Accepted by ***Water Resources Research*** on Mar. 12, 2014

Real-time Global Flood Estimation using Satellite-based Precipitation and a Coupled Land Surface and Routing Model

Huan Wu[1, 2*], Robert F. Adler[1, 2], Yudong Tian[1, 2], George J. Huffman[2], Hongyi Li[3], and JianJian Wang[1, 2]

[1] Earth System Science Interdisciplinary Center, University of Maryland, College Park, MD 20740
[2] NASA Goddard Space Flight Center, Greenbelt, MD 20771
[3] Pacific Northwest National Laboratory, Richland, Washington, USA.
*Email: huanwu@umd.edu

Abstract

A widely used land surface model, the Variable Infiltration Capacity (VIC) model, is coupled with a newly developed hierarchical dominant river tracing-based runoff-routing model to form the Dominant river tracing-Routing Integrated with VIC Environment (DRIVE) model, which serves as the new core of the real-time Global Flood Monitoring System (GFMS). The GFMS uses real-time satellite-based precipitation to derive flood-monitoring parameters for the latitude-band 50°N-50°S at relatively high spatial (~12km) and temporal (3-hourly) resolution. Examples of model results for recent flood events are computed using the real-time GFMS (http://flood.umd.edu). To evaluate the accuracy of the new GFMS, the DRIVE model is run retrospectively for 15 years using both research-quality and real-time satellite precipitation products. Evaluation results are slightly better for the research-quality input and significantly better for longer duration events (three-day events vs. one-day events). Basins with fewer dams tend to provide lower false alarm ratios. For events longer than three days in areas with few dams, the probability of detection is ~0.9 and the false alarm ratio is ~0.6. In general, these statistical results are better than those of the previous system. Streamflow was evaluated at 1,121 river gauges across the quasi-global domain. Validation using real-time precipitation across the tropics (30ºS–30ºN) gives positive daily Nash-Sutcliffe Coefficients for 107 out of 375 (28%) stations with a mean of 0.19 and 51% of the same gauges at monthly scale with a mean of 0.33. There were poorer results in higher latitudes, probably due to larger errors in the satellite precipitation input.

Keywords: flood modelling, flood detection, VIC, DRT, DRIVE, TRMM

1. Introduction

Floods are a leading natural disaster with worldwide, significant, negative social-economic impacts. According to World Disaster Report [2012], floods and associated landslides caused more than 55% (2,000) of a total of 3,600 significant natural disasters during 2002-2011 over the globe; they killed over 65,000 people, affected over 1.1 billion people and cost an estimated $280 billion (US Dollars in 2011). Most of these disasters occurred in densely populated and under-developed areas where an effective flood monitoring and forecasting system is lacking due to insufficient resources [*Wu et al.*, 2012a]. A reliable flood monitoring and forecasting system at a global scale is extremely desirable to a variety of national and international agencies for humanitarian response, hazard mitigation and management. Satellite remote sensing has opened a new era to pursue global flood estimation (particularly important for remote and trans-boundary areas) by providing: (1) flood extent mapping via direct observations using optical [e.g., *Brakenridge*, 2006; *Ordoyne and Friedl*, 2008] or Synthetic Aperture Radar imagery [e.g. *Horritt et al.*, 2003; *Mason et al.*, 2012]; and (2) flood monitoring and forecasting through the use of hydrologic models and observational inputs for precipitation, land cover, vegetation, topography, and hydrography etc. [e.g. *Shrestha et al.*, 2008; *Wu et al.*, 2012a, Alfieri et al., 2013], which is the subject of this paper.

Rainfall estimation is the most critical meteorological input of a hydrologic model for real-time flood estimation, and can be obtained through satellite remote sensing with reliable availability at relatively high spatial-temporal resolution and short lag time (hours). One such satellite-based precipitation product, the National Aeronautics and Space Administration (NASA) Tropical Rainfall Measuring Mission (TRMM) Multi-satellite Precipitation Analysis [TMPA; *Huffman et al.*, 2007], has been successfully applied in many hydrologic modelling applications [e.g., *Harris et al.*, 2007; *Su et al.*, 2008 and 2011]. The TMPA precipitation products are composed of multiple satellite estimates calibrated, or adjusted, to the information from the TRMM satellite itself, which carries both a radar and passive microwave sensor. An experimental Global Food Monitoring System (GFMS) using the real-time version of the TMPA precipitation information (3-hourly, with ~6 hour lag, 0.25º latitude–longitude resolution) for quasi-global (50ºS–50ºN) coverage was developed and improved [*Hong et al.*, 2007; *Yilmaz et al.*, 2010; *Wang et al.*, 2011; *Wu et al.*, 2012a] and has been running routinely for the last few years providing useful results for a number of organizations. Currently, this real-time flood estimation system is often the only source of quantitative information during significant flood events, when information is needed for relief efforts by humanitarian agencies, such as United Nations Office for the Coordination of Humanitarian Affairs (OCHA) and United Nations World Food Programme (WFP).

Evaluations of various hydrologic model-based flood estimation calculations using satellite precipitation data have been conducted with positive performances at local and regional scales (e.g., *Shrestha et al.*, 2008; *Pan et al.*, 2010; *Su et al.*, 2008 and 2011). On a larger, global scale, *Wu et al.* [2012a] evaluated the previous version of the GFMS, which was based on a grid-based hydrologic model [*Wang et al.*, 2011], driven by TMPA 3B42V6 research (non-real-time) rainfall product. They examined the performance in flood event detection against available flood inventories, showing that the GFMS flood detection performance improves with longer flood durations and larger affected areas. The presence of dams tended to result in more false alarms and longer false alarm duration. The statistics for this previous system for flood durations greater than three days and for areas without dams were around a Probability of Detection (POD) of ~ 0.70 and a False Alarm Ratio (FAR) of ~ 0.65 [*Wu et al.*, 2012].

These evaluations of our previous systems [*Yilmaz et al.*, 2010; Wu et al., 2012a] indicated pathways toward an improved approach with greater flexibility and accuracy. The key areas for potential improvement included consideration of sub-grid hydrologic processes, inclusion of cold season processes and improved routing that could lead to two-way interaction between the land-surface processes and the routing calculations. A Land Surface Model (LSM) can be used to effectively calculate land surface and subsurface runoff through its vertical water-energy processes, partitioning precipitation into infiltration, evapotranspiration and runoff components. However, a lateral process for runoff-routing is usually lacking within most LSMs, though an efficient and accurate runoff-routing scheme can have significant impacts on delineation of river basin water and energy budgets [*Decharme et al.*, 2011], and be critically important for flood simulation. For LSMs such as the Variable Infiltration Capacity (VIC) model (*Liang et al.,* 1994 and 1996), the traditional cell-to-cell or source-sink routing models based on widely used Unit Hydrograph methods, e.g. *Lohmann et al.* [1996] and *Wu et al.* [2012c] can be used to successfully simulate streamflow by post-processing the LSM runoff output. However, it is difficult (if even possible) to couple this type of routing model with an LSM (with feedbacks to the LSM online) for global-scale real-time flood calculation. This is because the convolution algorithms have to incorporate all upstream runoff information for multiple previous time steps to determine the streamflow for a specific downstream grid cell at a time step. For this study, we developed a new hydrologic module for the GFMS by coupling the widely used VIC land surface model with a recently developed physically-based hierarchical Dominant River Tracing [*Wu et al.*, 2011 and 2012b] based runoff-Routing (DRTR) model. This new coupled system, the Dominant river-tracing Routing Integrated with VIC Environment (DRIVE) model, is intended to provide improved global results and increased flexibility for implementation of future improvements.

In this paper we describe this new DRIVE-based version of the GFMS and evaluate the performance of the system on a global basis against stream flow observations and flood event archives, using satellite precipitation information from both the real-time and research products. Section 2 of this paper describes the methodology, particularly on the DRIVE coupled model system; Section 3 outlines the model data inputs and parameterization; Section 4 focuses on the model evaluation; and conclusions and future work are presented in Section 5.

2. Methodology

The new real-time GFMS (http://flood.umd.edu) combines the satellite-based precipitation estimation, runoff generation, runoff routing, and flood identification using the DRIVE coupled model system described in detail in Sections 2.1 and 2.2.

2.1 Variable Infiltration Capacity (VIC) model

Hydrologically oriented LSMs, such as the VIC model, solve for full water and energy balances with good skill for water budget estimation [*Peters-Lidard et al.*, 2011]. We selected the VIC model as a critical part of our GFMS for two additional reasons. First, significant community development has been carried out, and continued improvement will be maximized by being part of this larger community of land surface model development and testing. The VIC model has been successfully applied for many hydrologic simulations and water resource manangement studies, including flooding [e.g. *Hamlet and Lettenmaier*, 2007 and 2010; *Elsner et al.*, 2010; *Voisin et al.*, 2011]. Through these studies the VIC model has been generally well parameterized across the globe and thus provides a good starting point for global applications such as this study. Second, the VIC model includes a module for snow and soil frost dynamics [*Storck et al.*, 2002; *Cherkauer and Lettenmaier*, 2003], with good validation against streamflow observations in many snowmelt-dominated basins, particularly in mountainous areas [*Christensen et al.*, 2004; *Hamlet et al.*, 2005; *Elsner et al.*, 2010; *Wu et al.*, 2012c]. This will benefit the GFMS in forecasting spring streamflow and snowmelt-related floods and allow us to estimate floods in a large part of the globe with snowmelt-dominant basins.

Representation of complex physical processes at a spatial resolution commensurate with LSMs through sub-grid process is a good strategy to balance data availability, heavy computing loads, and model accuracy. Inclusion of sub-grid processes is a major feature of the VIC model contributing to its good performance in runoff generation calculations. The VIC model considers the sub-grid heterogeneity of infiltration capacity through statistical variable infiltration curves [*Zhao and Liu*, 1995], which have been demonstrated to work very well for large-scale applications [*Sivapalan and Woods*, 1995]. The VIC model also considers sub-grid parameterization and processes on fractional sub-grid areas for different land cover types and elevation bands. To use the VIC model for real-time runoff prediction, we made a significant effort to modify the VIC model from its original individual grid-cell–based mode to a mode that is able to simulate spatially distributed runoff at each time step, i.e., computing all the grid boxes at each time step. The modification was performed on the version of the VIC model (v4.1.1) in an efficient way without changing model physics, so that we can conveniently update our modified VIC model periodically using the updates from the VIC model community.

2.2 Dominant River Tracing-based runoff-Routing (DRTR) model and coupling with VIC model

For clarity, the term "runoff" hereafter stands for the excess water generated in each grid cell for routing with units of depth [mm], while "streamflow" and "discharge" are used interchangeably to indicate the routed flows in the channel/floodplain network with units of [m^3/s]. The function of a routing model is to transport water (runoff) downstream in a river basin system until the river empties into the ocean or a lake. A routing model consists of two main components: (1) the description of the river basin drainage system, i.e. simplifying the basin drainage system into a parameterized concept and (2) the physical and numerical models for computer simulation of stream flow and other variables with appropriate assumptions commensurate with the simplifications in the drainage basin concept. Recently developed and relatively advanced physically based routing schemes for large-scale applications [e.g. *Decharme* *et al.*, 2011; *Yamazaki et al.*, 2011; *Li et al.*, 2013] usually deploy similar governing equations taken from various forms of the classic St-Venant equations based on mass and momentum conservation, often using the kinematic wave and diffusion wave methods. The essential differences among routing models of this type lie in the levels at which a drainage system is abstracted and simplified, and the techniques used for parameterizing each element within the model conception.

In this study we implemented a physically based routing model based on the hierarchical DRT method [*Wu et al.*, 2011 and 2012b], which includes a package of hydrographic upscaling (from fine spatial resolution to coarse resolution) algorithms and resulting global datasets (flow direction, river network, drainage area, flow distance, slope, etc.) especially designed for large-scale hydrologic modelling. This DRT-based runoff-Routing (DRTR) model is grid based and convenient for coupling with the modified gridded VIC model to simulate spatially distributed streamflow.

2.2.1 The DRTR model concept and parameterization

Recently developed grid-based, large-scale (coarser resolution) routing models usually conceptualize a drainage system as connected stem rivers at grid resolution, but with major differences in subgrid process (routing) delineation. Given the generally well established mathematics and physics for land surface routing simulation, the major challenge to implementing a large-scale routing scheme lies in obtaining accurate parameterization of the model elements (particularly at sub-grid scale). For example, a recent large-scale routing model on a grid basis [*Li et al.*, 2013], deploying a kinematic wave type routing method, conceptualized the routing process by using a hypothetical sub-grid channel to link hillslopes and stem rivers which has a transport capacity equivalent to all tributaries combined, while linking the grids via the stem river network derived by the DRT upscaling algorithm by *Wu et al.* [2011 and 2012b]. Due to the scale-consistent stem river network derived by the DRT algorithm and the scale-consistent sub-grid routing parameterization, this large-scale routing model showed a consistent model performance across different spatial resolutions [*Li et al.*, 2013].

In this study, we implemented the DRTR routing model using a drainage system concept similar to *Li et al.* [2013], but with differences in sub-grid parameterization using the full strength of the DRT algorithms to allow more-detailed high resolution subgrid information that is aggregated for coarser resolution routing simulation and for numeric solutions of the governing equations. Under the gridded DRT framework, the hydrologic system of each river basin is conceptualized as a hierarchically-connected hillslope-river-lake or -ocean system. All grid cells are connected via the predominant river (or flow path) running through the grid cell, which forms the major drainage network for the river basin (red lines in Fig. 1a). For coarser spatial resolution (e.g. coarser than 1 km) hydrologic modelling, the DRT derives the predominant river (red lines) from the fine-resolution river network (blue lines; *Wu et al.*, 2011). Fig. 1b shows a typical real drainage system within an individual grid cell, represented by high-resolution river network data, with one predominant river (dark blue) collecting runoff from tributaries (light blue) and overland areas (blank), which is conceptualized as in Fig. 1c with simplified subgrid tributaries (light blue lines). At the subgrid scale, the predominant river within each grid cell is divided into one or multiple river intervals (purple ticks in Fig. 1c and d). Each dominant river interval can have one “effective tributary” (light blue lines in Fig.1c and d) collecting runoff from its overland contributing area even if there are multiple tributaries (defined from high resolution river network) connected to the dominant river interval. All secondary dominant rivers [*Wu et al.*, 2011] within a coarse grid cell, if any, are treated as tributaries. The overland area of each grid cell is divided into two parts: (1) areas nearby the dominant river and directly contributing runoff to the dominant river through overland flow (dark blue arrows in Fig. 1d); (2) areas contributing to the dominant river through tributaries (light blue arrows in Fig. 1d). Within each grid cell, runoff generated on hillslopes is routed to its corresponding tributary through overland flow and then is treated as channel flow to enter the relevant dominant river interval. The overland flow and the tributary flow are treated as evenly distributed along the tributary and predominant river interval as lateral flow input, respectively. Once water enters the dominant river intervals, the river routing calculations follow the hierarchical dominant river ordering sequence in the major river network. Floodplain, reservoir and lake elements are not included in the current model.

All the elements (hillslope, tributary and predominant river) (Fig. 1) are identified and parameterized by the DRT on a pixel-to-pixel basis tracing from the finer resolution river network (or flow path). In this study (model running at 1/8th degree resolution), we set the number of “effective tributaries” of each grid cell to one, while parameterizing the effective tributary (including tributary length, slope, width etc.) using the value averaged from all tributaries within that grid cell as shown in Fig. 1b. The channel width is estimated by an empirical relation to corresponding drainage area. The overland area within a grid cell directly contributing runoff to the corresponding dominant river is identified first using the DRT from high resolution flow direction map and the remaining area of the grid cell is assigned to the effective tributary. The DRT also uses the Strahler ordering system [*Strahler*, 1957] to define a hierarchical drainage network topology, e.g. for the upstream-downstream relationships and conjunctions connecting different river reaches. The model structure, based on the Strahler ordering system, is efficient for integrating numerical calculations established on each individual element for a better approximation of the characteristics of natural hierarchical runoff propagation.

2.2.2 DRTR routing scheme governing equations and numeric solutions

With the comprehensive parameterization provided by the DRT, the routing scheme can conveniently deploy different governing equations and numeric solutions to individual routing elements. In this study, we present a relatively simple method, i.e. applying the kinematic wave equations to both dominant rivers at grid level and tributaries at subgrid level, while assuming the overland surface runoff and baseflow enter the corresponding dominant river intervals and tributaries within each time step.

Rectangular cross-section is assumed for all channels. Eq. (1)-(3) are the governing equations adopted for the kinematic wave method [*Chow et al.*, 1988]:

Continutity equation $$\frac{\partial A}{\partial t}+\frac{\partial Q}{\partial x}=q_L \quad (1)$$

Momentum equation $$S_f=S_0 \quad (2)$$

Manning equation $$Q=\frac{S_0^{1/2}}{nP^{2/3}}A^{5/3} \quad (3)$$

where t is the time [s], x is the longitudinal flow distance [m], A is the wetted area [m^2] defined as the channel cross-section area below the water surface, and P is the wetted perimeter [m]. S_f is the friction slope which incorporates the impacts of the gravity force, friction force, inertia force and other forces on the water. If the topography is steep enough, the gravity force dominates over the others, and S_f can be approximated by the channel bottom slope S_0, which is the basic assumption for kinematic wave routing approaches [*Chow et al.*, 1988]. In Eq. (3), n is Manning's roughness coefficient, which is not directly measurable, but mainly controlled by surface roughness, type of bottom material and sinuosity of the flow path. In this study we applied a constant value of 0.03 globally for both predominant rivers and subgrid tributaries, although eventually it should be calibrated for local river basins. Q is the streamflow and discharge [m^3/s] and q_L is the lateral discharge in unit width [m^3/s/m]. The backward differential scheme of the eq. (1) is

$$\frac{A_{i+1}^{n+1}-A_{i+1}^{n}}{\Delta t}+\frac{Q_{i+1}^{n+1}-Q_{i}^{n+1}}{\Delta x}=\overline{q_L} \qquad (4)$$

where i and n are the spatial and temporal indexes, respectively. Rewriting the Manning equation, eq. (3), $A_{i+1}^{n+1}=\alpha(Q_{i+1}^{n+1})^{\beta}$ and $A_{i+1}^{n}=\alpha(Q_{i+1}^{n})^{\beta}$, substituting in eq. (4) we get

$$\frac{\Delta t}{\Delta x}Q_{i+1}^{n+1}+\alpha(Q_{i+1}^{n+1})^{\beta}=\frac{\Delta t}{\Delta x}Q_{i}^{n+1}+\alpha(Q_{i+1}^{n})^{\beta}+\Delta t\overline{q_L} \qquad (5)$$

where $\alpha=(nP^{2/3}/\sqrt{S_0})^{0.6}$ and $\beta=0.6$. The right side of eq. (5) is known, and the newton-iterative method is used to solve the unknown Q_{i+1}^{n+1} . The same numeric solution is also used for estimating channel water depth [mm] and thus for routed runoff (or land surface water storage, [mm]) calculations.

2.2.3 The coupling of the DRTR routing model with the VIC model

The vertical model processes of the VIC model run are calculated separately for each sub-grid area before they are aggregated to a grid-scale output at the end of each model time step. The routing scheme was implemented within the VIC model framework taking the VIC estimated runoff as input for the routing calculation of discharge and routed runoff at each time step. The VIC model was modified to match the DRTR routing model structure with all grid cell calculations completed at each time step in the Strahler order-based sequence. The routing time step can be finer than the VIC model time step assuming that the runoff generation by the VIC model has an even temporal distribution within each VIC model time step.

The DRTR routing scheme, implemented within the modified VIC model, can have bidirectional interactions with the VIC model. However, sub-grid floodplain delineation for appropriate redistribution of routed runoff is needed to really take advantage of the two-way coupling strategy. Therefore, in this study the routing scheme was used as a post-processor for the runoff routing after each time step from the VIC model. That is, there is no two-way interaction between VIC and the DRTR in the following calculations. We plan to test and implement this potential improvement in a future study.

3 Model setup and Data

We performed the long-term TRMM era retrospective simulations by running the DRIVE combined model using the TMPA 3B42V7 research (which contains monthly rain gauge data, from 1998 to present) and TMPA 3B42V7RT real-time precipitation data (which uses only a climatological gauge correction, from 2000 to present), at 3-hourly temporal and 1/8th degree spatial resolutions for the latitude band 50ºN-50ºS. Other forcing data (i.e. air temperature and wind speed) were taken from the NASA Modern-Era Retrospective analysis for Research and Applications (MERRA) reanalysis [*Rienecker et al.*, 2011]. The phase (liquid vs. solid) of the precipitation is determined based on a simple partitioning scheme using air temperature within the VIC model [*Hamlet et al.*, 2005]. For each grid cell at a time step, the satellite-based precipitation is assumed to be 100% snow when the air temperature is below -0.5°C, while it is 100% rain when the temperature is above 0.5°C. A linear relationship is assumed between the two extremes. The quarter-degree resolution global soil and vegetation parameters (provided by Justin Sheffield, University of Princeton) were simply projected (pixel replication) to 1/8th degree resolution. This dataset included the recent updated parameters for the VIC model improved through calibration efforts [*Troy et al.*, 2008]. The hydrographic parameters (e.g. flow direction, drainage area, flow length, channel width, channel slope, overland slope, flow fraction, river order) for the DRTR runoff-routing scheme were derived by applying the DRT to the HydroSHEDS [*Lehner et al.*, 2008] global 1 km baseline hydrographic data [*Wu et al.*, 2011, 2012b]. Based on the DRT algorithms, all parameters for subgrid tributaries and flow paths are derived by tracing each fine-resolution (i.e. 1 km) grid cell. For example, overland slope and channel (tributary and predominant river) slopes for a grid cell are estimated as the average slope of all overland flow paths and channel flow paths, respectively, within the grid cell (more details in Li et al., 2013). Hereafter, TMPA 3B42V7 research and real-time precipitation products are referred to as TMPA RP and TMPA RT respectively, while the DRIVE model driven by TMPA

RP and TMPA RT is referred to as DRIVE-RP and DRIVE-RT respectively. A 3-year model spin-up run was performed (1998-2000) using the DRIVE-RP data to define the initial conditions for the both scenarios (DRIVE-RP and DRIVE-RT). All model results presented in this study are based on model parameters either estimated directly from input data (e.g. through DRT algorithms) or from the VIC community (e.g. soil and vegetation parameters).

4. Model results and model performance evaluation

In order to evaluate the new GFMS performance in flood event detection and streamflow magnitude estimation, particularly for evaluating the status of the GFMS in real-time flood estimation at the global scale, we performed the following: (1) evaluating examples of recent flood events as seen by the real-time GFMS, which has been running the DRIVE model routinely at 3-hourly temporal and 1/8 degree spatial resolutions over the globe using the real-time precipitation data; (2) evaluating the system model performance using 2,086 archived flood events by Dartmouth Flood Observatory (DFO, http://floodobservatory.colorado.edu), according to the evaluation method used by *Wu et al.* [2012a] and (3) validating against observed daily streamflow data from the 1,121 gauges selected from the Global Runoff Data Centre (GRDC, http://grdc.bafg.de/) database.

4.1 Introduction of the major outputs of the DRIVE model and the real-time GFMS

The DRIVE model can calculate a large number of hydrologic variables (e.g. soil moisture, evaportranspiration, snow water equivalent), but the main focus in this paper is the two outputs from the routing model related directly to floods: (1) streamflow (or discharge, m^3/s); and (2) routed runoff (or surface water storage), which is the water depth [mm] at each grid cell on a dry ground basis, and statistical thresholds which were used for defining flood occurrence and intensity. According to *Wu et al.* [2012a], each grid cell is determined to be flooding at a time step when the routed runoff is greater than the flood threshold of that grid cell. In this study we calculated the flood threshold at each grid cell, based on the 11-year (2001-2011) DRIVE model retrospective simulation results, using the method from *Wu et al.* [2012a] with a slight modification, to make it relatively more reliable and easier to implement. Specifically, a grid cell is determined to be flooding when $R > P_{95} + 0.5 * \delta$ and $Q > 10$, where R is the routed runoff [mm] of that grid cell at a time step; P_{95} and δ are the 95th percentile value and the temporal standard deviation of the routed runoff derived from the retrospective simulation time series at the grid cell; and Q is the corresponding value of discharge [m^3/s].

By applying the flood threshold map to (subtracted from) the DRIVE model simulated routed runoff, the flood detection and intensity (i.e. the water depth above flood threshold, [mm]) is estimated for each grid cell of the globe at each time step. The real-time model results and precipitaton background information can be accessed at http://flood.umd.edu. Examples (screenshots) of the real-time GFMS major outputs (Routed runoff, streamflow, and flood detection/intensity) are shown in Fig. 2(a-c). An example of global TMPA 3B42 real-time rainfall input data (quarter degree) at a same time interval is also shown in Fig. 2d. For the flood detection/intensity parameter (depth above threshold), Fig. 2c(1-6) shows the evolution (at a daily interval) of the flood event in North India (north subbasins of Ganges River Basin) during Jun 15, 2013 to Jun 20, 2013. To interpret the flood detection and intensity results (Fig. 2c), areas with more than ~30 mm above the threshold (starting with blue) are ususally considered having significant flood, while other potential areas (i.e. green and light blue in Fig. 2c) with lower flood intensity indicate a possible developing flood. A wide-spread lower flood intensity usually occurs as a response to wide-spread rainfall events, often indicating a coming flood wave in downstream areas at a later time, which can serve as a warning signal. The North India floods were reported as killing more than 1,000 people. The GFMS generally captured the events but the accuracy was not validated because of the lack of observed data in real time for this case.

4.2 Recent floods in Mississippi upstream sub-basin rivers

Upstream sub-basins of the Mississippi River in Iowa, Ilinois, Missouri, Indiana, Ohio, and Kentucky flooded during April to June of 2013 (Figs. 3 and 4), with the location indicated in Fig. 2 as a red rectangle over the USA. The GFMS output successfully captured the occurrence of these events according to information from the Dartmouth Flood Observatory and the media (see flooding at Des Plaines, IL on April 19, 2013 in photograph in Fig. 4). Fig. 3a and 3b show the snapshots of the GFMS estimated flood detection and intensity parameter for the two major flood waves from Mississippi upstream tributary rivers originating in mid-April and early-June 2013, respectively. Both flood events were caused by wide-spread precipitation in this area as shown in Fig. 3c and 3d with previous 7-day accumulated precipitation prior to the flooding time (i.e. 09Z18Apr2013 and 09Z02Jun2013, respectively). Meanwhile, the spatially distributed streamflow information is also shown in Fig. 3e and 3f. All such information and more details are available from the GFMS website, e.g. animations for detailed (3-hourly time step) flood evolution within river basin drainage systems and time series data for any grid cell of interest.

In order to quantitatively validate the real-time GFMS performance in simulating these flood events, we compared the real-time calculations with 29 USGS streamflow gauges from the USGS WaterWatch program (http://waterwatch.usgs.gov; filled circles in Fig 4a) within the flood affected area (along the Iowa, Cedar, Wabash, Ilinoise, Ohio, Misouri, and Mississipi Rivers). The upstream drainage areas of these gauges range from 2,884 to 1,772,548 km^2. According to the metrics calculated based on the two-year retrospective period (2011-06-12 to 2013-06-12), there were 41% (12) out of 29 gauges showing positive daily Nash-Sutcliffe coefficient (NSC) [*Nash and Sutcliffe*, 1970] values with a mean of 0.23 as indicated as green points (rather than black) in Fig. 4a and 55% (16) of them showing positive monthly NSC values with a mean of 0.35. All these gauges showed fairly good correlation coefficients between observed and simulated streamflow with a mean of 0.55 and 0.70 at daily and monthly scale, respectively. Fig. 4 also shows the observed and simulated daily hydrographs for four of the gauges (locations indicated in Fig. 4a) during this Spring and early Summer flooding period (April 1 to Jun 9, 2013). These hydrographs explain the good performance of the GFMS in flood occurance detection (Section 4.3) as the system can generally capture the variation and magnitude of observed streamflow during the flooding season. There were biases in magnitude and shifts in timing as shown, but they have limited impacts on flood event detection. For these cases, the simulated floods tend to be faster than observed, which may be because the DRIVE model does not include floodplain and lake/reservoir processes. Hydrographic parameterization can also contribute to the timing error, e.g., overestimated channel width or underestimated surface roughness can also lead to faster flood waves. One can also see from Fig. 4 that in these cases the model consistently underestimated the snowmelt-related streamflow in early spring, which, however, is not typical for most years in our long-term retrospecitve simulation (not shown).

Overall, without model calibration and considering the impacts from man-made structures and regulated flow (many small dams in this area, Fig. 4a), the DRIVE model using the real-time satellite precipitation input gives a reasonable real-time detection of flood occurance and magnitude estimation.

4. 3 Flood event inventory based evaluation

Following the same methodology developed and used by *Wu et al.* [2012a], a similar evaluation of the new GFMS performance in flood event detection across the globe was conducted using the same reported flood event databases compiled mainly from news, reports and some satellite observations by the DFO. The flood event database used by *Wu et al.* [2012a] was extended through 2011 using the latest DFO database.

Based on a 2 × 2 contingency table (a =GFMS yes, reported yes; b = GFMS yes, reported no; c =GFMS no, reported yes; d =GFMS no, reported no), three categorical verification metrics, including probability of detection [POD; a/(a + c)], false alarm ratio [FAR; b/(a+b)], and critical success index [CSI; a/(a + b + c)], were calculated using the 11-year (2001-2011) retrospective simulations from both DRIVE-RP and DRIVE-RT, against the DFO flood inventary for the same time period.

4. 3.1 Flood Threshold maps by DRIVE-RP and DRIVE-RT and the corresponding background precipitation estimation

The flood threshold maps used for the Flood Detection/Intensity parameter are derived from the retrospective runs and the formulas given in a previous section. Both the DRIVE-RP and DRIVE-RT-based flood threshold maps have very similar spatial patterns and value ranges. The global flood threshold values by DRIVE-RP range from 0 to 14,349 mm with a mean of 17.7 mm, while the DRIVE-RT derived shreshold values range from 0 to 16,268 mm with a mean of 18.7 mm. Both flood threshold maps correspond well to the river basin drainage networks, with large values for river grid cells having large upstream drainage areas. Fig. 5a shows the DRIVE-RT-based flood threshold map, with the difference between the thresholds for DRIVE-RP and DRIVE-RT shown in Fig. 5b. Fig. 6a shows the mean annual precipitation distribution by TMPA RT from the same time period (2001-2011) and the difference map (Fig. 6b) in parallel to Fig. 5. There is a correlation coeficient of 0.98 between the two flood threshold maps, while the correlation coeficient of the two mean annual precipitation maps by TMPA RP and RT is also very high at 0.95. The global mean difference between the two flood threshold maps (DRIVE-RT minus DRIVE-RP) is 1.0 mm (5.9%), while the mean difference in the mean annual precipitation is 49.1 mm (5.4 %). Visually comparing of Fig. 5b and Fig. 6b clearly shows that the variations in the flood threshold values in the DRIVE-RT (relative to DRIVE-RP) are primarily controlled by the bias distribution in the precipitation. The DRIVE-RT flood thresholds usually show a consistent bias against those of DRIVE-RP, either low or high, within a basin or sub-basin (Fig. 5b). For example, from Fig. 5b and Fig. 6b, the DRIVE-RT flood threshold values and corresponding precipitaion are generally consistently higher than those of DRIVE-RP in the west-central U.S. (including the entire Missouri River basin and Colorado River basin). In contrast, they are generally lower in the eastern areas of the Mississippi River, with the result that flood threshold values are higher than for DRIVE-RP in the downstream part of the Mississippi stem river (as seen from the inset window in Fig. 5b). A similar situation happens in the Amazon river basin, while consistent higher threshold values by DRIVE-RT than DRIVE-RP were found in almost all Asian and Austrailian river basins, except for Southeast Asia and coastal areas. The entire Congo River and almost the entire Danube River basin and Nile River basin show lower DRIVE-RT thresholds than DRIVE-RP. A zoomed-in area for Asia of Fig. 5b is also shown as background in Fig. 7.

Fig. 6b also indicates the areas where improvements are needed for satellite-based real-time land precipitation estimation. The overestimation in the interiors of continents at higher latitudes may be related to false identification of surface effects as precipitation events in wintertime, while overestimation over the upper reaches of the Amazon may be related to overestimation of deep convective events. In coastal areas in middle latitudes the underestimation is most likely related to underestimation of shallow, orographic rainfall. Elimination of these precipitation biases will likely improve the flood statistics.

4.3.2 Flood event detection metrics

We used the same method developed by *Wu et al.* [2012a] to match the simulated and reported flood events for the evaluation. A brief introduction of the method is given below. For more details, one can refer to *Wu et al.* [2012a]. The DFO flood database provides the locations (latitudes/longitudes) and days of the reported floods. We assume the reported flood locations are located in the correct river basin, even though they may not be recorded with precisely correct latitude and longitude coordinates. A simulated flood event was defined within a local spatial window according to the reported location and a one-day (±24 hours) buffer surrounding the reported flood duration. The local spatial domain was defined, based on the DRT flow direction map, to be composed of all grid cells in the upstream drainage area within a limited flow distance (i.e. ~200 km) according to the reported location and the grid cells in the downstream stem river of the basin/sub-basin below the reported location within a limited distance (i.e. ~100 km). When there are more than three grid cells flooding (according to the method in Section 4.1) within the spatial domain for two continuous three-hour time intervals, we mark the entire area defined by the spatial domain as simulated flooding.

According to the flood event matching method discussed above, the DRIVE-RP and DRIVE-RT detected 1,820 (87.2%) and 1,799 (86.2%) out of total DFO reported 2,086 flood events over the entire study domain during the 11-year time period, respectively. The DRIVE-RP only has a slightly better performance than DRIVE-RT in detecting reported greater-than-one-day flood events, but both of them have a much higher POD than that of the previous version of the GFMS (~60%) [*Wu et al.,* 2012a]. The POD for flood events of greater than three-day duration is ~90%, as compared to ~80% for the previous system.

In order to evaluate the GFMS performance in terms of false alarms, 38 well-reported areas (shaded yellow in Fig. 7) are selected to further evaluate the flood detection performance POD, FAR and CSI, together. This approach is used to minimize the impact of unreported floods, especially in sparsely populated areas. Each of these well-reported areas, according to *Wu et al.* [2012a], is defined as a limited spatial window (based on reported flooding location) having at least six reported floods during the 11 years. Fig. 7 shows the distribution of these well-reported areas in South-East Asia for example, very similar to those identified using the reported flood inventory during a different time period (1998~2010) by *Wu et al.* [2012a]. Well-reported areas are also defined for the other continents. The metrics of POD, FAR and CSI vary across regions but with a generally consistent trend related to number of upstream dams. The dams (Fig. 7) are located according to the global large dam database [*Vörösmarty et al.*, 1997; *Vörösmarty et al.*, 2003]. Fig. 8 shows the statistical results for each well-reported area for floods longer than three days according to the DFO data. There are a total of 304 floods in this validation set. Along the bottom of the plots in Fig. 8 are the number of dams (from a more comprehensive Global Reservoir and Dam (GRanD) database [Lehner et al., 2011]) in each area, increasing toward the right side of the diagrams. For example, both DRIVE-RP and DRIVE-RT results show that the FAR tends to increase along with the increasing of number of dams in the upstream areas (Fig. 8). This trend is also clearly shown in Fig. 7, in which FAR tends to be smaller where there are fewer or no large dams (dots) upstream of a well-reported area. The POD score tends to be higher in well dammed and well-reported areas, though the signal is not consistent as for FAR.

These findings are consistent with and explained in detail in *Wu et al.* [2012a]. Dams tend to result in more false alarms since the DRIVE model does not included dam/reservoir operation information at this time.

The comparison between DRIVE-RP and DRIVE-RT results show very close performance for most of the selected well-reported areas indicating very similar precipitation information (in terms of ocurrence and relative magnitude) in the upstream basins of these well-reported areas by TMPA RP and TMPA RT. Generally DRIVE-RP showed somewhat better performance than DRIVE-RT according to all metrics. DRIVE-RP provided an overall slightly better mean POD of 0.93, FAR of 0.84 and CSI of 0.15 for all floods with duration greater than one day, compared to the DRIVE-RT with a mean POD of 0.90, FAR of 0.88 and CSI of 0.12 (Table 1). For floods with longer duration (i.e. $\geq$3 days), both DRIVE-RT and DRIVE-RP significantly decreased false alarms with a mean FAR of 0.73 and 0.65, resulting in higher CSI scores of 0.25 and 0.34 respectively (Table 2). Both DRIVE-RP and DRIVE-RT showed much better flood detection performance than the previous version of GFMS, which showed a mean POD of 0.70, FAR of 0.93 and CSI of 0.07 for floods with duration more than one day, and a mean POD of 0.78, FAR of 0.74 and CSI of 0.23 for floods with duration more than three days [*Wu et al.*, 2012a]. From Tables 1 and 2, the false alarm rates are significantly lower in WRAs with fewer dams than those with more dams. For floods more than three days in the 18 WRAs with fewer than five dams, the DRIVE-RP also showed an overall better mean POD of 0.92, FAR of 0.56 and CSI of 0.43, than the DRIVE-RT with a mean POD of 0.87, FAR of 0.66 and CSI of 0.32 (Table 2). The primary reason for improved detection results in the new system is surmised to be the improved runoff generation and routing with the DRIVE system, with a secondary factor possibly being improved precipitation estimation.

4. 4 Gauge streamflow based validation

Streamflow is arguably the best variable to be used to evaluate the overall performance of a hydrologic model because it represents the integrated results from all upstream water and energy processes and streamflow observations are much more available than other hydrologic variables (e.g. soil moisture, surface runoff) with relatively lower bias in observations. We evaluated the DRIVE model performance for streamflow simulation using observed streamflow data from 1,121 global river gauges from the GRDC database. The gauges were selected with the criteria: (1) gauge data have at least a one-year length of daily time series during the validation period

2001- 2011; (2) the gauge can be well located in the DRT upscaled river network, which serves as the geo-mask for organizing all model input and output data, so that the gauge observations can accurately represent the runoff-concentration from its upstream drainage area; (3) the gauge upstream drainage area >200 km^2; (4) the gauges are not close to the study domain boundaries (latitude 50°N and 50°S), since these gauges cannot accurately represent their full upstream drainage basins. A program from the DRT algorithm package was used to geo-locate the original GRDC gauges in the model domain for evaluation. For each selected gauge, the difference in upstream drainage area of the gauge location between the DRT dataset and the GRDC dataset is less than 10%. The selected river gauges are widely distributed across the study domain and provide a good representation of the diverse hydroclimate regions, e.g. arid, semiarid, and humid regions (Fig. 9). However, east Africa, and south and west Asia (particularly the area between 46°E - 97°E) are somewhat underrepresented for this evaluation.

Both DRIVE-RP and DRIVE-RT results for the same retrospective time period from Jan. 2001 to Dec. 2011 (132 months) were compared to observed daily streamflow data. Metrics including daily (N_d) and monthly (N_m) Nash-Sutcliffe coefficient (NSC) values, daily (R_d) and monthly (R_m) correlation coefficients, and Mean Annual Relative Error (MARE), all calculated based on the simulated and observed time series of streamflow (m^3/s).

4.4.1 Overall model performance in streamflow simulation over the globe

Overall, when compared against the observed daily streamflow data from 1,121 GRDC gauges, the DRIVE-RP showed that 60% (675) of the gauges had positive monthly NSC with a mean of 0.39, and 29% (322) of gauges had monthly NSC greater than 0.4 with a mean of 0.57 (Table 3). Meanwhile there were 38% (424) gauges having MARE within 30% with a mean of -0.3%. Good correlation between the model-simulated and observed streamflow time series at monthly scale exists in almost all the gauges with a mean correlation of 0.67. Fig. 9 shows the spatial distribution of the monthly NSC for the DRIVE-RP streamflow simulation results. It is shown in Fig. 9 that the model has a generally consistent performance across different regions. Fig. 10 shows the histogram distribution of the number of gauges with positive monthly and daily NSC metrics for DRIVE-RP and DRIVE-RT, which clearly indicates that DRIVE-RP outperforms DRIVE-RT at the monthly scale, while the difference in the performance between the DRIVE-RP and DRIVE-RT is smaller at the daily scale.

Model performance decreased, as expected, at the daily scale, e.g. 46% of the gauges with positive monthly NSC had negative daily NSC. However, 58% (655) of gauges had correlation coefficients greater than 0.4 between the model-simulated and observed streamflow at the daily scale with a mean of 0.57. The correlation is more important for flood event detection, in which the percentile-based skill mainly depends on the relative order of routed runoff (or streamflow) magnitudes [*Wu et al.*, 2012a]. The decrease of model skills at the daily scale is attributed to a combination of the precipitation input, model parameterization and the human impacts. The TMPA RP precipitation contains an adjustment using available rain gauge data at the monthly scale, which does not provide significant positive impact on the sub-monthly variability of precipitation because the sub-monthly depends on the sequence of short-interval precipitation events from the satellites. The model parameters (e.g. surface roughness) tend to lead to larger time lag bias at smaller time scales, e.g. a too fast flood wave simulation will have much more negative impact on daily evaluation metrics than on the monthly evaluation. Human impacts (particularly the effect of dam regulation) can significantly change the shape of the daily hydrograph of a natural river, while having less impact at seasonal scales. According to the global metrics (Table 3 and Fig. 9), the DRIVE model including only natural processes, driven by TMPA-RP precipitation and *a priori* parameter sets, shows an overall promising performance in reproducing streamflow for global rivers.

The generally good performance of DRIVE-RP can also provide a measure for evaluating the potential of the real-time GFMS performance when using TMPA-RT precipitation input. From Table 3 the DRIVE-RT has a generally consistently lower skill than DRIVE-RP as expected, and with lower NSCs and correlation coefficients at both daily and monthly scales, while also having larger MARE. However, there were 215 gauges (19%) with positive daily NSC with mean of 0.16 and 474 gauges (42%) having good correlations (> 0.4) between simulated and observed daily streamflow with a mean of 0.53. These types of variations in flood statistics that are a function of rainfall input indicate that improvement of the satellite precipitation information will lead directly to better flood determinations.

4.4.2 Seasonal and regional model performance in streamflow simulation

In order to further evaluate the variations of model performance in streamflow simulation, the same metrics as presented in Section 4.4.1 are derived based on the model results and observed data for different regions and seasons (Table 3-5).

Table 3 also shows the metrics calculated based on the full simulation time series (indicating the overall model performance) at several different latitude bands, i.e. deep tropics (10ºS to 10ºN), sub-tropics (10ºN to 30ºN and 10ºS to 30ºS), mid-latitudes (30ºN to 50ºN and 30ºS to 50ºS). To facilitate interpretation of the Table 3, for example, the percentage of gauges for which the DRIVE model showed positive daily NSCs is plotted for each latitude band, as seen in Fig. 11, from which the DRIVE-RT showed clearly model skill decay from the deep tropics toward higher latitudes in both hemispheres, probably in response to the TMPA RT precipitation quality. Similar decays occurred for other metrics, e.g. for DRIVE-RT results there are 57% of stations with positive monthly NSC with mean N_m of 0.36 in the deep tropics, dropping to 51% of gauges with a mean N_m of 0.33 for northern sub-tropics and 25% gauges with a mean N_m of 0.21 for northern mid-latitudes (Table 3). The DRIVE-RP showed generally consistently better model performance over all these regions than the DRIVE-RT, and similar model skill decay toward higher latitudes can also be seen in the DRIVE-RP results in Table 3 and Fig. 11. Interestingly, this decay pattern was modified slightly (Fig. 11) by the monthly gauge-based correction in the TMPA RP which leads to relatively better monthly scale performance in higher latitudes where more rain gauge data are available. For the northern mid-latitudes there are 66% gauges having positive N_m with mean of 0.38 with DRIVE-RP, while for northern sub-tropics there were 54% (23 out of 43) gauges having positive N_m with mean of 0.41.

The same metrics were also calculated for DRIVE-RP and DRIVE-RT results for these latitude bands but only based on summer (Table 4) and winter (Table 5) months respectively. The metrics calculated based on full time series, summer-only and winter-only months (Table 3, 4 and 5) indicate the same consistent relative model performance across different regions and between DRIVE-RP and DRIVE-RT. Seasonal metrics (Table 4 and 5) also show generally consistently better model performance in deep tropics and sub-tropics than mid-latitudes. Table 4 and 5 also show generally larger water balance bias (MARE), and relatively lower monthly correlation coefficients in streamflow between gauge observations and simulations in winter seasons than summer seasons, indicating a relative less quality of satellite based precipitation estimation for winter seasons. Although precipitation is not the only causation for the spatial variation of model performance, precipitation is probably the primary one and its signature is clearly visible in the results.

Fig. 12 shows an example of comparisons of model performance between DRIVE-RP and DRIVE-RT in South America (primarily in the Amazon River Basin with relatively fewer dams) according to daily NSC and MARE. One can see that the DRIVE model shows very similar statistical performance in terms of reproducing observed daily streamflow time series and annual water balance when driven by TMPA RP or RT data. For this region (Fig. 12) there were 76 gauges, out of total 205, showing a positive daily NSC with mean of 0.25 by the DRIVE-RP, while the DRIVE-RT derived 63 gauges with positive NSC with a mean of 0.22. There were 101 and 112 gauges with MARE<30% with mean of -2.3% and -5.4% by DRIVE-RP and DRIVE-RT respectively. This indicated a generally good real-time GFMS performance (relative to DRIVE-RP) for many areas. Note that all the results were derived from the DRIVE model without any further calibration. Appropriate calibration is expected to improve the model performance for many rivers particularly for those gauges (among green and purple points in Fig. 12c and 12d) with model-calculated negative NSCs and relative higher MARE, but being within a reasonable range of error (e.g. NSC>-1.0 and MARE within 50%). Of course, precipitation error reduction is probably even more important.

4.4.3 Examples of simulated hydrographs against observations

Two GRDC gauges (locations indicated as dark points in Fig. 6b) were selected as examples to show the simulated streamflow time series against observed hydrographs with monthly and daily intervals (Fig. 13). They were selected because they represent relatively natural river basins without dams and both DRIVE-RP and DRIVE-RT results show reasonable positive monthly and daily NSCs. The GRDC gauge 1577101 (8.38333N, 38.78333E) is on Awash River, Ethiopia, with a mean annual precipitation of 1,102 mm (according to TMPA RP observation from 1998 to 2012) for its upstream basin area of 7,656 km^2 (presented by the DRT with 40 $1/8^{th}$ degree grid cells). The gauge 3664100 (25.77389S, 52.93287W) is on Rio Chopim River, Brazil with a mean annual precipitation of 2,102 mm for its upstream drainage area of 6,756 km^2 (44 grid cells). Fig. 13 shows that the simulated hydrographs generally agree well against the observed hydrographs at both daily and monthly scales. DRIVE-RT results show systematically lower streamflow estimation than DRIVE-RP over the time period (2001-2009) at the Ethiopian gauge. However, at the Brazilian gauge, the DRIVE-RT and DRIVE-RP show very close results, while the DRIVE-RT estimated streamflow is overall slightly higher than that of DRIVE-RP. The streamflow biases (DRIVE-RT vs. DRIVE-RP) at both gauges are consistent with the precipitation bias (TMPA RT vs. RP, Fig 6b).

The time delay (in days) was calculated, based on the daily values, to evaluate the errors related to the time lag between the simulated and observed hydrographs. The time delay was calculated as the time lag where the correlation coefficient between the daily simulated and observed time series is at a maximum [*Paiva et al.*, 2013]. Positive (negative) time delay values indicate delayed (advanced) simulated hydrographs. A negative one day time delay was found at the two gauge locations for both DRIVE-RP and DRIVE-RT simulations, indicating the DRIVE model has faster flood wave simulations than observed at these two locations. Table 6 shows the model performances at the two gauges under different scenarios. A one-day delayed simulated hygrograph also resulted in significantly improved daily NSC metrics at the Brazilian gauge for both DRIVE-RP and DRIVE-RT. At this gauge, the original DRIVE-RT derived a daily NSC of 0.17 for the time period of 2002-2005, while the one-day time-lag corrected simulated hydrograph has a daily NSC of 0.43. As expected, a one-day time lag has minor impacts on monthly and annual metrics at both gauges. The Ethiopian gauge statistics improve only slightly with the one-day time-lag adjustment indicating the timing error is smaller (at sub-daily level) at this gauge, or that there are other effects. Simulated hydrographs that are too fast were found in many other locations. This general bias in timing may be related to the fact that a floodplain module is not included in the current version of the DRIVE model and the calibration of channel geometrics-related parameters (particularly the Manning roughness and channel width parameters) is lacking. The constant Manning roughness value of 0.03 used in this study is probably too low for many river basins. A simple increase of the Manning roughness to 0.035 resulted in significant improvements in DRIVE-RP for both gauges (Table 6). Fig. 14 shows the simulated and observed daily hydrographs (at gauge 3664100 [Brazil]) for a short time window as an example indicating the time delay error in the original DRIVE model simulation can be corrected through model calibration (here through a simple adjustment of the Manning roughness value).

The two examples indicate that improved calibration and better model parameterization will improve both runoff generation and runoff-routing modelling and should be a focus for the future. The major magnitude difference usually happens in flood season, which may indicate a seasonal oriented calibration, in addition to a floodplain module, might be required for more accurate flood magnitude estimation.

5. Discussion

In this study, we use a deterministic model for the real-time flood monitoring. Uncertainties can lie in both the model itself and model inputs. Many factors such as quality of precipitation estimation, human activities (particularly through reservoir/dam regulation, irrigation withdraw etc.), and model structure and parameterization can significantly impact model performance. Specifically for this study, satellite-based precipitation used here has generally good quality in the tropics, but with relatively more quality issues in higher latitudes, cool seasons and complex terrain; the DRIVE model in its current version doesn't include processes for man-made structures and human flow regulation, which exist extensively over the globe; even with only natural processes represented in the model, we have not performed any calibrations to tune the model toward reproducing better observations, though the model showed strong sensitivity to some parameters (e.g. Manning roughness). However, calibration of the hydrologic model can be problematic, if the observed discharge falls within the uncertainty of the simulated discharge [Biemans et al., 2009]. Calibration efforts in the future have to be implemented after an uncertainty analysis with particular attention paid to precipitation uncertainty for flood applications. Given a global domain in this study, the dominance of uncertainty sources will also be spatially dependent. Further work is needed to develop techniques or deploy existing ones from the literature [e.g. Beven and Freer, 2001; Renard et al., 2011; Demirel et al., 2013] for systematic uncertainty analysis. It is worth mentioning that the recent launch of the Global Precipitation Measurement (GPM) Core Observatory, a joint Earth-observing mission (as the follow-on of the TRMM mission) between NASA and the Japan Aerospace Exploration Agency (JAXA) [Hou et al., 2014], provides a good opportunity for further investigation of the uncertainties in our real-time flood modelling work. The DRIVE model is a participating hydrologic model in the GPM's Ground Validation (GV) Program to investigate the effects of precipitation uncertainty on model results and the uncertainty propagation in hydrologic processes by deploying various existing precipitation products (both conventional and satellite-based). We will report the results of that effort in a later paper.

Despite of the aforementioned uncertainties, we think the current model set-up and evaluation results provide a good basis for justification of the use of the GFMS for real-time flood monitoring, providing valuable information for flood analysis and for flood relief practice. Alfieri et al. [2013] recently performed a 21-year retrospective global hydrologic simulation driven by ERA-Interim reanalysis forcings at a 1/10th degree resolution. Their evaluation against streamflow observations at 620 GRDC gauges showed there were 58% of these gauges with positive daily NSC. In this study, we use satellite precipitation, and run the hydrologic model at 1/8th degree resolution while evaluating the model performance using 1,121 GRDC gauges (with more gauges with smaller upstream areas and shorter data time length). In our model performance statistics, we did not remove the gauges with upstream reservoirs as done by Alfieri et al. The validation metrics of the two studies are comparable. We also assume the uncertainties involved would not change the spatial-temporal pattern of the validation metrics derived in this study.

6. Summary and conclusions

An experimental real-time Global Flood Monitoring System (GFMS) using satellite-based precipitation information has been running routinely for the last few years with evaluations of previous versions [*Yilmaz et al.*, 2010; *Wu et al.*, 2012a] showing positive results, but indicating areas for additional improvement. In this paper we describe a new version of the system, present examples from the real-time system, and present an evaluation using a global flood event archive and streamflow observations. Real-time results from the system can be viewed at http://flood.umd.edu. For this new version of GFMS a widely used land surface model (LSM), the Variable Infiltration Capacity (VIC) model [*Liang et al.*, 1994 and 1996] is coupled with a newly developed hierarchical dominant river tracing-based runoff-routing (DRTR) model to form the Dominant river tracing-Routing Integrated with VIC Environment (DRIVE) model system. The DRTR routing model is a physically based routing model running on a grid system with parameterization of each routing model element (at either grid level or subgrid level) based on high resolution (1 km) hydrographic inputs through robust hierarchical DRT [*Wu et al.*, 2011 and 2012b]. The VIC model was modified, for real-time flood simulation, from its original individual grid cell based running mode to match the DRTR routing model structure with all grid cell calculations completed at each time step.

Examples from the GFMS real-time system over the North India are used to describe the flood detection/intensity algorithm, time history of regional maps of this parameter and present example of streamflow calculations. The validation and analysis based on the recent flood events over the upper Mississippi valley from the GFMS real-time system demonstrated that the real-time GFMS had a fairly good performance in flood occurrence detection, flood evolution and magnitude calculation according to observed daily streamflow data.

Results of 15-year retrospective calculations with the DRIVE system using research (TMPA-RP) and real-time (TMPA-RT) precipitation data sets indicate generally positive results. Global flood detection threshold maps based on the retrospective calculation of routed runoff at each grid location indicate a high level of correlation between the two rainfall data set inputs, with global and regional biases in the threshold related closely to differences in the mean rainfall. Using either rainfall data set the system detected about 87% of flood events of greater than one day duration across the globe. A further evaluation in 38 well-reported areas (to avoid under-reporting), also gave a POD of 0.90, with a false alarm ratio (FAR) of about 0.85 for flood events with duration greater than one day, which decreases to 0.70 for longer duration floods (greater than three days). Consistent with the findings of *Wu et al.* [2012a] in an evaluation of the previous version of our system, dams tended to undermine model skill in flood detection by leading to more false alarms. According to the statistics for the 18 WRAs with fewer than five dams (i.e., the most natural basins in our global comparison), the flood detection system being driven by the real-time precipitation information had a POD of 0.87, FAR of 0.66 and CSI of 0.32 for floods with duration longer than three days. Somewhat better statistics were achieved using the research quality precipitation information. In general, the new system provides improved statistics over the previous version of the GFMS when compared to the flood event inventory. This improvement is related primarily to the improved routing model and the use of a well-tested LSM (VIC), but also to some improvement to the real-time rainfall information.

The system was also tested against global streamflow observations from the Global Runoff Data Centre (GRDC). Using the research satellite precipitation information gave results of positive daily and monthly NSC values for 32% and 60% of the gauges with a mean of 0.22 and 0.39, respectively, which is promising considering the model was using only *a priori* parameters. The real-time precipitation data produced similar results in a parallel comparison, showing no significant difference at daily scale except in the northern mid-latitudes, where the research product produces better streamflow statistics than the real-time data, due to the positive influence of rain gauges in middle and higher latitudes. Validation using real-time precipitation across the tropics (30ºS–30ºN) gives positive daily Nash-Sutcliffe Coefficients for 107 out of 375 (28%) stations with a mean of 0.19 and 51% of the same gauges at monthly scale with a mean of 0.33. Better model performance was noted in deep tropics and sub-tropics as compared to mid-latitudes at monthly and daily scales. Analysis of individual observed vs. simulated hydrographs indicated that the simulated flood wave generally leads the observations by one day in the mean for the two selected gauges, possibly related to the current channel hydraulic parameter configurations and lack of floodplain delineation. The model appears sensitive to the Manning roughness coefficients. A sensitivity test with an increased Manning coefficient significantly reduced the lag and increased the NSC.

Uncertainties in the model inputs, model structure and parameter sets, and evaluation data can introduce considerable uncertainties in the results of this study. We'll investigate the uncertainty impacts on the flood estimation in future work, which is even more important in flood forecasting. However, both the flood event-based and the streamflow gauge-based evaluation indicated that even with the current quality of satellite-based precipitation, the model performance can likely be improved through hydrologic model development, particularly to include floodplain and reservoir/dam effects in the routing model (to decrease the false alarms) and better model parameterization and regional calibration. The model calibration strategy requires consideration of the uncertainty effects, particularly from the precipitation forcing. In addition to these directions, high-resolution (1 km) routing and water-storage calculations are being implemented for global real-time calculations, as well as combining the satellite precipitation information with precipitation forecasts from numerical weather prediction models to extend the real-time hydrological calculations into the future.

Acknowledgements

This research was supported by NASA's Applied Sciences Program. TMPA data used in this study were provided by the NASA/Goddard Space Flight Center's laboratory for Atmospheres and PPS. The TMPA real-time product is from ftp://trmmopen.gsfc.nasa.gov and the research product is from ftp://disc2.nascom.nasa.gov/. H.-Y. Li is supported by the Department of Energy Biological and Environmental Research (BER) Earth System Modeling (ESM) and Integrated Assessment Modeling (IAM) Programs through the Integrated Earth System Modeling (iESM). We gratefully acknowledge the historic discharge measurement provision by the Global Runoff

Data Centre. We also thank Justin Sheffield (University of Princeton) for sharing the VIC model set-up dataset.

Reference

Alfieri, L., Burek, P., Dutra, E., Krzeminski, B., Muraro, D., Thielen, J., and Pappenberger F. (2013), GloFAS - global ensemble streamflow forecasting and flood early warning, Hydrol. Earth Syst. Sci., 17, 1161-1175, doi:10.5194/hess-17-1161-2013.

Beven, K. J., and J. Freer (2001), Equifinality, data assimilation, and uncer- tainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology, J. Hydrol., 249(1–4), 11–29.

Biemans H, R.W. Hutjes, P. Kabat, B.J. Strengers, D. Gerten, S. Rost (2009), Effects of Precipitation Uncertainty on Discharge Calculations for Main River Basins. J. Hydrometeorol., 10(4):1011–1025. doi:10.1175/2008JHM1067.1.

Brakenridge, G.R. (2006), MODIS-based flood detection, mapping and measurement: The potential for operational hydrological applications, in Trans-boundary Floods: Reducing Risks Through Flood Management. New York: Springer-Verlag, pp. 1–12.

Christensen, N. S., Wood, A. W., Voisin, N., Lettenmaier, D. P., and Palmer, R. N. (2004), Effects of climate change on the hydrology and water resources of the Colorado River Basin, *Clim. Change*, 62, 337–363.

Christensen N. and D. P. Lettenmaier (2007), A multimodel ensemble approach to assessment of climate change impacts on the hydrology and water resources of the Colorado River basin. *Hydrol. Earth Syst. Sci*. 11:1417–1434.

Cherkauer, K. A. and D.P. Lettenmaier (1999), Hydrologic effects of frozen soils in the upper Mississippi River basin, *J. Geophys. Res.*, 104(D16), 19599-19610, 10.1029/1999JD900337.

Chow, Ven Te, D. R. Maidment, and L. W. Mays. 1988. Applied Hydrology. McGraw-Hill Inc., New York, New York.

Troy, T. J., E. F. Wood, and J. Sheffield (2008), An efficient calibration method for continental-scale land surface modeling, *Water Resour. Res.*, 44, W09411, doi:10.1029/2007WR006513.

Decharme, B., Alkama, R., Papa, F., Faroux, S., Douville, H., & Prigent, C. (2011), Global off-line evaluation of the ISBA-TRIP flood model. Climate Dynamics, 38(7-8), 1389–1412. doi:10.1007/s00382-011-1054-9

Demirel MC, M.J. Booij, A.Y. Hoekstra (2013), Effect of different uncertainty sources on the skill of 10 day ensemble low flow forecasts for two hydrological models. *Water Resour Res.* 49(April):n/a–n/a. doi:10.1002/wrcr.20294.

Elsner, M. M., L. Cuo, N. Voisin, J. Deems, A. F. Hamlet, J. A. Vano, K. E. B. Mickelson, S. Y. Lee, and D. P. Lettenmaier (2010), Implications of 21st century climate change for the hydrology of Washington State. *Climatic Change*, 102(1-2): 225-260, doi: 10.1007/s10584-010-9855-0.

Hamlet, Alan F., Philip W. Mote, Martyn P. Clark, Dennis P. Lettenmaier (2005), Effects of Temperature and Precipitation Variability on Snowpack Trends in the Western United States*. J. Climate, 18, 4545–4561. doi: http://dx.doi.org/10.1175/JCLI3538.1

Hamlet, A. F. and Lettenmaier, D. P. (2007), Effects of 20th century warming and climate variability on flood risk in the western U.S., *Water Resour. Res.*, 43, 1-17. doi: 10.1029/2006WR005099.

Hamlet, A. F., S. Y. Lee, K. E. B. Mickleson, and M.M. Elsner (2010), Effects of projected climate change on energy supply and demand in the Pacific Northwest and Washington State, *Climatic Change*, doi: 10.1007/s10584-010-9857-y.

Hamlet, A. F., Mote, P. W., Clark, M. P., and Lettenmaier, D. P. (2005), Effects of temperature and precipitation variability on snowpack trends in the western United States, *J. Climate*, 18, 4545–4561.

Harris, A., S. Rahman, F. Hossain, L. Yarborough, A. C. Bagtzoglou, and G. Easson (2007), Satellite-based flood modeling using TRMM-based rainfall products. *Sensors*, 7, 3416–3427.

Hong Y, R. F. Adler, F. Hossain, S. Curtis, and G. J. Huffman (2007), A first approach to global runoff simulation using satellite rainfall estimation. *Water Resour. Res.*, 43(8): W08502. doi: 10.1029/2006WR005739.

Horritt M.S., Mason D.C., Cobby D.M., Davenport I.J., Bates P.D. (2003), Waterline mapping in flooded vegetation from airborne SAR imagery, *Remote Sens. of Environ.*, 85 (3) , pp. 271-281.

Hou, A. Y., Kakar, R. K., Neeck, S., Azarbarzin, A. A., Kummerow, C. D., Kojima, M., ... & Iguchi, T. (2013), The Global Precipitation Measurement (GPM) Mission. Bulletin of the American Meteorological Society, e-View doi: http://dx.doi.org/10.1175/BAMS-D-13-00164.1

Huffman, G. J., R. F. Adler, D. T. Bolvin, G. Gu, E. J. Nelkin, K.P. Bowman, Y. Hong, E. F. Stocker, and D. B. Wolff (2007), The TRMM Multi-satellite Precipitation Analysis: Quasi-Global, Multi-Year, Combined-Sensor Precipitation Estimates at Fine Scale. *J. Hydrometeor.*, 8, 33-55.

Lehner, B., Verdin, K., Jarvis, A. (2008), New global hydrography derived from spaceborne elevation data. *Eos, Transactions*, AGU, 89(10), 93–94.

Lehner, B., et al. (2011), High-resolution mapping of the world's reservoirs and dams for sustainable river-flow management, Front. Ecol. Environ, 9(9), 494–502, doi:10.1890/100125.

Li, Hongyi, Mark S. Wigmosta, Huan Wu, Maoyi Huang, Yinghai Ke, André M. Coleman, L. Ruby Leung (2013), A Physically Based Runoff Routing Model for Land Surface and Earth System Models. *J. Hydrometeor*, 14, 808–828. doi: http://dx.doi.org/10.1175/JHM-D-12-015.1

Liang, X., D. P. Lettenmaier, E. F. Wood, and S. J. Burges (1994), A simple hydrologically based model of land surface water and energy fluxes for GSMs, *J. Geophys. Res.*, 99(D7), 14415–14428.

Liang, X., E. F. Wood, and D. P. Lettenmaier (1996), Surface soil moisture parameterization of the VIC-2L model: Evaluation and modifications, *Global Planet. Change*, 13, 195–206.

Lohmann, D., R. Nolte-Holube, and E. Raschke (1996), A large-scale horizontal routing model to be coupled to land surface parametrization schemes, *Tellus*, 48(A), 708-721.

Mason, D. C., Davenport, I. J., Neal, J. C., Schumann, G. P., & Bates, P. D. (2012), Near real-time flood detection in urban and rural areas using high-resolution synthetic aperture radar images. *Geoscience and Remote Sensing*, IEEE Transactions on, 50(8), 3041-3052.

Nash, J. E. and J. V. Sutcliffe (1970), River flow forecasting through conceptual models part I — A discussion of principles, *J. of Hydrol*., 10 (3), 282-290.

Ordoyne, C., and M. Friedl (2008), Using MODIS data to characterize seasonal inundation patterns in the Florida Everglades. *Remote Sens. of Environ.*, 112:4107–4119.

De Paiva, R. C. D., Buarque, D. C., Collischonn, W., Bonnet, M.-P., Frappart, F., Calmant, S., & Bulhões Mendes, C. A. (2013), Large-scale hydrologic and hydrodynamic modeling of the Amazon River basin. *Water Resour. Res.*, 49(3), 1226–1243. doi:10.1002/wrcr.20067

Peters-Lidard, C. D., Kumar, S. V., Mocko, D. M., & Tian, Y. (2011). Estimating evapotranspiration with land data assimilation systems. Hydrological Processes, 25(26), 3979–3992. doi:10.1002/hyp.8387

Reed, S., J. Schaake, and Z. Zhang (2007), A distributed hydrologic model and threshold frequency-based Method for flash flood forecasting at ungauged locations. *J. of Hydro.*, 337(3-4), 402-420. doi:10.1016/j.jhydrol.2007.02.015.

Renard, B., D. Kavetski, E. Leblois, M. Thyer, G. Kuczera, and S. W. Franks (2011), Toward a reliable decomposition of predictive uncertainty in hydrological modeling: Characterizing rainfall errors using conditional simulation, *Water Resour. Res.*, 47, W11516, doi: 10.1029/2011WR010643.

Rienecker, M.M., M.J. Suarez, R. Gelaro, R. Todling, J. Bacmeister, E. Liu, M.G. Bosilovich, S.D. Schubert, L. Takacs, G.-K. Kim, S. Bloom, J. Chen, D. Collins, A. Conaty, A. da Silva, et al. (2011), MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications. *J. Climate*, 24, 3624-3648, doi: 10.1175/JCLI-D-11-00015.1

Pan, M., H. Li, and E. Wood (2010), Assessing the skill of satellite-based precipitation estimates in hydrologic applications. *Water Resour. Res.*, 46, W09535, doi:10.1029/2009WR008290.

Shrestha, M. S., G. A. Artan, S. R. Bajracharya, and R. R. Sharma (2008), Using satellite-based rainfall estimates for streamflow modelling: Bagmati Basin. *J. of Flood Risk Man.*, 1(2): 89-99 DOI: 10.1111/j.1753-318X.2008.00011.x

Sivapalan, M. K. & Woods, R. A. (1995), Evaluation of the effects of general circulation model's subgrid variability and patchiness of rainfall and soil moisture and land surface water balance fluxes. In: *Advances in Hydrological Processes* (ed. by J. D. Raima & M. Sivapalan), 453-473. Wiley, Chichester, UK.

Strahler, A. N. (1957), "Quantitative analysis of watershed geomorphology", *Transactions of the American Geophysical Union* 38 (6): 913–920.

Storck, P., D. P. Lettenmaier, and S. Bolton (2002), Measurement of snow interception and canopy effects on snow accumulation and melt in mountainous maritime climate, Oregon, USA, *Water Resour. Res.*, 38, 1223–1238.

Su, F. G., Y. Hong, and D. P. Lettenmaier (2008), Evaluation of TRMM Multi-satellite Precipitation Analysis (TMPA) and its Utility in Hydrologic Prediction in La Plata Basin. *J. of Hydrometeor.*, 9(4), 622-640; DOI: 10.1175/2007JHM944.1

Su, F. G., H. Gao, G. J. Huffman, and D. P. Lettenmaier (2011), Potential Utility of the Real-Time TMPA-RT Precipitation Estimates in Streamflow Prediction. *J. of Hydrometeor.*, 12(3), 444-455; DOI: 10.1175/2010JHM1353.1

Voisin, N., F. Pappenberger, D. P. Lettenmaier, R. Buizza, and J. C. Schaake (2011), Application of a medium-range global hydrologic probabilistic forecast scheme to the Ohio River basin. *Wea. Forecasting*, 26, 425–446.

Vörösmarty, C. J., M. Meybeck, B. Fekete, K. Sharma, P. Green, and J. Syvitski (2003), Anthropogenic sediment retention: Major global-scale impact from the population of registered impoundments. *Global and Planetary Change*, 39: 169-190.

Vörösmarty, C. J., K. Sharma, B. Fekete, A. H. Copeland, J. Holden, J. Marble, and J. A. Lough (1997), The storage and aging of continental runoff in large reservoir systems of the world. *Ambio*, 26: 210-19.

Wang, J., H. Yang, L. Li, J. J. Gourley, I. K. Sadiq, K. K. Yilmaz, R. F. Adler, F. S. Policelli, S. Habib, D. Irwn, A. S. Limaye, T. Korme, and L. Okello (2011), The coupled routing and excess storage (CREST) distributed hydrological model. *Hydrol. Sci. J.*, 56(1), 84-98.

World Disasters Report. International Federation of Red Cross and Red Crescent Societies, 2012, 260-281.

Wu H., R. F. Adler, Y. Hong, Y. Tian, and F. Policelli (2012a), Evaluation of Global Flood Detection Using Satellite-Based Rainfall and a Hydrologic Model. *J. Hydrometeor*, 13, 1268–1284. doi: http://dx.doi.org/10.1175/JHM-D-11-087.1

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

Wu, H., J. S. Kimball, M. M. Elsner, N. Mantua, R. F. Adler, and J. Stanford (2012c), Projected climate change impacts on the hydrology and temperature of Pacific Northwest rivers, *Water Resour. Res.*, 48, W11530, doi: 10.1029/2012WR012082.

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

Yamazaki, D., Kanae, S., Kim, H., & Oki, T. (2011), A physically based description of floodplain inundation dynamics in a global river routing model. *Water Resour. Res.*, 47(4), 1–21. doi:10.1029/2010WR009726

Yilmaz, K. K., R. F. Adler, Y. Tian, Y. Hong, and H. F. Pierce (2010), Evaluation of a satellite-based global flood monitoring system. *Int. J. Remote Sens.*, 31, 3763-3782, doi: 10.1080/01431161.2010.483489.

Zhao, R. J., and X. R. Liu (1995), The Xinanjiang model. In: Computer Models of Watershed Hydrology (ed. by V. P. Singh), 215–232. Water Resources Publications, Colorado, USA.

Figure captions

Figure 1. The DRTR routing model concept on river basin drainage system at (a) grid and (b)-(d) subgrid scales using a real river basin (Mbemkuru river basin, Southeast of Tanzania) as example. The light blue lines in (a) is the baseline high resolution (1km) river network from HydroSHEDS and the red lines are the DRT-derived coarse-resolution rivers (1/8th degree in this case).

Figure 2. Example of the DRIVE model major outputs from the real-time GFMS with screenshots from http://flood.umd.edu. The examples show the model global outputs of routed runoff (a), streamflow (b), flood detection and intensity (water depth [mm] above flood threshold) (c) at a 3-hour time interval (15Z01Jul2013). An example of global TMPA 3B42V7 real-time rainfall input data at the same time interval is shown in (d). The example also shows the spatial-temporal evolution (at daily interval) of the flood event happened in North India during Jun 15, 2013 to Jun 20, 2013 (c1-6).

Figure 3. Snapshots from the real-time GFMS (online: http://flood.umd.edu) for major two flood waves, covering April to early Jun, 2013, in sub-basin rivers upstream of the Mississippi River, including (a-b) the flood detection and intensity (water depth above flood threshold), (c-d)

previous 7-day accumulated precipitation according TMPA RT, (e-f) streamflow. All data are at 1/8th (~12km) resolution.

Figure 4. (a) The DRIVE-RT simulated streamflow against observed data from 29 USGS gauges on the rivers of the upper Mississippi river basin for a two-year retrospective period (2011-06-12 to 2013-06-12). All USGS gauges are shown in filled circles, while their colors are turned into green when the model-estimated positive daily NSCs at the corresponding locations. (b)-(e) show the observed and simulated daily hydrographs for four of the gauges, with locations indicated in (a), during the Spring and early Summer flooding period (April 1 to Jun 9, 2013).

Figure 5. (a) Flood threshold map (according to routed runoff [mm]) based on 11-year (2001-2011) retrospective simulation by DRIVE-RT. (b) The difference between the flood threshold maps derived by the DRIVE-RT and DRIVE-RP (DRIVE-RT - DRIVE-RP).

Figure 6. (a) Mean annual precipitation map according to TMPA RT from 2001 to 2011; (b) the difference between the mean annual precipitation for TMPA RP and RT over the same period.

Figure 7. Example of well-reported areas (shaded yellow) and their corresponding FAR metrics (according to DRIVE-RT for all floods with duration greater than 1 day) in the part of Asia that tends to have more floods. The background image is the zoomed –in flood threshold difference (DRIVE-RT - DRIVE-RP) from Fig. 5b.

Figure 8. The flood detection metrics POD (a), FAR (b) and CSI (c) across 38 well-reported areas for DRIVE-RP and DRIVE-RT results for all floods with duration greater than three days, against DFO flood inventory data during 2001 to 2011. The numbers of dams upstream of each well-reported area are listed along the X-axis.

Figure 9. DRIVE-RP model performance (monthly NSC) in reproducing monthly streamflow during 2001-2011, when driven by TMPA RP research precipitation data, at 1,121 GRDC streamflow gauges across the globe. All GRDC gauges are shown as filled circles, while at each gauge if the model performance is of a positive value for monthly NSC, the gauge color turns into green or purple in accordance to the value of NSC.

Figure 10. Histogram distribution of the number of gauges with positive (a) monthly and (b) daily NSC values for DRIVE-RP and DRIVE-RT simulation for 2001-2011.

Figure 11. The percentage of gauges in each latitude band (defined in the Section 4.4.2) for which the DRIVE model showed positive daily NSCs using TMPA RP and TMPA RT precipitation input. The X-axis values are the central latitude for each band.

Figure 12. The daily NSC (a) and (b), and MARE (c) and (d) metrics for the region of South America from (a, c) DRIVE-RP and (b, d) DRIVE-RT model results.

Figure 13. Examples of the simulated and observed hydrographs at two gauges. The gauge locations are indicated as filled circles in Fig. 6b.

Figure 14. Example of hydrographs in a short time window (April 11, 2005 –December 31, 2005) computed by the DRIVE-RP. The red curve stands for the original DRIVE-RP modelling with Manning coefficient of 0.03 for both stem river and sub-grid tributaries; the black curve is from DRIVE-RP using a Manning coefficient of 0.035, while the green curve is negative one day corrected original DRIVE-RP simulated hydrograph.

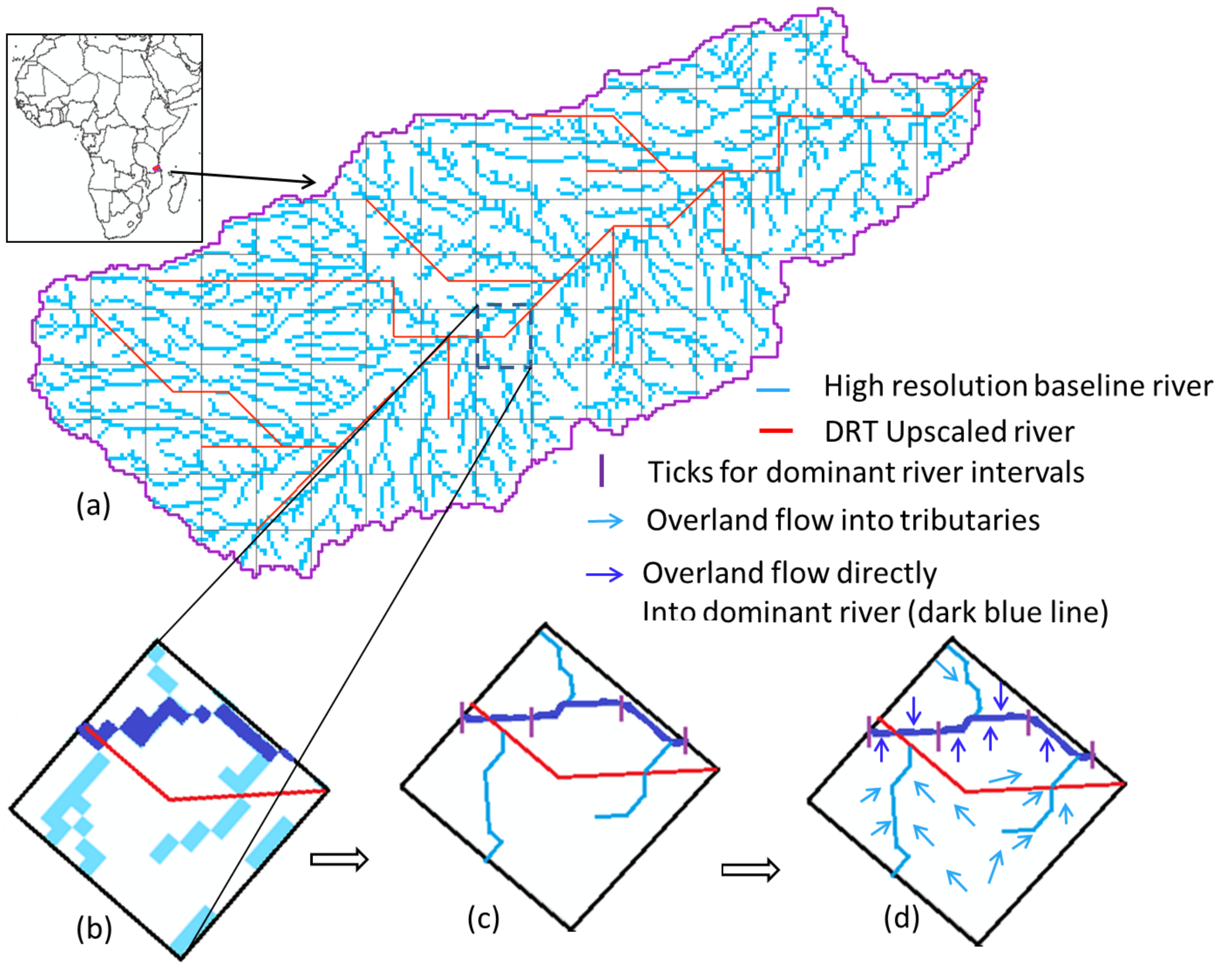

Figure 1. The DRTR routing model concept on river basin drainage system at (a) grid and (b)-(d) subgrid scales using a real river basin (Mbemkuru river basin, Southeast of Tanzania) as example. The light blue lines in (a) is the baseline high resolution (1km) river network from HydroSHEDS and the red lines are the DRT-derived coarse-resolution rivers (1/8th degree in this case).

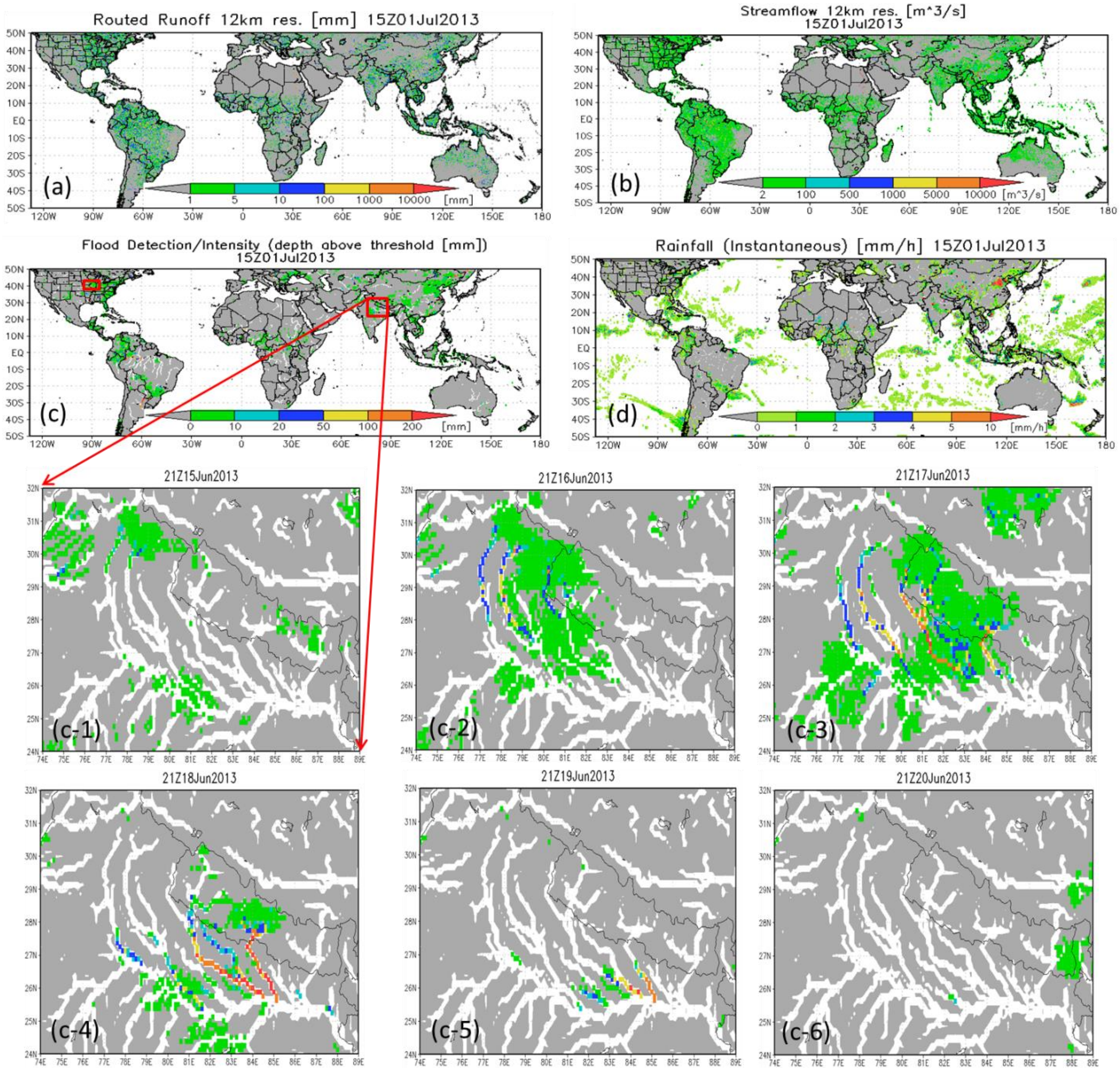

Figure 2. Example of the DRIVE model major outputs from the real-time GFMS with screenshots from http://flood.umd.edu. The examples show the model global outputs of routed runoff (a), streamflow (b), flood detection and intensity (water depth [mm] above flood threshold) (c) at a 3-hour time interval (15Z01Jul2013). An example of global TMPA 3B42V7 real-time rainfall input data at the same time interval is shown in (d). The example also shows the spatial-temporal evolution (at daily interval) of the flood event happened in North India during Jun 15, 2013 to Jun 20, 2013 (c1-6).

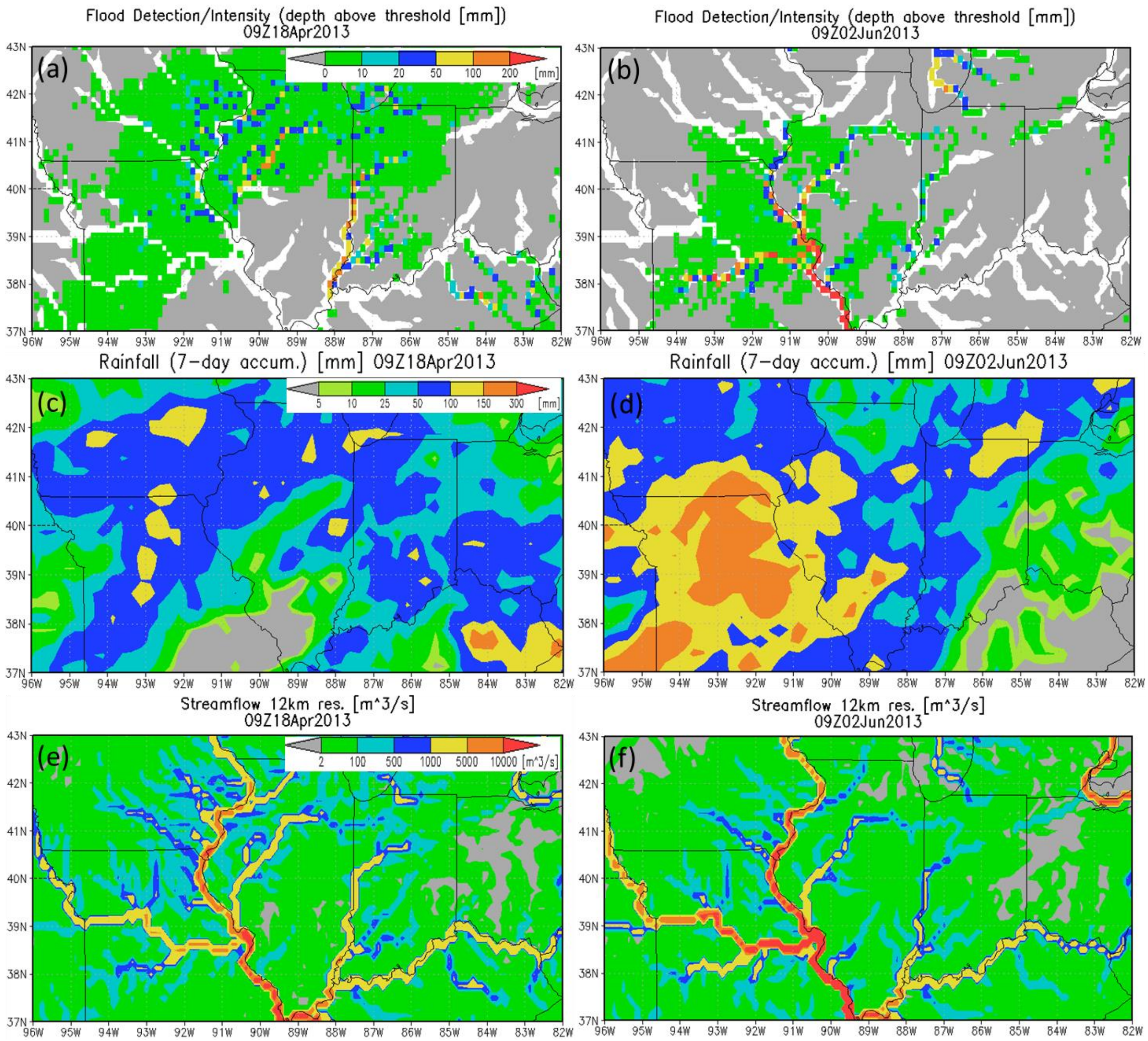

Figure 3. Snapshots from the real-time GFMS (online: http://flood.umd.edu) for major two flood waves, covering April to early Jun, 2013, in sub-basin rivers upstream of the Mississippi River, including (a-b) the flood detection and intensity (water depth above flood threshold), (c-d) previous 7-day accumulated precipitation according TMPA V7RT, (e-f) streamflow. All data are at 1/8th (~12km) resolution.

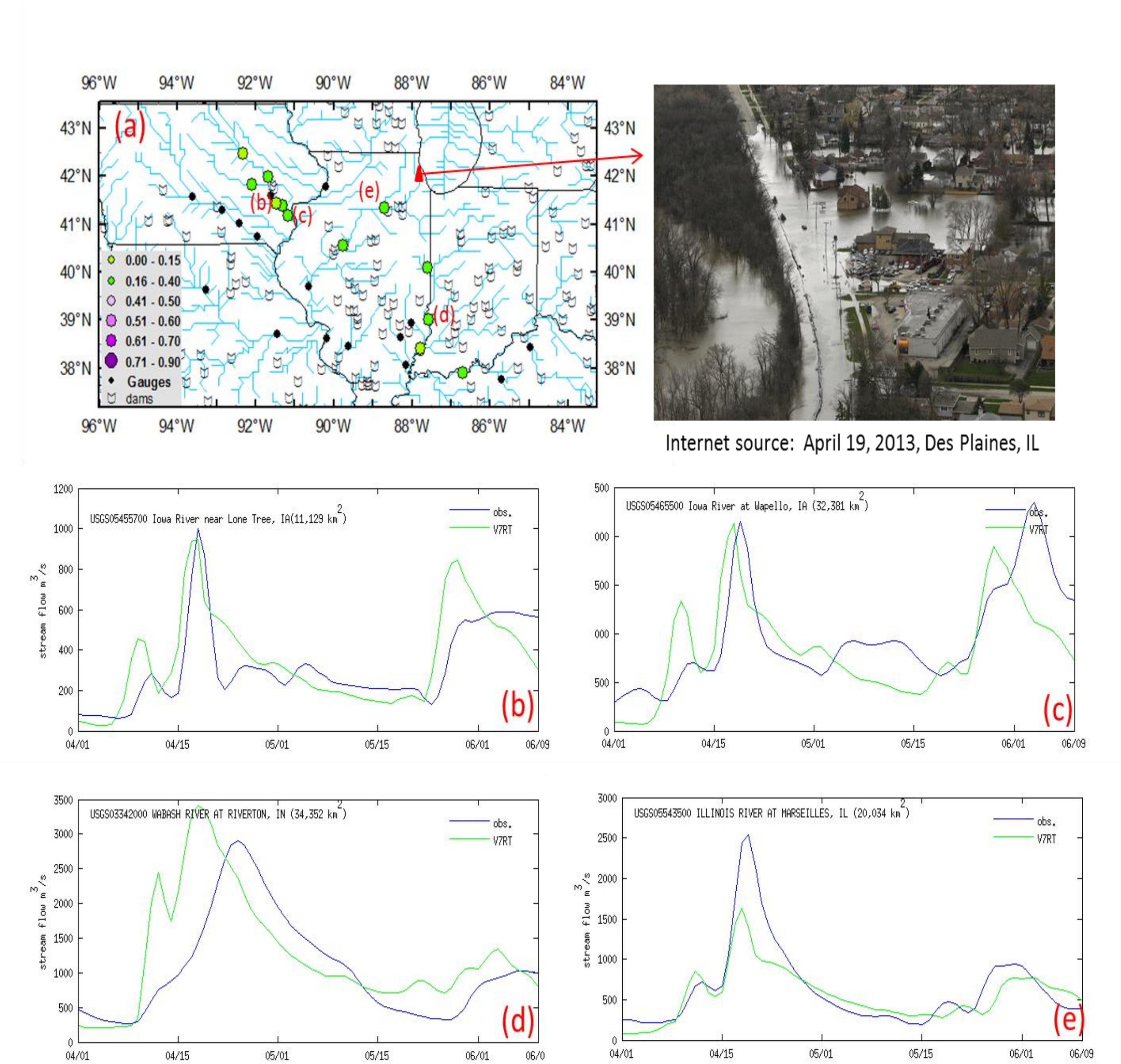

Figure 4. (a) The DRIVE-RT simulated streamflow against observed data from 29 USGS gauges on the rivers of the upper Mississippi river basin for a two-year retrospective period (2011-06-12 to 2013-06-12). All USGS gauges are shown in filled circles, while their colors are turned into green when the model-estimated positive daily NSCs at the corresponding locations. (b)-(e) show the observed and simulated daily hydrographs for four of the gauges, with locations indicated in (a), during the Spring and early Summer flooding period (April 1 to Jun 9, 2013).

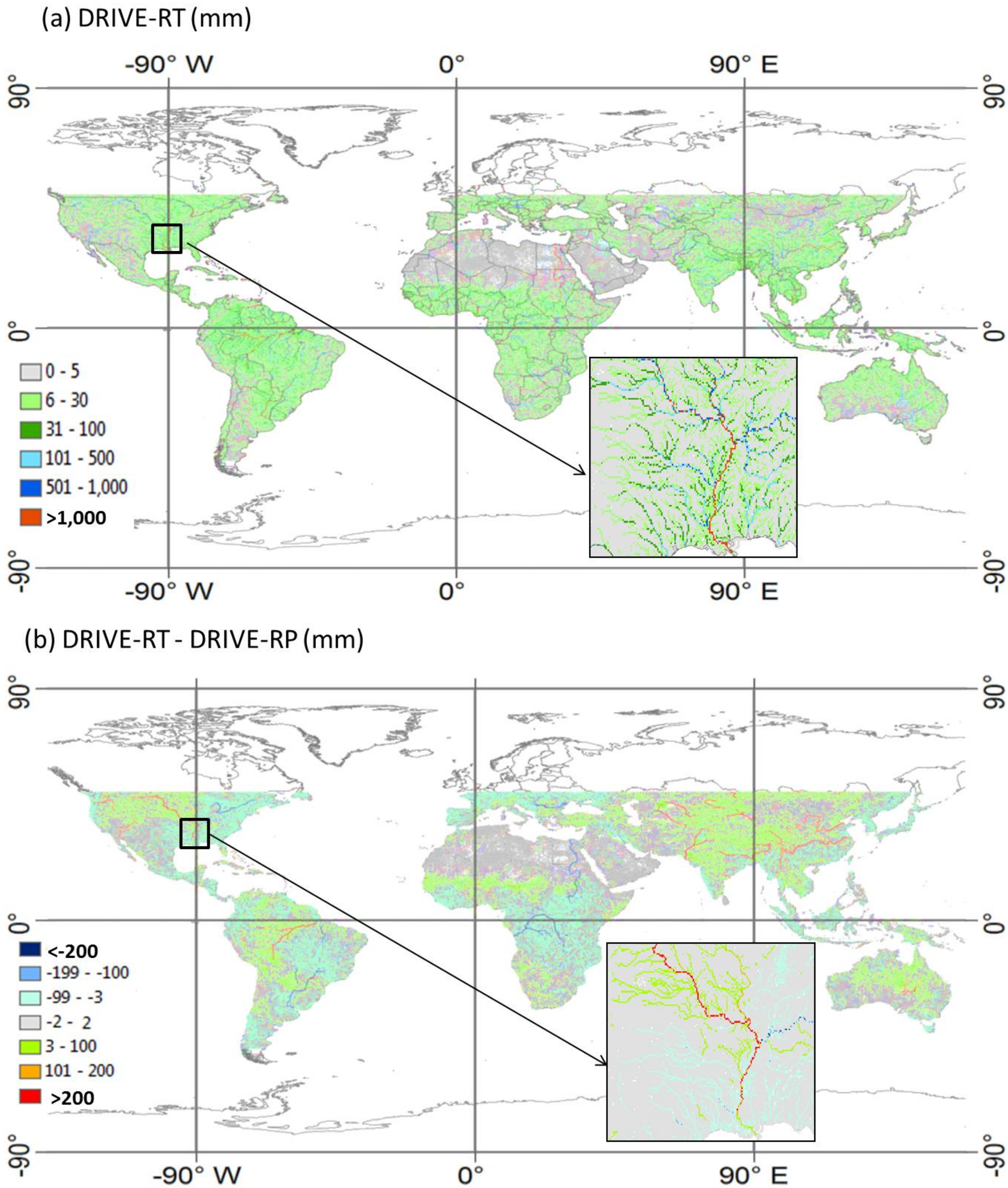

Figure 5. (a) Flood threshold map (according to routed runoff [mm]) based on 11-year (2001-2011) retrospective simulation by DRIVE-RT. (b) The difference between the flood threshold maps derived by the DRIVE-RT and DRIVE-RP (DRIVE-RT - DRIVE-RP).

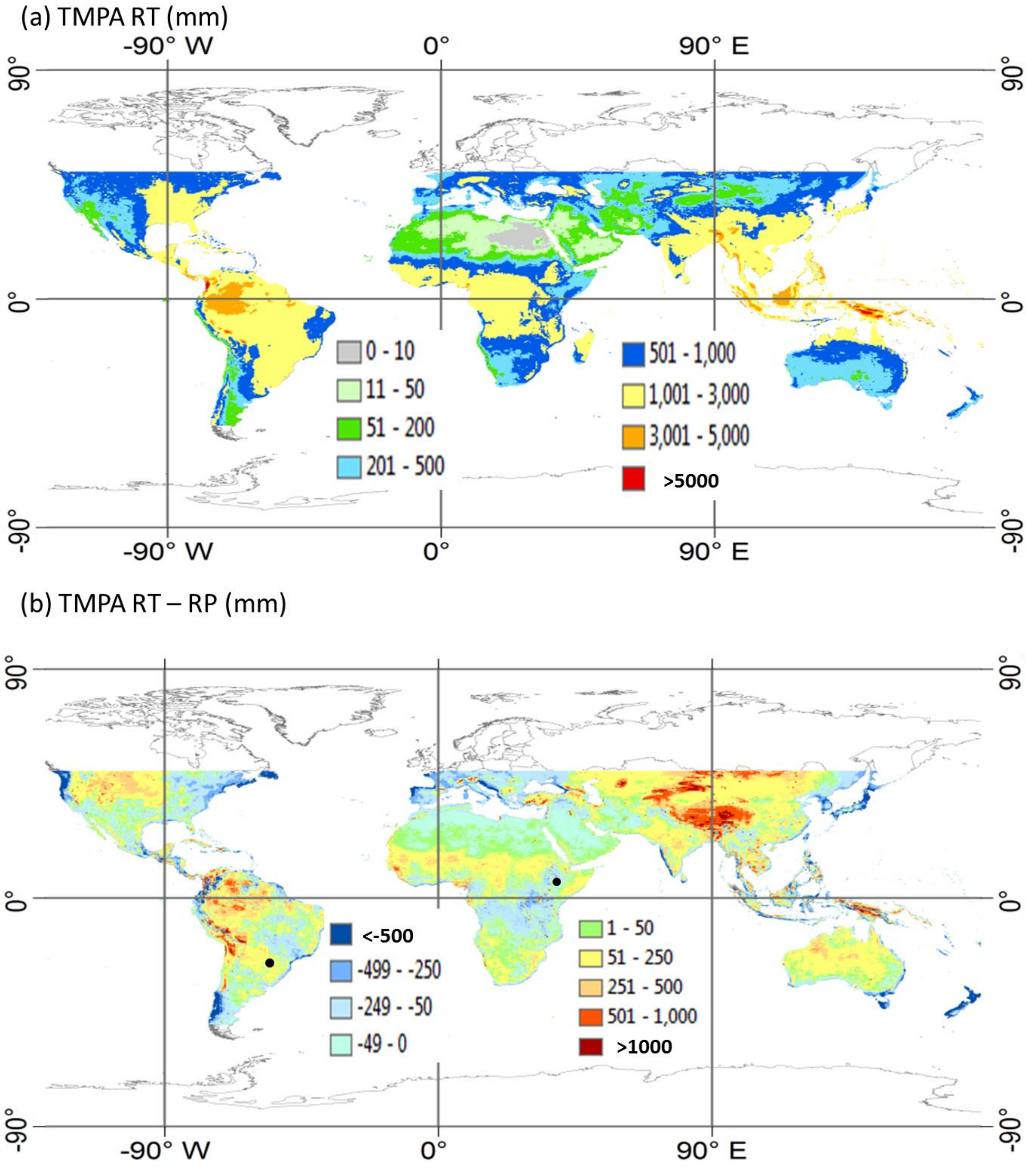

Figure 6. (a) Mean annual precipitation map according to TMPA RT from 2001 to 2011; (b) the difference between the mean annual precipitation for TMPA RP and RT over the same period.

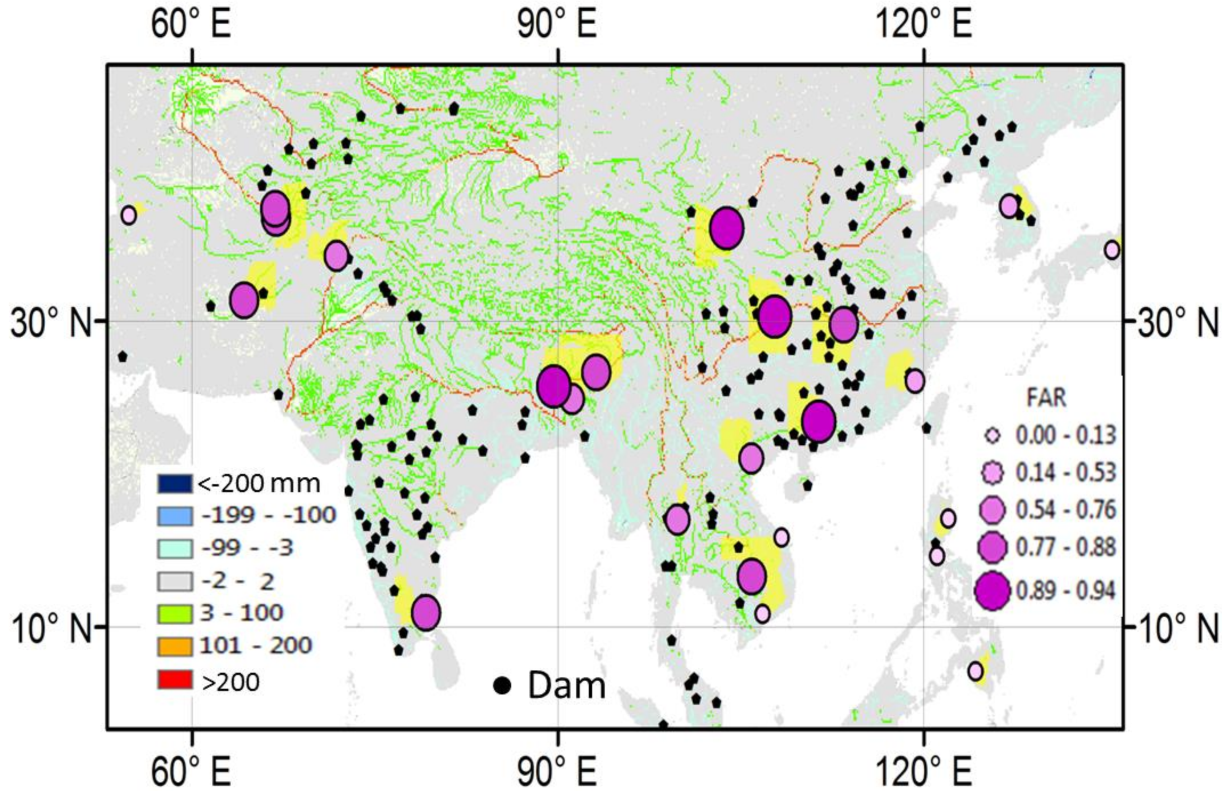

Figure 7. Example of well-reported areas (shaded yellow) and their corresponding FAR metrics (according to DRIVE-RT for all floods with duration greater than 1 day) in the part of Asia that tends to have more floods. The background image is the zoomed –in flood threshold difference (DRIVE-RT - DRIVE-RP) from Fig. 5b.

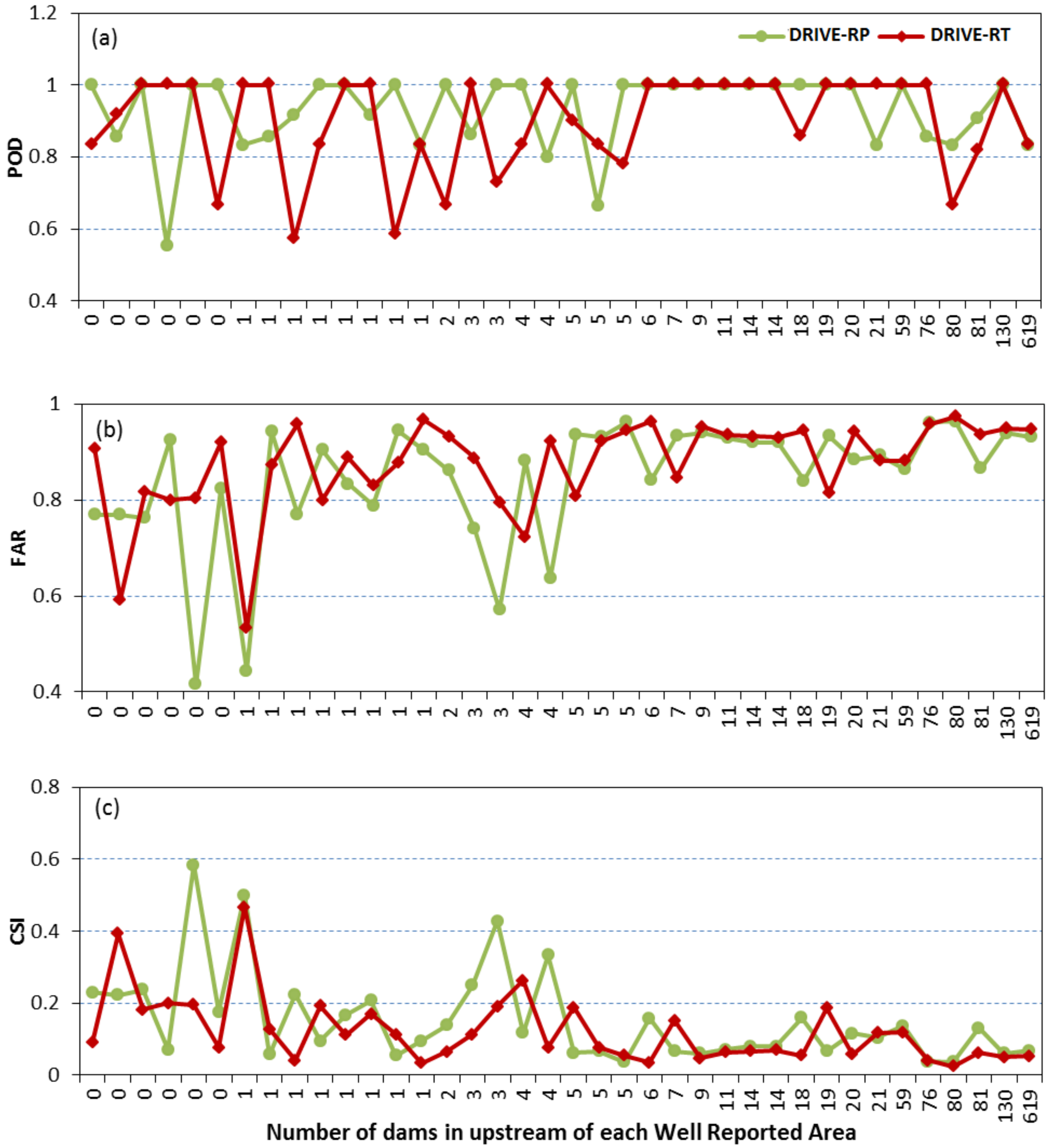

Figure 8. The flood detection metrics POD (a), FAR (b) and CSI (c) across 38 well-reported areas for DRIVE-RP and DRIVE-RT results for all floods with duration greater than three days, against DFO flood inventory data during 2001 to 2011. The numbers of dams upstream of each well-reported area are listed along the X-axis.

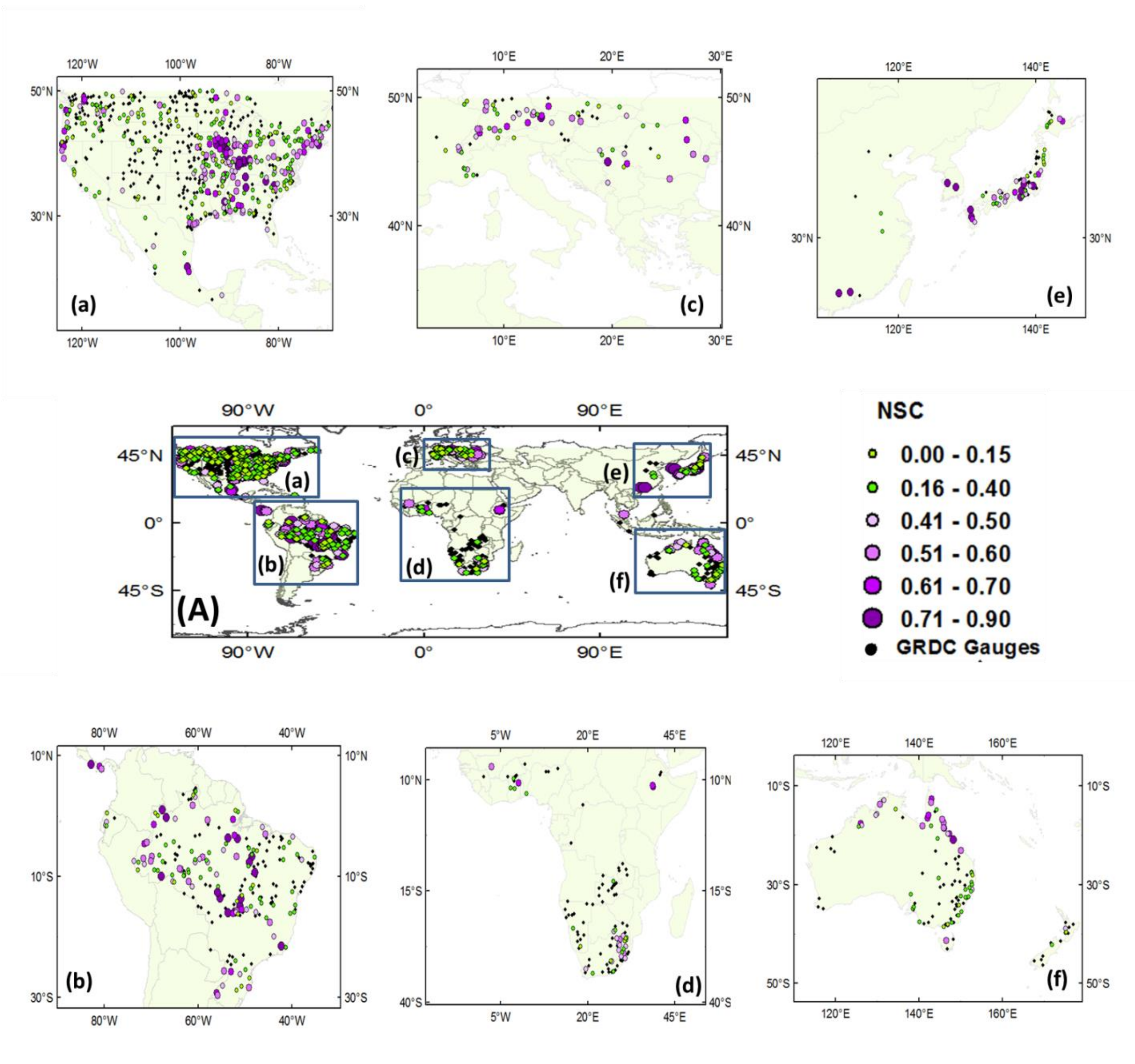

Figure 9. DRIVE-RP model performance (monthly NSC) in reproducing monthly streamflow during 2001-2011, when driven by TMPA RP research precipitation data, at 1,121 GRDC streamflow gauges across the globe. All GRDC gauges are shown as filled circles, while at each gauge if the model performance is of a positive value for monthly NSC, the gauge color turns into green or purple in accordance to the value of NSC.

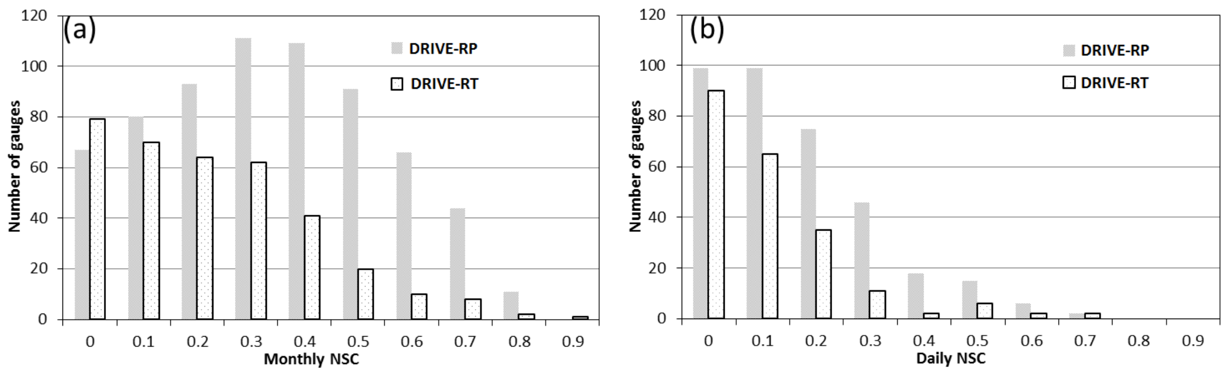

Figure 10. Histogram distribution of the number of gauges with positive (a) monthly and (b) daily NSC values for DRIVE-RP and DRIVE-RT simulation for 2001-2011.

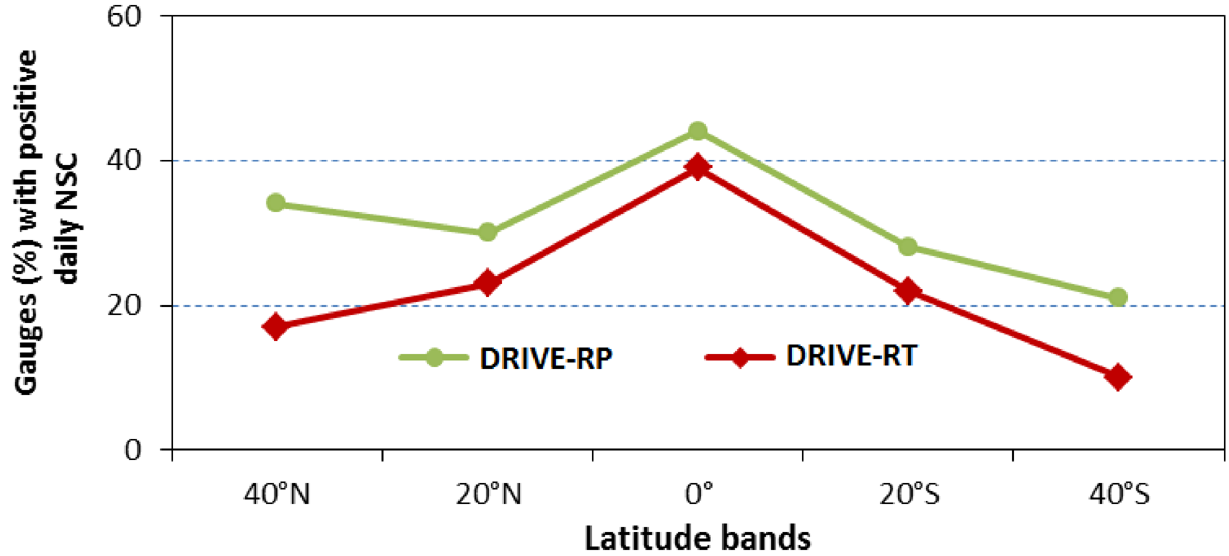

Figure 11. The percentage of gauges in each latitude band (defined in the Section 4.4.2) for which the DRIVE model showed positive daily NSCs using TMPA RP and TMPA-RT precipitation input. The X-axis values are the central latitude for each band.

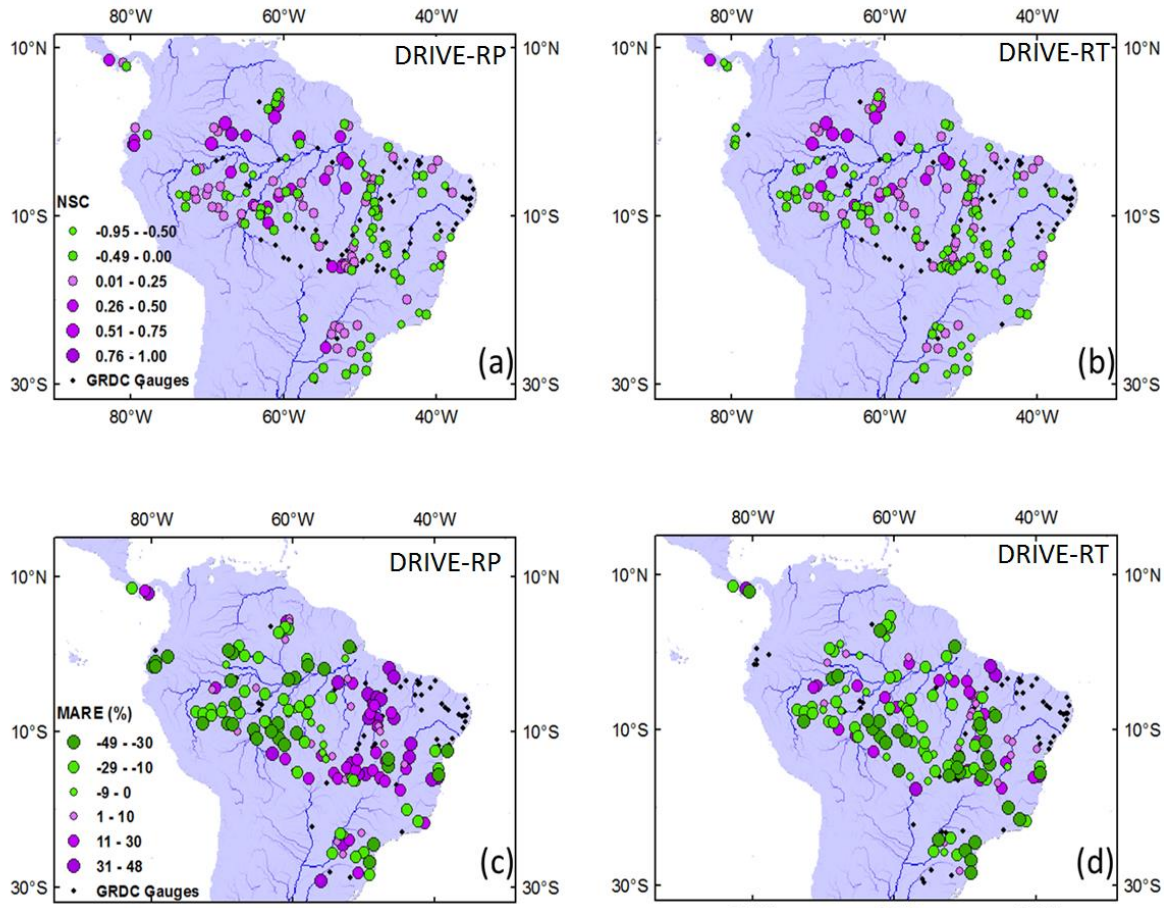

Figure 12. The daily NSC (a) and (b), and MARE (c) and (d) metrics for the region of South America from (a, c) DRIVE-RP and (b, d) DRIVE-RT model results.

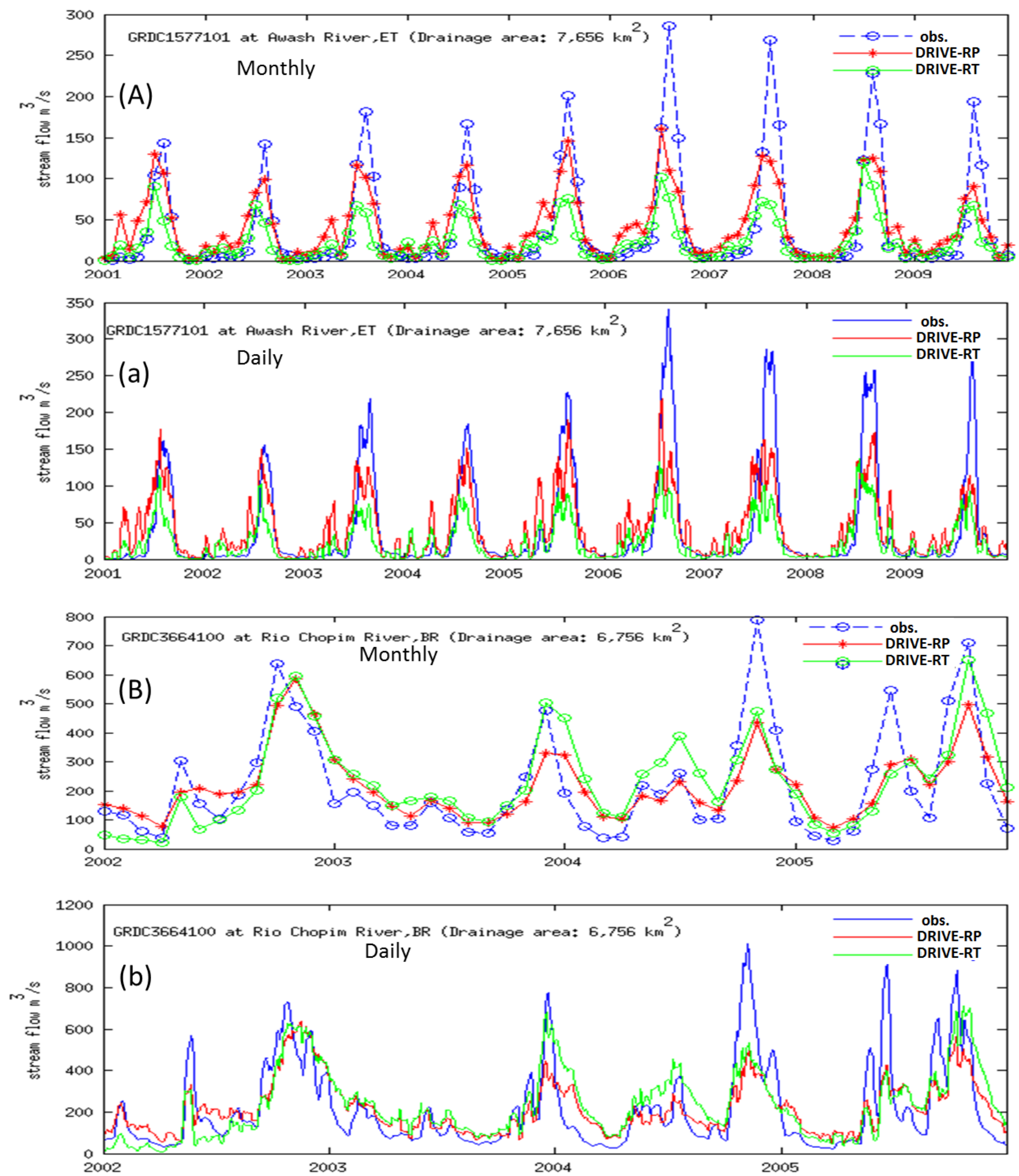

Figure 13. Examples of the simulated and observed hydrographs at two gauges. The gauge locations are indicated as filled circles in Fig. 6b.

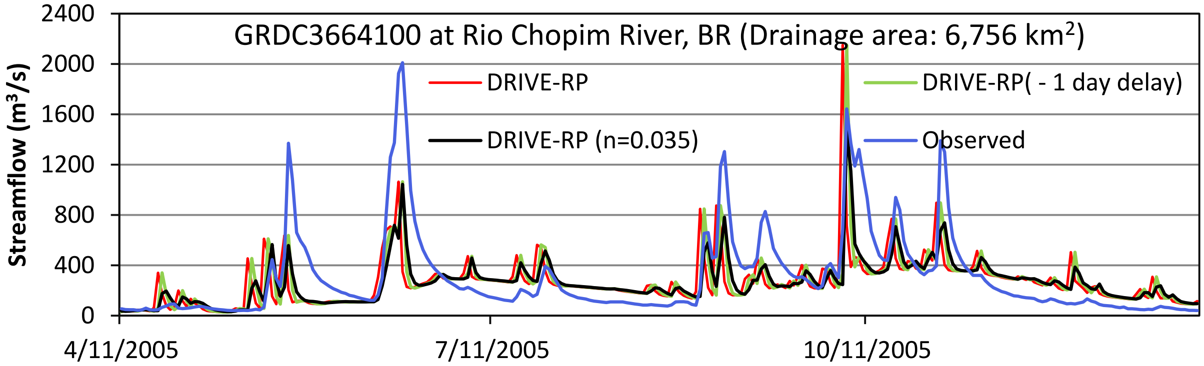

Figure 14. Example of hydrographs in a short time window (April 11, 2005 –December 31, 2005) computed by the DRIVE-RP. The red curve stands for the original DRIVE-RP modelling with Manning coefficient of 0.03 for both stem river and sub-grid tributaries; the black curve is from DRIVE-RP using a Manning coefficient of 0.035, while the green curve is negative one day corrected original DRIVE-RP simulated hydrograph.

Table 1. Flood detection verification against the DFO flood database over the 38 well reported areas (WRAs) for floods with duration more than 1 day.

Metrics	**POD**	**FAR**	**CSI**
Metrics averaged over all the 38 WRAs			
DRIVE-RT	0.90	0.88	0.12
DRIVE-RP	0.93	0.84	0.15
Metrics averaged over the 20 WRAs with ≥ 5 dam			
DRIVE-RT	0.93	0.92	0.08
DRIVE-RP	0.94	0.90	0.10
Metrics averaged over the 18 WRAs with<5 dam			
DRIVE-RT	0.86	0.83	0.17
DRIVE-RP	0.92	0.78	0.21

Table 2. The same as Table 1, but for floods with duration more than 3 days.

Metrics	**POD**	**FAR**	**CSI**
Metrics averaged over all the 38 WRAs			
DRIVE-RT	0.90	0.73	0.25
DRIVE-RP	0.93	0.65	0.34
Metrics averaged over the 20 WRAs with ≥ 5 dam			
DRIVE-RT	0.93	0.80	0.19
DRIVE-RP	0.94	0.73	0.26
Metrics averaged over the 18 WRAs with<5 dam			
DRIVE-RT	0.87	0.66	0.32
DRIVE-RP	0.92	0.56	0.43

Table 3. The metrics for model performance in streamflow simulation, at daily and monthly time intervals for continuous years, against 1,121 GRDC river gauges across the globe (-50°S to 50°N). Metrics are listed for global and regional areas (from deep tropics to higher latitudes). The time period of daily streamflow gauge data ranges in 1~11 years. N_d and N_m stand for daily and monthly NSC respectively. R_d and R_m stand for daily and monthly correlation coefficients respectively. MARE is the mean annual relative error.

		Daily NSC		Monthly NSC		Correlation Coeff.		MARE<30%
		N_d>0	N_d>0.4	N_m>0	N_m>0.4	R_d>0.4	R_m>0.4	
Global (-50°S to 50°N) with 1,121 gauges								
% of gauges	RP	32	4	60	29	58	99	38
	RT	*19*	*1*	*32*	*7*	*42*	*95*	*27*
Mean metrics	RP	0.22	0.52	0.39	0.57	0.57	0.67	-0.3%
	RT	*0.16*	*0.57*	*0.27*	*0.54*	*0.53*	*0.53*	*-2.9%*
-10°S~10°N with 141 gauges								
% of gauges	RP	44	9	62	31	76	99	44
	RT	*39*	*6*	*57*	*22*	*75*	*98*	*51*
Mean metrics	RP	0.25	0.55	0.41	0.58	0.64	0.70	-6.8%
	RT	*0.23*	*0.60*	*0.36*	*0.58*	*0.61*	*0.66*	*-5.5%*
10°N to 30°N with 43 gauges								
% of gauges	RP	30	5	54	28	51	95	37
	RT	*23*	*2*	*51*	*19*	*42*	*95*	*33*
Mean metrics	RP	0.17	0.47	0.41	0.59	0.58	0.72	-0.3%
	RT	*0.18*	*0.54*	*0.33*	*0.60*	*0.54*	*0.60*	*-0.6%*
30°N to 50°N with 671 gauges								
% of gauges	RP	34	4	66	31	61	99	41
	RT	*17*	*1*	*25*	*3*	*39*	*96*	*24*
Mean metrics	RP	0.21	0.52	0.38	0.56	0.56	0.66	1.1%
	RT	*0.13*	*0.53*	*0.21*	*0.50*	*0.51*	*0.45*	*-1.2%*
-10°S to -30°S with 191 gauges								
% of gauges	RP	28	1	52	28	59	99	34
	RT	*22*	*0*	*45*	*11*	*46*	*98*	*35*
Mean metrics	RP	0.17	0.46	0.30	0.56	0.54	0.46	2.0%
	RT	*0.11*	-	*0.29*	*0.50*	*0.52*	*0.56*	*-4.9%*
-30°S to -50°S with 75 gauges								
% of gauges	RP	21	0	44	8	5	96	20
	RT	*10*	*0*	*24*	*0*	*1*	*88*	*9*
Mean metrics	RP	0.05	-	0.25	0.46	0.52	0.57	-9.2%
	RT	*0.01*	-	*0.06*	-	*0.44*	*0.34*	*6%*

Table 4. The same as Table 3 but for summer seasons (i.e. JJA is used for deep tropic and Northern hemisphere while DJF is used for Southern hemisphere)

		Daily NSC		Monthly NSC		Correlation Coeff.		MARE<30%
		N_d>0	N_d>0.4	N_m>0	N_m>0.4	R_d>0.4	R_m>0.4	
-10°S~10°N with 141 gauges								
% of gauges	RP	14	5	31	11	51	84	33
	RT	*14*	*4*	*18*	*8*	*55*	*86*	*31*
Mean metrics	RP	0.32	0.68	0.32	0.59	0.65	0.64	-3.2%
	RT	*0.26*	*0.48*	*0.31*	*0.53*	*0.61*	*0.61*	*-2.7%*
10°N to 30°N with 43 gauges								
% of gauges	RP	19	0	28	14	37	86	23
	RT	*16*	*2*	*35*	*12*	*26*	*84*	*14*
Mean metrics	RP	0.10	-	0.31	0.54	0.56	0.65	0.1%
	RT	*0.16*	*0.43*	*0.30*	*0.52*	*0.56*	*0.62*	*-1%*
30°N to 50°N with 671 gauges								
% of gauges	RP	25	4	43	22	58	99	25
	RT	*10*	*1*	*19*	*3*	*30*	*92*	*21*
Mean metrics	RP	0.22	0.54	0.41	0.61	0.56	0.72	1.3%
	RT	*0.16*	*0.53*	*0.25*	*0.57*	*0.52*	*0.48*	*-1.4%*
-10°S to -30°S with 191 gauges								
% of gauges	RP	19	0	42	19	37	93	26.2
	RT	*13*	*0*	*26*	*6*	*19*	*85*	*31*
Mean metrics	RP	0.14	-	0.37	0.57	0.51	0.66	-3.5%
	RT	*0.10*	*-*	*0.26*	*0.49*	*0.48*	*0.48*	*1.4%*
-30°S to -50°S with 75 gauges								
% of gauges	RP	7	0	31	8	8	72	15
	RT	*8*	*0*	*11*	*0*	*3*	*63*	*11*
Mean metrics	RP	0.11	-	0.27	0.55	0.52	0.62	-3.7%
	RT	*0.03*	*-*	*0.06*	*-*	*0.49*	*0.37*	*2.3%*

Table 5. The same as Table 3 but for winter seasons (i.e. DJF is used for deep tropic and Northern hemisphere while JJA is used for Southern hemisphere)

		Daily NSC		Monthly NSC		Correlation Coeff.		MARE<30%
		N_d>0	N_d>0.4	N_m>0	N_m>0.4	R_d>0.4	R_m>0.4	
-10°S~10°N with 141 gauges								
% of gauges	RP	17	3	36	14	43	87	34
	RT	*15*	*4*	*26*	*11*	*23*	*89*	*37*
Mean metrics	RP	0.23	0.55	0.34	0.57	0.61	0.62	-2.8%
	RT	*0.24*	*0.55*	*0.31*	*0.53*	*0.60*	*0.47*	*-5.3%*
10°N to 30°N with 43 gauges								
% of gauges	RP	9	0	28	9	28	75	30
	RT	*14*	*0*	*26*	*2*	*21*	*63*	*28*
Mean metrics	RP	0.01	-	0.25	0.51	0.56	0.62	1.2%
	RT	*0.04*	*-*	*0.16*	*0.45*	*0.61*	*0.54*	*-2.3%*
30°N to 50°N with 671 gauges								
% of gauges	RP	22	3	34	16	48	92	39
	RT	*8*	*1*	*11*	*3*	*33*	*78*	*19*
Mean metrics	RP	0.02	0.12	0.40	0.62	0.55	0.61	-6.2%
	RT	*0.01*	*0.07*	*0.27*	*0.57*	*0.52*	*0.49*	*-5.5%*
-10°S to -30°S with 191 gauges								
% of gauges	RP	7	1	10	4	28	66	15
	RT	*5*	*1*	*7*	*3*	*15*	*56*	*14*
Mean metrics	RP	0.02	0.1	0.31	0.64	0.60	0.57	3.0%
	RT	*0.01*	*0.08*	*0.23*	*0.48*	*0.52*	*0.44*	*-1.6%*
-30°S to -50°S with 75 gauges								
% of gauges	RP	15	0	42	19	9	85	23
	RT	*15*	*0*	*21*	*1*	*8*	*76*	*11*
Mean metrics	RP	0.09	-	0.30	0.51	0.45	0.65	-8.9%
	RT	*0.09*	*-*	*0.22*	*0.44*	*0.47*	*0.44*	*-6.1%*

Table 6. DRIVE model streamflow simulation performance at two selected gauges. n is the Manning roughness coefficient, which was used uniformly globally for both the dominant rivers and tributaries. The metrics were also calculated by delaying the simulated streamflow time series by one day which resulted in the maximum correlation coefficient between simulated and observed hydrographs.

		N_d	N_m	R_d	MARE
GRDC 1577101 (2001-2009)	DRIVE-RP	0.35	0.67	0.62	5.4%
	DRIVE-RP(-1day)	0.35	0.67	0.63	5.4%
	DRIVE-RP ($n = 0.035$)	0.45	0.68	0.67	5.7%
	DRIVE-RT	0.29	0.40	0.60	-41%
	DRIVE-RT(-1day)	0.30	0.41	0.61	-41%
GRDC 3664100 (2002-2005)	DRIVE-RP	0.28	0.65	0.55	0.5%
	DRIVE-RP(-1day)	0.48	0.65	0.69	0.5%
	DRIVE-RP ($n = 0.035$)	0.55	0.64	0.75	0.6%
	DRIVE-RT	0.17	0.59	0.53	9.6%
	DRIVE-RT(-1day)	0.43	0.58	0.68	9.5%