AROMA_CoDa: Assessing Refugees' Onward Mobility through the Analysis of Communication Data

Harald Sterly[1] [0000−0001−8819−1638], Benjamin Etzold[2][0000−0002−1109−7640], Lars Wirkus[2][0000−0002−9795−3187], Patrick Sakdapolrak[3][0000−0001−7137−1552], Jacob Schewe[4][0000−0001−9455−4159], Carl-Friedrich Schleussner[5][0000−0001−8471−848X], Benjamin Hennig[6][0000−0002−5754−2455]

[1] University of Bonn, 53115 Bonn, Germany, sterly@giub.uni-bonn.de
[2] Bonn International Center for Conversion, Pfarrer- Byns-Straße 1, 53121 Bonn, Germany
[3] University of Vienna, Universitätsstraße 7/5, 1010 Vienna, Austria
[4] Potsdam Institute for Climate Impact Research, Telegrafenbg A56, 14473 Potsdam, Germany
[5] Climate Analytics gGmbH, Ritterstraße 3, 10969 Berlin, Germany
[6] University of Iceland, Askja, Sturlugata 7, 101 Reykjavík, Iceland

Abstract. Secondary or onward mobility of refugees can pose considerable challenges for targeted and timely humanitarian assistance, and for long-term integration. There is very little systematic knowledge on the onward migration of refugees after their initial flight to a country of reception in general, and specifically in Turkey. In the paper we describe how the analysis of mobile phone Call Details Records can help to better understand spatio-temporal patterns of refugees' onwards mobility. The analysis reveals some clear, large-scale mobility patterns (from South to North, from East to West, from Centre to the Coast, to large urban areas), and also some temporal patterns, but also shows that human mobility is complex and accordingly requires more advanced analytical tools. We conclude that it might be worth of reframing registration policies for refugees, given the highly mobile share of refugee population, and the important role that this mobility probably plays for livelihoods.
Keywords: Call data record, mobility, social integration, Turkey, refugees movements

1 Introduction: Why is secondary or onward mobility important?

By the end of the year 2017, 68.5 million people were forcibly displaced globally due to conflict, persecution or violence — an increase of almost 3 million compared to 2016 [1]. While many of the displaced find refuge at the places to where they flee, internationally or internally, many continue their journey — in order to reunite with family or kin, to escape poverty or to improve their livelihoods,

to flee further persecution, or due to other reasons. This *secondary* or *onward mobility* is an important issue due to several reasons.

First, and apart from often violating regulations for refugees or asylum seekers, it poses often practical challenges for humanitarian assistance and long-term integration of refugees: it is difficult to effectively address mobile persons with targeted and timely support for immediate needs (such as shelter, food or health), but even more so to socially and economically integrate mobile populations into the receiving society. On the other hand, unknown mobility patterns also challenge efficient planning and management of the assisting communities and institutions: investments of (scarce) resources in infrastructure and measures for support are likely to be misplaced if a target group moves on.

Second, there is very little knowledge about the secondary mobility of refugees – the size of moving populations, their routes and central nodes in mobility networks, the specific trajectories and timings of movements, as well as interim and final destinations. In general, however, it seems like refugees' journeys – even from the same regions of origin – have become not only more diverse, multi-directional and longer, but also fragmented as periods of mobility interchange with longer phases of immobility. With more than 3.5 million displaced persons [1], Turkey is not only a particularly important host country for a highly vulnerable, yet mobile population group, but also the most significant mobility hub for onward mobility to Europe and for the return to countries of origin.

2 Objectives, Data and Methods

Knowing larger scale patterns of onward mobility of refugees within Turkey helps to better assess where and how support (e.g. emergency shelter, health services) should be provided, and where sustainable investments in the physical, economic and social infrastructure (e.g. employment opportunities, educational facilities) are best allocated. Knowing key drivers of secondary mobility, moreover, helps to anticipate future mobility, to make use of refugees' flexibility and to enhance incentives for refugees to stay at places where their needs can be catered for adequately.

We aim to show how the analysis of mobile phone Call Detail Records (CDR) can yield valuable information on the spatial and temporal patterns of secondary mobility of refugees, including flows over time and spatial trajectories between important points of origin and destination. In addition, and with additional information from secondary data, CDR analysis can also help to better understand the drivers of spatio-temporal mobility patterns of refugees.

2.1 Data

We were granted access to a set of CDR from the Turkish mobile operator Türk Telekom. Access was granted in the context of a call for projects with the aim of investigating the potential of analysis of mobile call data for the improvement of the situation of Syrian refugees in Turkey ("Data for Refugees: The D4R

Challenge on Mobility of Syrian Refugees in Turkey", see [2]). The data was sampled from CDR over the period of one year, from 01.01.-31.12.2017 from 992.457 Türk Telekom customers, thereof 184.949 that were registered in the customer database as "refugees"; 75% of these "refugees" were registered as male [3], however gender is not stated in the dataset. It has to be noted that the flag "refugee" in the datasets includes migrants, asylum seekers and foreigners with a temporary protection status; and while the individual attribution of the refugee status to an individual caller ID might not be possible with certainty, it should however be possible to deduce general patterns from aggregate analyses [3].

The data consisted of three datasets: Dataset 1 "Antenna traffic between cell tower locations", consisting of the total exchange of calls and text messages between cell towers; Dataset 2 "Fine Grained Mobility" of about 65.000 users that were newly sampled and assigned random ID numbers for time periods of two weeks over the course of the year; and Dataset 3 "Coarse Grained Mobility", containing of the CDR of a subsample of users over the course of the whole 12 months but spatially aggregated on prefecture / district level [3].

We concentrated for the present (first) analysis on the analysis of Dataset 3, which consisted, after consolidation, of 56,433,358 entries of the form 'Caller ID', 'Timestamp (DD-MM-YYYY HH:MM)', 'District ID', 'City ID'. All steps of data analysis were carried out using the open source statistical software R (3.5.0), on a standard desktop computer. Visualization was done in R, MSExcel and in Adobe Illustrator.

2.2 Methods

Dataset 3 was of special interest for our analysis, as it allowed for a longer temporal overview of mobility patterns, and thus enabled us to differentiate between short-term movements (e.g. for visits), circular, seasonal or more longer-term migratory movements.

In a first step, the data was checked for consistency and was consolidated. This consisted of the removal of duplicate data (in total almost 9,6 million rows) and the combination of data for outgoing and incoming voice calls to improve temporal coverage, as the dataset showed considerable gaps, especially in February and March (see Figure 1).

In a next step, we aggregated the DS3 dataset over individual callers and days, resulting in a dataset with a combination of all callers and the days on which they actively called or were called, their first and last districts for each given day (dayn), as well as the first and last districts where the callers had been on the preceding day (dayn-1) and on the subsequent day (dayn+1). We take the location of the last districts where callers have been on these days as a proxy of their places of residence, as this usually refers to places in the evening hours.

We then calculated the distance between the first district where callers were using their phone, and the last district, i.e. where they "entered" and where they "exited" the dataset. Thus a caller who started at district A, moved to B, to C, etc., and finally returned to A would be showing a zero value here, whereas for a caller moving from A to B, to C, etc. and ending in X we would obtain

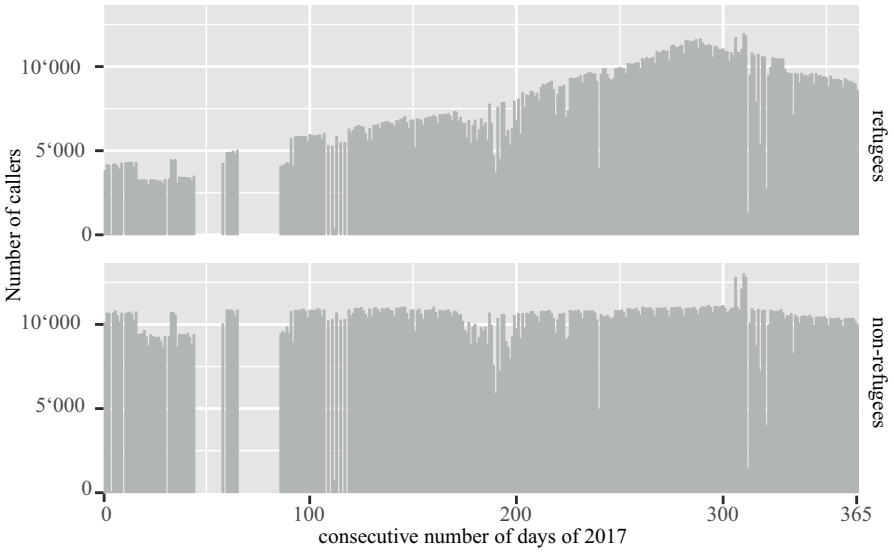

Fig. 1. Number of callers flagged as refugees and as non-refugees per day, over the whole of 2017. Note the large gaps in February and March, as well as also on single days later in the year.

the distance between A and X. Figure 2 shows the distribution of the callers according to this distance.

In a next step, a mobility (origin-source) matrix was created, containing the number of refugees moving between districts. From this, information on net migration flows between districts was calculated.

Based on this, a subset of refugees was selected who moved more than 100km between their districts of first and last appearance in the dataset. From an analysis of the mobility patterns of a small sample of callers we assume that a movement of more than 100km can be interpreted as a proxy for a temporary or permanent shift of residence, hence as migration as opposed to commuting.

Then we determined when/on which days these “migrating” refugees made movements of more than 100km, and for each day we summarized the number of the callers moving more than 100km on that day. Because of the high temporal variability of callers covered in the dataset (see Figure 1), we normalized the number of callers moving with the total number of callers on each particular day, in order to ensure comparability over time.

3 Results

Figure 3 shows an overview of “mobile” and “less mobile” (defined as moving more, respective less, than 100km between their first and their last appearance in the dataset) callers. If we consider this a relevant proxy for migration, then

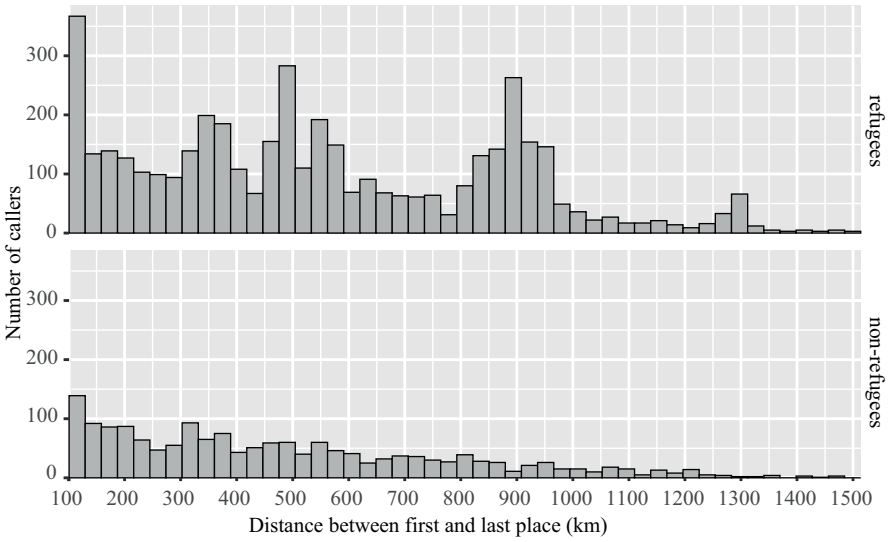

Fig. 2. Distances between first and last place of callers (1=refugees, 2=non-refugees, only dis-tances >100km shown)

about 14% of refugee callers in the whole of Turkey can be considered as onward migrating. In the border provinces to Syria (Hatay, Gaziantep, Sanliurfa, Kilis and Mardin) this ratio is about 17.7%.

If all movements of refugees of more than 100km are considered, regardless of the overall distance between their first and last district, a large number of bidirectional movements (between districts and back) becomes apparent, as Figure 4 shows. This indicates that refugees are highly mobile, even if they do not permanently (or at least over a longer period of time) change their place of residence.

This is also reflected in the high total cumulative distance (i.e. the sum of any movement, adding up also cyclical mobility, both migration and/or commuting) that many refugees travel over the time of their coverage in the dataset, compared to the (directed) total distance between the first and the last place when they are registered in the dataset (Figure 5).

When the total distance between the first and last districts of appearance in the dataset is taken into account (i.e. when "mobile persons" are defined as having a distance between the first and last place of more than 100km), and when the net flows of these "mobile persons" between places are calculated, a clearer pattern of origins and destinations emerges (see Figures 6 and 7). Generally, three larger migration systems seem to dominate: a) larger urban centres as destinations (notably Istanbul, Ankara, Adana, Antalya and others), b) a general direction from South to North and from East to West, and c) the movements within the Western Turkey-Syria border region, including movements back to border towns indicating return migration.

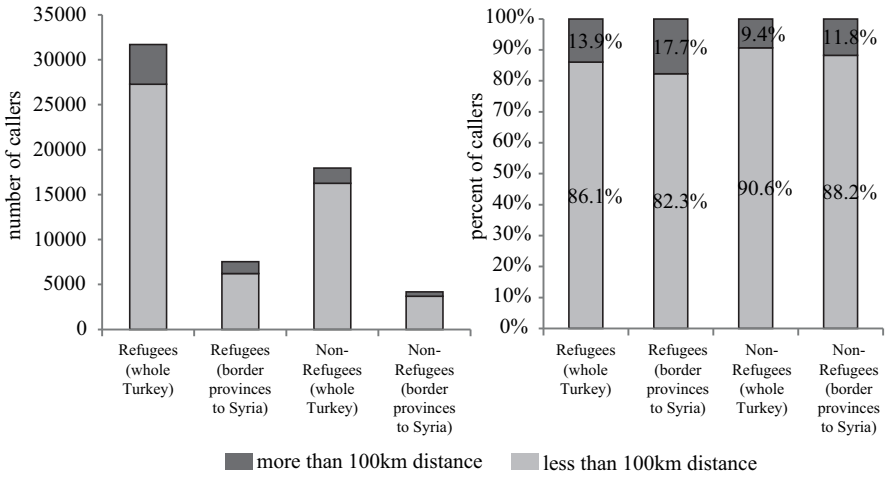

Fig. 3. "Mobile" and "less mobile" callers in the dataset (mobile = more than 100km between first and last district, less mobile = less than 100km), left: absolute numbers, right: percent; with "border provinces" we refer to Hatay, Gaziantep, Sanliurfa, Kilis and Mardin

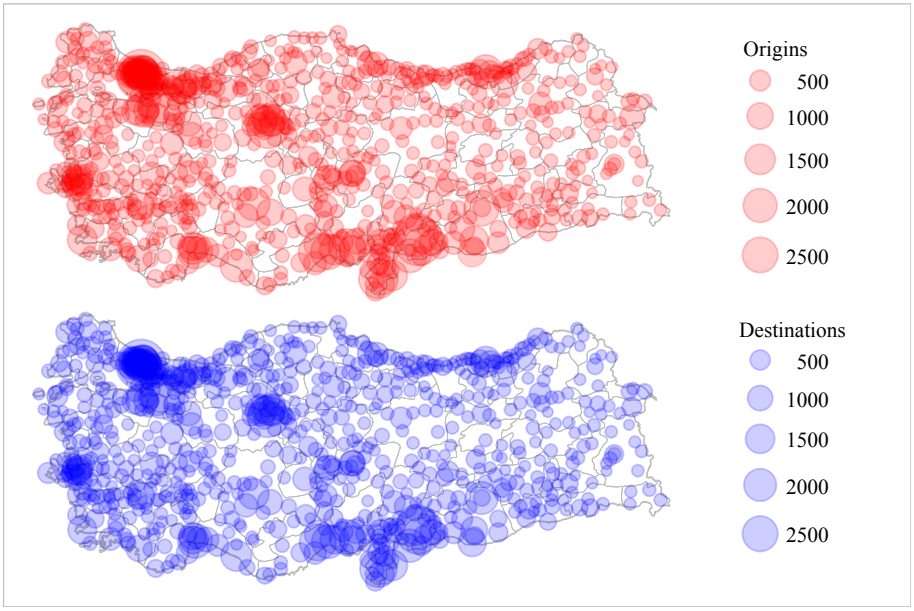

Fig. 4. Source and destinations of refugee movements (>100km) of all refugees over the total time of 2017. The similarity of the origin and destination maps (i.e. destinations are at the same time sources) indicate the high number of circular or back-and-forth movements of refugees.

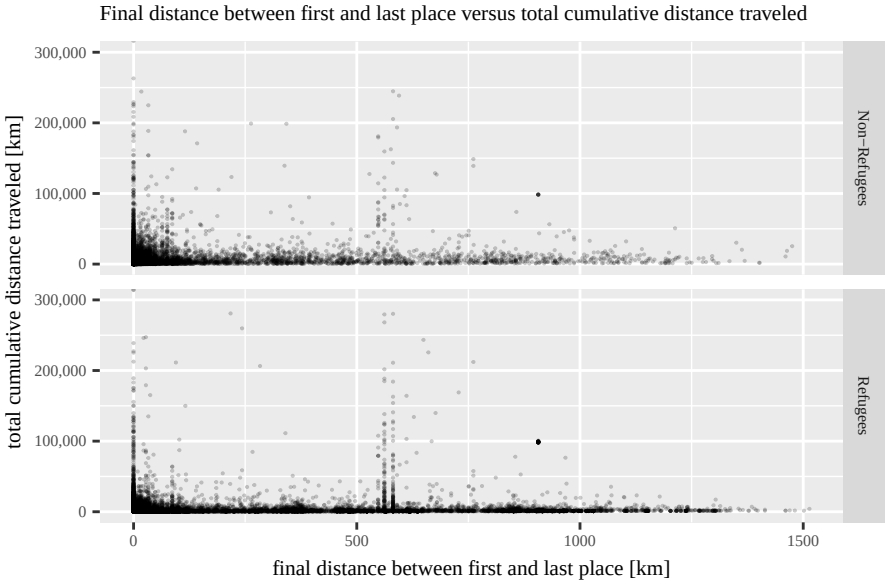

Fig. 5. Final distance between first and last place versus total cumulative distance traveled (1=refugees, 2=non-refugees), each dot equals one caller. The final distance refers to the distance between the first and the last district where a caller was registered in the dataset, the total cumulative distance includes all movements (one-time-migration, cyclical migration, commuting, visits, etc.).

Top 10 Provinces (in-migration)		Top 10 Provinces (out-migration)	
Istanbul	929	Mersin	262
Hatay	177	Mugla	175
Bursa	113	Izmir	166
Sanliurfa	64	Trabzon	157
Kayseri	28	Konya	143
Agri	22	Antalya	136
Kilis	14	Adana	109
Usak	7	Kocaeli	83
Karabük	3	K. Maras	81
Kinkkale	3	Sakarya	62

Table 1. Top 10 in-migration and out-migration provinces (number of refugees, according to their first and last appearance in the dataset)

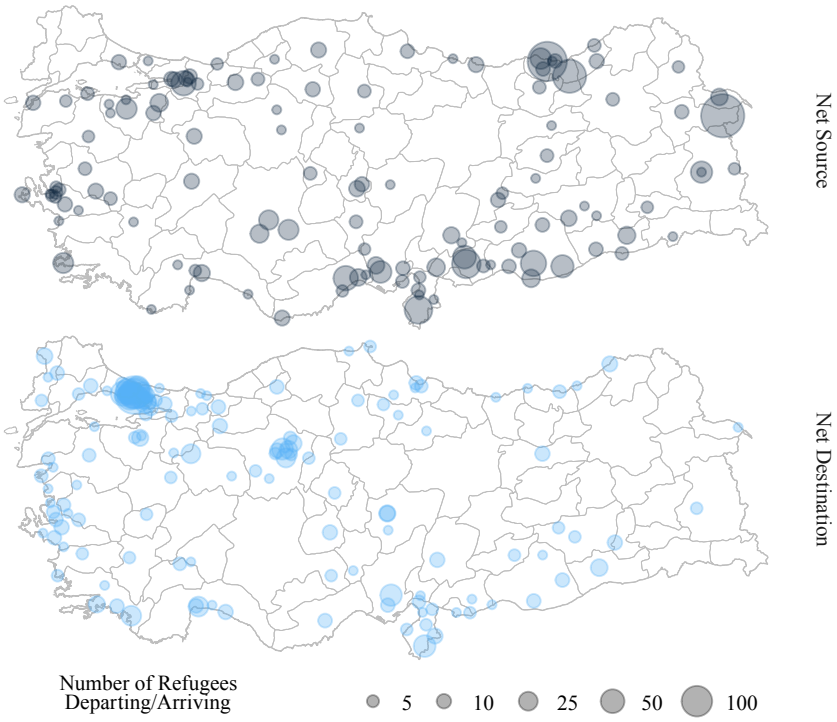

Fig. 6. Net sources and destinations of refugee movements (only refugee callers who moved more than 100km between their first and last district) over the total time of 2017 (by districts)

To determine the "migration intensity" over time, the days with movements of more than 100km of all refugees were identified. The number of refugees with such movements was summarized per day and normalized with the number of refugees appearing in the dataset on every day (see Figure 8). This data shows distinct patterns, for example the weekly mobility of the non-refugees, or the increased mobility during the two Eid festivals (both refugees, and especially non-refugees).

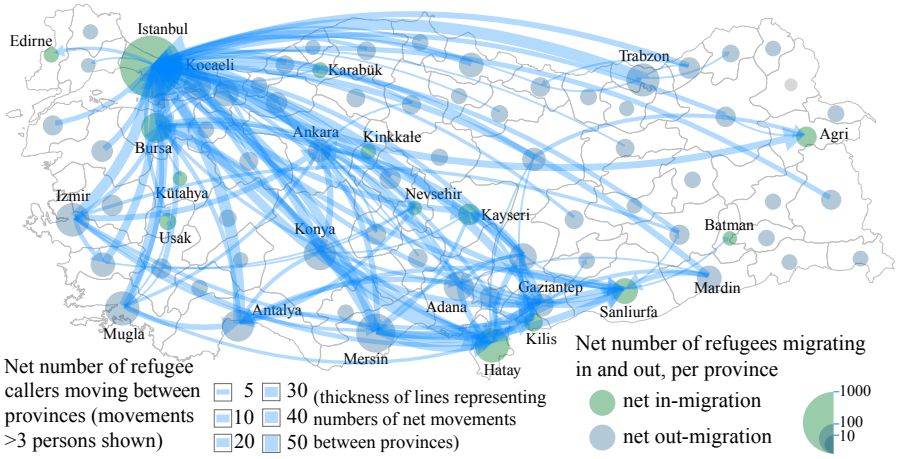

Fig. 7. Net flows (directed) of refugees, aggregated to flows between provinces of origin and destination (only flows of more than 3 individuals shown, for purposes of clarity)

4 Discussion

First it seems noticeable that the larger share of refugees do not *permanently move* more than 100km. This is in line with findings of existing research that many Syrian refugees stay either close to the border (in order to easily return when this is possible) or in areas where they have social networks and can find accommodation and jobs [4][5]. Also, refugees registered as temporary protection beneficiaries are required to stay in their assigned province and have to comply with reporting requirements [6].

At the same time, even refugees who cannot be considered as shifting their residence permanently do show a remarkably high degree of mobility. This is noteworthy especially insofar, as refugees would presumably have less financial resources that are necessary to travel and cover large distances, and it would be very interesting to investigate this issue further.

What becomes also obvious is that among those callers who *do move* more than 100km between their first and last place, refugees are *relatively more mobile*

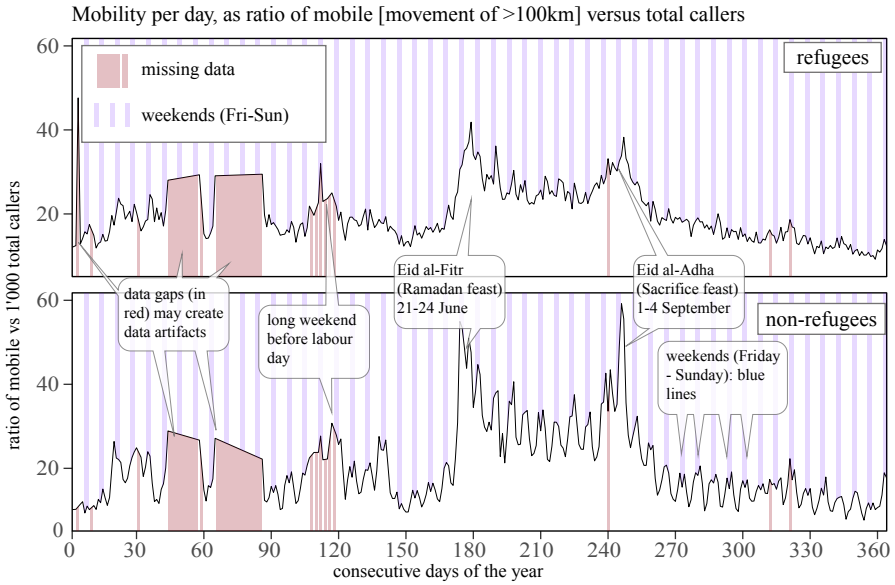

Fig. 8. Mobility per day, expressed as the ratio of mobile callers (moving more than 100km on that particular day) by total refugee / non-refugee callers on that day; note the data gaps in February/March.

than the comparable non-refugee population in the dataset (Fig 2 and Fig 3). While non-refugees cover almost twice the mileage than refugees (mostly through commuting and visits), refugees do move further, when the distance between their first and last place of appearance in the dataset is considered (Table 2).

Refugee Status	Total cumulative distance traveled	Absolute distance between first and last place
Refugees	6'365 km	588 km
Non-Refugees	12'195 km	475 km

Table 2. Total cumulative distance traveled, and absolute distance between first and last place, by refugee status

The dataset also shows clear patterns of larger distance movements to particular destinations, e.g. between places at the Syrian border, the Black Sea, and Istanbul, Ankara or Izmir.

Regarding mobility patterns in general, there seems to be a considerable back-and-forth movement, reflected in the differences of the total vs. the net flows (Fig 4 vs 6). The net flows (Fig 6 and 7) do reflect general movement patterns from the South to the North and from the East to the West, predominantly to Istanbul, Ankara and the Coast. There is also some movement to other urban areas, and

also bidirectional mobility to and from more rural places in central Turkey, indicating agricultural labour migration. There is also movement towards the Syrian border region, indicating possibly some return mobility.

The movement over time (Fig 8) clearly shows a growing mobility towards the two Eid festivals (Eid al-Fitr on 21st - 25th June and Eid al-Adha on 1st – 4th September 2017), which is then declining again. Mobility around the Eid festivals is much more pronounced among the non-refugee population, which might be explained by the better economic situation (allowing for leisure travel and family visits). Also visible is the stark contrast for weekend-mobility, which is strongly appearing in the non-refugee data and presumably related to commuting, and only very weak in the refugee-data. Although visible to some extent, there seems to be less seasonal mobility appearing in the data than we initially expected (with regard to temporal agricultural labour demand in central Turkey). This, and also other, smaller variations in mobility, especially for the refugees, need more explanation and require a more in-depth look into the data.

5 Conclusion and outlook

Methodologically: we can conclude that the analysis of CRD enables the unveiling of mobility and migration patterns to a hitherto unprecedented level of detail, both temporally as well as spatially. However, the complexity of human mobility requires also more advanced and in-depth analyses of mobility patters, including approaches such as most frequented locations or full temporal origin-destination-matrices [7][8]. Due to resource restrictions, we have applied a very simple approach of delineating migration, by assuming that those individuals can be regarded as migrants (permanently or semi-permanently shifting their residence), whose first and last districts of appearance in the dataset are located more than 100km apart from each other.

Knowledge on refugees' mobility: general patterns of mobility (South to North, East to West, to urban centers and to the coast, to and from agricultural areas) are clearly reflected in the data; however it becomes also evident that mobility is more complex, and that boundaries between commuting, visiting, temporary, seasonal and permanent migration might be more fluid than often conceptualized in migration research. Cultural motifs for movement (i.e. the Eid festivals) are more important than initially expected, and seasonal mobility due to agricultural labour demand seems to be less important than initially expected. However, the underlying motivations for mobility can only be inferred indirectly—either through additional data, or through making assumptions (i.e. that more intense connectivity as represented in dataset 1 might be a reason for mobility between persons).

Implications for refugees' wellbeing: although the majority of refugee callers is less mobile or at least appears not to permanently change location between the first and last appearance in the dataset, still a significant share of refugees (about 14%) does so. Within the existing registration system this implies limited or difficult, or even suspended, access to social services such as

healthcare, education, housing, and so on for these mobile refugees. Thus, it could be an important contribution to these refugees' wellbeing if changing the place of registration would be possible (more) easily and quickly — reflecting the mobile reality of their lives and livelihoods.

Open questions: As the analysis work was done with limited time resources, there still remains some work to do in a series of next steps:

- More in-depth analysis of general mobility patterns: it would be necessary to better differentiate between permanent and temporary migration (as a permanent or temporary shift of residence to another place), visiting mobility (as short-term change of place) and labour mobility (as longer term but onwards and finally returning mobility, e.g. following the changing places of agricultural labour demand), as these different kinds of mobility are resulting in different needs of the refugees in terms of support and service provision;
- A spatially more fine grained analysis of mobility would be beneficial, differentiating between mobility patterns from and to rural areas, smaller towns and larger urban centers, as this will also allow to link mobility with structural drivers (labour markets, existing social networks to refugees in the destination places, etc.);
- A spatially explicit analysis of mobility over time would help to disentangle different types of mobility (on the basis of the differentiation between the four mobility patterns mentioned above), and would allow for associating these types of movements with the spatio-temporality of seasonality, of weather and climate phenomena (precipitation, land surface temperature, drought indices, cf. [9] and [10]), conflict and political events (we have started working on [11] and [12]), and the important role of translocal connectivity of refugees to others (derived from the communication dataset DS1 provided through the D4R challenge).

On a more general level, it seems important to us to remark two issues: first, the inference of mobility patterns from CDR obviously poses challenges of data protection and privacy. In the context of this study, the organizers of the D4R challenge put a special emphasis on these issues, including anonymizing the datasets, a review process safeguarding the interests of refugees and setting up clear contractual agreements for the research teams using the data. The analysis of such data in other contexts would require similar standards. However, and secondly, the analysis of CDR can yield a wealth of fine-grained information on human mobility that is almost impossible to achieve with traditional means (e.g. surveys, registration or census data). Thus, given the safeguarding of privacy and data protection, researchers' access to CDR and similar data could benefit both science and development practice.

6 Acknowledgement

We would like to express our sincere gratitude to the D4R challenge organizers and Türk Telekom for setting up the challenge and for providing us with the datasets.

7 References

1. UNHCR: Global Trends – Forced Displacement in 2017, http://www.unhcr.org/5b27be547.pdf, last accessed 10.09.2018
2. D4R Homepage: http://d4r.turktelekom.com.tr/, last accessed 15.09.2018
3. Salah, A.A., Pentland, A., Lepri, B., Letouzé, E., Vinck, P., de Montjoye, Y.A., Dong, X. and Dağdelen, Ö., 2018. Data for Refugees: The D4R Challenge on Mobility of Syrian Refugees in Turkey. arXiv preprint arXiv:1807.00523.
4. Istanbul Policy Center (Ed) 2015, Urban Refugees: The Experiences of Syrians in Istanbul. Istanbul Policy Center. Instanbul.
5. Tuzcu, N., 2014, Syrian Urban Refugees in Turkey: Spatial & Social Segregation. Pub-lished by Displacement Research & Action Network, MIT Department of Urban Studies and Planning. Online available http://mitdisplacement.org/new-page-43/, last accessed on 05.09.2018.
6. AIDA - Asylum Information Database: Freedom of Movement, Turkey, http://www.asylumineurope.org/ movement-1, last ac-cessed 12.09.2018
7. González, M.A., Hidalgo, C.A., Barabási, A.-L., 2008, Understanding individual human mobility patterns, Nature Vol 453(5), pp. 779-782, doi: 10.1038/nature06958
8. Barbosa, H., Barthelemy, M., Ghoshal, G., James, C.R., Lenormand, M., Louail, T., Menezes, R., Ramasco, J.J., Simini, F., Tomasini, M., 2018, Human mobility: Models and applications, Physics Reports 734 (2018) 1–74, doi:10.1016/j.physrep.2018.01.001
9. PSD Gridded Climate Datasets, https://www.esrl.noaa.gov/psd/data/gridded/, accessed 12.09.2018
10. GHCN Gridded Products – Temperature, Precipitation and Drought, https://www.ncdc.noaa.gov/temp and-precip/ghcn-gridded-products/, last accessed 12.09.2018
11. UCDP Georeferenced Event Dataset (GED) Global version 18.1 (2017), http://ucdp.uu.se/downloads/, last accessed 17.08.2018
12. ACLED – Armed Conflict Location and Event Data Project, https://www.acleddata.com/, last accessed 17.08.2018