Probabilistic Framework for Integrating Multiple Data Sources to Estimate Disaster and Failure Events and Increase Situational Awareness

Cynthia Lee, S.M.ASCE[1]; and Iris Tien, Ph.D., A.M.ASCE[2]

Abstract: As data for monitoring the natural and built environments become increasingly prevalent, integrating information from varied data sources offers a fuller understanding of the impacts damaging events have on surrounding communities. In this paper, the authors present a probabilistic framework to integrate data from multiple sources to estimate disaster and failure events. The authors show how utilizing data from disparate sources, including physical sensors measuring environmental quantities and big data from social sensors reporting personal accounts and public perceptions within a community, contributes to increasing situational awareness during an event. The approach uses Bayesian updating to infer updated probabilities of event occurrence based on collected data and focuses on data fusion within first individual sensor networks and next across unrelated sensor types. The framework is flexible and applicable to estimate events in a variety of systems and environments using multiple, heterogeneous data sources. The authors apply the approach to estimate flood risks in Louisiana during a 4-d period in August 2016 by integrating physical sensor data from the United States Geological Survey and social media data from Twitter. The results show the change in estimated flood risks across the state as additional data is introduced and how multiple data sources increase the amount of updating possible in real-time event estimation. **DOI: 10.1061/AJRUA6.0000995.**

Introduction

For a community, situational awareness during disasters is the ability to collect and synthesize available information to fully understand vulnerabilities during times of crisis (Ireson 2009; Vieweg et al. 2008). This real-time and dynamic awareness of a community's surroundings is essential to support decision-making for response and mitigation of disaster and failure events and is critical for ensuring public safety and minimizing economic losses. Elements of these surroundings—both of the natural and built environments—can be difficult to estimate with uncertainties in hazards, responses, and impacts, and with continuous evolution in time. By better understanding the current state of the environment, a community can increase its absorptive capacities to improve its resilience to disaster events (Johansen et al. 2017).

Fast, accurate, and comprehensive monitoring of the natural and built environments is one way to do this. Physical sensors for monitoring these environments have increased capabilities as their technologies advance. Unfortunately, physical sensors lack widespread deployment in all communities, and data across available sensor networks is not always integrated to create a complete representation of potentially damaging conditions or events. At the same time, the rise of the web and social media has unlocked opportunities for gathering large amounts of information, observations, and opinions from the crowd in real time. Recent research has investigated the use of these social sensors to detect disaster events, e.g., Tien et al. (2016), Musaev et al. (2015), Vieweg et al. (2010), and Yin et al. (2012). Big data from social media is inherently noisy and unreliable, so discerning useful information from it can be difficult.

Variation in sensor types and the unique challenges associated with data analysis for each type must be considered to form a coherent view of a disaster or failure event from multiple sensor sources. In this paper, the authors present a probabilistic framework for integrating data from multiple and varied sensor types to estimate disaster and failure events. Such integration has the potential to improve situational awareness by providing more comprehensive and up-to-date evaluations of a community's surroundings than an individual sensor type alone. The approach uses Bayesian updating to infer updated probabilities of event occurrence from assumed or computed prior probabilities. The result is a posterior event probability given collected data. Bayesian updating is a well-known method for data fusion across multiple sensors, especially in the field of robotics (Durrant-Whyte and Henderson 2008). The novelty of the proposed approach is (1) in its considerations and updating of prior understandings of risk and (2) in integrating likelihoods of observing data points within individual sensor networks first before integrating that information across unrelated sensor types to monitor a community condition or event. The approach enables the framework to account for uncertainties in system states and to be applicable across events. It captures information specific to each data type, calculating specific sensor source likelihoods accounting for data heterogeneity, before combining information from all sources. It supports real-time and dynamic updating of estimated risks of an event as different data are observed and more data become available.

To illustrate the framework's use, the authors apply it to estimate flood risks in parishes in the state of Louisiana during a 4-d period in August 2016. Flooding during this time resulted in a Federal Emergency Management Agency (FEMA) disaster declaration in

[1]Ph.D. Student, School of Civil and Environmental Engineering, Georgia Institute of Technology, 631 Cherry St., Atlanta, GA 30332-0355 (corresponding author). ORCID: https://orcid.org/0000-0002-3581-6208. Email: cynthia.lee@gatech.edu

[2]Assistant Professor, School of Civil and Environmental Engineering, Georgia Institute of Technology, 631 Cherry St., Atlanta, GA 30332-0355. Email: itien@ce.gatech.edu

Note. This manuscript was submitted on October 27, 2017; approved on July 24, 2018; published online on October 11, 2018. Discussion period open until March 11, 2019; separate discussions must be submitted for individual papers. This paper is part of the ***ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering***, © ASCE, ISSN 2376-7642.

26 parishes on August 14, 2016, an estimated \$30 million in relief efforts from the American Red Cross, and over \$110 million of estimated losses for Louisiana's agriculture industry (Van Der Wiel et al. 2017). The data sources selected for integration for event estimation are stream gauges (a physical sensor type) from the United States Geological Survey (USGS) and tweets (microposts) from Twitter (a social sensor type). Prior probabilities of flood events in each parish are derived from FEMA flood risk maps. The results are validated by comparing the dates and locations of updated flood risks to the true flooding that occurred in Louisiana in August 2016.

The remainder of this paper is organized as follows. The first section describes the background and related work further while discussing the need for integrating data across sensor types to estimate events and the development of social media as a data source. This is followed by a description of the framework and the required calculations at each step of the approach. The authors next present an application of the framework for monitoring flood risks in Louisiana using data from two sensor types. In this section, the authors explain in detail the calculations to integrate the specific data sources for the application in order to explicate the framework's use. The next section presents the results from using the approach to integrate data from the two sources in a 4-d period in Louisiana in August 2016. Validation of the results and the use of the proposed probability updating for increasing situational awareness and decision-making are then discussed.

Background and Related Work

There are many different sensors available for monitoring conditions in the built and natural environments. For example, physical sensors such as strain gauges and accelerometers provide data to detect anomalies in civil structures, and air quality and meteorological sensors collect data on atmospheric conditions. While smart city initiatives are pushing forward the need for more connected wireless sensor networks, e.g., Chicago's Array of Things (Mone 2015), most interconnected sensor networks still lack heterogeneity to collect multiple types of data and require single or similar sensor types (Miorandi et al. 2012). Large-scale deployment of physical sensor networks must consider sensor size, cost, and configuration, and may be slow or impractical to implement depending on the application, limiting the availability of physical sensor data within a community (Rawat et al. 2014). Physical sensor networks for structural health monitoring are often designed to monitor single civil structures, as seen in studies such as Hackmann et al. (2014) and Pakzad et al. (2008), rather than groups of structures over larger distributed areas. Single physical sensor types alone are thus limited in the information they can reveal about community conditions affecting infrastructure and people.

Social media users offer a new type of sensor—a social sensor—which takes advantage of the constant information stream present today. Social sensors can fill in gaps of information about environmental conditions where physical sensors are not deployed or do not exist. They can also provide data that physical sensors cannot measure, such as information about societal conditions or public perception regarding major events. Research has been conducted using social sensors to examine community responses to events such as mass shootings (Vieweg et al. 2008) and to detect critical infrastructure failures (Tien et al. 2016) and natural disasters such as earthquakes (Sakaki et al. 2010). Social sensor data collection and analysis present their own set of challenges. Data from social sensors are noisy and unreliable, containing rumors or misinformation that are both intentional and unintentional (Alexander 2014). The majority of users do not share their locations (Leetaru et al. 2013), making it difficult to determine event locations and whether or not the user is truly observing the event of interest. Moreover, for social sensors to provide significant new information, specific topics need to garner enough attention from the crowd. While many physical sensors can provide continuous or near-continuous data, social sensors may vary in the amount of information provided on a topic at any given time.

Due to these challenges and limitations, including for both physical and social sensor sources, data from individual sensor networks alone may not truly describe the causes or impacts of a damaging event on a community. In this work, the authors investigate the ability of data integrated from multiple sensor sources to enhance resulting analyses, event detection, and overall situational awareness. Recent studies show potential in integrating different sensor sources to detect hazard events. For example, Musaev et al. (2015) map social media posts and data from physical sensor sources, including rainfall and earthquake data, to calculate the probability of detecting a landslide given a grid-based location. While this study and others, e.g., Jongman et al. (2015), use data from across sensor types, they do not provide a framework through which heterogeneous sensor likelihoods of observed data can be integrated to calculate a posterior probability of an event given that data. In addition, the framework in this paper considers prior probabilities of event occurrence to account for uncertainty in the disaster or failure events themselves. This enables the assessment of changing evaluations of risk as more data from different sources are collected.

Proposed Framework

For the remainder of this paper, the authors use the term sensor source to describe each type of sensor measuring a unique parameter, whether physical, social, or other, and the term sensor network to describe a group of sensors from the same source. In the proposed framework, data from sensors within one network are integrated first, and then data from across different sources are integrated. Let θ represent the occurrence of a disaster or failure event at a specified location or affecting a particular system component, and let $s_1, s_2, \ldots, s_k$ each represent data from a different sensor source, for k total sources. The prior probability of θ is denoted $P(\theta)$, with the prior probability of nonoccurrence of the event equal to $1 - P(\theta)$. The framework is applicable for systems or events with multiple states, as long as the states are mutually exclusive and collectively exhaustive. The authors refer to θ throughout this paper as the occurrence of the primary event of interest to be detected, $P(\theta)$ as the assessment to be updated, and $\bar{\theta}$ as the nonoccurrence of θ. Prior probabilities of θ can be based on historical data, physics-based analyses, previously updated probability distributions, expert judgment, or a combination of these. These probabilities represent a current understanding of risk for the event of interest.

To update prior probabilities, appropriate data sources, $s_1, s_2, \ldots, s_k$, must be identified that indicate θ. As no data collection is perfect, these sources indicate θ with some uncertainty. In the proposed approach, the data sources can be physical sensors, social sensors, a combination of the two, or others, and sensors that collect data continuously or provide information only at specific points in time. The data integration across sources, $s_1, s_2, \ldots, s_k$, is ultimately performed through Bayesian updating. Bayesian updating for data fusion combines data from sensors or experiments with prior probabilities of event occurrence to compute posterior occurrence probabilities (Khaleghi et al. 2013). The resulting posterior probabilities from the analysis are the conditional probabilities of events given observed data from the combined sources. Bayesian updating of θ using these data is shown in Eq. (1):

$$P(\theta|s_1, s_2, \dots, s_k) = \frac{P(s_1, s_2, \dots, s_k|\theta)P(\theta)}{P(s_1, s_2, \dots, s_k)} \tag{1}$$

where $P(s_1, s_2, \dots, s_k|\theta)$ is the joint conditional probability of observing data from all sensor sources for updating given event occurrence, representing the likelihood of all sources, and $P(s_1, s_2, \dots, s_k)$ is the joint probability of observing data from all sources. The resulting posterior probability, $P(\theta|s_1, s_2, \dots, s_k)$, is the probability of event occurrence given observed data from different data sources.

Calculating the likelihood of all sources and the joint probability of all data requires several intermediary calculations. First, the integrated likelihood of data given θ from each individual network, $P(s_i|\theta)$, must be calculated, where s_i represents the ith data source out of k, by combining the observations indicating θ within each network. These calculations are dependent on the specific application of interest, the nature of the data, any possible relationships between the data (including dependence or independence of observations within a network), and the accuracy or reliability of each source. The observed data from each source must correspond in terms of date, time, and/or location to the event of interest. Out of the sources considered, the number of sources k available for updating may vary in time. For instance, if on a day, only one of the considered sources provides data indicating θ, then $k = 1$, and k is subject to change the following day. The application example in the following section demonstrates how these calculations can be performed for two specific data sources. The calculations of $P(s_i|\theta)$ are intentionally left open and flexible so that the framework can be applied to different sensor sources outputting varying types of data with varying likelihood calculations for θ.

The next step of the proposed framework is to compute the joint likelihood of observing data from all sources, $P(s_1, s_2, \dots, s_k|\theta)$, from the integrated likelihoods of data within individual networks. The joint likelihood calculations are based on the assumption that individual source likelihoods are independent of each other when conditioned on the same event, location, or system component of interest, i.e., likelihoods are conditionally independent given θ. From this, the joint likelihoods are computed as the products of the individual source likelihoods for each event state, as in Eq. (2)

$$P(s_1, s_2, \dots, s_k|\theta) = \prod_{i=1}^{k} P(s_i|\theta) \tag{2}$$

The proposed framework next requires calculation of the joint probability of observing data from all sources, $P(s_1, s_2, \dots, s_k)$, using the total probability of data. This is possible because all states of the event are mutually exclusive and collectively exhaustive. Eq. (3) shows this calculation for two states of the event, θ and $\bar{\theta}$, i.e., occurred and not occurred. $P(s_1, s_2, \dots, s_k|\theta)$ is previously calculated from Eq. (2). The likelihood of data for the nonoccurrence of θ, $P(s_1, s_2, \dots, s_k|\bar{\theta})$, is computed in the same way, i.e., based on the likelihood of data from each source and assuming conditional independence. The likelihood of data from each source indicating $\bar{\theta}$, $P(s_i|\bar{\theta})$, depends on the application and source, just as $P(s_i|\theta)$ depends on them

$$P(s_1, s_2, \dots, s_k) = P(s_1, s_2, \dots, s_k|\theta)P(\theta) + P(s_1, s_2, \dots, s_k|\bar{\theta})P(\bar{\theta}) \tag{3}$$

Finally, the posterior probability of occurrence of the event of interest, θ, is updated given the data across sources by inputting the results of the previous calculations into Eq. (1). This posterior represents an updated understanding of risk for the disaster or failure event in real time as data become available.

Additionally, disaster or failure events in a community are often dynamically evolving in time. To consider data over multiple sequential time periods to estimate events, sequential Bayesian updating is employed in the steps above. Eq. (1) becomes Eq. (4) to compute a posterior probability using sequential data over two time periods. In the equation, s represents the first set of data from sensor sources $1, \dots, k$, and t, the second set from sources $1, \dots, m$. More time periods can be added as data are available, and it is not necessary to collect data from all of the same sources at every time period, as the availability of data from each source may change over time. Data output from across sensor networks is assumed to be conditionally independent from one time period to the next, with the joint likelihood of data from multiple time periods found by multiplying all available likelihoods. The purpose of sequential updating is to include data from multiple observed periods for updating events that develop and occur over time. Sequential updating in the proposed framework ends when updated probabilities approach 1, suggesting that occurrence of the event θ has been detected with near certainty, or when data indicating θ are no longer available. After obtaining posterior probabilities from updating or sequential updating, the proposed framework can continue to be used to detect θ in the future, with prior probabilities that remain the same or are re-evaluated based on the results of updating and the nature of the event of interest, e.g., if events occur as Poisson processes or with cumulative effects. For example, if damage is detected in an infrastructure component with a certain probability and no action is taken to remediate that damage, the updated probability can be assumed as the new prior. In contrast, if action is taken to repair that component, the prior probability of damage for future updating may be lower than the original prior. The same or additional data sources can be used

$$P(\theta|s_1, s_2, \dots, s_k, t_1, t_2, \dots, t_m) = \frac{P(s_1, s_2, \dots, s_k|\theta)P(t_1, t_2, \dots, t_m|\theta)P(\theta)}{P(s_1, s_2, \dots, s_k, t_1, t_2, \dots, t_m)} \tag{4}$$

In summary, the steps of the proposed approach are as numbered and described below:

1. Compute prior probability or probabilities of occurrence for the event(s) being considered, $P(\theta)$.
2. Identify the sensor sources available that indicate θ.
3. Determine which observations from each source indicate θ, with some uncertainty.
4. Integrate likelihoods of data from sensors within each sensor network to calculate the overall likelihood of data from each source, $P(s_i|\theta)$, $i = 1, \dots k$, where k is the number of sources. This process varies depending on the data output from each source. Repeat this for all sensor sources and states of the disaster or failure event.
5. Calculate the joint likelihood of data from all sources, $P(s_1, s_2, \dots, s_k|\theta)$. Assuming the sources are conditionally independent on θ, the joint likelihood is the product of all source likelihoods. Repeat this for all states of the event.
6. Calculate the joint probability of observed data, $P(s_1, s_2, \dots, s_k)$.
7. Calculate the final posterior probability of event occurrence updated given data from across sources, $P(\theta|s_1, s_2, \dots, s_k)$.

Application Example: Flood Event Estimation in Louisiana, United States

In this section, the authors present a specific example to illustrate application of the proposed framework. The application is to update flood risks in the state of Louisiana during August 10–13, 2016, by integrating data from both physical and social sensor sources using the proposed framework. Flood risk around the world is increasing due to climate change and other environmental factors (Hirabayashi et al. 2013), and Louisiana is known to be subject to high flood risk given its low elevation and coastal proximity (Groves et al. 2016). Flood events are distributed over large geographic areas that cannot be completely and continuously monitored in real time. In addition, uncertainty exists in terms of event timing and location based on a combination of factors in the natural and built environments (Morss et al. 2005). Therefore, there is the opportunity to integrate data from multiple sensor sources to increase situational awareness for these events. This application example is also chosen due to the availability of postevent data for validation of the approach for the flood events impacting Louisiana in August 2016. In the subsections that follow, each step in obtaining updated estimated flood risks based on collected data is explained within the structure of the proposed framework.

Steps 1–2: Prior Probabilities and Data Source Identification

For the application, let θ be defined as flood occurrence in a single parish (county) in Louisiana and $\bar{\theta}$ indicate nonoccurrence of a flood in the same parish with $P(\bar{\theta}) = 1 - P(\theta)$. The authors use FEMA Flood Insurance Risk Maps to derive a prior probability of risk, $P(\theta)$, and its complement for all 64 parishes. These maps, which are part of the National Flood Insurance Program, designate zones of the United States that are likely to be inundated in a flood event (Burby 2001). The base-level flood considered is the 100-year flood, or flood with a 1% probability of occurrence in a given year. The regions that will be inundated during the base flood event are named Special Flood Hazard Areas (SFHA) and are highlighted on the flood risk maps.

Fig. 1 shows an example flood risk map for Ascension Parish in Louisiana. This screenshot is from a publicly available interactive map developed by the Louisiana State University Agricultural Center, which allows users to view flood risks in Louisiana by parish (LSU AgCenter and LADOTD 2017). Some 64% of Ascension Parish is covered by a SFHA (shaded in the figure) and will be inundated in a 100-year flood event. The remaining area of each parish outside of the zone is expected not to be inundated in such an event.

An example event of interest $\theta_{Ascension}$ is flood occurrence in Ascension Parish. The authors assume the prior probability of this event to be 1% (the probability of flood inundation in a given year in a SFHA) multiplied by the fraction of the parish area covered by the SFHA. For Ascension Parish, this results in a probability of 0.64% for a prior knowledge of flood risk, and this is defined as $P(\theta_{Ascension})$. This process is repeated for all 64 parishes in Louisiana resulting in a range of prior probabilities of 0.05%–0.88% across all parishes. As with all probability estimations, there are potential errors in defining these priors. However, these represent typical assessments of flood risks in the United States, so the authors consider them sufficient to set an initial estimation of risk for each parish.

For data integration, the authors select data from two publicly available sources ($k = 2$) that, with some uncertainty, indicate a flood within each parish. The first data source (s_1) is physical sensor data from USGS, which monitors the conditions of the nation's streams and rivers with near-real-time data from stream gauges. A sudden increase in gauge height is selected as the measurement of interest to provide updating information on the probability of flood occurrence for each parish. The second data source considered (s_2) is social media big data from Twitter. The data collected from Twitter include the texts of tweets and metadata such as date, time, and tweet location, if available. Tweet relevance to flood events is selected as the metric of interest, which is determined using a machine learning classifier. The authors also select these sources to demonstrate how the framework can be used to integrate data from unrelated sources. By integrating information within each individual network before integrating data from across sources, the framework is able to consider unique data likelihoods for θ from each source.

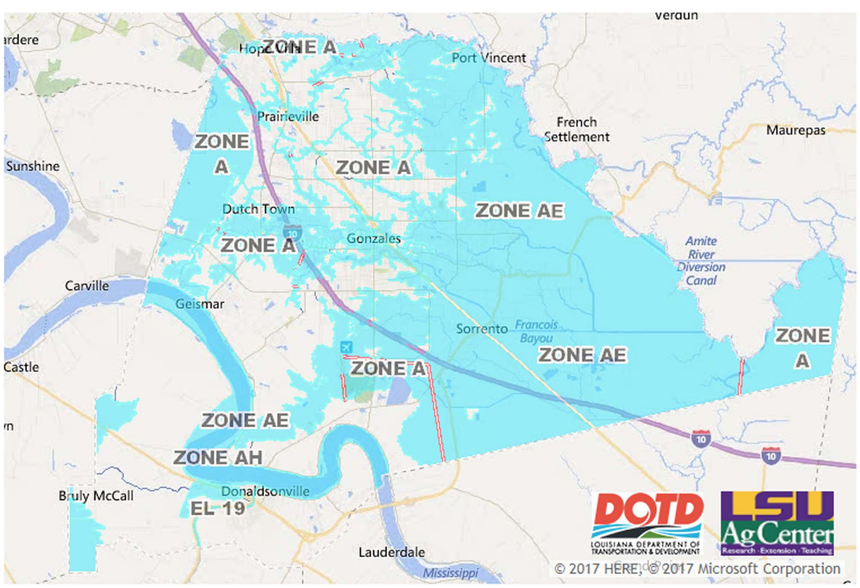

Fig. 1. Example flood risk map with 100-year floodplain shaded for Ascension Parish, Louisiana. (Map Data from © 2017 HERE, © 2017 Microsoft Corporation.)

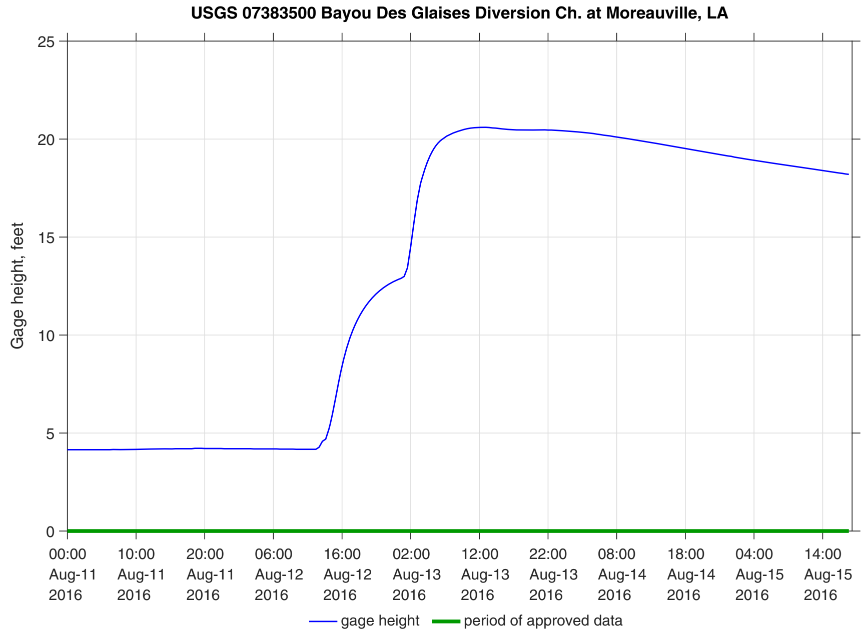

Fig. 2. Example gauge height data for a stream gauge in Louisiana. (Data from USGS 2017.)

Step 3: Data Collection from USGS and Twitter and Indications of θ

The authors next collect data from the two sources and determine which observations from each source indicate a flood event for each parish.

Gauge Height from USGS Stream Gauges

From the USGS stream gauges, data from all gauges outputting gauge height in Louisiana from June to August 2016 are collected. A total of 235 stream gauges in Louisiana provide downloadable gauge height data during the period of interest. Gauge height is reported at every 15-min or 30-min interval, depending on the gauge. Fig. 2 provides an example of available gauge height data from a stream gauge in Louisiana during a week in August.

Full flood predictions are based on a combination of topology, rainfall, streamflow, and gauge height data, as well as other hydrological and meteorological measurements. The objective of this study is to investigate the updating of prior estimations of event risk with data from multiple sources, rather than precise hydrologic modeling of flood systems. Therefore, a sudden increase in gauge height is taken as the indicator of flood occurrence.

To find sudden increases, daily average gauge heights are first computed over the period of data collection. From these averages, the percentage increase between each day is calculated. In the equation, y_j represents the percentage increase from the previous day to the current day for the jth stream gauge. A percentage increase in daily average gauge height y_j over 100%, i.e., the average gauge height at least doubled from one day to the next, indicates a potential flood event at that stream gauge. With this indicator of flood risk in the area surrounding a stream gauge, the data are binarized to indicate a flood event in an area on a particular day if $y_j > 100\%$. For this example, the authors do not consider data that do not indicate a flood event, where $y_j < 100\%$, but such data can be incorporated if the likelihoods based on such an indication can be calculated. Here, the authors focus on the differences in updating from unrelated data sources and choose only the $y_j > 100\%$ indication for the stream gauge source. The uncertainty of this indicator is accounted for when computing the overall likelihood of the sensor network, and the proposed framework allows for additional measurements and uncertainties from the same source indicating θ, such as multiple thresholds for a sudden increase in gauge height. Using other thresholds for indicating a flood event (e.g., $y_j > 150\%$) would add or remove data points from the analysis, with the overall process to calculate posterior probabilities remaining the same.

From the data, 66 stream gauges indicated a flood at least once during the three-month data collection period from June to August 2016, with several gauges indicating floods multiple times within the period. These are the observations that indicate occurrence of the event of interest θ and are used for updating in the following steps. Fig. 3 shows the locations of these stream gauges in Louisiana and the dates on which they indicated a flood for the period August 10–13, 2016.

Tweet Relevance to Flood Event

The Twitter data for this application are downloaded using Twitter's Streaming application programming interface (API) and by scraping historical tweets from Twitter's search page. To determine which observations from this data source indicate a flood event in each parish, a machine learning classifier is built to predict a tweet's relevance to a flood event, z. This prediction is binary: a tweet is either relevant ($z = 1$) or irrelevant ($z = 0$) to a flood event. The classifier is then used to determine which tweets from the period of interest are relevant to flood events. After the tweets are classified, they are sorted by location.

The classifier is a support vector machine (SVM) model built using Weka, a Java package for machine learning (Hall et al. 2009). The full database of tweets used to build the model consists of about 137,000 tweets, streamed or downloaded from various time periods from August 2016 to February 2017. To build the model, a small training set of tweets is first compiled from the database and manually labeled as relevant or irrelevant to flood events. This training set is assembled with a diverse set of tweets that refer to many different flood events and also unrelated phenomena or events to ensure the model generalizes and classifies relevance for tweets regarding any flood event. Relevance is defined as tweets

ASCE-ASME J. Risk Uncertainty Eng. Syst., Part A: Civ. Eng., 2018, 4(4): 04018042

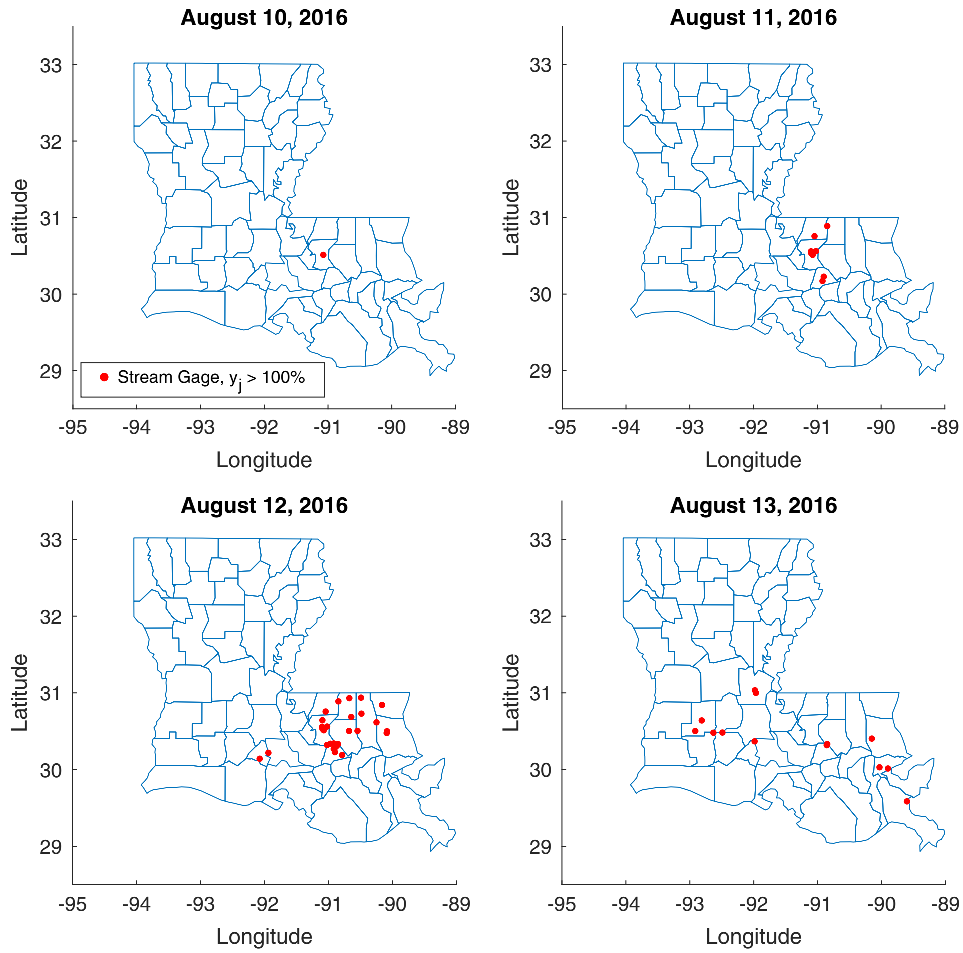

Fig. 3. Stream gauges in Louisiana indicating a flood, August 2016.

referring to current flood events, excluding updates about flood recovery efforts and expressions of sympathy from others.

The final training set consists of 496 tweets. In the training set, 125 tweets are relevant and 371 are irrelevant. A small testing set is also compiled to evaluate the model's performance at classifying new tweets. The testing set contains 214 tweets, manually classified with 55 relevant and 159 irrelevant tweets. Table 1 shows the confusion matrix for the model evaluated on the test set. The model's classification of a tweet is denoted by z, and the true value of a tweet's relevance is denoted by λ, where 1 indicates true relevance and 0 indicates true irrelevance.

The confusion matrix shows the performance of the model for evaluating the test set. The top row ($\lambda = 1$) shows 34 tweets were correctly classified as relevant, while 21 tweets were incorrectly classified as irrelevant. For $\lambda = 0$, the second row shows that 7 tweets were incorrectly classified as relevant although they were truly irrelevant. True positives are the tweets correctly classified as relevant by the model (the top left corner of the confusion matrix). Recall is calculated as the fraction of true positives out of the total number of truly relevant tweets, which is equal to 0.618. This represents a measure of accuracy expressing the likelihood of a correct tweet classification given true relevance to a flood, also represented by the conditional probability, $P(z = 1|\lambda = 1)$. The recall of the model in automatically classifying the testing set is used as the measure of the model's reliability. The authors use this value to calculate the likelihood of tweet relevance collected from Twitter in the next step of the updating process.

Table 1. Confusion matrix for support vector machine classifier tested for tweet relevance to flood events

Classified as $z = 1$	Classified as $z = 0$	True class
34	21	$\lambda = 1$
7	152	$\lambda = 0$

Using the built classification model, unseen tweets from August 10–13, 2016, filtered for the word *flood* are classified as relevant or irrelevant with the accuracy discussed above. As the event of interest θ is flooding in a parish, tweets are next filtered by location, if that metadata are available, to categorize them by parish in Louisiana. If location is not available, the text of each tweet is searched for cities and towns in Louisiana through a comprehensive list of municipalities and their respective parishes (Smith 2005). If the tweet contains one of the cities or towns on the list, it is considered a tweet relevant to a flood event for that parish. Of course, this means some tweets that are found may not truly be located in Louisiana (e.g., Iowa, Louisiana, is a town in Calcasieu Parish, but tweets found mentioning Iowa typically refer to the state of Iowa). Uncertainty in the data is accounted for by calculating the sensor source likelihood in the following section.

This process ultimately results in a list of tweets relevant to flood events for each parish. Fig. 4 shows an example of one of the tweets found relevant to a flood event in East Baton Rouge and Tangipahoa Parishes on August 12, 2016. There is no location

Date	Tweet Text
August 12, 2016	"Move to higher ground! Flash Flood Warning continues for Baton Rouge LA and Hammond LA until 1:00 PM CDT"

Fig. 4. Example of tweet classified as relevant and containing location indicators for Lousiana.

attached to the metadata of this tweet, so its associated locations are determined by searching the text for municipalities from the aforementioned list. The text of the tweet indicates Baton Rouge, Louisiana (East Baton Rouge Parish), and Hammond, Louisiana (Tangipahoa Parish). The tweet is therefore categorized to update the state estimations for these two parishes.

Step 4: Individual Source Likelihoods of Data from USGS and Twitter

Likelihood of $y_j > 100\%$

The daily likelihoods of $y_j > 100\%$ given a flood event in a parish θ, $P(y_j > 100\%|\theta)$, are now calculated, where y_j is the percent increase in daily average gauge height for the jth stream gauge and $j = 1, \ldots, n$ for n stream gauges indicating a flood event on the day in question. There are few validated empirical data that can be used to estimate the likelihoods of observing the data $y_j > 100\%$ given a flood event in a parish θ. Therefore, these likelihoods are calculated with a decaying function for $P(y_j > 100\%|\theta)$ that decreases in likelihood with each gauge's distance from the parish being considered. Fig. 5 shows the function used in this application, where distance is expressed in degrees of latitude and longitude. $P(y_j > 100\%|\theta)$ is assumed to be 1 when stream gauge j is in the parish considered by θ, i.e., a stream gauge will certainly show a daily average gauge height that doubles from one day to the next given a flood in a parish if it is located in that parish. As the distance between the parish and stream gauge increases, $P(y_j > 100\%|\theta)$ decreases. The distances are measured from each stream gauge to the nearest point on the border of the parish considered by θ.

As an example, on August 10, 2016, only one stream gauge in Louisiana reported $y_j > 100\%$ where $j = 1$ as shown in Fig. 3. That stream gauge's distance away from every parish was calculated and input into the function in Fig. 5. The results are the likelihoods of observing $y_1 > 100\%$ given flood events in each parish. For all stream gauges, the authors assume that any $P(y_j > 100\%|\theta)$ less than 0.01 is insignificant for updating and can be expressed as 0. Ultimately, there are $64 \cdot n$ likelihoods for each day, one from each stream gauge, out of n, for each of the 64 parishes.

Next, the likelihood of observed data for all stream gauges, $P(y > 100\%|\theta)$, is computed by combining the daily likelihoods of all stream gauges indicating a flood in a parish. The authors do not assume independence between these likelihoods and use the total probability theorem to integrate them. To do this, the authors introduce a variable, g_j, to represent the jth stream gauge out of all stream gauges indicating a flood per day. $P(g_j)$ is simply $1/n$, so that each stream gauge indicating a flood event is weighted equally. Let g_j be independent of θ and y_j, so $P(y_j > 100\%|g_j, \theta) = P(y_j > 100\%|\theta)$. The authors create independence for this variable to facilitate the integration of likelihoods of gauge height data. To eliminate this assumption of independence, more information on

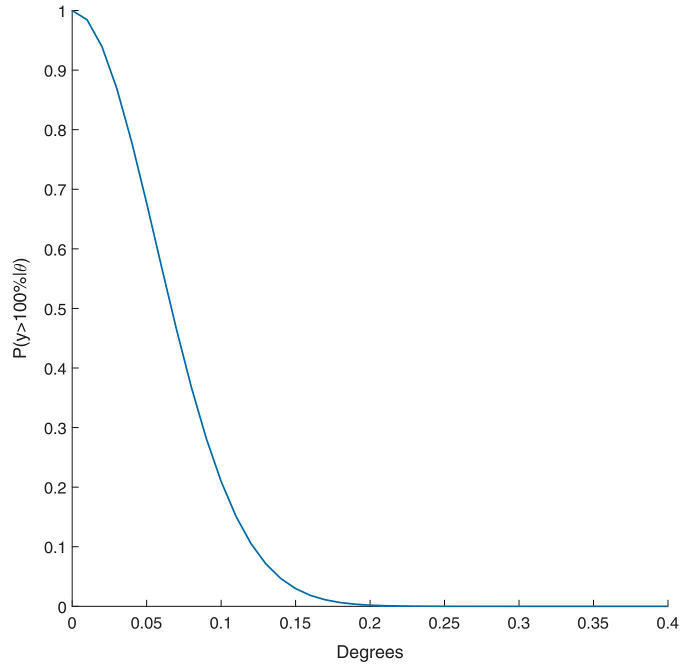

Fig. 5. Decaying probability function with distance for stream gauge sensor likelihood.

the combined likelihood of data from stream gauges indicating a flood for each parish is needed. The likelihood of observed data for all stream gauges is calculated as shown in Eq. (5):

$$P(y > 100\%|\theta) = \sum_{j=1}^{n} P(y_j > 100\%|g_j, \theta)P(g_j) \qquad (5)$$

If no data from other sources are available, as is the case for some parishes on August 10, 2016, the probability $P(y > 100\%)$ is computed, and the need to compute joint probabilities between multiple data sources in Step 6 is eliminated. To calculate $P(y > 100\%)$, first, $P(y_j > 100\%)$ is calculated for each stream gauge empirically by dividing the number of days the gauge read $y_j > 100\%$ by the total number of days on which data were collected. These values are then combined with total probability using g_j to result in $P(y > 100\%)$.

The likelihoods of the stream gauges indicating a flood given no flood occurrence in a parish are computed as shown in Eq. (6). The expression is derived using total probability

$$P(y_j > 100\%|\bar{\theta}) = \frac{P(y_j > 100\%) - P(y_j > 100\%|\theta)P(\theta)}{1 - P(\theta)} \qquad (6)$$

In assessing this sensor source, including more or fewer data points by changing the threshold y_j from 100% to another value would change the overall likelihood depending on the likelihoods from the individual stream gauges added or removed. For instance, if more stream gauges are included in the analysis, but some of those stream gauges are farther away from the parish of interest (i.e., with lower likelihoods), the overall likelihood will decrease even though more information is available. In other cases, decreasing the number of stream gauges available may increase the final posterior probabilities if those that remain are in or very close to that parish, making the overall likelihoods of stream gauge data at or close to 1. Due to the nature of the data collected, the individual likelihoods of each stream gauge are important in calculating this specific sensor source likelihood. This is in comparison to the calculation of the Twitter data source likelihood as discussed in the following section.

Likelihood of Tweet Relevance to a Flood Event ($z = 1$)

Tweets indicating a flood event in a parish during August 10–13, 2016, were collected from Twitter and classified by the built SVM model. The likelihood of tweet relevance to a flood event, $P(z = 1|\theta)$, is calculated using the number of tweets indicating a flood in each parish. $P(z = 1|\lambda = 1)$ is the recall calculated by the performance of the classifier. The number of tweets available is an indicator of classification accuracy (Musaev et al. 2014). Therefore, in Eq. (7), the recall accuracy metric is factored by the number of tweets indicating a flood in the considered parish N_P to compute the probability of tweet relevance given θ. The authors add this uncertainty to the likelihood calculation because most of the tweets are geolocated based on the presence of Louisiana city or town names in their texts, which does not guarantee correct categorization of indicated tweets by parish. Moreover, there are many fewer tweets available for each parish compared to the total number of tweets for each day. The classifier is therefore assumed to be less likely to predict the relevance of these specific tweets of interest correctly. In reducing the accuracy metric by $N_P/(N_P + 1)$, the likelihood of tweet relevance in a parish is higher when there are more relevant tweets in that parish. That is, if there are more relevant tweets mapped to a parish, the higher the source likelihood, $P(z = 1|\theta)$, will be. In contrast with the overall likelihood calculations for stream guage data previously described, the likelihood of this social sensor source given a flood event in a parish does not depend on varying likelihoods of individual tweets due to the nature of the data collection and observations

$$P(z = 1|\theta) = P(z = 1|\lambda = 1)\left(\frac{N_P}{N_P + 1}\right) \tag{7}$$

To calculate the probability the model will classify any tweet as relevant, $P(z = 1)$, the total number of tweets classified as relevant on a day is divided by the total number of tweets collected on each day, regardless of location. $P(z = 1)$ is taken as the same value for all parishes. This value is necessary on August 10, 2016, when several parishes are referred to by tweets classified as relevant to a flood, but no stream gauge data are available because those parishes are too far away from the only available stream gauge indicating a flood on that day. Therefore, Twitter is the only available data source for those parishes, and the joint probability calculation in Step 6 is replaced by $P(z = 1)$, as is the case when only stream gauge data are available and only $P(y > 100\%)$ is needed. Finally, $P(z = 1|\bar{\theta})$ is derived from total probability just as $P(y > 100\%|\bar{\theta})$ was calculated in Eq. (6) to calculate the likelihood of data for the complement of θ.

The authors use the calculations and integration of data likelihoods within each network in this application to demonstrate the required information to use in Step 4 of the proposed framework. The specific calculations for this step will vary depending on the nature of the data collected for each source. For the example, the gauge height data represents a source for which the likelihoods of individual observations can be integrated, while Twitter data represent a source for which the number of indications of θ and the accuracy of classifying individual observations can be used to obtain the overall likelihood of the source. The authors acknowledge assumptions and simplifications made in the analysis of both data sources for indicating flood events may not include other factors used in more comprehensive flood modeling and detection. The data collection and integration of information from these sources in the example explicates use of the framework for data with different likelihoods.

Steps 5–7: Probabilities of Data, Integration of Data Likelihoods, and Final Updating

The data from each source, gauge height and tweet relevance, are conditionally independent given the event of interest θ. The joint likelihood of data from both sources on each day is calculated using Eq. (8):

$$P(y > 100\%, z = 1|\theta) = P(y > 100\%|\theta)P(z = 1|\theta) \tag{8}$$

The joint probability of observing the data from both sources is calculated using Eq. (9), with information included on both states of a flood event in a parish: occurred, θ, and not occurred, $\bar{\theta}$

$$P(y > 100\%, z = 1) = P(y > 100\%|\theta)P(z = 1|\theta)P(\theta) + P(y > 100\%|\bar{\theta})P(z = 1|\bar{\theta})P(\bar{\theta}) \tag{9}$$

For the application, the authors also sequentially update data from August 11–12, 2016, and August 12–13, 2016. The probability of data from the 2-d intervals is calculated using Eq. (10). Subscript 1 refers to likelihoods calculated on one day, and subscript 2 refers to likelihoods calculated on the following day:

$$\begin{aligned} &P(y_1 > 100\%, z_1 = 1, y_2 > 100\%, z_2 = 1) \\ &\quad = P(y_1 > 100\%, z_1 = 1|\theta)P(y_2 > 100\%, z_2 = 1|\theta)P(\theta) \\ &\qquad + P(y_1 > 100\%, z_1 = 1|\bar{\theta})P(y_2 > 100\%, z_2 = 1|\bar{\theta})P(\bar{\theta}) \end{aligned} \tag{10}$$

For the application, the final posterior probabilities for each parish in Louisiana are computed using Eqs. (11) and (12), with Eq. (12) used for sequential updating. The authors limit the use of sequential updating here to two days because with the amount of data available during this time period, updating using more than two days of data results in posterior probabilities of nearly 1 in several parishes with the remaining parishes receiving little to no updating

$$P(\theta|y > 100\%, z = 1) = \frac{P(y > 100\%|\theta)P(z = 1|\theta)P(\theta)}{P(y > 100\%, z = 1)} \tag{11}$$

$$\begin{aligned} &P(\theta|y_1 > 100\%, z_1 = 1, y_2 > 100\%, z_2 = 1) \\ &\quad = \frac{P(y_1 > 100\%, z_1 = 1|\theta)P(y_2 > 100\%, z_2 = 1|\theta)P(\theta)}{P(y_1 > 100\%, z_1 = 1, y_2 > 100\%, z_2 = 1)} \end{aligned} \tag{12}$$

In cases when data from one source are completely unavailable (e.g., if there are no relevant tweets in a parish), the prior risks are updated with information from only the other source. When there are no data available from any source, the prior risk remains unchanged. In other applications with more than two sources, the joint probabilities of different combinations of sensor sources must be computed.

Results

Updated Probability Distributions and Flood Event Estimation over Time

Using the proposed framework for integrating data across multiple sources, the resulting updated probability distributions for flood risk by parish are mapped in Fig. 6 for each day of the 4-d period of investigation. The lowest probabilities in light yellow show the parishes that had little or no data with which to update their prior probabilities of flood events. From Fig. 6, few updated probabilities are computed on August 10. The highest updated probabilities of flood occurrence in a parish on each day from August 10–13, 2016, are 0.05 (Orleans Parish), 0.45 (Ascension Parish), 0.73 (St. Tammany Parish), and 0.81 (Livingston Parish), respectively. The largest changes in prior to posterior risks occurred on August 12.

Fig. 7 shows sequential updating results, from August 11 to 12 and from August 12 to 13, and the effect of combining data from multiple days. The largest updated probability of flood occurrence in a parish after sequential updating from August 11 to 12 was 0.998 (East Baton Rouge), and for sequential updating from August 12 to 13, 0.999 (East Baton Rouge and Livingston).

Updated Probability Distributions and Flood Event Estimation by Sensor Type

As an objective of this study is to integrate multiple data sources for event estimation, the authors investigate the effect of additional data on the estimation, specifically looking at the results using data from single sensor sources compared to combining the chosen data sets. Fig. 8 shows the results of updating probabilities of flood risk on August 12, 2016, by parish based on only stream gauge data, only Twitter data, and then using information both sensor sources.

Using information from both sources significantly increases the updated probabilities. For instance, in St. Tammany Parish, three stream gauges indicated a potential flood event based on increased gauge height and two tweets were classified as relevant to a flood event. The prior probability is computed to be 0.0065. The resulting updated probabilities are 0.17 and 0.08 for updating with gauge height data and tweet relevance data alone, respectively. When data from both sources are included in the inference, the updated probability of flood risk is 0.73. This is seen for all parishes that have data with which to update their prior probabilities of flood risk; updating with an individual data source does not result in a posterior probability higher than 0.20 on August 12, while both data sources combined update probabilities up to 0.73. This is due to the joint probabilities of observing data from multiple sources being smaller than the probabilities of observing data from one source alone. Therefore, even as more data are available from one source, the probabilistic estimations experience the most updating when multiple data sources indicate the same event θ.

Validation and Data Availability

The results of the approach and analyses are compared to the true flooding that occurred in August 2016 in Louisiana for validation. The floods caused a Major Disaster Declaration from FEMA, and 26 parishes were designated for Individual or Public Assistance, shaded in Fig. 9 (FEMA 2016). The results in Figs. 6 and 7 show increased flood risks after updating in most of the parishes listed in the Major Disaster Declaration. Of the 26 parishes with a declared Disaster Declaration, 17 are updated based on the data from August 12, and an additional six parishes not updated on August 12 are updated based on data from August 13. Fig. 10 shows

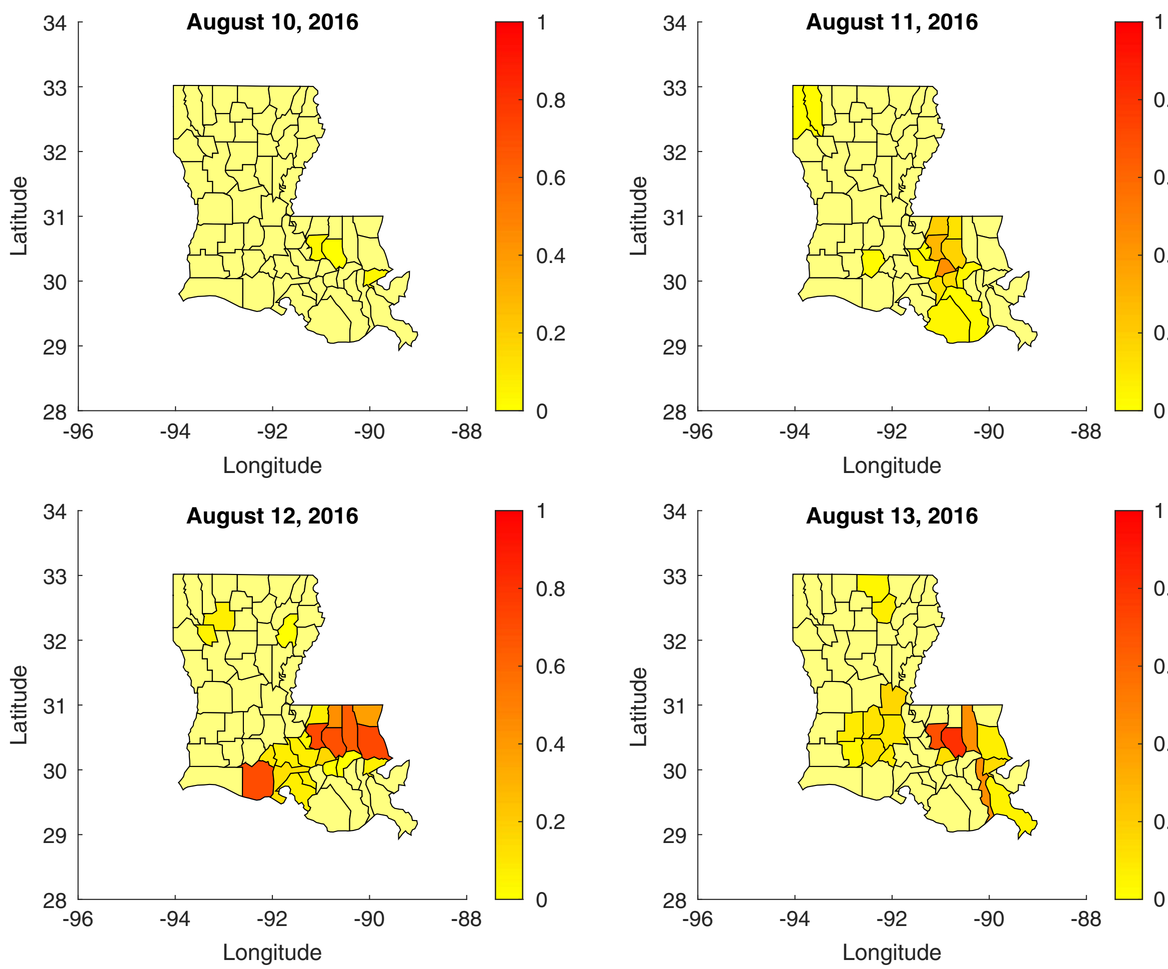

Fig. 6. Resulting updated probability distributions for August 10, 11, 12, and 13.

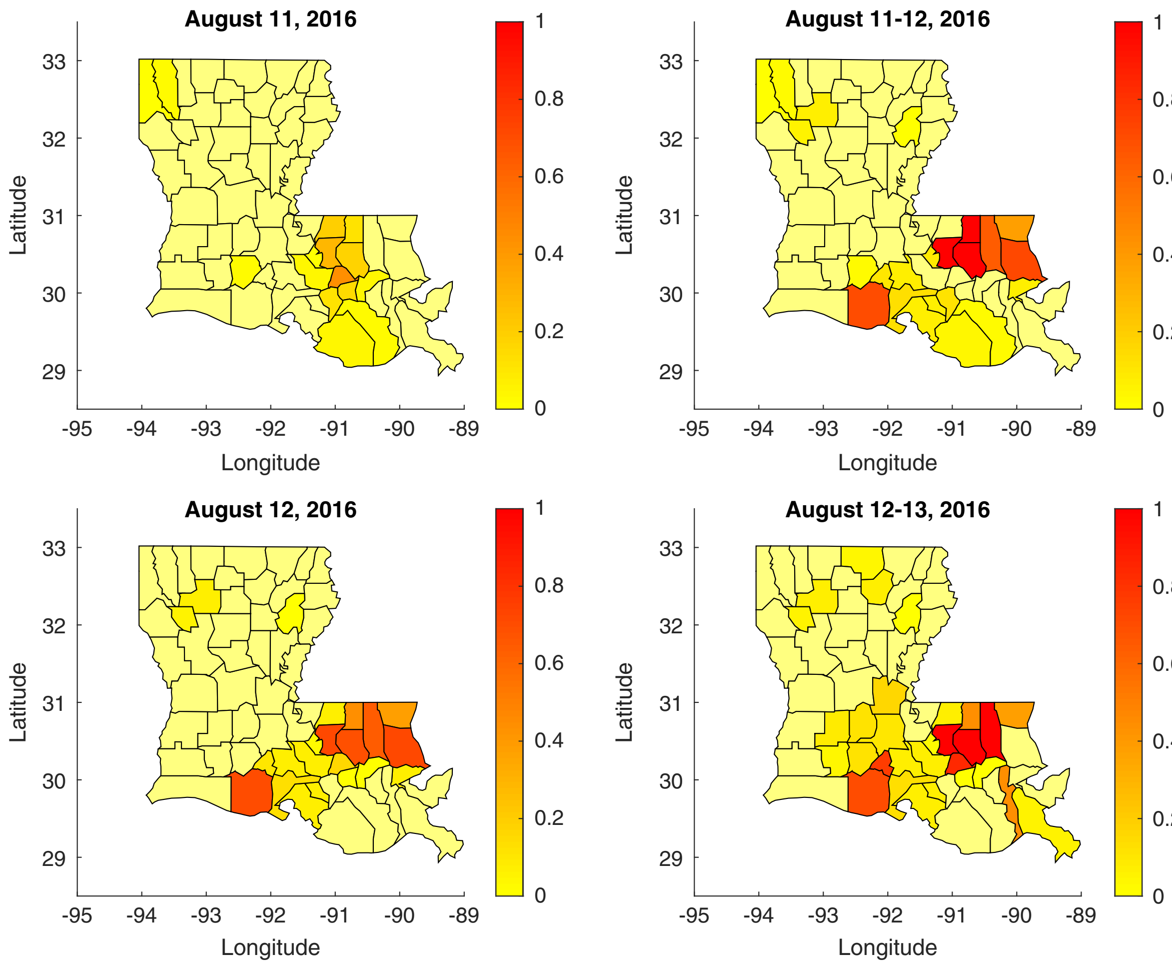

Fig. 7. Updating and sequential updating results with sensor observations from August 11 to August 11–12 and August 12 to August 12–13.

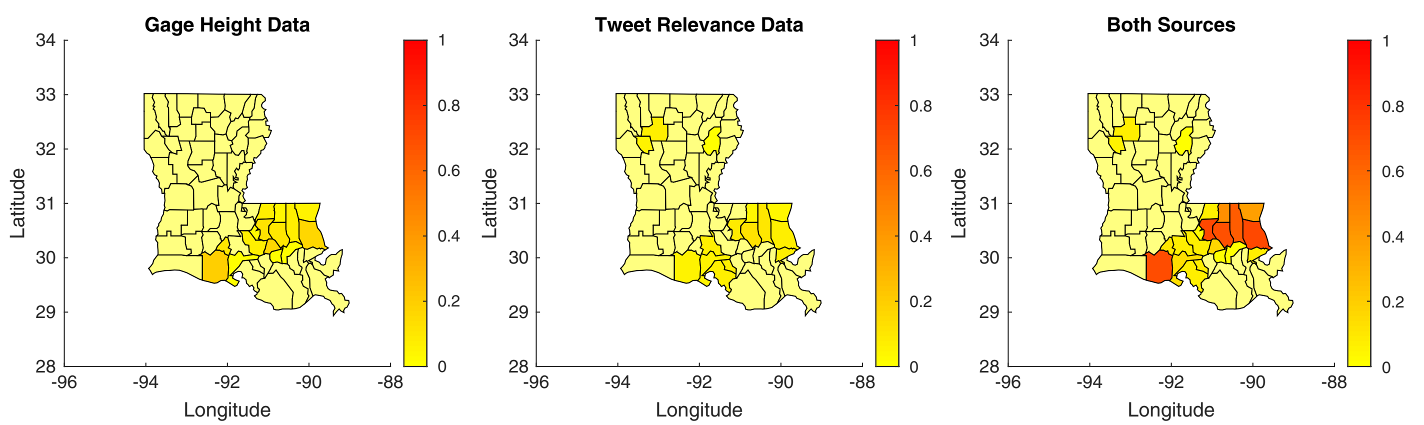

Fig. 8. Resulting updated probability distributions for August 12, 2016, using gauge height and tweet relevance data sources alone compared to estimation from integrating both data sources.

parishes with updated probabilities in dark blue and parishes without data to update priors in green for August 12 and 13, and after sequential updating from August 12–13.

The updated probabilities also include nine parishes not in FEMA's Disaster Declaration. This can be explained by a number of factors. FEMA considers numerous variables when deciding on areas for Disaster Declarations. Some of these factors include localized impacts, insurance coverage in force, estimated cost of assistance, and other federal assistance programs (FEMA 2017). Ultimately, flooding may have occurred in parishes updated by data, even if they are not part of the Disaster Declarations.

For those parishes included in the Disaster Declaration that did not have updated probabilities of flood risk from the application, this is due to a lack of availability of data indicating a flood event for those particular parishes. The presence of data is essential in the framework to estimate the event of interest. For any application,

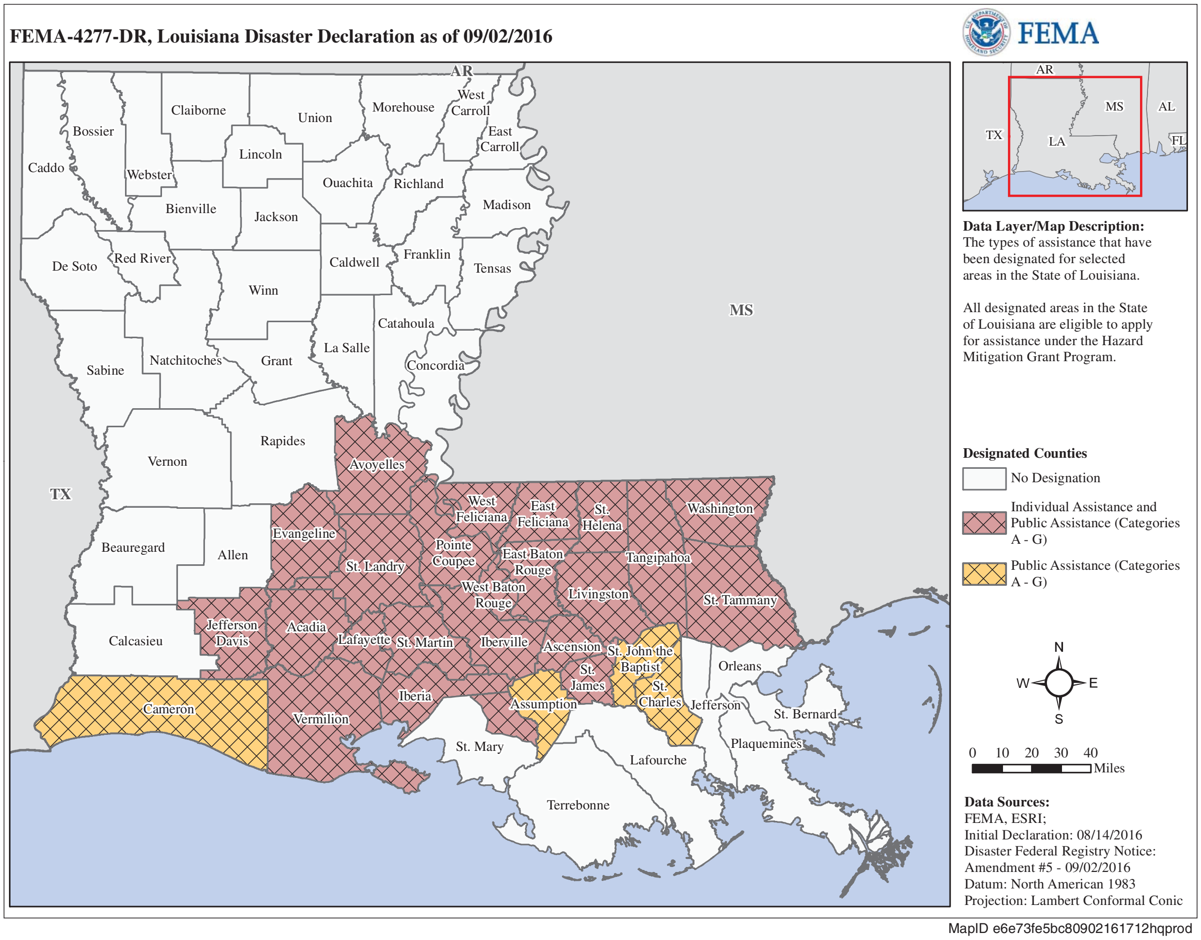

Fig. 9. Louisiana Disaster Declaration for flooding event in August 2016, FEMA-4277-DR. (Reprinted from FEMA 2016.)

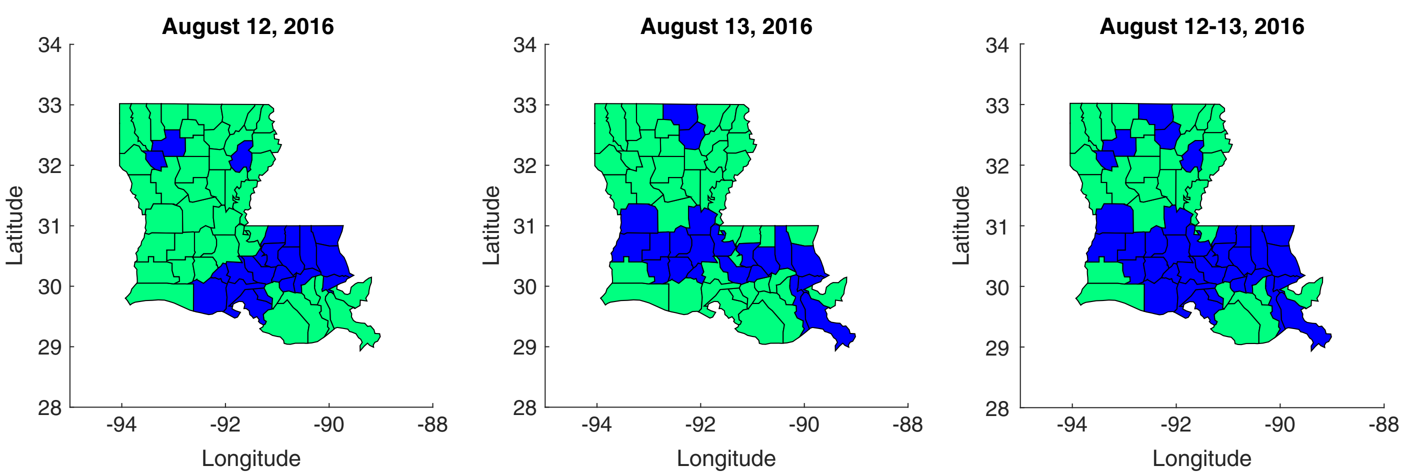

Fig. 10. Parishes with updated probabilities of flood risk (dark blue) from integrated data sources on the dates shown.

different sensor sources or different measures for which sensors indicate the event of interest would create different results. Fig. 11 shows the availability of data indicating θ over time for each parish. The number of available observations is shown, for counts of stream gauges with sudden increases in gauge height, number of tweets classified as relevant, and total daily data point counts from combining both sources.

To further validate the results, the authors examine USGS post-event maps to qualitatively evaluate the results of applying the framework to the application example. After the flood events in August 2016, USGS created a report to summarize the flooding and developed several flood inundation maps based on high-water marks. The report also included a map of cumulative rainfall across the state during August 11–14, 2016. This and the inundation map for Louisiana with shaded areas indicating inundation is shown in Fig. 12 (Watson et al. 2017).

Specifically, extensive inundation around the Amite and Tangipahoa Rivers affected parishes in the northeastern portion

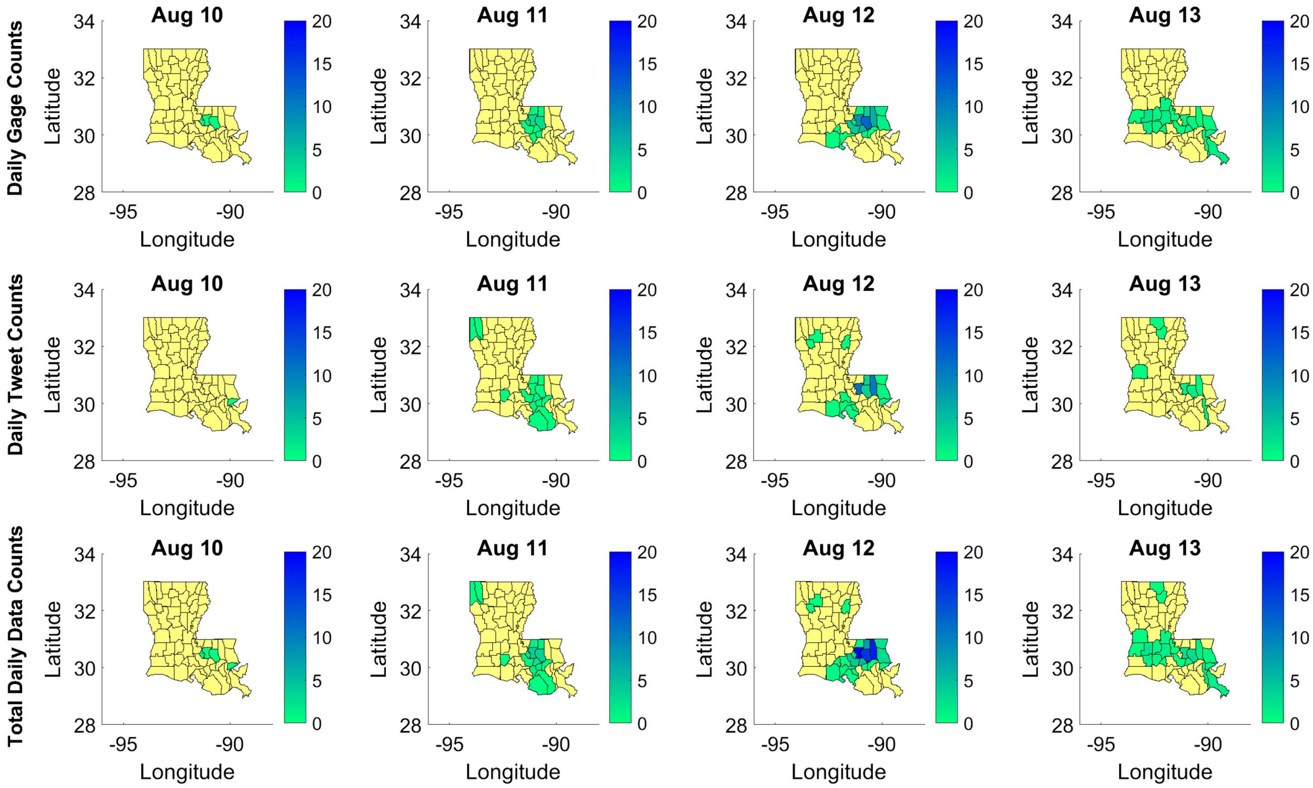

Fig. 11. Stream gauge and Twitter data availability by parish from August 10–13, 2016.

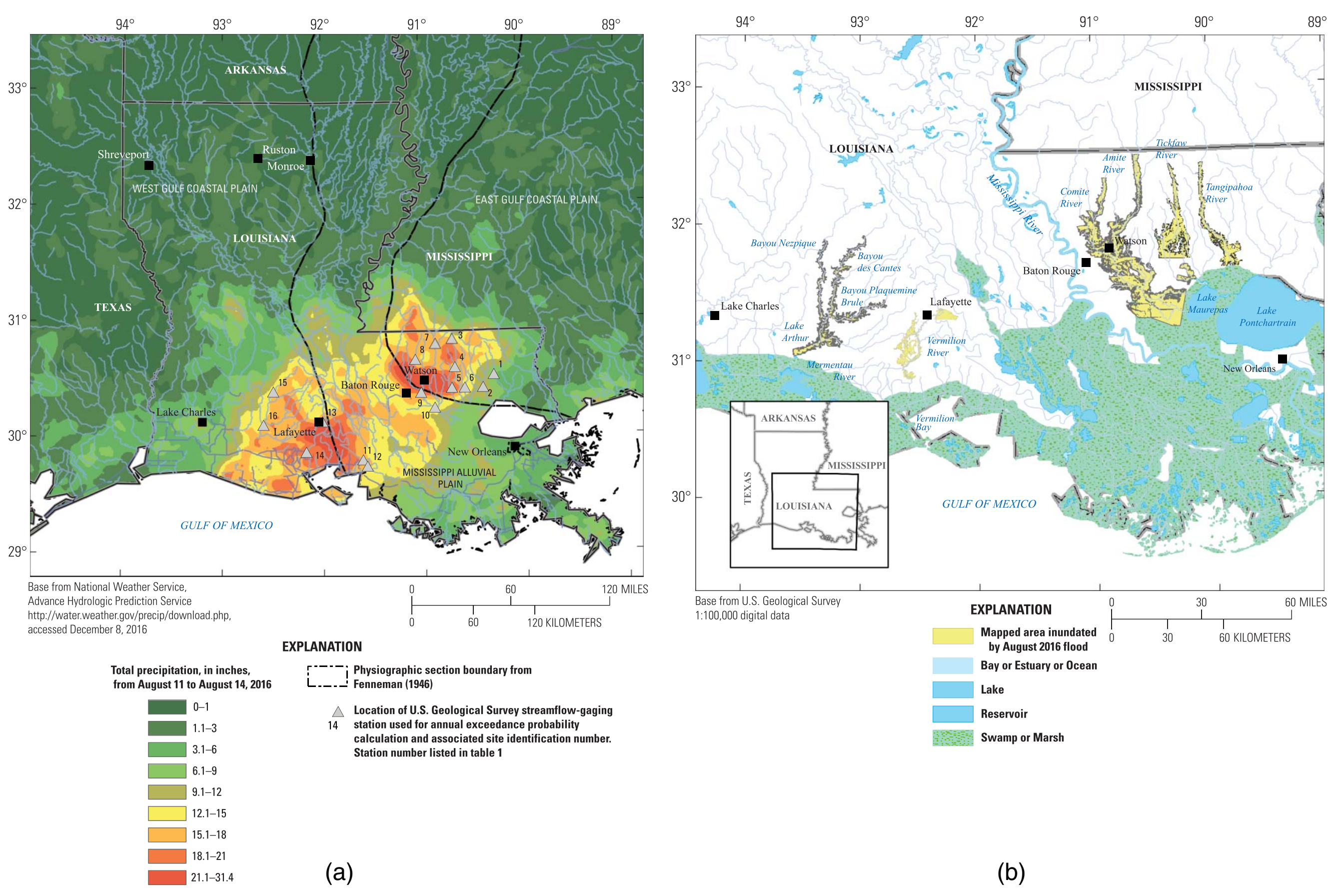

Fig. 12. (a) Precipitation in Louisiana from August 11–14, 2016 (physiographic boundaries from Fenneman 1946; base map courtesy of National Weather Service, Advance Hydrologic Prediction Service); and (b) inundation map for Louisiana. (Reprinted from Watson et al. 2017.)

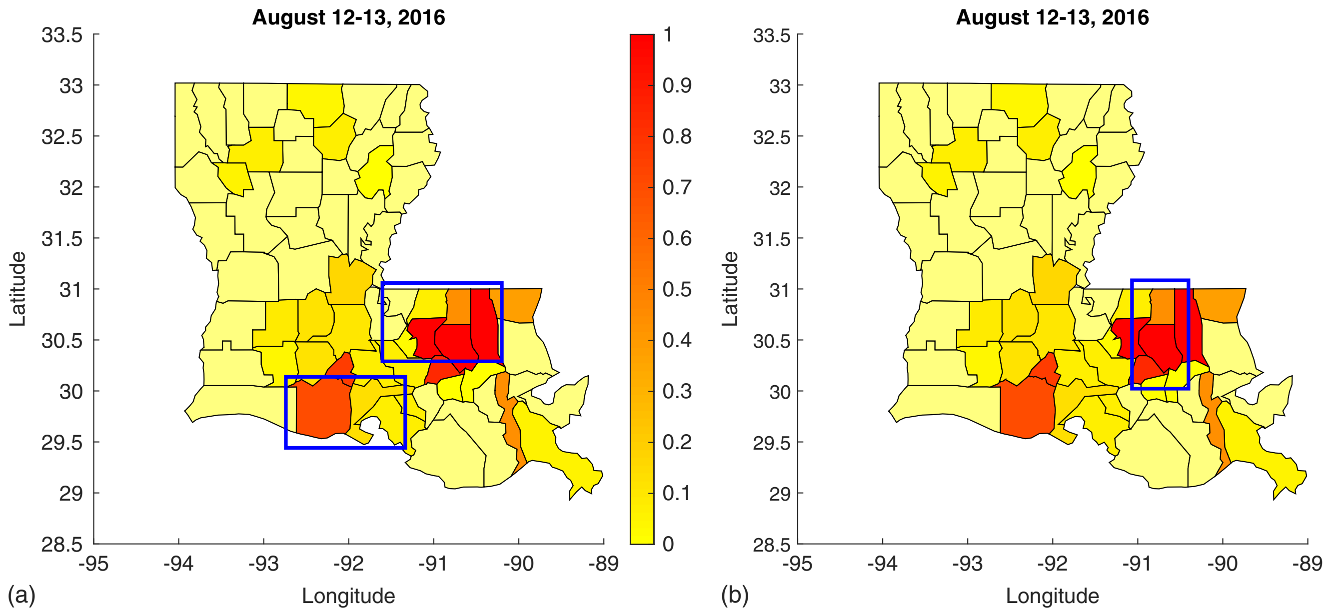

Fig. 13. Map of sequential updating from August 12–13, 2016, with regions of highest precipitation (a) and confirmed inundation (b) from Fig. 12 highlighted.

of Louisiana's boot region. These include three of the parishes with the highest updated probabilities of flood occurrence using the proposed approach: East Baton Rouge Parish, Livingston Parish, and St. Tammany Parish. These correspond to the areas of highest precipitation during the time period investigated and with areas of confirmed inundation. The integrated data during August 10–13, 2016, update the prior probabilities of flood occurrence from 0.0055, 0.008, and 0.0064 to 0.73, 0.81, and 0.73 for East Baton Rouge, Livingston, and St. Tammany parishes, respectively. Fig. 13 shows the results of sequential updating for the application from August 12–13, 2016, with regions of highest precipitation (1) and verified inundation (2) from Fig. 12 highlighted. The regions of highest precipitation [shown in dark orange and red in Fig. 12(a)] correspond to the boxed parishes in Fig. 13(a). The inundation around the Amite, Comite, Tickfaw, and Tangipahoa Rivers mapped in Fig. 12(b) correspond to the boxed region in Fig. 13(b).

Conclusions

The framework presented in this paper provides a unique probabilistic approach to integrating data from across sources to estimate the probability of disaster or failure event occurrence given observed data. It updates prior probabilities of event occurrence with both individual and combined data sources. The framework is able to include data from a wide range of sensor types with varied likelihoods and shows how prior risks of an event change as new, potentially anomalous data are introduced. It is applicable to general disaster or failure events, including natural disasters and structural or infrastructure system failures as long as data are available. The Bayesian updating approach for data integration does require the establishment of prior probabilities of event occurrence. If these are unknown, they are initially assumed with the potential use of uninformative priors to limit bias in the results.

In the application example, the authors apply the framework to estimate flood events in Louisiana in August 2016. Prior flood event risk in a parish is calculated based on FEMA risk map data. For updating in the application, physical stream gauge data and social Twitter data are used. Data from other sensor sources can be easily added using the same approach presented, with new likelihood calculations for each additional sensor type. While these likelihoods will be defined differently depending on the data output by each source, the general framework is flexible such that changing these will not change the implementation process as long as the sensor source likelihoods can be found.

The results from updating prior flood risks in Louisiana from August 10–13, 2016, show that additional data over time and from across both sensor sources increase the amount of updating possible in real-time event estimation. The results are validated by comparing the parishes with highest updated probabilities with FEMA Disaster Declarations and postevent inundation and precipitation maps from USGS. This showed similar regions of flooding indicated based on updating from the integrated data sets as from the true event.

Probabilistic updating using the proposed framework increases situational awareness and can be used to support community decision-making during and after disaster or failure events. The impacts of the results obtained from integrating data using the proposed framework are in three main areas:

1. Updating prior risk assessments based on integrated inferences from multiple data sources improves situational awareness, particularly if done in real-time, with updated probabilities indicating locations or components most likely to be experiencing the event at the time assessed. In the example application, communities have a more holistic view of their flood risk from monitoring both conditions from the natural environment and first-hand accounts from community members of a current event.
2. Based on the granularity of the estimation, comparing updated probabilities facilitates prioritization of resources by location and time during a disaster or failure event. In the application, directly comparing flood risks between parishes supports real-time decision-making and resource allocation during emergency response by identifying the most vulnerable parishes.
3. The framework enables assessment of the monitoring capabilities of different data sources. By integrating data first within networks then across sources, the updating approach reveals the

availability of data and the updated probabilities from individual compared to combined sources postevent. These results evaluate a community's monitoring capabilities, demonstrate what types of data are available throughout the community, and detail where multiple data sources can effectively supplement each other. The application shows how stream gauge data and Twitter data contribute to updating prior risks, both individually and particularly when combined.

Acknowledgments

The authors acknowledge support for this work from the National Science Foundation through Grant No. CNS-1541074.

References

Alexander, D. 2014. "Social media in disaster risk reduction and crisis management." *Sci. Eng. Ethics* 20 (3): 717–733. https://doi.org/10.1007/s11948-013-9502-z.

Burby, R. 2001. "Flood insurance and floodplain management: The US experience." *Global Environ. Change Part B: Environ. Hazard.* 3 (3): 111–122. https://doi.org/10.3763/ehaz.2001.0310.

Durrant-Whyte, H., and T. C. Henderson. 2008. "Multisensor data fusion." In *Springer handbook of robotics*, 585–610. Berlin: Springer.

FEMA (Federal Emergency Management Agency). 2016. "Louisiana severe storms and flooding (DR-4277)." Accessed September 20, 2017. https://www.fema.gov/disaster/4277-0.

FEMA (Federal Emergency Management Agency). 2017. "The disaster declaration process." Accessed September 20, 2017. http://www.fema.gov/declaration-process.

Fenneman, N. M. 1946. "Physical divisions of the United States: U.S. Geological Survey map, scale 1:7, 000, 000, 1 sheet." Accessed June 1, 2017. https://water.usgs.gov/lookup/getspatial?physio.

Groves, D. G., K. Kuhn, J. Fischbach, D. R. Johnson, and J. Syme. 2016. *Analysis to support Louisiana's flood risk and resilience program and application to the National Disaster Resilience Competition*. Rand Corporation 2016: Project Rep. Santa Monica, CA: RAND Corporation.

Hackmann, G., W. Guo, G. Yan, Z. Sun, C. Lu, and S. Dyke. 2014. "Cyber-physical codesign of distributed structural health monitoring with wireless sensor networks." *IEEE Trans. Parallel Distrib. Syst.* 25 (1): 63–72. https://doi.org/10.1109/TPDS.2013.30.

Hall, M., E. Frank, G. Holmes, and B. Pfahringer. 2009. "The WEKA data mining software: An update." *ACM SIGKDD Explor. Newsl.* 11 (1): 10–18. https://doi.org/10.1145/1656274.1656278.

Hirabayashi, Y., R. Mahendran, S. Koirala, L. Konoshima, D. Yamazaki, S. Watanabe, H. Kim, and S. Kanae. 2013. "Global flood risk under climate change." *Nat. Clim. Change* 3 (9): 816–821. https://doi.org/10.1038/nclimate1911.

Ireson, N. 2009. "Local community situational awareness during an emergency." In *Proc., 3rd IEEE Int. Conf. on Digital Ecosystems and Technologies*, 49–54. Piscataway, NJ: IEEE.

Johansen, C., J. Horney, and I. Tien. 2017. "Metrics for evaluating and improving community resilience." *J. Infrastruct. Syst.* 23 (2): 04016032. https://doi.org/10.1061/(ASCE)IS.1943-555X.0000329.

Jongman, B., J. Wagemaker, B. R. Romero, and E. C. de Perez. 2015. "Early flood detection for rapid humanitarian response: Harnessing near real-time satellite and Twitter signals." *ISPRS Int. J. Geo-Inf.* 4 (4): 2246–2266. https://doi.org/10.3390/ijgi4042246.

Khaleghi, B., A. Khamis, F. O. Karray, and S. N. Razavi. 2013. "Multisensor data fusion: A review of the state-of-the-art." *Inf. Fusion* 14 (1): 28–44. https://doi.org/10.1016/j.inffus.2011.08.001.

Leetaru, K. H., S. Wang, G. Cao, A. Padmanabhan, and E. Shook. 2013. "Mapping the global Twitter heartbeat: The geography of Twitter." *First Monday* 18 (5): 1–12. https://doi.org/10.5210/fm.v18i5.4366.

LSU AgCenter and LaDOTD (Louisiana State University Agricultural Center and Louisiana Department of Transportation and Development). 2017. "LA floodmaps." Accessed March 3, 2017. http://maps.lsuagcenter.com/floodmaps/.

Miorandi, D., S. Sicari, F. De Pellegrini, and I. Chlamtac. 2012. "Internet of things: Vision, applications and research challenges." *Ad Hoc Networks* 10 (7): 1497–1516. https://doi.org/10.1016/j.adhoc.2012.02.016.

Mone, G. 2015. "The new smart cities." *Commu. ACM* 58 (7): 20–21. https://doi.org/10.1145/2771297.

Morss, R. E., O. V. Wilhelmi, M. W. Downton, and E. Gruntfest. 2005. "Flood risk, uncertainty, and scientific information for decision making: Lessons from an interdisciplinary project." *Bull. Am. Meteorol. Soc.* 86 (11): 1593–1602. https://doi.org/10.1175/BAMS-86-11-1593.

Musaev, A., D. Wang, and C. Pu. 2014. "LITMUS: Landslide detection by integrating multiple sources." In *Proc., 11th Int. Conf. on Information Systems for Crisis Response and Management*, 677–686. State College, PA: University Park.

Musaev, A., D. Wang, and C. Pu. 2015. "LITMUS: A multi-service composition system for landslide detection." *IEEE Trans. Serv. Comput.* 8 (5): 715–726. https://doi.org/10.1109/TSC.2014.2376558.

Pakzad, S. N., G. L. Fenves, S. Kim, and D. E. Culler. 2008. "Design and implementation of scalable wireless sensor network for structural monitoring." *J. Infrastruct. Syst.* 14 (1): 89–101. https://doi.org/10.1061/(ASCE)1076-0342(2008)14:1(89).

Rawat, P., K. D. Singh, H. Chaouchi, and J. M. Bonnin. 2014. "Wireless sensor networks: A survey on recent developments and potential synergies." *J. Supercomput.* 68 (1): 1–48. https://doi.org/10.1007/s11227-013-1021-9.

Sakaki, T., M. Okazaki, and Y. Matsuo. 2010. "Earthquake shakes twitter users: Real-time event detection by social sensors." In *Proc., 19th Int. Conf. World Wide Web*, 851–60. New York: ACM.

Smith, L. L. 2005. *Louisiana directory of cities, towns, and villages*. Louisiana: Sims Memorial Library–Southeastern Louisiana Univ.

Tien, I., A. Musaev, D. Benas, A. Ghadi, S. Goodman, and C. Pu. 2016. "Detection of damage and failure events of critical public infrastructure using social sensor big data." In *Proc., Int. Conf. on Internet of Things and Big Data*, 435–440. Setúbal, Portugal: Science and Technology Publications.

USGS (United States Geological Survey). 2017. "National water information system (NWIS) data available on the World Wide Web (Water Data for the Nation)." Accessed May 7, 2017. https://nwis.waterdata.usgs.gov/usa/nwis/uv/?cb_00060=on&cb_00065=on&cb_63158=on&format=gif_default&site_no=07383500&period=&begin_date=2016-08-11&end_date=2016-08-15.

Van Der Wiel, K., S. B. Kapnick, G. Jan Van Oldenborgh, K. Whan, S. Philip, G. A. Vecchi, R. K. Singh, J. Arrighi, and H. Cullen. 2017. "Rapid attribution of the August 2016 flood-inducing extreme precipitation in south Louisiana to climate change." *Hydrol. Earth Syst. Sci.* 21 (2): 897–921. https://doi.org/10.5194/hess-21-897-2017.

Vieweg, S., A. L. Hughes, K. Starbird, and L. Palen. 2010. "Microblogging during two natural hazards events." In *Proc., 28th Int. Conf. on Human Factors in Computing Systems*, 1079–1088. New York: ACM.

Vieweg, S., L. Palen, S. B. Liu, A. L. Hughes, and J. Sutton. 2008. "Collective intelligence in disaster: An examination of the phenomenon in the aftermath of the 2007 Virginia Tech shootings." In *Proc., 5th Int. Conf. on Information Systems for Crisis Response and Management*. Boulder, CO: Univ. of Colorado.

Watson, K. M., J. B. Storm, B. K. Breaker, and C. E. Rose. 2017. *Characterization of peak streamflows and flood inundation of selected areas in Louisiana from the August 2016 Flood*. U.S. Geological Survey Scientific Investigations Rep. No. 2017-5005. Reston, VA: US Geological Survey.

Yin, Y., A. Lampert, M. Cameron, B. Robinson, and R. Power. 2012. "Using social media to enhance emergency situation awareness." *IEEE Intell. Syst.* 27 (6): 52–59. https://doi.org/10.1109/MIS.2012.6.