INTERFACE

royalsocietypublishing.org/journal/rsif

Research

Cite this article: Yabe T, Tsubouchi K, Fujiwara N, Sekimoto Y, Ukkusuri SV. 2020 Understanding post-disaster population recovery patterns. *J. R. Soc. Interface* **17**: 20190532.
http://dx.doi.org/10.1098/rsif.2019.0532

Received: 28 July 2019
Accepted: 21 January 2020

Subject Category:
Life Sciences–Physics interface

Subject Areas:
biophysics

Keywords:
human mobility, disaster resilience, population recovery, mobile phone data

Author for correspondence:
Satish V. Ukkusuri
e-mail: sukkusur@purdue.edu

Electronic supplementary material is available online at https://doi.org/10.6084/m9.figshare.c.4834668.

THE ROYAL SOCIETY PUBLISHING

Understanding post-disaster population recovery patterns

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

[1]Lyles School of Civil Engineering, Purdue University, West Lafayette, IN, 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

TY, 0000-0001-8967-1967; SVU, 0000-0001-8754-9925

Despite the rising importance of enhancing community resilience to disasters, our understandings on when, how and why communities are able to recover from such extreme events are limited. Here, we study the macroscopic population recovery patterns in disaster affected regions, by observing human mobility trajectories of over 1.9 million mobile phone users across three countries before, during and after five major disasters. We find that, despite the diversity in socio-economic characteristics among the affected regions and the types of hazards, population recovery trends after significant displacement resemble similar patterns after all five disasters. Moreover, the heterogeneity in initial and long-term displacement rates across communities in the three countries were explained by a set of key common factors, including the community's median income level, population, housing damage rates and the connectedness to other cities. Such insights discovered from large-scale empirical data could assist policymaking in various disciplines for developing community resilience to disasters.

1. Introduction

Following the series of natural hazards with unprecedented severity and magnitude including Tohoku Tsunami and Hurricanes Harvey, Irma and Maria, the concept of urban resilience has gained significant attention [1]. For many cities, it is of utmost importance to build institutional and infrastructural capacities to minimize economic loss and maintain the well-being of their citizens in case of extreme events [2,3]. Recent disasters have shown the existence of large variance in recovery trajectories across communities that have experienced similar damage levels [4,5]. We have witnessed manifold cases where cities experience significant drainage of population even with sufficient recovery of infrastructure systems [6,7]. Understanding the interplay between the recovery of infrastructure systems and population movement (displacement and return) after such large-scale disasters is essential for developing policies that could enhance effective population recovery in communities, and foster sustainable development in hazard-prone areas [8].

Human displacement and migration due to climate change and disasters have long been studied based on surveys and census data [5,6,9–15]. Studies have made future projections of migration patterns due to climate change and their implications for income inequality [16,17]. These studies reveal important factors that are associated with post-disaster displacement [18,19]; however, they fail to capture the detailed temporal patterns of recovery and their spatial heterogeneity. The increase in the availability of large mobility datasets including mobile phone call detail records (CDRs), GPS logs, and social media posts, has made it easier to collect spatio-temporally detailed observations of individual mobility from a large region [20,21]. Large-scale datasets are being used for inferring population distributions and migration patterns [22–26], which are applied in various domains to solve societal challenges including alleviating traffic congestion [27]

and preventing disease spread [28–30] . Several studies have used mobile phone data to understand the human mobility patterns during and after disasters such as earthquakes [31–34], hurricanes [35,36], and other anomalous events and shocks [37,38]. Lu *et al*. analysed mobile phone data to understand both the short-term and long-term migration patterns after a cyclone in Bangladesh [39]. Despite such progress, the current body of literature lacks a general understanding of population displacement and recovery patterns after disasters. More specifically, the following research questions are yet to be answered: *When do communities recover from population displacement, and why? Can we characterize population recovery patterns across different disaster events? How heterogeneous are recovery patterns across locations and disaster instances? Can we explain such heterogeneity using a common set of factors?*

In order to answer these questions and bridge the gaps in the current literature, we analysed large scale mobile phone GPS datasets collected before and after multiple disasters across different counties. We collaborated with three different companies across the US and Japan that collect GPS location data from mobile phones, and studied the movements of more than 1.9 million mobile phones of affected individuals over a six-month period. We studied the recovery patterns after Hurricane Maria (Puerto Rico, USA, 2017), Hurricane Irma (Florida, USA, 2017), Tohoku Tsunami (Tohoku area, Japan, 2011), Kumamoto Earthquake (Kyushu area, 2016) and Kinugawa Flood (Ibaraki area, Japan, 2015), shown in figure 1*a*. These five disasters, in total, destroyed more than 1.5 million residential buildings, caused power outages in more than 8 million households and caused more than $350 billion in economic loss (electronic supplementary material). The five disasters were diverse in various aspects including the type of disaster (tsunami, earthquake, hurricane, flood), location of occurrence (Puerto Rico, Florida, Tohoku, Kumamoto) and the socio-economic characteristics of the affected regions.

2. Data and methods

For each disaster, we analysed the longitudinal population recovery patterns in the affected areas. The affected areas were defined as the set of local government units (LGUs), which experienced damage to residential buildings due to the hazard. LGUs correspond to counties in Florida and Puerto Rico, and 'shichoson (cities/wards)' in Japan in this study.

There are mainly three 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 are available on the census tract level 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.

Housing damage data collected from official sources are used to understand the spatial extent of damage inflicted to each of the communities. For disasters in the US, the 'housing damage rate' of a given LGU refers to the rate of houses approved for the Individuals and Households Program of FEMA in each LGU [40]. For disasters in Japan, it refers to the rate of residential buildings classified as 'totally destroyed' or 'half destroyed' by the Cabinet Office of Japan (COJ) in each LGU [41]. Both datasets are publicly accessible. Seventy-eight LGUs in Puerto Rico, 49 LGUs in Florida, 30 LGUs in Tohoku, 33 LGUs in Kumamoto and 10 LGUs in Kinugawa were classified as affected areas with housing damage, and were included in the analysis (electronic supplementary material, table S1). Figure 1*a* shows the LGUs that were included in the analysis along with the housing damage rates in red colours.

Mobile phone location data for the five disasters were provided by three different companies in Japan and the USA Location data were collected by Yahoo Japan Corporation (https://www.yahoo.co.jp/) for Kumamoto Earthquake and Kinugawa Flood, by Zenrin Data Com (http://www.zenrin-datacom.net/toppage) for Tohoku Tsunami and Earthquake and Safegraph (https://www.safegraph.com/) for Hurricanes Irma and Maria. All companies obtained the location information (time, longitude, latitude) of mobile phones from users who agreed to provide their location data for research purposes, and all information was anonymized to protect the security of users. Each mobile phone user's home location was estimated by performing a weighted mean-shift clustering on the GPS location points observed during night-time prior to the disaster date [42,43]. As a result, a total of 1.9 million individual users were identified to be living in the affected areas before the disaster (electronic supplementary material, table S3). We refer to these users as 'affected users'. Correlations (both Pearson and Spearman rank correlations) between the number of affected mobile phone users in each LGU and the census population data were very high in all datasets (electronic supplementary material, figure S2). Thus, we assume that distribution of mobile phone users have little spatial bias, and that they are representative of the entire population on the macroscopic spatial scale, which is also shown in previous works using other mobile phone datasets [23,31,34]. The mobility trajectories of each user were tracked during and after the disaster, and were used to quantify the longitudinal population recovery patterns. The rate of displacement on a given day was defined as the rate of affected users who stayed outside their home LGU out of all affected users on that day. To capture the short-term fluctuations in the population recovery patterns, the raw observations of displacement rates were denoized using Gaussian process regression, which is a non-parametric probabilistic model for denoizing and regression [44]. To capture the general trend of population recovery, the raw observations were fitted using a negative exponential function $D(t) = (D_{160} - D_0)\exp(-(t/\tau)) + D_{160}$, where D_0, D_{160}, and τ denote the displacement rates on day 0, day 160 and recovery time parameter, respectively. Furthermore, the fitted negative exponential functions were normalized $\tilde{D}(t) = (D(t) - D_{160})/(D_0 - D_{160}) = e^{(-t/\tau)}$ to compare the speed of population recovery across different disasters.

3. Modelling population recovery patterns after disasters

Despite the differences in the disaster types and the heterogeneity in socio-economic characteristics among the affected regions, the recovery of displacement rates after the five disasters were all approximated well by a negative exponential function $D(t) = (D_0 - D_{160})\exp(-(t/\tau)) + D_{160}$, where D_0 and D_{160} denote the displacement rates on day 0 and day 160, respectively, and τ are the recovery time parameters.

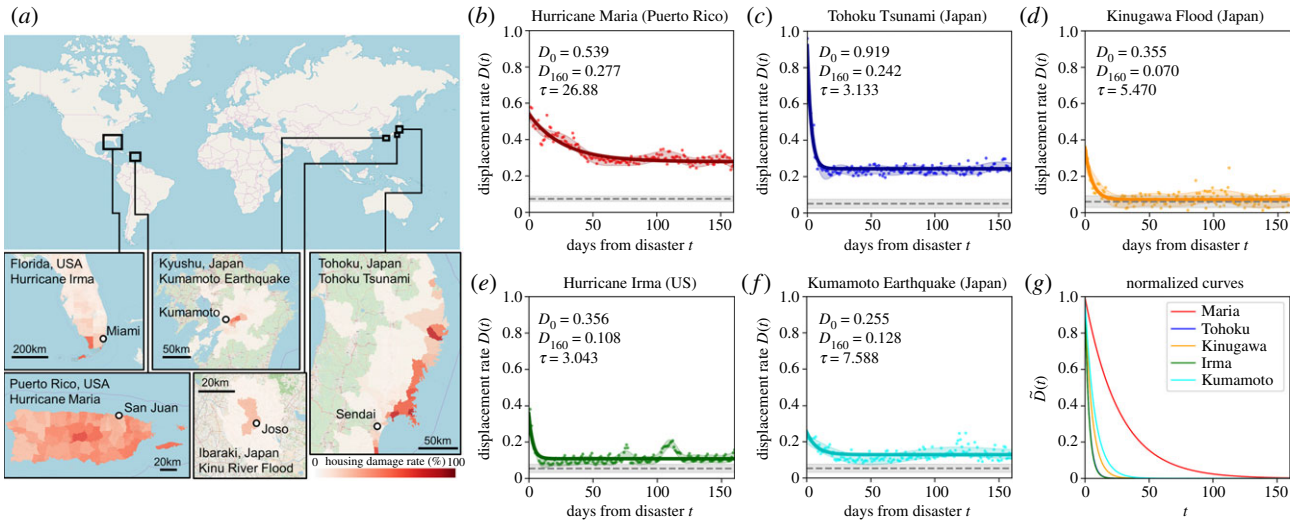

Figure 1. Similarity of macroscopic population recovery patterns across the five disasters. (*a*) Location, spatial scale and severity of disasters that were studied. Red colours indicate the percentages of houses that were severely damaged in each community. (*b–f*) Macroscopic population recovery patterns after each disaster. Raw observations of displacement rates were denoised using Gaussian process regression and were then fitted with a negative exponential function. D_0, D_{160} and τ denote the displacement rates on day 0, day 160 and recovery time parameter of each fitted negative exponential function. Black horizontal dashed line shows average displacement rates observed before the disaster. (*g*) Normalized population recovery patterns after Hurricane Maria (red), Tohoku Tsunami Tsunami (blue), Hurricane Irma (green), Kumamoto Earthquake (cyan), Kinugawa Flood (orange).

Figure 1*b–f* shows the observed daily displacement rates, smoothed trend estimated using Gaussian process regression, and the fitted exponential functions for each disaster. Goodness of fit measures were computed to show that the exponential functional form fits the data well, and that the estimation of parameters are robust (electronic supplementary material, figure S3). The observations were cut off on day 160 due to data limitation. Minor anomalies observed in the recovery patterns were due to national holidays such as Christmas (around day 100 of Hurricane Maria and day 110 of Hurricane Irma), Thanksgiving Holidays (around day 80 of Hurricane Irma), and 'Obon Breaks', which is a national holiday in Japan (around day 120 of Kumamoto Earthquake). The baseline (pre-disaster) displacement rates are shown in figure 1*b–f* in horizontal grey dotted lines (mean) and the grey shaded region (standard deviation) to compare the post-disaster displacement rates with the 'usual' displacement rates that are caused by activities such as travelling. In extreme disasters such as Hurricane Maria and Tohoku Tsunami, we observe a high long-term displacement even after 150 days from the disaster. We can infer that this population segment could have migrated out of the disaster affected areas to other locations. Figure 1*g* shows the normalized displacement rate observations $\tilde{D}(t)$ for each disaster in colours, along with the negative exponential function ($\tilde{D}(t) = e^{(-t/\tau)}$) shown in black. The closeness between the standard negative exponential function and the normalized population recovery patterns show that for all disasters, population recovery curves can be well approximated by a negative exponential function.

The negative exponential functional form of the population recovery patterns across the five disasters imply that the majority of users returned quickly within a couple of weeks from the disaster, but the rest of the users gradually returned over a longer time period. The exponential decay also indicates that for each day, a constant rate ($1/\tau$) of the remaining displaced population decides to return to their original home location. This variance in recovery timings can be explained by observing the relationship between the temporal duration and spatial distance of individual displacement mobility patterns. Figure 2*a* shows that the average evacuation duration increases with evacuation distance. Figure 2*b* shows the probability density plots of the maximum distance travelled from his/her estimated home location on a usual day before the disaster (grey), on the day of the disaster (brown), 10 days after the disaster (red) and one month after the disaster (orange). More people stayed further away (greater than 10^3 m) from their home locations after disasters compared to before the disaster due to evacuation activities. The distribution of evacuation distances is long tailed after all disasters at various time points, which indicates the majority of people evacuate short distances (thus short duration) and a small fraction evacuate extremely long distances (thus long duration). This explains why we observe the negative exponential function in population recovery patterns after all disasters. The recovery times after disasters that occurred in Japan and Florida were relatively short ($3 < \tau < 8$), but very long $\tau = 26.8$ after Hurricane Maria. The differences in recovery time parameter values τ across disasters can be explained by the differences in the speed of infrastructure recovery in each of the affected regions. In Japan and Florida, power was restored in over 90% of the households (that were not destroyed) within 10 days from the disaster, while it took more than 200 days for Puerto Rico (electronic supplementary material, figure S4).

4. Understanding the spatial heterogeneity in population recovery

We now downscale our analysis to LGUs (counties in the USA, cities/wards in Japan) within each affected area to understand the spatial heterogeneity in population recovery patterns. Since only one of few LGUs (10) was affected by the Kinugawa Flood, with most of them (9) having little housing damage (less than 1% housing damage rates), the Kinugawa Flood was not included in the LGU-scale analysis. The LGU-scale analysis was performed on the four major disasters

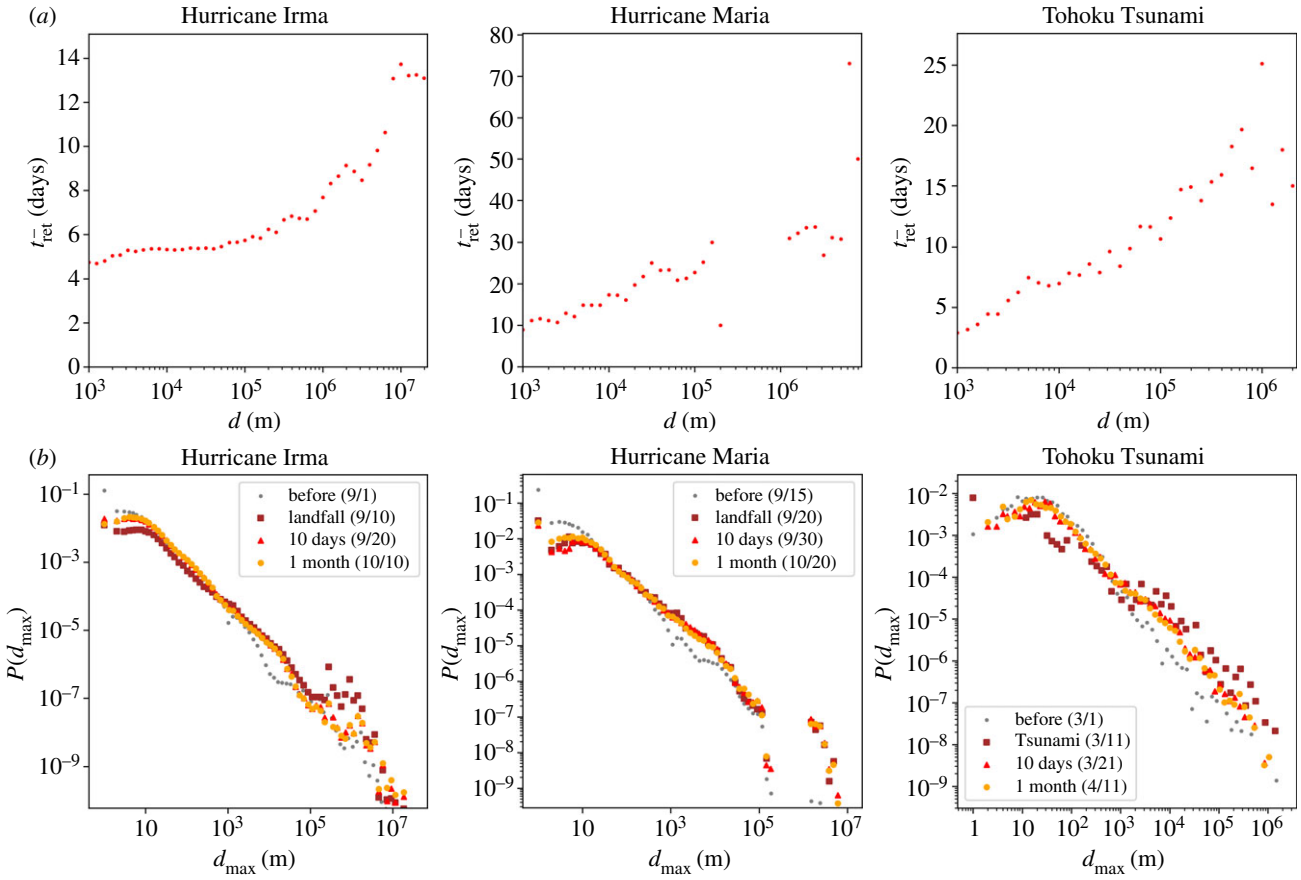

Figure 2. Relationship between displacement distance and duration after disasters. (*a*) The longer the evacuation distance, the longer the average evacuation duration. (*b*) Probability densities of maximum distance from home on a usual day (grey) and at various timings after the disaster day (brown: day of occurrence, red: 10 days after, orange: one month after) are all long-tailed. The long-tailed distribution of evacuation distances and the relationship between displacement distance and duration were common across disasters. The majority of people evacuating short distances (thus short duration) and a small fraction of the people evacuating extremely long distances (thus long duration) explains why similar negative exponential functions were observed after the disasters.

(Hurricanes Maria and Irma, Tohoku Tsunami, and Kumamoto Earthquake). In total, there are 200 LGUs with large diversity in socio-economic characteristics that were affected by the four disasters. The inset in figure 3*a* shows the large heterogeneity in recovery patterns across LGUs, even within each disaster. Figure 3*a* shows moderate correlation ($R = 0.612$) between D_0 and D_{160} for all LGUs.

To understand the effect of the independent variables on the displacement rates and the speed of recovery, we apply a generalized linear regression modelling 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.

$$\log\left(\frac{D(t)}{1 - D(t)}\right) = \beta^T x + \epsilon \tag{4.1}$$

and

$$\log(\tau) = \beta^T x + \epsilon. \tag{4.2}$$

In the regression models of population recovery, socio-economic data (population, median income, housing damage rates, power outage recovery time, connectedness to surrounding cities) were used as independent variables (electronic supplementary material, table S4). In addition to housing damage rates which directly quantify the magnitude of the disaster effect on each LGU, socio-economic variables (population and income) of LGUs were included in the model to seek any inequality between the urban and rural, and the rich and poor on the disaster recovery performances. Infrastructure recovery (power outage duration) was included in the model to assess the importance of the local agency's capacity to respond to extreme events. Moreover, we test whether the geographical configurations and accessibility between LGUs are important for post-disaster recovery, by including variables related to the proximity to large and wealthy cities. For Florida and Puerto Rico, population data were obtained from the US National Census (https://www.census.gov/), and median income data were obtained from the American Community Survey (https://www.census.gov/programs-surveys/acs). Similarly, for Japanese LGUs, population and income data were obtained from the Statistics Bureau (https://www.stat.go.jp/) of the Ministry of Internal Affairs and Communications of Japan. Power outage data of LGUs in Puerto Rico were collected from the website StatusPR (http://status.pr/), which is a government operated website that showed the recovery status of Puerto Rico after the Hurricane. Power outage data of Hurricane Irma were collected from the Florida Division of Disaster Management (https://www.floridadisaster.org/). Power outage information in the Japanese disasters were collected from the utility companies. The connectedness to surrounding cities were calculated by

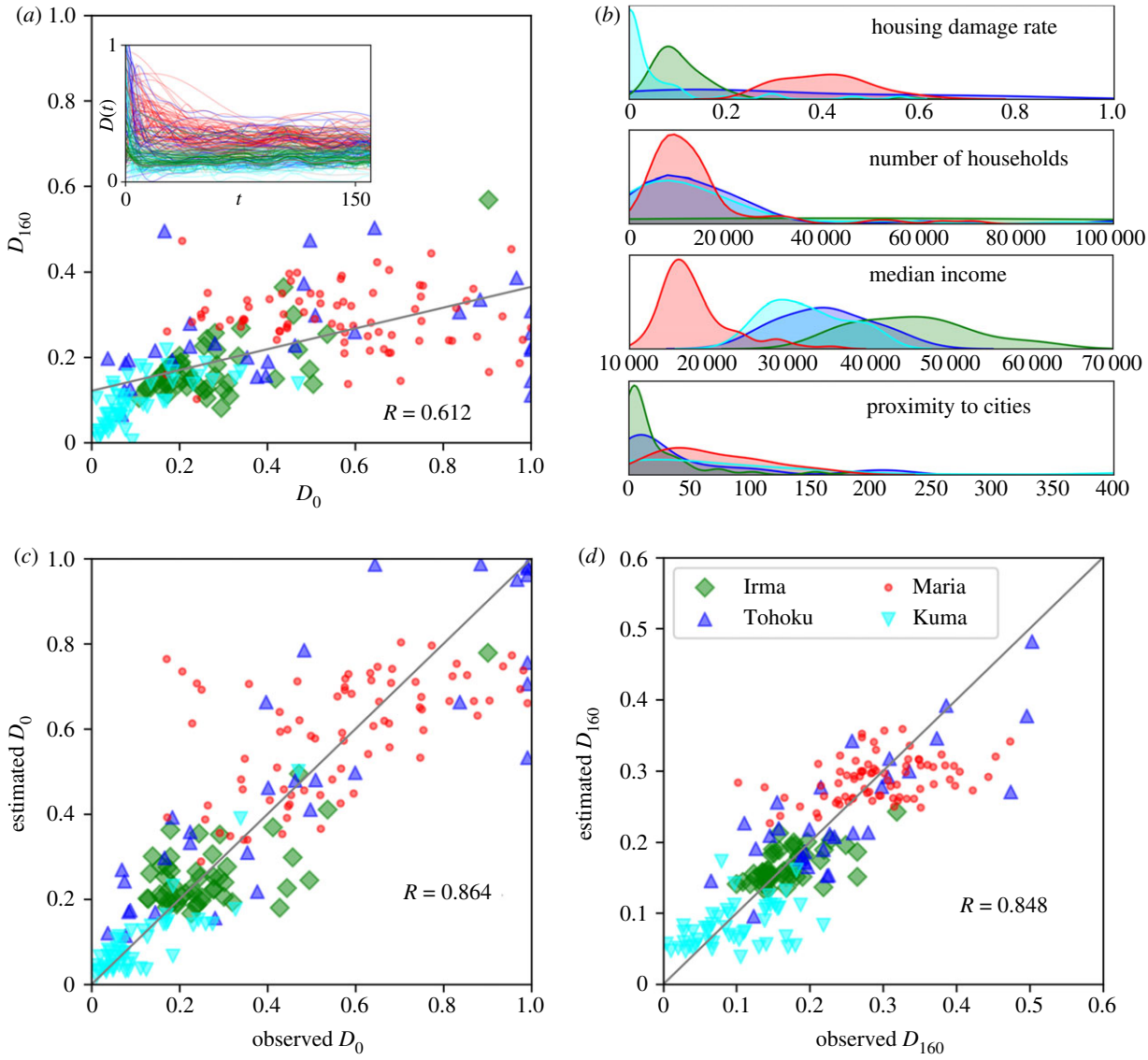

Figure 3. Explaining the spatial heterogeneity in population recovery using key common factors. (*a*) D_0 and D_{160} of all LGUs. Each trajectory corresponds to an LGU, and colours represent the disaster. D_0 and D_{160} have a moderate correlation of $R = 0.612$. The inset shows the high spatial heterogeneity of recovery trajectories. (*b*) Density plots of the four features of the affected LGUs, showing the heterogeneity in socio-economic characteristics. (*c*) Observed and estimated D_0 for all LGUs in all disasters had high correlation $R = 0.864$. (*d*) Observed and estimated D_{160} values had high correlation $R = 0.848$.

$d_p(i) = (\sum_{j \in S(i)} N_j)/(N_i)$, 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 h 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. Correlations among variables in all disasters were not significantly high, thus we included them in the regression analysis (electronic supplementary material, table S5). Power outage recovery time was excluded in the models for estimating D_0, since this information would not be available on day 0. The probability densities of the four attributes in each disaster are shown in figure 3*b*. Housing damage rates and median income levels significantly differ across the four disasters, however the number of households and the connectedness of cities have more similar distributions. The set of independent variables for the best model for each disaster was chosen based on the lowest AIC value and statistical significance ($p < 0.1$). Regression results are shown in detail in electronic supplementary material, tables S6–S9. Figure 3*c* plots the observed values and estimated D_0. Although we use only key variables in our model, the estimated values had high correlation with observed values ($R = 0.864$). For LGUs in Japan (Tohoku Tsunami and Kumamoto Earthquake) and Florida (Hurricane Irma), housing damage rates were good estimators of D_0 (figure 4*a–d*). On the other hand, housing damage rates had low and insignificant correlation with $\log (D_0/(1 - D_0))$ in Puerto Rico after Hurricane Maria. Rather, as shown in figure 4*e–g*, median income levels and number of households for each community had significant and stronger correlations with initial displacement rates. Median income and the proximity to wealthy cities had negative correlations with initial displacement rates, indicating that communities with lower incomes in isolated areas had higher initial displacement rates (electronic supplementary material, table S6). Figure 3*d* shows that D_{160} values

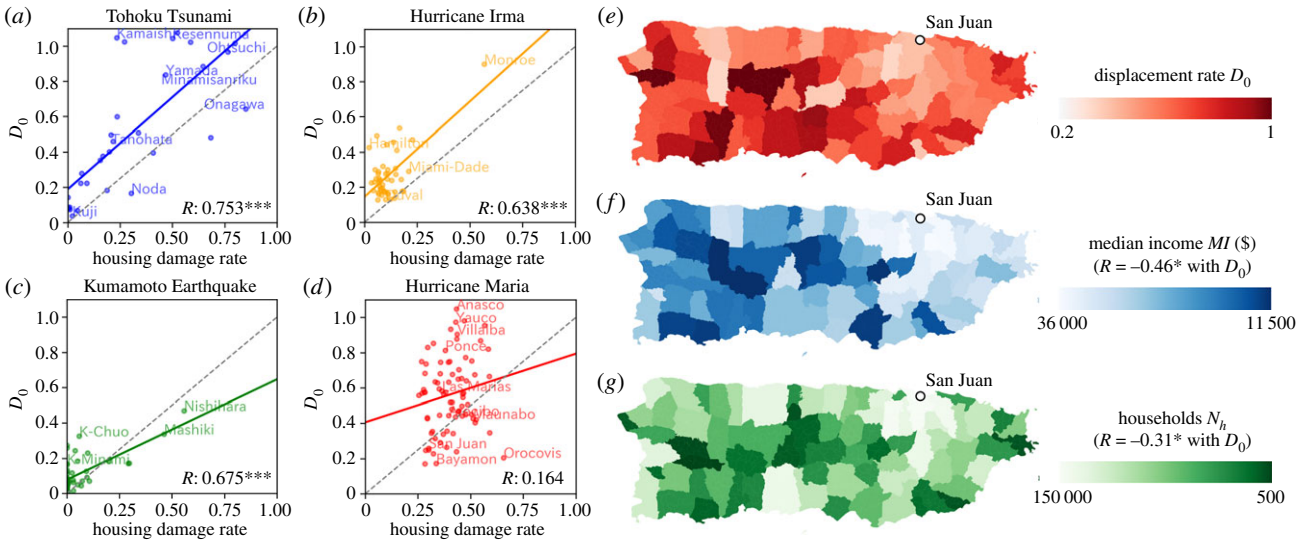

Figure 4. Initial displacement rates after disasters. (*a–d*) D_0 and housing damage rates for each LGU in Tohoku Tsunami (*a*), Hurricane Irma (*b*), Kumamoto Earthquake (*c*) and Hurricane Maria (*d*). Solid lines are regression results and the dotted black lines mark theoretical results if all people whose houses were damaged fled their hometown. Pearson's R values are shown in the corner of each panel, and their significance levels are noted by stars (***: $p < 0.01$). For disasters in Japan and mainland US, housing damage rates had a significant and strong correlation. However, the correlation was insignificant and smaller in Puerto Rico after Hurricane Maria. (*e–g*) Spatial distribution of D_0 (*e*), median income (*f*) and number of households (*g*) for all LGUs in Puerto Rico after Hurricane Maria. For Hurricane Maria, median income levels and number of households for each community had significant and negative correlations with D_0 than housing damage rates.

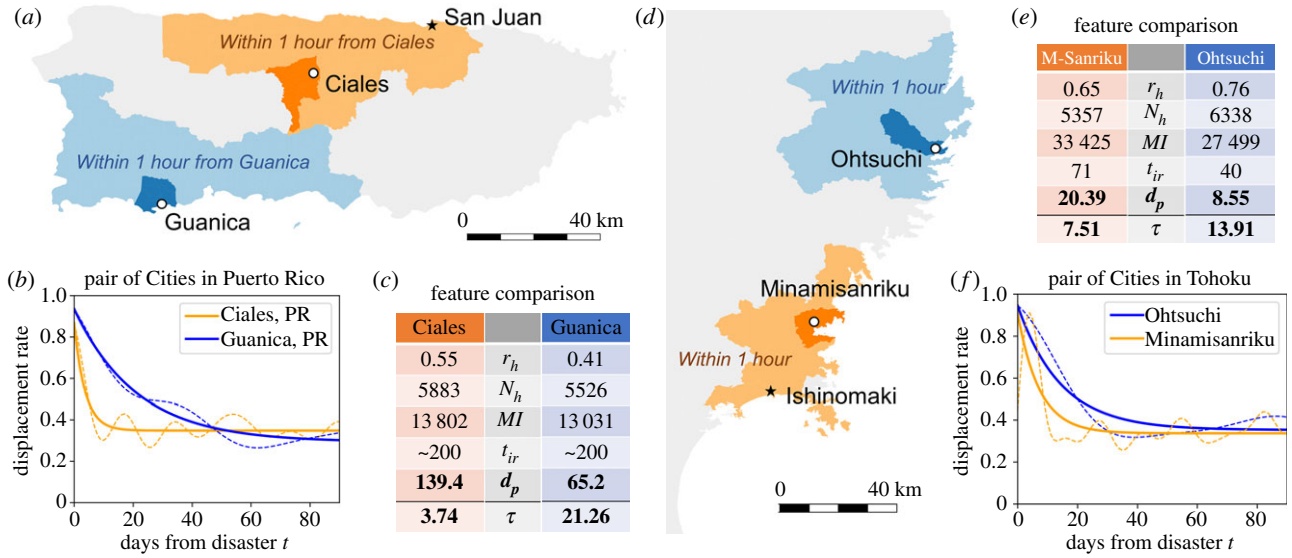

Figure 5. Connectedness to neighbouring cities as a key factor to recovery. (*a*) Map of Ciales and Guanica, Puerto Rico. Light coloured areas show the area that can be reached from each city within 1 h of driving time. (*b*) Recovery patterns of both communities, showing the faster recovery of Ciales. (*c*) Comparison of factors for both cities. Factors other than connectedness to other large cities (e.g. San Juan) were similar between the two communities. (*d–f*) Similar phenomenon was seen after Tohoku Tsunami in Japan. Minamisanriku City and Ohtsuchi City shared similar characteristics except for the connectedness to large cities (e.g. Ishinomaki), resulting in differences in recovery patterns.

were predicted by the five variables with high accuracy ($R = 0.848$). Median income and housing damage rates had positive effects on long-term displacement, implying that people with more income were able to evacuate from the affected regions (electronic supplementary material, table S7). In addition to such socio-economic variables, infrastructure recovery speed had a significant effect on long-term displacement rates. Recovery speed log (τ) had the lowest predictability out of all objective variables. The significant variables varied across different disasters, however, the connectedness to large cities and wealthy cities was a common variable with significant impact on recovery speed across three disasters (electronic supplementary material, table S8). The negative coefficient implies that if a city is surrounded by larger or wealthier cities, it has a shorter time needed for recovery. To check the temporal robustness of these findings, We performed the regression analysis on various time points (D_{10}, D_{20}, D_{30}, D_{60}, D_{90}, D_{120}), and summarized the results in electronic supplementary material, table S9. We found that the set of important variables generally stay similar for all disasters across different timepoints. However, as time progresses, infrastructure recovery variables become more significant while the significance of housing damage rates gradually decrease.

Figure 5 shows pairwise comparisons after Hurricane Maria and Tohoku Tsunami where a pair of similar LGUs with different levels of connectedness to neighbouring cities have distinct recovery outcomes, even though other socio-economic characteristics such as population, housing damage rates and income levels are similar.

5. Discussion and conclusion

In this study, we used large-scale mobile phone datasets from five disasters across the US, Puerto Rico and Japan to uncover the macroscopic population recovery patterns after disasters. We found that population recovery patterns across the five disasters can be approximated with a common negative exponential function, where the majority of the displaced users return to their residential areas quickly after the disaster, while some decide to stay away for longer periods. Previous studies of post-disaster human migration using household surveys [5,10,11,15] had failed to obtain a continuous and longitudinal understanding of the recovery dynamics due to limitations in data. Revealing the negative exponential function common across different disasters and locations could significantly contribute to the efforts in modelling and simulating human mobility patterns after disasters [45]. Further analysis using the individual evacuation mobility patterns showed that these patterns emerge because of the combined effect of long-tailed distributions in evacuation distances and positive correlation between evacuation distance and duration. The long-tailed distributions of evacuation distance had been observed in previous studies using other datasets from Haiti [31] and Japan [34], however, the latter relationship had not been shown in previous studies. A more detailed analysis on the population recovery patterns of 200 communities (LGUs) across different disasters and locations showed that the heterogeneity in short-term (day 0) to long-term (day 160) displacement rates can be well explained by key common factors including population, median income, housing damage rates and infrastructure recovery time. Previous studies on individual case studies have noted the relationship between such variables and population recovery (reentry) decisions. For example, studies on Hurricanes Katrina and Rita show that the rate of disadvantaged populations (characterized by variables including household income), density of the built environment, and housing damage contribute to migration and displacement [7]. This work contributes to the literature regarding disaster resilience and population migration by testing the insights obtained from individual case studies with multiple disasters in different locations. Moreover, the importance of connectivity to surrounding cities has been understudied in the current literature, despite its significant implications on policymaking for disaster resilience. This contradicts previous findings on non-disaster human mobility patterns (e.g. commuting), where out-migration increases with amount of opportunities available in surrounding cities [25]. This finding shows that after disasters, the existence of neighbouring cities act as catalysts that enhance recovery rather than attractors that drain population from damaged cities. This extends the theories on the importance of social capital and social support [4,15] to an inter-city scale. One example of this effect is how Tono City, an inland city close to the Tohoku area towns that were affected by the tsunami, acted as a recovery support hub after the Tohoku Tsunami [46]. The coastal cities were provided humanitarian, informational, and material support from surrounding nearby cities such as Tono City which experienced less damage due to the tsunami/earthquake. The effect of inter-city connectivity on community recovery is understudied in the urban resilience literature, and could have significant implications on the planning of inter-city networks to enhance the resilience of communities.

The presented empirical results should be considered in light of some limitations. First, the mobile phone data did not contain various attributes of households that are known to affect evacuation and return decisions such as age, gender, race [47], risk perception [48] and social network ties [4,15]. Although we were able to explain some of the variability in population recovery patterns using several key factors, including these variables by integrating mobile phone data with household survey data would be a valuable next step in future research. Second, although this study used mobile phone data from a diverse set of countries, the number of disaster cases need to be increased to be able to make robust conclusions. Increasing the number of study locations especially in the developing world would allow us to better compare and understand the recovery after Hurricane Maria in Puerto Rico. Moreover, although our study analysed the longitudinal population recovery until 160 days after the disaster, a longer study could provide more insights in the recovery of communities. Especially, for Puerto Rico after Hurricane Maria, observing more longitudinal data to determine whether the displacement rate permanently stays high after 160 days would provide valuable insights. With data for longer time periods from more instances of large-scale disasters in different countries, we will be able to obtain more generalizable findings on population recovery patterns after disasters. Third, in this study, regression analysis on the recovery parameters were conducted on the 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. The LGU scale was the finest scale in which we can obtain socio-economic data in Japan, unlike the USA, where data are 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. 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 (electronic supplementary material, figures S7 and S8 and table S10). The predictability of displacement rates were low compared to the county (LGU) level analysis (figure 3). The small sample size in each census tract (electronic supplementary material, figure S9) could be the reason for the noisy estimate. A more robust estimation using sparse mobile phone user samples would be needed to give better estimations on the census tract granularity. Finally, our primary focus of this paper was on the returning mobility of the displaced populations after disasters. Extending this study to not only returning behaviour but also incoming migration would be of interest to understand the recovery and further development of each community.

First, the understanding of the underlying recovery process using parsimonious models provides a way to characterize the recovery process using a small number of parameters (initial displacement and speed of recovery). Even when mobile phone data could not be obtained by policy makers in real-time, the negative exponential model can be combined using other (often spatio-temporally low resolution) data sources to make predictions of population recovery. For example, studies have shown that social media data (e.g. Twitter) can be used as a proxy to estimate human mobility [49]. Night light data collected from satellite images may also be used as a proxy for population recovery as well, after power has been restored in the area [50]. Second, the revealed common function of population recovery dynamics can be applied in developing agent-based simulations of evacuation and return mobility, which are commonly used in practice to predict post-disaster mobility and population recovery [45,51]. Predicted post-disaster mobility patterns can be used as a reference for policymakers to make decisions on the spatio-temporal allocation of resources and services such as evacuation shelters, emergency supplies, public utilities (e.g. water, electricity, gas). Third, the LGU-level analysis results can be used to estimate the impact of policies on population recovery. For example, it was shown that physical connectivity between LGUs was important for effective recovery, in addition to characteristics of individual LGUs. In the case of Puerto Rico after Hurricane Maria (figure 5*a*), the arterial road that runs through the central part of the island connecting communities in the southern part to the northern cities (e.g. San Juan) was not well designed, with inefficient road structures and paths. Similarly, communities in the northern part of Tohoku region (figure 5*d*) were severely isolated due to the mountainous terrains. This finding suggests that policymakers need to evaluate the collective capacity of the network of LGUs, rather than evaluating the resilience of each LGU independently.

Data accessibility. 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. Data collected from other sources are available from official documents that are openly accessible. Computer codes used to process and analyse the data are posted on the author's github page. (https://github.com/takayabe0505).

Authors' contributions. T.Y., K.T., N.F., Y.S. and S.V.U. designed the research; T.Y., K.T., N.F. and S.V.U. performed the research; T.Y. and K.T. analysed the data; T.Y., K.T., N.F., Y.S. and S.V.U. wrote the paper.

Competing interests. The authors declare that they have no competing financial interests.

Funding. The work of T.Y. and S.V.U. is partly funded by NSF grant no. 1638311 CRISP Type 2/Collaborative Research: Critical Transitions in the Resilience and Recovery of Interdependent Social and Physical Networks. N.F. was supported by JSPS KAKENHI grant number JP17H01742.

Acknowledgements. We thank collaborators in Safegraph, Mr Hodaka Kaneda of Zenrin Data Com and Mr Hiroshi Kanasugi of University of Tokyo for preparing the mobile phone GPS data used in this study.

References

1. De Bettencourt UM. 2013 Building resilience: integrating climate and disaster risk into development – the World Bank Group experience: Main report. See https://documents.worldbank.org/curated/en/762871468148506173/Main-report.
2. Eakin H *et al.* 2017 Opinion: urban resilience efforts must consider social and political forces. *Proc. Natl Acad. Sci. USA* **114**, 186–189. (doi:10.1073/pnas.1620081114)
3. Kousky C. 2014 Informing climate adaptation: a review of the economic costs of natural disasters. *Energy Econ.* **46**, 576–592. (doi:10.1016/j.eneco.2013.09.029)
4. Aldrich DP. 2012 *Building resilience: social capital in post-disaster recovery*. Chicago: University of Chicago Press.
5. Finch C, Emrich CT, Cutter SL. 2010 Disaster disparities and differential recovery in New Orleans. *Popul. Environ.* **31**, 179–202. (doi:10.1007/s11111-009-0099-8)
6. McCaughey JW, Daly P, Mundir I, Mahdi S, Patt A. 2018 Socio-economic consequences of post-disaster reconstruction in hazard-exposed areas. *Nat. Sustainability* **1**, 38–43. (doi:10.1038/s41893-017-0002-z)
7. Myers CA, Slack T, Singelmann J. 2008 Social vulnerability and migration in the wake of disaster: the case of Hurricanes Katrina and Rita. *Popul. Environ.* **29**, 271–291. (doi:10.1007/s11111-008-0072-y)
8. Aerts JCJH *et al.* 2018 Integrating human behaviour dynamics into flood disaster risk assessment. *Nat. Clim. Change* **8**, 193–199. (doi:10.1038/s41558-018-0085-1)
9. Cutter SL, Barnes L, Berry M, Burton C, Evans E, Tate E, Webb J. 2008 A place-based model for understanding community resilience to natural disasters. *Global Environ. Change* **18**, 598–606. (doi:10.1016/j.gloenvcha.2008.07.013)
10. DeWaard J, Curtis KJ, Fussell E. 2016 Population recovery in new orleans after Hurricane Katrina: exploring the potential role of stage migration in migration systems. *Popul. Environ.* **37**, 449–463. (doi:10.1007/s11111-015-0250-7)
11. Fussell E, Curtis KJ, DeWaard J. 2014 Recovery migration to the city of new orleans after Hurricane Katrina: a migration systems approach. *Popul. Environ.* **35**, 305–322. (doi:10.1007/s11111-014-0204-5)
12. Gray CL, Mueller V. 2012 Natural disasters and population mobility in Bangladesh. *Proc. Natl Acad. Sci. USA* **109**, 6000–6005.
13. Kniveton DR, Smith CD, Black R. 2012 Emerging migration flows in a changing climate in dryland Africa. *Nat. Clim. Change* **2**, 444–447. (doi:10.1038/nclimate1447)
14. Piguet E, Pécoud A, De Guchteneire P. 2011 Migration and climate change: an overview. *Refug. Surv. Q.* **30**, 1–23. (doi:10.1093/rsq/hdr006)
15. Sadri AM, Ukkusuri SV, Lee S, Clawson R, Aldrich D, Nelson MS, Seipel J, Kelly D. 2018 The role of social capital, personal networks, and emergency responders in post-disaster recovery and resilience: a study of rural communities in Indiana. *Nat. Hazards* **90**, 1377–1406. (doi:10.1007/s11069-017-3103-0)
16. Hauer ME. 2017 Migration induced by sea-level rise could reshape the US population landscape. *Nat. Clim. Change* **7**, 321–325. (doi:10.1038/nclimate3271)
17. Shayegh S. 2017 Outward migration may alter population dynamics and income inequality. *Nat. Clim. Change* **7**, 828–832. (doi:10.1038/nclimate3420)
18. Chen J, Mueller V. 2018 Coastal climate change, soil salinity and human migration in Bangladesh. *Nat. Clim. Change* **8**, 981–985. (doi:10.1038/s41558-018-0313-8)
19. Mueller V, Gray C, Kosec K. 2014 Heat stress increases long-term human migration in rural Pakistan. *Nat. Clim. Change* **4**, 182–185. (doi:10.1038/nclimate2103)
20. Blondel VD, Decuyper A, Krings G. 2015 A survey of results on mobile phone datasets analysis. *EPJ Data Sci.* **4**, 10. (doi:10.1140/epjds/s13688-015-0046-0)
21. Gonzalez MC, Hidalgo CA, Barabasi A-L. 2008 Understanding individual human mobility patterns. *Nature* **453**, 779–782. (doi:10.1038/nature06958)

22. Blumenstock JE. 2012 Inferring patterns of internal migration from mobile phone call records: evidence from Rwanda. *Inf. Technol. Dev.* **18**, 107–125. (doi:10.1080/02681102.2011.643209)
23. Deville P, Linard C, Martin S, Gilbert M, Stevens FR, Gaughan AE, Blondel VD, Tatem AJ. 2014 Dynamic population mapping using mobile phone data. *Proc. Natl Acad. Sci. USA* **111**, 15 888–15 893. (doi:10.1073/pnas.1408439111)
24. Lai S, zu Erbach-Schoenberg E, Pezzulo C, Ruktanonchai NW, Sorichetta A, Steele J, Li T, Dooley CA, Tatem AJ. 2019 Exploring the use of mobile phone data for national migration statistics. *Palgrave Commun.* **5**, 34. (doi:10.1057/s41599-019-0242-9)
25. Simini F, González MC, Maritan A, Barabási A-L. 2012 A universal model for mobility and migration patterns. *Nature* **484**, 96–100. (doi:10.1038/nature10856)
26. Wardrop NA, Jochem WC, Bird TJ, Chamberlain HR, Clarke D, Kerr D, Bengtsson L, Juran S, Seaman V, Tatem AJ. 2018 Spatially disaggregated population estimates in the absence of national population and housing census data. *Proc. Natl Acad. Sci. USA* **115**, 3529–3537. (doi:10.1073/pnas.1715305115)
27. Iqbal Md. S, Choudhury CF, Wang P, González MC. 2014 Development of origin–destination matrices using mobile phone call data. *Transp. Res. Part C: Emerging Technol.* **40**, 63–74. (doi:10.1016/j.trc.2014.01.002)
28. Bengtsson L, Gaudart J, Lu X, Moore S, Wetter E, Sallah K, Rebaudet S, Piarroux R. 2015 Using mobile phone data to predict the spatial spread of cholera. *Sci. Rep.* **5**, 8923. (doi:10.1038/srep08923)
29. Finger F, Genolet T, Mari L, de Magny GC, Manga NM, Rinaldo A, Bertuzzo E. 2016 Mobile phone data highlights the role of mass gatherings in the spreading of cholera outbreaks. *Proc. Natl Acad. Sci. USA* **113**, 6421–6426. (doi:10.1073/pnas.1522305113)
30. Wesolowski A, Eagle N, Tatem AJ, Smith DL, Noor AM, Snow RW, Buckee CO. 2012 Quantifying the impact of human mobility on malaria. *Science* **338**, 267–270. (doi:10.1126/science.1223467)
31. Lu X, Bengtsson L, Holme P. 2012 Predictability of population displacement after the 2010 Haiti earthquake. *Proc. Natl Acad. Sci. USA* **109**, 11 576–11 581. (doi:10.1073/pnas.1203882109)
32. Song X, Zhang Q, Sekimoto Y, Shibasaki R. 2014 Prediction of human emergency behavior and their mobility following large-scale disaster. In *Proc. of the 20th ACM SIGKDD Int. Conf. on Knowledge discovery and data mining*, pp. 5–14. ACM.
33. Wilson R *et al.* 2016 Rapid and near real-time assessments of population displacement using mobile phone data following disasters: the 2015 nepal earthquake. *PLoS Curr. Disasters* **8**. (doi:10.1371/currents.dis.d073fbece328e4c39087bc086d694b5c)
34. Yabe T, Sekimoto Y, Tsubouchi K, Ikemoto S. 2019 Cross-comparative analysis of evacuation behavior after earthquakes using mobile phone data. *PLoS ONE* **14**, e0211375. (doi:10.1371/journal.pone.0211375)
35. Wang Q, Taylor JE. 2014 Quantifying human mobility perturbation and resilience in Hurricane Sandy. *PLoS ONE* **9**, e112608. (doi:10.1371/journal.pone.0112608)
36. Wang Q, Taylor JE. 2016 Patterns and limitations of urban human mobility resilience under the influence of multiple types of natural disaster. *PLoS ONE* **11**, e0147299. (doi:10.1371/journal.pone.0147299)
37. Bagrow JP, Wang D, Barabasi A-L. 2011 Collective response of human populations to large-scale emergencies. *PLoS ONE* **6**, e17680. (doi:10.1371/journal.pone.0017680)
38. Toole JL, Lin Y-R, Muehlegger E, Shoag D, González MC, Lazer D. 2015 Tracking employment shocks using mobile phone data. *J. R. Soc. Interface* **12**, 20150185. (doi:10.1098/rsif.2015.0185)
39. Lu X *et al.* 2016 Unveiling hidden migration and mobility patterns in climate stressed regions: a longitudinal study of six million anonymous mobile phone users in Bangladesh. *Global Environ. Change* **38**, 1–7. (doi:10.1016/j.gloenvcha.2016.02.002)
40. FEMA. 2018 Fema housing assistance program data. https://www.fema.gov/media-library/assets/documents/34758. Accessed: 2018-09-07.
41. COJ. 2018 Disaster reports for disasters in Japan (Japanese). http://www.bousai.go.jp/ (accessed 7 September 2018).
42. Ashbrook D, Starner T. 2003 Using GPS to learn significant locations and predict movement across multiple users. *Pers. Ubiquitous Comput.* **7**, 275–286. (doi:10.1007/s00779-003-0240-0)
43. Kanasugi H, Sekimoto Y, Kurokawa M, Watanabe T, Muramatsu S, Shibasaki R. 2013 Spatiotemporal route estimation consistent with human mobility using cellular network data. In *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2013 IEEE Int. Conf. on*, pp. 267–272. IEEE.
44. Rasmussen CE, Williams CKI. 2006 *Gaussian processes for machine learning*. Cambridge, MA: The MIT Press. 38:715–719, 2006.
45. Ukkusuri SV, Hasan S, Luong B, Doan K, Zhan X, Murray-Tuite P, Yin W. 2017 A-RESCUE: an agent based regional evacuation simulator coupled with user enriched behavior. *Netw. Spatial Econ.* **17**, 197–223. (doi:10.1007/s11067-016-9323-0)
46. Shimizu M. 2012 Resilience in disaster management and public policy: a case study of the Tohoku disaster. *Risk, Hazards & Crisis in Public Policy* **3**, 40–59. (doi:10.1002/rhc3.17)
47. Dash N, Gladwin H. 2007 Evacuation decision making and behavioral responses: individual and household. *Nat. Hazards Rev.* **8**, 69–77. (doi:10.1061/(ASCE)1527-6988(2007)8:3(69))
48. Riad JK, Norris FH, Ruback RB. 1999 Predicting evacuation in two major disasters: risk perception, social influence, and access to resources 1. *J. Appl. Soc. Psychol.* **29**, 918–934. (doi:10.1111/j.1559-1816.1999.tb00132.x)
49. Jurdak R, Zhao K, Liu J, AbouJaoude M, Cameron M, Newth D. 2015 Understanding human mobility from twitter. *PLoS ONE* **10**, e0131469. (doi:10.1371/journal.pone.0131469)
50. Román MO *et al.* 2019 Satellite-based assessment of electricity restoration efforts in Puerto Rico after Hurricane Maria. *PLoS ONE* **14**, e0218883. (doi:10.1371/journal.pone.0218883)
51. Murray-Tuite P, Wolshon B. 2013 Evacuation transportation modeling: an overview of research, development, and practice. *Transp. Res. Part C: Emerg. Technol.* **27**, 25–45. (doi:10.1016/j.trc.2012.11.005)