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\textsuperscript{1,2,4}, Yang Hong\textsuperscript{1,2,4}, Xuguang Wang\textsuperscript{3,4}, Jonathan J. Gourley\textsuperscript{5}, Xianwu Xue\textsuperscript{1,2}, Manabendra Saharia\textsuperscript{1,2}, Guangheng Ni\textsuperscript{6}, Gaili Wang\textsuperscript{7}, Yong Huang\textsuperscript{8}, Sheng Chen\textsuperscript{1,2}, and Guoqiang Tang\textsuperscript{6}

\textsuperscript{1}School of Civil Engineering and Environmental Science, University of Oklahoma, OK, USA
\textsuperscript{2}Advanced Radar Research Center, University of Oklahoma, Norman, OK, USA
\textsuperscript{3}School of Meteorology, University of Oklahoma, OK, USA
\textsuperscript{4}Center for Analysis and Prediction of Storms, University of Oklahoma, Norman, OK, USA
\textsuperscript{5}NOAA/National Severe Storms Laboratory, National Weather Center, Norman, OK, USA
\textsuperscript{6}Department of Hydraulic Engineering, Tsinghua University, Beijing, China
\textsuperscript{7}State Key Laboratory of Severe Weather, Chinese Academy of Meteorological Science, Beijing, China
\textsuperscript{8}Anhui Meteorological Bureau, Hefei, China

Submitted electronically as a manuscript to

the Journal of Hydrometeorology

October 2014

Corresponding author address:

Dr. Yang Hong

National Weather Center ARRC Suite 4610, 120 David L. Boren Blvd., Norman, OK 73072 USA
Tel: 405-325-3644; Fax: 405-325-4217; email: yanghong@ou.edu; web: http://hydro.ou.edu
Abstract

Prediction and thus preparedness in advance of flood events is crucial for proactively reducing their impacts. In the summer of 2012, Beijing, the capital of China, experienced extreme rainfall and flooding causing economic losses to the tune of 1.6 billion dollars and 79 fatalities. 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) Multi-satellite Precipitation Analysis at near real-time and by the deterministic and ensemble precipitation forecast products from 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 four 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

Floods, 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 due to their high population densities and intensively developed infrastructure. Urban flooding affects structures such as buildings, bridges and roadways and may also induce severe water-borne diseases. On July 21, 2012, the capital of China, Beijing, and its surrounding areas experienced extreme rainfall and flooding. The storm lasted for around 16 hours and the rain rate reached as high as 215 mm/day 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 dollars in damages.

In the same year, the Gelendzhik, Novorossiysk and the Krymsk districts in Russia were affected by the Kuban flood in July and 171 people were killed. Three months later, New York City experienced Hurricane Sandy which flooded the streets, subways, and tunnels, and cut electricity in and around the city in October. Earlier in the previous years, the third biggest city in Australia, Brisbane, was inundated by floods from December 2010 to January 2011 by several separate storm events. As mentioned by (Adhikari et al. 2010), “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 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 they provide 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). In order 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 Hydrological Ensemble Prediction Experiment (HEPEX; (Schaake et al. 2007)), with its mission as “to demonstrate the added value of hydrological ensemble predictions (HEPS) for emergency management and water resources sectors to make decision 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; Bradley et al. 2004; Brown et al. 2012; Brown et al. 2010; Demargne et al. 2009; Demargne et al. 2013; 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 TRMM Multi-satellite Precipitation Analysis (TMPA) and NOAA’s 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? (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 set-up, are described. Then the study region and data sets 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 Global Hydrological Prediction System (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 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 real-time global hydrological simulation and flood monitoring demonstration system (http://eos.ou.edu) at the University of Oklahoma. Presently, it is drive by the TRMM 3B42 Real Time product (RT; (Huffman et al. 2010; Huffman et al. 2007)). The retrospective runs and forecast runs are driven by the post-real time research product from TRMM 3B42 Version 7 (V7; (Huffman et al. 2010; Huffman et al. 2007)) 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 degree based on a digital elevation model (DEM) with quasi-global coverage from 50°N to 50°S, providing near-real time runoff and streamflow simulation every three hours. Rainfall forcing comes from TRMM RT for real-time and by TRMM V7 for retrospective hydrological simulation since 1998.
Precipitation forecasts and subsequent flood forecasts are initialized at 00UTC every day with lead times up to 180 hours at each 1/8 degree grid cell. The model parameters are estimated a priori from Earth physical measurements (for detailed information about the parameter estimation, please refer to (Wang et al. 2011; Wu et al. 2012)). The physical parameters such as soil saturated hydraulic conductivity ($K_{sat}$), and the mean water capacity ($W_M$) in 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 > 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; Khan et al. 2011b; Wu et al. 2012; Yilmaz et al. 2010), proving its high cost-effectiveness in hydrological prediction.

Wu et al. (2013) applied the CREST model, but forced with TRMM V6 (the gauge-corrected research product) to run a retrospective streamflow simulation from 50°N to 50°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 three days in rivers that are not regulated by dams. The generally positive results indicate the potential value of this system forced by TRMM rainfall for global flood detection. However, Wu et al. (2013) 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 version of TRMM data, both TRMM RT (real time product) and TRMM V7 (the post-real time rain gauge-corrected research product), is applied for flood detection. Considering the improvement in satellite precipitation estimates from 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 (AET) at 1/8 degree spatial resolution as shown by the third column in Fig. 1.

In this study, soil states in the global CREST model are initialized by running the model using TRMM RT rainfall forcing from July 1, 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 surrounded by Heibei 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 July 21, 2012 using high-density rain gauge observations with hourly temporal resolution from 03 UTC on July 19, 2012 to 12 UTC on July 22, 2012 (Fig. 2). There are 2041 rain gauge stations in total within the Hebei province and 231 within the city of Beijing. The rain gauge data are interpolated onto a 0.25°-resolution grid using kriging and accumulated into 3-hourly rainfall accumulations in order to facilitate comparison with TMPA products. The TRMM Multi-satellite Precipitation Analysis (TMPA) near-real-time 3B42RT 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 3B42RT and 3B42 V7 products are quasi-global with coverage from 50°N to 50°S latitude at a spatial resolution of 0.25° and temporal resolution of 3 hours.

The deterministic model forecast (GFS hereafter, http://www.emc.ncep.noaa.gov/index.php?branch=GFS) and ensemble model forecast (GENS hereafter) 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 GENS forecasts were run in near-real time by NOAA Earth System Research Lab. The forecasts were initialized by the hybrid ensemble-variational data assimilation system developed based on NOAA NCEP operational data assimilation (Wang 2010) (Wang et al. 2013)). The GFS and GENS 20-member forecasts were initialized four times per day (00, 06, 12, and 18 UTC). The forecasts were produced at three-hourly intervals up to 180-hours of lead time for the GFS and 168 hours for GENS.
The spatial resolution of the forecasts was 0.5 degrees. In this study, only the GFS and GENS products initialized at 00UTC were applied to drive the hydrological forecasts. Both the deterministic and ensemble GFS members were interpolated to 0.25° in order to match the spatial resolution of the TRMM rainfall estimates.

4. Results and Discussion

4.1 Rainfall evaluation

Figure 3 shows the total precipitation accumulation (mm) on July 21, 2012 over the 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), GFS and GENS forecasts which initialized from four days to one 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 July 21 event, but they have limited spatial variability due to their coarse spatial resolution. Both GFS and GENS mean underestimate the daily accumulated precipitation amounts against gauge observation for the different lead times. In particular, the GENS mean has substantial underestimation.

The left panels of Figs. 4-7 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 four days in advance. As shown in the left panels of Figs. 4-5, the GFS exhibits strong run-to-run inconsistencies with significant underforecasts of precipitation according to the gauge observations at three- and one-day lead time, but performing much better at four and two days of lead time. The lag in the GFS precipitation forecast two days in advance at urban Beijing (Fig. 4e) and Fangshan (Fig. 5e) was only 6-9 hours following rainfall as observed by rain gauges. In contrast, for the forecast just one day prior to the event, the timing of the peak rainfall has improved, but there is significant underestimation with errors similar to those associated to the forecasts produced three days prior to the event.

The performance of TRMM V7 and RT are in agreement with one another throughout the July 21 event at urban Beijing and Fangshan (Figs. 4 and 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 July 21 event at Zhangfang and Zijingguan according to gauge observations (Figs. 6 and 7); however, both TRMM V7 and RT products estimate less than 30% of the observed rainfall at both locations.

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

Similar to rainfall evaluation, the right panels of Figs. 4-7 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 July 21 2012 at all four locations. Although the GFS performed reasonably well at a four-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 four days, the GFS-forced simulations matched well with gauge-forced peak runoff, but with around 6-9 hours delay in the timing of the peak (Fig. 5b). This indicates that the GHPS can potentially provide an early warning of up to four 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 four 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 the right panel of Fig. 6), although there is an approximate six-hour 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 four 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 (the right panel of Fig. 7). Interestingly, at two days of lead time, GFS-forced streamflow forecasts are more accurate than those from TRMM rainfall estimates in terms of magnitude, but the timing of the peak is lagged by about 12 hours. At one day of lead time, GFS-forced simulations also show advantages regarding both timing and peak magnitude relative to TRMM forcing.

In order to assess the applicability of the flood detection with the GHPS to ungauged basins over the globe, we used 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 until 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 years (orange dash line in Figs. 4-7) and 20 years (green dash line 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 streamgages. The results indicate there would be flooding with a return period of approximately 50 years in both urban Beijing and Fangshan four days in advance of the event (Figs. 4b and 5b). This analysis also indicates the possibility of near 20-year return
period flooding at Zhangfang four days in advance (Fig. 6b) and above 20-year return
period flooding at Zijingguan two days in advance (Fig. 7f).

In order 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 with Bias (%) and
CC (Correlation Coefficient) 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% four days prior to the Jul 21 event (Fig. 8b). Similarly, the Bias of
GFS-forced modeled streamflow relative to TRMM RT-based simulations is around -20%
with four 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.

4.3 Probabilistic hydrological evaluation

Probabilistic forecasts derived from ensemble members are considered to be much
more attractive for 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 2012)) is calculated to assess the overall performance of the probabilistic forecast exceeding various thresholds. The RPS is calculated via equation (1)

\[
RPS = \sum_{m=1}^{J} (F_m - O_m)^2
\]

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). Please be noted here that the deterministic forecast (i.e. GFS, TRMM RT, TRMM V7) only have the probabilities of 0 for no occurrence and 1 for occurrence. For a group of \( n \) forecasts, the RPS is the mean of \( n \) RPSs:

\[
\overline{RPS} = \frac{1}{n} \sum_{k=1}^{n} RPS_k
\]

A perfect forecast has an RPS value of zero. Higher RPS indicates larger forecast probability error.

Fig. 9 shows the RPSs of GFS, GENS, TRMM RT and TRMM V7 at different initializations with a time span of 168 hrs 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 performance of GFS and GENS are worse than TRMM RT and V7 for Beijing, Fangshan, and Zhangfang. For Zijingguan, the overall estimation of TRMM RT- and V7-forced streamflow relative to the 50- and 20-yr thresholds yield the same or worse performance than with GFS and GENS. Please note that the frequently applied ensemble streamflow verification metrics (e.g. POD, FAR, Reliability Diagram, Relative Operating Characteristic) are not applicable in this study.
due to 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.

In order to further assess the probability of ensemble forecast relative to the deterministic forecast, the Brier Skill Score (BSS, (Theis et al. 2005), Fig. 10) is calculated:

\[ BSS = 1 - \frac{BS_{\text{ensemble}}}{BS_{\text{reference}}} \] (3)

Where BS (Brier Score) is the Brier Score which is equivalent to the RPS in this study. The \( BS_{\text{reference}} \) here 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 for most of the cases, the ensemble GENS outperforms the deterministic GFS, especially for the rare and extreme event (i.e. 50-year return period event).

The ensemble predictive skill in terms of peak magnitude \( P^\#(PM) \), peak timing \( P^\#(PT) \), and both peak magnitude and timing \( P^\#(PMT) \) are defined as shown by equations (4), (5) and (6).

\[ P^\#(PM) = \frac{n(PM)^\#}{n} \] (4)

\[ P^\#(PT) = \frac{n(PT)^\#}{n} \] (5)

\[ P^\#(PMT) = \frac{n(PMT)^\#}{n} \] (6)

where \( n(PM)^\#, n(PT)^\#, n(PMT)^\# \) 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 detailed by (Bartholmes and Todini 2005), “the added benefit of ensemble forecast is not in quantitative flood forecasting (e.g. hydrograph predictions) but in the exceedance of warning levels.” In order to further examine the probability of the occurrence of an extreme event, the probabilities of the ensemble forecasts from the GENS exceeding the 50- and 20-year reoccurrence warning levels are calculated at the four locations as a function of lead time (Fig. 12). Recall that four days prior to the July 21 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 and 7b). The GFS deterministic forecasts are all well below the 50- and even 20-year recurrence thresholds (orange and green dashed lines in Figs. 6b and 7b), which indicates that early warnings based on the deterministic forecasts were unreliable at Zhangfang and Zijingguan. In contrast, the GENS ensemble hydrologic forecasts show probabilities of 20% (Fig. 12c) and 15% (Fig. 12d) for a 20-year event and 5% and 10% probabilities for a 50-year event at Zhangfang and Zijingguan with four days of lead time.

At Beijing (Fig. 4b) and Fangshan (Fig. 5b), the deterministic streamflow forecast from the GFS exceeds the 50-year recurrence threshold at a lead time of four days.
However, the prediction skill degrades at three days of lead time at Beijing (Fig. 4d) and Fangshan (Fig. 5d) by not exceeding the 20-year recurrence threshold. The probabilities for exceeding the 50- and 20-year streamflow recurrence thresholds by the GENS ensemble are 10% and 20% at Beijing, 15% and 15% at Fangshan with a lead time of three 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-year flood thresholds as lead time decreases. Particularly at one day of lead time, the negative bias and the small ensemble spread yield 0% of exceeding the 50-year flood threshold. Despite the negative trend in exceeding flooding thresholds as the event approached, the use of the GENS ensemble would have been beneficial especially when the GFS deterministic forecast had significant underestimation problems at lead times of one and three 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. Conclusion and Future Work

The results of this study indicate that the disastrous July 21 Beijing hydrometeorological extreme event was detectable by TRMM satellite precipitation estimates and predictable by deterministic GFS rainfall forecasts at least four days in advance. These conclusions are based on results from inputting the precipitation estimates and forecasts to the Global Hydrological Prediction System, 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 agencies would have adequate time for the preparation and thus would potentially reduce the impact of flooding, e.g. the loss of human life and property damages. The GHPS yielded mixed results with forecasts suggesting the likelihood of an extreme event with four and two days of lead time, but this signal was less obvious at three and one day of lead time. Thus, the GHPS still needs improvements, especially before engaging local stakeholders. The run-to-run inconsistency of the GFS forecast 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 spatiotemporal resolution and accuracy.
To further improve the Global Hydrological Prediction System for more accurate and reliable early flood warnings and responses, some activities are under 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 towards the envisioned future of GHPS forced with the ensemble GFS together with global parameterization. Last, data from current Aqua/AMSR-E and future SMAP (Soil Moisture Active and Passive, to be launched in 2014), with anticipated better soil moisture data in terms of coverage, accuracy, and resolutions, might be assimilated for improved hydrological predictions.

Acknowledgement:

Jeffrey Whitaker is acknowledged to provide GFS ensemble forecast data.
References:


Han, J., and H. L. Pan, 2011: Revision of convection and vertical diffusion schemes in the NCEP global forecast system. Weather and Forecasting, 26, 520-533.


Kanamitsu, M., and Coauthors, 1991: Recent changes implemented into the global forecast system at NMC. Weather and Forecasting, 6, 425-435.


List of Figures:


FIG.2. Study region over China. The zoomed in insets show urbain Beijing and its upstream Juna River basin and its topography.

FIG.3. Daily precipitation accumulation (mm) on July 21, 2012 from
(a) Rain Gauge stations; (b) TRMM V7; (c) TRMM RT;
(d) GFS initialized from July 18 2012 at 00 UTC;
(e) Ensemble mean initialized from July 18 2012 at 00 UTC;
(f) GFS initialized from July 19 2012 at 00 UTC;
(g) Ensemble mean initialized from July 19 2012 at 00 UTC;
(h) GFS initialized from July 20 2012 at 00 UTC;
(i) Ensemble mean initialized from July 20 2012 at 00 UTC;
(j) GFS initialized from July 21 2012 at 00 UTC;
(k) Ensemble mean initialized from July 21 2012 at 00 UTC;

FIG.4. Accumulated rainfall initialized from different dates at 00 UTC at urban Beijing (red dot in fig. 2) from different products: Gauge observation, TRMM RT, TRMM V7, GFS, GENS and GENS mean (left panel). GHPS predicted streamflow initialized from different dates at 00 UTC forced by different precipitation products: Gauge observation, TRMM RT, TRMM V7, GFS and GENS (right panel). Note: Rows 1 to 4 indicate lead time from four to one day. Orange dash line indicates 50-year return period surface runoff/streamflow threshold. Green dash line indicates 20-year return period surface runoff/streamflow threshold.

FIG.5. Accumulated rainfall initialized from different dates at 00 UTC at Fangshan (black dot in fig. 2) from different products: Gauge observation, TRMM RT, TRMM V7, GFS, GENS and GENS mean (left panel). GHPS predicted streamflow initialized from different dates at 00 UTC forced by different precipitation products: Gauge observation, TRMM RT, TRMM V7, GFS and GENS (right panel). Note: Rows 1 to 4 indicate lead time from four to one day. Orange dash line indicates 50-year return period surface runoff/streamflow threshold. Green dash line indicates 20-year return period surface runoff/streamflow threshold.
FIG. 6. Accumulated rainfall initialized from different dates at 00 UTC at Zhangfang gauge station (red triangle in fig. 2) from different products: Gauge observation, TRMM RT, TRMM V7, GFS, GENS and GENS mean (left panel). GHPS predicted streamflow initialized from different dates at 00 UTC forced by different precipitation products: Gauge observation, TRMM RT, TRMM V7, GFS and GENS (right panel). Note: Rows 1 to 4 indicate lead time from four to one day. Orange dash line indicates 50-year return period surface runoff/streamflow threshold. Green dash line indicates 20-year return period surface runoff/streamflow threshold. Red asterisk in the right panel indicates the reported streamflow peak and timing.

FIG. 7. Accumulated rainfall initialized from different dates at 00 UTC at Zijingguan gauge station (black triangle in fig. 2) from different products: Gauge observation, TRMM RT, TRMM V7, GFS, GENS and GENS mean (left panel). GHPS predicted streamflow initialized from different dates at 00 UTC forced by different precipitation products: Gauge observation, TRMM RT, TRMM V7, GFS and GENS (right panel). Note: Rows 1 to 4 indicate lead time from four to one day. Orange dash line indicates 50-year return period surface runoff/streamflow threshold. Green dash line indicates 20-year return period surface runoff/streamflow threshold. Red asterisk in the right panel indicates the reported streamflow peak and timing.

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

FIG. 9. Ranked probability score (RPS) of streamflow simulations forced by GFS, GENS, TRMM RT, and TRMM V7.

FIG. 10. (a) Brier Skill Score (BSS) of the GENS with reference to GFS for the threshold of 20yr return period; (b) BSS of the GENS with reference to GFS for the threshold of 50yr return period.

FIG. 11. The predictive skill $P^F(\%)$ of GENS in terms of Peak Magnitude (PM), Peak Timing (PT) and both Peak Magnitude and Timing (PMT) relative to GFS. Note that $P^F(\%)$ is the fraction of ensemble streamflow members that have more accurate streamflow forecasts relative to deterministic forecasts.

FIG. 12. The probability of exceeding 50- and 20-year return periods by the ensemble streamflow forecasts.
FIG. 2. Study region over China. The zoomed in insets show urban Beijing and its upstream Juma River basin and its topography.
FIG. 3. Daily precipitation accumulation (mm) on July 21, 2012 from

(a) Rain Gauge stations; (b) TRMM V7; (c) TRMM RT;
(d) GFS initialized from July 18 2012 at 00 UTC;
(e) Ensemble mean initialized from July 18 2012 at 00 UTC;
(f) GFS initialized from July 19 2012 at 00 UTC;
(g) Ensemble mean initialized from July 19 2012 at 00 UTC;
(h) GFS initialized from July 20 2012 at 00 UTC;
(i) Ensemble mean initialized from July 20 2012 at 00 UTC;
(j) GFS initialized from July 21 2012 at 00 UTC;
(k) Ensemble mean initialized from July 21 2012 at 00 UTC;
FIG. 4. Accumulated rainfall initialized from different dates at 00 UTC at urban Beijing (red dot in fig. 2) from different products: Gauge observation, TRMM RT, TRMM V7, GFS, GENS and GENS mean (left panel). GHPS predicted streamflow initialized from different dates at 00 UTC forced by different precipitation products: Gauge observation, TRMM RT, TRMM V7, GFS and GENS (right panel). Note: Rows 1 to 4 indicate lead time from four to one day. Orange dash line indicates 50-year return period surface runoff/streamflow threshold. Green dash line indicates 20-year return period surface runoff/streamflow threshold.
FIG. 5. Accumulated rainfall initialized from different dates at 00 UTC at Fangshan (black dot in fig. 2) from different products: Gauge observation, TRMM RT, TRMM V7, GFS, GENS and GENS mean (left panel). GHPS predicted streamflow initialized from different dates at 00 UTC forced by different precipitation products: Gauge observation, TRMM RT, TRMM V7, GFS and GENS (right panel). Note: Rows 1 to 4 indicate lead time from four to one day. Orange dash line indicates 50-year return period surface runoff/streamflow threshold. Green dash line indicates 20-year return period surface runoff/streamflow threshold.
FIG. 6. Accumulated rainfall initialized from different dates at 00 UTC at Zhangfang gauge station (red triangle in fig. 2) from different products: Gauge observation, TRMM RT, TRMM V7, GFS, GENS and GENS mean (left panel). GHPS predicted streamflow initialized from different dates at 00 UTC forced by different precipitation products: Gauge observation, TRMM RT, TRMM V7, GFS and GENS (right panel). Note: Rows 1 to 4 indicate lead time from four to one day. Orange dash line indicates 50-year return period surface runoff/streamflow threshold. Green dash line indicates 20-year return period surface runoff/streamflow threshold. Red asterisk in the right panel indicates the reported streamflow peak and timing.
FIG. 7. Accumulated rainfall initialized from different dates at 00 UTC at Zijingguan gauge station (black triangle in fig. 2) from different products: Gauge observation, TRMM RT, TRMM V7, GFS, GENS and GENS mean (left panel). GHPS predicted streamflow initialized from different dates at 00 UTC forced by different precipitation products: Gauge observation, TRMM RT, TRMM V7, GFS and GENS (right panel).

Note: Rows 1 to 4 indicate lead time from four to one day. Orange dash line indicates 50-year return period surface runoff/streamflow threshold. Green dash line indicates 20-year return period surface runoff/streamflow threshold. Red asterisk in the right panel indicates the reported streamflow peak and timing.
FIG. 8. (a) Meteorological predictability as indicated by bias (%) and correlation coefficient (CC) of GFS relative to Gauge and TRMM RT rainfall accumulations; (b) Hydrological predictability as indicated by bias (%) and correlation coefficient (CC) of GFS relative to Gauge and TRMM RT streamflow simulations.
FIG. 9. Ranked probability score (RPS) of streamflow simulations forced by GFS, GENS, TRMM RT, and TRMM V7.
FIG. 10. (a) Brier Skill Score (BSS) of the GENS with reference to GFS for the threshold of 20yr return period; (b) BSS of the GENS with reference to GFS for the threshold of 50yr return period.
FIG. 11. The predictive skill $P^\#(\%)$ of GENS in terms of Peak Magnitude (PM), Peak Timing (PT) and both Peak Magnitude and Timing (PMT) relative to GFS. Note that $P^\#(\%)$ is the fraction of ensemble streamflow members that have more accurate streamflow forecasts relative to deterministic forecasts.
FIG. 12. The probability of exceeding 50- and 20-year return periods by the ensemble streamflow forecasts.
List of Tables:

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

<table>
<thead>
<tr>
<th>Rainfall Products</th>
<th>Spatial Resolution</th>
<th>Time interval</th>
<th>Lead Time</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rain Gauge Observation</td>
<td>interpolated onto 0.25°</td>
<td>Hourly; then accumulated to 3 hourly</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>TRMM RT</td>
<td>0.25°</td>
<td>3 hourly</td>
<td>N/A</td>
<td>Huffman et al. 2007</td>
</tr>
<tr>
<td>TRMM RP</td>
<td>0.25°</td>
<td>3 hourly</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Deterministic GFS Precipitation</td>
<td>0.25°</td>
<td>3 hourly</td>
<td>180hr</td>
<td>Wang 2010; Wang et al. 2013</td>
</tr>
<tr>
<td>Ensemble GFS Precipitation</td>
<td>0.25°</td>
<td>3 hourly</td>
<td>168hr</td>
<td></td>
</tr>
</tbody>
</table>