Dynamic Forecasting of Bus Path Travel Time: Simplified Stochastic Approach

Suyun An[a], Huynho CHANG[b], Young-ihn LEE[c]
[a,b,c] *Graduate School of Environmental studies, Seoul National University, Seoul, 151-741, Korea*
[a] *E-mail: anssue@snu.ac.kr*
[b] *E-mail: nettrek@hanmail.net*
[c] *E-mail: yilee@snu.ac.kr*

Abstract: This study aims for the development of methodology on the prediction of bus path travel time considering dynamic traffic flow condition. For the progress of this study, we classified a bus route data with three standards: length of target area section, and traffic flow character by days of week and by season. The methodology development progressed is based on the probability density function, and the primary result is as follows. First, in general, when using the historical average method, the prediction rate was higher than the following stochastic process. Second, the stochastic method shows high prediction accuracy during peak time. Third, during peak time on weekends, the stochastic method was better than historical average method. Forth, the stochastic method is more flexible to forecast when the traffic flow is fluid. Fifth, when using the stochastic method, the section of larger standard deviation value was more accurate than relatively smaller one.

Keywords: Dynamic Forecasting, Stochastic Method, Bus Information, Path Travel Time

1. INTRODUCTION

The development of Bus Information System (BIS) of Korea has been a quantum leap as the intelligent transportation system (ITS) developed. It provides bus information within the maximum limit of 15 seconds update frequency. Moreover, through the rapid smartphone propagation of the twenty-first century, the utilization of smartphone is increasing.

Two basic needs of travelers who use public transportation are expected arrival time from the origin to destination, and required path travel time.

The time requirement information service does not provides smoothly on the current BIS and therefore, only arrival time information based on real time is offered.

The latest information system of travel time expectation, including the navigation system is based on the historical average time or the instantaneous path travel time, which does not reflect dynamic characteristic of travel. For this reason, there is a limitation on forecasting the path travel time in bus system with the present method. Furthermore, in the practical traffic field, an advanced model for forecasting can be misunderstood and misapplied to hands-on workers who do not have expertise in advanced mathematical model.

There are three conditions that useful forecasting systems require. First, having low load factor to the operate system. Second, being simple to calculate parameter and easy to manage. Third, producing a high accuracy in prediction (error rate within 5%). This paper focuses on the development of forecasting methodology, which satisfies these three conditions using the stochastic methods.

2. LITERATURE REVIEW

There are many researches of bus information forecasting. Chien (2002) researched for prediction of bus-transit arrival time with two artificial neural networks. This research contains the algorithms that prevision for prediction error. Zhou (2012) researched on forecasting bus arrival time using participatory sensing of mobile phone and provided a cost effective method.

While many researches are focusing on arrival time prediction, there has been no research for dynamic bus information except for a study done by H. Chang (2010). This research developed a dynamic model based on the Nearest Neighbor Non-Parametric Regression to forecast multi-interval path travel times between bus stops.

Therefore, there remains a need for an effect method that considers dynamic path travel time in bus information system through in bus stops, buss, and bus information applications.

3. PATH TRAVEL TIME PREDICTION METHODOLOGY

Using the stochastic method for the path travel time algorithm is as follows:

Step1. Divide 24 hours to a certain time unit by sequences, which are considered as the bus operating time. Each of the sequence should have enough time to include the bus operating time.

Step2. Construct a path travel time data from the designated origin to the destination by a departure time of each sequence.

Step3. Calculate the average and the standard deviation of each sequence. Figure 1. shows the travel time deviation. In this graph, each of the data follows the standard deviation, which means that each sequence has a different character as each of it has a different deviation and mean value. Figure 2.

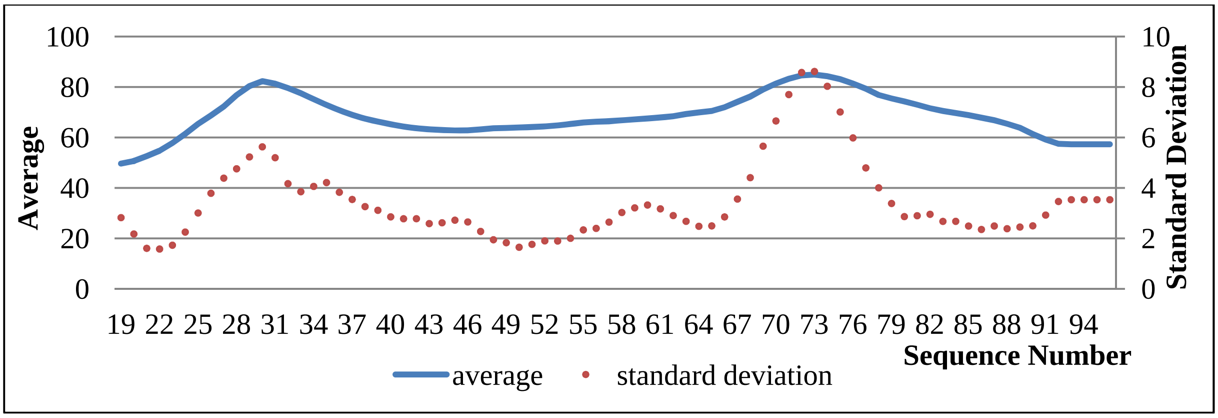

Figure 1. Temporal Variation of Average and Standard Deviation

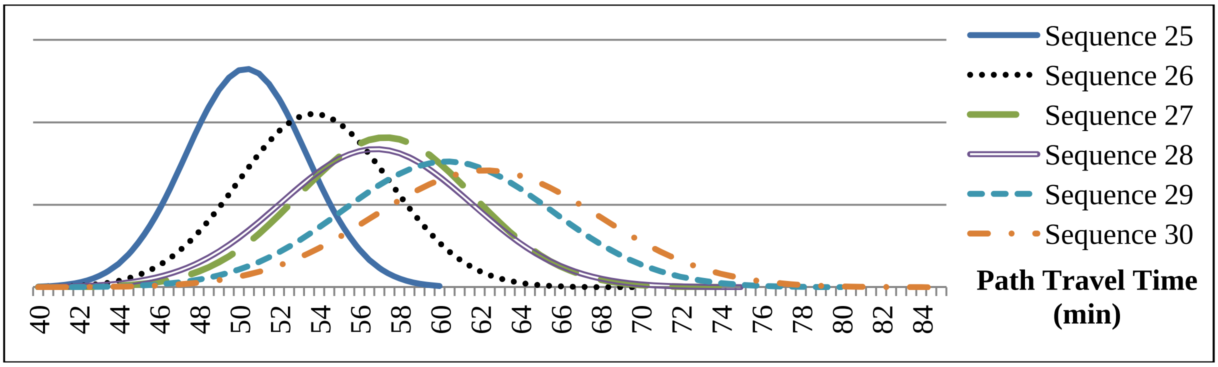

Figure 2. Temporal Variation of Distribution

Step4. Considering the character of each sequence, calculate each probability of path travel time by sequences using the integral calculus of probability density function, which is showed on Equation (1).

$$P(x) = \int_a^b \frac{1}{\sqrt{2\pi\sigma}} e^{\frac{1}{2}(\frac{x-\mu}{\sigma})^2} dx \tag{1}$$

where,

μ : mean value
σ : standard deviation
a, b : target path travel time

Step5. In the study, N is set as a number of sequences that are used for the prediction. The step for deciding the value of N is the optimal numbers of the application sequences that are needed, but in this study, we hypothesis N as 5 according to several tests. The forecasting is progressed based on optimal N of actual data. For example, a bus which departed at sequence “A” takes X1 minutes to pass the path. The X, mean the value of path travel time of sequences “A”, “A+1”, “A+2”, “A+3”, “A+4” is a clue to the forecast path travel time of sequence “A+5”. Described in terms of a mathematical equation of this step is Equation (2).

$$\frac{\sum_{i=1}^{N} X_i}{N} \tag{2}$$

where,

X : mean value
N : number of sequences

In this step, the data for forecasting path travel time is produced based on previous travel time, so the concept of “time lag” arises. In other words, a number of time sequences, “N” is produced, as “(A+4)×time unit” takes for pass the path.
To help understanding the conception of time lag is <Figure 3>.

Path Travel Time (min)	60	55	60	75	85	90	…	…	…	X	...

TT

Departure Time	06:00	06:15	06:30	06:45	07:00	07:15	07:30	...	08:45	09:00	...

DT

Figure 3. Conception of Time Lag

The key point in this step is the prediction value is changeable based on continuous changing time zones.

4. CASE STUDY

4.1 Data collection

Historical travel time of an intra-city bus route in Incheon, South Korea is used as test bed in this case study. In this paper, we processed data from two different angles: area and time. With subdividing the classes, test bed data was handled based on three criteria: length of target area section, and traffic flow character by days of week and by season.

To see accuracy difference of travel time forecasting between a long path and a short path, we divided target route: bus stop serial number 1 to 57 and 1 to 26.Additionally, we separated the data by its generation day of the week: weekdays and weekends. This study was progressed with 95 samples of weekdays: Monday to Friday, and 40 samples of weekends: Saturday and Sunday. In addition, the data sectionalized by seasons; August, which is generally regarded as a holiday season, and April, which is not. Thus, eight kinds of result were drawn in this study.

4.2 Data analysis

Since the test bed is operating 12-14 minutes interval by buses, we set up a time unit as 15 minutes, as we described in Section 3. After dividing 24 hours with 15 minutes, we have 96 sequences totally. Each sequence follows a normal distribution as we verified in Figure 3. The main conception of this paper is predicting path travel time of a bus with other travel times of which were generated short time ago in the same path. For this reason, some of travel time, which does not have enough previous data for forecasting, such as the first bus, only relies on the data of a prior day. This study deals with only 26 sequences travel time as data: A.M. 6:30 to P.M. 10:30 to minimize the possibility that can be ill-advised.

Equation (3) and Equation (4) are derived from the percentage of difference between the actual path travel time and the forecasting value. In this study, we used two methods for the prediction path travel time; the stochastic prediction value and the historical average; to compare and verify of accuracy between those two methods.

$$A = \frac{\sum_{1}^{n} \frac{|h-o|}{o}}{n} \times 100 \tag{3}$$

$$B = \frac{\sum_{1}^{n} \frac{|s-o|}{o}}{n} \times 100 \tag{4}$$

where,

- h : predicted value by historical average
- s : predicted value by stochastic method
- o : observed value
- n : number of samples

The results of each class by criterion are presented in <Figure 4> to <Figure 11>.

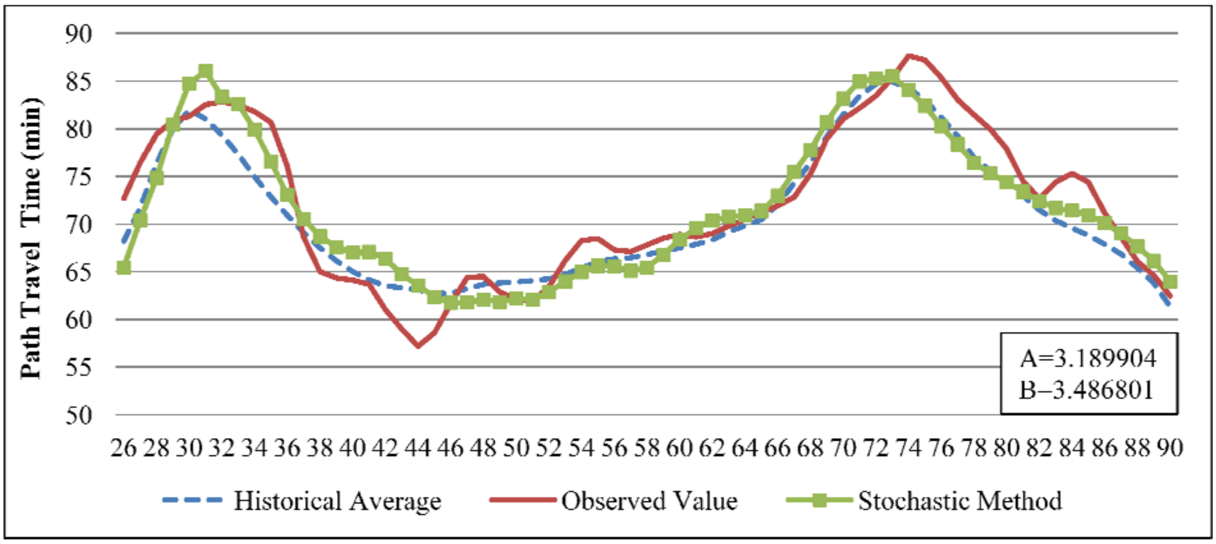

Figure 4. Long-distance Path, Weekdays, and Non-holiday Season Case

Figure 4 shows the data on the case of long-distance path, weekdays, and non-holiday season. Over all, the prediction parameter using the stochastic method is worse than using the historical average. The value of B is higher than A, and these value can be seen that the prediction gap was bigger when we applied stochastic method than historical average. It is estimated that the time lag is one of the reasons of this. It is likely that if the existence of the time lag is the main cause of the lower prediction parameter, the accuracy of forecasting using the stochastic method is higher as the bus running interval is shorter.

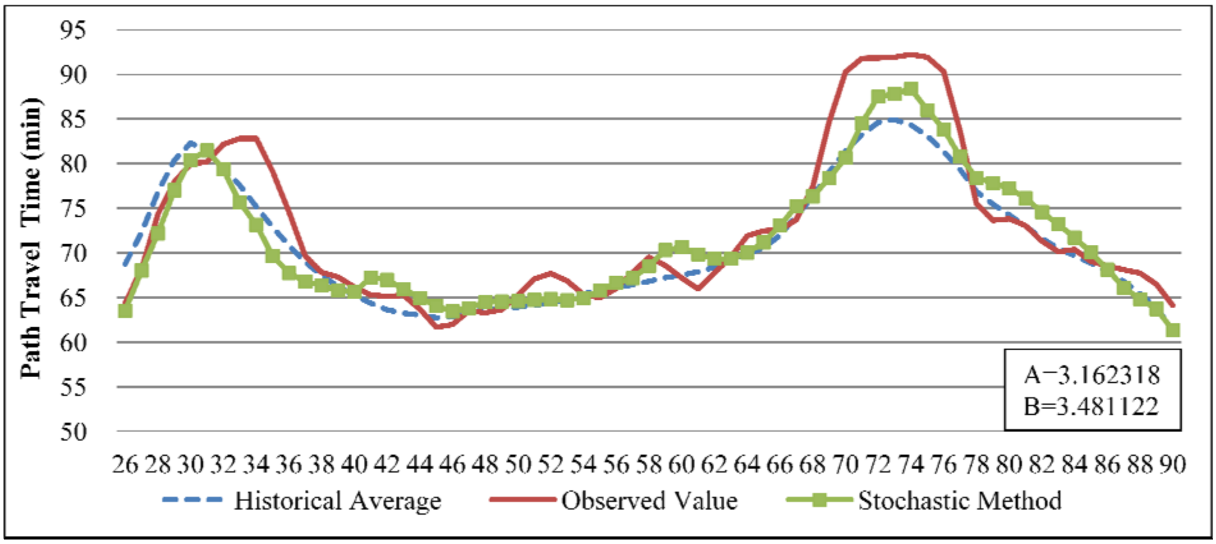

Figure 5. Long-distance Path, Weekdays, and Holiday Season Case

Figure 5 presents the data on the case of long-distance path, weekdays, and holiday season. In this case, the prediction parameter using the historical average was still little better than applying the stochastic method as we see the value A is slightly lower than of B. Notwithstanding, the prediction rate of using the stochastic method in this case was more accurate than non-holiday season with the same condition, except when the travel time gap between the sequences were conspicuously big. However, in cases when unusual traffic flow condition is continuing, during afternoon peak time for instance, the prediction rate of using the stochastic method was clearly higher than that of using the historical average method. This regards that it is able to forecast the travel time in dynamic traffic flow situation, when the pattern of travel time deviates from normal cases such as holiday season.

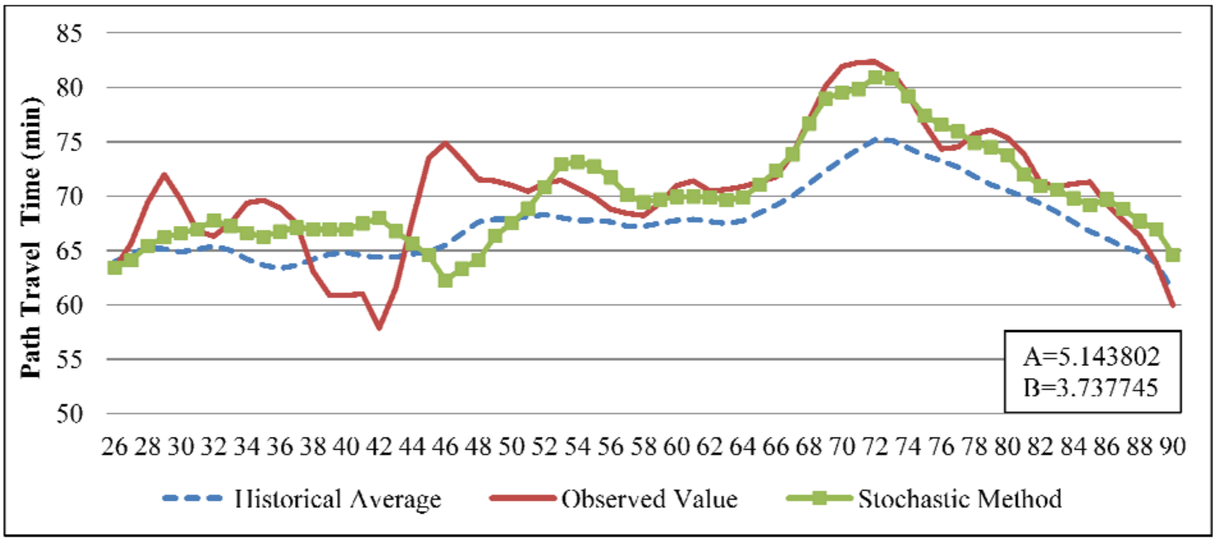

Figure 6. Long-distance Path, Weekends, and Non-holiday Season Case

Figure 6 reveals the data on the case of long-distance path, weekends, and non-holiday season. The prediction parameter in this case was generally worse than weekdays with the same condition. However, the prediction rate following the stochastic process was obviously higher than when using the historical average, peak time in particular. This case confirms that stochastic forecasting in fluid traffic flow situation might shows higher accuracy for prediction.

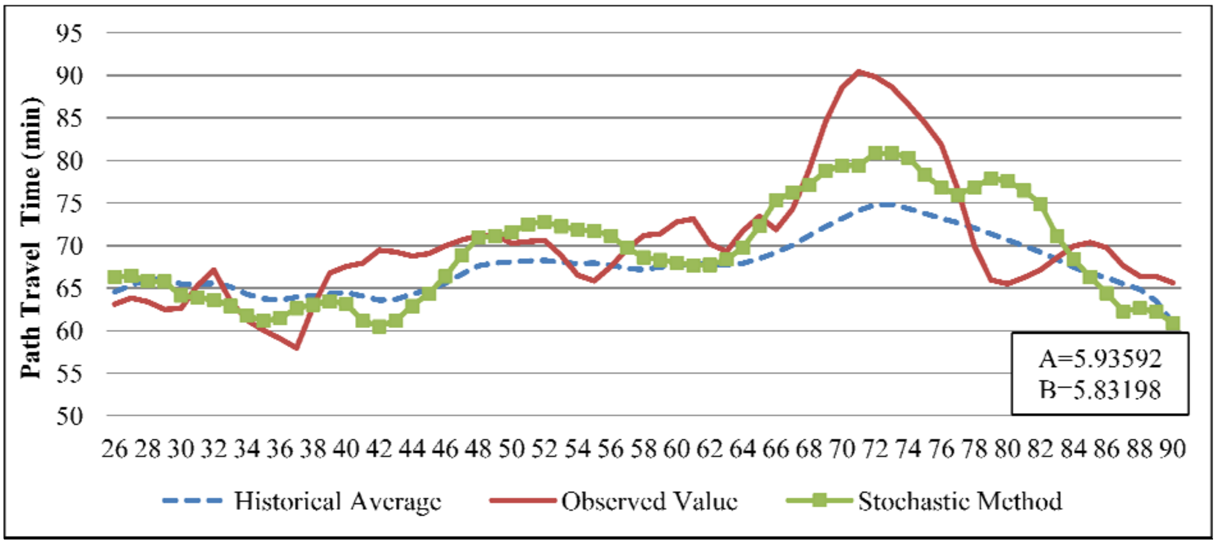

Figure 7. Long-distance Path, Weekends, and Holiday Season Case

Figure 7 reports the data on the case of long-distance path, weekends, and holiday season. Compared with the previous cases, the prediction rate with both methods was quite low in this case, but the prediction parameter of the stochastic method is slightly better than of the prediction parameter of historical average. This result can be analyzed with traffic flow characteristic of unusual condition such as weekends and holiday season.

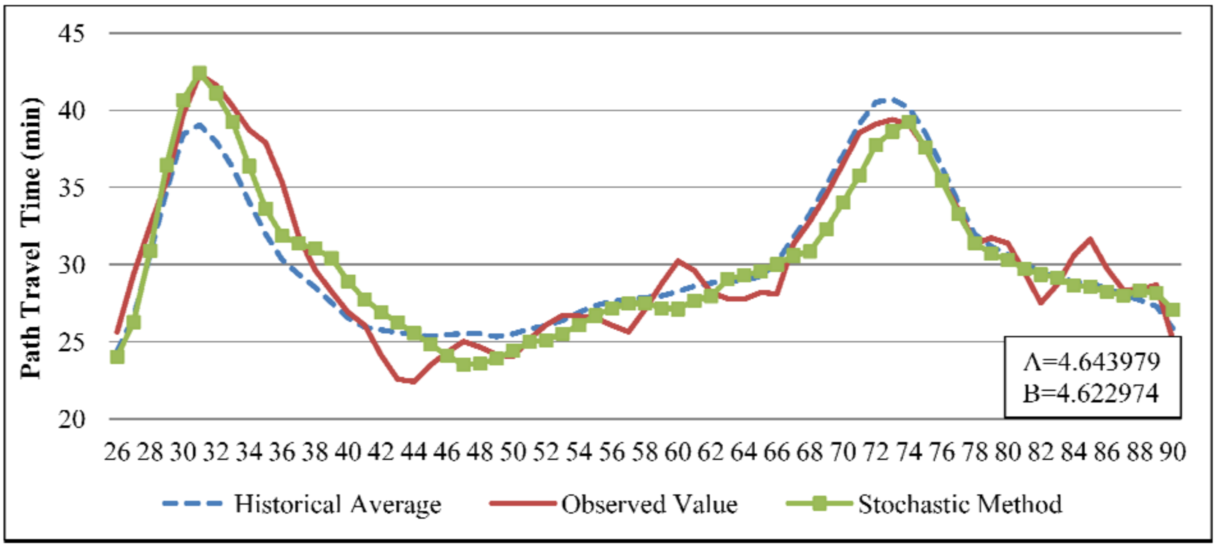

Figure 8. Short-distance Path, Weekdays, and Non-holiday Season Case

In cases of short path, the prediction rate was generally low, but the stochastic process method shows higher accuracy for prediction compared with the historical average method. Figure 8 illustrates the data on the case of short-distance path, weekdays, and non-holiday season. In this case, the prediction accuracy of both methods was similar. The stochastic process prediction was more accurate when only after few sequences of the travel time gap between sequences starts to swing and continue. Even though the graph has amplitude, if vibration does not stay out longer, the prediction was out of line.

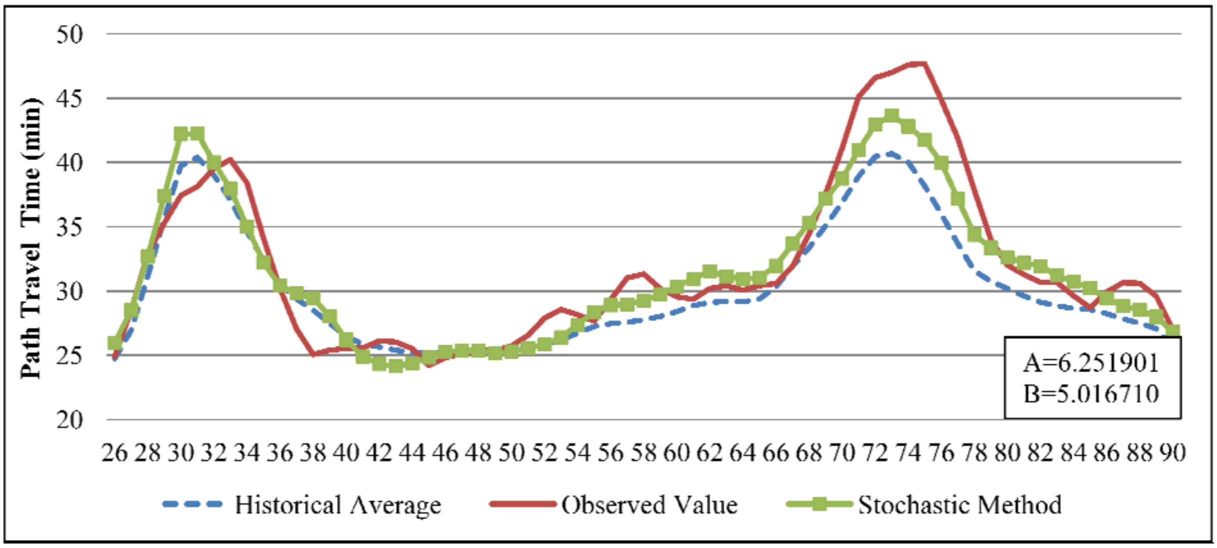

Figure 9. Short-distance Path, Weekdays, and Holiday Season Case

Figure 9 displays the data on a case of short-distance path, weekdays, and holiday season. The shape of this graph is not much different with the previous case, Figure 10. In spite of the prediction parameter of both methods in this condition were worse than non-holiday season with other same conditions, the stochastic method draw more accurate prediction than that of the historical method as we compared A value with B value.

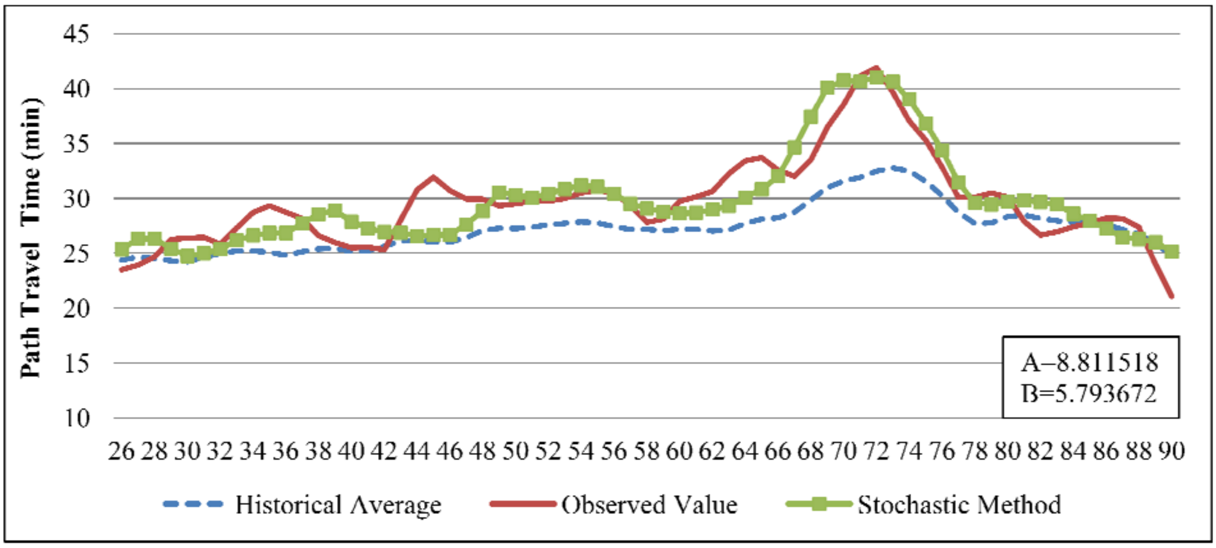

Figure 10. Short-distance Path, Weekends, and Non-holiday Season Case

Figure10 indicates the data on the case of short-distance path, weekends, and non-holiday season. Using both methods were very hard to forecast the path travel time except the afternoon peak time. The prediction accuracy in the following stochastic process is definitely higher than the prediction value of using the historical average method. The reason that this statue occurs seems to be due to the existence of time lag as well. In other words, because of the time lag, it does not show high accuracy when the traffic flow changes frequently.

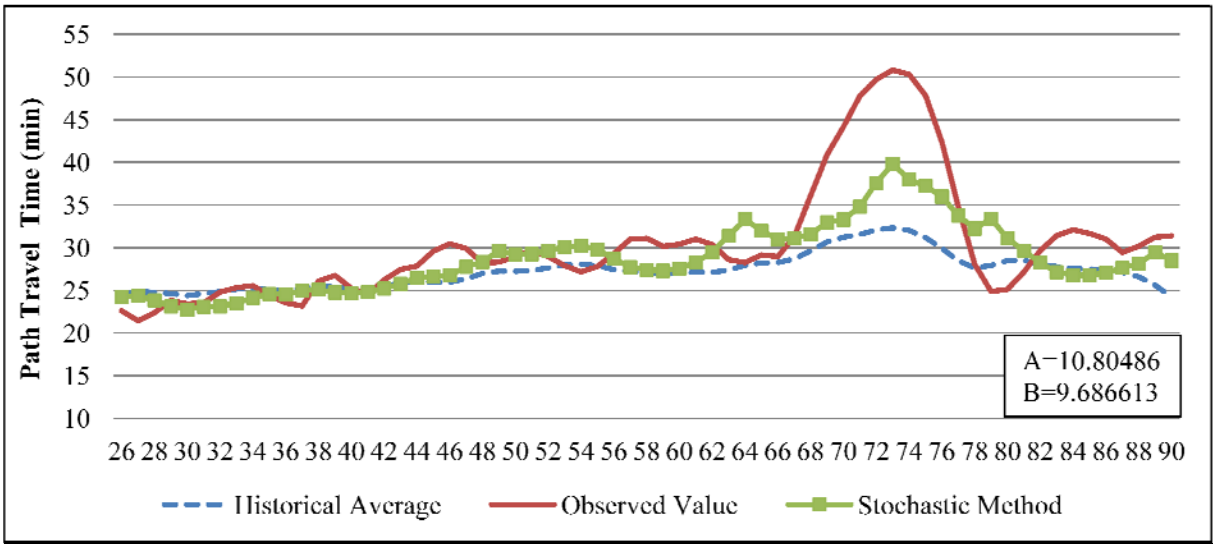

Figure 11. Short-distance Path, Weekends, and Holiday Season Case

Figure11 plots the data on the case of short-distance path, weekends, and holiday season. As it can be observed the value of A and B, this case presented the worst prediction parameter in total of this study. It might be caused by three conditions that causes comparative inferiority in this case, which are short-distance path, weekends, and holiday season.

5. CONCLUSION

The significant purpose of this study is finding a method which is simplified, lightening a load that is generated by bus information data, and forecasting a high accuracy travel time at the same time. In this paper, we have introduced a stochastic method for forecasting in dynamic traffic flow condition. The summary of the result of this paper is as follows.

First, a case of forecasting long-distance path, using historical average method tends to more accurate than using stochastic method, and the case of forecasting short-distance path is the complete opposite generally.

Second, during peak time, the stochastic process method shows more clear prediction accuracy than the historical average method. Forecasting of the stochastic method shows high predict rate when the tendency of fluctuation is continued. Nevertheless, when the path travel time by sequences have frequent changes, the prediction fall into disorder.

Third, following the stochastic process methods was more accurate than using the historical average method during peak time on weekends. However, during off-peak time on weekends, the prediction rate was very low.

Forth, compared to the historical average method, the stochastic method is more flexible to forecast when traffic flow condition is not ordinary, such as holiday seasons.

Fifth, when using stochastic method, the section of the larger standard deviation value was more accurate than relatively smaller standard deviation value.

In this paper, we set the number of N as 5 and the weight of each N equally, through several tests for set the N. However, this progress is not tested in this paper and therefore additional study, which accesses the optimal number of N and weight of it, is needed in the future.

REFERENCES

Zhou P., Zheng P., Li M., (2012) How long to wait: predicting bus arrival time with mobile phone based participatory sensing., *MobiSys*, 379-392

Chang H., Park D., Lee S., Lee H., and Baek S. (2010). Dynamic multi-interval bus travel time prediction using bus transit data., *Transportmetrica*, 6(1), 19-38

Chien, S., Ding, Y., and Wei, C. (2002). Dynamic Bus Arrival Time Prediction with Artificial Neural Networks. *J. Transp. Eng.*, 128(5), 429-48.