

The critical role of the routing scheme in simulating peak river discharge in global hydrological models

Fang Zhao, Ted Veldkamp, Katja Frieler, Jacob Schewe, Sebastian Ostberg, Sven Willner, Bernhard Schauberger, Simon Gosling, Hannes Müller Schmied, Felix Portmann, et al.

▶ **To cite this version:**

Fang Zhao, Ted Veldkamp, Katja Frieler, Jacob Schewe, Sebastian Ostberg, et al.. The critical role of the routing scheme in simulating peak river discharge in global hydrological models. Environmental Research Letters, 2017, 12 (7), pp.075003. 10.1088/1748-9326/aa7250 . hal-02165382

HAL Id: hal-02165382

https://hal.sorbonne-universite.fr/hal-02165382

Submitted on 25 Jun 2019

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.

LETTER • **OPEN ACCESS**

The critical role of the routing scheme in simulating peak river discharge in global hydrological models

To cite this article: Fang Zhao *et al* 2017 *Environ. Res. Lett.* **12** 075003

View the article online for updates and enhancements.

Related content

- Intercomparison of global river discharge simulations focusing on dam operation—multiple models analysis in two case-study river basins, Missouri–Mississippi and Green–Colorado
Yoshimitsu Masaki, Naota Hanasaki, Hester Biemans et al.

- Global assessment of agreement among streamflow projections using CMIP5 model outputs
Sujan Koirala, Yukiko Hirabayashi, Roobavannan Mahendran et al.

- Intercomparison of regional-scale hydrological models and climate change impacts projected for 12 large river basins worldwide—a synthesis
Valentina Krysanova, Tobias Vetter, Stephanie Eisner et al.

Recent citations

- Impacts of spatial resolution and representation of flow connectivity on large-scale simulation of floods
Cherry May R. Mateo *et al*

- Natural and human-induced terrestrial water storage change: A global analysis using hydrological models and GRACE
Farshid Felfelani *et al*



This content was downloaded from IP address 82.67.178.220 on 08/11/2017 at 07:40

Environmental Research Letters



OPEN ACCESS

RECEIVED
7 February 2017

REVISED
10 April 2017

ACCEPTED FOR PUBLICATION
10 May 2017

PUBLISHED
28 June 2017

Original content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence.

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.



LETTER

The critical role of the routing scheme in simulating peak river discharge in global hydrological models

Fang Zhao[1,22], Ted I E Veldkamp[2], Katja Frieler[1], Jacob Schewe[1], Sebastian Ostberg[1,3], Sven Willner[1], Bernhard Schauberger[1,4], Simon N Gosling[5], Hannes Müller Schmied[6,7], Felix T Portmann[6], Guoyong Leng[8], Maoyi Huang[8], Xingcai Liu[9], Qiuhong Tang[9], Naota Hanasaki[10], Hester Biemans[11,12], Dieter Gerten[1,3], Yusuke Satoh[13], Yadu Pokhrel[14], Tobias Stacke[15], Philippe Ciais[4], Jinfeng Chang[4], Agnes Ducharne[16], Matthieu Guimberteau[4], Yoshihide Wada[13,17,18,19], Hyungjun Kim[20] and Dai Yamazaki[19,21]

[1] Potsdam Institute for Climate Impact Research, Potsdam, Germany
[2] Institute for Environmental Studies (IVM), VU Amsterdam, Netherlands
[3] Geography Department, Humboldt-Universität zu Berlin, Berlin, Germany
[4] Laboratoire des Sciences du Climat et de l'Environnement, Gif-sur-Yvette, France
[5] School of Geography, University of Nottingham, Nottingham, United Kingdom
[6] Institute of Physical Geography, Goethe-University Frankfurt, Frankfurt, Germany
[7] Senckenberg Biodiversity and Climate Research Centre (BiK-F), Frankfurt, Germany
[8] Atmospheric Sciences and Global Change Division, Pacific Northwest National Laboratory, Richland, WA, United States of America
[9] Key Laboratory of Water Cycle and Related Land Surface Processes, Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, People's Republic of China
[10] National Institute for Environmental Studies, Tsukuba, Japan
[11] Earth System Science and Climate Change, Wageningen University and Research Centre, Wageningen, Netherlands
[12] Environmental Assessment Agency, Bilthoven, Netherlands
[13] International Institute for Applied Systems Analysis, Laxenburg, Austria
[14] Department of Civil and Environmental Engineering, Michigan State University, East Lansing, MI, United States of America
[15] Max-Planck-Institut für Meteorologie, Bundesstraße 53, 20146 Hamburg, Germany
[16] Sorbonne Universités, UMR 71619 METIS, UPMC/CNRS/EPHE, Paris, France
[17] Center for Climate Systems Research, Columbia University, New York, NY, United States of America
[18] NASA Goddard Institute for Space Studies, New York, NY, United States of America
[19] Department of Physical Geography, Utrecht University, Utretcht, Netherlands
[20] Institute of Industrial Science, The University of Tokyo, Tokyo, Japan
[21] Japan Agency for Marine-Earth Science and Technology, Kanagawa, Japan
[22] Author to whom any correspondence should be addressed.

E-mail: fangzhao@pik-potsdam.de

Keywords: ISIMIP, global hydrological models, peak river discharge, river routing, flood, daily runoff, GRDC

Supplementary material for this article is available online

Abstract

Global hydrological models (GHMs) have been applied to assess global flood hazards, but their capacity to capture the timing and amplitude of peak river discharge—which is crucial in flood simulations—has traditionally not been the focus of examination. Here we evaluate to what degree the choice of river routing scheme affects simulations of peak discharge and may help to provide better agreement with observations. To this end we use runoff and discharge simulations of nine GHMs forced by observational climate data (1971–2010) within the ISIMIP2a project. The runoff simulations were used as input for the global river routing model CaMa-Flood. The simulated daily discharge was compared to the discharge generated by each GHM using its native river routing scheme. For each GHM both versions of simulated discharge were compared to monthly and daily discharge observations from 1701 GRDC stations as a benchmark. CaMa-Flood routing shows a general reduction of peak river discharge and a delay of about two to three weeks in its occurrence, likely induced by the buffering capacity of floodplain reservoirs. For a majority of river basins, discharge produced by CaMa-Flood resulted in a better agreement with observations. In particular, maximum daily discharge was adjusted, with a multi-model averaged reduction in bias over about 2/3 of the analysed basin area. The increase in agreement was obtained in both managed and near-natural basins. Overall, this study demonstrates the importance of routing scheme choice in peak discharge simulation, where CaMa-Flood routing

© 2017 IOP Publishing Ltd

accounts for floodplain storage and backwater effects that are not represented in most GHMs. Our study provides important hints that an explicit parameterisation of these processes may be essential in future impact studies.

1. Introduction

Fluvial flooding is one of the most costly natural disasters in the world, claiming thousands of lives and causing billions of dollars in losses globally each year (Munich Re 2016). Most of the losses occur in Asia and Africa, where the frequency of flooding events is projected to increase under a warmer climate (Hirabayashi *et al* 2013, Dankers *et al* 2014, Winsemius *et al* 2015, Arnell and Gosling 2016). Fluvial flood occurs when river discharge exceeds the maximum flowing capacity at a river section. Accurate assessment and realistic projection of peak discharge is critical for risk assessment and adapting to climate change and designing flood protection infrastructures. Hydrological modelling has been a useful tool to provide flood projections under various climate scenarios (Lehner *et al* 2006, Prudhomme *et al* 2014). However, without a calibration for individual river basins global grid-based hydrological models (GHMs) generally perform worse in reproducing the observed discharge than calibrated regional models even at monthly timescale (Hattermann *et al* 2017, Veldkamp *et al* 2017). A first reason is the coarse spatial resolution and uncertainties in global precipitation datasets (Fekete *et al* 2004, Voisin *et al* 2008, Biemans *et al* 2009, Wanders and Wada 2015, Müller Schmied *et al* 2016, 2014). A second reason is the uncertainty related to the land surface model structure, namely the representation of local hydrological processes at the scale of grid cells, such as the partitioning of precipitation into evaporation and runoff, and representation of snow and groundwater (Haddeland *et al* 2011). A third factor is the routing of the simulated grid-cell runoff through the river network, which is a crucial step in simulating river discharge, a variable directly comparable to numerous observations. This physical process is commonly referred to as river routing. Although the effects of river routing scheme including the representation of drainage area (Alkama *et al* 2011) and flow velocity (Decharme *et al* 2010) are considered to be important for simulating realistic river discharge at global scale, there is thus far no comprehensive multi-model evaluation of the sensitivity of simulated discharge, especially of peak discharge, to using different routing schemes.

Routing schemes as a component of land surface models (part of Earth system models) were primarily developed to close the hydrological cycle and compute freshwater input into the ocean, which is important for ocean convection and circulation in atmosphere-ocean models (Manabe and Stouffer 1988). The first models of that kind were capable of simulating a rough water balance, by adopting strong simplifications of the fluvial hydrodynamic equations, often with linear reservoirs and spatially or temporally constant water transfer velocity (Liston *et al* 1994, Miller *et al* 1994, Sausen *et al* 1994, Coe 1997, 2000, Oki 1999). Improved schemes now better simulate flow velocity, e.g. by using a kinematic wave equation combined with the Manning's equation (Arora and Boer 1999, Oki 2003, Decharme *et al* 2010). While these simplified schemes are theoretically capable of representing flood wave delay and attenuation, they often lack water pressure gradient simulations and some of them do not account for processes such as floodplain storage (represented and shown to be important in e.g. Decharme *et al* 2008, 2012, Dadson *et al* 2010) and backwater effect (Meade *et al* 1991), thereby limiting the accuracy of river discharge especially in flat areas and during flooding episodes. The incorporation of routing schemes in GHMs has also allowed for assessments of future changes in hydrological regimes under different climate scenarios from an ensemble of models (e.g. Schewe *et al* 2014). In a continuous process of improvements of routing scheme in some GHMs, simulations have benefited from an increased spatial resolution, in both the topographical inputs and routing algorithm (e.g. Ducharne *et al* 2003, David *et al* 2011, Stacke and Hagemann 2012a, Lehner and Grill 2013) and improved hydrologically conditioned digital elevation models at high resolution, such as HydroSHEDS (Lehner *et al* 2008). This trend reduces the deviation from regional-scale hydraulic models, which have been progressively applied to simulate flood propagation in river systems of increasing size building on simplified or the full Saint-Venant equations (Rudorff *et al* 2014, Paiva *et al* 2013).

Recently, efficient and realistic global hydrodynamic simulations were made possible by the CaMa-Flood global river routing model, which relies on the HydroSHEDS topography and simulates floodplain dynamics and backwater effects by explicitly solving the diffusive wave equation (Yamazaki *et al* 2011), thus providing the potential for improved simulations of daily river discharge than GHMs' native routing schemes, especially for high flow simulations. CaMa-Flood is designed to use gridded runoff output from GHMs as forcing data and also allows for the calculation of flooded areas, flood depth and flood volume critical for global flood risk assessments (e.g. Hirabayashi *et al* 2013).

As mentioned above, differences between simulated and observed discharge could in principle stem

from (1) biases in the observed climate input data, (2) errors in the representation of runoff, and (3) limitations in the derivation of discharge. Here, we focus on an evaluation of the last component, and to reduce the influence of the first two components we (1) consider all of a large set of river basins with very different climatic conditions and potential bias of the observational climate data and (2) use runoff simulations from nine different hydrological models to test for a systematic increase of agreement across a wide range of runoff conditions. However, we can only reduce but not exclude the potential for error compensations that has for instance been found in a study evaluating the implementation of variable flow velocity scheme in a global river routing model TRIP 2.0 (Ngo-Duc *et al* 2007). So far, evaluations did not focus on daily maximum discharge, which is more relevant to flooding. Further, these evaluations only used a single GHM, and there has been no comprehensive, multi-model analysis examining the influence of a different routing scheme on the performance of peak discharge simulation down to daily scale. We perform simulations with CaMa-Flood to produce monthly and daily river discharge, driven by daily runoff from nine GHMs in a naturalized (e.g. without anthropogenic disturbances to the water cycle) experiment under the ISIMIP2a protocol (www.isimip.org). The GHMs additionally simulated monthly and daily river discharge from their own river routing schemes. Both native routing and CaMa-Flood routing discharge are then compared to *in situ* discharge observations worldwide. These comparisons test (1) for which models, and (2) in which regions CaMa-Flood routing brings discharge closer to observation compared to GHM's native routing. In that, evaluation metrics for the performance of the hydrological models focus on the amplitude (maximum discharge) and timing (correlation) of river discharge, both relevant for flood risk assessment. Representation of (1) flow velocities, (2) floodplain storage and (3) ground water storage are expected to be critical for the simulation of peak discharge. CaMa-Flood builds on a detailed implementation of the first two processes going beyond the simplified implementations included in the other routing schemes. However, CaMa-Flood itself does not represent ground water storage—a process that is included in the runoff projections by three of the GHMs. As direct anthropogenic disturbances such as dams and reservoirs also influence observed discharge, which are not represented in the model simulations, we test the agreement between observations and simulations separately for heavily managed basins and near-natural ones.

2. Data and methods

Nine offline routing simulations were performed with CaMa-Flood driven by daily runoff data from nine GHMs. The simulated daily and monthly discharges were then evaluated at 1701 globally distributed discharge gauges, using the data from the Global Runoff Data Centre (GRDC, 56068 Koblenz, Germany). Additionally, each GHM used in this study also provides daily discharge output from its own river routing scheme, which in general has simpler physics compared to CaMa-Flood. This group of discharge output was also evaluated against observations using the same performance metrics. For each GHM and at every discharge gauge location, the difference in performance metrics between the CaMa-Flood simulated discharge and the GHM's own discharge simulations were analysed. The following sections discuss the data and methods used here in detail.

2.1. Modelled runoff and discharge from GHMs' native routing schemes

We used daily and monthly runoff and discharge simulations (0.5° × 0.5° resolution) for the 1971–2010 period from nine GHMs: CLM (Leng *et al* 2015), DBH (Tang *et al* 2007, 2008), H08 (Hanasaki *et al* 2008b, 2008a), LPJmL (Rost *et al* 2008, von Bloh *et al* 2010), MATSIRO (Takata *et al* 2003, Pokhrel *et al* 2012, 2015), MPI-HM (Hagemann and Dümenil Gates 2003, Stacke and Hagemann 2012), ORCHIDEE (Guimberteau *et al* 2014), PCR-GLOBWB (Wada *et al* 2010, van Beek *et al* 2011, Wada *et al* 2011), and WaterGAP2 (Müller Schmied *et al* 2014, 2016), all setup under the framework of the Inter-Sectoral Impact Model Intercomparison Project phase 2a (ISIMIP2a). Each of the GHMs used in this analysis was forced by a daily (three-hourly for MATSIRO) observation-based historical climate dataset GSWP3 (http://hydro.iis.u-tokyo.ac.jp/GSWP3).

Supplementary table S2 (available online at stacks.iop.org/ERL/12/075003/mmedia) lists some general model characteristics regarding both the water and energy budget formulations, and routing schemes of the various GHMs, which mostly rely on linear reservoir algorithms. In the ISMIP2a framework, they all used the same 0.5° drainage direction map (DDM30, Döll and Lehner 2002), with the exception of CLM and PCR-GLOBWB. Note that the routing model of the MPI-HM model includes a wetland scheme with floodplains (Stacke and Hagemann 2012), and in order to have a consistent drainage direction map, the routing model of the MPI-HM model was also used to route the runoff from the ORCHIDEE model.

Except for WaterGAP2, none of the other GHMs has been calibrated with observed discharge data, allowing us to use GRDC observations as a relatively independent benchmark. As CaMa-Flood is not calibrated either, the WaterGAP group conducted additional no calibration (nc) runs (referred to as WaterGAP2nc) in order to allow a consistent comparison. The comparison procedure was conducted for both model versions, but only WaterGAP2nc was used for ensemble statistics and most of

the discussion. However, evaluating the calibrated version allows us to present the effect of model calibration compared to the effect of river routing.

GHMs used in this study were all run under so called naturalized conditions (referred to as 'NOSOC' in the ISIMIP2a protocol), meaning that no human impacts, such as dams and water abstractions on river flow were considered. This facilitates a consistent comparison focusing on the effect of routing scheme, given that CaMa-Flood does not account for human regulation of rivers. Also, for peak daily flow, a previous study showed that for some major basins, the shape of the hydrograph is not significantly different between natural and human impact experiments (Pokhrel *et al* 2012). Comparing the effect of human impacts on peak discharge is difficult at the current stage since not all the GHMs include reservoir operations; also, the rules of reservoir operations (when they are included) is substantially different among GHMs (Masaki *et al* 2017). However, we also used the 'VARSOC' simulations generated within the ISIMIP2a intercomparison to test the sensitivity of our main findings on the implementation of direct human influences. In the 'VARSOC' setting modellers were asked to account for time-varying dam construction, water abstraction and land use changes if their model offers the capacity to do so.

Under the ISIMIP protocol, runoff is defined as the sum of local surface and subsurface runoff within each grid box. Therefore, when routing the runoff along rivers with CaMa-Flood (see section 2.2), precipitation and evaporation over rivers or lakes are not included in the resulting discharge, while they can in some GHMs. Runoff should not be confused with river discharge, which is the result of routing gridded runoff along the river network. The latter is also generated with the same drainage direction map (DDM30, Döll and Lehner 2002) by the GHMs.

2.2. Simulating river discharge with CaMa-Flood

CaMa-Flood computes physically based hydrodynamics at continental-scale, based on a shallow water flow equation allowing for simulating backwater effect (Yamazaki *et al* 2013, 2011). Flow velocity and river discharge are calculated with the local inertial equation (Bates *et al* 2010) along a prescribed river network which is automatically generated with the Flexible Location of Waterways (FLOW) method (Yamazaki *et al* 2009). The 0.25° resolution (about 25 km) river network map was generated by upscaling the HydroSHEDS flow direction map (below 60 N; Lehner *et al* 2008) and the GDBD flow direction map (above 60 N; Masutomi *et al* 2009). The bathymetry of the basins (channel width and bank heights) was empirically determined as a function of maximum 30 day upstream runoff estimated by a land surface model (Kim *et al* 2009). The sub-grid topography of river channel and floodplain storage is explicitly parameterized in each unit-catchment corresponding to each grid box, and water level is assumed to be uniform within each unit-catchment (figure S5). Currently CaMa-Flood is the only open-source global river model available (http://hydro.iis.u-tokyo.ac.jp/~yamadai/cama-flood/) that is capable of simulating backwater effects with a reasonable computation time (7 hours on a 3.2 GHz Intel Xeon E5–2667 v3 processor for one single 40 year simulation), making it a popular choice in studies on the future projection of global flood risk (Pappenberger *et al* 2012, Hirabayashi *et al* 2013, Koirala *et al* 2014). Similar models such as ISBA-TRIP (Decharme *et al* 2012) and HyMAP (Getirana *et al* 2012) have other advantages such as differential time delays for surface and subsurface runoff, however they cannot simulate the backwater effect; their performance could be assessed in future studies.

In this study we used the default setup of CaMa-Flood version 3.4.4 following a common model spin-up procedure of five times repeating the first year (1971) of input daily runoff forcing. An adaptive time step scheme was used in the simulations, leading to a time step of about 10 minutes. The runoff data from the nine GHMs were finally spatially interpolated, conserving mass to the native resolution of CaMa-Flood (0.25°), which is twice as fine as the original GHM spatial resolution.

2.3. Comparison to the observed river discharge

Observed river discharge dataset from GRDC consists of more than 9000 stations in 160 countries, with an average record length of 42 years. We made a selection from this dataset according to the following criteria:

1. A minimum of 5 year coverage during the validation period of 1971–2010. While the data do not need to be continuous, we only use years with less than 10 missing days to avoid under-sampling any month. A threshold of 5 years instead of 10 years is adopted to allow for better spatial coverage, enabling assessment at some important basins such as Indus and Ganges in Asia, and Po in Europe, although a 10 year threshold would be more robust for inter-annual variations. Analyses using stations with a minimum of 10 year coverage show similar results (not shown), nevertheless.
2. A minimum catchment size of 9000 km^2 to omit catchments whose hydrological processes are not correctly represented by GHMs operating at 0.5° (Hunger and Döll 2008).

Using the MIRCA2000 irrigated crop area dataset (Portmann *et al* 2010) and the Global Reservoir and Dam (GRanD) Database (Lehner *et al* 2011), we separated GRDC stations with (near-)natural flow regimes from stations with highly managed flow, based on catchments characteristics. A catchment is classified as (near-) natural if its area subject to irrigation is

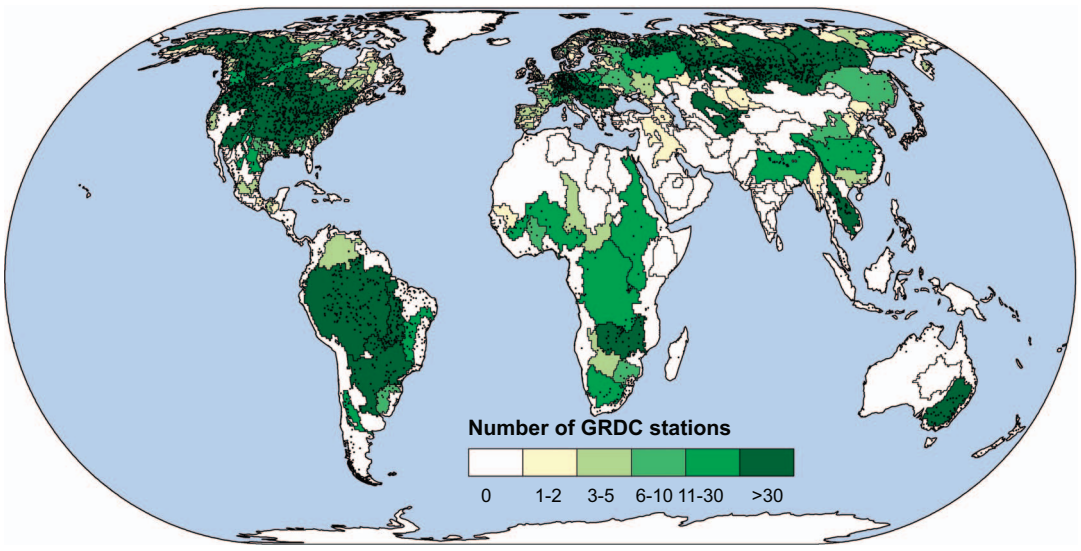


Figure 1. Map of the world's major river basins and selected GRDC stations (black dots) with at least five full years of monthly data (see text). Colours represent number of selected stations in each basin. Note the station scarcity in Asia, where most of flood hazards and exposure occurs.

<2%, and the total reservoir capacity in the catchment is <10% of its long-term mean annual discharge.

As detailed in section 2.4, we examined basin-average performance metrics, in addition to metrics at individual stations to avoid statistical effects from outliers; we chose not to focus on outlet stations, because for flood applications, it is important to have a reasonable representation of high flow at every stage of the river. We also analysed only stations located within major basins worldwide according to the DDM30 network. Grid cells in the DDM30 and CaMa-Flood's network corresponding to each GRDC station were determined through a semi-automatic procedure described in supplementary text C. In total, this selection procedure led to a set of 1701 stations with monthly data (figure 1) and 1205 stations with daily data. These stations belong to 198 (170 for stations with daily data) out of 321 major basins according to DDM30. These 198 basins sampled by at least one selected station cover 70 million km^2 (47% of global land area) area and host 3.8 billion people (55% of 2010 global total according to Gridded Population of the World, version 4, http://sedac.ciesin.columbia.edu/). Of all the selected GRDC stations, 538 (526 for daily) were identified as managed (i.e. having a managed upstream area), and 1130 (646 for daily) as (near-) natural, and the rest has no information regarding upstream area management.

While most models used the DDM30 drainage direction dataset provided in ISIMIP, different networks were used in CLM and PCR-GLOBWB (the latter use an independently adjusted version of DDM30). We nevertheless include them in the analysis for the purpose of maximizing the sample size while minimizing the impact of compromised consistency. For these two models we followed a similar procedure as with DDM30, keeping only stations with no more than 30% (approximately one grid cell for the smallest catchment) difference in upstream area with GRDC, and applying manual correction separately. This led to about 10% (40%) less corresponding cells selected for PCR-GLOBWB (CLM), respectively. As a result, for some stations, the ensemble statistics at station level contain only seven or eight GHMs. We checked that ensemble results excluding these two models are similar to the full ensemble, to make sure the different station sampling does not alter the conclusions of this study.

2.4. Performance metrics

Several metrics were computed at every station to assess the performance of GHM simulated discharge (D_H) and discharge using CaMa-Flood forced by GHM-runoff (D_C) versus GRDC observed daily and monthly discharge (D_O). The metrics include Pearson's correlation and percent biases of mean and standard deviation, as well as Nash–Sutcliffe Efficiency (NSE) for overall performance. We also included percent bias of multi-year maximum and mean annual maximum discharge as indicators of models' ability to simulate extreme discharge. Detailed description for the metrics is given in table 1.

After computing the metrics for pairs of GHM vs. observation and CaMa-Flood vs. observation, difference of each metric was assessed at GRDC stations. We consider NSE < 0, R < 0 and percent biases >100% as poor skill, and before computing the differences or basin-average metrics, we first capped NSE and R at 0 and the absolute value of percent biases at 100% (for example, change of NSE from -100 to -1 would be considered no change in performance, as both are poor). With this, performance difference at any station is limited within -1 and 1, preventing a single station to have a large effect in basin-average metrics.

Table 1. Description of performance metrics examined. $D_{H,C}$ represents discharge from GHM model output (D_H) or CaMa-Flood (D_C), computed separately with the same equations; D_O is GRDC-observed monthly or daily discharge. Time is represented by t (month or day) with a station-dependent step of T; N represents station-dependent record length in years. For BMAX, *max* indicates maximum discharge of the whole time series. For BMYM, max_i indicates annual maximum in year i. Overbar represents mean value of the whole time series.

Abbreviation	Name	Calculation procedure	Performance difference
NSE	Nash–Sutcliffe Efficiency	$\mathrm{NSE}_{H,C} = 1 - \frac{\sum_{t=1}^{T}\left(D_O^t - D_{H,C}^t\right)^2}{\sum_{t=1}^{T}\left(D_O^t - \overline{D_O}\right)^2}$	$\mathrm{dNSE} = \mathrm{NSE}_C - \mathrm{NSE}_H$
R	Pearson's Correlation	$R_{H,C} = \frac{\sum_{t=1}^{T}(D_O^t - \overline{D_O})(D_{H,C}^t - \overline{D_{H,C}})}{T\sigma_{H,C} * \sigma_O}$	$\mathrm{dR} = R_C - R_H$
BMEAN	Percent bias of means	$\mathrm{BMEAN}_{H,C} = \frac{\overline{D_{H,C}} - \overline{D_O}}{\overline{D_O}} * 100\%$	$\mathrm{dBMEAN} = \lvert\mathrm{BMEAN}_C\rvert - \lvert \mathrm{BMEAN}_H\rvert$
BSTD	Percent bias of standard deviation	$\mathrm{BSTD}_{H,C} = \frac{\sigma_{H,C} - \sigma_o}{\sigma_o} * 100\%$	$\mathrm{dBSTD} = \lvert\mathrm{BSTD}_C\rvert - \lvert\mathrm{BSTD}_H\rvert$
BMAX	Bias of overall maxima	$\mathrm{BMAX}_{H,C} = \frac{max(D_{H,C}) - max(D_O)}{max(D_O)} * 100\%$	$\mathrm{dBMAX} = \lvert\mathrm{BMAX}_C\rvert - \lvert\mathrm{BMAX}_H\rvert$
BMYM	Mean bias of annual maxima	$\mathrm{BMYM}_{H,C} = \frac{\sum_{i=1}^{N}(max_i(D_{H,C}) - max_i(D_O))}{\sum_{i=1}^{N} max_i(D_O)} * 100\%$	$\mathrm{dBMYM} = \lvert\mathrm{BMYM}_C\rvert - \lvert\mathrm{BMYM}_H\rvert$

Although this capping does not allow for assessing differences between GHM and CaMa-Flood at very poorly performing stations (22% of the stations do not contribute to NSE difference due to capping; for other metrics less than 3% stations are effectively excluded in computing the differences), it is unlikely to have a large impact in model applications if discharge is poorly simulated in both cases; e.g. NSE < 0 suggests the model skill is worse than using observed mean (Krause *et al* 2005).

Based on the location of the stations in the 198 major river basins, we then computed the average performance metrics for each GHM and the performance difference between GHM and CaMa-Flood over every river basin. We considered the stations as samples of the rivers and computed the basin averages without weighting, and it is worth noting that the sampling may not be representative especially in data sparse basins. Multi-model ensemble mean of the performance differences were also calculated, for daily and monthly values respectively. For biases, where values close to 0 indicate good agreement with observation, performance difference was computed based on absolute values (table 1), so that any positive difference indicates an improvement. Finally, multi-model mean and best-model performance metrics were computed to show where the models can simulate a reasonable discharge compared to the GRDC observations.

Additionally, before comparing to observations, we examined how daily river discharge simulated by CaMa-Flood differs from the GHMs' original discharge at the location of selected GRDC stations, both in timing and amplitude. In order to filter out inter-annual variability in the timing of annual maximum daily discharge, multi-year mean daily discharge for each of the 365 days (excluding February 29 from data) were computed to produce mean daily hydrograph during 1971–2010 (referred to as climatological daily discharge) at each station. To define a metrics for timing differences between discharge simulated by CaMa-Flood and by the GHMs' native routing schemes, differences in number of days on when the simulated climatological discharge maximum occurs in the mean daily hydrograph between CaMa-Flood and GHM were calculated for each GRDC station. Numbers are shifted by a year, if necessary, to yield differences of no more than half a year. At some stations, small upstream area or short record length may cause an unrealistic time lag between the two maxima in some models. To prevent this from affecting the model ensemble mean, all absolute difference values greater than 150 days were excluded when computing the ensemble mean. Regarding amplitude we simply compared the multi-year mean annual maximum discharges based on CaMa-Flood and GHMs' native routine schemes. The results are reported in section 3.1.

3. Results

3.1. Effect of CaMa-Flood routing on the timing and amplitude of the simulated discharge

For the majority of GHMs, CaMa-Flood routing generally produces a delayed climatological daily maximum discharge (figure 2) and a reduction in the amplitude of maximum river discharge at most stations (figure 3). Time series for individual river basins are shown in the supplementary data (figures S9–S11).

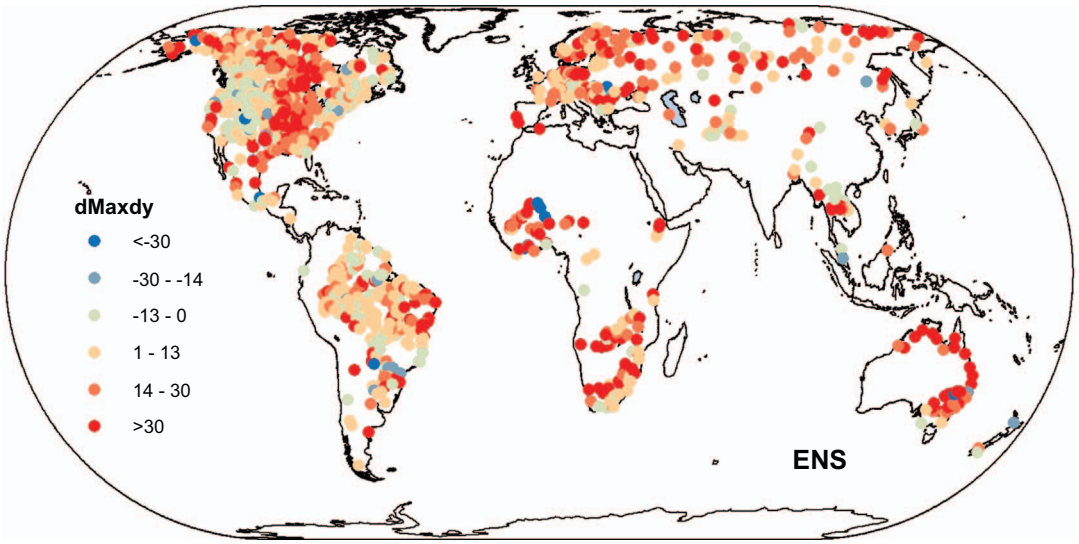


Figure 2. Multi-model ensemble mean changes in timing of climatological daily maximum discharge simulated by CaMa-Flood compared to GHMs' original discharge. A positive value indicates max discharge occurring later in CaMa-Flood.

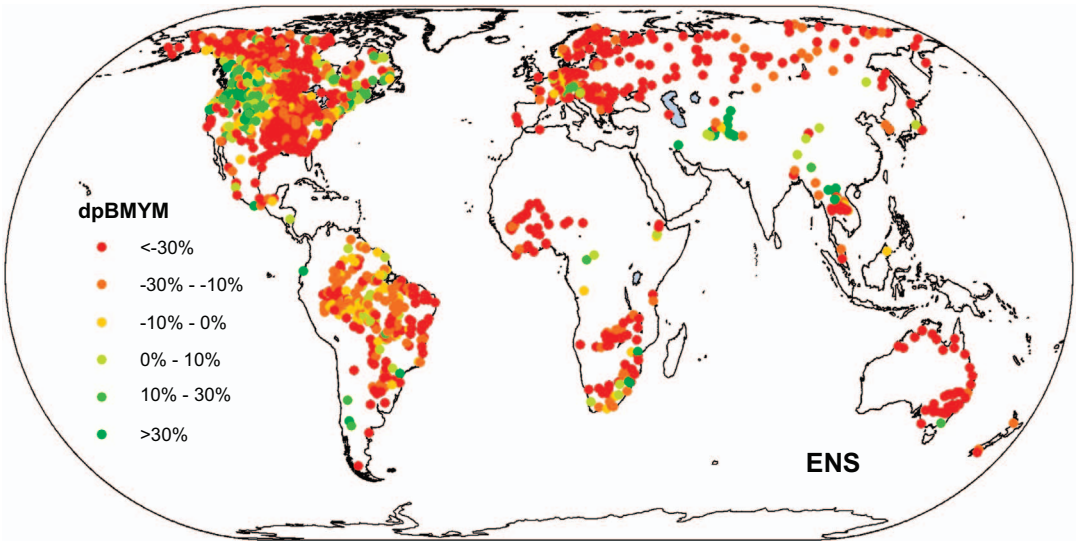


Figure 3. Multi-model ensemble mean percent changes of multi-year mean daily maximum discharge simulated by CaMa-Flood compared to GHMs' original discharge. A positive value indicates increased BMYM in CaMa-Flood.

Both effects are likely due to floodplain processes being present in CaMa-Flood and not in most of the GHMs examined (note that LPJmL and ORCHIDEE both have a version simulating floodplain processes but not used in ISIMIP simulations). The mean delay with CaMa-Flood is 18 days over all stations (with a large standard deviation of 24 days due to site specific characteristics), and the multi-year mean annual maximum discharge is 22% lower than in original GHMs. It is opposite for two of the models (ORCHIDEE and MPI-HM), where dynamic wetland processes, which include floodplain interactions, are already part of the native routing schemes (Stacke and Hagemann 2012), causing similar effects on timing and even stronger amplitude reduction of the peak discharge at many stations compared to CaMa-Flood's floodplain mechanism.

3.2. Change in performance of hydrological models when using CaMa-Flood routing

With alteration in both timing and amplitude of discharge shown in figures 2 and 3, noticeable performance changes can be identified with CaMa-Flood for most of the metrics (except BMEAN, for which no change is expected from different routing method when the upstream area is identical). We consider a change significant at the river basin scale, when the difference in the performance metrics was greater than 0.05 (or 5% for the percent changes). The ensemble mean of performance differences show that for all metrics related to peak flow (BSTD, BMAX, BMYM), most basins display a significant increase in agreement of more than 5% with CaMa-Flood routing. Basins where CaMa-Flood improved BMYM cover

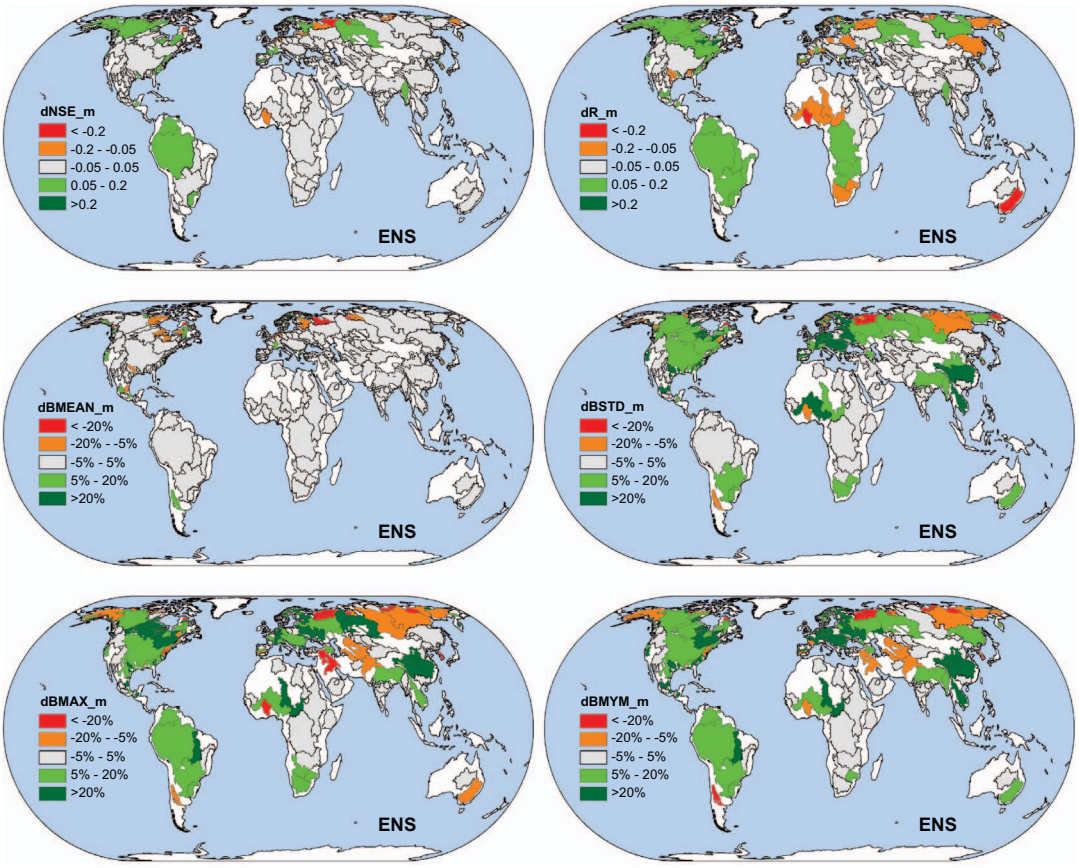


Figure 4. Multi-model ensemble mean performance differences compared to daily GRDC data, all shown as basin averages (denoted by _m). Grey colour shows differences <5% in basin-averaged performance metrics. Green colours show basins where a discharge metrics is improved with CaMa-Flood compared to native GHM routing.

Table 2. Percentage land area for which a significantly higher (left)/lower (right) agreement with observations in their basin-average representation of BMYM (over 0.05 or 5% difference) with CaMa-Flood routing is achieved compared to GHMs' native routing, using all studied stations, managed stations only, and (near-)natural stations only. Numbers in bold indicate better agreement with observations. The results for correlation (*R*) and NSE are presented in table S3 and S4.

	All stations (%)	Managed stations only (%)	Natural stations only (%)
CLM	**44**/31	39/**40**	**36**/31
DBH	**74**/15	**73**/1	**64**/25
H08	**46**/33	**54**/19	**43**/38
LPJmL	**79**/7	**80**/1	**80**/9
MATSIRO	42/42	32 / **47**	38/**44**
MPI-HM	24/**45**	26/**42**	29/**46**
ORCHIDEE	34/**37**	34/34	**37**/35
PCR-GLOBWB	**65**/26	**65**/16	**55**/33
WaterGAP2nc	**44**/35	**48**/31	33/**42**
WaterGAP2	22/**62**	**45**/36	17/**56**
ENS[a]	**63**/18	**69**/7	**60**/22

[a] Note ensemble (ENS) uses the uncalibrated (WaterGAP2nc) instead of calibrated version of WaterGAP2.

63% of all selected basins' area (figure 4). The improvement is more pronounced (>20% reduction of biases) in Southeast and East Asia, East Europe, and in some basins in tropical Africa and South America. Significant decreases in the peak flow metrics (CaMa-Flood routing worse than GHMs) are only found in a few basins (18% of the studied basin area) in Central and West Asia and in Eastern Siberia. In comparison, CaMa-Flood routing does not lead to a clear improvement on the correlation (*R*) metrics (table S3). Even though most GHMs in general simulate an earlier peak discharge compared to observations (figure S6), the delay in peak discharge with CaMa-Flood routing leads to a later peak (figure S7) that does not necessarily increase agreement with observations. Instead, the delay is probably too long, given that the model ensemble mean of absolute error in timing of peak discharge, averaged over 623 stations with over 30 years of observational record, becomes slightly larger (42 days) with CaMa-Flood compared to GHMs' native

routing schemes (36 days). As expected, little change in BMEAN is seen, and only a few basins display a significant change in NSE (table S4). Analyses on monthly discharge show similar results, although the improvement in peak discharge simulation is less pronounced (figure S8).

For the majority of GHMs, the larger part of the studied basin areas shows significantly better instead of worse agreement with observation in peak discharge simulation with CaMa-Flood (table 2, using BMYM as an example here as the statistics are similar for BSTD and BMAX). The largest improvement (and major contribution to ensemble mean) is observed in DBH and LPJmL. MATSIRO and ORCHIDEE show approximately the same area of increase and decrease in BMYM performance. The impact of CaMa-Flood routing on BMYM is largely negative for MPI-HM compared to its native routing scheme, and this is similar for ORCHIDEE (which is 'natively' routed by the MPI-HM scheme) to a lesser extent. Since MPI-HM includes a dynamic wetland scheme that could encompass larger buffering zones than floodplains, this result confirms that the gain in performance can be expected from better describing the hydraulic connection between the streams and the riparian areas. As a result, CaMa-Flood allows a large improvement of the simulated discharge if BMYM is originally large (too weak buffering with the original routing scheme), and slightly deteriorates the absolute bias in BMYM if GHMs' original discharge simulations are already relatively good (table S5). An interesting feature is that the area-weighted overall BMYM varies from 36 to 84% for the GHMs' original discharge, which spans a much larger inter-model range than the 43%–52% obtained with CaMa-Flood routing. This suggests that a significant portion of the variance among GHMs' performances in peak discharge simulation can be attributed to different routing schemes, including the effects of surface water bodies.

Similar conclusions can be drawn using only stations affected by human management, stations that measure near-natural discharges, or the full sample (table 2, figures not shown). For managed stations only, more models show larger area where CaMa-Flood routing leads to better agreement with observations. Notably for WaterGAP2, the larger part of the studied basin areas shows significantly better agreement with managed stations in peak discharge simulation with CaMa-Flood, but the opposite is true with natural stations only.

4. Discussion

In this study we assessed whether and where the CaMa-Flood global river routing scheme provides a closer agreement with observed discharge compared to the native routing schemes of nine GHMs. To our knowledge, this is a first multi-model comparison study using daily discharge observations, focusing on GHM performance in peak discharge simulations. Quite notable improvement of simulated peak discharge over large part of the world is found with CaMa-Flood routing compared to most of the GHMs' native routing schemes. As revealed by the daily hydrographs from three major river basins (supplementary text A), this improvement is primarily achieved through reducing the amplitude of peak discharge, in cases that peak discharge is overestimated by the GHMs, most likely due to floodplain mechanism in CaMa-Flood. Previous observational and modelling case studies have well documented the role of floodplains in reducing downstream peak discharge, largely through storage (or greatly reduced velocity) of a portion of the runoff on overbank surfaces (Ahilan *et al* 2016, Lininger and Latrubesse 2016, Acreman and Holden 2013, Woltemade and Potter 1994). CaMa-Flood represents this effect at global scale, with floodplain storage simulation and a higher roughness coefficient for floodplains. The variable flow speed simulation that accounts for backwater effect in CaMa-Flood may also play a role. Sensitivity experiments (Paiva *et al* 2013, Yamazaki *et al* 2011) showed that while backwater effect is important for representing more realistic floodplain storage and hydrological regimes (such as water surface profiles and flood extent) for Amazon, the simulated peak discharge with backwater effect is not drastically different compared to kinematic wave approach; in contrast, experiment with no floodplain produces daily discharge that is noisy and in advance, with a notably larger peak amplitude.

In addition, analyses on three selected GHMs (H08, LPJmL and PCR-GLOBWB) show that the CaMa-Flood simulated peak discharge compares favourably over the GHMs' native routing (supplementary text B), even when human impacts are included in the latter. These results indicate that an explicit representation of floodplains as included in CaMa-Flood may significantly improve the simulation of peak discharge. Model calibration (not implemented in CaMa-Flood) could potentially compensate for the lack of explicit floodplain hydrology as indicated by the good performances of the calibrated GHM WaterGAP2 (e.g. table 2). This is interesting, as the calibration scheme of WaterGAP2 only forces the simulated long-term average annual river discharge to observed values (Müller Schmied *et al* 2014) and is not directly designed to reproduce peak discharge at a daily time step. From a more general perspective, the differences in NSE or R between CaMa-Flood simulated discharge and discharge from GHMs' native schemes are relatively small for a majority of models. It is important to note that our assessment focuses on peak discharge while in some water resource applications low levels of discharge are more critical. The CaMa-Flood's representation of low flow conditions has to be assessed separately. As low flow

conditions could be particularly sensitive to human management (Veldkamp *et al* 2017), it may be essential to account for processes such as water withdrawal (agricultural, domestic, industrial, etc.) and flow alteration due to dams and reservoirs in simulating discharge. These effects are currently not included in CaMa-Flood.

While our results highlight the importance of representing floodplains in routing models, future research on the attribution of factors such as floodplain storage, roughness, connectivity and backwater effect in affecting peak flow will be helpful for further model improvement. Large differences among individual basins including varied rainfall regimes and river channel characteristics, distribution of floodplains, synchrony between tributary and main channel peaks (Acreman and Holden 2013), and channel connectivity, for example reported in observational studies on Amazon and Congo (Alsdorf *et al* 2010, Jung *et al* 2010), have to be taken into consideration. Nevertheless, we performed a simple analysis in order to further examine whether the storage capacity of floodplains is related to peak discharge reduction with CaMa-Flood routing. Specifically, we computed daily basin totals of flood storage and runoff for 34 basins ($>100\,000$ km^2) worldwide. For each basin we used the multi-year mean (1971–2010) of annual range of daily flood storage as a fraction of multi-year mean annual runoff to represent its floodplain storage capacity. Model ensemble medians of this capacity indicator and peak discharge (using dpBMYM here as in figure 3) changes at the outlets of the basins were then derived for five models that do not represent floodplains (DBH, H08, LPJmL, MATSIRO, WaterGAPnc; CLM and PCR-GLOBWB are excluded for this exercise due to their different routing networks). We saw a clear inverse relationship, suggesting that if the CaMa-Flood simulated annual fluctuation of flood storage is large relative to basin total runoff, a larger amplitude reduction of peak discharge at the basin outlet is more likely to be seen. Note that the relationship is likely not linear, as when the storage is large enough, its further increase may not lead to further reduction of peak discharge amplitude. Water stored on floodplain also does not always reduce peak flows especially if the storage is saturated during long wet period.

In addition to discharge, CaMa-Flood is capable of providing other key outputs for risk assessments such as flooded area, flood volume and flood depth. Some regional flood inundation models can represent physical processes in even more detail and could potentially be applied at the global scale (e.g. Sampson *et al* 2015). However, a global simulation with regional models is often limited to simulations of specific magnitude floods (e.g. a 100 year flood) due to computational constraints. When using runoff from multiple GHMs as input, CaMa-Flood can ensure that the simulated discharge follows a consistent drainage direction network that is similar to the established DDM30 network; this consistency alone is worth noting as it prevents potential location errors in using discharge results from multiple models. We also observed that the difference in overall performance among GHMs (between-GHMs spread) seems to be reduced with CaMa-Flood routing, suggesting that the routing scheme is an important contributing factor explaining the spread between models in simulated peak discharge. For all the GHMs we examined, CaMa-Flood generally leads to improved peak discharge simulations at a global scale except for MPI-HM and ORCHIDEE. In contrast to the other model specific routing schemes, these two models have their own representation of floodplain dynamics, which supports the hypothesis that such dynamics (also represented in CaMa-Flood) are critical for the representation of peak discharge. CaMa-Flood also enables the inclusion of more land or earth system models that do not have a routing scheme in global flood studies, in doing so accounting better for uncertainties in simulating land processes. Additionally, our study provides a model evaluation framework on peak discharge, which is often not directly addressed in global flood studies (e.g. Hirabayashi *et al* 2013). For an estimate on future flood hazards, it is advisable to use multiple GHMs as their performance can vary considerably depending on the basin of interest, due to their different process representation and generalized parameterisation. Recent research is exploring novel methods for utilizing ensemble model results (Zaherpour *et al* 2017); here we also showed that taking the best performing ensemble member at each station limits the basin-averaged biases in peak discharge to within 30% for most of the examined basins (figure 5). Therefore, even though global-scale flood simulations are still at an early stage, it is possible to gain insights using the current ensemble of modelling tools for future flood projections. Further improvement can be achieved by improving routing alone, for example, CaMa-Flood currently uses relatively low resolution in the terrain data north of 60°N for generating routing direction, limiting the accuracy of simulations in boreal regions; implementation of a higher quality topography input is currently under development. Neither evaporation over floodplains nor transmission loss is considered in CaMa-Flood, which could possibly lead to an overestimation of peak amplitude in basins under hot climate, such as in the inner Niger delta (Zwarts *et al* 2005), Okavango and Sudd (Sutcliffe and Parks 1989), and Parana (Decharme *et al* 2012). Another aspect is the globally universal parameterisation of processes such as floodplain inundation, which in principle cannot fit all basins due to their varied physical characteristics and hydrological processes (Jung *et al* 2010). Model calibration generally mitigates this problem and leads to better simulated results compared to observation (Hattermann *et al*

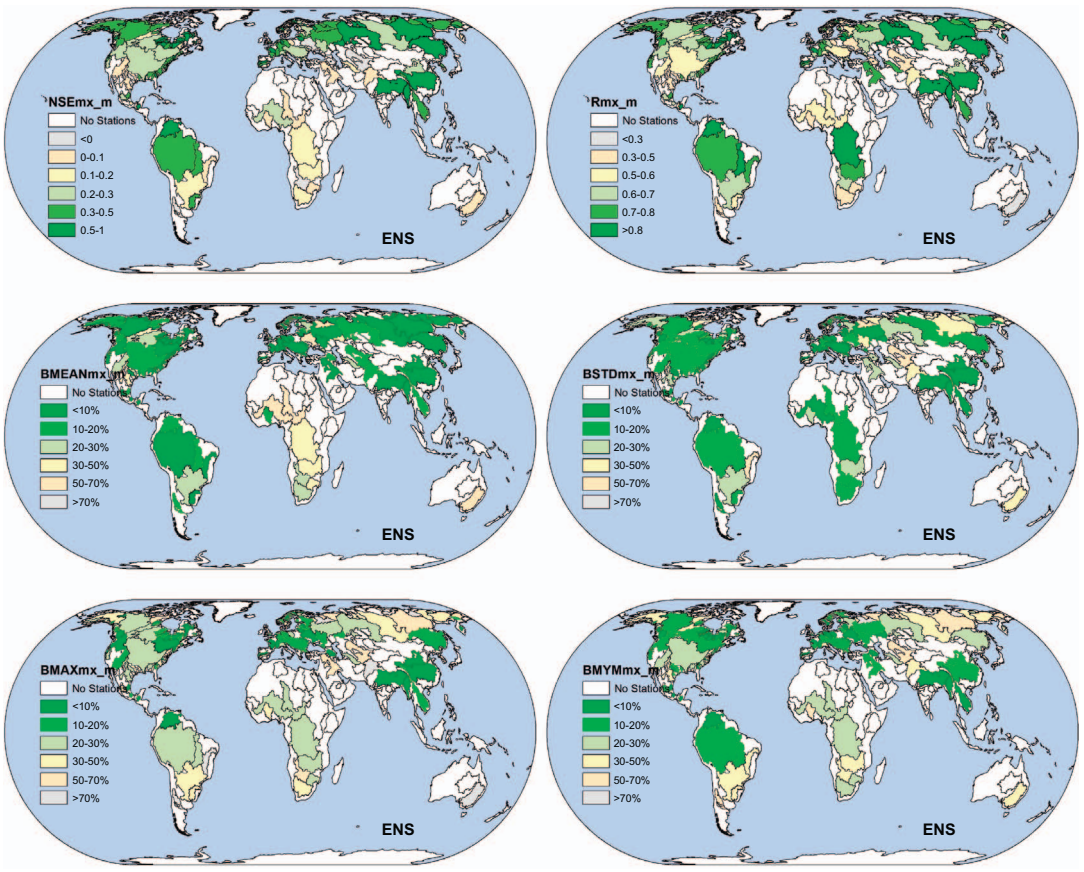


Figure 5. Basin-averaged best performance (denoted by mx_m) from ensemble of GHMs with CaMa-Flood routing for the six metrics (NSE, *R*, BMEAN, BSTD, BMAX, BMYM).

2017); this is also planned in further development of CaMa-Flood.

We also found that CaMa-Flood routing often shows considerable improvement in the cases where NSE<0 or biases >100% (results not shown), beyond which the performance metrics were capped for the analyses. It could be interesting from a model development perspective to also examine such improvement; however, regarding on the potential of model applications limited value can be gained if the improved NSE is still below 0. Additionally, we note that the GRDC stations used in this study are not evenly distributed across the globe, with relatively sparse samples in Asia and Central Africa, where flood risk is high. This prompted us to relax data requirement to a minimum of five years, which allowed us to perform analyses on about 10% more stations and cover a few more basins. However, results in data sparse basins are naturally less robust, have sampling bias and should be treated with caution. We also performed sensitivity analyses using only stations with a minimum record of ten years (results not shown) instead of five; here, our major conclusions still hold. The classification of managed and (near-) natural stations was based on two parameters (reservoir volume, irrigated area), while in the real world human impacts may be present in some other forms that are difficult to measure correctly. While currently being the best tool available, the GHMs' representation of time-varying human impacts is still at an early stage and will continue to improve in the future. Therefore, while our results show that human impacts implemented in the examined GHMs have a quite limited influence on peak discharge simulations, human interventions could be significant for certain regions in the real world.

5. Conclusions

The recently developed global river routing scheme CaMa-Flood, was shown to generally delay the timing and reduce the amplitude of daily peak discharge in most regions compared to the routing schemes in a majority of GHMs. The amplitude reduction led to an overall improvement in simulated peak discharge compared to the GHMs' native routing schemes. Similar levels of increased performance were observed when analysing managed or near-natural stations alone; also, sensitivity analysis displayed a very limited effect of human impacts on peak discharge. The improved peak discharge simulation is likely due to explicit simulation of the floodplain in CaMa-Flood that smoothens the daily hydrographs in a way similar to wetland implementations in the MPI-HM routing scheme. This buffering effect was also

found to largely reduce inter-model differences in overall performance.

Our analyses showed that the choice of routing scheme has a considerable influence on the simulated river discharge and its peak values, and suggested that GHMs could be useful tools in combination with routing schemes such as CaMa-Flood to simulate peak discharge. By adopting a multi-model approach, uncertainties in assessment with only a single model due to missing physical mechanisms and globally universal parameters could be largely mitigated. The benefit of a consistent routing also allows the inclusion of more models thus better accounting for structural model uncertainties in future flood-related studies. With a focus towards flood hazards simulations and projections, this study offers a comprehensive framework to assess how well the models simulate peak discharge, thus allowing a more informed interpretation on their future projections, which could be helpful for decision-makers. The improved multi-model database of historical daily discharge generated from this study will be made available and serve as a reference on the skills of models for studies projecting future flood disasters under a changing climate.

Acknowledgments

The Global Runoff Data Centre (GRDC, 56068 Koblenz, Germany) are thanked for providing the observed discharge data. This work has been conducted under the framework of phase 2 of the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP2a: www.isimip.org). The work was supported within the framework of the Leibniz Competition (SAW-2013 P IK-5), from EU FP7 project HELIX (grant no. FP7–603864-2), and by the German Federal Ministry of Education and Research (BMBF, grant no. 01LS1201A1 and 01LS1201A2). G Leng and M Huang were supported by the Integrated Assessment Research program through the Integrated Multi-sector, Multi-scale Modeling (IM3) Scientific Focus Area (SFA) sponsored by the Biological and Environmental Research Division of Office of Science, US Department of Energy (DOE). The Pacific Northwest National Laboratory (PNNL) is operated for the US DOE by Battelle Memorial Institute under contract DE-AC05-76RL01830. P Ciais and M Guimberteau were supported by the European Research Council Synergy grant ERC-2013-SyG 610028-IMBALANCE-P. The publication of this article was funded by the Open Access Fund of the Leibniz Association. We thank the two reviewers for their constructive and insightful comments.

References

Acreman M and Holden J 2013 How wetlands affect floods *Wetlands* **33** 773–86

Ahilan S, Guan M, Sleigh A, Wright N and Chang H 2016 The influence of floodplain restoration on flow and sediment dynamics in an urban river *J. Flood Risk Manag.* accepted (10.1111/jfr3.12251)

Alkama R, Decharme B, Douville H and Ribes A 2011 Trends in global and basin-scale runoff over the late twentieth century: methodological issues and sources of uncertainty *J. Clim.* **24** 3000–14

Alsdorf D, Han S C, Bates P and Melack J 2010 Seasonal water storage on the Amazon floodplain measured from satellites *Remote Sens. Environ.* **114** 2448–56

Arnell N W and Gosling S N 2016 The impacts of climate change on river flood risk at the global scale *Clim. Change* **134** 387–401

Arora V K and Boer G J 1999 A variable velocity flow routing algorithm for GCMs *J. Geophys. Res.* **104** 30965–79

Bates P D, Horritt M S and Fewtrell T J 2010 A simple inertial formulation of the shallow water equations for efficient two-dimensional flood inundation modelling *J. Hydrol.* **387** 33–45

van Beek L P H, Wada Y and Bierkens M F P 2011 Global monthly water stress: 1. Water balance and water availability *Water Resour. Res.* **47** W07517

Biemans H, Hutjes R W A, Kabat P, Strengers B J, Gerten D and Rost S 2009 Effects of precipitation uncertainty on discharge calculations for main river basins *J. Hydrometeorol.* **10** 1011–25

von Bloh W, Rost S, Gerten D and Lucht W 2010 Efficient parallelization of a dynamic global vegetation model with river routing *Environ. Model. Softw.* **25** 685–90

Coe M T 2000 Modeling terrestrial hydrological systems at the continental scale: testing the accuracy of an atmospheric GCM *J. Clim.* **13** 686–704

Coe M T 1997 Simulating continental surface waters: an application to Holocene Northern Africa *J. Clim.* **10** 1680–9

Dadson S J, Ashpole I, Harris P, Davies H N, Clark D B, Blyth E and Taylor C M 2010 Wetland inundation dynamics in a model of land surface climate: evaluation in the Niger inland delta region *J. Geophys. Res. Atmos.* **115** 1–7

Dankers R *et al* 2014 First look at changes in flood hazard in the inter-sectoral impact model intercomparison project ensemble *Proc. Natl Acad. Sci. USA* **111** 3257–61

David C H, Maidment D R, Niu G-Y, Yang Z-L, Habets F and Eijkhout V 2011 River network routing on the NHDPlus dataset *J. Hydrometeorol.* **12** 913–34

Decharme B, Alkama R, Douville H, Becker M and Cazenave A 2010 Global evaluation of the ISBA-TRIP continental hydrological system. Part II: Uncertainties in river routing simulation related to flow velocity and groundwater storage *J. Hydrometeorol.* **11** 601–17

Decharme B, Alkama R, Papa F, Faroux S, Douville H and Prigent C 2012 Global off-line evaluation of the ISBA-TRIP flood model *Clim. Dyn.* **38** 1389–412

Decharme B, Douville H, Prigent C, Papa F and Aires F 2008 A new river flooding scheme for global climate applications: off-line evaluation over South America *J. Geophys. Res. Atmos.* **113** 1–11

Döll P and Lehner B 2002 Validation of a new global 30 m in drainage direction map *J. Hydrol.* **258** 214–31

Ducharne A, Golaz C, Leblois E, Laval K, Polcher J, Ledoux E and De Marsily G 2003 Development of a high resolution runoff routing model, calibration and application to assess runoff from the LMD GCM *J. Hydrol.* **280** 207–28

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

Getirana A, Boone A, Yamazaki D, Decharme B, Papa F and Mognard N 2012 The hydrological modeling and analysis platform (HyMAP): evaluation in the Amazon Basin *J. Hydrometeorol.* **13** 1641–65

Guimberteau M, Ducharne A, Ciais P, Boisier J P, Peng S, De Weirdt M, Verbeeck H and Biology E 2014 Testing conceptual and physically based soil hydrology schemes against observations for the Amazon Basin *Geosci. Model Dev.* **7** 1115–36

Haddeland I *et al* 2011 Multimodel estimate of the global terrestrial water balance: setup and first results *J. Hydrometeorol.* **12** 869–84

Hagemann S and Dümenil Gates L 2003 Improving a subgrid runoff parameterization scheme for climate models by the use of high resolution data derived from satellite observations *Clim. Dyn.* **21** 349–59

Hanasaki N, Kanae S, Oki T, Masuda K, Motoya K, Shirakawa N, Shen Y and Tanaka K 2008a An integrated model for the assessment of global water resources—Part 1: model description and input meteorological forcing *Hydrol. Earth Syst. Sci.* **12** 1007–25

Hanasaki N, Kanae S, Oki T, Masuda K, Motoya K, Shirakawa N, Shen Y and Tanaka K 2008b An integrated model for the assessment of global water resources—Part 2: Applications and assessments *Hydrol. Earth Syst. Sci.* **12** 1027–37

Hattermann F *et al* 2017 Cross-scale intercomparison of climate change impacts simulated by regional and global hydrological models in eleven large river basins *Clim. Change* **141** 561–76

Hirabayashi Y, Mahendran R, Koirala S, Konoshima L, Yamazaki D, Watanabe S, Kim H and Kanae S 2013 Global flood risk under climate change *Nat. Clim. Change* **3** 816–21

Hunger M and Döll P 2008 Value of river discharge data for global-scale hydrological modeling *Hydrol. Earth Syst. Sci.* **12** 841–61

Jung H C, Hamski J, Durand M, Alsdorf D, Hossain F, Lee H, Azad Hossain A K M, Hasan K, Khan A S and Zeaul Hoque A K M 2010 Characterization of complex fluvial systems using remote sensing of spatial and temporal water level variations in the Amazon, Congo, and Brahmaputra rivers *Earth Surf. Process. Landforms* **35** 294–304

Kim H, Yeh P J F, Oki T and Kanae S 2009 Role of rivers in the seasonal variations of terrestrial water storage over global basins *Geophys. Res. Lett.* **36** 2–6

Koirala S, Hirabayashi Y, Mahendran R and Kanae S 2014 Global assessment of agreement among streamflow projections using CMIP5 model outputs *Environ. Res. Lett.* **9** 064017

Krause P, Boyle D P and Bäse F 2005 Advances in geosciences comparison of different efficiency criteria for hydrological model assessment *Adv. Geosci.* **5** 89–97

Lehner B, Döll P, Alcamo J, Henrichs T and Kaspar F 2006 Estimating the impact of global change on flood and drought risks in Europe: a continental, integrated analysis *Clim. Change* **75** 273–99

Lehner B and Grill G 2013 Global river hydrography and network routing: baseline data and new approaches to study the world's large river systems *Hydrol. Process.* **27** 2171–86

Lehner B *et al* 2011 High-resolution mapping of the world's reservoirs and dams for sustainable river-flow management *Front. Ecol. Environ.* **9** 494–502

Lehner B, Verdin K and Jarvis K 2008 New global hydrograghy derived from spaceborne elevation data *Eos, Trans. AGU* **89** 93–4

Leng G, Huang M, Tang Q and Leung L R 2015 A modeling study of irrigation effects on global surface water and groundwater resources under a changing climate *J. Adv. Model. Earth Syst.* **7** 1285–304

Lininger K B and Latrubesse E M 2016 Flooding hydrology and peak discharge attenuation along the middle Araguaia river in central Brazil *Catena* **143** 90–101

Liston G E, Sud Y C, Wood E F, Liston G E, Sud Y C and Wood E F 1994 Evaluating GCM land surface hydrology parameterizations by computing river discharges using a runoff routing model: application to the Mississippi Basin *J. Appl. Meteorol.* **33** 394–405

Manabe S and Stouffer R J 1988 Two stable equilibria of a coupled ocean-atmosphere model *J. Clim.* **1** 841–66

Masaki Y, Hanasaki N, Biemans H, Müller Schmied H, Tang Q, Wada Y, Gosling S N, Takahashi K and Hijioka Y 2017 Intercomparison of global river discharge simulations focusing on dam operation—multiple models analysis in two case–study river basins, Missouri–Mississippi and Green–Colorado *Environ. Res. Lett.* **12** 055002

Masutomi Y, Inui Y, Takahashi K and Matsuoka Y 2009 Development of highly accurate global polygonal drainage basin data *Hydrol. Process.* **23** 572–84

Meade R H, Rayol J M, Da Conceicao S C and Natividade J R G 1991 Backwater effects in the Amazon river basin of Brazil *Environ. Geol. Water Sci.* **18** 105–14

Miller J R, Russell G L and Caliri G 1994 Continental-scale river flow in climate models *J. Clim.* **7** 914–28

Müller Schmied H *et al* 2016 Variations of global and continental water balance components as impacted by climate forcing uncertainty and human water use *Hydrol. Earth Syst. Sci.* **20** 2877–98

Müller Schmied M, Eisner H, Franz S, Wattenbach D, Portmann M, Flörke F T and Döll P 2014 Sensitivity of simulated global-scale freshwater fluxes and storages to input data, hydrological model structure, human water use and calibration *Hydrol. Earth Syst. Sci.* **18** 3511–38

Munich Re 2016 Munich Re Insurance Company Geo Risks Research *NatCatSERVICE Database*

Ngo-Duc T, Oki T, Kanae S and Ngo T 2007 A variable stream flow velocity method for global river routing model: model description and preliminary results *Hydrol. Earth Syst. Sci.* **4** 4389–414

Oki T 1999 Assesment of annual runoff from land surface model using total runoff integration pathways (TRIP) *J. Meteor. Soc. Japan* **77** 235–55

Oki T 2003 Global water resources assessment under climatic change in 2050 using TRIP *Water Resources Systems—Water Availability and Global Change* ed S Franks, G Blöschl, M Kumagai, K Musiake and D Rosbjerg (Wallingford: IAHS Press) pp 124–33

Paiva R C D, Collischonn W and Buarque D C 2013 Validation of a full hydrodynamic model for large-scale hydrologic modelling in the Amazon *Hydrol. Process.* **27** 333–46

Pappenberger F, Dutra E, Wetterhall F and Cloke H L 2012 Deriving global flood hazard maps of fluvial floods through a physical model cascade *Hydrol. Earth Syst. Sci.* **16** 4143–56

Pokhrel Y, Hanasaki N, Koirala S, Cho J, Yeh P J-F, Kim H, Kanae S and Oki T 2012 Incorporating anthropogenic water regulation modules into a land surface model *J. Hydrometeorol.* **13** 255–69

Pokhrel Y N, Koirala S, Yeh P J-F, Hanasaki N, Longuevergne L, Kanae S and Oki T 2015 Incorporation of groundwater pumping in a global land surface model with the representation of human impacts *Water Resour. Res.* **51** 78–96

Portmann F T, Siebert S and Döll P 2010 MIRCA2000—Global monthly irrigated and rainfed crop areas around the year 2000: a new high-resolution data set for agricultural and hydrological modeling *Glob. Biogeochem. Cycles* **24** 1–24

Prudhomme C *et al* 2014 Hydrological droughts in the 21st century, hotspots and uncertainties from a global multimodel ensemble experiment *Proc. Natl Acad. Sci. USA* **111** 3262–7

Rost S, Gerten D, Bondeau A, Lucht W, Rohwer J and Schaphoff S 2008 Agricultural green and blue water consumption and its influence on the global water system *Water Resour. Res.* **44** 1–17

Rudorff C M, Melack J M and Bates P D 2014 Flooding dynamics on the lower Amazon floodplain: 1. Hydraulic controls on water elevation, inundation extent, and river-floodplain discharge *Water Resour. Res.* **50** 619–34

Sampson C C, Smith A M, Bates P D, Neal J C, Alfieri L and Jim E Freer 2015 A high-resolution global flood hazard model *Water Resour. Res.* **51** 7358–81

Sausen R, Schubert S and Dümenil L 1994 A model of river runoff for use in coupled atmosphere-ocean models *J. Hydrol.* **155** 337–52

Schewe J *et al* 2014 Multimodel assessment of water scarcity under climate change *Proc. Natl Acad. Sci. USA* **111** 3245–50

Stacke T and Hagemann S 2012 Development and evaluation of a global dynamical wetlands extent scheme *Hydrol. Earth Syst. Sci.* **16** 2915–33

Sutcliffe J V and Parks Y P 1989 Comparative water balances of selected African wetlands *Hydrol. Sci. J.* **34** 49–62

Takata K, Emori S and Watanabe T 2003 Development of the minimal advanced treatments of surface interaction and runoff *Glob. Planet. Change* **38** 209–22

Tang Q, Oki T, Kanae S and Hu H 2008 Hydrological cycles change in the Yellow River basin during the last half of the twentieth century *J. Clim.* **21** 1790–806

Tang Q, Oki T, Kanae S and Hu H 2007 The influence of precipitation variability and partial irrigation within grid cells on a hydrological simulation *J. Hydrometeorol.* **8** 499–512

Veldkamp T I E *et al* 2017 Human impact parameterization in global hydrological models improves estimates of monthly discharges and hydrological extremes: a multi-model validation study (in preparation)

Voisin N, Wood A and Dennis L 2008 Evaluation of precipitation products for global hydrological prediction *J. Hydrometeorol.* **9** 388–407

Wada Y, van Beek L P H, van Kempen C M, Reckman J W T M, Vasak S and Bierkens M F P 2010 Global depletion of groundwater resources *Geophys. Res. Lett.* **37** L20402

Wada Y, van Beek L P H, Viviroli D, Dürr H H, Weingartner R and Bierkens M F P 2011 Global monthly water stress: 2. Water demand and severity of water stress *Water Resour. Res.* **47** W07518

Wanders N and Wada Y 2015 Decadal predictability of river discharge with climate oscillations over the 20th and early 21st century *Geophys. Res. Lett.* **42** 10689–95

Winsemius H C, Aerts J C J H, Van Beek L P H, Bierkens M F P, Bouwman A F, Jongman B, Kwadijk J C J, Ligtvoet W, Lucas P L and van Vuuren D P 2015 Global drivers of future river flood risk *Nat. Clim. Change* **6** 381

Woltemade C J and Potter K W 1994 A watershed modeling analysis of fluvial geomorphologic influences on flood peak attenuation *Water Resour. Res.* **30** 1933–42

Yamazaki D, De Almeida G A M and Bates P D 2013 Improving computational efficiency in global river models by implementing the local inertial flow equation and a vector-based river network map *Water Resour. Res.* **49** 7221–35

Yamazaki D, Kanae S, Kim H and Oki T 2011 A physically based description of floodplain inundation dynamics in a global river routing model *Water Resour. Res.* **47** 1–21

Yamazaki D, Oki T and Kanae S 2009 Deriving a global river network map and its sub-grid topographic characteristics from a fine-resolution flow direction map *Hydrol. Earth Syst. Sci.* **13** 2241–51

Zaherpour J, Mount N and Gosling S N 2017 Optimising the combination of global-scale climate change impact models from a multi-model ensemble (in preparation)

Zwarts L, Cissé N and Diallo M 2005 Hydrology of the upper *Niger The Niger, a Lifeline: Effective Water Management in the Upper Niger Basin* ed L Zwarts, P van Beukering, B Kone and E Wymenga (Lelystad, Netherlands: RIZA) pp 14–41