교통량예측모형의 개발과 평가

김 장 균

교통개발연구원

교통량 예측모형의 개발과 평가

김 창균

1995 년 2 월 28 일

TRAFFIC-FLOW-PREDICTION SYSTEMS BASED ON UPSTREAM TRAFFIC

Changkyun Kim

ABSTRACT

Network-based model were developed to predict short term future traffic volume based on current traffic, historical average, and upstream traffic. It is presumed that upstream traffic volume can be used to predict the downstream traffic in a specific time period. Three models were developed for traffic flow prediction: a combination of historical average and upstream traffic, a combination of current traffic and upstream traffic, and a combination of all three variables. The three models were evaluated using regression analysis. The third model is found to provide the best prediction for the analyzed data. In order to balance the variables appropriately according to the present traffic condition, a heuristic adaptive weighting system is devised based on the relationships between the beginning period of prediction and the previous periods. The developed models were applied to 15-minute freeway data obtained by regular induction loop detectors. The prediction models were shown to be capable of producing reliable and accurate forecasts under congested traffic condition. The prediction systems perform better in the 15-minute range than in the ranges of 30- to 45-minute. It is also found that the combined models usually produce more consistent forecasts than the historical average.

BACKGROUND

Several traffic management strategies were deployed to relieve traffic congestion in urban areas. They range from increasing road capacity to managing demand. A strategy for effective diversion of traffic during congestion-causing events, a major issue of Advanced Traffic Management Systems (ATMS), could maximize the use of available capacities in the roadway systems, thus saving travel time and minimizing congestion cost. However, when areawide diversion is attempted, pre-planning of diversion routes is frequently difficult due to the presence of multiple alternate routes. Once the alternate routes for diversion are determined, one or more routes must be selected at a particular point in the system to divert the traffic (Hobeika et al., 1992 and 1993). By utilizing real-time traffic information from roadway facilities, the traffic may be appropriately assigned to each candidate route. However, since traffic flow patterns change all the time, a control strategy based on the previous traffic flow pattern may no longer be good a few minutes later. For example, even if an alternate route selected for diversion is not congested at the current time, one part of the chosen route may be congested by the time the driver reaches that part of the network. Thus, forecasting future traffic flow variables for each link along the diversion routes is a required process for selecting the most efficient alternate route (Oda, 1990). If a prediction system representing the traffic flow-fluctuation over time could be developed, it would help produce reliable forecasts for solving such route guidance problem. In this paper, the term forecasting has been used interchangeably to the term prediction.

CONCEPT

Traffic conditions in a network vary considerably over time. In order to look ahead, a prediction algorithm that describes the time-varying traffic conditions throughout the network is needed. Here, three variables are investigated to assess the variability in traffic conditions. The current traffic information (called *current*) obtained from roadway sensors is one of the essential means of surmising the traffic trend and is a component of the prediction model (Ahmed and Cook, 1979). This requires effective communication between traffic control center and roadway detectors to obtain this reliable traffic information. Practically, data from four previous time-series with 15-minute interval serve as current traffic information. Next, the historical average (called *hist. avg.*) represents time-of-day pattern. This variable is used to smooth the abrupt changes of the current traffic flow and avoid extreme forecasts (Nihan and Holmesland, 1980). Finally, the upstream traffic, called *upstream*, is added as a new variable of the network-based prediction model. It is introduced to represent the dynamic nature of traffic flow. The quality of the network-based prediction model depends on the way those three variables are combined. The adaptive weighting system combines the three variables based on current traffic condition.

Traffic flow must be investigated throughout the whole network for more prompt and efficient control of traffic under congestion conditions. A traffic prediction model (Kim and Hobeika, 1993) that only relies on traffic information from a specific link may not detect changes in traffic elsewhere in the network, thus resulting in wrong traffic-control decisions. Therefore, a study of the relationships between the segments along the network is required to help forecast the traffic more accurately. A main uni-direction freeway is chosen as a source of data for the new prediction systems. It has been assumed that upstream traffic for short-term prediction largely affects the concerned segment on the downstream, called *study link*. Traffic volume is used to represent the traffic conditions on the roadway. Speed data at the present time is only used to identify the upstream traffic affecting the prediction for the study link. The upstream segments that affect the downstream *study link* are first identified based on current travel times. The upstream link (called *origin link*) that includes the traffic affecting the downstream *study link* is determined by taking 15 minutes current travel times between the two links.

Traffic data sets collected by the INformation FOR Motorists (INFORM) systems are utilized for developing and evaluating the prediction model, and examining the upstream traffic for purposes of the prediction. The traffic volume, speed, and occupancy at the Long Island Expressway (L.I.E./Rt. 495) in New York from Hauppauge to Kew Gardens (about 30 miles) were collected at 15-minute intervals for 22 weekdays during June, 1993. The 17 links separated by two miles were selected for this study, because of the long interval chosen (15 minutes).

MULTIPLE COMBINED PREDICTION MODELS

In order to conduct the short-term prediction, three combined models were developed. The first model combines two variables ; upstream traffic and historical average.

$$Z^{i}_{t+1}=\alpha(a_1 U^{i+k}_{t}+a_2 U^{i+(k-1)}_{t}+a_3 U^{i+(k+1)}_{t})+\gamma H^{i}_{t+1}$$

where

Z^{i}_{t+1} = Forecasts at time t+1 on link i

U^{i+k}_{t} = Observed traffic at time t on link i+k (origin link)

U^{i+k-1}_{t} = Observed traffic at time t on link i+(k-1) (adjacent link to origin link)

U^{i+k+1}_{t} = Observed traffic at time t on link i+(k+1) (adjacent link to origin link)

H^{i}_{t+1} = Historical average in time t+1 at link i

a_1, a_2, a_3, α, and γ = Weighting parameters

For the upstream component, three sub-variables are chosen in the above model. Since fluctuation of traffic may affect the links around the *origin link* at the location where it takes 15 minutes to reach the study link, the adjacent two links of the beginning link are also chosen as upstream traffic components. The parameter a_1 among the three parameters (a_1, a_2, and a_3) has the heaviest weights because of the exact travel time between the origin link and the study link (a_1=1/2, a_2=1/3, and a_3=1/6). The historical average is obtained from the previous days data. For instance, if the historical average for 8 A.M. in a specific day is to be obtained, the values at 8 A.M. for 21 previous days are averaged. In this manner, the historical averages are predetermined and updated.

In the second model, upstream and current traffic conditions are combined. Four previous terms during one hour (15-minutes interval) are utilized as the current traffic component. The amount of previous intervals required for the current component is really a matter of the sampling rate. The 'b' parameters for the 'current' component in the second and third prediction models are also intuitively obtained for the current component. Since the 'b_1' component contains the most recent time-series data at the study link, it has the largest portion among the four parameters. The weighting values of the remaining 'b' parameters are determined over time. The second model is shown as follows:

$$Z_{t+1}^{i}=\alpha(a_1U_t^{i+k}+a_2U_t^{i+(k-1)}+a_3U_t^{i+(k+1)})+\beta(b_1C_t^{i}+b_2C_{t-1}^{i}+b_3C_{t-2}^{i}+b_4C_{t-3}^{i})$$

where

C_t^i = Current traffic in time t at link i

C_{t-1}^i = Current traffic in time t-1 at link i

C_{t-2}^i = Current traffic in time t-2 at link i

C_{t-3}^i = Current traffic in time t-3 at link i

$a_1, a_2, a_3, b_1, b_2, b_3, \alpha$, and β = Weighting parameters

The third model contains all three variables, which are upstream traffic, historical average, and current traffic. The model is as follows:

$$Z_{t+1}^{i}=\alpha(a_1U_t^{i+k}+a_2U_t^{i+(k-1)}+a_3U_t^{i+(k+1)})+\beta(b_1C_t^{i}+b_2C_{t-1}^{i}+b_3C_{t-2}^{i}+b_4C_{t-3}^{i})+\gamma H_{t+1}^{i}$$

CRITERIA FOR COMPARISONS

MAPE, RMSE, Accuracy Ratio were used to compare in the prediction performance of the three models. The Mean Absolute Percentage Error (MAPE) is defined as

$$\text{MAPE}=\left(\frac{1}{S}\right)\sum_{t=1}^{S}\left|\frac{Z_t-\hat{Z}_t}{Z_t}\right|$$

where

Z_t = observed traffic data

$\hat{Z}_t$ = forecasts,

where S is the total number of forecasts made. The MAPE is a measure of the expected error associated with an individual forecast (Ahmed, 1989). The accuracy ratio (Q-ratio) is to measure the quality of the modeled values (Saha, 1990). The rules that applies for that ratio are:

If the observed value is greater than the modeled value, then Q = observed / modeled.

If the modeled value is greater than the observed value, then Q = modeled / observed.

The Root Mean Square Error is

$$\mathrm{RMSE} = \left[\frac{\sum_{t=1}^{S} (Z_t - \hat{Z}_t)^2}{S} \right]^{1/2}$$

COMPARISONS OF THE THREE LOOK-AHEAD MODELS

The three models were evaluated using regression analysis. Since the number of independent variables contained in the three models differ, the reliability of each model had to be examined. As a result of developing regression models, it is interesting to compare the parameters of the upstream component among the three models in the three prediction ranges. The parameter for the upstream component of the first model in the 15-minute forecasting is the highest, which means it has the largest influence on the prediction performance. In contrast, there is a very small amount of weighting value portioned in the upstream component of the second model. It is due to high reliance on the current traffic component. The upstream parameter has been somewhat increased in the third model. It indicates that combining the three variables plays some role in providing more weights on the upstream traffic component. The upstream parameters in the 30- and 45-minutes ranges showed same trend as in the 15-minute range.

The regression model forecasts are compared to one another in terms of MAPE, RMSE, and Q-ratio in Table 1. A comparison of the quality of the forecasts are also shown in Table 1. As expected, the third model has the lowest mean error, variance, and standard error in MAPE. The second model including the 'upstream' and 'current' variables are also better than the first model including 'upstream' and 'historical average'. The distribution of mean, that is, Standard Error (S.E.) of mean tells that models II and III are almost always better than model I. Considering the above comparisons, it is noted that the second and third models provide better forecasts. However, the fixed parameters obtained from regression analysis may not be adaptive to real-time application due to the incapability of updating the model structure according to changes in current traffic flow. Thus, it is necessary to build an adaptation that accommodates changes over time and traffic conditions.

Table 1. Comparison of three regression models

Criteria		M I	M II	M III
MAPE	Mean (%)	7.64	5.87	5.84
	S.E.	0.01	0.0061	0.0071
	Variance	0.012	0.0046	0.0061
RMSE	Mean	498.4	352.7	367.2
Q-ratio	Mean	1.0788	1.0609	1.0602

ADAPTIVE WEIGHTING SYSTEMS FOR REAL-TIME APPLICATION

Time-responsive weighting systems were developed to obtain the most appropriate parameters for real-time application of the prediction. Since three variables (*upstream*, *current*, and *hist. avg.*) are all time-dependent data, they will be influenced by the adaptive/time-responsive weighting system. All possible weighting scenarios for model I and II are shown in Table 2.

Table 2. All possible weighting scenarios for model I and II

Scenarios	*Current (β) / Hist. avg (γ)*	*Upstream (α)*
1	0.9	0.1
2	0.8	0.2
3	0.7	0.3
4	0.6	0.4
5	0.5	0.5
6	0.4	0.6
7	0.3	0.7
8	0.2	0.8
9	0.1	0.9

The two combined models generate nine sets of scenarios in terms of the fluctuation of traffic condition over time. The next step is to devise a methodology for selecting the best scenario for the traffic condition at the current time interval. Two decision factors emerge:

$$f_1 = \frac{|C^i_t - C^i_{t-1}| + |C^i_{t-1} - C^i_{t-2}| + |C^i_{t-2} - C^i_{t-3}|}{3}$$

$$f_2 = |H^i_{t+1} - C^i_t|$$

$$L = \left|\frac{f_1 - f_2}{f_1}\right|$$

$$M = \text{Standard deviation of } [\,|C^i_t - C^i_{t-1}|,\ |C^i_{t-1} - C^i_{t-2}|,\ |C^i_{t-2} - C^i_{t-3}|\,]$$

The first factor is the percentile error (L) between the differentiate of the beginning of the prediction and the historical average (f_2), and the average of the differentiate of the four previous terms (f_1). The basic concept of the percentile error (L) is to measure how the traffic volume has fluctuated over the assigned interval. It is based on the supposition that reliable forecasts can be determined by carefully investigating the fluctuation of the traffic flow over time. If the differentiates of the traffic volume data in each interval are small then the historical average term must be heavily weighted compared to the other terms. The scenario 1 in Table 2 presents the case where historical averages are heavily weighted (γ=0.9 and α=0.1). Conversely, if there exist huge percentile errors (as in scenario 9 in Table 2) a weight that relies heavier on the upstream component than the historical average/current components is more suitable for forecasting. The selection numbers in Table 2 have been incremented as the percentile errors increase.

The second decision factor (M) is the standard deviation of the differentiates between the previous four intervals including the beginning of the prediction. It is devised to detect unusual trends in the recent past. If there are huge variances in recent traffic flow, the upstream traffic will have greater influence than the historical averages/current traffic in terms of the prediction performance. Thus, the parameters (α, β, and γ) are determined according to the combination of decision factors as shown in Table 3.

Table 3. Adaptive weighting system for model I and II

Decision Factors	Combination of the decision factors					Scenario
L (%)	0-49					1
M	0-49					
L (%)	0-49	50-99				2
M	50-99	0-49				
L (%)	0-49	50-99	100-199			3
M	100-149	50-99	0-49			
L (%)	0-49	50-99	100-199	200-299		4
M	150-199	100-149	50-99	0-49		
L (%)	0-49	50-99	100-199	200-299	>300	5
M	200-299	150-199	100-149	50-99	0-49	
L (%)	0-49	50-99	100-199	200-299	>300	6
M	>300	200-299	150-199	100-149	50-99	
L (%)		50-99	100-199	200-299	>300	7
M		>300	200-299	150-199	100-199	
L (%)			100-199	200-299	>300	8
M			>300	200-299	200-299	
L (%)				200-299	>300	9
M				>300	>300	

The two decision factors are empirically harmonized in Table 3. The percentile error ranges (L) have been diagonally incremented while the standard deviation ranges (M) has been vertically incremented. The combinations of the two decision factors are diagonally deployed by the order of selection number. The difference between models I and II in applying this procedure is that the current component in the second model replaces the historical average in the first model. Meanwhile, the third model, with three variables to combine, has twelve sets of possible scenarios as shown in Table 4.

Table 4. Adaptive weighting systems for third combined model in 15-minutes forecasts

L (%)	M	Scenario	Upstream	H. A	Current
0-24		1	0.2	0.6	0.2
25-49	0-149	2	0.1	0.5	0.4
25-49	>150	3	0.4	0.5	0.1
50-99	0-99	4	0.1	0.4	0.5
50-99	100-199	5	0.3	0.4	0.3
50-99	>200	6	0.5	0.4	0.1
100-299	0-99	7	0.1	0.3	0.6
100-299	100-199	8	0.3	0.3	0.4
100-299	>200	9	0.5	0.3	0.2
>300	0-99	10	0.2	0.2	0.6
>300	100-199	11	0.4	0.2	0.4
>300	>200	12	0.6	0.2	0.2

The *upstream* and *current* parameters are determined based on comparison with the *hist avg* parameter. The *current* is more weighted in a lower numbered scenario than in higher model within a specific historical average parameter chosen. For example, look at model 4, 5, and 6 from Table 4. The fourth model (lower-numbered model) has 0.5 for the *current*, which is much larger than 0.1 for the *upstream*. The percentile error (L) and standard deviations (M) are computed to select the most appropriate scenario for the current traffic condition. In the farther forecasts (30- and 45-minute intervals), the method of developing the systems and of obtaining the combination are exactly the same as the ones in the 15-minute forecasts as well as in the structure of the systems. There, however, is a small difference in combining the parameters and the decision factors. Since the *upstream* and *current* variables lose their reliability in that farther prediction, the historical average parameters ranging from 0.4 to 0.2 occupy only five out of 12 scenarios (In comparison to nine out of 12 scenarios in the 15-minute forecasts). In other words, reliance on the historical average becomes larger in this range. There is a little modification in combining two decision factors corresponding to the new combination sets of parameters. The weighting systems in 45-minute forecasts show a little different trend from the systems in 30-minute forecasts. They are also based on the importance of historical averages in performing prediction. The remarkable change in 45-minute forecasts is the disappearance of a combination set with 0.2 of historical average parameters and the expansion of the sets with 0.4 of historical average parameters. Correspondingly, the order of decision factors are determined.

EVALUATION OF THE MULTIPLE COMBINED MODELS

The forecasts were calculated using the multiple combined models with adaptive heuristic weighting systems. The three upstream traffic, four current traffic, and/or historical averages at the concerned time period are the variables for the prediction. The resultant forecasts were compared to one another, as well as to the historical averages.

-With peak-hour traffic data

The peak-hour traffic data between 6 A.M. and 9 A.M. were selected for evaluation of the multiple models. One-hundred forty randomly selected data sets were used. The mean values of three models in the three criteria are all lower than the historical averages as shown in Table 5. As expected, the third model has the lowest value among the three models. Also, the standard error for the mean of the three models are lower than the historical average (Ott, 1988). Intuitively, model III in MAPE has prevailed historical average 99 times out of 140 chances (70 percent). MAPE at model I and II have beaten 91 times (65 percent) and 97 times (68 percent) out of 140 chances, respectively. By the distribution of means and the intuitive comparison, the second and third model present almost equal accuracy of prediction performance.

Table 5. Comparison of three models and historical averages in 15-minute forecasts using peak-hour traffic data between 6 A.M. and 9 A.M.

		M I	M II	M III	H.A.
MAPE	Mean	6.8	6.34	5.86	8.51
	S.E.	0.0076	0.0070	0.0075	0.0096
	Variance	0.0081	0.0068	0.0078	0.013
Q-ratio	mean	1.0695	1.064	1.0593	1.0874
RMSE	mean	414.55	385.01	376.95	518.48

There is no distinct difference between 30-minute forecasts and 15-minute forecasts as shown in Table 6. Indications are that 30-minute forecasting performs as accurately as 15-minute forecasts. Intuitively, MAPE of model III prevailed 75 percent over historical average, while models I and II prevailed 60 percent and 64 percent over historical average, respectively.

Remarkably different results emerged from the 45-minute forecasts in Table 7. The first model is the best among the three combined models. The second and third models are even worse than the historical average. This is due to the low reliability of current and upstream traffic values used in this farther prediction. The two variables are vulnerable because of unexpected ramp traffic and other interruptions between the *study link* and the *origin link*. The estimated, not observed, current traffic values from 15- and 30-minute forecasts also interfere with good prediction performance. Probably, the historical average alone is good enough to predict the traffic in that range.

Table 6. Comparison of three models and historical averages in 30-minute forecasts using peak-hour traffic data between 6 A.M. and 9 A.M.

		M I	M II	M III	H. A.
MAPE	Mean (%)	6.3	6.2	5.4	7.4
	S.E.	0.0065	0.0059	0.0064	0.008
	Variance	0.006	0.005	0.006	0.01
Q-ratio	mean	1.0639	1.0541	1.0432	1.0683
RMSE	mean	396.32	377.54	359.09	438.30

Table 7. Comparison of three models and historical averages in 45-minute forecasts using peak-hour traffic data between 6 A.M. and 9 A.M.

		M I	M II	M III	H. A.
MAPE	Mean (%)	5.68	6.42	6.03	5.70
	S.E.	0.0138	0.0143	0.0141	0.0130
	Variance	0.0082	0.0089	0.0086	0.0080
Q-ratio	mean	1.0577	1.0596	1.0593	1.0601
RMSE	mean	357.95	383.30	375.68	371.74

-Non-peak-hour traffic data

Next, the multiple combined models are evaluated using non-peak hour traffic data. First, the traffic data between 11 A.M. and 3 P.M. were randomly selected for evaluation of prediction performance. The results from 15-minute forecasts are shown in Table 8. All three criteria brought more accurate forecasts than historical average alone. The third model, of course, resulted in the best forecasts out of three models. The unique finding of Table 8 was that the first model predicted better than the second model,

Table 8. Comparison of three models and historical averages in 15-minute forecasts with traffic data between 11 A.M. and 3 P.M.

		M I	M II	M III	H. A.
MAPE	Mean (%)	6.79	7.30	6.09	9.53
	S.E.	0.0044	0.0033	0.0036	0.0049
	Variance	0.0027	0.0016	0.0018	0.0035
Q-ratio	mean	1.0691	1.0767	1.0625	1.1028
RMSE	mean	322.18	333.96	280.59	452.95

which is in variation to the peak-hour test. It may be due to the absence of major change in traffic in the non-peak hour compared to the usual traffic flow. The historical average component of the first model plays a bigger role in prediction than the current traffic of the second model. Regardless of the invaluable current traffic component, the prediction performances have been improved by combining current traffic with the other two variables (*hist. avg.* and *upstream*).

In the MAPEs and RMSEs of Table 9, the forecasts of historical averages are better than the ones of the first and second models. It is noted that the combination of two variables in the 30-minutes range with non-peak hour data does not improve prediction performance. The variances of the MAPE in three models are also poorer than the values of historical average. The Q-ratios of the first and second models has equal or less values than historical average. Combining three variables has again improved the prediction performance.

Table 9. Comparison of three models and historical averages in 30-minute forecasts with traffic data between 11 A.M. and 3 P.M.

		M I	M II	M III	H. A.
MAPE	Mean	8.09	9.74	6.65	7.90
	S.E.	0.007	0.0088	0.0078	0.0059
	Var.	0.0076	0.0109	0.0087	0.005
Q-ratio	mean	1.0815	1.0974	1.0665	1.0816
RMSE	mean	304.86	347.81	270.96	336.75

A MODEL SELECTION RULE FOR BETTER PREDICTION

Throughout the evaluations and comparisons of the three combined models, the third model generally seems to result in the best forecasts due to the many variables it comprised. However, the first or second model is often better than the third model. Current traffic conditions must be considered to select the best model. In order to accomplish that, monthly average travel times along the network were compared to the travel times under current traffic conditions. If a specific day shows a traffic condition different from the monthly average traffic condition, the traffic on that day has to be considered in a different way for better prediction. The rule of selecting a model in terms of current traffic conditions is as follows :

"If current travel time along the network is 25 percent longer than the monthly average travel times, use the second model for the prediction performance. Otherwise, use the third model for the prediction performance."

This rule is based on the idea that under congested traffic conditions current traffic plays a heavier role than upstream and historical average traffic. The threshold value--25 percent--is determined by looking over the predictions from the evaluation of the models throughout various sets of data. The peak-hour traffic data is applied to this rule for 15-minutes forecasts in Table 10. The error rate from final decision rule ("Rule." in Table 10) presents the best forecasts among five models including historical average, which means that the rule was effective in improving the prediction performance.

Table 10. Evaluations of the values by final decision in 15-minute forecasts with traffic data between 6 A.M. and 9 A.M.

		I	II	III	H. A.	Rule
MAPE	Mean	9.03	8.45	8.12	10.4	7.00
	S.E.	0.010	0.009	0.010	0.012	0.009
	Var	0.015	0.012	0.014	0.022	0.012
Q	mean	1.09	1.08	1.08	1.11	1.07
RMSE	mean	511.8	464.9	471.3	625.4	433.4

COMPARISON OF THE HEURISTIC COMBINED MODEL WITH OTHER FORECASTING MODEL

Until now, the forecasts from the combined model developed here have been compared with historical average only. It is due to the finding that the historical average is as good as traditional forecasting model in terms of prediction performance. Urban traffic control is also dependent upon the historical traffic data. Attributes of the links or network used are a consideration for the comparison between the models. While the heuristic combined model focuses on the mainstream traffic on the freeway, some other models study the ramp traffic. Different time interval is another interruption for that comparison. Beyond the above reasons, there is a practical difficulty to compare the different two or more models. It is a lack of generally agreeable prediction model. Owing to the above reasons, the combined model was compared with historical average only. However, UTCS (Urban Traffic Control System) model was selected to compare with the combined models.

The UTCS-2 model is as follows:

$$\hat{V}_t = m_t + \gamma(m_{t-1} - f_{t-1}) + (1-\alpha)\sum_{s=0}^{t-1}\alpha^s(f_{t-s-1} - m_{t-s-1})$$
$$+\gamma(1-\alpha)\sum_{s=0}^{t-2}\alpha^s(f_{t-s-2} - m_{t-s-2})$$

where

$\hat{V}t$ = predicted volume at time t;
m_t = historical volume at time t;
f_t = measured volume at time t;
α = constant (0.2);
γ = smoothing coefficient (0.9).

The UTCS-2 model rely much on the difference of predicted volume and historical volume. The parameters such as α and γ are very important in combining the values as do in the combined model developed here. The values (0.2 and 0.9) of the parameters were captured from traffic data in Washington D.C. (Stephanedes, 1981). The area-specific parameters are not adaptive for current traffic flow. Let us look at the prediction performances between the UTCS-2 model and the heuristic combined models. The peak-hour traffic data were used for that comparison.

Table 11. Comparison of combined models and the UTCS model

	M1	M2	M3	H.A.	UTCS
MAPE(%)	8.57	7.93	7.58	10.45	14.22
Q-ratio	1.087	1.080	1.076	1.106	1.143
RMSE	511.8	464.9	471.4	625.5	760.2

It is obvious that the combined models are superior to the UTCS model in prediction performance. In the three criteria, the UTCS model is even worse than the historical average. The results indicate that the adaptive weighting system performs overall better than the fixed weighting parameters.

CONCLUSIONS

By incorporating *upstream* traffic into the three combined models better forecasts than historical average are usually produced. Especially, the combined models performed well in the 15- and 30-minutes forecasts under peak-hours traffic conditions. However, the predictions by the combined models under non-peak-hours traffic resulted in no improvements. Since there are not much changes in traffic on non-peak-hours compared to the daily average traffic, the models including two or three variables have not produced forecasts better than the historical average. Also, in the 45-minutes look-ahead, the forecasts from the models were worse than the historical average. It is due to the variables used in the models for farther forecasting. The *upstream* data that were selected based on the current travel times were not reliable because of many factors caused by long trip from *origin link* to *study link*.

Whenever the MAPE rates are greater than 20 percent, the model was better than the historical average in prediction performance. It represents that the combined models performs efficiently under congested traffic condition. The historical average is not able to detect the changes in traffic on the roadways caused by an incident. It indicates that the combined models will be effective for prediction under congested traffic condition. The prediction models developed in this study relied only on the current traffic information, not the future one. Even though a new variable--*upstream*--from an extended dimension was used for the network-based prediction model, the three variables were limited to current traffic information.

RECOMMENDATIONS FOR FURTHER STUDY

If all the link data throughout the network are available, it will be possible to evaluate more deeply the network-based prediction model. In order to conduct such an evaluation, an advanced traffic data collection system is required. There is no doubt that reliable and prompt traffic data through the advanced data collection system (Dickinson and Waterfall, 1984) are necessary for more efficient forecasting. In judging current traffic condition and identifying upstream traffic, speed data plays a big role. If data sets from various places are available in the future, the combined models developed can be evaluated in diverse ways. The availability of data collected from congested area caused by incidents will allow study of the vehicle movements under conditions caused by incidents (non-recurring congestion).

The 15-minute interval data were only used to develop and evaluate the network-based prediction model. The shorter interval data must be used to compare the prediction performance between the link-based model and the network-based model. In addition to that comparison, the traffic flow of the arterial must be studied separately. Although the ramp traffic along the main highways was ignored in this research, a complicated network including the ramps should be considered for better performance. The prediction models with the current traffic information may have a limitation to further improve the prediction performance. A new technique such as neural network analysis may be necessary to approach this problem in a different point-of -view.

REFERENCES

[1] S.A. Ahmed, *Operational Filtering of Traffic Data*, Journal of Forecasting, Vol. 8, 19-32, 1989.

[2] M. Ahmed and A. Cook, *Analysis of Freeway Traffic Time Series Data Using .Box and Jenkins Techniques*, Transportation Research Record, 722, 1-8, 1979.

[3] K.W. Dickinson and R.C. Waterfall, *Video Image Processing for Monitoring Road Traffic*, Institution of Electrical Engineers, International Conference on Road Traffic Data Collection 105-109, 1984.

[4] A.G. Hobeika, R. Sivanandan, C. Kim, R. Subramaniam, and K. Ozbay, *Real-Time Diversion Strategies for Congested Urban Networks,* U.S. DOT/VDOT Centers Program, Mid-Atlantic Universities Transportation Center (MAUTC), Phase 4 working paper, University Center for Transportation Research, Virginia Polytechnic Institute & State University, 1993.

[5] C. Kim and A.G. Hobeika, *A Short-Term Demand Forecasting Model from Real-Time Traffic Data*, Proceedings of Two Parallel Conference : Infrastructure Planning and Management, American Society of Civil Engineers (ASCE), Denver, Colorado, 540-549, 1993.

[6] N.L. Nihan and K.O. Holmesland, *Use of the Box and Jenkins Time Series Technique in Traffic Forecasting,* Transportation, 9, 125-143, 1980.

[7] T. Oda, *An Algorithm for Prediction of Travel Time Using Vehicle Sensor Data,* Third International Conference on Road Traffic Control, May, Institution of Electrical Engineers, 1990.

[8] L. Ott, *An Introduction to Statistical Methods and Data Analysis*, Third edition, PWS-Kent Publishing Company, Boston, 1988.

[9] S.K. Saha, *The Development of a Procedure to Forecast Traffic Volumes on Urban Segments of the State and Interstate Highway Systems*, Ph.D. Dissertation, Graduate School of Purdue University, 1990.

[10]T.J. Stephanedes, P. Michalopoulos, and R.A. Plum, *Improved Estimation of Traffic Flow for Real-Time Control,* Transportation Research Record, 795, 28-39, 1981.