Modeling Forced Migrant Populations Using ABMs

Melonie K Richey[1]

[1] George Mason University, Fairfax VA 22030, USA
mrichey2@masonlive.gmu.edu

Abstract. *The study of forced migration is neither new nor novel given the prominence of displacement events worldwide over the last several decades. Various computational models exist for analyzing and predicting refugee movement in disparate geographic regions. The construction of these models are neither automated nor are they generalizable across use cases. The purpose of the research presented in this paper is to automate the manual creation of a simulation environment and to convert a diffusion simulation approach to a true Agent-based model to study the phenomenon of forced migration from Syrian Arab Republic to the Republic of Turkey during the Spring 2019 timeframe. Using open source data and computational packages, this model presents a repeatable framework and codebase for the automated creation of a simulation environment, the inclusion of factors that influence forced migration, agent activation regimes, and agent decisionmaking.*

Keywords: migration, agent-based model

1 Introduction

The study of forced migration is neither new nor novel given the prominence of displacement events worldwide over the last several decades. Computational methods to model and predict these events, however, are increasing in prevalence and complexity. Academic groups, aid organizations, and even the US Intelligence Community are devising analytic methods and computational models to better understand and predict the movement the refugees, asylum-seekers, internally displaced persons (IDPs), returnees, and stateless persons worldwide. [1][2]

At the end of 2017, the Syrian Arab Republic alone hosted 6 million IDPs while almost 4 million of its citizens sought refuge in neighboring Turkey, 1 million in Lebanon, and 750,000 in Jordan. Equally tragic though lesser magnitude circumstances can be found in places such as Nigeria, Central African Republic (CAR), Democratic Republic of Congo (DRC), Colombia, Iran, Ukraine, and Pakistan. [3]

The objective of the research presented in this paper is to review extant computational methods for addressing forced migration (Section 2) and present a predictive agent-based model (ABM) for predicting migrant flows resulting from large-scale forced migration events (Section 3). The specific purpose of the model is to predict likely destinations of refugee agents as they move from areas of conflict to areas of asylum, which is shown in Section 4. Finally, in Section 5, a summary of the paper is provided and areas of further work are identified.

2 Background

2.1 Forced Migration Theory

Research around why migrants move and the so-called *push* and *pull* factors of migration predate any computational modeling of such drivers. The United Nations High Commissioner for Refugees [3], founded just after World War II to assist the world's displaced population, makes the strong distinction between factors that *explain* migration and factors that *influence* migration – the former originating in the "household/community/country of origin" such as inequality or conflict, and the latter more "destination-specific" such as the "presence of a co-ethnic community." This distinction aligns closely with the push/pull factor theory. [4][5]

Both statistical investigation and field research support the most influential factors are first and foremost social (presence of co-ethnic community in target destination), and, to a lesser extent, economic (presence of job opportunity), and physical (presence of infrastructure supporting the journey and absence of threat to persons while making that journey). [6][7][8] To the dominance of the first factor, the idea of "herd behavior" was introduced – "people go where they have observed others go" – a tenet which also seeks to explain phenomena such as "why migrants continue moving to areas with saturated labor markets, which would be expected to not occur when migration networks supply information on local economic conditions." [9] The dominance of this factor and the attention it has garnered, however, makes the case for the inclusion of social networks in migration models. [10] If people tend to go where others have gone, social networks and social media, then, are a migrant's glimpse into not only where other migrants have gone but where other migrants currently are and the social, political, and economic state of the target location.

2.2 Spatial Interaction Models

Traditional methods for modeling and, to a lesser extent, forecasting mass migration have made substantial use of Spatial Interaction Models (SIMs), which includes implementations of the gravity model (See [11] for one such review of statistical models). The gravity model is "an early form of spatial interaction model…which postulates that flows are likely to be, in some way, related to the sizes of the origin and destination areas, and in some way inversely related to distance between them." [12] The gravity model's application to the field of migration studies quickly led to not just one but a family of spatial interaction models.

As both spatial interaction models and Geographic Information Systems (GIS) matured in the late 80s and 90s, the ability to visualize migration using flow mapping techniques (and spatial movement in general) became more mainstream. [13] Flow mapping gave way to resistance mapping techniques, or modeling migration as the path of least resistance across a spatially explicit landscape filled with barriers to movement. [14][15]

Traditional methods focus on voluntary versus forced migration. While voluntary migration is important, the methods and techniques used to model voluntary versus forced decision processes are not the same. Traditional voluntary migration models tend to be econometric in nature weighting economic factors more highly than other factors in the model, if other factors are considered at all. Second, SIMs lack the granularity and control to be able to model migrant decisionmaking at the individual level. SIMs also do not account for the increasing importance of social media and social networks in migration modeling, nor do they provide a computational environment where individual migrants can make decisions non-deterministic decisions about the migration process. For this capability, we turn to agent-based modeling.

2.3 Computational Modeling of Forced Migration

Given that migration, whether forced or voluntary, is a very personal decision taken either as an individual or as a family unit, ABMs are an ideal modeling technique for this particular sociocultural phenomenon because, by allowing individual decisionmaking within the simulation, they provide an environment for the emergence of non-deterministic spatial and social patterns among refugees. Any truly canonical theory of the evolution of social phenomena must allow for path-dependent variation, or "multiple paths, not a singular causal path" to the emergence of the sociocultural event or situation. [16] ABMs are a natural approach in the computational analysis of forced migration because not only do they allow the simulation of "individual actions of diverse agents [while] measuring the resulting system behavior and outcomes over time." [17] Not only are ABMs created in a variety of computational environments, they can be spatially explicit and readily provide a quantitative means to include and test the influence of push/pull factors and social networks on forced migration.

Despite this powerful research potential, the use of ABM for experimenting and exploring geographical phenomena…is still in its infancy," as is the exploration of robust social network analysis techniques in migration networks. [17] Previous simulation efforts have modeled migration as a form of diffusion through network structure [18] and leveraged 'path of least resistance' modeling techniques. [15][19][20] These approaches are more akin to system dynamics models than true ABMs. [21]

The FLEE model, a network-based ABM, is likely the most robust of these recent efforts. [22] The FLEE model maps three data sources (Armed Conflict Location & Event Data (ACLED), geospatial base layers relevant for the area of interest, and UNHCR situation data) to four model factors (conflict zones, transit routes, refugee camps, and refugee populations) constituting the simulation environment. The simulations are run across three independent geographic areas: Burundi, Central African Republic (CAR), and Mali.

Where the model is lacking is in its use of refugee camp and population locations and ACLED conflict event locations for travel origins/destinations and, as is common with spatially-explicit simulations of refugee flows, its use of UNHCR refugee population maps as both an input to the simulation environment and as a means to validate simulation results.

2.4 Generalizability and Automation

Generally speaking, computational models of forced migration – ABMs or otherwise – are highly empirical and case-specific. Generalizable models are rare with the FLEE model serving as the most generalizable model to date. Two things contribute to the difficulty of creating a generalizable or truly global model of forced migration: 1) the region- and case-specific nature of the simulation environment and agent decisionmaking (e.g. in one area, religious identity may play a large role such as in CAR where it is less important or overruled by ethnic identity in other areas, such the Kurds in Turkey, Syria, Iraq, and Iran), 2) the relatively small sample of forced migration situations worldwide, especially over the previous decade during which time social media became prevalent. While the generalizable FLEE framework is sufficiently robust, a goal of this research is to create an alternative framework leveraging many of the decisionmaking mechanisms referenced above in addition to an alternative means of developing the simulation environment.

The second issue with using computational models to study forced migration is the amount of time required to create a case-specific simulation. The foundational research for this project leveraged simulation environments that were manually coded. A secondary goal of this research is to address this issue by creating an automated pipeline whereby data sources can be inserted as required to derive the simulation environment and the factors that influence agent decisionmaking in forced migration.

3 Methodology

A spatially explicit network-based ABM requires several components: 1) *geotagged origin and destination locations*, 2) knowledge of *factors that influence migration decisions* both temporally and spatially, 3) an *agent activation regime*, 4) and *agent decisionmaking*. The first two constitute the simulation environment while the second two comprise the simulation and model logic. Specifically, the third determines how the model is initialized, and the fourth determines how agents move during the course of the simulation. Agent heterogeneity and the inclusion of social networks are both optional, though the inclusion of the first would allow for multiple activation regimes and multiple decisionmaking mechanics, where inclusion of the second affects both the simulation environment and agent decisionmaking.

3.1 Computational Specifications

The model was created in Python 3.0 with the following packages: The Geospatial Data Abstraction library (GDAL), geopandas, networkx, numpy, and pyplot. Visualizations were created in Python and QGIS. Data for this initial research are from the Database of Global Administrative Areas (GADM), the UNHCR, the Turkish Statistical Institute, and from the Armed Conflict Location & Event Data Project (ACLED). [23] A list of model parameters is provided in Table 1.

Table 1. List of adjustable model parameters and their descriptions

Model Parameter	Parameter Range	Description
Percent_move_at_camp	0.00-1	Percent chance that refugees will move if located in a district node with one or more refugee camps
Percent_move_at_conflict	0.00-1	Percent chance that refugees will move if located in a district node with one or

		more conflict events
Percent_move_at_other	0.00-1	Percent chance that refugees will move if located in any other district
Seed_Refs	Integer	Number of refugees that cross at each border crossing each day
Border_crossing_list	Strings	List of border crossing nodes
Directionality_coords	Lat/Long	Lat/Long values of general end destination geographic area
Num_kin	Integer	Number of kin refugees in the social network for each individual refugee
Num_friend	Integer	Number of friend refugees for each refugee in the social network

3.2 Data Engineering of the Simulation Environment

The spatial extent for this research is the Level 0 Administrative area of Turkey. The first element of the simulation environment is the geotagged origin and destination locations. Suleimenova et al. elected to use geotagged conflict events from the ACLED database within the target country as origin locations and the locations of refugee camps as potential destination locations. This, however, makes the ABM somewhat of a self-fulfilling prophesy in that agents only have the option to move to refugee camp locations as opposed to anywhere in the target region. Cushman et al. 2010 present a more comprehensive method of overlaying the area of interest with a 10kmx10km grid and using each cell as a possible location for movement.

Developed here, however, is the most organic method for creating the spatially explicit simulation environment without adding any excess computational or data procurement burden. Since the data at Administrative Level 2 are readily available and statistical data (population or otherwise) is also frequently available with this granularity (or at Admin Level 1), it is logical to leverage the spatial centroid of each Level 2 Administrative Area as a refugee origin or destination (for our purposes, a 'node' in the network). Another advantage to using such an organically generated network architecture as the simulation environment is the visualization potential available to the researcher when seeking to analyze the evolving social networks of the migrants.

3.3 Simulation and Model Logic

The simulation is seeded at the four most prominent open border crossings from Syria to Turkey in February 2019. An additional random number of refugee agents between 75 and 150 appear at these four border crossings at every time step and join the existing refugee population in the simulation. While all agents are activated uniformly, they move in accordance with the probabilities associated with the move chance mechanic described below.

To satisfy *factors that influence migration decisions*, both push and pull factors are coded distinctly within the model. Push factors are coded in the refugee *move chance*, a probability between 0 and 1 that represents a refugee's likelihood to move from one node to the next. Move chance is set to 1 for those in a location with one or more conflict events (derived from the ACLED database for February 2019), a random number between 0.2 and 0.5 for those in a location with one or more refugee camps (derived from UNHCR database for February 2019), and a random number between 0.3 and 0.7 for all others. Note that input datasets are not updated dynamically throughout the course of the simulation. [22]

Alternatively, pull factors, or the things that make a location desirable to a refugee per the theory of intervening opportunities, are encoded in the node weights of the simulation network architecture. These include normalized pre-existing refugee population, general directionality of movement normalized across all nodes based on distance from most ideal refugee destination (in this case, London, UK), and locations of friends (halved) and kin (raw count) as represented in the simulation as the social network of each refugee agent. In this way, the desirability of a potential move location is unique to each refugee agent. *Agent decisionmaking*, or the decision of the location to which an agent moves, is based on the conflation of these three factors into each district node making it more or less desirable to a refugee agent.

One-time step is equivalent to one day, and the model is run for 30 total days with an assumed initiation in February 2019. For validation purposes, data from the UNHCR situation map for May 2019 (30 days after model initiation) was used.

4 Results

The number of refugees in-country at model initiation is 3,619,171 and the number of refugees at the end of the simulation is probabilistically determined, in this case at 3,824,171. Multiple representative runs of the model yielded similar spatial distributions and refugee populations by district. (See Figure 1). Across all model runs, refugees remain clustered near the Syrian border, which is to be expected as a) this is the model seed location and b) this is the location of the majority of the official refugee camps. Refugees also cluster around Istanbul. It is interesting to note that refugees do not move much further northwest beyond Istanbul indicating that the weight of the general directionality of movement encoded within the node weights is appropriate. What is not evident, however, is whether the success of the spatial distribution of refugees is due to the tendency to cluster around pre-existing refugee populations. So as not to allow this single factor to dominate despite its prominence in numerous theories of forced migration, the pre-existing refugee population was normalized between 0 and 1 before its input into the candidate location desirability score.

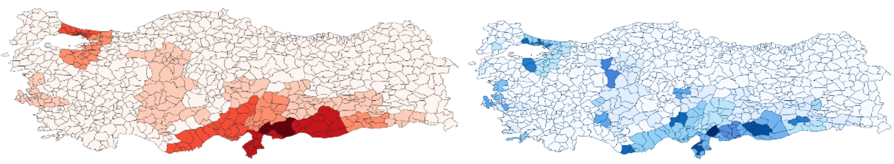

Fig. 1. Seeded simulation with refugee populations on model initiation (left). Ending refugee populations after 30 simulated days (right). Darker color indicates higher population.

4.1 Verification and Validation

Model verification was achieved through sensitivity testing of all model parameters to assess the effects of the various thresholds chosen. Additional verification was performed continuously throughout model development through spot checking of endogenously generated metrics (such as geocoordinates of district centroids or district-level refugee population) by writing out shapefiles to QGIS. During sensitivity testing, each parameter was set to min and max values, and the results validated, per the validation process described below, to assess the impact the selected factor had on model performance.

Model validation was performed using the known refugee populations in March 2019 as captured by the UNHCR 30 days after model initiation. Validation accuracy was captured by normalizing the actual refugee population in each district, normalizing the predicted refugee population in each district, and taking the absolute value of the difference. The mean and standard deviation of this normalized error, when averaged across the entire spatial extent, comprised the validation accuracy of the model as a whole (see Figure 2). The model validation accuracy achieved per the results presented above was 90% with a variance of 15%.

Fig. 2. Validation accuracy by district on a scale of 0 to 1. The darker the district shading, the greater the variance between simulation output and actual refugee population in March 2019.

5 Discussion

The model presented here is preliminary spatially-explicit, network-based ABM designed to model and predict refugee flows across geographic terrain in response to forced migration events. The utility of this model within the communities of international aid, public policy, and national security is evident. Beyond the specific case of Syrian refugees in Turkey, the model described above has been architected and automated in such a way that facilitates its application in other geographic locations and in response to other forced migration events. By using relevant data, setting the simulation duration, and tweaking some key model parameters, the hope is that the computational framework is extensible to other events and generalizable enough to be more broadly applied.

While sensitivity testing of existing model parameters was sufficient to establish internal validity, what remains outstanding is sensitivity testing of the push/pull factors that influence migration encoded within the model architecture itself. In addition to testing these factors, the inclusion of additional factors also remains to be explored. Fortunately, the model codebase is such that additional exogenous factors can easily be incorporated and mapped to the existing network architecture so long as the factors are derived from high-fidelity data sources with geospatial information.

It is clear that robust modeling and simulation efforts are greatly improving both analysis and prediction in refugee studies. There is still one very important piece that remains largely absent from current efforts owing largely to a data deficit that limits exploration. Through in-person interviews and surveys, it is well understood that refugees make use of a variety of social media platforms to identify border crossings, transit routes, means of transportation, and destination communities in host countries. [16] Official camps only account for a fraction of refugees in-country at any given time and social media is the primary means that refugees have for identifying alternate locations and communicating amongst themselves and those that provide aid once there. Popular social media and electronic communication platforms include WhatsApp, Facebook, and Twitter. [24] Given that these platforms are privately-owned, however, there is no way to access the communication data they contain to the level of fidelity required to do any real modeling. The addition that would most advance the work of modeling and simulation in refugee studies is access to this data (or robust means to collect it) and incorporation of it into extant models. [25] Within this model, the social media element can be proxied through the use of social networks that assume asynchronous communication across large geographic distances is possible. Further incorporation of refugee social media social networks, to include hubs with high measures of centrality, could provide additional insight into refugee communication networks and facilitate their inclusion into future modeling efforts. Of course, without access to this data, which is unlikely to come from the media companies themselves, a large gap will remain in the ability to truly include all factors that influence refugee movement into current models.

The ABM developed here provides a starting point and preliminary framework for modeling and predicting refugee flows resulting from forced migration events. As capabilities for modeling forced migration evolve and new data become available, both tradecraft and data alike can be iteratively incorporated into this framework to advance modeling and simulation efforts for refugee studies. Future research will focus on evolving this framework further.

References

1. Klabunde, A., & Willekens, F. (2016). Decision-making in agent-based models of migration: state of the art and challenges. *European Journal of Population*, *32*(1), 73-97.
2. Coats, D. (2019). The National Intelligence Strategy of the United States of America.
3. UNHCR Statistics: The World in Numbers. (2017). Retrieved May 11, 2019, from

http://popstats.unhcr.org/en/overview#_ga=2.118780737.939392455.1557574892-1504139322.1555790870

4. Richmond, A. H. (1993). Reactive migration: Sociological perspectives on refugee movements. Journal of Refugee Studies, 6(1), 7-24.
5. Castles, S. (2003). Towards a sociology of forced migration and social transformation. Sociology, 37(1), 13-34.
6. Adhikari, P. (2013). Conflict-Induced Displacement, Understanding the Causes of Flight. American Journal of Political Science, 57(1), 82-89.
7. Moore, W. H., & Shellman, S. M. (2006). Refugee or internally displaced person? To where should one flee? Comparative Political Studies, 39(5), 599-622.
8. Moore, W. H., & Shellman, S. M. (2007). Whither will they go? A global study of refugees' destinations, 1965–1995. International Studies Quarterly, 51(4), 811-834.
9. Epstein, G. S. (2008). Herd and network effects in migration decision-making. Journal of Ethnic and Migration Studies, 34(4), 567-583.
10. Tranos, E., Gheasi, M., & Nijkamp, P. (2015). International migration: a global complex network. Environment and Planning B: Planning and Design, 42(1), 4-22.
11. Sarra, A. L., & Del Signore, M. (2010). A dynamic origin-constrained spatial interaction model applied to Poland's inter-provincial migration. *Spatial Economic Analysis*, *5*(1), 29-41.
12. Disney, G., Wiśniowski, A., Forster, J. J., Smith, P. W., & Bijak, J. (2015). Evaluation of existing migration forecasting methods and models. *Report for the Migration Advisory Committee: Commissioned Research. ESRC Centre for Population Change, University of Southampton*.
13. Tobler, W. R. (1987). Experiments in migration mapping by computer. *The American Cartographer*, *14*(2), 155-163.
14. Cushman, S. A., McKelvey, K. S., Hayden, J., & Schwartz, M. K. (2006). Gene flow in complex landscapes: testing multiple hypotheses with causal modeling. *The American Naturalist*, *168*(4), 486-499.
15. Cushman, S. A., Chase, M., & Griffin, C. (2010). Mapping landscape resistance to identify corridors and barriers for elephant movement in southern Africa. In Spatial complexity, informatics, and wildlife conservation (pp. 349-367). Springer, Tokyo.
16. Cioffi-Revilla, C. (2005). A canonical theory of origins and development of social complexity. Journal of Mathematical Sociology, 29(2), 133-153.
17. Crooks, A., Castle, C., & Batty, M. (2008). Key challenges in agent-based modelling for geo-spatial simulation. Computers, Environment and Urban Systems, 32(6), 417-430.
18. Aral, S., Muchnik, L., & Sundararajan, A. (2009). Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. Proceedings of the National Academy of Sciences, 106(51), 21544-21549.

19. Richey, M. K. (2014). Geocultural Simulations: Populations and Predictions. In *Encyclopedia of US Intelligence-Two Volume Set* (pp. 1-13). Auerbach Publications.
20. Richey, M. K. (2015). From Crowds to Crystal Balls: Hybrid Analytic Methods for Anticipatory Intelligence. American Intelligence Journal, 32(1), 146-151.
21. Gilbert, N., & Troitzsch, K. G. (2005). System dynamics and world models. In *Simulation for the Social Scientist* (2nd ed., pp. 28-54). Glasgow: Bell & Bain.
22. Suleimenova, D., Bell, D., & Groen, D. (2017). A generalized simulation development approach for predicting refugee destinations. Scientific reports, 7(1), 13377.
23. Raleigh, C., Linke, A., Hegre, H., & Karlsen, J. (2010). Introducing ACLED: an armed conflict location and event dataset: special data feature. Journal of peace research, 47(5), 651-660.
24. Curry, T., Croitoru, A., Crooks, A., & Stefanidis, A. (2019). Exodus 2.0: crowdsourcing geographical and social trails of mass migration. *Journal of Geographical Systems*, *21*(1), 161-187.
25. Xu, Y., Holzer, A., Maitland, C., & Gillet, D. (2017, November). Community building with co-located social media: a field experiment with Syrian refugees. In *proceedings of the ninth international conference on information and communication technologies and development* (p. 16).