Evaluating Global Streamflow Simulations by a Physically Based Routing Model Coupled with the Community Land Model

HONG-YI LI AND L. RUBY LEUNG
Pacific Northwest National Laboratory, Richland, Washington

AUGUSTO GETIRANA
NASA Goddard Space Flight Center, Greenbelt, and Earth System Science Interdisciplinary Center, University of Maryland, College Park, College Park, Maryland

MAOYI HUANG
Pacific Northwest National Laboratory, Richland, Washington

HUAN WU
NASA Goddard Space Flight Center, Greenbelt, and Earth System Science Interdisciplinary Center, University of Maryland, College Park, College Park, Maryland

YUBIN XU, JIALI GUO, AND NATHALIE VOISIN
Pacific Northwest National Laboratory, Richland, Washington

(Manuscript received 21 April 2014, in final form 18 November 2014)

ABSTRACT

Accurately simulating hydrological processes such as streamflow is important in land surface modeling because they can influence other land surface processes, such as carbon cycle dynamics, through various interaction pathways. This study aims to evaluate the global application of a recently developed Model for Scale Adaptive River Transport (MOSART) coupled with the Community Land Model, version 4 (CLM4). To support the global implementation of MOSART, a comprehensive global hydrography dataset has been derived at multiple resolutions from different sources. The simulated runoff fields are first evaluated against the composite runoff map from the Global Runoff Data Centre (GRDC). The simulated streamflow is then shown to reproduce reasonably well the observed daily and monthly streamflow at over 1600 of the world’s major river stations in terms of annual, seasonal, and daily flow statistics. The impacts of model structure complexity are evaluated, and results show that the spatial and temporal variability of river velocity simulated by MOSART is necessary for capturing streamflow seasonality and annual maximum flood. Other sources of the simulation bias include uncertainties in the atmospheric forcing, as revealed by simulations driven by four different climate datasets, and human influences, based on a classification framework that quantifies the impact levels of large dams on the streamflow worldwide.

1. Introduction

The water cycle plays a central role in regional and global climate systems. Although the terrestrial water storage is small compared to the ocean, it has undergone substantial changes because of human activities such as land use/land cover change (Gordon et al. 2005) and groundwater use (Famiglietti et al. 2011). Through land–atmosphere interactions, changes in the terrestrial water cycle can have important influences on regional and global climate (e.g., Seneviratne et al. 2010; Taylor et al. 2012). Thus, the terrestrial water cycle has received increasing attention in the climate modeling community (Lettenmaier and Famiglietti 2006; Trenberth et al. 2007; Getirana et al. 2014), with a particular emphasis on
modeling evapotranspiration. Modeling soil hydrological processes, including runoff generation and surface–groundwater interactions, have also been more broadly addressed in recent studies (Koster et al. 2000; Miguez-Macho and Fan 2012a,b; Leung et al. 2011; Li et al. 2011; Zhou et al. 2012).

Although a small component in the global annual water cycle compared to precipitation and evapotranspiration, lateral routing and transport of surface water through the terrestrial drainage systems is vital to connecting the coupled atmosphere–land–ocean system. Through freshwater discharge, sediment, carbon and nitrogen fluxes to the ocean, and outgassing of CO₂ to the atmosphere, river transport plays an important role in the water and biogeochemical cycles of the coupled earth system (Butman and Raymond 2011; Coles et al. 2013). Surface water transport is also closely linked to human society, as water resources are managed to balance supply and demand through regulations of streamflow. Understanding and predicting natural hazards such as flooding and their hydrological and ecological consequences requires improved understanding and modeling of surface water dynamics through the drainage systems. River routing is also important for evaluating or calibrating hydrologic parameterizations because streamflow data are more widely available than observations of other hydrologic states or fluxes. For this purpose, minimizing uncertainty introduced by river routing would enable a more effective use of streamflow data to improve hydrologic modeling (Nijssen et al. 2001; Decharme et al. 2012; Yamazaki et al. 2011; Getirana et al. 2012; Li et al. 2013; Wu et al. 2014).

Previous analyses of global hydrology and global discharge into the oceans have used mostly simplified routing models, including simple annual discharge or unit hydrographs (Nijssen et al. 2001; Branstetter and Erickson 2003) to estimate water balance components at annual scales. For global modeling of floods and related hazards like surge, inundation, and sediment and debris transport and for representing the influence of water resources management, river transport must be simulated at finer spatial and temporal scales with adequate model complexity (Yamazaki et al. 2011; Getirana et al. 2012, 2013, 2014; Wu et al. 2014). Model complexity can be defined in many ways, and here we refer to the different aspects (e.g., spatial and temporal variability) of channel velocity being modeled, since it is one of the most important and direct controlling factors of flow rate in the streams. Channel velocity itself is affected by many factors, such as channel water storage, topography, and channel geometry. The Model for Scale Adaptive River Transport (MOSART) is a recently developed large-scale routing model with scale adaptive treatments of within- and between-grid routing processes. MOSART uses the kinematic wave equation for subgrid surface runoff and channel routing and for routing channel flow from one grid cell to another. Li et al. (2013) evaluated MOSART at the Columbia River basin in the U.S. Pacific Northwest, where excellent observation data are available. They showed that, under natural conditions, MOSART is capable of capturing the observed streamflow and river velocity quite well.

In this study, MOSART has been coupled with the Community Land Model, version 4 (CLM4), in the same manner as was used in coupling CLM4 to an existing runoff routing module, the River Transport Model (RTM; Branstetter and Erickson 2003; Oleson et al. 2010). The purpose of the current study is twofold. First, we aim to evaluate the impacts of the model complexity introduced in MOSART compared to the simpler approach adopted by RTM in a stepwise manner. This extends the limited comparison performed by Li et al. (2013) over the Columbia River basin to large river basins worldwide and adopts a more systematic framework to quantify the impacts of the spatial and temporal variability modeled by MOSART. Second, we aim to comprehensively evaluate the global implementation of the coupled CLM4–MOSART against streamflow observations.

Evaluation of large-scale river routing models is a challenging task because of unavoidable uncertainties from various sources (Widen-Nilsson et al. 2007; Dezetter et al. 2008; Pappenberger et al. 2010; Getirana et al. 2013): 1) uncertainty in the climatic forcings driving the land surface models, particularly related to precipitation and radiation, which affect runoff generation; 2) uncertainty from the runoff generation schemes used in the land surface models (model structure and parameters); 3) uncertainty from the routing model parameters, such as the preservation of the drainage boundaries and flow paths; and 4) impacts of human influences on streamflow, such as reservoir operation and irrigation. Runoff simulation bias due to the runoff generation scheme in CLM4 has been systematically discussed in our previous studies (Li et al. 2011; Hou et al. 2012; Huang et al. 2013), so it will not be included in this study. The uncertainty due to routing model parameters has been largely discussed in Wu et al. (2012) and Getirana et al. (2012) and therefore will not be included either. In this study, we investigate the uncertainties in streamflow simulations due to uncertainty in the climate forcing. We also investigate the impacts of model structure complexity and model skill in river basins that have varying degrees of human influences. The experimental approach to evaluate the impacts of model complexity is to perform multiple simulations using MOSART by reducing its complexities one at a time. To address the forcing uncertainties, we use
four different climate forcing datasets. Finally, we quantify the effect of human activities on the observed flow (extraction, regulation, and transfers) through statistical analysis to evaluate the biases between the simulated natural flow and the observed impounded flow in basins that have different degrees of human influence.

Section 2 introduces the routing model, MOSART, its coupling to CLM4, and its supporting global hydrography dataset. Section 3 describes the experiment design, the observation dataset, and evaluation metrics. Section 4 presents the results and discussion with respect to the routing model complexity, climate forcing, and human influences. Section 5 closes with a summary and remarks.

2. Model and datasets

a. MOSART

MOSART is a physically based runoff routing model designed for applications across watershed, regional, and global scales with relatively consistent performance at different resolutions. Each spatial unit can be conceptually divided into a single main channel that flows through the whole unit and connects it with other units, a number of tributaries that discharge into the main channel only without flowing out of the unit, and a number of hillslopes that discharge into these tributaries. Within each spatial unit (regular latitude–longitude grid in this study), surface runoff is first routed across hillslopes and then is discharged along with subsurface runoff into a “tributary subnetwork,” with a transport capacity that is equivalent to all tributaries combined and is thus scale adaptive to the number of tributaries within a spatial unit. The spatial units are linked via routing through the main channel network, which is constructed in a scale-consistent way to preserve the upstream drainage areas and flow distances across different spatial resolutions (Wu et al. 2011, 2012). All model parameters are physically based, and only a small subset of them require calibration.

In this study, MOSART has been coupled with CLM4 in the same manner as an existing runoff routing module, RTM (Branstetter and Erickson 2003; Oleson et al. 2010). In the coupling framework, most of the original CLM4 software structure is maintained, except for the physics part representing the routing processes, which is replaced by MOSART. This coupled modeling framework is called CLM4–MOSART herein. CLM4 simulates the surface runoff and base flow for each grid cell at each time step. The gridded CLM4-simulated surface runoff and base flow is transferred to MOSART at the end of the time step, and MOSART routes the runoff across hillslope and through subnetwork and main channels. While RTM uses globally uniform and constant river velocity, MOSART explicitly simulates both spatial and temporal variability of flow velocity. MOSART applies different time steps for hillslope routing and channel routing. While the time step for hillslope routing is similar to that of CLM4, for example, hourly, the time steps for subnetwork routing and main channel routing could be much smaller to assure numerical stability, for example, from minutes to seconds depending on local channel slope and upstream drainage areas. As such, different grids may have different time steps for the channel routing. Generally, a channel segment with a steeper channel slope and larger upstream drainage area would require a finer time step. Hence, MOSART is more demanding in terms of parameterization. It includes length of the channel, width, and slope for both the subnetwork and main channel networks. Readers are referred to the original publication (Li et al. 2013) and the following section that describes the development of the parameters’ dataset at the global scale.

To assess the impacts of the added model complexity, successive simulations are performed by turning off the subgrid routing and removing the temporal and spatial variability of channel flow velocities.

b. Hydrography dataset

One important contribution of this study is the development of a comprehensive, global hydrography dataset to support the implementation of large-scale, physically based river routing models at multiple spatial resolutions. The topographic parameters, including flow direction, channel length, and topographic and channel slopes, are derived with the Dominant River Tracing (DRT) algorithm (Wu et al. 2011, 2012). The DRT algorithm produces the topographic parameters in a scale-consistent way to preserve/upscale the key features of a baseline high-resolution hydrography dataset at multiple coarser spatial resolutions (hereafter, a grid at baseline high resolution is denoted as a grid, while a grid at coarser resolution is denoted as a cell). The baseline high-resolution hydrography dataset used in this study is the 1-km resolution Hydrological Data and Maps Based on Shuttle Elevation Derivatives at Multiple Scales (HydroSHEDS; Lehner and Döll 2004; Lehner et al. 2008).

Figure 1 shows the upstream drainage area, drainage density, and channel and topographic slope maps at 1/3° resolution, which is the resolution MOSART is applied in this study globally. Upstream drainage area is the total upstream source area of each coarse cell, and its estimation and evaluation is reported in Wu et al. (2011, 2012). Drainage density is the total channel length of the river network (including main channel and all tributaries) within a coarse grid cell divided by the local cell...
area. To delineate the river network, a threshold area of 5 km² has been used globally. Channel slope is the average slope of the dominant river flow path of the cell, and topographic slope is the average slope of all hillslope flow paths within a cell. The flow paths are all identified based on the 1-km digital elevation model of HydroSHEDS. For more details on the estimation of channel and topographic slope, please refer to Li et al. (2013).

Also shown in Fig. 1 are the channel geometry parameters and bankfull width and depth, estimated from empirical hydraulic geometry relationships as functions of the mean annual discharge. The mean annual discharge map used in this study is based on the composite runoff field from the Global Runoff Data Centre (GRDC; more details will be provided in section 3d). The Manning roughness coefficients for overland and channel flow are calculated as functions of land cover and water depth. For more details on the methodology to derive channel geometry and Manning’s roughness coefficients, please refer to Getirana et al. (2012). The full list of parameters included in this global hydrography dataset is provided in Table 1. Note that all of the
global hydrography parameters above have been derived at seven different resolutions, that is, $2^8$, $1^8$, $1/8^8$, $1/4^8$, $1/16^8$, $1/32^8$, and $1/64^8$. In this study, only those parameters at the $1/2^8$ resolution are used.

3. Experimental approach

**a. Coupled CLM4–MOSART global application**

The coupled CLM4–MOSART has been applied globally driven by the four different atmospheric forcing datasets described in section 3b. Note that all the land surface parameters are the same as those provided with the NCAR-I2000 configuration (see section 3b for explanation; for more details, please refer to the user’s guide of CLM4, available online at www.cesm.ucar.edu/models/ccsm4.0/clm/models/lnd/clm/doc/UsersGuide/book1.html). For all the numerical simulations described in the following, unless specified differently, the time step for CLM4 is 30 min and for MOSART is 1 h, and the final model results are saved at each hourly time step. The spatial resolution of the output runoff data is the same as the finite volume $1/2^8 \times 1/2^8$ grid defined for the NCAR-I2000 configuration. In CLM4–RTM, the spatial resolution of runoff routing and output streamflow is fixed at $1/2^8$ resolution. In the updated CLM4–MOSART, the spatial resolution of runoff routing can be adjusted based on the resolution of the supporting hydrography dataset (as described in section 2b). In this study, we use $1/2^8$ resolution for runoff routing. The runoff time series generated by CLM4 are remapped to the resolution of the routing model with a CLM4-embedded algorithm before feeding into MOSART.

**b. Sensitivity to climate forcings**

Four different atmospheric forcing datasets are used in this study to drive CLM4–MOSART simulations. The first dataset was created by Qian et al. (2006) using observed monthly precipitation, surface air temperature, cloud cover, and satellite solar radiation to constrain the monthly averages of the 6-hourly National Centers for Environmental Prediction–National Center for Atmospheric Research (NCEP–NCAR) reanalyses (hereafter denoted as the QIAN forcing dataset). The spatial resolution is determined by the NCEP T62 Gaussian gridding system (roughly equivalent to $17/8^8$). This dataset has been used in the official release of CLM4 by NCAR as part of the benchmarking configuration for CLM4 under CO2 level and aerosol deposition for the year 2000, which is denoted as NCAR-I2000. The QIAN forcing dataset includes air temperature, specific humidity, wind speed, surface pressure, precipitation, and incoming solar radiation for the period 1948–2004 at a 6-h time step.

The other three atmospheric forcing datasets were derived by Getirana et al. (2014) based on the global meteorological forcing dataset for land surface modeling developed by Sheffield et al. (2006). Getirana et al. 2014 preserved most of the forcing variables from Sheffield et al. (2006) but rescaled the most important variable, precipitation, with three different precipitation products: 1) the monthly Global Precipitation Climatology Centre (GPCC) full data product, version 6, on a $1/4^8$ grid (Schneider et al. 2014); 2) the monthly Global Precipitation Climatology Project (GPCP), version 2.2, at $21/8^8$ resolution (Adler et al. 2003); and 3) the daily Environmental Research Observatory–Geodynamical, Hydrological and Biogeochemical Control of Erosion/Alteration and Material Transport in the Amazon Basin (ORE-HYBAM) at $1^8$ resolution (Guimberteau et al. 2012). Note that the ORE-HYBAM precipitation data cover only the Amazon River basin; therefore, the resulting forcing dataset is the same as Sheffield et al. (2006) over most of the global domain, except for the Amazon River basin. For the sake of brevity, these three

---

### Table 1. List of parameters in the global hydrography dataset. Note that D8 refers to an extensively used algorithm that tracts surface runoff from each pixel to one of its eight neighbor pixels.

<table>
<thead>
<tr>
<th>Name</th>
<th>Unit</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$F_{dir}$</td>
<td>—</td>
<td>The D8 single flow direction for each coarse grid cell coded using 1 (east), 2 (southeast), 4 (south), 8 (southwest), 16 (west), 32 (northwest), 64 (north), and 128 (northeast)</td>
</tr>
<tr>
<td>$A_{total}$</td>
<td>km$^2$</td>
<td>The upstream drainage area of each coarse grid cell</td>
</tr>
<tr>
<td>$F_{dir}$</td>
<td>m</td>
<td>The dominant river length for each coarse grid cell</td>
</tr>
<tr>
<td>$S_{channel}$</td>
<td>—</td>
<td>The average channel slope for each coarse grid cell</td>
</tr>
<tr>
<td>$S_{topographic}$</td>
<td>—</td>
<td>The average topographic slope (for overland flow routing) for each coarse grid cell</td>
</tr>
<tr>
<td>$A_{local}$</td>
<td>km$^2$</td>
<td>The surface area for each coarse grid cell</td>
</tr>
<tr>
<td>$D_p$</td>
<td>m$^{-1}$</td>
<td>Drainage density, calculated as the total channel length within each coarse grid cell divided by the local cell area</td>
</tr>
<tr>
<td>$D_r$</td>
<td>m</td>
<td>The bankfull depth of main channel</td>
</tr>
<tr>
<td>$W_r$</td>
<td>m</td>
<td>The bankfull width of main channel</td>
</tr>
<tr>
<td>$D_t$</td>
<td>m</td>
<td>The average bankfull depth of tributary channels</td>
</tr>
<tr>
<td>$W_t$</td>
<td>m</td>
<td>The average bankfull width of tributary channels</td>
</tr>
<tr>
<td>$n_r$</td>
<td>—</td>
<td>Manning’s roughness coefficient for channel flow routing</td>
</tr>
<tr>
<td>$n_h$</td>
<td>—</td>
<td>Manning’s roughness coefficient for overland flow routing</td>
</tr>
</tbody>
</table>
forcing datasets are hereafter denoted as GPCP, GPCC, and HYBAM, respectively. The GPCP, GPCC, and HYBAM forcing datasets are all available for the period of 1979–2008 because of the availability of the above precipitation products.

Figure 2 shows the annual mean precipitation map from QIAN and the difference between GPCP, GPCC, or HYBAM and QIAN for the period of 1986–95. The choice of this period is to be consistent with the GRDC dataset. We note that for 1986–95, the QIAN precipitation is constrained by monthly observation data from rain gauges only. Among the four forcing datasets, only GPCP is derived as a combination of satellite and gauge data, while the other datasets are either constrained by or derived only from station data. The global averages of annual mean precipitation are 1.492, 1.632, 1.418, and 1.422 mm day\(^{-1}\) for the QIAN, GPCP, GPCC, and HYBAM datasets, respectively. The contrast between wet regions such as the Amazon and Southeast Asia with dry regions such as the Sahara Desert, the Arabian Peninsula, the Tibetan Plateau, and the southwestern United States is clear in the QIAN forcing data. Both GPCC and HYBAM are generally drier compared to QIAN, except for a large part of South America, especially in the northeast, the Amazon, and Brazil; south of the Tibetan Plateau; and for a much wetter Greenland in HYBAM. In contrast, GPCP is generally wetter than QIAN except in the Sahel, Greenland, and Antarctica, which may be caused by the undercatch correction procedure in the GPCP. Overall, the largest differences among the datasets are located in areas with complex topography and coastlines such as Southeast Asia, the Himalayas, the Alps, the Sierra Madre, and the northern Andes, partly reflecting differences in the spatial resolution of the original data from which these various forcing datasets were derived. Although these areas are relatively small, they may contribute disproportionately to precipitation that drives runoff in those basins.

c. Impacts of model complexity

The routing processes can be grouped into two categories: within- and between-grid processes. Within-grid routing processes include the overland flow routing across hillslopes and channel routing within the tributaries. Between-grid routing is dominated by the main channel processes, which can be represented with different levels of complexity depending on the parameterizations of channel velocity. In this study, we investigate the impacts of different routing processes on streamflow simulation by reducing the model...
complexity step by step. With a focus on the impacts of model complexity, only the QIAN forcing is used to conduct the following five numerical experiments. For experiment 0, the baseline simulation represents all within- and between-grid routing processes included in MOSART as described in Li et al. (2013). For experiment 1, the within-grid routing processes are turned off (by delivering the surface and subsurface runoff instantaneously into the main channels) but fully represent the spatiotemporal variation of the main channel velocity simulated by the kinematic wave method. For experiment 2, the within-grid routing processes are turned off and the temporal variation of channel velocity is removed. The spatial velocity map is derived by averaging the time series of channel velocity for each grid generated in experiment 0. For experiment 3, the within-grid routing processes are turned off and the spatiotemporal variation of channel velocity is removed. The spatially uniform and temporally constant velocity field is derived from the global average velocity value from experiment 2, which is $\sim 0.21 \text{ m s}^{-1}$. For experiment 4, there are similarities to experiment 3, but the velocity value is $0.35 \text{ m s}^{-1}$, the same as that used in the RTM algorithm in CLM4.

These five simulations are denoted as MO_baseline, MO_wgoff, MO_wgoff_vXY, MO_v0.21, and MO_vRTM, respectively. One should note that MO_vRTM is in fact a replication of the standard RTM included in CLM4, except for a different underlying network map [for comparison between the network maps, please see Wu et al. (2011, 2012)]. For the readers’ reference, a brief description of each simulation is listed in Table 2.

d. Evaluation metrics and observations

Runoff routing processes have two major effects on runoff after its generation: accumulation and dispersion. The former involves accumulation of surface and subsurface runoff from land surface in the river channels that goes to the basin outlets and can be quantified in terms of the long-term basin water balance. The latter involves dispersion of travel time distributions of water across the land surface and through the river channels, which very often manifest at monthly scales over large regions or at shorter scales over small areas. Three metrics are used here to capture different effects of the routing processes, annual mean streamflow (AMS), mean monthly streamflow, and annual maximum flood (AMF). Annual mean streamflow is most useful in describing the accumulating effect, which is largely affected by the annual water balance of the upstream drainage area (of a gauge station where the streamflow is measured or simulated). Mean monthly streamflow captures both the accumulation (particularly for small areas where the residence time of surface water is much less than a month) and dispersion (particularly for large regions where the residence time of surface water is close to a month). AMF (the maximum discharge in a calendar year, usually derived from daily or subdaily time series) captures both the accumulating and dispersion effects, particularly during major storm events.

Note that AMF is used here to indicate the magnitude of streamflow rate only, that is, not necessarily as an indicator of flooding that is also controlled by the channel geometry and topography. Mean monthly streamflow and AMF are expected to be more directly and significantly affected by human activities such as reservoir operations.

The observations to evaluate CLM4–MOSART simulations include monthly runoff maps and observed streamflow data from the GRDC. The runoff data were derived by the University of New Hampshire (UNH) through combining observed river discharge information from GRDC with a climate-driven water balance model and were peer-reviewed by the International Satellite Land Surface Climatology Project (Fekete et al. 1999, 2002; Fekete and Vörösmarty 2011; Hall et al. 2006). This runoff dataset is provided at a monthly scale for the period of 1986–95 at both $\frac{1}{2}^\circ$ and $1^\circ$ resolutions.
Hereafter, these runoff data are denoted as the UNH–GRDC composite runoff map. Figure 3 shows the annual mean runoff map at 1° resolution derived from the UNH–GRDC composite dataset. Comparing Fig. 3 with Fig. 2, one can see a large similarity between the spatial pattern of runoff and precipitation, that is, high runoff and precipitation over the Amazon and Southeast Asia regions and low runoff and precipitation over the Sahara region. Note that the UNH–GRDC composite runoff data do not include much of the Greenland and Antarctica areas.

The observed streamflow data from about 6900 stations were provided by GRDC at either monthly or daily scales. The geographic information of these stations, including upstream drainage areas and longitude and latitude values, is used to georeference these stations to the hydrography dataset at the 1/2° resolution. For each station, the upstream area value provided by GRDC $A_{GRDC}$ is compared to that estimated from the upstream area map shown in Fig. 1 $A_{DRT}$. If the two upstream area values differ by no more than 10%, as done by Wu et al. (2014), the station passes the georeferencing and is selected for the subsequent evaluation. As for the stations with $A_{GRDC}$ and $A_{DRT}$ differing substantially, those with smaller drainage area values (e.g., $A_{GRDC} < 10000$ km²) are simply dropped off; those with larger drainage area values are manually adjusted after a visual check, that is, slightly modifying the latitude–longitude values of these GRDC stations so that the resulting updated $A_{DRT}$ values are close enough to the $A_{GRDC}$ values. This way, the accumulation effect of runoff routing processes is preserved reasonably well by preserving the upstream drainage areas. There are 3195 GRDC stations that meet the upstream drainage area criterion.

The observed streamflow data, particularly those corresponding to large drainage areas, are often impacted by human activities on river systems, such as reservoir operation, surface water withdrawal, groundwater pumping, and irrigation. To better understand the impacts of human activities on model evaluation, we adopt the classification framework of Nilsson et al. (2005) to divide the global domain into three classes based on the level of flow regulation. Nilsson et al. (2005) estimated the total reservoir storage capacity and irrigation amount in each of 292 of the world’s largest river systems and classified these river systems as unaffected (120), moderately affected (68), or strongly affected (104), as shown in Fig. 4. Here, a large river system is defined as an integration of the river network (from headwaters all the way to the ocean) and its associated drainage area. In the following analysis, this classification system can partially explain the difference between the simulated and observed streamflow.

4. Results

The river systems classification of Nilsson et al. (2005) is used as a framework to evaluate the impacts of human activities. For this purpose, the period of 1995–2004 is chosen for display in most of the figures, unless otherwise specified. This 10-yr period is sufficiently long for model evaluation purposes, and we assume that most of the dam constructions as used in Nilsson et al. (2005) have been completed before this period. In the following, the analysis of model complexity impacts is conducted with a single forcing datum for all river basins so it is independent of forcing uncertainty and human activities.
a. Impacts of model complexity

The QIAN forcing dataset is used in the experiments to evaluate the impacts of model complexity. The simulated streamflows from the MO_baseline, MO_wgoff, MO_wgoff_vXY, MO_v0.21, and MO_vRTM simulations (Table 2) are first evaluated against the observed streamflow from GRDC within the period of 1980–2004. The choice of this period is to be consistent with the forcing datasets' availability. Recall that 1979–2004 is

b. GRDC stations within the selected large river systems only

c. GRDC stations with upstream drainage areas larger than 20,000 km²

FIG. 4. GRDC stream gauge stations used to evaluate simulated streamflow. (a) Distribution of 1674 GRDC stations (blue circles) overlaid with the classification map based on the level of streamflow regulation [adapted from Nilsson et al. (2005)]. (b) GRDC stations within the selected large river systems only. (c) GRDC stations with upstream drainage areas larger than 20,000 km², which are classified based on the level of streamflow regulation.

a. Impacts of model complexity

The QIAN forcing dataset is used in the experiments to evaluate the impacts of model complexity. The simulated streamflows from the MO_baseline, MO_wgoff, MO_wgoff_vXY, MO_v0.21, and MO_vRTM simulations (Table 2) are first evaluated against the observed streamflow from GRDC within the period of 1980–2004. The choice of this period is to be consistent with the forcing datasets' availability. Recall that 1979–2004 is
The RMSE is between the simulated and observed series.

<table>
<thead>
<tr>
<th>Model</th>
<th>AMS</th>
<th>AMF</th>
</tr>
</thead>
<tbody>
<tr>
<td>QIAN (MO_baseline)</td>
<td>0.855 0.694</td>
<td>0.832 0.896</td>
</tr>
<tr>
<td>MO_wgoff</td>
<td>0.856 0.703</td>
<td>0.830 0.897</td>
</tr>
<tr>
<td>MO_wgoff_vXY</td>
<td>0.856 0.703</td>
<td>0.805 0.781</td>
</tr>
<tr>
<td>MO_v0.21</td>
<td>0.856 0.701</td>
<td>0.758 0.599</td>
</tr>
<tr>
<td>MO_vRTM</td>
<td>0.856 0.702</td>
<td>0.780 0.697</td>
</tr>
<tr>
<td>GPCP (MO_baseline)</td>
<td>0.868 1.029</td>
<td>0.790 1.278</td>
</tr>
<tr>
<td>GPCC (MO_baseline)</td>
<td>0.873 0.956</td>
<td>0.792 1.160</td>
</tr>
<tr>
<td>HYBAM (MO_baseline)</td>
<td>0.870 0.985</td>
<td>0.783 1.163</td>
</tr>
</tbody>
</table>

The common period across the QIAN, GPCP, GPCC, and HYBAM forcings. The year 1979 is treated as the spinup period of MOSART and thus is not counted.

To evaluate both the simulated AMS and AMF, 1674 GRDC stations are selected with no less than 10 years of complete records of daily streamflow observation in 1980–2004, as shown in Fig. 4a. For each station, long-term averages of the mean and maximum values within each calendar year are then calculated for the years with complete daily streamflow observations (directly averaged with the streamflow values without invoking drainage area). Note here that any year with more than 360 daily streamflow records is considered a year with complete records. Linear regression has been done between the model simulations and observations for the selected 1674 GRDC stations.

Table 3 lists the $R^2$ and slope values (from linear regression) for AMS and AMF from these simulations. All the slope values are statistically significant based on a Student’s t test with a p value less than 0.001. CLM4–MOSART reproduces the long-term averaged AMS and AMF well, as indicated by the fact that all $R^2$ values are higher than 0.75. The slope values for the AMS differ slightly among the MO_baseline, MO_wgoff, MO_wgoff_vXY, MO_v0.21, and MO_vRTM simulations, but those for the AMF differ more substantially. Recall that the difference between MO_baseline and MO_wgoff is due to the effects of the subgrid routing processes, the difference between MO_wgoff and MO_wgoff_vXY is due to the effects of temporal variability of channel velocity, and the difference between MO_wgoff_vXY and the rest (MO_v0.21 and MO_vRTM) is due to the effects of spatial variability of channel velocity. For the simulation of AMF, the difference between the $R^2$ value from MO_wgoff (0.830) and that from MO_wgoff_vXY (0.805) is comparable to the difference between the $R^2$ value from MO_wgoff_vXY and those from MO_v0.21 (0.758) and MO_vRTM (0.78). Similar results are found for the slope and root-mean-square error (RMSE). One can then conclude that the effects of temporal variability of channel velocity on the simulated AMF are as important as those of the spatial variability of channel velocity and values of global uniform channel velocity.

The impacts of the subgrid routing processes appear to be much less than those of the spatial and temporal variability of channel velocity. This is mainly because the subgrid routing processes are mostly local, that is, affecting discharge from the local grid into the main channel, and thus have indirect impacts on main channel routing, while other processes have direct impacts on main channel routing. Note in this study the GRDC streamflow observation is exclusively provided at the gauges located on the main channels. RMSE values are also calculated between the observation and each model simulation. The major effects of both spatial and temporal variability of channel velocities are again confirmed by the RMSE values.

To better present the effects of model structure, the Taylor diagram (Taylor 2001) is used to visualize the statistical relationship of simulated monthly streamflow between the baseline simulation, MO_baseline, and other simulations, as shown in Fig. 5. This comparison highlights the impacts of model complexity rather than compares model biases between observed and simulated streamflow, which are more affected by factors such as biases in the runoff simulations and human impacts that are not directly related to model complexity. Three major statistics are captured in this diagram: correlation (shown as dashed straight lines), centered root-mean-square difference [proportional to the distance from the reference point on the horizontal axis (REF), shown as solid arc lines], and amplitude of variations (proportional to the distance from the origin, shown as dashed arc lines). Here the simulation of MO_baseline is the reference, and each marker represents comparison of monthly streamflow time series (period of 1995–2004) at one GRDC station simulated by MO_wgoff, MO_wgoff_vXY, MO_v0.21, or MO_vRTM, with the streamflow...
Simulated by MO_baseline. Overall, at the monthly time scale, the effects of within-grid processes are minor since all the red markers (corresponding to MO_wgoff) are close to the reference point. The blue (MO_wgoff_vXY), green (MO_v0.21), and golden (MO_vRTM) markers are increasingly farther away from the reference point, indicating increasing differences from the baseline simulation.

It is important to note that the increasing distances of the various plotted points from the origin are associated with both reductions in the correlations from 1.0 to less than 0.1, and reductions in the normalized standard deviations from 1.0 to less than 0.25, indicating increasing differences from the baseline simulation. More specifically, from MO_wgoff and MO_wgoff_vXY to MO_v0.21 or MO_vRTM, the distances of the markers from the origin consistently decrease. The lowest correlation and variability are associated with MO_vRTM and MO_v0.21. Clearly, variations of the river velocity are increasing removed by reducing the model structure complexity step by step. The time lag between MO_baseline and MO_wgoff is clearly less than that between MO_baseline and MO_wgoff_vXY, but it is not always clearly less than that between MO_baseline and MO_v0.21 or MO_vRTM. This suggests that turning off the temporal variability of channel velocity will certainly change the seasonal cycle compared to the baseline simulation, but further turning off the spatial variability might have some compensating effects on the timing of monthly streamflow. More investigation to elucidate the timing aspect is thus needed. Mean monthly streamflow curve, or regime curve, is a good indicator of the timing of streamflow at the monthly scale, which is discussed next.

Figure 6 shows the simulated mean monthly streamflow from the above five numerical experiments at 12 selected GRDC gauge stations. The gauge selection is done in two steps: 1) compare the outlet stations of river systems in each category as listed in Nilsson et al. (2005) with the list of GRDC stations described in section 2d and only select those that exist in both station lists (therefore, those stations are all located at the outlets of the river systems), and 2) based on the first step, select those stations with complete monthly streamflow records within the period of 1995–2004. After the two steps, we show the mean monthly streamflow at the outlet stations for four river systems within each category: not affected, moderately affected, and strongly affected.

The difference due to various model structures manifests consistently across all stations (Fig. 6), despite the fact that these stations are very diverse with respect to the geographic locations, climate regimes, and human influences. The mean monthly hydrographs from the MO_wgoff simulation show slightly higher peaks than those from the MO_baseline simulation. This is because neglecting the within-grid routing processes leads to reduced dispersion effect. The effects of temporal variation of channel velocity (indicated by the difference...
between the MO_wgoff_vXY and MO_wgoff simulations) are clearly more important than the within-grid routing processes at the monthly scale [the latter are important at daily or subdaily scales, as illustrated by Li et al. (2013)]. The effects of spatial variation of channel velocity (indicated by the difference between the MO_v0.21 and MO_wgoff_vXY simulations) are also apparent, as shown in different river systems. It is interesting that in the Indigirka, Pechora, Yana, and Orange River systems, the monthly mean streamflow simulated by the MO_wgoff_vXY and MO_vRTM simulations are pretty close. A first guess is that the actual long-term mean velocities averaged over all the grids within the Indigirka, Pechora, Amazon, Yana, and Orange River systems are close to 0.35 m s\(^{-1}\). The average velocities within the Yukon, Mackenzie, Mississippi, Lena, and Yenisey River systems are likely larger than 0.35 m s\(^{-1}\), as indicated by the earlier and higher monthly peak discharges. The impacts of different values of globally uniform and constant channel velocity are also clearly shown in all river systems, as expected. It is not feasible to use a single channel velocity to capture the diverse streamflow dynamics globally. Comparing the impacts of the different model

---

**FIG. 6.** Impacts of model structure on seasonality of streamflow. Note that the scales of the plots are not necessarily the same for the sake of clarity. River systems that are (left) not affected, (middle) moderately affected, and (right) strongly affected by flow regulation are shown.
complexity levels as demonstrated in Fig. 6, it is clear that the temporal variability of channel velocity is also an important factor in controlling the timing of monthly streamflow simulations besides the AMF.

Figure 6 shows the streamflow simulation only at the outlets of the large river systems. Spatially, however, streamflow can be highly variable, so it is interesting to find out how well the simulations reproduce the observed streamflow within large river systems. For this purpose, one more criterion of river system selection is added on top of those used in Fig. 6: only river systems including at least 10 GRDC stations with at least 8 years of complete daily streamflow records in 1995–2004 are selected. Only four river systems, Yukon, Mackenzie, Mississippi, and Danube, as shown in Fig. 4b, meet all the selection criteria. The impacts of routing model complexity on the AMS are expected to be negligible (as shown in Fig. 7 and Table 3, left) since the travel time of surface water (particularly through river channels) is usually much less than 1 year.

However, the impacts of routing model structure on the AMF peaks are clearly shown even from the log–log plots (Fig. 7, right). Note that each dot in Fig. 7 is for the simulation at one GRDC station, and the locations of these stations are shown in Fig. 4b. The temporal variability of channel velocity is again a dominant factor on the simulated magnitude of annual flood peaks, as one can tell from the difference between the simulated flood peaks from the MO_wgoff simulation and those from the MO_wgoff_vXY simulation. One may notice that the annual flood peaks simulated by the MO_vRTM case (uniform and constant velocity 0.35 m s\(^{-1}\)) are always larger than those by the MO_v0.21 case (uniform and constant velocity 0.2 m s\(^{-1}\)), which makes sense because larger channel velocity usually leads to larger flood peaks. This also explains the noticeable difference between the annual flood peaks simulated by MO_wgoff_vXY (spatially variable but temporally constant velocity) and MO_wgoff (spatiotemporally variable velocity), particularly in the Yukon and Mackenzie River systems, since the larger channel velocities associated with the flood peaks are well preserved in MO_baseline and MO_wgoff, while not in MO_wgoff_vXY, MO_v0.21, or MO_vRTM.

The difference in the simulated AMS and maximum flood caused by different levels of model complexity is further examined using the Wilcoxon rank-sum test (equivalent to the Mann–Whitney–Wilcoxon test, also known as the Mann–Whitney U test). This test, as used here, is a nonparametric statistical significance test with a null hypothesis that the two populations are the same. We used the standard MATLAB function “ranksum” to conduct the Wilcoxon rank-sum test to detect the statistical difference caused by different model structure, that is, difference among model simulations instead of between simulations and observations. For each simulation, the sample space is taken as the same as those in Table 3, that is, averaged AMS and AMF values over 1980–2004 for each of 1674 GRDC stations (so 1674 points for AMS and another 1674 points for AMF).

Table 4 lists the p values of the Wilcoxon rank-sum test on the difference between simulations under different model structures. Table 4 shows that the impacts of routing model structure are minimal on AMS, but statistically significant on AMF. This result is consistent with what we draw from Fig. 7.

b. Uncertainty due to climate forcing

Besides model complexity, uncertainty in the forcing dataset, particularly precipitation, is also an important source of streamflow simulation biases. To investigate the impacts of forcing uncertainty, three additional numerical experiments are conducted by using the GPCP, GPCC, and HYBAM forcing datasets to drive the coupled CLM4–MOSART. Note that the full capability of MOSART has been used in these experiments on climate forcing. These experiments are named after the forcing datasets, that is, QIAN, GPCP, GPCC, and HYBAM. Also note that the QIAN simulation is essentially the same as the previous MO_baseline simulation described in section 4a. A brief description of each additional numerical experiment is also listed in Table 2 for the readers’ reference. The uncertainties of the forcing dataset affect the streamflow simulation more indirectly through the runoff generation processes. Figure 8 shows the difference between the mean annual total runoff (sum of surface and subsurface runoff) simulated with the QIAN forcing and UNH–GRDC (Fig. 8, top left). Figure 8 also shows the relative differences between the mean annual total runoff from the CLM4 simulations driven by GPCP, GPCC, and HYBAM compared to that driven by QIAN. Note that the absolute differences between our simulated runoff and the GRDC runoff are generally much larger than differences caused by different forcing data. The global averages of annual mean total runoff are 0.480, 0.675, 0.518, and 0.509 mm day\(^{-1}\) for the simulations driven by the QIAN, GPCP, GPCC, and HYBAM datasets, respectively, for the period of 1986–95. The globally averaged annual mean total runoff from the UNH–GRDC composite runoff field is 0.722 mm day\(^{-1}\), higher than any of the simulations in this study. The differences among the simulated global mean total runoff are consistent with the differences among the global mean precipitation, with the highest global mean precipitation in GPCP producing the highest global mean total runoff.
Fig. 7. Impacts of model structure on (left) AMS and (right) AMF. Note that the scales of the plots are not necessarily the same for the sake of clarity.
The simulated runoff driven by the QIAN forcing shows a general dry bias in very wet regions, including parts of the Amazon and Congo rain forests and Southeast Asia, but broad areas with wet bias are found in East Asia and central Asia and southeastern and northwestern North America. Corresponding to the larger differences in precipitation among the forcing data in the Amazon, southeast Brazil, Southeast Asia, and central Africa, large differences are also found among the runoff from different simulations in the same regions. Larger precipitation differences in Greenland and Antarctica are inconsequential, as these regions are too cold to produce substantial liquid runoff. Overall, in the tropics and mid-latitudes, the simulated runoff differences largely reflect the precipitation forcing differences on an annual basis.

Figure 9 shows the Taylor diagrams comparing the spatial maps in Figs. 2 and 8. Unlike Fig. 5, which compares temporal variations, here we compare spatial distributions of annual mean precipitation or annual mean total runoff over the latitude–longitude grids, instead of time series as in Fig. 5. The annual precipitation field from the QIAN forcing is used as the reference for precipitation comparison, and the annual runoff field from the simulation driven by the QIAN forcing is used as the reference for runoff comparison. The marker of the GPCP precipitation (red dot label as ‘‘1’’) is farther from the reference point than the markers of GPCC and HYBAM, indicating the overall larger difference from the QIAN precipitation. The larger difference is mainly due to an increase in spatial variability rather than changes in the spatial correlation. It is interesting that this difference is amplified, again due to an increase in variability of magnitude rather than changes in timing, to about twice as large in the annual mean runoff field, reflecting the nonlinear effects of landscape heterogeneities in the CLM4 runoff generation scheme through parameters related to topography and soil hydraulic properties. As shown by the proximity between the markers for GPCC and HYBAM, the ORE-HYBAM precipitation in the Amazon basin has limited impacts on the global spatial distribution of runoff, though over the Amazon basin, the effects can be large locally, as seen in Fig. 8.

The model-simulated AMS and AMF from the GPCP, GPCC, and HYBAM simulations are also evaluated against the observations from 1674 GRDC stations as listed in Table 3. The high $R^2$ values (mostly over 0.7) suggest that CLM4–MOSART reproduces the annual mean and peak discharges reasonably well under different forcing datasets. Forcing uncertainty affects both simulated AMS and AMF, as indicated by the slope and $R^2$ values. Note that the $R^2$ values for AMF are consistently lower than those for AMS. This is largely because 1) the impacts of human activities such as
reservoir operation on streamflow are mostly significant at monthly or shorter time scales, but not as much at annual scales, and 2) the model used in this study does not explicitly represent human activities yet.

Similar to Fig. 5, Fig. 10 shows a Taylor diagram to visualize the statistical relationship of the simulated monthly streamflow using the baseline model structure, but driven by different forcings at four locations. Instead of comparing to the baseline simulation as the reference, however, Fig. 10 compares simulations driven by various forcings with observations. Comparing with Fig. 5, one would notice that the values of correlation in Fig. 10 fall within a narrower range, implying that the small timing difference among the forcings largely constrains the timing difference among the streamflow simulations. Comparing Fig. 10 with 9, however, the range of correlation values is larger, indicating the spatial variability in the precipitation is not only propagated to the streamflow, but also enhanced during the nonlinear runoff generation and routing processes due to highly spatial heterogeneous landscape properties. Compared to observed streamflow, most simulations have lower variability, but simulations with larger variability than the observed are also noted. The latter is mostly associated with class 2 (stars, moderately affected by flow regulation) and class 3 (triangles, strongly affected by flow regulation) GRDC stations, indicating that flow regulation impacts have also come into play. The difference between the simulated streamflow at the class 1 stations (not affected by flow regulation) and those at the class 2 or 3 stations (affected by flow regulation) is as big as that caused by the different forcings.

Figure 11 shows a comparison between the observed and simulated mean monthly streamflow averaged over the period of 1995–2004 for the 12 GRDC stations used previously. The GPCP simulation provides higher streamflow over the river systems such as Yukon, Indigirka, Pechora, and Yana, which are located in high-latitude regions. This is consistent with Fig. 8 that the runoff produced by GPCP simulation is higher over the high-latitude regions. In the Congo River system in Africa and the Amazon River system in South America, the QIAN simulation underestimates the streamflow, as also indicated by the lower runoff produced by the QIAN simulation, as shown in Fig. 8. The other three forcing datasets, nevertheless, lead to simulated monthly streamflow magnitude close to the observation in the Congo and Amazon River systems, indicating that these three forcing datasets may contain less precipitation bias than the QIAN forcing. In other words, this suggests the
underestimation of precipitation by the QIAN forcing dataset over central Africa and the Amazonas. The observed streamflow at the Mississippi, Orange, and Danube River systems are all apparently lower than those from all four simulations, indicating that human activities might play a role in reducing the total streamflow at least in the period of 1995–2004.

Figure 12 shows that the impacts of forcing uncertainty are substantial on AMS. The difference of AMF simulations under different forcings is largely consistent with that of AMS, indicating the dominating sensitivity of streamflow to forcing uncertainty. In Table 5, p values of the nonparametric Wilcoxon rank-sum test (over the 1674 GRDC stations for the period of 1980–2004) are listed to test the null hypothesis that simulations driven by different forcings are from the same population for each pair of simulations. The simulations under the GPCP forcing are evidently different from those under the other three forcings in terms of both AMS and AMF, as indicated by the very small p values. The difference among the QIAN, GPCC, and HYBAM simulations is not statistically significant, particularly in terms of AMS. These suggest that precipitation differences among the forcing datasets, which could be attributed to both magnitudes and spatial distribution and resolution, dominate the overall differences among the streamflow simulations. The impacts of the differences between other forcing variables such as radiation are likely a secondary controlling factor on the streamflow simulations.

c. Human influences on streamflow

Differences between the simulated and observed streamflow can only be partially attributed to the effects of model complexity and forcing datasets. Human activities also exert their impacts on streamflow directly (e.g., through dam regulation and surface water withdrawal) and indirectly (e.g., through land use change and irrigation). However, it is more challenging to attribute model–observation differences to the impacts of human influences than to uncertainties of climate forcing and model structure. Figure 12 shows such an attempt to compare the observed and simulated streamflow in large river systems influenced by different levels of human
influences [in this case, mainly dam regulation and irrigation according to Nilsson et al. (2005)]. Similar as in Fig. 7, five river systems—Yukon, Mackenzie, Amazon, Mississippi, and Danube—are selected for this purpose. Again, we include all GRDC stations within each river system as long as complete data records are available.

In the Yukon River system not affected by dam regulations (Fig. 12), the model-simulated mean annual streamflow and AMF are well aligned with the observed along the 1:1 line. In the Mackenzie and Amazon River systems, which are moderately affected by dam regulations, the simulated mean annual streamflow still largely aligns with the observations along the 1:1 line. The model-simulated AMFs, however, are noticeably larger than the observations. This is true under different atmospheric forcings. Given that one of the major functions of dams and reservoirs is to reduce flood peaks, it is reasonable to infer that this type of human influence has impacted the AMF in the Mackenzie and Amazon River systems. This is true under different atmospheric forcings. Given that one of the major functions of dams and reservoirs is to reduce flood peaks, it is reasonable to infer that this type of human influence has impacted the AMF in the Mackenzie and Amazon River systems. This is true under different atmospheric forcings. Given that one of the major functions of dams and reservoirs is to reduce flood peaks, it is reasonable to infer that this type of human influence has impacted the AMF in the Mackenzie and Amazon River systems.

Figure 12 represents the GRDC stations from five selected river systems, so the conclusions may not be extensible to other river systems. Also, if one roughly associates those dots of smaller streamflow values with smaller drainage areas, the relatively larger scattering over the GRDC stations with smaller streamflow values may be attributed partially to uncertainties associated with the area preservation in the hydrography dataset, which are expected to be more substantial over small drainage areas. Figure 13 shows a similar analysis but
over extended GRDC records compared to Fig. 12. The major differences between Figs. 12 and 13 are: 1) GRDC stations from all river systems with good observation records are considered in Fig. 13; 2) instead of plotting simulations driven by different forcings (Fig. 12), the ensemble mean of the four simulations driven by different forcing datasets are plotted in Fig. 13; 3) only those GRDC stations with a drainage area over 20000 km$^2$ are plotted in Fig. 13 to reduce uncertainty associated with the area preservation in the routing model; and 4) each dot in Fig. 13 now represents a calendar year at one GRDC station with complete daily streamflow records. 

Figure 4c shows the selected GRDC stations with a drainage area over 20000 km$^2$ and at least 1 year of complete daily streamflow observation records in the period of 1986–95. These stations are extensively distributed over the global domain and therefore are more representative compared to those used in Figs. 4b and 12, which are for selected large river systems only. One can observe from Fig. 13 that for the class 1 stations (not affected by flow regulation), most of the points are close to the 1:1 line, that is, a small difference between the
FIG. 12. Impacts of forcing on (left) AMS and (right) AMF. Note that the scales of the plots are not necessarily the same for the sake of clarity.
simulated and observed AMS and AMF values; for the class 2 and 3 stations (moderately or strongly affected), the blue points (for AMS) and red points (for AMF) are mostly above the 1:1 line. Assuming the forcing uncertainty is now constrained by the ensembles, Fig. 13 thus clearly reveals the two effects of reservoir operation and irrigation: reducing the total streamflow and, more importantly, reducing the flood peaks. Taken together, Figs. 12 and 13 show that the impacts of human influences should be considered in evaluating streamflow simulated by river routing models. Furthermore, our results suggest that a combination of modeling and data analysis may be used to detect human impacts on streamflow, although uncertainties in forcing and model structure must be considered to establish the robustness of the results.

5. Discussion and conclusions

This section summarizes the above analysis, discusses the limitations of this study, and points to future directions of improvements.

a. Summary and concluding remarks

In this study, a physically based routing model, MOSART, has been coupled with the Community Land Model and applied globally. A comprehensive global hydrography dataset has been constructed to support the implementation of the routing model at several spatial resolutions. Overall, it is shown that MOSART is able to simulate streamflow reasonably well across the global domain. The difference between the simulated

<table>
<thead>
<tr>
<th></th>
<th>AMS</th>
<th></th>
<th>AMS</th>
<th></th>
<th>AMS</th>
<th></th>
<th>AMS</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>QIAN</td>
<td>GPCP</td>
<td>GPCC</td>
<td>HYBAM</td>
<td>QIAN</td>
<td>GPCP</td>
<td>GPCC</td>
<td>HYBAM</td>
</tr>
<tr>
<td>QIAN</td>
<td>1.0000</td>
<td>&lt;0.0001</td>
<td>0.6471</td>
<td>0.8528</td>
<td>1.0000</td>
<td>&lt;0.0001</td>
<td>0.2342</td>
<td>0.3189</td>
</tr>
<tr>
<td>GPCP</td>
<td>&lt;0.0001</td>
<td>1.0000</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
<td>1.0000</td>
<td>&lt;0.0001</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>GPCC</td>
<td>0.6471</td>
<td>&lt;0.0001</td>
<td>1.0000</td>
<td>0.7887</td>
<td>0.2342</td>
<td>&lt;0.0001</td>
<td>1.0000</td>
<td>0.8453</td>
</tr>
<tr>
<td>HYBAM</td>
<td>0.8528</td>
<td>&lt;0.0001</td>
<td>0.7887</td>
<td>1.0000</td>
<td>0.3189</td>
<td>&lt;0.0001</td>
<td>0.8453</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

Fig. 13. Detection of human influences over extended GRDC records in 1995–2004. Qsim stands for the simulations streamflow, and Qobs stands for the observed streamflow.
and observed streamflow, in terms of annual and monthly mean streamflow and AMF, is attributed to the uncertainty of atmospheric forcing and human activities, and a physically based routing model structure with adequate complexity is important to capture different aspects of streamflow. Analysis suggests that representing the spatially and temporally varying flow velocities process in the routing model has important effects on simulating seasonality of streamflow and magnitude of AMF. Decharme et al. (2010) showed that realistic representation of channel velocity is important for global simulations of monthly streamflow. Here we further show that the temporal variability of channel velocity has no less influence than the spatial variability of channel velocity, particularly with respect to the simulated magnitude of AMF. As summarized in Table 4, each level of complexity enabled by MOSART compared to a simpler model can lead to statistically significant differences in simulating streamflow.

Compared to the standard RTM implemented in CLM4 (as replicated by the MO_vRTM simulation), the more process-oriented MOSART overall captures the AMF better over the global domain and, to a certain degree, also captures the seasonality of streamflow better. The above analysis is nevertheless confounded by forcing uncertainty and human influences. As differences among the forcing datasets used in this study are mainly associated with precipitation amount, their impacts on the streamflow primarily manifest in the annual means and AMF rather than the timing of monthly streamflow. Among the four forcing datasets, the simulations driven by the GPCP forcing are most different from those driven by the other forcings for both AMS and AMF, due to both magnitude and spatial distribution/resolution differences in the GPCP precipitation compared to the other three datasets. The impacts of human influences on streamflow are clearly different among the forcing datasets, which was considered the dominant factor of substantial runoff generation bias caused by the runoff scheme in CLM4, and Hou et al. (2012) and Huang et al. (2013) showed large sensitivity of CLM4 runoff simulations to parameters of the subsurface processes. More recently, Swenson et al. (2012) reported the overestimation of runoff over cold regions was due to not accounting for the effects of soil ice on the hydraulic properties, which was considered the dominant factor of the overestimated streamflow in the Lena and Yenisey River systems. This is indeed also the case in the Indigirka, Pechora, and Yana River systems, which are in cold regions. In these river systems, the bias associated with the runoff generation process dominates over the impacts of human activities, although the Lena River system is classified as moderately affected by dam regulation and the Yenisey River system is classified as strongly affected. Extensive investigation of the impacts of runoff generation modeling structure in CLM4 is beyond the scope of this study.

The impacts of forcings on AMS are clearly more substantial than those of the routing model structure. This is expected since the AMS mainly reflects the annual water balance within the upstream drainage area of a gauge station. Among the simulations, the simulation driven by the QIAN forcing has the largest drainage area of 41.5 W m\(^{-2}\), too high over global land areas compared to ground measurements. The negative bias in mean annual streamflow in the QIAN simulation is consistent with the high bias in downward shortwave radiation, based on surface energy and water balance. The impacts of forcings on AMF are also substantial, and again the QIAN simulation has a negative bias compared to the GRDC data, while the other simulations have positive biases.

The simulation bias associated with the runoff generation processes could propagate to the streamflow simulation since runoff time series are major inputs to the channel routing modules. Li et al. (2011) noted substantial runoff generation bias caused by the runoff scheme in CLM4, and Hou et al. (2012) and Huang et al. (2013) showed large sensitivity of CLM4 runoff simulations to parameters of the subsurface processes. More recently, Swenson et al. (2012) reported the overestimation of runoff over cold regions was due to not accounting for the effects of soil ice on the hydraulic properties, which was considered the dominant factor of the overestimated streamflow in the Lena and Yenisey River systems. This is indeed also the case in the Indigirka, Pechora, and Yana River systems, which are in cold regions. In these river systems, the bias associated with the runoff generation process dominates over the impacts of human activities, although the Lena River system is classified as moderately affected by dam regulation and the Yenisey River system is classified as strongly affected. Extensive investigation of the impacts of runoff generation modeling structure in CLM4 is beyond the scope of this study.

The modeling framework used in this study can be improved in several aspects. For example, the classification
of river systems by Nilsson et al. (2005) can be refined to improve the detection of human influences. Their classification categorizes a whole river system without differentiating the spatial heterogeneity of human influences within a river system. A more detailed, highly distributed classification map of human influences can be derived to elucidate the effects of human activities on local streamflow and downstream channels.

The impacts of reservoir operation and irrigation are demonstrated here by way of data analysis, but clearer mechanistic understanding might be gained by a numerical experiment conducted with an extended CLM4–MOSART with a reservoir operation scheme (Voisin et al. 2013a,b). In fact, coupling a reservoir module with CLM4–MOSART is an ongoing endeavor and its global application will be reported in the near future.

One may notice that MOSART is still not capturing the timing of streamflow well enough in several places such as Lena and Yukon. For those rivers in high-latitude areas, ice transport and jamming mechanisms are known to have important effects on the timing of streamflow, and these are not included in MOSART yet. In other river systems (e.g., the Amazon), representation of inundation dynamics, for example, exchange between the main channel and floodplain, will lead to more realistic simulations of average channel velocity and will help to improve the simulation of timing and magnitude of streamflow at short time scales such as daily or subdaily (Getirana et al. 2012). Extension of MOSART to incorporate inundation dynamics is thus the focus of another ongoing study.

Acknowledgments. This study was supported by the Office of Science of the U.S. Department of Energy as part of the Earth System Modeling (ESM) and Integrated Assessment Modeling (IAM) programs. Development of the datasets used in this study is partly supported by the PNNL Platform for Regional Integrated Modeling and Analysis (PRIMA) initiative. The Pacific Northwest National Laboratory is operated by Battelle for the U.S. Department of Energy under Contract DE-AC05-76RL01830. A. Getirana is funded by the NASA Postdoctoral Program (NPP) managed by Oak Ridge Associated Universities (ORAU). H. Wu is supported by NASA’s Applied Science Program. The first author also wants to thank C. Nilsson and C. A. Reidy for providing the data of river system classification.

REFERENCES


Hou, Z., M. Huang, L. Y. R. Leung, G. Lin, and D. M. Ricciuto, 2012: Sensitivity of surface flux simulations to hydrologic...


