



HAL
open science

Regional patterns of future runoff changes from Earth system models constrained by observation

Hui Yang, Feng Zhou, Shilong Piao, Mengtian Huang, Anping Chen, Philippe Ciais, Yue Li, Xu Lian, Shushi Peng, Zhenzhong Zeng

► **To cite this version:**

Hui Yang, Feng Zhou, Shilong Piao, Mengtian Huang, Anping Chen, et al.. Regional patterns of future runoff changes from Earth system models constrained by observation. *Geophysical Research Letters*, 2017, 44 (11), pp.5540-5549. 10.1002/2017GL073454 . hal-02903510

HAL Id: hal-02903510

<https://hal.science/hal-02903510>

Submitted on 6 May 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



RESEARCH LETTER

10.1002/2017GL073454

H.Y. and F.Z. contributed equally to this work.

Key Points:

- We use a Bayesian model ensemble, where higher weights are given to models with better historical performance, to predict runoff changes
- Our results suggest a more pronounced runoff decrease in Amazon and a weaker increase in northern high latitudes
- The large intermodel spread of runoff changes is mainly due to differences in modeled precipitation changes

Supporting Information:

- Supporting Information S1

Correspondence to:

S. Piao,
spiao@pku.edu.cn

Citation:

Yang, H., F. Zhou, S. Piao, M. Huang, A. Chen, P. Ciais, Y. Li, X. Lian, S. Peng, and Z. Zeng (2017), Regional patterns of future runoff changes from Earth system models constrained by observation, *Geophys. Res. Lett.*, 44, 5540–5549, doi:10.1002/2017GL073454.

Received 13 MAR 2017

Accepted 13 MAY 2017

Accepted article online 17 MAY 2017

Published online 12 JUN 2017

Regional patterns of future runoff changes from Earth system models constrained by observation

Hui Yang¹ , Feng Zhou¹ , Shilong Piao^{1,2} , Mengtian Huang¹, Anping Chen³ , Philippe Ciais⁴, Yue Li¹, Xu Lian¹ , Shushi Peng¹ , and Zhenzhong Zeng¹

¹Sino-French Institute for Earth System Science, College of Urban and Environmental Sciences, Peking University, Beijing, China, ²Key Laboratory of Alpine Ecology and Biodiversity, Institute of Tibetan Plateau Research, Center for Excellence in Tibetan Earth Science, Chinese Academy of Sciences, Beijing, China, ³The Woods Hole Research Center, Falmouth, Massachusetts, USA, ⁴LSCE, UMR CEA-CNRS, Gif-sur-Yvette, France

Abstract In the recent Intergovernmental Panel on Climate Change assessment, multimodel ensembles (arithmetic model averaging, AMA) were constructed with equal weights given to Earth system models, without considering the performance of each model at reproducing current conditions. Here we use Bayesian model averaging (BMA) to construct a weighted model ensemble for runoff projections. Higher weights are given to models with better performance in estimating historical decadal mean runoff. Using the BMA method, we find that by the end of this century, the increase of global runoff ($9.8 \pm 1.5\%$) under Representative Concentration Pathway 8.5 is significantly lower than estimated from AMA ($12.2 \pm 1.3\%$). BMA presents a less severe runoff increase than AMA at northern high latitudes and a more severe decrease in Amazonia. Runoff decrease in Amazonia is stronger than the intermodel difference. The intermodel difference in runoff changes is mainly caused not only by precipitation differences among models, but also by evapotranspiration differences at the high northern latitudes.

1. Introduction

River runoff is an essential source of freshwater, indispensable for maintaining food security and economic prosperity [Vörösmarty *et al.*, 2000; Hall *et al.*, 2014]. Both observations and theoretical analyses have shown that the ongoing climate change has a remarkable influence on large-scale hydrological processes, including river runoff [Milly *et al.*, 2005; Piao *et al.*, 2007; Gerten *et al.*, 2008]. Earth system models (ESMs) built on atmosphere-ocean general circulation models generally predict an increase in future global runoff under increased radiative forcing, with significant regional differences. It is generally predicted that river runoff will increase at high latitudes and in the humid tropics but decrease in most dry tropical regions [Cisneros *et al.*, 2014]. Nevertheless, projections of runoff changes are highly uncertain, especially at the regional scale (Figures S1 and S2 in the supporting information). For instance, the standard deviation of runoff change derived from multi-ESM ensemble in the Fifth Assessment Report (AR5) of the Intergovernmental Panel on Climate Change (IPCC) is 2 times larger than the mean runoff change across roughly half of the global land area. Multimodel uncertainty is especially large in the tropics [Collins *et al.*, 2013]. Reducing the uncertainty in future runoff change projections is important, given the need to manage finite water resources [Schewe *et al.*, 2014] to meet the growing human demand for water [Alcamo *et al.*, 2007].

The IPCC AR5 assessed scenario-dependent runoff changes with an ensemble of all ESMs [Taylor *et al.*, 2012]. This ensemble takes into account climate projections of individual ESMs and assigns equal weight to each model—an approach often referred as “model democracy” [Collins *et al.*, 2013]. The results of such a “one model-one vote” approach (hereafter AMA for arithmetic model averaging) should be viewed with caution, given the large intermodel spread [Knutti, 2010] and the likely distortion of the magnitude or sign of mean runoff change by a few outlier models with unrealistic values. Here we apply Bayesian model averaging (BMA) [Raftery *et al.*, 2005; Tebaldi *et al.*, 2005; Vrugt *et al.*, 2008] to recalculate the runoff using a weighted ensemble. Higher weights are given to those models with better performance at estimating historical decadal mean runoff.

2. Materials and Methods

2.1. Model Selection

We use ESMs from the Coupled Model Intercomparison Project Phase 5 (CMIP5) to perform the BMA analysis on modeled runoff. Most of the models have realistic representations (e.g., the representation

of vegetation dynamics and subgrid scale hydrology in some models) of processes related to surface hydrology [Moss *et al.*, 2010; Knutti and Sedláček, 2012; Flato *et al.*, 2013]. Only CMIP5 ESMs that provide monthly runoff of both historical simulations and scenario-dependent projections (Representative Concentration Pathway, RCP4.5 and RCP8.5) are selected in this study, resulting in a list of 31 ESMs (Table S1 in the supporting information). Outputs of the selected CMIP5 models are downloaded from the British Atmospheric Data Centre. To facilitate comparison with observations, all model outputs for the period 1986–2100 are regridded to the resolution of 1° latitude × 1° longitude using the first-order conservative remapping technique [Jones, 1999]. Although the regridded resolution of 1° × 1° is coarser than that provided by some models, it has negligible effect on determining BMA weight for each model (Text S1 and Figure S3 in the supporting information).

2.2. Data Sets

The observation-based “diagnostic” annual runoff from University of New Hampshire/Global River Discharge Center global composite runoff fields (UNH/GRDC GCRF) [Hall *et al.*, 2006; Fekete *et al.*, 2002; Fekete and Vorosmarty, 2011] is used to constrain the multimodel ensemble of CMIP5 models. This data set provides annual runoff at a resolution of 1° × 1° for the period 1986–1995 (Text S2). We also use mean annual discharge records from the Global River Discharge Center (GRDC) to further validate the accuracy of the two ensemble methods (BMA and AMA). From a previously defined set of top 200 large rivers [Dai and Trenberth, 2002], we select 28 rivers (Table S2 and Figure S4) that satisfy the following criteria: (i) discharge has been continuously measured for more than 10 years at the gauging station closest to the estuary, (ii) drainage area is greater than 100,000 km², and (iii) geographic information system shapefile can be acquired from the GRDC.

2.3. Weighted Ensemble

The skill of each CMIP5 model in simulating the UNH/GRDC GCRF runoff, measured by root-mean-square error (RMSE), varies greatly across different regions. We therefore use the BMA approach to determine the optimal weights for each CMIP5 model at spatial windows of 3° × 3°, each consisting of nine 1° × 1° grid cells. The idea of the BMA approach is to find a weighted model ensemble with the minimal RMSE when compared to observations [Raftery *et al.*, 2005]. Supposing that Q_i is annual runoff for the period 1986–1995 simulated by the i th CMIP5 model ($i = 1, \dots, 31$) for a spatial window of 3° × 3°, the probability density function (PDF) of annual runoff (Q) from UNH/GRDC data is thus defined as

$$p(Q|Q_1, \dots, Q_{31}) = \sum_{i=1}^{31} w_i g(Q|Q_i), \quad (1)$$

where w_i is the optimal weight for the i th model, $g(Q|Q_i)$ is a PDF of normal distribution with mean $a_i + b_i Q_i$ and standard deviation σ (Text S3). Parameters a_i and b_i are bias-correction terms that are estimated from the linear regression of annual runoff observations on Q_i for each of the 31 ensemble members. Thus, the BMA mean can be computed as $E[Q|Q_1, \dots, Q_{31}] = \sum_{i=1}^{31} w_i (a_i + b_i Q_i)$ and BMA variance is written as

$$\text{Var}[Q|Q_1, \dots, Q_{31}] = \sum_{i=1}^{31} w_i \left((a_i + b_i Q_i) - \sum_{j=1}^{31} w_j (a_j + b_j Q_j) \right)^2 + \sigma^2.$$

We estimate w_i and standard deviation σ with a maximum likelihood method. Here we used observation-based “diagnostic” annual runoff as constraints. Assuming model projection errors in space are independent, the log likelihood function of equation (1) for a spatial window of 3° × 3° (nine grid cells) is

$$l(w_1, \dots, w_{31}, \sigma^2) = \sum_{s=1}^S \log \left(\sum_{i=1}^{31} w_i g(Q_s|Q_{i,s}) \right), \quad (2)$$

where S is the total number of cell grids with data for the spatial window of 3° × 3° ($S \leq 9$). $g(Q_s|Q_{i,s})$ is the conditional PDF of annual runoff at the s th grid (Q_s), given that $Q_{i,s}$ is the runoff simulation from the i th model for that grid. The BMA method assumes that the conditional PDF $g(Q_s|Q_{i,s})$ can be approximated by a normal distribution noted above. The maximum likelihood estimator $l(w_1, \dots, w_{31}, \sigma^2)$ is then approximated through a Markov chain Monte Carlo algorithm [Vrugt *et al.*, 2008] (Text S4). Furthermore, we also estimate the BMA

weights w for spatial windows of $5^\circ \times 5^\circ$, and the results confirm that the weight for each model is independent of the size of spatial windows (Text S5 and Figure S5).

Temporal changes in runoff are mainly caused by variations in precipitation (P_{pt}) and evapotranspiration (ET). Since there are large differences among available ET products [Zeng *et al.*, 2014], we do not use any ET products to constrain runoff changes. Monthly precipitation data are from the Global Precipitation Climatology Centre (GPCC) for the same period 1986–1995 [Rudolf *et al.*, 2005]. We compared mean annual runoff estimates from model ensembles constrained by (1) only runoff, (2) only precipitation, and (3) both precipitation and runoff (Text S6). The results show that the model ensemble constrained by only runoff performs the best in reproducing historical runoff ($R = 0.92$; RMSE = 194.5 mm yr^{-1}) and very well in reproducing historical precipitation ($R = 0.90$; RMSE = 332.4 mm yr^{-1} ; Figure S6). Therefore, we use this model ensemble constrained by historical runoff to project future runoff changes.

Furthermore, considering the large variation among different precipitation products, especially in mountainous regions and the Arctic [Biemans *et al.*, 2009], we also test model-ensemble results constrained by five other precipitation products (i.e., CRU, GPCP, CPP, Udel, and Prec/L; Table S3 and Figure S7). Consistent with results constrained by the GPCC precipitation product, none of the weighted ensembles constrained by other precipitation data products can well reproduce the observed runoff, measured by the correlation coefficient (R ranges from 0.42 to 0.75) and RMSE (ranges from 337.2 to 464.2 mm yr^{-1}) between model-simulated and observed runoff.

2.4. A Metric for Robustness

Following the definition in Knutti and Sedláček [2012], metric ψ was applied to quantify the robustness of the signal (A_2) of future runoff changes relative to the noise (A_1). As shown in Figure S8, A_1 (noise) is the integral of the squared area (blue) between two cumulative distribution functions (CDFs) characterizing projected annual runoff from individual models (combined in a single CDF) and from the multimodel ensemble, respectively, and A_2 (signal) is the integral of the squared area between two CDFs characterizing historical (black solid line) and future projected (red solid line) annual runoff from the multimodel ensemble, respectively. Supposing that annual runoff over the historical (1986–2005) or future period (2081–2100) follows a Gaussian distribution, we estimate their CDF by using sample mean and variance [Weigel *et al.*, 2007; Knutti and Sedláček, 2012]. ψ is calculated as $\psi = 1 - A_1/A_2$ for each $1^\circ \times 1^\circ$ grid cell. $\psi = 1$ implies a significant change with high model agreement, and ψ approaching 0 indicates higher model spread or weak changes in runoff from the multimodel ensemble. The detailed description of this approach can be found in Knutti and Sedláček [2012].

3. Result and Discussion

First, we examine the model performance in reproducing historical runoff for each of the 31 ESMs used in the IPCC AR5. Annual runoff variation from 1986 to 1995 estimated by CMIP5 models is not correlated ($P > 0.05$; Figure S9) with the observation-based annual runoff over at least 95.3% of grid cells that have runoff data. The observation-based runoff is from UNH/GRDC GCRF. Given the high dependence of runoff variations on model initial conditions, it is difficult for ESMs to reproduce the observed interannual variability in runoff [Flato *et al.*, 2013]. As shown in Figure S10, annual precipitation modeled by individual models also insignificantly correlate ($P > 0.05$) with the GPCC data in more than 95.5% of the grid cells. Neither the UNH/GRDC GCRF data product nor the CMIP5 models show statistically significant runoff trends during the period of 1986–1995 (Figure S11). We thus calculated the BMA optimal weights (w_i) for each CMIP5 model based on their capability in reproducing observed mean annual runoff from 1986 to 1995.

Figure 1 shows the global pattern of runoff simulated by the weighted ensemble (BMA) for the period 1986–1995. Because the weighted ensemble is constrained by observations, the BMA ensemble is better than the AMA ensemble when plotted against UNH/GRDC GCRF data (BMA: $R = 0.92$, $P < 0.05$, Figure 1a; AMA: $R = 0.60$, $P < 0.05$, Figure S12c). The different capacity between BMA and AMA methods in reproducing observed runoff is particularly evident in high- or low-runoff regions (Figure 1b). BMA reproduced runoff is larger than that of AMA in most tropical regions but smaller in western North America, temperate Eurasia, the eastern Tibetan Plateau, and southern Africa (Figure S12f). We further compare BMA-based runoff

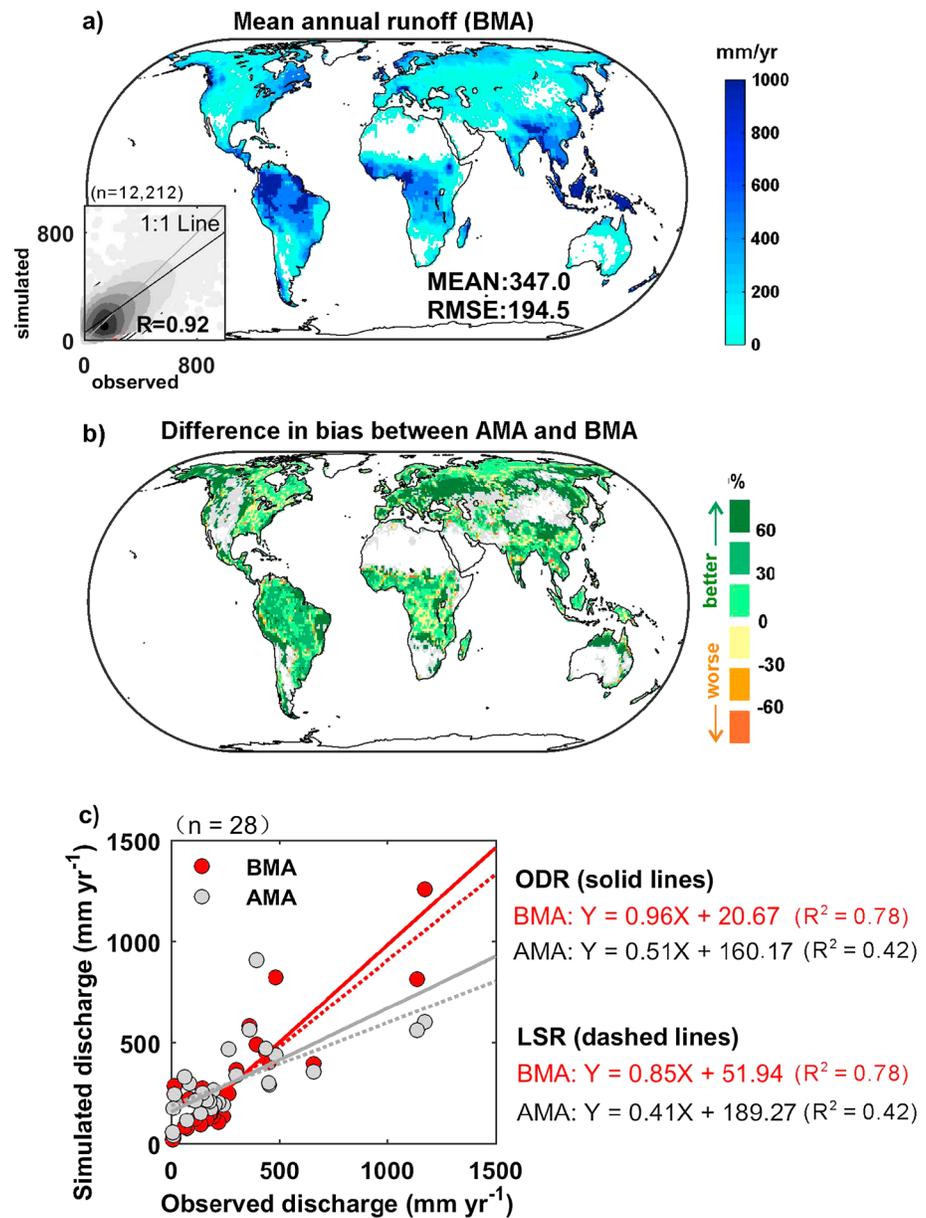


Figure 1. (a) Pattern of mean annual runoff (1986–1995) simulated by the BMA model ensemble, for which comparison of the simulated and observed values at the grid cell scale is shown in the inset at the bottom left of the map. The gray level indicates the density of points, with darker color representing the denser points. The white color in the map marks the areas without runoff in UNH/GRDC GCRF. (b) Pattern of the difference between the AMA’s biases in mean annual runoff and that of the BMA’s. Biases are computed as the ratio of absolute difference between simulated and observed mean annual runoff to observed value in each grid cell. (c) Comparison of the mean annual discharge from the BMA and AMA model ensemble against the observed annual discharge from gauge stations versus simulated over 28 river basins (Table S2). The orthogonal distance regression (ODR, solid lines) and ordinary least-square regression (LSR, dashed lines) are used to show the performance of model ensemble.

constrained by UNH/GRDC runoff with observed mean annual river discharges. Because not all CMIP5 models simulate river routing, we use the routing model of *Miller et al.* [1994] to convert gridded runoff to river discharge in the 28 river basins. The results show that BMA is also better than AMA in reproducing observed mean annual river discharges (Figures 1c and S13).

Figure 2 shows differences in future runoff changes between BMA and AMA under both RCP4.5 and RCP8.5. Compared with the reference period 1986–2005, BMA-based projection of global runoff for 2081–2100 will

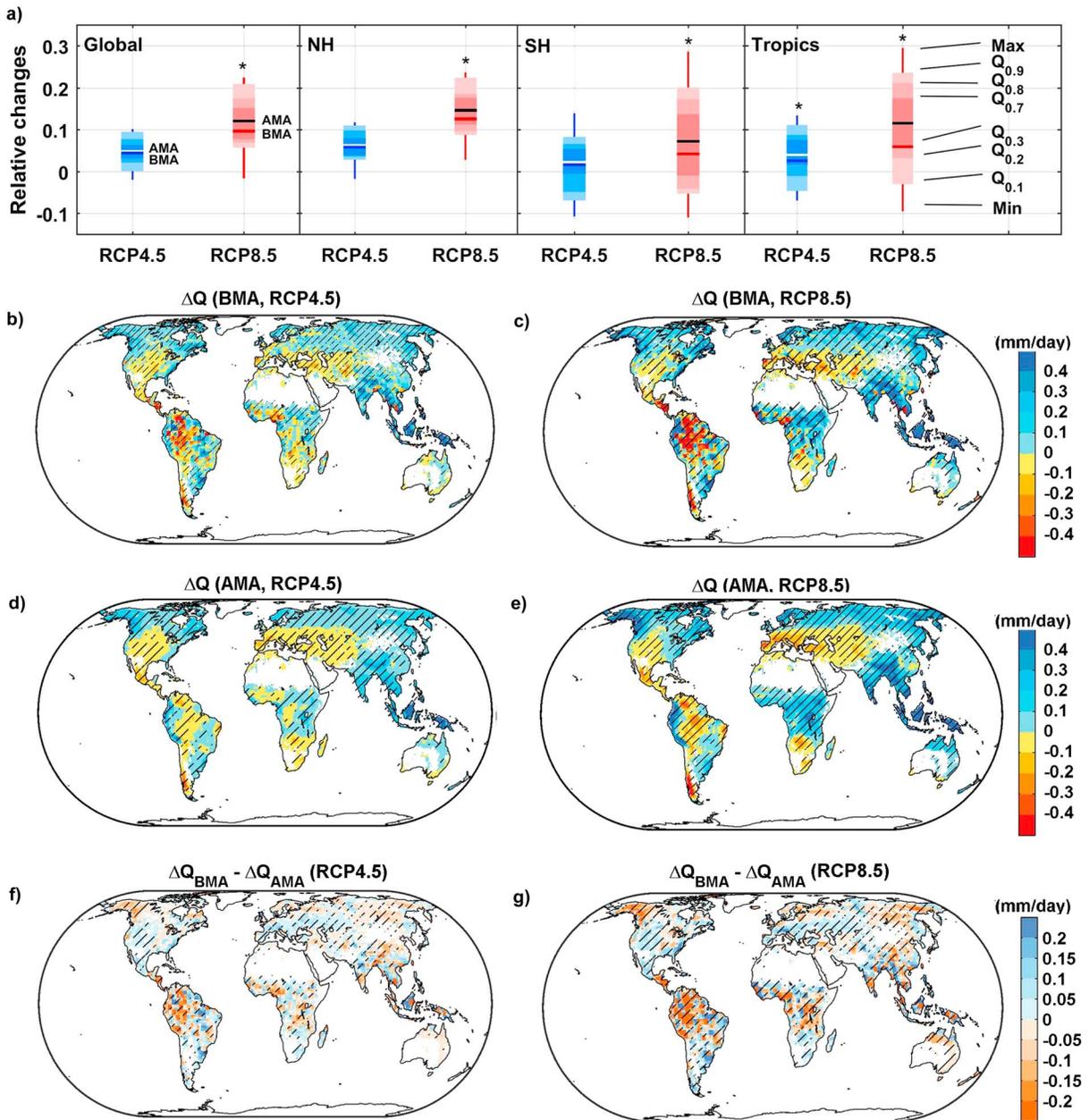


Figure 2. Patterns of projected runoff changes between the periods 2081–2100 and 1986–2005. (a) The BMA and AMA multimodel mean relative runoff change over the globe, Northern Hemisphere (NH), Southern Hemisphere (SH), and tropics (25°S–25°N) under RCP4.5 and RCP8.5. The whiskers and boxes indicate the minimum (Min); the 10th ($Q_{0.1}$), 20th ($Q_{0.2}$), and 30th ($Q_{0.3}$) percentiles; multimodel ensemble using BMA (blue and red) or AMA (white and black) approaches; the 70th ($Q_{0.7}$), 80th ($Q_{0.8}$), and 90th ($Q_{0.9}$) percentiles; and maximum (Max). The asterisks indicate significant difference in relative runoff changes by the BMA and AMA model ensemble according to the Wilcoxon signed-rank test ($P < 0.05$). Projected runoff changes (ΔQ) by the BMA ensemble under (b) RCP4.5 and (c) RCP8.5. (d and e) The same as for Figures 2b and 2c but by AMA ensemble. The difference in projected runoff changes between the BMA and AMA ensemble (i.e., $\Delta Q_{BMA} - \Delta Q_{AMA}$) under (f) RCP4.5 and (g) RCP8.5. Hatching marks significant changes according to the Wilcoxon signed-rank test ($P < 0.05$).

increase by $4.5 \pm 1.0\%$ (1.0% is the standard deviation of runoff changes over differing periods, 1986–2081, 1987–2082, ..., 2005–2100) under RCP4.5 and by $9.8 \pm 1.5\%$ under RCP8.5 (Figure 2a), whereas the AMA approach projects larger runoff increases of $5.0 \pm 0.7\%$ and $12.2 \pm 1.3\%$, respectively. Furthermore, such differences between BMA and AMA methods in future runoff change projections are significant over the globe, Northern Hemisphere (NH), Southern Hemisphere (SH), and tropics (25°S–25°N) under RCP8.5 (Wilcoxon signed-rank test, $P < 0.05$) but are only significant in tropical regions under the RCP4.5 scenario (Figure 2a).

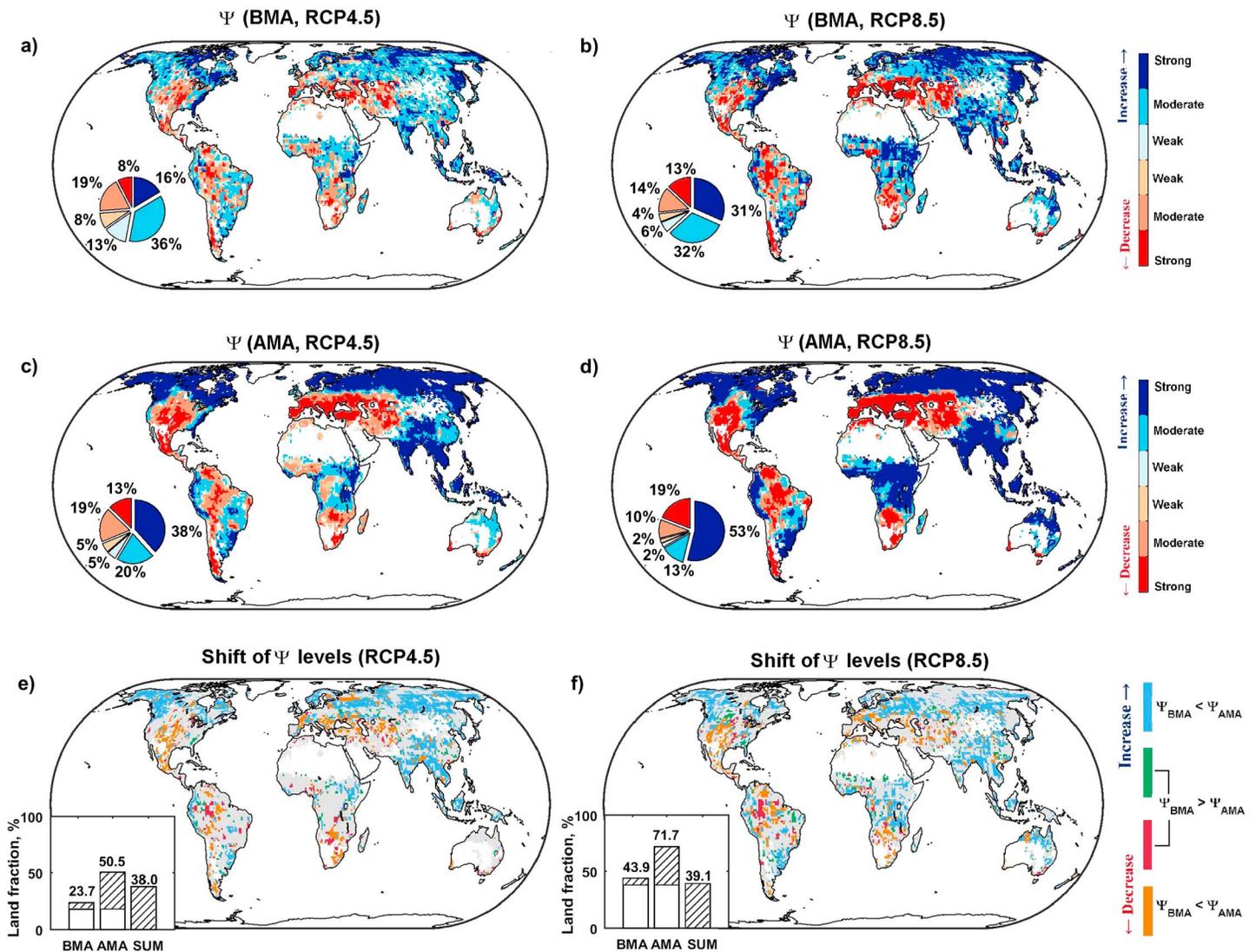


Figure 3. Patterns of robustness (ψ) of projected runoff changes. ψ for the runoff changes under RCP4.5 from (a) BMA model ensemble and (c) AMA model ensemble and (e) their difference. (b, d, and f) The same as for Figures 3a, 3c, and 3e but under RCP8.5. In Figures 3a–3d, the blue and red colors mark the regions experiencing increasing and decreasing runoff; the pie charts show the area fraction of the three robustness levels. In Figures 3e and 3f, the blue and orange colors mark the regions experiencing strong runoff changes in AMA ($\psi_{AMA} > 0.9$) but weak/moderate ones in BMA ($\psi_{BMA} < 0.9$). Conversely, the red and green mark the regions experiencing weak/moderate changes in AMA but strong changes in BMA. The gray color marks the regions where the levels of ψ remain unchanged. The bar plots demonstrate the fraction of land area where $\psi > 0.9$. The hatching shows the fraction of land areas with a ψ -level shift.

Regionally, the results of the BMA method (Figure 2b) indicate a significant increase of mean annual runoff ($P < 0.05$) under the RCP4.5 scenario in the Ganges and upper Brahmaputra river basins, Southeast Asia, and northern high-latitude basins. In contrast, runoff is projected to decrease in the Amazon Basin, sub-Saharan Africa, Central Asia, and Mediterranean region under the same RCP4.5 scenario (Figure 2b). Compared with results from the AMA method, runoff changes predicted by the BMA method are asymmetric with a more pronounced decrease in Amazonia and sub-Saharan Africa and a weaker increase in the high northern latitudes and Southeast and South Asia (Figure 2f). For example, for the Amazon Basin, a significant decrease in runoff under RCP4.5 ($P < 0.05$; Table S4) is found with BMA-based simulations, but not with AMA-based ones ($P = 0.56$). The magnitude of BMA projected future runoff decrease in Amazon Basin (-25.5 mm yr^{-1} in RCP4.5; -64.2 mm yr^{-1} in RCP8.5) is even larger than the intermodel standard deviation of projected runoff changes. For the high northern latitudes (e.g., Alaska and Siberia), the BMA and the AMA methods agree on the sign of future runoff changes but largely differ in their magnitudes. Three outlier models, GFDL-CM3 [Dunne et al., 2012], MIROC-ESM, and MIROC-ESM-CHEM [Sakamoto et al., 2012],

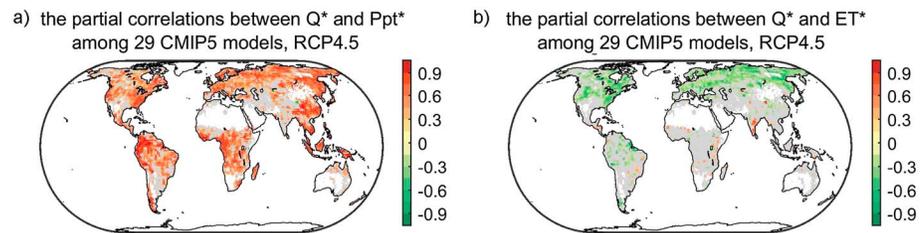


Figure 4. (a) Spatial pattern of partial correlation coefficients between relative runoff change (Q^*) from 2081–2100 to 1986–2005 and relative precipitation change (P_{pt}^*) of the same period for 29 CMIP5 models (except for CMCC-CMS and CESM1-WACCM) under RCP4.5. The corresponding relative changes in evapotranspiration (ET^*) are controlled for in the calculation. (b) Spatial pattern of partial correlation coefficients between Q^* and ET^* , while controlling for P_{pt}^* .

estimate a much higher increase in runoff ($197.5 \pm 15.3 \text{ mm yr}^{-1}$) over 2081–2100 than the other models ($120.4 \pm 30.1 \text{ mm yr}^{-1}$) in northern latitudes under RCP8.5 (Figure S2). These three models are assigned the same weight as all others in AMA but are given lower weights in BMA because of their poor skill in reproducing the UNH/GRDC data.

Following Knutti and Sedláček [2012], we use a ψ metric to diagnose the BMA-based or AMA-based runoff changes (analogous to “signal”) relative to the intermodel spread (analogous to “noise”). Following the definition in the IPCC AR5 [Collins et al., 2013], high values of ψ (>0.9) indicate a strong signal (larger than the noise) and $\psi < 0.5$ a weak signal. BMA projections under RCP4.5 display strong signal-to-noise ratios in the Amazon Basin, southern Chile, Mediterranean region, Quebec of Canada, Khatanga of Russia, Southeast Asia, South Asia, and Northeast China (marked as dark blue and dark red in Figure 3a), whereas signal-to-noise ratios are weak in central China, eastern Europe, Siberia, Alaska, the upper Mississippi, and Nelson basin (marked as light colors in Figure 3a). Overall, only 24% of the land area is diagnosed with a strong runoff change ($\psi > 0.9$) under RCP4.5 by BMA (Figure 3a), less than that by AMA (51%; Figure 3c). The values of ψ in the high northern latitudes, Southeast and South Asia, and midlatitude arid regions are high for the AMA, but not for the BMA ($\sim 33\%$ of the land area, marked in blue and orange in Figure 3e). By contrast, the values of ψ in the Amazon basin and sub-Saharan Africa are high for the BMA, but not for the AMA (covering $\sim 5\%$ land fractions; marked as green and red in Figure 3e). Similar shifts of ψ occur globally under the RCP8.5 scenario (Figures 3b, 3d, and 3f). In the high northern latitudes, shifts of ψ levels can be attributed to a weaker “signal” and stronger “noise” with BMA compared to AMA (Figure S14). In the Amazon Basin and Southeast Asia, BMA shows larger both signal and noise than AMA, with signal increasing faster than noise (Figure S14).

We then explore the source of the large intermodel differences in the future runoff projections. First, we calculate the relative change in runoff ($Q^* = Q_{\text{future}}/Q_{\text{historical}}$), precipitation ($P_{pt}^* = P_{pt \text{ future}}/P_{pt \text{ historical}}$), and evapotranspiration ($ET^* = ET_{\text{future}}/ET_{\text{historical}}$) from 29 CMIP5 models (except for CMCC-CMS and CESM1-WACCM where ET data are not available). We then apply partial correlation analyses to separate the influence of differences in P_{pt}^* and ET^* on the difference in Q^* across the 29 CMIP5 models at the grid cell. The results (Figures 4 and S15) suggest that Q^* and P_{pt}^* are positively correlated across most regions except in Central Asia, India, and Alaska, when ET^* is controlled for. P_{pt}^* may also influence ET^* . When controlling for the P_{pt}^* , we find a significant negative correlation between Q^* and ET^* at the high northern latitudes across models. It is worth noting that in these regions with highest warming rates, the large model spreads of temperature change projections may also result in large model differences in ET^* , which alone would lead to model differences in Q^* . Overall, our results suggest that the intermodel difference in precipitation changes is the prominent source of the difference in runoff changes. Model difference in evapotranspiration changes further makes important contributions to the intermodel uncertainty in model-projected runoff changes at the high northern latitudes. In addition, consistent with previous studies [Bring et al., 2015; Siam et al., 2013], we find the wide existence of nonzero terrestrial water storage changes (TWSCs) that differ across models (Figure S16). The nonzero TWSC suggests that water balance may not be conserved in many models, and the model difference in such water budget imbalance might be another source of model difference in the projected runoff changes. The exact cause(s) of the nonzero TWSC observed in models still remain an open question. Possible causes may include water leakage to deep underground, permafrost and glacier storage, stream-aquifer interactions, and groundwater dynamics [Clark et al., 2015].

In the most recent IPCC AR5, each Earth system model was assigned the same weight in multimodel ensemble runoff simulations (“one model-one vote”). However, results derived from this “democratic” approach should be treated with caution, given the large intermodel spread. Our approach, in which higher weights are given to models with better performances in reproducing past runoff, suggests a much lower increase in global runoff by the end of the 21st century, compared to the result of the traditional “democratic” approach. This weighted treatment is reasonable because models that are able to better reproduce historical runoff are usually based on verifiable physical processes [Flato *et al.*, 2013]. These physical processes that govern land surface hydrological dynamics are very likely to still function under future climate scenarios [Knutti *et al.*, 2017]. Therefore, although good model performance in the past is not a guarantee for the accuracy of runoff projections under future raised atmospheric greenhouse gas concentrations, one cannot expect that a poor-performed model in the past contains better representations of a set of key physical processes and would perform better than a well-performed model in the future runoff predictions. However, uncertainties in the projection of Bayesian weighted model ensemble are still high. This is because while the physical processes underlying hydrological dynamics are likely the same in the past versus in the future, some empirical parameters used for driving these processes and for training the models may go beyond their contemporary ranges. Nevertheless, our study provides a possible way to use historical observation to better constrain projections on future environmental changes and their potential impacts.

Also, our projections based on the Bayesian weighting scheme are only from the sense of statistical probability, but not based on process understanding. Process-based model evaluations would, in principle, be a useful approach from a mechanistic understanding perspective. However, because model behavior is determined by a large set of individual processes as well as by how these physical processes interlink together to produce the system-scale runoff response at large spatial scales [Clark *et al.*, 2011], there are three major challenges in process-based evaluations. First, such evaluation needs to pull together the multivariate and multiscale modeled and observed data from a wide range of hydroclimatic regimes. Although CMIP5 data set does provide a limited number of climatic variables, these are usually insufficient to fully understand what is happening within the governing equations for each land surface component. Moreover, even if a number of data streams are currently available, those data are also known to have issues like uncertainty in observations and the incommensurability between models and measurements. Second, it is difficult to devise meaningful ways (e.g., diagnostic metrics) for evaluation. In particular, a key problem is how to evaluate the impacts of interlinked physical processes on system-scale responses at large scale [Beven and Cloke, 2012; Wood *et al.*, 2012]. Third, it also requires substantial efforts (and even debates) on how to aggregate the multivariate and multiscale model evaluations and to use it for long-term projections. Nonetheless, further work is highly desired to better understand and improve the representation of relevant biophysical processes in model evaluation and projection. Before that, the statistical approach using Bayesian model ensemble will continue to provide us a useful tool in constraining multimodel runoff projections with observations.

4. Conclusion

Recently, climate scientists involved in the IPCC reports have been committed to reducing the uncertainty in projections of future river runoff. Benchmarking Earth system models provides information about confidence in their projections of future river runoff [Hoffman *et al.*, 2009; Clark *et al.*, 2015]. Our BMA-based projections of runoff change suggest a more pronounced decrease in Amazon and sub-Saharan Africa and a weaker increase in the high northern latitudes than those in the IPCC assessment (based on “model democracy”). This significantly different spatial pattern of future runoff, which emerges from adding the constraint of present-day observations, has considerable regional-scale consequences for impact assessment studies [Warszawski *et al.*, 2014; Schewe *et al.*, 2014; Prudhomme *et al.*, 2014]. However, it still remains an open question if the BMA projection of future runoff changes is more realistic than AMA. In other words, shall models that reproduce current runoff be trusted more in predicting the future when environmental conditions go beyond contemporary range of change? Furthermore, in addition to intermodel differences in structures and processes, natural variability in the climate system is also the main cause for the uncertainty of model projections [Deser *et al.*, 2012]. Further work on the role of internal variability in future runoff change, which is not included in this current study, is warranted to help reduce the uncertainty of regional and global runoff projections. Improving the representation of key processes, such as groundwater dynamics, stream-aquifer

interaction, and root water update, in land models needs to be a priority for the Earth system modeling community [Maxwell et al., 2014; Clark et al., 2015; Fatichi et al., 2016].

Acknowledgments

This study was supported by the National Natural Science Foundation of China (41561134016, 41530528, and 41671464), the 111 project, and the National Youth Top-notch Talent Support Program in China. We acknowledge the World Climate Research Program's (WCRP) Working Group on Coupled Modelling (WGCM), which is responsible for CMIP, and we thank the climate modeling groups for producing and making their model output available. CMIP5 ESM outputs to support this article are from the protocol of Coupled Model Intercomparison Project Phase 5 (CMIP5; <http://browse.ceda.ac.uk/browse/badc/cmip5/>). The observed river runoff data and discharge data are available from the Global Composite Runoff Fields (UNH/GRDC GCRF; <http://www.grdc.sr.unh.edu/index.html>) and the Global River Discharge Center (GRDC; <http://www.bafg.de/GRDC/>), respectively. Precipitation data are available from the UCAR Climate Data Guide (<http://climatedataguide.ucar.edu/>).

References

- Adler, R. F., et al. (2003), The version-2 Global Precipitation Climatology Project (GPCP) monthly precipitation analysis (1979–present), *J. Hydrometeorol.*, 4(6), 1147–1167, doi:10.1175/1525-7541(2003)004<1147:TVGPCP>2.0.CO;2.
- Alcamo, J., M. Flörke, and M. Märker (2007), Future long-term changes in global water resources driven by socio-economic and climatic changes, *Hydrol. Sci. J.*, 52(2), 247–275, doi:10.1623/hysj.52.2.247.
- Barkhordarian, A., H. von Storch, and J. Bhend (2013), The expectation of future precipitation change over the Mediterranean region is different from what we observe, *Clim. Dyn.*, 40(1–2), 225–244.
- Beven, K. J., and H. L. Cloke (2012), Comment on “Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth’s terrestrial water” by Eric F. Wood et al, *Water Resour. Res.*, 48, W01801, doi:10.1029/2011WR010982.
- 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.
- Bring, A., et al. (2015), Implications of freshwater flux data from the CMIP5 multimodel output across a set of Northern Hemisphere drainage basins, *Earths Future*, 3(6), 206–217.
- Chen, M., P. Xie, J. E. Janowiak, and P. A. Arkin (2002), Global land precipitation: A 50-yr monthly analysis based on gauge observations, *J. Hydrometeorol.*, 3(3), 249–266, doi:10.1175/1525-7541(2002)003<0249:GLPAYM>2.0.CO;2.
- Chen, M., W. Shi, P. Xie, V. B. Silva, V. E. Kousky, R. W. Higgins, and J. E. Janowiak (2008), Assessing objective techniques for gauge-based analyses of global daily precipitation, *J. Geophys. Res.*, 113, D04110, doi:10.1029/2007JD009132.
- Chylek, P., J. Li, M. K. Dubey, M. Wang, and G. Lesins (2011), Observed and model simulated 20th century Arctic temperature variability: Canadian Earth system model CanESM2, *Atmos. Chem. Phys. Discuss.*, 11(8), 22,893–22,907.
- Cisneros, B. E., et al. (2014), Freshwater resources, in *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by C. B. Field et al., pp. 229–269, Cambridge Univ. Press, Cambridge, U. K., and New York.
- Clark, M. P., D. Kavetski, and F. Fenicia (2011), Pursuing the method of multiple working hypotheses for hydrological modeling, *Water Resour. Res.*, 47, W09301, doi:10.1029/2010WR009827.
- Clark, M. P., et al. (2015), Improving the representation of hydrologic processes in Earth system models, *Water Resour. Res.*, 51, 5929–5956, doi:10.1002/2015WR017096.
- Collins, M., et al. (2013), Long-term climate change: Projections, commitments and irreversibility, 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. 1029–1136, Cambridge Univ. Press, Cambridge, U. K., and New York.
- 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.
- Deser, C., R. Knutti, S. Solomon, and A. S. Phillips (2012), Communication of the role of natural variability in future North American climate, *Nat. Clim. Change*, 2(11), 775–779.
- Dufresne, J. L., et al. (2013), Climate change projections using the IPSL-CM5 Earth system model: From CMIP3 to CMIP5, *Clim. Dyn.*, 40(9), 2123–2165, doi:10.1007/s00382-012-1636-1.
- Dunne, J. P., et al. (2012), GFDL’s ESM2 global coupled climate-carbon Earth system models. Part I: Physical formulation and baseline simulation characteristics, *J. Clim.*, 25(19), 6646–6665, doi:10.1175/JCLI-D-11-00560.1.
- Fatichi, S., et al. (2016), An overview of current applications, challenges, and future trends in distributed process-based models in hydrology, *J. Hydrol.*, 537, 45–60, doi:10.1016/j.jhydrol.2016.03.026.
- Fekete, B., C. J. Vorosmarty, and W. Grabs (2002), High-resolution fields of global runoff combining observed river discharge and simulated water balances, *Global Biogeochem. Cycles*, 16(3), 1042, doi:10.1029/1999GB001254.
- Fekete, B., and C. J. Vorosmarty (2011), ISLSCP II UNH/GRDC Composite Monthly Runoff, in *ISLSCP Initiative II Collection*, edited by F. G. Hall et al., Data set, available at <http://daac.ornl.gov/>, from Oak Ridge Natl. Lab., Distributed Active Archive Center, Oak Ridge, Tenn., doi:10.3334/ORNLAAC/994.
- 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. 1029–1136, Cambridge Univ. Press, Cambridge, U. K., and New York.
- Gent, P. R., et al. (2011), The community climate system model version 4, *J. Clim.*, 24(19), 4973–4991.
- 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.
- Hall, F. G., et al. (2006), ISLSCP Initiative II global data sets: Surface boundary conditions and atmospheric forcings for land-atmosphere studies, *J. Geophys. Res.*, 111, D22501, doi:10.1029/2006JD007366.
- Hall, J. W., et al. (2014), Coping with the curse of freshwater variability, *Science*, 346(6208), 429–430, doi:10.1126/science.1257890.
- Harris, I., P. D. Jones, T. J. Osborn, and D. H. Lister (2014), Updated high-resolution grids of monthly climatic observations—the CRU TS3.10 dataset, *Int. J. Climatol.*, 34(3), 623–642, doi:10.1002/joc.3711.
- Hoffman, F. M., et al. (2009), The Carbon-Land Model Intercomparison Project (C-LAMP): A prototype for coupled biosphere-atmosphere model benchmarking for the IPCC Fifth Assessment Report (AR5), *Mon. Weather Rev.*, 127(9), 2204–2210.
- Hurrell, J. W., et al. (2013), The community Earth system model: A framework for collaborative research, *Bull. Am. Meteorol. Soc.*, 94(9), 1339–1360.
- Jones, P. W. (1999), First- and second-order conservative remapping schemes for grids in spherical coordinates, *Mon. Weather Rev.*, 127(9), 2204–2210.
- Jungclaus, J. H., et al. (2010), Climate and carbon-cycle variability over the last millennium, *Clim. Past*, 6(5), 723–737, doi:10.5194/cp-6-723-2010.
- Knutti, R. (2010), The end of model democracy?, *Clim. Change*, 102(3–4), 395–404, doi:10.1007/s10584-010-9800-2.
- Knutti, R., and J. Sedláček (2012), Robustness and uncertainties in the new CMIP5 climate model projections, *Nat. Clim. Change*, 3(4), 369–373, doi:10.1038/nclimate1716.
- Knutti, R., J. Sedláček, B. M. Sanderson, R. Lorenz, E. M. Fischer, and V. Eyring (2017), A climate model projection weighting scheme accounting for performance and interdependence, *Geophys. Res. Lett.*, 44, 1909–1918, doi:10.1002/2016GL072012.

- Legates, D., and C. Willmott (1990), Mean seasonal and spatial variability in gauge-corrected, global precipitation, *Int. J. Climatol.*, *10*(2), 111–127, doi:10.1002/joc.3370100202.
- Li, L., et al. (2013), The flexible global ocean-atmosphere-land system model, Grid-point Version 2: FGOALS-g2, *Adv. Atmos. Sci.*, *30*, 543–560.
- Maxwell, R. M., et al. (2014), Surface-subsurface model intercomparison: A first set of benchmark results to diagnose integrated hydrology and feedbacks, *Water Resour. Res.*, *50*, 1531–1549, doi:10.1002/2013WR013725.
- 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:CSRFC>2.0.CO;2.
- 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.
- Mochizuki, T., et al. (2012), Decadal prediction using a recent series of MIROC global climate models, *J. Meteorol. Soc. Jpn.*, *90*, 373–383, doi:10.2151/jmsj.2012-A22.
- Moss, R. H., et al. (2010), The next generation of scenarios for climate change research and assessment, *Nature*, *463*(7282), 747–756.
- Piao, S., P. Friedlingstein, P. Ciais, N. de Noblet-Ducoudré, D. Labat, and S. Zaehle (2007), Changes in climate and land use have a larger direct impact than rising CO₂ on global river runoff trends, *Proc. Natl. Acad. Sci. U.S.A.*, *104*(39), 15,242–15,247, doi:10.1073/pnas.0707213104.
- Prudhomme, C., et al. (2014), Hydrological droughts in the 21st century, hotspots and uncertainties from a global multimodel ensemble experiment, *Proc. Natl. Acad. Sci. U.S.A.*, *111*(9), 3262–3267, doi:10.1073/pnas.1222473110.
- Qiao, F., Z. Song, and Y. Bao (2013), FIO-ESM model output prepared for CMIP5 historical, served by ESGF, World Data Center for Climate (WDCC).
- Raftery, A. E., T. Gneiting, F. Balabdaoui, and M. Polakowski (2005), Using Bayesian model averaging to calibrate forecast ensembles, *Mon. Weather Rev.*, *133*(5), 1155–1174, doi:10.1175/MWR2906.1.
- Rotstayn, L. D., S. J. Jeffrey, M. A. Collier, S. M. Dravitzki, A. C. Hirst, J. I. Suktus, and K. K. Wong (2012), Aerosol and greenhouse gas-induced changes in summer rainfall and circulation in the Australasian region: A study using single-forcing climate simulations, *Atmos. Chem. Phys.*, *12*(14), 6377–6404, doi:10.5194/acp-12-6377-2012.
- Rudolf, B., C. Beck, J. Grieser, and U. Schneider (2005), Global precipitation analysis products, Global Precipitation Climatology Centre (GPCC), DWD, Internet Publication: 1–8.
- Sakamoto, T. T., et al. (2012), MIROC4h—A new high-resolution atmosphere-ocean coupled general circulation model, *J. Meteorol. Soc. Jpn. Ser. II*, *90*(3), 325–359, doi:10.2151/jmsj.2012-301.
- Schewe, J., et al. (2014), Multimodel assessment of water scarcity under climate change, *Proc. Natl. Acad. Sci. U.S.A.*, *111*(9), 3245–3250, doi:10.1073/pnas.1222460110.
- Schmidt, G. A., et al. (2006), Present-day atmospheric simulations using GISS model E: Comparison to in situ, satellite, and reanalysis data, *J. Clim.*, *19*(2), 153–192, doi:10.1175/JCLI3612.1.
- Scoccimarro, E., et al. (2011), Effects of tropical cyclones on ocean heat transport in a high-resolution coupled general circulation model, *J. Clim.*, *24*(16), 4368–4384, doi:10.1175/2011JCLI4104.1.
- Siam, M. S., M. E. Demory, and E. A. Eltahir (2013), Hydrological cycles over the Congo and Upper Blue Nile Basins: Evaluation of general circulation model simulations and reanalysis products, *J. Clim.*, *26*(22), 8881–8894.
- Taylor, K. E., R. J. Stouffer, and G. A. Meehl (2012), An overview of CMIP5 and the experiment design, *Bull. Am. Meteorol. Soc.*, *93*(4), 485–498, doi:10.1175/BAMS-D-11-00094.1.
- Tebaldi, C., R. L. Smith, D. Nychka, and L. O. Mearns (2005), Quantifying uncertainty in projections of regional climate change: A Bayesian approach to the analysis of multimodel ensembles, *J. Clim.*, *18*(10), 1524.
- Tjiputra, J. F., et al. (2013), Evaluation of the carbon cycle components in the Norwegian Earth system model (NorESM), *Geosci. Model Dev.*, *6*(2), 301–325, doi:10.5194/gmd-6-301-2013.
- Voldoire, A., et al. (2013), The CNRM-CM5.1 global climate model: Description and basic evaluation, *Clim. Dyn.*, *40*(9–10), 2091–2121.
- Volodin, E. M., N. A. Dianskii, and A. V. Gusev (2010), Simulating present-day climate with the INMCM4.0 coupled model of the atmospheric and oceanic general circulations, *Izv. Atmos. Ocean. Phys.*, *46*(4), 414–431.
- Vörösmarty, C. J., P. Green, J. Salisbury, and R. B. Lammers (2000), Global water resources: Vulnerability from climate change and population growth, *Science*, *289*(5477), 284–288, doi:10.1126/science.289.5477.284.
- Vrugt, J. A., C. G. Diks, and M. P. Clark (2008), Ensemble Bayesian model averaging using Markov chain Monte Carlo sampling, *Environ. Fluid Mech.*, *8*(5), 579–595, doi:10.1007/s10652-008-9106-3.
- Warszawski, L., K. Frieler, V. Huber, F. Piontek, O. Serdeczny, and J. Schewe (2014), The Inter-Sectoral Impact Model Intercomparison Project (ISI-MIP): Project framework, *Proc. Natl. Acad. Sci. U.S.A.*, *111*(9), 3228–3232, doi:10.1073/pnas.1312330110.
- Wei, T., et al. (2012), Developed and developing world responsibilities for historical climate change and CO₂ mitigation, *Proc. Natl. Acad. Sci. U.S.A.*, *109*(32), 12,911–12,915, doi:10.1073/pnas.1203282109.
- Weigel, A. P., M. A. Liniger, and C. Appenzeller (2007), The discrete brier and ranked probability skill scores, *Mon. Weather Rev.*, *135*(1), 118–124.
- Wood, E. F., et al. (2012), Reply to comment by Keith J. Beven and Hannah L. Cloke on “Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth’s terrestrial water”, *Water Resour. Res.*, *48*, W01802, doi:10.1029/2011WR011202.
- Wu, T. (2012), A mass-flux cumulus parameterization scheme for large-scale models: Description and test with observations, *Clim. Dyn.*, *38*(3–4), 725–744.
- Yukimoto, S., et al. (2011), A new global climate model of the meteorological research institute: MRI-CGCM3-model description and basic performance, *J. Meteorol. Soc. Jpn.*, *90A*, 23–64, doi:10.2151/jmsj.2012-A02.
- Zeng, Z., T. Wang, F. Zhou, P. Ciais, J. Mao, X. Shi, and S. Piao (2014), A worldwide analysis of spatiotemporal changes in water balance-based evapotranspiration from 1982 to 2009, *J. Geophys. Res. Atmos.*, *119*, 1186–1202, doi:10.1002/2013JD020941.