Journal of Hydrology 470–471 (2012) 269–279

Contents lists available at SciVerse ScienceDirect

Journal of Hydrology

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

Benchmarking global land surface models against the observed mean annual runoff from 150 large basins

Xinyao Zhou [a,b,c], Yongqiang Zhang [b,*], Yingping Wang [d], Huqiang Zhang [e], Jai Vaze [b], Lu Zhang [b], Yonghui Yang [a], Yanchun Zhou [b,f]

[a] Key Laboratory of Agricultural Water Resources, Center for Agricultural Resources Research, The Chinese Academy of Sciences, Shijiazhuang 050021, China
[b] CSIRO Water for a Healthy Country National Research Flagship, CSIRO Land and Water, PO Box 1666, Canberra, ACT 2601, Australia
[c] Graduate University of the Chinese Academy of Sciences, Beijing, China
[d] CSIRO Marine and Atmospheric Research, Aspendale, Victoria 3195, Australia
[e] Centre for Australian Weather and Climate Research, A Partnership between Bureau of Meteorology & CSIRO, Melbourne, Victoria, Austria
[f] Faculty of Infrastructure Engineering, Dalian University of Technology, Dalian, China

ARTICLE INFO

Article history:
Available online 8 September 2012
This manuscript was handled by Konstantine P. Georgakakos, Editor-in-Chief, with the assistance of Aiguo Dai, Associate Editor

Keywords:
Mean annual runoff
Land surface models
Global
Hydrology
Precipitation

SUMMARY

Using the observed mean annual runoff for 1986–1995 from 150 large basins globally, we evaluate the performance of the 14 global land surface models (LSMs) and six Budyko-type hydrological models that are forced by the meteorological data from the second phase of the Global Soil Wetness Project (GSWP-2). The results show that both the 14 LSMs and six Budyko-type models can explain 55–70% of the spatial variations of mean annual runoff across the selected 150 basins. However, the 14 LSMs show larger biases in the simulated mean annual runoff than the Budyko-type models. The LSMs biases are caused by errors in forcing data, model structure and model parameterisation. The errors in the precipitation forcing data are found to be the main cause for biases in the simulated mean annual runoffs by the Budyko-types models, and most likely for biases in the 14 global land surface models too. The GSWP-2 precipitation is noticeably overestimated at Northern high-latitudes, which causes large positive biases for the LSMs in simulating mean annual runoff in these regions. The most LSMs show large biases in the regions with low mean annual precipitation. Underestimation of the GSWP-2 precipitation in Amazon and Orinoco basins results in significant underestimation in the simulated mean annual runoff by all LSMs and Budyko-type models for these regions. The LSMs with smaller biases generally show larger baseflow ratio in wet basins than in dry basins while the LSMs with larger biases generally show smaller baseflow ratio in wet basins than in dry basins. This indicates that errors in model structure can result in large biases in the simulated runoff. Several parameter sensitivity experiments for one LSM are carried out to investigate impacts on simulated mean runoff. The result indicates that ±20% changes in five key model parameters have relatively smaller impacts on the simulated mean annual runoff across the 150 basins, compared to errors in model structure.

1. Introduction

Surface runoff is one of the most important global water balance variables because it directly influences surface water availability and consequently human lives and activities. Numerous studies have used a variety of models, including global climate models coupled with land-surface models (LSMs), to estimate the spatial and temporal patterns of global runoff and its responses to climate change (Fung et al., 2011; IPCC, 2007; Milly et al., 2005; Nohara et al., 2006). Together with different types of hydrological models, LSMs are widely used for predicting surface water flux at regional or global scales under present or future climate conditions (Hurkmans et al., 2007; Levis et al., 1996; Sheffield et al., 2004).

Several studies have shown that LSMs can estimate runoff reasonably well for some selected large basins. For example, Lohmann et al. (1998) compared the simulated runoff, evapotranspiration, and soil moisture storage changes from 16 LSMs in the Project for Intercomparison of Land-surface Parameterization Schemes (PILPS) to the observations in Red-Arkansas River basin from 1980 to 1986. They found that all 16 models captured the spatial variability of mean annual runoff reasonably well. However, they did not analyse model biases and their potential causes. Gusev et al. (2008) investigated the ability of Soil–Water–Atmosphere–Plants (SWAP) model to simulate runoff in Mezen River basin from 1986 to 1995. They found that the SWAP model simulated monthly runoff reasonably well with a bias of 0.1% and the Nash and Sutc-

* Corresponding author. Tel.: +61 2 6246 5761; fax: +61 2 6246 5800.
E-mail address: yongqiang.zhang@csiro.au (Y. Zhang).

0022-1694/$ - see front matter
http://dx.doi.org/10.1016/j.jhydrol.2012.09.002

liffe Efficiency of monthly runoff of 0.94 using the Second Global Soil Wetness Project (GSWP-2) precipitation. Feng and Houser (2008) investigated the performance of three LSMs using the GSWP-2 forcings in the Mississippi River basin against observed data for 1986–1995. They found that the Noah and SSiB models overestimated annual runoff while the simulated runoff by the Community Land Model version 2 with TOPMODEL-based runoff scheme (CLM2-TOP) model was close to the gauge measurements. Zaitchik et al. (2010) evaluated the Global Land Data Assimilation System (GLDAS) by comparing simulated mean annual, daily, mean annual peak daily and interannual runoff from four LSMs to observations in global 66 basins over the period 1979–2007. Their study concluded that the Community Land Model (CLM) was the best model for estimating mean annual runoff while the Variable Infiltration Capacity (VIC) model was the most accurate for estimating the timing of peak flows. However, most of these studies only conducted evaluations for a limited number of LSMs and/or used a limited number of basins.

Outputs of land surface energy and water fluxes as simulated by several LSMs are available from a number of recent international projects, such as the GSWP-2. However the skills and uncertainties of those predictions are rarely evaluated against observations. To gain confidence in these predictions from the LSMs, their performance under present climate conditions needs to be thoroughly evaluated, including their ability in reproducing the observed hydrological responses at different temporal and spatial scales and their performance against some commonly used and simple hydrological models, such as Budyko-type hydrological models (Adams et al., 2009; Budyko, 1948; Fu, 1981; OL'Dekop, 1911; Pike, 1964; Schreiber, 1904; Turc, 1954; Zhang et al., 2001, 2012). The present study is designed to evaluate the performance of 14 LSMs used in the GSWP-2 over a large number of river basins across the globe. We use water fluxes simulated by 13 LSMs at a 1.0° spatial resolution, driven by the GSWP-2 forcings from 1986 to 1995, which is available as part of the GSWP2 study (Dirmeyer et al., 2006), and the simulations by the Community Atmosphere Biosphere Land Exchange (CABLE) model at the same spatial resolution over the same period (Dirmeyer et al., 2006). We then assess the skills of each LSM in reproducing the observed mean annual runoff.

One unique feature of this study is that we include a number of simple Budyko-type models in the analysis to assess if the process-based LSMs outperform the simple models for simulating regional and global water budgets. Budyko-type hydrological models (Adams et al., 2009; Teng et al., 2012; Zhang et al., 2012) are widely used in hydrology community for estimating long-term mean annual runoff from small catchments to large basins. The Budyko framework assumes that the water storage change in a catchment or basin is negligible and available energy and water are the primary factors controlling mean annual actual evaporation and runoff over a long time period. Thus, runoff coefficient can be estimated as a non-parametric function of aridity index. Several simple Budyko-type models are used in this study to evaluate the performances of LSMs.

The main objectives in this study include (1) benchmarking the LSMs against observed mean annual runoff from 150 large river basins under different climatic conditions; (2) comparing the performance of the LSMs and the simple Budyko-type models; (3) investigating potential causes for the biases in the simulated runoff by the LSMs.

2. Data and methods

2.1. Data

Dai et al. (2009) collated a comprehensive dataset of historical monthly streamflow at the farthest downstream stations for the world's 925 largest river basins. These data cover ~80% of global ocean-draining land area and most of the stations have relatively complete record for the period of 1948–2004. Observed monthly runoff data (Q_{obs}) for 150 large basins (>10,000 km^2) as collated by Dai et al. (2009) were summed from 1986 to 1995 to obtain the observed mean annual runoff. It was compared to the simulated mean annual runoff obtained from each LSM and Budyko-type model and from the same period (1986–1995) (Fig. 1a). The small basins (<10,000 km^2) were excluded because of the relative coarse resolution of model simulations (i.e., $1^\circ \times 1^\circ$). Each selected basin has at least 5-year runoff observations during the period from 1986 to 1995. Basin boundaries were provided by GRDC (Global Runoff Data Centre) (http://www.gewex.org/grdc.html) for calculating basin-average mean annual runoff obtained from each LSM and Budyko-type model. Some of the 150 basins were regulated by dams and/or reservoirs and therefore the temporal variations of the observed runoff from those basins were not comparable with the model simulations. However, dam/reservoir regulation mainly influences seasonal partition of runoff, but not observed mean annual runoff. Therefore, only the observed mean annual runoffs for the selected 150 basins were compared with the model simulations from 1986 to 1995.

The GSWP-2 is an ongoing environmental modelling research activity of the Global Land–Atmosphere System Study (GLASS) and the International Satellite Land-Surface Climatology Project (ISLSCP). Its objectives are to provide large-scale validation and quality check of the ISLSCP data sets and to provide the best global estimates of land surface state variables and fluxes (Zhao and Dirmeyer, 2003). The $1^\circ \times 1^\circ$ GSWP-2 atmospheric forcing data based on the regridded National Centers for Environmental Prediction/Department of Energy (NCEP/DOE) reanalysis were used in the ISLSCP initiative II. Further details could be found at the GSWP-2 website (http://www.iges.org/gswp/). The 14 LSMs (listed in Table 1) were all driven by the same forcing data. The model outputs for the 13 LSMs, except the CABLE model were sourced from the Center for Ocean–Land–Atmosphere Studies (COLA) group. We used the GSWP-2 forcing data to drive the CABLE model for a series of sensitivity experiments. We also used the GSWP-2 mean annual precipitation to drive the Budyko-type models.

To assess the uncertainty in the GSWP-2 precipitation, we regridded the resolution of the monthly precipitation (P_{GPCC}) data in the Global Precipitation Climatology Centre (GPCC, version 4, (Rudolf and Schneider, 2004)) from $0.5^\circ \times 0.5^\circ$ to $1^\circ \times 1^\circ$ resolution. Because the GPCC precipitation data were interpolated from the most complete gauge measurements over 64,000 stations globally (ftp://ftp-anon.dwd.de/pub/data/gpcc/html/fulldata_download.html), it is considered as a reliable global product for precipitation for most regions, but the accuracy may become quite low for regions with only sparse observations, such as at the Northern high-latitudes and Amazon basin (Rudolf and Schneider, 2004). The GSWP-2 precipitation is a hybrid product of reanalysis (NCEP/DOE) and observations (GPCC and CRU). Although GSWP-2 precipitation data were corrected using the GPCC precipitation data, significant differences still remain between the two products. This is because: (1) NCEP precipitation is higher than GPCC precipitation in most parts of the world (Fekete et al., 2004); (2) the hybrid precipitation product is corrected using Motoya's wind correction and (3) for the regions where gauge density is low, the Global Precipitation Climatology Project (GPCP) product is blended in (Zhao and Dirmeyer, 2003). To further assess the errors in the simulated runoff from the errors in precipitation for the regions where more accurate precipitation data are available, such as the Murray-Darling basin in Australia, we also use the local precipitation data from the Australian Bureau of Meteorology to drive the Budyko-type models for the region. The Budyko-type models need