COMPARISON OF THE VARIABILITY OF LINK TRAVEL TIME ESTIMATES BETWEEN DIFFERENT FREQUENCY PROBE VEHICLE DATA *

By Kai Liu** and Toshiyuki Yamamoto***

1. Introduction

The probe vehicles have been regarded as a promising source of providing link travel times within signalized networks used for DRGS (dynamic route guidance system). The successful large scale development of DRGS depends on high reliability of travel time estimation. Thus, a lot of studies have previously investigated the expected reliability of probe travel time reports. Sen et al. (1997) found that probe reports on link travel times are not independent, and that the sample mean, however large the sample size is, can not approach the population mean[1)]. Hellinga and Fu (1999) also found that bias in the probe sample leads to a sample mean that does not asymptotically approach the population mean[2)]. The variability of urban network link travel time always shows to be high which makes it vary difficult to find optimal routes, and lots of approach has been developed to reduce the variability. Torday and Dumont (2003) reduced effectively the variance by using the differentiation method[3)].

One new problem of low frequency probe data application became in practice one of the most obstacles to get the precise aggregated travel time estimation. It is important to know what level of travel time reliability can be achieved from a given frequency probe data. Higher frequency data provide a higher accuracy in map-matching and individual link travel time experienced by each vehicle, but it may include some noises specific to each vehicle. Especially when the sample size is too small, the noise problem due to individual trip condition seems to be very important. Lower frequency probe data are expected to diminish such noises by averaging the travel times of successive links. It is hypothesized that the accuracy of route travel time estimation and prediction using low frequency may be not as bad as the individual link information obtained. This paper focuses on the variability of link-based travel time estimates for different frequency data.

2. Study arterial links and data collection

The test bed for this study was Hirokoji-Dori in Imaike area in Nagoya which is a main east-west urban arterial directly to Nagoya station. 9 DRM (Digital Road Map) links (figure 1) from west to east were selected due to including relatively more probe records, the whole length of which is 877 meters. There are all signals between every two adjacent links except between link 5 and link 6, because link 5 is a very short link on a main crossing.

Figure 1: Study arterial links

* Keywords: variability, link travel time estimates, probe frequency
** 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: +81-52-789-3565, 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: +81-52-789-3565, E-mail: yamamoto@civil.nagoya-u.ac.jp)

Among all the 1500 probe taxies in our P-DRGS project, 90 taxies with the highest distance based frequency (at the length interval of 50 meters) were selected for this research. The data were collected over a twenty-four hour period for six months from October, 2002 to March, 2003. Only the weekday evening travel time data were employed in this study because the AM peak and PM peak with complicated network traffic condition could be avoided. The study period was defined as lasting four hours from 20:00 to 24:00 and the data collected over this period were used for this study. The four-hour period was divided into sixteen fifteen-minute aggregation time intervals to ensure that a reasonable number of observations were obtained for each of the links. There are total 438 trips during this four-hour period, with an average of approximately ten to forty seven vehicles per fifteen minute period traversed the whole study links.

Similar to our previous studies on the examination of map-matching accuracy and individual link travel time accuracy, we still use only high frequency data (with the length interval of 50 meters in this study) for obtaining the same travel condition and avoiding external effects, from which the relatively seven kinds of low frequency data were simulated by removing parts of records evenly according to the required frequency. So we got total eight sets of data with the successively distance lag of 50m, 100m, 150m, 200m, 250m, 300m, 350m, and 400m.

3. Variability of route travel time

It has been known that the degree of probe reports variability is critical in assessing the reliability of the sample mean. Understanding variability of vehicle travel times can do help to understand traffic flow on arterial streets. The problems of probe sample size and great variability of link travel time owing to complicated delay time at signal intersection seem to be the main obstacles on precise aggregated travel time estimation. This section will focus on determining the accuracy level using low frequency probe data, with which an aggregated route travel time based on sampled data, can match the aggregated travel time experienced by the highest frequency data set. Advantages and disadvantages of low frequency probe data on route travel time estimation are analyzed. Figure 2 and figure 3 clearly shows the difference of individual link travel time mean among different frequency data.

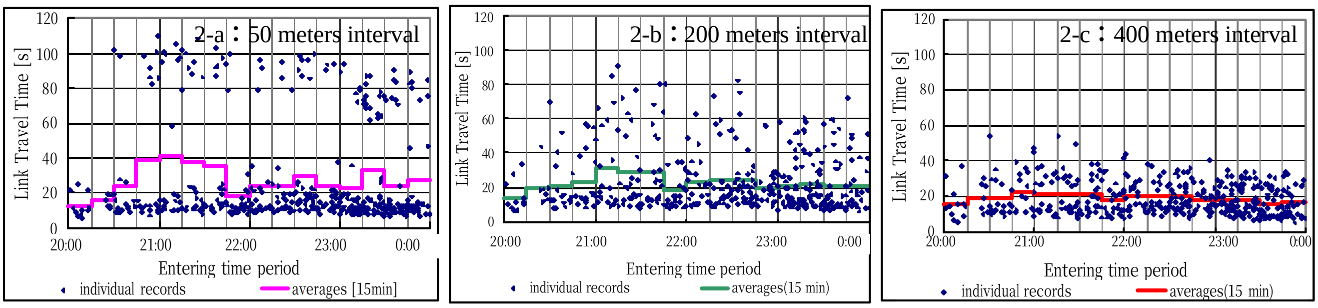

Figure 2: Individual and average travel time of Link 4 (the last link before the main signal)

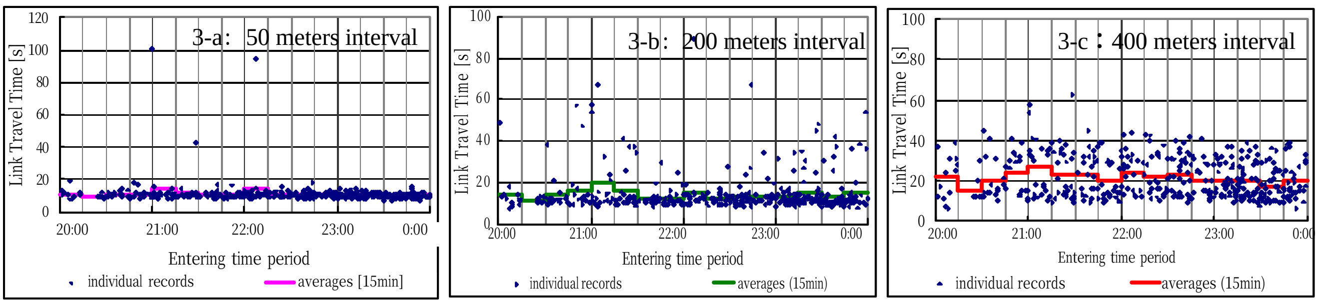

Figure 3: Individual and average travel time of Link 6 (the first link after the main signal)

Figure 2 gives one example (link4) of individual link travel time records provided by three kinds of frequency probe data with successively length interval as 50 meters, 200 meters and 400 meters. We can see from figure 2-a, reported from the probe with 50 meters interval, the probe can experience highly different individual travel time due to the unexpected red signal waiting time, even in the same condition of time period and the same upstream link and downstream link. Most of records ranged from 10 seconds to 25 seconds while others falling across red signal over 80 seconds, which result in during some periods the inexistent aggregated average travel times in practice. This contradiction can be smoothed using low frequency data, based on which the individual records congregated to around 20 seconds by evenly separating waiting time into adjacent links. Figure 3 shows a contradictory example (link6). All probes except only two went though this link rapidly without stopping; those records were all around 10 seconds. But the estimation using low frequency data exacerbate this variance between records because of containing some waiting time on adjacent links.

It may be noticed that the aggregated average travel times between the adjacent period using high frequency probe data has larger difference than those using low frequency probe data, and so does the variance, which are not the case in the real world. The noise problem owing to smaller sample size and consequently unexpected probe distribution is the major reasons of this difference. For example, the fact that the ratio of signal waiting probe and no waiting probe may be far different away the real ratio results in no signal waiting probe were recorded during the first two fifteen-minute periods. The standard deviation (indicator of variance) of high frequency probe is much higher than that of low frequency probe for link 4 in most periods(figure 4), while it is contradictory for link 6(figure 5). Compare of all 9 links showed that factors of relative position from the main signal crossing would affect this variance difference. Only link 5, link 6 and link 7 in this study showed clearly higher variance for low frequency probe which are the crossing link and the first and second link after the main crossing signal, while for other links the variance of low frequency probe data are relative lower in general.

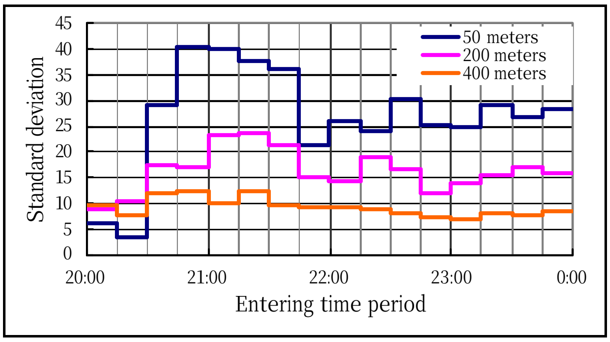

Figure 4: Travel time variance of Link 4

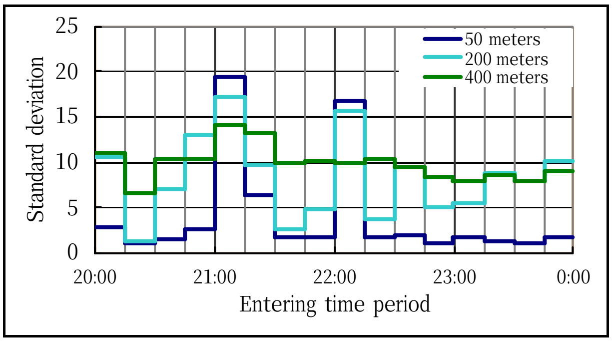

Figure 5: Travel time variance of Link 6

In fact, during the trip on all these 9 links, almost all probes experience one or two times of signal waiting, only few probes experience over three times of or no signal waiting. It is the difference on the stopping times and positions that leads to large difference on variance between different frequency probes and between different links.

An important practical issue then is to determine the route travel time estimation based on these link travel time estimates for all frequency probes can be achieved. In ITS applications, these route travel times are calculated by accumulating the average link travel time. Figure 6 showed the results of route travel time estimation for these three data sets, from which it is clear that before 23:00 the route travel time estimates by the highest frequency data are a little greater (ranging from 5 seconds to 20 seconds) than by the lowest frequency data and these values for three data sets after 23:00 are almost the same. The average standard deviation s of route travel time estimation for all 16 periods can be calculated by equation 1, where p means total numbers of periods; T_i means the actual route travel time reported by probe i; and $\overline{T}_j$ means the aggregated travel time mean for link j. Table 1 showed the results.

$$s = \frac{1}{p}\sum_{p}\sqrt{\sum_{i=1}^{n}\left(T_i - \sum_{j=1}^{9}\overline{T}_j\right)^2 / (n-1)} \quad (1)$$

It is clear that low frequency probe data can be used for route travel time estimation without accuracy descending too much compared to high frequency probe data, while a large amount of the transmitting cost can be saved.

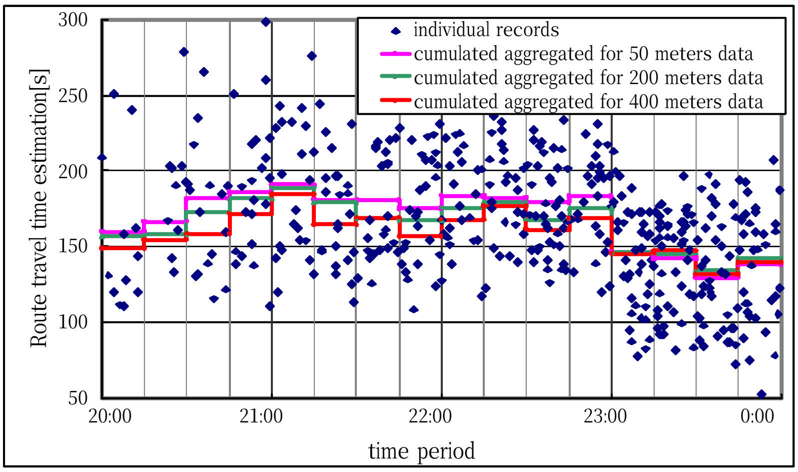

Figure 6: The aggregated average route travel times on link-based method

Table 1: Average standard deviation of route travel time estimation

Length interval	standard deviation
50m	40.1
100m	40.0
150m	40.2
200m	40.8
250m	41.1
300m	41.3
350m	41.9
400m	42.1

A very interesting and worthy advisement conclusion is that the mix data set of high frequency probe data and low frequency probe data might show relatively smaller variance than only high frequency data set without in fact improving the accuracy. That is to say only one indicator of variance is not enough to measure the accuracy.

5. Conclusion

The objective of this study is to examine the applicability of location information of taxis used for taxi dispatch control as probe information. The variance of sample mean has been used as the main criteria for identifying the value of low frequency data on link-based travel time estimation.

Although the high frequency probe data has the advantage of accurately reflecting the individual link traffic conditions, the low frequency data would support to obtain the route travel time estimation as accurate as high frequency data, which is more important than link travel times for travelers. But the low frequency probe data can still not be in applicable right now because of poorly map-matching which might be the main obstacle we should face with.

References

1) Sen, A., Thakuriah, P. Zhu, X.-Q. and Karr, A.: Frequency of probe reports and variance of travel time estimates, Journal of Transportation Engineering,Vol.123, No.4, pp.290-297, 1997.

2) Hellinga, B. and Fu L.: Assessing expected accuracy of probe vehicle travel time reports, Journal of Transportation Engineering, Vol.125, No.6, pp.524-530, 1999.

3) Torday, A. and Dumont, A.-G.: Link travel time estimation with probe vehicles in signalized networks, 3rd Swiss Transport Research Conference, Monte Verita / Ascona, March 19-21, 2003