Proceedings of the 2021 Winter Simulation Conference
S. Kim, B. Feng, K. Smith, S. Masoud, Z. Zheng, C. Szabo and M. Loper, eds.

PUBLIC DEMAND ESTIMATION FOLLOWING DISASTERS THROUGH INTEGRATING SOCIAL MEDIA AND COMMUNITY DEMOGRAPHICS

Yudi Chen
Wenying Ji

Department of Civil, Environmental, and Infrastructure Engineering
George Mason University
4400 University Drive
Fairfax, VA 22030, USA

ABSTRACT

Following disasters, a timely and reliable estimation of public demand—the number of individuals having demand—is essential for allocating relief resources properly. However, such an estimation is challenging as public demand varies significantly in dynamic disasters. To address the challenge, this research estimates public demand through proposing a data-driven approach that integrates social media and community demographics. In detail, social media is used to derive the percentage of a population having demand, and demographics are applied to normalize population differences amongst races/ethnicities. The proposed approach is capable of (1) eliminating the social media bias caused by racial disparities on social media platforms, and (2) modeling the uncertainty of social media-derived demand percentage. Hurricane Irma-induced food demand in Florida is studied to prove the feasibility of the proposed data-driven approach. In addition, the research sheds light on the use of partial information for deriving insights for the entire population.

1 INTRODUCTION

In recent decades, climate-related natural hazards have been occurring at an increasing frequency and intensity, which has brought devastating and widespread disaster impacts (Wallemacq et al. 2015). For example, from 1988 to 2017, disaster-hit countries reported direct economic losses of 2,908 billion US dollars and a death toll of 1.3 million people (UNISDR 2018). To reduce such devastating disaster impacts, relief agencies need to conduct effective disaster response that relies on a timely and reliable estimation of public demand. Here, the estimation refers to the quantification of public demand on relief resources, such as the number of individuals in need of food, water, and shelter. With such a quantitative estimation, agencies can appropriately allocate relief resources to satisfy public demand, thereby saving lives.

To reliably estimate public demand following disasters, researchers have developed various estimation approaches that can be categorized into two groups: knowledge-based approaches (Liu et al. 2012; Trivedi and Singh 2017) and statistical approaches (Ferbar Tratar et al. 2016; Trivedi and Singh 2017; Zhu et al. 2016). Knowledge-based approaches estimate public demand through investigating historical experiences of similar disasters with analytical methods, such as analytic hierarchy process (Trivedi and Singh 2017) and case-based reasoning (Liu et al. 2012). Taking earthquake as an example, public demand in a disaster-impacted region is derived using the experiences from regions that had similar ground-shaking parameters (e.g., seismic intensity) and community demographics (e.g., population) (Zhu et al. 2016). Although knowledge-based approaches perform well in many situations, their estimation performance highly relies on practitioners' domain knowledge, which makes it challenging for wide implementation. In contrast to knowledge-based approaches, statistical approaches explore the relationships between public demand and

978-1-6654-3311-2/21/$31.00 ©2021 IEEE

various features (e.g., disaster type and severity, community demographics, and building information) (Zhu et al. 2019). The most commonly used statistical approaches are autoregressive moving average (ARMA) (Holguín-Veras and Jaller 2012) and exponential smoothing models (Ferbar Tratar et al. 2016). However, these statistical approaches sometimes malfunction due to the strong nonlinearity and dynamics of public demand following disasters (Zhu et al. 2019). In addition, statistical approaches require reference data points for learning the relationship between demand and features, which makes them less capable in the scenario with limited demand information, such as the beginning of a disaster. Such limitations of existing approaches underscore the necessity of developing innovative and reliable approaches for estimating public demand following disasters.

Recently, researchers have broadly used social media to investigate human behavior, wherein social media users are assumed to be a representative sample set of that population (Chen et al. 2019, 2021; Li et al. 2017; Simon et al. 2015). In the disaster context, social media has been used for analyzing public demand in a rapid manner (Imran et al. 2018). Examples include (1) the identification of medical demand-related social media posts for optimizing the locations of temporary rescue centers (Schempp et al. 2019); (2) the analysis of demand-related sentiment for indicating the urgency of relief resources (Ragini et al. 2018); and (3) the modeling of demand distribution on various relief resources (food, water, etc.) for planning relief operations (Zhang et al. 2021). Compared to existing knowledge-based and statistical approaches, social media-based approaches are more efficient as they (1) require no profound disaster domain knowledge; (2) provide demand-related information in a near-real-time manner; and (3) require no reference data points for estimation. However, existing studies only analyze public demand for social media users, which makes them less capable of estimating public demand for the entire disaster-impacted population. Therefore, it is necessary to scale the demand-related information from social media users to the entire disaster-impacted population, but relevant investigations have yet been conducted.

This research aims to address the research gap through proposing a data-driven approach that integrates social media and community demographics, thereby achieving a rapid and reliable estimation of public demand for the entire disaster-impacted population. The proposed approach (1) removes the social media bias caused by racial disparities on social media platforms through independently modeling public demand percentage for each race/ethnicity; and (2) includes the sampling uncertainty of social media-derived demand percentage through proposing a Bayesian-based approach. To prove the feasibility and applicability of the proposed approach, hurricane Irma-induced food demand in Florida is studied. The remainder of this paper is organized as follows. First, a rationale is introduced to illustrate the challenges of using social media to derive insights on the entire population. Second, details of the research methodology are presented. Third, estimated public demand is presented and discussed. Finally, research contributions, limitations, and potential future work are concluded.

2 RATIONALE

Researchers and practitioners often use a sample set of a population to derive the characteristics of that population as experiments on a sample set are often practical, cost-effective, and time-efficient (Box and Tiao 2011). In this research, demand-related information posted by social media users, a sample population set, is used to estimate public demand for the entire population. To achieve a reliable estimation, we should address the following major concerns: social media data bias and sampling uncertainty.

In practice, researchers assume that the sample set has the same characteristics as the population, which is needed when using a sample set to derive insights into that population (Box and Tiao 2011). However, this assumption is often not true for social media users as they are biased by many demographic factors, such as race/ethnicity, age, and gender (Elliott and Pais 2006; Murthy et al. 2016; Ribeiro et al. 2020). Extensive studies have pointed out significant social media behavioral differences amongst races/ethnicities (e.g., White, Black, Asian, and Hispanic) (Elliott and Pais 2006; Yuan et al. 2021). For example, Hispanic and Black groups are more likely to express negative sentiment than the White group, while the race/ethnicity minority Asian is less likely to express negative sentiment than the White group (Yuan et al. 2021). And the accessibility to social media varies amongst racial groups in the US (Mislove et al. 2011). Such social media bias should be removed in order to reliably derive public demand on the disaster-impacted population using social media. To achieve the purpose, the race/ethnicity for each social media user is predicted, and the demand percentage is independently modeled for each race/ethnicity.

In statistics, sampling uncertainty incurs when a population's attributes are estimated from a sample set of that population (Box and Tiao 2011). The sampling uncertainty is inversely correlated with the sample size: the larger the sample size is, the smaller the sampling uncertainty is. Although social media's openness enables the collection of a large-scale dataset, the collected large-scale dataset is actually a collection of small datasets with various specific conditions (Ghahramani 2015). For example, millions of disaster-related tweets are collected in hurricane Harvey, but the highway-related tweets only account for an extremely small portion, 0.66% (Chen et al. 2020). Such a small social media dataset for a specific condition (public demand in this research) makes it necessary to include the sampling uncertainty when using social media data to derive insights on the entire population. In this research, a Bayesian-based method is proposed to model the sampling uncertainty of social media-derived demand percentage.

3 METHODOLOGY

This research proposes a systematic data-driven approach to estimate public demand through integrating social media and community demographics, as shown in Figure 1. It consists of three modules: data collection, social media data processing, and demand estimation. In the data collection module, social media data and community demographics are collected. In the social media data processing module, social media users' races/ethnicities (i.e., White, Black, Asian, and Hispanic) are predicted by matching their last names with the US Census data (US Census Bureau 2000), and demand-related social media posts are extracted through a set of predefined keywords. In the demand estimation module, public demand for the disaster-impacted population is derived through aggregating demand of all races/ethnicities. Demand percentage indicates the ratio of a population having demand, and it is modeled with a Bayesian-based method to include the sampling uncertainty. To remove the social media bias, the demand percentage is independently modeled for each race/ethnicity.

Public food demand in Florida following Hurricane Irma is studied to illustrate the feasibility and applicability of the proposed methodology. Hurricane Irma was an extremely powerful hurricane that caused widespread destruction across its path in September 2017. Hurricane Irma made landfall in Florida on September 10th, 2017 as a Category 4 hurricane, then traveled up, and finally moved away from Florida on September 11th, 2017 (Cangialosi et al. 2018). Hurricane Irma caused major damage to roads, sanitation, the water, medical, electricity, and fuel supply, which underscores the importance of timely and responsive disaster response. In this case, food demand is studied for illustration as they are directly related to life-sustaining measures.

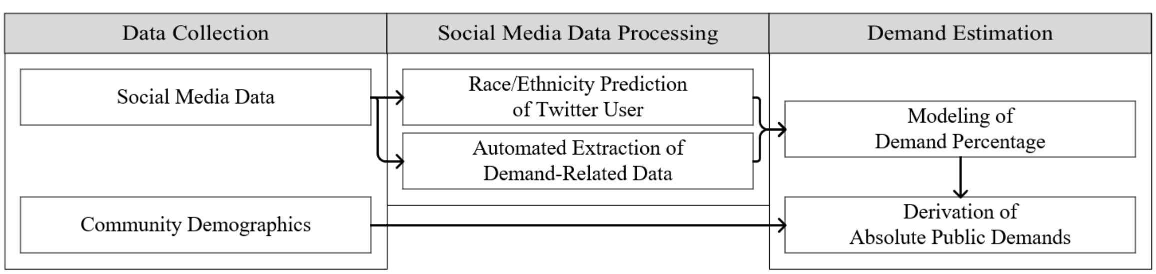

Figure 1: Research methodology.

3.1 Data Collection

3.1.1 Social Media Data

Social media posts were collected from Twitter due to its openness and easy access for collecting large-scale datasets (Kryvasheyeu et al. 2016). To obtain each social media user's location, we only collected geotagged tweets whose locations are characterized by a latitude and longitude pair. A location bounding box was used to ensure the collected tweets were posted in disaster-impacted regions. The bounding box comprises latitudes and longitudes of boundaries of disaster-impacted regions. The bounding box was created to cover the whole Florida state, which was severely impacted during hurricane Irma, with corresponding latitudes [24.754420, 31.099775] and longitudes [-79.536765, -87.758908]. The collection period is from September 10th, 2017, the landfall date of hurricane Irma in Florida, to September 15th, 2017, six days after the hurricane Irma landfall. The collected geotagged tweets were further mapped to each county in Florida based on their geotags. Finally, 14,791 tweets posted by 5,867 users were saved for further analysis.

3.1.2 Community Demographics

Community demographics are the statistical representations of community population, and they comprise many factors, such as population size, race/ethnicity, and household income. In this research, we collected the population size for each race/ethnicity. Race/ethnicity population information has been well-established by US Census (US Census Bureau 2000), and it was collected over the occurred year (2017) of hurricane Irma to ensure time consistencies.

3.2 Social Media Data Processing

In this section, social media users' race/ethnicity is predicted, and demand-related social media posts are efficiently extracted through a set of predefined demand-related keywords.

3.2.1 Race/Ethnicity Prediction of Twitter User

In this research, public demand is independently modeled for each race/ethnicity in order to eliminate the social media bias caused by the behavioral differences amongst races/ethnicities. To achieve the purpose, we need to obtain the Twitter users' races/ethnicities for grouping the collected social media posts into each race/ethnicity. In this research, Twitter users' race/ethnicity is predicted by matching the last name of Twitter user profile name and the last name database released by the US Census Bureau (US Census Bureau 2000). For each last name with over 100 individuals in the US, the Census Bureau releases the percentage distribution of race/ethnicity for that last name, as presented in Table 1. For example, the last name "Smith" was observed to correspond to White with a 70.90%, Black 23.11%, Asian 0.89%, and Hispanic 2.40%. Notably, the sum of the percentages for these races/ethnicities is not necessarily equal to 1 as other races/ethnicities (e.g., pacific islanders) are not included in this study due to their population sparseness.

Table 1: Percentage distribution of the last names.

Last name	Count	White	Black	Asian	Hispanic
Smith	2442977	70.90%	23.11%	0.89%	2.40%
Johnson	1932812	58.97%	34.63%	0.94%	2.36%
Kariuki	28311	3.32%	94.43%	0.00%	0.00%
Piedra	5098	5.00%	0.78%	0.12%	93.35%
Madosh	114	6.14%	0.00%	85.09%	0.00%

The race/ethnicity of a social media user is predicted as the dominant race/ethnicity of the user's last name in the database. Such a prediction method has been widely used in previous research studies (Dargin et al. 2021; Yuan et al. 2021). Due to Twitter user profile information's informality and the last name database's incompleteness, not all Twitter users' races/ethnicities are predictable. In this case, 1,668 of 5,867 users' races/ethnicities are predicted. The comparison of race/ethnicity distributions between Twitter users and the entire population in Florida is presented in Figure 2. Notably, the distributions are presented in a relative manner, wherein the percentage for each race/ethnicity is presented. The percentage value for a race/ethnicity in Twitter is calculated via dividing the number of users having that race/ethnicity with the total number of users. The percentage value of a race/ethnicity in Florida's population is calculated through dividing the number of individuals having that race/ethnicity by the entire population. The White group is more active as its occupied percentage on Twitter is higher than its actual percentage in Florida. In contrast, the Black group is less active as its occupied percentage on Twitter is lower than its actual percentage in Florida. Such distribution differences highlight the necessity of removing the social media bias caused by the behavioral differences amongst races/ethnicities.

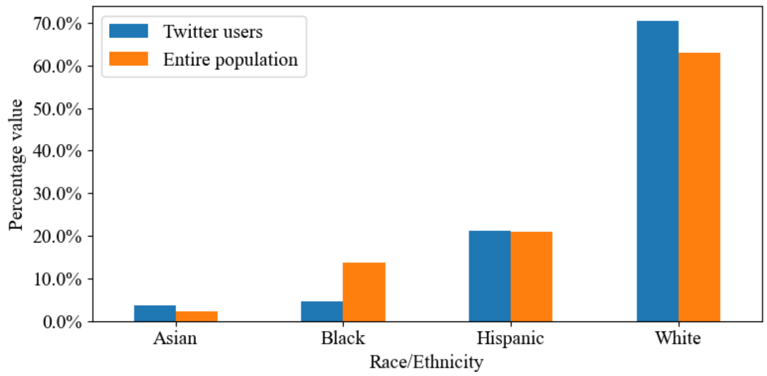

Figure 2: Race/Ethnicity distributions for Twitter users and the entire population in Florida.

3.2.2 Automated Extraction of Demand-related Data

To reliably extract public demand-related information, social media posts are processed through data cleaning and keyword determination. Data cleaning is essential in social media mining as social media data often contain excessive jargon and inconsistencies (e.g., spelling errors) (Zafarani et al. 2014). In this research, data cleaning is conducted in five steps:

- Step 1: Remove highly active social media users that are often bots and advertisements,
- Step 2: Remove useless URL links as this study is based on text information,
- Step 3: Tokenize a tweet text into separate unigrams based on commonly used delimiters,
- Step 4: Lemmatize the tokenized unigrams into their root terms, such as "injuries" to "injury", and
- Step 5: Combine the tweets posted by the same Twitter user on each studied date.

Notably, the threshold for determining highly active users is empirically identified. In this case, it is set as 20 to remove the Twitter users who posted more than 20 tweets in the studied period (i.e., from September 10th, 2017 to September 15th, 2017). With these cleaning steps, each tweet text is represented by a set of enhanced unigrams, which is critically needed for ensuring the efficiency and accuracy of extracting public demand from social media. Additionally, the same user's tweets on each date are combined as this research aims to estimate the number of individuals having demand. By doing so, each combined tweet indicates the information of a social media user at a specific date.

Public demand is extracted from the cleaned social media texts through a set of demand-related keywords. In this research, the demand-related keywords are empirically determined through checking the related information in social media. For food demand, the related keywords are "food", "meal", "noodle", "rice", and "hungry". Finally, 222 social media users are identified as having demand during the studied period (i.e., September 10th, 2017 to September 15th, 2017).

3.3 Demand Estimation

In this section, public demand for the entire population is estimated by modeling the demand percentage for each race/ethnicity and then aggregating the modeled public demand with population weights.

3.3.1 Modeling of Demand Percentage

In this research, demand percentage (DP) is modeled solely using social media, as shown in Equation (1).

$$DP_{r,c,d} = n_{r,c,d}/N_{r,c,d} \tag{1}$$

where n is the number of social media users having demand, and N is the total number of social media users. The subscripts r, c, d represent race/ethnicity, county, and date, respectively. Given that social media users are only a sample population set, the modeled demand percentage is an uncertain variable.

A Bayesian-based method is proposed to model the sampling uncertainty of the demand percentage. In Bayesian inference, the demand-related domain knowledge is modeled with a prior distribution, and the new demand observations from social media are modeled with a likelihood function. In this case, Beta distribution is used to model the prior demand-related domain knowledge due to the following reasons: (1) demand percentage is bounded within the range of [0, 1] that matches the natural boundaries of Beta distribution; and (2) parameters of the Beta distribution are intuitively and physically meaningful. The prior demand-related domain knowledge on demand percentage DP is modeled as Equation (2).

$$p(DP) = Beta(DP|a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} DP^{a-1}(1-DP)^{b-1} \tag{2}$$

where Γ is a Gamma function. a and b are the two shape parameters of the Beta distribution. In general, there are two types of priors: informative and non-informative priors. An informative prior provides specific and definite information, while a non-informative prior provides vague, flat, and diffuse information. The determination of an informative prior requires profound domain knowledge, which is out of this research scope. Therefore, in this research, a non-informative prior $Beta(0.5, 0.5)$ is used for illustration.

Given that each social media user is labeled as either demand-related or non-demand-related using keyword filtering, which is a binary problem, the Binomial distribution is used to model the new demand observations from social media. The number of demand-related social media users n follows the binomial distribution with parameters N and DP. The probability of getting exactly n successes in N independent Bernoulli trials is given by the probability mass function, as presented in Equation (3).

$$L(n|N, DP) = \binom{N}{n} DP^n (1-p)^{N-n} \tag{3}$$

The posterior of demand percentage is obtained through integrating the prior and the new demand observations from social media. Give that the prior Beta distribution is a conjugate prior to the likelihood Binomial distribution, the posterior is derived with an analytical form, as presented in Equation (4).

$$p(DP|n, N) = Beta(n + a, N - n + b) \tag{4}$$

Notably, the shape parameters $n + a$ and $N - n + b$ of the posterior are the simple summations of the corresponding parameters a and b in the prior distribution and the numbers of observations n and $N - n$ for demand-related and non-demand-related social media users. This straightforward property provides an easy and efficient way for deriving the posterior distribution to update the demand percentage. Notably, the demand percentage is modeled for each race/ethnicity, and a further aggregation across all races/ethnicities is needed to estimate public demand on the entire disaster-impacted population.

3.3.2 Derivation of Total Public Demand

Public demand on the entire disaster-impacted population is derived through aggregating the modeled demand percentages for all races/ethnicities with population weights, as presented in Equation (5).

$$D_{c,d} = \sum\nolimits_{r \in R} ps_{r,c} \times DP_{r,c,d} \tag{5}$$

where D represents the number of individuals having demand, $R = \{\text{White}, \text{Black}, \text{Asian}, \text{Hispanic}\}$, and ps is the population size for each race/ethnicity r in county c. Given that DP is an uncertain variable, Monte Carlo simulations are conducted to derive the final distribution of *D*. To ensure a reliable representation, 1,000 *DP* sets are sampled from the modeled posterior distributions of the four races, and each *DP* set derives one sample of *D*.

4 RESULTS

In this section, the modeled posteriors of demand percentage and the final derived total public food demand in Florida are presented to prove the feasibility and applicability of the proposed approach. The posteriors of food demand percentages for the four races/ethnicities are listed in Table 2. Only a small number of social media users express food demand-related information as the first parameters of the posterior Beta distributions are small. This indicates that even there are many individuals experiencing disasters, only a small number of individuals need food assistance. The aggregated public food demand in Florida is presented in Figure 3. For the studied period, public food demand has considerable uncertainties, which proves the necessity of incorporating the sampling uncertainty in demand percentage modeling. Public demand began to increase on September 12th, 2017, just the day following hurricane Irma, reached the peak on September 13th, 2017, and then decreased to a low level on September 15th, 2017. Such temporal trends are consistent with the intuitive assumptions: public food demand increases instantly following a disaster due to adverse disaster impacts and then decreases due to the distribution of food resources.

Table 2: The modeled posterior of the demand percentage for each race/ethnicity.

Date	**White**	**Black**	**Asian**	**Hispanic**
2017-09-10	Beta(22.5, 1303.5)	Beta(3.5, 69.5)	Beta(2.5, 55.5)	Beta(6.5, 380.5)
2017-09-11	Beta(20.5, 1086.5)	Beta(2.5, 54.5)	Beta(1.5, 50.5)	Beta(1.5, 297.5)
2017-09-12	Beta(23.5, 921.5)	Beta(2.5, 56.5)	Beta(1.5, 49.5)	Beta(6.5, 303.5)
2017-09-13	Beta(24.5, 1029.5)	Beta(2.5, 64.5)	Beta(2.5, 54.5)	Beta(13.5, 297.5)
2017-09-14	Beta(30.5, 997.5)	Beta(3.5, 76.5)	Beta(2.5, 51.5)	Beta(11.5, 330.5)
2017-09-15	Beta(34.5, 1127.5)	Beta(1.5, 87.5)	Beta(5.5, 61.5)	Beta(6.5, 347.5)

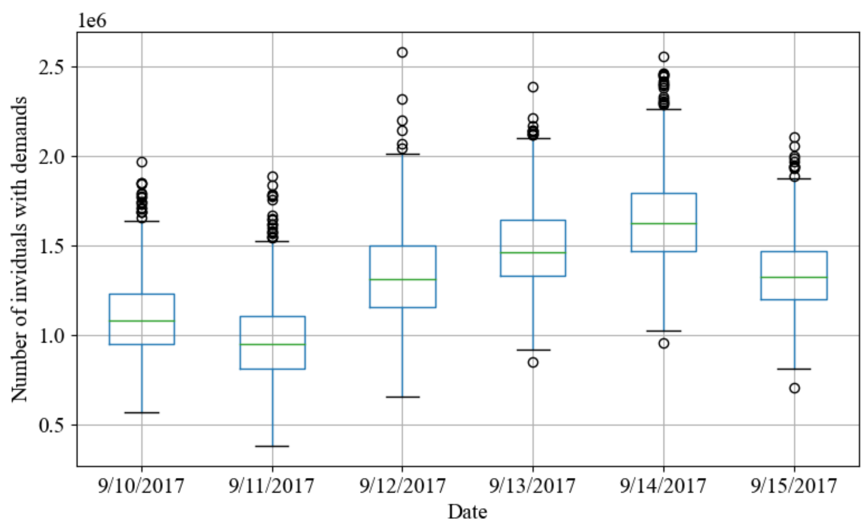

Figure 3: Hurricane Irma-induced public demand in Florida.

In reality, the number of social media users varies significantly amongst disaster-impacted regions. To examine the robustness of the proposed approach, we conducted a sensitivity analysis by modifying the number of social media posts. In this analysis, the demand percentages for all races/ethnicities on September 14th, 2017, which has the highest public demand, are selected. These demand percentages are kept constant, but the numbers of total social media users (the denominator of a demand percentage) and the social media users having demand (the numerator of a demand percentage) are scaled by multiplying a factor. The scale factor is set to cover both the areas with fewer Twitter activities (i.e., scale factor < 1.0) and the areas with more Twitter activities (i.e., scale factor > 1.0). By doing so, the applications of the proposed approach are simulated in regions with different numbers of social media posts. As shown in Figure 4, the uncertainty rapidly decreases with the increase of scales, which indicates that a large number of social media users improves the uncertainty. Additionally, the estimated number of individuals is always reasonable (i.e., >0) on all studied scales.

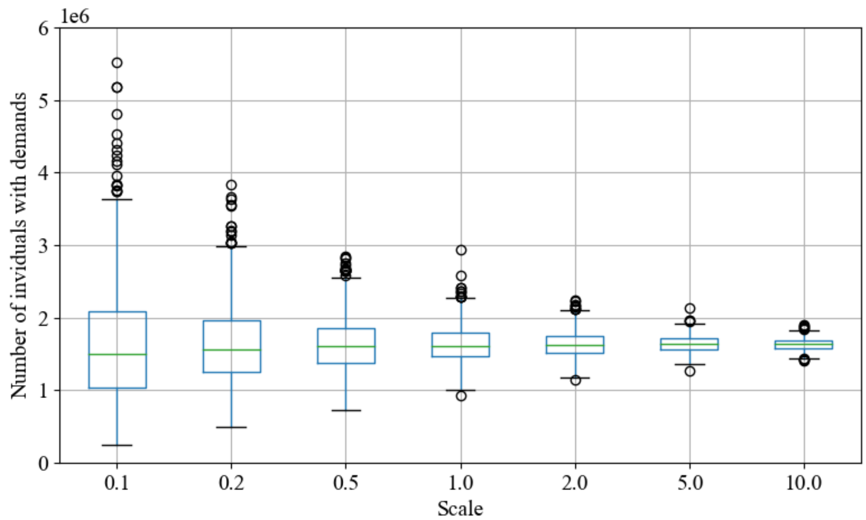

Figure 4: The estimated public demand with different scales.

5 CONCLUSIONS

This research proposes a systematic data-driven approach to derive public demand for the entire population through integrating social media and community demographics. In the proposed approach, the social media bias caused by racial disparities is removed by modeling demand percentage for each race/ethnicity independently. The uncertainty of social media-derived demand percentage is included by proposing a Bayesian-based method. Hurricane Irma-induced food demand in Florida is studied to illustrate the feasibility. Results show that the proposed approach is capable of estimating public demand in a timely manner (i.e., daily). Furthermore, sensitivity analysis is conducted to prove the robustness of the proposed approach. This research sheds light on the use of social media information for deriving insights into the entire population, which is a great benefit for researchers and practitioners to rapidly understand population characteristics. This study contributes to academia by (1) proposing a systematic data-driven approach to estimate public demand for the entire population; (2) removing the social media bias caused by racial disparities; and (3) developing a Bayesian-based method to include the sampling uncertainty of social media-derived demand percentage. Practically, the estimated public demand is expected to help practitioners make informed decisions, thereby promptly addressing public demand.

Although this presented research proposes a systematic methodology to derive public demand for the entire population using social media, it also has several limitations caused by the sparseness and bias of social media data and the informality of Twitter user profile information. First, demand-related information is sparse on social media, which makes the proposed approach limited in regions that have fewer social media activities, such as rural counties. Second, this research solely considers the social media bias caused by race/ethnicity. The investigation of social media bias caused by other demographic factors (e.g., age and gender) has not been investigated and removed. Third, the prediction accuracy of Twitter user race/ethnicity might be problematic due to the informality of Twitter user profile information (e.g., Twitter users are not using their real names) and the variation of last name's racial distributions amongst regions.

In the future, there are three potential research threads to enhance and extend this study. First, a major reason for the sparseness of demand-related social media data is that the employed geotagged tweets only account for 1%-2% of the entire tweets. In the future, data enrichment methods (profile-based (Kryvasheyeu et al. 2016) and content-based (Mao et al. 2019)) will be used to increase the available number of tweets. Second, a more holistic consideration of influencing factors (e.g., distance to hazard locations) will be investigated and included to enhance the reliability of estimated public demand. Third, systematic validation is also needed to prove the reliability of the proposed approach. And once the estimated demand is validated, they will be incorporated into the allocation of relief resources to enhance disaster response effectiveness.

ACKNOWLEDGMENTS

This study is supported by the Thomas F. and Kate Miller Jeffress Memorial Trust. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the Thomas F. and Kate Miller Jeffress Memorial Trust. The authors also gratefully acknowledge the support of 4-VA at Mason Collaborative Research Grants.

REFERENCES

Box, G. E. P., and Tiao, G. C. 2011. *Bayesian Inference in Statistical Analysis*. New Jersey: John Wiley & Sons.

Cangialosi, J. P., Latto, A. S., and Berg, R. 2018. Hurricane Irma Tropical Cyclone Report. https://www.nhc.noaa.gov/data/tcr/AL112017_Irma.pdf, assessed 5th April 2021.

Chen, Y., Umana, A., Yang, C., and Ji, W. 2021. "Condition Sensing for Electricity Infrastructure in Disasters by Mining Public Topics from Social Media." In *Proceedings of the 18th ISCRAM Conference*, May 23rd-May 26th, Blacksburg, VA.

Chen, Y., Wang, Q., and Ji, W. 2019. "A Bayesian-based Approach for Public Sentiment Modeling." In *Proceedings of the 2019 Winter Simulation Conference*, edited by N. Mustafee, K.-H.G. Bae, S. Lazarova-Molnar, M. Rabe, C. Szabo, P. Haas, and Y.-J. Son, 3053–3063. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Chen, Y., Wang, Q., and Ji, W. 2020. "Rapid Assessment of Disaster Impacts on Highways Using Social Media". *Journal of Management in Engineering*, 36(5), 04020068.

Dargin, J. S., Fan, C., and Mostafavi, A. 2021. "Vulnerable Populations and Social Media Use in Disasters: Uncovering the Digital Divide in Three Major U.S. Hurricanes". *International Journal of Disaster Risk Reduction*, 54(December 2020), 102043.

Elliott, J. R., and Pais, J. 2006. "Race, Class, and Hurricane Katrina: Social Differences in Human Responses to Disaster". *Social Science Research*, 35(2), 295–321.

Ferbar Tratar, L., Mojškerc, B., and Toman, A. 2016. "Demand Forecasting with Four-Parameter Exponential Smoothing". *International Journal of Production Economics*, 181, 162–173.

Ghahramani, Z. 2015. "Probabilistic Machine Learning and Artificial Intelligence". *Nature*, 521, 452–521.

Holguín-Veras, J., and Jaller, M. 2012. "Immediate Resource Requirements after Hurricane Katrina". *Natural Hazards Review*, 13(2), 117–131.

Imran, M., Castillo, C., Diaz, F., and Vieweg, S. 2018. "Processing Social Media Messages in Mass Emergency: Survey Summary". In *The Web Conference 2018 - Companion of the World Wide Web Conference*, April 23rd-27th, Lyon, France, 507–511.

Kryvasheyeu, Y., Chen, H., Obradovich, N., Moro, E., Van Hentenryck, P., Fowler, J., and Cebrian, M. 2016. "Rapid Assessment of Disaster Damage Using Social Media Activity". *Science Advances*, 2(3), 1–12.

Li, X., Wang, Z., Gao, C., and Shi, L. 2017. "Reasoning Human Emotional Responses from Large-Scale Social and Public Media." *Applied Mathematics and Computation*, 310, 182–193.

Liu, W., Hu, G., and Li, J. 2012. "Emergency Resources Demand Prediction Using Case-Based Reasoning". *Safety Science*, 50(3), 530–534.

Mao, H., Thakur, G., Sparks, K., Sanyal, J., and Bhaduri, B. 2019. "Mapping Near-Real-Time Power Outages from Social Media". *International Journal of Digital Earth*, 12(11), 1285–1299.

Mislove, A., Lehmann, S., Ahn, Y.-Y., Onnela, J.-P., and Rosenquist, J. N. (2011). "Understanding the Demographics of Twitter Users." In *International AAAI Conference on Web and Social Media*. July 17th-21st, Barcelona, Spain, 554-557.

Murthy, D., Gross, A., and Pensavalle, A. 2016. "Urban Social Media Demographics: An Exploration of Twitter Use in Major American Cities". *Journal of Computer-Mediated Communication*, 21(1), 33–49.

Ragini, J. R., Anand, P. M. R., and Bhaskar, V. 2018. "Big Data Analytics for Disaster Response and Recovery through Sentiment Analysis". *International Journal of Information Management*, 42, 13–24.

Ribeiro, F. N., Benevenuto, F., and Zagheni, E. 2020. "How Biased is the Population of Facebook Users? Comparing the Demographics of Facebook Users with Census Data to Generate Correction Factors". In *12th ACM Conference on Web Science*, Southampton, United Kingdom, 325–334.

Schempp, T., Zhang, H., Schmidt, A., Hong, M., and Akerkar, R. 2019. "A Framework to Integrate Social Media and Authoritative Data for Disaster Relief Detection and Distribution Optimization". *International Journal of Disaster Risk Reduction*, 39(August 2018), 101143.

Simon, T., Goldberg, A., and Adini, B. 2015. "Socializing in Emergencies - A Review of the Use of Social Media in Emergency Situations." *International Journal of Information Management*, 35(5), 609–619.

Trivedi, A., and Singh, A. 2017. "A Hybrid Multi-Objective Decision Model for Emergency Shelter Location-Relocation Projects Using Fuzzy Analytic Hierarchy Process and Goal Programming Approach". *International Journal of Project Management*, 35(5), 827–840.

UNISDR. 2018. UN 20-Year Review: Earthquakes and Tsunamis Kill More People While Climate Change is Driving up Economic Losses. https://www.undrr.org/news/un-20-year-review-earthquakes-and-tsunamis-kill-more-people-while-climate-change-driving, assessed 5th April 2021.

US Census Bureau. 2010. Frequently Occurring Surnames from Census 2010. https://www.census.gov/topics/population/genealogy/data/2010_surnames.html, assessed 5th April 2021.

Wallemacq, P., Guha-Sapir, D., and McClean, D. 2015. The Human Cost of Weather-Related Disasters. https://www.unisdr.org/2015/docs/climatechange/COP21_WeatherDisastersReport_2015_FINAL.pdf, assessed 5th April 2021.

Yuan, F., Li, M., Liu, R., Zhai, W., and Qi, B. 2021. "Social Media for Enhanced Understanding of Disaster Resilience During Hurricane Florence". *International Journal of Information Management*, 57(December 2020), 102289.

Zafarani, R., Abbasi, M. A., and Liu, H. 2014. *Social Media Mining: An Introduction*. Cambridge, United Kingdom: Cambridge University Press.

Zhang, T., Shen, S., Cheng, C., Su, K., and Zhang, X. 2021. "A topic Model-based Framework for Identifying the Distribution of Demand for Relief Supplies Using Social Media Data". *International Journal of Geographical Information Science*, 1–22.

Zhu, X., Sun, B., and Jin, Z. 2016. "A New Approach on Seismic Mortality Estimations Based on Average Population Density". *Earthquake Science*, 29(6), 337–344.

Zhu, X., Zhang, G., and Sun, B. 2019. "A Comprehensive Literature Review of the Demand Forecasting Methods of Emergency Resources from the Perspective of Artificial Intelligence." *Natural Hazards*, 97(1), 65–82.

AUTHOR BIOGRAPHIES

YUDI CHEN is a Ph.D. Candidate in the Department of Civil, Environmental & Infrastructure Engineering at George Mason University. His research interests include disaster management, infrastructure resilience, machine learning, and social media mining. Notably, he is focusing on the enhancement of infrastructure resilience through big data analytics. His e-mail address is ychen55@gmu.edu. His website is www.yudi-chen.com.

WENYING JI is an Assistant Professor in the Department of Civil, Environmental & Infrastructure Engineering, George Mason University. He received his Ph.D. in Construction Engineering and Management from the University of Alberta. He is an interdisciplinary scholar focused on the integration of advanced data analytics, complex system simulation, and construction management to enhance the overall performance of infrastructure systems. His e-mail address is wji2@gmu.edu. His website is http://mason.gmu.edu/~wji2/.