

University of Groningen

Using mobile phone data to map evacuation and displacement

Giardini, Francesca; Hadjidimitriou, Natalia Selini; Mamei, Marco; Bastardi, Giordano; Codeluppi, Nico; Pancotto, Francesca

Published in:
Scientific Reports

DOI:
[10.1038/s41598-023-48130-4](https://doi.org/10.1038/s41598-023-48130-4)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2023

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Giardini, F., Hadjidimitriou, N. S., Mamei, M., Bastardi, G., Codeluppi, N., & Pancotto, F. (2023). Using mobile phone data to map evacuation and displacement: a case study of the central Italy earthquake. *Scientific Reports*, 13(1), Article 22228. <https://doi.org/10.1038/s41598-023-48130-4>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.



OPEN

Using mobile phone data to map evacuation and displacement: a case study of the central Italy earthquake

Francesca Giardini^{1,5}✉, Natalia Selini Hadjidimitriou^{3,5}, Marco Mamei^{3,4,5}✉, Giordano Bastardi³, Nico Codeluppi³ & Francesca Pancotto^{2,5}

Population displacement is one of the most common consequences of disasters, and it can profoundly affect communities and territories. However, gaining an accurate measure of the size of displacement in the days and weeks following a major disaster can be extremely difficult. This study uses aggregated Call Detail Records as an inexpensive and efficient technique to measure post-disaster displacement in four Italian regions affected by repeated earthquakes in 2016–2017. By comparing post-disaster mobile phone count with a forecast computed before the earthquake hit, we can compute an index of change in the presence of mobile phones (MPE). This measure, obtained thanks to advanced analytical techniques, provides a reliable indication of the effect of the earthquake in terms of immediate and medium-term displacement. We test this measure against census data and in combination with other datasets. Looking into available data on economic activities and requests for financial support to rebuild damaged buildings, we can explain MPE and identify significant factors affecting population displacement. It is possible to apply this innovative methodology to other disaster scenarios and use it by policymakers who want to understand the determinants of population displacement.

The United Nations Office for Disaster Risk Reduction (UNDRR) defines a disaster as a “serious disruption of the functioning of a community or a society at any scale due to hazardous events interacting with conditions of exposure, vulnerability and capacity, leading to one or more of the following: human, material, economic and environmental losses and impacts” (UNDRR, n.d.). Quarantelli¹ (p. 682) explores the defining features of disasters, comprising the time frame (disasters are sudden-onset occasions), their disruptiveness towards collective routines, the adoption of unplanned courses of action to adjust to the disruption, and the danger posed to valued social objects. Quarantelli¹ (p. 682) also emphasized that disasters represent vulnerability, reflecting “weaknesses in social structures or social systems”² (p. 345). The conception of disasters as a weakness of social systems is firmly grounded in the social sciences tradition³. Since the relationships in the social system can cause vulnerability, the definition of disasters originates from the notion of social changes. Similarly, Alexander⁴ (p. 29) stresses the impossibility of defining “disasters with fixed events but by social constructs, and these are liable to change”. Other authors like⁵ support this perspective and refers to disasters as social events. Disasters result from a combination of hazards with vulnerability and exposure of people and assets. Vulnerability refers to the conditions which increase the susceptibility of individuals, communities and systems to the hazards’ impact. According to Wisner et al.⁶ (p. 11), vulnerability is “...the characteristics of a person or group and their situation that influence their capacity to anticipate, cope with, resist, and recover from the impact of a natural hazard”. Vulnerabilities vary by hazard type as they are contingent on several circumstances and unevenly distributed across individuals, households, communities, and locations^{7,8}.

Displacement and vulnerability are closely related. Displacement forces villages, towns, or regions to lose inhabitants who need to find alternative accommodations somewhere else⁹. There is extensive literature on the

¹Department of Sociology, University of Groningen, Grote Rozenstraat, 31, 9712 TG Groningen, The Netherlands. ²Dipartimento di Studi Linguistici e Culturali, Università di Modena e Reggio Emilia, Largo S. Eufemia, 19, 41121 Modena, Italy. ³Dipartimento di Scienze e Metodi dell’Ingegneria, Università di Modena e Reggio Emilia, Via Amendola, 2 - Pad. Morselli, 42122 Reggio Emilia, Italy. ⁴Artificial Intelligence Research and Innovation Center - AIRI, Università di Modena e Reggio Emilia, Via Vivarelli, 10, 41125 Modena, Italy. ⁵These authors contributed equally: Francesca Giardini, Natalia Selini Hadjidimitriou, Marco Mamei and Francesca Pancotto. ✉email: f.giardini@rug.nl; marco.mamei@unimore.it

risks to well-being generated by forced displacement (e.g.^{10–13}), but less is known about the increasing vulnerability of displacement that can unfold over different periods, from a few days to decades, and the displaced population can move very close to their original location, but also very far away. There is no univocal concept of displacement because it usually depends on the kind of crisis, disaster or conflict that triggered it, and the term itself is not uncontroversial¹⁴. Here we will use the definition developed by the UNDRR: Disaster displacement is one of the most common and immediate impacts of disasters. It refers to situations where people are obliged to leave their homes or places of habitual residence as a result of a disaster or in order to avoid the impact of an immediate and foreseeable natural hazard, including the adverse impacts of climate change, or a disaster triggered by human-made factors, such as large-scale industrial accidents. Displacement triggered by conflict is not considered disaster displacement¹⁵. If the citizens do not quickly return to their original places of residence and are involved in the reconstruction of the material and social environment, the loss of human, economic and social capital¹⁶ can lead to the abandonment of the area. When individuals, firms and service providers must relocate somewhere else, there are net losses in place of community resilience, with parallel increases in vulnerabilities. Infrastructure resilience, i.e., the extent to which there is a restoration of minimum services and functions in a relatively short amount of time, thus allowing people to return to their place of origin¹⁷, is a first step towards the restoration of communities. The importance of involving citizens and communities in the reconstruction process after the initial evacuation is clearly visible in the case of the L'Aquila earthquake. Imperiale and Vanclay¹⁸ show how the exclusion of local communities and the top-down planning approach chosen for the reconstruction led to more than 10,000 people still in temporary accommodations 10 years after the earthquake, with many citizens leaving the city and never going back. When population leaves the affected area in a permanent way, this contributes to a decrease in the capacity to cope with additional shocks^{16,19}. A vulnerable community can suffer from changes in population in response to direct and indirect impacts, and these variations are hardly stable over time. Usually, variations in population size in a given area are different between the aftermath of the disaster (short-term changes), some months later (mid-term changes), and in the long term, when the reconstruction phase has started because the emergency is over (in this study we are not considering permanent relocation, which might be necessary in some cases). This change is seldom visible in official census data because individuals can autonomously find accommodations without informing the authorities when they flee the disaster area. In other cases, the devastation is so profound that the census records are difficult to acquire and update. In this situation, it is fundamental to have an accurate picture of the number of displaced people and where they have relocated. If a system exhibits the capacity to 'return to equilibrium after a displacement'²⁰, then that system is considered resilient. Displacement is relevant in two of the four priorities outlined in the Sendai Framework for Disaster Risk Reduction 2015–2030 (Sendai Framework), the first agreement of the post-2015 development agenda that provides Member States with concrete actions to protect from the risk of disaster. This study responds to objective (h) in Priority 4 (Enhancing disaster preparedness for effective response and to “Build Back Better” in recovery, rehabilitation and reconstruction), which advocates disaster preparedness and the establishment of rapid and effective responses to disasters and related displacement.

More specifically, this study combines engineering methods to estimate post-disaster displacement using mobile phone data, with a social sciences perspectives on economic factors and community resilience. There are two reasons for this approach.

1. While there exist alternative approaches to estimate displacement, i.e., using aerial images²¹, night-light satellite images²², social network, crowd-sourced and damage assessment data^{23–25}, novel data sources can integrate existing information and provide further insights. Mobile phone data (often called Call Detail Records - CDRs) is a natural candidate in this regard as it combines ready available data with an almost complete population coverage.
2. Information about CDRs (short-term) can then be integrated with other elements, for instance information about the economic structure of the area or about the reconstruction process (mid-term). Understanding how factors other than the sheer magnitude of the disaster might be contributing to displacement is essential for the design of policies that can effectively counteract population displacement and can improve community resilience.

Thanks to this novel combination of data and methods, we aim to answer the following two research questions:

1. *How to quantify population displacement from fully aggregated mobile-phone data?* Most of the approaches reported in the literature^{26–31} derive displacement information directly from the data (i.e., looking at where each mobile phone moves after the disaster). Using fully-aggregated data, this kind of approach is unusable as privacy regulations do not allow to track phones but only to measure their presence in an area (see Section “Methods”). Therefore, in this study we had to develop a novel approach to quantify displacements. While we applied our method to a specific case study - the earthquake that struck Central Italy in 2016- the approach discussed in this study is more general, and can be implemented it in other emergencies once data are available—see Section “Methods”.
2. *What is the relationship between population displacement, economic activities and infrastructural damages in the selected case study?* After measuring displacement resulting from the Central Italy earthquake, it is interesting to understand the main factors impacting it. Proximity to the earthquake epicentre and the size of damage suffered naturally have a high impact, but there are other important factors. In particular, we focus on the geographical and economic characteristics of the affected areas. Because disasters affect various economic activities in a heterogeneous way, population changes are heterogeneous across municipalities. In our case study, manufacturing or logistics, for instance, require physical infrastructures that are easily

damaged by an earthquake. Moreover, these industries present more complex inter-dependencies among production plants and supply chains, all features heavily disrupted by a disaster. Therefore, municipalities with an high number of such companies are likely to suffer more displacement because of the destruction.

In this study, we focus on CDRs (Call Detail Records), which offer a quick way to obtain data about post-disaster mobility. CDRs are automatically and routinely collected by telecom operators and contain approximate information about the location of every mobile phone connected to their network^{32–36}. It is relevant to notice that while the use of mobile phone data at the *individual*-level (i.e., tracking individual mobile phones) is very much constrained due to privacy reasons, the use of *aggregated and fully anonymized* mobile phone data is easily accessible. Most telecommunication operators have commercial offers to provide aggregated and anonymous data about people's presence and mobility. The main idea is that mobile phones precisely localize a substantial fraction of the population with minimal cost and intervention. Unlike census data or data collected by municipalities at selected time points, CDRs provide a continuous estimate of the situation, allowing researchers to identify changes in relocation patterns and displacement as they happen.

In the literature there are many examples on the use of CDR data in the context of disaster risk reduction. Counter-intuitively, the vast majority of the works presented in the literature focus on individual-level data (that is hard to acquire), while they often disregard aggregated-level data.

Many scholars have focused on describing mobility patterns to understand post-disaster population displacement at different periods in time. In²⁸, mobile phone data allowed for predicting people's mobility during a disaster. The authors analysed data about mobile phones 42 days before and 341 days after the Haiti earthquake. Their results show that the earthquake caused an increase in population movement, an increase in daily distances in the affected areas and more heterogeneity in terms of average travel distance. They also found that people's movement was more predictable after the disaster because of the high correlation with the places where people had strong social relationships. Similarly, in another work on the Haiti earthquake²⁹, authors found that 20% of the population emigrate from Port-au-Prince during the first 19 days after the earthquake. To validate the results, the authors compared with the estimates of the National Civil Protection Agency based on the number of ships and buses leaving the area and the survey performed by the United Nations about migration after the earthquake. Thanks to mobile phone data, the authors found the regions where most of the population migrated, thus offering support to policymakers for planning and relief efforts. Similar analysis performed in the context of the Nepal earthquake in 2015³⁷, the landslide in Japan in 2014²⁶, or the Mozambique cyclone in 2019³⁸ show the evolution of population mobility patterns after the disasters and the return patterns to affected areas in the following days and weeks. In a recent study on mobile phone data published by³¹, the authors suggest the importance of inter-city connectivity in post-disaster recovery. Using mobile phone call data records, the authors show that the cities with high inflows and outflows of people before the disaster were more resilient and recovered faster.

Some recent works started comparative analysis among multiple disasters analyzing pre- and post-disaster individual mobile phone trajectories to test the generality of this approach^{39,40}. In²⁷, mobile phone data enabled the study of the population recovery patterns after extreme events. The authors found very similar behaviours in five different locations across world. They explained the differences using income, population, housing damage rate and connectivity with other cities. When they analysed the recovery time distribution, they found that most people returned to their homes within two weeks. Furthermore, they tried to explain the displacement rate based on socioeconomic variables, infrastructure recovery and accessibility. The results showed that, at different times, the infrastructure recovery became more relevant compared to the house's damage rates.

The works in^{41,42} analyze *aggregated* mobile phone data in the context of the earthquake that shook Ecuador on April 2016. The analyzed data contains information about the *home area* of the monitored phones, which allows estimating displacements more easily. Similarly, the work in⁴³ is one of the few dealing with aggregated mobile phone data. They identify anomalies in the data collected during multiple events affecting some US states in 2018 (e.g., hurricanes Florence and Michael).

Mobile phone data can indirectly measure the time needed for a given area to recover. The Resilience to Emergencies and Disaster Index proposed by⁴⁴ assesses the recovery period of neighbourhoods impacted by a storm to prioritize limited resource use. The authors considered mobile phone data as a proxy of neighbourhood activity, social infrastructure, community connectivity, physical infrastructure, economic strength and environmental conditions as distinct variables and computed the indicator as the normalized weighted sum of these variables. According to⁴⁵, mobile phone data alone does not explain the motivation why people move and relocate in a certain way after extreme events. Thus, it is necessary to perform data fusion on different datasets or to combine big data with smaller datasets, such as surveys. In this regard³⁰, analysed migration patterns based on mobile phone data and the national survey that included questions on temporary and seasonal migration. Furthermore, the authors underline the importance of considering short-term and long-term displaced people in the definition of migration. Their results evidence that temporary migration can be detected using mobile phone data.

Our research contributes to this consolidated body of work by providing new insights into data and methods. More specifically:

1. Unlike most papers discussed so far, we do not use data that can be associated with a single mobile phone because this would posit several limitations in terms of privacy and would hinder the actual applicability and reproducibility of our method and analysis. Our approach allows us to overcome this limitation by using anonymous and aggregated Call Detail Records (CDRs) that telecom companies routinely collect and sell for business/commercial purposes. For this reason, we acquired data from a wide area comprising four regions, 663 municipalities and spanning three years – such a large extent is uncommon in the literature. Such aggregated data required us to develop a novel method to measure population displacement based on

- non-autoregressive time-series forecasting. Although such a forecasting approach is not novel per-se, the application to mobile phone data is new (see Section “Methods”).
- In this study, we combine CDRs data with data about economic activities, companies’ information and government funding for reconstruction. Using a large dataset on commercial and industrial companies of any size and sector, we introduce in the model very detailed and relevant information on the kind of economic activities present in a given municipality. Information about the industrial composition at the municipality level is valuable to embed the effects of the disaster into the existing economic setting. As an additional step, we link CDRs with data about the funding requests for the reconstruction of residential, commercial and industrial buildings to understand damages and mid-term trends in people’s returning home. These datasets offer a multifaceted picture of post-disaster damage and displacement and suggest relevant directions for policymakers.

The case study

What is known as the ‘Central Italy earthquake’ is a sequence of different shocks happening over a period of five months in 2016⁴⁶. The first earthquake hit at 3.36 a.m. on August 24, 2016, with a 5.9 magnitude. The affected area, administratively divided under four Italian regions, consists of Abruzzo, Lazio, Marche and Umbria. During three days, above 2000 aftershocks were recorded following the main seismic event, extending along 25 kilometres through Amatrice and Norcia. The largest aftershock occurred near Norcia (5.3 Mw) at 04:33 on August 24, 2016. The Italian Civil Protection reported that 299 people died. On September 5, 2016, the number of people needing assistance was 4807. On October 26 and 30, new violent shocks rock the same area, especially between Umbria and Marche, already markedly affected by the August 24 earthquake. Specifically, the October 26 event consisted of two 5.4 and 5.9 magnitude earthquakes. On October 30, a new shake caused the collapse of buildings⁴⁷. On January 18, four earthquakes of magnitude higher than 5 hit the area, and Lazio and Abruzzo’s regions were the most affected. In particular, a 5.1 event was registered at 10:25, followed by other earthquakes: the second at 11:14 (5.5 Richter magnitude), the third at 11:25 (5.4 Richter) and the fourth at 14:33 (5 Richter). The affected territory is mountainous, with more than 70 per cent of the surface located at an altitude higher than 900 metres. This mountainous configuration means that there is an uneven distribution of urban areas: the majority of the inhabitants live in several small centres surrounded by widespread rural zones within a mountainous landscape, most of which are isolated and accessible only by minor roads⁴⁸. According to Decree-Law 17 October 2016, n. 189, 140 municipalities that fall into the “seismic crater” are entitled to government funding for private and public reconstruction. Overall, this sequence of earthquakes heavily affected the whole area, although the 140 municipalities close to the earthquakes’ epicentres—referred to as the *seismic crater*, were the most damaged. According to governmental reports⁴⁹, the total population living there before August 2016 was 581.885. Concerning the estimate of the costs of the Central Italy earthquakes, the Italian Government provided them in June 2022. The estimated request for funding at the time was of 22.695 units, corresponding to a value of 7.669.353.627 euros. The expected amount of further requests is 27.350 for 11.794.161.941 euros, reaching 50.045 units in total that need reconstruction⁴⁹.

Results

Our analytical strategy consists of the following steps:

- Measuring the impact of the earthquake in the number of mobile phones present in the area. The objective is to evaluate the relationship between this variation and the people’s presence in the area in the short-term. We propose the Mean Prediction Error (MPE) to compute the displacement in population, and assess how strongly this correlates with the observed changes in population according to census data.
- Analyzing the relationship between MPE and other important variables, namely reported damages, distance from the epicentre, structural characteristics, and economic activities, to explain variations in population displacement in the mid-term.

Monitoring population via CDRs

In the first set of experiments, we use mobile phone data to quantify people’s displacement after the earthquake. Figure 1A compares the number of residents in the municipalities of the area under study with the corresponding data from mobile phones (computed as the average number of phones at night – see Section “Methods”). There is a clear linear relationship among the two measures ($\beta = 1.17$ (0.01), $R^2 = 0.95$), confirming the representativeness of the mobile data. Figure 1B compares the percentage change in residents’ population as measured by official statistics between 2017 and 2016 and the change in residents’ population as measured by mobile phone data. In particular, we calculate the number of resident changes as measured by official statistics (OCHG—Official CHanGe) as follows:

$$OCHG = \frac{\text{n.residents end 2017} - \text{n.residents beginning 2016}}{\text{n.residents beginning 2016}}$$

Similarly, we calculate the change in residents’ population measured by mobile phone data (PCHG—Phone CHanGe) as:

$$PCHG = \frac{\text{avg. n. phone at night after earthquake} - \text{avg. n. phone at night before earthquake}}{\text{avg. n. phone at night before the earthquake}}$$

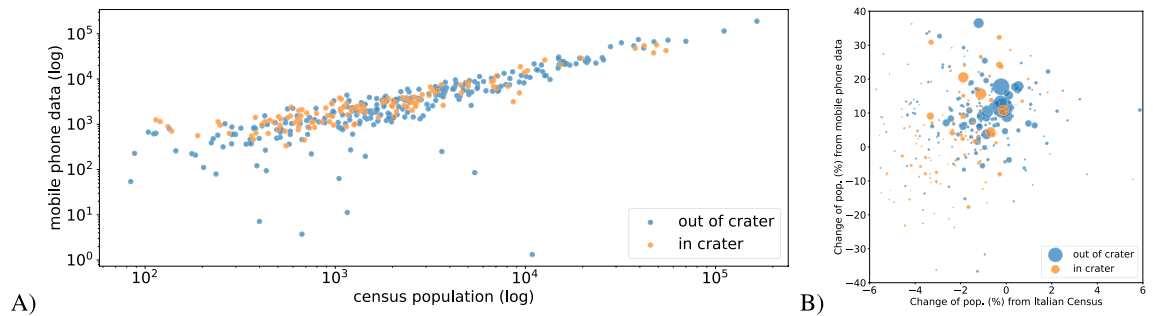


Figure 1. Each dot represents a municipality. Blue dots are outside the earthquake crater area, and orange dots are inside the crater area. **(A)** People count from the Italian census (ISTAT) and mobile phone data. **(B)** Population change from the Italian census (ISTAT) and mobile phone data, the size of dots represent the municipality population.

There is almost no relationship between the two measures. Looking at the small range of variation in official statistics and considering the destruction of some municipalities by the earthquake, we can safely assume that official statistics do not provide an accurate population dynamics estimate. In these extreme situations, people do not promptly communicate their change in residency. However, in the context of earthquakes and natural hazards, a timely representation of people displacement is critical to organize relief activities and prioritizing resources. Therefore our results show the importance of using CDRs *per-se* in a more precise manner for monitoring population dynamics.

To understand the impact of the earthquake in the area, we trained a non-autoregressive forecasting model to predict population dynamics as measured by mobile phones (see Section “Methods”). We trained the model on individual municipalities with data collected before the first earthquake hit the region (August 24, 2016). Then, we compared the prediction results with actual data, using ‘testing’ data after the first earthquake hit. Figure 2 illustrates the results in Accumoli, one of the municipalities hit the hardest by the earthquake. It is possible to see that forecast (orange) is much higher than the actual data (blue). The idea of this approach is that, since the forecasting algorithm receives only data *before* the earthquake happened, the forecast results (orange line) can be interpreted as the counterfactual scenario as if the earthquake had not occurred. In the “Discussion” Section, we discuss the limitations of this idea. The main advantage of MPE over PCHG is that MPE automatically deals with data seasonality.

Following this approach, for each municipality, we compute the mean (percentage) prediction error (MPE) as:

$$MPE = \frac{1}{N} \sum_t \frac{\text{actual}_t - \text{forecast}_t}{\text{forecast}_t}.$$

In this way, if forecast values are higher than actual ones, the *MPE* is negative and represents a percentage decrease in the population. Since there were multiple earthquakes spanning approximately six months, we further decompose MPE in two time-bound measures.

We compute *MPE INIT* focusing on ‘testing’ data from 25 August 2016 to 30 October 2016. This time frame coincides with the period between the aftermath of the initial shake (24 August 2016) and the two months after it. We also compute *MPE END* focusing on ‘testing’ data from 1 April 2017 to 30 October 2017. In this period, the medium-term effects of the earthquake become visible, allowing for an interesting comparison between the two-time frames. We chose these time frames to split between *MPE INIT* and *MPE END* according to data availability.

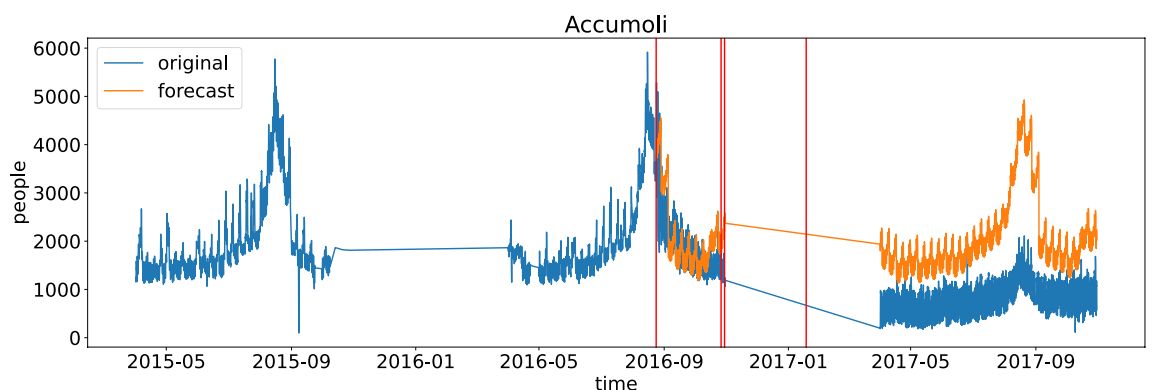


Figure 2. Mobile phone data was collected in the Accumoli municipality (blue). Forecast data based on the pattern exhibited before the earthquakes (orange). Red vertical lines are associated with the four main earthquakes that hit the region. Actual data is 50% lower than forecast ones.

Figure 3A shows the relationship between MPE and PCHG and the good linear relationship between the two variables ($\beta = 0.72$ (0.02), $R^2 = 0.64$). Figure 3B shows MPE results for all the municipalities. In this case, the variation in mobile phone data proves accurate. The area closer to the epicentre and mostly affected (the so called *seismic crater*) has the highest negative MPEs indicating that the municipalities had the highest decrease in population.

Similarly, Figure 3C shows a map of central Italy (left) and focuses on the earthquake crater area (right). It is possible to see that, in general, the crater area had the most significant reduction in population size, measured as the highest negative MPEs. However, there is high variability, and some municipalities in the seismic area witnessed an increase in their population.

Population change, economic activities and damage impact

In this second set of experiments, we create linear models to explain the variation of MPE in terms of geographic and economic variables. Given the variability in MPE (and PCGH) observed in Fig. 3, we created different models to explore the origin of this variability. Our intuition is that other than the earthquake magnitude and associated damages, the kind of economic activities in each municipality might impact population dynamics. For example, since earthquake usually affects built infrastructures, agricultural activities are likely to be less affected than manufacturing or logistics, but also because the latter typically involves more complex inter-dependencies among production plants, supply chains and logistics.

The following analysis provides evidence about this link between economic activities and displacement in municipalities. We developed three models (linear regression) to explain changes in MPE in the affected areas.

1. The model in Table 1 uses *geographic features* and *earthquake damage features* as covariates (independent variables). *Geographic features* of the municipality are the distance from the epicentre and median altitude (ALT MED). The epicentre distance can be a simple proxy of the effects of the earthquake on a given area

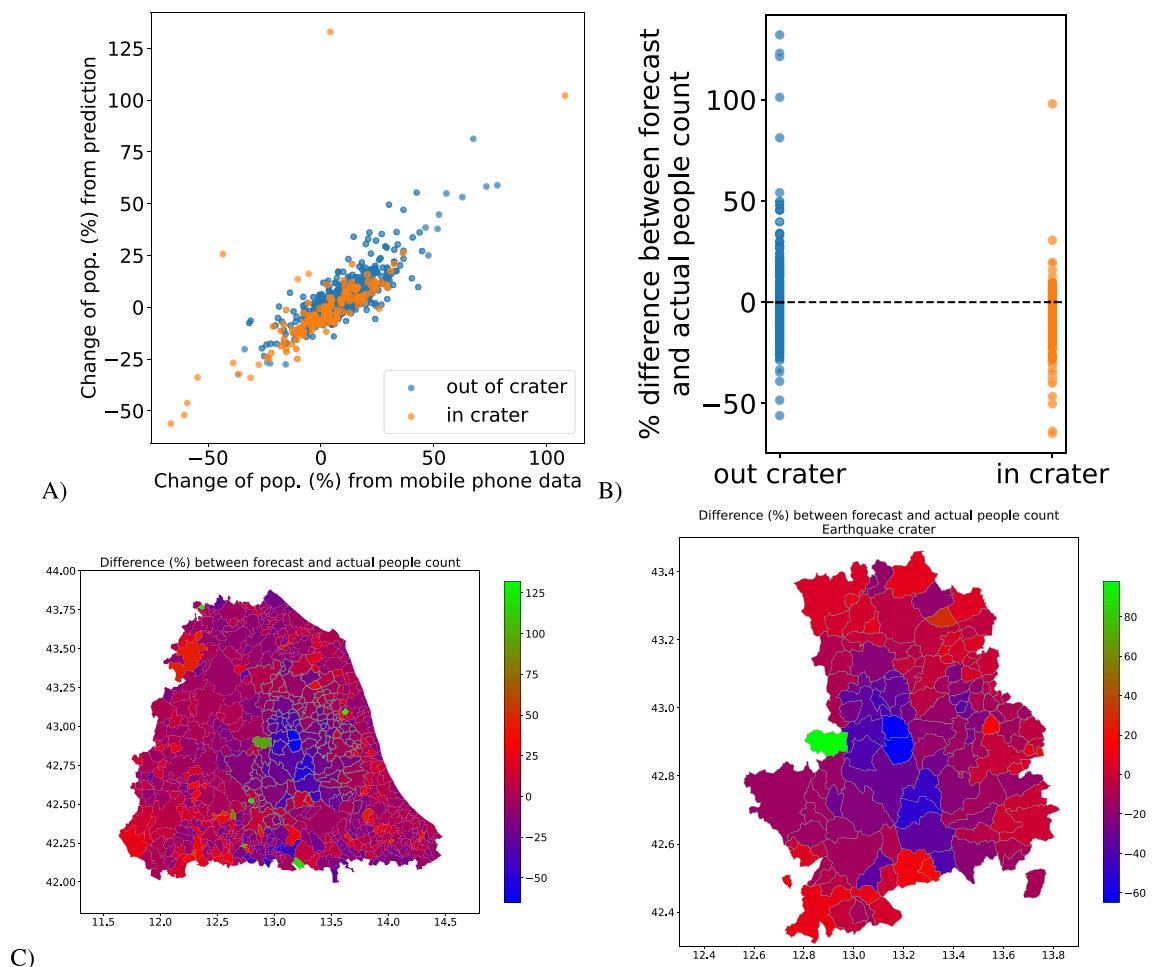


Figure 3. Each dot represents a municipality. Blue dots are municipalities outside the earthquake crater area, and orange dots are those inside the crater area. (A) Relationship between PCHG and MPE. (B) MPE for all municipalities under study. (C) Map, showing MPE for the Central Italy municipalities (left) and zoom in on the crater (right). Maps have been drawn using python libraries: *matplotlib* 3.5.2 and *geopandas* 0.2.12 – <https://www.anaconda.com>.

	MPE INIT	MPE END	MPE
Distance from epicenter	0.308***	0.602***	0.532***
	(0.083)	(0.168)	(0.134)
ALT MED	0.001	0.001	0.001
	(0.003)	(0.005)	(0.004)
RCR PRIV	-20.150*	-10.629	-12.904
	(11.831)	(23.369)	(18.611)
RCR PUB	-110.132**	-369.151***	-307.251***
	(48.978)	(99.943)	(79.503)
Observations	123	123	123
R ²	0.467	0.472	0.520
Adjusted R ²	0.449	0.454	0.504
Residual Std. Error	6.753	13.781	10.962
F Statistic	25.899***	26.349***	31.948***

Table 1. Model to explain MPE (MPE INIT, MPE END, MPE) in terms of geographic and damage features. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

- (of course, there are many more factors determining the consequences of the earthquake). In addition, given that the seismic area comprises several small centres surrounded by widespread rural and mountainous zones, most of which are isolated and accessible only by minor roads, we considered the municipality's (median) altitude as a simple proxy for municipality isolation. *Earthquake damage features* are the number of requests for rebuilding (RCR) from private (RCR PRIV) and public (RCR PUB) entities that are a proxy for earthquake destruction (see Section "Methods").
- The model in Table 2 uses *geographic features* and *economic sectors' features*. The considered classification of economic sectors is the ATECO (1-letter code) which contains unique identification codes for every economic activity. To limit the number of variables, we consider only those sectors contributing to at least 1% of the total number of companies and 1% of the total revenue of the area (see Section "Methods", also for comparison with the results with all the sectors considered).
 - The last model (Table 3) combines all the previous covariates: *geographic features*, *earthquake damage features* and *economic sectors' features*.

In Table 1, we see a positive relationship between the distance from the epicentre and MPE in all three measures, indicating the accuracy of MPE in capturing the displacement of inhabitants, both in the first response to the event as well as a later one. The number of requests for rebuilding from the public sector (RCR PUB) is significant and negative: the higher the amount of RCR, the lower the MPE (greater negative MPE, e.g., -50% in cell phone presence), meaning that there is a higher number of displaced people, probably because of the destruction. The number of requests for rebuilding from the private sector is significant (and negative) only for MPE INIT. RCR PRIV is a proxy for the damage to homes and private buildings. This damage implies the immediate displacement of people. Therefore, MPE INIT better captures displacement, while RCR PUB represents public buildings and infrastructures. Finally, MPE END represents the impact on the long-term habitability of the area. In this analysis, we considered the number of requests presented for funding. We conducted a similar study using the initial assessment of the damage of the total amount in Euros of the requests obtaining similar results (see Section "Methods").

In Table 2, we added the industrial sectors to explore what might affect MPE (a short description of the variables associated with industrial sectors in the regression is in Table 4). First, the distance from the epicentre positively affects MPE in all three measures. The altitude of the municipality is significant for MPE END and MPE, indicating that locations with higher altitudes (likely to be more rural and less connected) have a higher decrease in population. Concerning economic sectors, we observe that constructions, hospitality and manufacturing affect MPE INIT. The effect is present, also in MPE END for logistics, manufacturing, whole retail sales and finance, but not for hospitality and construction. These two sectors show no significant impact on MPE END. This effect may be because community resilience does not relate to hospitality and construction. However, these activities relate to the immediate responses to the disaster, such as rescue operations and activities to secure crumbling buildings and houses. Therefore, they can drive an increase in the number of people in the area. The evidence presented here suggests that more modern economic sectors (secondary and tertiary), which require supply chains, logistics and interconnections with other sectors and geographical areas, have been heavily affected by the disaster, and this effect is long-lasting. Furthermore, this suggests the need for more active policy interventions to support the recovery of these industries and economic sectors.

Finally, the coefficients in Table 3 support the previous conclusions. With more variables, the model has a greater explanatory power (R^2). Figure 4 shows the model result with the actual MPE, MPE INIT and MPE END and the predicted ones. The reported graph illustrates the fitting of the models. It is worth explaining that the model is fitted on 123 municipalities (observations) of the seismic crater, while official reports identified 140 in that area. This mismatch is due to missing data for 17 (12%) municipalities. We did not consider these municipalities in the model.

	MPE INIT	MPE END	MPE
Distance from epicenter	0.437*** (0.078)	0.833*** (0.158)	0.739*** (0.127)
ALT MED	-0.002 (0.003)	-0.012** (0.006)	-0.010** (0.005)
Agriculture	-5.893 (4.888)	-8.158 (9.838)	-7.617 (7.942)
Real estate	2.358 (9.329)	-18.597 (18.777)	-13.590 (15.158)
Hospitality	-11.096* (6.271)	-4.085 (12.621)	-5.760 (10.189)
Manufacturing	-6.869* (3.546)	-18.696** (7.138)	-15.870*** (5.762)
Tech services	-7.535 (11.358)	-26.066 (22.860)	-21.638 (18.454)
Whole/retail sales	-4.785 (3.998)	-14.255* (8.047)	-11.992* (6.496)
Constructions	-8.232** (3.801)	-5.361 (7.650)	-6.047 (6.176)
Business support services	-3.342 (22.615)	-6.245 (45.518)	-5.551 (36.744)
Communication and info serv.	19.008 (49.252)	-46.248 (99.131)	-30.653 (80.024)
Logistics	-12.423 (10.323)	-46.660** (20.776)	-38.478** (16.772)
Healthcare	-28.748 (33.740)	-25.464 (67.910)	-26.249 (54.820)
Finance	-2.998 (7.261)	-38.056** (14.614)	-29.678** (11.797)
Utilities	-9.577 (10.862)	-31.936 (21.863)	-26.593 (17.649)
Observations	123	123	123
R^2	0.467	0.485	0.518
Adjusted R^2	0.392	0.413	0.451
Residual Std. Error	7.098	14.287	11.533
F Statistic	6.239***	6.723***	7.671***

Table 2. Model to explain MPE (MPE INIT, MPE END, MPE) in terms of geographic and economic features. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Discussion

When sizeable portions of the population move out because of a disaster, massive economic and social losses are observed in the affected areas. Being able to quickly and reliably measure these changes inexpensively and unobtrusively is very important. This study shows that aggregated mobile phone data can represent a reliable indication of population variation in the aftermath of an earthquake. Moreover, the combination of CDR with the census, economic and administrative data from four Italian regions affected by repeated earthquakes in 2016 allows us to identify key characteristics that can help to explain displacement dynamics. Our results show that analyzing CDRs can be extremely useful in the aftermath of the disaster, but also that they can be fruitfully related to other data sources related to the reconstruction process.

Mobile phone data (CDRs) can naturally enrich and complement other approaches to estimate displacement (e.g. aerial images²¹, night-lights satellite images²², social network, crowd-sourced and damage assessment data^{23–25}) offering ready availability, extremely long monitoring periods, and almost complete coverage of the population. Differently than mobile phone data *at the individual level*, which are very hard to get due to privacy regulations, *aggregated* data are readily available without privacy concerns. The method proposed here for the analysis of CDRs data offers the opportunity of knowing soon after the disaster (the time frame depends on data availability) how many people moved out of the affected area (measured by the introduced MPE value).

The approach used to compute this displacement is one of the main novelties of our work. We calculate the value MPE as the mean percentage prediction error between the actual data and a non-autoregressive forecasting model trained with only data collected before the earthquake hit. As the model never receives data during or after the earthquake, we interpret its forecast as the counterfactual scenario as if the earthquake had not happened⁵⁰.

	MPE INIT	MPE END	MPE
Distance from epicenter	0.274*** (0.086)	0.560*** (0.171)	0.492*** (0.136)
ALT MED	0.001 (0.003)	-0.004 (0.006)	-0.003 (0.005)
RCR PRIV	-20.711* (11.799)	-11.338 (23.418)	-13.578 (18.624)
RCR PUB	-153.953*** (51.343)	-420.597*** (101.904)	-356.876*** (81.041)
Agriculture	-5.824 (4.653)	-10.518 (9.236)	-9.396 (7.345)
Real estate	1.440 (8.734)	-20.759 (17.335)	-15.454 (13.786)
Hospitality	-14.725** (5.942)	-10.620 (11.793)	-11.601 (9.379)
Manufacturing	-8.457** (3.348)	-22.601*** (6.645)	-19.221*** (5.285)
Tech services	-10.207 (10.681)	-33.978 (21.198)	-28.297* (16.858)
Whole/retail sales	-7.815** (3.836)	-21.728*** (7.613)	-18.403*** (6.054)
Constructions	-11.890*** (3.717)	-14.706** (7.378)	-14.033** (5.867)
Business support services	-9.999 (21.258)	-23.426 (42.192)	-20.218 (33.554)
Communication and info serv.	12.009 (46.200)	-52.880 (91.695)	-37.373 (72.923)
Logistics	-9.130 (9.693)	-40.085** (19.239)	-32.687** (15.300)
Healthcare	-41.674 (31.730)	-52.623 (62.976)	-50.006 (50.083)
Finance	-5.110 (6.880)	-45.172*** (13.654)	-35.598*** (10.859)
Utilities	-12.181 (10.190)	-38.209* (20.224)	-31.989** (16.083)
Observations	123	123	123
R^2	0.542	0.570	0.609
Adjusted R^2	0.467	0.500	0.546
Residual Std. error	6.643	13.184	10.485
F statistic	7.296***	8.180***	9.627***

Table 3. Model to explain MPE (MPE INIT, MPE END, MPE) in terms of geographic, damage and economic features. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

We do not claim this to be a full-fledged causal analysis, and we are aware that the results might be partial: there are likely some confounding factors not accounted for by our model. Nonetheless, we are confident that our analysis captures, at least, part of the counterfactual scenario.

This displacement estimate based on mobile phone data can be extremely useful in emergency intervention planning, especially in remote or widely spread areas. In the initial emergency stage, knowing how many people need alternative accommodations, how many should receive which forms of economic support and where there is a high expectation of population change are all relevant variables that can stir the emergency management phase in different directions.

Third, our study shows how fruitful it is to combine CDRs data with existing datasets to explore which characteristics of the disaster and the area can explain variations in the number of residents. In our study, the calculation of MPE and its combination with data about the economic activities in the affected areas shows that some areas are more vulnerable than others, and further studies would help understand better the extent of these vulnerabilities. The need to identify vulnerable areas is especially relevant in a country like Italy that, because of the particular morphology and geophysical context of its territory, is one of the five European countries with the highest probability of a disaster and related economic loss⁵¹. Readily available information about the impact of

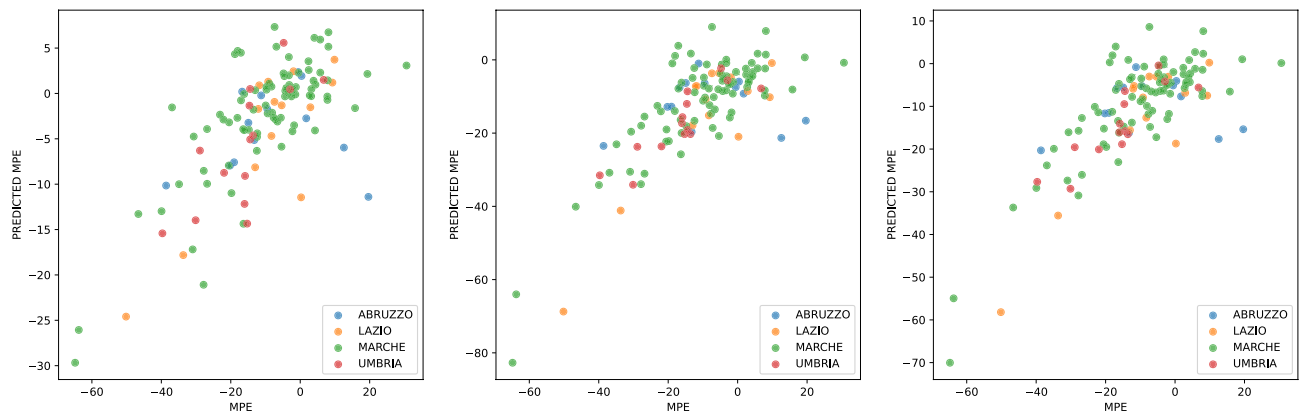


Figure 4. Actual and regressed MPE, MPE INIT and MPE END in the earthquake crater zone using the model with covariates: *geographic features, earthquake damage features and economic sectors' features.*

an earthquake in terms of immediate displacement can be part of strategic planning for temporary housing that identifies opportunities and constraints for temporary housing before the disaster⁵². Local, regional and national institutions face what has been called the “post-disaster recovery dilemmas”, i.e., the need to balance short-term and long-term needs for vulnerability reduction⁵². There are no obvious solution to such an issue, which greatly varies depending on the situation, but our results can inform policies aimed to support financially and logistically those industries most affected by the disaster, those that are supposedly more relevant for the area’s resilience and the country’s economic growth. Our results are limited in time, location, and kind of disaster, but they can provide a valuable avenue for future research. Even if our methodology needs to be further refined, we point out the relevance of the indications about the relationship between the kind of economic activities and the funding requests. In our data, hospitality is the sector that suffered the least, while the most affected municipalities are those characterized by construction and manufacturing activities. This finding could indicate that, in vulnerable areas, some economic activities could be more resilient to disasters, and their presence can positively affect the reconstruction.

In addition, by using administrative data about the requested funds to rebuild homes and combining it with economic data and MPE, we can offer an overview of the reconstruction process. By calculating MPE, the effective damage and the intention to stay (as the requests for funds necessary to rebuild or repair the damaged houses), we observe that MPE seems to be a good predictor of the latter variable. This finding confirms its efficacy as a medium-term indicator in a situation in which displaced citizens have relocated without registering as residents in another municipality, thus creating difficulties for institutional actors to know the actual magnitude of the displacement and to take adequate measures to assist the population.

For instance, when deciding where to locate aid centres and shelters, it is important to know where most of the population has moved. This data can enrich the vast pool of information (e.g., established emergency plans, conditions of buildings, roads, and transports) to guide and optimize the positioning of such facilities.

A further avenue for future research is using data about the post-earthquake reconstruction process that the Italian government has made available for three different earthquakes that happened in the last 15 years (L’Aquila, Emilia-Romagna and Central Italy). Our results indicate that using data about expected and actual funding requests for rebuilding could offer a valuable indication of displacement and its medium-term consequences. It is worth stressing that physical reconstruction does not necessarily lead to community recovery, and several factors can affect how much communities can “build back better”^{18,53}. However, our data and analysis do not allow us to draw any conclusions about community recovery. Qualitative and quantitative individual data, like interviews, focus groups and surveys, would be needed to investigate the relationship between infrastructural reconstruction and community restoration.

As a final remark, we want to point out that in this work, we use data from the Telecom operator in 2015–2017, as the earthquake hit the region during that period. Mobile phone data help analyze displacement right after the disaster and in the months and few years after the fact when official data are often missing or incomplete. Longer-term observations (e.g., using data available now - 6 years after the fact) are less useful as official statistics and data about the reconstruction take the lead role in such a time horizon. Nevertheless, the proposed methods can be applied to other disasters in other years without modifications, and data gathered after different kinds of disasters across diverse countries would be necessary to ground the validity of this approach.

Conclusions

From the above analysis, it is possible to draw the following conclusions which answer the stated research questions: (i) completely anonymized and aggregated mobile phone data can monitor population displacement after disaster scenarios. The proposed method computes displacement by comparing the actual mobile phone count with the forecast computed before the earthquake hit (see Fig. 3). (ii) Geographic features (epicentre distance and altitude) and damage-related features (public and private RCR) explain part of population displacement (see Table 1), (iii) also the industrial base and composition of the regions contribute to explaining part of population displacement (see Table 2). A better-performing model combines the above features (see Table 3).

Methods and supplementary analysis

Data

In this work, we processed four different sources of data to extract and analyze behavioural patterns before and after the earthquake in central Italy in 2016: (i) aggregated Call Detail Records; (ii) economic activity data; (iii) data about damage and recovery activities and (iv) census and demographic data.

On the one hand, aggregated Call Detail Records are the data at the basis of the proposed approach. They allow us to compute population displacement (MPE) at a fine-grained scale. In principle, they are available in quasi-real-time for emergencies and provide timely and valuable information. On the other hand: economic activity, damage, recovery, census and demographic data are functional to the analysis presented in the results to explain the factors affecting MPE.

Call detail records

We obtained from TIM, the largest mobile phone operator in Italy at the time of the data collected (30.2% market share⁵⁴), a large dataset of *aggregated* CDRs (Call Detail Records) measuring the number of cell phones localized by the network in a given area of the territory. Data covers the whole of Central Italy, divided into a matrix of small square areas of 150 meters side, called pixels. The dataset consists of the number of devices in each pixel every 15 minutes. The localization procedure is proprietary to the telecom operator. Localization is based on signal strength triangulation at the different antennas (see Fig. 5). The data contains only people count. It is not possible to reconstruct the movement of individuals. In addition, to remove the possibility of singling out individuals, the value of any pixel with fewer than N individuals (typically $N = 4$) is set to zero. As the data measures mobile phones from a single operator, we linearly scale numbers to consider the operators' market share in the area. Nevertheless, our data do not fully compensate for people not having mobile phones (e.g., the presence of children and elderly). Further research would be needed to address this aspect. However, for the presented results, it is worth considering that: (i) Italy is one of the countries with the higher number of mobile phone operations, (ii) our data strongly correlates with census data—see Fig. 1A— hinting to a limited effect of this bias.

The mobile phone dataset spans the following periods: from 2015-04-01 to 2015-10-30, from 2016-04-01 to 2016-10-30 and from 2017-04-01 to 2017-10-30, to capture presence before and after the earthquake.

To make the dataset more manageable and have areas compared with other datasets, we aggregated data at the municipality level and the 1-hour time interval. This procedure reduces possible biases due to setting small counts to zero, as the aggregation over whole municipalities leads to higher counts.

Overall, we have data for 663 municipalities with 14,420 hourly measures each. Figure 6 shows an example of the data collected for each municipality.

We want to emphasize again that we deployed fully aggregated data. While individual mobile phone traces are very privacy sensitive, as they allow tracking single phones and, therefore, users, aggregated CDRs counting the number of phones in a given area is privacy compliant as the aggregate number does not allow to single out individuals. The operator carried out the entire aggregation process in compliance with privacy regulation (GDPR), and most telecom operators have commercial activities to exploit this kind of data. The possibility to

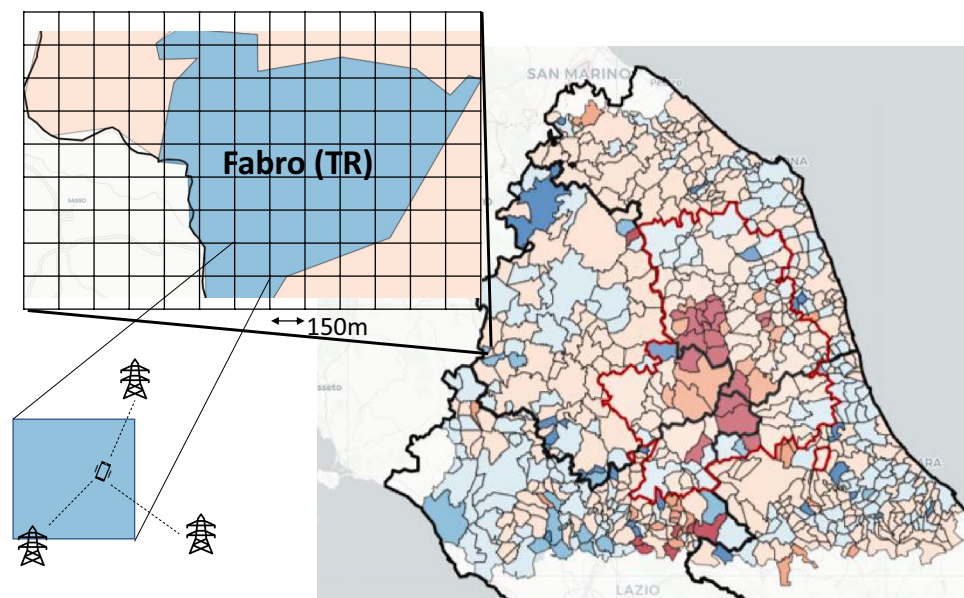


Figure 5. Cell phones are localized by the telecom operator in an area of 150 * 150 meters, called a pixel. Aggregated counts for each pixel is the data provided by the telecom operator in compliance with privacy regulation. Data have been further aggregated at the municipality level to make it more manageable and have areas comparable with other datasets. Maps have been drawn using python libraries: *matplotlib* 3.5.2, *geopandas* 0.2.12 and *folium* 0.14.0—<https://www.anaconda.com>.

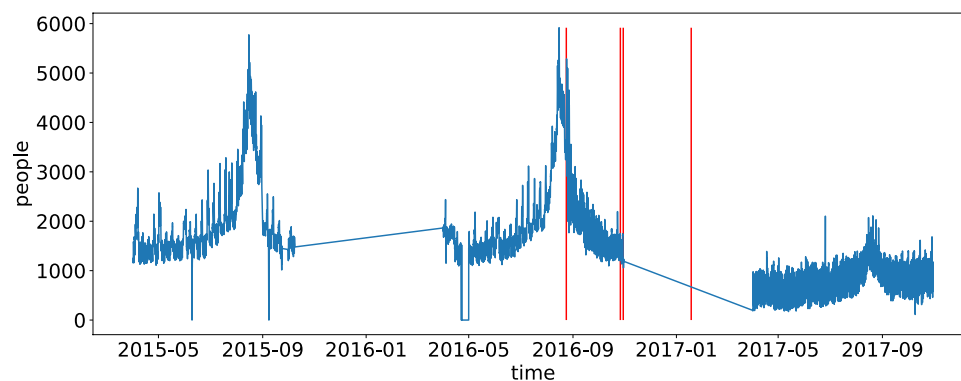


Figure 6. Mobile phone collected for the municipality of Accumoli. The time series records the number of mobile phones at an hourly resolution. Red vertical lines are associated with the four main earthquakes that hit the region.

ATECO	N	Revenue (K)	Employees (K)	Empl. per Comp.	Description
Whole/retail sales	9852 (18%)	28354 (31%)	66 (15%)	6	Whole and retail sales companies
Manufacturing	9660 (18%)	36901 (40%)	173 (39%)	17	Manufacturing activities
Constructions	9443 (17%)	5379 (5%)	35 (8%)	3	Constructions
Real estate	6251 (11%)	1054 (1%)	3 (0%)	0	Real estate companies
Tech services	3505 (6%)	2318 (2%)	14 (3%)	4	Professional, scientific, technical activities
Hospitality	3074 (5%)	1101 (1%)	21 (4%)	7	Hotels and restaurants
Business support services	2197 (4%)	1611 (1%)	26 (6%)	12	Business support services, renting
Comm. and info serv.	2085 (3%)	1027 (1%)	10 (2%)	5	Communication and information services
Logistics	1594 (2%)	3102 (3%)	22 (5%)	13	Transport and warehousing
Agriculture	1403 (2%)	1727 (1%)	9 (2%)	6	Agriculture, farming and fishing
Healthcare	1077 (2%)	1070 (1%)	23 (5%)	21	Healthcare activities
Finance	718 (1%)	3222 (3%)	13 (3%)	19	Finance and Insurance activities
Utilities	708 (1%)	1210 (1%)	1 (0%)	2	Utilities
Education	466 (0%)	155 (0%)	2 (0%)	5	Education
Entertainment	445 (0%)	106 (0%)	2 (0%)	4	Arts and entertainment
Waste MGMT	434 (0%)	2045 (2%)	11 (2%)	26	Waste management services
Other serv	217 (0%)	65 (0%)	1 (0%)	5	Other service activities
Mining	153 (0%)	249 (0%)	1 (0%)	8	Mining companies
Public admin	4 (0%)	0 (0%)	0 (0%)	0	Public administration
Extra	3 (0%)	0 (0%)	0 (0%)	0	Foreign organizations

Table 4. Overall composition of companies in the crater area. ATECO 1-letter classification.

deploy this kind of data is a strength for the actual applicability of our approach. While individual data would require strict scrutiny and control from privacy regulators, telecom operators sell aggregate data. Even more, the value recognition of this data for applications like the one described should push the government and public bodies to create partnerships with a data provider to ensure the prompt availability of this data.

Economic activity data

AIDA, created and distributed by Bureau van Dijk S.p.A., is a database, available upon subscription, containing the financial statements and personal and product data of all Italian companies, both active and bankrupt. For each company, the dataset includes the industry/sector, the number of employees and the legal and operating addresses, which allows for localizing them at the municipality level.

Table 4 shows the overall number of companies present in the seismic area.

In particular, we used this data to compute the industrial composition of each municipality. We considered all the companies active during the earthquake (August 2016). We grouped companies accordingly to their industrial sector as specified by standard ATECO 1-Letter codes (e.g., Agriculture, Arts and Entertainment, Buildings, Commerce). For each group, we computed the total revenue. Finally, we normalized the count to sum to 1. The result represents the industrial composition of a given municipality (e.g., Municipality A has: Agriculture=0.8,

Arts and Entertainment = 0, Construction = 0.2). We used the total revenue as a proxy of the relative weight of the company. We run experiments using the total number of employees, obtaining similar results.

Damage and recovery data

Governmental agencies in charge of designing policies for the reconstruction have devised a multi-step approach to collect data about damage and requests for recovery. Reporting these phases is essential to understand the nature of the data.

The Italian Civil Protection, the first responding agency to access the affected area during and after the population's evacuation, initiated the damage assessment (first phase). The experts of Civil Protection had to fill in a specific administrative file in which they reported whether it was safe to access the building and the level of damage. If the building was unsafe (in Italian *inagibile*), the owners could request governmental funds to cover the rebuilding costs. These technical reports provide a first assessment of the damage, and we used them to compute the potential forecast amount of funding requests for the reconstruction. We called them RCR Forecast (*Richiesta Contributo per la Ricostruzione*). These files helped to develop a preliminary assessment of the amount of funding needed for the reconstruction, but this first step must be followed by several others. Only after a second request compiled by another expert in compliance with all the regulations has been submitted by the owner of the damaged building (residential, commercial, or industrial) do the governmental agency and the municipality allocate the necessary funds for reconstruction. Once the expert has submitted the request, the administration evaluates the report and responds. The number of requests for funding presented after these stages is what we define RCR Submitted. It is worth stressing that RCR Forecast and RCR Submitted are not identical. Therefore the two datasets may differ. This difference is relevant because the initial forecast is based only on the assessed damage of a given building, regardless of the owners' decision to rebuild them. If the damage is considered negligible, homeowners may decide not to proceed with the request for funding because the procedure can be expensive and the funds received are insufficient to fix or renovate the damaged house. In addition, the damage assessment process can take between a few months and some years. In the meantime, individuals can decide to relocate and never come back. In this case, there is no willingness to continue the process to obtain funding, and a discrepancy between the initial damage assessment and the final one might emerge.

<https://sisma2016data.it> is the open data initiative (supported by the Italian Government) to monitor damage assessment and recovery plans for the earthquake in central Italy in 2016–2017. This dataset provides information about damage assessment and reconstruction, including the expected requests for rebuilding (RCR Forecast) and the actual ones presented by inhabitants (RCR Submitted) for each municipality. RCR Submitted is available both for private and public buildings in terms of absolute numbers of requests and total amount in euros.

1. **RCR Forecast.** The forecast number of requested funding (RCR) is calculated based on experts' first assessment of the damage.
2. **RCR Submitted** The number of requests presented, which will be able to access the funding (although not necessarily accepted at the current date). These RCRs can be private (private buildings, including commercial activities) and public (including buildings belonging to the state, like schools, and religious buildings, such as churches). For each category, available data are the absolute number and the requested amount in Euros.

All the measures are normalized based on the number of inhabitants of the relative municipality.

Census and demographic data

We used data from the official census data about the resident population and changes in residency in the municipalities under study. Thanks to the census data estimated by ISTAT between 2002 and 2018, we deployed the municipal demographic balance. We extracted the number of residents and the number of residency changes in 2015, 2016 and 2017 from the database (<https://demo.istat.it/ricostruzione/>). The database allows the extraction of municipalities for each province and each year. Therefore we aggregated data at the yearly and provincial levels to obtain the entire dataset on population variation.

Resident population from mobile phone

Several studies identified algorithms to extract resident population from mobile phone data^{55–57}. The main criteria to consider is the average number of mobile phones at night during weekdays as a proxy for the resident population. We applied this procedure also in this work considering mobile phone data from 1 am to 5 am as “night” data. The resulting estimate is naturally lower than census data as we monitor just one telecom operator, and there might be bias in the sample population (e.g., children who do not have phones). We partially compensate for those biases by scaling our results by the telecom operator market share.

Figure 1A, presented above, shows the relationship between the resident population monitored by mobile phone data and actual census counts. Figure 7A shows the number of mobile phones for each municipality before the earthquake. Figure 7B presents the percentage change in mobile phones before and after the earthquake (PCHG).

Forecasting mobile phone data

We created a forecasting model to predict population dynamics measured by mobile phones. The model trained on individual municipalities and macro-regions (crater and out-of-crater regions) with data before the first earthquake hit the area (24 August 2016). Then we used the model to forecast population dynamics.

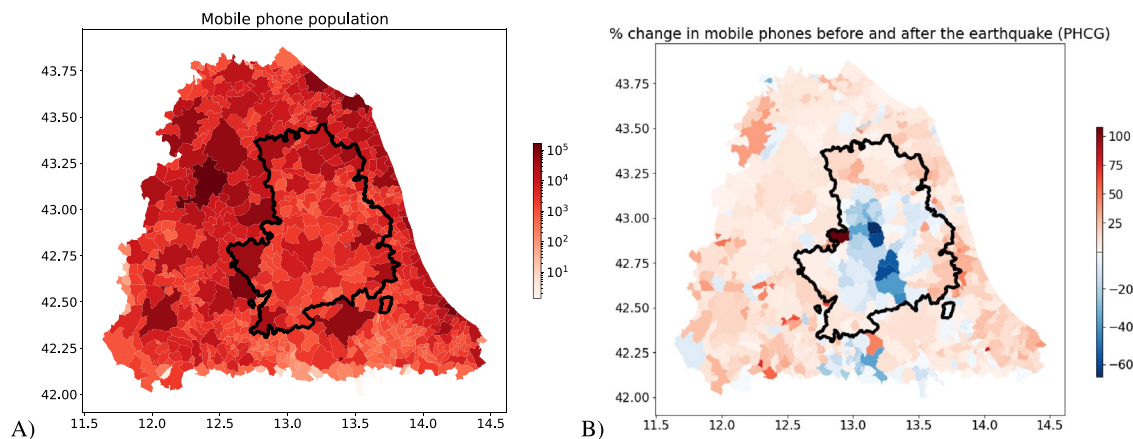


Figure 7. (A) Number of mobile phones for each municipality before the earthquake. (B) Percentage change of mobile phones before and after the earthquake (PHCG). Maps have been drawn using python libraries: *matplotlib 3.5.2, geopandas 0.2.12* – <https://www.anaconda.com>.

The main idea of the adopted forecasting approach is to use non-autoregressive models. In this way, the model never receives information about the earthquake hit. Therefore – although we are fully aware that there might be several confounding factors and this is not a complete causal analysis – we think of the forecast as a counterfactual scenario in which the earthquake did not happen.

The forecasting model uses only features associated with the timestamp: year (Y), week of the year (WoY), days of the week (DoW), and hour (H). Features other than a year are considered categorical (and therefore one-hot encoded). The model uses these features to forecast the number of people (cell phones) in the area. We tested both a linear model and a decision tree regressor.

For example, in the linear model, for each municipality and macro-region, we fit a model like the one in the equation below. The training procedure learn parameters $\alpha, \beta_i, \gamma_j, \delta_k$ using training data before the earthquake. Then, it uses the equation to forecast future time steps. The case of the decision tree regressor is perfectly analogous.

$$\text{forecast}_t = \alpha * Y + \sum_{i=1}^{\text{weeks}} \beta_i * WoY_i + \sum_{j=1}^7 \gamma_j * DoW_j + \sum_{k=1}^{24} \delta_k * H_k$$

Figure 8 shows the forecast result -linear model - applied to all the regions in and out of the crater area. These in the crater area see an average decline in the mobile phone (MPE) of -6.5%. These out of the crater see an MPE of -1.5%. Results from decision tree regressor exhibit a similar pattern.

Regression models

We developed several Ordinary Least Squares (OLS) models in our study. All the models try to explain MPE based on multiple data:

- We considered the industrial composition of each ATECO 1-Letter code (e.g., Agriculture, Arts and Entertainment, Buildings, Commerce) and the percentage of companies of that code. To limit the number of variables, we consider only those sectors contributing to at least 1% of the number of companies and 1% of the revenue of the area. Other sectors are discarded in Section “Results”. Nevertheless, we report the OLS with all the sectors in this Section.
- We computed the distance from the epicentre as the length from the closest epicentre of the earthquake.
- We considered the median altitude of the municipality as related to the municipality isolation.
- We considered information about damage assessment and reconstruction, including the expected requests for rebuilding (RCR Forecast) and the number of the actual request presented by inhabitants (RCR Submitted) for each municipality. RCR Submitted is available both for private and public buildings in terms of absolute numbers of requests and the total amount in euros.

In Section “Results”, we discussed the main findings of our model. In this Section, we show results analogous to Table 1, but using RCR forecast as a measure of damage assessment (see Table 5). The conclusions stated in Section “Results” hold in this case: RCR forecast contributes significantly and negatively to MPE, the more RCR (i.e., destruction) more the population displacement.

Similarly, we report results using all the ATECO codes. The introduction of the “minor” sectors (i.e., public administration, entertainment, mining, waste management, extraterritorial organization, education, and other services) does not affect the results significantly (see Table 6).

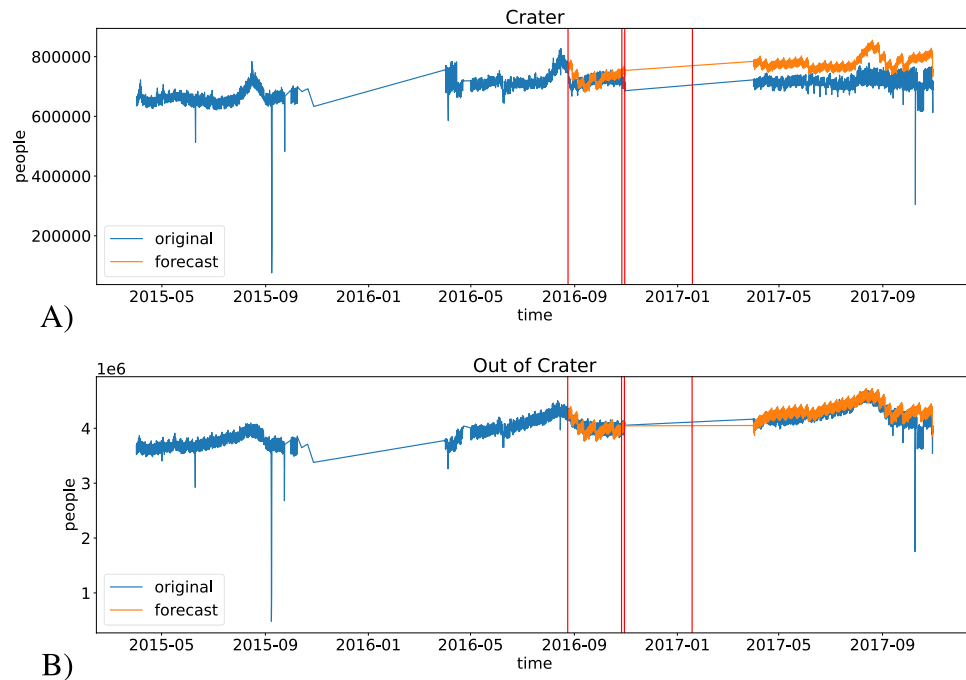


Figure 8. Forecast number of phones. (A) In crater regions. (B) Out of crater regions.

	MPE INIT	MPE END	MPE
D. from epicenter	0.292*** (0.081)	0.436*** (0.159)	0.402*** (0.126)
ALT MED	-0.001 (0.003)	0.001 (0.005)	0.000 (0.004)
RCR forecast	-6.090*** (1.747)	-18.244*** (3.407)	-15.339*** (2.711)
Observations	122	122	122
R^2	0.475	0.512	0.558
Adjusted R^2	0.461	0.500	0.547
Residual Std. error	6.706	13.079	10.410
F statistic	35.534***	41.304***	49.741***

Table 5. Model to explain MPE (MPE INIT, MPE END, MPE) in terms of geographic and forecast damage features. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Spatial auto-correlation

We tested the spatial auto-correlation for all the main variables involved. The presence of spatial auto-correlation is trivial given the localized nature of the earthquake, as it is also apparent in Fig. 3C. Figure 9 presents Moran's analysis of all the main considered variables. The presence of spatial auto-correlation hints at the influence of the earthquake on the monitored variable.

Predict variation

In this section, we propose an approach complementary to the one presented in the above results to analyze MPE after the earthquake. The idea is to create a set of variables summarizing (and reducing the dimensionality of) the features describing different municipalities. Then, we apply a clustering mechanism to the resulting factors.

We consider and transform the variables relevant to the population variation in the earthquake areas.

The considered features are calculated for each municipality as follows:

- The number of companies in the construction sector (*Constructions*);
- The number of companies in the accommodation and restaurant sector (*Hospitality*);
- The number of companies in the wholesale and retail trade sector (*Whole/retail sales*);
- The number of companies in the professional services sector (*Tech services*);
- The number of companies in the renting sector (*Business support services*);

	MPE INIT	MPE END	MPE
Distance from epicenter	0.265*** (0.087)	0.564*** (0.176)	0.492*** (0.140)
ALT MED	0.002 (0.003)	-0.005 (0.006)	-0.003 (0.005)
RCR PRIV	-21.315* (11.915)	-10.550 (24.000)	-13.123 (19.105)
RCR PUB	-165.588*** (53.468)	-413.376*** (107.695)	-354.161*** (85.732)
Agriculture	-6.563 (4.816)	-10.076 (9.701)	-9.237 (7.723)
Real estate	0.411 (8.887)	-20.000 (17.901)	-15.122 (14.250)
Hospitality	-15.972** (6.114)	-9.690 (12.315)	-11.191 (9.803)
Manufacturing	-8.780** (3.474)	-22.293*** (6.997)	-19.064*** (5.570)
Tech services	-9.865 (10.888)	-32.338 (21.931)	-26.967 (17.458)
Whole/retail sales	-8.931** (3.986)	-20.960** (8.029)	-18.085*** (6.391)
Constructions	-13.752*** (3.918)	-13.564* (7.891)	-13.609** (6.281)
Business support services	-10.799 (21.446)	-22.865 (43.196)	-19.981 (34.387)
Communication and info serv.	5.754 (47.455)	-59.279 (95.583)	-43.738 (76.090)
Logistics	-11.688 (9.902)	-38.591* (19.945)	-32.162** (15.877)
Finance	-6.289 (7.012)	-44.935*** (14.123)	-35.700*** (11.243)
Utilities	-11.470 (10.276)	-38.441* (20.698)	-31.996* (16.477)
Public admin	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Entertainment	-82.666 (128.636)	99.584 (259.099)	56.030 (206.259)
Mining	1.006 (79.348)	-18.761 (159.822)	-14.037 (127.229)
Waste MGMT	-14.306 (11.811)	8.356 (23.790)	2.940 (18.938)
Extra	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Healthcare	-35.299 (32.295)	-56.924 (65.049)	-51.756 (51.783)
Education	113.221 (93.092)	-51.497 (187.505)	-12.133 (149.267)
Other serv	1057.701 (1489.736)	627.047 (3000.615)	729.963 (2388.686)
Observations	123	123	123
R ²	0.557	0.572	0.610
Adjusted R ²	0.460	0.478	0.525
F statistic	5.717***	6.073***	7.120***

Table 6. Model to explain MPE (MPE INIT, MPE END, MPE) in terms of geographic, damage and economic features. All the ATECO sectors are considered. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

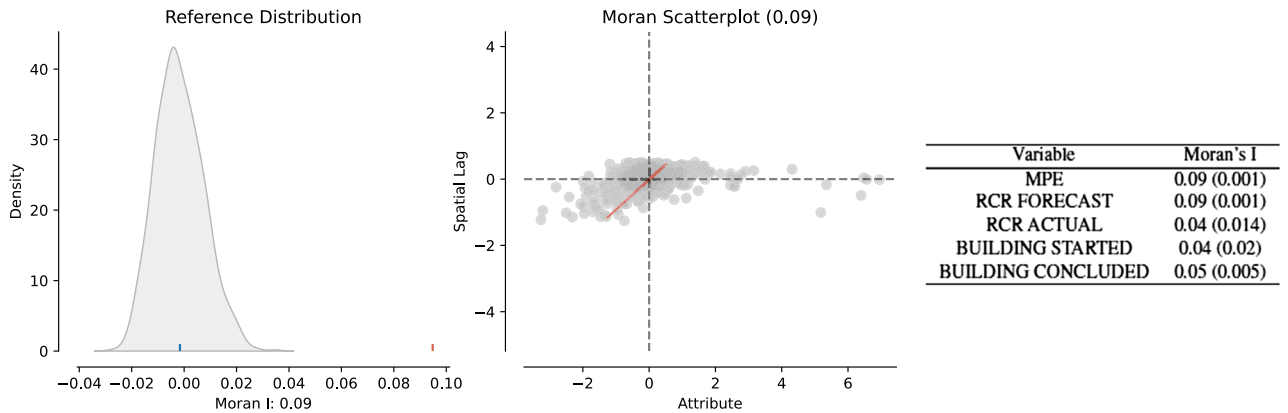


Figure 9. Spatial auto-correlation analysis. Moran's plot for MPE. In the table Moran's I for all the main variables.

- The total number of employees (*Employees*);
- The mean altitude in meters (*Alt med*);
- The province (*Province*);
- The municipality (*Municipality*);

Factor analysis has enabled us to transform and reduce the number of selected features. Initially, we considered a higher number of features. The final selection has been performed based on trial and error and according to the proportion of each variable's variance explained by the factors (commonalities).

We applied Kaiser Varimax rotation⁵⁸ on the above set of variables to identify factors that synthesize the considered variables. Based on the results, we identified three factors representing the variables. Figure 10 presents the results: Factor 1 has high factor loadings for *Constructions*, *Hospitality*, *Whole/retail sales*, *Tech services*, *Business support services* and *Employees*. Therefore, we named this factor as **Industrial_structure**. The second factor formed using the variable *Alt_med* was named **Physical_characteristics**. Finally, the third factor is the result

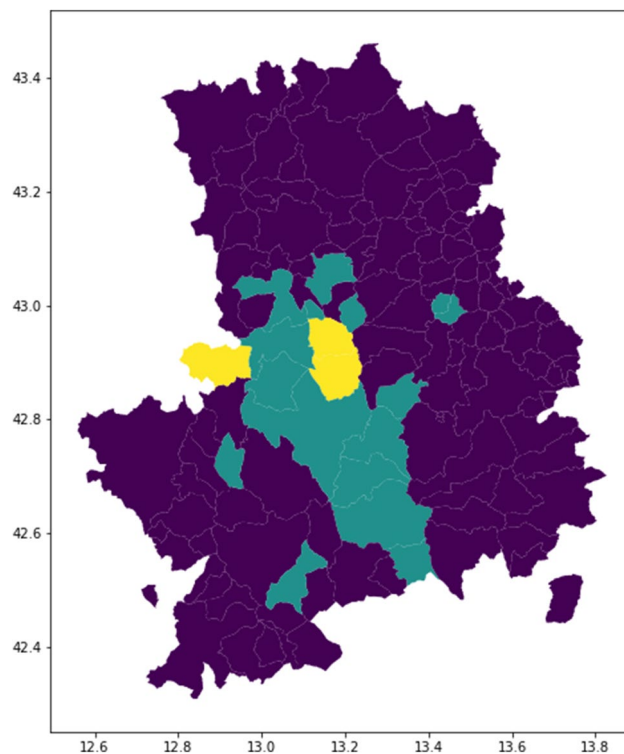


Figure 10. (left) Results of the factor analysis using a Varimax rotation of the axes. (right) Map representing the three clusters of municipalities created based on the new factors. Map has been drawn using python libraries: *matplotlib* 3.5.2, *geopandas* 0.2.12 – <https://www.anaconda.com>.

of *Province* and *Municipality* so that it has been named **Administrative_characteristics**. The commonalities of each variable are above 77%. Therefore it is possible to conclude that the identified factors reasonably and reliably explain the observed variables. We applied the factors to the entire dataset and created three new variables, summarizing the above features.

Based on the three newly created factors, we implemented an agglomerative cluster analysis to identify three clusters of municipalities. Figure 10 shows that one group of municipalities is concentrated in the crater area even though we did not deploy the distance from the epicentre to create the clusters. Furthermore, three have a substantial variation in the number of mobile phones. The similarity of this map with Figs. 3C and 7B confirm the impact of industrial, physical and administrative features to explain and predict mobile phone variation (MPE).

Data availability

Data and code to reproduce our experiments is available at <https://github.com/mmamei/earthquake2016>.

Received: 18 October 2022; Accepted: 21 November 2023

Published online: 14 December 2023

References

1. Quarantelli, E. L. *Emergencies, Disasters and Catastrophes are Different Phenomena* (Disaster Research Center, 2000).
2. Quarantelli, E. L. *What is a Disaster?: New Answers to Old Questions* (Xlibris Corporation, 2005).
3. Rodríguez, H. et al. *Handbook of Disaster Research* Vol. 643 (Springer, 2007).
4. Alexander, D. An interpretation of disaster in terms of changes in culture, society and international relations. In *What is a Disaster: New Answers to Old Questions* (Xlibris Press, 2005).
5. Mileti, D. *Disasters by Design: A Reassessment of Natural Hazards in the United States* (Joseph Henry Press, 1999).
6. Wisner, B., Blaikie, P. M., Blaikie, P., Cannon, T. & Davis, I. *At Risk: Natural Hazards, People's Vulnerability and Disasters* (Psychology Press, 2004).
7. Bankoff, G., Frerks, G., Hilhorst, T. & Hilhorst, D. *Mapping Vulnerability: Disasters, Development, and People* (Routledge, 2004).
8. Morrow, B. H. Identifying and mapping community vulnerability. *Disasters* **23**, 1–18 (1999).
9. Islam, M. R. & Khan, N. A. Threats, vulnerability, resilience and displacement among the climate change and natural disaster-affected people in South-East Asia: An overview. *J. Asia Pac. Econ.* **23**, 297–323 (2018).
10. Thomas, S. L. & Thomas, S. D. Displacement and health. *Br. Med. Bull.* **69**, 115–127 (2004).
11. Christensen, A. & Harild, N. *Forced Displacement: The Development Challenge* (World Bank, 2009).
12. Fiala, N. Economic consequences of forced displacement. *J. Dev. Stud.* **51**, 1275–1293 (2015).
13. DeJesus, K. M. Forced migration and displacement in Africa: Contexts, causes and consequences. *Afr. Geogr. Rev.* **37**(2), 79–82 (2018).
14. Oliver-Smith, A. Debating environmental migration: Society, nature and population displacement in climate change. *J. Int. Dev.* **24**, 1058–1070 (2012).
15. Addressing disaster displacement in disaster risk reduction policy and practice: a checklist. https://www.undrr.org/sites/default/files/2021-02/WiA_Displacement_Checklist_En_0.pdf (2021).
16. Aldrich, D. & Meyer, M. Social capital and community resilience. *Am. Behav. Sci.* **59**, 254–269 (2015).
17. Rand, K., Kurth, M., Fleming, C. H. & Linkov, I. A resilience matrix approach for measuring and mitigating disaster-induced population displacement. *Int. J. Disaster Risk Reduct.* **42**, 101310 (2020).
18. Imperiale, A. J. & Vanclay, F. Top-down reconstruction and the failure to build back better resilient communities after disaster: Lessons from the 2009 LAquila Italy earthquake. *Disaster Prev. Manag. Int. J.* **29**, 541–555 (2020).
19. Fantechi, F., Urso, G. & Modica, M. Can extreme events be an opportunity? Depopulation and resilience of rural communities in central Italy after the 1997 earthquake. *J. Rural Stud.* **79**, 311–321 (2020).
20. Norris, F. H., Stevens, S. P., Pfefferbaum, B., Wyche, K. F. & Pfefferbaum, R. L. Community resilience as a metaphor, theory, set of capacities, and strategy for disaster readiness. *Am. J. Community Psychol.* **41**, 127–150 (2008).
21. Gomez, C. & Purdie, H. UAV-based photogrammetry and geocomputing for hazards and disaster risk monitoring—a review. *Geoenviron. Disasters* **3**, 23 (2016).
22. Enenkel, M. et al. Emergencies do not stop at night: Advanced analysis of displacement based on satellite-derived nighttime light observations. *IBM J. Res. Dev.* **64**, 8:1–8:12 (2020).
23. Li, L., Bensi, M., Cui, Q., Baecher, G. B. & Huang, Y. Social media crowdsourcing for rapid damage assessment following a sudden-onset natural hazard event. *Int. J. Inf. Manag.* **60**, 102378 (2021).
24. Bonati, S., Nardini, O., Boersma, K. & Clark, N. Unravelling dynamics of vulnerability and social media use on displaced minors in the aftermath of Italian earthquakes. *Int. J. Disaster Risk Reduct.* **89**, 103632 (2023).
25. Petrovic, B., Scaini, C. & Parolai, S. The damage assessment for rapid response (DARR) method and its application to different ground-motion levels and building types. *Seismol. Res. Lett.* **94**, 1536–1555 (2023).
26. Wu, L., Chikaraishi, M., Nguyen, H. & Fujiwara, A. Analysis of post-disaster population movement by using mobile spatial statistics. *Int. J. Disaster Risk Reduct.* **54**, 102047 (2021).
27. Yabe, T., Tsubouchi, K., Fujiwara, N., Sekimoto, Y. & Ukkusuri, S. Understanding post-disaster population recovery patterns. *J. R. Soc. Interface* **17**, 20190532 (2020).
28. Lu, X., Bengtsson, L. & Holme, P. Predictability of population displacement after the 2010 Haiti earthquake. *Proc. Natl. Acad. Sci.* **109**, 11576–11581 (2012).
29. Bengtsson, L., Lu, X., Thorson, A., Garfield, R. & von Schreeb, J. Improved response to disasters and outbreaks by tracking population movements with mobile phone network data: A post-earthquake geospatial study in Haiti. *PLoS Med.* **8**, 1–9 (2011).
30. Blumenstock, J. Inferring patterns of internal migration from mobile phone call records: Evidence from Rwanda. *Inf. Technol. Dev.* **18**, 107–125 (2012).
31. Yabe, T., Ukkusuri, S. & Rao, P. Mobile phone data reveals the importance of pre-disaster inter-city social ties for recovery after hurricane maria. *Appl. Netw. Sci.* **5**, 4 (2020).
32. Calabrese, F., Ferrari, L. & Blondel, V. D. Urban sensing using mobile phone network data: A survey of research. *ACM Comput. Surv.* **47**, 1–20 (2014).
33. Blondel, V., Decuyper, A. & Krings, G. A survey of results on mobile phone datasets analysis. *EPJ Data Sci.* **4**, 10 (2015).
34. Netmob conference series. <https://netmob.org> (2011).
35. Szocska, M. et al. Countrywide population movement monitoring using mobile devices generated (big) data during the COVID-19 crisis. *Nat. Sci. Rep.* **11**, 5943 (2021).
36. Yabe, T., Jones, N. K., Rao, P. S. C., González, M. C. & Ukkusuri, S. V. Mobile phone location data for disasters: A review from natural hazards and epidemics. *Comput. Environ. Urban Syst.* **94**, 101777 (2021).

37. Wilson, R. *et al.* Rapid and near real-time assessments of population displacement using mobile phone data following disasters: The 2015 nepal earthquake. *PLoS Curr. Disasters* **8** (2016).
38. Cumbane, S. P. & Gidófalvi, G. Spatial distribution of displaced population estimated using mobile phone data to support disaster response activities. *ISPRS Int. J. Geo Inf.* **10**, 421 (2021).
39. Yabe, T., Sekimoto, Y., Tsubouchi, K. & Ikemoto, S. Cross-comparative analysis of evacuation behavior after earthquakes using mobile phone data. *PLoS ONE* **14**, e0211375 (2018).
40. Li, T., Dejby, J., Albert, M., Bengtsson, L. & Lefebvre, V. Estimating the resilience to natural disasters by using call detail records to analyse the mobility of internally displaced persons. [arXiv:1908.02381](https://arxiv.org/abs/1908.02381) (2019).
41. Andrade, X., Layedra, F., Vaca, C. & Cruz, E. Risc: Quantifying change after natural disasters to estimate infrastructure damage with mobile phone data. In *2018 IEEE International Conference on Big Data (Big Data)* 3383–3391 (2018).
42. Castillo-López, G., Guaranda, M.-B., Layedra, F. & Vaca, C. A place to go: Locating damaged regions after natural disasters through mobile phone data. In *Symposium on Information Management and Big Data* (2019).
43. Marzuoli, A. & Liu, F. Monitoring of natural disasters through anomaly detection on mobile phone data. *2019 IEEE International Conference on Big Data (Big Data)* 4089–4098 (2019).
44. Kontokosta, C. & Malik, A. The resilience to emergencies and disasters index: Applying big data to benchmark and validate neighborhood resilience capacity. *Sustain. Cities Soc.* **36**, 272–285 (2018).
45. Ford, J. *et al.* Opinion: Big data has big potential for applications to climate change adaptation. *Proc. Natl. Acad. Sci.* **113**, 10729–10732 (2016).
46. Stewart, J. P. *et al.* Reconnaissance of 2016 central Italy earthquake sequence. *Earthq. Spectra* **34**, 1547–1555 (2018).
47. Sextos, A. *et al.* Local site effects and incremental damage of buildings during the 2016 central Italy earthquake sequence. *Earthq. Spectra* **34**, 1639–1669 (2018).
48. Frigerio, I., Zanini, F., Mattavelli, M. & De Amicis, M. Understanding the interacting factors that influence social vulnerability: A case study of the 2016 central Italy earthquake. *Disasters* **43**, 867–890 (2019).
49. La ricostruzione post sisma 2016 in abruzzo, lazio, marche, umbria. https://sisma2016.gov.it/wp-content/uploads/2022/03/Rapporto2021_03-marzo_08-1.pdf (2021).
50. Pearl, J. Causal inference in statistics: An overview. *Stat. Surv.* **3**, 96–146 (2009).
51. Beck, e. a., M. World risk report 2012. Alliance development works in collaboration with unu/ehs, the nature conservancy (2012).
52. Johnson, C. Strategic planning for post-disaster temporary housing. *Disasters* **31**, 435–458 (2007).
53. Mannakkara, S. & Wilkinson, S. Re-conceptualising, building back better to improve post-disaster recovery. *Int. J. Manag. Proj. Bus.* **7**, 327–341 (2014).
54. Italia: ricavi e quote di mercato dei principali operatori mobili. <https://www.infodata.ilsole24ore.com/2017/05/23/italia-ricavi-quote-mercato-dei-principali-operatori-mobili/> (2017).
55. Douglass, R., Meyer, D., Ram, M., Rideout, D. & Song, D. High resolution population estimates from telecommunications data. *EPJ Data Sci.* **4**, 1–13 (2015).
56. Mamei, M., Colonna, M. & Galassi, M. Automatic identification of relevant places from cellular network data. *Pervasive Mob. Comput.* **31**, 147–158 (2015).
57. Vanhoof, M., Lee, C. & Smoreda, Z. Performance and sensitivities of home detection on mobile phone data. In *Big Data Meets Survey Science: A Collection of Innovative Methods* 245–271 (2015).
58. Kaiser, H. Computer program for varimax rotation in factor analysis. *Educ. Psychol. Measur.* **19**, 413–420 (1959).

Acknowledgements

We would like to thank Telecom Italia (TIM) for mobile phone data access. Work supported by: the University of Modena and Reggio Emilia FAR Project: Data driven methodologies to study social capital and its role for economic growth.

Author contributions

F.G., N.H., M.M., F.P. conceived the experiments, N.H., M.M., G.B., N.C. conducted the experiments and wrote code, F.G., N.H., M.M., F.P. analysed the results. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to F.G. or M.M.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023