Understanding Post-Disaster Population Recovery Patterns

Supplementary Information

Takahiro Yabe[1], Kota Tsubouchi[2], Naoya Fujiwara[3,4], Yoshihide Sekimoto[4], Satish V. Ukkusuri[1*]

[1]Lyles School of Civil Engineering, Purdue University, Indiana, USA
[2]Yahoo Japan Corporation, Tokyo, Japan
[3]Graduate School of Information Sciences, Tohoku University, Sendai, Japan
[4]Institute of Industrial Science, University of Tokyo, Tokyo, Japan
*sukkusur@purdue.edu

Contents

List of Figures

List of Tables

1 Description of disasters

The disasters that are studied in this paper are: Hurricanes Irma and Maria, Tohoku Tsunami and Earthquake, Kumamoto Earthquake, and Kinugawa (Kinu River) Flood. The five disasters all vary in intensity, type of disaster, and location of occurrence, which gives us a unique opportunity to conduct a cross-comparative analysis of the disasters. In addition to the spatial distributions of housing damage rates shown in Figure 1a, we provide a brief overview of all disasters, including the date and location of occurrence, severity, and outcomes inflicted to human activities.

Hurricane Irma made landfall on Florida on September 10th as a category 4 hurricane and traversed through the Florida peninsula, spawning storm surge and causing major inland flooding. Especially in the Florida Keys, 25 percent of the homes were destroyed and 65 percent were damaged. Many homes and businesses suffered damage or destruction, with more than 65,000 structures damaged to some degree in West Central and Southwest Florida alone. The hurricane caused more than 7.7 million homes and businesses to be out of power in the entire state of Florida, and at least 134 fatalities were confirmed [4]. Total economic losses are estimated to be $ 55 billion [1].

The tropical cyclone Hurricane Maria developed on September 16, 2017 in the Atlantic Ocean to the northeast of South America, and made landfall in southeastern Puerto Rico on September 20, 2017 with wind speeds of 155 miles per hour. Damage to Puerto Rico was severe and widespread following the Hurricane, with heavy rainfall, flooding, storm surge, and high winds causing considerable damage [14]. Various infrastructure systems were heavily damaged, causing power outages and water shortages for the entire island for months [17]. Fatalities as a consequence of Maria are still under investigation, however the most recent estimates suggest between 793 to 8,498 excess deaths occurred following the storm [11]. Total economic losses are estimated to be $ 92 billion.

The magnitude 9.0 earthquake that occurred in the western Pacific induced a huge tsunami that hit the coastal areas of the Tohoku area in Japan. The tsunami reached the eastern coast of Honshu, Japan within a couple of minutes after the quake and spilled into the interior to a maximum distance of 10km [13, 16]. The loses incurred by the earthquake and tsunami together were extremely severe. According to the Japan National Police Agency, there were intotal 13.392 people dead nationwide and 15,133 missing. It resulted in more than 335,000 refugees, damaged more than 190,000 buildings and caused 4.4 million households without power. The disaster caused total economic losses of approximately $ 171-183 billion [13].

Kumamoto prefecture located in the Kyushu island of Japan was shook by a magnitude 6.5 earthquake on April 14th 2016, which was followed by a magnitude 7.3 earthquake on April 16th, 2016. The earthquakes caused significant loss, destroying 8050 buildings and damaging 24,147. The total number of fatalities was 69, while 1,747 were injured. More than 180,000 people evacuated immediately after the earthquake, causing mass population displacements in the Kyushu region. The total

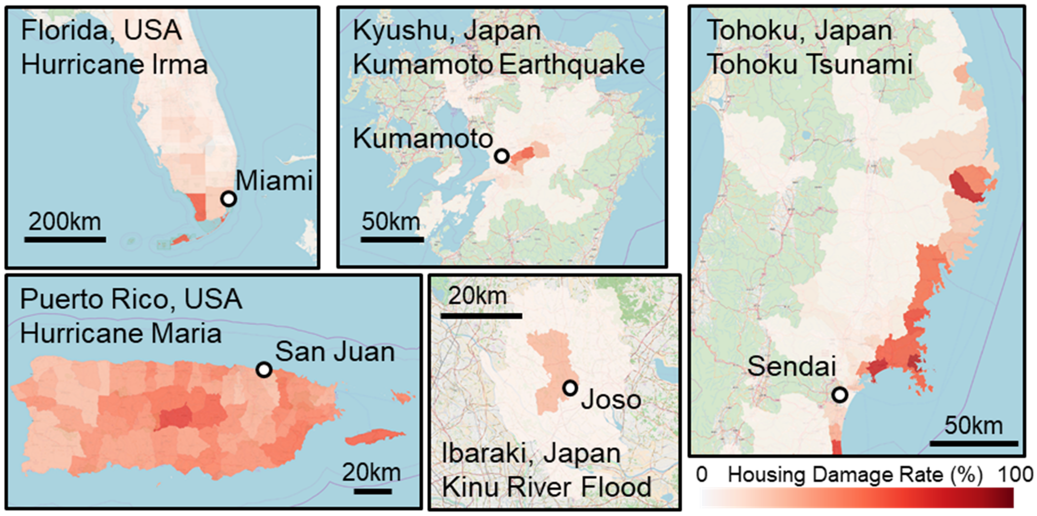

Figure S1: Housing damage rates due to the disaster. Local government units (LGUs) where housing damage was observed were classified as areas affected by the disaster. Mobile phone users who were estimated to be living in the affected LGUs were chosen for analysis.

economic loss was estimated to be $ 24-46 billion [8].

Heavy rain in the Northern part of the Kanto area caused severe flooding of the Kinugawa (Kinu River) on September 9th, 2015, which lasted for more than 3 days. More than 40 km^2 was flooded, causing 14 deaths and damage to over 22,000 houses in the Joso area. The total economic loss was estimated to be $ 2.9 billion [12].

2 Areas of study

For each disaster, we first define the areas of study based on the extent of damage caused by the disaster. The affected areas were defined as the set of local government units (LGUs) where housing damage was observed due to the disaster. LGUs refer to counties in Florida and Puerto Rico, and "shichoson (city/ward)" in Japan. Table S1 lists the names of all LGUs where housing damage was observed due to each disaster. 78 from Puerto Rico, 49 from Florida, 30 from Tohoku Tsunami, 33 from Kumamoto Earthquake, and 10 from Kinugawa Flood were included in the analysis. There are mainly 3 reasons to why we perform our analysis on the LGU scale. Firstly, due to the limitation in the number of mobile phone user samples, analysis at a further finer scale would yield statistically insignificant results especially in rural areas. Second, the LGU scale is the finest scale in which we can obtain socio-economic data in Japan, unlike the US where data is available on the census tract level

Table S1: Areas of study in the five disasters

Disaster	Local Government Units
Hurricane Maria (78)	Adjuntas, Aguada, Aguadilla, Aguas Buenas, Aibonito, Anasco, Arecibo, Arroyo, Barceloneta, Barranquitas, Bayamon, Cabo Rojo, Caguas, Camuy, Canovanas, Carolina, Catano, Cayey, Ceiba, Ciales, Cidra, Coamo, Comerio, Corozal, Culebra, Dorado, Fajardo, Florida, Guanica, Guayama, Guayanilla, Guaynabo, Gurabo, Hatillo, Hormigueros, Humacao, Isabela, Jayuya, Juana Diaz, Juncos, Lajas, Lares, Las Marias, Las Piedras, Loiza, Luquillo, Manati, Maricao, Maunabo, Mayaguez, Moca, Morovis, Naguabo, Naranjito, Orocovis, Patillas, Penuelas, Ponce, Quebradillas, Rincon, Rio Grande, Sabana Grande, Salinas, San German, San Juan, San Lorenzo, San Sebastian, Santa Isabel, Toa Alta, Toa Baja, Trujillo Alto, Utuado, Vega Alta, Vega Baja, Vieques, Villalba, Yabucoa, Yauco
Hurricane Irma (49)	Alachua, Baker, Bradford, Brevard, Broward, Charlotte, Citrus, Clay, Collier, Columbia, DeSoto, Dixie, Duval, Flagler, Gilchrist, Glades, Hamilton, Hardee, Hendry, Hernando, Highlands, Hillsborough, Indian River, Lafayette, Lake, Lee, Levy, Manatee, Marion, Martin, Miami-Dade, Monroe, Nassau, Okeechobee, Orange, Osceola, Palm Beach, Pasco, Pinellas, Polk, Putnam, St. Johns, St. Lucie, Sarasota, Seminole, Sumter, Suwannee, Union, Volusia
Tohoku Tsunami (30)	Hachinohe, Misawa, Oirase, Miyako, Ohfunato, Kuji, Rikuzentakata, Kamaishi, Ohtsuchi, Yamada, Iwaizumi, Tanohata, Noda, Hirono, Sendai-Miyagino, Sendai-Wakabayashi, Ishinomaki, Shiogama, Kesennuma, Natori, Tagajo, Iwanuma, E-Matsushima, Watari, Yamamoto, Matsushima, Shichigahama, Rifu, Onagawa, Minamisanriku
Kumamoto Earthquake (33)	Yanagawa, K-Chuo, Yatsushiro, Tamana, Yamaga, Kikuchi, Uto, Kamiamakusa, Uki, Aso, Koshi, Misato, Gyokuto, Ohzu, Kikuyoh, Minami-Oguni, Oguni, Ubuyama, Takamori, Nishihara, Minami-Aso, Mifune, Kashima, Mashiki, Kosa, Yamato, Hikawa, Ashikita, Beppu, Taketa, Yufu, Kokonoe, Shiiba
Kinugawa Flood (10)	Shimotsuma, Joso, Toride, Tsukuba, Moriya, Bando, Tsukuba-Mirai, Yachiyo, Noda, Kashiwa

through the American Community Survey. Third, government agencies often make policy decisions on the LGU scale, thus insights on that spatial scale would provide decision makers with relevant and useful insights.

In each of these affected LGUs, information on housing damage rates were collected. For disasters in the US, "housing damage rate" of a given county refers to the rate of houses approved for the Individuals and Households Program of FEMA in each LGU [7]. For disasters in Japan, it refers to the rate of houses classified as "totally destroyed" or "half destroyed" by the Cabinet Office of Japan (COJ) [5]. Both datasets are publicly accessible. Figure S1 shows the housing damage rates in all of the five disasters. Out of all the LGUs in Table S1, all 78 LGUs in Puerto Rico, 6 LGUs in Florida, 21 LGUs in Tohoku, 6 LGUs in Kumamoto, and 1 LGU in Kinugawa experienced extensive damage (housing damage rates of more than 10%).

3 Analysis of mobile phone data

3.1 Data description

Mobile phone location data for the five disasters were provided by 3 different companies in Japan and the US. Location data were collected by Yahoo Japan Corporation[1] for Kumamoto Earthquake and Kinugawa Flood, by Zenrin Data Com[2] for Tohoku Tsunami and Earthquake, and Safegraph[3] for Hurricanes Irma and Maria. All companies obtained the location information (time, longitude, latitude) of mobile phones via the Global Positioning Satellite (GPS) system, as shown in Table S2. The example ID is masked to protect privacy issues. GPS data were obtained from mobile phones of individuals who agreed to provide their location data for research purposes, and all information were anonymized to protect the security of users. Mobile phone location data are proprietary data owned by private companies. Although such data are not available for open access due to the users' privacy, we will obtain permission to post processed data that are sufficient to reproduce the results obtained in this study.

Table S3 shows the statistics of the mobile phone datasets. For each disaster, the area of study included LGUs that experienced any housing damage, as shown in Figure S1. Cell phone data for 6 months before and after the disaster date was observed for all disasters. Each user was observed at high frequency in all datasets, although there were differences in the average observations points per user per day. The average observations vary across datasets (33 to 97), however, given that we only observe the location where each user is staying each day over night, the granularity of these datasets

[1] https://www.yahoo.co.jp/
[2] http://www.zenrin-datacom.net/toppage
[3] https://www.safegraph.com/

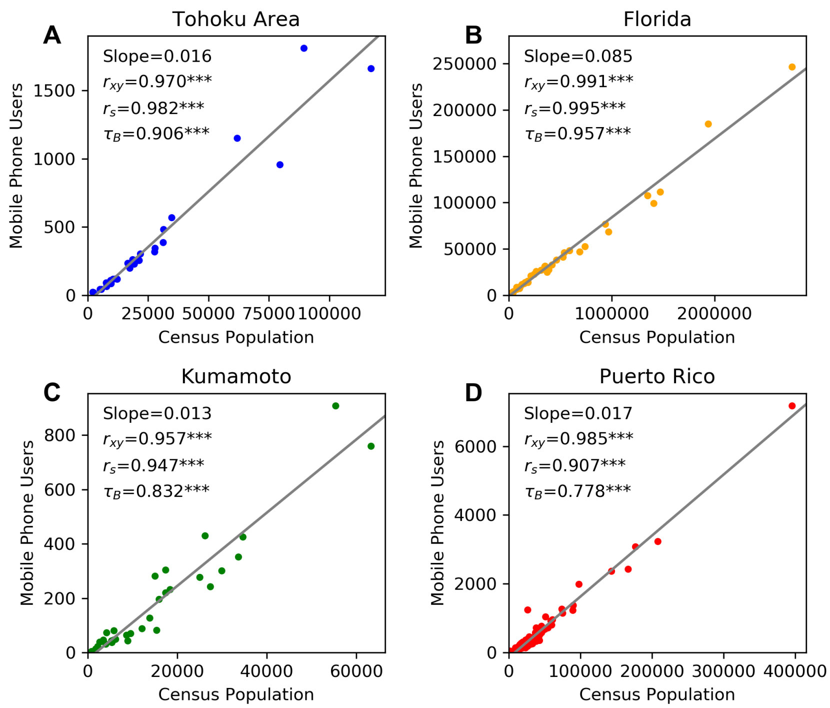

Figure S2: Representativeness of mobile phone users The correlation between the census population and the number of mobile phone users observed in each local government unit (counties in US, cities in Japan) is high. The slopes (sample rates) in each dataset is around 1.5% except for Florida, with 8.5%.

Table S2: Fields and content of mobile phone location data

Field	Description	Sample
User ID	Hashed unique user ID for each mobile phone user	###HA9K8FH6R
Timestamp	Unix time of observation	1505771147
Longitude	Longitude coordinate of the GPS observation	-82.0836
Latitude	Latitude coordinate of the GPS observation	27.4311

are high enough for our analyses. The total number of users in each dataset and the number of users in the affected areas are shown in brackets.

For our macroscopic study on recovery trends, it is important to account for the spatial bias in mobile phone users. To ensure that the data is not spatially biased, we plot the number of samples against census populations for each LGU in Figure S2 for all datasets. We observe that for all datasets, the number of mobile phone samples are highly correlated with census population for each county, with Pearson's correlation coefficient greater than 0.95, and also with high rank correlation (Spearman's correlation and Kendall's Tau). Thus, we conclude that the mobile phone data has little bias in the sample rates across different counties and that is representative of the entire population.

3.2 Home location estimation

Using the observed mobile phone location data, the home location of all users were estimated. It is well known that human trajectories show a high degree of temporal and spatial regularity, each individual having a significant probability to return to a few highly frequented locations, including his/her home location [9]. Due to this characteristic, it has been shown that home locations of individuals can be detected with high accuracy by clustering the individual's stay point locations over night [3]. Home locations of each individual was detected by applying mean-shift clustering to the nighttime staypoints (observed between 8PM and 6AM), weighted by the duration of stays in each location [2, 10]. Mean shift clustering was implemented using the scikit-learn package on Python[4] .

3.3 Fitting population recovery trends

A user was classified to be displaced if the user was estimated to have stayed outside his/her estimated home LGU. The displacement rate on a given day was calculated by dividing the number of displaced users observed on that day by the total number of users.

[4] `http://scikit-learn.org/stable/modules/generated/sklearn.cluster.MeanShift.html`

Table S3: Statistics of GPS data for all disasters. For all disasters, GPS location data of affected individuals were observed for approximately 6 months, including days before and after the disaster. All datasets had more than 30 datapoints per day for each individual on average, allowing us to accurately track where each individual stayed every night after the disaster.

Disaster	Main Study Area	Observed Period	Users (affected)	Observations (/day/user)
Hurricane Maria	Puerto Rico	2017/9/1-2018/3/15	53,511 (53,511)	82.8
Hurricane Irma	Florida, USA	2017/9/1-2018/3/1	1,730,326 (1,599,370)	97.0
Tohoku Tsunami	Tohoku, Japan	2011/3/1-2011/9/1	68,416 (10,697)	33.4
Kumamoto Earthquake	Kumamoto, Japan	2016/4/1-2016/10/1	80,933 (5,944)	40.7
Kinugawa Flood	Ibaraki, Japan	2015/9/1-2016/3/1	2,580 (437)	46.0

To capture the short term fluctuations in the population recovery trends, the raw observations of displacement rates were de-noised using Gaussian Process Regression. Gaussian Process Regression (GPR) is a non-parametric probabilistic model for denoising and regression [15]. Gaussian Processes (GPs) are extensions of multivariate Gaussian distributions to infinite dimensionality. GPs assume that values observed at $t = t_i$ and $t = t_j$ are jointly Gaussian with zero mean and covariance given by a covariance function $k(t_i, t_j)$. In this model, we use the squared exponential covariance function and we assume that the observed values y have i.i.d. Gaussian noise with variance σ_n^2 added on, shown in the following equation:

$$cov(f(t_i), f(t_j)) = k(t_i, t_j) = \exp\left(\frac{1}{2l^2}|t_i - t_j|^2\right) + \sigma_n^2 \delta_{i,j} \tag{1}$$

where δ_{ij} is a Kronecker delta which is 1 if $i = j$ and zero otherwise. In the GPR model, the hyperparameters are the length scale l and the scale of the Gaussian noise of the observed values σ_n. The model chooses the hyperparameters and covariances directly from the training data. To obtain such optimal hyperparameters, the log marginal likelihood $L(\theta)$, shown below, is maximized with respect to hyperparameters and noise level $\theta = l, \sigma_n$.

$$L = \log p(y|t) = -\frac{1}{2} y^T (K + \sigma_n^2 I)^{-1} - \frac{1}{2} \log |K + \sigma_n^2 I| - \frac{n}{2} \log 2\pi \tag{2}$$

The minimization of $L(\theta)$ is solved by conjugate gradients method [6]. To implement the GPR model, we used the package available on scikit learn and implemented the model using Python codes[5].

To capture the general trend of population recovery, the raw observations were fitted using a negative exponential function $D(t) = (D_0 - D_{160}) \exp(-\frac{t}{\tau}) + D_{160}$, where D_0, D_{160}, and τ denote the displacement rates on day 0, day 160, and recovery time parameter, respectively. The three parameters (D_0, D_{160}, and τ) were estimated using the Powell method, which is a widely used optimization algorithm. The robustness of the estimated exponential function parameters to temporal scales are shown in Figure S3. We tested fitting the function with the i) raw observations, moving average estimate of ii) 3 days window, iii) 7 days window, and iv) 14 day window. The parameter estimates and standard errors are shown in each figure. We can observe that the parameter estimates are robust against data variability, with very small standard error compared to the mean estimates. Moreover, the Pearson correlation between the daily observed displacement rates and the estimated displacement rates from the fitted exponential curves are high: Hurricane Maria: 0.85, Tohoku Tsunami: 0.95, Kinugawa Flood: 0.93, Hurricane Irma: 0.90, Kumamoto Earthquake: 0.73. Further, the fitted negative exponential functions were normalized $\tilde{D}(t) = \frac{D(t) - D_{160}}{D_0 - D_{160}}$, which should collapse to the same curve $e^{-\frac{t}{\tau}}$ if all curves follow a negative exponential function.

[5] http://scikit-learn.org/stable/modules/gaussian_process.html

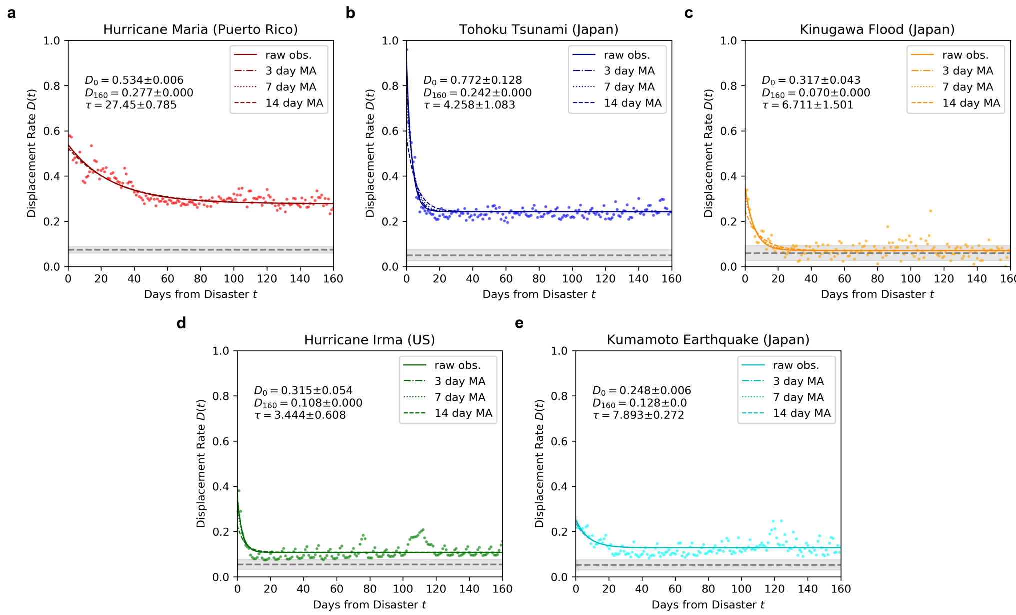

Figure S3: Robustness of exponential fit. Fitting results with i) raw observations, moving average estimate of ii) 3 days window, iii) 7 days window, and iv) 14 day window, for each disaster. The parameter estimates and standard errors are shown in each figure.

4 Regression Analysis

4.1 Models

To understand the effect of the independent variables on the displacement rates and the speed of recovery, we apply a generalized linear regression modeling framework. Because the displacement rates are probabilities, $0 < D(t) < 1$ holds for any t. Therefore, we apply a logit link function to the displacement rates in the regression model. Similarly, because the recovery times take only positive values ($0 < \tau$), we apply a log link function to the speed parameter. Equations (3) and (4) show the generalized linear regression model where β are the regression coefficients, x are the independent variables explained in the next section, and $\epsilon \sim \mathcal{N}(0, \sigma^2)$ is the error term. The model parameters are estimated via maximum likelihood estimation.

$$\log\left(\frac{D(t)}{1 - D(t)}\right) = \beta^T x + \epsilon \tag{3}$$

$$\log(\tau) = \beta^T x + \epsilon \tag{4}$$

4.2 Independent Variables

4.2.1 Socio-economic data

In the regression analyses of population recovery, socio-economic data were used as independent variables. Table S4 lists the independent variables used in the study. The number of households and median income data were log-transformed to truncate the skewness of the data (some cities such as Miami-Dade County have extremely large values). For Florida and Puerto Rico, population data were obtained from the US National Census[6], and median income data were obtained from the American Community Survey[7]. Similarly, for Japanese LGUs, population and income data were obtained from the Statistics Bureau[8] of the Ministry of Internal Affairs and Communications of Japan.

The proximity of city i to large cities is calculated by $d_p(i) = \frac{\sum_{j \in S(i)} N_j}{N_i}$, where N_i is the number of households in city i, and $S(i)$ is the set of cities that can be reached within 1 hour by vehicles from city i. d_p would be large for small cities that have large cities around it, and small for more isolated cities. For cities with similar population levels, d_p would be proportional to the total population of surrounding cities. Similarly, we propose the proximity to wealthy cities by using the median income value instead of the household number in the previous equation. This value would be large if the origin city has a relatively low income and it is surrounded by wealthier cities nearby. Note that these two complex variables capture not only the characteristics of the origin city, but that of the receptor cities.

4.2.2 Power damage

Figure S4 compares the recovery process of power outages in the four major disasters. The data of Puerto Rico were collected from the website `StatusPR`[9], which is a government operated website that showed the recovery status of Puerto Rico after the Hurricane. Data of Hurricane Irma were collected from the Florida Division of Diaster Management[10]. Power outage information in the Japanese disasters were collected from the utility companies. All data are publicly available and accessible. Figure S4 highlights the quick recovery of power in Florida and Japan (around 10 days until full recovery) compared to Puerto Rico after Hurricane Maria (more than 80 days and still around 60% availability).

[6] https://www.census.gov/

[7] https://www.census.gov/programs-surveys/acs

[8] https://www.stat.go.jp/

[9] http://status.pr/

[10] https://www.floridadisaster.org/

Table S4: Descriptions, variable transformations, and sources of socio-economic data

Variable	Description	Trans.	Data Source
Households	Number of households in LGU	log	Statistics Bureau (Japan) ACS* (USA)
Median income	Median household income in LGU	log	Statistics Bureau (Japan) ACS* (USA)
Housing damage	Rate of households damaged in LGU	-	Cabinet Office (Japan) FEMA (USA)
Proximity to large cities	$\frac{\sum_{j\in S(i)} N_j}{N_i}$ N_i: households in LGU i $S(i)$: nearby cities from i	-	-
Proximity to wealthy cities	$\frac{\sum_{j\in S(i)} (MI)_j}{(MI)_i}$ $(MI)_i$: income in LGU i $S(i)$: nearby cities from i	-	-
Infrastructure recovery	Days until power recovery in LGU	-	Local government reports

*ACS: American Communty Survey

4.3 Regression results

Table S5 shows the correlations among the independent variables. Correlations between all pairs of variables in all disasters were not significantly high, thus we included them in the regression analysis. Infrastructure recovery time was excluded in the model for estimating D_0, since this information would not be available on day 0. Table S6, Table S7, and Table S8 show the regression results of the best models for $\log\left(\frac{D_0}{1-D_0}\right)$, $\log\left(\frac{D_{160}}{1-D_{160}}\right)$, and $\log(\tau)$, respectively, for each disaster in each column. The set of independent variables for the best model for each disaster was chosen based on the lowest AIC value and statistical significance ($p < 0.1$). The prediction results from the best models are plotted in Figure 3c and 3d in the manuscript against the true values. For the initial displacement rates $\log\left(\frac{D_0}{1-D_0}\right)$, housing damage rates were the most significant variables as expected, (also shown in Figure 4 in the manuscript), with the exception of Puerto Rico. In Puerto Rico, median income values and proximity to wealthier cities were a better predictor of initial evacuation. For the long term ($t = 160$) displacement rates, the significant variables varied across different disasters. We see that median income has a positive association with long term displacements, which indicate that people with more income were able to evacuate and migrate to places that were not hit by the disaster. Recovery speed had the lowest predictability out of all objective variables. Again, the significant

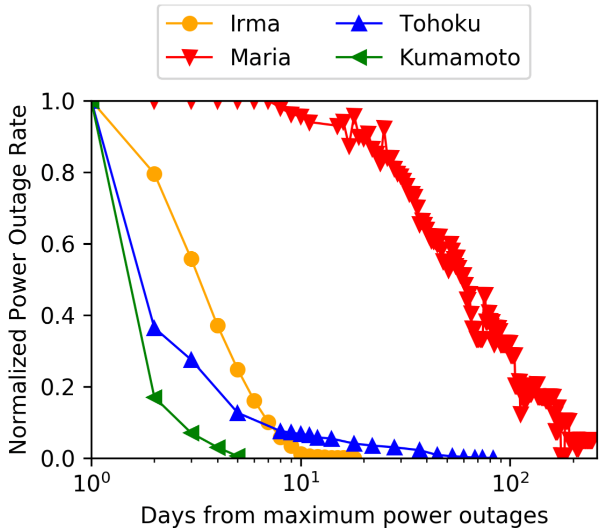

Figure S4: Recovery speed of power outages. Comparison of recovery from power outages across the four major disasters. For Hurricane Irma (Florida, USA), Tohoku Tsunami (Tohoku, Japan), and Kumamoto Earthquake (Kyushu, Japan), it took less than 10 days from the disaster day to restore more than 90% if the power outages. On the other hand, it took more than 200 days after Hurricane Maria for full recovery of power in Puerto Rico.

variables were different across different disasters. The findings and insights are discussed in the main manuscript in detail. Table S9 shows the regression results for various timings. We can observe that the effects of key variables in each disaster stay similar in general over time, with minor differences across timings.

4.4 Case study of intra-LGU variability in Miami-Dade County

In this study, regression analysis on the recovery parameters were conducted on the local government unit (LGU) scale (i.e. counties in Puerto Rico and Florida, and city/ward/towns in Japan). This was mainly due to the limitation in the number of mobile phone user samples; analysis at a further finer scale would yield statistically insignificant results especially in rural areas. Moreover, the LGU scale is the finest scale in which we can obtain socio-economic data in Japan, unlike the US where data is available on the census tract level through the American Community Survey.

Despite such limitations in the data, the intra-LGU variability in the socio-economic characteristics is of great importance in understanding spatial heterogeneity in population displacement and recovery patterns. Here we show a case study of Miami-Dade county after Hurricane Irma, which does not suffer from either of the data limitations; a large city with enough mobile phone user samples, located in the US with census-tract level socio-economic data. We were able to collect the number

Tohoku Tsunami	HDR	log(NH)	log(MI)	PLC	PWC	IRT
Housing Damage Rate	1					
log(Number of Households)	-0.22	1				
log(Median Income)	-0.14	0.51	1			
Proximity to Large Cities	0.22	-0.18	0.39	1		
Proximity to Wealthy Cities	0.7	0.082	-0.0026	-0.064	1	
Infrastructure Recovery Time	0.18	0.31	0.6	0.68	0.039	1

Hurricane Irma	HDR	log(NH)	log(MI)	PLC	PWC	IRT
Housing Damage Rate	1					
log(Number of Households)	0.095	1				
log(Median Income)	-0.036	0.53	1			
Proximity to Large Cities	-0.058	-0.65	-0.26	1		
Proximity to Wealthy Cities	0.72	-0.082	0.078	-0.077	1	
Infrastructure Recovery Time	-0.29	-0.37	-0.52	0.34	-0.4	1

Hurricane Maria	HDR	log(NH)	log(MI)	PLC	PWC	IRT
Housing Damage Rate	1					
log(Number of Households)	-0.34	1				
log(Median Income)	-0.25	0.48	1			
Proximity to Large Cities	0.28	-0.31	-0.0093	1		
Proximity to Wealthy Cities	0.19	0.061	-0.015	-0.18	1	
Infrastructure Recovery Time	0.15	0.39	0.14	0.59	-0.12	1

Kumamoto Earthquake	HDR	log(NH)	log(MI)	PLC	PWC	IRT
Housing Damage Rate	1					
log(Number of Households)	-0.22	1				
log(Median Income)	0.075	0.65	1			
Proximity to Large Cities	0.53	-0.44	-0.11	1		
Proximity to Wealthy Cities	0.48	-0.15	-0.072	0.064	1	
Infrastructure Recovery Time	0.44	0.018	0.18	0.6	0.092	1

Table S5: Correlation between the variables used for regression.

	Dependent Variable: $\log(\frac{D_0}{1-D_0})$			
	Tohoku Tsunami	Hurricane Irma	Hurricane Maria	Kumamoto Eq.
Intercept	-8.21** (3.46)	0.18 (0.78)	20.17*** (5.39)	-40.55*** (6.32)
Housing Damage Rate	7.35*** (1.27)	4.71*** (1.00)	-	4.37*** (0.81)
log(Households)	0.65* (0.34)	-0.12** (0.05)	-	-
log(Median Income)	-	-	-1.94*** (0.55)	3.75*** (0.61)
Proximity to Large Cities	-	-	-	-
Proximity to Wealthy Cities	-	-0.03* (0.02)	-0.018*** (0.006)	-
Observations	30	49	78	33
Adjusted R^2	0.51	0.44	0.22	0.59
AIC	126.67	79.91	233.60	102.32
Significance F	<0.01***	<0.01***	<0.01***	<0.01***

Table S6: Regression models with for displacement rates on day 0 for all disasters

	Dependent Variable: $\log(\frac{D_{160}}{1-D_{160}})$			
	Tohoku Tsunami	Hurricane Irma	Hurricane Maria	Kumamoto Eq.
Intercept	-16.36** (6.72)	-0.57** (0.28)	-6.29*** (1.87)	-23.54*** (7.49)
Housing Damage Rate	2.01*** (0.51)	0.91** (0.43)	1.06** (0.46)	-
log(Households)	-0.34*** (0.11)	-0.096*** (0.023)	-	-
log(Median Income)	1.76** (0.70)	-	0.53*** (0.186)	1.99*** (0.72)
Proximity to Large Cities	-0.0036** (0.00172)	-	-	-
Proximity to Wealthy Cities	-	-	-	-
Infrastructure recovery speed	-0.014* (0.0069)	-	-0.006* (0.0032)	-0.27** (0.12)
Observations	30	49	78	33
Adjusted R^2	0.48	0.31	0.12	0.17
AIC	42.02	1.648	60.71	118.84
Significance F	$<$ 0.01***	$<$ 0.01***	$<$ 0.01***	$<$ 0.01***

Table S7: Regression results for displacement rates on day 160

	Dependent Variable: $\log(\tau)$			
	Tohoku Tsunami	Hurricane Irma	Hurricane Maria	Kumamoto Eq.
Intercept	3.71*** (0.36)	1.98*** (0.21)	1.19** (0.53)	36.28*** (11.89)
Housing Damage Rate	-	-	3.47** (1.34)	-
log(Households)	-	-	-	-
log(Median Income)	-	-	-	-3.143*** (1.15)
Proximity to Large Cities	-	-0.0099* (0.0050)	-0.0075*** (0.0027)	-
Proximity to Wealthy Cities	-0.15*** (0.039)	-	-	-
Infrastructure recovery speed	-	-	-	-
Observations	30	49	78	33
Adjusted R^2	0.33	0.058	0.19	0.13
AIC	80.078	168.19	223.511	140.92
Significance F	< 0.01***	< 0.1**	< 0.01***	< 0.01***

Table S8: Regression results for recovery speed

Tohoku Tsunami	$t=0$	10	20	30	60	160
Intercept	-8.21**	-30.79***	-39.80***	-28.37***	-21.92***	-16.36**
Housing Damage	7.35***	4.34***	4.38***	3.45***	2.88***	2.01***
log(Households)	0.65*	-0.37***	-0.32*	-0.30***	-0.36***	-0.34***
log(Median Income)	-	3.13***	3.95***	2.84***	2.29***	1.76**
Proximity to Large Cities	-	-0.0069**	-0.011***	-0.0069***	-0.0056***	-0.0036**
Infrastructure recovery	-	-0.0147*	-0.025**	-0.019***	-0.018***	-0.014*
Adjusted R^2	0.51	0.60	0.72	0.78		0.48

Hurricane Irma	$t=0$	10	20	30	60	160
Intercept	0.18	0.17	0.29	-4.04*	-0.42	-0.57**
Housing Damage Rate	4.71***	2.14***	0.67*	0.72*	-	0.91**
log(Households)	-0.12**	-0.16***	-0.16***	-0.14***	-0.11***	-0.096***
log(Median Income)	-	-	-	0.37*	-	-
Proximity to Wealthy Cities	-0.03*	-	-	-	-	-
Adjusted R^2	0.44	0.51	0.46	0.48	0.38	0.31

Hurricane Maria	$t=0$	10	20	30	60	90	120	160
Intercept	20.17***	4.73***	4.28***	3.22***	-3.28*	-4.93***	-5.21***	-6.29***
Housing Damage Rate	-	-	-	-	1.26***	0.91**	-	1.06**
log(Households)	-	-0.46***	-0.43***	-0.35***	-0.22***	-0.10**	-0.27***	-
log(Median Income)	-1.94***	-	-	-	0.44**	0.49***	0.72***	0.53***
Proximity to Wealthy Cities	-0.018***	-	-	-	-	-	-	-
Infrastructure recovery speed	-	-	-	-	-	-	-	-0.006*
Adjusted R^2	0.22	0.26	0.32	0.29	0.31	0.20	0.17	0.12

Kumamoto Earthquake	$t=0$	10	20	30	60	160
Intercept	-40.55***	-34.02***	-27.18***	-23.51***	-24.50***	-23.54***
Housing Damage Rate	4.37***	4.16***	3.65***	3.16***	2.87***	-
log(Median Income)	3.75***	3.02***	2.35***	1.99***	2.09***	1.99***
Infrastructure recovery	-	-	-	-	-	0.27**
Adjusted R^2	0.59	0.52	0.44	0.36	0.34	0.21

Table S9: Regression results for displacement rates on various timepoints $\log(\frac{D(t)}{1-D(t)})$

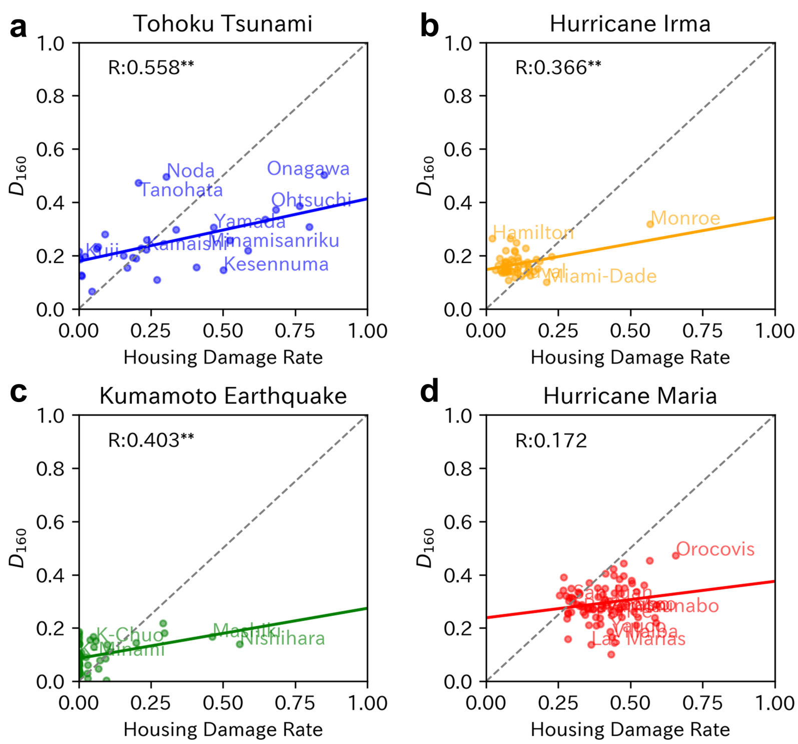

Figure S5: Long term displacement rates D_{160} plotted against housing damage rates for each community in the four major disasters. For all disasters, the correlation between housing damage rates and long term displacement rates are weaker than that with initial displacement rates, implying that recovery of communities are governed by more factors in addition to housing damage rates. In some communities such as Noda and Tanohata after the Tohoku Tsunami (a), displacement rates even increased compared to initial displacement rates after the Tohoku Tsunami.

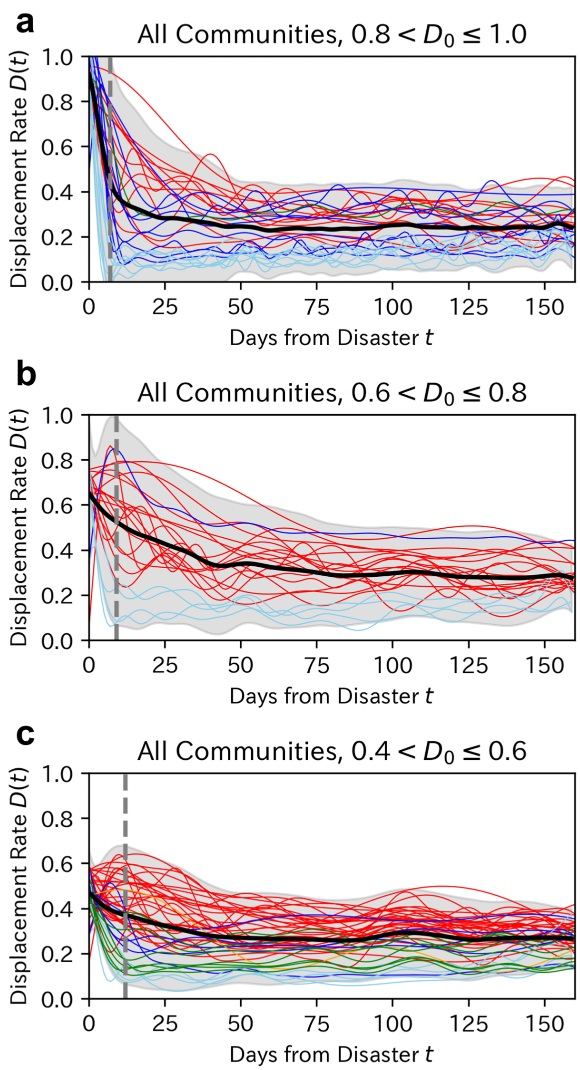

Figure S6: Heterogeneity of community recovery patterns given similar initial displacement rates. Recovery patterns $D(t)$ of each community plotted against number of days after the disaster for all disasters. Community recovery patterns are divided into three groups depending on the initial displacement rates D_0. Colors indicate the disaster, similar to previous figures. Broad black line shows the average values of the recovery patterns in the group. The results show that long term recovery outcomes are heterogeneous despite similar initial displacement conditions.

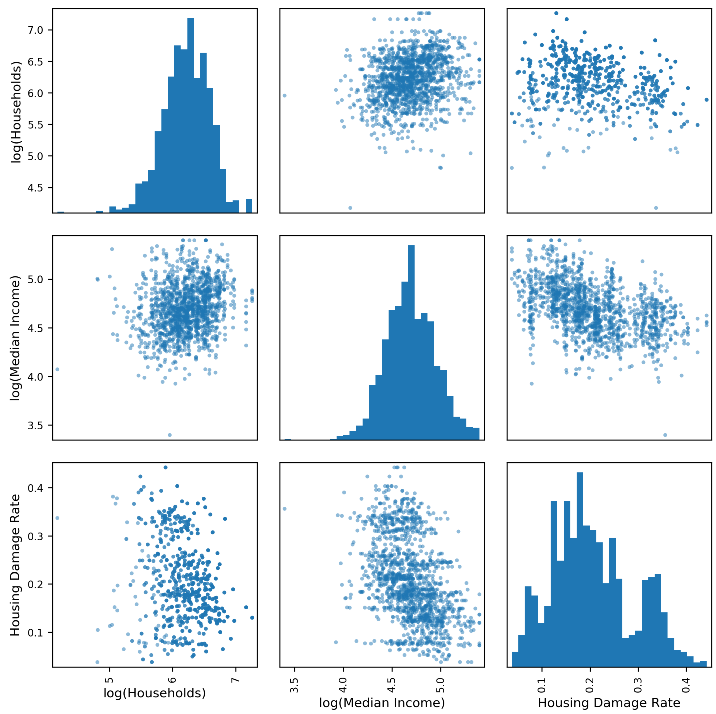

Figure S7: Correlation between census-tract level features in Miami-Dade

of households, median income, and housing damage rates in each census tract after Hurricane Irma. Figure S7 shows the correlations and scatter plots of each of the variables. The number of households and median income resemble a log-normal distribution (note the log-scale). Regression analysis were performed on displacement rates on dats $t = 0, 5, 10, 20, 60$ in a similar manner as the previous experiments, using the three aforementioned variables. Table S10 lists the regression coefficients, their significance, and the adjusted R^2 for the different timepoints. The most striking difference between the county-level analysis and the census tract level analysis was that housing damage rate had a negative correlation with displacement rates at all times on the census tract level. This was due to the strong negative correlation between median income levels and housing damage rates, and the fact that wealthier people were able to evacuate more from the hurricane. Figure S8 shows the observed and predicted displacement rates on the census tract level for 4 different timepoints ($t = 0, 5, 10, 60$). We can observe the low correlation coefficient in all timepoints compared to the county-level analysis (e.g. Figure 3), and that we are not able to explain the census tract level heterogeneity very well. This may be due to the small sample size in each census tract, which is an average of 40.7 (Figure S9). A more robust estimation using sparse mobile phone user samples would be needed to give better estimations on the census tract granularity.

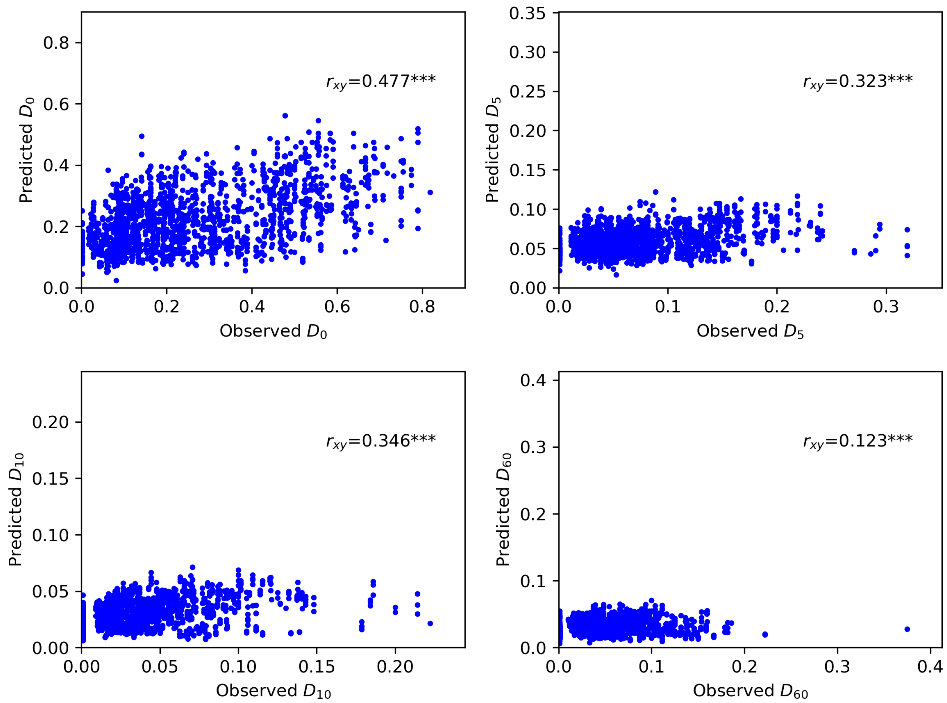

Figure S8: Prediction of displacement rates in Miami-Dade

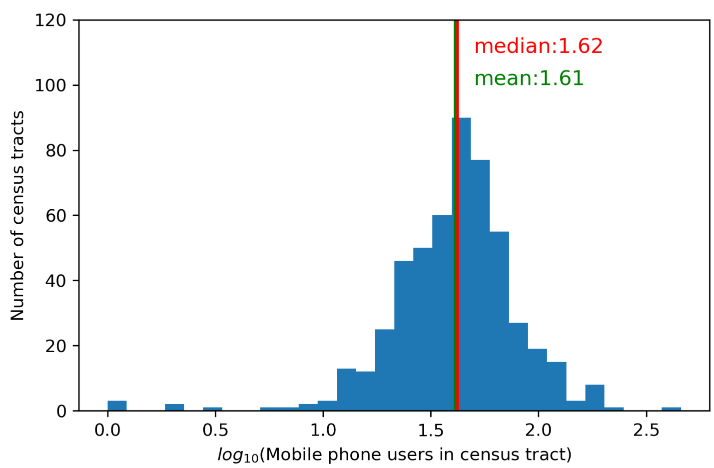

Figure S9: Histogram of mobile phone user samples in Miami-Dade census tracts

References

[1] AON Analytics. Hurricane irma event recap report. `http://thoughtleadership.aonbenfield.com/Documents/20180328-ab-if-hurricane-irma-recap.pdf`, 2018. Online; accessed 17 September 2018.

[2] Daniel Ashbrook and Thad Starner. Using gps to learn significant locations and predict movement across multiple users. *Personal and Ubiquitous computing*, 7(5):275–286, 2003.

[3] Francesco Calabrese, Giusy Di Lorenzo, Liang Liu, and Carlo Ratti. Estimating origin-destination flows using opportunistically collected mobile phone location data from one million users in boston metropolitan area. *IEEE Pervasive Computing*, 10(4):36–44, 2011.

[4] John P Cangialosi, Andrew S Latto, and Robbie Berg. National hurricane center tropical cyclone report: Hurricane irma. *National Hurricane Center*, 2018.

[5] COJ. Disaster reports for disasters in japan (japanese). `http://www.bousai.go.jp/`, 2018. Accessed: 2018-09-07.

[6] Mark Ebden et al. Gaussian processes for regression: A quick introduction. *The Website of Robotics Research Group in Department on Engineering Science, University of Oxford*, 2008.

[7] FEMA. Fema housing assistance program data. `https://www.fema.gov/media-library/assets/documents/34758`, 2018. Accessed: 2018-09-07.

[8] Katsuichiro Goda, Grace Campbell, Laura Hulme, Bashar Ismael, Lin Ke, Rebekah Marsh, Peter Sammonds, Emily So, Yoshihiro Okumura, Nozar Kishi, et al. The 2016 kumamoto earthquakes: cascading geological hazards and compounding risks. *Frontiers in built environment*, 2:19, 2016.

[9] Marta C Gonzalez, Cesar A Hidalgo, and Albert-Laszlo Barabasi. Understanding individual human mobility patterns. *nature*, 453(7196):779, 2008.

[10] Hiroshi Kanasugi, Yoshihide Sekimoto, Mori Kurokawa, Takafumi Watanabe, Shigeki Muramatsu, and Ryosuke Shibasaki. Spatiotemporal route estimation consistent with human mobility using cellular network data. In *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2013 IEEE International Conference on*, pages 267–272. IEEE, 2013.

[11] Nishant Kishore, Domingo Marqués, Ayesha Mahmud, Mathew V Kiang, Irmary Rodriguez, Arlan Fuller, Peggy Ebner, Cecilia Sorensen, Fabio Racy, Jay Lemery, et al. Mortality in puerto rico after hurricane maria. *New England journal of medicine*, 2018.

[12] Transport Ministry of Land, Infrastructure and Tourism of Japan. Economic losses of water-related disasters in 2015. `http://www.mlit.go.jp/common/001176247.pdf`, 2017. Online; accessed 17 September 2018.

[13] Okada Norio, Tao Ye, Yoshio Kajitani, Peijun Shi, and Hirokazu Tatano. The 2011 eastern japan great earthquake disaster: Overview and comments. *International Journal of Disaster Risk Science*, 2(1):34–42, 2011.

[14] Richard J Pasch, Andrew B Penny, and Robbie Berg. National hurricane center tropical cyclone report: Hurricane maria. *TROPICAL CYCLONE REPORT AL152017, National Oceanic And Atmospheric Administration and the National Weather Service*, pages 1–48, 2018.

[15] Carl Edward Rasmussen and Christopher KI Williams. Gaussian processes for machine learning. 2006. *The MIT Press, Cambridge, MA, USA*, 38:715–719, 2006.

[16] Anawat Suppasri, Shunichi Koshimura, Kentaro Imai, Erick Mas, Hideomi Gokon, Abdul Muhari, and Fumihiko Imamura. Damage characteristic and field survey of the 2011 great east japan tsunami in miyagi prefecture. *Coastal Engineering Journal*, 54(01):1250005, 2012.

[17] Infrastructure Security United States Department of Energy and Energy Restoration. Hurricanes nate, maria, irma, and harvey situation reports. `https://www.energy.gov/ceser/downloads/hurricanes-nate-maria-irma-and-harvey-situation-reports`, 2017. Online; accessed 17 September 2018.

Hurricane Irma; Miami-Dade	$t = 0$	5	10	20	60
Intercept	-8.69***	-7.43***	-7.17***	-10.48***	-9.11***
Housing Damage Rate	-3.34***	-1.32***	-3.97***	-1.31***	-2.02***
log(Households)	0.08**	0.07**	0.14***	0.28***	0.17***
log(Median Income)	0.63**	0.35***	0.21***	0.35***	0.32***
Adjusted R^2	0.19	0.072	0.12	0.12	0.06

Table S10: Regression results for displacement rates on various timepoints of census tracts in Miami-Dade County $\log(\frac{D(t)}{1-D(t)})$