

프로브 기반 교통정보 신뢰성 향상을 위한 평활화 기법 탐색

Exploring Smoothing Techniques for Reliable Travel-Time Information in Probe-Based Systems

장 진 환*

* 주저자 및 교신저자 : 한국건설기술연구원 도로연구소 수석연구원

Jinhwan Jang*

* Dept. of Highway Res., Korea Inst. of Civil Eng. and Building Tech.

† Corresponding author : Jinhwan Jang, jhjang@kict.re.kr

Vol.17 No.1(2018)

February, 2018

pp.79~88

ISSN 1738-0774(Print)

ISSN 2384-1729(On-line)

<https://doi.org/10.12815/kits.2018.17.1.79>

2018.17.1.79

Received 7 November 2017

Revised 1 December 2017

Accepted 19 December 2017

© 2018. The Korea Institute of Intelligent Transport Systems. All rights reserved.

요 약

하이패스 단말기의 확대 보급에 따라 DSRC 교통정보시스템이 지방부 도로를 중심으로 확대 설치되고 있다. 그러나 지방부 도로의 경우 고속도로와 달리 단말기 장착차량이 많지 않아 프로브 표본수가 충분하지 않는 경우가 종종 발생한다. 이 경우 프로브 통행시간은 적은 샘플수로 인해 많은 변동이 발생하고 이는 교통정보 오차를 증가시킨다. 본 연구는 부족한 샘플수로 인해 발생하는 통행시간 단기변동을 완화하여 신뢰성 있는 교통정보를 수집·제공하기 위해 프로브 통행시간 데이터에 이동평균, Loess, Savitzky-Golay 등 평활화 기법을 적용하였다. 그 결과, 낮은 샘플링(5%) 환경에서 통행시간 오차가 원시자료에 비해 45%까지 감소하는 결과를 보였다. 본 연구결과는 국내에서 운영 중인 프로브 기반 교통정보시스템에 적용되어 교통정보 신뢰도를 향상시킬 수 있을 것으로 기대된다.

핵심어 : 통행시간, 평활화, 이동평균, Loess, Savitzky-Golay 필터링

ABSTRACT

With the increasing popularity of electronic toll collection system using 5.8 GHz dedicated short-range communications (DSRC) technology, DSRC-based travel-time collection systems have been deployed on major urban and rural arterial routes in Korea. However, since probe sample sizes are frequently insufficient in probe-based systems, the gathered travel times from probe vehicles fluctuate significantly compared to those of the population; as a result, the accuracy of the collected travel times could decrease. To mitigate the fluctuations (also known as biases), smoothing techniques need to be applied. In this study, some smoothing techniques—moving average, the Loess, and Savitzky-Golay filtering—were applied to probe travel times. Resultantly, the error in the smoothed travel times at the lowest sampling plan (5%) decreased as much as 45% compared to those in non-smoothed travel times. The results of this study can be practically applied to probe-based travel-time estimation systems for providing reliable travel times along the travel corridor.

Key words : Travel time, Smoothing, Moving average, Loess, Savitzky-Golay filterin

I. INTRODUCTION

Travel-time information is an essential part of modern traffic systems for real-time traveler information or traffic monitoring. With this information, road users can divert their routes or rearrange their trip schedules, and road managers can develop and execute policies that enable the efficient alleviation of congestion on the road. In Korea, advanced traveler information systems (ATISs) have been steadily deployed across the nation since the early 1990s. As of 2016, travel-time collection devices have been deployed on every freeway, 20% of rural arterial routes, and urban arterial routes in more than 30 cities. Conventionally, the traffic detector—loop, video image, and radar—and/or automatic number plate recognition are used for collecting travel-time data in ATISs. However, low-cost probe systems that utilize dedicated short-range communications (DSRC) or the global positioning system (GPS) on smartphones are becoming popular.

However, due to the limited availability of probes, the number of samples is mostly insufficient in probe-based systems. In DSRC traffic information systems in Korea, probe rates only lie between 5-25% on rural highways. This low probe rate can cause travel times to fluctuate significantly compared to those of the population; which, in turn, can negatively affect the reliability of travel-time information provided(Eisele, 2012). In general, some forms of random variation are inherent to collected data taken over time. To reduce the variation by cancelling fluctuations, some statistical methods, also known as ‘smoothing techniques’ in industry, have been developed over the past few decades. Smoothing techniques, when appropriately applied, result in a more clear representation of the underlying trend(NIST, 2016). Hence, for probe-based travel-time systems to be effective, methodologies to resolve the problem of short-term variations need to be explored and their utility needs to be identified as well.

Some previous studies to estimate reliable travel times using probe-based systems have been performed. The Southwest Research Institute(SwRI 1998) developed the TransGuide algorithm, which eliminates outliers outside the predefined valid range. Clark et al.(2002) suggested a statistical method for cleaning outlying observations included in probe travel-time data gathered from license number plates. The traditional confidence interval concept with the median and quartile deviation was used. Dion et al.(2006) proposed a technique to estimate link travel times reliably. Their algorithm is based on the assumption that probe travel time follows a log-normal distribution. ITS Korea et al.(2008) developed a method that uses a simple confidence interval concept with an assumption that probe travel time follows a normal distribution. Ma et al.(2010) proposed a median filter approach that uses the median as a measure of location. Boxel et al.(2011) developed an innovative method of removing probe data gathered using Bluetooth scanners. They used the confidence interval concept (Kendall et al., 1973) using the standard residual of Greenshield’s speed-density model. They used the least median of squares to estimate parameters in the model to prevent it from being compromised by outliers. Eisele(2012) applied Loess statistical smoothing to estimate section travel times gathered using probe vehicles. Jang(2016) developed algorithms for filtering outliers in DSRC probe travel times on signalized rural arterial routes. His algorithm considered low sample size situations where previous methods had been proven not to operate effectively. Csiszar et al.(2016) estimated bus dwell times using multi-variate analysis followed by prediction of times based on factors describing certain situations.

Most previous studies focused mainly on filtering outliers to estimate reliable travel times with probe data. However, as stated above, the short-term bias caused by insufficient sampling is an imperative issue to be carefully

addressed although all the outliers were successfully removed. In this study, widely-recognized statistical smoothing techniques—moving average, Loess, and Savitzky-Golay filtering—to mitigate short-term bias are suggested. The performances of each technique are compared and investigated. The proposed techniques are outlined in the next section and are followed by a description of the data gathered in the field. Next, the techniques are applied to the field data and their performance characteristics are analyzed in detail. Finally, the author finish with principal contributions of this study and follow-up studies for practical applications of the consequences of this research.

II. SMOOTHING TECHNIQUES

1. Moving Average

A moving average (MA) filter smooths data by replacing each data point with the average of the neighboring data points defined within the span. This process is equivalent to low-pass filtering with the response of the smoothing given by the difference equation (1) (Mathworks, 2002).

$$y_s(i) = \frac{1}{2N+1} \{y(i+N) + y(i+N-1) + \dots + y(i-N)\} \quad (1)$$

where $y_s(i)$ = the smoothed value for the i -th data point, N = the number of neighboring data points on either side of $y_s(i)$, and $2N+1$ = the span for calculating the moving average.

The two fundamental and widely used MAs include a simple moving average, which is the simple average of values over a predefined span, and an exponential moving average, which places greater weight on more recent observations. MAs are most commonly applied to identify the fundamental trend and to resist any forms of unexpectedly short-term variations.

2. Loess

The name of ‘Loess’ is derived from the term ‘locally weighted scatter plot smooth,’ as it uses locally weighted linear regression to smooth observations. The smoothing procedure is performed locally because each smoothed value is determined by neighboring values lying within the predefined span. The procedure is weighted since a regression weight function, as shown in equation (2), is defined for the data points contained within the span (Mathworks, 2002).

$$w_i = \left[1 - \left| \frac{x - x_i}{d(x)} \right|^3 \right]^3 \quad (2)$$

where x = predictor value associated with the response value to be smoothed, x_i = the nearest neighbors of x as defined by the span, and $d(x)$ = distance along the x -axis from x to the most distant predictor value within the span.

The greatest advantage of the Loess procedure over many other methods is the fact that it does not require the

specification of a function to fit a model to all of the data in the sample. Instead, the analyst only has to provide a smoothing parameter value and the degree of the local polynomial. In addition, the Loess procedure is very flexible, making it ideal for modeling complex processes for which no theoretical models exist. These two advantages, combined with the simplicity of the method, make the Loess procedure one of the most attractive of the modern regression methods for applications that fit the general framework of least-squares regression but have a complex deterministic structure(NIST 2016).

3. Savitzky-Golay Filter

Savitzky and Golay(1964) defined a family of filters that are suitable for smoothing and/or differentiating sampled data. The data are assumed to be taken at equal intervals. The smoothing strategy is derived from the least-squares fitting of a lower order polynomial to a number of consecutive points. For example a cubic curve that is fit to five or more points in a least-squares sense can be viewed as a smoothing function. For instance, the data consists of a set of n (x_j, y_j) points ($j = 1, \dots, n$), where x_j is an independent variable and y_j is an observed value. They are treated with a set of m convolution coefficients, C_i , according to equation (3).

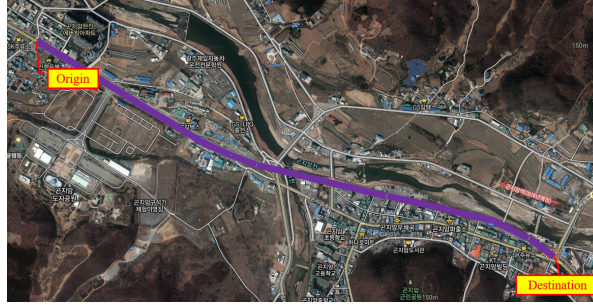
$$Y_j = \sum_{i=-\frac{m-1}{2}}^{\frac{m-1}{2}} C_i y_{j+i}, \quad \frac{m-1}{2} \leq j \leq n - \frac{m-1}{2} \quad (3)$$

where Y_j = the smoothed value for the j -th data point and m = the span for calculating the smoothing value.

The Savitzky-Golay (SG) filter is a digital filter that can be applied to a set of digital data points for the purpose of smoothing the data, that is, to increase the signal-to-noise ratio without greatly distorting the signal. This is achieved, in a process known as convolution, by fitting successive sub-sets of adjacent data points with a low-degree polynomial by the method of linear least squares. When the data points are equally spaced, an analytical solution to the least-squares equations can be found, in the form of a single set of "convolution coefficients" that can be applied to all data sub-sets, to give estimates of the smoothed signal, (or derivatives of the smoothed signal) at the central point of each sub-set(Guest, 2012).

III. DATA COLLECTION

Real-world data for this study were obtained by matching license plate numbers observed by field surveyors at two points on National Highway 3 in the vicinity of Seoul. The survey was conducted for one day (Monday) over a span of six hours (6:00 a.m. to noon) in September, 2015, and the license numbers of vehicles passing in the outer lane were recorded for convenience. Surveyors could not easily record the plate numbers of vehicles passing in the inner lane due to number plate occlusions. The section with 80 km/h speed limit and two lanes in each direction spans some 2 km of level terrain. Several intersections exist along the corridor. All these features represent typical rural highway in Korea. <Fig. 1> shows an aerial photograph of the study area on which indications of origin, destination, and route were superimposed.

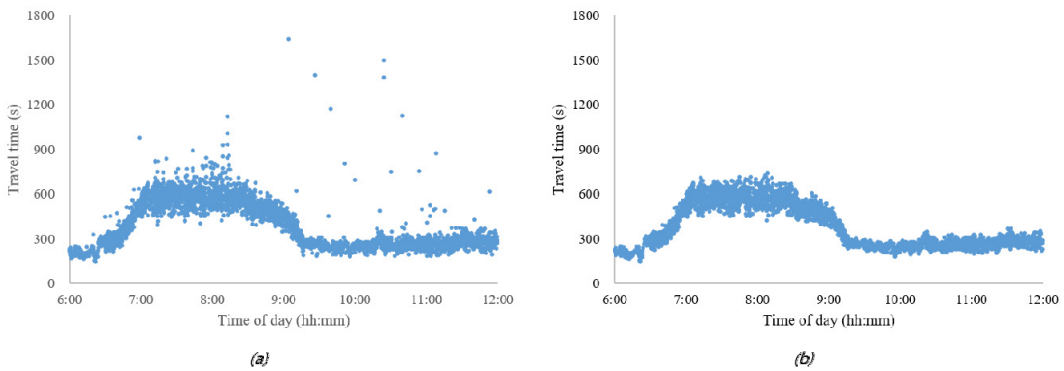


<Fig. 1> Study site

Around 4,000 matched probe travel times out of 5,312 vehicles observed at the destination point for six hours show congestion during the morning peak hours. The travel times (600 s) of congested flow are approximately three times higher than those (200 s) of free flow. Some outlying observations caused by exit/entry maneuvers on the corridor were removed using the Ferguson statistical test as expressed in equation (4). The individual probe travel times before and after outlier removal using the Ferguson test were plotted in <Fig. 2> The detailed procedure of the Ferguson test for removing outlying observations can be found elsewhere(Jang, 2016).

$$\sqrt{b_1} = \frac{\sqrt{n} \sum_{i=1}^n (x_i - \bar{x})^3}{\left[\sum_{i=1}^n (x_i - \bar{x})^2 \right]^{\frac{3}{2}}} \tag{4}$$

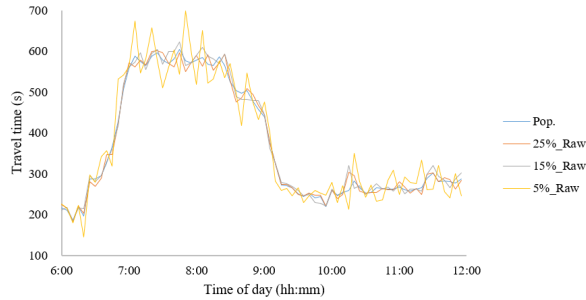
where $\sqrt{b_1}$ = statistic [refer to Ferguson(1961)], x_i and \bar{x} = observed values and the man.



<Fig. 2> Individual probe travel times: (a) before and (b) after removing outliers

To investigate the afore-mentioned travel-time estimation techniques, some samples were extracted from the population illustrated in <Fig. 2> Considering probe sample rates in DSRC traffic information systems in Korea, three sampling plans (25, 15, and 5%) using the simple random sampling scheme were adopted. Also, since most of the travel-time information systems around the world use 5-min aggregation interval, the sample data were extracted from the population aggregated into five minutes. <Fig. 3> shows the 5-min aggregated raw travel-time

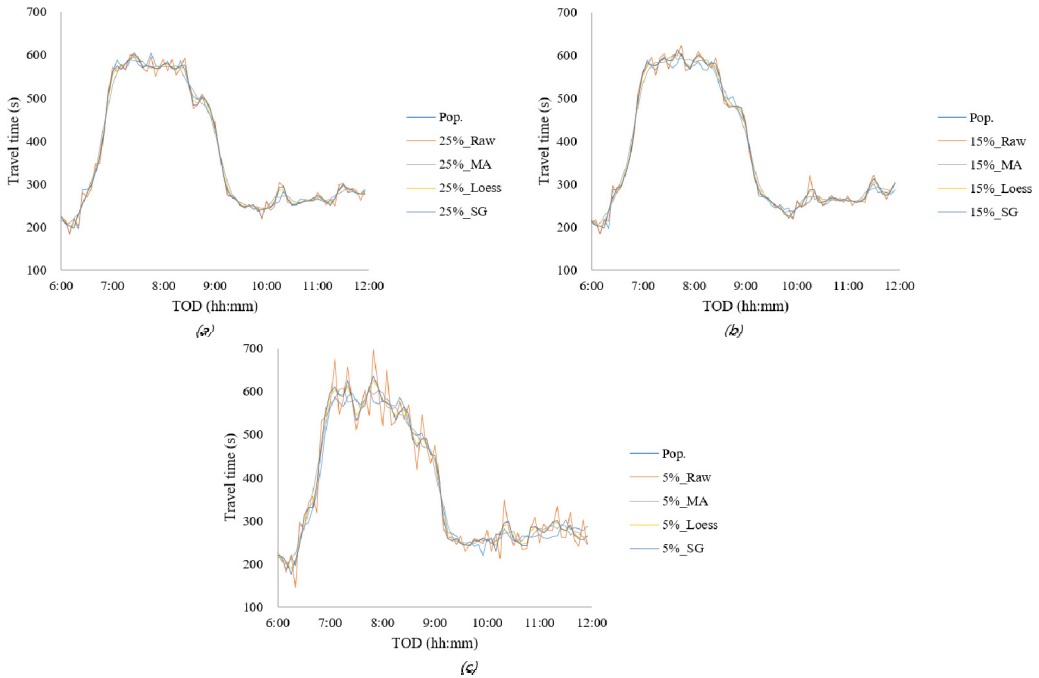
patterns of the three plans. As probe rates decrease, short-term biases (or fluctuations) become more significant; this phenomenon is more apparent in 5% sampling plan.



〈Fig. 3〉 5-min aggregated travel-time patterns

IV. APPLYING ESTIMATION TECHNIQUES TO PROBE TRAVEL TIMES

The three afore-mentioned smoothing techniques were applied to the 5-min aggregated travel times from the three sampling plans. Matlab was used for the application for efficiency and the span of seven consecutive data points that result in the best performances of each technique was chosen.



〈Fig. 4〉 Application of smoothing techniques: (a) 25%, (b) 15%, and (c) 5% sampling plan

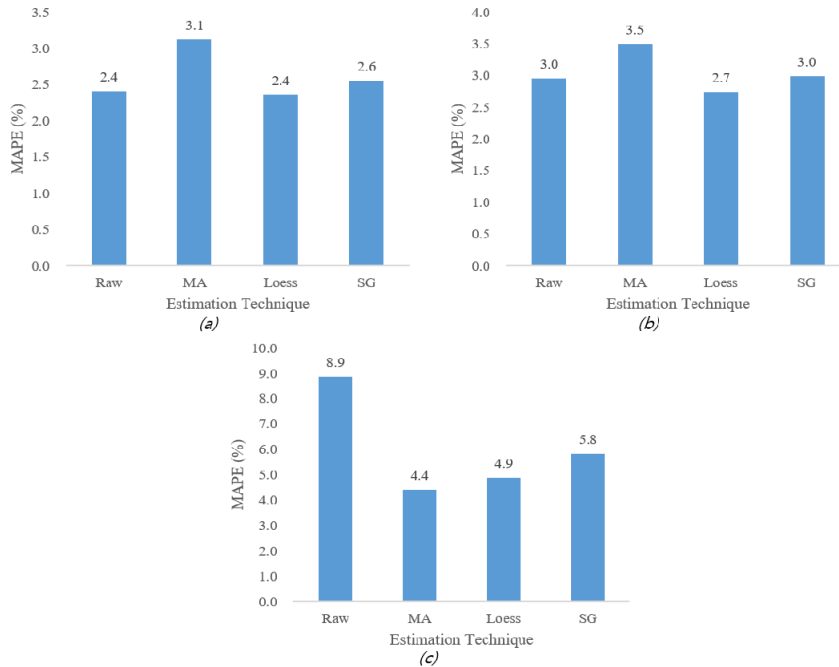
<Fig. 4> shows the 5-min aggregated travel times of population, raw samples, and samples smoothed by each of the three techniques (MA, Loess, and SG filter) in each sampling plan of 25, 15, and 5%. Compared to the 25 and 15% sampling plans, the raw travel times of the 5% sampling plan exhibit substantial fluctuations measured by coefficient of variation—3.0, 3.6, and 9.6% for 25, 15, and 5% sampling plan, respectively for a span of 7-8 a.m. period. However, this aspect was not observed in the smoothed travel times of the same sampling plan, indicating that the smoothing effects are probably conspicuous in low sampling conditions where travel times vary significantly compared to the population ones. Also, the smoothing effects are more apparent under congested flow conditions where greater emphasis is placed in terms of the real-time traveler information perspective. From this phenomenon, it should be stressed that an application of smoothing techniques is desirable to provide more accurate travel-time information in probe-based travel-time collection systems, especially under lower sampling rate conditions.

V. NUMERICAL ANALYSIS

To analyze the effects of the application of the smoothing techniques numerically, errors calculated against the population travel times were compared to the raw travel times of each sampling plan. A widely-recognized evaluation index of mean absolute percent error (MAPE) expressed in equation (5) was used for evaluation.

$$MAPE(\%) = 1/n \sum_{i=1}^n |x(t) - \hat{x}(t)| / x(t) \times 100 \tag{5}$$

where MAPE= mean absolute percent error; n= number of samples; x(t)= population travel time; and $\hat{x}(t)$ = probe travel time produced by each scheme (raw and smoothed travel times).



<Fig. 5> Percentage error in 5-min aggregated travel times : (a) 25%, (b) 15%, and (c) 5% sampling plan

<Fig. 5> shows the results of the evaluations on the raw and each smoothing technique under the three sampling conditions. Under the 25 and 15% sampling conditions, complicated consequences were revealed. In other words, the performances were similar, so apparent differences in percentage errors could not be easily discerned although Loess showed the best performances. The errors of the 5% sampling condition, however, were somewhat different from those of the 25 and 15% sampling plans. The raw sample exhibited much larger error compared to the errors from all the smoothing techniques. The error in the smoothed travel times at the lowest sampling plan (5%) decreased as much as 45% compared to those in the non-smoothed (raw) travel times.

<Table 1> Paired t-tests on differences in travel time errors

Sampling plan	Statistics	Observed	Smoothed		
		Raw	MA	Lowess	SG
25%	Mean (%)	2.41	3.12	2.36	2.55
	Std. dev.(%)	6.35	10.45	9.21	10.71
	Sample size	72	72	72	72
	t-statistic	na*	-1.48	0.14	-0.34
	Critical t-statistic at 5% sig. (one-sided)	na*	1.67	1.67	1.67
	p-value at 5% sig. (one-sided)	na*	0.07	0.44	0.37
15%	Mean (%)	2.95	3.50	2.74	2.99
	Std. dev.(%)	9.67	10.35	10.25	11.89
	Sample size	72	72	72	72
	t-statistic	na*	-1.14	0.55	-0.11
	Critical t-statistic at 5% sig. (one-sided)	na*	1.67	1.67	1.67
	p-value at 5% sig. (one-sided)	na*	0.13	0.29	0.45
5%	Mean (%)	8.88	4.41	4.88	5.82
	Std. dev.(%)	38.18	14.30	13.42	18.04
	Sample size	72	72	72	72
	t-statistic	na*	6.11	5.60	3.97
	Critical t-statistic at 5% sig. (one-sided)	na*	1.67	1.67	1.67
	p-value at 5% sig. (one-sided)	na*	2E-8	2E-7	9E-5

* Not available.

The phenomena shown in <Fig. 5> were numerically verified by paired t-tests, as described in <Table 1> The one-sided t-tests on the differences in errors obtained by the different sampling schemes at the 5% significance level show that no differences exist in the 25 and 15% sampling plans, but they do exist in the 5% sampling plan. The statistics (t-values) for the 25 and 15% sampling plans are lower than the critical t-value of 1.67, and those for the 5% sampling plan are higher than the critical t-value; p-values for the 25 and 15% sampling plans are higher than the critical value of 0.05, and those for the 5% sampling plan are significantly lower than 0.05. In addition to the improvement of average errors, the standard deviations decrease significantly in the 5% sampling plan, indicating that travel-time variation (or reliability), along with accuracy, can be enhanced through the application of smoothing techniques on probe-based travel times. The data analyzed in <Table 1> were sampled from around 4,000 matched probe travel times gathered manually for six hours including peak and non-peak hours. The sample

size of 72 is a consequence of the aggregation in the unit of five minutes for replication of real-world travel-time information systems. Hence, the implication of the statistics represented in <Table 1> should be emphasized. Although MA showed highest performance in the 5% sampling plan, it should be noted that Loess showed the stablest performance in terms of an average error of the three sampling plans—3.7, 3.3, and 3.8% for MA, Loess, and SG filter, respectively.

VI. CONCLUSIONS AND FUTURE STUDIES

Thanks to the development of wireless technologies, probe-based travel-time collection systems are anticipated to be deployed in wider areas worldwide. Compared to traditional point detectors, probe-based systems can obtain direct link travel times. However, under the condition of low sampling rates, the travel times can fluctuate significantly. Hence, producing travel-time information by simple aggregation of probe travel times to a certain time interval can deteriorate the quality of the processed data. To resolve this problem, statistical smoothing techniques need to be applied.

To identify the short-term biases in probe-based travel times, sample travel times by the simple random sampling scheme were extracted from the population travel times surveyed by license plate matching technique on a 2 km-long suburban arterial route in the vicinity of Seoul, Korea. With consideration of the real-world probe rates in DSRC traffic information systems in Korea, 25, 15, and 5% sampling plans are selected. Compared to the 25 and 15% samples, the 5% sample indicated significant fluctuation (or short-term bias) in travel times; the effect of the application of the smoothing techniques was also more conspicuous. In the other sampling plans, the difference in the errors of the raw and smoothed travel times did not show any statistical significance when analyzed by paired t-tests. It should be noted that one does not have to stick to the specific numbers of 25, 15, and 5%. The findings in this study should be considered as a general principle that low probe rates could cause a substantial short-term bias that can be mitigated with the application of smoothing techniques.

Outlier treatment techniques have been mostly investigated to estimate link travel times using probe data previously; short-term bias was inevitable. In this study, a negative effect from the bias was thoroughly investigated and methods to mitigate the bias were explored. As a consequence, Loess, although lacks a statistical ground for 25 and 15% sampling plans, is recommended to apply to probe travel times for improving accuracy of travel-time information thanks to its stable performances throughout the sampling plans.

The next research step would be the real-world application of the techniques proposed in this study. To do that, it is necessary to conduct abundant case studies with numerous data under various conditions—different sampling plans, aggregation intervals, traffic volume and so on. Travel-time prediction with a delicate methodology is another important issue to be explored.

ACKNOWLEDGEMENTS

This work was supported by a grant (No. 16TBIP-C111209-01) from the Korea Agency for Infrastructure Technology Advancement (KAIA). This paper is a revised version of the paper submitted to 2017 ITS World Congress in Montreal.

REFERENCES

- Boxel D. V., Schneider W. H. and Bakula C.(2011), “An Innovative Real-Time Methodology for Detecting Travel Time Outliers on Interstate Highways and Urban Arterials,” *TRB 2011 Annual Meeting CD-ROM*, Washington D.C.
- Clark S. D., Grant-Muller S. and Chen H.(2002), “Cleaning of Matched License Plate Data,” *Transportation Research Record*, no. 1804, Transportation Research Board.
- Cleveland W. S. and Devlin S. J.(1988), “Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting,” *Journal of the American Statistical Association*, vol. 83, pp.596-610.
- Cleveland W. S.(1979), “Robust Locally Weighted Regression and Smoothing Scatterplots,” *Journal of the American Statistical Association*, vol. 74, pp.829-836.
- Csiszár C. and Sándor Z.(2016), “Method for analysis and prediction of dwell times at stops in local bus transportation,” *Transportation*, vol. 32, pp.302-313. DOI: 10.3846/16484142.2016.1190402.
- Dion F. and Rakha H.(2006), “Estimating Dynamic Roadway Travel Times Using Automatic Vehicle Identification Data,” *Transportation Research Part B*, vol. 40. no. 9, pp.745-766.
- Eisels W.(2012), *Estimating corridor Travel Time Using Point and Probe Detector Data (Implications for Emerging Intelligent Transportation Systems Data Sources and Performance measurement)*, Lambert Academic Publishing.
- Ferguson T. S.(1961), “Rules for Rejection of Outliers,” *Revue Inst. Int. de Stat. (RINSA)*, vol. 29, no. 3, pp.29-43.
- Guest P. G.(2012), *Numerical Methods of Curve Fitting*, Cambridge University Press, p.147.
- ITS Korea, Hitecom System and Aju University(2008), *Development of Practical Technology for DSRC Traffic Information System, Korea Expressway Corporation*, Korea Expressway Corporation.
- Jang J.(2016), Data-Cleaning Technique for Reliable Real-Life Travel Time Estimation (Use of Dedicated Short-Range Communications Probes on Rural Highways), *Transportation Research Record*, no. 2593.
- Ma X. and Koutsopoulos H.(2010), “Estimation of the Automatic Vehicle Identification Based Spatial Travel Time In-formation Collected in Stockholm,” *IET Intelligent Transport Systems*, vol. 4, Iss. 4, pp.298-306.
- MathWorks(2004), *Curve Fitting Toolbox User’s Guide*, Version 1.
- National Institute of Standards and Technology, NIST/SEMATECH e-Handbook of Statistical Methods, <http://www.itl.nist.gov/div898/handbook/>, accessed July, 2016.
- Savitzky A. and Golay M. J. E.(1964), “Smoothing and Differentiation of Data by Simplified Least Squares Procedures,” *Analytic Chemistry*, vol. 36, no. 8, pp.1627-1639.
- Soltész T., Kózel M., Csiszár C., Centgráf T. and Benyó B.(2011), “Information System for Road Infrastructure Booking,” *Periodica Polytechnica Transportation Engineering*, vol. 39, no. 2, pp.55-62.
- Southwest Research Institute(1998), *Automatic Vehicle Identification Model Deployment Initiative - System Design Document*, Texas Department of Transportation.