Climate Migration

Modeling human movement

Assignments

Keep working on your projects!

Each team will need to check in with me during our next class.

Discussion question: What is the most impactful or useful thing you learned in this class?

Due Friday 5pm

Advice on data preprocessing

Do the bare minimum to be able to run a model

Then tweak and feature engineer from there

➢ You should be training models by now!!

Climate change in the news

Climate change in the news

Climate advocates call for NYU trustee Larry Fink's expulsion from board

A group of student climate activists called on NYU to expel a member of its board of trustees for his company's investments in fossil fuels and private prisons.

Sheridan Smith

https://nyunews.com/news/2023/04/21/laurence-fink-nyu-trustees-removal-blackrock/

"On the West 4th Street Starbucks window are the words: 'Sustainability is not a spectator sport' and yet NYU's administration, and you, the board, are sitting back and watching us do all the work," the petition reads. "It is time NYU stops placing the burden of climate action on students, who individually have little impact on greenhouse gas emissions."

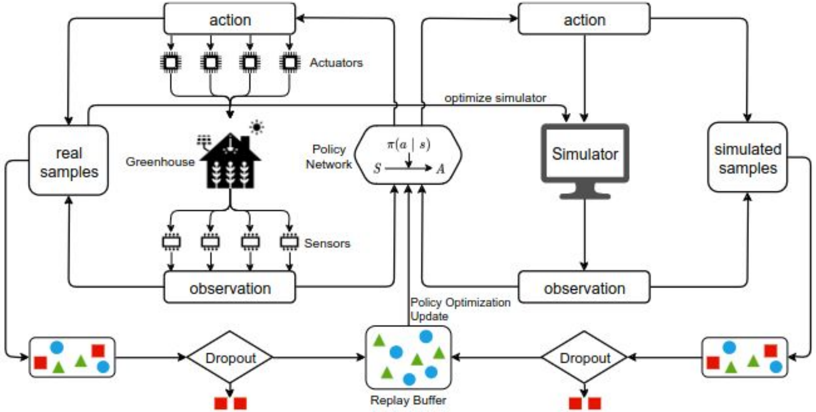
The protest was organized by the Climate Care Collective NYC, a group of climate advocates, some of whom are NYU students. Gigi Weisberg, a sophomore at NYU who is a member of the group, said she thinks NYU is not doing enough to fight climate change, and that the university's 2040 Now week feels insincere due to Fink's association with BlackRock.

BlackRock's investments are tied to hundreds of millions of tons of greenhouse gas emissions, and it is one of the largest shareholders in the world, with $8.5 trillion in assets under management. Fink's company is also the largest shareholder of both CoreCivic and Geo Group, which operate private prisons and immigrant detention centers around the country and the world. Currently, BlackRock owns 16.34% of CoreCivic and 14.67% of Geo Group, holdings that constitute over $316 million in shares for the two companies combined.

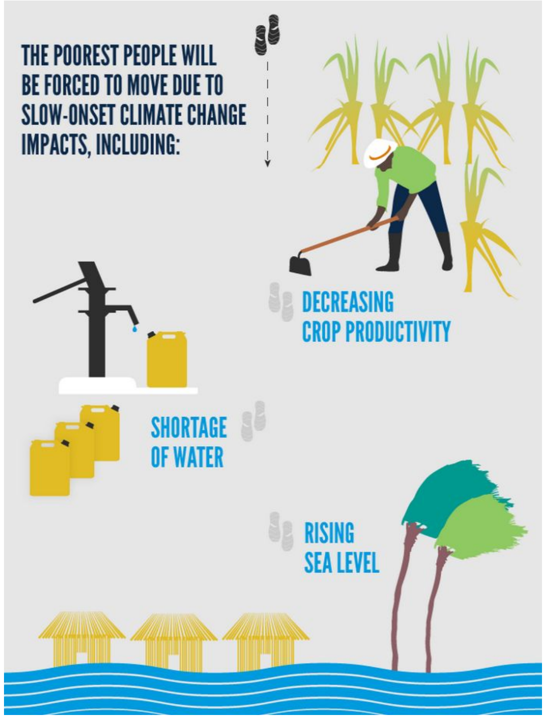
In recent years, BlackRock has set some sustainability goals, including reducing 67% of emissions from directly owned sources and 40% from unowned but indirectly influenced assets by 2030. The company also recently published a report of its climate-related finances, including its greenhouse gas emissions impact.

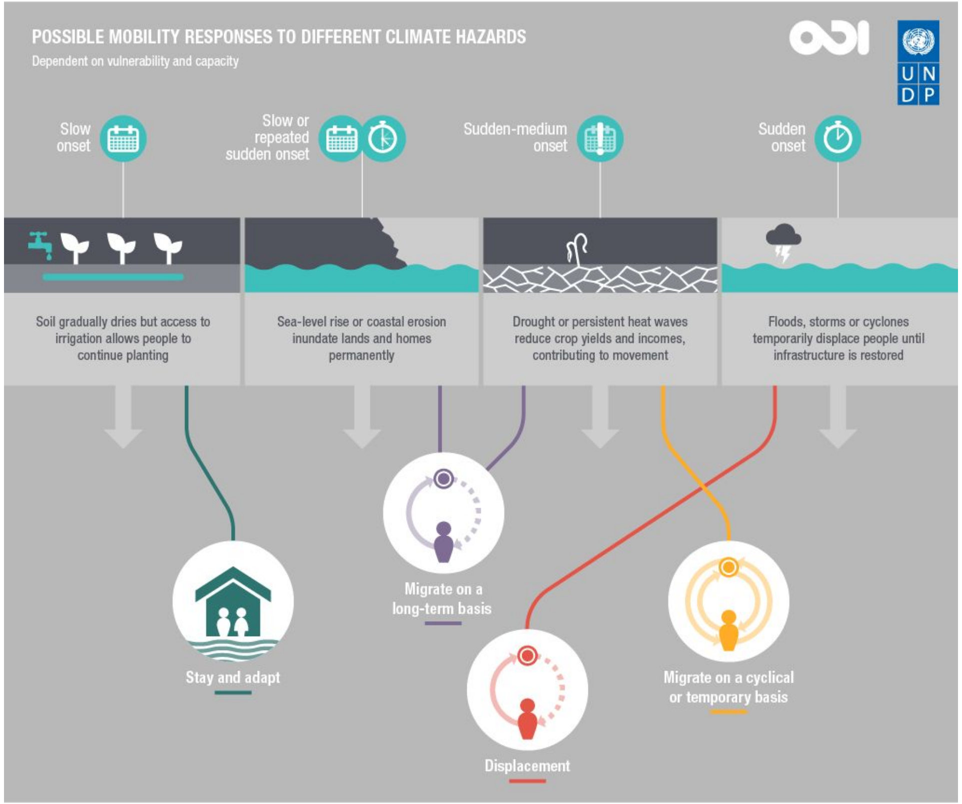
Recap


National Farm Safety and Health Week September 18-24, 2022
Top 5
Farm-Related Injuries
1 Overturning tractors and heavy machinery
2 Falls
3 Toxic chemical exposure to pesticides
4 Suffocation
5 Heat stress
$8.3 Billion
Annual costs of occupational injuries in agriculture**
100 ag workers
daily suffer a lost-work-time injury*
410
farmworkers who died in a work-related injury In 2019
70,000
suffered heat-related injuries during last 25 years*
20x
U.S. crop workers more likely to die of heat-related illness*
EFI provides training and a place for workers to have a voice in the farming operation. Worker-manager collaborative teams play a vital role in creating safer and healthier workplaces. Learn more at equitablefood.org.
Sources: *National Institute for Occupational Safety and Health (NIOSH); **OSHA


action
Actuators
real samples
Greenhouse
Sensors
observation
optimize simulator
Policy Network
$\pi(a \mid s)$
$S \rightarrow A$
action
Simulator
simulated samples
observation
Policy Optimization Update
Dropout
Replay Buffer
Dropout

Climate migration





Occurs when climate change or its consequences make a region unlivable for a person

Climate Migration is already happening

In Southeast Asia, where increasingly unpredictable monsoon rainfall and drought have made farming more difficult, the World Bank points to **more than eight million people** who have moved toward the Middle East, Europe and North America.

In the African Sahel, **millions of rural people** have been streaming toward the coasts and the cities amid drought and widespread crop failures.

https://www.nytimes.com/interactive/2020/07/23/magazine/climate-migration.html

Migration is politically fraught

An injection of new people into an aging work force could be to everyone’s benefit.

Northern nations can relieve pressures on the fastest-warming countries by ***allowing more migrants*** to move north across their borders.

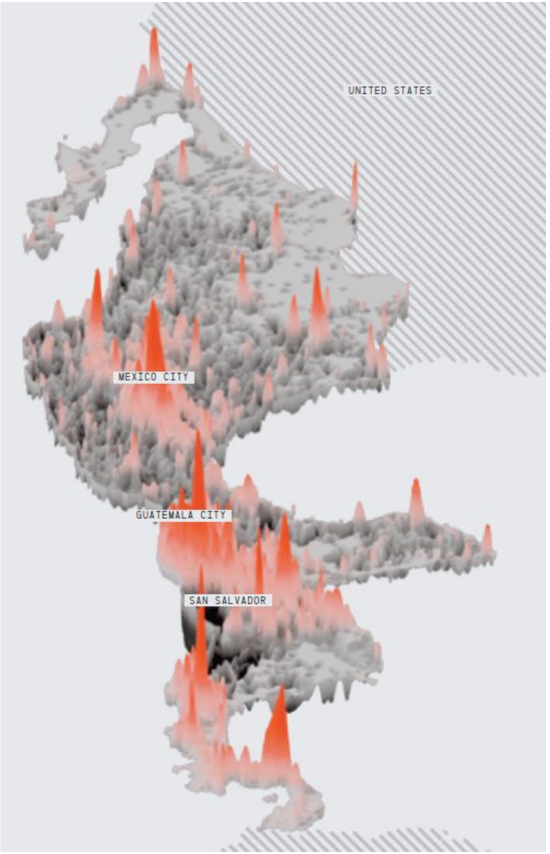
The best outcome requires not only good will and the careful management of turbulent political forces; without preparation and planning, ***the sweeping scale of change could prove wildly destabilizing***.

https://www.nytimes.com/interactive/2020/07/23/magazine/climate-migration.html

Migration grows as temperatures do

If governments take modest action to reduce climate emissions, about ***680,000 climate migrants*** might move from Central America and Mexico to the United States between now and 2050.

If emissions continue unabated, leading to more extreme warming, that number jumps to ***more than a million people***. (None of these figures include undocumented immigrants, whose numbers could be twice as high.)



https://www.nytimes.com/interactive/2020/07/23/magazine/climate-migration.html

Planning can help

Planning can help communities prepare and adapt to changing climate, or prepare to accept migrants.

Paper Deep Dive

PLOS ONE

OPEN ACCESS PEER-REVIEWED

RESEARCH ARTICLE

Modeling migration patterns in the USA under sea level rise

Caleb Robinson, Bistra Dilkina, Juan Moreno-Cruz

Published: January 22, 2020 • https://doi.org/10.1371/journal.pone.0227436

https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0227436

Motivation

Human migration is a natural response to these climate change pressures, and is one of many adaptation measures that people will take in response to climate change.

Understating how human migration will be affected by climate change is therefore a critical input in the ***decision making process of many governments and organizations***.

In particular, it is important to understand how ***climate change driven migration will differ from "business as usual"*** forms and motivations humans have to migrate.

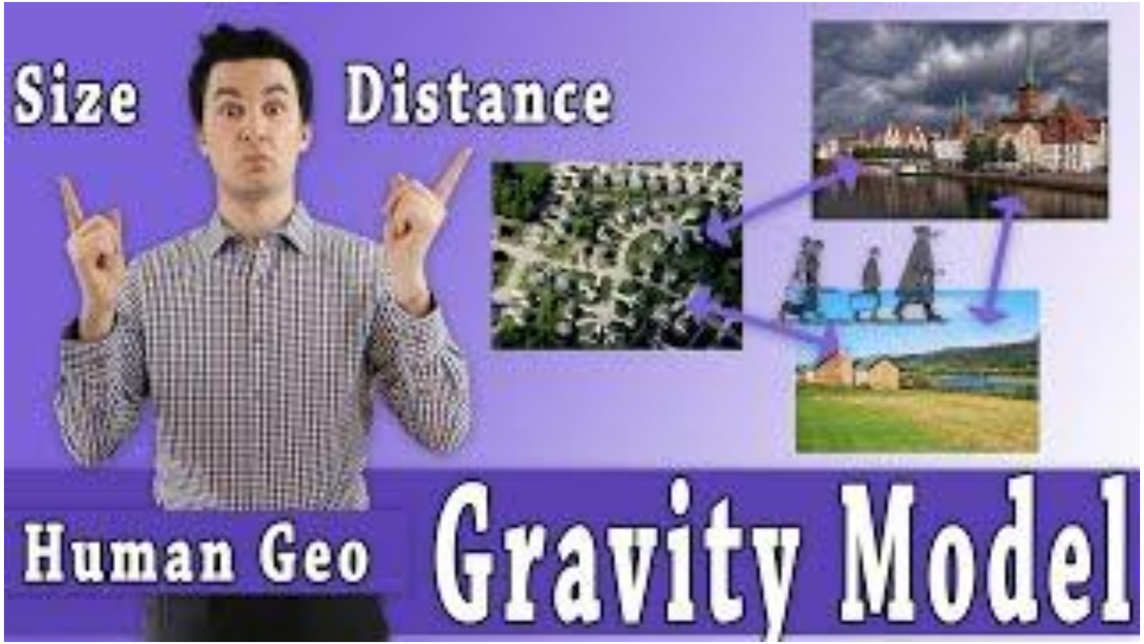
Background

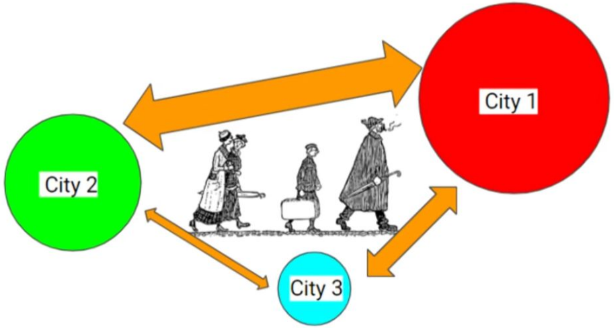
Sea level rise (SLR) will affect millions of people living in coastal areas.

Different studies have highlighted likely scenarios of sea level rise by 2100, varying in their projections of severity. According to the IPCC Fifth Assessment Report, in the “worst-case” Representative Concentration Pathways (RCP) scenario, RCP 8.5, where greenhouse gas emissions continue to rise throughout the 21st century, a global mean sea level (GMSL) rise between **0.52 to 0.98 meters (m) is likely by 2100** [11]. Other estimates, using statistical instead of process based models of GMSL, project a rise in the range of **0.75 m to 1.9** m by 2100 [12]. Recent research from the National Oceanic and Atmospheric Administration (NOAA) has even suggested a **2.5 m upper bound** of GMSL rise by 2100 for an ‘extreme’ SLR scenario, and a 2 m GMSL rise for a ‘high’ scenario [13].

By the year 2100, a projected **13.1 million people in the United States** alone would be living on land that will be considered flooded with a SLR of 6 feet (1.8 m)

Background

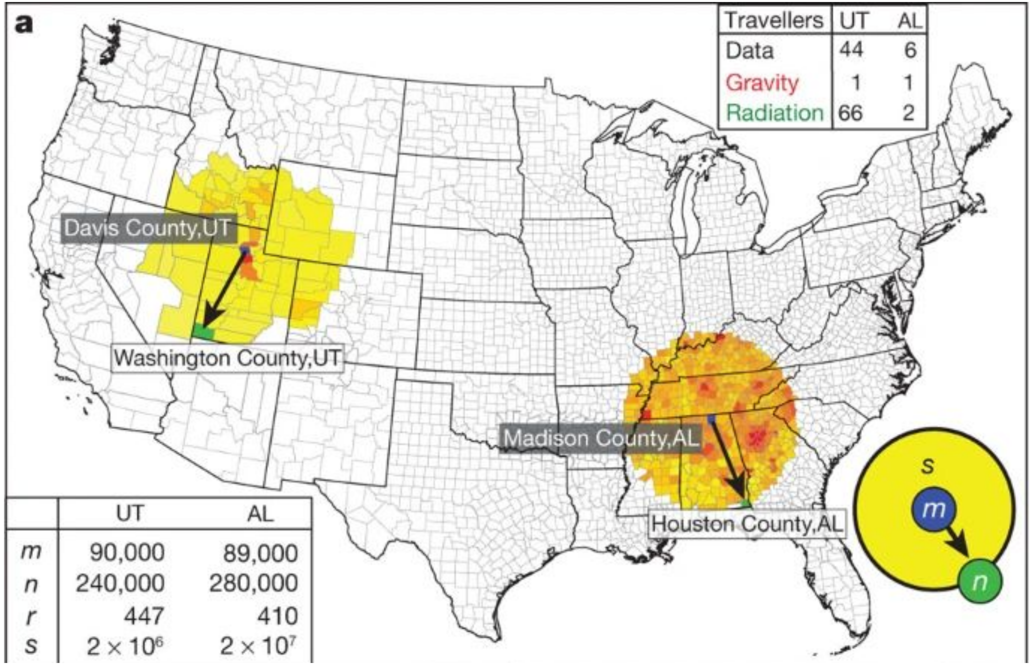




Existing methods don't incorporate all possible features

Background

Radiation models take in information about intervening locations



Goal

In particular, our framework assesses the broader impacts of climate change on population, by explicitly considering the effects on migration on populations ***directly affected by climate change and indirectly affected by the change*** in migration patterns induced by climate change.

Our framework, by incorporating separate models of migration choices for climate change driven versus “business-as-usual” migration, recognizes that **the patterns of climate migrants will not necessarily follow patterns observed in historical migration data**.

Brainstorm

What kind of data would you want to have to be able to approach this problem?

What kind of methods would you apply?

How would you measure success?

What difficulties might you face?

Framing the problem

For a given time, t, we will consider climate change features, $\mathbf{x}^t$, and a list of n spatial zones, $\boldsymbol{\theta}^t = [\theta_1^t, \ldots, \theta_n^t]$, which includes a spatial definition and features associated with each zone. Using this information, we want to compute a *migration matrix* $\boldsymbol{T}^t$, where an entry T_{ij}^t represents the number of migrants that travel from zone i to j at time t under a given climate impact model.

To:

From:

	City A	City B	City C	City D	City E
Cty A	0.69	0.03	0.15	0.03	0.11
City B	0.41	0.38	0.12	0.00	0.09
City C	0.11	0.24	0.50	0.00	0.15
City D	0.30	0.00	0.25	0.32	0.14
City E	0.22	0.06	0.20	0.06	0.46

Framing the problem

Break the problem down into 2 modules:

`CLIMATE` module. This module uses a "climate impacts model", *CLIMATE*($\boldsymbol{\theta}$, $\mathbf{x}$) to partition each input zone into two new zones: the affected portion and the unaffected portion. Using the best available data, this module should also divide the *features* from the original zone (θ_i) between the affected-portion zone (θ_i^A) and the unaffected-portion zone (θ_i^U). For example, if we have high-resolution spatial population data, then we can split population between the two partitions based on the spatial extent of flooding given a SLR model.

The Climate module identifies the locations (and corresponding populations) that will be affected by sea level rise

Framing the problem

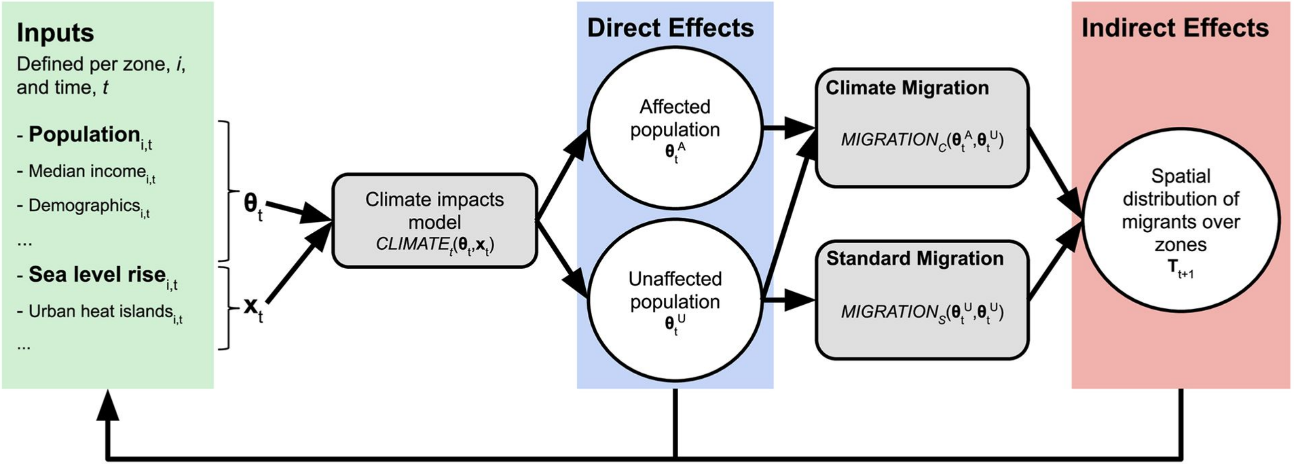
Break the problem down into 2 modules:

The Migration module calculates the transition probability by considering regular migration and migration from sea level rise.

MIGRATION module. This module calculates $\boldsymbol{T}$ using the two sets of zones from the CLIMATE module. Specifically, this module uses two migration models: 1.) a model for migrations from affected zones $\boldsymbol{\theta}^A$ to unaffected zones $\boldsymbol{\theta}^U$ with the function $MIGRATION_C(\boldsymbol{\theta}^A, \boldsymbol{\theta}^U) = \mathbf{T}'$, where migration is a forced process driven by climate change; and 2.) a model of migrations from unaffected zones to unaffected zones with the function $MIGRATION_S(\boldsymbol{\theta}^U, \boldsymbol{\theta}^U) = \mathbf{T}''$, where migration happens as usual. Finally, this module should aggregate migrant flows from the two migration functions, $\mathbf{T} = \mathbf{T}' + \mathbf{T}''$.

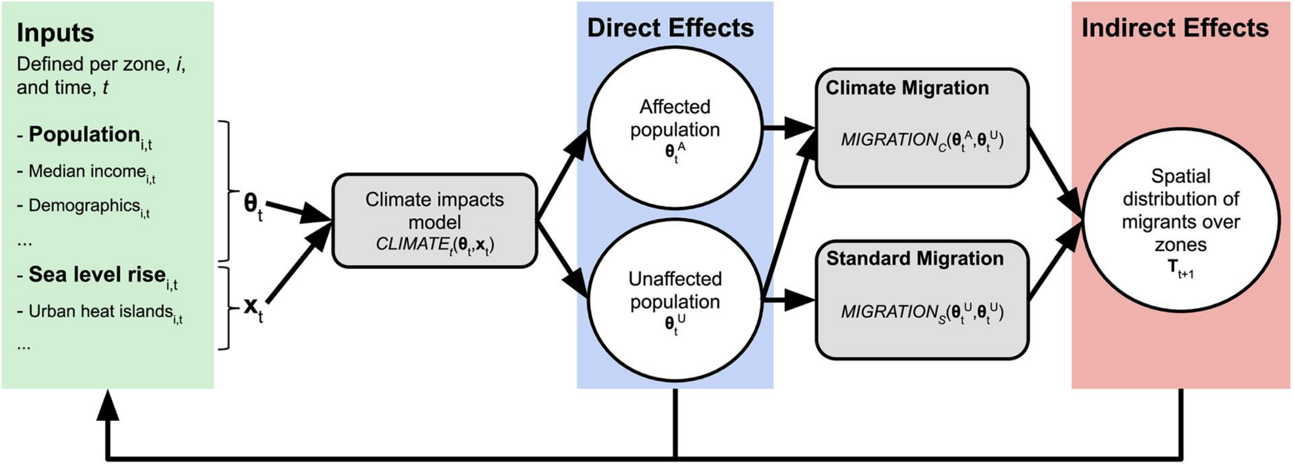
Framing the problem

Break the problem down into 2 modules:



Framing the problem

Break the problem down into 2 modules:



Benefits of this framing: can use different models/data for each section, explicitly limits migration to affected regions, can separately study the impacts of each section, recursive nature models long term and indirect impacts

Framing the problem

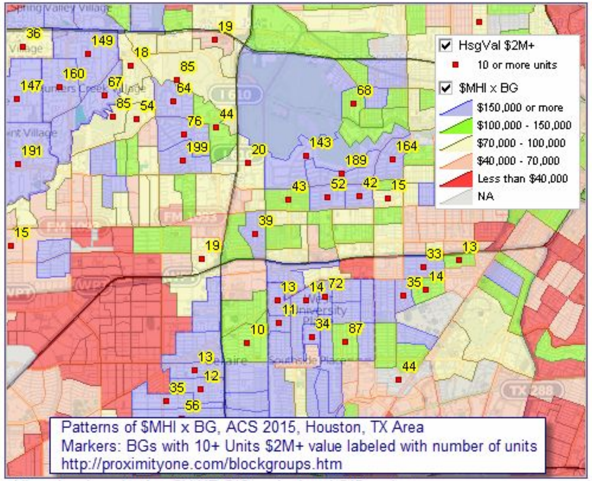
Units = census block groups

Census block groups are the smallest geographical unit for which the U.S. census bureau publishes sample data. Typically, Block Groups have a population of **600 to 3,000 people**. In this data year, there are 216,330 block groups across the U.S.

Can aggregate on county level (3,143).

Patterns of Economic Prosperity by Block Group

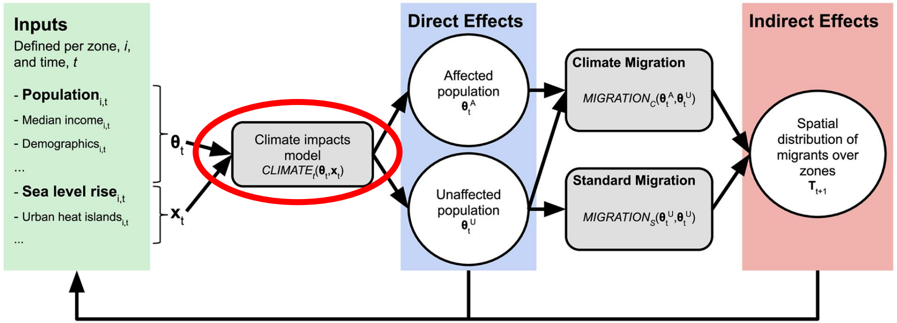
The following graphic shows patterns of median household income by block group in the Houston, TX area. Markers show block groups with 10 or more housing units having value of $2 million or more. Markers are labeled with the number of housing units having value of $2 million or more in that block group. Click graphic for larger view, more detail and legend color/data intervals. This map illustrates the geographic level of detail available using block group demographics and the relative ease to gain insights using geospatial data analytics tools.



- View developed using CV XE GIS and related GIS project.

Models

Climate Impacts Model: turn information about sea level rise into an estimate of the number of affected and unaffected people in each census block



Models

Climate Impacts Model: turn information about sea level rise into an estimate of the number of affected and unaffected people in each census block

Use the Digital Coast dataset. Creates a detailed map of flooding for a given global sea level rise based on land and sea features.

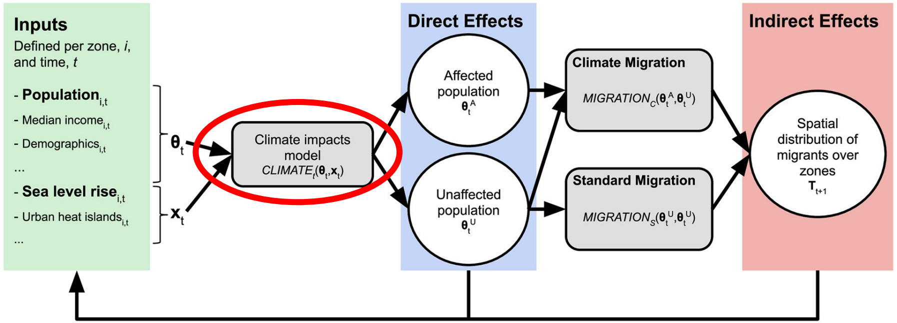
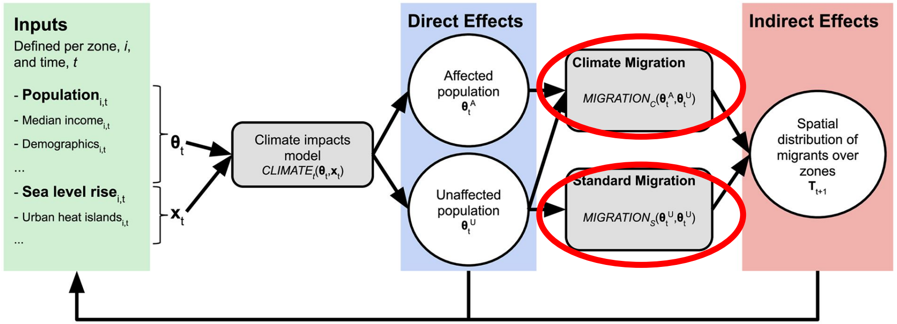


Figure 1. SLR and Coastal Flooding Impacts Viewer showing 4 feet (1.2 meters) of SLR above MHHW in Galveston, TX. Local impacts of this amount of water at local landmarks can be seen in simulation photos. This is one of 5 features of the tool. The tool can be accessed at the following URL: http://www.csc.noaa.gov/digitalcoast/tools/slrviewer/.

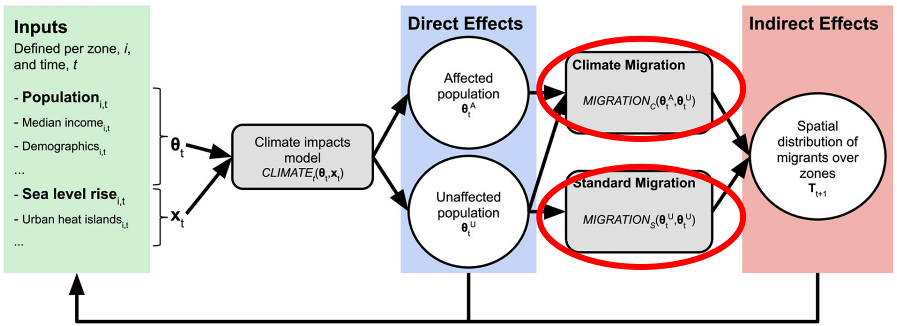
Models

Migration Models: estimate probability of person moving from county A to county B using county features



Models

Migration Models: estimate probability of person moving from county A to county B using county features



A Machine Learning Approach to Modeling Human Migration

Caleb Robinson
School of Computational Science and Engineering
Georgia Institute of Technology
Atlanta, Georgia
dcrobins@gatech.edu

Bistra Dilkina
Department of Computer Science
University of Southern California
Los Angeles, California
dilkina@usc.edu

https://arxiv.org/pdf/1711.05462.pdf

Model

A Machine Learning Approach to Modeling Human Migration

Caleb Robinson
School of Computational Science and Engineering
Georgia Institute of Technology
Atlanta, Georgia
dcrobins@gatech.edu

Bistra Dilkina
Department of Computer Science
University of Southern California
Los Angeles, California
dilkina@usc.edu

“Our goal is to estimate a function f, which takes the features of zone i and j, as well as the joint features between them, as input, and directly outputs the estimated number of migrants that travel from i to j.”

Compare:

XGBoost and **ANN**s to traditional models (radiation and gravity)

Model

A Machine Learning Approach to Modeling Human Migration

Caleb Robinson
School of Computational Science and Engineering
Georgia Institute of Technology
Atlanta, Georgia
dcrobins@gatech.edu

Bistra Dilkina
Department of Computer Science
University of Southern California
Los Angeles, California
dilkina@usc.edu

ANN loss function is based on migration evaluation metric, CPC

Common Part of Commuters (*CPC*) This metric directly compares numbers of travelers between the predicted and ground truth matrices. It will be 0 when the two matrices have no entries in common, and 1 when they are identical. We note that this metric, as used in previous studies of commuting flows, is identical to the Bray-Curtis similarity score used to compare abundance data in ecological studies [9, 15].

$$CPC(\mathbf{T}, \hat{\mathbf{T}}) = \frac{2\sum_{i,j=1}^{n} min(T_{ij}, \hat{T}_{ij})}{\sum_{i,j=1}^{n} T_{ij} + \sum_{i,j=1}^{n} \hat{T}_{ij}} \quad (1)$$

with. This loss function is given as:

$$L(y, \hat{y}) = 1 - \frac{2\sum_{i=1}^{n} min(y_i, \hat{y}_i)}{\sum_{i=1}^{n} y_i + \sum_{i=1}^{n} \hat{y}_i} \quad (5)$$

Where y_i is a migration flow entry from $\mathbf{T}$. The gradient update for this loss function is:

This loss function was found to perform better than mean-squared error

Data

A Machine Learning Approach to Modeling Human Migration

Caleb Robinson
School of Computational Science and Engineering
Georgia Institute of Technology
Atlanta, Georgia
dcrobins@gatech.edu

Bistra Dilkina
Department of Computer Science
University of Southern California
Los Angeles, California
dilkina@usc.edu

The **USA Migration** dataset consists of yearly intra-county migrations in the USA between 3106 counties from the IRS Tax-Stats datafor the 11 years in the range from 2004 to 2014.

We supplement the migration data with the following **7 per-county features**: population, land area, population density, median household income, county water area, is a coastal county, and number of neighboring counties; and **6 between-county features**: distance, intervening population, intervening land area, intervening number of counties, intervening population density, and intervening median income.

The **Global Migration** dataset consists of decadal inter-country migration data between 207 countries from the World Bank Global Bilateral Migration Database. The Global Migration dataset contains 5 timesteps, one every 10 years from 1960 to 2000. In the Global Migration dataset we add the following **5 per-country features**: population, population density, population growth, agricultural emissions, and land area. Additionally, we include **3 between-country features**: distance, intervening population, and intervening land area.

Data Problem!

A Machine Learning Approach to Modeling Human Migration

Caleb Robinson
School of Computational Science and Engineering
Georgia Institute of Technology
Atlanta, Georgia
dcrobins@gatech.edu

Bistra Dilkina
Department of Computer Science
University of Southern California
Los Angeles, California
dilkina@usc.edu

“Migration data is heavily zero-inflated, where in any given year, most pairs of zones do not have any migrants traveling between them. For example, Considering migrations between US counties, less than 1% of the possible pairs of counties have migrations between them”

Data Problem!

A Machine Learning Approach to Modeling Human Migration

Caleb Robinson
School of Computational Science and Engineering
Georgia Institute of Technology
Atlanta, Georgia
dcrobins@gatech.edu

Bistra Dilkina
Department of Computer Science
University of Southern California
Los Angeles, California
dilkina@usc.edu

“Migration data is heavily zero-inflated, where in any given year, most pairs of zones do not have any migrants traveling between them. For example, Considering migrations between US counties, less than 1% of the possible pairs of counties have migrations between them”

To address this problem, when creating a training dataset we undersample “negative” samples between pairs of zones for which there are no observed migrations.
This is a naïve technique that will necessarily throw out available information. To offset this, we introduce a hyperparameter k that determines the number of “negative” examples of migrations to train with.

Training

A Machine Learning Approach to Modeling Human Migration

Caleb Robinson
School of Computational Science and Engineering
Georgia Institute of Technology
Atlanta, Georgia
dcrobins@gatech.edu

Bistra Dilkina
Department of Computer Science
University of Southern California
Los Angeles, California
dilkina@usc.edu

Hyperparameter search

For ANN: varying loss function, number of layers, layer width, number of training epochs, k, and training mini-batch, size.

For XGBoost: varying maximum tree depth, number of estimators, k, and learning rate.

To select the hyperparameters of the models that we described in Sections 2 and 4, we consider triplets of "years" of data as training, validation, and testing sets. Specifically, for three years of data $\{(X_{t-2}, Y_{t-2}), (X_{t-1}, Y_{t-1}), (X_t, Y_t)\}$, we call (X_{t-2}, Y_{t-2}) the training set, (X_{t-1}, Y_{t-1}) the validation set, and (X_t, Y_t) the test set. We

Evaluation

A Machine Learning Approach to Modeling Human Migration

Caleb Robinson
School of Computational Science and Engineering
Georgia Institute of Technology
Atlanta, Georgia
dcrobins@gatech.edu

Bistra Dilkina
Department of Computer Science
University of Southern California
Los Angeles, California
dilkina@usc.edu

Common Part of Commuters (*CPC*) This metric directly compares numbers of travelers between the predicted and ground truth matrices. It will be 0 when the two matrices have no entries in common, and 1 when they are identical. We note that this metric, as used in previous studies of commuting flows, is identical to the Bray-Curtis similarity score used to compare abundance data in ecological studies [9, 15].

$$CPC(\mathbf{T}, \hat{\mathbf{T}}) = \frac{2\sum_{i,j=1}^{n} min(T_{ij}, \hat{T}_{ij})}{\sum_{i,j=1}^{n} T_{ij} + \sum_{i,j=1}^{n} \hat{T}_{ij}} \tag{1}$$

Common Part of Commuters Distance Variant (CPC_d) This metric measures how well a predicted migration matrix recreates trips at the same distances as the ground truth data. In this definition, N is a histogram where a bin N_k contains the number of migrants that travel between $2k-2$ and $2k$ kilometers. It will be 0 when the two matrices do not have any migrations at the same distance, and 1 when all fall within the same distances.

$$CPC_d(\mathbf{T}, \hat{\mathbf{T}}) = \frac{2\sum_{k=1}^{\infty} min(N_k, \hat{N}_k)}{\sum_{k=1}^{\infty} N_k + \sum_{k=1}^{\infty} \hat{N}_k} \tag{2}$$

Root mean squared error (*RMSE*) This is a standard regression measure that will "punish" larger errors more than small errors. This score ranges from 0 in a perfect match, to arbitrarily large values as the predictions become worse.

$$RMSE(\mathbf{T}, \hat{\mathbf{T}}) = \sqrt{\frac{1}{n}\sum_{i,j=1}^{n}(T_{ij} - \hat{T}_{ij})^2} \tag{3}$$

Coefficient of determination (r^2) This score measures the goodness of fit between a set of predictions and the ground truth values. This score ranges from 1, in a perfect fit, to arbitrarily negative values as a fit becomes worse, and is 0 when the predictions are equivalent to the expectation of the ground truth values.

$$r^2(\mathbf{T}, \hat{\mathbf{T}}) = 1 - \frac{\sum_{i,j=1}^{n}(T_{ij} - \hat{T}_{ij})^2}{\sum_{i,j=1}^{n}(T_{ij} - \bar{T})^2} \tag{4}$$

Results

A Machine Learning Approach to Modeling Human Migration

Caleb Robinson
School of Computational Science and Engineering
Georgia Institute of Technology
Atlanta, Georgia
dcrobins@gatech.edu

Bistra Dilkina
Department of Computer Science
University of Southern California
Los Angeles, California
dilkina@usc.edu

USA Migrations	Metrics on full matrix			
Production Function	*CPC*	CPC_d	*RMSE*	r^2
Gravity Model Exponential Decay	0.53 +/- 0.01	0.66 +/- 0.02	87.4 +/- 9.0	-1.48 +/- 0.28
Gravity Model Power Law Decay	0.56 +/- 0.01	0.78 +/- 0.02	75.7 +/- 8.0	-0.86 +/- 0.26
Radiation Model	0.53 +/- 0.01	0.76 +/- 0.02	47.6 +/- 5.0	0.26 +/- 0.09
Extended Radiation Model	0.58 +/- 0.01	0.83 +/- 0.01	35.6 +/- 3.0	0.59 +/- 0.03
No Production Function	*CPC*	CPC_d	*RMSE*	r^2
XGBoost model - traditional features	0.54 +/- 0.11	**0.99 +/- 0.02**	18.5 +/- 6.1	0.88 +/- 0.08
ANN model - traditional features	0.63 +/- 0.02	0.88 +/- 0.06	35.3 +/- 3.5	0.60 +/- 0.04
XGBoost model - extended features	0.62 +/- 0.04	**0.99 +/- 0.02**	**13.0 +/- 1.5**	**0.94 +/- 0.02**
ANN model - extended features	**0.69 +/- 0.01**	0.93 +/- 0.05	28.0 +/- 3.6	0.75 +/- 0.03

Global Migrations	Metrics on full matrix			
Production Function	*CPC*	CPC_d	*RMSE*	r^2
Gravity Model Exponential Decay	0.16 +/- 0.00	0.16 +/- 0.00	62,218 +/- 5,341	0.02 +/- 0.03
Gravity Model Power Law Decay	0.16 +/- 0.00	0.15 +/- 0.00	61,523 +/- 5,278	0.05 +/- 0.00
Radiation Model	0.16 +/- 0.00	0.14 +/- 0.00	62,173 +/- 5,277	0.02 +/- 0.00
Extended Radiation Model	0.16 +/- 0.00	0.14 +/- 0.00	62,108 +/- 5,299	0.03 +/- 0.00
No Production Function	*CPC*	CPC_d	*RMSE*	r^2
XGBoost model - traditional features	0.33 +/- 0.02	0.59 +/- 0.03	52,729 +/- 5,455	0.26 +/- 0.26
ANN model - traditional features	0.33 +/- 0.01	0.37 +/- 0.04	56,005 +/- 882	0.20 +/- 0.11
XGBoost model - extended features	**0.43 +/- 0.03**	**0.64 +/- 0.02**	**47,329 +/- 5,073**	**0.42 +/- 0.13**
ANN model - extended features	0.40 +/- 0.02	0.43 +/- 0.02	50,921 +/- 3,556	0.33 +/- 0.13

Traditional features are only those normally used in gravity or radiation based models.

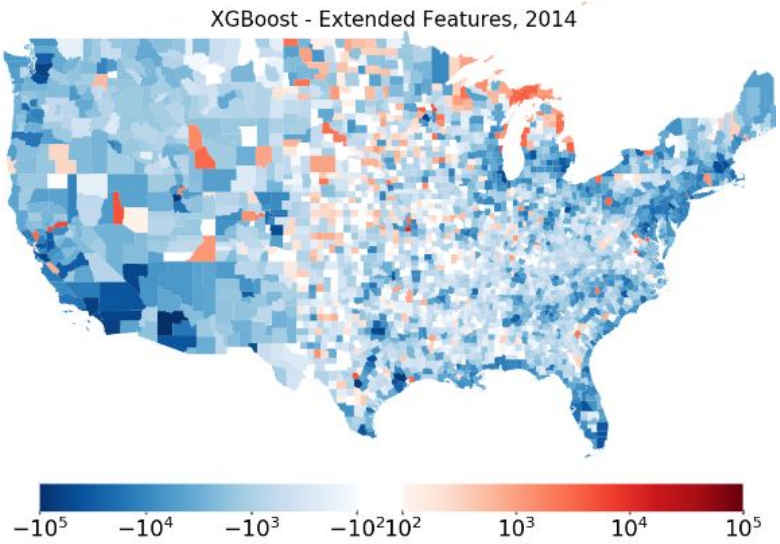
XGBoost with the full feature set is consistently high performing

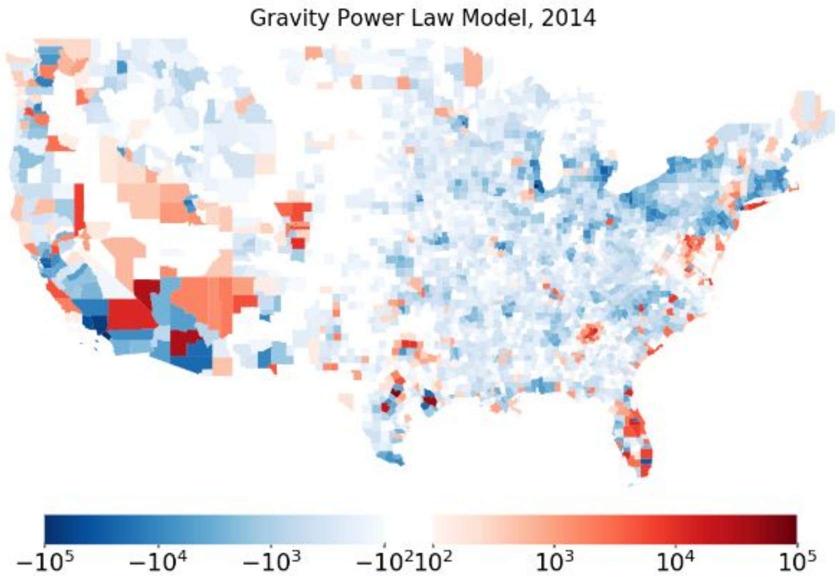
Results

A Machine Learning Approach to Modeling Human Migration

Caleb Robinson
School of Computational Science and Engineering
Georgia Institute of Technology
Atlanta, Georgia
dcrobins@gatech.edu

Bistra Dilkina
Department of Computer Science
University of Southern California
Los Angeles, California
dilkina@usc.edu





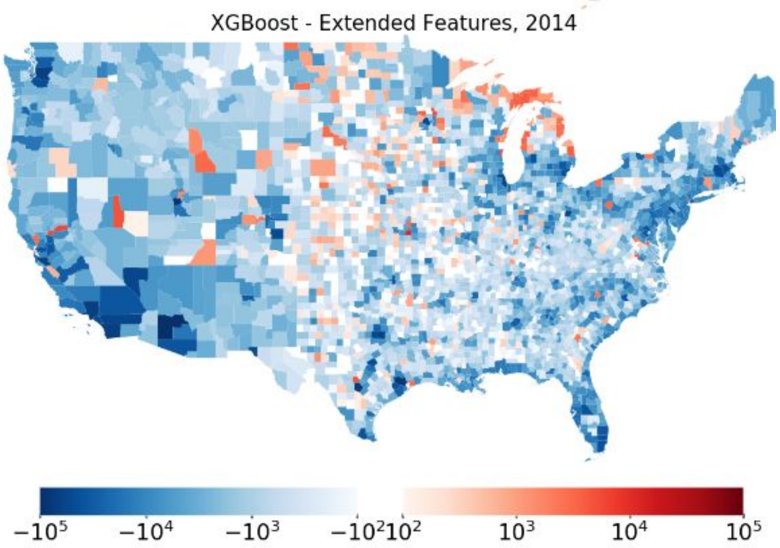
Different models make very different immigration errors

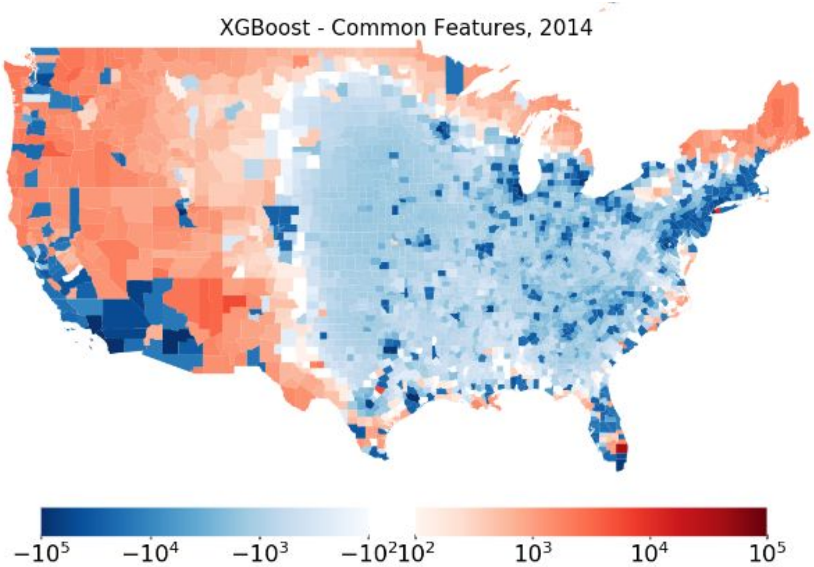
Results

A Machine Learning Approach to Modeling Human Migration

Caleb Robinson
School of Computational Science and Engineering
Georgia Institute of Technology
Atlanta, Georgia
dcrobins@gatech.edu

Bistra Dilkina
Department of Computer Science
University of Southern California
Los Angeles, California
dilkina@usc.edu





Comparing same model with different data

Results

A Machine Learning Approach to Modeling Human Migration

Caleb Robinson
School of Computational Science and Engineering
Georgia Institute of Technology
Atlanta, Georgia
dcrobins@gatech.edu

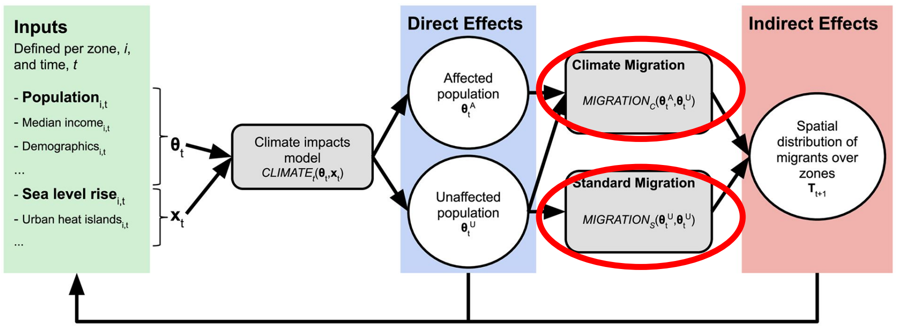
Bistra Dilkina
Department of Computer Science
University of Southern California
Los Angeles, California
dilkina@usc.edu

USA Features	Importance	Global Features	Importance
Intervening number of counties	25.3% +/- 2.4%	Population growth of origin	19.5% +/- 16.0%
Population of origin (trad)	15.7% +/- 1.7%	Intervening population (trad)	12.3% +/- 3.7%
Population of destination (trad)	14.2% +/- 0.9%	Agricultural emissions of destination	10.6% +/- 5.8%
Intervening population (trad)	6.1% +/- 1.2%	Intervening land area	8.7% +/- 5.6%
Is destination coastal	4.3% +/- 4.6%	Population growth of destination	7.9% +/- 6.2%
Distance between counties (trad)	3.7% +/- 0.9%	Population of destination (trad)	6.9% +/- 0.8%
Intervening area	3.6% +/- 0.9%	Distance between counties (trad)	6.6% +/- 1.9%
Area of destination	3.5% +/- 0.5%	Population of origin (trad)	6.1% +/- 1.6%
Number of neighbors destination	3% +/- 1.6%	Population density of destination	5.7% +/- 4.5%
Water area of origin	3% +/- 1.3%	Land area of origin	5.2% +/- 3.1%

Table 4: Top 10 most important (extended) features in both the *USA Migration* and *Global Migration* datasets. The values in the table show the average and standard deviations of the information gain feature importances from an XGBoost model trained on the extended feature set for each timestep of data.

Many of the important features are not traditionally included in migration models. ML models can accommodate any features.

Models



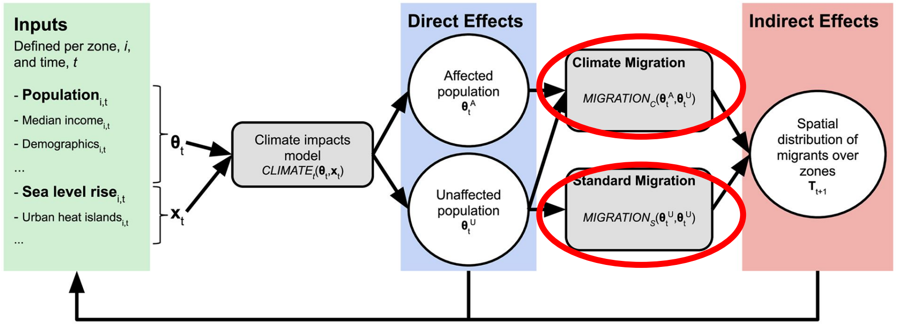
Migration Models: estimate probability of person moving from county A to county B using county features.

Climate migration model models migration from affected regions to unaffected regions

Standard migration model models migration from unaffected to unaffected regions

Models

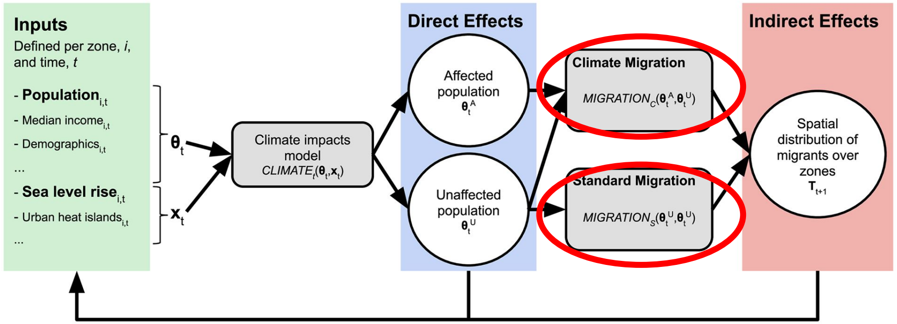
Migration Models: estimate probability of person moving from county A to county B using county features.



Climate migration model is fit on data from 7 counties severely affected by Hurricanes Katrina and Rita, from 2004-2014.

Models

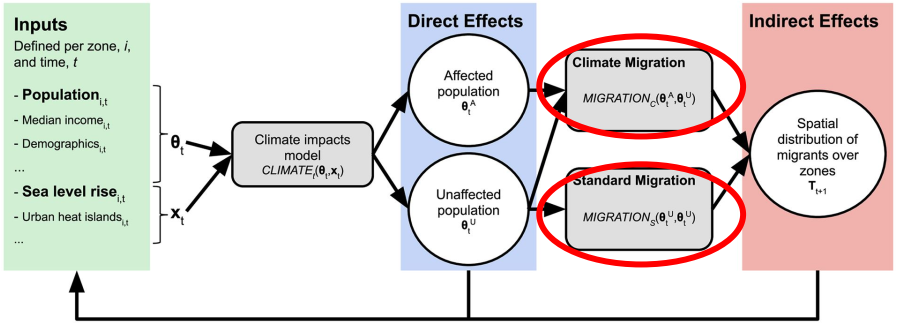
Migration Models: estimate probability of person moving from county A to county B using county features.



Standard migration model is fit on data from remaining counties, from 2004-2014.

Models

Migration Models: estimate probability of person moving from county A to county B using county features.



Both models are ANNs, which on average outperform standard gravity and radiation models.

XGBoost??

Evaluation

Individual models have been independently validated

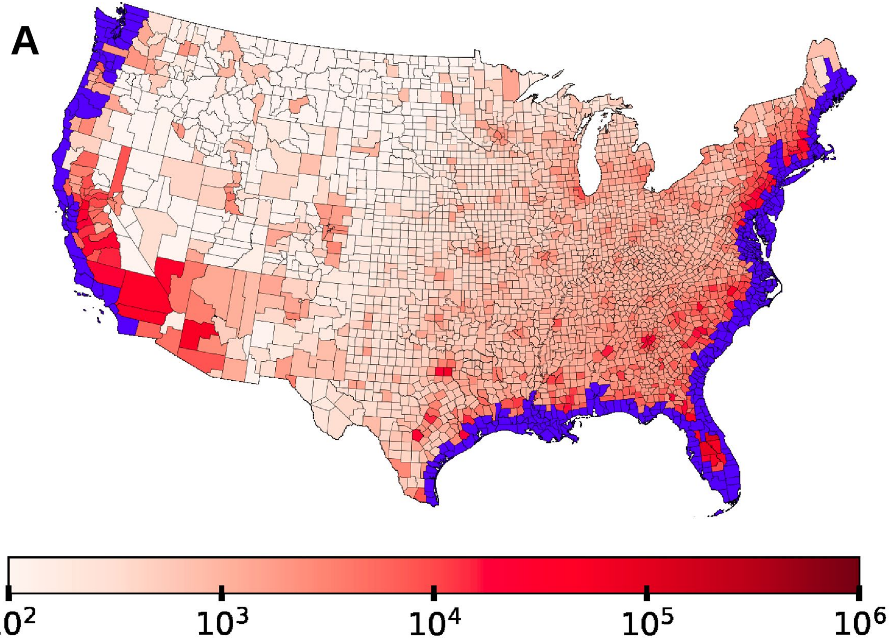
Results are described in terms of direct and indirect effects of sea level rise.

“A zone is marked as **indirectly** affected if the difference between the number of incoming migrants in the CLIMATE scenario and the baseline scenario is greater than a percentage *d*% of the population of that zone. By varying *d* we can see different intensities of indirect effects.”

Results

All counties that experience flooding under 1.8m of SLR by 2100 in blue

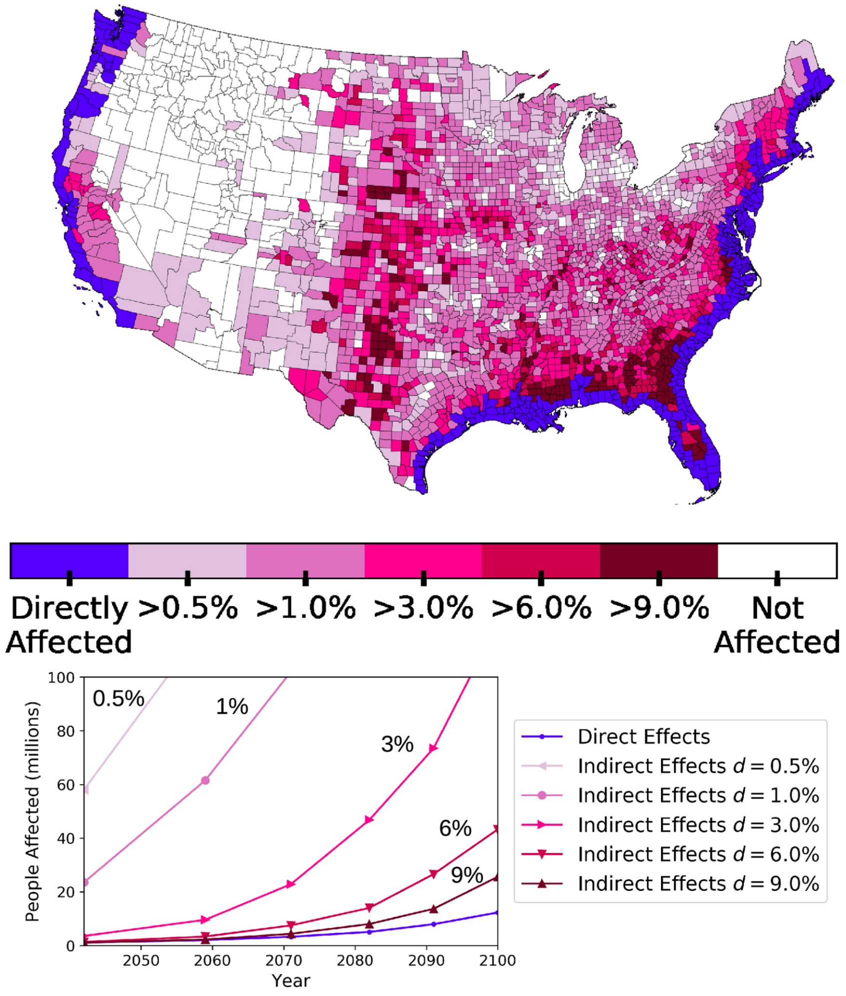
The remaining counties are colored based on the number of additional incoming migrants per county that there are in the SLR scenario over the baseline



1.8m sea rise will have strong impact on migration

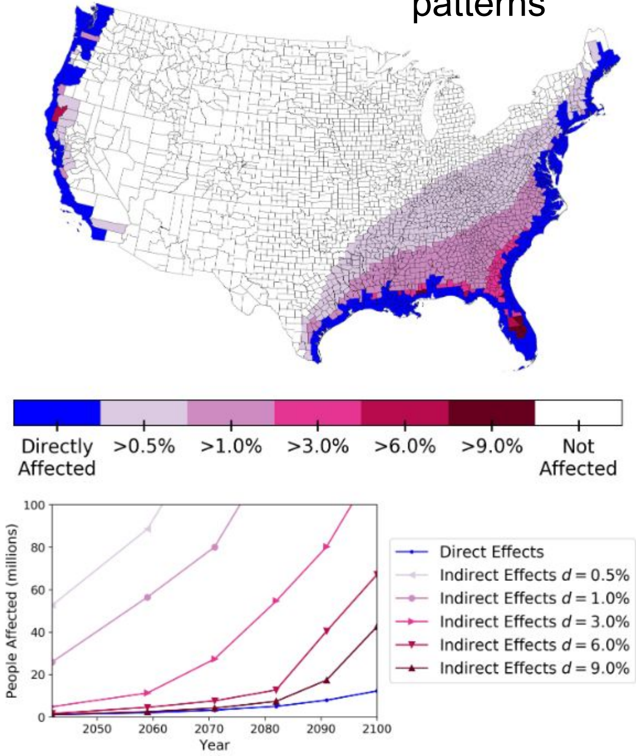
Results

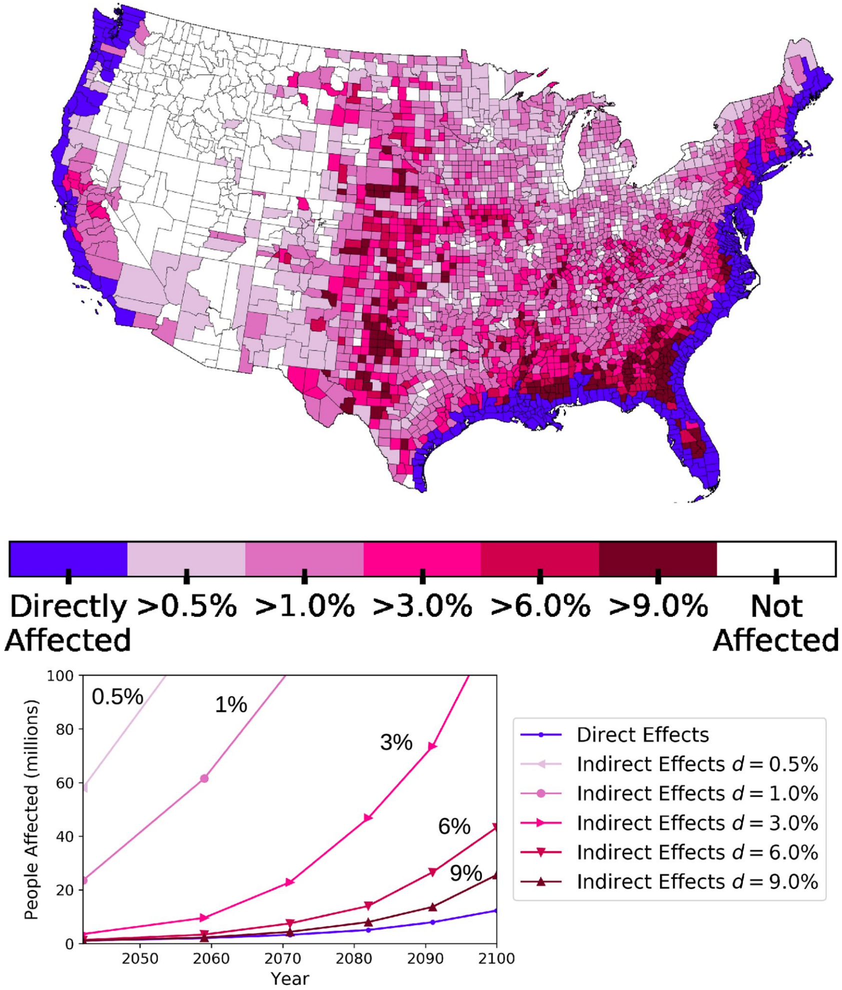
Much of the impact will be indirect



Results

Extended radiation model gives very different spatial patterns





Results

In general, we find that **previously “unpopular” migrant destinations** (areas with relatively low numbers of incoming migrants) would be more popular solely due to their close proximity to counties that experience “direct effects”.

The **East Coast will experience larger effects** than the West coast because of the large coastal population centers and shallower coastlines, indeed, all counties adjacent to coastal counties on the East coast are marked as *indirectly* affected.

Existing urban areas will receive the largest magnitudes of migrants, as they represent the most attractive destinations, which will **accelerate the existing trends of urbanization**.

Limitations

While our framework is flexible enough in theory, in practice it will require assumptions specific to each application. These assumptions can be the result of knowledge gaps, data limitations, or large uncertainties in the individual component models.

In our application to SLR, we have made several such assumptions that affect how we interpret the results of the coupled SLR/migration models. For example, we have **assumed that people move due to SLR only when their homes are permanently flooded**. It is of course possible, that people will move because their business or jobs are affected. According to the 'Nuisance Hypothesis' from Keenan et al. [40] housing prices are **affected by people's perceptions as to whether or not a property is at risk of flooding**. This could impact "pull" factors of migration.

While we cannot consider this channel in our current application due to data limitations, **our framework allows to expand in this direction once more data is available**.

Further resources

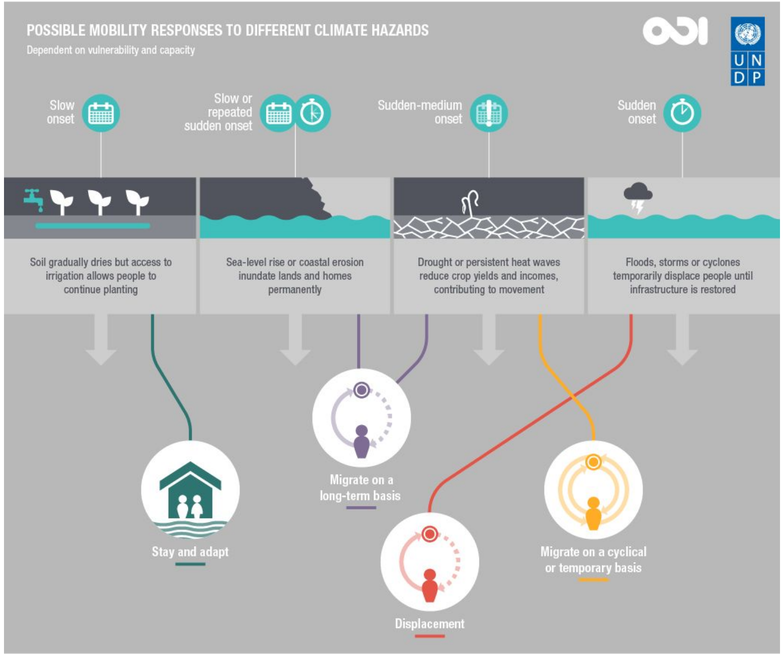
https://www.whitehouse.gov/wp-content/uploads/2021/10/Report-on-the-Impact-of-Climate-Change-on-Migration.pdf

ANNEX 2: EXAMPLES OF CLIMATIC TRENDS AND EXTREMES THAT INFLUENCE MIGRATION

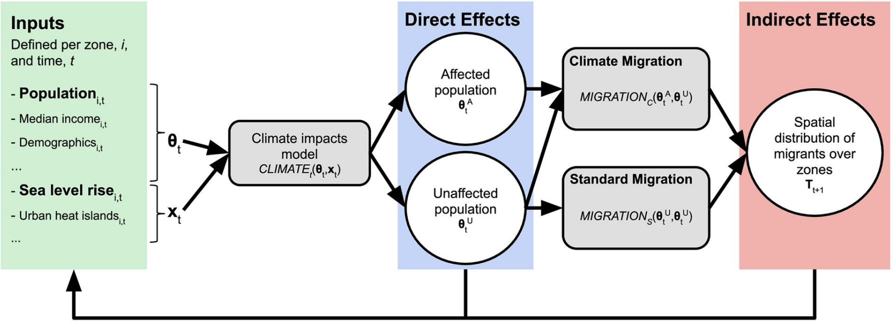
Manifestations of climate change	Environmental change	Pathways/effects that influence human security	Observed or potential influence on migration
Increasing warm days and nights	Decreases in surface and groundwater; desertification	Arable/grazing land degradation; drought stress on flora and fauna; lack of water for human settlements and agriculture; local economic decline; limitations to outdoor activities under extreme temperatures	Decline in pastoral land use, African Sahel; rural to urban migration in Malawi concurrent with increasing frequency and severity of drought since 1970
Heat waves increasing over land and ocean	Increase in temperature extremes, exceeding heat stress tolerance level of humans and ecosystems	Increase in excess death rate; impacts on food safety and changing ecological patterns of vector-borne, zoonotic, and environmentally sourced (e.g., from water-, soil-, or dust-borne pathogens) infectious diseases; increase in wildfire; coral bleaching events; more frequent harmful algal blooms	Heat wave deaths in India (2015), Europe (2019); impacts of weather extremes in highly vulnerable economies (e.g. Dominican Republic, Jamaica); agricultural land degradation
Increases in the intensity and duration of drought	Declining lake storage, streamflow, and groundwater	Water resource shortages and food insecurity; land degradation; reduction in crop, forest, and livestock production; increase in wildfire	Migration and conflict over water in Burkina Faso (ongoing); migration from drought-stricken lands in Ethiopia, Iraq and Somalia (2019)
Increase in heavy precipitation events	Flooding, erosion, channel modification, debris flows	Loss of life; impacts on homes and infrastructure; damage to crops and increase flood insecurity	Flash floods in Nepal (1993, 2020); monsoonal floods and abrupt migration in Bangladesh (1987-1988, 2004, 2007)

https://www.weforum.org/agenda/2022/11/how-ai-can-help-climate-migration/

Summary



POSSIBLE MOBILITY RESPONSES TO DIFFERENT CLIMATE HAZARDS
Dependent on vulnerability and capacity
ODI
UN DP
Slow onset
Slow or repeated sudden onset
Sudden-medium onset
Sudden onset
Soil gradually dries but access to irrigation allows people to continue planting
Sea-level rise or coastal erosion inundate lands and homes permanently
Drought or persistent heat waves reduce crop yields and incomes, contributing to movement
Floods, storms or cyclones temporarily displace people until infrastructure is restored
Stay and adapt
Migrate on a long-term basis
Displacement
Migrate on a cyclical or temporary basis



Inputs
Defined per zone, *i*, and time, *t*
- **Population**$_{i,t}$
- Median income$_{i,t}$
- Demographics$_{i,t}$
...
- **Sea level rise**$_{i,t}$
- Urban heat islands$_{i,t}$
...
θ_t
x_t
Climate impacts model
$CLIMATE_t(\theta_t, x_t)$
Direct Effects
Affected population θ_t^A
Unaffected population θ_t^U
Climate Migration
$MIGRATION_C(\theta_t^A, \theta_t^U)$
Standard Migration
$MIGRATION_S(\theta_t^U, \theta_t^U)$
Indirect Effects
Spatial distribution of migrants over zones T_{t+1}