



# AMERICAN METEOROLOGICAL SOCIETY

*Weather and Forecasting*

## **EARLY ONLINE RELEASE**

This is a preliminary PDF of the author-produced manuscript that has been peer-reviewed and accepted for publication. Since it is being posted so soon after acceptance, it has not yet been copyedited, formatted, or processed by AMS Publications. This preliminary version of the manuscript may be downloaded, distributed, and cited, but please be aware that there will be visual differences and possibly some content differences between this version and the final published version.

The DOI for this manuscript is doi: 10.1175/WAF-D-12-00124.1

The final published version of this manuscript will replace the preliminary version at the above DOI once it is available.

If you would like to cite this EOR in a separate work, please use the following full citation:

Clark III, R., J. Gourley, Z. Flamig, Y. Hong, and E. Clark, 2013: CONUS-wide Evaluation of National Weather Service Flash Flood Guidance Products. *Wea. Forecasting*. doi:10.1175/WAF-D-12-00124.1, in press.



# CONUS-wide Evaluation of National Weather Service Flash Flood Guidance

## Products

Robert A. Clark<sup>1,2,3</sup>, Jonathan J. Gourley<sup>\*2</sup>, Zachary L. Flamig<sup>2,3</sup>, Yang Hong<sup>3,4</sup>, Edward Clark<sup>5</sup>

1 – Cooperative Institute for Mesoscale Meteorological Studies, University of Oklahoma, Norman, OK

2 – NOAA/National Severe Storms Laboratory, Norman, OK

3 – Advanced Radar Research Center, University of Oklahoma, Norman, OK

4 – Department of Civil Engineering and Environmental Science, University of Oklahoma, Norman, OK

5 – NOAA/National Weather Service Office of Climate, Water, and Weather Services, Silver Spring, MD

## Revised Draft

*Submitted electronically as an article to: Weather and Forecasting*

**22 October 2013**

*\*Address for Correspondence:* Jonathan Gourley, National Weather Center, 120 David L. Boren Blvd, Norman, OK 73072-7303

Tel: 405 325-6472

Email: jj.gourley@noaa.gov

## Abstract

This study quantifies the skill of the National Weather Service's (NWS) Flash Flood Guidance (FFG) product. Generated by River Forecast Centers (RFCs) across the US, local NWS Weather Forecast Offices compare estimated and forecast rainfall to FFG to monitor and assess flash flooding potential. A national flash flood observation database consisting of NWS *Storm Data* reports and United States Geologic Survey (USGS) stream gauge measurements is used to determine the skill of FFG over a four-year duration. FFG skill is calculated at several different precipitation-to-FFG ratios for both observation datasets. Although a ratio of 1.0 nominally indicates a potential flash flooding event, this study finds that FFG can be more skillful when ratios other than 1.0 are considered. When the entire continental US is considered, the highest observed critical success index (CSI) with 1-hr FFG is 0.20 for the USGS dataset, which should be considered a benchmark for future research that seeks to improve, modify, or replace the current FFG system. Regional benchmarks of FFG skill are also determined on an RFC-by-RFC basis. When evaluated against NWS *Storm Data* reports, the regional skill of FFG ranges from 0.00 to 0.19. When evaluated against USGS stream gauge measurements, the regional skill of FFG ranges from 0.00 to 0.44.

## 1. Introduction

Flash floods are the second most deadly weather-related hazard in the United States behind extreme heat (Ashley and Ashley 2008). The US National Weather Service (NWS) Glossary (cited 2012) defines a “flash flood” as follows:

A flash flood is a rapid and extreme flow of high water into a normally dry area, or a rapid water level rise in a stream or creek above a predetermined flood level, beginning within six hours of the causative event (e.g., intense rainfall, dam failure, ice jam).

This six-hour threshold is used within the NWS to divide hydrologic forecasting and monitoring responsibility between regional River Forecast Centers (RFCs), who deal with fluvial floods that take place over longer time-scales, and local Weather Forecast Offices (WFOs), who deal with flash floods that happen on shorter time-scales (Gourley et al. 2012).

Despite recent advances made in hydrologic modeling, quantitative precipitation estimation, and numerical weather prediction, some components of the system the NWS uses to forecast and monitor dangerous flash flood events are forty years old (RFC Development Management Team 2003). This system, which includes flash flood guidance (FFG), was originally implemented after a deadly 1969 flash flood in Ohio (Schmidt et al. 2007). FFG was significantly modified in 1992 (Sweeney and Baumgardner 1999) with additional changes undertaken in the last ten years (Schmidt et al. 2007; Smith 2003). Henceforth, the pre-1992 product will be referred to as “original FFG”. The product used between 1992 and the early 2000s (and still used at some RFCs as of 2012) will be referred to as “lumped FFG” (LFFG). Finally, additional FFG

products developed at the Colorado Basin River Forecast Center (CBRFC) in 2003, at the Arkansas Red-Basin River Forecast Center (ABRFC) in 2005, and at the Middle Atlantic River Forecast Center (MARFC) will be called “flash flood potential index” (FFPI), “gridded flash flood guidance” (GFFG), and “distributed flash flood guidance” (DFFG), respectively. “Flash flood guidance” (FFG) is a broader term that encompasses all of these above methods and refers to the general product issued by the RFCs for the purposes of helping WFOs monitor and forecast flash flooding events.

FFG is defined as the amount of rain required in a given time and area to produce bank full conditions on small streams; these conditions are considered to be associated with flash flooding. FFG is produced at twelve RFCs located throughout the continental United States (see Fig. 1). Regardless of the exact flavor of FFG being produced, this RFC-derived FFG is then delivered to the NWS’s network of WFOs. WFOs overlay their most accurate and timely precipitation estimates onto FFG values and areas where precipitation exceeds FFG are potential candidates for flash flood warnings or other actions on the part of WFO forecasters (Gourley et al. 2012).

Although the NWS maintains verification statistics regarding flash flood warnings, there is presently no system to provide feedback to WFOs or RFCs about the effectiveness of the FFG product that is typically used to issue these flash flood warnings (RFC Development Management Team 2003). Therefore, this study will, for the first time, provide these verification statistics for FFG over the entire continental United States (CONUS) from 1 October 2006 to 31 August 2010. The general methodology is similar to that outlined in Gourley et al. (2012); they produced similar statistics over the ABRFC area of responsibility for the 2006-2008 period. The specific study objectives are

to determine the benchmark skill of FFG across the entire CONUS, to determine the benchmark skill of FFG at each CONUS RFC, to offer specific recommendations to the NWS regarding improvements to the use of FFG at RFCs and WFOs, and to provide the research community with information about the current state of US flash flood forecasting and monitoring. The next section provides a historical context of FFG leading up to the current status in the NWS. We provide the technical details of each of the operational FFG-generation methods in section 3 and the analysis methodology in section 4. Section 5 presents the results of this study, and concluding remarks are supplied in section 6.

## **2. The History and Current Status of Flash Flood Guidance**

From the 1940s to the 1970s, the average annual number of deaths due to flash flooding tripled, while the monetary damages due to flash flooding increased more than six-fold (Mogil et al. 1978). The NWS flash flood warning program was not deployed nationally until 1971, but severe thunderstorms and tornadoes had national warning programs for years or decades before that. In the early 1970s, RFCs already produced an early FFG product – original FFG – based “on drainage basin configuration and past rainfall” (Mogil et al. 1978). Through the 1970s, methods of estimating rainfall from convective activity varied from office to office within the NWS while additional local programs, including flash flood alarm systems, were explored. However, the “critical element” in any of these local programs remained the RFC-generated FFG product (Mogil et al. 1978).

During the 1970s and 1980s, the NWS developed the NWS River Forecast System (NWSRFS) (RFC Development Management Team 2003). This system was initially only used to produce forecasts for larger-scale fluvial floods, but provided national consistency between the RFCs for those particular products. By the 1980s, the use of NWSRFS to produce FFG was being explored, as well, due to the local and regional differences between the FFG being produced at each RFC. This work eventually resulted in “modernized FFG” or lumped FFG (LFFG) (Sweeney 1992).

There were two main impetuses behind the development of LFFG: the deployment of the Advanced Weather Interactive Processing System (AWIPS) and the more-accurate and higher-resolution precipitation estimates available from the WSR-88D network (Sweeney and Baumgardner 1999). An additional benefit was the increased consistency in FFG generation method at each RFC (Sweeney 1992). This modernization project also put FFG generation into the same framework as the RFC river stage forecast system that had been developed during the 1980s, and now national standards were available to guide RFCs in the process of generating FFG products. In 2003, the RFC Development Management Team issued a report regarding the state of FFG at that time as well as several recommendations regarding the future direction of the program. The most significant advance described in the report is the delineation of small, truly flash flood-scale basins. The National Basin Delineation Project (NBDP) used geographic information system (GIS) technology to produce flash flood scale basin datasets for each NWS WFO (Arthur et al. 2005). These basins then are used as part of the FFMP (Flash Flood Monitoring and Prediction) system, which was deployed as part of the AWIPS software package. The scale of these small flash flood basins is much more similar to the

scale of precipitation estimates from the WSR-88Ds. The average basin area traced out by the NBDP is around ten km<sup>2</sup> (RFC Development Management Team 2003) and the minimum size is five km<sup>2</sup> (Davis 2007). This does not eliminate the resolution gap between WSR-88D precipitation estimates and the lumped FFG basins used at the RFCs (300 – 5,000 km<sup>2</sup>). In other words, the FFG was still representative of processes on the large basin scale, not the newly computed small basins. The RFC Development Management Team (2003) recognized this limitation, but due to computational requirements and scientific limitations, this particular issue was still partially unresolved at the end of the study period.

The RFC Development Management Team (2003) primarily focused on suggesting small changes to the FFG system rather than major modifications or a complete overhaul. Several issues are described in that report are still observed in the FFG mosaics used in this study (1 October 2006 – 31 August 2010). Some HRAP grid cells always have missing FFG values; this problem occurs within RFC domains and on the boundaries between domains (see Fig. 2). Other grid cells on RFC boundaries have multiple overlapping (and different) FFG values. Additionally, FFG values can exhibit sharp gradients along RFC boundaries, in many cases for no hydrologic reason. These problems are due to software and hardware limitations, hydrologic model parameter differences between RFCs, or even model differences between RFCs. Most seriously, no national verification program for FFG has ever been developed (RFC Development Management Team 2003).

During and immediately after the RFC Development Management Team recommendations, some RFCs began to modify or replace the lumped FFG product.



However, lumped FFG was still being produced at multiple RFCs at the end of the study period. Starting in 2003, the Colorado Basin RFC began testing a replacement for LFFG. This replacement is a flash flood potential index method (FFPI). In 2005, the Arkansas-Red Basin RFC deployed a method known as gridded flash flood guidance (GFFG). Over the next few years, both FFPI and GFFG methods were implemented at additional RFCs. A fourth method is used at the Middle Atlantic RFC and is referred to as distributed flash flood guidance, or DFFG, in this paper. Technical details about all four FFG generation methods are provided in Section 3. The history of the NWS flash flood warning program involved only minor alterations for its first twenty years of existence. With the advent of the modernized FFG program at the beginning of the 1990s, RFC methodologies were mostly standardized across the US. However, several problems continued to be noted by forecasters and others throughout the next ten years. In the last decade, hydrologists and meteorologists at some RFCs have developed their own FFG products. This patchwork of different methods of FFG generation (LFFG, FFPI, GFFG, and DFFG) was the state of flash flood operations in the NWS as of 2010.

### **3. Technical Details of Flash Flood Guidance**

#### *a. Lumped Flash Flood Guidance*

Sweeney and Baumgardner (1999) describe the methodology used to generate LFFG values. Rainfall-runoff models are normally used to determine the amount of runoff generated by a given amount of rainfall and a particular soil moisture condition. In LFFG, a rainfall-runoff model is run in reverse – the FFG value transmitted to a WFO is the amount of rainfall required to cause bankfull (i.e., flooding) conditions on small streams

at the basin outlet. Thus, in this process, it is necessary to know the state of two variables: soil moisture and threshold runoff (ThreshR). Soil moisture data for LFFG comes from the same information used by RFCs to produce river stage forecasts on large basins. ThreshR is a function of basin geography; ThreshR values are most easily determined at gauged basin outlets but some RFCs have undertaken field campaigns to determine ThreshR at ungauged locations, as well. Because RFCs must produce FFG over large areas, ThreshR values are usually contoured between gauged points to produce areal averages (Gourley et al. 2012). In some RFCs, a single ThreshR value is assigned to entire states while in others, each county has its own ThreshR value assigned. Other offices delineated basins on the order of 1,000 km<sup>2</sup>, calculated ThreshR at the headwaters of these basins, then averaged these on a county-by-county basis. Thus, ThreshR values are not always representative of small basin hydrology. Additional information on the available threshold runoff calculation methods can be found in Carpenter et al. (1999).

LFFG works with any sort of soil moisture information and does not require a specific rainfall-runoff model, though most RFCs utilize the Sacramento Soil Moisture Accounting (SAC-SMA) model. Although the exact issuance schedule of LFFG varies, if precipitation data is available on schedule at the RFCs, the soil moisture information used for RFC river stage forecasts is updated every six hours and thus LFFG can be updated every six hours, as well (Sweeney and Baumgardner 1999). A major limitation of LFFG is that model parameters are constant for the basins of 300-5,000 km<sup>2</sup> over which the various rainfall-runoff models run (RFC Development Management Team 2003). However, flash flooding events are often observed on basins of much smaller size and the lumped-parameter method does not allow for variability in soil moisture conditions

within each lumped basin. Additionally, even though the rainfall-runoff models used in the generation process are calibrated with six-hour time steps, flash flood events take place on time scales of less than six hours and so these rainfall-runoff models may not be best suited for flash flood forecasting (RFC Development Management Team 2003).

*b. Flash Flood Potential Index*

The Colorado Basin RFC covers an area where flash flooding is not necessarily associated with bank full conditions on small streams (Smith 2003). Additionally, in this area, soil moisture is believed to be a less important component of determining when and where a flash flood might occur (RFC Development Management Team 2003). Therefore, in 2003 and 2004, CBRFC undertook a project to develop FFPI (Flash Flood Potential Index), designed as a replacement for LFFG. FFPI uses gridded physiographic information (soil characteristics, vegetation cover [including forest density], slope, land use and urbanization, and seasonal effects like wildfire) to determine the relative likelihood of flash flooding in a given FFMP basin (Smith 2003). The FFPI method developed at CBRFC was eventually deployed at the California Nevada RFC (CNRFC) in 2008; the Northwest RFC (NWRFC) also uses a similar method.

Each piece of gridded physiographic information referenced above was resampled to a consistent resolution and then each HRAP grid cell was assigned a flash flood potential index on a scale of 1 (least hydrologically sensitive to rainfall) to 10 (most hydrologically sensitive to rainfall) for each layer of data. An average of these potentials yields the final FFPI product. Initially, all layers were given equal weight except for the slope parameter, which was weighted above the other data layers.

The slope data are derived from a USGS digital elevation model dataset and sampled at a resolution of 400 meters; thus, the smallest basins that can be defined from this process will have a drainage area of roughly 60 km<sup>2</sup>. Although this represents a significant resolution improvement over the larger lumped FFG basins, it is still coarser than the FFMP basins or an individual HRAP grid cell. The second piece of data used in FFPI generation is soil type. A total of sixteen possible soil types are defined by the STATSGO (State Soil Geographic data from the Natural Resources Conservation Service) dataset. Land use information comes from the Landsat satellite program and forest density was obtained from satellite imagery.

It is possible for FFPI values to change over seasonal timescales. Some of these changes are due to WFO requests to alter the FFPI of specific basins. Other changes are due to wildfires; after such events, FFPI can be modified to reflect changes in soil permeability and forest cover. Still other variations in the FFPI grid are due to seasonal changes in vegetation cover and snow cover (RFC Development Management Team 2003).

The original FFPI product was interpolated to FFMP basins over the CBRFC area of responsibility and was used at individual WFOs as a supplement to LFFG, not as a replacement for LFFG. Positive forecaster feedback resulted in CBRFC eventually replacing LFFG with FFPI. In operations at the Colorado Basin RFC, FFPI basin susceptibilities are used to adjust a one inch-per-hour (2.54 mm/h) rainfall rate; this modified rule of thumb is then used as FFG by the applicable WFOs. In the NWRFC and CNRFC, FFPI values are similarly used for assigning initial basin susceptibilities and are then scaled to produce the final flash flood guidance values for transmission to WFOs.

c. *Gridded Flash Flood Guidance*

In 2005 and 2006, the ABRFC deployed a new method of producing FFG known as “gridded FFG” (GFFG) (Schmidt et al. 2007). This method imitates LFFG but increases FFG’s spatial resolution to that of the HRAP grid (the cells are nominally four km on a side, though the actual dimensions vary with latitude). This resolution is much closer to the resolution of the FFMP basins and thus mitigates, but does not eliminate, the issue of scale mismatch between LFFG basins and FFMP basins noted earlier in this paper. Gourley et al. (2012) found that the distribution of GFFG values over the ABRFC domain from 2006 to 2008 was roughly comparable with the distribution of LFFG values over the same area and time period, which was an anticipated result as discussed in Schmidt et al. (2007). In 2007 and 2008, the GFFG method was extended to other RFCs, including the Lower Mississippi RFC (LMRFC), the Southeast RFC (SERFC), and the West Gulf RFC (WGRFC). By the end of 2008, GFFG was in use across the entire southeastern and south-central United States.

GFFG uses a distributed hydrologic model to monitor the soil moisture component of FFG, unlike the older lumped model used in LFFG (Schmidt et al. 2007). Like LFFG, the GFFG method requires a soil moisture model, a rainfall-runoff model, and the determination of ThreshR values. Land use and soil type datasets are combined to yield an NRCS (Natural Resources Conservation Service) curve number (CN). Higher curve numbers are associated with larger runoff generation potential and therefore with a greater flash flood potential. These curve numbers are then adjusted to account for recent soil moisture conditions (Schmidt et al. 2007). In GFFG, this is accomplished by calculating a saturation ratio for each grid cell. The HL-RDHM (NWS Hydrology

Laboratory Research Distributed Hydrologic Model) is run in continuous mode and the upper zone tension and free water contents (UZTWC and UZFWC) parameters are obtained (Gourley et al. 2012). Then the maximum possible values for each of these parameters are estimated using a method outlined in Koren et al. (2000). The two model parameters are added and the ratio of these parameters to their maximum possible values becomes the saturation ratio at each grid cell (Schmidt et al. 2007). This saturation ratio is used to adjust the NRCS CN to a final value.

The final remaining variable in the GFFG system is ThreshR. In this method, a three-hour design rainfall event that corresponds to a five-year return period is used to produce the precipitation for input to the curve number model. The runoff produced by this model is then treated as the flow at flood stage (Schmidt et al. 2007). The unit hydrograph peak is found using the NRCS curve number method (Gourley et al. 2012), and requires basin slope, rainfall duration, soil moisture conditions, basin area, rainfall duration, and other characteristics (Schmidt et al. 2007). Then, as in the LFFG method, ThreshR is the ratio of flow at flood stage to the unit hydrograph peak. Schmidt et al. (2007) note that the GFFG ThreshR is lower in high elevations and higher in low elevations and contains greater spatial variability than the legacy ThreshR values.

The final calculation of GFFG values is accomplished using the adjusted curve number,  $S$ , and the ThreshR values at each grid cell. The following equation:

$$Q = \frac{(P - 0.2S)^2}{P + 0.8S} \quad (1)$$

is solved for  $P$ , the precipitation, where  $Q$  is the ThreshR value, and  $S$  is the soil moisture-adjusted curve number.  $P$  is the final gridded flash flood guidance value (Schmidt et al. 2007).

298 *d. Distributed Flash Flood Guidance*

299 A fourth type of FFG is generated at the Middle Atlantic RFC. It uses the continuous-  
300 API (Antecedent Precipitation Index) model on the HRAP grid for the soil moisture  
301 component of FFG. The net result is a spatially distributed FFG product (called “DFFG”  
302 in this study) with similar spatial variability to the GFFG method derived at ABRFC.

303  
304 **4. Methodology**

305 *a. Study Domain*

306 The spatial domain of the study covers the entire CONUS (see Fig. 1) while the  
307 temporal range of the study is 1 October 2006 to 31 August 2010. The start time of the  
308 study was fixed due to the availability of flash flooding reports from the online NWS  
309 Performance Management system; only reports of flash flooding starting on or after 1  
310 October 2006 are available through that website. The end date of the study was set by end  
311 of the date range of the USGS stream flow measurements that make up the flash flood  
312 event database described in Gourley et al. (2013).

313 *b. Datasets*

314 Four major datasets were obtained to complete this study – quantitative precipitation  
315 estimates (QPE), the FFG products, and two observation datasets – one derived from  
316 NWS *Storm Data* flash flood reports and another derived from USGS stream gauge  
317 measurements.

318 The QPE data used in this project are the hourly Stage IV products generated  
319 operationally at all CONUS RFCs. Hourly Stage IV accumulations were available for all  
320 of the CONUS except for the state of Washington, the northern third of Oregon, and the

Idaho Panhandle. Stage IV precipitation estimates include information from the WSR-88Ds and rain gauges and benefit from manual quality control procedures undertaken by RFC personnel (Gourley et al. 2012). Algorithms used to generate Stage IV products are not consistent across the country. At most RFCs, the multisensor precipitation estimator (MPE) is used operationally. However, the three western RFCs use the Mountain Mapper methodology and the ABRFC uses the P1 methodology (Lin 2007). Differences in the algorithms used to produce Stage IV precipitation estimates across the country likely contribute to some of the skill differences observed between FFPI and the other three FFG methods. It is also important to note that, in operations, the WFOs use radar-derived precipitation estimates and not Stage IV to monitor basins for FFG exceedance, so this study cannot exactly replicate the operational conditions under which FFG is normally used. However, Stage IV is a manually quality-controlled product and is the most accurate national precipitation estimate archived by NCEP (Lin 2007). The NCEP archive contained Stage IV mosaics for more than 99.7 percent of the hours in the study period.

Mosaics of operational FFG were obtained from the National Precipitation Verification Unit (NPVU). The FFG mosaics cover the time period from 1 October 2006 to 31 August 2010. FFG mosaics are produced at NPVU every six hours (at 00, 06, 12, and 18 UTC), but the regular issuance schedule varies between RFCs (see Table 1). A method was developed to fill in the gaps in the national FFG mosaic for times and RFC domains where FFG was not issued or was not available. In this procedure, the most recent valid FFG issuance for a given RFC was copied forward to the next mosaic time if no new FFG issuance was available. For instance, in cases when ABRFC did not issue



FFG at 06 UTC, their FFG from 00 UTC was copied forward in time and used to populate the 06 UTC FFG mosaic. In some situations, a RFC may not have issued any FFG product in a twenty-four hour period. In that case, those areas were left blank in the FFG mosaics for that day. NPVU archives contained national FFG mosaics for more than 97 percent of the total time in the study period (see Fig. 2). The dark areas in the western US are due to situations where no FFG products were issued on certain days.

Additionally, some RFCs changed their methods for generating FFG during the study period (see Table 2). These dates and times were established via visual examination of the FFG grids in GIS software and will be used to divide the nation geographically and spatially into “LFFG”, “DFFG”, “GFFG”, and “FFPI” areas in the results section of this paper.

All flash flooding reports recorded by the National Weather Service in *Storm Data* between 1 October 2006 and 31 December 2011 were downloaded from the online NWS Performance Management system and processed for use in this study. This resulted in 19,419 reports from across the entire US. After filtering the database to include only those flash floods caused by heavy rains, occurring in the CONUS, and during the study period, 14,827 *Storm Data* events remained (see Fig. 3). These reports are recorded by NWS forecasters and are used by the NWS to produce verification statistics regarding flash flood warnings. The reports contain information about timing (start and end time), location (WFO, state, county, NWS region, time zone, latitude, longitude, and distance to nearest place name), meteorological conditions, injuries, fatalities, and monetary damages. The source of these reports varies, but most are from emergency management officials, law enforcement, trained spotters, off-duty NWS employees, broadcast media,

and the public. The NWS observations fall into two categories: point-based and storm-based. Because storm-based observations are not of a standard size, all NWS observations (both point-based and storm-based) were normalized to a circular area of 49 HRAP grid cells for this study. Gourley et al. (2013) contains additional information about the character and development of this NWS report database.

Gourley et al. (2013) obtained an archive of stream flow data from July 1927 to September 2010 for 10,106 gauges operated by the USGS. We consider only small basins on the flash flood scale (contributing drainage area of less than 260 km<sup>2</sup>), and those to which the NWS has assigned an “action stage”. “Action Stage” is defined by Helble (2010) in NWS Manual 10-950 as “the stage which when reached by a rising stream, lake, or reservoir represents the level where the NWS or a partner/user needs to take some type of mitigation action in preparation for possible significant hydrologic activity.” Action stage is used as the stage height of interest in this study because it allows for the consideration of more events than could have been included if “minor flood” stage was used instead. Although action stage is defined differently than bankfull conditions, the two are extremely close at many gauged sites. On average, for the gauges selected for use in this study for which both bankfull and action stage information is available, the action stage height is 0.05 meters less than the bankfull height. Only those sites with at least one action stage exceedance during the study period are included. This resulted in a total of 244 gauges being used in the analysis. Each action stage exceedance is treated as an observed flash flood; 2,244 of these events make up the final dataset used in the USGS analysis portion of this study. These USGS events occur in all twelve of the CONUS RFC domains and the included stream gauge sites are located at different elevations and

in different hydroclimatic regimes (see Fig. 4). Each gauge is associated with an average of nine flooding events during the roughly four-year study period. The USGS data does not contain floods associated with ungauged basins and overland flow.

### *c. Flash Flood Guidance Evaluation Procedure*

In normal operations, forecasters update FFG values several times a day (see Table 1). FFG values are predicated upon changes in soil moisture occurring as a result of previous rainfall. Consider an FFG mosaic nominally valid at 0000 UTC. If heavy rainfall and/or flooding are ongoing at the time this mosaic was issued, a timeseries of FFG values will exhibit a sharp dip. This is because the new FFG mosaic (0000 UTC) takes into account six hours (or possibly more, if the RFC in question updates FFG less than every six hours) of antecedent rainfall not included in the previously-issued FFG mosaic. The heavy rainfall occurring between 1800 and 0000 UTC has the effect of reducing the FFG, because correspondingly less rainfall would be needed after 0000 UTC to cause flash flooding. Problems arising from these sharp differences in FFG values from issuance to issuance are mitigated in operations by simply resetting the QPE to which FFG is compared every time a new FFG grid is issued. However, in this evaluation, QPE cannot be “reset” with each new FFG grid because 3- and 6-hr FFG is compared to a rolling sum of hourly Stage IV QPE grids. Therefore, a precipitation-weighted FFG interpolation procedure outlined in Gourley et al. (2012) is used to create interim hourly mosaics of 1-, 3-, and 6-hr FFG. This procedure eliminates the problem of sharp jumps or changes in FFG values from issuance to issuance. In all cases, original FFG grids are preserved as they were originally issued. In the example above, the 1800

and 0000 UTC grids remain unchanged and only the interim products valid at 1900, 2000, 2100, 2200, and 2300 UTC are affected by the interpolation procedure. The following equation explains the interpolation process:

$$FFG_{current} = FFG_{next} w_t + FFG_{previous} (1 - w_t) \quad (2)$$

where  $FFG_{current}$  represents the interpolated FFG at that exact hour,  $FFG_{next}$  is the FFG product issued by the RFC at the next normal issuance time,  $FFG_{previous}$  is the FFG product issued by the RFC at the previous normal issuance time, and  $w_t$  is a precipitation weight given by the following equation:

$$w_t = \frac{\sum QPE \text{ from previous FFG issuance time to current time}}{\sum QPE \text{ from previous FFG issuance time to next FFG issuance time}} \quad (3)$$

The result of this is a series of national hourly FFG mosaics. For each hour, the Stage IV precipitation estimate is divided by the corresponding interpolated FFG mosaic. As in Gourley et al. (2012), FFG skill at QPE-to-FFG ratios of 0.5, 0.75, 1.0, 1.25, 1.5, 2.0, 2.5, and 3.0 is examined in this study.

The evaluation of FFG proceeds on a case-by-case basis (see Fig. 5). For each QPE-to-FFG ratio of interest, we search the interpolated hourly ratio mosaics for collections of adjacent HRAP grid cells where the ratio of interest is exceeded; each of these collections of grid cells is saved as a forecast flash flood event (also referred to hereafter as an “FFG event”). The ratios of interest are exceedance thresholds, not discreet intervals. Then the next ratio mosaic in time is searched, and if two events in subsequent mosaics overlap at all, they are combined into one event. This procedure continues throughout the entire study period. In this manner, a series of FFG events is saved for 1-, 3-, and 6-hr FFG products at each of the eight QPE-to-FFG ratios of interest. The HRAP grid cell in which

each FFG event is centered is recorded for comparison with the actual reported NWS flash flooding events.

In the USGS stage height evaluation, the rasterized USGS basins plotted on the HRAP grid are compared to the interpolated hourly QPE-to-FFG ratio grids. Ratio data for the HRAP grid cells that comprise a given USGS basin are extracted and stored for each hour. Then the mean QPE-to-FFG ratio is calculated at each hour for each USGS basin in the analysis. When the basin-mean ratio exceeds a threshold of interest (0.5, 0.75, 1.0, and the other values listed above), the start time, end time, and basin in which this exceedance occurred are recorded.

For both the NWS and USGS evaluations, the list of FFG events can simply be compared in space and time with the list of reported NWS *Storm Data* events or with the list of recorded USGS events. If an FFG event centroid falls within the search radius (the normalized 49 HRAP grid cell area described above) of an NWS event centroid, a “hit” is recorded. “Misses” occur when no FFG event centroid falls within the search radius of an NWS event centroid, and “false alarms” occur when there is an FFG event with no associated NWS event. For USGS events, no search radius is used. Instead, the centroid of an FFG event must be located within the drainage area of the gauge where the flooding event was recorded. As in Gourley et al. (2012), time buffers are applied to each USGS and NWS report. Eight hours are added to the start time of each report (two hours of this are for general uncertainty in the timing of the flooding reports and estimated rainfall and six hours of this allow heavy rainfall to translate into surface flooding impacts), and two hours are added to the end time of each report (again for uncertainty).

From hits, misses, and false alarms, a contingency table is populated (see Table 3 and Table 4). Then the standard metrics of probability of detection (POD), false alarm rate (FAR), and critical success index (CSI) (also called “skill” in this study) are computed.

$$POD = \frac{Hits}{Hits + Misses} \quad (4)$$

$$FAR = \frac{FalseAlarms}{Hits + FalseAlarms} \quad (5)$$

$$CSI = \frac{Hits}{Hits + Misses + FalseAlarms} \quad (6)$$

A CSI value of 1.0 indicates perfect forecast skill while a CSI of 0.0 indicates the forecast had no skill.

## 5. Results

### a. CONUS Skill of Operational Flash Flood Guidance

Skill indices are presented for operational 1-, 3-, and 6-hr FFG running over the CONUS from October 1, 2006 to August 31, 2010 at eight different QPE-to-FFG ratios and evaluated against NWS *Storm Data* reports (see Fig. 6). The CONUS-wide CSI of FFG ranges between 0.01 and 0.07; the latter value is achieved by 1-hr FFG at a QPE-to-FFG ratio of 1.5. In particular, for the higher QPE-to-FFG ratios (2.0, 2.5, and 3.0), the FFG tool fails to forecast, or catch, a large number of the observed *Storm Data* events. This means the probability of detection is much better for low QPE-to-FFG ratios but the effect is counterbalanced by a higher false alarm rate at those ratios. The net result is, in the NWS *Storm Data* analysis, that FFG performs best when considering moderate ratios. In general, 1- and 3-hr FFG display similar skill values at all ratios examined while 6-hr FFG is less skillful.

A similar analysis, but where FFG is evaluated using USGS stream gauge flood stage heights instead of NWS *Storm Data* reports, shows slightly higher skill values (see Fig. 7). Now the CSI of FFG ranges between 0.01 and 0.20, where the highest skill occurs with the 3-hr FFG product at a ratio of 0.5. Unlike in the NWS *Storm Data* analysis, here the FFG skill is highest at low QPE-to-FFG ratios and declines sharply with increasing ratio. Additionally, the 6-hr product is the most skillful and the 1-hr product is the least skillful, with 3-hr FFG generally falling somewhere in between the other two. Gourley et al. (2012) also reported these differences between USGS and NWS analyses of FFG skill, where the skill curves using USGS reports are pushed up and to the left compared with the NWS *Storm Data* skill curves. The high false alarm rates in the NWS analysis can be partially explained by underreporting of flash flood events in sparsely populated regions of the US. The authors are more confident about conclusions drawn from the NWS *Storm Data* analysis because there are many more events available in that dataset. Not all small basins gauged by the USGS have defined action stages, so the sample size of gauges and events available for the USGS analysis is much smaller than desired.

*b. Flash Flood Guidance Skill by River Forecast Center*

FFG values are generated at the RFC level, and different methods of generating FFG were in operational use during the study period. The number of NWS *Storm Data* events in each RFC domain varies, as does the number of events normalized for the area of each RFC domain (see Table 5). The three western RFCs (Northwest, California Nevada, and Colorado Basin) have the lowest number of events per 1,000 km<sup>2</sup>. This is probably due to underreporting of events—these areas have low population densities relative to the rest of

the US, and FFG in these areas displays very high false alarm rates, which could indicate that forecast events are not being observed. The event densities in the other nine RFCs display less variance suggesting underreporting of events is less of a problem in those areas.

When NWS *Storm Data* reports are used to evaluate FFG at a QPE-to-FFG ratio of 1.0, the DFFG product developed at the Middle Atlantic RFC performs well, with an observed CSI of 0.15 (see Fig. 8a and Table 6). The RFCs in the western US, using FFPI, exhibit skills ranging from 0.00 to 0.04. RFCs in roughly the northern half of the CONUS east of the Rockies, generally using LFFG, have CSIs between 0.07 and 0.12. Finally, those RFCs using GFFG, generally located in the southern half of the CONUS east of the Rockies, have CSIs between 0.05 and 0.07.

If the evaluation is expanded to include any QPE-to-FFG ratios (see Fig. 8b and Table 6), the MARFC's skill improves to 0.19. The skill of the western RFCs improves slightly, now ranging from 0.00 to 0.05. Improvement is also noted in the RFCs using LFFG (0.07 to 0.16) and those using GFFG (0.08 to 0.13).

If USGS flood stage height reports are used to evaluate FFG, a different picture emerges. At the standard QPE-to-FFG ratio of 1.0, the three western RFCs all have a CSI of 0.00 (see Fig. 9a and Table 7). The CSI of DFFG in the MARFC domain (0.16 here) is roughly the same as it was in the NWS *Storm Data* analysis (0.15). The skill indices of the RFCs running GFFG all improve, some in dramatic fashion. In this analysis, those values now range from 0.12 to 0.33. Finally, some RFCs running LFFG improve their CSI numbers in this analysis, while others decline, ranging from 0.04 to 0.19,



If the analysis is expanded to include all QPE-to-FFG ratios, FFPI skill at the three western RFCs does not improve (see Fig. 9b and Table 7). Some improvement in GFFG areas is noted when additional ratios are considered, with CSIs ranging from 0.18 to 0.43. There is also improvement over the LFFG domains, where CSIs now range from 0.16 to 0.27. Finally, the DFFG product used in the MARFC has a skill of 0.22 in this analysis. The USGS data suffers from a small sample size (see Table 8), and are less likely to include the difficult forecast locations associated with very small basins, urban runoff, wildfire scars, and overland flows. For these reasons, more credence should be given to conclusions drawn from the NWS *Storm Data* analysis.

#### *c. Flash Flood Guidance Skill by Generation Method*

Because this study evaluates whichever version of FFG was being produced operationally at the various RFCs, each generation method cannot be directly compared, since each type of FFG was running over different regions and different times. However, we divide the CONUS into four spatial and temporal regions; these represent the times and places where FFPI, LFFG, GFFG, and DFFG were operational. We provide statistics about each method without providing any judgment on the relative utility of each.

FFPI was operational at the Northwest RFC and the Colorado Basin RFC during the entire study period and at the California Nevada RFC after November 2008. Using NWS *Storm Data* flash flooding reports ( $N = 820$ ) to verify the skill of FFPI reveals the minimal skill of the product at all QPE-to-FFG ratios considered (see Fig. 10b). The maximum skill of FFPI ( $CSI = 0.02$ ) is achieved when the 3-hr product is used at a ratio of 0.75. At all ratios, high false alarm rates, ranging from 99% to 100%, and low

probabilities of detection, ranging from 1% to 40%, contribute to the low CSI of FFPI. The poor skill of the FFPI method can only be partially explained by the low population densities and less frequent reporting of events. Poor radar coverage over the western US results in difficulties in the production of the Stage IV precipitation estimates used in the study. Some areas of the West rely on climatological precipitation altered by rain gauge data to produce Stage IV estimates. Additionally, flash floods in this region are often caused by meteorological systems of a different character than the organized convective systems in the central and eastern CONUS. This also helps explain the low skill of FFPI, which was observed in both the NWS *Storm Data* and USGS flood stage height analyses.

LFFG was operational at eight of the twelve CONUS RFCs at the start of the study period, but was replaced by GFFG at three of these offices and by FFPI at a fourth. A total of 7,760 NWS *Storm Data* reports are included in this analysis. LFFG is most skillful at ratios of 1.5 or 2.0 (see Fig. 10c) and the 1- and 3-hr products are generally more skillful than the 6-hr ones. False alarm rates for this product range from 85% to 99%, and probabilities of detection range from 8% to 40%.

GFFG was developed at the Arkansas Red Basin RFC in 2005 and then later deployed at the West Gulf RFC, Lower Mississippi RFC, and Southeast RFC. In these domains, during the times in which GFFG was operational, a total of 5,530 flash flooding events were recorded in NWS *Storm Data*. At all QPE-to-FFG ratios (see Fig. 10d), the 1-hr GFFG product is the most skillful, with 3-hr GFFG less skillful and 6-hr GFFG even less skillful than 3-hr GFFG. The best GFFG skill is observed when the high ratios (over 2.0) are considered, though once a 2.0 ratio of is reached, neither 2.5 nor 3.0 provides an improvement or a drop-off in skill.

DFFG, which was only produced at the Middle Atlantic RFC during the study period, has the same spatial variability as GFFG but is generated using a different hydrologic model. There were 710 reported NWS *Storm Data* flash flood events in the MARFC domain, which is a much smaller dataset compared to the other analyses. Nonetheless, at QPE-to-FFG ratios between 1.0 and 2.0, DFFG performs well (see Fig. 10a). The best DFFG skill in the NWS *Storm Data* evaluation occurs when using 1-hr DFFG at ratio of 1.25, which results in a CSI of 0.19.

## 6. Conclusions

This study establishes the benchmark skill of the operational flash flood guidance (FFG) product used by the National Weather Service (NWS) to forecast, monitor, and warn the public about dangerous flash flooding events. Although flash flood guidance has been produced by the NWS for over forty years, little literature about its performance outside of isolated case studies exists. Using a CONUS-wide observational database of flash flooding events consisting of two separate sources, FFG was evaluated on a national scale, on an RFC scale, and by the various methods used to generate the product operationally. This evaluation covers a four-year period with events occurring in all but one state of the CONUS (Washington), including over 2,200 instances of flash flooding recorded by USGS stream gauges and over 14,000 instances of flooding recorded by NWS forecasters.

The NWS *Storm Data* reports include thousands of observed events because that database is intended to be comprehensive. Subjectivity due to the human element must be considered in any analysis relying upon NWS *Storm Data* reports. The USGS reports, on

the other hand, suffer from small sample sizes. Only a portion of the available USGS stream gauges was used in this study, because not all gauged locations have defined action stages and not all of those locations experienced an action stage exceedance during the study period. However, those flash floods that were recorded will be highly reliable because of the automated nature of the USGS observations. For this reason, the best approximation of the skill of FFG at forecasting flash floods and near-flash floods is the CONUS-wide USGS analysis.

Using NWS *Storm Data* reports as verification, FFG (all methods combined) achieved a maximum skill ( $CSI=0.07$ ) using the 1-hr accumulation product at a QPE-to-FFG ratio of 1.5. When using the USGS flood stage heights as the verification source, a maximum skill of 0.20 occurred with the 1-hr accumulation product for a QPE-to-FFG ratio of 0.5. This latter value should serve a benchmark skill for FFG on the national scale in subsequent research.

Four different methods for deriving FFG exist within the River Forecast Centers (RFCs). We grouped them according to FFG generation method being either distributed (DFFG), lumped (LFFG), gridded (GFFG), or flash flood potential index (FFPI), and compared their skill. This intercomparison is not objective because each method was running in different locations and so the events and sample sizes are quite different. When using the NWS *Storm Data* reports of flash flooding, LFFG, the oldest current method of FFG generation in use at the end of the study, performed best when its 3-hr version was used with a ratio of 1.5 or 2.0. GFFG, a newer method with some higher-resolution components, was most skillful when its 1-hr version was used at ratios of 2.0 and higher. DFFG, though only used at the Middle Atlantic RFC, reached a skill of 0.19

when the 1-hr product was evaluated against NWS *Storm Data* reports while using a ratio of 1.25. FFPI, used in the West, had CSI values below 0.02 for all ratios. Since these methods were not operating at the same times and in the same places, factors like topography, radar coverage, and population density prevent us from ranking the relative skill of any of them.

Future research in this area should continue, specifically looking at new ways of generating FFG with advanced distributed hydrologic models. More observational datasets, including additional small-scale USGS gauged basins, could be used to produce more detailed evaluations of FFG. Finer-scale observations like those collected by the Severe Hazards Analysis and Verification Experiment (Gourley et al. 2010) could also be used to build upon this study and may also support the more widespread adoption of modern, distributed methods of generating products to supplement or replace the current FFG system.

National Weather Service forecasters working in areas covered by the FFPI method of FFG generation should continue to use a wide range of information in the flash flood monitoring and warning process as relying solely on FFPI-generated FFG may have undesirable results. RFCs currently running the LFFG system should consider transitioning to GFFG or DFFG since the overall distribution of forecast values is similar, but GFFG and DFFG produce higher resolution information on a scale similar to that used in radar precipitation estimates and in the flash flood basins used by the Flash Flood Monitoring and Prediction program. NWS forecasters should remain aware of locations when a QPE-to-FFG ratio of 1.0 is exceeded while recognizing that the skill of the guidance product is potentially maximized at either a higher or lower ratio depending on

the part of the country and the FFG generation method being considered. Additionally, NWS forecasters can be trained to modify FFG within their county warning area in an effort to increase the usefulness of the current product until more permanent improvements can be made. These modifications are currently undertaken at many WFOs across the United States, particularly in urban areas, but they are not centrally archived and thus were not evaluated as a part of this study.

Over more than forty years, flash flood guidance has been a critical link in the system that protects Americans and their property from the most dangerous storm-related hazard. Many years of modifications have resulted in a patchwork of different generation methods and ideas about how FFG should work. The results of this study are not intended to discourage the storm-scale hydrologic community or to disparage the current state of FFG generation. Instead, the groundwork is being laid for the meteorological and hydrological community to explore large-scale improvements to operational FFG in the hope of improving scientific understanding of flash flooding events and of making flash flood forecasts more specific, more accurate, and more useful.

**Acknowledgements**

Funding was provided by the NOAA/Office of Oceanic and Atmospheric Research under NOAA–University of Oklahoma Cooperative Agreement NA17RJ1227, the U.S. Department of Commerce and the National Weather Service’s Advanced Hydrologic Prediction System funds. We appreciate the former NWS National Precipitation Verification Unit for making the QPE and GFFG grids available to us. We are also grateful to the USGS for access to the database of streamflow observations. Comments from three anonymous reviewers improved the content and readability of the manuscript. We appreciate their time.

**References**

- Arthur, A., G. Cox, N. Kuhnert, D. Slayter, and K. Howard, 2005: The National Basin Delineation Project. *Bull. Amer. Meteor. Soc.*, **86**, 1443-1452.
- Ashley, S. and W. Ashley, 2008: Flood fatalities in the United States. *J. Appl. Meteor. Climatol.*, **47**, 806-818.
- Carpenter, T., J. Sperflage, K. Georgakakos, T. Sweeney, and D. Fread, 1999: National threshold runoff estimation utilizing GIS in support of operational flash flood warning systems. *J. Hydrol.*, **224**, 21-44.
- Davis, R., 2003: Some practical applications of Flash Flood Monitoring and Prediction. Preprints, *17<sup>th</sup> Conf. on Hydrology*, Long Beach, CA, Amer. Meteor. Soc., J4.7. [Available online at <http://www.ams.confex.com/ams/pdfpapers/55695.pdf>.]
- , 2004: Locally modifying flash flood guidance to improve the detection capability of the Flash Flood Monitoring and Prediction program. Preprints, *18<sup>th</sup> Conf. on Hydrology*, Seattle, WA, Amer. Meteor. Soc., J1.2. [Available online at <http://www.ams.confex.com/ams/pdfpapers/68922.pdf>.]



688 -----, 2007: Detecting the entire spectrum of stream flooding with the flash flood  
 689 monitoring and prediction (FFMP) program. Preprints, *21<sup>st</sup> Conf. on Hydrology*,  
 690 San Antonio, TX, Amer. Meteor. Soc., 6B.1.  
 691  
 692 Gourley, J. J., J. Erlingis, T. Smith, K. Ortega, and Y. Hong, 2010: Remote collection and  
 693 analysis of witness reports on flash floods. *J. Hydrol.*, **394**, 53-62.  
 694  
 695 -----, -----, Y. Hong, and E. Wells, 2012: Evaluation of tools used for monitoring  
 696 and forecasting flash floods in the United States. *Wea. Forecasting*, **27**, 158-173.  
 697  
 698 -----, Y. Hong, Z. Flamig, A. Arthur, R. Clark, M. Calianno, I. Ruin, T. Ortel, M.  
 699 Wieczorek, E. Clark, and W. Krajewski, 2013: A unified flash flood database  
 700 over the US. *Bull. Amer. Meteor. Soc.*, **94**, 799-805.  
 701  
 702 Helble, T., 2010: Operations and services: Hydrologic services program, NWSPD 10-9:  
 703 Definitions and general terminology. NWS Manual 10-950, 5 pp. [Available  
 704 online at <http://www.nws.noaa.gov/directives/sym/pd01009050curr.pdf>.]  
 705  
 706 Koren, V., M. Smith, D. Wang, and Z. Zhang, 2000: Use of soil property data in the  
 707 derivation of conceptual rainfall-runoff model parameters. Preprints, *15<sup>th</sup> Conf. on*  
 708 *Hydrology*, Long Beach, CA, Amer. Meteor. Soc., 103-106.  
 709

710 Lin, Y., 2007: Q&A about the new NCEP Stage II/Stage IV. Mesoscale Modeling  
 711 Branch, Environmental Modeling Center, National Centers for Environmental  
 712 Prediction. [Available online at  
 713 <http://www.emc.ncep.noaa.gov/mmb/ylin/pcpanl/QandA>]  
 714  
 715 Maddox, R., 1980: Mesoscale convective complexes. *Bull. Amer. Meteor. Soc.*, **61**, 1374-  
 716 1387.  
 717  
 718 Mogil, H., J. Monro, and H. Groper, 1978: NWS's flash flood warning and disaster  
 719 preparedness programs. *Bull. Amer. Meteor. Soc.*, **59**, 690-699.  
 720  
 721 National Weather Service, cited 2012: NWS Glossary. [Available online at  
 722 <http://w1.weather.gov/glossary/index.php>.]  
 723  
 724 River Forecast Center Development Management Team, 2003: Flash flood guidance  
 725 improvement team – Final report. Report to the Operations Subcommittee of the  
 726 NWS Corporate Board, 47 pp. [Available online at  
 727 <http://www.nws.noaa.gov/oh/rfcdev/docs/ffgitreport.pdf>]  
 728  
 729 Schmidt, J., A. Anderson and J. Paul, 2007: Spatially-variable, physically-derived, flash  
 730 flood guidance. Preprints, *21<sup>st</sup> Conf. on Hydrology*, San Antonio, TX, Amer.  
 731 Meteor. Soc., 6B.2. [Available online at  
 732 <http://ams.confex.com/amspdfpapers/120022.pdf>]

733

734 Smith, G., 2003: Flash flood potential: Determining the hydrologic response of FFMP  
735 basins to heavy rain by analyzing their physiographic characteristics. Report to  
736 the NWS Colorado Basin River Forecast Center, 11 pp. [Available online at  
737 [http://www.cbrfc.noaa.gov/papers/ffp\\_wpap.pdf](http://www.cbrfc.noaa.gov/papers/ffp_wpap.pdf)]

738

739 Sweeney, T., 1992: Modernized areal flash flood guidance. NOAA Tech. Rep. NWS  
740 HYDRO 44, Hydrologic Research Laboratory, National Weather Service, NOAA,  
741 Silver Spring, MD, 21 pp. and an appendix.

742

743 ----- and T. Baumgardner, 1999: Modernized flash flood guidance. Report to NWS  
744 Hydrology Laboratory, 11 pp. [Available online at  
745 <http://www.nws.noaa.gov/oh/hrl/ffg/modflash.htm>]

746

## List of Tables

TABLE 1 – Flash flood guidance issuance schedule for each RFC.

TABLE 2 – Type of flash flood guidance produced by each RFC, where changes in generation method are noted if applicable.

TABLE 3 – Contingency table used to evaluate flash flood forecasts.

TABLE 4 – Contingency table for the CONUS-wide NWS *Storm Data* evaluation of FFG at a QPE-to-FFG ratio of 1.0.

TABLE 5 – Number of events per RFC in NWS *Storm Data* during the study period.

TABLE 6 – Statistics, based on NWS *Storm Data* reports, used in development of Figure 8. The three leftmost columns show the best CSI and corresponding POD and FAR associated with a QPE-to-FFG ratio of 1.0. The three rightmost columns should the best CSI and corresponding POD and FAR associated with any QPE-to-FFG ratio.

TABLE 7 – Statistics, based on USGS stream gauges, used in development of Figure 9.

The three leftmost columns show the best CSI and corresponding POD and FAR associated with a QPE-to-FFG ratio of 1.0. The three rightmost columns should the best CSI and corresponding POD and FAR associated with any QPE-to-FFG ratio.

770     TABLE 8 – Number of events per RFC in the USGS state height exceedance database.

771

772 TABLE 1 – Flash flood guidance issuance schedule for each RFC.

<b>RFC</b>	<b>00 UTC</b>	<b>06 UTC</b>	<b>12 UTC</b>	<b>18 UTC</b>
<i>Arkansas Red-Basin</i>	Yes	When necessary	Yes	Yes
<i>California Nevada</i>	Stopped in 2007	Stopped in 2007	Yes	Stopped in 2007
<i>Colorado Basin</i>			Yes	
<i>Lower Mississippi</i>	Yes	Yes	Yes	Yes
<i>Middle Atlantic</i>	Yes	When necessary	Yes	Yes
<i>Missouri Basin</i>	Yes	When necessary	Yes	Yes
<i>North Central</i>	Yes	Yes	Yes	Yes
<i>Northeast</i>	Yes	Stopped in 2008	Yes	Yes
<i>Northwest</i>			Yes	
<i>Ohio</i>	Yes	When necessary	Yes	Yes
<i>Southeast</i>	Yes	Yes	Yes	Yes
<i>West Gulf</i>	Yes	Yes	Yes	Yes

773

774

775 TABLE 2 – Type of flash flood guidance produced by each RFC, where changes in  
 776 generation method are noted if applicable.

<b>RFC</b>	<b>Study Period Start (Oct 1 2006)</b>	<b>Study Period End (Aug 31 2010)</b>	<b>Date/Time of Switch</b>
<i>Arkansas Red-Basin</i>	GFFG	GFFG	Did not switch
<i>California Nevada</i>	LFFG	FFPI	16 Aug 2007 12 UTC
<i>Colorado Basin</i>	FFPI	FFPI	Did not switch
<i>Lower Mississippi</i>	LFFG	GFFG	13 Aug 2007 12 UTC
<i>Middle Atlantic</i>	DFFG	DFFG	Did not switch
<i>Missouri Basin</i>	LFFG	LFFG	Did not switch
<i>North Central</i>	LFFG	LFFG	Did not switch
<i>Northeast</i>	LFFG	LFFG	Did not switch
<i>Northwest</i>	FFPI	FFPI	Did not switch
<i>Ohio</i>	LFFG	LFFG	Did not switch
<i>Southeast</i>	LFFG	GFFG	6 Mar 2008 12 UTC
<i>West Gulf</i>	LFFG	GFFG	23 Oct 2007 18 UTC

777

778

779      TABLE 3 – Contingency table used to evaluate flash flood forecasts.

		<b>Was the event observed (by either the NWS or the USGS)?</b>	
		Yes	No
<b>Was the event forecast by FFG?</b>	Yes	Hit	False alarm
	No	Miss	Correct negative

780

781



782 TABLE 4 – Contingency table for the CONUS-wide NWS *Storm Data* evaluation of  
 783 FFG at a QPE-to-FFG ratio of 1.0.

		<b>Was the event observed (by the NWS)?</b>	
		Yes	No
<b>Was the event forecast by FFG?</b>	Yes	5751 Hits	88,560 False alarms
	No	9076 Misses	
<b>Total:</b>		14827 Events	

784

785

786 TABLE 5 – Number of events per RFC in NWS *Storm Data* during the study period.

<b>RFC</b>	<b>Number of NWS Storm Data Events</b>	<b>Area of RFC Domain (km<sup>2</sup>)</b>	<b>Number of Events per 1,000 km<sup>2</sup></b>
<i>Arkansas Red-Basin</i>	1,413	185,000	7.6
<i>California Nevada</i>	189	226,000	0.8
<i>Colorado Basin</i>	549	273,000	2.0
<i>Lower Mississippi</i>	2,482	175,000	14
<i>Middle Atlantic</i>	710	76,400	9.3
<i>Missouri Basin</i>	2,014	512,000	3.9
<i>North Central</i>	1,896	337,000	5.6
<i>Northeast</i>	540	103,000	5.2
<i>Northwest</i>	90	321,000	0.3
<i>Ohio</i>	1,551	161,000	9.6
<i>Southeast</i>	1,156	210,000	5.5
<i>West Gulf</i>	2,140	333,000	6.4
<b><i>Totals:</i></b>	<b><i>14,730</i></b>	<b><i>2,912,400</i></b>	<b><i>5.1</i></b>

787

788

TABLE 6 – Statistics, based on NWS *Storm Data* reports, used in development of Figure 8. The three leftmost columns show the best CSI and corresponding POD and FAR associated with a QPE-to-FFG ratio of 1.0. The three rightmost columns should the best CSI and corresponding POD and FAR associated with any QPE-to-FFG ratio.

RFC	Best CSI at a Ratio of 1.0			Best CSI at any Ratio		
	CSI	POD	FAR	CSI	POD	FAR
<i>Arkansas Red-Basin</i>	0.05	0.43	0.94	0.09	0.27	0.88
<i>California Nevada</i>	0.00	0.10	1.00	0.00	0.09	1.00
<i>Colorado Basin</i>	0.04	0.12	0.95	0.05	0.09	0.92
<i>Lower Mississippi</i>	0.06	0.42	0.93	0.10	0.27	0.86
<i>Middle Atlantic</i>	0.15	0.37	0.80	0.19	0.35	0.71
<i>Missouri Basin</i>	0.07	0.39	0.92	0.12	0.19	0.77
<i>North Central</i>	0.12	0.32	0.84	0.14	0.23	0.72
<i>Northeast</i>	0.07	0.17	0.90	0.07	0.23	0.92
<i>Northwest</i>	0.01	0.11	0.99	0.02	0.10	0.98
<i>Ohio</i>	0.08	0.34	0.91	0.16	0.24	0.66
<i>Southeast</i>	0.07	0.45	0.92	0.13	0.24	0.77
<i>West Gulf</i>	0.06	0.42	0.93	0.08	0.21	0.88

795 TABLE 7 – Statistics, based on USGS stream gauges, used in development of Figure 9.  
 796 The three leftmost columns show the best CSI and corresponding POD and FAR  
 797 associated with a QPE-to-FFG ratio of 1.0. The three rightmost columns should the best  
 798 CSI and corresponding POD and FAR associated with any QPE-to-FFG ratio.

RFC	Best CSI at a Ratio of 1.0			Best CSI at any Ratio		
	CSI	POD	FAR	CSI	POD	FAR
<i>Arkansas Red-Basin</i>	0.33	0.36	0.20	0.43	0.79	0.52
<i>California Nevada</i>	0.00	0.00	1.00	0.00	0.00	1.00
<i>Colorado Basin</i>	0.00	0.00	1.00	0.00	0.00	1.00
<i>Lower Mississippi</i>	0.17	0.26	0.68	0.18	0.40	0.75
<i>Middle Atlantic</i>	0.16	0.22	0.60	0.22	0.38	0.66
<i>Missouri Basin</i>	0.19	0.22	0.46	0.27	0.39	0.53
<i>North Central</i>	0.15	0.18	0.48	0.24	0.46	0.65
<i>Northeast</i>	0.04	0.05	0.80	0.16	0.42	0.80
<i>Northwest</i>	0.00	0.00	1.00	0.00	0.00	1.00
<i>Ohio</i>	0.14	0.22	0.72	0.17	0.36	0.76
<i>Southeast</i>	0.12	0.15	0.63	0.21	0.32	0.61
<i>West Gulf</i>	0.23	0.35	0.59	0.25	0.43	0.63

799  
 800

801      TABLE 8 – Number of events per RFC in the USGS stage height exceedance database.

<b>RFC</b>	<b>Number of USGS Events</b>	<b>Area of RFC Domain (km<sup>2</sup>)</b>	<b>Number of Events per 1,000 km<sup>2</sup></b>
<i>Arkansas Red-Basin</i>	33	185,000	0.18
<i>California Nevada</i>	11	226,000	0.05
<i>Colorado Basin</i>	30	273,000	0.11
<i>Lower Mississippi</i>	86	175,000	0.49
<i>Middle Atlantic</i>	290	76,400	3.8
<i>Missouri Basin</i>	227	512,000	0.44
<i>North Central</i>	464	337,000	1.4
<i>Northeast</i>	208	103,000	2.0
<i>Northwest</i>	7	321,000	0.02
<i>Ohio</i>	264	161,000	1.6
<i>Southeast</i>	361	210,000	1.7
<i>West Gulf</i>	243	333,000	0.72
<b><i>Totals:</i></b>	<b><i>2,224</i></b>	<b><i>2,912,400</i></b>	<b><i>0.76</i></b>

802

803

804

805

806

807

808

809

810

811

812

813

814

815

816

817

## List of Figures

FIG. 1. Map of the twelve CONUS River Forecast Centers with the domains of each outlined in bold.

FIG. 2. Percentage of the study period in which timely FFG data was available.

FIG. 3. NWS *Storm Data* CONUS flash flooding observations occurring between 1 Oct 2006 and 31 Aug 2010. Point reports (generally recorded prior to 30 Sep 2007) are plotted in purple and storm-based polygon reports (generally recorded after 1 Oct 2007) are plotted in blue. RFC domain boundaries are in bold.

FIG. 4. Locations of USGS stream gauges used in the study are plotted with orange marks. Included gauges have a contributing drainage area of less than 260 km<sup>2</sup>, have been assigned “action stage” stage heights by the NWS, and had at least one instance of the action stage being exceeded during the study period. RFC domain boundaries are in bold.

FIG. 5. General schematic of the FFG event selection and evaluation process using NWS *Storm Data* reports. In the top left panel, 1-hr Stage IV precipitation for the state of Oklahoma is displayed for 19 Aug 2007 06z. At top right, the 1-hr flash flood guidance (FFG) product valid at the same time is displayed. The precipitation grid is divided by the FFG grid to produce the ratio grid in the bottom panel. All contiguous ratio grid cells over 1.0 are selected as a FFG event. Then all NWS *Storm Data* events recorded within

two hours before, or eight hours after, the valid time of the ratio grid are used to determine the performance of FFG. *Storm Data* reports are shown in black, with a circle around each to represent the search radius used. The single black arrows refer to events where FFG correctly forecast the flash flood. The single red arrows refer to events where FFG failed to properly forecast the flash flood.

FIG. 6. The CONUS-wide skill of flash flood guidance for a variety of exceedance ratios. Observations are from reports of flash flooding in NWS *Storm Data* between 1 Oct 2006 and 31 Aug 2010.

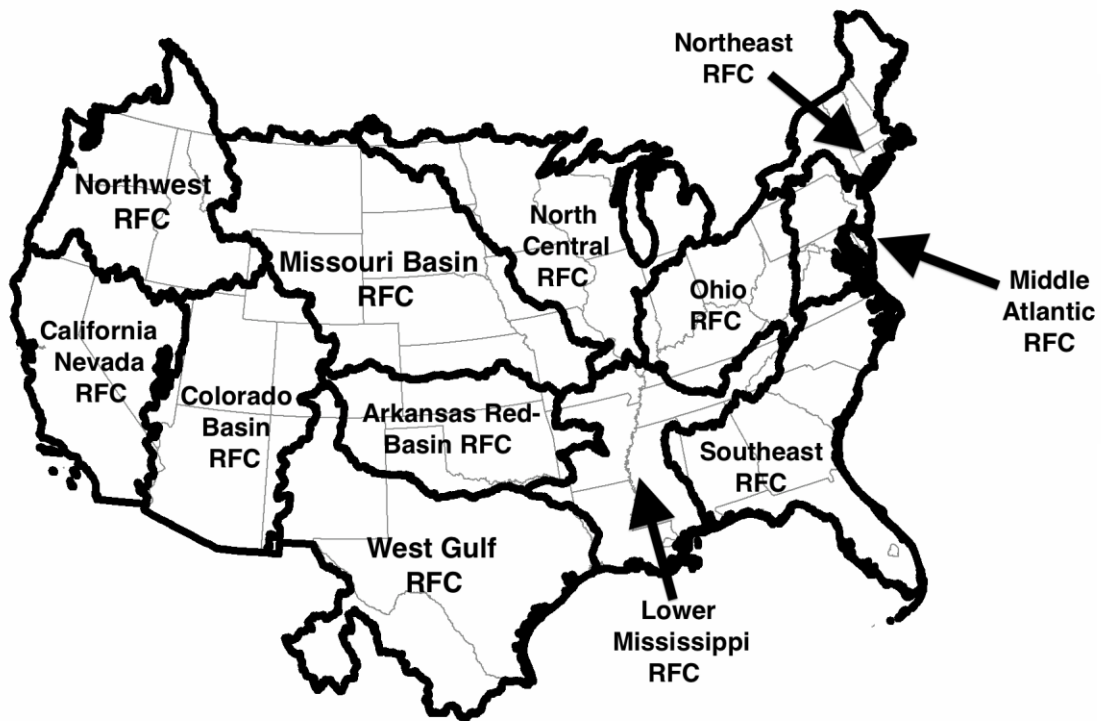
FIG. 7. As in Fig. 6, but flash flood observations are from exceedances of action stage stage heights at USGS stations plotted in Fig. 4 for basin-mean QPE to FFG ratio.

FIG. 8. Map of FFG skill as verified by NWS *Storm Data* flash flooding reports when (a) a QPE to FFG ratio of 1.0 is considered and (b) when any QPE to FFG ratio is considered.

FIG. 9. Map of FFG skill as verified by USGS stream gauge measurements when (a) a QPE to FFG ratio of 1.0 is considered and (b) when any QPE to FFG ratio is considered.

FIG. 10. The skill of (a) DFFG, (b) FFPI, (c) LFFG, and (d) GFFG for a variety of exceedance ratios. Observations are from reports of flash flooding in NWS *Storm Data* between 1 Oct 2006 and 31 Aug 2010.

864



865

866 FIG. 1. Map of the twelve CONUS River Forecast Centers with the domains of each  
867 outlined in bold.



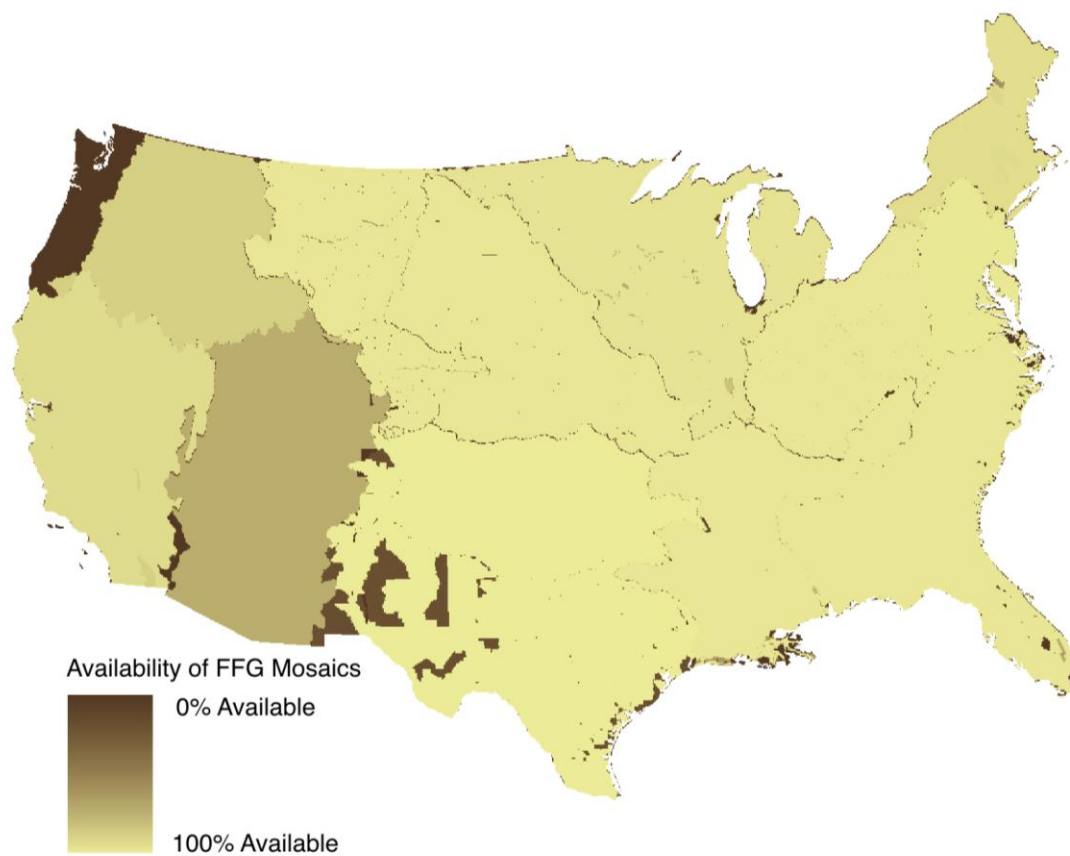
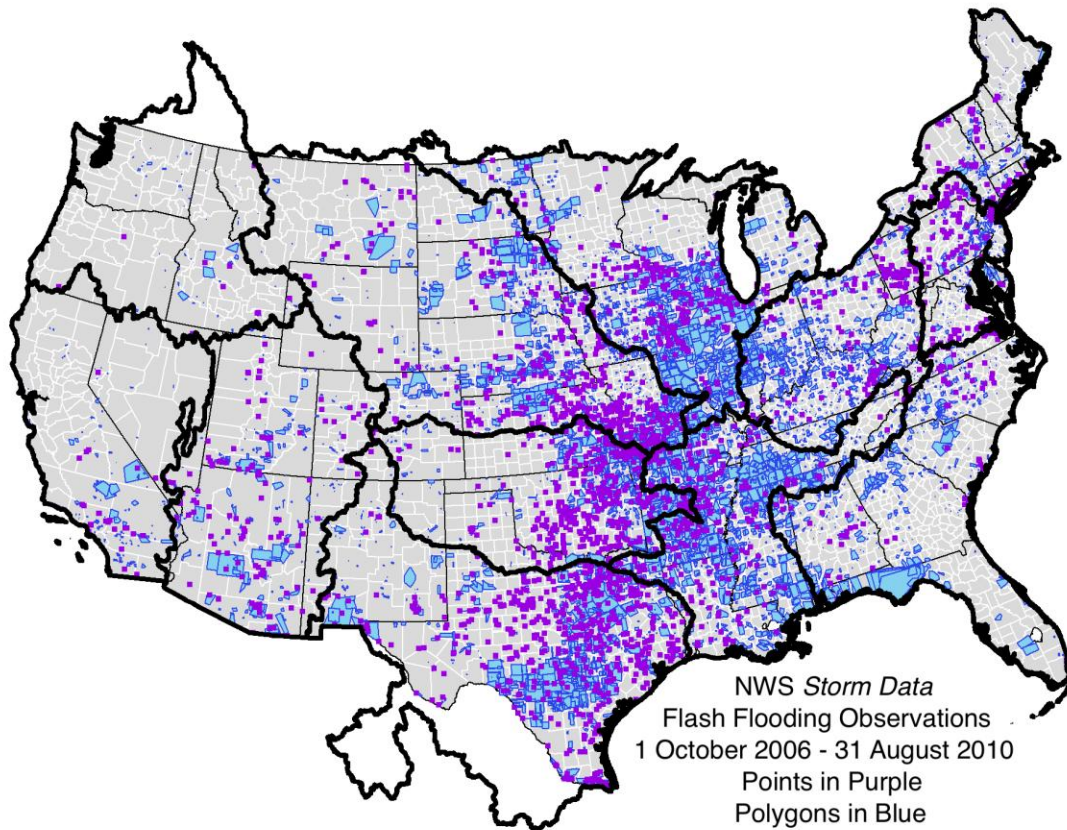
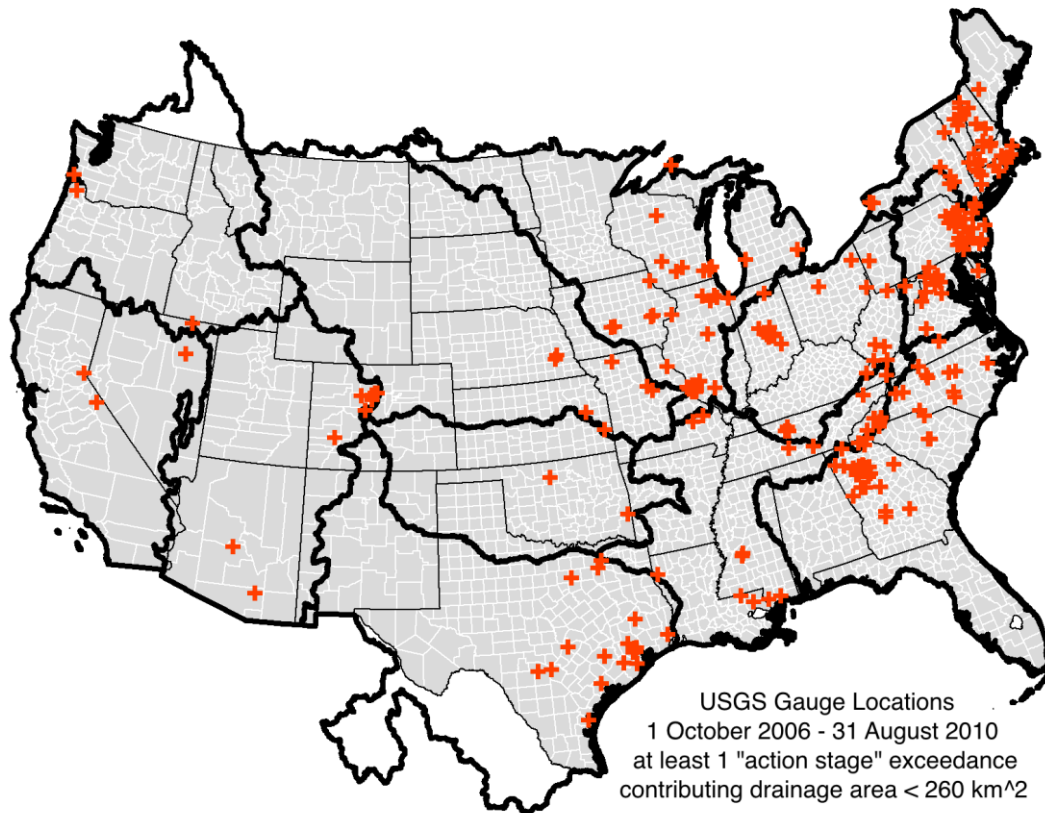


FIG. 2. Percentage of the study period in which timely FFG data was available.



870

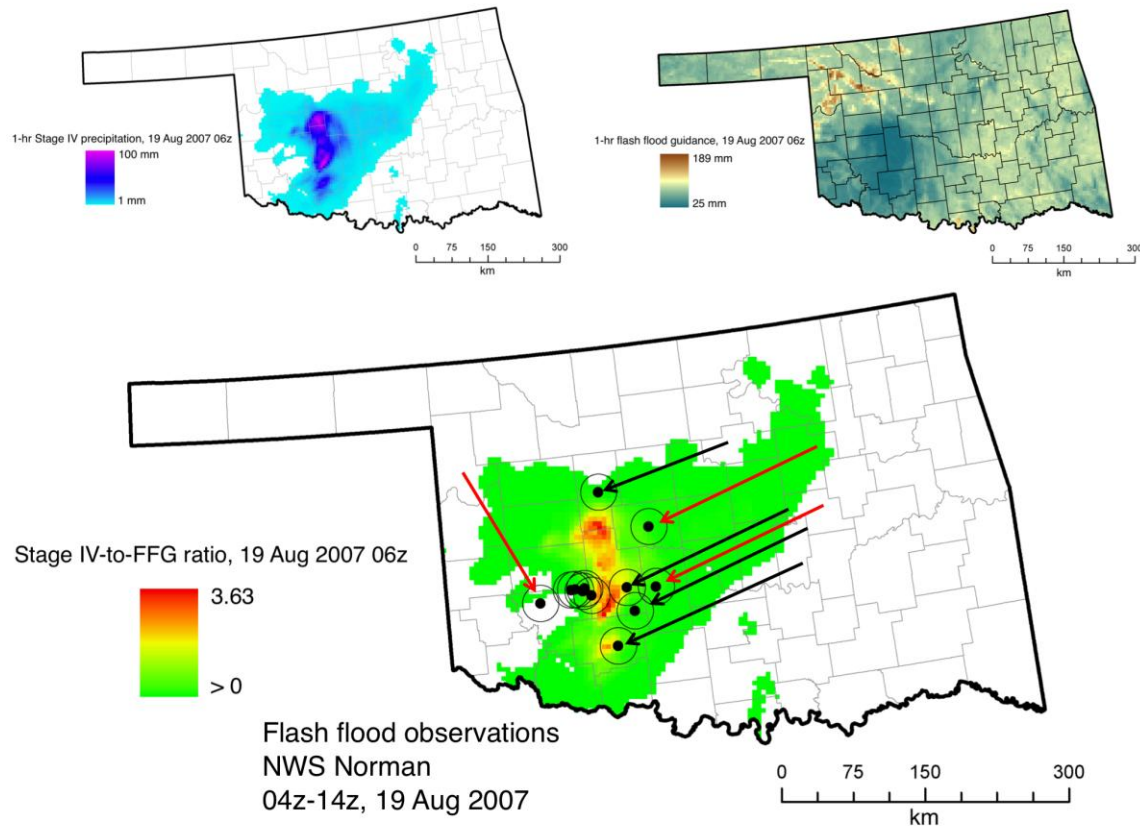
871 FIG. 3. NWS *Storm Data* CONUS flash flooding observations occurring between 1 Oct  
 872 2006 and 31 Aug 2010. Point reports (generally recorded prior to 30 Sep 2007) are  
 873 plotted in purple and storm-based polygon reports (generally recorded after 1 Oct 2007)  
 874 are plotted in blue. RFC domain boundaries are in bold.



875

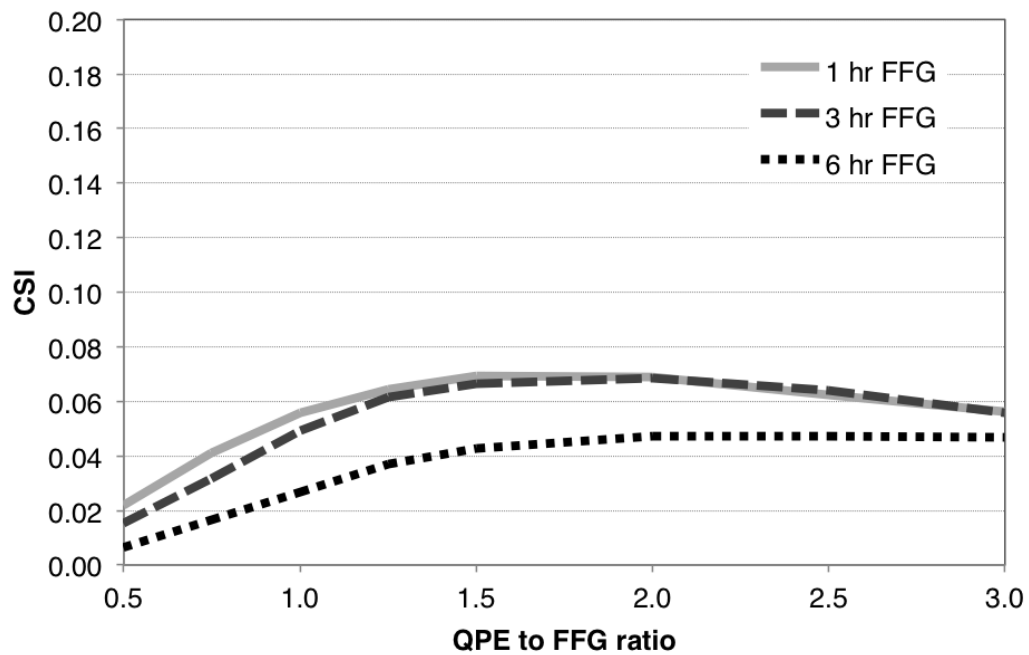
876 FIG. 4. Locations of USGS stream gauges used in the study are plotted with orange  
 877 marks. Included gauges have a contributing drainage area of less than 260 km<sup>2</sup>, have been  
 878 assigned “action stage” stage heights by the NWS, and had at least one instance of the  
 879 action stage being exceeded during the study period. RFC domain boundaries are in bold.  
 880 244 gauges with a total of 2,244 events are included, for an average of 9 events per gauge  
 881 during the study period.

882



883

884 FIG. 5. General schematic of the FFG event selection and evaluation process using NWS  
 885 *Storm Data* reports. In the top left panel, 1-hr Stage IV precipitation for the state of  
 886 Oklahoma is displayed for 19 Aug 2007 06z. At top right, the 1-hr flash flood guidance  
 887 (FFG) product valid at the same time is displayed. The precipitation grid is divided by the  
 888 FFG grid to produce the ratio grid in the bottom panel. All contiguous ratio grid cells  
 889 over 1.0 are selected as a FFG event. Then all NWS *Storm Data* events recorded within  
 890 two hours before, or eight hours after, the valid time of the ratio grid are used to  
 891 determine the performance of FFG. *Storm Data* reports are shown in black, with a circle  
 892 around each to represent the search radius used. The single black arrows refer to events  
 893 where FFG correctly forecast the flash flood. The single red arrows refer to events where  
 894 FFG failed to properly forecast the flash flood.



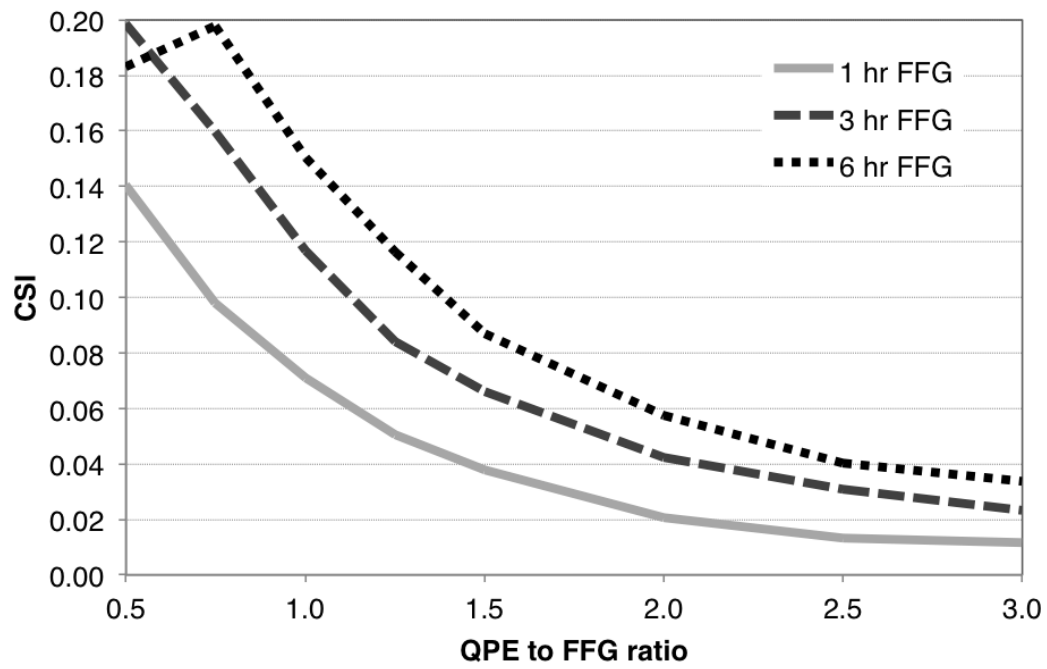
895

896 FIG. 6. The CONUS-wide skill of flash flood guidance for a variety of exceedance ratios.

897 Observations are from reports of flash flooding in NWS *Storm Data* between 1 Oct 2006

898 and 31 Aug 2010.

899



900

901 FIG. 7. As in Fig. 6, but flash flood observations are from exceedances of action stage

902 stage heights at USGS stations plotted in Fig. 4 for basin-mean QPE to FFG ratio.



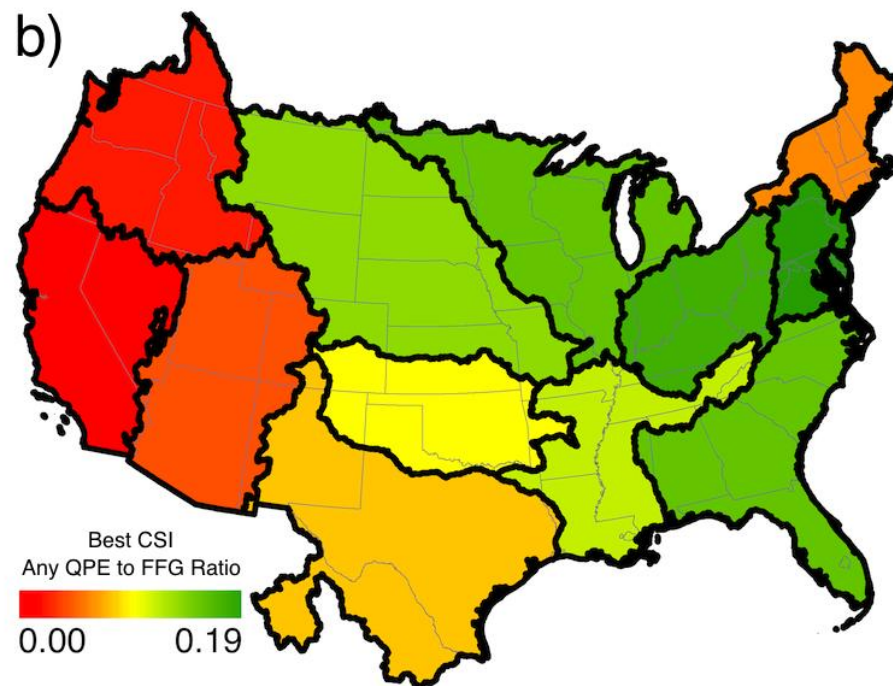
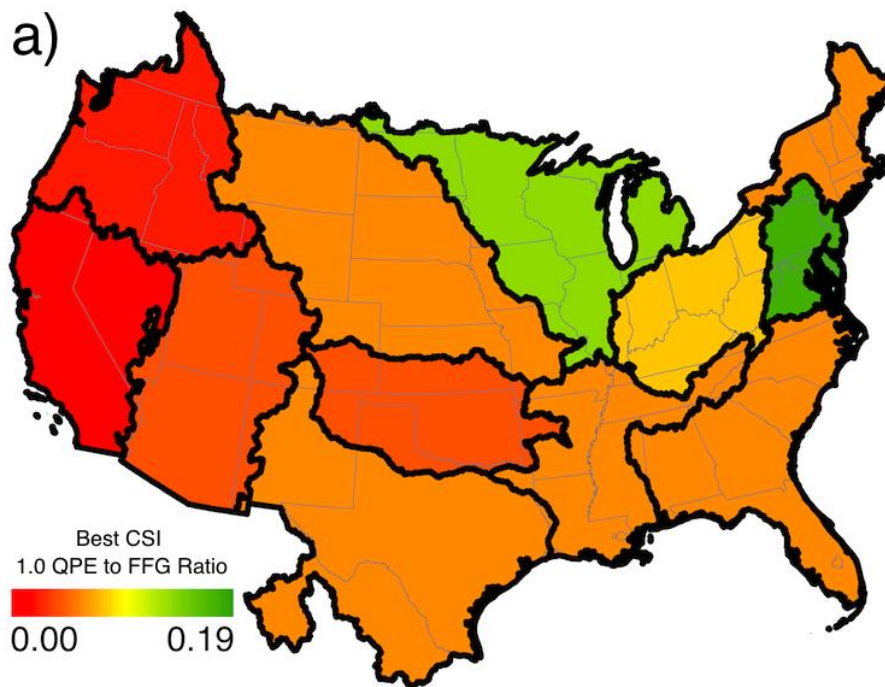
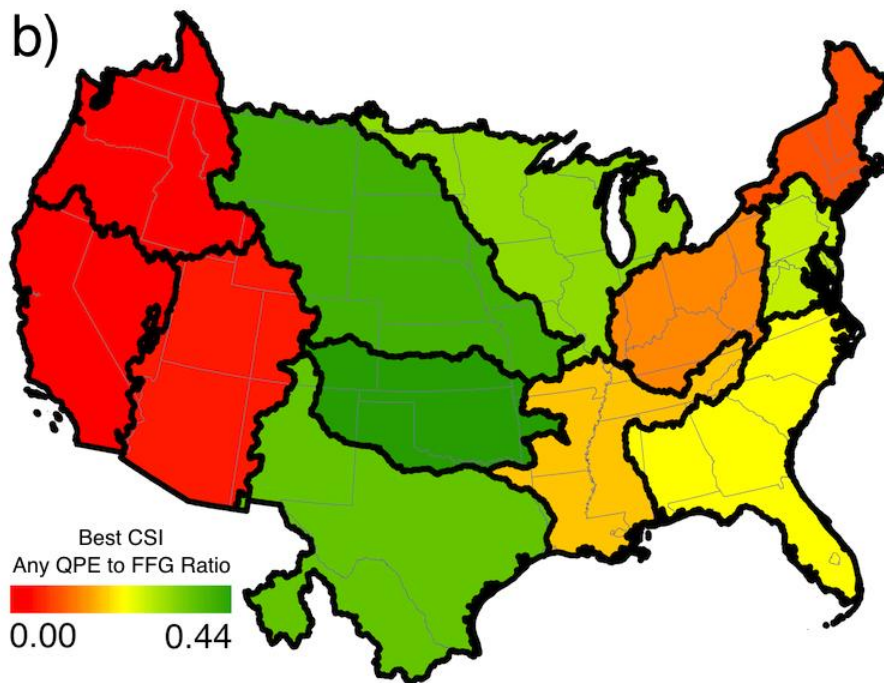
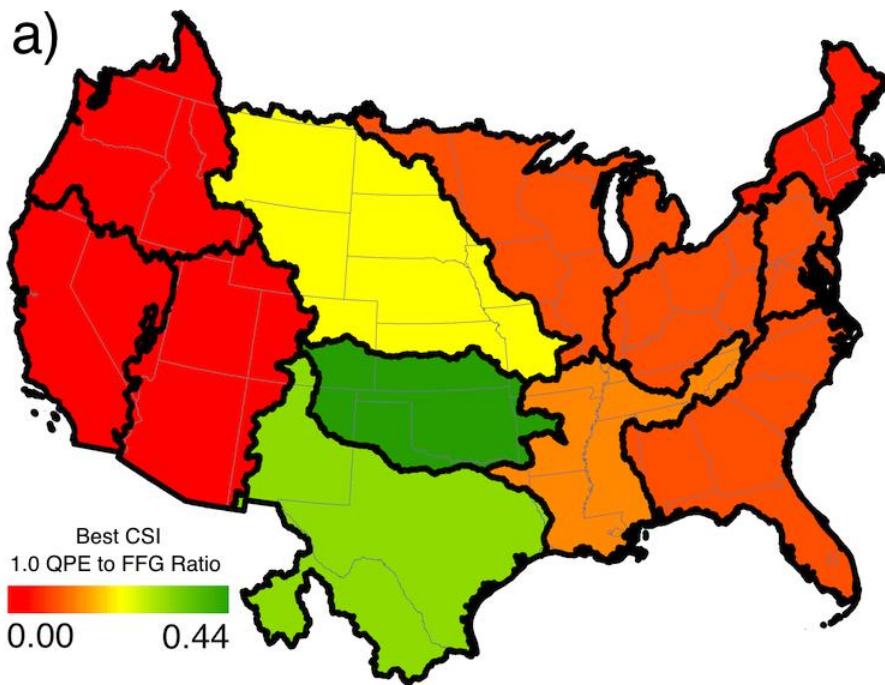


FIG. 8. Map of FFG skill as verified by NWS *Storm Data* flash flooding reports when (a) a QPE to FFG ratio of 1.0 is considered and (b) when any QPE to FFG ratio is considered.



907

908 FIG. 9. Map of FFG skill as verified by USGS stream gauge measurements when (a) a

909 QPE to FFG ratio of 1.0 is considered and (b) when any QPE to FFG ratio is considered.



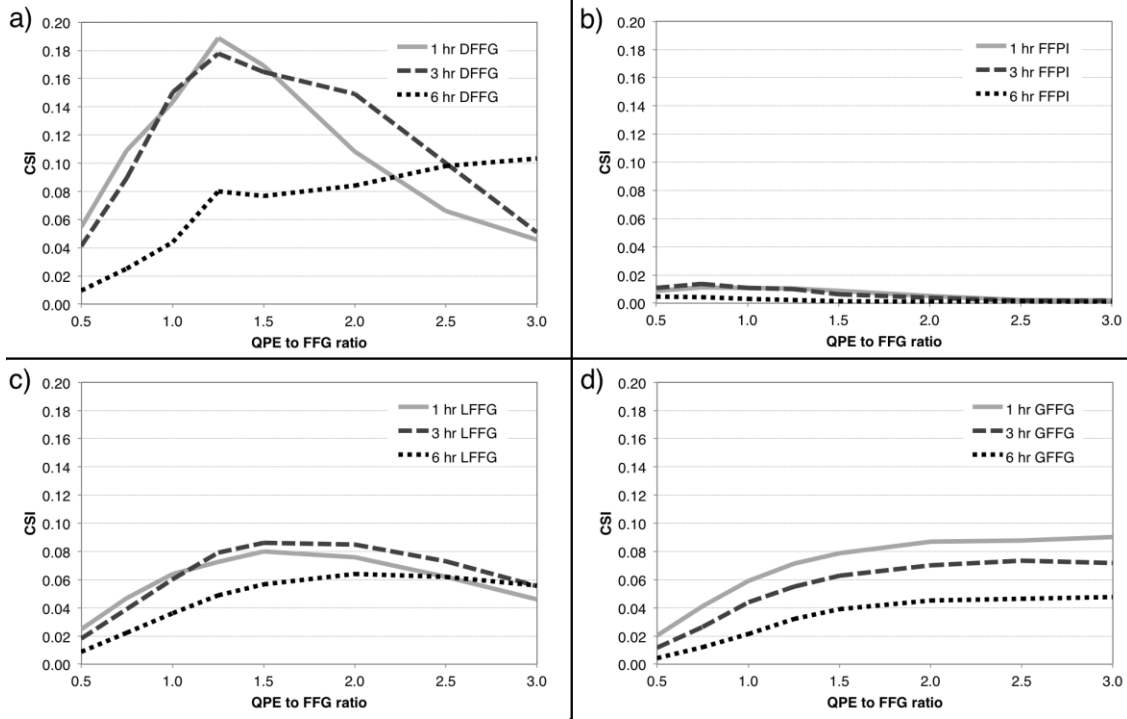


FIG. 10. The skill of (a) DFFG, (b) FFPI, (c) LFFG, and (d) GFFG for a variety of exceedance ratios. Observations are from reports of flash flooding in NWS *Storm Data* between 1 Oct 2006 and 31 Aug 2010.