

## Constructing Two-layered Freight Traffic Network Model From Truck Probe Data

Takayoshi Yokota<sup>\*1</sup> and Dai Tamagawa<sup>\*1</sup>

*Advanced Transport Logistics (Hanshin Expressway),*

*Department of Urban Management, Kyoto University,<sup>\*1</sup>*

*(C-1, Kyoto Daigaku Katsura, Nishikyo-ku, Kyoto-shi, 615-8530, Japan, TEL: +81-383-3140,*

*E-mail: tyokota@kiban.kyoto-u.ac.jp, d-tamagawa@kiban.kyoto-u.ac.jp)*

The road network in the Keihanshin area (the second most populated urban region in Japan encompassing the metropolitan areas of the cities of Osaka in Osaka prefecture, Kobe in Hyōgo Prefecture, and Kyoto in Kyoto prefecture) is modeled as two layers from the viewpoint of the frequency of use of roads by freight vehicles. This model was derived, tested and validated from the probe car data of 300 trucks operated by 21 freight companies in the Keihanshin area. The two layers are called high-frequency-network and low-frequency-network layers. Characteristics, such as the density of probe data, average driving time, and speed of both network layers, were estimated. This analysis revealed that trucks spend a duration about 6 times longer on the high-frequency network even though it is 1/4 the size of the low-frequency network. The instantaneous velocity histograms indicated a significant difference in speeds between the two network layers of 25.03 km/h in the high-frequency layer and 16.87 km/h in the low-frequency layer. Service-level evaluations, reliability analysis, and route-selection modeling should become easier to deal with in the future by introducing these layered networks.

**Keywords:** *Probe car, freight traffic, network model, truck*

### 1. Introduction

We tried to identify frequently used road networks, or the distribution of probe data in other words, traveled by freight vehicles from truck probe data. This was done through an experiment involving 300 trucks from 21 freight companies in the Keihanshin area of Japan. If we can understand frequently used networks traveled by freight vehicles, we can study the behavior of freight vehicles that choose routes more efficiently than that in studies without knowledge on frequently used networks. There have been several papers that have discussed the spatial coverage problem and the construction of driver route-choice models based on probe car data. However, very few studies have been done on the spatial distribution of probe cars. To estimate the density of probe data, the target area in the Keihanshin area was divided into small cells of 20 by 20 m and the probe data that fell into these cells were counted. This count was defined as probe density. We found that the distribution in the number of cells with respect to the minimum probe density followed that of the sum of two different exponentially decaying functions. We assumed that the road network could be separated into two distinctive layers according to the frequency of the use of road links acquired from the probe data. These layers reflected the actual truck traffic flow and were generated from the viewpoint of the frequency of truck probe car traffic. Two layers, i.e., high-frequency-network and low-frequency-network layers, were generated based on

these distributions. This assumption was tested, validated and we propose a formulation of the model. This model should contribute to useful evaluations of road-network-service levels and reliability. Route-selection models should be able to be formulated based on the high-frequency-network layer in the future.

### 2. Literature Review

There have been several papers that may be related to the topic of this paper. One direction of research has been aimed at estimating the travel time based on probe data. Early work on probe car based measurements of travel time appeared in the late 1990s. For instance, Yokota [1] described the principles underlying a method of estimating the travel time based on the up-link travel time sent from vehicles via the communication capabilities of roadside infra-red vehicle detectors. He also described methods of estimating and predicting travel times from an insufficient number of probe data [2]. Fushiki [3] formulated a relationship between area coverage, the update cycle of travel time, and the traffic volume of probe cars.

There has been another direction in probe car research, i.e., concerning the identification of driver behavior in choosing routes and modeling. Kitamura [4] and Miwa [5], [6] discussed driver behavior in choosing routes from taxi-probe data assuming several different levels of drivers cognition of travel time on everyday roads. So-called data cleansing is an important issue in

processing probe-car data. Sarvi [7] proposed a method of data cleansing and identifying the end of a trip; however, the technology he proposed was for buses and taxis and was not intended to be applied to freight vehicles. Most of this work was based on taxi probe data and they did not intend to analyze the behavior of freight vehicles. This means freight vehicles have barely been studied.

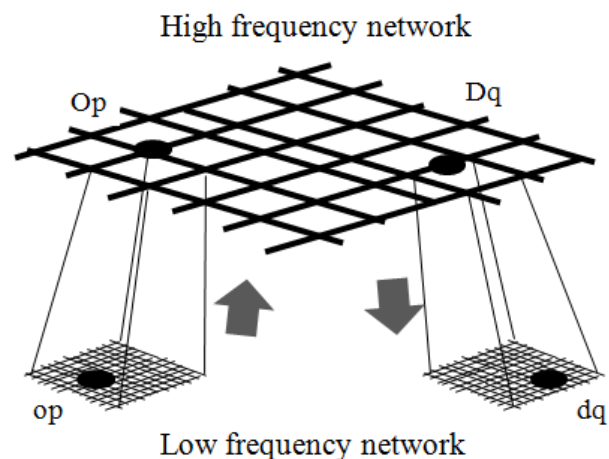
Several studies have been done, however, concerning the trip chains or tours of commercial vehicles. Figliozzi [8] proposed a set of continuous approximation models for four different types of tours. The likely impact of policies or networks changes by using the models he discussed. Figliozzi [9] also formulated a tour model and analyzed the impact of congestion with respect to different conditions. He categorized tours into three classes based on the average distance per stop and the percentage of time spent driving with an example of empirical tour data in Sydney. Greaves [10] introduced probe data based analysis of commercial-vehicle-tour data and described his pilot survey in Melbourne. He also discussed potential applications such as estimating O-D matrices and constructing trip-length distributions. However, this pilot survey studied few vehicles (30 trucks) and was not intended to obtain unbiased results. Holguin-Veras [11] discussed the trip chain behavior of commercial vehicles and estimated the probability of the trip's purpose with respect to the vehicle category and the number of trip chains, average number of stops, and length of trip chains from data on the basis of travel diaries in Denver. Brown [12] reviewed urban-freight survey techniques in the UK since the 1970s to the present mostly based on questionnaires and not based on probe data. Christian [13] reviewed urban-freight surveys in France with case studies carried out in Marseilles, Bordeaux, and Dijon. Similarly, the surveys were mostly based on questionnaires administered to drivers. Anderson et al. [14] presented survey results obtained in Norwich and London based on questionnaires. Vleugel and Janic [15] statistically investigated the route-choice behavior by freight drivers based on interviews and questionnaires in Dutch cities.

To the best of the authors' knowledge, none of the existing research has covered road choice by freight vehicles or route-choice behavior based on actual probe data. There are many road links in urban cities such as those in the Keihanshin area; however, which roads are mainly used by freight vehicles is not very well understood. It needs to be emphasized that it is important to know which roads are frequently used by freight vehicles before we proceed to investigate route-choice behavior. Understanding the frequency characteristics of road use is an approach to understanding the characteristics of freight vehicles.

### 3. Constructing Road Network Layer from Probe Data

We conducted a probe-car experiment from October 1 to 31 in which almost 300 trucks belonging to 21 different transport operators participated. We carefully chose and negotiated with the companies to diversify the locations of origin and destination. Almost 270 truck probe cars contributed to the probe experiment on weekdays while the number decreased on Saturdays, Sundays, and a public holiday (October 12). Furthermore, the total mileage of the probe cars that was counted was approximately 40,000 km on weekdays, while it decreased to about 10,000 km on Sundays and the public holiday. Layer models of road networks are often used in car-navigation systems mostly to reduce the time required in calculating routes under several criteria such as the shortest distance and the shortest time. In this case, the layers of the network were defined according to the types of road classes or kinds of empirical criteria.

Probe data were accumulated for 31 days and their density was calculated. The target area of about 20 by 20 km of central Osaka was divided into cells of 20 by 20 m. This amounted to 1 million cells. The size of the cells was chosen so that it was almost the same as the GPS error diameter. The number of probe data, which consisted of vehicle identification codes, timestamps, latitude and longitude, which fell into each cell, were counted. The interval of the probe data was 2 sec with a slight possibility of data being missed when the GPS reception was not completed. The number of probe data that fell into each cell was defined as the probe data density with units of "counts/cell·month". The concept underlying the layered network is outlined in Figure 1.



**Figure 1**  
**Concept behind two-layered road network regarding frequency of road use by trucks**

The upper layer network represents a high frequency layer network that consists of roads frequently or commonly used by freight vehicles such as highways

and major roads, while the lower layer is the low frequency layer, which consists of roads rarely used by freight vehicles. We assumed that when a truck started a trip, the origin was usually in a low frequency network such as a local road near a factory, distribution center, or a delivery destination ( $o_p$  in lower left of Figure 1). We also assumed the vehicle accessed a high frequency network such as a highway using a low frequency network (lower left network in Figure 1) and after driving along the highway for a while it exited (egress) the highway near its destination. After exiting the highway, the vehicle reached its destination using the low frequency network (lower right network in Figure 1). Each trip was assumed to take a series of transitions between the two layers in this way, starting from an origin,  $o_p$ , to the destination,  $d_q$ , in the low frequency network via  $O_p$  and  $D_q$  in the high frequency network, as illustrated in Figure 1.

#### 4 Model of Probe Density

Probe density varies from place to place. We analyzed its distribution in the following way. Figure 2 plots the profile of the number of cells with at least some designated probe density (x-axis). The number of cells over four meshes covering about 20 by 20 km of the central Osaka area, viz., 513573, 513574, 523503, and 523504 mesh codes, (JIS X 0410) is summed up. The actual number of cells, indicated by the solid line in Figure 2, is approximated by the sum of two exponential decaying functions. We estimated the optimum coefficients for the two exponential curves by minimizing the mean squared error and obtained the model of Eq. (1)

$$G(p) = 151,110 \cdot e^{-0.126957 \cdot p} + 53,539 \cdot e^{-0.007419 \cdot p} \quad (1)$$

where  $p$  is the designated minimum probe density.

The coefficient of determination  $R^2$  of the model is 0.9918, which is quite close to 1.0. The decaying coefficient of the high frequency network is -0.126957, which appears in the first term on the right hand side of Eq. (1), and this is about 17 times more significant than that (-0.007419) of the low frequency network in the second term on the right hand side of Eq. (1).

The model defined by Eq. (1) is also indicated by the dashed lines in Figures 2 and 3. Figure 3 is an enlarged graph of Figure 2 with respect to the x-axis, and the profiles of each model layer, which are the two terms on the right hand side of Eq. (1), have been illustrated separately. The model in Figure 3 represents the summation of the dotted lines of the high frequency and low frequency models. The model seems to closely fit the actual profile.

It was a little surprising that the probe distribution followed a model quite as simple as this one and we are studying why the distribution followed a simple summation of the two different exponential distributions.

The probe density function,  $G(p)$ , in Eq. (1) is defined as the integration of what we call the “cell distribution” function,  $K(q)$ , which we also want to model as defined by Eq. (2).

$$G(p) = \int_{q=p}^{\infty} K(q) dq \quad (2)$$

Here, we assume both  $G(p)$  and  $K(q)$  take the following forms, i.e., both are sums of different speed exponential distributions.

$$G(p) = A \cdot e^{-a \cdot p} + B \cdot e^{-b \cdot p} \quad (3)$$

$$K(p) = C \cdot e^{-c \cdot p} + D \cdot e^{-d \cdot p} \quad (4)$$

$$\text{where we assume } A, B, C, D, a, b, c, d > 0 \quad (5)$$

After integrating  $K(q)$  in Eq.(2) and comparing both sides, we obtain Eqs. (6) and (7) as

$$A \cdot e^{-a \cdot p} + B \cdot e^{-b \cdot p} = \left(\frac{C}{c}\right) \cdot e^{-c \cdot p} + \left(\frac{D}{d}\right) \cdot e^{-d \cdot p} \quad (6)$$

$$c = a, d = b, C = A \cdot a, D = B \cdot b \quad (7)$$

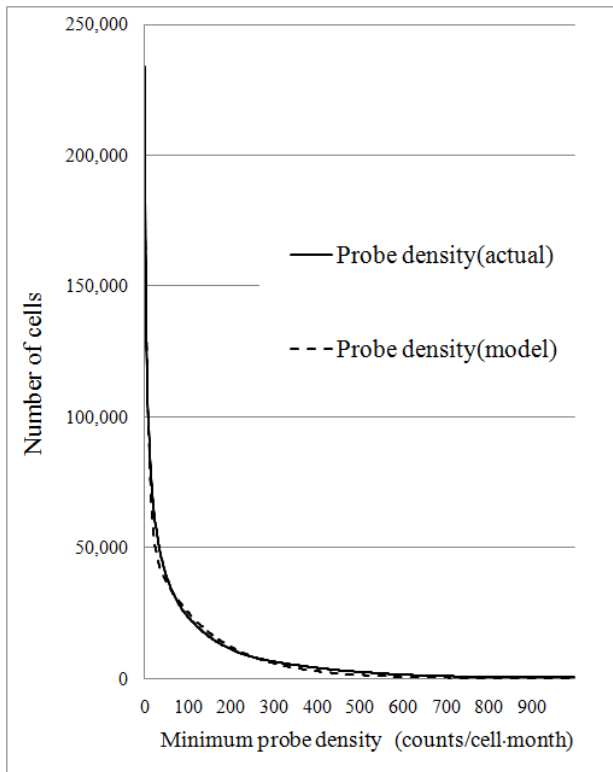
After substituting the values of  $A=151,110$ ,  $B=53,539$ ,  $a=0.126957$  and  $b=0.007419$ , respectively, for the four unknowns of  $C$ ,  $D$ ,  $c$ , and  $d$  in Eq. (7), we finally obtain the equation for  $K(p)$  as

$$K(p) = 19,187 \cdot e^{-0.126957 \cdot p} + 397 \cdot e^{-0.007419 \cdot p} \quad (8)$$

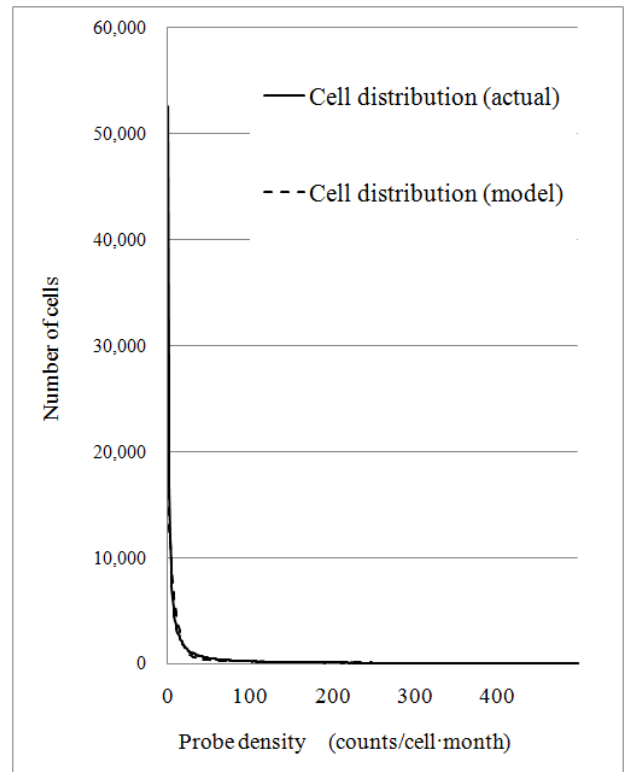
where  $p$  is the designated probe density.

The profiles of the cell distribution function in Eq. (8) are plotted in Figures 4 and 5.

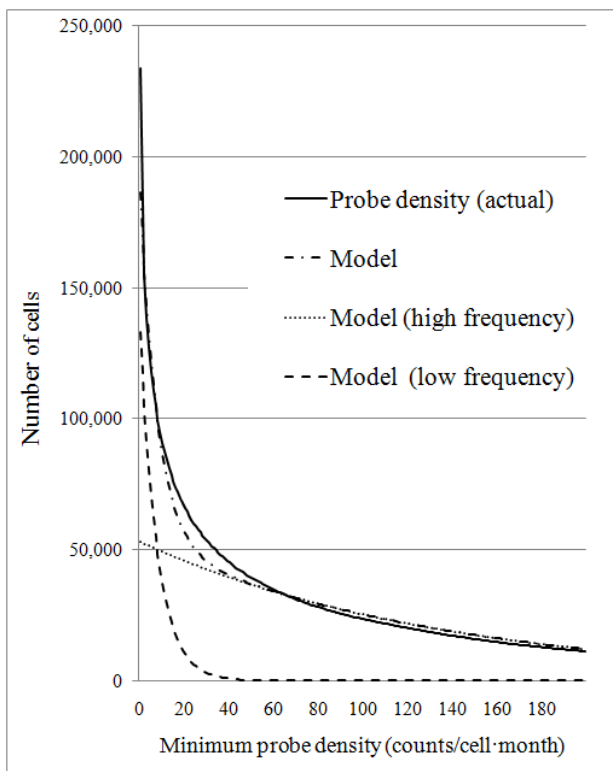
To check if the distribution model in Figure 4 and Figure 5 was statistically valid, we tried to apply a chi-square test; however, it was rejected. This was because a chi-square test is not suitable for these types of problems with thousands of degrees of freedom. The number of cases  $N$  was 2583. With these thousands of samples, small differences between the model and the actual data can easily reject the test even though the goodness of fit is as good as  $R^2=0.9918$ . Ransom and Cramer [18] pointed out a similar issue, i.e., a chi-square test only allows for sampling variations, and thus tests the hypothesis that the model distribution holds exactly. Since our model was not exactly the same as the real distribution, we thought that this was not an



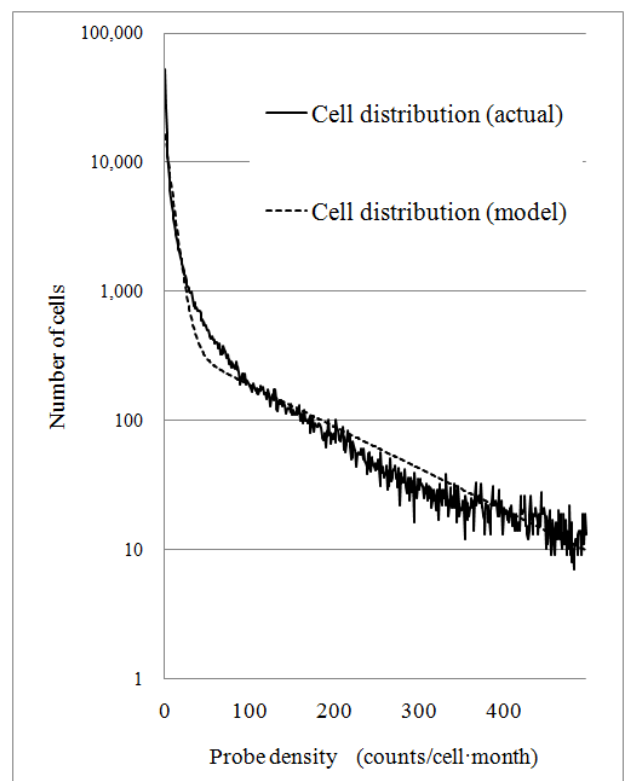
**Figure 2 Actual probe density and model probe density  $G(p)$**



**Figure 4 Actual cell distribution and model distribution  $K(p)$**



**Figure 3 Enlarged graph of actual probe density and model distribution**



**Figure 5 Actual cell distribution and model distribution  $K(p)$  on log scale**

important issue. We calculated Kullback–Leibler divergence [19] in Eq. (9) as another approach. It yielded values of 0.051 nat or 0.074 bit for the model defined in Eq.(8), which enabled us to conclude that the model fitted quite well from the viewpoint of information theory



**Figure 6 Digital road map:  
All roads with minimum width of 5.5 meters**



**Figure 7 Probe density map:  
Black lines are high frequency network  
and gray lines are low frequency network**

$$D_{KL}(P||K) = \sum_{i=1}^N P(i) \log \left\{ \frac{P(i)}{K(i)} \right\} \quad (9)$$

where  $P(i)$  is the observed normalized distribution indicated by the solid line in Figure 4, and  $K(i)$  is the model distribution derived from Eq. (8), which is plotted by the dashed line in Figure 4 and Figure 5. If we make the model more elaborate by introducing more factors to it, this may contribute more to the goodness of fit; however, it will become more complex to interpret the model. The same examples have been presented to enable the differences between the two layers to be compared. The digital roadmap of mesh code 523503 is in Figure 6 where all roads over 5.5 meters wide have been shown. There is a probe density map in Figure 7 where the solid black lines correspond to the high frequency network and the gray lines correspond to the low frequency road network. The threshold that separates the high frequency cells from those of low frequency is 30 probes/cell-month, which roughly means that at least one of 300 trucks visits these links every day on average.

## 5. Difference in Characteristics Between Two Network Layers

Our next concern was to find whether there were any significant characteristic differences between the high frequency and low frequency networks. One serious concern may be to find how long the driver is staying in each network. The number of probes with respect to probe density of each layer is shown in Figure 8. Here, the distribution is calculated with the cell distribution model in Eq. (8) and multiplying it by the probe density. Table 1 summarizes the characteristics of the probe data in each network.

**Table 1  
Comparison of two layered networks**

|                        | Cells   | Cell ratio | Probes    | Time Spent (hr) | Probe Ratio |
|------------------------|---------|------------|-----------|-----------------|-------------|
| High frequency network | 53,539  | 26.16%     | 7,189,732 | 3,994           | 86.56%      |
| Low frequency network  | 151,110 | 73.84%     | 1,116,289 | 620             | 13.44%      |

Since each piece of probe data was recorded every 2 sec, the number is proportional to the time trucks were driven in each network. The probe density map such as that in Figure 7 indicates the actual frequency that trucks use road networks. Instead of counting the actual

probe data in each network, we counted the number of cells and probe density, because, we do not currently think that map matching is necessary for analyzing network characteristics. In map-matching processing, each item of probe data, which has longitude and latitude, is assigned to a corresponding nearby point on the link of a digital road map. This is useful for calculating the link travel time or congestion or for any link-based studies. For example, once the link travel time is obtained, studies on choices of routes or applications such as calculating the fastest route can be dealt with. However, map-matching processing especially that based only on GPS data is quite difficult in metropolitan areas such as in the Keihanshin area. There are many cases where elevated roads coexist with parallel roads running underneath them. In these cases, it is quite difficult to distinguish roads only from GPS probes that do not have adequate height resolution.

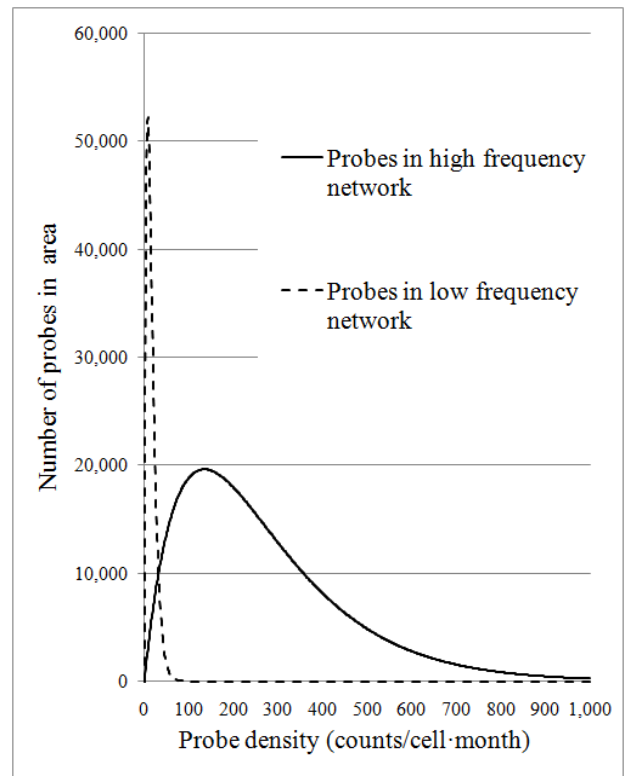
Hence, we currently propose cell-based analysis in this paper instead of link-based analysis until a map-matching technology with sufficient reliability is developed. How to overcome this problem will be another focus of our research in the future. Although the ambiguity of roads in dense areas has been left unresolved, it is sufficient to check whether the proposed layered network model is valid or not.

The discovered network sizes in terms of cells, i.e., the number of cells in the high frequency network, is estimated to be 53,539 while that in the low frequency network is 151,110. The ratio is about 26% to 74%, as illustrated in Figure 9. The total amount of time drivers spend in each network is derived by integrating the profiles plotted in Figure 8. As seen in Figure 10, 86.56% of the time is spent in the high frequency network while only 26.16% of cells are in the network. Therefore, trucks spend durations that are about 6 times longer in the high frequency network even though it is 1/3 the size of the low frequency network. The ratio of the number of cells can be substituted for the ratio of the road length, which we could not calculate from the digital roadmap data base because we did not do map-matching in our study. It should be emphasized that roughly 90% of the time spent by trucks was in the high frequency network. Thus, improving the reliability of the high frequency network, its efficiency, and safety deserves a higher priority.

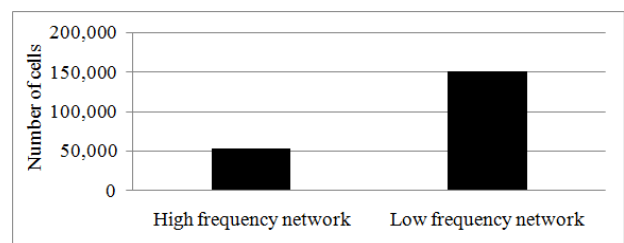
Because this high frequency network is derived from the probe data, it is adaptive to actual truck probe data. A reasonable threshold for the two layers can be the point at which the two curves cross in Figure 8, i.e., about 30 counts/cell·month in this case.

Figure 11 shows the extracted high frequency network covering about 20 by 20 km in central Osaka. We can clearly see that only limited networks are used by trucks. Figure 12 shows both of the two layered network cells. The high frequency network cells are in black and the low frequency network cells are in gray. Both kinds of cells form corresponding networks. To see if there is a

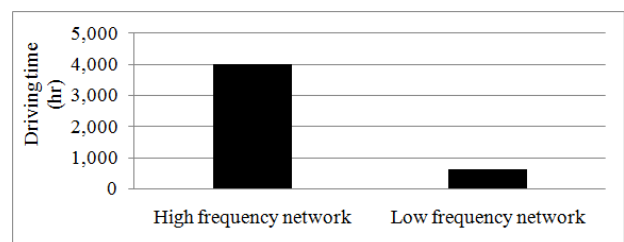
difference in the velocity characteristics between the two layers, the instantaneous velocity data of 244 trucks on Monday, October 1st, 2009 were analyzed. The probe data, which were recorded every 2 sec, were data cleansed by re-sampling every 20 sec according to the method explained in the next section to eliminate stops related to truck operation.



**Figure 8 Probe density distribution in both network layers**



**Figure 9 Number of cells in both layers**



**Figure 10 Driving time in each layer**

The results are in the graphs in Figures 13 and 14. These were calculated for four meshes of 513573, 513574, 523503, and 523504, as illustrated in Figures 11. The velocity distribution for the high frequency network has slightly shifted to a higher speed than that of the low frequency network. The mean value is 16.87 km/h in the low frequency network and 25.03 km/h in the high frequency network. The difference is quite obvious.



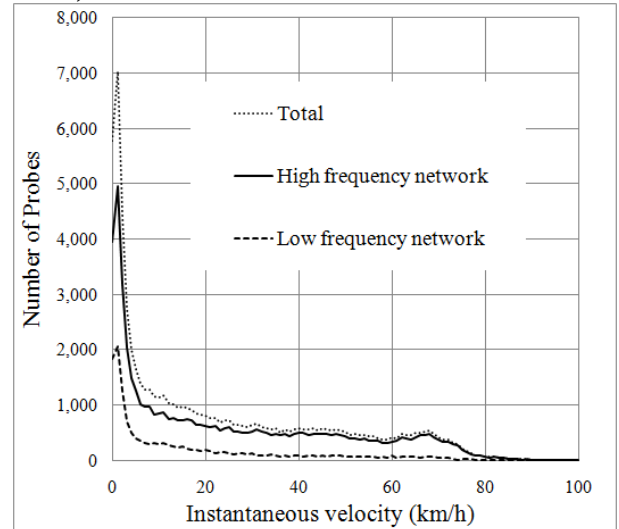
**Figure 11 High frequency network cells for 4 meshes**



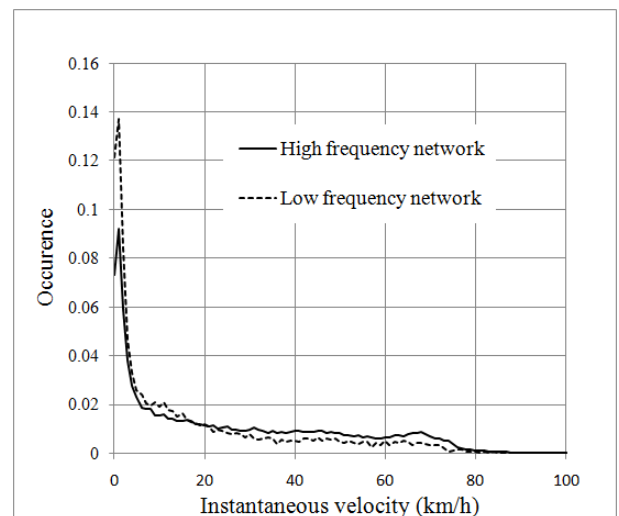
**Figure 12 High frequency network cells and low frequency network cells**  
Solid lines indicate high frequency network and gray lines indicate low frequency network

The number of re-sampled probes on October 1st that fell into the high frequency network was 53,843, while

there were 15,092 in the low frequency network. The percentage for the former is 78.1% and that for the latter is 21.9%. This is close to the percentages in Table 1, which were derived with the model obtained from the one month survey (86.6% for the former and 13.4% for the latter.)



**Figure 13 Velocity graph of both layered networks**

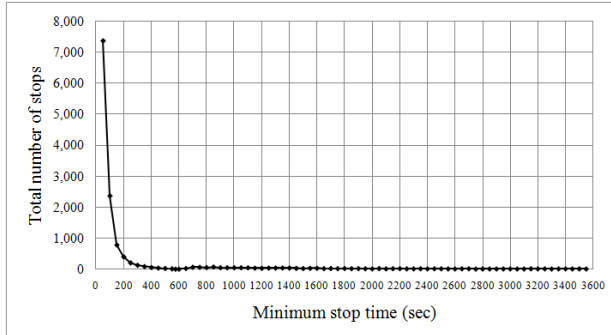


**Figure 14 Normalized velocity graph of both layered networks**

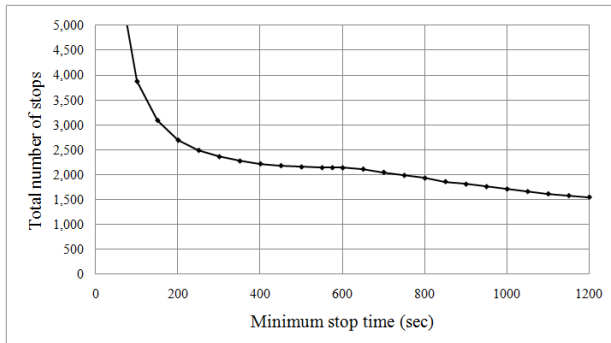
## 6. Extraction of Stopping Points for Trucks

The number of probe data for 300 trucks during one month in October 2009 consisted of 36,761,274 records. The data consisted of vehicle ids, time stamps, longitude, and latitude. To evaluate the velocity characteristics of the road network, we must distinguish stops, which are due to inherent truck operations such as loading or unloading goods, resting, or stopping at intersections.

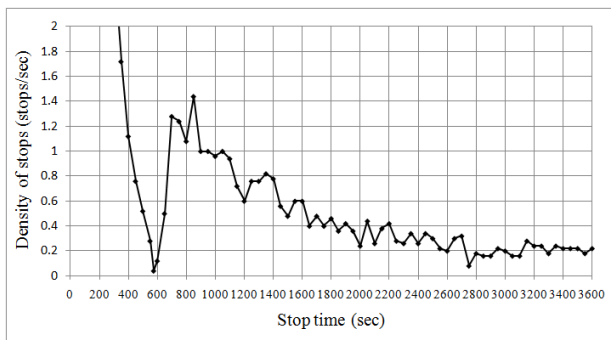
Figures 15(a) and (b) plot the distributions for the estimated number of vehicle stop points versus the minimum stop time in all the probe data in all areas of the Keihanshin area where there were probe data including the two layered networks on a typical weekday on October 1. The only difference between Figures 15(a)



**Figure 15(a) Estimated Number of Stop Points**



**Figure 15(b) Estimated number of stop points**



**Figure 16 Estimated density of stop Points**

and (b) is in the shown scales. We chose 2.0 km/h as the threshold value to determine that a vehicle had stopped based on the results in the Appendix, i.e., a 95% confidence interval for the velocity corresponding to actual stops, taking GPS positioning fluctuations into account. We can see that there is a saddle point in the range of minimum stop times that ranges from 200 to

600 sec in Figures 15(a) and (b). Figure 16 plots the density of the number of stop points versus the duration of stop time. This was obtained by differentiating the data of Figure 15(b). Based on these results, we set the threshold time of stops related to truck operation at 600 sec. This meant that any stops that lasted for less than 600 sec were regarded as stops caused by traffic congestion although situations where vehicles waited for 600 sec at intersections or in congested traffic were very rare. However, this is not critical if we chose 200 sec as the threshold value from Figure 15(a).

## 7. Conclusion

We presented results that we obtained by analyzing truck probe data. The six main findings were:

- 1) The road network could be modeled as a two layered network, and both layers could be adaptively constructed according to the frequency of road use, which was calculated with probe data.
- 2) Trucks in the Keihanshin area spent about 6 times longer on the high frequency network even though it was 1/4 the size of the low frequency network. It should be emphasized that 87% (model) or 78% (actual) of the time spent by trucks was in the high frequency network in the Osaka area. Thus, improving the reliability, efficiency, and safety of the high frequency network deserves a higher priority.
- 3) The instantaneous velocity histograms did reveal a significant difference between the two layers as one of their main characteristics, e.g., the mean value was 16.87 km/h in the low frequency network and 25.03 km/h in the high frequency network.
- 4) There was a saddle point in the profile of the number of stop points for trucks. The saddle point ranged from 200 to 600 sec. This means that trucks' inherent stops ranged from about 600 seconds.
- 5) Evaluation of services, analysis of reliability and modeling of route selection should become easier to deal with and become simpler and more effective based on this two-layered network.
- 6) The concept of the layered network model may also be applied to traffic involving passenger vehicles.

## 8. Acknowledgements

The authors would like to thank the Japan Institute of Logistics Systems, Osaka Trucking Association, Kansai Economic Federation, and all the 21 companies who participated in the probe experiment.

## 9. References

- [1] T. Yokota et al., "Evaluation of Advanced Mobile Information Systems (AMIS) for the Winter Olympic Games in Nagano, Japan 1998", 10<sup>th</sup> World Congress on Intelligent Transport Systems, Seoul, 1999.

- [2] T. Yokota, "Estimation and Prediction for Floating Car Based Traffic Information System", *Journal of Japanese Society for Artificial Intelligence*, Vol. 22, No. 4, 2007, pp. 523–528.
- [3] T. Fushiki et al. "Study on density of probe cars sufficient for both levels of area coverage and traffic information update cycles", 11th ITS worlds 2nd Annual National Urban Freight Conference, Long Beach, CA. Dec., 2007.
- [4] S. Kitamura et al., "Descriptive Analysis of Detouring Traffic through Neighbourhood Streets Using Probe Car Data", 10<sup>th</sup> World Congress on Intelligent Transport Systems, Madrid, 2003.
- [5] T. Miwa et al., "Route Identification and Travel Time Prediction Using Probe-Car Data", *International Journal of ITS Research*, Vol. 2, 2008, No. 1, pp. 21–28.
- [6] T. Miwa and T. Morikawa, "Analysis on Route Choice Behaviour Based on Probe-Car Data", 10th World Congress on Intelligent Transport Systems, Madrid, 2003.
- [7] M. Sarvi, "A Methodology for Data Cleansing and Trip End Identification of Probe Vehicles", *Proceedings of the JSCE Conference of Infrastructure Planning*, Vol. 26, Nov. 2002.
- [8] M. Figliozzi, "Analysis of the Efficiency of Urban Commercial Vehicle Tours Data Collection Methodology, and Policy Implications", *Transportation Research Part B: Methodological*, Vol. 41, No. 9, 2007, pp. 1014–1032.
- [9] M. Figliozzi, "The impacts of congestion on commercial vehicle tour characteristics and costs", *Transportation Research Part E: Logistics and Transportation Review*, Volume 46, Issue 4, Jul. 2010, pp. 496–506.
- [10] S. P. Greaves, "Collecting Commercial Vehicle Tour Data with Passive Global Positioning System Technology", *Transportation Research Record* 2049, pp. 158–166.
- [11] J. Holguin-Veras, "Observed Trip Chain Behavior of Commercial Vehicles", *Transportation Research Record* Vol. 1906, 2005, pp. 74–80.
- [12] Browne et al. "Analysing the results of UK urban freight studies", pp. 71–99, *Proceedings of the 6<sup>th</sup> International Conference on City Logistics*, Puerto Vallarta, 2009.
- [13] A. Christian et al., "Urban freight establishment and tour-based surveys for policy-oriented modelling", pp. 161–177, *Proceedings of the 6<sup>th</sup> International Conference on City Logistics*, Puerto Vallarta, 2009.
- [14] S. Anderson, M. Browne, and P. Jones "A framework for considering policies to encourage sustainable urban freight traffic and goods/service flows", Summary Report, University of Westminster, 2000.
- [15] J. Vleugel and M. Janic, "Route choice and the impact of 'Logistics Routes', in *Logistics Systems for Sustainable Cities*", Elsevier, Amsterdam, Netherlands, 2004.
- [16] M. R. Ransom and J. S. Cramer, "Income Distribution Functions with Disturbances", *European Economic Review* 22, pp. 363–372, 1983.
- [17] Kullback, S. and Leibler, R. A., "On information and sufficiency", *Annals of Mathematical Statistics* 22: pp. 79–86, 1951.

The copyright for the digital road map data used in drawing Figure 6 is owned by Sumitomo Electric Industries, Ltd.



**Takayoshi Yokota** is a professor of Advanced Transport Logistics (Hanshin Expressway) in the Department of Urban Management at the Graduate School of Engineering at Kyoto University, Japan. He received his B.S. M.S. and Ph. D from the Tokyo Institute of Technology in 1979, 1981, and 1984. He was born in 1956 in Ibaraki prefecture, Japan. He joined the Hitachi Research Laboratory of Hitachi Ltd. in 1984. He was the chief researcher responsible for Hitachi's R&D of intelligent transport systems and car navigation systems from 1993–2008. He joined the faculty of Kyoto University in April 2009. His current research interests include analysis, modeling, and spatio-temporal information processing of urban freight-transport systems.



**Dai Tamagawa** is an assistant professor of Advanced Transport Logistics (Hanshin Expressway) in the Department of Urban Management at Kyoto University, Japan. He was born in 1974 in Hiroshima prefecture, Japan. He received his B.S., M.S. and Ph.D from Kyoto University in 1997, 1999, and 2007. He worked for the Hanshin Expressway Company Ltd. from 1999–2009. His current research interests are city logistics measures and network evaluation by using probe car data. He has focused on applying multi-agent

simulation to evaluating city logistics measures taking into consideration several stakeholders associated with urban freight transport.

## APPENDIX

### Criteria for Determining Stops Taking into Account GPS Positioning Error

The probe data suffers from measurement errors. The GPS errors in this experiment were measured for the purpose of setting an optimum threshold to determine whether a vehicle had stopped. Figure 17 shows a typical example of the trajectory for a truck that stood still for 460 sec. The latitude and longitude fluctuated and their histograms for two orthogonal directions are in Figures 18(a) and (b). The absolute GPS positioning error is plotted in Figure 19. From this, we determined that 90% percentile error was within about a 10-m radius (20-m diameter) and 95% percentile error was within 12 m radius. We need to set a threshold value for instantaneous velocity to determine whether a vehicle has stopped. Because of fluctuations in GPS measurements, the instantaneous velocity that was calculated from adjacent-position data was not free from error. This error becomes significant as the sampling interval of GPS data decreases. Figure 19 plots the occurrence of velocity error calculated with different sampling intervals. The probe data are the same as those used in Figures 17 and 18, i.e., the vehicle is known to be standing still. From this, it is obvious that the velocity error decreases as the interval increases. This is because the GPS positioning error is upper bounded while the sampling interval can be increased. However, the temporal resolution, which is important for analyzing the characteristics of the movement of vehicles, will reduce as the interval decreases. We finally set the interval value to 20 sec as the best compromise. In this case, the 95% error percentile of velocity error is about 2 km/h which is below human walking velocity and quite a reasonable value to distinguish stops from running.

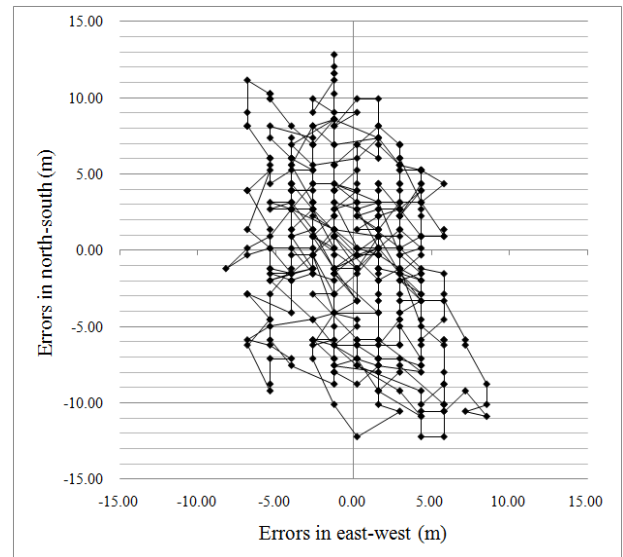


Figure 17 Typical GPS fluctuations

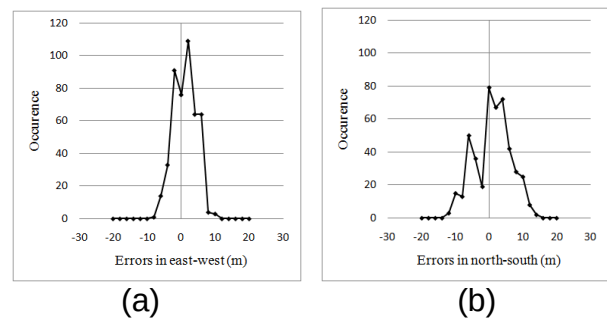


Figure 18 Graph of GPS fluctuations

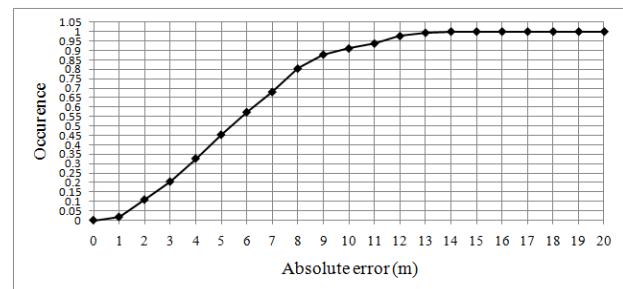


Figure 19 Typical GPS absolute error distribution

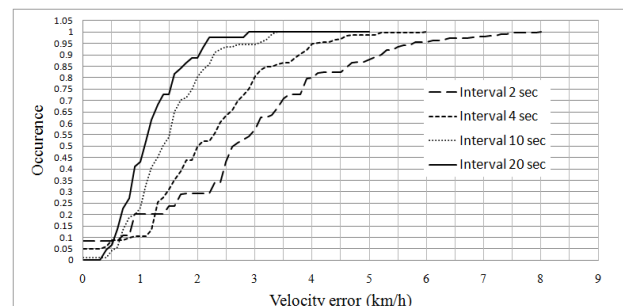


Figure 20 GPS Velocity error distributions