University of Reading

Multimodel estimate of the global terrestrial water balance: setup and first results

Article

Published Version

Haddeland, I., Clark, D. B., Franssen, W., Ludwig, F., Voß, F., Arnell, N. W. ORCID: https://orcid.org/0000-0003-2691-4436, Bertrand, N., Best, M., Folwell, S., Kabat, P., Koirala, S., Oki, T., Polcher, J., Stacke, T., Viterbo, P., Weedon, G. P., Yehm, P., Gerten, D., Gomes, S., Gosling, S. N., Hagemann, S., Hanasaki, N., Harding, R. and Heinke, J. (2011) Multimodel estimate of the global terrestrial water balance: setup and first results. Journal of Hydrometeorology, 12 (5). pp. 869-884. ISSN 1525-7541 doi: https://doi.org/10.1175/2011JHM1324.1 Available at https://centaur.reading.ac.uk/24011/

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

To link to this article DOI: http://dx.doi.org/10.1175/2011JHM1324.1

Publisher: American Meteorological Society

Publisher statement: © Copyright 2011 American Meteorological Society (AMS). Permission to use figures, tables, and brief excerpts from this work in scientific and educational works is hereby granted provided that the source is acknowledged. Any use of material in this work that is determined to be "fair use" under Section 107 of the U.S. Copyright Act or that satisfies the conditions specified in Section 108 of the U.S. Copyright Act (17USC §108, as revised by

P.L. 94-553) does not require the AMS's permission. Republication, systematic reproduction, posting in electronic form, such as on a web site or in a searchable database, or other uses of this material, except as exempted by the above statement, requires written permission or a license from the AMS. Additional details are provided in the AMS Copyright Policy, available on the AMS Web site located at (http://www.ametsoc.org/) or from the AMS at 617-227-2425 or copyright@ametsoc.org.

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the End User Agreement.

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

Multimodel Estimate of the Global Terrestrial Water Balance: Setup and First Results

INGJERD HADDELAND,[a] DOUGLAS B. CLARK,[b] WIETSE FRANSSEN,[c] FULCO LUDWIG,[c] FRANK VOß,[d] NIGEL W. ARNELL,[e] NATHALIE BERTRAND,[f] MARTIN BEST,[g] SONJA FOLWELL,[b] DIETER GERTEN,[h] SANDRA GOMES,[i] SIMON N. GOSLING,[j] STEFAN HAGEMANN,[k] NAOTA HANASAKI,[l] RICHARD HARDING,[b] JENS HEINKE,[h] PAVEL KABAT,[c] SUJAN KOIRALA,[m] TAIKAN OKI,[m] JAN POLCHER,[f] TOBIAS STACKE,[k] PEDRO VITERBO,[i] GRAHAM P. WEEDON,[g] and PAT YEH[m]

[a] *Norwegian Water Resources and Energy Directorate, Oslo, Norway, and Wageningen University and Research Centre, Wageningen, Netherlands*
[b] *Centre for Ecology and Hydrology, Wallingford, United Kingdom*
[c] *Wageningen University and Research Centre, Wageningen, Netherlands*
[d] *Center for Environmental Systems Research, University of Kassel, Kassel, Germany*
[e] *Walker Institute for Climate System Research, University of Reading, Reading, United Kingdom*
[f] *Laboratorie de Meteorologie Dynamique, Paris, France*
[g] *Met Office Hadley Centre, Joint Centre for Hydrometeorological Research, Wallingford, United Kingdom*
[h] *Potsdam Institute for Climate Research, Potsdam, Germany*
[i] *Centro de Geofisica da Universidade de Lisboa, Lisbon, Portugal*
[j] *School of Geography, University of Nottingham, Nottingham, United Kingdom*
[k] *Max Planck Institute for Meteorology, Hamburg, Germany*
[l] *National Institute for Environmental Studies, Tsukuba, Japan*
[m] *University of Tokyo, Tokyo, Japan*

(Manuscript received 4 June 2010, in final form 4 February 2011)

ABSTRACT

Six land surface models and five global hydrological models participate in a model intercomparison project [Water Model Intercomparison Project (WaterMIP)], which for the first time compares simulation results of these different classes of models in a consistent way. In this paper, the simulation setup is described and aspects of the multimodel global terrestrial water balance are presented. All models were run at 0.5° spatial resolution for the global land areas for a 15-yr period (1985–99) using a newly developed global meteorological dataset. Simulated global terrestrial evapotranspiration, excluding Greenland and Antarctica, ranges from 415 to 586 mm yr^{-1} (from 60 000 to 85 000 km^3 yr^{-1}), and simulated runoff ranges from 290 to 457 mm yr^{-1} (from 42 000 to 66 000 km^3 yr^{-1}). Both the mean and median runoff fractions for the land surface models are lower than those of the global hydrological models, although the range is wider. Significant simulation differences between land surface and global hydrological models are found to be caused by the snow scheme employed. The physically based energy balance approach used by land surface models generally results in lower snow water equivalent values than the conceptual degree-day approach used by global hydrological models. Some differences in simulated runoff and evapotranspiration are explained by model parameterizations, although the processes included and parameterizations used are not distinct to either land surface models or global hydrological models. The results show that differences between models are a major source of uncertainty. Climate change impact studies thus need to use not only multiple climate models but also some other measure of uncertainty (e.g., multiple impact models).

Corresponding author address: Ingjerd Haddeland, Norwegian Water Resources and Energy Directorate, P.O. Box 5091 Maj., 0301 Oslo, Norway.
E-mail: ingjerd.haddeland@nve.no

DOI: 10.1175/2011JHM1324.1

© 2011 American Meteorological Society

1. Introduction

The global water balance has been the subject of modeling studies for decades, both from a climate perspective where the main interest is the influence of the water balance on surface heat fluxes and from a hydrological perspective focusing on water availability and use. However, there are still many uncertainties in our understanding of the current water cycle, and to date the results of land surface models (LSMs) and global hydrology models (GHMs) have not been compared in a consistent way. LSMs, which can be coupled to atmospheric models, tend to describe the vertical exchanges of heat, water, and sometimes carbon, in considerable detail. In contrast, GHMs are traditionally more focused on water resources and lateral transfer of water.

There have been several previous model intercomparisons: for example, Spatial Variability of Land Surface Processes (SLAPS) (Polcher et al. 1996), the Project to Intercompare Land surface Parameterization Schemes (PILPS) (Henderson-Sellers et al. 1995; Pitman and Henderson-Sellers 1998), and the Global Soil Wetness Project (GSWP) (Dirmeyer et al. 1999, 2006). The focus in these projects has been on LSMs and the simulations of surface water and energy balances. Results on water availability and stress from different GHMs have appeared in the scientific literature (e.g., Alcamo et al. 2003; Arnell 2004), as have results on anthropogenic water uses at the global scale (e.g., Döll and Siebert 2002; Hanasaki et al. 2008b; Rost et al. 2008). However, comparison of these numbers, their uncertainties, and the causes thereof has been limited. The GHM community has recently started the process of systematically compiling and comparing results through the GWSP and the Green Blue Water Initiative (Voß et al. 2008; Hoff et al. 2010).

The Water and Global Change (WATCH) project, funded under the European Union (EU) Sixth Framework Programme (FP6), brings together the hydrological, water resources, and climate communities to analyze, quantify, and predict the components of the current and future global water cycles and related water resources states. An important part of WATCH is a model intercomparison project in which both LSMs and GHMs participate. WATCH and GWSP have recently combined their model intercomparison efforts in a joint project called the Water Model Intercomparison Project (WaterMIP). WaterMIP includes both LSMs and GHMs, and many of the participating models include the possibility of taking into account anthropogenic impacts such as water withdrawals and dams. Hence, WaterMIP provides an opportunity to compare results of LSMs and GHMs, focusing on differences between the two model strategies, while additionally investigating the effects of anthropogenic impacts on the global terrestrial water balance. Estimates of water availability and stress, as well as the uncertainties thereof, will also be compared for both current and future conditions. Using a range of model simulations, the aim is to improve our understanding of current and future water availability and water stress at the global scale, with an emphasis on the available water resources of major river systems at the subannual time scale. Water demands involve strong seasonal variations; hence, both annual water volumes and seasonal timing are important factors. Through integrated model intercomparison and evaluation, participating models will improve the parameterization of human interactions with the global terrestrial water cycle. In related activities within WATCH, global consumptive water use in different sectors—not only for irrigation but also for domestic, manufacturing, and livestock farming purposes—will be considered.

This paper is the first in a series presenting the results of WaterMIP. It gives an overview of the participating models, describes the experimental setup, and discusses the results of naturalized model simulations (i.e., without taking water management like reservoirs and water withdrawals into account) for historic climate. It also identifies reasons for some of the differences between model results. Understanding how the models perform differently for naturalized conditions and current climate provides important information with which to understand why some models might respond differently in future runs using climate projections. The models participating in WaterMIP cover a wide range of characteristics, ranging from physically based models run at subhourly time steps to more conceptual models run at daily time steps. An objective of WaterMIP is to bring together researchers from the climate and water resources communities, because there have been few comparisons of water balance results between these communities. The main hypothesis tested in this paper is whether there is a consistent difference in simulations of the global terrestrial water cycle between LSMs and GHMs. Explaining all the differences is beyond the scope of this paper. Subsequent papers will present results of model simulations including human influences and the impacts of climate change on global water resources.

2. Simulation setup and model descriptions

In this first stage of WaterMIP, we assess the components of the contemporary global terrestrial water balance under naturalized conditions: that is, human impacts such as storage in man-made reservoirs and

TABLE 1. Participating models, including their main characteristics.

Model name[a]	Model time step	Meteorological forcing variables[b]	Energy balance	ET scheme[c]	Runoff scheme[d]	Snow scheme	Reference(s)
GWAVA	Daily	P, T, W, Q, LW_{net}, SW, SP	No	Penman–Monteith	Saturation excess/beta function	Degree-day	Meigh et al. 1999
H08	6 h	R, S, T, W, Q, LW, SW, SP	Yes	Bulk formula	Saturation excess/beta function	Energy balance	Hanasaki et al. 2008a
HTESSEL	1 h	R, S, T, W, Q, LW, SW, SP	Yes	Penman–Monteith	Infiltration excess/Darcy	Energy balance	Balsamo et al. 2009
JULES	1 h	R, S, T, W, Q, LW, SW, SP	Yes	Penman–Monteith	Infiltration excess/Darcy	Energy balance	Cox et al. 1999; Essery et al. 2003
LPJmL	Daily	P, T, LW_{net}, SW	No	Priestley–Taylor	Saturation excess	Degree-day	Bondeau et al. 2007; Rost et al. 2008
MacPDM	Daily	P, T, W, Q, LW_{net}, SW	No	Penman–Monteith	Saturation excess/beta function	Degree-day	Arnell 1999; Gosling and Arnell 2010
MATSIRO	1 h	R, S, T, W, Q, LW, SW, SP	Yes	Bulk formula	Infiltration and saturation excess/groundwater	Energy balance	Takata et al. 2003; Koirala 2010
MPI-HM	Daily	P, T	No	Thornthwaite	Saturation excess/beta function	Degree-day	Hagemann and Gates 2003; Hagemann and Dümenil 1998
Orchidee	15 min	R, S, T, W, Q, SW, LW, SP	Yes	Bulk formula	Saturation excess	Energy balance	De Rosnay and Polcher 1998
VIC	Daily/3h	P, T_{max}, T_{min}, W, Q, LW, SW, SP	Snow season	Penman–Monteith	Saturation excess/beta function	Energy balance	Liang et al. 1994
WaterGAP	Daily	P, T, LW_{net}, SW	No	Priestley–Taylor	Beta function	Degree-day	Alcamo et al. 2003

[a] Model names written in bold are classified as LSMs in this paper; the other models are classified as GHMs.
[b] R = rainfall rate; S = snowfall rate; P = precipitation (rain or snow distinguished in the model); T = air temperature; T_{max} = maximum daily air temperature; T_{min} = minimum daily air temperature; W = wind speed; Q = specific humidity; LW = longwave radiation flux (downward); LW_{net} = longwave radiation flux (net); SW = shortwave radiation flux (downward); and SP = surface pressure.
[c] Bulk formula: Bulk transfer coefficients are used when calculating the turbulent heat fluxes.
[d] Beta function: Runoff is a nonlinear function of soil moisture.

agricultural water withdrawal are not included in the model runs. The spatial resolution of the forcing data and the model simulations is 0.5° in latitude and longitude, covering the land area defined by the Climate Research Unit of the University of East Anglia (CRU) global land mask. The land mask does not include Antarctica. Models that include lateral routing of streamflow all use the DDM30 routing network (Döll and Lehner 2002), which was slightly modified to match the CRU land mask. A total of 11 models participated in this round of WaterMIP (see Table 1, which includes a description of the models' main characteristics). The models use their default soil and vegetation information and no attempt was made to standardize these parameters.

A key difference between the models is whether they solve both the water and the energy balances at the land surface or only the water balance. The models that solve the energy balance have to be run using a subdaily time step, whereas participating models run in water balance mode alone all run with daily time steps. The models differ in their choice of evapotranspiration (ET) and runoff schemes (see Table 1) and vary substantially in complexity. For example, there are differences in the number of components of evapotranspiration that are considered: for example, interception evaporation, vegetation transpiration, open-water evaporation, and the level of detail given to vegetation description and processes. Other model differences concern the complexity of the representation of runoff processes, groundwater, snow, and frozen soil. The snow schemes are based on either the degree-day approach, which is used by all models run at daily time step, or an energy balance approach, which is used by all models run at subdaily time steps. Detailed information on each participating model can be found in the references listed in Table 1. Although, traditionally, LSMs have been developed within the climate community and GHMs have been developed within the hydrologic community, there are similarities in particular areas between individual models from the different groups; thus, the grouping shown in Table 1 is a useful device but is not necessarily definitive. Other classifications are undoubtedly possible by other aspects of the

models, but, in this paper, models that can solve both the surface energy and water balances are classified as LSMs, whereas the models solving the water balance only are classified as GHMs. This differentiation means that six LSMs and five GHMs participated in this round of WaterMIP. The sample size is fairly small, and certain results are strongly affected by results from a subset of models; consequently, analyses based on individual models and grouped results are both presented.

All models used the meteorological data described by Weedon et al. (2010, 2011), but they do not all use the same variables or model time step (Table 1). The meteorological data, called the WATCH forcing data (Weedon et al. 2010, 2011), are available at both daily and subdaily time steps. The WATCH forcing variables are taken from the 40-yr European Centre for Medium-Range Weather Forecasts (ECMWF) Re-Analysis (ERA-40) as described by Uppala et al. (2005). The 1° ERA-40 product was interpolated to ½° resolution on the CRU land mask, adjusted for elevation changes where needed, and bias corrected using monthly observations. Temperature, surface pressure, specific humidity, and downward longwave radiation were adjusted sequentially in that order because they are interdependent via the elevation adjustment. Diurnal air temperature was bias corrected with CRU data (New et al. 1999, 2000; Mitchell and Jones 2005). Shortwave downward radiation (SW) was corrected using CRU cloud cover fractions, having found the gridpoint-specific correlations between monthly average SW and ERA-40 cloud fraction. SW was also adjusted in clear sky and cloudy sky for the effects of tropospheric and stratospheric aerosol loading. Precipitation was adjusted using both a wet-day correction from CRU and precipitation totals from the GPCCv4 full data product (Rudolf and Schneider 2005; Schneider et al. 2008; Fuchs 2009), and it was corrected for undercatch (snowfall and rainfall separately) based on Adam and Lettenmaier (2003). For detailed information on the forcing data, see Weedon et al. (2010, 2011).

The simulation period is 1985–99, preceded by a spinup period of at least 5 yr, and results are submitted at a monthly time scale. Requested output variables include the main water balance states and fluxes, and components of these fluxes (e.g., interception evaporation and vegetation transpiration). The variables were submitted in Network Common Data Format (NetCDF), following the definitions and units of the Assistance for Land surface Modeling Activities (ALMA) data convention (Polcher et al. 2000). The modeling protocol, including detailed information on requested variables, is available online (at http://www.eu-watch.org/watermip). A single model, WaterGAP, has applied a correction factor on cell runoff to match observed river discharge, and evapotranspiration is adjusted accordingly. All the other participating models are uncalibrated for this exercise, although they may have been calibrated for previous studies.

3. Results and discussion

a. Global analyses

Mean annual averages of total precipitation, evapotranspiration, and runoff fraction and the coefficient of variation (CV; i.e., standard deviation divided by the mean) of the model means of snowfall, evapotranspiration, and runoff are presented in Fig. 1. Global terrestrial mean precipitation in the period 1985–99 was, according to the WATCH forcing data, 872 mm yr^{-1} (or 126 000 km^3 yr^{-1}). A few models reduce precipitation when simulated snow water equivalent (SWE) exceeds a given level, which influences precipitation numbers in some northern areas. Because of this and because very few models include a glacier scheme, the numbers presented in this paper do not include Greenland. However, Greenland is included in all model simulations and included in the maps in Fig. 1. The global land area is calculated assuming that the earth is a sphere with radius 6371 km, meaning the total land area according to the CRU land mask is 1.46×10^8 km^2 (or 1.44×10^8 km^2 when Greenland is excluded).

The models show a significant spread of the partitioning of precipitation into snowfall and rainfall and the further partitioning of precipitation into evapotranspiration and runoff (throughout this paper, runoff refers to the combined surface and subsurface runoff) (see Figs. 1, 2). Simulated global evapotranspiration over land ranges from 415 to 586 mm yr^{-1} (from 60 000 to 85 000 km^3 yr^{-1}) and simulated runoff ranges from 290 to 457 mm yr^{-1} (from 42 000 to 66 000 km^3 yr^{-1}), with the global mean model simulated runoff fraction ranging from 0.33 to 0.52. The runoff fractions are calculated as runoff divided by precipitation. Both the mean and the median runoff fractions for the LSMs are lower than the corresponding GHM values, although the LSMs show a larger spread in the predicted runoff fraction than the GHMs (Fig. 2b). From a water availability point of view, this means that the model predicting the most runoff on average has about 57% or nearly 25 000 km^3 yr^{-1} more surface water available globally than the model simulating the least runoff, which can dramatically influence subsequent studies of water stress.

Biemans et al. (2009) compared seven different global terrestrial precipitation datasets and reported a global terrestrial mean precipitation between 743 and 926 mm yr^{-1}. Compared to previous estimates, precipitation values are at the upper end in this study, which is mainly due to undercatch correction factors (Weedon et al. 2010, 2011).

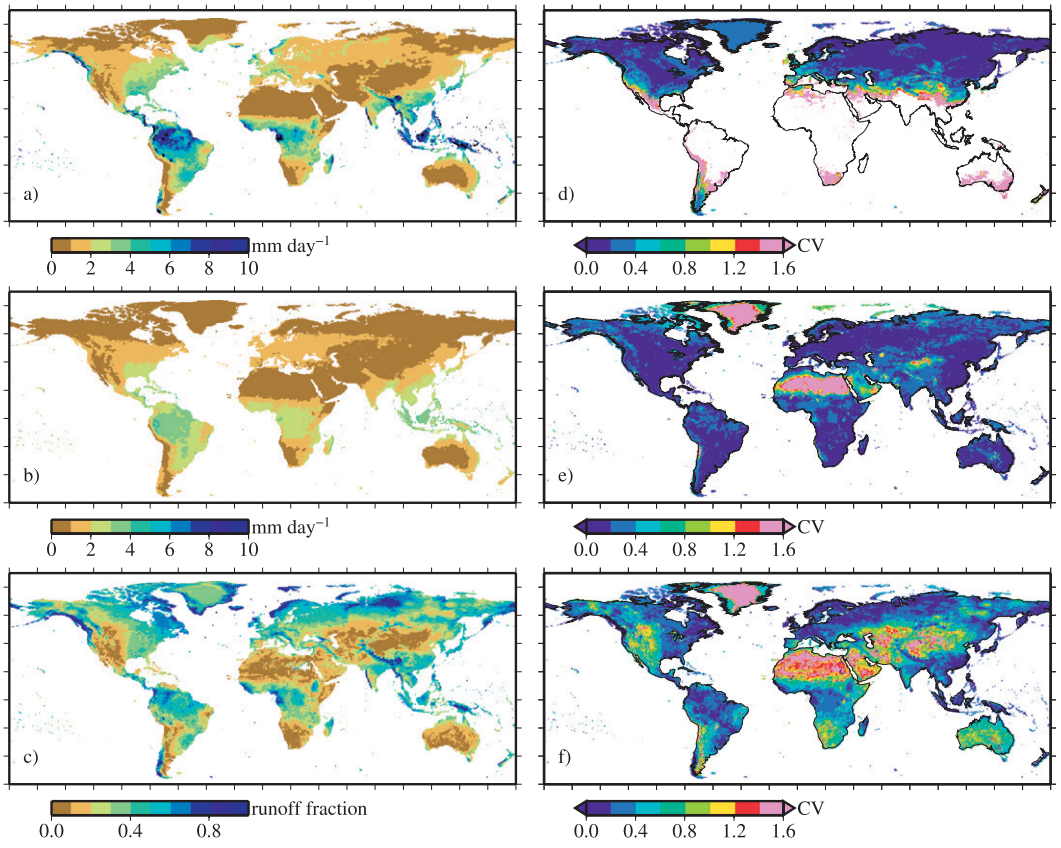

FIG. 1. (a) Mean annual precipitation; multimodel mean annual (b) ET and (c) runoff fraction; and CV of the model means of (d) snowfall, (e) ET, and (f) runoff.

Based on streamflow data from the world's largest rivers, combined with estimates for the ungauged areas, Dai and Trenberth (2002) estimated continental runoff of about 37 000 km^3 yr^{-1}, which is similar to the Fekete et al. (2000) estimate of about 38 000 km^3 yr^{-1}. The runoff volumes found here, 42 000–66 000 km^3 yr^{-1}, are therefore higher. However, Fekete et al. (2000) reports using a land mask that covers 1.33×10^8 km^2 of the world, which is only 92% of the area reported herein (Greenland excluded). Hence, the runoff volume differences can partly be attributed to the land mask used. Dai and Trenberth (2002) do not report the area of the land mask used in their study or globally averaged runoff numbers in millimeters per year, but Fekete et al. (2000) report globally averaged continental runoff of 299 mm yr^{-1}. Given that the land mask of Fekete et al. (2000) and the land mask used in this study do not overlap, the runoff numbers should not be compared directly, but most of the models included in this study simulate higher global terrestrial runoff (290–457 mm yr^{-1}) than the 299 mm yr^{-1} reported in Fekete et al. (2000). Section 3b discusses reasons why most models participating in this study possibly overestimate global terrestrial runoff.

The undercatch correction of precipitation and the aerosol correction of shortwave radiation will in some areas lead to higher runoff values than if these corrections were not implemented. For example, analyses performed for the Amazon and Congo River basin with the Orchidee model indicate on the order of 10% more runoff when using the aerosol-corrected shortwave radiation (Weedon et al. 2010, 2011) than when using shortwave radiation that is not corrected for aerosols (J. Polcher 2010, personal communication).

The long-term intermodel range in predicted water balance terms is larger than the interannual model mean range (Fig. 2a). Also, in the 15-yr simulation period, the interannual variation in multimodel mean predicted global runoff is much larger; both in absolute and relative terms, than the interannual variation in multimodel mean predicted global evapotranspiration. This indicates that, globally averaged, the majority of the interannual variation in precipitation feeds directly through to the runoff and that the evaporation is constrained by other atmospheric factors such as temperature, radiation, and humidity. No major difference in the interannual variations have been found (not shown) between

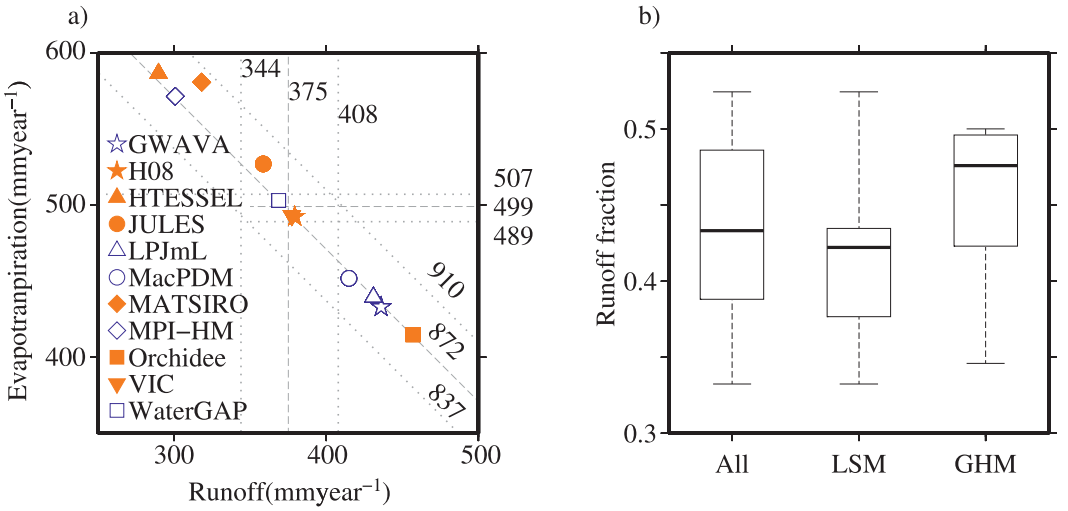

FIG. 2. (a) Global terrestrial mean model predicted runoff vs ET values (mm yr^{-1}; excluding Antarctica and Greenland). The diagonal, vertical, and horizontal lines show long-term multimodel mean annual values and the interannual range of multimodel mean precipitation, runoff, and ET, respectively. LSMs are represented by solid orange symbols, and GHMs are represented by open blue symbols. (b) Box plots illustrating the smallest simulated runoff fractions, lower quartiles, medians, upper quartiles, and the largest simulated runoff fractions for all participating models, the LSMs, and the GHMs.

the models run at daily or subdaily time steps or between models using different evapotranspiration or runoff schemes.

Snow accumulation and ablation influence the shape of the hydrograph significantly in many parts of the world, and so the representation of these processes is an important factor in water availability studies. The amount of snowfall is fairly consistent among the models in the northernmost and coldest areas of the world (Fig. 1d). However, in areas where winter temperatures are closer to 0°C, the models show a large spread in how precipitation is partitioned into rainfall and snowfall. In this study, all models run at subdaily time steps use the provided rainfall and snowfall values directly (see Table 1). The models run at daily time steps partition total precipitation into rainfall and snowfall, using a threshold temperature (typically 0° or 1°C), or into a combination of snow and rain between an upper and lower threshold temperature. Consequently areas experiencing temperatures around these threshold values show larger variations in snowfall amounts. In addition, subgrid elevation schemes influence subgrid air temperatures and hence grid mean snowfall amounts in GWAVA and WaterGAP, which partly explains why the coefficient of variation of snowfall is fairly high in parts of the Rocky Mountains, the Andes, and the Himalayas.

HTESSEL, H08, JULES, MATSIRO, Orchidee, and VIC all use snow schemes based on a physically based energy balance approach, whereas the other models use schemes based on the conceptual degree-day approach. In this study, it appears that the degree-day approach in most places results in higher SWE values than the energy balance approach both in the winter season [December–February (DJF)] and in spring [March–May (MAM)] (Fig. 3). It is important to note that there are also differences between the snow energy balance approaches: for example, number of snow layers, snow albedo values, and how much liquid water can be retained within the snowpack. In the Himalayan region, snow accumulates over the years in several models, which contributes to the model differences illustrated in Fig. 3. The model simulating the lowest SWE numbers, H08, uses a relatively simple one-layer snow scheme and fairly low snow albedo values. In H08, snow albedo varies between 0.6 and 0.45, whereas several other models use snow albedo values up to 0.8. This leads to increased net radiation at the snow surface compared to many other models. The conclusions are not dependent on the H08 results alone, though, and the pattern of Fig. 3 does not change much when excluding the H08 results (not shown). For degree-day snow schemes, there are similar differences: for example, threshold temperature and degree-day factor used and whether melted snow percolates through the snow and directly into the soil or can be retained in the snowpack. The highest global mean SWE values (Himalayas and Greenland excluded) are simulated by the MPI-HM model, which uses threshold temperatures for rain/snow of −1.1° and 3.3°C, respectively; uses a degree-day factor of 3.22 mm $°C^{-1}$ day^{-1}; and assumes that 6% of the SWE can be retained

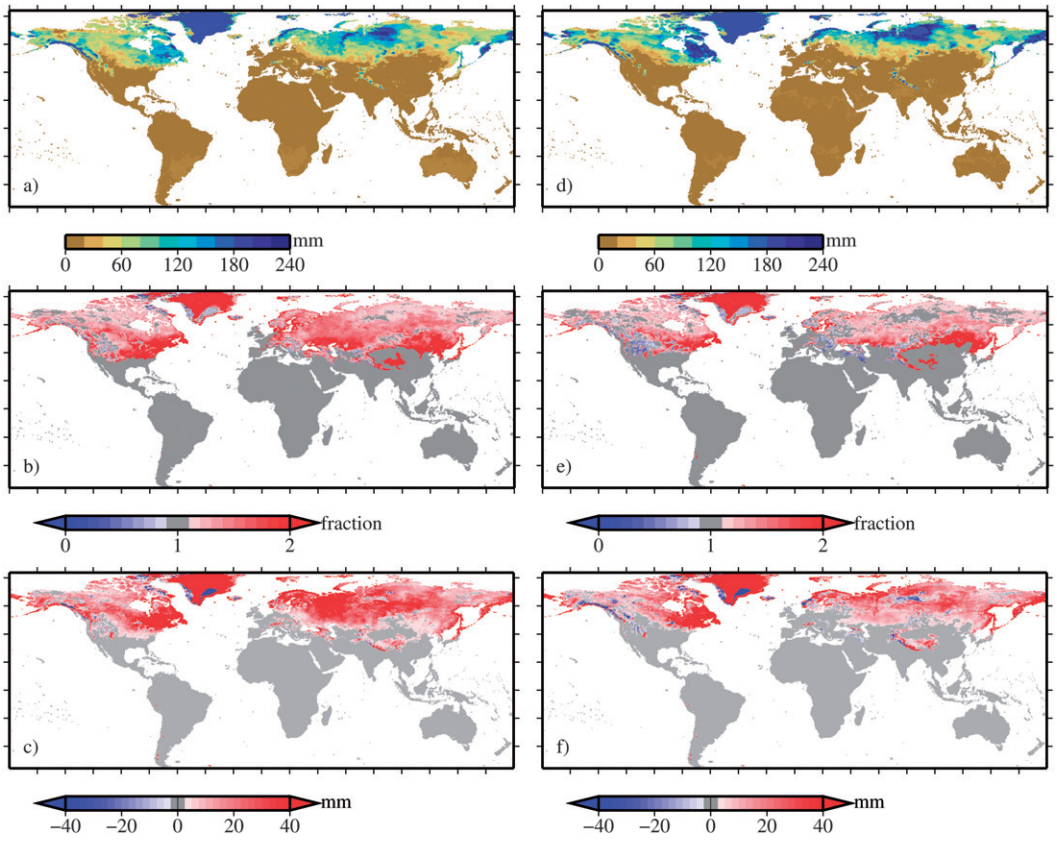

FIG. 3. Comparison of SWE values simulated by degree-day and energy balance models. (a) Mean winter (DJF) SWE, all models; (b) degree-day results divided by energy balance results (DJF); and (c) degree-day results minus energy balance results (DJF). (d) Mean spring (MAM) SWE, all models; (e) degree-day results divided by energy balance results (MAM); and (f) degree-day results minus energy balance results (MAM).

as liquid water in the snowpack. A simple degree-day equation was used to study the sensitivity to the range of threshold temperatures and degree-day factors used among the participating models; and the conclusion is that the threshold temperature influences SWE amounts more than the degree-day factor does. For midlatitude basins, averaged maximum monthly SWE can be 50% higher using a fixed rainfall–snowfall threshold temperature of 1°C compared to 0°C.

The relative differences between models can be expressed by the CV. In most areas, the CV is much higher for simulated runoff than for simulated evapotranspiration (Fig. 1) because runoff values are generally smaller. In arid and semiarid areas, the spread of simulated runoff and evapotranspiration is relatively large, and the CV is high for both evapotranspiration and runoff. Also noticeable is the high CV around the Laurentian Great Lakes in North America, which is a result of the models handling the presence of lakes very differently. The parameterizations of evapotranspiration and runoff vary substantially between the models (see Table 1), and the complicated interactions between the various processes make it infeasible to explain the causes of many simulation differences in detail, as noted in previous model intercomparisons (e.g., Koster and Milly 1997).

b. *Basin analyses*

Some general conclusions can be made based on results from river basins representing contrasting climate characteristics (see locations in Fig. 4). The interannual variations in the main water flux terms (evapotranspiration and runoff) are fairly similar among the models both globally and in the river basins studied, and hence only mean annual and mean monthly results are presented here.

Predicted potential evapotranspiration (PET) values for five large river basins, representing wet and arid or semiarid basins, for participating GHMs using the Penman–Monteith equation (GWAVA and MacPDM) and using the Priestley–Taylor equation (LPJmL and

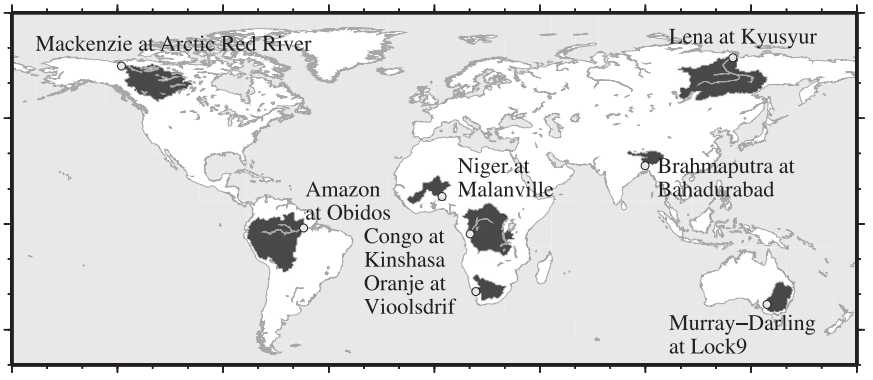

FIG. 4. Location of river basins and discharge gauges.

WaterGAP) are compared in Fig. 5. Previous studies have noted the differences resulting from the Penman–Monteith and Priestley–Taylor equations in wet and dry climates (Weiß and Menzel 2008; Kingston et al. 2009). In general, those studies conclude that simulated PET using the Priestley–Taylor equation tends to be higher than when using the Penman–Monteith equation in humid regions, whereas the opposite is true in dry areas. This is also reported by Weedon et al. (2011), who calculated PET for reference crops globally using the WATCH forcing data (i.e., the same meteorological forcing data that are used here). Not all models compute PET, and fewer models are included in Fig. 5 than what would be expected based on Table 1. Although Fig. 5 does not show that PET in humid climates using the Priestley–Taylor equation is higher than when using the Penman–Monteith equation, it does indicate that the spread in simulated PET is lower in wet than in dry basins. The models represented in Fig. 5 describe the vegetation in the basins somewhat differently, and there are also differences in approach: for example, PET calculated by WaterGAP is dependent on land cover albedo, whereas MacPDM in addition takes both LAI and stomatal resistance into account. LPJmL accounts for stomatal conductance and also for dynamical vegetation changes and it computes PET using a modified Priestley–Taylor formulation that accounts for boundary layer dynamics. Hence, the PET and resulting ET differences cannot be attributed solely to the choice of equation. These differences in the details of the implementations most likely explain why the results presented here are somewhat different from those of Weedon et al. (2011).

The simulated mean annual water balance and runoff fraction statistics for eight large river basins are presented in Fig. 6, which also includes information on mean annual observed discharge in the basins and the range in observed annual discharge. Discharge values are obtained from the Global Runoff Data Centre [data are available from the GRDC in the Bundesanstalt für Gewaesserkunde, 56068 Koblenz, Germany (see http://grdc.bafg.de)], and converted to millimeters per year using the area upstream of the gauge according to the DDM30 river network. This means that an area correction factor is applied to the GRDC discharge data to account for the fact that the river network, which is at 0.5° spatial resolution, may not perfectly overlap with the river basin boundaries. The resulting statistics are fairly sensitive to individual model results and the grouping chosen, and hence individual, grouped, and mean model results are presented in Fig. 6. Figure 7 shows simulated multimodel mean monthly runoff values, the range of model means and interannual multimodel mean results. Not all participating models have a routing scheme included, and hence Fig. 7 shows mean basin runoff values and not discharge at the basin outlets. Some models do not sit on the mean basin (Fig. 6)

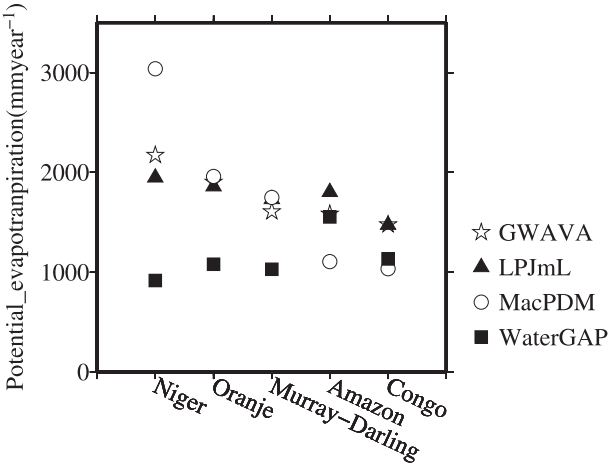

FIG. 5. Simulated mean annual PET in the Niger, Oranje, Murray–Darling, Amazon, and Congo River basins for a subset of GHMs. For calculating PET, GWAVA, and MacPDM, use the Penman–Monteith equation (open symbols); for calculating LPJmL and WaterGAP, use the Priestley–Taylor equation (solid symbols).

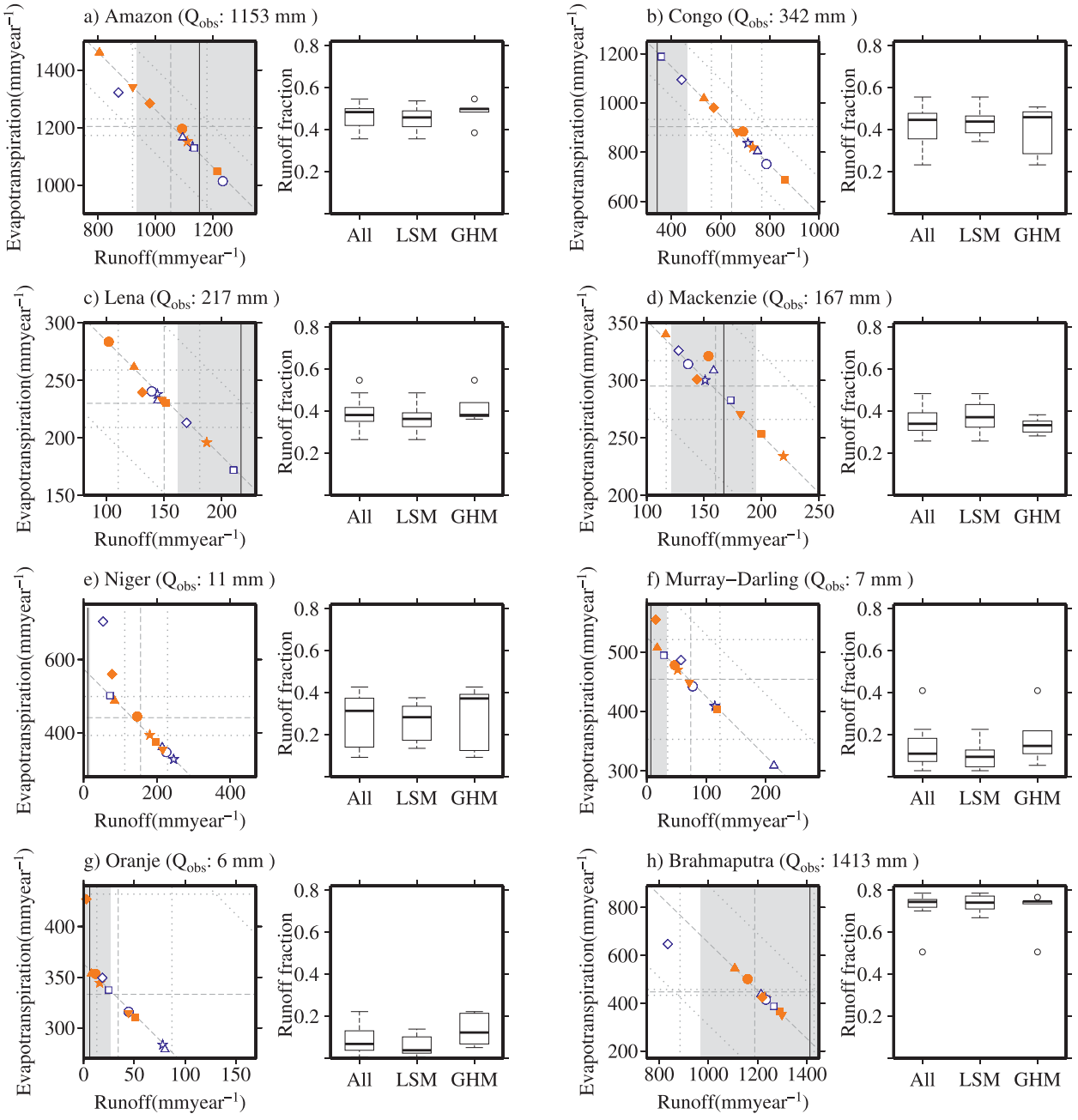

FIG. 6. River basin mean model predicted runoff and ET values (mm yr^{-1}). LSMs are represented with solid orange symbols, GHMs are represented with open blue symbols, and the same symbols as in Fig. 2 are used. The dashed gray lines show long-term multimodel mean annual runoff, ET, and precipitation, and the dotted lines show the range in multimodel mean annual runoff. Observed mean annual runoff (vertical black line; mm yr^{-1}) for the 15-yr simulation period is included for all basins except the Congo and Murray–Darling River basins, where the long-term average is used because there were no or insufficient data available for the period in question. The shaded area indicates the range in observed runoff for the period in question. For the Amazon, Congo, Lena, and Brahmaputra basins, the maximum or minimum observed annual runoff falls outside the runoff range included on the *x* axis. The box plots represent runoff fractions for all models combined and for the LSMs and GHMs separately, and they illustrate the smallest simulated runoff fractions, lower quartiles, medians, upper quartiles, and the largest simulated runoff fractions. Outliers are represented by circles. All terms are calculated for the basin area upstream of the discharge gauge.

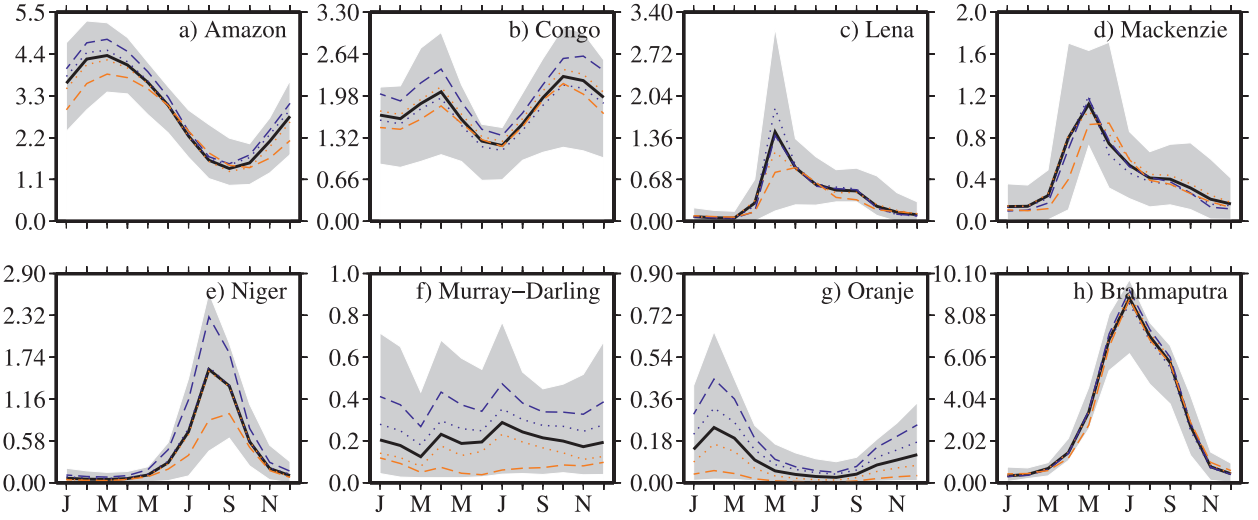

FIG. 7. Multimodel mean monthly runoff values, in millimeters per day. The multimodel mean runoff values are represented by a solid black line, and the shaded area represents the range of the model mean runoff values. Blue dotted lines represent the mean of the GHMs, and orange dotted lines represent the mean of the LSMs. Blue and orange dashed lines represent the means of the GHMs GWAVA, LPJmL, and MacPDM, and the LSMs HTESSEL, JULES, and MATSIRO, respectively. The runoff values are calculated for the area upstream of the discharge gauges.

or global (Fig. 2) precipitation line. This is caused by changes in the water stores between the start and end of the run and, for the JULES model, by nonconservation of water for lake surfaces.

Figures 6 and 7 both show that the simulated runoff varies substantially between the models, and to some extent so does the timing of runoff over the year. The patterns seen in Figs. 6 and 7—namely large absolute differences in runoff in the tropics, with relative differences being larger in the drier basins—reflect the mean annual values and CVs presented globally in Fig. 1. Because all simulations were for naturalized conditions, meaning that dams and water withdrawals that change the dynamics of the water cycle are not taken into account, it is not appropriate to compare the models with observed discharge at subannual time scales in all basins. However, at the annual time scale, Fig. 6 shows that most models clearly overestimate runoff in the semiarid and arid basins, such as the Niger, Murray–Darling, and Oranje River basins. This likely can be explained in part by water extractions in these areas, which will be explored further in future WaterMIP analyses, but it is also likely that the models miss out two key processes. The first is transmission loss along the river channel, which is very significant along major rivers in arid zones and means that it is arguably inappropriate to compare observed streamflow and simulated runoff. The second is the reinfiltration and subsequent evaporation of surface runoff generated in part of the catchment. Overprediction of runoff in the Congo and Niger River basins is possibly linked to the complicated wetland dynamics in these basins (see also discussion in Taylor 2010). In the Brahmaputra River basin, the neglect of water use in the model simulations might be expected to lead to overestimation of runoff, but all models underpredict runoff in this basin. The results of the other basins are fairly mixed: for example, there is no consistent overprediction or underprediction in the Arctic river basins (see results for the Mackenzie and Lena River basins in Fig. 6).

The differences between the models in each of the classes (LSMs or GHMs) are larger than the interclass differences for all of the basins presented in Fig. 6. However, there are some subgroups that show more consistent behavior. The global average runoff fractions are lower for the three LSMs HTESSEL, JULES, and MATSIRO than for most other models (Fig. 2), and this behavior is also found when looking at most of the individual basins presented in Fig. 6, particularly the Oranje and Murray–Darling basins, where the LSMs on average predict runoff values closer to the observed values than do the GHMs. The global hydrological models GWAVA, LPJmL, and MacPDM agree well on the runoff fraction in most basins but have relatively high runoff fractions compared to most other models. The results for these two subgroups (i.e., three LSMs and three GHMs) are also presented in Fig. 7, which shows clear differences between these subgroups in terms of the runoff from some basins and that the relative difference is especially high in the arid and semiarid basins (Niger, Murray–Darling, and Oranje). In these

arid and semiarid basins, the runoff ratio is low, and small differences in evaporation result in large relative differences in runoff. The runoff differences are probably not due to differences in radiation: for example, in the Niger, Murray–Darling, and Oranje River basins, net radiation in LPJmL is higher than net radiation in the LSMs, although resulting evapotranspiration is lower from LPJml. Actual evapotranspiration in LPJmL is constrained by a physiological maximum (Rost et al. 2008), which, together with the use of the Priestley–Taylor method (see above), may partly explain the rather high runoff in dry regions. Also, it may still be that the differences are caused by the temporal resolution and energy balance implemented in the LSMs, compared to daily time steps and no closure of the energy balance in the GHMs. For the Lena basin, Fig. 7 indicates that the lower SWE values predicted in this basin by the LSMs (Fig. 3) result in lower spring runoff volumes than are predicted by the GHMs (see also below).

Globally, Orchidee predicts the highest runoff fraction (Fig. 2), and in most basins Orchidee predicts runoff fractions that are among the highest (Fig. 6). In basins dominated by snow accumulation and melt (here represented by the Lena and Mackenzie River basins), the H08 model tends to have relatively higher runoff fractions, compared to the other models, than elsewhere. Hence, the models characterized as LSMs are represented at both the dry and wet ends of the range of simulated runoff fraction in many basins. It may also be noted that H08 and VIC in many basins are closer to the global hydrological models GWAVA, LPJmL, and MacPDM than to the other LSMs. MPI-HM stands out slightly from the other GHMs by having higher evaporation both globally (Fig. 2) and in some basins (Fig. 6). This is linked to the evaporation scheme used, which will be further discussed below. WaterGAP appears at both the low and high end of the runoff fraction ranges. WaterGAP is the only model that is calibrated, which also explains why the WaterGAP simulated basin runoff values are closer to the observations than all other model results. With some exceptions, the findings in Fig. 6 agree with the GSWP-2 results (Dirmeyer et al. 1999) that most of the models behaved consistently between the basins, and the same models routinely appeared at either end of the runoff distribution.

In the Lena River basin, winter temperatures are well below freezing and hence all models agree fairly well on snowfall amounts (Fig. 8a). However, even in this basin there are differences in snowfall amounts between the models, especially in October, when models that directly use the provided snowfall and rainfall data (HTESSEL, H08, JULES, and MATSIRO) have about 8 mm less snowfall than models partitioning daily precipitation into rainfall and snowfall based on daily mean air temperature. In particular, H08 predicts relatively high runoff values in the fall, which might be attributed to rainfall percolating through the snowpack (no water is retained in the snowpack in H08) into the soil and producing runoff. Also, as mentioned in section 3a, snow albedo values in H08 are fairly low, which influences net radiation to the snowpack. Given the low winter temperatures in the Lena River basin and hence little snowmelt, one might expect that modeled SWE would be fairly similar throughout the winter. However, the difference in simulated SWE in the peak month (March) is actually about 50 mm (Fig. 8b) and the lowest simulated SWE is approximately 50% of the maximum SWE in March. Snow throughfall (i.e., melted snow that leaves the snowpack) is nearly zero in the Lena River basin until April (not shown) and hence cannot explain the differences in maximum SWE. However, the differences in SWE can partly be explained by looking at snow sublimation and evaporation (Fig. 8c), which have also been found to influence runoff in previous model intercomparison projects (Bowling et al. 2003). In some of the models (H08, HTESSEL, and JULES) about 30 mm of water is lost to snow sublimation and evaporation in the snow accumulation season, and these models correspond to the models simulating the lowest SWE values in March.

The inclusion of soil frost is very likely to influence runoff in the Lena basin and other Arctic basins, and this hypothesis was tested when analyzing runoff volume and timing in the basin. However, it was not possible to reach a definite conclusion that the inclusion of soil frost results in a higher runoff peak in the spring, because runoff is influenced by so many factors. However, for any one model, snowmelt and spring season runoff will be higher with soil frost than if frost is not included in the model.

In the Amazon River basin, all models agree closely on the shape of the annual runoff distribution, although simulated runoff amounts are significantly different (Figs. 6a, 8f). In most basins, MATSIRO predicts somewhat less seasonal variation in runoff than the other models, and this is true in the Amazon River basin. MATSIRO has a deep groundwater reservoir, and this clearly influences the timing of runoff. HTESSEL, MATSIRO, and VIC simulate the highest canopy evaporation in the Amazon River (Fig. 8d) and also the lowest vegetation transpiration. It has previously been pointed out that canopy evaporation amounts can affect the seasonal cycle of soil moisture. Demory and Vidale (2009) showed that reduced canopy interception capacity—and hence reduced canopy evaporation—leads to higher soil moisture variations in JULES. However, although this relationship between canopy evaporation and soil moisture amplitudes is likely

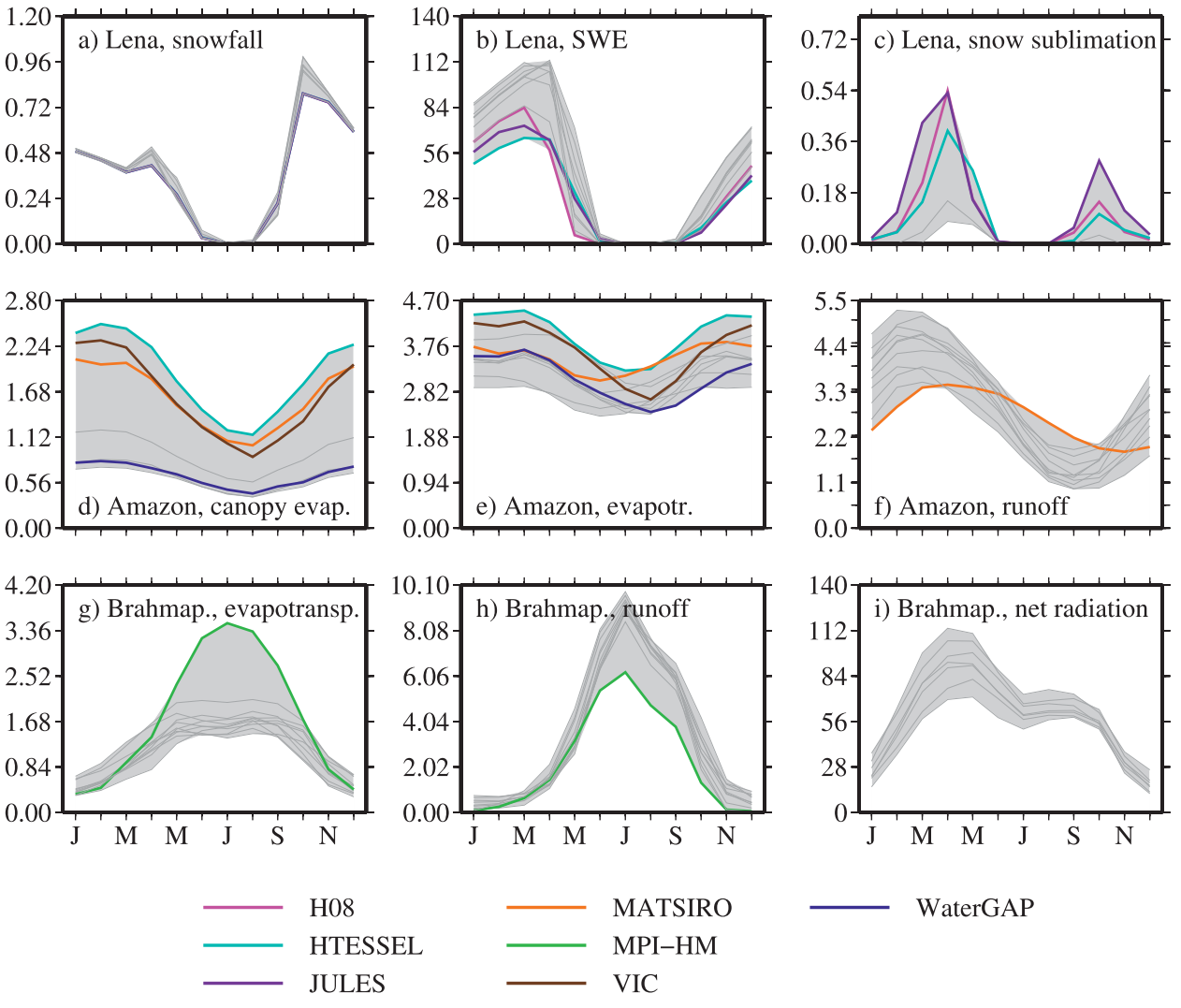

FIG. 8. Mean monthly river basin results. To highlight main simulation differences, only selected model results are shown in each panel. Model results mentioned in the text are represented by colored lines; the others are represented by gray lines. Shaded area shows the range of the model results. (a) Snowfall, (b) SWE, and (c) snow sublimation and evaporation values in the Lena River basin. (d) Canopy evaporation, (e) ET, and (f) runoff in the Amazon River basin. (g) ET, (h) runoff, and (i) net radiation in the Brahmaputra River basin. Water fluxes are in mm day^{-1}, storage terms are in mm, and radiation values are in W m^{-2}.

to hold for individual models, the WaterMIP results for the Amazon show that it is not universally applicable. Canopy interception capacity varies substantially between the models: for example,, VIC and WaterGAP have canopy interception capacities of 0.1 × leaf area index (LAI) and 0.3 × LAI (mm), respectively. Despite having lower canopy interception capacity, canopy evaporation in VIC is much higher than in WaterGAP in the Amazon River basin (Fig. 8d), although LAI values are broadly similar, and indeed is higher in all river basins studied. This is at least partly attributed to WaterGAP using less vegetation-specific information when calculating evapotranspiration than VIC does: for example, WaterGAP does not take vegetation height and its influence on aerodynamic resistance into account.

In the Brahmaputra River basin, the effects on evapotranspiration of reduced incoming solar radiation and high humidity during the Indian monsoon is clearly visible in the results of all models other than MPI-HM (Fig. 8g). MPI-HM is the only model using the Thornthwaite evapotranspiration equation, meaning potential evapotranspiration is calculated based on air temperature only. In periods when shortwave radiation or humidity limits evapotranspiration in models using, for example, the Penman–Monteith or Priestley–Taylor equation, the MPI-HM estimated evapotranspiration can be substantially higher

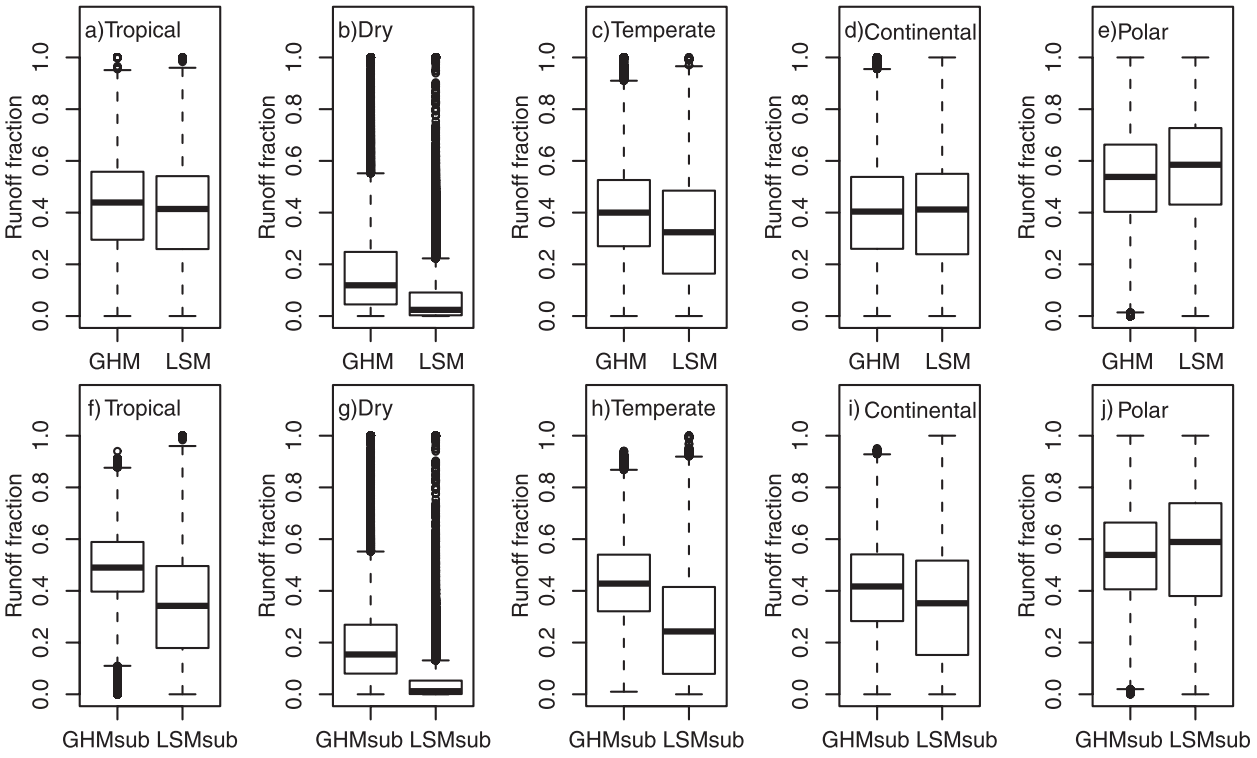

FIG. 9. Box plots illustrating the smallest simulated runoff fractions, lower quartiles, medians, upper quartiles, and the largest simulated runoff fractions for all cells within the five main Köppen climate zones. (a)–(e) All participating models are included: that is, the same grouping as in Table 1 is used. (f)–(j) The same subgroups as in Fig. 7 are used: that is, GWAVA, LPJmL, and MacPDM are included in the GHM subgroup (GHMsub) and HTESSEL, JULES, and MATSIRO are included in the LSM subgroup (LSMsub).

than that estimated by other models. This is especially noticeable in the Himalayan region during the Indian monsoon and is also apparent in the results for the Chang Jiang basin (not shown).

In sum, model differences result in significant differences in predicted runoff values at both annual and monthly time scales (Figs. 6, 7). In many basins the intermodel runoff range is larger than the interannual mean model range, and during parts of the year the model range is substantial, especially in low-flow periods. Although the differences per unit area can appear small (e.g., mm day^{-1} for fluxes and mm for stores in Figs. 6, 8), these are large volumes of water when aggregated over the basins. In the Lena River basin, for example, 1 mm of SWE amounts to 2.4 km^3 basin total and if melted on one day equals nearly 30 000 m^3 s^{-1} at the basin outlet. Such a runoff difference of 1 mm day^{-1} is well within the model range for many of the basins presented in Fig. 7, and in some basins the differences are much larger during parts of the year.

c. *Köppen climate zone analyses*

Some general results and distinct differences between models are presented in the global and basin analyses in sections 3a and 3b. The basin results presented in Figs. 6 and 7 indicate that the relative runoff differences between LSMs and GHMs are largest in arid or semiarid basins (e.g., Murray–Darling and Oranje Rivers), whereas the differences between the model classes are less prominent in other basins and there was no clear signal for the Arctic basins studied (Mackenzie and Lena). Figure 9 shows the results of a more comprehensive analysis in which runoff fractions were analyzed separately across Köppen climate zones rather than a few basins. The model groupings used in Fig. 7 were also applied in these analyses: that is, one grouping in which all the models were included (Figs. 9a–e) and one subgrouping that included three GHMs and three LSMs (Figs. 9f–j). When all model results are included in the analyses (Figs. 9a–e), there are few differences in the runoff fraction statistics in climate zones other than the dry areas (Fig. 9b). However, for the subgrouping of models, the differences are generally larger (Figs. 9f–j), as was found for the individual basins presented in Fig. 7. In the tropical and temperate Köppen climate zones (Figs. 9a,c), the differences become particularly noticeable, which for the tropical zone is consistent with the results for the Amazon and Congo Rivers in Figs. 7a,c. At least for

these subgroups of models, there are systematic differences between the groups across broad climate zones.

4. Conclusions

Results from 11 land surface and global hydrological models demonstrate a large range in global and regional water flux and storage terms. Globally, the simulated range in runoff values is nearly 25 000 $km^3\ yr^{-1}$ (or 45% of the mean simulated runoff), with the results of the LSMs appearing at both the wet and dry ends of the range. However, both the mean and median LSM runoff fractions are lower than the corresponding GHM values. In the 15-yr simulation period, the interannual variation in multimodel mean predicted global runoff is much larger; both in absolute and relative terms, than the interannual variation in multimodel mean predicted global evapotranspiration over land. As regards the interannual variation in runoff and evapotranspiration, no major differences have been found between the models run at daily or subdaily time steps or between models using different evapotranspiration or runoff schemes.

The largest absolute runoff differences are found in the tropics, whereas the largest relative differences are found in arid areas. The models generally overpredict runoff for the arid and semiarid basins, but some of the energy balance models are closer to the observations for these basins. Models using a physically based energy balance approach in general predict lower snow water equivalent than models using a conceptual degree-day approach, which at least partly can be explained by snow sublimation, which is accounted for only in the energy balance models. For evapotranspiration and runoff no major differences have been found between the LSMs and GHMs. Some of the differences in model predicted water fluxes and storage terms can be attributed to specific model parameterizations, although the complexity of the models makes it infeasible to explain all differences. The results indicate that differences in simulated PET tend to be smaller in wet climates than in dry climates. Results also show that, in some areas, calculating evapotranspiration based only on temperature can lead to significantly different results than if radiation and humidity are also considered.

The impact of climate change on the global terrestrial water cycle and water resources is an important research question relevant to many policy areas. Many of the models participating in WaterMIP are being used for climate change impact studies. This model intercomparison shows that there are considerable differences in simulated evaporation and runoff, which can have a large impact on the available water resources in some regions. Studies of the climate change predicted by climate models show considerable differences between models, particularly for precipitation, and there is now a growing consensus that climate change impact studies should consider results from a range of climate models (Covey et al. 2003; Meehl et al. 2009). Our results show that differences between hydrological model results are also a major source of uncertainty. When studying the impacts of climate change on the global terrestrial water cycle and water resources, definite conclusions should not be based on the results of a single model realization. Climate change impact studies have for some time used multiple climate models and should preferably also start using multiple impact models. Alternatively, other approaches to assess hydrological model uncertainty must be considered (see, e.g., Lawrence and Haddeland 2011).

The next step in WaterMIP will be multimodel analyses of simulated historical water use and water stress, for which the models will include representations of dams and water used for agriculture. Thereafter, hydrologic simulations using future climate projections (Hagemann et al. 2011) will be performed, with and without taking anthropogenic impacts into account. More information about WaterMIP and related modeling activities within WATCH, including information on the protocol and possibilities of obtaining forcing data and modeling results, can be found online (at http://www.eu-watch.org/watermip).

Acknowledgments. This research was undertaken as part of the European Union (FP6) funded Integrated Project called WATCH (Contract 036946), in collaboration with the Global Water System Project (GWSP). Martin Best and Graham Weedon were supported by the Joint DECC and Defra Integrated Climate Programme, DECC/Defra (GA01101). Thanks to Niko Wanders at the Wageningen University and Research Centre for providing the Köppen zones calculated based on the WATCH forcing data. Thanks also to three anonymous reviewers whose comments helped us improve the paper.

REFERENCES

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

Alcamo, J., P. Döll, T. Henrichs, F. Kaspar, B. Lehner, T. Rösch, and S. Siebert, 2003: Development and testing of the WaterGAP 2 global model of water use and availability. *Hydrol. Sci. J.,* **48,** 317–333.

Arnell, N. W., 1999: A simple water balance model for the simulation of streamflow over a large geographic domain. *J. Hydrol.,* **217,** 314–335.

——, 2004: Climate change and global water resources: SRES emissions and socio-economic scenarios. *Global Environ. Change,* **14,** 31–52.

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

Biemans, H., R. W. A. Hutjes, P. Kabat, B. J. Strengers, D. Gerten, and S. Rost, 2009: Effects of precipitation uncertainty on discharge calculations for main river basins. *J. Hydrometeor.,* **10,** 1011–1025.

Bondeau, A., and Coauthors, 2007: Modelling the role of agriculture for the 20th century global terrestrial carbon balance. *Global Change Biol.,* **13,** 679–706.

Bowling, L. C., and Coauthors, 2003: Simulation of high-latitude hydrological processes in the Torne–Kalix basin: PILPS phase 2(e): 1. Experiment description and summary intercomparisons. *Global Planet. Change,* **38,** 1–30.

Covey, C., K. M. AchutaRao, U. Cubasch, P. Jones, S. J. Lambert, M. E. Mann, T. J. Phillips, and K. E. Taylor, 2003: An overview of results from the Coupled Model Intercomparison Project. *Global Planet. Change,* **37,** 103–133.

Cox, P. M., R. A. Betts, C. B. Bunton, R. L. H. Essery, P. R. Rowntree, and J. Smith, 1999: The impact of new land surface physics on the GCM simulation of climate and climate sensitivity. *Climate Dyn.,* **15,** 183–203.

Dai, A., and K. E. Trenberth, 2002: Estimates of freshwater discharge from continents: Latitudinal and seasonal variations. *J. Hydrometeor.,* **3,** 660–687.

Demory, M. E., and P. L. Vidale, 2009: Does overestimated canopy interception weaken the UK land surface model response to precipitation events? *iLEAPS Newsletter*, No. 7, iLEAPS International Project Office, Helsinki, Finland, 14–16.

De Rosnay, P., and J. Polcher, 1998: Modeling root water uptake in a complex land surface scheme coupled to a GCM. *Hydrol. Earth Syst. Sci.,* **2,** 239–256.

Dirmeyer, P. A., A. J. Dolman, and N. Sato, 1999: The Global Soil Wetness Project: A pilot project for global land surface modeling and validation. *Bull. Amer. Meteor. Soc.,* **80,** 851–878.

——, X. Gao, M. Zhao, Z. Guo, T. Oki, N. Hanasaki, 2006: GSWP-2: Multimodel analysis and implications for our perception of the land surface. *Bull. Amer. Meteor. Soc.,* **87,** 1381–1397.

Döll, P., and B. Lehner, 2002: Validation of a new global 30-minute drainage direction map. *J. Hydrol.,* **258,** 214–231.

——, and S. Siebert, 2002: Global modeling of irrigation water requirements. *Water Resour. Res.,* **38,** 1037, doi:10.1029/2001WR000355.

Essery, R. L. H., M. J. Best, R. A. Betts, P. M. Cox, and C. M. Taylor, 2003: Explicit representation of subgrid heterogeneity in a GCM land surface scheme. *J. Hydrometeor.,* **4,** 530–543.

Fekete, B. M., C. J. Vörösmarty, and W. Grabs, 2000: Global composite runoff fields based on observed river discharge and simulated water balances. Global Runoff Data Centre Rep. 22, 120 pp. [Available online at http://www.grdc.sr.unh.edu/html/paper/ReportUS.pdf.]

Fuchs, T., 2009: GPCC annual report for year 2008: Development of the GPCC data base and analysis products. DWD Rep., 13 pp. [Available online at http://www.dwd.de/bvbw/generator/DWDWWW/Content/Oeffentlichkeit/KU/KU4/KU42/en/Reports__Publications/GPCC__annual__report__2008,templateId=raw,property=publicationFile.pdf/GPCC_annual_report_2008.pdf.]

Gosling, S. N., and N. W. Arnell, 2010: Simulating current global river runoff with a global hydrological model: Model revisions, validation, and sensitivity analysis. *Hydrol. Processes,* **25,** 1129–1145, doi:10.1002/hyp.7727.

Hagemann, S., and L. Dümenil, 1998: A parameterization of the lateral waterflow for the global scale. *Climate Dyn.,* **14,** 17–31.

——, and L. D. Gates, 2003: Improving a subgrid runoff parameterization scheme for climate models by the use of high resolution data derived from satellite observations. *Climate Dyn.,* **21,** 349–359.

——, C. Chen, J. O. Haerter, D. Gerten, J. Heinke, and C. Piani, 2011: Impact of a statistical bias correction on the projected hydrological changes obtained from three GCMs and two hydrology models. *J. Hydrometeor.,* **12,** 556–578.

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

——, ——, ——, ——, ——, ——, ——, and ——, 2008b: An integrated model for the assessment of global water resources—Part 2: Applications and assessments. *Hydrol. Earth Syst. Sci.,* **12,** 1027–1037.

Henderson-Sellers, A., A. J. Pitman, P. K. Love, P. Irannejad, and T. H. Chen, 1995: The Project for Intercomparison of Land Surface Parameterization Schemes (PILPS): Phases 2 and 3. *Bull. Amer. Meteor. Soc.,* **76,** 489–503.

Hoff, H., M. Falkenmark, D. Gerten, L. Gordon, L. Karlberg, and J. Rockström, 2010: Greening the global water system. *J. Hydrol.,* **384,** 177–184.

Kingston, D. G., M. C. Todd, R. G. Taylor, J. R. Thompson, and N. W. Arnell, 2009: Uncertainty in the estimation of potential evapotranspiration under climate change. *J. Geophys. Res.,* **36,** L20403, doi:10.1029/2009GL040267.

Koirala, S., 2010: Explicit representation of groundwater process in a global-scale land surface model to improve hydrological predictions, Ph.D thesis, University of Tokyo, 208 pp.

Koster, R. D., and P. C. D. Milly, 1997: The interplay between transpiration and runoff formulations in land surface schemes used with atmospheric models. *J. Climate,* **10,** 1578–1591.

Lawrence, D., and I. Haddeland, 2011: Uncertainty in catchment-scale HBV modelling of climate change impacts on peak flows in Norway. *Hydrol. Res.,* in press.

Liang, X., D. P. Lettennmaier, E. F. Wood, and S. J. Burges, 1994: A simple hydrologically based model of land surface water and energy fluxes for general circulation models. *J. Geophys. Res.,* **99** (D7), 14 415–14 428.

Meehl, G. A., and Coauthors, 2009: Decadal prediction. *Bull. Amer. Meteor. Soc.,* **90,** 1467–1485.

Meigh, J. R., A. A. McKenzie, and K. J. Sene, 1999: A grid-based approach to water scarcity estimates for eastern and southern Africa. *Water Resour. Manage.,* **13,** 85–115.

Mitchell, T. D., and P. D. Jones, 2005: An improved method of constructing a database of monthly climate observations and associated high-resolution grids. *Int. J. Climatol.,* **25,** 693–712.

New, M., M. Hulme, and P. Jones, 1999: Representing twentieth-century space–time climate variability. Part I: Development of

a 1961–90 mean monthly terrestrial climatology. *J. Climate,* **12,** 829–856.

——, ——, and ——, 2000: Representing twentieth-century space–time climate variability. Part II: Development of 1901–96 monthly grids of terrestrial surface climate. *J. Climate,* **13,** 2217–2238.

Pitman, A. J., and A. Henderson-Sellers, 1998: Recent progress and results from the project for the intercomparison of landsurface parameterization schemes. *J. Hydrol.,* **212–213,** 128–135.

Polcher, J., K. Laval, L. Dümenil, J. Lean, and P. R. Rowntree, 1996: Comparing three land surface schemes used in GCMs. *J. Hydrol.,* **180,** 373–394.

——, and Coauthors, 2000: GLASS: Global Land-Atmosphere System Study. *GEWEX News,* Vol. 10, No. 2, International GEWEX Project Office, Silver Spring, MD, 3–5.

Rost, S., D. Gerten, A. Bondeau, W. Lucht, J. Rohwer, and S. Schaphoff, 2008: Agricultural green and blue water consumption and its influence on the global water system. *Water Resour. Res.,* **44,** W09405, doi:10.1029/2007WR006331.

Rudolf, B., and U. Schneider, 2005: Calculation of gridded precipitation data for the global land-surface using in-situ gauge observations. *Proc. Second Workshop of the Int. Precipitation Working Group,* Monterey, CA, GPCC, 231–247. [Available online at http://www.dwd.de/bvbw/generator/DWDWWW/Content/Oeffentlichkeit/KU/KU4/KU42/en/Reports__Publications/Calculation,templateId=raw,property=publicationFile.pdf/Calculation.pdf.]

Schneider, U., T. Fuchs, A. Meyer-Christoffer, and B. Rudolf, 2008: Global precipitation analysis products of the GPCC. GPCC Rep., 12 pp. [Available online at http://www.dwd.de/bvbw/generator/DWDWWW/Content/Oeffentlichkeit/KU/KU4/KU42/en/Reports__Publications/GPCC__intro__products__2008,templateId=raw,property=publicationFile.pdf/GPCC_intro_products_2008.pdf.]

Takata, K., S. Emori, and T. Watanabe, 2003: Development of the minimal advanced treatments of surface interaction and runoff. *Global Planet. Change,* **38,** 209–222.

Taylor, C. M., 2010: Feedbacks on convection from an African wetland. *Geophys. Res. Lett.,* **37,** L05406, doi:10.1029/2009GL041652.

Uppala, S. M., and Coauthors, 2005: The ERA-40 Re-Analysis. *Quart. J. Roy. Meteor. Soc.,* **131,** 2961–3012.

Voß, F., and Coauthors, 2008: First results from intercomparison of surface water availability modules. WATCH Tech. Rep. 1, 19 pp.

Weedon, G. P., S. Gomes, P. Viterbo, H. Österle, J. C. Adam, N. Bellouin, O. Boucher, and M. Best, 2010: The WATCH forcing data 1958-2001: A meteorological forcing dataset for land surface and hydrological models. WATCH Tech. Rep. 22, 41 pp.

——, and Coauthors, 2011: Creation of the WATCH forcing data and its use to assess global and regional reference crop evaporation over land during the twentieth century. *J. Hydrometeor.,* **12,** 823–848.

Weiß, M., and L. Menzel, 2008: A global comparison of four potential evapotranspiration equations and their relevance to stream flow modelling in semi-arid environments. *Adv. Geosci.,* **18,** 15–23.