# Modeling human migration patterns during drought conditions in La Guajira, Colombia

Sibren Isaacman
Loyola University Maryland
isaacman@cs.loyola.edu

Vanessa Frias-Martinez
College of Information Studies
University of Maryland
vfrias@umd.edu

Enrique Frias-Martinez
Telefonica Research
Madrid, Spain
enrique.friasmartinez@telefonica.com

## ABSTRACT

Modeling human mobility is key for a variety of applications such as migratory flows, epidemic modeling or traffic estimation. Recently, cell phone traces have been successfully used to model aggregated human mobility, in particular during natural disasters such as earthquakes or flooding. Climate-related environmental change brings a decline of productive agricultural land and livestock which will push rural residents to migrate. As a result, it also has the potential of causing changes in human mobility and cause migrations that have a wider and long standing impact. In this study, using anonymized and aggregated cell phone traces, we model the migrations that happened during a severe drought that happened in La Guajira, Colombia, in 2014. Our results indicate a linear reduction of the population of 10 percent during the 6 months considered for this study. Furthermore, predicting these migrations has about a 60% success rate for both the total number of people that migrate and to where they migrate. We also introduce a modification of the Radiation model in order to capture weather as one of the factors driving mobility, showing a RSS and RMSE reduction of 4.5% when compared with the standard models.

## CCS CONCEPTS

**•Social and professional topics → Geographic characteristics; •Applied computing →** *Sociology;* **•Computing methodologies → Modeling methodologies;**

**ACM Reference format:**
Sibren Isaacman, Vanessa Frias-Martinez, and Enrique Frias-Martinez. 2018. Modeling human migration patterns during drought conditions in La Guajira, Colombia. In *Proceedings of ACM SIGCAS Conference on Computing and Sustainable Societies (COMPASS), Menlo Park and San Jose, CA, USA, June 20–22, 2018 (COMPASS '18),* 9 pages.
DOI: 10.1145/3209811.3209861

## 1 INTRODUCTION

Natural disasters, such as earthquakes, flooding, storms and snowstorms cause a sudden change in mobility as people try to find safe areas and/or follow government instructions. This sudden change in mobility has been captured and modeled already using cell phone traces [10, 23, 24] and other Location Based Social Networks (LBSN) mobility sources, such as geolocated twitter [26, 29].

As opposed to sudden environmental changes, climate change can be defined as the variation in weather patterns for a long period of time. In the mid- to long-term, these variations also have the potential of causing changes in human mobility and can be one of the main factors that cause migrations. Drought, coastal erosion and flooding are the main climate change reasons for causing these climate migrations. A climate or environmental migrant is an individual that is forced to leave their local environment due to long or sudden weather changes. The effects of global environmental change are creating new patterns of human migration and alter the existing ones [3, 13]. A recent study in Rwanda [4] has demonstrated that cell phone traces can be used to help find these new internal migration patterns.

A link between migration and climate change has already been documented in several studies that focused on developing economies [12, 13, 22]. In Burkina Faso [12], residents of dry areas tend to migrate to rural regions with greater rainfall. The study, using questionnaires on migration done to about 10k individuals, reports that only short-distance moves appear affected by climactic factors, since international migration tends to be less common in a period of rainfall shortage, perhaps because of the investment required for an international move. In the case of Ethiopia, the work by Meze-Hausken [22], after evaluating the response of 100 migrants, concluded that migration may be a long-term response to the threat of recurrent droughts. In any case the study reports that other survival strategies are used first in case of drought (such as selling farm equipment or borrowing food), and only when those are exhausted people migrate to new areas. In general, the literature has established that in rural areas, migration rises both immediately and as a long-term response to the threat of recurrent droughts[8].

In this paper we study and model the change in population caused by the severe drought that happened in La Guajira, Colombia, in 2014. La Guajira is a department of Colombia located in the northwest tip of the country, bordering Venezuela. It has an area of 20,848km$^2$ and an estimated population of 900,000 according to the projections of the last census available of 2005, belonging predominantly to the Watuu native people. According to the UN Office for the Coordination of Humanitarian Affairs (OCHA) [9], 2014 an extreme drought has affected La Guajira since the start of 2014. Rain is seasonal in the region, but repeated El Niño weather systems triggered a severe and long-lasting drought that, combined with corruption and poor management, further increased the effects on the population [5]. In some areas, the drought lasted for


*COMPASS '18, Menlo Park and San Jose, CA, USA*

978-1-4503-5816-3/18/06…$15.00
DOI: 10.1145/3209811.3209861

over three years. The drought caused a declaration of the state of public calamity in the municipality of Uribia (La Guajira) in February of 2014 [6]. Uribia has a total of 8,200km$^2$ and a population of 186,000 (as projected by the census of 2005). Less than 10% of the population lives in urban areas; the rest live in rural farms. An estimated 65,000 people have been affected by the severe droughts in La Guajira. Consequences to the population mainly regard malnutrition, especially for infants, with extreme consequences for the agricultural and livestock sectors.

For this study we use anonymized and aggregated cell phone records to study the mobility of the population during the droughts in order to answer two questions. First, if there is a change in the actual number of individuals living in the affected area, and, second, if and to what extent can we model the migration caused by the drought. To the best of our knowledge it is the first time in the literature that climate change migration has been studied and modeled using cell-phone records.

Section 2 presents the characteristics of the cell phone traces and the weather data used. Section 3 presents the home detection algorithm used to identify changes in population and the results obtained during the period considered. In Section 4 we review the two main approaches used to model migration and apply them to our context proposing a variation of the models to capture some indication of weather change. Section **??** presents the hypothesis of the mobility; i.e., that migrations are partially caused by the drought, and presents a modified migration model that captures changes in weather and rainfall to model mobility. Finally we finish with the Conclusion section.

# 2 DATA

## 2.1 Cell Phone Traces

Cell phone networks are built using a set of base transceiver stations (BTS) that are responsible for communicating cell phone devices within the network. Each BTS or cellular tower is identified by the latitude and longitude of its geographical location. The coverage area of each individual BTS is called a cell and is typically divided in three sectors each one covering 120°. Nevertheless it is possible for a BTS to have just one-directional sector or more than three sectors to handle areas with high density of population. The geographical area covered by a BTS depends mainly on the power of the individual antennas. Depending on population density, BTS coverage typically ranges from less than 1 km$^2$, in dense urban areas, to more than 4 km$^2$, in rural areas. For simplicity, it is common in the literature to assume that the cell of each BTS is a 2-dimensional non-overlapping polygon, which is typically approximated using Voronoi diagrams. This approach gives an acceptable approximation of the coverage area of each BTS. In practice, to build the firealfi diagram of coverage, one has to consider several factors, including the power and orientation of each antenna. In order to optimize signaling, BTS are grouped in Location Area Networks (LACs), which contain multiple BTSs. LACs determine the current location of a mobile phone without having to go down to the BTS level. Figure 1(left) presents a set of BTS with the original coverage for each cell, (center) the simulated coverage obtained using Voronoi diagrams and (right) the grouping of BTS into LACs.

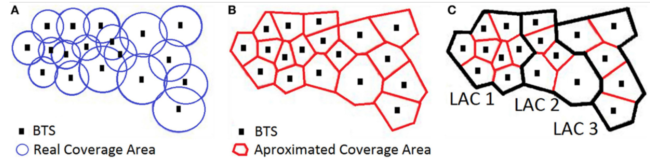



**Figure 1: (left) Original coverage areas of BTS, (center) approximation of coverage areas by Voronoi diagram and (right) geographical representation of LACs.**

Call Detail Records (CDRs) are generated whenever a cell phone connected to the network makes or receives a phone call or uses a service (e.g., SMS, MMS). In the process, the BTS details are logged, giving an indication of the geographical position of the user at the time of the call. Note that no information about the exact position of a user in a cell phone is known, i.e., we do not know the whereabouts of a phone within the coverage area of a BTS. Also, if the phone is not actively using (making or receiving) mobile network services, there will be no information in the CDRs.

In this study we use a 6-month (December 2013 through May 2014) aggregated and anonymized CDR dataset from Colombia. In order to preserve privacy, all the information presented is aggregated and original records are encrypted. From all the information contained in a CDR, our study only considers the encrypted originating number, the time and date of the interaction, and the BTS and LAC that the cell phone was connected to when the call was placed. No contract or demographic data was considered or available for this study and none of the authors of this paper collaborated in the extraction and the encryption of the original data.

## 2.2 Weather Data

Weather data covering total rainfall and average temperature per day was provided by IDEAM (Instituto de Hidrologia, Meteorologia y Estudios Ambientales), the Colombian National Weather Service. The information was provided for each of the weather stations based on the Colombian airport network. We have information from each of the 33 departments in which the country is divided (32 departments plus Bogotá Federal District). The time window covered depended on each weather station considered but in all cases covered the period under study with CDRs, a year before it and a year after it. The information was presented per month and in a text format.

# 3 MEASURING POPULATION CHANGE

In this section we measure the impact of the drought on the number of inhabitants of both Uribia and La Guajira. For that reason, we will apply a home detection algorithm to identify home on a weekly basis and measure the change in home location each week.

## 3.1 Home Detection

Home Detection algorithms are a specific kind of a wider group of algorithms used to identify meaningful places from mobility information. Although the bulk of the state of the art focuses on CDRs, the algorithms can be used to any set of non-continuous location traces. Usually, important places can be classified as home, work

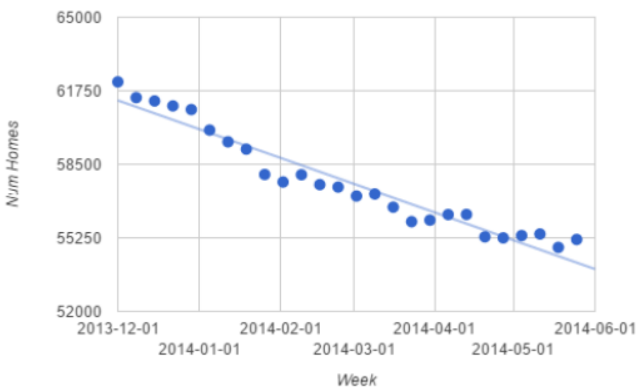



**Figure 2: Number of Home locations in La Guajira from Dec 13 to June 14**

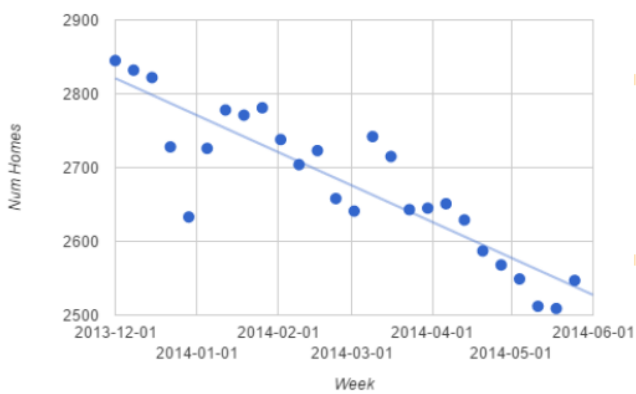



**Figure 3: Number of Home locations in Uribia from Dec 13 to June 14**

or other, where work defines any repetitive behavior taking place during typical work hours. The main idea behind all the different approaches consists of using some criteria to define time slots for home, work an other activities and from there processing the geo-information available. Ahas et al. [1] used an anchor-point model to identify home and work and validated it with the actual geography of the population finding a high level of correlation. Kang et al. [15] and Isaacman at al. [14] use a similar approach consisting on clustering towers (or active points) to identify home and work, but while Kang et al [15] does it for WiFi Signals, Isaacman [14] presented the study for CDR records. Frias-Martinez et al. [11] proposed a genetic algorithm approach to identify the time slot that had to be used to better characterize home and work. Last, Liao et al. [21] used hierarchically structured random fields to generate a model of a person's places. This approach had the advantage of taking into account the high level context in order to identify the relevance of each place.

In order to choose which approach to take, the main elements to consider are the amount of information available and the need for training and validating data. The amount of information available is usually expressed as the average number of calls for the specific time frame in which the home detection wants to be applied. The literature has reported how developing economies tend to have less interactions, both due to economic factors and to the general characteristics of the types of contract and pre-paid modes available. Rubio et al.[25] compared phone usage in a developed and developing economy showing that the advanced economy tend to use cell phones much more than in the developing economy. In fact, almost 75% of individuals in the advanced economy make/receive on average at least 2 calls per day, while only 27% of the population in the developing economy averages the same number of calls per day. This is our case in La Guajira, where the average number of calls per week is 5.8.

The distribution of towers also has an effect on the accuracy of home detection. In the case of La Guajira, the low density of population combined with the reduced number of urban areas, builds a map of towers where each urban area has a reduced set of towers (in general no more than 10). As our goal is to identify changes between municipalities, the architecture of the network helps to easily identify those changes. Figure 4 presents the BTS distribution in La Guajira. Uribia is in the north, at has only one BTS, while the big cluster of towers correspond to Riohacha (the Capital) and Distraccion.

As a result of the previous characteristics, we took the simplest approach, which consisted on identifying home as the most used BTS tower between the hours of 9pm and 8am Mon trough Thursday each week. If there was not enough information to identify a home tower during a week the last known position was assigned. Weeks prior to first available call are assumed to have the home location of the first call. We only considered a home change if there was a change in municipality. The home detection algorithm was applied for any cell phone that was present (i.e. has at least one phone call) in La Guajira during the 6 month period of study. Note that all calls independently of where they were made have been considered when applying the home detection algorithm, i.e., homes could be assigned anywhere in Colombia.

## 3.2 Home Changes

Figure 2 and Figure 3 presents the number of cell phones whose home location was assigned in the towers that cover the state of La Guajira and the municipality of Uribia from December 2013 to June 2014 during each week. Both results are similar, showing a linear decrease of 10% of the population during the period of study. Both cases can be model with a linear regression showing an $r^2$ of .78 and .93 for Uribia and La Guajira respectively. The last census available in Colombia is from 2005, and it provides an estimation of population given for both La Guajira and Uribia. In both cases the prediction was an increase in population and no reduction was presented or considered, in contrast with the actual numbers that the home detection algorithm has identified. In the case of La Guajira for the years 2013, 2014 and 2015 the estimated population was respectively 902,386, 930,165 and 957,814 and for Uribia 162,362, 168,286 and 174,281.

Regarding where those climate migrants go, roughly 90% percent of them stay in La Guajira relocating to other municipalities where access to help, food and water is probably easier. The other 10%

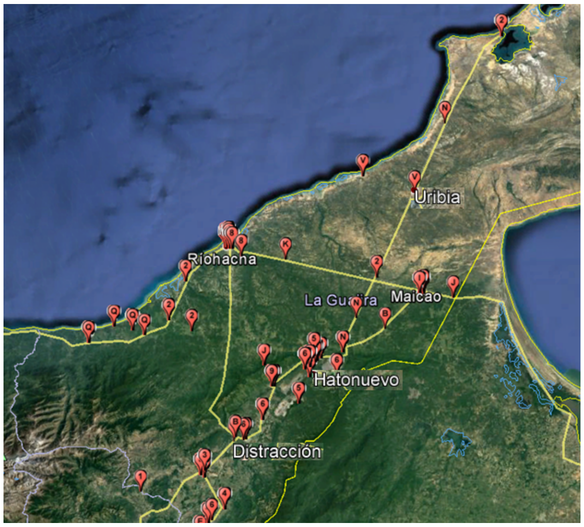



**Figure 4: BTS distribution in La Guajira,Uribia is in the north, at has only one BTS, while the big cluster of towers correspond to Riohacha (the Capital) and Distraccion.**

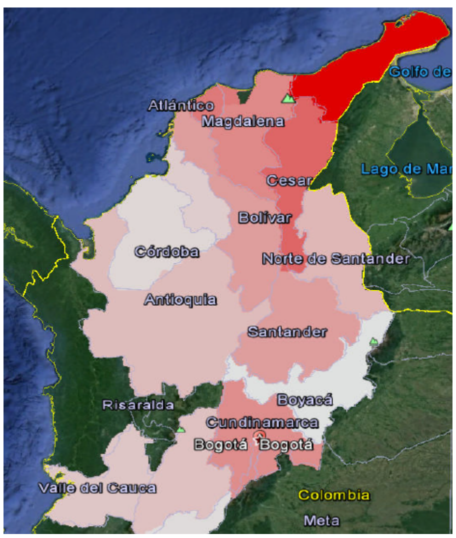



**Figure 5: Heatmap of the Relocation of individuals changing home during the second week of January 14 at a department level**

move to other departments. In general, the closer the department to La Guajira the higher the number of people relocating is, with the exception of Bogotá. This behavior suggest a gravity model in which migrants go to neighboring states or states with high population density (as Bogotá). Figure 5 presents in a red color scale where the migrants move from the second week of January 2014 at a department level. This result is in agreement with other studies that report that only short-distance moves appear affected by climactic factors [12].

## 4 MIGRATION MODELS

Although migration and commuting models were proposed as early as the forties [28, 31], recently, due to the explosion in geolocation information and the pervasiveness of cell phones, new ways of tackling the problem have been proposed. These new data sources gather human mobility datasets to a scale not previously available and have allowed for new model proposals both at an individual and at aggregated level of human mobility.

Currently, there are two main families most frequently used for modeling migration patterns: Gravity models [31] and Intervening opportunities models [28]; which have evolved into Radiation models[27]. Each approach is based on underlying physics formulas. While Gravity assumes that the number of trips between two locations decreases with distance, intervening opportunities assumes that the flow between two locations depends on the number of opportunities. With this approach the decision to migrate is not related to the distance of the two places, but to capability of satisfying the reason for the migration. As the concept of opportunity can be very complex to define, in its simplest form it us typically defined as the total number of locations between two points. Both approaches have been compared by the literature in a variety of studies and in general the results depend on each particular case and the characteristics of the data [19, 20, 27].

### 4.1 Gravity Models

In this family of models the population of cities is used as mass would be used in a gravity formula from physics. Due to its simplicity and ability to estimate flows, it is commonly used in transport [7], estimating volume of phone calls between urban areas[18], sociology [16] or epidemic spreading [2]. When using this approach, the probability,$p$, that a person commutes from location $i$ to location $j$ is proportional to the product of the populations ($m_i m_j$) and inversely proportional to the cost of travel between them.

$$p_{ij} \propto m_i m_j f(d_{ij}) \tag{1}$$

We may model the cost as exponential decay (which we will term the *GravExp* model)

$$f(d_{ij}) = e^{-\beta d_{ij}} \tag{2}$$

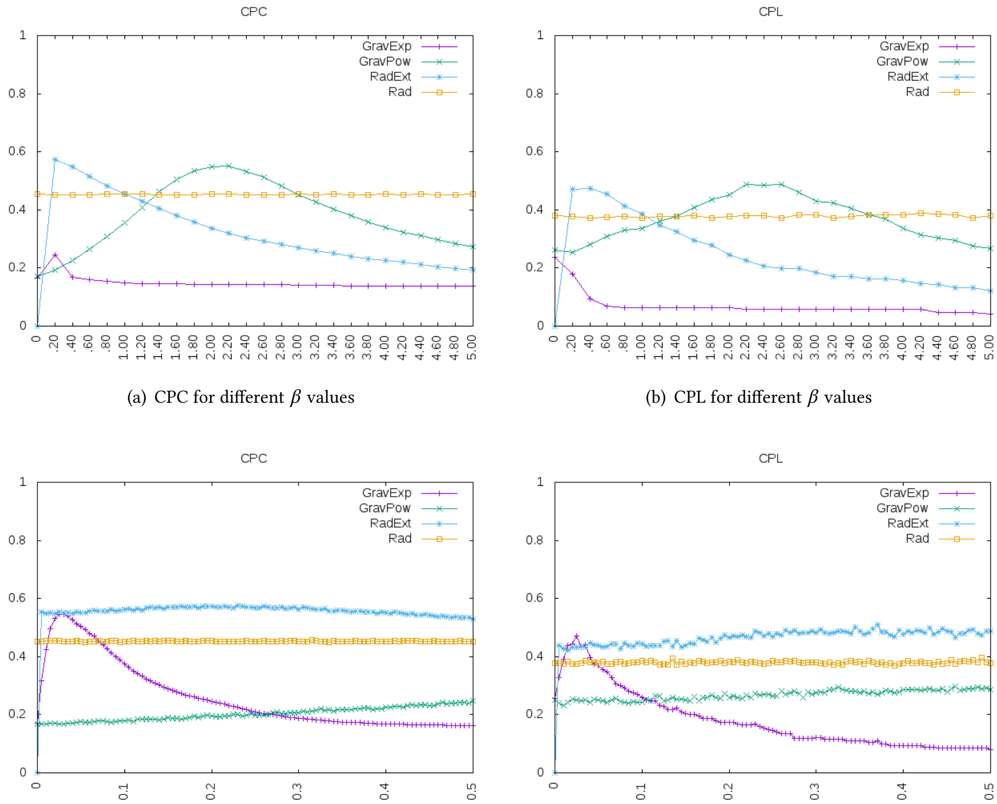
of as a power decay function (which we will term the *GravPow* model)

$$f(d_{ij}) = d_{ij}^{-\beta} \tag{3}$$

In general the main limitations of this Gravity models are (1) that it requires, at least, the estimation and calibration of $\beta$ which makes it sensitive to its changes; and (2) that for doing that the system needs empirical data of the actual movements which is not necessarily available for all cases. As a result of the previous limitations this approach is a strong simplification of the actual flows so the results may not reflect the actual mobility.

### 4.2 Radiation Models

In order to solve some of the limitations of the Gravity model, the radiation model [27] its an evolution of the intervening opportunities model[28] where enters are modeled as radiation and absorption bodies. In this case flows are modeled without parameters and the only input is the distribution of the population. The concept of




(a) CPC for different $\beta$ values

(b) CPL for different $\beta$ values

(c) CPC for $\beta$ values with a finer granularity

(d) CPL for $\beta$ values with a finer granularity

**Figure 6: CPC and CPL values for a given week for each migration model considered, Gravity Exponential(GravExp), Gravity Power Decay(GravPow), Radiation(Rad) and Radiation Extended(RadExt), with $\beta$ ranging from 0 to 5 in .2 increments (top row) and $\beta$ ranging from 0 to 0.5 in 0.05 increments (bottom row).**

"intervening opportunities" ($s_{ij}$) reflects which other population centers in a radius between the two points of interest $i$ and $j$ serve to reduce the probability that a migrant opts to migrate from $i$ to $j$. Radiation models have been reported to better captures long term migrations patterns and to have high degree of accuracy at the intra-county scale [27]. The basic radiation model (which we will call *Rad*) is formulated as:

$$p_{ij} = \frac{m_i m_j}{(m_j + s_{ij})(m_i + m_j + s_{ij})} \tag{4}$$

The radiation model as stated has no calibration parameters. Taking a hint from survival analysis, researchers have been able to extend the radiation model to be more general [30], and proposes the "Extended Radiation Model" *RadExt*. The formula, including the tunable parameter $\beta$ is

$$p_{ij} = \frac{((s_{ij} + m_i + m_j)^{\beta} - (s_{ij} + m_i)^{\beta}) * (m_i^{\beta} + 1)}{((s_{ij} + m_i + m_j)^{\beta} + 1) * ((s_{ij} + m_i)^{\beta} + 1)} \tag{5}$$

In all cases but *Rad*, we must explore the space to determine an appropriate value for $\beta$.

## 4.3 Goodness of Fit Metrics

In order to tune our parameters and determine the success of our chosen models, we need to determine a metric for goodness of fit. We choose two common goodness of fit metrics [20]: common part of commuters (*CPC*) and common part of links (*CPL*).

**CPC:** The percentage of good predictions for the number of people that migrate to other municipalities. It varies from 0, when no agreement is found, to 1, when the two networks are identical.

**CPL:** The percentage of the topological structure maintained for those that migrate; i.e. the ability of the models to recover the topological structure of the original network.

An extended description of these metrics can be found in [19, 20].

## 4.4 Efficacy of current mobility models

In order to evaluate the accuracy of the migration models, we used the 2014 projections of population from the most recent (2005) census in Colombia. The distance between states is taken as the distance between its centroids. The ground truth was defined by the location of the homes as identified with the CDR data. The accuracy is implemented by comparing people with a home in La

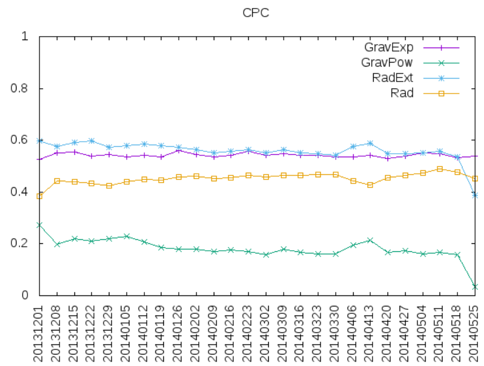



**Figure 7: CPC across each week of the dataset**

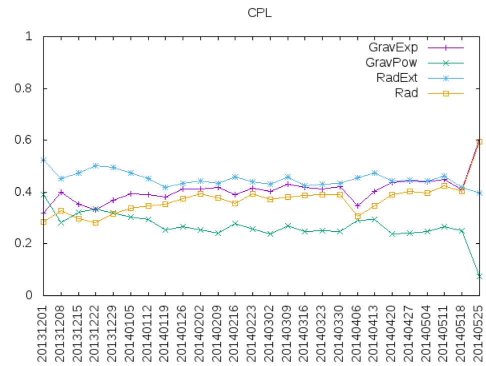



**Figure 8: CPL across each week of the dataset**

Guajira in week $w$ and in a different municipality in week $w+1$. We compare the number and location of people that have moved against those predicted to commute by the model.

The first step is to tune an appropiate value for $\beta$. In order to do that we scan $\beta$ over an arbitrary week in the 6 month window for both the CPL and CPC metrics. Figures 6(a) and 6(b) present the values of CPL and CPC for each one of the migration model considered: Gravity Exponential(GravExp), Gravity Power Decay(GravPow), Radiation(Rad) and Radiation Extended(RadExt) with $\beta$ values ranging between 0 and 5 in .2 increments. Both graphs show that CPL and CPC exhibit similar behavior over the range of $\beta$ values and each one of the migration models considered. In any case at this granularity, it is not clear which model outperforms the others. Figures 6(c) and 6(d) present the same models but for finer $\beta$ granularities, rangig from 0 to .5 with .01 increments. At this scale, it appears that RadExt with a $\beta$ of .05 consistently outperforms the other approaches for both the actual number of migrants (CPL) and the network of the migrations (CPC).

Having selected $\beta$, we can see the efficacy of prediction across all the weeks of our dataset. Figures 7 and 8 present the CPC and CPL values for each week considered for each of the models. In general, both metrics appear relatively stable across the whole period for each of the models. Again, RadExt achieves better accuracy both using CPC and CPL. The CPC goodness of fit metric achieves up to 60% accuracy while CPL achieves up to 52%. In general this values are in agreement with the literature, where RadExt tends to better capture the complexities of migrations [19, 27].

Also, these results seem to indicate that the state of the art in migration models is able to capture, to a large extent, the particularities of human flows caused by droughts. This fact also implies that, because RadExt basically assumes economic opportunities to model migration flows, migrations caused by droughts have economic motivation. Previous literature has already reported that the main motivation for migrations caused by droughts in rural areas (such as La Guajira) are economic: the drought affects the basic livelyhood of the locals [12, 13, 17]. In the following section we will introduce some modifications in the RadExt model to try to improve its prediction ability by incorporating rainfall.

## 5 WEATHER-BASED MIGRATION MODEL

Our working hypothesis is that the correlation between the drought and the migration of people is not coincidental, meaning that one of the main reason for people to relocate is to move to areas where there is no drought or at least where there are resources to access water and as a result economic prosperity is feassible. To that end, we propose and explore modifications to the mobility models with weather statistics related to the drought. Considering the nature of the migration, we focus on the rainfall information provided by the Colombian National Weather Service (IDEAM).

Radiation models are based in part in the concept of "intervening opportunities", meaning that the likelihood of migrating is influenced most by the opportunities to settle at the destination and less by distance or population. Even more, in this case the opportunities are used to modulate the effects of population and distance. We propose to design a mechanism to further modulate the effects of distance and population but based on the differences in rainfall between origin and each possible destination of the migration.

The new model builds a matrix for each week that reflects the difference in rainfall between any two departments ($r_{ij}$) during that week. Note that the original data reflects the rainfall per day and total rainfall for that week was obtained as the sum of all days. Considering that we have the information for all the airports, there is at least one weather station for each department, and in the cases where more than one airport is located in a department the average of the total rain during that week in each weather station is considered. In order to build a new intervening opportunities matrix, a scaling factor $\alpha$ is introduced:

$$s'_{ij} = \alpha * r_{ij} + s_{ij} \tag{6}$$

This new $s'_{ij}$ is used in the RadExt model from equation 5. Note that this new intervening opportunities matrix reflects the hypothesis that when migrating during drought conditions, a factor that should be included for modeling the destination of the migration is rainfall. In this context, when deciding where to move, within the same conditions, there would be a tendency towards migrating to rainier places.

Also, the decision of where to migrate is not necessarely based only on last week rainfall but on a constant difference in rainfall

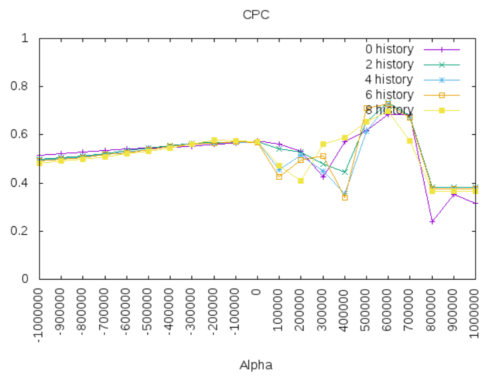



**Figure 9: CPC for difference Alpha values using an additive rainfall history. Coarse grain sweep.**

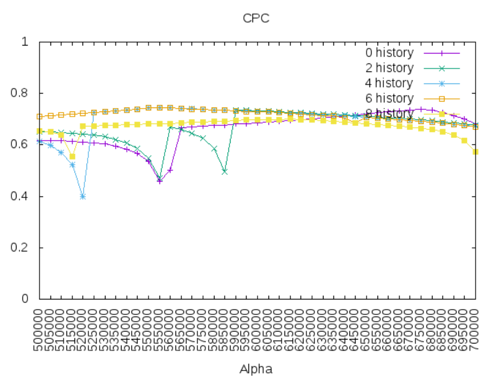



**Figure 10: CPC for difference Alpha values using an additive rainfall history.**

during some time before the actual migration, i.e. on a perception of a better environment. As a result we introduce a factor of additive rainfall history in $r_{ij}$ that reflects the perception of better weather conditions. In the base case, 0 history, we consider only the current week. As the history increases, $r_{ij}$ becomes:

$$r_{ij} = \sum_{w=0}^{history} r_{ij}(curr - w) \tag{7}$$

where $r_{ij}(x)$ is the rainfall difference in week x.

## 5.1 Evaluating Weather-based Radiation Model

In order to determine the optimum $\alpha$ value with various levels of rainfall history, Figure 9 shows, for a randomly selected week, the results of the sweep over a large range of alphas and considering 0, 2, 4, 6 and 8 weeks for rainfall history. This range was narrowed once an appropriate $\alpha$ range was determined. Figure 10 shows this reduced range and that an $\alpha$ value of 675000 and considering no history provides an optimum CPC value of .78. This result in itself is a marked improvement over the baseline model that did not include rainfall.

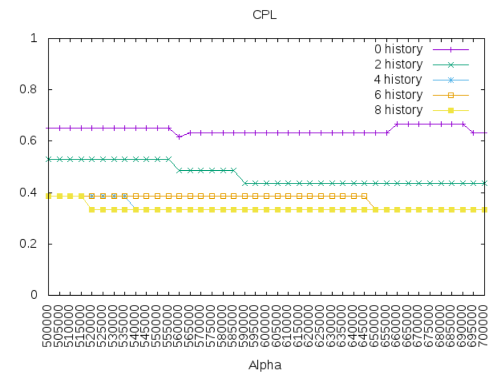



**Figure 11: CPL for difference Alpha values using an additive rainfall history.**

Figure 11 presents the CPL values when considering different $\alpha$ values. In this case the optimum value is achieved also with a history of 0 (i.e. no memory) and $\alpha$ value of 675000. For any other case, the CPL value decreases. Even in this scenario, the optimal $\alpha$ value produces a CPL of 0.68 which is still an improvement over the baseline RadExt Model.

These results indicate that an $\alpha$ value of 675000 and our previous $\beta$ value produce a significant improvement over the baseline RadExt. Also they show that the concepto of rainfall history does not seem to improve the prediction capabilities of the model. Nevertheless, as can be seen in Figure 12 in the RadExt with Rainfall plot, such improvement is not necessarely constant throughout the whole time window considered. CPC values fluctuate within a wide range of values, and although the addition of rainfall can produce significant improvements in some weeks, in others it does not seem to have the same impact. The main reason for that fluctuation is that $\alpha$ is not unitless and as such depends on the total rainfall for that week. As a result of that it is difficult to generalize $\alpha$ across weeks and find a consistently appropriate value.

## 5.2 Normalized Weather-based Radiation Model

The previous approach constructed an additive model that defined $s'_{ij}$ as the addition of all rainfall to create an "intervening opportunities" matrix. While there was an improvement in CPC and CPL over the baseline model in some weeks, that improvement was not necessarely constant from week to week due to the fact that $\alpha$ is not unitless. To minimize this effect we proposse a normalized Extended Radiation model. Formally:

$$R = max(r_{ij}) \tag{8}$$

$$S = \sum_i m_i \tag{9}$$

$$m'_{i,j} = \frac{m_{i,j}}{S} \tag{10}$$

$$s'_{ij} = \frac{s_{ij}}{S} + \alpha * \frac{r_{ij}}{R} \tag{11}$$

Figure 12 presents CPC values of the two rainfall-based RadExt models for each week considered. As mentioned earlier, before

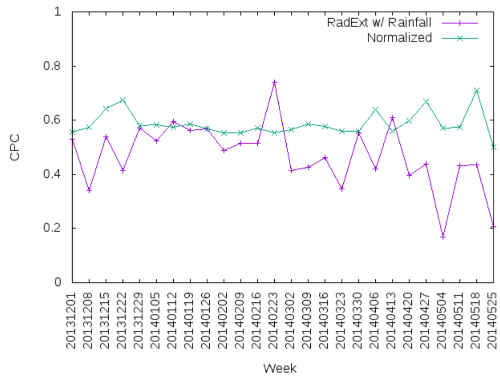



**Figure 12: Model performance over all weeks in the dataset using the RadExt w/ Rainfall and the Normalized Models.**

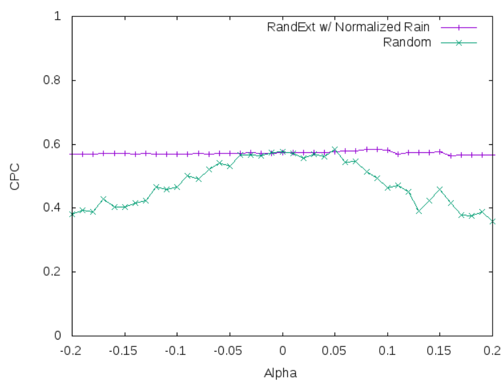



**Figure 13: Average performance of CPC over all weeks compared to scaled random noise.**

normalization, CPC values fluctuate within a wide range of values. However, modifying the model with a normalizing factor stabilizes the model and provides better average performance.

This normalization allows us to better explore the behavior across weeks and rule out the possibility that rainfall is just adding random noise to the model. Figure 13 shows the effect of adding random noise in place of rainfall data. The noise is generated uniformly in the range of 0 to 10% of the maximum rainfall for each week. The data is average CPC across all weeks. We note two things: (1) random noise is highly detrimental to the behavior of the model and (2) despite the irregularity in improvement from week to week, adding rainfall to the model with a (normalized) alpha factor of 0.09 improves the Extended radiation model across all weeks.

Considering the previous results we can now evaluate the impact that the inclusion of rainfall has on the normalized RadExt model when compared with the standard RadExt model presented in the "Efficacy of current mobility models" subsection. Such comparisson is relevant as indirectly it is a meassuare of the relevance of rainfall for migrating. Figure 14 compares the effect of including normalized rainfall in the Radiation model with the standard radiation model and considers the information provided by the CDR as the real mobility data. The graphs present the PDF of (a) the migration distance and (b) the population that migrated. Although both models present a good approach to the actual mobility the inclusion of normalized rainfall results in a reduction of RSS (Residual Sum of Squares) of 4.5% . This improvement in RSS partially validates the assumption that the migrations observed in La Guajira during the droughts are to a certain extent caused by the lack of rainfall.

Observing Figure 14, both in (a) and (b), there is a big peak in the real mobility data (at $X$ value of 150km in Figure 14(a)) that is not captured by any of the models. This is a result of an artifact caused by how distance is calculated. For the models, the distance between two states is given by the centroid of the states. In the particular case of that peak, and focusing on Figure 14(a), it corresponds to a neighboring state of La Guajira, Cesar. The north of Cesar has a city where people from southern La Guajira have migrated (Valledupar, the capital of Cesar). This migration is roughly 150 km if obtained from CDRs. However, the migration model is based on the distance between states which is obtained from their centroids. In general, for any two states that are not close to each other, this is not a problem, however this fact can create artifacts when the actual migration is simply crossing a border. As a result the models assign a longer range to that migration than it actually is.

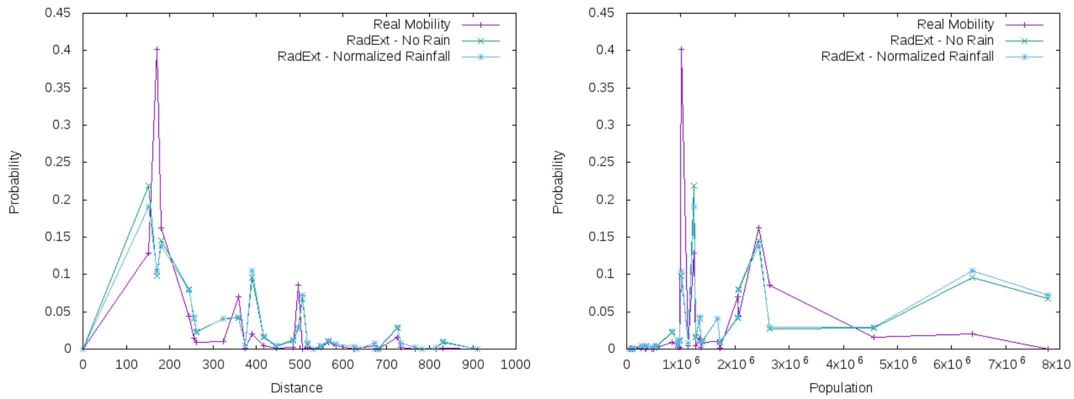
## 6 CONCLUSIONS AND FUTURE WORK

Climate change-based migrations have been presented as one of the main reasons for human mobility, especially in dry land areas of less developed countries [13]. In these areas, environmental change has immediate and direct effects on the health and well-being of households that depend on natural resources for their basic livelihoods [17] which motivates the migration.

This study has evaluated using cell phone traces to study the impact of and model mobility during extreme drought conditions in La Guajira, Colombia. Our results support the fact that CDRs are a powerful source of information for modeling climate-change migrations. Typically natural disasters produce a sudden change in mobility but for climate change migrations, and drought in particular, the change is slower over time, and the use of points of interest algorithms (and home detection algorithms in particular) opens the door to use CDRs for modeling these types of human flows.

From a population perspective we have shown that there is a 10% reduction in population both at a municipal and state level. Also, the majority of people that migrate tend to do so within their own state, in agreement with other studies that report that only short-distance moves appear affected by climactic factors [12]. Using state of the art migration models we are able to model the migration with CPC and CPL values in agreement with some of the studies presented in the literature that use the Radiation model [19, 20, 27]. This indicates that climate change migration to a large extent can be explained by economic factors captured by "intervening opportunities." In the end, climate-change in rural areas affects rural way of living which pushes people for looking new jobs.

In order to try to capture the specificities of a weather-based migrations we have proposed a Radiation model with normalized rainfall integrated in the intervening opportunities matrix. Our results indicate an improvement of 4.5% in RSS when considering rainfall, which can also be an interpreted as the importance given to the weather when migrating to a new place.




(a) PDF for migration by Distance

(b) PDF for migration by Population

**Figure 14: PDF of (a) the distance and (b) the population for the Radiation model and the Radiation Model+Normalized Rainfall. The information provided by the CDR is used as the standard for comparison of the results.**

We believe that there is still promising work to be done in a number of different directions. First, both approaches to the addition of rainfall have been focused on summation. It is reasonable to question this assumption and instead propose multiplicative modifications. Second, in addition to rainfall, we have collected temperature data from weather stations throughout the country. It seems likely that rainfall and temperature may act in tandem.

## ACKNOWLEDGMENTS

This work is supported in part by the National Science Foundation under Grant Nos. NSF-CNS 1750102 and NSF-EFMA 1441177.

## REFERENCES

[1] R. Ahas, S. Silm, O. Järv, E. Saluveer, and M. Tiru. 2010. Using mobile positioning data to model locations meaningful to users of mobile phones. *Journal of urban technology* 17, 1 (2010), 3–27.

[2] D. Balcan, V. Colizza, B. Gonçalves, H. Hu, J.J. Ramasco, and A. Vespignani. 2009. Multiscale mobility networks and the spatial spreading of infectious diseases. *Proc. Nat. Academy of Sciences* 106, 51 (2009), 21484–21489.

[3] R. Black, S. Bennett, S. Thomas, and J.R. Beddington. 2011. Climate change: Migration as adaptation. *Nature* 478, 7370 (2011), 447.

[4] Joshua E Blumenstock. 2012. Inferring patterns of internal migration from mobile phone call records: evidence from Rwanda. *Information Technology for Development* 18, 2 (2012), 107–125.

[5] Natalio Cosoy. 2015. Colombia's Alta Guajira region struggles with drought. *BBC News* (2015).

[6] Ayuntamiento de Uribia. 2014. Declaration of Public Calamity in Uribia (in Spanish). In *Decreto Calamidad Publica de Uribia*.

[7] S. Erlander and N.F. Stewart. 1990. *The gravity model in transportation analysis: theory and extensions*. Vol. 3. Vsp.

[8] Sally E Findley. 1994. Does drought increase migration? A study of migration from rural Mali during the 1983-1985 drought. *International Migration Review* (1994), 539–553.

[9] UN Office for the Coordination of Humanitation Affairs. 2014. Colombia Disaster Snapshot 2014.

[10] C. Frias-Martinez, V.and Soguero and E. Frias-Martinez. 2012. Estimation of urban commuting patterns using cellphone network data. In *ACM SIGKDD Intl. Workshop on Urban Computing (pp. 9-16)*.

[11] V. Frias-Martinez, J. Virseda, A. Rubio, and E. Frias-Martinez. 2010. Towards large scale technology impact analyses: Automatic residential localization from mobile phone-call data. In *4th ACM/IEEE Intl. Conf. Inf. and Communication technologies and development*. ACM, 11.

[12] S. Henry, B. Schoumaker, and C. Beauchemin. 2004. The impact of rainfall on the first out-migration: A multi-level event-history analysis in Burkina Faso. *Population and Environment* 25, 5 (2004), 423–460.

[13] Lori M Hunter. 2007. Climate change, rural vulnerabilities, and migration. *Population Reference Bureau* (2007).

[14] S. Isaacman, R. Becker, R. Cáceres, S. Kobourov, J. Martonosi, M.and Rowland, and A. Varshavsky. 2011. Identifying important places in peoplefis lives from cellular network data. In *Int. Conf. on Pervasive Computing*. Springer, 133–151.

[15] J.H. Kang, W. Welbourne, B. Stewart, and G. Borriello. 2005. Extracting places from traces of locations. *ACM SIGMOBILE Mobile Computing and Communications Review* 9, 3 (2005), 58–68.

[16] D. Karemera, V. Oguledo, and B. Davis. 2000. A gravity model analysis of international migration to North America. *Applied Economics* 32, 13 (2000), 1745–1755.

[17] I. Koziell and J.(Eds.) Saunders. 2001. *Living off biodiversity: Exploring livelihoods and biodiversity issues in natural resources management*. Iied.

[18] G. Krings, F. Calabrese, C. Ratti, and V.D. Blondel. 2009. Urban gravity: a model for inter-city telecommunication flows. *J.of Statistical Mechanics: Theory and Experiment* 2009, 07 (2009), L07003.

[19] M. Lenormand, A. Bassolas, and J.J. Ramasco. 2016. Systematic comparison of trip distribution laws and models. *J. Transport Geography* 51 (2016), 158–169.

[20] M. Lenormand, S. Huet, F. Gargiulo, and G. Deffuant. 2012. A universal model of commuting networks. *PloS one* 7, 10 (2012), e45985.

[21] L. Liao, D. Fox, and H. Kautz. 2007. Extracting places and activities from gps traces. *The International Journal of Robotics Research* 26, 1 (2007), 119–134.

[22] E. Meze-Hausken. 2000. Migration caused by climate change: how vulnerable are people inn dryland areas? *Mitigation and Adaptation Strategies for Global Change* 5, 4 (2000), 379–406.

[23] B. Moumni, V. Frias-Martinez, and E. Frias-Martinez. 2013. Characterizing social response to urban earthquakes using cell-phone network data: the 2012 oaxaca earthquake. In *ACM Conf. Pervasive and Ubiquitous Computing*.

[24] D. Pastor-Escuredo, A. Morales-Guzman, E. Torres-Fernndez, A. Frias-Martinez, and V. Frias-Martinez. 2014. Flooding through the lens of mobile phone activity. In *IEEE Humanitarian Technology Conference (GHTC)*.

[25] A. Rubio, V. Frias-Martinez, E. Frias-Martinez, and N. Oliver. 2010. Human Mobility in Advanced and Developing Economies: A Comparative Analysis.. In *AAAI Spring Symposium: Artificial Intelligence for Development*.

[26] T. Sakaki, M. Okazaki, and Y. Matsuo. 2010. Earthquake shakes Twitter users: real-time event detection by social sensors. In *Proc. 19th Int. Conf. WWW*. ACM, 851–860.

[27] F. Simini, M.C. González, A. Maritan, and A.lbert-László Barabási. 2012. A universal model for mobility and migration patterns. *Nature* 484, 7392 (2012), 96.

[28] S.A. Stouffer. 1940. Intervening opportunities: a theory relating mobility and distance. *American sociological review* 5, 6 (1940), 845–867.

[29] Q. Wang and J.E. Taylor. 2014. Quantifying human mobility perturbation and resilience in Hurricane Sandy. *PLoS one* 9, 11 (2014), e112608.

[30] Y. Yang, C. Herrera, N. Eagle, and M.C. González. 2014. Limits of predictability in commuting flows in the absence of data for calibration. *Scientific reports* 4 (2014), 5662.

[31] G.K. Zipf. 1946. On the intercity movement of persons. *American sociological review* 11, 6 (1946), 677–686.