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

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

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 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 modelling 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 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 modelling 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 observation-based WFD/WFDEI climate data. Discharge computed by a calibrated model version using monthly CRU 3.2 and GPCC 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 short wave 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 current.

1 Introduction

The estimation of global scale freshwater fluxes, in particular river discharge, is essential to assess e.g. availability and scarcity of water resources for humans and the environment both in recent times (Hoekstra et al., 2012; Oki and Kanae, 2006; Prudhomme et al., 2014) and in future scenarios (Döll and Müller Schmied, 2012; Masaki et al., 2014; Schewe et al., 2014). Global 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) were estimated as well as groundwater related fluxes like 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 (Wada et al., 2010).

There are different ways to estimate global scale freshwater fluxes and storages. Interpolation of in-situ measurements works well with a dense monitoring network, as for precipitation products (e.g. GPCC (Schneider et al., 2014), CRU (Harris et al., 2014) and many more) or, even with less dense point measurements in combination with other data sources like remote sensing for evapotranspiration (Jung et al., 2010). In particular, remote sensing is used to derive spatio-temporal 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). Spatio-temporal 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, solve the energy balance (Haddeland et al., 2011) and have some limitations, esp. 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 (Nijssen et al., 2001), WBM (Vörösmarty et al., 1998), Mac-PDM (Gosling and Arnell, 2011), WASMOD-M (Widén-Nilsson et al., 2007), H08 (Hanasaki et al., 2008), Water – Global Assessment and Prognosis (WaterGAP) (Alcamo et al., 2003; Döll et al., 2003) and PCR-GLOBWB (Sperna Weiland et al., 2010).

Results of LSMs and GHMs are highly uncertain. Epistemic uncertainty due to a lack of knowledge and understanding is of particular importance at the global scale (e.g. see discussion in Beven and Cloke (2012) and Wood et al. (2011, 2012)). Generally, three sources of uncertainty can be distinguished: spatially distributed input data (e.g. climate forcing, water use, land cover), model structure (or modeling approach) and model parameters.

Uncertainties due to the choice of climate forcing were 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 (esp. to precipitation and radiation), and concluded that this uncertainty on land surface hydrology is as large as the variations 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 with these precipitation forcings and concluded with an average uncertainty in discharge of about 90%. Even though climate forcing is of such importance, only few studies are available reflecting this uncertainty in a global hydrological model setup.

Uncertainties in terms of model structure are related to the design of the model, i.e. the (number of) 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 its uncertainties. Model intercomparison efforts in which identical climate forcing is used to drive all investigated models (e.g. WATCH WaterMIP, ISI-MIP) have shown the effects of different model structures (Gudmundsson et al., 2012a, 2012b; Haddeland et al., 2011; 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 multi-model 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 assessed 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-PDF.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 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).

The goal of this study is to analyze the impact of the uncertainty of 1) spatially distributed input data and 2) model structure and modeling approach on water fluxes and storages at the global scale, using the most recent version of the GHM WaterGAP 2.2. As previous studies (Kaspar, 2003; Schumacher et al., 2014; Werth and Güntner, 2010) have already investigated both parameter sensitivity and uncertainty for WaterGAP, and due to length issues, this is not focus of this study. The study was motivated by newly available climate forcing and land cover input data as well as the significant modifications of the WaterGAP model structure during the last decade.

In particular, we will answer the following research questions:

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

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

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

iv) 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 study 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 with a spatial resolution of 0.5° by 0.5° (55 km by 55 km at the equator) for the whole land area of the Earth except Antarctica (66896 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 basin (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 semi-arid / 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 Study 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 modeled time span is from 1901 to 2009. In this paper, model results for 1971-2000 are evaluated.

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 different climate forcing datasets on calibration and subsequently on freshwater fluxes, two climate forcings were used to force both the 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 data (WFD) (Weedon et al., 2011) for the years 1901 to 1978 (the years 1901 to 1957 are based on reordered reanalysis data) and the WATCH Forcing Data methodology applied to ERA-Interim data (WFDEI) (Dee et al., 2011; Weedon et al., 2010, 2011) for the years 1979 to 2009 was chosen. Switching the climate input dataset in 1979 leads to inconsistencies in terms of AET (much 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 TS2.1 for WFD and to CRU TS 3.1 for WFDEI as well as adjustment of monthly precipitation sum to GPCC v4 (WFD) and GPCC v5 (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 dataset CRU TS 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 datasets (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 dataset; 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 short wave radiation as well as outgoing long wave 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.

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 dataset 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 land cover information was used.

In variant LANDCOVER, a combination of the GLCC based on the years 1992/1993 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 (Alcamo et al., 1998) model outputs (as in previous model versions). Both input datasets 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 m s^{-1} and the meandering ratio is set to 1, 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 arc minute sub-grid (Schulze and Döll, 2004)) which should lead to less snow dynamics.
- Finally, there is no distinction in groundwater recharge for semi-arid / arid and humid regions (in contrast to Döll and Fiedler (2008) all regions are treated like humid regions) resulting in higher groundwater recharge in semi-arid / 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 study, 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

As described in Appendix B, WGHM is calibrated in a basin-specific manner, against mean annual discharge by adjusting, in all grid cells within each of the 1319 calibration basins, a runoff coefficient that affects the outflow from the soil compartment, and – if necessary to simulate mean annual discharge within 1% of the observed value – two additional correction factors. 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 corrections factors are set to 1 in all grid cells. The comparison of NoCal to e.g. 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 Figure 3 it is important to know that actual evapotranspiration (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 and 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 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 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, P and Q are both increased by around 1900 km^3 yr^{-1} whereas global AET sums are nearly equal. When partitioning the increased Q into calibrated and uncalibrated grid cells, most additional Q (1546 of overall 1906 km^3 yr^{-1}) is generated in non-calibrated 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 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, col NoUse and STANDARD). Q of the other model variants and hence RWR increase 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 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 that is based on ERA-Interim, AET is around 70 000 $km^3\ yr^{-1}$, compared to 65 000 $km^3\ yr^{-1}$ in case of WFD. This is caused by differences in the shortwave downward radiation (much higher in WFDEI) which impacts the net radiation as 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 (around five) 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 global average of the latter two storages but decreasing since 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 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, the 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 case of CLIMATE (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 PET 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 ¾ of PET. 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 Great Lakes in Fig. 3).

Differences between NoCal and STANDARD are resulting due to the calibration parameter γ which differs from 1.0 (NoCal) in most cases in STANDARD (and the other model variants). For example, there are blue patterns in China and South America. In both regions, γ is less 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. north eastern 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 wetlands 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 the outflow cell, hence AET can become very large as the value in mm 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 of Fig. 3 of the previous section. Compared to STANDARD, 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 (Table 2). The global sum of RWR from CLIMATE is 4.7% higher but with large spatial spread. RWR decreases 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 decreases (increases) 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 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 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 e.g. 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, 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 River, 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 river basin is also frozen during winter time 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 consists of mixed forest or cropland / natural vegetation mosaic which both have a lower albedo, resulting in more evaporation and less discharge especial in the summer months. Additional effects occur due to deeper roots at mixed forest class. Only for the Mackenzie, Lena and Yangtze, mean monthly river discharges of NoCal within the range of all other variants in some months. The NoCal values for the Orange river are so high that throughout the year, they are higher than the highest observed value (and the values of the other variants) (Fig. 5). This supports the use of 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 1319 gauging stations used for calibration.

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

with 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, E_{NS} of monthly discharge increases in all calibrated model variants as compared to the NoCal variant, as E_{NS} is sensitive to both 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 in case of STANDARD and NoUse, E_{NS} is larger than 0.5 in 53.5% of the basins. Comparing 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, in particular 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, effects on TWS variations are strong (e.g. in Southern Congo or in 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 water sources) 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 (e.g. High Plains Aquifer in central USA, Iran and Northwestern India) (Fig. 7d and 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 large 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 the NoCal variant, which has very low AET and high discharge values (Tables 2 and 5). Difficulties with such comparisons can occur if different time spans are used. In addition, different land area is used, e.g. Mu et al. (2011) is based on remote sensing data and neglects bare land surfaces (their area: 109.03×10^6 km^2) whereas Mueller et al. (2013) covers 130.92×10^6 km^2 (which is also a reason for a larger AET).

Discharge estimates differ due to the applied estimation method and precipitation data set. Mueller et al. (2013) do not consider precipitation undercatch correction and assume a global precipitation of ~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). Regarding WaterGAP, estimated of global discharge, model refinements have led to an increase of discharge. The value of STANDARD is approx. 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 and 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 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 (Fig. 4e) dominates all other modifications within this study 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. AET is largely reduced in one half of the basin (and vice versa) at the river basin Yenisey at station Igarka (western Siberian Plain) when using alternative climate forcing. Transferring the correction factor CFS (which is, if necessary, calculated 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 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. 59% of the basins in Fig. 8 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 boundaries minimum (0.1) or maximum (5.0) value and the influence of input data and model structure, which were modified in this study, is insignificant. On the other hand, 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. Anyhow, 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 (multi-parameter calibration) and/or by integrating additional reference data, e.g. GRACE based data as was shown by Werth and Güntner (2010) (multi-objective 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, 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 when the belonging attributes were also modified. Forced with CRU 3.2 and GPCC v6 instead of WFD/WFDEI input, AET is 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 decreases by the same amount as AET increased (e.g. South East Asia, Australia, Saudi

Arabia). In other regions, no clear effect on RWR is detectable (e.g. North America). In some parts of Europe, RWR increases by at least 10 mm yr^{-1} even if AET increases. Here, besides radiation (affecting AET), the amount of precipitation is of great importance (affecting RWR).

In regions where the climate forcing datasets differ significantly (e.g. Danube River Basin), the impact on discharge is large (Fig. 5 bottom center). 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). 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. For example, the different elevations of the 100 subgrids used for the improved snow modeling (Schulze and Döll, 2004) lead to different temperatures (see Appendix A2) and thus to more differentiated snow melting within one 0.5° grid cell in STANDARD as compared to STRUCTURE where snow within the whole cell either melted or not on any day. In many basins in the alpine region in central Europe, E_{NS} of STRUCTURE ranks behind STANDARD (Fig. 9d, red colors) reflecting too early snow melt in STRUCTURE. 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, 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 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 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 and 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 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 (Appendix A5) decreases evaporation from surface water bodies. If the impact of human water use on river discharge were not considered, van Beek et al. (2011) showed lower performance in general. Within our study, 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, 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 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.

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 and 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 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). For the modeling experiment, 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 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 (short wave 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. section 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 the observed on, 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 semi-arid and arid regions are poor, and improved modeling of evaporation from ephemeral ponds is planned.

Appendix

Appendix A describes the WaterGAP Global Hydrology Model (WGHM) in its current version 2.2. In the order of processing, the single storage compartments and 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 sub-models, 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 [-1] 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 (α = 1.26) and semi-arid / arid (α = 1.74) areas. Grid cells were defined as semi-arid / arid if 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

$$S_{c,\max} = m_c L \quad \text{(A4)}$$

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:

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

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 semi-arid and arid regions, it is necessary that at least 0.5 mm daily precipitation occurs to keep the growing season ongoing. If the condition for 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 [$^{-1}$] is calculated as

$$\frac{dS_{sn}}{d_t} = P_{sn} - M - E_{sn} \quad \text{(A6)}$$

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

Snow accumulation and melt are modeled on a 3 arc minute sub-grid (100 sub-grid cells per 0.5°) using a degree day algorithm (Schulze and Döll, 2004). Mean sub-grid elevation was derived from GTOPO30 (U.S. Geological Survey, 2003). The daily temperature for each sub-grid 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 sub-grid.

At temperatures below 0 °C, all precipitation is assumed to fall and accumulate as snow. At sub-grid 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:

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

Instead of using one specific albedo for snow as in previous versions (α = 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 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 × 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 [$^{-1}$] is calculated as one layer as:

$$\frac{dS_s}{dt} = P_{eff} - R_l - E_s \tag{A8}$$

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.

$$P_{eff} = P_t - P_{sn} + M \tag{A9}$$

with P_t is through fall [mm d^{-1}], (see Sect A1), P_{sn} is precipitation falling as snow [mm d^{-1}] and M is snow melt [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

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

where $E_{p,\max}$ is 20 mm d^{-1} in semi-arid 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

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

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 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 $\mathrm{NA_g}$ [mm d^{-1}] (Section C), which can also act as inflow (e.g. as additional groundwater recharge due to irrigation with surface water).

$$\frac{dS_g}{dt} = R_g - Q_g - \mathrm{NA_g} \tag{A12}$$

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 of (Döll and Fiedler, 2008) as that groundwater recharge estimation is confirmed with experts within the WHYMAP (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

$$Q_g = k_g S_g \tag{A13}$$

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 by 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:

$$\frac{dS}{dt} = I - Q + B - \mathrm{NA}_s \tag{A14}$$

Outflow is in principle modeled like groundwater outflow (Sect. 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

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

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 dataset on lake volumes is available, the maximum storage capacity is determined by multiplying the surface area with an "active" depth (set to 5 m 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 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. River bed slopes were calculated based on the HydroSHEDS 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 non-irrigation 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 in Döll et al. (2009), assuring that at least some water is released e.g. for downstream ecosystem demands.

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. 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 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 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 sub-basins. 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 with 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 next the downstream basin, the modeled discharge is corrected to the measured discharge in the grid cell where the discharge station is located by multiplying with the station correction factor CFS (Hunger and Döll, 2008).

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: 28.09.2012) were included in the calibration setup if they fulfilled three main criteria: (1) an upstream area of at least 9000 km², (2) a time series of at least four (complete) years, and (3) an inter-station catchment area of at least 30 000 km². 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 to 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 to 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 pre-processing steps to the WGHM, the global water use sub-models (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 Global Irrigation Model (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 non-rice crops, based on modelled cropping patterns (Döll and Siebert, 2002). Consumptive livestock water use is calculated as a function of animal numbers per grid cell and water requirements per capita for ten 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 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., 2014). Return flows from irrigation to either groundwater or surface water are computed as a function of 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., 2014).

Acknowledgements

The authors thank the Global Runoff Data Center, D-56002 Koblenz, Germany for providing the discharge data used in this study. The authors acknowledge the comments of two anonymous referees that helped to improve the manuscript significantly.

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., 443–476, Water Resources Publications;, Lone Tree, USA, 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.

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/jhydrol.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. Dyn. 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.

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. R. Meteorol. Soc., 137(656), 553–597, doi:10.1002/qj.828, 2011.

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., 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, 2014.

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.

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, Glob. Environ. Chang., 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. Dyn., 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, 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, 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, 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, 1068–72, 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. U. S. A., 111, 3262–7, 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. U. S. A., 111, 3245–50, 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. Chang., 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?ds_id=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.

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, (available at: www.eu–watch.org/publications), 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.

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.

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 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 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)

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.

nr	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	-7	-7	-8	-7	-88	-7
	($P – AET – Q – WC_a – dS/dt$) 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] total water storage (TWS) of 31. December 2000 minus

TWS of 31. December 1970 divided by 30 years, [h] STANDARD but no subtraction of water use; discharge into oceans and inland sinks equals renewable water resources.

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

compartment	STANDARD	NoUse[a]	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

[a] 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

Table 4. Number of calibration basins per E_{NS} category and Köppen-Geiger climate zone[a].

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

[a] 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 basis for Köppen-Geiger climate calculation is CRU TS 3.2 and GPCC v6 instead of WFD/WFDEI climate input for all other variants.

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)
70576[c]	STANDARD	40458[c]	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

Table 6. The three model variants with the largest differences to STANDARD variant (dSTA) regarding global freshwater fluxes (Q and AET) and total water storages 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

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	% 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

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)

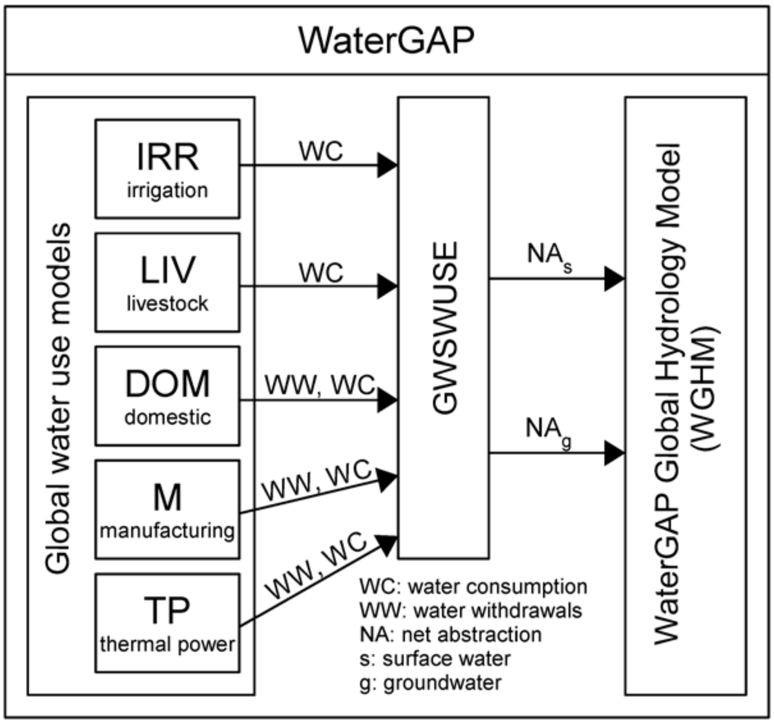

Figure 1. Schematic of WaterGAP 2.2. The output of five water use models is translated into net abstractions from groundwater NA_g and surface water NA_s by the submodel GWSWUSE, which allows computing the impact of human water use on water flows and storages by WGHM. For details see Döll et al. (2012).

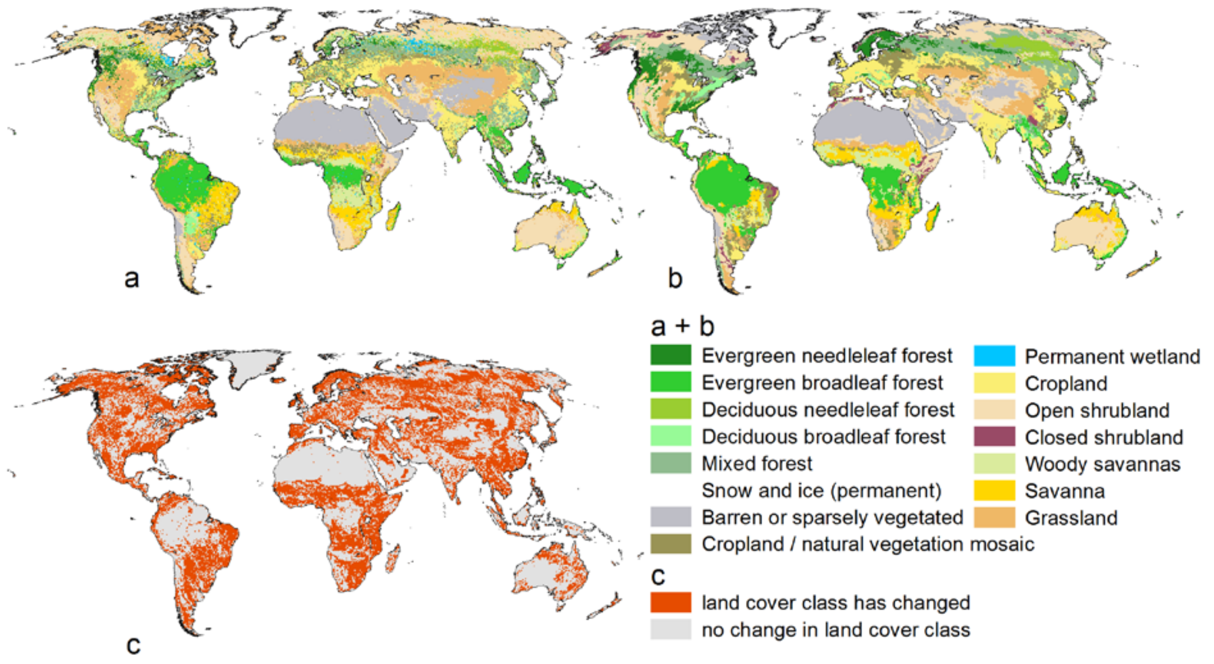

Figure 2. Land cover maps with a spatial resolution of 0.5° used as WaterGAP input based on MODIS observations for the year 2004 (variant STANDARD) (a), land cover derived from USGS GLCC but CORINE for Europe reflecting land cover distribution around the year 2000 (variant LANDCOVER) (b), and identification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

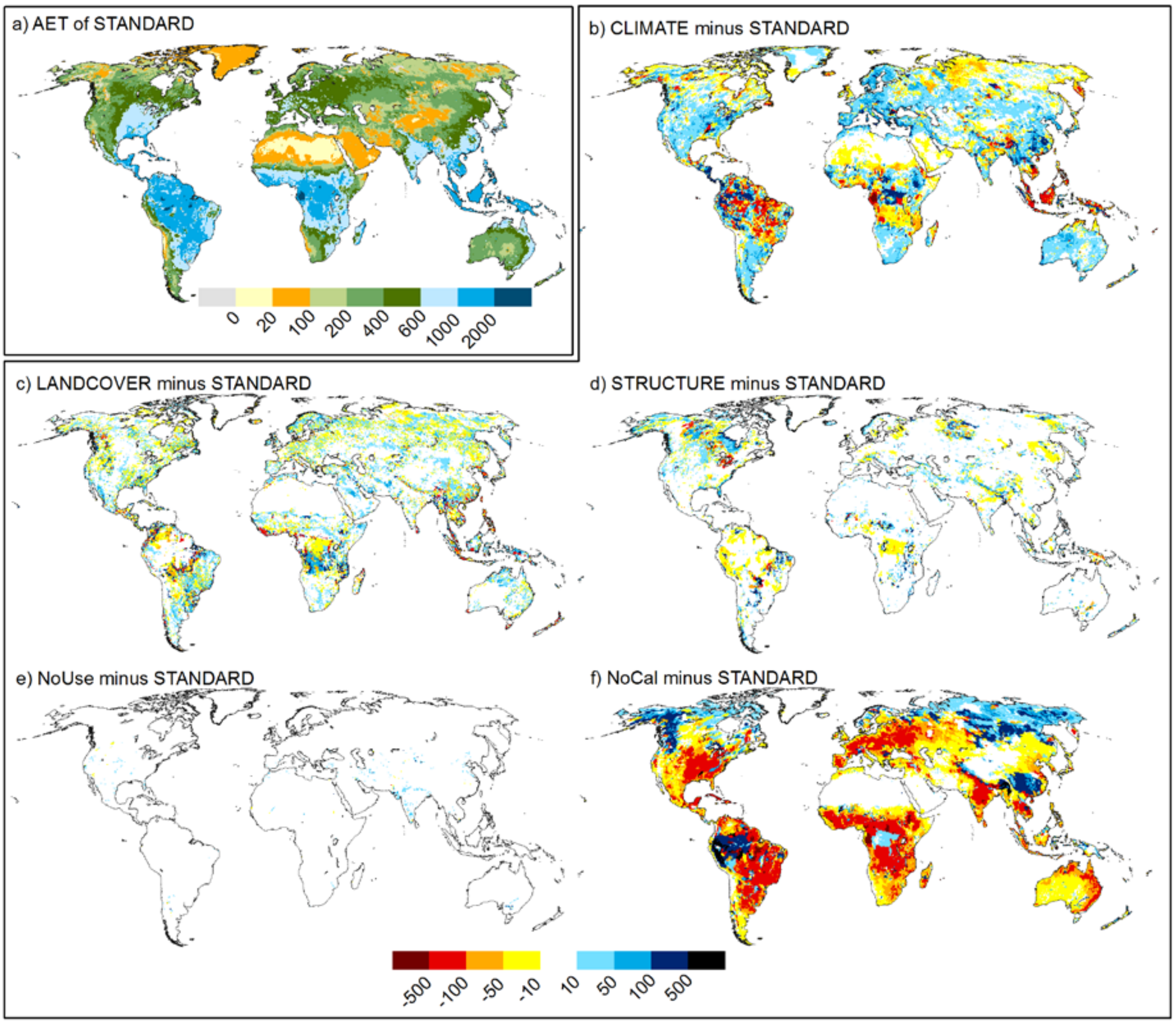

Figure 3. Actual evapotranspiration AET for STANDARD (mean value 1971-2000, in mm yr^{-1}) (a) and differences between the model variants and STANDARD in mm yr^{-1} (b-f).

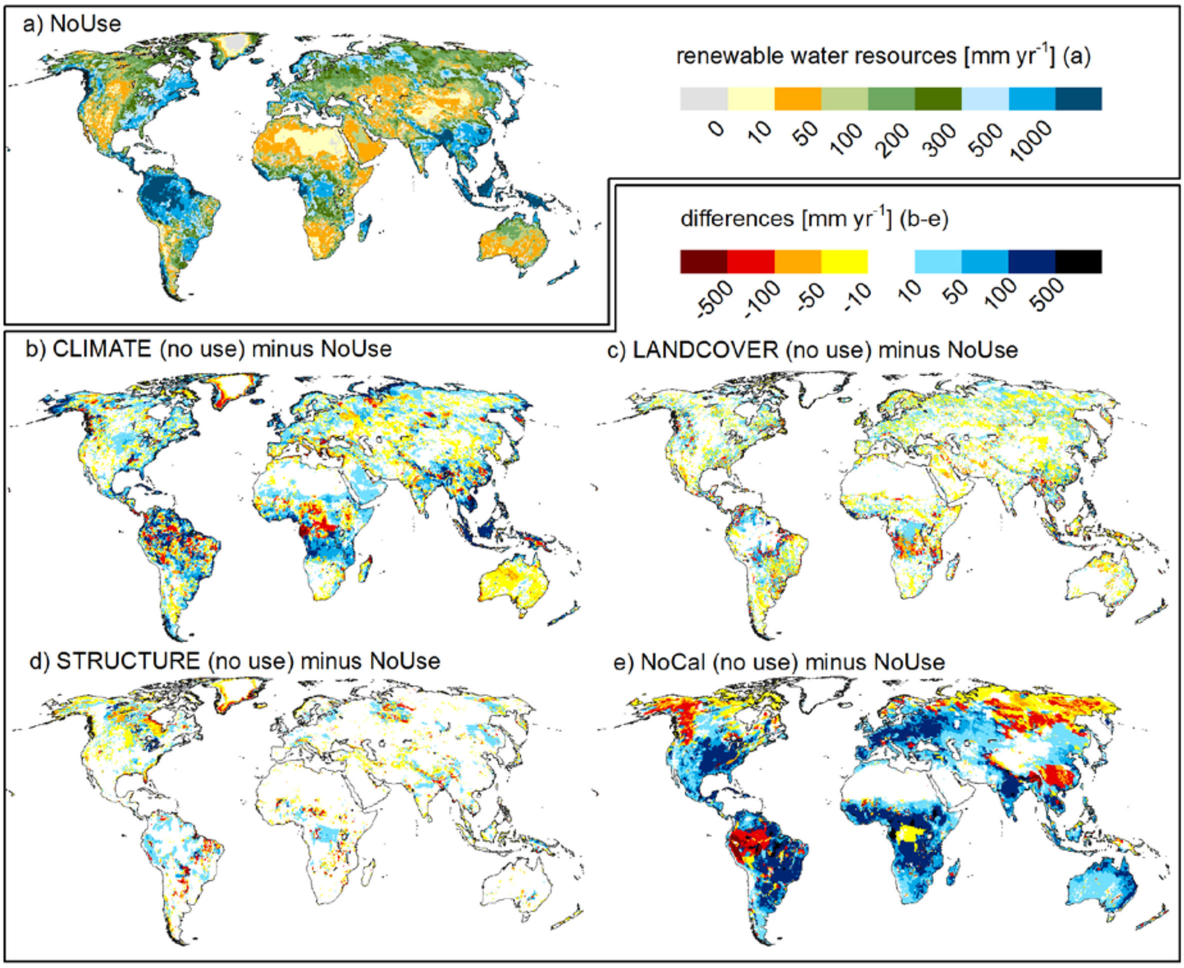

Figure 4. Renewable water resources (mean annual runoff from each cell if water use is neglected) calculated by WaterGAP 2.2 NoUse variant (a) and differences to other variants (variants here run without considering water use) (b-e).

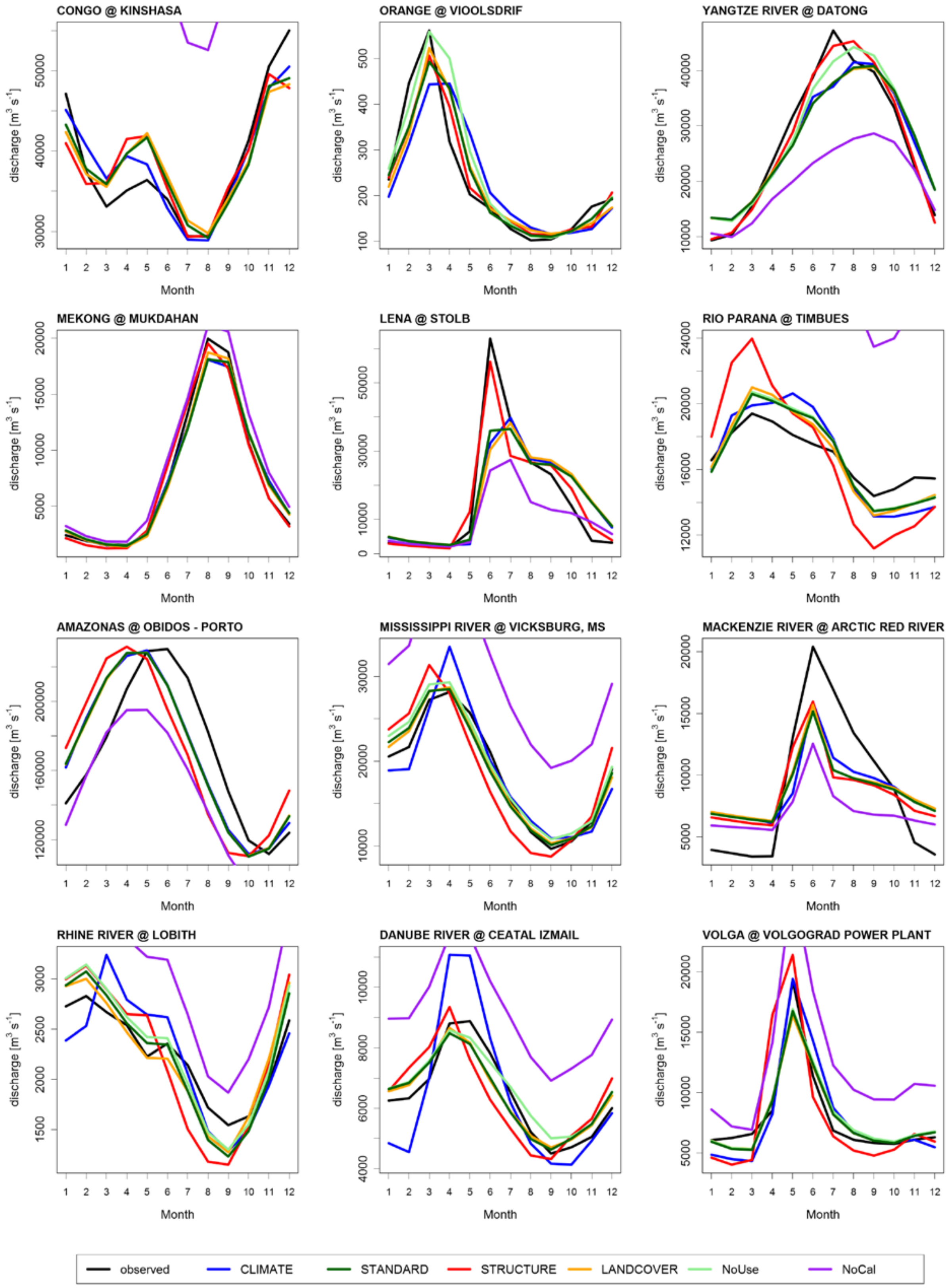

Figure 5. Discharge seasonality for selected basins and the calibrated model variants. Values for NoCal are only visible if they are in the range of calibrated model variants.

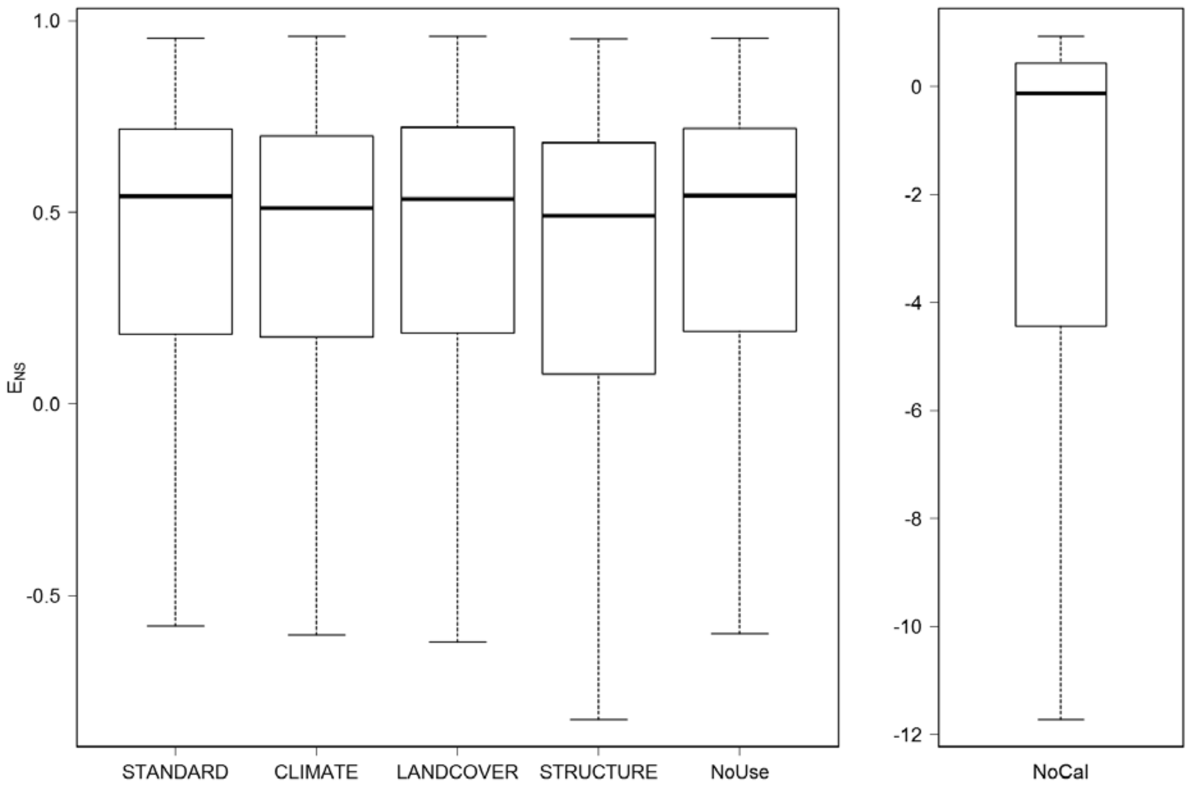

Figure 6. Nash-Sutcliffe efficiencies E_{NS} (excluding outliers) of monthly observed and simulated discharge at 1319 stations used for calibration.

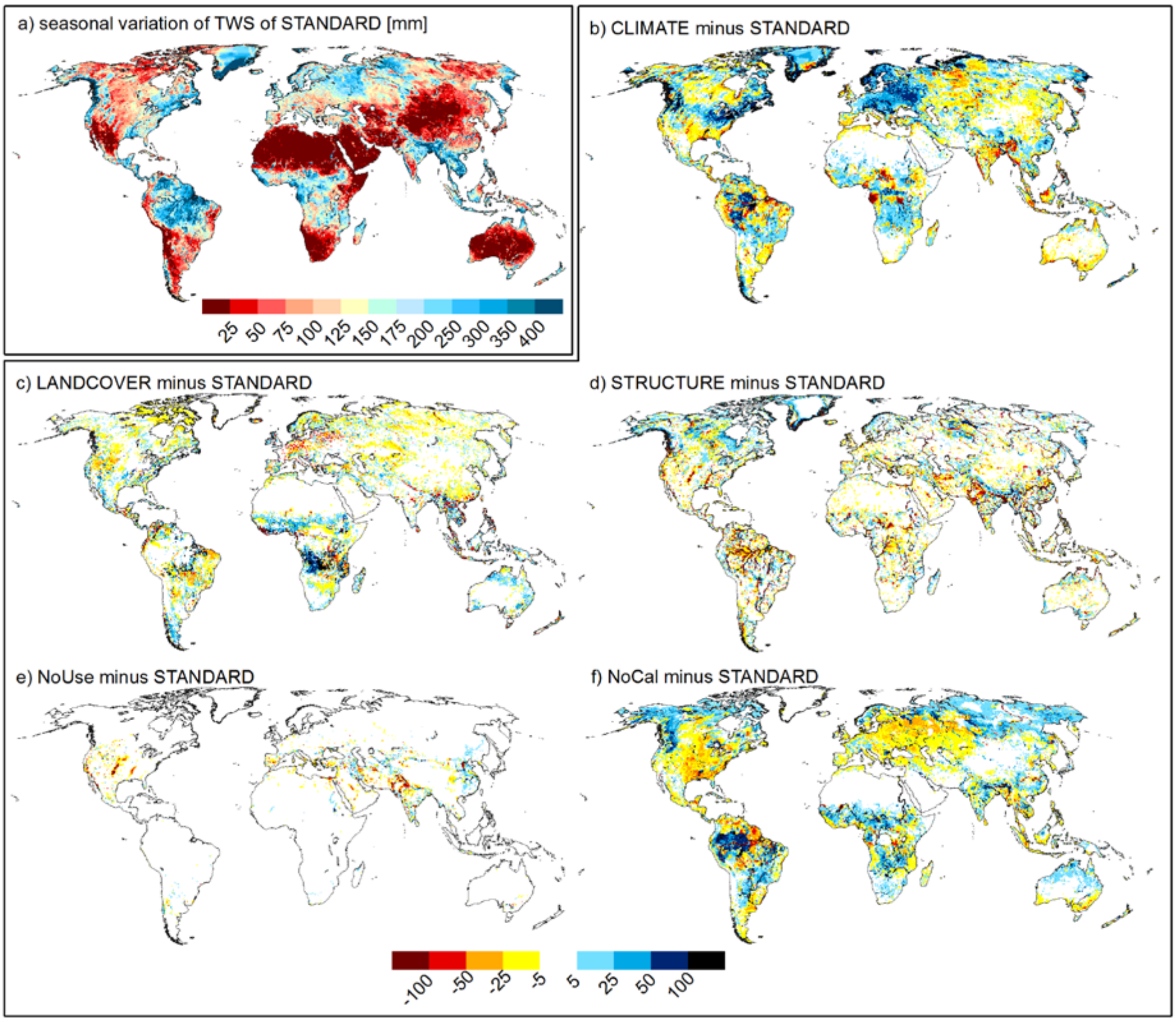

Figure 7. Seasonal variation of total water storage (TWS) for STANDARD (a) and as difference maps [mm] to all other model variants (b-f).

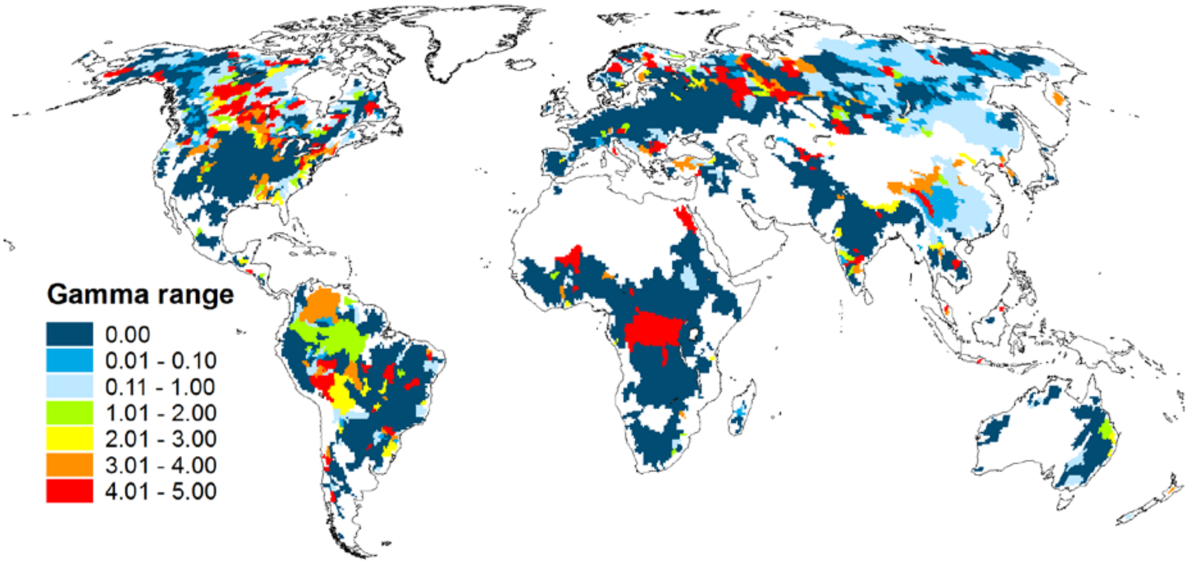

Figure 8. Range of calibration parameter γ through all four calibrated model variants (calculated as $\gamma_{\max} - \gamma_{\min}$) showing the general sensitivity to input data and model structure. White colors indicate uncalibrated regions.

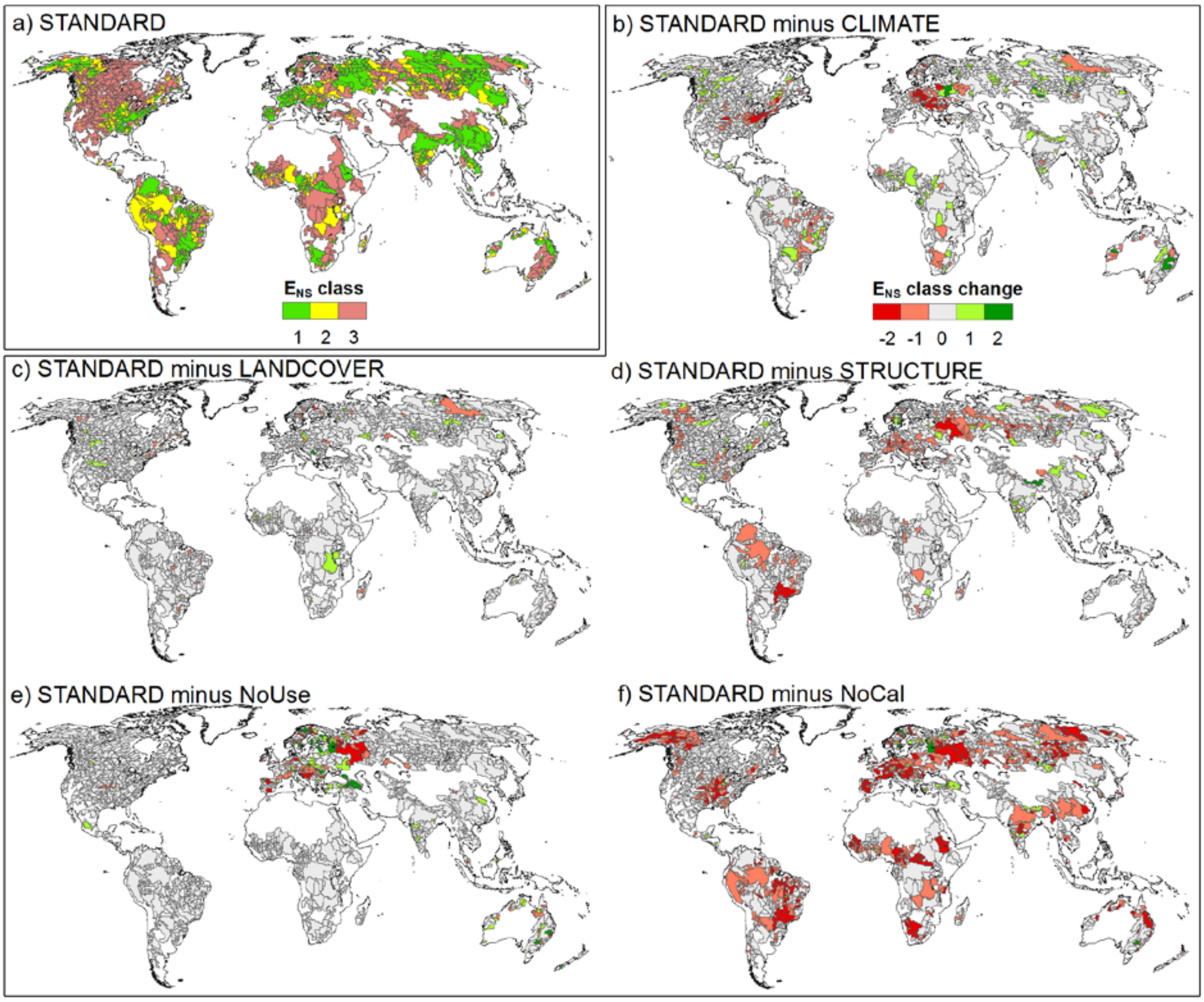

Figure 9. Spatial distribution of Nash-Sutcliffe efficiency E_{NS} classes (from Table 4, 1: E_{NS} > 0.7, 2: 0.5 < E_{NS} < 0.7, 3: E_{NS} < 0.5) for STANDARD (a), and differences of model variants (calculated as STANDARD E_{NS} class minus that of the model variant) (b-f). Red colors indicate a decrease,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

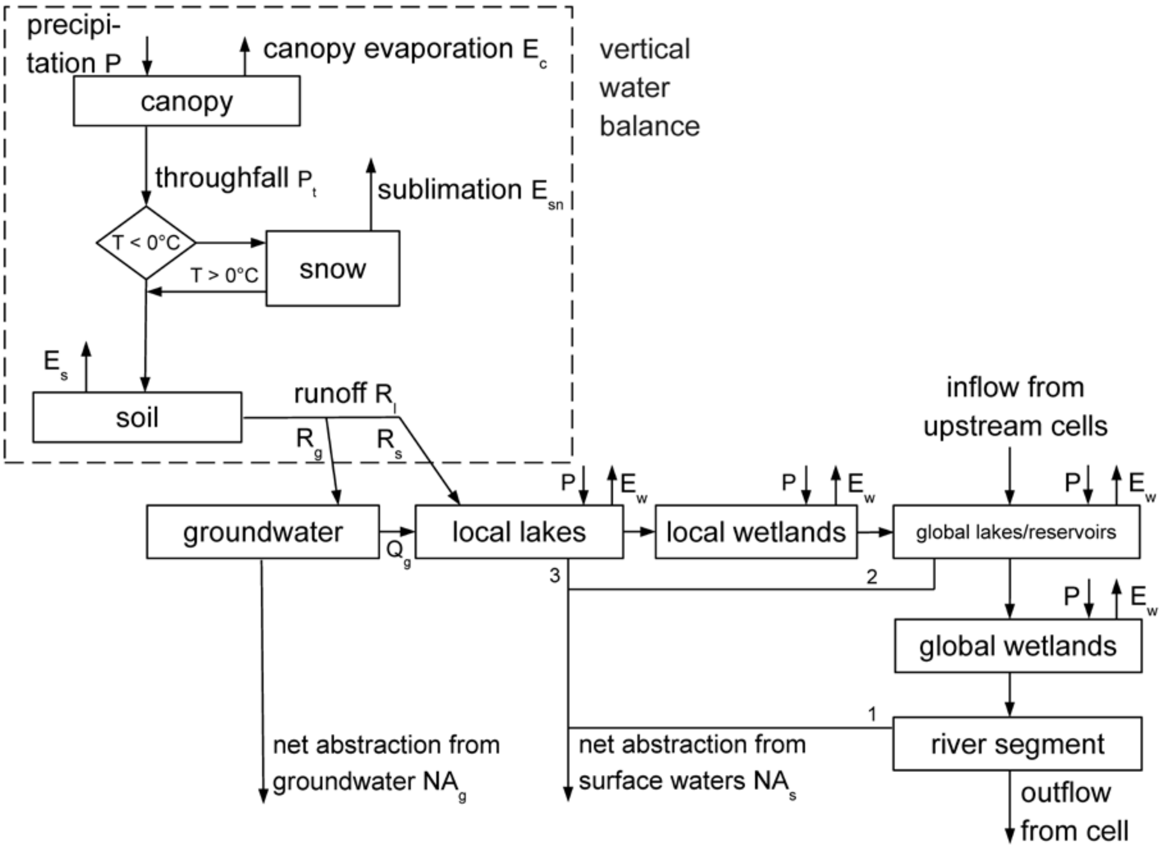

Figure A1. Schematic structure of the water fluxes and storages as computed by WaterGAP Global Hydrology Model (WGHM) within each 0.5° grid cell. Boxes represent water storage compartments, arrows water fluxes (inflows, outflows). Numbers at net abstraction from surface waters (NA_s) are the order from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

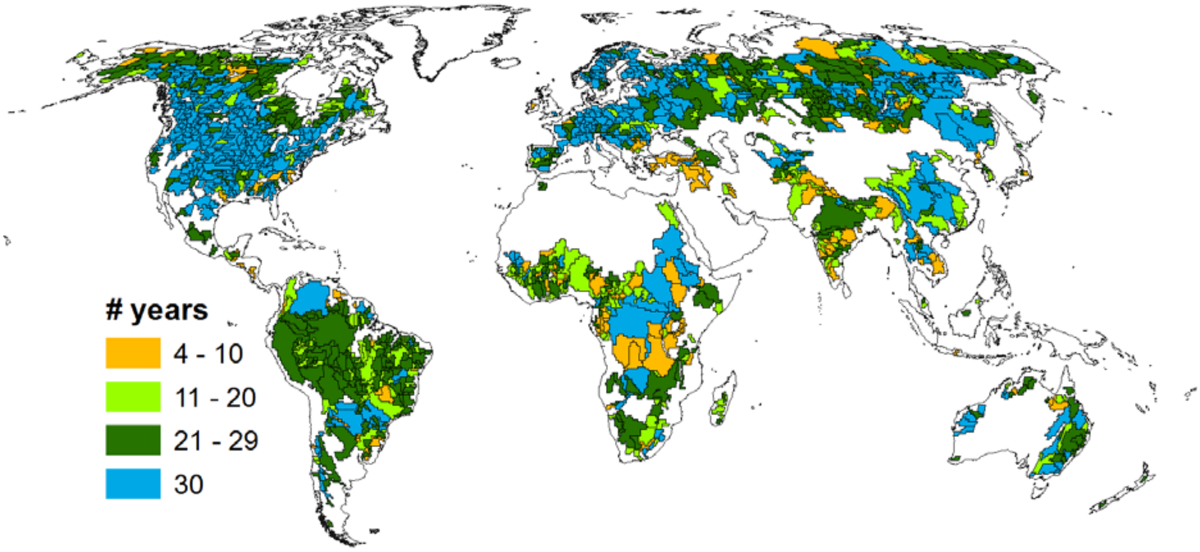

Figure B1. Calibration basins of WaterGAP 2.2 with number of years with discharge observations used for calibration.