

# Hydrometeorological Analysis and Remote Sensing of Extremes: Was the July 2012 Beijing Flood Event Detectable and Predictable by Global Satellite Observing and Global Weather Modeling Systems?

YU ZHANG,\* YANG HONG,\* XUGUANG WANG,<sup>+</sup> JONATHAN J. GOURLEY,<sup>#</sup> XIANWU XUE,<sup>@</sup>  
MANABENDRA SAHARIA,<sup>@</sup> GUANGHENG NI,<sup>&</sup> GAILI WANG,\*\* YONG HUANG,<sup>++</sup>  
SHENG CHEN,<sup>@</sup> AND GUOQIANG TANG<sup>&</sup>

*\* School of Civil Engineering and Environmental Science, and Advanced Radar Research Center, and Center for Analysis and Prediction of Storms, University of Oklahoma, Norman, Oklahoma*

*<sup>+</sup> School of Meteorology, and Center for Analysis and Prediction of Storms, University of Oklahoma, Norman, Oklahoma*

*<sup>#</sup> NOAA/National Severe Storms Laboratory, National Weather Center, Norman, Oklahoma*

*<sup>@</sup> School of Civil Engineering and Environmental Science, and Advanced Radar Research Center, University of Oklahoma, Norman, Oklahoma*

*<sup>&</sup> Department of Hydraulic Engineering, Tsinghua University, Beijing, China*

*\*\* State Key Laboratory of Severe Weather, Chinese Academy of Meteorological Science, Beijing, China*

*<sup>++</sup> Anhui Meteorological Bureau, Hefei, China*

(Manuscript received 24 February 2014, in final form 3 October 2014)

## ABSTRACT

Prediction, and thus preparedness, in advance of flood events is crucial for proactively reducing their impacts. In the summer of 2012, Beijing, China, experienced extreme rainfall and flooding that caused 79 fatalities and economic losses of \$1.6 billion. Using rain gauge networks as a benchmark, this study investigated the detectability and predictability of the 2012 Beijing event via the Global Hydrological Prediction System (GHPS), forced by the NASA Tropical Rainfall Measuring Mission (TRMM) Multisatellite Precipitation Analysis at near-real time and by the deterministic and ensemble precipitation forecast products from the NOAA Global Forecast System (GFS) at several lead times. The results indicate that the disastrous flooding event was detectable by the satellite-based global precipitation observing system and predictable by the GHPS forced by the GFS 4 days in advance. However, the GFS demonstrated inconsistencies from run to run, limiting the confidence in predicting the extreme event. The GFS ensemble precipitation forecast products from NOAA for streamflow forecasts provided additional information useful for estimating the probability of the extreme event. Given the global availability of satellite-based precipitation in near-real time and GFS precipitation forecast products at varying lead times, this study demonstrates the opportunities and challenges that exist for an integrated application of GHPS. This system is particularly useful for the vast ungauged regions of the globe.

## 1. Introduction

Flooding, considered as one of the most hazardous disasters in both rural and urban areas, accounts for about one-third of all geophysical hazards globally (Adhikari et al. 2010; Smith and Ward 1998). Urban areas are more vulnerable to floods and their associated damages than rural areas because of their high population densities and intensively developed infrastructure. Urban

flooding affects structures such as buildings, bridges, and roadways and may also induce severe waterborne diseases. On 21 July 2012, the capital of China, Beijing, and its surrounding areas experienced extreme rainfall and flooding. The storm lasted for around 16 h and the rain rate reached as high as 215 mm day<sup>-1</sup> in the urban areas. It was reported as the heaviest storm event since 1951, and the return periods for flooding were estimated at 60 years in Beijing and 100 years in the surrounding Fangshan suburban area. It inundated roadways, bridges, and sewage systems, causing houses to collapse, damage to cars, and even debris flows in Fangshan. Overall, the flooding event resulted in 79 fatalities and around \$1.6 billion in damages.

---

*Corresponding author address:* Dr. Yang Hong, National Weather Center, ARRC Suite 4610, 120 David L. Boren Blvd., Norman, OK 73072.

E-mail: yanghong@ou.edu

In the same year, the Gelendzhik, Novorossiysk, and Krymsk districts in Russia were affected by the Kuban flood in July, and 171 people were killed. Three months later, in October, New York City experienced Hurricane Sandy, which flooded the streets, subways, and tunnels and cut electricity in and around the city. From December 2010 to January 2011, the third biggest city in Australia, Brisbane, was inundated by floods from several separate storm events. As mentioned by Adhikari et al. (2010, p. 406), “the International Flood Network indicates that from 1995 to 2004, natural disasters caused 471,000 fatalities worldwide and economic losses totaling approximately \$49 billion USD, out of which approximately 94,000 (20%) of the fatalities and \$16 billion USD (33%) of the economic damages were attributed to floods alone.” In the coming decades, as urban population grows rapidly, especially in fast-growing developing countries, urban areas will likely become increasingly vulnerable to hydrometeorological extremes.

The increasing adverse worldwide impact from floods indicates this is not only a regional or national-level issue but also a global problem, which motivates the development of a global flood detection and prediction system in coordination with research institutions and government decision makers. Currently, several satellite remote sensing–based, flood-monitoring systems exist at global scales and provide predictions in near–real time (Brakenridge et al. 2007; Hong et al. 2007; Westerhoff et al. 2013; Wu et al. 2012; Yilmaz et al. 2010). Improvement of global flood early warning systems is appealing to decision makers as it provides forecasts several days in advance for better planning and response to emerging disasters. The traditional approach to forecasting streamflow at the outlet of a basin often depends on observed rainfall for observations of flooding from an upstream stream gauge. In this case, the lead time is often limited by the catchment concentration time (Bartholmes and Todini 2005). To extend the hydrological forecast horizon, numerical weather prediction (NWP) products (e.g., temperature and precipitation) can be coupled with hydrological rainfall–runoff models, which is of great importance to rivers without upstream discharge observations and to those smaller, ungauged rivers with shorter response times (Hopson and Webster 2010).

Ensemble forecast products from NWP models are becoming an increasingly popular option for hydrologic modeling and quantifying the uncertainties in the forecasts. NWP-based hydrologic ensembles provide an attractive option for flood forecasting systems by estimating the probability that an extreme flooding event will occur (Cloke and Pappenberger 2009). In particular, the Hydrologic Ensemble Prediction Experiment (HEPEX;

Schaake et al. 2007; <http://hepex.irstea.fr/about-hepex/>), with its mission “to demonstrate the added value of hydrological ensemble predictions (HEPS) for emergency management and water resources sectors to make decisions that have important consequences for economy, public health and safety,” has developed a community with experts from meteorology to hydrology in order to improve ensemble forecasts (Bradley et al. 2003, 2004; Brown et al. 2010, 2012; Demargne et al. 2009, 2014; Gneiting et al. 2007; Pappenberger et al. 2008; Seo et al. 2006; Zappa et al. 2013). Recently, a review paper (Cloke and Pappenberger 2009) showed the potential of using ensemble streamflow forecasts to further improve early warning systems.

This study evaluates a prototype of a real-time Global Hydrological Prediction System (GHPS), which is driven by the Tropical Rainfall Measuring Mission (TRMM) Multisatellite Precipitation Analysis (TMPA) and the National Oceanic and Atmospheric Administration’s (NOAA) Global Forecast System (GFS) deterministic and ensemble precipitation forecasts. We intend to address the following questions: 1) Was the July 2012 Beijing flood event detectable and predictable by global satellite observing and weather modeling systems in its initial stage without site-specific calibration? and 2) How much “added value” does the ensemble streamflow forecast contribute to the hydrological prediction in the probabilistic domain for this specific case study?

This study is organized as follows. In section 2, the core part of GHPS, a distributed hydrological model and its setup, are described. Then, the study region and datasets applied for this particular case study are introduced in section 3. In section 4, the hydrologic predictions conditioned on forcing from satellite remote sensing and NWP model forecasts are assessed in both deterministic and probabilistic domains. Finally, the results are summarized in section 5 along with concluding remarks.

## 2. Global Hydrological Prediction System

The GHPS (Fig. 1), with the Coupled Routing and Excess Storage (CREST; Wang et al. 2011) distributed hydrological model as its core, is applied to investigate the detectability and predictability of flooding using precipitation estimates from TMPA and precipitation forecasts from the GFS. The CREST model is modified from the Variable Infiltration Capacity (VIC) model and concepts originally represented in the Xinanjiang model (Liang et al. 1994; Nijssen et al. 1997) and has added a distributed grid-to-grid routing scheme. The CREST model is currently running within the Near Realtime Global Hydrological Simulation and Flood Monitoring Demonstration System (<http://eos.ou.edu>)

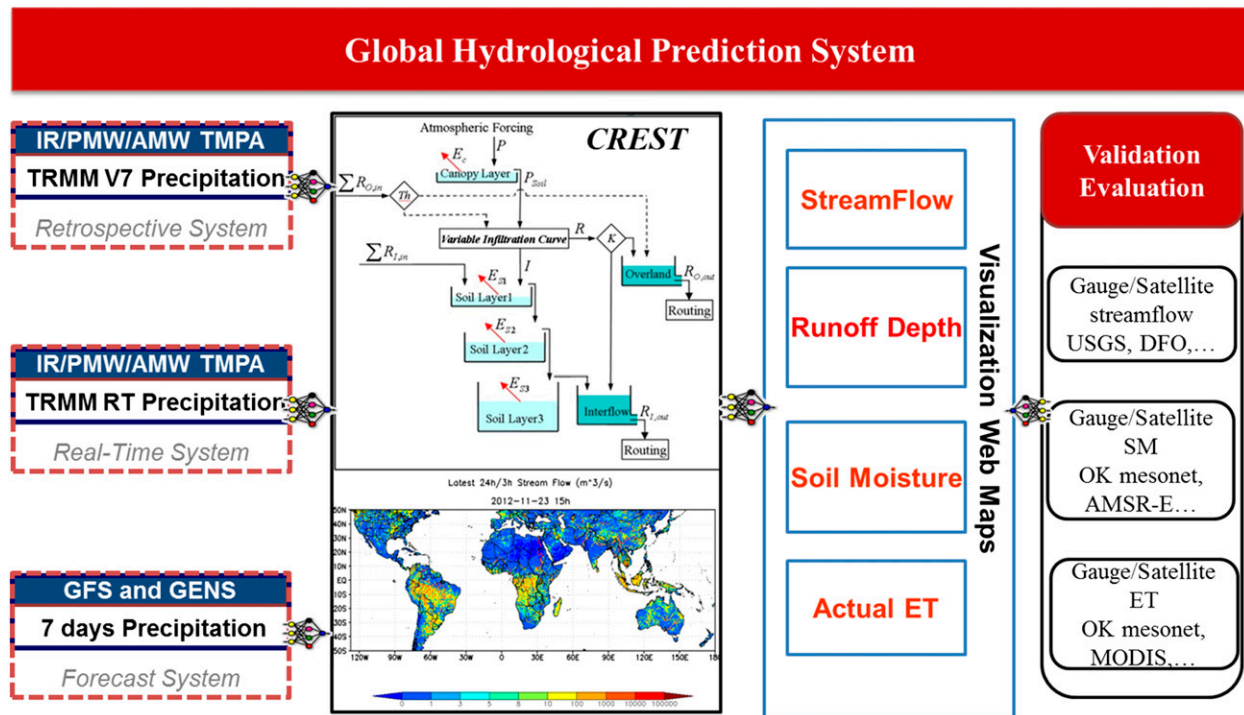


FIG. 1. Structure of the GHPS.

at the University of Oklahoma. Presently, it is driven by the TRMM 3B42 real-time product (RT; Huffman et al. 2007, 2010). The retrospective runs and forecast runs are driven by the post-real-time research product from TRMM 3B42, version 7 (V7; Huffman et al. 2007, 2010), and NOAA GFS precipitation forecasts (Han and Pan 2011; Kanamitsu et al. 1991; Wang 2010; Wang et al. 2013; Yang et al. 2006), respectively. For detailed information of the forcing data, please refer to section 3 and the corresponding references.

In GHPS, the CREST model is set up at  $1/8^\circ$  based on a digital elevation model (DEM) with quasi-global coverage from  $50^\circ\text{N}$  to  $50^\circ\text{S}$ , providing a near-real-time runoff and streamflow simulation every 3 h. Rainfall forcing comes from TRMM RT (for real time) and from TRMM V7 (for retrospective) hydrological simulations since 1998. Precipitation forecasts and subsequent flood forecasts are initialized at 0000 UTC every day with lead times up to 180 h at each  $1/8^\circ$  grid cell. The model parameters are estimated a priori from Earth's physical measurements [for detailed information about the parameter estimation, please refer to Wang et al. (2011) and Wu et al. (2012)]. The physical parameters such as soil-saturated hydraulic conductivity  $K_{\text{sat}}$  and mean water capacity  $W_M$  in the CREST model can be estimated based on the soil type, land cover, and DEM data. The soil states in the CREST model have been warmed up (initialized) using more than 10 years of TRMM V7 rainfall forcing.

The CREST model has been evaluated and implemented at both global and regional scales (Khan et al. 2011a,b; Wu et al. 2012; Yilmaz et al. 2010), proving its high cost effectiveness in hydrological prediction.

Wu et al. (2012) applied the CREST model, but forced with TRMM, version 6 (V6; the gauge-corrected research product) to run a retrospective streamflow simulation from  $50^\circ\text{N}$  to  $50^\circ\text{S}$  during the period 1998–2010. In general, the results show that the probability of detection (POD) is around 0.70 for floods with durations longer than 3 days in rivers that are not regulated by dams. The generally positive results indicate the potential value of this system forced by TRMM for global flood detection. However, Wu et al. (2012) did not specifically address or investigate the extreme or rare events such as the Beijing event investigated in this paper. Therefore, this paper is the first assessment of prediction skill from GHPS in a local setting for an extreme event. In this study, the updated versions of TRMM data, both TRMM RT (real-time product) and TRMM V7 (the post-real-time rain gauge-corrected research product) are applied for flood detection. Considering the improvement in satellite precipitation estimates from the TRMM V6 to V7 product, the GHPS is expected to have better flood prediction skill. In addition to streamflow and runoff depth, the GHPS can also provide gridded soil moisture and actual evapotranspiration rates (AETs) at  $1/8^\circ$  spatial resolution, as shown by the third column in Fig. 1.

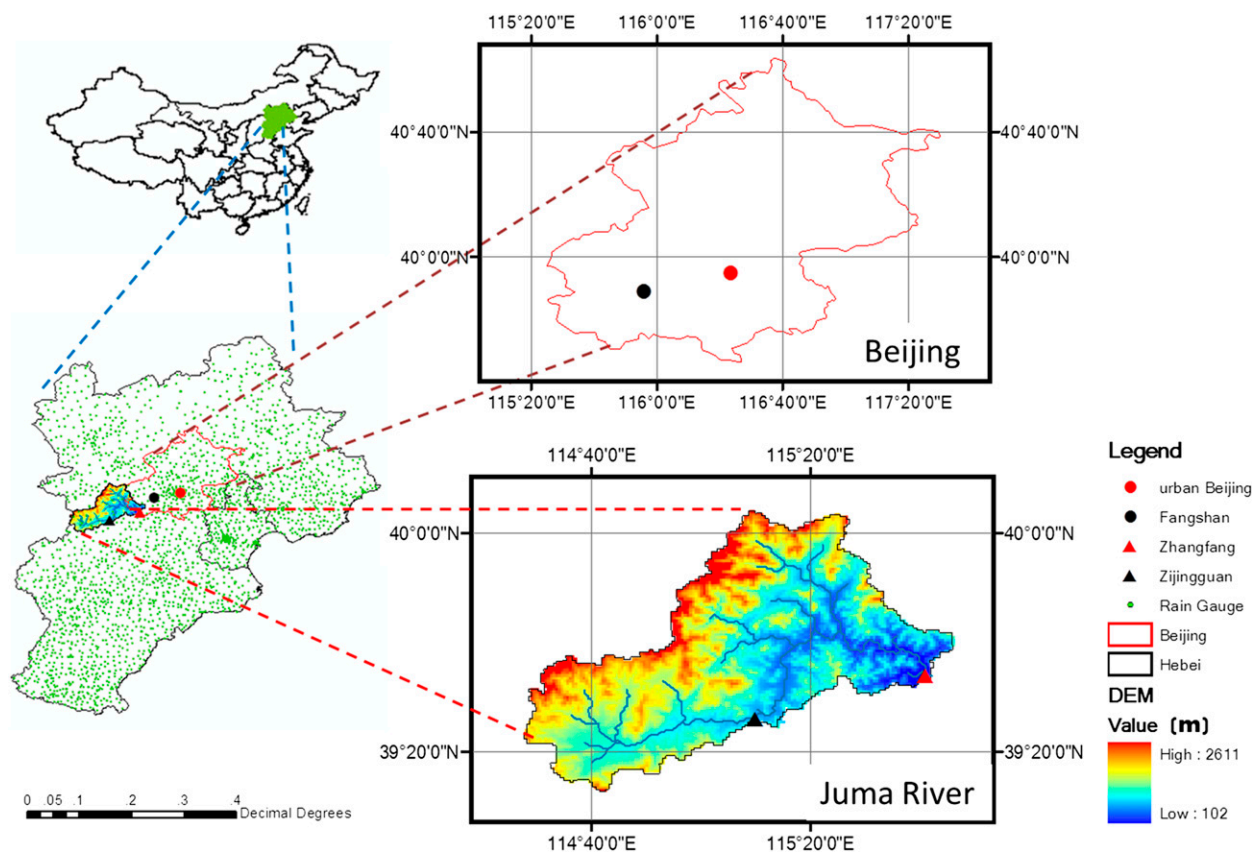


FIG. 2. Study region over China. The zoomed-in insets show urban Beijing and the upstream Juma River basin and its topography.

In this study, soil states in the global CREST model are initialized by running the model using TRMM RT rainfall forcing from 1 July 2012 until the initial time of each experiment. Then the model is forced by rain gauge observations, TRMM RT, TRMM V7, and both GFS deterministic and ensemble precipitation forecasts at different initializations (with different lead times) to simulate the hydrological predictions of surface runoff in urban areas and streamflow in the watersheds. Although the CREST model includes a parameter describing the degree of imperviousness of the surface, which is quite distinct in urban regions, the model physics do not explicitly account for evapotranspiration, surface runoff generation, routing, and drainage processes that are specific to the urban environment. A detailed discussion regarding the detectability and predictability of surface runoff depths and streamflows using the GHPS, even with the simplified natural environment assumption, will be discussed in [section 4](#).

### 3. Research region and input data

For this case study, Beijing and its upstream Juma River basin are selected as the research region, as shown

in [Fig. 2](#). Beijing is located in the northern part of China and is surrounded by Hebei Province. It is the most densely populated metropolis in the world. The dense population of Beijing makes it vulnerable to impacts of rainfall and flood extremes, which often lead to huge economic loss and fatalities.

Four precipitation products (see [Table 1](#)) are evaluated in and around Beijing on 21 July 2012 using high-density rain gauge observations with hourly temporal resolution from 0300 UTC 19 July to 1200 UTC 22 July 2012 ([Fig. 2](#)). There are 2041 rain gauge stations in total within Hebei Province and 231 within the city of Beijing. The rain gauge data are interpolated onto a  $1/4^\circ$ -resolution grid using kriging and are accumulated into 3-hourly rainfall accumulations in order to facilitate comparison with TMPA products. The TMPA near-real-time 3B42 RT uses a combination of active and passive microwave and infrared measurements from TRMM and other satellites ([Huffman et al. 2007](#)) to estimate precipitation. The TMPA post-real-time 3B42 V7 product adjusts the rainfall accumulation using monthly rain gauge accumulations. Both 3B42 RT and 3B42 V7 products are quasi-global, with coverage from  $50^\circ\text{N}$  to  $50^\circ\text{S}$  latitude at a spatial resolution of  $1/4^\circ$  and temporal resolution of 3 h.



TABLE 1. Summary of characteristics of the precipitation products used in the study.

Rainfall products	Spatial resolution	Time interval	Lead time (h)	References
Rain gauge observation	Interpolated onto $1/4^\circ$	Hourly, then accumulated to 3 hourly	—	—
TRMM RT	$1/4^\circ$	3 hourly	—	Huffman et al. (2007)
TRMM V7	$1/4^\circ$	3 hourly	—	Huffman et al. (2007)
Deterministic GFS precipitation	$1/4^\circ$	3 hourly	180	Wang (2010); Wang et al. (2013)
Ensemble GFS precipitation	$1/4^\circ$	3 hourly	168	Wang (2010); Wang et al. (2013)

The deterministic model forecast (GFS; [www.emc.ncep.noaa.gov/index.php?branch=GFS](http://www.emc.ncep.noaa.gov/index.php?branch=GFS)) and the global ensemble model forecast (GENS) from NOAA were used to drive the global hydrological forecasts. Please refer to Wang (2010) and Wang et al. (2013) for the details of the system and the algorithm. The global ensemble model forecasts were run in near-real time by the NOAA/Earth System Research Laboratory. The forecasts were initialized by the hybrid ensemble-variational data assimilation system developed based on NOAA–National Centers for Environmental Prediction (NCEP) operational data assimilation (Wang 2010; Wang et al. 2013). The GFS and GENS 20-member forecasts were initialized four times per day (0000, 0600, 1200, and 1800 UTC). The forecasts were produced at 3-hourly intervals up to 180 h of lead time for the GFS and 168 h for GENS. The spatial resolution of the forecasts was  $1/2^\circ$ . In this study, only the GFS and GENS products initialized at 0000 UTC were applied to drive the hydrological forecasts. Both the deterministic and ensemble GFS members were interpolated to  $1/4^\circ$  in order to match the spatial resolution of the TRMM estimates.

## 4. Results and discussion

### a. Rainfall evaluation

Figure 3 shows the total precipitation accumulation (mm) on 21 July 2012 over Hebei Province (dark outline), which contains the Beijing region (white outline), based on rain gauges (Fig. 3a); TRMM V7 (Fig. 3b); TRMM RT (Fig. 3c); and GFS and GENS forecasts, which initialized from 4 days to 1 day in advance of the event (Figs. 3d–k). Although TRMM V7 and RT slightly underestimate the daily accumulated precipitation amounts in the center of the field compared to the gauge observations, the main characteristics of TRMM precipitation products capture the observed precipitation patterns well. The GFS daily precipitation accumulations and GENS daily accumulation mean with different lead times resemble the general patterns of the 21 July event, but they have limited spatial variability because of their coarse spatial resolution. Both GFS and GENS

means underestimate the daily accumulated precipitation amounts against gauge observation for the different lead times. In particular, the GENS mean has substantial underestimation.

Figures 4–7 (left) show the rainfall accumulation time series from rain gauges, TRMM V7, TRMM RT, GFS, and GENS precipitation forecasts, initialized at different dates for different locations. GFS forecasts indicate an impending storm event over these regions 4 days in advance. As shown in Figs. 4 and 5 (left), the GFS exhibits strong run-to-run inconsistencies with significant underforecasts of precipitation according to the gauge observations at 3- and 1-day lead time, but it performs much better at 4- and 2-days lead time. The lag in the GFS precipitation forecast 2 days in advance at urban Beijing (Fig. 4e) and Fangshan (Fig. 5e) was only 6–9 h following rainfall as observed by rain gauges. In contrast, for the forecast just 1 day prior to the event, the timing of the peak rainfall has improved, but there is significant underestimation, with errors similar to those associated with the forecasts produced 3 days prior to the event.

The performance of TRMM V7 and RT are in agreement with one another throughout the 21 July event at urban Beijing and Fangshan (Figs. 4, 5); both products capture the timing of the rainfall peak quite well but underestimate the rainfall magnitude compared to the gauge observations. Similar to Beijing and Fangshan, the performance of TRMM V7 and RT are in agreement throughout the 21 July event at Zhangfang and Zijinguan according to gauge observations (Figs. 6, 7); however, both TRMM V7 and RT products estimate less than 30% of the observed rainfall at both locations.

In summary, although the GFS underforecasts rainfall amounts and had a lag of approximately 6–9 h in reaching the maximum rainfall rates, the model provided informative prognostic skill up to 4 days in advance of the 21 July Beijing flooding event (e.g., Figs. 4a, 5a). Because GFS and GENS show almost no skill at lead times from 5 to 7 days, the corresponding runoff and streamflow simulation plots at these long lead times have been omitted.

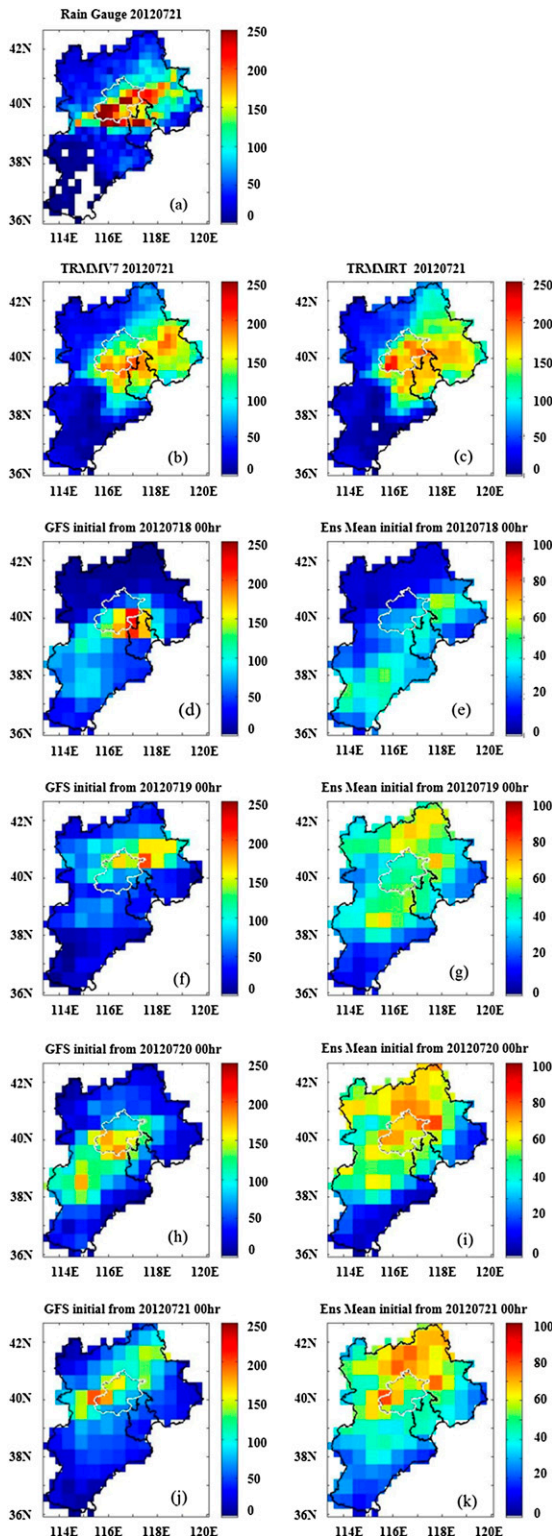


FIG. 3. Daily precipitation accumulation (mm) on 21 Jul 2012 from (a) rain gauge stations; (b) TRMM V7; (c) TRMM RT; GFS initialized at 0000 UTC (d) 18 Jul, (f) 19 Jul, (h) 20 Jul, and (j) 21 Jul 2012; and GENS initialized at 0000 UTC (e) 18 Jul, (g) 19 Jul, (i) 20 Jul, and (k) 21 Jul 2012.

### b. Deterministic hydrological evaluation

Similar to rainfall evaluation, Figs. 4 (right) show the temporal evolution of GHPS-simulated surface runoff at urban Beijing and Fangshan and then simulated discharge at Zhangfang and Zijingguan, the latter two of which are located on the Juma River upstream of Beijing (Fig. 2). For urban Beijing and Fangshan, where the flood peaks were not recorded, the rain gauge–forced simulations are taken as the reference. In general, the simulated surface runoff and streamflows are underestimated when using the GFS forcing compared to the rain gauge–forced results on 21 July 2012 at all four locations. Although the GFS performed reasonably well at a 4-day lead time, the GFS-forced runoff peak at urban Beijing reached about 60% of the rain gauge–forced runoff peak, and there was also a slight delay in peak timing (Fig. 4b). For suburban Beijing–Fangshan, at the same lead time of 4 days, the GFS-forced simulations matched well with gauge-forced peak runoff, but with around 6–9 h delay in the timing of the peak (Fig. 5b). This indicates that the GHPS can potentially provide an early warning of up to 4 days in advance when forced with GFS rainfall forecasts, but the performance does not exhibit run-to-run consistency as the lead time decreases. Similarly, Hlavcova et al. (2006) concluded that there is considerable forecast variability with deterministic forecasts such that a clear signal with 4 days of lead time may not be sufficient for taking preventative actions.

At Zhangfang, the peak of rain gauge–forced simulated streamflow is in agreement with the gauge-reported peak [red asterisk in Fig. 6 (right)], although there is an approximate 6-h timing offset. The potential of GHPS when forced by rain gauges is demonstrated, but the lead time of flooding is limited by the basin response time to observed rainfall. GFS-forced streamflow simulations with 4 days of lead time at Zhangfang show a relatively accurate forecast of peak timing compared to the reported peak, but with obvious underestimation in magnitude. For the Zijingguan gauge station, streamflow simulations conditioned on different forcings (i.e., rain gauge observations, TRMM V7, TRMM RT, and GFS) all underestimated the peak flows compared to the reported peak [Fig. 7 (right)]. Interestingly, at 2 days of lead time, GFS-forced streamflow forecasts are more accurate than those from TRMM estimates in terms of magnitude, but the timing of the peak is lagged by about 12 h. At 1 day of lead time, GFS-forced simulations also show advantages regarding both timing and peak magnitude relative to TRMM forcing.

To assess the applicability of the flood detection with the GHPS to ungauged basins over the globe, we used

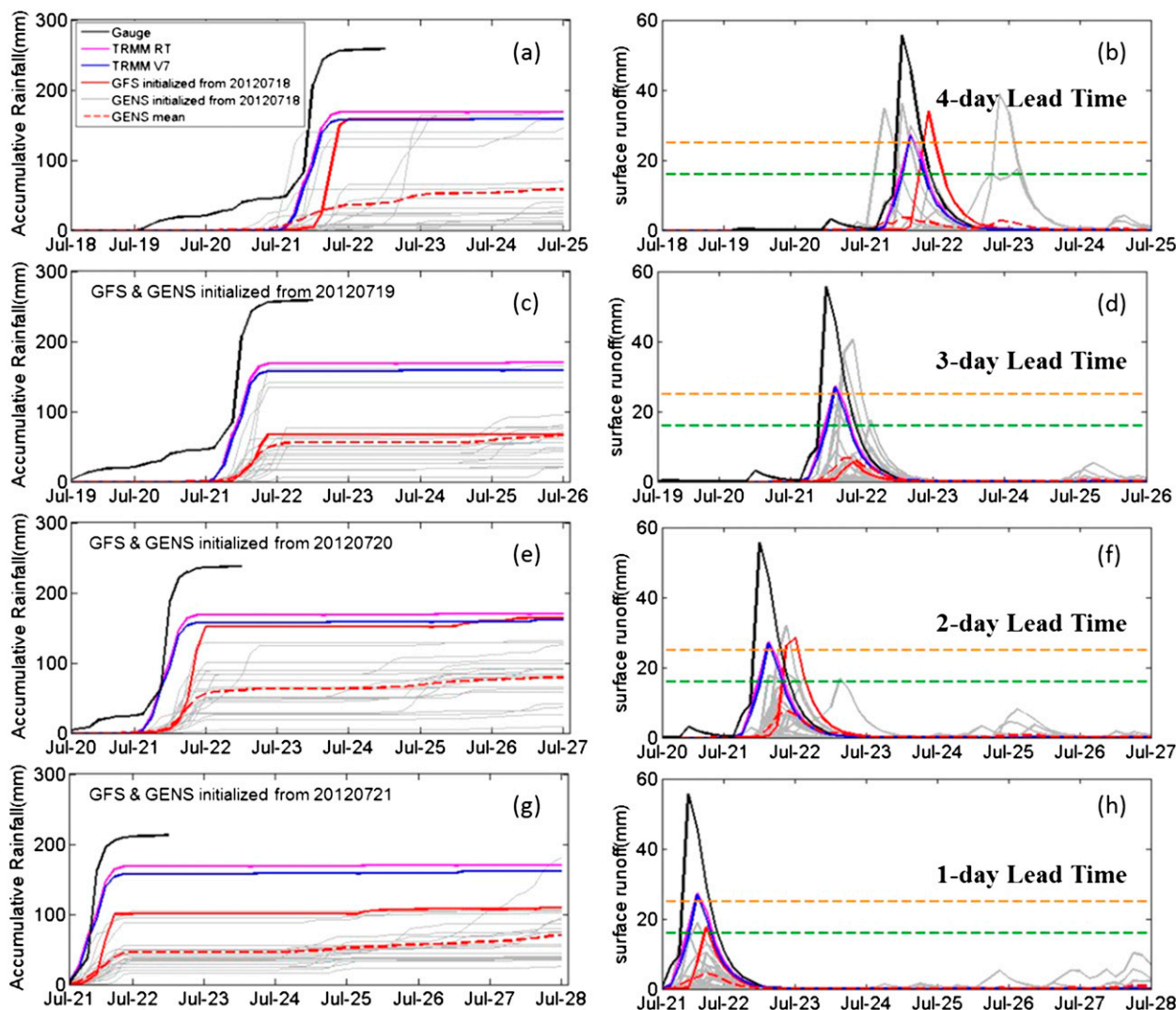


FIG. 4. (left) Accumulated rainfall initialized from different dates at 0000 UTC at urban Beijing (red dot in Fig. 2) from different products: gauge observation, TRMM RT, TRMM V7, GFS, GENS, and GENS mean. (right) GHPS predicted surface runoff initialized from different dates at 0000 UTC forced by different precipitation products: gauge observation, TRMM RT, TRMM V7, GFS, and GENS. From top to bottom, lead times are from 4 days to 1 day. Orange dashed line indicates 50-yr return period surface runoff–streamflow threshold. Green dashed line indicates 20-yr return period surface runoff–streamflow threshold.

a historical database of TRMM RT rainfall estimates. The global CREST model was driven by TRMM RT for its archive of 10 years to yield a retrospective hydrological simulation from 2002 to 2011 at each grid point. Then, the annual peaks were extracted and used to estimate the parameters of a log Pearson type III distribution. This enabled us to estimate the modeled surface runoff and streamflow corresponding to return periods of 50 (orange dashed lines in Figs. 4–7) and 20 years (green dashed lines in Figs. 4–7). This technique enables the GHPS to provide useful early detection information on the basis of its historical database without requiring rain gauges or stream gauges. The results indicate there

would be flooding with a return period of approximately 50 years in both urban Beijing and Fangshan 4 days in advance of the event (Figs. 4b, 5b). This analysis also indicates the possibility of near-20-yr return period flooding at Zhangfang 4 days in advance (Fig. 6b) and above-20-yr return period flooding at Zijingguan 2 days in advance (Fig. 7f).

To further assess the predictability of GHPS driven by GFS precipitation forecasts for this event, taking rain gauge observations as ground truth and TRMM RT as an additional benchmark because rain gauge observations are limited on the global scale, both the meteorological and hydrological predictabilities are evaluated



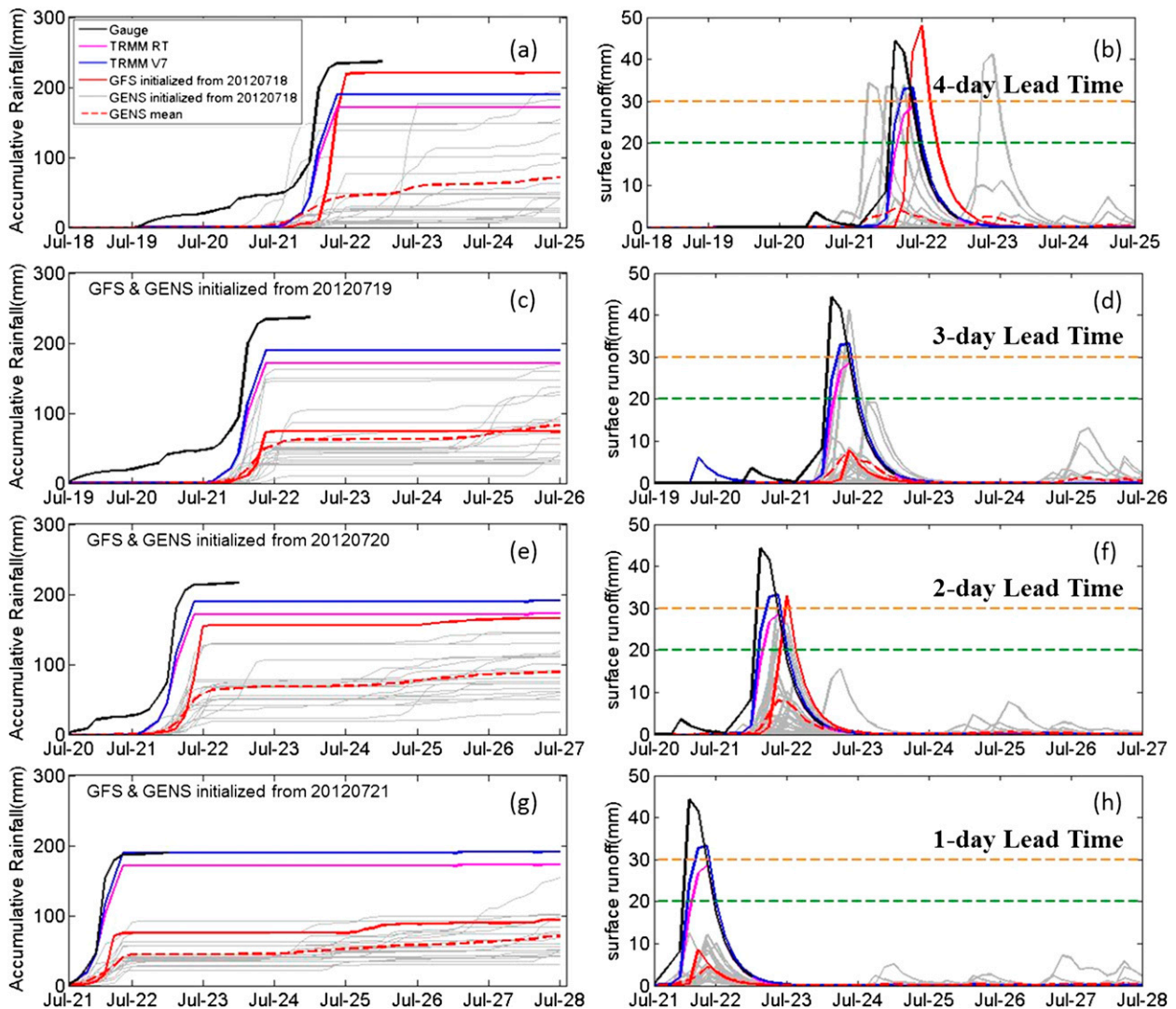


FIG. 5. As in Fig. 4, but for Fangshan (black dot in Fig. 2).

with bias (%) and correlation coefficient (CC) as a function of lead time. In Fig. 8, the bias (%) and CC values of GFS rainfall relative to TRMM RT are calculated for different initialization times. For the meteorological predictability, the bias and CC, as functions of lead time, are calculated over the Beijing area (red outline in Fig. 2). For the hydrological predictability, the bias and CC are calculated for urban Beijing, Fangshan, Zhangfang, and Zijingguan by combining the four series into a mean (Fig. 8b). The GFS has a general trend of increased prediction skill with shorter lead time in terms of both the meteorological (Fig. 8a) and hydrological (Fig. 8b) aspects relative to gauge observations and TRMM RT. The bias of GFS-forced simulations relative to gauge-forced simulations is approximately  $-60\%$  4 days prior to the 21 July event (Fig. 8b). Similarly, the

bias of GFS-forced modeled streamflow relative to TRMM RT-based simulations is around  $-20\%$  with 4 days of lead time. This result shows the potential of the hydrological prognostic capability of GFS-forced GHPS relative to the using TRMM RT in real time.

### c. Probabilistic hydrological evaluation

Probabilistic forecasts derived from ensemble members are considered to be much more attractive for the flood forecasting system because they can provide additional information than the deterministic forecast in regards to the uncertainty or probability of extreme and rare events (Buizza 2008; Gouweleeuw et al. 2005). In this section, the “added value” of ensemble streamflow forecasting contributed by the GHPS is evaluated. First, the ranked probability score (RPS; Jolliffe and Stephenson



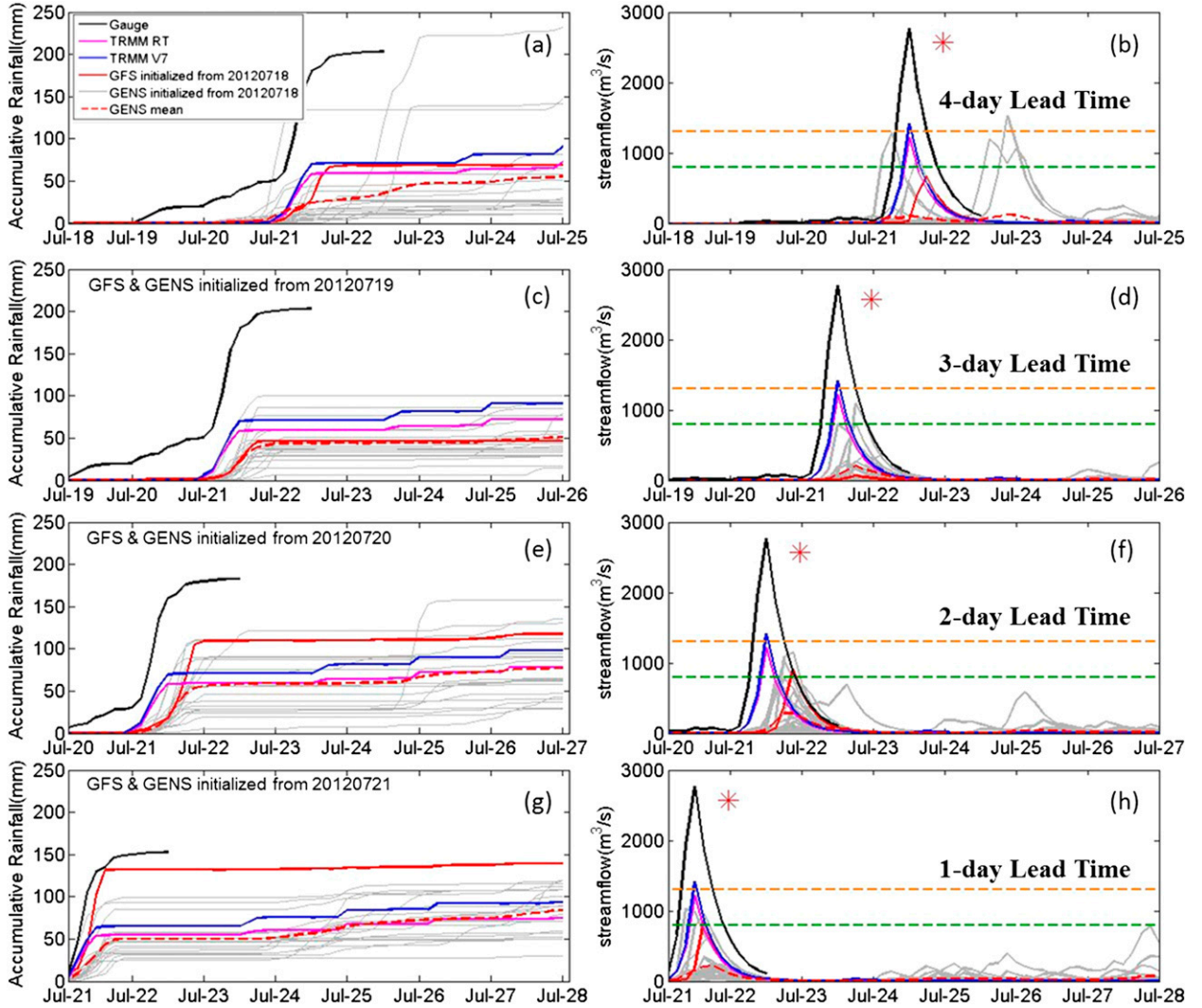


FIG. 6. As in Fig. 4, but for the Zhangfang gauge station (red triangle in Fig. 2) and with streamflow (rather than runoff) given (right). Red asterisk (right) indicates the reported streamflow peak and timing.

2011) is calculated to assess the overall performance of the probabilistic forecast exceeding various thresholds. The RPS is calculated via Eq. (1):

$$\text{RPS} = \sum_{m=1}^J (F_m - O_m)^2, \quad (1)$$

where  $J$  is the number of thresholds,  $F_m$  is the probability of the forecast that exceeds the  $m$ th threshold, and  $O_m$  is the probability of the observation (0 if the observation did not exceed the  $m$ th threshold, 1 if it did). It should be noted here that the deterministic forecast (i.e., GFS, TRMM RT, TRMM V7) only has the probabilities of 0 for no occurrence and 1 for occurrence. For a group of  $n$  forecasts,  $k$  is the  $k$ th group of forecast, and the  $\overline{\text{RPS}}$  is the mean of  $n$  RPSs:

$$\overline{\text{RPS}} = \sum_{k=1}^n \frac{1}{n} \text{RPS}_k. \quad (2)$$

A perfect forecast has an  $\overline{\text{RPS}}$  value of 0. Higher RPS indicates larger forecast probability error.

Figure 9 shows the RPSs of GFS, GENS, TRMM RT, and TRMM V7 at different initializations with a time span of 168 h at the different locations. In this case, the 50- and 20-yr return period thresholds were applied to calculate the RPS for the evaluated rainfall products. Generally, the overall performances of GFS and GENS are worse than TRMM RT and V7 for Beijing, Fangshan, and Zhangfang. For Zijianguan, the overall estimation of TRMM RT- and V7-forced streamflow relative to the 50- and 20-yr thresholds yields the same or worse performance than with GFS

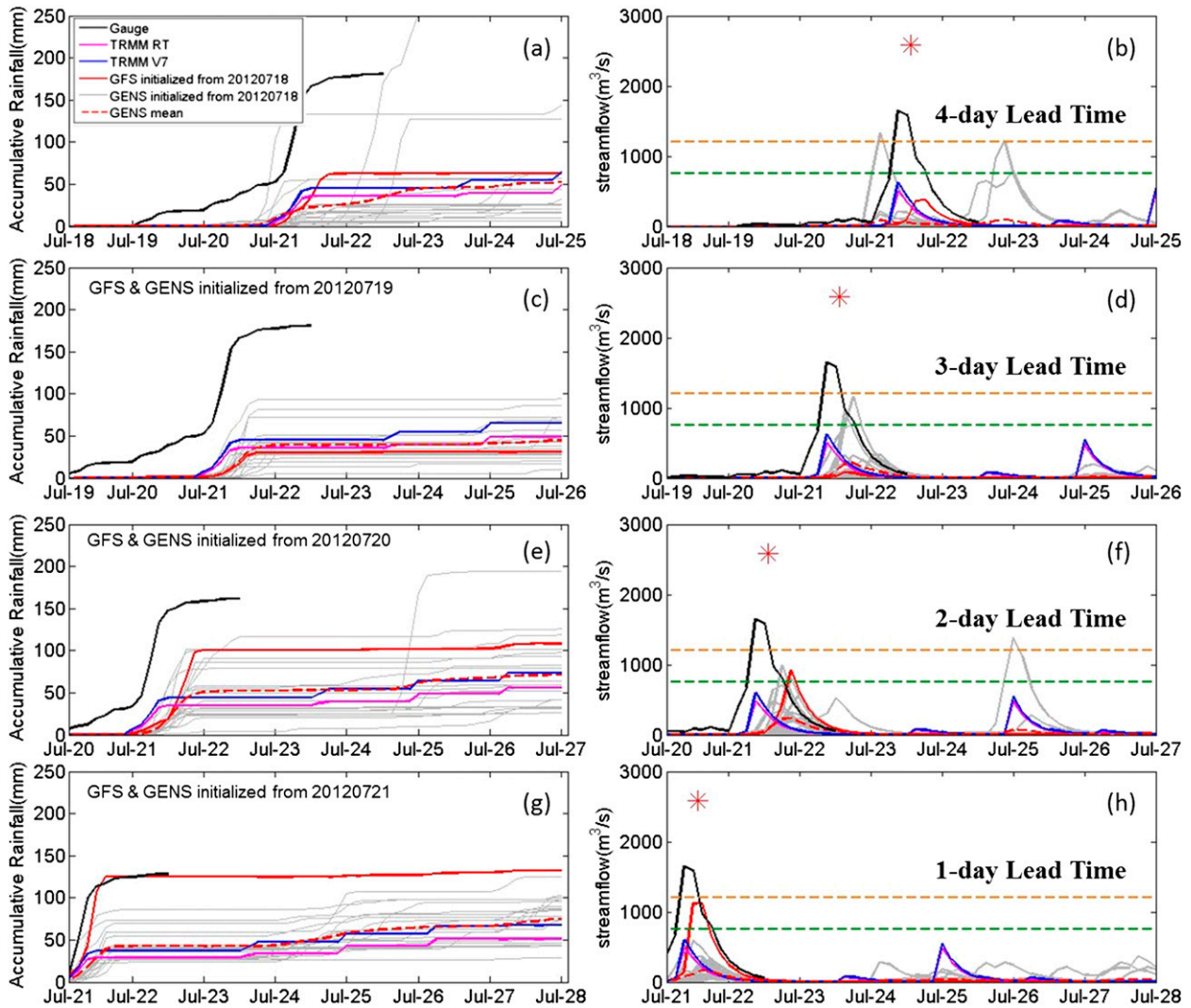


FIG. 7. As in Fig. 6, but for the Zijianguan gauge station (black triangle in Fig. 2).

and GENS. Please note that the frequently applied ensemble streamflow verification metrics (e.g., POD, false alarm rate, reliability diagram, and relative operating characteristic) are not applicable in this study because of the low sample size limitation (Brown et al. 2010; Cloke and Pappenberger 2009). So, the ensemble predictive skill in terms of peak magnitude is investigated to evaluate the ensemble streamflow forecasts (from GENS) relative to the deterministic ones (from GFS), thus delivering additional useful information.

To further assess the probability of ensemble forecast relative to the deterministic forecast, the Brier skill score (BSS; Theis et al. 2005) is calculated:

$$\text{BSS} = 1 - \frac{\text{BS}_{\text{ensemble}}}{\text{BS}_{\text{reference}}}, \quad (3)$$

where Brier score (BS) is equivalent to the RPS in this study and  $\text{BS}_{\text{reference}}$  is the BS of the GFS. The perfect value of BSS is 1. As shown in Fig. 10, most of the BSS exhibits positive value, which indicates that, for most of the cases, the ensemble GENS outperforms the deterministic GFS, especially for the rare and extreme event (i.e., 50-yr return period event).

The ensemble predictive skill in terms of peak magnitude  $P^{\#}(\text{PM})$ , peak timing  $P^{\#}(\text{PT})$ , and both peak magnitude and timing  $P^{\#}(\text{PMT})$  are defined as shown by Eqs. (4)–(6):

$$P^{\#}(\text{PM}) = \frac{n(\text{PM})^{\#}}{n}, \quad (4)$$

$$P^{\#}(\text{PT}) = \frac{n(\text{PT})^{\#}}{n}, \quad (5)$$

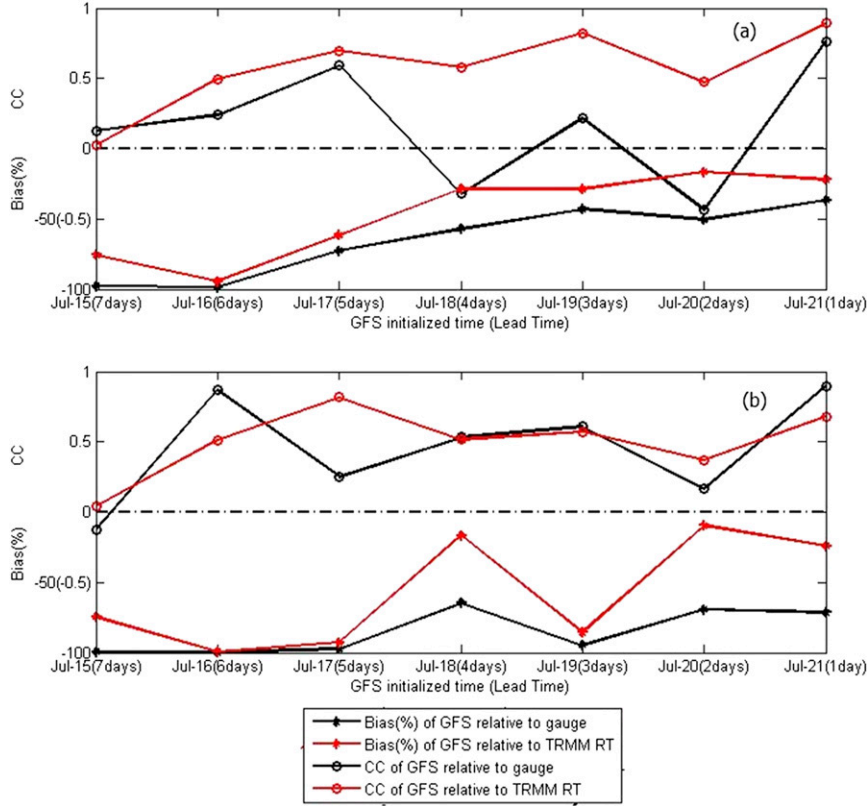


FIG. 8. (a) Meteorological predictability as indicated by bias (%) and CC of GFS relative to gauge and TRMM RT rainfall accumulations; (b) hydrological predictability as indicated by bias (%) and CC of GFS relative to gauge and TRMM RT streamflow simulations.

and

$$P^{\#}(\text{PMT}) = \frac{n(\text{PMT})^{\#}}{n}, \quad (6)$$

where  $n(\text{PM})^{\#}$ ,  $n(\text{PT})^{\#}$ , and  $n(\text{PMT})^{\#}$  are the total number of ensemble streamflow members that have more accurate streamflow forecasts in terms of peak magnitude, peak timing, and both peak magnitude and timing relative to the deterministic forecast, taking the rain gauge–forced streamflow simulation as the benchmark. The results shown in Fig. 11 demonstrate the possibility that ensemble forecasts can deliver more accurate and reliable early warning information regarding the peak magnitude and timing of a rare event. The predictive capabilities of the GENS–forced streamflow ensembles are also demonstrated in relation to the deterministic forecast forced by the GFS.

As indicated by Bartholmes and Todini (2005) that the added benefit of ensemble forecast is not in quantitative flood forecasting (e.g., hydrograph predictions) but in the exceedance of warning levels or the probabilistic forecast. To further examine the probability of

the occurrence of an extreme event, the probabilities of the ensemble forecasts from GENS exceeding the 50- and 20-yr reoccurrence warning levels are calculated at the four locations as a function of lead time (Fig. 12). Recall that 4 days prior to the 21 July 2012 Beijing extreme event, the deterministic streamflow forecasts from all rainfall products at Zhangfang and Zijingguan showed substantial underestimation of the reported observations (red asterisks in Figs. 6b, 7b). The GFS deterministic forecasts are all well below the 50- and even 20-yr recurrence thresholds (orange and green dashed lines in Figs. 6b, 7b), which indicates that early warnings based on the deterministic forecasts were unreliable at Zhangfang and Zijingguan. In contrast, GENS shows probabilities of 20% (Fig. 12c) and 15% (Fig. 12d) for a 20-yr event and 5% and 10% probabilities for a 50-yr event at Zhangfang and Zijingguan with 4 days of lead time.

At Beijing (Fig. 4b) and Fangshan (Fig. 5b), the deterministic streamflow forecast from the GFS exceeds the 50-yr recurrence threshold at a lead time of 4 days. However, the prediction skill degrades at 3 days of lead time at Beijing (Fig. 4d) and Fangshan (Fig. 5d) by not exceeding the 20-yr recurrence threshold. The probabilities for

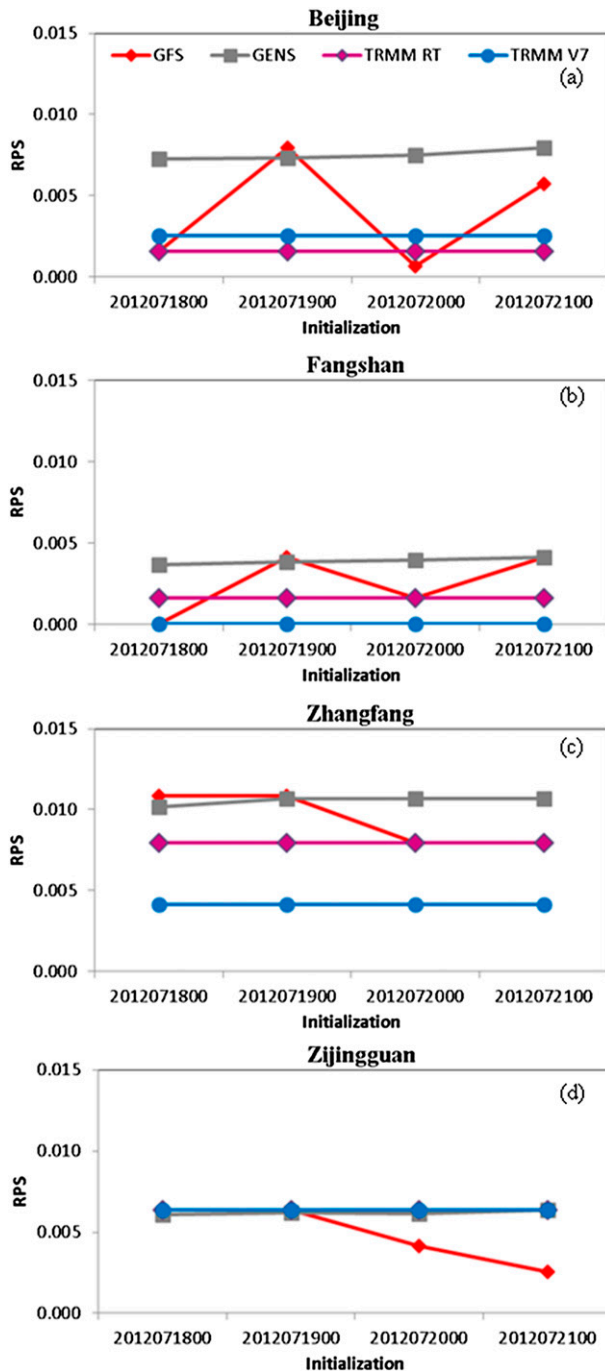


FIG. 9. RPS of streamflow simulations forced by GFS, GENS, TRMM RT, and TRMM V7.

exceeding the 50- and 20-yr streamflow recurrence thresholds by GENS are 10% and 20% at Beijing and 15% and 15% at Fangshan with a lead time of 3 days, which provides potentially useful information for decision makers to issue early warnings. However, the large negative bias with the ensemble forecasts reduces the probability of exceeding the 20- and 50-yr flood thresholds as

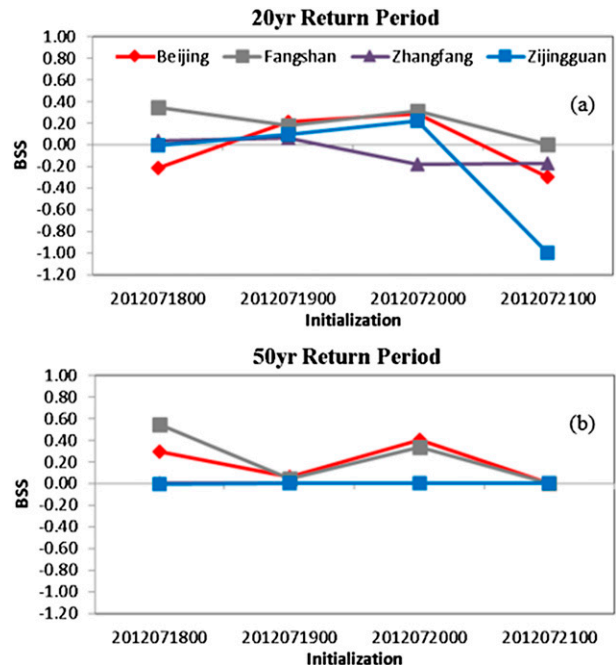


FIG. 10. BSS of the GENS with reference to GFS for the threshold of (a) 20- and (b) 50-yr return period.

lead time decreases. Particularly at 1 day of lead time, the negative bias and the small ensemble spread yield 0% probability of exceeding the 50-yr flood threshold. Despite the negative trend in exceeding flooding thresholds as the event approached, the use of GENS would have been beneficial, especially when the GFS deterministic forecast had significant underestimation problems at lead times of 1 and 3 days. While beyond the scope of the present study, the results also support the use of time-lagged ensembles (i.e., those that incorporate forecasts from prior initializations) in hydrologic forecasting of extreme events.

## 5. Conclusions and future work

The results of this study indicate that the disastrous 21 July Beijing hydrometeorological extreme event was detectable by TRMM satellite precipitation estimates and predictable by deterministic GFS rainfall forecasts at least 4 days in advance. These conclusions are based on results from inputting the precipitation estimates and forecasts to the Global Hydrological Prediction System (GHPS), which has been trained through the use of a decade-long retrospective simulation using TRMM RT rainfall. If the operational hydrological forecast forced by reliable meteorological precipitation forecast products were available and accessible by local stakeholders and integrated into Beijing emergency planning and response decision-making systems, governmental



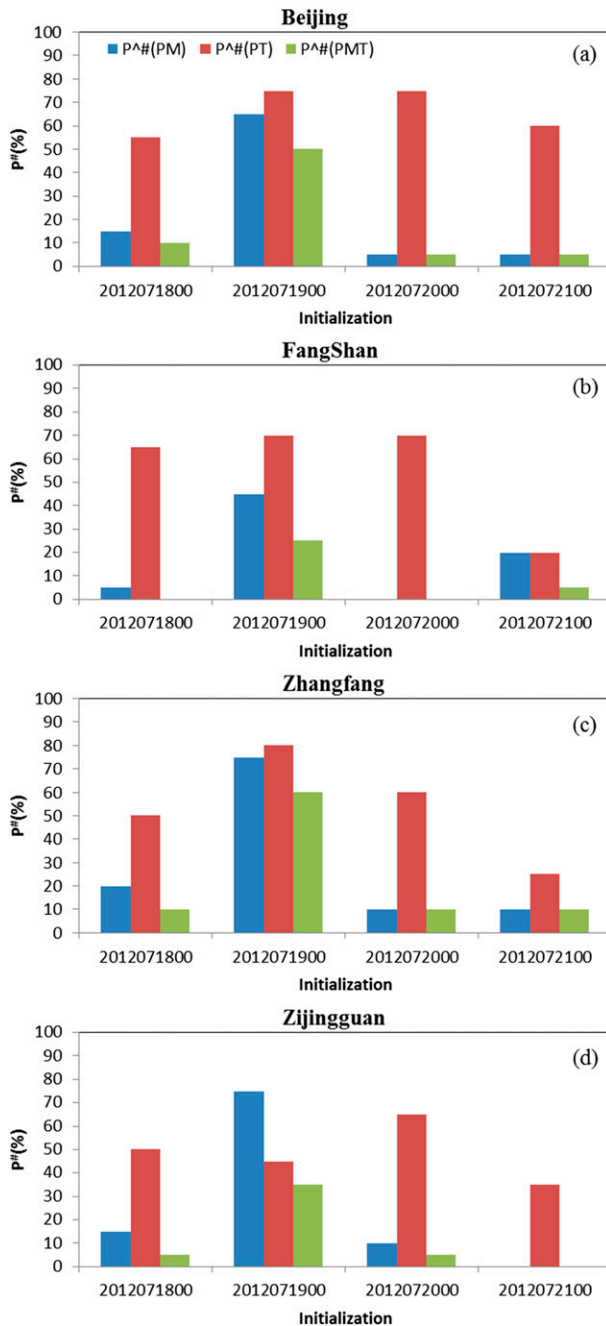


FIG. 11. The predictive skill  $P^{\#}(\%)$  of GENS in terms of peak magnitude, peak timing, and both peak magnitude and timing relative to GFS. Note that  $P^{\#}(\%)$  is the fraction of ensemble streamflow members that have more accurate streamflow forecasts relative to deterministic forecasts.

agencies would have adequate time for preparation and thus would potentially reduce the impacts of flooding, such as the loss of human life and property damages. The GHPS yielded mixed results with forecasts suggesting the likelihood of an extreme event with 4 and 2 days of lead time, but this signal was less obvious at 3 days and

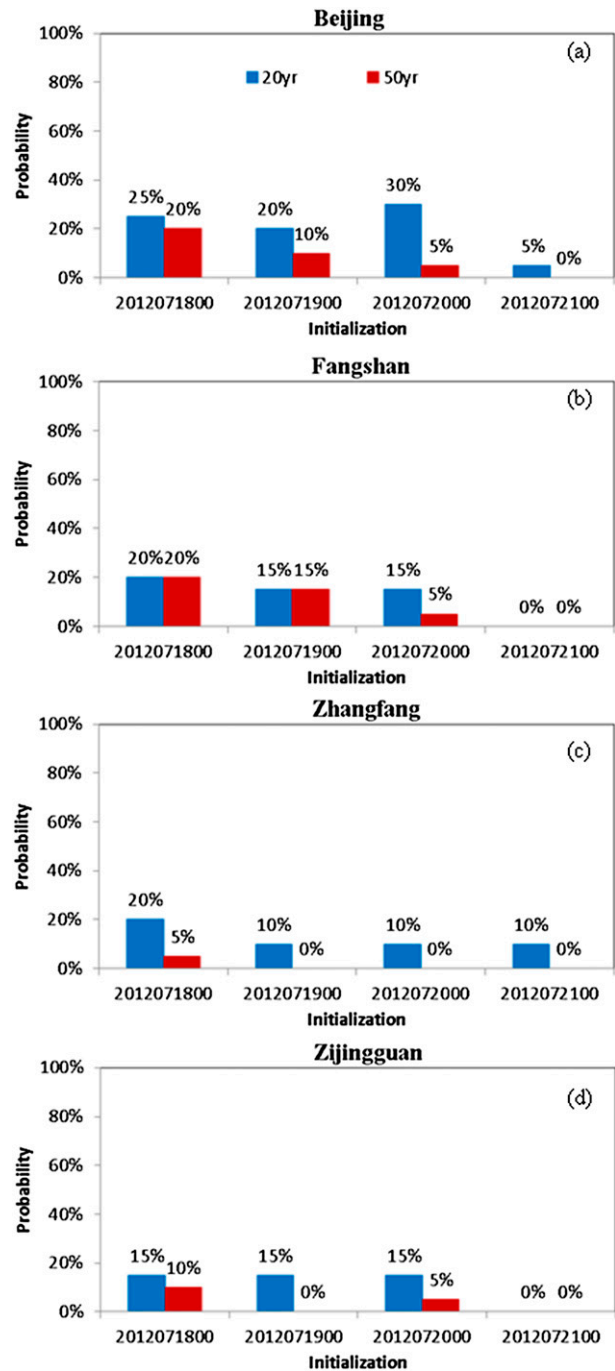


FIG. 12. The probability of exceeding 50- and 20-yr return periods by the ensemble streamflow forecasts.

1 day of lead time. Thus, the GHPS still needs improvements, especially before engaging local stakeholders. The run-to-run inconsistency of the GFS products supports the development and future investigation of ensembles that employ members from prior initializations (i.e., time-lagged ensembles).

This study explored the additional value of the GENS precipitation ensembles in forecasting the probability of an extreme event from the perspective of a global hydrological forecasting system. Given the global availability of such satellite-based precipitation observing system and GENS precipitation forecasting products, this study demonstrates the opportunities and challenges that exist for an integrated application of GHPS and GENS precipitation for flood prediction, systematically over the globe. The method of forecasting rare flooding situations by referencing a decade-long retrospective simulation enables the forecasts to be applied in basins without the requirement of rain gauges or stream gauges. And the hydrological performance is expected to be improved with the recently launched Global Precipitation Mission (GPM) that will yield higher spatio-temporal resolution and accuracy.

To further improve the Global Hydrological Prediction System for more accurate and reliable early flood warnings and responses, some activities are in progress or in planning. First, the regionalization of this system with historical GFS precipitation as input for those areas with high occurrence of flooding events is being explored so that it can locally improve the predictive skill with local expert knowledge as well as data availability. Second, a much more extensive evaluation with a longer period (not only an extreme case study) will be conducted to demonstrate the predictive skill of this system over the globe. We have recently investigated both the deterministic and ensemble GFS precipitation forecasts for a summer season, which is the first stepping stone toward the envisioned future of GHPS forced with the ensemble GFS together with global parameterization. Last, data from current *Aqua*/Advanced Microwave Scanning Radiometer for Earth Observing System (AMSR-E) and future Soil Moisture Active Passive (SMAP; to be launched in early 2015), with anticipated better soil moisture data in terms of coverage, accuracy, and resolutions, might be assimilated for improved hydrological predictions.

**Acknowledgments.** Jeffrey Whitaker is acknowledged for providing GFS ensemble forecast data.

#### REFERENCES

- Adhikari, P., Y. Hong, K. Douglas, D. Kirschbaum, J. Gourley, R. 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](https://doi.org/10.1007/s11069-010-9537-2).
- Bartholmes, J., and E. Todini, 2005: Coupling meteorological and hydrological models for flood forecasting. *Hydrol. Earth Syst. Sci.*, **9**, 333–346, doi:[10.5194/hess-9-333-2005](https://doi.org/10.5194/hess-9-333-2005).
- Bradley, A. A., T. Hashino, and S. S. Schwartz, 2003: Distributions-oriented verification of probability forecasts for small data samples. *Wea. Forecasting*, **18**, 903–917, doi:[10.1175/1520-0434\(2003\)018<0903:DVOPFF>2.0.CO;2](https://doi.org/10.1175/1520-0434(2003)018<0903:DVOPFF>2.0.CO;2).
- , S. S. Schwartz, and T. Hashino, 2004: Distributions-oriented verification of ensemble streamflow predictions. *J. Hydrometeorol.*, **5**, 532–545, doi:[10.1175/1525-7541\(2004\)005<0532:DVOESP>2.0.CO;2](https://doi.org/10.1175/1525-7541(2004)005<0532:DVOESP>2.0.CO;2).
- 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](https://doi.org/10.1029/2006WR005238).
- Brown, J. D., J. Demargne, D.-J. Seo, and Y. Liu, 2010: The Ensemble Verification System (EVS): A software tool for verifying ensemble forecasts of hydrometeorological and hydrologic variables at discrete locations. *Environ. Modell. Software*, **25**, 854–872, doi:[10.1016/j.envsoft.2010.01.009](https://doi.org/10.1016/j.envsoft.2010.01.009).
- , D.-J. Seo, and J. Du, 2012: Verification of precipitation forecasts from NCEP's Short-Range Ensemble Forecast (SREF) system with reference to ensemble streamflow prediction using lumped hydrologic models. *J. Hydrometeorol.*, **13**, 808–836, doi:[10.1175/JHM-D-11-036.1](https://doi.org/10.1175/JHM-D-11-036.1).
- Buizza, R., 2008: The value of probabilistic prediction. *Atmos. Sci. Lett.*, **9**, 36–42, doi:[10.1002/asl.170](https://doi.org/10.1002/asl.170).
- Cloke, H., and F. Pappenberger, 2009: Ensemble flood forecasting: A review. *J. Hydrol.*, **375**, 613–626, doi:[10.1016/j.jhydrol.2009.06.005](https://doi.org/10.1016/j.jhydrol.2009.06.005).
- Demargne, J., and Coauthors, 2009: Application of forecast verification science to operational river forecasting in the U.S. National Weather Service. *Bull. Amer. Meteor. Soc.*, **90**, 779–784, doi:[10.1175/2008BAMS2619.1](https://doi.org/10.1175/2008BAMS2619.1).
- , and Coauthors, 2014: The Science of NOAA's operational Hydrologic Ensemble Forecast Service. *Bull. Amer. Meteor. Soc.*, **95**, 79–98, doi:[10.1175/BAMS-D-12-00081.1](https://doi.org/10.1175/BAMS-D-12-00081.1).
- Gneiting, T., F. Balabdaoui, and A. E. Raftery, 2007: Probabilistic forecasts, calibration and sharpness. *J. Roy. Stat. Soc.*, **69B**, 243–268, doi:[10.1111/j.1467-9868.2007.00587.x](https://doi.org/10.1111/j.1467-9868.2007.00587.x).
- Gouweleeuw, B., J. Thielen, G. Franchello, A. D. Roo, and R. Buizza, 2005: Flood forecasting using medium-range probabilistic weather prediction. *Hydrol. Earth Syst. Sci.*, **9**, 365–380, doi:[10.5194/hess-9-365-2005](https://doi.org/10.5194/hess-9-365-2005).
- Han, J., and H. L. Pan, 2011: Revision of convection and vertical diffusion schemes in the NCEP global forecast system. *Wea. Forecasting*, **26**, 520–533, doi:[10.1175/WAF-D-10-05038.1](https://doi.org/10.1175/WAF-D-10-05038.1).
- Hlavcova, K., J. Szolgay, R. Kubes, S. Kohnova, and M. Zvolensky, 2006: Routing of numerical weather predictions through a rainfall–runoff model. *Transboundary Floods: Reducing Risks through Flood Management*, J. Marsalek, G. Stancalie, and G. Balint, Eds., Springer, 79–90, doi:[10.1007/1-4020-4902-1\\_8](https://doi.org/10.1007/1-4020-4902-1_8).
- Hong, Y., R. F. Adler, F. Hossain, S. Curtis, and G. J. Huffman, 2007: A first approach to global runoff simulation using satellite rainfall estimation. *Water Resour. Res.*, **43**, W08502, doi:[10.1029/2006WR005739](https://doi.org/10.1029/2006WR005739).
- Hopson, T. M., and P. J. Webster, 2010: A 1–10-day ensemble forecasting scheme for the major river basins of Bangladesh: Forecasting severe floods of 2003–07. *J. Hydrometeorol.*, **11**, 618–641, doi:[10.1175/2009JHM1006.1](https://doi.org/10.1175/2009JHM1006.1).
- 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, doi:[10.1175/JHM560.1](https://doi.org/10.1175/JHM560.1).
- , R. F. Adler, D. T. Bolvin, and E. J. Nelkin, 2010: The TRMM multi-satellite precipitation analysis (TMPA). *Satellite Rainfall Applications for Surface Hydrology*,

- M. Gebremichael and F. Hossain, Eds., Springer, 3–22, doi:[10.1007/978-90-481-2915-7\\_1](https://doi.org/10.1007/978-90-481-2915-7_1).
- Jolliffe, I. T., and D. B. Stephenson, Eds., 2011: *Forecast Verification: A Practitioner's Guide in Atmospheric Science*. 2nd ed. John Wiley & Sons, 292 pp.
- Kanamitsu, M., and Coauthors, 1991: Recent changes implemented into the global forecast system at NMC. *Wea. Forecasting*, **6**, 425–435, doi:[10.1175/1520-0434\(1991\)006<0425:RCITG>2.0.CO;2](https://doi.org/10.1175/1520-0434(1991)006<0425:RCITG>2.0.CO;2).
- Khan, S. I., and Coauthors, 2011a: Hydroclimatology of Lake Victoria region using hydrologic model and satellite remote sensing data. *Hydrol. Earth Syst. Sci.*, **15**, 107–117, doi:[10.5194/hess-15-107-2011](https://doi.org/10.5194/hess-15-107-2011).
- , and Coauthors, 2011b: Satellite remote sensing and hydrologic modeling for flood inundation mapping in Lake Victoria basin: Implications for hydrologic prediction in ungauged basins. *IEEE Trans. Geosci. Remote Sens.*, **49**, 85–95, doi:[10.1109/TGRS.2010.2057513](https://doi.org/10.1109/TGRS.2010.2057513).
- 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**, 14 415–14 428, doi:[10.1029/94JD00483](https://doi.org/10.1029/94JD00483).
- Nijssen, B., D. P. Lettenmaier, X. Liang, S. W. Wetzel, and E. F. Wood, 1997: Streamflow simulation for continental-scale river basins. *Water Resour. Res.*, **33**, 711–724, doi:[10.1029/96WR03517](https://doi.org/10.1029/96WR03517).
- Pappenberger, F., K. Scipal, and R. Buizza, 2008: Hydrological aspects of meteorological verification. *Atmos. Sci. Lett.*, **9**, 43–52, doi:[10.1002/asl.171](https://doi.org/10.1002/asl.171).
- Schaake, J. C., T. M. Hamill, R. Buizza, and M. Clark, 2007: HEPEX: The Hydrological Ensemble Prediction Experiment. *Bull. Amer. Meteor. Soc.*, **88**, 1541–1547, doi:[10.1175/BAMS-88-10-1541](https://doi.org/10.1175/BAMS-88-10-1541).
- Seo, D. J., H. D. Herr, and J. C. Schaake, 2006: A statistical post-processor for accounting of hydrologic uncertainty in short-range ensemble streamflow prediction. *Hydrol. Earth Syst. Sci. Discuss.*, **3**, 1987–2035, doi:[10.5194/hessd-3-1987-2006](https://doi.org/10.5194/hessd-3-1987-2006).
- Smith, K., and R. Ward, 1998: *Floods: Physical Processes and Human Impacts*. Wiley, 394 pp.
- Theis, S., A. Hense, and U. Damrath, 2005: Probabilistic precipitation forecasts from a deterministic model: A pragmatic approach. *Meteor. Appl.*, **12**, 257–268, doi:[10.1017/S1350482705001763](https://doi.org/10.1017/S1350482705001763).
- Wang, J., and Coauthors, 2011: The coupled routing and excess storage (CREST) distributed hydrological model. *Hydrol. Sci. J.*, **56**, 84–98, doi:[10.1080/02626667.2010.543087](https://doi.org/10.1080/02626667.2010.543087).
- Wang, X., 2010: Incorporating ensemble covariance in the grid-point statistical interpolation variational minimization: A mathematical framework. *Mon. Wea. Rev.*, **138**, 2990–2995, doi:[10.1175/2010MWR3245.1](https://doi.org/10.1175/2010MWR3245.1).
- , D. Parrish, D. Kleist, and J. Whitaker, 2013: GSI 3DVar-based ensemble-variational hybrid data assimilation for NCEP Global Forecast System: Single-resolution experiments. *Mon. Wea. Rev.*, **141**, 4098–4117, doi:[10.1175/MWR-D-12-00141.1](https://doi.org/10.1175/MWR-D-12-00141.1).
- Westerhoff, R. S., M. P. H. Kleuskens, H. C. Winsemius, H. J. Huizinga, G. R. Brakenridge, and C. Bishop, 2013: Automated global water mapping based on wide-swath orbital synthetic-aperture radar. *Hydrol. Earth Syst. Sci.*, **17**, 651–663, doi:[10.5194/hess-17-651-2013](https://doi.org/10.5194/hess-17-651-2013).
- Wu, H., R. F. Adler, Y. Hong, Y. Tian, and F. Policelli, 2012: Evaluation of global flood detection using satellite-based rainfall and a hydrologic model. *J. Hydrometeor.*, **13**, 1268–1284, doi:[10.1175/JHM-D-11-087.1](https://doi.org/10.1175/JHM-D-11-087.1).
- Yang, F., H. L. Pan, S. K. Krueger, S. Moorthi, and S. J. Lord, 2006: Evaluation of the NCEP Global Forecast System at the ARM SGP site. *Mon. Wea. Rev.*, **134**, 3668–3690, doi:[10.1175/MWR3264.1](https://doi.org/10.1175/MWR3264.1).
- 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](https://doi.org/10.1080/01431161.2010.483489).
- Zappa, M., F. Fundel, and S. Jaun, 2013: A ‘Peak-Box’ approach for supporting interpretation and verification of operational ensemble peak-flow forecasts. *Hydrol. Processes*, **27**, 117–131, doi:[10.1002/hyp.9521](https://doi.org/10.1002/hyp.9521).