Intro to Predictive Analytics (PA) in forced displacement settings

Understanding key concepts differences

Prediction:
Forecast + explainability

Predictive analytics does not tell you what **will exactly happen** in the future.

Instead, it forecasts what **might happen in the future** with an acceptable level of **reliability/certainty,** and includes **what-if scenarios and risk assessment by describing the factors (predictors) that influence the prediction.**

Forecasting:
a process of predicting or **estimating future events based on past and present data** and most commonly by **analysis of trends**

Weather forecast?

or

Weather prediction?

Simulation:
a **representation** of the **dynamic responses of a system** by **reproducing the behaviour** of the elements of the system.

But the same...

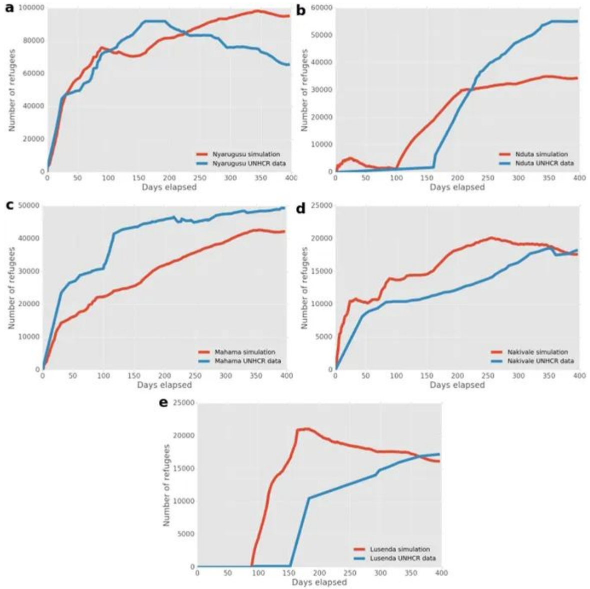
$$X = \begin{bmatrix} y_{N-1} & y_{N-2} & \cdots & y_{N-n-1} \\ y_{N-2} & y_{N-3} & \cdots & y_{N-n-2} \\ \vdots & \vdots & \vdots & \vdots \\ y_n & y_{n-1} & \cdots & y_1 \end{bmatrix}$$

ALL forecasting, PA and simulations have:

1. **An output**
   a. predict movement or natural disasters
2. **Target variable (dependent)**
   a. # of arrivals, % of food insecure people, or probability/likelihood for a typhoon to hit x area
3. **Predictors (independent variables) or rules (simulation)**
   a. Factors affecting the dependent variable and/or the agent (human)
4. **Data**
   a. And you need lots of it for ANY chosen method, see below
5. **Method/Modeling**
   a. Statistical modeling, forecasting
   b. ML-based (supervised, semi-supervised or unsupervised)
   c. Computer simulation with agent-rules

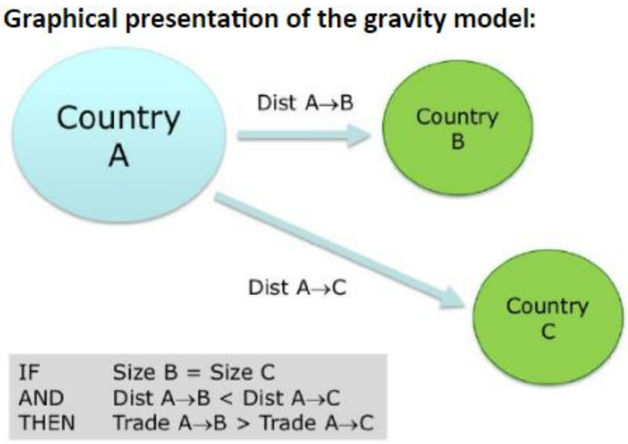
Simulation: *Agent-based modeling (ABM)* Brunel University

- **Output:** Movement and distribution of refugees across camps in conflict situation
- **Target variable:** number of displaced
- **Simulation Method:** Simulation (agent-based modelling)
- **Data:** UNHCR, the Armed Conflict Location & Event Data Project (ACLED)



Suleimenova, D., Bell, D. & Groen, D. A generalized simulation development approach for predicting refugee destinations. *Sci Rep* 7, 13377 (2017) doi:10.1038/s41598-017-13828-9

Forecasting: Online search data to predict migration flows (Kiel Institute for the World Economy)

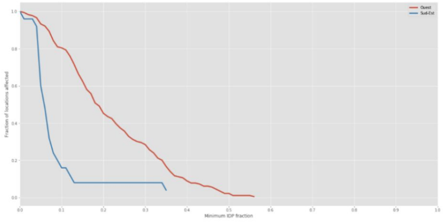


- **Output:** migration flows
- **Target variable:** number of migrants
- **Simulation Method:** Gravity model
- **Data:** online search behavior (google trends)

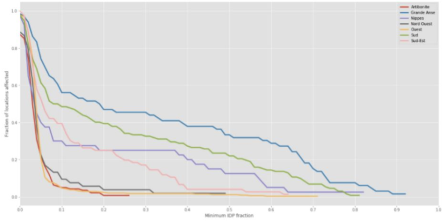
Google Trends	(1) None	(2) Migration	(3) Economic	(4) Mig+Econ
Log GDP (origin)	-0.641***	-0.399**	-0.445**	-0.344*
	(0.231)	(0.182)	(0.198)	(0.176)
Log Population (origin)	2.161***	1.626***	1.793***	1.432**
	(0.597)	(0.561)	(0.672)	(0.612)
GTI Migration keywords (37)		√		√
GTI Economic keywords (37)			√	√
Origin FE	√	√	√	√
Year FE	√	√	√	√
Observations	1,068	1,068	1,068	1,068
Joint significance GTI keywords (p-value)	–	0.000	0.0002	0.000
within-R^2	0.077	0.2080	0.167	0.258
Number of Origins	98	98	98	98

Böhme, Marcus H., et al. *"Searching for a Better Life: Predicting International Migration with Online Search Keywords."* Journal of Development Economics, 2019, p. 102347.

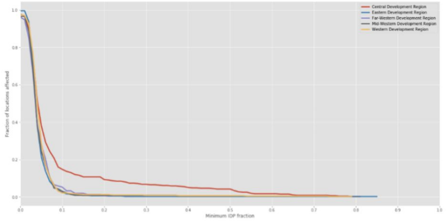
Forecasting: Detecting internal displacement with mobile CDR data (Flowminder)



(a) Haiti earthquake



(b) Hurricane Matthew.



(c) Nepal earthquake.

Detecting individual internal displacements following a sudden-onset disaster using time series analysis of call detail records

Tracey Li*, Jesper Dejby, Maximilian Albert, Linus Bengtsson, Véronique Lefebvre†

Flowminder Foundation

17th May, 2019

Abstract

We present a method for analysing mobile phone call detail records to identify individuals whom we believe to be have been internally displaced as a result of a sudden-onset disaster. We model each anonymous individual's movement trajectory as a piecewise-constant time series signal, assume that a disaster-induced displacement is exhibited as a level shift from an individual's 'normal' location, and then apply a step detection algorithm to identify level shifts in the signal. In contrast to typical methods that are used to analyse mobility patterns from call detail records, where the aggregate movements of large groups of individuals are analysed, our method offers the advantage that no assumptions regarding the destination or duration of an individual's displacement are necessary. We have applied the method to the datasets from three disasters - the 2010 earthquake in Haiti, the 2015 Gorkha earthquake in Nepal, and Hurricane Matthew in Haiti in 2016. Our results demonstrate that this method can facilitate improvements in the analysis and modelling of the mobility of internally displaced persons in post-disaster scenarios, using call detail records. Such analyses can be used to complement traditional survey methods to assess the scale and characteristics of disaster-induced displacements in a timely manner.

FLOWMINDER.ORG PROVIDING PRICELESS INFORMATION FOR FREE FOR THE BENEFIT OF THOSE WHO NEED IT THE MOST ABOUT WHA

Home Publications FlowKit: Unlocking the power o…

FlowKit: Unlocking the power of mobile data for humanitarian and development purposes

Daniel Power, Martin Thom, Jonathan Gray, Maximilian Albert, Sophie Delaporte, Tracey Li, James Harrison, Joshua Greenhalgh, Nick Thorne, and Linus Bengtsson

Flowminder, Flowminder-DIAL White Paper

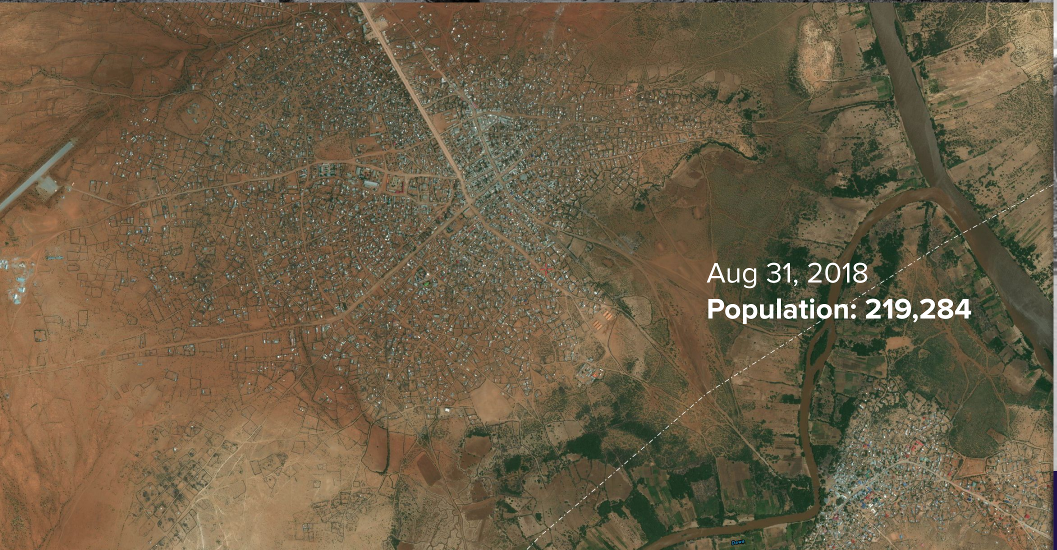
The Flowminder Toolkit:
how to use mobile call detail records (CDR) data

Intro to Project Jetson

Welcome to Project Jetson.

Jetson is a machine learning-based experiment that provides predictions on the movement(s) of displaced people. This experimental project combines data science, statistical processes, design-thinking techniques, and qualitative research methods. Jetson actively seeks new data sources, new narratives, and new collaborations in order to keep iterating, and improving. It has further underlined the importance of partnership, of collaboration, and of transparency.

Operational Challenge



Emergency: July 19, 2011 **Population: 115,224**

The Challenge: asking us interesting question(s)

UNHCR Somalia operation: IDPs

- Where are they moving? = categorical variable
- When are they arriving? = numerical variable
- **How many** people are moving? = **numerical**

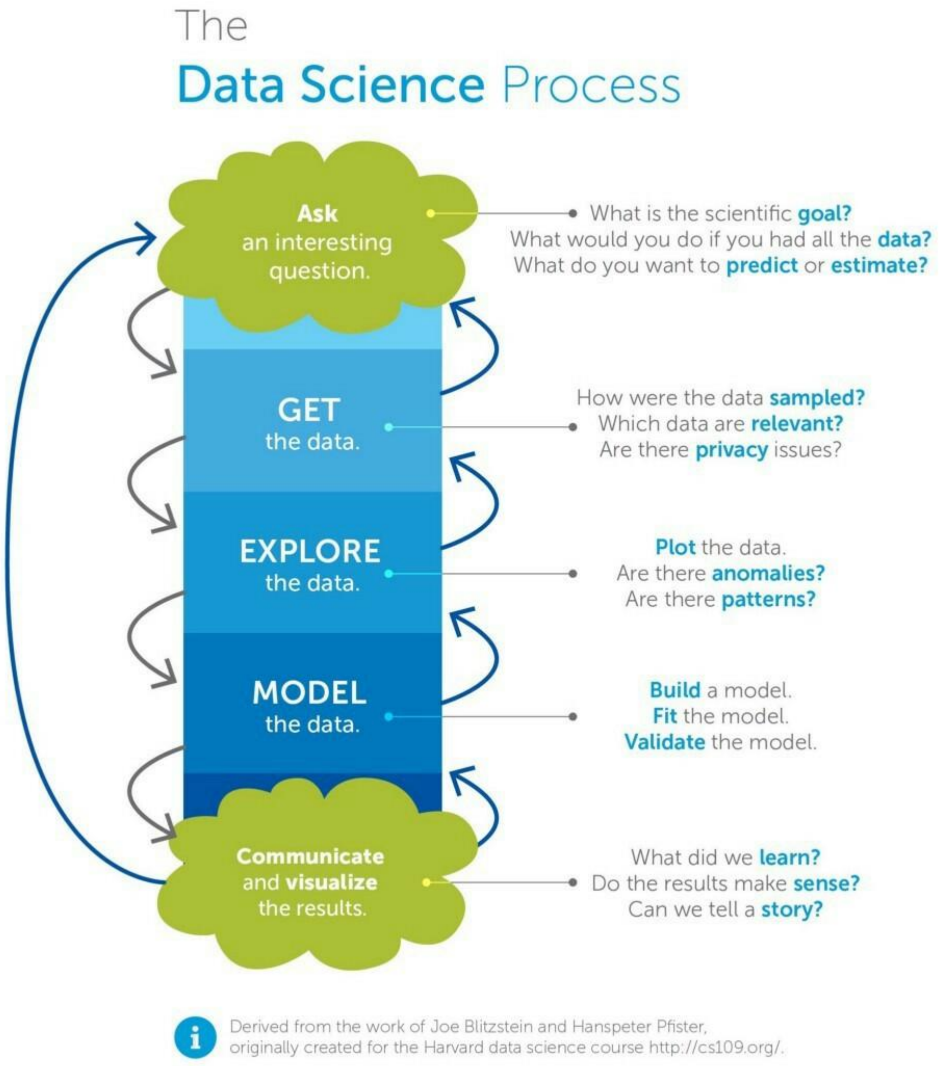
UNHCR Ethiopia operation: Refugees

- Dollo Ado, Melkadida sub-office: given 2017 conditions (drought/conflict) and with prior institutional memory of displacement from 2011 with similar conditions...
  - Are we going to receive the **same amount** of refugees? = **numerical**

Why Jetson? (Rationale why we got involved in 2017)

- Help 2 UNHCR Operations make evidence-based decisions and adequately plan/prepare contingencies
- Set a precedent for this type of predictive analytics work in humanitarian sector for coordination and compilation of data
- Open data, including models for other data scientists, statisticians, computer scientists, programmers/students can edit, add, and improve upon

Methodology



Project Jetson: creating a recipe of PA applied to forced displacement settings

Ingredients

Preparation

Results

Party?

Challenge scope

What can machine learning do to solve this **challenge**?

How do we deal with lack of data?

Can we predict at least for one month in advance?

Can we detect push and pull factors that drive forced displacement?

Brainstorming challenge: finding predictors

- **Desktop research: push/pull factors for**
  - Climate anomalies
    - Rain
    - River levels
  - Violent conflict
- **Historical numbers of arrivals: People movement**
  - To Dollo (Ethiopia)
  - Within Somalia
- **Unexpected/policy changes**
  - Closure of borders (porous)
  - Changes in asylum policy (e.g. Dadab, Kenya)

Recognizing Bias and Building Assumptions

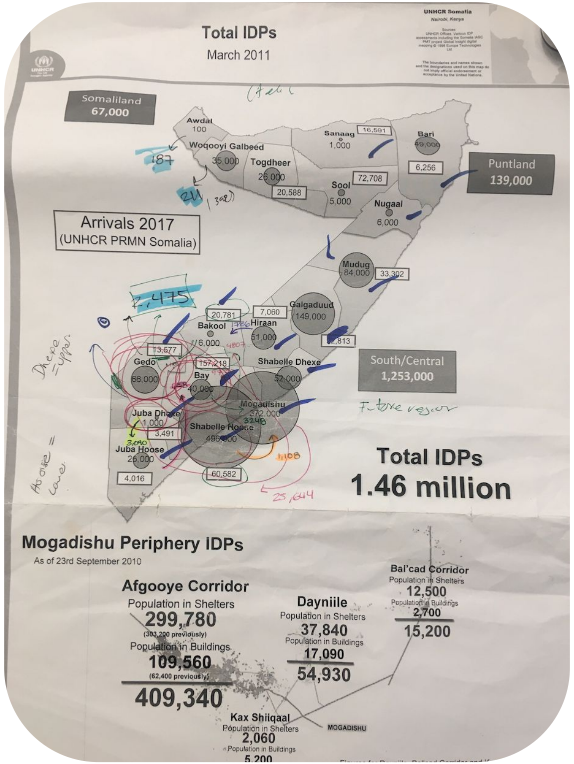
Assumptions:

“Persons of Concern (PoCs) in Somalia are fleeing from conflict areas”

“PoCs movement is also affected by external factors (e.g. drought/floods)”

“PoCs in Somalia are going to places where humanitarian assistance is being provided”

“A machine/computer program could help us predict how many people are going to be arriving/moving to a particular region”



Bias:

- Collector bias: ACLED, UNHCR-PRMN
- Representativeness: UNHCR-PRMN Methodology

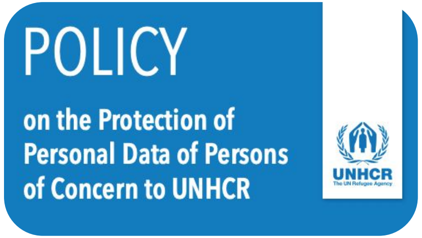
Research Assumptions

Project Phase	Assumption	Way we tested	Result
Challenge Definition	*"PoCs in Somalia are fleeing from conflict areas"*	• Desktop research: literature review on historical conflict in Somalia • Data on conflict (ACLED) graph trend	• **True:** 7 years of data historical conflict doubled and all literature pointed out on ethnic, inter-tribal and extreme groups conflict
	"PoCs in Somalia are going to places where humanitarian assistance is being provided"	• Desktop research (data on humanitarian assistance) • Semi-structured interviews (phone calls) with partners on their activities in certain regions	• **Inconclusive**: not enough data to conclude assistance is a pull factor. However, partners in their interviews mentioned the fact cash is actually preferred by PoC (qualitative/phone call)
	"PoCs movement is also affected by external factors (e.g. drought/floods)"	• Focus groups with PoCs • Interviews with different operational partners	• **True AND** found out because of focus groups we were missing a variable: **goats/market prices**
[Solution] Ideation	*"A machine/computer program could help us predict how many people are going to be arriving/moving to a particular region"*	• Desktop research: academic papers on machine learning • Read blogs on other agencies trying to do similar work (predictive analytics)	• **True**: academic research and financial sector research on predictive analytics by analyzing historical data and getting accurate results.
Experimentation	*"The machine can predict PoC arrivals with 75-100% of accuracy"*	• Build the computer program and ran it - literally - 65 times, until it provided a good result.	• **False:** after running it for 65 times, the machine can provide the arrivals on a 60% to 110% accuracy (difference between actual arrivas vs. machine predicted)
Implementation	*"The predictions will help Somalia operation to do better planning/preparedness" (e.g. food distribution, shelter planning)"*	• Calls with Somalia operation on a monthly basis	• **False:** the operation uses it to fact-check if the arrivals are actually true then they allocate resources (send protection partners to specific areas)
Scaling	*"The predictions can help the humanitarian sector to share their data and do more regions or predict other variables"*	• Workshop with partners in Kenya (TBC)	Working on it :) you're part of it


Ingredients

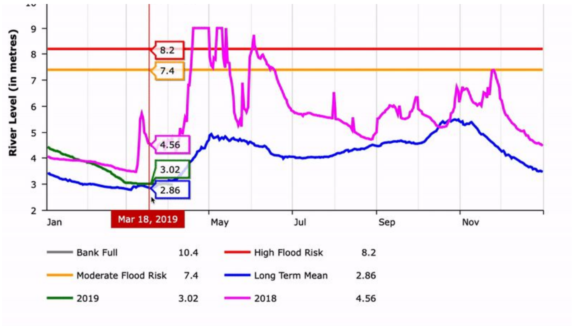

Humanitarian data issues

Legally Open



Aggregate
Anonymized

Technically open



1/1/2018	1.81	1.77	1.21	2.27	3.59	1.78	1.00
2/1/2018	1.17	1.15	0.61	1.87	3.72	1.44	1.00
3/1/2018	2.15	1.59	2.66	2.39	4.32	2.24	1.88
4/1/2018	4.75	3.07	3.69	4.86	6.27	3.74	3.56
5/1/2018	7.94	6.64	4.64	6.58	7.14	4.45	3.97
6/1/2018	5.37	4.63	4.65	6.99	7.79	4.68	4.42
7/1/2018	4.18	3.23	4.31	4.59	5.66	NA	2.76
8/1/2018	3.20	2.12	4.16	3.40	5.06	NA	2.32
9/1/2018	4.24	3.79	4.31	NA	5.96	3.06	2.83
10/1/2018	2.71	2.32	3.72	3.38	5.54	2.81	2.61

Machine readable format

Data Quality

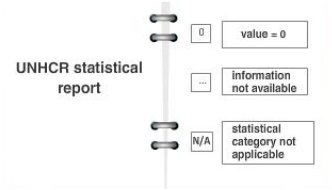
Issue 1: manual collection *[automation issues]*

*"We normally have a delay in getting all the data from rainfall gauges around the country **as the ministries and partners in Somalia have to collect and share with us,** August data, we will have the **full dataset after 10th of September.** Will share with you once it is received"*

Issue 2: Humanitarian access [*lack of*]

*"Please find attached available rainfall data for stations within South and Central stations. I am afraid **we do not have data** for Wajid, Afogoye and Genale as of now - **efforts to reach our gauge readers has been futile".***

Issue 3: Data quality



Data sources: who has what in Somalia?

Humanitarian:

- **Displacement (y variable):** Protection & Return Monitoring Network (PRMN), UNHCR-NRC, 7 years data
- **Conflict:** ACLED Data, 7 years data

Love-to-have (and N/A)

Remittances: Dahabshiill

CBI Data: partially in HDX (only 1.5 years)

Development:

Climate & Weather anomalies

- **Rainfall:** FAO SWALIM, 7 years data
- **River Levels:** FAO SWALIM, 7 years data

Market Prices

- **Local commodities:** FAO FSNAU, 7 years data

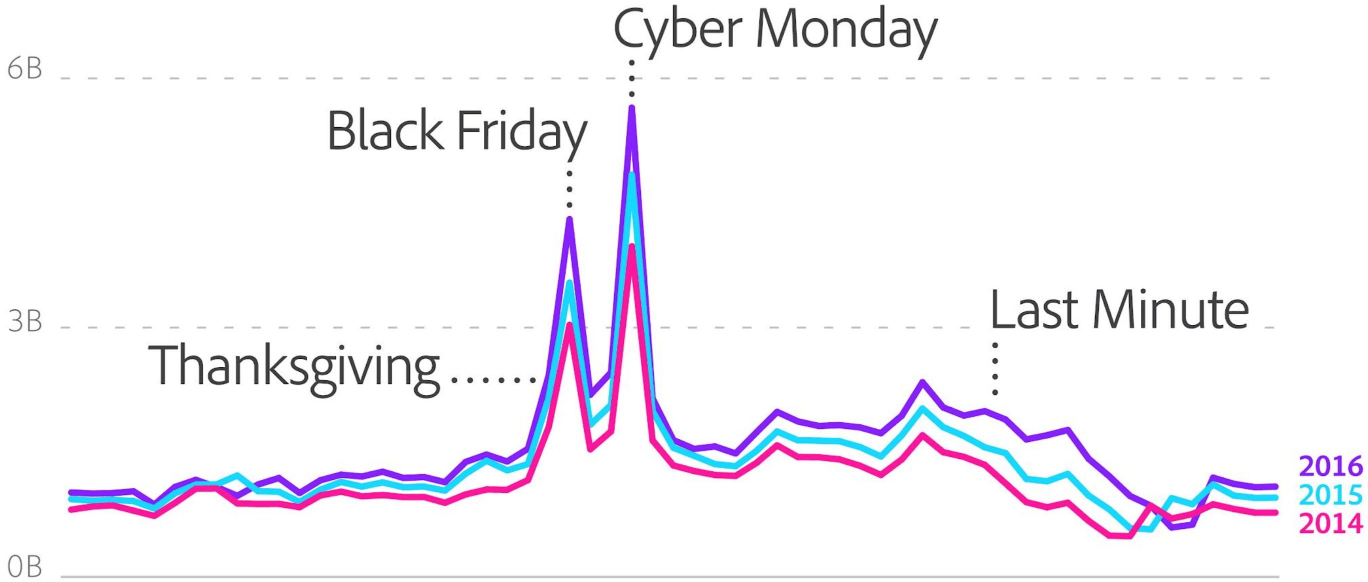
Additional References:

- WMO-ICPAC rainfall dekadal reports
- Indian Ocean Dipole forecasts

Get Data

Organization - Unit	Topic	Data type	Frequency	Format
ACLED	Conflict	• Violent incidents • Fatalities	• Weekly	• CSV, with python API
FAO - SWALIM	Climate	• River levels/discharge • Rain	• Daily (sensor data)	• Web table ○ *We built parser to extract (R)*
FAO- FSNAU	Market Prices	• Local Goat price • Water drum price	• Monthly	• CSV
WMO- ICPAC	Weather	• Weather Forecast	• Dekadal • Monthly • Seasonal	• PDF report
UNHCR PRMN - Somalia	Population	• IDP movement	• Daily	• CSV ○ *We built a parser to aggregate*
UNHCR Ethiopia		• Registration data	• Daily	• CSV ○ *We built a parser to aggregate*

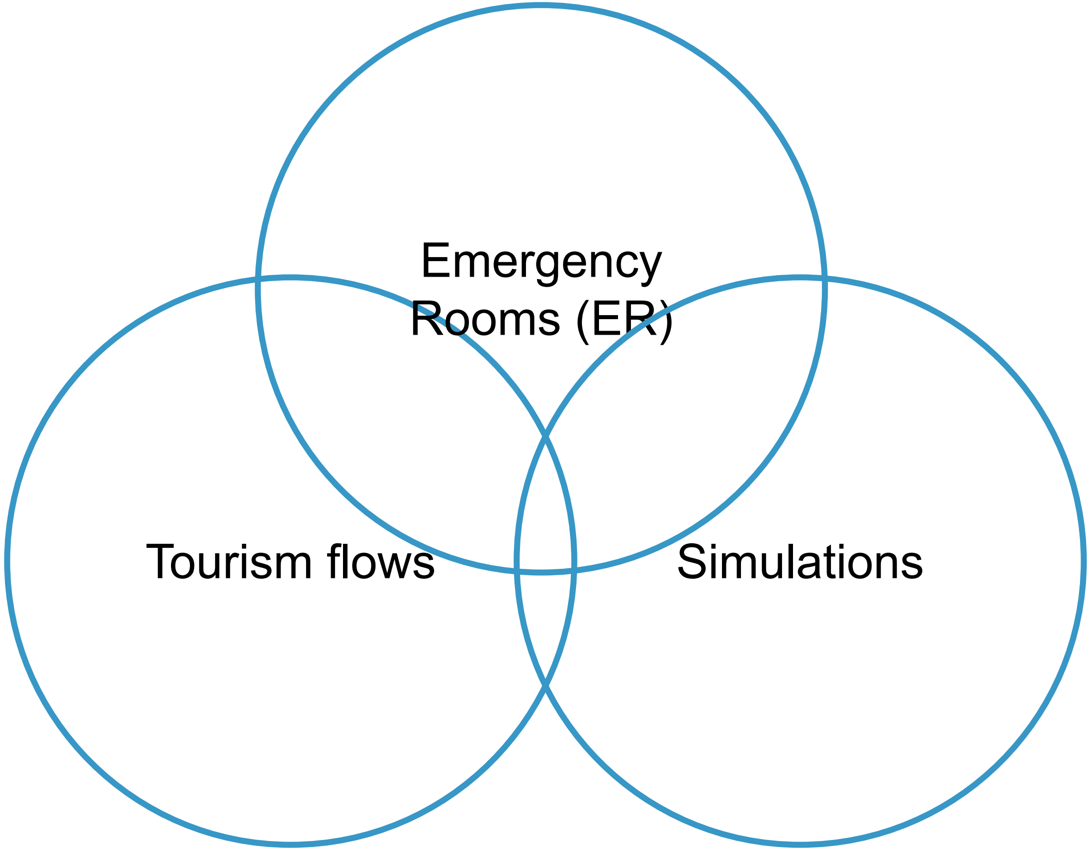
Intro to ML-based Predictive Analytics in Forced displacement



Artificial Intelligence Research and Theories

- Disassociated theories - cognitive process, economic giving receiving region
- Geographical theories – distance, cost of transportation
- Post–phenomenon theories – cannot assist for prediction

Population Flow studies



Machine Learning on Time Series Forecasting

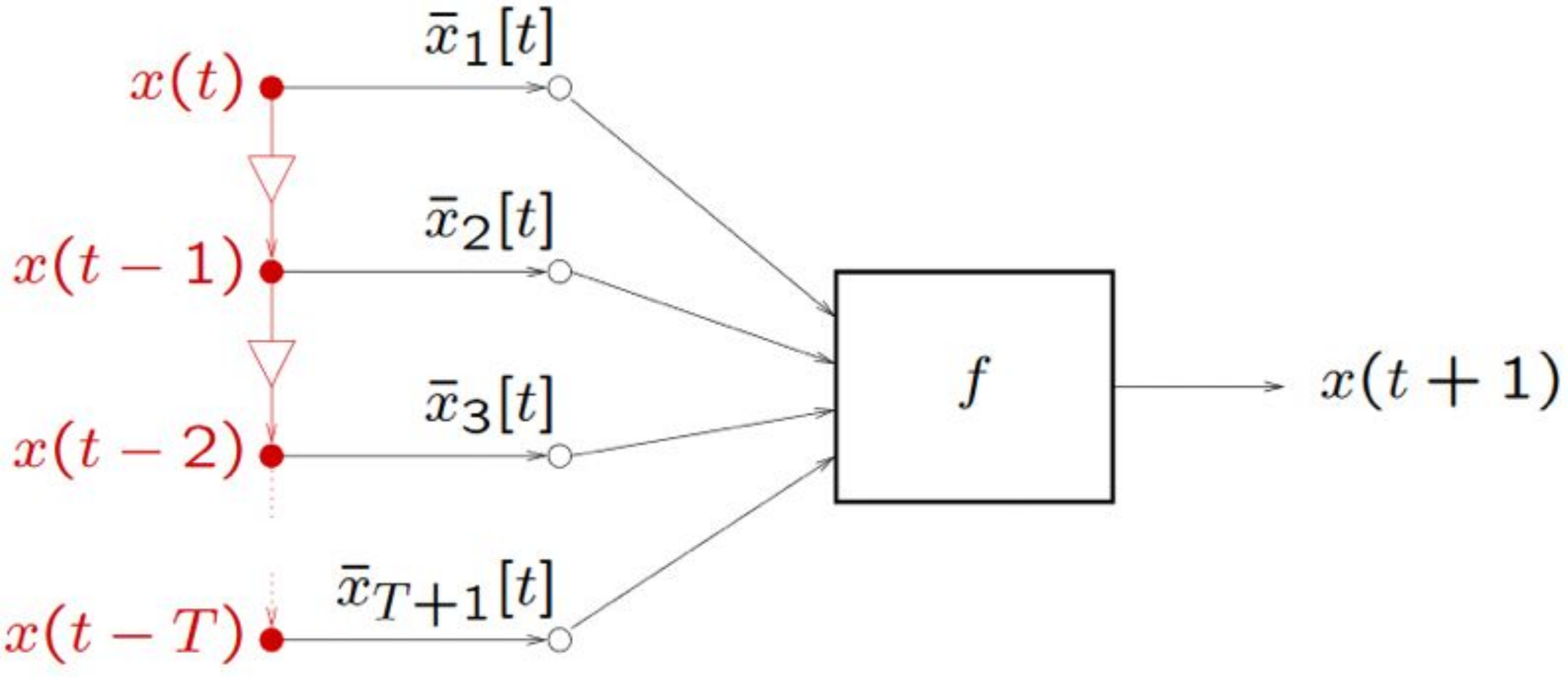
- Short - Long Term Memory

  Recurrent Neural Networks

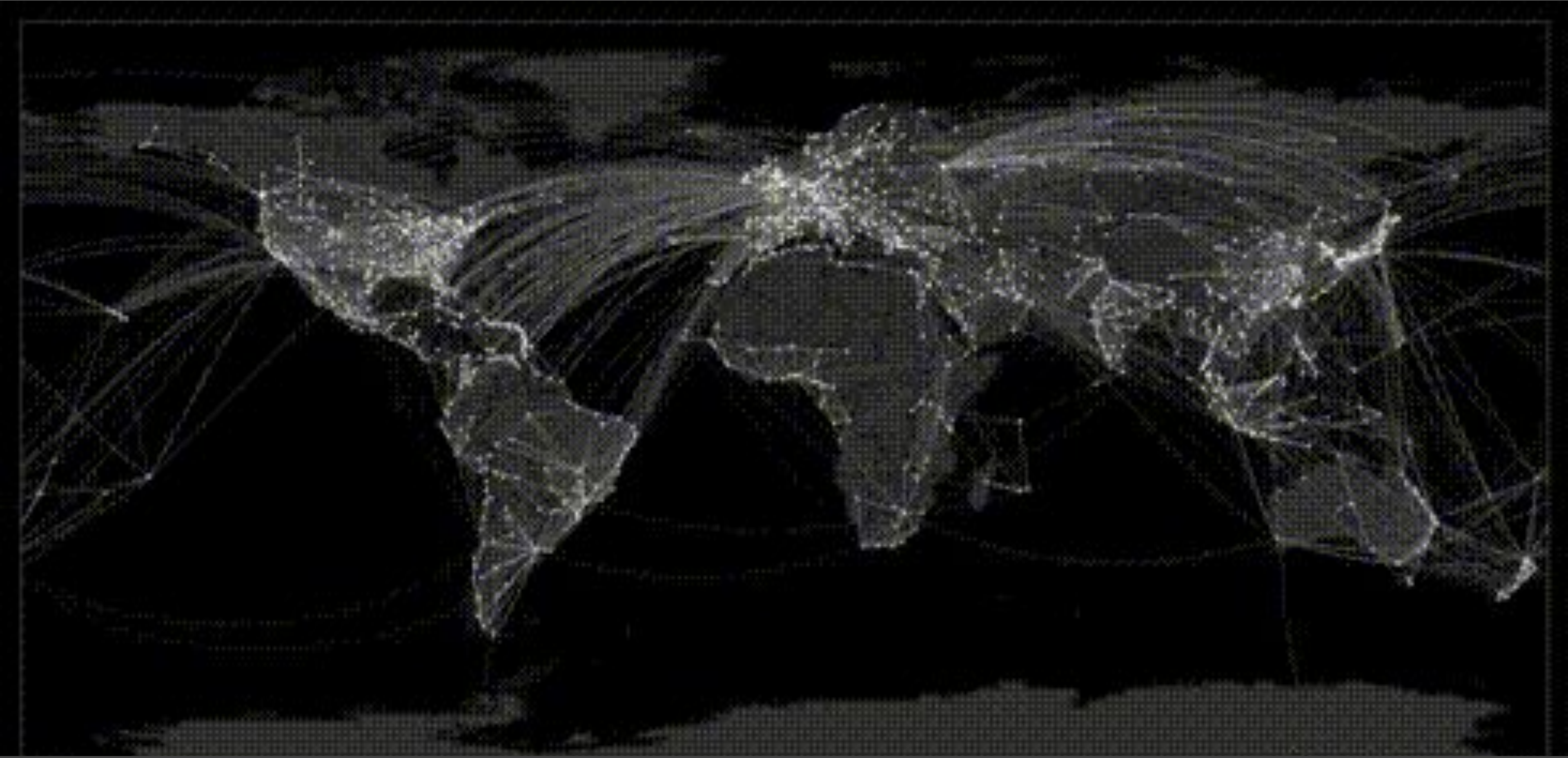
- Genetic and Evolutionary

  Algorithms

Multivariate time series forecasting



Current status: population flow research in displacement



2017 Secondary Data Review: lack of TA with ML in population movement

Author/Organization	Title	Topic (Data Science/Stats)
IDMC	Assessing drought displacement risk for Kenyan, Ethiopian and Somali pastoralists	Agent-based model (ABM)
Brunel University London ● Groen, D. Bell, D, & Suleimenova, D. *(currently working as partners)*	A generalized simulation development approach for predicting refugee destinations	Agent-based model (ABM)
Carnegie Mellon/Singapore Management University ● Cheng, S., Lin, L. & Carley, K.	An agent-based approach to human migration movement	Agent-based model (ABM)
U.S.A. National Research Council	Model-based approaches to estimating migration flows	Bayesian models Regression models Simulation models
George Mason University ● Gulden, T., Harrison, J., Crooks, A.	Modeling Cities and Displacement through an Agent-based Spatial Interaction Model	Agent-based model (ABM)
University of Pennsylvania ● Mellers, et.al, Merkle, E. & Tetlock, P.	The Psychology of Intelligence Analysis: Drivers of Prediction Accuracy in World Politics	Forecasting vs. Intelligence Analysis
Berkeley University ● Hsiang, S., Burke, M. & Miguel, E.	Quantifying the Influence of Climate on Human Conflict	Regression


Preparation

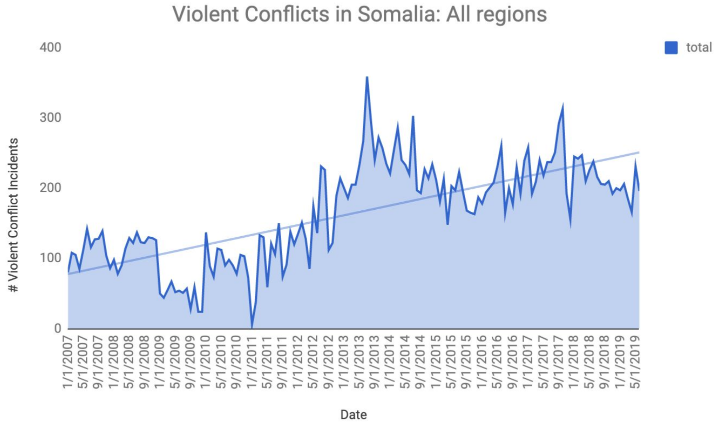

Jetson Experiments timeline

1. **Experiment #1 (mid-2017 to 2018):** predicting forced displacement **1 month in advance.**
   a. First experiment, using off-the-shelf software for multivariate TSA, combined with open-source scripts for ML (R and Python)
      i. *Training set* January 2010 - June 2017 or January 2010 to September 2017
      ii. *Testing set:* October 2017 - March 2018
      iii. *Cross-validation*: March 2018 to whatever we had available (new each month)

2. **Experiment #2 (2018 and onwards):** predicting forced displacement **1 month to 3 months in advance, testing 6 and 12 too.**
   a. Second experiment, using full open-source scripting for multivariate TSA for ML (R and Python) and also exploring the use for gravity model
      i. *Training set:* January 2010 to September 2017
      ii. *Testing set:* October 2017 - October 2018
      iii. *Cross-validation:* November 2018 and onwards

3. **Experiment #3 (planned, 2019):** finding t**he nexus between climate, conflict and forced displacement,** building correlations among NDVI and displacement/conflict data

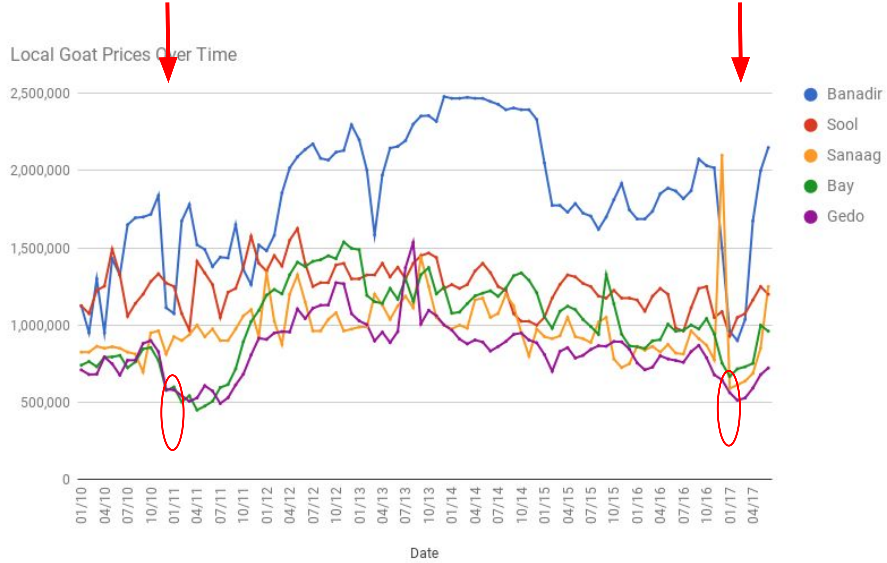
Explore Data - Descriptive Analytics: *The importance of predicting other agencies predictors*



Area monthly IDPs correlation with Region MaximumWater Drum Price with (1) Month Lage

	Awdal	Bakool	Banadir	Bari	Bay	Galgaduud	Gedo	Hiiraan	Jubbada_Dhexe	Jubbada_Hoose	Mudug	Nugaal	Sanaag	Shabeallaha_Dhexe	Shabeellaha_Hoose	Sool	Togdheer	Woqooyi_Galbeed
Awdal_WaterDrumPrice	0.13	0.42	0.15	0.25	0.067	-0.25	0.089	0.13	0.081	-0.018	-0.15	0.08	0.11	0.12	0.0015	0.046	-0.17	-0.045
Bakool_WaterDrumPrice	0.25	0.29	0.43	0.17	0.34	0.18	0.19	-0.24	-0.026	-0.067	-0.13	-0.11	0.18	0.11	0.13	0.0055	0.19	0.2
Banadir_WaterDrumPrice	-0.69	0.043	-0.1	-0.023	0.0068	0.18	-0.027	-0.072	0.00026	0.14	0.099	0.083	0.87	0.052	0.1	0.058	0.1	-0.86
Bari_WaterDrumPrice	-0.15	-0.19	-0.21	-0.21	-0.1	-0.15	-0.046	0.28	-0.072	-0.049	-0.097	0	0	-0.05	-0.09	-0.043	-0.035	-0.095
Bay_WaterDrumPrice	0.12	0.16	0.29	0.049	0.34	0.24	0.24	-0.08	0.24	0.039	0.076	0.12	0.31	0.48	0.26	0.16	0.37	0.16
Galgaduud_WaterDrumPrice	-0.21	-0.44	-0.3	-0.48	-0.19	-0.15	-0.094	-0.19	-0.13	-0.00066	-0.094	0.1	0.54	0.016	0.26	-0.16	-0.056	-0.006
Gedo_WaterDrumPrice	-0.4	-0.41	-0.38	-0.5	-0.25	-0.13	-0.11	0.016	-0.16	-0.12	-0.22	-0.39	-0.11	0.033	-0.21	-0.23	-0.18	-0.23
Hiiraan_WaterDrumPrice	0.065	0.037	0.25	-0.24	-0.056	0.13	-0.092	-0.28	-0.029	-0.076	-0.081	-0.11	-0.0018	0.17	-0.046	0.021	0.14	-0.064
Jubbada_Dhexe_WaterDrumPrice	-0.1	-0.0026	0.044	-0.095	-0.046	-0.023	8.5e-05	-0.068	-0.19	-0.23	-0.052	-0.13	0.0091	0.23	0.00055	-0.17	0.056	-0.075
Jubbada_Hoose_WaterDrumPrice	-0.00069	0.28	0.3	0.36	0.11	-0.17	0.15	0.12	0.22	0.056	-0.14	-0.2	-0.49	0.17	-0.31	0.049	-0.42	-0.12
Mudug_WaterDrumPrice	0.057	-0.16	-0.017	0.082	0.051	0.42	0.05	-0.081	0.12	0.54	0.12	-0.051	0	-0.038	0.053	-0.021	0.02	-0.014
Nugaal_WaterDrumPrice	0.07	0.42	0.35	0.44	0.2	-0.19	0.22	0.11	0.22	-0.012	-0.14	-0.054	-0.51	0.13	-0.29	0.023	-0.38	-0.094
Sanaag_WaterDrumPrice	0.24	0.52	0.42	0.51	0.26	0.27	0.11	-0.037	0.17	0.089	0.097	0.09	-0.28	0.16	-0.11	0.19	0.17	-0.1
Shabeallaha_Dhexe_WaterDrumPrice	0.41	-0.28	0.081	-0.17	-0.0045	-0.036	0.041	-0.52	-0.28	-0.3	-0.039	-0.23	0.061	-0.049	-0.03	-0.23	-0.02	0.7
Shabeellaha_Hoose_WaterDrumPrice	0.024	0.38	0.25	0.26	0.26	0.14	0.13	0.026	0.23	0.0055	-0.24	0.11	0.099	0.24	-0.34	0.17	0.3	-0.05
Sool_WaterDrumPrice	0.32	-0.06	0.3	0.21	0.24	0.42	0.05	-0.2	0.096	0.26	0.1	-0.11	0.71	0.17	0.19	0.3	0.4	0.35
Togdheer_WaterDrumPrice	0.17	0.46	0.3	0.4	0.12	-0.14	0.13	0.15	0.15	-0.0069	-0.19	-0.19	-0.7	0.14	-0.22	0.018	-0.32	-0.13
Woqooyi_Galbeed_WaterDrumPrice	0.021	0.35	0.19	0.059	0.17	-0.049	0.14	0.4	0.17	-0.059	0.0064	-0.11	-0.075	0.04	-0.19	0.24	0.025	-0.17

“Find your goats”: nesting behavior for movement



How many goats did you sell?

“All of 50 of them… the goats don’t survive the [flight] journey”

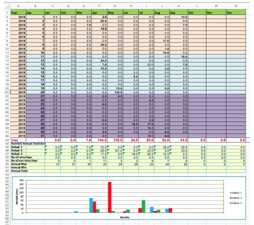
Mission to Dollo Ado, Ethiopia (May 2017)

Results

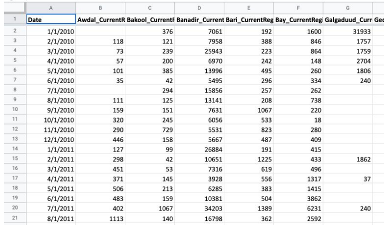
Ideal Recipe

Jetson Recipe

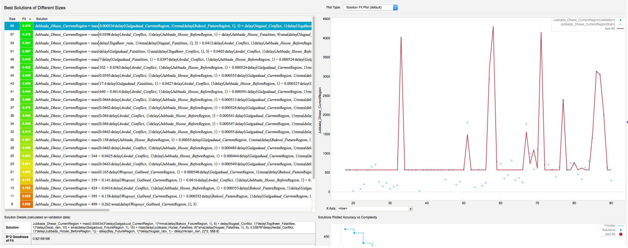
Preparation: Jetson recipe



Step 1: Data wrangling



Step 2: Open data repository (g-sheets)

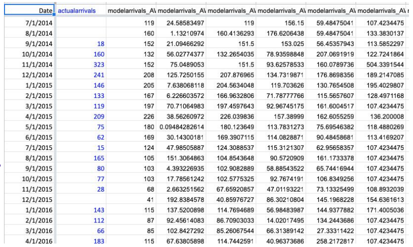


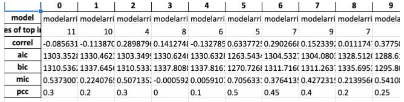
Step 3: Statistical modelling: exploration

(R^2 + Mean Absolute Error)

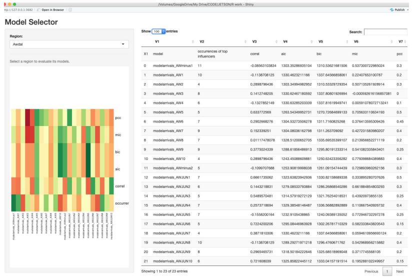
Step 4: Optimus prime

(algorithm transformation: math, equation, string packages python)



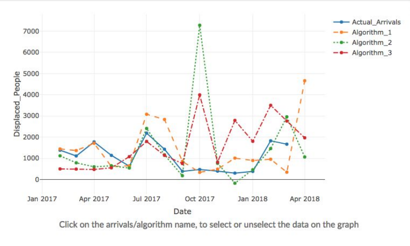


Step 4: models output machine vs. actuals



Step 5: model evaluation

(human selection via R-Shiny app: BIC, AIC, Correlations)



Step 6: Data Visualization + Automatic Report

The first test: Bay (region) model - weeks in advance

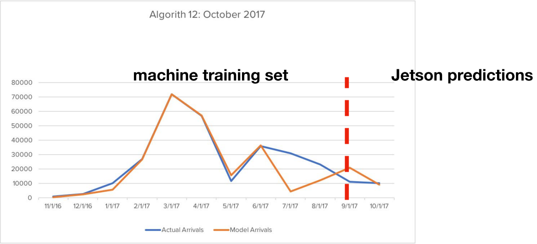
Algorithm example (to be read by the machine)

Model produced by machine:

Bay_CurrentRegion = max(max(delay(Awdal_CurrentRegion, 1), 0.000127177740817059*delay(Bakool_WaterDrumPrice, 1)*delay(Awdal_CurrentRegion, 1) + log(delay(Jubbada_Hoose_rain, 1))*smm(delay(Bari_CurrentRegion, 1), 11) + 2.11983998917739e-6*delay(Mudug_BeforeRegion, 1)*delay(Shabeellaha_Dhexe_BeforeRegion, 1)*sma(delay(Nugaal_FutureRegion, 1), 10) + delay(Sool_BeforeRegion, 15)) - delay(Jubbada_Hoose_FutureRegion, 1), 0.805982905429227*delay(Awdal_Conflict, 1)^2*delay(Sool_Fatalities, 1)*delay(Bari_FutureRegion, 1))

```
rrivals_BAY13arrivals <- function(start, end){
tart = 20
PI <- PA <- PD <- rep(NA, end)
for (t in start:end){

  A<- current.long[(t- 1),"Awdal_CurrentRegion"]
  B<- water.long[(t- 1),"Bakool_WaterDrumPrice"]
  C<- current.long[(t- 1),"Awdal_CurrentRegion"]
  D<- rain.long[(t- 1),"Jubbada_Hoose_rain"]
  E<- median(current.long[(t-12):(t- 1),"Bari_CurrentRegion"], na.rm=TRUE)
  G<- before.long[(t- 1),"Mudug_BeforeRegion"]
  H<- before.long[(t- 1),"Shabeellaha_Dhexe_BeforeRegion"]
  I<- mean(future.long[(t-11):(t- 1),"Nugaal_FutureRegion"], na.rm=TRUE)
  J<- before.long[(t- 15),"Sool_BeforeRegion"]
  K<- future.long[(t- 1),"Jubbada_Hoose_FutureRegion"]
  L<- conflicts.long[(t- 1),"Awdal_Conflict"]
  M<- fatalities.long[(t- 1),"Sool_Fatalities"]
  N<- future.long[(t- 1),"Bari_FutureRegion"]
  O<- log(D)
  P<- max(A,sum( 0.000127177740817059*B*C , O*E , 2.11983998917739e-6*G*H*I ,
                J,na.rm=TRUE),na.rm=TRUE)
  Q<- max(sum(P , -K,na.rm=TRUE), 0.805982905429227*L^2*M*N,na.rm=TRUE)
  FIN <-Q
  PA[t] <- FIN
  #Bay_Incidents
  PI[t] <- 0
  #Bay_Departures
  PD[t] <- 0
}
write.csv(as.data.frame(PA[1:length(PA)]), file ="results.csv",row.names = FALSE)
eturn(PA)
```

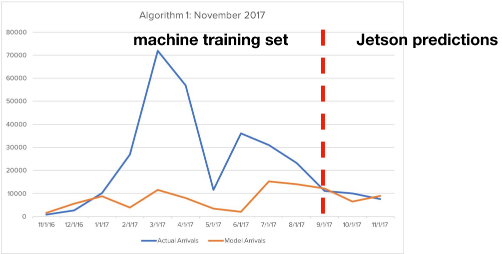


Predicting October

Time prior: 1 week

Jetson: **9,062**

Actual Arrivals: **10,003**

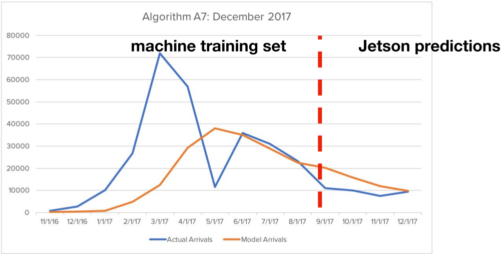


Predicting November

Time prior: 2 weeks

Jetson: **8,928**

Actual Arrivals: **7,750**



Predicting December

Time prior: 3 weeks

Jetson: **9,468**

Actual Arrivals: **9,881**

Model Evaluation: Example

Model #: **Region:**	BN_8 Banadir	**Average arrivals** ***(past 12 months)***	40,280.92	**R-square:** **MAE:**	0.86930941 3035.7865	**Test vs. training set**	50%/50% until Sep 2017
Date	Oct-17	Nov-17	Dec-17	Jan-18	Feb-18	Mar-18	Apr-18
Accuracy ***(PCC, percentage correctly classified)***	123%	112%	112%	98%	94%	39%	156%

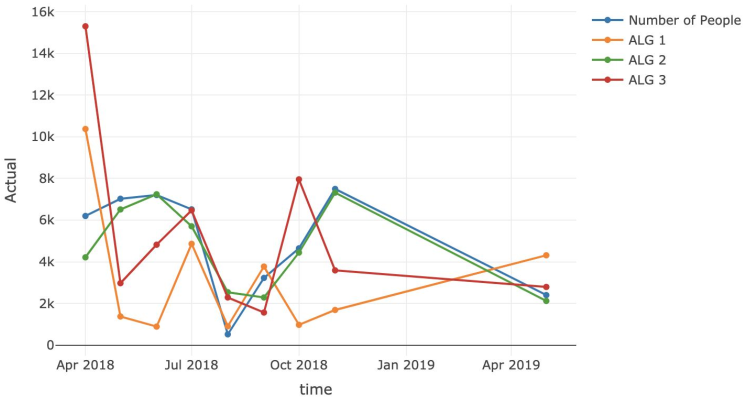
2017-2018 Experiment #2: Predicting 1 month in advance

Predictive Engine: Project JETSON

Jetson is a project aimed at providing better data analytics to make better decisions to adequately prepare for contingencies in forced displacement situations. The Predictive Analytics Engine (Jetson) is an applied predictive analytics experiment taking concrete steps to provide insights on the future of displacement.
The data behind the engine is anonymized, aggregated per month and per region. Project Jetson uses machine-learning for building a nonparametric algorithm (model) for each region. The models used for each region represent the best 'fit' that can explains the behaviour of seven years of historical data.

Choose region

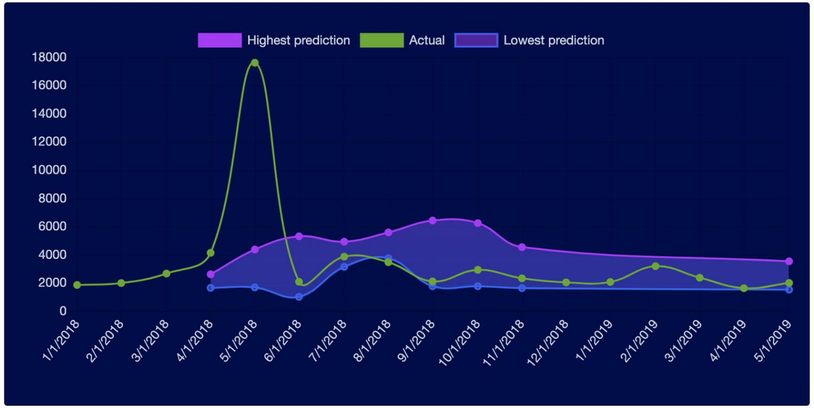
- Awdal
- Bay
- Bakool
- Banadir
- Gedo
- Middle_Juba
- Lower_Juba
- Middle_Shabelle
- Lower_Shabelle
- Hiiraan
- Galgaduud
- Mudug
- Nugaal
- Bari
- Sanaag
- Sool
- Togdheer
- Woqooyi_Galbeed



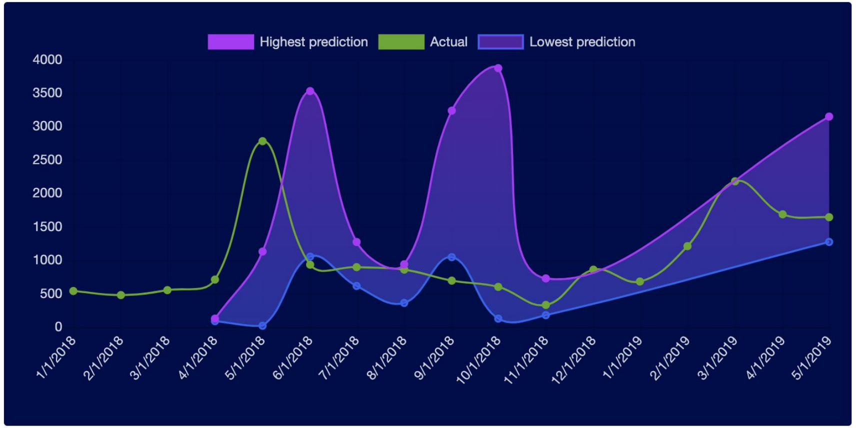
Click on the arrivals/algorithm name, to select or unselect the data on the graph

Predictions

Jubbada Hoose



Sanaag

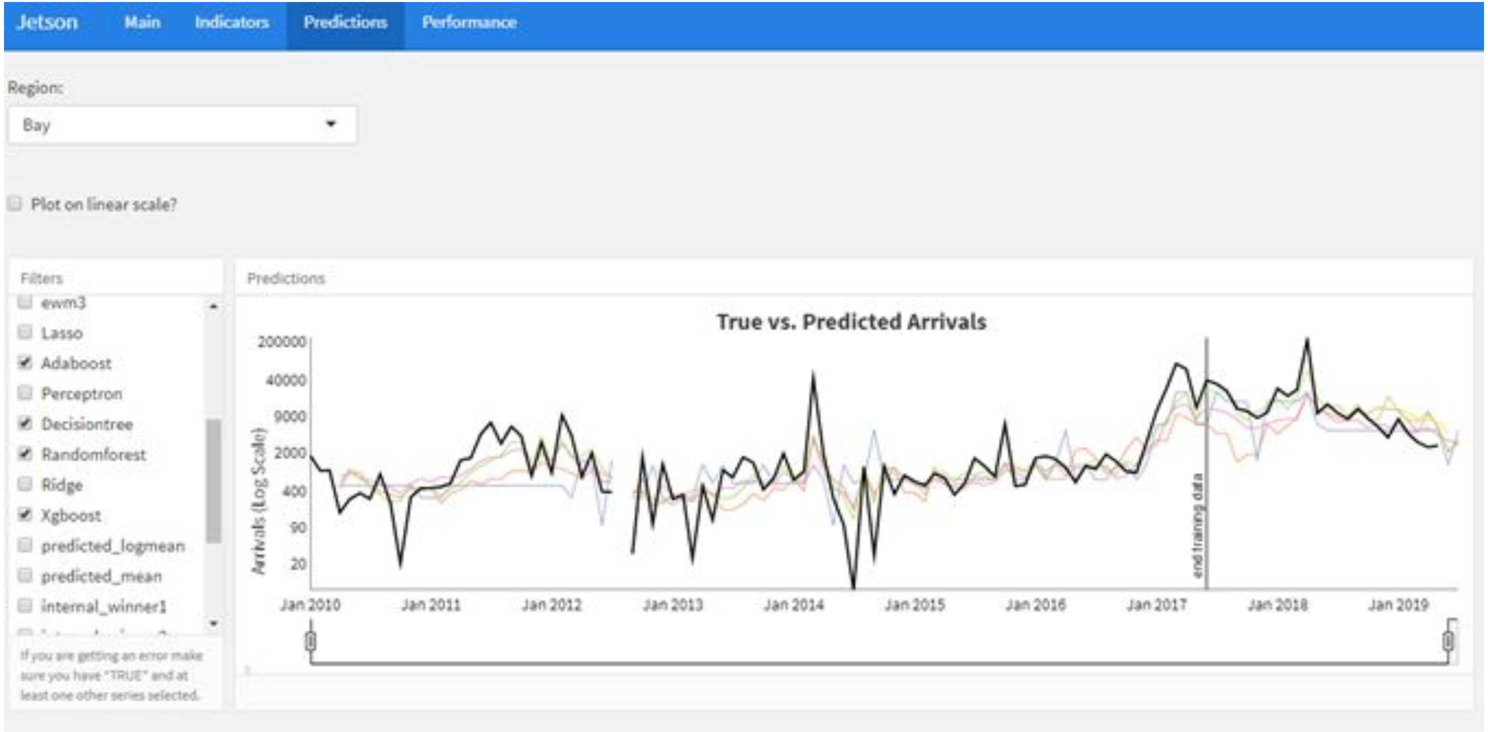


Transparency in A.I. black box: sensitivity map

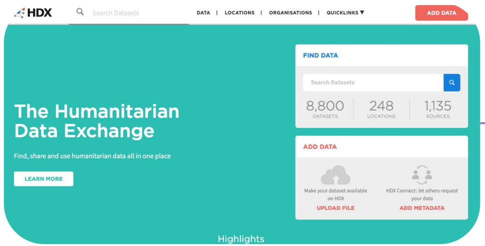
Variables Prevalence in 380 Algorithms						
PREDICTORS (what variables are influential in these regions displacement, either in their own region or neighboring regions?)						
REGION	Violent Conflict	Fatalities	Anomalies Rain patterns	Anomalies Rivers affectations	Goat prices*	Water drum prices
Awdal		yes	yes			
Bakool	yes	yes		yes		
Banadir	yes	yes	yes	yes	yes	yes
Bari	yes	yes	yes			
Bay	yes	yes		yes		yes
Galgaduud	yes	yes	yes	yes	yes	
Gedo	yes	yes	yes	yes		
Hiiraan	yes	yes	yes			
Middle_Juba	yes	yes				
Lower_Juba	yes	yes	yes	yes	yes	yes
Mudug	yes	yes			yes	
Nugaal	yes	yes		yes	yes	yes
Sanaag	yes			yes		yes
Middle_Shabelle	yes	yes	yes	yes		
Lower_Shabelle	yes	yes	yes	yes		yes
Sool	yes	yes				
Togdheer	yes	yes				yes
Woqooyi_Galbeed	yes	yes		yes		

2019: Experiment #2.1: predicting 3 months in advance

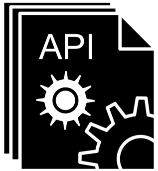
(in partnership with UN Global Pulse Data fellows program)



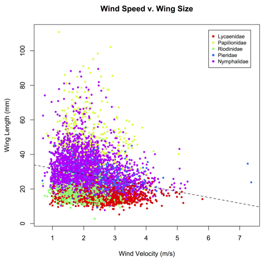
Our ideal PA recipe on forced displacement using ML-based TA



Data legally + technically open (repository)

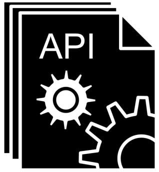


Automated Parsing



Exploratory projects (basic data relationships)

Machine-Learning algorithm(s) production (open source)



Model automatic extraction + automated transformation scripts

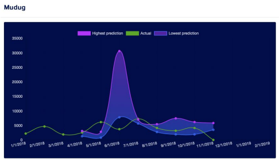
Sensitivity analysis

NPV, $ million

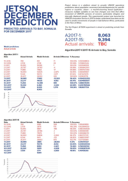
Variable two \ Variable one	-20%	-10%	Base case	+10%	+20%
-20%	60	70	80	90	100
-10%	70	80	90	100	110
Base case	80	90	100	110	120
+10%	90	100	110	120	130
+20%	100	110	120	130	140

Variables sensitivity analysis and/or other explainability elements (unravelling A.I. black box)

Automated visualization

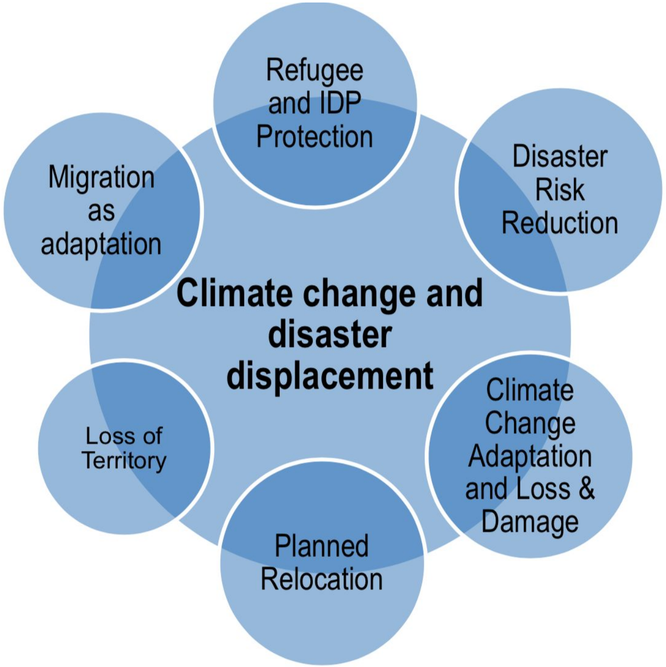


Automated reporting/publishing



2020: Climate Change and Forced Displacement

- Myth-busting of 'climate refugees'
- 1951 Refugee Convention defines a refugee as:
  - *someone with "a well-founded fear of persecution" on the basis of one of five grounds: race, religion, nationality, membership in a particular social group, and political opinion*
- Nonetheless, persons displaced in CCDD contexts may be refugees
  - Nexus with conflict
  - Denial of disaster assistance to minority group
- Regional instruments :*"events seriously disturbing public order"* in OAU Convention, Cartagena Declaration



Policy Space: Research and Knowledge Production

[Climate Change and Disaster Displacement Report, Chapter 4, section 2](#)

4.2 IMPROVE TOOLS AND METHODS

UNHCR also improves tools and methods utilized to collect and analyse displacement data to consider disaster dimensions. For example, as part of the steering group for the Mixed Migration Monitoring Mechanism Initiative (4Mi) in the Horn of Africa region, UNHCR provides inputs on additional questions about role of disasters as secondary drivers of displacement.

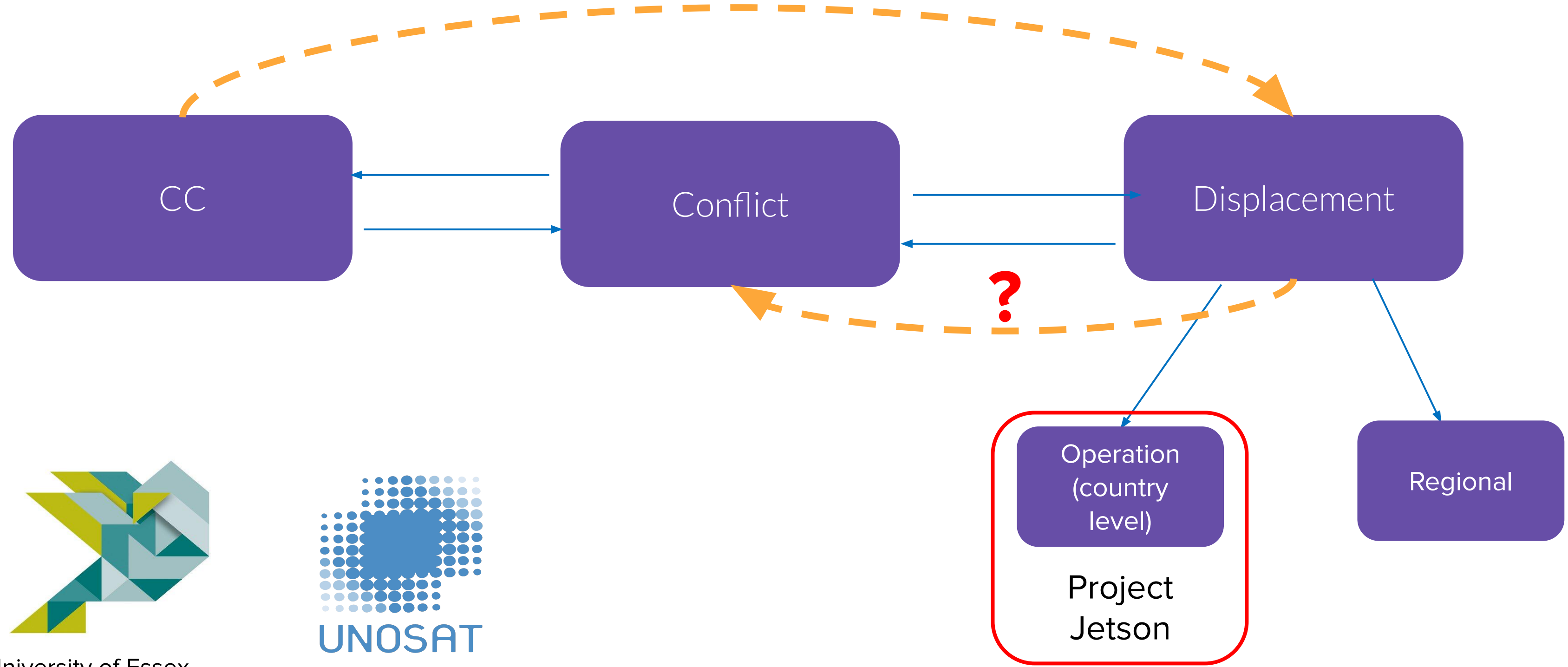
4.3 FOSTER DIALOGUE AND COORDINATION OF ACADEMIC AND POLICY COMMUNITIES

Notably, UNHCR actively participated in research dialogue and coordination through the Consultative Committee of the Nansen Initiative and now participates in the Advisory Committee of the Platform on Disaster Displacement. Additionally, in 2016 UNHCR participated in the launch of a new international Association for Study of Environmental Migration, and the launch of the 'Hugo Observatory', a research center dedicated to environment and mobility in Liege Belgium.

A motorbike travels along a flooded road in Jhati Tersil, Thatta District, Sindh, Pakistan on September 28, 2011. ; In August 2011, Heavy monsoon rains triggered flooding in lower parts of Sindh and northern parts of Punjab, making it difficult to quickly deliver aid to flood affected communities due to damaged roads. To date, the Government of Pakistan reports that more than 5.3 million people have been affected, Over 300 people have lost their lives, over 4.2 million acres of land flooded and 1.59 million acres of crops destroyed.
© UNHCR/Sam Phelps

UNHCR also contributes to disperse and communicate research outcomes with the broader academic and public community. For instance, with the support of the European Union, UNHCR commissioned an Issue of the Forced Migration Review (FMR 49) on 'Disasters and displacement in a changing climate' that published in May 2015 and includes 36 articles that gather latest available evidence on climate change, disasters and displacement.

UNHCR Innovation [ongoing]: finding the nexus
Machine-learning meets satellite data

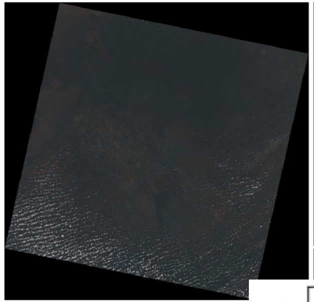


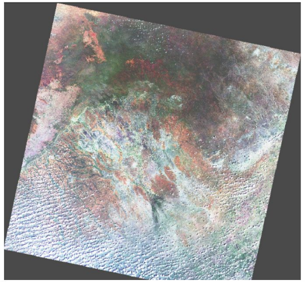
University of Essex
HRBDT Centre

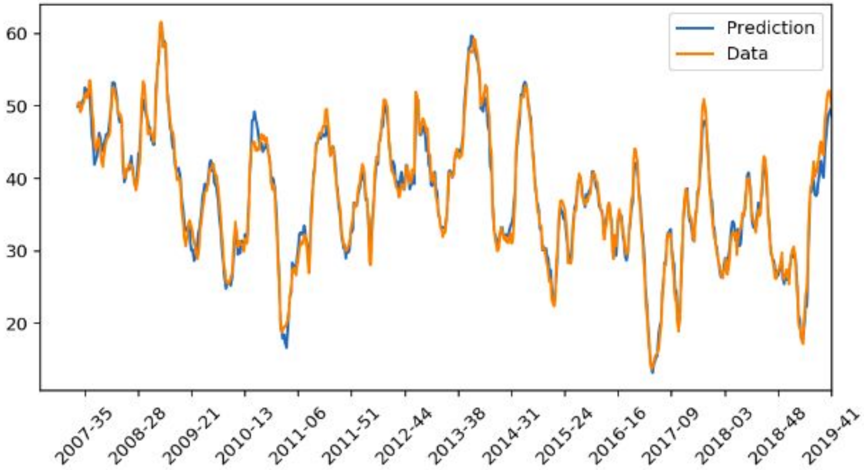
UNOSAT

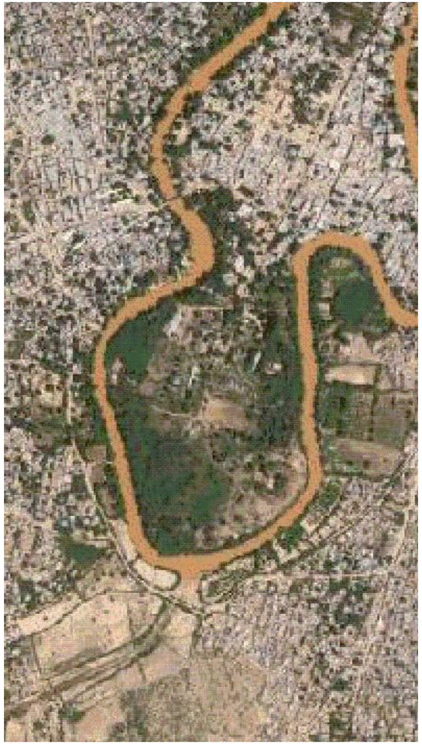
2020 Experiment #3: finding the nexus between climate, conflict & displacement using satellite imagery using NN

(in partnership with Omdena Foundation: Displacement Challenge)









Predictive Analytics: *Project Jetson*

- **Output:** IDP and refugee displacement figures (# of people)
- **Target variable:** number of displaced
- **PA Method:** ML-based supervised time series analysis
- **Data:** UNHCR-DRC PRMN, ACLED, FAO (SWALIM, FSNAU)
- **Research partners:** UN Global Pulse, University of Essex, WMO

Privacy

Europe is using smartphone data as a weapon to deport refugees

European leaders need to bring immigration numbers down, and metadata on smartphones could be just what they need to start sending migrants back

For future reflection: new trends in forced displacement

Recommendations for PA in forced displacement

ALL require both technical capacity (skills, data) and inter-agency coordination

1. **Recommendation #1:** collect and systematically publish different data sources (stress variables) in **one central repository with a <u>public facing-version of the figures</u>**.
    i. e.g. use FSNAU dashboard (a public-facing version) and/or use HDX platform to create this public version with an API.

2. **Recommendation #2:** Collect, anonymize and aggregate 1) **cash-assistance data and 2) remittances data** [7 years needed, only 2018 year available in HDX for cash and *nothing* available for remittances, both from WB or Dahabshiil]. Assumption: pull/push factor for forced displacement
    a. ideally also include non-traditional datasets or big data (e.g. satellite imagery of UNOSAT)

3. **Recommendation #3:** work cross-borders and cross-regions. Have regular communication and data exchange with your agency team's in the other side of the border/region. Alert them if something unusual is happening. Publish these alerts and its data in an internal-facing dashboard (e.g. FSNAU)

4. **Recommendation #4:** use non-traditional data sources to find additional movement. Use satellite imagery and/or work with mobile network operators to release anonymized call detail records data (CRDs) to understand movement. A good example is Turktelekom challenge.

OCHA Humanitarian Data Centre: PA stream

centre for humdata

FOCUS AREAS STORIES PRESS RESOURCES ABOUT US GET INVOLVED

Catalogue Of Predictive Models In The Humanitarian Sector

To improve transparency, the Centre has created a catalogue of predictive models with information on 'who is doing what, where and when'. We hope that this effort will make it easier for partners to get a quick overview of the models that are available and their current state of development. If your organization would like to be listed in the catalogue, please fill in this form.

Partners involved *: **Topic *:** CLEAR FILTERS

Search: **Show** All **entries**

Model name	Partners involved *	Topic *	Geographical scope *
510 Typhoon impact model	Netherlands Red Cross - 510	Coordination and context	The Philippines
African Risk Capacity	African Union, Africa Risk Capacity	Affected people	33 African Union member states
Artemis Model	World Bank	Food security	Pilot in 5 countries to be expanded to 21 additional countries
Asset Impact Monitoring System (AIMS)	WFP	Geography and infrastructure	Pilots in Niger, Aghanistan, Sudan, South Sudan, Tajikistan
Conflict Early Warning and Response Mechanism (CEWARN)	Intergovernmental Authority on Development (IGAD)	Conflict Migration	Djibouti, Eritrea, Ethiopia, Kenya, Somalia, South Sudan, Sudan, Uganda
Crisis Computing/ Artificial Intelligence for Digital Response	Qatar Computing Research Institute	Coordination and context	Global coverage
CrisisWatch	The International Crisis Group	Conflict	Global coverage

PEER REVIEW FRAMEWORK FOR
PREDICTIVE ANALYTICS IN
HUMANITARIAN RESPONSE

MARCH 2020

THE CENTRE FOR
HUMANITARIAN DATA

OCHA centre for humdata

UNHCR
The UN Refugee Agency

Thank you

unhcr.org/innovation
innovation@unhcr.org
jetson.unhcr.org