Transportation Research Part C 165 (2024) 104753

Contents lists available at ScienceDirect

Transportation Research Part C

journal homepage: www.elsevier.com/locate/trc

Predicting hurricane evacuation behavior synthesizing data from travel surveys and social media

Tanmoy Bhowmik [a,*], Naveen Eluru [b], Samiul Hasan [c], Aron Culotta [d], Kamol Chandra Roy [e]

[a] Department of Civil and Environmental Engineering, Portland State University, Portland, OR, United States
[b] Department of Civil, Environmental & Construction Engineering, University of Central Florida, Orlando, FL, United States
[c] Department of Civil, Environmental, and Construction Engineering, University of Central Florida, Orlando, FL, United States
[d] School of Science & Engineering, Tulane University, New Orleans, LA, United States
[e] Infrastructure Analytics Unit, CPCS Transcom Inc., 1000 Potomac St NW #500, Washington, DC 20007, United States

A R T I C L E I N F O

Keywords:
Hurricane
Evacuation
Twitter
NHTS
Data fusion
Probabilistic matching

A B S T R A C T

Evacuation behavior models estimated using post-disaster surveys are not adequate to predict real-time dynamic population response as a hurricane unfolds. With the emergence of ubiquitous technology and devices in recent times, social media data with its higher spatio-temporal coverage has become a potential alternative for understanding evacuation behaviour during hurricanes. However, these data are often associated with selection bias and population representativeness issues. To that extent, the current study proposes a novel data fusion algorithm to combine heterogeneous data sources from transportation systems and social media, in a unified framework to understand and predict real-time population response during hurricanes. Specifically, Twitter data of 2300 users are collected for evacuation response during Hurricane Irma and augmented behaviourally (probabilistically) with a representative National Household Travel Survey (NHTS) data, thus creating a hybrid dataset to improve the representativeness as well as provide a rich set of explanatory variables for understanding the evacuation behavior. The fusion process is conducted using a probabilistic matching method based on a set of common attributes across NHTS and Twitter. The fused dataset is employed to estimate the evacuation model and a comparison exercise is conducted to evaluate the performance of the model via fusion. The model fitness measures clearly demonstrate the improvement in data fit for the evacuation model through the proposed fusion algorithm. Further, we conduct a prediction assessment to illustrate the applicability of the proposed fusion technique and the results clearly highlight the improvement in the evacuation prediction rate achieved through the fused models. The proposed data-driven methods will enhance our ability to predict time-dependent evacuation demand for better hurricane response operations such as targeted warning dissemination and improved evacuation traffic management, allowing emergency plans to be more adaptive.

* Corresponding author.
E-mail addresses: tbhowmik@pdx.edu (T. Bhowmik), naveen.eluru@ucf.edu (N. Eluru), samiul.hasan@ucf.edu (S. Hasan), aculotta@tulane.edu (A. Culotta), roy.kamol@knights.ucf.edu (K.C. Roy).

https://doi.org/10.1016/j.trc.2024.104753
Received 15 February 2023; Received in revised form 3 July 2024; Accepted 4 July 2024
Available online 7 July 2024

1. Background

In recent years, Hurricanes Harvey, Irma, Maria, Michael, and Florence have disrupted the lives of millions of people in the United States (Hasan and Foliente, 2015). For instance, Hurricane Irma, one of the deadliest hurricane events in recent times, tore through Florida resulting in 87 fatalities (7 direct and 80 indirect) and $50 billion in economic damages (Kanner and Pintalgua, 2018). An important mechanism to avoid catastrophic losses due to a hurricane requires an effective evacuation plan and its execution. However, planning a hurricane evacuation requires the knowledge of many interdependent processes including hurricane dynamics, traffic flows in transportation networks, emergency services, and population response to warnings (Dash and Gladwin, 2007; Gladwin et al., 2001; Murray-Tuite and Wolshon, 2013). Researchers have studied population response by mainly modeling household decisions on: whether or not to evacuate (Gladwin et al., 2001; Hasan, 2010; Kang et al., 2007), time of evacuation (Hasan et al., 2013; Lindell, 2008), and choice of destination and routes (Mesa-Arango et al., 2013; Sadri et al., 2015). Earlier research has indicated that these decisions depend on many unobserved factors including a household's risk perception, hurricane awareness, reliance on information sources, and socio-economic characteristics. However, evacuation decision making is a complex, dynamic process that also depends on various time-dependent factors such as hurricane strength and projected trajectory, state of the infrastructure (such as traffic congestion), evacuation warning from local agencies, and traffic prediction (Davidson et al., 2020). Thus, modeling evacuation decision processes must consider dynamics and situational contexts related to the process.

Most of our current knowledge of hurricane evacuation behavior comes from post-disaster surveys. Specifically, researchers have developed behavioral models based on post-disaster surveys to predict household evacuation decisions (see (Lindell, 2008; Martín et al., 2017; Sarwar et al., 2018) for details). These frameworks were used to model a household's decision-making process in the presence of an extreme event i.e., they provide an understanding of reported behavior due to the event. For instance, current approaches to estimate hurricane evacuation demand rely on travel demand models estimating origin to destination flows. While such an understanding is quite useful, it is still inadequate to predict population behavior as a hurricane unfolds in real time. Furthermore, studies based on post-event surveys have limited transferability across different hurricane contexts and regions (Hasan et al., 2012; Martín et al., 2017). Further, behavioral models estimated over post-disaster surveys may not be adequate to predict dynamic population response as a disaster happens in real-time. Real-time prediction of evacuation behavior is very critical in disaster events. For example, weather agencies warned about Hurricane Irma well before its landfall providing enough time for Florida residents to decide whether to evacuate or not. In addition, close to landfall, several counties in Florida issued mandatory evacuation orders. Under such scenarios, typical demand estimates will not accurately capture the traffic due to the proactive decisions made by residents. At the same time, transportation agencies cannot assume full compliance from a region under a mandatory evacuation order. Many residents may not have access to information, may be unable to evacuate, or evacuate very late or may simply decide not to follow the evacuation orders. Thus, to prepare and respond in a timely manner, agencies should monitor and predict population response in real-time.

To that extent, social media data has attracted significant interests in recent times for modeling human behavior (Beiró et al., 2016; Duan et al., 2016; Roy et al., 2019). With better insights from real-time data, disaster management tools/practices can make better and quicker decisions. Advanced data science methods, predicting trends and patterns, have the potential to enhance evacuation decision support systems. While at first it may seem that online social media data would not be available before and during a crisis event due to infrastructure damage/disruptions, numerous prior studies (Guan et al., 2016; Martín et al., 2017) provide evidence that social media platforms are a vital source of information during disasters, particularly in the US. Researchers have been using social media data for analyzing disaster response from various perspectives (Guan et al., 2016; Kumar and Ukkusuri, 2018; Lazer et al., 2009; Martín et al., 2017). For instance, using Twitter and New York City subway and taxicab data, dynamics of population response has been analyzed during Hurricane Sandy (Guan et al., 2016). A recent study by Martin and his colleagues (Martín et al., 2017) has estimated evacuation compliance behavior during Hurricane Matthew. However, social media data analysis often faces several issues such as limited sample size, representativeness, and geo-tagged information (Li et al., 2022; Olteanu et al., 2019; Roy et al., 2021; Roy and Hasan, 2021). For instance, Twitter's public API does not provide all the tweets posted in the platform; purchasing the data can provide a larger sample size. In addition, errors might be introduced in the aggregate estimates due to the lack of sample representativeness and the selection biases present in the data. Individuals across different age and ethnicity categories are less likely to use the same social media platforms and maintain similar levels of activities. Thus, drawing data from social media is likely to result in a sample that potentially is biased. This is particularly problematic in understanding population behavior as we do not want emergency management decisions to be made that are not reflective of the diverse individuals involved in an unfolding crisis. To draw inferences on aggregate impact of individual decisions, a representative population is essential.

In our current study, we address the potential bias in social media data by combining the data with a representative dataset. Specifically, we propose an innovative information fusion approach to relate social media users to representative survey respondents. The fusion analysis is conducted based on a set of common attributes and a probabilistic matching method. This unique approach will generate a fused dataset allowing us to tag social media users with appropriate weights (from representative dataset) that can be employed to estimate population-level demand from social media platforms. The process begins with acquiring the real time evacuation decisions from a social media platform. The social media data is augmented with a nationally representative survey data. Each representative survey record is tagged to multiple social media user based on their attribute similarity. By matching social media users to respondents in national surveys, we can accomplish two goals: (1) reweight the social media sample to be more representative; and (2) augment the user attributes inferred from social media data with the attributes from the survey dataset. Thus, the process will improve data representation as well as provide a rich set of explanatory variables for understanding evacuation decision process.

For the current analysis, we collect Twitter data during Hurricane Irma for the evacuation response and employ the national household travel survey (NHTS) data administered by Federal Highway Administration (FHWA) for the representative dataset. Using our proposed fusion algorithm, we merge the NHTS, and Twitter datasets based on behavioural similarity to create a hybrid fused dataset that provides real time evacuation information as well as outcomes that are representative of the real population.

The rest of the paper is organized as follows: Section 2 provides an overview of data fusion approaches in transportation while describing the proposed research approach. The modeling framework of the current study is outlined in Section 3. The experimental setup of the fusion algorithm is summarized in Section 4. Section 5 describes the model findings (results and prediction assessment) and finally, concluding thoughts are presented in Section 6.

2. Research approach

Data fusion has been widely researched and employed in various fields including transportation, statistics, business analysis, chemical engineering, and navigation industry (Data Fusion, 2022). There is a large body of literature devoted to applying data fusion techniques within different sectors of transportation field including travel demand (Huang et al., 2018; Iqbal et al., 2014; Montero et al., 2019; Palubinskas and Runge, 2008; Pan et al., 2013; Wang and Chen, 2018; Wu et al., 2015), mobility pattern analysis (Yang et al., 2020), ride sourcing (Bi and Ye, 2021), freight movement modeling (Liao, 2010; Momtaz et al., 2020; Zhao et al., 2020) and safety (Yang et al., 2021; Yasmin et al., 2015). These research efforts focus on improving the quality of the data available for analysis.

Recently, in the travel demand modeling sector, mobile phone data is combined with other data sources including transportation data (such as global positioning system, GPS) (Huang et al., 2018), traffic count data (Iqbal et al., 2014), traffic sensor data (Wu et al., 2015) or sighting data (Wang and Chen, 2018) to estimate the origin–destination matrices more effectively. Pan et al., 2013 linked social media data with GPS data to create a complete picture of the real time traffic anomalies to offer alternative route guidance to drivers during congested time periods. Palubinskas et al. (Palubinskas and Runge, 2008) focused on fusing remote sensing data with roadway geographic file to improve our understanding of traffic congestion. In Yang et al. (2020), the authors identified potential underreporting of non-major activities by survey respondents (as highlighted in Bohte and Maat, 2009). The authors fused the travel survey data to mobile phone data to achieve improved accuracy in predicting activity patterns. Bi and Ye, 2021 combined ride sourcing trip data with the point of interests (POIs) data through a data driven approach labeled Latent Dirichlet Allocation (LDA) model to examine the behaviour of the shared mobility user. Momtaz et al., 2020 fused two commonly used freight databases – Freight Analysis Framework (FAF) and Transearch (TS) data – to realize transportation network flows at a fine spatial resolution (county level). Yasmin et al., 2015 adopted a data fusion algorithm in crash safety literature merging two crash databases- Fatality Analysis Reporting System (FARS) and Generalized Estimates System (GES) -to undertake injury severity analysis with a very refined characterization of fatality along with other injury severity levels.

Greaves (2006), a study of particular relevance for our work, proposes a method to simulate travel survey data by blending local sociodemographic information from sources such as a census with probability distributions of activity-travel patterns from other surveys. For example, if a specific percentage of individuals in a certain age group and employment category are identified to engage in work-related trips using public transportation, these patterns are assigned to matching characteristics in the simulated data. This technique creates a synthetic sample of households and individuals to represent the broader population, offering a cost-effective alternative to traditional travel surveys. The discussion above review clearly highlights the prevalence of data fusion algorithm across various transportation sectors.

The current study is geared towards proposing a behavioral fusion algorithm to combine two different datasets without any common identifier.[1] In this section, we will summarize the data fusion technique that is being adopted in the analysis and highlight the contributions of the study. The current approach is focused on a data fusion algorithm that augments NHTS data (source dataset) with social media data (Twitter, donor dataset) with a focus on developing an enriched dataset for evacuation prediction model development. The main motivation behind our matching approach is that NHTS data contains detailed socio-demographics, daily travel patterns and/or time use information and location information. Thus, if we can find Twitter users that match (defined by matching variables deterministically or probabilistically) with these survey respondents, we can create a representative data sample. The fused data set will carry the weights (from NHTS) that can be used to generate population-level estimates from twitter data by correcting for the selection bias present in twitter data. Furthermore, the fused dataset will have additional variables that can enrich the developed models. Thus, by merging real-time Twitter data with survey data, we can build a model to estimate evacuation responses for Twitter users with population weights and detailed characteristics available from survey data. This will improve the quality of the evacuation prediction model.

Towards implementing this fusion, we will collect a large number of social media users from the target location (the region affected by Hurricane Irma). Then, we will match each user to the NHTS survey respondent. Please note that the objective of the fusion is not to identify the social media users who participated in the NHTS survey. The idea is to find users who are similar (as determined by their attributes including demographics, and location) to a survey respondent. To compute similarity, a number of attributes have been inferred for Twitter users (demographic and behavioral), which are probabilistically matched with the survey respondents. Since twitter user attribute determination may be noisy and only partially observable, we will develop data fusion algorithms that match individuals across these datasets probabilistically.

In our current analysis, we find several common attributes between Twitter and NHTS datasets including age, gender, ethnicity, income, educational attainment, and presence of a child in the house. Initially we attempted to undertake the fusion using all common

[1] The fusion process is recently developed in Bhowmik et al., 2023 to augment two distinct datasets: Residential Energy Consumption Survey (RECS) and NHTS to improve the quality of the energy model.

attributes. When attempting to match the two datasets based on all six variables, a match was not found for approximately 93 % of the records. Hence, we employ an approach where we choose a subset of common attributes for matching. The fusion process starts with the selection of a sample of these common matching attributes to identify records from the twitter dataset that are likely candidates for fusion with the NHTS user. For example, let us consider two attributes: age and gender for matching. We augment each NHTS record with multiple users from the twitter data based on similar age and gender. The fusion will result in the addition of evacuation data column from Twitter data and several variables from NHTS in the fused dataset. Now, it is quite possible that that several such matches exist| (say K) and so in this case, the same NHTS user will be duplicated K times. The rationale behind this step is that these multiple social media records will collectively represent an NHTS individual. Of course, to accommodate for the duplication, we add an additional weight in the fused data ensuring all the K records represent only 1 record from the NHTS data. Now, this weight can be estimated following two approaches: deterministic and probabilistic. In the deterministic approach, all attributes are weighted equally, and all the matching records will have an equal weight (1/K, for K repetitions). This approach will work efficiently with a small set of matching variables. As the number of matching variables increases, the number of potential matches could reduce very quickly. Therefore, we propose a probabilistic approach for estimating the weight. In the probabilistic approach, we relax that all attributes have to be exactly matched thus allowing us to consider a larger set of matching variables. We require exclusive matching on a small number of attributes, and we estimate differential weights across records based on what factors are unmatched. Specifically, the weight is parameterized as a function of exogenous variables that is present in both datasets but not being used for fusion. This process translates to estimating the actual value of weight to improve data fit of the evacuation choice model as opposed to fixing it to (1/K). The parameters estimated will inform us about the ranking of the various matching factors on their impact on evacuation modeling. The reader would note that the exact number of replications (K) will be tested for different values based on evacuation prediction model performance.

To illustrate the data fusion process, an example is presented in Fig. 1. The NHTS survey data provides information for two persons with their age, gender, household size (number of people in their households), household renting status (own/rent), vehicle ownership and their race/ethnicity. The Social media data compiles information on age, gender, race and evacuation information (Evacuate: Yes/No). Now, let's assume that we matched twitter user to the NHTS respondents based on their age and gender. Based on this, comparing the two datasets, we find 2 matches from Twitter data for Person 1 in the NHTS database. Now, using the matched records, a fused dataset is created with two repetitions of Person 1 with NHTS data with the twitter data columns for Evacuation status and ethnicity (from twitter). Here, additional variables available in the survey data (e.g., car ownership) and missing from the social media data are also transferred to the fused data, thus potentially providing several additional independent variables for model development. In the deterministic weight approach, we will use a weight value of 0.5 (1/2) for the two repetitions to ensure that the fused data matches with the survey data weights. For the probabilistic weight method, we first identify the common attributes across these two datasets which are not used for fusion. For the example presented in Fig. 1, a common attribute race is present in both NHTS and Twitter database but is not used for fusion. We will compute the similarities in race across the two datasets (if matched then 1 otherwise 0) and use this similarity to parameterize the weight function. The hypothesis behind the probabilistic weight variable is that matched records for variables not considered for matching will result in higher weights. This is represented in the Fig. 1 as a higher weight for first matched record. Please note that the numbers provided in Fig. 1 are for illustration purposes and will actually be estimated on the data in our model.

In summary, the current analysis makes a threefold contribution to the evacuation choice literature. First, we are leveraging large-scale real-time data to predict evacuation demand. Second, we propose a novel fusion algorithm by combining heterogeneous data sources from transportation systems and social media, in a unified framework providing better information for modeling dynamic population behavior during hurricanes. Third, instead of strictly matching records, we propose a flexible differential weighting method

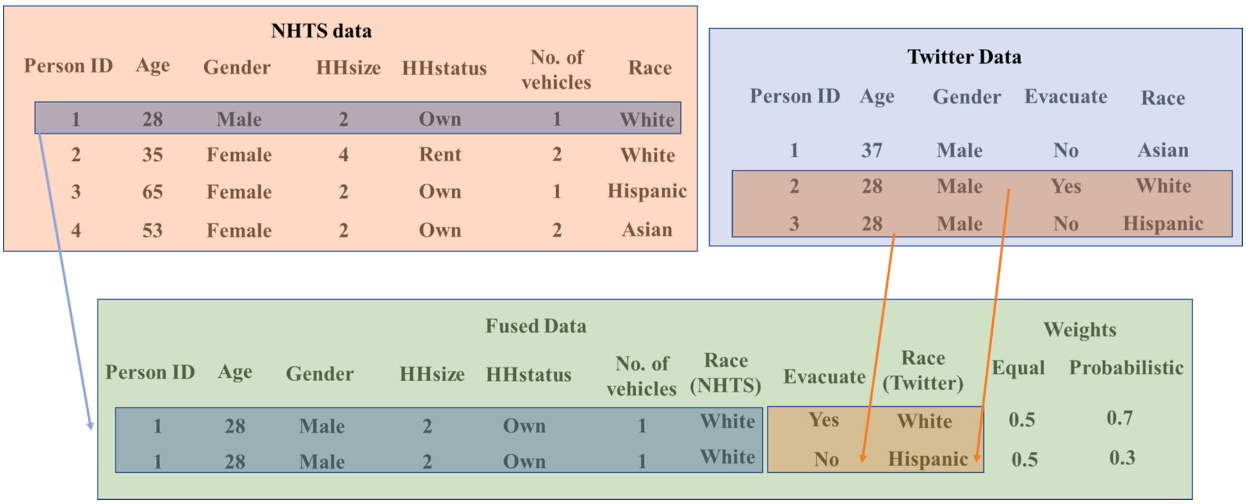


Fig. 1. NHTS and Twitter Data Fusion Illustration.

(probabilistic) based on attribute similarity (or dissimilarity) across the common attributes for the two datasets. Finally, extensive testing is conducted with multiple variables, fusion size and weight functions to evaluate the efficacy of the proposed fusion technique.

3. Methodology

In this section, we present the methodological framework adopted in the study for analyzing individual's evacuation decision during Hurricane Irma. The model structure has a decision model component (evacuate-yes/no) and a weight component. In the decision model component, a binary logit formulation for the evacuation variable is considered as we have only two alternatives in the choice dimension. A brief description of the proposed econometric fusion approach is given below.

Let us assume that there are i (1,2, …, N, N = 2,246) individuals in the NHTS survey data and K possible matches from the Twitter dataset. With this notation, the evacuation propensity component takes the following form:

$$v_{ik}^{*} = \beta' X_{ik} + \gamma' S_{ik} + \varepsilon_{ik}, v_{ik} = 1 if v_{ik}^{*} > 0; v_{ik} = 0 otherwise \tag{1}$$

where v_{ik}^{*} represents the propensity of the individual i for the k^{th} fused record to evacuate during Hurricane Irma. v_{ik} is the observed evacuation choice, that is 1 if the person i in NHTS for the k^{th} fused record from Twitter had evacuated during the disaster and 0 otherwise. X_{ik} is a vector of attributes from the NHTS dataset that influence the evacuation choice and β' is the corresponding coefficients to be estimated (including a scalar constant). S_{ik} is the vector of attributes from the Twitter dataset that affect the evacuation decision and γ' is the corresponding vector of coefficients to be estimated. ε_{ik} is an idiosyncratic error term assumed to be identically and independently standard logistic distributed. Based on this, the probability for person i for the k^{th} fused records to evacuate is given by:

$$P_{ik} = \frac{\exp\left(v_{ik}^{*}\right)}{1 + \exp\left(v_{ik}^{*}\right)} \tag{2}$$

The corresponding probability for non-evacuation is computed as:

$$Q_{ik} = 1 - P_{ik} \tag{3}$$

Thus the choice probability expression for the binary logit model can be expressed as

$$C_{ik} = (P_{ik})^{v_{ik}} * (Q_{ik})^{1-v_{ik}} \tag{4}$$

The weight component takes the form of a latent multinomial logit structure allocating the probability for each Twitter user being paired with the NHTS respondent. The matched weightage propensity is determined based on a latent probability value estimated using a multinomial logit model as follows:

$$w_{ik} = \frac{\exp(\alpha Z_{ik})}{\sum_{k=1}^{K} \exp(\alpha Z_{ik})} \tag{5}$$

where Z_{ik} is a column vector of attributes for individual i and fused record k that influences the propensity of matching the Twitter data with the NHTS data. To be specific, Z_{ik} represents the variables that are present in both datasets but not used for fusion. α is the corresponding vector coefficients to be estimated. Based on this notation, the overall weighted probability can be written as:

$$L_i = \sum_{k=1}^{K} w_{ik} x C_{ik} \tag{6}$$

where L_i is the weighted probability the person i in the NHTS dataset has for the corresponding evacuation choice C_{ik}. This matching, when executed, will provide us a relationship between the NHTS and Twitter datasets. Specifically, employing Equation (6), several additional variables from the NHTS dataset will be employed to develop an improved evacuation choice model. Finally, the log-likelihood function for the fused dataset is defined as:

$$LL = \sum_{i=1}^{N} \log(L_i) * wt_i \tag{7}$$

where wt_i represents the person weight in the NHTS data to represent the population. The proposed matching algorithm has been estimated using a maximum likelihood based econometric model. We have used the GAUSS Matrix Programming software for estimating the models.

4. Experimental setup

Prior to developing the fusion algorithm, we need to consider several aspects that affect the fusion process: 1) among the common variables present in both dataset, which variable/variable groups will be used to link the two dataset and which variables will be used in the weight component; 2) how many records (number of Twitter users) will be matched with each NHTS respondent; and 3) how will we assess the impact of randomness of fusion process on parameter stability.[2] In this section, we illustrate the experimental setup documenting the structure of how the fusion process will be tested (see Fig. 2).

As discussed earlier, both datasets have several common attributes such as: age, gender, race, education, income and presence of a child in the household. We can use either a single variable (like matching only age) or variable group with multiple variables (match based on similar age and gender) for fusion. We start the fusion process with one combination, that is select one variable/variable group to match the NHTS and Twitter dataset. The variables not considered for matching are tested in the weight function to allow for probabilistic weighting described in Section 2. Please note that this matching might result in multiple records form the Twitter dataset to be linked with each person in the NHTS. However, estimating models with all these fused records increases the computational burden. Hence, we select a fixed number of records to be fused (say $K = 5$) and repeat the sampling process several times (say $N = 10$, so we fused 5 records randomly and generate 10 samples). Now, we estimate our evacuation model across these fused samples considering attributes from both NHTS and Twitter dataset and evaluate the average log-likelihood improvement (over N samples) compared to the evacuation model estimated on the Twitter dataset only. After selecting the optimal variable (X) with superior improvement, we continue along the flow chart to identify the optimal number of records (K) to be matched between the NHTS and Twitter dataset. Specifically, we test at what K (3, 5, 10, 15, 20, 40, 50) do we obtain the highest improvement in average log-likelihood. Finally, in the last part of the fusion, we evaluate the consistency of our evacuation model for the fused dataset (fuse with X variable/variable group and K records) by comparing model parameters across N samples. If we find any variation across the samples, then it lends evidence to instability in parameter magnitudes and signs. Hence, we eliminate that variable/variable group from the fusion and proceed with other variables. The process is repeated for each potential matching variable (and variable combination) and the combination that offers the highest improvement while satisfying other criteria is selected as the final fusion variable, size and fused dataset.

4.1. Data description

During Hurricane Irma (Sept. 5 to Sept. 14, 2017), we collected around 1.81 million tweets (248,763 users) using Twitter's streaming API. To find user activities during a pre-hurricane period, we also collected historical data using Twitter's REST API for the users who were active for at least 3 days between the first evacuation order and landfall days. By examining tweets that contain geotags (latitude/longitude values), we inferred home and workplace locations based on each user's most visited places during office hours and nighttime, respectively. If a user tweeted more than 200 miles from their home during the evacuation period, we assume that the user has evacuated (see (Roy and Hasan, 2021) for details). We have also filtered out non-Florida users and international visitors. Our final dataset includes 2,357 Florida users with 552 users evacuated during hurricane Irma. We counted the number of Irma or evacuation related tweets posted by the evacuees against the time before they evacuated (Fig. 3a). This indicates that users' intents to evacuate can potentially be detected much earlier than their actual evacuation times. For example, one tweet stating "Me and my family evacuating Florida right after my class on Friday" was posted on a Tuesday night, September 5th, about 2.5 days prior to the user's inferred evacuation time (which indeed corresponded to Friday). Given that Irma did not make its landfall in Florida until Sunday (Sept. 10), such tweets provide ample lead time to aid forecasting estimates. Additionally, we identified the most likely evacuation and return times and the origins and destinations of the evacuated users (Fig. 3b). We found that Florida residents mainly evacuated to Georgia, Alabama, South Carolina, and North Carolina.

To estimate user demographics, we use a classifier from prior work (Culotta et al., 2016, 2015) which was trained to predict user demographics based on a sample of their tweets and users that they follow. The model is trained using distantly labeled data based on audience measurement data for 1,500 websites. While undoubtedly such a classifier is imperfect (and oversimplifies gender and race/ethnicity categories), prior results indicate that it exhibits high concordance both with labeled data and with population level characteristics (e.g., 0.73 average correlation with panels of matched web traffic demographics). We have also used it previously to study hurricane evacuation social media data (Li et al., 2022). Fig. 4a represent the summary statistics for the demographic variables for the Twitter users (orange bar).

The 2017 NHTS data in addition to the demographic variables documents detailed information on several important variables like home renting status, location (urban/rural), region (core based statistical region), no. of people, drivers, adults, and workers in the house that might influence the evacuation choice of a person during hurricanes. The NHTS dataset collects information for 264,234 individuals from 129,696 households sampled from all over the country. The corresponding person weights provided in the NHTS data were employed to generate population representative values for the individuals. As our analysis is focused on the evacuation behaviour during Hurricane Irma, we confined our attention to the State of Florida in the NHTS dataset. Within the state of Florida, the NHTS dataset provides information for 2,399 individuals. After processing and cleaning the data (several variables are missing for some respondents), the final NHTS sample consists of total 2,246 individuals representing around 14 million people in the state. A brief summary of the NHTS 2017 data sample employed in our analysis is presented Fig. 4a (blue bar) and 4b.

[2] The reader would note that different samples of fused records can potentially result in different parameters. To address this variation, we test the parameters from multiple samples to ensure that the results are stable across samples. The model parameters across these samples are compared using a modified Wald *t*-test to ensure that parameter variation is small across the samples i.e., there is no statistically significant difference in parameters across samples.

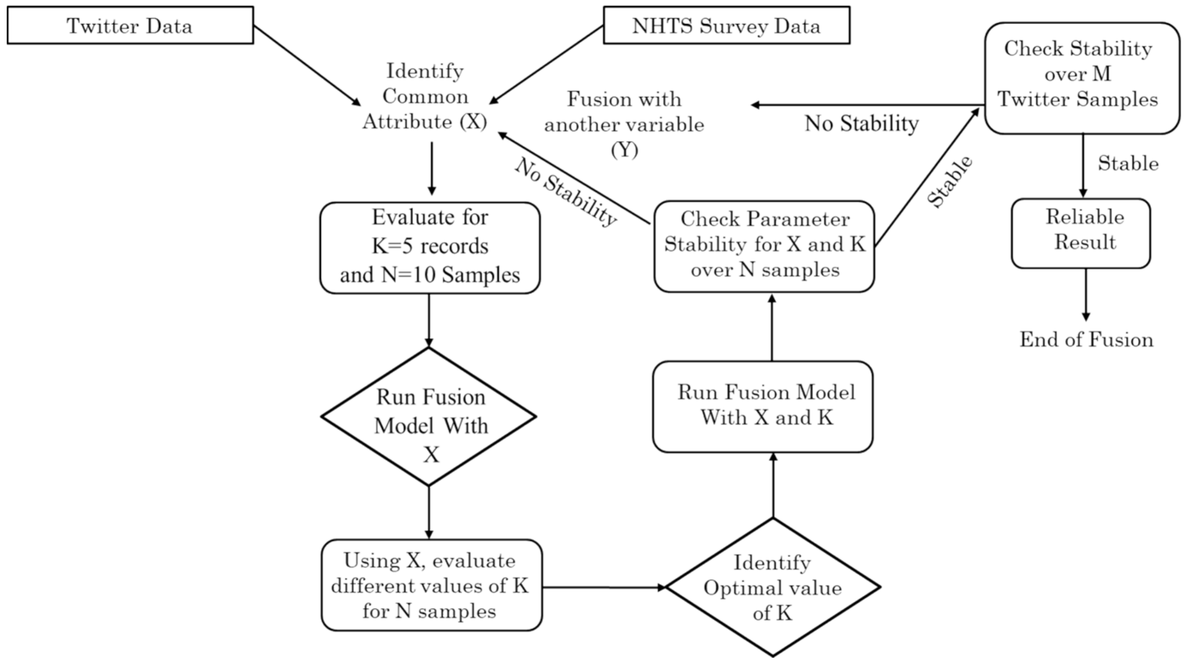
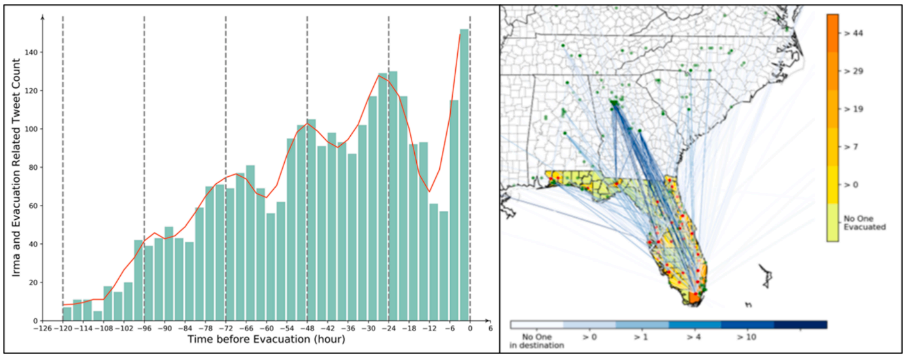


Fig. 2. Flow Chart Showing the Framework of the Fusion Algorithm.



a. No. of Irma and evacuation related post of evacuees before their evacuation

b. Origin Destination of Irma evacuees

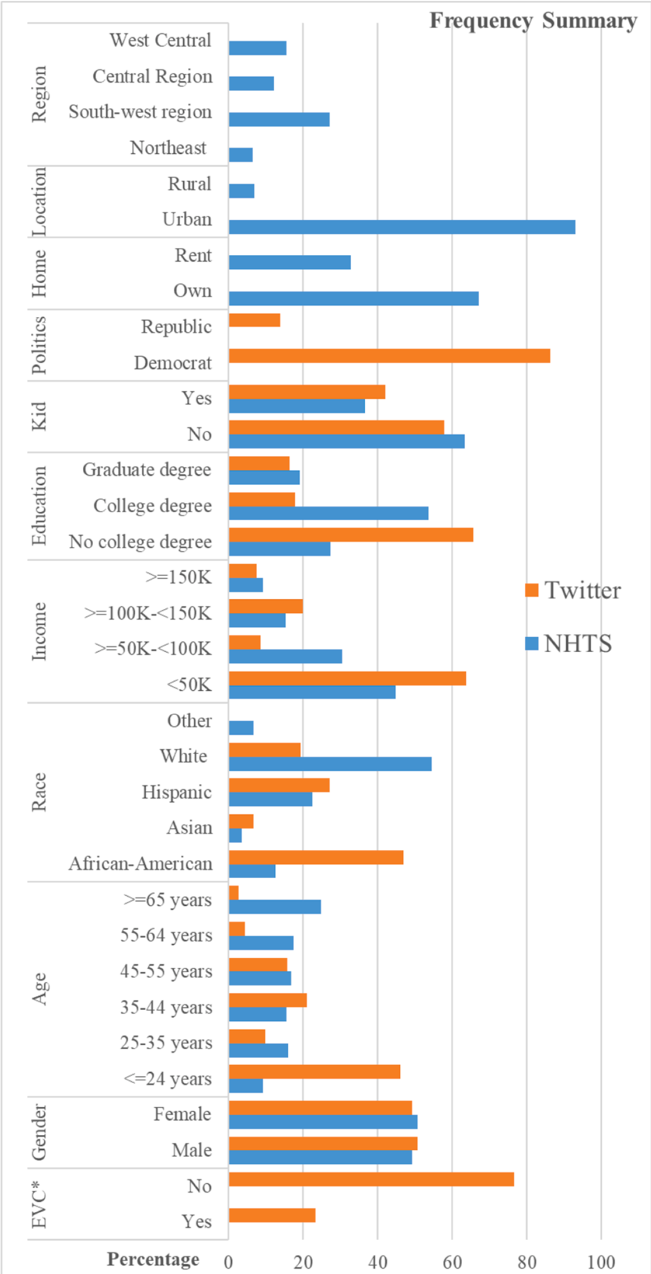
Fig. 3. Twitter Data Analysis.

4.2. Selecting variables for fusion

In the current analysis, we considered several combinations of variables (single and multiple variables) for linking the two datasets and within each combination, we estimated two models: 1) an evacuation model considering variables only from the Twitter dataset and 2) an evacuation model based on the fused dataset considering variables from both NHTS and Twitter. Then, for each combination, we calculate the improvement in average LL (we consider N = 10 samples) for the fused evacuation model (model employed on the fused dataset) relative to the Twitter only evacuation model (model employed on the Twitter dataset). Finally, the combination providing the superior improvement is selected. Fig. 5 represents the average LL improvement measures across different combinations of variables. The Fig. 5 clearly highlights the superior improvement in LL for the fused evacuation model when each respondent in NHTS is linked with multiple Twitter users based on matching age, race and income. Therefore, we select this variable group for fusing the two datasets and proceed to the next step.

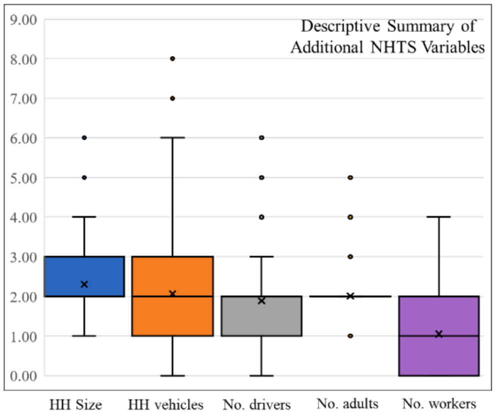
4.3. Selecting the number of matching records for fusion

Based on the results obtained in the first step, we merged the two datasets based on matching age, race and income and created N = 10 fused databases using multiple matching records of K including 3, 5, 10, 15, 20, 30, 40, and 50 (see Fig. 6). Similar to the variable



*EVC: Evacuation Decision (dependent variable)

a) Frequency Summary of NHTS and Twitter Dataset



b) Descriptive Summary of additional variables in NHTS Dataset

Fig. 4. Descriptive Summary of Twitter and NHTS Variables.

selection criteria for fusion, we again resort to the improvement in model fit measures (average LL improvement) for identifying the optimal number for fusion. However, here we also focused on one additional factor: the complexity of the model in terms of run times. For instance, if we matched 25 records instead of 5, that might give us better model fit but at the same time, the complexity of the model will increase as it will require higher run times for convergence. Obviously, there is a trade-off between these two factors and thus, we need to select an optimal value of the number of records for matching which will give us a relatively better model fit while being simple from the estimation perspective. From Fig. 6, we can clearly see the substantial improvement in LL as the number of fusions keep increasing, particularly up to 15. Interestingly, the improvement becomes stable or even worsens beyond 15 records per individual in NHTS. Also, the improvement is marginal (6.5 %) from 10 to 15 fusion while the computational run times is approximately 60 % higher for fusing 15 records relative to fusing 10 records. Therefore, we select K = 10 as the optimal value, meaning for

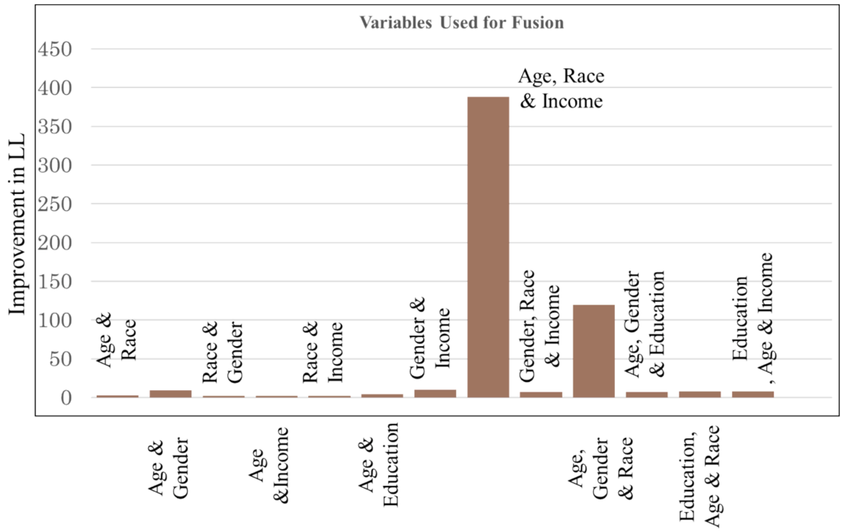


Fig. 5. Model Fit Summary Across Different Variable Group Used for Fusion.

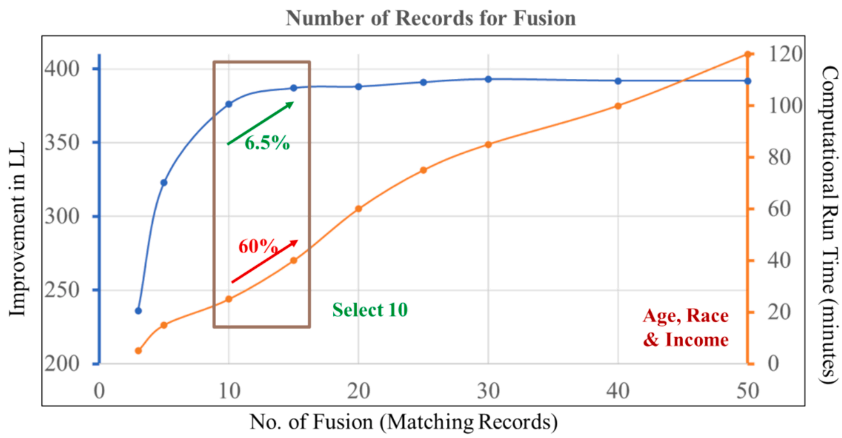


Fig. 6. Model Fit Summary Across Different Number of Fusion.

each respondent in NHTS, 10 records from the Twitter data will be fused based on similar age, race and income.

4.4. Parameter stability

After selecting the variables and the number of records to be used for fusion, the next step is to evaluate the stability of the parameters of the evacuation model estimated using the fused data.

For instance, let's say for the common age, race and income variable, we have 100 matching records in the Twitter dataset that can be fused with each NHTS respondent. Out of these 100 matching records, we consider 10 (as indicated by previous section) through a random sampling process and estimate our model based on this samples. However, it is quite possible that considering a different sample with different 10 matching records might alter the parameter estimates (particularly the new variables from NHTS and variables in the weight component). Therefore, we carried out our proposed fusion algorithm multiple times (10 to be specific) and test the hypothesis whether the parameter estimates obtained from the model exhibit any significant differences across the samples. If the hypothesis is rejected, it will indicate that the parameter estimates are highly dependent on the randomness in sampling. On the other hand, if the hypothesis is not rejected, it will inform us that the coefficients do not significantly differ across the samples, thus reinforcing the statistical validity of our proposed fusion approach.

To test the hypothesis, we first calculate the estimates (B_s) and its corresponding standard deviations (SD_s) specific to each fused variable and variables in the weight component across the 10 samples. Then, we calculate the mean of these estimates ($B_m = \frac{1}{N} * \sum_{N=1}^{10} B_s$) and standard deviations ($SD_m = \frac{1}{N} * \sum_{N=1}^{10} SD_s$). Then we simply calculate an approximate t-statistics specific for each variable across each sample as follows:

$$t_s = Abs(\frac{(B_m - B_s)}{\sqrt{SD_m^2 + SD_s^2}}) \tag{6}$$

This t-statistics value will highlight the statistical differences for the variables across the samples. If this value is above (below) 1.65 (critical t-stat at 90 % confidence level), then we will reject (not reject) the hypothesis that the estimates are similar across samples. For each variable, we will have 10 t-statistics (due to 10 samples). The comparison results, for the ease of presentation, are presented in a box plot (see Fig. 7). The boxplots clearly illustrate significant stability in the parameters estimated. In fact, the computed approximate t-statistic does not reach 1.65 for even one parameter across all samples. From the results, it is quite clear that we have obtained a reasonably stable estimate for all parameters in the fused dataset which further provides support to the statistical validity of our proposed fusion algorithm. The reader would also notice the significant number of variables added to the fused model from NHTS that would not have been available in Twitter dataset without fusion.

5. Empirical analysis

5.1. Model fit

We apply our data fusion algorithm proposed in the study to combine the two datasets (NHTS and Twitter) based on the common attribute present in both datasets. Specifically, for each respondent in NHTS, we fused 10 records from the Twitter dataset based on the same age, race and income. As we did not find 10 matches for all records, the final fused sample had 16,055 records matched with 2,246 NHTS users. Using this fused dataset of 2,246 users, the empirical analysis involved a series of model estimations. First, we develop a simple binary logit model employing the fused dataset while considering variables only from the Twitter dataset. Of course, we have to use a weight component to ensure all the 10 records represent only 1 record from the NHTS survey data. Here, we used a fixed equal weight (1/10) and thus named the model as equal weight simple binary logit model (EWSBL). This EWSBL model is used as a baseline to evaluate the performance of the other models. Second, we improve the EWSBL model by considering additional variables from NHTS data while retaining the equal weight assumption and labelling the second model as equal weight fused binary logit model (EWFBL). In the final step, we relax this equal weight assumption and allow the weight to vary across the fused record based on the similarity (dissimilarity) between the common variables present in both datasets that are not used for fusion. For example, let's assume out of the 10 fused records, 5 of the records have same educational attainment across the two datasets. The hypothesis behind our proposed latent weight approach is these 5 records should offer higher weights relative to the other 5 fused records in contributing to the evacuation decision model. Now, using this latent weight approach, we again estimate the binary logit framework and named it as Latent weight fused binary logit model (LWFBL). Please note that the different names are given to each model just for the sake of representation.

For evaluating the performance of the models, we employ the Bayesian Information Criterion (BIC) as it is the most widely used approach for comparing non-nested models.[3] The BIC (LL) values for the final specifications of the three models are: 1) EWSBL model (with 4 parameters) – 5961.36 (−2961.31); 2) EWFBL model (with 9 parameters) – 5,903.88 (−2908.36); and 3) LWFBL model (with 10 parameters) – 5242.82 (−2572.99). The comparison exercise highlights two important observation. First, models incorporating additional information from the NHTS dataset provides improved performance as indicated by the lower BIC value associated with the EWFBL model. The result is a clear indication of how new variables from the NHTS dataset helps in understanding the evacuation behaviour and trend. Second, the probabilistic weight framework (LWFBL) across the fused records clearly outperforms the equal weight framework (EWFBL) which provides evidence that fused record's contribution can be optimized using the weight function based on the similarity/dissimilarity of the common attributes. In summary, the model fit measures clearly demonstrates the improvement in data fit for the evacuation model through the proposed fusion algorithm and the weighted contribution estimation.

5.2. Estimation results

The model fit measures clearly indicate the superior performance of the LWFBL model in analyzing the evacuation behavior. In this section, we discuss the effects of variables on the evacuation decision obtained from the LWFBL model only. In estimating the model, we tested for several variables and functional forms and retain those that provides the best result in terms of data fit and interpretation. In the final specification, we remove all the insignificant variables in a systematic process based on statistical significance (90 % significance level). The estimation results are presented in Table 1. A positive (negative) sign in the Table indicates a higher (lower) evacuation likelihood for the corresponding individual. For ease of presentation, we discuss the results by variable groups.

5.2.1. Constant

The negative sign of the constant term suggests a lower preference for evacuation in the population during Hurricane Irma. The result follows the observed evacuation trend where one in four Floridians preferred not to evacuate during Hurricane Irma and the other three will only evacuate in case of a category 3 or stronger hurricane (FirstCoast News, 2022).

[3] In addition to BIC measures, we also conducted the log-likelihood ratio tests across each pair of models for further evaluating the model performance. To conserve space, we have detailed the results of these log-likelihood ratio tests in the Appendix (Section A.2 in the Appendix).

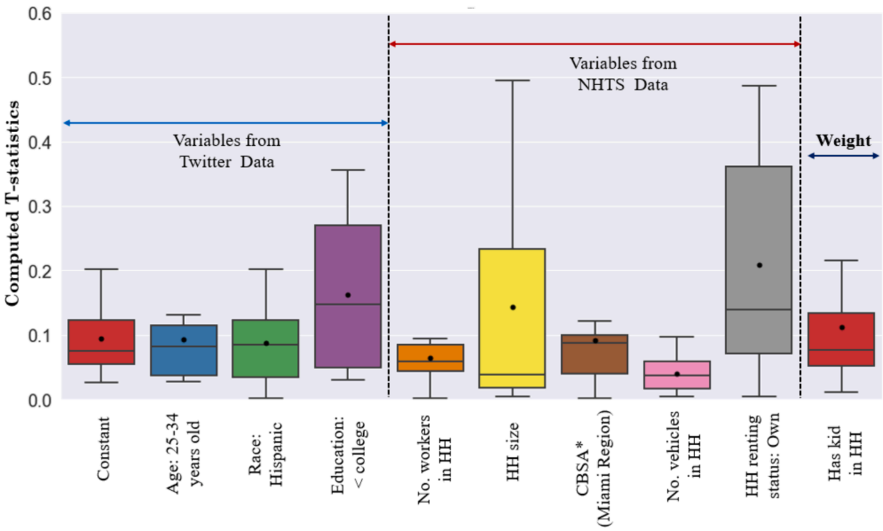


Fig. 7. Test Statistics (t-statistics) for Parameter Estimates Across Samples for each Variables and Models. *: CBSA 2 is respondent's core based statistical area code in the southwest region (Miami-Fort Lauderdale-West Palm Beach).

Table 1
Latent Weight Fused Binary Logit (LWFBL) Model Estimation Results.

Sample Size (2,246 NHTS user)		
Variables	Estimates	T-stat
Constant	−13.378	−7.716
Variables from Twitter		
Age between 25 and 34 years (base: others)	5.599	5.138
Race: Hispanic (base: others)	3.810	4.121
Education: no college degree (base: college degree/grad school)	2.613	3.108
Variables from NHTS		
HH size	−2.554	−4.374
No. workers in HH	−0.957	−2.312
No. vehicles in HH	2.298	5.324
Home owned (base: home rent)	−1.143	−1.798
Residing in south-west region (base: other location)	3.062	3.413
Weight Component		
Presence of Child matched in both data	2.249	18.956

5.2.2. *Twitter variables*

As discussed earlier, we generate several demographic variables including age, gender, race, education, political profile (democrat or not) and presence of a child in the HH for each Twitter user. Among these factors, we find three variables to influence the evacuation decision. Consistent with earlier literature (Martín et al., 2020), we find that individuals aged between 25 and 34 years are more likely to evacuate during hurricanes relative to the other age groups. The result is understandable since people from this age group are less constrained (not married, usually no child, or face health related issues). With respect to ethnicity, the estimation results indicate an increased evacuation propensity among Hispanic population perhaps because the hurricane path traversed through regions with a higher share of Hispanic population. Martin and his colleagues (Martín et al., 2020) also point to this higher evacuation likelihood among the Hispanic population during hurricanes. The final variable specific to the educational attainment reveals a significant role in influencing the evacuation decision. In particular, people with less than high school education are more likely to evacuate during hurricane compared to people with higher education. Given that education attainment is closely associated with household wealth, these populations are likely to be low income and live in housing that might be affected by hurricanes (Martín et al., 2020). We also tested for several hurricane related factors such as the distance of the hurricane from user's home location, hurricane's intensity and wind speed; however, none of the variables are found to significantly influence the evacuation behaviour.[4]

5.2.3. NHTS variables

We tested for several variables and the results clearly highlighted the significant impact of different variables from the NHTS data affecting people' evacuation decision. For instance, as indicated by Table 1, the number of people and the number of workers in the household offers a negative association with the evacuation decision suggesting a lower willingness of a person to evacuate with more people as well as more workers in the house. The results are expected as evacuating larger households is resource intensive (see (Dixon et al., 2017; Rambha et al., 2021) for similar results). The presence of multiple workers also reduces evacuation likelihood as these individuals are possibly concerned about time off and also are likely to be residing in safer homes (The Atlantic, 2017). Interestingly, number of vehicles in the household is positively associated with the evacuation decision highlighting that when everything else is same, individuals with access to vehicles are more likely to evacuate (see Yabe and Ukkusuri (Yabe and Ukkusuri, 2020) for a similar finding) Moreover, during hurricanes, vehicle access is quite important as rental companies usually run out of rental cars (Planetizen, 2005) and individual access to vehicles provides flexibility for a family to evacuate. From Table 1, we also find that homeowners are less likely to evacuate as evidenced by the negative coefficient. Home owners are potentially concerned about damage to the home and staying can allow them to fix damage faster (Hasan et al., 2011). Finally, people residing in the southwest region are more likely to evacuate as indicated by the positive coefficient in the Table. The result is intuitive as IRMA was directed at the southwest region.

5.2.4. Weight component

As the fusion is conducted based on matching age, ethnicity and income, the other three common variables across the datasets including gender, education and presence of a child in the house are the potential candidates in the weight component. After extensive testing, we find one variable, similarity in the presence of a child in the house to exert a positive influence on the weight contribution. The result indicates that within the fused records, records with similar information on the presence of a child in the house across both datasets will contribute more to the evacuation model relative to the other records. This significant effect in the weight component further reinforces our hypothesis on using a differential weighting method (probabilistic) as opposed to equal weight assumption.

5.3. Validation analysis

The model fit measures presented in Section 5.1 underscores the improvement in the evacuation model though the data fusion algorithm and the latent weight contribution. However, it is also essential to assess the real-world applicability of the fusion technique. This involves employing the fusion algorithm to predict the evacuation rate and comparing it the report evacuation rates. In the prediction analysis, we conduct the assessment focusing on the NHTS data, considering all three models: a Twitter-only model, model achieved through the fusion technique with equal weight assumption, and model achieved through the fusion technique with latent weight assumption.[5] The purpose of using only NHTS data is to utilize the national representative dataset to estimate the potential evacuation rates on a broader scale and compare these predictions with actual evacuation figures. Specifically, we predicted the percentage of people in the state of Florida who evacuated (using all three models) during Hurricane Irma and compared the results with the actual values (SRESP, 2021). To robustly assess the prediction performance, 50 data samples with 300 records (in NHTS) are employed to generate evacuation rate predictions for each sample. This sampling approach ensures that our model evaluations are not dependent on a single dataset and helps to capture the variability in predictions providing a comprehensive evaluation of model performance. Additionally, we included the performance of a trivial classifier with a 50 % probability of evacuation as a baseline for comparison. It helps to demonstrate whether our sophisticated models offer a significant improvement over a trivial classifier. The prediction results are presented in Fig. 8.

From the figure, three important observations can be made: First, the Twitter-only model significantly overpredicts the evacuation rate, supporting our hypothesis that the data is not representative of the overall population. Second, through the fusion technique, we observe a substantial improvement in the prediction rate, as indicated by the lower error rate in both variants of the fused model. The results clearly highlight how new variables from the NHTS dataset contribute to improvement in predicting evacuation behaviour. Finally, among the fused models, the latent weight model demonstrates slightly superior performance compared to the equal weight model. In summary, the validation results further provide evidence on the enhanced explanatory power and predictive capability achieved through the proposed fusion algorithm and latent weight contribution in the context of evacuation modeling.

[4] The reader would note that as a majority of the Twitter data were not geocoded, we did not employ inference algorithms for home (or work) location as they are likely to introduce potential stochasticity. In our analysis, we employed Core Based Statistical Areas (CBSA) information from the NHTS data as a surrogate for home location information.

[5] The reader will note that the analysis involves predicting evacuation rate for a single record with the NHTS dataset (as opposed to the ten record fusion exercise). Hence, we need to recalibrate the constant in all three models (Twitter-only, equal weight fusion, and latent weight fusion) to adjust from a 10-record fusion basis to a single-record basis. This recalibration ensures that our predictions accurately reflect the individual-level evacuation behavior.

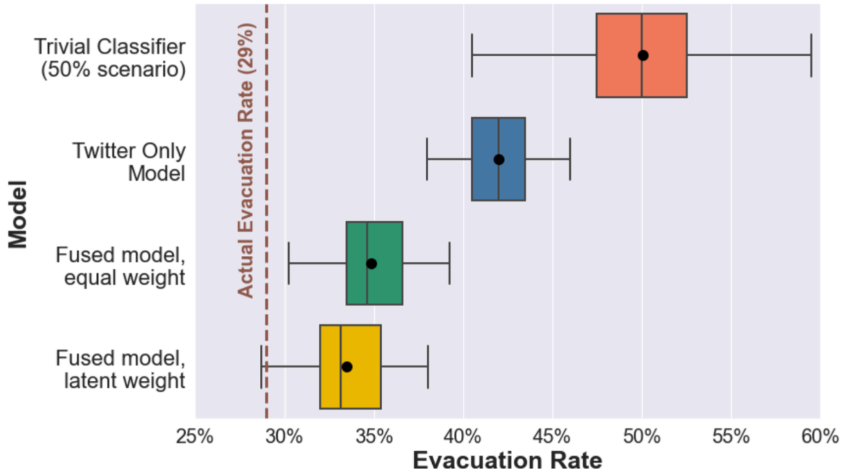


Fig. 8. Prediction Assessment for Hurricane Irma Evacuation Rate.

6. Conclusion

The proposed research is geared towards developing new data science methods to predict population behavior under a hurricane threat. Evacuation decision making is a complex, dynamic process that depends on various temporal and spatial factors like hurricane path, intensity, projected trajectory and prevailing traffic conditions. Modeling population evacuation behaviour using traditional data sources (like post-disaster survey) are not sensitive to such underlying contextual dynamics. Social media data, on the other hand offers tremendous potential for understanding hurricane response due to its real time data availability and higher spatio-temporal coverage. However, these data (social media) are prone to selection bias and representativeness issue and thus needs advanced filtering and fusion techniques to be reliably used for predicting large-scale population behavior, specially during disasters. To that extent, the current analysis is focused on developing an innovative fusion technique to connect the social media users with traditional national survey respondents, thus providing more complete representation of user preferences and expected behaviours during hurricanes. In the current study context, we collect large scale Twitter data for evacuation response during hurricane Irma and augment the Twitter user with NHTS respondents based on behavioural similarity, thus creating a hybrid synthesized dataset with real time evacuation information and representative population coverage.

The fusion process is conducted using a probabilistic matching method based on a set of common attributes across NHTS and Twitter. There are several common attributes between these two datasets including age, gender, ethnicity, income, educational attainment and presence of a child in the house. Further, multiple twitter users can be matched with one NHTS respondent based on the similarity and hence, we employ a weight function to accommodate for the duplication. The weight is estimated considering two approaches: 1) deterministic approach where all attributes and matching records are weighted equally, (1/K, for K repetitions) and 2) a probabilistic approach where we relax the equal weight assumption and allow the weight to vary across the fused record based on the similarity (dissimilarity) between the common variables present in both datasets that are not used for fusion. The efficacy of the proposed fusion method is rigorously tested with a well-crafted experimental design evaluating the influence of multiple independent variables for matching and fusing, fusion sample sizes and weight functions. We estimate multiple models to serve as a benchmark and the model fit measures clearly highlight the improvement in data fit for the evacuation model through the proposed fusion algorithm and the weighted contribution estimation. Further, the estimation results provide strong evidence that relevant variables from the NHTS dataset have significant impact on the evacuation behaviour. Moreover, to demonstrate the applicability of the fused model, we conduct a prediction assessment to estimate the potential evacuation rates on a broader scale (in the state of Florida) considering all the three models and compare these predictions with actual evacuation figures. The validation results further provide evidence on the enhanced explanatory power and predictive capability achieved through the proposed fusion algorithm and latent weight contribution in the context of evacuation modeling. In summary, by merging real-time Twitter data with NHTS data, we develop a model to estimate evacuation responses in real time (from Twitter). Drawing on population weights from NHTS data, the approach provides a complete picture of the evacuation behaviour of the population as well as improve the quality of the evacuation prediction model.

Findings of this research can help facilitate improved decision making for emergency agencies in hurricane evacuation and disaster management. The improvements will have substantial societal and financial benefits by reducing loss of lives and resources during hurricanes. In future research, the analysis could be extended to explore the joint decision of evacuation as well as the timing dynamics. This could offer a comprehensive understanding of evacuation behavior during emergency scenarios, enhancing the overall effectiveness of emergency management strategies. Further, the research findings can be integrated with agent-based travel demand modeling software to generate evacuation rates for the entire state or urban region. However, it would be useful to consider data from multiple hurricanes to represent hurricane specific information (such as intensity, wind speed and path) on evacuation decisions. This will be a fruitful avenue for future research. In addition, this impact is not limited to hurricanes or specific geography only. By synthesizing multiple datasets augmenting our knowledge and how to incorporate it in developing predictive models, it is possible to build applications for other disasters such as wildfire and flood evacuations. In all these disaster contexts, the need to predict real-time population dynamics naturally arises. Finally, the proposed research will also contribute to the methodological advances in data science that are relevant across various fields. In our analysis, we employed Twitter as the source for evacuation data. It is possible that due to the changes in how Twitter data is shared, other researchers might not be able to access similar data in the future. However, the methods proposed in the paper can be applied to other data including location-based smart phone data and data from other social media platforms such as Facebook's "Data for Good" resource (Meta, 2023). Another important application area is in the transportation planning field where travel demand analysis is conducted while relying on household travel survey data. However, these survey data come with limited spatial coverage as well as are resource intensive. We can apply our fusion algorithm to combine the survey data with location-based services (LBS) data and create a fused dataset that offers improved spatiotemporal coverage and a comprehensive list of variables for developing the travel demand framework.

7. Author contribution statement

The authors confirm contribution to the paper as follows: study conception and design: Naveen Eluru, Samiul Hasan, Tanmoy Bhowmik and Aron Culotta; data collection: Aron Culotta, Tanmoy Bhowmik, Samiul Hasan, Naveen Eluru, Kamol Roy; model estimation and validation: Tanmoy Bhowmik, Naveen Eluru, Samiul Hasan; analysis and interpretation of results: Tanmoy Bhowmik, Naveen Eluru, Samiul Hasan; draft manuscript preparation: Tanmoy Bhowmik, Naveen Eluru, Samiul Hasan, Aron Culotta, Kamol Roy. All authors reviewed the results and approved the final version of the manuscript.

CRediT authorship contribution statement

Tanmoy Bhowmik: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Naveen Eluru:** Writing – review & editing, Writing – original draft, Validation, Supervision, Investigation, Funding acquisition, Conceptualization. **Samiul Hasan:** Writing – review & editing, Supervision, Formal analysis, Conceptualization. **Aron Culotta:** Supervision, Formal analysis. **Kamol Chandra Roy:** Formal analysis, Data curation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgment

This study was supported by the U.S. National Science Foundation through the grant CMMI #1917019 titled as "Collaborative Research: Predicting Real-time Population Behavior during Hurricanes Synthesizing Data from Transportation Systems and Social Media". However, the authors are solely responsible for the facts and accuracy of the information presented in the paper.

Appendix

A.1. Model performance summary for different variables, fusions and samples

The first table (Table A1) presents the average log-likelihood improvement for various variable combinations X at different K values while considering a fixed number of samples (N = 10). Following that, the second table (Table A2) highlights the average log-likelihood improvement considering the selected values of X and K from the first table while varying the sample size (N). Further, we also include the model run time as an additional performance indicator to provide a comprehensive understanding of the model's performance under different conditions.

Table A1
Model Performance Summary Across Different Variable Group and Number of Fused Records Used for Fusion.

Variable Combination	No. Fusion records (Run time)				
	3 (5 min)	5 (10 min)	10 (25 min)	15 (40 min)	20 (60 min)
Age, Race	0.00	1.31	2.85	2.86	2.86
Age, Gender	3.87	7.34	12.91	12.91	9.54

(continued on next page)

Table A1 (*continued*)

Variable Combination	No. Fusion records (Run time)				
	3 (5 min)	5 (10 min)	10 (25 min)	15 (40 min)	20 (60 min)
Race, Gender	0.00	0.00	2.03	3.11	2.76
Age, Income	0.00	1.11	1.89	1.96	1.96
Race, Income	0.00	0.87	2.35	2.39	2.43
Age, Education	0.00	0.00	3.93	3.91	3.91
Gender, Income	3.89	8.75	11.13	15.16	16.17
Age, Race, Income	**235.81**	**322.78**	**375.98**	**402.28**	**402.80**
Gender, Race, Income	6.57	7.65	8.31	11.12	9.11
Age, Gender, Race	67.21	89.57	120.00	122.67	122.43
Age, Gender, Education	0.00	2.31	7.11	9.65	9.65
Age, Education, Income	7.03	7.31	7.65	9.12	9.83

Table A2
Model Performance Summary Under Different Fusion Records and Sample Size.

Sample	5 fusion		10 fusion		15 fusion	
	Improvement	Run Time (Minutes)	Improvement	Run Time (Minutes)	Improvement	Run Time (Minutes)
1	298.31	15	378.32	25	397.32	40
3	304.98	45	367.57	75	397.91	120
5	306.05	75	374.68	125	401.60	200
10	**324.40**	**150**	**387.86**	**250**	**402.28**	**400**
15	327.61	225	388.75	375	405.12	600

From Tables A1 and A2, three observations can be made as follows: First, the best improvement in model performances is observed when the datasets are fused based on matching age, race and income **(X)**. Second, regarding the number of fused records, it is observed that matching 15 twitter records for each individual in the NHTS dataset leads to the best performance. However, this improvement comes at the cost of additional computational run time. Considering a trade-off between performance and computational efficiency, we identify 10 **(K)** as the optimal number of matched records for the subsequent analysis. Finally, analyzing the impact of sample size (N) on model performance, we observe that performance improves with higher sample size with an associated increase in run times. Considering the trade-off between the run time and model improvement, we select N = 10 as the sample size for the current analysis. Further, it is important to note that the purpose of varying N is specifically geared towards conducting the stability test of the fusion algorithm. Once we identify our optimal values for X and K through the initial analysis, it becomes imperative to ascertain the reliability of our model results. To achieve this, we systematically run our analysis on multiple samples (N) and perform a stability test. The primary goal of this test is to investigate whether there are any significant differences in model parameter estimates across different N samples. If the results consistently demonstrate stability, indicating no significant deviations in parameter estimates, it signifies statistical validity of our proposed fusion approach.

A.2. Log-Likelihood Ratio (LLR) test results

In addition to the Bayesian Information Criterion (BIC) as our primary measure for model comparison, we have also conducted log-likelihood ratio tests (LLR) across each pair of models for further evaluating the model performances. The results are presented in Tables A3. The results clearly highlight the enhanced performance of the evacuation model through the proposed fusion algorithm and weighted contribution estimation proposed in the paper.

Table A3
LL ratio Results.

Model*	Sample Size	No. Variables	Log-Likelihood	LL ratio
EWSBL	2,246	4	−2961.31	—
EWFBL	2,246	9	−2908.36	105.9 (from EWSBLE) critical value is 11.07 (5DF)
LWFBL	**2,246**	**10**	**−2572.99**	**670.74 (from EWFBL)** **critical value is 3.84 (1DF)**

Note: EWSBL = equal weight simple binary logit model; EWFBL = equal weight fused binary logit model.
LWFBL = Latent weight fused binary logit model.

References

Beiró, M.G., Panisson, A., Tizzoni, M., Cattuto, C., 2016. Predicting human mobility through the assimilation of social media traces into mobility models. EPJ Data Sci. 5 (1) https://doi.org/10.1140/epjds/s13688-016-0092-2.

Bhowmik, T., Iraganaboina, N.C., Eluru, N., 2023. A novel maximum likelihood based probabilistic behavioral data fusion algorithm for modeling residential energy consumption. Under Review in PLOS ONE.

Bi, H., Ye, Z., 2021. Exploring ridesourcing trip patterns by fusing multi-source data: a big data approach. Sustain. Cities Soc. 64, 102499 https://doi.org/10.1016/j.scs.2020.102499.

Bohte, W., Maat, K., 2009. Deriving and validating trip purposes and travel modes for multi-day GPS-based travel surveys: a large-scale application in the Netherlands. Transport. Res. Part C: Emerg. Technol. 17 (3), 285–297. https://doi.org/10.1016/j.trc.2008.11.004.

Culotta, A., Ravi, N.K., Cutler, J., 2015. Predicting the demographics of Twitter users from website traffic data. In: Proceedings of the National Conference on Artificial Intelligence, pp. 72–78. doi: 10.1609/aaai.v29i1.9204.

Culotta, A., Ravi, N.K., Cutler, J., 2016. Predicting twitter user demographics using distant supervision from website traffic data. J. Artif. Intell. Res. 55, 389–408. https://doi.org/10.1613/jair.4935.

Dash, N., Gladwin, H., 2007. Evacuation decision making and behavioral responses: individual and household. Nat. Hazard. Rev. 8 (3), 69–77.

Data Fusion - an overview | ScienceDirect Topics [WWW Document], n.d. URL https://www.sciencedirect.com/topics/computer-science/data-fusion (accessed 7.24.21).

Davidson, R.A., Nozick, L.K., Wachtendorf, T., Blanton, B., Colle, B., Kolar, R.L., Leonardo, N., 2020. An integrated scenario ensemble-based framework for hurricane evacuation modeling: Part 1—decision support system. Risk Anal. 40 (1), 97–116.

Dixon, D.S., Mozumder, P., Vásquez, W.F., Gladwin, H., 2017. Heterogeneity within and across households in hurricane evacuation response. Netw. Spat. Econ. 17, 645–680.

Duan, Y., Yisheng, L.V., Wang, F.Y., 2016. Travel time prediction with LSTM neural network. In: 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC). IEEE, pp. 1053–1058.

FirstCoast News, 2022. 25% of Floridians won't evacuate their homes for a hurricane | firstcoastnews.com [WWW Document]. URL https://www.firstcoastnews.com/article/news/local/survey-25-percent-of-floridians-wont-evacuate-their-homes-for-hurricane/77-980465d8-54d0-4cd1-a866-6b332e2b2755 (accessed 7.19.22).

Gladwin, C.H., Gladwin, H., Peacock, W.G., 2001. Modeling hurricane evacutaion decisions with ethnographic methods. Int. J. Mass Emerg. Disasters 19 (2), 117–143.

Greaves, S.P., 2006. Simulating household travel survey data. In: Travel Survey Methods: Quality and Future Directions. Emerald Group Publishing Limited, pp. 625–650.

Guan, X., Chen, C., Work, D., 2016. Tracking the evolution of infrastructure systems and mass responses using publically available data. PLoS One 11 (12), e0167267.

Hasan, S., Foliente, G., 2015. Modeling infrastructure system interdependencies and socioeconomic impacts of failure in extreme events: emerging R&D challenges. Nat. Hazards 78 (3), 2143–2168.

Hasan, S., Ukkusuri, S., Gladwin, H., Murray-Tuite, P., 2011. Behavioral model to understand household-level hurricane evacuation decision making. J. Transp. Eng. 137 (5), 341–348.

Hasan, S., Mesa-Arango, R., Ukkusuri, S., Murray-Tuite, P., 2012. Transferability of hurricane evacuation choice model: joint model estimation combining multiple data sources. J. Transp. Eng. 138 (5), 548–556.

Hasan, S., Mesa-Arango, R., Ukkusuri, S., 2013. A random-parameter hazard-based model to understand household evacuation timing behavior. Transport. Res. Part C: Emerg. Technol. 27, 108–116.

Hasan, S., 2010. A Behavioral Model to Understand Household Level Hurricane Evacuation Decision Making (Doctoral dissertation, Virginia Tech).

Huang, Z., Ling, X., Wang, P., Zhang, F., Mao, Y., Lin, T., Wang, F.Y., 2018. Modeling real-time human mobility based on mobile phone and transportation data fusion. Transport. Res. Part C: Emerg. Technol. 96, 251–269.

Iqbal, M.S., Choudhury, C.F., Wang, P., González, M.C., 2014. Development of origin–destination matrices using mobile phone call data. Transport. Res. Part C: Emerg. Technol. 40, 63–74.

Kang, J.E., Lindell, M.K., Prater, C.S., 2007. Hurricane evacuation expectations and actual behavior in Hurricane Lili 1. J. Appl. Soc. Psychol. 37 (4), 887–903.

Kanner & Pintalgua, 2018. The Lasting Damages of Hurricane Irma | Kanner & Pintaluga [WWW Document]. URL https://hurricanedamage.com/blog/damages-cost-hurricane-irma/(accessed 7.20.22).

Kumar, D., Ukkusuri, S.V., 2018, April. Utilizing geo-tagged tweets to understand evacuation dynamics during emergencies: a case study of Hurricane Sandy. In: Companion Proceedings of the Web Conference 2018, pp. 1613–1620.

Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A.L., Brewer, D., Van Alstyne, M., 2009. Computational social science. Science 323 (5915), 721–723.

Li, X., Hasan, S., Culotta, A. 2022, May. Identifying hurricane evacuation intent on Twitter. In: Proceedings of the International AAAI Conference on Web and Social Media, Vol. 16, pp. 618–627.

Liao, C.F., 2010. Fusing public and private truck data to support regional freight planning and modeling. In: TRB Strategic Highway Research Program 2 Symposium, p. 5.

Lindell, M.K., 2008. EMBLEM2: An empirically based large scale evacuation time estimate model. Transp. Res. A Policy Pract. 42 (1), 140–154.

Martín, Y., Li, Z., Cutter, S.L., 2017. Leveraging Twitter to gauge evacuation compliance: spatiotemporal analysis of Hurricane Matthew. PLoS One 12 (7), e0181701.

Martín, Y., Cutter, S.L., Li, Z., 2020. Bridging twitter and survey data for evacuation assessment of Hurricane Matthew and Hurricane Irma. Nat. Hazard. Rev. 21 (2), 04020003.

Mesa-Arango, R., Hasan, S., Ukkusuri, S.V., Murray-Tuite, P., 2013. Household-level model for hurricane evacuation destination type choice using hurricane Ivan data. Nat. Hazard. Rev. 14 (1), 11–20.

Meta, 2023. https://dataforgood.facebook.com/(accessed 1.29.24).

Momtaz, S.U., Eluru, N., Anowar, S., Keya, N., Dey, B.K., Pinjari, A., Tabatabaee, S.F., 2020. Fusing freight analysis framework and transearch data: econometric data fusion approach with application to Florida. J. Transport. Eng., Part A: Syst. 146 (2), 04019070.

Montero, L., Ros-Roca, X., Herranz, R., Barceló, J., 2019. Fusing mobile phone data with other data sources to generate input OD matrices for transport models. Transp. Res. Proc. 37, 417–424.

Murray-Tuite, P., Wolshon, B., 2013. Evacuation transportation modeling: an overview of research, development, and practice. Transport. Res. Part C: Emerg. Technol. 27, 25–45.

Olteanu, A., Castillo, C., Diaz, F., Kıcıman, E., 2019. Social data: biases, methodological pitfalls, and ethical boundaries. Front. Big Data 2, 13.

Palubinskas, G., Runge, H., 2008, June. Detection of traffic congestion in SAR imagery. In 7th European Conference on Synthetic Aperture Radar. VDE, pp. 1–4.

Pan, B., Zheng, Y., Wilkie, D., Shahabi, C., 2013. Crowd sensing of traffic anomalies based on human mobility and social media. In: Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, pp. 344–353.

Planetizen, 2005. Car-less in the Eye of Katrina | Planetizen Features [WWW Document]. URL https://www.planetizen.com/node/17255 (accessed 7.19.22).

Rambha, T., Nozick, L.K., Davidson, R., 2021. Modeling hurricane evacuation behavior using a dynamic discrete choice framework. Transp. Res. B Methodol. 150, 75–100.

Roy, K.C., Hasan, S., 2021. Modeling the dynamics of hurricane evacuation decisions from twitter data: an input output hidden Markov modeling approach. Transport. Res. Part C: Emerg. Technol. 123, 102976.

Roy, K.C., Cebrian, M., Hasan, S., 2019. Quantifying human mobility resilience to extreme events using geo-located social media data. EPJ Data Sci. 8 (1), 1–15.

Roy, K.C., Hasan, S., Culotta, A., Eluru, N., 2021. Predicting traffic demand during hurricane evacuation using real-time data from transportation systems and social media. Transport. Res. Part C: Emerg. Technol. 131, 103339.

Sadri, A.M., Ukkusuri, S.V., Murray-Tuite, P., Gladwin, H., 2015. Hurricane evacuation route choice of major bridges in Miami Beach, Florida. Transp. Res. Rec. 2532 (1), 164–173.
Sarwar, M.T., Anastasopoulos, P.C., Ukkusuri, S.V., Murray-Tuite, P., Mannering, F.L., 2018. A statistical analysis of the dynamics of household hurricane-evacuation decisions. Transportation 45, 51–70.
SRESP, 2021. Florida SRESP: East Central Florida Regional Planning Council https://portal.floridadisaster.org/preparedness/RES/Studies/Shared%20Documents/Supporting%20Documents/Region-Specific%20Folders/East%20Central%20FL/2021_SRESP_BehavioralStudy_ECFRPC_Final.pdf (accessed 12.19.23).
The Atlantic, 2017. The Uncertainties of Being Asked to Work During a Hurricane - The Atlantic [WWW Document]. URL https://www.theatlantic.com/business/archive/2017/09/hurricane-irma-work/539295/(accessed 7.19.22).
Wang, F., Chen, C., 2018. On data processing required to derive mobility patterns from passively-generated mobile phone data. Transport. Res. Part C: Emerg. Technol. 87, 58–74.
Wu, C., Thai, J., Yadlowsky, S., Pozdnoukhov, A., Bayen, A., 2015. Cellpath: fusion of cellular and traffic sensor data for route flow estimation via convex optimization. Transp. Res. Procedia 7, 212–232.
Yabe, T., Ukkusuri, S.V., 2020. Effects of income inequality on evacuation, reentry and segregation after disasters. Transport. Res. Part D: Transp. Environ. 82, 102260.
Yang, D., Xie, K., Ozbay, K., Yang, H., 2021. Fusing crash data and surrogate safety measures for safety assessment: development of a structural equation model with conditional autoregressive spatial effect and random parameters. Accid. Anal. Prev. 152, 105971.
Yang, C., Zhang, Y., Zhan, X., Ukkusuri, S.V., Chen, Y., 2020. Fusing mobile phone and travel survey data to model urban activity dynamics. J. Adv. Transp. 2020, 1–17.
Yasmin, S., Eluru, N., Pinjari, A.R., 2015. Pooling data from fatality analysis reporting system (FARS) and generalized estimates system (GES) to explore the continuum of injury severity spectrum. Accid. Anal. Prev. 84, 112–127.
Zhao, D., Balusu, S.K., Sheela, P.V., Li, X., Pinjari, A.R., Eluru, N., 2020. Weight-categorized truck flow estimation: a data-fusion approach and a Florida case study. Transport. Res. Part E: Logist. Transport. Rev. 136, 101890.