

# IBM Research Report

## Predicting Personal Mobility with Individual and Group Travel Histories

Giusy Di Lorenzo<sup>1,2</sup>, Jonathan Reades<sup>3</sup>, Francesco Calabrese<sup>1,2</sup>, Carlo Ratti<sup>2</sup>

<sup>1</sup>IBM Research  
Smarter Cities Technology Centre  
Mulhuddart  
Dublin 15, Ireland

<sup>2</sup>Senseable City Laboratory  
MIT  
77 Massachusetts Avenue  
Cambridge, MA USA

<sup>3</sup>Centre for Advanced Spatial Analysis  
University College London  
1-19 Torrington Place  
London WC1E 7HB UK



Research Division

Almaden - Austin - Beijing - Cambridge - Haifa - India - T. J. Watson - Tokyo - Zurich

# Predicting Personal Mobility with Individual and Group Travel Histories

## Abstract

Understanding and predicting human mobility is a crucial component of a range of administrative activities, from transportation planning to tourism and travel management. In this paper we propose a new approach that predicts the location of a person over time based on both individual and collective behaviors. The system draws on both previous trajectory histories and the features of the region – in terms of geography, land use, and points of interest – which might be ‘of interest’ to travellers. We test our approach’s effectiveness using a massive data set of mobile phone location events compiled for the Boston metropolitan region, and experimental results suggest that the predictions are accurate to within  $1.35km$  and demonstrate the significant advantages of incorporating collective behavior into individual trip predictions.

## 1 Introduction

Predicting individual and group travel patterns is a crucial component of transportation modelling and management; historically, these have been approached in two ways: trip-based models, in which mobility is typically considered in aggregate as area-based Origin/Destination matrices, and activity-based models, in which individual agents are assigned ‘resources’ to obtain and some number of trips are needed to access them (see, for instance, Gendreau and Marcotte, 2002). However, the rapid spread of the Global Positioning System (GPS) and other high-resolution tracking tools enables us to adopt a very different approach to travel prediction, one that is rooted in data mining methodologies and that considers individual and group preferences simultaneously in order to deliver spatially- and temporally-sensitive forecasts.

Our proposed system extracts historical patterns of movement from a large database of mobile phone subscriber records, and employs a weighted mix of individual and aggregate statistical models to predict the most likely next location of each member of the group. We test the approach’s performance against a randomly-selected population of 2,000 users and find that it is accurate, on average, to within  $1.35km$  across an entire American metropolitan region. By taking into account the specific social, economic, and geographical characteristics of places within the study area, we are able to move beyond the more abstract mobile-derived models of Song et al. (2010a), and the results show how such data can underpin a range of concrete services. We envision a system able to support applications such as vehicle routing to avoid predicted congestion, and communications from business or government at a time when they are most likely to influence individual and household choice.

## 2 Context

The idea of using collective behavior to inform travel predictions is not new, Adler and Ben-Akiva (1979) set out a model in which a ‘pool’ of alternatives for a household was conditioned on the choices made by other households in the same origin district (see also Axhausen and Gärling, 1992, for a good overview). However, the increasing availability of fine-grained, but extensive data sets from GPS devices and mobile phone networks is providing us with a more comprehensive view of activity and mobility at the urban scale than travel diaries can possibly do on their own (Ahas et al., 2010b, pp.52-53). This expanded capacity has already been used to explore the behaviours of large groups (see, for instance, Li et al., 2008; Liao et al., 2007; Zheng et al., 2008), but it also enables us to shed light on hitherto invisible *intra*-personal variation in travel activity, which has “played a minor role in travel behaviour research in comparison with research on interpersonal variability...” (Schlich and Axhausen, 2003, p.14).

This historical blind spot is crucial since it has been widely expected that people’s behaviours will become increasingly complex through the freedom afforded by mobile telecommunications and increased personal mobility. The point is that a decrease in the constraints operating on individual and household choice is likely to lead to an increased diversity in travel patterns (Schlich and Axhausen, 2003, p.14). Moreover, even before the introduction of technologies that decreased the costs of being ‘out’ there was strong evidence to suggest that people value the ability to perform – and enjoy – unplanned activities (Axhausen and Gärling, 1992, p.333). As a result, some researchers have argued that we should move away from normative terms such as ‘home’ and ‘work’ towards the more flexible and indeterminate terminology of ‘meaningful places’ and ‘anchor points’ as a better way to think about location (Ahas et al., 2010c, p.5).

Of the two new technologies being employed, GPS was the first to be used extensively in travel studies since it could be linked to individual activity at very high resolutions. The power consumption and weight profiles of early GPS units meant that vehicular travel was the principal focus of pioneering studies such as those of Krumm (2006) and Krumm and Horvitz (2006). Technological improvements have changed this picture dramatically, enabling even children to participate in such studies (see, for instance, Gong and Mackett, 2009). However, the use of a separate tracking device introduces several new challenges: that the device is not charged or not switched on by the study participant; that it fails to function indoors or in urban environments; or that it is accidentally (or deliberately) not carried by the participant in ‘exceptional’, but analytically important, circumstances such as going on holiday or meeting colleagues for an after-work event.

These challenges mean that there remain serious reliability and management issues associated with the use of dedicated GPS systems for longitudinal or large-scale behavioural research (Ahas et al., 2010b, p.53). This dynamic helps to explain the significant interest amongst researchers in moving towards the use of mobile phones as study platform. An important intermediate stage in this process was embodied in Eagle and Pentland’s (2006; 2009) ‘reality mining’ approach, which offered participants from M.I.T. University the use of a free mobile phone with a tracking and logging tool enabled for the duration of the study. However, this method does not avoid the basic cost of subsidising devices and so Pentland’s SenseNetworks (2008)

has shifted to a crowd-sourced approach, and most research is now predicated on extensive relationships with operators able to collect locational data on a much larger scale from the network itself.

The sheer volume of data generated by modern digital networks has been shown to have great potential for analysing human society (see, for instance, Onnela et al., 2007; Lazer et al., 2009). What has emerged from this type of work is that a high level of predictability exists at the individual level, suggesting that it may be possible to derive universal ‘laws’ to characterise personal mobility for the vast majority of a country’s population (González et al., 2008; Barabási, 2010; Song et al., 2010b). At the finer scale, Ahas et al. (2010c) are perhaps the most notable exemplar of those working with network-collected data to understand individual behaviour: they are able to infer significant sites – such as home, work, and summer cottage locations – for Estonian residents and visitors by examining how location changes over time and factoring in the role of weekends, holidays, and even seasonal variation (e.g. Ahas et al., 2007). And while these strands of research demonstrate the essential validity of mobile network data for understanding spatial and temporal movement patterns, they are not particularly concerned with predicting specific trajectories for a given individual.

### 3 Predictive model

To what extent individual trajectories are predictable on any spatial and temporal scale? Song et al. (2010b) find that Zipf’s law is a suitable approximation of an individual’s cumulative place-frequency distribution, and that the probability of returning to a place is correlated with the number of times that same location has been visited in the past. However, this descriptive approach does not directly bind the findings to the underlying reason for moving, which is obviously going to be an important factor in a predictive model. Fortunately, Ahas et al. (2010a) observe that temporal patterns found at hourly, daily, and weekly scales can be connected not only to specific activities, but also to their geographical distribution across a metro region. For example, the further an employee is from an office in the Central Business District, the earlier their daily commute to work begins.

The broad implication for our system is that near randomness in individual trajectories of the sort that Barabási (2010) documents for Hasan Elahi is quite rare, and that most of us follow a fairly regular routine that is structured by the interaction between activity, space, and time. With this research in mind, our basic hypothesis is therefore that if we have a population of  $m$  individuals, then for each member  $u \in U_m = \{1, 2, \dots, m\}$  there exists some function  $x(u)_k = i$  to predict their location  $i$  at time  $k$ . Furthermore, we also postulate that their location at time  $k + 1$ , as given by the function  $x(u)_{k+1}$ , is in some way conditioned by their prior location at time  $k$ .

In and of itself this basic formulation will not yield a useful model; however, the frequency distributions found by Song et al. (2010b) and Ahas et al. (2010a) enable us to connect this problem to observations of recurrent behaviours such as travelling to work, going out for dinner on a weekend, or collecting the children from school. We may therefore *also* anticipate that our predictions can be improved by using information about where an individual was to be found at times  $k - T, k - 2T, \dots$ , where  $T$  is the relevant periodicity in an individual’s travel pattern, be it daily, weekly,

or yearly. Such a formulation is in line with Song et al.’s (2010b, p.1021) finding that “... despite our deep-rooted desire for change and spontaneity, our daily mobility is, in fact, characterized by a deep-rooted regularity.”

Of course, the persistent search for novelty – trying a new restaurant, taking a new route to work, or checking out a new nightclub, for instance – is also important for any kind of prediction platform, and our approach presupposes that land use, points of interest, and group behaviour will help us with this process, even when the individual probability of genuinely unpredictable behaviour becomes almost vanishingly small. Consequently, our final hypothesis is that we can further improve the accuracy of the predicted destination by incorporating the aggregate past behavior of the group  $U$  as a whole and the distribution of possible ‘activities’ and ‘attractors’.

For instance, we might infer that the majority of people setting out from their ‘home’ location between 6 a.m. and 10 a.m. are headed to work. In the case of a tourist this would obviously be an error, but if we can narrow the individual’s reference group  $U$  to include only other tourists then we can rule out unlikely destinations (e.g. suburban office locations). This process enables us to fine-tune our predictions based on those destinations that are most likely to be relevant to the visitor, commuter, or stay-at-home parent given the place, time-of-day, demographics, etc. (Calabrese et al., 2010b).

### 3.1 Individual behavior

Let us begin to develop the model by dividing the study area  $N$  into  $n$  grid cells such that any location can be denoted as  $i \in N_n = \{1, 2, \dots, n\}$ . We can now model the conditional probability  $P_I$  that an individual, having been in location  $i$  at time  $k$ , will move to location  $j$  at time  $k + 1$  as follows:

$$P_I(x_{k+1} = j | x_k = i) = \frac{\sum_{k_t=1}^{\lfloor k/T \rfloor} f_I(x_{k-Tk_t+1} = j | x_{k-Tk_t} = i)}{\lfloor k/T \rfloor} \quad \forall j \in N_n, \quad (1)$$

where  $\lfloor a \rfloor = \max\{n \in N : n < a\}$ ,  $\forall a \in R$ .

The frequency  $f_I$  on the right-hand side of the equation is defined as:

$$f_I(x_{k+1} = j | x_k = i) = \begin{cases} 1 & \text{if } x_{k+1} = j \text{ and } x_k = i \\ 0 & \text{elsewhere} \end{cases} .$$

What we are suggesting here is that the probability that cell  $j$  is the *next* destination of a given traveller is equal to the frequency with which they *previously* visited that cell when starting from cell  $i$  during the earlier periods  $k - T + 1, k - 2T + 1, \dots$ , and so on. If the selected individual has never been in cell  $i$  during those previous periods, then the frequency can be computed as:

$$f_I(x_{k+1} = j | x_k = i) = \begin{cases} 1 & \text{if } x_{k+1} = j \\ 0 & \text{elsewhere} \end{cases} .$$

## 3.2 Collective behavior

We now have a working model for personal mobility prediction, but one of our hypotheses is that we can use aggregate movement data to improve our predictions. Empirical support for this approach comes from Song et al. (2010b, p.1021), whose findings showed very little sensitivity to individual demographics. Conceptually, there are two ways in which this feature of group behavior reinforces the predictive component: first, aggregate mobility data can help us to predict when an individual is likely to change location and over what distance this movement is likely to occur; and second, it can aid us in predicting the type of place that he or she is likely to visit next. In other words, we can design the model so that the probability that an individual chooses a given destination as their *next* ‘port-of-call’ is, in part, a function of what the population  $U$  as a whole chooses as their next destination. We can also structure the selection of likely destinations so that it is influenced by the distance of travel and the presence of points of interests (POIs) similar to the ones that ‘the collectivity’ has visited.

### 3.2.1 Distance

Many models assume that trip-distance follows gravity-like decay such that the probability of a trip of length  $d$  is inversely proportional to  $d^2$  or some other exponential function. As discussed in the Context section, more recent analyses have suggested very high degrees of consistency in how far people travel and how frequently they return to just a few, key locations (González et al., 2008). Furthermore, this consistency appears to be largely ‘scale free’ in that people covering hundreds of kilometres in a week are just as predictable in statistical terms as those covering only dozens (Song et al., 2010b, p.1020).

Given these characteristics of travel, we can improve our predictions if we recognize that, in any given interval, some ‘translations’ are effectively impossible – or, at least, very unlikely – because of the distances (in the form of travel time) involved. We therefore constrain the model using a distance-based probability function:

$$P_D(x_{k+1} = j | x_k = i) = f_d(d_{ij}, k),$$

What we are suggesting here is that the conditional probability of trip  $i \rightarrow j$  is, at least in part, a function of the distance  $d_{ij}$  between cells  $i$  and  $j$ .  $f_d(d, k)$  is therefore the normalized frequency of *all* trips undertaken by the reference population  $U$  which covered distance  $d$  at times  $k - T, k - 2T, \dots$ . Note too that this function should include the possibility that the person does *not* change location at all between times  $k$  and  $k + 1$  such that:

$$f_d(0, k) = \frac{1}{m \lfloor k/T \rfloor} \cdot \sum_{u=1}^m \sum_{k_t=1}^{\lfloor k/T \rfloor} P(x(u)_{k-Tk_t+1} = x(u)_{k-Tk_t}).$$

where  $P(x(u)_{k-Tk_t+1} = x(u)_{k-Tk_t})$  is the ‘non-moving’ probability.

### 3.2.2 Time

The models of González et al. (2008) and Song et al. (2010b,a) incorporate time as an ordering of locational events, and it is therefore quite surprising that Song et al. obtain 93% predictability for an individual’s ‘next location’ using nothing more than the frequency of visits to a given place (which they find to follow a power law) and their sequencing. However, while many of our daily activities are strongly ordered – getting up, going to work, going to the shops, going home... – this is a far cry from predicting when we’ll next go on holiday or head off to visit our family.

In other words, although day-of-week and time-of-day have important effects on travel (Schlich and Axhausen, 2003, p.26) and upon activity participation (Axhausen and Gärling, 1992, p.335), time-of-year Ahas et al. (2010a) also plays a role. By way of an example, at the more extreme end of travel periodicity is long-distance travel during the summer months in Scandinavia and North America to ‘cottage country’ (e.g. Williams and Kaltenborn, 1999; Ahas et al., 2007). Schlich and Axhausen (2003, p.26) also note that the failure to consider ‘activity chains’ and their temporal scale is problematic when we start thinking about average distances and the variability around this metric, and this is another reason why a more nuanced conception of time and periodicity is useful in making accurate predictions.

### 3.2.3 Points of Interest

Another short-coming of most purely statistical approaches to mobility is that many of them seem to assume that movement is a kind of random process so that, while frequency of visits affects probability, the next location is always a throw of the dice. Since this is clearly not the case, the third part of the model is structured by the distribution of points of interest (POIs, which we take as a rough proxy for the range of possible activities) across the study area. If we associate a set of POIs, belonging to  $Q$  categories, with each location  $i$ , then we can characterize a place with a vector  $poi_i$  whose definition is given by  $\{poi_i(1), \dots, poi_i(Q)\}$ . Using the collective movement traces, we can infer the probability that someone will be found near a POI of category  $q$  at time  $k$  as:

$$f_{POI}(q, k) = \frac{\sum_{k_t=1}^{\lfloor k/T \rfloor} \left( \frac{\sum_{u=1}^m poi_{x(u)_{k-Tk_t}}(q)}{\sum_{u=1}^m \sum_{q'=1}^Q poi_{x(u)_{k-Tk_t}}(q')} \right)}{\lfloor k/T \rfloor}.$$

From this, it follows that the probability of finding a person in cell  $j$  as a function of the POIs available can be written as follows:

$$P_{POI}(x_k = j) = \sum_{q=1}^Q f_{POI}(q, k) \frac{poi_j(q)}{\sum_{q'=1}^Q poi_j(q')}.$$

### 3.2.4 Land use

A similar approach can be applied when using land use data in place of POI information. In this sense, land use data supplements the POI component of the model that we set out above by reinforcing linkages between between places and times. For instance, proximity to school land will likely be factor in predicting the whereabouts of the principal childminder in a family at the close of the school day (Ahas et al.,

2010b, p.48). Consequently, the same equations can be derived for a grid cell  $i$  with a descriptive vector  $lu_i$  of land use data given by  $\{lu_i(1), \dots, lu_i(R)\}$  such that:

$$P_{LU}(x_k = j) = \sum_{r=1}^R f_{LU}(r, k) \frac{lu_j(r)}{\sum_{r'=1}^R lu_j(r')}.$$

### 3.3 Combined behavior

Combining the components outlined above, we obtain the following probability function:

$$P_C(x_{k+1} = j | x_k = i) = P_D(x_{k+1} = j | x_k = i) P_{POI}(x_{k+1} = j) P_{LU}(x_{k+1} = j) \cdot \left( \sum_{j'=1}^n P_D(x_{k+1} = j' | x_k = i) P_{POI}(x_{k+1} = j') P_{LU}(x_{k+1} = j') \right)^{-1} \quad \forall j \in N_n. \quad (2)$$

This equation simply states that the conditional probability of aggregate movement between locations  $i$  and  $j$  is given by the joint probability: that a journey of distance  $d_{ij}$  can occur between times  $k$  and  $k + 1$ ; that a journey to a location  $j$  with POIs  $q$  will happen at time  $k + 1$ ; and, finally, that a journey to a location  $j$  with land use  $r$  will happen at time  $k + 1$ . The second part of the equation is a scaling factor which ensures that the sum of all probabilities is equal to one since  $\sum_{j=1}^n P_C(x_{k+1} = j | x_k = i) = 1$ .

To put this in more concrete terms, we're suggesting that the probability of a given trip is related to: the distance people usually travel (and can possibly travel) at that time of day; the general land use in the area to which they are travelling; and the points of the interest available to them when they get there. For instance, people, on the whole, do not travel  $200km$  at 6 p.m. to a place whose primary land use is industrial and where the main point of interest is a refinery. They do, however, often 'nip out to the shops' to purchase last-minute food items at a nearby shopping center – i.e. retail land use – containing such points of interest as a grocery, a bakery and, perhaps, a DVD rental shop.

We now combine the individual (1) and group (2) probability functions in such a way as to predict an individual's future behavior as a function of both sets of inputs. This prediction is given by the equation:

$$P(x_{k+1}(u) = j | x_k(u) = i) = (1 - \alpha(k)) P_I(x_{k+1}(u) = j | x_k(u) = i) + \alpha(k) P_C(x_{k+1} = j | x_k = i), \quad \forall j \in N_n, \quad (3)$$

Note that the combinatorial parameter  $\alpha \in [0, 1]$  can vary over time, enabling us to adjust the model so that it reflect periods where individual behavior is more important and periods where collective behavior is better able to model future locational decisions.



### 3.4 Further considerations

The proposed model could also be extended to take into account the ways in which the distribution of activity may change with time, and the impact that the selection of the reference population used to calculate  $P_C$  can have on the resulting predictions. Although these two features are not integral to the model *per se*, they may well prove useful for refining the predictions made for an individual's location at time  $k + 1$  since they both serve to remove noise from the system.

#### 3.4.1 Forgetting old behaviour

As currently framed, the model begins with any time  $k > T$  and all periodicities  $T$  are taken into account in calculating the probability distribution function. However, over time it may become useful to add a ‘forgetting factor’ which reflects the fact that more recent data points contain more timely information than older ones. For instance, a user's preferences may change with time.

The ‘forgetting factor’ ( $\lambda$ ) can be introduced by modifying the formula (1) as follows:

$$P_I(x_{k+1} = j | x_k = i) = \frac{1}{\lfloor k/T \rfloor \sum_{k_t=1}^{\lfloor k/T \rfloor} \lambda^{\lfloor k/T \rfloor - k_t}} \cdot \sum_{k_t=1}^{\lfloor k/T \rfloor} \lambda^{\lfloor k/T \rfloor - k_t} f(x_{k-Tk_t+1} = j | x_{k-Tk_t} = i), \forall j \in N_n.$$

Similar changes would also be required to the equations covering aggregate movement probability. However, the key point here is that  $\lambda$  ranges between 0 and 1, and that the closer it is to 0, the less older samples are factored into the resulting probabilities. Generally speaking, samples older than  $\tau = \frac{1}{1-\lambda}$  carry a weight that is less than  $\cong 0.3$ .

#### 3.4.2 Selecting an appropriate reference population

As we noted above, it is also clearly going to be of importance that an appropriate sample is selected for the aggregate behavior portion of the model. For individuals without a travel history, it might be useful to start with the general population's activity patterns; where additional data can be brought to bear, the model can be customized such that the reference group is similar to the person whose movement we are trying to predict.

To put it more simply, if we are trying to model the future activities of a businessperson then it is logical to select as inputs to  $P_D$ ,  $P_{POI}$ , and  $P_{LU}$  a collection of people who also commute for work purposes. Similarly, to model the mobility of elderly retirees it makes very little sense to include in the group model the activities of executives and students. The better the identification of a ‘dominant identity’ for a person, the more easily we can select a useful cohort whose behaviour can be incorporated into the prediction system.

In cases where the dominant behavioural pattern of an individual is uncertain or displays divergent tendencies, then varying collectivity weightings could be included in the combined model. Higher weights would be given to populations whose behaviors

are more similar to the observed individual’s own pattern of activity. For instance, a two-group model of equation (3) can be given by:

$$\begin{aligned}
P(x_{k+1}(u) = j | x_k(u) = i) = & \\
(1 - \alpha_1(k) - \alpha_2(k))P_I(x_{k+1}(u) = j | x_k(u) = i) + & \\
+ \alpha_1(k)P_{C_1}(x_{k+1} = j | x_k = i), & \\
+ \alpha_2(k)P_{C_2}(x_{k+1} = j | x_k = i), & \\
\forall j \in N_n. &
\end{aligned}$$

Here,  $\alpha_1$  and  $\alpha_2$  govern the relative weight of the two populations being used to drive the group travel probability function.

## 4 Case study: Boston Metro Area

To test our model we employed a mobile phone mobility data set covering the American state of Massachusetts, and we focussed in particular on the eight counties – Middlesex, Suffolk, Essex, Worcester, Norfolk, Bristol, Plymouth, Barnstable – that constitute the Boston metropolitan area. The data consists of pseudonymous location estimates collected by AirSage<sup>1</sup>, which uses cellular signalling data to generate real-time and long-term traffic analysis for individuals, firms, and planning/engineering departments. In all, the data set contains more than 800 million pseudonymous location estimations for nearly 1 million unique devices (20% of the metro area’s population of 5.5 million assuming one device per person) and covers 4 months’ worth of activity.

AirSage’s system collects and automatically anonymises the ‘locational events’ generated each time a mobile device connects to a cellular network. These events can be triggered by many types of network activity: when a call is placed or received (both at the beginning and at the end of a call); when a text message is sent or received; and when the user’s phone connects to the Internet, regardless of whether this is deliberate (e.g. to browse the web) or automatic (e.g. to check for new email). The rise of smartphones using Apple’s iOS or Google’s Android operating systems is thus especially important to this type of research since their ‘always on’ data connections trigger high-frequency locational updates.

The events captured by AirSage’s collection system are a superset of what are ordinarily contained in Call Detail Records (CDRs) such as those considered in White and Wells (2002) and González et al. (2008). This sampling difference is likely to account for the lower interval that we observe between network events in this data set when compared with what is reported in González et al. (2008): we find a mean of 260 minutes between events for the entire study population, and a median of 84 minutes. Note, however, that this is the frequency across a 24-hour period and that, since there are far fewer connections in the period between 11 p.m. and 6 a.m., the true temporal resolution during the daytime is quite a bit higher.

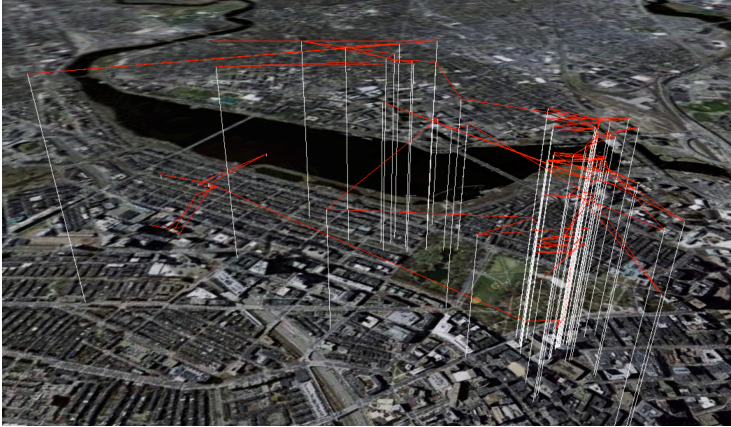
Moreover, the focus of the AirSage system on real-time traffic analysis means that the locational data has a higher resolution than what is normally available via

---

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

purely passive collection (cf. Ahas et al., 2010c), and that it comes with an estimation of a user’s current position within the cell through triangulation with a ‘Wireless Signal Extraction’ technology. Consequently, each location measurement  $m_i \in M$  is characterized by a position  $p_{m_i}$  that is expressed in latitude and longitude. When paired with the timestamp  $t_{m_i}$ , the data can then be linked together in a sequence of movements  $\{m_1 \rightarrow m_2 \rightarrow \dots \rightarrow m_n\}$ . From this it is, for instance, fairly easy to construct a ‘time geography’ (Hägerstrand, 1970) of the type shown in Figure 1, where the z-axis represents the time of day or week.

Figure 1: Time Geography of a Mobile Phone (October 2009)



## 4.1 Methodology

Of course, 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 that is further away but less heavily-loaded. To address this potential source of problems we applied a low-pass filter with a resampling rate of 10 minutes to the raw data; this follows an approach that was successfully tested against ground truth data from Rome, Italy (Calabrese and Ratti, 2006; Calabrese et al., 2011). In addition, since lesser localization errors might still generate fictitious trips, we also adapted a pre-processing step employed in the analysis of GPS traces (see Krumm, 2006; Krumm and Horvitz, 2006) 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:

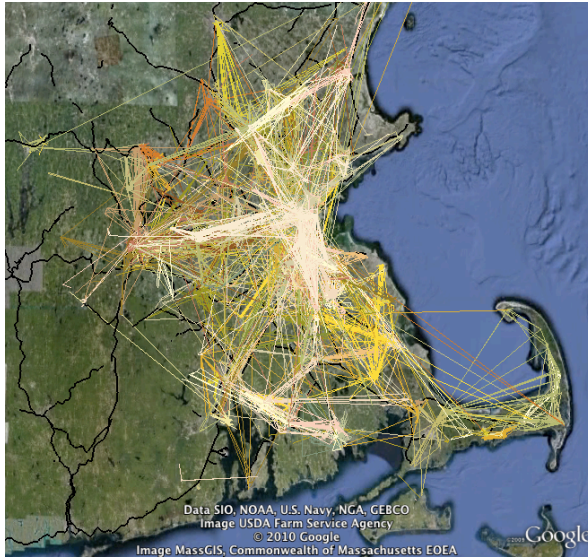
- We begin with a measurement series  $M_s = m_q, m_{q+1}, \dots, m_z \in M^{z-q-1}$  derived from a series of network connections over a certain time interval  $\Delta T = t_{m_z} - t_{m_q} > 0$ .
- We define an area with radius  $\Delta S$  – in this case, 1km to take into account the localization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

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

- Where this condition holds, then the points  $M_s = m_q, m_{q+1}, \dots, m_z \in M^{z-q-1}$  can be fused together such that the centroid becomes a ‘virtual location’ ( $p_s = (z - q)^{-1} \sum_{i=q}^z p_{m_i}$ , 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 location is associated to a  $500m \times 500m$  cell of a grid covering the greater Boston metropolitan area.

## 4.2 Results

Figure 2: Unique paths of 2,000 selected users

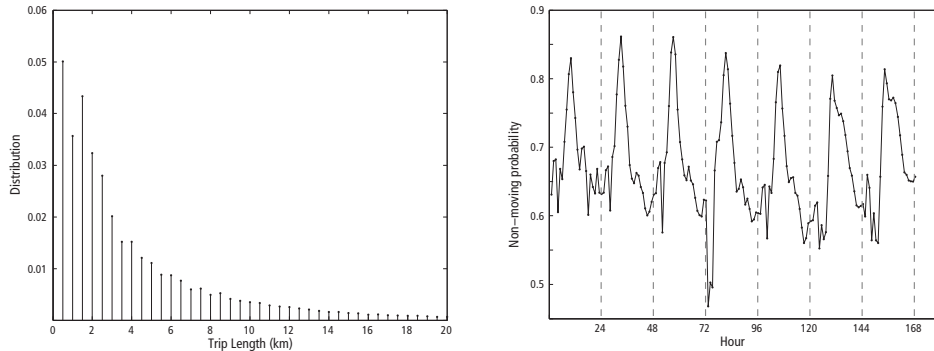


To test the prediction model, we extracted the complete traces of 2,000 randomly-selected users, each of whom made at least 100 network connections per day, with an individual average inter-event time of less than 1 hour in 75% of cases. The users’ raw locational data was then processed using the methodology outlined in Calabrese et al. (2010b) to obtain paths with sampling rates of 1 hour. After the clustering and filtering processes, users who still had more than one distinct identifiable location were positioned at the last one logged in that one hour window. The entirety of the group’s travel over the course of the study period is shown in Figure 2, and the density of paths clearly picks out major ‘destinations’ (in terms of employment, goods and services) and travel corridors such as the Boston/Braintree axis, Worcester/Framingham, and Providence (Rhode Island)/New Bedford. With the exception of this last corridor, which was excluded because the users passed out of the state of Massachusetts, these trajectories constitute the principal inputs into equation (1) above (see page 5).

Taken together, the 2,000 path histories also supply the aggregate data used in equation (2) on page 8. The normalized frequency of trips by distance travelled ( $f_d(d_{ij}, k)$ ,  $d_{ij} > 0$ , averaged over all values of  $k$ ) is shown in Figure 3a, and next to it the time-varying ‘non-moving probability’ ( $f_d(0, k)$  for  $k$  of 1 hour, and  $T = 1$  week)

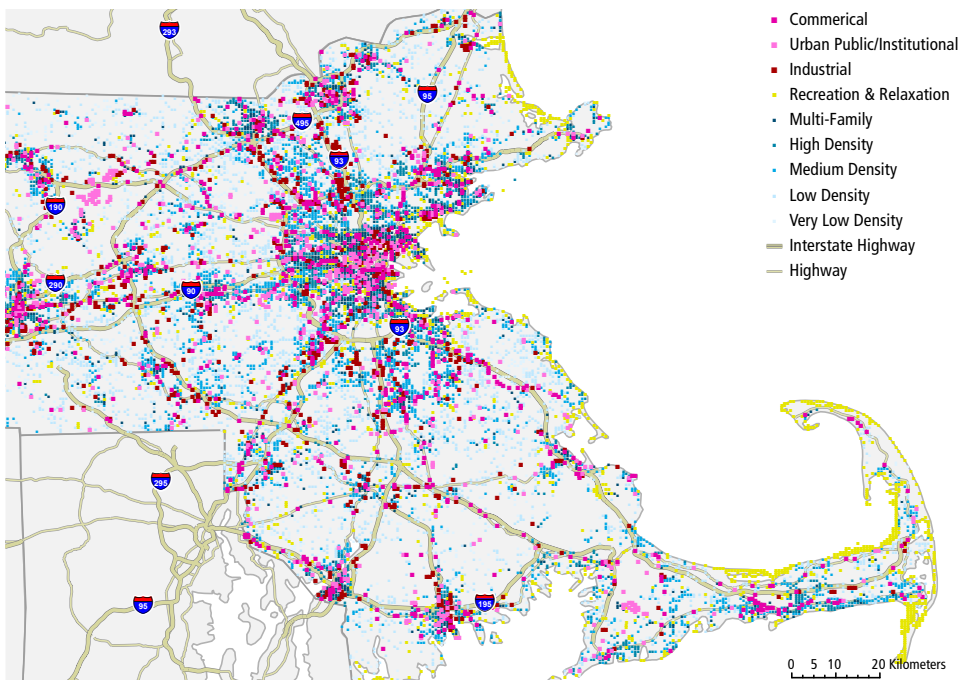
Figure 3: Collective Moving & Non-Moving Distributions

(a) Normalized frequency of trips as function of distance. (b) Non-moving probability in one week period (starting Monday 12am).



that is shown in Figure 3b. Figure 3a is organised so that the sum of all trip lengths is 1; however, since the proportion of trips of less than  $2\text{km}$  is just 16%, it is clear that the overall range of trip lengths is quite broad. The period of 168 hours (i.e. 1 week) is consistent with earlier results presented in Reades et al. (2007) and Calabrese et al. (2010a).

Figure 4: Map of Dominant Land Uses in Central and Eastern Massachusetts

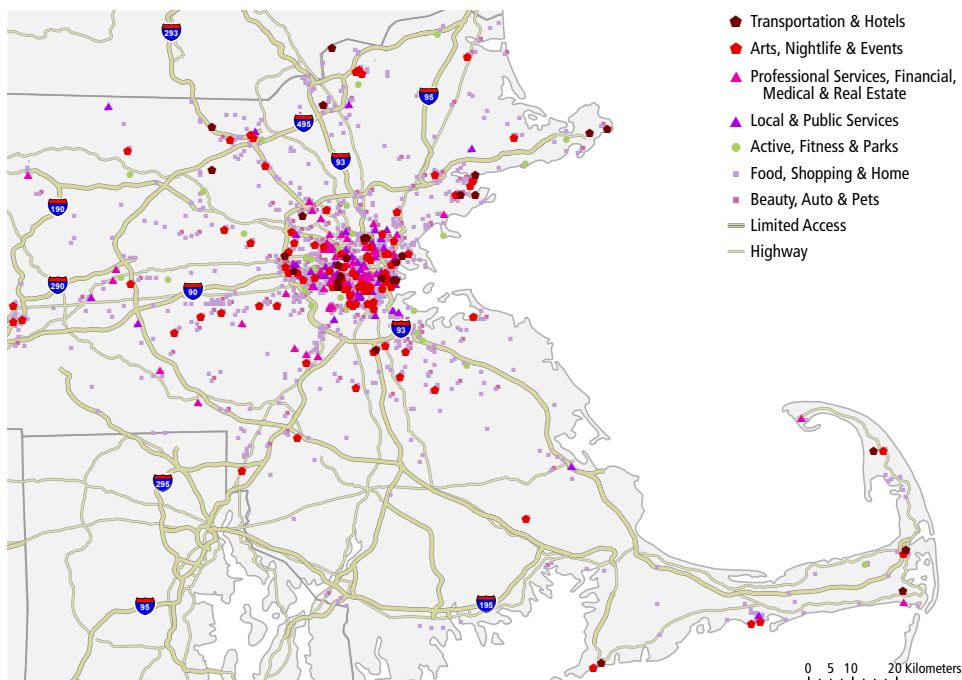


Of course, we can not directly infer individual or group activities from GPS or mobile phone traces since many activities could be performed at a same location – shopping, banking, and having a dental checkup, for instance. Consequently, in this model we will use information about an area’s resources as a broad proxy for the types of activities that might occur there. As discussed above, this information is drawn from two sources: land use and POI data. Since the total number of land uses and POIs is quite high, Figures 4 and 5 show only the most relevant categories and super-

categories (i.e. combinations of primary categories); we have suppressed categories such as ‘Cranberry Bog’ and ‘Waste Disposal’ for the sake of legibility, though they were included in the running of the model.

Land use data was drawn from Massachusetts’s MassGIS platform<sup>2</sup>, and was grouped into uniform  $500m \times 500m$  parcels. Figure 10 shows the average preference of users for each of the main super-categories (there were 33 underlying primary categories), where  $1/T \sum_{k=1}^T f_{LU}(r, k), r = 1, \dots, R$ . Taken together, multi-family and high-density residential areas account for almost 50% of the the most visited areas, followed by urban institutional and commercial areas. As might be expected, land use preferences (in terms of time spent in an area of a particular category) change over time, and Figure 10 shows how the distribution differs for two different times of day on a typical Monday in the data set.

Figure 5: Map of POIs in Central and Eastern Massachusetts



POIs were extracted from Yelp’s web service<sup>3</sup> with the same resolution as the land use data, and grouped into 22 categories. Figure 5 shows the spatial distribution of points of interest, while Figure 11 shows the average preference of users for the different categories  $1/T \sum_{k=1}^T f_{POI}(q, k), q = 1, \dots, Q$ . Looking at the most visited areas, food-related POIs cover around 50% of the preferences, followed by beauty, shopping and medical. Other categories have very low impact. In much the same way that land use varies with time, POI preferences change as well, and Figure 11 shows the principal POIs for the same time periods that are shown in Figure 10.

<sup>2</sup><http://www.mass.gov/mgis/lus2005.htm>

<sup>3</sup><http://www.yelp.com/boston/>

## 5 Prediction

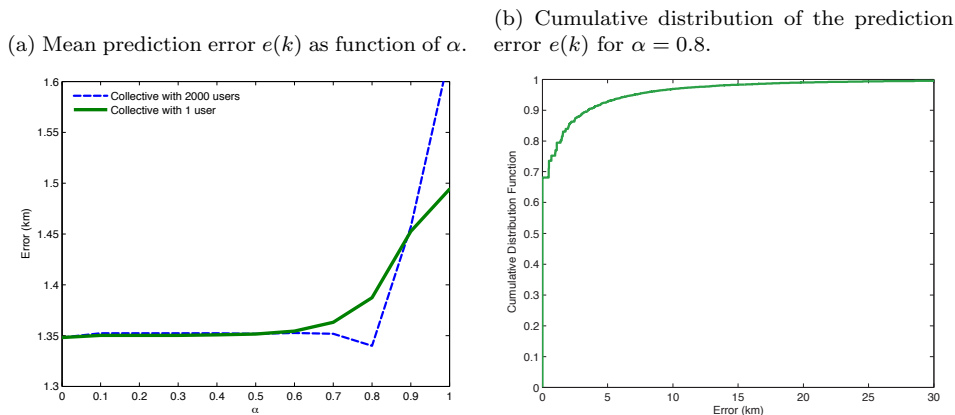
### 5.1 Prediction Error

To test the accuracy of the model, we compared its predictions of the next location for each of the 2,000 individuals in the data sub-sample against their next actual location. By evaluating the difference between the predicted location ( $x_{k+1}^P(u) = P(x_{k+1} = j | x_k = i)$ ) and the actual location ( $x_{k+1}(u)$ ) we can derive a measure of the error using the Euclidean distance between the two parcel centroids:

$$e(k) = |x_{k+1}(u) - x_{k+1}^P(u)|.$$

Recall too that the model contains a tuneable  $\alpha$  parameter, ranging from 0 (individual only) to 1 (group only), which balances the effect of the individual’s unique trajectory history against the aggregate habits of a larger population. By repeatedly re-running the model with different values of  $\alpha$ , we obtained the results shown in Figure 6. To confirm that it is the group preference or habit that is improving the error rate, Figure 6a also shows the results of predictions made when the ‘collectivity’ is just one user, the same one whose location we are attempting to predict: the results show an increased error due to the absence of a global influence from other users.

Figure 6: Prediction Errors

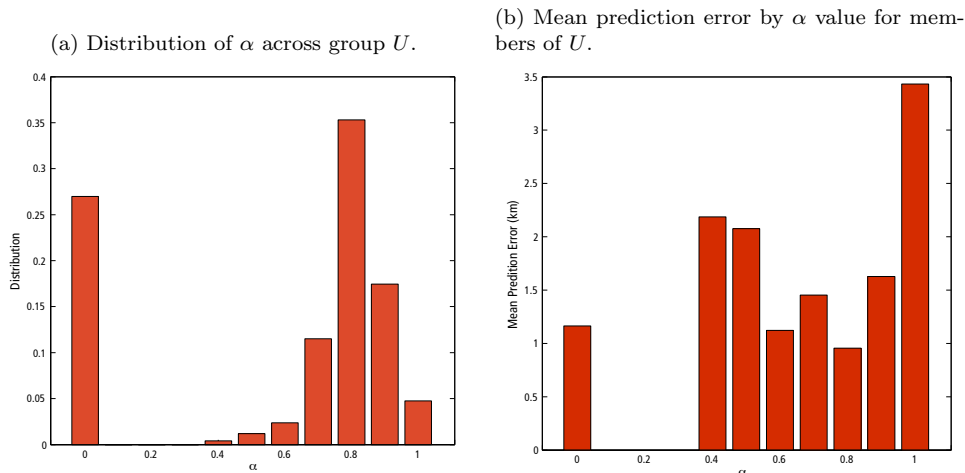


This approach also substantially outperforms a naïve null model in which we predicted that people could always be found in their ‘most visited’ cells: in this latter case the mean error was  $2.8km$ , which also gives a sense of how much people move, on average. Figure 6b shows the cumulative error that results when we use this optimal value: in 60% of cases the predictive error is zero (i.e. we are able to correctly estimate the user’s next location), and in more than 90% of cases the error is less than  $5km$  over a total possible area of some  $27,000km^2$ .

Another way of looking at the model is to consider the distribution of  $\alpha$  across population  $U$  as a means of understanding to what extent the population is self-similar. Figure 7 groups users according to their optimal  $\alpha$ ; in Figure 7a we show the proportional membership in each decile, and in Figure 7b we show how this impacts the prediction error for each member of that decile. In Figure 7a two obvious features emerge: the largest single group of users has  $\alpha = 0.8$  (the same value as the

global optimum obtained above), but there is a secondary group containing 27% of the sample that has an optimal  $\alpha$  of 0. What this suggests is that *most* people – the three deciles clustered around 0.8 that account for more than 50% of users – are fairly predictable using group behavior, but that for a significant minority aggregate patterns are of no use at all.

Figure 7: Prediction Errors and  $\alpha$



The results outlined in Figure 7a might appear to suggest that the proposed model will fail to produce meaningful predictions, but the distribution of errors in Figure 7b makes it clear that this is not the case. At  $\alpha = 0$  this approach actually produces very low mean errors, suggesting that these individuals are still predictable *even if* they are not like the rest of the reference population. Similarly, the range  $0.6 \leq \alpha \leq 0.8$  also produces predictions with a low mean error and, when combined with the overall weight of this segment in the population, accounts for the global optimum of 0.8. Unsurprisingly, the largest prediction errors occur where  $\alpha$  tends towards 1.

## 5.2 Individual Prediction

To further test the model, and to help clarify how it could be employed in a real-world context, we will here present a (fictional) platform that predicts the likely destination of a ‘user’ at any given time of day. The platform could have many uses, but the basic concept is that it is possible to anticipate the user’s next location and to use this prediction to influence their behaviour. Some potential applications of this platform would include: trip planning assistance for tourists to help them navigate between attractions; route guidance for regular commuters that could help them to avoid areas that will *become* congested around the time that they will be passing through; and timely offers from merchants that could be made in advance of the user departing their current location.

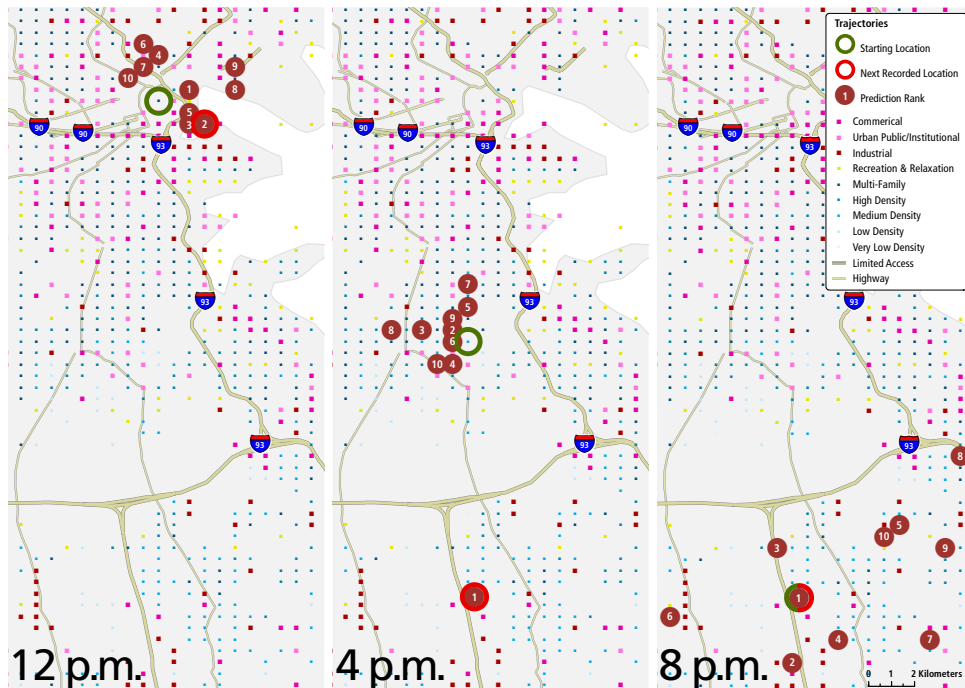
This approach addresses a fundamental weakness that can be seen in most commercial LBS proposals: the idea that a user passing by a shop will suddenly and radically change their current behaviour in order to take advantage of a highly-localised special offer. In this model, because predication is directly connected to land use and POIs, it is possible to use the platform to *infer* the most likely activity in which the



user *will* be engaged. This capacity, in turn, opens up the possibility of influencing user and consumer choices for the purpose of a public good – such as proposing an alternate route that will avoid adding to existing congestion – or a commercial gain – such as an inducement to try out a newly-opened shop, bar, or venue.

Figure 8 illustrates how the system calculates a set of probabilities for a large area and can then select and make an ordered set of predictions (e.g. top three, top ten, ...). In this example, comparisons between predicted and actual behaviour show that, in two cases the highest-rated location was, in fact, the next destination of the user, and that in the third case it was the second highest-rated location that was ‘selected’. This figure also makes clear the spatial and temporal inter-dependence of the predictions: at 12 p.m. the predictions are focussed on downtown Boston; at 4 p.m. they show a pattern intuitively consistent with evening commutes home; and at 8 p.m. there is a wider dispersal across a residential area that would be consistent with a range of more social activities (e.g. being at home, going out the movies, etc.).

Figure 8: Comparison of Predictions & Actual Behaviour of a Randomly-Selected User



This type of platform could clearly also have applications in tourism and leisure management where an obvious issue is that different types of visitors – businesspeople, frequent travellers, first-time arrivals, and relatively local households – have different interests, as well as different levels of familiarity with the area and differing degrees of mobility. By distinguishing between different classes of users, and by ‘learning’ from their collective and individual preferences, the system outlined here could gradually, and continuously, improve its predictions with minimal input from administrators.

For instance, new users of the platform outlined here could be provisionally assigned to a generic ‘tourist’ group with the model’s  $\alpha$  value set to 1 (group only) so that it is able to offer some level of predictive capacity from the very beginning of the visit. Over time, the  $\alpha$  parameter can be tuned downwards to reflect an improved

‘understanding’ of the individual’s preferences and habits. Alternatively, a single system deployed across many cities could leverage an understanding of the preferences of a user in one city to ‘predict’ their likely itinerary in another. The focus on land use and POIs provides a useful degree of abstraction within the model since the handle for behavioural analysis is similarity of activity, and not pure geography. So the platform could as easily draw on an individual’s visits to *other* cities to develop a set of predictions, as draw on what people like that individual who from the *same* city have done during similarly structure visits.

Figure 9: Prediction system performance

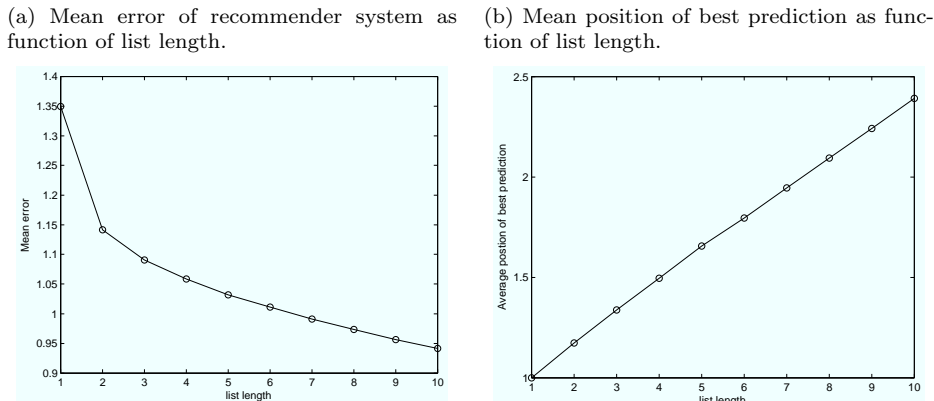


Figure 9 shows the relationship between the ‘prediction list’ length and the prediction error. As the length of the list increases, the error obviously decreases (see Figure 9a), but with just three elements in the list the system has already achieved a mean error of just  $\approx 1.1km$ . In fact, as Figure 9b shows, the best prediction is almost always amongst the first three most probable places. Of course, achieving perfect prediction of where the user would have gone anyway is not necessarily helpful, so it’s worth noting that, in addition to the opportunities set out above, the platform could also be designed to allow the user to investigate what other users with similar behaviours are doing and so identify ‘missed opportunities’ for novel activities. In a sense, the prediction platform could even be extended to perform a ‘recommendation’ function whose purpose is to push the user to explore new locations which fall outside of their historical pattern of activity, but which nonetheless still fall within the set of places and times that people in their dominant ‘group’ like to visit.

## 6 Discussion & Limitations

The results presented above enable us to shed light on a variety of challenges facing the fields of transportation research and urban planning, but the novelty of the data source and techniques also raise some important questions that require further consideration. The most basic question concerns the validity of the data itself: can mobile phone signalling data substitute for existing methods of data collection and urban simulation? The simple answer to this is obvious: it cannot, because mobile phone usage does not enable us to directly model the decision-making processes involved in trip-generation. But this answer is overly simplistic, and our results demonstrate

that it is possible to use this type of locational data to predict the *outcomes* of those decisions at an incredibly fine-grained level. Moreover, the layering in of land use and POI data not only enables us to improve individual-level predictions, it obviously enables us to draw some inferences about the types of activity in which the user is most likely to be engaged.

In this sense then, the approach presented in this article marks a step-wise change in the ways in which we can approach and validate existing models rooted in methods such as activity scheduling and agent-based modelling (ABM). And we should note too that this type of data-driven approach side-steps the collection and simulation limits embedded in both more qualitative and more abstract (i.e. computationally-driven) research methods: although we have only here reported results for 2,000 simultaneous users, the system is inherently scalable to tens, or even hundreds, of thousands of individuals. This degree of scalability compares well with the typical population sizes used in more traditional approaches to the field where samples can range from a few hundred (cf. Schlich and Axhausen, 2003) down to the teens. Since users tend to naturally keep their phone on and charged, the use of the mobile phone as a research tool also provides spatially- and temporally-extensive coverage that addresses some of the limitations of higher-resolution devices such as handheld GPS units.

The single most important limitation of our approach has to do with the variability in the sampling rate for each user. We should note that less active users – such as the elderly, some minority groups, and ‘late adopters’ (Ahas et al., 2010c, p.24) – on older phones will tend to be under-represented in our data set and to have ‘inter-event’ periods of more than an hour. Not only does this mean that locational data for such users will be limited, but also that the data we do receive is biased (Ahas et al., 2010c, p.23). The issue here is the well-known fact that people often place more calls when in a transition between two stable places than they do when not moving, suggesting that this population is more likely be localised to travel infrastructure than to real origins and destinations. And while active positioning methods address some of these sampling issues (Ahas et al., 2010b, p.46), the costs associated with collecting this type of data are substantial and the results are not necessarily any more accurate than those produced by AirSage’s passive Wireless Signal Extraction system. This is why the rise of smartphones is so important to this type of research: because such phones tend to be in near-constant contact with the network (to check for new email, to browse the web, to download updates, etc.) they generate detailed traces that can be used to construct path histories with high fidelity across across long periods of time.

A final question worth addressing in more detail here is the degree to which group behaviour is relevant to individual prediction at a time when our ability to coordinate and reschedule activities on-the-fly would seem to be making individual activity sequencing increasingly idiosyncratic. The conclusion that has emerged in this work, and which builds on the findings of Song et al. (2010b) and Ahas et al. (2010c), is that behaviour is very definitely not becoming less predictable and that *many* individuals follow paths that are largely predictable from aggregate models. Where a significant minority does appear to have little in common with the population as a whole, this work is currently not at the stage where we could reasonably determine if this is simple a case of not having selected an appropriate reference population (which seems likely as the users were selected at random). The selection process is thus one area

that will require particular attention going forward.

However, the existence of a globally optimal  $\alpha$  at 0.8 and the overall distribution of  $\alpha$  and its prediction error shown in Figure 7 for the population of 2,000 users highlight the fact that group behavior is a very useful starting point for prediction in the absence of individual history. Moreover, even as individual history becomes available it proves essential to prediction in less than 25% of the sample population. The tuneable  $\alpha$  parameter also enables us to avoid an issue that would arise using a single weighted mixing of individual and group behavior: that large differences between the user and the group would produce correspondingly large prediction errors. Finally, the data-mining approach has demonstrated that prediction must be informed by the ‘activity space’ – defined here in terms of land use and POIs– available to the user upon arrival at a given location.

## 7 Conclusions

In recently-published work Ahas et al. (2010b, p.45) used calling activity to draw inferences about the functional diversity and quality of space. Here we have, in a sense, done the opposite by using the range of activities available at a given point in space to help us infer who will travel to it and when. By blending individual and collective behavior with ‘points of interest’ and the range of possible activities defined by land use, we are able to calculate with a high degree of accuracy ( $\leq 1.34km$  globally) the probability that a person will move to given location in the next measurement interval. Using a mobile phone location dataset, we tested the model for 2,000 users living in the Boston metropolitan area and obtained an optimal group weighting ( $\alpha$ ) of 0.8, but also found that, while this value worked well for the majority of users, for a minority of individuals a group weighting of 0 provided much better results overall.

Although we have elaborated this model in the context of a prediction service – where will the user go next? what can we do to influence their choice of route and of the activities in which they engage once they reach their destination? – it is also clear that this type of data and analysis has value to more traditional planning domains. Combining land use and points of interest with long-term trajectory histories for individuals and groups also enables planners to explore the temporal sequencing of events or activities, as well as the changing spatial relationships between them; this field has been explored on a conceptual level in terms of ‘mobility environments’ (cf. Bertolini and Dijst, 2003; Bertolini, 2005) and the ‘new mobilities paradigm’ (cf. Sheller and Urry, 2006), but the backing data that would support the detailed analysis required to begin understanding and designing such environments has been lacking. We feel that this model opens up a new avenue into this domain of research, and future work will concentrate on improving the model and integrating more sophisticated activity-based models.

## 8 Acknowledgments

The authors gratefully acknowledge AirSage for providing the mobile phone location dataset. Jon Reades wishes to also acknowledge the support of the EPSRC (Grant #EP/I018433/1) and of the European Commission (Complexity-NET/FP6

ERANET) for this research .

## 9 Appendix

Figure 10: Collective Land Use Interest Preferences by Time of Day and Average

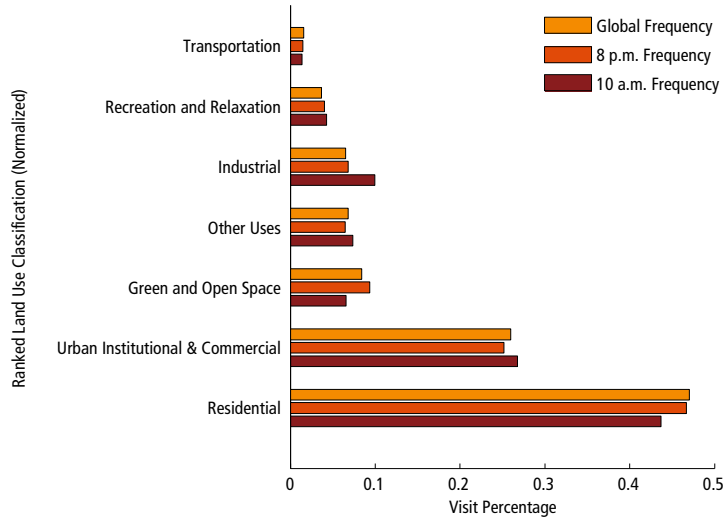
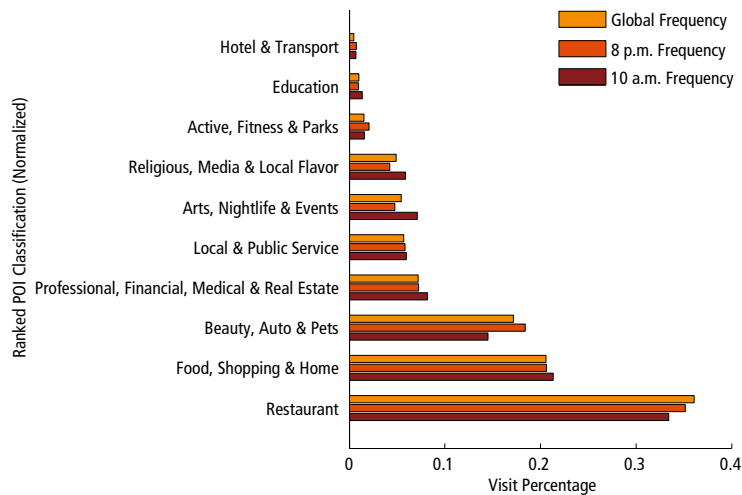


Figure 11: Collective POI Interest Preferences by Time of Day and Average



## References

- T. Adler and M. Ben-Akiva. A theoretical and empirical model of trip chaining behavior. *Transportation Research Part B: Methodological*, 13(3):243–257, 1979.
- R. Ahas, A. Aasa, U. Mark, T. Pae, and A. Kull. Seasonal tourism spaces in Estonia: Case study with mobile positioning data. *Tourism Management*, 28(3):898–910, 2007. ISSN 0261-5177.

- R. Ahas, A. Aasa, S. Silm, and M. Tiru. Daily rhythms of suburban commuters' movements in the Tallinn metropolitan area: Case study with mobile positioning data. *Transportation Research Part C*, 18(1):45–54, 2010a.
- R. Ahas, A. Aasa, S. Silm, and M. Tiru. Daily rhythms of suburban commuters' movements in the tallinn metropolitan area: Case study with mobile positioning data. *Transportation Research Part C: Emerging Technologies*, 18(1):45–54, 2010b.
- R. Ahas, S. Silm, O. Järv, E. Saluveer, and M. Tiru. Using Mobile Positioning Data to Model Locations Meaningful to Users of Mobile Phones. *Journal of Urban Technology*, 17(1):3–27, 2010c.
- K.W. Axhausen and T. Gärling. Activity-based approaches to travel analysis: conceptual frameworks, models, and research problems. *Transport Reviews*, 12(4):323–341, 1992. ISSN 0144-1647.
- A. L. Barabási. *Bursts: The Hidden Pattern Behind Everything We Do*. Dutton, 2010.
- L. Bertolini. Cities and Transport. In L. Albrechts and S. J. Mandelbaum, editors, *The network society: a new context for planning?*, The networked cities series, pages 146–160. Routledge, 2005.
- L. Bertolini and M. Dijst. Mobility Environments and Network Cities. *Journal of Urban Design*, 8(1):27–44, 2003.
- F. Calabrese and C. Ratti. Real Time Rome. *Networks and Communications Studies*, 20(3–4):247–258, 2006.
- F. Calabrese, J. Reades, and C. Ratti. Eigenplaces: segmenting space through digital signatures. *IEEE Pervasive Computing*, 9(1):78–84, 2010a.
- 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*, 12(1):141–151, 2011.
- Francesco Calabrese, Francisco C. Pereira, Giusy Di Lorenzo, Liu Liang, and Carlo Ratti. The Geography of Taste: Analyzing Cell-Phone Mobility and Social Events. In *Pervasive*, pages 22–37, 2010b.
- N. Eagle and A. Pentland. Reality mining: sensing complex social systems. *Personal and Ubiquitous Computing*, 10(4):255–268, 2006.
- N. Eagle and A. Pentland. Eigenbehaviors: Identifying Structure in Routine. *Behavioral Ecology and Sociobiology*, 63(7):1057–1066, 2009.
- M. Gendreau and P. Marcotte. *Transportation and Network Analysis: Current Trends*. Springer, 2002.
- Y. Gong and R. Mackett. Visualizing Children's Walking Behaviour Using Portable Global Positioning (GPS) Units and Activity Monitors. In Lin H. and M. Batty, editors, *Virtual Geographic Environments*, pages 295–310. Science Press, 2009.
- M. C. González, C. A. Hidalgo, and A. L. Barabási. Understanding individual human mobility patterns. *Nature*, 453(7196):779–782, 2008.

- T. Hägerstrand. What about people in regional science? *Papers in Regional Science*, 24(1):6–21, 1970.
- J. Krumm. Real Time Destination Prediction Based on Efficient Routes. In *Society of Automotive Engineers (SAE) 2006 World Congress*, 2006.
- J. Krumm and E. Horvitz. Predestination: Inferring Destinations from Partial Trajectories. In *UbiComp*, pages 243–260, 2006.
- D. Lazer, A. Pentland, L. Adamic, S. Aral, A. L. Barabási, D. Brewer, N. Christakis, N. Contractor, J. Fowler, M. Gutmann, T. Jebara, G. King, M. Macy, D. Roy, and Marshall Van Alstyne. Computational Social Science. *Science*, 323:3, 2009.
- Q. Li, Y. Zheng, X. Xie, Y. Chen, W. Liu, and W.-Y. Ma. Mining user similarity based on location history. In *GIS '08: Proceedings of the 16th ACM SIGSPATIAL international conference on Advances in geographic information systems*, pages 1–10, New York, NY, USA, 2008. ACM.
- L. Liao, D. J. Patterson, D. Fox, and H. Kautz. Learning and inferring transportation routines. *Artificial Intelligence*, 171(5-6):311–331, 2007.
- J. P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, and A. L. Barabási. Structure and tie strengths in mobile communication networks. *Proceedings of the National Academy of Sciences*, 104(18):7332, 2007.
- J. Reades, F. Calabrese, A. Sevtsuk, and C. Ratti. Cellular census: Explorations in urban data collection. *IEEE Pervasive Computing*, 6(3):30–38, July–September 2007.
- R. Schlich and K.W. Axhausen. Habitual travel behaviour: Evidence from a six-week travel diary. *Transportation*, 30(1):13–36, 2003.
- Sense Networks. Citysense, January 2008. URL <http://www.sensenetworks.com/citysense.php>.
- M. Sheller and J. Urry. The new mobilities paradigm. *Environment and Planning A*, 38:207–226, 2006.
- C. Song, T. Koren, P. Wang, and A. L. Barabási. Modelling the scaling properties of human mobility. *Nature Physics*, 2010a. ISSN 1745-2473.
- C. Song, Z. Qu, N. Blumm, and A. L. Barabási. Limits of Predictability in Human Mobility. *Science*, 327(5968):1018–1021, 2010b.
- J. White and I. Wells. Extracting origin destination information from mobile phone data. In *Road Transportation and Control*, 2002.
- D.R. Williams and B.P. Kaltenborn. Leisure places and modernity: The use and meaning of recreational cottages in Norway and the USA. *Leisure practices and geographic knowledge*, pages 214–230, 1999.
- Y. Zheng, Q. Li, Y. Chen, X. Xie, and W.-Y. Ma. Understanding mobility based on GPS data. In *UbiComp '08: Proceedings of the 10th International Conference on Ubiquitous Computing*, pages 312–321, New York, NY, USA, 2008. ACM.