Multicriteria evaluation of discharge simulation in Dynamic Global Vegetation Models

Hui Yang, Shilong Piao, Zhenzhong Zeng, Philippe Ciais, Yi Yin, Pierre Friedlingstein, Stephen Sitch, Anders Ahlström, Matthieu Guimberteau, Chris Huntingford, et al.

AGU PUBLICATIONS

JGR

Journal of Geophysical Research: Atmospheres

RESEARCH ARTICLE
10.1002/2015JD023129

Special Section:
Fast Physics in Climate Models: Parameterization, Evaluation and Observation

Multicriteria evaluation of discharge simulation in Dynamic Global Vegetation Models

Hui Yang[1], Shilong Piao[1,2], Zhenzhong Zeng[1], Philippe Ciais[1,3], Yi Yin[3], Pierre Friedlingstein[4], Stephen Sitch[4], Anders Ahlström[5,6], Matthieu Guimberteau[3], Chris Huntingford[7], Sam Levis[8,9], Peter E. Levy[10], Mengtian Huang[1], Yue Li[1], Xiran Li[1], Mark R Lomas[11], Philippe Peylin[3], Ben Poulter[12], Nicolas Viovy[3], Soenke Zaehle[13], Ning Zeng[14], Fang Zhao[14], and Lei Wang[2]

[1]Sino-French Institute for Earth System Science, College of Urban and Environmental Sciences, Peking University, Beijing, China, [2]Institute of Tibetan Plateau Research, Chinese Academy of Sciences, Beijing, China, [3]Laboratoire des Sciences du Climat et de l'Environnement, CEA CNRS UVSQ, Gif-sur-Yvette, France, [4]College of Engineering, Computing and Mathematics, University of Exeter, Exeter, UK, [5]Department of Earth System Science, School of Earth, Energy and Environmental Sciences, Stanford University, Stanford, CA, USA, [6]Department of Physical Geography and Ecosystem Science, Lund University, Lund, Sweden, [7]Centre for Ecology and Hydrology, Wallingford, UK, [8]National Center for Atmospheric Research, Boulder, Colorado, USA, [9]Now at The Climate Corporation, San Francisco, California, USA, [10]Centre for Ecology and Hydrology, Bush Estate, Midlothian, UK, [11]Department of Animal & Plant Sciences, University of Sheffield, Sheffield, UK, [12]Department of Ecology, Montana State University, Bozeman, Montana, USA, [13]Max Planck Institute for Biogeochemistry, Jena, Germany, [14]Department of Atmospheric and Oceanic Science, University of Maryland, College Park, Maryland, USA

Key Points:
- The seasonal cycle of river discharge is well reproduced in low and middle latitudes
- DGVMs generally underestimate annual mean discharge
- Most DGVMs correctly reproduce observed interannual variability of discharge

Supporting Information:
- Figure S1 and Table S1

Correspondence to:
S. Piao,
slpiao@pku.edu.cn

Citation:
Yang, H., et al. (2015), Multicriteria evaluation of discharge simulation in Dynamic Global Vegetation Models, *J. Geophys. Res. Atmos.*, *120*, 7488–7505, doi:10.1002/2015JD023129.

Received 18 JAN 2015
Accepted 28 JUN 2015
Accepted article online 30 JUN 2015
Published online 8 AUG 2015

Abstract In this study, we assessed the performance of discharge simulations by coupling the runoff from seven Dynamic Global Vegetation Models (DGVMs; LPJ, ORCHIDEE, Sheffield-DGVM, TRIFFID, LPJ-GUESS, CLM4CN, and OCN) to one river routing model for 16 large river basins. The results show that the seasonal cycle of river discharge is generally modeled well in the low and middle latitudes but not in the high latitudes, where the peak discharge (due to snow and ice melting) is underestimated. For the annual mean discharge, the DGVMs chained with the routing model show an underestimation. Furthermore, the 30 year trend of discharge is also underestimated. For the interannual variability of discharge, a skill score based on overlapping of probability density functions (PDFs) suggests that most models correctly reproduce the observed variability (correlation coefficient higher than 0.5; i.e., models account for 50% of observed interannual variability) except for the Lena, Yenisei, Yukon, and the Congo river basins. In addition, we compared the simulated runoff from different simulations where models were forced with either fixed or varying land use. This suggests that both seasonal and annual mean runoff has been little affected by land use change but that the trend itself of runoff is sensitive to land use change. None of the models when considered individually show significantly better performances than any other and in all basins. This suggests that based on current modeling capability, a regional-weighted average of multimodel ensemble projections might be appropriate to reduce the bias in future projection of global river discharge.

1. Introduction

Climate change and human management of land ecosystems imprint the regional and global hydrological cycle and are expected to continue to do so in the coming decades. Natural and anthropogenically induced changes in river discharge, and especially any related extreme hydrological events of major droughts or flooding, are of concern given their damaging effects on human societies [*Kim*, 2005; *Milly et al.*, 2005; *Oki and Kanae*, 2006; *Gerten et al.*, 2008; *Stocker et al.*, 2013]. However, the scarcity and heterogeneity of river discharge measurements make it difficult to evaluate the ability of models to reproduce present-day conditions and identify biases that affect required future projections [*Oki et al.*, 2001; *Alkama et al.*, 2011]. Several statistical methods, e.g., wavelet time-series decomposition or cumulative discharge analysis, have been used to reconstruct time series of discharge [*Labat et al.*, 2004; *Milliman et al.*, 2008]. However, the reliability of these methods is questionable given the limited global data available [*Legates et al.*, 2005; *Alkama et al.*, 2011].

Dynamic Global Vegetation Models (DGVMs) served as the land surface component of many of the general circulation models (GCMs) that were used in Intergovernmental Panel on Climate Change (IPCC) Fifth Assessment Report (AR5). DGVMs have been developed to explicitly model the vegetation responses to climate change, rising atmospheric CO_2 (and N deposition in some models) as well as to land use change

Table 1. The Description of the Seven Terrestrial Ecosystem Models

Model name	Abbreviation	Spatial Resolution	Period	Snowmelt	Reference
Community Land Model 4	CLM4CN	2.5° × 1.875°	1901–2010	Y	*Oleson et al.* [2010]; *Lawrence et al.* [2011]
Lund-Potsdam-Jena	LPJ	0.5° × 0.5°	1901–2010	N	*Sitch et al.* [2003]
LPJ-GUESS	LPJ-GUESS	0.5° × 0.5°	1901–2010	N	*Smith et al.* [2001]
ORCHIDEE	ORCHIDEE	0.5° × 0.5°	1981–2010	Y	*Krinner et al.* [2005]
ORCHIDEE-CN	OCN	3.75° × 2.5°	1901–2010	Y	*Zaehle et al.* [2010]
Sheffield-DGVM	SDGVM	3.75° × 2.5°	1901–2010	N	*Woodward et al.* [1995]; *Woodward and Lomas* [2004]
TRIFFID	TRIFFID	3.75° × 2.5°	1901–2010	Y	*Cox* [2001]; *Best et al.* [2011]; *Clark et al.* [2011]

[*Murray et al.*, 2013]. In addition, physical land hydrology processes is built to generate global and regional discharge estimates. In general terms of comparing climate models, such DGVMs land surface components have been generally regarded as equally good in performance in reproducing changes in runoff, yet this fundamental assumption ("model democracy") has never been tested and evaluated [*Knutti*, 2010; *Flato et al.*, 2013]. A benefit of DGVMs is that they can also be used "offline" to quantitatively evaluate the separate contributions of different environmental factors, such as climate, CO_2, and land use, to changes in regional and global discharge. For instance, with the land surface scheme in Hadley Center climate models, *Gedney et al.* [2006] using the HadCM3 model could identify an effect of elevated CO_2 induced stomatal closure (operating in parallel to other forcings), causing global discharge over the period 1901–1994 to be higher than if that effect was not present. Despite this, *Piao et al.* [2007] using the ORCHIDEE model demonstrated that observed rainfall increase and land use change (forest area decrease) explained most of global positive discharge trends and that higher LAI offsets decreased stomatal conductance per leaf area. Using the LPJmL model, *Gerten et al.* [2008] confirmed that the increasing trend of global discharge was the result of multiple factors including precipitation changes, global warming, land use change, rising atmospheric CO_2, and irrigation water withdrawals. As such, the source of differences in model diagnostics, and thus implications of individual magnitudes of effects, does still contain uncertainties. For this reason, we have performed a systematic evaluation of model runoff simulation across available DGVMs to understand current and projected changes in global and regional discharge.

In this study we used an ensemble of DGVMs from the TRENDY model intercomparison project [*Sitch et al.*, 2013] in which all models were forced by the same historical climate forcing and atmospheric CO_2 data. This ensemble is used to investigate model performance in simulating basin level discharge. Although comparative evaluations have already been used to show model skill in reproducing discharge at a global scale [*Cramer et al.*, 2001; *Blyth et al.*, 2011; *Ringeval et al.*, 2012], this does not capture the various environmental factors controlling directly or indirectly discharge in specific different river basins [*Alkama et al.*, 2011]. For this reason, our model performance evaluation based on single basins or regional average rather than on global numbers.

We evaluate discharge outputs from seven DGVMs [*Sitch et al.*, 2013] (Table 1) against river discharge observations in 16 large river basins (Table 2) to compare regional differences in the performance of individual models. Our objectives include (1) to quantify how well DGVMs represent variability and trends in river discharge at regional scales and assess whether there are consistently better models; (2) to investigate the sources of bias between simulated and observed discharge, including uncertainties associated with precipitation forcing, soil hydrology parameterizations of DGVMs controlling runoff, by coupling a routing model to DGVM runoff to calculate discharge; and (3) to discuss whether considering more processes, e.g., land use change, helps to narrow any noted biases. Multiple model evaluation metrics and a skill score based on overlapping of Epanechnikov kernel-based probability density functions (PDFs) is applied to quantify the agreement between model simulation and observations. Model comparison with observed discharge is decomposed into different time scales: the seasonal cycle, the annual mean discharge and its trend over the last three decades, and the interannual variability (see the next section).

2. Material and Methods

2.1. Model Description

The seven DGVMs are LPJ [*Sitch et al.*, 2003], ORCHIDEE [*Krinner et al.*, 2005], Sheffield-DGVM [*Woodward et al.*, 1995; *Woodward and Lomas*, 2004], TRIFFID [*Cox*, 2001; *Best et al.*, 2011; *Clark et al.*, 2011], LPJ-GUESS

Table 2. The Description of the River Basins and Gauging Stations in This Study

Grouping	River Basin	Station Name (GRDC)	Catchment Area (km^2)	Studied Period
Low latitude	Amazon	OBIDOS - LINIGRAFO; FORTALEZA; ALTAMIRA; ARAPARI; SAO FRANCISCO	5,557,618	1981–1997, 1999–2002
	Tocantins	TUCURUI	742,300	1981–1983, 1985–1995, 1998–2009
	Congo	KINSHASA	3,475,000	1981–2010
	Fitzroy	THE GAP	135,757	1981–2006
Mid latitude	Yellow	HUAYUANKOU	730,036	1981–1987, 2008–2010
	Nelson	LONG SPRUCE GENERATING	1,060,000	1987–2008
	Danube	CEATAL IZMAIL	807,000	1981–1995, 1997–2008
	Mississippi	VICKSBURG, MS	3,088,653	1981–1998, 2008
	Rhine	LOBITH; MEGEN DORP	189,780	1997–2007
	Yangtze	DATONG	1,705,383	1981–1985, 1987, 2002–2004, 2006–2010
	Amur	KOMSOMOLSK	1,730,000	1981–1983, 1985–1990, 2000–2004
High latitude	Lena	KYUSYUR (KUSUR)	2,430,000	1981–2003
	Mackenzie	ARCTIC RED RIVER; ABOVE FORT MCPHERSON	1,730,600	1981–1986, 1988–1996, 1999–2009
	Ob	SALEKHARD	2,949,998	1981–1999, 2001–2003
	Yenisei	IGARKA	2,440,000	1981–2003
	Yukon	PILOT STATION, AK	831,390	1981–1995

[*Smith et al.*, 2001], NCAR_CLM4CN [*Oleson et al.*, 2010; *Lawrence et al.*, 2011], and OCN [*Zaehle et al.*, 2010] (Table 1). The models runs were performed with the protocol of TRENDY (http://dgvm.ceh.ac.uk/, accessed 11 July 2013) project, which defines a set of factorial DGVM simulations (see below) over the period 1901–2010. Compared with detailed hydrological models, DGVMs generally simulate the water balance of grid cells without river routing and human withdrawal of water—and where such a balance can subsequently be used to drive routing descriptions. DGVMs include important vegetation-hydrological processes such as the response of stomatal conductance to climate and CO_2, transpiration controlled by energy and soil moisture availability, and in some instances, nutrient mineralization changes affecting transpiration or photosynthesis. These effects are explicitly represented in the TRENDY process-oriented DGVMs [*Sitch et al.*, 2013].

All models use the same meteorological forcing files, in which historical climate data are from the CRUNCEP v4 data set (http://dods.extra.cea.fr/data/p529viov/cruncep/). This is a merging of the CRU data set monthly anomalies at 0.5° resolution and the 6-hourly NCEP/NCAR reanalysis data at 2.5° resolution [*Kalnay et al.*, 1996]. Global atmospheric CO_2 concentration data originate from ice core and the National Oceanic and Atmospheric Administration (NOAA) measurements [*Keeling and Whorf*, 2005; *Keeling et al.*, 2009], and the land use condition is from the Hyde database [*Klein Goldewijk et al.*, 2011]. In TRENDY simulations named S1, models are forced with elevated atmospheric CO_2, while other factors are held constant. In simulations S2, models are forced with elevated atmospheric CO_2 concentrations and varying climate. In simulations S3, atmospheric CO_2 concentrations, climate, and land use are all varied. Hence, the effect of land use change can be distinguished by examining the differences between simulations S3 and S2. All the diagnostics and statistics are computed after regridding all model outputs into a $0.5° \times 0.5°$ grid, regardless of their original spatial resolution (Table 1).

2.2. Gauging Station Data Set

The criteria used to choose river basins for this study are (1) basin area larger than 100,000 km^2 to ensure selected river basins are larger than the area of one grid box at the original resolution of models [*Arora and Boer*, 2001] and (2) river basins are in different continents to ensure global representativeness. In total 16 river basins fulfilled this set of two criteria (Figure 1).

Daily river discharge data were obtained from the Global River Discharge Center (GRDC; in Koblenz, Germany, http://www.bafg.de/GRDC/) measured at gauging stations. The gauging stations located in river mouths and main tributary mouths were selected, for their close approximation of total catchment discharge. We chose 22 gauging stations for the selected 16 rivers, including their branches (Table 2 and Figure 1). More than one station was selected for the Amazon, Mackenzie, and Rhine basins because river discharge data from the gauging station in the main channel did not include discharge of some tributaries. Hence, we could also

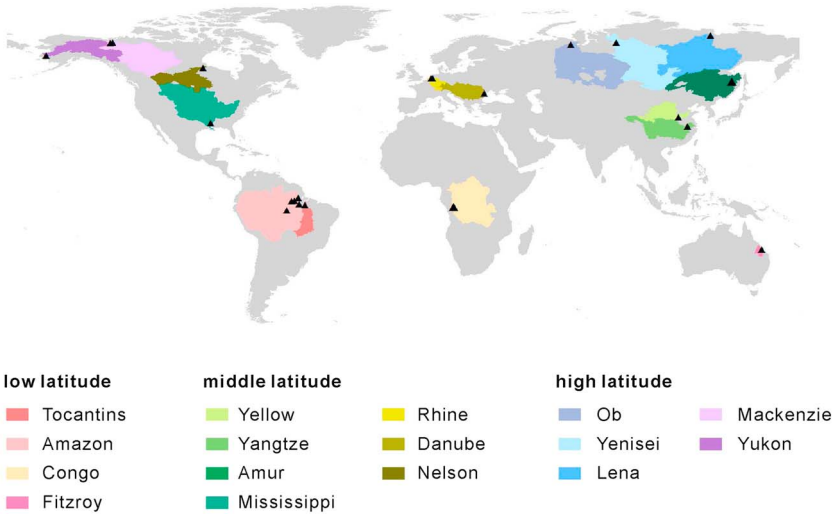

Figure 1. The location of the 16 river basins and the 22 gauging stations for discharge measurements (black triangles).

include discharge data from such tributaries using gauging stations in them as well. The analysis is focused on the period 1981–2010. Only months with over 50% of daily discharge data were included, monthly discharge being calculated from averaged daily data.

2.3. River Routing Model

The surface runoff simulated by DGVMs on each grid cell was fed into the river routing model [*Miller et al.*, 1994], which was operated with a 6 h time step at 0.5° resolution. The global Dominant River Tracing (DRT) [*Wu et al.*, 2011, 2012], a global river network data set, including flow direction, flow distance, and river channel slope on a 0.5° by 0.5° grid for macroscale hydrological modeling, was used. The transport equations of the routing model are

$$\frac{dS}{dt} = R + \sum F_{in} - F_{out} \tag{1}$$

where S represents the water storage, R the surface runoff, and F_{in} and F_{out} the inflow (for summation, and from upstream grid boxes) and the outflow (to the next grid) of the river discharge, respectively. The water flow F is calculated as

$$F = S \times \frac{u}{d} \tag{2}$$

where d is the distance between grid cells and u is the effective flow velocity of water from a grid to its downstream grids according to the network map. The value of u is related to actual flow direction and river channel slope by

$$u = 0.35\sqrt{i/i_0} \tag{3}$$

where i is the channel slope within the grid box and i_0 is the reference topography gradient ($i_0 = 0.00005$). It is noted that u is limited to the range of 0.15 to 5 m/s [*Miller et al.*, 1994].

2.4. Evaluation Metrics

We use four statistical indexes to evaluate seasonal discharge, which individually place an emphasis on mean bias, overall fitting, error index, and agreement of variability. We then develop and use a skill score metrics that combines the four indices to define an integrated overall model score.

Firstly we use the percentage bias (PBIAS, equation (4)) to measure the mean bias of modeled monthly runoff.

$$PBIAS = \frac{\sum_{i=1}^{n}(Q_i^{obs} - Q_i^{sim})}{\sum_{i=1}^{n} Q_i^{obs}} \tag{4}$$

where Q_i^{obs} is the i^{th} monthly mean observation, Q_i^{sim} is the i^{th} monthly mean simulated value, and so n is the total number of observations (n = 12, i.e., all months). The best achievable value of PBIAS is zero, and the lower the value is, the less biased is a given model [*Moriasi et al*., 2007].

RMSE is a commonly used error index [0.0, +∞] (equation (5)), a perfect model taking a RMSE of zero. To make RMSE independent of PBIAS, for this statistic our observed and simulated data have been centered for analyzing. Sometimes a high RMSE value may be caused simply by low standard deviation in observations and does not necessarily indicates good model performance [*Singh et al*., 2005]. Thus, we also use the ratio of RMSE to the observations interannual standard deviation (hereafter RSR, equation (6)). RMSE is a good evaluation index over a single basin, whereas RSR is better when time series from different basins are grouped together as a single statistic. In fact, RSR is similar to the Nash-Sutcliffe Efficiency criteria (NSE, see equation (7)), which is widely used for hydrological models [*Nash and Sutcliffe*, 1970].

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Q_i^{obs} - Q_i^{sim}\right)^2} \tag{5}$$

where $\overline{Q^{obs}}$ is the mean of observed data.

$$RSR = \frac{\sqrt{\sum_{i=1}^{n}\left(Q_i^{obs} - Q_i^{sim}\right)^2}}{\sqrt{\sum_{i=1}^{n}\left(Q_i^{obs} - \overline{Q^{obs}}\right)^2}} \tag{6}$$

$$NSE = 1 - \left[\frac{\sum_{i=1}^{n}\left(Q_i^{obs} - Q_i^{sim}\right)^2}{\sum_{i=1}^{n}\left(Q_i^{obs} - \overline{Q^{obs}}\right)^2}\right] \tag{7}$$

The correlation coefficient, r, between either monthly or annual simulations and respective observations is a good index to estimate the agreement for the seasonal and interannual variability (r, equation (8)). Here r = 1 implies that the phase of the variability of the model perfectly matches the observed one [*Legates and McCabe*, 1999].

$$r = \frac{\sum_{i=1}^{n}\left(Y_i^{sim} - \overline{Y^{sim}}\right)\left(Y_i^{obs} - \overline{Y^{obs}}\right)}{\sqrt{\sum_{i=1}^{n}\left(Y_i^{sim} - \overline{Y^{sim}}\right)^2}\sqrt{\sum_{i=1}^{n}\left(Y_i^{obs} - \overline{Y^{obs}}\right)^2}} \tag{8}$$

The above metrics can be displayed in a Taylor diagram [*Taylor*, 2001] in order to visually compare the model performances.

The above three evaluation statistics (PBIAS, RSR, and r) are combined into one single score to quantify overall model performance. PBIAS < -1 or PBIAS > 1 is considered as unacceptable model performance, and so we defined a normalized PBIAS score as one minus the absolute value of PBIAS with a score of 0 if PBIAS < -1 or PBIAS > 1. For RSR, *Moriasi et al*. [2007] concluded that model simulation can be considered as satisfactory if RSR < 0.7. Therefore, to define an RSR_{score} between 0 and 1, we used a linear transform for RSR [*Moriasi et al*., 2007; *Chen et al*., 2012] (a score of 1 corresponds to RSR = 0 and a score of 0.8 to the acceptable threshold RSR = 0.7; transform as equation (9)). For correlation coefficient, the score is the value of r, with negative r values being set to 0. The average of the above three scores is considered to be the overall skill score which is used to qualify model performance of simulated seasonal cycle.

$$RSR_{score} = \begin{cases} 1 - 2/7 \times RSR, \text{for } RSR \le 7/2 \\ 0, \text{for } RSR > 7/2 \end{cases} \tag{9}$$

Finally we used one further skill score to evaluate annual discharge, by comparing the Epanechnikov kernel-based probability density functions (PDFs) of models with reference values (i.e., observed discharge records).

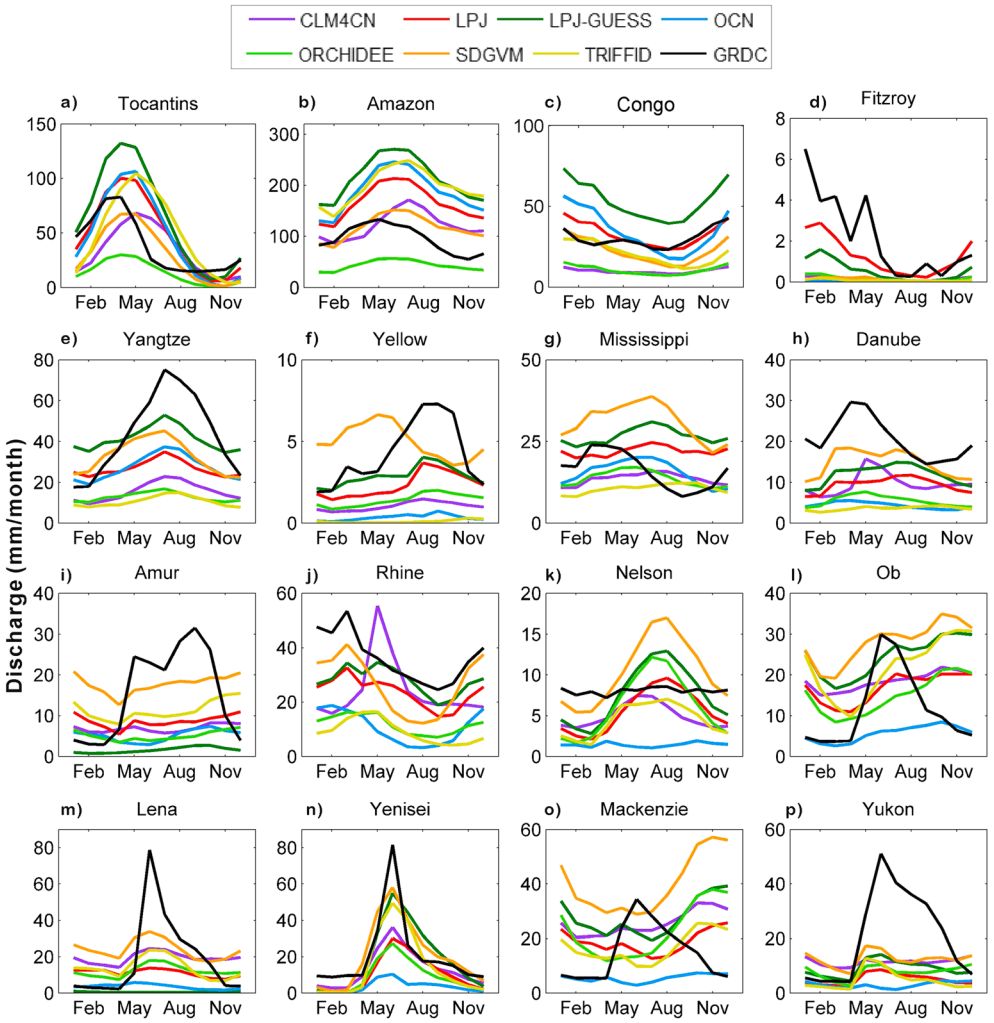

Figure 2. Observed (black) and simulated seasonal cycle of river discharge for 16 river basins. (a–d) The river basins in the low latitudes, (e–k) the river basins in the midlatitudes, and (l–p) the river basins in the high latitudes.

This method is powerful, as it can show the discrepancy between simulated and observed data in both mean value and interannual variability by calculating the common area under the two PDFs [*Maxino et al.*, 2008; *Anav et al.*, 2013]. For a perfect model simulation, the skill score would equal 1. The skill score (S_{score}) is given by equation (10).

$$S_{score} = \sum_{1}^{n} \text{minimum}(Z_{sim}, Z_{obs}) \tag{10}$$

where n is the number of bins used to calculate the PDF (n = 600 in this study), and Z_{sim} and Z_{obs} are the frequency of values (i.e., PDF value) in a given bin from the model and from the observed data, respectively. The larger the imperfections of a model, the closer S_{score} becomes to zero.

3. Results

3.1. Seasonal Cycle

The comparison of models against observations of mean annual cycle of river discharge curves over the 16 selected river basins is shown in Figure 2. In the low latitudes (Figures 2a–2d), most models produce a positive bias in rivers Tocantins, Amazon, and Congo (except for ORCHIDEE) but a slightly negative bias in the Fitzroy River (except for the LPJ model). In the midlatitudes (Figures 2e–2k) most models underestimate discharge for all months. It is particularly noticeable that all models have a poor

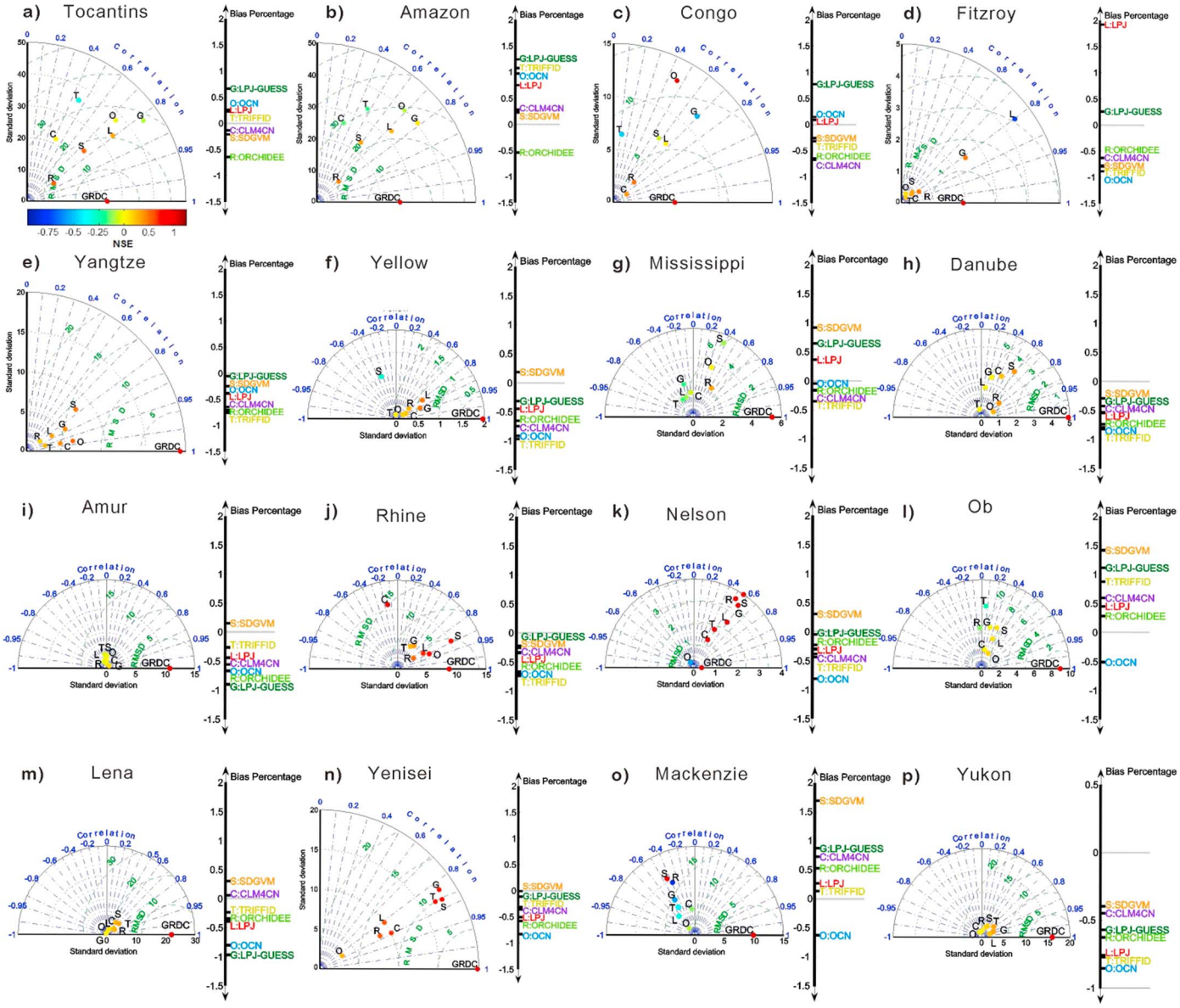

Figure 3. Taylor diagrams for the 1981–2010 seasonal discharge, computed over 16 river basins (see Figure 2). The *y* axis on the right panel shows the mean bias percentage of each model, and on the left panel reference data are plotted along the abscissa as a baseline. The radial distance indicates the standard deviation. The azimuthal angle is proportional to the correlation coefficient. The distance from the point of reference data is related to the centered RMSE. And the color of point indicates the centered NSE.

performance in the Nelson River (which flows mainly in Canada). In the high latitudes (Figures 2l–2p), observed discharge increases rapidly in April with snow and glacier melt and then decreases slowly in June or July. Most models could not reproduce this spring peak, especially for the Ob and Mackenzie Rivers.

The centered RMSE, correlation coefficient, and standard deviation of each model for different river basins are shown in a Taylor diagram, and also marked is mean percentage bias (Figure 3). The results show a complex and large variability of models' performances among different basins. In Tocantins, Amazon, and Congo Rivers, SDGVM has the best or second best performance among all models, with a small mean bias and the best match of the seasonal amplitude (referring to the ratio of standard deviation of simulations and observations being close to 1), the lowest RMSE and a high correlation coefficient (nearly 0.8). In contrast, TRIFFID model has a poorer performance. For the Nelson and Mackenzie Rivers, all the models show a poor skill in reproducing the seasonality of river discharge. In the Amur, Yukon, Yellow, and Fitzroy Rivers,

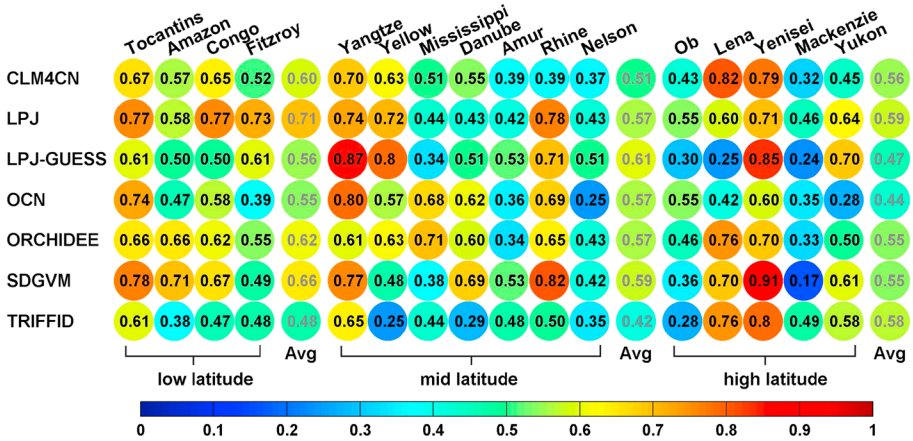

Figure 4. The skill scores (see color bar) of seven models in simulating river discharge seasonal cycle over 16 river basins. Seasonal skill score is the average of four evaluation metrics: mean bias percentage, the centered RSR, the centered NSE, and correlation coefficient. A score close to 1 indicates a good performance of models in reproducing discharge seasonal variation.

the different models show a similar degree of performance simulating seasonal discharge. SDGVM and CLM4CN models perform better at high latitudes. Better performance is likely with the CLM4CN model as it includes a snowmelt module, and that of SDGVM may have been calibrated [*Mao et al.*, 2007; *Murray et al.*, 2013].

Figure 4 presents the integrated skill score. In general, all the models cluster around 0.5 and ORCHIDEE has the best performance (score $>$ 0.5 over 12 basins). However, different models perform differently in different latitudes, and no single model is universally best across all river basins. The seasonal variation of discharge is well captured in the low latitudes (average score of 0.6 between the models; ranging from 0.48 to 0.71), but not in the high latitudes. In the low latitudes, the best scores are obtained for LPJ (average score of 0.69) and the worst one for TRIFFID (score of 0.48). In the middle and high latitude basins, all the models poorly simulate the seasonal variation of discharge in the Amur, Nelson, Ob, and Mackenzie Rivers. This is consistent with the low correlation coefficient values in Figures 3i–3k and 3n. LPJ-GUESS performs better in the mid-latitudes (average score of 0.61), while TRIFFID has a better performance in the high latitude basins (average score of 0.58). TRIFFID (score of 0.42) and OCN (0.44) show the poorest scores in the mid and high latitudes, respectively. Overall, a skill score greater than 0.8 indicates very good model performance [*Moriasi et al.*, 2007; *Gupta et al.*, 2009; *Chen et al.*, 2012]. LPJ-GUESS and SDGVM scores are over 0.8 for two basins.

In the Nelson River, observations show almost no seasonal cycle, whereas models do exhibit a pronounced seasonality. One possible reason is that the Nelson is large river and one that has experienced considerable anthropogenic disturbances such as dams, diversions, and/or reservoirs linked to power production, particularly during the wet season [*Déry et al.*, 2011]. A large amount of water may have been withdrawn in certain periods, and which dampens variability and in a way that models may not be currently taking into account. In the high latitude region, most models underestimate the peak discharge and spring discharge (Figure 2). A previous study demonstrated, as expected, that the modeling of snow amount and timing of snowmelt is crucially important to determine the timing and magnitude of raised discharge due to snowmelt [*Liston*, 1999]. Three out of seven DGVMs generally use a simplified water balance and discharge generation without snow accumulation and melt scheme (Table 1). CLM4CN, ORCHIDEE, TRIFFID and OCN, which have snowmelt scheme, show no significant improvement. It is likely due to the unknowns of hydrological processes in snow-dominated regions.

3.2. Annual Mean Discharge and Its Trend

We compare in this section simulated and observed annual mean discharge, and their trends, for the 16 basins in the period 1981–2010 (Figure 5). Generally, most models underestimate annual mean discharge, with a few exceptions. In the high and middle latitude basins, SDGVM has a more realistic estimation on annual mean discharge; while OCN, ORCHIDEE, and TRIFFID perform less well, especially for the Ob, Amur,

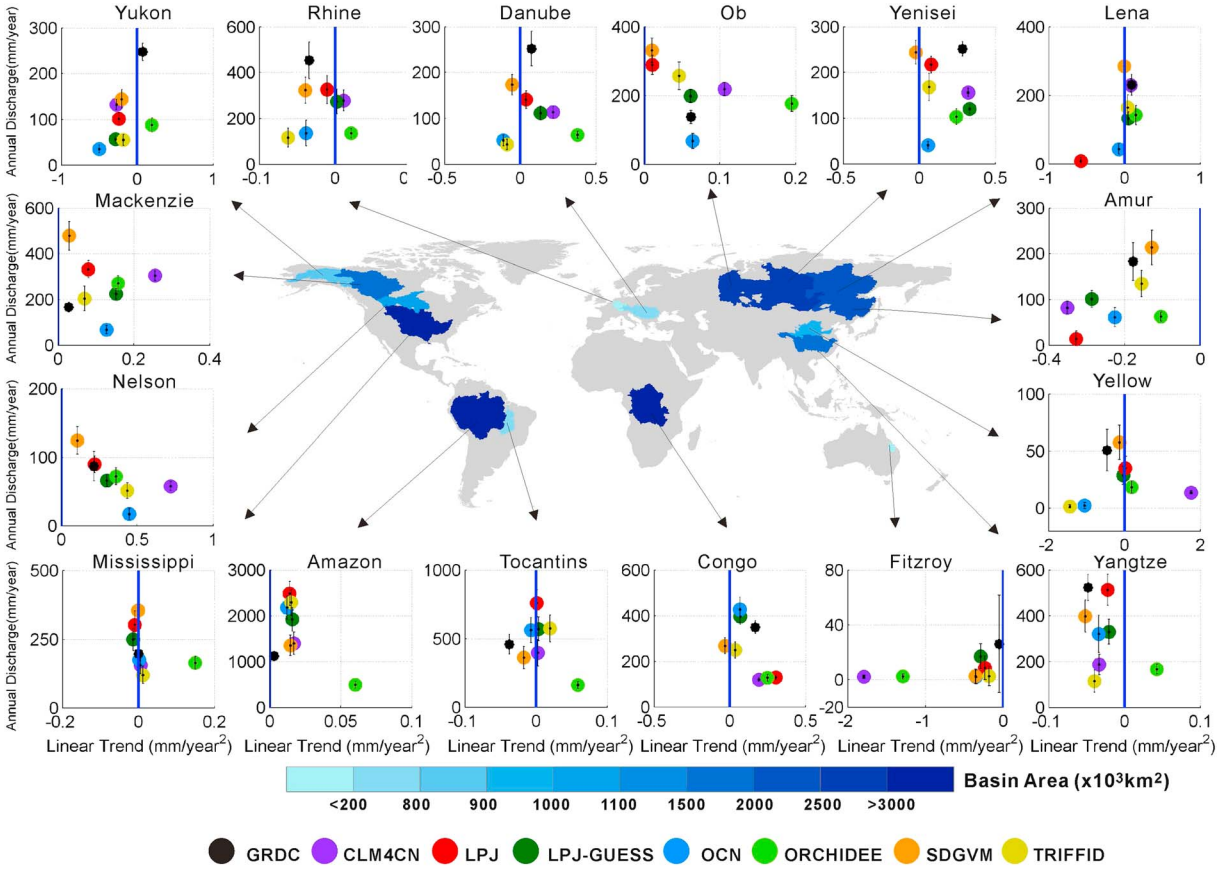

Figure 5. In each subplot, *y* axis indicates annual mean discharge and the error bar indicates its standard deviation (interannual variation of annual discharge), *x* axis indicates the long-term trend of discharge estimated by linear regression. Different dots representing observed (black) and model simulated data (color). On the map, the studied basins are also shown with the color representing their catchment areas.

Yellow, and Rhine Rivers. In the low latitudes, two models out of seven have a PBIAS larger than 50%. However, much of these biases of annual discharge may be caused by the effect of land use change, vegetation cover change, and human interventions, which were not considered in the S2 simulation of the TRENDY protocol, yet was shown to affect discharge trends [*Piao et al.*, 2007; *Gerten et al.*, 2008]. Furthermore, the overall underestimate of annual discharge by the seven DGVMs forced by the same climate data may also reflect to systematic uncertainties in the forcing climate (mainly precipitation) data [*Fekete et al.*, 2004]. This highlights the need to consider the implications for estimates of discharge by DGVMs when driven by different forcing data. Finally, the availability and reliability of observed discharge data vary from one river basin to another, and the differences between observed and simulated discharge of different basins may be related to the configuration of gauging stations and uncertainties in the rating curves [*Gudmundsson et al.*, 2012].

To characterize the discharge trends over period 1981–2010, we fitted a linear trend to both modeled and observed annual data (Figure 5). The observed trend of discharge is generally positive in most of the high and middle latitude river basins except for Amur, Yangtze, and Yellow River basins, while it is negative in the low latitudes. The sign of the trends in discharge can be reproduced by most models, except for the Yukon River. However, the magnitude of trends is generally underpredicted. The positive trend in high-latitude rivers discharge was explained by an increasing snowmelt and precipitation over the high latitudes in the past decades [*Barnett et al.*, 2005; *Solomon et al.*, 2007; *Stocker et al.*, 2013]. Results show that any (possibly greenhouse gas-induced) changes in surface meteorological conditions, and therefore discharge, is not fully captured by the DGVMs studied. In temperate arid regions, snowmelt and precipitation increase soil moisture first rather than being channeled into river discharge, so a decreasing trend in discharge can be found [*Bates et al.*, 2008]. Moreover, increasing human water uses through dams

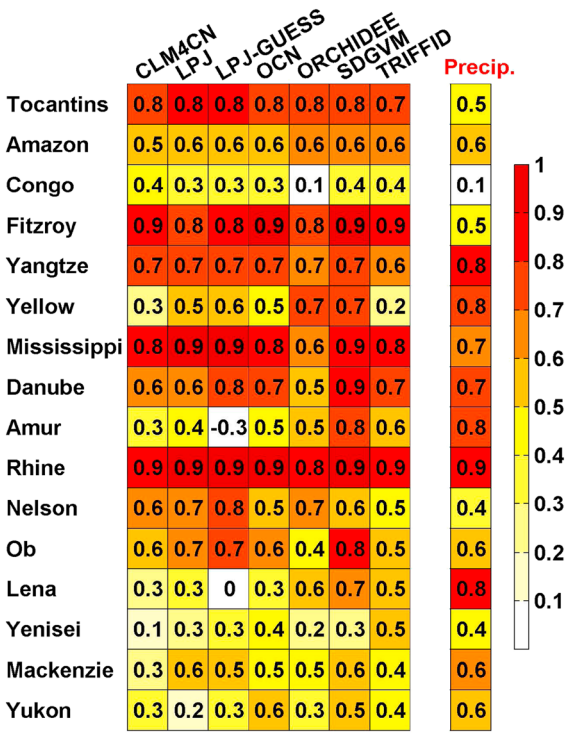

Figure 6. Matrix displaying correlation coefficients (r) (see color bar) of detrended interannual discharge variability between the simulated and observed data at 16 river basins for the seven DGVMs. The rightmost column represents the correlation coefficient between the detrended discharge observations and the detrended precipitation used to force the models. An r value close to 1 indicates a well model reproduced interannual variability.

and irrigation likely amplify the decreasing trend in Yangtze River, Yellow River, and Amur [*Dai and Trenberth*, 2002]. Those effects, however, have not been fully considered in these DGVMs.

3.3. Interannual Variability

To check how models simulate the interannual variation of discharge (Figure S1), we calculated the correlation coefficients between detrended discharge simulations and observations. Figure 6 shows that most models capture the interannual discharge variability ($r > 0$), but the correlation coefficient values remain lower in some high latitude basins (r ranges from 0 to 0.7 for Lena, Yenisei, and Yukon). Some models perform better than others, but no single model performs better than all other models when considering across all the different river basins. SDGVM has a good interannual correlation ($r \geq 0.5$) for 14 out of 16 basins. LPJ-GUESS simulates badly ($r \leq 0$) for the interannual variability of the Amur and Lena River discharge, whereas other models could capture it better (r values from 0.3 to 0.8).

The skill score matrix of the interannual variability of discharge (equation (10); Figure 7) indicates an overall low score and a large mean bias for the OCN model, and especially in the middle and high latitudes. However, SDGVM and LPJ scores are generally high which is related to their good performance in reproducing the interannual variability ($r > 0.5$). LPJ-GUESS has good performances in some midlatitude basins (e.g.,

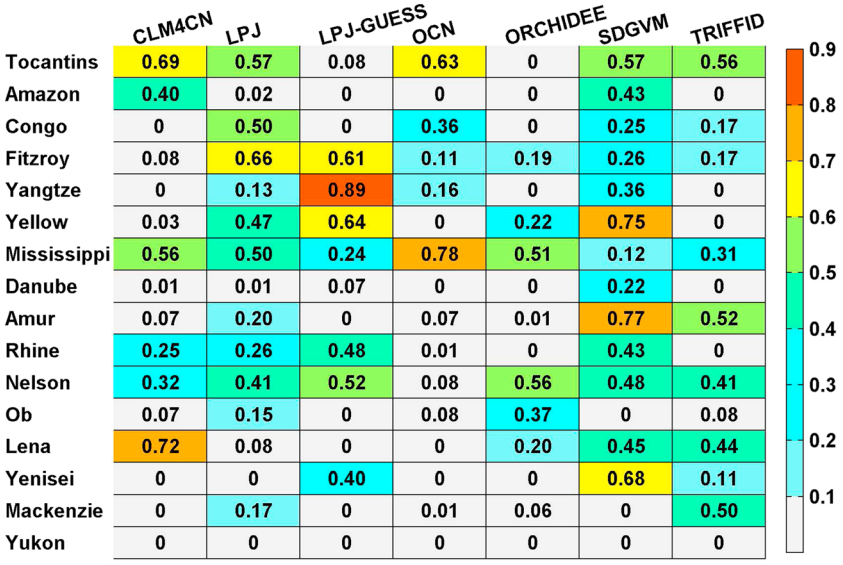

Figure 7. Interannual variability skill score matrix based on overlapping of PDFs for 16 river basins, respectively. A score of 0 indicates a poor model performance in reproducing the mean level and shape of discharge interannual variation, while a perfect performance score is equal to 1.

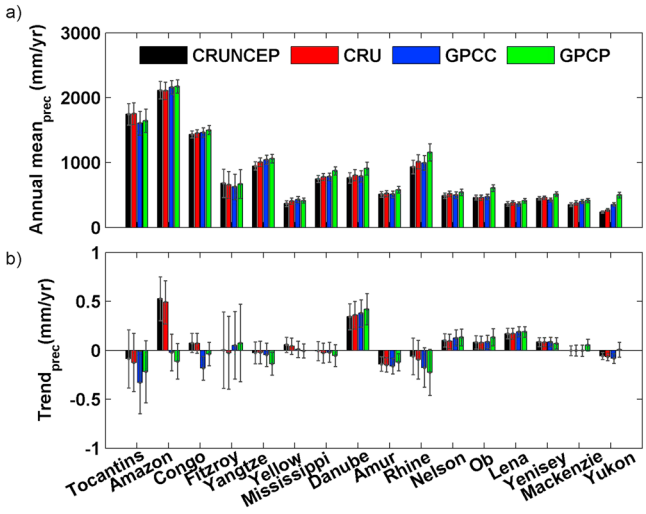

Figure 8. Comparison of (a) mean annual and (b) multiyear trend of precipitation from CRUNCEP, CRU, GPCC, and GPCP over the period 1981–2010. Note that the CRUNCEP data set was made from the CRU data set and the NCEP-NCAR Reanalysis data.

Yangtze, Yellow, Rhine, and Nelson Rivers) and achieves the highest score in the Yangtze River basin. In general, the performances of CLM4CN and TRIFFID lie in the middle of the ensemble. In the Yukon, Danube, Mackenzie, Ob, and Amazon basin, most models have a poor score (most scores are less than 0.1) due to their negative bias in the mean discharge simulation (Figure 3) and/or poor performance in reproducing the interannual variability (Figure 6).

Our definition of an overall skill score is not a perfect tool to characterize DGVM performance [see also *Anav et al*., 2013]. For example, a model that can simulate the tails of distribution curves well would be valuable for assessment of extreme events, even if it simulates the center area of the distribution curves badly and therefore achieve a low skill score. Overall, our results imply that both the model-based estimates interannual variability and the seasonal cycle of simulated discharge data in both the tropics and the high latitudes seem to be problematic.

4. Discussion

4.1. Uncertainty Associated With Precipitation Forcing

The uncertainty in precipitation forcing data strongly affects simulated discharge outputs through modeled runoff generation processes. *Biemans et al*. [2009] used seven data sets of precipitation as inputs to the single LPJmL model and found that the uncertainty in precipitation translates into at least the same uncertainty in river discharge simulations. This is by comparing our CRUNCEP in situ based observations with the GPCP precipitation products [*Adler et al*., 2003; *Schneider et al*., 2011]. GPCP is a combination of satellite-based precipitation estimates and GPCC gauge observations.

Firstly, we compared directly with each other the mean annual and multi-year trends of precipitation from different data sets for the period 1981–2010. Although mean annual precipitation from different data sets appears to be consistent between CRUNCEP and GPCP for most river basins, the multi-year trends show large differences (Figure 8). Most notably, the direction of multiyear trends shows an inconsistency over the wet tropics and dry regions. For the Amazon and Congo basins, precipitation from CRUNCEP (based on CRU data set) shows an upward trend, while GPCP (based on GPCC data set) actually shows an opposite decreasing trend. For the Fitzroy Basin and some northern basins (e.g., the Mackenzie and Yukon basins), the downward trend of precipitation from CRUNCEP is opposite to the increasing trend of GPCC and GPCP. In the tropics, CRUNCEP may be of poor quality because stations are sparsely distributed [*Fekete et al*., 2004]. In dry regions and several northern basins, inconsistency in the precipitation data is more likely to occur due to changes in precipitation that too small to detect and to high spatial variability, making extrapolation of station data problematic [*Fekete et al*., 2004; *Greve et al*., 2014]. Therefore, uncertainty in precipitation trends may translate into larger errors in runoff and discharge trends, particularly in the tropics and some dry regions, whereas the mean value of precipitation is close between different data sets.

Secondly, to check whether the precipitation is related to the interannual variation of simulated discharge, we calculated the correlation coefficient (r) between the detrended discharge observations and the detrended precipitation forcing (Figure 6). The results show that the interannual variability of precipitation forcing is consistent with variability in the discharge observations for 15 of 16 basins (the exception being Congo Basin). The precipitation variability explains a large fraction of the variance of river discharge variability in the mid latitudes. Again, for the Congo Basin, the quality of precipitation data

might be poor because of few rain gauges in that region [*Fekete et al*., 2004]. Precipitation may also increase soil moisture first, before producing runoff, in arid regions and more generally in places where soils are far from saturation. In some catchments such as Nelson, Ob, and Yenisei, the inconsistency between simulated interannual discharge and observations (average $r = 0.47$) may result from non-modeled river management (damming) and from the presence of wetlands and floodplains, acting as a buffer in the transit of water from each grid cell to the river mouth as large evaporative area. Of particular note is that for the Yellow River, Amur, and Lena catchments, the interannual variability of precipitation forcing and discharge observations are consistent with each other (average $r = 0.8$), but the simulated and observed discharges are not. This implies that at present, a purely nonphysical statistical model actually performs better than a process-based DGVM. Simulated water residence time in soils in DGVMs, or the water transit time in the routing model, worsen the phase of the discharge. In sum, precipitation forcing biases and limitations in the structure of DGVM (e.g., the fact that runoff does not depends on terrain slope in most models) and of the routing model might lead to failure in reproducing the interannual variability of discharge.

4.2. Uncertainty Associated With the River Routing Model

The observed discharge records are from the gauging stations located at the river mouths, and a river routing model was applied to convert gridded runoff within a catchment to discharge. River routing models are critical when focusing on the seasonality of discharge, especially for large rivers [*Oki et al*., 2001; *Nohara et al*., 2006; *Guimberteau et al*., 2014]. Hence for most basins, including river routing processes results in simulations generates an improvement of modeled discharge in DGVMs compared to simply summing the runoff of all the relevant grid cells (Figure 9). The river routing model can capture the seasonal variation of discharge from the majority of the world river. However, several sources of uncertainty are related to the river routing model itself.

Some uncertainties are driven by the parameterization of routing processes [*Miller et al*., 1994; *Pitman*, 2003]. For the Ob, Amur, and Mackenzie River (Figure 9) basins, the observed discharge increases sharply in April or May, in response to massive snow melt or spring ice drift. However, the discharge simulation in most models when linked to our routing model has a flatter seasonal cycle, with a weak peak during summer. For those rivers in the middle or high latitudes, the existence of large snow and ice quantities, as well as the extensive presence of wetlands, modifies the relationship between river flow velocity and channel slope. This is through ice transport and jamming mechanisms, which are not considered in the current routing model [*Li et al*., 2013, 2014]. Using another routing scheme, *Ringeval et al*. [2012] obtained better discharge values through the ORCHIDEE model for Russian rivers and showed the importance of incorporating the residence time of water in wetlands to capture the timing of peak discharge by high latitude rivers. Hence, a limitation of our approach over the Ob, Amur, and Mackenzie River basins is the use of predefined flow velocities. The flow velocity is related to the slope of the river basins and then leads to a change in discharge through a river routing scheme [*Materia et al*., 2010]. The findings of another previous study supports that the spatial variation of the channel velocity has a different influence on the seasonality of river flow [*Li et al*., 2014]. For the Mississippi River (Figure 9), the simulated seasonal discharge without the routing scheme is similar to observation. However, when the routing model is incorporated, this causes the simulated peak timing to shift 2 or 3 months later. This bias is likely due to the underestimation of flow velocity in the routing model. In reality, the flow velocity is related to the quantity of flow, but this is not included in the routing model.

A second potential deficiency is that our routing model (in common with most routing models) only considers surface runoff as an input. Drainage from deeper soil layers is also routed with a longer water residence time [*Guimberteau et al*., 2012], and so lack of inclusion of this effect may additionally result in the biases in simulated discharge. Thirdly, in line with results from earlier studies [*Piao et al*., 2007; *Gerten et al*., 2008], we found that the change of land use is probably accelerating the river flow in the Mississippi Basin. In the TRENDY simulation S2, DGVMs were forced with fixed (present-day) land use [*Sitch et al*., 2008; *Sitch et al*., 2013]. Therefore, DGVMs forced with the changing land use are expected to reproduce better. And the simulated discharge without routing would be earlier than the observations and the simulated discharge with routing would be closer to the observations. In addition, the discharge of some DGVMs might be calibrated without considering a river routing scheme. This potential introduces "compensating errors," which may result in the discharge without being closer to the observations than when using more physically based routing descriptions.

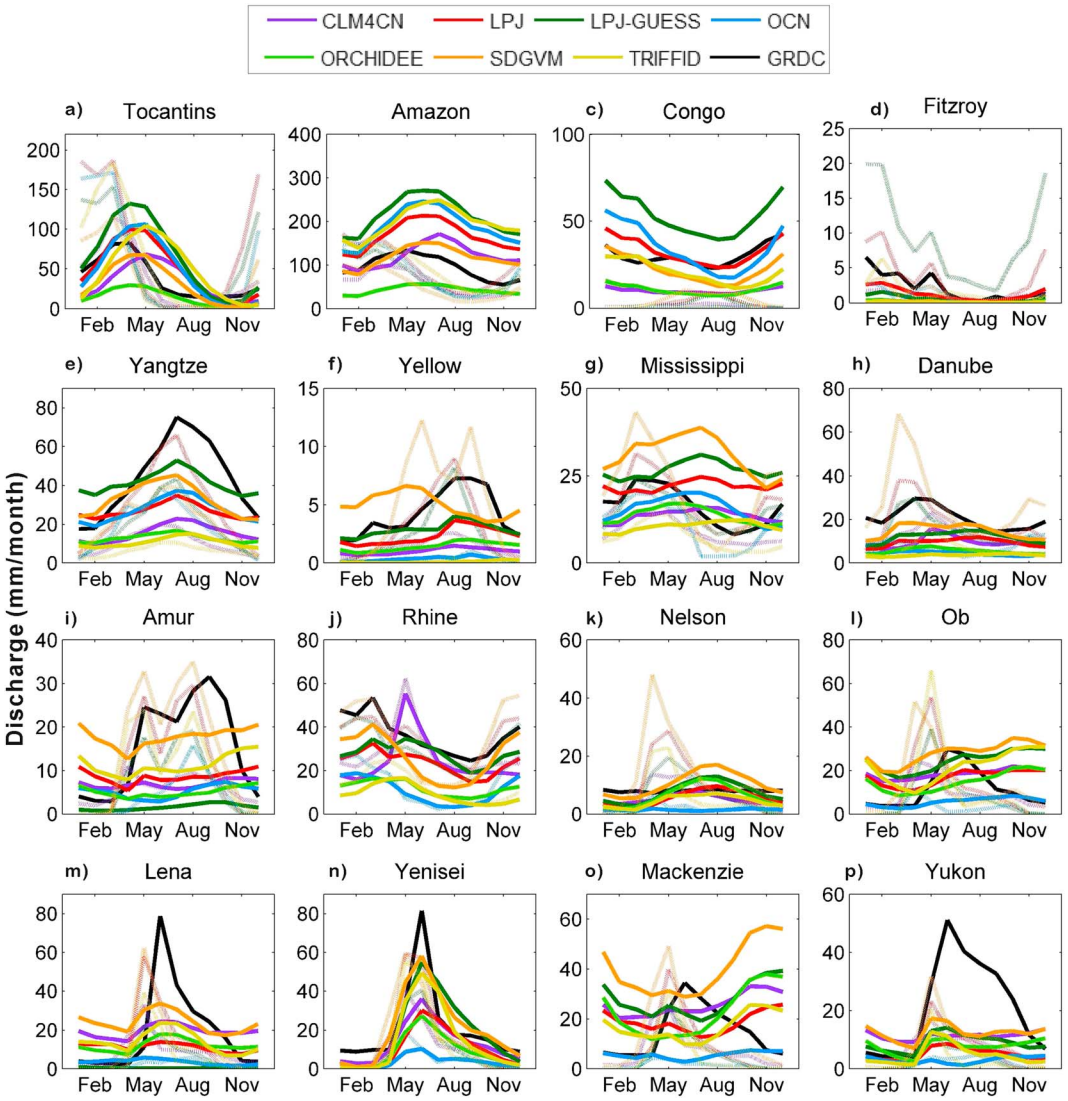

Figure 9. The monthly mean modeled discharge with routing (color solid lines), without routing (color dotted lines), and observed river flow (black solid line) over the 16 basins.

4.3. Uncertainty Associated With Scoring Metrics

There is uncertainty in the statistical computation of the scores, caused by relatively small sample sizes. A bootstrap resampling approach was used to estimate the uncertainty in scoring metrics. We took 1000 random samples of size 10 from the 12 model-data pairs of monthly discharge data. The correlation coefficient, PBIAS, and centered RSR were calculated for each of these samples (equations (4), (6), and (8)), and the average of the three indexes was taken as a score for each sample. Finally, the standard deviations of the scores from the bootstrap approach were computed to estimate the uncertainty in the mean score. The uncertainty was 0.20 overall across all basins. However, uncertainties in scores were different for different basins (ranging from 0.07 in the Yenisei Basin to 0.32 in the Mackenzie Basin).

4.4. Effects of Land Use Change and CO_2

Land use and land cover change (LULCC) impacts on simulation of runoff through evapotranspiration and soil moisture, where the latter is changed by different nutrient cycling, root depth distributions, and water consumption [*Jackson et al.*, 2005; *Piao et al.*, 2007; *Mango et al.*, 2011]. For instance, in temperate regions, short vegetation (such as crops) can transpire up to five times less water to the atmosphere than forests. Hence, observed discharge should account for the influence of LULCC over the basin, as well as for

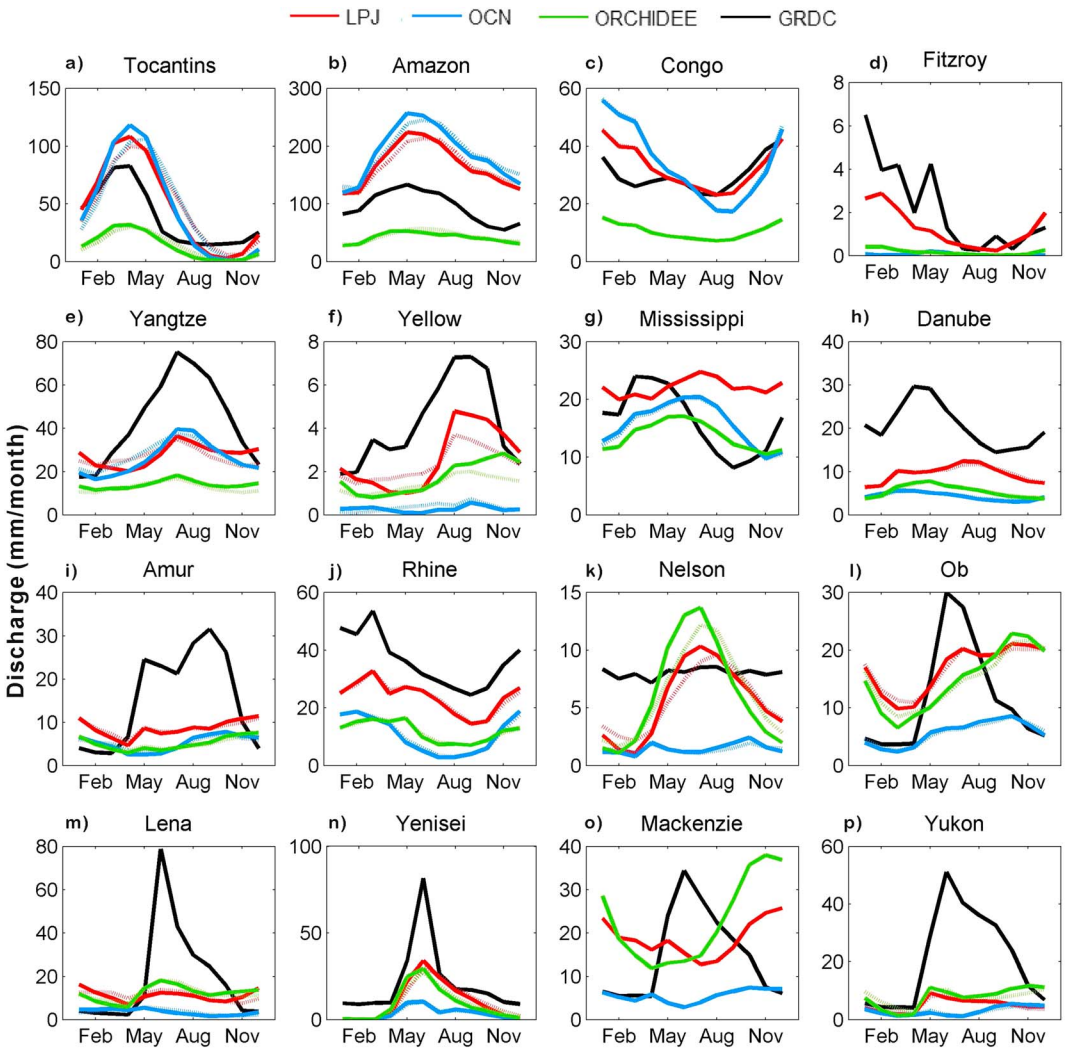

Figure 10. The monthly mean modeled discharge under S3 simulation (color solid lines), under S2 simulation (color dotted lines), and observed river flow (black solid line) over the 16 basins. Models were forced with time-invariant land use in simulation S2 while they were forced with changing land use in simulation S3.

irrigation and river damming. We therefore investigated if biases between simulated and observed discharges can be explained by nonmodeled land use change in S2, by comparing the model outputs of S2 with the ones of S3 where LULCC is included.

Land use change effects on the seasonal variation of discharge are illustrated in Figure 10. This result shows that there are only minor differences in the seasonality of monthly mean and annual mean river discharge due to land use change over the past 30 years, according to its representation in our DGVM results. However, we did find when considering differences S3 minus S2, a slight but significant downward trend in river discharge for most basins in the low and mid latitudes, which we believe to be due to changes in ET that result from land use. This result supports the conclusions of *Piao et al*. [2007] and *Gerten et al*. [2008].

Overall, therefore, this suggests that both the seasonal cycle and annual mean of discharge have been little affected by land use change over the basins but that the runoff trends are sensitive to land use change (Figures 10 and 11). We found that including land use change processes in the model slightly narrows the bias between simulated and observed discharges. One reason might be that the S3 simulation is forced with smaller land use trends over the last decade than over the last century, and hence, models differ less because of land use change if compared over the last decade. In addition, expansion of agriculture is considered as a major factor in land use change, which occurs in earlier decades, and has now started to

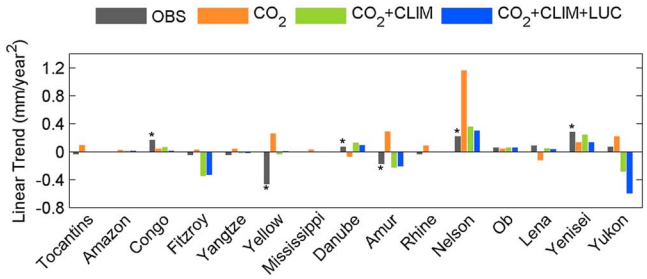

Figure 11. Discharge trends (mm/year^2) estimated by DGVMs for 15 river basin regions from 1981 to 2010. DGVM scenario simulations include "CO_2" (S1), "CO_2+CLIM" (S2), and "CO_2+CLIM+LUC" (S3). * Statistical significance at the 95% ($P < 0.05$) level. The trend for each model scenario was calculated as the median value of the trends estimated by different DGVM. (The modeled discharge data are unavailable in the Mackenzie Basin under S1 simulation.)

stabilize [*Klein Goldewijk et al.*, 2011]. Moreover, the effects of irrigation and damming on river discharge are generally not considered in most large-scale land surface models. Further studies are needed to further analyze these processes.

Although raised atmospheric carbon dioxide is believed to be influencing temperature and rainfall patterns, thus altering river discharge, there is also a direct CO_2 effect on vegetation which modifies river discharge. On the one hand, the direct effect of elevated CO_2 leads to reduced stomatal conductance, decreased ET, and increased runoff, especially in cases where soils are close to saturation [*Gedney et al.*, 2006]. If soils are undersaturated, soil moisture is maintained to higher value by elevated CO_2 but runoff may not necessarily increase. On the other hand, rising CO_2 can increase photosynthesis, resulting in increased leaf area index (LAI) and transpiration, leading to decreased river discharge [*Alkama et al.*, 2010; *Shi et al.*, 2011]. This balance of competing effects on river discharge can be estimated from the S1 simulation performed by all models in the TRENDY DGVM intercomparison protocol. In response to elevated atmospheric CO_2 (S1), river discharge generally shows a significant upward trend for 13 of 15 basins, exceptions being the Danube and Lena basins where the river discharge shows a downward trend (Figure 11). This result is similar to previous studies [*Alkama et al.*, 2010; *Betts et al.*, 2007; *Shi et al.*, 2011; *Tao et al.*, 2014] and supports the findings of *Gedney et al.* [2006].

5. Conclusion

In this study different evaluation metrics and a skill score based on the overlapping of PDFs reconstructed using kernel functions were used to assess performance of model in simulating the seasonal cycle, annual mean value, trends, and interannual variability of large river basin discharge during the last three decades. Systematic intercomparison of DGVMs is intended to identify the advantages and disadvantages of individual models for different basins. We found that DGVMs generally underestimate annual mean discharge and discharge trends. Both the seasonal and interannual variation of the discharge can be well reproduced by the DGVMs across basins in low and middle latitudes, with the exception of the Congo Basin. However, the skill scores of models are sensitive to the choice of river basins. Model performance shows a strong regional variation, and the best models for one particular region do not always surpass the others in other regions. We also found that for annual mean and interannual variability, the bias between simulated and observed discharges are mainly attributable to uncertainties in precipitation forcing and the physical construction of models, whereas for the seasonal cycle in high latitudes, bias may be caused by limitations of the river routing model chosen (no ice transport, jamming, wetlands, and floodplains). The trend of discharge was found to be sensitive to land use change but not the mean value and interannual discharge variability.

This study mainly discussed the uncertainty resulting from precipitation forcing, and limitations on soil hydrology parameters of DGVM models, but a single routing model was used. A comprehensive evaluation of large-scale river routing models remains a challenging task that requires investigating both uncertainty of processes controlling runoff (e.g., precipitation, transpiration, and soil water holding capacity) and water transit in a catchment (e.g., change in aquifers, routing, wetland and floodplains, and damming). Some caveats must be noted. We have performed a simple statistical evaluation of model performance. Further work is required to assess the complex interactions between discharge and precipitation, temperature, and evapotranspiration, to assess the model sensitivity to these external meteorological input parameters, and to further assess the uncertainty due to land surface parameterization schemes.

Acknowledgments
This study was supported by the National Natural Science Foundation of China (41125004 and 31321061), Chinese Ministry of Environmental Protection grant (201209031), and the 111 Project (B14001), and the National Youth Top-notch Talent Support Program in China. The National Center for Atmospheric Research is sponsored by the U.S. National Science Foundation. P.C. and M.G. acknowledge support from the Imbalance-P ERC-synergy grant. Data to support this article are from the protocol of TRENDY (http://dgvm.ceh.ac.uk/, accessed 11 July 2013) project and the Global River Discharge Center (GRDC; in Koblenz, Germany, http://www.bafg.de/GRDC/).

References

Adler, R. F., et al. (2003), The Version 2 Global Precipitation Climatology Project (GPCP) Monthly Precipitation Analysis (1979–Present), *J. Hydrometeorol.*, *4*, 1147–1167.

Alkama, R., M. Kageyama, and G. Ramstein (2010), Relative contributions of climate change, stomatal closure, and leaf area index changes to 20th and 21st century runoff change: A modelling approach using the Organizing Carbon and Hydrology in Dynamic Ecosystems (ORCHIDEE) land surface model, *J. Geophys. Res.*, *115*, D17112, doi:10.1029/2009JD013408.

Alkama, R., B. Decharme, H. Douville, and A. Ribes (2011), Trends in global and basin-scale runoff over the late twentieth century: Methodological issues and sources of uncertainty, *J. Clim.*, *24*(12), 3000–3014.

Anav, A., P. Friedlingstein, M. Kidston, L. Bopp, P. Ciais, P. Cox, C. Jones, M. Jung, R. Myneni, and Z. Zhu (2013), Evaluating the land and ocean components of the global carbon cycle in the CMIP5 earth system models, *J. Clim.*, *26*, 6801–6843, doi:10.1175/JCLI-D-12-00417.1.

Arora, V. K., and G. J. Boer (2001), Effects of simulated climate change on the hydrology of major river basins, *J. Geophys. Res.*, *106*(D4), 3335, doi:10.1029/2000JD900620.

Bates, B. C., Z. W. Kundzewicz, S. Wu, J. P. Palutikof, H. Gitay, A. Suárez, and R. T. Watson (2008), Climate change and water, Technical Paper of the Intergovernmental Panel on Climate Change, IPCC Secretariat, Geneva.

Barnett, T. P., J. C. Adam, and D. P. Lettenmaier (2005), Potential impacts of a warming climate on water availability in snow-dominated regions, *Nature*, *438*(7066), 303–309.

Best, M. J., et al. (2011), The Joint UK Land Environment Simulator (JULES), model description—Part 1: Energy and water fluxes, *Geosci. Model Dev.*, *4*(3), 677–699.

Betts, R. A., et al. (2007), Projected increase in continental runoff due to plant responses to increasing carbon dioxide, *Nature*, *448*(7157), 1037–1041.

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. Hydrometeorol.*, *10*(4), 1011–1025.

Blyth, E., D. B. Clark, R. Ellis, C. Huntingford, S. Los, M. Pryor, and S. Sitch (2011), A comprehensive set of benchmark tests for a land surface model of simultaneous fluxes of water and carbon at both the global and seasonal scale, *Geosci. Model Dev.*, *4*(2), 255–269, doi:10.5194/gmd-4-255-2011.

Chen, H., C. Y. Xu, and S. L. Guo (2012), Comparison and evaluation of multiple GCMs, statistical downscaling and hydrological models in the study of climate change impacts on runoff, *J. Hydrol.*, *434–435*, 36–45, doi:10.1016/j.jhydrol.2012.02.040.

Clark, D. B., et al. (2011), The Joint UK Land Environment Simulator (JULES), model description—Part 2: Carbon fluxes and vegetation dynamics, *Geosci. Model Dev.*, *4*(3), 701–722.

Cox, P. M. (2001), Description of the TRIFFID dynamic global vegetation model Technical Note 24 Hadley Centre, Met Office.

Cramer, W., A. Bondeau, F. I. Woodward, I. C. Prentice, R. A. Betts, V. Brovkin, and C. Young-Molling (2001), Global response of terrestrial ecosystem structure and function to CO_2 and climate change: Results from six dynamic global vegetation models, *Global Change Biol.*, *7*(4), 357–373, doi:10.1046/j.1365-2486.2001.00383.x.

Dai, A., and K. E. Trenberth (2002), Estimates of freshwater discharge from continents: Latitudinal and seasonal variations, *J. Hydrometeorol.*, *3*(6), 660–687, doi:10.1175/1525-7541(2002)003<0660:EOFDFC>2.0.CO;2.

Déry, S. J., T. J. Mlynowski, M. A. Hernández-Henríquez, and F. Straneo (2011), Interannual variability and interdecadal trends in Hudson Bay streamflow, *J. Mar. Syst.*, *88*(3), 341–351.

Fekete, B. M., C. J. Vörösmarty, J. O. Roads, and C. J. Willmott (2004), Uncertainties in precipitation and their impacts on runoff estimates, *J. Clim.*, *17*(2), 294–304.

Flato, G., et al. (2013), Evaluation of climate models, in *Climate Change 2013. The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by T. F. Stocker et al., pp. 741–866, Cambridge Univ. Press, Cambridge, U. K., and New York.

Gedney, N., P. M. Cox, R. A. Betts, O. Boucher, C. Huntingford, and P. A. Stott (2006), Detection of a direct carbon dioxide effect in continental river runoff records, *Nature*, *439*(7078), 835–838, doi:10.1038/nature04504.

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

Greve, P., B. Orlowsky, B. Mueller, J. Sheffield, M. Reichstein, and S. I. Seneviratne (2014), Global assessment of trends in wetting and drying over land, *Nat. Geosci.*, *7*(10), 716–721.

Gudmundsson, L., et al. (2012), Comparing large-scale hydrological model simulations to observed runoff percentiles in Europe, *J. Hydrometeorol.*, *13*(2), 604–620, doi:10.1175/JHM-D-11-083.1.

Guimberteau, M., et al. (2012), Discharge simulation in the sub-basins of the Amazon using ORCHIDEE forced by new datasets, *Hydrol. Earth Syst. Sci.*, *16*, 911–935, doi:10.5194/hess-16-911-2012.

Guimberteau, M., A. Ducharne, P. Ciais, J. P. Boisier, S. Peng, M. De Weirdt, and H. Verbeeck (2014), Testing conceptual and physically based soil hydrology schemes against observations for the Amazon Basin, *Geosci. Model Dev.*, *7*, 1115–1136.

Gupta, H. V., H. Kling, K. K. Yilmaz, and G. F. Martinez (2009), Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, *J. Hydrol.*, *377*(1–2), 80–91, doi:10.1016/j.jhydrol.2009.08.003.

Jackson, R. B., E. G. Jobbágy, R. Avissar, S. B. Roy, D. J. Barrett, C. W. Cook, K. A. Farley, D. C. le Maitre, B. A. McCarl, and B. C. Murray (2005), Trading water for carbon with biological carbon sequestration, *Science*, *310*(5756), 1944–1947.

Kalnay, E., et al. (1996), The NCEP/NCAR 40-year reanalysis project, *Bull. Am. Meteorol. Soc.*, *77*(3), 437–471.

Keeling, C. D., and T. P. Whorf (2005), Atmospheric carbon dioxide record from Mauna Loa, in *Trends: A Compendium of Data on Global Change. Carbon Dioxide Information Analysis Center*, pp. 1–6, Oak Ridge Natl. Lab., Oak Ridge, Tenn.

Keeling, R. F., S. C. Piper, A. F. Bollenbacher, and J. S. Walker (2009), Atmospheric CO_2 records from sites in the SIO air sampling network, in *Trends: A Compendium of Data on Global Change. Carbon Dioxide Information Analysis Center*, pp. 1–6, Oak Ridge Natl. Lab., U.S. Dep. of Energy, Oak Ridge, Tenn., doi:10.3334/CDIAC/atg.035.

Kim, J. (2005), A projection of the effects of the climate change induced by increased CO_2 on extreme hydrologic events in the western US, *Clim. Change*, *68*(1–2), 153–168, doi:10.1007/s10584-005-4787-9.

Klein Goldewijk, K., A. Beusen, G. van Drecht, and M. de Vos (2011), The HYDE 3.1 spatially explicit database of human induced land use change over the past 12,000 years, *Global Ecol. Biogeogr.*, *20*(1), 73–86, doi:10.1111/j.1466-8238.2010.00587.x.

Knutti, R. (2010), The end of model democracy?, *Clim. Change*, *102*(3–4), 395–404, doi:10.1007/s10584-010-9800-2.

Krinner, G., N. Viovy, N. de Noblet-Ducoudré, J. Ogée, J. Polcher, P. Friedlingstein, P. Ciais, S. Sitch, and I. C. Prentice (2005), A dynamic global vegetation model for studies of the coupled atmosphere-biosphere system, *Global Biogeochem. Cycles*, *19*, GB1015, doi:10.1029/2003GB002199.

Labat, D., Y. Goddéris, J. L. Probst, and J. L. Guyot (2004), Evidence for global runoffrunoff increase related to climate warming, *Adv. Water Resour.*, *27*(6), 631–642, doi:10.1016/j.advwatres.2004.02.020.

Lawrence, D. M., et al. (2011), Parameterization improvements and functional and structural advances in version 4 of the Community Land Model, *J. Adv. Model. Earth Syst.*, *3*, M03001, doi:10.1029/2011MS00045.
Legates, D. R., and G. J. McCabe (1999), Evaluating the use of "goodness-of-fit" Measures in hydrologic and hydroclimatic model validation, *Water Resour. Res.*, *35*(1), 233–241, doi:10.1029/1998WR900018.
Legates, D. R., H. F. Lins, and G. J. McCabe (2005), Comments on "Evidence for global runoff increase related to climate warming" by Labat et al., *Adv. Water Resour.*, *28*(12), 1310–1315, doi:10.1016/j.advwatres.2005.04.006.
Li, H. Y., M. S. Wigmosta, H. Wu, M. Huang, Y. Ke, A. M. Coleman, and L. Ruby Leung (2013), A physically based runoff routing model for land surface and earth system models, *J. Hydrometeorol.*, *14*, 808–828, doi:10.1175/JHM-D-12-015.1.
Li, H.-Y., L. R. Leung, A. Getirana, M. Huang, H. Wu, Y. Xu, J. Guo, and N. Voisin (2014), Evaluating global streamflow simulations by a physically based routing model coupled with the community land model, *J. Hydrometeorol.*, *16*, 948–971, doi:10.1175/JHM-D-14-0079.1.
Liston, G. E. (1999), Interrelationships among snow distribution, snowmelt, and snow cover depletion: Implications for atmospheric, hydrologic, and ecologic modeling, *J. Appl. Meteorol.*, *38*(10), 1474–1487.
Mango, L. M., A. M. Melesse, M. E. McClain, D. Gann, and S. G. Setegn (2011), Land use and climate change impacts on the hydrology of the upper Mara River Basin, Kenya: Results of a modeling study to support better resource management, *Hydrol. Earth Syst. Sci.*, *15*(7), 2245–2258.
Materia, S., P. A. Dirmeyer, Z. Guo, A. Alessandri, and A. Navarra (2010), The sensitivity of simulated river discharge to land surface representation and meteorological forcings, *J. Hydrometeorol.*, *11*(2), 334–351, doi:10.1175/2009JHM1162.1.
Mao, J., B. Wang, Y. Dai, F. I. Woodward, P. J. Hanson, and M. R. Lomas (2007), Improvements of a dynamic global vegetation model and simulations of carbon and water at an upland-oak forest, *Adv. Atmos. Sci.*, *24*(2), 311–322, doi:10.1007/s00376-007-0311-7.
Maxino, C. C., B. J. McAvaney, A. J. Pitman, and S. E. Perkins (2008), Ranking the AR4 climate models over the Murray-Darling Basin using simulated maximum temperature, minimum temperature and precipitation, *Int. J. Climatol.*, *28*(8), 1097–1112, doi:10.1002/joc.1612.
Miller, J. R., G. L. Russell, and G. Caliri (1994), Continental-scale river flow in climate models, *J. Clim.*, *7*(6), 914–928, doi:10.1175/1520-0442(1994)007<0914:CSRFIC>2.0.CO;2.
Milliman, J. D., K. L. Farnsworth, P. D. Jones, K. H. Xu, and L. C. Smith (2008), Climatic and anthropogenic factors affecting river discharge to the global ocean, 1951–2000, *Global Planet. Change*, *62*(3), 187–194, doi:10.1016/j.gloplacha.2008.03.001.
Milly, P. C., K. A. Dunne, and A. V. Vecchia (2005), Global pattern of trends in streamflow and water availability in a changing climate, *Nature*, *438*(7066), 347–350, doi:10.1038/nature04312.
Moriasi, D. N., J. G. Arnold, M. W. Van Liew, R. L. Bingner, R. D. Harmel, and T. L. Veith (2007), Model evaluation guidelines for systematic quantification of accuracy in watershed simulations, *Trans. ASABE*, *50*(3), 885–900.
Murray, S. J., I. M. Watson, and I. C. Prentice (2013), The use of dynamic global vegetation models for simulating hydrology and the potential integration of satellite observations, *Prog. Phys. Geogr.*, *37*(1), 63–97, doi:10.1177/0309133312460072.
Nash, J., and J. V. Sutcliffe (1970), River flow forecasting through conceptual models part I: A discussion of principles, *J. Hydrol.*, *10*(3), 282–290, doi:10.1016/0022-1694(70)90255-6.
Nohara, D., A. Kitoh, M. Hosaka, and T. Oki (2006), Impact of climate change on river discharge projected by multimodel ensemble, *J. Hydrometeorol.*, *7*(5), 1076–1089, doi:10.1175/JHM531.1.
Oki, T., and S. Kanae (2006), Global hydrological cycles and world water resources, *Science*, *313*(5790), 1068–1072, doi:10.1126/science.1128845.
Oki, T., Y. Agata, S. Kanae, T. Saruhashi, D. Yang, and K. Musiake (2001), Global assessment of current water resources using total runoff integrating pathways, *Hydrol. Sci. J.*, *46*(6), 983–995, doi:10.1080/02626660109492890.
Oleson, K. W., D. M. Lawrence, B. Gordon, M. G. Flanner, E. Kluzek, J. Peter, S. Levis, S. C. Swenson, E. Thornton, and J. Feddema (2010), Technical description of version 4.0 of the Community Land Model (CLM), NCAR/TN-478+STR, NCAR, Boulder, Colo.
Piao, S., P. Friedlingstein, P. Ciais, N. de Noblet-Ducoudre, D. Labat, and S. Zaehle (2007), Changes in climate and land use have a larger direct impact than rising CO_2 on global river runoff trends, *Proc. Natl. Acad. Sci. U.S.A.*, *104*(39), 15,242–15,247, doi:10.1073/pnas.0707213104.
Pitman, A. J. (2003), The evolution of, and revolution in, land surface schemes designed for climate models, *Int. J. Climatol.*, *23*(5), 479–510, doi:10.1002/joc.893.
Ringeval, B., B. Decharme, S. L. Piao, P. Ciais, F. Papa, N. D. Noblet-Ducoudré, and A. Ducharne (2012), Modelling sub-grid wetland in the ORCHIDEE global land surface model: Evaluation against river discharges and remotely sensed data, *Geosci. Model Dev.*, *5*(1), 683–735, doi:10.5194/gmdd-5-683-2012.
Schneider, U., A. Becker, P. Finger, A. Meyer-Christoffer, B. Rudolf, and M. Ziese (2011), GPCC full data reanalysis version 6.0 at 0.5: Monthly land-surface precipitation from rain-gauges built on GTS-based and historic data, doi:10.5676/DWD_GPCC.FD_M_V6_050.
Singh, J., H. V. Knapp, J. G. Arnold, and M. Demissie (2005), Hydrological modeling of the iroquois river watershed using HSPF and SWAT, *J. Am. Water Resour. Assoc.*, *41*(2), 343–360, doi:10.1111/j.1752-1688.2005.tb03740.x.
Sitch, S., B. Smith, I. C. Prentice, A. Arneth, A. Bondeau, W. Cramer, J. Kaplan, S. Levis, W. Lucht, and M. Sykes (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, doi:10.1046/j.1365-2486.2003.00569.x.
Sitch, S., et al. (2008), Evaluation of the terrestrial carbon cycle, future plant geography and climate-carbon cycle feedbacks using five Dynamic Global Vegetation Models (DGVMs), *Global Change Biol.*, *14*(9), 2015–2039, doi:10.1111/j.1365-2486.2008.01626.x.
Sitch, S., et al. (2013), Trends and drivers of the regional-scale sources and sinks of carbon dioxide over the past two decades, *Biogeosciences*, *10*, 20,113–20,177, doi:10.5194/bgd-10-20113-2013.
Shi, X., J. Mao, P. E. Thornton, F. M. Hoffman, and W. M. Post (2011), The impact of climate, CO2, nitrogen deposition and land use change on simulated contemporary global river flow, *Geophys. Res. Lett.*, *38*, L08704, doi:10.1029/2011GL046773.
Smith, B., I. C. Prentice, and M. T. Sykes (2001), Representation of vegetation dynamics in the modelling of terrestrial ecosystems: Comparing two contrasting approaches within European climate space, *Global Ecol. Biogeogr.*, *10*(6), 621–637, doi:10.1046/j.1466-822X.2001.t01-1-00256.x.
Solomon, S., et al. (2007), The physical science basis, in *Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, pp. 235–337, Cambridge Univ. Press, Cambridge, U. K., and New York.
Stocker, T. F., Q. Dahe, and G. K. Plattner (2013), Climate change 2013: The physical science basis, in *Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Summary for Policymakers (IPCC, 2013)*, pp. 159–254, Cambridge Univ. Press, Cambridge, U. K., and New York.
Tao, B., H. Tian, W. Ren, J. Yang, Q. Yang, R. He, W. Cai, and S. Lohrenz (2014), Increasing Mississippi river discharge throughout the 21st century influenced by changes in climate, land use, and atmospheric CO_2, *Geophys. Res. Lett.*, *41*, 4978–4986, doi:10.1002/2014GL060361.
Taylor, K. E. (2001), Summarizing multiple aspects of model performance in a single diagram, *J. Geophys. Res.*, *106*(D7), 7183–7192, doi:10.1029/2000JD900719.

Woodward, F. I., and M. R. Lomas (2004), Simulating vegetation processes along the Kalahari transect, *Global Change Biol.*, *10*(3), 383–392, doi:10.1046/j.1365-2486.2003.00697.x.

Woodward, F. I., T. M. Smith, and W. R. Emanuel (1995), A global land primary productivity and phytogeography model, *Global Biogeochem. Cycles*, *9*(4), 471–490, doi:10.1029/95GB02432.

Wu, H., J. S. Kimball, N. Mantua, and J. Stanford (2011), Automated upscaling of river networks for macroscale hydrological modeling, *Water Resour. Res.*, *47*, W03517, doi:10.1029/2009WR008871.

Wu, H., J. S. Kimball, H. Li, M. Huang, L. R. Leung, and R. F. Adler (2012), A new global river network database for macroscale hydrologic modeling, *Water Resour. Res.*, *48*, W09701, doi:10.1029/2012WR012313.

Zaehle, S., A. D. Friend, P. Friedlingstein, F. Dentener, P. Peylin, and M. Schulz (2010), Carbon and nitrogen cycle dynamics in the O-CN land surface model: 2. Role of the nitrogen cycle in the historical terrestrial carbon balance, *Global Biogeochem. Cycles*, *24*, GB1006, doi:10.1029/2009GB003522.