Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2015-527, 2016
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Published: 15 January 2016

Hydrology and Earth System Sciences
Open Access
Discussions

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

H. Müller Schmied[1,2], L. Adam[1], S. Eisner[3], G. Fink[3], M. Flörke[3], H. Kim[4], T. Oki[4], F. T. Portmann[1], R. Reinecke[1], C. Riedel[1], Q. Song[1], J. Zhang[1], and P. Döll[1]

[1]Institute of Physical Geography, Goethe-University Frankfurt, Frankfurt, Germany
[2]Biodiversity and Climate Research Centre (BiK-F) & Senckenberg Research Institute and Natural History Museum, 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

Received: 7 December 2015 – Accepted: 13 December 2015 – Published: 15 January 2016

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

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

Abstract

When assessing global water resources with hydrological models, it is essential to know the methodological uncertainties in the water resources estimates. The study presented here quantifies effects of the uncertainty in the spatial and temporal patterns of meteorological variables on water balance components at the global, continental and grid cell scale by forcing the global hydrological model WaterGAP 2.2 (ISI-MIP 2.1) with five state-of-the-art climate forcing input data-sets. While global precipitation over land during 1971–2000 varies between 103 500 and 111 000 $km^3\,yr^{-1}$, global river discharge varies between 39 200 and 42 200 $km^3\,yr^{-1}$. Temporal trends of global water balance components are strongly affected by the uncertainty in the climate forcing (except human water abstractions), and there is a need for temporal homogenization of climate forcings (in particular WFD/WFDEI). On about 10–20 % of the global land area, change of river discharge between two consecutive 30 year periods was driven more strongly by changes of human water use including dam construction than by changes in precipitation. This number increases towards the end of the 20th century due to intensified human water use and dam construction. The calibration approach of WaterGAP against observed long-term average river discharge reduces the impact of climate forcing uncertainty on estimated river discharge significantly. Different homgeneous climate forcings lead to a variation of Q of only 1.6 % for the 54 % of global land area that are constrained by discharge observations, while estimated renewable water resources in the remaining uncalibrated regions vary by 18.5 %. Uncertainties are especially high in Southeast Asia where Global Runoff Data Centre (GRDC) data availability is very sparse. By sharing already available discharge data, or installing new streamflow gauging stations in such regions, water balance uncertainties could be reduced which would lead to an improved assessment of the world's water resources.

HESSD

doi:10.5194/hess-2015-527

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 Global Runoff Data Centre (GRDC), http://grdc.bafg.de/), model-based or hybrid (i.e. incorporating historic discharge observations) approaches to estimate discharge and other water balance components are of increasing importance. The International Association of Hydrological Sciences (IAHS) decade on Predictions in Ungauged Basins (PUB) from 2003–2012 put a special focus on model application in data-sparse regions (Sivapalan et al., 2003). Since the 1980s, global hydrological models (GHMs) have been developed to calculate the water balance on global and/or continental scale. Recent reviews of developments and current state 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 (meteorological forecast models which assimilate all available up-to-date data for present time step) and/or remote sensing data (e.g. for radiation). Within the last two 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) and are available for historical time periods (from as early as 1901 until recent years). These climate forcings may differ among each other and thus may lead to different water resources estimates by GHMs.

HESSD
doi:10.5194/hess-2015-527

Humans have altered the global water cycle with an increasing intensity e.g. due to agricultural 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 leads to wrong assumptions for water resources assessment, in particular in regions with high water consumption relative to the renewable resources (e.g. High Plains Aquifer, Indus, Ganges-Brahmaputra). For example, groundwater depletion as observed in wells and by GRACE satellite observations of gravity variations can only be modelled 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 GHM ability to include 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), climate forcing and its underlying station measurement errors (Biemans et al., 2009) as well as uncertainties in discharge observations (Coxon et al., 2015; McMillan et al., 2012). In addition to these uncertainties, water resources estimates may differ due to different reference periods.

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

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

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")?

HESSD
doi:10.5194/hess-2015-527

To answer these questions, we used four (plus one modified) state-of-the-art climate forcings and the GHM WaterGAP in its version 2.2 (ISI-MIP 2.1). The model, data and methods are described in Sect. 2, followed by the results in Sect. 3. The research questions are answered in the discussion (Sect. 4). Finally, conclusions are drawn and an outlook is presented in Sect. 5.

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 (ISI-MIP) in its current phase 2.1 (https://www.pik-potsdam.de/research/climate-impacts-and-vulnerabilities/research/rd2-cross-cutting-activities/isi-mip/about/isi-mip2) and a fifth homogenized forcing.

2.1 GHM WaterGAP 2.2 (ISI-MIP 2.1)

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 water use sectors (irrigation, domestic, manufacturing, cooling water for electricity generation, and livestock) that are processed by the Ground Water Surface Water Use (GWSWUSE) submodule to establish net water abstraction from surface and groundwater resources (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

HESSD
doi:10.5194/hess-2015-527

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, named WaterGAP 2.2 (ISI-MIP 2.1), includes the following modifications of WaterGAP 2.2, mainly done to fulfil requirements of the ISI-MIP project, phase 2.1:

- 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 centre).
- 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. 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 neighbouring cell, while some other coastal cells disappeared), and 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.

HESSD
doi:10.5194/hess-2015-527

– For lakes, the reduction of evaporation due to decreasing lake area is calculated as 2 × maximum lake storage (S_{max}) in the denominator as in Hunger and Döll (2008, their Eq. 1), for a more realistic storage behaviour.

2.2 Calibration of WaterGAP 2.2 (ISI-MIP 2.1) against observed streamflow

The initial purpose of WaterGAP development was the assessment of (renewable) water resources as a prerequisite for human and ecosystem well-being. To obtain meaningful estimates of water resources despite different sources of uncertainty related to GHMs, a calibration scheme was applied (see Döll et al., 2003; Hunger and Döll, 2008; Müller Schmied et al., 2014). The calibration routine in WaterGAP 2.2 (ISI-MIP 2.1) forces the long-term annual simulated 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 observation being 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 WFD_WFDEI forcing. Observations stations were selected such that the upstream area and the interstation catchment area had a minimum size of 9000 and 30 000 km^2, respectively, and only if a minimum of four complete years of data was available.

HESSD
doi:10.5194/hess-2015-527

2.3 Climate forcing data sets

Within the ISI-MIP project phase 2.1, 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 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.

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) done with the NCEP atmosphere land model (Compo et al., 2011) which has relatively low spatial resolution ($\sim 2.0^{\circ}$) and long term availability (140 years) was dynamically downscaled into the global T248 ($\sim 0.5^{\circ}$) 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 and Climate Research Unit (CRU) TS3.21 observational data 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).

HESSD
doi:10.5194/hess-2015-527

2.3.2 Princeton Global Meteorological Forcing Dataset (PGFv2)

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 in its current form the period 1901–2012 (http://hydrology.princeton.edu/data.pgf.php). *P* is bias corrected to monthly CRU TS3.21 and is not undercatch corrected (different to the previous version). 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 University of Maryland (by Rachel Pinker) developed within 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, 2015.

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 (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 to 2001 and on the reordered ERA-40 data for the period 1901–1957. The variables from ERA-40 are interpolated (incorporating elevation) and some are corrected to monthly observation data, e.g. monthly bias-correction for *P* using GPCC version 4 (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).

HESSD
doi:10.5194/hess-2015-527

2.3.4 Combined WFD and WFDEI (WFD_WFDEI)

The WFDEI dataset 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 to 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. There is another P variable available until to 2012, but corrected to CRU TS3.1/TS3.21 (not used in this study). In WFDEI, SWD in northern Africa up to central Europe is much larger than in WFD 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 ISI-MIP 2.1 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), the actual evapotranspiration AET increased globally by $\sim 5000\,km^3\,yr^{-1}$, 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 similar to that of Haddeland et al. (2012) was applied to the daily data for three climatic variables. For the variables SWD and LWD, a multiplicative approach was applied (Eq. 1). T was homogenized with an additive approach due to possible zero values (Eq. 2). P was not homogenized as only marginal differences in continental and global sums occur (Table 4).

HESSD
doi:10.5194/hess-2015-527

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

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

with V_{hom} being the homogenized daily variable (1901–2001), V_{WFD} the original daily variable from WFD (1901–2001), $\overline{V_{WFDEI}}(m)$ and $\overline{V_{WFD}}(m)$ the long-term mean monthly variable from WFDEI and WFD for the overlapping time period 1979-2001, applied to the current month. The final homogenized daily WFDEI_hom time series consists of homogenized WFD data until 1979 and of WFDEI data afterwards.

2.4 Calculation of global averages and indicators

2.4.1 Global climate and water balance components

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 represents inland sinks. For T, SWD and LWD, area-weighted averages were calculated. Global Q was calculated by summing up the Q of all grid cells that are outflow cells into the ocean according to the drainage direction map DDM30 (Döll and Lehner, 2002) and the Q into all grid cells that represent inland sinks.

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 period is either more likely due to the change of P in the upstream river basin or more likely due to the change of the anthropogenic impact on Q by human water use and dam construction, a number of

HESSD
doi:10.5194/hess-2015-527

indicators were developed and combined. In the equations below, Q represents simulated discharge under anthropogenic conditions, whereas Q_{nat} is the discharge that would occur without human water use as well as reservoirs or regulated lake regulation by dams.

First, we assume that P change cannot be the more dominant driver than the change of anthropogenic impacts if P increases while Q decreases (and vice versa). For each grid cell, the indicator A_n is computed as

$$A_n = \frac{P_{\mathrm{bas}(n),t_2} - P_{\mathrm{bas}(n),t_1}}{Q_{n,t_2} - Q_{n,t_1}} \quad (3)$$

where A_n [–] is the indicator for consistency between changes in cumulative upstream precipitation of grid cell n ($P_{\mathrm{bas}(n)}$, [$\mathrm{km^3\,yr^{-1}}$]) and simulated river discharge of the grid cell (Q_n, [$\mathrm{km^3\,yr^{-1}}$]) between the time periods t_1 (e.g. 1941–1970) and t_2 (e.g. 1971–2000). If changes of P and Q have the same sign, A_n is positive, and the P change 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 significant driver of the change in Q.

Similar to A_n, indicator B_n serves to check whether the change of the anthropogenic impact on river discharge, expressed as the difference between Q and Q_{nat}, is consistent with the change of Q. An increasing difference should lead to a decrease of Q.

$$B_n = \frac{Q_{n,t_2} - Q_{n,t_1}}{\left(Q_{\mathrm{nat}_{n,t_1}} - Q_{n,t_1}\right) - \left(Q_{\mathrm{nat}_{n,t_2}} - Q_{n,t_2}\right)} \quad (4)$$

where B_n [–] is the indicator for consistency between changes in Q and changes in the anthropogenic impact on river discharge. If e.g. Q increases between the two time periods but the difference between Q_{nat} and Q increases e.g. due to increased human water use among the time periods, B_n gets negative, indicating that anthropogenic effects cannot be a significant driver of change in Q.

HESSD
doi:10.5194/hess-2015-527

In case of both A_n and B_n are equal to or greater than zero, I_{varprec}, I_{varant} and $I_{\text{varprecdom}}$ are calculated as

$$I_{\text{varprec},n} = \left| \left(Q_{n,t_2} - Q_{n,t_1} \right) - C_{QP,n} \left(P_{\text{bas}(n),t_2} - P_{\text{bas}(n),t_1} \right) \right| \quad (5)$$

where I_{varprec} [km^3 yr^{-1}] is an absolute indicator for the amount that the change in Q results from the change in P, translated into Q via the runoff coefficient $C_{QP,n}$ [–], calculated as

$$C_{QP,n} = 0.5 \left(\frac{Q_{\text{nat}_{n,t_1}}}{P_{\text{bas}(n),t_1}} + \frac{Q_{\text{nat}_{n,t_2}}}{P_{\text{bas}(n),t_2}} \right) \quad (6)$$

and assuming that the runoff coefficient remains constant over the two time periods.

The smaller I_{varprec}, the more likely that the change in P is a strong driver of the change in Q. The indicator for anthropogenic effect I_{varant} [km^3 yr^{-1}] is calculated as

$$I_{\text{varant},n} = \left| \left(Q_{n,t_2} - Q_{n,t_1} \right) - \left(\left(Q_{\text{nat}_{n,t_1}} - Q_{n,t_1} \right) - \left(Q_{\text{nat}_{n,t_2}} - Q_{n,t_2} \right) \right) \right|. \quad (7)$$

The smaller I_{varant}, the more likely that the change in the anthropogenic impact is a strong driver of the change in Q. Finally, the indicator of relative dominance $I_{\text{varprecdom}}$ [km^3 yr^{-1}] is calculated as

$$I_{\text{varprecdom},n} = I_{\text{varant},n} - I_{\text{varprec},n}. \quad (8)$$

The change in P is a more dominant driver than the change in anthropogenic impact if $A_n > 0$ and $B_n < 0$, or if $I_{\text{varprecdom},n} > 0$. The change in anthropogenic impact is a more dominant driver than the change in P, if $A_n < 0$ and $B_n > 0$, or if $I_{\text{varprecdom},n} < 0$. If both $A_n < 0$ and $B_n < 0$, changes in Q are not consistent with either changes due to P or due to anthropogenic effects, and are caused by changes of other drivers, e.g. T.

HESSD
doi:10.5194/hess-2015-527

3 Results

3.1 Variation of estimated global water balance components for different temporal aggregations

Figure 2 shows the importance of temporal aggregation and reference periods for the assessment of global scale climatic variables and water balance components. Even for globally aggregated components, there are strong year-to-year fluctuations. Decadal (10 years) aggregations show either fluctuation around the 30 year aggregations or, in case of variables with a significant temporal trend (T and actual water consumption WCa), a decade appears to be an appropriate aggregation period to clearly show the trend.

Increase of global averages of T during the last three decades is comparable among the five climate forcings but PGFv2 shows higher T in the first two decades of the time series which affects long-term trends. Large differences occur for SWD, for which WFD forcing shows an offset of around $-15\,\mathrm{W\,m^{-2}}$. This also affects the combined standard WFD_WFDEI, resulting in an implausible inhomogeneity 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. The variable LWD shows different variations among the climate forcings (e.g. between GSWP3 and PGFv2), 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. Using land surface parameters, WaterGAP calculates the outgoing components of radiation and subsequently net radiation which is then used to calculate potential evapotranspiration. The temporal variation of net radiation (not shown), is relatively low, except for the WFD_WFDEI combination, and WFD net radiation is comparatively low in absolute numbers. Generally, P is increasing over time, visible in the 30 year climate normal, but varying in smaller time aggrega-

HESSD
doi:10.5194/hess-2015-527

tions (e.g. decadal variation is not negligible). For century means (1901–2000), forcings that are bias corrected to GPCC data (all except PGFv2) show nearly the same value. The century mean for PGFv2 is around 7000 $km^3 yr^{-1}$ smaller due to the not used undercatch adjustment in this version. The trend of Q does not necessarily follow P (e.g. GSWP3), while AET does (except WFD_WFDEI where the AET pattern resembles the inverted Q pattern).

For Q (AET), the inhomogeneity in WFD_WFDEI leads to an implausible decrease (increase) of around 5000 $km^3 yr^{-1}$. This error can be reduced by homogenization to around 2000 $km^3 yr^{-1}$ (WFDEI_hom). The trend of Q varies among the forcings (increase in PGFv2, constant in GSWP3, decrease in the others). The decrease of Q from 1981 onwards which is visible in the WFD_WFDEI and WFDEI_hom forcings, is likely related to differences of the ERA-Interim reanalysis vs. the NCEP (GSWP3, PGFv2) and ERA-40 (WFD) forcings. As neither P (more or less constant) nor radiation (which starts to have an offset already in the previous decade) is not related to the decrease in Q, regional variation of climate forcing variables are likely to be responsible. Century means of Q differ remarkably (up to 5500 $km^3 yr^{-1}$). WCa is continuously increasing in all model variants with a more pronounced trend since the 1950s. The separation of water uses into the different sectors as well as into water withdrawals and consumptive use are shown by Müller Schmied et al. (2015), their Fig. 3.

3.2 Global water balance components (1971–2000)

For the global land area (excluding Antarctica, Greenland and grid cells representing inland sinks) and the five model variants for the time period 1971–2000, totals of water resources are shown in Table 1, with a separation by calibration areas and non-calibration areas in Table 2. Row 3 in Tables 1 and 2 show Q into oceans and inland sinks, and runoff produced in all grid cells within the calibration and non-calibration basins (runoff computed as grid cell-specific difference between Q flowing into and out of the cell), respectively.

HESSD
doi:10.5194/hess-2015-527

The 1971–2000 averaged global P differs among the model forcing variants, with the largest difference found between CRU based product (PGFv2) and the GPCC based products (all other forcings) amounting up to 7000 $km^3\,yr^{-1}$ (Table 1). Even the GPCC-based forcings vary by up to 1400 $km^3\,yr^{-1}$ (exceeding the amount of WCa), obviously due to the different correction of measurement errors, such as snowfall undercatch. While the calibrated grid cells cover 54 % of global land area (excluding Antarctica and Greenland), they receive 61.3 % of P (Table 2). Müller Schmied et al. (2015) shows the proportions of AET, Q, WCa and the change of total water storage of the five climate forcings in their Table 2. The difference among the forcing variants (calculated as (maximum P – minimum P)/mean P) is with 7.5 % in calibrated basins slightly higher than in non-calibrated basins (6.1 %).

Global Q varies among the five forcings by around 3000 $km^3\,yr^{-1}$ (Table 1). As expected, the difference among the forcing variants (calculated as (maximum Q – minimum Q)/mean Q) is with 18.4 % higher in non-calibrated basins than in calibrated basins, where it is only 2.8 %. For the homogeneous forcings (less WFD_WFDEI), Q differs in calibration regions with 18.5 % and in non-calibrated regions by 1.6 %. That there is any discharge variation in calibrated basins is due various reasons. Calibration forces the simulated mean annual discharges to be equal (within 1 %) to the observed ones for the calibration period. Outside the calibration period, the different forcings cause the computed Q to vary. In the homogeneous forcings that are undercatch adjusted by using GPCC data (GSWP3, WFD, WFDEI_hom), Q differs by only 0.1 % in calibrated regions and 10.5 % in non-calibrated regions. This highlights the agreement of spatio-temporal P distribution at this scale and the effect of calibration for these forcings, and the low influence of the smaller number of years of available climate data for calibrating WFD. For the forcings GSWP3 and PGFv2, the same discharge data were used for calibration (if available, 30 years of data between 1920 and 2009, preferable between 1971–2000). To avoid the offset problematic in the WFD_WFDEI forcing, WFDEI_hom was calibrated with a preferable time period of 1980–2009 and the calibration parameters were transferred to WFD_WFDEI. As WFD

HESSD
doi:10.5194/hess-2015-527

ends in 2001, only station data until 2001 could be used (same preference time). For this reason, seven out of the 1319 calibration stations were excluded. Nevertheless, Q for calibration regions and WFD forcing is within the range of the other forcings that are driven by GPCC data (WFDEI_hom, GSWP3) which is a result of higher calibration parameter values for WFD forcing.

Around 55.2 % of Q and 65.0 % of AET occur in calibrated regions (Table 2). Main differences of Q among the four homogeneous forcings in non-calibrated regions occur in South East Asia (see Sect. 4.3, Fig. 5b and d). Anthropogenic water use is (in total numbers) around 100 $km^3\,yr^{-1}$ higher in calibrated areas compared to non-calibrated regions. The WFD forcing leads to much higher Q in the non-calibrated area (Table 2) than the other forcings, consistent with the relatively low incoming shortwave radiation which results in a comparable low AET. The deviation of WCa among all forcing variants are with 5.9 % in calibration basins higher than in non-calibration basins (4.4 %) (Table 2).

As Q is forced to be nearly equal within the calibration region, different amount of P is reflected in large differences of AET, in particular in calibration regions (with higher absolute values than P differs, or 12.2 %). In contrast, AET differs by 8.8 % (and lower absolute amount than P) in non-calibration regions (both numbers for all forcings). PGFv2 has the lowest global AET but the highest WCa of all five forcings (Table 1), 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.

3.3 Continental water balance components

Table 3 presents the continental values of P, AET, Q, and WCa as computed by the four homogeneous forcing variants (GSWP3, PGFv2, WFD, WFDEI_hom). Both the ensemble mean and the percent deviation among the forcing variants (min and max in % of mean) are shown. Deviations in P are mainly due to PGFv2 that uses CRU TS3.21 observation-based data for bias correction and no undercatch correction,

HESSD
doi:10.5194/hess-2015-527

while the other forcings use GPCC data and two types of undercatch correction. For P, Oceania (with the lowest absolute value) has the lowest deviation among the forcings. The largest deviations are found in North America, Europe, and Africa. In North America and Europe, where the station density is reasonable high and GPCC versions agree very well (Table 4), the different approaches to undercatch correction in the forcings (after Adam and Lettenmaier (2003) for WFD and WFDEI and using gauge type corrections (Hirabayashi et al., 2008) for GSWP3), or neglecting undercatch adjustment (PGFv2) leads to the large deviations of WaterGAP outputs. The AET deviation follows the P deviation. For Q, deviation among forcings is highest in Africa. Here, some areas with high amounts of P (and Q) are in non-calibration regions (e.g. Madagascar, see Fig. 5). Besides, the runoff coefficient (Q/P) of Africa is with 0.21 the lowest of all continents; for the other continents it varies 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). The largest amount of water resources is calculated in South America followed by Asia. The highest WCa value and the lowest relative deviation can be found in Asia; the low deviation might be explained by averaging out differences in grid cell values due to the climate forcings over the large number of grid cells in Asia with irrigation water use. Overall, highest relative global uncertainty due to climate forcing was identified for AET.

3.4 Drivers of change in 30 year average river discharge

Figure 3 shows where the change of long-term average Q between the time period 1941–1970 and the time period 1971–2000 is either more likely due to the change of P in the upstream river basin (blue colours) or more likely due to the change of the anthropogenic impact on Q by human water use and dam construction (red colours, see Sect. 2.4.2). Results for WaterGAP as driven by each of the four homogeneous climate forcings GSWP3, PGFv2, WFD and WFDEI_hom are shown. In most regions, changes in P is the more important driver of changes in Q than changes of the anthropogenic

HESSD
doi:10.5194/hess-2015-527

impact. On the other hand, in areas with remarkable water consumption, human water use or/and the construction of dams lead to a relative dominance of such effects (compared to P). Note that the developed indicators only compare the relevance of two drivers of change. In the blue and red grid cells, other drivers such as T or radiation could be even stronger drivers of the simulated change in Q. However, in those regions where indicators A_n and B_n are both negative (green colours), other drivers (and not P or anthropogenic effects) are the main reason for changes in Q.

Changes in long-term average Q between the time periods 1911–1940 and 1941–1970 are, in most world regions, less dominated by changes in the anthropogenic impact on river discharge (Fig. 4). This is consistent with the accelerating human water use and dam construction throughout the 20th century (Fig. 2). In the earlier analysis period, most occurrences of anthropogenic domination of changes in Q are restricted to 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. 3 and 4).

The fraction of land area among the forcings where Q changes are dominantly due to human water use and dam construction is 10–13 % for the period 1911–1940/1941–1970 and increases to 13–20 % in the later period. P domination increases also (from 53–54 to 58–65 %) but at the same time, areas which cannot be calculated decrease from ~ 30 to ~ 20 %. Fractions where both, P and human water use is not the dominant driver, are in both time periods around 4 %.

4 Discussion

4.1 Comparison of simulated water balance components to previous estimates

The simulated global water balance components AET and Q are within the range of estimates reported in the literature (see values in Müller Schmied et al. (2014), their Table 5). Values for AET of this study (63 500–70 000 $km^3\,yr^{-1}$) are within this

HESSD
doi:10.5194/hess-2015-527

range. Global values for Q (39 200–42 200 $km^3 yr^{-1}$) are at the upper level of values from literature. Compared to the STANDARD WaterGAP2.2 model variant driven by WFD_WFDEI (Müller Schmied et al. (2014), their Table 2, hereafter referred to as STANDARD), global P equals to the WFD_WFDEI forcing used here (Table 1). The data basis for P is the same, and the additional grid cells of the WATCH-CRU land-ocean mask (global, and including Greenland, with 134.6 million km^2, comp. to 134.5 million km^2 for STANDARD) with many additional (tropical) and mostly small islands as well as coastal cells used here, does not influence global P remarkably. WCa is with 936 $km^3 yr^{-1}$ lower in this study (compared to 1031 $km^3 yr^{-1}$ in STANDARD) due to the assumed deficit irrigation (Sect. 2.1). This also explains the smaller depletion of total water storages in this study (−215 $km^3 yr^{-1}$ in STANDARD, −74 $km^3 yr^{-1}$ in this study). Groundwater storage change of STANDARD was −125 $km^3 yr^{-1}$ compared to −75 $km^3 yr^{-1}$ with WFD_WFDEI in this study. In addition, due to filling up storages when operation year for individual reservoirs are reached in the model version used here, change in reservoir storage is positive (53 $km^3 yr^{-1}$) compared to −43 $km^3 yr^{-1}$ in STANDARD, where all reservoirs were assumed to be in operation over the entire simulation period. Global AET and Q are comparable and differ only by around 1 % between both studies.

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

To assess the importance of the choice of time period and temporal aggregation of particular water balance components or climatic variables, value ranges (max–min) and percentual change of this range were computed. The smaller the range, the less important is the choice. For the water balance components AET, Q and P, the range of values is reduced by 55–70 % when moving from 1 to 10 year; by 40–54 % when moving from 10 to 30 year and by 73–83 % when moving from 1 to 30 year temporal aggregation scale. The highest range for annual aggregation occurs for Q and WFD_WFDEI

HESSD
doi:10.5194/hess-2015-527

(12 500 $km^3 yr^{-1}$), which is higher than P and AET (both 11 000 $km^3 yr^{-1}$). For 30 year aggregates of AET and Q, WFD has the lowest and WFD_WFDEI the highest range. Homogenization of WFD and WFDEI (comparing WFD_WFDEI and WFDEI_hom) reduces the range in percent for 1, 10 and 30 years and AET (Q) by 35, 54, 55 (31, 62, 64), which is then close to the ranges of the other forcing variants. Due to the strong trend in WCa, differences in range between 1 and 10 year are comparable among the model variants (5–15 % difference) but higher when going from 10 to 30 years (25–38 %).

The T ranges are 1.5 °C (1 year), 1.1 °C (10 years) and 0.4 °C (30 years). The ranges of the radiation variables SWD, LWD, and net radiation are comparable among all forcings (reduction of 44–49 % when going from 1 to 10 years, 46–55 % from 10 to 30 years and 69–77 % from 1 to 30 years).

This analysis supports, that when assessing water balance with small trends (e.g. P, Q, AET), temporal aggregation of 30 years are meaningful for comparisons among models and/or climate forcings. For variables with strong trends (e.g. WCa, T), a reasonable temporal aggregation is the 10 year period, as this is a trade-off between keeping fluctuations but not balancing trends out.

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

If precipitation and radiation data differ among state-of-the-art climate forcings as strong as in the forcings used here, the effect of choosing a climate forcing can be more important than the selection of the time aggregation. Uncertainty in global-scale P estimates is caused by different reanalysis products used, different observation-based bias correction methods/data-sets, and different approaches for undercatch adjustment. Unexpectedly, the largest differences among the forcings occur in data-rich continents (Europe, North America). For global Q, differences of up to 3000 $km^3 yr^{-1}$ can occur (1971–2000). Most climate forcing-induced variation in continental Q occurs in Africa.

HESSD
doi:10.5194/hess-2015-527

Calibration reduces Q variations significantly at continental and global scales. AET has large uncertainties/variability both at global and continental scale (esp. in Europe and North America). As AET is not directly calibrated (but water balance is closed and thus influenced by calibration), this is related to the differences in climatic variables. Differences in WCa are, in absolute numbers, comparably low at global scale but differ largely among the continents. 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. Again, Europe and North America have high uncertainties in continental assessments due to climate forcing uncertainty/variability.

The low global values for SWD in WFD forcing leads to less AET and higher Q (2000 $km^3\,yr^{-1}$) compared to the homogenized forcing WFDEI_hom. As the calibration approach of WaterGAP forces the simulated long term averaged annual Q to match the observed one, higher calibration parameters (esp. area and station correction factors, see Müller Schmied et al. (2014) for calibration details) were needed with WFD forcing. For non-calibrated regions (Table 2) as well for GHMs which do not use a calibration scheme, water resources are likely to be overestimated when using WFD forcing. The low SWD in WFD can also be interpreted as one reason of the high absolute values of Q in Haddeland et al. (2011) (42 000–66 000 $km^3\,yr^{-1}$) compared to previous estimates (e.g. Fekete, 2002) or values from this study.

Figure 5 shows the range of Q at grid cell level in $km^3\,yr^{-1}$ for calibration and non-calibration regions. In calibration regions, most differences occur in coastal regions at the equator (where generally high water resources are simulated) and upstream mountainous areas, where runoff is high and forcing data sets are likely to vary more strongly than elsewhere (Fig. 5a). In non-calibrated regions, Q varies strongly among the forcing variants in particular in coastal cells in South East Asia (mainly Indonesia, Borneo and Papua New Guinea) and around the Bay of Bengal (Fig. 5b). Average deviation of Q on grid cell level are equal in calibrated regions and non-calibrated regions with about 0.27 $km^3\,yr^{-1}$.

HESSD
doi:10.5194/hess-2015-527

The precipitation undercatch adjustment method (if applied at all) strongly impacts water resources components. Again, for calibration regions, Q is nearly equal (except probably regional differences between CRU and GPCC precipitation pattern), but AET differs largely. For non-calibration regions, Q is clearly lower (1500 $km^3 yr^{-1}$) when precipitation undercatch is not considered (PGFv2 comp. to WFDEI_hom, Table 2). Considering or neglecting P undercatch affects therefore global scale water resources assessments in the same order of magnitude than using climate forcing with e.g. comparable low radiation input (WFD).

4.4 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?

For most areas of the globe, changes in long-term average river discharge between two consecutive 30 year periods (t_1: 1911–1940, t_2: 1941–1970, t_3: 1971–2000) were driven more strongly by changes in P than by changes in human water use and dam construction. The fraction of global land area where changes in Q is driven more strongly by the latter anthropogenic drivers increases with time, indicating an acceleration of the human impact around 1970 compared to 1940. Taking India and GSWP3 forcing as example, P increases for all time steps (t_1–t_2: +49 $km^3 yr^{-1}$, t_2–t_3: +30 $km^3 yr^{-1}$), while Q (+9 $km^3 yr^{-1}$) and WCa (+26 $km^3 yr^{-1}$) increases between t_1 and t_2, whereas Q (−39 $km^3 yr^{-1}$) decreases between t_2 and t_3 and WCa (+81 $km^3 yr^{-1}$) increases stronger than between t_1 and t_2. In India, the intensified water use and changed signs between P and Q lead to the indication that anthropogenic effects dominates the change in Q (comp. Figs. 3 and 4). The effects of human water use and dam construction to Q variations cannot be distinguished by those indicators. While dam construction leading to new reservoirs decreases long-term average Q, human water consumption is expected to be more important in most grid cells.

HESSD
doi:10.5194/hess-2015-527

The four climate forcings used to compute the dominance indicator $I_{varprecdom}$ affect the spatial pattern of the dominance indicators. For example, with the forcings based on ECMWF reanalyses (WFD, WFDEI_hom, Fig. 3c–d), large parts in southeast Australia are driven by anthropogenic effects which is not (PGFv2, Fig. 3b) or less (GSWP3, Fig. 3a) the case for the forcings based on NCEP reanalyses. For WFD, the anthropogenic dominance is considerably higher in Mexico (Fig. 3). If using PGFv2 forcing, the area around the North China Plain is dominated by *P* whereas in the other forcings it is dominated by anthropogenic effects (Fig. 4). Even if mean global values, e.g. for *P* and *Q*, compare well (Fig. 2, Table 1), regional differences in the climate forcings (and underlain reanalysis) result in these different spatial patterns of model output (e.g. *Q*).

5 Conclusions

This study presents a model-based assessment of water balance components on different temporal and spatial scales. The GHM WaterGAP was forced with four (plus one homogenized) state-of-the-art climate forcings. 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 time periods should be used. For variables that (until now) show a small trend only (like *P* and *Q*), the widely used 30 year aggregation period appears to be useful for comparison purposes, while variables showing a stronger trend such as *T* and WCa, decadal aggregation appear to be more appropriate. Of great importance is the choice of the climate forcing. Differences in water balance components are high, reflecting the uncertainty of underlain meteorological data and interpolation methods. Homogenization of climate forcing is required when generating time series of meteorological variables from different sources, as is the case if WFD and WFDEI are combined to cover the time period since 1901 until recent times.

At grid cell, continental and global scales, most of the differences in *Q* among forcings occur in areas where WaterGAP cannot be calibrated due to a lack of information

HESSD
doi:10.5194/hess-2015-527

in the GRDC database (esp. South East Asia). Uncertainty of long-term average Q aggregated over all uncalibrated areas for the homogeneous forcings (GSWP3, PGFv2, WFD, WFDEI_hom) was 18.5 % for the period 1971–2000 but only 1.6 % for the calibrated areas. This supports the many calls for improving the situation of in-situ Q observations (e.g. Fekete et al., 2015) and sharing the already available Q data (e.g. Hannah et al., 2011). Satellite observations can help to constrain the global water balance assessment (Rodell et al., 2015; Tang et al., 2009), but even if the Surface Water and Ocean Topography (SWOT) proposes discharge observations for river widths > 50 m and could thus improve spatial data coverage, satellite observation used to derive Q strongly rely on in-situ gauge measurements (Pavelsky et al., 2014).

Humanity affects the global water cycle. In global numbers for water consumption, the influence is comparably low (Table 1). But at regional (grid cell) level, human water use and dam construction can be more important for changes in Q than P (Figs. 3 and 4). However, in this study only the influence on long-term averaged discharge was analysed, not taking into account seasonal impacts e.g. due to reservoir operation (Adam et al., 2007; Döll et al., 2009).

For future water resources modelling studies (see also Döll et al., 2015), the impact of the uncertainty of meteorological 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. ISI-MIP 2.1) or are already finished (e.g. WATCH water 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. Finally, the assessment of global water resources can be improved when existing high-quality Q data are made available and currently ungauged basins are instrumented (Fekete et al., 2015). In addition, determination of Q from satellite observations could help in the future, but would still require in-situ data.

HESSD
doi:10.5194/hess-2015-527

Data availability

The WaterGAP output will become freely available for the public within the framework of the ISI-MIP project phase 2.1 but it is not yet clarified where the data will be hosted (please check https://www.pik-potsdam.de/research/climate-impacts-and-vulnerabilities/research/rd2-cross-cutting-activities/isi-mip/for-modellers/isi-mip-phase-2 for updates). The homogenized climate forcing WFDEI_hom is not included within the ISI-MIP 2.1a project. All model outputs used in this study are available on request from the corresponding author.

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. 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.

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.

HESSD

doi:10.5194/hess-2015-527

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., 1–55, 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.

HESSD

doi:10.5194/hess-2015-527

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., doi:10.1007/s10712-015-9347-x, in press, 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. 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.

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.

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.

HESSD

doi:10.5194/hess-2015-527

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., 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, Proc. Assoc. Hydrol. Sci., submitted, 2015.
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.
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.

HESSD
doi:10.5194/hess-2015-527

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:10.13019/M20599, 2015.

Sivapalan, M., Takeuchi, K., Franks, S. W., Gupta, V. K., Karambiri, H., Lakshmi, V., Liang, X., McDonnell, J. J., Mendiondo, E. M., O'Connell, P. E., Oki, T., Pomeroy, J. W., Schertzer, D., Uhlenbrook, S., and Zehe, E.: IAHS Decade on Predictions in Ungauged Basins (PUB), 2003–2012: Shaping an exciting future for the hydrological sciences, Hydrolog. Sci. J., 48, 857–880, doi:10.1623/hysj.48.6.857.51421, 2003.

Sood, A. and Smakhtin, V.: Global hydrological models: a review, Hydrolog. Sci. J., 60, 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, 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 Pt. 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.

HESSD
doi:10.5194/hess-2015-527

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., 7505–7514, doi:10.1002/2014WR015638, 2014.

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

HESSD

doi:10.5194/hess-2015-527

Table 1. Global sums of water balance components for land area (except Antarctica and Greenland) [$km^3\ yr^{-1}$] (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	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	63 415	67 588	69 907	68 887
3	Discharge into oceans and inland sinks Q[b]	40 678	39 173	42 200	40 216	41 298
4	Water consumption (actual) (rows 5 + 6) WCa	933	960	915	949	932
5	Net abstraction from surface water (actual)[c]	1050	1071	1026	1070	1044
6	Net abstraction from groundwater[d]	−117	−111	−108	−121	−112
7	Change of total water storage dS/dt[e]	−14	−29	−20	−25	−74
8	Long-term-averaged yearly volume balance error (P – AET – Q – 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 the number of 30 years.

HESSD
doi:10.5194/hess-2015-527

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

	Calibration regions					Non-calibration regions				
No.	GSWP3	PGFv2	WFD	WFDEI_hom	WFD_WFDEI	GSWP3	PGFv2	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

HESSD
doi:10.5194/hess-2015-527

Table 3. Continental ensemble-mean water balance components [$km^3\,yr^{-1}$] and min/max deviation from mean (in %) over all four homogenous forcing variants (GSWP3, PGFv2, WFD, WFDEI_hom), for six continental regions and the global total, and water balance components P, AET, Q, and WCa (1971–2000).

	P			AET			Q			WCa		
Region	Mean	Δmin	Δmax	Mean	Δmin	Δmax	Mean	Δmin	Δmax	Mean	Δmin	Δmax
Africa	20 457	−5.6	3.9	16 194	−5.5	4.7	4183	−5.9	10.9	71	−5.9	2.5
Asia	24 501	−3.0	2.6	13 506	−4.0	3.7	10 276	−3.6	4.4	612	−1.0	1.7
Europe	13 026	−7.0	3.7	6942	−8.7	3.7	6138	−5.0	4.0	87	−8.5	3.6
NAmerica	16 177	−6.9	3.7	9573	−8.0	5.4	6507	−5.4	3.8	121	−5.3	5.7
Oceania	5939	−2.4	1.9	3887	−3.0	3.2	2033	−3.4	5.0	16	−2.8	2.5
SAmerica	28 623	−4.0	2.1	17 132	−5.2	3.4	11 428	−2.2	2.5	32	−2.8	2.8
Global	108 724	−4.8	2.1	67 234	−5.7	4.0	40 566	−3.4	4.0	939	−2.6	2.2

HESSD
doi:10.5194/hess-2015-527

Table 4. 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: CRU TS3.21) and P outputs of WaterGAP using the undercatch adjusted forcings (except PGFv2 which is not adjusted).

Variable	Continent	Africa	Asia	Europe	NAmerica	Oceania	SAmerica	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	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

HESSD
doi:10.5194/hess-2015-527

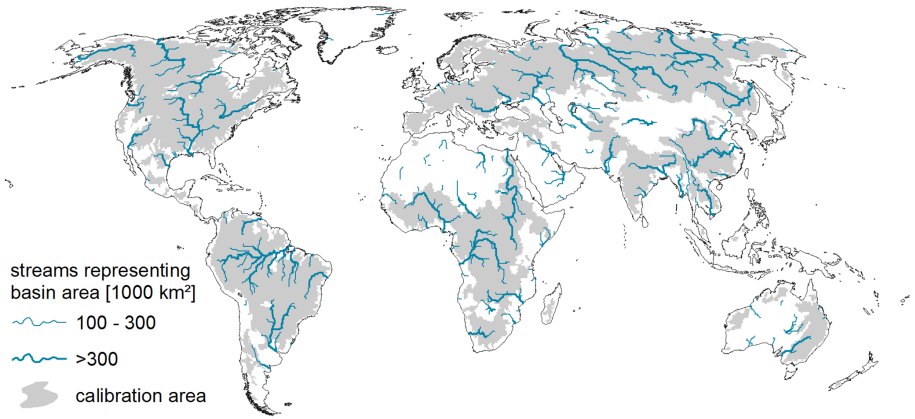
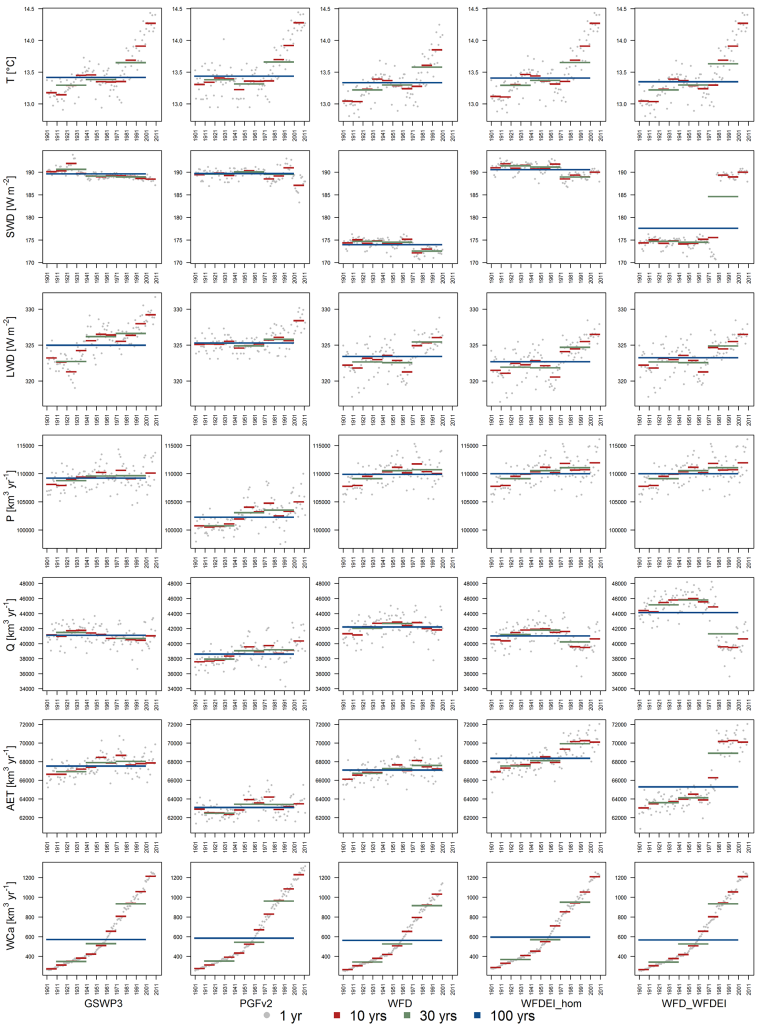


Figure 1. Global land area affected by WaterGAP 2.2 (ISI-MIP 2.1) calibration (grey shading) against observed long-term average river discharge. Streamflow directions and flow accumulation are based on the drainage direction map DDM30 with 0.5° resolution (Döll and Lehner, 2002).

HESSD
doi:10.5194/hess-2015-527



HESSD

doi:10.5194/hess-2015-527

Variations of global and continental water balance components

H. Müller Schmied et al.

Figure 2. Global sums (means) of climatic variables and water balance components for five climate forcings (GSWP3: 1901–2010, PGFv2: 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).

HESSD
doi:10.5194/hess-2015-527

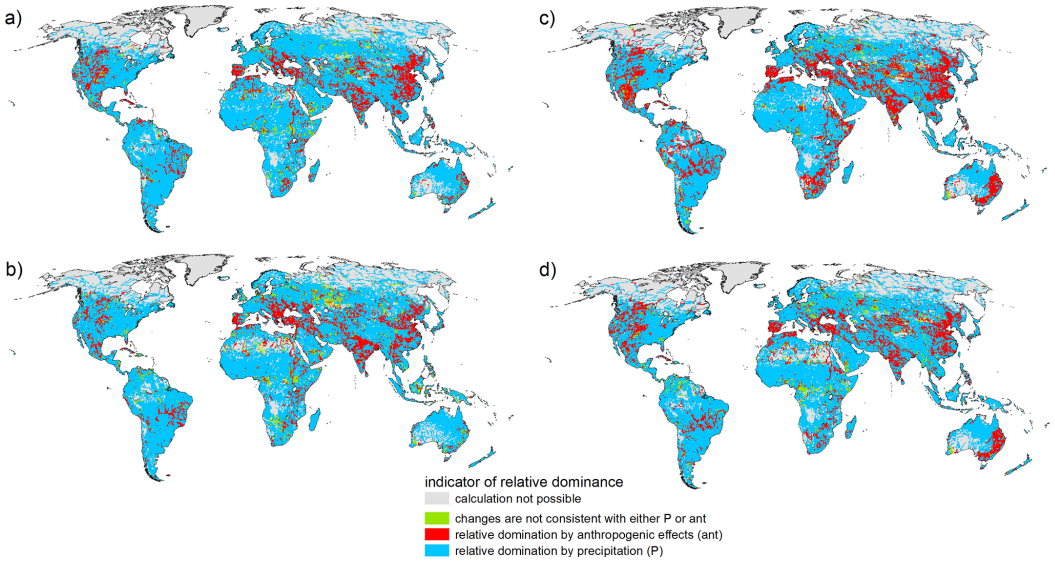


Figure 3. Relative dominance of determinants of change in long-term average Q between 1941–1970 and 1971–2000. Blue indicates that change in P is more dominant than the anthropogenic impact due to water abstraction and dam construction, red indicates the opposite. In green areas, other drivers are dominant. In grey areas a calculation is not possible as the denominator of indicators A_n and/or B_n (see Sect. 2.4.2) is zero. Results are shown for WaterGAP as driven by the meteorological forcings GSWP3 **(a)**, PGFv2 **(b)**, WFD **(c)** and WFDEI_hom **(d)**.

HESSD
doi:10.5194/hess-2015-527

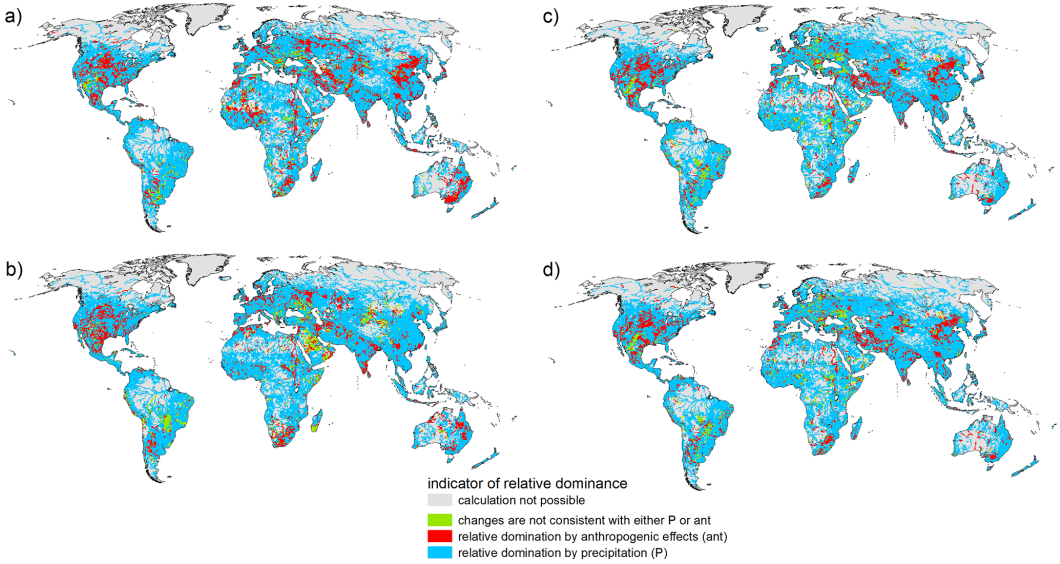


Figure 4. Relative dominance of determinants of change in long-term average Q between 1911–1940 and 1941–1970. Blue indicates that change in P is more dominant than the anthropogenic impact due to water abstraction and dam construction, red indicates the opposite. In green areas, other drivers are dominant. In grey areas a calculation is not possible as the denominator of indicators A_n and/or B_n (see Sect. 2.4.2) is zero. Results are shown for WaterGAP as driven by the meteorological forcings GSWP3 **(a)**, PGFv2 **(b)**, WFD **(c)** and WFDEI_hom **(d)**.

HESSD
doi:10.5194/hess-2015-527

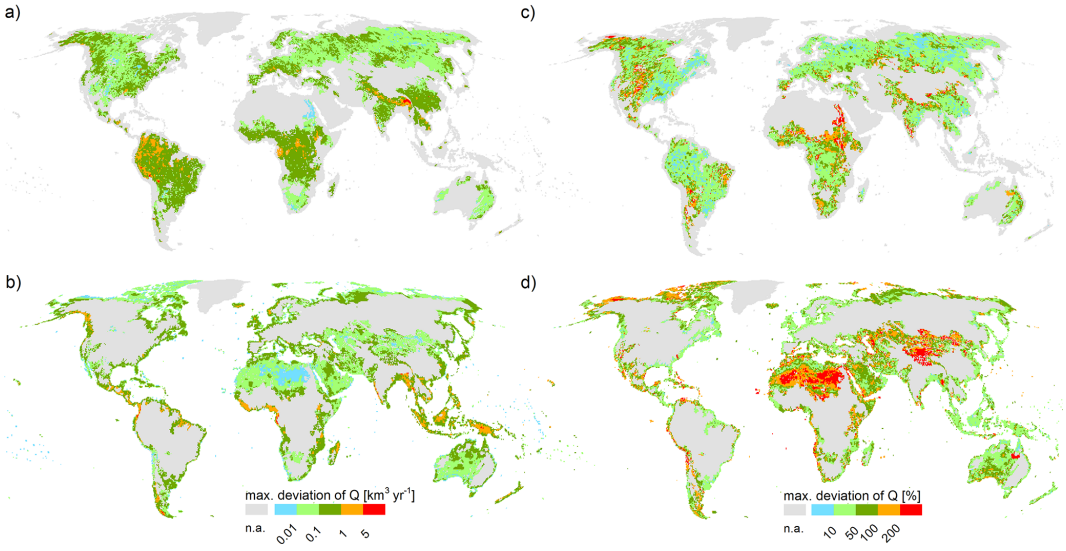


Figure 5. Maximum difference of long term averaged (1971–2000) *Q* among the four homogeneous climate forcings (GSWP3, PGFv2, WFD, WFDEI_hom), expressed as deviation [$km^3\,yr^{-1}$] **(a, b)** and relative deviation **(c, d)** separately for calibration **(a, c)** and non-calibration **(b, d)** regions. Grey areas contain either no discharge or are outside the region of interest, i.e. non-calibration areas are grey in **(a)** and **(c)** and vice versa.

HESSD
doi:10.5194/hess-2015-527