



**HAL**  
open science

## A Unified Flash Flood Database over the US

J. J. Gourley, Y. Hong, Z. L Flamig, A. Arthur, R. Clark, M. Calianno, I. Ruin, T. Ortel, M. E Wiczorek, P.-E. Kirstetter

► **To cite this version:**

J. J. Gourley, Y. Hong, Z. L Flamig, A. Arthur, R. Clark, et al.. A Unified Flash Flood Database over the US. Bulletin of the American Meteorological Society, 2013, 94 (6), pp.799-805. 10.1175/BAMS-D-12-00198.1 . hal-03091825

**HAL Id: hal-03091825**

**<https://hal.science/hal-03091825>**

Submitted on 14 Jan 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# A Unified Flash Flood Database across the United States

BY JONATHAN J. GOURLEY, YANG HONG, ZACHARY L. FLAMIG, AMI ARTHUR, ROBERT CLARK,  
MARTIN CALIANNO, ISABELLE RUIN, TERRY ORTEL, MICHAEL E. WIECZOREK,  
PIERRE-EMMANUEL KIRSTETTER, EDWARD CLARK, AND WITOLD F. KRAJEWSKI

Flash floods are rapid surface water responses over normally dry land to intense rainfall or a sudden release of water from a dam break or ice jam, and have significant impacts on transportation, infrastructure, and human safety. While in recent decades hydrometeorologists have significantly advanced our comprehension of synoptic and mesoscale environments conducive to intense rainfall, tools used by forecasters to predict the hydrologic response, location, timing, and magnitude of the social impact have not progressed commensurately. Predictability

is determined by the development and refinement of a physical system model, and it is strongly dependent on the available data. In this case, data involve a unified observational dataset of flash floods, which holds great potential for advancing the science and predictability of flash flooding. This article serves the following purposes: 1) to announce a recently assembled U.S.-wide database on flash flooding available in multiple, common formats to reach a wide range of users, and 2) to encourage others to submit additional, related datasets that can be incorporated into the database.

There is no single source of information that holistically describes flash flooding in the United States. Perhaps it is the diverse and discontinuous nature of flash-flooding impacts that makes them difficult to observe and subsequently catalogue in a consistent database. Flash floods differ from other weather-related hazards (e.g., tornadoes, hail) in that their impacts are strongly controlled by surface properties, infrastructure, and social exposure factors. Streamflow measurements operated and maintained by the U.S. Geological Survey (USGS) benefit from automation and high temporal resolution, resulting in long-term, continuous records at each gauge site. These instruments require electrical power and road access for communications, regular instrument maintenance, and manual measurements to empirically establish a rating curve (i.e., the relationship between the measured stage and the desired discharge). The costs associated with these requirements (on the order of \$10,000 per year per gauge site) imply that automated streamflow measurements are relatively uncommon in small basins where flash floods occur.

Forecasters at local National Weather Service (NWS) offices throughout the United States routinely collect reports of flash flooding from trained spotters, local authorities, and emergency management officials within their areas of responsibility. The

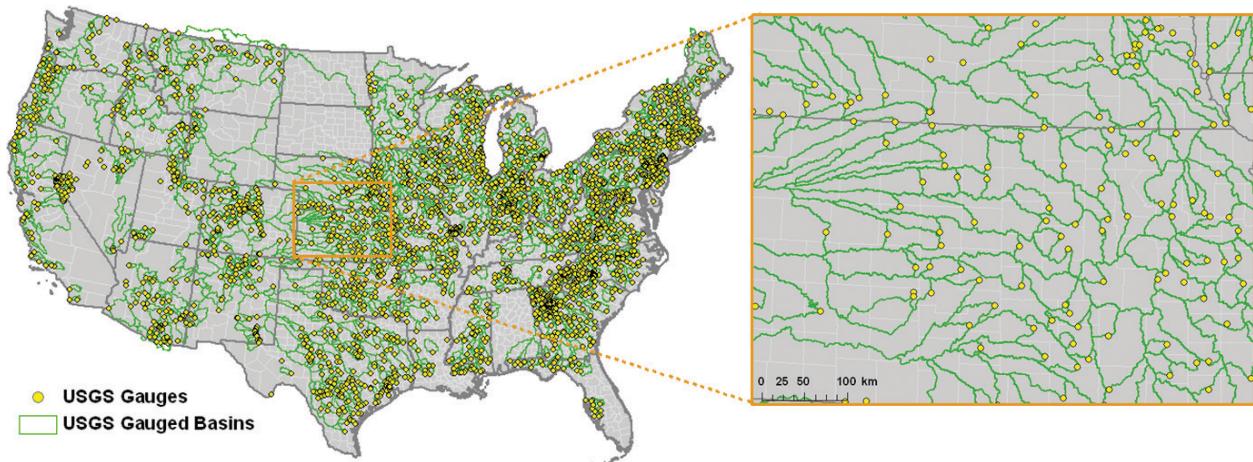
**AFFILIATIONS:** GOURLEY—NOAA/National Severe Storms Laboratory, Norman, Oklahoma; HONG—Department of Civil Engineering and Environmental Science, and Atmospheric Radar Research Center, University of Oklahoma, Norman, Oklahoma; FLAMIG AND KIRSTETTER—NOAA/National Severe Storms Laboratory, and Atmospheric Radar Research Center, University of Oklahoma, Norman, Oklahoma; ARTHUR—NOAA/National Severe Storms Laboratory, and Cooperative Institute for Mesoscale Meteorological Studies, University of Oklahoma, Norman, Oklahoma; R. CLARK—NOAA/National Severe Storms Laboratory, Atmospheric Radar Research Center, University of Oklahoma, and Cooperative Institute for Mesoscale Meteorological Studies, University of Oklahoma, Norman, Oklahoma; CALIANNO AND RUIN—Laboratoire d'étude des Transferts en Hydrologie et Environnement, Grenoble, France; ORTEL—United States Geologic Survey, Illinois Water Science Center, Urbana, Illinois; WIECZOREK—United States Geologic Survey, Maryland—DE-DC Water Science Center, Baltimore, Maryland; E. CLARK—National Weather Service Office of Climate, Water, and Weather Services, Silver Spring, Maryland; KRAJEWSKI—IIHR-Hydroscience & Engineering, The University of Iowa, Iowa City, Iowa

**CORRESPONDING AUTHOR:** Jonathan J. Gourley, National Weather Center, 120 David L. Boren Blvd, Norman, OK 73072-7303

E-mail: [jj.gourley@noaa.gov](mailto:jj.gourley@noaa.gov)

DOI:10.1175/BAMS-D-12-00198.1

©2013 American Meteorological Society



**FIG. 1. USGS streamflow stations (yellow dots) and their contributing drainage areas (outlined in green) for 2,948 gauges with flood events from Jul 1927 to Sep 2010. The time-series data have been converted to event data based on flows exceeding their predefined action stages. In total, there are 98,668 events comprising the USGS component of the flood database.**

NWS Storm Events Database is essential for evaluating and improving operational forecast products and procedures. Details contained in the reports, such as information about the meteorological environment, a dollar estimate of the damages, and number of fatalities, have yielded very useful information to the research community as well. Limitations of the database include 1) subjectivity in the reported event locations defined by often imprecise, bounding polygons; 2) times that are often related to the meteorological event rather than the flooding impacts; 3) reports are dependent on a person to witness the event; and 4) reports provide little or no information about the site's societal exposure or antecedent conditions.

The third database considered in this study comes from the Severe Hazards Analysis and Verification Experiment (SHAVE), which was conducted across the United States during the summers of 2008–10 at the National Severe Storms Laboratory (NSSL). Student callers obtained details on the specific type of impact, magnitude, and frequency of flash flooding directly from members of the public who responded to a telephone-based questionnaire. This georeferenced, survey-based approach overcomes some of the imprecision noted with the NWS Storm Events Database, but it is based on unreliable reports from the general public. Also, NSSL researchers designed the SHAVE data collection methodology for high-resolution, storm-targeted reports, and thus the dataset does not encompass all events at a given time.

The assembly of the three aforementioned datasets into a unified, consistent database retains the inherent

limitations associated with each one, yet the database combines the high-resolution details from SHAVE with the broad spatial coverage and event narratives from the NWS storm reports with the automated streamflow measurements from USGS to provide a more complete depiction of flash flooding across the United States. The database is freely available to the public at [www.nssl.noaa.gov/projects/flash/database.php](http://www.nssl.noaa.gov/projects/flash/database.php), and we provide it in three different formats for a variety of users who may be interested in quick-and-easy plots, detailed spatial investigations, or statistical analysis using the raw data.

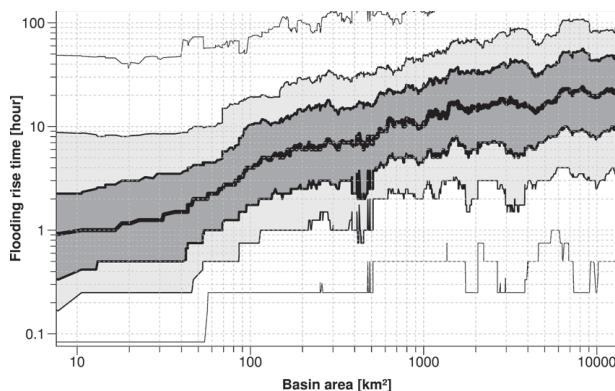
#### **DATA SOURCES AND PROCESSING STEPS.**

**USGS discharge measurements.** We obtained the USGS archive of instantaneous streamflow data from 10,106 gauges with records dating from July 1927 through September 2010 in a MySQL database format. The USGS maintains the instantaneous data for most stations beginning in the mid-1980s, and with data intervals commonly ranging from 5 to 60 min. The public can directly access this archive at <http://ida.water.usgs.gov/ida>. While applications such as calibrating and evaluating hydrologic models for water budget studies and water resources management require time-series data, these data require additional processing in order to define when flooding events occurred and to determine the magnitude of flooding. NWS offices have defined stage heights associated with stream conditions (bankfull; action; and minor, moderate, and major flooding) for 3,490 stream gauge locations. These thresholds were

defined in coordination with the local emergency management and stakeholder community and were based on impacts to lives and/or property. Often, the bankfull stage is the same as the minor flood stage (often referred to as the flood stage). However, in more rural areas, the flood stage may be greater than the bankfull stage due to the lack of infrastructure situated in close proximity to the streams.

We converted the USGS database to an “event-based” database for flood studies in the following manner. We identified all events that exceeded their predefined action stage for each station; this is most often the lowest stage height threshold defined by the NWS. According to the NWS Directive 10–950–Hydrologic Services Program Definitions and General Terminology, action stage is “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” ([www.nws.noaa.gov/directives/sym/pd01009050curr.pdf](http://www.nws.noaa.gov/directives/sym/pd01009050curr.pdf)). In total, there were 98,668 events in the database that exceeded action stages at 2,948 of the gauges in the USGS archive (see Fig. 1); 665 of these gauges have catchment areas less than 250 km<sup>2</sup>. For each event, we provide the following information: USGS Gauge ID, latitude (decimal degrees), longitude (decimal degrees), start time (UTC) at which the flow exceeded the action stage threshold, end time (UTC) when the flow dropped below the threshold, peakflow magnitude (m<sup>3</sup> s<sup>-1</sup>), peak time (UTC) at which peakflow occurred (UTC), and the difference between start time and peak time (in hours).

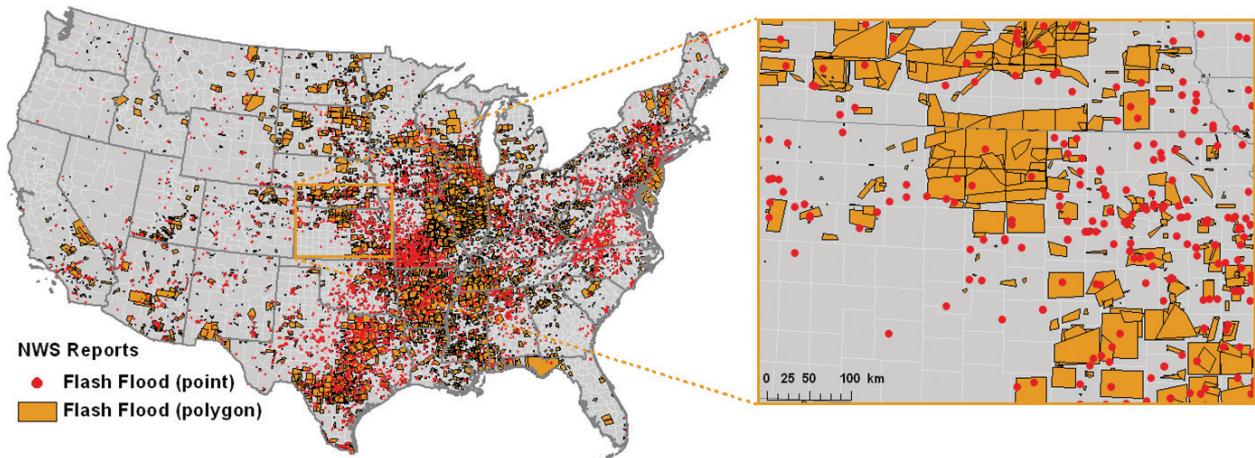
This latter variable, referred to hereafter as the flooding rise time, is a proxy for the time-to-rise and is plotted against basin catchment area in Fig. 2 for all 98,668 events combined. We see there is a clear relationship between the proxy “flashiness” of an event and catchment area. The spread, represented by the gray-shaded interquantile areas, results from different antecedent conditions, variable basin geomorphologies that impact their responses to rainfall, and events with rainfall that fell near the basin outlet. While there is no formal method to segregate flash floods from river floods, we recommend users apply a threshold to the flooding rise time rather than the catchment area. From the figure, we see that applying a threshold of 6 h corresponds to a median basin catchment area of approximately 250 km<sup>2</sup>, and is the recommended flooding rise time threshold for flash flood events.



**FIG. 2. Quantile plot of flooding rise time (in h), defined as the time of peak flow minus the time at which the action stage was reached, as a function of basin catchment area (in km<sup>2</sup>) for all 98,668 events. The thick black line represents the median (50% quantile), the dark gray-shaded region represents the area between the 25% and 75% quantiles, the light gray-shaded region represents the area between the 10% and 90% quantiles, and the thin lines represent the 1% and 99% quantiles.**

Along with the events dataset, we supply metadata for each station containing static information about the USGS station’s ID, latitude (decimal degrees), longitude (decimal degrees), hydrologic unit code (HUC), agency, degree of regulation, gauge name, drainage area (km<sup>2</sup>), contributing drainage area (km<sup>2</sup>), computed flows (m<sup>3</sup> s<sup>-1</sup>) for recurrence intervals for 2, 5, 10, 25, 50, 100, 200, and 500 yr, and computed flows (m<sup>3</sup> s<sup>-1</sup>) for action stage, minor, moderate, and major flooding. The USGS has previously computed flows for recurrence intervals (return periods) from 2 to 500 years using a Log-Pearson Type III distribution for those stations with at least 10 years of record and no significant changes in the record due to urbanization, diversion, or regulation. The degree of regulation field comes directly from USGS metadata for peakflow data and has values of either “Yes,” “No,” or “Undefined.”

Each station’s event data and metadata are grouped by a first-level, two-digit hydrologic unit code (HUC), which represents a basin scale at the regional level. We provide the processed USGS flood event database for HUC basins in the United States and Caribbean in the following three formats: 1) comma-delimited text file, 2) GIS shapefile, and 3) KMZ file for Google Earth. Use of KMZ files yields quick-and-easy displays, while the provision of shapefiles enables more in-depth spatial analysis using GIS software. The comma-delimited files can be read by a number of



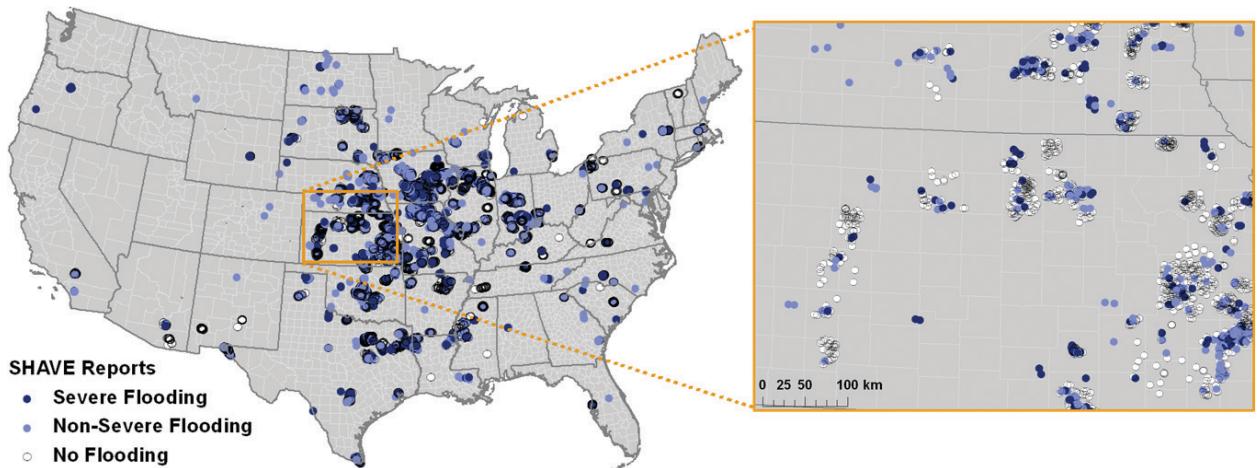
**FIG. 3. National Weather Service reports of flash flooding from 01 Oct 2006 to 31 Dec 2011 from the Stormdat program. The database includes 19,419 flash-flood reports.**

commonly available statistical software packages. Some users may also wish to access the text files directly for use in originally developed code and scripts. Data for Alabama, North Dakota, South Dakota, and Wyoming were not available when we obtained the database. Station data for these states will be added in future updates to the database (to be done on an annual basis).

**NWS storm reports.** We obtained all NWS reports of flash flooding from 1 October 2006 to 31 December 2011 from the database managed by the NWS Performance Branch. Prior to 1 October 2007, the NWS recorded storm reports by county; any instance of flash flooding yielded a recorded event for the county. Because counties are delineated primarily according to political, rather than physical, geographic boundaries and their sizes vary considerably across the United States, this procedure lacks the necessary spatial resolution for meaningful hydrological interpretation. NWS forecasters now report the locations of impacted regions using bounding polygons defined by as many as eight vertices. There was a transitional period in implementing the new procedures, so it is not uncommon for polygons to have the same shape as a county or to be recorded as a point, especially in 2006–20. In addition to the reports nominally being stored as a single county-wide point prior to 1 October 2007, there are also instances in which event locations were recorded with only two vertices. In the latter case, we converted the two points into a single polygon feature by assigning the vertices as the diagonal of a bounded rectangle.

According to the NWS *Storm Data* Directive, a recorded flash flood must have posed a potential threat to life or property and had a report of moving water with a depth greater than 0.15 m (6 in.) or more than 0.91 m (3 ft) of standing water. Typical situations that meet these criteria include rivers and streams out of their banks, evacuations, rescues, road closures, and floodwaters in an above-ground residence (i.e., not a basement). Each report from the NWS database contains a unique ID, the three-letter abbreviation of the NWS forecast office (WFO) that reported the event, beginning and ending time of event (UTC), state, county, NWS region, direct/indirect fatalities and injuries (if applicable), a dollar estimate of property and crop damage (if applicable), details about the event including its cause (e.g., heavy rain), source of report (e.g., law enforcement), event and episode narratives, and vertex coordinates in decimal degrees of latitude and longitude as well as the range (miles) and azimuth (e.g., NE) from the nearest city. We used the entries in the vertex coordinate fields to create individual polygons for display and analysis in Google Earth and GIS software. In this conversion, we noted that the maximum allowable characters in the event and episode narrative fields were often exceeded for shapefile and KMZ formats. In some cases, these narratives can be several sentences long. The full narratives are preserved in the comma-delimited format.

Figure 3 shows the locations of all 19,419 flash flood reports contained in the NWS database. The limitations inherent in this database include poor precision and accuracy in both the timing and spatial extents of flash flooding. Often, the meteorological



**FIG. 4. SHAVE reports of no flooding, nonsevere flooding, and severe flooding obtained from the public during the summers of 2008–10. In total, there are 9,369 reports contained within the SHAVE storm-targeted dataset.**

event timing is taken as flash flood timing. Regarding spatial extent, it's not clear that bounding polygons are appropriate for delineating impacted regions. Unlike rainfall amounts, flash-flood impacts are often spatially discontinuous and may be associated with difficult-to-contour features such as road networks. Consistent with the processing of the USGS dataset, we segregated the NWS flash-flood reports into regional, two-digit HUC basins. We provide files separately for point-based reports versus polygons. The same file formats used for the USGS dataset (i.e., comma-delimited text, GIS shapefile, and KMZ) are utilized for the NWS flash flooding reports.

*SHAVE questionnaire responses.* During the summers of 2008–10, the NSSL employed 5–6 undergraduate meteorology students to collect unique details on flash-flooding impacts at very high resolution. The experiment designers utilized WSR-88D-based warning products and flash-flood warnings issued by the NWS and displayed them in Google Earth to guide the SHAVE callers where flash flooding may have just occurred. Students initiated calls if rainfall exceeded flash-flood guidance, there was a NWS warning or advisory, or a survey response for a different hazard (i.e., hail) suggested flash flooding was a problem. Then, callers employed a purposeful sampling strategy to better refine the spatial extent of flash flooding. They used a georeferenced telephone database to call the public and initiate a questionnaire designed to obtain details about flash-flooding impacts, including the depth and movement of flood waters, lateral extent of water out of the stream, incidence of rescues

and evacuations, start and end times of impacts, respondent-estimated frequency of event, and types of impacts. The callers also included detailed comments to assess the uncertainty and validity of the reports as well as to include other anecdotal responses that didn't readily fit into one of the survey questions. The SHAVE dataset was postprocessed in order to better classify the impact types and to incorporate additional geographical attributes into each report, including land use, local terrain slope, contributing drainage area, compound topographic index (relates to dominant runoff process), and population density. Future efforts planned for SHAVE in 2013 include the collection of social science data on human behavior prior to and during flash-flooding events.

Figure 4 shows the spatial distribution of all 9,369 reports collected during SHAVE. SHAVE was operational only when undergraduate students were available on a full-time basis during the warm season from early May through the end of August in 2008–10. The SHAVE database differs from the NWS storm reports in that it is storm-targeted and point-based. The NWS reports are meant to encompass all flash-flooding events across the United States from 2006 to 2011, while the SHAVE reports are for specific storms. The SHAVE dataset provides for the assessment of false alarm rates (i.e., forecast of flood with no observed event) because it includes reports of "no flooding." In fact, this class comprises 73% of the total reports. Accounts from human reports, especially the untrained public, are subject to uncertainty due to perceptions and occasional embellishments, and must be used with caution. Users are encouraged to refer

to the supplied metadata for additional information about each field. In concert with the other databases, we provide the SHAVE data in comma-delimited text format, KMZ, and GIS shapefile format for each of the regional HUC basins.

*Other candidate data sources.* An additional goal of this article is to reach out to other agencies, universities, and companies who maintain datasets related to the observation of flash flooding. For example, active and passive microwave sensors positioned on the ground and in space (e.g., 37-GHz channel onboard Tropical Rainfall Measurement Mission, Moderate Resolution Imaging Spectroradiometer onboard the *Terra* and *Aqua* satellites, Advanced Spaceborne Thermal Emission and Reflection Radiometer onboard *Terra*) have shown the potential for monitoring inundated areas near streams and provide reasonable estimates of surface water fluxes and depths (see e.g., <http://oas.gsfc.nasa.gov/floodmap>). Local and regional networks of cameras or ultrasonic sensors mounted near bridges and tunnels can be very useful for detecting flash floods. There are also myriad networks of stream gauges and rain gauges, many of which are used by NWS offices, operated by cities, counties, districts, water boards, bureaus, private companies, and tribes. For example, Community Collaborative Rain, Hail and Snow Network (CoCoRaHS) observers are encouraged to report flooding along with their rainfall reports. In 2011, the Iowa Flood Center based at The

University of Iowa developed and deployed more than 100 fully automated and autonomous stream-stage sensors on local bridges. NWS forecasters use these data in real time, and we plan to add the data to the U.S. flash-flood database.

Postevent surveys serve as the primary basis for a flash-flood database already built for Europe. For instance, while the NWS storm events dataset generally contains surveyed events, further details from field investigations would certainly be useful to estimate peakflows. There is a wealth of information on flash-flood impacts that could be made available by insurance agencies. Thus far, social science data such as human exposure, behavior, perceived risks, and responses constitutes a significant missing component in the unified flash-flood database. Once social science data are included in SHAVE—a planned activity for 2013—we will be able to gain a more comprehensive understanding of flash-flood impacts through an end-to-end analysis of the physical and societal components. Social media have also been shown to provide potential for volunteer reporting by the public, including submission of georeferenced photographs of flooded lands. Finally, incorporation of long-term archives (decadal) of gridded rainfall datasets from WSR-88D-based products will be essential for understanding the role of the causative rainfall to the quickness and intensity of basin responses. We encourage those who maintain datasets like the ones described here to consider making them a part of the U.S. unified flash-flood database for freely accessible, community use.

**SUMMARY.** This article describes the data sources and processing steps to create a unified database of flash-flood observations across the United States that is now available to the community at [www.nssl.noaa.gov/projects/flash/database.php](http://www.nssl.noaa.gov/projects/flash/database.php). The database comprises streamflow observations maintained by the USGS, storm reports collected by the NWS from trained personnel, and public survey responses to a questionnaire developed for the SHAVE experiment. We have rigorously postprocessed all datasets for consistency in terms of data



THE UNIVERSAL SOUNDING DECODE & ANALYSIS PROGRAM

Great User Manual.  
Great Help Files.  
and now...

**GREAT VIDEOS!**

[www.raob.com/videos.php](http://www.raob.com/videos.php)

RAOB...for the most versatile and accurate sounding analyses available!  
Environmental Research Services, LLC • 570-491-4689 • [www.raob.com](http://www.raob.com)

formats, time formats, geographic projection, and units. The database will be updated once per year in order to include recent USGS streamflow data, storm reports from the NWS, and SHAVE.

We expect the announcement of this unified database will result in the inclusion of additional datasets relevant to flash-flood observation, specifically those that provide social science data. The overall goal is to provide a comprehensive observational database on flash-flood impacts, which will enhance the research community's understanding of the social, physical, and economic effects of flash flooding.

**ACKNOWLEDGMENTS.** 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 Advanced Hydrologic Prediction Service of the NOAA/National Weather Service. WFK acknowledges the support of the Iowa Flood Center.

## FOR FURTHER READING

Brakenridge, G. R., S. V. Nghiem, E. Anderson, and S. Chien, 2005: Space-based measurement of river runoff. *Eos, Trans. Amer. Geophys. Union*, **86** (19), 185–188.

Calianno, M., I. Ruin, and J. J. Gourley, 2013: Supplementing flash flood reports with impact classifications. *J. Hydrol.*, **477**, 1–16.

Gaume, E., and M. Borga, 2008: Post-flood field investigations in upland catchments after major flash floods: Proposal of a methodology and illustrations. *J. Flood Risk Manage.*, **1**, 175–189, doi:10.1111/j.1753-318X.2008.00023.

—, and Coauthors, 2009: A compilation of data on European flash floods. *J. Hydrol.*, **367**, 70–78.

Gourley, J. J., J. M. Erlingis, T. M. Smith, K. L. Ortega, and Y. Hong, 2010: Remote collection and analysis of witness reports on flash floods. *J. Hydrol.*, **394**, 53–62, doi:10.1016/j.jhydrol.2010.05.

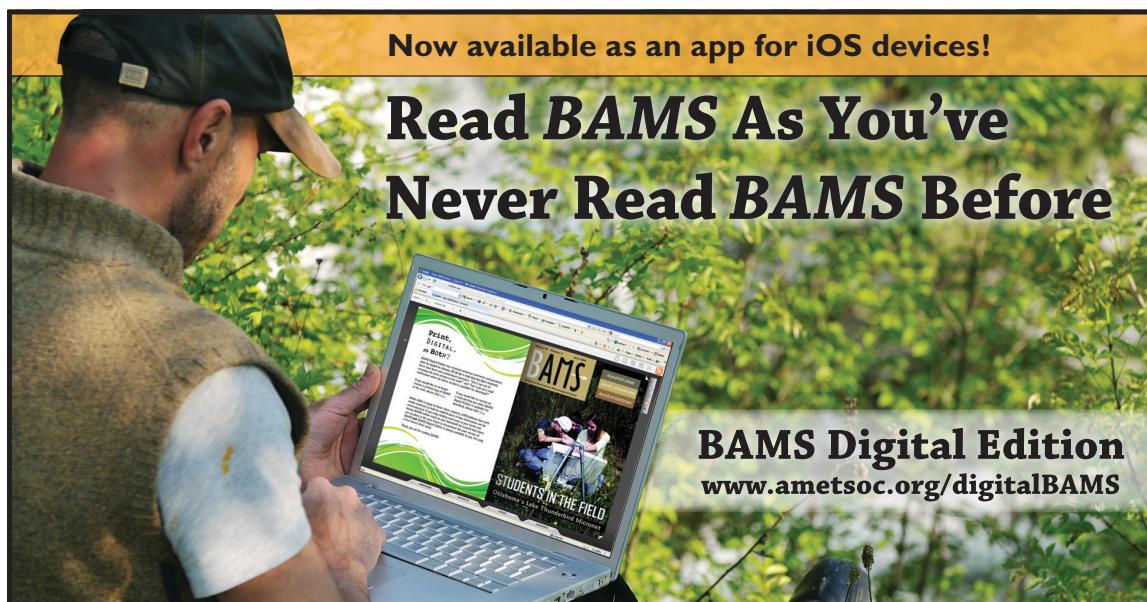
—, —, Y. Hong, and E. Wells, 2012: Evaluation of tools used for monitoring and forecasting flash floods in the United States. *Wea. Forecasting*, **27**, 158–173, doi:10.1175/WAF-D-10-05043.1.

Gruntfest, E., 2009: Editorial. *J. Flood Risk Manage.*, **2**, 83–84.

Kim, Y., M. Muste, A. Hauet, W. F. Krajewski, A. Kruger, and A. Bradley, 2008: Stream discharge using mobile large-scale particle image velocimetry: A proof of concept. *Water Resour. Res.*, **44**, W09502, doi:10.1029/2006WR005441.

Ortega, K. L., T. M. Smith, K. L. Manross, K. A. Scharfenberg, A. Witt, A. G. Kolodziej, and J. J. Gourley, 2009: Severe hazards analysis and verification experiment. *Bull. Amer. Meteor. Soc.*, **90**, 1519–1530, doi:10.1175/2009BAMS2815.1.

Xiao, X., P. Dorovskoy, C. Biradar, and E. Bridge, 2011: A library of georeferenced photos from the field. *Eos, Trans. Amer. Geophys. Union*, **92** (49), 453–454.



Now available as an app for iOS devices!

# Read *BAMS* As You've Never Read *BAMS* Before

**BAMS Digital Edition**  
[www.ametsoc.org/digitalBAMS](http://www.ametsoc.org/digitalBAMS)

# SHOP

the AMS online bookstore



Use this **easy-to-navigate** site to review and purchase new and classic titles in the collection of AMS Books—including general interest weather books, histories, biographies, and monographs—plus much more.

**View** tables of contents, information about the authors, and independent reviews.

As always, **AMS members receive deep discounts** on all AMS Books.

[www.ametsoc.org/amsbookstore](http://www.ametsoc.org/amsbookstore)

The AMS online bookstore is now open.

**Booksellers and wholesale distributors** may set up accounts with our distributor, The University of Chicago Press, by contacting Karen Hyzy at [khyzy@press.uchicago.edu](mailto:khyzy@press.uchicago.edu), 773-702-7000, or toll-free at 800-621-2736.

**AMS BOOKS**

RESEARCH APPLICATIONS HISTORY

Unauthenticated | Downloaded 01/14/21 05:55 AM UTC