Ecological Modelling 303 (2015) 87–96

Contents lists available at ScienceDirect

Ecological Modelling

journal homepage: www.elsevier.com/locate/ecolmodel

How well do terrestrial biosphere models simulate coarse-scale runoff in the contiguous United States?

C.R. Schwalm[a,b,*], D.N. Huntzinger[b,c], R.B. Cook[d], Y. Wei[d], I.T. Baker[e], R.P. Neilson[f], B. Poulter[g], Peter Caldwell[h], G. Sun[h], H.Q. Tian[i], N. Zeng[j]

[a] Center for Ecosystem Science and Society, Northern Arizona University, Flagstaff, AZ 86011, USA
[b] School of Earth Sciences and Environmental Sustainability, Northern Arizona University, Flagstaff, AZ 86011, USA
[c] Department of Civil Engineering, Construction Management, and Environmental Engineering, Northern Arizona University, Flagstaff, AZ 86011, USA
[d] Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA
[e] Department of Atmospheric Sciences, Colorado State University, Fort Collins, CO 80523, USA
[f] Pacific Northwest Research Station, USDA Forest Service, Corvallis, OR 97331, USA
[g] Department of Ecology, Montana State University, Bozeman, MT 59717, USA
[h] Southern Research Station, USDA Forest Service, Raleigh, NC 27606, USA
[i] International Center for Climate and Global Change Research School of Forestry and Wildlife Sciences, Auburn University, Auburn, 36849 AL, USA
[j] Department of Atmospheric and Oceanic Science, University of Maryland, College Park, MD 20742, USA

ARTICLE INFO

Article history:
Received 4 July 2014
Received in revised form 23 January 2015
Accepted 9 February 2015
Available online 11 March 2015

Keywords:
Terrestrial biosphere models
Runoff
Intercomparison
North American Carbon Program
Regional

ABSTRACT

Significant changes in the water cycle are expected under current global environmental change. Robust assessment of present-day water cycle dynamics at continental to global scales is confounded by shortcomings in the observed record. Modeled assessments also yield conflicting results which are linked to differences in model structure and simulation protocol. Here we compare simulated gridded (1° spatial resolution) runoff from six terrestrial biosphere models (TBMs), seven reanalysis products, and one gridded surface station product in the contiguous United States (CONUS) from 2001 to 2005. We evaluate the consistency of these 14 estimates with stream gauge data, both as depleted flow and corrected for net withdrawals (2005 only), at the CONUS and water resource region scale, as well as examining similarity across TBMs and reanalysis products at the grid cell scale. Mean runoff across all simulated products and regions varies widely (range: 71 to 356 mm yr^{-1}) relative to observed continental-scale runoff (209 or 280 mm yr^{-1} when corrected for net withdrawals). Across all 14 products 8 exhibit Nash–Sutcliffe efficiency values in excess of 0.8 and three are within 10% of the observed value. Region-level mismatch exhibits a weak pattern of overestimation in western and underestimation in eastern regions—although two products are systematically biased across all regions—and largely scales with water use. Although gridded composite TBM and reanalysis runoff show some regional similarities, individual product values are highly variable. At the coarse scales used here we find that progress in better constraining simulated runoff requires standardized forcing data and the explicit incorporation of human effects (e.g., water withdrawals by source, fire, and land use change).

1. Introduction

Water balance calculations are becoming increasingly important for Earth system studies and link directly to the amount of reusable water available for wildland and managed environments, as well as human society. Both a general intensification of the hydrological cycle (Schwalm et al., 2011) and, more specifically, an increase in runoff are expected under climate change (Gerten et al., 2008). While numerous attempts (e.g., Alkama et al., 2011; Dai et al., 2009; Gerten et al., 2008; Haddeland et al., 2011; Milliman et al., 2008; Munier et al., 2012; Syed et al., 2010; Walling and Fang, 2003) have been made to observationally constrain continental to global runoff values large uncertainties remain; linked to *inter alia* spatiotemporal gaps in the observed record and the overall heterogeneity of discharge measurements.

A standard approach to address inconsistent observational records is the use of modeling. However, simulated runoff is similarly variable (e.g., Alkama et al., 2011; Gedney et al., 2006;

* Corresponding author at: Northern Arizona University, Center for Ecosystem Science and Society, Flagstaff, AZ 86011, USA. Tel.: +1 928 523 8413; fax: +1 928 523 7423.
E-mail address: christopher.schwalm@nau.edu (C.R. Schwalm).

http://dx.doi.org/10.1016/j.ecolmodel.2015.02.006
0304-3800/

Fig. 1. CONUS water resource regions. USGS water resource regions and major rivers in the CONUS domain. Prior to aggregation of product runoff, regions coverage converted from polygon (http://water.usgs.gov/GIS/huc.html) to 1° raster using the region with the maximum area of overlap for each grid cell.

Haddeland et al., 2011; Shi et al., 2011). Furthermore, large variation in runoff persists even when models use the same meteorological input data (Haddeland et al., 2011). A key source of this ambiguity is the diversity in how models simulate runoff in relation to global environmental change, e.g., changes in precipitation, temperature, net radiation, land cover/use, nitrogen deposition, fire regime, atmospheric concentrations of greenhouse gases, and irrigation (Caldwell et al., 2012; Gerten et al., 2008; Neilson, 1995; Sun et al., 2011). Model forcing data also plays a significant role in simulated runoff magnitude, with the choice of precipitation dataset alone altering simulated region-scale runoff estimates of up to 30% (Biemans et al., 2009). Furthermore, uncertainty in precipitation fields (inter-product spread) may propagate to a similar or greater magnitude of uncertainty in runoff estimates (Fekete et al., 2004).

As a first step to resolve ambiguity in simulated runoff, models must be confronted with observational records. Thus, the objective of this study is to evaluate a suite of modeled runoff estimates in a region with a dense network of stream gauges, the water resource regions (WRRs) of the contiguous United States (CONUS). Specifically, we use TBM simulations from the Regional and Continental Interim-Synthesis (RCIS); a synthesis activity part of the U.S. North American Carbon Program (NACP) with emphasis on the 2000 to 2005 time period (Huntzinger et al., 2012). To provide additional context for RCIS TBM runs we extend our intercomparison with reanalysis- and surface station-based estimates of runoff for the same spatiotemporal domain. Throughout this study we emphasize the coarse scales (from grid cell to CONUS) that typify TBM use as opposed to pristine watersheds and finer scales.

2. Data and methods

We compare observed runoff to 14 modeled runoff estimates. Observed runoff is derived from the *c.* 7400 currently operating stream gauges maintained by the United States Geological Survey (USGS)[1]. This network of stream gauges is organized by hydrologic unit codes (HUC[2]) using a standardized six-level nested hierarchy that, nationally for the United States, varies from 21 water resource regions (WRR) at level one (also called 2-digit HUC or HUC2) to *c.* 160,000 subwatersheds at level six[3]. For this study the 18 WRRs in the CONUS domain (Fig. 1) are used.

Monthly WRR runoff is available directly from the USGS WaterWatch[4] portal. The WRR values are based on daily flow data collected at stream gauges, stream gauge drainage basins, and HUC boundaries. Monthly runoff is computed for each basin by dividing average daily flow (scaled by days per month) by basin drainage area. These stream gauge-specific values are assigned to the target HUC level using area weighting and assuming runoff is uniform across each stream gauge basin. Thus, the WRR values leverage the full network of currently operating stream gauges.

The gridded (1° spatial resolution) monthly TBM runoff values (Table 1) span 2001 to 2005 on a water year basis. A water year is the period from October 1 to September 30 with the water year designation (e.g., 2005) corresponding to the year of the ending date. TBM output is taken from the NACP RCIS (Huntzinger et al., 2012), as well as an additional TBM, WaSSI, that simulates the same spatiotemporal domain as the RCIS but not gridded. The six TBM simulations, an ensemble of opportunity (Allen and Ingram, 2002), are comprised of model output generated from ongoing NACP and related studies. While this precludes attributing model-data mismatch to model structure and/or differences in driver data,

[1] http://waterwatch.usgs.gov/public/flow_stats/dv01d.zip.
[2] http://water.usgs.gov/GIS/huc.html.
[3] ftp://ftp-fc.sc.egov.usda.gov/NCGC/products/watershed/hu-standards.pdf.
[4] http://waterwatch.usgs.gov.

Table 1
Summary of runoff algorithms for modeled products. All products are monthly and resampled to 1° spatial resolution for the CONUS domain; WaSSI runoff available only in ungridded format at the 8-digit sub-basin level.

Model	Algorithm	Precipitation	Citation
DLEM	Runoff curve number method; function of effective precipitation (precipitation minus interception, plus snow melt), potential maximum soil moisture retention after runoff, and antecedent water in the soil column. Soil water in excess of saturation in the first soil layers becomes runoff	NARR	Liu et al. (2012), Tian et al. (2010)
LPJ-wsl	Sum of surface runoff from the top soil layer, subsurface runoff from the lower soil layer, and water percolating down through the lower soil layer. The surface and subsurface runoff are defined as the excess water above field capacity of the top and lower soil layers	CRU-TS 3.0[a]	Gerten et al. (2004), Sitch et al. (2003)
MC1	Sum of surface runoff, macropore (rapid through-flow via roots, cracks, etc.) flow, rapid through-flow and baseflow. Baseflow is a fraction of precipitation as modified by losses to transpiration or direct percolation by soil layer. Losses to transpiration are driven by a simplified version of Penmon-Montieth and transpiration by soil moisture factor for each plant functional type	PRISM[b]	Bachelet et al. (2001)
SiB3.1	Precipitation (scaled to GPCP) minus evapotranspiration. SiB3.1 natively calculates runoff using a defined allowable surface interception storage (puddle) depth, which accumulates as precipitation strikes the ground directly or runs off from the canopy. There is a maximum allowable puddle depth; any water accumulating above this is transferred to runoff, and is immediately in the ocean. As native runoff is unphysical (→0) scaled precipitation minus evapotranspiration is used instead	NCEP II[c] (precipitation scaled to GPCP[d])	Baker et al. (2010)
SLand	Sum of surface and subsurface runoff. Surface runoff is precipitation minus interception loss scaled by a non-linear function of relative soil wetness. Subsurface runoff is a nonlinear function of relative soil wetness and subsurface runoff at saturation. SLand is the land surface model component of the dynamic vegetation/carbon model VEGAS as used in the NACP RCIS	PREC/L[e]	Zeng et al. (2000)
WaSSI	Runoff is the sum of overland lateral flow, subsurface, and groundwater flow by an empirical method	PRISM[b]	Sun et al. (2011)
GLDAS NLDAS	Sum of surface and subsurface runoff. Surface runoff is a function of infiltration capacity and excess precipitation (non-evaporated inflow in excess of storage capacity by layer). Subsurface runoff is a linear function of subsurface moisture content above a minimum threshold. Runoff is simulated using the Noah version 2.8 land surface model as implemented in GLDAS-2 (Rodell et al., 2004) and NLDAS-2 (Xia et al., 2012)	Princeton[f] gauge-only CPC[g] (orographic adjust-ment using PRISM)	Ek et al. (2003), Schaake et al. (1996)
NARR	Sum of surface and subsurface runoff. Surface runoff is a function of infiltration capacity and excess precipitation (non-evaporated inflow in excess of storage capacity by layer). Subsurface runoff is a linear function of subsurface moisture content above a minimum threshold. Water budget does not close due to assimilation of precipitation and snow	–	Mesinger et al. (2006), Schaake et al. (1996)
NARR [GPCP]	Scaled NARR precipitation (rescaled GPCP[d] v2.1 data) minus NARR evapotranspiration. Rescaling aggregates all NARR grid cell values within a given 2.5° GPCP grid cell to area-averaged monthly values. These values are then linearly rescaled to monthly GPCP precipitation		
NARR [P–E]	NARR precipitation minus NARR evapotranspiration		
MERRA MERRA LAND	Sum of precipitation and spurious water source (non-zero due to land-atmosphere interface inconsistencies) minus evapotranspiration and changes in surface and sub-surface water (including interception reservoir, soil moisture, and snow)	–	Koster et al. (2000), Rienecker et al. (2011)
UDel	Precipitation minus evapotranspiration (from a modified Thornthwaite water-budget equation)	–	Willmott et al. (1985)

[a] CRU-TS—climatic research unit (CRU) time-series datasets (http://badc.nerc.ac.uk/browse/badc/cru/data/cru_ts_observation_databases).
[b] PRISM—parameter-elevation regressions on independent slopes model (http://www.prism.oregonstate.edu/).
[c] NCEP II—(Kanamitsu et al., 2002; http://www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysis2.html).
[d] GPCP—global precipitation climatology project (Huffman et al., 2009; http://www.ncdc.noaa.gov/oa/wmo/wdcamet-ncdc.html).
[e] PREC/L—precipitation reconstruction over land (Chen et al., 2002; http://www.esrl.noaa.gov/psd/data/gridded/data.precl.html).
[f] Princeton—princeton forcing dataset (Sheffield et al., 2006; http://hydrology.princeton.edu/data.php).
[g] gauge-only CPC—climate prediction center (Xie and Arkin, 1996; http://ftp.cpc.ncep.noaa.gov/precip/CPC_UNI_PRCP/).

it better mimics current practice as each run represents the "best estimate" of runoff for each TBM. A further concern is the time period of the TBM simulations. This five-year period limits the time horizon and resolution of the intercomparison (all other products are multi-decadal), is not sufficient to derive climatological mean runoff statistics, and could be biased by extreme events (although this bias would be present in all products). However, our goal is to evaluate the ability of the off-the-shelf NACP RCIS runs to simulate runoff during the RCIS study period. We make no claim that TBM evaluations transfer to other areas or periods of interest but also would highlight the need to understand the utility of such "as is" ensembles of opportunity regularly generated by the broader land surface modeling community.

To contextualize the NACP RCIS runs we also evaluate runoff derived from a North America-specific reanalysis, third generation reanalysis products from coupled models, and land-based reanalysis products (Table 1):

(1) Native runoff from the National Center for Environmental Prediction (NCEP) North American Regional Reanalysis product (NARR; Mesinger et al., 2006);
(2) runoff as NARR precipitation minus NARR evapotranspiration (hereafter NARR [P–E]);
(3) runoff as scaled NARR precipitation minus NARR evapotranspiration (hereafter NARR [GPCP]), where the scaling (Table 1) merges the spatial texturing of NARR with GPCP bias corrections for wind, gauge wetting, and gauge evaporation (Wei et al., 2013);

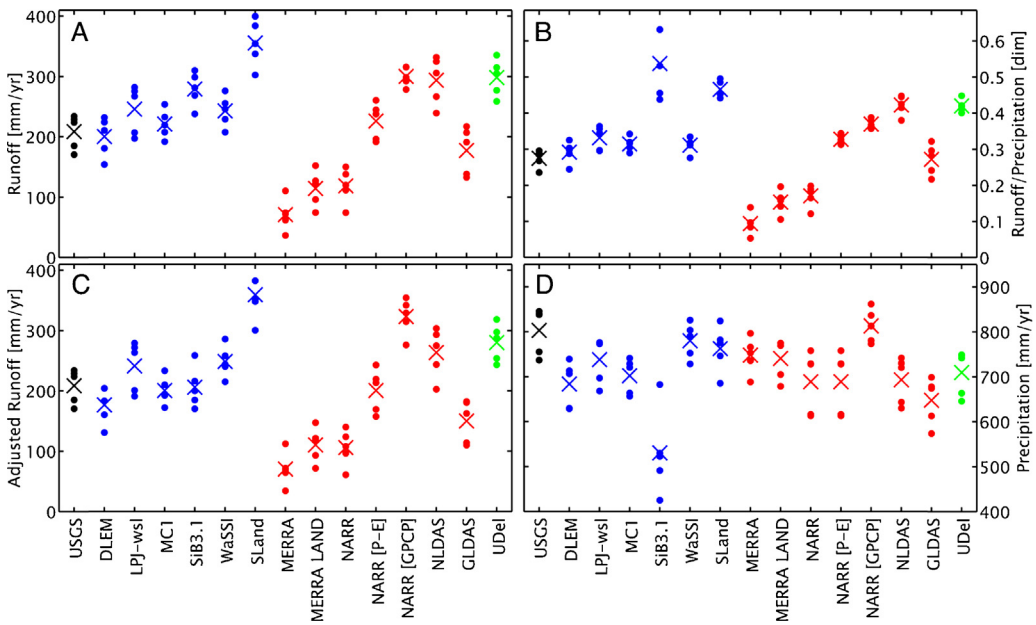

Fig. 2. Mean CONUS runoff, adjusted runoff, runoff normalized by precipitation, and precipitation. USGS precipitation is based on GPCC, an independent product (not used elsewhere). Adjusted runoff is runoff multiplied by the ratio of forcing precipitation by a standard precipitation reference, GPCC. Data values are mean 2001 to 2005 value, water year basis (cross) and individual years (circles). Color coding denotes product type: USGS stream gauge observations (black), TBMs (blue), reanalyses (red), and the UDel surface station based product (green). NARR variants are: NARR [GPCP]; NARR precipitation scaled to GPCP minus NARR precipitation, and NARR [P–E]; NARR precipitation minus NARR evapotranspiration. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

(4) native runoff from the NASA Modern Era Reanalysis for Research and Applications product (MERRA; Rienecker et al., 2011);
(5) native runoff from MERRA LAND (Reichle et al., 2011), an off-line land-only replay of the MERRA land model with precipitation forced using native MERRA precipitation merged with the NOAA Climate Prediction Center gauge-based data product (Xie and Arkin, 1996) and using the Fortuna-2.5 version of the catchment land surface model as opposed to the native MERRA version[5]; and
(6) native runoff simulated by the Noah version 2.8 land surface model (Ek et al., 2003) driven by version 2 of both the North American Land Data Assimilation System (NLDAS; Xia et al., 2012) and the Global Land Data Assimilation System (GLDAS; Rodell et al., 2004).

In addition to the seven variants of reanalyzed runoff we use an estimate of runoff derived from monthly water-budget fields (Table 1) calculated by the Center for Climatic Research, Department of Geography at the University of Delaware[6]. This estimate (hereafter UDel) is based on surface station records of temperature and precipitation. Both are first interpolated in space using Shepard's method and in time using climatologically aided interpolation (Willmott and Robeson, 1995). These interpolated estimates are then used as inputs in a modified Thornthwaite water-budget equation, assuming a soil water holding capacity of 150 mm, to estimate evapotranspiration (Willmott et al., 1985). Evapotranspiration is subtracted from precipitation to estimate runoff (1° spatial resolution).

Before analysis, all runoff products are aggregated to annual values on a water year basis (October to September) from 2001 to 2005, i.e., the temporal extent of NACP RCIS model runs. We use annual values as none of the TBM runs evaluated include river routing and effectively discharge all runoff into the ocean immediately. After integration in time, runoff is spatially aggregated, apart from USGS values which are already WRR-level estimates. For the comparison using WRRs, modeled runoff is aggregated to the relevant region (Fig. 1); for the CONUS-wide analysis, spatial aggregation is across all WRRs. We also compare TBMs (except WaSSI) and reanalysis products (including UDel) to each other at the 1° by 1° grid scale. For the 2001 to 2005 period water withdrawal data is available for 2005 only (Barber et al., 2009; Caldwell et al., 2012). We use this water use data to correct depleted USGS runoff to naturalized flow, the quantity estimated by the 14 runoff products, to evaluate 2005 model-data agreement. This correction is based on adding, by WRR, net withdrawals to reported USGS runoff. Model skill is quantified using bias and Nash–Sutcliffe efficiencies (NSE; Nash and Sutcliffe, 1970); the latter metric ranges from negative infinity to unity where unity indicates perfect model-data agreement.

3. Results

The reanalysis and UDel runoff values are in poor agreement with observed continental-scale runoff (Fig. 2A). Relative to the average CONUS stream gauge runoff of 209 mm yr^{-1}, mean modeled runoff from 2001 to 2005 is 166 mm yr^{-1} and 298 mm yr^{-1} for the reanalyses and UDel respectively. In a relative sense, the reanalysis products underestimate CONUS runoff by *c.* 25% while the annual runoff derived from the UDel product is almost 1.5 times greater than observed. In contrast, the mean value across all six TBMs is within 25% (258 mm yr^{-1}) of average stream gauge runoff. In addition to being less consistent with observations, the variability (standard deviation: 91 mm yr^{-1}) of reanalysis estimates is also *c.* 1.75 times greater than for TBMs (55 mm yr^{-1}).

Normalizing runoff by precipitation, or runoff as a proportion of precipitation, can be viewed as a control for the differences in precipitation data (Fig. 2D). For USGS data we use precipitation data not used in conjunction with any estimated runoff values, the Global Precipitation Climatology Centre Full Data Reanalysis

[5] http://gmao.gsfc.nasa.gov/research/merra/merra-land.php.
[6] http://climate.geog.udel.edu/~climate.

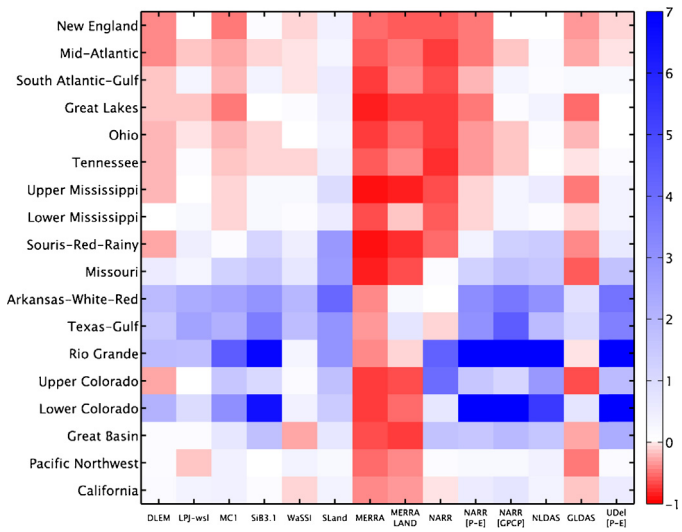

Fig. 3. Relative bias in runoff by water resource region and data product. Relative bias is calculated as $(\hat{y} - y)/y$ where y is USGS stream gauge runoff for a given region averaged over 2001 to 2005 (water year basis) and $\hat{y}$ is the corresponding simulated value. Red denotes underestimation; blue overestimation. Color range is bound by −1 to 7 to accommodate the lower limit of relative bias (−1) and its positive skew. Off-scale values are Rio Grande: NARR [P–E] (10), NARR [GPCP] (11) and Lower Colorado: NARR [P–E] (10), NARR [GPCP] (10), UDel (8). NARR variants are: NARR [GPCP]; NARR precipitation scaled to GPCP minus NARR precipitation, and NARR [P–E]; NARR precipitation minus NARR evapotranspiration. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

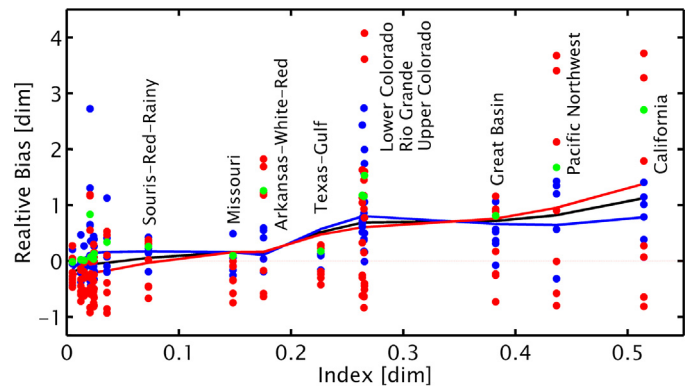

Fig. 4. Relative bias by water use. Relative bias is calculated as $(\hat{y} - y)/y$ where y is USGS stream gauge runoff for the 2005 calendar year (corrected to naturalized flow) and $\hat{y}$ is the corresponding simulated value. Water use is a dimensionless index of net withdrawals normalized by corrected USGS runoff. Each circle represents one CONUS water resource region (inset label for index >0.05 only) by product. The Lower Colorado, Rio Grande, and Upper Colorado differ by less than 0.01 in water use index. Circle color coding denotes product type: TBMs (blue), reanalyses (red), and the UDel surface station based product (green). Thick lines (loess smooth) shows increase in relative bias with increasing water use by product type: TBMs (blue), reanalyses – including UDel – (red), and all products (black). Thin red horizontal line is zero bias reference line. NARR variants are: NARR [GPCP]; NARR precipitation scaled to GPCP minus NARR precipitation, and NARR [P–E]; NARR precipitation minus NARR evapotranspiration. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

(GPCC[7]; Schneider et al., 2011). While normalization does decrease interannual variability for all estimated products except SiB3.1 (Fig. 2B), the general pattern of overestimation vs. underestimation remains largely unchanged. As another means to control for the differences in precipitation data we adjust runoff to match a precipitation standard. The adjustment is based on multiplying modeled runoff by the ratio of model-specific precipitation to the independent GPCC product. Similar to normalization this results in only minor changes (Fig. 2C). A large discrepancy between runoff and adjusted runoff would suggest that precipitation data differences confound simulated runoff. However, this is not apparent for either GPCC (Fig. 2D) or the forcing datasets used (not shown).

We acknowledge that normalization or adjustment cannot render an ensemble of opportunity into a standardized ensemble, as other non-standardized determinants of model output and skill, e.g., boundary conditions, spin-up procedures (Huntzinger et al., 2013), and analysis structure (Schwalm et al., 2013), are not amenable to a simple adjustment. However, the minimal changes in estimated runoff values strongly suggest that the variability in precipitation forcing datasets (Fig. 2D) does not confound model-data mismatch. Therefore, we limit our discussion to unadjusted runoff.

The region level mismatch between stream gauge and modeled runoff implies a geographic divide; WRRs east of and including the Mississippi river are generally underestimated whereas western WRRs are overpredicted (Fig. 3; Table A1). However, SLand systematically overpredicts all regions. The remaining TBMs also show a tendency toward positive biases, especially in the Rio Grande and Lower Colorado WRRs where SiB3.1 overestimates both by a factor of *c.* 7 (Fig. 3). Among the reanalysis and UDel products, MERRA underestimates runoff in every WRR while NLDAS overestimates all and UDel all but two (New England and Mid-Atlantic). Furthermore, there is no relationship between mismatch and WRR size (not shown).

[7] ftp://ftp.dwd.de/pub/data/gpcc/html/fulldata_v6_doi_download.html.

The weak east-west pattern suggests that water use, which is generally larger in the western CONUS (Table 1), may degrade consistency with USGS stream gauge observations (Caldwell et al., 2012). To investigate this potential dependency we use corrected USGS runoff values and an index of water use, i.e., net withdrawals normalized by corrected USGS runoff. Here, most (8 of 14) products show a dependency ($p < 0.05$) of relative bias on the index of water use. LPJ-wsl, MERRA, MERRA LAND, NARR, SLand, and WaSSI show no relationship ($p > 0.2$ but $p = 0.11$ for NARR). More generally, relative bias is highly sensitive to increases in water use across all products (Fig. 4). For every 0.10 increase in the water use index, relative bias increases, on average, by 0.25 ($p < 0.0001$). Also, WRRs with higher index values are not only preferentially overestimated but also show a greater spread in estimated values. The increase in relative bias with increasing water use similarly holds when using a given withdrawal source (e.g., surface vs. groundwater withdrawals and eight standard categories [Barber et al., 2009]) to index water use. Only industrial and thermoelectric power water use have negative sensitivities (−0.2 and −1.6 respectively). Also, relative bias is more sensitive to groundwater (0.28) than surface water (0.18) withdrawals.

Using corrected USGS runoff we find that eight of the 14 products evaluated exceed the customary NSE threshold for "good" model-data agreement (≥0.8) (Fig. 5). This includes four of the six TBMs evaluated. Furthermore, TBMs exhibit a mean NSE of 0.78 vs. 0.61 for reanalyses products (although the land-based reanalysis products have NSE values of 0.85 and 0.86) and 0.94 for the UDel product. These NSE values are higher than when using depleted USGS runoff values as comparator; where only six of the 14 products exceed the 0.8 threshold. Products showing less skill, i.e., lower NSE values, exhibit the largest gains in NSE when switching to corrected USGS values (not shown). This does not hold for MC1 and WaSSI where using corrected USGS values results in a 0.08 and 0.01 loss in NSE respectively.

While mean gridded runoff from TBMs and reanalysis products (including UDel) are within 6 mm month^{-1} or *c.* 25% (22 and 17 mm month^{-1} for TBMs and reanalysis respectively), these composite values (Fig. 6) mask highly variable individual product

estimates of runoff and spatial gradients in grid cell level differences (Fig. A1). Across the Great Plains and western CONUS both TBMs and reanalysis products show similar means. Despite this broad similarity, reanalysis products tend to simulate slightly more runoff west of the Mississippi (save the Pacific Northwest), with the largest relative differences located in the Southwest. In contrast, grid cells in six eastern WRRs (Lower Mississippi, Ohio, Tennessee, South Atlantic-Gulf, Mid-Atlantic, and New England) and the lower Great Lakes regions are not in agreement, with TBMs simulating more runoff than reanalysis products and UDel. The spread in reanalysis runoff (coefficient of variation) averages 79% in space with the highest values in the Southwest (Fig. 6). TBMs also exhibit a large degree of spread (mean coefficient of variation: 53%) with the largest variability in the Mountain West and Southwest.

4. Discussion

This study provides an evaluation of model and observationally derived continental, WRR, and grid cell-scale surface annual runoff in the CONUS domain from 2001 to 2005. At the region-scale, the 14 products are consistent in their lack of agreement with stream gauge values. There is neither a coherent spatial pattern across WRRs, nor a region where all products exhibit a uniformly high (or low) level of consistency with stream gauge data. General agreement is seen only at the grid cell level in composite means (averages across all TBMs or reanalysis products) west of the Mississippi, but individual products are highly variable. Nonetheless, we can use CONUS-wide runoff to place the 14 products into three generic tiers based on model skill, with the first tier products generally agreeing the closest with USGS observations.

The first tier products comprise TBMs (DLEM, LPJ-wsl, SiB3.1, WaSSI), land-based reanalyses (GLDAS, NLDAS), and precipitation minus evapotranspiration estimates (NARR [GPCP], UDel). All of these products show "good" model-data agreement (NSE ≥ 0.8) and generally low relative biases (≤0.15 and 0.25 in magnitude for TBMs and other first tier products respectively) for CONUS-wide runoff. The second tier of skill is occupied by NARR [P–E] and MC1. These show intermediate NSE values of *c.* 0.7 and negative relative biases ≤0.2. It is noteworthy that the products based on based on precipitation minus evapotranspiration (UDel, NARR [P–E], and NARR [GPCP]) have reasonable skill levels; NSE range: 0.73 to 0.94 and relative biases ≤0.2. Compared to TBMs, which require substantial infrastructure to implement and run, precipitation minus evapotranspiration is trivial to estimate using readily accessible data products suggesting utility in large-scale diagnostic runoff studies.

The lowest tier consisting of reanalysis from coupled models (MERRA, MERRA LAND, and NARR) and SLand exhibits little to no skill (NSE range: 0.15 to 0.55) with the three reanalysis products biased low and SLand high. These patterns hold when either uncorrected or corrected USGS runoff is used as comparator. The underestimation by NARR has been previously documented using both CONUS River Forecast Center regions (Sheffield et al., 2012) and the Mississippi River basin (Kumar and Merwade, 2011). In contrast, both MERRA variants have shown higher model skill relative to USGS data (Reichle et al., 2011) than in this study. Methodological differences in evaluating MERRA and MERRA LAND skill (e.g., scale mismatch of watersheds, different temporal extent and granularity) preclude reconciliation of these findings although both studies show MERRA LAND outperforming MERRA.

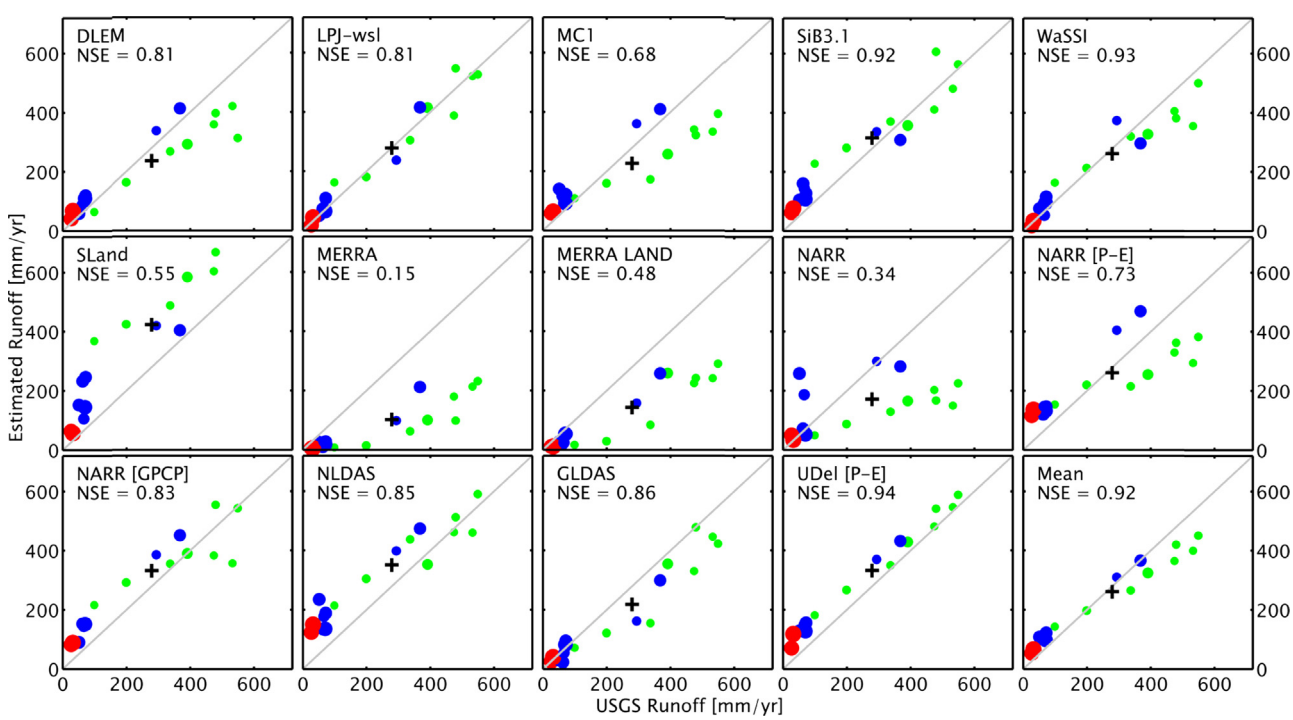

Fig. 5. Estimated and observed corrected mean runoff for 2005 only. Depleted flows from USGS are corrected to naturalized flow by adding back net withdrawals in each of the 18 CONUS water resource regions. Estimated runoff product and Nash–Sutcliffe efficiency (NSE) given in upper-left of each panel. NARR variants are: NARR [GPCP]; NARR precipitation scaled to GPCP minus NARR precipitation, and NARR [P–E]; NARR precipitation minus NARR evapotranspiration. Each circle represents the 2001 to 2005 (water year basis) mean for one of the 18 CONUS water resource regions. Mean is average across all 14 products by WRR. Symbol size is proportional to net withdrawals by region (i.e., larger symbols indicate larger net withdrawals relative to corrected USGS runoff). Symbol color coding denotes geography: eastern (green), from the Souris-Red-Rainy, Upper Mississippi, and Lower Mississippi regions eastward or numbers 1 to 9, and western (blue) regions, from the Missouri, Arkansas-White-Red, and Texas-Gulf regions westward or numbers 10 to 18. The Lower Colorado and Rio Grande are colored red and have the highest ratios (in excess of 0.4) of net withdrawals to corrected flow. Plus sign shows CONUS-wide runoff. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

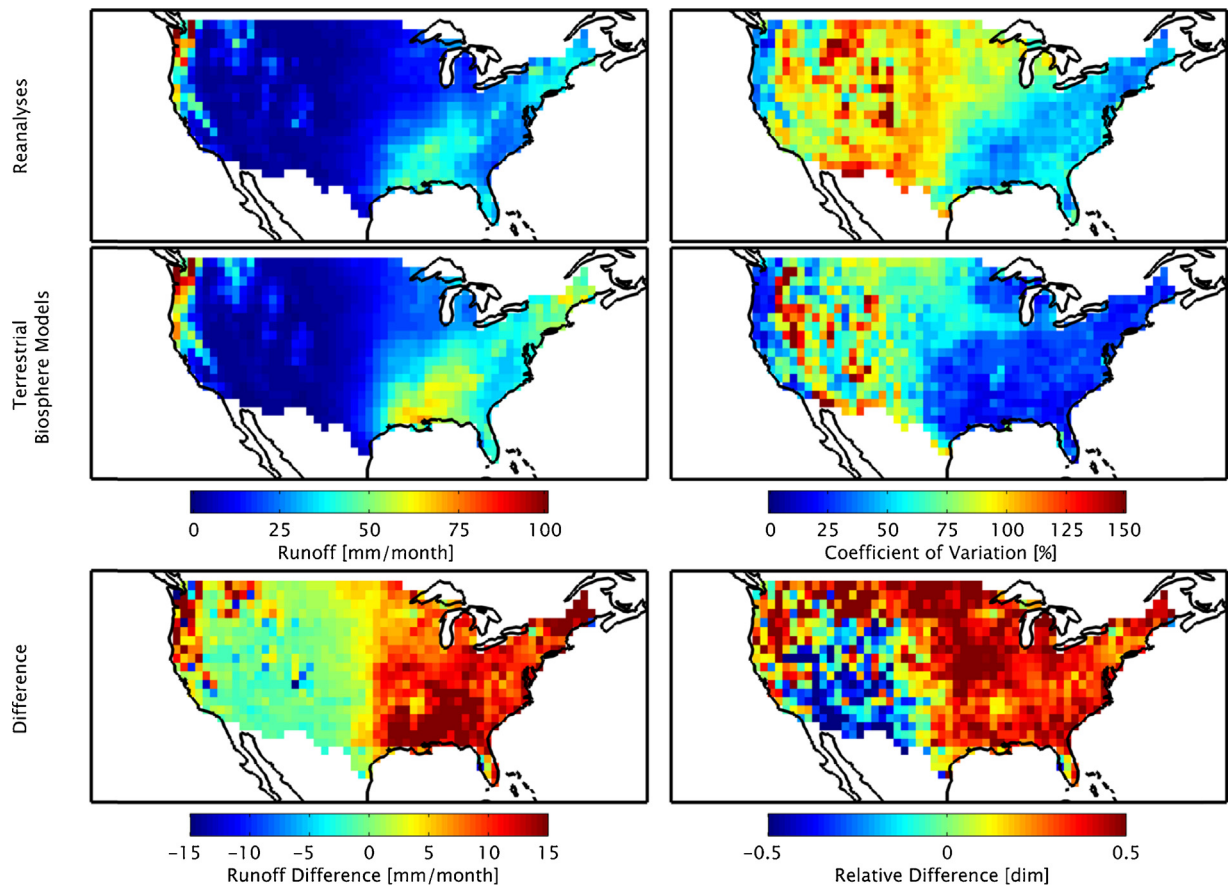

Fig. 6. Spatial patterns of runoff. Maps show monthly mean runoff and its coefficient of variation (mean/standard deviation in %) by grid cell from 2001 to 2005 (water year basis). Difference maps show runoff difference (left; terrestrial biosphere model mean minus the reanalysis product mean) and relative difference (right; runoff difference normalized by the mean across all 13 gridded products). Reanalysis includes all seven reanalysis products and UDel. WaSSI is not included as its output is catchment-scale and not gridded. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Identifying the root causes of model-data mismatch in runoff estimates is an ongoing challenge, especially for short-term ensembles of opportunity like the NACP RCIS. Such ensembles do not control for possible confounding factors such as forcing data (Biemans et al., 2009) and simulation protocol (Huntzinger et al., 2013). Furthermore, while USGS runoff can be corrected for water withdrawals – to better mimic the 14 products evaluated here – the correction is imprecise. Water use estimates for some withdrawal sources are based on ancillary data across multiple years (Barber et al., 2009); thus degrading model-data intercomparison as a function of time. In addition, USGS withdrawal data does not include water use for hydroelectric power generation and reclaimed wastewater (an important water source for thermoelectric power generation in arid CONUS regions) (Barber et al., 2009). Similarly, mining, although a standard source category, does not account for all mining-related activities; aquifer dewatering, water use for smelting, refining petroleum, or slurry pipeline operations are excluded (Barber et al., 2009).

Beyond this unspecified but known water use, it is not a given that all withdrawals decrease runoff. The net change in runoff depends on surface-subsurface interactions (Leung et al., 2010) as well as withdrawal sources, which may have contrasting impacts on runoff (Leng et al., 2014). As an example, ground water mining may act to increase runoff assuming the aquifer source is not connected to surface water (Caldwell et al., 2012). In this study industrial and thermoelectric power water use are shown to increase runoff, for industrial use this applies for both ground and surface water separately as well. Runoff may also increase due to fire (Certini, 2005; Verkaik et al., 2013) and land use change (Foley et al., 2005). None of these factors are tracked explicitly by any product.

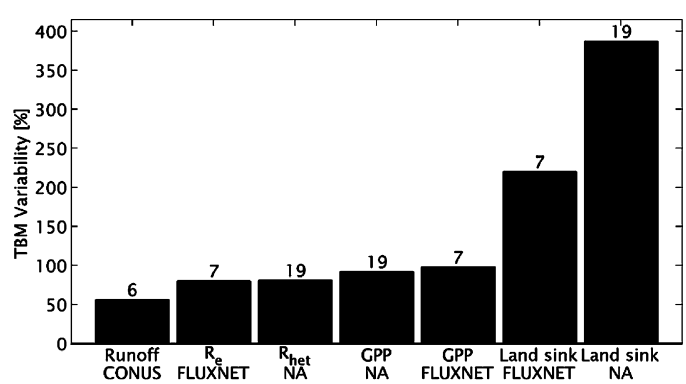

Fig. 7. Variability in TBM estimates. Variability is defined as range normalized by mean value expressed as a percentage. Inset numbers: number of NACP RCIS TBMs evaluated. Bar labels refer to variable (upper label: ecosystem respiration [R_e]; heterotrophic respiration [R_{het}]; gross primary productivity [GPP]) and spatial domain (lower label: contiguous United States [CONUS]; grid cells co-located with flux towers across the North American domain [FLUXNET]; full North American domain [NA]).

5. Conclusions

The TBMs examined here are, in general, able to reproduce observed patterns in CONUS-wide runoff over the 2001 to 2005 analysis period. However, several products exhibit profound biases

and spatial heterogeneity in model skill. Diagnosing mismatch between stream gauge runoff and any given product is confounded by their coarse scale and off-the-shelf nature. Runoff is fundamentally a process that occurs on the catchment scale and multiple catchments within a large WRR may act in a compensatory manner that is not resolvable at a 1° spatial resolution or regional scale. Similarly, mismatch may also be influenced by the choice of forcing data used in a particular TBM. As such it is difficult to attribute differences between simulated and observed values solely to intrinsic characteristics of the models themselves. While this suggests that a constrained ensemble (Huntzinger et al., 2013) is preferable to an ensemble of opportunity it also reveals a clear dilemma: Even though several factors that may impact model-data mismatch are only resolvable at finer scales, this does not obviate the need to evaluate runoff simulated at coarser scales for ensembles of opportunity (such as the NACP RCIS).

A further complication is the tendency of several products, especially the TBMs, to exhibit the "right answers for the wrong reasons" (Kirchner, 2006). This is most visible using uncorrected USGS runoff for western WRRs with high water use, e.g., Rio Grande. That is, TBM estimates show consistency with depleted USGS stream gauge data without explicit consideration of anthropogenic perturbations to the water cycle (e.g., land use change, fire effects, and water withdrawals). This compensatory effect can be related to model formulation, e.g., bulk parameters compensating for a lack of physically-based equations at relevant scales (Kirchner, 2006), or overall complexity in large heterogeneous systems at coarser scales as studied here (McDonnell et al., 2007). Alternatively, model formulation of those processes included can compensate for those processes excluded.

For TBMs specifically, contrasting model skill and model spread for the carbon and water cycles offers a bridge between terrestrial and hydrological research. Model-model (Huntzinger et al., 2013) and model-data (Rackza et al., 2013) carbon flux intercomparisons have been conducted using NACP RCIS runs. Although each intercomparison varies in analytical framework both can be contrasted with TBM outputs evaluated in this study. Given the lack of gridded observational products that can serve as unambiguous references (Luo et al., 2012) we examine model spread, i.e., variability across the TBM ensemble (calculated as range normalized by mean value expressed as a percentage). This also allows diverse land surface processes to be compared directly. Using grid cells with co-located flux towers (Raczka et al., 2013) and the full North American domain (Huntzinger et al., 2012) TBM carbon cycle variability ranges from 81% to 387% (Fig. 7). For runoff the corresponding CONUS-wide variability is markedly lower (=56%). Although the water and carbon cycles are tightly coupled, this suggests that the limited resources available for TBM development should preferentially target key constraints of the carbon cycle. Ultimately, improving the characterization of the carbon cycle will improve model-data agreement for the water cycle and vice versa. However, model-model and model-data studies that simultaneously address both the water and carbon cycles are a necessary precondition to realize this goal.

More generally, there is a pressing need to confront TBMs (indeed, all simulated products) with observations at multiple scales (Gerten, 2013), particularly the characteristic scale at which the process occurs, such as the catchment scale for runoff. Furthermore, intercomparisons of model structure relative to skill (Schwalm et al., 2010) will help inform model development and reduce ensemble spread and error. For runoff specifically, the use of standardized forcing data (especially for TBMs; Wei et al., 2013), the explicit incorporation of human effects (e.g., water cycle management, fire, land use change), higher quality water withdrawals data (Caldwell et al., 2012), more realism in the representation of subgrid variability, and river routing schemes are all necessary to improve model-data agreement. Large-resource model-intercomparison projects that use a constrained protocol (e.g., Huntzinger et al., 2013) hold great promise in furthering our understanding of Earth system dynamics at multiple scales and provide a framework to evaluate model skill and spread of carbon and water cycles simultaneously.

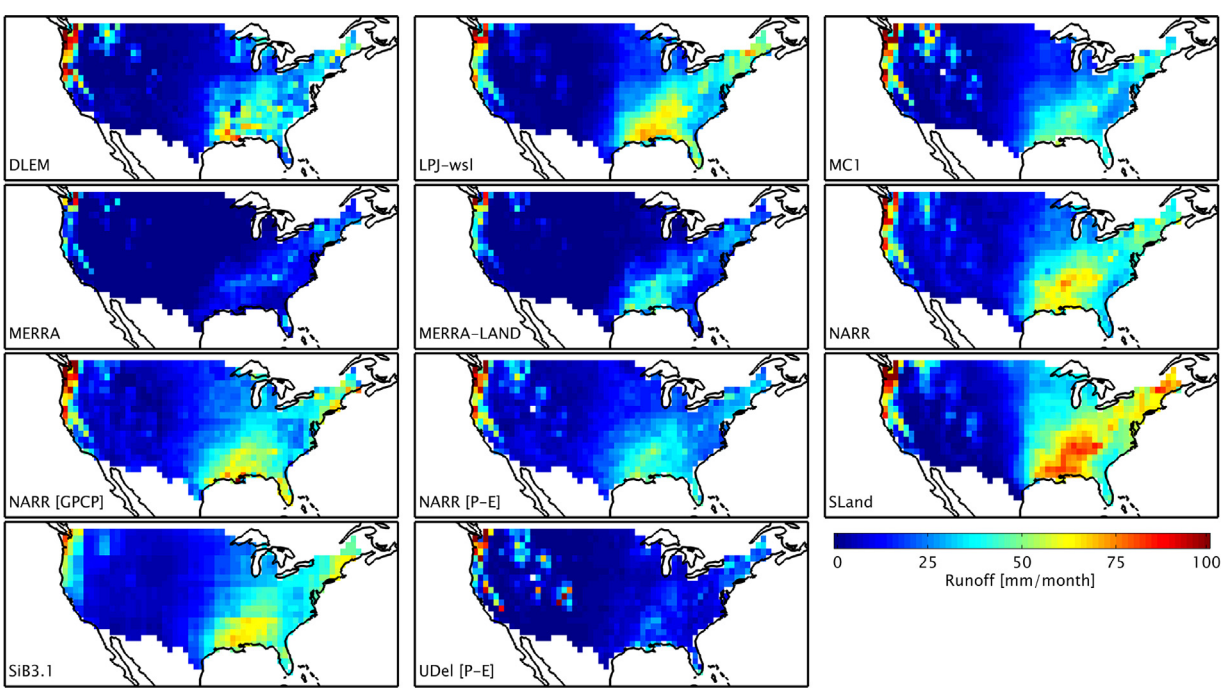

Fig. A1. Spatial patterns of runoff. Maps show monthly mean runoff by product (alphabetical from left to right) and grid cell from 2001 to 2005 (water year basis). USGS and WaSSI are not displayed as these products are catchment-scale and not gridded.

Table A1
Mean runoff (mm/yr) from 2001 to 2005 (water year basis) and water use (2005 calendar year only) by water resource region (WRR). Water use expressed as an index; the ratio of gross or net withdrawals (taken from Caldwell et al., 2012) to USGS depleted runoff.

WRR	Product															Water use	
	DLEM	LPJ-wsl	MC1	SiB3.1	WaSSI	SLand	MERRA	MERRA LAND	NARR	NARR [P-E]	NARR [GPCP]	NLDAS	GLDAS	UDel [P-E]	USGS	Gross Withdrawal	Net Withdrawal
New England	339	598	312	623	548	811	263	236	254	309	606	595	399	545	580	0.05	0.01
Mid-Atlantic	307	459	378	486	520	689	205	273	166	292	448	604	373	506	528	0.23	0.02
South Atlantic-Gulf	352	541	339	561	403	652	130	249	139	345	552	471	519	518	423	0.17	0.03
Great Lakes	285	295	170	339	363	497	59	87	101	171	371	458	160	351	335	0.37	0.02
Ohio	397	501	391	470	538	715	152	232	159	337	424	567	414	541	516	0.20	0.01
Tennessee	489	659	512	563	570	899	259	366	136	388	518	626	606	713	605	0.27	0.01
Upper Mississippi	184	234	203	282	273	444	16	40	80	210	293	354	125	298	225	0.33	0.03
Lower Mississippi	534	655	459	637	571	792	168	428	215	489	673	567	480	717	520	0.25	0.11
Souris-Red-Rainy	61	119	92	180	123	312	7	16	38	113	191	199	48	134	82	0.09	0.03
Missouri	71	60	102	116	75	171	8	16	50	103	125	118	18	123	46	0.88	0.36
Arkansas-White-Red	162	185	195	220	166	286	32	69	57	230	261	222	103	270	56	0.63	0.34
Texas-Gulf	166	234	207	300	184	257	44	114	60	261	359	190	133	293	66	0.78	0.41
Rio Grande	26	26	50	72	12	36	5	8	49	97	115	75	9	74	9	2.47	1.25
Upper Colorado	26	38	95	76	44	100	10	12	185	94	79	139	12	105	37	1.00	0.37
Lower Colorado	26	16	34	63	12	20	3	4	14	95	96	53	14	78	8	3.39	1.97
Great Basin	34	35	53	87	23	53	10	9	85	81	94	84	24	103	32	0.67	0.36
Pacific Northwest	299	232	387	297	381	348	125	162	303	343	345	392	156	323	278	0.22	0.08
California	237	271	280	220	184	270	117	130	189	310	351	264	173	313	199	0.64	0.42

Acknowledgements

Support for this activity came from the US National Aeronautics and Space Administration (NASA) Terrestrial Ecology Program (NNH10AN68I and NNX11AO08A) and from the Terrestrial Ecosystem Science Focus Area at the Office of Biological and Environmental Research, US Department of Energy (DOE). Oak Ridge National Laboratory is managed by UT-Battelle for DOE under contract DE-AC05-00OR22725. IB was sponsored by the National Science Foundation (NSF) (ATM-0425247, AGS-0425247-13 and AGS1049041), the Department of Commerce/National Oceanic and Atmospheric Administration (NA08AR4320893), NASA (NNX06AC75G, NNX08AM56G, NNX12AP86G, NNX10AT41G, and NNX11AB87G), DOE (DE-FG02-06ER64317), and the National Institute for Climate Change Research (MTU050516Z14). PC and GS were partially supported by the NSF EaSM program (AGS-1049200) at North Carolina State University. HT was supported by the NASA Interdisciplinary Science Program (NNX11AD47G and NNX10AU06G). We thank Mac Post for his contributions to this research. This is a contribution of the North American Carbon Program.

Appendix A. Appendix

Fig. A1 and Table A1

References

Alkama, R., Decharme, B., Douville, H., Ribes, A., 2011. Trends in global and basin-scale runoff over the late twentieth century: methodological issues and sources of uncertainty. J. Clim. 24, 3000–3014, http://dx.doi.org/10.1175/2010JCLI3921.1.

Allen, M.R., Ingram, W.J., 2002. Constraints on future changes in climate and the hydrologic cycle. Nature 419 (6903), 224–232.

Bachelet, D., Lenihan, J.M., Daly, C., Neilson, R.P., Ojima, D.S., Parton, W.J., 2001 June. MC1. A dynamic vegetation model for estimating the distribution of vegetation and associated ecosystem fluxes of carbon, nutrients, and water. In: USFS GTR PNW-GTR-508. USDA Forest Service Pacific Northwest Research Station, Portland, Oregon, USA.

Baker, I.T., Denning, A.S., Stockli, R., 2010. North American gross primary productivity: regional characterization and interannual variability,. Tellus Ser. B 62B, 533–549, http://dx.doi.org/10.1111/j. 1600-0889.2010.00492.x.

Barber, N.L., Hutson, S.S., Linsey, K.S., Lovelace, J.K., Maupin, M.A., 2009. Estimated Use of Water in the United States in 2005. US Geological Survey, Reston, VA, pp. 52.

Biemans, H., Hutjes, R., Kabat, P., Strengers, B., Gerten, D., Rost, S., 2009. Impacts of precipitation uncertainty on discharge calculations for main river basins. J. Hydrometeorol. 10, 1011–1025.

Caldwell, P.V., Sun, G., McNulty, S.G., Cohen, E.C., Moore Myers, J.A., 2012. Impacts of impervious cover, water withdrawals, and climate change on river flows in the conterminous US. Hydrol. Earth Syst. Sci. 16, 2839–2857, 10.5194/hess-16-2839-2012.

Certini, G., 2005. Effects of fire on properties of forest soils: a review. Oecologia 143 (1), 1–10.

Chen, M., Xie, P., Janowiak, J.E., Arkin, P.A., 2002. Global land precipitation: a 50-yr monthly analysis based on gauge observations. J. Hydrometeorol. 3, 249–266.

Dai, A., Qian, T., Trenberth, K.E., Milliman, J.D., 2009. Changes in continental freshwater discharge from 1948–2004. J. Clim. 22, 2773–2791.

Ek, M.B., Mitchell, K.E., Lin, Y., Rogers, E., Grunman, P., Koren, V., Gayno, G., Tarpley, J.D., 2003. Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model. J. Geophys. Res. 108, 8851, http://dx.doi.org/10.1029/2002JD003296.

Fekete, B.M., Vörösmarty, C.J., Roads, J.O., Willmott, C.J., 2004. Uncertainties in precipitation and their impacts on runoff estimates. J. Clim. 17, 294–304.

Foley, J.A., DeFries, R., Asner, G.P., Barford, C., Bonan, G., Carpenter, S.R., et al., 2005. Global consequences of land use. Science 309 (5734), 570–574.

Gedney, N., Cox, P.M., Betts, R.A., Boucher, O., Huntingford, C., Stott, P.A., 2006. Detection of a direct carbon dioxide effect in continental river runoff records. Nature 439, 835–838.

Gerten, D., Schaphoff, S., Haberlandt, U., et al., 2004. Terrestrial vegetation and water balance—hydrological evaluation of a dynamic global vegetation model. J. Hydrol. 286, 249–270.

Gerten, D., Rost, S., von Bloh, W., Lucht, W., 2008. Causes of change in 20th century global river discharge. Geophys. Res. Lett. 35, L20405, http://dx.doi.org/10.1029/2008GL035258.

Gerten, D., 2013. A vital link: water and vegetation in the Anthropocene. Hydrol. Earth Syst. Sci. 17 (10).

Haddeland, I., Clark, D.B., Franssen, W., Ludwig, F., Voß, F.R.A.N.K., Arnell, N.W., et al., 2011. Multimodel estimate of the global terrestrial water balance: setup and first results. J. Hydrometeorol. 12 (5), 869–884.
Huffman, G.J., Adler, R.F., Bolvin, D.T., Gu, G., 2009. Improving the global precipitation record: GPCP version 2.1. Geophys. Res. Lett. 36, L17808, http://dx.doi.org/10.1029/2009GL040000.
Huntzinger, D.N., Post, W., Michalak, A., Wei, Y., Jacobson, A., West, T.O., Baker, I., Chen, J., Davis, K., Hayes, D., Hoffman, F., Jain, A., Liu, S., McGuire, D., Neilson, R., Raczka, B., Poulter, B., Tian, H., Thornton, P., Tomelleri, E., Viovy, N., Xiao, J., Zeng, N., Zhao, M., Cook, R., 2012. North American Carbon Project (NACP) regional interim synthesis: terrestrial biospheric model intercomparison. Ecol. Modell. 232, 144–157, http://dx.doi.org/10.1016/j.ecolmodel.2012.02.004.
Huntzinger, D.N., Schwalm, C., Michalak, A.M., Schaefer, K., King, A.W., Wei, Y., Jacobson, A., Liu, S., Cook, R.B., Post, W.M., Berthier, G., Hayes, D., Huang, M., Ito, A., Lei, H., Lu, C., Mao, J., Peng, C.H., Peng, S., Poulter, B., Riccuito, D., Shi, X., Tian, H., Wang, W., Zeng, N., Zhao, F., Zhu, Q., 2013. The North American Carbon Program Multi-scale synthesis and Terrestrial Model Intercomparison Project—Part 1: Overview and experimental design. Geosci. Model Dev. Discuss. 6, 2121–2133, http://dx.doi.org/10.5194/gmd-6-2121-2013.
Kanamitsu, M., Ebisuzaki, W., Woollen, J., Yang, S.-K., Hnilo, J.J., co-authors., 2002. NCEP-DOE AMIP-II reanalysis (R-2). Bull. Am. Meterol. Soc. 83 (11), 1631–1643.
Kirchner, J.W., 2006. Getting the right answers for the right reasons: linking measurements, analyses, and models to advance the science of hydrology. Water Resour. Res. 42, W03S04, http://dx.doi.org/10.1029/2005WR004362.
Koster, R.D., Suarez, M.J., Ducharne, A., Stieglitz, M., Kumar, P., 2000. A catchment-based approach to modeling land surface processes in a general circulation model. 1. Model structure. J. Geophys. Res. 105, 24809–24822.
Kumar, S., Merwade, V., 2011. Evaluation of NARR and CLM3.5 outputs for surface water and energy budgets in the Mississippi River Basin. J. Geophys. Res. 116, D08115, http://dx.doi.org/10.1029/2010JD014909.
Leng, G., Huang, M., Tang, Q., Gao, H., Leung, L.Y.R., 2014. Modeling the effects of groundwater-fed irrigation on terrestrial hydrology over the conterminous United States. J. Hydrometeorol. 15 (3), 957–972, http://dx.doi.org/10.1175/JHM-D-13-049.1.
Leung, L.Y.R., Huang, M., Qian, Y., Liang, X., 2010. Climate-soil-vegetation control on groundwater table dynamics and its feedbacks in a climate model. Clim. Dyn. 36 (1–2), 57–81, http://dx.doi.org/10.1007/s00382-010-0746-x.
Luo, Y.Q., Randerson, J.T., Abramowitz, G., Bacour, C., Blyth, E., Carvalhais, N., et al., 2012. A framework for benchmarking land models. Biogeosciences 9 (10), 3857–3874, http://dx.doi.org/10.5194/bg-9-3857-2012.
Liu, M., Tian, H., Lu, C., Xu, X., Chen, G., Ren, W., 2012. Effects of multiple environment stresses on evapotranspiration and runoff over the eastern China. J. Hydrol. 426–427, 39–54, http://dx.doi.org/10.1016/j.jhydrol.2012.01.009.
McDonnell, J.J., et al., 2007. Moving beyond heterogeneity and process complexity: a new vision for watershed hydrology. Water Resour. Res. 43, W07301, http://dx.doi.org/10.1029/2006WR005467.
Mesinger, F., et al., 2006. North American regional reanalysis. Bull. Am. Meteorol. Soc. 87, 343–360.
Milliman, J.D., Farnsworth, K.L., Jones, P.D., Xu, K.H., Smith, L.C., 2008. Climatic and anthropogenic factors affecting river discharge to the global ocean, 1951–2000. Global Planet. Change 62, 187–194.
Munier, S., Palanisamy, H., Maisongrande, P., Cazenave, A., Wood, E.F., 2012. Global runoff anomalies over 1993–2009 estimated from coupled Land–Ocean–Atmosphere water budgets and its relation with climate variability. Hydrol. Earth Syst. Sci. 16 (10), 3647–3658.
Nash, J., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models Part I—A discussion of principles. J. Hydrol. 10 (3), 282–290.
Neilson, R.P., 1995. A model for predicting continental-scale vegetation distribution and water balance. Ecol. Appl. 5, 362–385, http://dx.doi.org/10.2307/1942028.
Raczka, B.M., Davis, K.J., Huntzinger, D., Neilson, R.P., Poulter, B., Richardson, A.D., et al., 2013. Evaluation of continental carbon cycle simulations with North American flux tower observations. Ecological Monographs 83 (4), 531–556.
Reichle, R.H., Koster, R.D., De Lannoy, G.J., Forman, B.A., Liu, Q., Mahanama, S.P., Touré, A., 2011. Assessment and enhancement of MERRA land surface hydrology estimates. J. Clim. 24, 6322–6338, http://dx.doi.org/10.1175/JCLI-D-10-05033.1.
Rienecker, Michele M., Suarez, Max J., Gelaro, Ronald, et al., 2011. MERRA—NASA's modern-era retrospective analysis for research and applications. J. Clim. 24 (14), 3624–3648.
Rodell, M., Houser, P.R., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C.-J., Arsenault, K., Cosgrove, B., Radakovich, J., Bosilovich, M., Entin, J.K., Walker, J.P., Lohmann, D., Toll, D., 2004. The global land data assimilation system. Bull. Am. Meteor. Soc. 85 (3), 381–394.
Schaake, J.C., Koren, V.I., Duan, Q.-Y., Mitchell, K., Chen, F., 1996. Simple water balance model for estimating runoff at different spatial and temporal scales. J. Geophys. Res. 101, 7461–7475.
Sheffield, J., Goteti, G., Wood, E.F., 2006. Development of a 50-yr high-resolution global dataset of meteorological forcings for land surface modeling. J. Clim. 19 (13), 3088–3111.
Sheffield, J., Livneh, B., Wood, E.F., 2012. Representation of terrestrial hydrology and large-scale drought of the continental United States from the North American Regional Reanalysis. J. Hydrometeor. 13, 856–876, http://dx.doi.org/10.1175/JHM-D-11-065.1.
Schneider, Udo, Andreas, Becker, Peter, Finger, Anja, Meyer-Christoffer, Bruno, Rudolf, Markus, Ziese, 2011. GPCC Full Data Reanalysis Version 6.0 at 1.0: Monthly Land-Surface Precipitation from Rain-Gauges built on GTS-based and Historic Data., http://dx.doi.org/10.5676/DWD_GPCC/FD_M_V6_100.
Schwalm, C.R., Williams, C.A., Schaefer, K., Anderson, R., Arain, M.A., Baker, I., et al., 2010. A model-data intercomparison of CO2 exchange across North America: results from the North American Carbon Program site synthesis. J. Geophys. Res. Biogeosci. 11, 5, http://dx.doi.org/10.1029/2009JG001229.
Schwalm, C.R., Williams, C.A., Schaefer, K., 2011. Carbon consequences of global hydrologic change, 1948–2009. J. Geophys. Res. 116, G03042, http://dx.doi.org/10.1029/2011JG001674.
Schwalm, C.R., Huntzinger, D.N., Michalak, A.M., Fisher, J.B., Kimball, J.S., Mueller, B., et al., 2013. Sensitivity of inferred climate model skill to evaluation decisions: a case study using CMIP5 evapotranspiration. Environ. Res. Lett. 8 (2), 024028.
Shi, X., Mao, J., Thornton, P.E., Hoffman, F.M., Post, W.M., 2011. The impact of climate, CO_2, nitrogen deposition and land use change on simulated contemporary global river flow. Geophys. Res. Lett. 38, L08704, http://dx.doi.org/10.1029/2011GL046773.
Sitch, S., Smith, B., Prentice, I.C., Arneth, A., Bondeau, A., Cramer, W., et al., 2003. Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model. Global Change Biol. 9 (2), 161–185.
Sun, G., Caldwell, P., Noormets, A., Cohen, E., McNulty, S.G., Treasure, E., Domec, J.-C., Mu, Q., Xiao, J., John, R., Chen, J., 2011. Upscaling key ecosystem functions across the conterminous United States by a water-centric ecosystem model. J. Geophys. Res. 116, G00J05, http://dx.doi.org/10.1029/2010JG001573.
Syed, T.H., Famiglietti, J.S., Chambers, D.P., Willis, J.K., Hilburn, K., 2010. Satellite-based global-ocean mass balance estimates of interannual variability and emerging trends in continental freshwater discharge. Proc. Natl. Acad. Sci. 107 (42), 17916–17921.
Tian, H., Chen, G., Liu, M., Zhang, C., Sun, G., Lu, C., Xu, X., Ren, W., Pan, S., Chappelka, A., 2010. Model estimates of ecosystem net Primary productivity, evapotranspiration, and water use efficiency in the southern United States during 1895–2007. For. Ecol. Manage. 259, 1311–1327, http://dx.doi.org/10.1016/j.foreco.2009.10.009.
Verkaik, I., Rieradevall, M., Cooper, S.D., Melack, J.M., Dudley, T.L., Prat, N., 2013. Fire as a disturbance in Mediterranean climate streams. Hydrobiologia 719 (1), 353–382.
Walling, D.E., Fang, D., 2003. Recent trends in the suspended sediment loads of the world's rivers. Global Planet. Change 39 (1), 111–126.
Wei, Y., Liu, S., Huntzinger, D.N., Michalak, A.M., Viovy, N., Post, W.M., Schwalm, C.R., et al., 2013. The North American carbon program multi-scale synthesis and terrestrial model intercomparison project—Part 2: Environmental driver data. Geosci. Model Dev. Discuss. 6, 5375–5422, http://dx.doi.org/10.5194/gmdd-6-5375-2013.
Willmott, C.J., Robeson, S.M., 1995. Climatologically aided interpolation (CAI) of terrestrial air temperature. Int. J. Climatol. 15 (2), 221–229.
Willmott, C.J., Rowe, C.M., Mintz, Y., 1985. Climatology of the terrestrial seasonal water cycle. J. Climatol. 5, 589–606.
Xia, Y., et al., 2012. Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparis on and application of model products. J. Geophys. Res. 117, D03109, http://dx.doi.org/10.1029/2011JD016048.
Xie, P., Arkin, P.A., 1996. Global precipitation: a 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs. Bull. Am. Meteor. Soc. 78, 2539–2558.
Zeng, N., Neelin, J.D., Chou, C., 2000. The first quasi-equilibrium tropical circulation model-implementation and simulation. J. Atmos. Sci. 57, 1767–1796.