International Journal of ITS Research, Vol. 5, No.1, October 2007

Mapping Personal Trip OD from Probe Data

Edward Chung*1 Masao Kuwahara*2
*Institute of Industrial Science, University of Tokyo *1*
(EPFL ENAC ICARE LAVOC, Station 18, CH 1015 Lausanne, Switzerland, +41-21-693 2343, edward@iis.u-tokyo.ac.jp)
*Institute of Industrial Science, University of Tokyo*2*
(4-6-1 Komaba, Meguro-ku, Tokyo 〒153-8505, Japan, +81-3-5452-6418, kuwahara@iis.u-tokyo.ac.jp)

The paper analyses the expected value of OD volumes from probe with fixed error, error that is proportional to zone size and inversely proportional to zone size. To add realism to the analysis, real trip ODs in the Tokyo Metropolitan Region are synthesised. The results show that for small zone coding with average radius of 1.1km, and fixed measurement error of 100m, an accuracy of 70% can be expected. The equivalent accuracy for medium zone coding with average radius of 5km would translate into a fixed error of approximately 300m. As expected small zone coding is more sensitive than medium zone coding as the chances of the probe error envelope falling into adjacent zones are higher. For the same error radii, error proportional to zone size would deliver higher level of accuracy. As over half (54.8%) of the trip ends start or end at zone with equivalent radius of ≤ 1.2 km and only 13% of trips ends occurred at zones with equivalent radius ≥2.5km, measurement error that is proportional to zone size such as mobile phone would deliver higher level of accuracy. The synthesis of real OD with different probe error characteristics have shown that expected value of >85% is difficult to achieve for small zone coding with average radius of 1.1km. For most transport applications, OD matrix at medium zone coding is sufficient for transport management. From this study it can be drawn that GPS with error range between 2 and 5m, and at medium zone coding (average radius of 5km) would provide OD estimates greater than 90% of the expected value. However, for a typical mobile phone operating error range at medium zone coding the expected value would be lower than 85%. This paper assumes transmission of one origin and one destination positions from the probe. However, if multiple positions within the origin and destination zones are transmitted, map matching to transport network could be performed and it would greatly improve the accuracy of the probe data.

Keywords: *Probe, Mobile phone, GPS, Origin destination, Location based system*

1. Introduction

The planning and management of transport network require information about where people and goods are moving. Knowing the origin-destination (OD) of all trips would enable infrastructure needs to be planned and traffic better managed. Traditionally, origin-destination (OD) surveys were conducted with questionnaires. It could be a simple survey only asking drivers about their origin and destination, or a more comprehensive survey in the form of a diary. A travel diary requires the respondent to fill in all the trips made that day including other details such as mode, trip purpose, start and end time of trip. This method of collecting OD information is expensive and can only be carried out every 5 or 10 years. Furthermore, the sample size using questionnaires survey is fairly small and also it is time consuming to code the data once the survey forms are collected.

On the other hand, using location base system such as GPS installed on-board a vehicle or mobile phone carried by driver (from hereon referred to as probe) could be used to estimate OD continuously by time of day and day of week, and a much lower cost. Almost every adult carries a mobile phone and potentially the sample size is very large. While it is feasible to ascertain the OD of a trip, other information such as trip purpose will be unknown.

Previous research have focused on how to use mobile phone to collect travel time[1]–[4] and also to collect movement of people[5]. However, there were less fundamental studies on how the basic parameters such as accuracy of probe, rate of data transmission, density of urban network and mode of transport affect the application results such as travel time and origin-destination estimations.

2. Issue of using probe data for OD estimation

The authors[6] previous fundamental study looked at the issues using probe data for OD estimation and route identification. Theoretical analysis of a linear city with measurement error, *e* found that the same estimate of trip generation has the same variance under uniform error and population distributions. However, in the real world, the population (trip generation points) is usually distributed in ways different from the error distribution. In such realistic conditions, larger *e* cannot capture the actual distribution of population but it may skew the distribution by the error distribution itself. This is because, within a certain range of error, we assume the actual probe location is distributed according to the error distribution rather than the population distribution.

Although all the equations to estimate OD are derived, the practical calculation using this theoretical method can be a little tedious. Hence, simulation analysis examining the probability that OD information provided by the probe is in fact correct was applied. The scenarios simulated were an area divided into square grids of length *l* with uniform, central and fringe activities as shown in Figure 1.

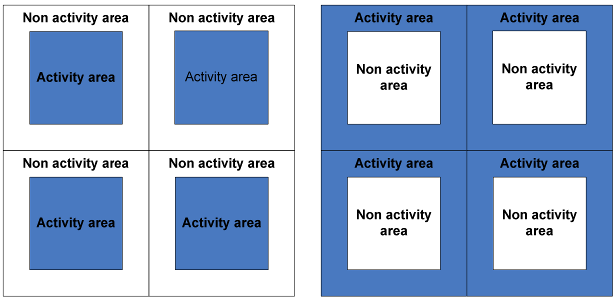


Figure 1 – Centre and fringe activity distribution

The results from the simulation can be expressed in terms of ratio between the zone size *l* and measurement error *r* (i.e. radius of circle). Figure 2 shows for activities occurring at the centre of the zone, even when $r/l=0.2$, the expected value is still 1. However, for the same r/l value if all activities occur at the fringe, the expected value is only 0.55. Note that all the expected values converge to 0.1 at $r/l=1$ (i.e. measurement error radius equals the zone size).

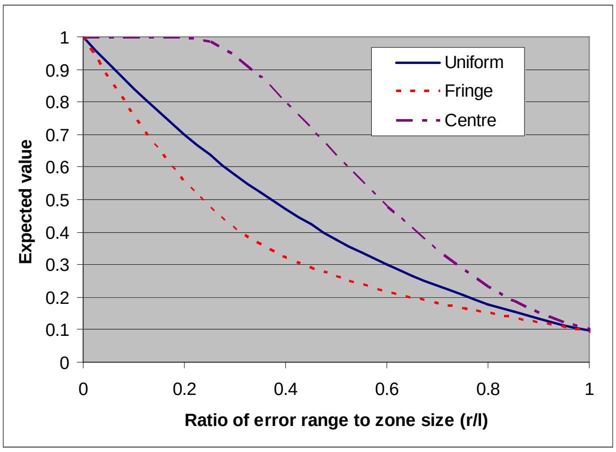


Figure 2 – OD analysis for different density distributions

This paper is an extension of the fundamental study presented above and uses the Tokyo Metropolitan Region (TMR) area and real OD to study the feasibility and issues of using probe for OD estimation. Section 3 describes the OD data used and the zones for TMR, followed by discussion about characteristics of probe in Section 4. The synthesis of the OD performed is explained in Section 5 and finally the results and conclusion in Section 6 and 7, respectively.

3. Tokyo Metropolitan Region Person Trip Survey

The Tokyo Metropolitan Region (TMR) making up of Tokyo and four other prefectures has a population of 34 million inhabitants. Every decade, the TMR Transportation Planning Committee carried out its large scale transportation fact-finding survey (Person Trip Survey), to understand the movement of people in TMR and to solve transportation issues in various localities. The first Person Trip Survey (PTS) began in 1968 and the latest was conducted in October 1998.

Households are randomly selected and people who are 5 years or more are surveyed. About 880,000 people participated in the last questionnaire survey. This represents a sample size of approximately 3%.

PTS is basically a person trip diary where the respondents enter the following information for all the trips made on the day.

1. trip start time
2. trip origin
3. modes of travel (e.g. train, bus, private car)
4. trip purpose (e.g. work, shopping, education)
5. trip destination
6. trip end time.

3.1. Zone coding

Tokyo Metropolitan Region is divided into zones and each trip's origin and destination (OD) are coded so the pinpoint locations cannot be identified for privacy reasons. Each zone is given a 5-digit code and the TMR is divided into 1648 zones (small zones) of irregular shapes. Two further zoning sizes i.e. medium and large, which are aggregated from the small zones are also defined. The first 2 digits and 3 digits of the 5 digit zone ID represent the large and medium zone IDs respectively. There are 144 medium and 52 large zones. The medium and large zones are made up of small zone though not always contiguous small zones. Keeping that in mind, the corresponding average radius of small, medium and large zones are approximately 1.1, 5 and 10 km, respectively. In this paper, small and medium zones are used for the analysis.

4. Probe properties

For transport applications, a probe could be anything that gives the position of the subject. It ranges from the common tools such as GPS, PHS (personal handy-phone system), and mobile phone with and without GPS, to less use tools such as Pedestrian Navigation Module (a combination of miniaturized low-power inertial measurement units coupled with GPS receivers and other sensors that can provide accurate position in both indoor and outdoor situations)[7]. Each type of probes has different measurement accuracy and typical values of some probes are shown in Table 1. It is important to note that different probes have different characteristics. For example, the accuracy of a mobile phone as positioning tool depends on the cell size i.e. the area of coverage of the base station that is serving the mobile phone. In areas where there are large number of activities or high level of mobile phone usage, microcell base stations are installed usually on external walls of existing buildings every few hundred metres to provide additional radio capacity. Picocell base stations are sometimes used to improved coverage and capacity within 50 metres range inside buildings, in tunnels and other difficult-to-reach environment. This means that the cell sizes are smaller in major activity centres such as city centres. In fringe areas where there are few tall buildings to block signals, macrocell base stations mounted on ground-based mast and rooftops can be used to provide coverage for a few kilometres. Therefore less accurate position of the mobile phone is expected. An illustration of the coverage of the three types of base stations is shown in Figure 3. There are various positioning methods to improve the accuracy of mobile phone location such as Cell ID, Angle of Arrival, Time of Arrival, and Time Difference of Arrival [8-11] and typical accuracy for some of these technologies is listed in Table 2.

As for GPS, the opposite outcome is observed because GPS requires clear sighting of the satellites and in urban city centres, satellite signals are often obstructed by high rise buildings. However, high accuracy is expected in areas where greater amount of open space e.g. fringe areas. In other words, the accuracy of mobile phone is higher in activity centres and less accurate in the fringe areas, whereas generally the opposite is true for GPS (lower accuracy in activity centres).

Table 1 – Typical Accuracy of Various Probes

Technology	Accuracy
GPS	2–5 m
Differential GPS	5–30 cm
Personal Handy-phone System (PHS)	20–50m
Mobile phone	50m – 2 km
Mobile phone with GPS	10–20m

In this paper, fixed measurement errors (i.e. the same measurement error regardless of location) and variable measurement errors (i.e. measurement error is a function of the zone size) are studied. The later is more typical characteristics of mobile phone and GPS, as explained above.

Table 2 – Accuracy of various technologies for mobile phone[8]

Technology	Suburban	Urban	Indoor
Cell ID	1–10 km Typical: 5km	Macrocells: 500 m – 5km Typical: 2 km Microcells: 50-500m Typical: 200m	10m – 50m (if picocells are used)
E-CGI[†]	250m–2.5km	50–550m	Highly variable
E-OTD	50–150m	50–150m	Good
A-GPS	10–20m	10–100m	Variable

E-CGI – Enhanced Cell Global ID
E-OTD – Enhanced Observed Time Difference
A-GPS – Assisted GPS

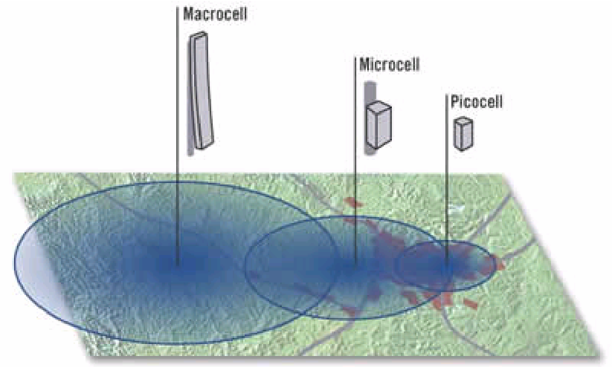


Figure 3 – Macrocell, microcell and picocell coverage Source[12]

5. Synthesising Real OD

To analyse the accuracy of probe with a given error radius for realistic ODs, the 1998 PT survey data is used. For each trip's OD, the position of trip ends are randomly generated within the origin and destination zones (see Figure 4). These random locations are where a probe would say where the person's trip ends are. Given that a probe cannot give its precise location, its true position could be anywhere within its measurement error. The measurement error is assumed to be circular in shape. If the entire circle falls within the zone, it can be concluded that the position given by the probe is definitely in the zone. However, if only a part of the circle is within the zone, the proportion of the circle inside the zone is the probability that the position given by the probe is correct.

Three different types of error are studied and described in the following sub-sections.

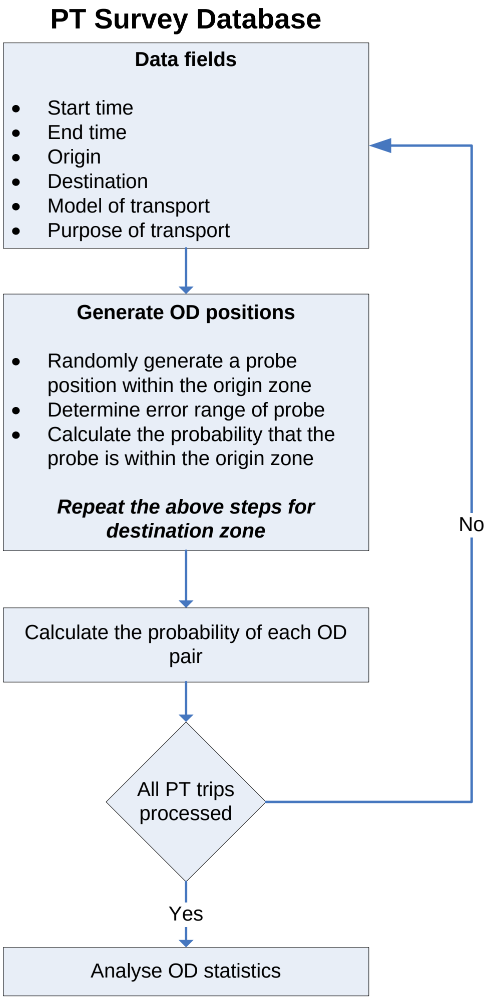


Figure 4 – Generation of Person Trips trip ends

5.1. Fixed error

Probe measurement error is assumed to be fixed regardless of the location of the probe. The error ranges from 10m to 500m

5.2. Error proportional to zone size

Probe error is a function of the zone size. The assumption here is that more accuracy position can be obtained from small zone. This is especially true for mobile phone probe where microcell base stations are installed at places with high building density and high mobile usage demand. These usually occur at small zone area.

The measurement error is show in **Figure 5**. For zone with equivalent radius <= 1.2 km a minimum error radius is used and for equivalent radius >= 2.5 km a maximum error radius is used. A linearly increasing error radius is adopted for equivalent radius between >1.2 km and <2.5 km, i.e. error increase from minimum to maximum error radius. Four combinations of error radius are studied for small zones coding and are listed in Table 3.

Table 3 – Measurement error radius used for proportional to zone size scenario

Case	Minimum error radius (m)	Maximum error radius (m)
P1	20	100
P2	50	150
P3	50	500
P4	500	2000

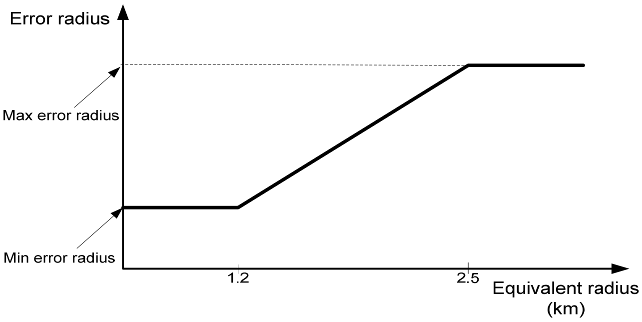


Figure 5 – Error radius proportion to zone size

5.3. Error inversely proportional to zone size

This is the opposite of the above. In this case, the measurement error is at its maximum and minimum when the equivalent radius is ≤1.2 km and ≥2.5 km, respectively. This is applicable for GPS receiver where the measurement error can be quite large in build-up areas. Two combinations of error radius are studied for small zones coding and are listed in Table 4

Table 4 – Measurement error radius used for inversely proportional to zone size scenario

Case	Maximum error radius (m)	Minimum error radius (m)
Inv1	100	20
Inv2	50	5

6. Results

The results show that for small zone coding and fixed measurement error of 100m, an accuracy of 70% can be expected. The equivalent accuracy for medium zone coding would translate into a fixed error of approximately 300m (see Figure 6). As expected small zone coding is more sensitive to medium zone coding as the chances of the probe error envelope falling into adjacent zones are higher.

For the case where measurement error is proportional to zone size i.e. mobile phone case, useful OD estimates can only be achieved for small zone coding if the probe accuracy ranges between 20m and 100m (Table 5). Most mobile phone service providers would not be able to offer such service.

For "GPS" scenario i.e. error inversely proportional to zone size (Cases Inv1 and Inv2), smaller error radii are used. Both cases are in the operating range of an average standard GPS receiver, and the expected value range from 72.2 to 84.9 for small zone coding.

Cases P1 and *Inv1* have the same error range but a mirror image of the other with error radius of 20m and 100m. Note that the expected value for P1 (88.6%) is much higher than Inv1 (72.2%). This is because over half (54.8%) of the trip ends start or end at zone with equivalent radius of ≤ 1.2 km. Only 13% of trips ends occurred at large zone (equivalent radius ≥2.5km). Therefore in terms of overall quality, for the same range of error radius, probe with error proportional to zone size (e.g. mobile phone) is more effective. Certainly, the accuracy of mobile is not as good as GPS which needs to be taken into account.

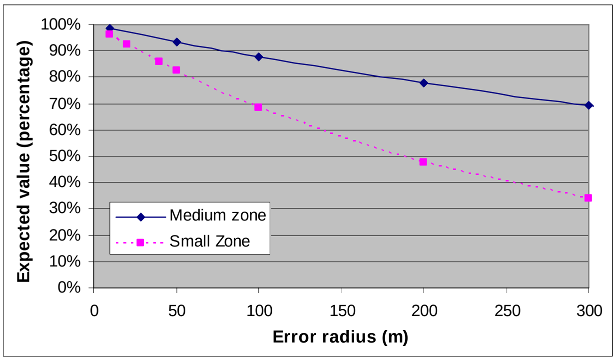


Figure 6 – Expected value from fixed measurement error radius

Table 5 – Expected value from measurement errors that are proportional and inversely proportional to zone size

Case	Expected value (%)
P1	88.6
P2	78.1
P3	66.2
P4	8.6
Inv1	72.2
Inv2	84.9

7. Conclusion

The synthesis of real OD with different probe error characteristics have shown that expected value of > 85% is difficult to achieve for small zone coding with measurement error greater than 50m. For most transport applications, OD matrix at medium zone coding is sufficient for transport management. From Figure 6 it can be drawn that GPS at medium zone coding would provide OD estimates greater than 90% the expected value. However, for a typical mobile phone operating error range at medium zone coding the expected value would be lower than 85%. Another important consideration if different probes were to be used for collect OD data is the cost of data transmission. With GPS the owner of the GPS would have to pay for the data transmission. However, for mobile phone, potentially, the data is available in the mobile phone system which could be extracted and used at low cost. There is also the privacy issue which needs to address more.

This paper assumes transmission of one origin and one destination positions from the probe. However, if multiple positions within the origin and destination zones are transmitted, map matching to transport network could be performed and it would greatly improve the accuracy of the probe data. The continuous transmission of probe data is increasingly being used for travel time estimation such as Intellione™ using mobile phone[13]. No doubt increasing measurement accuracy would make probe an efficient tool for OD survey in the very near future.

8. References

[1] Bolla, R., Davoli, F. "Road Traffic Estimation from Location Tracking Data in the Mobile Cellular Network", *Proc of IEEE 2000*, Lyon France. http://ieeexplore.ieee.org/iel5/7252/19572/00904783.pdf?arnumber=904783

[2] Holm, C., Anzek, M., Kastela, S. "Travel Time Information Service Utilizing Mobile Phone Tracking", *Promet-Traffic- Traffico*, Vol. 16, No. 4, pp 211-216, 2004.

[3] Schneider, W. and Mrakotsky, E. "Mobile Phones as a Basis for Traffic State Information", *Proc. of the 8th International IEEE Conference on ITS*, Vienna, Austria, September 13-16, 2005, pp 782-784.

[4] Ygnace, J.L., et al. "Travel time/speed estimates on the French Rhone Corridor Network Using Cellular Phones as Probes", *Final Report, STRIP (System for Traffic Information and Positioning) Project*, 2001.

[5] Asakura, Y., Hato, E. "Tracking Survey for Individual Travel Behaviour using Mobile Communication Instruments", *Transportation Research Part C: Emerging Technologies*, 12(3-4) pp 273-291, 2004.

[6] Kuwahara, M., Chung, E. and Tomotaka, I. "Fundamental study on the issues of using probe data for

OD estimation and route identification". *11th ITS World Congress*, (CD), Nagoya, Japan, 2004.

[7] Vectronix AG. "Pedestrian Navigation Module Reliable Navigation in GPS-Denied Environments", http://www.vectronix.ch/files/PNM_en.pdf.

[8] GSM Association. "Location based services", *GSM Association Reference Document SE 2.3*, Ver. 3.1.0, January 2003, pp 49-53, 2003.

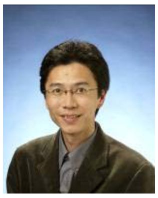
[9] Favey, E., Ammann, D., Bürgi, C. "Hybrid Positioning using GPS and GSM Ranging Measurements", *Proc. of the 17th International Technical Meeting of the Satellite Division of the Institute of Navigation*, ION GNSS 2004, pp 1918-1923, 2004.

[10] Buchanan, W., J., Munoz, J., Raja, K. A. "Critical Analysis and Error determination of Locating-finding Techniques in GSM", *International Journal of Mobile Communications*, 3 (2), pp. 165-182, 2005.

[11] Trevisani, E., Vitaletti, A., "Cell-ID Location Technique, limits and benefits: an experimental survey", *Proceedings of the Sixth IEEE Workshop on Mobile Computing Systems and Applications (WMCSA'04)*, Volume 00, pp 51-60, 2004.

[12] Mobile Operator Association. "Base stations and masts" http://www.mobilemastinfo.com/information/masts.htm

[13] National Science Foundation. "Real-Time Traffic Routing from the Comfort of Your Car" Press Release 06-124, August 30, 2006 http://www.nsf.gov/news/news_summ.jsp?cntn_id=107972&org=NSF&from=news



Edward Chung is an Adjunct Professor at the Collaborative Research Centre for Advanced Mobility, University of Tokyo. He received his Bachelor degree and PhD in Civil Engineering from Monash University. Edward is leading an ITS group at the Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland and his research interests include traffic performance, traffic safety, transport management, transport emission and traffic operations.

Masao Kuwahara is a Professor at the Institute of Industrial Science, University of Tokyo. He received his Ph.D in Civil Engineering from University of California, Berkeley. Kuwahara has about 20 years experience in teaching and researching on traffic engineering in University of Tokyo. Kuwahara's research interests include dynamic network analysis, traffic simulation, highway capacity, traffic signal control, transport emission, and transport planning.

- *Received date*
- *Received in revised form (if any)*
- *Accepted date*

- *Editor*