Project Jetson: Predicting Forced Displacement in Somalia

Rebeca Moreno Jiménez[1], Sofia Kyriazi[1], and Katherine Hoffmann Pham[2,3]

[1] UNHCR Innovation [2] NYU Stern School of Business [3] United Nations Global Pulse

Motivation

Predicting arrivals of refugees and Internally Displaced Persons (IDPs) is of critical interest in humanitarian emergencies. Field operations teams are interested in anticipating these arrivals in order to **provide adequate assistance** in the form of food, shelter, and protection-related services.

Objectives

We identify three types of information needs in humanitarian emergencies: (1) description, (2) prediction, and (3) simulation. We focus on **predicting** arrivals by region 1 and 3 months in advance.

Context

- **70.8m** people displaced worldwide (58% IDP)
- **2.6m** IDPs in Somalia alone (21% of population)

Sources: UNHCR 2019, OCHA 2019

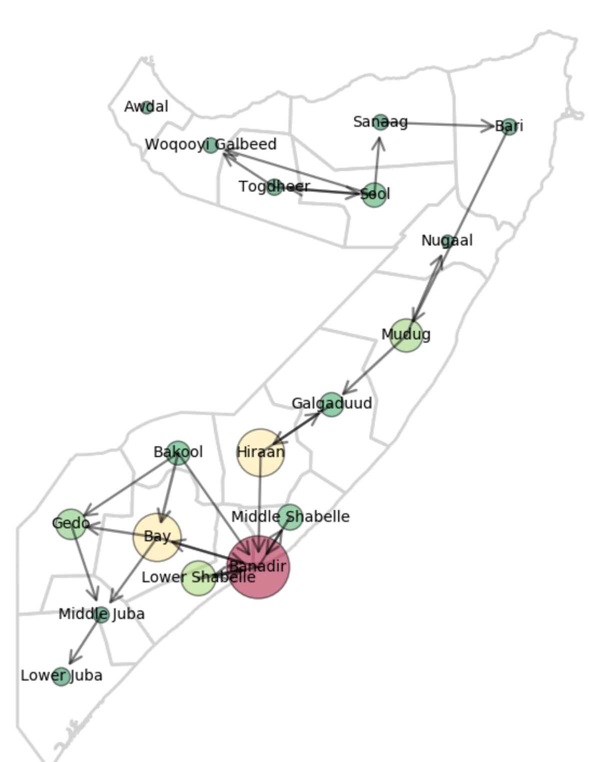

Figure 1: Inflows averaging at least 100 people per month

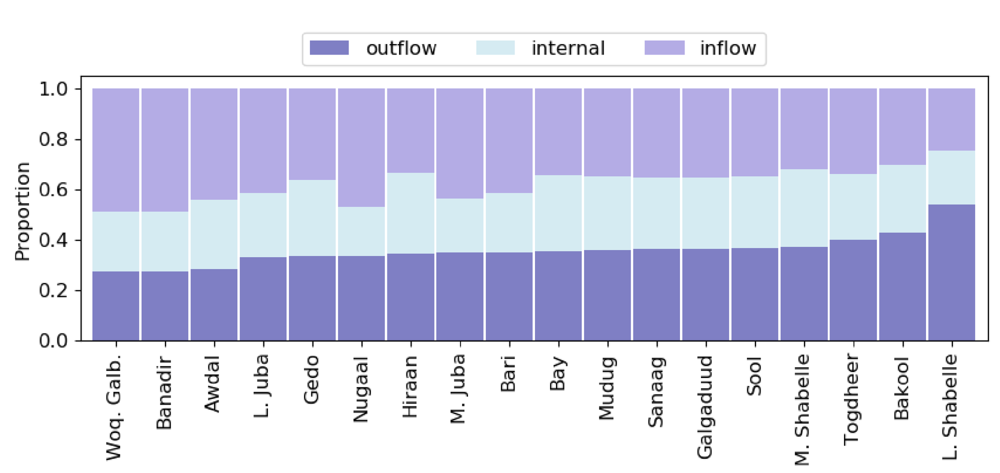

Figure 2: Inflows, outflows, and within-region displacement

Research Questions

- **How best to forecast** displacement across 18 regions of Somalia using historical data?
- **What level of performance** is achievable?
- **How does performance vary** w/ forecast horizon, # features, # historical observations?

Data

We aggregate the data at the level of region and month. We observe 107 months' worth of data on 18 regions, covering the period from January 2011 through November 2019. We collect a diverse set of political, economic, environmental, health, and geographic features. Sources: UNHCR, ACLED, FSNAU, OSM

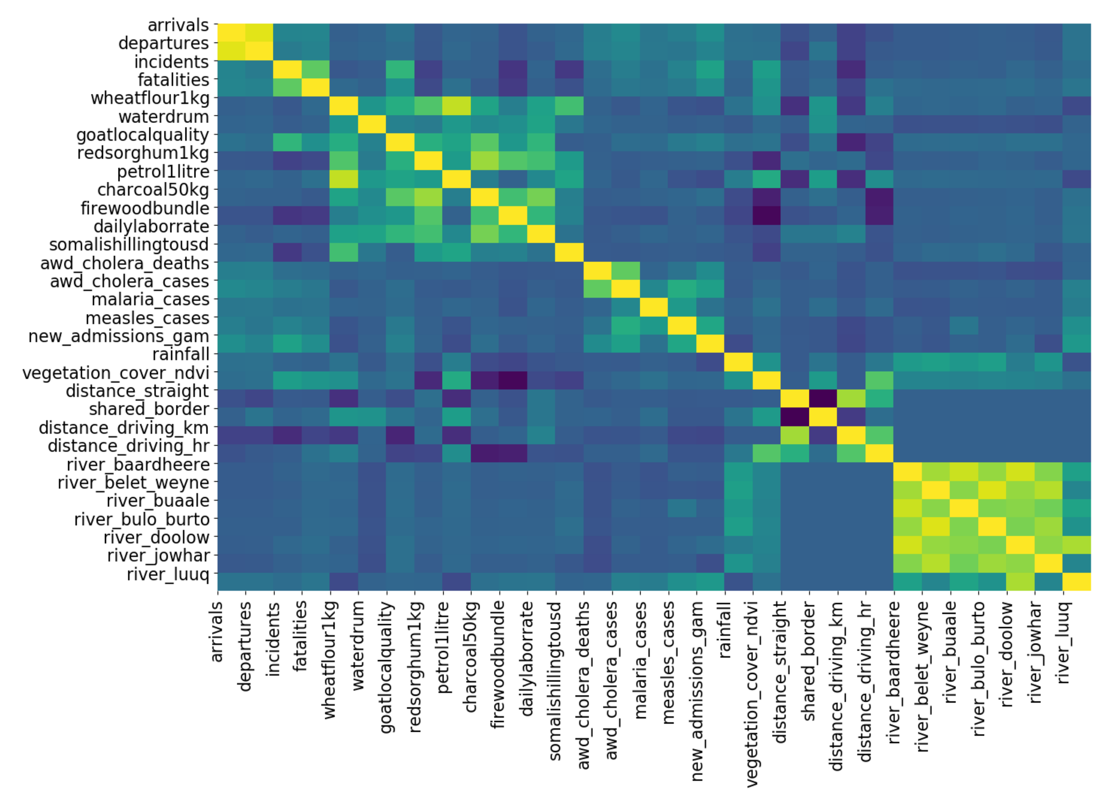

Figure 3: Correlations between variables in the dataset

Practical Challenges

- Integrating/updating data from **various sources**
- Target variable has missing values (57% in Sanaag)
- Average displacement varies by region (∼600-13,000)
- Variability is **growing over time**
- Recording of arrivals can be **unreliable**

Models

Data through December 2018 is used for training, and data from 2019 forms the test set. We use 10-fold time-series cross-validation for tuning the regression models. We test:

- **Naïve models:** lags, historical means, exponentially weighted means, expanding mean;
- **Classic regression models:** lasso/ridge regressions, decision trees, random forests, XGBoost, AdaBoost;
- **Neural network models:** perceptrons, LSTM

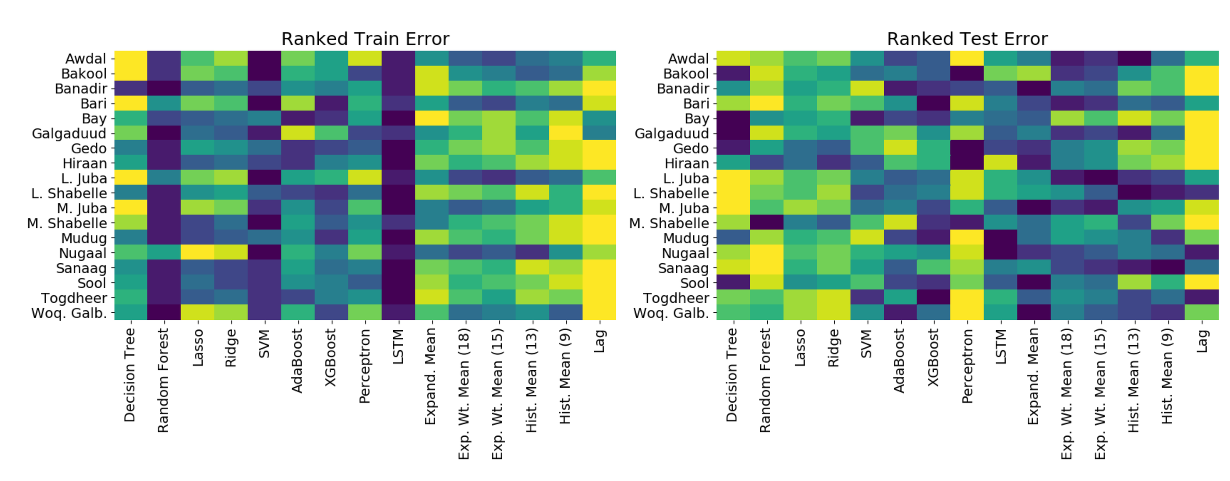

Figure 4: Ranked error by model on the train and test set (1-month)

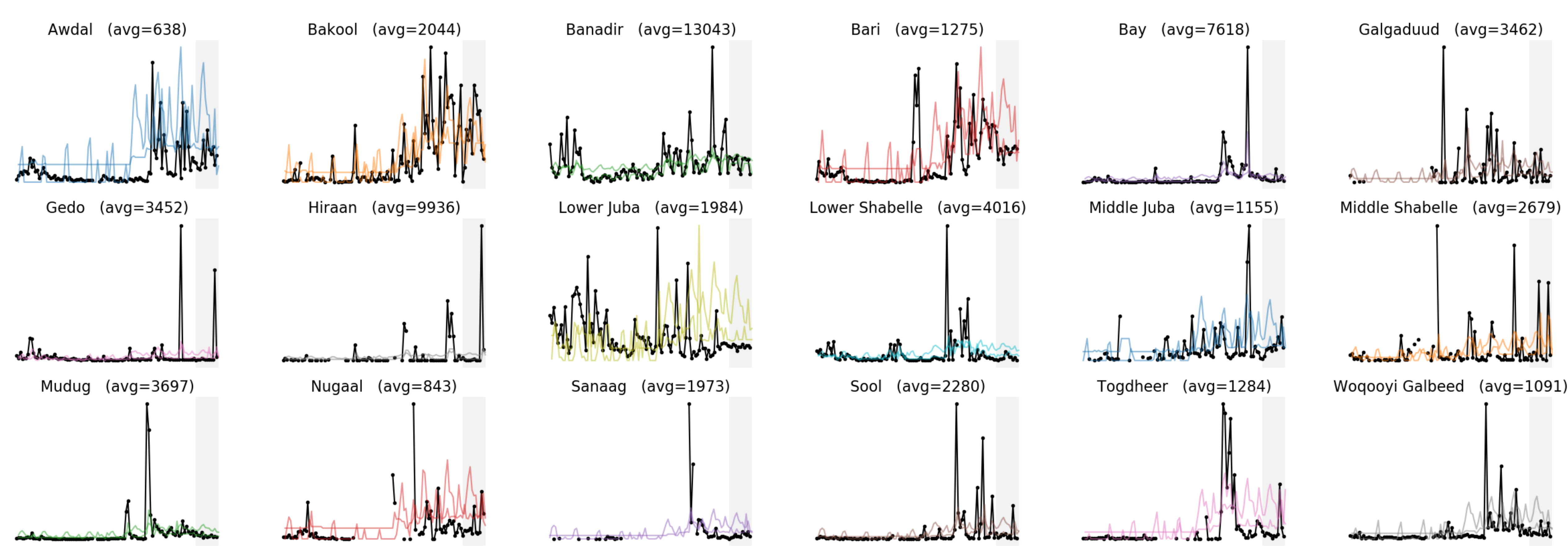

Figure 5: Scaled arrivals by region (true values are shown in black; colors illustrate XGBoost/Perceptron predictions with 1-month forecast horizon)

Work in Progress

- Further **tune** models to improve performance
- Explore **sub-regional** and **gravity models**
- Add additional models (e.g. econometric)
- Reframe as **classification** problem (red flags)

Preliminary Results

The best 1-month RMSE achieved was 18,021, relative to a mean of 3,555 arrivals, a standard deviation of 11,278, and a maximum of 235,069. Currently, the models struggle with overfitting and naïve benchmarks are competitive. Interestingly, we don't see performance declining much with the forecast horizon.

Model	Test - 1mo	Train - 1mo	Test - 3mo	Train - 3mo
Perceptron	**18,021**	11,423	18,229	11,508
Expand. Mean	18,083	12,338	18,319	12,871
Ridge	18,271	11,065	18,376	11,366
Lasso	18,328	11,043	18,391	11,182
Exp. Wt. Mean	18,341	11,776	19,015	12,920
Random Forest	18,388	9,624	18,851	9,687
XGBoost	18,454	9,479	**18,027**	11,700
SVM	18,460	11,043	18,267	11,171
Decision Tree	18,544	11,432	19,321	9,662
LSTM	18,639	**4,905**	-	-
Hist. Mean	18,654	12,167	19,299	12,929
AdaBoost	18,671	8,265	18,292	**9,147**
Lag	26,349	12,083	26,349	12,083

Table 1: Root Mean Square Error (RMSE) by model

Implementation

We have designed a dashboard for exploring the data and models, and plan to hand the project over to the regional team for long-term development. Open questions include:

1. What level of performance should be required before release? How to quantify/communicate **uncertainty**?
2. How to empower users to explore and interrogate? Can we use **explainability** tools to aid understanding?
3. When should a **warning** be triggered? What predicted % increase in arrivals is critical?

As we learn more about the problem, we are developing a broader literature review on forecasting displacement. We are also supporting discussions of how predictive analytics can be applied in other critical contexts (e.g. the Sahel).