Deprem sonrası iç göç tahminleri için cep telefonu verisinin kullanılması
Prediction of internal migration after an earthquake with call detail records

1st Jingkang Hu
Information and Computing Sciences
Utrecht University, Utrecht, The Netherlands
j.jingkanghu@gmail.com

2nd Bilgeçağ Aydoğdu
Information and Computing Sciences
Utrecht University, Utrecht, The Netherlands
b.aydogdu@uu.nl

3rd Subhi Güneş
Turkcell Tech
Istanbul, Türkiye
subhi.gunes@turkcell.com.tr

4th Semiha Nur Yağcıklı
Turkcell Tech
Istanbul, Türkiye
semiha.yagcikli@turkcell.com.tr

5th Albert Ali Salah
Information and Computing Sciences, Utrecht University
Utrecht, The Netherlands
a.a.salah@uu.nl

***Özetçe* —2023 Kahramanmaraş Depremi 55.000'den fazla ölüme yol açtı. Depremlerde can kayıplarını önlemek ve kaynak tahsisini düzenlemek için etkili tahliye yönetimi gereklidir. Bu çalışmada, kriz anlarında, özellikle deprem sonrasında cep telefonu Detaylı Çağrı Kayıtlarının (CDR) kullanımını inceliyoruz. Çalışmamızda deprem sonrası tahliye davranışlarına ve nüfus hareketlerinin tahmin edilmesine odaklanıyoruz. Tahmin uygulamasında uzaklığa göre ölçeklenmiş özelliklerle yapay öğrenme modelleri kullanıyoruz. Deneylerimiz kurduğumuz modelin farklı ilçeler arasında tahliye akışlarını tahmin edebildiğini göstermektedir. Ana bulgularımız, nüfus dağılımı ve deprem şiddetinin tahliye örüntülerinin en önemli faktörleri olduğudur. Deprem bölgesindeki Türk nüfusu ile Suriyeli göçmen nüfus arasındaki karşılaştırmalı analiz, benzer öznitelik önemi sıralaması gösterse de, bazı farklı örüntü dağılımları sergilemektedir.**

Anahtar Kelimeler—Mobil veri, deprem sonrası hareketlilik, yapay öğrenme

***Abstract*—The 2023 Kahramanmaraş Earthquake resulted in over 55,000 deaths. Efficient evacuation management can significantly reduce secondary casualties and optimize resource allocation. This study explores the application of Call Detail Records (CDR) in times of crisis, with a particular focus on the consequences of earthquakes. We focus on two issues: the post-earthquake evacuation behavior and prediction of population movements after earthquakes. We use machine learning models with distance-normalized features inspired by the Gravity model to predict population movements immediately after earthquake. The experiments show that our model can predict evacuation flows between different districts. Our main findings are that population distribution and earthquake intensity are the primary factors of evacuation patterns. Comparative analysis between Turkish and Syrian populations shows the same feature importance rankings, but distinct pattern distributions.**

Keywords—Mobile data, post-earthquake mobility, machine learning

I. Introduction

Natural disasters such as earthquakes can have a huge impact on people's lives and lead to large-scale population movements[1]. In recent years, mobile call detail records (CDR) have been proposed as a unique data source for analyzing post-disaster population movement [1]. CDR data are collected by telecommunications operators, and contain information on the time and duration of all cell phone calls, as well as the geographic location of senders and receivers [2]. During an earthquake, mobile networks usually maintain some functionality even if part of the infrastructure is damaged, and CDR provides a means to observe population movements in great detail. If the infrastructure is there, this can be done in real-time, which can help rescue and resource planning efforts.

In this work, we use CDRs to construct a migration model to predict the evacuation flows in the aftermath of the Kahramanmaraş Earthquake. Our research questions are: 1) How can we use mobile phone data to supplement other data sources to predict the movement of people inside the country after an earthquake? 2) Which features will influence evacuation behaviors? 3) Did Turkish and Syrian people respond to similar or different factors in their evacuation decisions?

Using carefully anonymized and aggregated data from a major telecommunications company, we analyzed post-earthquake evacuation patterns and found that approximately half of all displacement flows occurred within the same city. Moreover, the vast majority of evacuation destinations remained within earthquake-affected areas. Our city-level analysis of both Turkish and Syrian populations revealed distinct evacuation behaviors: Turkish evacuees showed more dispersed destination choices, while Syrian evacuees demonstrated more

[1]This is the uncorrected author proof. Copyright with IEEE. Please cite as "Hu, J., B. Aydoğdu, S. Güneş, S.N. Yağcıklı, A.A. Salah, Prediction of internal migration after an earthquake with call detail records, 33rd IEEE Conf. on Signal Processing and Communications Applications, Istanbul, 2025."

concentrated destination preferences. Our experiments validated the use of gravity-transformed features in predicting post-earthquake population movements.

II. Related Work

Call Detail Records (CDR) are data collected by mobile network operators when providing their services. These records include call initiation time, duration, phone numbers of both parties, call type, and possible device location information [3]. Using CDR, it is possible to infer population dynamics, mobility, social networks, and socio-demographics [4]. The spatial trajectory of each user can be obtained based on the original CDR. Furthermore, the night-time activity range can be used to infer the user's home location.

Studies using mobile phone location data to study disasters are categorized into three main categories [5]: population displacement and evacuation modeling (our focus), long-term recovery analysis, and inverse inferences about damage to the built environment, respectively. CDR has a lot of promise for counteracting and responding to the adverse effects of disasters [6]. A post-earthquake study in Nepal used CDR data to rapidly assess mobility patterns after the earthquake, particularly the mass exodus from the Kathmandu Valley [7]. [8] found that the destinations of people displaced after the earthquake were highly correlated with their prior mobility patterns. In 2016, Kargel et al. analyzed behavioral patterns of 12 million mobile phone users following an earthquake, and this was the first time that large-scale mobile location data significantly aided disaster relief [9].

Yabe et al. used CDR data of over 1.9 million users before and after five natural disasters, including the 2017 Puebla Earthquake, to investigate population recovery trends [10] . After all five disasters, the majority of users returned quickly within a few weeks of the disaster, with the remainder returning gradually over a longer period. Therefore, data from the weeks following a disaster are most important in a study to understand the pattern of population recovery after a disaster. Our present study investigates population recovery patterns for one and a half months after the 2023 earthquake.

III. Methodology

A. Data used for the study

We use a fine-grained CDR dataset for a total of three months between 01/01/23-31/03/23, coming from a major telecom operator in Turkey. The original dataset, which was not shared, consists $325,000$ customers recorded with timestamp of calls (aggregated at the hour level), base station site information (for caller and callee), randomized numbers representing customer IDs and customer segment information, expressed with 1 for Turkish and 2 for Syrian. Data are anonymized and aggregated within the operator before being shared for research[2], along established guidelines to remove all personal information [11], [2].

We use the site ID with the highest call frequency for each subject from 19:00 to 7:00 as the home location of that day. The home location with the highest frequency each week will be regarded the home location for that week, and the same applies to each month. The home location of most consumers in the data set has not changed in three months, so they can be regarded as not migrating. Within our dataset, $140,750$ people stayed in the same city, and $15,817$ people displaced to another city during the three months.

We supplement the CDR with census data TurkStat[3], which includes the total population of each city and district in Turkey. Syrian populations per city is obtained from the General Directorate of Migration Management in 2023[4]. The earthquake data were extracted from the USGS Earthquake Hazard Program [5]. The dataset provides the shaking intensity in increments of 0.2. In Figure 1, the distribution of the seismic intensities is shown. Areas with intensities above 4.5 were classified as severely impacted zones, while those below 4.5 were designated as less severely impacted zones.

Table I: Populations and sample sizes in the CDR dataset

City	Population	CDR Sample	Percentage %
GAZIANTEP	2,164,131	25,427	1.175
SANLIURFA	2,213,964	19,144	0.865
HATAY	1,544,640	11,787	0.763
ADANA	2,270,298	9,570	0.422
KAHRAMANMARAS	1,116,618	4,256	0.381
KAYSERI	1,445,683	151	0.0105
KILIS	155,179	4,979	3.209
OSMANIYE	557,666	1,324	0.238
MALATYA	742,725	4,044	0.545
ADIYAMAN	604,978	966	0.160

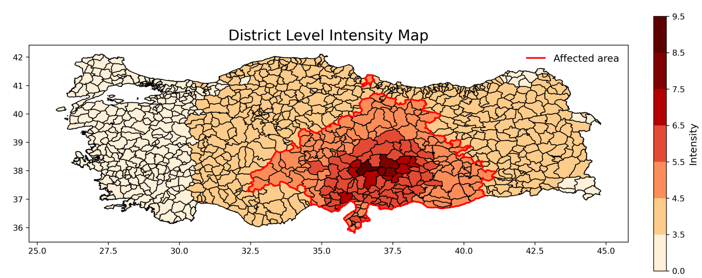

Figure 1: District-level intensity map of the earthquake

We use data on buildings damaged by the earthquake to assess the varying impacts across different regions[6]. From Figure. 2, buildings damaged by the earthquake are predominantly located in major cities near the earthquake's epicenter. In contrast, cities situated farther from the epicenter primarily report slight damage to buildings, with fewer instances of buildings experiencing more severe levels of damage.

Finally, we use the Relative Wealth Index (RWI), which was developed to accurately assess the economic status of low- and middle-income countries[7]. RWI combines publicly

[2]See https://hummingbird-h2020.eu/images/projectoutput/d6-1.pdf

[3]https://data.tuik.gov.tr/Kategori/GetKategori?p=nufus-ve-demografi-109

[4]https://multeciler.org.tr/eng/number-of-syrians-in-turkey/

[5]https://earthquake.usgs.gov/earthquakes/eventpage/us6000jllz/shakemap/intensity

[6]https://sheltercluster.org/turkiye-earthquake-2023/pages/damaged-buildings

[7]https://dataforgood.facebook.com/dfg/tools/relative-wealth-index

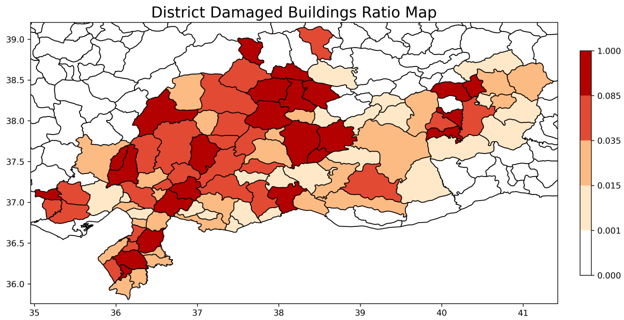

Figure 2: District-level ratio of damaged buildings

available survey data with non-traditional predictive data via machine learning to predict absolute and relative wealth for 2.4 square kilometer grid cells.

B. *Features and models*

After [12] demonstrated that the amount of migration is inversely proportional to distance, [13] formally proposed to apply the law of gravity to population movements between two locations. Isaacman et al. simulated migration during the severe 2014 drought in Guajira, Colombia, using anonymized and aggregated CDR with the Gravity model [14]. The results showed that the prediction had a success rate of about 60 percent for the total number of people who migrated and the locations where they migrated.

The model assumes that the flow between two areas decreases with distance. The basic equation is as follows:

$$T_{ij} \propto \frac{m_i \cdot m_j}{rij} \tag{1}$$

where T_{ij} is the amount of flow from area i to area j, m_i and m_j are the populations of the two areas respectively, and r_{ij} is the distance between the two areas. The attractiveness of area j is proportional to m_j, but there is a cost of distance traveled:

$$T_{ij} = K \cdot m_i \cdot m_j \cdot f(r_{ij}) \tag{2}$$

where K is a constant, and $f(r_{ij})$ is known as the deterrence function, a decreasing function of distance. The distance function $f(r_{ij})$ is usually modeled as a power law or exponential [15]:

$$f(r_{ij}) = \alpha \cdot r_{ij}^{-\beta} \cdot e^{-r_{ij}/r_c} \tag{3}$$

where α and β are adjustable exponents. Although it looks simple, the Gravity model fits internal migration data well [16].

Our approach is inspired by [17], which uses distance-based feature normalization inspired by the Gravity model to predict post-hurricane evacuation flows. All features described in Section III-A, namely population intensity, damaged buildings ratio, and RWI, are first processed through a formula conversion. Compared to directly using the characteristics of any given two cities to predict the evacuation flow between them, the formula in the form of a Gravity model takes into account the influence of the distance between the two cities.

Given a district i, we represent the set of features as D_i, and for two districts i and j, the joint feature using a gravity-inspired transformation is written as follows:

$$g_{ij,k} = \frac{d_{i,k} * d_{j,k}}{dist_{i,j}} \tag{4}$$

where $d_{i,k}$ is the k-th feature of district i, $dist_{i,j}$ is the distance between districts, and $g_{ij,k}$ is the k-th joint feature. We do the same normalization for features at the city level.

We tested Support Vector Machines, Random Forests, and XGBoost as supervised machine learning models, and used XGBoost in the final approach [18] Given that our district-level dependent variable contains missing values and exhibits a right-skewed distribution, we adopted Tweedie regression as a loss function, which is a special case of the exponential dispersion family [19]. [20] successfully applied Tweedie regression loss function in XGBoost for hurricane evacuation prediction, achieving promising results. To predict evacuation flows after the earthquake at district-level, we used the (normalized) population size, RWI, and damaged building ratio features as input variables, and evacuation flows as the output variables. In order to avoid poor prediction results caused by sample imbalance, we used stratified k-fold cross-validation.

IV. EXPERIMENTAL RESULTS

A. *Description of flows*

In terms of evacuation flows, most origin districts are located in Hatay, Gaziantep, Kilis, Adana, Kahramanmaras, and Malatya. The flow counts from Hatay and Gaziantep are the highest, reaching 21.6% and 21.1% respectively. Our data show that flows within the same city account for 40.74% of all flows. For Hatay city, flows within the city reach 48.3%. Flow from the Antakya district of Hatay was the largest among the origin districts, with a total of 2152, accounting for 5.1% of the total. Most people on the move did not leave the affected area completely, but have gone to relatively less affected areas and adjacent cities. Among the destination districts that are not part of the affected area, Mezitli, Erfemli, Toroslar districts of Mersin, Selcuklu district of Konya, and Cankaya district of Ankara are the top five destinations.

Turkish and Syrian subpopulations showed similar origin cities, with Hatay, Gaziantep, and Sanliurfa being the top three. However, notable differences emerged in destination city patterns. Turkish citizens showed a preference for Ankara, while Syrians favored Istanbul. Antalya ranked as the third destination for Turkish citizens, but did not appear among the top five destinations for Syrians. Both populations generally showed significant intra-city movement across different districts within their origin cities, with Gaziantep's Syrian population being a notable exception, where only one-fifth remained within the city.

B. *Predictive models*

We tested our supervised learning approach to see how much we could have predicted the mobility, given a certain damage profile. Table II shows our results, with simple Gravity model and XGBoost without feature normalization as baselines, in terms of R^2, Root mean square error (RMSE), and mean average error (MAE). Furthermore, a SHAP value analysis shows the importance of factors[8] for the predictive modeling in Figure 3.

Table II: Prediction results for models across evaluation metrics

Model	R2	RMSE	MAE
XGBoost with normalization	0.601	6.921	1.257
Simple Gravity model	0.148	15.950	4.487
XGBoost without normalization	0.063	16.487	3.662

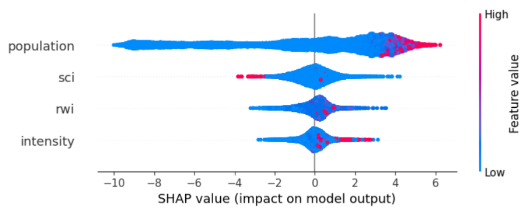

Figure 3: Feature importances via SHAP method.

V. Conclusions

This initial work provides some descriptive mobility results, which need further investigation via qualitative studies, and a first attempt at predictive modeling of post-earthquake evacuation flows. We show that CDR has potential to contribute to both descriptive and predictive analyses.

The biggest limitation of our study is the access to CDR. CDR is stored by telecommunications companies and is not publicly available. Special arrangements and legal agreements are required for governmental or non-governmental organizations to access anonymized and aggregated versions of these valuable datasets. Turk Telekom established the Data for Refugees (D4R) initiative between 2016-2019, and made CDR data public to support scientific studies that aimed at enhancing the living conditions of Syrian refugees in Turkey [2]. We have used a similarly prepared but private dataset in this study, coming from a different telecom operator.

Privacy protection in CDR is an important concern, with the need to ensure anonymity and confidentiality of personal information during data analysis. There are established guidelines to achieve this, such as the application of "privacy by design and default" [21] to ensure that no individual can be tracked or identified in the data. Our project has been screened by university-level and international ethics committees to adhere to legal and ethical guidelines.

CDR also has possible biases. Telecom operators have different penetration rates in each area, and in very low-income areas, only people who own and use a SIM card are included in the dataset [22]. These should be considered in evaluating the results of analysis.

[8]SCI denotes Facebook social connectivity index, which had negligible impact and is not further discussed here.

Acknowledgment

This study is supported by European Union's Horizon 2020 Research and Innovation Programme under grant agreement No 870661.

References

[1] J. E. Blumenstock, "Using mobile phone data to measure the ties between nations," in *Proc. IConference*, 2011, pp. 195–202.

[2] A. A. Salah, A. Pentland, B. Lepri, E. Letouzé, Y. De Montjoye, X. Dong, and P. Vinck, *Guide to mobile data analytics in refugee scenarios*. Springer, 2019.

[3] E. Letouzé, P. Vinck, and L. Kammourieh, "The law, politics and ethics of cell phone data analytics," *Data-Pop Alliance*, 2015.

[4] S. Dujardin, D. Jacques, J. Steele, and C. Linard, "Mobile phone data for urban climate change adaptation: Reviewing applications, opportunities and key challenges," *Sustainability*, vol. 12, no. 4, p. 1501, 2020.

[5] T. Yabe, N. K. Jones, P. S. C. Rao, M. C. Gonzalez, and S. V. Ukkusuri, "Mobile phone location data for disasters: A review from natural hazards and epidemics," *Computers, Environment and Urban Systems*, vol. 94, p. 101777, 2022.

[6] M. N. I. Sarker, Y. Peng, C. Yiran, and R. C. Shouse, "Disaster resilience through big data: Way to environmental sustainability," *International Journal of Disaster Risk Reduction*, vol. 51, p. 101769, 2020.

[7] R. Wilson *et al.*, "Rapid and near real-time assessments of population displacement using mobile phone data following disasters: The 2015 Nepal earthquake," *PLoS currents*, vol. 8, 2016.

[8] X. Lu, L. Bengtsson, and P. Holme, "Predictability of population displacement after the 2010 Haiti earthquake," *PNAS*, vol. 109, no. 29, pp. 11 576–11 581, 2012.

[9] J. S. Kargel *et al.*, "Geomorphic and geologic controls of geohazards induced by Nepal's 2015 Gorkha earthquake," *Science*, vol. 351, no. 6269, p. aac8353, 2016.

[10] T. Yabe, K. Tsubouchi, N. Fujiwara, Y. Sekimoto, and S. V. Ukkusuri, "Understanding post-disaster population recovery patterns," *Journal of the Royal Society Interface*, vol. 17, no. 163, p. 20190532, 2020.

[11] V. Blondel *et al.*, "Mobile phone data for development-analysis of mobile phone datasets for the development of Ivory Coast," *Orange D4D Challenge*, 2013.

[12] E. G. Ravenstein, *The laws of migration*. Royal Statistical Society, 1885.

[13] J. Q. Stewart, "The development of social physics," *American Journal of Physics*, vol. 18, no. 5, pp. 239–253, 1950.

[14] S. Isaacman *et al.*, "Modeling human migration patterns during drought conditions in La Guajira, Colombia," in *ACM SIGCAS Conf. on Computing and Sustainable Societies*, 2018.

[15] H. Barbosa *et al.*, "Human mobility: Models and applications," *Physics Reports*, vol. 734, pp. 1–74, 2018.

[16] J. Poot, O. Alimi, M. P. Cameron, and D. C. Maré, "The gravity model of migration: the successful comeback of an ageing superstar in regional science," *Investigaciones Regionales*, no. 36, pp. 63–86, 2016.

[17] L. Hong and V. Frias-Martinez, "Modeling and predicting evacuation flows during hurricane Irma," *EPJ Data Science*, vol. 9, no. 1, p. 29, 2020.

[18] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proc. ACM KDD*, 2016, pp. 785–794.

[19] M. C. Tweedie *et al.*, "An index which distinguishes between some important exponential families," in *Statistics: Applications and new directions: Proc. Indian statistical institute golden Jubilee International conference*, vol. 579, 1984, pp. 579–604.

[20] P. K. Anyidoho, X. Ju, R. A. Davidson, and L. K. Nozick, "A machine learning approach for predicting hurricane evacuee destination location using smartphone location data," *Computational Urban Science*, vol. 3, no. 1, p. 30, 2023.

[21] A. A. Salah *et al.*, "Introduction to the data for refugees challenge on mobility of Syrian refugees in Turkey," in *Guide to Mobile Data Analytics in Refugee Scenarios*. Springer, 2019, pp. 3–27.

[22] T. Li, J. Dejby, M. Albert, L. Bengtsson, and V. Lefebvre, "Detecting individual internal displacements following a sudden-onset disaster using time series analysis of call detail records," *arXiv preprint arXiv:1908.02377*, 2019.