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

H. Müller Schmied¹, S. Eisner², D. Franz³, M. Wattenbach³, F. T. Portmann¹,4,5, M. Flörke², and P. Döll¹

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

*now at: Inorganic and Isotope Geochemistry, GFZ Potsdam, Potsdam, Germany

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

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

Abstract. Global-scale assessments of freshwater fluxes and storages by hydrological models under historic climate conditions are subject to a variety of uncertainties. Using the global hydrological model WaterGAP (Water – Global Assessment and Prognosis) 2.2, we investigated the sensitivity of simulated freshwater fluxes and water storage variations to five major sources of uncertainty: climate forcing, land cover input, model structure/refinements, consideration of human water use and calibration (or no calibration) against observed mean river discharge. In a modeling experiment, five variants of the standard version of WaterGAP 2.2 were generated that differed from the standard version only regarding the investigated source of uncertainty. The basin-specific calibration approach for WaterGAP was found to have the largest effect on grid cell fluxes as well as on global AET (actual evapotranspiration) and discharge into oceans for the period 1971–2000. Regarding grid cell fluxes, climate forcing ranks second before land cover input. Grid cell fluxes, climate forcing ranks second before land cover input. Global water storage trends are most sensitive to model refinements (mainly modeling of groundwater depletion) and consideration of human water use. The best fit to observed time series of monthly river discharge or discharge seasonality is obtained with the standard WaterGAP 2.2 model version which is calibrated and driven by daily reanalysis-based WFD/WFDEI (combination of Watch Forcing Data based on ERA40 and Watch Forcing Data based on ERA-Interim) climate data. Discharge computed by a calibrated model version using monthly CRU TS (Climate Research Unit time-series) 3.2 and GPCC (Global Precipitation Climatology Center) v6 climate input reduced the fit to observed discharge for most stations. Taking into account uncertainties of climate and land cover data, global 1971–2000 discharge into oceans and inland sinks ranges between 40 000 and 42 000 km³ yr⁻¹. Global actual evapotranspiration, with 70 000 km³ yr⁻¹, is rather unaffected by climate and land cover uncertainties. Human water use reduced river discharge by 1000 km³ yr⁻¹, such that global renewable water resources are estimated to range between 41 000 and 43 000 km³ 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 present.

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

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

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

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

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

Uncertainties in terms of model structure are related to the design of the model, i.e., the processes considered and their representation by conceptual approaches. To consider this kind of uncertainty, Butts et al. (2004), Clark et al. (2008), Refsgaard et al. (2006) and Song et al. (2011) developed approaches to diagnose different structures of hydrological models and related uncertainties. The model intercomparison efforts WATCH WaterMIP (Water and Global Change Water Model Intercomparison Project; Haddeland et al., 2011) and ISI-MIP (Inter-Sectoral Impact-MIP; Warszawski et al., 2014) have shown the effects of different model structures (Gudmundsson et al., 2012a, b; Hagemann et al., 2013; Van Loon et al., 2012; Prudhomme et al., 2014; Schewe et al., 2014) even though this was not explored systematically. For example, values for global annual evapotranspiration between 60,000 and 85,000 km³ yr⁻¹ were reported in the WATCH WaterMIP study (Haddeland et al., 2011). In such multimodel studies, many completely different models are participating, which makes it very difficult to identify the reasons for different model behavior. A sensitivity study using basically the same model but with a refined model structure can therefore be of benefit (e.g., Thompson et al., 2013).

Model parameters are used to represent system dynamics in solvable equations, in particular when the hydrological process cannot be described physically. These parameters are generally not measurable and, hence, are a source of uncertainty that can influence model results to varying degrees. Within the GCM community, the perturbed physics ensemble
approach has been used to assess this kind of uncertainty in a structured way (Collins et al., 2006; Rowlands et al., 2012). Global-scale hydrology applications also assess parameter uncertainty. For example, Gosling and Arnell (2011) used seven sets of parameter perturbations for two model parameters of the GHM Mac-PDM.09 (Macro-scale – Probability-Distributed Moisture model.09). For the GHM WaterGAP, Kaspar (2003) investigated the impact of uncertainty of 38 model parameters on simulated river discharge by conducting various model runs with a sampling of parameter values within specific ranges. He found that major uncertainties are related to evapotranspiration parameters and land cover specific attributes. Schumacher et al. (2014) confirmed the sensitivity of model output (here: monthly total water storage) to radiation calculation and related parameters in WaterGAP which, together with a river roughness coefficient and precipitation, dominate uncertainty in many of the 33 investigated river basins. Groundwater-related parameters and soil parameters were found to be important for the timing and variation of total water storages in WaterGAP (Werth and Güntner, 2010).

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

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

In particular, we will answer the following research questions:

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

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

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

2 Methods and experiment setup

2.1 Description of WaterGAP 2.2

The global hydrology and water use model WaterGAP (Fig. 1) consists of two major parts: the water use models for five different sectors (Appendix C) and the WaterGAP Global Hydrology Model (WGHM, Fig. A1). The submodel GroundWater Surface Water USE (GWSWUSE) (Appendix D) is used to distinguish water use from groundwater and surface water sources and computes net abstractions from both sources which are an input to WGHM (Fig. 1). Using a number of water storage equations (change of storage over time equals to inflow minus outflows; Appendix A), WGHM calculates daily water flows and storages at a spatial resolution of 0.5° by 0.5° (55 km by 55 km at the Equator) for the whole land area of Earth except Antarctica (66 896 cells). WaterGAP 2.2 is calibrated against mean annual river discharge at 1319 gauging stations, and the adjusted calibration factor is regionalized to grid cells outside the calibration basins (Appendix B).
Since the initial publication of WaterGAP 2.1d (Döll et al., 2003), major changes were made to keep the model up to date. For example, algorithms of reservoir operation were included (Döll et al., 2009), groundwater recharge was optimized by distinguishing semiarid/arid regions from humid regions (Döll and Fiedler, 2008), a variable flow velocity algorithm was included (Verzano et al., 2012) and the source of abstracted water was considered (Döll et al., 2012).

2.2 Experiment setup

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

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

2.2.1 Climate input

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

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

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

2.2.2 Land cover input data

The distribution of land cover classes and associated attributes are affecting simulated fluxes in terms of radiation energy balance (albedo and emissivity), snow dynamics (degree-day factor $D_p$), available soil water capacity (rooting depth) and interception capacity ($L$) (for details see Appendix A). To estimate the effect of different, homogeneous-source land cover data, two input maps were used (Fig. 2). Attributes and model parameters associated to land cover classes were derived from literature or previous model versions (Table A2) and left equal in both variants.
Table 1. Overview of the model variants.

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

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

In variant STANDARD we used the gridded MODIS (Moderate Resolution Imaging Spectroradiometer) land cover product (MOD12Q1) for the year 2004. The product MOD12Q1 (1 km resolution, global coverage up to 80° N) was used with land cover type 1 according to the International Geosphere-Biosphere Programme (IGBP) classification. After resampling to 0.5° spatial resolution, the data set was reclassified to fit to the WaterGAP land cover classification system (Table A2). As water bodies (from the global lakes and wetlands database, GLWD; Lehner and Döll, 2004) and percentage of urban area (from previous model versions) are obtained by additional input files, the second land cover class was appointed in case of “water” or “urban and built-up” as primary land cover. For coastal grid cells, which are not fully covered by MODIS and north of 80° N, a combination of Global Land Cover Characteristics database GLCC (USGS, 2008) + CORINE (Coordination of Information on the Environment) land cover information was used.
In variant LANDCOVER, a combination of the GLCC based on the years 1992 and 1993 and, for Europe, CORINE land cover based on the year 2000 (European Environment Agency, 2004) was used as land cover information, as also in a previous WaterGAP version (Haddeland et al., 2011). The idea was to use an IGBP-based classification scheme and a remote-sensing-based land cover distribution instead of IMAGE (Integrated Model to Assess the Global Environment; Alcamo et al., 1998) model outputs (as in previous model versions). Both input data sets have a resolution of 1 by 1 km and were aggregated to the 0.5° model resolution by assigning the majority land cover type.

### 2.2.3 Structural model changes

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

- Flow velocity is globally set to 1 m s\(^{-1}\) and the meandering ratio is set to 1.0, instead of the variable flow velocity algorithm of Verzano et al. (2012) in the other variants.

- Reservoirs are treated as global lakes, i.e., the reservoir operation algorithm of Döll et al. (2009) is not used, which should result in a more dynamic discharge downstream of reservoirs.

- Water for human water use is abstracted only from surface water bodies; i.e., there are no groundwater abstractions as introduced by Döll et al. (2012).

- Evaporation from lakes/wetlands is not adjusted by reduction factors (Hunger and Döll, 2008) resulting in evaporation at potential rate even at low storage.

- Snow accumulation and melt are modeled on 0.5° (instead of the 3 arcmin subgrid (Schulze and Döll, 2004)) which should lead to less snow dynamics.

- Finally, there is no distinction in groundwater recharge for semiarid/arid and humid regions (in contrast to Döll and Fiedler, 2008, all regions are treated like humid regions) resulting in higher groundwater recharge in semiarid/arid regions.

### 2.2.4 Human water use

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

### 2.2.5 Calibration

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

### 3 Results

#### 3.1 Global water balance

Table 2 lists global values for various components of the global water balance and changes in total water storages (TWS) (calculated excluding Antarctica, Greenland and inland sinks) as estimated by the different model variants. Global values vary mainly due to calibration and selected climate forcing. For interpreting Table 2 and Fig. 3 it is important to mention that AET does not include additional evapotranspiration caused by irrigation and other human water use. This part of evapotranspiration is called actual water consumption (WC\(_a\)). For computing global values of AET and renewable water resources (RWR), the values were adjusted in calibration basins using the station correction factor (CFS) such that a closed global water balance is achieved (for calibration details see Appendix B). Grid cell values of AET and RWR (Figs. 3, 4), however, do not reflect CFS to avoid physically implausible values that likely result from inconsistencies between precipitation data and observed river discharge. Global precipitation \(P\) is about 1900 km\(^2\) 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 \(E_c\), \(E_m\), \(E_a\) and \(E_w\), see Appendix A) does not vary considerably among the variants. In general, discharge to oceans and inland sinks is lower by the amount of change in AET. WC\(_a\) (row 4 in Table 2) varies due to the demand of surface
Table 2. Long-term-average (1971–2000) freshwater fluxes from global land area (except Antarctica and Greenland) of WaterGAP 2.2 (in km³ yr⁻¹). Cells representing inland sinks were excluded but discharge into inland sinks was included.

<table>
<thead>
<tr>
<th>No.</th>
<th>Component</th>
<th>STANDARD</th>
<th>NoUse h</th>
<th>CLIMATE</th>
<th>LANDCOVER</th>
<th>STRUCTURE</th>
<th>NoCal</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Precipitation Pᵃ</td>
<td>111 070</td>
<td>111 070</td>
<td>112 969</td>
<td>111 070</td>
<td>111 070</td>
<td>111 070</td>
</tr>
<tr>
<td>2</td>
<td>Actual evapotranspiration AETᵇ</td>
<td>69 803</td>
<td>69 934</td>
<td>69 842</td>
<td>70 012</td>
<td>70 217</td>
<td>63 344</td>
</tr>
<tr>
<td>3</td>
<td>Discharge into oceans and inland sinks Qᶜ</td>
<td>40 458</td>
<td>41 216</td>
<td>42 364</td>
<td>40 250</td>
<td>40 002</td>
<td>46 822</td>
</tr>
<tr>
<td>4</td>
<td>Water consumption (actual) (rows 5 + 7) WCᵃ</td>
<td>1031</td>
<td>0</td>
<td>927</td>
<td>1029</td>
<td>983</td>
<td>1054</td>
</tr>
<tr>
<td>5</td>
<td>Net abstraction from surface water (actual)ᵈ</td>
<td>1102</td>
<td>0</td>
<td>960</td>
<td>1102</td>
<td>983</td>
<td>1126</td>
</tr>
<tr>
<td>6</td>
<td>Net abstraction from surface water (demand) NAˢ</td>
<td>1154</td>
<td>0</td>
<td>1000</td>
<td>1154</td>
<td>1082</td>
<td>1154</td>
</tr>
<tr>
<td>7</td>
<td>Net abstraction from groundwater NAᶠ</td>
<td>−72</td>
<td>0</td>
<td>−33</td>
<td>−72</td>
<td>0</td>
<td>−72</td>
</tr>
<tr>
<td>8</td>
<td>Change of total water storage dS/dtᵍ</td>
<td>−215</td>
<td>−73</td>
<td>−156</td>
<td>−214</td>
<td>−44</td>
<td>−143</td>
</tr>
<tr>
<td>9</td>
<td>Long-term-averaged yearly volume balance error (P − AET − Q − WCᵃ − dS/dt) deviation to P</td>
<td>−0.006 %</td>
<td>−0.006 %</td>
<td>−0.007 %</td>
<td>−0.006 %</td>
<td>−0.08 %</td>
<td>−0.006 %</td>
</tr>
</tbody>
</table>

ᵃ Mean annual P (1979–2001) is 110 309 km³ yr⁻¹ in WFD and 110 812 km³ yr⁻¹ in WFDEI;ᵇ AET does not include evapotranspiration caused by human water use, i.e., actual water consumption WCᵃ;;c including anthropogenic water use (except NoUse);ᵈ if not enough water is available, demand is not completely satisfied;ᵉ demand that needs to be satisfied (water use model output);ᶠ negative values indicate that return flows from irrigation with surface water exceed groundwater abstractions;ᵍ TWS of 31 December 2000 minus TWS of 31 December 1970 divided by 30 years;ʰ STANDARD but no subtraction of water use; i.e., discharge into oceans and inland sinks equals renewable water resources.

Figure 3. Actual evapotranspiration AET for STANDARD (mean value 1971–2000, in mm yr⁻¹) (a) and differences between the model variants and STANDARD (in mm yr⁻¹) (b–f).
Figure 4. Renewable water resources (mean annual runoff from each cell if water use is neglected) calculated by the WaterGAP 2.2 NoUse variant (a) and differences to other variants (variants here run without considering water use) (b–e).

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

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

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

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

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

The decreasing trends of total water storage are mainly caused by groundwater depletion, except in variants NoUse.
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 $\gamma$ which is on average lower in case of NoCal. The lower $\gamma$, the more water leaves the soil and can subsequently contribute to groundwater recharge. Note that water abstractions from groundwater are taken directly from the groundwater storage and also return flows are added directly to groundwater storage (without passing the soil compartment). Hence, there is no difference in soil water storage between STANDARD and NoUse (Table 3).

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

### 3.2 Actual evapotranspiration

Mean AET shows the highest values around the Equator, consistent with available energy, except for the Pacific Rim of South America (Fig. 3a). Among the variants, the largest differences to STANDARD occur in the case of the uncalibrated version NoCal (Fig. 3f). As the calibration approach also affects grid cells outside of the 1319 calibration basins due to the regionalization (Appendix B3), all grid cells are affected. In most regions, calibration leads to higher AET but in the upstream Amazon, Congo, Arctic river basins and some other basins, the opposite is true. The global sum of AET of NoCal is 9.2 % lower than estimated with STANDARD (Table 2). Notable differences in AET also occur when using an alternative climate input (Fig. 3b). AET increases in CLIMATE on 42.6 % of the land surface by more than 10 mm yr$^{-1}$ and decreases by more than 10 mm yr$^{-1}$ on 30.5 % of the land surface. It increases (decreases) by more than 100 mm yr$^{-1}$ on 5.4 % (5.6 %) of the land surface. When summed globally, only minor changes in AET occur in the CLIMATE variant (increase of 0.06 % or 39 km$^3$ yr$^{-1}$; Table 2). In contrast, AET differences of the STRUCTURE variant are higher for the global sum (increase of 0.6 % or 414 km$^3$ yr$^{-1}$) but occur on an overall smaller area (increase by more than 10 mm yr$^{-1}$ on 11.9 % of the land surface, decrease on 14.2 %). The effect of STRUCTURE is visible in areas with surface water bodies and in snow-dominated areas. On the one hand, an increase in net radiation in snowy regions leads to a slight increase of AET but in small absolute numbers as total AET is comparatively low. On the other hand, effects due to the evaporation reduction factor for surface water bodies are visible. In all variants, except STRUCTURE, evaporation is limited when the surface water body storage is reduced to mimic the shrinking of surface area. Hence, in regions with a high percentage/volume of surface water bodies, AET is increased. In addition, more complex
As RWR are approximately the difference between precipitation and AET, the difference maps (Fig. 4b–e) represent more or less the inverted difference maps in Fig. 3. Compared to STANDARD, the largest differences occur in model variant NoCal. In contrast to AET, calibration leads in many cases to lower RWR. The global sum of RWR of NoCal is 15.8 % higher than with STANDARD. The global sum of RWR from CLIMATE is 4.7 % higher but with a large spatial spread. RWR decrease in CLIMATE on 21.4 % of land surface by more than 10 mm yr$^{-1}$ and increase by more than 10 mm yr$^{-1}$ on 29.9 % of the land surface. RWR decrease (increase) by more than 100 mm yr$^{-1}$ on 4.7 % (9.0 %) of the land surface. The differences in LANDCOVER mainly follow differences in net radiation (not shown). In snow-dominated regions, RWR are lower in STRUCTURE because snow-cover dynamics are less intense than in STANDARD. In grid cells with (large) surface water bodies, RWR are lower in STRUCTURE (as AET is unlimited here even if storages are nearly empty).

### 3.4 River discharge

#### 3.4.1 River discharge seasonality

River discharge is the integral result of runoff generation, water losses by evaporation from surface water bodies, positive or negative net abstractions from surface water bodies and groundwater, and routing processes. It is one of the most important diagnostic variables in water resources. In many different climatic zones and levels of anthropogenic influence (Fig. 5). Climate input and model structure influence modeled discharge seasonality more than land cover changes for the selected river basins. Where seasonality of climate is high, like in the monsoon-dominated Mekong Basin, only marginal differences occur due to land cover and model structure. Structural model refinements have also important effects on discharge seasonality. For example, the constant flow velocity of STRUCTURE (in contrast to variable flow velocity in the other variants) leads to a higher peak in the Lena. Here, the variable flow velocity algorithm underestimates flow velocity in the lower reaches where bed slopes are very small. This leads to a strong underestimation of peak flow (which explains the improved seasonality of STRUCTURE compared to observed discharge in the Lena). The reservoir algorithm which is not enabled in STRUCTURE has impacts at the Yangtze, Rio Parana, Mississippi and the Volga rivers in terms of smoothing the discharge. For the Rio Parana, this is the main influence in the STRUCTURE variant. The representation of snow in STANDARD leads
to a more heterogeneous snow coverage as compared to the STRUCTURE variant. The strongest impact occurs for the Rhine, where the snow algorithm is the dominant reason for the differences to STRUCTURE. In STRUCTURE, the snow water storage of the Rhine headwater (Alps) is generally lower. In particular between May and October (the Alps are modeled as snow-free between June and September), this leads to a decrease of discharge as snowmelt cannot contribute any longer as it does, for example, in STANDARD. The importance of the climate forcing can be seen

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

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

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

### 3.4.2 Monthly time series

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

\[
E_{NS} = 1.0 - \frac{\sum_{i=1}^{n} (O_i - S_i)^2}{\sum_{i=1}^{n} (O_i - \bar{O})^2},
\]

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

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

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

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

### 3.5 Variations of total water storages

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

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

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

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

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

4 Discussion

4.1 Comparison of simulated freshwater fluxes to other estimates

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

4.2 Advantages and limitations of the calibration approach

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

www.hydrol-earth-syst-sci.net/18/3511/2014/
Table 5. Comparison of diverse estimates of global actual evapotranspiration and discharge (in km$^3$ yr$^{-1}$).

<table>
<thead>
<tr>
<th>Actual evapotranspiration</th>
<th>Discharge</th>
</tr>
</thead>
<tbody>
<tr>
<td>62 800 Mu et al. (2011)</td>
<td>34 406 Mueller et al. (2013)</td>
</tr>
<tr>
<td>64 512$^a$ Mueller et al. (2013)</td>
<td>36 200 Wada et al. (2010)</td>
</tr>
<tr>
<td>65 000 Jung et al. (2010)</td>
<td>36 687 Döll et al. (2003)</td>
</tr>
<tr>
<td>66 000 Sterling et al. (2012)</td>
<td>38 587 Baumgartner and Reichenel (1975)</td>
</tr>
<tr>
<td>71 000 Baumgartner and Reichenel (1975)</td>
<td>38 605 Widén-Nilsson et al. (2007)</td>
</tr>
<tr>
<td>72 000 Korzun (1978)</td>
<td>39 307 Fekete et al. (2002)</td>
</tr>
<tr>
<td>75 981$^b$ Mueller et al. (2011)</td>
<td>39 414 Döll and Fiedler (2008)</td>
</tr>
<tr>
<td>60 000–85 000 Haddeland et al. (2011)</td>
<td>44 560 Korzun (1978)</td>
</tr>
<tr>
<td>70 576$^c$ STANDARD</td>
<td>45 500 Oki and Kanae (2006)</td>
</tr>
<tr>
<td>45 000–66 000 Haddeland et al. (2011)</td>
<td>42 000–66 000 Haddeland et al. (2011)</td>
</tr>
<tr>
<td>40 458 STANDARD</td>
<td>40 458 STANDARD</td>
</tr>
</tbody>
</table>

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

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

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

However, the correction of total cell runoff using CFA and CFS that is required to force simulated mean annual river discharge values to be equal to observed values is not ideal and has undesirable effects on estimated AET and RWR. In the Yenisey Basin, upstream of Igarka (western Siberian Plain), AET is largely reduced in one-half of the basin (and vice versa) when using alternative climate forcing. Transferring the correction factor CFS (which is, if necessary, calculated
at the outflow grid cell of the basin) to the upstream grid cells can lead to unrealistic high positive and negative values for AET if precipitation is too low in these parts of the basin to simulate observed discharge or if the AET of surface water bodies has to be reduced by CFA. This is the reason for some artificial patterns in Fig. 3 and consequently in Fig. 4. These kinds of consistency errors can be found in some more basins where cumulative AET is low and parts of the basins are covered with surface water bodies. Nevertheless, the approach ensures a closed water balance for the whole basin.

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

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

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

In regions where the climate forcing data sets differ significantly (e.g., Danube River basin), the impact on discharge is large (Fig. 5, bottom center panel). Here, differences in temperature and precipitation amounts lead to a poor fit compared to observed discharge when using the CLIMATE variant which is also reflected in the $E_{NS}$ criterion (Fig. 9b).
Figure 9. Spatial distribution of Nash–Sutcliffe Efficiency ($E_{NS}$) classes (from Table 4, 1: $E_{NS} > 0.7$, 2: $0.5 < E_{NS} < 0.7$, 3: $E_{NS} < 0.5$) for STANDARD (a), and differences of model variants (calculated as STANDARD $E_{NS}$ class minus that of the model variant) (b–f). Red colors indicate a decrease, green an increasing $E_{NS}$ when using the model variant compared to STANDARD.

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

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

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

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

The calibration procedure reduces simulated river discharge and water resources on most of the land area and increases the AET (Figs. 3, 4; Table 2). Without calibration, global AET and discharge would rank at the lower and higher end of the published values, respectively (Table 5). In addition, the fit to observed monthly river discharge time series as quantified using the $E_{NS}$ criterion would worsen almost
everywhere (Fig. 9f). The impact of calibration on freshwater fluxes and water storages is higher than those of alternative climate forcings and land cover data, and of a more sophisticated model structure. This confirms the strong benefit of calibration. However, as $E_{NS}$ is affected by mean discharge as well as discharge variations, the calibration approach improves this criterion.

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

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

The answer to this question depends on the type of fluxes and the spatial aggregation. Regarding selected global sums of freshwater fluxes ($Q$ and AET) and mean annual total water storage trends $d$TWS/$dt$ (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.

<table>
<thead>
<tr>
<th>Variable</th>
<th>STANDARD</th>
<th>Rank 1</th>
<th>dSTA</th>
<th>Rank 2</th>
<th>dSTA</th>
<th>Rank 3</th>
<th>dSTA</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Q$</td>
<td>69 803</td>
<td>NoCal</td>
<td>848</td>
<td>NoCal</td>
<td>862</td>
<td>NoCal</td>
<td>870</td>
</tr>
<tr>
<td>AET</td>
<td>60 458</td>
<td>NoCal</td>
<td>6459</td>
<td>STRUCTURE</td>
<td>414</td>
<td>LANDCOVER</td>
<td>209</td>
</tr>
<tr>
<td>$d$TWS/$dt$</td>
<td>−214</td>
<td>STRUCTURE</td>
<td>169</td>
<td>NoCal</td>
<td>140</td>
<td>NoCal</td>
<td>71</td>
</tr>
<tr>
<td>Median $E_{NS}$</td>
<td>0.54</td>
<td>NoCal</td>
<td>0.66</td>
<td>STRUCTURE</td>
<td>−0.05</td>
<td>CLIMA TE</td>
<td>−0.03</td>
</tr>
</tbody>
</table>

### 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.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Variant</th>
<th>Percentage of area affected by changes above 10th percentile</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>NoCal</td>
<td>60.5 13.5</td>
</tr>
<tr>
<td>2</td>
<td>CLIMA TE</td>
<td>45.5 3.2</td>
</tr>
<tr>
<td>3</td>
<td>LANDCOVER</td>
<td>24.2 1.2</td>
</tr>
<tr>
<td>4</td>
<td>STRUCTURE</td>
<td>13.6 1.1</td>
</tr>
<tr>
<td>5</td>
<td>NoUse</td>
<td>0.9 0.03</td>
</tr>
</tbody>
</table>

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, 4). The fraction of the global land area that is affected by significant differences of AET and river discharge between a certain model variant and the STANDARD variant is largest in the case of NoCal, followed by CLIMA TE, 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

*www.hydrol-earth-syst-sci.net/18/3511/2014/*

Hydrol. Earth Syst. Sci., 18, 3511–3538, 2014
uncertainty which is a major additional source of uncertainty (Kaspar, 2003).

5 Conclusions

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

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

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

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

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

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

A1 Canopy

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

\[
\frac{dS_c}{dt} = P - P_t - E_c, \quad (A1)
\]

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

\[
P_t = \begin{cases} \frac{P}{S_c \geq S_{c,\text{max}}} \\ 0 \quad S_c < S_{c,\text{max}} \end{cases}, \quad (A2)
\]

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

\[
E_c = E_p \left( \frac{S_c}{S_{c,\text{max}}} \right)^{\frac{3}{2}}, \quad (A3)
\]

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

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

Appendix B provides information on the calibration and regionalization approach WaterGAP is based on. Appendix C gives a brief introduction of the water use submodels, and the GWSWUSE module is described in Appendix D.

A2 Snow

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

\[
\frac{dS_{sn}}{dt} = P_{sn} - M - E_{sn}, \quad (A6)
\]

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

Snow accumulation and melt are modeled on a 3 arcmin subgrid (100 subgrid cells per 0.5°) using a degree-day algorithm (Schulze and Döll, 2004). Mean subgrid elevation was derived from GTOP030 (US Geological Survey, 2003). The daily temperature for each subgrid cell is calculated from the temperature of the 0.5° cell, applying an adiabatic lapse rate of 0.6°C per 100 m. To avoid excessive snow accumulation, temperature does not decrease if a snow water equivalent of 1000 mm is reached in one subgrid.

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

\[
M = \begin{cases} \frac{D_F (T - T_m)}{T > T_m}, \quad S_{sn} > 0 \\ 0 \quad \text{other} \end{cases}. \quad (A7)
\]

Instead of using one specific albedo for snow as in previous versions (\( \alpha = 0.4 \)), land cover specific snow albedo values are used to account for differences in reflective properties between the land use classes under snow-covered conditions (Table A2). The albedo value switches to snow albedo if the snow water equivalent of the grid cell exceeds 3 mm, i.e., a closed snow cover is assumed. Sublimation \( E_{sn} \) is modeled like potential evaporation rate but applying a latent heat of 2.835 (MJ kg\(^{-1}\)) for temperatures below 0°C and 2.501–0.002361 × \( T \) (MJ kg\(^{-1}\)) above 0°C.
A3 Soil

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

$$\frac{dS_s}{dt} = P_{\text{eff}} - R_l - E_s,$$  \hspace{1cm} (A8)

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

$$P_{\text{eff}} = P_t - P_{\text{sn}} + M,$$  \hspace{1cm} (A9)

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

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

$$E_s = \min[(E_p - E_c), (E_{p,\text{max}} - E_c) \frac{S_s}{S_{s,\text{max}}}],$$  \hspace{1cm} (A10)

where $E_{p,\text{max}}$ is 20 mm d$^{-1}$ in semiarid and arid regions whereas 10 mm d$^{-1}$ in grid cells classified as humid, $S_{s,\text{max}}$ 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

$$R_l = P_{\text{eff}} \left(\frac{S_s}{S_{s,\text{max}}}\right)^\gamma.$$  \hspace{1cm} (A11)

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

A4 Groundwater

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

$$\frac{dS_g}{dt} = R_g - Q_g - NA_g$$  \hspace{1cm} (A12)

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

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

where $R_{g,\text{max}}$ 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, $R_{g,\text{max}}$ and the precipitation threshold, are adapted to the climate forcing used (WFD) aiming to reach comparable groundwater recharge patterns to Döll and Fiedler (2008) as that groundwater recharge estimation is confirmed by experts within the WHYMAP (World-wide Hydrogeological Mapping and Assessment Programme: http://www.whymap.org) efforts. Within CLIMATE, the original values 5/3/1.5 for $R_{g,\text{max}}$ and 10 mm d$^{-1}$ as precipitation threshold were used.

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

$$Q_g = k_g S_g.$$  \hspace{1cm} (A13)

The runoff from land $R_l$, which is not groundwater recharge $R_g$, represents the fast surface runoff $R_s$ and is routed, together with $Q_g$, through a series of different storages representing wetlands, lakes and reservoirs until reaching the river segment (Fig. A1).
A5 Surface water bodies

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

\[
\frac{dS}{dt} = I - Q + B - N_{A_s}. \quad (A14)
\]

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

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

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

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

The distribution of wetlands is derived from GLWD (Lehner and Döll, 2004) as percentage of cell coverage. Locations and attributes of lakes and reservoirs are based on a combination of GLWD and a preliminary version of the GRanD (Global Reservoir and Dam) database (Döll et al., 2009; Lehner et al., 2011). In total, 6553 reservoirs, 52 regulated lakes (lakes whose outflow is regulated by a dam) (from GRanD) and 242 798 unregulated lakes (from GLWD) were considered. Out of these, 1386 large lakes (area \( \geq 100 \text{ km}^2 \)), 1110 large reservoirs (storage capacity \( \geq 0.5 \text{ km}^2 \)) and 52 regulated lakes (area \( \geq 100 \text{ km}^2 \) or storage capacity \( \geq 0.5 \text{ km}^2 \)) 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 \( R_I = R_l - R_g \) is routed to the surface storages without any delay, whereas baseflow \( Q_g \) is a function of groundwater storage (Fig. A1, Appendix A4). Due to limited information on groundwater flow between grid cells, the groundwater recharge can only contribute to groundwater runoff of the same grid cell.

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

The reservoir algorithm of Hanasaki et al. (2006), distinguishing irrigation and nonirrigation reservoirs and considering 1109 reservoirs was implemented and improved by Döll et al. (2009) and slightly adapted in WaterGAP 2.2: if reservoir storage falls below 10% of storage capacity, the release coefficient is set to 0.1 instead of 0.0 as in Döll et al. (2009), assuring that at least some water is released, e.g., for downstream ecosystem demands.
Table A1. Parameters of the leaf area index model.

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

$^a$ $L_{\text{max}}$ 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.

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

$^a$ Adapted from the IMAGE model (Alcamo et al., 1998); $^b$ Wilber et al. (1999); $^c$ Maniak (1997), WMO (1994).
Appendix B: Calibration and regionalization

B1 Calibration approach

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

B2 Discharge stations used

Observed discharge time series were provided by the Global Runoff Data Center (GRDC). Following Hunger and Döll (2008), gauging stations listed in the GRDC catalogue (http://grdc.bafg.de/; download date: 28 September 2012) were included in the calibration setup if they fulfilled three main criteria: (1) an upstream area of at least 9000 km$^2$, (2) a time series of at least 4 (complete) years, and (3) an interstation catchment area of at least 30000 km$^2$. All in all, a number of 1319 stations, covering 53.6% of the global land area except Antarctica and Greenland, was used for calibration (Fig. B1). If available, the 30-year period 1971–2000 was chosen as calibration years.

B3 Regionalization

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

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

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

Appendix D: Description of GWSWUSE

In the water use models, the source of the abstracted water is not distinguished. This is done in the WaterGAP submodel GWSWUSE (Döll et al., 2012). Based on the results of the water use models, GWSWUSE computes net abstractions (abstractions minus return flows) from groundwater and net abstraction from surface water bodies that serve as input to WGHM (Fig. 1). As a first step within GWSWUSE, the time series of consumptive water use in irrigation, which is computed by GIM for temporally constant irrigation areas but changing climate variables, is scaled by using an annual time series of irrigated area by country uses information (Döll et al., 2012). Then, groundwater use fractions for irrigation (Siebert et al., 2010), domestic and manufacturing water use are applied, and irrigation water abstractions are determined by dividing consumptive use by irrigation water use efficiencies. In contrast to Döll et al. (2012), irrigation water use efficiencies differ between surface water and groundwater use in WaterGAP 2.2. While for surface water irrigation country-specific values are still used, irrigation water use efficiency was set to 0.7 worldwide in case of groundwater irrigation (Döll et al., 2014a). Return flows from irrigation to either groundwater or surface water are computed as a function of the cell-specific artificial drainage fraction (Döll et al., 2012). In WaterGAP 2.2, the fraction of irrigation return flows that recharge groundwater was increased as compared to Döll et al. (2012) and is computed as 0.95–0.75 times the cell-specific artificial drainage fraction. Due to return flows, net abstractions can be positive (water is abstracted from storage) or negative (water is added to storage) (see Fig. 1 of Döll et al., 2014a).
Acknowledgements. The authors thank the Global Runoff Data Center, 56068 Koblenz, Germany, for providing the discharge data used in this study. The authors acknowledge the comments and helpful suggestions of two anonymous referees and the editor that helped to improve the manuscript significantly.

Edited by: K. Stahl

References


H. Müller Schmied et al.: Sensitivity of simulated global-scale freshwater fluxes and storages 3535

H. Müller Schmied et al.: Sensitivity of simulated global-scale freshwater fluxes and storages


