Extended abstract: Producing population-level estimates of internal displacement in Ukraine using GPS mobile phone data

Rodgers Iradukunda[1], Francisco Rowe[1*], Elisabetta Pietrostefani[1]

[1]Geographic Data Science Lab, Department of Geography and Planning, University of Liverpool, Liverpool.

*Corresponding author(s). E-mail(s): f.rowe-gonzalez@liverpool.ac.uk;

Abstract

Nearly 110 million people are forcibly displaced people worldwide. However, estimating the scale and patterns of internally displaced persons in real time, and developing appropriate policy responses, remain hindered by traditional data streams. They are infrequently updated, costly and slow. Mobile phone location data can overcome these limitations, but only represent a population segment. Drawing on an anonymised large-scale, high-frequency dataset of locations from 25 million mobile devices, we propose an approach to leverage mobile phone data and produce population-level estimates of internal displacement. We use this approach to quantify the extent, pace and geographic patterns of internal displacement in Ukraine during the early stages of the Russian invasion in 2022. Our results produce reliable population-level estimates, enabling real-time monitoring of internal displacement at detailed spatio-temporal resolutions. Accurate estimations are crucial to support timely and effective humanitarian and disaster management responses, prioritising resources where they are most needed.

Keywords: internal population displacement, GSP mobile phone data, Ukraine

Introduction

The forced displacement of individuals, including refugees, asylum-seekers and internally displaced people (IDP), creates considerable humanitarian, social and economic costs. Recent estimates indicates that the number of forcibly displaced populations has significantly grown as result of persecution, conflict, violence, human rights violations and disasters. As of June 2023, the United Nations High Commissioner for Refugees (UNHCR) estimated 110 million of forcibly displaced people worldwide, with the number of IDP (62.5 million) accounting for the largest share of these displacements [1]. The Russian full-scale invasion of Ukraine is estimated to have created the fastest global displacement crisis, and one of the largest, since the Second World War [2].

Forcibly displaced population data are key to inform operational plans, humanitarian responses and long-term policy making. By understanding the scale and locations where people are forcibly fleeing and the extent of their return, government agencies, aid organisations and local community groups can better prioritise and allocate resources and services where they are most needed in the required quantities. Highly granular geographical data tracking population displacements in real time are therefore critical to support these efforts [3].

Traditional data systems are constrained to render information at such high temporal and geographical resolution and speed. Over the years, UNHCR and the Internal Displacement Monitoring Centre (IDMC) have made significant efforts triangulating various data sources to improve and deliver global databases that enable the monitoring and management of forced population displacements. However, they have also identified persistent challenges in the production of reliable estimates of forcibly displaced populations [4]. Traditional data systems are not regularly updated, costly and characterised by slow data collection and release [5]. Particularly in conflict areas, humanitarian partners and data collectors often face access restrictions due to violence and insecurity preventing data gathering [1]. Data streams may also have gaps collecting data on displacement during short-term evacuations or spontaneous movements resulting from conflict and violence [6]. The danger and challenging nature of field work in conflict zones can also disrupt continuous engagement in data collection by humanitarian and development agencies [7].

Novel digital footprint data have emerged as a key source of information offering an opportunity to capture human population movements at highly granular geographical and temporal scales [5, 8]. These data are automatically and continuously generated avoiding exposure of data collectors to hazardous areas and minimising potential data gaps [6]. Mobile phone location data have increasingly been used to monitor population movements during crises, particularly measuring exposure to ambient pollutant exposure [9, 10], transport patterns [11], recreational behaviour [12], disaster-induced displacement (e.g. flooding and earthquakes) [13] and the spread of diseases - notably during the COVID-19 pandemic [14]. Yet, limited work has been undertaken to estimate the scale and patterns of IDP using mobile phone data. Additionally, differences in the access and use of mobile phone technology and applications used to collect location data prevent the production of reliable population-level mobility estimates. Most existing work based on mobile phone data has thus constrained to offer rough signals about population movements (e.g. spatial concentration), trends (e.g. increasing) and changes (e.g. low to high) [5].

To address these gaps, we propose an approach to produce high-frequency population-level estimates of internal displacement drawing on location data from 25 million unique devices. Our first contribution is methodological and illustrates how high-frequency footprint data can enable the generation of population-level estimates of internal displacement correcting for differences in mobile phone-derived and actual population counts, moving beyond providing rough signals. Most prior work leveraging on digital footprint data to estimate population displacement relies on social media or call detail records, with location being inferred resulting in reduced precision [15]. We use data collected via GPS technology which provides greater precision data on location [16].

Our second contribution is to provide evidence of the scale and spatial patterns of population displacement in Ukraine during the first year of the invasion. The Russian full-scale invasion of Ukraine has created the fastest global displacement crisis, and one of the largest, since the Second World War [2]. Recent estimates suggest that nearly one-third of Ukrainian residents are estimated to have been forced from their homes [2]. As of 25 September 2023, 3.67 million people were estimated to have been displaced internally within Ukrainian borders [17]. These estimates are based on a random digit dial telephone survey aiming at generating a nationally representative sample of 2,000 individuals at each monthly round [17]. While consistent with high frequency estimates based on Facebook data [18], these estimates cannot deliver population-level estimates of population displacement for subnational areas at high granularity, or high-temporal frequency. Our approach offers high frequency population displacement estimates to complement data derived from traditional data streams.

Methods

We use Global Positioning System (GPS) location data from mobile phones, in combination with WorldPop and UNHCR refugee data, to measure internal population displacement.

Data sources

GPS location data. The primary source consists of GPS location data from 25 million unique mobile phone devices. The data include daily GPS locations (longitude and latitude) in Ukraine, their accuracy and time stamps from January 1st 2022 to August 31st 2022. Data from digital mobile phone applications are known to contain biases as they typically represent the behaviour of a segment of the population. To mitigate any potential biases from the use of information from a single source, we use data collected from a range of mobile applications comprising a variety of users and purposes.

Baseline population data. We use $100m^2$ gridded population data to establish the baseline population before the conflict in Ukraine in 2020. We utilise unconstrained population estimates from worldpop.org. These were the most up-to-date population estimates available for our analysis. We spatially aggregate the WorldPop population counts to create baseline population datasets at the raion and oblast levels. These population estimates are then used to derive population-level estimates of internal displacement, as described below.

Refugee data. We use UNHCR daily counts of people entering and leaving Ukraine. We accessed these data from an archived version of the UNHCR website available via The Internet Archive. These data included daily cross-border movement records from the start of the full-scale invasion in Ukraine until August 16 2022. We calculate a cumulative net count by subtracting the number of people entering Ukraine from those leaving the country. We use this count to more accurately estimate the number of internally displaced people in Ukraine by discounting the number of people who moved overseas from the baseline population.

Computation of population-level displacement estimates

We obtain population-level estimates of internal displacement by correcting population counts derived from the identified home location based on our smartphone GPS data, to make them representative of the overall population. That is, we correct mobile phone-derived population estimates to account for differences in the use of mobile phone technology across locations in Ukraine and over time. To this end, we adapted a deterministic model proposed by Leasure and colleagues [18]. Intuitively the approach involves first establishing our baseline population; that is the pre-war population of Ukraine. We use population data from WorldPop for 2020. Second, we identify the baseline number of mobile phone users in Ukraine before the start of the full-scale invasion by aggregating the number of unique devices in each home location based on our GPS mobile phone data. Third, these two sets of baseline estimates are used to compute the baseline mobile phone penetration rate in each location before the start of the full-scale invasion.

Results

We first estimate the extent of daily internal population displacement at the oblast and raion level (Figure 1). We estimate that over 5 million people were internally displaced from their oblast of residence by April 2022 reaching an average of about 10 million in late July and August 2022. Figure 1 reveals a drop in population displacement during mid-June and mid-July, coinciding with a pattern of return displacements primarily to the cities of Kiev and Kharkhiv. In addition to return movements, subsequently higher but fluctuating levels of movement after mid July seem to reflect the shifting dynamics of the armed conflict towards southeastern Ukraine where war fire intensified during this period.

Our contribution is to generate geographically granular estimates of internal displacement at the raion level leveraging the high spatial precision of GPS data. As anticipated, the levels of raion-level displacement consistently exceeds those of oblast-level displament as they reflect movements that cannot be captured at higher levels of spatial aggregation: raions within the same oblast's boundaries capturing the fact that most displacement tends to occur over short distances. Our raion-level estimates indicate a rise and peak of over 17 million displaced people in mid June 2022 following the start of the Russian invasion of Severodonetsk. Around 90% of the buildings and infrastructure is estimated to have been destroyed or damaged after the capture of Severodonetsk. From mid July, our estimates indicate a rise in population displacement at the oblast level, but such increase is not reflected at the raion level, indicating that the most displacement that took place during this time tended to occur over long distances involving a cross of oblast boundaries displaying distance distributions).

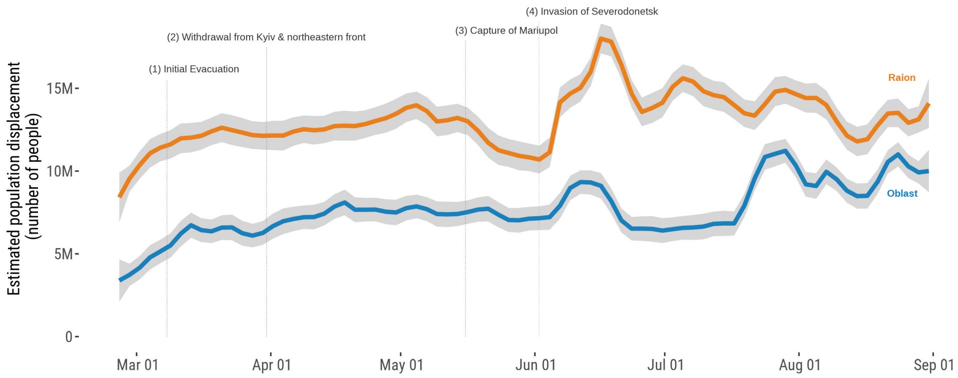

Figure 1: **Estimated daily number of displaced population. February to August 2022.** The number of displaced people is estimated as the difference between the population in a region for a given day after the start of the war and the population before the war in 2020 - see Section 1.

Our findings are consistent with existing estimates. We compare our oblast-level displacement estimates with existing estimates derived from an UN-IOM survey and Facebook data. The shape of the temporal evolution of population displacement is remarkably consistent. Though, we identify some discrepancies. Our estimates tend to be higher than those produced by Leasure et al. by approximately 250 thousand people across the time series. The difference can be explained by Leasure et al.'s estimates are affected by power outages in the Donetsk and Luhansk regions resulting in zero or small numbers for various dates. Similarly, our oblast-level estimates are noticable greater than the IOM figures in June and August. We assume that this is because our estimates include data from Crimea, and there was significant movement from and to Crimea to Russian-occupied Ukrainian territory and Russia during these months. This is as Russia started a "volunteer mobilisation" and deployed new troops and logistics to support an a frontline extending from Zaporizhzhia to Kherson, along the Dnieper River. If we exclude Crimea, our estimates are much closer to IOM and Leasure et al.'s estimates.

Future work

Future work aims to identify the spatial patterns of internal population displacement in Ukraine. We seek to determine the key areas of origin and destination by examining the patterns of net migration gains and losses. Additionally, we aim to measure the national extent of return movement as the conflict has progressed. We will validate our estimates against published estimates of internal population displacement by the IOM and Leasure et al. based on survey and Facebook data, respectively. Our approach will demonstrate how GPS data can be used to produce population-level estimates of internal population displacement. Our evidence will illustrate the utility of GPS-based estimates of internal population displacement to inform humanitarian assistance efforts identifying places concentrating displaced individuals.

References

[1] UNHCR. Mid-year trends 2023 (2023). URL https://www.unhcr.org/mid-year-trends-report-2023.

[2] UNHCR. *Global trends. Forced displacement in 2022.* (United Nations High Commissioner for Refugees. UNHCR, Copenhagen, Denmark, 2022). URL https://www.unhcr.org/refugee-statistics.

[3] Rowe, F. Using digital footprint data to monitor human mobility and support rapid humanitarian responses. *Regional Studies, Regional Science* **9**, 665–668 (2022). URL http://dx.doi.org/10.1080/21681376.2022.2135458.

[4] Sarzin, Z. I. Stocktaking of global forced displacement data. *World Bank Policy Research Working Paper* (2017).

[5] Rowe, F. *Big data*, 42–47 (Edward Elgar Publishing, 2023). URL http://dx.doi.org/10.4337/9781800883499.ch09.

[6] Salah, A. A., Pentland, A., Lepri, B. & Letouzé, E. (eds) *Guide to Mobile Data Analytics in Refugee Scenarios* (Springer International Publishing, 2019). URL http://dx.doi.org/10.1007/978-3-030-12554-7.

[7] Hoglund, K. & Oberg, M. *Understanding Peace Research Methods and Challenges* (Routledge, 2011). URL http://dx.doi.org/10.4324/9780203828557.

[8] Checchi, F., Stewart, B. T., Palmer, J. J. & Grundy, C. Validity and feasibility of a satellite imagery-based method for rapid estimation of displaced populations. *International journal of health geographics* **12**, 1–12 (2013). URL https://doi.org/10.1186/1476-072X-12-4.

[9] Nyhan, M. *et al.* "Exposure Track" - The impact of mobile-device-based mobility patterns on quantifying population exposure to air pollution. *Environmental science & technology* **50**, 9671–9681 (2016). URL https://doi.org/10.1021/acs.est.6b02385.

[10] Dewulf, B. *et al.* Dynamic assessment of exposure to air pollution using mobile phone data. *International journal of health geographics* **15**, 1–14 (2016). URL https://doi.org/10.1186/s12942-016-0042-z.

[11] Huang, H., Cheng, Y. & Weibel, R. Transport mode detection based on mobile phone network data: A systematic review. *Transportation Research Part C: Emerging Technologies* **101**, 297–312 (2019). URL https://doi.org/10.1016/j.trc.2019.02.008.

[12] Kim, J. Y., Kubo, T. & Nishihiro, J. Mobile phone data reveals spatiotemporal recreational patterns in conservation areas during the COVID pandemic. *Scientific Reports* **13**, 20282 (2023). URL https://doi.org/10.1038/s41598-023-47326-y.

[13] Lu, X. *et al.* Unveiling hidden migration and mobility patterns in climate stressed regions: A longitudinal study of six million anonymous mobile phone users in Bangladesh. *Global Environmental Change* **38**, 1–7 (2016). URL https://doi.org/10.1016/j.gloenvcha.2016.02.002.

[14] Grantz, K. H. *et al.* The use of mobile phone data to inform analysis of COVID-19 pandemic epidemiology. *Nature communications* **11**, 4961 (2020). URL https://doi.org/10.1038/s41467-020-18190-5.

[15] Ranjan, G., Zang, H., Zhang, Z.-L. & Bolot, J. Are call detail records biased for sampling human mobility? *ACM SIGMOBILE Mobile Computing and Communications Review* **16**, 33–44 (2012). URL http://dx.doi.org/10.1145/2412096.2412101.

[16] Grantz, K. H. *et al.* The use of mobile phone data to inform analysis of covid-19 pandemic epidemiology. *Nature Communications* **11** (2020). URL http://dx.doi.org/10.1038/s41467-020-18190-5.

[17] IOM. Ukraine internal displacement report - general population survey - round 11 (25 november - 5 december 2022). Tech. Rep., International Organization for Migration (IOM) (2022).

[18] Leasure, D. R. *et al.* Nowcasting daily population displacement in Ukraine through social media advertising data. *Population and Development Review* **49**, 231–254 (2023). URL https://doi.org/10.1111/padr.12558.