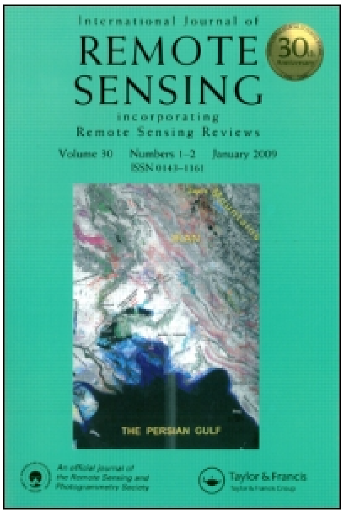

International Journal of Remote Sensing
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713722504

Evaluation of a satellite-based global flood monitoring system

Koray K. Yilmaz[ab]; Robert F. Adler[ab]; Yudong Tian[bc]; Yang Hong[d]; Harold F. Pierce[be]
[a] Earth System Science Interdisciplinary Center, University of Maryland, College Park, MD, USA [b] NASA Goddard Space Flight Center, Greenbelt, MD, USA [c] Goddard Earth Sciences and Technology Center, University of Maryland, Baltimore County, Baltimore, MD, USA [d] School of Civil Engineering & Environmental Sciences, University of Oklahoma, Norman, OK, USA [e] Science Systems and Applications, Inc., Lanham, MD, USA

Online publication date: 04 August 2010

To cite this Article Yilmaz, Koray K. , Adler, Robert F. , Tian, Yudong , Hong, Yang and Pierce, Harold F.(2010) 'Evaluation of a satellite-based global flood monitoring system', International Journal of Remote Sensing, 31: 14, 3763 — 3782
To link to this Article: DOI: 10.1080/01431161.2010.483489
URL: http://dx.doi.org/10.1080/01431161.2010.483489

International Journal of Remote Sensing
Vol. 31, No. 14, 20 July 2010, 3763–3782

Evaluation of a satellite-based global flood monitoring system

KORAY K. YILMAZ*†‡, ROBERT F. ADLER†‡, YUDONG TIAN‡§, YANG HONG¶ and HAROLD F. PIERCE‡|

†Earth System Science Interdisciplinary Center, University of Maryland, College Park, MD 20740, USA

‡NASA Goddard Space Flight Center, Code 613.1, Greenbelt, MD 20771, USA

§Goddard Earth Sciences and Technology Center, University of Maryland, Baltimore County, Baltimore, MD 21250, USA

¶School of Civil Engineering & Environmental Sciences, University of Oklahoma, Norman, OK 73019, USA

|Science Systems and Applications, Inc., Lanham, MD 20706, USA

This study provides an initial evaluation of a global flood monitoring system (GFMS) using satellite-based precipitation and readily available geospatial datasets. The GFMS developed by our group uses a relatively simple hydrologic model, based on the run-off curve number method, to transform precipitation into run-off. A grid-to-grid routing scheme moves run-off downstream. Precipitation estimates are from the TRMM Multi-satellite Precipitation Analysis (TMPA). We first evaluated the TMPA algorithm using a radar/gauge merged precipitation product (Stage IV) over south-east USA. This analysis indicated that the spatial scale (and hence the basin size) as well as regional and seasonal considerations are important in using the TMPA to drive hydrologic models. GFMS-based run-off simulations were evaluated using observed streamflow data at the outlet of two US basins and also using a global flood archive. Basin-scale analysis showed that the GFMS was able to simulate the onset of flood events produced by heavy precipitation; however, the simulation performance deteriorated in the later stages. This result points out the need for an improved routing component. Global-scale analysis indicated that the GFMS is able to detect 38% of the observed floods; however, it suffers from region-dependent bias.

1. Introduction

Floods are the most widespread and frequent natural disaster and are responsible for significant loss of lives and property each year. Recent statistics show that the number of people affected by floods has been rising rapidly (both in absolute terms and relative to other forms of natural disasters) not only due to extreme weather conditions but also due to increasing urbanization and inadequate disaster response (IFNet Action Report 2006). It has been established that flood early warning systems are the most effective way to mitigate flood induced hazards. The possibility and reliability of such early warning systems depend heavily on the availability of good-quality

*Corresponding Author. Email: yilmaz@agnes.gsfc.nasa.gov

International Journal of Remote Sensing
ISSN 0143-1161 print/ISSN 1366-5901 online

http://www.tandf.co.uk/journals
DOI: 10.1080/01431161.2010.483489

precipitation estimates. Difficulties in estimating precipitation arise in many remote parts of the world and particularly in developing countries where ground-based measurement networks (rain gauges or weather radar) are either sparse or nonexistent, mainly due to the high costs of establishing and maintaining infrastructure. This situation imposes an important limitation on the possibility and reliability of flood early warning systems in these regions.

Recent improvements in the ability of satellite-based precipitation retrieval algorithms (e.g. Sorooshian *et al*. 2000, Hong *et al*. 2004, Joyce *et al*. 2004, Huffmann *et al*. 2007) to produce near-real time estimates (with quasi-global coverage) at high space and time resolutions make them potentially attractive for flood monitoring. Among others, Yilmaz *et al*. (2005), Hong *et al*. (2006), Artan *et al*. (2007), Harris and Hossain (2008) and Su *et al*. (2008) investigated the utility of satellite-based precipitation estimates for hydrologic applications. Their main conclusion was that although satellite-based precipitation estimates contain considerable error, the ongoing improvements and future planned satellite missions (such as the Global Precipitation Measurement (GPM) mission) make them potentially useful for hydrologic modelling of large basins after model calibration. Being restricted to a single or a limited number of watersheds, the above studies provide hydrologic insight into flood monitoring only at the local or regional scale. The increasing availability of precipitation estimates and geospatial datasets covering the globe at scales useful for hydrologic applications increase the possibility of establishing global flood monitoring systems. However, the science of how to implement, parameterize and calibrate hydrologic and hydraulic models at space (sub-degree) and time (daily to sub-daily) scales suitable for a global flood monitoring system is not yet well understood.

Hong *et al*. (2007) developed an initial satellite-based near-real-time *global flood monitoring system* (GFMS) which is operationally available at the Tropical Rainfall Measuring Mission (TRMM) website (http://trmm.gsfc.nasa.gov/). In this system, a relatively simple hydrologic model, based on the run-off curve number (CN) and antecedent precipitation index methods, transforms precipitation into run-off. A simple grid-to-grid routing scheme is used to move the run-off downstream. The key input to the current system is the precipitation estimates from the NASA-based TRMM Multi-satellite Precipitation Analysis (TMPA; Huffman *et al*. 2007, 2010). The objective of this study is to perform an initial evaluation of this GFMS in an effort to provide useful insights into its current strengths and limitations, and point toward potential improvements necessary for increasing its reliability (i.e. precision) and accuracy (i.e. unbiasedness).

The paper is organized as follows: section 2 presents the study area, the datasets and the hydrologic model. Section 3 outlines the methods used in the analysis. Section 4 initially provides an evaluation of the TMPA algorithm using a radar/gauge merged precipitation product over south-east USA; following this analysis an initial evaluation of the GFMS-based run-off simulation using observed discharge data (basin scale analysis) and an archive of global large flood events (global scale analysis) is presented. Finally, in section 5 we conclude and discuss the main findings and give insights into future research directions.

2. Study area, datasets and hydrological model

2.1 *Study area*

The study area used for the regional evaluation of the TMPA precipitation is the relatively humid south-eastern USA (figure 1). The basin-scale flow analysis includes

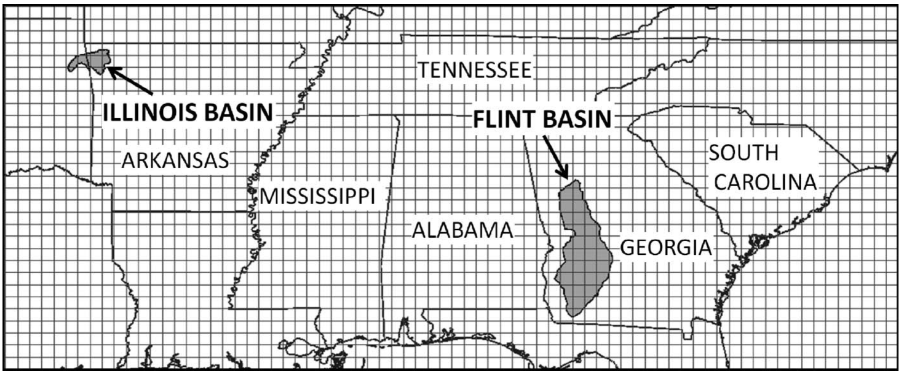

Figure 1. The study area. The grids represent the spatial resolution of the TMPA precipitation dataset (0.25° × 0.25°).

Table 1. Study basin characteristics and relevant information.

Basin ID	Basin name	Elevation (m)	Area (km^2)	P* (mm)	Q* (mm)
ILLINOIS	Illinois River near Tahlequah, OK	202.4	2484	1259	445
FLINT	Flint River near Bainbridge, GA	17.7	19606	993	197

P = mean annual precipitation from radar/gauge; Q = mean annual run-off.
*based on the July 2007–August 2008 time period.

two basins of varying size and geographic location within the study area (figure 1; table 1). The basins are free of snow and well instrumented with outlet stream gauges, rain gauges and weather radar and are therefore suitable for initial limited evaluation of the global flood monitoring system. The global-scale evaluation utilizes a large flood events archive and focuses on the 50° N–S latitude band.

2.2 *Radar/gauge merged precipitation estimates*

Six-hourly NCEP (National Centers for Environmental Prediction) gridded Stage IV precipitation estimates (Lin and Mitchell 2005) are available on the National Hydrologic Rainfall Analysis Project (HRAP) grid (~4 km × 4 km). NCEP Stage IV is a mosaic of the Stage III analyses produced by the National Weather Service (NWS). NWS uses a multivariate optimal estimation procedure to incorporate hourly raingauge data into the radar estimates (Seo 1998) which is followed by a quality control. Hereafter Stage IV precipitation dataset will be called as RADG (for RADar and Gauge). Although the NCEP Stage IV product is widely used in the literature for testing satellite-based algorithms, it has its own limitations and cannot be considered as truth (see Stellman *et al.* 2001 and Yilmaz *et al.* 2005).

2.3 *Satellite based precipitation estimates*

The Tropical Rainfall Measuring Mission (TRMM) Multi-satellite Precipitation Analysis (TMPA) (Huffman *et al.* 2007, 2010) provides precipitation estimates by combining information from multiple satellites as well as rain gauges where feasible, and is available at 3-hourly, 0.25° × 0.25° latitude–longitude spatial resolution covering the globe between the 50° N–S latitude band. The real-time product makes use of TRMM's highest quality observations, along with a high quality passive microwave-based rain estimates from three to seven polar-orbiting satellites and IR estimates from the international constellation of geosynchronous earth orbit satellites to fill in any remaining gaps. The resulting 3-hourly product is made up of about 80% microwave-based rain estimates and 20% IR-based rain estimates, all calibrated by information from TRMM. There are two TMPA products: (1) an experimental real-time monitoring product which is available approximately 9 hours after real-time; (2) a post-real-time research quality product available nearly 10 to 15 days after the end of each month. We will refer to these products as TMPA-RT and the TMPA research product, respectively. The TMPA research product differs from TMPA-RT mainly in two ways: (1) it incorporates monthly rain gauge analysis for bias correction; (2) it uses the TRMM Combined Instrument (TCI) precipitation product for the calibrating information, as opposed to the TRMM Microwave Imager (TMI) used in TMPA-RT. In this study, we used the TMPA research product because a recent upgrade (February 2009) to the TMPA-RT algorithm employs monthly climatological adjustments (also incorporates additional satellite data sources) to approximate the bias characteristics of the TMPA research product (Huffmann *et al*. 2010). We will simply refer to the TMPA research product as TMPA hereafter.

2.4 *Flood datasets*

Daily observed streamflow data for the study basins are obtained from the US Geological Survey website (http://www.waterdata.usgs.gov). Note that the observed streamflow is likely subject to errors due, for example, to rating curve transformation. However, for the purposes of investigating the performance of the global hydrological model used in this study, we assumed that these errors could be neglected.

An archive of large flood events over the globe was determined from the Dartmouth Flood Observatory (DFO) website (www.dartmouth.edu/~floods). DFO compiles information on large floods from a variety of news, governmental, instrumental and remote sensing sources. The information in the DFO archive comprises the location (latitude and longitude of the flood centroid), begin–end dates, damage (loss of life and property), main cause, areal extent and magnitude of large floods. Of course, the DFO flood archive has limitations. It is likely that many floods that occurred were not included in the database because (1) floods occurred in remote areas and were not recorded; (2) floods were recorded in a local database that is not available at the international level; (3) floods were recorded in foreign languages not known to the person creating the archive. In addition, start–end date and location of the floods may not be known precisely. In any case, we expect that extreme flood events are fairly well represented in the DFO archive and the comparison of the two datasets is worthwhile.

2.5 *Hydrologic model*

Hong *et al*. (2007) developed a relatively simple precipitation–run-off model based on the Natural Resources Conservation Service (NRCS) run-off curve number (CN) approach that converts the TMPA-based precipitation estimate into run-off at 0.25° × 0.25° latitude–longitude spatial resolution every 3 h. The resulting quasi-global (latitude band 50° N–S) run-off map is operationally available at http://trmm.gsfc.nasa.gov/.

The main advantage of the NRCS-CN approach is that it is a simple and well established methodology having only one parameter, the run-off curve number, which can be estimated using geospatial datasets. The NRCS-CN approach estimates surface run-off as a function of precipitation, soil type, land cover and antecedent moisture conditions. The run-off curve number is estimated from the area's hydrologic soil group (HSG), land use/cover and hydrologic condition. Hong and Adler (2008) proposed an approach to estimate CN using a global HSG map derived from Food Agricultural Organization soil dataset in conjunction with MODIS-derived land cover classification map. Using the standard look-up table for the 'fair' hydrologic conditions, Hong *et al*. (2007) estimated the time-variation of CN values under changing surface moisture conditions (dry or wet) by a concept based on antecedent precipitation index. A simple grid-to-grid routing scheme is then used to move the surface run-off downstream. The routing scheme makes use of the elevation, slope and flow direction information given by the HYDRO1k dataset to estimate the grid-to-grid flow velocity and direction. The current model set-up does not take into account groundwater storage and snow processes. For details of the hydrologic model, see Hong *et al*. (2007), and Hong and Adler (2008).

3. Methods

The initial objective was to evaluate the TMPA precipitation estimates using the RADG precipitation estimates because any error in the TMPA precipitation estimates will directly translate into the hydrologic model output.

The August 2006–July 2008 study period was selected based on data availability. Gridded datasets were aggregated into daily periods. Focusing on the spatial domain of the south-eastern USA, a scaling experiment was carried out to quantify the changes in the spatial skill of the TMPA dataset as compared to the RADG over increasing averaging scales. The scaling experiment starts with the 0.25° (TMPA native resolution) by upscaling the 4-km resolution RADG dataset via the box averaging method. The datasets were then incrementally upscaled to coarser resolutions (0.5, 1.0 and 2.0°) again with the box averaging method. Differences in the spatial distribution of daily TMPA and RADG precipitation estimates were investigated using quantitative and categorical statistics. The quantitative statistics include correlation coefficient (CORR), percentage bias (%Bias) and normalized root mean squared error (NRMSE). The latter two statistics were calculated for every grid as follows:

$$\%\mathrm{Bias} = \left(\frac{\sum_{i=1}^{n} T_i - R_i}{\sum_{i=1}^{n} R_i} \right) \cdot 100, \tag{1}$$

$$\mathrm{NRMSE} = \left(\sqrt{\frac{\sum_{i=1}^{n} (T_i - R_i)^2}{n}} \right) \bigg/ \left(\frac{\sum_{i=1}^{n} R_i}{n} \right), \tag{2}$$

where T and R represents the TMPA and the RADG precipitation estimates respectively, and $i = 1, 2, \ldots, n$ is the number of daily precipitation data pairs for each grid.

Categorical statistics were calculated for the gridded daily precipitation to evaluate the TMPA estimates in detecting rain events at 25th and 75th percentile precipitation thresholds; the latter threshold was selected to evaluate the TMPA skill in detecting large rain events which are more likely to cause flooding. The percentile (probability of exceedance) values were calculated for rain events larger than 1 mm day^{-1}. We used the percentile values (instead of the absolute values) as precipitation thresholds to make the categorical skill results consistent over different scales and hence to better assess the evolution in categorical skills with change in scale. Note that the percentile values are the same for different spatial scales but the absolute values change. Categorical statistics include probability of detection (POD) and false alarm ratio (FAR). These are based on a 2 × 2 contingency table [*a*—TMPA yes, RADG yes; *b*—TMPA yes, RADG no; *c*—TMPA no, RADG yes; and *d*—TMPA no, RADG no]. The POD [$= a/(a + c)$] gives the fraction of rain events that were correctly detected and ranges from 0 to 1; 1 being the perfect score. The FAR [$= b/(a + b)$] measures the fraction of rain events that were actually false alarms and ranges from 0 to 1; 0 being the perfect score.

Seasonal differences in precipitation events were also investigated by comparing daily mean basin precipitation from the TMPA and the RADG datasets. Mean basin precipitation was estimated by area-averaging the grid-based precipitation values over the Illinois basin and the Flint basin. The analysis was performed using scatter plots and quantitative statistics (calculated using the mean basin precipitation) for cold (December, January, February) and warm (June, July, August) periods.

Our second objective was to evaluate the GFMS simulation performance. A basin-scale analysis was carried out to compare the observed discharge values at the outlet of the two study basins with the flow values simulated by the GFMS. For each study basin, the GFMS output calculated for the model grid overlaying the basin outlet was used. During the course of this analysis it was found that due to current limitations (e.g. coarse spatial resolution and constant grid-to-grid flow velocity), the routing component was unable to adequately represent attenuation and delay mechanisms exerted by the river network on the run-off. As a result, the output of the GFMS more closely represents the run-off. Therefore, in this study, the GFMS output will be termed as run-off. Although some progress has been made (Oki and Sud 1998, Naden *et al.* 1999, Fekete *et al.* 2001, Gong *et al.* 2009), the science of up-scaling fine resolution river networks to a coarser resolution in a consistent and effective way while maintaining the necessary attenuation and delay mechanisms is not yet well understood. Recent studies focusing on scale independent approaches (e.g. Gong *et al.* 2009) show some promise; however they require intense labour time which currently limits their application at the global scale.

A global scale analysis that helps to understand, if present, the dependency of the GFMS performance on the geographic location is particularly important for model improvement. To investigate this, we tested the ability of the model to detect historical large flood events provided by the DFO. Focusing on the 16-month period of April 2007–July 2008, we selected the flood events that are caused by heavy precipitation and excluded those events caused by snow melt and man-made features such as dams. We defined simulated flood events in the following manner. In a moving window of size 2.25° × 2.25°—i.e. the window extends four grids (1°) away from the centre grid—over the globe, if there exist at least two contiguous model grids with GFMS-simulated run-off higher than a selected run-off threshold, that window location was labelled as 'flooded' for that 3-h time step.

We decided to choose a 2.25° spatial window for the following reasons. The flood archive (compiled by the Dartmouth Flood Observatory) is mainly based on news sources such as newspapers, and organizations such as the United Nations. In these sources the floods are often described by a city, district or neighbourhood name and lack the exact geographical coordinates. Therefore flood location is approximate. We, therefore, extended the window 1° away from the centre grid (hence the window size becomes 2.25°). Note also that the global flood archive contains extreme flood events with large spatial scale. We therefore defined the simulated floods with at least two contiguous model grids having run-off values over the run-off threshold.

The run-off threshold at each 2.25° $\times$ 2.25° grid was defined in a way to simply account for hydro-climatic variations across the globe. In the analysis, five different geographical zones were defined based on the Koppen climate classification and the run-off threshold for each zone was selected as the 0.98 exceedance probability of 3-hourly run-off in each zone during the study time period. The run-off thresholds for each zone are listed in table 2. It can be seen that Zone 3 (Europe) has the lowest run-off threshold, whereas Zone 5 (Asia–Australia) has the largest.

Once the simulated flood events were defined, the next step was to identify the measures to evaluate the performance of the GFMS in detecting historical floods. For this task, we used the probability of detection (POD) and the false alarm count categorical performance measures. In the POD calculation, we considered an observed flood event successfully detected by the GFMS (hit), if within the 2.25° $\times$ 2.25° moving window there exists an observed flood, and within a $\pm$1-day temporal window of that observed flood duration there also exists at least one time step (3 h) with simulated flood. We decided to set a $\pm$1-day temporal window in flood timing because the beginning and ending times of the floods may not be exact in a global archive. In addition, there are time zone differences in reporting of floods over the globe. In each geographical zone, the ratio between the number of successfully detected flood events (hits) to the total number of observed flood events is the POD measure for that zone and provides insight into the predictive capacity of the hydrologic model.

For a complete analysis, a complementary measure indicative of the falsely predicted floods is necessary, e.g. the false alarm count. The obvious difficulty in detecting model-simulated false alarms using a global flood archive is that the archive may not contain all the flood events that occurred during the study time period. To define the false alarms, we divided the study time period into 7-day periods which we termed as ‘events’. The 7-day time period corresponds to the median value of the duration of the observed floods within the five regional zones defined earlier. For each 7-day event

Table 2. Number of observed floods in the Dartmouth Flood Observatory archive and the GFMS performance statistics. Numbers in parenthesis indicate the number of GFMS detected events (hits). See figure 7 for the extent of each zone.

Zone	1	2	3	4	5	Global
Region (approximate)	North America	South America	Europe	Africa	Asia-Australia	
Number of observed flood events	13	34	24	47	129	247
Run-off threshold (mm/3 h)	80	88	63	90	128	—
POD	0.38 (5)	0.35 (12)	0.33 (8)	0.34 (16)	0.40 (52)	0.38 (93)

we performed the following: in a 2.25° × 2.25° moving window, if there exist at least eight simulated flood time steps (corresponds to one day) and if there exists no observed flood then that event was marked as a 'false event' for that 2.25° × 2.25° grid. The false alarm count measure at each 2.25° × 2.25° grid is simply the count of false 7-day events over the study time period.

4. Results

4.1 *Comparison of the precipitation estimates*

This comparison study aims at evaluating the TMPA skill in characterizing the spatial variation, occurrence and magnitude correspondence of daily precipitation events over south-east USA (including the Flint basin and the Illinois basin) at various spatial scales. The RADG dataset was used as the reference dataset. The TMPA skill was assessed using maps showing spatial distribution of quantitative (figure 2) and categorical (figure 3) statistics and using box-plots showing the summary of the skill distribution. In a box-plot, the box contains horizontal lines at the 25th, 50th and 75th percentiles of the distribution and vertical lines extend from each end of the box to show the extent of the rest of the data (in this case 1.5 times the 25^{th}–75th percentile range). The mean values are represented by circles inside the box. Outliers are represented by '+' markers.

The magnitude correspondence of daily precipitation events and their seasonal dependency was analysed using scatter plots and quantitative statistics constructed using mean areal precipitation estimates over the Flint basin and the Illinois basin.

Figures 2(*a*)–2(*c*) show the spatial distribution of the quantitative statistics calculated between the TMPA and the RADG precipitation datasets over south-east USA at 0.25° spatial resolution. Box-plots in figures 2(*d*)–2(*f*) summarize these statistics as a function of spatial scale. Starting with the %Bias statistic, figure 2(*a*) shows that TMPA underestimates precipitation by 10% to 30% as compared to RADG over the east and west parts of the domain. The underestimation is most significant in the north-east region with %Bias values in the order of –50%. In the centre and centre-east regions TMPA shows a +10% to +30% bias, hence indicating higher precipitation compared to RADG. Over the Flint basin, TMPA predominantly overestimates precipitation (up to +30% Bias), specifically in the upstream region (north), which in turn will likely lead to higher GFMS-simulated run-off estimates. In the Illinois basin, %Bias varies between –10% and +10%, indicating that TMPA both underestimates and overestimates the RADG precipitation. Box-plots in figure 2(*d*) show that as the spatial grid resolution is increased incrementally from 0.25° to 2° the mean %Bias (black-circle-marker) between the TMPA and the RADG only slightly changes from –7.5% bias at 0.25° spatial resolution to –9% bias at 2° spatial resolution while the spread of the %Bias distribution becomes narrower and closer to zero bias (zero bias is the perfect match), as expected. For example, at the 0.25° grid resolution the %Bias statistic generally varies between +40% and –50% while at the 2° grid resolution, the %Bias statistic generally varies between +18% and –40%. We note, however, that at coarse spatial resolutions, the less frequent extreme precipitation events—that are likely to trigger floods—will be averaged out and will in turn potentially deteriorate the flood detection performance of the GFMS simulations.

Figure 2(*b*) shows the spatial distribution of the time correlation statistic (CORR) calculated between the TMPA and the RADG precipitation estimates. The time

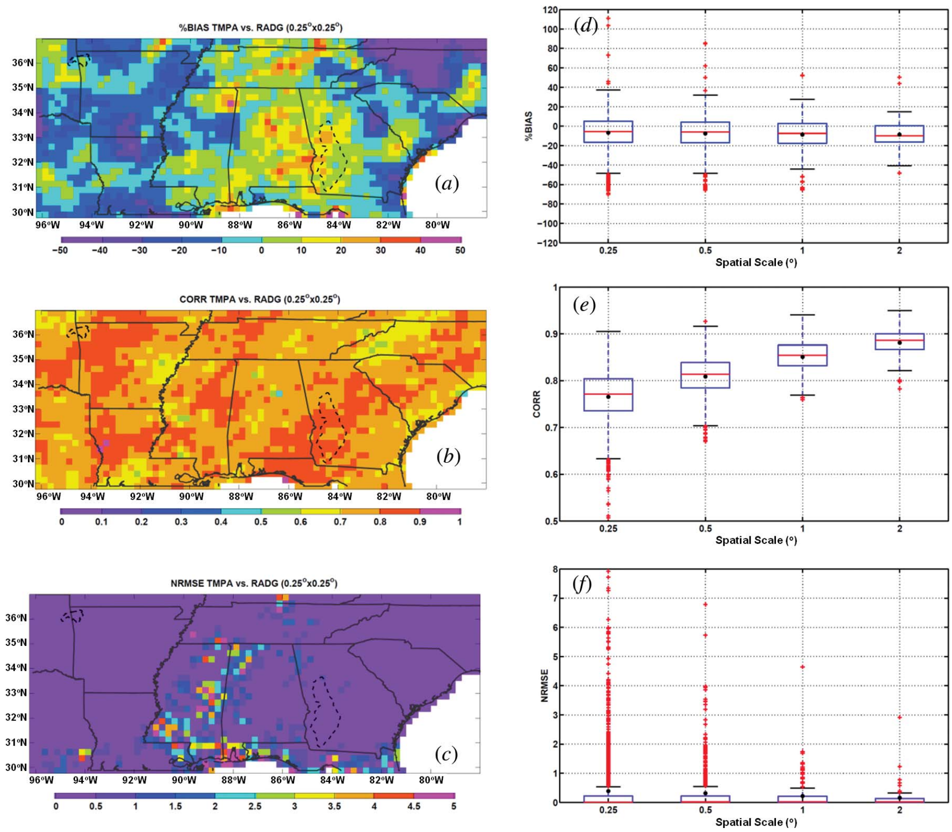

Figure 2. (*a*) Percentage bias, (*b*) correlation coefficient and (*c*) normalized root mean squared error between the TMPA and the RADG precipitation estimates (mm day^{-1}) at 0.25° spatial scale over the two-year study period. The two dashed-line polygons represent the boundaries of the Illinois basin (smaller polygon) and the Flint basin (larger polygon). Box-plots (*d*), (*e*) and (*f*) show, respectively, the summary statistics for %Bias, correlation coefficient and normalized root mean squared error as a function of spatial scale.

correlation statistic is relatively high over the domain and generally varies between 0.7 and 0.8 with patches of values at 0.6–0.7 and 0.8–0.9. Relatively high correlations indicate a good match in the timing of the TMPA and the RADG estimated precipitation events. Over the Flint basin correlations predominantly vary between 0.8 and 0.9 and over the Illinois basin the correlations vary between 0.7 and 0.8, indicating relatively good timing of TMPA precipitation events compared to those of RADG. Box-plots in figure 2(*e*) indicate that the time correlation between TMPA and RADG increases as the spatial grid is incrementally increased from 0.25° to 0.5°, 1° and 2° scale. At the 0.25° grid resolution, the mean correlation value between the TMPA and the RADG is 0.76. As the spatial grid scale is increased to 0.5°, 1° and 2°, the mean correlation between the TMPA and the RADG is significantly increased to the values of 0.8, 0.85 and 0.88 respectively; meanwhile the spread of the distributions becomes smaller (indicated as narrower boxes) as expected due to scale averaging.

Figure 2(*c*) shows the spatial distribution of the normalized root mean square statistic (NRMSE) between the TMPA and the RADG precipitation estimates. NRMSE is relatively low over the domain with values less than 0.2. An exception

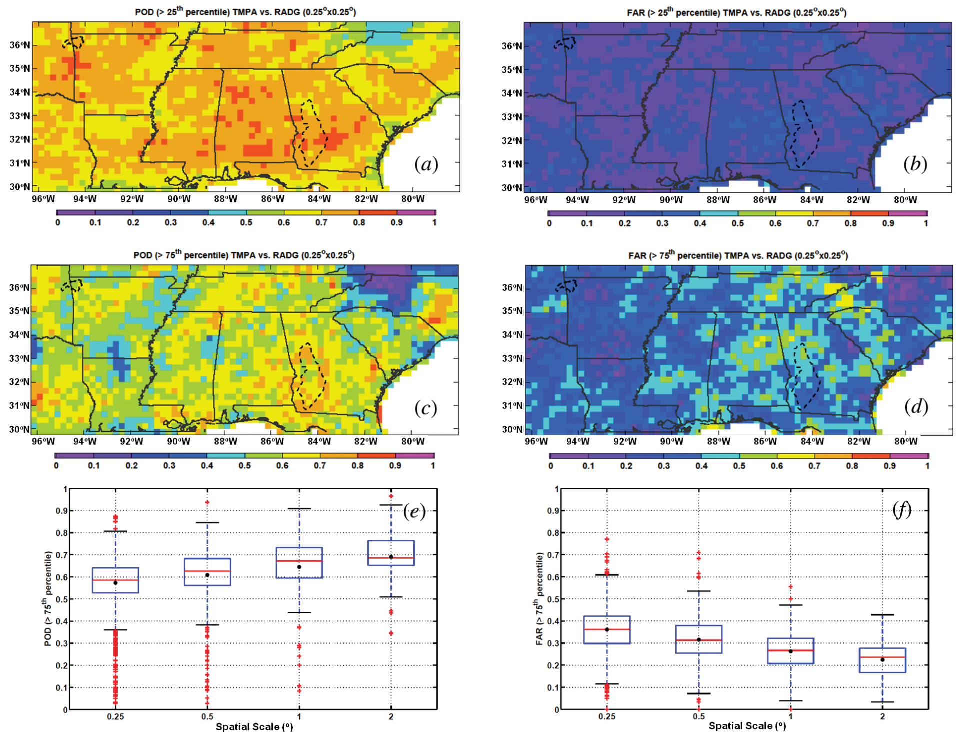

Figure 3. Spatial distribution of TMPA daily precipitation, (*a*) POD and (*b*) FAR calculated using a 25th percentile precipitation threshold at 0.25° spatial scale. Spatial distribution of TMPA daily precipitation, (*c*) POD and (*d*) FAR calculated using a 75th percentile precipitation threshold at 0.25° spatial scale. The two dashed-line polygons represent the boundaries of the Illinois basin (smaller polygon) and the Flint basin (larger polygon). Box-plots showing the summary statistics for (*e*) POD (>75th percentile) and (*f*) FAR (>75th percentile) as a function of spatial scale.

is the central region, which is characterized by NRMSE values that vary generally between 1.5 and 2 and occasionally as high as 8. Higher NRMSE values in the central region indicate possible problems in detecting the magnitude of the precipitation events in this region. Over the Flint basin and the Illinois basin the NRMSE values are lower than 0.2, indicating relatively good performance of TMPA in detecting precipitation events. Box-plots in figure 2(*f*) indicate that, as the spatial grid scale is increased from 0.25° to 0.5, 1 and 2°, the mean NRMSE statistic slightly decreases from 0.39 to the values of 0.31, 0.22 and 0.16 respectively; meanwhile the spread of the NRMSE distribution reduces as expected due to scale averaging.

Next, we present the spatial distribution of the categorical statistics (POD and FAR) to evaluate the TMPA performance in detecting daily precipitation events as compared to the RADG precipitation dataset. Figure 3(*a*) shows the spatial distribution of the POD statistic calculated for the 25th percentile precipitation threshold. The POD values over the domain generally vary between 0.6 and 0.8, indicating that TMPA has a good ability to detect precipitation events above the 25th percentile. Central and eastern regions of the domain are characterized by patches of POD values varying between 0.8 and 0.9. Over the northeast and eastern coast however, lower POD values (between 0.4 and 0.6) can be seen. Tian *et al.* (2007) indicated that low POD over the north-east and eastern coast of south-east USA could be due to the decreased amount of convective precipitation at higher latitudes and also due to joining the two different passive microwave retrieval algorithms—over land and over ocean—along the coast (Adler *et al.* 1993). Over the Flint and Illinois basins the POD values generally vary between 0.7 and 0.8, indicating the relatively good ability of TMPA to detect precipitation events at daily scales at 0.25° spatial resolution. Figure 3(*b*) shows the spatial distribution of the FAR statistic calculated for the 25th percentile precipitation threshold. The domain is characterized by FAR values varying between 0.1 and 0.3, indicating that TMPA has only a slight false precipitation event problem in the study area during the selected time period. There also exist a few patches of FAR values varying between 0.3 and 0.4 without a clear trend. Over the Flint and Illinois basins the FAR values are generally between 0.1 and 0.3, indicating that TMPA has only a slight false precipitation problem in these basins at the 0.25° spatial resolution.

As our main focus is flood detection, it is important to investigate the ability of TMPA to detect large precipitation events which are more likely to generate floods. Figure 3(*c*) shows the spatial distribution of POD calculated for the 75th percentile precipitation threshold at 0.25° spatial scale. POD (>75th percentile) values over the domain show a decrease when compared to POD calculated for the 25th percentile precipitation threshold. The POD generally varies between 0.4 and 0.7 and is noticeably low in the north-east and coastal regions. Over the Flint basin POD generally varies between 0.5 and 0.7 with presence of values as high as 0.7 and 0.8. Over the Illinois basin, however, the POD values are lower and vary between 0.5 and 0.7. Boxplots in figure 3(*e*) indicate that, as the spatial grid scale is increased from 0.25° to 0.5°, 1° and 2°, the mean POD (>75th percentile) statistic slightly increases from 0.57 to the values of 0.61, 0.65 and 0.69 respectively, indicating slight TMPA skill increase in detecting precipitation events at larger spatial averaging scales.

Figure 3(*d*) shows the spatial distribution of the FAR statistic calculated for daily precipitation events higher than the 75th percentile. The FAR (>75th percentile) statistic generally varies between 0.2 and 0.6 over the domain with the lowest (best) values in the north-east. Over the Illinois basin FAR predominantly varies between 0.3 and 0.4 while over the Flint basin the FAR values are higher and vary between 0.3 and 0.6. At the 0.25° spatial scale, the mean FAR (>75th percentile) value over the domain is 0.36, which decreases to the values of 0.32, 0.26 and 0.22 as the spatial scale is incrementally increased to 0.5°, 1° and 2° (figure 3(*f*)). In summary, the above analysis indicates that regional considerations as well as spatial scale—and hence basin size—are important in using the TMPA product as input to hydrological models.

Scatter plots in figure 4 facilitate a visual comparison of the magnitude correspondence of daily RADG and TMPA precipitation events as averaged over the study basins. The diagonal line indicates a 'perfect' correspondence. The study period is divided into cold (December, January, February) and warm (June, July, August) seasons to examine the seasonal behaviour between TMPA and RADG. Comparison of figures 4(*a*) and 4(*b*) indicates that TMPA overestimates the precipitation in the Flint basin regardless of the season (indicated by positive %Bias; see also figure 2(*a*)). Based on the CORR and %Bias statistics, the agreement between the RADG and TMPA

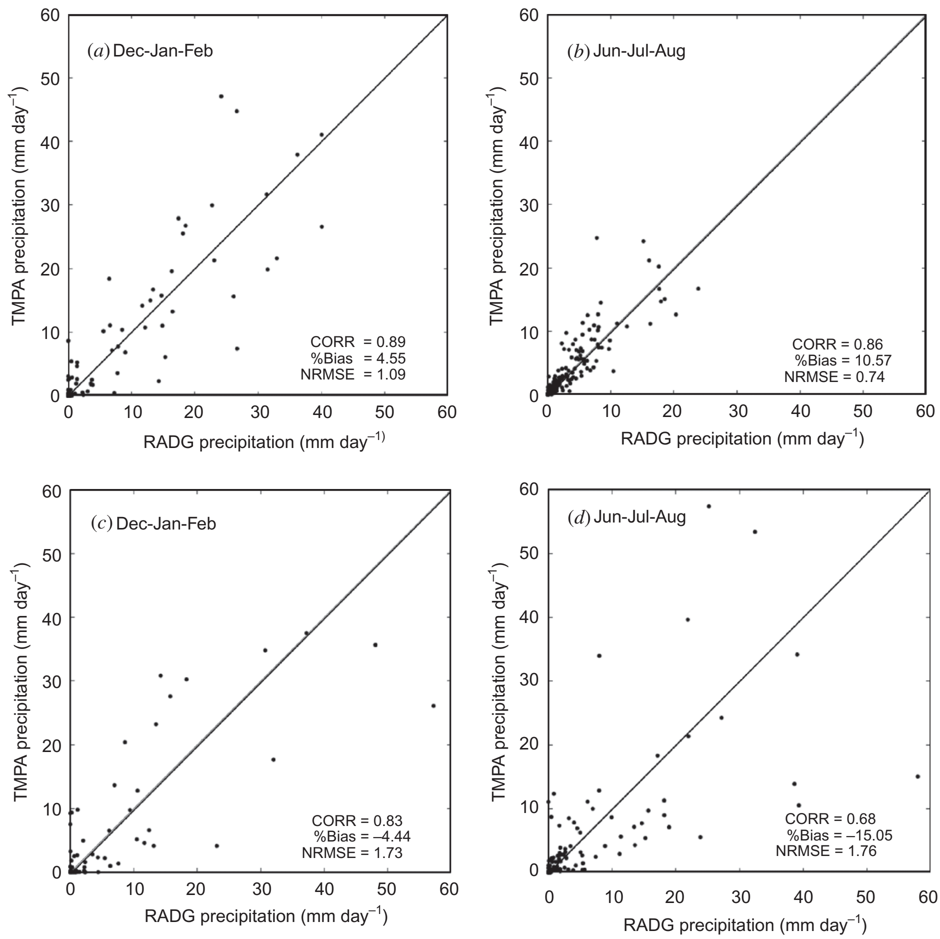

Figure 4. Scatter plots for the Flint basin in (*a*) winter and (*b*) summer; and for the Illinois basin in (*c*) winter and (*d*) summer.

estimates is more pronounced during the cold season (0.89 CORR and +4.55 %Bias) compared to the warm season (0.86 CORR and +10.57 %Bias). NRMSE shows opposite behaviour with a better statistic in summer (0.74 NRMSE) than winter (1.09 NRMSE). However the NRMSE is likely to be influenced by the large precipitation events during winter (despite the normalization). In the Illinois basin (figures 4(*c*) and 4(*d*)) TMPA tends to underestimate precipitation compared to RADG regardless of the season (indicated by negative %Bias; see also figure 2(*a*)). This behaviour is more pronounced in the warm season as indicated by –15% bias compared to –4.4% bias in the cold season. Both CORR and NRMSE statistics deteriorate (the former more significantly) in the summer season (0.68 CORR and 1.76 NRMSE) compared to the cold season (0.83 CORR and 1.73 NRMSE), most likely due to the local character of convective precipitation in summer. Local, small size precipitation events may not be captured by the large footprint of the TMPA product. These findings agree with Tian *et al.* (2007) who compared the TMPA and Stage IV radar precipitation estimates over the USA for 3-year seasonal accumulations at 0.25° spatial grids. Tian *et al.* (2007)

reported that the TMPA overestimates Stage IV precipitation over the Flint basin regardless of the season. Over the Illinois basin their results showed the TMPA underestimating Stage IV precipitation in summer. In winter however, they found both overestimation and underestimation between the TMPA and Stage IV precipitation estimates for the grids within the Illinois basin. Similar to our findings, Tian *et al.* (2007) reported higher (lower) correlations in winter (summer) between the TMPA and Stage IV estimates. In the scatter plots shown in figure 4, the large precipitation events along either x-axis or y-axis are particularly important if these estimates are to be used for flood detection. Points closely located to the x-axis (y-axis) represent undetected (falsely predicted) large precipitation events by TMPA. This behaviour can be seen in figures 4(*a*), 4(*c*) and 4(*d*) for precipitation magnitudes of up to 10 mm day^{-1}. One reason might be that the satellite-based precipitation estimation relies on a single snapshot in a 3-hour window, and hence it is likely that the intermittent sampling by the satellites missed the peak in precipitation; whereas radar estimates are calculated by merging more frequent samples within a 3-hour window. For larger precipitation events RADG and TMPA agree well on the occurrence of the events but with relatively large random errors (these are the points away from both axis lines).

In summary, the above analysis indicates that seasonal consideration is important in using TMPA-based precipitation estimates for flood prediction.

4.2 *Evaluation of the run-off simulation performance*

Figures 5(*c*) and 5(*d*) facilitate a comparison between the observed discharge at the Flint basin outlet (normalized by the basin area) and the GFMS-simulated run-off. It can be seen that most of the simulated high run-off events correspond relatively well with the observed high discharge events. However, the simulated events are earlier and flashier than the observed events due likely to the inadequacy of the routing component in representing the attenuation and delay mechanisms exerted by the river network on the run-off (see §3 for a discussion). As an example, in figure 5(*c*) (5(*d*)), the numbers in parenthesis indicate the difference, in days, between the peak precipitation time and the peak discharge (run-off) time. It can be seen that, in the Flint basin, observed peak flow arrives on average four days after the peak precipitation. Note that the GFMS-simulated run-off occurs at the same day or one day later than the precipitation peak. Therefore a robust routing component that can adequately represent the smoothing and delay mechanisms in the basin outlet flow is seen as one major direction of improvement in the current GFMS system.

In addition, the hydrologic model does not properly represent the observed baseflow and water storage in the Flint basin (compare the January–May 2008 period in figures 5(*c*) and 5(*d*)). This is a limitation of the CN-based approach which only calculates an estimate of the excess precipitation in each model grid and does not consider groundwater storage in the soil column as well as interflow and baseflow from the soil column. Due to lack of baseflow representation and limitations in the flow routing there is a magnitude difference and timing mismatch between the observed discharge and simulated run-off values given in figures 5(*c*) and 5(*d*). These limitations are being considered in our effort to develop a new, more physically-based hydrologic model with groundwater representation and an improved routing scheme. Note also that there are several simulated high run-off events that are not evident in the observed discharge (e.g. mid-October, late November, mid-December 2007). Comparison of the RADG and the TMPA rainfall estimates (figures 5(*a*) and

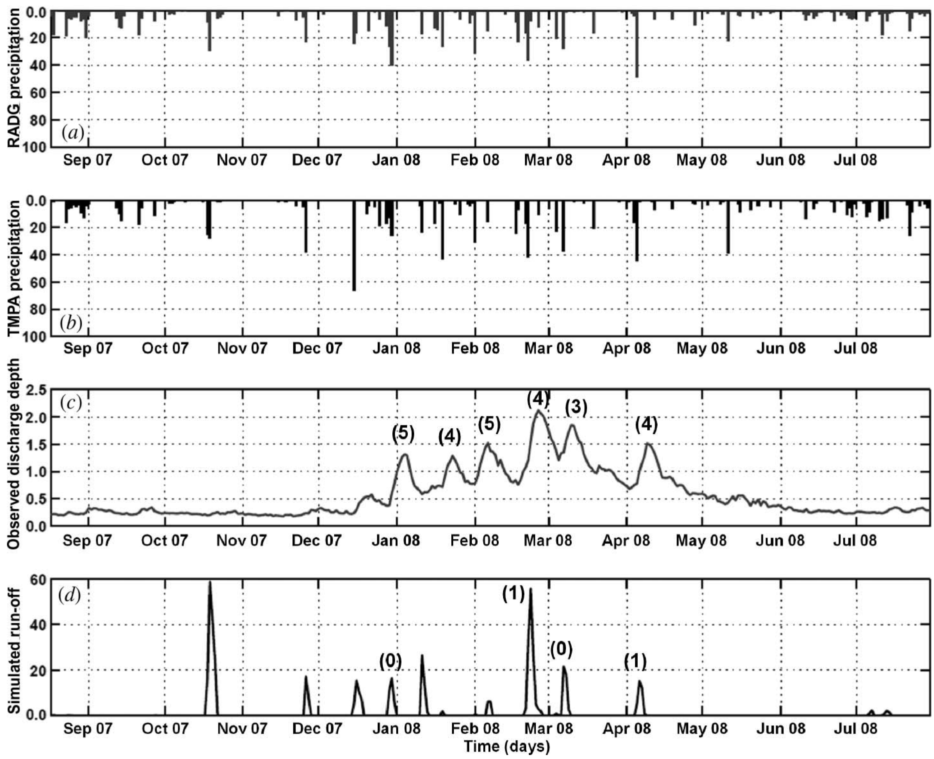

Figure 5. (*a*) RADG precipitation, (*b*) TMPA precipitation, (*c*) observed discharge depth and (*d*) simulated run-off values for the Flint basin (all units are in mm day^{-1}). Note: numbers in parenthesis indicate the difference (in days) between peak precipitation timing and peak discharge timing.

5(*b*)) for these time steps indicate that these events are in part due to positive bias in the TMPA precipitation estimates.

In the Illinois basin, observed discharge (figure 6(*c*)) is flashier and is comprised of a smaller baseflow component compared to the Flint basin partly due to the smaller basin size. For example, in figure 6(*c*) (6(*d*)), the numbers in parenthesis indicate the difference, in days, between the peak precipitation timing and the peak discharge (run-off) timing. It can be seen that, in the Illinois basin, observed peak flow arrives one day or two days after the peak precipitation. The GFMS-simulated peak run-off occurs at the same day or one day after the precipitation peak. Again, the hydrologic model lacks the attenuation and delay mechanisms, but this problem is less pronounced for the smaller Illinois basin compared to the larger Flint basin. This behaviour is expected because the larger Flint basin (1) introduces more attenuation and delay while converting run-off into discharge and (2) has more baseflow component. Therefore, the effect of the routing component is more prominent in large basins. The simulated run-off contains both overestimated events (e.g. late August 2007 and early March and early May 2008) and underestimated events (e.g. mid-March 2008) compared to the observations, which in part is due to the bias in the satellite-based precipitation estimates (compare figures 6(*a*) and 6(*b*) for these time periods).

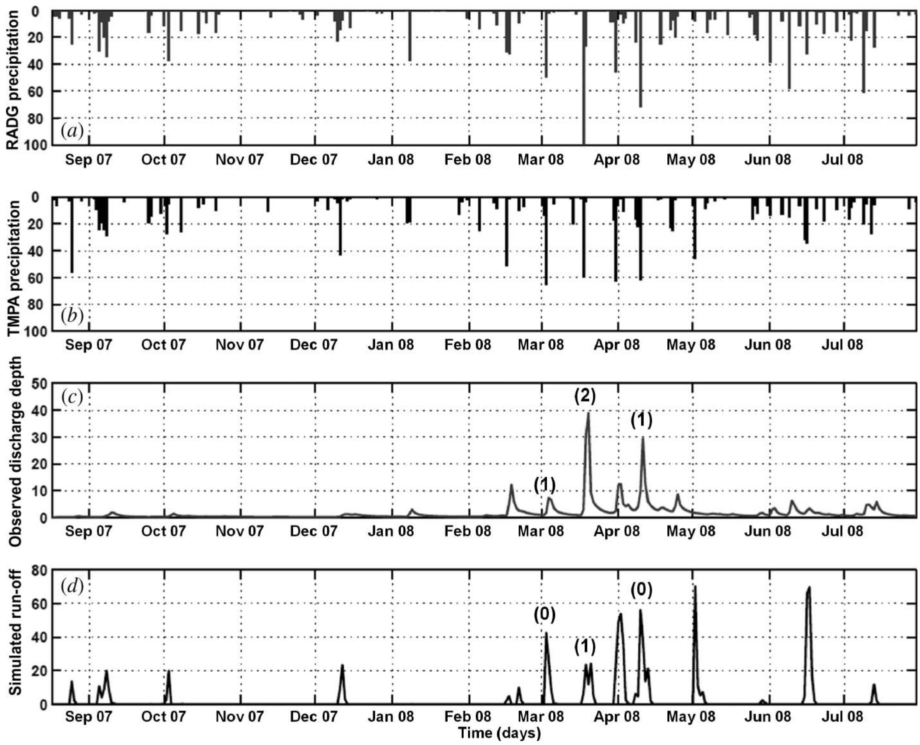

Figure 6. (*a*) RADG precipitation, (*b*) TMPA precipitation, (*c*) observed discharge depth and (*d*) simulated run-off values for the Illinois basin (all units are in mm day^{-1}). Note: numbers in parenthesis indicate the difference (in days) between peak precipitation timing and peak discharge timing.

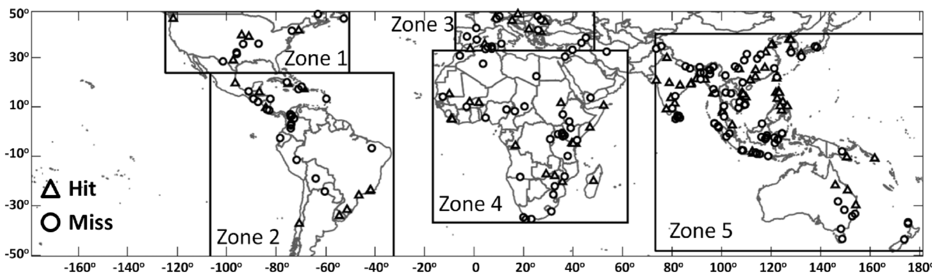

Figure 7. Locations of the historical flood events compiled by the Dartmouth Flood Observatory. Observed flood events detected by the hydrologic model (hits) are shown as triangle markers. Undetected flood events (misses) are shown as circle markers. Boxes denote the zones with different run-off threshold.

Focusing now on the global evaluation of the GFMS performance, figure 7 shows the locations of the historical observed flood events (all markers) during the April 2007–July 2008 period and table 2 lists the observed flood count in each zone. Following the POD calculation procedure described in §3, the triangle markers and circle markers in figure 7 represent those flood events that were successfully detected (‘hit’) and undetected (‘missed’) by the GFMS respectively. Over the globe, the GFMS was capable of detecting 93 out of 247 events with a global POD value of 0.38 during the study time period (table 2). The POD value calculated for each zone (table 2) shows similarity and varies between 0.33 for Zone 3 (Europe) and 0.40 for Zone 5 (Asia–Australia). Zone 5 has the highest POD value possibly due to the fact that the floods in this region are generally caused by the large scale tropical and monsoonal heavy precipitation systems; these events can be relatively easily detected by the TMPA algorithm. In other zones the GFMS skill in detecting the observed flood events is similar.

The colour coded map in figure 8 shows the number of false 7-day events simulated by GFMS based on the false alarm count description given in §3. GFMS has a tendency to simulate a high number of false flood events over the globe, specifically within the tropics. The highest number of simulated false flood events is generally located in the proximity of the outlet of the large river basins of the globe, such as the Amazon, the Ganges, the Niger, the Mississippi and the Pearl Rivers. Zone 2 (South America) has the highest number of false flood events, which are mainly located in the Amazon basin and the La Plata basin. This result is in parallel with Su *et al.* (2008) who evaluated the TMPA product over the La Plata basin using gauge-based precipitation and forced a hydrology model with TMPA to simulate streamflow. They found that, over the La Plata basin, both the TMPA precipitation estimates and the simulated streamflow contain positive bias regardless of the season when compared to the observations. Figure 7 shows that GFMS successfully detected the observed floods towards the south of Zone 2; however the GFMS skill deteriorated in the remaining parts. Note also that although the Amazon basin is characterized by a high number of simulated false events, there exist no observed floods within the basin. We suspect that the floods occurring in regions with sparse population are not reported and hence not archived in these remote regions of the world. Language barrier is another limiting factor for flood reports; flood information is mainly obtained from news agencies that report in English. The possible deficiencies in the GFMS system, including the hydrological model and the TMPA data, as well as the potential reporting problems such as those listed above, all contribute to the high number of false flood events simulated by the GFMS. In Zone 4 (Africa) GFMS also shows a tendency to overestimate the flood events, mainly in the tropics (figure 8). Therefore

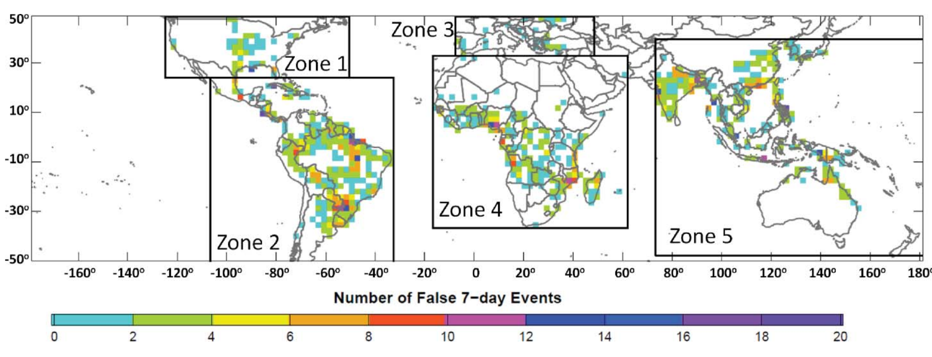

Figure 8. Colour map showing the number of simulated false seven-day events over the globe. Boxes denote the zones with different run-off threshold.

GFMS showed a fairly good skill in detecting the observed floods in the tropics, but the observed floods in the drier northern and southern parts were not detected by the GFMS (figure 7). Zone 5 (Asia and Australia), specifically India, is characterized by a high number of GFMS-simulated false flood events (figure 8), which in part explains the relatively good flood detection performance of GFMS in this region (see figure 7 and table 2). In Zone 1 (North America) and Zone 3 (Europe), the number of GFMS-simulated false flood events was significantly less. In addition, GFMS seems to be incapable of detecting floods in arid and semi-arid regions (e.g. North Africa, central Asia, the Middle East), and regions with complex topography (central America, central Asia).

The limitations in flood simulations described above are possibly due to a combination of factors related to the hydrological model as well as the TMPA-based precipitation estimates. First, the hydrological model does not consider snow processes and its relatively simple structure and coarse resolution may not be adequate to resolve complex topography and represent run-off generation and flow routing mechanisms. Second, since the gauge adjustment of the TMPA research product is performed at the *monthly* time scale, it is likely that the intense precipitation, which is more frequent in the warm season, at the 3-hourly time scale is scaled down. Therefore, considering that the study time period (April 2007–July 2008) is dominated by the warm seasons, the TMPA research product is likely to provide underestimated precipitation into the hydrological model. Other possibilities include that the gauge analysis missed the small-scale high precipitation events, which might even affect the monthly accumulations and unnecessarily force the TMPA precipitation estimates down. Last but not the least, there is a possibility that the intermittent sampling by the satellites missed the peaks in precipitation (George Huffmann, NASA Goddard Space Flight Center, personal communication).

5. Conclusions and discussion

We presented an initial evaluation of a satellite-based near real-time global flood monitoring system (GFMS) developed by Hong *et al.* (2007) and operationally available at http://trmm.gsfc.nasa.gov/. Our first objective was to evaluate the TMPA-based satellite precipitation estimates (the key input to the system) using NCEP Stage IV radar product. Spatial evaluation of the TMPA product over south-east USA indicated that regional and seasonal considerations as well as spatial scale (and hence basin size) are important factors in using the TMPA-based precipitation estimates as input to hydrologic models. For example, at the 0.25° spatial scale (the TMPA native scale), the TMPA precipitation estimates showed relatively good time correlation (generally CORR >0.7) with the NCEP Stage IV radar product, but suffered from relatively large bias (%Bias values vary between –40% and +40%) that can potentially translate into hydrologic simulations. A spatial scaling experiment showed that generally better performance can be obtained at larger averaging scales, indicating that the TMPA product is promising for use in hydrologic predictions in basins of large size.

Our second objective was to evaluate the simulation performance of the GFMS. The evaluation consisted of observed streamflow data at two basins and a global flood archive. The basin-scale analysis indicated that the GFMS was able to simulate the onset of flood events produced by heavy rainfall; however, the simulation performance deteriorated in the later stages of the flood events. Hence, the GFMS output is more similar to run-off at this stage of the GFMS development. A major finding, therefore, was that there is a need for an improved routing component, which is able to characterize flow attenuation and delay mechanisms exerted by the river network in a consistent manner. The global analysis showed that the GFMS performance varies by the geographical region, with best performances in tropical regions of East Asia. Our research into the science of how to implement, parameterize and calibrate hydrologic and hydraulic models at space (sub-degree to a few kilometres) and time (daily to sub-daily) scales suitable for a global flood monitoring system is still in its early stages and significant progress is yet to come. For example, a new hydrologic model, with an improved physical representation and routing component, is currently under development for use in this project and will likely lead to improved evaluation results. The results presented here will serve as a benchmark for future model improvements.

To realize the potential of global flood monitoring systems, simple and robust flow routing schemes that contain minimal calibration parameters are needed. Investigation into the spatial scales required to implement these routing schemes in global scale hydrologic monitoring could be performed considering the trade-off between model run-time and its capacity to consistently represent river network structure, flow attenuation and delay mechanisms at short (daily to sub-daily) time scales. The up-scaling procedures for extracting consistent coarse resolution river network structure from fine resolution networks (Naden *et al.* 1999, Fekete *et al.* 2001, Davies and Bell 2009, Gong *et al.* 2009) is foreseen as a key direction of research.

A difficulty in evaluation of a global flood monitoring system driven by remotely sensed precipitation is the fact that the errors in the hydrologic model output are the result of a combination of errors in the model structure itself and errors in the precipitation estimates. For example, our analysis indicated that the TMPA-based precipitation estimates may result in missed and/or false precipitation events, which directly translates into the hydrological model output. It is often difficult to isolate the contribution of precipitation error from the model output error. In this regard, the sensitivity of the GFMS output on the input precipitation error could be investigated with an analysis that drives the GFMS system with two different precipitation inputs, namely the TMPA product and a more reliable ground-based product, such as rain-gauge and/or radar network. This analysis could be further extended to basins located in various hydroclimatic regions in which reliable ground-based network are available (e.g. USA, Europe and Australia). We note that since the TMPA product used in this study contains monthly gauge information, areas with fewer or poor quality monthly rain gauges may have a large bias error which then may result in larger GFMS output error than shown here for south-east USA.

We also note that the satellite-based operational flood monitoring systems will always have a latency that is at least equal to the latency of the satellite-based precipitation product (currently 9 hours for the TMPA real time product). Therefore such systems can be used to monitor floods for basins with concentration times higher than this latency. We do not expect major improvements in the latency of the satellite-based precipitation products even with the planned future satellite missions, such as the Global Precipitation Mission (GPM). Nevertheless, we expect that the GPM project will lead to greatly improved and accessible satellite-based precipitation estimates and increased efforts for implementation of global flood monitoring systems.

Acknowledgements
This research is supported by NASA's Applied Sciences Program (Michael Goodman, Andrea Donellan and Stephen Ambrose) and NASA's Precipitation Measurement Missions (PMM) Program (Ramesh Kakar).

References

Adler, R.F., Keehn, P.R. and Hakkainen, I.M., 1993, Estimation of monthly rainfall over Japan and surrounding waters from a combination of low-orbit microwave and geosynchronous IR data. *Journal of Applied Meteorology*, **32**, pp. 335–356.

Artan, G., Gadain, H., Smith, J.L., Asante, K., Bandaragoda, C.J. and Verdin, J., 2007, Adequacy of satellite derived rainfall data for stream flow modelling. *Natural Hazards*, **43**, pp. 167–185.

Davies, H.N. and Bell, V.A., 2009, Assessment of methods for extracting low-resolution river networks from high-resolution digital data. *Hydrological Sciences Journal*, **54**, pp. 17–28.

Fekete, B.M., Vörösmarty, C.J. and Lammers, R.B., 2001, Scaling gridded river networks for macroscale hydrology: development, analysis, and control of error. *Water Resources Research*, **7**, pp. 1955–1967.

Gong, L., Widén-Nilsson, E., Halldin, S. and Xu, C.-Y., 2009, Large-scale runoff routing with an aggregated network-response function. *Journal of Hydrology*, **368**, pp. 237–250.

Harris, A. and Hossain, F., 2008, Investigating the optimal configuration of conceptual hydrologic models for satellite-rainfall-based flood prediction. *IEEE Geoscience and Remote Sensing Letters*, **5**, pp. 532–536.

Hong, Y. and Adler, R.F., 2008, Estimation of Global NRCS-CN (Natural Resource Conservation Service Curve Numbers) using satellite remote sensing and geospatial data. *International Journal of Remote Sensing*, **29**, pp. 471–477.

Hong, Y, Hsu, K.L., Gao, X. and Sorooshian, S., 2004, Precipitation estimation from remotely sensed imagery using artificial neural network – cloud classification system (PERSIANN-CCS). *Journal of Applied Meteorology*, **43**, pp. 1834–1853.

Hong, Y., Hsu, K.-L., Moradkhani, H. and Sorooshian, S., 2006, Uncertainty quantification of satellite precipitation estimation and Monte Carlo assessment of the error propagation into hydrologic response. *Water Resources Research*, **42**, W08421, doi: 10.1029/2005WR004398.

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

Huffman, G. J., Adler, R.F., Bolvin, D.T., Gu, G.J., Nelkin, E.J., Bowman, K.P., Hong, Y., Stocker, E.F. and Wolff, D.B., 2007, The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-global, multiyear, combined sensor precipitation estimates at fine scales. *Journal of Hydrometeorology*, **8**, pp. 38–55.

Huffman, G.J., Adler, R.F., Bolvin, D.T. and Nelkin, E.J., 2010, The TRMM Multi-satellite Precipitation Analysis (TMPA). In *Satellite Applications for Surface Hydrology*, F. Hossain and M. Gebremichael (Eds), (Berlin: Springer Verlag), accepted. ISBN: 978-90-481-2914-0, pp. 3–22.

IfNet Action Report, 2006, Good practices for substantial human loss reduction from floods. International Flood Network. Available online at: http://www.internationalfloodnetwork.org/index.html (accessed 12 May 2009)

Joyce, R.J., Janowiak, J.E., Arkin, P.A. and Xie, P., 2004, CMORPH: a method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. *Journal of Hydrometeorology*, **5**, pp. 487–503.

Lin, Y. and Mitchell, K.E., 2005, The NCEP stage II/IV hourly precipitation analyses: development and applications. Preprints, *19th Conference on Hydrology* 9–13 January 2005, San Diego, CA. Available on CD-ROM, 1.2.

NADEN, P., BROADHURST, P., TAUVERON, N. and WALKER, A., 1999, River routing at the continental scale: use of globally-available data and an *a priori* method of parameter estimation. *Hydrology and Earth System Sciences*, **3**, pp. 109–124.

OKI, T. and SUD, Y.C., 1998, Design of total runoff integrating pathways TRIP – a global river channel network. *Earth Interactions*, **2**, pp. 1–37.

SEO, D.J., 1998, Real-time estimation of precipitation fields using radar precipitation and rain gauge data. *Journal of Hydrology*, **208**, pp. 37–52.

SOROOSHIAN, S., HSU, K.-L., GAO, X., GUPTA, H., IMAM, B. and BRAITHWAITE, D., 2000, Evaluation of PERSIANN system satellite-based estimates of tropical rainfall. *Bulletin of the American Meteorological Society*, **81**, pp. 2035–2046.

STELLMAN, K.M., FUELBERG, H.E., GARZA, R. and MULLUSKY, M., 2001, An examination of radar and rain gauge–derived mean areal precipitation over Georgia watersheds. *Weather and Forecasting*, **16**, pp. 133–144.

SU, F., HONG, Y. and LETTENMAIER, D.P., 2008, Evaluation of TRMM Multisatellite Precipitation Analysis (TMPA) and its utility in hydrologic prediction in the La Plata Basin. *Journal of Hydrometeorology*, **9**, pp. 622–640.

TIAN, Y., PETERS-LIDARD, C.D., CHOUDHURY, B.J. and GARCIA, M., 2007, Multitemporal analysis of TRMM-based satellite precipitation products for land data assimilation applications. *Journal of Hydrometeorology*, **8**, pp. 1165–1183.

YILMAZ, K.K., HOGUE, T.S., HSU, K.-L., SOROOSHIAN, S., GUPTA, H.V. and WAGENER, T., 2005, Intercomparison of rain gauge, radar, and satellite-based precipitation estimates with emphasis on hydrologic forecasting. *Journal of Hydrometeorology*, **6**, pp. 497–517.