DATA FOR POLICY PROCEEDINGS PAPER

Mobile phone data for anticipating displacements: practices, opportunities, and challenges

Bilgeçağ Aydoğdu[1] , Özge Bilgili[3], Subhi Güneş[2] and Albert Ali Salah[1]

[1]Dept. Information and Computing Sciences, Universiteit Utrecht, Utrecht, The Netherlands
[2]Turkcell Technology, Istanbul Turkey
[3]Dept. Interdisciplinary Social Science, Universiteit Utrecht, Utrecht, The Netherlands
Corresponding author: Bilgeçağ Aydoğdu; Email: b.aydogdu@uu.nl

Abstract

The global number of individuals experiencing forced displacement has reached its highest level in the past decade. In this context, the provision of services for those in need requires timely and evidence-based approaches. How can mobile phone data (MPD) based analyses address the knowledge gap on mobility patterns and needs assessments in forced displacement settings? To answer this question, in this paper, we examine the capacity of MPD to function as a tool for anticipatory analysis, particularly in response to natural disasters and conflicts that lead to internal or cross-border displacement. The paper begins with a detailed review of the processes involved in acquiring, processing, and analyzing MPD in forced displacement settings. Following this, we critically assess the challenges associated with employing MPD in policy-making, with a specific focus on issues of user privacy and data ethics. The paper concludes by evaluating the potential benefits of MPD analysis for targeted and effective policy interventions and discusses future research avenues, drawing on recent studies and ongoing collaborations with mobile network operators.

Policy Significance Statement

This article aims to provide a comprehensive overview to policymakers and practitioners on mobile phone data (MPD) sources and analysis techniques. By discussing the practices, challenges, and opportunities of MPD-based anticipatory research, we seek to push forward evidence-based policy-making in the field of forced displacement settings.

---

1. Introduction

According to the United Nations Refugee Agency (UNHCR), more than 117.3 million individuals were forcibly displaced in 2023 (UNHCR 2023). As indicated by the Global Trends of UNHCR, "This constitutes a rise of 8% or 8.8 million people compared to the end of 2022 and continues a series of year-on-year increases over the last 12 years" (UNHCR 2023, p. 6). Of these individuals, 68.3 million were internally displaced while 38.5 million were refugees or asylum seekers. The absolute number of forcibly displaced individuals has reached its highest point in the past years due to the ongoing war in

Ukraine, the conflicts in Syria and Afghanistan and the growing number of natural disasters. In fact, according to the Internal Displacement Monitoring Centre (IDMC), in 2023, more than 56% of internal displacements were caused by natural disasters (IDMC 2024). The disruptive effects of climate-change and environmental degradation clearly contribute to the growing trend of more individuals having to leave their homes in search of safety.

While forced displacement is far from being a new phenomenon and there is a wealth of experience in understanding its impacts on individuals and communities, addressing the challenges caused by forced displacement remains a major challenge for both national and local governments as well as humanitarian agencies. These challenges are multi-faceted and include preparing for and responding to the short-and long-term needs of displaced populations. Ranging from providing shelter, food and nutrition, and health care services to relieving the suffering of the most vulnerable, to being effective, humanitarian aid relies on significant infrastructural investment, strong governance, and adequate financial resources. Important to note however that more than 75% of forcibly displaced individuals[1] seek refuge in low and middle-income countries which themselves may face economic and infrastructure challenges. In these contexts, accommodating displaced populations and delivering essential humanitarian aid can impose additional pressures on host communities and their resources, further complicating the implementation of effective solutions.

For the development and implementation of well-targeted and efficient measures and programs based on evidence, it is crucial to anticipate forced displacements, understand their underlying mechanisms and characteristics, and be able to predict and monitor their multifaceted effects on those who are forcibly displaced and the host communities. Evidence-based approaches to humanitarian aid have for decades relied primarily on rapid needs assessments and key informant interviews to quickly gather information on displacement situations. Despite the fact that these data collection methods have been crucial for immediate action, they are not without limitations. Due to their retrospective nature, conventional data collection methods can quickly be outdated and they do not always have the in-depth insights for predictive analysis on human mobility and behavior.

To address these limitations, big data are increasingly used as complementary data sources by researchers and policymakers. For example, the analysis of satellite imagery helps identify new settlements or changes in existing ones (Tatem et al. 2020). Social media data, particularly from platforms like Twitter and Facebook, are utilized to gain insights into displacement patterns and perceptions. In addition to these, mobile phone data (MPD) is increasingly recognized as one of the most prominent data sources for analyzing forced displacements, including data collected by mobile network operators and location data collected by mobile application owners.

MPD sources have a high potential to address various issues faced in the realm of data collection on forced displacements. The scale, temporal, and spatial granularity, and the rich attributes of MPD can provide the much-needed displacement indicators in a timely and cost-effective manner. These indicators can then be used in models for predicting or explaining various aspects of displacements. Consequently, in recent years, policy makers and practitioners have begun to show interest in integrating MPD sources in displacement-related statistics. However, the potential opportunities and challenges of this endeavor are not yet fully understood. We believe that a comprehensive discussion is needed in terms of how MPD can support anticipatory decision-making and facilitate translating MPD-based information to policy and program making in forced displacement settings.

In this paper, we aim to outline the practices, opportunities, and challenges of using MPD in forced displacement settings, with a particular focus on predicting the populations potentially vulnerable to crises and natural disasters. Although the literature is rich in discussing the variety of indicators that show the population movements through MPD, a great majority of these analyses remain retrospective and descriptive. We discuss the existing approaches that are used for analyzing displacement events

---

[1] https://www.unhcr.org/refugee-statistics/insights/explainers/refugee-hosting-metrics.html, Accessed 19 April 2024.

through MPD sources in two categories; ex-ante, which tries to create useful models to assess the potential future crisis-induced population movements, and ex-post, which looks at past historical data for a better understanding of what has happened after forced displacements occurred. In this framework, anticipatory approaches correspond to ex-ante approaches, and we will refer to them as such from now on. By doing so, we seek to contribute to a better understanding of how MPD can be used for anticipating displacements.

Ex-ante approaches using MPD received less attention than ex-post approaches, especially because of the challenge in the validation of the predictions, yet they are important for preemptive crisis management and improved decision-making. Despite the limited number of studies, there is a great diversity in MPD-based indicators and features used, as well as methodological approaches followed in existing ex-ante studies. To better understand these practices, we review the commonly used MPD sources (Section 3), features and indicators (Section 4), as well as methods, which we discuss under forecasting and scenario-based approaches (Section 5). We summarize the steps in data processing for ex-ante approaches in one common framework described in Figure 1. We also discuss the opportunities, such as the potential of new data processing platforms, as well as the challenges of using MPD-based approaches for forward-looking decision and policy making, including hurdles of data access and governance (Section 6).

2. Anticipating displacements

There are multiple drivers of forced displacement. These include not only conflict, human rights abuses, violence, and prosecution but also a diverse set of factors that relate to acute poverty, effects of climate change, natural disasters, and environmental degradation. Accordingly, forced displacements vary in complexity, severity, duration, and spatial coverage (Cantor 2024). In the face of such diverse crises, studying mobility as an adaptation strategy and identifying forced displacement patterns have been an empirical and conceptual challenge for academics, and led to various comparative studies and typologies (Cantor 2024, Steele 2019). It is largely accepted that the individual characteristics of those who can respond to and seek refuge elsewhere show great variation in the face of macro-level factors (Adhikari 2013, Miller 2020). Accordingly, systematic and comprehensive (individual and contextual) data are
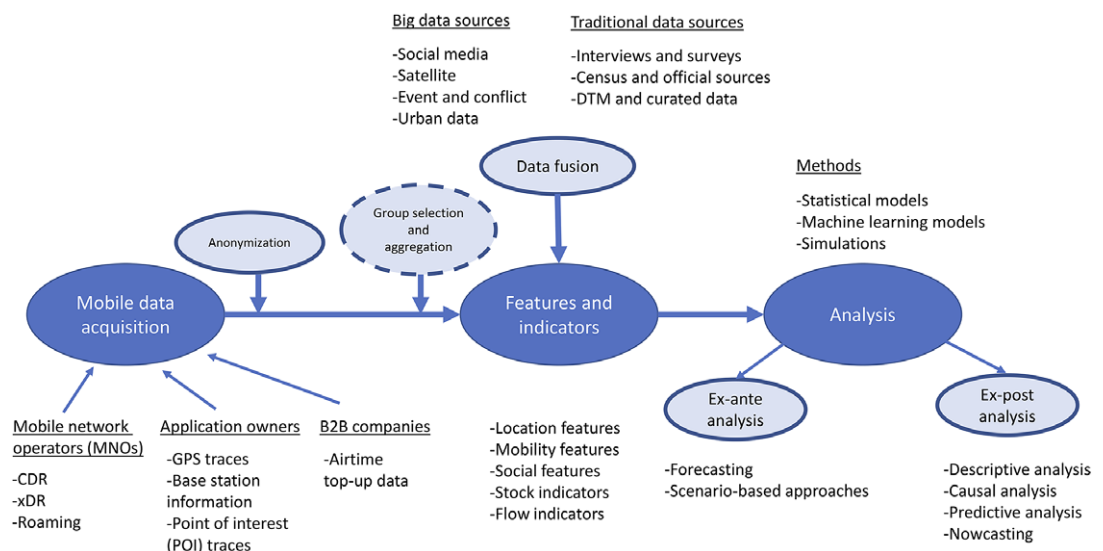

Figure 1. *An overview of the MPD processing pipeline for analyzing displacements.*

necessary to identify the needs of diverse groups of displaced persons (DPs)[2] and to enable the delivery of complex sets of services they need during their displacement.

Policy makers and actors managing displacement crises need to anticipate both short-term evacuation needs and long-term resettlement or integration needs to effectively reduce humanitarian impacts (Thalheimer et al. 2022). Data collaboratives can be very valuable in addressing their information needs. These are public-private partnerships for collaboration and information sharing across sectors and actors, permitting the use of private datasets, such as MPD, stored within mobile network operator (MNO) and company servers (Verhulst & Young 2019). Ex-ante analysis of displacements via MPD can (1) improve forecasting, (2) provide guidance for crisis management, (3) help create more realistic crisis-induced mobility scenarios. A large number of studies focus on ex-ante approaches either forecast or simulate temporary and often rapid movement of people away from immediate threats or hazards. Forecasting displacements at longer periods of time is much more difficult, as it requires factoring in complex issues around prolonged displacements such as return migration.

Disasters and armed conflicts are two major sources of displacement. In particular, in the past two decades, there has been a prominent debate on how to incorporate climate-induced displacements in human rights and protection dimensions (Bradley & Cohen 2010, Ferris 2007, McAdam 2012, Goodwin-Gill & McAdam 2017). However, the binary distinction between conflict and disaster-induced displacement can be considered rather artificial. In reality, forced mobilities are caused and reinforced by the interaction between different factors (Abel et al. 2019). Namely, while climate change-related problems (e.g. water insufficiency, natural hazards, and environmental degradation) can lead to new conflicts that cause displacement, in return these factors can also intensify existing displacement situations that were in the first place caused by conflict. Moreover, displacement conditions are also affected by other economic, social, and political drivers that generally influence migration patterns. Therefore, it is important not to see these movements as discrete and independent categories of movements and situate them within a wider global development and humanitarian context (Mitchell & Pizzi 2021, Morales-Muñoz et al. 2020).

Having said this, an overview of the literature shows that empirical research on two determinants remains rather distinct, especially when we look at research using big data. As it is the focus of this paper, it is important to note that MPD-based research is much more conducted in disaster settings. During armed conflict situations, access to satellite or mobility data can be blocked by the data owners to obstruct the malicious use of data by the conflicting parties. A recent example of this was observed during the onset of the Russia-Ukraine war when Google disabled the live traffic features in Google Maps in Ukraine to protect users' safety [3]. Furthermore, the locations of mobile base stations can be considered critical communication infrastructure information and may be hidden. Studies using MPD for conflict-induced displacement are often published after some time to minimize these risks, and hence they are not anticipatory. In contrast, in the aftermath of a large natural disaster, the incentives for protecting and hiding data are significantly lower. Policy makers who are responsible for managing the crisis are much more motivated to share and work with MPD data. In this case, the use of data to alleviate the immediate impact of the disaster more easily outweighs privacy-related concerns.

In Figure 1, we introduce an overview of the MPD processing pipeline, which starts with mobile data acquisition, followed by anonymization and aggregation to remove sensitive and private information. All designated sources are potentially large-scale, and require special legal provisions to be shared, due to

---

[2] If forced displacement occurs within the borders of a nation state, those individuals who seek flight are considered as "internally displaced persons" (IDP). Individuals who are displaced across national borders are asylum seekers and refugees. Asylum seekers apply for formal recognition of becoming a refugee, often through UN agencies (UNHCR 1951). Both IDPs and people displaced across borders are subject to various forms of international protection and may be eligible for certain rights and services where they seek refuge, though the specificities of these rights and services vary based on their legal status (Berchin et al. 2017).

[3] https://www.business-humanrights.org/en/latest-news/google-disables-maps-traffic-data-in-ukraine-to-protect-citizens-during-russian-military-invasion/, Accessed 19 April 2024.

both privacy risks and commercial value of such data. We will not discuss the legal aspects of this work, but point out several past data collaboratives, which established guidelines and legal precedents, to facilitate new initiatives.

3. Mobile data acquisition

3.1. *MPD types and sources*

MPD encompasses a range of data types collected by various stakeholders, such as mobile application owners, MNOs, GPS data providers, and companies that sell services to MNOs. An early important work that pointed out the potential of mobile data was the reality mining study, which collected application data from phones distributed to 100 subjects (Eagle and Pentland 2006). This mode of data collection continues to be important, as mobile phone applications can ask explicit permission from their users, and then access information like GPS traces, wireless router signatures, mobile base tower details, phone status, and usage information. Furthermore, a number of approaches are available for determining the location of the user through applications, including simple cell activity (i.e. base tower locations), GPS sensors, additional signal strength, and time or arrival differences (Ghahramani et al. 2020).

MPD-based studies in displacement literature use overwhelmingly two MPD sources; call detail records (CDR) and GPS data, respectively. CDR data are primarily collected by MNOs for operational and billing purposes. It contains meta-information of calls and text messages, documenting elements such as caller and receiver identifiers, call duration, timestamp, and the antennas used during the communication. Every phone call and message routed through the network of the operator generates one line of data in the CDR database. Most mobility analyses have used cell tower locations contained in the CDR as indicating the broad area of where a user is at the time of the record (Blondel et al. 2012, Salah, Pentland, Lepri, Letouzé et al. 2019), but some approaches include mobility modeling to improve these estimates (Lind et al. 2017).

GPS data, on the other hand, are collected by companies, like Safegraph[4], Foursquare[5], and Cuebiq[6], as well as big tech companies (Meta, Google, etc.). It includes the precise location of users recorded by latitude, longitude, and a timestamp. GPS data differs from the data collected through CDR in terms of volume, velocity, veracity, and data biases. GPS has a higher temporal frequency and more spatial granularity compared to CDR, so the types of features and indicators that can be developed with GPS differ from CDR. Some tech companies sell or share GPS-based data insights as indicators without always being fully transparent about their calculation methods. We will not report on such pre-calculated metrics in this paper, as their opaqueness is a distinct drawback for policy makers. There are also MPD sources rarely used in the literature with high potential to anticipate displacements; those we discuss in Section 6.

MPD sources are primarily relevant to anticipating displacements as they collect information on the location of users. As we will explain in Section 4, the location information is used for developing displacement indicators. Many studies have explored the capabilities of mining CDR and GPS in deriving indicators of migration and displacements by approximating users' home, work, and visited locations. Both data sources perform well on these indicators, although GPS gives more detailed insights into people's whereabouts. As CDR includes call information, it is much better for deriving indicators of social ties and communication patterns, which are important factors influencing the destination choices of DPs. GPS is frequently used in ex-ante studies in anticipating immediate mobility after crises. The high spatio-temporal frequency of GPS enables researchers to mine features useful for forecasting and simulating evacuations. CDR has been also used in ex-ante studies, but most of its applications have been ex-post.

---

[4] https://www.safegraph.com/, Accessed 19 April 2024.
[5] https://location.foursquare.com/resources/blog/news/geolocation-101/, Accessed 19 April 2024.
[6] https://www.cuebiq.com/, Accessed 19 April 2024.

3.2. Data anonymization, aggregation, and group selection

MPD are sensitive, as it is possible to track the movements of individuals in great detail and depending on the resolution, to infer a large amount of personal information about the users. For this reason, MPD should be carefully anonymized and aggregated to the level that individuals cannot be identified and tracked. De Montjoye et al. (2016) investigated anonymization of CDR in detail and proposed to perform data aggregation spatiotemporally within the MNO (i.e. before sharing data) to ensure privacy. CDR data are typically aggregated hourly, and the spatial resolution is at the base tower level for short-term tracking, and at the district level for long-term tracking of individuals. In no scenario are the actual phone numbers shared.

Other approaches include question-answer models, where the raw data remains behind company firewalls, and only authorized queries are run to obtain summary results (Lepri et al. 2018). Lastly, sharing pre-computed indicators is a common approach that is proposed for finding the balance between privacy protection and utility (De Montjoye et al. 2018). This method is especially useful in the context of migration and displacement research but requires bringing domain expertise together with expertise on processing MPD. There are a few organizations, such as DataPop Alliance[7] and Flowminder[8], working on sustainable partnership models with MNOs in developing countries to enable sharing of indicators and data insights to policy makers in a sustainable way. Flowminder offers a privacy-preserving indicator calculating tool called Flowkit that can be installed on servers of MNOs (Power et al. 2019). Following their approach requires a substantial amount of technical, financial, and human resources, which is not easy to gather quickly in case of a humanitarian crisis.

In order to provide a more granular analysis, groups may be identified in the data based on demographics, status (e.g. refugee), or specific vulnerabilities. Group selection entails filtering and sampling of groups of interest in MPD sources. MPD can give insights on whether the subject is living in a poor or expensive neighborhood (wealth), regularly visits a mosque on Fridays (religion), and mobility during special days and events (ethnicity). User registration information, app usage, mobility patterns, and incoming and outgoing communication patterns are used for creating demographic proxies on wealth, gender, age, nationality, and religion. Such proxies then can be used for selecting the groups for which indicators and features will be calculated. Adding and inferring more demographic information poses higher risks in terms of privacy issues, therefore the data have to be always aggregated and scaled. It is possible to perform aggregation close to the source, i.e. as a part of data collection. Demographic flags and indicators computed by the data company can be directly used for grouping if they are useful for the analysis. Another approach would be to use clustering during the analysis stage. In our own research on the 2023 Hatay Earthquake, we could discern groups of DPs going to different cities via MPD, but further qualitative insights were necessary to interpret such "discovered" groups.

MNOs collect data via cell towers, which are more vulnerable to local infrastructure damage and GPS-based data. These vulnerabilities can hamper data collection and decrease the quality of the data, negatively impacting the accuracy of predictions and nowcasting efforts.

3.3. Debiasing

MPD proves to be a valuable resource in developing indicators of mobility thanks to its spatio-temporal granularity (described further in Section 4), but it faces challenges in terms of certain irregularities and biases, which are partially addressed via debiasing approaches.

The first bias that MPD may have is in data selection. The mobile phone penetration rate in the country of analysis is one indicator, and a low rate means that a portion of the population will systematically miss from MPD. Sample statistics should be appropriately scaled to account for this (Sekara et al. 2019). Conversely, in countries with high-quality internet coverage, there is a growing trend of using instant messaging platforms as opposed to calls for communication. This is a harder factor to consider that

---

[7] https://datapopalliance.org/, Accessed 19 April 2024.
[8] https://www.flowminder.org/, Accessed 19 April 2024.

particularly affects CDR, but not GPS. In general, as CDR is event-driven, it may underestimate features like travel distances and movement entropy (Zhao et al. 2016). Additionally, the penetration of the mobile phones per area, and the market share of MNOs may be spatially non-homogeneous (i.e. an MNO may be more popular in a particular city or for a particular community). These factors are important as the data are typically collected from a single MNO. To deal with market share biases, the estimates should be scaled with debiasing coefficients obtained from population statistics and market share reports of the MNO itself.

In developing countries, network coverage may be limited to urban areas, or certain services might not be accessible in certain rural areas. As rural areas in general have less cell tower coverage, location estimates via CDR will be less accurate. Very poor areas may have reduced phone ownership, or phone-sharing may be a common practice. The latter will be more difficult to deal with, as it becomes difficult to estimate home and work locations for shared phones. Qualitative insights are required to estimate to what extent this can affect the analysis results. Furthermore, disasters and conflict situations can exacerbate issues via damage caused to the network infrastructure; destroyed base stations will affect CDR collection, and mobile base stations may be substituted temporarily. GPS will be less affected, but malfunctioning electricity grids may prevent Internet and phone usage, causing data gaps for all MPD, as well as social media sources (Leasure et al. 2023).

The user base of MNOs may be potentially very large, ensuring good population coverage, but certain demographic groups will be under-represented. For example, elderly people use mobile phones much less than younger individuals, and small-aged children are completely excluded, as they do not (officially) have access to phones. Conversely, GPS from mobile application data will be biased towards people actually using the application and will require post-stratification. In some regions, men are more likely than females to own phones or females share phones (Marshall et al. 2016). In some countries or cultures, even when females own phones, they may be predominantly registered under their husbands' or father's name (Sekara et al. 2019). This makes it more difficult to use demographic tags already stored with the original data in group-based aggregation. Some of these tags may be noisy. This means that aggregated analysis should be preferred over analysis of individual records. In the context of disasters and conflicts, users added to the dataset after the crisis onset can be filtered out in user selection (Lu et al. 2016). Furthermore, when working with MPD, seasonal trends and increased activities around special days and events should be taken into account.

Apart from data collection biases, the data aggregation approach itself may introduce issues by removing cells with too low mobile activity to protect the privacy of the individuals in those cells. This is called the confidentiality threshold (Vespe et al. 2021), and suppression of small cells can bias counts downwards systematically (Kohli et al. 2024). To debias data aggregated in this way, low-frequency counts can be boosted and high-frequency counts can be flattened (Guan et al. 2024).

4. Features and indicators of MPD

4.1. MPD features

We distinguish between MPD features, which are input variables that are used in ex-ante models, and MPD indicators, which are stock and flow measures of DPs. MPD features are obtained via feature extraction, which refers to transforming raw MPD into meaningful variables for analysis, and feature selection, which refers to choosing of most relevant variables for further analysis. The MPD-based features consist of quantitative measures and indices. They are often easier to calculate than stock and flow indicators, as they have simple quantitative definitions. MPD features are commonly used for predicting evacuation routes, and destinations both for exploring scenarios. The choice of features for ex-ante analysis should be guided by the type of MPD, and the methodologies that will be used. The data processing pipeline described in Figure 1 need to be adjusted based on the chosen methodology to extract the necessary features, as many features are calculated before aggregation.

Both GPS and CDR can be mined for mobility-related and location-related features, whereas CDR can also be mined for communication-related features. In Table 1, we give an overview of commonly used

Table 1. *Summary of commonly used displacement features and indicators for anticipatory analysis*

	Predictive potential	Limitations	Use in ex-ante studies
Features			
Home location	Shifts in home-location are the first signals of displacements.	Mobile-usage discrepancies can create bias and reduce accuracy in predictive models.	It is the main input for flow indicators (Pappalardo et al. 2021).
Important places	High potential for predicting individual displacement behavior and destination choices.	Longitudinal data is needed to accurately identify important places (work, social, etc.). Usually GPS works better than cellular data types.	Behavior modeling and evacuation simulations (Song et al. 2016).
Mobility metrics	Pre-crisis mobility patterns give insights into individual wealth and employment.	The relationships between pre-crisis and post-crisis mobility are less straightforward than other features.	ML models (Khaefi et al. 2018)
Social relationships	Can predict likely evacuation destinations and support networks at individual or area levels.	Requires careful handling of privacy. Constructing social networks is to be avoided.	Incorporated in destination choice models (Song et al. 2016).
Indicators			
Origin–destination matrices	Useful for predicting destination choices.	OD matrices tend to be sparse due to the specificity of corridors with the flow of DPs.	In mathematical models (Isaacman et al. 2018, Anyidoho et al. 2023), in ML models (Khaefi et al. 2018), in ABS (Yin et al. 2020), and in RL (Fan et al. 2021).
Population density maps	Can help predict the impact of crises in scenario-based approaches.	Requires integration with other data sources for predictive power.	Population maps are often used in scenario-based approaches (Balistrocchi et al. 2020, Sudo et al. 2016, Pang et al. 2020).

MPD features and indicators, their predictive potential, and limitations for ex-ante studies. Home locations are the most essential feature for flow indicators, calculated with rule-based algorithms that capture the location of signals when users are most likely to spend time at home (Pappalardo et al. 2021). In addition to home location, places of work, leisure, and other important places (such as schools, hospitals, or religious places visited) can be mined from users' routine mobility behavior, and behavior around special days (national and religious holidays, festivals, weekends, etc.) can be modeled. Important place features are extensively used in ex-ante studies. For simulation approaches, identifying important

places helps to realistically generate mobility behavior in the aftermath of disasters. Mobility-related features include radius of gyration, which refers to the measure of the typical distance a person travels from their average location (usually home location) (Gonzalez et al. 2008), daily traveled distances, which is a direct measure of displacement at the individual level (Lu et al. 2012), random entropy, which captures the degree of predictability if each location is visited with equal probability (Song et al. 2010).

Some measures of human mobility use physical formulas to describe the extent and frequency of mobility patterns. Pre-crisis mobility is indicative of various attributes of the individual such as wealth level, and employment status, therefore they are important features to include in ex-ante models. In addition, various ex-post studies compare distributions of mobility-related features across time and space to depict the extent of displacements (Lu et al. 2012, Yabe, Tsubouchi, Fujiwara, Sekimoto & Ukkusuri 2020). Lastly, CDR is used for mining social relationship-related features using the interactions between users through messages and calls. For instance, it is possible to estimate the amount of connectedness between different cities through mutual call statistics. This may be relevant for predicting crisis-induced displacements from one city to another (Song et al. 2016).

4.2. MPD indicators

The displacement indicators are usually reported as flows, which refer to the number of DPs who move due to a crisis, and stocks, which refer to the total number of DPs in a given area (EGRIS 2018). Instead of following this terminology, MPD-based studies tend to use mobility-related features to give a measure of displacements. However, such metrics are rarely easily accessible to policymakers, which creates a disconnection between the potential users of MPD and the existing practices. To bridge this gap, and make the MPD studies more accessible, we group MPD-based indicators of displacement into two categories, namely, flow and stock indicators, respectively. In addition, MPD sources are also usable for assessing social aspects such as DP integration and social segregation, but these analyses are out of the scope of this paper. Displacement indicators are usually what interests the policymakers for anticipatory decision-making. The level at which the indicators are calculated can change depending on the data source and aggregation steps taken. The indicators can be aggregated at different geographical levels (e.g. sites, grids, administrative units), and at different temporal levels (hours, days, weeks). While MPD offers rich data for developing these indicators, we recognize that displacement prediction remains a complex task. Events like conflicts or natural disasters can dramatically alter normal mobility patterns. Despite the chaotic aspects of crises, some studies (Lu et al. 2012, Wang & Taylor 2016) indicated that there is consistency between crisis and normal time mobility patterns. Such predictability of the statistical properties of features and indicators is what allows the creation of ex-ante models.

4.2.1. Flow indicators

Flow indicators estimate the movement of individuals from one location to another over a specified duration. To measure the flow of DPs the first step is to calculate the home location of users across a specific period, and identify the sustained shifts that occurred between different locations. Then the individual level changes in home locations are aggregated into (1) origindestination (OD) matrices, and (2) inflow and outflow indicators. These indicators can be disaggregated demographically, if some groups were identified previously, enabling a more detailed overview of flows.

One challenge in identifying the flow of DPs is the attribution error, which refers to falsely attributing the home location changes to permanent displacements, even when they are caused by other factors such as seasonal mobility. In acute conflicts and slow-onset disasters, the attribution error is a bigger concern than sudden-onset disasters, as displacement effects tend to be highly visible with the latter type of crises. One common way of reducing this error is to calculate a baseline level of the flow indicator from the previous periods and compare it to the period with crisis-induced mobility. This requires a long data collection period that pre-dates the crisis.

The most commonly used flow indicator is the OD matrix, which is a two-dimensional array that represents the volume of movement or flow of DPs between pairs of geographic locations. OD matrices

allow for predicting the destinations both for short-term evacuation patterns and for long-term displacements. They are commonly used in forecasting with population-level mathematical models such as gravity and radiation models. For instance, Anyidoho et al. (2023) used OD matrices to predict the evacuation destinations in the aftermath of various hurricanes. Similarly, Isaacman et al. (2018) used OD matrices as ground truth to evaluate and refine gravity and radiation models to anticipate drought-induced mobility. These models predict directly by using the OD matrix as a target variable, whereas simulation models predict the displacements at the individual level and aggregate the predictions into OD matrices (Fan et al. 2021, Yin et al. 2020).

In addition to OD matrices, there are also inflow and outflow (IO) indicators measuring the number of DPs leaving from or arriving at a region in a given time period. IO indicators are commonly used in ex-post studies to demonstrate the effect of a crisis on the dispersal of the affected population. For instance, Shibuya et al. (2024) calculated indicators such as the relative frequency of in- and outflows using GPS to assess internal displacement patterns in Ukraine during the Russia-Ukraine war. In ex-ante studies, IO indicators are not commonly used in prediction models, as they are highly aggregated and tend not to perform well with data-intensive methods. MPD-based IO indicators are also used for comparisons with official figures, which are usually shared in a similar format.

4.2.2. Stock indicators

Stock estimation refers to the calculation of the size of DPs in a given location within a specified time period. To create stock indicators during rapid-onset crises, the simplest approach is to define DPs based on how the crisis affects different areas over time and count the number of these DPs present at various destination locations. Marzuoli & Liu (2018) used this approach to calculate the changes in the stock of people, who were displaced after the hurricanes Michael and Florence, across various areas in Florida. Similarly, Bengtsson et al. (2011) used this approach to map the stocks of DPs across the country, a couple of weeks after the Haiti earthquake in 2010. This approach is often used in ex-post, and less in ex-ante approaches. Another method involves using MNO data to estimate the population shifts near cell towers, as such data can estimate local population densities dynamically. Deville et al. (2014) demonstrated how to use CDR and satellite data to create detailed maps of the population, which can help dynamically map the population in different time periods (weekdays vs. weekends, day vs. night, etc.). Population density maps have been used to demonstrate the changes in the stock of DPs in the context of floods (Balistrocchi et al. 2020). These maps are usually compared to a baseline level to assess the level of displacements. In addition, MPD-based estimations of dynamic population levels are very useful in ex-ante studies following scenario-based approaches.

4.3. Data fusion

MPD, by itself, can indicate some patterns of mobility, but it is typically not detailed enough for building predictive models for displacements. Additionally, the interpretation of findings of research that employ MPD requires secondary sources for contextualization. Other data sources including other big data (e.g. social media, news sources, satellite data, and points of interest), and more traditional data sources should be integrated for ex-ante analyses. The data collected in the aftermath of the crises on damage assessment, disaster/conflict intensity, government declarations, violence-related variables, news reports etc. are essential for accurate predictions on displacements. This type of features are not reported in Table 1, because such data need to be carefully curated for each study, considering the predictors of a given specific crisis. The data fusion process is a major step for three main reasons.

Firstly, the patterns observed in big data-driven research have to be compared against official figures and other data sources for triangulation. For example, IOM collects data on displacements through its displacement tracking matrix (DTM), and the Internal Displacement Monitoring Centre (IDMC)[9]

---

[9] https://www.internal-displacement.org/database/methodology, Accessed 19 April 2024.

compiles and aggregates data on displacements by country and year. However, official data sources also have "data gaps," due to inconsistencies in key definitions and data collection methodologies across countries, lack of adequate statistics, and most notably, ignorance of new data sources such as MPD (Bircan et al. 2020). Nonetheless, when available, these sources should be used for validation and triangulation, which ensures that the patterns observed in big data are not driven by the quirks of the data, such as technical issues, business decisions, or sample biases. Official figures can be complemented with MPD or other big data sources to address missing information or data quality issues. For instance, WorldPop has worked to triangulate IOM and UNHCR estimates of displacements using satellite imagery (Dooley et al. 2021). Li et al. (2019) used DTM data collected at sites where DPs were sheltered to compare their MPD-based estimations of the displaced population in the aftermath of hurricane Matthew in Haiti. They found an 80% match between DTM- and CDR-based estimations in the origin provinces of DPs at urban sites.

Secondly, the interpretation of findings of research that employ MPD requires secondary sources for contextualization. A typical example is an anomaly detection framework, where the mobile phone data exhibits detectable anomalous patterns (Gundogdu et al. 2016). The cause of the anomaly, however, is impossible to read directly from the MPD. Global events will cause anomalies across all base stations in a country, whereas a highly localized event will cause anomalies in the base stations covering that location. Once the time and location of the anomaly are known, it may be easy to cross-reference it with event data or social media data to find probable causes. Event datasets are slowly updated, and will not allow real-time response, whereas social media responses can be very fast.

Thirdly, secondary sources can be used for feature extraction and feature selection to be used in further analysis, especially in prediction models. For instance, features extracted from remote sensing data will improve the models that predict wealth, compared to using only CDR features (Steele et al. 2017). Some social media companies publish internally aggregated indicators (such as Meta's Social Connectedness Index[10]) or allow population indicators related to advertisement targeting available. The latter was successfully used in predicting mobility in conflict situations (Minora et al. 2022, Leasure et al. 2023). The massive user base of such platforms makes these indicators detailed and useful.

5. Methods and applications

Anticipating displacements requires (1) predicting the events that are going to cause the displacements, (2) predicting the mobility of people in response to it. In both cases, the timeframe of the prediction is a key factor, affecting the difficulty of the prediction, as well as the policy response.

Predicting displacement-inducing events can be difficult depending on the type of crisis. For instance, with the current technology, it is not possible to accurately predict the timing, intensity, and location of earthquakes[11], whereas prediction of various attributes of storms (formation, path, intensity, etc.) (Alemany et al. 2019, Chen et al. 2020) and floods (water level, inundation extent, flow rates, etc.) (Mosavi et al. 2018) are possible and improving. Predicting violent conflicts that cause displacements has shown some promise (Ward et al. 2013, Hegre et al. 2017), though models tend to struggle with predicting specific instances of conflict onset, especially rare or unexpected events. A recent study found that the inclusion of a comprehensive set of variables on socioeconomic conditions, political factors, and past violence events, does not significantly improve the model predictions on new surges or intensification of violence, as the violence predictability seem to be "largely a function of time-invariant and location-specific risks" (Bazzi et al. 2019, p. 26). The second prediction task, predicting the mobility of affected populations in response to the crisis, requires historical data on mobility behavior, where features and indicators developed by MPD sources can play an important role.

---

[10] https://dataforgood.facebook.com/dfg/tools/social-connectedness-index, Accessed 19 April 2024.

[11] According to The United States Geological Survey (USGS), it is only possible to predict the probability of occurrence of a significant earthquake in a certain area within a certain time period: https://www.usgs.gov/faqs/can-you-predict-earthquakes , Accessed 19 April 2024.

A great majority of academic studies are using MPD in response to specific crises, with ex-post, rather than ex-ante methods, as the latter are difficult to verify, even with historical data. Ex-post methodologies consist of descriptive, causal, and predictive approaches. Descriptive approaches entail summarizing and describing the statistical or visual properties of the features and indicators. Causal approaches refer to a large set of quasi-experimental and probabilistic methods that use MPD to establish a causal link between crises and mobility, economic activity (Yabe, Zhang & Ukkusuri 2020, Giardini et al. 2023), segregation (Yabe et al. 2021), wealth inequality, etc.

The output of descriptive and causal analyses is an important input for building successful ex-ante models. For example, Lu et al. (2012) analyzed the movements of people in the aftermath of the Haiti earthquake in 2010 and showed that the mobility metrics after the event were highly correlated with the behavior before the event. Similar insights on the regularity of movements and persistence of important locations were used in the study of Song et al. (2016), where Markov models were trained from pre-disaster behavior and locations to forecast post-disaster behaviors in the aftermath of several disasters in Japan.

Ex-ante approaches using MPD can be categorized into two main types based on their objectives; (1) forecasting, and (2) scenario-based approaches[12]. Forecasting generally refers to short-term, immediate predictions based on real-time or very recent data about the specific event. The event itself is known to happen, but there are challenges of rapidly modeling mobility within a limited time and with limited data. Scenario-based approaches, on the other hand, explore multiple hypothetical disaster scenarios using historical MPD and additional data sources. They incorporate uncertainty in both the crisis events and mobility behavior by exploring multiple possibilities, providing a broader but less specific view of potential outcomes.

5.1. *Forecasting*

Forecasting is the process of making predictions about future events based on historical and current information. Forecasting population movements can be done reactively, in response to a crisis event, with the aim of facilitating preemptive action in affected areas and destination locations, or proactively, before a crisis event occurs. Machine learning and big data sources offer new possibilities for proactively forecasting displacements. For instance, Carammia et al. (2022) have developed a framework for forecasting asylum-related migration flows using data sources such as GDELT, Google Trends, and time series on asylum applications collected by EU countries. In order to train such a system, some ground truth historical data are required, and the chosen model complexity should be in line with the amount of data available. Hoffmann Pham & Luengo-Oroz (2023) provide an overview of data sources and models to create predictive models of forced displacement and empirically illustrate that more powerful (or flexible) models can easily overfit the data, which is a well-known problem in machine learning.

Integrating MPD sources into forecasting frameworks is difficult, as extensive mobile phone datasets are rarely available for research purposes. Many MNOs only retain CDR data for a year or less due to the enormous volume of raw data. This limitation makes it challenging to account for the inherent uncertainties of future events' proactive forecasting strategies with MPD. Therefore, forecasting displacements with MPD tends to be reactive; typically starting shortly after a crisis event hits, when displacements begin to unfold in reaction to the event.

The selection of an appropriate forecasting method depends on various factors, including the nature of the available data, the forecast's context (e.g. short-term vs. long-term), the extent of historical data, and the required level of accuracy (Hoffmann Pham & Luengo-Oroz 2023). Forecasting can be conducted at either the individual or group level. Individual-level analysis begins with data collected and processed for each person, including anonymization, and can later be aggregated to derive group-level insights while preserving individual privacy. The primary objective of decision makers to use forecasting strategies is to

---

[12] For anticipating migration, these two categories are emerging as distinct approaches. For more discussions on the issue, see: https://data4migration.org/blogseries/anticipating-migration/, Accessed 19 April 2024.

predict the number of DPs, along with their key characteristics, for specific time horizons and origin-destination pairs.

Forecasting displacements with MPD include mathematical modeling, machine learning, and simulation approaches. These approaches are not used mutually exclusively, they can be used in combination. In particular, simulation models are often mixed with machine learning approaches. We will look at each of those approaches in turn.

5.1.1. Mathematical models

Mathematical models to forecast human mobility and migration include population-level and individual-level models. At the individual level, we have 1) continuous time random walk (CTRW) models, which treat movements as discrete jumps with variable waiting times; 2) recency models, which consider recently visited locations as more likely destinations (Barbosa et al. 2015); and 3) social-based models (Dugundji & Walker 2005), which incorporate social relationships into mobility patterns (Barbosa et al. 2018). These models can also provide some underlying rules for simulations, where multiple virtual individuals are created to observe emergent population-level patterns. While not explicitly designed for crisis scenarios, these models can be adapted to anticipate mobility responses to crises. Their ability to capture individual variability and contextual factors makes them potentially valuable tools for predicting diverse displacement behaviors in the aftermath of crises, though careful calibration with crisis-specific data would be necessary for accurate forecasting.

At the population level, deterministic migration models, such as gravity, radiation, and intervening opportunities models, are often employed for forecasting mobility between pairs of locations (Barbosa et al. 2018). The traditional gravity model predicts migration flows as proportional to the population sizes of origin and destination areas and inversely proportional to the distance in between (Zipf 1946). While gravity models are straightforward to implement, they tend to oversimplify migration dynamics and require parameter fitting. Radiation models, in contrast, offer parameter-free predictions based on the population distribution around origins and destinations, often outperforming gravity models (Simini et al. 2012). Intervening opportunities models consider the number of opportunities (e.g., jobs) between an origin and destination. These population-level models can be calibrated and validated using aggregated mobile phone data, providing insights into large-scale mobility patterns (Barbosa et al. 2018). The prediction horizon for population-level mathematical models depends on the temporal frequency of the indicators used. They can potentially make short-term, as well as long-term predictions.

MPD is used mostly for calibrating population-level models with empirical observations and for validation of the results. For instance, in a study analyzing drought-related migrations in La Guajira, Colombia, Isaacman et al. (2018) used a weather-based radiation model incorporating rainfall data to predict migration patterns, and validated predictions against migration flows estimated through CDR. As their model includes drought-specific push and pull factors, it can forecast displacements more successfully than a simple radiation model. Gravity models are also used in the context of hurricanes (Anyidoho et al. 2023). Hurricanes go through multiple periods (formation, intensification, landfall, etc.) and the predictability of these periods allows for the application of forecasting methods to estimate how many people would evacuate from one district to another. Anyidoho et al. (2023) compared a traditional gravity model with a machine-learning approach for predicting hurricane evacuation flows using GPS data. They found that the machine learning approach performed better, but it is worth noting that they did not include hurricane-related push factors in their model, which could potentially have improved its performance. The importance of including crisis-specific factors in mobility models is highlighted by other studies. For example, Luca et al. (2022) show that the inclusion of pandemic-related push and pull factors could improve the estimation of international travel estimated by roaming data at the beginning of the COVID-19 pandemic. However, we note that in the aftermath of sudden-onset crises such as earthquakes, it can be difficult to use gravity models to capture crisis-related pull and push factors in a timely manner due to the sudden and large impact of some disasters. In such cases, nowcasting and mixed approaches can be considered for effective crisis response.

5.1.2. Machine learning (ML) approaches

MPD has been successfully used for predicting routine mobility such as forecasting the future locations of individuals or flows of people using traditional ML and more recent deep learning methods (Luca et al. 2021). These methods capture the regular patterns and habits in people's living through temporal and spatial dynamics of the data to make future predictions. Sequence prediction models can successfully predict the next location from hours to a day, but after that, the accuracy of such models decays significantly (Pang et al. 2020). The use cases of such approaches focus on urban settings, for optimizing and planning public transportation networks, reducing traffic congestion, and improving health outcomes. However, these approaches do not perform as well in the context of crises, as the pre-crisis data may not accurately reflect the post-crisis situation.

The literature using MPD follows reactive forecasting strategies to predict crisis-induced mobility, primarily due to the limited size of historical MPD sources. In addition, many data collaborations tend to start in response to a crisis, rather than to prepare for crises. When this happens, researchers are usually left with mobility data on a single crisis event. If these events are expected to continue in the future, or in other similar contexts, the models can give predictions for future crises. Otherwise, ML models trained in the aftermath of a single-event have limited generalizability and are hardly useful for crises in other contexts. For example, Khaefi et al. (2018) employed ML to predict evacuation destinations during the 2017 Mount Monaro eruption, based on CDR-based features. Their model has the potential to inform future mobility behavior in case the volcano activity increases in Mount Monaro, but cannot be directly used in another context without historical data. Additionally, some studies apply ML methods to gain insights into the relationship between the characteristics of DPs and their destination choices or resettlement processes (Li et al. 2019). Despite not being fully ex-ante, the feature importance analysis can help to understand the relative contribution of input variables and can still inform future models.

An example study for generalization between similar crises is the study of Anyidoho et al. (2023), which forecasts evacuation flows between metropolitan statistical areas (MSAs), as well as incoming evacuees at each destination during hurricanes. They use GPS data from three past hurricanes (namely, Florence, Michael, and Dorian) to train their models using cross-validation. The model incorporates various features including hurricane characteristics, socioeconomic factors, evacuation policies, as well as joint features inspired by the gravity equation. It generalizes well, as demonstrated on a holdout set from Hurricane Ida.

5.1.3. Mixed approaches

There are more complex approaches to forecast displacements in the aftermath of disasters using ML in combination with simulation approaches. Mixed approaches were tested for predicting the evacuation routes, simultaneously estimating specific routes people will take, and transportation modes people will use when moving between locations. Song et al. (2016) used ML in combination with a simulation approach to learn general mobility patterns in response to past disasters. Their proposed model learns to recognize how people choose and move between different places (home, work, social relationships, unknown) depending on the state of the crisis and given the restrictions of the urban environment. This approach relied on a large feature and indicator set including individual level GPS indicators from 1.6 million users from Japan, collected over 3 years, as well as earthquake intensity and levels of damage over 4 years, news report data, and transportation network data. To estimate the behavior states, they used a Hidden Markov Model (HMM), which is a statistical model that predicts a sequence of hidden states based on observable data (Rabiner 1989), and a Markov Decision Process (MDP) for modeling the decision making to move between locations. These models worked together, with the HMM accounting for individual behaviors by predicting likely actions based on personal patterns and current conditions, while the MDP predicted individual routes by modeling how people might decide to move through an area during a disaster. Thanks to extensive historical data, the proposed approach reached about 70% accuracy when predicted and actual paths of the people were compared.

5.2. Scenario-based approaches

Before a potential crisis event hits, there are many unknowns concerning the location, the timing, the scale, and the severity of the potential disruptive event. In such cases, scenario-based approaches can be very useful, as such approaches are designed to handle uncertainty. Unlike forecasting, which relies on predicting the most probable future based on the current data, scenario planning explores multiple potential futures. We classified existing scenario-based approaches into (1) agent-based simulations and (2) reinforcement learning (RL) approaches, respectively.

5.2.1. Agent-based simulations

Agent-based simulations (ABS), one of the most commonly used approaches for simulations, is based on setting up an environment in which various types of agents with different characteristics can interact between themselves and with the environment. In the context of crisis-induced mobility, ABS can be used for simulating immediate evacuation behavior, destination choices of people fleeing crisis areas, as well as longer-term dynamics of migration, such as return migration. Simulations enable feeding different parameters to the models and illustrating the effects thereof. Realistically simulating these behaviors, especially over longer time periods, is extremely challenging. The most successful applications of ABS have been related to simulating short-term evacuations in emergency situations. ABS approaches were used for predicting evacuation routes and destinations, in the aftermath of tsunamis, and earthquakes (Chen et al. 2023), as well as floods (Chapuis et al. 2019). ABS are also increasingly used in predicting forced displacements in conflict scenarios. For instance, Suleimenova et al. (2017) built an ABS-based model that predicts the destination of refugees in conflict zones by synthesizing data from multiple sources. Recently, Mehrab et al. (2024) used ABS to predict the initial refugee outflows from Ukraine in the aftermath of the Russian invasion.

There are emerging use cases of MPD with ABS. The statistical properties of MPD such as temporal and spatial distributions can be used for creating realistic simulations with ABS. Yin et al. (2020) constructed an ABS-based evacuation simulation that is in communication with real-time population distributions from MPD, as well as with a knowledge database that stores typical population distributions in order to create the most realistic scenarios in the face of a crisis. Similarly, Sudo et al. (2016) developed a method that combines ABS with real-time mobile phone data to estimate population movements following a major disaster. Their model consistently outperformed standard simulation approaches that did not use MPD.

Some of the mixed approaches that are used for forecasting crisis-induced mobility explained in Section 5.1.3 can be also used in scenario-based approaches. For instance, the approach developed by Song et al. (2016) can be used to simulate human movements under hypothetical scenarios if the models are fed with synthetic data (earthquake intensity, damage data, etc.). Another potentially useful tool to guide simulations is the dynamic population density map (Deville et al. 2014), which uses MPD to predict population distributions across a country at specific time periods. When used in combination with hazard maps (i.e. maps showing risk levels of potential hazards in areas), these can inform models on the extent of the exposure (Balistrocchi et al. 2020).

5.2.2. Reinforcement learning (RL) approaches

There is an emerging branch of literature that combines RL[13] with agent-based simulations to predict human mobility after crises. These studies are characterized by adaptive agents that learn from data and interact within complex environments (Pang et al. 2020, Fan et al. 2021). The main advantages of these approaches include their ability to learn from historical data, adapt to new scenarios, and potentially capture more realistic human behavior in crisis situations. However, if the state and action spaces are too

---

[13] Reinforcement learning refers to a ML technique in which agents learn to make decisions as a result of performing sequences of actions and receiving rewards or penalties in return (Sutton and Barto 2018).

big, RL approaches require very large datasets to properly learn models. Subsequently, proper quantization of these spaces becomes important, sometimes forcing coarser granularity of analysis.

In an example study, Fan et al. (2021) predict the destinations and trajectories of people affected by crises. Their model combines simulation with RL by first predicting destinations using historical mobility data, then determining optimal routes using a reward-based system that can adapt to crisis conditions like flooding. As their model was trained on real mobility data from normal conditions, their simulated agents made realistic decisions even when faced with crisis scenarios not present in the original training data. Another example is the model developed by Pang et al. (2020) which used inverse RL to extract human mobility preferences from observed trajectories in a source city, applying these learned preferences to simulate population movements in a target city. This approach enables the generation of realistic daily movement patterns and rare event scenarios in cities lacking historical data (Pang et al. 2020).

6. Opportunities and challenges

Ex-ante approaches with MPD are still emerging and their use-cases are largely at the research level at the moment. In this section, we briefly point out some future directions that are promising for MPD usage, as well as challenges that require solutions to increase the effectiveness of MPD for policy makers.

6.1. Opportunities

6.1.1. Increasing data for social good initiatives

There is a growing availability of data products related to the estimation of population displacements in a timely manner (Yabe et al. 2022). In recent years, data for social good initiatives of big tech companies such as Meta and Google have started to share aggregated and anonymized mobility indicators on an ongoing basis. For instance, Meta regularly publishes population counts and crisis-related movements based on Facebook user population for major disasters across the globe that are available to decision-makers and researchers (Maas et al. 2019). Google regularly shared community mobility reports during the COVID-19 pandemic that gave insights on the effectiveness of lockdown measurements (Aktay et al. 2020).

There are also efforts to share aggregated pre-computed CDR-based indicators that can potentially inform conflict and disaster-induced displacements. For instance, Flowminder calculates various CDR-based stock and flow indicators in Haiti, a country that experiences regular violent clashes, as well as large-scale natural disasters, such as hurricanes and earthquakes. These indicators are regularly shared through a data mobility platform with researchers and decision makers[14]. Creating more incentives for data collaboratives and nurturing data stewards that can guide the process in companies with access to valuable data are the keys to generating more initiatives (Verhulst & Young 2019).

6.1.2. Open source platforms

Mobility indicators, when fused with other data sources, can provide the necessary input for applying ex-ante methods. There are open-source platforms such as OpenStreetMap[15], demographic data sets built at high spatial granularity by WorldPop[16], as well as publicly available satellite data sources such as Copernicus[17], which is a program of the European Union (EU). In addition to such resources, publicly available libraries and tools such as Scikit-mobility (Pappalardo et al. 2022), BigGIS-RTX (Lwin et al. 2018) and Bandicoot (De Montjoye et al. 2016) can be used for processing mobility data for calculating various indicators and features. Data collected by organizations such as IOM and UNHCR can be complemented with openly available MPD and big data sources, using these processing tools.

---

[14] https://haiti.mobility-dashboard.org/, Accessed 19 April 2024.
[15] https://www.openstreetmap.org/, Accessed 19 April 2024.
[16] https://www.worldpop.org, Accessed 19 April 2024.
[17] https://www.copernicus.eu/en, Accessed 19 April 2024.

Increasing the availability of data sources and processing tools allows researchers with a broader expertise to contribute to MPD analysis. Several data challenges organized by MNOs, such as data for development (Blondel et al. 2012, 2015), Telecom Italia (Barlacchi et al. 2015), and data for refugees (Salah et al. 2018, Salah, Pentland, Lepri, Letouzé et al. 2019) have opened large CDR datasets to research groups, and enabled multifaceted analyses of data from a specific country, and led to policy recommendations in areas like education, health, security, and unemployment (Salah, Altuncu, Balcisoy, Frydenlund et al. 2019).

6.1.3. Novel MPD sources

Apart from CDR and GPS data, large-scale mobile phone data sources include Extended Detail Records (xDR), Inbound Roaming (IR), Outbound Roaming (OR), and Airtime Top-up Transfers (ATT). Most of these are collected by MNOs. xDR marks data exchanges between devices and the network. The signals in xDR tend to be more frequent than CDR, which is valuable for many mobility and migration indicators. CDR and xDR are highly useful MPD sources for understanding the internal mobility of users. IR and OR data offer insights into user movements between different network operators, both domestically and internationally, with variables recording elements such as timestamps, used antennae, and the country of origin or visitation for the user. These sources miss the social connections between users, which is a unique aspect of the CDR.

There are some previous studies that demonstrated the use of roaming data in the context of cross-border commuting (Järv et al. 2021), and tourism (Ahas et al. 2007), yet to our knowledge, these have not been used in crisis-induced displacement research. ATT data documents the transfer of phone credits between MNOs, and can be collected by MNOs or by companies that facilitate such transfers, but they are scarcely utilized in migration and mobility research (Aydoğdu et al. 2022). Aydoğdu et al. (2023) showed that ATT has a very high correlation with migrant stocks and is a strong signal on the presence and origin of migrants in an area. Blumenstock et al. (2016) demonstrated that natural disasters induce airtime transfers to the affected regions. Consequently, ATT data can provide predictive insights about the post-disaster resilience of migrant communities.

6.2. Challenges

6.2.1. Data access

Data access is a highly challenging issue in the context of MPD, due to ethical, legal, and technical complexity of the process. Timely and cost-effective response using MPD requires a data collaborative in place before the crisis (Oliver et al. 2020). For example, the World Bank initiative that aimed to use CDR indicators in response to the COVID-19 outbreak faced various challenges while trying to access CDR, including lengthy processes, regulatory restrictions, insufficient capacity of data users (governments in their case), high costs, funding gaps, and the need for coordination across various actors (Milusheva et al. 2021). These challenges are expected, as MNOs employ robust internal security measures to protect data storage and access. Data sharing even between units within these companies must navigate complex internal policies, often involving conflicting interests among stakeholders, so the process of acquiring MPD is lengthy.

Conversely, policy-making institutions frequently lack personnel with the skills necessary for managing complex MPD partnerships, which complicates the process of integrating MPD sources into existing systems. Additionally, the operational processes and objectives of policy-making bodies and MNOs might not be mutually transparent. Consequently, successful projects demand commitment from both parties and good communication.

Data collaboratives can potentially solve some of the above-mentioned challenges and help to create value in terms of enhancing the analysis of displaced persons, improving predictions, and refining the impact assessment of interventions (Verhulst & Young 2019). An example is the Hummingbird EU project[18], which initiated a data collaborative with the largest MNO in Turkey, with the goal of using MPD

---

[18] https://hummingbird-h2020.eu/, Accessed 19 April 2024.

sources to develop indicators of migration and conflict-driven refugees, eventually to be used by decision-makers (Aydoğdu et al. 2022). Ongoing challenges in Turkey arise from hosting a large population of refugees (approx. 4 million, mainly from Syria and Afghanistan), and large-scale natural disasters (such as the 2023 Türkiye-Syria earthquake that killed over 50 K people and demolished about 1 million residences) necessitate the use of alternative sources for effective preparedness and prevention policies.

6.2.2. Data governance

MPD sources have many advantages for policy making in the area of migration and mobility, but they come with some risks. It is crucial to handle the data in a way that preserves the privacy of individuals and groups. In the EU, personal data processing is regulated by the General Data Protection Regulation (GDPR), but many MPD partnerships occur in developing countries, where similar legal frameworks are absent. Furthermore, handling large-scale mobile data necessitates consideration of group privacy, an area not fully covered by the GDPR.

Besides legal issues, ethical concerns about potential harm to individuals and groups are important. Ethical standards and established procedures from earlier initiatives can provide some guidance in addressing these concerns (Blondel et al. 2012, Vinck et al. 2019, Salah et al. 2022). Additionally, the OCHA Centre for Humanitarian Data's Data Responsibility Guidelines address the ethical considerations of using the data of affected people in humanitarian contexts (OCHA 2021, p. 8). They define data responsibility as "the safe, ethical and effective management of personal and non-personal data for operational response, in accordance with established frameworks for personal data protection," and provide actionable recommendations for a range of issues including diagnosis of risks, creating and maintaining documentation and procedures for oversight, and data management best practices.

7. Moving forward: MPD for anticipatory policy making

In this paper, we introduced a unified approach to processing various MPD sources for anticipating displacements. Existing studies in the MPD literature largely focus on disaster-induced displacements and follow ex-post descriptive, rather than ex-ante predictive methodologies. We focused on the latter in this work. Based on the literature and our own experience with various MPD sources, we reviewed the methodological approaches for anticipating displacements.

Currently, a great majority of academic studies that use MPD to estimate population displacements are conducted with CDR and GPS, and data access remains a hurdle for wider integration of MPD sources. The analyzed work clearly demonstrates that data fusion with complementary data sources, including in-house data at policy institutions, is paramount for building successful models.

Forecasting displacements is an extremely challenging task due to the complexity and diversity of characteristics and causes of displacements, as well as the lack of trusted gold standard data (Franklinos et al. 2021). For future endeavors to use MPD in anticipatory policy making, we recommend following simulation approaches, fueled by detailed spatio-temporal data that MPD sources can provide, as a way of allowing detailed scenario-based analysis, which can complement statistical and machine learning based predictive models that are useful for providing statistically aggregated indicators of human behavior.

While the prediction methods discussed in this work have been tested in different crisis events, their accuracy in practice is difficult to estimate. Machine learning-based prediction should ideally be tested with independent data, not seen during model and parameter selection. We have reported studies using leave-one-crisis-out cross-validation, but even this approach is not fully resistant to producing optimistic results. A standard approach in the machine learning community is to introduce public data challenges with well-defined predictive goals and sequestered testing data (i.e. not disclosed to participants) to provide completely independent testing of models. This should be considered as a future direction in this domain.

Finally, data collaboratives can help policy institutions to enhance their capacity to integrate MPD-based indicators into their existing data sources and to develop tools that can support future-looking

decision-making in the context of displacements. Having sustainable collaboratives with legal and technical preliminaries handled before a crisis strikes is a big asset for policy makers, as MPD can provide near real-time data for crisis managements under such conditions.

References

Abel GJ, **Brottrager M**, **Cuaresma JC and Muttarak R** (2019), Climate, conflict and forced migration. *Global Environmental Change 54*, 239–249.

Adhikari P (2013), Conflict-induced displacement, understanding the causes of flight. *American Journal of Political Science 57*(1), 82–89.

Ahas R, **Aasa A**, **Mark Ü**, **Pae T and Kull A** (2007) Seasonal tourism spaces in Estonia: Case study with mobile positioning data. *Tourism M 28*(3), 898–910.

Aktay A, **Bavadekar S**, **Cossoul G**, **Davis J**, **Desfontaines D**, **Fabrikant A**, **Gabrilovich E**, **Gadepalli K**, **Gipson B**, **Guevara M** *et al.* (2020) Google covid-19 community mobility reports: anonymization process description (version 1.1), *arXiv preprint arXiv:2004.04145*.

Alemany S, **Beltran J**, **Perez A and Ganzfried S** (2019) Predicting hurricane trajectories using a recurrent neural network, *in* The Thirty-Third AAAI Conference on Artificial Intelligence (AAAI-19). Association for the Advancement of Artificial Intelligence, Honolulu, Hawaii, pp. 468–474.

Anyidoho PK, **Ju X**, **Davidson RA and Nozick LK** (2023) A machine learning approach for predicting hurricane evacuee destination location using smartphone location data. *Computational Urban Science 3*(1), 30.

Aydoğdu B, **Momeni R**, **Subhi G**, **Balcik C**, **Bircan T and Salah AA** (2022) Hummingbird interim report on migration and mobility indicators developed with mobile phone data, Technical report, HumMingBird Project Deliverable D6.3, Leuven: HumMingBird project 870661 – H2020. **URL:** https://hummingbird-h2020.eu/publications

Aydoğdu B, **Samad H**, **Bai S**, **Abboud S**, **Gorantis I and Salah AA** (2023) Analysing international airtime top-up transfers for migration and mobility. *International Journal of Data Science and Analytics*. **URL:** https://doi.org/10.1007/s41060-023-00396-7

Balistrocchi M, **Metulini R**, **Carpita M and Ranzi R** (2020) Dynamic maps of human exposure to floods based on mobile phone data. *Natural Hazards and Earth System Sciences 20*, 3485–3500. **URL:** https://doi.org/10.5194/nhess-20-3485-2020

Barbosa H, **Barthelemy M**, **Ghoshal G**, **James CR**, **Lenormand M**, **Louail T**, **Menezes R**, **Ramasco JJ**, **Simini F and Tomasini M** (2018) Human mobility: Models and applications. *Physics Reports 734*, 1–74. **URL:** https://doi.org/10.1016/j.physrep.2018.01.001

Barbosa H, **de Lima-Neto FB**, **Evsukoff A and Menezes R** (2015) The effect of recency to human mobility. *EPJ Data Science 4*, 1–14.

Barlacchi G, **De Nadai M**, **Larcher R**, **Casella A**, **Chitic C**, **Torrisi G**, **Antonelli F**, **Vespignani A**, **Pentland A and Lepri B** (2015) A multi-source dataset of urban life in the city of Milan and the Province of Trentino. *Scientific Data 2*(1), 1–15.

Bazzi S, **Blair RA**, **Blattman C**, **Dube O**, **Gudgeon M and Peck, RM** (2019) The promise and pitfalls of conflict prediction: Evidence from Colombia and Indonesia, NBER Working Paper Series 25980, National Bureau of Economic Research. **URL:** http://www.nber.org/papers/w25980

Bengtsson L, **Lu X**, **Thorson A**, **Garfield R and von Schreeb J** (2011) Improved response to disasters and outbreaks by tracking population movements with mobile phone network data: A post-earthquake geospatial study in Haiti. *PLoS Medicine 8*(8), 1–9.

Berchin II, **Valduga IB**, **Garcia J and de Andrade Guerra JBS** (2017) Climate change and forced migrations: An effort towards recognizing climate refugees. *Geoforum 84*, 147–150.

Bircan T, **Purkayastha D**, **Ahmad-Yar AW**, **Lotter K**, **Dello Iakono C**, **Göler D**, **Stanek M**, **Yilmaz S**, **Solano G and Ünver Ö** (2020) Gaps in migration research. Review of migration theories and the quality and compatibility of migration data on the national and international level, Technical report, HumMingBird Project Deliverable D2.1, Leuven: HumMingBird project 870661 – H2020. **URL:** https://hummingbird-h2020.eu/publications

Blondel VD, **Decuyper A and Krings G** (2015) A survey of results on mobile phone datasets analysis. *EPJ Data Science 4*(1), 1–55.

Blondel VD, **Esch M**, **Chan C**, **Clérot F**, **Deville P**, **Huens E**, **Morlot F**, **Smoreda Z and Ziemlicki C** (2012) Data for development: the D4D challenge on mobile phone data. *arXiv preprint* arXiv:1210.0137.

Blumenstock JE, **Eagle N and Fafchamps M** (2016) Airtime transfers and mobile communications: Evidence in the aftermath of natural disasters. *Journal of Development Economics 120*, 157–181.

Bradley M and Cohen R (2010) Disasters and displacement: Gaps in protection. *Journal of International Humanitarian Legal Studies 1*(1), 95–142.

Cantor DJ (2024) Divergent dynamics: disasters and conflicts as drivers of internal displacement?. *Disasters 48*(1), e12589.

Carammia M, **Iacus SM and Wilkin T** (2022) Forecasting asylum-related migration flows with machine learning and data at scale. *Scientific Reports 12*(1), 1457.

Chapuis K, **Elwaqoudi TA**, **Brugière A**, **Daudé E**, **Drogoul A**, **Gaudou B**, **Nguyen-Ngoc D**, **Nghi HQ and Zucker J-D** (2019), An agent-based co-modeling approach to simulate the evacuation of a population in the context of a realistic flooding event: a case study in Hanoi (Vietnam), in Mohd, M.H., Misro, M.Y., Ahmad, S., Nguyen Ngoc, D., (eds) *Modelling, Simulation and Applications of Complex Systems. CoSMoS 2019. Springer Proceedings in Mathematics & Statistics*, vol *359*. Springer, Singapore. https://doi.org/10.1007/978-981-16-2629-6_5 , pp. 79–108.

Chen C, **Koll C**, **Wang H and Lindell M K** (2023) An interdisciplinary agent-based evacuation model: integrating the natural environment, built environment, and social system for community preparedness and resilience. *Natural Hazards and Earth System Sciences 23*(2), 733–749.

Chen R, **Zhang W and Wang X** (2020) Machine learning in tropical cyclone forecast modeling: A review. *Atmosphere 11*(7), 676.

De Montjoye Y-A, **Gambs S**, **Blondel V**, **Canright G**, **De Cordes N**, **Deletaille S**, **Engø-Monsen K**, **Garcia-Herranz M**, **Kendall J**, **Kerry C** *et al* (2018) On the privacy-conscientious use of mobile phone data. *Scientific Data 5*(1), 1–6.

De Montjoye Y-A, **Rocher L and Pentland AS** (2016) Bandicoot: A python toolbox for mobile phone metadata. *Journal of Machine Learning Research 17*, 1–5.

Deville P, **Linard C**, **Martin S**, **Gilbert M**, **Stevens FR**, **Gaughan AE**, **Blondel VD and Tatem AJ** (2014) Dynamic population mapping using mobile phone data. *Proceedings of the National Academy of Sciences 111*(45), 15888–15893.

Dooley C, **Jochem W**, **Leasure D**, **Sorichetta A**, **Lazar A**, **Tatem A and Bondarenko M** (2021) South Sudan 2020 gridded population estimates from census projections adjusted for displacement, version 20. *WorldPop, University of Southampton.* **URL:** https://doi.org/10.5258/SOTON/WP00709 *[Dataset]*

Dugundji ER and Walker JL (2005) Discrete choice with social and spatial network interdependencies: an empirical example using mixed generalized extreme value models with field and panel effects. *Transportation Research Record 1921*(1), 70–78.

Eagle N and Pentland A (2006) Reality mining: sensing complex social systems. *Personal and Ubiquitous Computing 10*, 255–268.

EGRIS (2018), Technical report on statistics of internally displaced persons: Current practice and recommendations for improvement, Technical report, United Nations Statistical Commission. Expert Group on Refugee and IDP Statistics (EGRIS). Report submitted to the 49th session of the UN Statistical Commission. **URL:** https://egrisstats.org/resource/technical-report-on-statistics-of-internally-displaced-persons/

Fan C, **Jiang X and Mostafavi A** (2021) Evaluating crisis perturbations on urban mobility using adaptive reinforcement learning. *Sustainable Cities and Society 75*, 103367.

Ferris E (2007) *Making sense of climate change, natural disasters, and displacement: a work in progress*. Bern Universität: Calcutta Research Group Winter Course.

Franklinos LH, **Parrish R**, **Burns R**, **Caflisch A**, **Mallick B**, **Rahman T**, **Routsis V**, **López AS**, **Tatem AJ and Trigwell R** (2021)] Key opportunities and challenges for the use of big data in migration research and policy. *UCL Open Environment 3*.

Ghahramani M, **Zhou M and Wang G** (2020) Urban sensing based on mobile phone data: Approaches, applications, and challenges. *IEEE/CAA Journal of Automatica Sinica 7*(3), 627–637.

Giardini F, **Hadjidimitriou NS**, **Mamei M**, **Bastardi G**, **Codeluppi N and Pancotto, F** (2023) Using mobile phone data to map evacuation and displacement: a case study of the central italy earthquake. *Scientific reports 13*(1), 22228.

Gonzalez MC, **Hidalgo CA and Barabasi A-L** (2008) Understanding individual human mobility patterns. *Nature 453*(7196), 779–782.

Goodwin-Gill GS and McAdam J (2017) Unhcr and climate change, disasters and displacement. *The United Nations Refugee Agency (UNHCR)*, Geneva, Switzerland.

Guan V, **Guépin F**, **Cretu A-M and De Montjoye Y-A** (2024) A zero auxiliary knowledge membership inference attack on aggregate location data. *arXiv preprint arXiv:2406.18671*

Gundogdu D, **Incel OD**, **Salah AA and Lepri B** (2016) Countrywide arrhythmia: emergency event detection using mobile phone data. *EPJ Data Science 5*(1), 1–19 **URL:** https://doi.org/10.1140/epjds/s13688-016-0086-0

Hegre H, **Metternich NW**, **Nygård HM and Wucherpfennig J** (2017) Introduction: Forecasting in peace research. *Journal of Peace Research 54*(2), 113–124.

Hoffmann Pham K and Luengo-Oroz M (2023) Predictive modelling of movements of refugees and internally displaced people: towards a computational framework. *Journal of Ethnic and Migration Studies 49*(2), 408–444.

IDMC (2024) *Global Report on Internal Displacement 2024*, Internal Displacement Monitoring Centre, Geneva, Switzerland. **URL:** https://www.internal-displacement.org/global-report/grid2024

Isaacman S, **Frias-Martinez V and Frias-Martinez E** (2018) Modeling human migration patterns during drought conditions in La Guajira, Colombia. *Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies, COMPASS 2018.*

Järv O, **Tominga A**, **Müürisepp K and Silm S** (2021) The impact of COVID-19 on daily lives of transnational people based on smartphone data: Estonians in Finland. *Journal of Location Based Services* 15(3), 169–197.

Khaefi M R, **Prahara P J**, **Rheza M**, **Alkarisya D and Hodge G** (2018) Predicting evacuation destinations due to a natural hazard using mobile network data, *in 2018 2nd International Conference on Informatics and Computational Sciences (ICICoS)*. IEEE, pp. 1–6.

Kohli N, **Aiken E and Blumenstock JE** (2024) Privacy guarantees for personal mobility data in humanitarian response. *Scientific Reports* 14(1), 28565.

Leasure DR, **Kashyap R**, **Rampazzo F**, **Dooley CA**, **Elbers B**, **Bondarenko M**, **Verhagen M**, **Frey A**, **Yan J**, **Akimova ET** *et al* (2023) Nowcasting daily population displacement in Ukraine through social media advertising data. *Population and Development Review* 49(2), 231–254.

Lepri B, **Oliver N**, **Letouzé E**, **Pentland A and Vinck P** (2018) Fair, transparent, and accountable algorithmic decision-making processes: The premise, the proposed solutions, and the open challenges. *Philosophy and Technology* 31, 611–627.

Li T, **Dejby J**, **Albert M**, **Bengtsson L and Lefebvre V** (2019) Contributing to a better understanding of human mobility in crisis, Technical report, Flowminder Foundation.

Lind A, **Hadachi A and Batrashev O** (2017) A new approach for mobile positioning using the CDR data of cellular networks, *in 2017 5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*. IEEE, pp. 315–320.

Lu X, **Bengtsson L and Holme P** (2012) Predictability of population displacement after the 2010 Haiti earthquake. *Proceedings of the National Academy of Sciences of the United States of America* 109(29), 11576–11581.

Lu X, **Wrathall DJ**, **Sundsøy PR**, **Nadiruzzaman M**, **Wetter E**, **Iqbal A**, **Qureshi T**, **Tatem A J**, **Canright GS**, **Engø-Monsen K and Bengtsson L** (2016) Detecting climate adaptation with mobile network data in Bangladesh: anomalies in communication, mobility and consumption patterns during cyclone Mahasen. *Climatic Change* 138(3–4), 505–519. **URL:** https://doi.org/10.1007/s10584-016-1753-7

Luca M, **Barlacchi G**, **Lepri B and Pappalardo L** (2021) A survey on deep learning for human mobility. *ACM Computing Surveys (CSUR)* 55(1), 1–44.

Luca M, **Lepri B**, **Frias-Martinez E and Lutu A** (2022) Modeling international mobility using roaming cell phone traces during COVID-19 pandemic. *EPJ Data Science* 11(1), 1–17.

Lwin KK, **Sekimoto Y and Takeuchi W** (2018) Development of GIS integrated big data research toolbox (BigGIS-RTX) for Mobile CDR data processing in disasters management. *Journal of Disaster Research* 13(2), 380–386.

Maas P, **Iyer S**, **Gros A**, **Park W**, **McGorman L**, **Nayak C and Dow PA** (2019) Facebook disaster maps: Aggregate insights for crisis response and recovery, *in Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Vol. *19*, p. 3173.

Marshall JM, **Touré M**, **Ouédraogo AL**, **Ndhlovu M**, **Kiware SS**, **Rezai A**, **Nkhama E**, **Griffin JT**, **Hollingsworth TD**, **Doumbia S** *et al.* (2016) Key traveller groups of relevance to spatial malaria transmission: a survey of movement patterns in four sub-Saharan African countries. *Malaria Journal* 15, 1–12.

Marzuoli A and Liu F (2018) A data-driven impact evaluation of Hurricane Harvey from mobile phone data. *IEEE Int Conf on Big Data* pp. 3442–3451.

McAdam J (2012) Creating new norms on climate change, natural disasters and displacement: international developments 2010-2013. *Refuge* 29, 11.

Mehrab Z, **Stundal L**, **Venkatramanan S**, **Swarup S**, **Lewis B**, **Mortveit HS**, **Barrett CL**, **Pandey A**, **Wells CR**, **Galvani AP** *et al* (2024) An agent-based framework to study forced migration: A case study of ukraine. *PNAS nexus* 3(3), pgae080.

Miller SD (2020) Conflict-induced displacement and development, *in Routledge Handbook of Migration and Development*. Routledge, pp. 363–372.

Milusheva S, **Lewin A**, **Gomez TB**, **Matekenya D and Reid K** (2021) Challenges and opportunities in accessing mobile phone data for covid-19 response in developing countries. *Data and Policy* 3, e20.

Minora U, **Bosco C**, **Iacus SM**, **Grubanov-Boskovic S**, **Sermi F and Spyratos S** (2022) The potential of Facebook advertising data for understanding flows of people from Ukraine to the European Union. *EPJ Data Science* 11(1), 59.

Mitchell SM and Pizzi E (2021) Natural disasters, forced migration, and conflict: The importance of government policy responses. *International Studies Review* 23(3), 580–604.

Morales-Muñoz H, **Jha S**, **Bonatti M**, **Alff H**, **Kurtenbach S and Sieber S** (2020) Exploring connections—environmental change, food security and violence as drivers of migration—a critical review of research. *Sustainability* 12(14), 5702.

Mosavi A, **Ozturk P and Chau K-w** (2018) Flood prediction using machine learning models: Literature review. *Water* 10(11), 1536.

OCHA (2021) *OCHA Data Responsibility Guidelines*, United Nations Office for the Coordination of Humanitarian Affairs Centre for Humanitarian Data, The Hague, The Netherlands **URL:** https://centre.humdata.org/data-responsibility/

Oliver N, **Lepri B**, **Sterly H**, **Lambiotte R**, **Deletaille S**, **De Nadai M**, **Letouzé E**, **Salah A A**, **Benjamins R**, **Cattuto C** *et al* (2020) Mobile phone data for informing public health actions across the COVID-19 pandemic life cycle. *Science Advances* 6(23), eabc0764.

Pang Y, **Tsubouchi K**, **Yabe T and Sekimoto Y** (2020) Intercity simulation of human mobility at rare events via reinforcement learning, *in* Proceedings of the 28th International Conference on Advances in Geographic Information Systems. pp. 293–302.

Pappalardo L, **Ferres L**, **Sacasa M**, **Cattuto C and Bravo L** (2021) Evaluation of home detection algorithms on mobile phone data using individual-level ground truth. *EPJ Data Science 10*(1), 29.

Pappalardo L, **Simini F**, **Barlacchi G and Pellungrini R** (2022) scikit-mobility: A Python library for the analysis, generation, and risk assessment of mobility data. *Journal of Statistical Software 103*, 1–38.

Power D, **Thom M**, **Gray J**, **Albert M**, **Delaporte S**, **Li T**, **Harrison J**, **Greenhalgh J**, **Thorne N and Bengtsson L** (2019) *FlowKit : Unlocking the power of mobile data for humanitarian and development purposes*, Flowminder Foundation DIAL Whitepaper. **URL:** https://www.flowminder.org/resources/publications-reports/flowkit-unlocking-the-power-of-mobile-data-for-humanitarian-and-development-purposes

Rabiner LR (1989) A tutorial on hidden Markov models and selected applications in speech recognition *Proceedings of the IEEE 77*(2) 257–286.

Salah AA, **Altuncu MT**, **Balcisoy S**, **Frydenlund E** *et al* (2019) Policy implications of the D4R challenge, *in* AA Salah, A Pentland, B Lepri, E Letouzé, Y-A De Montjoye, X Dong and P Vinck (eds), *Guide to mobile data analytics in refugee scenarios*. Springer Cham, pp. 477–495.

Salah AA, **Canca C and Erman B** (2022) Ethical and legal concerns on data science for large scale human mobility, *in* AA Salah, E E Korkmaz and T Bircan (eds), *Data Science for Migration and Mobility*. Proceedings of the British Academy, British Academy / Oxford University Press, London, UK.

Salah AA, **Pentland A**, **Lepri B**, **Letouzé E**, **Vinck P**, **de Montjoye Y-A**, **Dong X and Dagdelen, O** (2019) Guide to Mobile Data Analytics in Refugee Scenarios: The 'Data for Refugees Challenge' study, Springer Cham.

Salah A, **Pentland A**, **Lepri B**, **Letouzé E** *et al* (2018) *Guide to Mobile Data Analytics in Refugee Scenarios: The Data for Refugees Challenge study*, Springer Cham.

Sekara V, **Omodei E**, **Healy L**, **Beise J**, **Hansen C**, **You D**, **Blume S and Garcia-Herranz M** (2019) Mobile phone data for children on the move: Challenges and opportunities, *in* AA Salah, A Pentland, B Lepri, E Letouzé, Y-A De Montjoye, X Dong and P Vinck (eds), *Guide to Mobile Data Analytics in Refugee Scenarios*. Springer, Cham pp. 53–66.

Shibuya Y, **Jones N and Sekimoto Y** (2024) Assessing internal displacement patterns in Ukraine during the beginning of the Russian invasion in 2022. *Scientific Reports 14*(1), 11123.

Simini F, **González MC**, **Maritan A and Barabási A-L** (2012) A universal model for mobility and migration patterns. *Nature 484* (7392), 96–100.

Song C, **Qu Z**, **Blumm N and Barabási A-L** (2010) Limits of predictability in human mobility. *Science 327*(5968), 1018–1021.

Song X, **Zhang Q**, **Sekimoto Y**, **Shibasaki R**, **Yuan N J and Xie X** (2016) Prediction and simulation of human mobility following natural disasters. *ACM Transactions on Intelligent Systems and Technology (TIST) 8*(2), 1–23.

Steele A (2019) Civilian resettlement patterns in civil war. *Journal of Peace Research 56*(1), 28–41

Steele JE, **Sundsøy PR**, **Pezzulo C**, **Alegana VA**, **Bird TJ**, **Blumenstock J**, **Bjelland J**, **Engø-Monsen K**, **De Montjoye Y-A**, **Iqbal A M** *et al* (2017) Mapping poverty using mobile phone and satellite data. *Journal of The Royal Society Interface 14*(127), 20160690.

Sudo A, **Kashiyama T**, **Yabe T**, **Kanasugi H and Sekimoto Y** (2016) Human mobility estimation following massive disaster using filtering approach. *Journal of Disaster Research 11*(2), 217–224.

Suleimenova D, **Bell D and Groen D** (2017) A generalized simulation development approach for predicting refugee destinations. *Scientific reports 7*(1), 13377.

Sutton RS (2018) *Reinforcement learning: An introduction*, 2nd ed., the MIT Press: Cambridge, MA.

Tatem AJ, **Dooley CA**, **Lai S**, **Woods D**, **Cunningham A and Sorichetta A** (2020) Geospatial data integration to capture small-area population dynamics. *Harnessing Data Innovation for Migration Policy: A Handbook for Practitioners* pp. 10–24.

Thalheimer L, **Simperingham E and Jjemba EW** (2022) The role of anticipatory humanitarian action to reduce disaster displacement. *Environmental Research Letters 17*(1), 014043.

UNHCR (1951) *Convention and Protocol Relating to the Status of Refugees*, United Nations High Commissioner for Refugees, Geneva. **URL:** https://www.unhcr.org/media/convention-and-protocol-relating-status-refugees

UNHCR (2023) Global trends report 2023. United Nations High Commissioner for Refugees, Geneva. **URL:** https://www.unhcr.org/global-trends-report-2023

Verhulst SG and Young A (2019) The potential and practice of data collaboratives for migration, *in* AA Salah, A Pentland, B Lepri, E Letouzé, Y-A De Montjoye, X Dong and P Vinck (eds) *Guide to Mobile Data Analytics in Refugee Scenarios*. Cham: Springer, pp. 465–476.

Vespe M, **Iacus SM**, **Santamaria C**, **Sermi F and Spyratos S** (2021) On the use of data from multiple mobile network operators in Europe to fight COVID-19. *Data and Policy 3*, e8.

Vinck P, **Pham PN and Salah, AA** (2019) "Do no harm" in the age of Big Data: Data, ethics, and the refugees, *in* A A Salah, A Pentland, B Lepri, E Letouzé, Y-A De Montjoye, X Dong and P Vinck (eds), *Guide to Mobile Data Analytics in Refugee Scenarios*. Cham: Springer, pp. 87–99.

Wang Q and Taylor JE (2016) Patterns and limitations of urban human mobility resilience under the influence of multiple types of natural disaster. *PLoS one 11*(1), e0147299.

Ward MD, **Metternich NW**, **Dorff CL Gallop M**, **Hollenbach FM**, **Schultz A and Weschle S** (2013) Learning from the past and stepping into the future: Toward a new generation of conflict prediction. *International Studies Review 15*(4), 473–490.

Yabe T, **Jones NK**, **Lozano-Gracia N**, **Khan MF**, **Ukkusuri SV**, **Fraiberger S and Montfort A** (2021) Location Data Reveals Disproportionate Disaster Impact Amongst the Poor: A Case Study of the 2017 Puebla Earthquake Using Mobilkit. *arXiv preprint arXiv:2107.13590.*

Yabe T, **Jones NK**, **Rao PSC**, **Gonzalez MC and Ukkusuri SV** (2022) Mobile phone location data for disasters: A review from natural hazards and epidemics. *Computers, Environment and Urban Systems 94*, 1–18.

Yabe T, **Tsubouchi K**, **Fujiwara N**, **Sekimoto Y and Ukkusuri SV** (2020) Understanding post-disaster population recovery patterns. *Journal of the Royal Society Interface 17*(163), 20190532.

Yabe T, **Zhang Y and Ukkusuri SV** (2020) Quantifying the economic impact of disasters on businesses using human mobility data: a Bayesian causal inference approach. *EPJ Data Science 9*(1), 36.

Yin L, **Chen J**, **Zhang H**, **Yang Z**, **Wan Q**, **Ning L**, **Hu J and Yu Q** (2020) Improving emergency evacuation planning with mobile phone location data. *Environment and Planning B: Urban Analytics and City Science 47*(6), 964–980.

Zhao Z, **Shaw S-L**, **Xu Y**, **Lu F**, **Chen J and Yin, L** (2016) Understanding the bias of call detail records in human mobility research. *International Journal of Geographical Information Science 30*(9), 1738–1762.

Zipf G =K (1946) The P 1 P 2/D hypothesis: on the intercity movement of persons. *American Sociological Review 11*(6), 677–686.