

Evaluation of Global Flood Detection Using Satellite-Based Rainfall and a Hydrologic Model

HUAN WU AND ROBERT F. ADLER

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

YANG HONG

School of Civil Engineering and Environmental Sciences, and Atmospheric Radar Research Center, University of Oklahoma, Norman, Oklahoma

YUDONG TIAN

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

FRITZ POLICELLI

NASA Goddard Space Flight Center, Greenbelt, Maryland

(Manuscript received 22 July 2011, in final form 2 April 2012)

ABSTRACT

A new version of a real-time global flood monitoring system (GFMS) driven by Tropical Rainfall Measuring Mission (TRMM) Multisatellite Precipitation Analysis (TMPA) rainfall has been developed and implemented using a physically based hydrologic model. The purpose of this paper is to evaluate the performance of this new version of the GFMS in terms of flood event detection against flood event archives to establish a baseline of performance and directions for improvement. This new GFMS is quantitatively evaluated in terms of flood event detection during the TRMM era (1998–2010) using a global retrospective simulation (3-hourly and $1/8^\circ$ spatial resolution) with the TMPA 3B42V6 rainfall. Four methods were explored to define flood thresholds from the model results, including three percentile-based statistical methods and a Log Pearson type-III flood frequency curve method. The evaluation showed the GFMS detection performance improves [increasing probability of detection (POD)] with longer flood durations and larger affected areas. The impact of dams was detected in the validation statistics, with the presence of dams tending to result in more false alarms and greater false-alarm duration. The GFMS validation statistics for flood durations >3 days and for areas without dams vary across the four methods, but center around a POD of ~ 0.70 and a false-alarm rate (FAR) of ~ 0.65 . The generally positive results indicate the value of this approach for monitoring and researching floods on a global scale, but also indicate limitations and directions for improvement of such approaches. These directions include improving the rainfall estimates, utilizing higher resolution in the runoff-routing model, taking into account the presence of dams, and improving the method for flood identification.

1. Introduction

Floods are a leading natural disaster, common and costly, and responsible for about one-third of natural catastrophes (Smith and Ward 1998). Losses caused by

floods have been rising rapidly because of extreme weather conditions, urbanization, and inadequate disaster response. Hydrologic model-based flood forecasting systems have been regarded as the most effective way for flood early warning and monitoring and subsequent hazard mitigation and management (e.g., Dutta et al. 2000; Al-Sabhan et al. 2003; Hong et al. 2007; Reed et al. 2007; Yilmaz et al. 2010; among many others). However, almost all these existing flood forecasting systems are established at local or regional scales (e.g., Reed et al.

Corresponding author address: Huan Wu, Earth System Science Interdisciplinary Center, University of Maryland, College Park, 5825 University Court, Suite 4001 College Park, MD 20740-3823.
E-mail: huanwu@umd.edu

2007; Cloke and Pappenberger 2009; Pappenberger and Buizza 2009; Voisin et al. 2011), usually in developed regions, where sufficient resources are available, while many remote, ungauged regions and regions with transboundary basins remain without such systems. Ongoing improvements in global remote sensing data for estimating precipitation and delineating land surface characteristics (e.g., land cover, vegetation, topography, and hydrography) have augmented hydrological simulations on a wide range of scales, including the global scale. Developing global flood forecasting systems based on hydrologic models driven by remote sensing data at relatively high spatial and temporal resolution is now practical and has the potential for providing useful information for flood estimation and management, especially for underdeveloped or remote regions. However, challenges remain in accurate precipitation estimation, globally distributed parameterization for hydrological modeling, etc. But with improved accuracy, coverage, and resolution from satellite-based rainfall estimation (Adler et al. 2003), these products have been used in many hydrologic modeling applications with positive performance (e.g., Hong et al. 2006; Artan et al. 2007; Shrestha et al. 2008; Su et al. 2008; Pan et al. 2010; Su et al. 2011; among others). One such satellite rainfall product, the National Aeronautics and Space Administration (NASA) Tropical Rainfall Measuring Mission (TRMM) Multi-satellite Precipitation Analysis (TMPA; Huffman et al. 2007), has been used extensively and provides quasi-global (50°S–50°N) precipitation analyses at 3-hourly, 0.25° latitude–longitude resolution, with all satellite estimates calibrated or adjusted to the information from the TRMM satellite itself, which carries both a radar and passive microwave sensor.

Using the real-time version of the TMPA rainfall information, an experimental global flood monitoring system (GFMS) was developed (Hong et al. 2007) and has been running routinely for the last few years with results being displayed at the NASA TRMM website (<http://trmm.gsfc.nasa.gov/>). In this original GFMS, a simplified hydrologic infiltration module using a curve number (CN) approach and an antecedent precipitation index method as soil moisture proxy is used to partition rainfall and a linear slope–based flow speed and direction scheme is used to route runoff in order to predict potential floods over the quasi-globe in near real time. This original GFMS was evaluated in terms of detecting flood events by Yilmaz et al. (2010), which showed that the simplified CN-based hydrologic approach has some skill in detecting floods, especially during the early stages of flood events, but has low performance in flood event detection metrics (e.g., probability of detection) and delineation (e.g., flood evolution in the river network). Both

studies concluded that a relatively more physically based hydrologic model may improve the GFMS performance (Hong et al. 2007; Yilmaz et al. 2010). The Coupled Routing and Excess Storage (CREST) hydrologic model, later developed (Wang et al. 2011) for this purpose, is the subject of the evaluation in this paper.

The purpose of this paper is to evaluate the performance of the new version of the GFMS in flood detection against available flood event archives to indicate the skill and limitations of the system. This paper is organized as follows. In section 2, we describe the method used in this study. The results of the evaluation are described and discussed in section 3, including the GFMS flood detection performance at various scales and the impacts of dams on the results. Conclusions are in section 4 and future work is discussed in section 5.

2. Methodology

The GFMS combines the satellite-based estimates of precipitation, runoff generation, runoff routing, and flood identification. A unified algorithm for flood event identification and matching between modeled results and reported floods was developed. Four different flood threshold definition methods using GFMS output were developed and utilized to evaluate the sensitivity of the results to this variation.

a. Hydrologic model and data

The new (current) version of the GFMS uses the CREST model (Wang et al. 2011) to simulate the spatial and temporal variation of land surface and subsurface water fluxes and storages by cell-to-cell simulation, considering canopy interception, infiltration, and evapotranspiration processes. However, there are no cool season processes (e.g., snow or frost) considered in the model at this time. The CREST model calculates infiltration and surface runoff using the variable infiltration capacity curve similar to the Xinanjiang model (Zhao and Liu 1995) and the Variable Infiltration Capacity (VIC) model (Liang et al. 1994, 1996). It employs a vertical, parallel, multilinear reservoir module adapted from Xinanjiang model (Zhao and Liu 1995) coupled with a simplified cell-to-cell routing scheme with high computing efficiency. The CREST model main inputs include rainfall (e.g., TMPA), potential evapotranspiration (Famine Early Warning Systems Network; <http://igskmncnwb015.cr.usgs.gov/global/>) and hydrography. The hydrography data include 1/8° resolution flow direction and drainage area derived by the hierarchical dominant river tracing (DRT) algorithm by Wu et al. (2011) using 30 arc-second-resolution Hydrological Data and Maps Based on Shuttle Elevation Derivatives at

Multiple Scales (HydroSHEDS; Lehner et al. 2008) as baseline fine-resolution hydrography inputs. For this exercise and for the current real-time application we do not calibrate the CREST model because of the difficulty of doing so across the globe at $1/8^\circ$ resolution; more importantly, we assume the hydrologic model still has skill in ranking events at locations, even though the model-simulated flood magnitudes may be locally biased relative to observed data (Reed et al. 2007). All the model parameters were either directly estimated from input data, or used as a priori parameters (see detailed parameter estimation and description by Wang et al. 2011).

We performed the evaluation based on the retrospective simulation results of the CREST model forced by the TMPA version 6 (V6) research quality data from 1998 to 2010 at 3-hourly time resolution and $1/8^\circ$ latitude–longitude spatial resolution over the TRMM quasi-global domain. The TMPA (V6) rainfall, which is only available about a month after observation time, is used for this study because of its consistency during the 13-yr TRMM period used, as compared to the real-time version of the product (TMPA RT), which has changed significantly over that period. The current version of the real-time TMPA (RT) rainfall, used for the real-time GFMS, uses monthly and regional climatological adjustments to produce real-time estimates close to the after-the-fact V6, which includes monthly rain gauge information used for bias adjustments of the satellite rainfall estimates (Huffman et al. 2009). While the model outputs major hydrologic variables including discharge ($\text{m}^3 \text{s}^{-1}$), routed runoff (mm), evapotranspiration and soil water (mm), etc., the evaluation was performed mainly using the routed runoff variable (depth of water in each grid cell at each time), which represents the total amount of water stored in each grid cell surface at each time interval, routed from its upstream drainage area. The routed runoff variable was chosen for the evaluation because it directly represents the magnitude (depth) of water stored on the dry land surface for each grid cell regardless of flow conditions of inbank or overbank. The routed runoff and discharge can be calculated from one another at each grid cell. The simulated routed runoff was stored for each grid cell at every 3-h time interval for the simulated 13 years. These routed runoff results were then used to determine the flood definition statistics for each grid cell.

There are several global flood event databases available for comparison with the model. Most of them are online resources—for example, Emergency Events Database (EM-DAT) by the Centre for Research on the Epidemiology of Disasters (CRED) (<http://www.emdat.be/>), Global Identifier Number (GLIDE) disaster database

[Asian Disaster Reduction Center (ADRC); <http://www.glidnumber.net/>], Financial Tracking Service (FTS) global, real-time database [U.N. Office for Coordination of Humanitarian Affairs (OCHA); <http://fts.unocha.org/>], Dartmouth Flood Observatory (DFO) (<http://floodobservatory.colorado.edu/>), European Commission Joint Research Center (JRC) Global Disaster Alert and Coordination System (GDACS) (<http://www.gdacs.org/flooddetection/>), and the International Flood Network (IFNET) (<http://www.internationalfloodnetwork.org/>). However, although most databases record flood date, duration, and country, more detailed information such as geographical location (latitude and longitude) and river basin are often not recorded. This type of information is critically needed for the evaluation in this study. Since 2006, the DFO has begun to record geographical locations of flood events based on the center of a polygon enclosing the inundated area. A longer-period global flood inventory (GFI) based on DFO, EM-DAT, FTS, and IFNET was compiled for 11 years (1998–2008), coinciding with the availability of TRMM precipitation products (Adhikari et al. 2010). Geographical locations of flood events in GFI were mainly taken from DFO (for 2006 ~ 2008) with additional reports, aerial photographs, and remote sensing images used through tedious verification and cross-checking processes with Google Earth (Adhikari et al. 2010). We employed the DFO (2006 ~ 2010) and GFI (1998 ~ 2008) flood event database (referred to as the flood database) as the reference for the quantitative evaluation of the GFMS performance in flood detection, as they both provide both flood location and duration. Affected areas of flood events are also available from DFO flood database. There are 929 and 2672 reported flood events within the study domain (50°S – 50°N) by DFO and GFI, respectively, after removal of flood events caused by dam failure and snowmelt, which are not represented in the current GFMS formulation. A combined flood database was created using GFI (1998 ~ 2008) and DFO (2009 ~ 2010) for the evaluation.

b. Flood threshold definition

The Log Pearson type-III (LP3) distribution presented in Bulletin 17B (B17) (IACWD 1982) is the method currently recommended by United States federal agencies for flood frequency analysis. The LP3 distribution is recommended to fit the observed flood flow data using three sample moments (mean, standard deviation, and skew) calculated from the logarithmic transformed data. Magnitudes can be derived from the analytical LP3 distribution fit for floods with various return periods, and these magnitudes can be used as thresholds for flood definition. Although B17 tries to use LP3 to promote a consistent, uniform approach to flood frequency determination (Chow

TABLE 1. The definition of threshold values to define flood from the four methods. The unit for θ is mm and for FAC is km².

Methods	Thresholds	Methods	Thresholds
Method 1	Log Pearson type-III 2-yr	Method 3	$P_{95} + \sigma + \theta(\text{FAC}) P_{98}$
Method 2	return flood $P_{95} + 30$ (mm)	Method 4	
	$\theta = 6$ (FAC ≤ 3000); $\theta = 10$ ($3000 < \text{FAC} \leq 1.0 \times 10^6$); $\theta = 20$ (FAC $\geq 1.0 \times 10^6$)		

et al. 1988), it also indicates that flood events do not fit any one specific, known statistical distribution (IACWD 1982). With only historic streamflow, it is difficult to derive thresholds for identifying floods (Hirsch 1987), and no single probability distribution is the best to fit flood events under all situations in terms of variations in space and time. To define generalized reliable thresholds for flood identification for global-scale applications is even more challenging. However, in this study we are focused on the problem of flood detection but not flood intensity. Instead, in this evaluation of flood detection we are mainly focused on differentiating flood flow (either overbank or even more severe) from normal flow (below fullbank) given historic hydrologic data (from the simulations), regardless of the length of the return period or magnitude of an identified flood. This makes the flood definition problem in this study relatively easier. In addition to using the LP3 method, we also performed a series of experiments using statistic percentile-based method to determine alternate thresholds for flood identification.

1) LP3 DISTRIBUTION METHOD

The LP3 distribution has been widely used for hydrological data analysis in many applications. The third parameter of LP3 (skew) permits the fitting of asymmetric distribution. When the coefficient of skewness is zero, the LP3 becomes identical to the lognormal distribution. Flood magnitudes estimated by the LP3 distribution are very sensitive to the value of coefficient of skewness. Because the coefficient of skewness is very sensitive to the size of the sample and difficult to accurately estimate from small samples, B17 recommends a generalized estimator for coefficient of skewness by combining the station skew with a regional skew generalized from annual maximum streamflow using the inverse of their mean square errors as weights (IACWD 1982). As the generalized global map for skew is not available and difficult to use, we simply derived the LP3 for each grid cell over the globe from each grid cell's corresponding 13-yr annual maximum routed runoff (converted to discharge in units of m³ s⁻¹), following procedures by Chow et al. (1988) as in described in Eqs. (1)–(6):

$$X_T = \mu + K_T\sigma, \tag{1}$$

$$K_T = z + (z^2 - 1)k + \frac{1}{3}(z^3 - 6z)k^2 - (z^2 - 1)k^3 + zk^4 + \frac{1}{3}k^5, \tag{2}$$

$$k = C_s/6, \tag{3}$$

$$z = w - \frac{2.515517 + 0.802853w + 0.010328w^2}{1 + 1.432788w + 0.189269w^2 + 0.001308w^3}, \tag{4}$$

$$w = \left[\ln\left(\frac{1}{p^2}\right) \right]^{1/2}, \text{ and} \tag{5}$$

$$p = 1/T, \tag{6}$$

where X_T is the magnitude (logarithmic transformed) of a flood flow with return period of T years, μ is the mean and σ is the standard deviation and C_s is the skew calculated from the annual maximum discharges (converted from routed runoff and 10-based log transformed), K_T is a frequency factor approximated by Eq. (2), z is the standard normal variable approximated by Eq. (4), and p is the exceedance probability. The X_T related to the 2-yr return period, after being logarithmic back transformed and converted back to routed runoff in units of depth (mm), was selected to define the threshold to define flood in this study. On average, rivers are fullbank about every 2 years (Carpenter et al. 1999; Reed et al. 2007). Therefore, the magnitude of flood corresponding to a 2-yr return period was selected from the LP3 method as the threshold to define floods (Table 1). We used the 2-yr return period flood threshold estimated from the 13 years of data to define all floods. The LP3 method was also adopted by Reed et al. (2007) to estimate flood frequency using an 8-yr simulation for flash flood forecasting. However, we used the LP3 only as a binary indicator of flood occurrence tuned against the reported flood inventory, which should be reliable. Hereafter, the LP3 method is referred to as method 1.

2) PERCENTILE-BASED METHOD

Model-derived routed runoff absolute values are strongly determined by model assumptions and calibration, but relative values such as percentile statistics (probability of exceedance) can be used, especially for extreme

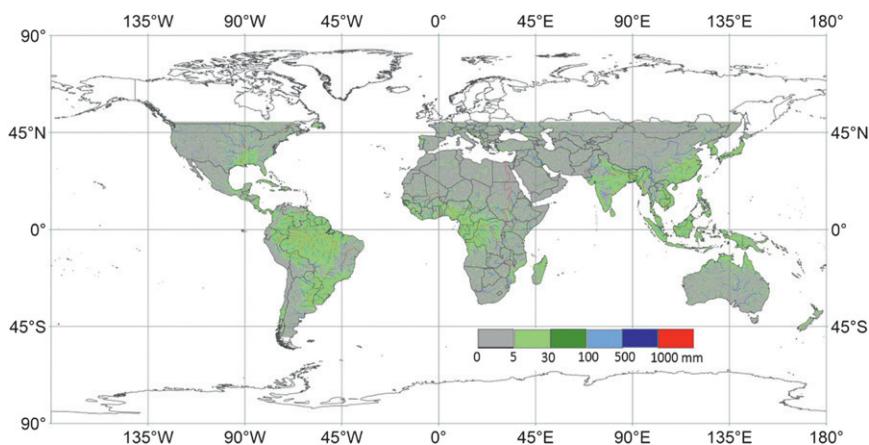


FIG. 1. Quasi-global 95th percentile routed runoff (mm) map derived from 13-yr retrospective simulation.

events, to effectively compare simulated and reported flood events. Brakenridge et al. (2007) developed a methodology for satellite-based flood detection by thresholding the Advanced Microwave Scanning Radiometer for Earth Observing System (AMSR-E) passive microwave signal of water surface change using 95th percentile values. In the evaluation of the initial GFMS by Yilmaz et al. (2010), five different geographical zones were defined considering hydroclimatic variations and the runoff threshold for each zone was defined as the 0.98 exceedance probability of 3-hourly runoff in each zone during the 1-yr study time period. As percentile value represents the relative rank of routed runoff for each grid cell, spatially distributed percentile values can be used to determine thresholds to differentiate flood flow from normal flow status. We assume the 95th percentile routed runoff (referred to as P_{95}) represents the streamflow within bank, while some higher percentile of routed runoff value represents the river in the fullbank status. We will determine a grid cell is flooding when its routed runoff is greater than a threshold value, which represents a river in full-bank status. Because of spatial heterogeneity of climate and landscape characteristics across the globe and their effects on hydrological response, using a uniform percentile (e.g., 95th and 98th) based threshold to define flood over globe may not be suitable. Instead of seeking a spatially distributed percentile map to define floods, we employed the 95th percentile routed runoff value of each grid cell as the starting point to define thresholds for flood identification—that is, we use the 95th percentile routed runoff value plus an additional threshold value to approximate fullbank routed runoff value for each grid cell (Table 1). The 95th percentile routed runoff values derived for each grid cell over the globe (Fig. 1) show a distributed spatial pattern generally consistent with the natural river network, with increasing

value along river flow path. Remember the routed runoff variable is the depth of water from dry ground (river bottom) at the $\frac{1}{8}^\circ$ scale.

In method 2, we used the P_{95} plus a constant value (i.e., 30 mm for this study) to represent fullbank status over the globe (Table 1). The 30-mm value was chosen based on experiments as to what value subjectively gave a reasonable number of flood events as compared to the DFO flood database (2006–10). As seen in Fig. 1, headwater and overland areas of basins in the map of P_{95} are separated from more downstream portions of rivers. However, the P_{95} , even with the constant 30 mm added as in method 2, cannot effectively separate rivers with high interannual or seasonal variations from rivers with low variations. To account for these variations in method 3, additional parameters are added. These additional parameters are used to take into account river hydrograph variations and are needed to increase the P_{95} to represent the fullbank status. Generally, a smaller (larger) range of difference between the P_{95} and routed runoff value at which river is fullbank is expected for rivers with less (more) interannual or seasonal variations. Standard deviation (σ) of the routed runoff over the 13 years represents the variation or dispersion from the mean and can be used to measure the interannual and seasonal variability of streamflow. Larger rivers (in terms of magnitude of streamflow) tend to require relatively larger absolute additional routed runoff threshold value above the P_{95} to reach the fullbank status than smaller rivers. Similar to P_{95} , σ usually increases as the mean increases with a distributed spatial pattern generally consistent to the natural river network with increasing value along river flow path (not shown). Therefore, although σ has a very high correlation coefficient to P_{95} from the global statistics, we used σ locally at each grid cell to form the

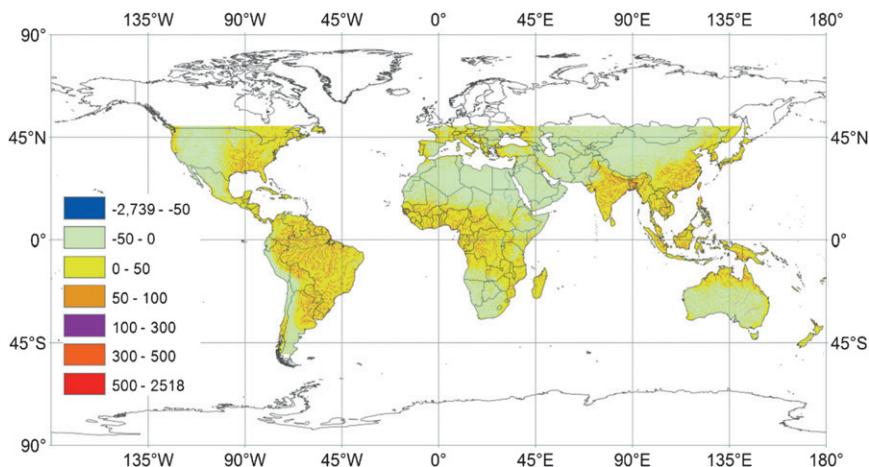


FIG. 2. The difference of thresholds derived by methods 3 and 1 (method 3 – method 1).

additional threshold in method 3. Because of σ being too small for some rivers with low streamflow (e.g., rivers in up basins or arid areas) or low seasonality (i.e., with a flatter monthly hydrograph), an upstream flow accumulation area (or upstream basin area, referred to as FAC, km^2) dependent additional threshold (θ) was also added in method 3. Because it is very difficult to derive an appropriate analytic relation between the additional thresholds and the FACs, arbitrary values for three FAC bands (Table 1) were adopted to define the θ in method 3 by which subjectively a reasonable number of flood events were defined as compared to the DFO flood database (2006–10). Thresholds directly using the 98th percentile routed runoff value as used by Yilmaz et al. (2010) were also investigated in this study and this higher percentile method is referred to as method 4 (Table 1).

The four methods were employed to define flood occurrence from the simulated results for each grid cell. A grid cell is determined as flooding at a time interval when the routed runoff for this time is greater than the threshold at the grid cell defined by method in question. The thresholds derived by each method (Table 1) are spatially distributed, with method 1 having the highest spatial variability while methods 2 and 4 have the lowest. The differences between thresholds derived by the methods are mainly reflected in up–low basin areas and wet–dry areas. There are large spatial variations in various thresholds, while there is no method that consistently produces the largest or least threshold across the study domain. Figure 2 shows differences between the thresholds defined by methods 1 and 3 and indicates that method 3 generally has higher thresholds for wet areas and lower ones for dry areas. In most of the study domain, the difference in thresholds between methods 1 and 3 ranges between -50 to $+50$ mm (green and yellow in Fig. 2). However, in stem rivers of the Amazon

basin and the Nile basin, method 1 derives larger thresholds than method 3 and the difference tends to be larger toward the river mouth. In many downstream areas of basins, the threshold values from methods 1 and 3 are much larger than those of method 2. There are differences among the finally determined threshold values of the method because they deviate from the exact “full-bank” or 2-yr return period when tuned (e.g., the additional threshold in percentile-based methods) against the flood database to obtain better detection performance of the system. However, instead of adjusting a single method, we explored the four methods to see the sensitivity of the flood detection results to the differences among the thresholding methods. However, the evaluation of these methods is not the primary focus of this paper.

c. Flood matching between simulated and archived databases

Although estimated flood events can be calculated for each grid cell from the retrospective simulated results according to the method in section 2b and there are locations (latitude and longitude) reported in both flood event databases, matching the flood events between these simulated and reported events based on a single grid cell is not appropriate. Both flood databases consist mainly of news reports and the assigned locations and days of the reported floods are not always accurate (Yilmaz et al. 2010). To make the evaluation more meaningful, we further developed the flood event identification method by Yilmaz et al. (2010), who used a $2.25^\circ \times 2.25^\circ$ moving spatial window based on the reported flood location and a 1-day (± 24 h) buffer surrounding the reported flood duration for matching the simulated and reported flood events. For this study, a spatial window (yellow area in Fig. 3) was defined for matching a simulated flood to

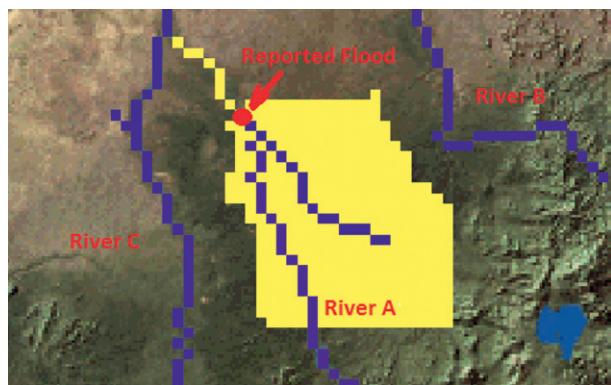


FIG. 3. Definition of spatial window for matching between simulated and reported flood events.

a reported flood according to the reported flood location and drainage network. The spatial window was defined 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 (red dot in Fig. 3). We also extended the spatial window definition by including the grid cells in the downstream stem river of the basin/subbasin below the reported location within a limited distance (i.e., ~ 100 km). In some cases the reported locations of floods are not located in rivers (i.e., with $FAC < 2$), and for these cases we moved the reported location downstream along the flow path a distance of two grid cells (~ 30 km) within the river basin. On the simulation side, we mark the entire area defined by the spatial window described above as simulated flooding when there are more than three grid cells flooding (according to the method in section 2b) within the spatial window for two continuous 3-h time intervals. The advantage of the spatial window definition is that the flood matching can be constrained in the same basin—that is, the simulated (reported) floods in neighboring basins and subbasins will not be incorrectly matched to the flood event reported (simulated) in the interested basin. We assume the reported flood locations are located in the correct basin, even though they may not be recorded with precisely correct latitude and longitude coordinates. If a flood is reported at a location in a stem river just downstream of a confluence while the flood actually occurred in the subbasin just upstream of the confluence, the flood identification algorithm we developed will check the upstream drainage area within a distance according to the reported location that contains the subbasin where the flood actually happened. In the other situation, if a flood is reported at a location within a subbasin just upstream from a confluence, while the flood actually occurred in the stem river where the confluence is, the flood identification algorithm will not miss the

match because it also checks the river segments downstream to the reported location for an extended 100 km. Therefore, the algorithm will check the stem river where the flood actually happened.

3. Results and discussion

Using the four flood definition methods, simulated floods for each 3-h time interval were derived globally and compared to the flood inventory data. Subjective evaluation of the results indicates that the model results often capture flood occurrence and general flood evolution reasonably well, responding to rainfall events with the start, development, and recession of flooding along the drainage networks (<http://trmm.gsfc.nasa.gov/>). Statistical results of the evaluation are presented in the following sections. To quantitatively evaluate the GFMS performance in flood event detection, we calculated three classic categorical verification metrics—that is, probability of detection [POD; $a/(a + c)$], false-alarm ratio [FAR; $b/(a + b)$], and critical success index [CSI; $a/(a + b + c)$], 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).

a. Model flood detection performance

An algorithm was developed to search the flood events in the simulated results according to the thresholds and method discussed in sections 2b and 2c to attempt to match with reported flood events in the flood databases. We determine that a reported flood event is hit by the GFMS if a reported flood event can be found in the simulated results within the spatial-temporal window associated with the reported flood event. The global PODs were calculated using the four flood definition methods for the two global flood databases separately and combined (Table 2). The calculation of POD by each method used the same flood event matching rules except for the four different threshold values. Results in Table 2 indicate that the POD values are basically independent of which reported flood inventory is used and therefore the two databases can be combined for the overall evaluation. Figure 4 shows the global flood events detected by the GFMS using method 3 (with a POD of 0.59) during 1998–2010 against the combined flood database, which indicates a reasonable geographic distribution and overlap of simulated and reported floods.

The reported and estimated number of floods decreases as a function of flood duration (Fig. 5a) and the POD of the GFMS increases with longer duration floods (Fig. 5b), with a gradual increase to an asymptote at 10–20 days, depending on the method used in identifying floods in the simulations. In the combined flood database,

TABLE 2. The POD performance by the four methods based on global statistics.

Flood database						
Name	Time	No.	M1	M2	M3	M4
GFI	1998–2008	2672	0.55	0.42	0.57	0.90
DFO	2006–2010	929	0.60	0.44	0.62	0.96
Combined	1998–2010	2949	0.56	0.42	0.59	0.90
Combined	Duration ≤ 3 days		0.38	0.26	0.42	0.79
Combined	Duration > 3 days		0.77	0.67	0.78	0.95

there are 1032 (35% of total 2949) short-term floods with flood duration ≤ 3 days. The GFMS has difficulty detecting these short-term floods with PODs of 0.38, 0.26, 0.42, and 0.79 (Table 2) for methods 1, 2, 3, and 4, respectively. However, the PODs increase to 0.77, 0.67, 0.78, and 0.95 (Table 2) for all floods with reported duration > 3 days. The POD model performance for flood detection also steadily increases as the flood-affected area increases (Fig. 6). These relations are almost certainly related to the limitations in the satellite rainfall data. The TMPA has a 3-h time resolution and 0.25° spatial resolution and these resolutions will certainly limit the definition of small-scale rain events. However, random sampling errors will decrease with spatial and temporal averaging and this tends to translate into better hydrologic model and flood calculations for larger (and longer) events. Larger floods also have higher possibility of meeting the flood definition and the matching rules discussed in sections 2b and 2c, with longer durations (larger temporal window) and more affected grid cells (more potential flooding individual grid cells). Larger floods are relatively easier to detect, while thresholds defined by the different methods may have difficulty

detecting smaller floods. Both Fig. 5 and Fig. 6 showed relatively larger differences of the POD performance between the methods (especially method 2 and other methods) for short-term floods or smaller affected areas, while the difference steadily decreases as the flood scale increases with longer duration or larger affected area. There are 35% (934 out of total 2672) of flood events in the GFI flood database that are reported as short-term (duration ≤ 3 days) floods, compared to 25% (231 out of total 929) in the DFO flood databases. This leads to consistently higher PODs for DFO as compared to the GFI database in Table 2.

There is a relatively sharp peak in the number (i.e., 156) of reported floods with duration of 15 days (Fig. 5a) in the combined flood database. However, this 15-day flood event peak only appears in the GFI flood database. As discussed in section 2c, to calculate the POD, the simulated floods are searched to match the record only when there is a flood is recorded in the flood event database. Therefore, the number of simulated floods for 15-day floods is found to increase for each method in Fig. 5a. However, the POD performance decreases for all the four methods for that specific duration of flood. The reason for the peak in flood events of this duration is not known, but may be related to human estimates tending to peak at 2 weeks (14 days) or one-half of a month. Similarly, the DFO flood database showed relatively more floods reported in some flood-affected-area ranges—for example, 131 out of total 929 flood events with affected area of $200\,000 \sim 300\,000 \text{ km}^2$ (Fig. 6a). The reason for this is also unknown, but may be related to the analysts preferentially picking a certain size of event.

Using the combined 13-yr flood database, a relatively large range of POD values from 0.42 to 0.90 is noted

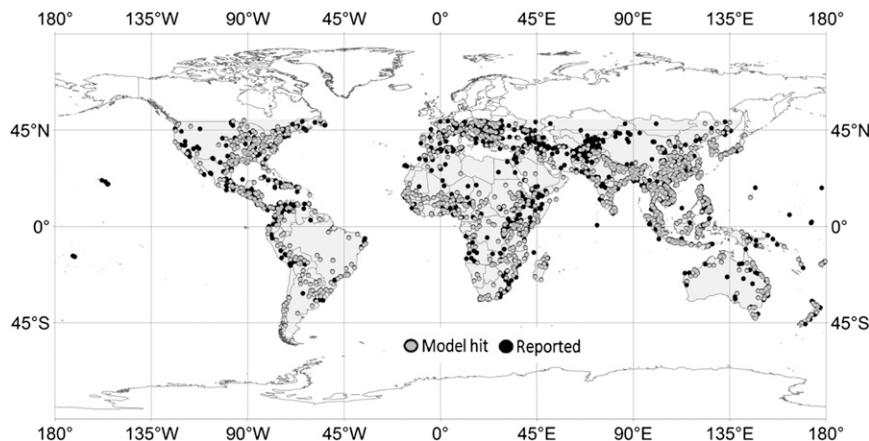


FIG. 4. Global flood events detected by the GFMS using method 3 during 1998–2010 against combined flood database. The dark balls are reported flood events in the database. When the model successfully hits a reported flood event, the dark ball turns to gray. The gray shaded part of the map is the TRMM-based study domain.

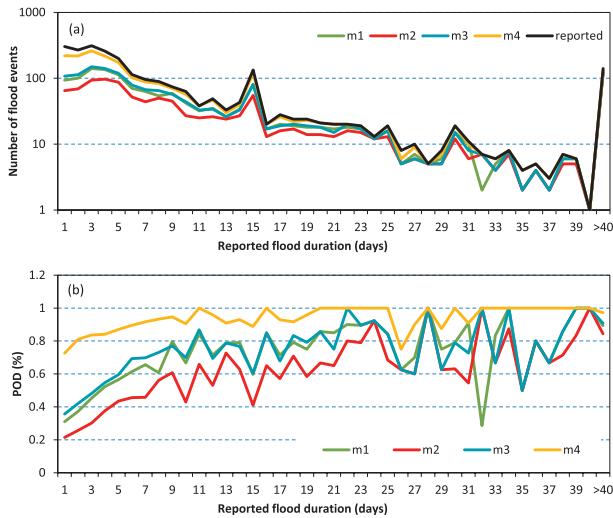


FIG. 5. The GFMS performance of flood detection in terms of flood duration against the combined flood database using the four flood definition methods.

(Table 2), with the values increasing and the range of values narrowing somewhat when only floods of greater than 3 days are included. However, a complete evaluation must include other statistics (e.g., FAR).

b. Model false-alarm performance

The false alarms in the predictions are of equal importance as the successful model hits, as they determine the flood forecast reliability and efficiency. A higher POD performance can be achieved by using lower thresholds or larger temporal and spatial windows to match the simulated and reported flood events. However, for a specific flood definition method, high POD usually comes with a larger number of false alarms.

Although the same four methods were used to define flood thresholds, the algorithm used to evaluate false alarms is different from the one for POD calculation, because FAR cannot be calculated straightforwardly like the calculation of POD by directly searching for flood events in the three-dimensional (latitude, longitude, and time) simulated results according to each reported flood event. To derive the FAR, flood events had to be identified first in the simulated results, and then those identified simulated flood events were used to compare with reported flood events. Many floods not only occur in local subbasins but also move downstream along river networks, creating a larger affected area within the entire river basin, while the location of a reported flood is a specific point with only a latitude and longitude coordinate available in the flood databases. The reported location for a flood event may not be precise; for example, the reported location could be adjacent to the actual

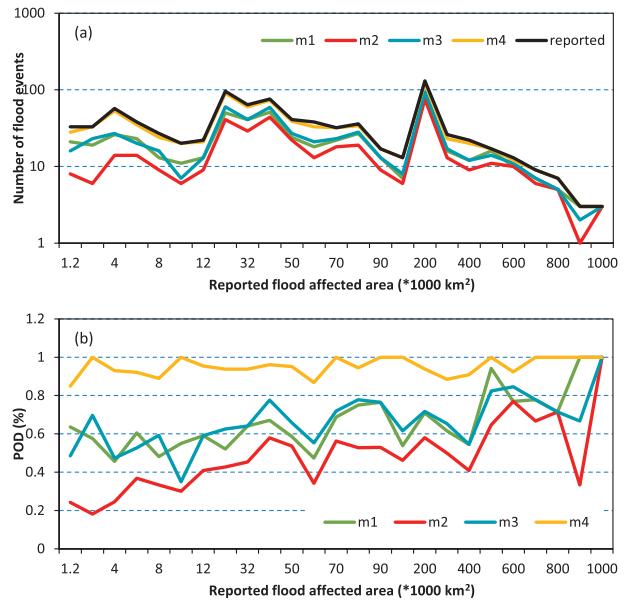


FIG. 6. The GFMS performance of flood detection in terms of affected area against DFO flood database using the four flood definition methods.

place where the flood actually happened, according to the news report. Given a specific subbasin or local river reach, the reported flood events also have errors in assigned times. Furthermore, floods are likely underreported in both the GFI and DFO archives, because floods tend to be reported in high-population areas while underreported in remote areas—for example, the Amazon basin (Fig. 7e). In addition, larger floods causing more damage tend to be reported, while smaller floods tend to be missed. Therefore, in order to evaluate the model performance in FAR as objectively as possible, we calculated the FAR by comparing the simulated flood events to reported floods in 53 selected well-reported areas (WRA) over the globe (yellow areas in Fig. 7). The WRA are defined according to the 13-yr combined flood database by the following procedure: 1) the same method for definition of the spatial window (section 2c) was applied to each reported flood location in the combined flood database; 2) if there are multiple reported flood events in the spatial window, the reported location with the largest upstream drainage area was selected and used to define a new spatial window; and 3) if there are more than six flood events reported during the 13 years in the new spatial window, we determine the new spatial window as a WRA. All the WRAs are located in wet and/or high-population regions (Fig. 7). A large proportion of the well-reported areas are located in South Asia (Fig. 7c) and Africa (Fig. 7b). The numbers of well-reported areas for each continent are 24 (Asia), 16 (Africa), 6 (North America), 4 (Europe), and 3 (South

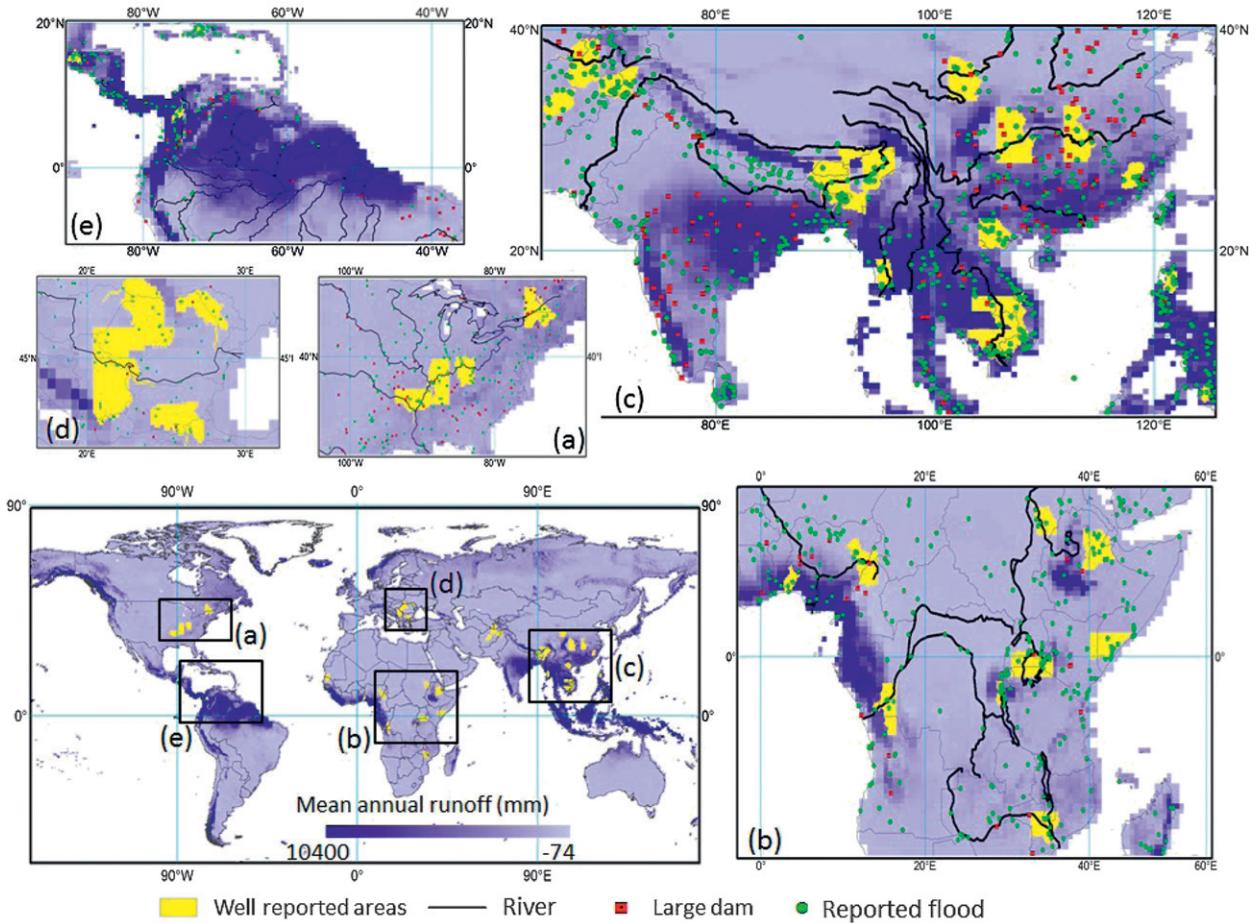


FIG. 7. The spatial distribution of the 53 well-reported areas (according to the combined flood database) over the TRMM global domain, with 5 regions selected to zoom in. The background image is the mean annual runoff (precipitation minus evapotranspiration) from NASA’s Modern-Era Retrospective Analysis for Research and Applications (MERRA) reanalysis data for the satellite era (Bosilovich et al. 2006).

America). There are a total of 490 flood events reported for the 53 WRAs from 1998 to 2010. The number of reported flood events ranges from 6 to 25 with a mean value of 9.

To calculate the FAR, the simulated floods were identified by checking each grid cell in a selected WRA for every modeling time step. A simulated flood event is identified for each time step for which there are at least three grid cells flooding at each time step (same as the POD calculation in section 3a). Then, if the flood duration of a simulated flood event overlaps with the duration of a reported flood in the same selected WRA, we determine the reported flood is successfully detected by the GFMS. All simulated floods having no overlap in time and space with any reported floods are regarded as false alarms. By this method the number of hits, misses, and false alarms and simulated flood durations were derived for each WRA. When a reported flood had a

long duration and it was hit by the model-based result multiple times, each match is recorded. However, two neighboring (in time) simulated events were considered independent events only when they were 2 days apart. When a simulated flood event had a long time period and overlapped with more than one reported flood, it was simply divided into events by 15-day periods. However, this type of case did not happen in this evaluation.

All the three verification metrics by all methods vary from one WRA to another because of the small number of cases (6–25) in each area (Figs. 8 and 9). The mean PODs over the 53 WRAs are 64%, 54%, 70%, and 89% by methods 1, 2, 3, and 4, respectively, based on the combined flood database (Table 3). The PODs by methods using absolute magnitude thresholds (i.e., methods 1, 2, and 3) from the WRAs are higher than those for the whole globe (Table 2), because the WRAs are mostly located in wet areas, where the flood identification is relatively

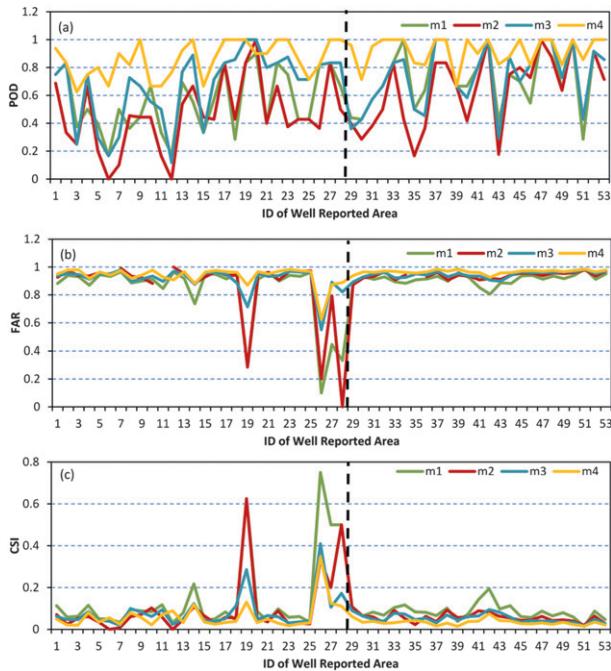


FIG. 8. The GFMS flood detection performance against the combined flood database for floods with all durations (≥ 1 day) over the 53 well-reported areas. The WRAs with identification from 1 to 28 (left to the vertical dash line) are with no dams and the WRAs with identification >28 (right to the vertical dash line) are with dams.

easier than in dry regions. In arid or up-basin areas where the routed runoff is smaller, the additional threshold (θ) in methods 2 and 3 tends to reduce the identification of floods in these areas, while method 4, which uses a relative rank, was not affected significantly.

There are a number of factors that can lead to false alarms in the model results, including errors in the precipitation estimation, impacts of flow control structures (e.g., dams and levees), missing reports, limits in the flood definition methods, and errors in the hydrologic model.

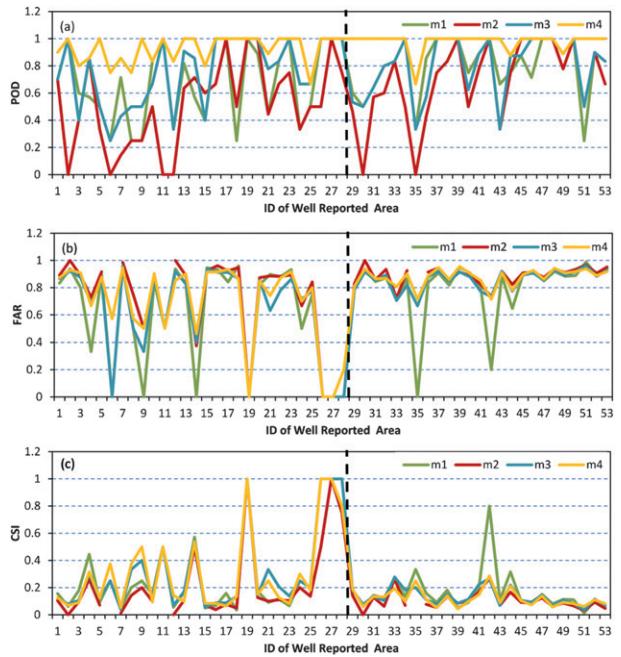


FIG. 9. As in Fig. 8, but for longer-term floods (duration > 3 days).

As all these factors are probably contributing, the FAR statistics appear poor at first glance. The mean FARs over the 53 WRAs for all floods with duration ≥ 1 day are 87%, 89%, 93%, and 95% by methods 1, 2, 3, and 4, respectively (Table 3), with only a few areas showing lower FAR values (Fig. 8b). However, 35% of floods in the reported flood databases are short-term floods (172/490 floods with duration ≤ 3 days over the 53 WRAs). When these short-term floods are removed from the analysis, the GFMS has significantly better performance, with lower FARs, and also higher PODs (Fig. 9 and Table 3).

Three of the techniques have similar CSIs of 22%–23% for floods greater than 3-day durations. Method 4 (the 98th percentile method) has a very high POD (95%) for the longer floods, with a FAR of 78%. The greatly

TABLE 3. Flood detection verification against the combined flood database over the 53 well-reported areas by the four methods.

Metrics	M1	M2	M3	M4
Metrics averaged over total 53 WRAs for all floods with duration ≥ 1 day				
POD	0.64	0.54	0.70	0.89
FAR	0.87	0.89	0.93	0.95
CSI	0.12	0.08	0.07	0.05
Metrics averaged over the 53 WRAs for floods with duration ≤ 3 days				
POD	0.42	0.35	0.52	0.77
FAR	0.93	0.95	0.97	0.98
CSI	0.06	0.04	0.03	0.02
Metrics averaged over the 53 WRAs for floods with duration > 3 days				
POD	0.74	0.60	0.78	0.95
FAR	0.70	0.80	0.74	0.78
CSI	0.23	0.16	0.23	0.22

TABLE 4. Flood detection verification against the combined flood database over the 28 WRAs without dam and the 25 WRAs with dams by the four methods.

Metrics	M1	M2	M3	M4
Metrics averaged over the 28 WRAs with dam for floods with duration ≤ 3 days				
POD	0.51	0.40	0.54	0.79
FAR	0.96	0.97	0.97	0.98
CSI	0.04	0.02	0.02	0.01
Metrics averaged over the 28 WRAs without dam for floods with duration ≤ 3 days				
POD	0.35	0.31	0.50	0.75
FAR	0.91	0.93	0.96	0.97
CSI	0.07	0.05	0.04	0.03
Metrics averaged over the 25 WRAs with dam for floods with duration > 3 days				
POD	0.80	0.69	0.83	0.98
FAR	0.81	0.89	0.85	0.88
CSI	0.16	0.10	0.14	0.12
Metrics averaged over the 25 WRAs without dam for floods with duration > 3 days				
POD	0.68	0.51	0.74	0.93
FAR	0.60	0.71	0.63	0.68
CSI	0.30	0.23	0.32	0.31

increased POD and decreased FAR values for longer-term flood detection indicates the GFMS is more reliable for larger-scale floods, which is not surprising considering the resolutions of the precipitation data and the hydrologic model. Uncertainties in the data (especially the rainfall) and the model may produce noisiness in the model flood identification, leading to a large number of small-scale false alarms. For example, among the 5759 simulated floods identified for the 53 WRAs by method 3, 73% are short-term floods (< 3 days), so that many of the false alarms are associated with small-scale events. Part of the reason for the high false alarms for short-term floods may be related to a greater likelihood for missed reports for smaller events.

c. Impact of dams on false-alarm validation statistics

In Fig. 9, which is for floods lasting greater than 3 days, the distribution of FAR values (Fig. 9b) is very different than the distribution for all floods in Fig. 8b. This variation in FAR among the WRAs is related to the presence of dams, and becomes clearest when the short-term floods are ignored in the analysis. A global large dam database (Vörösmarty et al. 1997, 2003; <http://wwdrii.sr.unh.edu/download.html>) was employed to investigate the dam effects on the false alarm over the 53 WRAs. To investigate the effects of dams on the false-alarm statistics, we divided the 53 WRAs into two groups. The first group consists of the 28 WRAs with no dams (left of the vertical dashed lines in Figs. 8 and 9), referred to as no-dam group. The second group consists of the 25 WRAs with dams (right of the vertical dashed lines in Figs. 8 and 9), referred to as the dam group. One can see immediately in Fig. 9b that the lower FAR values tend to be associated with the WRAs in the no-dam group, whereas the areas

with dams tend to have high FAR values, indicating a clear relation between the presence of large dams and the false-alarm statistics. The FAC of the two groups vary over a similar range of values, which also indicates that the comparison is valid. Table 4 shows the flood detection verification metrics based on short-term and long-term floods derived for the no-dam group and dam group separately by averaging over the WRAs in the two groups. For the short-term floods (top half of table) there is only a slight difference between the dam and no-dam groups. For longer-term floods (> 3 days; bottom half of Table 4) the FAR values are much lower on average for the no-dam group, although the POD values are somewhat lower also. The resulting CSI values are generally higher for the no-dam areas. The higher PODs in the dam group may reflect that larger floods are relatively easier to detect for the GFMS and the reported floods in the dam group may be relatively larger (because of the presence of dams).

The results in Table 4 indicate that the GFMS has better performances in areas without dams, which is as expected since the hydrologic model does not include a reservoir module to represent dam operations. The GFMS shows very good performance in detecting floods with duration > 3 days in nondam situations, with relatively high POD and low FAR leading to relatively higher CSI (Table 4). This result indicates that dam effects on the GFMS flood detection ability highly depend on the flood scale—that is, dams prevent many small simulated floods from actually occurring and being reported, leading to lower GFMS performance metrics, while the statistics are better for longer duration events, even for WRAs with dams, because the larger rainfall events can still produce actual floods, even in areas with dams, though the dams

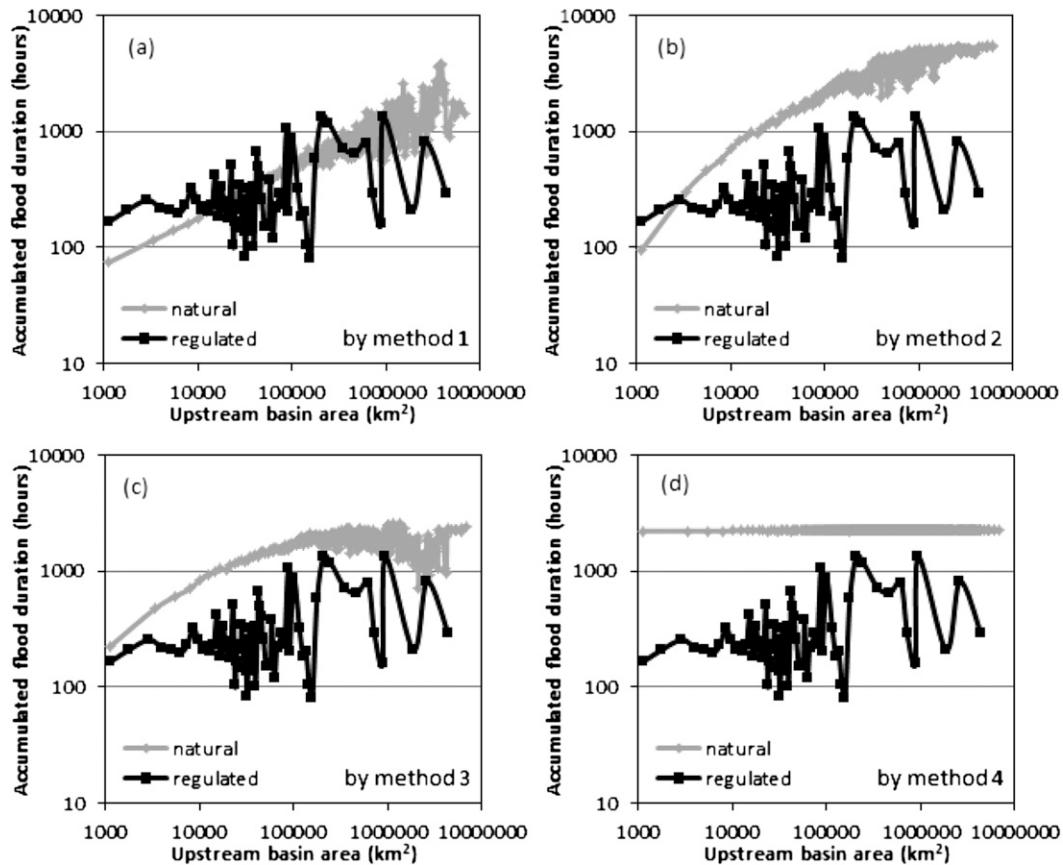


FIG. 10. The accumulated flood duration changes with upstream basin area in natural (by model) and regulated (reported) scenarios.

would likely decrease flood peaks and damages. The GFMS, using relatively coarse-resolution rainfall information and hydrologic modeling, and without any method to take into account the effect of dams, should be expected to have reasonable statistical results for events of at least a few days' duration in areas not affected by large dams. Results summarized in Table 4 (fourth panel) indicate that this is the case. For flood duration greater than 3 days in areas without large dams the POD is ~ 0.7 , the FAR is ~ 0.6 , and the CSI is ~ 0.3 . These are good results for this stage of GFMS development.

d. Flood duration statistics

Accumulated flood duration (AFD) during the 13 TRMM-era years was calculated for each grid cell from the simulated results for each flood definition method. The simulated (natural—no dams) and reported (regulated—including basins with dams) AFD histograms based on FAC were derived respectively from the simulated results and the 2949 reported flood events in the combined flood database. The AFD in each histogram column was calculated as the average of the AFDs

from all grid cells with their FAC values falling into the FAC band indicated. Natural floods progress from upstream to downstream along a drainage network and thus increase the flood duration in lower parts of river basins. By methods 1, 2, and 3, the simulated AFD generally increases downstream along the drainage network with a similar spatial pattern to FAC and basin drainage network (gray in Figs. 10a–c). This indicates that the GFMS generally maintains the natural spatial pattern of AFD with the hydrologic model in which only natural processes are considered. Of course, the current routing scheme does not consider the presence of dams. However, this type of AFD curve was not generated by method 4, which uses the 98th percentile uniformly for each grid cell and derives a spatially uniform AFD (gray in Fig. 10d). The uniform 98th percentile applied to each grid cell determines the same number of time intervals (2% of the time) flooding for each grid cell, resulting in the uniform AFD. Independently defining floods for each single grid cell solely using a uniform percentile threshold value cannot take into account the fact that floods in upstream rivers add more flood risk in downstream areas

as floods propagate along the drainage network. However, percentile plus an additional threshold (e.g., 30 mm by method 2) defines floods not only in a relative manner, but also by an absolute threshold. In this way floods are defined only when the runoff is accumulated with a large enough magnitude, which significantly reduces the number of flood identifications in upstream basins and in relatively drier areas leading to, on average, larger AFD values in areas with larger FAC magnitudes.

From Fig. 10, method 1 derives relatively smaller AFD in up basins and method 4 derives the most, probably mostly contributed by short-term floods. Simulated AFD by method 2 increases relatively more steadily toward downstream than other methods while it is higher than other methods that are relatively closer to the AFD magnitudes based on the reported floods in downstream areas. This is consistent to the comparisons between verification metrics (e.g., Table 3), which indicate that methods 1, 3, and 4 have closer and better performance for long-term floods (tending to occur in downstream basins) than method 2. The 30-mm additional threshold used in method 2 is too large for presenting bank-full status in up-basin areas and too small for many downstream basins (Fig. 10b), but it generally captures the spatial pattern of flood duration for natural scenarios. As floods are probably largely underreported, it is difficult to draw a strong conclusion on which method derives the closest AFD to reality. However, if reliable reported flood duration information is available, the relation between AFD and FAC (Fig. 10) can provide a useful reference to find more reasonable thresholds for flood definition.

The AFD is well related to FAC based on the global statistics from the simulated results by all the methods, except method 4. Unlike the simulated results (except by method 4), there is no strong increase in the reported AFD as FAC increases and the variability range in reported AFD is much larger than the simulated. This could be partly caused by the bias in the reported flood duration. However, the good relation between the AFD and FAC may exist only in a natural scenario. When dams stop floods and change the flood duration, the AFD spatial pattern could be changed. If the bias in the reported flood duration does not significantly change the relation between the AFD and FAC in reality, Fig. 10 may cast another hint of dam impacts on floods, leading to more false alarms. Dams and artificial structures decrease the possibility of flooding in their downstream areas, while they might also increase the possibility of flooding in upstream areas, thus flood duration in upstream (downstream) basins might also increase (decrease). However dam effects are difficult to quantify without a reservoir module in the hydrologic model.

e. Duration of false alarms

POD and FAR statistics represent how well the technique detects individual flood events, no matter what the durations of the actual and estimated floods. Another measure of the quality and usefulness of the model-based flood estimates in this study is the mean duration of the false alarms. Therefore, as a further evaluation of the overall model-based technique, the four methods were evaluated in terms of the lengths of the false alarms. The evaluation showed that method 1 has the longest average false-alarm flood duration (9.7 days) based on all the simulated floods with durations >3 days from all WRAs, while methods 3 and 4 derived the least (6.4 days). On average, based on floods with durations >3 days, the false flood duration for each WRA per year is 22.8, 19.9, 14.4, and 20.5 days, while the average number of false alarms for each WRA per year are 2.3, 2.6, 2.2, and 3.2 by methods 1, 2, 3, and 4, respectively. For short-term floods, all methods showed similar average flood duration of 1.5–1.7 days per event. Although there needs to be additional analysis in this area, the different flood identification thresholding approaches used in the four methods all give reasonable results. However, the type of approach used for method 3, taking into account basin size through the FAC parameter and the seasonal and interannual variability through the use of the flood variance parameter (σ) seems to provide the best framework for future work in this area.

4. Conclusions

This paper describes an evaluation of a new version of a global flood monitoring system (GFMS) using an improved hydrologic model (the CREST model) driven by Tropical Rainfall Measuring Mission (TRMM) Multi-satellite Precipitation Analysis (TMPA) rainfall. The new GFMS was quantitatively evaluated on flood event detection during the TRMM era (1998–2010) based on a global retrospective simulation (3-hourly and $1/8^\circ$ spatial resolution) using the satellite rainfall for that period. Four methods were explored to define flood threshold from the simulated results to compare against the flood events in reported archives, including three statistic percentile-based methods and a log Pearson type-III flood frequency curve-based flood definition. The GFMS performance was evaluated with regard to flood occurrence using three classic categorical verification metrics (POD, FAR, and CSI). Balanced POD and FAR results are necessary for this type of system to be useful in applications. Flood duration statistics as related to false-alarm rates were also examined in relation to the utility of the simulated results.

In this study, flood matching rules (e.g., the spatial-temporal window) remain the same for all the methods. Therefore the differences of the GFMS flood detection performances interpreted are caused by threshold values and the spatial distribution defined by these methods. The verification metrics vary across WRAs (Figs. 8 and 9) with all the flood detection methods showing roughly similar results. The evaluation of the GFMS in this study showed two key results independent of the specific flood identification method used. First, the statistics clearly showed that the results improve with flood duration. That is, both POD and FAR improve when the evaluation is confined to longer-term floods—in this case, >3 day durations. This result is reasonable considering the time resolution of the satellite rainfall data (~3 h), the spatial resolution of the hydrologic model ($1/8^\circ$), and the limitations of the flood inventory data used for comparison. The GFMS is therefore best utilized for floods of over a day or a few days' duration and should not be expected to consistently detect shorter-term floods (e.g., floods with duration <1 day).

Second, the impact of dams can be detected in the validation statistics, with areas without dams showing a much lower FAR, as one would expect. The hydrologic model used treats the water flow in a strictly natural mode, following the terrain, without taking into account man-made structures or water management. More dams tend to result in more false alarms and false-alarm duration. However, dam effects highly depend on flood scale with more negative effects on detection for short-term floods. Global comparison of accumulated flood duration between natural (by model) and regulated (reported) flood events also indicates dam and artificial structures play important roles leading to more false alarms and false-alarm duration. Therefore, the GFMS statistics for flood durations >3 days and for areas without dams give a good estimate of the overall status of the approach at this time. The statistics vary across the four identification methods, but center around a POD of ~0.7 and a FAR of ~0.6.

The evaluation of the current system, both subjective and quantitative, indicates an improvement over the earlier, simpler system evaluated by Yilmaz et al. (2010). However, although this evaluation of the earlier flood identification technique was done in a similar manner (with an overlap time period: April 2007–July 2008), full and direct comparison is difficult because of differences in techniques used, spatial resolution, and the shorter length of record used. But, in terms of POD the current GFMS seems to have somewhat higher values (0.9 for method 4 versus 0.38 for the same threshold technique used by Yilmaz). The other flood detection methods in this study have lower PODs (~0.7) that are still higher

than the earlier technique. However, FAR statistics were not calculated by Yilmaz, although he noted significant regions of numerous false alarms. Subjectively, the new GFMS seems to improve both the flood detection performance and the presentation of flood evolution (start, development, and recession) in the drainage network. This overall better flood detection performance in the current version of GFMS is probably due to both the hydrologic model and the flood identification algorithms. The precipitation input is identical, so that is not an issue in any difference. However, the key conclusion is that the current system performs in an understandable fashion and reasonably well against global flood event information. These important results allow us to proceed to further improvements and more detailed evaluation and validation. The new GFMS has replaced the old one and is operationally available at <http://oas.gsfc.nasa.gov/CREST/global>.

5. Future work

This model development and evaluation provides a pathway forward for continued improvement in the future. First, the improvements brought by the new hydrologic model encourage us to use more physically based hydrologic models to potentially achieve better flood forecasting capability and performance in future endeavor, though very likely with much higher computational cost. The NASA Land Information System (LIS; Kumar et al. 2006; Peters-Lidard et al. 2007) provides a series of state-of-art large-scale land surface processes models and therefore gives a good opportunity for efforts in this direction. Second, to realize the potential of global flood monitoring systems, simple and robust flow routing schemes that contain minimal calibration parameters wherever possible are needed (Yilmaz et al. 2010), in addition to the a priori parameters. Although the routing scheme in the CREST hydrologic model used in this study has advantages in computing efficiency, it requires additional efforts in model regional calibration. Improved routing techniques taking into account within cell routing will be implemented in the near future. Third, the evaluation on effects by dams on flood detection indicates the limitations of the current GFMS in flood detection without accounting for dams and levees. Without an explicit module for representing the function of flood control by reservoir operations in the hydrologic model, the effects of the spatial distribution of dams (in upstream stem river and/or tributaries) and large reservoirs on the false alarms remain unknown. Implementation of a reservoir module in the routing scheme should also have a high priority in future work. Furthermore, the continuation and improvement of multisatellite precipitation observations through

NASA's Global Precipitation Measurement (GPM) mission will provide the GFMS with more accurate precipitation analyses utilizing space–time interpolations and improvements for shallow, orographic rainfall systems, and snow. A more precise and detailed flood observation database is also very desirable for future evaluations. Once acceptable performance in POD and FAR is achieved, the GFMS can also be used to reconstruct historical flood events for climate variation studies. Thus, the next stage of the GFMS development will focus on precisely quantifying flood properties including flood timing, magnitude, stage, inundation depth, extent, etc.

Acknowledgments. This work was supported by NASA's Applied Sciences Program (Michael Goodman) and NASA's Precipitation Measurement Missions (PMM) Program (Ramesh Kakar).

REFERENCES

- Adhikari, P., Y. Hong, K. R. Douglas, D. B. Kirschbaum, J. J. Gourley, R. F. Adler, and G. R. Brakenridge, 2010: A digitized global flood inventory (1998–2008): Compilation and preliminary results. *Nat. Hazards*, **55**, 405–422, doi:10.1007/s11069-010-9537-2.
- Adler, R. F., and Coauthors, 2003: The Version-2 Global Precipitation Climatology Project (GPCP) Monthly Precipitation Analysis (1979–present). *J. Hydrometeorol.*, **4**, 1147–1167.
- Al-Sabhan, W., M. Mulligan, and G. A. Blackburn, 2003: A real-time hydrological model for flood prediction using GIS and the WWW. *Comput. Env. Urban Syst.*, **27**, 9–32.
- Artan, G., H. Gadain, J. Smith, K. Asante, C. J. Bandaragoda, and J. Verdin, 2007: Adequacy of satellite derived rainfall data for streamflow modeling. *Nat. Hazards*, **43**, 167–185, doi:10.1007/s11069-007-9121-6.
- Bosilovich, M., and Coauthors, 2006: NASA's Modern Era Retrospective-Analysis for Research and Applications (MERRA). *U.S. CLIVAR Variations*, Vol. 4, No. 2, U.S. CLIVAR Project Office, Washington, DC, 5–8.
- Brakenridge, G. R., S. V. Nghiem, E. Anderson, and R. Mic, 2007: Orbital microwave measurement of river discharge and ice status. *Water Resour. Res.*, **43**, W04405, doi:10.1029/2006WR005238.
- Carpenter, T. M., J. A. Spersflage, K. P. Georgakakos, T. Sweeney, and D. L. Fread, 1999: National threshold runoff estimation utilizing GIS in support of operational flash flood warning systems. *J. Hydrol.*, **224**, 21–44.
- Chow, V. T., D. R. Maidment, and L. W. Mays, 1988: *Applied Hydrology*. McGraw-Hill, 572 pp.
- Cloke, H. L., and F. Pappenberger, 2009: Ensemble flood forecasting: A review. *J. Hydrol.*, **375** (3–4), 613–626, doi:10.1016/j.jhydrol.2009.06.005.
- Dutta, D., S. Herath, and K. Musiak, 2000: Flood inundation simulation in a river basin using a physically based distributed hydrologic model. *Hydrol. Processes*, **14**, 497–519.
- Hirsch, R. M., 1987: Probability plotting position formulas for flood records with historical information. *J. Hydrol.*, **96**, 185–199.
- Hong, Y., K. Hsu, H. Moradkhani, and S. Sorooshian, 2006: Uncertainty quantification of satellite precipitation estimation and Monte Carlo assessment of the error propagation into hydrologic response. *Water Resour. Res.*, **42**, W08421, doi:10.1029/2005WR004398.
- , 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**, W08502, doi:10.1029/2006WR005739.
- Huffman, G. J., and Coauthors, 2007: The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *J. Hydrometeorol.*, **8**, 38–55.
- , R. F. Adler, D. T. Bolvin, and E. J. Nelkin, 2009: The TRMM Multi-Satellite Precipitation Analysis (TMPA). *Satellite Rainfall Applications for Surface Hydrology*, M. Gebremichael and F. Hossain, Eds., Springer Verlag, 3–22.
- IACWD, 1982: Guidelines for determining flood flow frequency. Interagency Advisory Committee on Water Data, Hydrology Subcommittee Bulletin 17-B (revised and corrected), 194 pp.
- Kumar, S. V., and Coauthors, 2006: Land information system: An interoperable framework for high resolution land surface modeling. *Environ. Modell. Software*, **21**, 1402–1415.
- Lehner, B., K. Verdin, and A. Jarvis, 2008: New global hydrography derived from spaceborne elevation data. *Eos, Trans. Amer. Geophys. Union*, **89**, 93–94.
- 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 general circulation models. *J. Geophys. Res.*, **99** (D7), 14 415–14 428.
- , 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.
- 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.
- Pappenberger, F., and R. Buizza, 2009: The skill of ECMWF precipitation and temperature predictions in the Danube basin as forcings of hydrological models. *Wea. Forecasting*, **24**, 749–766.
- Peters-Lidard, C. D., and Coauthors, 2007: High-performance Earth system modeling with NASA/GSFC's Land Information System. *Innovations Syst. Software Eng.*, **3**, 157–165.
- 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. Hydrol.*, **337** (3–4), 402–420, doi:10.1016/j.jhydrol.2007.02.015.
- 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. Flood Risk Manage.*, **1**, 89–99, doi:10.1111/j.1753-318X.2008.00011.x.
- Smith, K., and R. Ward, 1998: *Floods: Physical Processes and Human Impacts*. Wiley, 394 pp.
- Su, F. G., Y. Hong, and D. P. Lettenmaier, 2008: Evaluation of TRMM Multisatellite Precipitation Analysis (TMPA) and its utility in hydrologic prediction in La Plata basin. *J. Hydrometeorol.*, **9**, 622–640.
- , H. Gao, G. J. Huffman, and D. P. Lettenmaier, 2011: Potential utility of the real-time TMPA-RT precipitation estimates in streamflow prediction. *J. Hydrometeorol.*, **12**, 444–455.
- 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., 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**, 210–219.
- , 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 Planet. Change*, **39**, 169–190.
- Wang, J., and Coauthors, 2011: The Coupled Routing And Excess Storage (CREST) distributed hydrological model. *Hydrol. Sci. J.*, **56**, 84–98.
- 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.
- 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. *Computer Models of Watershed Hydrology*, V. P. Singh, Ed., Water Resources Publications, 215–232.