Individual level drivers of seasonal agricultural mobility in Turkey

Bilgeçağ Aydoğdu[1], David Natarajan[1], Subhi Güneş[2], Albert Ali Salah[1,3]
[1] Utrecht University, The Netherlands
[2] Turkcell Technology, Türkiye
[3] Boğaziçi University, Türkiye
Emails: {b.aydogdu,a.a.salah}@uu.nl, d.j.natarajan@students.uu.nl, subhi.gunes@turkcell.com.tr

I. Introduction

Seasonal agricultural migration in Turkey is a complex phenomenon, largely influenced by social, demographic and economic factors. Over the past decade, the landscape of seasonal agricultural employment in Turkey has undergone a substantial shift, primarily driven by the massive influx of refugees from Syria. With the onset of the Syrian refugee crisis in 2013, the wages got lowered and working conditions got worsened in the informal agricultural industry, which pushed some natives out of seasonal agricultural employment [1].

In this work, we aim at developing a model for predicting seasonal mobility from Istanbul, the most populous city in Turkey, to the northeastern hazelnut-producing cities of Ordu, Giresun, and Trabzon between July and August 2020. We also assess the drivers of seasonal mobility using extended detail records (xDR) from the largest telecom operator in Turkey. Recent research done with a survey collected in 2018 [2] demonstrated that the agriculture sector is one of the primary employment areas, especially for refugee women. Surveys and census studies give insights on the characteristics of seasonal workers, but are costly to apply.

The contribution of this study is to (1) develop xDR-based indicators of seasonal mobility and (2) show the usefulness of xDR-based mobility and census-based features in generating insights on the characteristics of seasonal agricultural workers. Previous studies have analyzed the usefulness of call detail records (CDR) [3] [4] for measuring seasonal migration, yet xDR is an underused data source in migration analysis. Temporal frequency of signals in xDR is higher than CDR, which helps for creating more complete trajectories and more accurate mobility and migration indicators, thus closing some data gaps in seasonal agricultural mobility.

II. Data

The xDR datase we use is prepared within the scope of the HummingBird EU H2020 Project [5], following the fine-grained mobility approach (and accompanying ethical framework for data anonymization and aggregation) suggested in Data for Development (D4D) [6] and Data for Refugees (D4R) challenges [7]. It includes user id, timestamp, and the id of the cell tower used by the user. We enriched the xDR with (noisy) nationality and sex flags using the telco indicators, and the noise in these flags serves further anonymization. Up to thirty percent of users flagged as ”male” may be females. The datasets are shared with our research group under strict data use agreements and ethical approvals.

We have approximately 100,000 users subsampled with replacement from a pool of 4 million users every two weeks throughout 2020. The short sampling period is used to prevent profiling of users. We combine these data with various other data sources, such as the night light satellite data created from the Earth Observation Group (EOG) sources, and neighborhood level indicators collected in Istanbul as a part of the Mahallem Istanbul project[1]. We used spatial scaling methods to recalculate neighborhood-level indicators around the cell towers. These indicators give insights into the demographic characteristics of the Turkish population living around the cell tower such as education, age, and maritial status.

III. Methodology

We processed the anonymized fine grained mobility data sets to identify the users whose home location (based on activity during 18:00-06:00) is in Istanbul, and who have been to the Northeastern cities in the two week sampling period. The number of travellers originating from Istanbul to the harvesting cities peaks in the first week of August. In order to analyze the drivers of this mobility behavior, we focus on the trips that took place between 15/07/2020 and 31/08/2020, the high season for harvesting hazelnuts[2]. Although we cannot know whether trips are indeed associated with agricultural work with high certainty, this approach optimizes the chances that is the case. The gender and nationality breakdown of the groups are given in Table I.

We used the scikit-mobility library [8] to calculate six different mobility-related indicators, namely, the radius of gyration, maximum distance from home, number of visits, number of locations, maximum distance travelled, and total distance travelled calculated in straight linesseperately for night (between 18:00 and 06:00) and day for all days that the user spent in Istanbul.

This work was funded by the European Union Horizon 2020 Research and Innovation funding Program, project number 870661.

[1]https://www.kalkinmakutuphanesi.gov.tr/dokuman/mahallem-istanbul/554
[2]https://arastirma.tarimorman.gov.tr/findik/Sayfalar/Detay.aspx?SayfaId=32

	Likely Agricultural Migrant	Likely Non-Migrant
Turkish Male	532	25614
Turkish Female	185	11310
Syrian Male	188	16432
Syrian Female	76	6747
Total	981	60103

Table I: Number of users grouped by category.

	F1 (mean)	F1 (std)	AUC-ROC (mean)	AUC-ROC (std)
Turkish Male	0.68	0.03	0.75	0.04
Turkish Female	0.63	0.05	0.70	0.05
Syrian Male	0.72	0.05	0.80	0.04
Syrian Female	0.67	0.09	0.76	0.08

Table II: Performance metrics across subsets.

Following [9], we first develop a model to predict seasonal mobility from urban mobility related metrics (which indicates employment) and census-driven indicators, and then we analyze feature importances to understand the drivers. We develop four different prediction models for each demographic group. For Syrians, we do not use census features, as they were collected for the Turkish population. Including census features improves the prediction model for Turkish group only slightly, whereas it lowers the performance for Syrians.

In our data set, mobility related metrics were highly correlated. We used a principal component analysis (PCA) to day and night mobility related variables seperately to reduce the dimensionality to four by keeping 80 percent of the variance. Furthermore, we dropped various demographic features that were in high correlation with other features (correlation estimate higher than 0.8) based on a collinearity test.

The challenge of our data is the great imbalance between the number of users who have been and who have not been to the harvest cities. To address this, we randomly downsampled the non-migrant group to the same numbers with the migrant group [9], for a hundred times. We applied 5-fold cross validation in each such sample. We used logistic regression, as our aim is to have simple and interpretable model. The feature importances are computed for each fold, in each sample.

IV. Results and Discussion

We share the model performances in Table II, measured by using F1 and Area Under the Receiver Operating Characteristic Curve (AUC-ROC) scores that are averaged across all experimental samples. The model we developed has the highest performance for predicting the seasonal mobility of Syrian males, and the AUC-ROC scores are comparable to similar studies in the literature [9].

To understand feature importances we report the odds ratio (OR) for all features averaged across all samples in Table III. In logistic regression, OR measures the impact of features on the individual decision to go to the harvesting regions during the high season (value above 1 indicates the chances of going). In Table III, we report the top three most positively (blue) and negatively (red) influential features for each group.

The results show that urban mobility related principal components (PC) are highly influential in decision making for sea-

	Turkish M	Turkish F	Syrian M	Syrian F
Day mobility PC 1	1.15	1.04	1.43	1.24
Day mobility PC 2	0.63	0.59	0.55	0.62
Night mobility PC 1	1.86	1.96	2.27	1.95
Night mobility PC 2	0.74	0.91	0.77	0.62
Nightlight brightness	1.08	1.14	1.24	1.21
Percentage nocturnal	0.96	0.96	0.91	0.84
Average Age	0.84	0.83	-	-
Age Dependency ratio	1.47	0.70	-	-
Baby boomer ratio	0.58	1.10	-	-
Illiterate ratio	1.31	1.36	-	-
Literate but no education ratio	0.42	0.94	-	-
Average education dur. (men)	1.10	1.16	-	-
Married ratio	1.29	1.46	-	-
Total female population	1.02	1.22	-	-
Population density	1.08	0.86	-	-

Table III: Mean Odds Ratios per subset.

sonal mobility for all groups. While the ”maximum distance” has a positive association, ”radius of gyration” and ”number of locations” have a negative association with seasonal mobility (i.e. seasonal migrants tend to be less mobile, but travel greater distances). For users with less mobility, the unemployment probability is higher. The illiterate population ratio around the home location cell tower is positively associated with the target variable for the Turkish group. For women, we also found that ”married ratio” is associated positively, but ”average age” is associated negatively with seasonal mobility.

In this study, we show that the xDR data can help to improve our understanding of the drivers of seasonal mobility. Both for Syrians and Turkish, the biggest driver of seasonal migration is the unemployment. For Turkish population, we also found that the illiteracy can be one of the biggest drivers. We argue that these results can complement the information collected on the seasonal agricultural workers through surveys.

References

[1] Z. Bayramoglu and M. Bozdemir, “Dış göçlerin mevsimlik tarım işçiliği üzerine etkilerinin değerlendirilmesi,” *Journal of the Institute of Science and Technology*, vol. 9, no. 2, pp. 1164–1176, 2019.

[2] M. Demirci and M. G. Kırdar, “The labor market integration of syrian refugees in turkey,” *World Development*, vol. 162, p. 106138, 2023.

[3] P. J. Zufiria *et al.*, “Identifying seasonal mobility profiles from anonymized and aggregated mobile phone data. application in food security,” *PloS one*, vol. 13, no. 4, p. e0195714, 2018.

[4] S. Turper Alışık *et al.*, “Seasonal labor migration among Syrian refugees and urban deep map for integration in Turkey,” in *Guide to Mobile Data Analytics in Refugee Scenarios*. Springer, 2019, pp. 305–328.

[5] B. Aydogdu, A. A. Salah, O. Ones, and B. Gurbuz, “Description of the mobile CDR database,” *HumMingBird Project Deliverable*, vol. 6, no. 1, 2021. [Online]. Available: https://hummingbird-h2020.eu/publications

[6] V. D. Blondel *et al.*, “Data for development: the D4D challenge on mobile phone data,” *arXiv preprint arXiv:1210.0137*, 2012.

[7] A. Salah, A. Pentland, B. Lepri, and E. Letouzé, *Guide to Mobile Data Analytics in Refugee Scenarios*. Springer, 2019.

[8] L. Pappalardo, F. Simini, G. Barlacchi, and R. Pellungrini, “scikit-mobility: A python library for the analysis, generation and risk assessment of mobility data,” *arXiv preprint arXiv:1907.07062*, 2019.

[9] V. Dias, L. Fernando, Y. Lin, V. Frias-Martinez, and L. Raschid, “Framework to study migration decisions using call detail record (CDR) data,” *IEEE Transactions on Computational Social Systems*, 2022.