I. Project Information*

Project Director	Jamie Padgett
Project Title	Understanding and Overcoming Data Biases to Enable More Equitable and Reliable Flood Alert
Project Location	Houston
Project Summary	This project addresses a need for timely and reliable information on flood conditions and their impacts, which hamper response, public health and safety in communities affected by severe storms. Via community-engaged research conducted in partnership with the City of Houston's Health Department (HHD), we derive and pilot a new framework for Equitable and Reliable Flood Alert Systems (ERFAS). This new framework targets the biases and inequities in data quality and availability in underserved communities by providing a smart system that fuses data from diverse sources (e.g. radar-based, authoritative, camera, social media, sensor) while positioning flood estimates in the context of the community. Four thrusts underpin this research: 1) Thrust 1 measures social vulnerability, including a new flood-related health hazards composite metric (FHSV), providing a tailored lens through which to view social equity of information. 2) Thrust 2 characterizes the data sources that may underpin flood alert systems in terms of their availability, reliability, and biases to equitably predict flood conditions and their impacts such as access to critical facilities. 3) Thrust 3 overcomes the limitations of traditional systems by fusing data for robust, real-time, equitable estimates of flooding and its impacts. 4). Thrust 4 engages community partners and stakeholders in ERFAS development and testing, including informant interviews and scenario immersion. While piloted in the underserved, minority, low income region of Hunting Bayou in Houston, the methods are portable to address injustices that other communities face regarding data biases to support decision-making related to public health and safety during floods.

II. Progress Report Questions

1. Please revisit your proposal and review your goals and the outcomes you were seeking to achieve through this grant. How successful were you in meeting your goals? Please assess your success against the criteria you set in your proposal and use any combination of anecdotes, stories, graphs, charts, visuals as well as data to explain your success. Upload supporting files if you choose.*

Please see attached file for response to Q1

Optional File Upload

Q1_ERFAS Report_Final_JEP.pdf

Filename: Q1_ERFAS Report_Final_JEP.pdf **Size:** 1.4 MB

2. How has your work benefited your organization, professional field, community, or other stakeholders?*

The multidisciplinary work presented in this report makes several key contributions to smart resilience, situational awareness, emergency response, and data equity research areas. In addition to offering methodological contributions highlighted in our answer to Question 1, this study generated multiple data sets, open-source data processing workflows, and a web tool. Models and methods proposed in this research offer a practical pathway for flood-prone communities to repurpose their existing data sources to acquire equitable and reliable situational awareness data during storms--- a vital contribution in an epoch of exacerbating flood risk. Insights gained from this study will be used to develop an urban digital twin that will use resilience models, urban data analytics, and multi-sensor data fusion to track flood impacts in a coupled social-physical-cyber system to inform emergency response decision-making. Disciplinary barriers were broken and our team of social scientists, engineers, and computer scientists were able to make great strides.

This study helped us expand our work on flood situational awareness and deepen partnerships between Rice University, the Houston Health Department, and the Harris County Flood Control District. These partner organizations have a strong interest in expanding the modeling platform across our region and pursuing deep-dive verification and validation with past/future event data. This project laid a solid foundation for future testing and implementation. However, there is much work left to do on this front before a publicly deployed tool is ready.

3. Are there any other successes related more broadly to this project that you would like to share with us?*

The OpenSafe Fusion framework and the accompanying web tool developed through this project have garnered interest among policymakers in Houston and globally. Specifically, Harris County Flood Control District is interested in partnering with us to deploy the OpenSafe Fusion framework in Houston, Texas. Similarly, The National Transportation Planning and Research Centre (NATPAC) (a research, development, and planning center in India) is exploring pathways to adopt OpenSafe Fusion in Kochi, a major port city in Southern India.

4. What did you learn (positive or negative) as a result of this grant? What lessons would you share with other organizations or the field at large?*

A paradigm shift in community resilience is underway, primarily enabled by recent technological developments in artificial intelligence, connected communities, and big data. The transition to smart resilience offers both opportunities and challenges. As shown in our research, intelligent technologies could significantly improve community resilience with minimal physical infrastructure investment, albeit they require considerable investment in emerging technologies. Indiscriminate application of machine learning models, often trained on biased or limited data, could lead to biased models that could exacerbate inequities and perpetuate a cycle of unjust investments. Consequently, developing equity- and bias-aware models and frameworks is necessary to ensure the equitable distribution of benefits from the paradigm shift to intelligent technologies and to facilitate equitable community resilience. There are many unexplored opportunities, if not necessities, in this space.

We also learned that community engagement and user-centered design are essential for guiding research to address inequities in urban resilience. Specifically, interactions with community stakeholders will provide insights into key challenges and help streamline research focus to address complex challenges with significant societal impacts.

5. How do you characterize your relationship with the GRP and what suggestions do you have for improvement?*

We had a pleasant experience working with GRP. The community-centric work of GRP is a good fit for our research interests, and we look forward to future collaborations. We appreciated the opportunity to meet and learn from other projects funded in our cohort, and the valuable discussions and connections that the GRP staff offered at our check-ins. Furthermore, GRP did a fantastic job organizing the panel discussion titled “Bridging Diverse Knowledge Systems to Address Flood Risk in Northeast Houston Communities” (26-27 April 2023). The event was very helpful in developing new connections that helped inform our research. Continuing to organize events to bring researchers and community stakeholders together would be excellent.

6. Please provide any other feedback or comments you have for the GRP.*

This initial funding provided a valuable start—both in relationship development, integrated academic-community stakeholder partnership, data collection, research methods advancement and tool development. However, truly impactful community engaged work requires a much more significant timeline and level of funding. If there is an opportunity for a second phase of funding, we would be delighted to pursue it! Such continuation would allow us to extend and deepen this community-engaged work, afford future testing and feedback, and expand the work beyond our case study area. Further, it would allow us to develop tools, methodologies, and datasets that will serve as the foundation for a broader Gulf Coast regional or even national framework for reliable and equitable situational awareness during storms that is accessible to all communities.

7. If applicable, please identify and describe the ways you or your organization leveraged GRP’s grant (e.g., other funders, volunteers who worked on the program, in-kind donations etc.) Please specify the value and/or number/hours of volunteers if possible.

(No response)

Q1. Please revisit your proposal and review your goals and the outcomes you were seeking to achieve through this grant. How successful were you in meeting your goals? Please assess your success against the criteria you set in your proposal and use any combination of anecdotes, stories, graphs, charts, visuals as well as data to explain your success.

Response: The objectives of the research project were to understand and overcome biases in mobility-centric flood situational awareness data to enable more equitable and reliable flood alerts. We successfully achieved the goals identified in the project proposal for all key focus areas. Our research activities were organized into four research thrusts, and these thrusts are:

1. Contextualizing flood impacts and social vulnerability in support of public health and safety
2. Characterizing data quality, biases, and data-to-knowledge pipelines
3. Fusing data for robust, real-time, equitable estimates regarding flooding and its impacts
4. Conducting community-engaged research and evaluating impact of an equity-informed flood alert system

Under Thrust 1, we focused on developing methods to measure social vulnerability within the target community. Our initial focus was on developing a new flood-related health hazards composite social vulnerability metric (FHSV).

- From the literature review on the social vulnerability metrics and extensive interactions with stakeholders in Houston (e.g., the City of Houston and Houston Galveston Area Council), we concluded that an index-based approach alone (as proposed in our proposal) is insufficient. Indices usually presented in a unified scale (e.g., CDC/ATSDR Social Vulnerability Index (SVI) is a value between 0 and 1) provide little insight into the community characteristics and their needs---thus failing to provide actionable input or inform risk mitigation or adaption decision-making. For example, consider a census tract with a 0.7 value in the CDC/ATSDR SVI. Since the value is on the higher end of the 0 to 1 SVI scale, it conveys that the community might be vulnerable. However, it does not provide any information on the characteristics of the population or their needs. For informing decision-making, additional context on the population characteristics is required.
- We proposed a new clustering-based alternative to indices such as CDC SVI. The newly developed clustering approach divides communities into homogenous clusters based on social, demographic, and locational attributes. The identified clusters provide insights into the characteristics of the community, including its residents and built environment, and suggest the potential for differing community needs in the context of flooding. Figure 1 shows a case study application of the proposed method in Houston. As an example, Cluster 5, shown in purple, are communities that tend to have older housing and a high proportion of residents with disabilities. Residents in such communities may have limited ability to self-evacuate during a flood event and need assistance for evacuation.

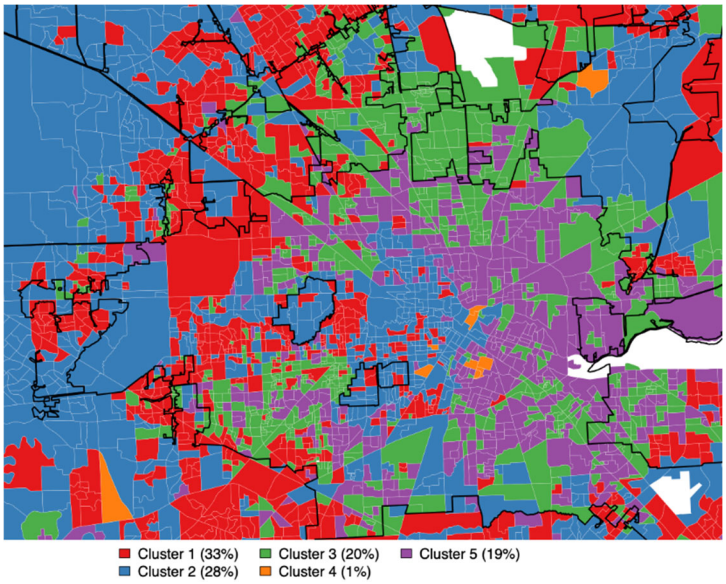

Figure 1. A case-study application of the developed cluster-based approach for characterizing social vulnerability in Houston.

In Thrust 2, we focused on characterizing data quality, biases in data sources, and developing data-to-knowledge pipelines. We identified data sources in Houston, designed data processing workflows, and characterized data sources on quality, availability, and bias.

- We successfully identified diverse data sources in Houston that can provide real-time road condition data. Example sources include physical sensors (USGS water level gages, traffic cameras, traffic speed data), social sensors (crowdsourcing, social media, citizen service portals), mathematical models (physics-based flood models), authoritative sources (traffic alerts), and hybrid models (human-in-the-loop framework). We collaborated with different organizations to gain access to real-time and historical data. For example, the Texas Department of Transportation and Houston TranStar helped us gain access to real-time data such as traffic speed, traffic alerts, and traffic camera data. This research shows that communities have access to diverse data sources that they could use to enhance situational awareness.
- We developed and tested source-specific data collection and processing workflows to infer road conditions from existing data sources. For example, we used deep learning computer vision models to automatically identify flooded roads from traffic camera data, thus overcoming the need to manually process data from 700+ traffic cameras updated every 5-10 mins. Similarly, we developed a natural language processing workflow to extract flooded entities from tweets and leveraged spatial analysis to infer road conditions from co-located water level sensors. This project showed that cities could repurpose their existing data sources to improve situational awareness and enhance community resilience. Moreover, while significant additional testing might be required, many of the data processing workflows developed in this study may be transferrable to other regions.
- Finally, we characterized data sources using historical observations from Hurricane Harvey (2017) in Houston. Our study indicates that the characteristics of data sources should be considered before leveraging them for decision-making. For example, data from crowdsourcing (Fig 2) significantly underrepresented socially vulnerable regions in Houston.

Consequently, the inequity in situational awareness data generation could translate to inequitable emergency response and recovery resource allocation, thus perpetuating a cycle of unjust recovery assistance.

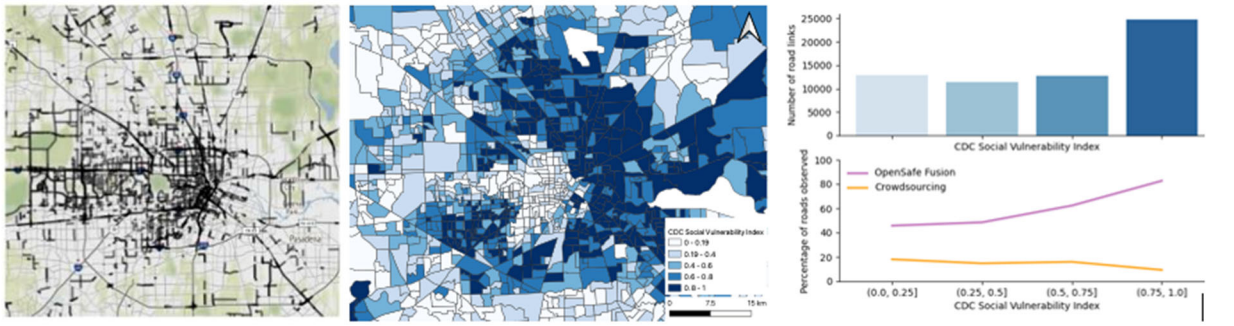

Figure 2. Bias in situational awareness data availability during Hurricane Harvey in Houston. Even though socially vulnerable neighborhoods were flooded, limited situational awareness data were available in these regions.

In Thrust 3, we focused on developing the methodological underpinning of a data fusion framework that can fuse real-time observations from diverse data sources to provide robust, reliable, and equitable situational awareness data on roadway flood impacts.

- We developed and validated a new situational awareness framework for sensing real-time roadway conditions during storms. The open-source situational awareness framework for mobility using data fusion (OpenSafe Fusion) (Fig. 3) is a modular framework composed of five components: data acquisition and processing, data fusion, data augmentation, impact assessment, and communication. During the data acquisition step (Fig.3a), real-time data from select sources are acquired, processed to infer road conditions, and geolocated. During the data fusion step (Fig.3b), observations for each road link are used to infer the road conditions by explicitly accounting for the characteristics of the data sources. Similarly, during the optional data augmentation step (Fig.3c), observed roadway status in the current time step is used to infer the state of roads for which direct observations are unavailable. Next, the impact assessment step (Fig.3d) estimates the network-level impacts of roadway flooding on access to select facilities. Finally, the results are communicated to stakeholders using a web dashboard (Fig.3e) and REST API (Fig.3f).
- A case study application of the framework in Houston showcases that fusing observations from existing data sources is an effective way to overcome inequity in situational awareness data availability. For example, Fig. 4 contrasts the road condition data availability of different sources (items a-g) with OpenSafe Fusion (item h) during Hurricane Harvey in Houston. From Fig. 4, it is clear that adopting OpenSafe Fusion can significantly improve data availability. While extensive additional testing is required, our initial case study analysis in Houston indicates that OpenSafe Fusion provides reliable situational awareness data for informing emergency response decision-making. Moreover, OpenSafe Fusion can partially overcome systemic biases in data availability, such as the data sparsity in socially vulnerable regions observed in sources such as crowdsourcing (item c).

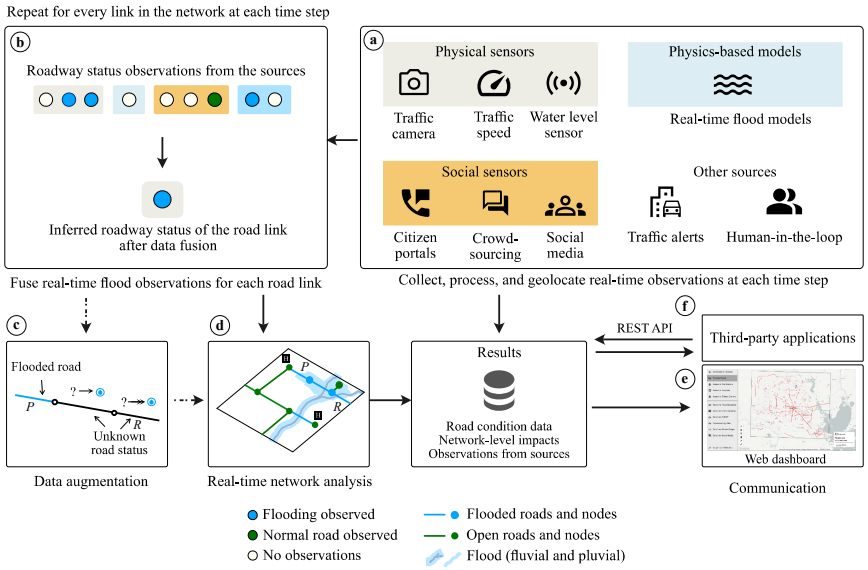

Figure 3. Overview of the OpenSafe Fusion methodology

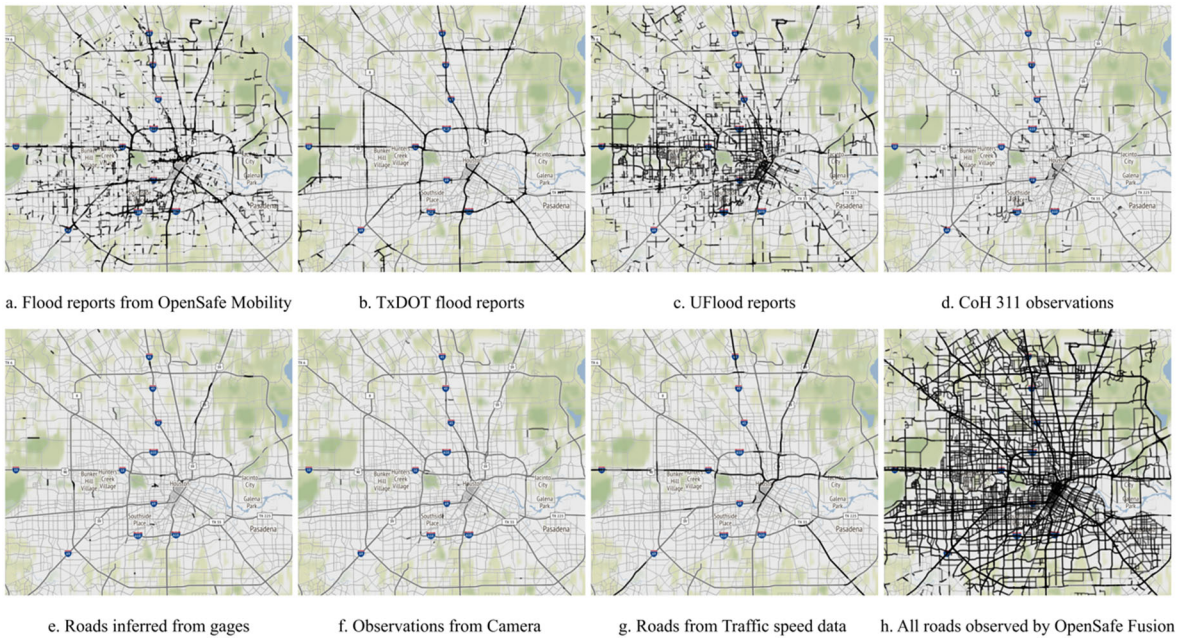

Figure 4. Spatial distribution of data availability during Hurricane Harvey. The OpenSafe Fusion framework proposed in this study can improve data availability as well as partially overcome inequities in situational awareness data generation.

- This thrust addressed the methodological challenges in fusing observations from disparate sources while accounting for data type heterogeneity, spatial and temporal resolution mismatch, and time lag. Moreover, the proposed method shows promising results in providing equitable and reliable situational awareness data leveraging existing data sources available in communities. Flood-prone regions could leverage the workflows and the open-source framework developed in this project to enhance their ability to sense and respond to flood events---a meaningful contribution in the epoch of climate-exacerbated hazards.

In Thrust 4, we aimed to inform the development of the proposed equity-aware flood alert system through user-centered design and community-engaged research.

- We adapted the user-centered design philosophy to identify stakeholder needs through extensive informant interviews. Through one-on-one informal interviews with Houston Health Department personnel involved in flood response, we gleaned insights on situational awareness data requirements and unearthed perceived imbalances in data availability and accuracy.
- The insights gained during the informal interviews informed all research thrusts. For example, the insights gained during the interviews were used to design OpenSafe Fusion and the accompanying web tool. Figure 5 shows the components of the web tool designed considering the extensive stakeholder input. We are currently working on scenario immersion studies with the developed pilot tool to evaluate model performance and perform rigorous usability studies before a potential deployment.

In summary, we achieved key research objectives identified in the project proposal.

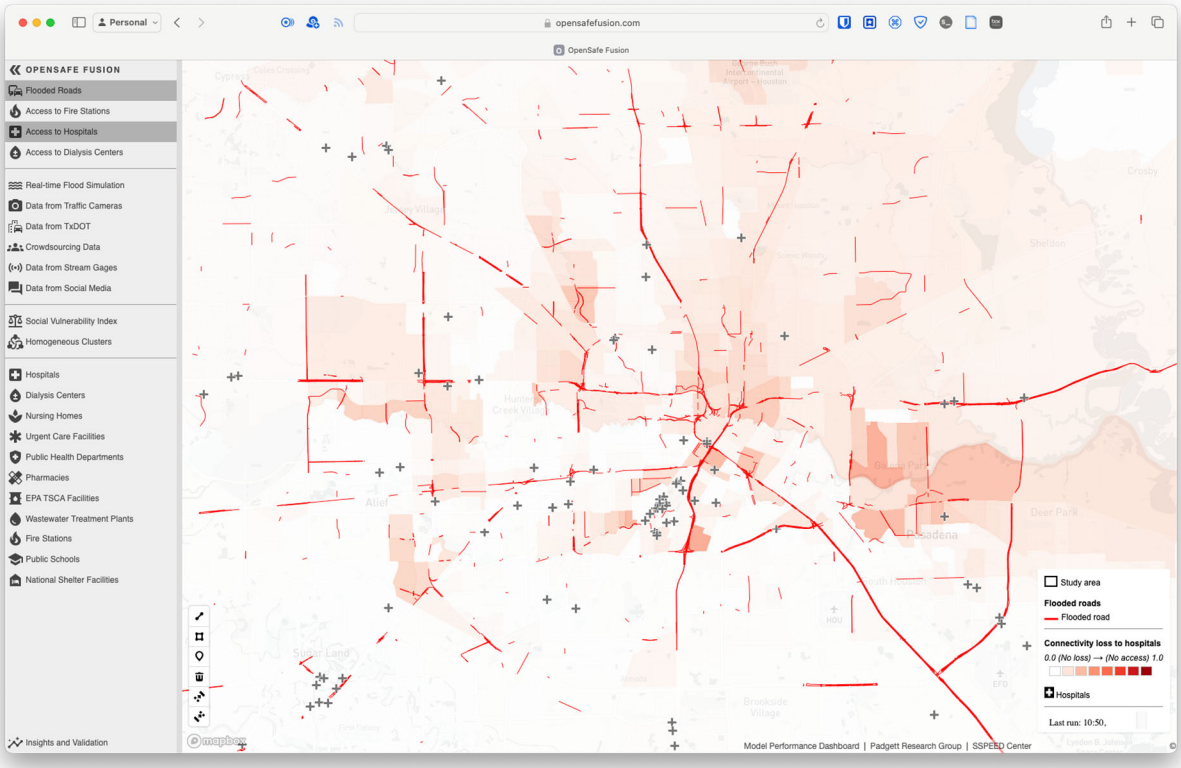

Figure 5. The OpenSafe Fusion web-tool. The tool provides real-time information on flooded roads and flood impacts on access to critical facilities such as hospitals. The results shown here are for a select time during Hurricane Harvey in Houston.