New Approaches to Hydrological Prediction in Data-parse Regions (Proc. of Symposium HS.2 at the Joint IAHS & IAH Convention, Hyderabad, India, September 2009). IAHS Publ. 333, 2009.

Evaluation of a satellite-based near real-time global flood prediction system

KORAY K. YILMAZ[1,2], ROBERT F. ADLER[1,2], YANG HONG[3] & HAROLD F. PIERCE[2,4]

1 *Earth System Science Interdisciplinary Center, University of Maryland, College Park, Maryland 20740, USA* yilmaz@agnes.gsfc.nasa.gov

2 *NASA Goddard Space Flight Center, Code 613.1, Greenbelt, Maryland 20771, USA*

3 *School of Civil Engineering & Environmental Sciences, University of Oklahoma, Norman, Oklahoma 73019, USA*

4 *Science Systems and Applications, Inc., Lanham, Maryland 20706, USA*

Abstract This study provides an evaluation of a global flood prediction (GFS) system utilizing satellite-based rainfall and readily available geospatial data sets. The GFS, developed by our group, uses a relatively simple hydrological model, based on the runoff curve number method to transform rainfall into runoff. A grid-to-grid routing calculates the flow. Rainfall estimates are from TRMM Multi-satellite Precipitation Analysis (TMPA). An evaluation of the TMPA algorithm using a radar/gauge merged rainfall product over two basins in the southeast USA indicated that seasonal and regional considerations as well as basin size are important in using TMPA to drive hydrological models. GFS-based flood predictions were evaluated using observed streamflow data, MODIS-based inundation maps and a flood database. The GFS was able to simulate the onset of flood events produced by heavy rainfall; however, the prediction deteriorated in the later stages. This result points out the need for an improved routing component. The model showed dependency by the geographical region. A new hydrological model, with an improved physical representation and routing component is currently under development and will likely lead to improved validation results.

Key words satellite-based rainfall estimation; hydrological models; global flood prediction

INTRODUCTION

Many hydrological simulation studies, whether related to climate change scenarios, flood forecasting, or water management, depend heavily on the availability of good-quality 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 hydrological forecasting and 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; Huffmann *et al*., 2007) to produce near-real-time estimates (with global coverage) at high space and time resolutions make them potentially attractive for hydrological forecasting in ungauged basins (e.g. Yilmaz *et al*., 2005; Su *et al*., 2008). This study provides an evaluation of an initial satellite-based near real-time global flood prediction system developed by Hong *et al*. (2007) and operationally available at our website (http://trmm.gsfc.nasa.gov/). In this system, a relatively simple hydrological model, based on the runoff curve number (CN) and antecedent precipitation index methods, transforms rainfall into runoff. Runoff is then routed grid-to-grid to estimate flow. The key input to the current system is the rainfall estimates from the NASA-based Tropical Rainfall Measuring Mission (TRMM) Multi-satellite Precipitation Analysis (TMPA; Huffman *et al*., 2007). This paper initially focuses on the evaluation of the TMPA algorithm using a radar/gauge merged rainfall product over two basins in the southeast USA. This analysis is followed by the evaluation of the runoff predictions using observed discharge data and other sources of flood information. This initial evaluation of the global flood prediction system is expected to provide useful insights into strengths and limitations of the initial global flood prediction system and point toward potential improvements necessary for increasing its reliability and accuracy.

STUDY AREA, DATA SETS AND HYDROLOGICAL MODEL

Study area

The study area includes two basins of varying size and geographic location within the relatively humid southeastern USA (Fig. 1; Table 1). The basins are free of snow. The study area is well instrumented with outlet streamgauges, raingauges and weather radar, and therefore it is suitable for the evaluation of the global flood prediction system.

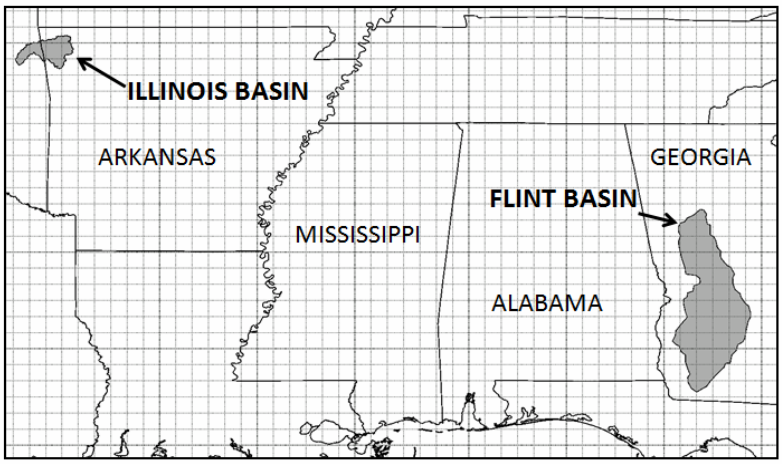

Fig. 1 Study area. TMPA product grids with 0.25° × 0.25° resolution are also shown.

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	2 484	1259	445
FLINT	Flint River near Bainbridge, GA	17.7	19 606	993	197

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

Radar/gauge merged precipitation estimates

Six-hourly NCEP (National Centers for Environmental Prediction) gridded Stage IV precipitation estimates (Lin & 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. Mean areal precipitation for a basin was calculated by averaging the HRAP grids contained in the basin. Hereafter, this precipitation data set will be called RADG. Although 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; Yilmaz *et al*., 2005).

Satellite-based precipitation estimates

The Tropical Rainfall Measuring Mission (TRMM) Multi-satellite Precipitation Analysis (TMPA) (Huffman *et al*., 2007) provides precipitation estimates by combining information from multiple satellites, as well as raingauges where feasible, and is available at 3-hourly, 0.25° × 0.25° latitude–longitude spatial resolution covering the globe between latitude band 50°N–S. 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 all the geo-synchronous IR sensors. Low Earth-orbit microwave data have a strong physical relationship to the hydrometeors that result in surface precipitation, but each individual microwave sensor provides a very sparse sampling of the time–space occurrence of precipitation. In contrast, the geosynchronous IR data provide excellent time–space coverage but poorly correlated to surface precipitation. In TMPA algorithm IR data are calibrated using the microwave data to improve the rainfall estimates (see Huffmann *et al*., 2007).

A post-real-time product also incorporates monthly raingauge analysis for bias correction. We used the post-real-time TMPA product because a recent upgrade (February 2009) to the real-time algorithm incorporates additional satellite data sources and employs monthly climatological adjustments to approximate the bias characteristics of the research quality post-real-time product (Huffmann *et al*., 2009). Mean areal precipitation (TMPA, hereafter) for a basin was calculated by area averaging the 0.25° × 0.25° grids contained in the basin.

Flood data sets

Daily observed streamflow data for the study basins are obtained from the US Geological Survey website (http://www.waterdata.usgs.gov). A list of large flood events was determined from the Dartmouth Flood Observatory (DFO) website (http://www.dartmouth.edu/~floods). The DFO compiles information on large floods from a variety of news, governmental, instrumental and remote sensing sources. The archive contains statistics related to location (latitude and longitude of the flood centroid) name of rivers, begin–end dates, damage (loss of life and property), main cause, areal extent and magnitude. MODIS-based inundation maps provided by DFO were also used in the evaluation.

Hydrological model

A relatively simple rainfall–runoff model based on the Natural Resources Conservation Service (NRCS) runoff curve number (CN) approach converts TMPA-based satellite rainfall into runoff estimate at 0.25° × 0.25° latitude–longitude spatial resolution every 3 h (Hong *et al*., 2007). The resulting quasi-global (between latitude band 50°N–S) runoff map is operationally available at http://trmm.gsfc.nasa.gov/. The NRCS-CN approach estimates surface runoff as a function of precipitation, soil type, land cover and antecedent moisture conditions. The only parameter, the runoff curve number (CN), is estimated from the area's hydrological soil group (HSG), land-use/cover and hydrological condition. Hong *et al*. (2007) proposed an approach to estimate CN using a global HSG map derived from the Food and Agricultural Organization soil data set in conjunction with a MODIS-derived land cover classification map. They also suggested a concept based on antecedent precipitation index to estimate the time variation of CN values under changing surface moisture conditions (dry or wet). A simple grid-to-grid routing scheme is then used to move the surface runoff downstream. For details see Hong *et al*. (2007).

METHODS

The first objective was to analyse the differences between TMPA and RADG precipitation estimates. The August 2006–July 2008 study period was selected based on data availability. Data sets were aggregated into daily periods and the time periods that are missing from either of the precipitation estimates have been excluded from the analysis. Cumulative precipitation graphs were constructed to analyse the bias behaviour between data sets throughout the study time period. Differences in daily precipitation estimates were further investigated using scatter plots constructed for cold (DJF: December, January, February) and warm (JJA: June, July, August) seasons, and quantitative statistics including percent bias (difference in TMPA and RADG estimates as a fraction of RADG estimate, times 100) and correlation coefficient. Categorical

statistics were calculated for the daily precipitation to evaluate the TMPA estimates in detecting rain events at different precipitation thresholds over the two study basins. Categorical statistics include probability of detection (POD), false alarm ratio (FAR) and the frequency bias index (FBI). 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. The FBI (= (a + b)/(a + c)) is the ratio of the number of estimated to observed rain events; it can indicate whether there is a tendency to underestimate (FBI < 1) or overestimate (FBI > 1) rainy events. It ranges from 0 to infinity; 1 being the perfect score.

The second objective was to evaluate the runoff predictions. A basin-scale analysis was carried out to compare observed discharge values at the outlet of each study basin with the runoff calculated by the hydrological model. Due to current limitations of the routing component (e.g. coarse spatial resolution and constant flow velocity), simulated basin-scale runoff was calculated by area averaging the grid-runoff depths over the basin. The capacity of the model to predict the spatial extent of the flood was evaluated using the MODIS-based inundation maps provided by the Dartmouth Flood Observatory. Finally, the ability of the model to detect historical large flood events (provided by DFO) during April 2007–July 2008 (16 months) was tested. From the DFO flood archive we used coordinates of the flood centroid, together with the beginning and ending dates of flood events, to evaluate the predictive capacity of the hydrological model. We considered a flood event successfully detected if within 1 degree distance from the flood centroid and within a ±1-day temporal window there exists at least one model grid with simulated runoff higher than 50 mm.

RESULTS

Comparison of the precipitation estimates

Figure 2 shows the cumulative precipitation estimates for the basins over the study period. In Flint basin (Fig. 2(a)), TMPA produces more precipitation compared to RADG (indicated by the upwards deviation of the TMPA precipitation accumulations) with almost linearly increasing positive bias accumulation throughout the study period (dash-dot line). Overall bias between TMPA and RADG estimates during the study period is positive, 13%. In Illinois basin (Fig. 2(b)) TMPA indicates less precipitation compared to RADG with a negative 8% overall bias. Note also that the

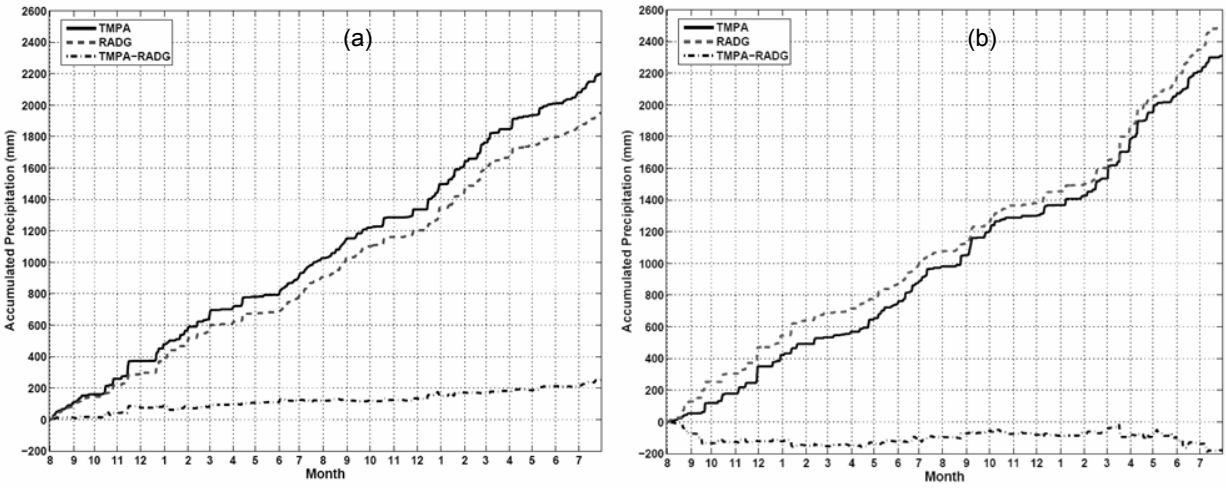

Fig. 2 Cumulative precipitation estimates from TMPA and RADG, for: (a) Flint basin, and (b) Illinois basin. The difference between TMPA and RADG estimates is shown as the dash-dot line.

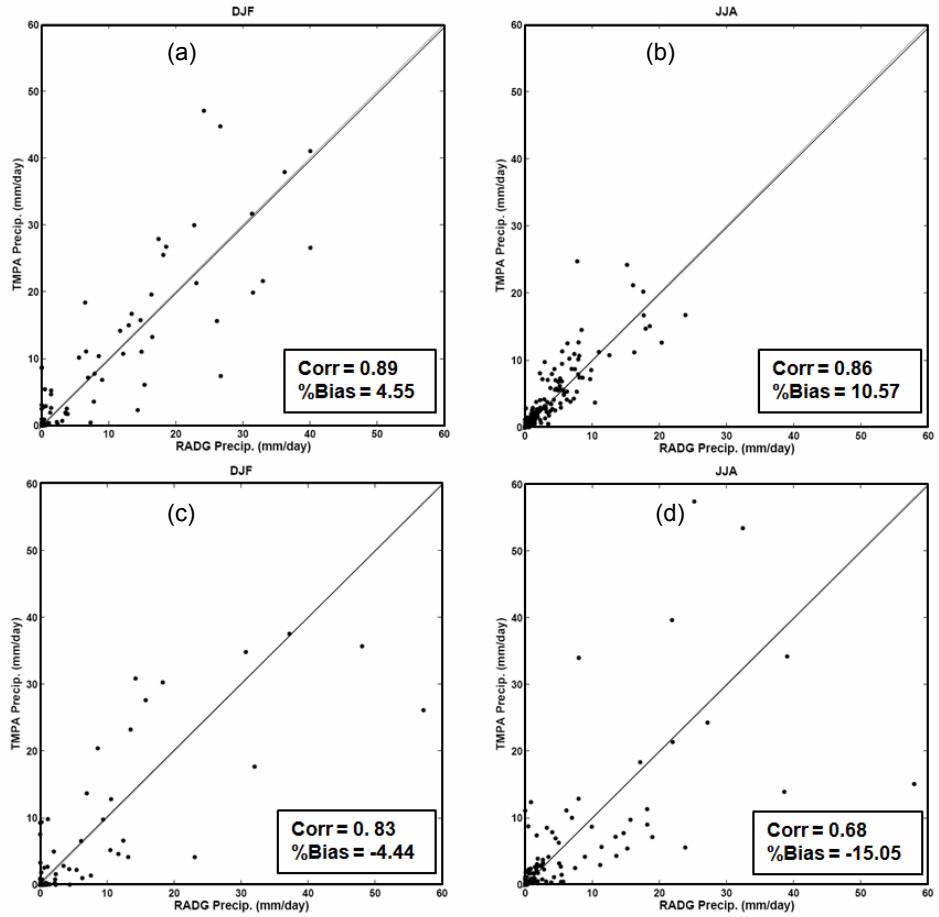

Fig. 3 Scatter plots for winter (DJF) and summer (JJA) seasons for (a,b) Flint, and (c,d) Illinois basins, respectively.

accumulated bias (dash-dot line) magnitude alternates between more negative and less negative values throughout the time period,, indicating the presence of periods with positive and negative bias between TMPA and RADG. This changing behaviour of satellite rainfall estimates in study basins might indicate regional trend in precipitation estimated by the TMPA algorithm. Scatter plots in Fig. 3 facilitate a visual comparison of the magnitude correspondence of daily RADG and TMPA precipitation. The diagonal line indicates a "perfect" correspondence. The study period is divided into cold (DJF) and warm (JJA) seasons to examine the seasonal behaviour. Comparison of Fig. 3(a) and (b) indicates that TMPA overestimates the rainfall in Flint basin regardless of the season. The agreement between RADG and TMPA estimates is more pronounced during the cold season (correlation coefficient is 0.89 and %Bias is 4.55) compared to the warm season (correlation coefficient is 0.86 and %Bias is 10.57). In Illinois basin (Fig. 3(c) and (d)), TMPA tends to underestimate precipitation compared to RADG regardless of the season. This behaviour is more pronounced in the warm season as indicated by negative 15% bias compared to negative 4.4% bias in the cold season. The correlation coefficient significantly deteriorates in summer season (0.68) compared to the cold season (0.83), most likely due to local character of convective precipitation in summer. These findings agree with Tian *et al*. (2007), who compared TMPA and Stage IV radar rainfall estimates over the USA for 3-year seasonal accumulations at 0.25° spatial grids. Tian *et al*. (2007) reported that TMPA overestimates Stage IV rainfall over Flint basin region regardless of the season. Over Illinois basin region their results showed TMPA underestimating Stage IV rainfall in summer. In winter, however, they found both over- and underestimation between TMPA and Stage IV rainfall estimates for the grids within Illinois basin region. Similar to our findings, Tian *et al*. (2007) reported higher (lower) correlations in winter

(summer) between TMPA and Stage IV estimates. In the scatter plots shown in Fig. 3, the large precipitation points along either *x*-axis or *y*-axis are particularly important if these estimates are to be used for flood detection. Points along the *x*-axis (*y*-axis) represent undetected (falsely predicted) large rain events. This behaviour can be seen in Fig. 3(a), (c) and (d) up to 10 mm rainfall. One reason might be that the satellite-based rainfall is based on a snapshot in a 3-h window, while radar estimates are calculated from more frequent estimates. However, for larger rain events RADG and TMPA agree well on the occurrence of the events but with large random errors (these are the points away from both axis lines). Figure 4 shows the categorical evaluation measures calculated for various rain rate thresholds. Figure 4 indicates better rainfall detection correspondence between TMPA and RADG estimates in the larger Flint basin compared to the smaller Illinois basin, as indicated by higher POD and lower FAR values. Focusing on Flint basin, high POD values (~0.9) up to 10 mm rainfall thresholds are followed by a pronounced decrease (0.78) at 20 mm threshold. There is a tendency to increase in FBI and FAR values with increasing threshold indicating that the TMPA tends to increasingly overestimate the number and magnitude of events as rain rate increases in Flint basin. In the Illinois basin, the categorical statistics present similar behaviour as the threshold increases, while providing a general deterioration trend in statistics compared to Flint basin. Deterioration in the statistics towards the higher rainfall thresholds points may undermine flood detection efforts. In summary, this analysis indicates that seasonal and regional considerations, as well as basin size, are important in using satellite-based rainfall estimates as input to the hydrological models. The analysis indicates that in smaller watersheds (such as Illinois) the errors in precipitation estimates may not be reduced by area-averaging as much as in the large basins (i.e. correspondence between TMPA and RADG is somewhat better in the larger basin). This is reflected by the somewhat better bias, correlation coefficient and categorical statistic values obtained for Flint basin compared to Illinois basin. Good values for POD (~0.8–0.9), FAR (~0.2–0.3) and FBI (~1–1.1) obtained for Flint basin indicate that the TMPA product is more promising for use in hydrological predictions in relatively large basins.

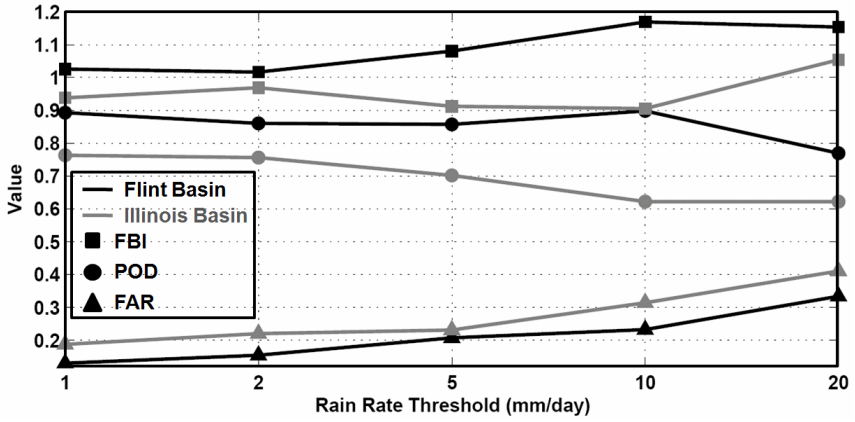

Fig. 4 Categorical evaluation measures for Flint basin (black line) and Illinois basin (grey line).

Evaluation of the flood predictions

Figure 5(a) and (b) facilitates a comparison between the observed discharge at the Flint basin outlet and model simulated runoff, respectively. Figure 5 shows that high simulated runoff events correspond relatively well with the high observed discharge events. However, simulated events are earlier and flashier than the observed events, indicating that improvements are needed in the model to better represent the smoothing and delay mechanisms exerted by the basin on the excess rain water entering the stream before the routing process. However, observed discharge includes the real stream routing process. For this reason there is a magnitude difference between fall. In this study, due to limitations of the routing component, runoff more closely represents the

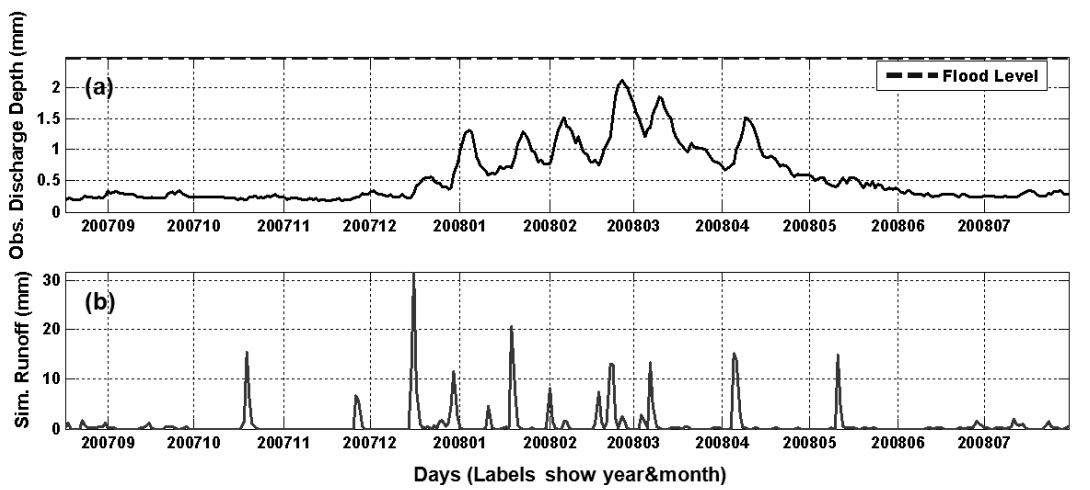

Fig. 5 (a) Observed discharge, and (b) simulated runoff values for Flint basin.

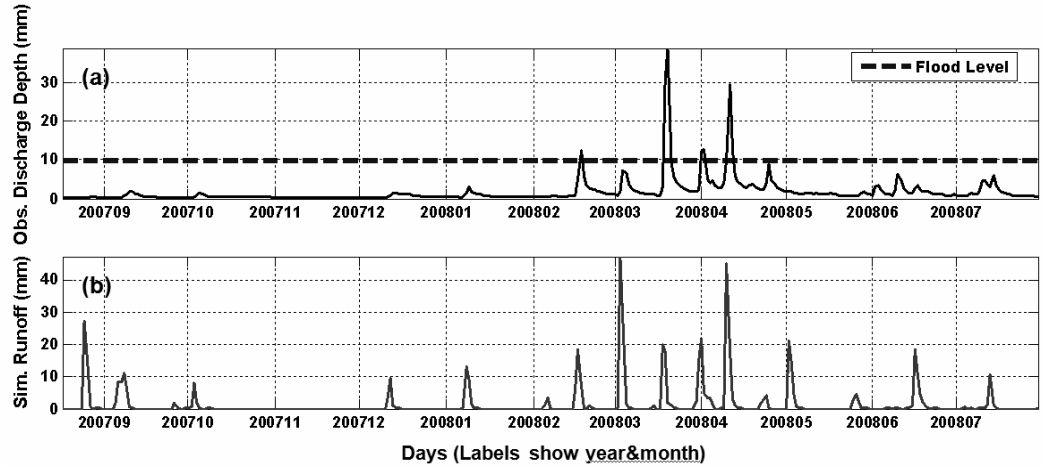

Fig. 6 (a) Observed discharge, and (b) simulated runoff values for Illinois basin.

discharge and runoff values given in Fig. 5. Here we are more interested in the general behaviour of these quantities rather than the magnitude differences.

In addition, observed baseflow and water storage (January–May 2008) in the Flint basin could not be represented by the hydrological model. This is a limitation of the curve number based approach which only estimates the excess precipitation and does not consider groundwater storage. These limitations have already been considered in the development of a new physically-based hydrological model with groundwater representation and an improved routing scheme (Wang *et al*., 2009). Note also that several simulated high runoff events do not correspond to high flow events in the observations (e.g. mid-October and mid-December 2007, mid-May 2008). Comparison of the TMPA and RADG rainfall estimates (not shown) indicated that these events are mostly due to positive bias in the TMPA at these time periods. In the Illinois basin (Fig. 6), observed discharge is flashier and contains a smaller baseflow component. High observed flow events are better matched with high simulated runoff events. Again, the hydrological model lacks the smoothing and delay mechanisms, but this problem is less pronounced for this smaller basin compared to the larger Flint basin. This behaviour is expected because the larger basin (a) introduces more smoothing and delay to convert runoff into discharge, and (b) has more baseflow component. Simulated runoff contains both overestimated events (e.g. late August 2007 and early March 2008) and underestimated events (mid-March 2008) compared to the observations, in part due to the bias in the satellite-based precipitation estimates.

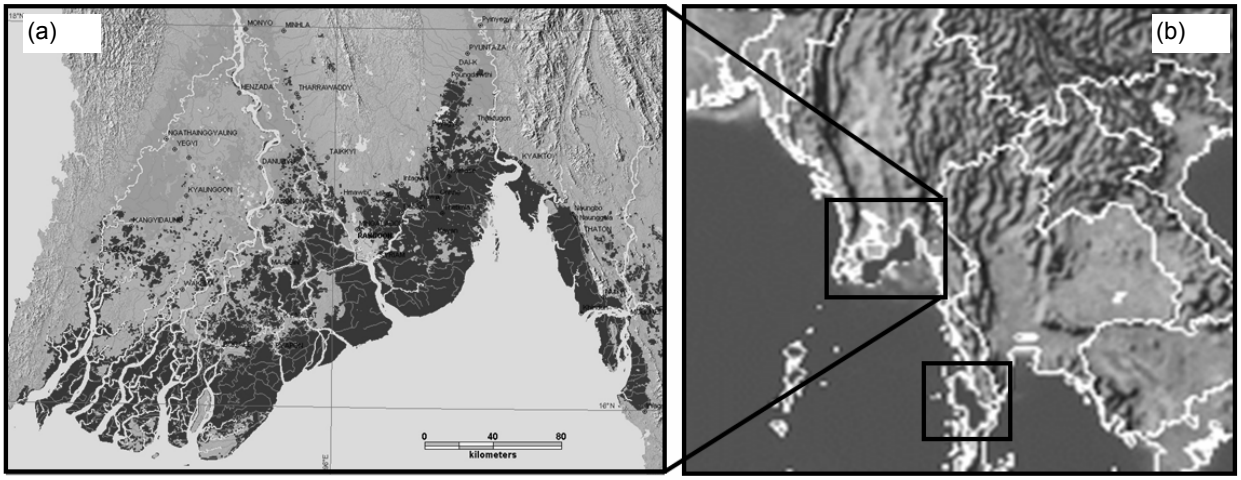

Fig. 7 Spatial extent of the Myanmar flood caused by the Cyclone Nargis on 5 May 2008: (a) MODIS inundation map for 5 May 2008 provided by the Dartmouth Flood Observatory (dark grey regions show inundated land), and (b) severely flooded regions estimated by the hydrological model (dark grey regions inside the boxes) at time 03:00 GMT.

Figure 7 facilitates evaluation of the spatial extent of the flooding produced by the hydrological model. Figure 7(a) shows the MODIS inundation map provided by DFO after-the-fact (2–4 days delay depending on the cloud cover) and Fig. 7(b) shows the flood extent prediction by the hydrological model as soon as the satellite rainfall estimates were available (6–9 hours delay for the real-time TMPA product). It can be seen that the spatial extent of the flood is well simulated by the model. Our analysis including other events showed that the model is capable of estimating the spatial extent during the onset of the floods caused by heavy rainfall; however, in the later stages the estimates deteriorate. This is due to limitations in the routing component (channel hydraulics), which is responsible for moving the water downstream and holding the excess water on the ground.

Figure 8 shows the locations of historical flood events (all markers) during the April 2007–July 2008 period. Detected flood events are shown as a black circles and undetected flood events are shown as white squares. Among 263 flood events, the model was capable of detecting 175 events (67% success rate). This analysis is also valuable in analysing the dependency of the model predictions in geographic locations. The model is capable of detecting flood events located in the tropical zone in Asia. The model also detected a majority of flood events in Australia, South

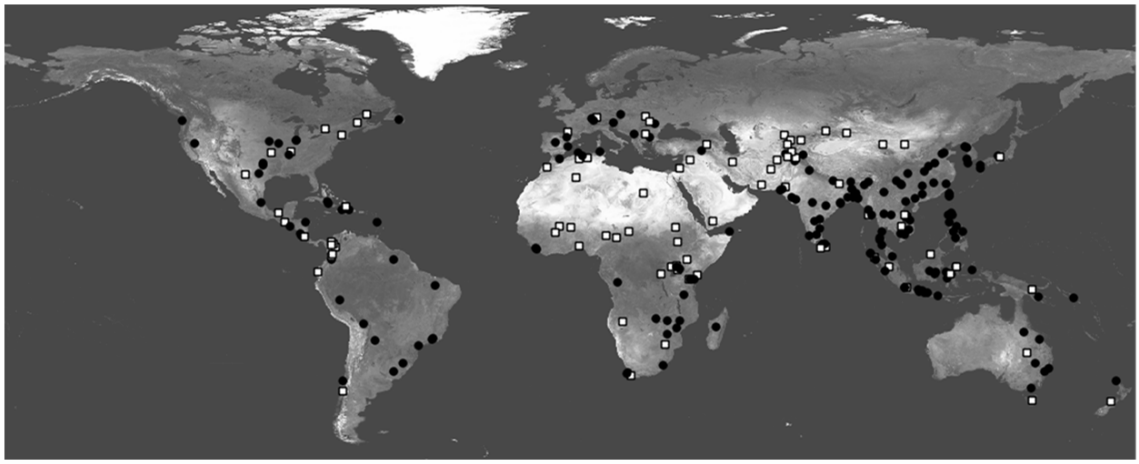

Fig. 8 Locations of the historical flood events compiled by DFO. Flood events detected by the hydrological model are shown as black circles. Undetected flood events are shown as white squares.

Africa, Europe and the Americas. The model seems to be incapable of detecting floods in arid and semi-arid regions (e.g. North Africa, Central Asia), or regions with snow (higher latitudes) and complex topography (Central America, Central Asia). Several factors contribute to these limitations, such as: (a) satellite-based precipitation algorithms have limitations in regions with snow and complex topography; and (b) the hydrological model does not consider snow processes and its coarse resolution may not be adequate to resolve complex topography. This analysis only focused on detection of historical floods. Of course, a natural and complimentary extension would be to analyse the false flood events predicted by the model. This is currently underway, but is further complicated by the fact that the DFO archive most likely does not contain all the flood events that occurred during this time period.

SUMMARY AND CONCLUSIONS

This study evaluated an initial satellite-based near real-time global flood prediction system developed at NASA Goddard (Hong *et al*., 2007) and operationally available at our website (http://trmm.gsfc.nasa.gov/). The first step was to evaluate TMPA-based satellite precipitation estimates (the key input to the system) using the NCEP Stage IV radar rainfall product. Evaluation over two basins in the southeastern USA indicated that basin size and seasonal considerations are important in using satellite-based rainfall estimates as input to the hydrological models. Generally better quantitative statistics (bias and correlation coefficient) and categorical statistics (POD, FAR, FBI) were obtained for the larger Flint basin. Relatively good values for these statistics indicated that the TMPA product is promising for use in hydrological predictions in basins of large size. The second step was to evaluate the flood predictions calculated by the hydrological model using observed streamflow data, MODIS-based inundation maps and historical flood archive. The analysis indicated that the model was able to simulate the onset of flood events produced by heavy rainfall; however, the prediction deteriorated in the later stages of the flood events. This result points out the need for an improved routing component, specifically to reflect flow smoothing and delay mechanisms exerted by the natural river basin. The model also showed dependency by the geographical region, with best performances in tropical regions of East Asia. A new hydrological model, with an improved physical representation and routing component is currently under development and will likely lead to improved validation results.

Acknowledgements This research is carried out with support from NASA's Applied Sciences Programme (Stephen Ambrose) and PMM (Ramesh Kakar) of NASA Headquarters.

REFERENCES

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

Hong, Y., Hsu, K. L., Gao, X. & Sorooshian, S. (2004) Precipitation estimation from remotely sensed imagery using artificial neural network – cloud classification system (PERSIANN-CCS). *J. Appl. Met.* **43**(12), 1834–1853.

Huffman, G. J. & co-authors (2007) The TRMM Multisatellite Precipitation Analysis (TMPA): quasi-global, multiyear, combined sensor precipitation estimates at fine scales. *J. Hydromet.* **8**, 38–55.

Huffman, G. J., Adler, R. F., Bolvin, D. T. & Nelkin, E. J. (2009) The TRMM Multi-satellite Precipitation Analysis (TMPA). In: *Satellite Applications for Surface Hydrology* (ed. by F. Hossain & M. Gebremichael), Springer Verlag (in revision).

Lin, Y. & Mitchell, K. E. (2005) The NCEP stage II/IV hourly precipitation analyses: development and applications. Preprints, *19th Conf. on Hydrology* (San Diego, California, USA), Am. Met. Soc. 1.2.

Seo, D. J. (1998) Real-time estimation of precipitation fields using radar precipitation and rain gauge data. *J. Hydrol.* **208**, 37–52.

Sorooshian, S., Hsu, K.-L., Gao, X., Gupta, H., Imam, B. & Braithwaite, D. (2000) Evaluation of PERSIANN system satellite-based estimates of tropical rainfall. *Bull. Am. Met. Soc.* **81**, 2035–2046.

Stellman, K. M., Fuelberg, H. E., Garza, R. & Mullusky, M. (2001) An examination of radar and rain gauge-derived mean areal precipitation over Georgia watersheds. *Weather Forecasting,* **16**, 133–144.

Su, F., Hong, Y. & Lettenmaier, D. P. (2008) Evaluation of TRMM Multisatellite Precipitation Analysis (TMPA) and its utility in hydrologic prediction in the La Plata Basin. *J. Hydromet.* **9**, 622–640.

Tian, Y., Peters-Lidard, C. D., Choudhury, B. J. & Garcia, M. (2007) Multitemporal analysis of TRMM-based satellite precipitation products for land data assimilation applications. *J. Hydromet.* **8**, 1165–1183.

Wang, J. & co-authors (2009) Implementation of a distributed hydrologic model for global flood guidance: TRMM-based OU-Goddard Flood Modeling System (to be submitted).

Yilmaz, K. K., Hogue, T. S., Hsu, K.-L., Sorooshian, S., Gupta, H. V. & Wagener, T. (2005) Intercomparison of rain gauge, radar, and statellite-based precipitation estimates with emphasis on hydrologic forecasting. *J. Hydromet.* **6**, 497–517.