Earth Syst. Sci. Data, 9, 389–413, 2017
https://doi.org/10.5194/essd-9-389-2017

Open Access Earth System Science Data

A global water resources ensemble of hydrological models: the eartH2Observe Tier-1 dataset

Jaap Schellekens[1], Emanuel Dutra[2,a], Alberto Martínez-de la Torre[3], Gianpaolo Balsamo[2], Albert van Dijk[4], Frederiek Sperna Weiland[1], Marie Minvielle[5], Jean-Christophe Calvet[5], Bertrand Decharme[5], Stephanie Eisner[6], Gabriel Fink[6], Martina Flörke[6], Stefanie Peßenteiner[7], Rens van Beek[7], Jan Polcher[8], Hylke Beck[9,b], René Orth[10], Ben Calton[11], Sophia Burke[12], Wouter Dorigo[13], and Graham P. Weedon[14]

[1]Deltares, Rotterdamseweg 185, 2629 HD Delft, the Netherlands
[2]European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading RG2 9AX, UK
[3]Centre for Ecology and Hydrology, Wallingford, Oxfordshire, OX10 8BB, UK
[4]Fenner School of Environment and Society, Australian National University, Canberra, ACT 0200, Australia
[5]CNRM/GAME, Météo-France, CNRS, UMR 3589, 42 avenue Coriolis, 31057 Toulouse CEDEX 1, France
[6]Center for Environmental Systems Research (CESR), University of Kassel, Wilhelmshöher Allee 47, 34117 Kassel, Germany
[7]Department of Physical Geography, Utrecht University, Heidelberglaan 2, 3584 CS Utrecht, the Netherlands
[8]Laboratoire de Météorologie Dynamique (LMD, CNRS), Ecole Polytechnique, 91128 Palaiseau, France
[9]European Commission, Institute for Environment and Sustainability, Joint Research Centre, Ispra, VA, Italy,
[10]Institute for Atmospheric and Climate Science, ETH Zurich, Universitätstrasse 16, 8092 Zurich, Switzerland
[11]PML Applications Ltd, Prospect Place, The Hoe, Plymouth, UK
[12]AmbioTEK Community Interest Company, Essex, UK
[13]Department of Geodesy and Geoinformation, Vienna University of Technology, Vienna, Austria
[14]Met Office, Joint Centre for Hydrometeorological Research, Maclean Building, Benson Lane, Crowmarsh Gifford, Wallingford, OX10 8BB, UK
[a]now at: Instituto Dom Luiz, Faculdade de Ciências, Universidade de Lisboa, 1749-016 Lisbon, Portugal
[b]now at: Princeton University, Civil and Environmental Engineering, Princeton, NJ, USA

Correspondence to: Jaap Schellekens (jaap.schellekens@deltares.nl)

Received: 8 November 2016 – Discussion started: 6 December 2016
Revised: 17 May 2017 – Accepted: 25 May 2017 – Published: 3 July 2017

Abstract. The dataset presented here consists of an ensemble of 10 global hydrological and land surface models for the period 1979–2012 using a reanalysis-based meteorological forcing dataset (0.5° resolution). The current dataset serves as a state of the art in current global hydrological modelling and as a benchmark for further improvements in the coming years. A signal-to-noise ratio analysis revealed low inter-model agreement over (i) snow-dominated regions and (ii) tropical rainforest and monsoon areas. The large uncertainty of precipitation in the tropics is not reflected in the ensemble runoff. Verification of the results against benchmark datasets for evapotranspiration, snow cover, snow water equivalent, soil moisture anomaly and total water storage anomaly using the tools from The International Land Model Benchmarking Project (ILAMB) showed overall useful model performance, while the ensemble mean generally outperformed the single model estimates. The results also show that there is currently no single best model for all variables and that model performance is spatially variable. In our unconstrained model runs the ensemble mean of total runoff into the ocean was $46\,268\,\mathrm{km}^3\,\mathrm{yr}^{-1}$ ($334\,\mathrm{kg\,m}^{-2}\,\mathrm{yr}^{-1}$), while the ensemble mean of total evaporation was $537\,\mathrm{kg\,m}^{-2}\,\mathrm{yr}^{-1}$. All data are made available openly through a Water Cycle Integrator portal (WCI, wci.earth2observe.eu), and via a direct http and ftp download. The portal follows the protocols of the open geospatial consortium such as OPeNDAP, WCS and WMS. The DOI for the data is https://doi.org/10.5281/zenodo.167070.

Published by Copernicus Publications.

1 Introduction

Water security concerns all global economies, rich and poor (Collins et al., 2009; McDonald et al., 2011; Hansen et al., 2012). At the same time water availability in many areas is decreasing due to demographic and climatic changes (Faures, 2006) which can influence agriculture (Rijsberman, 2006) but also industry and energy through its influence on cooling water and hydropower (Van Vliet et al., 2011; van Vliet et al., 2016). This combination stresses the need for a holistic (integrated) approach to water resources management and decision making for flood protection, food and water security, energy and large-scale (re)forestation. Such an approach requires the integration of information on water availability, demand and quality at all scales, and must be supported by an improved assessment of water resources and predictive understanding of the water and energy cycles (UN, 2016). Yet, the availability of this information is lacking in many regions of the world (Pozzi et al., 2013). To capture the uncertainty that stems from the simplifications and assumptions in our models, a global reanalysis dataset of water resources should contain a large number of global hydrological and land surface models that assimilate the most important satellite-based products (Sood and Smakhtin, 2015). This should be available as a reference for local studies and for the support of policy and decision making for transboundary watersheds and for global applications such as flood risk analysis (Trigg et al., 2016) for which the global land surface and hydrological models are the key components.

Only a limited number of global reanalysis datasets that can support water resources analysis is available. Pioneered by GLDAS, (Houser et al., 2001; Rodell et al., 2004) several other systems followed (e.g. GSWP-2 Dirmeyer et al., 2006, MERRA-Land Reichle (2012), ERA-Land Balsamo et al. (2015) but also van Dijk et al. (2014) and WATCH, Haddeland et al. (2011)). In these, Haddeland et al. (2011) combines both global land surface models and global hydrological models into a single multi-model ensemble of which some are also used in this study. The WATCH programme (Harding et al., 2011) used the WATCH Forcing Data which, for January 1958 to December 2001, was created by bias correcting ERA40 reanalysis data (Uppala et al., 2005) using gridded in situ meteorological observations (Weedon et al., 2011). For January 1901 to December 1957 the WATCH Forcing Data applied the same system of bias correction, but applied to randomly selected years of the ERA40 1958–2001 data (Weedon et al., 2011). During the EMBRACE programme the WATCH Forcing Data (1958–2001) methodology was applied to the more recent ERA-Interim reanalysis (Dee et al., 2011) to create the WFDEI (WATCH Forcing Data methodology applied to ERA-Interim reanalysis; Weedon et al., 2014).

We use the WFDEI dataset to force a set of 10 global models, both land surface models (LSMs) and global hydrological models (GHMs). By using a sizeable set of models we take steps to mitigate some of the errors and uncertainties that are introduced in individual models by the simplified representation of spatially heterogeneous real world processes like water and energy balances, river routing and seasonal varying vegetation cover (Beven and Binley, 1992; Vrugt et al., 2005; Gosling et al., 2010). As a general principle, this is superior to the results of any individual model and as good as or better than the best model at each point and time (Dirmeyer et al., 2006; Harding et al., 2011). However, this does not necessarily mean that this is the case for the set of models that we use or that some models do not perform considerably better in specific locations, climatic conditions or for specific variables (e.g. runoff) than others.

The multi-model ensemble presented here inherits a number of models from the WATCH project supplemented by additional models, a new forcing dataset (WFDEI), a WFDEI-derived reference potential evapotranspiration dataset and a new modelling protocol. Furthermore, we introduce the data repository where the results are stored in an open format including all data needed for other groups to perform a similar exercise. In the end this can contribute to a better understanding of the characteristics of the increasing number of global models (Bierkens, 2015). The repository comes with (downscaling) tools and river basin management models such as WaterWorld (Mulligan, 2012) to increase usage of the data at the basin scale outside of the research community.

In this paper we present the first version of the dataset, which is based on the current state of the art of the contributing modelling systems and will provide a benchmark to evaluate improvements made to the models and forcing data in the coming years. The main goal of this paper is to provide a multi-decadal dataset of water balance components from an ensemble of models that is open and of use for further research and applications. Secondly, we investigate whether the ensemble mean in this dataset is superior to the individual models given the diverse set of models, and if so, for which variables.

First, we describe the methods and models we have used. Secondly, we investigate the characteristics of the resulting dataset using the multi-model signal-to-noise ratio (SNR) to investigate multi-model agreement and the tools from The International Land Model Benchmarking Project (ILAMB, Luo et al., 2012; Mu et al., 2016) to compare the model output to reference datasets for evapotranspiration, total water storage, soil moisture, snow water equivalent and snow cover. Thirdly, the terrestrial water budget is used to compare the results with previous efforts. Finally we present conclusions and an outlook for further versions of this dataset.

2 Methodology and modelling protocol

Each of the models used produced results for the period 1979–2012 based on the provided meteorological forcing. In total 10 models were used, both large-scale hydrological models and land surface models with extended hydrological schemes (see the list below and Table 1), all running offline (i.e. not connected to an atmospheric model) while driven by the same reanalysis-based forcing dataset. Given the different nature of the models a single spin-up procedure was not feasible. The spin-up procedure was chosen for each model individually to match each model's requirements with the goal to best represent the climatic conditions over the simulation period.

HTESSEL-CaMa represents the Hydrology Tiled ECMWF Scheme for Surface Exchanges over Land (Balsamo et al., 2009). The runoff produced by the land surface scheme is routed with the Catchment-based Macro-scale Floodplain model CaMa-Flood (Yamazaki et al., 2011). A 10-year spin-up was carried out: an initial run from 1 January 1979 to 1 January 1989, while the land surface state of January 1989 was used to initialize the main simulation.

JULES is the Joint UK Land Environment Simulator, (Best et al., 2011; Clark et al., 2011), a community land surface model that has evolved from the Met Office Surface Exchange Scheme (MOSES). It includes an additional saturation excess runoff production using a probabilistic distributed model (Moore, 2007) approach. A 10-year spin-up was carried out: an initial run from 1 January 1979 to 1 January 1989, while the land surface state of January 1989 was used to initialize the main simulation.

LISFLOOD (Van Der Knijff et al., 2010) is a spatially distributed, grid-based rainfall–runoff and channel routing model that has been designed primarily for the simulation of the water cycle in large river basins. The model is made up of a two-layer soil water balance sub-model, sub-models for the simulation of groundwater and subsurface flow, a sub-model for the routing of surface runoff to the nearest river channel, and a sub-model for the routing of channel flow. The model was initialized by running the full 1979–2012 period before starting the main run.

ORCHIDEE is a land surface scheme resulting from the coupling of the SECHIBA land surface scheme and the carbon and vegetation model STOMATE. It consists of a hydrological module (Krinner et al., 2005) and a routing (Ngo-Duc et al., 2007) and floodplain module (d'Orgeval et al., 2008). The model was spun up with a simulation from 1 January 1979 to 31 December 1990. This simulation started with an average soil moisture and empty aquifers. After the 12 years of spin-up, river discharges have reached equilibrium.

PCR-GLOBWB (Bierkens and Van Beek, accessed 2015; van Beek et al., 2011; Wada et al., 2014) is a leaky bucket global hydrological model providing a regular grid-based representation of terrestrial hydrology. The routing is based on a computationally efficient travel time approach, where volumes of water are transported over a characteristic distance along the drainage system (Deursen, 1995). A 68-year spin-up was carried out by performing two initial back-to-back runs from 1979 to 2012 prior to the definite run.

SURFEX-TRIP uses the ISBA multi-layer land surface model to compute the soil/snow/vegetation energy and water budgets (Decharme et al., 2010, 2013) and the TRIP river routing model to simulate the river flow at the global scale. A 20-year spin-up was carried out using the 1979–1988 period two times.

SWBM (Simple Water Balance Model) is a global model that derives soil moisture, evapotranspiration (ET) and runoff from meteorological information alone, i.e. does not use any information on soil or vegetation characteristics (Orth and Seneviratne, 2013). The model parameters have been determined by calibrating the model against multiple reference datasets in Europe. These spatially uniform parameters were applied globally to derive the eartH2Observe simulations. The spin-up was done by running the first year 5 times. The resulting soil moisture and snow fields were then used to start the actual simulation.

W3RA (worldwide water resources assessment) is based on the landscape hydrology component model of the AWRA system (AWRA-L version 1.0; Van Dijk, 2010; van Dijk et al., 2014). AWRA-L can be considered a hybrid between a simplified grid-based land surface model and a nonspatial (or so-called "lumped") catchment model applied to individual grid cells (model code available at http://www.wenfo.org/wald). Spin-up was carried out using the entire 1979–2012 modelling period before the final runs.

WaterGAP3, Water – Global Assessment and Prognosis-3 is a grid-based, integrative global freshwater resources assessment tool. It consists of a spatially distributed rainfall–runoff model, five sectorial water use models, and a large-scale water quality model (Flörke et al., 2013; Döll et al., 2009). Storage compartments were initialized by re-running the model with the first year of available meteorological forcing 10 times.

HBV-SIMREG is the simple conceptual HBV hydrological model (Lindström et al., 1997) with optimized parameters derived using a novel regionalization

Table 1. Overview of models and summary of processes included.

Model	Interception	Evaporation	Snow	Soil layers	Groundwater	Runoff	Reservoirs/lakes	Routing	Water use	Time step
HTESSEL-CaMa	Single reservoir, potential evaporation	Penman–Monteith	Energy balance, 1 layer	4	No	Saturation excess	No	CaMa-Flood	No	1 h
JULES	Single reservoir, potential evaporation	Penman–Monteith	Energy balance, 3 layers	4	No	Saturation and infilt. excess	No	No	No	1 h
LISFLOOD	Single reservoir, potential evaporation	Penman–Monteith	Degree-day, 1 layer	2	Yes	Saturation and infilt. excess	Yes	Double kinematic wave	Yes	1 day
ORCHIDEE	Single reservoir structural resistance to evaporation	Bulk PET (Barella-Ortiz et al., 2013)	1 moisture layer, 1–5 thermodynamic layers	11	Yes	Green-Ampt infiltration	No	linear cascade of reservoirs (sub-grid)	irrigation only	900 s energy balance, 3 h routing
PCR-GLOBWB	Single layer, subject to open water evaporation	Hamon (tier 1) or imposed as forcing	Temperature based melt factor	2	Yes	Saturation excess	Tier 1 only lakes	Travel time approach	Not in tier 1	1 day
SURFEX-TRIP	Single reservoir, potential evaporation	Penman–Monteith	Energy and mass balance, 12 layers	14	Yes	Saturation and infilt. excess	No	TRIP with stream	No	900 s for ISBA, 3600 s for TRIP
SWBM	No	Inferred from net radiation	Degree-day, 1 layer	1	No	Inferred from precipitation and soil moisture	No	No	No	1 day
W3RA	Gash event-based model	Penman–Monteith	Degree-day, 1 layer	3	Yes	Saturation and infiltration excess	No	Cascading linear reservoirs	No	1 day
WaterGAP3	Single reservoir	Priestley–Taylor	Degree-day, 1 layer	1	Yes	Beta function	Yes	Manning–Strickler	Yes	1 day
HBV-SIMREG	No	Penman 1948	Degree-day, 1 layer	1	No	Beta function	No	No	No	1 day

scheme in which calibrated parameters are transferred to grid cells with similar characteristics to produce parameter maps with global coverage (Beck et al., 2016b). For each grid cell, we used calibrated parameters from the 10 most similar catchments and averaged the ensemble of model outputs. The model was initialized using the first 10 years of the forcing data before starting the main run.

HBV-SIMREG, SWBM, LISFLOOD and WaterGAP3 all have been calibrated in previous studies based on observed runoff data, although these calibration efforts were done with different forcing datasets (see the respective model papers listed above). The other models rely on a priori parameter estimation alone.

The data used to force the models were from the WFDEI dataset (Weedon et al., 2011, 2014) that comprised the period between and including 1979 and 2012 and contains both 3-hourly time intervals and daily time intervals. WFDEI is based on the ECMWF ERA-Interim reanalysis (Dee et al., 2011) with a spatial resolution of 0.5°, and is corrected with the CRU dataset (Harris et al., 2014b) using a sequential elevation correction of surface meteorological variables plus monthly bias correction from gridded observations. Compared to the original WFDEI dataset we applied several data formatting changes to facilitate its usage, storage and dissemination. In order to avoid land–sea mask problems with different models all oceans were filled with data from the original ERA-Interim for all variables apart from precipitation for which the ERA-Interim/Land dataset was used (Balsamo et al., 2015). Furthermore, the files were reformatted to netCDF4 and some metadata attributes have been modified to comply with the Climate and Forecast (CF) conventions. Table 2 provides an overview of the WFDEI variables used in this study.

A list of the most important output variables is presented in Table 3. If a model does not represent a process, the associated variables are not available in the dataset. The water and energy fluxes follow the mathematical convention, i.e. positive onto the surface and negative away from the surface (see details in Table 3). For example, runoff has a negative signal. Although most models can only supply a subset of the requested variables listed in Table 3 we have defined this rather large number of variables so that specific fluxes/stores from models that can supply those can be used for analysis later on.

Similar to other global forcing datasets (Li and Ma, 2010; Rust et al., 2014; Sheffield et al., 2006; Rienecker et al., 2011), the WFDEI forcing contains a number of problems. For WFDEI we identified five issues. (i) The rainfall in Gabon in Africa seems to be unrealistically high. A likely reason for this could be a unit error in the reported precipitation. (ii) There are some concerns about the energy forcing terms of WFDEI (SWdown, LWdown) over the Amazon region (an underestimation of SWdown and overestimation of LWdown, coming from the ERA-Interim data). (iii) In a number of time stamps there is some incoming radiation noise at night time (about $0.05\,\mathrm{W\,m^{-2}}$), also inherited from the original ERA-Interim data, resulting from the processing of the fluxes archived by the atmospheric model in ERA-Interim. (iv) Large positive values of SWdown $> 5\,\mathrm{W\,m^{-2}}$ at night time in some islands and coastal points introduced during the WFDEI data processing. (v) A total of nine grid points out of 67 209 were found with a substantial conversion of liquid rainfall (in ERA-Interim) into snowfall (in WFDEI). For some models, this can lead to a continuous accumulation of snow over those points. Although some of these problems could have been addressed it was decided to keep the original WFDEI dataset to guarantee consistency with other studies in the literature.

The simulations were performed from 1 January 1979 to 31 December 2012 in a continuous run. With respect to static fields (e.g. soil physical parameters, land cover type) each modelling group used their own datasets, as this is considered to be part of the modelling system, and exchanging these fields between models is not straightforward. Two simple quality control tests were applied to the data: (i) generic metadata and quality control (including mass balance checks) and (ii) comparison of minimum, maximum and mean fields. The first test is automatic (script available at: https://github.com/earth2observe/project-tools), while the second relies on the inspection of aggregated statistics.

To ensure uniform input and use of the data the project's servers have been configured to host the forcing data and also provide an interchange platform for the project using a THREDDS data server (Domenico et al., 2006). This server is also used to distribute the data to the rest of the world and includes an interactive portal (http://wci.eartH2Observe.eu). The flow of data for the current model runs is depicted in Fig. 1. Direct access to the THREDDS server is available at http://wci.earth2observe.eu/thredds/catalog.html while a mirror of the data is hosted at http://al-tc002.xtr.deltares.nl:8080/thredds/catalog.html.

All model outputs passed the metadata consistency checks. The simulations from WaterGAP3, PCR-GLOBWB and JULES have grid points with residuals above the defined threshold ($5.0 \times 10^{-6}\,\mathrm{kg\,m^{-2}\,s^{-1}}$) due to the water transport in the river network that is not accounted for in the water balance calculations, but the water balance is closed on a basin scale. For the second data quality check the temporal and global spatial minimum, maximum, mean and standard deviation were computed for all variables and compared among the different models using the monthly data. This allowed the identification of several problems (e.g. different signal conventions for the fluxes, numerical/rounding errors) that were corrected directly in the data server or by each institution.

Table 2. Overview of the meteorological forcing used in the simulations, and the corrections applied to the original ERA-Interim during the WFDEI processing (Weedon et al., 2014).

Variable	Standard name	Definition	Units	Corrections
Wind	wind_speed	Wind speed at a reference level near the surface – 10 m	$\mathrm{m\,s^{-1}}$	None
Tair	air_temperature	Temperature at a reference level near the surface – 2 m	K	Elevation using lapse rate; CRU average Tair and average diurnal temperature range
Qair	specific_humidity	Specific humidity at reference level near the surface – 2 m	$\mathrm{kg\,kg^{-1}}$	Via changes in Tair and PSurf
PSurf	surface_air_pressure	Pressure at the surface	Pa	Via changes in Tair
SWdown	surface_downwelling_shortwave_flux_in_air	Average incident radiation in the shortwave part of the spectrum	$\mathrm{W\,m^{-2}}$	CRU average cloud cover and effects of inter-annual changes in atmospheric aerosol loading
LWdown	surface_downwelling_longwave_flux_in_air	Average incident radiation in the longwave part of the spectrum	$\mathrm{W\,m^{-2}}$	Via fixes in relative humidity and changes in Tair, PSurf and Qair
Rainf	rainfall_flux	Average rainfall (only liquid phase)	$\mathrm{kg\,m^{-2}\,s^{-1}}$	CRU number of wet days and precipitation totals
Snowf	snowfall_flux	Average snowfall (only solid phase)	$\mathrm{kg\,m^{-2}\,s^{-1}}$	CRU number of wet days and precipitation totals

Table 3. List of most important output variables and conventions. If a standard name is not available the name will be used in the respective netCDF files.

Name	long_name (attribute)	standard_name (attribute)	Units (attribute)	Definition	Positive direction
Precip	Total precipitation	precipitation_flux	$\mathrm{kg\,m^{2}\,s^{-1}}$	Average of total precipitation (Rainf+Snowf)	downwards
Evap	Total evapotranspiration	water_evaporation_flux	$\mathrm{kg\,m^{2}\,s^{-1}}$	Sum of all evaporation sources, averaged over a grid cell	downwards
Runoff	Total runoff	runoff_flux	$\mathrm{kg\,m^{2}\,s^{-1}}$	Average total liquid water draining from land (specific runoff)	into grid cell
RivOut	River discharge	NA	$\mathrm{m^{3}\,s^{-1}}$	Water volume leaving the cell	downstream
SnowFrac	Snow cover fraction	surface_snow_area_fraction	–	Fraction of each grid cell covered with snow (0–1)	–
SWE	Snow water equivalent	liquid_water_content_of_surface_snow	$\mathrm{kg\,m^{-2}}$	Total water mass of the snowpack (liquid or frozen), averaged over a grid cell	–
SurfStor	Surface water storage	NA	$\mathrm{kg\,m^{-2}}$	Total liquid water storage, other than soil, snow or interception storage (i.e. lakes, river channel or depression storage)	–
CanopyInt	Canopy interception depth	–	$\mathrm{kg\,m^{-2}}$	Depth of intercepted water on the canopy	–
SnowDepth	Depth of snow layer	surface_snow_thickness	m	Total snow depth	–
SurfMoist	Surface soil moisture	NA	$\mathrm{kg\,m^{-2}}$	5 cm depth or first model layer	–
RootMoist	Root zone soil moisture	NA	$\mathrm{kg\,m^{-2}}$	Total soil moisture available for evapotranspiration (or up to 1 m depth if not defined)	–
TotMoist	Total soil moisture	NA	$\mathrm{kg\,m^{-2}}$	Vertically integrated total soil moisture	–
GroundMoist	Ground water	NA	$\mathrm{kg\,m^{-2}}$	Ground water not directly available for evapotranspiration	–

NA = not available

3 Dataset characteristics

3.1 Multi-model SNR

An important component of a multi-model dataset is the possibility to characterize the multi-model agreement or consistency. While such characteristics do not imply quality or skill, it can provide an overview of the regions and variables where datasets strongly disagree. This information can be used either by the modelling community to focus on particular aspects of their models or by users as a first order uncertainty estimate of the multi-model ensemble. The agreement

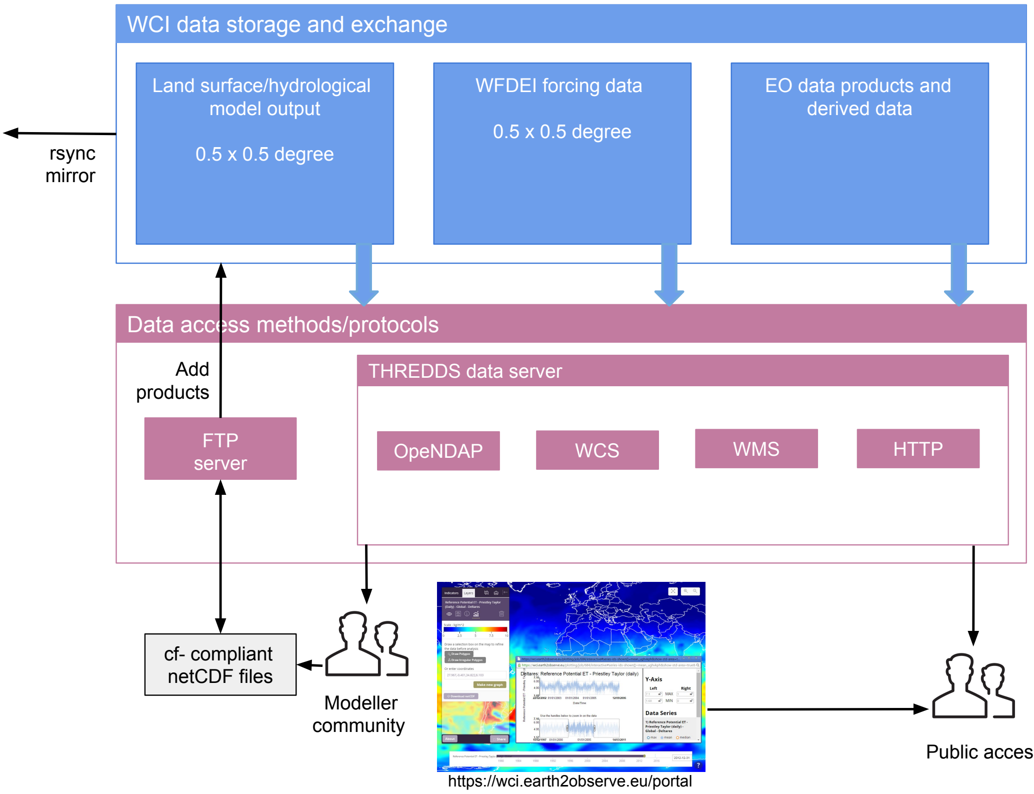
WCI data storage and exchange
Land surface/hydrological model output
0.5 x 0.5 degree
WFDEI forcing data
0.5 x 0.5 degree
EO data products and derived data
rsync mirror
Data access methods/protocols
Add products
FTP server
THREDDS data server
OpenNDAP
WCS
WMS
HTTP
cf- compliant netCDF files
Modeller community
Public access
https://wci.earth2observe.eu/portal

Figure 1. Flow of input and model data via the eartH2Observe Water Cycle Integrator (WCI). All data are accessible to users via a number of open protocols and a tailor-made user interface at http://wci.earth2observe.eu.

metric selected here is the SNR, which compares the signal to noise levels by relating the ensemble's variance to that of the individual members, which has been widely used as a classical measure of predictability in seasonal forecasting (Kumar and Hoerling, 2000). SNRs were calculated for three model variables: evapotranspiration, runoff and root zone soil moisture plus an ensemble of global precipitation datasets. See Appendix A for a detailed description of the SNR calculations.

We performed the calculations with monthly mean anomalies to focus on the model agreement in terms of intra-seasonal to inter-annual variability. Since all models were driven by the same atmospherical conditions, low values of SNR can be directly associated with differences in the representation of processes such as energy partitioning and runoff generation – i.e. ensemble uncertainty. However, since one single forcing was used, the ensemble is missing an important source of uncertainty: the driving data. Precipitation is likely the main source of uncertainty; it is very important for the terrestrial water balance while at the same time it is difficult to observe both locally and remotely. To put our results into perspective, we also computed the SNR of an ensemble of precipitation datasets including three atmospheric reanalysis datasets (ERA-Interim: Dee et al., 2011, JRA55: Kobayashi et al., 2015, MERRA: Rienecker et al., 2011), two datasets based on rain gauges (GPCC: Schneider et al., 2011, CRU TS3.10: Harris et al., 2014a) and two datasets derived from remote sensing data (GPCP: Adler et al., 2003, CMAP: Xie and Arkin, 1997).

The SNRs were computed for the period January 1980 to December 2012 by removing the mean annual cycle in each grid point from each ensemble member such that $\overline{y}$ and $\overline{y}_i$ are both equal to zero; see Appendix A. Due to the differences in the representation of soil moisture storage among the models, the ensemble mean is dominated by those models with a higher water holding capacity and hence larger absolute soil moisture variability. Therefore, the root zone soil moisture was first transformed to percentiles before performing the calculations for each model (Wang et al., 2011). The SNR varies between 0 and $+\infty$ with values below 1 indicating that the inter-model variability is larger than the ensemble mean variability, i.e. low inter-model agreement, and values above 1 indicating a high inter-model agreement.

The multi-model consistencies in terms of inter-annual variability evaluated by the SNR are shown in Fig. 2a–c for the different variables, and aggregated by climate types in Fig. 3a–c. The results highlight regions with a low inter-model agreement over (i) snow-dominated regions (runoff, evapotranspiration and root zone soil moisture) and (ii) tropical rainforest and monsoon regions (evapotranspiration), whereas the temperate areas show a high inter-model agreement. Comparing these results with the precipitation datasets' agreement (Figs. 2d and 3d), which were not included in the driving data, the large uncertainty in the tropical areas is not reflected in the runoff or root zone soil moisture. On the other hand, there is little disagreement in the precipitation datasets in cold regions, which could be caused by the fact that in these regions we rely mostly on reanalysis data sources, while the multi-model ensemble contains a large spread. The SNRs suggest that over cold regions the multi-model ensemble is generating a large spread (likely due to the different treatment of cold processes among the models) while over the tropical areas some of the multi-model agreement might be underestimating the actual uncertainty by neglecting the driving data uncertainty in the ensemble generation.

3.2 Verification with external datasets

We use the ILAMB system (The International Land Model Benchmarking Project; Luo et al., 2012; Mu et al., 2016) to compare the model results against benchmark data mostly derived from satellite remote sensing, of evapotranspiration (ET), terrestrial water storage anomaly (TWSA), soil moisture anomaly (SMA), snow water equivalent (SWE) and snow cover fraction (SC). For evapotranspiration both the GLEAM-V2a and V3b datasets (Miralles et al., 2011; Martens et al., 2017) and the MODIS evapotranspiration estimates (Mu et al., 2011) were used while for soil moisture anomaly we used the combined active+passive microwave ESA CCI soil moisture dataset (Dorigo et al., 2015, 2012). The snow cover dataset was obtained from the Interactive Multisensor snow and ice mapping System (IMS; Ramsay, 1998; Helfrich et al., 2007) while for SWE we used GLOBSNOW-2 (Takala et al., 2011; Pulliainen, 2006). The TWSA dataset was obtained from GRACE data (JPL; Landerer and Swenson, 2012). The complete benchmark results are available at the dataset storage entry page at https://github.com/earth2observe/water-resource-reanalysis-v1 while a summary is given in Appendix B, Table B1 to B6. A project report presenting in-depth verification is provided in the Supplement.

ILAMB provides a scoring system to relate modelled results to reference datasets. In the ILAMB system multiple performance metrics are calculated, and additionally these metrics are converted to scores ranging between 0 and 1 to facilitate comparison and averaging. In this study three performance metrics are calculated for each of the five model variables evaluated (ET, TWSA, SMA, SWE, SC): total bias, root mean square error (RMSE) and phase difference (difference in months between peak values); furthermore a total of five 0–1 scores are calculated, for global bias, RMSE, seasonal cycle, spatial distribution and inter-annual variability, plus a 0–1 overall score that summarizes them. The metrics and scoring are explained in detail in the ILAMB documentation (http://earth2observe.github.io/water-resource-reanalysis-v1/assets/pdf/ILAMB_metrics_document.pdf). For the case of anomaly variables (TWSA and SMA), the bias and spatial distribution scores are calculated over the standard deviations of the variables, and not directly over the anomalies. The current study is not an exhaustive test of the performance of the different models, instead we focus on the multi-model ensemble, commenting on specific models only when this is thought to be important to the ensemble as a whole. Apart from the global level, the ILAMB results are also available for a number of predefined regions (biomes; see Fig. 4).

Although there are a number of uncertainties associated with TWSA as estimated by GRACE measurements (Long et al., 2014; Riegger et al., 2012), resulting from the uncertainty of the GRACE data itself and the leakage corrections, the results provide an independent means of evaluating our model results. Both the peak month phase difference and the ensemble mean bias are negative for most models compared to GRACE (see Fig. 5) which could indicate that these models lack a store such as rivers or lakes (Kim et al., 2009) or that the size of the included stores is inadequate. Table 4 lists the models and the variables used to calculate TWSA in this study. Results for the WaterGAP3 model should be interpreted with caution because it lacks total moisture which is a large store in many models. RMSE is largest in high-precipitation regions and lowest in dry areas (Fig. 5, bottom panel). Overall, total scores for TWSA are close between the models, while the highest values are recorded for the ensemble mean, suggesting that the lowest-scoring models do not have a large detrimental effect on the ensemble mean. Although the scores for the TWSA (ranging between 0.49 for SWBM and 0.60 for HTESSEL) are similar at the global level, there are larger differences between the models at the regional level. The negative phase difference for nearly all models indicates that the peak in the anomaly in the models occurs earlier than in the GRACE data. This could point to a general underestimation of the TWSA (for example the groundwater component) resulting in a system that reacts too quickly. In addition, the negative phase difference is strongest in the cold regions indicating that snow modelling might be an important factor here. There are large regional differences in the phase difference performance showing best results in Africa (below the Sahel), Australia, India, and South America north of 25° S. There is no single model that performs best in all regions.

For evapotranspiration our results compare better to the GLEAM products (mean model total score of 0.83 for both products) but less so to the MODIS product (mean model score of 0.78; see Table 5). Most models (and the ensemble mean) evaporate more water than the remote sensing based estimates. As shown by Miralles et al. (2016) it is still difficult to determine the quality of global ET products. The

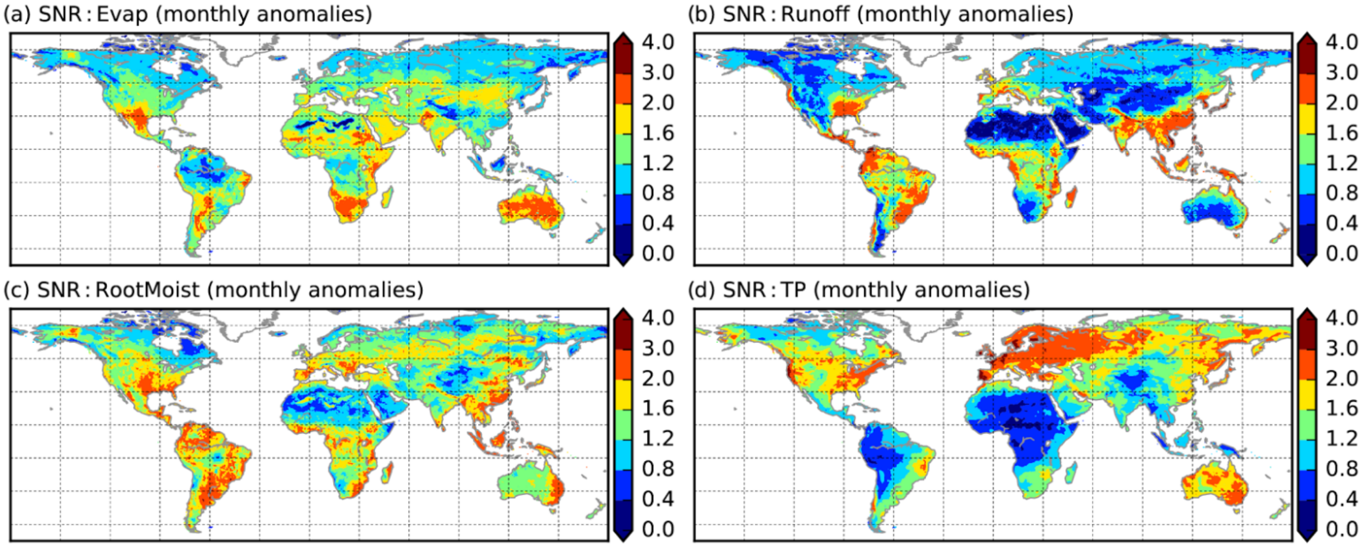
(a) SNR : Evap (monthly anomalies)
(b) SNR : Runoff (monthly anomalies)
(c) SNR : RootMoist (monthly anomalies)
(d) SNR : TP (monthly anomalies)
4.0
3.0
2.0
1.6
1.2
0.8
0.4
0.0

Figure 2. Signal-to-noise ratio of monthly mean anomalies of evapotranspiration **(a)**, runoff **(b)**, root zone soil moisture **(c)** and precipitation **(d)**.

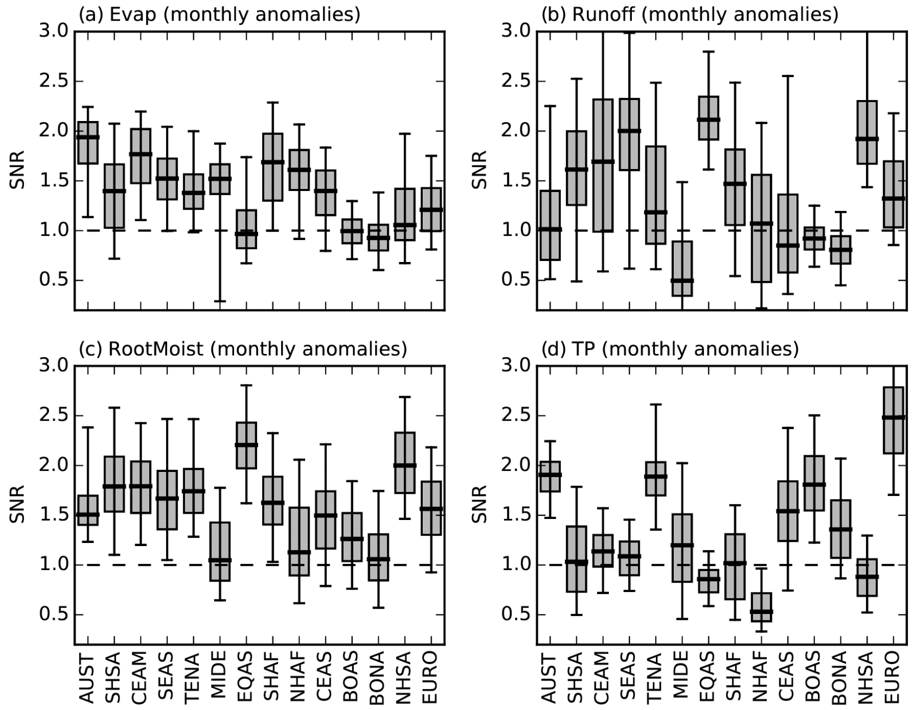

Figure 3. Distribution of the SNR of monthly anomalies over different BIOMES (horizontal axis; see Fig. 4) for evapotranspiration **(a)**, runoff **(b)**, root zone soil moisture **(c)** and precipitation **(d)**. The boxplots represent the spatial variability of the individual pixels of SNR in each biome extending from percentile 5 to 95 (whiskers), percentiles 25 to 75 (box) and median (horizontal line).

driving force behind the ET estimates by the multi-model ensemble is provided by the WFDEI (based on ERA-Interim) which is shown to have relatively high ET (Miralles et al., 2016). It is beyond the scope of the current study to examine the differences in ET between the models, but the choice of calculation method of potential ET within the models may already account for a large spread (see e.g. Weiland et al., 2015, who used the WFDEI forcing to calculate FAO Penman–Monteith reference evapotranspiration (ET) (Allen et al., 1998), Priestley–Taylor reference ET (Priestley and Taylor, 1972) and Hargreaves reference ET (Hargreaves and Samani, 1982; Hargreaves and Allen, 2003; Sperna Weiland

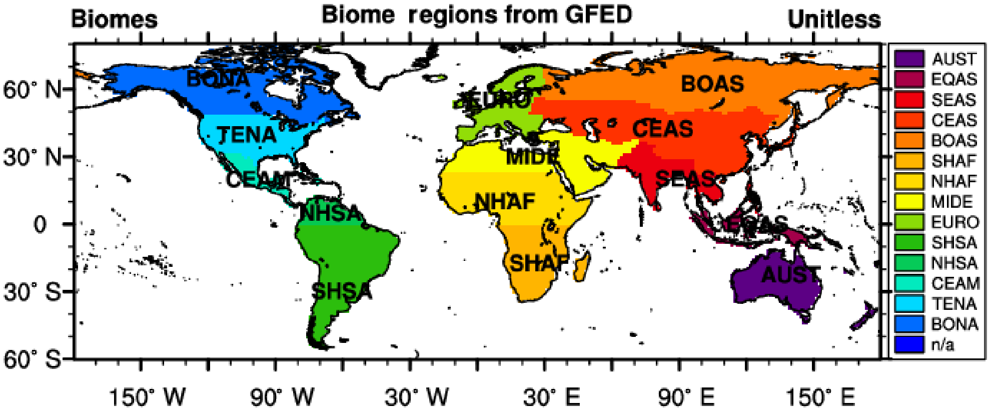

Figure 4. BIOMES used in calculating regional averages. These are: AUST (Australia), EQAS (equatorial Asia), SEAS (Southeast Asia), CEAS (central Asia), BOAS (boreal Asia), SHAF (Southern Hemisphere Africa), NHAF (Northern Hemisphere Africa), MIDE (Middle East), EURO (Europa), SHSA (Southern Hemisphere South America), NHSA (Northern Hemisphere South America), CEAM (Central America), TENA (temperate North America) and BONA (boreal North America).

Table 4. Components used in total water storage estimation for each model. The definition of the variables can be found in Table 3.

	SWE	CanopInt	SurfStor	TotMoist	GroundMoist
HTESSEL-CaMa	x	x	x	x	–
JULES	x	x	–	x	–
LISFLOOD	x	–	–	x	x
ORCHIDEE	x	–	x	x	–
PCR-GLOBWB	x	x	x	x	x
SURFEX-TRIP	x	x	x	x	–
SWBM	x	–	–	x	–
W3RA	x	–	–	x	x
WaterGAP3	x	x	x	–	–
HBV-SIMREG	x	–	–	x	x

et al., 2012)). The method used to estimate net radiation may also play a large role. Although the results show that the ensemble mean provides best overall performance, the spread in ET is large (between 1.66 and 1.33 mm day^{-1}).

All models provided SWE, while only six models provided SC. Total performance against the reference dataset was highest for ORCHIDEE, WaterGAP3 and LISFLOOD although the bias is fairly large for all models (see Fig. 6). The phase difference seems to be influenced most by rather poor scores in the Himalaya. The total ensemble mean score for SWE is 0.67, which is lower than the highest model score of 0.74, suggesting that in this case the model mean should be used with care. However, GlobSnow has been shown to miss early season snow and SWE levels can drop too rapidly in spring (Takala et al., 2011). As most models show a later spring melt than GlobSnow it remains unclear if this is a model deficiency. A number of models show an unrealistic build-up of snow over time in Europe (HBV-SIMREG and PCR-GLOBWB), boreal North America (HTESSEL, HBV-SIMREG, PCR-GLOBWB), central Asia (HTESSEL, HBV-SIMREG) and Southeast Asia (HTESSEL, JULES, LISFLOOD). This may be caused by the fact that the models have been driven by a different dataset (with different temperature and radiation characteristics) than what they have been developed with. Snow cover fraction (SC) estimated by the models compares well to the IMS results. Here SURFEX-TRIP performs best but performance in general is much closer compared to the SWE results and the ensemble mean seems to provide a good estimate. Getting the phase correct in the Himalayan region seems to be the most challenging parameter for the models (see Fig. 7).

Although current satellite-derived surface soil moisture products that cover a long period have a number of limitations (Su et al., 2016; Loew et al., 2013), and are also not completely model-independent, they have been shown to capture intra- and inter-annual soil moisture variability, and this variability is not dependent on an external atmospheric forcing dataset. In addition, comparison with land surface models and in situ data showed good correlation (Albergel et al., 2013). However, the long soil moisture record is not homogeneous because of sensor degradation as well as differences in sensor characteristics, algorithms and calibration

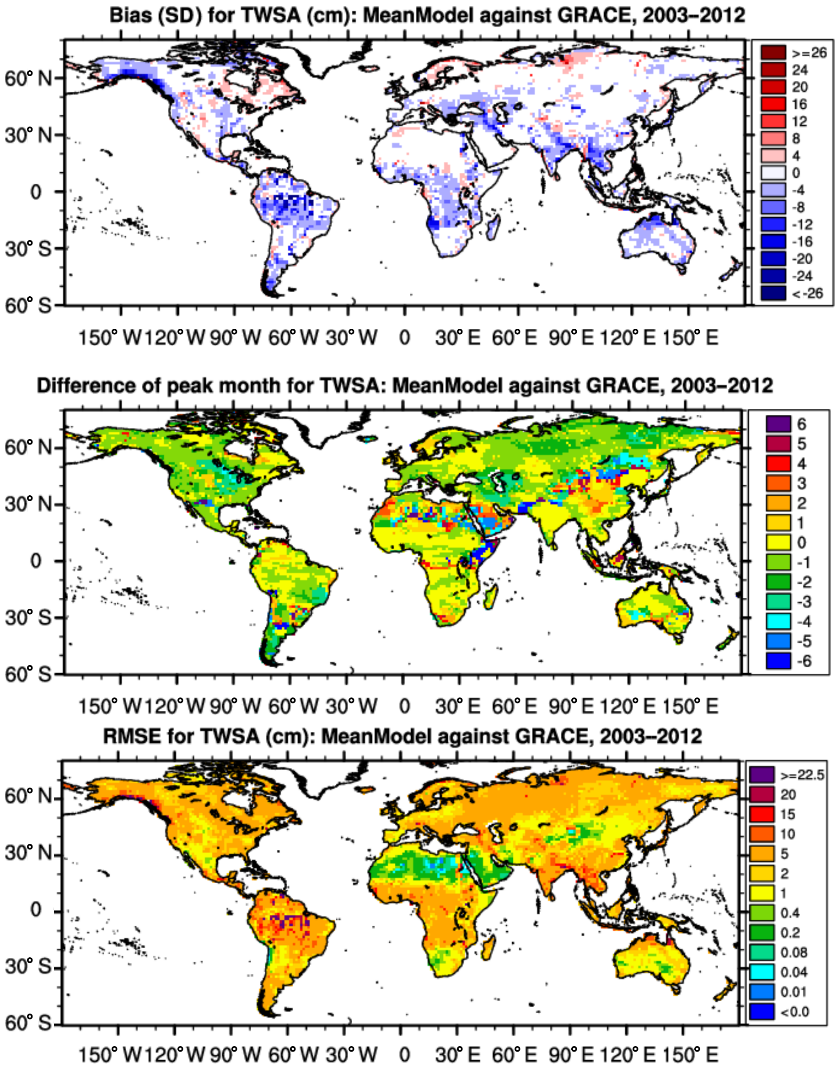
Bias (SD) for TWSA (cm): MeanModel against GRACE, 2003–2012
60° N
30° N
0
30° S
60° S
150° W 120° W 90° W 60° W 30° W 0 30° E 60° E 90° E 120° E 150° E
>=26
24
20
16
12
8
4
0
-4
-8
-12
-16
-20
-24
<-26
Difference of peak month for TWSA: MeanModel against GRACE, 2003–2012
60° N
30° N
0
30° S
60° S
150° W 120° W 90° W 60° W 30° W 0 30° E 60° E 90° E 120° E 150° E
6
5
4
3
2
1
0
-1
-2
-3
-4
-5
-6
RMSE for TWSA (cm): MeanModel against GRACE, 2003–2012
60° N
30° N
0
30° S
60° S
150° W 120° W 90° W 60° W 30° W 0 30° E 60° E 90° E 120° E 150° E
>=22.5
20
15
10
5
2
1
0.4
0.2
0.08
0.04
0.01
<0.0

Figure 5. Terrestrial water storage anomaly metrics for the ensemble mean, from top to bottom: SD of bias, phase difference (months) and root mean square error.

Table 5. Model mean evapotranspiration compared to the MODIS and GLEAM-V2a/GLEAM-V3b products. The difference in model mean annual ET in the last three rows is due to different periods used for the comparison (GLEAM-V2a 1980–2011, GLEAM-V3b 2003–2012, MODIS 2000–2012).

	Annual mean	Bias	RMSE	Phase difference	Overall score
	mm day^{-1}	mm day^{-1}	mm day^{-1}	months	–
GLEAM-V2a	1.31	–	–	–	–
GLEAM-V3b	1.27	–	–	–	–
MODIS	1.28	–	–	–	–
Model mean: GLEAM-V2a period	1.46	0.15	0.37	−0.31	0.83
Model mean: GLEAM-V3b period	1.48	0.21	0.30	−0.06	0.83
Model mean: MODIS period	1.48	0.20	0.49	−0.23	0.78

(Liu et al., 2012). Therefore we used the period 2002–2012 only. Because only a limited number of models was able to provide surface soil moisture we have chosen to evaluate root zone soil moisture from the models with the remote-sensing product (see Table 6). Although this may seem to be a large mismatch, the fact that we compare monthly aver-

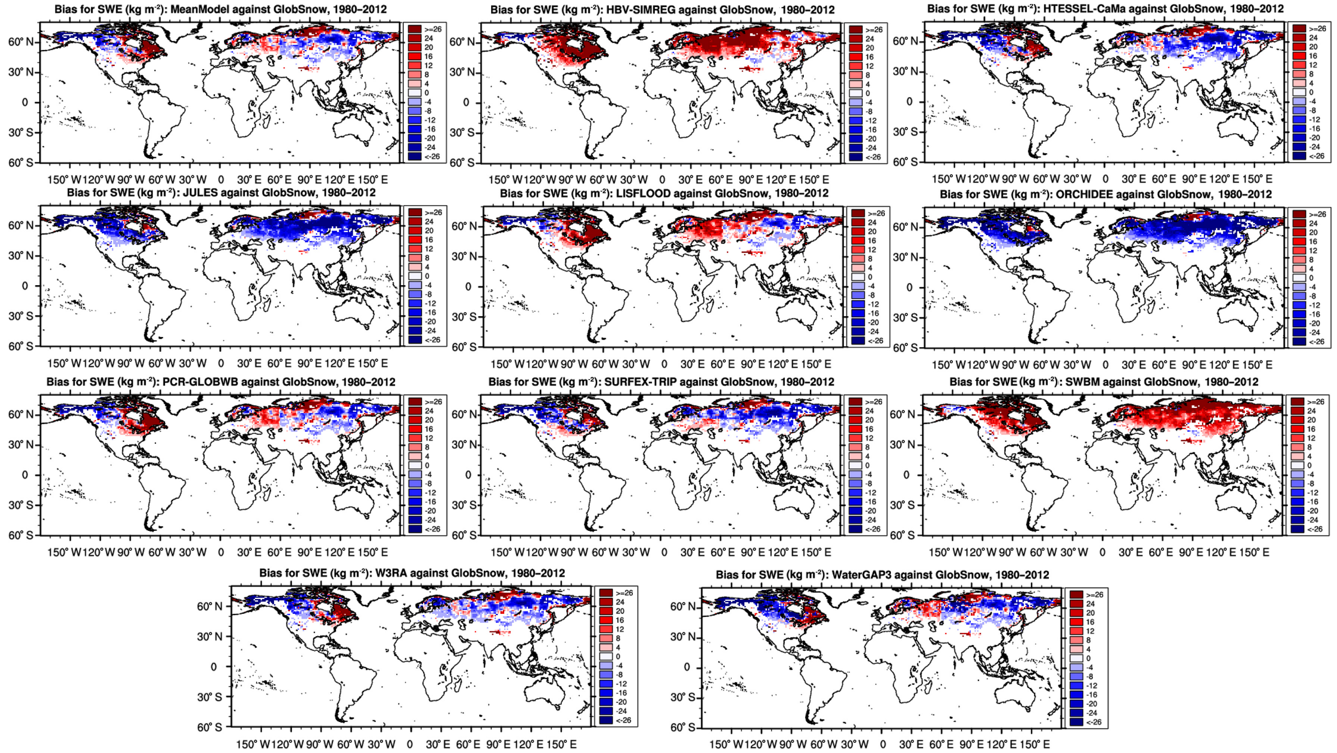
Bias for SWE (kg m⁻²): MeanModel against GlobSnow, 1980–2012
Bias for SWE (kg m⁻²): HBV-SIMREG against GlobSnow, 1980–2012
Bias for SWE (kg m⁻²): HTESSEL-CaMa against GlobSnow, 1980–2012
Bias for SWE (kg m⁻²): JULES against GlobSnow, 1980–2012
Bias for SWE (kg m⁻²): LISFLOOD against GlobSnow, 1980–2012
Bias for SWE (kg m⁻²): ORCHIDEE against GlobSnow, 1980–2012
Bias for SWE (kg m⁻²): PCR-GLOBWB against GlobSnow, 1980–2012
Bias for SWE (kg m⁻²): SURFEX-TRIP against GlobSnow, 1980–2012
Bias for SWE (kg m⁻²): SWBM against GlobSnow, 1980–2012
Bias for SWE (kg m⁻²): W3RA against GlobSnow, 1980–2012
Bias for SWE (kg m⁻²): WaterGAP3 against GlobSnow, 1980–2012
60° N
30° N
0
30° S
60° S
150° W 120° W 90° W 60° W 30° W 0 30° E 60° E 90° E 120° E 150° E
>=26
24
20
16
12
8
4
0
-4
-8
-12
-16
-20
-24
<-26

Figure 6. Bias in kg m^{-2} between GlobSnow and the model results. The top left figure shows the ensemble mean. Reds denote a positive bias, blues a negative bias.

age data should filter out the fast topsoil moisture fluctuation and make it more comparable. When comparing the results of the global models, the obtained differences are a result of the imperfections in the meteorological forcing, the model uncertainty – including parameterization and representative depth of the soil moisture – and the uncertainty in the soil moisture product. Global scores range from 0.45 for SWBM to 0.61 for HTESSEL and JULES with the ensemble mean score being very close to the best model at 0.60. Figure 8 shows the results for Australia and Southeast Asia. A recent study that used remotely sensed soil moisture to update the state of the PCR-GLOBWB hydrological model over a catchment in Australia (López López et al., 2016) showed significant improvement in simulated discharge after including the remotely sensed soil moisture, demonstrating that the hydrological models can benefit from the remotely sensed surface soil moisture. Conversely, Orth et al. (2013) showed that calibrating the SWBM against discharge only, yielded well represented soil moisture dynamics.

Beck et al. (2016a) compared the set of model outputs described in this paper with discharge from 966 medium-sized catchments and demonstrated that the calibrated models showed best performance and that, on average, for the uncalibrated models the hydrological models performed better than the land surface models in snow-dominated regions. They also show that for example ORCHIDEE performs well in cold regions but tends to underestimate runoff in the other parts of the globe. This seems to be confirmed by the SC and SWE results for ORCHIDEE, and by its ET results that indicate that ORCHIDEE overestimates ET in high-ET regions. Combining the results of Beck et al. (2016a) and the current results also shows that the models that have been calibrated on discharge do not necessarily give the best results

Table 6. Averaging depth of surface moisture and root zone moisture in the models.

	SurfMoist	RootMoist
HTESSEL-CaMa	0.07 m	1 m
JULES	0.1 m	1 m
LISFLOOD	–	var[b]
ORCHIDEE	0.092 m[a]	2 m
PCR-GLOBWB	var[c]	1 m
SURFEX-TRIP	0.04 m	var[b]
SWBM	–	var[c]
W3RA	0.05 m	1 m
WaterGAP3	–	var[b]
HBV-SIMREG	var[c]	var[c]

[a] using the sixth layer in SoilMoist, [b] variable depth supplied with model and available on server, [c] bucket model, depth = 0.970 m/(theta_fc-theta_wp), where 0.970 m is the capacity of the bucket for SWBM and using the FC parameter for HBV-SIMREG.

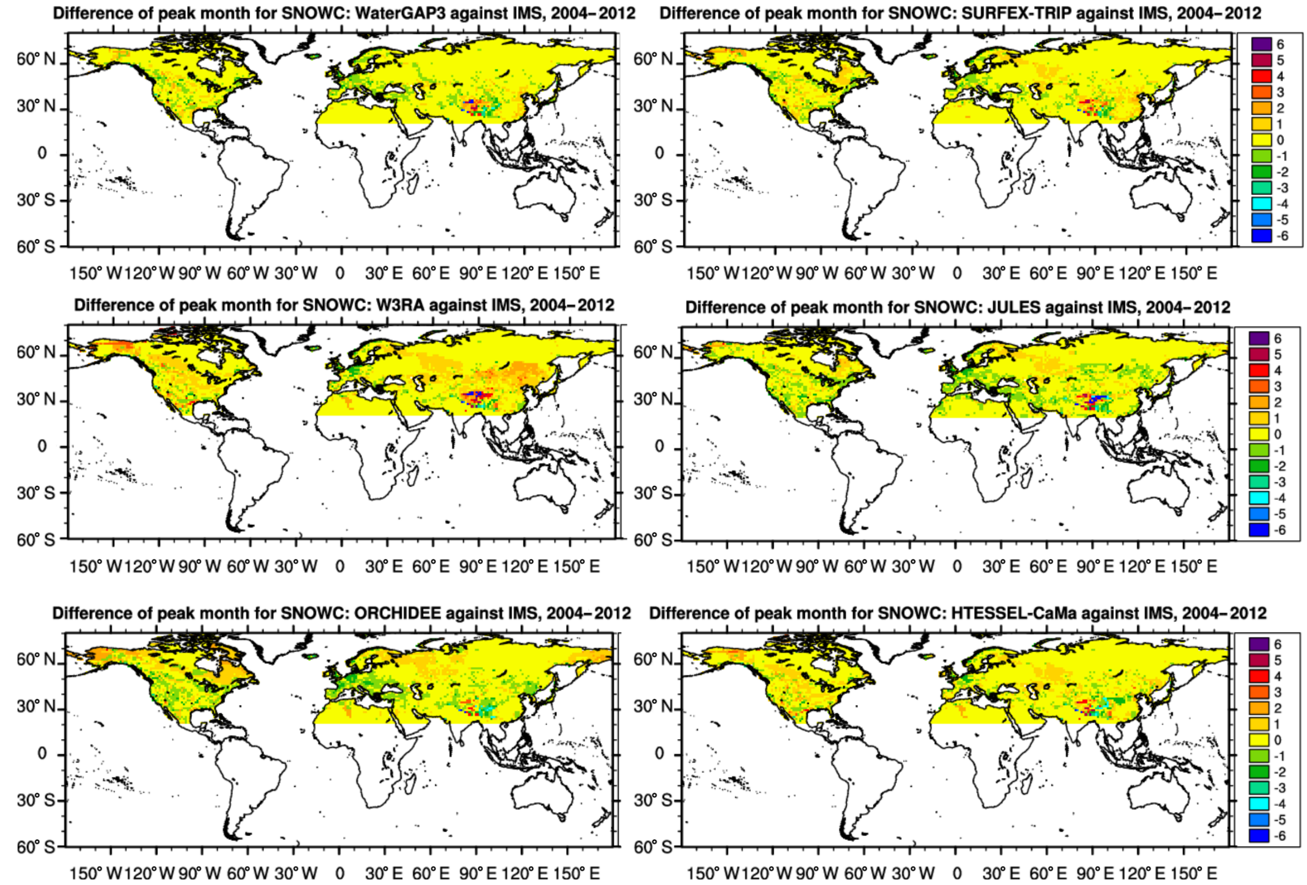
Difference of peak month for SNOWC: WaterGAP3 against IMS, 2004–2012
Difference of peak month for SNOWC: SURFEX-TRIP against IMS, 2004–2012
Difference of peak month for SNOWC: W3RA against IMS, 2004–2012
Difference of peak month for SNOWC: JULES against IMS, 2004–2012
Difference of peak month for SNOWC: ORCHIDEE against IMS, 2004–2012
Difference of peak month for SNOWC: HTESSEL-CaMa against IMS, 2004–2012
60° N
30° N
0
30° S
60° S
150° W 120° W 90° W 60° W 30° W 0 30° E 60° E 90° E 120° E 150° E
6
5
4
3
2
1
0
-1
-2
-3
-4
-5
-6

Figure 7. Difference in peak snow cover (SC) month for the models compared to the IMS dataset.

for the variables we have used here. This is especially true for the HBV-SIMREG model which provides best results for discharge but has overall lowest scores when comparing it to the other datasets.

3.3 Terrestrial water budget

Table 7 summarizes the global water budget of all the models. The terrestrial runoff totals for all models apart from ORCHIDEE (where runoff also includes the excess water flowing off floodplains and irrigated areas: de Rosnay et al., 2003) have been derived from the specific runoff per grid cell. As such the runoff into internal basins that do not drain into the ocean (endorheic basins) is included in the estimates and evaporation and abstractions from the routed water are not included (except for ORCHIDEE, where the evaporation in floodplains and abstraction for irrigation are included). The spread in runoff is fairly large (see Fig. 9) and must originate from the difference in model concepts and parameterization (including the available energy partitioning) since the atmospherical forcing data used is identical for all models. Runoff is increasing after 1997 in all models. As can be seen from the top panel of Fig. 9, this is due to the elevated precipitation during the same period making more moisture available for both evaporation and runoff. As demonstrated by Fig. 10 the results plot closely to the precipitation minus runoff line with the LSMs generally showing more evapotranspiration and less runoff compared to the GHMs.

Table 8 presents the results of this study together with a selection of previous studies. Although results are not always directly comparable due to differences in land mask and techniques used, current results compare reasonably well with previous estimates. Yearly terrestrial runoff (excluding Antarctica and Greenland) from the 10 models ranges between 38 652 and 55 877 $\mathrm{km^3\,yr^{-1}}$ with an ensemble mean of 46 268 $\mathrm{km^3\,yr^{-1}}$. Rodell et al. (2015) presented an optimized estimate of global terrestrial runoff of 45 900 $\mathrm{km^3\,yr^{-1}} \pm 4400\,\mathrm{km^3\,yr^{-1}}$ for the period 2000–2010. Furthermore, the lower estimates compare well with findings from Clark et al. (2015) ($44\,200 \pm 2660\,\mathrm{km^3\,yr^{-1}}$), while the ensemble mean compares well with the WATCH-based simulations of 49 680 $\mathrm{km^3\,yr^{-1}}$ (Clark et al., 2015) and the results by Haddeland et al. (2011) (42 000 to 66 000 $\mathrm{km^3\,yr^{-1}}$), but are higher than estimates by van Dijk et al. (2014) (20 909 $\mathrm{km^3\,yr^{-1}}$, based on 430 basins estimated to cover 90 % of global runoff) and Dai et al. (2009) (37 288 $\mathrm{km^3\,yr^{-1}}$). The relatively high runoff in the estimates

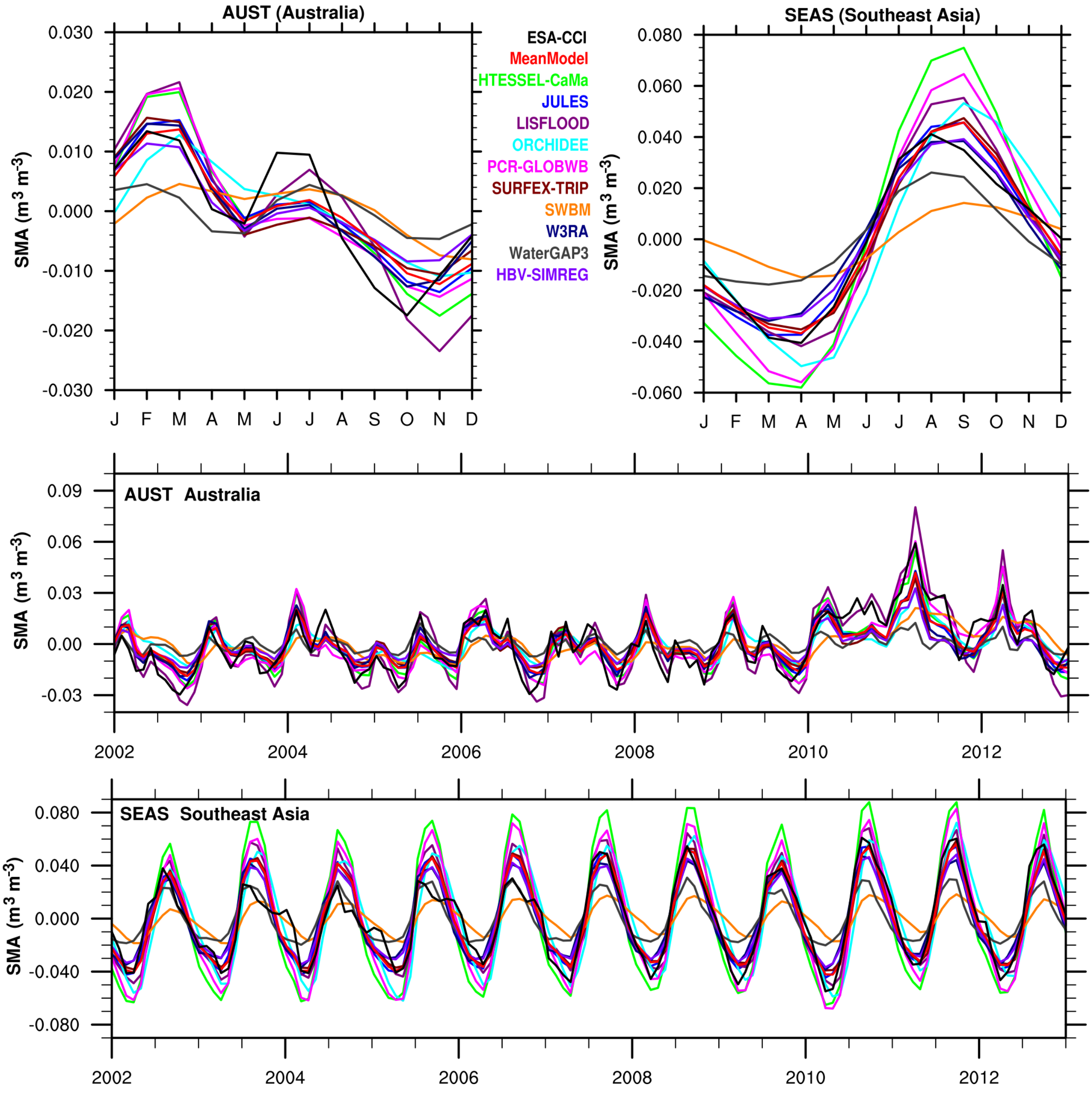

Figure 8. Soil moisture anomaly dynamics and climatology over Australia and Southeast Asia for all models compared to ESA CCI SM.

that rely on models, such as in this study, may in part be caused by the fact that they include small islands (Syed et al., 2009) which are not represented in the gauge- and GRACE-based estimates.

4 Data availability

All data are made available via the eartH2Observe server which can be accessed via the WCI portal (http://wci.earth2observe.eu; see Fig. 1) which offers plotting and collaboration features, or direct via a THREDDS server allowing access via OPeNDAP, WCS WMS and direct HTTP download (ftp is also supported). The main servers are hosted at PLMA-Ltd (Plymouth, United Kingdom) and a mirror server is hosted at Deltares (Delft, the Netherlands). Data are stored on the server in netCDF-cf compliant files. All data generated for this paper are freely available via the OCD Open Database Licence (http://opendatacommons.org/licenses/odbl/summary/). The DOI for the data is https://doi.org/10.5281/zenodo.57760.

5 Conclusions and outlook

For most of the variables we have found that the ensemble of land-surface and hydrological models gives satisfactory results and the agreement between the models is good for large

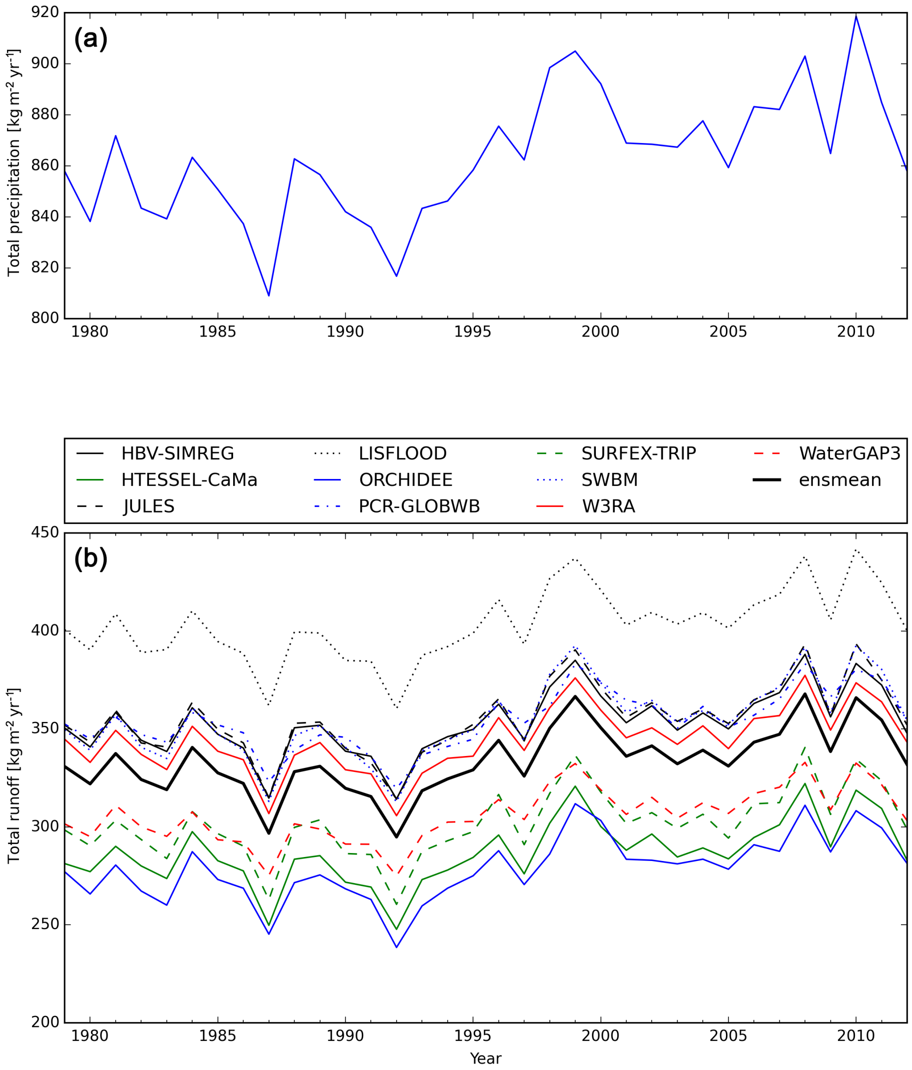

Figure 9. Global mean yearly precipitation **(a)** and runoff ($\mathrm{kg\,m^{-2}\,yr^{-1}}$) for all models **(b)**. The thick black line represents the ensemble mean.

parts of the terrestrial earth. The multi-model agreement in terms of monthly anomalies using the SNR provided an insight into the main regions/variables where the dataset shows a reduced multi-model agreement: (i) snow-dominated regions (in all three variables – evapotranspiration, runoff and root zone soil moisture) and (ii) tropical rainforest and monsoon regions (only for evapotranspiration). Furthermore, the SNR of an ensemble of precipitation datasets was calculated, indicating a large uncertainty of precipitation in the tropics, which is not reflected in the ensemble runoff from the models. In cold regions the precipitation uncertainty derived from the available datasets is small compared to the uncertainty of the multi-model simulations. This suggests that the model cold processes are an important factor in this multi-model disagreement. However, in these regions there are no satellite estimates and a limited number of rain gauges, which means that the current global datasets most probably underestimate the precipitation uncertainty in those regions.

The ability of the multi-model ensemble to model total water storage dynamics at the scale of the GRACE data is generally good although models predict the peak in total water storage earlier in all regions. The fact that the phase difference is largest in the cold zones also indicates that there are difficulties in modelling the snow pack. This is in line with the observation of Beck et al. (2016a) that the models tend to produce the snowmelt runoff peak too early. Getting SWE right is difficult for the models, and the results of the verification with the GlobSnow data strengthen the results of the SNR analysis (and the TWSA analysis), which point to cold regions as regions with low inter-model agreement. It is one of the few cases where the multi-model performance against the reference dataset is markedly lower than the performance of the best model.

Table 7. Mean evaporation and runoff for the whole period compared to the change in storage of the total moisture component of each model. Mean precipitation for the whole period using the common land surface mask was 863 ($\text{kg m}^{-2}\,\text{yr}^{-1}$). Surface water storage and storage in snow and glaciers is not taken into account. A positive change in storage indicates the model lost water storage during the simulation period.

Model	Total evaporation ($\text{kg m}^{-2}\,\text{yr}^{-1}$)	Runoff ($\text{kg m}^{-2}\,\text{yr}^{-1}$) ($\text{km}^3\,\text{yr}^{-1}$)	Δ storage (kg m^{-2})
HBV-SIMREG	529	353 (48 945)	1.5
HTESSEL-CaMa	576	287 (39 785)	21.5
JULES	524	355 (49 239)	11.5
LISFLOOD	480	403 (55 877)	−9.3
ORCHIDEE[a]	598	278 (38 652)	9.4
PCR-GLOBWB	511	354 (49 096)	3.2
SURFEX-TRIP	561	301 (41 818)	−8.8
SWBM	519	354 (49 129)	7.9
W3RA	518	344 (47 721)	3.1
WaterGAP3[b]	549	306 (42 415)	–
Ensemble mean	537	334 (46 268)	4.4

[a] Runoff results have been obtained using the routed discharge and not the grid cell specific runoff as in the other models. [b] Change in storage for WaterGAP3 is not shown here because it only supplies root-zone storage.

Table 8. Comparison of mean annual total of terrestrial precipitation, evapotranspiration and runoff with previous studies.

	Runoff ($\text{km}^3\,\text{yr}^{-1}$)	Total evaporation ($\text{km}^3\,\text{yr}^{-1}$)	Precipitation ($\text{km}^3\,\text{yr}^{-1}$)
This study	46 268	74 457	119 659
Rodell et al. (2015)	45 900	70 600	116 500
Clark et al. (2015)	44 200	–	–
Haddeland et al. (2011)	54 186	72 103	12 6000
Syed et al. (2009)	30 354	–	–
van Dijk et al. (2014)	20 909	–	–
Dai et al. (2009)	37 288	–	–
Trenberth et al. (2007)	37 300	73 000	112 600

Although the scores indicate a good performance of the ensemble, the evapotranspiration estimates are higher than those by the benchmark datasets. This, combined with the large spread within the ensemble itself, indicates that the ET estimates have a large uncertainly and further work is needed to improve the results. It also shows that in future versions of the dataset potential ET (PET) and net radiation should also be reported by all models as the choice of PET calculation method and net radiation estimate may be large contributors to the recorded spread in ET estimates.

The current study shows a wide spread in runoff into the oceans derived from the set of models used. The large range stems from a combination of different total evaporation values and different storage dynamics in the models due to the different concepts and parameterization of runoff generation. Given the large spread it seems plausible that the ensemble mean provides the most reliable estimate of the global water fluxes although there is no independent way of testing this assumption.

At the global level the multi-model ensemble mean provides the best (or close to the best) performance for most of the variables we investigated using the ILAMB system although caution should always be used. Beck et al. (2016a) concluded similarly in their investigation of the current ensemble with respect to global discharge. Beck et al. (2016a) also show that the ensemble mean comes close to the best (calibrated) models with respect to discharge. The main exception is SWE, where the ensemble mean is not the best performer. Furthermore, the results for TWSA for WaterGAP3 should probably be discarded as we do not have all the required information. At the regional level the picture is less clear. This means that although the ensemble mean can be regarded as a best first estimate, a look at the regional results is required for basin-scale applications of the current results. Nevertheless, the above demonstrates that the ensemble mean of the present dataset could be used to evaluate water resources.

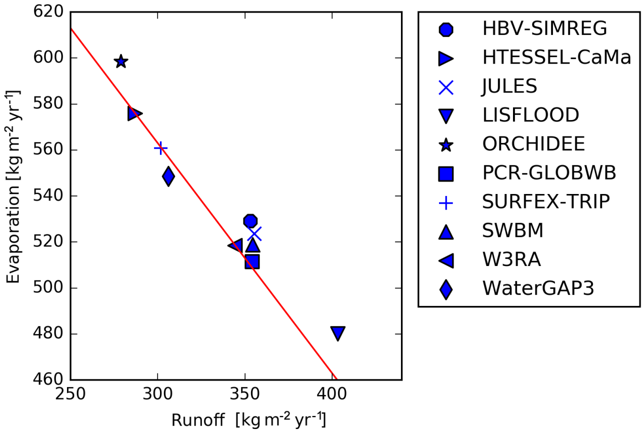

Figure 10. Average runoff plotted against average total evaporation (both expressed in kg m^2 yr^{-1}) for all models. The solid line represents the input precipitation minus runoff. For this line we used 863 kg m^{-2} yr^{-1}, which is the average calculated over all grid cells that have values for each model.

The above shows a couple of areas of importance for further development of global models and datasets and the current set in particular: precipitation estimates in the tropics, cold weather processes and evapotranspiration losses. This does not mean that other processes are already properly represented in the global models and that the influence of these processes is not important or not reflected in the current results. In particular for snow precipitation we rely on reanalysis mostly and the uncertainty in the SWE estimates could also stem from snow input. Work on improving the precipitation estimates has been started by creating a merged precipitation product (Beck et al., 2017) that may help to improve the forcing input.

Constraining models with soil moisture may reduce the spread in evapotranspiration rates and discharge estimates (see e.g. López López et al., 2016) while van Dijk et al. (2014) demonstrated that the use of satellite-derived total water storage can be successfully used to constrain global hydrological models. When combining different data sources estimating the errors associated with them becomes very important. This may be done using error models that allow for error propagation for various scenarios of data assimilation and data source sampling (Anagnostou et al., 2010). We plan to combine this in a future version of the multi-model ensemble that includes uncertainty envelopes (Nikolopoulos et al., 2010) and error estimates for runoff and other hydrological variables.

One way of making the forcing data and model results more relevant for basin-scale studies is by including higher-resolution model runs. Several of the models will be running at a higher resolution in a future set of runs and the common resolution will be increased to 0.25° resolution. In addition, the WaterWorld model (a web-based hydrological model based entirely on provided global datasets; Mulligan, 2012) will be run at 10 km resolution at the global scale and will form part of the continuing model inter-comparisons beyond this paper.

A literature search reveals that the data we produced have already been used by several other researchers including a study investigating vegetation–atmosphere coupling (Zscheischler et al., 2015). This demonstrates the value of open data that is easy to access and comes with little restrictions on its use. Furthermore, the hydrological simulations that are performed within the eartH2Observe project can be reproduced by other groups by accessing the forcing data on the data server.

Appendix A: SNR

The SNR of the multi-model ensemble was computed as the ratio between the external variance (V_{ext}) and the internal variance (V_n) as

$$\text{SNR} = \frac{V_{\text{ext}}}{V_n}, V_{\text{ext}} = \sqrt{V^2{}_h + V^2{}_n}. \tag{A1}$$

Here V_h is the total variance with the internal and total variances computed as

$$V^2{}_h = \frac{1}{NM-1}\sum_{i=1}^{N}\sum_{j=1}^{M}\left(y_{i,j} - \overline{y}\right)^2 \tag{A2}$$

$$V^2{}_n = \frac{1}{NM-N}\sum_{i=1}^{N}\sum_{j=1}^{M}\left(y_{i,j} - \overline{y}_i\right)^2 \tag{A3}$$

and $\overline{y}$ is defined as

$$\overline{y} = \frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M} y_{i,j} \tag{A4}$$

with $\overline{y}_i$:

$$\overline{y}_i = \frac{1}{N}\sum_{j=1}^{N} y_{i,j}. \tag{A5}$$

In the above, $y_{i,j}$ is the value for the ensemble member j and time i, with M the number of ensemble members and N the length of the time series, $\overline{y}$ is the total temporal and ensemble mean and $\overline{y}_i$ is the ensemble mean.

Appendix B: Summary of ILAMB results

Table B1. Global variables summary.

	Mean Model	HTESSEL-CaMa	JULES	LISFLOOD	ORCHIDEE	PCR-GLOBWB	SURFEX-TRIP	SWBM	W3RA	WaterGAP3	HBV-SIMREG
Soil moisture anomaly	0.60	0.61	0.61	0.60	0.57	0.59	0.60	0.45	0.60	0.51	0.58
Evapotranspiration	0.81	0.79	0.80	0.80	0.76	0.78	0.77	0.78	0.81	0.70	0.78
Snow water equivalent	0.67	0.64	0.66	0.71	0.74	0.60	0.67	0.61	0.67	0.74	0.48
Snow cover	0.88	0.87	0.88	–	0.83	–	0.89	–	0.85	0.85	–
Terrestrial water storage anomaly	0.63	0.62	0.59	0.56	0.53	0.59	0.59	0.52	0.62	0.53	0.60

Table B2. Diagnostic summary for soil moisture anomaly: model vs. ESA-CCI.

	Annual mean (m^3 m^{-3})	Bias (SD)	RMSE (m^3 m^{-3})	Phase difference (months)	Global bias	RMSE	Seasonal cycle	Spatial distribution	Interannual variability	Overall
MeanModel	0.02	−0.01	0.02	−0.23	0.63	0.4	0.70	0.81	0.48	0.6
HTESSEL-CaMa	0.04	0.00	0.02	−0.33	0.72	0.38	0.68	0.74	0.53	0.61
JULES	0.02	−0.01	0.02	−0.47	0.67	0.38	0.67	0.84	0.05	0.61
LISFLOOD	0.03	0.00	0.02	0.30	0.72	0.38	0.67	0.76	0.49	0.6
ORCHIDEE	0.02	−0.01	0.03	−0.3	0.61	0.37	0.64	0.73	0.51	0.57
PCR-GLOBWB	0.04	0.00	0.02	0.06	0.68	0.40	0.69	0.65	0.51	0.59
SURFEX-TRIP	0.02	−0.01	0.02	−0.58	0.65	0.38	0.67	0.78	0.51	0.6
SWBM	0.01	−0.02	0.02	−0.14	0.47	0.39	0.67	0.28	0.42	0.45
W3RA	0.02	−0.01	0.02	−0.41	0.62	0.42	0.7	0.80	0.45	0.60
WaterGAP3	0.01	−0.02	0.02	−0.61	0.52	0.39	0.64	0.57	0.41	0.51
HBV-SIMREG	0.02	−0.01	0.02	−0.17	0.58	0.42	0.72	0.72	0.45	0.58

Table B3. Diagnostic summary for evapotranspiration: model vs. GLEAM-V3B.

	Annual mean (mm day^{-1})	Bias (mm day^{-1})	RMSE (mm day^{-1})	Phase difference (months)	Global bias	RMSE	Seasonal cycle	Spatial distribution	Interannual variability	Overall
MeanModel	1.48	0.21	0.30	−0.06	0.86	0.83	0.81	0.95	0.69	0.83
HTESSEL-CaMa	1.60	0.33	0.36	−0.10	0.84	0.80	0.77	0.96	0.66	0.81
JULES	1.45	0.18	0.35	−0.27	0.84	0.80	0.77	0.93	0.70	0.81
LISFLOOD	1.33	0.06	0.37	−0.21	0.84	0.79	0.79	0.93	0.69	0.81
ORCHIDEE	1.66	0.39	0.44	−0.22	0.81	0.76	0.70	0.94	0.67	0.77
PCR-GLOBWB	1.42	0.15	0.39	−0.52	0.83	0.78	0.77	0.95	0.65	0.79
SURFEX-TRIP	1.55	0.28	0.39	0.25	0.85	0.78	0.75	0.96	0.61	0.79
SWBM	1.43	0.16	0.43	0.11	0.83	0.76	0.80	0.89	0.69	0.79
W3RA	1.44	0.17	0.32	0.13	0.86	0.82	0.85	0.95	0.66	0.83
WaterGAP3	1.45	0.18	0.58	−0.19	0.78	0.69	0.82	0.77	0.58	0.72
HBV-SIMREG	1.47	0.20	0.39	0.16	0.84	0.78	0.81	0.94	0.64	0.80

Table B4. Diagnostic summary for snow water equivalent: model vs. GLOBSNOW.

	Annual mean (kg m^{-2})	Bias (kg m^{-2})	RMSE (kg m^{-2})	Phase difference (months)	Global bias	RMSE	Seasonal cycle	Spatial distribution	Interannual variability	Overall
MeanModel	29.50	20.20	7.90	0.70	0.74	0.65	0.95	0.32	0.73	0.67
HTESSEL-CaMa	49.90	40.60	7.20	0.50	0.73	0.67	0.95	0.06	0.73	0.64
JULES	21.30	12.00	5.90	0.30	0.67	0.68	0.96	0.29	0.69	0.66
LISFLOOD	21.10	11.80	9.30	0.90	0.70	0.57	0.92	0.79	0.73	0.71
ORCHIDEE	6.90	−2.40	5.70	−0.20	0.62	0.66	0.97	0.86	0.66	0.74
PCR-GLOBWB	53.00	43.70	9.80	0.90	0.72	0.60	0.92	0.03	0.72	0.60
SURFEX-TRIP	27.10	17.90	6.50	0.50	0.73	0.68	0.95	0.26	0.73	0.67
SWBM	26.50	17.20	13.80	0.90	0.58	0.42	0.93	0.56	0.75	0.61
W3RA	17.90	8.60	6.30	0.80	0.75	0.67	0.94	0.26	0.74	0.67
WaterGAP3	14.50	5.20	7.20	0.80	0.70	0.62	0.94	0.86	0.70	0.74
HBV-SIMREG	57.30	48.00	22.80	1.00	0.55	0.41	0.91	0.03	0.57	0.48

Table B5. Diagnostic summary for snow cover: model vs. IMS.

	Annual mean (snow/snow+land)	Bias (snow/snow+land)	RMSE (snow/snow+land)	Phase difference (months)	Global bias	RMSE	Seasonal cycle	Spatial distribution	Interannual variability	Overall
MeanModel	0.14	−0.01	0.04	0.27	0.88	0.84	0.95	0.98	0.77	0.88
HTESSEL-CaMa	0.14	-0.02	0.05	0.07	0.88	0.83	0.97	0.98	0.74	0.87
JULES	0.16	0.00	0.05	−0.03	0.90	0.83	0.97	0.98	0.76	0.88
LISFLOOD	–	–	–	–	–	–	–	–	–	–
ORCHIDEE	0.13	–0.03	0.07	0.10	0.82	0.75	0.95	0.97	0.74	0.83
PCR–GLOBWB	–	–	–	–	–	–	–	–	–	–
SURFEX–TRIP	0.15	–0.01	0.04	0.02	0.90	0.84	0.98	0.99	0.77	0.89
SWBM	–	–	–	–	–	–	–	–	–	–
W3RA	0.13	–0.02	0.06	0.27	0.85	0.8	0.95	0.97	0.71	0.85
WaterGAP3	0.15	–0.01	0.06	–0.07	0.84	0.8	0.98	0.96	0.73	0.85
HBV–SIMREG	–	–	–	–	–	–	–	–		

Table B6. Diagnostic summary for terrestrial water storage anomaly: model vs. GRACE.

	Annual mean (m^3 m^{-3})	Bias (SD)	RMSE (m^3 m^{-3})	Phase difference (months)	Global bias	RMSE	Seasonal cycle	Spatial distribution	Interannual variability	Overall
MeanModel	5.73	−1.17	3.43	−0.43	0.63	0.47	0.83	0.62	0.58	0.63
HTESSEL-CaMa	6.62	−0.28	3.55	−0.53	0.61	0.47	0.82	0.67	0.54	0.62
JULES	5.14	−1.76	3.59	−0.59	0.62	0.47	0.81	0.51	0.55	0.59
LISFLOOD	4.77	−2.13	4.04	−0.42	0.60	0.43	0.81	0.39	0.55	0.56
ORCHIDEE	6.12	−0.78	3.98	−0.66	0.60	0.46	0.79	0.25	0.53	0.53
PCR-GLOBWB	8.88	1.98	4.08	0.00	0.55	0.43	0.82	0.69	0.46	0.59
SURFEX-TRIP	8.40	1.50	4.91	−0.49	0.59	0.41	0.82	0.60	0.54	0.59
SWBM	7.58	0.68	4.82	−0.42	0.50	0.36	0.80	0.44	0.49	0.52
W3RA	6.36	−0.54	3.97	−0.54	0.63	0.42	0.82	0.64	0.57	0.62
WaterGAP3	2.81	−4.09	5.07	−0.86	0.49	0.40	0.75	0.53	0.46	0.53
HBV-SIMREG	5.81	−1.09	4.19	−0.56	0.61	0.41	0.82	0.61	0.55	0.60

The Supplement related to this article is available online at https://doi.org/10.5194/essd-9-389-2017-supplement.

Author contributions. JS wrote most of the text, performed the global land surface water budget analysis and interpreted the validation with external datasets. ED performed the SNR analysis, the HTESSEL-CaMa runs and supplied the model information. AMdlT did the JULES runs, supplied the model information and performed the evaluation with the ILAMB system; GB supported with the HTESSEL-CaMa runs; AvD performed the W3RA runs, supplied the model information and supported in editing the paper; FSW supported editing the paper and performed the PET calculations; MM, J-CC and BD performed the SURFEX-TRIP runs and supplied the model information. SE, GF and MF performed the WaterGAP3 runs and supplied the model information. SP and RvB performed the PCR-GLOBWB runs and supplied the model information. JP performed the ORCHIDEE runs, supplied the model information and assisted in the terrestrial water budget calculation. HB performed the LISFLOOD and HBV-SIMREG runs and supplied the model information. RO performed the SWBM runs and supplied the model information. BC developed and maintained the data server and portal and supplied the portal information. SB supplied the WaterWorld information. WD supplied the soil moisture data and assisted in the analysis. GPW supplied the WFDEI forcing data and information and assisted in the analysis.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. This research received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 603608, "Global Earth Observation for integrated water resource assessment": eartH2Observe. The National Snow and Ice Data Center is acknowledged for providing the IMS snow cover data. We thank Tim Stockdale for the suggestions on the use of the SNR analysis.

Edited by: David Carlson
Reviewed by: two anonymous referees

References

Adler, R. F., Huffman, G. J., Chang, A., Ferraro, R., Xie, P.-P., Janowiak, J., Rudolf, B., Schneider, U., Curtis, S., Bolvin, D., Albergel, C., Dorigo, W., Reichle, R. H., Balsamo, G., De Rosnay, P., Muñoz-Sabater, J., Isaksen, L., De Jeu, R., and Wagner, W.: The version-2 global precipitation climatology project (GPCP) monthly precipitation analysis (1979-present), J. Hydrometeorol., 4, 1147–1167, 2003.

Albergel, C., Dorigo, W., Reichle, R. H., Balsamo, G., de Rosnay, P., Muñoz-Sabater, J., Isaksen, L., de Jeu, R., and Wagner, W.: Skill and Global Trend Analysis of Soil Moisture from Reanalyses and Microwave Remote Sensing, J. Hydrometeorol., 14, 1259–1277, https://doi.org/10.1175/JHM-D-12-0161.1, 2013.

Allen, R. G., Pereira, L. S., Raes, D., and Smith, M.: Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56, FAO, Rome, 300, 6541, 1998.

Anagnostou, E. N., Maggioni, V., Nikolopoulos, E. I., Meskele, T., Hossain, F., and Papadopoulos, A.: Benchmarking high-resolution global satellite rainfall products to radar and rain-gauge rainfall estimates, IEEE T. Geosci. Remote 48, 1667–1683, 2010.

Balsamo, G., Beljaars, A., Scipal, K., Viterbo, P., van den Hurk, B., Hirschi, M., and Betts, A. K.: A revised hydrology for the ECMWF model: Verification from field site to terrestrial water storage and impact in the Integrated Forecast System, J. Hydrometeorol., 10, 623–643, 2009.

Balsamo, G., Albergel, C., Beljaars, A., Boussetta, S., Brun, E., Cloke, H., Dee, D., Dutra, E., Muñoz-Sabater, J., Pappenberger, F., de Rosnay, P., Stockdale, T., and Vitart, F.: ERA-Interim/Land: a global land surface reanalysis data set, Hydrol. Earth Syst. Sci., 19, 389–407, https://doi.org/10.5194/hess-19-389-2015, 2015.

Barella-Ortiz, A., Polcher, J., Tuzet, A., and Laval, K.: Potential evaporation estimation through an unstressed surface-energy balance and its sensitivity to climate change, Hydrol. Earth Syst. Sci., 17, 4625–4639, https://doi.org/10.5194/hess-17-4625-2013, 2013.

Beck, H. E., van Dijk, A. I. J. M., de Roo, A., Dutra, E., Fink, G., Orth, R., and Schellekens, J.: Global evaluation of runoff from 10 state-of-the-art hydrological models, Hydrol. Earth Syst. Sci., 21, 2881–2903, https://doi.org/10.5194/hess-21-2881-2017, 2017a.

Beck, H. E., van Dijk, A. I. J. M., de Roo, A., Miralles, D. G., McVicar, T. R., Schellekens, J., and Bruijnzeel, L. A.: Global-scale regionalization of hydrologic model parameters, Water Resour. Res., 52, 3599–3622, https://doi.org/10.1002/2015WR018247, 2016b.

Beck, H. E., van Dijk, A. I. J. M., Levizzani, V., Schellekens, J., Miralles, D. G., Martens, B., and de Roo, A.: MSWEP: 3-hourly 0.25° global gridded precipitation (1979–2015) by merging gauge, satellite, and reanalysis data, Hydrol. Earth Syst. Sci., 21, 589–615, https://doi.org/10.5194/hess-21-589-2017, 2017.

Best, M. J., Pryor, M., Clark, D. B., Rooney, G. G., Essery, R. L. H., Ménard, C. B., Edwards, J. M., Hendry, M. A., Porson, A., Gedney, N., Mercado, L. M., Sitch, S., Blyth, E., Boucher, O., Cox, P. M., Grimmond, C. S. B., and Harding, R. J.: The Joint UK Land Environment Simulator (JULES), model description – Part 1: Energy and water fluxes, Geosci. Model Dev., 4, 677–699, https://doi.org/10.5194/gmd-4-677-2011, 2011.

Beven, K. and Binley, A.: The Future of Distributed Models: Model calibration and uncertainty prediction, Hydrol. Process., 6, 279–298, 1992.

Bierkens, M.: Global hydrology 2015: State, trends, and directions, Water Resour. Res., 51, 4923–4947, https://doi.org/10.1002/2015WR017173, 2015.

Bierkens, M. F. P. and Van Beek, L.: The Global Hydrological Model PCR-GLOBWB, Tech. rep., Utrecht University, available at: http://vanbeek.geo.uu.nl/suppinfo/vanbeekbierkens2009.pdf, last access: 2015.

Clark, D. B., Mercado, L. M., Sitch, S., Jones, C. D., Gedney, N., Best, M. J., Pryor, M., Rooney, G. G., Essery, R. L. H., Blyth, E., Boucher, O., Harding, R. J., Huntingford, C., and Cox, P.

M.: The Joint UK Land Environment Simulator (JULES), model description – Part 2: Carbon fluxes and vegetation dynamics, Geosci. Model Dev., 4, 701–722, https://doi.org/10.5194/gmd-4-701-2011, 2011.

Clark, E. A., Sheffield, J., van Vliet, M. T., Nijssen, B., and Lettenmaier, D. P.: Continental Runoff into the Oceans (1950–2008), J. Hydrometeorol., 16, 1502–1520, https://doi.org/10.1175/JHM-D-14-0183.1, 2015.

Collins, R., Kristensen, P., and Thyssen, N.: Water resources across Europe-confronting water scarcity and drought, no. 2 in EEA report, Office for Official Publications of the European Communities, 2009.

Dai, A., Qian, T., Trenberth, K. E., and Milliman, J. D.: Changes in continental freshwater discharge from 1948 to 2004, J. Climate, 22, 2773–2792, https://doi.org/10.1175/2008JCLI2592.1, 2009.

Decharme, B., Alkama, R., Douville, H., Becker, M., and Cazenave, A.: Global evaluation of the ISBA-TRIP continental hydrological system. Part II: Uncertainties in river routing simulation related to flow velocity and groundwater storage, J. Hydrometeorol., 11, 601–617, https://doi.org/10.1175/2010JHM1212.1, 2010.

Decharme, B., Martin, E., and Faroux, S.: Reconciling soil thermal and hydrological lower boundary conditions in land surface models, J. Geophys. Res.-Atmos., 118, 7819–7834, https://doi.org/10.1002/jgrd.50631, 2013.

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

de Rosnay, P., Polcher, J., Laval, K., and Sabre, M.: Integrated parameterization of irrigation in the land surface model ORCHIDEE. Validation over Indian Peninsula, Geophys. Res. Lett., 30, 1986–1990, 2003.

Deursen, W. P. A.: Geographical Information Systems and Dynamic Models: development and application of a prototype spatial modelling language, Faculteit Ruimtelijke Wetenschappen, Universiteit Utrecht, available at: http://library.wur.nl/WebQuery/hydrotheek/907249 (last access: 2 February 2016), 1995.

Dirmeyer, P. A., Gao, X., Zhao, M., Guo, Z., Oki, T., and Hanasaki, N.: GSWP-2: Multimodel analysis and implications for our perception of the land surface, B. Am. Meteorol. Soc., 87, 1381–1397, https://doi.org/10.1175/BAMS-87-10-1381, 2006.

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

Domenico, B., Caron, J., Davis, E., Kambic, R., and Nativi, S.: Thematic real-time environmental distributed data services (thredds): Incorporating interactive analysis tools into nsdl, Journal of Digital Information, 2, available at: https://journals.tdl.org/jodi/index.php/jodi/article/viewArticle/51 (last access: 12 January 2016), 2006.

d'Orgeval, T., Polcher, J., and de Rosnay, P.: Sensitivity of the West African hydrological cycle in ORCHIDEE to infiltration processes, Hydrol. Earth Syst. Sci., 12, 1387–1401, https://doi.org/10.5194/hess-12-1387-2008, 2008.

Dorigo, W., Jeu, R., Chung, D., Parinussa, R., Liu, Y., Wagner, W., and Fernández-Prieto, D.: Evaluating global trends (1988–2010) in harmonized multi-satellite surface soil moisture, Geophys. Res. Lett., 39, 18, https://doi.org/10.1029/2012GL052988, 2012.

Dorigo, W. A., Gruber, A., De Jeu, R. A. M., Wagner, W., Stacke, T., Loew, A., Albergel, C., Brocca, L., Chung, D., Parinussa, R., and Kidd, R.: Evaluation of the ESA CCI soil moisture product using ground-based observations, Remote Sens. Environ., 162, 380–395, 2015.

Faures, J. M.: Mapping Existing Global Systems & Initiatives Background Document – August 2006, prepared by FAO on behalf of the UN-Water Task Force on Monitoring Stockholm, 21 August 2006.

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

Gosling, S. N., Bretherton, D., Haines, K., and Arnell, N. W.: Global hydrology modelling and uncertainty: running multiple ensembles with a campus grid, Philos. T. R. Soc. Lond. A, 368, 4005–4021, https://doi.org/10.1098/rsta.2010.0164, 2010.

Haddeland, I., Clark, D. B., Franssen, W., Ludwig, F., Voß, F., Arnell, N. W., Bertrand, N., Best, M., Folwell, S., Gerten, D., Gomes, S., Gosling, S. N., Hagemann, S., Hanasaki, N., Harding, R., Heinke, J., Kabat, P., Koirala, S., Oki, T., Polcher, J., Stacke, T., Viterbo, P., Weedon, G. P., and Yeh, P.: Multimodel Estimate of the Global Terrestrial Water Balance: Setup and First Results, J. Hydrometeorol., 12, 869–884, https://doi.org/10.1175/2011JHM1324.1, 2011.

Hansen, J., Sato, M., and Ruedy, R.: Perception of climate change, P. Natl. Acad. Sci. USA, 109, E2415–E2423, 2012.

Harding, R., Best, M., Blyth, E., Hagemann, S., Kabat, P., Tallaksen, L. M., Warnaars, T., Wiberg, D., Weedon, G. P., van Lanen, H., Ludwig, F., and Haddeland, I.: WATCH: Current knowledge of the terrestrial global water cycle, J. Hydrometeorol., 12, 1149–1156, https://doi.org/10.1175/JHM-D-11-024.1, 2011.

Hargreaves, G. H. and Allen, R. G.: History and evaluation of Hargreaves evapotranspiration equation, J. Irrig. Drain. E.-ASCE, 129, 53–63, https://doi.org/10.1061/(ASCE)0733-9437(2003)129:1(53), 2003.

Hargreaves, G. H. and Samani, Z. A.: Estimating potential evapotranspiration, Journal of the Irrigation and Drainage Division, 108, 225–230, 1982.

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

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

Helfrich, S. R., McNamara, D., Ramsay, B. H., Baldwin, T., and Kasheta, T.: Enhancements to, and forthcoming developments in the Interactive Multisensor Snow and Ice

Mapping System (IMS), Hydrol. Process., 21, 1576–1586, https://doi.org/10.1002/hyp.6720, 2007.

Houser, P. R., Rodell, M., Jambor, U., Gottschalck, J., Cosgrove, B., Radakovich, J., Arsenault, K., Bosilovich, M., Entin, J. K., Walker, J. P., Mitchell, K., Pan, H. L., and Meng, C.-J.: The global land data assimilation system, GEWEX News, 11, 11–13, available at: http://users.monash.edu.au/~jpwalker/papers/tmp/GEWEX_news_01.pdf (last access: 5 March 2017), 2001.

Kim, H., Yeh, P. J.-F., Oki, T., and Kanae, S.: Role of rivers in the seasonal variations of terrestrial water storage over global basins, Geophys. Res. Lett., 36, 17, https://doi.org/10.1029/2009GL039006, 2009.

Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., Onogi, K., Kamahori, H., Kobayashi, C., Endo, H., Miyaoka, K., and Takahashi, K.: The JRA-55 reanalysis: general specifications and basic characteristics, J. Meteorol. Soc. Jpn., 93, 5–48, https://doi.org/10.2151/jmsj.2015-001, 2015.

Krinner, G., Viovy, N., N., de Noblet-Ducoudré, N., Ogée, J., Polcher, J., F riedlingstein, P., Ciais, P., Stich, S., and Prentice, I. C.: A dynamic global vegetation model for studies of the coupled atmosphere-biosphere system, Global Biogeochem. Cy., 19, 1, https://doi.org/10.1029/2003GB002199, 2005.

Kumar, A. and Hoerling, M. P.: Analysis of a conceptual model of seasonal climate variability and implications for seasonal prediction, B. Am. Meteorol. Soc., 81, 255–264, 2000.

Landerer, F. and Swenson, S.: Accuracy of scaled GRACE terrestrial water storage estimates, Water Resour. Res., 48, 4, https://doi.org/10.1029/2011WR011453, 2012.

Li, M. and Ma, Z.: Comparisons of simulations of soil moisture variations in the Yellow River basin driven by various atmospheric forcing data sets, Adv. Atmos. Sci., 27, 1289–1302, https://doi.org/10.1007/s00376-010-9155-7, 2010.

Lindström, G., Johansson, B., Persson, M., Gardelin, M., and Bergström, S.: Development and test of the distributed HBV-96 hydrological model, J. Hydrol., 201, 272–288, https://doi.org/10.1016/S0022-1694(97)00041-3, 1997.

Liu, Y. Y., Dorigo, W. A., Parinussa, R. M., de Jeu, R. A. M., Wagner, W., McCabe, M. F., Evans, J. P., and van Dijk, A. I. J. M.: Trend-preserving blending of passive and active microwave soil moisture retrievals, Remote Sens. Environ., 123, 280–297, https://doi.org/10.1016/j.rse.2012.03.014, 2012.

Loew, A., Stacke, T., Dorigo, W., de Jeu, R., and Hagemann, S.: Potential and limitations of multidecadal satellite soil moisture observations for selected climate model evaluation studies, Hydrol. Earth Syst. Sci., 17, 3523–3542, https://doi.org/10.5194/hess-17-3523-2013, 2013.

Long, D., Longuevergne, L., and Scanlon, B. R.: Uncertainty in evapotranspiration from land surface modeling, remote sensing, and GRACE satellites, Water Resour. Res., 50, 1131–1151, https://doi.org/10.1002/2013WR014581, 2014.

López López, P., Wanders, N., Schellekens, J., Renzullo, L. J., Sutanudjaja, E. H., and Bierkens, M. F. P.: Improved large-scale hydrological modelling through the assimilation of streamflow and downscaled satellite soil moisture observations, Hydrol. Earth Syst. Sci., 20, 3059–3076, https://doi.org/10.5194/hess-20-3059-2016, 2016.

Luo, Y. Q., Randerson, J. T., Abramowitz, G., Bacour, C., Blyth, E., Carvalhais, N., Ciais, P., Dalmonech, D., Fisher, J. B., Fisher, R., Friedlingstein, P., Hibbard, K., Hoffman, F., Huntzinger, D., Jones, C. D., Koven, C., Lawrence, D., Li, D. J., Mahecha, M., Niu, S. L., Norby, R., Piao, S. L., Qi, X., Peylin, P., Prentice, I. C., Riley, W., Reichstein, M., Schwalm, C., Wang, Y. P., Xia, J. Y., Zaehle, S., and Zhou, X. H.: A framework for benchmarking land models, Biogeosciences, 9, 3857–3874, https://doi.org/10.5194/bg-9-3857-2012, 2012.

Martens, B., Miralles, D. G., Lievens, H., van der Schalie, R., de Jeu, R. A. M., Fernández-Prieto, D., Beck, H. E., Dorigo, W. A., and Verhoest, N. E. C.: GLEAM v3: satellite-based land evaporation and root-zone soil moisture, Geosci. Model Dev., 10, 1903–1925, https://doi.org/10.5194/gmd-10-1903-2017, 2017.

McDonald, R. I., Douglas, I., Revenga, C., Hale, R., Grimm, N., Grönwall, J., and Fekete, B.: Global Urban Growth and the Geography of Water Availability, Quality, and Delivery, AMBIO, 40, 437–446, https://doi.org/10.1007/s13280-011-0152-6, 2011.

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

Miralles, D. G., Jiménez, C., Jung, M., Michel, D., Ershadi, A., McCabe, M. F., Hirschi, M., Martens, B., Dolman, A. J., Fisher, J. B., Mu, Q., Seneviratne, S. I., Wood, E. F., and Fernández-Prieto, D.: The WACMOS-ET project – Part 2: Evaluation of global terrestrial evaporation data sets, Hydrol. Earth Syst. Sci., 20, 823–842, https://doi.org/10.5194/hess-20-823-2016, 2016.

Moore, R. J.: The PDM rainfall-runoff model, Hydrol. Earth Syst. Sci., 11, 483–499, https://doi.org/10.5194/hess-11-483-2007, 2007.

Mu, M., Randerson, J. T., Riley, W. J., and Hoffman, F. M.: International land Model Benchmarking (ILAMB) Package v001. 00, https://doi.org/10.18139/ILAMB.v001.00/1251597, 2016.

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

Mulligan, M.: WaterWorld: a self-parameterising, physically-based model for application in data-poor but problem-rich environments globally, Hydrol. Res., 44, 748–769, 2012.

Ngo-Duc, T., Laval, K., Ramillien, G., Polcher, J., and Cazenave, A.: Validation of the land water storage simulated by Organising Carbon and Hydrology in Dynamic Ecosystems (ORCHIDEE) with Gravity Recovery and Climate Experiment (GRACE) data, Water Resour. Res., 43, 4, https://doi.org/10.1029/2006WR004941, 2007.

Nikolopoulos, E. I., Anagnostou, E. N., Hossain, F., Gebremichael, M., and Borga, M.: Understanding the scale relationships of uncertainty propagation of satellite rainfall through a distributed hydrologic model, J. Hydrometeorol., 11, 520–532, https://doi.org/10.1175/2009JHM1169.1, 2010.

Orth, R. and Seneviratne, S. I.: Predictability of soil moisture and streamflow on subseasonal timescales: A case study, J. Geophys. Res.-Atmos., 118, 10–963, https://doi.org/10.1002/jgrd.50846, 2013.

Orth, R., Koster, R. D., and Seneviratne, S. I.: Inferring soil moisture memory from streamflow observations using a simple water balance model, J. Hydrometeorol., 14, 1773–1790, https://doi.org/10.1175/JHM-D-12-099.1, 2013.

Pozzi, W., Sheffield, J., Stefanski, R., Cripe, D., Pulwarty, R., Vogt, J. V., Heim Jr., R. R., Brewer, M. J., Svoboda, M., Westerhoff, R., van Dijk, A. I . J. M., Lloyd-Hughes, B., Pappenberger, F., Werner, M., Dutra, E., Wetterhall, F., Wagner, W., Schubert, S., Mo, K., Nicholson, M., Bettio, L., Nunez, L., van Beek, R., Bierkens, M., Gustavo Goncalves de Goncalves, L., de Mattos, J. G. Z., and Lawford, R.: Toward Global Drought Early Warning Capability: Expanding International Cooperation for the Development of a Framework for Monitoring and Forecasting, B. Am. Meteorol. Soc., 94, 776–785, https://doi.org/10.1175/BAMS-D-11-00176.1, 2013.

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

Pulliainen, J.: Mapping of snow water equivalent and snow depth in boreal and sub-arctic zones by assimilating space-borne microwave radiometer data and ground-based observations, Remote Sens. Environ., 101, 257–269, 2006.

Ramsay, B. H.: The interactive multisensor snow and ice mapping system, Hydrol. Process., 12, 1537–1546, 1998.

Reichle, R.: The MERRA-land data product, The MERRA-Land Data Product. GMAO Office Note 3, 38 pp., available at: http://gmao.gsfc.nasa.gov/pubs/docs/Reichle541.pdf, last access: 10 May 2015.

Riegger, J., Tourian, M. J., Devaraju, B., and Sneeuw, N.: Analysis of grace uncertainties by hydrological and hydro-meteorological observations, J. Geodynam., 59, 16–27, 2012.

Rienecker, M. M., Suarez, M. J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., Bosilovich, M. G., Schubert, S. D., Takacs, L., Kim, G.-K., Bloom, S., Chen, J., Collins, D., Conaty, A., da Silva, A., Gu, W., Joiner, J., Koster, R. D., Lucchesi, R., Molod, A., Owens, T., Pawson, S., Pegion, P., Redder, C. R., Reichle, R., Robertson, F. R., Ruddick, A. G., Sienkiewicz, M., and Woollen, J.: MERRA: NASA's modern-era retrospective analysis for research and applications, J. Climate, 24, 3624–3648, https://doi.org/10.1175/jcli-d-11-00015.1, 2011.

Rijsberman, F. R.: Water scarcity: Fact or fiction?, Agr. Water Manage., 80, 5–22, https://doi.org/10.1016/j.agwat.2005.07.001, 2006.

Rodell, M., Houser, P. R., Jambor, U. E. A., Gottschalck, J., Mitchell, K., Meng, C. J., Arsenault, K., Cosgrove, B., Radakovich, J., Bosilovich, M., Entin, J. K., Walker, J. P., Lohmann, D., and Toll, D.: The global land data assimilation system, B. Am. Meteorol. Soc., 85, 381–394, https://doi.org/10.1175/BAMS-85-3-381, 2004.

Rodell, M., Beaudoing, H. K., L'Ecuyer, T. S., Olson, W. S., Famiglietti, J. S., Houser, P. R., Adler, R., Bosilovich, M. G., Clayson, C. A., Chambers, D., Clark, E., Fetzer, E. J., Gao, X., Gu, G., Hilburn, K., Huffman, G. J., Lettenmaier, D. P., Liu, W. T., Robertson, F. R., Schlosser, C. A., Sheffield, J., and Wood, E. F.: The observed state of the water cycle in the early twenty-first century, J. Climate, 28, 8289–8318, https://doi.org/10.1175/JCLI-D-14-00555.1, 2015.

Rust, H. W., Kruschke, T., Dobler, A., Fischer, M., and Ulbrich, U.: Discontinuous Daily Temperatures in the WATCH Forcing Datasets, J. Hydrometeorol., 16, 465–472, https://doi.org/10.1175/JHM-D-14-0123.1, 2014.

Schneider, U., Becker, A., Finger, P., Meyer-Christoffer, A., Rudolf, B., and Ziese, M.: GPCC full data reanalysis version 6.0 at 0.5: monthly land-surface precipitation from raingauges built on GTS-based and historic data, , FD_M_V6_050, https://doi.org/10.5676/DWD_GPCC , 2011.

Sheffield, J., Goteti, G., and Wood , E. F.: Development of a 50-year high-resolution global dataset of meteorological forcings for land surface modeling, J. Climate, 19, 3088–3111, doi10.1175/JCLI3790.1, 2006.

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

Sperna Weiland, F. C., Tisseuil, C., Dürr, H. H., Vrac, M., and van Beek, L. P. H.: Selecting the optimal method to calculate daily global reference potential evaporation from CFSR reanalysis data for application in a hydrological model study, Hydrol. Earth Syst. Sci., 16, 983–1000, https://doi.org/10.5194/hess-16-983-2012, 2012.

Su, C.-H., Ryu, D., Dorigo, W., Zwieback, S., Gruber, A., Albergel, C., Reichle, R. H., and Wagner, W.: Homogeneity of a global multi-satellite soil moisture climate data record, Geophys. Res. Lett., 43, 21, https://doi.org/10.1002/2016GL070458, 2016.

Syed, T. H., Famiglietti, J. S., and Chambers, D. P.: GRACE-based estimates of terrestrial freshwater discharge from basin to continental scales, J. Hydrometeorol., 10, 22–40, https://doi.org/10.1175/2008JHM993.1, 2009.

Takala, M., Luojus, K., Pulliainen, J., Derksen, C., Lemmetyinen, J., Kärnä, J.-P., Koskinen, J., and Bojkov, B.: Estimating northern hemisphere snow water equivalent for climate research through assimilation of space-borne radiometer data and ground-based measurements, Remote Sens. Environ., 115, 3517–3529, 2011.

Trenberth, K. E., Smith, L., Qian, T., Dai, A., and Fasullo, J.: Estimates of the Global Water Budget and Its Annual Cycle Using Observational and Model Data, J. Hydrometeorol., 8, 758–769, https://doi.org/10.1175/JHM600.1, 2007.

Trigg, M. A., Birch, C. E., Neal, J. C., Bates, P. D., Smith, A., Sampson, C. C., Yamazaki, D., Hirabayashi, Y., Pappenberger, F., Dutra, E., Ward, P. J., Winsemius, H. C., Salamon, P., Dottori, F., Rudari, R., Kappes, M. S., Simpson, A. L., Hadzilacos, G., and Fewtrell, T. J.: The credibility challenge for global fluvial flood risk analysis, Environ. Res. Lett., 11, 094014, https://doi.org/10.1088/1748-9326/11/9/094014, 2016.

UN: UN Water: Task Force on Indicators, Monitoring and Reporting (2008–2010), available at: http://www.unwater.org/activities/task-forces/indicators/en/, last access: 2016.

Uppala, S. M., Kallberg, P. W., Simmons, A. J., Andrae, U., Bechtold, V., Fiorino, M., Gibson, J. K., Haseler, J., Hernandez, A., Kelly, G. A., Li, X., Onogi, K., Saarinen, S., Sokka, N., Allan, R. P., Andersson, E., Arpe, K., Balmaseda, M. A., Beljaars, A. C. M., Van De Berg, L., Bidlot, J., Bormann, N., Caires, S., Chevallier, F., Dethof, A., Dragosavac, M., Fisher, M., Fuentes, M., Hagemann, S., Hólm, E., Hoskins, B. J., Isaksen, L., Janssen, P. A. E. M., Jenne, R., Mcnally, A. P., Mahfouf, J.-F., Morcrette, J.-J., Rayner, N. A., Saunders, R. W., Simon, P., Sterl, A., Trenberth, K. E., Untch, A., Vasiljevic, D., Viterbo, P., and Woollen, J.: The ERA-40 re-analysis, Q. J. Roy. Meteor. Soc., 131, 2961–3012, https://doi.org/10.1256/qj.04.176, 2005.

van Beek, L. P. H., Wada, Y., and Bierkens, M. F. P.: Global monthly water stress: 1. Water balance and water availability, Water Resour. Res., 47, w07517, https://doi.org/10.1029/2010WR009791, 2011.

Van Der Knijff, J. M., Younis, J., and De Roo, A. P. J.: LISFLOOD: a GIS-based distributed model for river basin scale water balance and flood simulation, Int. J. Geogr. Inf. Sci., 24, 189–212, https://doi.org/10.1080/13658810802549154, 2010.

Van Dijk, A.: The Australian water resources assessment system, Version 0.5, 3, available at: http://www.clw.csiro.au/publications/waterforahealthycountry/2010/wfhc-awras-evaluation-against-observations.pdf (last access: 2 June 2015), 2010.

van Dijk, A. I. J. M., Renzullo, L. J., Wada, Y., and Tregoning, P.: A global water cycle reanalysis (2003–2012) merging satellite gravimetry and altimetry observations with a hydrological multi-model ensemble, Hydrol. Earth Syst. Sci., 18, 2955–2973, https://doi.org/10.5194/hess-18-2955-2014, 2014.

Van Vliet, M. T. H., Ludwig, F., Zwolsman, J. J. G., Weedon, G. P., and Kabat, P.: Global river temperatures and sensitivity to atmospheric warming and changes in river flow, Water Resour. Res., 47, 2, https://doi.org/10.1029/2010WR009198, 2011.

van Vliet, M. T. H., van Beek, L. P. H., Eisner, S., Flörke, M., Wada, Y., and Bierkens, M. F. P.: Multi-model assessment of global hydropower and cooling water discharge potential under climate change, Global Environ. Change, 40, 156–170, https://doi.org/10.1016/j.gloenvcha.2016.07.007, 2016.

Vrugt, J. A., Diks, C. G., Gupta, H. V., Bouten, W., and Verstraten, J. M.: Improved treatment of uncertainty in hydrologic modeling: Combining the strengths of global optimization and data assimilation, Water Resour. Res., 41, 1, https://doi.org/10.1029/2004WR003059, 2005.

Wada, Y., Wisser, D., and Bierkens, M. F. P.: Global modeling of withdrawal, allocation and consumptive use of surface water and groundwater resources, Earth Syst. Dynam., 5, 15–40, https://doi.org/10.5194/esd-5-15-2014, 2014.

Wang, A., Lettenmaier, D. P., and Sheffield, J.: Soil moisture drought in China, 1950–2006, J. Climate, 24, 3257–3271, 2011. 2014.

Weedon, G. P., Gomes, S., Viterbo, P., Shuttleworth, W. J., Blyth, E., Österle, H., Adam, J. C., Bellouin, N., Boucher, O., and Best, M.: Creation of the WATCH Forcing Data and Its Use to Assess Global and Regional Reference Crop Evaporation over Land during the Twentieth Century, J. Hydrometeorol., 12, 823–848, https://doi.org/10.1175/2011JHM1369.1, 2011.

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

Weiland, F. S., Lopez, P., Van Dijk, A., and Schellekens, J.: Global high-resolution reference potential evaporation, in: MODSIM 2015, Conference Proceedings, Broadbeach, Queensland, Australia, available at: http://www.mssanz.org.au/modsim2015/OZEWEX/spernaweiland.pdf (last access: 6 May 2016), 2015.

Xie, P. and Arkin, P. A.: Global precipitation: A 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs, B. Am. Meteorol. Soc., 78, 2539–2558, 1997.

Yamazaki, D., Kanae, S., Kim, H., and Oki, T.: A physically based description of floodplain inundation dynamics in a global river routing model, Water Resour. Res., 47, 4, https://doi.org/10.1029/2010WR009726, 2011.

Zscheischler, J., Orth, R., and Seneviratne, S. I.: A submonthly database for detecting changes in vegetation-atmosphere coupling, Geophys. Res. Lett., 42, 9816–9824, 2015.