A Multi-Scale Approach to Data-Driven Mass Migration Analysis

Mohammed N. Ahmed[2], Gianni Barlacchi[1,4], Stefano Braghin[1], Francesco Calabrese*[5], Michele Ferretti[3], Vincent Lonij[1], Rahul Nair[1], Rana Novack[2], Jurij Paraszczak[5], and Andeep S. Toor[2]

[1] IBM Research - Ireland
[2] Global Business Services, IBM Corporation
[3] King's College London
[4] University of Trento
[5] New York University

Abstract. A system for scenario analysis and forecasting of mass migration is presented. The system consists of a family of multi-scale models to address the need of responding agencies for better situational awareness, short and medium-term forecasts of migration patterns, and assess impact of changes on the ground. Such insights allow for better planning and resource allocations to address migrant needs. The analytical framework consists of three separate models (a) a global push-pull model to estimate macro-patterns, (b) a time-series prediction model for estimating future boundary conditions of crisis regions, and (c) a detailed network flow model that models population diffusion within the crisis region and allows for scenario modeling. The paper presents the framework using the European refugee crisis as a case study. In addition, overall system design, practical considerations, end-user applications, and limitations of the modeling approach are discussed.

1 Introduction

65.3 million people in the world today are forcibly displaced[6]. There are a further 232 million migrants worldwide who live outside their country of origin[7]. With these growing numbers, governments, international organizations, NGO's and other stakeholders face an increasing challenge in responding to migration crisis, such as the one in Europe recently. If necessary data and tools were available to forecast displacement crisis, response actions can be better coordinated. In this paper, we present a data-driven approach to enable responders to manage mass migration events.

Migration is an inherently complex and uncertain process. Direct observations of migration patterns are typically partial and inaccurate. Paths and destinations for migration are influenced by a range of human factors. Information

* Currently at Vodafone, Italy
[6] http://www.unhcr.org/figures-at-a-glance.html
[7] http://www.unfpa.org/migration

sources for some factors may exist and come in different forms. For example, structured data from organizations such as registration counts at selected sites or refugee camps may be available. For locations without observers, one may need to rely on unstructured data such as news reports. Past data on migration movements and different data sources can be used to learn patterns of movements. Governments policies impacting migration patterns can change over the course of time. The types of changes and the response of migrants to such changes can be difficult to ascertain from past observations. For such cases, approaches that model the underlying processes are required. The need for hybrid approaches that merge purely data-driven (capture past patterns) and ones that model the physics (capture aggregate interactions and exogenous inputs) are necessary.

The current climate of global displacement has inspired a growing awareness that an innovation gap exists in addressing social issues within the humanitarian sector and that creative methods of partnership between the public and private sectors are needed. While corporate philanthropy programs and social entrepreneurship have undoubtedly made inroads in this regard, in order to more comprehensively design, develop, apply solutions, and leverage the vast skill sets in big data and analytics, we must better incentivize competition and thereby critical thinking among the private sector.

The challenges in doing so are surmountable, but not trivial. We must overcome obstacles on both sides that are cultural and come to a mutual understanding that a symbiotic relationship between the humanitarian sector and revenue-driven organizations can not only exist, but will have an exponentially greater impact on social issues than when such projects are treated exclusively as philanthropic.

This research addresses the following questions. What information signals on mass migration are available? What analytical framework and models help in developing forecasts of mass migration? What scenarios are likely to occur and what is their impact on migration flows? The subsequent sections present the models that are designed to (a) enhance situational awareness using multiple data sources, (b) provide short and medium-term forecasts on migration patterns to aid operational decision making, and (c) enable 'what-if' scenario analysis for agencies looking to study the impact of exogenous factors.

The model development is presented using the European migration crisis as a case study. The choice is motivated by availability of different datasets, for example the large volume of news reports on the crisis and detailed registration data at various sites. The dynamics of the crisis were notably complex, as migrants were crossing several international boundaries. The patterns of flight shifted as conditions across borders changed.

The forecasting and analysis tool for mass migration addresses the gap that the humanitarian sector has related to forecasting and scenario analysis. The developed models are presented in Section 4 with the system that implements these models is described in Section 5. The tool is aimed at agencies, governments, and NGOs that respond to crisis through the different states of refugee and migration movements.

2 Related Work

Models dealing with the spatial movement of people have sparked the interest of researchers for more than a hundred years, since Ravenstein's "Laws of Migration" [15]. The 20^{th} century saw a rise of the more quantitatively oriented models, among which Zipf's early intuition of spatial interactions as shaped by population size and distance (an analogy with the Gravitation Law), based on a previous formulation by Gaspard Monge [17]. More recent approaches have considered tools from econometrics and demography [1]. Improvements in computing technologies allowed for better algorithms to be applied in this field [6, 8]. Several examples of such developments can be found in the literature [23, 2, 14, 22, 4, 9, 5].

Current trends in quantitative mobility models have been influenced in part by (a) the availability of personal digital traces such as mobile phone data, and (b) new information signals, such as satellite imagery and crowd-sourced initiatives.

Prediction of human mobility has benefited from the large amount of personal digital traces. Various methods have been proposed in the literature, such as entropy rate of sequence of locations [18], travel distance functions [11], and stochastic processes [17]. Intuitively, such methods may appear to be suited to 'regular' mobility, such as those describing data-to-day movements at city-scale. However, predictability of movements have been shown to persist even in the wake of disasters. For example, Xin Lu et al. [13] analyzed movement of 1.9 million mobile phone users before and after the 2010 Haiti earthquake and concluded that mobility is correlated to social ties that existed before the earthquake. Tools built on digital traces for humanitarian applications have found to be an 'operationally valuable platform' [24].

Additional information signals that have been leveraged to understand migration recently include satellite imagery and crowd-sourcing efforts. Image classification routines on satellite images have been used to quantify size and growth of refugee camps. Examples include techniques using feature detection and extraction methodologies [21, 19, 12, 10] and deep learning and pattern recognition [16, 25]. Such techniques have been advocated for [3, 20] as a promising path to map temporary settlements and refugee camps.

Crowd-sourced information has been critical in providing better situational awareness for humanitarian agencies and NGOs. OpenStreetMap (OSM) provides an ecosystem of tools, events and volunteers that enable crowd-sourced gathering of information before and during crisis with focused initiatives such as a Humanitarian OSM Team[8]. These efforts are vital for relief efforts, estimation of refugee statistics, and community involvement.

With the European refugee crisis in 2015, there have been efforts aiming to fuse different data sources and provide migration awareness. The British Red

[8] https://hotosm.org/

Cross[9] provide a dashboard to monitor and forecast migration flows using daily arrivals information, border statuses, and news reports.

To address needs to end-user agencies, this work builds on these previous efforts by taking a multi-scale (global and local) approach that allows for scenario analysis. The multi-scale approach allows inclusion of factors at the global scale, while preserving detailed factors within the crisis region. Modeling scenarios allows for analysts to account for exogenous factors (e.g. repercussions of potential international policies).

3 Information Signals

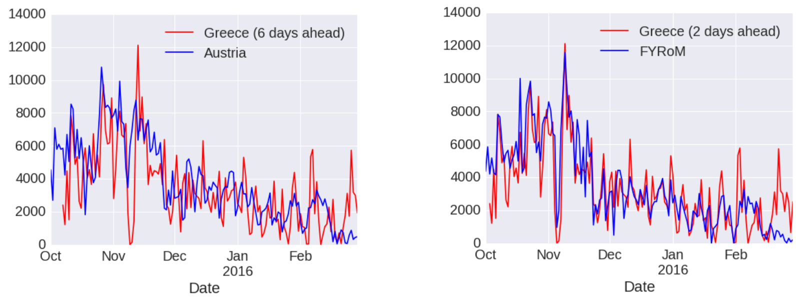
Complex processes like migration depend on a wide range of factors. Data sources that describe the socio-demographics, overall context, conflict conditions (in the case of forced displacements), and changing conditions over a period of time are all determinants of migrations and paths of flight. There is no single functional relationship between these variables, so its essential to understand key information signals that can add value to data-driven modeling. Using the European crisis as a case, some common information sources are described. For multi-scale approaches, the data sources related to both operational (e.g. daily) and strategic (e.g. annual) characteristics.

Migrant Registration Data Registration data on migrant arrivals at key sites, such as the Greek islands provides daily, weekly or monthly arrival rates at various points along the crisis region. The UNHCR, UN's Refugee Agency, provides open data that is compiled from several sources[10]. The time series of arrivals at each of the sites can be used for several features in forecasting (see Section 4.1). Additionally, computing the cross-correlation function between the registration, provides an estimate of transit times through the network (used in Section 4.2). Figure 1 shows examples of time series shift that give estimates of transit times between select countries.

Weather Data Information on weather conditions and seasonal factors influence movements. Adverse conditions are likely to restrict possible migration paths. For sea faring refugees arriving in the Greek islands, factors such as wind speed were found to be negatively correlated with arrivals over the winter months. Figure 2 shows the Pearson correlation coefficient between arrivals in different countries and different weather-related variables. Weather data is sourced from the Weather Underground[11] service.

News Data In order to capture information about policy changes (e.g. close of the Hungary border) and other external events, we used the news data provided by the GDELT Project [12]. It allows users to monitor the web news around

[9] http://brcmapsteam.github.io/europe-15-situational-awareness/
[10] http://data.unhcr.org/mediterranean/regional.php
[11] https://www.wunderground.com
[12] http://gdeltproject.org/



(a) Arrivals in Greece (6 days ahead) and Austria

(b) Arrivals in Greece (2 days ahead) and FYRoM

Fig. 1: Estimating transit time using a cross-correlation function

the world and extract valuable information from text such as entities (e.g people locations, organizations), counts, emotions and events in over 100 different languages. Documents are annotated applying a combination of state-of-the-art natural language processing techniques. Moreover, it is worth to mention that for each article Global Knowledge Graph (GKG) provides a variable called *locations* in which they report the places mentioned in the article. This is an important information since it allow us to localize the potential country of the news. Bilateral similarity scores, as shown in Figure 3, are used to train an affinity function (Section 4.3).

Country-level data Aggregate statistics at the country level were collected from sources such as the World Bank. Socio-economic indicators such as GDP, population, conflict status were considered. Datasets from the gravity model literature[13] where included for dyadic features such as common languages, historical ties, trade volumes, and geographic distances.

4 Analysis Framework

Consider a crisis region of interest where dynamics of migration need to be evaluated. In the European crisis, this includes the regions around the Mediterranean Sea where paths of flight are concentrated. In addition, the region of interest should include potential transit paths that may be considered by migrants. The region could also include sources of forced displacement or intended destinations, e.g. Syria or Germany respectively.

The analysis framework consists of three distinct models that jointly provide forecasts. The main crisis region is modeled as a network flow model (Section 4.2). The representation is detailed and the models the rate of movements from one node to another. The boundary conditions are handled by the other two

[13] http://www.cepii.fr/

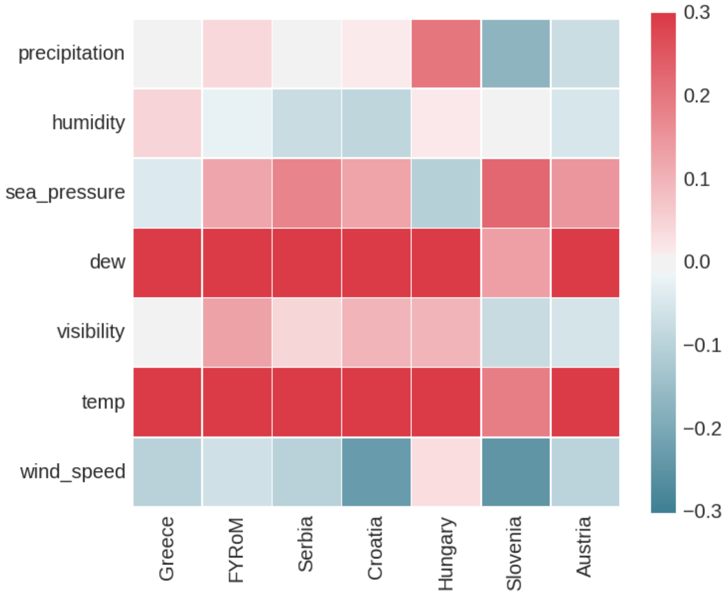


Fig. 2: Pearson correlation coefficient between refugee/migrant arrivals and weather conditions for each of the involved countries.

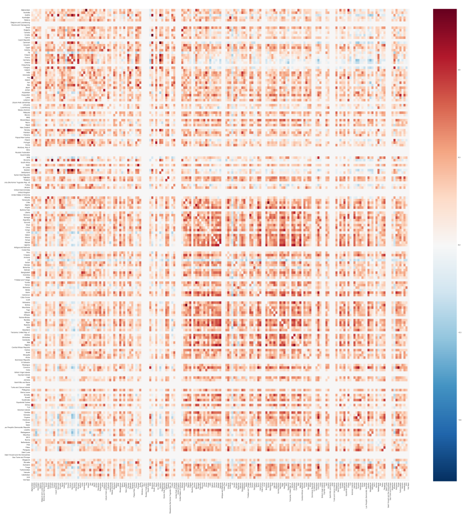


Fig. 3: Country-level bilateral similarity scores based on word embeddings (word2vec)

models. An arrivals model (Section 4.1) is a time series forecasting model that predicts daily arrival rates at nodes on the edge of the crisis region. To determine exit rates and consider intended destinations, a macro push-pull model (Section 4.3) is used to estimate fraction of migrants for each likely destination. Figure 4 shows this framework.

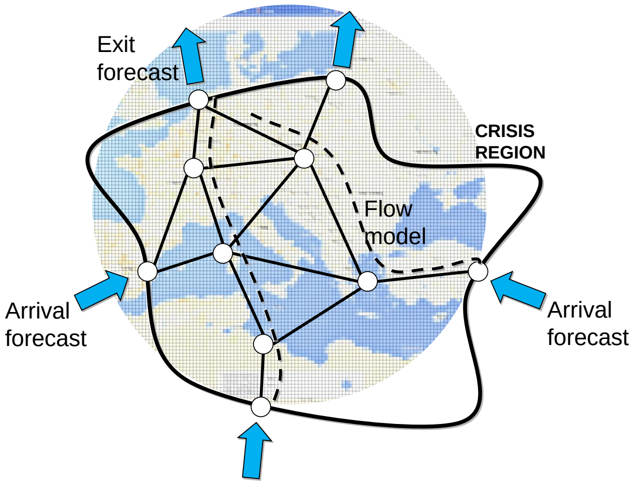


Fig. 4: Analytical framework showing delineation of a crisis region of interest. Arrival forecasts at boundary nodes are estimated using arrival prediction models (Section 4.1). Exit nodes forecasts are estimated from macro-migration model (Section 4.3). A detailed network flow model (Section 4.2) models the population movement within the crisis region.

4.1 Arrival Prediction Model

The aim of this model is to predict the number of arrivals at the entry-points of the crisis region. Future arrivals are dependent on external variables such as weather, as well trends due to (de-) escalation of the crisis in question. To capture these effects we used data from weather reports and news reports and experimented with multiple machine learning models to predict day-ahead arrivals of refugees to the Greek Islands. Around 180 features were extracted from the following datasets:

Autoregressive Features (AF) Previously observed value describe the trend of arrivals until the time of observation. In the model we use as a feature the number of arrivals one day before as well as rolling statistic such as mean, variance, maximum and minimum computed over the previous week.

Weather Features (WF) We used six different weather variables: wind speed, temperature, sea level visibility, dew point, sea level pressure, humidity, and precipitation. To generate features, we aggregated the data computing the min, max, sum and mean of each variable over the day.

News Features (NF) To obtain news features, we queried the GKG historical data using the GKG Exporter [14]. We filtered the news selecting those in which the word *refugee* and either *migration* or *border* appear. Then, we selected only articles that match the location field with one of the countries involved in the migration crisis. Starting from these articles, three types of features are computed: (i) number of occurrences of relevant countries name (e.g. Syria and Greece) in the news, (ii) number of occurrences of relevant words related to humanitarian crisis (e.g. refugee camp) and, (iii) number of articles about humanitarian themes (e.g. migration). The former has been computed by applying a keyword-based approach where an article is assigned to a theme considering the presence of predefined words like *border* or *humanitarian*.

We applied four different machine learning models: LASSO, Linear Regression (LR), Ridge Regression (RR) and Gradient Boosting of Regression Trees (GBRT). In order to reduce the trend in the signals, we trained our model on the number of arrivals minus the rolling mean of the arrivals in the previous seven days. We added this value to the prediction before computing error metrics. The hyper-parameters of each of the models were determined through cross-validation; training the models on data between October 2015 and January 2016 and testing it on February 2016. Moreover, we normalized the feature values subtracting the mean and dividing by the variance. In Table 1 the results of the tested models are reported. Results of a persistence model prediction are also shown as a baseline. This model predicts arrivals each day to be equal to arrivals the previous day.

Due to the high number of features, we experimented with a feature selection step. To select features we first fitted one of the regression models. Then we ranked the features by the magnitude of their weight in the model (i.e. the slope). We used this approach on the results of LASSO, Linear Regression, and Ridge Regression. We also tried recursive feature elimination (RFE), and simple correlation.

We evaluated the quality of the feature selection by again through cross validation, training with the reduced set of features on the training set. RFE produced the worst set of features. The other methods produced similar sets of features with similar performance.

The smaller set of features lead to better model performance. After feature selection, each of the models shows improved performance, but the difference between the models is small. The best performance are obtained using the LASSO model on a subset of 10 features that included weather, autoregressive and news data. In particular, from the weather features we used the forecasting mean and max for wind speed, the forecasting gust speed and the forecasting mean for hu-

[14] http://analysis.gdeltproject.org/module-gkg-exporter.html

midity, temperature and sea level pressure. Regarding autoregressive features, we used the difference in number of arrivals between one and two days before, the mean and the minimum in number of arrivals for last three days, the number of arrivals one day before and the sign of the derivative in the last two days. Finally, the model also includes the number of news (for the arrival country) about international organization for migration. Table 1 shows the results.

	No Feature Selection		Feature Selection	
Models	RMSE	Error Reduction(%)	RMSE	Error Reduction(%)
Baseline	1429.075	-	-	-
Linear Regression	$> 10^9$	$>> 100$	1111.214	22.24
Lasso	1197.096	16.23	1110.503	22.30
GBRT	1266.349	11.38	1257.430	12.01
Ridge Regression	1258.782	11.91	1116.311	21.88

Table 1: Arrivals model results for February in the Greek islands.

For all of the models, the weather features are the most present, followed by the historical data. This could be due to the fact that we tested our model on arrivals in the Greek Islands where weather conditions influence arrivals. Figure 5 shows a sample of forecasts for the Greek islands. There is not a significant difference in terms of results between the linear models after the feature selection step.

4.2 Network Flow Model

The challenge in modeling the paths of refugees through the crisis region is two-fold. First, measured arrivals at different location along a path through the crisis region are known to be inaccurate. Second, the movement patterns can change due to changing conditions on the ground, such as border closures. In this section we present a network flow model that can address both issues.

To mitigate the effect of inaccurate measurements, we developed a model that can impose some common sense boundary conditions on the prediction. Specifically, arrivals in each country must correspond to departures in a different country. Departures from any country cannot exceed the total number of refugees present in that country.

Our model represents locations along the path of movement as nodes on a graph. The edges that connect the nodes i and j represent the likelihood that refugees will travel from node i to node j. As an example we will model travel of refugees from Greece to Austria through FYROM, Slovenia, Hungary, Croatia, and Serbia.

The number of people traveling from node i to node j at time t, denoted by $F_{ij}(t)$, is given by

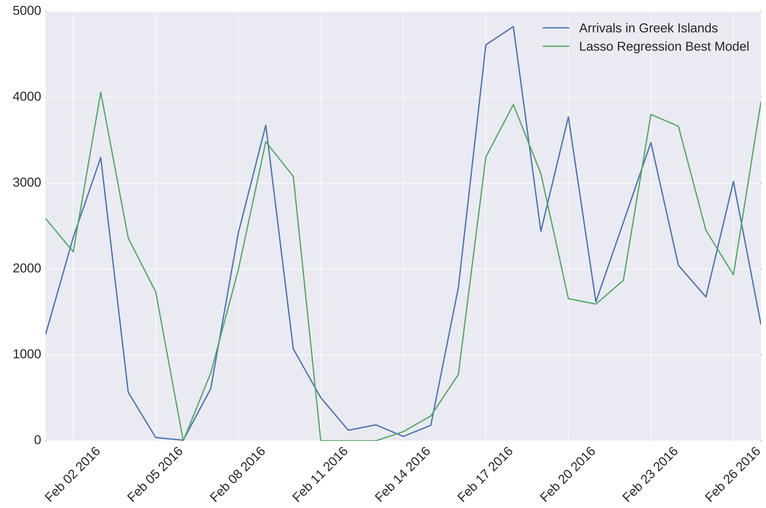


Fig. 5: Models predictions in February in the Greek islands.

$$F_{ij}(t) = P_i(t) f_{ij} \tag{1}$$

where $P_i(t)$ is the number of people present at node i, and f_{ij} is a constant that we will determine from historical arrivals data. The f_{ij} can be interpreted as split-fractions for the departures of each node. The net flow from node i to node j is then given by

$$N_{ij}(t) = P_i(t) f_{ij} - P_j(t) f_{ji} \tag{2}$$

To simplify the problem, we consider only these net flows. Arrivals ($A_i(t)$), and departures ($D_i(t)$) at node i are given by

$$A_i(t) = \sum_j \max(0, N_{ji}(t)) \tag{3}$$

$$D_i(t) = \sum_j \max(0, N_{ij}(t)) \tag{4}$$

The populations of the next time step can then be calculated as

$$P_i(t+1) = P_i(t) + A_i(t) - D_i(t) + E_i(t) \tag{5}$$

where E_i are exogenous arrivals, to be specified independently. Equations 1-5 can then be used iteratively to compute future flows.

In the present case we use E_i to specify arrivals to Greece from Turkey and departures from Austria to the rest of Europe. For arrivals to Greece, we use the outputs of the arrival prediction model discussed in the previous section. For departures from Austria, we assume all people present in the country, depart each day.

To determine the values f_{ij}, we optimized the following loss function.

$$L = \sum_{it}(A_i(t) - M_i(t))^2 + \alpha \sum_{ij} |f_{ij}| \tag{6}$$

where M_i are the measured arrivals at node i, and α is a regularization parameter. The second term in this equation is an L1 regularization. This regularization imposes sparsity in the possible paths.

The model is trained at the beginning of each week on the preceding 30 days. Forecasts are then generated for the subsequent 2 weeks. This means that for each day, there are two forecast values; the mean of those two values is shown in Figures 6 and 7.

Changing conditions on the ground can be defined through adjustment of the exogenous arrivals. For example, starting mid February there were several policy changes that reduced the number of arrivals in Greece. This impacted arrivals in the other nodes of the network as well. We modeled this change by manually setting (exogenous) arrivals in Greece to 0 after February 16. Figures 6 and 7 show both scenarios compared to measured arrivals. Scenario 1 assumes no change to the arrivals in Greece, Scenario 2 assumes no more arrivals after February 16. This method can be used to make more accurate predictions if policy changes are known. In addition, this methods could be used to do counterfactual scenario analysis.

4.3 A push-pull macro migration model

To derive destination preferences, a global model of migration that seeks to determine bilateral flows between countries is considered. The model seeks to estimate the fraction of refugees from each country that are likely to migrate to any other country. Such an approach is necessary to estimate macro-level movement patterns and serves to project intended destinations of migrants within the detailed network flow model.

The model assumes there are *push* ("repulsion") factors at a home country along with *pull* ("enticing") factors at the destination. Push factors at origin and pull factors at destination cause migration. Movement is also a function of distance/affinity between two countries. Countries with higher affinity are likely have more migration.

Affinity metrics can be defined in several ways. Classical approaches are using spatial properties such as geographic distance. In this work, we have considered an affinity function trained on past migration data and considering distance metrics, exogenous variables such as Gross Domestic Product (GDP), colonial relationships, commonality of language, contiguity, and conflict status of origin countries. One endogenous variable in bilateral migration in previous years (if available) was included to model 'social pull' factors. News articles were also considered to include more current reports of migration.

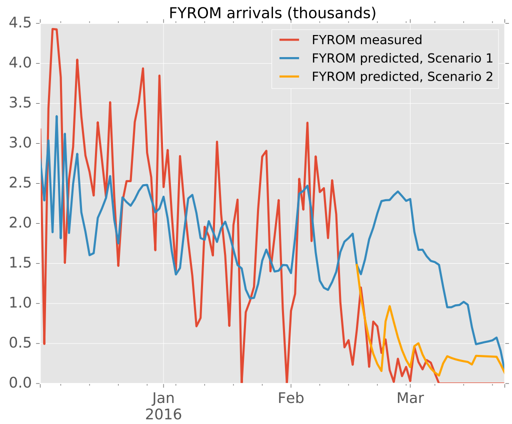


Fig. 6: Output from network flow model - Predicted arrivals in the Former Yugoslav Republic of Macedonia (FYROM). The figure shows the measured arrivals as well as predicted arrivals in two different scenarios. Scenario 1 assumes the "status quo", that is, past data is representative of future movements. Scenario 2 assumes no more arrivals to Greece after February 16. Scenario 2 is more representative of the changed circumstances after several border closures. Forecasts are generated at the beginning of each week, for the subsequent two weeks.

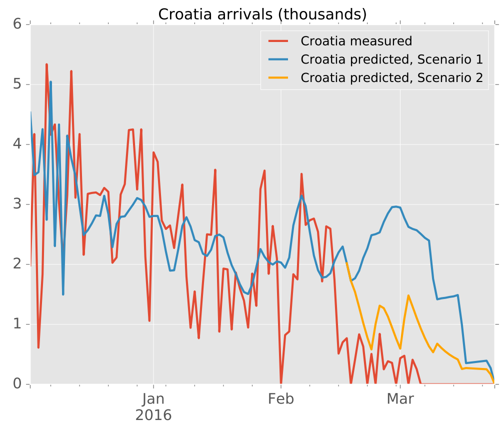


Fig. 7: Output from network flow model - Predicted arrivals in Croatia. The figure shows the measured arrivals as well as predicted arrivals in two different scenarios. Scenario 1 assumes the "status quo", that is, past data is representative of future movements. Scenario 2 assumes no more arrivals to Greece after February 16. Scenario 2 is more representative of the changed circumstances after several border closures. Forecasts are generated at the beginning of each week, for the subsequent two weeks.

Since the true affinity measure is not known, a log-transformed linear model with these features is fitted on past bilateral migration and the set of features described above [5]. The coefficients from this model are used to estimate the affinity metric.

More formally, we use the formulation presented in [7]. Let N be a set of countries. Given a vector of inflows $I_j \, \forall j \in N$, a vector of outflows $O_i \, \forall i \in N$, and a matrix of distances/affinity metrics d_{ij} for each pair of sites, the following quadratic program seeks to estimate bilateral flows between a set of sites N (where internal migration is ignored).

$$\min \sum_{i \in N} \sum_{j \in N} M_{ij}^2 d_{ij} \tag{7}$$

subject to:

$$\sum_{j \in N} M_{ij} = O_i \qquad \forall i \in N \tag{8}$$

$$\sum_{i \in N} M_{ij} = I_j \qquad \forall j \in N \tag{9}$$

The model computes M_{ij}, the migration flows between sites i and j. The numerical push-pull estimates can be ascertained by solving a linear system of equations, once all outflows and inflows are known.

$$\sum_{j \in S, j \neq i} M_{ij} = R_i \sum_{j \in S, j \neq i} \frac{1}{d_{ij}} + \sum_{j \in S, j \neq i} \frac{E_j}{d_{ij}} = O_i \, \forall i \tag{10}$$

$$\sum_{i \in S, i \neq j} M_{ij} = \sum_{i \in S, i \neq j} \frac{R_i}{d_{ij}} + E_j \sum_{i \in S, i \neq j} \frac{1}{d_{ij}} = I_j \, \forall j, \tag{11}$$

where R_i is the 'repulsion' (push) factor and E_j are estimates of the 'enticing' (pull) factor. For distance-based affinity functions, the units associated with the push-pull quantities can be interpreted as 'person-kilometers'. However, for more complex affinity functions, the values are relative and cannot be interpreted directly.

We have tested 4 cases for the affinity function. The first case considers geographic distances only. Case 2 additionally considers exogenous factors such as GDPs, common languages, contiguity, conflict status. Case 3 considers additionally "social pull" proxies, such as historical migration. Case 4 includes similarity measures based on news sources (see Section 3). Forecast errors were measured for years when the bilateral flows were known (currently 2013). The resulting split fractions serve as input to the flow model.

5 Description of Prototype

The models and tools are currently being deployed in an instance of IBM Bluemix[15]. A limited release is planned with a suitable agency for testing and validation.

[15] http://www.ibm.com/cloud-computing/bluemix/

Case	Features	RMSE
Case 1	Distance	11882.58
Case 2	Distance + exogenous variables	9494.30
Case 3	Distance + exo. vars + past migration	814.21
Case 4	Distance + exo. vars + past migration + news	814.21

Table 2: RMSE for different cases in persons per year (2013)

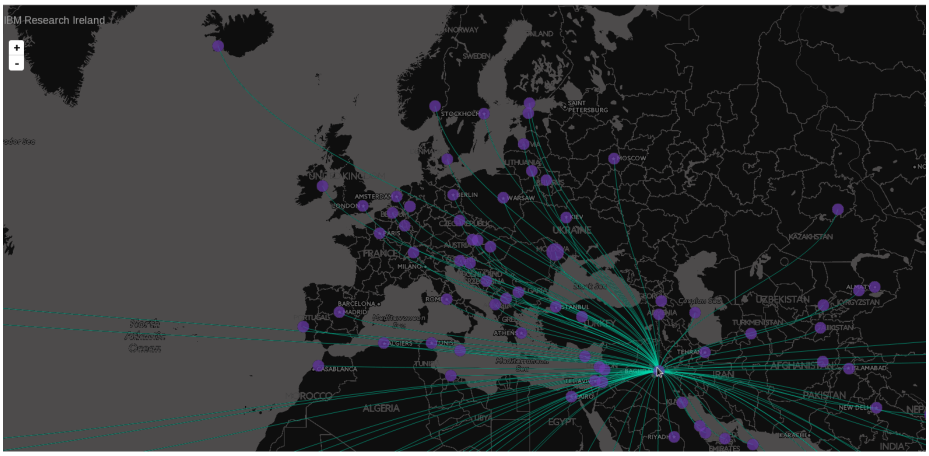


Fig. 8: Prototype of the global flow model

The prototype follows a client-server architecture in a scalable and efficient manner. The backend is implemented as a REST service and exposes the functionality of the models described in Section 4. The frontend application serves to allow for user-specified scenario modeling and displaying model results. The functionality exposed by the REST services is modular, so the general purpose functions can be consumed and by multiple applications. The system can therefore integrate easily with existing applications used by NGOs across the globe.

Methods associated with the push-pull model (Section 4.3) exposes only a training functionality. The computed model represents the actual migration flows between the observed countries, is presented to the user via the User Interface (UI). The predicted affinity function can be periodically updated to reflect the political, social and economic changes.

Methods associated with the network flow and arrivals prediction models (Sections 4.1 and 4.2 respectively) are for training and forecasting. The training methods returns a serialization of the computed model which can be used, by the prediction method, to produce a forecast of arrivals or arrivals distribution in the network, respectively.

Figure 8 shows the current version of the user interface for the prototype. It allows the user to travel across different spatial and temporal scales, as well as to fluidly integrate additional external resources. Using this interface an operator can visualize the flows of migrants at different granularity. Similarly, the operator can interact with the models to alter assumptions and change model parameters.

The UI visualizes the predicted arrivals of migrants in the various nodes of the network. An NGO is then able to visually assess the criticality level for each area of a given crisis region, and to better orchestrate on-the-ground operations.

6 Discussion

Data-aware tools, processes and methods have the potential to improve operations in the humanitarian sector. Forecasting and scenario modeling tools, such as those presented here, aid agencies to move to more proactive operations. There are several challenges to be addressed for wider uptake. The classical philanthropic engagement model with the private sector needs to shift to a more collaborative approach, where varied expertise across organizations can be tapped and models and methods refined.

The work presented has the following limitations. Since each migration crisis is unique, impact of some features used to model the process will be different across different contexts. Past observed factors may have little or no role in future crises. For example, wind speeds used to forecast migration arrivals in Greece will not yield useful information for land-based migration.

Partly relying on physical models and enabling scenario analysis helps mitigate some of these limitations. However, such models are based on assumptions that may also be specific to a particular context. Ensuring that the right assumptions are considered and appropriated incorporated within the models is key.

Precise measurements, and in turn forecasts, remain a challenge on the ground and for models. While relative measures provide indications on how resources could be potentially deployed, the absolute numbers may be critical for some applications.

References

1. Aleshkovski, I., Iontsev, V.: Mathematical Models of Migration. In: Livchits, V.N., Tokarev, V.V. (eds.) Systems analysis and modeling of integrated world systems, vol. II. Encyclopedia of Life Support Systems (EOLSS) (2002)
2. Batty, M., Mackie, S.: The calibration of gravity, entropy, and related models of spatial interaction. Environment and Planning A 4(2), 205–233 (1972), http://epn.sagepub.com/content/4/2/205.abstract
3. Bjorgo, E.: Using very high spatial resolution multispectral satellite sensor imagery to monitor refugee camps. International Journal of Remote Sensing 21(3), 611–616 (2000), http://www.tandfonline.com/doi/abs/10.1080/014311600210786$\backslash$nhttp://dx.doi.org/10.1080/014311600210786
4. Boyle, P.J., Flowerdew, R., Shen, J.: Modelling inter-ward migration in Hereford and Worcester: the importance of housing growth and tenure. Regional studies 32(2), 113–32 (1998), http://www.ncbi.nlm.nih.gov/pubmed/12293518

5. Cohen, J.E., Roig, M., Reuman, D.C., GoGwilt, C.: International migration beyond gravity: A statistical model for use in population projections. Proceedings of the National Academy of Sciences 105(40), 15269–15274 (2008)
6. Dennett, A., Wilson, A.: A multilevel spatial interaction modelling framework for estimating interregional migration in Europe. Environment and Planning A 45(6), 1491–1507 (2013)
7. Dorigo, G., Tobler, W.: Push-pull migration laws. Annals of the Association of American Geographers 73(1), 1–17 (1983)
8. Edwards, S.: Computational tools in predicting and assessing forced migration. Journal of Refugee Studies 21(3), 347–359 (2008)
9. Flowerdew, R., Aitkin, M.: A method of fitting the Gravity Model based on the Poisson distribusion. Journal of Regional Science 22(2), 191–202 (1982), http://dx.doi.org/10.1111/j.1467-9787.1982.tb00744.x
10. Giada, S., De Groeve, T., Ehrlich, D., Soille, P.: Information extraction from very high resolution satellite imagery over Lukole refugee camp, Tanzania. International Journal of Remote Sensing 24(22), 4251–4266 (2003), http://www.tandfonline.com/doi/abs/10.1080/0143116021000035021
11. González, M.C., Hidalgo, C.A., Barabási, A.L.: Understanding individual human mobility patterns. Nature 453(7196), 779–782 (2008), http://www.nature.com/nature/journal/v453/n7196/full/nature06958.html
12. Laneve, G., Santilli, G., Lingenfelder, I.: Development of automatic techniques for refugee camps monitoring using very high spatial resolution (VHSR) satellite imagery. In: 2006 IEEE International Symposium on Geoscience and Remote Sensing. pp. 841 – 845. IEEE, Denver, CO (2006)
13. Lu, X., Bengtsson, L., Holme, P.: Predictability of population displacement after the 2010 Haiti earthquake. Proceedings of the National Academy of Sciences of the United States of America 109(29), 11576–81 (2012), http://www.pnas.org/content/109/29/11576
14. Openshaw, S.: Neural network, genetic, and fuzzy logic models of spatial interaction. Environment and Planning A 30(10), 1857–1872 (1998)
15. Ravenstein, E.: On the Laws of Migration. The effects of brief mindfulness intervention on acute pain experience: An examination of individual difference 48(2), 167–235 (1885)
16. Roemheld, L.: Humanitarian Mapping with Deep Learning (2010)
17. Simini, F., González, M.C., Maritan, A., Barabási, A.L.: A universal model for mobility and migration patterns. Nature 484(7392), 96–100 (2012), http://dx.doi.org/10.1038/nature10856$\backslash$npapers3://publication/doi/10.1038/nature10856
18. Song, C., Qu, Z., Blumm, N., Barabási, A.L.: Limits of predictability in human mobility. Science 327(5968), 1018–1021 (2010)
19. Tiede, D., Lang, S., Hölbling, D., Füreder, P.: Transferability of obia rulesets for idp camp analysis in darfur. Geobia 2006 (2010), http://geobia.ugent.be/proceedings/papersproceedings/Tiede{_}137{_}TransferabilityofOBIARulesetsforIDPCampAnalysisinDarfur.pdf
20. UNHCR: Refugees, Asylum-Seekers, Returnees, Internally Displaced and Stateless Persons. Tech. rep., UNHCR (2009), papers2://publication/uuid/F9B96CAD-C064-46AA-A23A-DF5335716633
21. Wang, S., So, E., Smith, P.: Detecting tents to estimate the displaced populations for post-disaster relief using high resolution satellite imagery. International Journal of Applied Earth Observation and Geoinformation 36, 87–93 (2015)

22. Willekens, F.: Monitoring international migration flows in Europe - Towards a statistical data base combining data from different sources. European Journal of Population 10(1), 1–42 (1994)
23. Wilson, A.: Entropy in urban and regional modelling. Pion Ltd, London, UK, monographs edn. (1970)
24. Wilson, R., zu Erbach-Schoenberg, E., Albert, M., Power, D., Tudge, S., Gonzalez, M., Guthrie, S., Chamberlain, H., Brooks, C., Hughes, C., Pitonakova, L., Buckee, C., Lu, X., Wetter, E., Tatem, A., Bengtsson, L.: Rapid and Near Real Time Assessments of Population Displacement Using Mobile Phone Data Following Disasters : The 2015 Nepal Earthquake. PLoS Currents (1), 1–26 (2016)
25. Xie, M., Jean, N., Burke, M., Lobell, D., Ermon, S.: Transfer Learning from Deep Features for Remote Sensing and Poverty Mapping. arXiv.org preprint p. 16 (2015), http://arxiv.org/abs/1510.00098