



# MIT Open Access Articles

## *Estimating Origin-Destination flows using opportunistically collected mobile phone location data from one million users in Boston Metropolitan Area*

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

<b>Citation</b>	Calabrese, Francesco, Giusy Di Lorenzo, Liang Liu, and Carlo Ratti. "Estimating Origin-Destination Flows Using Mobile Phone Location Data." IEEE Pervasive Computing 10, no. 4 (April 2011): 36-44.
<b>As Published</b>	<a href="http://dx.doi.org/10.1109/mprv.2011.41">http://dx.doi.org/10.1109/mprv.2011.41</a>
<b>Publisher</b>	Institute of Electrical and Electronics Engineers (IEEE)
<b>Version</b>	Original manuscript
<b>Citable link</b>	<a href="http://hdl.handle.net/1721.1/101623">http://hdl.handle.net/1721.1/101623</a>
<b>Terms of Use</b>	Creative Commons Attribution-Noncommercial-Share Alike
<b>Detailed Terms</b>	<a href="http://creativecommons.org/licenses/by-nc-sa/4.0/">http://creativecommons.org/licenses/by-nc-sa/4.0/</a>



# *senseable* city lab:...

# Estimating Origin-Destination flows using opportunistically collected mobile phone location data from one million users in Boston Metropolitan Area

Francesco Calabrese, *Member, IEEE*, Giusy Di Lorenzo, Liang Liu, Carlo Ratti

## Abstract

In this paper, we present an algorithm for the analysis of opportunistically collected mobile phone location data to estimate a population's travel demand in terms of origins and destinations of individual trips. Aggregating the trips from millions individual mobile phone users in the Boston Metropolitan area, we show that the estimated Origin-Destination flows correlate well with the US Census estimates at both the county and census tract levels. Moreover, compared to traditional census survey data, our estimations allow capturing weekday and weekend patterns as well as seasonal variations. These features could make methods for Origin-Destination flow estimation based on opportunistically collected mobile phone location data a critical component for transportation management and emergency response.

## I. INTRODUCTION

Origin Destination (OD) matrices represent one of the most important sources of information used for strategic planning and management of transportation networks. A precise calculation of OD matrices is an essential component for enabling administrative authorities to optimize the use of their transportation networks, not only for the benefit of users on their daily journeys but also with a view to the investments required to adapt these infrastructures to envisaged future needs. Traditionally, urban planning and transportation engineering rely on household questionnaires or census and road surveys conducted every 5-10 years and develop methodologies for OD matrices estimation. This approach has two main drawbacks:

- the process involved in the calculation of an OD matrix, from the initial data-gathering to the exploitation of the first results, is lengthy and may take years to only get a snapshot of the travel demand;
- the collected data has shortcomings both in terms of spatial and temporal scale.

Sensor-based OD estimation methods have also been developed in the past few years, making use of street sensors such as loop detectors and video cameras together with traffic assignment models. Analogous methods have been developed using probe vehicles, where vehicles traces are used as data sources [1], [2]. Those methods are, however, limited by the fact that models are often underdetermined because the number of parameters to be estimated is typically larger than the number of monitored network links [3].

On the other hand, the wide deployment of pervasive computing devices (e.g. mobile phone, smart cards, GPS devices and digital cameras) provide unprecedented digital footprints, telling where people are and when they are there. In former projects, different methodologies for detecting the presence and movement of crowds through their digital footprint (flickr photo, mobile phone logs, smart card record and taxi/bus GPS traces) were developed, see for instance [4]–[6]. This fine grained analysis can potentially make a big leap in terms of understanding the use of space and daily commuting flows for the purposes of urban mobility planning and management. Thus, it is no surprise that the idea of using mobile phones to monitor traffic conditions is not new. A fair number of studies relating to this matter have been published in recent

years. Bolla et al. [7] presented a model for estimating traffic by means of an algorithm that calculates traffic parameters on the basis of mobile phone location data. A case study was developed in Rome for real time urban monitoring using aggregated mobile phone data to monitor traffic and movement of vehicles and pedestrians [8]. Cayford et al. analyzed the main parameters to be taken into account, namely precision, metering frequency and the number of localizations necessary to achieve accurate traffic descriptions [9]. Several companies worldwide, including ITIS Holdings (Britain), Delcan (Canada), CellInt (Israel), as well as AirSage and IntelliOne (USA), have begun developing commercial applications of mobile phone based traffic monitoring.

With the specific goal of measuring origin-destination flows, different mobile phone signaling datasets have been considered and simulated to evaluate the feasibility of estimating trips. Initial work was done by [10] using billing data, consisting of cell phone tower information every time a phone received or made a call. In [11] the authors used mobile phone positions every two hours to infer trips. In [12] the authors studied the use of location updates to infer mobile phone movement. In [13] the authors used cell phone tower handover information acquired every time, during a call, a phone switches a tower it is connected to. In the latest effort, [14] estimated the daily OD demand using simulated cellular probe trajectory information (extracted from location updates, handover, and transition of Timing Advance values) and tested the methodology via the VISSIM simulation.

Though these results show great potential for using cellular probe trajectory information as a means to estimating travel demand, all methods have several shortcomings before they can be put into practice. Indeed, as mentioned in [14], field tests are needed for the following reasons:

- real coverage areas of cellphone towers are very different from the simulated ones, and vary from urban to rural areas;
- validations of methods to determine origin and destination of trips should be performed using real individual mobility data;
- real mobility and calling patterns should be included in the analysis, as they crucially influence the performance of the methods;
- existing OD matrices should be used as ground truth to verify the correctness of the estimated results.

In this paper, we design a methodology that makes use of opportunistically collected mobile phone location data to estimate dynamic OD matrices. We address all above concerns using a real mobility and calling dataset from 1 million mobile phone users. We use the Boston Metropolitan area as a case study and validate our methodology using census survey data for both county and census tract levels [15]. Both the methodology developed and the data precision and amount are thus far novel and unique to our knowledge.

The paper is structured as follows. Section II describes the mobile phone dataset considered. Section III describes the OD estimation method. Section IV shows the application of the method to a real case study in the Boston Metropolitan area, and comparison of the estimated OD matrices with Census commuting flows. Section V shows some new potentials for dynamically updated OD matrices. Finally, discussion and conclusion are given.

## II. MOBILE PHONE DATASET

The considered dataset consists of anonymous location measurements generated each time a device connects to the cellular network, including:

- when a call is placed or received (both at the beginning and end of a call);
- when a short message is sent or received;
- when the user connects to the internet (e.g. to browse the web, or through email programs that periodically check the mail server).

In the remainder of the paper we will call these events *network connections*. These events represent a superset of the ones contained in the Call Details Records, previously considered in [10], [16]. In this research we have been able to analyze 829 million mobile location data for 1 million device collected

by AirSage<sup>1</sup>. Not only the id of the cell tower the mobile phone is connected to was available, but also an estimation of its position within the cell is generated through triangulation by means of AirSage's Wireless Signal Extraction technology. Each location measurement  $m_i \in M$  is characterized by a position  $p_{m_i}$  expressed in latitude and longitude and a timestamp  $t_{m_i}$ .

In order to infer trips from these measurements, we first characterized the individual calling activity and verified whether that is frequent enough to allow monitoring the user's movement over time with a fine enough resolution. For each user we measured the interevent time i.e. the time interval between two consecutive network connections (similar to what was measured in [16]). The average interevent time measured for all the whole population was 260 minutes, much lower than the one found in [16] (500 minutes) as we are also considering mobile internet connections. Since the distribution of interevent times for an user spans over several temporal scales, we further characterized each calling activity distribution by its first and third quantile and the median. Fig. 1 shows the distribution of the first and third quantile and the median for all users available into the dataset. The arithmetic average of the medians is 84 minutes (the geometric average of the medians is 10.3 minutes) with results small enough to detect changes of location where the user stops as low as 1.5 hours.

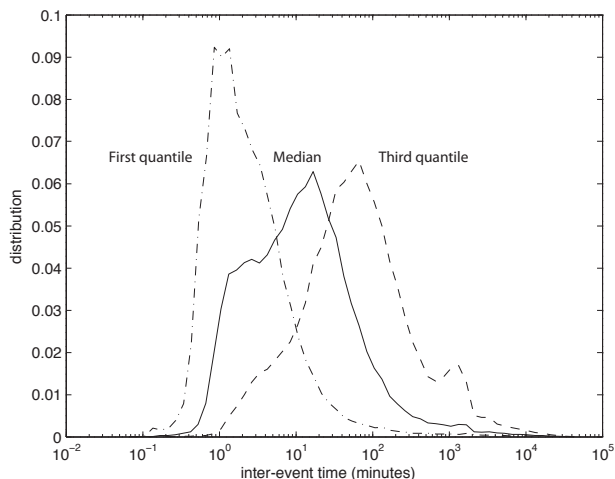


Fig. 1. Characterization of individual calling activity for the whole population. Median (solid line), first quantile (dash-dotted line) and third quantile (dashed line) of individual interevent time.

Mobile phone-derived location data has lower resolution than GPS data: internal and independent testing suggests an average uncertainty radius of 320 meters, and a median of 220 meters. Moreover, at some peak usage periods additional locational error may be introduced when users are automatically transferred by the network from the closest cellular tower to one which is further away but less heavily-loaded.

### III. ORIGIN-DESTINATION ESTIMATION METHOD

The procedure for estimating dynamic OD matrices is composed of two steps: trips determination and origin-destination estimation.

To alleviate the effects of localization errors and event-driven location measurements on the determination of individual trips, we propose the following method: we apply a low-pass filter with a resampling rate of 10 minutes to the raw data, this follows an approach tested with data from Rome, Italy [8]. In addition, since lesser localization errors might still generate fictitious trips, we adapt a pre-processing step employed in the analysis of gps traces, which uses clustering to identify minor oscillations around a common location. In more detail, the approach employed to handle locational errors and identify meaningful locations in a user's travel history can be understood as follows:

<sup>1</sup><http://www.airsage.com/>

- We begin with a measurement series  $M_s = \{m_q, m_{q+1}, \dots, m_z\} \in M^{z-q-1}$ ,  $q > z$ , derived from a series of network connections over a certain time interval  $\Delta T = t_{m_z} - t_{m_q} > 0$ .
- We define an area with radius  $\Delta S$  – in this case,  $1km$  to take into account the localization errors estimated by AirSage – such that

$$\max distance(p_{m_i}, p_{m_j}) < \Delta S \quad \forall \quad q \leq i, j \leq z$$

- All the consecutive points  $p_j \in M_s$  for which this condition holds can be fused together such that the centroid becomes a ‘virtual location’ ( $p_s = (z - q)^{-1} \sum_{i=q}^{i=z} p_{m_i}$ , the centroid of the points) that is the origin or destination of a trip.
- Once the virtual locations are detected, we can evaluate the stops (virtual locations) and trips as paths between users’ positions at consecutive virtual locations. Each trip  $trip(u, o, d, t)$  is characterized by user id  $u$ , origin location  $o$ , destination location  $d$  and starting time  $t$ .

Section IV presents some statistics on the trips estimated using the proposed method comparing it with reference statistics, showing how the method performs well in estimating trips in our case study.

Once trips are extracted, the procedure to derive Origin-Destination flows is the following:

- 1) The geographical area under analysis is divided into regions:  $region_i$ ,  $i = 1, \dots, n$ .
- 2) Origin and destination regions, together with starting time are extracted for each trip of each user  $trip(u, o, d, t)$ .
- 3) Trips with the same origin and destination regions are grouped together at different temporal windows  $tw$  e.g. weekly, daily, hourly:

$$m(i, j, tw) = \sum_{o \in region_i, d \in region_j, t \in tw} trip(u, o, d, t).$$

The result is a three-dimensional matrix  $M \in \mathfrak{R}^3$  whose element  $m(i, j, tw)$  represents the number of trips from origin region  $i$  to destination region  $j$  starting within the time window  $tw$ . The potentials of using adaptive time windows will be shown in Section V-A.

#### IV. CASE STUDY IN THE BOSTON REGION AND COMPARISON WITH CENSUS COMMUTING FLOWS

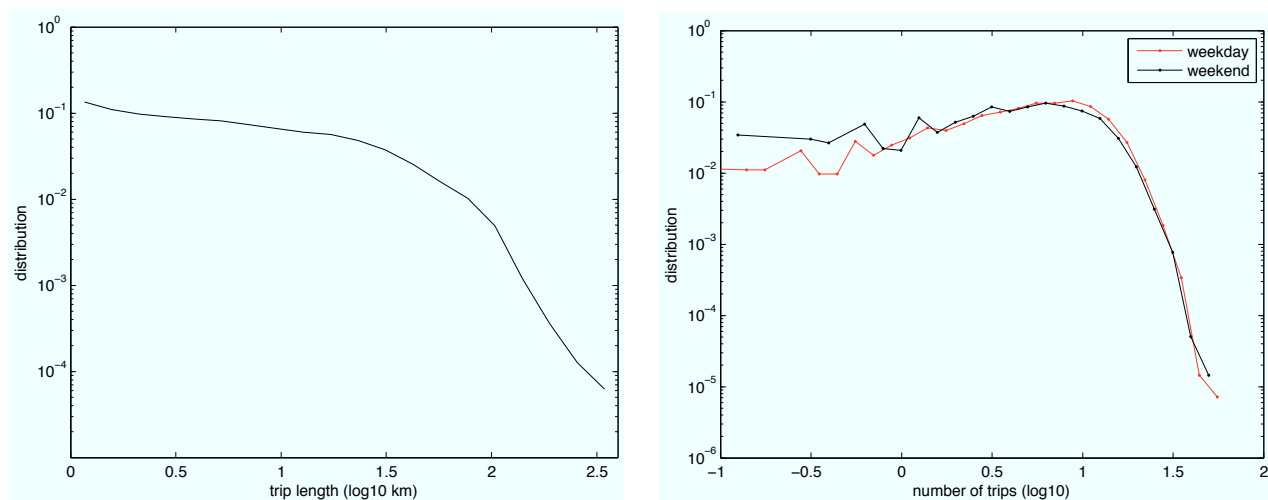
In this section we study the effectiveness of the methodology in a real case study in the Boston region. Based on the area covered by the mobile phone locations dataset, we analyzed the movements among areas in 8 counties in east Massachusetts (Middlesex, Suffolk, Essex, Worcester, Norfolk, Bristol, Plymouth, Barnstable) with an approximate population of 5.5 million people. To simplify the analysis, we extracted traces for 25% randomly selected users among the available ones.

##### A. Characterization of trips

As a first analysis we studied the trip length distribution (see Figure 2(a)), showing that trips range from 1 to 300 Km. We determined the trip length  $x$  by calculating the Euclidean distances among trip’s origin and destination. The distribution is well approximated by  $P(x) = (x + 14.6)^{-0.78} \exp(-x/60)$  with  $R^2 = 0.98$ , which confirms what was found in [16]. The slightly different coefficients found in this case could be attributed to the different built environment in Europe and US, see [17]. To check the plausibility of our segmentation of the trajectory in trips, we compute same statistics computed on the number of individual trips per day. The distribution over the whole population is shown in Figure 2(b), separating weekday and weekend trips. We obtain an average of 5 trips per day during the weekday, and 4.5 during the weekend. This number is reasonable when compared to the US National Household Travel Survey<sup>2</sup> which evaluated this number to be between 4.18 during weekdays and 3.86 during weekends<sup>3</sup>.

<sup>2</sup><http://nhts.ornl.gov/>

<sup>3</sup>The sources of differences can be associated to several reasons, including the several years of difference between when the two datasets have been collected, and the fact that NHTS is based on a sample over all US population, so not focused on the behavior of people in the Boston Metropolitan area.



(a) Trip length distribution. Curve interpolated with  $P(x) = (x + 14.6)^{-0.78} \exp(-x/60)$  with  $R^2 = 0.98$ .

(b) Trips per day distribution

Fig. 2. Statistics on the detected trips.

To evaluate whether we have sampling biases in our data, we computed the home locations distribution estimated from the mobile phone data, and compared it with data from the US 2000 Census. To detect the home location, we first group together geographic regions that are close in space, creating a grid in space where the side of every cell is 500 meters. For each cell we evaluate the number of nights the user connects to the network in the night time interval while in that cell, and select as a home location the cell with the greatest value<sup>4</sup>.

To validate the home location distribution, we then compared it with population data from the US 2000 Census, at the level of the census tract [18]. In the selected 8 counties, we have 1171 distinct census tracts, with populations ranging from 70 to 12 thousand people (on average 4705), and an area ranging from 0.08 to 203  $km^2$  (on average 10.8  $km^2$ ). The census tract population estimated using mobile phone users' home locations scales linearly with the Census population, as shown in Figure 3(b), corresponding to an average 4.3% of the population being monitored.

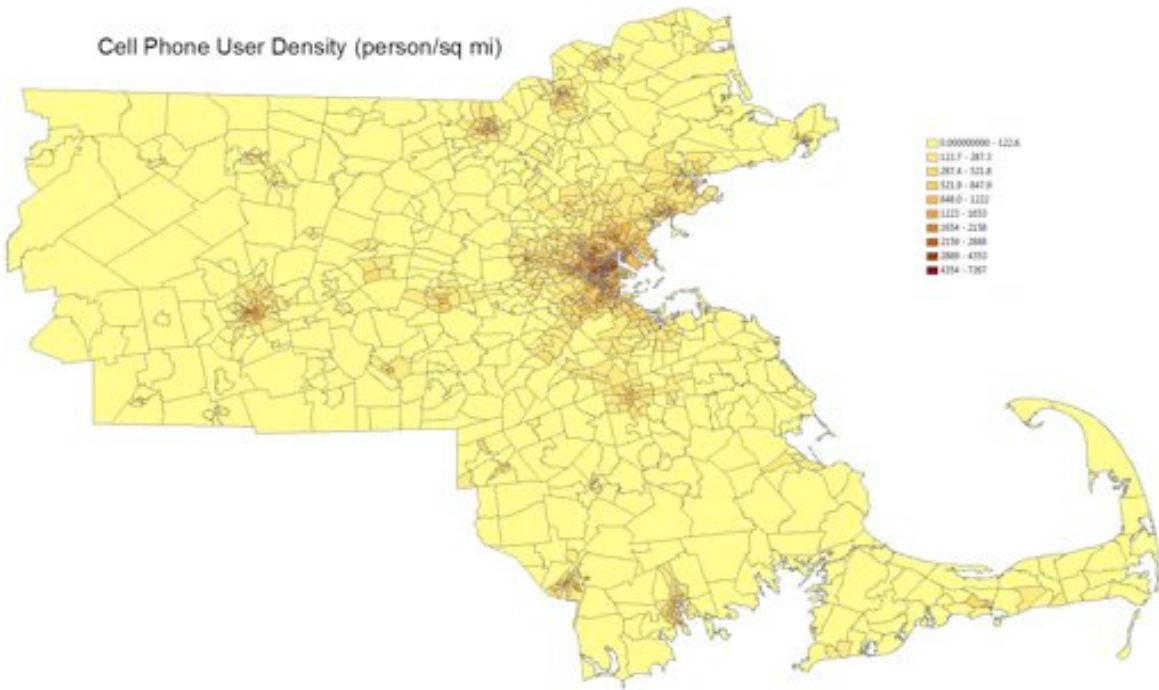
### B. Characterization of OD flows

To validate the accuracy of the OD matrices produced using the mobile phone traces, we used the most recent Tract-Tract Worker Flows dataset from Census Transportation Planning Package [15]. CTPP is a special tabulation of responses from households completing the Census long form. It is the only Census product that summarizes data by place of work and tabulates the flow of workers between home and work.

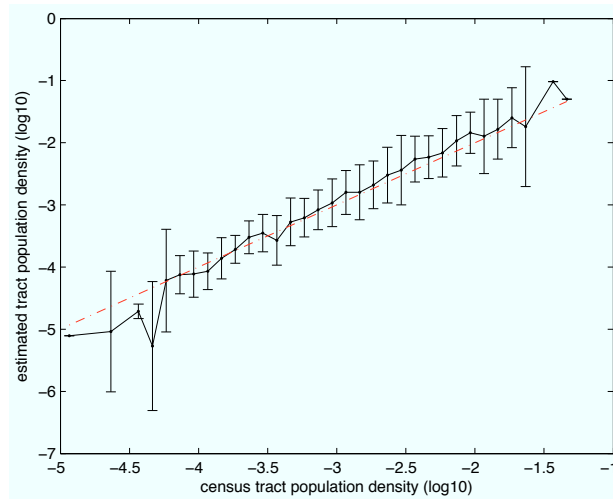
The *Tract-Tract Worker Flows* data shows the number of workers in each tract of work by tract of residence. Workers are defined as people age 16 years old and over who were employed and at work, full time or part time, during the Census reference week (generally the last week of March). The data contains the number of workers in the flow who were allocated to tract, place, and county of work.

Given the two levels of granularity (tract and county) available in the CTPP dataset, we computed our OD estimates at two levels of aggregation. Since commuting flow generally accounts for two trips (home to work and work to home), we considered undirected flows between two locations to compare our OD

<sup>4</sup>The considered night time interval is 6pm-8am and has been defined considering the statistics available in the American Time Use Survey, <http://www.bls.gov/tus/charts/work.htm>



(a) Census tract population density derived from Cellphone users' estimated home locations



(b) Population density comparison. Cellphone-based population density has been multiplied by  $100/4.3$  to take into account the percentage of population being monitored. Error bars represent the standard error.

Fig. 3. Census tract population density derived from US 2000 Census compared to Cellphone users' estimated home locations density

estimations. For each granularity, we computed the average daily number of trips:

$$m_{All}(i, j) = \frac{K_{All}}{\#days} \sum_{tw=day} (m(i, j, tw) + m(j, i, tw)),$$

$$i = 1, \dots, n, \quad j = 1, \dots, i - 1,$$

where  $K_{All}$  is a scaling factor we use to compare them with the Census estimations.

Moreover, since according to the definition, the census dataset includes only commuting trips, we evaluated the average daily number of trips made only on weekdays mornings (6-10am) from the estimated



home to estimated work location<sup>5</sup>:

$$m_{WM}(i, j) = \frac{K_{WM}}{\#weekdays} \sum_{tw=wm} (m(i, j, tw) + m(j, i, tw)),$$

$$i = 1, \dots, n, \quad j = 1, \dots, i - 1,$$

where  $K_{WM}$  is a scaling factor. Finally, we also considered the well known and widely used gravity model [19] to compare our predictions with:

$$m_{Gravity}(i, j) = K_G \frac{P_i \cdot P_j}{d_{i,j}^2}, \quad i = 1, \dots, n, \quad j = 1, \dots, i - 1,$$

where  $K_G$  is a scaling factor, and  $d_{i,j}$  is the Euclidean distance (in kilometers) between the centroids of the regions. The results at the county level are shown in Figures 4(a). The plots correspond to models which minimize the least square errors, using:  $K_{All} = 16.9$  for the prediction made with the average number of trips in a day  $m_{All}$ ;  $K_{WM} = 71.4$  for the prediction made with the average number of trips on weekday mornings  $m_{WM}$ . and  $K_G = 58.4$  for the gravity model  $m_{Gravity}$ .

Correlations show very encouraging results, with  $R^2 = 0.59$  for the gravity model,  $R^2 = 0.73$  for the prediction made with all trips, and the best result  $R^2 = 0.76$  for predictions made considering only weekday morning trips. The resulting high correlation shows that the estimated OD matrices are able to resemble very well OD matrices generated using completely different information.

Using the best model  $m_{WM}$ , we compared our results with the tract level census data. At this level, noise is more evident (see Figure 4(b)), but still we can see on average a very good linear relationship between census estimation and our estimation.  $R^2 = 0.36$  in this case, which is however very high compared to the  $R^2 = 0.10$  of the gravity model<sup>6</sup>. The relatively low value of  $R^2$  compared to the county level analysis is partially due to the fact that the relationship seems less linear for cases when the census estimates less than 10 trips from tract to tract. This might be explained by the fact that census flows are estimated from a subsample, that might result in very small numbers for particular pairs of census tracts. Moreover, census estimates were not available for the same year as the mobile phone data, and origins and destinations of trips might have slightly changed (at this high level of spatial detail) between the two monitored periods.

We note that the scaling factor  $K_{WM}$  used for the last model  $m_{WM}$  corresponds to a share of monitored trips which is about 1.4% compared to the census estimations. This factor can be explained by the percentage of mobile phones selected (about 4.3%) and by the calling activity which is not very high in the morning. Other elements such as the fact that we are monitoring not only commuting flows might explain the remaining difference. Estimating  $K_{WM}$  allows to extrapolate the ODs computed using the mobile phone data to the whole population.

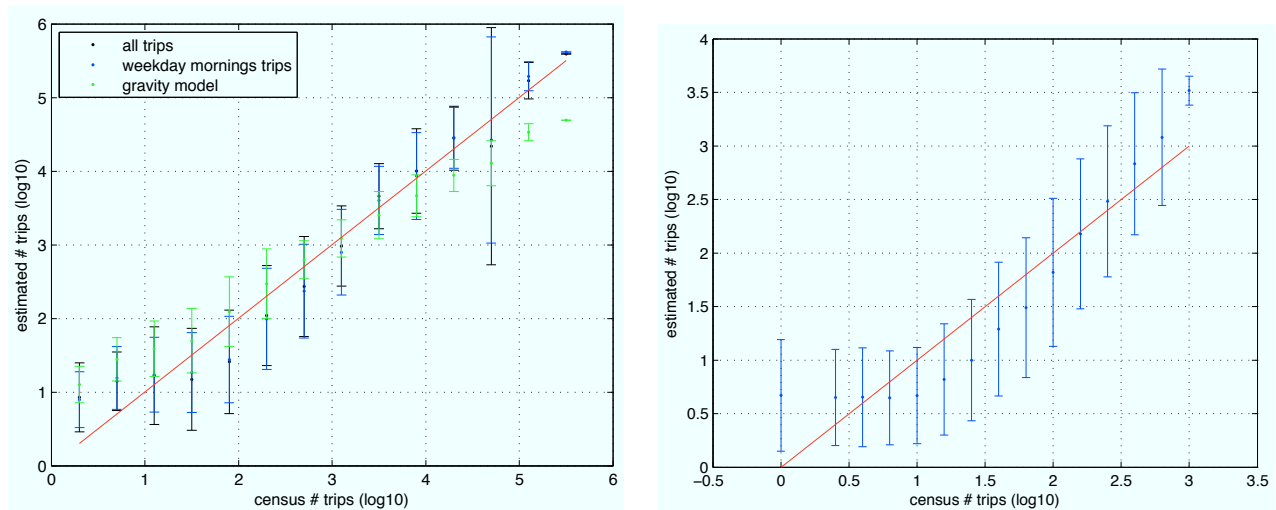
## V. NEW POTENTIALS

Origin-destination flows data estimated through census surveys have the following limitations (see [15]):

- The decennial census monitors "usual" days to avoid local or regional anomalies such as transit strike or severe weather, on a single sampling day. However, this tends to hide the less common uses, such as telecommuting once every two weeks or carpooling once a week due to the ever-changing life and work patterns.
- According to the definition, the census dataset does not include non-work trips, and modelers have to develop relationships between work and non-work trips.
- The census data is based on a fixed point "snapshot" approach, and so transportation planners can only interpret data over geographic space, rather than over time.

<sup>5</sup>The work location has been estimated as the most frequent stop area on weekday morning 8-10am.

<sup>6</sup>We have also evaluated more sophisticated gravity-like models by optimizing the  $d$  exponent and substituting the populations with the total estimated number of trips outgoing or incoming an area, but have still obtained  $R^2 < 0.3$ .



(a) County level. All trips, weekday morning trips and gravity model.

(b) Tract level. Weekday morning trips

Fig. 4. Comparison between mobile phone and census OD estimates. Error bars are showing one standard deviation from the average.

Compared with traditional census data, our methodology to detect OD matrices from mobile phone traces has several advantages:

- It can capture the weekday and weekend patterns as well as seasonal variations.
- It can capture work trips and non-work trips, which is essential for trip chaining and activity based modeling.

For these reasons, they could then be used to complement traditionally generated OD matrices providing a very fine grain spatialtemporal patterns of mobility.

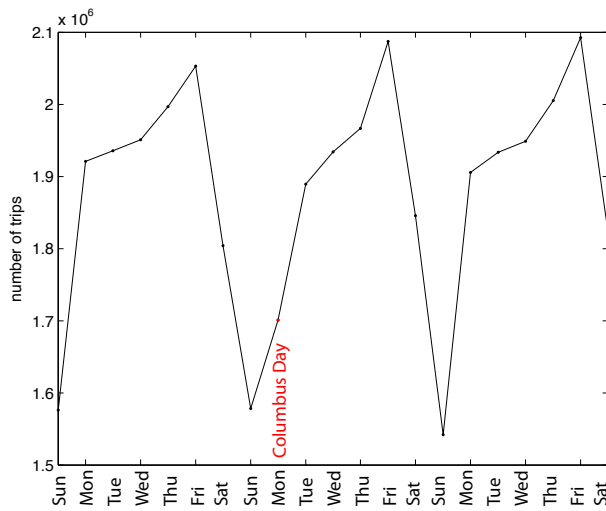
In the following subsections, examples of these potentials are shown.

### A. Temporal analysis

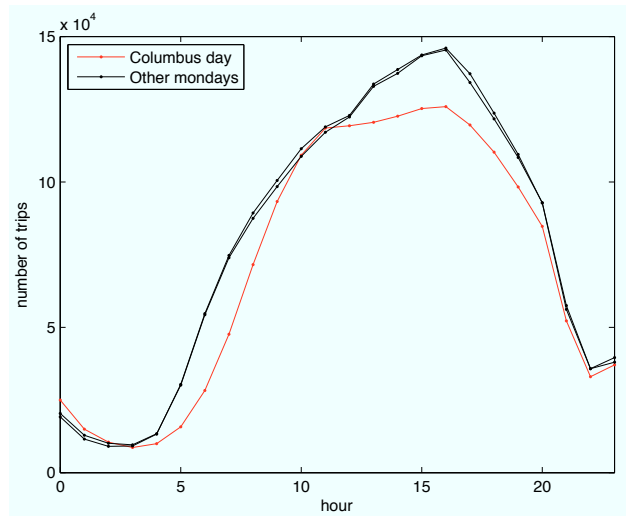
While the census gives only a static information about origin-destination flows, the OD matrices derived from mobile phone data allows us to appreciate the differences in travel demand over time. Figure 5(a) shows the total *daily* travel demand for 3 different weeks in October 2009. A weekly pattern clearly appears in the travel demand, with the minimum over weekends (especially sundays) and a maximum over fridays. Moreover, Figure 5(a) shows a particular change in travel demand in the second monday (day number 9 in figure), corresponding to Columbus Day. For a better look at this pattern, we plot the *hourly* travel demand for Columbus Day compared to the other mondays (see Figure 5(b)). We clearly see a higher travel demand in the first 2 hours of the day, followed by lower demand from 4 to 9, and from 12 to 20, due to the holiday.

### B. Spatiotemporal analysis

Our methodology can capture very fine grain OD matrices in both spatial and temporal scale, essential data for understanding transport demand and transport modeling especially during special events. For example, Figure 6 compares the incoming flows toward the Boston Baseball stadium Fenway Park. We compare two different days: Sunday October 11th where the local baseball team the Red Sox played against the Angels in a postseason game, and an average sunday without events. As it can be seen from the figures, we are able to capture the increasing incoming flow due to the special event, both in terms of new origins of trips, and in volumes of flow. Further studies with the same dataset have also shown regular spatial patterns of attendee origins based on the type of event, information that would be very valuable for event management [5].



(a) Number of trips per day, over a 3 weeks interval.



(b) Number of trips per hour, over 3 different Mondays.

Fig. 5. Temporal variation in the number of trips.



(a) Normal Sunday.



(b) Day of the Red Sox game.

Fig. 6. Incoming trips in the Fenway Park area. Flow volume is represented by the thickness of the line.

## VI. DISCUSSION AND CONCLUSION

As shown in this study, pervasive datasets such as mobile phone traces provide rich information to support transportation planning and operation. Meanwhile, some related limitations should also be addressed when applying these datasets in mobility analysis. A crucial parameter to take into account is the localization error, which limits the minimum size of the regions that can be considered. Other elements that can affect the statistical results include: 1) the market share of the mobile phone operator from which the dataset is obtained, 2) the potential non-randomness of the mobile phone users (e.g. teenagers), 3) calling plans which can limit the number of samples acquired at each hour or day, 4) number of devices that each person carries. Moreover, due to the fact that the considered dataset is event-driven (location measurements available only when the device makes network connections) the connection patterns of users can affect the possibility to capture more or less trips. This last limitation could be solved by continuous location readings from GPS devices, which would however require the users consent. An hybrid approach could be envisioned, integrating both event-driven and continuous location measurements, as the current method can be easily generalized to different datasets with different spatio-temporal resolutions. Nonetheless, the analysis performed on the inter-event time, the spatial distribution of mobile phone users, and comparisons with census estimations confirm that the mobile phone data represent a reasonable proxy for human mobility.

Apart from reproducing data derived by means of expensive census surveys, our methodology to detect OD matrices from mobile phone traces has several advantages: 1) It can capture the weekday and weekend patterns as well as seasonal variations. 2) It can capture work and non work trips. 3) It can produce real time, continuous OD matrices which can capture the very fine grain spatialtemporal patterns of urban mobility.

Future work will involve reproducing the analysis for other cities, in order to understand which parameters influence the scaling factors to be used to extrapolate the ODs computed using the mobile phone data to the whole population. The research output will give transport planners an automatic and systematic way to understand the dynamics of daily mobility in a real complex metropolitan area.

## REFERENCES

- [1] X. Zhou and H. S. Mahmassani, "Dynamic origin-destination demand estimation using automatic vehicle identification data," *IEEE Transactions on Intelligent Transportation Systems*, vol. 7, no. 1, pp. 105–114, 2006.
- [2] S. Baek, Y. Lim, S. Rhee, and K. Choi, "Method for estimating population od matrix based on probe vehicles," *KSCE Journal of Civil Engineering*, vol. 14, no. 2, pp. 231–235, 2010.
- [3] M. L. Hazelton, "Some comments on origin-destination matrix estimation," *Transportation Research Part A: Policy and Practice*, vol. 37, no. 10, pp. 811 – 822, 2003.
- [4] F. Girardin, F. Calabrese, F. D. Fiore, C. Ratti, and J. Blat, "Digital footprinting: Uncovering tourists with user-generated content," *IEEE Pervasive Computing*, 2008.
- [5] F. Calabrese, F. Pereira, G. DiLorenzo, and L. Liu, "The geography of taste: analyzing cell-phone mobility and social events," in *International Conference on Pervasive Computing*, 2010.
- [6] D. Quercia, G. DiLorenzo, F. Calabrese, and C. Ratti, "Mobile phones and outdoor advertising: Measurable advertising," *IEEE Pervasive Computing*, vol. 10, no. 2, pp. 28–36, 2011.
- [7] R. Bolla and F. Davoli, "Road traffic estimation from location tracking data in the mobile cellular network," in *IEEE Wireless Communications and Networking Conference*, vol. 3, 2000, pp. 1107 –1112.
- [8] F. Calabrese, C. Ratti, M. Colonna, P. Lovisolo, and D. Parata, "Real-time urban monitoring using cell phones: A case study in rome," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 1, pp. 141–151, 2011.
- [9] R. Cayford and T. Johnson, "Operational parameters affecting use of anonymous cell phone tracking for generating traffic information," in *Institute of transportation studies for the 82th TRB Annual Meeting*, 2003.
- [10] J. White and I. Wells, "Extracting origin destination information from mobile phone data," in *Road Transportation and Control*, 2002.
- [11] C. Pan, J. Lu, S. Di, and B. Ran, "Cellular-based data-extracting method for trip distribution," *Journal of Transportation Research Board*, pp. 33–39, 2006.
- [12] N. Caceres, J. Wideberg, and F. Benitez, "Deriving origin destination data from a mobile phone network," *Intelligent Transport Systems, IET*, vol. 1, no. 1, pp. 15 –26, 2007.
- [13] K. Sohn and D. Kim, "Dynamic origin-destination flow estimation using cellular communication system," *Vehicular Technology, IEEE Transactions on*, vol. 57, no. 5, pp. 2703 –2713, sept. 2008.
- [14] Y. Zhang, X. Qin, S. Dong, and B. Ran, "Daily o-d matrix estimation using cellular probe data," in *Transportation Research Board Annual Meeting*, 2010.
- [15] CTPP, "Us department of transportation, census transportation planning products," <http://www.fhwa.dot.gov/ctpp/>, 2010.
- [16] M. Gonzalez, C. Hidalgo, and A.-L. Barabasi, "Understanding individual human mobility patterns," *Nature*, vol. 453, no. 7196, pp. 779–782, 2008.
- [17] L. Liu, F. Calabrese, A. Biderman, and C. Ratti, "The law of inhabitant travel distance distribution," in *European Conference on Complex Systems*, Warwick, UK, September 2009.
- [18] MassGIS, "Census 2000 tracts datalayer description," <http://www.mass.gov/>, 2010.
- [19] J. Anderson, "A theoretical foundation for the gravity equation," *The American Economic Review*, vol. 69, pp. 106–116, 1979.