative climate forcing	Similar to STANDARD but CRU TS 3.2 and GPCC v6 for precipitation as monthly climate input.
LANDCOVER	Alternative land cover data	Similar to STANDARD but a combination of GLCC and CORINE (for Europe) was used as land cover input.
STRUCTURE	Alternative model structure	Similar to STANDARD but with a less-refined process representation (comparable to Döll et al., 2003).
NoUse	No water use	Similar to STANDARD but without considering water use.
NoCal	No calibration	Similar to STANDARD but without calibration to mean annual river discharge. Calibration parameter and correction factors are globally set to 1.0 (for details see Appendix B).

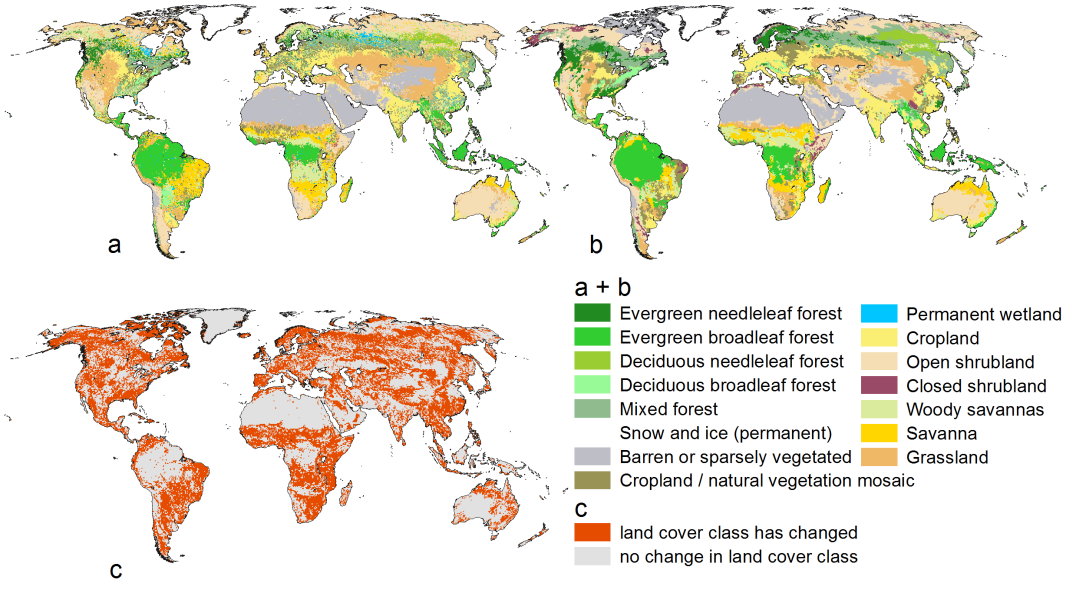


Figure 2. Land cover maps with a spatial resolution of 0.5° used as WaterGAP input based on MODIS observations for the year 2004 (variant STANDARD) **(a)**, land cover derived from USGS GLCC, but CORINE for Europe, reflecting land cover distribution around the year 2000 (variant LANDCOVER) **(b)**, and identification of grid cells where land cover class has changed due to different input data **(c)**.

In variant STANDARD we used the gridded MODIS (Moderate Resolution Imaging Spectroradiometer) land cover product (MOD12Q1) for the year 2004. The product MOD12Q1 (1 km resolution, global coverage up to 80° N) was used with land cover type 1 according to the International Geosphere-Biosphere Programme (IGBP) classification. After resampling to 0.5°spatial resolution, the data set was reclassified to fit to the WaterGAP land cover classification system (Table A2). As water bodies (from the global lakes and wetlands database, GLWD; Lehner and Döll, 2004) and percentage of urban area (from previous model versions) are obtained by additional input files, the second land cover class was appointed in case of "water" or "urban and built-up" as primary land cover. For coastal grid cells, which are not fully covered by MODIS and north of 80° N, a combination of Global Land Cover Characteristics database GLCC (USGS, 2008) + CORINE (Coordination of Information on the Environment) land cover information was used.

In variant LANDCOVER, a combination of the GLCC based on the years 1992 and 1993 and, for Europe, CORINE land cover based on the year 2000 (European Environment Agency, 2004) was used as land cover information, as also in a previous WaterGAP version (Haddeland et al., 2011). The idea was to use an IGBP-based classification scheme and a remote-sensing-based land cover distribution instead of IMAGE (Integrated Model to Assess the Global Environment; Alcamo et al., 1998) model outputs (as in previous model versions). Both input data sets have a resolution of 1 by 1 km and were aggregated to the 0.5° model resolution by assigning the majority land cover type.

2.2.3 Structural model changes

During the last 10 years, the WaterGAP model was subject to several revisions and improvements in terms of hydrologic process representation, resulting in an overall more complex model structure. To assess the sensitivity of simulated freshwater fluxes to model complexity, one model variant with a simplified structure comparable to Döll et al. (2003) (variant STRUCTURE) was set up which was run with the same input data as all other model variants. Differences of variant STRUCTURE as compared to the other variants are as follows.

- Flow velocity is globally set to $1\,\mathrm{m\,s^{-1}}$ and the meandering ratio is set to 1.0, instead of the variable flow velocity algorithm of Verzano et al. (2012) in the other variants.
- Reservoirs are treated as global lakes, i.e., the reservoir operation algorithm of Döll et al. (2009) is not used, which should result in a more dynamic discharge downstream of reservoirs.
- Water for human water use is abstracted only from surface water bodies; i.e., there are no groundwater abstractions as introduced by Döll et al. (2012).
- Evaporation from lakes/wetlands is not adjusted by reduction factors (Hunger and Döll, 2008) resulting in evaporation at potential rate even at low storage.
- Snow accumulation and melt are modeled on 0.5° (instead of the 3 arcmin subgrid (Schulze and Döll, 2004)) which should lead to less snow dynamics.
- Finally, there is no distinction in groundwater recharge for semiarid/arid and humid regions (in contrast to Döll and Fiedler, 2008, all regions are treated like humid regions) resulting in higher groundwater recharge in semiarid/arid regions.

2.2.4 Human water use

In many areas of the globe, human water use significantly affects water flows and storage. In this experiment, all model variants except NoUse and STRUCTURE are taking into account water use from surface water and groundwater resources. In variant NoUse it is assumed that there are no water abstractions at all, while in STRUCTURE, water is only abstracted from surface water (as formerly no information on the source of water abstractions was available).

2.2.5 Calibration

WGHM is calibrated against mean annual discharge in a basin-specific manner by adjusting a runoff coefficient that affects the outflow from the soil compartment in each grid cell of the 1319 calibration basins. If necessary to simulate mean annual discharge within 1 % of the observed value, two additional correction factors are adjusted (for details see Appendix B). All other parameters are globally uniform (or land cover class dependent), based on literature or experiences from past studies; i.e., there is no basin or region specific modification. All model variants except NoCal are independently calibrated to the same observational data. In variant NoCal, the runoff coefficient and both correction factors are set to 1.0 in all grid cells. The comparison of NoCal to, for example, STANDARD allows for a direct quantification of the effect of calibration on simulated water fluxes and storages.

3 Results

3.1 Global water balance

Table 2 lists global values for various components of the global water balance and changes in total water storages (TWS) (calculated excluding Antarctica, Greenland and inland sinks) as estimated by the different model variants. Global values vary mainly due to calibration and selected climate forcing. For interpreting Table 2 and Fig. 3 it is important to mention that AET does not include additional evapotranspiration caused by irrigation and other human water use. This part of evapotranspiration is called actual water consumption (WC_a). For computing global values of AET and renewable water resources (RWR), the values were adjusted in calibration basins using the station correction factor (CFS) such that a closed global water balance is achieved (for calibration details see Appendix B). Grid cell values of AET and RWR (Figs. 3, 4), however, do not reflect CFS to avoid physically implausible values that likely result from inconsistencies between precipitation data and observed river discharge. Global precipitation P is about $1900\,\mathrm{km^3\,yr^{-1}}$ (or 1.7 %) higher when using the CLIMATE model variant which results in an equal increase of discharge compared to STANDARD. Except for NoCal, global AET (calculated as sum of E_c, E_{sn}, E_s and E_w, see Appendix A) does not vary considerably among the variants. In general, discharge to oceans and inland sinks is lower by the amount of change in AET. WC_a (row 4 in Table 2) varies due to the demand of surface

Table 2. Long-term-average (1971–2000) freshwater fluxes from global land area (except Antarctica and Greenland) of WaterGAP 2.2 (in $km^3\ yr^{-1}$). Cells representing inland sinks were excluded but discharge into inland sinks was included.

No.	Component	STANDARD	NoUse[h]	CLIMATE	LANDCOVER	STRUCTURE	NoCal
1	Precipitation P[a]	111 070	111 070	112 969	111 070	111 070	111 070
2	Actual evapotranspiration AET[b]	69 803	69 934	69 842	70 012	70 217	63 344
3	Discharge into oceans and inland sinks Q[c]	40 458	41 216	42 364	40 250	40 002	46 822
4	Water consumption (actual) (rows 5 + 7) WC_a	1031	0	927	1029	983	1054
5	Net abstraction from surface water (actual)[d]	1102	0	960	1102	983	1126
6	Net abstraction from surface water (demand) NA_s[e]	1154	0	1000	1154	1082	1154
7	Net abstraction from groundwater NA_g[f]	−72	0	−33	−72	0	−72
8	Change of total water storage dS/dt[g]	−215	−73	−156	−214	−44	−143
9	Long-term-averaged yearly volume balance error ($P - AET - Q - WC_a - dS/dt$)	−7	−7	−8	−7	−88	−7
	deviation to P	−0.006 %	−0.006 %	−0.007 %	−0.006 %	−0.08 %	−0.006 %

[a] Mean annual P (1979–2001) is 110 309 $km^3\ yr^{-1}$ in WFD and 110 812 $km^3\ yr^{-1}$ in WFDEI; [b] AET does not include evapotranspiration caused by human water use, i.e., actual water consumption WC_a; [c] including anthropogenic water use (except NoUse); [d] if not enough water is available, demand is not completely satisfied; [e] demand that needs to be satisfied (water use model output); [f] negative values indicate that return flows from irrigation with surface water exceed groundwater abstractions; [g] TWS of 31 December 2000 minus TWS of 31 December 1970 divided by 30 years; [h] STANDARD but no subtraction of water use; i.e., discharge into oceans and inland sinks equals renewable water resources.

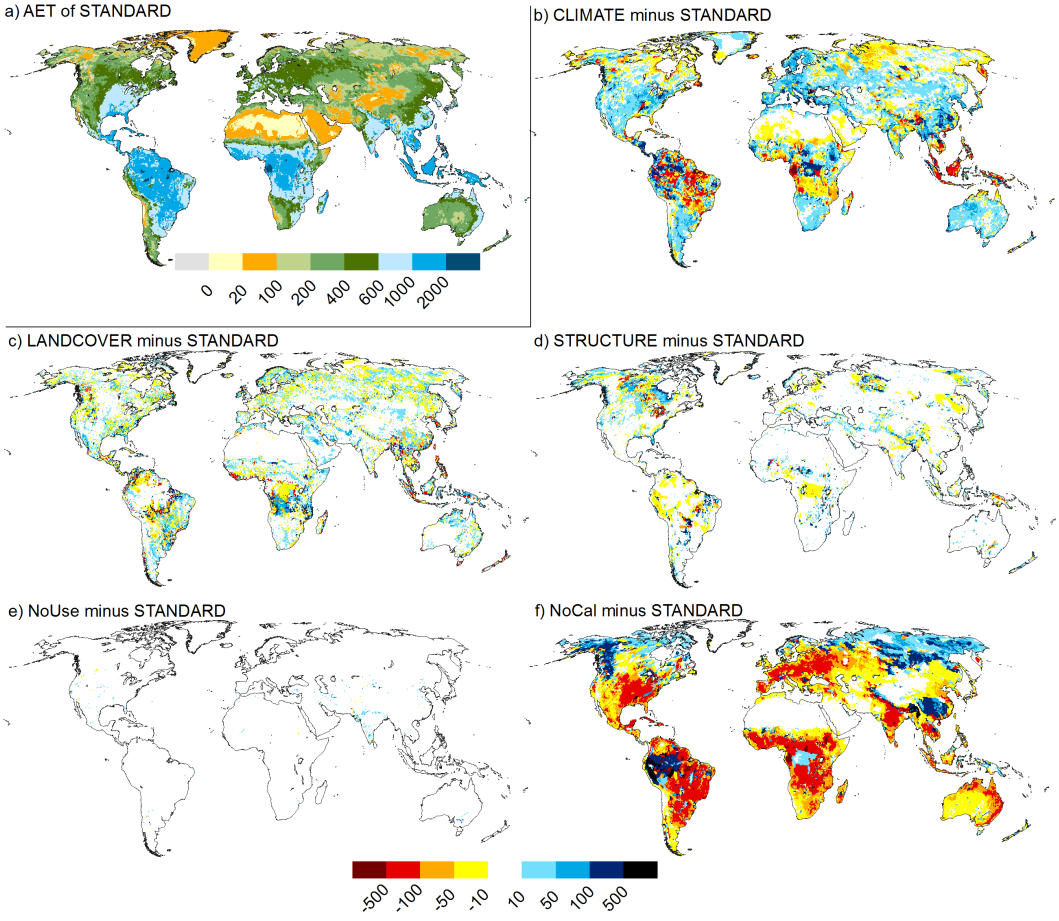


Figure 3. Actual evapotranspiration AET for STANDARD (mean value 1971–2000, in $mm\ yr^{-1}$) **(a)** and differences between the model variants and STANDARD (in $mm\ yr^{-1}$) **(b–f)**.

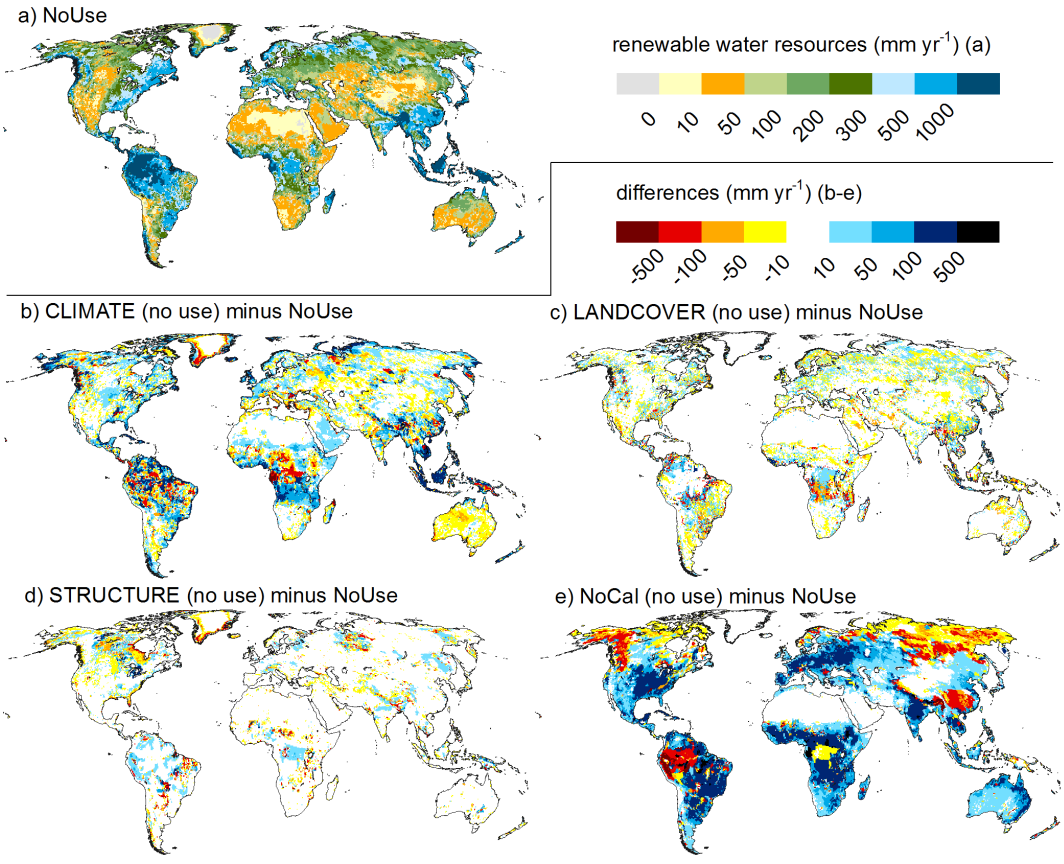


Figure 4. Renewable water resources (mean annual runoff from each cell if water use is neglected) calculated by the WaterGAP 2.2 NoUse variant **(a)** and differences to other variants (variants here run without considering water use) **(b–e)**.

water abstractions and groundwater abstractions (which differs in CLIMATE due to the forcing of GIM (Appendix C) and in STRUCTURE where water demand is entirely extracted from surface water resources) and due to the different water availability for abstractions. In all cases, a large share of the total water demand could be satisfied (between 90 % in STRUCTURE and 96 % in CLIMATE).

When human water use is not taken into account (NoUse), AET increases by 131 $km^3\,yr^{-1}$ because evaporation from open water bodies increases as they are not depleted by water uses and additional evapotranspiration of irrigated crops is not included in AET (but quantified within WC_a; row 4 in Table 2). As expected, river discharge is higher (by 758 $km^3\,yr^{-1}$) in NoUse. Changes in total water storages (142 $km^3\,yr^{-1}$ less storage decrease) are also visible, especially due to no groundwater withdrawals in this variant (Table 3). The sum of these differences between STANDARD and NoUse is 1031 $km^3\,yr^{-1}$ which equals to WC_a (row 4 for STANDARD in Table 2).

The calibration has a strong effect on freshwater fluxes. Global discharge to oceans and inland sinks (Q) in NoCal is about 6400 $km^3\,yr^{-1}$ (or 15.7 %) higher than in STANDARD, meaning that the main effect of calibration is lowering discharge. In many river basins, the calibration parameter γ is higher than the value 1.0 globally used in NoCal which reduces the share of effective precipitation actually contributing to runoff. Consequently, AET is lower by nearly the same amount.

When comparing CLIMATE to STANDARD, both P and Q increased by approximately 1900 $km^3\,yr^{-1}$ whereas global AET sums are nearly equal. Most additional Q (1546 of overall 1906 $km^3\,yr^{-1}$) is generated in noncalibrated grid cells mostly because of an increased P (which explains 1200 of the additional 1546 $km^3\,yr^{-1}$) and a reduced AET (which explains 282 of the additional 1546 $km^3\,yr^{-1}$) in these grid cells.

RWR equal the long-term-averaged discharge to oceans and inland sinks (Q in Table 2) but without considering human water withdrawals. For the STANDARD model variant, RWR are 1.9 % higher than with WC_a (row 3 in Table 2, column NoUse and STANDARD). Q of the other model variants and hence RWR increase by about a similar value (NoCal 2.0 %, LANDCOVER and STRUCTURE 1.9 %, CLIMATE 1.6 %; values not shown in Table 2).

The decreasing trends of total water storage are mainly caused by groundwater depletion, except in variants NoUse

Table 3. Mean change in water storage in different compartments between 31 December 1970 and 31 December 2000 (in $km^3\,yr^{-1}$; global sum except Antarctica and Greenland). Cells representing inland sinks were excluded.

Compartment	STANDARD	NoUse*	CLIMATE	LANDCOVER	STRUCTURE	NoCal
Total water storage	−214.8	−73.7	−156.4	−214.8	−44.5	−143.0
Canopy	−0.05	−0.05	0.002	−0.05	−0.05	−0.05
Snow	−3.0	−3.0	−6.3	−3.3	−1.3	−3.0
Soil	−21.6	−21.6	−0.9	−20.6	−20.9	−20.0
Groundwater	−124.9	8.6	−126.9	−125.4	9.7	−82.7
Local lake	−1.9	−1.5	−0.3	−1.9	−2.1	−1.1
Local wetland	−4.9	−4.3	1.9	−5.1	−8.4	−2.2
Global lake	−3.5	−3.4	−1.1	−3.4	−8.2	−3.8
Reservoirs	−43.1	−37.5	−23.2	−43.1	**	−21.4
Global wetlands	−4.9	−4.3	1.9	−5.1	−8.4	−2.2
River	−6.7	−6.0	2.7	−6.7	−4.6	−4.3

* In WaterGAP, increase of soil water storage by irrigation is not taken into account such that storage values for STANDARD and NoUse variants are the same. ** Not applicable as reservoirs are treated as global lakes.

and STRUCTURE where no groundwater abstraction is modeled. Interestingly, NoCal shows a smaller decrease in groundwater storage than STRUCTURE. This is also due to the calibration parameter γ which is on average lower in case of NoCal. The lower γ, the more water leaves the soil and can subsequently contribute to groundwater recharge. Note that water abstractions from groundwater are taken directly from the groundwater storage and also return flows are added directly to groundwater storage (without passing the soil compartment). Hence, there is no difference in soil water storage between STANDARD and NoUse (Table 3).

Except for groundwater and snow, CLIMATE shows less storage depletion than all other variants that are forced by WFD/WFDEI (Table 3). The strong decrease in case of WFD/WFDEI is an artifact caused by combining WFD before 1979 with WFDEI after 1979. With WFDEI based on the ERA-Interim, AET is approximately 70 000 $km^3\,yr^{-1}$, compared to 65 000 $km^3\,yr^{-1}$ in the case of WFD. This is caused by differences in the shortwave downward radiation (much higher in WFDEI) which impacts the net radiation as the main input for calculating potential evapotranspiration after Priestley and Taylor (1972). As all model runs are started in 1901, the storages are more or less in equilibrium until 1978. AET is increased in the following 22 years by ca. 10 %, which leads to a higher water loss and therefore to a reduction of all storage compartments. For all storages except snow, reservoirs and groundwater, a new equilibrium is achieved a few (approximately 5) years after 1979 on a lower level (STANDARD variant). Whereas snow storage is not influenced at all, groundwater storage is affected by groundwater depletion and reservoirs by water use and obvious limitations of the reservoir algorithm. Thus, an equilibrium is not reached in the global average of the latter two storages but decreasing after 1901.

3.2 Actual evapotranspiration

Mean AET shows the highest values around the Equator, consistent with available energy, except for the Pacific Rim of South America (Fig. 3a). Among the variants, the largest differences to STANDARD occur in the case of the uncalibrated version NoCal (Fig. 3f). As the calibration approach also affects grid cells outside of the 1319 calibration basins due to the regionalization (Appendix B3), all grid cells are affected. In most regions, calibration leads to higher AET but in the upstream Amazon, Congo, Arctic river basins and some other basins, the opposite is true. The global sum of AET of NoCal is 9.2 % lower than estimated with STANDARD (Table 2). Notable differences in AET also occur when using an alternative climate input (Fig. 3b). AET increases in CLIMATE on 42.6 % of the land surface by more than 10 mm yr^{-1} and decreases by more than 10 mm yr^{-1} on 30.5 % of the land surface. It increases (decreases) by more than 100 mm yr^{-1} on 5.4 % (5.6 %) of the land surface. When summed globally, only minor changes in AET occur in the CLIMATE variant (increase of 0.06 % or 39 $km^3\,yr^{-1}$; Table 2). In contrast, AET differences of the STRUCTURE variant are higher for the global sum (increase of 0.6 % or 414 $km^3\,yr^{-1}$) but occur on an overall smaller area (increase by more than 10 mm yr^{-1} on 11.9 % of the land surface, decrease on 14.2 %). The effect of STRUCTURE is visible in areas with surface water bodies and in snow-dominated areas. On the one hand, an increase in net radiation in snowy regions leads to a slight increase of AET but in small absolute numbers as total AET is comparatively low. On the other hand, effects due to the evaporation reduction factor for surface water bodies are visible. In all variants, except STRUCTURE, evaporation is limited when the surface water body storage is reduced to mimic the shrinking of surface area. Hence, in regions with a high percentage/volume of surface water bodies, AET is increased. In addition, more complex

effects occur. The Great Lakes, for example, evaporate with potential evapotranspiration E_P (Appendix A1) in STRUCTURE, even when the lake storage is relatively low. This results in a relatively low modeled discharge which fits well to the observed ones. Hence, no correction factor (neither CFA nor CFS) is required in the Great Lakes Basin. However, in STANDARD, the reduction factor reduces evaporation by up to three-fourths of E_P. The resulting higher modeled discharge has to be reduced by an increased AET in STANDARD (and in the other model variants) on the land around the lakes as compared to STRUCTURE (red areas around the Great Lakes in Fig. 3).

Differences between NoCal and STANDARD are caused by the calibration parameter γ which differs from 1.0 (NoCal) in most river basins of STANDARD (and of the other model variants). For example, there are blue patterns in China and South America. In both regions, γ is lower than 1.0 in STANDARD which results in higher runoff and less modeled AET. In many other regions (red areas), γ is greater than 1.0 in STANDARD.

AET differences between LANDCOVER and STANDARD (Fig. 3c) are caused by changes in net radiation in energy-limited areas (not shown) as well as changes in rooting depth. In general, minor differences occur (except in some basins; see explanation below). In some regions, an increasing net radiation results in an increasing AET, e.g., in parts of Angola. In water-limited areas (e.g., northeastern Brazil), insignificant changes of AET occur even if net radiation strongly increases. In northern Australia, AET increases even when net radiation is reduced. Here, large parts are defined in STANDARD as open shrubland (rooting depth of 0.5 m) and in LANDCOVER as savanna (rooting depth of 1.5 m). As soil storage capacity is a function of rooting depth, even with more energy available for evapotranspiration, only half of the soil water can be evapotranspirated due to the limited rooting depth. Neglecting human water abstraction in variant NoUse would lead to an overestimation of AET in regions where water abstraction for irrigation leads to reduction of wetland areas (Fig. 3e), and a global AET overestimation by less than 0.2 % (Table 2).

In WaterGAP 2.2, AET can become negative in some (mostly snow dominated) regions, where precipitation input is too low to reproduce observed discharge (grey colors in Fig. 3a). The total water balance of each large water body is calculated in its outflow cell; hence, AET can become very large as the value in millimeters is calculated by dividing AET over the whole lake by grid cell area.

3.3 Renewable water resources

RWR (mean annual runoff of the grid cell to the river without consideration of human water use) are dominantly influenced by the calibration (NoCal) and subsequently by input data and model structure (Fig. 4).

As RWR are approximately the difference between precipitation and AET, the difference maps (Fig. 4b–e) represent more or less the inverted difference maps in Fig. 3. Compared to STANDARD, the largest differences occur in model variant NoCal. In contrast to AET, calibration leads in many cases to lower RWR. The global sum of RWR of NoCal is 15.8 % higher than with STANDARD. The global sum of RWR from CLIMATE is 4.7 % higher but with a large spatial spread. RWR decrease in CLIMATE on 21.4 % of land surface by more than 10 mm yr^{-1} and increase by more than 10 mm yr^{-1} on 29.9 % of the land surface. RWR decrease (increase) by more than 100 mm yr^{-1} on 4.7 % (9.0 %) of the land surface. The differences in LANDCOVER mainly follow differences in net radiation (not shown). In snow-dominated regions, RWR are lower in STRUCTURE because snow-cover dynamics are less intense than in STANDARD. In grid cells with (large) surface water bodies, RWR are lower in STRUCTURE (as AET is unlimited here even if storages are nearly empty).

3.4 River discharge

3.4.1 River discharge seasonality

River discharge is the integral result of runoff generation, water losses by evaporation from surface water bodies, positive or negative net abstractions from surface water bodies and groundwater, and routing processes. It is one of the most important diagnostic variables in water resources. In many regions, river discharges have been observed for decades, providing an important data source for model evaluation. A good representation of modeled seasonality in comparison to the observed one is therefore a criterion for model evaluation. We compared observed and modeled discharge seasonality at the outflow of 12 large river basins, covering different climatic zones and levels of anthropogenic influence (Fig. 5). Climate input and model structure influence modeled discharge seasonality more than land cover changes for the selected river basins. Where seasonality of climate is high, like in the monsoon-dominated Mekong Basin, only marginal differences occur due to land cover and model structure. Structural model refinements have also important effects on discharge seasonality. For example, the constant flow velocity of STRUCTURE (in contrast to variable flow velocity in the other variants) leads to a higher peak in the Lena. Here, the variable flow velocity algorithm underestimates flow velocity in the lower reaches where bed slopes are very small. This leads to a strong underestimation of peak flow (which explains the improved seasonality of STRUCTURE compared to observed discharge in the Lena). The reservoir algorithm which is not enabled in STRUCTURE has impacts at the Yangtze, Rio Parana, Mississippi and the Volga rivers in terms of smoothing the discharge. For the Rio Parana, this is the main influence in the STRUCTURE variant. The representation of snow in STANDARD leads

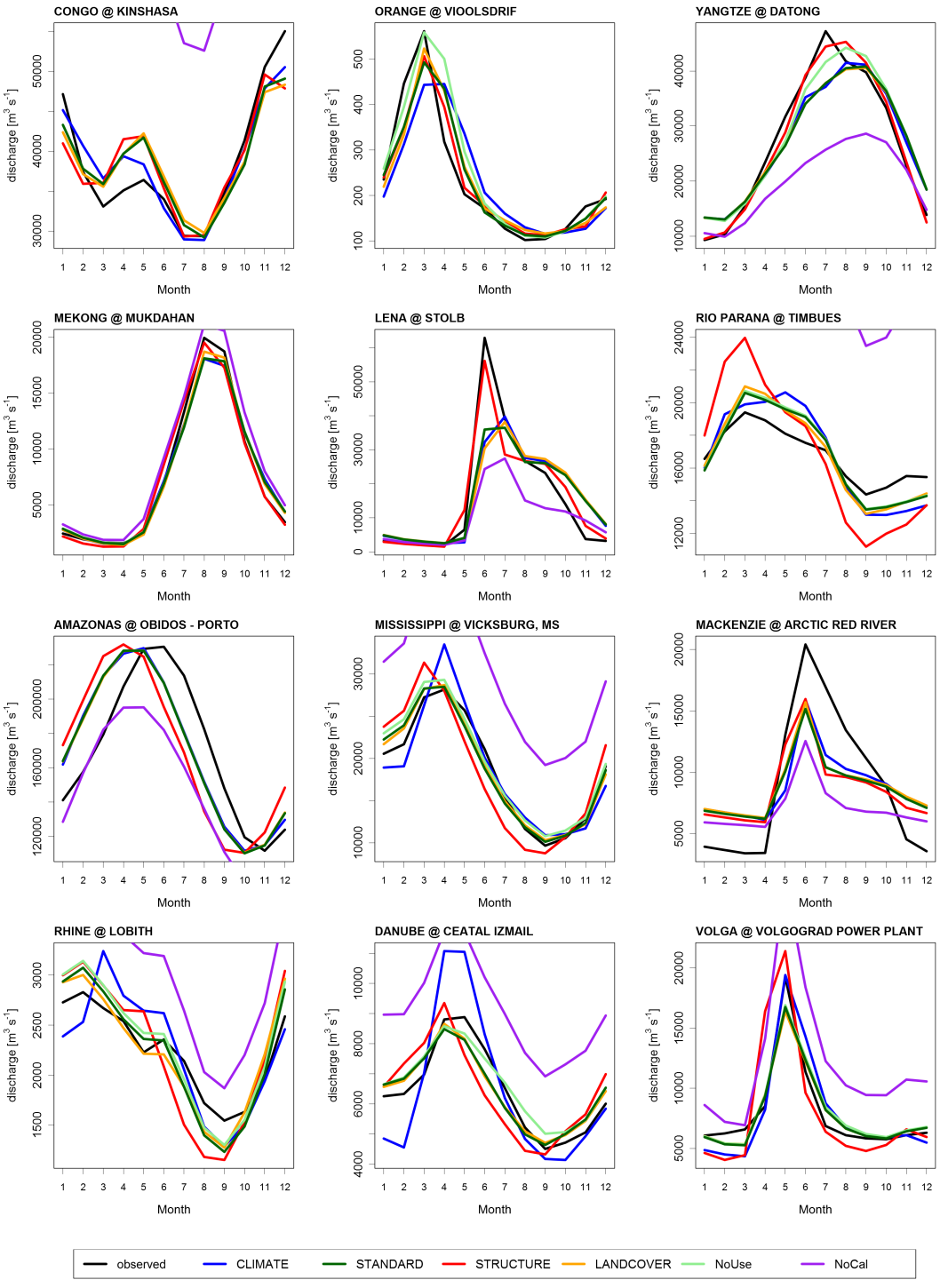


Figure 5. Discharge seasonality for selected basins and the calibrated model variants. Values for NoCal are only visible if they are in the range of calibrated model variants.

to a more heterogeneous snow coverage as compared to the STRUCTURE variant. The strongest impact occurs for the Rhine, where the snow algorithm is the dominant reason for the differences to STRUCTURE. In STRUCTURE, the snow water storage of the Rhine headwater (Alps) is generally lower. In particular between May and October (the Alps are modeled as snow-free between June and September), this leads to a decrease of discharge as snowmelt cannot contribute any longer as it does, for example, in STANDARD. The importance of the climate forcing can be seen

in the Mississippi and the Rhine where CLIMATE results in overestimated peak seasonal discharge. In the Danube, the WFD/WFDEI climate input (in STANDARD) is particularly beneficial, as the fit to observed seasonality is much better than with CRU TS 3.2/GPCC v6 climate (in CLIMATE).

For the Mackenzie, all model variants are close to each other but far away from observations. Here, freezing and thawing of the river are not reproduced as none of the model variants represents these processes. Interestingly, the Lena basin is also frozen during wintertime but here low flows are simulated quite well. In the Amazon, the model variants underestimate the delay of peak discharge which might be explained by the lack of modeling dynamic floodplain inundation.

The impact of alternative land cover is only slightly influencing discharge seasonality. Most effects occur at the Rhine, where CORINE-based land cover (variant LANDCOVER) consists dominantly of cropland. Many grid cells in the other model variants consist of mixed forest or cropland/natural vegetation mosaics which have a lower albedo, resulting in more evaporation and less discharge especially in the summer months. Additional effects occur due to deeper roots in the mixed forest class. Only for the Mackenzie, Lena and Yangtze, are mean monthly river discharges of NoCal within the range of all other variants in some months. The NoCal values for the Orange are higher than any observed value (and the values of the other variants) throughout the year (Fig. 5). This highlights the need for a calibrated model for discharge analyses.

3.4.2 Monthly time series

Nash–Sutcliffe efficiencies E_{NS} (Eq. 1; Nash and Sutcliffe, 1970) were calculated for time series of monthly river discharges at the gauging stations used for calibration.

$$E_{NS} = 1.0 - \frac{\sum_{i=1}^{n} (O_i - S_i)^2}{\sum_{i=1}^{n} \left(O_i - \overline{O}\right)^2}, \quad (1)$$

where O_i is observed discharge, S_i is simulated discharge and $\overline{O}$ is mean observed discharge (all units in $km^3\,month^{-1}$).

By adjusting the mean annual river discharge as done in our calibration approach, the E_{NS} of monthly discharge increases in all calibrated model variants as compared to the NoCal variant because E_{NS} is sensitive to both the mean and variances (Fig. 6). Among all calibrated variants, STANDARD and NoUse achieve the highest mean E_{NS} values, while variant STRUCTURE shows a distinctly lower model performance (Fig. 6). This is further confirmed by the E_{NS} distribution per Köppen–Geiger region (Table 4, column "sum"), where for the STANDARD and NoUse variants E_{NS} is larger than 0.5 in 53.5 % of the basins. Comparing

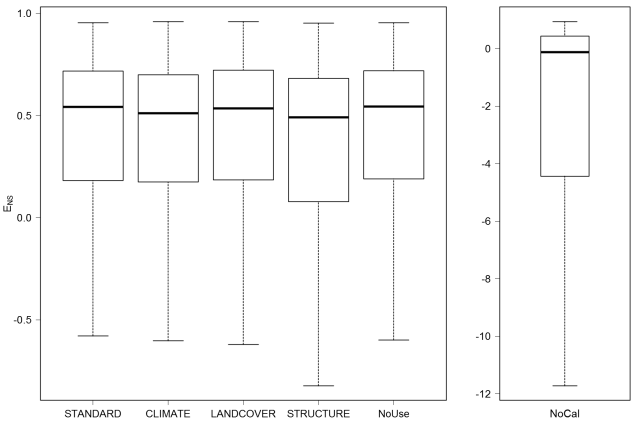


Figure 6. Nash–Sutcliffe efficiency (E_{NS}) (excluding outliers) of monthly observed and simulated discharge at 1319 stations used for calibration.

STANDARD and STRUCTURE, model development clearly improved simulation results in A, C and D climates. The CLIMATE variant performs better in cold areas but overall performs worse than STANDARD, particularly in temperate climate. No significant differences occur when using an alternative land cover input (LANDCOVER). Performance of all variants is very poor in arid (B) climate.

3.5 Variations of total water storages

Simulated temporal variations of TWS, i.e., the total amount of water in all continental water storage compartments (Fig. A1), are used widely in the context of analyzing information derived from the Gravity Recovery and Climate Experiment (GRACE). The dominant seasonal changes of TWS can be characterized by the difference between the minimum and the maximum value of mean monthly TWS (1971–2000). The spatial distribution of seasonal TWS variations (Fig. 7a) is similar to that derived with an earlier version of WaterGAP (see Fig. 4 in Güntner et al., 2007). Seasonal TWS variations are affected most strongly by the climate forcing (Fig. 7b). For example, in Europe and eastern US, they are more than 25 mm higher in case of CRU/GPCC climate forcing. This finding is consistent with the impact of climate forcing on river discharge, e.g., of the Danube (Fig. 5). The calibration approach leads to a decrease of TWS variation in areas where runoff is overestimated (Fig. 7f). Where land cover attributes vary significantly due to different land cover classes in LANDCOVER, the effects on TWS variations are strong (e.g., in the southern Congo or southern Amazon). Neglecting groundwater abstractions (as done in NoUse, which neglects any human water use, and in STRUCTURE, where water is only abstracted from surface waters) leads to lower seasonal TWS variations in areas of groundwater abstractions (if in case of STRUCTURE surface water is not able to satisfy water uses) and groundwater depletion

Table 4. Number of calibration basins per E_{NS} category and Köppen–Geiger climate zone*.

Variant	Class	E_{NS}	A	B	C	D	E	Sum
STANDARD	1	> 0.7	75	19	117	129	29	369
	2	0.5–0.7	100	17	68	134	18	337
	3	< 0.5	110	91	83	282	47	613
CLIMATE	1	> 0.7	67	8	77	145	30	327
	2	0.5–0.7	116	31	68	107	26	348
	3	< 0.5	104	79	127	293	41	644
LANDCOVER	1	> 0.7	77	20	117	128	32	374
	2	0.5–0.7	94	16	68	132	15	325
	3	< 0.5	114	91	83	285	47	620
STRUCTURE	1	> 0.7	63	20	85	99	27	294
	2	0.5–0.7	101	16	84	132	22	355
	3	< 0.5	121	91	99	314	45	670
NoUse	1	> 0.7	77	15	109	138	30	369
	2	0.5–0.7	97	26	68	130	17	338
	3	< 0.5	111	86	91	277	47	612
NoCal	1	> 0.7	17	5	39	61	12	134
	2	0.5–0.7	28	4	32	80	11	155
	3	< 0.5	240	118	197	404	71	1030

* Calculated by WaterGAP after Kottek et al. (2006); A: equatorial climate, B: arid climate, C: warm temperate climate, D: snow climate and E: polar climates. Note that the number of basins per climate zone differs for CLIMATE as here the bases for Köppen–Geiger climate calculation are CRU TS 3.2 and GPCC v6 instead of WFD/WFDEI climate input for all other variants.

(e.g., High Plains Aquifer in central US, Iran and northwestern India) (Fig. 7d, e). In these two variants, seasonal groundwater storage variations are solely driven by seasonal variations of groundwater recharge. Without simulating water use, some areas with extensive surface water irrigation have higher seasonal variations than with water use because large return flows during the dry (irrigation) season smooth natural groundwater storage variations.

In addition, seasonal TWS variations in STRUCTURE differ from STANDARD particularly along large rivers (Fig. 7d), mostly with a smaller range in STRUCTURE. There, the flow velocity (variable in STANDARD) is lower than the constant $1\,m\,s^{-1}$ in STRUCTURE, resulting in increased river storage. In many cold areas, the simpler snow algorithm in STRUCTURE leads to increased TWS seasonality.

4 Discussion

4.1 Comparison of simulated freshwater fluxes to other estimates

The modeled AET and discharge to the oceans and inland sinks for all model variants are within the range of published values except for the NoCal variant, which has very low AET and high discharge values (Tables 2, 5). Discharge estimates differ due to the applied estimation method and precipitation data set. Mueller et al. (2013) do not consider a precipitation undercatch correction and assume a global precipitation of $\sim 99\,000\,km^3\,yr^{-1}$ which is low compared to recent estimates of Schneider et al. (2014) ($117\,000\,km^3\,yr^{-1}$) or the values used in this study (Table 2). Compared to previous WaterGAP results, model refinements have led to an increase of discharge. The value of STANDARD is approximately $450\,km^3\,yr^{-1}$ higher than for STRUCTURE (Table 2), and previous estimates (Döll et al., 2003) are even lower as precipitation undercatch was not taken into account.

4.2 Advantages and limitations of the calibration approach

The applied calibration approach is clearly beneficial as it leads to a better fit of simulated to observed monthly river discharge time series (Fig. 6; Table 4). Consequently, the basin-specific adjustment of 1–3 parameters (γ, CFA and CFS; see Appendix B1) based on observed mean annual discharge has been a part of the WaterGAP modeling approach since the beginning. Calibration allows to a certain degree compensating errors in input data and effective model parameters. Also, structural problems of the model, e.g., due to the simplified representation of hydrological processes at a half-degree grid cell, may be balanced out. The effect of calibration on modeled renewable water resources

Table 5. Comparison of diverse estimates of global actual evapotranspiration and discharge (in $km^3\,yr^{-1}$).

Actual evapotranspiration		Discharge	
62 800	Mu et al. (2011)	34 406	Mueller et al. (2013)
64 512[a]	Mueller et al. (2013)	36 200	Wada et al. (2010)
65 000	Jung et al. (2010)	36 687	Döll et al. (2003)
65 500	Oki and Kanae (2006)	37 288	Dai and Trenberth (2002)
66 000	Sterling et al. (2012)	38 587	Baumgartner and Reichel (1975)
71 000	Baumgartner and Reichel (1975)	38 605	Widén-Nilsson et al. (2007)
72 000	Korzun (1978)	39 307	Fekete et al. (2002)
75 981[b]	Mueller et al. (2011)	39 414	Döll and Fiedler (2008)
60 000–85 000	Haddeland et al. (2011)	44 560	Korzun (1978)
		45 500	Oki and Kanae (2006)
		42 000–66 000	Haddeland et al. (2011)
70 576[c]	STANDARD	40 458	STANDARD

[a] $1.35\,mm\,d^{-1}$ based on a land area of $130\,922 \times 10^6\,km^2$. [b] $1.59\,mm\,d^{-1}$ based on a land area of $130\,922 \times 10^6\,km^2$ (value taken from Mueller et al., 2013 as no area is given in Mueller et al., 2011). [c] Sum of AET and WC_a.

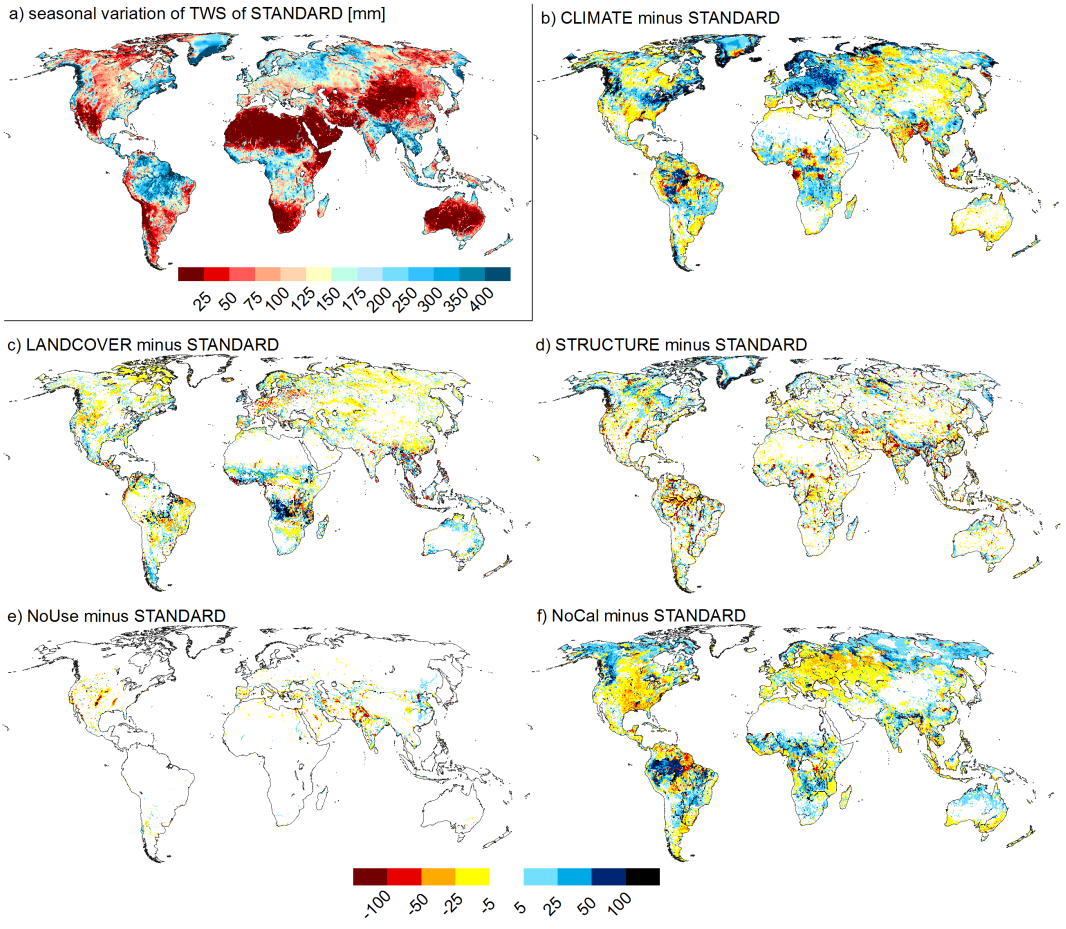


Figure 7. Seasonal variation of total water storages (TWS) for STANDARD **(a)** and as difference maps (mm) to all other model variants **(b–f)**.

(Fig. 4e) dominates all other modifications within this experiment setup.

However, the correction of total cell runoff using CFA and CFS that is required to force simulated mean annual river discharge values to be equal to observed values is not ideal and has undesirable effects on estimated AET and RWR. In the Yenisey Basin, upstream of Igarka (western Siberian Plain), AET is largely reduced in one-half of the basin (and vice versa) when using alternative climate forcing. Transferring the correction factor CFS (which is, if necessary, calculated

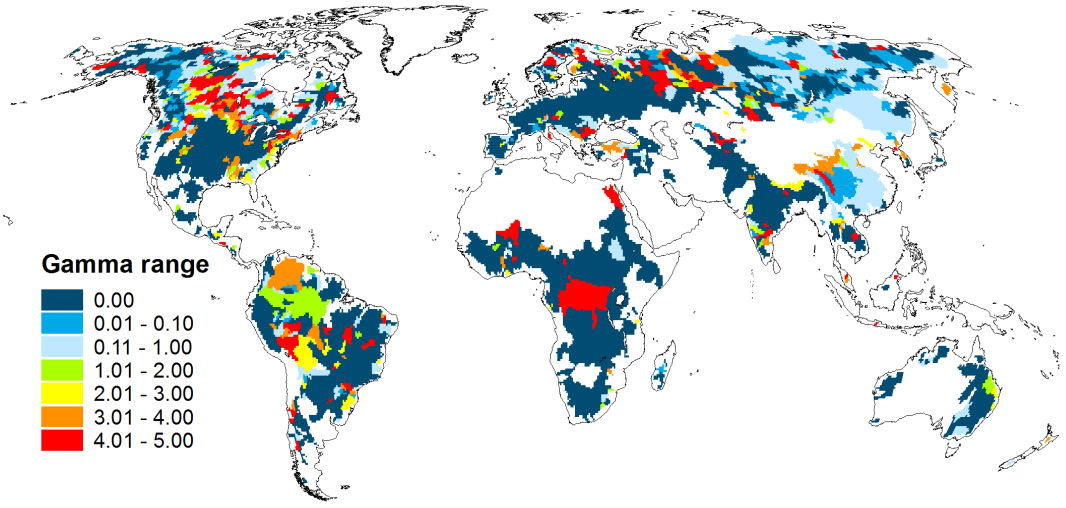


Figure 8. Range of calibration parameter γ through all four calibrated model variants (calculated as $\gamma_{max} - \gamma_{min}$) showing the general sensitivity to input data and model structure. White colors indicate uncalibrated regions.

at the outflow grid cell of the basin) to the upstream grid cells can lead to unrealistic high positive and negative values for AET if precipitation is too low in these parts of the basin to simulate observed discharge or if the AET of surface water bodies has to be reduced by CFA. This is the reason for some artificial patterns in Fig. 3 and consequently in Fig. 4. These kinds of consistency errors can be found in some more basins where cumulative AET is low and parts of the basins are covered with surface water bodies. Nevertheless, the approach ensures a closed water balance for the whole basin.

Obviously, one parameter is not sufficient to calibrate the model. In many basins the γ parameter is not sensitive to input data and model structure in the current calibration approach as the range of γ through all four variants (NoCal is not considered, NoUse has the same value as STANDARD) is rather small. Of the basins in Fig. 8, 59 % are colored dark blue which means that the calibration parameter γ has the same value in all model variants. Here, γ is at its artificial minimum (0.1) or maximum (5.0) value, and the influence of input data and model structure, which were modified in this experiment, is insignificant. However, in 21 % of the basins, γ is differing by > 1 (green, yellow and red colors). In these basins, the calibration parameter is sensitive to input data and model structure. Therefore, within future model development, one task is to restructure the calibration approach with the aim to avoid correction factors or rather to introduce and test alternative calibration objectives. This could be achieved by either including more parameters (multiparameter calibration) and/or by integrating additional reference data, e.g., GRACE-based data as was shown by Werth and Güntner (2010) (multiobjective calibration). In addition, remote-sensing-based input data with global coverage have been available for a decade. Especially for land cover characteristics (e.g., land cover type, L, and albedo; see Appendix A), a more realistic representation of dynamics (integration of time series as input data instead of static input maps) can reduce the input data and model parameter uncertainty.

4.3 How sensitive are freshwater fluxes and water storages to spatially distributed input data (climate forcing, land cover)?

In general, more differences occur due to the alternative climate input than due to the alternative land cover data. The major freshwater fluxes (AET, Fig. 3, and RWR, Fig. 4) as well as river discharge (Fig. 5) show in many cases that land cover input has much less impact (except for some areas where the attributes of a changed land cover type differ significantly). The effect of different land cover input would probably increase if the associated attributes were also modified. Forced with CRU 3.2 and GPCC v6 instead of WFD/WFDEI input, AET increased by at least 10 mm yr^{-1} in large parts in the world (light blue colors in Fig. 3b). In those regions with similar precipitation amounts but different radiation, RWR decreased by the same amount that AET increased (e.g., Southeast Asia, Australia, Saudi Arabia). In other regions, no clear effect on RWR is detectable (e.g., North America). In some parts of Europe, RWR increased by at least 10 mm yr^{-1} even if AET increased. Here, besides radiation (affecting AET), the amount of precipitation is of great importance (affecting RWR).

In regions where the climate forcing data sets differ significantly (e.g., Danube River basin), the impact on discharge is large (Fig. 5, bottom center panel). Here, differences in temperature and precipitation amounts lead to a poor fit compared to observed discharge when using the CLIMATE variant which is also reflected in the E_{NS} criterion (Fig. 9b).

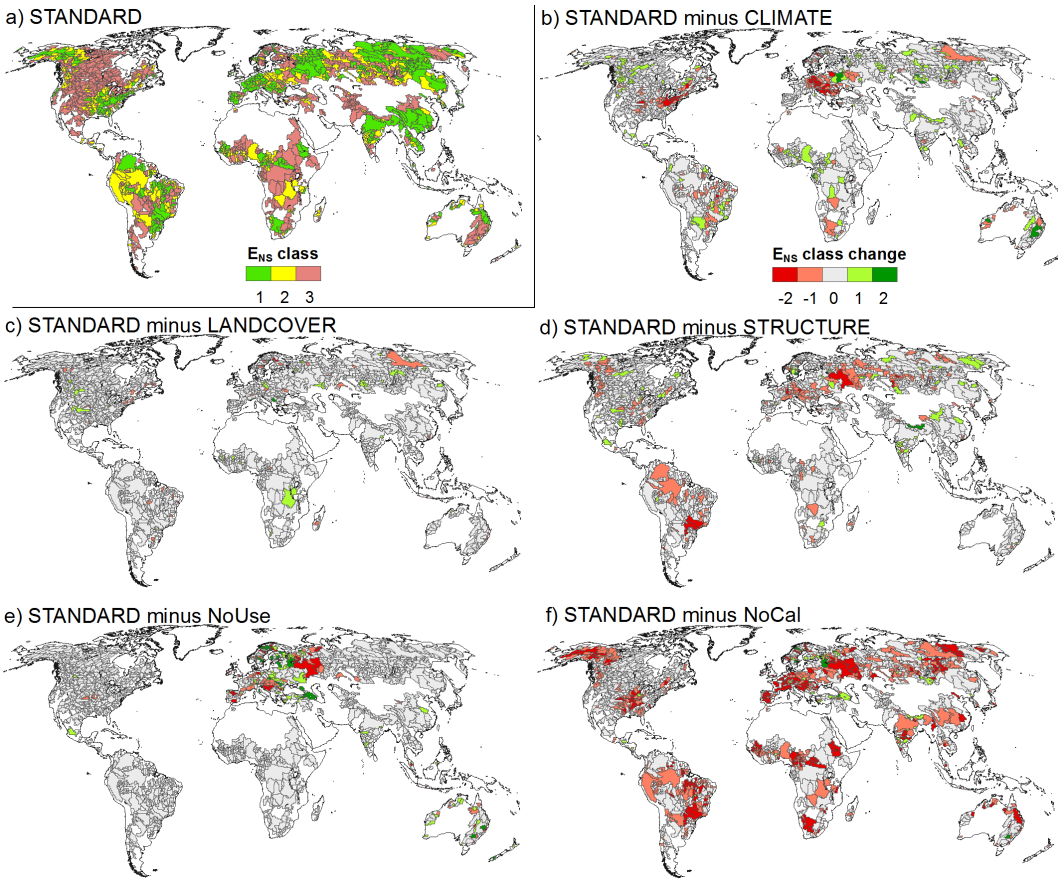


Figure 9. Spatial distribution of Nash–Sutcliffe Efficiency (E_{NS}) classes (from Table 4, 1: $E_{NS} > 0.7$, 2: $0.5 < E_{NS} < 0.7$, 3: $E_{NS} < 0.5$) for STANDARD **(a)**, and differences of model variants (calculated as STANDARD E_{NS} class minus that of the model variant) **(b–f)**. Red colors indicate a decrease, green an increasing E_{NS} when using the model variant compared to STANDARD.

Also, the two land cover input data sets used here result in the same E_{NS} classes, with only a few exceptions (Fig. 9c).

4.4 What are the benefits of WaterGAP model structure refinements implemented during the last decade?

In general, WaterGAP 2.2 STANDARD leads to improved results compared to the reduced model version STRUCTURE that is comparable to the Döll et al. (2003) model version. In many basins in the Alpine region in central Europe, the E_{NS} of STRUCTURE ranks behind that of STANDARD (Fig. 9d, red colors) reflecting a refined simulation of snow-cover dynamics on the 3 arcmin subgrid. In some basins, the reservoir algorithm improves E_{NS} (and discharge seasonality). For example, the Volga at station Volgograd Power Plant (see also Fig. 5) and basins in Brazil show a much better E_{NS} (Fig. 9d) in STANDARD compared to STRUCTURE. However, the E_{NS} of some basins with $E_{NS} < 0.5$ in STANDARD is improved in STRUCTURE. In summary, integrating more complex and refined process descriptions (see Sect. 2.2.3) in the past decade has led to the improved simulation of monthly time series of river discharge with WaterGAP. However, discharge before calibration tends to be higher with the implemented structural changes, e.g., due to the storage-dependent reduction of surface water evaporation. This together with the use of more calibration stations (Hunger and Döll, 2008) and the introduction of a bias correction for observed precipitation (Döll and Fiedler, 2008) has had the problematic consequence that correction factors to lower simulated river discharge have increasingly been required to ensure that simulated mean annual river discharges are equal to observed values.

4.5 How does the modeling approach (calibration procedure, consideration of human water use) affect freshwater fluxes and water storages?

The calibration procedure reduces simulated river discharge and water resources on most of the land area and increases the AET (Figs. 3, 4; Table 2). Without calibration, global AET and discharge would rank at the lower and higher end of the published values, respectively (Table 5). In addition, the fit to observed monthly river discharge time series as quantified using the E_{NS} criterion would worsen almost

Table 6. The three model variants with the largest differences to the STANDARD variant (dSTA) regarding global freshwater fluxes (Q and AET) and total water storage trends (dTWS/dt) (from Table 2; values in km^3 yr^{-1}) as well as median E_{NS} for monthly time series of river discharge at the 1319 calibration basins.

Variable	STANDARD	Rank 1	dSTA	Rank 2	dSTA	Rank 3	dSTA
Q	40 458	NoCal	6364	CLIMATE	1906	NoUse	758
AET	69 803	NoCal	−6459	STRUCTURE	414	LANDCOVER	209
dTWS/dt	−214	STRUCTURE	169	NoUse	140	NoCal	71
Median E_{NS}	0.54	NoCal	−0.66	STRUCTURE	−0.05	CLIMATE	−0.03

everywhere (Fig. 9f). The impact of calibration on freshwater fluxes and water storages is higher than those of alternative climate forcings and land cover data, and of a more sophisticated model structure. This confirms the strong benefit of calibration. However, as E_{NS} is affected by mean discharge as well as discharge variations, the calibration approach improves this criterion.

Compared to the other variants, the consideration of human water use does not have large effects on freshwater fluxes and storages at the global scale. In regions with intense water use, in particular from surface water bodies (e.g., in Pakistan), AET without considering additional evaporation from WC_a (Table 2) is reduced due to human water use (Fig. 3e). This effect occurs because human water uses decrease surface water storages and thus the reduction factor decreases evaporation from surface water bodies (Appendix Sect. A5). If the impact of human water use on river discharge were not considered, van Beek et al. (2011) showed there would be a lower performance in general. Within our experiment, higher correction factors would be necessary in basins with large abstractions from surface water bodies or significant decreases of baseflow due to groundwater abstractions. Still, the E_{NS} of basins with high amounts of human water use is generally lower than those without human water use (not shown). In some basins, mainly in northeastern Europe, E_{NS} improves when neglecting human water use (Fig. 9e). This obviously reflects uncertainties in water use models.

4.6 Which type of uncertainty is dominant for specific fluxes and variations of total water storage?

The answer to this question depends on the type of fluxes and the spatial aggregation. Regarding selected global sums of freshwater fluxes (Q and AET) and mean annual total water storage trends dTWS, dominant uncertainties can be determined by computing differences between the values computed with a certain model variant and STANDARD. As already shown above, global values of AET and Q as well as the fit of simulated to observed river discharge time series (E_{NS}) are most sensitive to whether the model is calibrated or not (Table 6). STRUCTURE and NoUse have the strongest impact on the global TWS trend (Table 6) as these model variants cannot reflect groundwater depletion. More refined model algorithms rank second regarding global AET sums and E_{NS}, and alternative climate forcings rank second regarding river discharge and third regarding median E_{NS}. The alternative land cover input data sets have the overall lowest impact on computed freshwater fluxes and storages.

Table 7. Rank of model variants where global land area (except Greenland and Antarctica) is affected most based on a threshold which represents the 10th percentile of averaged (1971–2000) global grid cell values for AET and discharge.

Rank	Variant	Percentage of area affected by changes above 10th percentile	
		AET	Discharge
1	NoCal	60.5	13.5
2	CLIMATE	45.5	3.2
3	LANDCOVER	24.2	1.2
4	STRUCTURE	13.6	1.1
5	NoUse	0.9	0.03

Regarding grid-cell-specific differences that are more relevant than global values for most applications, the ranking of dominant uncertainties is quite different. Patterns of seasonal TWS variations are affected most strongly by the climate forcing (Fig. 7b), while climate forcings show the second largest impact on the spatial distribution of AET and RWR, after calibration (Figs. 3, 4). The fraction of the global land area that is affected by significant differences of AET and river discharge between a certain model variant and the STANDARD variant is largest in the case of NoCal, followed by CLIMATE, LANDCOVER, STRUCTURE and NoUse. Thus, both global and grid cell values are most sensitive to calibration. The larger sensitivities to climate forcings and land cover input at the grid cell level (Table 7) cancel when globally averaged. The larger sensitivities of globally aggregated values (Table 6) to structural changes and the consideration of water use is due to unidirectional changes for all affected grid cells, but different to alternative climate and land cover data, structural changes and water use only affect a limited number of grid cells. This discussion on the dominant type of uncertainty does not take into account parameter

uncertainty which is a major additional source of uncertainty (Kaspar, 2003).

5 Conclusions

We studied the sensitivity of freshwater fluxes and storages as computed by the GHM WaterGAP 2.2 to spatially distributed input data (climate forcing and land cover input) as well as model structure (model refinements during the last decade), consideration of human water use and calibration (or no calibration). We designed five model variants in addition to the standard variant. In each model variant, one component or feature was modified with respect to the standard variant. Sensitivity of different freshwater fluxes and water storage variations to the five types of uncertainty were analyzed and ranked considering both global sums and grid cell values, taking into account also the capability of the model variants to simulate time series of observed river discharge. Basin-specific calibration to mean annual river discharge was found to have the strongest impact on fluxes and storage variation and is the dominant reason for an improved simulation of observed monthly river discharge time series (as characterized by the Nash–Sutcliffe criterion). Uncertainty due to alternative climate forcing, and to a lesser extent, land cover input, leads to significant variations of grid cell fluxes (actual evapotranspiration, renewable water resources and river discharge) and storages (seasonal range of total water storage) even if the model variants are individually calibrated. However, these uncertainties largely cancel each other out at the global scale while the more refined model structure, and to a lesser extent water use, are more important for global sums of river discharge and actual evapotranspiration but also for an improved fit to observed monthly time series of river discharge.

The STANDARD variant of WaterGAP 2.2 leads to the best fit to observed river discharge (monthly time series, Fig. 6 and Table 4, and seasonality, Fig. 5). We conclude that the daily WFD/WFDEI data set as climate forcing is preferable to using a combination of the monthly CRU 3.2 and GPCC v6 data sets as done for model variant CLIMATE. However, we found that it is problematic to combine the WFD climate data set (covering 1901–2001) with the only seemingly consistent WFDEI data set (covering 1979–2009) due to a radiation bias (shortwave downward radiation component) between the two data sets. This results in a steep increase of actual evapotranspiration in 1979, and a water storage decrease between 1971 and 2000 that is an artifact of the combination of the two climate data sets (comp. Sect. 3.1). It would be very beneficial for an improved estimation of global freshwater fluxes and storages to have a consistent daily climate forcing that covers the whole 20th and the early 21st century.

The calibration approach of WaterGAP is necessary to compensate uncertainties of spatially distributed input data, parameters and model structure. However, a calibration of only one parameter related to soil water balance is not sufficient and correction factors have to be applied in a number of basins. Therefore, a redesign of the calibration approach, with additional observations (e.g., including TWS variations as derived from GRACE gravity fields), other calibration objectives and adjustment of more model parameters (without correction factors) is planned.

The improved representation of hydrological processes of WaterGAP within the last decade led to a more complex model structure. In most cases, those modifications resulted in a better fit to observed river discharge. However, in some parts of the world, model performance is still not satisfactory due to an inappropriate modeling of certain processes such that further changes of the model structure are required. For example, the modeled discharge seasonality in the Amazon basin is shifted compared to that observed, which is suspected to be caused by inappropriate modeling of the temporal variations of inundations and the neglect of backwater effects. The reservoir operation algorithm does not yet take into account the construction year of the dam. Moreover, model results in semiarid and arid regions are poor, and improved modeling of evaporation from ephemeral ponds is planned.

Overview of Appendix

Appendix A describes the WGHM in its current version 2.2. In the order of processing, the single-storage compartments and the belonging in- and outflows are explained. Appendix B provides information on the calibration and regionalization approach WaterGAP is based on. Appendix C gives a brief introduction of the water use submodels, and the GWSWUSE module is described in Appendix D.

Appendix A: Description of the WaterGAP Global Hydrology Model (WGHM)

A1 Canopy

The change of canopy storage S_c (mm) over time t (d) is calculated as

$$\frac{dS_c}{dt} = P - P_t - E_c, \tag{A1}$$

where precipitation P (mm d^{-1}) is the inflow and the amount of throughfall P_t (mm d^{-1}) and canopy evaporation E_c (mm d^{-1}) are the outflows.

Throughfall P_t is calculated as

$$P_t = \begin{cases} P & S_c \geq S_{c,max} \\ 0 & S_c < S_{c,max} \end{cases}. \tag{A2}$$

Following Deardorff (1978), canopy evaporation E_c is calculated as

$$E_c = E_p \left(\frac{S_c}{S_{c,max}} \right)^{\frac{2}{3}}, \tag{A3}$$

where E_p is potential evapotranspiration (mm d^{-1}).

E_p is calculated according to the Priestley–Taylor model (Priestley and Taylor, 1972), differentiating atmospheric water demand between humid ($\alpha = 1.26$) and semiarid/arid ($\alpha = 1.74$) areas. Grid cells were defined as semiarid/arid if the long-term-average (1971–2000) precipitation is less than $0.5 \times E_p$ (UNEP, 1992).

S_c is limited between 0 and maximum canopy storage $S_{c,max}$, which is calculated as

$$S_{c,max} = m_c L, \tag{A4}$$

where m_c is 0.3 (mm) and L is the leaf area index (−). L is calculated based on a modified growth model described in Kaspar (2003) and is limited to minimum and maximum values. Maximum L values per land cover class (Table A1) are based on literature (Schulze et al., 1994; Scurlock et al., 2001). Minimum L values per land cover class are calculated as

$$L_{min} = 0.1 f_{d,lc} + \left(1 - f_{d,lc}\right) c_{e,lc} L_{max}, \tag{A5}$$

where $f_{d,lc}$ is the fraction of deciduous plants (−) and $c_{e,lc}$ is the reduction factor for evergreen plants (−) (Table A1). Development of L is simulated as a function of daily temperature and precipitation. The growing season starts when the daily temperature is above 8 °C for a land cover specific number of days (Table A1) and cumulative precipitation is at least 40 mm. During the growing season, L increases linearly until it reaches L_{max} after 30 days. In semiarid and arid regions, it is necessary that at least 0.5 mm daily precipitation occurs to keep the growing season going. If the condition for the growing season is not fulfilled anymore, the senescence phase is initiated; i.e., L is degraded to L_{min} linear within 30 days.

A2 Snow

The change of snow water storage S_{sn} (mm) over time t (d) is calculated as

$$\frac{dS_{sn}}{dt} = P_{sn} - M - E_{sn}, \tag{A6}$$

where P_{sn} is precipitation, falling as snow at temperatures below 0 °C (mm d^{-1}), M is snowmelt (mm d^{-1}) and E_{sn} is sublimation (mm d^{-1}).

Snow accumulation and melt are modeled on a 3 arcmin subgrid (100 subgrid cells per 0.5°) using a degree-day algorithm (Schulze and Döll, 2004). Mean subgrid elevation was derived from GTOPO30 (US Geological Survey, 2003). The daily temperature for each subgrid cell is calculated from the temperature of the 0.5° cell, applying an adiabatic lapse rate of 0.6 °C per 100 m. To avoid excessive snow accumulation, temperature does not decrease if a snow water equivalent of 1000 mm is reached in one subgrid.

At temperatures below 0 °C, all precipitation is assumed to fall and accumulate as snow. At subgrid temperatures T (°C) above melting temperature T_m (0 °C) and if snow storage is present, snow melts with land cover specific degree-day factor D_F (mm d^{-1} °C^{-1}) (Table A2) as

$$M = \begin{cases} D_F (T - T_m) & T > T_m,\ S_{sn} > 0 \\ 0 & \text{other} \end{cases}. \tag{A7}$$

Instead of using one specific albedo for snow as in previous versions ($\alpha = 0.4$), land cover specific snow albedo values are used to account for differences in reflective properties between the land use classes under snow-covered conditions (Table A2). The albedo value switches to snow albedo if the snow water equivalent of the grid cell exceeds 3 mm, i.e., a closed snow cover is assumed. Sublimation E_{sn} is modeled like potential evaporation rate but applying a latent heat of 2.835 (MJ kg^{-1}) for temperatures below 0 °C and 2.501–$0.002361 \times T$ (MJ kg^{-1}) above 0 °C.

A3 Soil

Like snow and canopy, the change of soil water storage S_s (mm) over time t (d) is calculated as one layer as

$$\frac{dS_s}{dt} = P_{eff} - R_l - E_s, \quad \text{(A8)}$$

with effective precipitation P_{eff} (mm d^{-1}) as inflow and runoff from land R_l (mm d^{-1}) and actual evapotranspiration E_s (mm d^{-1}) as outflows.

$$P_{eff} = P_t - P_{sn} + M, \quad \text{(A9)}$$

with P_t is throughfall (mm d^{-1}), (see Sect. A1), P_{sn} is precipitation falling as snow (mm d^{-1}) and M is snowmelt (mm d^{-1}).

Actual evapotranspiration from the soil E_s (mm d^{-1}) is a function of potential evapotranspiration from the soil E_p (mm d^{-1}) minus the already evaporated water from the canopy E_c (mm d^{-1}), actual soil water content in the effective root zone S_s (mm) and total available soil water capacity $S_{s,max}$ (mm) as

$$E_s = \min\left(\left(E_p - E_c\right), \left(E_{p,max} - E_c\right) \frac{S_s}{S_{s,max}} \right), \quad \text{(A10)}$$

where $E_{p,max}$ is 20 mm d^{-1} in semiarid and arid regions whereas 10 mm d^{-1} in grid cells classified as humid, $S_{s,max}$ is the product of total available water capacity in the upper meter of the soil (Batjes, 1996) and the land cover specific rooting depth (Table A2).

Runoff from land R_l (mm d^{-1}) is calculated after Bergström (1995) as

$$R_l = P_{eff} \left(\frac{S_s}{S_{s,max}} \right)^{\gamma}. \quad \text{(A11)}$$

Dependent on the soil water storage S_s, a part of effective precipitation P_{eff} becomes runoff. If the soil water storage is empty, $R_l = 0$. If the soil is completely saturated (at $S_{s,max}$), runoff equals effective precipitation. Between these points, the runoff coefficient γ determines the amount of precipitation that converts to runoff. This parameter is used for calibration (see Sect. B1). In urban areas (defined as a separate input map from IMAGE 2.2), 50 % of P_{eff} is directly passed to the river.

A4 Groundwater

Inflow to groundwater storage S_g (mm) is groundwater recharge R_g (mm d^{-1}), whereas outflows are baseflow Q_g (mm d^{-1}) and net abstractions from groundwater NA_g (mm d^{-1}) (Appendix C), which can also act as inflow (e.g., as additional groundwater recharge due to irrigation with surface water).

$$\frac{dS_g}{dt} = R_g - Q_g - NA_g \quad \text{(A12)}$$

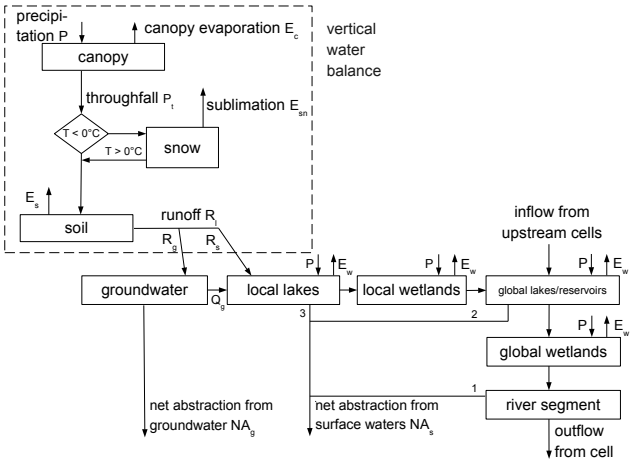


Figure A1. Schematic structure of the water fluxes and storages as computed by WGHM within each 0.5° grid cell. Boxes represent water storage compartments, arrows water fluxes (inflows, outflows). Numbers at net abstraction from surface waters (NA_s) are the order in which storage water is abstracted until demand is satisfied.

Groundwater recharge R_g (mm d^{-1}) is calculated as a fraction of runoff from land:

$$R_g = \min\left(R_{g,max},\ f_g R_l\right),$$

where $R_{g,max}$ is soil texture specific maximum groundwater recharge (mm d^{-1}) (with values of 7/4.5/2.5 for sandy/loamy/clayey soils) and f_g is the groundwater recharge factor (ranging between 0 and 1) related to relief, soil texture, aquifer type and the existence of permafrost or glaciers. For a detailed description see Döll and Fiedler (2008). If a grid cell is defined as arid and has coarse (sandy) soil, groundwater recharge will only occur if precipitation exceeds a critical value of 12.5 mm d^{-1}. Both values, $R_{g,max}$ and the precipitation threshold, are adapted to the climate forcing used (WFD) aiming to reach comparable groundwater recharge patterns to Döll and Fiedler (2008) as that groundwater recharge estimation is confirmed by experts within the WHYMAP (World-wide Hydrogeological Mapping and Assessment Programme; http://www.whymap.org) efforts. Within CLIMATE, the original values 5/3/1.5 for $R_{g,max}$ and 10 mm d^{-1} as precipitation threshold were used.

The outflow is modeled with $k_g = 0.01$ d^{-1} as

$$Q_g = k_g S_g. \quad \text{(A13)}$$

The runoff from land R_l, which is not groundwater recharge R_g, represents the fast surface runoff R_s and is routed, together with Q_g, through a series of different storages representing wetlands, lakes and reservoirs until reaching the river segment (Fig. A1).

A5 Surface water bodies

Surface water bodies (inland freshwater such as wetlands, lakes and reservoirs) play an important role in the hydrologic cycle, e.g., for evaporation and the lateral transport. In general, surface water body storages S (m^3) increase with inflow I ($m^3\,d^{-1}$) from other storages or from upstream (see Fig. A1), and are reduced by the outflow Q ($m^3\,d^{-1}$). Additionally, the water balance of the water body itself B ($m^3\,d^{-1}$) is calculated as $B = P - E_w$, where P is precipitation ($m^3\,d^{-1}$) and E_w is potential evaporation of open water surfaces ($m^3\,d^{-1}$) applying an albedo of 0.08. Finally, net abstractions of surface water NA_s ($m^3\,d^{-1}$) are considered, resulting in the storage equation:

$$\frac{dS}{dt} = I - Q + B - NA_s. \quad \text{(A14)}$$

Outflow is in principle modeled like groundwater outflow (Appendix A4) for "local" lakes and wetlands, whereas "global" lakes and wetlands are linear storages whose equations are solved analytically.

WaterGAP 2.2 does not consider variable land/water fractions as would be expected when a lake is shrinking due to evaporation and land surface increases; thus Hunger and Döll (2008) introduced a reduction parameter which reduces the evaporation when lake/wetland storage is low. In WaterGAP 2.2, for all surface water bodies the reduction factor r (−) is calculated as

$$r = 1 - \left(\frac{|S - S_{max}|}{S_{max}}\right)^p, \quad \text{(A15)}$$

where S is actual water body storage (m^3), S_{max} is maximum water body storage (m^3) and p is the reduction exponent (−). As no truly global data set on lake volumes is available, the maximum storage capacity is determined by multiplying the surface area with an "active" depth (set to 5 and 2 m for lakes and wetlands, respectively). Values for p are 3.32 for lakes and wetlands which means a reduction of evaporation by 10 % if storage is halved and 2.81 for reservoirs, which means a reduction of 15 % if storage is half of the maximum storage capacity (and a reduction of 50 % if storage is reduced to 20 % of storage capacity).

The distribution of wetlands is derived from GLWD (Lehner and Döll, 2004) as percentage of cell coverage. Locations and attributes of lakes and reservoirs are based on a combination of GLWD and a preliminary version of the GRanD (Global Reservoir and Dam) database (Döll et al., 2009; Lehner et al., 2011). In total, 6553 reservoirs, 52 regulated lakes (lakes whose outflow is regulated by a dam) (from GRanD) and 242 798 unregulated lakes (from GLWD) were considered. Out of these, 1386 large lakes (area $\geq 100\,km^2$), 1110 large reservoirs (storage capacity $\geq 0.5\,km^2$) and 52 regulated lakes (area $\geq 100\,km^2$ or storage capacity $\geq 0.5\,km^2$) were classified as "global"; i.e., they receive inflow not only from the grid cell itself but also from upstream ("global" wetlands are defined in the same way; see Fig. A1). All other surface water bodies were classified as "local". If "global" lakes or reservoirs cover more than one grid cell, the water balance of the whole surface water body is calculated at the outflow cell.

A6 Lateral routing

The global drainage direction map DDM30 (Döll and Lehner, 2002) is used to route the discharge through the stream network until it reaches the ocean or an inland sink. Fast runoff $R_s = R_l - R_g$ is routed to the surface storages without any delay, whereas baseflow Q_g is a function of groundwater storage (Fig. A1, Appendix A4). Due to limited information on groundwater flow between grid cells, the groundwater recharge can only contribute to groundwater runoff of the same grid cell.

Verzano et al. (2012) improved the routing by introducing a variable flow velocity approach based on the Manning–Strickler equation. The roughness coefficient is calculated after Cowan (1956) by using different physiographic parameters and information about rural and urban areas. The hydraulic radius is calculated using actual discharge of the cell and empirical relationships of river width and depth at bankfull flow conditions. Bankfull conditions are assumed to correspond to the 1.5-year maximum series annual flow (Schneider et al., 2011) and were accordingly calculated from daily discharge time series for the global land surface. Riverbed slopes were calculated based on the HydroSHEDS (Hydrological data and maps based on SHuttle Elevation Derivatives at multiple Scales) drainage direction map (Lehner et al., 2008) and a meandering ratio (method is described in Verzano et al., 2012).

The reservoir algorithm of Hanasaki et al. (2006), distinguishing irrigation and nonirrigation reservoirs and considering 1109 reservoirs was implemented and improved by Döll et al. (2009) and slightly adapted in WaterGAP 2.2: if reservoir storage falls below 10 % of storage capacity, the release coefficient is set to 0.1 instead of 0.0 as in Döll et al. (2009), assuring that at least some water is released, e.g., for downstream ecosystem demands.

Table A1. Parameters of the leaf area index model.

No.	Land cover type	L_{max} (−)	Fraction of deciduous plants $f_{d,lc}$	L reduction factor for evergreen plants $c_{e,lc}$	Initial days to start/end with growing season (d)
1	Evergreen needleleaf forest	4.02[a]	0	1	1
2	Evergreen broadleaf forest	4.78[b]	0	0.8	1
3	Deciduous needleleaf forest	4.63	1	0.8	10
4	Deciduous broadleaf forest	4.49[c]	1	0.8	10
5	Mixed forest	4.34[d]	0.25	0.8	10
6	Closed shrubland	2.08	0.5	0.8	10
7	Open shrubland	1.88	0.5	0.8	10
8	Woody savanna	2.08	0.5	0.3	10
9	Savanna	1.71	0.5	0.5	10
10	Grassland	1.71	0	0.5	10
11	Permanent wetland	6.34	0	0	10
12	Cropland	3.62	0	0.1	10
13	Cropland/natural vegetation mosaic	3.62	0.5	0.5	10
14	Snow and ice	0	0	0	0
15	Bare ground	1.31	0	1	10

[a] L_{max} is assumed to be the mean value of land cover classes of Scurlock et al. (2001), TeENL and BoENL; [b] only value for TrEBL and not TeEBL (Scurlock et al., 2001) as in WaterGAP this class is mainly in the tropics; [c] mean value from TeDBL and TrDBL (Scurlock et al., 2001); [d] mean value of all forest classes. Fraction of deciduous plants and L reduction factor for evergreen plants based on IMAGE (Alcamo et al., 1998), initial days to start/end with growing season are estimated.

Table A2. Attributes for IGBP land cover classes used in WaterGAP 2.2 for all model variants, compiled from various literature sources. Water has an albedo of 0.08, snow 0.6.

No.	Land cover type	Rooting depth[a] (m)	Albedo[a] (−)	Snow albedo (−)	Emissivity[b] (−)	Degree-day factor D_F [c] (mm d^{-1} °C^{-1})
1	Evergreen needleleaf forest	2	0.11	0.278	0.9956	1.5
2	Evergreen broadleaf forest	4	0.07	0.3	0.9956	3
3	Deciduous needleleaf forest	2	0.13	0.406	0.99	1.5
4	Deciduous broadleaf forest	2	0.13	0.558	0.99	3
5	Mixed forest	2	0.12	0.406	0.9928	2
6	Closed shrubland	1	0.13	0.7	0.9837	3
7	Open shrubland	0.5	0.2	0.7	0.9541	4
8	Woody savanna	1.5	0.2	0.558	0.9932	4
9	Savanna	1.5	0.3	0.7	0.9932	4
10	Grassland	1	0.25	0.7	0.9932	5
11	Permanent wetland	1	0.15	0.2	0.992	4
12	Cropland	1	0.23	0.376	0.9813	4
13	Cropland/natural vegetation mosaic	1	0.18	0.3	0.983	4
14	Snow and ice	1	0.6	0.7	0.9999	6
15	Bare ground	0.1	0.35	0.7	0.9412	6

[a] Adapted from the IMAGE model (Alcamo et al., 1998); [b] Wilber et al. (1999); [c] Maniak (1997), WMO (1994).

Appendix B: Calibration and regionalization

B1 Calibration approach

WGHM is calibrated against mean annual discharge by adjusting the runoff coefficient γ (Eq. A11) for all grid cells of each calibration basin and – if necessary – two additional correction factors. The calibration procedure of WGHM is described in Döll et al. (2003) and Hunger and Döll (2008). As WaterGAP was developed to quantify water resources and water stress, calibration forces simulated discharge to be, during the calibration period, between 99 and 101 % of observed river discharge. It is implicitly assumed that the model should be robust enough to reproduce intra- and interannual variability. The main reasons for calibration are the uncertainty of input data, parameters and model structure as well as the scale of the model and grid cell heterogeneity. To overcome overparameterization and to keep the calibration as simple as possible, calibration is performed by adjusting the one free parameter γ (Eq. A11) within the limits of 0.1 and 5.0. With low γ, runoff is high even if the soil is at low saturation, and with a high value, runoff is small even with nearly saturated soils. However, in many basins, adjustment of the soil water balance alone does not lead to a fit of simulated discharge to observed discharge for various reasons. These include uncertainty of climate forcing, underestimation of evaporation losses in dry areas caused by neglecting the formation of ephemeral ponds and neglecting of streambed losses. In these cases, the area correction factor (CFA) is computed, which adjusts net cell runoff of each cell in the subbasins. With limits between 0.5 and 1.5, cells with positive (precipitation > evapotranspiration) and negative (water body evapotranspiration > precipitation, e.g., global lakes which are fed by upstream inflow) are multiplied by a value symmetric around 1.0 (Hunger and Döll, 2008). In some basins, however, the adaptation of both γ and CFA is not sufficient for a successful calibration; i.e., the deviation between simulated and observed long-term-average discharge remains larger than 1 %. Possible reasons are discussed in Hunger and Döll (2008). To avoid error propagation to the next downstream basin, the modeled discharge is corrected to the measured discharge in the grid cell where the discharge station is located by multiplying it by the station correction factor CFS (Hunger and Döll, 2008).

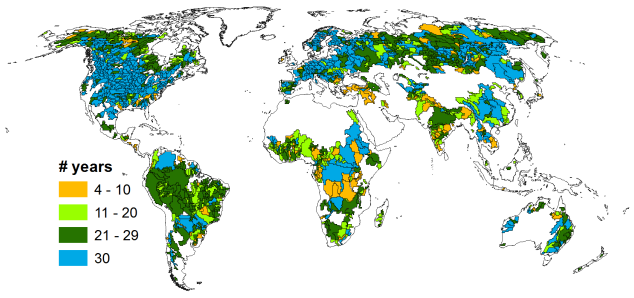


Figure B1. Calibration basins of WaterGAP 2.2 with number of years with observations of discharge used for calibration.

B2 Discharge stations used

Observed discharge time series were provided by the Global Runoff Data Center (GRDC). Following Hunger and Döll (2008), gauging stations listed in the GRDC catalogue (http://grdc.bafg.de/; download date: 28 September 2012) were included in the calibration setup if they fulfilled three main criteria: (1) an upstream area of at least 9000 km^2, (2) a time series of at least 4 (complete) years, and (3) an interstation catchment area of at least 30 000 km^2. All in all, a number of 1319 stations, covering 53.6 % of the global land area except Antarctica and Greenland, was used for calibration (Fig. B1). If available, the 30-year period 1971–2000 was chosen as calibration years.

B3 Regionalization

In order to transfer the calibrated γ values to ungauged basins, the parameter is regionalized using a multiple linear regression approach relating the natural logarithm of the calibrated γ values to the following basin descriptors: mean annual temperature, mean available soil water capacity, fraction of open water bodies, mean basin land surface slope, fraction of permanent snow and ice, and the aquifer-related groundwater recharge factor. Like in calibration basins, the regionalized parameter values are constrained to the range 0.1–5.0. CFA and CFS are not regionalized but are set to 1.0 in uncalibrated basins.

Appendix C: Description of water use models

In preprocessing steps to the WGHM, the global water use submodels (left side of Fig. 1) provide water withdrawal and water consumption (the part of withdrawn water that is not returned to the system but evaporated or incorporated in products) for five sectors: irrigation, livestock farming, domestic use (households and small businesses), manufacturing industries and thermal power plant cooling.

Irrigation water consumption is calculated on daily time steps for each grid cell by the GIM on the basis of gridded area equipped for irrigation (Siebert et al., 2005, 2007) and climate as full irrigation (the difference between potential evapotranspiration and effective precipitation) of paddy rice and nonrice crops, based on modeled cropping patterns (Döll and Siebert, 2002). Consumptive livestock water use is calculated as a function of the number of animals per grid cell and water requirements per capita for 10 different livestock types, while national values of domestic and manufacturing water use are downscaled to the grid cells using population density (Flörke et al., 2013). Cooling water use per grid cell accounts for the location of more than 60 000 power plants, their cooling and combustion type, and their electricity production (Flörke et al., 2013; Vassolo and Döll, 2005). Temporal development of domestic, manufacturing, and cooling water use is calculated as water use intensity per capita or unit of industrial output (considering structural and technological change over time), multiplied by the driving force of water use, either population (for domestic use), national manufacturing output (as gross value added, which is a share of gross domestic product), or national thermal electricity production (Flörke et al., 2013). While WGHM uses aggregated monthly time series of irrigation consumptive use, the other sectoral water uses are distributed equally throughout the year.

Appendix D: Description of GWSWUSE

In the water use models, the source of the abstracted water is not distinguished. This is done in the WaterGAP submodel GWSWUSE (Döll et al., 2012). Based on the results of the water use models, GWSWUSE computes net abstractions (abstractions minus return flows) from groundwater and net abstraction from surface water bodies that serve as input to WGHM (Fig. 1). As a first step within GWSWUSE, the time series of consumptive water use in irrigation, which is computed by GIM for temporally constant irrigation areas but changing climate variables, is scaled by using an annual time series of irrigated area by country uses information (Döll et al., 2012). Then, groundwater use fractions for irrigation (Siebert et al., 2010), domestic and manufacturing water use are applied, and irrigation water abstractions are determined by dividing consumptive use by irrigation water use efficiencies. In contrast to Döll et al. (2012), irrigation water use efficiencies differ between surface water and groundwater use in WaterGAP 2.2. While for surface water irrigation country-specific values are still used, irrigation water use efficiency was set to 0.7 worldwide in case of groundwater irrigation (Döll et al., 2014a). Return flows from irrigation to either groundwater or surface water are computed as a function of the cell-specific artificial drainage fraction (Döll et al., 2012). In WaterGAP 2.2, the fraction of irrigation return flows that recharge groundwater was increased as compared to Döll et al. (2012) and is computed as 0.95–0.75 times the cell-specific artificial drainage fraction. Due to return flows, net abstractions can be positive (water is abstracted from storage) or negative (water is added to storage) (see Fig. 1 of Döll et al., 2014a).

Acknowledgements. The authors thank the Global Runoff Data Center, 56068 Koblenz, Germany, for providing the discharge data used in this study. The authors acknowledge the comments and helpful suggestions of two anonymous referees and the editor that helped to improve the manuscript significantly.

Edited by: K. Stahl

References

Adam, J. C. and Lettenmaier, D. P.: Adjustment of global gridded precipitation for systematic bias, J. Geophys. Res., 108, 4257, doi:10.1029/2002JD002499, 2003.

Alcamo, J., Leemans, R., and Kreileman, E. (Eds.): Global Change Scenarios of the 21st Century – Results from the IMAGE 2.1 Model, Pergamon, Oxford, 1998.

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

Batjes, N. H.: Development of a world data set of soil water retention properties using pedotransfer rules, Geoderma, 71, 31–52, doi:10.1016/0016-7061(95)00089-5, 1996.

Baumgartner, A. and Reichel, E.: The World Water Balance: Mean Annual Global, Continental and Maritime Precipitation, Evaporation and Runoff, Elsevier, Amsterdam, 1975.

Bergström, S.: The HBV model, in: Computer Models of Watershed Hydrology, edited by: Singh, V. P., Water Resources Publications, Lone Tree, USA, 443–476, 1995.

Beven, K. J.: Rainfall-Runoff Modelling. The Primer, John Wiley & Sons Ltd., Chichester, 2001.

Beven, K. J. and Cloke, H. L.: 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, doi:10.1029/2011WR010982, 2012.

Biemans, H., Hutjes, R. W. A., Kabat, P., Strengers, B. J., Gerten, D., and Rost, S.: Effects of precipitation uncertainty on discharge calculations for main river basins, J. Hydrometeorol., 10, 1011–1025, doi:10.1175/2008JHM1067.1, 2009.

Bondeau, A., Smith, P. C., Zaehle, S., Schaphoff, S., Lucht, W., Cramer, W., Gerten, D., Lotze-Campen, H., Müller, C., Reichstein, M., and Smith, B.: Modelling the role of agriculture for the 20th century global terrestrial carbon balance, Global Change Biol., 13, 679–706, doi:10.1111/j.1365-2486.2006.01305.x, 2007.

Butts, M. B., Payne, J. T., Kristensen, M., and Madsen, H.: An evaluation of the impact of model structure on hydrological modelling uncertainty for streamflow simulation, J. Hydrol., 298, 242–266, doi:10.1016/hydrol.2004.03.042, 2004.

Clark, M. P., Slater, A. G., Rupp, D. E., Woods, R. A., Vrugt, J. A., Gupta, H. V., Wagener, T., and Hay, L. E.: Framework for Understanding Structural Errors (FUSE): a modular framework to diagnose differences between hydrological models, Water Resour. Res., 44, W00B02, doi:10.1029/2007WR006735, 2008.

Collins, M., Booth, B. B. B, Harris, G. R., Murphy, J. M., Sexton, D. M. H., and Webb, M. J.: Towards quantifying uncertainty in transient climate change, Clim. Dynam., 27, 127–147, doi:10.1007/s00382-006-0121-0, 2006.

Corzo Perez, G. A., van Huijgevoort, M. H. J., Voß, F., and van Lanen, H. A. J.: On the spatio-temporal analysis of hydrological droughts from global hydrological models, Hydrol. Earth Syst. Sci., 15, 2963–2978, doi:10.5194/hess-15-2963-2011, 2011.

Cowan, W.: Estimating hydraulic roughness coefficients, Agr. Eng., 37, 473–475, 1956.

Dai, A. and Trenberth, K. E.: Estimates of freshwater discharge from continents: latitudinal and seasonal variations, J. Hydrometeorol., 3, 660–687, doi:10.1175/1525-7541(2002)003<0660:EOFDFC>2.0.CO;2, 2002.

Deardorff, J. W.: Efficient prediction of ground surface temperature and moisture, with inclusion of a layer of vegetation, J. Geophys. Res., 83, 1889, doi:10.1029/JC083iC04p01889, 1978.

Döll, P. and Fiedler, K.: Global-scale modeling of groundwater recharge, Hydrol. Earth Syst. Sci., 12, 863–885, doi:10.5194/hess-12-863-2008, 2008.

Döll, P. and Lehner, B.: Validation of a new global 30-min drainage direction map, J. Hydrol., 258, 214–231, doi:10.1016/S0022-1694(01)00565-0, 2002.

Döll, P. and Müller Schmied, H.: How is the impact of climate change on river flow regimes related to the impact on mean annual runoff? A global-scale analysis, Environ. Res. Lett., 7, 014037, doi:10.1088/1748-9326/7/1/014037, 2012.

Döll, P. and Siebert, S.: Global modeling of irrigation water requirements, Water Resour. Res., 38, 8-1–8-10, doi:10.1029/2001WR000355, 2002.

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

Döll, P., Fiedler, K., and Zhang, J.: Global-scale analysis of river flow alterations due to water withdrawals and reservoirs, Hydrol. Earth Syst. Sci., 13, 2413–2432, doi:10.5194/hess-13-2413-2009, 2009.

Döll, P., Hoffmann-Dobrev, H., Portmann, F. T., Siebert, S., Eicker, A., Rodell, M., Strassberg, G., and Scanlon, B. R.: Impact of water withdrawals from groundwater and surface water on continental water storage variations, J. Geodyn., 59–60, 143–156, doi:10.1016/j.jog.2011.05.001, 2012.

Döll, P., Fritsche, M., Eicker, A., and Müller Schmied, H.: Seasonal water storage variations as impacted by water abstractions: comparing the output of a global hydrological model with GRACE and GPS observations, Surv. Geophys., doi:10.1007/s10712-014-9282-2, in press, 2014a.

Döll, P., Müller Schmied, H., Schuh, C., Portmann, F. T., and Eicker, A.: Global-scale assessment of groundwater depletion and related groundwater abstractions: Combining hydrological modeling with information from well observations and GRACE satellites, Water Resour. Res., 50, 5698–5720, doi:10.1002/2014WR015595, 2014b.

European Environment Agency: Corine land cover 2000: mapping a decade of change, Copenhagen, 2004.

Fekete, B. M., Vörösmarty, C. J., and Grabs, W.: High-resolution fields of global runoff combining observed river discharge and simulated water balances, Global Biogeochem. Cy., 16, 15-1–15-10, doi:10.1029/1999GB001254, 2002.

Flörke, M., Kynast, E., Bärlund, I., Eisner, S., Wimmer, F., and Alcamo, J.: Domestic and industrial water uses of the past 60 years as a mirror of socio-economic development: A global simulation study, Global Environ. Change, 23, 144–156, doi:10.1016/j.gloenvcha.2012.10.018, 2013.

Gosling, S. N. and Arnell, N. W.: Simulating current global river runoff with a global hydrological model: model revisions, validation, and sensitivity analysis, Hydrol. Process., 25, 1129–1145, doi:10.1002/hyp.7727, 2011.

Gudmundsson, L., Tallaksen, L., Stahl, K., Clark, D., Dumont, E., Hagemann, S., Bertrand, N., Gerten, D., Heinke, J., Hanasaki, N., Voss, F., and Koirala, S.: Comparing large-scale hydrological model simulations to observed runoff percentiles in Europe, J. Hydrometeorol., 13, 604–620, doi:10.1175/JHM-D-11-083.1, 2012a.

Gudmundsson, L., Wagener, T., Tallaksen, L. M., and Engeland, K.: Evaluation of nine large-scale hydrological models with respect to the seasonal runoff climatology in Europe, Water Resour. Res., 48, 1–20, doi:10.1029/2011WR010911, 2012b.

Güntner, A., Stuck, J., Werth, S., Döll, P., Verzano, K., and Merz, B.: A global analysis of temporal and spatial variations in continental water storage, Water Resour. Res., 43, W05416, doi:10.1029/2006WR005247, 2007.

Guo, Z., Dirmeyer, P. A., Hu, Z.-Z., Gao, X., and Zhao, M.: Evaluation of the second global soil wetness project soil moisture simulations: 2. Sensitivity to external meteorological forcing, J. Geophys. Res., 111, D22S03, doi:10.1029/2006JD007845, 2006.

Haddeland, I., Clark, D. B., Franssen, W., Ludwig, F., Voß, F., Arnell, N. W., Bertrand, N., Best, M., Folwell, S., Gerten, D., Gomes, S., Gosling, S. N., Hagemann, S., Hanasaki, N., Harding, R., Heinke, J., Kabat, P., Koirala, S., Oki, T., Polcher, J., Stacke, T., Viterbo, P., Weedon, G. P., and Yeh, P.: Multimodel estimate of the global terrestrial water balance: setup and first results, J. Hydrometeorol., 12, 869–884, doi:10.1175/2011JHM1324.1, 2011.

Hagemann, S., Chen, C., Clark, D. B., Folwell, S., Gosling, S. N., Haddeland, I., Hanasaki, N., Heinke, J., Ludwig, F., Voss, F., and Wiltshire, A. J.: Climate change impact on available water resources obtained using multiple global climate and hydrology models, Earth Syst. Dynam., 4, 129–144, doi:10.5194/esd-4-129-2013, 2013.

Hanasaki, N., Kanae, S., and Oki, T.: A reservoir operation scheme for global river routing models, J. Hydrol., 327, 22–41, doi:10.1016/j.jhydrol.2005.11.011, 2006.

Hanasaki, N., Kanae, S., Oki, T., Masuda, K., Motoya, K., Shirakawa, N., Shen, Y., and Tanaka, K.: An integrated model for the assessment of global water resources – Part 1: Model description and input meteorological forcing, Hydrol. Earth Syst. Sci., 12, 1007–1025, doi:10.5194/hess-12-1007-2008, 2008.

Harris, I., Jones, P. D., Osborn, T. J., and Lister, D. H.: Updated high-resolution grids of monthly climatic observations – the CRU TS3.10 Dataset, Int. J. Climatol., 34, 623–642, doi:10.1002/joc.3711, 2014.

Hoekstra, A. Y., Mekonnen, M. M., Chapagain, A. K., Mathews, R. E., and Richter, B. D.: Global monthly water scarcity: blue water footprints versus blue water availability, PLoS One, 7, e32688, doi:10.1371/journal.pone.0032688, 2012.

Hunger, M. and Döll, P.: Value of river discharge data for global-scale hydrological modeling, Hydrol. Earth Syst. Sci., 12, 841–861, doi:10.5194/hess-12-841-2008, 2008.

Jasechko, S., Sharp, Z. D., Gibson, J. J., Birks, S. J., Yi, Y., and Fawcett, P. J.: Terrestrial water fluxes dominated by transpiration, Nature, 496, 347–350, doi:10.1038/nature11983, 2013.

Jung, M., Reichstein, M., Ciais, P., Seneviratne, S. I., Sheffield, J., Goulden, M. L., Bonan, G., Cescatti, A., Chen, J., de Jeu, R., Dolman, A. J., Eugster, W., Gerten, D., Gianelle, D., Gobron, N., Heinke, J., Kimball, J., Law, B. E., Montagnani, L., Mu, Q., Mueller, B., Oleson, K., Papale, D., Richardson, A. D., Roupsard, O., Running, S., Tomelleri, E., Viovy, N., Weber, U., Williams, C., Wood, E., Zaehle, S., and Zhang, K.: Recent decline in the global land evapotranspiration trend due to limited moisture supply, Nature, 467, 951–954, doi:10.1038/nature09396, 2010.

Kaspar, F.: Entwicklung und Unsicherheitsanalyse eines globalen hydrologischen Modells, Ph.D. thesis, University of Kassel, Kassel, Germany, 2003.

Koirala, S., Yeh, P. J.-F., Hirabayashi, Y., Kanae, S., and Oki, T.: Global-scale land surface hydrologic modeling with the representation of water table dynamics, J. Geophys. Res.-Atmos., 119, 75–89, doi:10.1002/2013JD020398, 2014.

Korzun, V. I.: World water balance and water resources of the world, UNESCO Stud. Reports Hydrol., 25, 663, 1978.

Kottek, M., Grieser, J., Beck, C., Rudolf, B., and Rubel, F.: World map of the Koppen-Geiger climate classification updated, Meteorol. Z., 15, 259–263, doi:10.1127/0941-2948/2006/0130, 2006.

Legates, D. R.: A climatology of global precipitation, Publ. Climatol. 40, University of Delaware, Newark, USA 1987.

Lehner, B. and Döll, P.: Development and validation of a global database of lakes, reservoirs and wetlands, J. Hydrol., 296, 1–22, doi:10.1016/j.jhydrol.2004.03.028, 2004.

Lehner, B., Verdin, K., and Jarvis, A.: New global hydrography derived from spaceborne elevation data, EOS T. Am. Geophys. Un., 89, 93–94, doi:10.1029/2008EO100001, 2008.

Lehner, B., Liermann, C. R., Revenga, C., Vörösmarty, C., Fekete, B., Crouzet, P., Döll, P., Endejan, M., Frenken, K., Magome, J., Nilsson, C., Robertson, J. C., Rödel, R., Sindorf, N., and Wisser, D.: High-resolution mapping of the world's reservoirs and dams for sustainable river-flow management, Front. Ecol. Environ., 9, 494–502, doi:10.1890/100125, 2011.

Maniak, U.: Hydrologie und Wasserbewirtschaftung, 4th Edn., Springer, Berlin, 1997.

Masaki, Y., Hanasaki, N., Takahashi, K., and Hijioka, Y.: Global-scale analysis on future changes in flow regimes using Gini and Lorenz asymmetry coefficients, Water Resour. Res., 50, 4054–4078, doi:10.1002/2013WR014266, 2014.

Miralles, D. G., Holmes, T. R. H., De Jeu, R. A. M., Gash, J. H., Meesters, A. G. C. A., and Dolman, A. J.: Global land-surface evaporation estimated from satellite-based observations, Hydrol. Earth Syst. Sci., 15, 453–469, doi:10.5194/hess-15-453-2011, 2011.

Mu, Q., Zhao, M., and Running, S. W.: Improvements to a MODIS global terrestrial evapotranspiration algorithm, Remote Sens. Environ., 115, 1781–1800, doi:10.1016/j.rse.2011.02.019, 2011.

Mueller, B., Seneviratne, S. I., Jimenez, C., Corti, T., Hirschi, M., Balsamo, G., Ciais, P., Dirmeyer, P., Fisher, J. B., Guo, Z., Jung, M., Maignan, F., McCabe, M. F., Reichle, R., Reichstein, M., Rodell, M., Sheffield, J., Teuling, A. J., Wang, K., Wood, E. F., and Zhang, Y.: Evaluation of global observations-based evapotranspiration datasets and IPCC AR4 simulations, Geophys. Res. Lett., 38, L06402, doi:10.1029/2010GL046230, 2011.

Mueller, B., Hirschi, M., Jimenez, C., Ciais, P., Dirmeyer, P. A., Dolman, A. J., Fisher, J. B., Jung, M., Ludwig, F., Maignan, F., Miralles, D. G., McCabe, M. F., Reichstein, M., Sheffield, J., Wang, K., Wood, E. F., Zhang, Y., and Seneviratne, S. I.: Benchmark products for land evapotranspiration: LandFlux-EVAL multi-data set synthesis, Hydrol. Earth Syst. Sci., 17, 3707–3720, doi:10.5194/hess-17-3707-2013, 2013.

Nash, J. and Sutcliffe, J.: River flow forecasting through conceptual models part I: A discussion of principles, J. Hydrol., 10, 282–290, 1970.

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

Oki, T. and Kanae, S.: Global hydrological cycles and world water resources, Science, 313, 1068–1072, doi:10.1126/science.1128845, 2006.

Pokhrel, Y., Hanasaki, N., Koirala, S., Cho, J., Yeh, P. J.-F., Kim, H., Kanae, S., and Oki, T.: Incorporating anthropogenic water regulation modules into a land surface model, J. Hydrometeorol., 13, 255–269, doi:10.1175/JHM-D-11-013.1, 2012.

Portmann, F. T., Döll, P., Eisner, S., and Flörke, M.: Impact of climate change on renewable groundwater resources: assessing the benefits of avoided greenhouse gas emissions using selected CMIP5 climate projections, Environ. Res. Lett., 8, 024023, doi:10.1088/1748-9326/8/2/024023, 2013.

Priestley, C. H. B. and Taylor, R. J.: Assessment of surface heat flux and evaporation using large-scale parameters, Mon. Weather Rev., 100, 81–92, doi:10.1175/1520-0493(1972)100<0081:OTAOSH>2.3.CO;2, 1972.

Prudhomme, C., Parry, S., Hannaford, J., Clark, D. B., Hagemann, S., and Voss, F.: How well do large-scale models reproduce regional hydrological extremes in Europe?, J. Hydrometeorol., 12, 1181–1204, doi:10.1175/2011JHM1387.1, 2011.

Prudhomme, C., Giuntoli, I., Robinson, E. L., Clark, D. B., Arnell, N. W., Dankers, R., Fekete, B. M., Franssen, W., Gerten, D., Gosling, S. N., Hagemann, S., Hannah, D. M., Kim, H., Masaki, Y., Satoh, Y., Stacke, T., Wada, Y. and Wisser, D.: Hydrological droughts in the 21st century, hotspots and uncertainties from a global multimodel ensemble experiment, P. Natl. Acad. Sci. USA, 111, 3262–3267, doi:10.1073/pnas.1222473110, 2014.

Refsgaard, J. C., van der Sluijs, J. P., Brown, J., and van der Keur, P.: A framework for dealing with uncertainty due to model structure error, Adv. Water Resour., 29, 1586–1597, doi:10.1016/j.advwatres.2005.11.013, 2006.

Rowlands, D. J., Frame, D. J., Ackerley, D., Aina, T., Booth, B. B. B., Christensen, C., Collins, M., Faull, N., Forest, C. E., Grandey, B. S., Gryspeerdt, E., Highwood, E. J., Ingram, W. J., Knight, S., Lopez, A., Massey, N., Mcnamara, F., Meinshausen, N., Piani, C., Rosier, S. M., Sanderson, B. M., Smith, L. A., Stone, D. A., Thurston, M., Yamazaki, K., Yamazaki, Y. H., and Allen, M. R.: Broad range of 2050 warming from an observationally constrained large climate model ensemble, Nat. Geosci., 5, 256–260, doi:10.1038/ngeo1430, 2012.

Schewe, J., Heinke, J., Gerten, D., Haddeland, I., Arnell, N. W., Clark, D. B., Dankers, R., Eisner, S., Fekete, B. M., Colón-González, F. J., Gosling, S. N., Kim, H., Liu, X., Masaki, Y., Portmann, F. T., Satoh, Y., Stacke, T., Tang, Q., Wada, Y., Wisser, D., Albrecht, T., Frieler, K., Piontek, F., Warszawski, L., and Kabat, P.: Multimodel assessment of water scarcity under climate change., P. Natl. Acad. Sci. USA, 111, 3245–3250, doi:10.1073/pnas.1222460110, 2014.

Schmidt, R., Schwintzer, P., Flechtner, F., Reigber, C., Güntner, A., Döll, P., Ramillien, G., Cazenave, A., Petrovic, S., Jochmann, H., and Wünsch, J.: GRACE observations of changes in continental water storage, Global Planet. Change, 50, 112–126, doi:10.1016/j.gloplacha.2004.11.018, 2006.

Schneider, C., Flörke, M., Eisner, S., and Voss, F.: Large scale modelling of bankfull flow: an example for Europe, J. Hydrol., 408, 235–245, doi:10.1016/j.jhydrol.2011.08.004, 2011.

Schneider, U., Becker, A., Finger, P., Meyer-Christoffer, A., Ziese, M., and Rudolf, B.: GPCC's new land surface precipitation climatology based on quality-controlled in situ data and its role in quantifying the global water cycle, Theor. Appl. Climatol., 115, 15–40, doi:10.1007/s00704-013-0860-x, 2014.

Schulze, E. D., Kelliher, F. M., Korner, C., Lloyd, J., and Leuning, R.: Relationships among maximum stomatal conductance, ecosystem surface conductance, carbon assimilation rate, and plant nitrogen nutrition: A global ecology scaling exercise, Annu. Rev. Ecol. Syst., 25, 629–662, doi:10.1146/annurev.es.25.110194.003213, 1994.

Schulze, K. and Döll, P.: Neue Ansätze zur Modellierung von Schneeakkumulation und -schmelze im globalen Wassermodell WaterGAP, in: Tagungsband zum 7. Workshop zur großskaligen Modellierung in der Hydrologie, edited by: Ludwig, R., Reichert, D., and Mauser, W., Kassel University Press, Kassel, 2004.

Schumacher, M., Eicker, A., Kusche, J., Müller Schmied, H., and Döll, P.: Covariance analysis and sensitivity studies for GRACE assimilation into WGHM, in: IAG Symposia Series: Proceedings of the IAG Scientific Assembly 2013, accepted, 2014.

Scurlock, J. M., Asner, G. P., and Gower, S. T.: Worldwide Historical Estimates of Leaf Area Index, 1932–2000, Oak Ridge National Library, Oak Ridge, USA, 2001.

Shuttleworth, W. J.: Evaporation, in: Handbook of Hydrology, edited by: Maidment, D. R., McGraw-Hill, New York, 4.1–4.53, 1993.

Siebert, S., Döll, P., Hoogeveen, J., Faures, J.-M., Frenken, K., and Feick, S.: Development and validation of the global map of irrigation areas, Hydrol. Earth Syst. Sci., 9, 535–547, doi:10.5194/hess-9-535-2005, 2005.

Siebert, S., Döll, P., Feick, S., Hoogeveen, J., and Frenken, K.: Global map of irrigation areas version 4.0.1, CD-ROM, FAO Land and Water Digital Media Series 34, Rome, Italy, FAO, 2007.

Siebert, S., Burke, J., Faures, J. M., Frenken, K., Hoogeveen, J., Döll, P., and Portmann, F. T.: Groundwater use for irrigation – a global inventory, Hydrol. Earth Syst. Sci., 14, 1863–1880, doi:10.5194/hess-14-1863-2010, 2010.

Song, X., Zhan, C., Kong, F., and Xia, J.: Advances in the study of uncertainty quantification of large-scale hydrological modeling system, J. Geogr. Sci., 21, 801–819, doi:10.1007/s11442-011-0881-2, 2011.

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

Sterling, S. M., Ducharne, A., and Polcher, J.: The impact of global land-cover change on the terrestrial water cycle, Nat. Clim. Change, 3, 385–390, doi:10.1038/nclimate1690, 2012.

Thompson, J. R., Green, A. J., Kingston, D. G., and Gosling, S. N.: Assessment of uncertainty in river flow projections for the Mekong River using multiple GCMs and hydrological models, J. Hydrol., 486, 1–30, doi:10.1016/j.jhydrol.2013.01.029, 2013.

UNEP: World Atlas of Desertification, Edward Arnold, Sevenoaks, 1992.

US Geological Survey: GTOPO30 Digital Elevation Model, available at: http://webmap.ornl.gov/wcsdown/dataset.jsp?dsid=10003 (last access: 12 December 2013), 2003.

USGS: Global Land Cover Characterization, available at: http://edc2.usgs.gov/glcc/glcc.php (last access: 22 October 2012), 2008.

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

Van Loon, A. F., Van Huijgevoort, M. H. J., and Van Lanen, H. A. J.: Evaluation of drought propagation in an ensemble mean of large-scale hydrological models, Hydrol. Earth Syst. Sci., 16, 4057–4078, doi:10.5194/hess-16-4057-2012, 2012.

Vassolo, S. and Döll, P.: Global-scale gridded estimates of thermoelectric power and manufacturing water use, Water Resour. Res., 41, W04010, doi:10.1029/2004WR003360, 2005.

Verzano, K., Bärlund, I., Flörke, M., Lehner, B., Kynast, E., Voß, F., and Alcamo, J.: Modeling variable river flow velocity on continental scale: current situation and climate change impacts in Europe, J. Hydrol., 424-425, 238–251, doi:10.1016/j.jhydrol.2012.01.005, 2012.

Vinukollu, R. K., Wood, E. F., Ferguson, C. R., and Fisher, J. B.: Global estimates of evapotranspiration for climate studies using multi-sensor remote sensing data: evaluation of three process-based approaches, Remote Sens. Environ., 115, 801–823, doi:10.1016/j.rse.2010.11.006, 2011.

Vörösmarty, C., Federer, C., and Schloss, A.: Potential evaporation functions compared on US watersheds: possible implications for global-scale water balance and terrestrial ecosystem modeling, J. Hydrol., 207, 147–169, doi:10.1016/S0022-1694(98)00109-7, 1998.

Wada, Y., van Beek, L. P. H., van Kempen, C. M., Reckman, J. W. T. M., Vasak, S., and Bierkens, M. F. P.: Global depletion of groundwater resources, Geophys. Res. Lett., 37, L20402, doi:10.1029/2010GL044571, 2010.

Wang, K. and Liang, S.: An improved method for estimating global evapotranspiration based on satellite determination of surface net radiation, vegetation index, temperature, and soil moisture, J. Hydrometeorol., 9, 712–727, doi:10.1175/2007JHM911.1, 2008.

Warszawski, L., Frieler, K., Huber, V., Piontek, F., Serdeczny, O., and Schewe, J.: The Inter-Sectoral Impact Model Intercomparison Project (ISI-MIP): project framework, Proc. Natl. Acad. Sci. USA, 111, 3228–3232, doi:10.1073/pnas.1312330110, 2014.

Weedon, G. P., Gomes, S., Viterbo, P., Shuttleworth, W. J., Blyth, E., Österle, H., Adam, J. C., Bellouin, N., Boucher, O., and Best, M.: Creation of the WATCH Forcing Data and its use to assess global and regional reference crop evaporation over land during the twentieth century, J. Hydrometeorol., 12, 823–848, doi:10.1175/2011JHM1369.1, 2011.

Weedon, G. P., Balsamo, G., Bellouin, N., Gomes, S., Best, M. J., and Viterbo, P.: The WFDEI meteorological forcing data set: WATCH Forcing Data methodology applied to ERA-Interim reanalysis data, Water Resour. Res., doi:10.1002/2014WR015638, accepted, 2014.

Werth, S. and Güntner, A.: Calibration analysis for water storage variability of the global hydrological model WGHM, Hydrol. Earth Syst. Sci., 14, 59–78, doi:10.5194/hess-14-59-2010, 2010.

Widén-Nilsson, E., Halldin, S., and Xu, C.: Global water-balance modelling with WASMOD-M: parameter estimation and regionalisation, J. Hydrol., 340, 105–118, doi:10.1016/j.jhydrol.2007.04.002, 2007.

Wilber, A. C., Kratz, D. P., and Gupta, S. K.: Surface emissivity maps for use in satellite retrievals of longwave radiation, NASA/TP-1999-209362, Langley Research Center, Hampton, USA, 1999.

WMO: Guide to hydrological practices, Geneva, 1994.

Wood, E. F., Roundy, J. K., Troy, T. J., van Beek, R. L. P. H., Bierkens, M. F. P., Blyth, E., de Roo, A., Döll, P., Ek, M., Famiglietti, J., Gochis, D., van de Giesen, N., Houser, P., Jaffé, P. R., Kollet, S., Lehner, B., Lettenmaier, D. P., Peters-Lidard, C., Sivapalan, M., Sheffield, J., Wade, A., and Whitehead, P.: Hyperresolution global land surface modeling: meeting a grand challenge for monitoring Earth's terrestrial water, Water Resour. Res., 47, W05301, doi:10.1029/2010WR010090, 2011.

Wood, E. F., Roundy, J. K., Troy, T. J., van Beek, R. L. P. H., Bierkens, M. F. P., Blyth, E., de Roo, A., Döll, P., Ek, M., Famiglietti, J., Gochis, D., van de Giesen, N., Houser, P., Jaffe, P., Kollet, S., Lehner, B., Lettenmaier, D. P., Peters-Lidard, C. D., Sivapalan, M., Sheffield, J., Wade, A. J., and Whitehead, P.: Reply to comment by Keith J. Beven and Hannah L. Cloke on "Hyperresolution global land surface modeling: meeting a grand challenge for monitoring Earth's terrestrial water", Water Resour. Res., 48, W01802, doi:10.1029/2011WR011202, 2012.

4 Variations of global and continental water balance components as impacted by climate forcing uncertainty and human water use

Full citation: Müller Schmied, H., Adam, L., Eisner, S., Fink, G., Flörke, M., Kim, H., Oki, T., Portmann, F. T., Reinecke, R., Riedel, C., Song, Q., Zhang, J. and Döll, P.: Variations of global and continental water balance components as impacted by climate forcing uncertainty and human water use, Hydrol. Earth Syst. Sci., 20, 2877-2898, doi:10.5194/hess-20-2877-2016, 2016.

Declaration of portion of own work

H. Müller Schmied has shared with S. Eisner and F. T. Portmann model code development. The model experiment setup was shaped mainly by H. Müller Schmied, P. Döll and C. Riedel. All model runs (except water use, which was done by S. Eisner and F. T. Portmann), analyses, figures and tables were done by H. Müller Schmied. The draft of the manuscript was written by H. Müller Schmied and revised by the co-authors. In overall percentage, H. Müller Schmied contributed around 80% of this manuscript.

Hannes Müller Schmied

Hydrol. Earth Syst. Sci., 20, 2877–2898, 2016
www.hydrol-earth-syst-sci.net/20/2877/2016/
doi:10.5194/hess-20-2877-2016

Hydrology and Earth System Sciences
Open Access
EGU

Variations of global and continental water balance components as impacted by climate forcing uncertainty and human water use

Hannes Müller Schmied[1,2], Linda Adam[1], Stephanie Eisner[3], Gabriel Fink[3], Martina Flörke[3], Hyungjun Kim[4], Taikan Oki[4], Felix Theodor Portmann[1], Robert Reinecke[1], Claudia Riedel[1], Qi Song[1], Jing Zhang[1], and Petra Döll[1]

[1]Institute of Physical Geography, Goethe-University Frankfurt, Frankfurt, Germany
[2]Senckenberg Biodiversity and Climate Research Centre (BiK-F), Frankfurt, Germany
[3]Center for Environmental Systems Research (CESR), University of Kassel, Kassel, Germany
[4]Institute of Industrial Science, The University of Tokyo, Tokyo, Japan

Correspondence to: Hannes Müller Schmied (hannes.mueller.schmied@em.uni-frankfurt.de)

Received: 7 December 2015 – Published in Hydrol. Earth Syst. Sci. Discuss.: 15 January 2016
Revised: 25 April 2016 – Accepted: 29 June 2016 – Published: 18 July 2016

Abstract. When assessing global water resources with hydrological models, it is essential to know about methodological uncertainties. The values of simulated water balance components may vary due to different spatial and temporal aggregations, reference periods, and applied climate forcings, as well as due to the consideration of human water use, or the lack thereof. We analyzed these variations over the period 1901–2010 by forcing the global hydrological model WaterGAP 2.2 (ISIMIP2a) with five state-of-the-art climate data sets, including a homogenized version of the concatenated WFD/WFDEI data set. Absolute values and temporal variations of global water balance components are strongly affected by the uncertainty in the climate forcing, and no temporal trends of the global water balance components are detected for the four homogeneous climate forcings considered (except for human water abstractions). The calibration of WaterGAP against observed long-term average river discharge Q significantly reduces the impact of climate forcing uncertainty on estimated Q and renewable water resources. For the homogeneous forcings, Q of the calibrated and non-calibrated regions of the globe varies by 1.6 and 18.5 %, respectively, for 1971–2000. On the continental scale, most differences for long-term average precipitation P and Q estimates occur in Africa and, due to snow undercatch of rain gauges, also in the data-rich continents Europe and North America. Variations of Q at the grid-cell scale are large, except in a few grid cells upstream and downstream of calibration stations, with an average variation of 37 and 74 % among the four homogeneous forcings in calibrated and non-calibrated regions, respectively. Considering only the forcings GSWP3 and WFDEI_hom, i.e., excluding the forcing without undercatch correction (PGFv2.1) and the one with a much lower shortwave downward radiation SWD than the others (WFD), Q variations are reduced to 16 and 31 % in calibrated and non-calibrated regions, respectively. These simulation results support the need for extended Q measurements and data sharing for better constraining global water balance assessments. Over the 20th century, the human footprint on natural water resources has become larger. For 11–18% of the global land area, the change of Q between 1941–1970 and 1971–2000 was driven more strongly by change of human water use including dam construction than by change in precipitation, while this was true for only 9–13 % of the land area from 1911–1940 to 1941–1970.

1 Introduction

Assessment of global-scale water resources and water balance components is of importance for water resources management at global, continental, and river basin scales (Vörösmarty et al., 2015). Many data-based, model-based, and hybrid approaches exist in order to quantify macro-scale water balance components (Baumgartner and Reichel, 1975; Fekete et al., 2002; Haddeland et al., 2011; Müller Schmied et al., 2014; Oki and Kanae, 2006). For water resources management, especially the estimation of renewable freshwater resources (long-term average runoff or river discharge) is of

importance, as it is the source for both human and ecosystem needs. As adequate discharge observations are available only at selected locations (see the catalogue of the Global Runoff Data Centre (GRDC), http://grdc.bafg.de/), model-based or hybrid (i.e., incorporating historical discharge observations) approaches to estimating discharge and other water balance components are of increasing importance. Since the 1980s, global hydrological models (GHMs) have been developed to calculate the water balance on global and/or continental scales. Recent reviews of such models are presented by Bierkens (2015), Sood and Smakhtin (2015), and Trambauer et al. (2013).

All GHMs are driven by climate forcing input data sets (hereafter called climate forcings), based on station observations (e.g., for precipitation and air temperature), reanalysis (global circulation models for numerical weather prediction, which assimilate all available up-to-date data for current time step), and/or remote sensing data (e.g., for radiation). Within the last 2 decades, numerous climate forcings were developed with a current standard of at least daily temporal resolution and 0.5° by 0.5° spatial resolution (the common GHM spatial resolution), providing data from as early as 1901 until recent years. These climate forcings differ among each other and thus may lead to different water resources estimates by GHMs.

Humans have altered the global water cycle with an increasing intensity, e.g., due to irrigation or industrial water use (Döll and Siebert, 2002; Döll et al., 2012; Flörke et al., 2013; Siebert et al., 2015; Wada et al., 2010). A number of GHMs (but not all) are able to incorporate human water use in their calculations (see Table 2 in Bierkens, 2015). Neglecting anthropogenic water consumption prevents meaningful water resources assessments, at least in regions with high water consumption relative to renewable resources (e.g., High Plains aquifer, Indus, Ganges–Brahmaputra). For example, groundwater depletion as observed by falling groundwater heads in wells and by GRACE satellite observations of gravity variations can only be modeled when human water use is considered (Döll et al., 2014).

Simulated water balance components vary considerably due to various uncertainties of GHMs (Haddeland et al., 2011; Schewe et al., 2014) including human water use, model improvements over time (e.g., see the different results of the Water Global Assessment and Prognosis (WaterGAP) model in Müller Schmied et al. (2014), their Table 5), and climate forcing (Biemans et al., 2009; Voisin et al., 2008) as well as uncertainties in discharge observations (Coxon et al., 2015; McMillan et al., 2012). In addition to these uncertainties, water resources estimates differ due to different reference periods (Wisser et al., 2010).

This study contributes to the assessment of water balance components on a global and continental scale by answering the following research questions.

1. What is the impact of climate forcing uncertainty on water balance components at global, continental, and grid-cell scale?

2. What is the variation of estimated global water balance components for different temporal aggregations: year, decade, 30 years, and century?

3. What determines variations of long-term average river discharge between consecutive 30-year periods more strongly: either change of precipitation or change of human water use and dam construction creating reservoirs and regulated lakes (anthropogenic impact)?

To answer these questions, we conducted a modeling experiment. The model, data, and methods are described in Sect. 2. Results are presented and discussed in Sect. 3. Finally, conclusions are drawn and an outlook is given.

2 Data and methods

In this study, the global water availability and water use model WaterGAP (Alcamo et al., 2003; Döll et al., 2003) was applied in a modified version of WaterGAP 2.2 (Müller Schmied et al., 2014) in two water use and management variants (including and excluding anthropogenic effects). The model was driven by four state-of-the-art climate forcings provided by the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP) in its phase 2a (https://www.isimip.org/about/#simulation-rounds-isimip2a) and a fifth homogenized forcing.

2.1 GHM WaterGAP 2.2 (ISIMIP2a)

The spatial resolution of WaterGAP is 0.5° by 0.5° (∼ 55 km by 55 km at the equator), and the model uses daily time steps for calculation. The WaterGAP water use models compute water use estimates for five sectors (irrigation, domestic, manufacturing, cooling water for electricity generation, and livestock) that are processed by the GroundWater Surface Water USE (GWSWUSE) submodule to quantify both net water abstractions from surface water and from groundwater (Fig. 1 in Müller Schmied et al., 2014). Taking into account the net abstractions, the WaterGAP Global Hydrology Model (WGHM) calculates changes in water storage compartments as well as water flows between these compartments based on water balance equations, including groundwater recharge, evapotranspiration, and river discharge. A description of model version WaterGAP 2.2 can be found in Müller Schmied et al. (2014). The version used for this study is named WaterGAP 2.2 (ISIMIP2a), and differences to WaterGAP 2.2 mainly consider requirements of the ISIMIP project phase 2a as described in Appendix A.

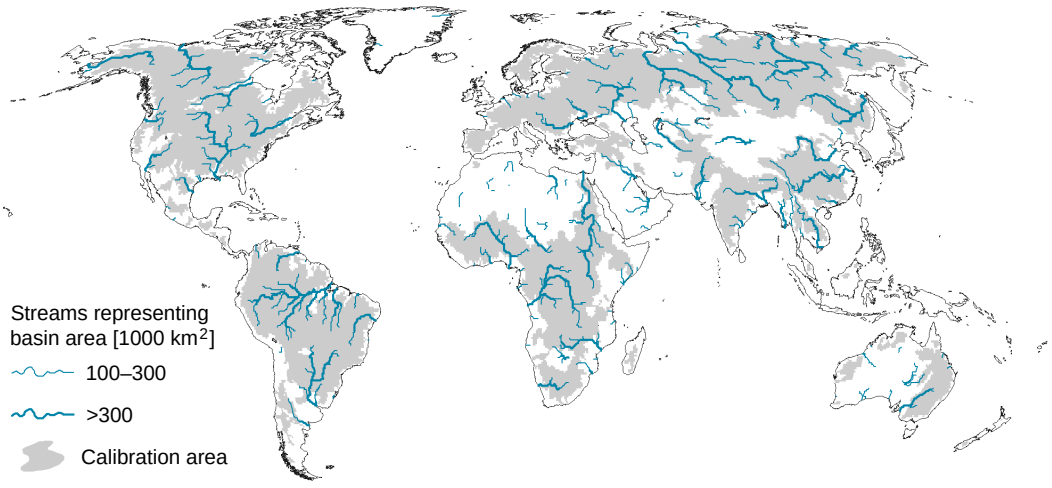


Figure 1. Global land area affected by WaterGAP 2.2 (ISIMIP2a) calibration (grey shading) against observed long-term average river discharge. Streamflow directions and flow accumulation are based on the drainage direction map DDM30 with 0.5° resolution (Döll and Lehner, 2002).

2.2 Calibration of WaterGAP 2.2 (ISIMIP2a) against observed streamflow

The purpose of WaterGAP has been to provide a best estimate of renewable water resources worldwide. To obtain meaningful estimates of water resources despite different sources of uncertainty related to GHMs, a calibration routine was applied (see Döll et al., 2003; Hunger and Döll, 2008; Müller Schmied et al., 2014). The calibration routine in WaterGAP 2.2 (ISIMIP2a) forces the long-term annual simulated river discharge (Q) to be equal (within ±1 %) to observed long-term annual discharge at grid cells representing calibration stations, for the period of observations (with a maximum of 30 years of observations considered). With alternative climate forcings, basin-scale differences in Q and (subsequent) actual evapotranspiration (AET) therefore occur especially in catchments without calibration stations or during years without observed discharge. Figure 1 shows the land grid cells that are affected by calibration in this study, incorporating around 54 % of the global land surface (excluding Antarctica and Greenland). We calibrated the model for each of the four climate forcings GSWP3, PGFv2, WFD and WFDEI_hom (descriptions of acronyms in Sect. 2.3) against mean annual discharge at 1319 discharge observation stations from the Global Runoff Data Centre (GRDC) catalogue, except for WFD, where due to the earlier end of the forcing time series, only 1312 stations could be used. The calibration parameters of WFDEI_hom were then used for the WFD_WFDEI forcing. Observation stations were selected such that the upstream area was a minimum of 9000 km^2. To avoid including stations that are located very close to each other along a river, the minimum interstation catchment area was set to 30 000 km^2. Furthermore, a station was selected only if a minimum of 4 complete years of data were available.

2.3 Climate forcing data sets

Within the ISIMIP project phase 2a, four state-of-the-art climate forcings were made available through the coordinating Potsdam Institute for Climate Impact Research (PIK): GSWP3, PGFv2, WFD, and WFD_WFDEI. For each forcing, daily values of the variables surface-level (raingage-level) precipitation (P), 2 m air temperature (T), shortwave downward radiation at the surface level (SWD), and longwave downward radiation at the surface level (LWD) were used to run WaterGAP. Due to inhomogeneity problems during overlapping periods of WATCH Forcing Data based on ERA-40 (WFD data set, 1901–2001) and WFD methodology applied to ERA-Interim (WFDEI data set, 1979–2010), a data homogenization method was applied. This resulted in a fifth homogenized climate forcing (WFDEI_hom). The name of the climate forcing is used to name the model variant. In all data sets, daily precipitation estimates were obtained by bias correcting output of weather models by monthly precipitation data sets that had been derived from monthly precipitation observed at raingages. These monthly data sets were optimized for spatial coverage, i.e., using, for each month, the available number of gauging stations. The temporally variable number of precipitation observations makes the applied precipitation data sets less suitable for the analysis of temporal variations. While a temporally homogeneous data set of observation-based monthly precipitation exists at least for the time period 1950–2000, it is based on less than 10 000 gauging stations and therefore provides a spatially less accurate representation of global-scale precipitation (Beck et al.,

2005) than the data sets used in this study, which include up to 50 000 gauging stations (Schneider et al., 2015).

2.3.1 Global Soil Wetness Project 3 (GSWP3)

For the third phase of Global Soil Wetness Project (GSWP), a century-long (1901–2010) high-resolution global climate data was developed (http://hydro.iis.u-tokyo.ac.jp/GSWP3). The 20th Century Reanalysis (20CR) project done with the NCEP atmosphere land model (Compo et al., 2011) which has a relatively low spatial resolution ($\sim 2.0°$) and long-term availability (140 years) was dynamically downscaled into the global T248 ($\sim 0.5°$) resolution using Experimental Climate Prediction Center (ECPC) Global Spectral Model (GSM) by spectral nudging data assimilation technique (Yoshimura and Kanamitsu, 2008). Also, Global Precipitation Climatology Centre (GPCC) version 6 (for P), Climate Research Unit (CRU) TS3.21 (for T), and Surface Radiation Budget project (SRB, for SWD/LWD) were used for bias correction to reduce model-dependent uncertainty. Wind-induced P undercatch correction is applied depending on gauge type and their global distribution according to Hirabayashi et al. (2008).

2.3.2 Princeton Global Meteorological Forcing Dataset (PGFv2.1)

The Princeton Global Meteorological Forcing Dataset, version 2 (PGFv2) is an update of the forcing described by Sheffield et al. (2006). It blends reanalysis data (NCEP-NCAR) with station and satellite observations and covers the period 1901–2012 in its current form (http://hydrology.princeton.edu/data.pgf.php). P is bias corrected to monthly CRU TS3.21 but is not undercatch corrected (different to its previous version 1). Daily T is adjusted to match CRU TS3.21 monthly values by shifting. SWD is adjusted for systematic biases at monthly scale (using a product from the University of Maryland (by Rachel Pinker) developed within the NASA MEaSUREs project) and then for trends using CRU TS3.21 cloud cover. LWD is scaled to match the mean and variability of the University of Maryland data (see SWD) but retains the year-to-year variation of the NCEP data. All information on PGFv2 is based on personal communication with J. Sheffield in 2015. During the first review process of this paper, we were informed about an error in the T data for the period 1901–1947 for certain regions. We therefore present results below of a WaterGAP run driven by the corrected version PGFv2.1 but with calibration parameters determined by using PGFv2, as no significant effect of the erroneous T data on calibration is expected, because calibration periods start after 1947 (except for 21 basins that are all located in regions where the error effect is small).

2.3.3 WATCH Forcing Data (WFD)

The WATCH Forcing Data (WFD) was developed by Weedon et al. (2010, 2011) in the scope of the European FP6-funded Water and Global Change (WATCH) project (http://www.eu-watch.org). The data set is based on the European Centre for Medium-Range Weather Forecasts (ECMWF) 40-year reanalysis product (ERA-40) for the period 1958–2001 and on the reordered ERA-40 data for the period 1901–1957. The variables from ERA-40 are interpolated (taking into account elevation) and some are corrected to monthly observation data, e.g., P is corrected using GPCC version 4 observations (details in Weedon et al., 2010, 2011). Monthly P is corrected for wind-induced undercatch according to Adam and Lettenmaier (2003). Monthly T is corrected to CRU TS2.1 and SWD is corrected to cloud cover of CRU TS2.1, whereas LWD is not bias corrected (Weedon et al., 2010).

2.3.4 Combined WFD and WFDEI (WFD_WFDEI)

The WFDEI data set was created by applying the WFD methodology to the newer ERA-Interim reanalysis data of ECMWF, which is improved compared to ERA-40, especially for SWD (Weedon et al., 2014). WFDEI is available for the period 1979–2010, with P bias corrected to GPCC version 5 (and version 6 for 2010) and using ratios from Adam and Lettenmaier (2003) for correction of P undercatch. SWD in WFDEI is larger than SWD of WFD almost everywhere on the globe, with differences between 15 and $100\,\mathrm{W\,m^{-2}}$ in most of Africa and Europe, due to changes in aerosol distribution in ERA-Interim as compared to ERA-40 (Dee et al., 2011; Weedon et al., 2014). Monthly values for T are bias corrected to CRU TS3.1/3.21 and SWD to cloud cover of CRU TS3.1/3.21. WFD_WFDEI, as provided by ISIMIP2a, is a simple time-consecutive combination of WFD (1901–1978) and WFDEI (1979–2010), which can be problematic when not checking for offsets (Weedon et al., 2014). Müller Schmied et al. (2014) used the same concatenating approach and found considerable offsets in WaterGAP simulated water balance components. Due to the strong global increase in SWD in WFDEI relative to WFD for overlapping periods (1979–2001), global AET increased by $\sim 5000\,\mathrm{km^3\,yr^{-1}}$, which affects resulting water storages and global sums of Q (Müller Schmied et al., 2014).

2.3.5 Homogenized combined WFD and WFDEI (WFDEI_hom)

To overcome the offset in selected climatic variables between WFD and WFDEI, a homogenization approach analog to the bias correction approach in Haddeland et al. (2012) was applied to the daily data for three climatic variables (SWD, LWD, and T). For SWD and LWD, a multiplicative approach was applied (Eq. 1), whereas T was homogenized with an

additive approach due to possible zero values (Eq. 2), and P was not homogenized as only marginal differences in continental and global sums occur (Table 4).

$$V_{\mathrm{hom}} = V_{\mathrm{WFD}} \cdot \frac{\overline{V_{\mathrm{WFDEI}}}(m)}{\overline{V_{\mathrm{WFD}}}(m)} \quad (1)$$

$$V_{\mathrm{hom}} = V_{\mathrm{WFD}} + \overline{V_{\mathrm{WFDEI}}}(m) - \overline{V_{\mathrm{WFD}}}(m), \quad (2)$$

with V_{hom} being the homogenized daily variable (1901–2001), V_{WFD} the original daily variable from WFD (1901–2001), and $\overline{V_{\mathrm{WFDEI}}}(m)$ and $\overline{V_{\mathrm{WFD}}}(m)$ the long-term mean monthly variable from WFDEI and WFD for the overlapping time period 1979–2001, applied to the current month (m). The final homogenized daily WFDEI_hom time series consists of homogenized WFD data until 1979 and of WFDEI data afterwards. As the averages of SWD and T during the overlapping period are larger for WFDEI than for WFD, WFDEI_hom values until 1978 are larger than respective original WFD values, also included in WFD_WFDEI time series. The opposite is true for LWD, which is furthermore only slightly adjusted compared to SWD.

2.4 Calculation of spatial averages and indicators

2.4.1 Calculation of spatial averages

The calculation of global averages for climate forcing variables as well as water balance components are based on all land grid cells excluding Antarctica (not represented), Greenland, and those grid cells that represent inland sinks. For T, SWD, and LWD, area-weighted averages were calculated. Q was calculated for global totals by summing up Q of all grid cells that are outflow cells into the ocean according to the drainage direction map DDM30 (Döll and Lehner, 2002) and Q into all grid cells that represent inland sinks. The same procedure was used for the continental assessment (with all of the Russian Federation considered to belong to Europe in this study). For the calibrated and non-calibrated regions, the sum of net cell runoff (Q flowing out of the grid cell minus Q flowing into the grid cell) was used.

2.4.2 Indicator for relative dominance of precipitation or anthropogenic impact on discharge variability

To answer research question 3, i.e., to determine whether the change of long-term average discharge between two consecutive 30-year periods is caused mainly by the change of P in the upstream river basin or by the change of anthropogenic impact on Q by human water use and dam construction, two indicators were developed and combined. In the equations below, Q represents simulated discharge under anthropogenic conditions, whereas *Qnat* is the discharge that would occur with neither human water use nor reservoirs or regulated lake regulation by dams.

First, we assume that P change cannot be a more dominant driver than change of anthropogenic impacts if P increases while Q decreases (and vice versa), expressed by the ratio of differences in Eq. (3). Furthermore, the runoff coefficient scales the ratio. Thus, indicator A_n is computed as

$$A_n = C_{QP,n} \frac{P_{\mathrm{bas}(n),t2} - P_{\mathrm{bas}(n),t1}}{Q_{n,t2} - Q_{n,t1}}, \quad (3)$$

where A_n (–) is the indicator for dominance of P of grid cell n with $P_{\mathrm{bas}(n)}$ (km^3 yr^{-1}) as sum of P for the upstream area (contributing basin area) and Q_n (km^3 yr^{-1}) the simulated river discharge of the grid cell between the time periods $t1$ (e.g., 1941–1970) and $t2$ (e.g., 1971–2000). The runoff coefficient $C_{QP,n}$ (–) is calculated as the averaged mean runoff coefficient of the two time periods under consideration

$$C_{QP,n} = \mathrm{avg}\left(\frac{Qnat_{n,t1}}{P_{\mathrm{bas}(n),t1}}, \frac{Qnat_{n,t2}}{P_{\mathrm{bas}(n),t2}}\right), \quad (4)$$

where $Qnat_n$ (km^3 yr^{-1}) is the simulated river discharge of the grid cell of the model runs without human water abstractions and reservoir operation. The runoff coefficient is independently calculated for the two time periods.

If changes of P and Q have the same sign, A_n is positive, and the change in P may be a significant driver of the Q change. If A_n is negative, it can be excluded that the change of P is a dominant driver of the change in Q.

Indicator B_n quantifies the anthropogenic impact on river discharge, expressed as the change in the difference between Q and *Qnat* compared to the change in Q. An increasing difference between Q and *Qnat* between the periods should lead to a decrease of Q.

$$B_n = \frac{\left(Q_{n,t2} - Qnat_{n,t2}\right) - \left(Q_{n,t1} - Qnat_{n,t1}\right)}{Q_{n,t2} - Q_{n,t1}}, \quad (5)$$

where B_n (–) is the indicator for dominance of anthropogenic impact on river discharge ranging from negative values, zero (for $Q = Qnat$), to positive values. If, e.g., Q increases between the two time periods but the difference between Q and *Qnat* decreases, e.g., due to decreased human water use among the time periods, B_n becomes negative, indicating that anthropogenic effects cannot be the dominant driver of change in Q.

The larger $A_n(B_n)$, the more likely P (anthropogenic effects) is the dominant driver of Q change, since the change in P (anthropogenic effects) is large. Consequently, P is a more dominant driver than change in anthropogenic impact if $A_n > B_n$ and $A_n > 0$. The change in anthropogenic impact is the more dominant driver than change in P if $B_n > A_n$ and $B_n > 0$. If both $A_n <= 0$ and $B_n <= 0$, changes in Q are neither consistent with changes in P nor with changes in anthropogenic impact, and Q change is caused by other drivers, e.g., T. No assessment is possible if there is no change in Q. To illustrate the indicator of relative dominance approach, Table 1 lists indicator values and underlying data for the example of four grid cells representing discharge of large rivers near the outlet to the ocean.

Table 1. Examples of indicator calculation (Sect. 2.4.2) for four large river basins at grid cells located near the outflow to the ocean for the forcing GSWP3 and changes from 1941–1970 ($t1$) to 1971–2000 ($t2$). Values for latitude and longitude in decimal degrees, values for A_n and B_n are dimensionless, other numbers are in $km^3\ yr^{-1}$. Explanations of variables other than lat and long, see Sect. 2.4.2.

	Rhine River	Congo River	Colorado River	Yellow River
Lat	4.25	12.25	−114.75	133.25
Long	52.25	−6.25	31.75	48.25
$P_{bas(n),t1}$	169.36	5735.52	191.24	771.92
$P_{bas(n),t2}$	176.43	5469.11	206.56	771.91
$Qnat_{n,t1}$	69.27	1370.46	1.53	215.28
$Qnat_{n,t2}$	75.19	1251.09	1.92	209.94
$Q_{n,t1}$	67.83	1370.46	0.62	213.41
$Q_{n,t2}$	72.63	1250.67	0.10	203.68
A_n	0.61	0.52	−0.26	0.00
B_n	−0.23	0.00	1.76	0.45
Dominant driver	$A_n > B_n$ and $A_n > 0$: precipitation		$B_n > A_n$ and $B_n > 0$: human impact	

3 Results and discussion

3.1 Water balance components as impacted by climate forcing uncertainty

In this section, uncertainties of climate forcing are described first, followed by uncertainties of model output variables stemming from climate forcing uncertainty (Fig. 2). Spatial scales range from global (Table 2) to continental (Table 3) and to grid cells (Fig. 3). In addition, we differentiate between calibrated and non-calibrated regions (Table 5). Finally, values of water balance components are compared to values from other studies (Table 6).

3.1.1 Uncertainty of global climate forcings

The 1971–2000 global P differs among the model forcing variants, with the largest difference found between the CRU-based product (PGFv2.1) and the GPCC-based products (all other forcings) amounting up to $7500\,km^3\,yr^{-1}$ (Table 2). Even the GPCC-based forcings vary by up to $1400\,km^3\,yr^{-1}$ (exceeding the amount of actual water consumption WCa). Oceania (with the lowest absolute value) has the lowest deviation among the forcings (Table 3). The largest deviations are found in North America, Europe, and Africa. In North America and Europe, where the station density is comparably high and GPCC versions agree very well (Table 4) but in winter precipitation falls often as snow (with strong undercatch in gauging devices), the different approaches to undercatch correction of P lead to large P deviations among the climate forcings. In case of WFD and WFDEI, monthly precipitation data are undercatch corrected according to Adam and Lettenmaier (2003); in the case of GSWP3, a correction described in Hirabayashi et al. (2008) is applied, while there is no undercatch correction in PGFv2.1. While the calibrated grid cells cover 53.9 % (53.7 % for WFD) of global land area (excluding Antarctica and Greenland), they receive 61.0–61.5 % of P (for all forcings, Table 5). The variation among the forcing variants, calculated as (maximum P minus minimum P) divided by mean P, is with a value of 7.5 % slightly higher in calibrated basins than in non-calibrated basins (6.1 %) (Table 5).

Global averages of T for 1971–2000 are very similar for all forcings, which is not surprising as all of them are bias corrected to (different) versions of the CRU time series. Global annual averages over the 30 years differ between the warmest (PGFv2.1) and coldest (WFD) forcing by only 0.08 °C.

SWD is the forcing variable which has large differences throughout the forcings (Fig. 2). Remarkably lower values are found for WFD (compared to GSWP3 and PGFv2.1) which is a result of the underlying reanalysis and dominantly affects Africa and Europe (Sect. 2.3.4, Table 3, Weedon et al., 2014). The concatenation approach (which is also used in the ISI-MIP project phase 2a) of WFD_WFDEI leads to a very strong increase (on average $\sim 15\,W\,m^{-2}$) starting in 1979. Homogenizing WFD eliminates this effect (WFDEI_hom, Fig. 2). Variations of global LWD are rather low (Table 3).

3.1.2 Uncertainty of simulated water balance components due to climate forcing uncertainty

Climate forcing uncertainty propagates to all water balance components simulated by WaterGAP. For the period 1971–2000, global Q varies among the five forcings by about $3400\,km^3\,yr^{-1}$ (Table 2). On the continental scale, the strongest climate-forcing-induced variation of Q occurs in Africa (Table 3). Here, some areas with high amounts of P (and Q) are in non-calibrated regions (e.g., Madagascar, see Fig. 3). Besides, the runoff coefficient (Q/P) of Africa, with a value of 0.21, is the lowest compared to all other continents, which vary between 0.34 (Oceania) and 0.47 (Europe). A low runoff coefficient leads to the translation of a small precipitation deviation (in percent of mean) to a relatively large discharge deviation, as can also be seen for Oceania (Table 3).

While calibrated basins cover 54 % of the global land area excluding Greenland and Antarctica (Fig. 1), 53–58 % of global Q flows out of calibrated basins (Table 5). Most of the Q from non-calibrated basins is simulated to occur in tropical regions, particularly in Indonesia and other parts of southeast Asia. As expected, the sum of Q from all non-calibrated basins varies more strongly among the forcing variants (18.4 %) than the sum of Q from all calibrated basins (2.8 %, Table 5). Variation of Q from non-calibrated regions is reduced to 10.5 % if the PGFv2.1 variant (the only forcing without precipitation undercatch correction) is excluded, while Q variation in the calibrated regions remains the same. If only the four homogeneous forcings (without

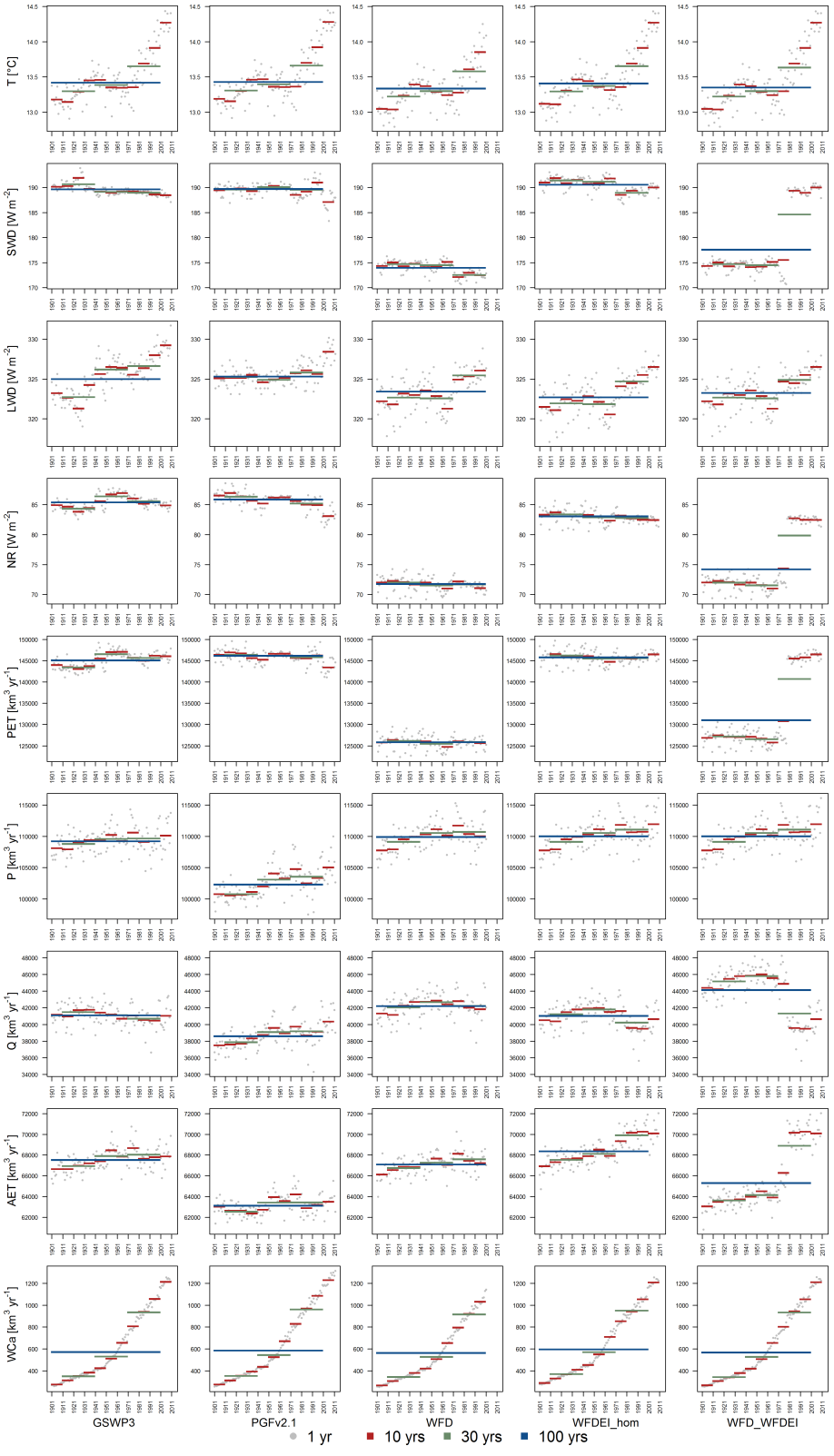


Figure 2. Global sums (means) of climatic variables and water balance components for five climate forcings (GSWP3: 1901–2010, PGFv2.1: 1901–2012, WFD: 1901–2001, WFDEI_hom: 1901–2010, WFD_WFDEI: 1901–2010) for different temporal aggregation periods of 1, 10, 30, and 100 years. Displayed are temperature (T), shortwave downward radiation (SWD), longwave downward radiation (LWD), precipitation (P), discharge into the ocean or inland sinks (Q), actual evapotranspiration (AET), and (actual) water consumption from surface water resources (which could be smaller than the demand, depending on water availability) and groundwater resources (WCa).

Table 2. Global sums of water balance components for land area (except Antarctica and Greenland) ($km^3\,yr^{-1}$) from WaterGAP (same sorting as Table 2 in Müller Schmied et al., 2014) for the five model variants and the years 1971–2000. Cells representing inland sinks were excluded but discharge into inland sinks was included.

No.	Component	GSWP3	PGFv2.1	WFD	WFDEI_hom	WFD_WFDEI
1	Precipitation P	109 631	103 525	110 690	111 050	111 050
2	Actual evapotranspiration AET[a]	68 026	63 416	67 588	69 907	68 887
3	Discharge into oceans and inland sinks Q[b]	40 678	39 173	42 200	40 213	41 298
4	Water consumption (actual) (rows 5 and 6) WCa	933	960	915	949	932
5	Net abstraction from surface water (actual)[c]	1050	1071	1023	1070	1044
6	Net abstraction from groundwater[d]	−117	−111	−108	−121	−112
7	Change of total water storage dS/dt[e]	−14	−30	−20	−25	−74
8	Long-term-averaged yearly volume balance error ($P - \mathrm{AET} - Q - \mathrm{WCa} - dS/dt$)	6	6	7	6	6

[a] AET does not include evapotranspiration caused by human water use, i.e., actual water consumption WCa. [b] Taking into account anthropogenic water use. [c] Satisfied demand from surface waters. [d] Negative values indicate that return flows from irrigation with surface water exceed groundwater abstractions. [e] Total water storage (TWS) of 31 December 2000 minus TWS of 31 December 1970, divided by 30 years.

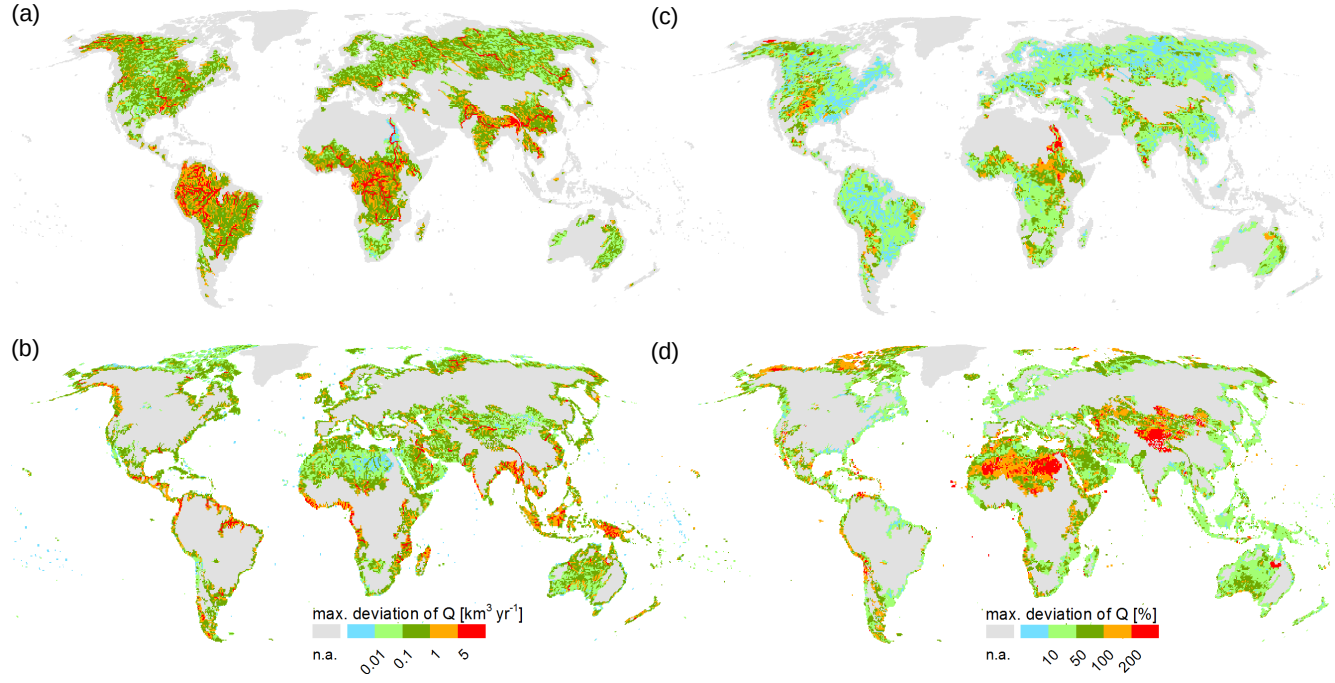


Figure 3. Spatial distribution of the maximum difference of long-term average (1971–2000) Q among the four homogeneous climate forcings (GSWP3, PGFv2.1, WFD, WFDEI_hom), expressed as absolute deviation ($km^3\,yr^{-1}$) **(a, b)** and relative deviation **(c, d)** separately for calibrated **(a, c)** and non-calibrated **(b, d)** regions. Grey areas contain either no discharge or are outside the region of interest, i.e., non-calibrated regions are grey in **(a)** and **(c)** and vice versa.

WFD_WFDEI) are considered, Q varies by 18.5 % for the non-calibrated region and by 1.6 % for the calibrated one (Table 5, row 3). Q variation in calibrated basins is due to various reasons. Calibration forces the simulated mean annual discharges in the cells with discharge stations to be equal (within 1 %) to the observed ones for the calibrated period. Outside the calibrated period, the different forcings cause the computed Q to vary. In case of the homogeneous forcings that are undercatch corrected and bias corrected against GPCC data (GSWP3, WFD, WFDEI_hom), Q differs by only 0.1 % in calibrated regions and 10.5 % in non-calibrated regions. The low value for the calibrated regions indicates neglectable influence of the different calibration periods and the smaller number of calibration stations in the case of WFD. The Q variation for the discharge produced in non-calibrated regions appears to be large in particular because all forcings are bias corrected against monthly observations of temperature from CRU, and of P from GPCC (Ta-

Table 3. Continental climate forcing variables (T, SWD, LWD, P) and water balance components (AET, Q, WCa). Ensemble mean and min/max deviation from mean (in percent, also for T) over all four homogenous forcing variants (GSWP3, PGFv2.1, WFD, WFDEI_hom), for six continental regions, and the global total for the time period 1971–2000.

		Africa	Asia	Europe*	N America	Oceania	S America	Global
T (°C)	Mean	24.1	14.6	−1.6	4.2	21.7	22.1	13.6
	Δmin	−0.2	0.0	−5.3	−3.0	−0.2	−0.1	−0.4
	Δmax	0.1	0.1	14.0	1.3	0.4	0.2	0.2
SWD (W m^{-2})	Mean	229	196	117	156	229	197	185
	Δmin	−11.1	−5.6	−9.9	−1.9	−2.4	−6.8	−6.7
	Δmax	3.9	2.9	4.5	2.3	1.2	4.4	2.5
LWD (W m^{-2})	Mean	365	322	266	285	351	381	326
	Δmin	−0.2	−0.7	−1.0	−1.2	−0.9	1.4	−0.3
	Δmax	0.4	1.0	0.9	1.2	0.6	1.5	0.3
P [km^3 yr^{-1}]	Mean	20 457	24 501	13 026	16 177	5939	28 623	108 724
	Δmin	−5.6	−3.0	−7.0	−6.9	−2.4	−4.0	−4.8
	Δmax	3.9	2.6	3.7	3.7	1.9	2.1	2.1
AET [km^3 yr^{-1}]	Mean	16 194	13 506	6942	9573	3887	17 132	67 234
	Δmin	−5.5	−4.0	−8.7	−8.0	−3.0	−5.2	−5.7
	Δmax	4.7	3.7	3.7	5.4	3.2	3.4	4.0
Q [km^3 yr^{-1}]	Mean	4183	10 276	6138	6507	2033	11 428	40 566
	Δmin	−5.9	−3.6	−5.0	−5.4	−3.4	−2.2	−3.4
	Δmax	10.9	4.4	4.0	3.8	5.0	2.5	4.0
WCa [km^3 yr^{-1}]	Mean	71	612	87	121	16	32	939
	Δmin	−5.9	−1.0	−8.5	−5.3	−2.8	−2.8	−2.6
	Δmax	2.5	1.7	3.6	5.7	2.5	2.8	2.2

* includes all of Russian Federation.

Table 4. Average density of precipitation gauging stations and P sums (km^3 yr^{-1}) for 1971–2000 of the original P data that were used for bias correction (WFD: GPCCv4, WFDEI: GPCCv5, GSWP3: GPCCv6, PGFv2.1: CRU TS3.21) and P outputs of WaterGAP using the undercatch adjusted forcings (except PGFv2.1 which is not adjusted).

Variable	Data source	Africa	Asia	Europe	N America	Oceania	S America	Global
Stations per 0.5° grid cell	CRU TS3.21	0.12	0.09	0.06	0.12	0.17	0.06	0.09
	GPCCv4	0.30	0.23	0.61	0.32	1.05	0.61	0.44
	GPCCv5	0.31	0.30	0.66	0.54	1.82	0.63	0.57
	GPCCv6	0.31	0.32	0.68	0.60	1.85	0.71	0.60
P totals (without undercatch correction)	CRU TS3.21	19 595	24 040	12 128	15 160	5958	27 611	104 492
	GPCCv4	19 745	24 062	11 858	15 073	5732	28 135	104 605
	GPCCv5	19 729	24 044	11 852	15 095	5688	28 201	104 610
	GPCCv6	19 724	24 066	11 861	15 116	5694	28 085	104 546
P totals (WaterGAP)	PGFv2.1	19 318	23 756	12 112	15 065	5799	27 475	103 525
	WFD	21 102	24 519	13 232	16 732	5960	29 146	110 690
	WFDEI_hom	21 250	24 597	13 256	16 779	5945	29 223	111 050
	GSWP3	20 160	25 133	13 505	16 131	6053	28 649	109 631

ble 4). This indicates a dissimilar spatiotemporal distribution of SWD and LWD radiation components. The larger deviation of WFD_WFDEI in calibrated regions (Table 5, row 3) can be explained by the fact that in order to deal with the offset problem in the WFD_WFDEI forcing, the WFDEI_hom calibration parameters were also used for the model variant that was driven by WFD_WFDEI (see Appendix A). Due to the much lower values of SWD in WFD as compared to

Table 5. Global sums of water balance components for land area ($km^3\ yr^{-1}$) (except Antarctica, Greenland, and inland sinks) (component numbers as in Table 2) for the model variants and the years 1971–2000, divided in calibrated and non-calibrated grid cells.

	Calibrated regions					Non-calibrated regions				
No.	GSWP3	PGFv2.1	WFD	WFDEI_hom	WFD_WFDEI	GSWP3	PGFv2.1	WFD	WFDEI_hom	WFD_WFDEI
1	66 825	63 290	68 039	68 288	68 288	42 806	40 235	42 651	42 762	42 762
2	43 996	40 112	45 232	45 482	44 903	24 031	23 303	22 356	24 425	23 984
3	22 291	22 619	22 286	22 269	22 893	18 388	16 554	19 915	17 944	18 405
4	523	546	515	531	523	411	414	400	418	410
5	582	598	572	594	581	468	473	451	476	463
6	−59	−52	−58	−62	−59	−57	−59	−50	−58	−53
7	18	15	9	9	−28	−32	−44	−29	−33	−46
8	−3	−3	−2	−3	−2	9	8	9	8	8

WFDEI, GSWP3, or PGFv2.1, Q, as computed with WFD, has the highest value of all variants in non-calibrated regions, and is 11 % larger than Q computed with WFDEI_hom for 1971–2000 (Table 5). One may conclude that GHMs without a calibration routine overestimate Q if driven by WFD; this may be one reason for the comparably high multi-model Q estimate of 42 000–66 000 $km^3\ yr^{-1}$ reported in Haddeland et al. (2011) which is much higher than previous estimates (e.g., Baumgartner and Reichel, 1975; Fekete et al., 2002) or this study.

Figure 3 shows the uncertainty range of Q at grid-cell level for calibrated and non-calibrated regions caused by the four homogeneous forcings. In both calibrated and non-calibrated regions, the highest absolute differences occur in cells with large discharge, either in the downstream part of large rivers (e.g., Nile in Fig. 3a) or in areas with high precipitation (e.g., coast of Alaska or in Papua New Guinea in Fig. 3b). The lowest relative differences in the calibrated regions occur upstream and downstream of the 1319 discharge gauging stations that were used for model calibration (Fig. 3c). The effect of calibration is also visible in non-calibrated regions downstream of a gauging station, e.g., in the Amazon downstream of Obidos (Fig. 3d). Even in most areas of the globe which are calibrated, i.e., in grid cells upstream of calibration stations, relative Q variations due to variations in climate forcings exceed 10 % (Fig. 3c). In many cells, not only in dry regions, variations exceed 50 %. In non-calibrated regions, grid cells with relative Q variations below 10 % are very rare unless they are located downstream of a calibration station (Fig. 3d). In general, relative variations of Q are often higher in non-calibrated (Fig. 3d) than in calibrated regions (Fig. 3c) mainly because dry areas are less likely to have calibration stations. However, humid Iceland, for example, also exhibits simulated Q variations of more than 50 %. When averaged over all grid cells globally (with $Q > 0$), variation of Q due to variation of the four homogeneous forcings is 55 % (1.3 $km^3\ yr^{-1}$). For calibrated regions, the variation reduces to 37 % (1.6 $km^3\ yr^{-1}$), while it increases to 74 % (1.0 $km^3\ yr^{-1}$) in non-calibrated regions. When considering net cell runoff R in all cells with positive values, i.e., the runoff added to upstream discharge within a cell, variations due to the climate forcings grow to an average of 64 % in calibrated regions and an average of 92 % in non-calibrated regions. When considering only GSWP3 and WFDEI_hom, i.e., additionally excluding the forcings without undercatch correction (PGFv2.1) and with a much lower SWD than the others (WFD), the Q (runoff) variations are reduced to 16 % (27 %) and 31 % (38 %) in calibrated and non-calibrated regions, respectively. Reduction due to excluding PGFv2.1 is larger than reduction due to excluding WFD.

Global AET is the variable with the highest relative uncertainty due to climate forcing (Table 2). As Q within the calibrated region is forced to be nearly equal for all climate data sets, different values of P (as well as T and radiation) lead to large differences in aggregated AET (with higher absolute differences than P differences, or 12.2 %). In contrast, AET differs by only 8.8 % (and lower absolute differences than the P differences) in non-calibrated regions (both numbers for all forcings, Table 5). A total of 63–67 % of AET occurs in calibrated regions (Table 5, row 2). In WFD forcing, the low global values for SWD lead to relatively low AET and higher Q (2000 $km^3\ yr^{-1}$) compared to the homogenized forcing WFDEI_hom. PGFv2.1 has the lowest global AET but the highest WCa of all five forcings (Table 2), even though WCa includes mainly evaporation of irrigation water that is driven by the same climatic variables as AET. This reflects the variations in the spatial pattern of the climatic variables among the five forcing data sets.

For the period 1971–2000, global WCa varies among the five forcings by 45 $km^3\ yr^{-1}$ (Table 2), i.e., the range is less than 5 %. A total of 56.0–56.9 % of global WCa occurs in calibrated regions (Table 5, row 4). Among all forcing variants, deviation of WCa is higher in calibrated regions (5.9 %) than in non-calibrated regions (4.4 %) (Table 5). WCa uncertainty due to climate forcings differs strongly among the continents (Table 3). For Asia, the continent with the highest water use, variation among the model variants is very low, indicating good agreement of climate forcing for the irrigation sub-model and/or averaging out differences in climate forcings over the large number of grid cells in Asia with ir-

Table 6. Global and continental estimates of WaterGAP water balance components compared to literature values ($km^3\ yr^{-1}$). WaterGAP results are analyzed for the same time span and spatial coverage as the reference and are comparable in terms of precipitation undercatch (see footnotes).

Source	Coverage	Time span	*P*		AET		*Q*	
			WaterGAP	Reference	WaterGAP	Reference	WaterGAP	Reference
Wisser et al. (2010)	Global, w/ G	1901–1925	102 110[a]	105 298	63 319[a,b]	68 274	37 974[a]	36 888
		1926–1950	102 653[a]	105 675	63 081[a,b]	67 826	38 837[a]	37 092
		1951–1975	105 444[a]	108 081	64 693[a,b]	68 550	39 914[a]	38 864
		1976–2002	104 436[a]	106 764	64 337[a,b]	69 917	39 421[a]	36 813
		1901–2002	103 676[a]	106 461	63 867[a,b]	68 480	39 044[a]	37 401
Hanasaki et al. (2010)	Global, w/ G, w/ A	1984–1999	106 012[a,c]	113 900	64 281[a,b,d]	72 080	40 876[a,e]	41 820
Rodell et al. (2015)	Global, w/o A	2000–2010	113 341[f]	114 300	71 554[b,f]	70 500	41 309[f]	43 800
	N America, w/ G		17 983[f]	17 717	10 339[b,f]	9911	6604[f]	7894
	S America		29 153[f]	29 587	17 573[b,f]	17 286	11 579[f]	12 301
	Africa		21 323[f]	20 629	17 307[b,f]	16 809	4029[f]	3820
Müller Schmied et al. (2014)	Global, w/o G, w/o A	1971–2000	111 050[g]	111 070[h]	69 819[b,g]	70 576[b,h]	41 298[g]	40 458[h]

[a] PGFv2.1. [b] Including WCa. [c] Including Antarctica (as 2.1 % of global value). [d] Including Antarctica (as 4.9 % of global value). [e] Including Antarctica (as 0.2 % of global value, all percentages based on Rodell et al., 2015). [f] WFDEI_hom. [g] WFD_WFDEI. [h] STANDARD model variant; G in column coverage: Greenland, A in column coverage: Antarctica.

rigation water use. Again, Europe and North America have high uncertainties in continental assessments due to climate forcing uncertainty/variability.

3.1.3 Comparison with other studies

Global sums of AET and *Q* for the five climate forcings used in this study are within the range of estimates reported in the literature (see values from various sources in Müller Schmied et al. (2014), their Table 5). Values for AET of this study (64 400–70 800 $km^3\ yr^{-1}$ including WCa) are well within this range. Global values for *Q* (39 200–42 200 $km^3\ yr^{-1}$) are at the upper end of values from literature (except Haddeland et al., 2011).

Table 6 shows a comparison to global and continental estimates of AET and *Q* of this study to four recent reference studies. Time span and spatial coverage of WaterGAP results is the same as in the respective references, and the climate forcing variant of WaterGAP was selected such that studies using *P* without undercatch were compared to results of WaterGAP using PGFv2.1. Wisser et al. (2010) used the WBMplus model with CRU forcing plus three different precipitation data sets for an uncertainty analysis. Even though their *P* was not undercatch corrected and also scaled to CRU observations (like PGFv2.1), global *P* of PGFv2.1 is 2.7 % lower for the time period 1901–2002 (and lower between 2.2 and 3.1 % in the different time periods analyzed). For this period, WaterGAP simulates around 7.2 % less AET and 4.2 % more *Q* compared to Wisser et al. (2010), but differences vary for the other time periods analyzed (AET: 6.0–8.7 %, *Q*: 2.6–6.6 %).

Hanasaki et al. (2010) used a climate forcing that is scaled to CRU TS 2.1 and not undercatch corrected for the time span 1985–1999. Therefore, results of WaterGAP driven by PGFv2.1 were used for the comparison in Table 6. Their values probably also included Antarctica, as they mention a land area of 144 000 km^2, so a direct comparison is not straightforward. Based on the assessment of Rodell et al. (2015) (see next paragraph), Antarctica's share in global *P*, AET, and *Q* is about 2.1, 0.2, and 4.9 %, respectively, and these percentages were added to the WaterGAP results. Surprisingly, global *P* of PGFv2.1 is 7.4 % lower than *P* of Hanasaki et al. (2010). As a consequence, AET (by 12.1 %) and *Q* (by 2.3 %) are also lower for WaterGAP forced with PGFv2.1 compared to Hanasaki et al. (2010) (Table 6).

Rodell et al. (2015) provide an optimized consistent set of global and continental water fluxes during 2000–2010 by combining satellite products and outputs from a number of models in an optimization routine that enforced multiple water and energy budget constraints simultaneously. Compared to WFDEI_hom (this study), global *P* is nearly equal (0.8 % lower). WaterGAP simulated AET slightly higher (1.5 %) but *Q* was 6.0 % lower compared to Rodell et al. (2015) (Table 6). As the definition of continents differs partly between Rodell et al. (2015) and this study, only North and South America as well Africa can be compared. PGFv2.1 continental estimates for *P* are 1.5 and 3.3 % higher for North America and Africa, and 1.5 % lower for South America, with WaterGAP AET being higher (1.6–4.1 %). Large differences occur for *Q*, where WaterGAP estimated 19.5 and 6.2 % lower values for North and South America, and 5.2 % higher values for Africa. North America and Africa are the continents which show high variations in *Q* also in this study among the forcings (Table 3).

Considering the water balance component values of Wisser et al. (2010), Hanasaki et al. (2010), and Rodell et al. (2015), there is no water balance component for which WaterGAP values are consistently too high or too low. Even

when the climate forcings used in these studies are similar to one of the climate forcings used here (e.g., regarding undercatch and bias correction), global P values differ, which in itself leads to different model output. Therefore, the approach of the many model intercomparison studies to use the same climate forcing for all models helps to assess the differences of the models themselves (Haddeland et al., 2012).

The WaterGAP 2.2 (ISIMIP2a) water balance components using WFD_WFDEI climate input (Table 2) differ from those of the STANDARD WaterGAP 2.2 model runs that was also driven by WFD_WFDEI as presented in Müller Schmied et al. (2014) (their Table 2) due to the seven model modifications listed in Appendix A. Global P is insignificantly affected by the different ocean–land mask. Global AET and Q are comparable and differ only by 1–2 % between both studies (Table 6). Due to the assumed deficit irrigation in groundwater depletion areas (Sect. 2.1), global WCa during 1971–2000 is estimated as $936\,km^3\,yr^{-1}$ as compared to $1031\,km^3\,yr^{-1}$ in STANDARD. Deficit irrigation also explains the smaller decrease of groundwater storage in this study, with an average of $75\,km^3\,yr^{-1}$ during the period 1971–2000 compared to $125\,km^3\,yr^{-1}$ in STANDARD (Müller Schmied et al. (2014), their Table 3). In the applied WaterGAP 2.2 (ISIMIP2a) version, reservoirs are filled up with water in their construction year. This leads to a net increase of reservoir storage ($53\,km^3\,yr^{-1}$) compared to a decrease of $43\,km^3\,yr^{-1}$ in STANDARD, where reservoirs are assumed to have been in operation over the entire simulation period. Thus, total water storage decreased less than in STANDARD, with $74\,km^3\,yr^{-1}$ instead of $215\,km^3\,yr^{-1}$.

3.2 Variation of estimated global water balance components across temporal aggregation and reference periods

Figure 2 shows the importance of temporal aggregation and reference periods for the assessment of global-scale climatic variables and water balance components during the time period 1901–2010 (2001 for WFD, 2012 for PGFv2.1). Even for globally aggregated components, there are strong year-to-year fluctuations. To assess (next to the visual interpretation) the importance of the choice of temporal aggregation for the different climatic variables or water balance components on their variability during the simulation period, the ranges of their global values at temporal aggregations of 1, 10, and 30 years were first computed as the difference between the maximum and the minimum value during the whole time period. Then the effect of temporal aggregation was quantified by calculating the ratio of the ranges at the different temporal aggregations. For all climate variables and water balance components, except those with a significant trend, the ranges (Fig. 2) and ratios vary strongly among the forcing variants. To achieve an approximate but robust representation of the effect of temporal aggregation on variability, we present only the median of the ratios among the four homogeneous forcings. Regarding the radiation variables SWD and LWD, their range is approximately halved when going from 30 years to 10 years or from 10 years to 1 year, and consequently reduced by a factor of 3–4 when going from 30 years to 1 year. Global P and AET range is reduced by a factor of about 2 when going from 30 years to 10 years or by a factor of 3 when going from 10 years to 1 year. Regarding global discharge, the corresponding ratios are approximately 2 and 4. Here, the variation among the four forcings is 1.6–2.7 for the reduction of variability when going from 30 years to 10 years and 3.0–5.4 when going from 10 years to 1 year. Quantifying temporal variability of global WCa, which has a significant trend (Fig. 2), the range of 1-year and 10-year aggregates is very similar, while the range is reduced by a factor of 1.6 when going from 30 years to 10 years. Considering the variability of T, the ranges during the simulation period are around 1.5 °C (1 year), 1.1 °C (10 years), and 0.4 °C (30 years).

Regarding the choice of reference period, its importance is obvious in case of the variables with a strong temporal trend like T and WCa. The increase of global averages of T during the last 3 decades is comparable among the five climate forcings as they are all bias corrected to almost the same observation-based product (CRU TS, but different versions). Large differences occur for 100-year average SWD, for which WFD forcing shows an offset of around $-15\,W\,m^{-2}$. This also affects the combined WFD_WFDEI, resulting in an implausible discontinuity from 1978 to 1979. The monthly homogenized series (WFDEI_hom) reduces this offset, but the (smaller) offset within WFD since 1973 (integration of first NOAA VTPR satellite data, Uppala et al., 2005) cannot be reduced by this method. LWD shows different variations among the climate forcings at annual, decadal, and 30-year aggregations (e.g., between GSWP3 and PGFv2.1), while the 100-year averages are relatively close to each other. Again, in WFD (and consequently WFD_WFDEI and WFDEI_hom) the usage of satellite data in the ERA-40 reanalyses from 1973 onwards leads to an offset in LWD, which is clearly visible in the 30-year averages (1971–2000) in all three forcings. Except PGFv2.1, all climate forcings indicate an increase of LWD in the last decades which fits to increasing T. Using land surface parameters and T, WaterGAP calculates the outgoing components of radiation and subsequently net radiation which is then used to calculate potential evapotranspiration. In WFD, net radiation is much lower than in the other data sets (century mean $72\,W\,m^{-2}$ compared to $83\,W\,m^{-2}$ for WFDEI_hom and $86\,W\,m^{-2}$ for GSWP3 and PGFv2.1) (Fig. 2). Considering the four homogeneous forcings only, temporal variations of net radiation are low but rather different among the forcings, and there is no significant trend, except for PGFv2.1 with a decreasing trend in the last 30 years. Global PET has an even smaller variation, and no trend during the century either. Global P seems to be slightly smaller before 1940 than afterwards but this may be due to the lower number of rain gauges available during this time period. After 1940, 30-year averages of global P are al-

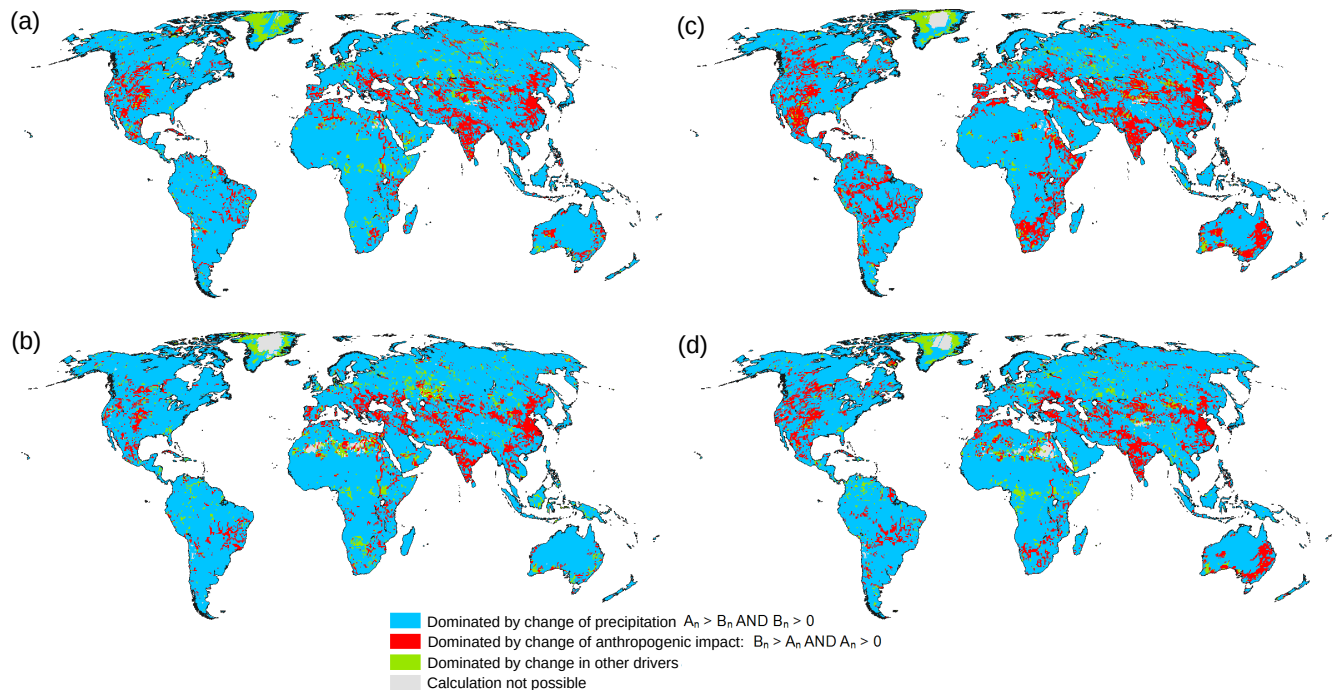


Figure 4. Relative dominance of drivers of change of long-term average Q between 1941–1970 and 1971–2000 (Sect. 2.4.2). Blue indicates that change in P is more dominant than change in anthropogenic impact due to water abstraction and dam construction, red indicates the opposite. In green areas, other drivers are dominant. In grey areas, a calculation is not possible as the denominator of indicators A_n and B_n is zero (no change in long-term average Q). Results are shown for WaterGAP as driven by the meteorological forcings GSWP3 **(a)**, PGFv2.1 **(b)**, WFDEI_hom **(c)**, and WFD **(d)**.

most constant in time. This is supported by Beck et al. (2005) who found no significant trend in global P for 1950–2000 when utilizing observations from the same set of rain gauges over the whole analysis period.

Neither can trends of global AET or Q be detected. The decadal or 30-year variations vary strongly among the forcings. For Q (AET), the inhomogeneity in WFD_WFDEI leads to an implausible decrease (increase) of around 5000 $km^3\ yr^{-1}$. Among the homogeneous forcings, WFDEI_hom shows low Q (high AET) during the last 3 decades as compared to the previous decades and as compared to the other forcings, even though PET of all those forcings does not show a trend. This might be related to differences in spatial patterns among the forcings. The results of this study confirm the finding of the IPCC Fifth Assessment Report that "the most recent and most comprehensive analyses of river runoff do not support the IPCC Fourth Assessment Report (AR4) conclusion that global runoff has increased during the 20th century" (Stocker et al, 2013, p. 44). Century means of global Q from GSWP3 and WFDEI are very similar (like their P and PET values), while Q is smaller in case of PGFv2.1 due to lower P (compare Sect. 3.1) and higher in WFD due to lower SWD (and thus PET).

WCa is the only water balance component with a strong temporal trend (strong increase since the 1950s) and only a small variation of annual values around the trend that is mainly caused by expansion of irrigated land. Interannual variability is due to climate variability affecting irrigation water use. Temporal aggregation over a decade appears to be appropriate to clearly show the trend. The separation of total water use into the different sectors as well as into water withdrawals and consumptive use is presented by Müller Schmied et al. (2016, their Fig. 3).

When comparing the output of different GHMs, the climate forcing used as model input is a very strong determinant of model output (see Sect. 3.1). When GHMs driven by (more or less) the same climate forcing are compared (see comparison of WaterGAP to Wisser et al. (2010), in Table 6), the choice of reference period matters. Differences for global P, AET, and Q among the four roughly 25-year time periods are 3.2, 2.5, and 4.9 %, respectively, for WaterGAP in this study, and 2.6, 3.0, and 5.5 % for Wisser et al. (2010).

3.3 Dominant drivers of temporal variations of 30-year mean annual river discharge: precipitation or human water use and dam construction

Figure 4 shows where the change of long-term average Q between the time period 1941–1970 and the time period 1971–2000 is either caused mainly by the change of P in the upstream river basin (blue colors) or by the change of the anthropogenic impact on Q by human water use and dam con-

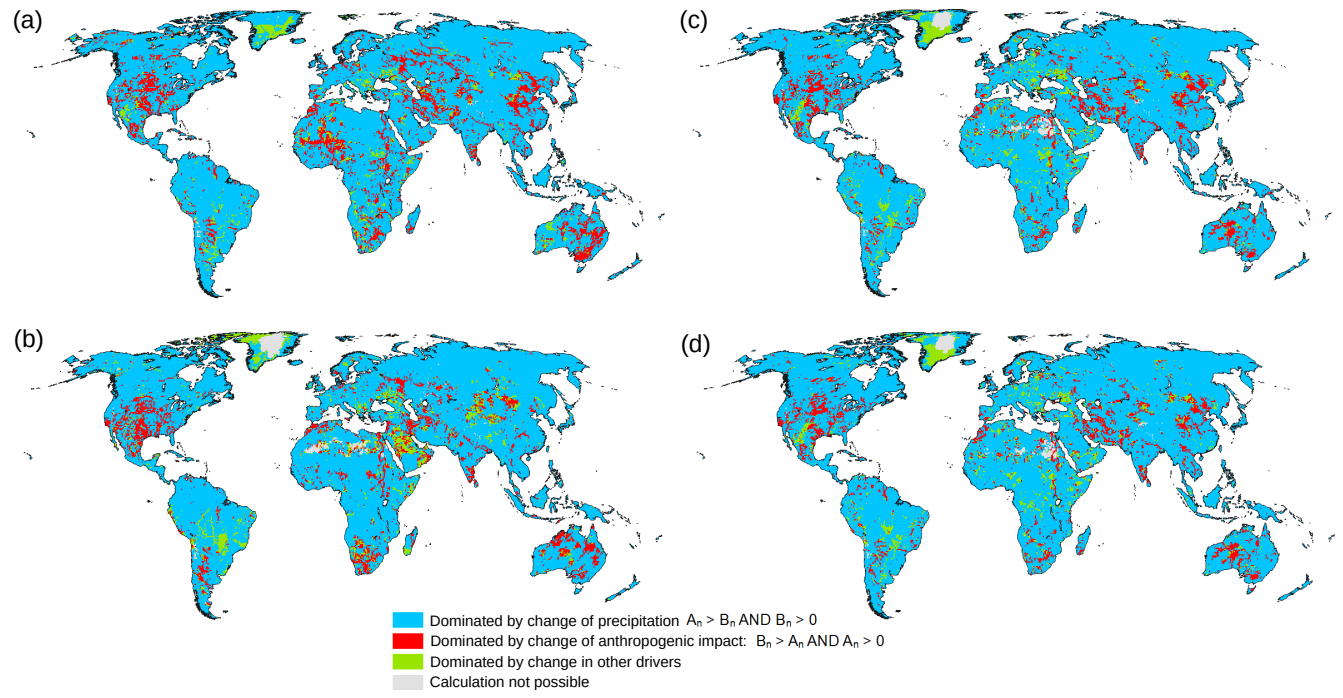


Figure 5. Relative dominance of drivers of change of long-term average Q between 1911–1940 and 1941–1970 (Sect. 2.4.2). Blue indicates that change in P is more dominant than change in anthropogenic impact due to water abstraction and dam construction, red indicates the opposite. In green areas, other drivers are dominant. In grey areas, a calculation is not possible as the denominator of indicators A_n and B_n is zero (no change in long-term average Q). Results are shown for WaterGAP as driven by the meteorological forcings GSWP3 **(a)**, PGFv2.1 **(b)**, WFDEI_hom **(c)**, and WFD **(d)**.

struction (red colors, see Sect. 2.4.2). Results for WaterGAP as driven by each of the four homogeneous climate forcings GSWP3, PGFv2.1, WFDEI_hom, and WFD are shown. In most regions, change in P is the more important driver of change in Q than change in the anthropogenic impact. It is areas with high water consumption or/and the construction of dams where change in anthropogenic impact is more important than change in P for explaining temporal Q changes. Note that the developed indicators only compare the relevance of two drivers of change. Even in the blue and red grid cells, other variables such as T or radiation may be even stronger drivers of the simulated change in Q. In grid cells where indicators A_n and B_n are both negative or zero (green colors), however, other drivers (and not P or anthropogenic effects) are certainly the main reason for changes in Q.

Changes in long-term average Q between the time periods 1911–1940 ($t1$) and 1941–1970 ($t2$) are, in most world regions, less dominated by changes in the anthropogenic impact on river discharge (Fig. 5). Anthropogenic impact increases in the time periods 1941–1970 and 1971–2000 ($t3$), which is consistent with the acceleration of human water use (Fig. 2) and dam construction throughout the 20th century. In the earlier analysis period, anthropogenic activity dominates Q change only in small parts of North America and Asia (around the North China Plain and inflows to the Caspian Sea). It is only in the later analysis period that anthropogenic impact dominates over P impact in India, southeast China, Spain, and Turkey (compare Figs. 4 and 5). Taking India and GSWP3 forcing as an example, P increases for both time steps ($t1$–$t2$: $+49\,\mathrm{km^3\,yr^{-1}}$ and $t2$–$t3$: $+30\,\mathrm{km^3\,yr^{-1}}$). However, Q ($+9\,\mathrm{km^3\,yr^{-1}}$) and WCa ($+26\,\mathrm{km^3\,yr^{-1}}$) increases between $t1$ and $t2$, while between $t2$ and $t3$ Q ($-39\,\mathrm{km^3\,yr^{-1}}$) decreases and WCa ($+81\,\mathrm{km^3\,yr^{-1}}$) increases more strongly than between $t1$ and $t2$. In India, the intensified water use and changed signs between P and Q lead to the indication that anthropogenic effects dominate the change in Q (compare Figs. 4 and 5).

Human water use and dam construction are the dominant drivers for changes in long-term Q averages on 9–13 % of the land area for the time periods 1911–1940 and 1941–1970, and increases to 11–18 % of the land area for the time periods 1941–1970 and 1971–2000. The fraction with P domination decreases, from 82–84 to 77–82 %. At the same time, the area for which the indicators A_n and B_n cannot be calculated (due to similar long-term Q averages and thus zero in the denominators of Eqs. 3 and 4) is rather constant (1.1 to 0.9 %). The land fractions where neither driver dominates decreases slightly from 6 to 5 %. Figures B1 and B2 in Appendix B shows A_n and Figs. B3 and B4 shows B_n.

The four climate forcings affect the spatial pattern of dominance. They lead to different changes of P and different changes of human water use as the globally dominant irri-

gation water use is computed as a function of climate. For example, with the forcings based on ECMWF reanalyses (WFDEI_hom, WFD, Fig. 4c and d), large parts in southeast Australia are driven by anthropogenic effects, whereas for the forcings based on NCEP reanalyses this is not the case (PGFv2.1, Fig. 4b) or is applicable to a lesser extent (GSWP3, Fig. 4a). For WFDEI_hom, the anthropogenic dominance is considerably higher in Mexico (Fig. 4). If using PGFv2.1 forcing, the area around the North China Plain is dominated by P changes, whereas in the other forcings it is dominated by anthropogenic effects (Fig. 5). Even if mean global values, e.g., for P and Q, compare well (Fig. 2, Table 2), regional differences in the climate forcings (and underlying reanalysis) result in these different spatial patterns of GHM output.

The effects of human water use and dam construction on Q variations cannot be separated by the applied indicator approach. While dam construction leading to new reservoirs decreases long-term average Q (e.g., due to additional evaporation), human water consumption is expected to be more important in most grid cells (see also Döll et al., 2009).

4 Conclusions

This study presents a model-based assessment of water balance components considering different temporal (year to century) and spatial (0.5° grid cell to global) aggregations. The GHM WaterGAP 2.2 (ISIMIP2a) was forced with an ensemble of four (plus one homogenized) state-of-the-art climate forcings with daily data. These forcings differ by the underlying reanalyses, the observational data sets used for bias correction, and whether precipitation observations were corrected for undercatch. At global scale and for 1971–2000, P differs among the forcing by 7500 $km^3\,yr^{-1}$ and Q about 3000 $km^3\,yr^{-1}$. Estimated Q differs most among climate forcings where WaterGAP cannot be calibrated due to a lack of river discharge observations in the GRDC database, in particular in southeast Asia (Indonesia and Papua New Guinea). Variations among the four homogeneous forcings (GSWP3, PGFv2.1, WFD, WFDEI_hom) result, for 1971–2000, in a variation of long-term average Q aggregated over all non-calibrated areas of 18.5 % but only in a variation of 1.6 % for the calibrated areas. This supports the many calls for extending (or maintaining) in situ Q observations (e.g., Fekete et al., 2015) and for sharing the already available Q data (e.g., Hannah et al., 2011). Certainly, satellite observations have the potential to support river discharge estimation (Tang et al., 2009). The Surface Water and Ocean Topography (SWOT) mission, for example, proposes discharge observations for river widths > 50 m but all remote sensing methods for deriving Q strongly rely on in situ measurements (Pavelsky et al., 2014).

On continental scale, most differences for P and Q among the homogeneous forcings (GSWP3, PGFv2.1, WFD, WFDEI_hom) occur in Africa and, due to snow undercatch of rain gauges, also in the data-rich continents Europe and North America. Variations of Q at the grid-cell scale due to uncertainty in meteorological data are large, except in a few grid cells upstream and downstream of calibration stations, with on average 37 and 74 % variation among the four homogeneous forcings. These large forcing-induced uncertainties are disturbing because the actual forcing data set uncertainty may not fully be represented by the ensemble and uncertainty due to the choice of hydrological model and its parameters is neglected.

The study underlined that the level of temporal aggregation of water balance components is of importance, such that for comparison purposes, the same temporal aggregation and identical reference periods should be used. However, for all variables except T and WCa, due to the uncertainty of climate data, the choice of the climate forcing affects climate variables and water balance components computed by GHMs more strongly than the choice of reference period. For global variables that (until now) showed no significant trend (like P and Q), the widely used 30-year aggregation period is suitable for comparison purposes, while for variables showing a strong trend, i.e., T and WCa, decadal aggregation is recommended. Ranges of climate forcing variables and water balance components are reduced roughly by a factor of 2 when going from 30 years to 10 years (and 10 years to 1 year) and consequently by a factor of 3–4 when going from a 30-year to a 1-year assessment.

Homogenization of climate forcing is required when concatenating time series of meteorological variables from different sources, as in the case of WFD and WFDEI (which are based on two different reanalyses), are combined to cover the time period since 1901 until recent times. Even within the homogenized WFDEI_hom climate forcing there remains an offset in SWD and LWD data in 1973 that stems from the ERA-40 reanalysis; therefore, it is recommended to start analysis if possible only after 1978 when ERA-Interim data are available. Regardless, none of the four homogeneous climate forcings appears to be suitable for trend analyses as they are all bias corrected against gridded monthly data derived from observations of precipitation and temperature where the number of observation stations varies over time.

Humans affect the global water cycle increasingly. When comparing global sums of human water consumption to river discharge into oceans and internal sinks (or to renewable water resources), human impact seems to be small (Table 2). However, on 9–18 % of global land area, human water consumption and dam construction were more important drivers of change in river discharge in the 20th century than precipitation (Figs. 4 and 5). In this study, however, only the impact on long-term averaged discharge was analyzed, while possible seasonal impacts, e.g., due to reservoir operation, were not considered (Adam et al., 2007; Döll et al., 2009).

For future water resources modeling studies (see also Döll et al., 2016), the impact of the uncertainty of meteoro-

logical variables should be considered by applying various (equally) plausible climate forcings. Using more than one GHM may add additional robustness. Such model intercomparison projects are currently on the way (e.g., ISIMIP2a, eartH2Observe (http://www.earth2observe.eu/), The Agricultural Model Intercomparison and Improvement Project AgMIP (http://www.agmip.org/), Land Surface, Snow and Soil Moisture Model Intercomparison Project LS3MIP (http://www.climate-cryosphere.org/activities/targeted/ls3mip) or already finished (e.g., WATCH model intercomparison; Haddeland et al., 2011, ISI-MIP Fast Track; Schewe et al., 2014). They may improve the quantification of the world's water resources and guide investigation of various sources of uncertainty. Development of an improved method for correcting the global state-of-the-art precipitation products, by building on the work of Fuchs et al. (2001), would enable a better quantification of global precipitation.

5 Data availability

The WaterGAP output will become freely available for the public within the framework of the ISI-MIP project phase 2a but it is not yet known where the data will be hosted (please check https://www.isimip.org/outputdata/ for updates). The homogenized climate forcing WFDEI_hom is not included within the ISIMIP2a project phase. All model outputs used in this study are available on request from the corresponding author.

Appendix A: Modification of WaterGAP 2.2 (ISIMIP2a) compared to WaterGAP 2.2

- A new land cover input based on MODIS data from the year 2004 (using the dominant land cover class per 0.5° cell instead of the land cover class at the grid center).
- Updated lake and wetland inputs based on the Global Lakes and Wetlands Database (GLWD) (Lehner and Döll, 2004) and the Global Reservoir and Dam database (GRanD) version 1.01 (Lehner et al., 2011) as well as information on operation years from available electronic resources.
- Different ocean–land mask: while WaterGAP 2.2 uses the ocean–land mask from the IMAGE model (Alcamo et al., 1998), being the standard for WaterGAP development and covering 66 896 grid cells, here the WATCH-CRU ocean–land mask with 67 420 grid cells is used. The main differences occur in coastal areas (for which static attributes, such as soil moisture capacity, of the standard land mask are transferred to the new neighboring cell, while some other coastal cells disappeared), and due to the inclusion of many more islands in the Pacific Ocean (that obtained attributes values from nearest grid cells).
- Deficit irrigation based on Döll et al. (2014), with only 70 % of irrigation water demand in grid cells which have a groundwater depletion of at least 5 mm yr^{-1} during 1980–2009 and where the fraction of water withdrawals for irrigation is larger than 5 % of total water withdrawals for the same time period.
- Man-made reservoirs are no longer assumed to exist over the whole simulation period but only from the year of their construction onward. This includes also regulation of the outflow of natural lakes by dams.
- For lakes, reduction of evaporation due to decreasing lake area is calculated according to Eq. (1) in Hunger and Döll (2008), resulting in a lower but more realistic lake area and thus evaporation reduction with decreasing lake storage.
- For WaterGAP calibration, we used observed streamflow data from 30 years. For GSWP3, PGFv2.1, and WFD, we used data from 1971 to 2000 if available for the time period. Due to the offset in radiation of WFD_WFDEI forcing (and consequences for model results, see Müller Schmied et al., 2014), we calibrated WFDEI_hom using preferably the period 1980–2009 and used these calibration parameters for the WFD_WFDEI simulation.

Appendix B: Indicators A_n and B_n

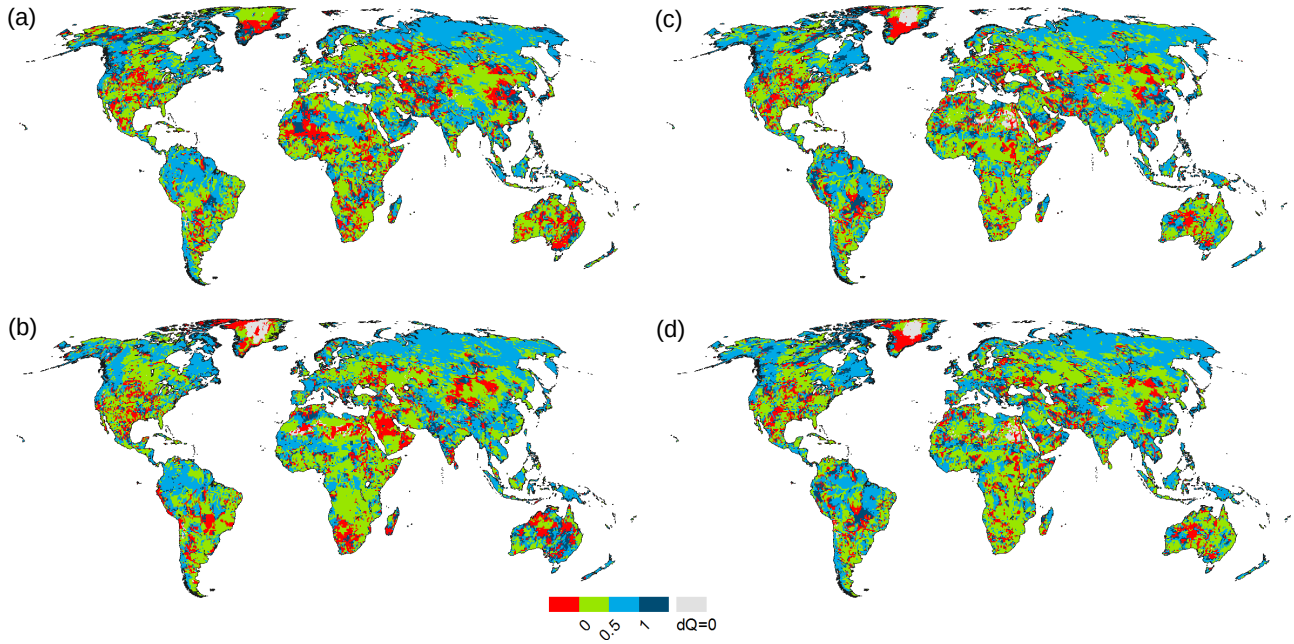


Figure B1. Indicator A_n representing dominance of the change of P for the change in Q (Eq. 3) from 1911–1940 to 1941–1970. Grey color indicates that the change in Q is zero, such that A_n cannot be computed. Red color indicate areas where A_n is negative, i.e., change in P had the opposite sign of the change in Q; therefore, P was not the dominant driver for change in Q. Results are shown for WaterGAP as driven by the climate forcings GSWP3 **(a)**, PGFv2.1 **(b)**, WFDEI_hom **(c)**, and WFD **(d)**.

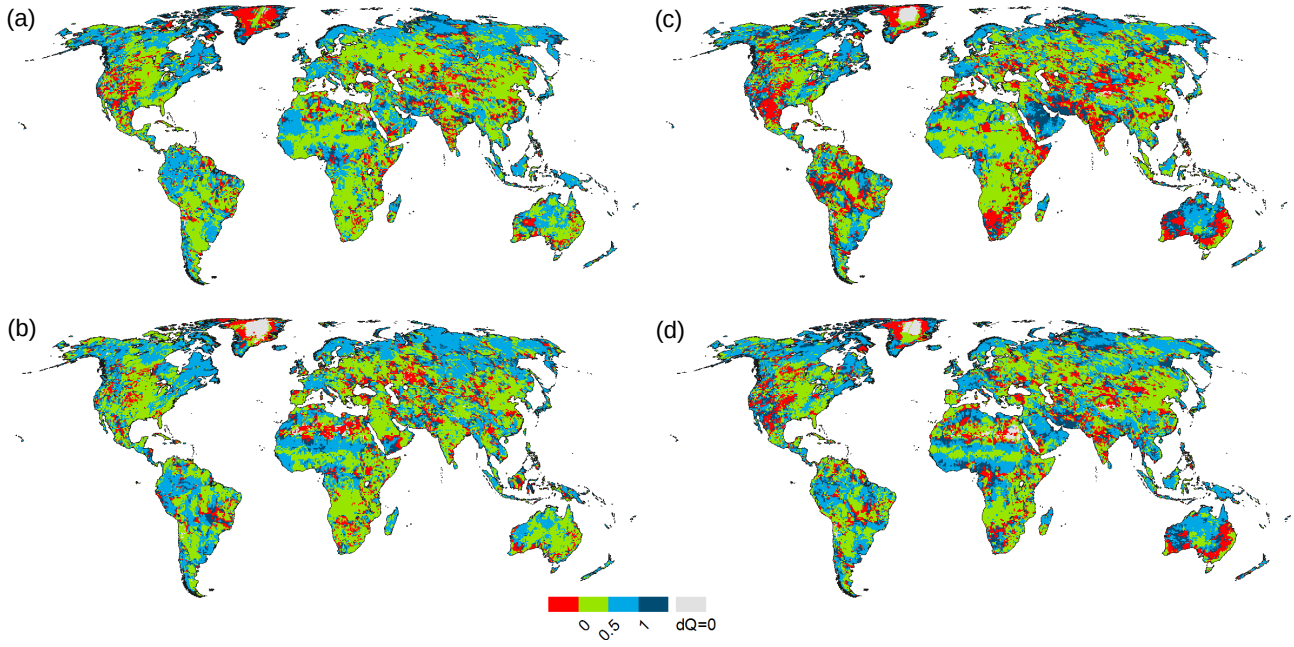


Figure B2. Indicator A_n representing dominance of the change of P for the change in Q (Eq. 3) from 1941–1970 to 1971–2000. Grey color indicates that the change in Q is zero, such that A_n cannot be computed. Red color indicate areas where A_n is negative, i.e., change in P had the opposite sign of the change in Q; therefore, P was not the dominant driver for change in Q. Results are shown for WaterGAP as driven by the climate forcings GSWP3 **(a)**, PGFv2.1 **(b)**, WFDEI_hom **(c)**, and WFD **(d)**.

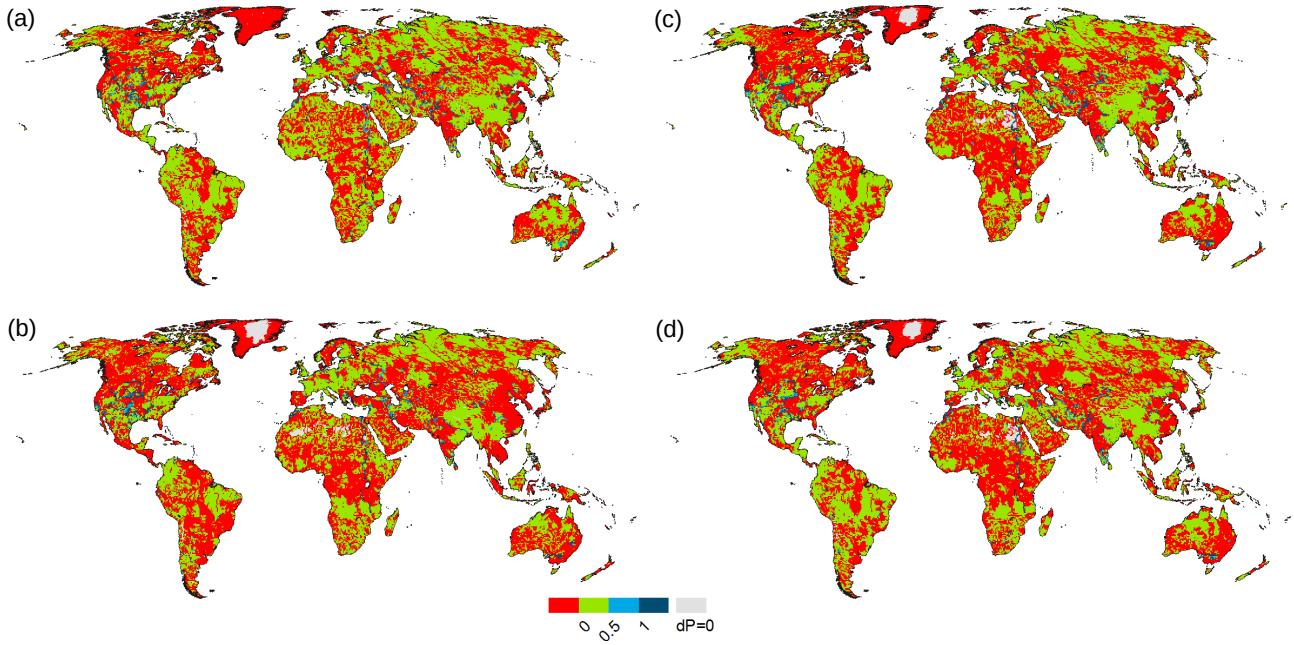


Figure B3. Indicator B_n quantifying the relative dominance of anthropogenic impact on Q change (i.e., $Q - Qnat$) as compared to the change in Q (Eq. 5) from 1911–1940 to 1941–1970. Grey color indicates that the change in Q is zero, such that B_n cannot be computed. Red color indicates areas where B_n is less than 0, and the change in anthropogenic impact is not consistent with the change in Q; therefore, the anthropogenic impact is not the dominant driver for change in Q. Results are shown for WaterGAP as driven by the climate forcings GSWP3 **(a)**, PGFv2.1 **(b)**, WFDEI_hom **(c)**, and WFD **(d)**.

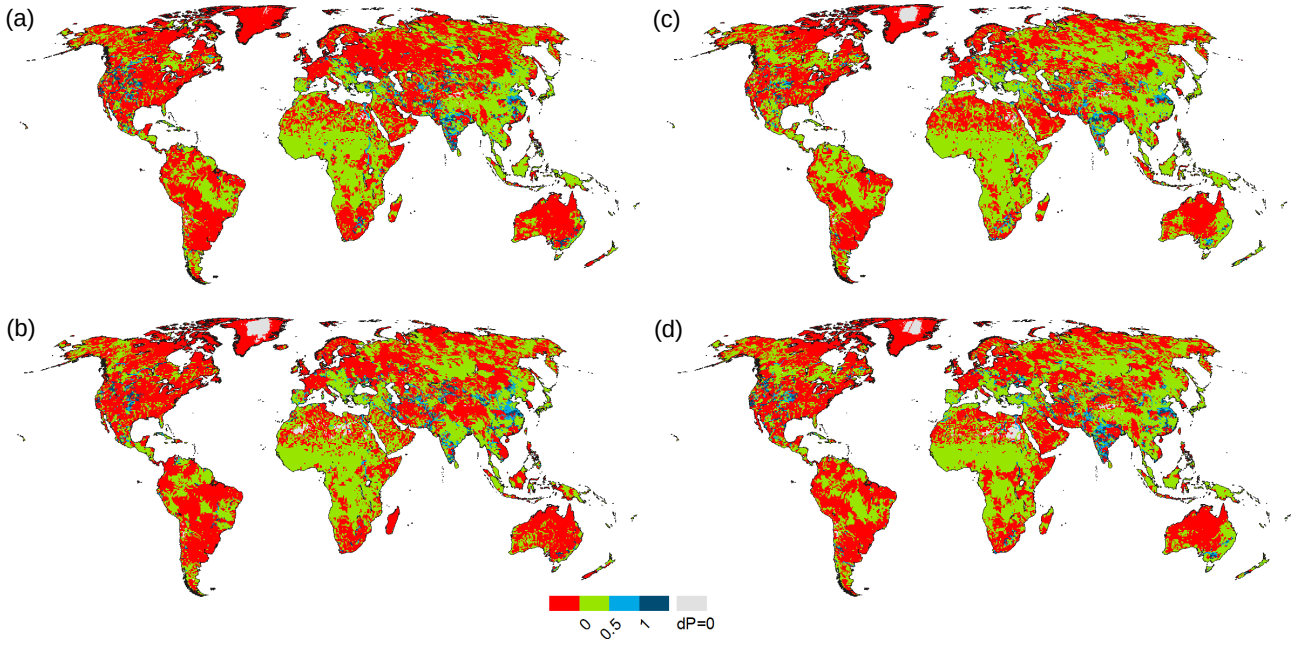


Figure B4. Indicator B_n quantifying the relative dominance of anthropogenic impact on Q change (i.e., $Q - Qnat$) as compared to the change in Q (Eq. 5) from 1941–1970 to 1971–2000. Grey color indicates that the change in Q is zero, such that B_n cannot be computed. Red color indicates areas where B_n is less than 0, and the change in anthropogenic impact is not consistent with the change in Q; therefore, the anthropogenic impact is not the dominant driver for change in Q. Results are shown for WaterGAP as driven by the climate forcings GSWP3 **(a)**, PGFv2.1 **(b)**, WFDEI_hom **(c)**, and WFD **(d)**.

Acknowledgements. The authors thank the Global Runoff Data Centre (GRDC, http://grdc.bafg.de), in Koblenz, Germany, for providing the discharge data used in this study. We are also grateful to the ISIMIP coordination team as well as the leaders of the ISIMIP global water sector (Simon Gosling and Rutger Dankers) for providing the climate forcings and for their support. Furthermore, we thank Wolfgang Grabs for organizing the international conference "Water Resources Assessment & Seasonal Prediction" (13–16 October 2015 in Koblenz, Germany) where some content of this paper was presented. Finally, we thank three anonymous reviewers for their valuable suggestions that improved the manuscript significantly.

Edited by: Ross Woods
Reviewed by: three anonymous referees

References

Adam, J. C. and Lettenmaier, D. P.: Adjustment of global gridded precipitation for systematic bias, J. Geophys. Res., 108, 4257, doi:10.1029/2002JD002499, 2003.

Adam, J. C., Haddeland, I., Su, F., and Lettenmaier, D. P.: Simulation of reservoir influences on annual and seasonal streamflow changes for the Lena, Yenisei, and Ob' rivers, J. Geophys. Res., 112, 1–22, doi:10.1029/2007JD008525, 2007.

Alcamo, J., Leemans, R., and Kreileman, E., Eds.: Global change scenarios of the 21st century – Results from the IMAGE 2.1 model, Pergamon, Oxford, 1998.

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

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

Beck, C., Grieser, J., and Rudolf, B.: A new monthly precipitation climatology for the global land areas for the period 1951 to 2000, Climate Status Report KSB 2004, German Weather Service, Offenbach, Germany, 181–190, 2005.

Biemans, H., Hutjes, R. W. A., Kabat, P., Strengers, B. J., Gerten, D., and Rost, S.: Effects of precipitation uncertainty on discharge calculations for main river basins, J. Hydrometeorol., 10, 1011–1025, doi:10.1175/2008JHM1067.1, 2009.

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

Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Matsui, N., Allan, R. J., Yin, X., Gleason, B. E., Vose, R. S., Rutledge, G., Bessemoulin, P., Brönnimann, S., Brunet, M., Crouthamel, R. I., Grant, a. N., Groisman, P. Y., Jones, P. D., Kruk, M. C., Kruger, a. C., Marshall, G. J., Maugeri, M., Mok, H. Y., Nordli, Ø., Ross, T. F., Trigo, R. M., Wang, X. L., Woodruff, S. D., and Worley, S. J.: The Twentieth Century Reanalysis Project, Q. J. Roy. Meteorol. Soc., 137, 1–28, doi:10.1002/qj.776, 2011.

Coxon, G., Freer, J., Westerberg, I. K., Wagener, T., Woods, R., and Smith, P. J.: A novel framework for discharge uncertainty quantification applied to 500 UK gauging stations, Water Resour. Res., 51, 5531–5546, doi:10.1002/2014WR016532, 2015.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Q. J. Roy. Meteorol. Soc., 137, 553–597, doi:10.1002/qj.828, 2011.

Döll, P. and Lehner, B.: Validation of a new global 30-min drainage direction map, J. Hydrol., 258, 214–231, doi:10.1016/S0022-1694(01)00565-0, 2002.

Döll, P. and Siebert, S.: Global modeling of irrigation water requirements, Water Resour. Res., 38, 8-1–8-10, doi:10.1029/2001WR000355, 2002.

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

Döll, P., Fiedler, K., and Zhang, J.: Global-scale analysis of river flow alterations due to water withdrawals and reservoirs, Hydrol. Earth Syst. Sci., 13, 2413–2432, doi:10.5194/hess-13-2413-2009, 2009.

Döll, P., Hoffmann-Dobrev, H., Portmann, F. T., Siebert, S., Eicker, A., Rodell, M., Strassberg, G., and Scanlon, B. R.: Impact of water withdrawals from groundwater and surface water on continental water storage variations, J. Geodyn., 59–60, 143–156, doi:10.1016/j.jog.2011.05.001, 2012.

Döll, P., Müller Schmied, H., Schuh, C., Portmann, F. T., and Eicker, A.: Global-scale assessment of groundwater depletion and related groundwater abstractions: Combining hydrological modeling with information from well observations and GRACE satellites, Water Resour. Res., 50, 5698–5720, 2014.

Döll, P., Douville, H., Güntner, A., Müller Schmied, H., and Wada, Y.: Modelling freshwater resources at the global scale: Challenges and prospects, Surv. Geophys., 37, 195–221, doi:10.1007/s10712-015-9347-x, 2016.

Fekete, B. M., Vörösmarty, C. J., and Grabs, W.: High-resolution fields of global runoff combining observed river discharge and simulated water balances, Global Biogeochem. Cy., 16, 15-1–15-10, doi:10.1029/1999GB001254, 2002.

Fekete, B. M., Robarts, R. D., Kumagai, M., Nachtnebel, H.-P., Odada, E., and Zhulidov, A. V.: Time for in situ renaissance, Science, 349, 685–686, doi:10.1126/science.aac7358, 2015.

Flörke, M., Kynast, E., Bärlund, I., Eisner, S., Wimmer, F., and Alcamo, J.: Domestic and industrial water uses of the past 60 years as a mirror of socio-economic development: A global simulation study, Global Environ. Change, 23, 144–156, doi:10.1016/j.gloenvcha.2012.10.018, 2013.

Fuchs, T., Rapp, J., Rubel, F., and Rudolf, B.: Correction of synoptic precipitation observations due to systematic measuring errors with special regard to precipitation phases, Phys. Chem. Earth B, 26, 689–693, doi:10.1016/S1464-1909(01)00070-3, 2001.

Haddeland, I., Clark, D. B., Franssen, W., Ludwig, F., Voß, F., Arnell, N. W., Bertrand, N., Best, M., Folwell, S., Gerten, D., Gomes, S., Gosling, S. N., Hagemann, S., Hanasaki, N., Harding, R., Heinke, J., Kabat, P., Koirala, S., Oki, T., Polcher, J., Stacke,

T., Viterbo, P., Weedon, G. P., and Yeh., P.: Multi-model estimate of the global terrestrial water balance: Setup and first results, J. Hydrometeorol., 12, 869–884, doi:10.1175/2011JHM1324.1, 2011.

Haddeland, I., Heinke, J., Voß, F., Eisner, S., Chen, C., Hagemann, S., and Ludwig, F.: Effects of climate model radiation, humidity and wind estimates on hydrological simulations, Hydrol. Earth Syst. Sci., 16, 305–318, doi:10.5194/hess-16-305-2012, 2012.

Hanasaki, N., Inuzuka, T., Kanae, S., and Oki, T.: An estimation of global virtual water flow and sources of water withdrawal for major crops and livestock products using a global hydrological model, J. Hydrol., 384, 232–244, doi:10.1016/j.jhydrol.2009.09.028, 2010.

Hannah, D. M., Demuth, S., van Lanen, H. A. J., Looser, U., Prudhomme, C., Rees, G., Stahl, K., and Tallaksen, L. M.: Large-scale river flow archives: Importance, current status and future needs, Hydrol. Process., 25, 1191–1200, doi:10.1002/hyp.7794, 2011.

Hirabayashi, Y., Kanae, S., Motoya, K., Masuda, K., and Döll, P.: A 59-year (1948–2006) global meteorological forcing data set for land surface models. Part II: Global snowfall estimation, Hydrol. Res. Lett., 2, 65–69, doi:10.3178/hrl.2.65, 2008.

Hunger, M. and Döll, P.: Value of river discharge data for global-scale hydrological modeling, Hydrol. Earth Syst. Sci., 12, 841–861, doi:10.5194/hess-12-841-2008, 2008.

Lehner, B. and Döll, P.: Development and validation of a global database of lakes, reservoirs and wetlands, J. Hydrol., 296, 1–22, 2004.

Lehner, B., Liermann, C. R., Revenga, C., Vörösmarty, C., Fekete, B., Crouzet, P., Döll, P., Endejan, M., Frenken, K., Magome, J., Nilsson, C., Robertson, J. C., Rödel, R., Sindorf, N., and Wisser, D.: High-resolution mapping of the world's reservoirs and dams for sustainable river-flow management, Front. Ecol. Environ., 9, 494–502, doi:10.1890/100125, 2011.

McMillan, H., Krueger, T., and Freer, J.: Benchmarking observational uncertainties for hydrology: rainfall, river discharge and water quality, Hydrol. Process., 26, 4078–4111, doi:10.1002/hyp.9384, 2012.

Müller Schmied, H., Eisner, S., Franz, D., Wattenbach, M., Portmann, F. T., Flörke, M., and Döll, P.: Sensitivity of simulated global-scale freshwater fluxes and storages to input data, hydrological model structure, human water use and calibration, Hydrol. Earth Syst. Sci., 18, 3511–3538, doi:10.5194/hess-18-3511-2014, 2014.

Müller Schmied, H., Adam, L., Eisner, S., Fink, G., Flörke, M., Kim, H., Oki, T., Portmann, F. T., Reinecke, R., Riedel, C., Song, Q., Zhang, J., and Döll, P.: Impact of climate forcing uncertainty and human water use on global and continental water balance components, P. Assoc. Hydrol. Sci., accepted, 2016.

Oki, T. and Kanae, S.: Global hydrological cycles and world water resources, Science, 313, 1068–1072, doi:10.1126/science.1128845, 2006.

Pavelsky, T. M., Durand, M. T., Andreadis, K. M., Beighley, R. E., Paiva, R. C. D., Allen, G. H., and Miller, Z. F.: Assessing the potential global extent of SWOT river discharge observations, J. Hydrol., 519, 1516–1525, doi:10.1016/j.jhydrol.2014.08.044, 2014.

Rodell, M., Beaudoing, H. K., L'Ecuyer, T. S., Olson, W. S., Famiglietti, J. S., Houser, P. R., Adler, R., Bosilovich, M. G., Clayson, C. A., Chambers, D., Clark, E., Fetzer, E. J., Gao, X., Gu, G., Hilburn, K., Huffman, G. J., Lettenmaier, D. P., Liu, W. T., Robertson, F. R., Schlosser, C. A., Sheffield, J., and Wood, E. F.: The observed state of the water cycle in the early 21st century, J. Climate, 28, 8289-8928, doi:10.1175/JCLI-D-14-00555.1, 2015.

Schewe, J., Heinke, J., Gerten, D., Haddeland, I., Arnell, N. W., Clark, D. B., Dankers, R., Eisner, S., Fekete, B. M., Colón-González, F. J., Gosling, S. N., Kim, H., Liu, X., Masaki, Y., Portmann, F. T., Satoh, Y., Stacke, T., Tang, Q., Wada, Y., Wisser, D., Albrecht, T., Frieler, K., Piontek, F., Warszawski, L., and Kabat, P.: Multimodel assessment of water scarcity under climate change, P. Natl. Acad. Sci. USA, 111, 3245–3250, doi:10.1073/pnas.1222460110, 2014.

Schneider, U., Becker, A., Finger, P., Meyer-Christoffer, A., Rudolf, B., and Ziese, M.: GPCC Full Data Reanalysis Version 7.0 at 0.5°: Monthly land-surface precipitation from rain-gauges built on GTS-based and historic data, Global Precipitation Climatology Centre (GPCC, http://gpcc.dwd.de/, last access: July 2016), Deutscher Wetterdienst, Offenbach/Main, doi:10.5676/DWD_GPCC/FD_M_V7_050, 2015.

Sheffield, J., Goteti, G., and Wood, E. F.: Development of a 50-year high-resolution global dataset of meteorological forcings for land surface modeling, J. Climate, 19, 3088–3111, doi:10.1175/JCLI3790.1, 2006.

Siebert, S., Kummu, M., Porkka, M., Döll, P., Ramankutty, N., and Scanlon, B. R.: A global data set of the extent of irrigated land from 1900 to 2005, Hydrol. Earth Syst. Sci., 19, 1521–1545, doi:10.5194/hess-19-1521-2015, 2015.

Sood, A. and Smakhtin, V.: Global hydrological models: a review, Hydrolog. Sci. J., 60, 549–565, doi:10.1080/02626667.2014.950580, 2015.

Stocker, T. F., Qin, D., Plattner, G.-K., Alexander, L. V., Allen, S. K., Bindoff, N. L., Bréon, F.-M., Church, J. A., Cubasch, U., Emori, S., Forster, P., Friedlingstein, P., Gillett, N., Gregory, J. M., Hartmann, D. L., Jansen, E., Kirtman, B., Knutti, R., Krishna Kumar, K., Lemke, Pl, Marotzke, J., Masson-Delmotte, V., Meehl, G. A., Mokhov, I. I., Piao, S., Ramaswamy, V., Randall, D., Rhein, M., Rojas, M., Sabine, C., Shindell, D., Talley, L. D., Vaughan, D. G., and Xie, S.-P.: Technical Summary, in: Climate Change 2013: The Physical Science Basis, Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P. M., Cambridge University Press, Cambridge, UK and New York, NY, USA, 84 pp., 2013.

Tang, Q., Gao, H., Lu, H., and Lettenmaier, D. P.: Remote sensing: hydrology, Prog. Phys. Geogr., 33, 490–509, doi:10.1177/0309133309346650, 2009.

Trambauer, P., Maskey, S., Winsemius, H., Werner, M., and Uhlenbrook, S.: A review of continental scale hydrological models and their suitability for drought forecasting in (sub-Saharan) Africa, Phys. Chem. Earth A/B/C, 66, 16–26, doi:10.1016/j.pce.2013.07.003, 2013.

Uppala, S. M., Kållberg, P. W., Simmons, A. J., Andrae, U., Bechtold, V. D. C., 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., Berg, L. Van De, 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., and Woollen, J.: The ERA-40 re-analysis, Q. J. Roy. Meteorol. Soc., 131, 2961–3012, doi:10.1256/qj.04.176, 2005.

Voisin, N., Wood, A. W., and Lettenmaier, D. P.: Evaluation of Precipitation Products for Global Hydrological Prediction, J. Hydrometeorol., 9, 388–407, doi:10.1175/2007JHM938.1, 2008.

Vörösmarty, C. J., Hoekstra, A. Y., Bunn, S. E., Conway, D., and Gupta, J.: Fresh water goes global, Science, 349, 478–479, doi:10.1126/science.aac6009, 2015.

Wada, Y., van Beek, L. P. H., van Kempen, C. M., Reckman, J. W. T. M., Vasak, S., and Bierkens, M. F. P.: Global depletion of groundwater resources, Geophys. Res. Lett., 37, L20402, doi:10.1029/2010GL044571, 2010.

Weedon, G. P., Gomes, S., Viterbo, P., Österle, H., Adam, J. C., Bellouin, N., Boucher, O. and Best, M.: The WATCH Forcing Data: a meteorological forcing dataset for land surface- and hydrological models, Watch Techn. Rep. 22, p. 41, available at: http://www.eu-watch.org/media/default.aspx/emma/org/10376311/WATCH+Technical+Report+Number+22+The+WATCH+forcing+data+1958-2001+A+meteorological+forcing+dataset+for+land+surface-+and+hydrological-models.pdf (last access: 22 December 2015), 2010.

Weedon, G. P., Gomes, S., Viterbo, P., Shuttleworth, W. J., Blyth, E., Österle, H., Adam, J. C., Bellouin, N., Boucher, O., and Best, M.: Creation of the WATCH Forcing Data and its use to assess global and regional reference crop evaporation over land during the twentieth century, J. Hydrometeorol., 12, 823–848, doi:10.1175/2011JHM1369.1, 2011.

Weedon, G. P., Balsamo, G., Bellouin, N., Gomes, S., Best, M. J., and Viterbo, P.: The WFDEI meteorological forcing data set: WATCH Forcing Data methodology applied to ERA-Interim reanalysis data, Water Resour. Res., 50, 7505–7514, doi:10.1002/2014WR015638, 2014.

Wisser, D., Fekete, B. M., Vörösmarty, C. J., and Schumann, A. H.: Reconstructing 20th century global hydrography: a contribution to the Global Terrestrial Network- Hydrology (GTN-H), Hydrol. Earth Syst. Sci., 14, 1–24, doi:10.5194/hess-14-1-2010, 2010.

Yoshimura, K. and Kanamitsu, M.: Dynamical global downscaling of global reanalysis, Mon. Weather Rev., 136, 2983–2998, doi:10.1175/2008MWR2281.1, 2008.

5 Impact of climate forcing uncertainty and human water use on global and continental water balance components

Full citation: Müller Schmied, H., Adam, L., Eisner, S., Fink, G., Flörke, M., Kim, H., Oki, T., Portmann, F. T., Reinecke, R., Riedel, C., Song, Q., Zhang, J. and Döll, P.: Impact of climate forcing uncertainty and human water use on global and continental water balance components, Proc. Int. Assoc. Hydrol. Sci., 374(7), 53–62, doi:10.5194/piahs-374-53-2016, 2016.

Declaration of portion of own work

The model experiment setup was framed mainly between H. Müller Schmied and P. Döll. All model runs, analyses, figures and tables were done by H. Müller Schmied. The draft of the manuscript was written by H. Müller Schmied and revised by the co-authors. In overall percentage, H. Müller Schmied contributed around 90% of this manuscript.

Hannes Müller Schmied

Proc. IAHS, 374, 53–62, 2016
proc-iahs.net/374/53/2016/
doi:10.5194/piahs-374-53-2016

Impact of climate forcing uncertainty and human water use on global and continental water balance components

Hannes Müller Schmied[1,2], Linda Adam[1], Stephanie Eisner[3], Gabriel Fink[3], Martina Flörke[3], Hyungjun Kim[4], Taikan Oki[4], Felix Theodor Portmann[1], Robert Reinecke[1], Claudia Riedel[1], Qi Song[1], Jing Zhang[1], and Petra Döll[1]

[1]Institute of Physical Geography, Goethe University Frankfurt, Frankfurt, Germany
[2]Senckenberg Biodiversity and Climate Research Centre (BiK-F), Frankfurt, Germany
[3]Center for Environmental Systems Research (CESR), University of Kassel, Kassel, Germany
[4]Institute of Industrial Science, The University of Tokyo, Tokyo, Japan

Correspondence to: Hannes Müller Schmied (hannes.mueller.schmied@em.uni-frankfurt.de)

Published: 17 October 2016

Abstract. The assessment of water balance components using global hydrological models is subject to climate forcing uncertainty as well as to an increasing intensity of human water use within the 20th century. The uncertainty of five state-of-the-art climate forcings and the resulting range of cell runoff that is simulated by the global hydrological model WaterGAP is presented. On the global land surface, about 62 % of precipitation evapotranspires, whereas 38 % discharges into oceans and inland sinks. During 1971–2000, evapotranspiration due to human water use amounted to almost 1 % of precipitation, while this anthropogenic water flow increased by a factor of approximately 5 between 1901 and 2010. Deviation of estimated global discharge from the ensemble mean due to climate forcing uncertainty is approximately 4 %. Precipitation uncertainty is the most important reason for the uncertainty of discharge and evapotranspiration, followed by shortwave downward radiation. At continental levels, deviations of water balance components due to uncertain climate forcing are higher, with the highest discharge deviations occurring for river discharge in Africa (−6 to 11 % from the ensemble mean). Uncertain climate forcings also affect the estimation of irrigation water use and thus the estimated human impact of river discharge. The uncertainty range of global irrigation water consumption amounts to approximately 50 % of the global sum of water consumption in the other water use sector.

1 Introduction

The interest in global-scale water resources assessments has increased in the last two decades. There has been an increasing number of publications in this field (Web of Science, topic "global scale" AND "water resources", 1981–1990: 0 entries; 1991–2000: 6 entries; 2001–2010: 64 entries; 2011–2015: 85 entries), and a number of global hydrological models (GHMs) have been developed (Bierkens, 2015). The UN and other international organizations require global-scale information on water resources and their use, e.g. UNESCO's World Water Assessment Programme (www.unesco.org/water/wwap) or the Transboundary Waters Assessment Programme (TWAP, http://www.geftwap.org/twap-project), which can only be provided by modelling approaches due to a lack of observations with global coverage. Such model-based assessments require meteorological variables as climate forcing input. Currently, a number of state-of-the-art global-scale climate forcings are available that are all based on weather models and differ in terms of methodology including the underlying reanalysis and in terms of observation data used for bias correction. Different climate forcings result in large differences in simulated water fluxes and states as has already been shown by Biemans et al. (2009) for precipitation uncertainty and by Müller Schmied et al. (2014) regarding the uncertainty caused by two climate forcings that differ with respect to other climate variables. Analyses of the impact of different climate forcings are currently the focus of

Table 1. Summary of climate forcing characteristics used in this study. Abbreviations: precipitation – P, temperature – T, shortwave downward radiation – SWD, longwave downward radiation – LWD.

Name	Time span	Basis	Bias correction	Reference
GSWP3	1901–2010	20th Century Reanalysis using NCEP atmosphere land model	GPCC v6 (P), and undercatch correction (Hirabayashi et al., 2008) CRU TS3.21 (other variables)	http://hydro.iis.u-tokyo.ac.jp/GSWP3
PGFv2	1901–2012	NCEP-NCAR reanalysis	CRU TS3.21 (P, T), no precipitation undercatch correction, U Maryland, CRU TS3.21 cloud cover (SWD), U Maryland (LWD)	Updated version of Sheffield et al. (2006), information based on personal communication with J. Sheffield (2015)
WFD	1901–2001	ERA-40 reanalysis	GPCCv4 (P), undercatch correction using Adam and Lettenmaier (2003), CRU TS 2.1 cloud cover (SWD), CRU TS 2.1 temperature (T)	Weedon et al. (2010)
WFD_WFDEI	1901–2010	WFD 1901–1978, WFDEI (based on ERA-Interim reanalysis) afterwards	GPCC v5 (v6 for 2010) (P), undercatch correction using Adam and Lettenmaier (2003), CRU TS 3.1/3.21 cloud cover (SWD), CRU TS 3.1/3.21 temperature (T)	Weedon et al. (2014)
WFDEI_hom	1901–2010	As WFD_WFDEI, but WFD homogenized using a multiplicative approach for SWD and LWD and additive approach for T	Homogenization: Haddeland et al. (2012), Müller Schmied et al. (2016)	

model intercomparison studies such as the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP) in its phase 2a, where (among other sectors) several global and regional water models are driven by four state-of-the-art climate forcings and compared to historical observations of, for example, discharge and actual evapotranspiration. In particular, the propagation of climate forcing uncertainty at multiple scales (grid-cell level, continental, global) is one topic to be addressed in ISIMIP2a.

Humans increasingly influence the water cycle through water abstractions (Oki and Kanae 2006), in particular for irrigation (e.g. Siebert et al., 2015) but also for other purposes like thermal power plant cooling, manufacturing, livestock production and domestic sectors (Flörke et al., 2013). Quantification of sectoral water abstractions and consumptive water use (also called water consumption, the amount of the abstracted water that evapotranspires during human water use or is incorporated in products), and in particular of the source of water, is highly uncertain due to lack of data (Döll et al., 2016). In some regions, irrigation by groundwater leads to groundwater depletion problems (Döll et al., 2014a; Wada, 2016), and it has been estimated that in groundwater depletion areas, farmers irrigate with only 70 % of the optimal amount of water (Döll et al., 2014a).

Given the large uncertainties, we aim to answer the following research questions by using the Water Global Assessment and Prognosis (WaterGAP) GHM in its version 2.2 (ISIMIP2a):

1. How does climate forcing affect computed runoff at the grid-cell level?

2. How does climate forcing uncertainty and human water use affect long-term average water balance components (including human water use) on global and continental scales?

In Sect. 2 we briefly present the model and climate forcings used in this study. Results are presented and discussed in Sect. 3. The paper ends with a conclusion (Sect. 4), where we answer the research questions, followed by an outlook.

2 Data and methods

The global water availability and water use model WaterGAP (Alcamo et al., 2003; Döll et al., 2003; Müller Schmied et al., 2014) was applied using version WaterGAP 2.2 (ISIMIP2a). The main model characteristics of version 2.2 are described in Müller Schmied et al. (2014), and the differences to the ISIMIP2a version are described in Müller Schmied et

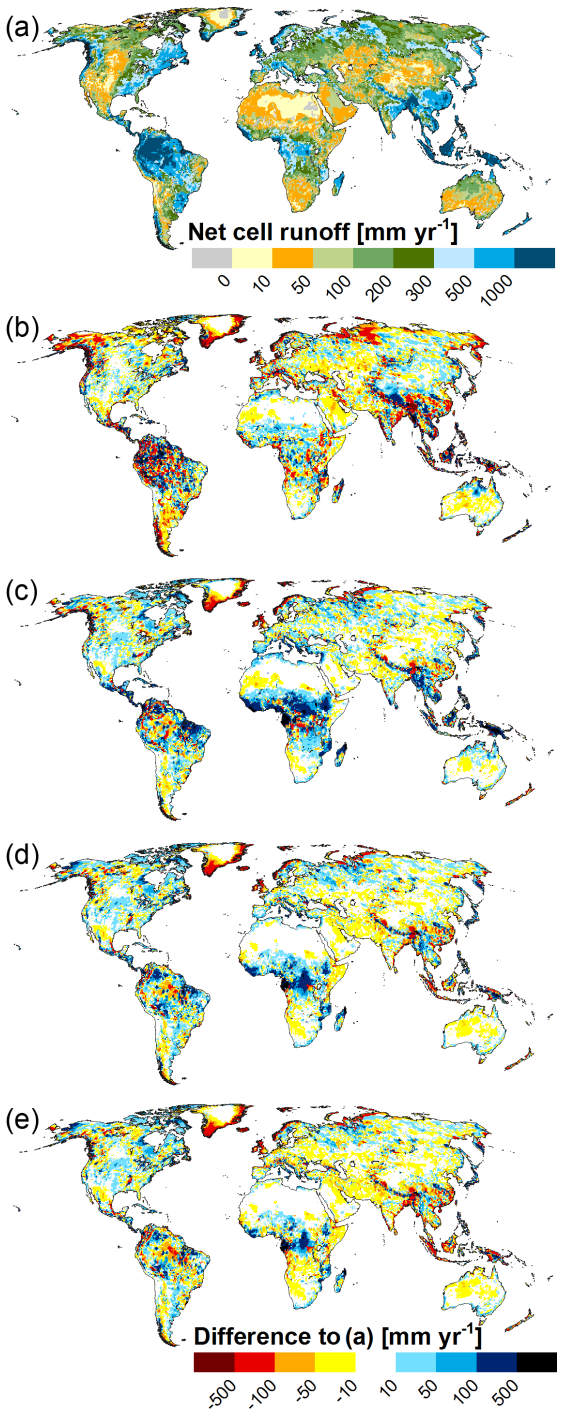


Figure 1. Long-term (1971–2000) average net cell runoff of the model variants, displayed as absolute numbers for GSWP3 **(a)** and differences to the other forcings, computed as PGFv2 minus GSWP3 **(b)**, WFD minus GSWP3 **(c)**, WFD_WFDEI minus GSWP3 **(d)** and WFDEI_hom minus GSWP3 **(e)**. Negative values in **(a)** indicate that water inflow into cell from upstream and by precipitation is larger than outflow due to evaporation from surface water bodies. All units in mm yr^{-1}.

al. (2016). WaterGAP has a spatial resolution of $0.5 \times 0.5°$ (i.e. 55×55 km at the Equator) and consists of five water use models that are linked through the Ground Water Surface Water Use submodule with the WaterGAP global hydrology model (WGHM). Within WGHM, water storage changes in several compartments and freshwater fluxes are modelled at a daily time step. WGHM is calibrated to match long-term average discharge at 1319 observation points (from GRDC database) within 1 % deviation by adjusting one to three parameters (calibration details in Müller Schmied et al., 2014).

Four state-of-the-art climate forcings provided by the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP) in its current phase 2a (https://www.isimip.org/about/#simulation-rounds-isimip2a) plus a fifth homogenized forcing were used to force WaterGAP. Table 1 summarizes the main characteristics of all five climate forcing datasets. For a detailed description, the reader is referred to Müller Schmied et al. (2016). The names of the model runs are similar to the names of the climate forcings.

3 Results and discussion

3.1 Global runoff at grid-cell level

Net cell runoff (computed as outflow minus inflow of each grid cell) differs considerably between the different climate forcing datasets (Fig. 1). This can be attributed to large differences between climate forcings at grid-cell level, in particular with respect to precipitation. Different observational datasets are used to bias-correct P (PGFv2 based on CRU and the others based on different versions of GPCC). This results in large differences (in both directions) at the regional scale for South America and South East Asia. Obviously, the unequal P gauging networks underlying the observational datasets and/or varying regionalization approaches lead to the large differences. In addition, PGFv2 is not corrected for undercatch of solid precipitation (J. Sheffield, personal communication, 2015) while all other datasets are. As undercatch correction (e.g. Adam and Lettenmaier, 2003) leads to the highest P increases in northern (snow-dominated) latitudes, P (and consequently net cell runoff) is lower for PGFv2 (red areas in the northern latitudes in Fig. 1b).

The main reason for the large discrepancies between GSWP3 and WFD in equatorial regions (Fig. 1c) is attributable to systematically smaller SWD (Fig. A3c) in energy-limited areas for the WFD dataset. This effect is lessened in the combined WFD_WFDEI dataset (Fig. 1d) and even more in the homogenized forcing WFDEI_hom (Fig. 1e), as WFDEI shows systematically higher SWD than WFD. The higher SWD in parts of Asia, western Africa and Australia (Fig. A3d, e) does not influence net cell runoff significantly because these regions are water-limited: evaporation and runoff are mainly controlled by precipitation and not by available energy. In many regions where SWD is in-

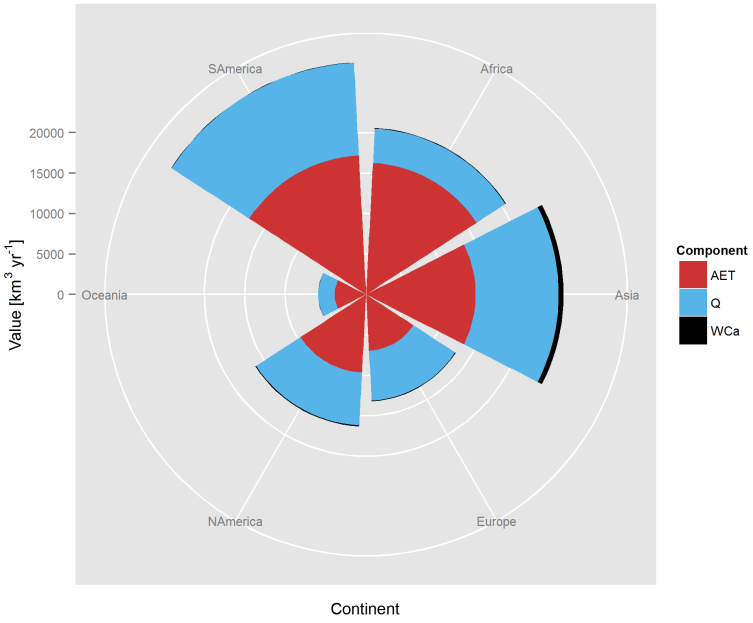


Figure 2. Continental water balance components 1971–2000 ($km^3\ yr^{-1}$); ensemble mean of the five climate forcings. Abbreviations: actual evapotranspiration (AET), discharge into oceans and inland sinks (Q), actual water consumption (WCa).

Table 2. Global water balance components for land area (except Antarctica and Greenland) in % of precipitation (row 1) for the five model variants and 1971–2000. Cells representing inland sinks were excluded but discharge into inland sinks was included.

No.	Component	GSWP3	PGFv2	WFD	WFDEI_hom	WFD_WFDEI
1	Precipitation P ($km^3\ yr^{-1}$)	109 631	103 525	110 690	111 050	111 050
2	Actual evapotranspiration AET[a]	62.0	61.3	61.1	63.0	62.0
3	Discharge into oceans and inland sinks Q[b]	37.1	37.8	38.1	36.2	37.2
4	Water consumption (actual) WCa	0.9	0.9	0.8	0.9	0.8
5	Change of total water storage dS / dt[c]	−0.01	−0.03	−0.02	−0.02	−0.07

[a] AET does not include evapotranspiration caused by human water use, i.e. actual water consumption WCa. [b] Taking into account anthropogenic water use. [c] Total water storage (TWS) of 31 December 2000 minus TWS of 31 December 1970, divided by the number of 30 years.

creased, LWD is decreased (and vice versa), which reduces the effect on net radiation. T effects model results via the equation for potential evapotranspiration (Priestley and Taylor, 1972), via snow dynamics as well as the leaf area index model that affects canopy evaporation (details in Müller Schmied et al., 2014). As T differs only little between the forcing datasets (Fig. A2), effects of T differences on simulated net cell runoff are expected to be relatively small.

3.2 Continental water balance components

Figure 2 displays the continental-scale partitioning of precipitation into actual evapotranspiration AET, river discharge Q, and human water consumption WCa. South America and Africa have nearly the same absolute amount of AET, but values for Q differ strongly. As a consequence of extensive irrigated agriculture especially in India and China (Siebert et al., 2015), the highest water consumption occurs in Asia, where 2.5 % of precipitation is evapotranspired, mainly due to irrigation. For the other continents, water consumption plays – relative to the other water balance components – only a marginal role. The lowest runoff coefficient (Q/P) is found in Africa (0.21), whereas runoff coefficients vary between 0.34 (Oceania) and 0.47 (Europe) for the other continents. Hence, differences in P result in higher relative uncertainties of estimated water resources for Africa. The deviation from the mean continental value for Q among the climate forcings is between −5.9 (calculated as min Q / mean Q) and 10.9 % (max Q / mean Q) for Africa, whereas for all other continents deviations are lower (−5.4 and −2.2 as minimum Q, 2.5–5.0 % as maximum Q).

3.3 Global water balance components

Compared to the continental-scale deviation of water balance components, the impact of climate forcing uncertainty

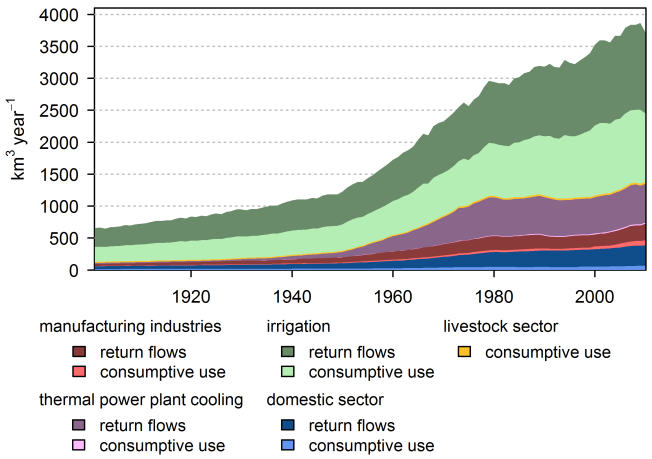


Figure 3. Development of water abstractions (sum of return flows and consumptive use) and water consumption (the amount of water that is evapotranspired or incorporated in products, light colours) of the five water use sectors considered in WaterGAP for 1901–2010. Values for irrigation (modelled with 70 % of demand in grid cells with groundwater depletion) are averaged across the five climate forcings; other sectors are modelled independently of climate forcing and taken from Flörke et al. (2013).

is levelling out at the global scale to a certain degree (Table 2). Global runoff coefficients vary between 0.362 and 0.381, and deviation of global Q from the ensemble mean is −3.8 to 3.7 %. Actual evapotranspiration is estimated to range between 61.1 and 63.0 % of global P (Table 2). The lowest value for AET (and highest value for Q) is computed when using WFD climate forcing. Here, average global SWD is 15 W m^{-2} lower compared to the other forcings (see also Müller Schmied et al., 2016, their Fig. 1). In absolute numbers, differences in AET and Q resulting from the five climate forcings are considerable. For example, global discharge values range from 39 200 to 42 200 km^3 yr^{-1}: the uncertainty range is equal to thrice the total water consumption ($\sim$ 930 km^3 yr^{-1}).

On the global scale, sectoral water uses have strongly increased since 1901 (Fig. 3). Whereas overall water abstractions (consumptive use) are about 650 (260) km^3 yr^{-1} in the year 1901, values are about 5 times higher with 3700 (1250) km^3 yr^{-1} in 2010. In contrast to Müller Schmied et al. (2016, their Fig. 1), where water consumption of each climate forcing is presented using different time step aggregations, Fig. 3 shows the proportion of potential (if water were available without limitation) consumptive water use components (light colours) and the amount of return flows (dark colours). The most important water use sector regarding both abstraction and consumption is the irrigation sector. The sum of potential water consumption of all water use sectors (except irrigation) throughout the period 1971–2000 is 112 km^3 yr^{-1}, whereas the sums of potential irrigation water consumption vary between 834 and 894 km^3 yr^{-1} depending on the climate forcings. Together with the other potential water uses (manufacturing, cooling of thermal power plants, domestic and livestock sector), the demand of consumptive water uses ranges from 946 to 1006 km^3 yr^{-1}. Due to limited water availability to satisfy the demand, actual water consumption (WCa) ranges between 915 (WFD) and 960 (PGFv2) km^3 yr^{-1} (all numbers 1971–2000). Hence, water availability reduces the impact of climate forcing uncertainty when modelling water use demand. The uncertainty range of estimated global irrigation water consumption due to the climate forcing is therefore about 50 % of the sum of all the other water use sectors.

4 Conclusions

Within this study, the WaterGAP 2.2 (ISIMIP2a) model was used to assess water balance components on grid-cell, continental and global scale as well as the development of human water use on the global scale. The research questions can be answered as follows:

1. How does climate forcing affect computed runoff at the grid-cell level?

On the grid-cell level, the effect of climate forcing uncertainty on computed runoff is very large. In particular, usage of different observation-based products to bias-correct reanalysis data affects the spatial distribution of runoff. Furthermore, undercatch correction (or the lack thereof) of P leads to differences in model estimates. Whereas T uncertainty does not lead to clearly visible spatial differences in computed runoff, SWD uncertainty was found to have a large impact in energy-limited regions like tropical Africa. For water-limited areas, this is not the case.

2. How does climate forcing uncertainty and human water use affect long-term average water balance components on global and continental scales?

Climate forcing uncertainty is high (Figs. A1–A4), and most important are differences in P and SWD. At the continental scale, these uncertainties lead to large differences in calculated water balance components, in particular in regions with high P uncertainty and low runoff coefficient (e.g. Africa). Global-scale values vary less in relative terms as deviations even out with spatial aggregation. The uncertainty range of estimated global irrigation water consumption due to uncertain climate forcing is around 50 % of the water consumption in the other water use sectors.

Multi-model hydrological assessments as done by the ISIMIP initiative for both historical periods (e.g. Haddeland et al., 2011) and future scenarios (e.g. Schewe et al., 2014) will help to relate the uncertainties of water balance components at different scales of aggregation that are caused by different climate forcings to uncertainties due to the hydrological models themselves. To constrain both types of uncertainty, model calibration not only of mean annual river

discharge (as done for the WaterGAP model) but also of remote-sensing-based data like total water storage variations from GRACE (Eicker et al., 2014; Döll et al., 2014b, 2016) is promising, but collection and sharing of in situ data remains crucial (Fekete et al., 2015).

5 Data availability

The WaterGAP output will become freely available to the public within the framework of the ISIMIP project phase 2a, but it is not yet clarified where the data will be hosted (please check https://www.isimip.org/outputdata/ for updates). The homogenized climate forcing WFDEI_hom is not included within the ISI-MIP project. All model outputs used in this study are available on request from the corresponding author.

Appendix A

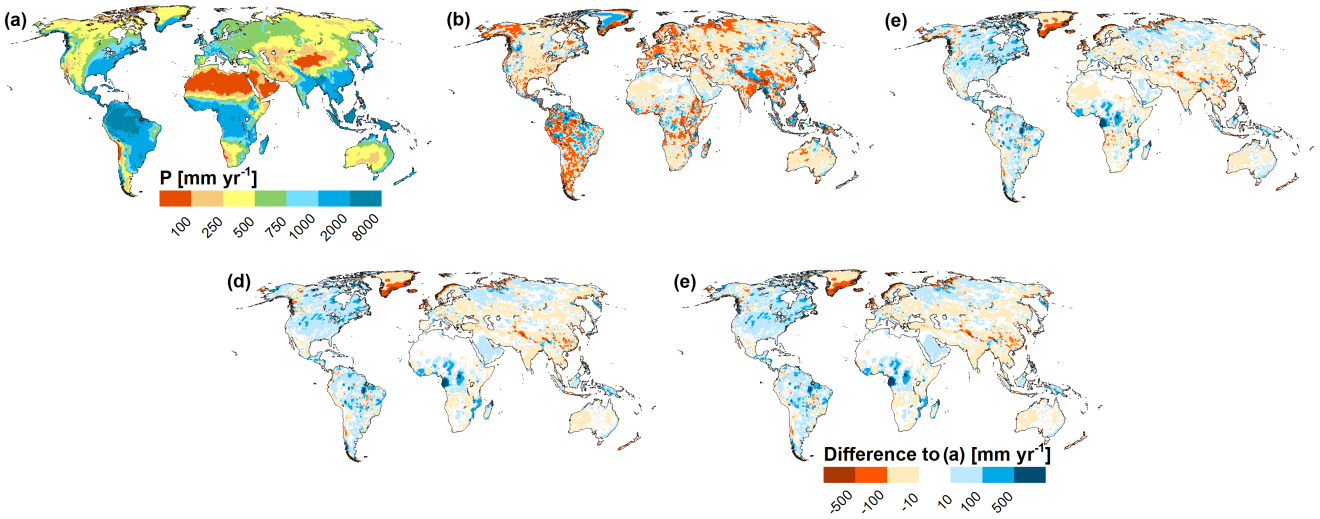


Figure A1. Long-term (1971–2000) average precipitation of the model variants, displayed as absolute number for GSWP3 **(a)** and differences to the other forcings, computed as PGFv2 minus GSWP3 **(b)**, WFD minus GSWP3 **(c)**, WFD_WFDEI minus GSWP3 **(d)**, and WFDEI_hom minus GSWP3 **(e)**. All units in mm yr^{-1}.

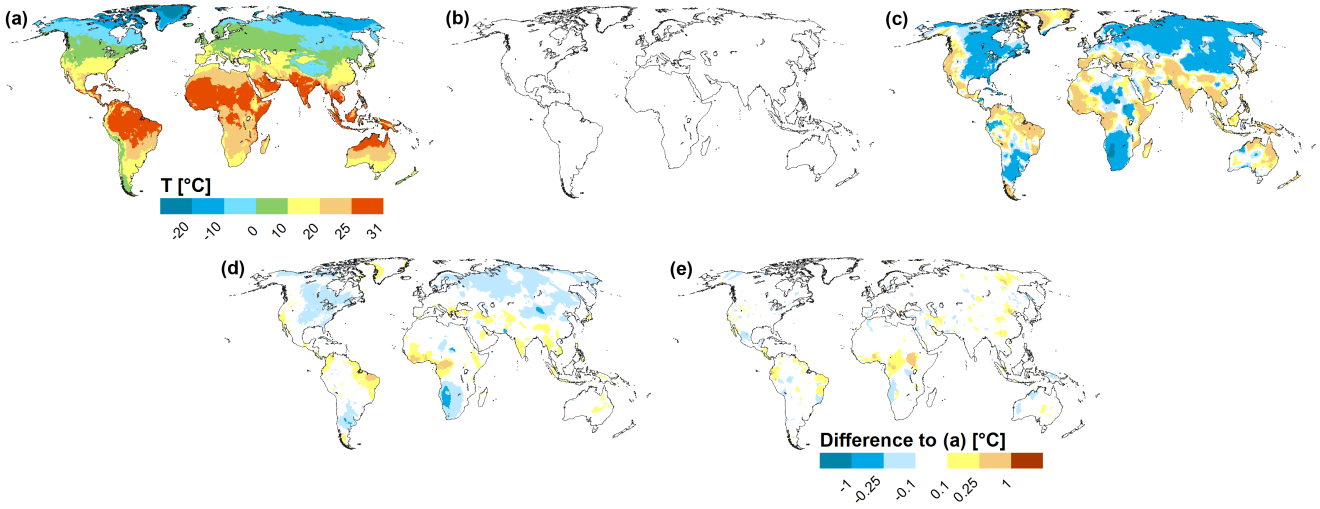


Figure A2. Long-term (1971–2000) average temperature of the model variants, displayed as absolute number for GSWP3 **(a)** and differences to the other forcings, computed as PGFv2 minus GSWP3 **(b)**, WFD minus GSWP3 **(c)**, WFD_WFDEI minus GSWP3 **(d)**, and WFDEI_hom minus GSWP3 **(e)**. All units in °C.

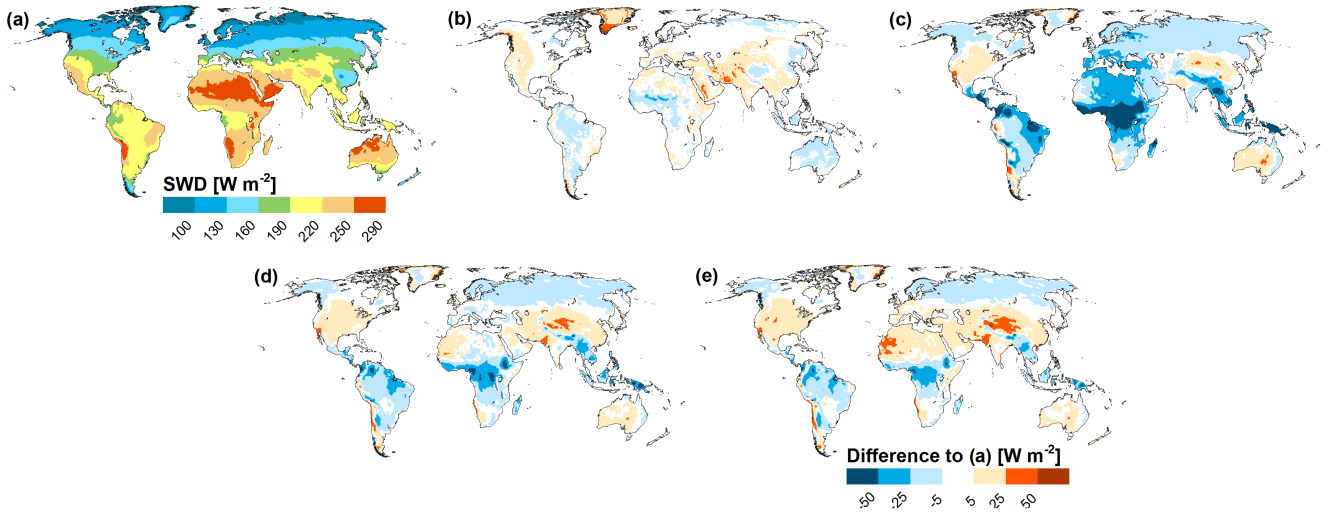


Figure A3. Long-term (1971–2000) average shortwave downward radiation of the model variants, displayed as absolute number for GSWP3 **(a)** and differences to the other forcings, computed as PGFv2 minus GSWP3 **(b)**, WFD minus GSWP3 **(c)**, WFD_WFDEI minus GSWP3 **(d)**, and WFDEI_hom minus GSWP3 **(e)**. All units in W m^{-2}.

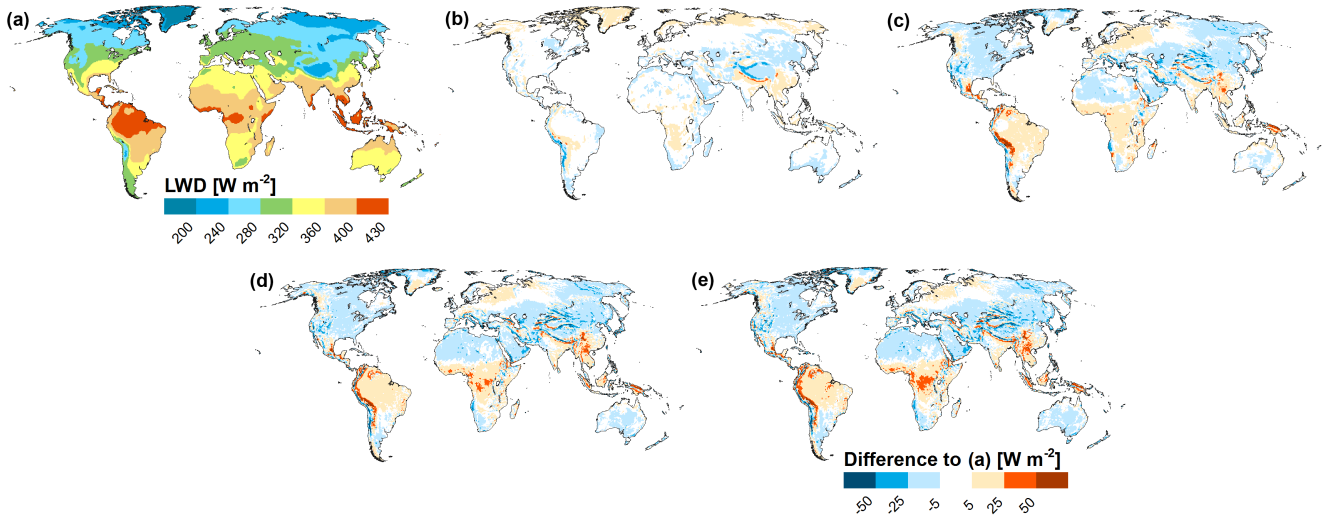


Figure A4. Long-term (1971–2000) average longwave downward radiation of the model variants, displayed as absolute number for GSWP3 **(a)** and differences to the other forcings, computed as PGFv2 minus GSWP3 **(b)**, WFD minus GSWP3 **(c)**, WFD_WFDEI minus GSWP3 **(d)**, and WFDEI_hom minus GSWP3 **(e)**. All units in W m^{-2}.

Acknowledgements. The authors thank the Global Runoff Data Centre (GRDC, http://grdc.bafg.de), 56068 Koblenz, Germany, for providing the discharge data used in this study for calibrating the model. We are also grateful to the ISI-MIP coordination team as well as the leaders of the water sector (Simon Gosling and Rutger Dankers) for providing the climate forcings and the support. Furthermore, we thank Wolfgang Grabs for organizing the international conference "Water Resources Assessment & Seasonal Prediction" (13–16 October in Koblenz, Germany), where some content of this paper was presented.

References

Adam, J. C. and Lettenmaier, D. P.: Adjustment of global gridded precipitation for systematic bias, J. Geophys. Res., 108, 4257, doi:10.1029/2002JD002499, 2003.

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

Biemans, H., Hutjes, R. W. A., Kabat, P., Strengers, B. J., Gerten, D., and Rost, S.: Effects of precipitation uncertainty on discharge calculations for main river basins, J. Hydrometeorol., 10, 1011–1025, doi:10.1175/2008JHM1067.1, 2009.

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

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

Döll, P., Müller Schmied, H., Schuh, C., Portmann, F. T., and Eicker, A.: Global-scale assessment of groundwater depletion and related groundwater abstractions: Combining hydrological modeling with information from well observations and GRACE satellites, Water Resour. Res., 50, 5698–5720, 2014a.

Döll, P., Fritsche, M., Eicker, A., and Müller Schmied, H.: Seasonal water storage variations as impacted by water abstractions: Comparing the output of a global hydrological model with GRACE and GPS Observations, Surv. Geophys, 35, 1311–1331, doi:10.1007/s10712-014-9282-2, 2014b.

Döll, P., Douville, H., Güntner, A., Müller Schmied, H., and Wada, Y.: Modelling freshwater resources at the global scale: Challenges and prospects, Surv. Geophys., 37, 195–221, doi:10.1007/s10712-015-9347-x, 2016.

Eicker, A., Schumacher, M., Kusche, J., Döll, P., and Müller Schmied, H.: Calibration/data assimilation approach for integrating GRACE data into the WaterGAP Global Hydrology Model (WGHM) using an Ensemble Kalman Filter: First results, Surv. Geophys., 35, 1285–1309, doi:10.1007/s10712-014-9309-8, 2014.

Fekete, B. M., Robarts, R. D., Kumagai, M., Nachtnebel, H.-P., Odada, E., and Zhulidov, A. V.: Time for in situ renaissance, Science, 349, 685–686, doi:10.1126/science.aac7358, 2015.

Flörke, M., Kynast, E., Bärlund, I., Eisner, S., Wimmer, F., and Alcamo, J.: Domestic and industrial water uses of the past 60 years as a mirror of socio-economic development: A global simulation study, Global Environ. Chang., 23, 144–156, doi:10.1016/j.gloenvcha.2012.10.018, 2013.

Haddeland, I., Clark, D. B., Franssen, W., Ludwig, F., Voß, F., Arnell, N. W., Bertrand, N., Best, M., Folwell, S., Gerten, D., Gomes, S., Gosling, S. N., Hagemann, S., Hanasaki, N., Harding, R., Heinke, J., Kabat, P., Koirala, S., Oki, T., Polcher, J., Stacke, T., Viterbo, P., Weedon, G. P., and Yeh., P.: Multi-model estimate of the global terrestrial water balance: Setup and first results, J. Hydrometeorol., 12, 869–884, doi:10.1175/2011JHM1324.1, 2011.

Haddeland, I., Heinke, J., Voß, F., Eisner, S., Chen, C., Hagemann, S., and Ludwig, F.: Effects of climate model radiation, humidity and wind estimates on hydrological simulations, Hydrol. Earth Syst. Sci., 16, 305–318, doi:10.5194/hess-16-305-2012, 2012.

Hirabayashi, Y., Kanae, S., Motoya, K., Masuda, K., and Döll, P.: A 59-year (1948–2006) global meteorological forcing data set for land surface models. Part II: Global snowfall estimation, Hydrol. Res. Lett., 2, 65–69, doi:10.3178/hrl.2.65, 2008.

Müller Schmied, H., Eisner, S., Franz, D., Wattenbach, M., Portmann, F. T., Flörke, M., and Döll, P.: Sensitivity of simulated global-scale freshwater fluxes and storages to input data, hydrological model structure, human water use and calibration, Hydrol. Earth Syst. Sci., 18, 3511—3538, doi:10.5194/hess-18-3511-2014, 2014.

Müller Schmied, H., Adam, L., Eisner, S., Fink, G., Flörke, M., Kim, H., Oki, T., Portmann, F. T., Reinecke, R., Riedel, C., Song, Q., Zhang, J., and Döll, P.: Variations of global and continental water balance components as impacted by climate forcing uncertainty and human water use, Hydrol. Earth Syst. Sci., 20, 2877–2898, doi:10.5194/hess-20-2877-2016, 2016.

Oki, T. and Kanae, S.: Global hydrological cycles and world water resources, Science, 313, 1068–1072, doi:10.1126/science.1128845, 2006.

Priestley, C. H. B. and Taylor, R. J.: Assessment of surface heat flux and evaporation using large-scale parameters, Mon. Weather Rev., 100, 81–92, doi:10.1175/1520-0493(1972)100<0081:OTAOSH>2.3.CO;2, 1972.

Schewe, J., Heinke, J., Gerten, D., Haddeland, I., Arnell, N. W., Clark, D. B., Dankers, R., Eisner, S., Fekete, B. M., Colón-González, F. J., Gosling, S. N., Kim, H., Liu, X., Masaki, Y., Portmann, F. T., Satoh, Y., Stacke, T., Tang, Q., Wada, Y., Wisser, D., Albrecht, T., Frieler, K., Piontek, F., Warszawski, L., and Kabat, P.: Multimodel assessment of water scarcity under climate change, P. Natl. Acad. Sci. USA, 111, 3245–3250, doi:10.1073/pnas.1222460110, 2014.

Sheffield, J., Goteti, G., and Wood, E. F.: Development of a 50-year high-resolution global dataset of meteorological forcings for land surface modeling, J. Climate, 19, 3088–3111, doi:10.1175/JCLI3790.1, 2006.

Siebert, S., Kummu, M., Porkka, M., Döll, P., Ramankutty, N., and Scanlon, B. R.: A global data set of the extent of irrigated land from 1900 to 2005, Hydrol. Earth Syst. Sci., 19, 1521–1545, doi:10.5194/hess-19-1521-2015, 2015.

Wada, Y.: Modeling groundwater depletion at regional and global scales: Present state and future prospects, Surv. Geophys., 37, 419–451, doi:10.1007/s10712-015-9347-x, 2016.

Weedon, G. P., Gomes, S., Viterbo, P., Österle, H., Adam, J. C., Bellouin, N., Boucher, O., and Best, M.: The WATCH Forcing Data: a meteorological forcing dataset for land surface- and hydrological models, Watch Techn. Rep., 22, 41 pp., available at: http://www.eu-watch.org/media/default.aspx/emma/org/10376311/

WATCH+Technical+Report+Number+22+The+WATCH+forcing+data+1958-2001+A+meteorological+forcing+dataset+for+land+surface-+and+hydrological-models.pdf (last access 29 July 2016), 2010.

Weedon, G. P., Balsamo, G., Bellouin, N., Gomes, S., Best, M. J., and Viterbo, P.: The WFDEI meteorological forcing data set: WATCH Forcing Data methodology applied to ERA-Interim reanalysis data, Water Resour. Res., 50, 7505–7514, doi:10.1002/2014WR015638, 2014.

6 Global-scale assessment of groundwater depletion and related groundwater abstractions: Combining hydrological modeling with information from well observations and GRACE satellites

Full citation: Döll, P., Müller Schmied, H., Schuh, C., Portmann, F. T. and Eicker, A.: Global-scale assessment of groundwater depletion and related groundwater abstractions: Combining hydrological modeling with information from well observations and GRACE satellites, Water Resour. Res., 50(7), 5698–5720, doi:10.1002/2014WR015595, 2014.

Declaration of portion of own work

P. Döll designed the experiment setup and wrote the text. H. Müller Schmied developed the model code, took part in concept development of groundwater recharge below surface water bodies and conducted the model runs. Together with the other co-authors, he revised the manuscript. In overall percentage, H. Müller Schmied contributed around 20% of this manuscript.

Hannes Müller Schmied

AGU PUBLICATIONS

Water Resources Research

RESEARCH ARTICLE
10.1002/2014WR015595

Key Points:
- Groundwater depletion is simulated by a global model
- Seventy percent deficit irrigation is likely in groundwater depletion areas
- About 15% of global groundwater abstractions are from nonrenewable sources

Correspondence to:
P. Döll,
p.doell@em.uni-frankfurt.de

Citation:
Döll, P., H. Müller Schmied, C. Schuh, F. T. Portmann, and A. Eicker (2014), Global-scale assessment of groundwater depletion and related groundwater abstractions: Combining hydrological modeling with information from well observations and GRACE satellites, *Water Resour. Res.*, *50*, 5698–5720, doi:10.1002/2014WR015595.

Received 17 MAR 2014
Accepted 16 JUN 2014
Accepted article online 18 JUN 2014
Published online 12 JUL 2014

Global-scale assessment of groundwater depletion and related groundwater abstractions: Combining hydrological modeling with information from well observations and GRACE satellites

Petra Döll[1], Hannes Müller Schmied[1], Carina Schuh[1], Felix T. Portmann[1], and Annette Eicker[2]

[1]Institute of Physical Geography, Goethe University Frankfurt, Frankfurt, Germany, [2]Institute of Geodesy and Geoinformation, University of Bonn, Bonn, Germany

Abstract Groundwater depletion (GWD) compromises crop production in major global agricultural areas and has negative ecological consequences. To derive GWD at the grid cell, country, and global levels, we applied a new version of the global hydrological model WaterGAP that simulates not only net groundwater abstractions and groundwater recharge from soils but also groundwater recharge from surface water bodies in dry regions. A large number of independent estimates of GWD as well as total water storage (TWS) trends determined from GRACE satellite data by three analysis centers were compared to model results. GWD and TWS trends are simulated best assuming that farmers in GWD areas irrigate at 70% of optimal water requirement. India, United States, Iran, Saudi Arabia, and China had the highest GWD rates in the first decade of the 21st century. On the Arabian Peninsula, in Libya, Egypt, Mali, Mozambique, and Mongolia, at least 30% of the abstracted groundwater was taken from nonrenewable groundwater during this time period. The rate of global GWD has likely more than doubled since the period 1960–2000. Estimated GWD of 113 km^3/yr during 2000–2009, corresponding to a sea level rise of 0.31 mm/yr, is much smaller than most previous estimates. About 15% of the globally abstracted groundwater was taken from nonrenewable groundwater during this period. To monitor recent temporal dynamics of GWD and related water abstractions, GRACE data are best evaluated with a hydrological model that, like WaterGAP, simulates the impact of abstractions on water storage, but the low spatial resolution of GRACE remains a challenge.

1. Introduction

The temporal development of groundwater storage depends on (1) the inflow into the groundwater body from the soil or from surface water bodies (recharge) and (2) the outflow from the groundwater to surface water bodies (base flow) or due to groundwater abstractions. All groundwater abstractions initially lead to a decrease of stored groundwater volume within the cone of depression of hydraulic heads around the pumping well. The cone of depression can stabilize if the lowered hydraulic heads have induced a decrease of base flow or an increase of groundwater recharge that balance the abstractions [*Zhou*, 2009]. Then, groundwater storage stabilizes at a lower equilibrium level as compared to the situation without groundwater abstractions. If, however, groundwater abstractions cannot be balanced by increased recharge and decreased discharge over a number of years, a long-term decline of hydraulic heads and groundwater storage, i.e., groundwater depletion (GWD), will result. Given the stochastic nature of climate, consideration of a time period of 10–30 years appears appropriate for defining GWD. If GWD occurs, groundwater is no longer exploited as a renewable resource but as a nonrenewable one. The rate of renewal is smaller than the rate of consumption, and the abstracted water very often stems from a wetter time period with higher recharge than today.

Even where groundwater systems are not affected by groundwater abstractions, long-term average groundwater recharge is not exactly equal to long-term average base flow, such that long-term changes in groundwater storage may occur. This is due to slow hydraulic head changes in response to temporally changing groundwater recharge. While large aquifers show unsteady conditions over thousands of years (e.g., in formerly wet but now desert regions like the Nubian aquifer system in North Africa which still discharges groundwater [*Gossel et al.*, 2004]), smaller aquifers are affected by decadal trends of groundwater recharge due to anthropogenic climate change. However, in most groundwater systems not affected by groundwater abstractions, long-term average base flow is at least roughly equal to long-term average groundwater

recharge [*Bredehoeft*, 2002]. Therefore, base flow reductions can only balance groundwater abstractions that are not larger than natural groundwater recharge, unless recharge from surface water bodies can be induced. Most cases of GWD occur in semiarid and arid regions, where runoff and groundwater recharge from soils are small and large amounts of irrigation water are required for crop production [*Margat et al*., 2006].

GWD is generally considered to be problematic with respect to (1) maintaining the natural capital for future generations, (2) land subsidence [*Konikow and Kendy*, 2005], (3) negative ecological effects of decreased base flow that may even lead to drying-up of wetlands and rivers, (4) saltwater intrusion, (5) increasing pumping costs, and (6) declining water supplies for some of the world's major agricultural areas, putting sustained crop production at risk. In addition, the water volume no longer stored in the ground ends up in the oceans, leading to sea level rise [*Konikow and Kendy*, 2005].

GWD has been quantified in some regions using well observations (e.g., High Plains Aquifer, United States [*McGuire*, 2011]) or modeling (e.g., Central Valley Aquifer, United States [*Faunt*, 2009]), or by subtracting modeled or observed estimates of other water storage compartments from satellite-based GRACE estimates of total water storage (TWS) change (e.g., Northern India and adjacent areas [*Tiwari et al*., 2009]; Central Valley, United States [*Famiglietti et al*., 2011]). The amount of global GWD is disputed. An average GWD of 127 and 195 km^3/yr for the time period 1993–2008 was estimated by *Konikow* [2011] and *Wada et al*. [2012a], respectively, while *Pokhrel et al*. [2012] computed a much higher value of 362 km^3/yr for 1981–2007. While the first two studies estimated a strong increase of global annual GWD over time, the value remains almost constant between 1950 and 2000 in the study of *Pokhrel et al*. [2012]. The three studies used different estimation approaches. *Konikow* [2011] derived global GWD based on (1) rather detailed information from local modeling or observations for the United States, (2) less precise information from a mix of methods for five aquifer systems outside the United States, and (3) an extrapolation approach for the rest of the world where depletion was determined by multiplying estimated groundwater withdrawals with the ratio of groundwater depletion to withdrawals in the United States around 2000 (0.15). *Wada et al*. [2012a] used a global hydrological model to compute GWD, with a spatial resolution of $0.5^\circ \times 0.5^\circ$, by subtracting groundwater abstractions from the sum of natural groundwater recharge and return flow from irrigation [*Wada et al*., 2012b]. Country-specific groundwater abstractions (not sector-specific) from the IGRAC GGIS database (http://www.un-igrac.org) were downscaled using total cell-specific water demand and surface water availability. To achieve a better fit to independent estimates, a correction was applied to nonarid grid cells such that global depletion was reduced to 80% of the computed value. Base flow reduction due to groundwater depletion was not modeled. *Pokhrel et al*. [2012] also applied a global hydrological model but only used information on total water abstractions, not on groundwater abstractions. They assumed cell-specific GWD whenever no "near-surface" water derived from current runoff was available to satisfy computed water demand [*Pokhrel et al*., 2012]. The larger depletion estimates may be due to an underestimation of availability of "near-surface" water and the fact that in case of irrigated areas equipped with surface irrigation only, any lack of surface water supplies can in reality not be compensated by groundwater abstractions but will lead to deficit irrigation.

With the study presented here, we wanted to (1) identify GWD areas world-wide in a consistent manner, taking into account that in semiarid and arid areas surface water bodies may be the most important source of groundwater recharge, (2) derive improved values of cell-specific, country, and global GWD, and (3) concurrently estimate the degree of deficit irrigation (irrigation below optimal). In addition, we wanted to assess the use of global-scale quantifications of TWS changes from GRACE satellites to monitor GWD.

Using the global hydrological and water use model WaterGAP, it is possible to simulate GWD by computing groundwater storage change as the (daily) difference between groundwater recharge and the sum of base flow and net groundwater abstraction, with base flow declining with decreasing groundwater storage [*Döll et al*., 2012]. Net groundwater abstractions are based on more detailed information on groundwater and surface water abstractions compared to *Wada et al*. [2012a]. In particular, information on irrigated areas equipped with surface water or groundwater irrigation was compiled for more than 15,000 national and subnational administrative units [*Siebert et al*., 2010], and information about artificial drainage was used to model irrigation return flows to groundwater [*Döll et al*., 2012]. Nevertheless, it was found that WaterGAP overestimates GWD in the High Plains Aquifer HPA (United States) for which the globally best GWD data, based on a very large number of well observations, are available [*Döll et al*., 2012]. When some amount of

groundwater recharge from the playas (small wetlands) in the HPA was assumed, overestimation decreased. The remaining overestimation was suspected to be caused by an overestimation of groundwater abstraction as WaterGAP assumes that farmers irrigate optimally while in reality they may apply less water in case of water scarcity.

While in humid areas groundwater can be assumed to discharge into surface water bodies, in semiarid and arid areas like the HPA it is more likely that surface water bodies recharge the groundwater. In the HPA, there are 61,000 playas, small seasonal wetlands with groundwater recharge rates that are 1–2 orders of magnitude larger than diffuse groundwater recharge from soil [*Gurdak and Roe*, 2010]. However, the standard version of WaterGAP 2.2 assumes that groundwater discharges into surface water in all grid cells. For this study of global GWD, we modified WaterGAP 2.2 such that in all semiarid and arid regions of the globe, groundwater is recharged from lakes, reservoirs, and wetlands. In addition, we adjusted the routing of water between the storage compartments soil, groundwater, surface water bodies, and river.

We then used the reliable independent information for the HPA to design model variants which varied by the amount of irrigation water use and the quantity of groundwater recharge from surface water bodies (section 2). We apply selected variants to simulate GWD world-wide and compare the results to independent GWD estimates (section 3). For the first time at the global scale, time series of GRACE total water storage TWS are compared to modeled TWS to assess GWD (section 3). Based on these comparisons, we compute a best estimate of GWD and related groundwater abstractions (section 3). In sections 4 and 5, we discuss the results and draw conclusions.

2. Methods

2.1. Description of WaterGAP 2.2a

With a spatial resolution of 0.5° × 0.5° (55 km × 55 km at the equator), WaterGAP 2.2 simulates continental water flows and storages as well as human water use for all land areas of the globe excluding Antarctica [*Alcamo et al*., 2003; *Döll et al*., 2003; *Müller Schmied et al*., 2014]. Sectoral water uses for irrigation, livestock, households, manufacturing, and cooling of thermal power plants are computed by separate models. With a daily time step, the irrigation water use model GIM [*Döll and Siebert*, 2002] computes consumptive water use for the fraction of each grid cell that is irrigated, consumptive use being defined as the part of the withdrawn water that evapotranspires during use. For all other sectors, both water withdrawals and consumptive water use are quantified by the water use models. Temporal development of domestic, manufacturing, and cooling water use is calculated as water use intensity (water use per capita or unit industrial output, considering structural and technological change over time) multiplied by the driving force of water use, either population (for domestic use), national manufacturing output (as Gross Value Added, which is a share of Gross Domestic Product), or national thermal electricity production [*Flörke et al*., 2013]. Temporal development of consumptive irrigation water use is driven by time series of climate and annually changing values of irrigated area per country [*Döll et al*., 2012].

Using monthly time series of irrigation water use and annual time series for the other water uses, taking into account information on the source of water, and making assumptions on irrigation water use efficiencies and return flows, the submodel GWSWUSE computes net abstractions from groundwater (NA_g) and from surface water (NA_s) [*Döll et al*., 2012]. The groundwater fractions of sectoral water uses are assumed to be constant in time due to lack of data. Net abstractions are computed as the difference of water withdrawals (gross water abstractions) from the specific source and the return flows from water use sites. Net abstractions are negative if withdrawals are smaller than return flows. For NA_g, this can only occur in case of irrigation from surface water.

Monthly time series of NA_g and NA_s are inputs to the WaterGAP Global Hydrology Model (WGHM), together with climate and physiogeographic data. With a daily time step, WGHM computes, for each 0.5° grid cell, flows among the water storage compartments canopy, snow, soil, groundwater, lakes, man-made reservoirs, wetlands, and rivers (Figure 1). "Global" lakes, reservoirs, and wetlands are distinguished from "local" ones by the fact that they are recharged not only by the runoff of the cell but also by the streamflow from upstream cells (Figure 1). If there are multiple lakes (or reservoirs or wetlands) within a grid cell, they are lumped into one. While global lakes are distinguished from global reservoirs, local lakes and reservoirs are lumped together into one local lake. The water balance of global lakes and reservoirs, which can cover

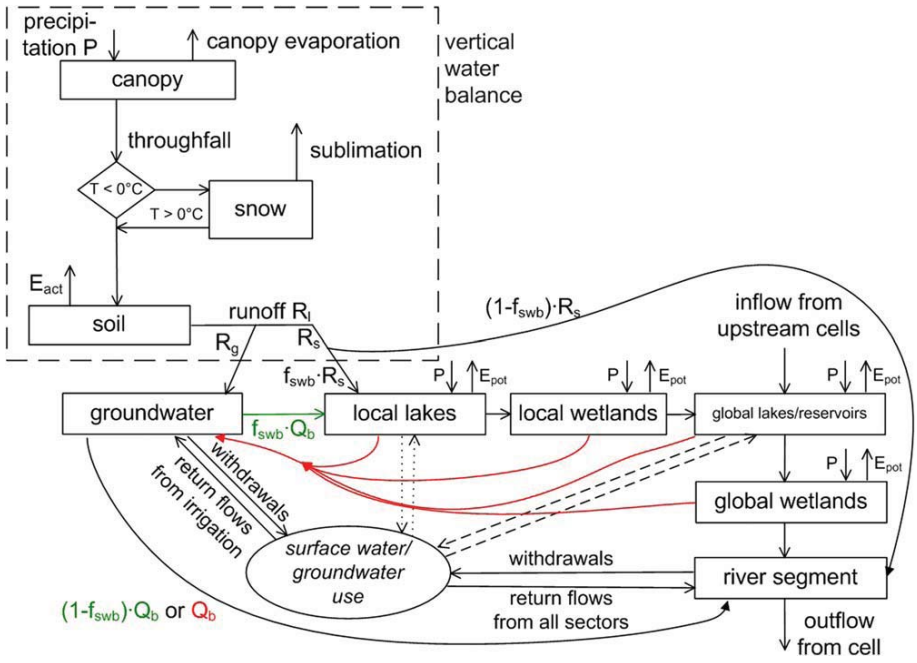


Figure 1. Schematic of water storage compartments (boxes) and flows (arrows) within each 0.5° grid cell of the WaterGAP Global Hydrology Model, including the simulation of water use impacts on water storage in groundwater and surface water. Compared to WaterGAP 2.1h [*Döll et al.*, 2012, Figure 1], routing of surface runoff and base flow was modified to simulate that not the whole runoff from land R_l is routed through the surface water bodies. In addition, in semiarid and arid regions, groundwater recharge is assumed to occur beneath surface water bodies (lakes, wetlands, and reservoirs but not rivers), while in humid areas, groundwater recharges surface water bodies. Flows that are only modeled in semiarid and arid regions are shown in red and those occurring only in humid regions are shown in green.

more than one grid cell, is computed in the grid cell where the outflow of the lake or reservoir is located. Monthly values of NA_g and NA_s are distributed evenly to all days of the months and subtracted from groundwater and surface water bodies, respectively. Flows include actual evapotranspiration, total runoff, groundwater recharge, and streamflow. Total runoff as computed by the vertical water balance is heuristically partitioned into fast surface and subsurface runoff R_s and groundwater recharge R_g as a function of relief, soil texture, hydrogeology, and the existence of glaciers or permafrost [*Döll and Fiedler*, 2008]. Both runoff components are routed to the river storage compartment of the grid cell applying a newly developed fractional routing algorithm (section 2.1.3) and are transported as streamflow to the downstream grid cell (Figure 1). WaterGAP 2.2 is calibrated against mean annual streamflow at 1319 gauging stations by adjusting one to three parameters in the upstream cells, and the adjusted main calibration parameter "runoff coefficient" is regionalized to grid cells outside the calibration basins based on climatic and physiogeographic basin characteristics. WaterGAP 2.2a is based on a slightly different precursor version of WaterGAP 2.2 [*Müller Schmied et al.*, 2014].

2.1.1. Modeling of Groundwater Recharge From Surface Water Bodies

In WaterGAP 2.1h [*Döll et al.*, 2012] and 2.2 [*Müller Schmied et al.*, 2014], there is no groundwater recharge from surface water bodies as groundwater is assumed to discharge into surface water bodies [*Döll et al.*, 2012, Figure 1]. In these model versions, groundwater storage changes are computed as

$$\frac{dGWS}{dt} = R_g - Q_b - NA_g \tag{1}$$

where GWS denotes volume of water stored in the groundwater compartment, t time, R_g diffuse groundwater recharge from soil, Q_b base flow from groundwater to surface water bodies, and NA_g net groundwater abstraction (groundwater withdrawals minus return flow from irrigation with both surface water and groundwater). Q_b is proportional to GWS and is set to zero if GWS becomes negative. At the start of the

simulation (in 1901), *GWS* is set to zero everywhere, and in case of relatively small NA_g, increases quickly. Negative *GWS* values can only be reached where NA_g is positive and larger than R_g and correspond to the situation where the hydraulic head of the groundwater has fallen below the surface water table. NA_g is negative if due to surface water irrigation, irrigation return flows exceed groundwater withdrawals [*Döll et al*., 2012].

In WaterGAP 2.2a, it is assumed that in semiarid and arid areas, groundwater recharge from surface water bodies (lakes, wetlands, and reservoirs) R_{g_swb} occurs. R_{g_swb} recharges the groundwater storage compartment (red arrows in Figure 1) and

$$\frac{dGWS}{dt} = R_g + R_{g,wb} - Q_b - NA_g \tag{2}$$

Semiarid and arid areas are defined as those areas where mean annual precipitation does not exceed 50% of potential evapotranspiration (except areas north of 60°N). Even though f_{gwr_swb}, defined as R_{g_swb} per unit area of surface water body, is assumed to be constant, the groundwater recharge from surface water bodies in each grid cell varies temporally with the dynamically varying size of the surface water bodies which depends on the water storage in the surface water body. In WaterGAP 2.2a, rivers are assumed to always receive base flow from groundwater, and evaporation from rivers is simulated.

2.1.2. Fractional Routing

In WaterGAP 2.1h and 2.2, base flow Q_b leaving the groundwater storage of the grid cell is discharged first into the lakes, reservoirs, and wetlands of the grid cell (if existing). If "local" lakes exist, for example, the sum of Q_b and R_s flows into the "local" lakes storage compartment of a grid cell, and the total outflow of "local" lakes flows into "local" wetlands [*Döll et al*., 2012, Figure 1]. Only if there are no lakes, reservoirs, and wetlands in a grid cell, Q_b directly discharges into the river storage. With this approach, if surface water bodies recharge groundwater, the recharged groundwater would again recharge the surface water bodies as Q_b, thus creating a short circuit. In addition, if all surface water flows generated within the 0.5° grid cell were added to the local lakes (or local wetlands), too much water would be retained in surface water body storage, resulting in an unrealistically high R_{g_swb}. Therefore, a so-called fractional routing scheme has been implemented in WaterGAP 2.2a. This new scheme for routing water within each grid cell assumes that (1) in arid and semiarid areas, base flow Q_b directly flows into the river and (2) only part of the surface runoff (and the base flow in case of humid conditions) recharges the surface water bodies, while the rest directly recharges the river storage compartment (Figure 1). The fraction reaching the surface water bodies f_{swb} is proportional to the area of all surface water bodies (except rivers) as a fraction of the cell area. In this study, we assume that the drainage basin of the surface water bodies is 20 times larger than the maximum extent of the surface water bodies (but not larger than the cell area).

To conserve mass, land area in percent of cell area is made variable in WaterGAP 2.2a. Evaporation from surface water bodies that varies with water storage is now interpreted as a change in surface water body area which needs to be balanced by a corresponding change of land area. Due to technical constraints, however, this cannot be implemented for global lakes and reservoirs as they may cover more than one grid cell. In addition, the total volume of precipitation over surface water bodies is adjusted to the changed surface water body area.

2.1.3. Modification of the Computation of Net Groundwater and Surface Water Abstractions

In WaterGAP 2.1h, irrigation water use efficiency, the ratio of consumptive use to abstractions, was assumed to be the same for surface water and groundwater irrigation. Country-specific values were compiled from a large number of studies without distinction of the source of water [*Döll et al*., 2012]. However, based on limited data, we suspect that efficiencies are higher in the case of groundwater abstractions. In WaterGAP 2.2 and 2.2a, water use efficiency for groundwater irrigation is set to 0.7 world-wide, a value estimated for India [*Central Ground Water Board (CGWB)*, 2009, p. 18], while the compiled country-specific values are applied for surface water irrigation (e.g., 0.378 for India and 0.5 for United States). This leads to reduced estimates of net and gross groundwater abstractions for irrigation, as consumptive irrigation water use from surface water and from groundwater is computed in WaterGAP and then divided by the efficiencies to calculate water abstractions. With this adjustment and assuming optimal irrigation (see variant IRR100 in Table 1) or deficit irrigation at only 70% of optimal irrigation (see variant IRR70_S in Table 1) throughout the country,

Table 1. Model Variants of WaterGAP 2.2a

Abbreviation of Variant	Groundwater Recharge per Unit Area of Surface Water Body f_{gwr_swb} (mm/d)	Consumptive Water Use *CU* for Irrigation as a Fraction of Optimal Water Use (%)
IRR100	0	100
IRR100_S	10	100
IRR70_S	10	70
NOUSE_S	10	0, and no other water use
IRR70_S3	3	70
IRR70_S1	1	70

we estimate groundwater abstraction in India to have been between 209 and 290 km^3/yr during 2000–2009, which fits well to the independent estimate of 250 km^3/yr in 2010 of *Margat and van der Gun* [2013, Table 5.3]. With the lower (groundwater) irrigation water use efficiencies of WaterGAP 2.1h, a much higher estimate for groundwater abstractions in India of 459 km^3/yr (2000–2009, optimal irrigation) would be computed.

The computation of return flow was modified such that a larger fraction of irrigation return flow reaches the groundwater. In WaterGAP 2.2 and 2.2a, this fraction is computed as 0.95–0.75 f_{drain} (fraction of grid cell that is artificially drained), instead of 0.8–0.6 f_{drain}. For India, for example, it was estimated that 10% of return flow from irrigation contribute to surface water resources [*Gupta*, 2004]. Both adjustments in the computation of net abstractions result in lower NA_g values as compared to WaterGAP 2.1h. They also lead to a smaller increase of streamflow due to return flows in regions with groundwater irrigation, which is more consistent with the fact that even in regions like the HPA with very intensive groundwater irrigation, such increases have not been reported.

2.2. Model Variants

In the presented modeling study, we analyzed various model variants to identify which model variant is best suited to simulate GWD at the global scale. The design of the variants, with different degrees of deficit irrigation and f_{gwr_swb} (groundwater recharge per unit area from surface water bodies in semiarid grid cells), was inspired by information about the HPA. The HPA in central United States can be considered to be the best observed groundwater depletion area world-wide, regarding irrigated area, water use, groundwater recharge, and GWD estimates. In the HPA, like in many semiarid and arid regions, deficit irrigation is likely to occur due to water scarcity at least in the central and southern part, i.e., the volume of water applied is less than the amount that would allow the plants to evapotranspire at the optimal potential evapotranspiration rate. Experts estimate that in the HPA the amount of applied irrigation water is only 70% of optimal (B. Scanlon, personal information, 2011). Therefore, in some variants of WaterGAP we reduced consumptive irrigation water use *CU* (sum of NA_g and NA_s) to 70% of its standard (optimal) value (IRR70 runs in Table 1). "_S" refers to the implementation of groundwater recharge from surface water bodies, with f_{gwr_swb} of either 1, 3, or 10 mm/d (based on *Gurdak and Roe* [2010]; see section 4.1). IRR100 assumes optimal irrigation and no R_{g_swb}. In the NOUSE_S variant, no water use at all is assumed to occur. Table 2 shows global values of groundwater and surface water use 2003–2009 (period for which we evaluated GRACE data of TWS) as computed with consumptive irrigation water use either being optimal or at 70% of optimal. Please note that the actual value of global irrigation water use can be expected to be between the two values. Irrigation is by far the most important water use sector, in particular in terms of *CU*. NA_g is slightly negative in case of optimal irrigation which means that globally averaged there is a net recharge of groundwater due the high return flow from irrigation with surface water. Compared to WaterGAP 2.1h [*Döll et al.*, 2012, Table 1], total water abstractions for irrigation as well as NA_g are strongly reduced in both the IRR70_S and IRR100_S variant.

Table 2. Global Water Use 2003–2009 as Computed by WaterGAP 2.2a (Rounded, Except Net Groundwater Abstractions NA_g and Net Surface Water Abstractions NA_s)[a]

Sector	Total Water Abstractions (km^3/yr)	GW Fraction (%)	CU (km^3/yr)	GW Fraction (%)	NA_g (km^3/yr)	NA_s (km^3/yr)
Irrigation (optimal)	2400	29	1100	42		
Irrigation (70% of opt.)	1700	29	800	42		
Livestock	30	0	30	0		
Domestic	350	36	60	36		
Manufacturing	340	26	140	24		
Thermal power	590	0	16	0		
Total (IRR100_S)	3700	24	1400	39	−37	1422
Total (IRR70_S)	3000	23	1050	38	38	1003

[a]GW: groundwater; CU: consumptive use.

For all model variants the monthly precipitation monitoring product GPCC v6 [*Schneider et al*., 2013] as well as monthly data of temperature, cloud cover, and number of wet days of the CRU TS 3.10 data set [*Harris et al*., 2014] were used. Monthly precipitation was distributed evenly to all wet days in the month, while the other climate variables were interpolated to daily values.

2.3. Comparison to Independent GWD Estimates

Year-to-year change of groundwater storage *GWS* per grid cell was determined by subtracting mean *GWS* in year i from mean *GWS* in year i + 1. Please note that *GWS* is not the total amount of groundwater storage in each grid cell (which is unknown) but a relative value that develops according to equation (2) if groundwater storage is set to zero at the beginning of the model runs in 1901. The trend of *GWS* and thus GWD was determined by averaging the year-to-year changes over the respective years. A positive GWD is equivalent to a decrease in *GWS* over time. Independent estimates of GWD in a number of aquifers and regions world-wide were taken from literature. These estimates were derived by a wide variety of methods, including (1) observations of groundwater well levels combined with estimates of storage capacity, (2) local-scale modeling, and (3) translation of regional GRACE TWS changes to GWD by estimating changes in other storage compartments by models or observations. When *GWS* decreases over only a few years, this may be due to climatic variability and is not necessarily related to a long-term depletion of the groundwater by human water use. For the comparison, however, all decreasing trends of *GWS* are called GWD.

2.4. Comparison to GRACE TWS

Starting in 2003, the GRACE satellites allow the determination of TWS variations with a monthly time resolution and a low spatial resolution of a few 100 km. If TWS as seen by GRACE is continuously declining over a long period of time, GWD can be suspected to occur. However, TWS is not only affected by groundwater storage but also by water storage in snow and ice, soil, lakes, and wetlands. On a short time scale of a few years, declining TWS can be due to GWD but also to climatic drought conditions which lead to a decline of water storage in soil, surface water bodies, and groundwater (e.g., in Australia [*Leblanc et al*., 2009]). On a decadal time scale, a TWS decline can be caused by GWD but also by climate change, with a loss in snow and ice storage or, in case of decreased runoff, loss in water storage in all other compartments. Therefore, a global hydrological model is necessary to interpret GRACE-based TWS declines, by identifying which storage compartment contributes how much to the observed TWS decline.

To assess TWS variations from GRACE satellite observations, monthly gravity field models provided by three different analysis centers were taken into account: GFZ-RL05 [*Dahle et al*., 2013], CSR-RL05 [*Bettadpur*, 2012], and ITG-Grace2010 [*Mayer-Gürr et al*., 2010]. The GRACE solutions are available as spherical harmonic expansion up to a maximum degree of 60 (CSR), 90 (GFZ), and 120 (ITG). To isolate the hydrologic contribution to the TWS changes, background models for ocean, Earth, and pole tides, atmospheric and oceanic mass variations (AOD1B-RL04 for ITG and RL05 for GFZ and CSR [*Flechtner*, 2007]), and glacial isostatic adjustment were reduced either directly during the data processing or in a postprocessing step. As GRACE cannot observe geocenter motion, the coefficients of degree 1 (geocenter motion including trend) were replaced using the time series provided by *Rietbroek et al*. [2012a, 2012b]; degree 2 coefficients, which are strongly influenced by tidal aliasing errors, were substituted according to *Cheng* [2004]. The gravitational potential provided by GRACE was converted to masses in terms of equivalent water heights following *Wahr et al*. [1998]. All three GRACE solutions were smoothed by applying the nonisotropic filter DDK3 [*Kusche*, 2007], the isotropic part of which can be compared to a 240 km Gaussian filter [*Kusche et al*., 2009]. To allow a comparison to WaterGAP TWS, the hydrological model output, converted to a spherical harmonic expansion up to degree of 120, was smoothed using the same procedure. Based on annual mean TWS from both GRACE and WaterGAP, TWS trends were computed like GWS trends. As ITG-Grace2010 data were only available from January 2003 to August 2008, the ITG-Grace2010 trend (2003–2009) was computed by multiplying the ITG-Grace2010 trend (2003–2008) by the GFZ-RL05 trend (2003–2009) to GFZ-R05 trend (2003–2008) ratio.

3. Results

3.1. High Plains Aquifer

GWD simulated for the High Plains Aquifer HPA can be compared to GWD derived from water level observations at more than 9000 wells measured in winter or early spring (here assumed to be in January) when

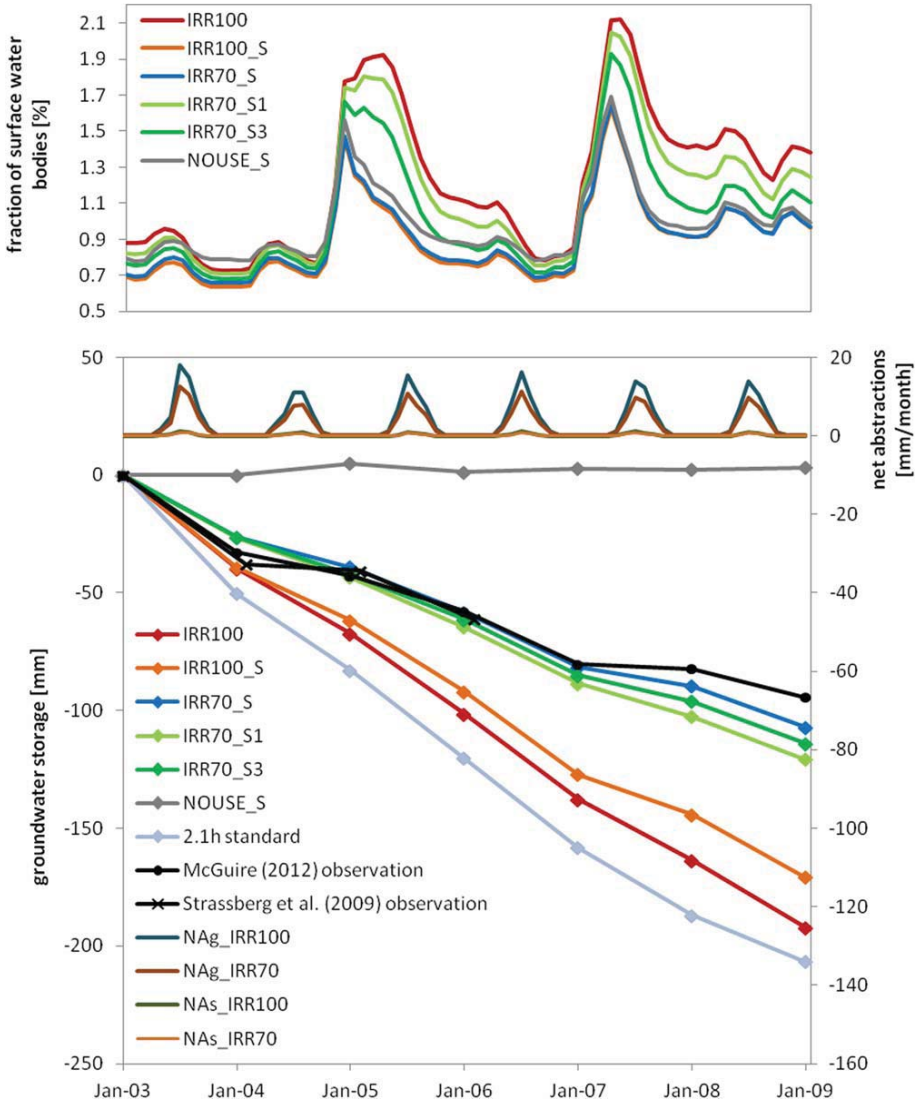


Figure 2. Groundwater depletion (GWD) in the High Plains Aquifer (United States) from January 2003 to January 2009. Comparison of groundwater storage (GWS) as computed by variants of the WaterGAP 2.2a model to *GWS* estimates based on observations wells [*McGuire*, 2011]. *GWS* in January of each year is shown as groundwater levels observed in January before the pumping season were evaluated in *McGuire* [2011]. Observational values of *Strassberg et al.* [2009] refer to January–March periods. "2.1h standard" refers to model version of *Döll et al.* [2012]. In addition, (top) monthly net abstractions NA_g and NA_s as well as monthly areal fractions of surface water bodies are shown.

irrigation wells were not pumping and water levels had recovered from pumping during the previous irrigation season [*McGuire*, 2007, 2009, 2011]. To translate water level changes to TWS changes, McGuire used the area-weighted average-specific yield of 0.15. *Strassberg et al.* [2009] analyzed seasonal groundwater levels at approximately 1000 wells, applying the same average-specific yield. Of all six WaterGAP 2.2 variants (Table 1), IRR70_S, with irrigation at 70% of optimal and a groundwater recharge per unit area of surface water body of 10 mm/d, shows the best fit to observed depletion (Figure 2). All other variants (except NOUSE_S) overestimate observed depletion. While average GWD between January 2003 and 2009 was 16 mm/yr according to observations, IRR70_S results in 18 mm/yr (Table 3).

If the recharge from surface water bodies, here mostly local wetlands, is assumed to be only 3 or 1 mm/d, GWD increases slightly. More runoff is retained in the local wetlands, such that the fraction of surface water

Table 3. Water Balance Components of the High Plains Aquifer (United States) 2003–2008[a]

Water Balance Component	Period	Independent Estimate	IRR100	IRR100_S	IRR70_S
R_g	2003–2008	n.a.	20	20	20
R_{g_swb}	2003–2008	n.a.	0	6	6
NA_g	2003–2008	n.a.	43	43	30
Q_b	2003–2008	n.a.	9	12	14
dGWS	2003–2008	−16[b]	−32	−29	−18
R_g	1971–2000	32[c]	24	24	24
R_{g_swb}	1971–2000	n.a.	0	6	6
GW abstractions	2005	48[d]	63	63	45
Irrigated area	2005	66,379[d]	66,187	66,187	66,187

[a]All values in mm/yr, except irrigated area (in km^2). Modeled change in groundwater storage dGWS is computed from changes of mean annual groundwater storage, with negative values meaning a decrease over time. GW: groundwater.
[b]*McGuire* [2011] (note: dGWS evaluated every January from 2003 to 2009).
[c]*Scanlon et al.* [2012].
[d]*USGS* [2009].

bodies remains higher than in the variants with 10 mm/d (Figure 2, top). In the IRR100 variants, GWD is strongly overestimated with 32 mm/yr in case of IRR100. R_{g_swb} of 10 mm/d decreases GWD by only 3 mm/yr (Table 3). Still, variant IRR100 of WaterGAP 2.2a results in a smaller depletion than WaterGAP 2.1h [*Döll et al.*, 2012, Figure 2] because modeled groundwater abstractions are lower and return flows higher in WaterGAP 2.2 and 2.2a due to changed assumptions on irrigation water use efficiency and return flows (section 2.1.3). The spatial distribution of modeled GWD, with highest values in the southern HPA, fits well to the map of GWD since predevelopment (beginning of significant irrigation) of *McGuire* [2011] (not shown). *Schuh* [2013] showed that the spatial distribution of GWD 1950–2009 computed with IRR70_S fits quite well to the respective map based on well observations by *McGuire* [2011].

GWD does not only depend on NA_g and R_{g_swb}, but also on diffuse groundwater recharge R_g and base flow Q_b. Table 3 (upper part) shows the groundwater balance of the HPA for the years 2003–2008. R_g is the same for all three variants. Q_b increases from variant IRR100 to IRR100_S because of the additional R_{g_swb}, while it increases from variant IRR100_S to IRR70_S because of decreased NA_g. Thus, GWD differences among the three variants are less than would be expected if only NA_g reduction or increase of total groundwater recharge would be considered. Table 3 also shows a comparison of independent estimates of R_g, groundwater abstraction, and irrigated area with the respective WaterGAP values. Average R_g of WaterGAP is 25% lower than the value estimated by *Scanlon et al.* [2012] from groundwater chloride data. R_{g_swb} is calculated as 6 mm/yr averaged over the whole HPA but unfortunately no independent estimate is available. Groundwater abstraction in case of IRR70_S is close to the *USGS* [2009] estimate for 2005, as is the total irrigated area in the HPA (Table 3). From this comparison, we conclude that variant IRR70_S is suited best to simulate GWD in the HPA.

3.2. Comparison of Simulated GWD to Independent GWD Estimates for Selected Regions Around the World

To evaluate GWD as simulated by WaterGAP 2.2a variants, we compared simulated GWD to independently estimated values of GWD for selected regions. Where possible, comparison was done against carefully estimated GWD during a 40–50 year long period at the end of the 20th century, and during a period after 2000 (Table 4). Outlines of the comparison regions can be seen in Figure 3.

As IRR70_S was designed to fit to HPA conditions for 2003–2008, it is not surprising that it fits well to independent GWD estimates for 2000–2009 and reasonably well for 1950–2000. GWD in the Central Valley is best simulated with IRR100 for the period 1961–2000, but strongly underestimated for 2001–2009. This can be explained by the fact that WaterGAP assumes a temporally constant fraction of consumptive water use to be taken from groundwater. This assumption is appropriate if farmers are equipped with either surface water or groundwater irrigation infrastructure. In the Central Valley, however, farmers have the infrastructure to switch, in case of surface water scarcity, to groundwater. This was done from 2006 onward when drought conditions prevailed in the Central Valley [*Famiglietti et al.*, 2011]. GWD in the Western Volcanic Systems of the United States is underestimated, while the stronger GWD in the Western Alluvial Basins is captured quite well at least for 1960–2000. The small GWD in the deep confined bedrock aquifers of the United

Table 4. Comparison of Groundwater Depletion (GWD) Computed by Three WaterGAP 2.2a Variants to Independent Estimates for Selected Regions (See Figure 3)[a]

Study Region	Area (1000 km^2)	Period	Independently Estimated GWD (mm/yr (km^3/yr))	Modeled GWD (mm/yr (km^3/yr))		
				IRR100	IRR100_S	IRR70_S
United States						
High Plains aquifer	488	1950–2000[1]	10 (5) W	25 (12)	23 (11)	15 (7)
		2000–2009[1,2]	21 (10) W	35 (17)	32 (15)	20 (10)
Central Valley aquifer	53	1961–2000[3]	26 (1) M	23 (1)	20 (1)	14 (0.8)
		2001–2009[4,5]	73 (4) M, G	31 (2)	27 (1)	21 (1)
Western volcanic systems	172	1961–2000[4]	0.5 (0.1) O	0.6 (0.1)	0.1 (0)	0.1 (0)
		2001–2008[4]	2 (0.4) O	1 (0.2)	0.3 (0.1)	0.3 (0)
Western alluvial basins	480	1961–2000[4]	5 (2) O	4 (2)	3 (2)	3 (1)
		2001–2008[4]	0.6 (0.3) O	6 (3)	6 (3)	5 (3)
Deep confined bedrock aquifers	521	1961–2000[4]	0.6 (0.3) O	0.6 (0.3)	0.4 (0.2)	0.4 (0.2)
		2001–2008[4]	0.6 (0.3) O	0.7 (0.4)	0.7 (0.4)	0.7 (0.4)
Gulf coastal plain	518	1961–2000[4]	8 (4) O	1 (0.5)	1 (0.5)	0.8 (0.4)
		2001–2008[4]	16 (8) O	0 (0)	−0.1 (−0.1)	−0.2 (−0.1)
Atlantic coastal plain	265	1961–2000[4]	0.7 (0.2) O	4 (1)	4 (1)	4 (1)
		2001–2008[4]	1 (0.3) O	4 (1)	4 (1)	4 (1)
Mississippi basin	3418	2002–2005[6]	7 (24) W	4 (13)	3 (12)	2 (8)
Northern India and Adjacent Areas						
Three-states region	410	2002–2008[7]	43 ± 11 (18 ± 5) G	75 (31)	67 (27)	45 (18)
Tiwari study region	2070	2002–2008[8]	26 ± 4 (54 ± 9) G	25 (53)	23 (47)	14 (29)
Jacob study region	1843	2003–2009[9]	19 (35) G	28 (52)	25 (46)	16 (30)
Bangladesh	130	1985–2007[10]	2 (0.2) W	17 (2)	17 (2)	6 (0.8)
		2003–2006[10]	10 ± 2 W/5–31 G (1 ± 0.2 W/0.6–4 G)	31 (4)	31 (4)	15 (2)
		2003–2007[10]	7 ± 1 W/3–19 G (0.9 ± 0.2 W/0.4–3 G)	31 (4)	31 (4)	15 (2)
Irrigated Indus basin	350	2007[11]	89 (31) M	60 (21)	55 (19)	39 (14)
Northeastern China						
North China Plain	140	2000–2008[12]	27 (4) M	155 (22)	152 (21)	127 (18)
Hai river basin	330	1958–1998[13]	7 (2) W	33 (11)	32 (11)	24 (8)
		1999–2006[13]	13 (4) W	73 (24)	71 (23)	58 (19)
		2003–2006[14]	23 W/12 G (8 W/4 G)	73 (24)	71 (23)	59 (20)
		2003–2010[15]	25 ± 3 (8 ± 1) G	83 (27)	81 (27)	67 (22)
Middle East and North Africa						
Saudi Arabia	1725	1980–2000[16]	6 (10) O	6 (10)	6 (10)	4 (7)
TEWI	740	2003–2006[17]	−5 ± 3 (−4 ± 2) G	10 (8)	10 (7)	7 (5)
		2007–2009[17]	35 ± 5 (26 ± 3) G	13 (10)	14 (10)	10 (8)
Nubian sandstone aquifer system	1965	1960–2000[18]	0.4 (0.8) M	0.8 (2)	0.8 (2)	0.5 (1)
North Western Sahara aquifer system	918	1960–2000[19]	2 (1) O	0.6 (0.6)	0.4 (0.3)	0.2 (0.2)

[a]GWD is listed both in mm/yr and km^3/yr (in parentheses). In the former case, the area of the grid cells used for determining the WaterGAP estimates (column 2) was also applied to the independent estimates. The letters in the column "independently estimated GWD" refer to the following. W: groundwater well observations; M: modeling; G: derived from GRACE TWS data; O: other methods. The Tiwari and Jacob study areas are essentially the same and cover Northern India and adjacent areas. TEWI: Tigris-Euphrates plus Western Iran. References: 1, *McGuire et al.* [2003]; 2, *McGuire* [2011]; 3, *Faunt* [2009] cited in *Konikow* [2011]; 4, *Konikow* [2011]; 5, *Famiglietti et al.* [2011]; 6, *Rodell et al.* [2007]; 7, *Rodell et al.* [2009]; 8, *Tiwari et al.* [2009]; 9, *Jacob et al.* [2012]; 10, *Shamsudduha et al.* [2012]; 11, *Cheema et al.* [2014]; 12, *Liu et al.* [2011]; 13, *Liu et al.* [2010]; 14, *Moiwo et al.* [2009]; 15, *Feng et al.* [2013]; 16, *Abderrahman* [2006] in *Foster and Loucks* [2006]; 17, *Voss et al.* [2013]; 18, *Bakhbakhi* [2002] cited in Table 6.3 of *Margat and van der Gun* [2013]; 19, *Mamou et al.* [2006].

States is simulated well but the temporal development is not (not shown). The strong GWD in the Gulf Coastal Plain, which mainly occurs in the Mississippi Embayment Aquifer System, is significantly underestimated by WaterGAP because WaterGAP likely overestimates groundwater recharge R_g there. In the Atlantic Coastal Plain, all variants of WaterGAP strongly overestimate depletion. The independent GWD estimate for the large Mississippi basin is based on water levels in only 58 monitoring wells and can therefore be regarded to be quite uncertain, and likely an overestimation given the more reliable data of *Konikow* [2011]. When summing up the GWD estimates of the US depletion regions for the first decade of the 21st century (except for Mississippi basin) and comparing it to the IRR70_S value, the independent estimate is approximately 50% higher, while for the second half of the 20th century, the independent estimate is approximately 10% lower than the IRR70_S value.

The largest GWD world-wide has happened in the three-states region in Northern India (Figure 3), for which *Rodell et al.* [2007] derived GRACE-based estimates of GWD. Clearly, the results of IRR70_S fit very well to this estimate. With IRR100, GWD would be overestimated by two thirds. Taking into account groundwater

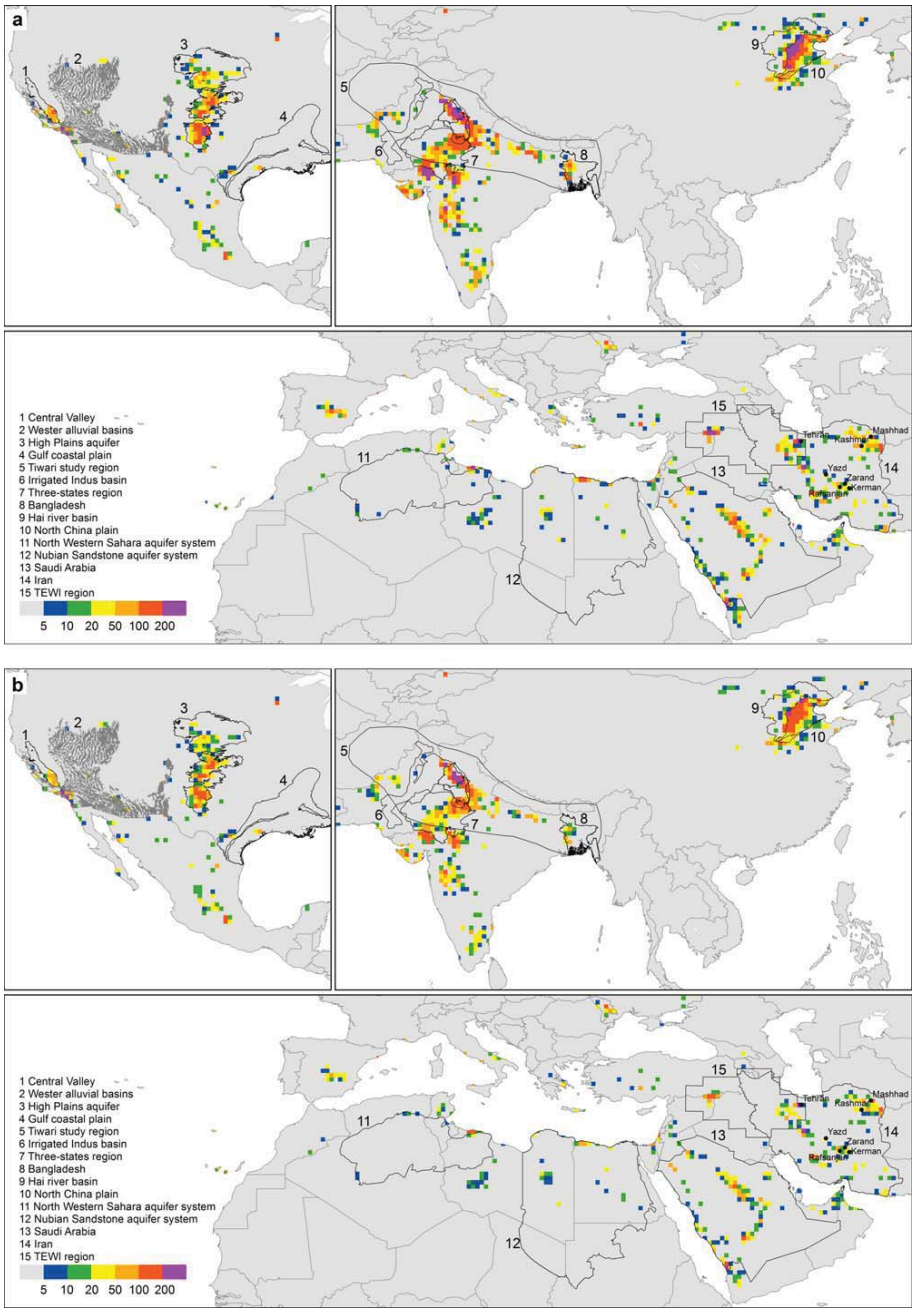


Figure 3. Groundwater depletion 1980–2009 computed as the difference between (a) IRR100_S and NOUSE_S and (b) IRR70_S and NOUSE_S (mm/yr). Regions for which independent GWD estimates are available are delineated.

recharge from surface water bodies, in IRR100_S, would still result in an overestimation by one half. There are two independent GRACE-based GWD estimates for the larger "Tiwari study region" and for the very similar "Jacob study region" which stretch from Afghanistan to Bangladesh and encompass the three-states region (Figure 3). For a comparable time period, the estimated average annual GWD are 54 [*Tiwari et al.*, 2009] and 35 km^3/yr [*Jacob et al.*, 2012], respectively. While the latter value is similar to the simulated IRR70_S value, the former is similar to the IRR100 value. However, when looking at only one of four subregions of *Tiwari et al.* [2009], at subregion b which approximately covers the Indus basin, estimated GWD is

10 km^3/yr while the IRR70_S value is 13 km^3/yr (not shown in Table 4). The independent GWD estimate for the irrigated part of the Indus basin during the wet year 2007 computed by *Cheema et al*. [2014] is higher than the highest WaterGAP estimate obtained by IRR100 (Table 4). However, in our opinion, the value obtained by *Cheema et al*. [2014] strongly overestimates actual GWD. With 31 km^3/yr, it is 3 times larger than the GWD estimate of 10 km^3/yr of *Tiwari et al*. [2009]. The high value of GWD is caused by assuming, in the model of *Cheema et al*. [2014], a base flow of 20 km^3/yr. If base flow were zero, GWD would be reduced to 11 km^3/yr, which is similar to the value of IRR70_S and of *Tiwari et al*. [2009]. As GWD computed by *Cheema et al*. [2014] would correspond to a water table decline of more than 1 m/yr in most areas within the Indus basin, it is unlikely that there could be any significant base flow under these conditions, i.e., the model assumptions made by *Cheema et al*. [2014] seem implausible. In Bangladesh, IRR70_S falls into the very large observational range for the time period after 2000, better than IRR100, while long-term GWD appears to be overestimated least with IRR70_S (Table 4).

WaterGAP strongly overestimates a number of independent estimates of GWD in Northeastern China, by approximately a factor of 4 (Table 4). This is mainly caused by an underestimation of R_g, which according to *Liu et al*. [2011] is 92 mm/yr in the North China Plain while WaterGAP computes only 12 mm/yr. GWD in Saudi Arabia is equal to the independent estimate if optimal irrigation is assumed (Table 4), while GWD in the Nubian Sandstone Aquifer System fits best to the independent estimate in case of deficit irrigation. Even assuming optimal irrigation, the independent estimate of GWD in the North Western Sahara Aquifer System is underestimated by WaterGAP (Table 4). The strong increase of GWD after 2007 that is estimated for the Tigris-Euphrates-Western-Iran (TEWI) region using GRACE TWS (Table 4) is assumed to be caused by intensive well construction in Iraq due to a strong drought and thus decreased surface water availability in 2007 [*Voss et al*., 2013]. However, *Voss et al*. [2013] stated that they overestimate GWD after 2007 due to underestimating storage decline in surface water bodies. The strong GRACE-based decline of groundwater storage 2007–2009 is not represented well by WaterGAP, possibly because a constant fraction of groundwater use is assumed in the model.

Assuming that consumptive irrigation water use is approximately equal to NA_g in GWD areas, GWD that would be caused by other degrees of deficit irrigation can roughly be estimated from the difference DIFF70 between GWD for optimal irrigation and GWD for 70% deficit irrigation (Table 4). For irrigation at only x percent of optimal, the difference between GWD for optimal irrigation and GWD for x percent deficit irrigation can be approximated as DIFF70·(100-x)/30. Comparing the independent GWD estimates to the GWD values computed with the three WaterGAP variants (Table 4) and to rough estimates of GWD derived for other degrees of deficit irrigation with the approximation above, it appears that WaterGAP 2.2a variant IRR70_S fits mostly best to reliable independent estimates where WaterGAP is at all able to compute GWD reasonably. No conclusions about deficit irrigation can be drawn where R_g of WaterGAP 2.2a is implausible, or where, in contrast to the model assumption, the fraction of groundwater use varies strongly over time (like in the Central Valley and TEWI).

3.3. Simulated GWD at the Global Scale

GWD is generally considered to be driven by groundwater abstractions. However, groundwater storage can also decline over a certain period if groundwater recharge is decreasing due to climate variability. To isolate the impact of groundwater abstractions on groundwater storage, we compute GWD (as shown in the following maps and tables) as the difference between GWD in model variants with abstraction (IRR100_S, IRR70_S) minus GWD in the model variant without any water use NOUSE_S. Figure 3 shows average annual GWD during the time period 1980–2009 as computed by WaterGAP 2.2a, for the two model variants IRR100_S and IRR70_S. In those areas of the globe not shown in Figure 3, about 20 grid cells display GWD values of more than 5 mm/yr, many of them in grid cells where large cities are located. Results of variant IRR100, i.e., without additional groundwater recharge from surface water bodies in dry areas, look rather similar to IRR100_S results and are therefore not shown.

Clearly, GWD with annual rates of at least 5 mm/yr is concentrated in few semiarid and arid regions around the globe (all nongray grid cells in Figure 3). The most intensive GWD can be found in those regions that are well known for GWD and are listed in Table 4. The following additional GWD areas can be identified from Figure 3: large parts of central and southern India, large parts of Iran, southeastern part of Spain, parts of Libya and Tunisia, parts of Turkey and Syria, and parts of northern and central Mexico, and of the

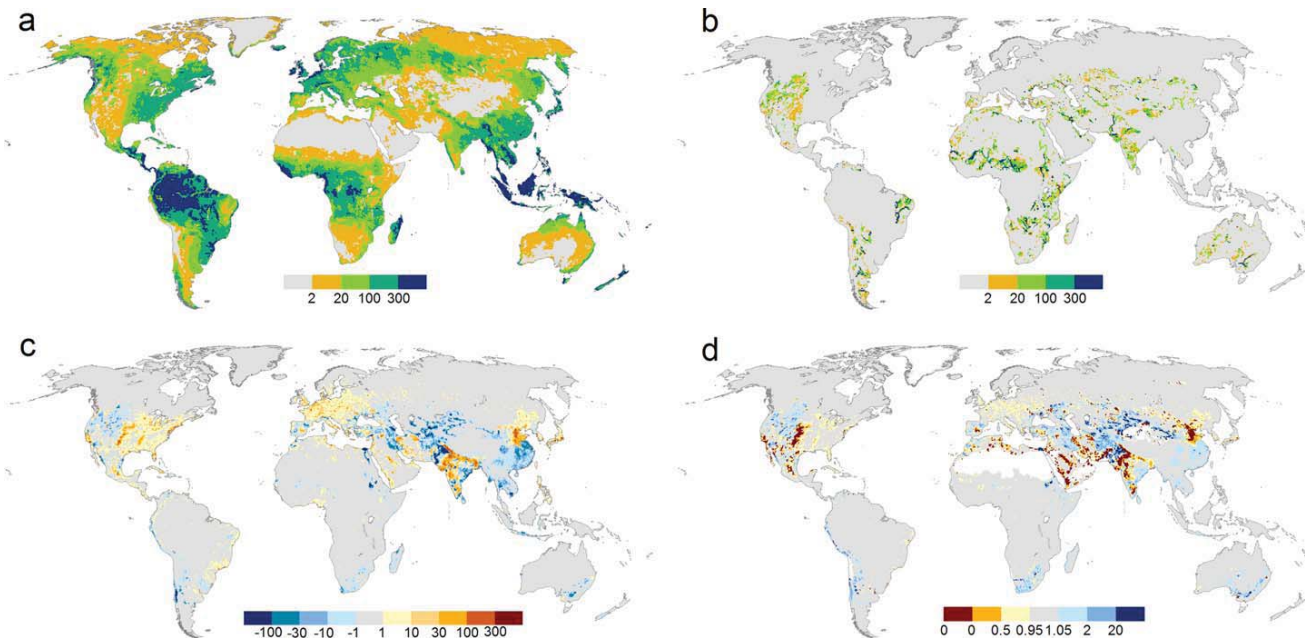


Figure 4. Water flows determining groundwater storage changes, for 1980–2009, for IRR70_S, in mm/yr. (a) Diffuse groundwater recharge R_g, (b) groundwater recharge from surface water bodies R_{g_swb}, (c) net abstraction from groundwater NA_g, (d) base flow as a fraction of total natural groundwater recharge ($R_g + R_{g\_swb}$). White areas in Figure 4d indicate grid cells where total groundwater recharge is zero.

Californian coast. Groundwater depletion in the southeastern part of Spain as well as on the Canary Islands, in northwestern Mexico (Hermosillo) and in southern Libya (Murzuk), is confirmed by *Custodio* [2010]. According to our computations, average annual GWD during 1980–2009 has exceeded 50 mm/yr in many cells even with deficit irrigation (orange, red, and purple grid cells in Figure 3b). Such groundwater storage losses could correspond to groundwater table declines of more than 0.5 m/yr averaged over the whole grid cell (approximately 2500 km^2). Please note that GWD shown in Northeastern China overestimates actual GWD by a factor of about 4 (compare Table 4).

To understand the spatial pattern of GWD, its drivers, i.e., the flows into and out of the groundwater storage compartment, are shown in Figure 4. In GWD areas, R_g is generally less than 100 mm/yr. A comparison of R_g (Figure 4a) with R_{g_swb} (Figure 4b) shows that the latter dominates total groundwater recharge in parts of Northeastern Brazil, the Sahel region, South Africa, Pakistan, India, China, Central Asia, and Australia. Irrigation with surface water causes artificial groundwater recharge and balances groundwater depletion (grid cells shown in blue colors in Figure 4c). Only grid cells shown in brown colors are subject to a positive net groundwater abstraction. The impact of human water use on ecologically relevant base flow is shown in Figure 4d. In WaterGAP 2.2a, positive NA_g values (Figure 4c) lead to reduced base flow to surface water bodies as compared to the sum of diffuse groundwater recharge and recharge from surface water bodies (Figure 4d), reaching a reduction to zero in GWD areas. Negative NA_g values, indicating groundwater recharge due to irrigation with surface water, lead to base flow that is larger than groundwater recharge from soil by precipitation (and not irrigation) and from surface water bodies.

3.4. Applicability of Global-Scale GRACE Estimates of Total Water Storage (TWS) Variations to Monitor GWD

Figure 5 compares *GWS* and TWS trends computed by WaterGAP 2.2a (IRR70_S). GWD areas (red in Figure 5a) clearly correspond to areas with negative TWS trend (Figure 5b). However, there are additional areas with strong decreasing TWS trends of more than 20 mm/yr (e.g., in Florida and the southeastern South America) which represent climate-related decreases of surface water storage. In the green grid cells in Figure 5c, the TWS trend is almost equal to the *GWS* trend. These areas correspond well with the GWD areas seen in Figure 5a, in particular the intensively exploited HPA (United States), three-states region (Northern India), Hai river basin (Northeastern China), and Saudi Arabia. GRACE data, however, are not available with a

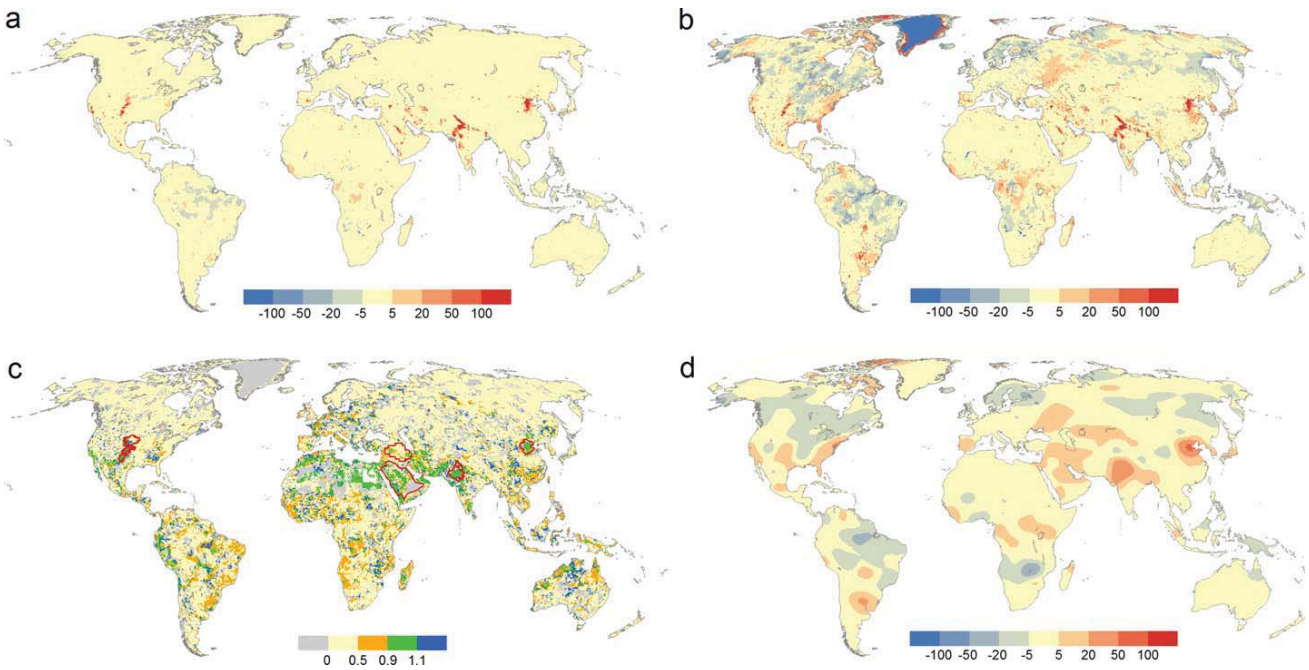


Figure 5. (a) GWS trend IRR70_S (mm/yr), (b) TWS trend IRR70_S (mm/yr), (c) GWS trend as a fraction of TWS trend, (d) TWS trend IRR70_S filtered (mm/yr), for 2003–2009. Positive trends indicate a decrease of GWS or TWS over time. Red polygons in Figure 5c show the main GWD areas/countries HPA (United States), Saudi Arabia, TEWI region, three-states region (Northern India), and Hai river basin (Northeastern China).

spatial resolution of 0.5°. When TWS computed by WaterGAP with a spatial resolution of 0.5° is filtered with the DDK3 filter used for GRACE data (see section 2.4), the pattern of TWS trend is significantly smoothed, and the localization of GWD areas is hampered (compare Figure 5d to Figure 5b). In all light and dark orange grid cells shown in Figure 5c, the *GWS* trend is smaller than the TWS trend. Where water storage trends are not caused by overabstraction of groundwater but climatic variations, trends in the compartments snow, soil, and surface water bodies can be as large as or larger than trends in groundwater storage. This shows that a TWS trend seen by GRACE should not be interpreted as being mainly caused by a *GWS* trend.

3.5. Comparison of Modeled Trends of TWS to GRACE Observations

The ensemble mean GRACE TWS trend during 2003–2009 (Figure 6a) can be compared to the filtered WaterGAP trend (variant IRR70_S) (Figure 5d). Over most of the globe, the sign of the trends agree between WaterGAP and GRACE. This can also be verified in Figure 6b where areas with different signs are shown in light yellow. For example, in the United States, decreasing trends are identified by both for eastern, central (HPA), and western United States. The trend in the eastern region is, according to WaterGAP (Figure 5a), not related to GWD. In the HPA, both GRACE and WaterGAP show an increasing TWS trend in the northern part and a decreasing trend in the central and southern part. Further coinciding negative trends can be found in Spain, Central Asia (from the Caspian Sea to Northwestern China), Iraq, and Iran, and also in the three other main GWD areas world-wide (Saudi Arabia, three-states region in Northern India, and Hai river basin in Northeastern China). In addition, there is a clearly coinciding negative trend in Argentina that is not related to GWD but to meteorological drought. The strong negative GRACE trend in Greenland is not reflected by WaterGAP as the glacier and ice sheet dynamics are not simulated by WaterGAP and climate observations are poor.

For 27% of the grid cells (outside Greenland), WaterGAP IRR70_S TWS trends are within the range of the three GRACE TWS trends from the different analysis centers (green in Figure 6b). Simulated trends are mostly smaller than the GRACE ensemble mean trend (in particular in Australia). This, however, is not the case in the main GWD areas except TEWI (Table 5). While two of the three GRACE trends show an increase of TWS for the HPA, the ITG-Grace2010 trend is very close to the WaterGAP IRR70_S TWS trend, i.e., the

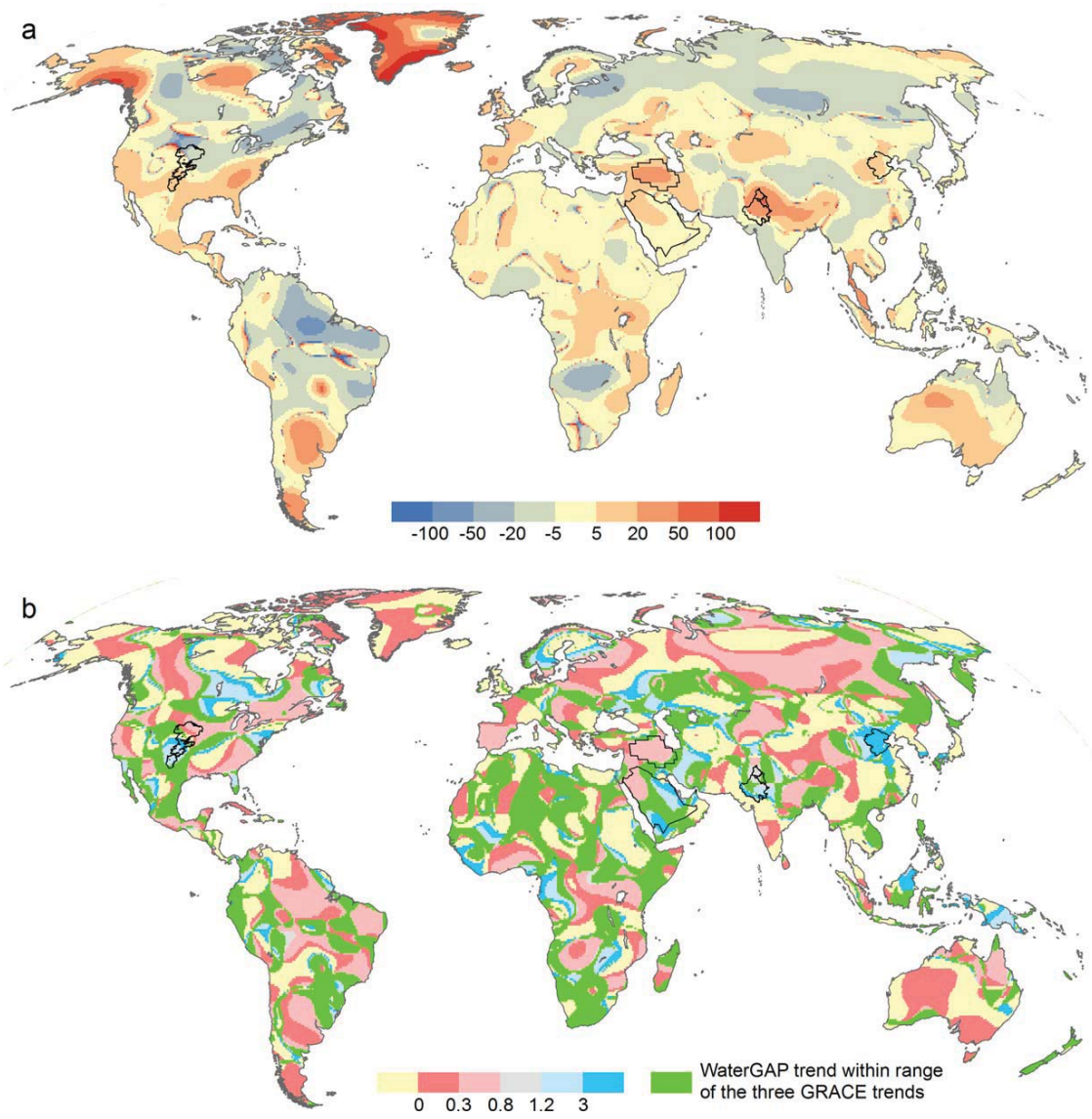


Figure 6. (a) Ensemble mean TWS trend (mm/yr) from GRACE (GFZ-RL05, CSR-RL05, and ITG-Grace2010) 2003–2009, and (b) ratio of IRR70_S trend and ensemble mean GRACE trend. In Figure 6a, a positive trend indicates a decrease of TWS over time. In Figure 6b, areas where GRACE and WaterGAP trends have the opposite sign are shown in yellow. In black, outlines of the main GWD areas/countries HPA (United States), Saudi Arabia, TEWI region, three-states region (Northern India), and Hai river basin (Northeastern China) are shown.

Table 5. Comparison of Trends in Simulated TWS in Main GWD Areas/Countries to Three Different (Filtered) GRACE Solutions, for 2003–2009 (mm/yr)[a]

	WaterGAP IRR70_S Unfiltered	WaterGAP IRR70_S Filtered	WaterGAP IRR100_S Filtered	GRACE GFZ-RL05	GRACE CSR-RL05	GRACE ITG-Grace2010
High Plains aquifer	11.9	4.5	9.1	−0.2	−3.5	4.9
Hai river basin (Northeast China)	60.4	42.2	49.2	6.5	7.1	5.9
Three-states region (Northern India)	44.6	32.5	46.9	24.1	25.4	29.4
Tiwari study region	16.5	17.4	25.1	17.9	16.2	17.3
Saudi Arabia	5.3	5.5	7.2	6.2	3.2	−7.5
TEWI (2003–2009)	12.9	11.8	14.1	22.3	19.8	18.9
TEWI (2003–2006)	6.6	7.7	10.2	10.8	−3.8	−2.5
TEWI (2007–2009)	22.0	18.8	21.3	40.6	45.3	n.a.

[a]Positive trends indicate a decrease of TWS over time. Location of the areas is shown in Figure 5c.

Table 6. Average Annual Groundwater Depletion (GWD) and Its Relation to Total Groundwater Abstractions in Countries Modeled by WaterGAP 2.2a (GWD IRR70_S Minus GWD NOUSE_S, Without Grid Cells With Big Cities in which Irrigation Water Withdrawals Are Less than 5% of Total Water Withdrawals)[a]

Country	GWD 1960–2000 (km^3/yr)	GWD/Groundwater Abstractions 1960–2000	GWD 2000–2009 (km^3/yr)	GWD/Groundwater Abstractions 2000–2009
Afghanistan	0.14	0.04	0.28	0.09
Algeria	0.16	0.11	0.32	0.12
Argentina	0.11	0.04	0.06	0.02
Australia	0.09	0.04	0.22	0.06
Bangladesh	0.24	0.03	1.92	0.10
China	2.42	0.04	6.27	0.07
Egypt	1.07	0.29	1.45	0.30
Greece	0.08	0.04	0.29	0.08
India	21.70	0.16	43.13	0.21
Iran	4.81	0.20	9.98	0.29
Iraq	0.09	0.08	0.36	0.21
Israel	0.11	0.23	0.18	0.31
Italy	0.12	0.02	0.12	0.02
Jordan	0.06	0.10	0.19	0.20
Kazakhstan	0.08	0.07	0.07	0.06
Kyrgyzstan	0.39	0.23	−0.21	−0.21
Libya	0.98	0.52	2.33	0.59
Mexico	2.02	0.23	3.21	0.24
Moldova	0.60	0.29	−0.05	−0.08
Morocco	0.13	0.03	0.17	0.03
Oman	0.21	0.69	0.37	0.71
Pakistan	2.44	0.07	3.61	0.08
Romania	0.23	0.12	−0.10	−0.09
Saudi Arabia	4.30	0.64	9.04	0.68
Spain	0.51	0.09	1.17	0.13
Syria	0.79	0.19	1.92	0.25
Tunisia	0.12	0.12	0.29	0.15
Turkey	0.20	0.02	0.49	0.04
Ukraine	0.44	0.09	−0.14	−0.08
United Arab Emirates	0.21	0.58	0.99	0.62
United States of America	10.66	0.17	23.47	0.19
Yemen	0.37	0.27	0.90	0.35
Other countries	0.17	0.00	0.25	0.00
Global	56	0.12	113	0.17

[a]Values for China were obtained by multiplying the WaterGAP estimate for the Hai river basin by 0.25, to account for the obvious overestimation of WaterGAP compared to independent estimates (Table 4). Values for the United States were obtained by multiplying the WaterGAP value by 0.9 for 1960–2000 and by 1.5 for 2000–2009, based on a comparison to independent estimates for individual GWD regions in the United States (Table 4 except *Rodell et al.* [2009]). Only countries with GWD larger than 0.06 km^3/yr for 1960–2000 are listed. Negative values for the period 2000–2009 indicate an increase of groundwater storage due to reduced groundwater abstractions.

trend of the WaterGAP version that fits very well to GWD as observed from groundwater well observations. With optimal irrigation (IRR100_S), the TWS decrease would be higher than estimated by any of the three GRACE solutions. Thus, for the HPA, WaterGAP IRR70_S total and groundwater storage trends results are supported by both GRACE data and groundwater well observations. Regarding the Hai river basin, the extreme overestimation of independently estimated GWD (Table 4; derived only partly from GRACE data) by all variants of WaterGAP is confirmed by GRACE TWS trends (Table 5). For the globally most intensive GWD in the three-states region, GRACE TWS trends confirm, as does the GRACE-based GWD estimate of *Rodell et al.* [2009] (Table 4) that deficit irrigation at 70% of optimal irrigation is consistent with GRACE observations. This is also true for the larger Tiwari study region (Figure 3). This supports that GWD in this region is quantified well by both WaterGAP and *Jacob et al.* [2012], but overestimated by *Tiwari et al.* [2009] (Table 4). In case of Saudi Arabia, the IRR70_S TWS trend falls within the range of the three GRACE solutions, while IRR100_S overestimates the decline of GRACE TWS. This differs from the result of comparison to GWD independent estimates where IRR100_S resulted in a better fit, albeit for the earlier time period 1980–2000. In the TEWI region, WaterGAP TWS decrease is within the range of the three GRACE solutions for the period 2003–2006, but underestimates the 2007–2009 GRACE trends by more than 50%, which could be explained by a massive construction of groundwater wells in Iraq in the drought year 2007 (see section 3.2).

3.6. Best Estimate of Global Groundwater Depletion

Based on the comparison of GWD and TWS trends simulated by WaterGAP 2.2a variants with a large range of independent estimates from the literature, but also taking into account the plausibility of water fluxes constraining GWD (sections 3.1–3.5), we conclude that it is likely that farmers perform deficit irrigation in GWD regions. The deficit irrigation model variant IRR70_S, with only 70% of optimal irrigation requirement, leads to declining groundwater and total water storages that correspond reasonably well to independent estimates in many regions world-wide. To derive a best estimate of global groundwater depletion and depletion in individual countries due to human water use, we compute GWD, like for Figure 3, as the difference between GWD of IRR70_S minus GWD of NOUSE_S (without any water abstractions) but do not consider GWD in those cells where big cities are located (ESRI shapefile) and water abstraction for irrigation accounted for less than $<$5% of total water abstraction during the time period 1960–2000. This was done to correct for the erroneous model assumption that the complete groundwater demand also of big cities is supplied from the grid cell in which the city is located. Finally, the country values for China and the United States were corrected based on independent GWD estimates that are very likely to be more accurate than WaterGAP results (Table 4). In the case of China, the GWD simulated in the Hai river basin was multiplied by 0.25, while the GWD values for the United States were multiplied by 0.9 for the time period 1960–2000 and by 1.5 for the period 2000–2009 (compare section 3.2).

Table 6 provides average annual GWD for those 32 countries for which GWD in the period 1960–2000 was computed to exceed 0.06 km^3/yr. Average annual GWD has increased from 1960–2000 to 2000–2009 in all countries except Italy and five countries which all have been subject to a strong decrease of irrigated areas after the end of the Soviet Union. In each of the five countries with the largest GWD, i.e., in India, United States, Iran, Saudi Arabia, and China (in decreasing order), GWD has more than doubled from 1960–2000 to 2000–2009. GWD in Bangladesh, Greece, Iraq, Jordan, and United Arab Emirates has even more than tripled according to our computations. GWD as a fraction of groundwater abstractions shows to what extent a country depends on nonrenewable groundwater for its groundwater supply. These fractions have also increased over time but not as strongly as GWD since renewable groundwater abstractions have increased, too (Table 6). The fraction reaches a maximum of 0.68 for Saudi Arabia, which corresponds quite well to the estimate that "nonrenewable groundwater resources supplied about 66% of the total national needs in 2000" [*Abderrahman*, 2006].

Global GWD is estimated to have doubled from 56 km^3/yr during the period 1960–2000 to 113 km^3/yr during 2000–2009. In the two time periods, 10–12% and 13–17%, respectively, of the globally abstracted groundwater was taken from nonrenewable groundwater; the lower values would result under the assumption that irrigation is done at the optimal rate everywhere, while the upper values would be obtained in case of 70% deficit irrigation world-wide. Groundwater depletion outside the 32 countries was very small (Table 6). However, in four countries with relatively small groundwater abstractions, the relative reliance on nonrenewable groundwater is high. In Mongolia, Mozambique, Qatar, and Mali, 30–74% of the abstracted groundwater was nonrenewable in 2000–2009, increasing from the earlier period.

4. Discussion

4.1. Global-Scale Modeling of Groundwater Recharge From Surface Water Bodies

For this study of global GWD, we selected a simple approach for simulating the fact that in semiarid and arid areas, surface water bodies tend to recharge the groundwater while in humid areas, groundwater mostly recharges surface water bodies via base flow. With this approach, it is not possible to simulate the temporal dynamics of water exchange between surface water bodies and groundwater that are due to changing hydraulic gradients. Only the dynamics due to changing areas of surface water bodies are modeled. Different from representation in our model, there are no distinct boundaries between areas of losing and gaining surface water bodies. For example, we found that due to our delineation of semiarid and arid areas, the Sandhill wetlands in the northern HPA, north of the playa regions, are modeled as losing surface water bodies while in reality they are recharged by groundwater [*LaGrange*, 2005].

Unfortunately, it was not possible to constrain the value of the parameter f_{gwr_swb} by modeling GWD in the HPA. While the combination of 70% deficit irrigation with groundwater recharge per unit surface water area $f_{gwr\_swb} = 10$ mm/d yielded the best fit of all variants (Figure 2), the decrease from 100% to 70% of optimal irrigation lowered GWD by 11 mm/yr, while assuming groundwater recharge from surface water bodies but

optimal irrigation lowered GWD by only 3 mm/d. So at least for the HPA, with a strong groundwater overdraft in some southern and central grid cells, GWD is much more sensitive to the degree of deficit irrigation than to the assumed f_{gwr_swb}. *Gurdak and Roe* [2010] compiled a large number of observed groundwater recharge estimates from the playas of the southern HPA. They cite infiltration rates beneath playas f_{gwr_swb} ranging from 1.5 to 43 mm/d. Besides, they point out that infiltration is highest immediately after filling of the seasonal playas and that a lack of evaporite minerals in the playas indicates a significant infiltration.

Groundwater recharge from large lakes or reservoirs in dry areas can be expected to be high. However, due to the permanent existence of the surface water body, hydraulic gradients may be lower. Various studies on groundwater recharge from Lake Nasser identified a significant recharge of the underlying groundwater from Lake Nasser, but quantitative estimates vary extremely. When steady state conditions of groundwater recharge are assumed, observed hydraulic gradients in the adjacent aquifer and estimates of hydraulic conductivity lead to very small values of groundwater recharge per unit of lake area of 5–50 mm/yr [*Metwaly et al*., 2006; *Ghoubachi*, 2012]. Dynamic groundwater modeling showed that recharge rates decrease for decades after the reservoir was filled. For 1970–2000, groundwater recharge from Lake Nasser is computed to be 1000 mm/yr (or 3 mm/d) and will decrease to 14% of this value for the period 2000–2030 [*Kim and Sultan*, 2002]. Based on these three studies, we conclude that we possibly overestimate groundwater recharge from permanent water bodies if we assume a daily groundwater recharge of 10 mm.

4.2. Monitoring GWD Using GRACE

When assessing GWD using GRACE gravity changes in regions with oil production, the amount of extracted oil must be subtracted from total continental mass loss to obtain TWS changes. However, we found that even on the Arabian Peninsula, oil production is small compared to consumptive irrigation water use. Based on oil production data from *BP* [2009], we estimated that in Saudi Arabia and in Oman, oil production in 2002–2008 was about 4% of consumptive water use for irrigation. Only in the small countries Qatar and Kuwait, both fluxes are of similar magnitude.

GRACE TWS are certainly very useful for monitoring GWD in particular where the dynamics of groundwater and surface water withdrawals are high and not well monitored. To derive GWD, GRACE TWS should not only be combined with model-based estimates of soil and snow storage as it has been done in most studies [e.g., *Tiwari et al*., 2009; *Jacob et al*., 2012] but should be jointly analyzed together with models like WaterGAP that simulate surface water and groundwater storage variations as impacted by humans as well. Then, GRACE data can contribute better to an improved understanding of GWD and its causes, and climatic reasons for groundwater storage decreases can be distinguished from GWD due to human water abstractions.

In such a joint analysis, uncertainties of the hydrological model (see section 4.3) and of GRACE data need to be taken into account. One important uncertainty stems from the low spatial resolution of GRACE and thus the need of filtering, with the resulting leakage errors [*Werth and Güntner*, 2010]. A comparison of trends of filtered and unfiltered WaterGAP TWS for major GWD areas (Table 5 and Figures 5b and 5d) shows that trends for GWD regions with intensive GWD as large as the HPA or the three-states regions in Northern India have significantly decreased due to filtering, while this is not the case for even larger regions like Saudi Arabia or the Tiwari study region. The TEWI region shows that the effect of filtering on trends may vary over time, with filtering either increasing or decreasing the trend (Table 5).

Another uncertainty stems from the different approaches to process GRACE data. A part of the uncertainty of GRACE TWS trends can be quantified by comparing GRACE data from different analysis centers (Table 5). Different TWS trends were determined for river basins by *Jensen et al*. [2013] depending on the method to evaluate GRACE data (e.g., fingerprint method with and without using data on sea level, or an approach very similar to our method for deriving TWS values). For example, decreasing trends ranged between 16 and 23 km^3/yr for the Ganges basin and between 2 and 17 km^3/yr for the Indus basin (period August 2002 to July 2009). For the TEWI region, *Voss et al*. [2013] determined an average TWS decline of 27 mm/yr for the period 2003–2009, as compared to the range of 21–24 mm/yr for the three GRACE products of this study (derived from values in Table 5 by correcting the filtering effect by multiplying with the ratio of unfiltered to filtered IRR70_S TWS). Looking at shorter time periods, GRACE CSR results (also used by *Voss et al*. [2013]) show a stronger change in TWS trends before and after 2007 than GRACE GFZ, the latter fitting better to WaterGAP results than GRACE CSR (Table 5). According to WaterGAP, climate-driven TWS decrease in 2007–2009 is 15 mm/yr (computed using variant NOUSE_S), and human water use driven TWS decrease

only 7 mm/yr. Comparing WaterGAP results to GRACE GFZ-RL05 trends, one could deduce that human water use has caused a TWS decrease of 29 mm/yr instead of only 7 mm/yr during this period. However, it is not possible to deduce if the additionally depleted water stems from groundwater or surface water use.

4.3. Uncertainties of Derived GWD and Deficit Irrigation

Given the groundwater storage balance (equation (2)), simulated GWD as well as the estimated degree of deficit irrigation strongly depend on computed groundwater recharge R_g and net groundwater abstractions NA_g in case of optimal irrigation. Estimates of optimal irrigation are a function of irrigated area, crops, growing season, climate, and irrigation water use efficiency. Based on the computed net and gross irrigation water requirements under optimal conditions, NA_g is computed as a function of the fractions of groundwater and surface water use and the assumptions about return flows [*Döll et al*., 2012]. R_g depends on knowledge about daily climate variables, land cover, relief, and soils. This makes computation of R_g and NA_g and thus GWD highly uncertain. *Gleeson and Wada* [2013] found that much of the uncertainty of groundwater stress in the United States and India is related to groundwater recharge estimates. It is, however, not possible to conclude from their study whether the uncertainty of water use estimates is more or less important than the uncertainty of recharge estimates as only two water use estimates were used that were not independent of each other.

As an example, the strong overestimation of GWD in the North China Plain/Hai river basin is very likely due to a strong underestimation of R_g, with 12 mm/yr for the North China Plain (for 2000–2008) in WaterGAP (plus 5 mm/yr groundwater recharge from surface water bodies), as compared to 94 mm/yr estimated by *Liu et al*. [2011]. As no streamflow observations are available for the North China Plain, a regionalized value for the runoff coefficient is applied in WaterGAP (section 2.1). If regionalization of the runoff coefficient would be done based on proximity instead of basin characteristics, the runoff coefficient would be 1 instead of 4, and R_g would increase to almost 40 mm/yr. Interestingly, with a former version of WaterGAP with a different regionalization algorithm (but also based on basin characteristics) and driven by different climate data, R_g in the North China Plain is estimated to be almost 40 mm/yr in 2000–2008 and 60 mm/yr in 1961–1990 [*Döll and Fiedler*, 2008]. Modeled and independently estimated NA_g values are similar (140 mm/yr versus 130 mm/yr, time period 2000–2008). This supports the adjustment of the NA_g computation made for WaterGAP 2.2a. Assuming a higher irrigation efficiency for groundwater use (resulting in reduced groundwater abstractions for irrigation) and a decreased return flow fraction to surface water bodies in WaterGAP 2.2a as compared to 2.1h [*Döll et al*., 2012] resulted in a 20% decrease of NA_g in the HPA.

Even though modeled long-term annual average GWD over regions may fit well to independent estimates, this can be due to a balancing of errors over time or space. The same is true for modeled water abstractions. As an example, GWD in the HPA is underestimated by IRR70_S in 2003 but overestimated in 2004, compared to well observation in 2003–2004 (Figure 2). IRR70_S groundwater abstractions in the whole HPA fit very well to estimates of *USGS* [2009], supporting the assumption of a 70% deficit irrigation. However, for the Northern HPA, optimal irrigation leads to the best fit to USGS data, while it is 70% deficit irrigation for the Central HPA and even 50% deficit irrigation in the dry Southern HPA that would reflect USGS groundwater abstractions [*Schuh*, 2013, Table 4]. Obviously, our best estimate of 70% deficit irrigation can only be regarded a rough overall estimate given spatial and temporal heterogeneity.

To validate the spatial pattern of modeled GWD at the 0.5° grid scale, which strongly depends on the irrigated areas in the grid cell, it is not possible to rely on GRACE due to its coarse spatial resolution. Satellite radar altimetry observations of land subsidence can be compared to spatial patterns of modeled GWD in case of compressible underground material. Based on radar altimetry, land subsidence was detected for areas around the seven Iranian cities shown in Figure 3, and groundwater tables had been observed to decline (but not enough data are available to estimate GWD) [*Motagh et al*., 2008].

Temporal development of groundwater abstractions and GWD only reflects changes in total abstractions due to climate, extension of irrigated areas, population growth, etc. Changes in irrigation water use efficiencies, cropping pattern, and groundwater abstractions as a fraction of total abstractions are not taken into account in WaterGAP. While an increase of groundwater abstractions as a fraction of total abstractions has been documented for United States and India [*Siebert et al*., 2010], there is little information globally. This increase has been estimated at the global scale only by a simulation approach that calculates groundwater abstractions as a function of total water withdrawals and surface water availability [*Wada et al*., 2014]. The grid-specific but temporally constant fractions of sectoral groundwater use applied in WaterGAP may

slightly underestimate groundwater abstractions after 2000, and more strongly overestimate those in the 1960s, but uncertainty of groundwater abstractions due to the unresolved temporal development of the groundwater fractions are expected to be only a small part of the overall uncertainty.

An additional uncertainty of computed GWD stems from the fact that WaterGAP does not simulate groundwater dynamics by considering hydraulic head gradients but by using a simple water balance approach. On one hand, groundwater recharge from surface water bodies induced by lowered hydraulic heads due to groundwater abstractions is not simulated in the model. This should lead to an overestimation of simulated GWD by WaterGAP. On the other hand, depending on the aquifer properties and the distance between abstractions and surface water bodies, the cone of depression and thus the reduction of base flow may take a very long time to fully develop, maybe more than 100 years [*Bredehoeft*, 2002]. As WaterGAP assumes that base flow reduces in proportion to groundwater storage, WaterGAP should underestimate GWD in cases of long lag times for base flow reduction, as is particularly the case for groundwater abstractions from deep confined aquifers.

4.4. Comparison to Other Estimates of Global GWD

We derived that average global GWD during the periods 1960–2000 and 2000–2009 was 56 and 113 km^3/yr, respectively. For 1993–2008, the average GWD was 95 km^3/yr (no correction done for United States). This value is much smaller than the rather implausible value (362 km^3/yr during 1987–2007) of *Pokhrel et al.* [2012] (see section 1). It is smaller than the values of *Konikow* [2011] and *Wada et al.* [2012a] who estimate an average GWD of 121 km^3/yr (1991–2008) and 195 km^3/yr (1993–2008), respectively. The corresponding sea level rise caused by GWD according to this study is 0.26 mm/yr only, compared to 0.34 and 0.54 mm/yr. The differences cannot be explained by the fact that we subtracted GWD that was probably due to erroneous modeling of the water supply of big cities. This would add only 5 km^3/yr to the GWD values of this study. Taking into account the natural climate-induced groundwater storage changes by subtracting groundwater storage changes computed in case of no water abstractions decreases GWD by 2–3 km^3/yr, depending on the time period considered. Had we not corrected the IRR70_S values for China and the United States, the determined global GWD would be 14 km^3/yr higher at most.

One reason for the discrepancy to both studies is the GWD estimate for the Tiwari study region (Northern India and surrounding regions) with an average GWD of 54 km^3/yr during 2002–2008 according to the GRACE-based study of *Tiwari et al.* [2009]. This value, which accounts for almost half of the global GWD estimate of this study during the first decade of the 21st century, is used directly by *Konikow* [2011] and indirectly by *Wada et al.* [2012a] because the latter adjust GWD in all nonarid grid cells by a correction factor that was heavily influenced by the *Tiwari et al.* [2009] value and reduces the computed value for the Tiwari study region from 71 km^3/yr to the GRACE-based value of 54 km^3/yr (compare Figure S4 of *Wada et al.* [2012a] to their Figure 1). However, a more recent GRACE-based study for a very similar area by *Jacob et al.* [2012] only estimated a GWD of 35 km^3/yr, to which GWD simulated with IRR70_S fit well (Table 4). In addition, the TWS trend of IRR70_S falls into the range of the three GRACE solutions analyzed in this study.

Consideration of the overestimated GWD from the study by *Tiwari et al.* [2009] can explain two thirds of the difference between the global GWD estimates of *Konikow* [2011] and this study. In addition, as there are no quantitative studies on GWD for many areas of the globe, 30% of Konikow's global GWD is derived by extrapolation of the relation between GWD and total water withdrawals in the United States, and total withdrawals in those areas not covered by studies [*Konikow*, 2011].

The much higher global GWD estimate of *Wada et al.* [2012a] could be related to differences in estimated groundwater recharge and abstractions, but also to the fact that they do not model three processes that decrease GWD, and which are simulated by the version of WaterGAP applied in this study: (1) deficit irrigation, (2) base flow reductions due to groundwater abstractions (as noted by *Konikow* [2011]), and (3) groundwater recharge from surface water bodies. Relative overestimation of GWD due to neglecting baseflow reduction is expected to be particularly high in areas where net groundwater abstraction does not strongly exceed groundwater recharge. In these areas, net abstraction can be balanced by decreased base flow.

5. Conclusions

Applying an improved version of the global hydrological and water use model WaterGAP 2.2a that takes into account groundwater recharge from surface water bodies in semiarid and arid regions, groundwater

depletion GWD was quantified. Groundwater abstractions are based on spatially highly resolved information on the fraction of irrigated areas that is equipped with surface water or groundwater irrigation which was, however, assumed to be constant over time due to lack of data. Reduction of base flow due to water abstractions as well as return flows from irrigation with surface water was simulated (see sections 2.1.1 and 2.1.3). Final GWD estimates were corrected for the natural temporal variability of groundwater storage and the likely overestimation of GWD in grid cells with big cities. By comparing model results to a large number of independent regional-scale estimates of GWD and estimates of relevant water fluxes like groundwater recharge and groundwater abstractions, it was found that GWD can be modeled best if it is assumed that deficit irrigation occurs in GWD areas, with only 70% of the optimal irrigation water amount.

The modeled spatial distribution of GWD areas around the globe on the continents fits well to information from the literature. To derive a best estimate of GWD at the country and global scales, model results were corrected for the United States and China based on reliable independent estimates (section 3.6). According to our estimation, India, United States, Iran, Saudi Arabia, and China are the countries with the highest GWD rates in the first decade of the 21st century. In the countries on the Arabian Peninsula, and in Libya, Egypt, Israel, Mali, Mozambique, and Mongolia, at least 30% of the abstracted groundwater was taken from nonrenewable groundwater during this time period. In all but a few countries, average annual GWD has strongly increased from the period 1960–2000 to the period 2000–2009. We compute that global GWD has doubled from an average 56 km^3/yr to an average 113 km^3/yr, but considering the increase in the groundwater fraction of total water abstractions over time, it is likely that it more than doubled between 1960–2000 and 2000–2009. The fraction of global groundwater abstractions from nonrenewable groundwater has increased from approximately 11% to 15%.

With 95 km^3/yr for 1993–2008, our best estimate of global GWD is somewhat smaller than the previous estimate by *Konikow* [2011], and much smaller than the two estimates that are, like ours, derived from global hydrological modeling [*Wada et al*., 2012a; *Pokhrel et al*., 2012]. Advantages of our GWD estimates are a detailed groundwater use database, the consideration of deficit irrigation in GWD areas based on comparison to independent groundwater well and GRACE-based GWD estimates, the simulation of base flow reduction by groundwater abstraction and the simulation of groundwater recharge from surface water bodies. Nevertheless, modeling uncertainties remain high, as discussed in sections 4.1 and 4.3. According to our best estimate, contribution of groundwater depletion to sea level rise was only 0.26 mm/yr for 1993–2008 (and 0.31 mm/yr for 2000–2009).

While it was found that the seasonal variation of TWS as observed by GRACE cannot be used to quantify groundwater abstractions [*Döll et al*., 2014], GRACE TWS trends can help to constrain (ground)water abstraction in conjunction with a hydrological and water use model like WaterGAP. In particular, monitoring of recent temporal dynamics of water abstraction not (yet) documented could be achieved. Still, the low spatial resolution of GRACE data remains a challenge as GWD is often quite local.

Acknowledgments
Part of this study was funded by the German Research Foundation (DFG Priority Program SPP 1257 Mass Transport and Mass Distribution in the System Earth). We thank Laura Foglia for her advice, and the five reviewers for their thoughtful comments which helped to improve the manuscript. WaterGAP model output referred to in this paper is available at http://www2.uni-frankfurt.de/49903932/7_GWdepletion.

References

Abderrahman, W. A. (2006), Saudi Arabia aquifers, in *Non-Renewable Groundwater Resources: A Guidebook on Socially-Sustainable Management for Water-Policy Makers*, edited by S. Foster and D. P. Loucks, vol. 10, pp. 63–67, U. N. Educ., Sci. and Cult. Organ., Paris.

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

Bakhbakhi, M. (2002), Hydrological framework of the Nubian sandstone aquifer system, in *GW-MATE/UNESCO Expert Group Meeting on Socially Sustainable Management of Groundwater Mining form Aquifer Storage*, 6 pp., Paris..

Bettadpur, S. V. (2012), Insights into the Earth system mass variability from CSR-RL05 GRACE gravity fields, in *EGU General Assembly Conference Abstracts*, Univ. of Tex., Austin. [Available at http://meetingorganizer.copernicus.org/ EGU2012/EGU2012–6409.pdf.].

BP (2009), *Statistical Review of World Energy* 2009, BP, London.

Bredehoeft, J. D. (2002), The water budget myth revisited: Why hydrogeologists model, *Ground Water*, *40*, 340–345, doi:10.1111/j.1745–6584.2002.tb02511.x.

Central Ground Water Board (CGWB) (2009), Groundwater resource estimation methodology, vol. 18, report of the Ground Water Resource Estimation Committee, Ministry of Water Resour., Gov. of India, New Delhi. [Available at http://cgwb.gov.in/documents/gec97.pdf, last accessed 9 Oct. 2013.].

Cheema, M., W. Immerzeel, and W. Bastiaanssen (2014), Spatial quantification of groundwater abstraction in the irrigated Indus basin, *Ground Water*, *52*, 25–36, doi:10.1111/gwat.12027.

Cheng, M. (2004), Variations in the Earth's oblateness during the past 28 years, *J. Geophys. Res.*, *109*, B09402, doi:10.1029/2004JB003028.

Custodio, E. (2010), Intensive groundwater development: A water cycle transformation, a social revolution, a management challenge, in *Rethinking Water and Food Security*, edited by L. Martínez-Cortina et al., pp. 259–277, CRC Press, Leiden, Netherlands.

Dahle, C., F. Flechtner, C. Gruber, D. König, R. König, G. Michalak, and K.-H. Neumayer (2013), GFZ GRACE Level-2 Processing Standards Document for Level-2 Product Release 0005: revised edition, January 2013, (Scientific Technical Report - Data; 12/02 rev. ed.), Potsdam: Deutsches GeoForschungs Zentrum GFZ, *21*, doi:10.2312/GFZ.b103-1202-25.
Döll, P., and K. Fiedler (2008), Global-scale modeling of groundwater recharge, *Hydrol. Earth Syst. Sci.*, *12*, 863–885, doi:10.5194/hess-12–863-2008.
Döll, P., and S. Siebert (2002), Global modeling of irrigation water requirements, *Water Resour. Res.*, *38*(4), 1037, doi:10.1029/2001WR000355.
Döll, P., F. Kaspar, and B. Lehner (2003), A global hydrological model for deriving water availability indicators: Model tuning and validation, *J. Hydrol.*, *270*, 105–134, doi:10.1016/S0022-1694(02)00283-4.
Döll, P., H. Hoffmann-Dobrev, F. T. Portmann, S. Siebert, A. Eicker, M. Rodell, G. Strassberg, and B. Scanlon (2012), Impact of water withdrawals from groundwater and surface water on continental water storage variations, *J. Geodyn.*, *59–60*, 143–56, doi:10.1016/j.jog.2011.05.001.
Döll, P., M. Fritsche, A. Eicker, and H. Müller Schmied (2014), Seasonal water storage variations as impacted by water abstractions: Comparing the output of a global hydrological model with GRACE and GPS observations, *Surv. Geophys.*, doi:10.1007/s10712-014-9282-2.
Famiglietti, J. S., M. Lo, S. L. Ho, J. Bethune, K. J. Anderson, T. H. Syed, S. C. Swenson, C. R. de Linage, and M. Rodell (2011), Satellites measure recent rates of groundwater depletion in California's Central Valley, *Geophys. Res. Lett.*, *38*, L03403, doi:10.1029/2010GL046442.
Faunt, C. C. (Ed.) (2009), Groundwater availability of the Central Valley Aquifer, *U.S. Geol. Surv. Prof. Pap.*, 1776, 225 p.
Feng, W., M. Zhong, J.-M. Lemoine, R. Biancale, H.-T. Hsu, and J. Xia (2013), Evaluation of groundwater depletion in North China using the Gravity Recovery and Climate Experiment (GRACE) data and ground-based measurements, *Water Resour. Res.*, *49*, 2110–2118, doi:10.1029/2010GL046442.
Flechtner, F. (2007), *AOD1B product Description Document for Product Releases 01 to 04*, Geoforschungszentrum, Potsdam, Germany.
Flörke, M., E. Kynast, I. Bärlund, S. Eisner, F. Wimmer, and J. Alcamo (2013), Domestic and industrial water uses of the past 60 years as a mirror of socio-economic development: A global simulation study, *Global Environ. Change*, *23*, 144–156, doi:10.1016/j.gloenvcha.2012.10.018.
Foster, S., and P. D. Loucks (Eds.) (2006), *Non-Renewable Groundwater Resources: A Guidebook on Socially-Sustainable Management for Water-Policy Makers, IHP-VI Ser. Groundwater*, U. N. Educ., Sci. and Cult. Organ., Paris.
Ghoubachi, S. Y. (2012), Impact of Lake Nasser on the groundwater of the Nubia sandstone aquifer system in Tushka area, South Western Desert, Egypt, *J. King Saud Univ. Sci.*, *24*, 101–109, doi:10.1016/j.jksus.2010.04.005.
Gleeson, T., and Y. Wada (2013), Assessing regional groundwater stress for nations using multiple data sources with the groundwater footprint, *Environ. Res. Lett.*, *8*, 44010, doi:10.1088/1748-9326/8/4/044010.
Gossel, W., A. M. Ebraheem, and P. Wycisk (2004), A very large scale GIS-based groundwater flow model for the Nubian sandstone aquifer in Eastern Sahara (Egypt, northern Sudan and eastern Libya), *Hydrogeol. J.*, *12*, 698–713, doi:10.1007/s10040-004-0379-4.
Gupta, S. D. R. (2004), Water for India in 2050: First-order assessment of available options, *Curr. Sci.*, *86*, 1216–1224.
Gurdak, J. J., and C. D. Roe (2010), Review: Recharge rates and chemistry beneath playas of the high plains aquifer, USA, *Hydrogeol. J.*, *18*, 1747–1772, doi:10.1007/s10040-010-0672-3.
Harris, I., P. Jones, T. Osborn, and D. Lister (2014), Updated high-resolution grids of monthly climatic observations—The CRU TS3.10 dataset, *Int. J. Climatol.*, *34*, 623–642, doi:10.1002/joc.3711.
Jacob, T., J. Wahr, W. T. Pfeffer, and S. Swenson (2012), Recent contributions of glaciers and ice caps to sea level rise, *Nature*, *482*, 514–518, doi:10.1038/nature10847.
Jensen, L., R. Rietbroek, and J. Kusche (2013), Land water contribution to sea level from GRACE and Jason-1 measurements, *J. Geophys. Res.*, *118*, 212–226, doi:10.1002/jgrc.20058.
Kim, J., and M. Sultan (2002), Assessment of the long-term hydrological impacts of Lake Nasser and related irrigation projects in Southwestern Egypt, *J. Hydrol.*, *262*, 68–83, doi:10.1016/S0022-1694(02)00013-6.
Konikow, L. F. (2011), Contribution of global groundwater depletion since 1900 to sea-level rise, *Geophys. Res. Lett.*, *38*, L17401, doi:10.1029/2011GL048604.
Konikow, L. F., and E. Kendy (2005), Groundwater depletion: A global problem, *Hydrogeol. J.*, *13*, 317–320, doi:10.1007/s10040-004-0411-8.
Kusche, J. (2007), Approximate decorrelation and non-isotropic smoothing of time-variable GRACE-type gravity field models, *J. Geodyn.*, *81*, 733–749, doi:10.1007/s00190-007-0143-3.
Kusche, J., R. Schmidt, S. Petrovic, and R. Rietbroek (2009), Decorrelated GRACE time-variable gravity solutions by GFZ, and their validation using a hydrological model, *J. Geodyn.*, *83*, 903–913, doi:10.1007/s00190-009-0308-3.
LaGrange, T. (2005), *Guide to Nebraska's Wetlands and Their Conservation Needs*, Nebr. Game and Parks Comm, Lincoln, Nebraska, USA.
Leblanc, M. J., P. Tregoning, G. Ramillien, S. O. Tweed, and A. Fakes (2009), Basin-scale, integrated observations of the early 21st century multiyear drought in southeast Australia, *Water Resour. Res.*, *45*, W04408, doi:10.1029/2008WR007333.
Liu, J., D. Qin, H. Wang, M. Wang, and Z. Yang (2010), Dualistic water cycle pattern and its evolution in Haihe River basin, *Chin. Sci. Bull.*, *55*, 1688–1697, doi:10.1007/s11434-010-3043-5.
Liu, J., G. Cao, and C. Zheng (2011), Sustainability of groundwater resources in the North China Plain, in *Sustaining Groundwater Resources: A Critical Element in the Global Water Crisis*, edited by J. Anthony and A. Jones, pp. 69–87, Springer, Dordrecht, Netherlands.
Mamou, A., B. Besbes, B. Abdous, D. J. Latrech, and C. Fezzani (2006), North Western Sahara Aquifer System (NWSAS), in *Non-Renewable Groundwater Resources: A Guidebook on Socially-Sustainable Management for Water-Policy Makers*, edited by S. Foster and D. P. Loucks, vol. 10, pp. 68–74, U. N. Educ., Sci. and Cult. Organ., Paris.
Margat, J., and J. van der Gun (2013), *Groundwater Around the World*, CRC Press, London.
Margat, J., S. Foster, and A. Droubi (2006), Concept and importance of non-renewable resources, in *Non-Renewable Groundwater Resources: A Guidebook on Socially-Sustainable Management for Water-Policy Makers*, vol. 10, edited by S. Foster and D. P. Loucks, pp. 13–24, U. N. Educ., Sci. and Cult. Organ., Paris.
Mayer-Gürr, T., E. Kurtenbach, and A. Eicker (2010), ITG-Grace2010 gravity field model, Inst. für Geodäsie und Geoinf., Univ. Bonn, Bonn, Germany. [Available at http://www.igg.uni-bonn.de/apmg/index.php?id=itg-grace201.]
McGuire, V. L. (2007), Water-level changes in the High Plains aquifer: Predevelopment to 2005 and 2003 to 2005, *U.S. Geol. Surv. Sci. Invest. Rep. 2006–5324*. [Available at http://pubs.usgs.gov/sir/2006/5324/.], U.S. Geological Survey, Reston, Va.
McGuire, V. L. (2009), Water-level changes in the High Plains aquifer: Predevelopment to 2007, 2005-06 and 2006-07, *U.S. Geol. Surv. Sci. Invest. Rep. 2009–5019*. [Available at http://pubs.usgs.gov/sir/2009/5019/.], U.S. Geological Survey, Reston, Va.
McGuire, V. L. (2011), Water-level changes in the High Plains aquifer: Predevelopment to 2009, 2007–08, and 2008–09, and change in water in storage, predevelopment to 2009, *U.S. Geol. Surv. Sci. Invest. Rep. 2011–5089*. [Available at http://pubs.usgs.gov/sir/2011/5089/.], U.S. Geological Survey, Reston, Va.

McGuire, V. L., M. Johnson, R. Schieffer, J. Stanton, S. Sebree, and I. M. Verstraeten (2003), Water in storage and approaches to groundwater management, High Plains Aquifer, 2000, *U.S. Geol. Surv. Circ.*, 1243, 51 p.
Metwaly, M., M. Khalil, E.-S. Al-Sayed, and S. Osman (2006), A hydrogeophysical study to estimate water seepage from northwestern Lake Nasser, Egypt, *J. Geophys. Eng.*, *3*, 21–27, doi:10.1088/1742-2132/3/1/003.
Moiwo, J. P., Y. Yang, H. Li, S. Hanand, and Y. Hu (2009), Comparison of GRACE with in situ hydrological measurement data shows storage depletion in Hai River basin, Northern China, *Water SA*, *35*, 663–670.
Motagh, M., T. R. Walter, M. A. Sharifi, E. Fielding, A. Schenk, J. Anderssohn, and J. Zschau (2008), Land subsidence in Iran caused by widespread water reservoir overexploitation, *Geophys. Res. Lett.*, *35*, L16403, doi:10.1029/2008GL033814.
Müller Schmied M., S. Eisner, D. Franz, M. Wattenbach, F. T. Portmann, M. Flörke, and P. Döll (2014), Sensitivity of simulated global-scale freshwater fluxes and storages to input data, hydrological model structure, human water use and calibration, *Hydrol. Earth Syst. Sci. Discuss.*, *11*, 1583–1649, doi:10.5194/hessd-11-1583-2014.
Pokhrel, Y. N., N. Hanasaki, P. J.-F. Yeh, T. J. Yamada, S. Kanae, and T. Oki (2012), Model estimates of sea-level change due to anthropogenic impacts on terrestrial water storage, *Nat. Geosci.*, *5*, 389–392, doi:10.1038/NGEO1476.
Rietbroek, R., S.-E. Brunnabend, J. Kusche, and J. Schröter (2012a), Resolving sea level contributions by identifying fingerprints in time-variable gravity and altimetry, *J. Geodyn.*, *59–60*, 72–81, doi:10.1016/j.jog.2011.06.007.
Rietbroek, R., M. Fritsche, S.-E. Brunnabend, I. Daras, J. Kusche, J. Schröter, F. Flechtner, and R. Dietrich (2012b), Global surface mass from a new combination of GRACE, modelled OBP and reprocessed GPS data, *J. Geodyn.*, *59–60*, 64–71, doi:10.1016/j.jog.2011.02.003.
Rodell, M., J. Chen, H. Kato, J. S. Famiglietti, J. Nigro, and C. R. Wilson (2007), Estimating groundwater storage changes in the Mississippi River basin (USA) using GRACE, *Hydrogeol. J.*, *15*, 159–166, doi:10.1007/s10040-006-0103-7.
Rodell, M., I. Velicogna, and J. S. Famiglietti (2009), Satellite-based estimates of groundwater depletion in India, *Nature*, *460*, 999–1002, doi:10.1038/nature08238.
Scanlon, B. R., C. C. Faunt, L. Longuevergne, R. C. Reedy, W. M. Alley, V. L. McGuire, and P. B. McMahon (2012), Groundwater depletion and sustainability of irrigation in the US High Plains and Central Valley, *Proc. Natl. Acad. Sci. U. S. A.*, *109*, 9320–9325, doi:10.1073/pnas.1200311109.
Schneider, U., A. Becker, P. Finger, A. Meyer-Christoffer, M. Ziese, and B. Rudolf (2013), GPCC's new land surface precipitation climatology based on quality-controlled in situ data and its role in quantifying the global water cycle, *Theor. Appl. Climatol.*, *115*, 15–40, doi:10.1007/s00704-013-0860-x.
Schuh, C. (2013), Evaluating groundwater depletion as computed by a global water model, B.Sc. thesis, Inst. of Phys. Geogr., Univ. of Frankfurt, Frankfurt, Germany.
Shamsudduha, M., R. G. Taylor, and L. Longuevergne (2012), Monitoring groundwater storage changes in the highly seasonal humid tropics: Validation of GRACE measurements in the Bengal Basin, *Water Resour. Res.*, *48*, W02508, doi:10.1029/2011WR010993.
Siebert, S., J. Burke, J. M. Faures, K. Frenken, J. Hoogeveen, P. Döll, and F. T. Portmann (2010), Groundwater use for irrigation—A global inventory, *Hydrol. Earth Syst. Sci.*, *14*, 1863–1880, doi:10.5194/hess-14-1863-2010.
Strassberg, G., B. R. Scanlon, and D. Chambers (2009), Evaluation of groundwater storage monitoring with the GRACE satellite: Case study of the High Plains aquifer, central United States, *Water Resour. Res.*, *45*, W05410, doi:10.1029/2008WR006892.
Tiwari, V. M., J. Wahr, and S. Swenson (2009), Dwindling groundwater resources in northern India, from satellite gravity observations, *Geophys. Res. Lett.*, *36*, L18401, doi:10.1029/2009GL039401.
USGS (2009), Estimated use of water in the United States. County-level data for 2005, *U.S. Geol. Surv. Water Resour*. [Available at http://water.usgs.gov/watuse/data/2005/].
Voss, K. A., J. S. Famiglietti, M. Lo, C. de Linage, M. Rodell, and S. C. Swenson (2013), Groundwater depletion in the Middle East from GRACE with implications for transboundary water management in the Tigris-Euphrates-Western Iran region, *Water Resour. Res.*, *49*, 904–914, doi:10.1002/wrcr.20078.
Wada, Y., L. P. H. van Beek, F. C. S. Weiland, B. F. Chao, Y.-H. Wu, and M. F. P. Bierkens (2012a), Past and future contribution of global groundwater depletion to sea-level rise, *Geophys. Res. Lett.*, *39*, L09402, doi:10.1029/2012GL051230.
Wada, Y., L. P. H. van Beek, and M. F. P. Bierkens (2012b), Nonsustainable groundwater sustaining irrigation: A global assessment, *Water Resour. Res.*, *48*, W00L06, doi:10.1029/2011WR010562.
Wada, Y., D. Wisser, and M. F. P. Bierkens (2014), Global modeling of withdrawal, allocation and consumptive use of surface water and groundwater resources, *Earth Syst. Dyn.*, *5*, 15–40, doi:10.5194/esd-5-15-2014.
Wahr, J., M. Molenaar, and F. Bryan (1998), Time variability of the Earth's gravity field: Hydrological and oceanic effects and their possible detection using GRACE, *J. Geophys. Res.*, *103*, 30,205–30,229, doi:10.1029/98JB02844.
Werth, S., and A. Güntner (2010), Calibration analysis for water storage variability of the global hydrological model WGHM, *Hydrol. Earth Syst. Sci.*, *14*, 59–78.
Zhou, Y. (2009), A critical review of groundwater budget myth, safe yield and sustainability, *J. Hydrol.*, *370*, 207–213, doi:10.1016/j.jhydrol.2009.03.009.

7 How is the impact of climate change on river flow regimes related to the impact on mean annual runoff? A global-scale analysis

Full citation: Döll, P. and Müller Schmied, H.: How is the impact of climate change on river flow regimes related to the impact on mean annual runoff? A global-scale analysis, Environ. Res. Lett., 7(1), 014037, doi:10.1088/1748-9326/7/1/014037, 2012.

Declaration of portion of own work

Experiment design and text writing were mainly done by P. Döll. All model runs, analyses, figures and tables were done by H. Müller Schmied, as was manuscript revision. In overall percentage, H. Müller Schmied contributed around 40% of this manuscript.

Hannes Müller Schmied

Environ. Res. Lett. **7** (2012) 014037 (11pp) doi:10.1088/1748-9326/7/1/014037

How is the impact of climate change on river flow regimes related to the impact on mean annual runoff? A global-scale analysis

Petra Döll and Hannes Müller Schmied

Institute of Physical Geography, Goethe University Frankfurt, Frankfurt am Main, Germany

E-mail: p.doell@em.uni-frankfurt.de

Received 24 December 2011
Accepted for publication 2 March 2012
Published 26 March 2012
Online at stacks.iop.org/ERL/7/014037

Abstract
To assess the impact of climate change on freshwater resources, change in mean annual runoff (MAR) is only a first indicator. In addition, it is necessary to analyze changes of river flow regimes, i.e. changes in the temporal dynamics of river discharge, as these are important for the well-being of humans (e.g. with respect to water supply) and freshwater-dependent biota (e.g. with respect to habitat availability). Therefore, we investigated, in a global-scale hydrological modeling study, the relation between climate-induced changes of MAR and changes of a number of river flow regime indicators, including mean river discharge, statistical low and high flows, and mean seasonal discharge. In addition, we identified, for the first time at the global scale, where flow regime shifts from perennial to intermittent flow regimes (or vice versa) may occur due to climate change. Climate-induced changes of all considered river flow regime indicators (except seasonal river flow changes) broadly follow the spatial pattern of MAR changes. The differences among the computed changes of MAR due to the application of the two climate models are larger than the differences between the change of MAR and the change of the diverse river flow indicators for one climate model. At the sub-basin and grid cell scales, however, there are significant differences between the changes of MAR, mean annual river discharge, and low and high flows. Low flows are projected to be more than halved by the 2050s in almost twice the area as compared to MAR. Similarly, northern hemisphere summer flows decrease more strongly than MAR. Differences between the high emissions scenario A2 (with emissions of 25 Gt C yr^{-1} in the 2050s) and the low emissions scenario B2 (16 Gt C yr^{-1}) are generally small as compared to the differences due to the two climate models. The benefits of avoided emissions are, however, significant in those areas where flows are projected to be more than halved due to climate change. If emissions were constrained to the B2 scenario, the area with ecologically relevant flow regime shifts would be reduced to 5.4%–6.7% of the global land area as compared to 6.3%–7.0% in A2. In particular, under the B2 scenario, fewer rivers will change from perennial to intermittent (or transitional) river flows.

Keywords: climate change, river flow, flow regime, runoff

1. Introduction

When assessing the impact of climate change on freshwater resources at the global scale, it is most common to look at changes of mean annual runoff (MAR), i.e. the difference between the long-term averages of annual precipitation and evapotranspiration (Bates *et al* 2008). While changes in MAR are of major interest as they represent changes of total renewable water resources, the assessment of these changes alone is not sufficient for supporting sustainable water management. According to Gosling *et al* (2011), 'the common use of MAR as a measure of the response of hydrological systems to climate change is oversimplistic'. More specific analyses of climate change impacts on runoff components like groundwater recharge (Döll 2009), or on river discharge, i.e. runoff accumulated along the drainage direction, are required.

River discharge (also called river flow) varies over space and time. The relevant temporal scales range from minutes (e.g. in the case of flash floods) to decades (e.g. in the case of water resource assessments). River flow regimes describe the temporal patterns of flow variability. Knowledge about changing river flow regimes is paramount for assessing climate change risks related to freshwater. Estimation of changes in seasonality, interannual variability, statistical low and high flows, and floods and droughts is required to understand the impact of climate change on humans and freshwater ecosystems. River flow regime alterations affect humans with respect to water supply, navigation, hydropower generation and flooding, and they affect ecosystems with respect to habitat suitability for freshwater-dependent biota (Poff and Zimmerman 2010). Of particular relevance for freshwater ecosystems are shifts from perennial to intermittent flow regimes or vice versa, as they have a strong impact on biota living in the river (Bond *et al* 2010) and on riparian vegetation (Stromberg *et al* 2005).

Climate change impact studies for individual drainage basins mostly focus on changes of river discharge and aspects of its temporal variability, in particular seasonality (Kundzewicz *et al* 2007). Global-scale studies on the impact of climate change on river flow regimes are still rare (e.g. Milly *et al* 2002, Hirabayashi *et al* 2008, Döll and Zhang 2010). Our study was inspired by Gosling *et al* (2011) who applied a global hydrology model (GHM) (without the capability to route water downstream) and six catchment models to compute the impact of climate change on MAR in six catchments (one catchment model per catchment). In addition, they simulated the impact of climate change on monthly runoff as well as on monthly low and high flows, and related the changes of these regime indicators to changes in MAR. For the Rio Grande and the Okavango, they found that flow seasonality would not change under climate change, while it would change significantly in the other four catchments. Changes in low and high flows were found to be generally similar to changes of mean annual flows, with some exceptions.

The objective of this study was to better understand, at the global level, how climate-induced changes of indicators of the river flow regime, which are important for the well-being of humans and ecosystems, relate to changes of MAR. We wanted to find out how significant the results of resource-intensive ensemble studies on the impact of climate change on MAR would be for assessing impacts on river flow regimes. We used the state-of-the-art global water model WaterGAP (Alcamo *et al* 2003, Döll *et al* 2003) to translate four climate scenarios (as computed by two global climate models) into scenarios not only of MAR but also of a number of flow regime indicators, and of shifts between perennial and intermittent flow regimes. As flow regimes are affected by human water use, predominantly by irrigation, we also took into account the impact of climate change on irrigation water use. In section 2, both the WaterGAP model and the climate data are described. In addition, we introduce the investigated flow regime indicators, in particular how perennial and intermittent flow regimes are identified globally by WaterGAP. In sections 3 and 4, computational results are presented and discussed. Finally, conclusions are drawn.

2. Methods

2.1. Model description

With a spatial resolution of 0.5° × 0.5° (55 km × 55 km at the equator), the global water resources and use model WaterGAP simulates water flows and storages as well as human water use for all land areas of the globe excluding Antarctica (Alcamo *et al* 2003, Döll *et al* 2003). For this study, WaterGAP 2.1g was applied. Water use, i.e. water withdrawals and consumptive water use, is estimated by separate models for the sectors' irrigation (Döll and Siebert 2002), livestock, households and industry. The WaterGAP Global Hydrology Model (WGHM) computes groundwater recharge, total runoff generation and river discharge, taking into account the impact of human water use and man-made reservoirs on river discharge (Döll 2009, Döll and Fiedler 2008). For each grid cell, the vertical water balance is computed, and the resulting runoff is routed laterally within the cell through a groundwater store and various surface water stores (so-called 'local' surface water bodies that are only fed by runoff of the cell, and so-called 'global' surface water bodies that also receive water from upstream cells). All groundwater generated within one cell is assumed to return to surface water within each grid cell. The effect of surface water bodies on water balance and flow dynamics is modeled by first routing the runoff generated within the grid cell through 'local' lakes, reservoirs and wetlands. The difference between precipitation and potential evapotranspiration is computed for each surface water type within the grid cell, thus taking into account the effect of the surface water balance on cell runoff. The water volume resulting from grid cell runoff is added to the discharge from the upstream grid cell and routed through 'global' lakes, reservoirs and wetlands, and through the river storage compartment. The total runoff of a grid cell may be negative as it includes the water balance of surface water bodies which may be negative. Negative surface water balances occur if water bodies fed from upstream evaporate

more water than falls as precipitation on the water bodies in the cell. Grid cell discharge is assumed to represent discharge in the largest river within the grid cell, and is routed to the next downstream cell according to the global drainage direction map DDM30 (Döll and Lehner 2002). WaterGAP is tuned in a basin-specific manner against long-term average discharge at 1235 gauging stations (discharge data provided by GRDC, grdc.bafg.de), by adjusting 1–3 model parameters (Hunger and Döll 2008).

Important WaterGAP inputs include time series of monthly values of climate variables as well as information on soil and land cover. Monthly climate data are downscaled to daily data, in the case of precipitation using the available number of wet days per month. Monthly climate data, except for precipitation, are provided by the CRU TS 2.1 data set (Mitchell and Jones 2005). As precipitation input, 0.5° gridded monthly time series of the GPCC Full Data Product Version 3 (Rudolf and Schneider 2005) were used, together with the number of wet days from the CRU TS 2.1 data set.

2.2. Modeling climate change impacts

We considered four different climate change scenarios, comparing runoff and river flow regimes resulting from the climate during the time period 1961–90 to runoff and flow regimes resulting from the climate during 2041–70 (the 2050s). The two IPCC greenhouse gas emissions scenarios A2 and B2 (Nakicenovic and Swart 2000) were translated into climate change scenarios by two state-of-the-art global climate models, the ECHAM4/OPYC3 model (Röckner *et al* 1996, hereafter referred to as ECHAM4) and the HadCM3 model (Gordon *et al* 2000). In the A2 scenario, emissions increase from 11 Gt C yr^{-1} (CO_2-equivalent) in 1990 to 25 Gt C yr^{-1} in the 2050s, but they only increase to 16 Gt C yr^{-1} in the case of scenario B2. Due to large climate model uncertainties, the same emissions scenarios are translated to rather different climate scenarios, in particular regarding precipitation.

Changes in mean monthly precipitation and temperature between the periods 1961–90 and 2041–70 as computed by the climate models were used to scale the grid cell values of observed monthly precipitation and temperature between 1961–90 that drive WaterGAP in the control run (delta change method). In a first step, the climate model data were interpolated from their original resolutions to the WaterGAP resolution of 0.5° × 0.5°. Then, in the case of temperature, observed values were scaled by adding to them the difference of the climate model values of future (2041–70) and present-day (1961–90) temperature. The 30 yr perturbed precipitation time series was produced by multiplying observed values by future climate model precipitation as a ratio of the present-day precipitation. If present-day monthly precipitation was less than 1 mm, precipitation was scaled additively, like temperature. The impact of changed interannual variability and the predicted increased variability of daily precipitation could not be taken into account in this study due to the delta change method. Finally, the impact of altered radiation, humidity or wind speed on evapotranspiration was also neglected.

We simulated not only the impact of climate change on runoff but also on irrigation water use. Thus, the future flow regimes analyzed in this study are impacted by climate change impacts on both runoff and irrigation water use. Irrigation water use in each grid cell is computed as a function of irrigated area (Siebert *et al* 2005) and the 30 yr temperature and precipitation time series. To isolate the impact of climate change, we kept the irrigated area as well as domestic, industrial and livestock water uses constant at the level of the year 2002 in all simulations. Equally, dams remained constant at the 2002 level. Differently from river discharge, computed runoff is essentially unaffected by human water use and thus by the impact of climate change on irrigation water use (except for some effect of water use on the extent of surface water bodies which affects evapotranspiration).

2.3. Flow regime indicators

All indicators were computed based on 30 yr of monthly values. We considered the impact of climate change on MAR (R_{mean}) and on the following indicators of the river flow (discharge) regime in each 0.5° grid cell.

- Q_{mean}: mean annual river discharge.
- Q_{90}: monthly discharge that is exceeded in nine out of ten months (low flow).
- Q_{10}: monthly discharge that is exceeded in one out of ten months (high flow).
- Q_{DJF}: mean river discharge in December to February season.
- Q_{MAM}: mean river discharge in March to May season.
- Q_{JJA}: mean river discharge in June to August season.
- Q_{SON}: mean river discharge in September to November season.

All discharge indicators represent the actual discharge as affected by human water withdrawals. In addition, we also analyzed the climate impact on mean annual river discharge that would occur without any human water withdrawals.

Furthermore, we computed shifts from perennial to intermittent river flow regimes and vice versa due to climate change. There are no general definitions for perennial rivers versus intermittent rivers. Wilhelm (1997) stated that intermittent rivers fall dry at least one month per year. In the US National Hydrography Dataset (USGS 2000), rivers are classified as 'intermittent' if they 'contain water for only part of the year, but more than just after rainstorms and at snowmelt', and 'perennial' if they 'contain water throughout the year, except for infrequent periods of severe drought'. Based on monthly discharge during 1961–90 (with anthropogenic impact, i.e. reservoirs and water withdrawals), we defined three regime types and determined the regime type of each grid cell using WaterGAP. If more than 350 out of the 360 months had a discharge value larger than a threshold value of 0.0001 km^3/month (0.04 m^3 s^{-1}), the cell was classified

as 'strictly perennial'. If this was the case in less than 321 months, it was classified as 'strictly intermittent', otherwise as 'transitional'. The classification algorithm was derived by trial-and-error based on digital data sets that represent information on perennial and intermittent river reaches from maps, the VMAP0 and ESRI Big Rivers data sets. The VMAP0 data set is an updated version of the Digital Chart of the World and contains 'streams/rivers' which are either 'non-perennial/intermittent/fluctuating' or 'perennial/permanent' (VMAP0 hydrography layer, file hydro-water-course-l.shp, field HYC-DESCRI). Also based on the Digital Chart of the World, ESRI produced a global Big Rivers layer with the attributes 'perennial' and 'intermittent' in which only large rivers are included. The threshold value may represent a situation where river flow velocity is 0.1 m s^{-1}, wetted width is 20 m and water depth is 2 cm, i.e. when the river is essentially dry. A flow regime shift was identified if classification differed for the present and future climates, and the number of months with discharge below the threshold changed by at least 3.

3. Results

3.1. Climate change impacts on MAR and discharge

Averaged over 30 yr, runoff is approximately the difference between mean precipitation P and mean actual evapotranspiration (AET), and is essentially unaffected by human water withdrawals. Figure 1(a) shows the per cent change of MAR until the 2050s (2041–70) as compared to the period 1961–90, in case of emissions scenario A2, using climate scenarios derived by either the climate model ECHAM4 or HadCM3. In figure 1(a), green and pink color mostly indicate cells with a negative mean runoff, which only occurs if there are lakes or wetlands that are fed by discharge from upstream. Surprisingly, there are a number of cells where MAR (P–AET) will become less negative, i.e. increase, even though the surrounding positive runoff values decrease (due to decreased precipitation and increased temperature), e.g. in northeastern Brazil and southern Africa and southern Australia (figure 1(a), in green). This is caused by decreased inflow into the surface water bodies from upstream cells, which leads to a reduction of the effective open water area and thus a reduced cell AET. If this decrease of AET exceeds the decrease of P due to climate change, MAR becomes less negative.

Maps of MAR changes like those of figure 1(a) do not show changes in river flow volumes unless the grid cell is at the upstream edge of the river basin and thus does not receive any inflow from an upstream cell. For all other grid cells, climate change impact on river discharge is different from climate change impact on runoff because discharge aggregates runoff from cell to cell along the drainage direction. Hence the locations of many rivers that are fed by large upstream areas are clearly visible in the maps that show the impact of climate change on discharge (figure 1(b)–(d)) but not in the runoff maps (figure 1(a)). Examples include the Amazon, São Francisco (northeastern Brazil), Nile, Zambezi, Mekong and Siberian rivers where changes in upstream runoff are obviously carried downstream. The effect of flow accumulation and lateral routing can be analyzed by comparing the impact of climate change on mean runoff (figure 1(a)) with the impact of climate change on mean discharge that would occur if there were no human withdrawals (figure 1(b)). In the ECHAM4 A2 scenario, for example, mean river discharge of the Amazon in the downstream (eastern) section would increase even though runoff in these grid cells would decrease. Discharge change reflects more strongly the increased runoff in the upstream (western) part of the basin. In the HadCM3 A2 scenario, both discharge and runoff decrease in the downstream section of the Amazon, but discharge less than runoff as runoff decrease in the upstream part of the basin is smaller than in the downstream part. In this scenario, it is the São Francisco in northeastern Brazil that is predicted to have an increased mean river discharge in its downstream section even though runoff is reduced very strongly there (figure 1(b)). At a smaller spatial scale, the strong runoff decrease at the North American Pacific coast, for the HadCM3 A2 scenario, translates into a smaller decrease of discharge due to more runoff in the headwater area.

Actual mean annual river discharge can be much smaller than natural mean annual discharge that results from the aggregation of upstream runoff. Around the year 2000, mean annual discharge had decreased by more than 10% on one sixth of the global land area (excluding Antarctica and Greenland) compared to natural conditions without man-made reservoirs and water withdrawals (Döll 2009). As the same absolute changes to lower values result in higher per cent changes, per cent changes of actual river discharge due to climate change (figure 1(c)) are higher than per cent changes of river discharge that would occur if there were no human water withdrawals (figure 1(b)). For example, if river discharge is decreased by 50% due to water withdrawals, any per cent flow increase due to climate change would double as compared to considering natural flows. Due to the coarse change classes in figure 1, this effect is visible only in areas with strong anthropogenic discharge reductions (e.g. central USA and Canada, Spain, Turkey). For 7% (ECHAM A2) and 5% (HadCM A2) of the global land area (excluding Antarctica and Greenland), per cent changes of river discharge altered by human water withdrawals are at least 10 percentage points higher (e.g. 30% instead of 20%) than changes of discharge that is assumed to be unaffected by water withdrawals.

3.2. Climate change impacts on low and high flow

The broad global patterns of climate change impact on monthly low flows Q_{90} (figure 1(d)) are similar to the patterns for mean annual discharge Q_{mean} (figure 1(c)). The pattern of change of Q_{90} derived from the output of either of the two climate models is more similar to the pattern of change of Q_{mean} computed with the same climate model output than to the pattern of change of Q_{90} derived from the output of the other climate model. However, at smaller scales, there are significant differences between climate-induced changes of mean annual discharge and low flows. Per cent decreases

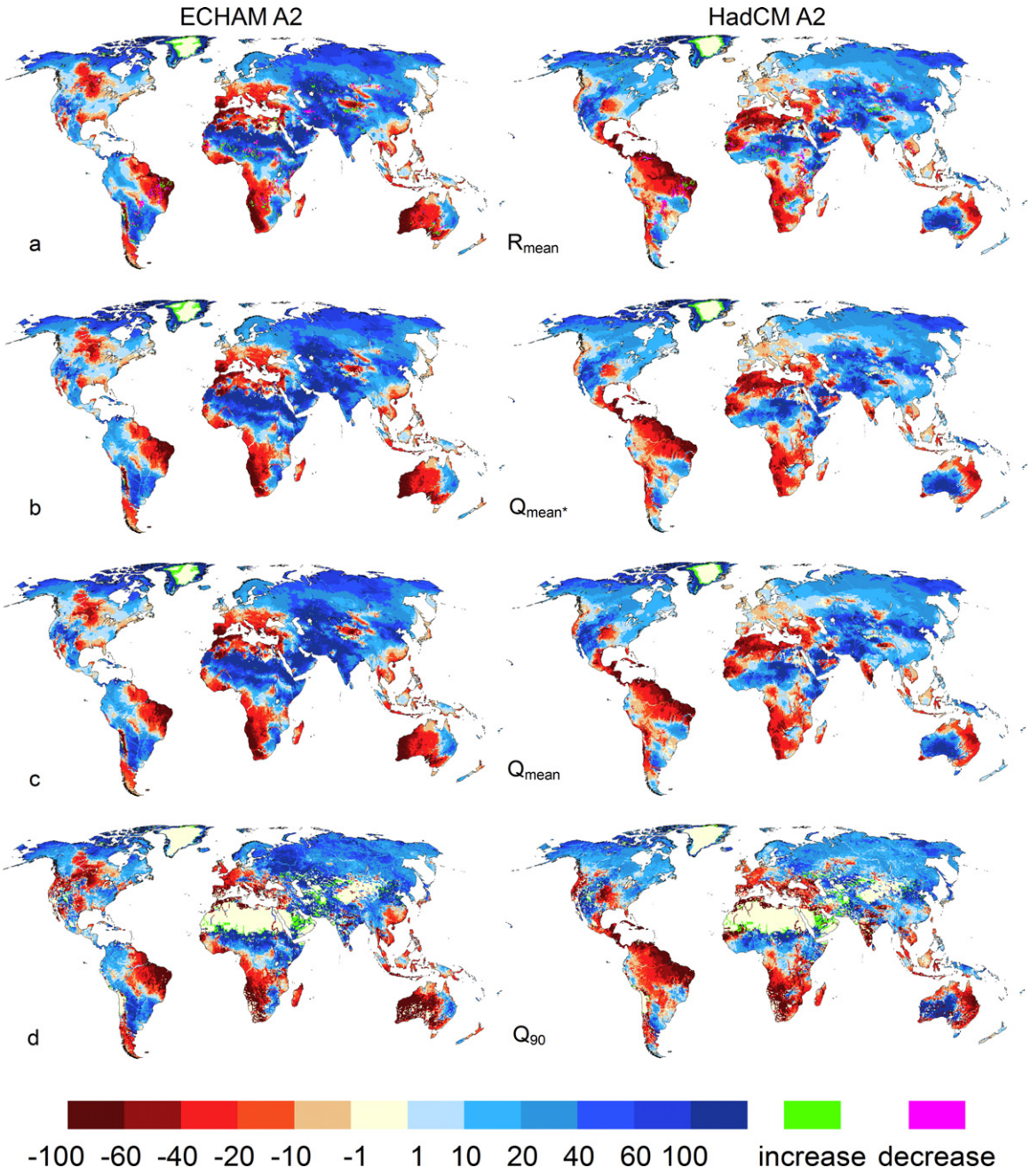


Figure 1. Impact of climate change on MAR R_{mean} (a), on mean annual river discharge Q_{mean*} that would occur without water withdrawals (b), on mean annual river discharge Q_{mean} (with water withdrawals) (c) and on low flows Q_{90} (with water withdrawals) (d) by the 2050s. Per cent changes between 1961–90 and 2041–70, as computed by WaterGAP applying climate change scenarios computed by the climate models ECHAM4 and HadCM3, each interpreting the IPCC greenhouse gas emissions scenario A2. For cells where MAR (a) is less than or equal to zero in 1961–90, runoff changes are indicated by green in the case of increasing future runoff (or less negative difference between precipitation and actual evapotranspiration) and pink in the case of decreasing (more negative) future runoff, while no change is represented by light yellow (class −1 to 1). For cells where discharge is equal to zero in 1961–90, increasing future discharge is indicated by green, while no change is represented by light yellow ((b)–(d)).

of Q_{90} are often higher than per cent decreases of Q_{mean}, e.g. in Australia and South America, and there are areas where Q_{90} decreases, while Q_{mean} increases (e.g. in Great Britain for ECHAM4 A2 or in western France for HadCM3 A2, figures 1(c) and (d)). The strong differences between the climate-induced per cent changes of mean annual flows and low flows become obvious if per cent change of Q_{90} as a fraction of per cent change of Q_{mean} is considered (relative change, figure 2(a)). There are not many cells where the per cent changes of Q_{mean} and Q_{90} are rather similar (green color). Predominantly, the per cent change of Q_{90} is larger than that of Q_{mean} (dark yellow). Where Q_{mean} decreases in the future (indicated by cross-hatching in figure 2), this means that the relative discharge variability increases. For ECHAM4 A2, for example, this is the case in southern and northeastern South America, southern Africa, most of Australia, France

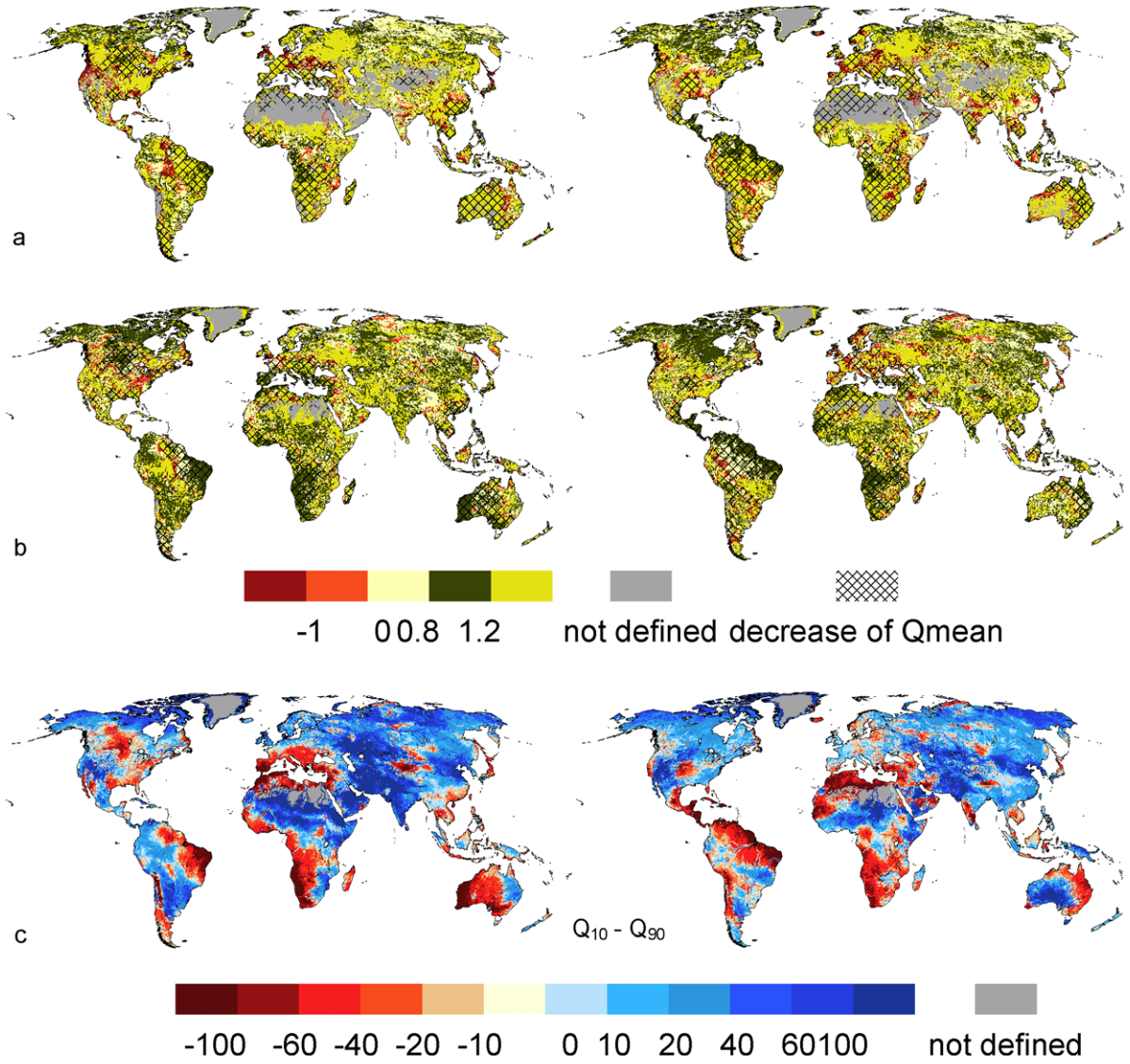


Figure 2. Climate change impact on low flows Q_{90} (a) and high flows Q_{10} (b) as compared to the climate change impact on mean annual river discharge Q_{mean}, and climate change impact on monthly flow variability (c), for climate scenarios ECHAM A2 and HadCM A2. Per cent changes of Q_{90} (Q_{10}) as a fraction of per cent changes of Q_{mean} between 1961–90 and 2041–70. Negative values indicate that the direction of change differs between Q_{mean} and Q_{90} (Q_{10}) ((a) and (b)). Per cent change of Q_{10} minus Q_{90} between 1961–90 and 2041–70 (c).

and South China (figure 2(a) left). Where Q_{mean} increases, dark yellow in figure 2(a) indicates that discharge variability is reduced. According to both climate models, Q_{90} increases more strongly than Q_{mean} in the eastern USA, Scandinavia and the European part of Russia. There, winter low flows will likely increase due to more winter precipitation as rain. However, in most of the Asian part of Russia and in some parts of Canada Q_{90} increases less then Q_{mean} (light yellow).

On a significant fraction of land all around the globe, the directions of change differ between Q_{90} and Q_{mean} (orange and red colors in figure 2(a)). Q_{90} is computed to decrease but Q_{mean} to increase in northern India, parts of the USA, the central part of South America and the northern Nile upstream of the Assuan High Dam (depending on the climate scenario). Cells where Q_{90} increases and Q_{mean} decreases are more rare and are located, among others, in Europe, e.g. in the Alps where winter low flows strongly increase due to higher temperatures and less water storage as snow.

The relative climate change patterns of the monthly high flows Q_{10} are similar to those of Q_{90} with respect to the dominance of grid cells with the same direction of change, but the percentage changes of Q_{10} and Q_{mean} are more similar, in particular where Q_{mean} decreases (figure 2(b)). Where Q_{mean} increases, the per cent change of Q_{10} is predominantly larger than that of Q_{mean} (dark yellow), indicating more strongly increased high flow discharge values. Grid cells with opposite signs of change for Q_{10} are as dispersed as in the case of Q_{90}. In some parts of Siberia, where Q_{mean} increases due to increased annual precipitation, decreased snow accumulation during winter leads to an absolute (red/orange) or relative (light yellow) decrease of Q_{10}. Q_{10} increases in the upstream areas of the Amazon even though Q_{mean} decreases significantly (HadCM3 A2, figure 1(c)). Where Q_{90} and Q_{mean} show opposite signs of change, Q_{10} and Q_{mean} mostly change in the same direction (e.g. western USA and India), and vice versa (e.g. eastern USA, compare

Table 1. Land area (in per cent of total land area of the Earth excluding Antarctica and Greenland) affected by changes of long-term average runoff and river flow regime indicators between 1961–90 and 2041–70, computed by WaterGAP applying climate change scenarios from the ECHAM4 and HadCM3 climate models, each interpreting the IPCC greenhouse gas emissions scenarios A2 and B2. In each field, the range of land area due to the use of the two climate models is given. The extreme classes of changes were chosen in order to see areas where values are at least halved (−50%) or doubled (>100%). The column 'not defined' refers to the area where the 1961–90 values of indicators are less than (only for R_{mean}) or equal to zero. P: mean annual precipitation (climate model output downscaled with delta change method). R_{mean}: MAR. Q^*_{mean}: mean annual river discharge (without water withdrawals). Q_{mean}: mean annual river discharge. Q_{90}: monthly discharge that is exceeded in nine out of ten months (low flow); here, the land area in the −10 to 10% change range also includes the 13–14% of the land area where both current and future Q_{90} is zero. The 'not defined' area refers to the 1–2% of the global land area where future Q_{90} stops being zero. Q_{10}: monthly discharge that is exceeded in one out of ten months (high flow). Q_{DJF}: long-term average river discharge in December to February season. Q_{MAM}: long-term average river discharge in March to May season. Q_{JJA}: long-term average river discharge in June to August season. Q_{SON}: long-term average river discharge in September to November season.

Change (%)	−100 to −50		−50 to −10		−10 to 10		10–100		>100		Not defined
Scenario	A2	B2	A2	B2	A2	B2	A2	B2	A2	B2	
P	0–0	0–0	13–15	11–15	44–53	48–56	31–41	29–40	0–2	0–2	0
R_{mean}	5–5	4–4	17–18	16–19	21–26	23–29	42–43	42–44	5–9	3–8	4
Q^*_{mean}	4–5	3–4	18–19	16–19	22–27	24–31	45–46	43–47	4–9	3–8	1
Q_{mean}	5–5	3–4	18–19	16–20	22–27	23–31	44–44	42–46	5–10	3–9	1
Q_{90}	8–9	6–7	16–18	15–18	28–32	29–35	34–36	33–38	6–9	5–10	1–2
Q_{10}	5–5	4–5	18–18	16–18	22–26	23–29	41–42	42–43	6–11	4–11	2
Q_{DJF}	6–7	6–7	17–18	16–17	18–25	19–28	42–42	41–45	7–14	5–12	2
Q_{MAM}	5–5	4–4	20–20	18–18	18–22	18–23	41–41	43–46	10–16	7–14	1
Q_{JJA}	6–7	4–5	24–27	22–26	24–29	26–30	30–34	33–35	5–10	5–11	2
Q_{SON}	8–9	6–7	20–26	18–29	19–22	19–23	36–40	35–44	6–13	5–11	2

figures 2(a) and (b)). The difference between Q_{10} and Q_{90} can be regarded as a measure of temporal variability of monthly river flows, with 8 out of 10 months having discharge within this range (figure 2(c)). It is strongly correlated with per cent change of Q_{mean}, with some exceptions in snow-influenced regions.

Globally, Q_{90} increases less with climate change than Q_{mean}. By the 2050s, Q_{90} will have increased by 10–100% on only 34–36% for emissions scenario A2 (33–38% for B2) of the land area while the corresponding area for Q_{mean} is 44–44% (43–46%) (table 1). The area where Q_{90} decreases to less than 50% of its 1961–90 value is almost twice as large as the corresponding area for Q_{mean}. Differently from Q_{90}, the land area fractions of changes in Q_{10} are very similar to those of Q_{mean} (table 1). Besides, the land area fractions of changes in Q_{mean} are very similar to those of R_{mean}, but not to those of precipitation P (table 1). On about half the global land area, precipitation is projected to change by less than 10%, while this is the case for only about a quarter of the global land area for MAR and discharge (table 1). Extreme decreases by more than 50% are not projected at all for precipitation.

3.3. Climate change impacts on seasonal river discharge

Figure 3 shows the very heterogeneous pattern of per cent change of seasonal river discharges as a fraction of the per cent change of Q_{mean} (relative seasonal discharge changes). Climate change will lead to strong changes in the seasonality of discharge almost everywhere. Only on a very small part of the land area will seasonal discharge change approximately like annual discharge (shown in dark green in figure 3). For both climate models, seasonality will remain stable with increasing Q_{mean} in parts of Canada and Australia (while, for Australia, regions with increasing discharge strongly differ between the two climate models, compare figure 1(c)). In northeastern Brazil and southern Africa, seasonality will remain stable with strongly decreasing Q_{mean}. In most regions north of 30–35°N (southern rim of the Mediterranean Sea), summer discharge Q_{JJA} is projected to decrease. In areas with decreasing Q_{mean} (like in central Europe), Q_{JJA} will decrease more (dark yellow in figure 3), while even in many areas with increasing Q_{mean}, like in northern Europe, Q_{JJA} will decrease (orange and red in figure 3). For winter flows, the opposite is visible (figure 3). In Mexico, a shift from winter/spring to autumn discharge can be recognized (in particular for ECHAM4 A2). In South Asia, Q_{JJA} is projected to increase more strongly than Q_{mean} due to a strengthening of the monsoon, while Q_{DJF} may even decline in some areas.

Patterns of relative seasonal river discharge changes are related to patterns of relative low and high flow changes (figures 2(a) and (b)). If, for example, low flows during 1961–90 occur in northern hemisphere winter, like in the Alps and the European part of Russia, both Q_{DJF} and Q_{90} increase due to increased temperatures and thus rainfall and runoff in winter. In Siberia, where the temperature rise is not sufficient to convert snowfall into rainfall, the increases of both Q_{DJF} and Q_{90} are much smaller. The decrease of spring discharge in eastern USA (with increasing Q_{mean}, figure 3) is connected (at least for ECHAM4 A2) with a clear increase in Q_{90} (figure 3). Globally, more land area will suffer from decreased Q_{JJA} than from decreased Q_{mean} (table 1). This is also true for Q_{SON} but here the two climate models differ more strongly.

3.4. Shifts of river flow regimes from perennial to intermittent or vice versa due to climate change

To assess the capability of WaterGAP to identify the flow regime types 'perennial' and 'intermittent', we compared the

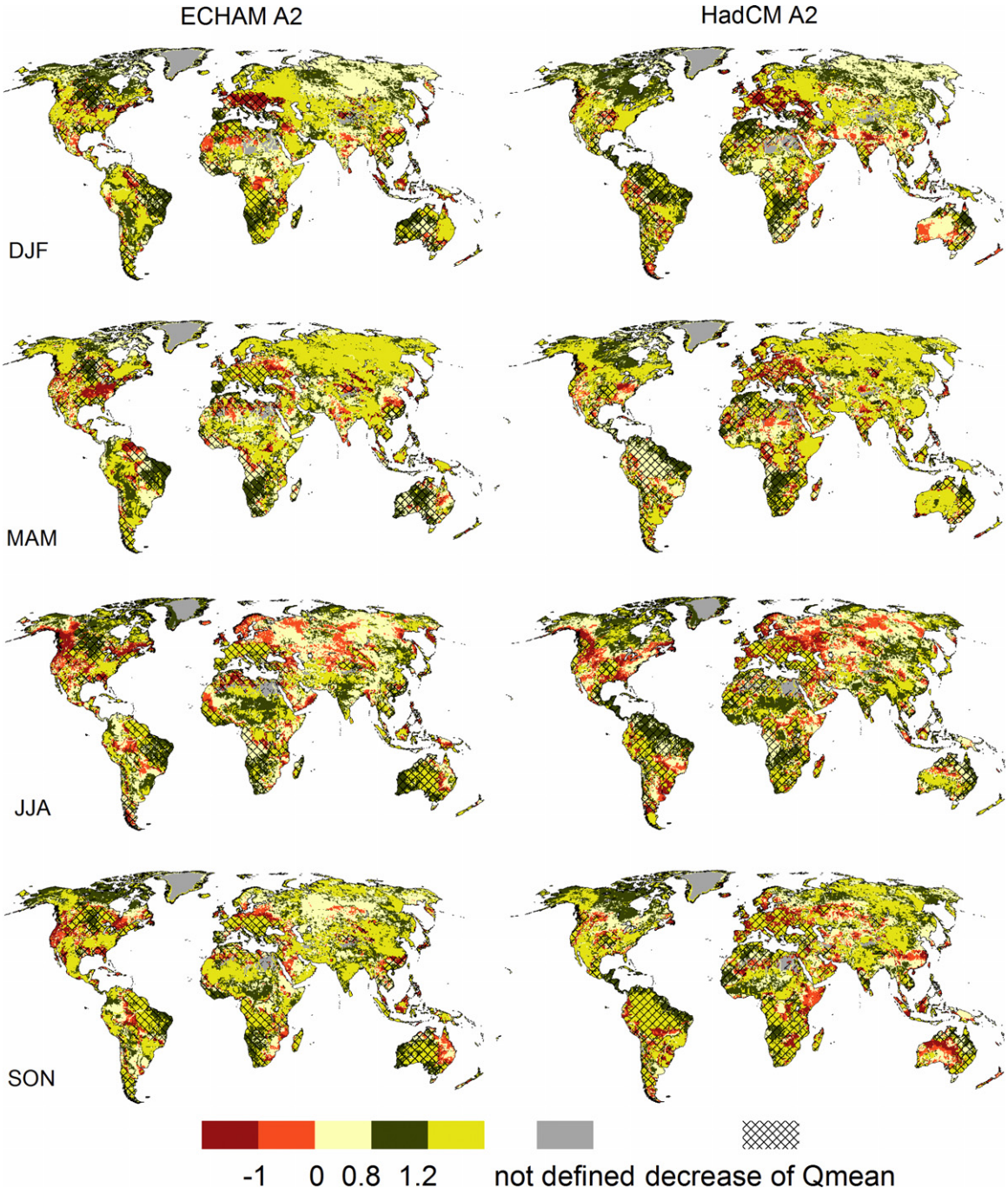


Figure 3. Climate change impact on mean seasonal river flows as compared to the climate change impact on long-term average annual river discharge Q_{mean} for climate scenarios ECHAM A2 and HadCM A2. Per cent changes of mean river discharge from December to February (DJF), March to May (MAM), June to August (JJA) or September to November (SON) are shown, as a fraction of per cent changes of Q_{mean}, between 1961–90 and 2041–70. Negative values indicate that the direction of change differs between Q_{mean} and the seasonal discharges.

WaterGAP map with information from maps as captured in the VMAP0 and ESRI Big Rivers data sets. The spatial patterns derived from WaterGAP are similar to the patterns on the maps. WaterGAP appears to underestimate intermittency in northeastern Brazil, northern, northeastern and central Australia, Ethiopia and central India, while it identifies intermittent rivers in Spain, different from the maps. However, in northeastern Brazil, many large rivers have been made perennial by dams. While dam effects on river flows are simulated by WaterGAP, the impact of dams may not be reflected in maps, in particular if they predate dam construction. The intermittent grid cells identified by WaterGAP in Siberia are caused by a lack of water flows during the cold period, and Russian maps might not label such arctic rivers as intermittent.

The global pattern of flow regime shifts under the impact of the HadCM3 A2 climate scenario (figure 4) reflects both the current flow regime and the future climate changes. Flow regime shifts from perennial to transitional or intermittent until the 2050s are projected, for example, for

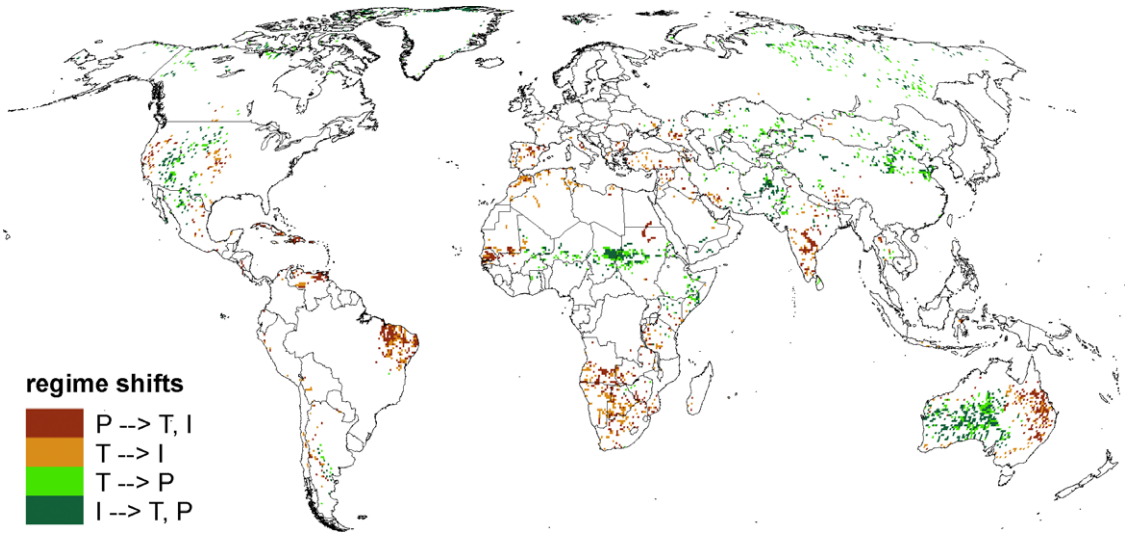


Figure 4. Regime shifts among perennial (P), transitional (T) and intermittent (I) river flow regimes occurring between 1961–90 and 2041–70 due to climate change (as computed with WaterGAP, climate scenario HadCM3 A2).

California, the Caribbean, southern Africa, West Africa and around the Mediterranean Sea (figure 4). Such shifts are also computed for northeastern Brazil and northeastern Australia but, as discussed above, the current regime may already be intermittent (misclassified by WaterGAP). Flow regime shifts to the perennial regime may occur in western USA, parts of China, western and central Australia and on the southern rim of the Sahara. Flow is computed to become perennial in parts of Siberia, Canada and Alaska due to warmer winters. There is certainly a significant correlation between changes in R_{mean} and flow regime shifts under a given climate scenario (compare figure 1(a) right-hand side to figure 4). An even higher correlation is observed between changes in statistical low flow Q_{90} (figure 1(d) right-hand side) and flow regime shifts as Q_{90} takes into account changes in temporal variability (consider, for example, Spain and northern India). With the HadCM3 A2 climate scenario, 6.3% of the global land area (except Greenland and Antarctica) will have experienced flow regime shifts by the 2050s, with the corresponding impacts on freshwater-dependent biota. On 1.8% of the land area, the flow regime will no longer be perennial, and also on 1.8% it will no longer be intermittent. Transitional regimes shift to intermittent ones on 1.4% of the land area, and to perennial ones on 1.3%. Extreme regime shifts from perennial to intermittent, where the fraction of months with (almost) no discharge increases from less than 3% to more than 11% (section 2.3), occur on 0.4% of the land area, e.g. in Venezuela, the Caribbean and India. The reverse occurs on 0.2% of the land area, e.g. in China, Pakistan and the Sahel. With the ECHAM4 A2 scenario, flow shifts are identified on 7.0% of the global land area, with dominant shifts to wetter conditions. In the case of B2 emissions, flow shifts are projected to occur on only 5.4% (HadCM3) or 6.7% (ECHAM4) of the global land area. In particular, the area where flow will no longer be perennial would be particularly reduced. This is true for both climate models.

4. Discussion

It is well known that different global climate models translate the same greenhouse gas emissions scenario into strongly differing climate change scenarios, with precipitation changes being more uncertain than temperature changes (Bates *et al* 2008). Therefore, the presented climate change impacts, which are based on the scenarios of only two climate models, can only be considered to be illustrative but do not show the full range of uncertainty. The aim of this study was not to derive quantitative estimates of climate change impact but to improve our understanding of the relation between the impact of climate change on MAR and the impact of climate change on river flow regime indicators that are more relevant for water management purposes. In this section, we wish to discuss the quality of the modeled impacts on the river flow regime for a given climate change scenario, considering the capability of WaterGAP to simulate river flow regimes under historical climate (section 4.1), and the way in which GCM scenarios are translated into climate input for WaterGAP (section 4.2). In addition, we discuss the relevance of the computed flow regime indicators for the rivers within each grid cell (section 4.3).

4.1. Quality of the modeled flow regime indicators as related to the quality of the hydrological model

Because WaterGAP is tuned against mean annual discharge, it can model MAR and mean annual discharge reasonably well for most grid cells that are located within the 1235 tuning basins covering 48.7% of the global land area excluding Greenland and Antarctica (Hunger and Döll 2008). However, model results for individual grid cells can be completely wrong, in particular if they are in semi-arid/arid areas outside calibration basins (e.g. in western and central Australia). Model calibration also leads to more realistic values for statistical monthly Q_{90} and Q_{10} and mean monthly discharges, but this is mainly caused by adjustment of the mean and not of

the variance. Nevertheless, monthly Q_{90} and Q_{10} were shown to be estimated quite well for most observation stations (Döll *et al* 2003). A comparison of observed and modeled flow regime indicators for four river discharge gauging stations (not shown) indicated that Q_{90} and Q_{10} are modeled quite well, while mean monthly and seasonal discharges may suffer from an inadequate simulation of snow and ice, or water storage in surface water bodies.

4.2. Quality of the modeled flow regime indicators as related to the translation of GCM climate change scenarios into input for the hydrological model

It cannot be assessed how well WaterGAP (or any other hydrological model) can translate climate change to change in runoff and discharge, as appropriate observations are still lacking. Certainly, consideration of the method to produce climate input for WaterGAP under conditions of changed climate is crucial for understanding the uncertainty of the computed flow regime changes. The applied delta change method does not take into account future changes in the interannual variability of climate variables, as mean monthly changes are added to (or multiplied by) time series of observed climate. This will likely lead to an underestimation of the changes in Q_{90} and Q_{10}, and of shifts between perennial and intermittent flow regimes, as interannual climate variability is projected to increase in the future on all temporal scales (Bates *et al* 2008). The delta change method can also lead to implausible shifts in the seasonality of precipitation if the climate model does not model the observed seasonality of precipitation reasonably well. This affects the analysis of seasonal river flows. Alternative methods may avoid these problems but have other drawbacks. Statistical bias correction of climate model output against observed climate, for example, can represent future change of climate variability. However, bias correction alters the climate change signal for specific locations and months (Hagemann *et al* 2011). For most river basins, the long-term average temperature change was found to decrease by bias correction as compared to the original climate model runs, and the precipitation change was found to increase for most river basins (Hagemann *et al* 2011). The uncertainty due to statistical bias correction may be of the same order of magnitude as the uncertainty related to the choice of the GCM or the applied global hydrological model (Hagemann *et al* 2011). An additional constraint of our study is that we did not take into account changes of radiation, humidity or wind speed.

4.3. Relevance of the computed river flow regime indicators

As runoff can be conceptualized as a vertical flow of water per unit land area, with units of, for example, mm yr^{-1}, the runoff computed for any 0.5° grid cell can be regarded as an average value of the grid cell. In contrast, river discharge is defined as the lateral flow of water along a river channel. In the case of global-scale models, it is assumed that there exists exactly one river channel in each grid cell, while in reality there may be many (in particular small tributaries to one main river). While computed grid cell discharge can be thought of as the sum of the discharges in all river channels within the grid cell, the temporal variability of river discharge is expected to vary strongly among the diverse river channels. For example, small tributaries are expected to have a much higher temporal variability than the main river, e.g. a larger Q_{90}-to-Q_{mean} ratio. Regarding the definition of perennial and intermittent flow regimes, a grid cell identified as perennial in WaterGAP may contain small intermittent tributaries. This has to be taken into account when a map of regime shifts like the one shown in figure 4 is interpreted. Computed river flow regime indicators and their changes should be assumed to be relevant only for the main river within each grid cell.

5. Conclusions

How is the impact of climate change on river flow regimes related to the impact on mean annual runoff (MAR)? With the exception of seasonal flows, climate-induced future changes of the considered river flow regime indicators broadly follow the spatial pattern of increases and decreases of MAR generated in the grid cell. However, there are important differences at the sub-basin and grid cell scales. In grid cells where river flow is fed from upstream cells, the impact of climate change on mean river discharge represents the aggregated response of upstream runoff to climate change, and differs from the impact on MAR. Downstream reaches of major rivers often show an opposite sign of change with respect to MAR and mean annual discharge. Where natural river discharge has been reduced significantly due to human water withdrawals, relative changes (expressed in per cent change) of actual mean river discharge are higher than relative changes of natural mean river discharge or mean runoff. On about 5% of the global land area, the impact of climate change by the 2050s would be underestimated by at least 10 percentage points if the influence of water withdrawals were neglected.

MAR is projected to increase by more than 10% on 50% of the global land area (excluding Antarctica and Greenland) by the 2050s. This is also true for mean annual river discharges and high flows but not for low flows. The area where low flows will increase by more than 10% is approximately eight percentage points smaller. On a significant part of the land area, low flows may decrease even though mean annual river discharges increase.

Climate-induced changes of seasonal river flows are not correlated with changes of MAR. Our global-scale study confirms projected changes of seasonal river flow dynamics in basins where winter precipitation currently is dominated by snowfall. In these basins, winter or spring flows are likely to increase, while summer flows are likely to decrease. In most regions north of 35°N, summer discharge is projected to decrease. Of all four seasonal discharges, June–July–August discharge shows the largest global land fraction with significant river flow decreases and the smallest fraction with significant flow increases, even though in the monsoon areas it is projected to increase.

Flow regime shifts among perennial, transitional and intermittent regimes indicate strong changes in habitat conditions for freshwater biota and therefore a strong impact of climate change on freshwater ecosystems. Flow regime shifts by the 2050s may occur on 6.3–7.0% (A2) or 5.4–6.7% (B2) of the global land area, mainly in semi-arid areas as well as in some cold areas (from intermittent to transitional or transitional to perennial only) where during 1961–90 there was (almost) no river flow in the winter months due to freezing. Shifts from perennial to intermittent (intermittent to perennial) flow regimes correlate with decreases (increases) in mean runoff, but even more with decreases (increases) in statistical low flows.

This study has improved our understanding of the relation between climate-induced changes of MAR, the major indicator of the impact of climate change on freshwater resources, and river flow regime indicators that are relevant for the well-being of humans and freshwater ecosystems. We found that the differences between changes of MAR related to the use of two different climate models are still larger than the differences between the change of MAR and the change of the investigated river flow indicators (except for seasonal discharge). If larger ensembles of climate models were considered, the spread of computed MAR changes would be even larger (Gosling *et al* 2010). Therefore, broad conclusions about climate-induced changes of river flow regimes can be derived from ensembles of mean annual and seasonal runoff. Unfortunately, it seems impossible to define general rules for translating changes of MAR into changes of river flow regime indicators because the relation between these changes depends on the climate model applied.

Acknowledgments

We thank the two reviewers who helped us to improve the quality of our paper, in particular to focus and shorten it.

References

Alcamo J, Döll P, Henrichs T, Kaspar F, Lehner B, Rösch T and Siebert S 2003 Development and testing of the WaterGAP 2 global model of water use and availability *Hydrol. Sci. J.* **48** 317–38

Bates B C, Kundzewicz Z W, Wu S and Palutikof J P (ed) 2008 Climate change and water *Technical Paper of the Intergovernmental Panel on Climate Change (IPCC Secretariat)* p 210

Bond N, McMaster D, Reich P, Thomson J R and Lake P S 2010 Modelling the impacts of flow regulation on fish distributions in naturally intermittent lowland streams: an approach for predicting restoration responses *Freshwater Biol.* **55** 1997–2010

Döll P 2009 Vulnerability to the impact of climate change on renewable groundwater resources: a global-scale assessment *Environ. Res. Lett.* **4** 035006

Döll P and Fiedler K 2008 Global-scale modeling of groundwater recharge *Hydrol. Earth Syst. Sci.* **12** 863–85

Döll P, Kaspar F and Lehner B 2003 A global hydrological model for deriving water availability indicators: model tuning and validation *J. Hydrol.* **270** 105–34

Döll P and Lehner B 2002 Validation of a new global 30 min drainage direction map *J. Hydrol.* **258** 214–31

Döll P and Siebert S 2002 Global modeling of irrigation water requirements *Water Resour. Res.* **38** 1037–46

Döll P and Zhang J 2010 Impact of climate change on freshwater ecosystems: a global-scale analysis of ecologically relevant river flow alterations *Hydrol. Earth Syst. Sci.* **14** 783–99

Gordon C, Cooper C, Senior C A, Banks H, Gregory J M, Johns T C, Mitchell J F B and Wood R A 2000 The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments *Clim. Dyn.* **16** 147–68

Gosling S N, Bretherton D, Haines K and Arnell N W 2010 Global hydrology modelling and uncertainty: running multiple ensembles with a campus grid *Phil. Trans. R. Soc.* A **368** 1–17

Gosling S N, Taylor R G, Arnell N W and Todd M C 2011 A comparative analysis of projected impacts of climate change on river runoff from global and catchment-scale hydrological models *Hydrol. Earth Syst. Sci.* **15** 279–94

Hagemann S, Chen C, Haerter J O, Heinke J, Gerten D and Piani C 2011 Impact of a statistical bias correction on the projected hydrological changes obtained from three GCMs and two hydrology models *J. Hydrometeorol.* **12** 556–78

Hirabayashi Y, Kanae S, Emori S, Oki T and Kimoto M 2008 Global rojections of changing risks of floods and droughts in a changing climate *Hydrol. Sci. J.* **53** 754–73

Hunger M and Döll P 2008 Value of river discharge data for global-scale hydrological modeling *Hydrol. Earth Syst. Sci.* **12** 841–61

Kundzewicz Z W, Mata L J, Arnell N W, Döll P, Kabat P, Jiménez B, Miller K A, Oki T, Sen Z and Shiklomanov I A 2007 Freshwater resources and their management *Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* ed M L Parry, O F Canziani, J P Palutikof, P J van der Linden and C E Hanson (Cambridge: Cambridge University Press) pp 173–210

Milly P C D, Wetherald R T, Dunne K A and Delworth T L 2002 Increasing risk of grate floods in a changing climate *Nature* **415** 514–7

Mitchell T D and 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

Nakicenovic N and Swart R (ed) 2000 Emission scenarios *IPCC Special Report on Emission Scenarios* (Cambridge: Cambridge University Press) www.ipcc.ch

Poff N L and Zimmerman J K H 2010 Ecological responses to altered flow regimes: a literature review to inform the science and management of environmental flows *Freshwater Biol.* **55** 194–205

Röckner E, Arpe K, Bengtsson L, Christoph M, Claussen M, Dümenil L, Esch M, Giorgetta M, Schlese U and Schulzweida U 1996 The atmospheric general circulation model ECHAM-4: Model description and simulation of present day climate *MPI-Report* (Hamburg: MPI für Meteorologie) No 218

Rudolf B and Schneider U 2005 Calculation of gridded precipitation datafor the global land-surface using *in situ* gauge observations *IPWG: Proc. 2nd Workshop of the Int. Precipitation Working Group (Monterey, October 2004)*

Siebert S, Döll P, Hoogeveen J, Faures J-M, Frenken K and Feick S 2005 Development and validation of the global map of irrigation areas *Hydrol. Earth Syst. Sci.* **9** 535–47

Stromberg J C, Bagstad K J, Leenhouts J M, Lite S J and Makings E 2005 Effects of stream flow intermittency on riparian vegetation of a semiarid region river (San Pedro River, Arizona) *River Res. Appl.* **21** 925–38

USGS 2000 *The National Hydrography Dataset: Concepts and Contents* http://nhd.usgs.gov/chapter1/chp1_data_users_guide.pdf (last visited 13 September 2011)

Wilhelm F 1997 *Hydrogeographie* (Braunschweig: Westermann)

8 Answers to the research questions

In the following, and based on the six papers (Chapters 2–7) as well as Appendix A1–3, the research questions from Chapter 1.5 are answered. Furthermore, and where appropriate, an outlook for future model development or further studies is given.

8.1 Model evaluation

The first part of the research topic deals with the evaluation of the global hydrological model WaterGAP and is divided into two parts.

8.1.1 Evaluation of net radiation and its components with station-based observations

Net radiation as the main driver of the calculation of potential evapotranspiration is subject to parameter uncertainty (e.g., land-cover-dependent attributes) as well as uncertainty of climate input. Within paper 1, a thorough assessment of radiation components as simulated by WaterGAP (or taken from climate input data) was carried out via comparison to station-based observations. Three main questions (RQs 1–3) were answered in this context.

RQ 1 How is the performance of simulated upward radiation components of WaterGAP compared to observations?

WaterGAP considers land cover dependent values for albedo and emissivity together with climate input variables to calculate the upward radiation components. Within paper 1, WaterGAP output as forced with two climate input data sets (WFDEI, PGFv2) and two setups of ERA-Interim reanalysis was compared to observation data from 16 stations of the Baseline Surface Radiation Network (BSRN, both short- and longwave upward fluxes) as well as to data from 43 (shortwave) and 14 (longwave) stations of the Global Energy Balance Archive (GEBA). Two efficiency metrics (one absolute, one relative) were used to assess model performance. Upward shortwave radiation is directly dependent on downward shortwave radiation and is only affected by surface albedo, which varies in WaterGAP with land cover type as well as snow dynamics, indicating that upward flux cannot be evaluated without considering downward flux. For shortwave downward flux, a systematic overestimation of the mean value was found together with very high correlations (close to 1 with a bit more spread for GEBA stations) and a slight underestimation of variability. For shortwave upward radiation, the Kling–Gupta efficiency metric varies greatly between the reference data (for BSRN around 0.7, for GEBA around 0.5) and is in general relatively low. There is no systematic over- or underestimation, but the spread is the largest among all radiation components. Hence, the performance of

shortwave upward flux depends strongly on the specific locations of the observation stations in terms of mean bias. Obviously, there is a mismatch of albedo values for the 0.5° grid cell with those of the station observations, which is reasonable, as the measurement footprint of the observations is small (and in many cases artificially constrained to a standard vegetation) and the land cover heterogeneity of the grid cell cannot be represented. Mean correlations are moderate (ranging from 0.8 to 0.9), whereas the mean variability of the simulations is lower than those of the observations, which is consistent with the underestimated variability of shortwave downward radiation of the two climate input data and ERA-Interim. As the albedo value is already coupled to snow dynamics in WaterGAP, there is little room for improvements in the current model structure. Nevertheless, albedo varies only with snow dynamics, and it could be tested whether a dynamic albedo variation (e.g., dependent upon vegetation or soil saturation dynamics) leads to improved model results in terms of mean bias and correlations.

Upward longwave radiation is simulated by considering a land-cover-based (but static) emissivity value and temperature provided by the climate input data set. Kling–Gupta efficiency metrics are around 0.9 and thus the highest among the radiation components. The mean bias indicator is slightly below 1 and nearly without variation; hence simulations tend to underestimate longwave upward radiation systematically by a small amount. Higher emissivity values of WaterGAP could thus improve modeling of this component. Correlation is 1 for BSRN stations and close to 1 for GEBA stations. In contrast, the variability of simulated upward longwave radiation flux is lower than that of the observations. This is likely related to the static emissivity values taken into account in WaterGAP simulations. Hence, a coupling of emissivity values to snow dynamics (see the discussion in paper 1) could help to improve the variability indicator at least for the stations in areas where snow falls.

RQ 2 Is it beneficial to substitute standard climate input data and calculation of upward radiation components by ERA-Interim reanalysis?

Due to a closed energy balance and the additional observations that are included via data assimilation techniques, the usage of reanalysis is of interest not only for downward radiation components but also for the upward components. In paper 1, two setups of ERA-Interim reanalysis were compared to station observations as well as WaterGAP simulations forced with two standard climate input data sets (WFDEI, PGFv2). Within ERAID, only downward radiation components from ERA-Interim were taken into account, whereas in ERAIN the upward components were also considered.

Upward shortwave radiation of ERAIN does not perform better than that from WFDEI. ERAIN underestimates mean value, and PGFv2 variability performs better than ERAIN (WFDEI is close to ERAIN for BSRN). There is no difference for correlations. In summary, the

Kling–Gupta efficiency metric does not vary among ERAIN, PGFv2 and WFDEI. Even though the mean absolute bias of ERAIN is slightly lower than of the WaterGAP calculations, a clear improvement is not visible.

Longwave upward radiation of ERA-Interim performs better than the WaterGAP estimates for all indicators and reference data sets (except mean absolute bias and GEBA stations). This can be for two reasons. Firstly, surface temperature within ERA-Interim could be closer to that of the station data. However, this is not very likely as temperature is a frequently (in space and time) measured meteorological variable and also basis for the climate input data. Furthermore, it was found that temperature is not the main driver of longwave upward radiation. Secondly, emissivity values of WaterGAP could be too low compared to ERA-Interim. This would explain the slight underestimation of the mean value in comparison to the observations. Here, a land-cover-specific comparison of simulated and observed upward radiation could be helpful. The variability performance indicator of ERA-Interim longwave upward radiation matches the observations but is underestimated in the case of WaterGAP calculations. The static emissivity value of WaterGAP could be a reason for this, and efforts to, e.g., couple emissivity with snow dynamics could improve the variability of the simulations.

However, the low amount of available station data measurements and the questionable representativeness compared to the 0.5° grid cell (see discussion in paper 1) hamper a general assessment of the radiation fluxes. Even though longwave upward radiation could be improved by using ERA-Interim reanalysis data, the consistency of WaterGAP simulations is lost when that is done. For example, the snow dynamics are simulated within WaterGAP and coupled to the albedo value. At least for regions with relevant snow dynamics, different states of this compartment (ERA-Interim vs. WaterGAP) will lead to inconsistencies and could influence water resources modeling.

Long-term average potential evapotranspiration (PET) values increase when upward radiation components are used from ERA-Interim reanalysis, especially in energy-limited regions (7% higher than for WFDEI climate input data). Those regions are of interest as variations in PET are directly related to altered values of actual evapotranspiration and thus water resources simulation. However, to assess the plausibility of simulated PET values (due to radiation input data as well as the PET approach), efforts to validate those estimations, e.g., to evaporation pan observations, are needed.

RQ 3 What is the likely uncertainty of net radiation and its components?

It remains a challenge to answer this question due to a lack of ubiquitous ground-based and highly qualitative radiation measurements. Nevertheless, for the two station-based observation databases that were considered in paper 1 and based on the mean absolute bias efficiency metric, some initial conclusions can be drawn. Independent of calculations

of WaterGAP or ERA-Interim reanalysis, the (median of the) mean absolute bias for net radiation is, at around 20 Wm^{-2}, the highest uncertainty among the radiation components. The mean absolute bias of the other radiation components shows a dependency on the calculation method (climate input data and WaterGAP vs. ERA-Interim reanalysis) and the reference database. Shortwave downward radiation uncertainty is, at 12 to 18 Wm^{-2}, higher than longwave downward radiation uncertainty (8 to 10 Wm^{-2}). This is somewhat surprising, as global averages of longwave downward radiation are (in absolute numbers) a factor of 2 higher than shortwave downward radiation, indicating that there is a more significant relative bias of shortwave downward radiation. Both radiation components are best covered with observations (number of stations/number of months), thus allowing relatively strong conclusions to be drawn. The assessment of the remaining radiation components is conducted with limited confidence due to the small number of observation stations, and mean absolute bias varies greatly between the calculation method and reference data set between 7 and 20 Wm^{-2}.

8.1.2 Evaluation of the sensitivity of simulated water balance components to climate input data

One major focus of this dissertation is the sensitivity of simulated water balance components to climate input data and the role of calibration in that regard. In the following, RQ 4 and RQ 5 are answered.

RQ 4 What is the sensitivity of simulated water balance components, including human water use, to state-of-the-art climate input?

Climate input data have a huge impact on model outputs, as indicated both in global numbers and spatial patterns in the papers 2–4. In paper 2, and taking WaterGAP 2.2 as well as two climate input data as an example, the difference in global-scale precipitation P (~ 2000 $km^3\ yr^{-1}$) is translated into a strongly differing global estimate of river discharge into oceans and inland sinks Q in the same quantity (paper 2, Table 2). Even though both precipitation variables of that study are based on the same data source family, differences in global-scale water balance components exist due to the approach of handling precipitation undercatch. Despite global sums of actual evapotranspiration (AET) being equal among the two climate input data, the spatial pattern differs greatly (especially in the tropics). For AET, climate input data uncertainty has the second-largest impact after calibration (or no calibration) in that assessment (paper 2, Figure 3).

Within papers 3 and 4, the number of investigated climate input data was increased to five state-of-the-art data sets, including one homogenized climate input data set. As shown in paper 3 (Table 4), global sums for the original P data sets are (despite station density differing significantly) very similar both for the continents and the global scale. Different

handling of precipitation undercatch (or neglecting it) leads here to significant variation in P (~ 8000 $km^3\ yr^{-1}$, Table 4). In contrast to paper 2, AET differs by 5500 $km^3\ yr^{-1}$ at the global scale, whereas Q ranges about 3000 $km^3\ yr^{-1}$. Mean global temperatures are very close among the climate input data of this study (paper 3, Fig. 2), and there are only small differences in spatial distribution (paper 4, Fig. A2), meaning that this variable can be seen as consistent regardless of the climate input. With Fig. 2 of paper 3 and Figs. A3 and A4 of paper 4 taken into account, reasonable differences exist for downward radiation fluxes in the climate input data, resulting in inconsistencies when simply concatenating two reanalysis-based climate input data. Net radiation differs globally by up to 15 Wm^{-2} with corresponding consequences for PET (20 000 $km^3\ yr^{-1}$), but interestingly there is not much impact on AET (1000 $km^3\ yr^{-1}$), compared to the homogenized climate input data where net radiation and PET are similar to the other climate input data. Nevertheless, the homogenization of climate input data avoids offsets in radiation, water fluxes and storages to a large degree (as described in paper 2).

Human water use is, in global sum, not very sensitive to climate input data uncertainty. Despite only one water use sector (irrigation) being modeled using climate input data (among other data sources), irrigation has the largest proportion of overall water withdrawals and water consumption (paper 4, Fig. 3). Hence, uncertainty in simulated irrigation water use amounts to approximately 50% (or around 50 $km^3\ yr^{-1}$) of the global sum of water consumption of the other water use sectors (paper 4). Anyhow, global-scale actual water consumption from five climate input data varies only by about 5% (paper 3, Table 2) and human water use accounts for only roughly 1% of precipitation (paper 4, Table 2).

In summary, the largest proportion of climate input data uncertainty in simulated global freshwater fluxes is related to precipitation, and especially due to the different handling of precipitation undercatch effects. For future creation of climate input data sets, this kind of uncertainty needs to be addressed carefully (e.g., as in Beck et al., 2017) to enable a best estimate of water resources.

RQ 5 How does the WaterGAP calibration approach affect the uncertainty of climate input data?

The calibration using long-term average observed river discharge in ~54% of global land area (excluding Antarctica and Greenland) greatly reduces the uncertainty of climate input data in these regions. In calibration areas, uncertainty of climate input data is reduced to 1.6% for the sums of simulated river discharge, and for the homogeneous climate input data sets. In contrast, for non-calibrated regions, this variation is 18.5% (paper 3, Table 5). For the two alternative climate input data in paper 2 (WFDEI, CRU), 82% of the related uncertainty is generated in non-calibration areas. Variation of river discharge on the grid-cell level is, at 37% for calibration and 74% for non-calibration regions, still high. As

simulated river discharge is forced to observed values in calibration regions, the uncertainty of climate input data is transferred to actual evapotranspiration there, which differs greatly in those regions and less for non-calibrated regions (paper 3, Table 5). Due to limited water availability, the human water use demand cannot be fully satisfied, which reduces the effect of climate forcing uncertainty on human water use estimations (paper 4).

8.2 Model modification

The focus of the second part of this dissertation is to analyze the effect of model modifications (in terms of including or refining process description, input data sets or algorithms) on corresponding model results as well as model parameters. The following five research questions deal with consequences of model development in general (RQ 6), the effect of newly introduced land cover information (RQ 7), improved process representation (integrating reservoir commissioning years, RQ 8), updated soil water capacity information (RQ 9) and a modified calibration scheme (RQ 10).

RQ 6 What is the effect of WaterGAP development on water flows and storages between version 2.1d and version 2.2?

Between the publication of WaterGAP model version 2.1d and 2.2, more than a decade of model development took place, resulting in various model versions integrating dozens of new features/data sets (Table 1). The improvement of the simulation of water fluxes and storages is (apart from specific project-related issues) the driving force of WaterGAP development. In sum, this resulted in a more complex hydrological model structure on the one hand and (hopefully) a better assessment of water flows and storages on the other. In order to quantify the cumulative effect of model development, one part of paper 2 deals with the comparison of WaterGAP 2.2 with a version that is similar to version 2.1d. Specifically, water balance components were compared in terms of global sums as well as spatial distribution. The long-term average global sum of river discharge (actual evapotranspiration) is calculated to be about 450 $km^3\ yr^{-1}$ lower (higher) for the modeling structure of version 2.1d. Spatial differences in actual evapotranspiration (and also of renewable water resources) are visible especially in areas with surface water bodies and in snowy regions (Fig. 3d and 4d of paper 2) but are in sum moderate as values increase (decrease) by more than 10 mm yr^{-1} only at 12% (14%) of the land surface. The general increase of simulated river discharge between versions 2.1d and 2.2 (Table 5 of paper 2) is in line with that result, even though different climate input data were used in the assessments of Table 5. Satisfied water consumption is only around 50 $km^3\ yr^{-1}$ lower in the old model structure, meaning that model structure does not have a strong effect on this water balance component (despite the many modifications in that component). For both model variants, a large share (around 90%) of water use demand could be satisfied.

Regarding model performance, the widely used Nash–Sutcliffe efficiency was used to investigate river basins with aggregations of Köppen–Geiger climate zones (paper 2, Table 4). In sum, there are significantly more basins with higher model performance in model version 2.2, which holds true also for the climate zones, especially for temperate C and continental D climates, indicating that the model development led to more realistic simulation of monthly river discharges in those climate zones. However, there are still challenges left, especially for the dry B climates. Change of total water storage during a 30-year period differs significantly and is mainly related to the lack of a decrease of groundwater storage in the model structure of 2.1d (due to the absence of groundwater abstractions). Spatial patterns of total water storage amplitudes differ mainly in snow-dominated regions (due to snow algorithm), along large rivers (due to variable river velocity approach) and in areas where water is withdrawn from the groundwater (not considered in 2.1d). In terms of water balance closure, the model structure of 2.1d (as emulated in version 2.2 of paper 2) has an error which is a magnitude higher than those of version 2.2.

In general, model development from version 2.1d to 2.2 leads to improved results in terms of monthly river discharge, and it is expected that this also leads to a more realistic assessment of renewable water resources as well as total water storage changes. However, simulated river discharge on the global scale tends to be higher with model structures of version 2.2 and afterwards.

RQ 7 What is the impact of an alternative land cover map on simulated water flows and storages?

An updated land cover map, based on satellite information from MODIS from the year 2004, was introduced in WaterGAP 2.2. As part of paper 2, the effect of this alternative land cover input was investigated by comparison to a model variant with the previous (map-based) input. The effects on global sums of actual evapotranspiration as well as river discharge is, at around 200 $km^3\ yr^{-1}$, lower than those of the model structure (see RQ 6), and no other global-scale values differ (paper 2, Table 2), even though land cover class differs in a large number of grid cells (paper 2c, Figure 2). The effect of varying land cover attributes on radiation calculation (and consequently actual evapotranspiration as well as renewable water resources) is visible especially in energy-limited regions. Additional impact is visible in areas where total soil water capacity changed due to different rooting depths (paper 2, Fig. 3c, Fig. 4c). Large differences in river discharge seasonality exist only in areas where land cover changed in a significant number of grid cells, e.g., for the Rhine basin (paper 2, Fig. 5). Also, changes in Nash–Sutcliffe efficiency metrics are only marginal (paper 2, Table 4), indicating that a modified land cover input map has no large effect on simulated river discharge, even though differences exist for a substantial number of grid cells in terms of land-cover-based attributes. The low sensitivity to changes in land cover input limits the ability of WaterGAP to simulate hydrological effects due to land cover

change, but this is also hard to assess for other modeling approaches (e.g., Li et al., 2009). On the other hand, variations of total water storage are impacted by land cover attributes strongly in those regions with significant changes (e.g., for the Yangtze, see paper 2, Fig. 7c). Hence, the question of whether the modified land cover input leads to improvements in total water storage variations should be studied basin by basin via comparing to satellite gravimetry (GRACE) data.

RQ 8 How does the integration of reservoir commissioning years affect the model results?

Within model version 2.2 (ISIMIP2a), the year in which each reservoir was commissioned was implemented, i.e., the year when the dam started its regulation activities. This required thorough preparation of input data as well as code modifications. The effect of this model modification on simulated water fluxes is described in detail in Appendix A1. In a model experiment study, one model variant with reservoir commissioning years was considered as well as one that assumes that all reservoirs have been in operation since the beginning of the 20th century (the handling of all model versions except 2.2 (ISIMIP2a)). The global sum of maximum (and actual) reservoir storage volume reflects the historical dynamics when commissioning years are considered (Fig. A1). Model performances of simulated to observed monthly river discharge were compared for the years before and after commissioning and especially for those 214 river basins that have a considerably high relative reservoir storage volume. Remarkable differences exist for the time period before the upstream reservoir(s) began operation, and the model variant that considers commissioning years performs significantly better for all investigated model efficiency metrics (Fig. A2). Only small differences in model efficiency exist for the period after reservoir(s) began operation, and model performance is (in most cases) lower than for the pre-reservoir-commissioning time period. This indicates that the generic reservoir algorithm is not able to capture the management of the reservoirs in a realistic way. However, that is not expected as reservoir management is highly dependent on economic, political and other decisions which can hardly be captured with a generic approach, e.g., that of Hanasaki et al. (2006). In summary, the implementation of reservoir commissioning years improved the simulation of monthly river discharge especially in those river basins with a considerable amount of reservoir storage volume and for the time period before the reservoir began operation.

RQ 9 What is the effect of an updated spatial input map of total available soil water capacity?

The availability and quality of spatially distributed data sets on soil properties have improved in the past. For example, the number of soil profiles in the World Inventory of Soil Emission Potentials (WISE) database has doubled in the past 15 years, and they contain depth-varying attributes such as available water capacity for each 20 cm (Batjes, 2012).

Updating the spatial input data of WaterGAP therefore ensures that the latest available information from the scientific communities is used to generate such data sets. WaterGAP 2.2b was used to test the effect of an updated total available soil water content (TAWC) data set. Two non-calibrated model variants were run, one with the input map of Batjes (1996) and the other one with data from Batjes (2012). This experiment is described in Appendix A2 and allows this research question to be answered.

Global average TAWC for the top meter increase from 105 mm m^{-1} to 112 mm m^{-1} when the updated data set is taken into account. In most regions of the global land surface, TAWC increases, but for northeastern Eurasia values decrease (Fig. A4). Similar patterns are visible when multiplying TAWC by rooting depth to obtain the maximum soil water storage. Due to the overall increased maximum soil water storage, the mean amplitude of soil water storage increased from 59 to 62 mm when the updated TAWC map is used. This is of particular interest, as this storage compartment shows, in comparison to other models, relatively small amplitudes (e.g., Khandu et al., 2016). Significant regional differences in runoff exist (more runoff in northeastern Eurasia, less in many other regions, Fig. A6) which could have a positive influence on calibration parameters. For the large areas where the calibration parameter γ is at the maximum value (indicating too much simulated runoff) the greater amount of water that is kept in the soil water storage (and evaporates) could help to decrease runoff and therefore the γ value (see also Appendix A3). For northeastern Eurasia, where γ is as its minimum value (Fig. 8 in Appendix A3), a decrease of maximum soil water storage leads to an increased calibration parameter. The regional differences in runoff and actual evapotranspiration are balanced out at the global scale. Runoff (evapotranspiration) decreased (increased) only slightly by 3 mm yr^{-1}. In summary, the update of TAWC lead to slight increases of soil water storage variation. Regional differences in runoff and actual evapotranspiration with in slight general decrease of global-scale runoff are also resulting due to the updated TAWC input.

RQ 10 How does river discharge measurement uncertainty influence calibration parameters and simulated freshwater fluxes?

The common standard in (catchment) hydrological modeling is to use a calibration routine that adapts model parameters with the aim of reproducing, e.g., observed river discharge with a reasonable agreement. In GHMs, such approaches are only sparsely used. WaterGAP is the only model that considers a basin-specific calibration routine to match the simulated to the observed long-term average river discharge (see also Chapter 1). This was mainly done to allow for water resources assessment with high accuracy. However, as reported in the literature (e.g., Coxon et al., 2015), measurement uncertainty of discharge observations can be significant, which is currently not considered in the WaterGAP calibration routine. In addition, the usage of additional calibration parameters can lead to physically implausible values on the grid-cell level (e.g., see the discussion in paper 2), which needs

to be avoided. Therefore, an optional 10% uncertainty of long-term average measured river discharge was taken into account in the WaterGAP calibration scheme. Appendix A3 describes this approach and the resulting effects on calibration parameters as well as simulated freshwater fluxes.

The assumed 10% uncertainty of observed river discharge is implemented in a four-step calibration approach (for details see Appendix A3). The assessment of model parameters showed that the usage of additional calibration parameters could be reduced when the observation uncertainty is considered (Table A7), and in most cases the parameter values are less extreme (Figs A7, A8). However, for semi-arid and arid regions WaterGAP still over-estimates river discharge by at least a factor of 2. The choice of climate input data greatly influences the calibration scheme, and the homogenized WFD/WFDEI climate input data lead to the least usage of extreme calibration parameters (Table A7). Discharge seasonality is influenced in those cases where the observed river discharge from the specific (or upstream) basin is varied by 10%, and the modified calibration approach can have significant impacts on discharge amplitudes (e.g., Rhine in Fig. A8). Model efficiencies are affected by the calibration scheme only slightly. Global sums of river discharge and evapotranspiration are affected by roughly 250 $km^3\ yr^{-1}$ with an overall increase of river discharge when discharge uncertainty is taken into account. This effect was traced back as a complex interplay of upstream and downstream basins especially in the calibration areas, but minor effects also occur in the non-calibrated areas where the calibration parameter γ is regionalized.

In summary, the modified calibration scheme of WaterGAP leads to less extreme calibration parameters and only slight changes in model efficiencies as well as global-scale water balance components. Significant differences in discharge seasonality are visible in the basins where the uncertainty of observed river discharge became effective in the calibration routine.

8.3 Model application

The third and last part of the thesis consists of three applications of WaterGAP. In particular, a best estimate of water balance components at the global and continental scale is provided (RQ 11), an integrated study of how to assess groundwater depletion is presented (RQ 12) and the question is answered of how river flow regimes will likely be altered as a consequence of climate change (RQ 13).

RQ 11 What is the current best estimate of global and continental freshwater fluxes and storages with WaterGAP?

There is a long tradition of assessing global-scale freshwater fluxes (e.g., Baumgartner and Reichel, 1975; Korzun, 1978, and references therein). By using GHMs that are constrained by a closed water balance, global (and grid-cell-specific) water balance components can be of great value in assessing e.g., the river discharge, as well as variations in total (and single) water storage compartments. With all the model improvements as well as the updated calibration structure taken into account, the most recent model version 2.2b can be seen as the WaterGAP model version to determine the best estimate of global (continental)-scale water balance components. The water balance components as calculated with WaterGAP 2.2b and driven by homogenized WFD/WFDEI climate input for the continents and the global land surface is shown in Table 3 for the standard climate period 1971–2000 and in Table 4 for the IPCC AR5 reference period 1986–2005. As shown in paper 3, the choice of time period for such assessments is important due to the inter-annual (and inter-decadal) variations of meteorological variables and corresponding hydrological fluxes. This is also visible when Tables 3 and 4 are compared. Differences in simulated global river discharge Q between the two time periods are, at around 800 km^3 yr^{-1}, close to the estimation of consumptive water use. The comparison of Q from Table 3 with the previous WaterGAP versions (and other references in papers 2 and 3) shows that values are well in the range of other estimates. Specific reasons for the higher simulated values of WaterGAP 2.2 (ISIMIP2a) include the land-ocean mask and are discussed in paper 3.

Table 3: Continental and global scale water balance components (except for those of Antarctica and Greenland) from WaterGAP for homogenized WFD/WFDEI climate input data and the years 1971–2000. P: precipitation; AET: actual evapotranspiration; Q: discharge into oceans and inland sinks (computed as sum of net cell runoff); WCa: water consumption (actual, sum of NAs and NAg); NAs: net abstraction from surface water (actual); NAg: net abstraction from groundwater; dS/dt: change of total water storage. All values are given in km^3 yr^{-1}.

Component	Africa	Asia	Europe[f]	N America	Oceania	S America	Global
P	21315	24834	13254	16894	5978	29328	111602
AET[a]	17131	14126	7304	10275	4002	17785	70623
Q[b]	4123	10122	5873	6525	1954	11491	40088
WCa	74	632	94	134	17	31	983
NAs[c]	111	790	60	75	24	47	1106
NAg[d]	-37	-157	35	59	-7	-16	-123
dS/dt[e]	-13	-46	-17	-41	5	21	-92

[a] AET does not include evapotranspiration caused by human water use, i.e., WCa. [b] Taking into account anthropogenic water use. [c] Satisfied demand from surface waters. [d] Negative values indicate that return flows from irrigation with surface water exceed groundwater abstractions. [e] Total water storage (TWS) of 31 December 2000 minus TWS of 31 December 1970, divided by 30 years. [f] Includes the entire Russian Federation.

Table 4: Continental and global scale water balance components (except for those of Antarctica and Greenland) from WaterGAP for homogenized WFD/WFDEI climate input data and the years 1986–2005. P: precipitation; AET: actual evapotranspiration; Q: discharge into oceans and inland sinks (computed as sum of net cell runoff); WCa: water consumption (actual, sum of NAs and NAg); NAs: net abstraction from surface water (actual); NAg: net abstraction from groundwater; dS/dt: change of total water storage. All values are given in $km^3\ yr^{-1}$.

Component	Africa	Asia	Europe[f]	N America	Oceania	S America	Global
P	21205	24776	13375	16863	5798	28917	110934
AET[a]	17266	14154	7384	10311	3909	17696	70720
Q[b]	3873	9960	5913	6432	1867	11182	39227
WCa	84	724	97	144	20	36	1105
NAs[c]	122	890	68	85	29	56	1251
NAg[d]	-39	-166	29	59	-9	-21	-146
dS/dt[e]	-17	-62	-20	-25	2	3	-119

[a] AET does not include evapotranspiration caused by human water use, i.e., WCa. [b] Taking into account anthropogenic water use. [c] Satisfied demand from surface waters. [d] Negative values indicate that return flows from irrigation with surface water exceed groundwater abstractions. [e] Total water storage (TWS) of 31 December 1985 minus TWS of 31 December 2005, divided by 20 years. [f] Includes the entire Russian Federation.

RQ 12 How can groundwater depletion at the global scale being assessed using multiple sources of observations jointly with hydrological modeling?

If "(…) groundwater abstractions cannot be balanced by increased recharge and decreased discharge over a number of years, a long-term decline of hydraulic heads and groundwater storage, i.e., groundwater depletion (GWD), will result" (Döll et al., 2014a: 5698). Hot spot GWD areas are e.g., the High Plains Aquifer in the US and the Ganges Delta of Bangladesh and northwestern India. In the literature, many approaches exist to characterize the anthropogenic pressure on groundwater aquifers (e.g., Gleeson et al., 2012; Wada et al., 2010). The novelty of the approach in paper 5 is that it combines data from global hydrological modeling, groundwater well observations, independent estimates and GRACE satellite total water storage anomalies to determine GWD areas (and intensities) for the global scale in a consistent way. WaterGAP 2.2a was applied, including a newly integrated algorithm to compute groundwater recharge from surface water bodies in semi-arid/arid regions. That reflects the hydrological process that occurs frequently in dry regions when surface water bodies "lose" water to the groundwater and do not receive it from the groundwater (as in the case of humid areas). As GWD is related to the amount of water that is artificial pumped e.g., for extensive crop irrigation, the quantification of water abstractions from groundwater plays a key role. Some assumptions while calculating water withdrawals have to be made (e.g., constant fractions of areas equipped with surface water or groundwater irrigation) due to the lack of data. By comparison to independent estimates (including observations) of GWD and groundwater recharge, it was found that a model

version in which a deficit irrigation of 70% from the optimal irrigation is assumed resembles observed depletion. This is also supported by the likely local reduction of irrigation intensities when water supply is limited. Global-scale GWD was quantified as having doubled (from 56 to 113 $km^3\ yr^{-1}$) between the time periods 1960–2000 to 2000–2009 and is smaller than previous estimates. The countries with the highest GWD rate in 2000–2009 are India, the United States, Iran, Saudi Arabia and China. Furthermore it was found that trends of GRACE total water storage anomalies can, together with a global water availability and water use model (like WaterGAP), help to monitor temporal dynamics of water abstractions, but the low spatial resolution of GRACE remains a challenge. The spatially explicit quantification of human water abstractions is still subject to assumptions due to partly missing statistical data (for both time and space) and leads to uncertainties in model-based assessments of GWD. For example, groundwater recharge in the Hai River basin (China) are underestimated with WaterGAP, which leads to a significantly higher GWD estimate than GRACE and regional assessments.

RQ 13 What is the impact of climate change on river flow regimes?

It is expected that climate change will have substantial impacts on freshwater ecosystems and thus it is of interest to quantify those effects. In paper 6, a new indicator for detecting changes of river flow regimes was developed. It extends the widely used quantification of mean (runoff or river discharge) change by seasonality and statistical low-flow/high-flow components. This provides better insight into, e.g., the seasonal pattern of flow regime change in relation to mean changes. WaterGAP 2.1g was forced with two bias-corrected GCMs and two greenhouse gas emission scenarios (namely A2 and B2; see paper 6 for details). The statistical low-flow indicator Q_{90} was assessed to follow broadly the change of mean river discharge Q_{mean} despite significant differences being visible; e.g., the per-cent decreases of Q_{90} are often higher than those of Q_{mean} (e.g., in Australia and South America), and there are some areas where Q_{90} decreases while Q_{mean} increases (and vice versa). In other areas (e.g., the eastern USA, Scandinavia and the European part of Russia), Q_{90} increase is greater than Q_{mean} increase. In general, this means that relative discharge variability increases with the climate change projections. Also, seasonal river discharge patterns will likely differ with climate change in some areas.

In addition, a new indicator was developed that aims to quantify river flow regime shifts. Based on map information of perennial and intermittent river segments, generic thresholds for the specific river stream characteristics were estimated and used to quantify shifts in river flow regimes due to climate change. When the two GCMs and emission scenarios are taken into account, significant flow regime shifts in the 2050s are calculated to take place in 6.3–7.0% (A2 scenario) and 5.4–6.7% (B2 scenario) of global land area. Substantial changes occur mainly in semi-arid areas and cold regions. Flow regime shifts from perennial

to intermittent (intermittent to perennial) correlate with decreased (increased) mean runoff and especially with decreased (increased) Q_{90}.

The development and application of those new indicators have allowed for new insights into spatial variability of river flow regime as impacted by climate change. However, the assessment is limited to one GHM and only two GCMs (with two emission scenarios each). It is expected that more robust conclusions can be drawn when multiple GHMs are driven with multiple GCMs under different emission scenarios. Such a multi-model, multi-scenario assessment could be done within the framework of the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP), e.g., in the upcoming Phase ISIMIP2b (Frieler et al., 2016).

9 Conclusions and outlook

The assessment of water balance components and related indicators by means of global hydrological models (GHMs) has evolved significantly over the previous 25 years. Due to the closed water balance and inclusion of hydrological processes as well as human influence on the terrestrial water cycle, GHMs are valuable tools for water resources assessment. Furthermore, GHMs are frequently used to quantify water-related impacts of projected climate change. One GHM, namely the freshwater availability and use model Water – Global Assessment and Prognosis (WaterGAP,) is the subject of detailed analyses in this thesis. The dissertation contains six peer-reviewed papers along with a total 13 research questions that are structured in three parts in accordance with the title "Evaluation, modification and application of a global hydrological model". The majority of the research questions are related to the various sources of uncertainty in global water assessments and to challenges that still remain for the current generation of GHMs.

Different ways to calculate net radiation within WaterGAP (or taking from a reanalysis) as the basis for potential (and actual) evapotranspiration are the focus of the first part. Based on efficiency metrics with station-based observations, it is shown that the upward shortwave radiation calculation in WaterGAP is as plausible as from a state-of-the-art reanalysis (ERA-Interim) and performs well overall compared to observations, even though the energy balance is not closed in WaterGAP. The longwave upward radiation component of ERA-Interim reanalysis fits better to observations than those calculated by WaterGAP, which could be improved if the emissivity value were increased and coupled to snow dynamics. The mean absolute uncertainty of net radiation is, independent of reference data set and modeling approach, around 20 Wm^{-2} and the highest among the radiation components. Thus, and given the impact of net radiation uncertainty on water resources assessment (Döll et al., 2016), quantification of net radiation with high accuracy remains a subject for research.

The uncertain climate input data are the focus of the second part of the model evaluation. Assuming that the considered state-of-the-art climate input data sets are equally likely to be correct, the sensitivity of simulated freshwater fluxes and water storages to alternative climate input data was assessed in model experiments. In general, the range of simulated discharge increases with the number of alternative data sets considered. It was shown that the major uncertainty remains due to different handling of precipitation undercatch in the climate input data sets. Two of the common global precipitation data sets (CRU, GPCC) are similar in their continental and global values even though station densities vary significantly. Appropriate handling of precipitation undercatch is thus a very important challenge and is currently subject to improvements (Beck et al., 2017). Inconsistent combination of two of the most frequently used climate input data sets (WFD, WFDEI) leads to significant impacts on simulated water storage and fluxes. It was shown that a

homogenization approach could reduce this inconsistency greatly. The sensitivity of global-scale sums of simulated irrigation water use to different climate input data sets is about 50% of the sum of all the other water use sectors. Given the relation of absolute values, this can be seen as relatively low sensitivity, but differences might be very significant at grid-cell or regional level. Water availability limitations that prevent fulfillment of the full water demand are reducing this sensitivity even further. The calibration approach of WaterGAP reduces the uncertainty due to climate input in calibrated areas to around 1.6%, compared to 18.5% in non-calibrated areas, but variations among the climate input data remain at the grid-cell level. The lack of river discharge observations (or the sharing of such data) in some parts of the Earth limits more constrained assessments of global water resources.

The second part of the thesis deals with the effect of model modification on model output as well as model parameters. The basic objective was to quantify the impact of implementing new data sets, new/refined processes and an updated calibration structure in a total of five research questions.

The modifications of model structure from an initial WaterGAP version to version 2.2 led to significant improvements in the simulation of monthly river discharge, especially for temperate and continental Köppen–Geiger climate zones. It is expected that other components of the water balance are also improved with the updated model structure. However, challenges for model improvement remain, especially in the dry climates. WaterGAP tends to overestimate river discharge in many of those basins by at least a factor 2. Therefore, future model development should focus on those areas. The inclusion of capillary rise from groundwater could increase the amount of evaporation. In the current model structure, exfiltration and subsequent evaporation from temporary ponds in semi-arid/arid regions are not implemented, and the first efforts have been made to investigate the potential of this additional contribution (Nagel, 2012). As WaterGAP contains only one soil layer (with varying thickness according to the rooting depth), the actual soil water content is typically low in those regions. Even though precipitation increases the soil water storage, the relative soil water storage is still very low, which limits the ability to evaporate water from the soil. Hence, the inclusion of a second soil layer could increase the evaporation of the top layer and thus help to reduce the simulated runoff in those regions. However, model complexity would strongly increase as water transport between the soil layers has to be parameterized.

The integration of the calendar year in which reservoirs were commissioned into WaterGAP improves the simulated river discharge significantly, especially for the time period before the reservoir began operation. The analyses showed that the generic reservoir algorithm is subject to improvements as model performance is substantially lower after the reservoirs are commissioned and the timing of peak flows (e.g., in Nile River basin downstream of the

Aswan Dam) does not match. However, reservoir management rules are typically not publically available, which limits the capability of improving generic or individual reservoir algorithms.

Two spatial data sets were updated in order to allow the integration of state-of-the-art input data. The effect was assessed by setting up model experiments with two versions that differ only in that specific data set. A satellite (MODIS)-based land cover map replaced map-based input data, resulting in a change in land cover type for a large part of the land surface. However, the effect of updating the input data was moderate and much lower than, e.g., the effect of calibration (or a lack thereof) or forcing the model with alternative climate input data. In particular, global sums of river discharge to the oceans and inland sinks as well as actual evapotranspiration vary by only ~200 $km^3\ yr^{-1}$. At least for some river basins with wide-spread changes in land cover, differences in river discharge seasonality were also found (e.g., the Yangtze River basin). Updating the land cover input data led to improvements in simulated river discharge in some regions and to a lower performance in other regions.

The spatial input data on total available water capacity were updated to a database with a much higher number of in situ data. Global mean values of total available water capacity increased by 7%. In most regions of the land surface (except for northeastern Eurasia) the values increased, and this pattern holds true for the maximum soil water storage, which is calculated by multiplying the updated input data (for the first meter) by land-cover dependent rooting depth. The mean amplitude of soil water storage increased slightly, and significant spatial differences in runoff exist; however the effect on global-scale averages is very small. The model output improves in the sense that runoff is slightly decreased (resulting in lower calibration demand) and variations of total water storage are increased.

The calibration routine of WaterGAP was modified to consider a potential 10% uncertainty of long-term average river discharge observations in a four-step scheme. As a result, the number of basins in which additional calibration parameters are needed was reduced by 14%. With this approach, in 43% to 50% of the river basins (varying among climate input data) no additional correction factors were needed. Furthermore, for most of the basins where those additional parameters are still required, parameter values are less extreme. The model performance compared to non-modified observed river discharge reduced only slightly, and global-scale river discharge and evapotranspiration differ by roughly 250 $km^3\ yr^{-1}$: the same order of magnitude as the modified land cover input data. However, in around the half of the river basins, additional calibration parameters are still required. Future research should focus therefore on developing alternative calibration approaches, and first experiences on a multi-criterial calibration and data assimilation scheme with WaterGAP are in progress.

The third and final part of the thesis contains three applications of the WaterGAP model. The most current WaterGAP version 2.2b, forced with homogenized climate input data, was used to assess global-scale and continental-scale water balance components for two time periods. For the standard climate period 1971–2000, river discharge to the oceans and inland sinks is computed to be 40 000 $km^3\ yr^{-1}$ and actual evapotranspiration as 70 500 $km^3\ yr^{-1}$. Water consumption is calculated to be roughly 1% (1000 $km^3\ yr^{-1}$) of the global-scale precipitation, and > 60% is consumed in Asia. For the IPCC AR5 reference period 1986–2005, the numbers differ only slightly. The 700 $km^3\ yr^{-1}$ less precipitation in that period translates into a 850 $km^3\ yr^{-1}$ lower river discharge and 100 $km^3\ yr^{-1}$ higher actual evapotranspiration. In addition, annual averages of water consumption increase by 120 $km^3\ yr^{-1}$ in the later time period. As WaterGAP is calibrated to river discharge observations for 54% of the global land surface (excluding Antarctica and Greenland), it is expected that this assessment is reasonably well constrained. By enhancing highly qualitative river discharge observations and sharing available data, the assessment of water resources and water balance components could be improved in the future.

Assessing groundwater depletion at the global scale is the focus of the second WaterGAP application. For semi-arid and arid regions, the process of groundwater recharge below surface water bodies was integrated, which helps to improve the agreement of simulated groundwater depletion as compared to independent estimates. Furthermore, it was found, that in areas where groundwater depletion already occurs, farmers likely decrease the irrigation to 70% of optimal irrigation demand. For many study regions where independent observations or estimates were available, the model configuration with both groundwater recharge below surface water bodies and deficit irrigation fits best to observations or other (model- or data-based) estimates. The uncertain data and assumptions for human water use limit such model-based assessments (see also Döll et al., 2016). Global-scale groundwater depletion was estimated to be lower than reported in previous studies and has doubled from 1960–2000 to 2000–2009. Data from the GRACE satellite were tested regarding their ability to monitor temporal dynamics (and trends) of water abstractions, but the low spatial resolution of GRACE, as well as the fact that GRACE observes changes in the total water column and does not directly observe groundwater or groundwater depletion, remains challenging. This study showed that a combination of hydrological modeling and observation data could increase the plausibility of groundwater depletion simulation.

In the third example, WaterGAP was applied for an impact study of climate change on river flow regimes. Several new indicators were developed which allow for a more in-depth view on changes in discharge variability and seasonality for ecologically relevant river flow regimes. With two GCMs with two greenhouse gas emission scenarios taken as input data, WaterGAP simulates substantial changes in river discharge variation. In addition, an indicator for defining perennial and intermittent rivers was developed and applied

regarding shifts under climate change. Significant shifts in river flow regime were calculated for roughly 6% of the global land surface in the 2050s. For future research, a multi-model and multi-scenario study should be conducted in order to strengthen the robustness of such ecologically highly relevant flow regime shifts.

Specific research needs according to the research questions have already been touched upon in the sections above. Based on the findings of the thesis, future research ideas and applications with the WaterGAP global hydrological model are

- assessment of the role of potential evapotranspiration equations for simulating potential water flux to the atmosphere, and validating with reference data, e.g. pan evaporation observations;
- integration of a gradient-based groundwater model for an improved representation of surface water-groundwater interaction, a more physically based groundwater depletion assessment, the opportunity to include capillary rise and thus increase evaporation, and the quantification of reservoir impoundments, e.g., according to Chao et al. (2008);
- improvement of simulations in semi-arid and arid regions by, e.g., including the exfiltration process and subsequent evaporation in temporary ponds, as well as capillary rise from the groundwater and a second shallow soil layer on the top to increase the relative soil storage saturation after precipitation events, which would increase evaporation and reduce river discharge;
- development of a standardized evaluation framework (e.g., to assess model modifications) including multiple observations for model validation, taking into account data-driven products (e.g., Fekete et al., 2002; Gudmundsson and Seneviratne, 2016);
- advancement and application of the calibration and data assimilation approach of Eicker et al. (2014) and Schumacher et al. (2016b) to assimilate multiple sources of observation (GRACE, river discharge, altimetry) for the global scale.

In general, global hydrological models provide valuable assessments of global-scale water fluxes and storages. They are used as integrated tools for assessing water resources in transboundary river basins where data sharing among countries is often limited due to political issues and could be of benefit for international initiatives such as the World's Large Rivers Initiative (http://unesco-chair.boku.ac.at/index.php/wlri.html). Some ideas for a more robust assessment of the terrestrial hydrological cycle could comprise

- the development of consistent, up-to-date climate input data including a plausible way to deal with the problem of precipitation undercatch;
- conducting consistent model intercomparison studies under historical and climate change conditions with the aim of investigating the reasons for varying model

output as found in previous studies (Haddeland et al., 2011; Schewe et al., 2014) in a structured way;

- increasing the plausibility of historical and projected water use assessments, e.g., under the framework of the Water Futures and Solutions initiative (WFaS; Wada et al., 2016a);
- the development of a multi-model seasonal forecast system, e.g., for drought warnings in a community-driven approach;
- understanding the HyperHydro initiative (http://www.hyperhydro.org/) as a call for joint and community-driven development of a global hydrological model with high resolution to support the initial ideas of Wood et al. (2011) and Bierkens et al. (2015), who defined a potential application in water management for data-scarce regions.

As models are only simplified representations of the reality and some of the uncertainties that are described within this thesis will persist in the near future, there is still room for improving global hydrological models, as well as indicators and applications that are based on the model output. By the evaluation, modification and application of the global hydrological model WaterGAP, valuable insights are given in this highly interesting topic. These will hopefully build a strong basis for further research in the scientific community.

10 Schlussfolgerung und Ausblick

In den letzten 25 Jahren gewann die Quantifizierung von globalskaligen Wasserhaushaltskomponenten sowie wasserrelevanten Indikatoren mittels globaler hydrologischer Modelle (GHMs) wesentlich an Bedeutung. Die geschlossene Wasserbilanz, der explizite Fokus auf die Abbildung terrestrischer hydrologischer Prozesse sowie (in den meisten Fällen) die Berücksichtigung des menschlichen Einflusses auf den Wasserkreislauf, etwa in Form von Wasserentnahmen für den Bewässerungsbedarf, stellen wesentliche Eigenschaften von GHMs dar. In dieser Dissertation wird das Wasserverfügbarkeits- und Nutzungsmodell Water – Global Assessment and Prognosis (WaterGAP) für detaillierte Analysen verwendet. Die Arbeit beinhaltet sechs begutachtete Artikel in Fachzeitschriften und gibt Antworten auf insgesamt 13 Forschungsfragen, die hinsichtlich des Titels in „Auswertung, Weiterentwicklung und Anwendung eines globalen hydrologischen Modells" dreigeteilt dargestellt sind. Hauptsächlicher Fokus der Forschungsfragen ist der Umgang mit verschiedenen Aspekten der Unsicherheit der globalskaligen Berechnung von Wasserhaushaltskomponenten und der Benennung von damit zusammenhängenden Herausforderungen der derzeitigen Generation von GHMs.

Verschiedene Wege zur Berechnung der Nettostrahlung oder die direkte Verwendung aus sogenannten Reanalysen als Grundlage für die Berechnung der potentiellen (und aktuellen) Evapotranspiration bilden den ersten Teil dieses Teilgebietes. Basierend auf dem Vergleich mit stationsbasierten Beobachtungsdaten wurden Gütemaße berechnet. Es konnte gezeigt werden, dass die Berechnung der ausgehenden kurzwelligen Strahlungs-komponente von WaterGAP ähnlich plausibel ist wie der aktuelle Reanalyse-Ansatz von ERA-Interim, und eine insgesamt zufriedenstellende Übereinstimmung mit Messwerten vorliegt. Das ist insofern von Bedeutung, als dass, im Gegensatz zu WaterGAP, Reanalysen eine geschlossene Energiebilanz aufweisen und auf diverse Beobachtungsdaten gestützt sind. Es konnte gezeigt werden, dass die langwellige ausgehende Strahlungskomponente der ERA-Interim-Reanalyse besser mit den Beobachtungen übereinstimmt. WaterGAP könnte verbessert werden, indem der Emissionsgrad erhöht sowie eine Kopplung an die Schneedynamik entwickelt wird. Die mittlere absolute Unsicherheit der besonders relevanten Nettostrahlung wurde mit etwa 20 W m^{-2} quantifiziert. Im Vergleich mit den anderen Strahlungskomponenten ist diese die höchste. Daher, und in Bezug zur Sensitivität der simulierten Wasserressourcen auf Unsicherheiten der Nettostrahlung (Döll et al., 2016), bleibt die plausible Berechnung der Nettostrahlung eine Herausforderung für zukünftige Forschung.

Die Unsicherheiten der klimatischen Antriebsdaten (im Folgenden lediglich Antriebsdaten genannt) sind Schwerpunkt des zweiten Teils der Evaluation. Unter Annahme, dass die verwendeten Antriebsdaten in gleichem Maße Anspruch auf Plausibilität und Fehlerfreiheit haben, wurde die Sensitivität von simulierten Wasserflüssen und -speichern auf verschiedene Antriebsdaten in Modellexperimenten untersucht. Es konnte gezeigt werden, dass die Unsicherheit der Antriebsdaten hauptsächlich auf die unterschiedlich angewandte Methodik der Niederschlagskorrektur insbesondere des Schneeniederschlages zurück zu führen ist. Daneben wurden für zwei der am häufigsten verwendeten globalen

Niederschlagsdatensätze (CRU, GPCC), trotz sehr unterschiedlicher Stationsdichten, große Übereinstimmungen der kontinentbezogenen und globalen langjährigen Niederschlagsmengen gefunden, die allerdings nach Anwendung der beschriebenen Niederschlagskorrektur signifikante Differenzen aufweisen. Daher ist ein plausibler Ansatz der Niederschlagskorrektur für solche Eingangsdaten von essentieller Bedeutung für die globalskalige Modellierung der Wasserressourcen und sollte in Zukunft weiter erforscht werden. Erste Ansätze etwa von Beck et al. (2017) zeigen, dass die Kombination von Beobachtungsdaten und Modellergebnissen zu diesem Zweck von Bedeutung sein kann. Inkonsistenzen der Strahlungskomponenten bei der Kombination von zwei der am häufigsten verwendeten Antriebsdaten (WFD, WFDEI) wirken sich signifikant auf die Simulation der Wasserspeicher und -flüsse aus. Ein vorgestellter Homogenisierungsansatz reduziert diese Inkonsistenzen erheblich. Die Unsicherheit der Antriebsdaten wirkt sich nur in relativ geringem Maße auf die globalen Summen (nicht aber notwendigerweise auf Gitterzellen-Ebene bzw. Regionen) des Bewässerungswasserbedarfes aus und wird durch die Verfügbarkeit der Wasserressourcen zusätzlich verringert. Darüber hinaus führt der Kalibrierungsansatz von WaterGAP zu einer Reduzierung der Antriebsdaten-Unsicherheit auf etwa 1,6 % in den Gebieten, wo langjährige Messreihen des Durchflusses zur Verfügung stehen. Im Gegensatz dazu, ist die Unsicherheit in nicht-kalibrierten Gebieten mit 18,5 % bedeutend höher. Nichtsdestotrotz sind Unterschiede auf Gitterzellen-Ebene in beiden Gebieten erkennbar. Fehlende Durchflussmessungen oder der nicht erfolgte Datenaustausch auf etwa der Hälfte der Landoberfläche verringern die Möglichkeit, durch einen solche Kalibrierung die Wasserressourcen besser auf Beobachtungen zu stützen.

Der zweite Teilaspekt der vorliegenden Arbeit behandelt die Auswirkung der Modellentwicklung auf simulierte Ergebnisse sowie auf Modellparameter. Das Ziel war, zu zeigen, wie sich die Einbeziehung neuer Datensätze, neuer oder veränderter Prozessabbildung sowie ein aktualisierter Kalibrieransatz auswirken. Dieser Teil umfasst ebenfalls 5 Forschungsfragen.

Die Weiterentwicklung der Modellstruktur der initialen WaterGAP-Version 2.1d bis zur Version 2.2 führte zu entscheidenden Verbesserungen in der Simulation von monatlichen Durchflüssen, insbesondere in warmgemäßigten Regenklimaten sowie borealen Klimaten. Es ist davon auszugehen, dass durch diese Entwicklung die Berechnung anderer Wasserhaushaltskomponenten verbessert wurde. Nichtsdestotrotz bleibt die Verbesserung der Simulationen in Trockenklimaten eine große Herausforderung der Modellentwicklung, denn hier wird häufig der Durchfluss um den Faktor zwei überschätzt. Die Berücksichtigung des kapillaren Aufstieges aus dem Grundwasser könnte sich auf die Erhöhung der Evaporation auswirken. Zudem sind in der derzeitigen Modellstruktur spezielle Prozesse aus Trockengebieten nicht integriert. Erste Grundlagen zur Darstellung von Exfiltrationsprozessen und nachfolgender Evaporation von temporären Wasserflächen in semi-ariden / ariden Gebieten wurden durch Nagel (2012) gelegt. In WaterGAP ist der Bodenspeicher vertikal nicht differenziert, sondern in Abhängigkeit der Durchwurzelungstiefe unterschiedlich mächtig. Daher wird in trockenen Gebieten trotz eintretender Niederschläge oft eine sehr geringe relative Bodenfeuchte simuliert, die entsprechend geringe Evaporationsverluste zur Folge hat. Die vertikale Differenzierung in

zwei Bodenspeicher könnte die relative Wassersättigung im oberen Speicher infolge von Niederschlagsereignissen und damit die Evaporation erhöhen, was wiederum zu einer Reduktion des simulierten Abflusses (Durchflusses) führen kann. Natürlich ist dies mit einer weitaus komplexeren Modellstruktur verbunden, da z.B. Austauschprozesse zwischen den Bodenspeichern entsprechend parametrisiert werden müssen.

Die Berücksichtigung des Kalenderjahres, in dem das Management neugebauter Stauseen beginnt, führte zu einer wesentlichen Verbesserung der Durchfluss-Simulation, insbesondere in dem Zeitraum, bevor der Stausee existierte. Die Analyse hat jedoch auch gezeigt, dass der generische Stausee-Algorithmus Verbesserungspotential hat, da die Modellgüte für die Zeitreihe nach Inbetriebnahme der Stauseen signifikant schlechter ist als unter unbeeinflussten Bedingungen vor der Inbetriebnahme. Dies wurde exemplarisch am zeitlich verschoben simulierten Monat der höchsten Durchflüsse im Nil unterhalb des Assuan-Staudammes gezeigt. Das Verbesserungspotential von generischen (oder auch von individuellen) Stausee-Algorithmen ist jedoch eingeschränkt, da Bewirtschaftungsregeln von Staudämmen typischerweise nicht öffentlich einsehbar sind.

Die Auswirkung der Aktualisierung zweier räumlich verteilter Eingangsdaten auf Modellergebnisse wurde in Simulationsexperimenten analysiert. Ein satellitenbasierter Landbedeckungs-Datensatz (MODIS) hat dabei eine kartenbasierte räumliche Information ersetzt. Es hat sich gezeigt, dass in einer Vielzahl an Gitterzellen der Landbedeckungstyp modifiziert wurde. Allerdings wirkt sich dies im Vergleich zu der Unsicherheit von Antriebsdaten oder der Frage ob kalibriert wird (oder nicht) nur gering auf Modellergebnisse aus. Globale Summen des Durchflusses in Ozeane und Inlandsenken sowie die aktuelle Evapotranspiration unterscheiden sich nur um etwa 200 $km^3\ a^{-1}$. Änderungen in der Durchflusssaisonalität treten lediglich in Einzugsgebieten mit großflächigen Änderungen der Landbedeckung (z.B. Yangtze) auf. Bezogen auf die Modellgüte gibt es Regionen, in denen sich der simulierte Durchfluss verbessert hat, aber auch Gebiete, in denen das Gegenteil der Fall ist.

Auch die räumlich verteilte Information zum verfügbaren Bodenwasserspeicher im obersten Meter wurde auf eine Datenbasis aktualisiert, in die wesentlich mehr Profildaten eingegangen sind. Die mittleren Werte des verfügbaren Bodenwasserspeichers im oberen Meter vergrößerten sich um 7 %. Mit Ausnahme von Nordost-Eurasien erhöhten sich in den meisten Gebieten der Landfläche diese Werte, ebenso der maximal verfügbare Bodenwasserspeicher unter Berücksichtigung der Landbedeckungs-spezifischen Durchwurzelungstiefe. Die mittlere Amplitude des Bodenwasserspeichers erhöhte sich, und signifikante räumliche Unterschiede im Abfluss wurden ermittelt, auch wenn die Auswirkung auf die Summe globalskaliger Wasserhaushaltskomponenten nur gering ausfällt. Insgesamt wurde eine Verbesserung im Sinne einer leichten Verringerung des Abflusses (und entsprechende Erhöhung der aktuellen Evapotranspiration) festgestellt, die Amplituden vom Gesamtwasserspeicher erhöhten sich ebenfalls.

Der Kalibrieransatz von WaterGAP wurde angepasst, um bei Bedarf bei den Durchfluss-messungen eine mittlere Messunsicherheit in Höhe von 10 % zu berücksichtigen. Ein vierstufiger Ansatz wurde implementiert, der zu einer deutlichen Reduzierung der Anzahl

von Einzugsgebieten mit zusätzlichen Korrekturfaktoren (um 14 %) geführt hat. Mit diesem Ansatz konnte in 43 % bis 50 % der Einzugsgebiete (die Variation entsteht durch alternative Antriebsdaten) auf zusätzliche Korrekturfaktoren verzichtet werden. Darüber hinaus liegen die Werte der zusätzlichen Korrekturfaktoren für die meisten der Einzugsgebiete mit einem solchen Bedarf näher an dem Wert 1, sodass eine solche Anpassung weniger extrem ist. Die Modellgüte reduziert sich mit dem Ansatz nur geringfügig und globalskalige Summen des Durchflusses in Ozeane und Inlandsenken sowie die aktuelle Evapotranspiration unterscheiden sich lediglich um etwa 250 $km^3\ a^{-1}$ im Vergleich zur Kalibrierung ohne Messunsicherheit. Trotz der wesentlichen Verbesserungen werden noch in etwa der Hälfte der Einzugsgebiete zusätzliche Kalibrierfaktoren benötigt. Es besteht also Bedarf an alternativen Kalibrieransätzen. Erste Schritte hinsichtlich einer multikriteriellen Kalibrierung und Daten-Assimilierung von WaterGAP konnten bereits erfolgreich durchgeführt werden.

Der dritte und letzte Teil dieser Dissertation beinhaltet drei Anwendungsbeispiele von WaterGAP. Mit der derzeit aktuellsten WaterGAP Version 2.2b und unter Verwendung eines homogenisierten Antriebsdatensatzes (WFD/WFDEI) wurden die Wasserhaushaltskomponenten für zwei Zeitschritte und die globale Landfläche sowie die Kontinente berechnet. Für die Klimanormale 1971 bis 2000 wurde der Zufluss zu den Ozeanen sowie in Inlandsenken mit 40000 $km^3\ a^{-1}$ berechnet, die aktuelle Evapotranspiration mit 70500 $km^3\ a^{-1}$. Der konsumtive Verbrauch durch Wassernutzung wurde mit etwa 1 % (1000 $km^3\ a^{-1}$) des Niederschlages auf der globalen Skala quantifiziert. Mehr als 60 % der Wassernuztung findet in Asien statt. Für die im IPCC AR5 verwendete Referenzzeitreihe 1986 bis 2005 sind die Werte ähnlich. Die etwa 700 $km^3\ a^{-1}$ geringeren Niederschläge führen zu etwa 850 $km^3\ a^{-1}$ niedrigeren Durchflüssen und einer 100 $km^3\ a^{-1}$ höheren Evapotranspiration. Darüber hinaus steigt die konsumtive Wassernutzung um 120 $km^3\ a^{-1}$ in dieser Zeitreihe. Da WaterGAP in etwa 54 % der globalen Landfläche (mit Ausnahme der Antarktis und Grönland) auf mittlere Jahreswerte des gemessenen Durchflusses kalibriert ist, ist davon auszugehen, dass die Quantifizierung dieser Wasserhaushaltskomponenten gut durch Messungen gestützt wird. Eine qualitativ hochwertige Durchflussmessung von bislang nicht erfassten Einzugsgebieten, sowie der Austausch von vorhandenen Daten kann eine bessere Abschätzung der globalen Wasserressourcen und -haushaltskomponenten mit solchen Modellen in Zukunft weiter befördern.

Die globalskalige Abschätzung der Grundwasserzehrung (groundwater depletion) ist Gegenstand des zweiten WaterGAP-Anwendungsbeispieles. Zu diesem Zweck wurde in semi-ariden und ariden Gebieten der Prozess der Grundwasserneubildung unterhalb von Oberflächengewässern entwickelt, was zu einer Verbesserung der Simulation von Grundwasserzehrungen im Vergleich zu unabhängigen Messungen führte. Zusätzlich wurde erarbeitet, dass in Gebieten mit Grundwasserzehrungen eine Defizitbewässerung mit 70 % des optimalen Wasserbedarfes wahrscheinlich ist. In vielen Gebieten, in denen unabhängige Beobachtungen oder Abschätzungen vorlagen, hat die Kombination aus Grundwasserneubildung unter Oberflächengewässern sowie die Defizitbewässerung zu der besten Übereinstimmung geführt. Die unsichere Datengrundlage und zahlreiche Annahmen in Bezug auf die Wassernutzung wirken sich limitierend auf eine bessere

Abschätzung der Grundwasserzehrung aus (siehe auch Döll et al., 2016). Nach den WaterGAP-Berechnungen verdoppelte sich die globalskalige Grund-wasserzehrung zwischen 1960 bis 2000 und 2000 bis 2009. Schwerefelddaten der GRACE-Satelliten wurden hinsichtlich der Beobachtung von zeitlicher Dynamik und Trends von Wasserentnahmen untersucht. Die grobe Auflösung von GRACE, und weil GRACE Änderungen im Gesamtwasserspeicher beobachtet, nicht jedoch direkt die vom Grundwasserspeicher, bleiben wesentliche Hindernisse in diesem Zusammenhang. Insgesamt hat dieses Beispiel zeigen können, dass eine Verknüpfung von hydrologischer Modellierung mit Beobachtungsdaten die Plausibilität der Abschätzung von Grundwasser-zehrung erhöhen kann.

Im dritten Anwendungsbeispiel wurde der Einfluss des Klimawandels auf Fließregime untersucht. Eine Reihe von Indikatoren wurden entwickelt, die einen Einblick in Veränderungen von Durchflussvariabilität und -saisonalität hinsichtlich ökologisch relevanter Fließregime erlauben. Unter Verwendung von Daten zweier Klimamodelle und zweier Emmissionsszenarien wurden mit WaterGAP wesentliche Veränderungen in der Durchflussvariabilität berechnet. Zusätzlich wurde ein neuer Indikator entwickelt, mit dem der Charakter des Fließregimes (perennierend, intermittierend) bestimmt wird. Es konnte gezeigt werden, dass etwa um das Jahr 2050 mit Änderungen des Fließregimes auf rund 6 % der globalen Landfläche zu rechnen ist. Zukünftige Forschung sollte für eine robustere Quantifizierung dieser (nicht nur für die Ökologie) relevanten Fließregime-Änderungen ein multi-Modell- und multi-Szenario-Ansatz beinhalten.

In den vorangegangenen Ausführungen wurden bereits spezifische Forschungsbedarfe und Herausforderungen dargestellt. Basierend auf den Erkenntnissen dieser Dissertation können folgende Forschungsideen und Anwendungen für WaterGAP definiert werden:

- Bewertung der Rolle von Ansätzen zur Berechnung der potentiellen Evapo-transpiration unter Verwendung von Verdunstungskesseln als Referenzdaten;
- Die Einbindung eines gradientenbasierten Grundwassermodells dürfte zu einer verbesserten Simulation der Oberflächen-Grundwasser-Interaktion, einer physikalisch besser beschriebenen Grundwasserzehrung, der Möglichkeit der Einbeziehung von kapillarem Aufstieg (und der damit verbundenen erhöhten Evaporation des Bodens) sowie der Möglichkeit der Quantifizierung der Grundwassererhöhung durch Stauseen (z.B. nach dem Ansatz von Chao et al., 2008) führen;
- Die Simulation von Wasserressourcen in semi-ariden und ariden Gebieten sollte vordringlich verbessert werden. Dies könnte durch die Einbeziehung von Exfiltrationsprozessen und nachfolgender Evaporation in temporären Wasser-flächen, dem kapillaren Aufstieg von Grundwasser und der Einführung eines zweiten Bodenspeichers (um die relative Wassersättigung nach Niederschlagsereignissen und damit die Evaporation zu erhöhen) erreicht werden;
- Eine standardisierte Validierungsumgebung, die verschiedenartige Beobachtungs-daten umfasst (u.a. datenbasierte Produkte z.B. Fekete et al., 2002; Gudmundsson and Seneviratne, 2016) kann die Bewertung von Modelländerungen unterstützen;

- Das Kalibrierungs/Datenassimilations-Werkzeug von Eicker et al. (2014) und Schumacher et al. (2016b) sollte weiterentwickelt werden, um verschiedene Beobachtungsdaten (z.B. GRACE, Durchflussmessungen, Altimeterdaten) auf der globalen Skala anzuwenden.

Im Allgemeinen liefern globale hydrologische Modelle einen wertvollen Beitrag zur Quantifizierung globalskaliger Wasserflüsse und -speicher. Daher sollten diese Modelle als integrierte Werkzeuge für internationale Initiativen wie die World‘s Large Rivers Initiative (http://unesco-chair.boku.ac.at/index.php/wlri.html) oder etwa die Quantifizierung von Wasserressourcen in grenzüberschreitenden Flusseinzugsgebieten (in denen der Datenaustausch oftmals aus politischen Gründen eingeschränkt ist) verstärkt verwendet werden. Folgende Vorschläge zur robusteren Berechnung des terrestrischen hydrologischen Kreislaufes umfassen die:

- Entwicklung von konsistenten und zeitlich aktuellen Antriebsdaten unter Berücksichtigung eines plausiblen Ansatzes zur Niederschlagskorrektur;
- Durchführung von möglichst konsistenten Modellvergleichsstudien (sowohl historisch als auch mit Szenarien) mit dem Ziel, die Gründe für unterschiedliche Modellausgaben (siehe Haddeland et al., 2011; Schewe et al., 2014) zu untersuchen;
- Erhöhung der Plausibilität von historischen und szenario-gestützten Wassernutzungen zum Beispiel im Rahmen der Water Futures and Solutions Initiative (WFaS, Wada et al., 2016a);
- Entwicklung eines multi-Modell-Ansatzes für saisonale Vorhersagen (z.B. von Dürren) als gemeinschaftliche Aufgabe der globalen hydrologischen Modelle;
- Übersetzung der Ziele der HyperHydro-Initiative (http://www.hyperhydro.org/) (Wood et al., 2011; Bierkens et al., 2015) als Aufruf zu einer gemeinschaftlichen Entwicklung eines Modellsystems mit hoher räumlicher Auflösung zur Unterstützung des Wassermanagements in datenarmen Gebieten.

Alle Modelle bilden die Realität teils stark vereinfacht ab, und einige der Unsicherheiten, die in dieser Arbeit beschrieben sind, werden auch in naher Zukunft bestehen bleiben. Es konnte gezeigt werden, dass weiterhin Entwicklungsbedarf für globale hydrologische Modelle besteht und auch modell-basierte Indikatoren sowie Anwendungen weiterentwickelt werden sollten. Durch die Evaluation, Entwicklung und Anwendung des globalen hydrologischen Modells WaterGAP konnten einige Einblicke in dieses Themengebiet gegeben werden. Diese werden hoffentlich die Entwickler und Anwender von globalen hydrologischen Modellen bei der weiteren Arbeit unterstützen.

Appendix A: Impact of model modification on simulated water storages and fluxes

In this chapter, three examples of model modification and the respective impacts on simulation results are described. The first example (Appendix A1) focuses on an improved process representation, namely the inclusion of reservoir commissioning years. The second (Appendix A2) shows the effects of an updated total soil water capacity input map. Finally, a modified calibration scheme and its effect on model parameters and simulated water fluxes are presented in Appendix A3.

A1 Integrating reservoir commissioning years

A1.1 Background

In all model versions except 2.2 (ISIMIP2a), dams and their corresponding reservoirs are treated as being present over the entire simulation period. When the model is run without the reservoir algorithm, reservoirs are treated as (global) lakes. Especially when starting the model runs in the early 20th century (at which time in reality only a few reservoirs were already in operation), this can lead to an unrealistic simulated river discharge seasonality (compared to the observed one) due to 1) the presence of a water body (where in reality none was present) and 2) the reservoir algorithm (which leads to water storage and release processes that would not occur if the dam were not yet in operation). In the context of the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP), phase 2a, the participating modeling groups were asked to account for the year reservoirs were commissioned in their modeling efforts; hence this feature was included in the model version 2.2 (ISIMIP2a).

A1.2 Implementation

Based on information from the Global Reservoir and Dam database (GRanD; Lehner et al., 2011) version 1.1 as well as electronic resources (e.g., websites from the operation companies or http://www.wikipedia.org), the calendar year was obtained in which each of the 1109 currently implemented reservoirs was commissioned. In cases where more than one reservoir outflow was assigned to a grid cell, the year of the reservoir with the larger storage capacity was considered. The WaterGAP code was modified in such a way that, in the year when the dam was commissioned, reservoir area increases to its full extent (thus land area fraction is adjusted), the reservoir starts filling and the reservoir algorithm (after Hanasaki et al., 2006, described in Döll et al. (2009)) is enabled. The storage capacity of those reservoirs that are already in operation in the model initialization year is set to the maximum value. In the case of regulated lakes, in contrast, a natural water body is already present before dam construction. Several new input data were created (e.g., a file with the operation year and yearly files with spatial information on reservoir area), and a thorough

check of outflow cells and their representation in the drainage network was undertaken. In addition, the precision of input files (reservoir area and volume) was changed from integer to float values, which is especially beneficial for representing the smaller reservoirs with reasonable accuracy.

A1.3 Study setup and evaluation design

WaterGAP 2.2 (ISIMIP2a) was applied to three socio-economic scenarios, following the ISIMIP modeling protocol (https://www.isimip.org/protocol/#isimip2a). In the variable socio-economic (varsoc) scenario, reservoir operation years, as well as time-varying human water use, were taken into account to mimic the historical dynamics of these features. In the present-day socio-economic (pressoc) scenario, human water use was set to the intensity of the year 2000 for the entire simulation period, and only those reservoirs commissioned until (and including) the year 2000 are in operation for the full simulation period. The pressoc scenario is intended to represent "present-day" human impact. The model was calibrated using the varsoc scenario as described in the appendix of Müller Schmied et al. (2014), and calibrated parameter values were used for the pressoc scenario.

In order to show the effect of the reservoir operation years on simulated river discharge and derived model efficiency, discharge observation stations from the Global Runoff Data Centre (GRDC) were selected which are located downstream of reservoirs that have a considerable effect on river discharge. The frequently used degree-of-regulation index (expressed as the ratio of upstream storage capacity to mean river discharge without human impact) (Eisner, 2015; Lehner et al., 2011) was calculated for each of the 1319 grid cells that contain calibration stations. Firstly, those grid cells with an indicator value larger than 0.5 were selected, which corresponds to 257 stations. Such an indicator value means that natural river discharge can be stored theoretically for half a year within the artificial upstream reservoir(s) and should thus have a considerable impact on discharge seasonality. As the effect of river regulation should be most visible at the next station downstream of a reservoir, the closest downstream station was selected (i.e., in the case of equal upstream reservoir storage capacity the grid cell with the lowest mean naturalized river discharge). This reduced the number of evaluation stations to 214. The observed discharge time series was divided into a pre-commissioning time period and a period in which the reservoir is in operation if at least one reservoir is within the basin. If more than one reservoir is present in a basin, the commissioning year of the reservoir with the largest storage volume was taken into account (except for the two downstream stations on the Nile, where the year of the Aswan reservoir was used instead of that of Lake Victoria). The Nash–Sutcliffe efficiency E_{NS} (Nash and Sutcliffe, 1970) and the Kling–Gupta efficiency metric E_{KG} as well as the latter's components (E_{KGbeta} as a measure of mean bias, $E_{KGgamma}$ as a measure of variability and E_{KGr} as a measure of timing) (for details of E_{KG} see paper 1 and Gupta et

al., 2009; Kling et al., 2012) were used to evaluate the simulated monthly river discharge against observed values separately for the pre- and post-commissioning time period.

A1.4 Results

The development of global-scale reservoir storage for the varsoc and pressoc scenario is shown in Fig. A1. Especially between the 1950s and 1980s, storage capacity (solid blue line) strongly increased. The simulated actual reservoir storage differs considerably between the two scenarios up to the 2000s and converges afterward.

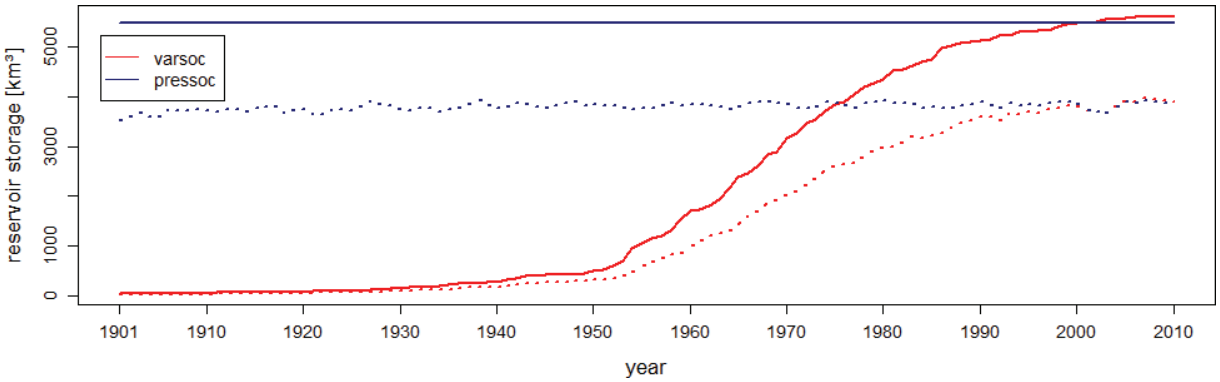


Fig. A1: The global sum of maximum storage capacity and simulated actual storage for two scenarios: starting the model run with reservoir status of the year 2000 (comparable with model runs without implementation of the operation year pressoc, blue color) and integrating reservoir operation years (varsoc, red color). Solid lines represent the maximum reservoir storage capacity, whereas dashed lines show the actual storage. Values are given in km^3.

The effect of implementing reservoir commission years into WaterGAP on model efficiencies is shown in Fig. A2. Only the 214 basins with a high impact from reservoirs are displayed, and the time series are divided into pre- and post-commissioning years. In general, the efficiencies of simulations before the reservoir was commissioned are better for both scenarios compared to the time series after the dam was commissioned. In addition, the varsoc scenario fits better than the pressoc in the pre-commission time series. This is reflected by better E_{NS} and E_{KG} values for varsoc and pre-reservoir-commission time series. The single components of E_{KG} are also improved in that time period. For the years after dams began operation, the efficiencies of varsoc and pressoc are very similar, as the reservoir algorithm is enabled in both cases, and the simulated river discharge of each is similar. The generic reservoir algorithm is not perfect for the majority of reservoirs because the efficiency criteria of the varsoc scenario before reservoirs were commissioned are better than after applying the reservoir algorithm (Fig. A2).

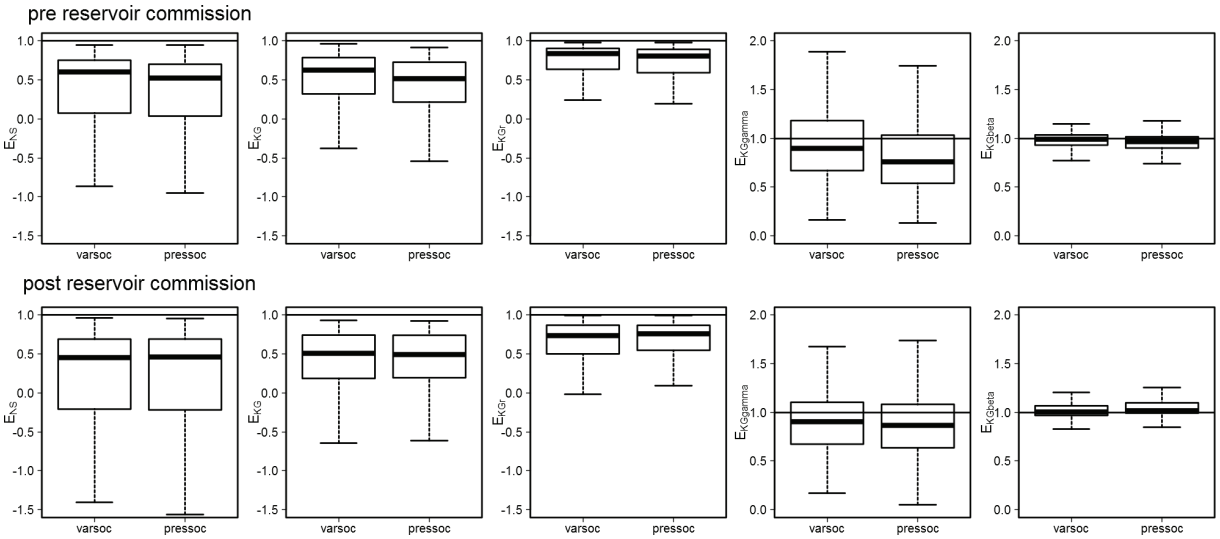


Fig. A2: Nash–Sutcliffe and Kling–Gupta efficiency metrics as well as their components for the 214 gauging stations considerably impacted by reservoir flow regulation. Results are shown for varying historical human impact (varsoc) and constant present-day human impact (pressoc) for the time period before reservoirs were commissioned (top) and afterwards (bottom).

The effect of the inclusion (or lack of inclusion) of the commissioning year is visualized for three reservoirs (Table A1) in Fig. A3. For the Nile, the timing of peak flow is much improved for varsoc compared to pressoc for the pre-commissioning years. Afterwards, model simulations are comparable and the reservoir algorithm is not able to represent the reservoir operation scheme of the Aswan Dam as peak flows are shifted by around 6 months. The later simulated peak flow is also visible in the pre-commissioning time period for the varsoc scenario. In the Volta basin, the filling-up of the reservoir after 1965 is well simulated using the varsoc scenario. For both time periods, varsoc seasonality is closer to the observed values, but differences are smaller than those of the Nile. A clear benefit of the varsoc scenario and the pre-commissioning time period is visible for the Salado gauging station. Even though there is a shift in timing, peak, and low flows are well captured when the year of commissioning is taken into account.

Table A1: Characteristics of selected river discharge stations and upstream reservoirs.

River	Station	Lat	Lon	Upstream reservoir capacity [km³]	Reservoir operation year *	Indicator of flow regulation
Nile	Aswan Dam	23.75	32.75	376	1970	5.6
Volta	Senchi	6.25	0.25	155	1965	4.7
Salado	El Arenal	-25.75	-63.75	3	1973	4.1

* of the reservoir with largest storage capacity in the case of more than one upstream

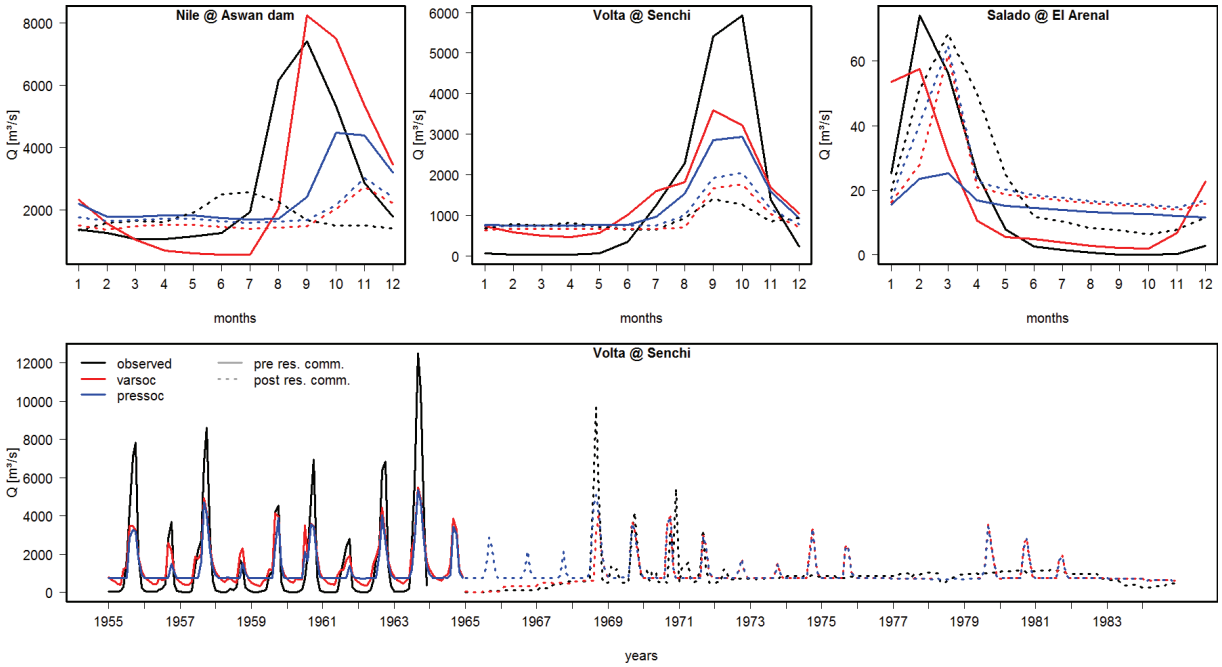


Fig. A3: River discharge seasonality and monthly river discharge for selected stations (Table A1) where the reservoir began operation during the calibration time series (i.e., the time series in which discharge observations were available).

A1.5 Summary

The implementation of the reservoir operation years generally improves the simulation of river discharge for the years before the reservoir was commissioned, which is visible in better agreement to observed river discharge both for efficiency metrics and for discharge seasonality. However, the generic reservoir algorithm is not able to handle all reservoirs sufficiently.

A2 Adapting input map of total soil water capacity

A2.1 Background

In WaterGAP, the maximum soil water storage S_{max} is calculated by using a spatially distributed input map of total available water capacity (TAWC) for the top meter, which is then multiplied by the land-cover-specific effective rooting depth of the specific grid cell (see the appendix in Müller Schmied et al. (2014) for description). The original TAWC map was derived from Batjes (1996) based on the World Inventory of Soil Emission Potentials (WISE) database. Within this version of WISE, 4353 globally distributed soil profiles were included which were used to derive pedotransfer functions (PTFs) to calculate TAWC. The updated WISE database contains information on 10250 soil profiles (Batjes, 2012), is available at 5-arc-minute resolution and also includes information on available water capacity that is now available for different soil depth intervals.

A2.2 Implementation

The dominant TAWC unit per 5-arc-minute grid cell was taken into account. For each of the five 20 cm soil layers, TAWC (provided in cm m^{-1}) was divided by 5 (to derive TAWC per 0.2 m layer), and TAWC was subsequently summed up over the profile and converted to mm m^{-1}. Those layers without TAWC values (e.g., due to shallow soils or too many rocks underground) were ignored in the calculation. The average TAWC of the 0.5 x 0.5° grid cell was obtained using zonal statistics within ESRI ArcGIS software. Around 2500 grid cells did not contain values (e.g., because they are fully covered with water or rocks). Here, the values from Batjes (1996) were used instead.

A2.3 Study setup and evaluation design

The differences between the two input versions are displayed in Fig. A4, which shows TAWC in the first meter of the soil profile for Batjes (2012) and its differences to Batjes (1996), as well as those values for the maximum soil water storage (where TAWC is simply multiplied by rooting depth of the land cover class) (Fig. A4). To assess the effect on storage variations, the mean annual amplitude of the soil water storage was calculated. WaterGAP 2.2b with a homogenized combination of WATCH Forcing Data based on ERA40 (until 1978) and WATCH Forcing Data based on ERA-Interim afterward was run as a “NoCal” scenario; i.e., all calibration parameters were set to 1.0 globally. The only difference between both model runs is the TAWC input. Mean amplitude was obtained from monthly maximum minus monthly minimum soil water storage per year for the time period 1981–2010 (Fig. A5). Finally, annual average runoff from land (same time period) was assessed (Fig. A6).

A2.4 Results

Global average TAWC of Batjes (2012) is 112 mm m^{-1}, whereas it is 105 mm m^{-1} for Batjes (1996). The red colors in Fig. A4 (c) show areas where the updated version has larger values than the previous version (vice versa for bluish colors). Except for large parts of northeastern Eurasia, most of the global land areas have increased TAWC values in the updated version. This general picture is still the same when considering the rooting depth, i.e., S_{max} (Fig. A4 b, d).

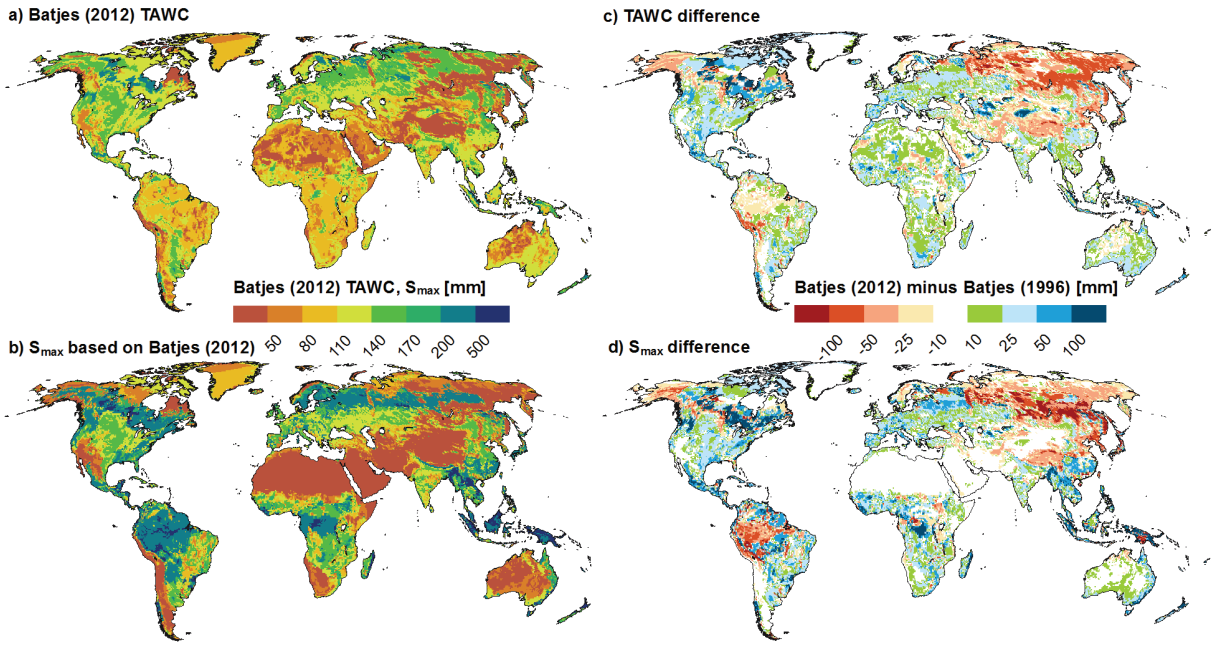


Fig. A4: Total available soil water capacity (TAWC) from Batjes (2012) (a); maximum soil water storage S_{max}, which is derived by multiplying TAWC with a land cover dependent rooting depth (b); and absolute difference between TAWC (c) and S_{max} (d) calculated as Batjes (2012) minus Batjes (1996). Red colors indicate higher values in Batjes (1996).

The increasing S_{max} corresponds to an increased soil water storage amplitude at least for some areas (including the tropics except for the upper Amazon, eastern North America, Europe) and to a decrease in the upper Amazon and northeastern Eurasia (Fig. A5). The global mean amplitude of soil water storage increased from 58.7 mm (Batjes, 1996) to 61.8 mm (Batjes, 2012).

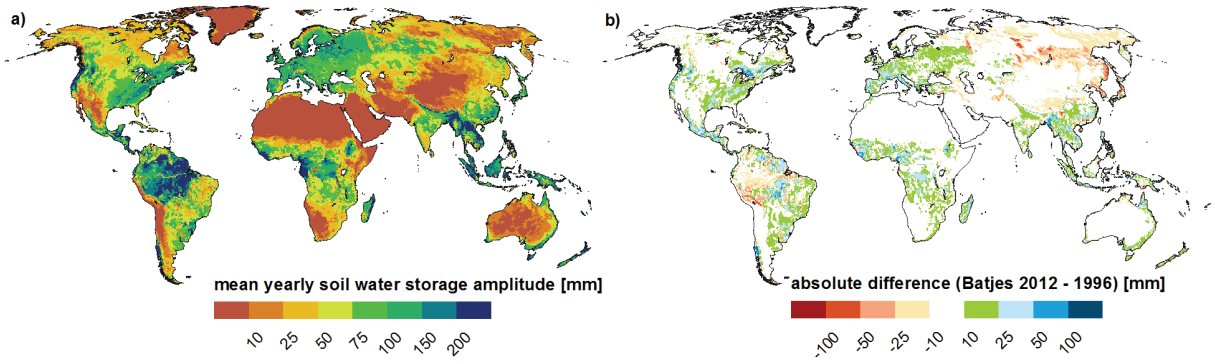


Fig. A5: Mean yearly soil water storage amplitude based on a non-calibrated WaterGAP 2.2b run and the TAWC input of Batjes (2012) (a), and absolute differences to the version of Batjes (1996) (red colors indicate higher values in Batjes (1996)) (b). Units are given in mm.

When analyzing mean runoff from land for the same time period (Fig. A6), the spatial pattern differs from the previous figures. Western North America, southern Africa, main parts of Australia and parts of Europe show reduced runoff values when TAWC from Batjes (1996) is used instead of the updated version. In these regions, evapotranspiration increases when the updated TAWC input map is used. Global average runoff (evapotranspiration) with the updated TAWC version is estimated at 328 (348) mm yr^{-1}, compared to 331 (345) mm yr^{-1} for Batjes (1996). Hence, regional differences are balanced out when summed up for the global scale.

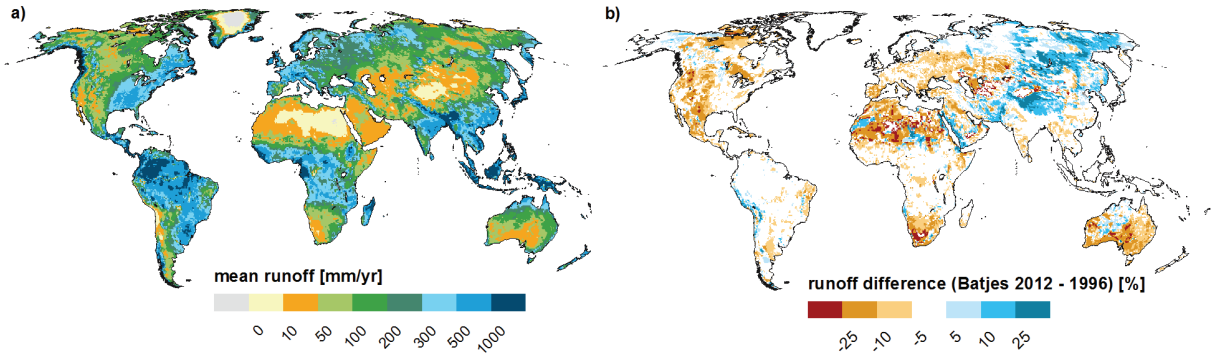


Fig. A6: Mean yearly runoff from land [mm yr^{-1}] with a non-calibrated WaterGAP 2.2b run and the TAWC input of Batjes (2012) (a) and relative differences [%] to the version of Batjes (1996) (red colors indicate higher values in Batjes (1996)) (b).

A2.5 Summary

Updating the spatial input map of total available soil water capacity leads to spatial differences in maximum soil water storage, the annual amplitude of soil water storage and in average annual runoff. In global numbers, the effect – a 3 mm difference – is not very large. However, evapotranspiration increases with increasing maximum soil water storage and vice versa. A comparison with observed river discharge by using a calibrated version with each of the input data could help to assess which input map might be more accurate.

A3 A modified calibration approach to account for uncertainty in river discharge measurements

A3.1 Background

So far, WaterGAP has employed a calibration scheme (Cal_{std}) that forces simulated long-term average river discharge to match the observed value within a maximum permitted deviation of ± 1% (Döll et al., 2003; Hunger and Döll, 2008). During calibration, a maximum of three parameters are adjusted: (1) the basin-wide uniform parameter γ (or β in Bergström, 1995) determines the amount of water draining the soil column within a given time step. High parameter values reduce outflow even if soil water storage is nearly saturated, and vice versa (Müller Schmied et al., 2014, their appendices A3, B1). (2) The area correction factor (CFA) is computed if adjusting γ in the limits 0.1 and 5.0 did not yield an acceptable runoff estimate. Allowed to range between 0.5 and 1.5, CFA adjusts the runoff of each grid cell (so-called net cell runoff, computed as the outflow of the grid cell minus the inflow to it) in each grid cell of the basin. Net cell runoff is multiplied by the value symmetrically around 1.0, dependent upon the sign of the long-term water balance in the individual grid cell (see also Hunger and Döll, 2008). (3) In cases where γ and CFA are not sufficient to match the observed river discharge, the station correction factor (CFS) is applied that simply multiplies river discharge in the grid cell in which the observation station is located with a factor (without value limitation) to match the observed long-term average annual river discharge. This is mainly done to avoid error propagation to the next downstream basin.

The calibration approach described above implicitly assumes that there is no measurement uncertainty in the long-term average annual river discharge data used for parameter estimation which was obtained from the Global Runoff Data Center (GRDC). A recent study of Coxon et al. (2015), however, showed that even for presumably high-quality discharge stations in the United Kingdom measurement uncertainty of, e.g., high-flows or low-flows mostly ranges between 20 and 40% but may be as high as 80% in single cases. It is to be expected that this uncertainty range holds true (or is even higher) for other discharge observation stations as well, especially in regions where high-quality measurements are not available. Forcing WaterGAP to a maximum deviation of ±1% therefore leads to over-justifying the accuracy of discharge observations. Furthermore, Eicker et al. (2014) and Schumacher et al. (2016a) use WaterGAP in a calibration/data assimilation environment including model parameter sets in an ensemble approach. Hence, it is required to include an uncertainty range of the model parameters and/or data sets. Finally, previous analysis has shown that the usage of CFA and CFS may lead to physically implausible (negative) values for actual evapotranspiration for the grid-cell level, which needs to be avoided.

The modification of the calibration routine ($Cal_{10\%}$) now considers (optionally) a 10% measurement uncertainty in annual observed river discharge values. The value of 10% was chosen as an arbitrary value as Coxon et al. (2015) did not provide errors at the annual level and assuming that uncertainties level out for yearly averages.

A3.2 Implementation

The calibration routine was modified to a four-stage scheme that is indicated by a specific calibration status (CS):

1) The basin is sufficiently calibrated to long-term average annual observed river discharge Q_{obs} (99–101%) by varying γ in its assigned limits only.
2) A 10% uncertainty (referring to the 90–110% range) of Q_{obs} is assumed and is again tested by varying γ only.
3) CFA has to be applied to match Q_{obs} with 10% uncertainty (referring to the 90–110% range).
4) A CFS value is needed to match Q_{obs} with 10% uncertainty (referring to the 90–110% range).

For this analysis, WaterGAP 2.2b was calibrated using a homogenized combination of WATCH Forcing Data and WATCH Forcing Data based on ERA-Interim (WFDEI_hom; see Müller Schmied et al., 2016b). In addition, the impact of climate input uncertainty was considered by using alternative climate input data from CRU TS 3.23, as well as CRU TS 3.23 plus GPCC v7 precipitation.

A3.3 Effect on calibration parameters

Including a 10% Q_{obs} uncertainty (calibration status > 1) in the calibration scheme leads to a slightly higher mean γ of 3.93, compared to 3.89 (WFDEI_hom) which means that more precipitation is kept as soil water. For 57 (61) basins, γ values are smaller (larger) with $Cal_{10\%}$ which is a result of a complex interplay of higher calibration demand of downstream basins in cases where Q_{obs} uncertainty is applied upstream. CFA is not required at all for 659 basins with $Cal_{10\%}$ (sum of CS1 and CS2 in Table A7), whereas this is the case for only 512 basins with Cal_{Sta}. In 540 basins, CFA is closer to unity with $Cal_{10\%}$ (one basin with Cal_{Sta}); for 778 basins, CFA is equal in both calibration setups (thereof 512 basins with CFA = 1.0, 236 basins with CFA = 0.5 and 30 basins with CFA = 1.5). The modified calibration approach leads to a general reduction of CFS usage in the case of WFDEI_hom to 386 river basins (compared to 462 basins in Cal_{Sta}). In 452 basins, CFS values are closer to unity for $Cal_{10\%}$, whereas this is the case for 12 basins and Cal_{Sta} (for 855 basins, CFS values remains one among the two calibration setups). As shown in Table A7, for 43 to 50% of the river basins, no additional calibration parameter (CFA, CFS) is needed in $Cal_{10\%}$. Using the modified approach, additional calibration parameters are avoided in around 14% of the river basins. The effect

of the calibration scheme on the parameters is shown graphically in Fig. A7. Parameter values for γ are close to the 1:1 line and thus not very sensitive (except for the 9% of the basins with changes). In contrast, CFA and CFS are strongly influenced by the assumed 10% uncertainty.

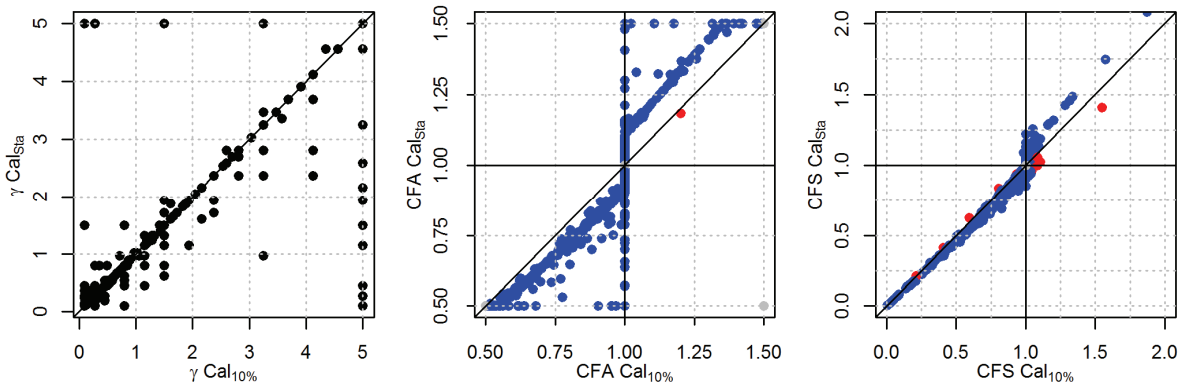


Fig. A7: Scatter plots of calibration parameters γ (left), CFA (center) and CFS (right) for calibration considering ($Cal_{10\%}$) and not considering (Cal_{Sta}) 10% uncertainty in observed river discharge. Blue colors for CFA and CFS indicate basins with a value closer to unity, and red colors the opposite. Grey colors (CFA) indicate no changes in the parameter value. Horizontal and vertical lines for CFA and CFS indicate the value of unity, whereas the diagonal line in all three plots shows the 1:1 line.

The spatial pattern of calibration parameters is shown in Fig. A8 for the WFDEI_hom climate input data. In the majority of river basins γ is at the lower or upper limit (210 basins with γ = 0.1, 675 with γ = 5.0) for $Cal_{10\%}$, whereas these numbers reduce to 199 and 665 for Cal_{Sta}. This effect results from uncertainty propagation to downstream basins, i.e., when observed discharge in the upstream basin is increased or reduced by 10%. The differences in γ as calculated with $Cal_{10\%}$ minus Cal_{Sta} are shown in the top row of Fig. A8. CFA values are generally closer to unity with $Cal_{10\%}$ but still are at the minimum value of 0.5 for 232 basins and at the maximum value of 1.5 for 32 basins (284 and 50 basins with Cal_{Sta} respectively). Values for CFS are also closer to unity with $Cal_{10\%}$ (see also Fig. A7). Hence, the modified calibration scheme reduces the need for strong calibration parameters (CFA/CFS). However, CFS values are below 0.5 in many semi-arid and arid regions which means that WaterGAP over-estimates river discharge by at least a factor of 2.

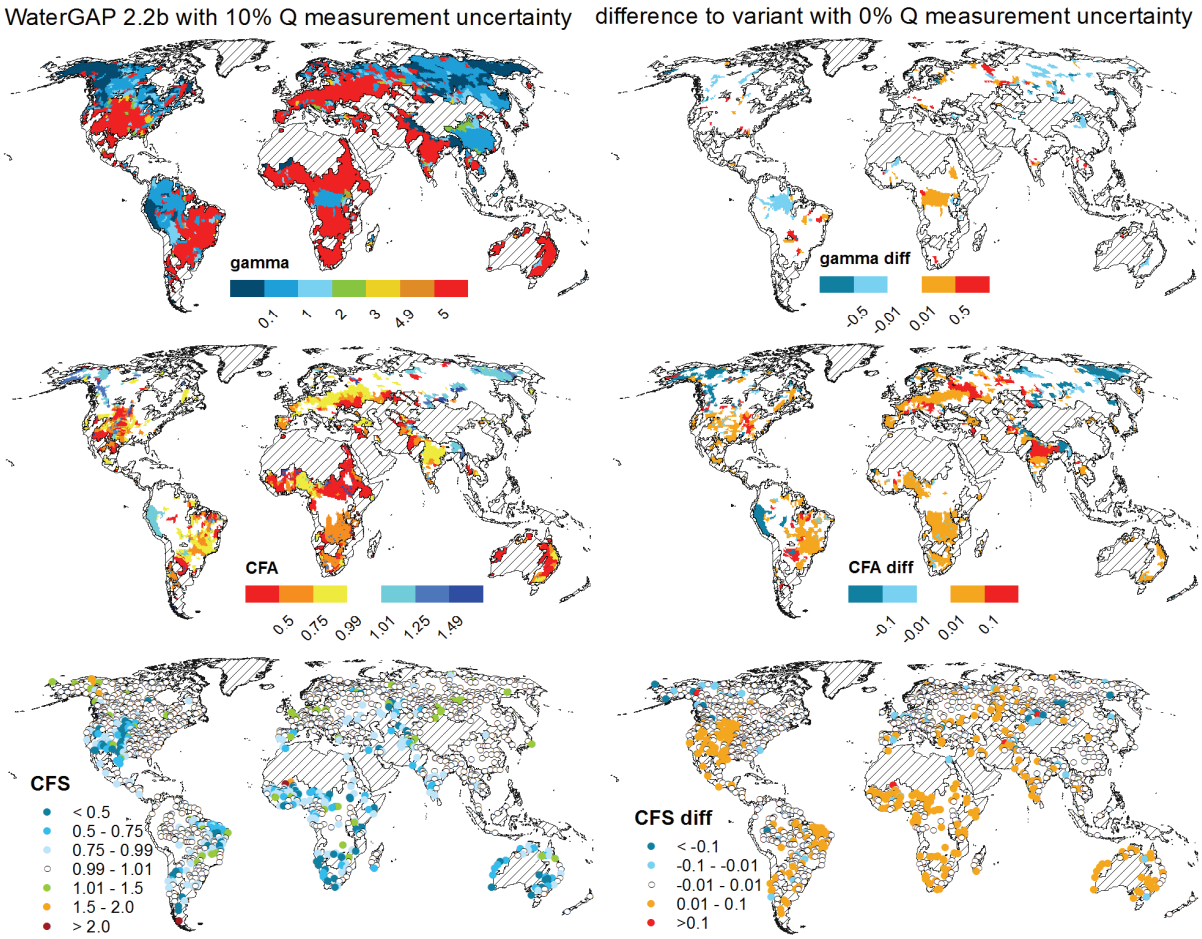


Fig. A8: Spatial distribution of calibration parameters γ (top row), CFA (center) and CFS (bottom) for WaterGAP 2.2b calibration forced with WFDEI_hom (left column), considering a 10% uncertainty of mean annual observed river discharge and differences to a calibration not considering this uncertainty (right column, calculated as $Cal_{10\%}$ minus Cal_{Sta}). Reddish colors in the right column indicate higher parameter values for $Cal_{10\%}$. Hatched areas represent non-calibrated regions.

The effect of climate input data on calibration status is shown in Table A7. Similar to the calibration with the homogenized WFD/WFDEI climate input data, WaterGAP 2.2b was calibrated with CRU TS 3.23 (CRU) as well as with CRU TS 3.23 but GPCCv7 for precipitation (CRUGPCC; except for the year 2014, when CRU precipitation was used). The number of basins with specific calibration status varies greatly, implying a strong effect of climate input data on calibration parameters. From the three climate input data sets investigated here, WFDEI_hom has the smallest number of basins that require CFA and CFS.

Table A7: Calibration status (CS see Appendix A3.2) with number of calibration basins (#) and corresponding percentage (%) of all 1319 calibration basins of the WaterGAP 2.2b calibration forced with three climate input data sets (homogenized WFD+WFDEI (WFDEI_hom), CRU TS 3.23 (CRU), CRU TS 3.23 but GPCCv7 for precipitation (CRUGPCC)).

CS	WFDEI_hom		CRU		CRUGPCC	
	#	%	#	%	#	%
1	475	36	399	30	394	30
2	184	14	165	13	183	14
3	274	21	326	25	356	27
4	386	29	429	33	386	29

A3.4 Effect on discharge seasonality, model efficiency and water balance components

Discharge seasonality is affected by the modified calibration routine in most cases where the calibration status > 1 was assigned to the selected basins (that are used in Müller Schmied et al., 2014). In Fig. A8, the two seasonality graphs are very close for calibration status 1. As here the modification of the γ parameter alone is sufficient to match long-term average river discharge, differences in the graphs occur only downstream of basins where the observed discharge was modified by 10% (e.g., in the case of the Congo). The differences in the simulated discharge seasonality are dependent on the calibration status and the value of the parameters CFA and CFS. In some cases, simulated river discharge seasonality fits better to the observed ones with $Cal_{10\%}$; in others it does not. For example, the peak flow of the Orange River fits perfectly to the observed value, but, e.g., the peak flow of the Rio Paraá River is significantly overestimated when the 10% uncertainty of discharge observations is considered. For the Rhine basin, the modified calibration scheme leads to even higher seasonal amplitude than Cal_{Sta} and the observations. Hence, the choice of the calibration scheme influences not only the parameter values but also the seasonality.

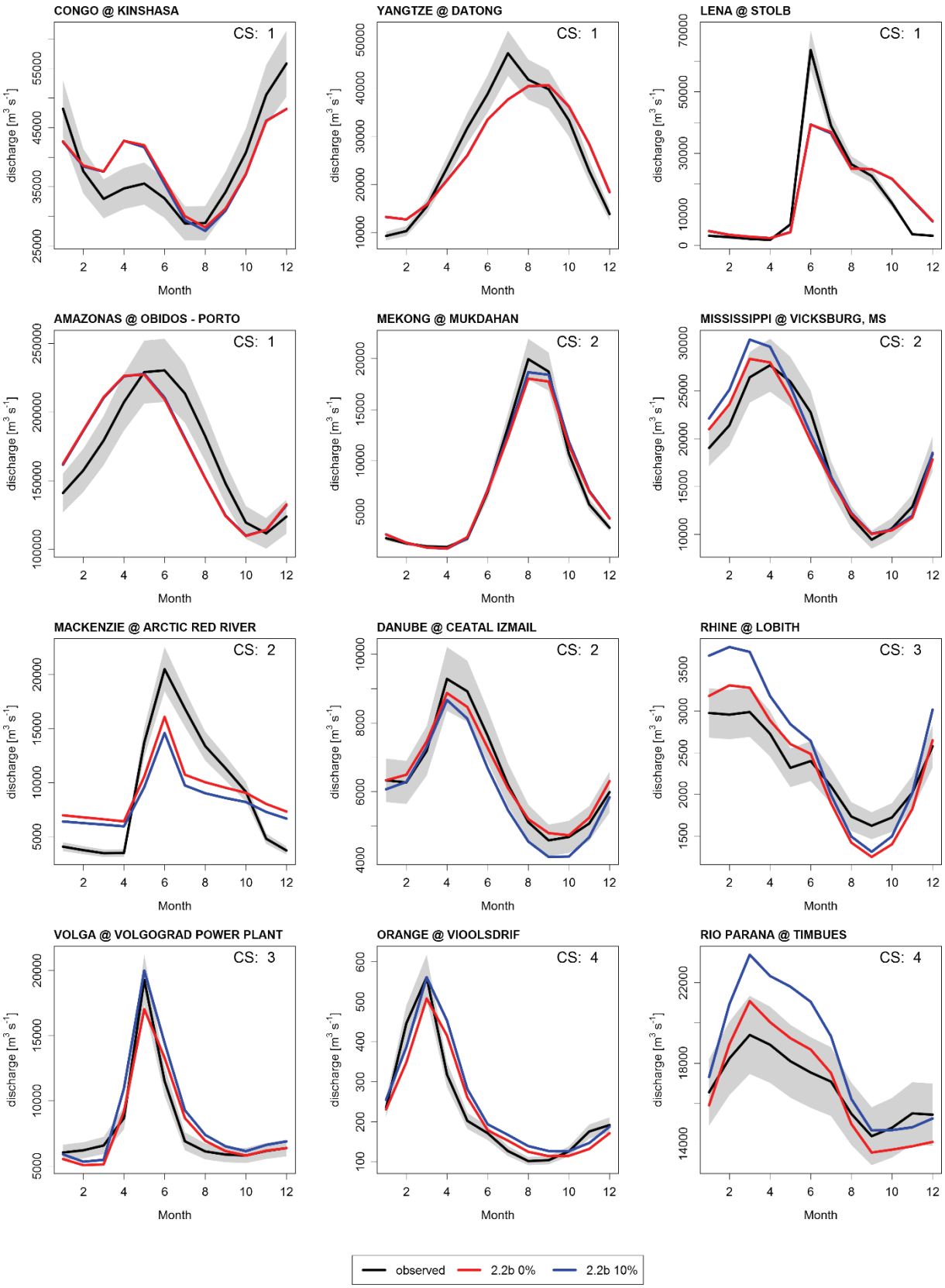


Fig. A8: Observed and simulated discharge seasonality of WaterGAP 2.2b runs with $Cal_{10\%}$ (blue color), Cal_{Sta} (red) and observed values (black) including a 10% variation of mean monthly values (grey) and an indicator of calibration status (CS; see Sect. A3.2) for the large river basins that were used in Müller Schmied et al. (2014).

Because differences in seasonality were observed (Fig. A8), the popular Nash-Sutcliffe model efficiency (NSE) was calculated and aggregated separately as mean values for each calibration status (Table A8) to test whether model performance differences between the calibration schemes exist. Note that WaterGAP is calibrated for the stations used in this assessment but not for monthly flows. Thus, seasonal dynamics and monthly variations are not calibrated using the discharge observations. The comparison was made in two ways. Firstly, the observed river discharge was taken as a reference of a common way of comparing modeling output with observations. However, as the observed river discharge can be modified during calibration, this also leads to a problematic evaluation of the effect of the calibration scheme itself. In order to allow for a fair comparison of NSE values, observed monthly river discharge was increased (reduced) by 10% for $Cal_{10\%}$ in the basins where CS is larger than 1 and γ is at the upper (lower) limit. The values in Table A8 show only slight differences between the model results when observed river discharge is modified accordingly to the calibration. In contrast, when the original observations were taken as a reference, model performance decreased. Interestingly, CS 2 and 3 provide the best median NSE values, whereas the basins where all calibration parameters are needed show the weakest performance.

Table A8: Nash–Sutcliffe efficiency for WaterGAP 2.2b including/excluding a 10% uncertainty of observed mean annual river discharge according to calibration status (CS). In column "Q_{orig}", monthly observed river discharge was taken from the GRDC database, whereas in column "Q_{mod}" the monthly river discharge values of a specific station were modified by 10% in case this was done during the calibration.

CS	$Cal_{10\%}$		Cal_{Sta}
	Q_{orig}	Q_{mod}	Q_{orig}
1	0.52	0.52	0.53
2	0.60	0.63	0.62
3	0.59	0.62	0.62
4	0.41	0.44	0.45
All basins	0.53	0.55	0.54

The general calibration approach of WaterGAP varies the proportion of evapotranspiration and runoff via the parameters γ and the area correction factor (CFA Müller Schmied et al., 2014, their eq. A11, Sect. B1). Therefore, the actual evapotranspiration and river discharge (where CFS is also modifying simulation results) are modified, albeit only to a minor extent (Table A9). The slightly higher (lower) river discharge (actual evapotranspiration) in the case of $Cal_{10\%}$ could be related to a complex interplay of calibration parameters in calibrated and non-calibrated regions (due to regionalization of γ). In calibrated (non-calibrated) regions, actual evapotranspiration is 186 (48) $km^3\ yr^{-1}$ lower for $Cal_{10\%}$, whereas river discharge is higher, at 186 (46) $km^3\ yr^{-1}$. In both regions, γ has slightly higher mean values for $Cal_{10\%}$. As

it is expected that higher γ would lead to less runoff (discharge), and this is in contrast to the increased discharge of $Cal_{10\%}$, a thorough assessment of model parameter behavior was done at the basin scale. In basins with high γ values (> 4) for Cal_{Sta}, γ for $Cal_{10\%}$ is lower or remain the same, irrespective of whether observed river discharge is modified by 10% or not. In contrast, in basins with low γ values (< 1) for Cal_{Sta}, γ for $Cal_{10\%}$ is higher only when observed river discharge has to be adapted by 10%. In all cases, mean CFA and CFS are closer to 1. Hence, the overall increase of river discharge in $Cal_{10\%}$ relates to basin-specific characteristics during the calibration. It is assumed that for non-calibrated regions basin-specific characteristics also lead to a slight increase in river discharge for $Cal_{10\%}$ even though mean γ increases.

Table A9: Global-scale water balance components (except for those of Antarctica and Greenland) for WaterGAP 2.2b for the two calibration schemes and 1971–2000. All units are given in $km^3\ yr^{-1}$.

Variable	$Cal_{10\%}$	Cal_{Sta}
Precipitation	111 602	111 602
River discharge*	40 088	39 855
Actual evapotranspiration	70 623	70 857
Actual use from surface water	1 106	1 106
Net abstraction from groundwater	-123	-123
Actual water consumption	983	982
Change in total water storage	-92	-93
Absolute water balance error	0	0

* computed as sum of net cell runoff

A3.5 Summary

The modified calibration routine for WaterGAP 2.2b takes a 10% uncertainty of mean annual observed river discharge in a four-stage calibration approach into account. Compared to a version that does not consider this uncertainty, fewer additional calibration parameters (CFA, CFS) are required; and if they are required, mean parameter values are closer to unity. This behavior can be seen as successful, as the γ parameter (which can be described physically) itself has a greater influence on model results and physically implausible effects should reduce when disaggregating the additional calibration parameters to the grid-cell level. However, calibration parameters and the need to modify observed river discharge are strongly influenced by climate input data uncertainty. Discharge seasonality differs in basins where the uncertainty of observed river discharge is assumed. The Nash–Sutcliffe model efficiency is not affected significantly, and global-scale river discharge is increased slightly when discharge uncertainty is taken into account.

References

Abtew, W., Obeysekera, J. and Iricanin, N.: Pan evaporation and potential evapotranspiration trends in South Florida, Hydrol. Process., 25(6), 958–969, doi:10.1002/hyp.7887, 2011.

Adam, J. C. and Lettenmaier, D. P.: Adjustment of global gridded precipitation for systematic bias, J. Geophys. Res., 108(D9), 4257, doi:10.1029/2002JD002499, 2003.

Alcamo, J., Döll, P., Henrichs, T., Kaspar, F., Lehner, B., Rösch, T. and Siebert, S.: Development and testing of the WaterGAP 2 global model of water use and availability, Hydrol. Sci. J., 48(3), 317–337, doi:10.1623/hysj.48.3.317.45290, 2003.

Allen, R. G., Pereira, L. S., Howell, T. A. and Jensen, M. E.: Evapotranspiration information reporting: I. Factors governing measurement accuracy, Agric. Water Manag., 98(6), 899–920, doi:10.1016/j.agwat.2010.12.015, 2011.

Al-Yaari, A., Wigneron, J. P., Ducharne, A., Kerr, Y., de Rosnay, P., de Jeu, R., Govind, A., Al Bitar, A., Albergel, C., Muñoz-Sabater, J., Richaume, P. and Mialon, A.: Global-scale evaluation of two satellite-based passive microwave soil moisture datasets (SMOS and AMSR-E) with respect to Land Data Assimilation System estimates, Remote Sens. Environ., 149, 181–195, doi:10.1016/j.rse.2014.04.006, 2014.

Aus Der Beek, T. and Teichert, E.: Global simulation of permafrost distribution in the past, present, and future using the frost number method, in Proceedings of the Ninth International Conference on Permafrost, vol. 1, edited by D. L. . Kane and K. M. Hinkel, pp. 71–76, University of Alaska Fairbanks, 2008.

Austin, A. T., Yahdjian, L., Stark, J. M., Belnap, J., Porporato, A., Norton, U., Ravetta, D. A. and Schaeffer, S. M.: Water pulses and biogeochemical cycles in arid and semiarid ecosystems, Oecologia, 141(2), 221–235, doi:10.1007/s00442-004-1519-1, 2004.

Batjes, N. H.: Development of a world data set of soil water retention properties using pedotransfer rules, Geoderma, 71(1–2), 31–52, doi:10.1016/0016-7061(95)00089-5, 1996.

Batjes, N. H.: ISRIC-WISE derived soil properties on a 5 by 5 arc-minutes global grid (ver. 1.2), ISRIC Rep., (2012/01), 52, 2012.

Baumgartner, A. and Reichel, E.: Preliminary results of new investigations of world's water balance, in World Water Balance, Proc. Symp. on the World's Water Balance (Reading, July 1970), vol. 3, pp. 580–592, IAHS Publication 93, 1972.

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

Beck, H. E., van Dijk, A. I. J. M., Levizzani, V., Schellekens, J., Miralles, D. G., Martens, B. and de Roo, A.: MSWEP: 3-hourly 0.25° global gridded precipitation (1979-2015) by merging gauge, satellite, and reanalysis data, Hydrol. Earth Syst. Sci., 21, 589-615, doi:10.5194/hess-21-589-2017, 2017.

Berger, M., van der Ent, R., Bach, V., Brochnow, K. and Finkbeiner, M.: A new method for assessing impacts of water use in life cycle assessment, Proc. Vth Conf. Life Cycle Assess., 22–26, 2013.

Bergström, S.: The HBV model, in Computer models of watershed hydrology, edited by V. P. Singh, pp. 443–476, Water Resources Publications, Lone Tree, USA, 1995.

Bernacchi, C. J. and VanLoocke, A.: Terrestrial ecosystems in a changing environment: A dominant role for water, Annu. Rev. Plant Biol., 66(1), 599–622, doi:10.1146/annurev-arplant-043014-114834, 2015.

Beven, K.: A manifesto for the equifinality thesis, J. Hydrol., 320(1–2), 18–36, doi:10.1016/j.jhydrol.2005.07.007, 2006.

Beven, K. J.: Rainfall-Runoff Modelling. The Primer, John Wiley & Sons Ltd., Chichester, 2001.

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

Bierkens, M. F. P.: Global hydrology 2015: State, trends, and directions, Water Resour. Res., 51(7), 4923–4947, doi:10.1002/2015WR017173, 2015.

Bierkens, M. F. P., Bell, V. A., Burek, P., Chaney, N., Condon, L. E., David, C. H., de Roo, A., Döll, P., Drost, N., Famiglietti, J. S., Flörke, M., Gochis, D. J., Houser, P., Hut, R., Keune, J., Kollet, S., Maxwell, R. M., Reager, J. T., Samaniego, L., Sudicky, E., Sutanudjaja, E. H., van de Giesen, N., Winsemius, H. and Wood, E. F.: Hyper-resolution global hydrological modelling: what is next? “Everywhere and locally relevant,” Hydrol. Process., 29(2), 310–320, doi:10.1002/hyp.10391, 2015.

Boulay, A.-M., Motoshita, M., Pfister, S., Bulle, C., Muñoz, I., Franceschini, H. and Margni, M.: Analysis of water use impact assessment methods (part A): evaluation of modeling choices based on a quantitative comparison of scarcity and human health indicators, Int. J. Life Cycle Assess., 20(1), 139–160, doi:10.1007/s11367-014-0814-2, 2015.

Buma, W., Lee, S.-I. and Seo, J.: Hydrological evaluation of Lake Chad Basin using space borne and hydrological model observations, Water, 8(5), 205, doi:10.3390/w8050205, 2016.

Butts, M. B., Payne, J. T., Kristensen, M. and Madsen, H.: An evaluation of the impact of model structure on hydrological modelling uncertainty for streamflow simulation, J. Hydrol., 298(1–4), 242–266, doi:10.1016/j.jhydrol.2004.03.042, 2004.

Chao, B. F., Wu, Y. H. and Li, Y. S.: Impact of artificial reservoir water impoundment on global sea level, Science, 320(5873), 212–214, doi:10.1126/science.1154580, 2008.

Chen, J. L., Wilson, C. R., Tapley, B. D., Scanlon, B. and Güntner, A.: Long-term groundwater storage change in Victoria, Australia from satellite gravity and in situ observations, Glob. Planet. Change, 139, 56–65, doi:10.1016/j.gloplacha.2016.01.002, 2016.

Clark, E. A., Sheffield, J., van Vliet, M. T. H., Nijssen, B. and Lettenmaier, D. P.: Continental runoff into the oceans (1950–2008), J. Hydrometeorol., 16(4), 1502–1520, doi:10.1175/JHM-D-14-0183.1, 2015.

Coxon, G., Freer, J., Westerberg, I. K., Wagener, T., Woods, R. and Smith, P. J.: A novel framework for discharge uncertainty quantification applied to 500 UK gauging stations, Water Resour. Res., 51(7), 5531–5546, doi:10.1002/2014WR016532, 2015.

Decharme, B., Alkama, R., Papa, F., Faroux, S., Douville, H. and Prigent, C.: Global off-line evaluation of the ISBA-TRIP flood model, Clim. Dyn., 38(7–8), 1389–1412, doi:10.1007/s00382-011-1054-9, 2012.

Döll, P.: Vulnerability to the impact of climate change on renewable groundwater resources: a global-scale assessment, Environ. Res. Lett., 4(3), 1–13, doi:10.1088/1748-9326/4/3/035006, 2009.

Döll, P. and Fiedler, K.: Global-scale modeling of groundwater recharge, Hydrol. Earth Syst. Sci., 12(3), 863–885, doi:10.5194/hess-12-863-2008, 2008.

Döll, P. and Lehner, B.: Validation of a new global 30-min drainage direction map, J. Hydrol., 258(1–4), 214–231, doi:10.1016/S0022-1694(01)00565-0, 2002.

Döll, P. and Zhang, J.: Impact of climate change on freshwater ecosystems: a global-scale analysis of ecologically relevant river flow alterations, Hydrol. Earth Syst. Sci., 14(5), 783–799, doi:10.5194/hess-14-783-2010, 2010.

Döll, P., Kaspar, F. and Lehner, B.: A global hydrological model for deriving water availability indicators: model tuning and validation, J. Hydrol., 270(1–2), 105–134, doi:10.1016/S0022-1694(02)00283-4, 2003.

Döll, P., Fiedler, K. and Zhang, J.: Global-scale analysis of river flow alterations due to water withdrawals and reservoirs, Hydrol. Earth Syst. Sci., 13(12), 2413–2432, doi:10.5194/hess-13-2413-2009, 2009.

Döll, P., Hoffmann-Dobrev, H., Portmann, F. T., Siebert, S., Eicker, A., Rodell, M., Strassberg, G. and Scanlon, B. R.: Impact of water withdrawals from groundwater and surface water on continental water storage variations, J. Geodyn., 59–60, 143–156, doi:10.1016/j.jog.2011.05.001, 2012.

Döll, P., Müller Schmied, H., Schuh, C., Portmann, F. T. and Eicker, A.: Global-scale assessment of groundwater depletion and related groundwater abstractions: Combining hydrological modeling with information from well observations and GRACE satellites, Water Resour. Res., 50(7), 5698–5720, doi:10.1002/2014WR015595, 2014a.

Döll, P., Fritsche, M., Eicker, A. and Müller Schmied, H.: Seasonal water storage variations as impacted by water abstractions: Comparing the output of a global hydrological model with GRACE and GPS observations, Surv. Geophys., 35(6), 1311–1331, doi:10.1007/s10712-014-9282-2, 2014b.

Döll, P., Douville, H., Güntner, A., Müller Schmied, H. and Wada, Y.: Modelling freshwater resources at the global scale: challenges and prospects, Surv. Geophys., 37(2), 195–221, doi:10.1007/s10712-015-9343-1, 2016.

Dorigo, W., de Jeu, R., Chung, D., Parinussa, R., Liu, Y., Wagner, W. and Fernández-Prieto, D.: Evaluating global trends (1988-2010) in harmonized multi-satellite surface soil moisture, Geophys. Res. Lett., 39(18), L18405, doi:10.1029/2012GL052988, 2012.

Dudgeon, D., Arthington, A. H., Gessner, M. O., Kawabata, Z.-I., Knowler, D. J., Lévêque, C., Naiman, R. J., Prieur-Richard, A.-H., Soto, D., Stiassny, M. L. J. and Sullivan, C. A.: Freshwater biodiversity: importance, threats, status and conservation challenges, Biol. Rev. Camb. Philos. Soc., 81(2), 163–82, doi:10.1017/S1464793105006950, 2006.

Durand, M., Gleason, C. J., Garambois, P. A., Bjerklie, D., Smith, L. C., Roux, H., Rodriguez, E., Bates, P. D., Pavelsky, T. M., Monnier, J., Chen, X., Di Baldassarre, G., Fiset, J.-M., Flipo, N., Frasson, R. P. M., Fulton, J., Goutal, N., Hossain, F., Humphries, E., Minear, J. T., Mukolwe, M. M., Neal, J. C., Ricci, S., Sanders, B. F., Schumann, G., Schubert, J. E. and Vilmin, L.: An intercomparison of remote sensing river discharge estimation algorithms from measurements of river height, width, and slope, Water Resour. Res., 51(2), 1333–1352, doi:10.1002/2015WR018434, 2016.

Easterling, D. R., Meehl, G. A., Parmesan, C., Changnon, S. A., Karl, T. R. and Mearns, L. O.: Climate extremes: Observations, modeling, and impacts, Science, 289(5487), 2068–2074, doi:10.1126/science.289.5487.2068, 2000.

Eicker, A., Schumacher, M., Kusche, J., Döll, P. and Müller Schmied, H.: Calibration/Data Assimilation approach for integrating GRACE data into the WaterGAP Global Hydrology Model (WGHM) using an Ensemble Kalman Filter: First results, Surv. Geophys., doi:10.1007/s10712-014-9309-8, 2014.

Eisner, S.: Comprehensive evaluation of the WaterGAP3 model across climatic, physiographic, and anthropogenic gradients, PhD thesis, Kassel University, 2015.

Evaristo, J., Jasechko, S. and McDonnell, J. J.: Global separation of plant transpiration from groundwater and streamflow, Nature, 525, 91–94, doi:10.1038/nature14983, 2015.

Famiglietti, J. S., Cazenave, A., Eicker, A., Reager, J. T., Rodell, M. and Velicogna, I.: Satellites provide the big picture, Science, 349(6249), 684–685, doi:10.1126/science.aac9238, 2015.

Fekete, B. M., Vörösmarty, C. J. and Grabs, W.: High-resolution fields of global runoff combining observed river discharge and simulated water balances, Global Biogeochem. Cycles, 16(3), 15-1-15–10, doi:10.1029/1999GB001254, 2002.

Fekete, B. M., Robarts, R. D., Kumagai, M., Nachtnebel, H.-P., Odada, E. and Zhulidov, A. V.: Time for in situ renaissance, Science, 349(6249), 685–686, doi:10.1126/science.aac7358, 2015.

Fernández, A. and Mark, B. G.: Modeling modern glacier response to climate changes along the Andes Cordillera: A multiscale review, J. Adv. Model. Earth Syst., 8(1), 467–495, doi:10.1002/2015MS000482, 2016.

Fiedler, K. and Döll, P.: Global modelling of continental water storage changes - Sensitivity to different climate data sets, Adv. Geosci., 11, 63–68, doi:10.5194/adgeo-11-63-2007, 2007.

Flörke, M., Kynast, E., Bärlund, I., Eisner, S., Wimmer, F. and Alcamo, J.: Domestic and industrial water uses of the past 60 years as a mirror of socio-economic development: A global simulation study, Glob. Environ. Chang., doi:10.1016/j.gloenvcha.2012.10.018, 2012.

Foley, J. A., Defries, R., Asner, G. P., Barford, C., Bonan, G., Carpenter, S. R., Chapin, F. S., Coe, M. T., Daily, G. C., Gibbs, H. K., Helkowski, J. H., Holloway, T., Howard, E. A., Kucharik, C. J., Monfreda, C., Patz, J. A., Prentice, I. C., Ramankutty, N. and Snyder, P. K.: Global consequences of land use, Science, 309(5734), 570–4, doi:10.1126/science.1111772, 2005.

Forman, B. A., Reichle, R. H. and Rodell, M.: Assimilation of terrestrial water storage from GRACE in a snow-dominated basin, Water Resour. Res., 48(1), W01507, doi:10.1029/2011WR011239, 2012.

Frieler, K., Betts, R., Burke, E., Ciais, P., Denvil, S., Deryng, D., Ebi, K., Eddy, T., Emanuel, K., Elliot, J., Galbraith, E., Gosling, S., Halladay, K., Hattermann, F., Hickler, T., Hinkel, J., Huber, V., Jones, C., Krysanova, V., Lange, S., Lotze, H., Lotze-Campen, H., Mengel, M., Mouratiadou, I., Müller Schmied, H., Ostberg, S., Piontek, F., Popp, A., Reyer, C. P. O., Schewe, J., Stevanovic, M., Suzuki, T., Thonicke, K., Tian, H., Tittensor, D. P., Vautard, R., van Vliet, M., Warszawski, L. and Zhao, F.: Assessing the impacts of 1.5°C global warming - simulation protocol of the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP2b), Geosci. Model Dev. Discuss., doi: 10.5194/gmd-2016-229, in review, 2016.

Gerten, D.: A vital link: water and vegetation in the Anthropocene, Hydrol. Earth Syst. Sci., 17(10), 3841–3852, doi:10.5194/hess-17-3841-2013, 2013.

Gleeson, T., Wada, Y., Bierkens, M. F. P. and van Beek, L. P. H.: Water balance of global aquifers revealed by groundwater footprint., Nature, 488(7410), 197–200, doi:10.1038/nature11295, 2012.

Gleeson, T., Befus, K. M., Jasechko, S., Luijendijk, E. and Cardenas, M. B.: The global volume and distribution of modern groundwater, Nat. Geosci., 9(2), 161–167, doi:10.1038/ngeo2590, 2015.

Gosling, S. N. and Arnell, N. W.: A global assessment of the impact of climate change on water scarcity, Clim. Change, 1–15, doi:10.1007/s10584-013-0853-x, 2013.

Gudmundsson, L. and Seneviratne, S. I.: Observation-based gridded runoff estimates for Europe (E-RUN version 1.1), Earth Syst. Sci. Data, 8(2), 279–295, doi:10.5194/essd-8-279-2016, 2016.

Güntner, A., Stuck, J., Werth, S., Döll, P., Verzano, K. and Merz, B.: A global analysis of temporal and spatial variations in continental water storage, Water Resour. Res., 43(5), W05416, doi:10.1029/2006WR005247, 2007.

Gupta, H. V., Kling, H., Yilmaz, K. K. and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, J. Hydrol., 377(1–2), 80–91, doi:10.1016/j.jhydrol.2009.08.003, 2009.

Haddeland, I., Clark, D. B., Franssen, W., Ludwig, F., Voß, F., Arnell, N. W., Bertrand, N., Best, M., Folwell, S., Gerten, D., Gomes, S., Gosling, S. N., Hagemann, S., Hanasaki, N., Harding, R., Heinke, J., Kabat, P., Koirala, S., Oki, T., Polcher, J., Stacke, T., Viterbo, P., Weedon, G. P. and Yeh, P.: Multimodel estimate of the global terrestrial water balance: Setup and first results, J. Hydrometeorol., 12(5), 869–884, doi:10.1175/2011JHM1324.1, 2011.

Haddeland, I., Heinke, J., Voß, F., Eisner, S., Chen, C., Hagemann, S. and Ludwig, F.: Effects of climate model radiation, humidity and wind estimates on hydrological simulations, Hydrol. Earth Syst. Sci., 16(2), 305–318, doi:10.5194/hess-16-305-2012, 2012.

Hanasaki, N., Kanae, S. and Oki, T.: A reservoir operation scheme for global river routing models, J. Hydrol., 327(1–2), 22–41, doi:10.1016/j.jhydrol.2005.11.011, 2006.

Hannah, D. M., Demuth, S., van Lanen, H. A. J., Looser, U., Prudhomme, C., Rees, G., Stahl, K. and Tallaksen, L. M.: Large-scale river flow archives: Importance, current status and future needs, Hydrol. Process., 25(7), 1191–1200, doi:10.1002/hyp.7794, 2011.

Heathwaite, A. L.: Multiple stressors on water availability at global to catchment scales: understanding human impact on nutrient cycles to protect water quality and water availability in the long term, Freshw. Biol., 55(Suppl. 1), 241–257, doi:10.1111/j.1365-2427.2009.02368.x, 2010.

Hertel, T. W.: The challenges of sustainably feeding a growing planet, Food Secur., 185–198, doi:10.1007/s12571-015-0440-2, 2015.

Hirabayashi, Y., Kanae, S., Motoya, K., Masuda, K. and Döll, P.: A 59-year (1948-2006) global near-surface meteorological data set for land surface models. Part I: Development of daily forcing and assessment of precipitation intensity, Hydrol. Res. Lett., 2(April), 36–40, doi:10.3178/hrl.2.36, 2008.

Hoekstra, A. Y., Chapagain, A. K., Aldaya, M. M. and Mekonnen, M. M.: The water footprint assessment manual. Setting the global standard, Earthscan, London, Washington, DC, 2011.

Hoekstra, A. Y., Mekonnen, M. M., Chapagain, A. K., Mathews, R. E. and Richter, B. D.: Global monthly water scarcity: blue water footprints versus blue water availability, PLoS One, 7(2), e32688, doi:10.1371/journal.pone.0032688, 2012.

Huang, J., Halpenny, J., van der Wal, W., Klatt, C., James, T. S. and Rivera, A.: Detectability of groundwater storage change within the Great Lakes Water Basin using GRACE, J. Geophys. Res. Solid Earth, 117(8), B08401, doi:10.1029/2011JB008876, 2012.

Hunger, M. and Döll, P.: Value of river discharge data for global-scale hydrological modeling, Hydrol. Earth Syst. Sci., 12(3), 841–861, doi:10.5194/hess-12-841-2008, 2008.

Jasechko, S.: Partitioning young and old groundwater with geochemical tracers, Chem. Geol., 427, 35–42, doi:10.1016/j.chemgeo.2016.02.012, 2016.

Jasechko, S. and Taylor, R. G.: Intensive rainfall recharges tropical groundwaters, Environ. Res. Lett., 10(12), 124015, doi:10.1088/1748-9326/10/12/124015, 2015.

Jasechko, S., Kirchner, J. W., Welker, J. M. and Mcdonnell, J. J.: Substantial proportion of global streamflow less than three months old, Nat. Geosci., 9(2), 126–129, doi:10.1038/ngeo2636, 2016.

Jovanovic, B., Jones, D. A. and Collins, D.: A high-quality monthly pan evaporation dataset for Australia, Clim. Change, 87(3–4), 517–535, doi:10.1007/s10584-007-9324-6, 2008.

Jung, M., Reichstein, M., Ciais, P., Seneviratne, S. I., Sheffield, J., Goulden, M. L., Bonan, G., Cescatti, A., Chen, J., de Jeu, R., Dolman, A. J., Eugster, W., Gerten, D., Gianelle, D., Gobron, N., Heinke, J., Kimball, J., Law, B. E., Montagnani, L., Mu, Q., Mueller, B., Oleson, K., Papale, D., Richardson, A. D., Roupsard, O., Running, S., Tomelleri, E., Viovy, N., Weber, U., Williams, C., Wood, E., Zaehle, S. and Zhang, K.: Recent decline in the global land evapotranspiration trend due to limited moisture supply, Nature, 467(7318), 951–954, doi:10.1038/nature09396, 2010.

Kaspar, F.: Entwicklung und Unsicherheitsanalyse eines globalen hydrologischen Modells, PhD thesis, University of Kassel, 2003.

Kelley, C. P., Mohtadi, S., Cane, M. A., Seager, R. and Kushnir, Y.: Climate change in the Fertile Crescent and implications of the recent Syrian drought, Proc. Natl. Acad. Sci., 112(11), 3241–3246, doi:10.1073/pnas.1421533112, 2015.

Khandu, Forootan, E., Schumacher, M., Awange, J. L. and Müller Schmied, H.: Exploring the influence of precipitation extremes and human water use on total water storage (TWS) changes in the Ganges-Brahmaputra-Meghna River Basin, Water Resour. Res., 52(3), 2240–2258, doi:10.1002/2015WR018113, 2016.

Kling, H., Fuchs, M. and Paulin, M.: Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios, J. Hydrol., 424–425, 264–277, doi:10.1016/j.jhydrol.2012.01.011, 2012.

Knieper, C. and Pahl-Wostl, C.: A comparative analysis of water governance, water management, and environmental performance in river basins, Water Resour. Manag., 30(7), 2161–2177, doi:10.1007/s11269-016-1276-z, 2016.

Kool, D., Agam, N., Lazarovitch, N., Heitman, J. L., Sauer, T. J. and Ben-Gal, A.: A review of approaches for evapotranspiration partitioning, Agric. For. Meteorol., 184, 56–70, doi:10.1016/j.agrformet.2013.09.003, 2014.

Korzun, V. I.: World water balance and water resources of the world, UNESCO Stud. Reports Hydrol., 25, 663, 1978.

Lehner, B., Liermann, C. R., Revenga, C., Vörösmarty, C., Fekete, B., Crouzet, P., Döll, P., Endejan, M., Frenken, K., Magome, J., Nilsson, C., Robertson, J. C., Rödel, R., Sindorf, N. and Wisser, D.: High-resolution mapping of the world's reservoirs and dams for sustainable river-flow management, Front. Ecol. Environ., 9(9), 494–502, doi:10.1890/100125, 2011.

Li, Z., Liu, W., Zhang, X. and Zheng, F.: Impacts of land use change and climate variability on hydrology in an agricultural catchment on the Loess Plateau of China, J. Hydrol., 377(1–2), 35–42, doi:10.1016/j.jhydrol.2009.08.007, 2009.

Liu, C., Zhang, X. and Zhang, Y.: Determination of daily evaporation and evapotranspiration of winter wheat and maize by large-scale weighing lysimeter and micro-lysimeter, Agric. For. Meteorol., 111(2), 109–120, doi:10.1016/S0168-1923(02)00015-1, 2002.

Lo, M.-H., Famiglietti, J. S., Yeh, P. J.-F. and Syed, T. H.: Improving parameter estimation and water table depth simulation in a land surface model using GRACE water storage and estimated base flow data, Water Resour. Res., 46(5), W05517, doi:10.1029/2009WR007855, 2010.

Lvovitch, M. J.: River runoff. Water regime types of rivers (of the world), in Physico-geographical atlas of the world, pp. 58–61, 1964.

Marcinek, J.: Der Abfluß von den Landflächen der Erde und seine Verteilung auf 5°-Zonen, Mitt. Inst. Wasserwirtschaft Berlin, 21, 202, 1964.

de Marsily, G. and Abarca-del-Rio, R.: Water and food in the twenty-first century, Surv. Geophys., 37(2), 503–527, doi:10.1007/s10712-015-9335-1, 2016.

Matsoukas, C., Benas, N., Hatzianastassiou, N., Pavlakis, K. G., Kanakidou, M. and Vardavas, I.: Potential evaporation trends over land between 1983-2008: Driven by radiative fluxes or vapour-pressure deficit?, Atmos. Chem. Phys., 11(15), 7601–7616, doi:10.5194/acp-11-7601-2011, 2011.

McMillan, H., Krueger, T. and Freer, J.: Benchmarking observational uncertainties for hydrology: rainfall, river discharge and water quality, Hydrol. Process., 26(26), 4078–4111, doi:10.1002/hyp.9384, 2012.

McVicar, T. R., Van Niel, T. G., Li, L., Hutchinson, M. F., Mu, X. and Liu, Z.: Spatially distributing monthly reference evapotranspiration and pan evaporation considering topographic influences, J. Hydrol., 338(3–4), 196–220, doi:10.1016/j.jhydrol.2007.02.018, 2007.

McVicar, T. R., Roderick, M. L., Donohue, R. J., Li, L. T., Van Niel, T. G., Thomas, A., Grieser, J., Jhajharia, D., Himri, Y., Mahowald, N. M., Mescherskaya, A. V., Kruger, A. C., Rehman, S. and Dinpashoh, Y.: Global review and synthesis of trends in observed terrestrial near-surface wind speeds: Implications for evaporation, J. Hydrol., 416–417, 182–205, doi:10.1016/j.jhydrol.2011.10.024, 2012.

Miralles, D. G., De Jeu, R. A. M., Gash, J. H., Holmes, T. R. H. and Dolman, A. J.: Magnitude and variability of land evaporation and its components at the global scale, Hydrol. Earth Syst. Sci., 15(3), 967–981, doi:10.5194/hess-15-967-2011, 2011.

Montanari, A. and Di Baldassarre, G.: Data errors and hydrological modelling: The role of model structure to propagate observation uncertainty, Adv. Water Resour., 51, 498–504, doi:10.1016/j.advwatres.2012.09.007, 2013.

Mu, Q., Zhao, M. and Running, S. W.: Brief introduction to MODIS evapotranspiration data set (MOD16), Water Resour. Res., (2005), 0–3, 2011a.

Mu, Q., Zhao, M. and Running, S. W.: Improvements to a MODIS global terrestrial evapotranspiration algorithm, Remote Sens. Environ., 115(8), 1781–1800, doi:10.1016/j.rse.2011.02.019, 2011b.

Mueller, B., Hirschi, M., Jimenez, C., Ciais, P., Dirmeyer, P. A., Dolman, A. J., Fisher, J. B., Jung, M., Ludwig, F., Maignan, F., Miralles, D. G., McCabe, M. F., Reichstein, M., Sheffield, J., Wang, K., Wood, E. F., Zhang, Y. and Seneviratne, S. I.: Benchmark products for land evapotranspiration: LandFlux-EVAL multi-data set synthesis, Hydrol. Earth Syst. Sci., 17(10), 3707–3720, doi:10.5194/hess-17-3707-2013, 2013.

Müller Schmied, H., Eisner, S., Franz, D., Wattenbach, M., Portmann, F. T., Flörke, M. and Döll, P.: Sensitivity of simulated global-scale freshwater fluxes and storages to input data, hydrological model structure, human water use and calibration, Hydrol. Earth Syst. Sci., 18(9), 3511–3538, doi:10.5194/hess-18-3511-2014, 2014.

Müller Schmied, H., Müller, R., Sanchez-Lorenzo, A., Ahrens, B. and Wild, M.: Evaluation of radiation components in a global freshwater model with station-based observations, Water, 8(10), 450, doi:10.3390/w8100450, 2016a.

Müller Schmied, H., Adam, L., Eisner, S., Fink, G., Flörke, M., Kim, H., Oki, T., Portmann, F. T., Reinecke, R., Riedel, C., Song, Q., Zhang, J. and Döll, P.: Impact of climate forcing uncertainty and human water use on global and continental water balance components, Proc. Int. Assoc. Hydrol. Sci., 374(7), 53–62, doi:10.5194/piahs-374-53-2016, 2016b.

Müller Schmied, H., Adam, L., Eisner, S., Fink, G., Flörke, M., Kim, H., Oki, T., Portmann, F. T., Reinecke, R., Riedel, C., Song, Q., Zhang, J. and Döll, P.: Variations of global and continental water balance components as impacted by climate forcing uncertainty and human water use, Hydrol. Earth Syst. Sci., 20(7), 2877–2898, doi:10.5194/hess-20-2877-2016, 2016c.

Nagel, A.: Entwicklung einer GIS-basierten semi-automatisierten Methode zur Detektion von Hohlformen in ariden und semi-ariden Regionen der Erde, MSc thesis, Goethe-University Frankfurt, 2012.

Nash, J. and Sutcliffe, J.: River flow forecasting through conceptual models part I — A discussion of principles, J. Hydrol., 10(3), 282–290, doi:10.1016/0022-1694(70)90255-6, 1970.

Neeck, S. P., Kakar, R. K., Durning, J. F. and Hou, A. Y.: Global precipitation measurement (GPM) progress, edited by R. Meynart, S. P. Neeck, and H. Shimoda, p. 59780F, doi:10.1117/12.627672, 2005.

Nijssen, B., O'Donnell, G. M., Lettenmaier, D. P., Lohmann, D. and Wood, E. F.: Predicting the discharge of global rivers, J. Clim., 14(15), 3307–3323, doi:10.1175/1520-0442(2001)014<3307:PTDOGR>2.0.CO;2, 2001.

Oki, T. and Kanae, S.: Global hydrological cycles and world water resources., Science, 313(5790), 1068–72, doi:10.1126/science.1128845, 2006.

Oki, T. and Sud, Y. C.: Design of Total Runoff Integrating Pathways (TRIP)—A global river channel network, Earth Interact., 2(1), 1–1, doi:10.1175/1087-3562(1998)002<0001:DoTRIP>2.0.CO;2, 1998.

Oudin, L., Hervieu, F., Michel, C., Perrin, C., Andréassian, V., Anctil, F. and Loumagne, C.: Which potential evapotranspiration input for a lumped rainfall–runoff model?, J. Hydrol., 303(1–4), 290–306, doi:10.1016/j.jhydrol.2004.08.026, 2005.

Pan, M., Sahoo, A. K., Troy, T. J., Vinukollu, R. K., Sheffield, J. and Wood, E. F.: Multisource estimation of long-term terrestrial water budget for major global river basins, J. Clim., 25(9), 3191–3206, doi:10.1175/JCLI-D-11-00300.1, 2012.

Piao, S., Friedlingstein, P., Ciais, P., de Noblet-Ducoudré, N., Labat, D. and Zaehle, S.: Changes in climate and land use have a larger direct impact than rising CO_2 on global river runoff trends., Proc. Natl. Acad. Sci. U. S. A., 104(39), 15242–7, doi:10.1073/pnas.0707213104, 2007.

Pokhrel, Y., Hanasaki, N., Koirala, S., Cho, J., Yeh, P. J.-F., Kim, H., Kanae, S. and Oki, T.: Incorporating Anthropogenic Water Regulation Modules into a Land Surface Model, J. Hydrometeorol., 13(1), 255–269, doi:10.1175/JHM-D-11-013.1, 2012.

Pokhrel, Y. N., Hanasaki, N., Wada, Y. and Kim, H.: Recent progresses in incorporating human land-water management into global land surface models toward their integration into Earth system models, Wiley Interdiscip. Rev. Water, doi:10.1002/wat2.1150, 2016.

Polcher, J., Bertrand, N., Biemans, H., Clark, D. B., Floerke, M., Gedney, N., Gerten, D., Stacke, T., van Vliet, M., Voss, F., Vilet, M. Van and Voss, F.: Improvements in hydrological processes in general hydrological models and land surface models within WATCH, WATCH Technical Report 34, 2011.

Ramillien, G. L., Frappart, F., Gratton, S. and Vasseur, X.: Sequential estimation of surface water mass changes from daily satellite gravimetry data, J. Geod., 89(3), 259–282, doi:10.1007/s00190-014-0772-2, 2015.

Rana, G. and Katerji, N.: Measurement and estimation of actual evapotranspiration in the field under Mediterranean climate: A review, Eur. J. Agron., 13(2–3), 125–153, doi:10.1016/S1161-0301(00)00070-8, 2000.

Rast, M., Johannessen, J. and Mauser, W.: Review of understanding of Earth's hydrological cycle: Observations, theory and modelling, Surv. Geophys., 35(3), 491–513, doi:10.1007/s10712-014-9279-x, 2014.

Refsgaard, J. C., van der Sluijs, J. P., Brown, J. and van der Keur, P.: A framework for dealing with uncertainty due to model structure error, Adv. Water Resour., 29(11), 1586–1597, doi:10.1016/j.advwatres.2005.11.013, 2006.

Richts, A., Struckmeier, W. F. and Zaepke, M.: WHYMAP and the Groundwater Resources Map of the World 1:25,000,000 BT - Sustaining groundwater resources: A critical element in the global water crisis, edited by A. J. A. Jones, pp. 159–173, Springer Netherlands, Dordrecht, doi:10.1007/978-90-481-3426-7_10, 2011.

Ričko, M., Birkett, C. M., Carton, J. A. and Crétaux, J.: Intercomparison and validation of continental water level products derived from satellite radar altimetry, J. Appl. Remote Sens., 6(1), 61710, doi:10.1117/1.JRS.6.061710, 2012.

Riedel, C. and Döll, P.: Global-scale modeling and quantification of indicators for assessing transboundary aquifers: a contribution to the GEF Transboundary Waters Assessment Programme (TWAP), Frankfurt Hydrology Paper 14, Institut für Physische Geographie, Frankfurt am Main, Germany, 2016.

Ringler, C., Bhaduri, A. and Lawford, R.: The nexus across water, energy, land and food (WELF): potential for improved resource use efficiency?, Curr. Opin. Environ. Sustain., 5(6), 617–624, doi:10.1016/j.cosust.2013.11.002, 2013.

Rodell, M., Beaudoing, H. K., L'Ecuyer, T. S., Olson, W. S., Famiglietti, J. S., Houser, P. R., Adler, R., Bosilovich, M. G., Clayson, C. A., Chambers, D., Clark, E., Fetzer, E. J., Gao, X., Gu, G., Hilburn, K., Huffman, G. J., Lettenmaier, D. P., Liu, W. T., Robertson, F. R., Schlosser, C. A., Sheffield, J. and Wood, E. F.: The observed state of the water cycle in the early twenty-first century, J. Clim., 28(21), 8289–8318, doi:10.1175/JCLI-D-14-00555.1, 2015.

Rosenzweig, C., Elliott, J., Deryng, D., Ruane, A. C., Müller, C., Arneth, A., Boote, K. J., Folberth, C., Glotter, M., Khabarov, N., Neumann, K., Piontek, F., Pugh, T. A. M., Schmid, E., Stehfest, E., Yang, H. and Jones, J. W.: Assessing agricultural risks of climate change in the 21st century in a global gridded crop model intercomparison, Proc. Natl. Acad. Sci. U. S. A., 111(9), 3268–3273, doi:10.1073/pnas.1222463110, 2014.

Ruhoff, A. L., Paz, A. R., Aragao, L. E. O. C., Mu, Q., Malhi, Y., Collischonn, W., Rocha, H. R. and Running, S. W.: Assessment of the MODIS global evapotranspiration algorithm using eddy covariance measurements and hydrological modelling in the Rio Grande basin, Hydrol. Sci. J., 58(8), 1658–1676, doi:10.1080/02626667.2013.837578, 2013.

Sanchez-Lorenzo, A., Vicente-Serrano, S., Wild, M., Calbó, J., Azorin-Molina, C. and Peñuelas, J.: Evaporation trends in Spain: a comparison of Class A pan and Piché atmometer measurements, Clim. Res., 61(3), 277–288, doi:10.3354/cr01255, 2014.

Schaller, N., Kay, A. L., Lamb, R., Massey, N. R., van Oldenborgh, G. J., Otto, F. E. L., Sparrow, S. N., Vautard, R., Yiou, P., Ashpole, I., Bowery, A., Crooks, S. M., Haustein, K., Huntingford, C., Ingram, W. J., Jones, R. G., Legg, T., Miller, J., Skeggs, J., Wallom, D., Weisheimer, A., Wilson, S., Stott, P. A. and Allen, M. R.: Human influence on climate in the 2014 southern England winter floods and their impacts, Nat. Clim. Chang., 6(6), 627–634, doi:10.1038/nclimate2927, 2016.

Schewe, J., Heinke, J., Gerten, D., Haddeland, I., Arnell, N. W., Clark, D. B., Dankers, R., Eisner, S., Fekete, B. M., Colón-González, F. J., Gosling, S. N., Kim, H., Liu, X., Masaki, Y., Portmann, F. T., Satoh, Y., Stacke, T., Tang, Q., Wada, Y., Wisser, D., Albrecht, T., Frieler, K., Piontek, F., Warszawski, L. and Kabat, P.: Multimodel assessment of water scarcity under climate change., Proc. Natl. Acad. Sci. U. S. A., 111(9), 3245–50, doi:10.1073/pnas.1222460110, 2014.

Schlesinger, W. H. and Jasechko, S.: Transpiration in the global water cycle, Agric. For. Meteorol., 189–190, 115–117, doi:10.1016/j.agrformet.2014.01.011, 2014.

Schneider, U., Becker, A., Finger, P., Meyer-Christoffer, A., Ziese, M. and Rudolf, B.: GPCC's new land surface precipitation climatology based on quality-controlled in situ data and its role in quantifying the global water cycle, Theor. Appl. Climatol., 115(1–2), 15–40, doi:10.1007/s00704-013-0860-x, 2014.

Schulte, P., van Geldern, R., Freitag, H., Karim, A., Négrel, P., Petelet-Giraud, E., Probst, A., Probst, J.-L., Telmer, K., Veizer, J. and Barth, J. A. C.: Applications of stable water and carbon isotopes in watershed research: Weathering, carbon cycling, and water balances, Earth-Science Rev., 109(1–2), 20–31, doi:10.1016/j.earscirev.2011.07.003, 2011.

Schulze, K. and Döll, P.: Neue Ansätze zur Modellierung von Schneeakkumulation und -schmelze im globalen Wassermodell WaterGAP, in Tagungsband zum 7. Workshop zur großskaligen Modellierung in der Hydrologie, edited by R. Ludwig, D. Reichert, and W. Mauser, pp. 145–154, Kassel University Press, Kassel, 2004.

Schulze, K., Hunger, M. and Döll, P.: Simulating river flow velocity on global scale, Adv. Geosci., 5(1), 133–136, doi:10.5194/adgeo-5-133-2005, 2005.

Schumacher, M., Kusche, J. and Döll, P.: A systematic impact assessment of GRACE error correlation on data assimilation in hydrological models, J. Geod., 90(6), 537–559, doi:10.1007/s00190-016-0892-y, 2016a.

Schumacher, M., Eicker, A., Kusche, J., Müller Schmied, H. and Döll, P.: Covariance analysis and sensitivity studies for GRACE assimilation into WGHM, in IAG 150 Years, Proceedings of the IAG Scientific Assembly in Potsdam, Germany, 2013, edited by C. Rizos and P. Willis, pp. 241–247, Springer International Publishing, doi:10.1007/1345_2015_119, 2016b.

Schuol, J., Abbaspour, K. C., Yang, H., Srinivasan, R. and Zehnder, A. J. B.: Modeling blue and green water availability in Africa, Water Resour. Res., 44(7), 1–18, doi:10.1029/2007WR006609, 2008.

Siebert, S., Kummu, M., Porkka, M., Döll, P., Ramankutty, N. and Scanlon, B. R.: A global data set of the extent of irrigated land from 1900 to 2005, Hydrol. Earth Syst. Sci., (19), 1521–1545, doi:10.13019/M20599, 2015.

Simmons, A., Fellous, J.-L., Ramaswamy, V., Trenberth, K., Asrar, G., Balmaseda, M., Burrows, J. P., Ciais, P., Drinkwater, M., Friedlingstein, P., Gobron, N., Guilyardi, E., Halpern, D., Heimann, M., Johannessen, J., Levelt, P. F., Lopez-Baeza, E., Penner, J., Scholes, R. and Shepherd, T.: Observation and integrated Earth-system science: A roadmap for 2016–2025, Adv. Sp. Res., 57(10), 2037–2103, doi:10.1016/j.asr.2016.03.008, 2016.

Smith, P., House, J. I., Bustamante, M., Sobocká, J., Harper, R., Pan, G., West, P. C., Clark, J. M., Adhya, T., Rumpel, C., Paustian, K., Kuikman, P., Cotrufo, M. F., Elliott, J. A., McDowell, R., Griffiths, R. I., Asakawa, S., Bondeau, A., Jain, A. K., Meersmans, J. and Pugh, T. A. M.: Global change pressures on soils from land use and management, Glob. Chang. Biol., 22(3), 1008–1028, doi:10.1111/gcb.13068, 2016.

Sood, A. and Smakhtin, V.: Global hydrological models: a review, Hydrol. Sci. J., 60(4), 549–565, doi:10.1080/02626667.2014.950580, 2015.

Tang, Q., Gao, H., Lu, H. and Lettenmaier, D. P.: Remote sensing: hydrology, Prog. Phys. Geogr., 33(4), 490–509, doi:10.1177/0309133309346650, 2009.

Tapley, B. D., Bettadpur, S., Ries, J. C., Thompson, P. F. and Watkins, M. M.: GRACE measurements of mass variability in the Earth system, Science, 305(5683), 503–5, doi:10.1126/science.1099192, 2004.

Verstraeten, W. W., Veroustraete, F. and Feyen, J.: Assessment of evapotranspiration and soil moisture content across different scales of observation, Sensors, 8(1), 70–117, doi:10.3390/s8010070, 2008.

Verzano, K., Bärlund, I., Flörke, M., Lehner, B., Kynast, E., Voß, F. and Alcamo, J.: Modeling variable river flow velocity on continental scale: Current situation and climate change impacts in Europe, J. Hydrol., 424–425, 238–251, doi:10.1016/j.jhydrol.2012.01.005, 2012.

Vinukollu, R. K., Meynadier, R., Sheffield, J. and Wood, E. F.: Multi-model, multi-sensor estimates of global evapotranspiration: climatology, uncertainties and trends, Hydrol. Process., 25(26), 3993–4010, doi:10.1002/hyp.8393, 2011.

Voisin, N., Wood, A. W. and Lettenmaier, D. P.: Evaluation of precipitation products for global hydrological prediction, J. Hydrometeorol., 9(3), 388–407, doi:10.1175/2007JHM938.1, 2008.

Vörösmarty, C. J., McIntyre, P. B., Gessner, M. O., Dudgeon, D., Prusevich, A., Green, P., Glidden, S., Bunn, S. E., Sullivan, C. A., Liermann, C. R. and Davies, P. M.: Global threats to human water security and river biodiversity, Nature, 467(7315), 555–561, doi:10.1038/nature09440, 2010.

Wada, Y., van Beek, L. P. H., van Kempen, C. M., Reckman, J. W. T. M., Vasak, S. and Bierkens, M. F. P.: Global depletion of groundwater resources, Geophys. Res. Lett., 37(20), L20402, doi:10.1029/2010GL044571, 2010.

Wada, Y., Wisser, D. and Bierkens, M. F. P.: Global modeling of withdrawal, allocation and consumptive use of surface water and groundwater resources, Earth Syst. Dyn., 5(1), 15–40, doi:10.5194/esd-5-15-2014, 2014.

Wada, Y., Lo, M., Yeh, P. J.-F., Reager, J. T., Famiglietti, J. S., Wu, R.-J. and Tseng, Y.-H.: Fate of water pumped from underground and contributions to sea-level rise, Nat. Clim. Chang., (May), 8–13, doi:10.1038/nclimate3001, 2016a.

Wada, Y., Flörke, M., Hanasaki, N., Eisner, S., Fischer, G., Tramberend, S., Satoh, Y., van Vliet, M. T. H., Yillia, P., Ringler, C., Burek, P. and Wiberg, D.: Modeling global water use for the 21st century: The Water Futures and Solutions (WFaS) initiative and its approaches, Geosci. Model Dev., 9(1), 175–222, doi:10.5194/gmd-9-175-2016, 2016b.

Wahr, J., Molenaar, M. and Bryan, F.: Time variability of the Earth's gravity field: Hydrological and oceanic effects and their possible detection using GRACE, J. Geophys. Res. Solid Earth, 103(B12), 30205–30229, doi:10.1029/98JB02844, 1998.

Weedon, G. P., Gomes, S., Viterbo, P., Shuttleworth, W. J., Blyth, E., Österle, H., Adam, J. C., Bellouin, N., Boucher, O. and Best, M.: Creation of the WATCH Forcing Data and its use to assess global and regional reference crop evaporation over land during the twentieth century, J. Hydrometeorol., 12(5), 823–848, doi:10.1175/2011JHM1369.1, 2011.

Weedon, G. P., Balsamo, G., Bellouin, N., Gomes, S., Best, M. J. and Viterbo, P.: The WFDEI meteorological forcing data set: WATCH Forcing Data methodology applied to ERA-Interim reanalysis data, Water Resour. Res., 50(9), 7505–7514, doi:10.1002/2014WR015638, 2014.

Weiß, M. and Menzel, L.: A global comparison of four potential evapotranspiration equations and their relevance to stream flow modelling in semi-arid environments, Adv. Geosci., 18, 15–23, doi:10.5194/adgeo-18-15-2008, 2008.

Werth, S. and Güntner, A.: Calibration analysis for water storage variability of the global hydrological model WGHM, Hydrol. Earth Syst. Sci., 14(1), 59–78, doi:10.5194/hess-14-59-2010, 2010.

Werth, S., Güntner, A., Petrovic, S. and Schmidt, R.: Integration of GRACE mass variations into a global hydrological model, Earth Planet. Sci. Lett., 277(1–2), 166–173, doi:10.1016/j.epsl.2008.10.021, 2009.

Willmott, C. J. and Johnson, M. L.: Resolution errors associated with gridded precipitation fields, Int. J. Climatol., 25(15), 1957–1963, doi:10.1002/joc.1235, 2005.

Wisser, D., Fekete, B. M., Vörösmarty, C. J. and Schumann, A. H.: Reconstructing 20th century global hydrography: a contribution to the Global Terrestrial Network- Hydrology (GTN-H), Hydrol. Earth Syst. Sci., 14(1), 1–24, doi:10.5194/hess-14-1-2010, 2010.

Wood, E. F., Roundy, J. K., Troy, T. J., van Beek, R. L. P. H., Bierkens, M. F. P., Blyth, E., de Roo, A., Döll, P., Ek, M., Famiglietti, J., Gochis, D., van de Giesen, N., Houser, P., Jaffé, P. R., Kollet, S., Lehner, B., Lettenmaier, D. P., Peters-Lidard, C., Sivapalan, M., Sheffield, J., Wade, A. and Whitehead, P.: Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth's terrestrial water, Water Resour. Res., 47(5), W05301, doi:10.1029/2010WR010090, 2011.

Yang, H. and Yang, D.: Climatic factors influencing changing pan evaporation across China from 1961 to 2001, J. Hydrol., 414–415, 184–193, doi:10.1016/j.jhydrol.2011.10.043, 2012.

Zarfl, C., Lumsdon, A. E., Berlekamp, J., Tydecks, L. and Tockner, K.: A global boom in hydropower dam construction, Aquat. Sci., 77(1), 161–170, doi:10.1007/s00027-014-0377-0, 2014.

Zhang, Y., Pan, M. and Wood, E. F.: On creating global gridded terrestrial water budget estimates from satellite remote sensing, Surv. Geophys., 37(2), 249–268, doi:10.1007/s10712-015-9354-y, 2016.

Acknowledgements

First of all, I would like to thank Petra Döll for her eyes-opening supervision, the intensive discussions over the past years and the family-friendly environment, all of which has made working with her and writing the thesis a great joy. I would also like to thank Andreas Güntner for preparing the second review and the collaboration. For the discussions about radiation and the link to the world of meteorology, I would like to thank Bodo Ahrens. I enjoy working in the Frankfurt hydrology group and thank all current and former members and student assistants for working together, exchanging ideas and sharing the workload. I also acknowledge the members of the Institute of Physical Geography for the nice collaboration in research and teaching – it is a pleasure to work in such a productive and friendly environment. In particular, I would like to acknowledge the technical support by our system administrators, Ralph Strobl and Hans-Peter Rulhof-Döll. Their quick reactions and solutions were tremendously helpful to meeting certain deadlines. I am grateful for the administrative support of Christiane Bergmann and Rita Peters. It is still amazing how just two people manage the whole institute. I have really enjoyed working together with Stephanie Eisner, our joint participation at the many conferences, and her great and goal-oriented ideas. Without her, I think the past years would only have been half the fun. I thank my colleague and flatmate Dirk Nowacki for sharing his flat during my visits in Frankfurt and the nice discussions apart from the university's cosmos. Stephanie Eisner, Maike Schumacher, Claudia Riedel and Katja Jusch Neumann gave very helpful comments on the thesis – thanks a lot! I greatly acknowledge Copernicus GmbH and especially Arne Richter, who introduced me to the "scientific circus" of the EGU meetings. I really like how my former classmate Martin Rasmussen and his team come up with new ideas every year for improving the meetings and also the publishing system. I sincerely thank my parents, Elisabeth and Martin for their great support over the past years of my studies and the PhD. Finally, and exceptionally, I would like to thank my partner, Jusch and my kids, Till and Lotta for filling my life with so much joy and reminding me what is really important.

Curriculum Vitae

Name:	Müller Schmied, Hannes
Address:	Juri-Gagarin-Straße 16, 07743 Jena, Germany
Date of birth:	19 May 1980
Place of birth:	Erfurt
Marital status:	In partnership
Children:	2 (2010, 2014)
E-mail:	hannes.mueller.schmied@em.uni-frankfurt.de

Education

1998	Staatliches Gymnasium Arnoldischule, Gotha (Abitur)
10/1999–9/2008	Study of geography with physical geography (Prof. Dr. R. Mäusbacher) and geoinformatics (Prof. Dr. W.-A. Flügel) as major subject and geosciences and ecology as minor subjects at Friedrich Schiller University Jena (degree: Diplom-Geograph)
Diploma thesis:	"Integrative Ableitung hydrologischer Funktionen von Feucht-gebieten am Beispiel des „Wipfragrund", oberes Gera-Einzugsgebiet, Thüringen" / Integrated assessment of hydrological functions of wetlands at the "Wipfragrund", upper Gera catchment, Thuringia (supervisors: Prof. Dr. W.-A. Flügel; Dr. J. Helmschrot)
10/2008	Degree as Diplom-Geograph with grade of 1.3 (1.0 is best grade)
10/2008–3/2009	PhD student at Friedrich Schiller University (Prof. Dr. W.-A. Flügel)
Since 5/2010	PhD student at FB11 (Geosciences/Geography) at Goethe University Frankfurt (Prof. Dr. P. Döll)

Work experience

9/1998–6/1999	Military service (air surveillance)
2002	Internship at Environmental Research Centre (UFZ) Leipzig
2000–2006	Tutor and student assistant at Department for Geoinformatics, Geohydroloy and Modelling, Friedrich Schiller University Jena
2001, 2004–2006	Student assistant at Max Planck Institute for Biogeochemistry Jena
2003–2004	Contract work for engineering consultancy Jena-GEOS GmbH
2005	Internship at remote sensing company Jena-Optronik GmbH
Since 2/2009	Research associate at Goethe University Frankfurt (Prof. Dr. P. Döll)
Since 2/2015	Research associate at Senckenberg Biodiversity and Climate Research Centre (BiK-F) (Prof. Dr. P. Döll)

Active memberships on university boards and in scientific societies

2000–2004	Member of student representative board of Geography, Friedrich Schiller University Jena
2003–2007	Student representative on faculty board of the Faculty of Chemistry and Geosciences, Friedrich Schiller University Jena
2005–2007	President of the Thuringian Geographical Society
Since 2015	Member of the working committee for young hydrologists of the German Hydrological Society (DHG e.V.)

Certificates, grants, awards

2008	Award for the best diploma thesis (Thuringian Geographical Society)
2012	"eLearning-certificate" (studiumdigitale of Goethe University Frankfurt)
2013	"certificate academic teaching" (Interdisziplinäres Kolleg Hochschuldidaktik (IKH) of Goethe University Frankfurt)
2013	Research grant (€5000) for developing e-learning modules
2015	Water Resources Research 2015 Editor's Choice Award
2017	Postdoctoral Fellow within Junior Fellowship project of Dr. Jamiat Nanteza, funded from VolkswagenStiftung

Miscellaneous

- 15 publications with peer review (published/accepted; status: 11 Mar 2017)
- 41 contributions at national and international meetings, thereof 17 as first author (status: 11 Mar 2017)
- 17 reviews for scientific journals (status: 11 Mar 2017; https://publons.com/a/471812)
- Co-supervision 11 bachelor's, master's and diploma theses at Goethe University Frankfurt
- Lecturer of 37 courses at various levels at Goethe University Frankfurt

List of publications

Journal papers

Müller Schmied, H., Adam, L., Eisner, S., Fink, G., Flörke, M., Kim, H., Oki, T., Portmann, F. T., Reinecke, R., Riedel, C., Song, Q., Zhang, J. and Döll, P.: Variations of global and continental water balance components as impacted by climate forcing uncertainty and human water use, Hydrol. Earth Syst. Sci., 20(7), 2877–2898, doi:10.5194/piahs-93-1-2016, 2016.

Müller Schmied, H., Eisner, S., Franz, D., Wattenbach, M., Portmann, F. T., Flörke, M. and Döll, P.: Sensitivity of simulated global-scale freshwater fluxes and storages to input data, hydrological model structure, human water use and calibration, Hydrol. Earth Syst. Sci., 18(9), 3511–3538, doi:10.5194/hess-18-3511-2014, 2014.

Müller Schmied, H., Müller, R., Sanchez-Lorenzo, A., Ahrens, B. and Wild, M.: Evaluation of radiation components in a global freshwater model with station-based observations, Water, 8(10), 450, doi:10.3390/w8100450, 2016.

Döll, P., Douville, H., Güntner, A., **Müller Schmied, H.** and Wada, Y.: Modelling Freshwater Resources at the Global Scale: Challenges and Prospects, Surv. Geophys., 37(2), 195–221, doi:10.1007/s10712-015-9343-1, 2016.

Döll, P., Fritsche, M., Eicker, A. and **Müller Schmied, H.**: Seasonal water storage variations as impacted by water abstractions: Comparing the output of a global hydrological model with GRACE and GPS observations, Surv. Geophys., 35(6), 1311–1331, doi:10.1007/s10712-014-9282-2, 2014.

Döll, P. and **Müller Schmied, H.**: How is the impact of climate change on river flow regimes related to the impact on mean annual runoff? A global-scale analysis, Environ. Res. Lett., 7(1), 14037, doi:10.1088/1748-9326/7/1/014037, 2012.

Döll, P., **Müller Schmied, H.**, Schuh, C., Portmann, F. T. and Eicker, A.: Global-scale assessment of groundwater depletion and related groundwater abstractions: Combining hydrological modeling with information from well observations and GRACE satellites, Water Resour. Res., 50(7), 5698–5720, doi:10.1002/2014WR015595, 2014.

Frieler, K., Betts, R., Burke, E., Ciais, P., Denvil, S., Deryng, D. Ebi, K., Eddy, T., Emanuel, K., Elliot, J., Galbraith, E., Gosling, S., Halladay, K., Hattermann, F., Hickler, T., Hinkel, J., Huber, V., Jones, C., Krysanova, V., Lange, S., Lotze, H., Lotze-Campen, H., Mengel, M., Mouratiadou, I., **Müller Schmied, H.**, Ostberg, S., Piontek, F., Popp, A., Reyer, D., Schewe, J., Stevanovic, M., Suzuki, T., Thonicke, K., Tian, H., Tittensor, D. T., Vautard, R., van Vliet, M., Warszawski, L, and Zang, F.: Assessing the impacts of 1.5°C global warming – simulation

protocol of the Inter-Sectoral Model Intercomparsion Project (ISMIP2b), Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-229, in review, 2016.

Khandu, Forootan, E., Schumacher, M., Awange, J. L. and **Müller Schmied, H.**: Exploring the influence of precipitation extremes and human water use on total water storage (TWS) changes in the Ganges-Brahmaputra-Meghna River Basin, Water Resour. Res., 52(3), 2240–2258, doi:10.1002/2015WR018113, 2016.

Masaki, Y., Hanasaki, H., Biemans, H., **Müller Schmied, H.**, Tang, Q., Wada, Y., Gosling, S., Takahashi, K. and Hijioka, Y.: Intercomparison of global river discharge simulations focusing on dam operation – Part II: Multiple models analysis in two case-study river basins: Missouri-Mississippi and Green-Colorado. Environ. Res. Lett., doi:10.1088/1748-9326/aa57a8, 2017 (in press).

Samaniego, L., Kumar, R., Thober, S., Rakovec, O., Zink, M., Wanders, N., Eisner, S., **Müller Schmied, H.**, Sutanudjaja, E. H., Warrach-Sagi, K. and Attinger, S.: Towards seamless hydrologic predictions across scales. Hydr. Earth Syst. Sci. Disc., doi:10.5194/hess-2017-89, in review, 2017.

Book chapters

Döll, P., Douville, H., Güntner, A., **Müller Schmied, H.** and Wada, Y.: Modelling freshwater resources at the global scale: Challenges and prospects, in Remote Sensing and Water Resources, Space Sciences Series of ISSI, vol. 55, edited by A. Cazenave, N. Champollion, J. Benveniste, and J. Chen, pp. 5–31, Springer International Publishing, Cham, doi:10.1007/978-3-319-32449-4, 2016.

Conference proceedings

Müller Schmied, H., Adam, L., Eisner, S., Fink, G., Flörke, M., Kim, H., Oki, T., Portmann, F. T., Reinecke, R., Riedel, C., Song, Q., Zhang, J. and Döll, P.: Impact of climate forcing uncertainty and human water use on global and continental water balance components, Proc. Int. Assoc. Hydrol. Sci., 374(7), 53–62, doi:10.5194/piahs-374-53-2016, 2016.

Müller Schmied, H., Helmschrot, J. and Flügel, W. A.: Hydrological functioning of a small wetland patch within a headwater environment in Thuringia, Germany, in Wetlands and Climate Change: New Challenges for Wetland Research, edited by U. Mander, pp. 70–72, Department of Geography, University of Tartu, Tartu, 2008.

Bäse, F., Helmschrot, J., **Müller Schmied, H.** and Flügel, W.: Die Auswirkungen von Landnutzungsänderungen auf die hydrologische Dynamik im semiariden Einzugsgebiet des Tsitsa, Südafrika, Zentralblatt für Geol. und Paläontologie, Tl. 1, (1–4), 251–267, 2006.

Bäse, F., Helmschrot, J., **Müller Schmied, H.** and Flügel, W. A.: The impact of land use change on the hydrological dynamic of the semi arid Tsitsa catchment in South Afrika, in Global Change Issues in Developing and Emerging Countries. Proceeidngs of the 2nd Göttingen GIS and Remote Sensing Days 2006, 4th to 6th October, Göttingen, Germany, edited by M. Kappas, C. Kleinn, and B. Sloboda, pp. 257–268, Universitätsverlag Göttingen, Göttingen, 2007.

Pechstädt, J., Zander, F., Bartosch, A., **Müller Schmied, H.** and Flügel, W. A.: Development of a River Basin Information System for a sustainable development in the Upper Brahmaputra River Basin, Sustain. Dev. - a Chall. Eur. Res. 26-28 May, Brussels, Belgium, 2009.

Schumacher, M., Eicker, A., Kusche, J., **Müller Schmied, H.** and Döll, P.: Covariance Analysis and Sensitivity Studies for GRACE Assimilation into WGHM, in IAG 150 Years, Proceedings of the IAG Scientific Assembly in Potsdam, Germany, 2013, edited by C. Rizos and P. Willis, pp. 241–247, Springer International Publishing, doi:10.1007/1345_2015_119, 2016.

Conference contributions (selection)

Müller Schmied, H. and Döll, P.: Assessment of the terrestrial water balance using the global water availability and use model WaterGAP – status and challenges. EGU General Assembly, 23.-28. April 2017, Vienna, Austria, 2017.

Müller Schmied, H., Adam, L., Eisner, S., Fink, G., Flörke, M., Kim, H., Oki, T., Portmann, F. T., Reinecke, R., Riedel, C., Song, Q., Zhang J. and Döll, P.: Variations of global and continental water balance components as impacted by climate forcing uncertainty and human water use. International conference on "Water Resources Assessment and Seasonal Predictions", 13.-16. October, Koblenz, Germany, 2015.

Müller Schmied, H., Möller, M. and Döll, P.: Evaluating the importance of climate input and calculation approach for estimating global scale potential evaporation. EGU General Assembly, 12.-17. April 2015, Vienna, Austria, 2015.

Müller Schmied, H., Adam, L., Döll, P., Eisner, S., Flörke, M., Güntner, A., Kynast, E., Portmann, F. T., Riedel, C., Schneider, C., Song, Q., Wattenbach, M. and Zhang, J.: Modelling global freshwater resources using WaterGAP 2.2 - model overview, selected results and applications. EGU General Assembly, 27. April - 2. May 2014, Vienna, Austria, 2014.

Müller Schmied, H.: Mehr Arbeit oder Mehrwert? Über Entwicklung und Einsatz von eLearning im frühen Geographiestudium. eLearning-Netzwerktag am 9.12.2013 an der Goethe Universität Frankfurt am Main, 2013.

Müller Schmied, H., Helmschrot, J. and Flügel, W.-A.: Analysis of flow duration curves and hydrochemistry for the assessment of wetland dynamics in Thuringian headwaters, Germany. 30th SWS-Conference, June 2008, Madison, Wisconsin, USA, 2009.

Helmschrot, J. and **Müller Schmied, H.**: Characterization and hydrological functioning of man-made wetland patches and shallow ponds in the Thuringian Forest, Germany. 27 th annual meeting of the Society of Wetland Scientists, Cairns, Australia, 2006. *(awarded with the „Honourable Mention Award" for best poster presentation)*

Shamsudduha, M., Taylor, R., Owor, M., Todd, M., Wada, Y., Lo, M.-H., **Müller Schmied, H.**, Döll, P., Rodell, M., Jasechko, S., Favreau, G., MacDonald, A. and Scanlon, B.: A pan-African inter-comparison of the relationship between precipitation and groundwater recharge from in-situ observations and large-scale models. 43rd IAH Congress, 25.-29. September 2016, Montpellier, France, 2016.

Scanlon, B., Zhang, Z., Sun, A., Save, H., **Müller Schmied, H.**, Wada, Y, Döll, P. and Eisner, S.: Comparison of total water storage anomalies from global hydrologic and land surface models and new GRACE satellite solutions. AGU Fall Meeting, 12.-16. December 2016, San Francisco, USA, 2016.

Schumacher, M., Forootan, E., van Dijk, A. I. J. M., **Müller Schmied, H.**, Crosbie, R. S., Kusche, J. and Döll, P.: GRACE assimilation into hydrological model improves representation of drought-induced groundwater trend over Murray-Darling Basin, Australia. EGU General Assembly, 18.-22. April 2016, Vienna, Austria, 2016.

Schumacher, M., Kusche, J., Eicker, A., Longuevergne, L., Franz, K., Döll, P. and **Müller Schmied, H.**: The influence of climate input uncertainties on the assimilation of GRACE data into the WaterGAP Global Hydrology Model. AGU Fall Meeting, 9.-30. December 2013, San Francisco, USA, 2013.

Sutanudjaja, E., Bosmans, J., Chaney, N., Clark, M., Condon, L., David, C., De Roo, A. P. J., Döll, P., Drost, N., Eisner, S., Famiglietti, J., Flörke, M., Gilbert, J., Gochis, D., Hut, R., Keune, J., Kollet, S., Maxwell, R., Pan, M., Rakovec, O., Reager, J., Samaniego, L., **Müller Schmied, H.**, Trautmann, T., van Beek, L., van de Giesen, N., Wood E. and Bierkens, M.: The HyperHydro (H^2) experiment for comparing different large-scale models at various resolutions. AGU fall meeting, 14.-18. December 2015, San Francisco, USA, 2015.