ANDRES SEVTSUK AND CARLO RATTI,
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

MOBILE SURVEYS

In recent years, a new approach for estimating people's movement in cities has emerged through mobile phone positioning. As opposed to the more traditional methods of traffic surveys, automated counts, or individual counters on streets, the use of aggregated and anonymous cellular network log files has shown promise for large-scale surveys with notably smaller efforts and costs (Reades, Calabrese, Sevtsuk & Ratti, 2007). In addition, a frequent data feed from the cellular network has also been argued to demonstrate fine grain over-time variation in urban movements, which are lacking from the traditional prediction methods (Ratti, Pulselli, Williams & Frenchman, 2006). However, despite the positivist approach to the new methodology, additional evidence is needed to show how cellular network signals correlate with the actual presence of vehicles and pedestrians in the city. The purpose of this paper is to address this shortcoming by presenting the results of a survey effectuated in Rome, Italy in January 2007. Using the results of the two-day experiment, we will employ statistical models to investigate the relationship between empirical pedestrian and traffic counts on the streets of Rome with the simultaneous Telecom Italia Mobile (TIM) network signal and traffic prediction. Secondly, we will explore whether the mobile network data demonstrates the significant time-dependent variation that is missing from traditional fixed predictors like Space Syntax choice and integration analysis and could thus describe cities dynamically over time. Finally, we will also outline some general issues of accuracy in using aggregate mobile network data for estimating people's movement in cities.

INTRODUCTION

As more than half of the world's population now inhabits cities, and more cities are being designed and constructed than ever before, the need for analyzing and understanding how cities function in daily life increases. Most of the knowledge available to us on this topic is concerned with the social demographics and the constructed elements of cities – streets, buildings, public spaces, parks etc. Much less is known about how these elements are actually used and what the temporary patterns of people's movements between the fixed points look like. As people's movements and actions are hard to predict and difficult to track, it has been traditionally challenging to study the usage dynamics of large cities.

In recent decades, urban mobility modeling has witnessed a rise due to advances in computational capacity and the development of virtual simulation environments. Several theories have been developed to explain people's mobility. Hillier et al. have interestingly shown that people's movements in cities are remarkably tied to the geometric configuration of the street network (Hillier, Hanson & Peponis, 1987). Their quantitative technique of measuring topological graphs for predicting mobility flows, called Space Syntax, has gained particular popularity amongst architects and urban designers over the past decades. Others have argued that streets are more than topological networks and their usage is largely dependent on the morphology of the urban area (Anderson, 1986; Ellis, 1986; Mathema, 2000). Their approaches argue that streets are not only channels that allow people to flow between the nodes of an urban network, but also places of encounter and interaction; they are transit spaces as well as places. Both approaches suggest, however, that the usage of a street is strongly dependent on urban form: on the one hand on the arrangement of connections, on the other, its content and shape. Transportation researchers have studied the impact of amenities, land use and transit connections on transportation flows (Cervero, 1989; Handy, 1996; Rodriguez & Woo, 2002; Zegras, 2005) and developed quantitative models to predict pedestrian volumes of an area (Chung, 2003). Yet other scholars have claimed that physical neighborhood characteristics only have a minor impact on travel behaviour (Crane, 1996). More recent developments in discrete choice modeling (DCM) have started simulating pedestrian and vehicular movement dynamically in time (Antonini, Bierlaire, Webber, 2006; Ben-Akiva et al., 1997). Even though most researchers argue that urban form does affect pedestrian activity, and accept that mobility patterns can be at least partially predicted, there does not seem to exist an agreement on how and to what extent. Overall, the scientific explanations of mobility flows present multiple views and a lack of consensus and an empirical methodology for confirming the theories is much awaited.

Recent developments in portable communication have rendered mobile phones increasingly affordable and popular, which makes mobile network logs appealing for aggregate analysis of urban population flows. The work on the geographic analysis of mobile phone log files effectuated thus far has mainly centered on two aspects: the positioning of individual users for a variety of location based services (LBS), social networking and individual tracking purposes (Ohmori, Harata, Nakazato, 2005; Laineste, 2003) and probing the mobile data for vehicular traffic analysis (Kummala, 2002; Rose, 2006; cf. chapter 11). Relatively little work has been done to analyze the aggregate movements of people in the city through mobile phone data. This has been attempted in a few projects at MIT's SENSEable City Laboratory (Ratti, Pulselli, Williams & Frenchman, 2006; Ratti, Sevtsuk, Huang & Pailer, 2005; Calabreses, Reades & Ratti, 2007), and Tartu University in Estonia (Ahas, Aasa, Mark, Pae & Kull 2007) . This lack is probably caused by the difficulty to obtain the data and the uncertainty about how well mobile traffic data represents the actual mobility of people in cities. Yet, for urban planners, this is the most awaited kind of analysis, which could possible yield empirical evidence for explaining large scale dynamics of an urban population.

The mobile network data also has its reservations. Compared to the more detailed information available in origin and destination surveys, and DCM models, aggregated analysis of mobile network logs does not yet inform us where people's journeys start and end. Instead, the data only describes how many callers are where at any given time. However, this qualitative shortcoming is counterbalanced by a quantitative advantage: the network logs can be obtained from a city-wide telecom system that is already in place with relatively minor effort. Furthermore, the data can theoretically be sampled in real-time for a period of any length. These are the promising advantages of mobile network data. The purpose of this paper is to analyze how accurately and reliably the data describes the actual presence of people on city streets. Using Rome as a case study, we will be addressing two main questions. First, do the vehicle estimates (BSC data) and overall network activity measures (Erlang data) on Telecom Italia's network correlate with empirical observations of vehicles and pedestrians on specific streets in Rome? And secondly, does mobile phone data predict additional over time variation in mobility patterns that is missing from traditional fixed predictors?

MEASURES

Predictors

In Italy, there are now more registered mobile phones than people [1]. The data used in this paper comes from Telecom Italia (TIM), the largest service provider in the country. Besides TIM, there are three other large service providers in Rome: Omnitel Vodafone, Wind and Blue. TIM is currently market leader in the city, supplying about 40.3% of the share. This constitutes approximately one million users in Rome, less than half of the city's population. However, TIM's data used in this study does not describe the activity of all the registered users in Rome, but only those who were actively engaged in phone calls during the measurement periods. In particular, we will be using two distinct sources of TIM's network data: 1) traffic information from Erlang measurements of individual antennae and 2) aggregate vehicular traffic predictions from Base Station Controllers (BSC). We use Erlang in the models as a question predictor for pedestrian counts, and BSC_veh as a question predictor for vehicle counts.

Erlang measurements are commonly used for assessing aggregate mobile network traffic in particular cells. An Erlang measure is essentially a use multiplier per unit time. The use of one mobile phone for one hour in a particular cell constitutes one Erlang, whereas the use of two phones for half an hour each also constitutes one Erlang. Each network cell has a unique Erlang value at any given time, depending on the amount and length of calls processed by that cell. In this study, Erlang data from selected cells in Rome was measured at 15 minute intervals during two days as shown in the sample in **Illustration 8.1**.

Illustration 8.1 - Example of Erlang values

	Erlang Values				
Cell ID	**RM25**	**RY36**	**RK38**	**RJ98**	**RM01**
12/01 08:00..08:15	16277	1219	16277	7056	3730
12/01 08:15..08:30	22543	1170	22543	9393	6350
12/01 08:30..08:45	29100	1302	29100	12571	8722
12/01 08:45..09:00	38321	1636	38321	12338	8320
12/01 09:00..09:15	53393	2224	53393	16715	8689
12/01 09:15..09:30	62582	3124	62582	23185	11347
12/01 09:30..09:45	64678	2843	64678	21049	10999
12/01 09:45..10:00	72001	3316	72001	25904	13380
12/01 10:00..10:15	69800	2905	69800	24433	12184
12/01 10:15..10:30	76793	4313	76793	30870	12107

For the purpose of this analysis, we transformed the Erlang values in TIM's dataset by dividing the raw measures by the area of the given cell and the amount of antennae in that cell. This resulted in an Erlang measure that shows how much call volume is processed in each cell per equal unit area and a single antenna.

In addition, estimates for the number of calls originating from vehicles were obtained from two base-station controllers (BSC) in the north-eastern part of the city center. Each BSC monitored the activity in a set of individual cells and aggregated the information into a continuous rectangular grid of 250x250 meters, covering the area of multiple cells. Thus each BSC measurement estimated how many calls within each 250x250 meter "pixel" in a 15 minute time period had occurred, and filtered out those that had a mean speed above 8km/h during the call, categorizing the latter as the number of vehicles. Unlike Erlang, which was generated by all calls that were processed by a given cell, BSC measures were obtained by anonymously triangulating all clients who engaged in calls, determining which pixel they were located in. The callers' position was obtained by using a combination of methods including cell Id, angle of arrival, timing advance, and signal strength triangulation. The raw information was presented in matrix form, where each value corresponded to the estimated number of calls from vehicles in a certain pixel, as shown in **Illustration 8.2**.

Illustration 8.2 - BSC estimates for calls originating from vehicles

HOUR 8; MINUTES 15
YEAR 2007; MONTH 1; DAY 12
CITY Roma
NROWS 48; NCOLS 56
ULXMAP 785566.000000; ULYMAP 4650585.000000
XDIM 250.0; YDIM 250.0

0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	1	0	0	0	0	0	0	0	0	1	0	1	1	2	1	0	0	0	0	0	0	3	0	0	0	0	0	1	1	1	2
0	0	0	0	0	0	0	0	1	1	2	2	3	3	4	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	2	0	0	0	0	0	0	0	1	0	1	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

In addition to the two question predictors from the mobile network, a few additional control predictors were used. *Major_arterial* was a dichotomous predictor indicating whether a street was classified as a major arterial road in the Navteq street database or not. Two commonly used space syntax measures were also included: integration with a radius of 400m (int_r400) and choice with a radius of 3200 m (choice _r3200) (a more detailed explanation of these variables can be found in Hillier (1984 (1989)). The value for the integration analysis with the radius of 400 meters (int_r400) is commonly used in space syntax as a predictor of average pedestrian flow on a particular street. The choice analysis with a radius of 3200 meters is commonly used for vehicular traffic predictions. Since the choice values are computed with an exponential distribution, we followed a common practice of transforming the choice_r3200 values by a log2. Finally, *Weekend* was a dichotomous predictor indicating if the count was measured on a Friday or Saturday. **Illustration 8.3** presents the descriptive statistics of the predictors and outcomes used in the models.

Illustration 8.3 Descriptive statistics of the outcomes and predictors

	Mean	Std. Deviation	Range	Min.	Max.	N
Pedestrians	198.28	216.63	1046.00	10.00	1056.00	396
Vehicles	141.97	128.26	634.00	2.00	636.00	374
BSC_veh	0.78	0.83	3.50	0.00	3.50	396
Major_arterial	0.33	0.47	1.00	0.00	1.00	396
Raw cell_Erlang	35017.26	29893.93	92737.00	116.00	92853.00	396
Cell_Erlang_norm (question)	0.05	0.04	0.12	0.00	0.13	396
Int_400	33.42	19.37	74.61	-1.00	73.61	374
Log2_ch_r3200	15.51	3.72	14.05	4.17	18.22	374
Weekend	**0.50**	**0.50**	**1.00**	**0**	**1**	**374**

Outcomes

Based on the availability and geographic distribution of Erlang and BSC data, 9 locations in the neighborhood around the Termini train station in Rome were selected for comparative pedestrian and traffic counts. This area was attractive for the study because it was well covered in TIM's dataset and it contained a uniformly dense urban fabric with ostensibly many pedestrians and cars. The counts were effectuated on January 12th and 13th, a Friday and a Saturday. On both days, six counters counted vehicles and pedestrians from 8am to 8pm on 18 street segments using tally-sheets. Pairs of two counters circulated between three particular street intersections in a continuous loop, counting the two intersecting streets at each intersection for 15 minutes at a time, returning to the same place once every hour. The intersections and street segments were chosen after a personal site visit, so that the seemingly largest intersection in each cell of interest gave a representative estimate for the pedestrian and vehicular traffic in that cell. The counting period was chosen to match the 15-minute Erlang and BSC measurement periods. **Illustrations 8.4 and 8.5** below illustrate the chosen streets that were counted and the corresponding Erlang cells and BSC pixels.

ANALYSIS

Naturally the 9 intersections that were chosen varied in exact size. As the mobile phone data estimates referred to the whole coverage area of network cells or BSC pixels, and not individual streets, then the representational "weight" of each counted street was different. Furthermore, all streets where counts occurred had a slightly different importance relative to the total set of

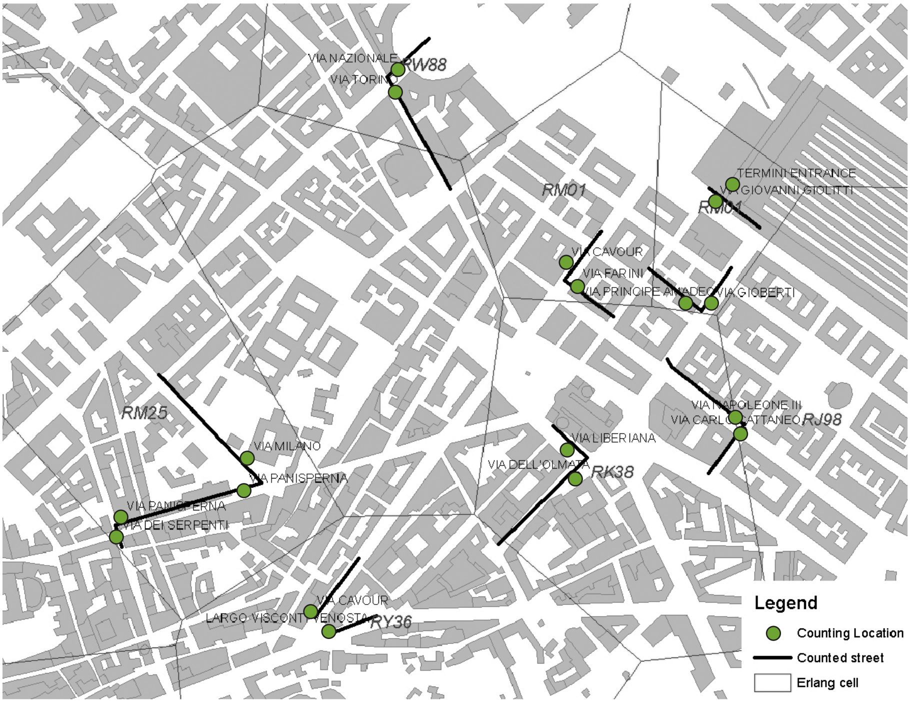

Illustration 8.4
Map of counted streets and Erlang cells.

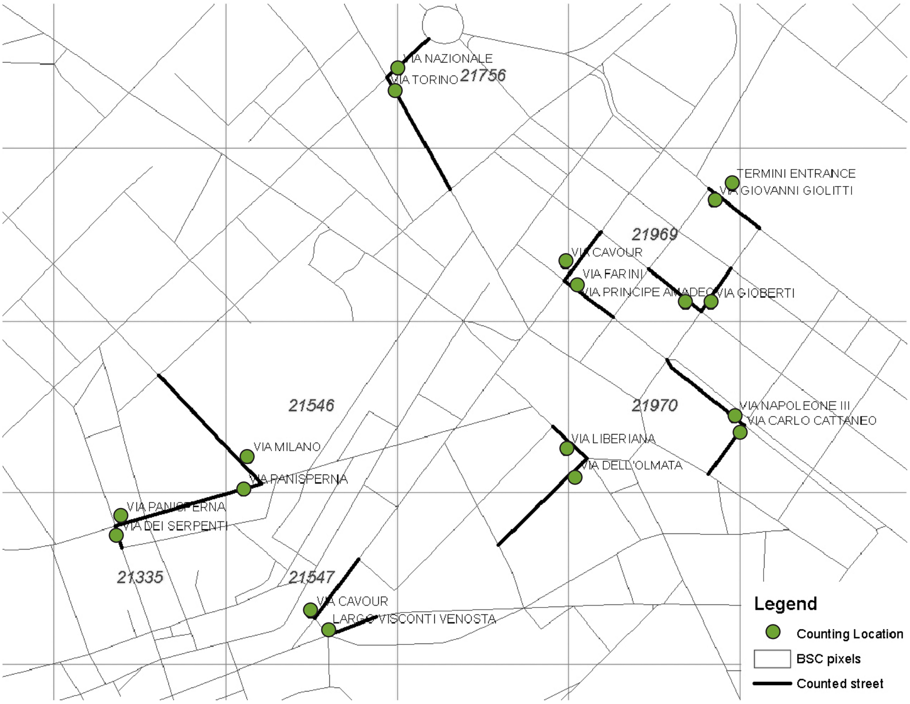

Illustration 8.5
Map of counted streets and BSC pixels.

streets within each cell or pixel. Some streets were clearly the busiest ones out of all the streets within a given cell, whereas others were less hierarchal compared to the rest of the streets in the pixel. Therefore some count locations naturally captured a larger proportion of traffic in their area than others. These location differences and the repeated nature of the counted data (every location was counted 11 times a day, and 22 times during both days) demanded that the similarity of variance at the same locations during different counting times be taken into account. For specifying a statistical model this suggested that a multilevel structure of repeated measurements be used. A multilevel structure was therefore specified in the SAS software MIXED procedure using a random effects model. Using this model, we will first proceed with the analysis of the vehicle prediction and subsequently turn to pedestrian counts.

TIM's Vehicle Traffic Prediction

First, the unconditional model M1 was specified as a base line against which we could compare the gains of adding predictors from the mobile network. The unconditional model allowed the intercepts to have random effects at different counting locations. The estimated parameters of this model are shown in **Illustration 8.6**. The fixed effect of the intercept across all locations showed that the average count during the first observation period (8.00-8.15am) was 141.97 vehicles. The random effects part of the models show how the intercepts and residuals vary across the counting locations. We see that intercepts did vary significantly between locations ($p < 0.01$), which indicates that a large portion of the variation (81%) can systematically be attributed to differences amongst counting locations. The remaining 19% of all variation was attributable to the residuals of individual streets, which suggests that the differences between streets were a much greater cause of variation than the hourly fluctuations of vehicles on particular streets.

After specifying the unconditional model, we added only BSC_veh as a predictor to model M2 in order to test if BSC_veh alone could explain variations in the counted traffic data and thus be singularly used as a reliable estimate for vehicle traffic on particular streets. The results showed that BSC_veh did have a significant fixed linear effect on the vehicle counts: a one point change in the BSC_veh estimate was associated with a 20.01 point change in vehicle counts ($p < 0.001$).

Looking at the random effects, we see that the improvement in prediction between different locations attributable to TIM's vehicle prediction in M2 can be summarized by the proportional decline in the between-street residual variance. This shows that Telecom Italia's vehicle prediction could alone explain 31% of the variation in our traffic counts. This correlation was quite good, considering that most (81%) of the variation in the counts came from differences

Illustration 8.6
Taxonomy of fitted multilevel models predicting vehicle counts.

Taxonomy of fitted multilevel models describing the relationship between the empirical vehicle counts and TIM's vehicle traffic estimates, cntrolling for road-classes and Space Syntax choice estimates (n streets= 18, n counts= 342)

Predictor	Model M1 (uncond.)	M2	M3	M4
Fixed Effects:				
Intercept	141.97*** (28.80)	127.22*** (24.16)	7.84 (92.48)	56.35 (75.84)
BSC_veh		20.01* (8.06)		-36.74 (29.63)
Road class			150.78** (46.90)	136.58** (38.26)
Log2_ch_r3200			6.64 (6.03)	3.71 (4.93)
Weekend			-43.96*** (5.46)	-48.52*** (7.75)
BSC_veh * Road class				18.89 (17.47)
BSC_veh * Log2_ch_r3200				2.02 (1.98)
BSC_veh*Weekend				12.39 (9.72)
Random Effects:				
σ_{11} (intercept)	**13950****	**9649.54****	**7661.08****	**4948.68****
σ_{22} (BSC_veh slope)		**746.90***		**682.04***
σ_ε^2 (residual)	**3286.03*****	**2764.02*****	**2787.65*****	**2415.83*****
Fit Statistics:				
-2LL	**4157.6**	**4098**	**4066.2**	**4000**
AIC	**4161.6**	**4106**	**4070.2**	**4008**

Significance level ~p<0.10, *p<0.05, **p<0.01, *** p<0.001
Cell etries are coefficients and standard errors

between streets. The within-street residual variance declined by 16% with the addition of BSC_veh, which means that Telecom's traffic estimate also explained 16% of the traffic over time variation on the same streets. In addition we see that the random effect of BSC data ($\sigma 22$ BSC_veh slope) was also significant at a 95% confidence level, suggesting that the relationship between TIM's traffic estimate and the actual counts varied from street to street. We thus conclude that a significant relationship of BSC_veh and empirical counts is found, and 31% of the between-street variations as well as 16% of the within-street variation in vehicular traffic are explained by TIM's vehicle prediction. These findings are encouraging given the very small mean and range of BSC_veh values. Furthermore, we know that BSC_veh is an estimate for an urban area of 250x250 meters, which includes many different streets, whereas the counting data was collected from only specific streets within each square. Based on these results we can hypothesize that using the BSC_veh values for estimating traffic at a larger scale than singular streets (for instance, block level, neighborhood level or even district level) might result in even better predictions.

Illustration 8.7
Taxonomy of fitted pedestrian count models.

Taxonomy of fitted multilevel models describing the relationship between the empirical pedestrian counts and TIM's Erlang estimates, cntrolling for road-classes and Space Syntax integration values (n streets= 18, n counts= 342)

Predictor	Model M1b (uncond.)	M2b	M3b	M4b
Fixed Effects:				
Intercept	164.98*** (37.61)	143.15** (37.79)	73.16 (76.66)	25.09 (77.80)
Erlang_norm		493.43** (177.10)		926.68* (441.92)
Road class			43.62 (80.86)	92.83 (81.00)
Integration_R400			2.25 (1.99)	1.25 (2.00)
Weekend			2.68 (7.71)	3.78 (12.5)
Road class * Erlang_norm				328.28 (499.08)
Integration_R400 * Erlang_norm				3.31 (9.33)
Weekend*Erlang_norm				985.51** (299.85)
Random Effects:				
σ_{11} (intercept)	**23791****	**22999****	**23925****	**22988*****
σ_{22} (Erlang slope)		**0**		**0**
σ_{ε}^2 (residual)	**5544.84*****	**5441.53*****	**5558.52*****	**5201.73*****
Fit Statistic:				
-2LL	**4352.9**	**4332.8**	**4331.1**	**4257**
AIC	**4356.9**	**4338.8**	**4335.1**	**4263**

~p<0.10, *p<0.05, **p<0.01, *** p<0.001
Cell etries are coefficients and standard errors

We next explored whether the addition of TIM's traffic estimate to a different conditional model, which already contained three static predictors: road-class, radius 3200m choice estimate from Space Syntax analysis, and a dummy variable for weekend, could explain any dynamic variation in the outcome. Since both road classifications and space syntax estimates provide a fixed average value for each street, regardless of the fluctuations in traffic that occur on the street in an hourly, daily or weekly scale, we expected the BSC_veh estimates, collected at 15min intervals, to improve the model significantly

The results showed that the addition of the BSC data to the conditioned model that already included road_classes, space syntax choice values and *weekend*, decreased the within-street residuals by 13%. Thus we find that there was indeed a component of over time variations in the traffic counts explained by TIM's traffic predictions, which was not be predicted through the fixed predictors. Again, the results are encouraging even though the majority of over time residual variations on individual streets still remained unexplained with TIM's predictor in the model and we conclude that the BSC_veh estimates do show promise for illustrating how vehicular traffic flux oscillate at particular streets during different hours of the day.

ERLANG AND EMPIRICAL PEDESTRIAN COUNTS.

We will now turn to the analysis of how the antenna level Erlang measures predicted the pedestrian flux on particular streets. Since the analysis is very similar to the vehicular traffic analysis above, we can directly turn to the findings.

The unconditional model for predicting pedestrian flux through TIM's Erlang measurements was specified similarly to the unconditional vehicle model. The results of the pedestrian models are shown in **Illustration 8.7**.

The intercept of the unconditional model, used as a baseline, showed that during the first counting period (8.00 – 8.15am) there were on average 164.98 pedestrians on a street. The random effect of the intercept shows that the counts at different locations did vary significantly ($p < 0.01$).

We next added the Erlang predictor to the model and tested whether Erlang alone would reveal a significant relationship with the pedestrian counts. The resulting estimated Erlang fixed effect captured the initial relationship between the normalized Erlang and pedestrian counts across all locations. This effect was significant and showed that between 8.00 and 8.15AM, locations that differed in one point with respect to normalized Erlang, differed by 493.43 pedestrians ($p < 0.001$). However, the random effect of Erlang ($\sigma 22$) was zero, which implies that there is no evidence to suggest in our sample that the relationship between Erlang values and empirical pedestrian counts varied between locations. Rather, we only see the fixed effect of Erlang suggesting that this relationship is uniform across all counted streets.

Looking at the residual changes, we find that the between-street residual variance decreased by only 4% with the introduction of Erlang to the model and the within-street variation decreased by 2%. This led us to conclude that Erlang did not improve the between-streets or within-street predictions of pedestrians much. Given that Erlang measures were distributed across all streets in the whole cell to explore the activity on specific single streets, it was not surprising to find very little correlation between the two. Furthermore, Erlang measures were generated by all calls from within the cell, not only from pedestrians on the streets, but also calls from parks, buildings, vehicles and so on. We concluded that in the given sample, the Erlang measure alone predicted little pedestrian activity on specific streets.

Lastly, we also explored whether the addition of Erlang data could explain the dynamic variations of pedestrians at specific streets, which a model with fixed predictors could not account for. The fixed predictors that were added to the unconditional model were Road Class and the radius 400m integration value from space syntax. A dummy variable for the weekend was also included. The estimated coefficients of this model (M3b) are shown in **Illustration 8.7**.

While the fixed effects of the above mentioned control predictors remained insignificant, Erlang in M4b did have a significant relationship to the outcome. The Erlang coefficient suggested that a one point change in Erlang values was on average associated with a 926.68 ($p < 0.05$) point change in pedestrian counts, controlling for weekends, road classes and choice values. The decrease in residual variance suggested that the addition of Erlang to a controlled model did improve the pedestrian prediction, but not much: Erlang accounted for 4% of the between-street residual variance and 6% for the within-street residual variance. The relationship between Erlang values and empirical pedestrian counts thus led to the conclusion that Erlang measurements were only marginally successful in predicting pedestrian flux on the chosen streets.

This makes intuitive sense. An important complication in relating Erlang values to actual population distribution is the changing nature of callers' behaviour. When the total call volume increases, for instance, then the sample size that generates an Erlang value also increases and we should expect a more accurate prediction from Erlang during the hours of large call volume. The same concern applies to the changes in average call length. If average calls are systematically longer during certain hours of a day, then Erlang values would increase, but the actual population distribution could remain the same. In addition, the average call length could also vary by area- people in residential areas might make longer calls than people in a noisy shopping area during the same hour. Unfortunately the data on call lengths and user behaviour was not available to us in this study and remains to be tested in future research.
As a preliminary test for these variations, the students in Rome also counted the number of passing pedestrians who were using cell phones. The analysis of these data showed that there was a significant positive correlation between Erlang and the number of pedestrians observed talking on the phone during the counts ($R2 = 6\%$, $t=5.01$; $p < 0.0001$) and that the proportion of pedestrians using mobile phones did significantly vary during different hours of the day as well as across different locations ($p < 0.05$). This suggests that when a larger proportion of pedestrians use their phones on the street, Erlang values are significantly larger, whereas the amount of pedestrians can remain similar. In future work using Erlang values for estimating callers' distribution, it would thus be important to account for how the total number of calls and the average call length vary throughout a typical day and to consider these variations when interpreting the relationship between Erlang and the caller's distribution.

CONCLUSION

Does the usage of a mobile phone network at a particular time reflect the distribution of people in a city? And could the data from the network dynamically predict the amount of vehicles and pedestrians on particular streets?

Indeed, to at least some extent, suggest the findings of our study, which compared two types of mobile network indicators with empirical counts of vehicles and pedestrians on the streets of Rome. First, we used the BSC measures which positioned each phone call geographically and measured the caller's speed of movement during the call, as a predictor for vehicular traffic. Secondly, we used the Erlang measures, which indicate the total usage intensity of each network antenna, as a predictor for pedestrian traffic.

Our analysis showed that Telecom Italia's estimate for mobile phone calls from vehicles was successful in predicting approximately a third of the traffic patterns at specific streets. This finding corroborates the idea of using vehicle estimates from a mobile-phone network in future transportation analysis. Telecom Italia has in fact already started using these measures to estimate the traffic conditions on the highways around Rome. For urban planners, this data now opens up an opportunity to investigate the driving dynamics in different neighborhoods with much less effort than offered by the traditional origin destination surveys and traffic counts. The data could also be used for assessing the role of urban design, land use and public amenities in travel mode choice in different neighborhoods (cf. chapters 7, 10, 11 and 13).

Erlang values were marginally, though significantly, successful in predicting pedestrian flux at particular streets. Only 2-4% of the pedestrian flow was forecast by the data. This finding was discouraging, but intuitively reasonable. Erlang measures at antennae provide an overall indication of call volumes processed by the particular cells and give no information on whether the callers were still or moving, indoors or outdoors, and are therefore less likely to correlate with a specific subset of callers, such as pedestrians used in this study. Since the measures are taken at the antenna level, the data also provides less accuracy and certainty for predicting the true location of callers. However, the relative ease of access and ubiquity of the Erlang in any mobile phone network still make the data very attractive for further study. Using a larger study area and accounting for the typical calling behaviour could potentially yield more accurate predictions of the actual population distribution at a particular time. Another interesting alley of research already emerging, is to cluster Erlang patterns at cells that are used similarly over time into sensible groups, in order to functionally describe urban areas that behave similarly over time.

Mobile networks thus remain interesting and valuable resources of information for urban analysis. Cell phones, though only as proxies, can be used to describe other human activities than calling, and their continuously expanding adoption worldwide provides an unprecedented sample in studying the daily activities of urban dwellers. Unlike other electronic devices that register usage patters (i.e. cashier counters, ATMs, subway gates etc.) in fixed positions, mobile phones travel along with people as they inhabit the city in daily life, thus describing not only how places are used, but also the personal experience of people in cities. The availability of mobile network data offers an entirely new type of evidence and scale for the dynamic study of cities.

NOTES

1 1.24 mobile phones per person, CIA World Factbook 2005.

REFERENCES

- Ahas, R., Aasa, A., Mark, U., Pae, K., Kull, T. (2007) Seasonal Tourism Spaces in Estonia: Case study with mobile positioning data, *Tourism Management*, 28(3): 898-910.
- Anderson S. (ed., 1986), *On Streets*, MIT Press, Cambridge (Mass.).
- Antonini, G., Bierlaire, M. and Weber, M. (2006) Discrete choice models of pedestrian walking behaviour, Transportation Research Part B 40(8): 667-687.
- Ben-Akiva, M. et al. (1997) Modeling Methods for Discrete Choice Analysis, *Marketing Letters*, 8:3.
- Calabreses, F., Reades, J., Ratti, C. (2007) Eigenplaces: analyzing cities using the space-time structure of the mobile phone network, *Environment and Planning B* (submitted).
- Cervero, R. (1989) *America's Suburban Centers: The Land Use - Transportation Link*, Unwin Hyman, Boston (Mass.).
- Chung, J.-S. (2003) *Pedestrian Environment as Urban Indicator: Developing a GIS Model for Measuring Pedestrian Friendliness*. MCP Thesis, MIT.
- Ellis, C.W. (1986) The Spatial Structure of Streets. In: S. Anderson, ed., *On Streets*, MIT Press, Cambridge (Mass.).
- Handy, S. (1996) Urban Form and Pedestrian Choices: Study of Austin Neighborhoods, *Transportation Research Record*, 1552, pp. 135-144.
- Hillier, B. and Hanson, J. (1984 (1989)) *The Social Logic of Space*, Cambridge University Press, Cambridge (UK).
- Hillier, B., Hanson, J. and Peponis, J. (1987) Syntactic Analysis of Settlements, *Architecture et Comportement/Architecture and Behaviour*, 3 (3). pp. 217-231.
- Mathema, K.B (2000) *Transit Useage and Urban Form: Lessons for Transit reliant neighborhoods*, MIT SMArchS thesis.
- Ratti, C., Pulselli, R.M., Williams, S. and Frenchman, D. (2006) Mobile Landscapes: using location data from cell phones for urban analysis, *Environment and Planning B: Planning and Design* 33, pp. 727-748.
- Ratti, C., Sevtsuk, A., Huang, S., Pailer, R. (2005) Mobile Landscapes: Graz in Real Time, *Proceedings of the 3rd Symposium on LBS & TeleCartography*, 28-30 November, Vienna, Austria.
- Reades J., Calabrese F., Sevtsuk A., Ratti C. (2007) Cellular Census: Explorations in Urban Data Collection, *IEEE Pervasive Computing*.
- Rodriguez, D. and Woo, J. (2002) *A case study of the relationship between non-motorized travel behaviour and the local physical environment*. Paper presented at ACSP conference.
- Zegras, C. (2005) *Sustainable urban mobility: exploring the role of the built environment*. PhD Thesis, MIT.