AN ANALYSIS ON ACCURACY OF REPRODUCED ROUTE AND SPEED BY TRANSMITTED DATA AT DIFFERENT FREQUENCY FROM PROBE VEHICLE *

by Kai Liu** and Toshiyuki Yamamoto***

1. Introduction

Probe vehicles are playing an increasing role in the field of ITS, while showing high value for applications in route guidance, travel time prediction, traffic control, congestion monitoring. However the diffusion of probe vehicles faced a serious difficulty. Different from other detector data, probe vehicle data are dynamic and successive, which include a great deal of information but need enormous expense. It is the high cost of transmitting the information from the vehicle that limited its expansion.

In order to expand the scope and to improve the density of probe vehicles, with every corner of the city having applicable and enough probe vehicles data, the most emergent and important thing that should be done is reducing the cost. To drop the frequency down to a reasonable extent is in practice one of the effective methods. Unquestionable the higher the frequency of probe vehicle data, the more information can we get under the same condition of technology and equipment. With the frequency dropping down, the information reduced which may affect the accuracy of observation.

However few studies have been undertaken on the accuracy of different frequency in the transmission of the information from probe vehicle, especially the diversity between high frequency data and low frequency data. The main purpose of this study is to examine and estimate the accuracy of different frequency data.

2. Study Objects

The highest frequency data that we now have in Nagoya ITS project is at the time lag of 5 seconds. Comparison of accuracy should be carried on under the same condition, for the sake of eliminating external effect. We use only high frequency data from which the relatively low frequency data were taken out by removing parts of records evenly according to the required frequency. For example, when selecting 1/2 data, delete the second one from every successive two records, with the starting point and ending point for every trip preserved which are required for map-matching. We got eight groups of data for each vehicle, and also for each trip, with the according interval time as: 5 sec., 10 sec., 15 sec., 20 sec., 25 sec., 30 sec., 35 sec. and 40 sec. We select seven places in Nagoya as origin and destination, between which there are totally 181 trips.

It is hardly to measure the absolute accuracy value of all different frequency probe vehicle data, while it is simply and still valuable to calculate the accurate ratio relative to the most accurate one in practice, which is of course the one with the highest frequency and the group of data with 5 seconds interval in this study.

3. Accuracy of Map-matching Results

The accuracy of map-matching using probe vehicle data nowadays directly determines the accuracy of traffic navigation.

---

* Keywords: probe vehicle data, frequency, accuracy
** Student Member of JSCE, M.Eng., Department of Civil Engineering, Nagoya University,
(Furo-cho, Chikusa-ku, Nagoya 464-8603 JAPAN, Tel: +81-52-789-3729, Fax: +81-52-789-3738, E-mail: liukai@trans.civil.nagoya-u.ac.jp)
*** Member of JSCE, Dr. Eng., Department of Civil Engineering, Nagoya University,
(Furo-cho, Chikusa-ku, Nagoya 464-8603 JAPAN, Tel: +81-52-789-4636, Fax: +81-52-789-3738, E-mail: yamamoto@civil.nagoya-u.ac.jp)

Map-matching using higher frequency data which include more information can get obviously more precise route identification than using lower frequency data in the same condition of deviation with the same technology and equipment. In most cases, there exist many routes to be matched if the interval time between two adjacent records is too long (low frequency), which results possibly in the incorrect map-matching and limits the application of probe vehicle data. There must have a level of frequency that the data to be matched to the correct route with higher probability can meet the requirement of map-matching.

Assuming that the highest frequency data as real-time data are always correct and catch the most accuracy information, the accuracy of low frequency data can be measured by comparing with the high frequency data. That is to say, imaging that map-matching results using the group of data with the lag of 5 seconds always catch the real route, other groups of data may bring to some map-matching errors during the whole trip, the scope of which can be seen as the measuring of inaccuracy. The map-matching algorism we used has been proved to be relatively highly accurate[1)], assuring the incorrect route identification being based on the low frequency of data.

We got eight routes for every trip, which may be identical or partly different from each other, by map-matching using eight groups of different frequency data, with which we can compare the accuracy indices of map-matching. The indices are as follows: Correct Trip Numbers (CTN), means trip numbers of absolutely correct route identification; Correct Rate of Trip Numbers (CRTN), means the rate of CTN; Correct Link Numbers (CLN), means the total numbers of completely identical links of 181 trips; Correct Rate of Link Numbers (CRLN), means the rate of CLN; Correct Length (CL), means the total length of completely identical links of all 181 trips; Correct Rate of Length (CRL), means the rate of CL.

The results of total 181 trips are showed in Table 1, which shows that all the six indices drop along with the increasing interval time. The lowest accuracy of map-matching results is the data with the interval of 40 seconds whose indices of CRTN, CRLN, CRL are 60.8%, 93.8% and 93.9% respectively. That is to say, the map-matching program got about a mean of 93.9% correct information when the interval time increase from 5 seconds to 40 seconds while the expense decrease sharply to about 1/8. Figure 1, showing the change of CRLN and CRL according to the interval time, may help us have a draft of accuracy about different frequency data. We can see from this figure that the dropping slope of accuracy become precipitous and unstable when the time lag is over than 25 seconds, from which about 97% of correct information can be caught, while the decreasing slope of accuracy from 5 seconds to 25 seconds are relatively gentle and equivalent.

Table 1：The accuracy difference of map-matching results

	CTN	CRTN	CLN	CRLN	CL(meters)	CRL
5seconds	181	100%	6924	100%	617403	100%
10seconds	156	86.2%	6831	98.7%	611030	99.0%
15seconds	152	84.0%	6819	98.5%	608371	98.5%
20seconds	134	74.0%	6749	97.5%	603762	97.8%
25seconds	130	71.8%	6686	96.6%	600168	97.2%
30seconds	121	66.9%	6599	95.3%	591139	95.7%
35seconds	118	65.2%	6591	95.2%	588392	95.3%
40seconds	110	60.8%	6496	93.8%	579853	93.9%

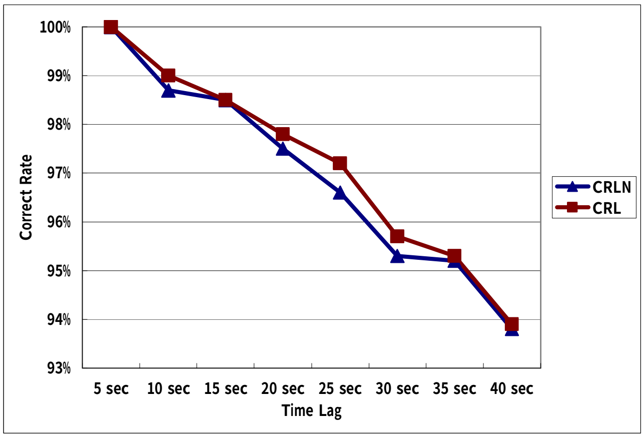

Figure 1：The decreasing accuracy of different frequency data

Further the differences between the map-matching accuracy for the trips of different length using the same frequency data are examined. The all 181 trips were divided into four groups according to the distance, the first group is below 2km which including 32 trips, the second group is from 2km to 4km including 116 trips, the third group is from 4km to7km including 17 trips, and the last group is over 7km which including 16 trips. Three of accuracy indices of these four groups are showed separately in Table 2. The accuracy values of group 1 based on all frequency data are always high and have little differences, and so do the two segments of frequency with the time lag of 10 seconds and 15 seconds for all length groups. It seems that all frequency data are more appropriate in the group 1 of trips than other groups and the data with the time lags of 10 seconds and 15 seconds are more appropriate for all groups of trips than other frequency data do. Actually it is the problem of map-matching methods which can be solved by dividing a long length trip into several short length trips.

Table 2: The accuracy difference of map-matching results based on different length of trip

	0-2km (group 1)			**2-4km (group 2)**			**4-7km (group 3)**			**over 7km (group 4)**		
	CRTN	**CRLN**	**CRL**	**CRTN**	**CRLN**	**CRL**	**CRTN**	**CRLN**	**CRL**	**CRTN**	**CRL.N**	**CRL**
10seconds	**87.5%**	**99.1%**	**98.5%**	**89.7%**	**99.0%**	**99.1%**	**94.1%**	**99.0%**	***99.6%***	**50.0%**	**96.9%**	**98.6%**
15seconds	**90.6%**	**98.9%**	**99.3%**	**83.6%**	**98.3%**	**98.6%**	**88.2%**	**99.7%**	***99.6%***	**68.8%**	**97.6%**	**97.7%**
20seconds	**78.1%**	**97.7%**	**98.5%**	**80.2%**	**98.0%**	**98.3%**	**64.7%**	**97.9%**	**97.9%**	**31.3%**	**95.2%**	**96.7%**
25seconds	**78.1%**	**98.2%**	**98.8%**	**75.0%**	**97.1%**	**97.8%**	**58.8%**	**93.8%**	**94.1%**	**50.0%**	**96.0%**	**97.2%**
30seconds	**71.9%**	**96.8%**	**98.0%**	**71.6%**	**95.3%**	**95.8%**	**52.9%**	**94.2%**	**94.1%**	**37.5%**	**95.3%**	**95.7%**
35seconds	**78.1%**	**98.0%**	**98.1%**	**69.0%**	**95.9%**	**96.3%**	**47.1%**	**92.3%**	**92.8%**	**31.3%**	**93.7%**	**93.9%**
40seconds	**75.0%**	**97.3%**	**98.4%**	**65.5%**	**94.8%**	**95.6%**	**41.2%**	**92.7%**	**92.8%**	**18.8%**	**89.4%**	**90.1%**

We must note that all accuracy values above are relative accuracy but not absolute accuracy because of the assumption of the always correct map-matching results for 5 seconds data which we know is not true, in some cases there still are some errors in the map-matching results using 5 seconds data or even higher frequency data. However, with this relatively accuracy results on our minds, it became easier to select suitable frequency in different situation.

4. The Accuracy Difference on Real-time Information

Although the map-matching program can be improved to have relaxed confinement on the frequency of data, the difference between high frequency data and low frequency data is still enormous due to the massive information that were included in the high frequency data. For example, in the situation of getting and forecasting the continuous real-time length and real-time speed in one trip, the higher the frequency of data, the more accurate information can we get, and also the more accurate estimation. That is to say that there are some deviations for each frequency data on the accuracy of the information about real-time position and real-time speed, the extent of which are different according to the frequency of probe vehicle data. This study evaluates these deviations from flowing two aspects.

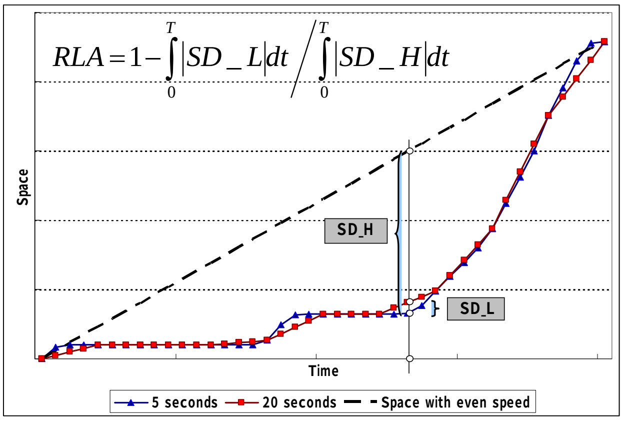

Figure 2：Sketch map of RLA calculation method for one trip

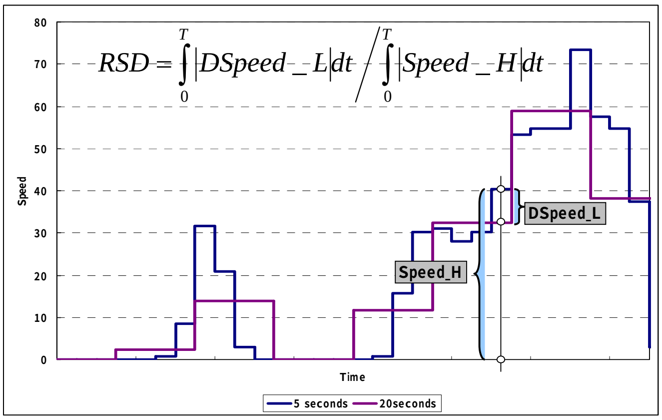

Figure 3 Sketch map of RSD calculation method for one trip

The first aspect is to what extent that the data with a given frequency can get the accurate running real-time length, which is measured by the rate of relative accuracy of real-time length named as Real-time Length Accuracy (RLA). And the second aspect is to what extent that the data with a certain frequency can get the accurate running real-time speed, with the index of the rate of speed deviation during the total trips named as Real-time Speed Deviation (RSD). Figure 2 and figure 3 explain the actual meaning of these indices using the data with the interval of 20 seconds as example.

These will be done under the assumption that the map-matching results using all groups of data with different frequency are all correct and the probe vehicle data with the time lag of 5 seconds can always get all the information when estimate the real-time length and real-time speeds. The two accuracy indices under this assumption reflect the estimation accuracy with the larger value of RLA meaning high accuracy and the smaller value of RSD meaning little estimation deviation.

The results are shown in Table 3. We can see from Table 3 that the changing extents of these ratios are larger than the accordingly accuracy of map-matching results especially the value of RSD ranging from 0.0% to 41.55%. As the real-time length for granted, the index of RLA, ranging from 100% to 87.55%, decreased about 2 percents if the interval time of adjacent records increase 5 seconds, which means that the deviation of real-time length increased only about 2 percents of the distance difference between the real-time length and the mean length of the whole trip. These values are relatively high and most of these frequency data may satisfy the estimation. Unfortunately the values of RSD change rapidly especially for the relative high frequency data with the interval time being 10 seconds, 15seconds, 20seconds in this study. These results showed that the frequency dropping down for probe vehicle data have great effect on the estimation of real-time speeds and other attributes relied on the information of speed such as the travel time estimation.

Table 3: The accuracy difference on real-time information among different frequency data

	5 seconds	**10 seconds**	**15 seconds**	**20 seconds**	**25 seconds**	**30 seconds**	**35 seconds**	**40 seconds**
RLA	**100%**	**98.76%**	**97.46%**	**95.73%**	**93.96%**	**91.75%**	**89.77%**	**87.55%**
RSD	**0.0%**	**15.98%**	**21.49%**	**26.57%**	**30.90%**	**35.48%**	**38.50%**	**41.55%**

5. Conclusion and Future Research

This study focused on the accuracy difference among the different frequency data and estimated the accuracy value from two aspects, one is map-matching results and another is the real-time attributes. Although the low frequency data showed relatively high accuracy when map-matching, their values of accuracy when reflecting the real-time information were fairly low which might influence traffic navigation and travel time forecasting in the next steps.

It is really difficult to say what extent of frequency for the probe vehicle data is reasonable. However from the aspect of map-matching, the applicable interval time should be much better if no more than 25 seconds, over 25 seconds the map-matching appears complicated and have high error rate because the average links (road segments) length is shorter than 200 meters in Nagoya which need normally about 15~20 seconds.

Future analysis will be focused on the accuracy difference of the travel time estimating of every link with different frequency data, which is a very important section of traffic navigation. And then another kind of high frequency data at the length lag of 50meters will be examined, same as the analysis on the time frequency data.

References

1) Miwa, T., Sakai, T. and Morikawa, T.: Route identification and travel time estimation using probe-car data, Proceedings of Symposium on ITS 2003, pp.277-282, 2003 (in Japanese).