13B.1

USING CROWD-SOURCED DATA TO IMPROVE ANALYSES OF FLASH FLOOD EVENTS

W. Scott Lincoln[1]
1. NOAA/NWS Lower Mississippi River Forecast Center, Slidell, LA

1. INTRODUCTION

1.1 Why care about flash flooding?

National Weather Service (NWS) hazard statistics continue to indicate flooding as one of the top causes of weather-related fatalities in the United States (OCWSS, 2014). When subdividing floods into river floods and flash floods, OCWSS data indicates that flash flooding caused the second highest number of short fuse weather fatalities and property damage, second only to tornadoes, over the 1994-2013 period when detailed statistics are available (Figure 1).

1.2 Why improve flash flooding warnings?

Despite the threat that flash flooding presents, warnings for flash floods are typically not as useful as warnings for tornadoes, hail, and straight-line winds. Tornado/severe thunderstorm warnings typically include scientific reasoning, a specific location and movement of the threat, an estimate of the threat severity, and expected evolution of the threat over the next few minutes. In contrast, flash flood warnings frequently cover very large areas, are in effect for several hours, often lack estimates of event severity, and often lack mention of specific areas or waterways that will be impacted. Some of this discrepancy in warning information can be tied to the prediction and monitoring tools utilized for issuing flash flood warnings.

The most commonly-used method used by forecasters to predict flash flooding is Gridded Flash Flood Guidance (GFFG) in the Flash Flood Monitoring and Prediction (FFMP) software. GFFG is the estimated amount of rainfall over a set duration that, once exceeded, should correspond to the onset of flash flooding. FFMP indicates the locations where this GFFG is exceeded and thus flash-flood-producing runoff is expected to occur (Figure 2), although this is not necessarily the same location where the *flooding* will occur because FFMP lacks any simulation of surface flow routing. GFFG/FFMP also do not provide any estimates of flash flood severity.

1.3 How can flash warnings be improved?

To improve flash flood warnings, new models and techniques will need to not only indicate that flash flooding is likely but also 1) provide estimates of the event magnitude, and 2) project the evolution of the threat. Different methods are currently being developed which may improve both of these categories, although challenges with verification remain.

Current flash flood reports come from the NWS Local Storm Reports (LSRs) which consist of reports made by storm spotters, law enforcement, and sometimes members of the public that are sent to the local NWS Weather Forecast Office (WFO). Unlike straight-line wind and tornado events where damage is typically surveyed afterward by NWS staff, flash flood events are rarely surveyed and quality control of reports is limited. Another complication is that flash flooding sometimes leaves fewer clues behind after the event than high-end wind events (Figure 3).

This paper explores one method of improving flash flood reports. When possible, information documenting flash flooding was obtained from social media (Facebook, Twitter, Youtube), news media (TV, radio, newspaper websites), and other web sources (including Google Traffic and blog

* *Corresponding author address:* W. Scott Lincoln, NOAA/NWS/Lower Mississippi River Forecast Center, 62300 Airport Road, Slidell, LA 70460; *scott.lincoln@noaa.gov*

posts). When possible, pictures and videos were geo-referenced using aerial imagery and street-level imagery, and were also assigned a relative magnitude.

In the follow sections, the various sources of crowd-sourced flash flood reports will discussed in more detail, and then six (6) events will be presented as case studies in using this methodology.

2. SOURCES FOR ADDITIONAL REPORTS

For six (6) separate events, the addition of crowd-sourced reports of flash flooding has significantly improved the information available for post-event analysis at the NWS Lower Mississippi River Forecast Center (LMRFC). A description of these events is provided by Table 1. For each event, reports begin with the NWS LSRs only, and then crowd-sourced reports are added from social media, news media, and other web sources. Reports were subsequently categorized by impact. The Pensacola flash flood event from April 2014 will be used to provide a step-by-step illustration of this process (Figure 4).

2.1 Social Media

As soon as possible after the flash flood events, social media was data-mined for any additional information on possible impacts. Social media platforms utilized to search for additional flash flood reports included Facebook, Twitter, and Youtube.

Of the three platforms, Facebook was the easiest to use if several days had passed since the flash flood event being studied. Facebook's Timeline functionality allowed for improved navigation to the time period in question (as opposed to continuously scrolling to move backward in time through posts). First, the Facebook pages of popular media outlets were checked for additional information, which may include posts by the media outlet itself or by others in the "posts to page" section. Photos were found to be most helpful because with a photo it was easier to estimate flash flood severity and clues from the image could be used to help identify the precise location. When the precise location could be easily determined, a comment was added to the photo asking for additional information to either be posted as a follow-up comment or as an email (in the event that the photo author did not want to share personal location information). Sometimes the location of photos could be determined by using the caption to narrow down the search area, then by using aerial imagery to find a match (Figure 5); this is described further in section "2.4 Geo-referencing Reports".

Twitter was also used in a few instances to provide additional information for post-event analysis. In one notable example, a Twitter user began posting the storm total rainfall as reported by his personal weather station a few hours after the Pensacola flash flood event began, and then continued this effort until the rainfall had ended. The combination of text reports and images posted to his Twitter account over a nine (9) hour period was sufficient to create an event hyetograph similar to what could be produced from official rain gauge data (Figure 6).

2.2 News Media

During and after flash flood events, news articles posted to the websites of local television stations and local newspapers were monitored for descriptions of flood impacts. Useful information included lists of flooded roadway sections as well as photo galleries attached to the articles. As with social media pictures that cannot be pinned to a precise location, aerial imagery and street-level imagery was used to determine the location, when possible.

2.3 Other Web Sources

There are also a few other locations that may sometimes provide information on the location of flash flooding, or at least suggest that flooding may have occurred. A few blog posts discussing flooding impacts were found when doing an advanced web search (specifying a date range) within a few days of the event. Most of the time, posts to blogs were not conclusive enough to produce a report of their own, but could corroborate other information as well as provide additional information to make the location more precise.

Another indirect method of identifying areas impacted by flash flooding was to use traffic speed data from Google Maps. For the April 2014 Pensacola event, traffic maps indicated flooded bridges by showing a lack of traffic data right at a waterway crossing coincident with stopped traffic on both ends of the crossing (Figure 7). Although this information was not conclusive by itself that flooding caused the traffic data anomaly, it was consistent with what would occur if a roadway is impassible, and was worthy of a report based upon anecdotes from other sources.

2.4 Geo-referencing Reports

As mentioned in previous sections, often times the reports of flooding were ambiguous with regards to the location. Aerial imagery and street-level imagery (typically from Google) were used to try and determine the location of flooding reports. Locations were corroborated with additional information when possible.

Once geo-referencing was complete for the Pensacola flash flood example, the number of reports jumped from 20 to 147 (Figure 4, middle). The vast majority of reports (108) came from social media (Figure 8, top). Of the social media based reports, almost all (101) came from Facebook (Figure 8, bottom).

2.5 Categorizing Reports by Severity

Reports were categorized by relative impact using the methodology in “Analysis of the 15 June 2013 Isolated Extreme Rainfall Event in Springfield, Missouri” (Lincoln, 2014). Table 2 describes the severity categories and the necessary criteria for each. In some cases, the description of flooding was too vague to assign a category, and a label of “unknown” was used. In the Pensacola flash flood example, severity of flash flood impacts varied widely across the entire area (Figure 4, bottom).

3. CASE STUDIES

3.1 Ouachita Mountains, AR: June 2010

Very heavy rainfall impacted portions of western Arkansas on June 11th, 2010, causing significant flooding in portions of the Ouachita Mountains. The heaviest 3 hours of rainfall exceeded the 50-year (2% annual chance) event in a few areas.

Due to the relatively remote location of the heavy rainfall, only a few reports were sent to the local NWS office. Using crowd-sourcing, the number of flash flood reports was increased from 5 to 11. Most reports of flooding were within, or just downstream of, the areas experiencing the most significant rainfall (Figure 9).

3.2 Eastern TN: August 2012

Very heavy rainfall impacted portions of eastern Tennessee on August 5th, 2012, causing significant flooding near Johnson City. The heaviest 3 hours of rainfall exceeded the 1000-year (4% annual chance) event in a few areas.

Only a few reports were sent to the local NWS office during and soon after the event. Using crowd-sourcing, the number of flash flood reports was increased from 3 to 20. Almost all reports of flooding were within, or just downstream of, the areas experiencing the most significant rainfall (Figure 10).

3.3 Coastal MS: May 2013

Very heavy rainfall impacted portions of southeastern Mississippi on May 5th, 2012, causing widespread reports of flooding. The heaviest 3 hours of rainfall exceeded the 50-year (2% annual chance) event in a few areas.

Numerous reports were sent to the local NWS office during and soon after the event, although reports from the areas of heaviest rainfall were sparse. Using crowd-sourcing, the number of flash flood reports was increased from 12 to 23. Most reports of flooding were within, or just downstream of, the areas experiencing the most significant rainfall (Figure 11).

3.4 Springfield, MO: June 2013

Very heavy rainfall impacted portions of southwestern Missouri on June 15th, 2013, causing significant flooding in portions of Springfield. The heaviest 3 hours of rainfall exceeded the 500-year (0.2% annual chance) event in a few areas.

Numerous reports were sent to the local NWS office during and soon after the event. Using crowd-sourcing, the number of flash flood reports was increased further from 8 to 17. Most reports of flooding were within, or just downstream of, the areas experiencing the most significant rainfall (Figure 12).

3.5 Southwest MS: March 2014

Very heavy rainfall impacted portions of southwestern Mississippi on March 28th, 2014, causing significant flooding in the town of Crosby as well as nearby rural areas. The heaviest 3 hours of rainfall exceeded the 100-year (1% annual chance) event in a few areas.

Few reports were sent to the local NWS office during and soon after the event. Using crowd-sourcing, the number of flash flood reports was increased substantially from 2 to 26. All reports of flooding were within, or just downstream of, the areas experiencing the most significant rainfall (Figure 13).

4. CONCLUSIONS AND FUTURE WORK

New techniques for nowcasting flash flooding locations and relative severity are currently being developed. These techniques will need to be verified, which will require reports of flooding that are improved compared to current NWS LSRs. A crowd-sourcing approach, utilizing the data mining of social media, news media, and other web sources is one way to accomplish this goal. This paper illustrated an example of this crowd-sourcing approach using six (6) events. The number of flash flooding reports were increased significantly (Table 1), and information about the relative severity was added.

Future work on this effort should include methods to automate the data mining process so that the number of cases can be greatly increased. With the small number of cases presented here (6), definitive conclusions on the output from new nowcasting techniques will not be possible.

5. REFERENCES

Lincoln, W. S. (2014). Analysis of the 15 June 2013 Isolated Extreme Rainfall Event in Springfield, Missouri. *J. Operational Meteor.*, 233-245.

OCWSS. (2014). *National Hazard Statistics*. (NOAA NWS Office of Climate, Water, and Weather Services) Retrieved November 2014, from NWS Weather Fatality, Injury and Damage Statistics: http://www.nws.noaa.gov/om/hazstats.shtml

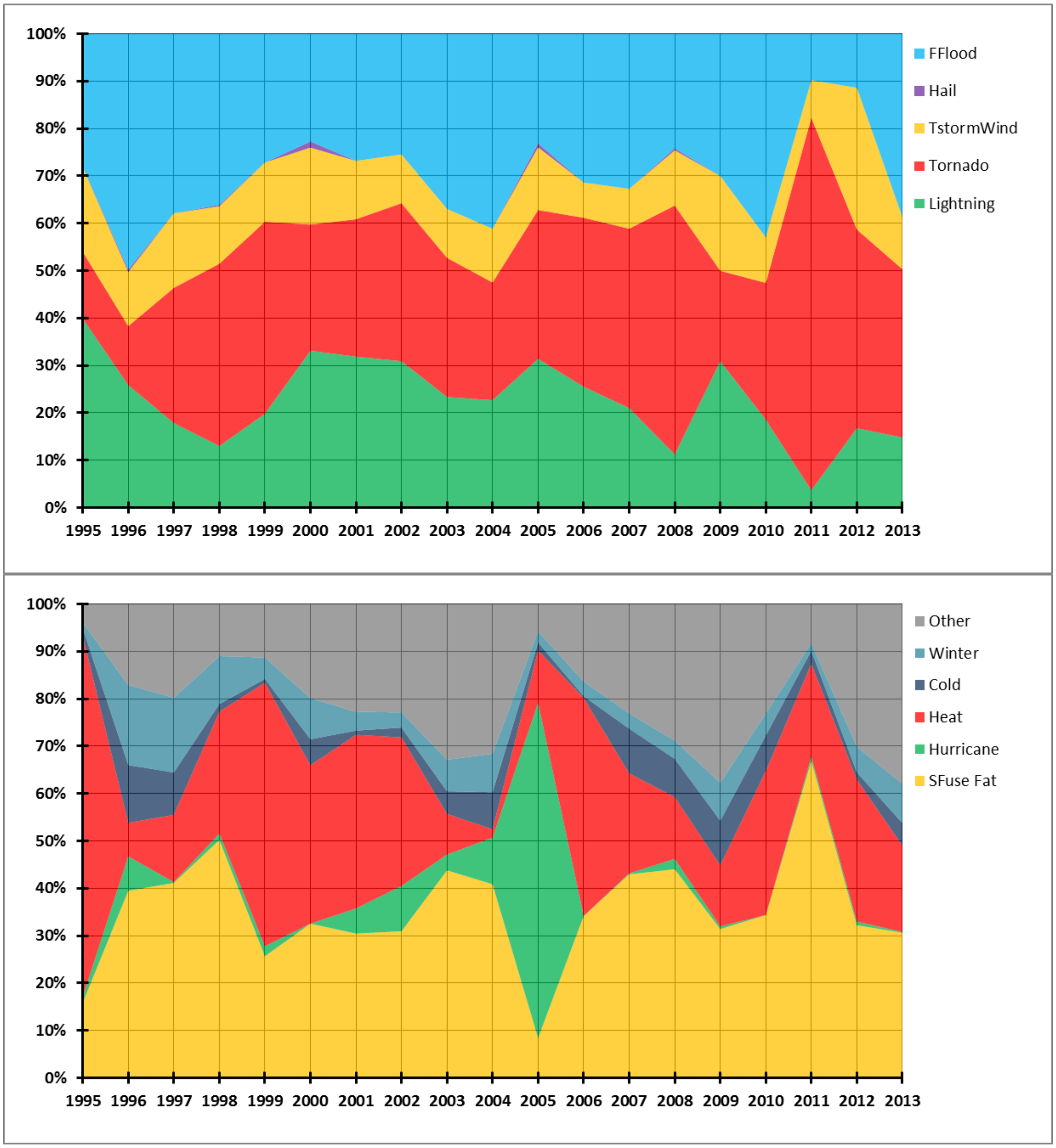


Figure 1. Fatalities in the U.S. due to all weather events (top) and only "short-fuse" weather events (bottom) as reported by the National Weather Service Office of Climate, Water, and Weather Services. Fatalities due to flash flooding have typically ranged from 20-40% of all short-fuse weather fatalities over the 1995-2013 period.

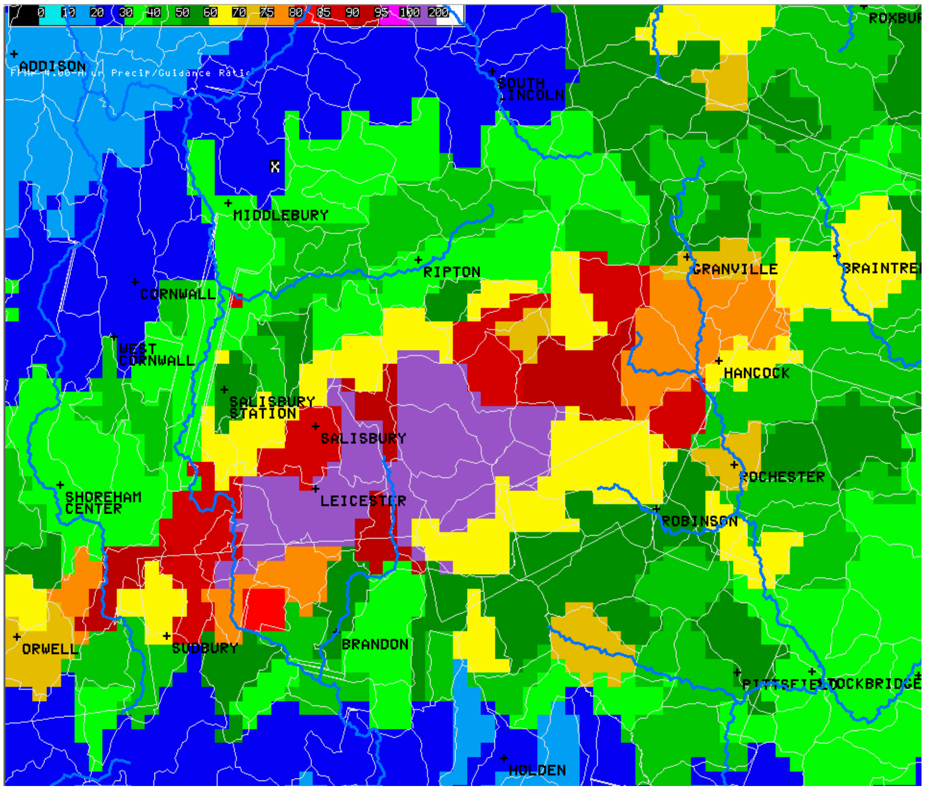


Figure 2. Ratio of rainfall estimates to the GFFG value. Ratios of 1.0 or greater correspond to where runoff should exceed the threshold needed to begin causing flash flooding.

Figure 3. Two photographs of areas that experienced severe flash flooding in the Lake Charlene neighborhood of Pensacola, FL, during the April 2014 event. Photos were taken within about a week of the water receding. In the top photo, water was high enough to necessitate rescue of persons through a home's attic. In the bottom photo, water was reportedly as high as the top of the fence (blue line added for clarity). In both photos, note the lack of flood debris, mud marks, or other clear high water marks. *Photo credit: Jenna Beall.*

Table 1. Flash flood events where crowd-sourced data was collected and added to available NWS LSRs. The maximum flooding severity is based upon the criteria presented in Table 2. Because rainfall climatology can vary greatly between case study locations, the maximum 3-hour rainfall average recurrence interval (ARI) is used to “normalize” the data.

Date	Location	Maximum Flooding Severity	Maximum 3-hour Rainfall ARI
June 2010	Ouachita Mountains, AR	100-yr floodplain inundated	50 year
August 2012	Johnson City, TN, area	Roadway washed out	1000 year
May 2013	Pascagoula, MS, area	Roadway washed out	50 year
June 2013	Springfield, MO, area	500-yr floodplain inundated	500 year
March 2014	Crosby, MS, area	Residences flooded	100 year
April 2014	Pensacola, FL, area	Dam failure	--

Figure 4. Reports of flash flooding from the NWS LSRs only (top), the NWS LSRs plus crowd-sourced reports (middle), and reports from both sources categorized by relative severity, all for the April 2014 Pensacola flash flood event.

Figure 5. Photograph posted to Facebook showing a small dam break that occurred near Crestview, FL, as a result of the April 2014 flash flood event in the Pensacola area (top). The approximate location of the photograph was determined using the brief description of the photograph and aerial imagery. Black circles show locations that were matched between the photgraph and aerial imagery, and the red square indicates the location of the dam break. *Photo credit: Neda Burtman, Aerial Imagery Credit: Google.*

Figure 6. Tweets of private weather station data posted to Twitter (top) by user Rob McGahen was of sufficient frequency to create an event hyetograph (bottom) for the the night of April 30th, 2014 during the Pensacola flash flood event.

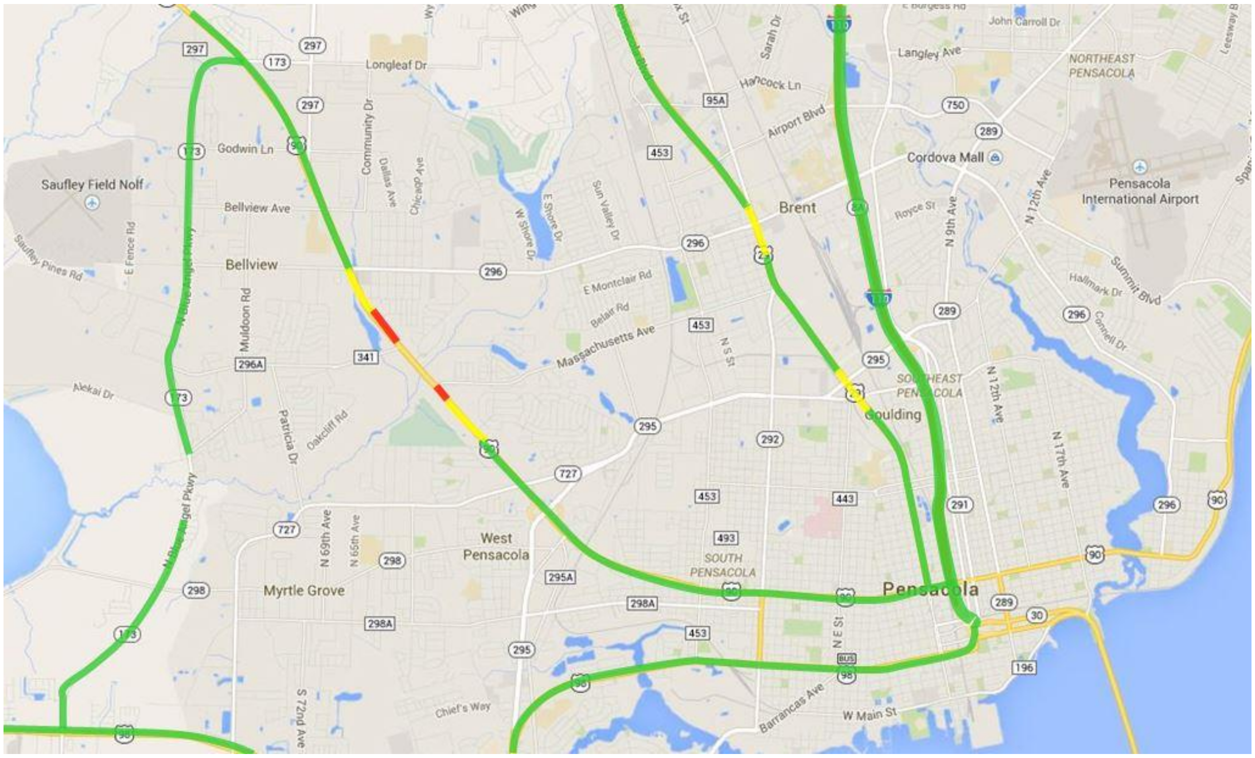
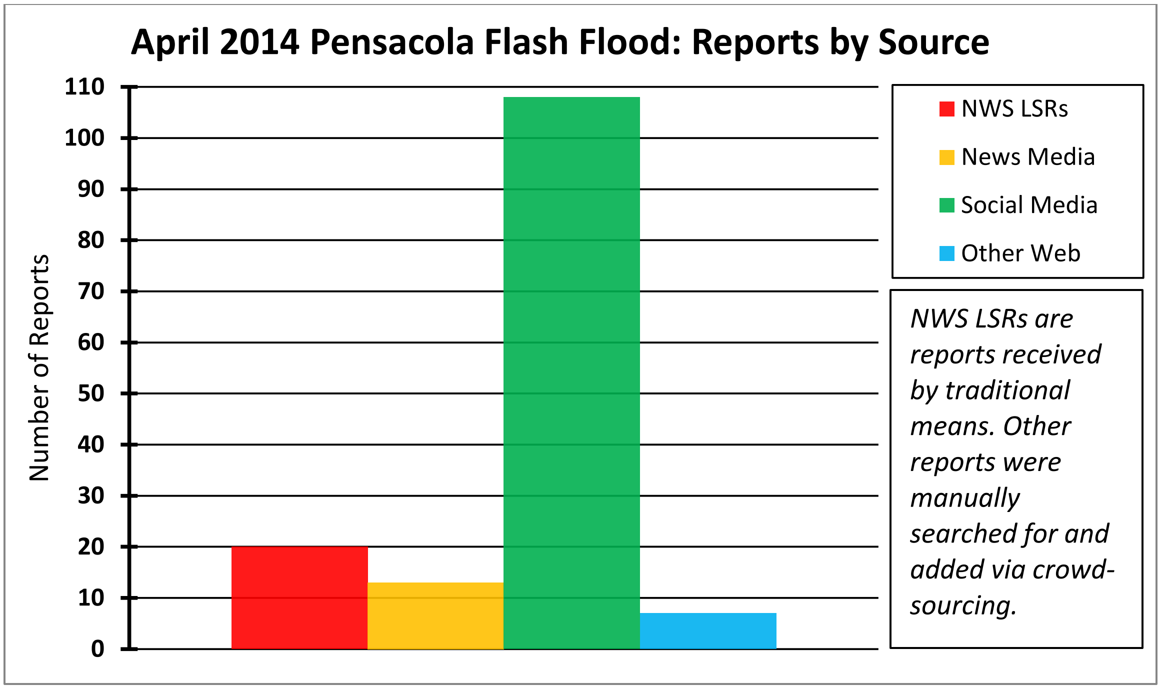


Figure 7. Recreation of the traffic display from Google Maps for the morning of April, 30th, 2014, during the Pensacola flash flood event. Areas of green indicate free-flowing traffic, yellow indicates slow-moving traffic, and red indicates stop-and-go traffic.



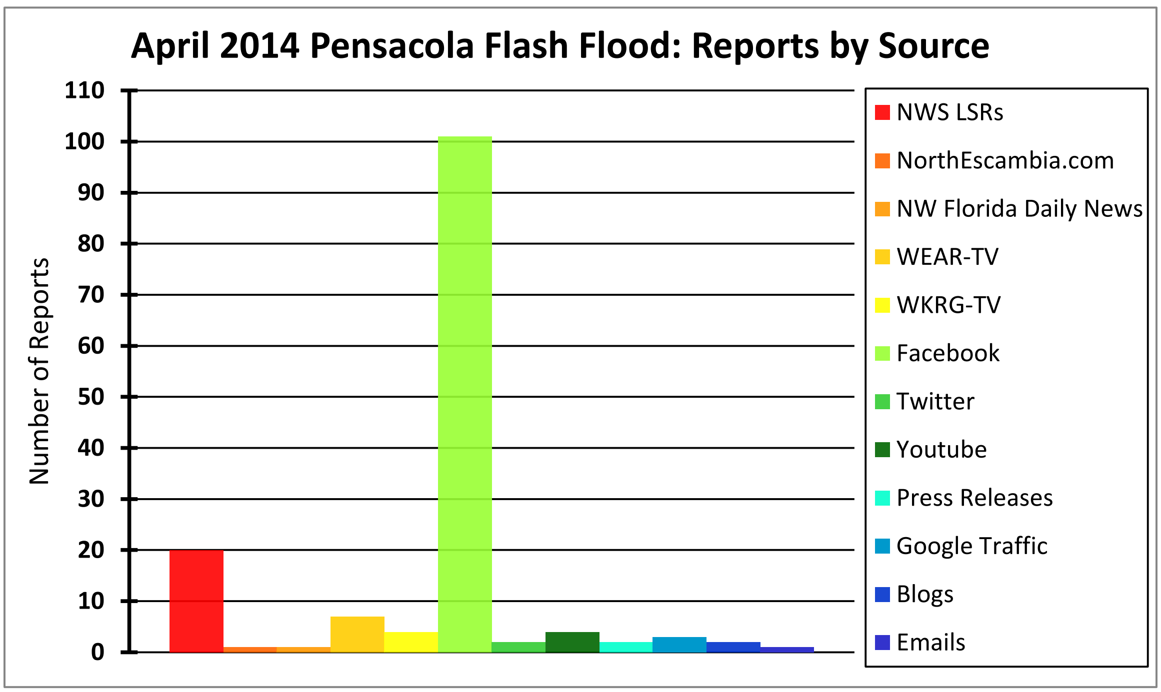


Figure 8. Reports of flash flooding from the April 2014 Pensacola event organized by source. First, the reports were organized by broad category (top). Next, the reports were organized by the specific source (bottom).

Table 2. Severity categories used to qualify the relative magnitude of flash flood impacts. Adapted from Lincoln (2014).

Severity Label	Description/Criteria
Roadway flooded	Minor nuisance flooding of roadways
Roadway flooded (major)	Flooding of roadways deep enough to stall cars, or overtopping of bridges along major highways of modern design standards
Water rescue	Reports of persons needing to be rescued from residences or their vehicles
100-yr floodplain	Flood inundation reaches the extent of the 100-yr (1% annual chance) floodplain, or river gauge reaches the 100-yr (1% annual chance) event.
Structure flooded	Residences or businesses flooded
Washout	Roadways or culverts completely washed away
Dam failure	Dam eroded away to allow impounded water to release uncontrolled
500-yr floodplain	Flood inundation reaches the extent of the 500-yr (0.2% annual chance) floodplain, or river gauge reaches the 500-yr (0.2% annual chance) event.
Unknown	Flooding reported but little additional information provided

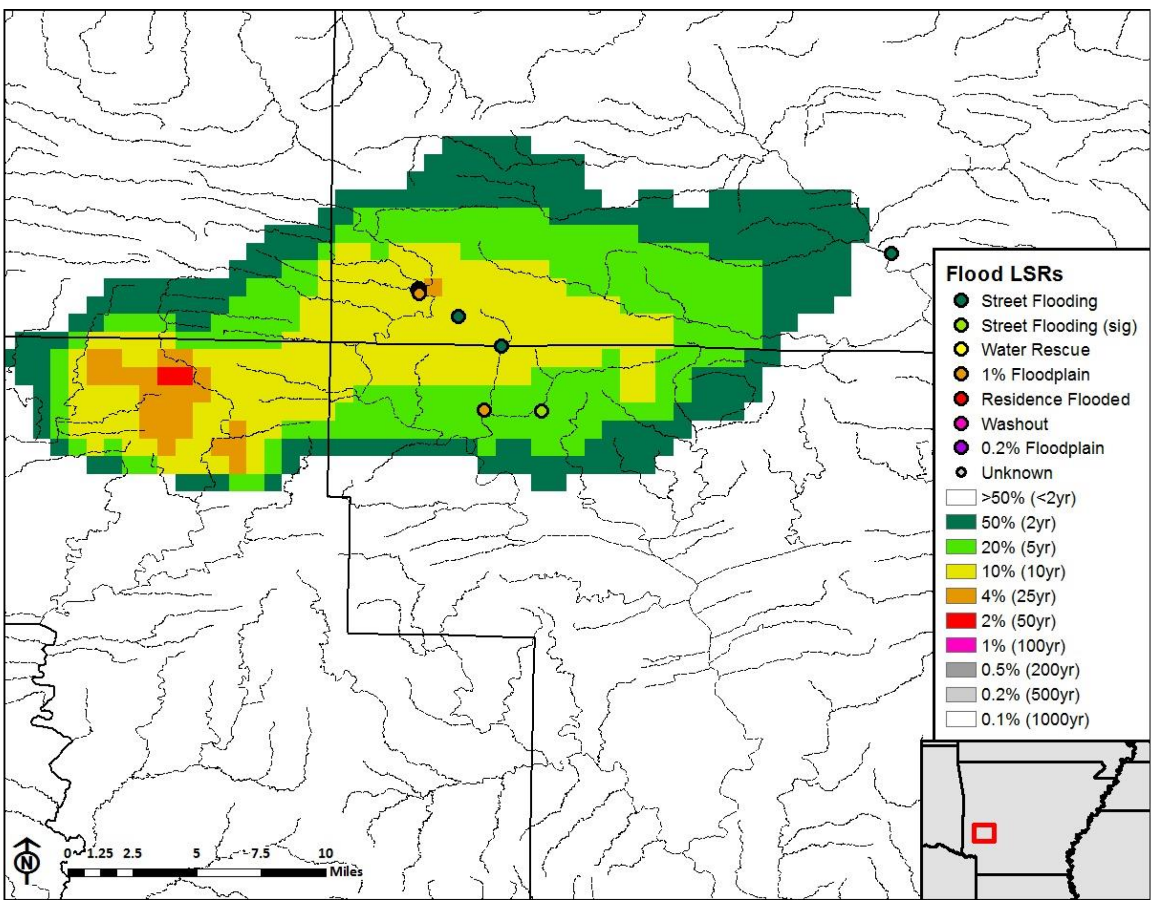


Figure 9. Reports of flash flooding from NWS LSRs and crowd-sourcing for the June 2010 Ouachita Mountains, AR, flash flood event. Gridded underlay is the ARI for the maximum 3-hour rainfall.

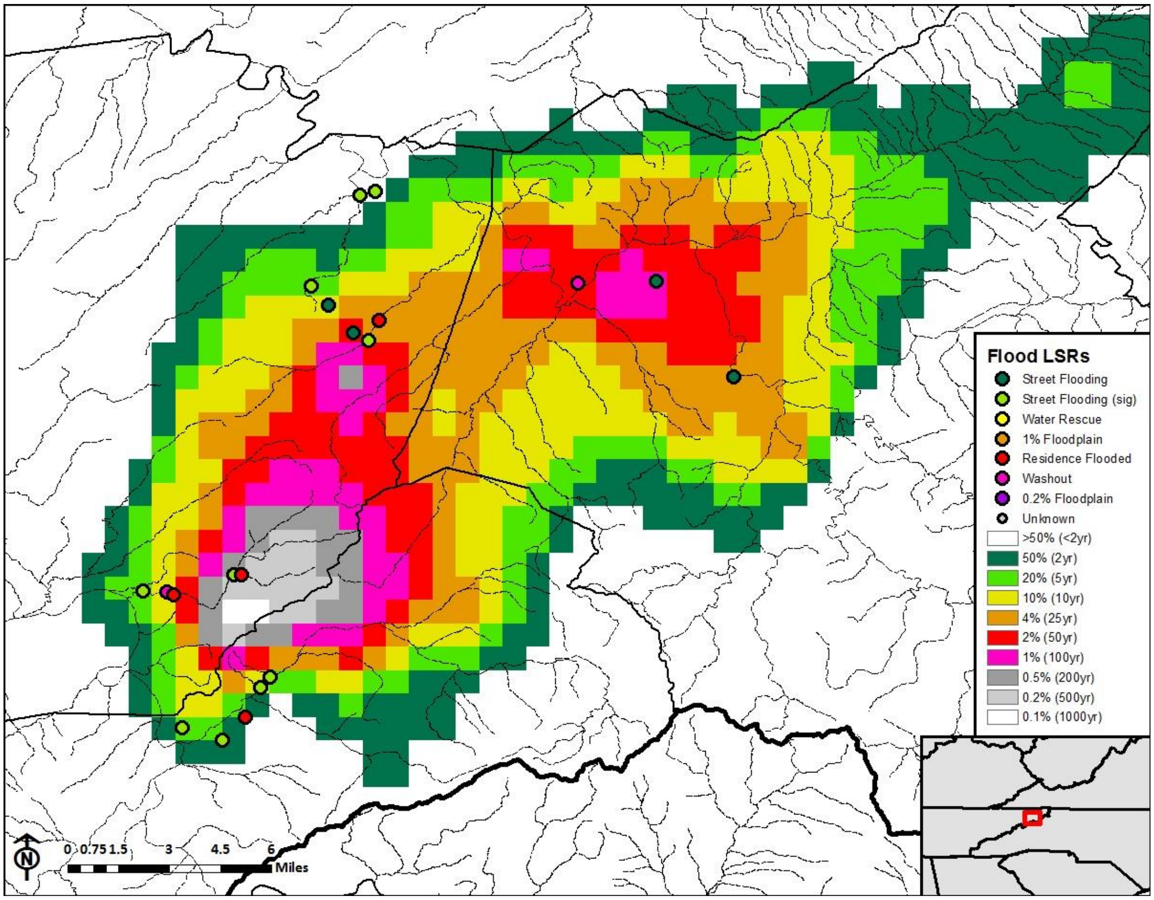


Figure 10. Reports of flash flooding from NWS LSRs and crowd-sourcing for the August 2012 Eastern TN flash flood event. Gridded underlay is the ARI for the maximum 3-hour rainfall.

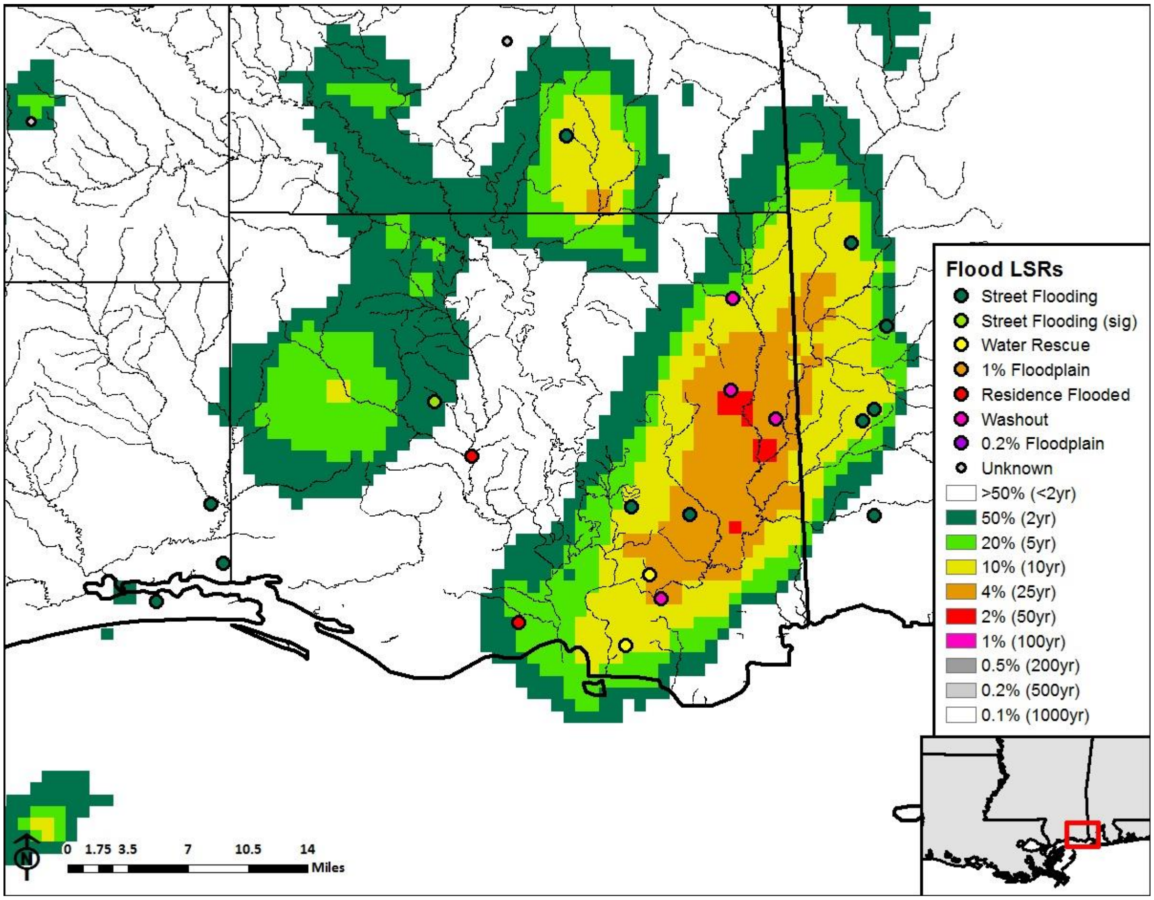


Figure 11. Reports of flash flooding from NWS LSRs and crowd-sourcing for the May 2013 coastal MS flash flood event. Gridded underlay is the ARI for the maximum 3-hour rainfall.

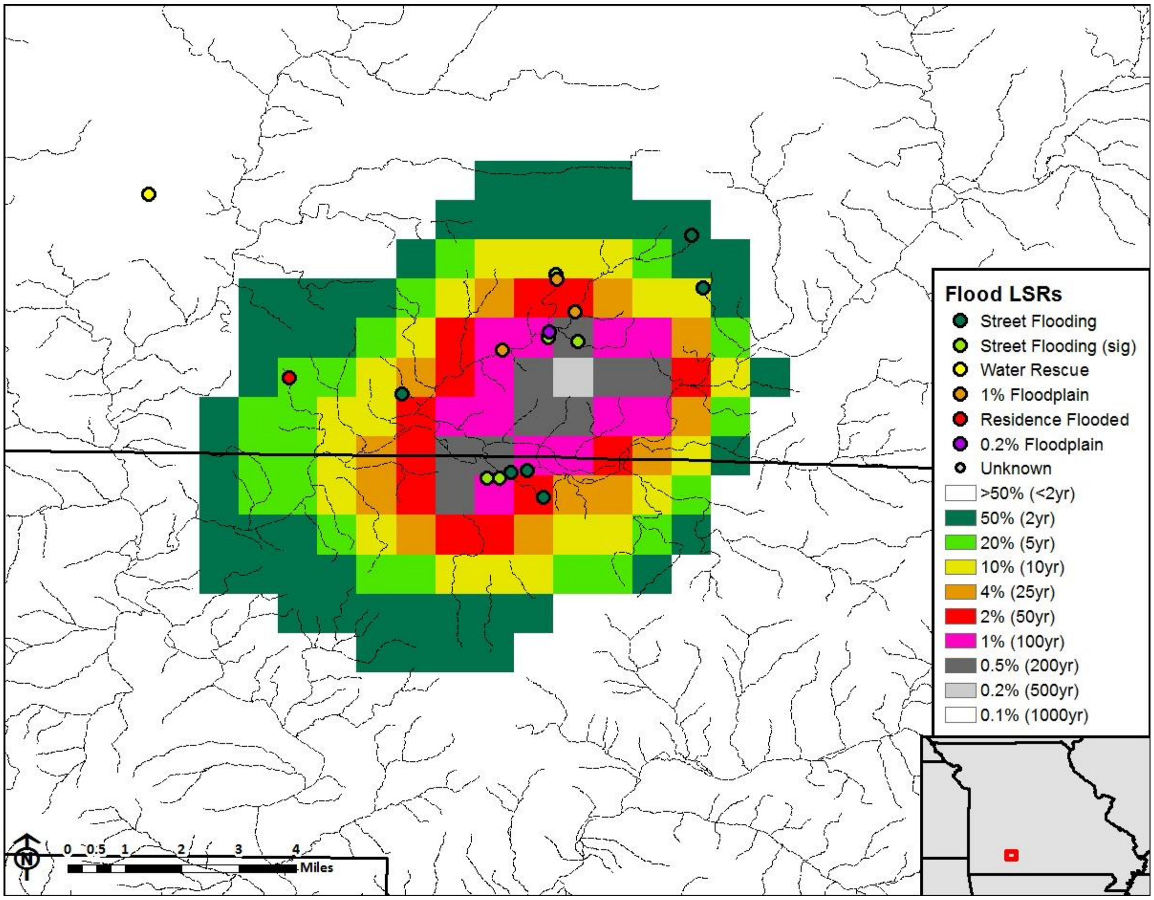


Figure 12. Reports of flash flooding from NWS LSRs and crowd-sourcing for the June 2013 Springfield, MO, flash flood event. Gridded underlay is the ARI for the maximum 3-hour rainfall.

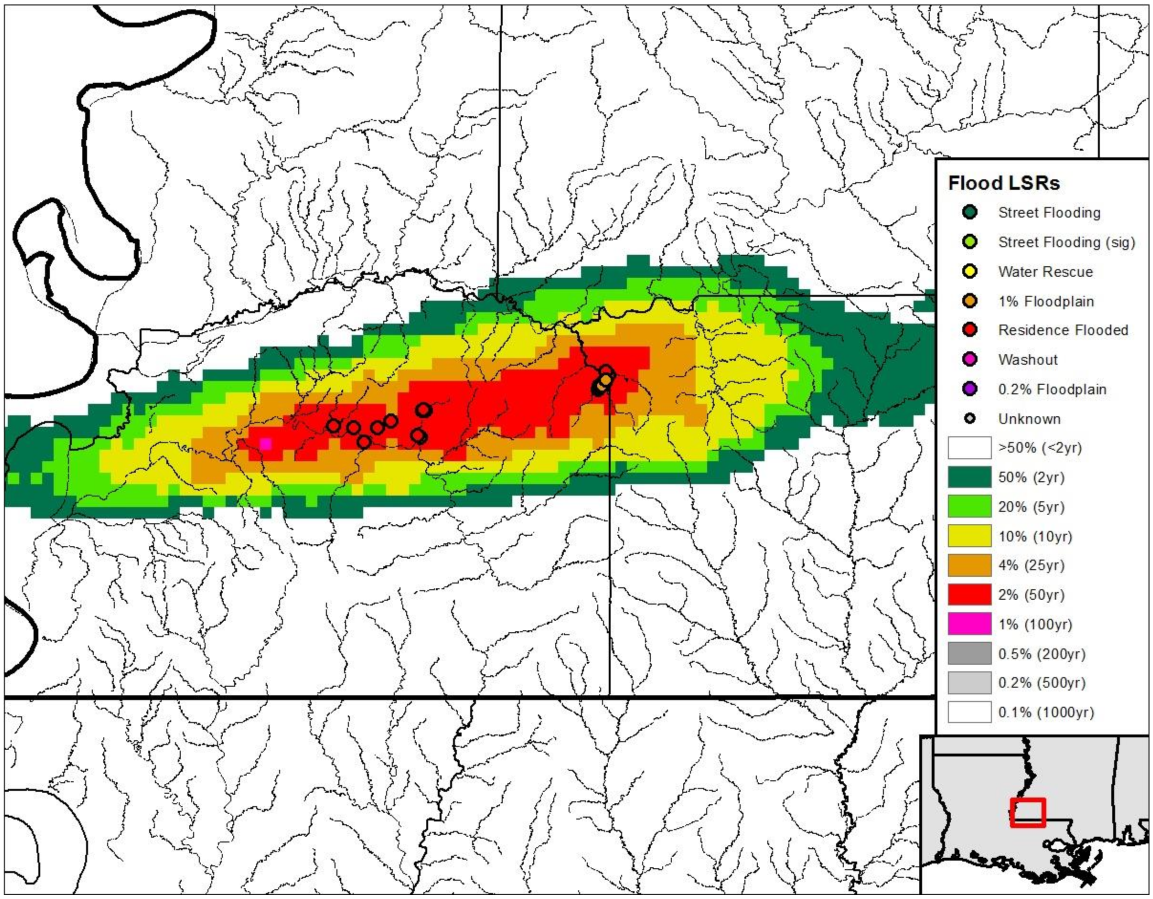


Figure 13. Reports of flash flooding from NWS LSRs and crowd-sourcing for the March 2014 Southwest MS flash flood event. Gridded underlay is the ARI for the maximum 3-hour rainfall.

Table 3. Summary of the improvement in quantity of flash flood reports for each event.

Date	**Location**	**Number of Reports**		**Change**	
		LSRs Only	**LSRs & Crowd-Sourced**	*(Combined – LSR)*	*(Difference/LSR)*
June 2010	Ouachita Mountains, AR	5	11	6	+120%
August 2012	Johnson City, TN, area	3	20	17	+567%
May 2013	Pascagoula, MS, area	12	23	11	+92%
June 2013	Springfield, MO, area	8	17	9	+113%
March 2014	Crosby, MS, area	2	26	24	+1200%
April 2014	Pensacola, FL, area	20	147	127	+635%