Prediction of Freeway Traffic Flows Using Kalman Predictor in Combination With Time Series

It is essential to predict traffic flow rates dynamically and accurately for traffic engineers to efficiently control traffic flows and reduce traffic delays. This paper introduces a method for prediction of freeway traffic flows. The method combines the combination of the Kalman control theory and the times series theory into a tool for traffic flow prediction. It is illustrated that the combination method provides more accurate traffic flow prediction than using either one of the two theories individually. With the prediction model, the traffic flow on a given freeway in the next time interval (five to 15 minutes) can be predicted using traffic data at the current and past time intervals. Dynamic traffic predictions with the developed model can be performed for individual lanes as well as for all the lanes of each travel direction. It is also shown that a dynamic prediction of traffic flow rate with this prediction model would also constitute a dynamic prediction of traffic congestion if the traffic capacity was given.

by Yi Jiang

Traffic congestion occurs when traffic flow exceeds the capacity of the roadway. Consequently, during congestion vehicles travel the roadway at reduced speeds and with fluctuated traffic flow rates. Motorists endure considerably greater traffic delays under congested traffic conditions than under uncongested conditions. The ability to dynamically predict traffic flow rates is essential for highway/traffic engineers to maintain smooth traffic flows. It would enable them to apply traffic control measures to prevent traffic congestion rather than to deal with traffic problems after traffic congestion already occurred. Methods of adaptive forecasting of traffic flow have been explored by many researchers. Ahmed and Cook (1982) applied time series methods to provide a short-term forecast of traffic occupancies for incident detection. Okutani and Stephanedes (1984) employed the Kalman filtering theory in dynamic prediction of traffic flow. Davis et al. (1990) used pattern recognition algorithms to forecast freeway traffic congestion. Lu (1990) developed a model of adaptive prediction of traffic flow based on the least-mean-square algorithm.

As part of the effort to study the traffic characteristics on Indiana freeways, a dynamic traffic prediction model was developed using the combination of the Kalman predictor theory and the time series theory. Different from the previous prediction models that all utilized a single theory or method for traffic flow prediction, this model combines two theories to formulate a dynamic prediction algorithm. This paper presents the development of the prediction model. The accuracy of predictions when the Kalman filtering theory and the time series theory are used in combination are compared to the prediction accuracy of the time series theory alone. The applications of the prediction model are illustrated through numerical

Journal of the Transportation Research Forum. Published in *Transportation Quarterly,*
Vol. 57, No. 2, Spring 2003 (99–118).

examples with actual traffic flow data. Dynamic traffic predictions with the developed model can be performed for individual lanes as well as for all the lanes of each travel direction. Therefore, the prediction model can be used as an efficient tool for traffic control. It is also shown that a dynamic prediction of traffic flow rate with this prediction model would also constitute a dynamic prediction of traffic congestion if the traffic capacity was given.

DATA COLLECTION

The traffic data used in this study included the data collected with traffic counters and the data from the Weigh-In-Motion (WIM) stations on Indiana freeways. Ten freeway sections across Indiana were selected for data collection with traffic counters. Traffic flow rate, vehicle speed, and classification were recorded at five-minute or 10-minute time intervals during high volume hours and at one-hour intervals during low traffic volume hours. The vehicle counters were set up to classify the detected vehicles into three groups: (1) passenger cars, (2) heavy trucks, and (3) buses. The traffic counter data was used to develop the model of dynamic prediction of freeway traffic flow rates. There are 20 WIM stations on Indiana freeways. WIM traffic data was collected from the 20 interstate WIM stations to study the traffic characteristics on Indiana freeways for 12 months. The traffic data covered a 13-month period, between January 1, 1998, to January 31, 1999; however, the data for March 1998 was not available because of problems with the WIM software. It was found that two of the 20 stations did not properly function at all during the 13 months and therefore could not provide useful data for this study. The other 18 WIM stations worked properly at least for one month during the 12 months. Thus, the traffic data from the 18 WIM stations was used in this study.

FREEWAY CAPACITY

Capacity is defined in terms of the maximum rate of flow that can be accommodated by a given road under prevailing conditions (TRB 2000). Traffic congestion occurs when traffic flow exceeds the capacity of the roadway. Consequently, during congestion, vehicles travel at reduced speeds and with fluctuating traffic flow rates. Motorists endure considerably greater traffic delays under congested traffic conditions than under uncongested conditions. There are two types of traffic congestion, nonrecurrent congestion and recurrent congestion. Nonrecurrent congestion is unanticipated congestion due to the random nature of traffic flows and incidents. Recurrent congestion often occurs at specific locations, such as at bottle neck locations, due to regular rush hour traffic and problems with highway layout or design. This study deals with only nonrecurrent traffic congestion.

The reported maximum one-way volumes in the *2000 Highway Capacity Manual* range from 2,446 vehicles per hour per lane (veh/h/ln) to 2,552 veh/h/ln for four-lane freeways, and from 2,500 veh/h/ln to 2,664 veh/h/ln for six-lane freeways. The manual recommends a rate of flow of 2,400 passenger cars per hour per lane (pc/h/ln) for freeways with free-flow speeds of 70 to 75 miles per hour (mph) and 2,300 pc/h/ln for freeways with free-flow speeds of 65 mph as the capacity under base conditions.

A study (Jiang 1999) was conducted to determine the freeway capacity values in Indiana. It was observed during the study that in Indiana, traffic flows changed from uncongested to congested conditions always with a sharp speed drop. This observation validates the research findings based on the catastrophe theory by Persaud and Hall (1989). Their research indicated that the transitions from uncongested to congested traffic conditions are characterized by a fairly gentle change in occupancy, and a fairly constant flow, but a sudden and sharp change in speed. Therefore, freeway capacity is identified in this study as the maximum observed hourly volume before a substantial speed drop. To express freeway capacity in passenger cars per hour, the traffic flow rate was converted to hourly volume and the adjustment factors from the *2000 Highway Capacity Manual* were used to convert heavy vehicles to passenger car equivalents. The observed capacity values on Indiana's four-lane freeways range from 1,489 to 2,006 pc/h/ln with an average value of 1,767 pc/h/ln. The observed capacity values on Indiana's six-lane freeways range from 1,463 to 2,093 pc/h/ln with an average value of 1,778 pc/h/ln. The Indiana study recommended to use the average capacity values to represent the Indiana freeway capacities in traffic analysis. It should be noted that Indiana's capacity values are based on the number of freeway lanes while the freeway capacity values in the *2000 Highway Capacity Manual* are based on the freeway free-flow speeds. This is because the early version of the manual, the *1994 Highway Capacity Manual* (TRB 1994), reported freeway capacity values in terms of the number of freeway lanes, and the Indiana study was conducted before the publication of the 2000 manual.

DYNAMIC PREDICTION OF TRAFFIC FLOW

Traffic Flow Prediction Using Time Series: Given the capacity values, it was desired to develop methods for predicting traffic flow and congestion so that appropriate traffic control strategies could be applied to avoid traffic congestion and to reduce traffic delay. Traffic flow rates constantly change with time on any given highway section. To predict traffic conditions, the relationship between traffic flow and time must be studied. The time series theory (Cryer 1986; Bowerman and O'Connell 1979) is a frequently used tool to study the traffic and time relationship. One of the time series models is the *autoregressive process* {Z(t)}. A *p*th-order autoregressive process, AR(p), satisfies the following equation (Bowerman and O'Connell 1979):

$$(1)\quad Z(t) = \phi_1 Z(t-1) + \phi_2 Z(t-2) + \cdots + \phi_p Z(t-p) + \varepsilon_t$$

where:

$Z(t)$ = value of the process Z at time t;

ϕ_i = unknown parameters; i = 1, 2, 3, ..., p

ε_t = a random variable with zero mean and variance σ_w^2.

This equation requires that the mean of the series has been subtracted out so that $Z(t)$ has a zero mean. This time series implies that the current value of the series $Z(t)$ is a linear combination of the p most recent past values of itself plus an error term ε_t.

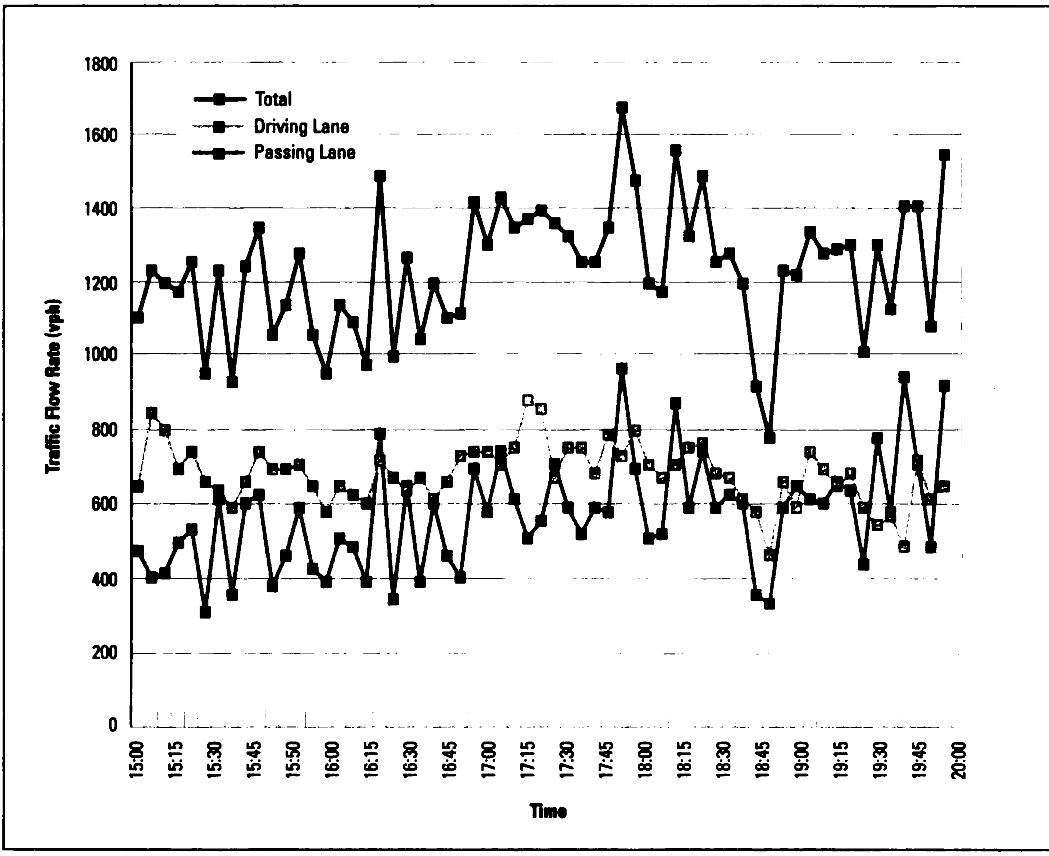
To demonstrate the development of a model of dynamically predicting traffic flow rates, traffic data recorded with traffic counters on Interstate 65 (I-65) at about one mile south of State Road 47 (SR-47) was used. Figure 1 shows the observed traffic flow rates in order of time.

With the traffic flow data, an AR(1) model was fitted using the MINITAB (Minitab 1996) computer software. The AR(1) equation for the traffic flow rate is expressed as follows:

$$(2)\quad f(t) = \phi_1 f(t-1) + \varepsilon_t$$

In Equation 2, $f(t)$ denotes the traffic flow rate at time t. As expressed by the equation, the traffic flow rate at time t, $f(t)$, can be predicted from the traffic flow rate observed at the most recent past time point t-1, $f(t-1)$. It should be noted that the mean of the series of traffic flow rates must be subtracted from $f(t)$ as required by the autoregressive model of Equation 1. The actual prediction is then the

Figure 1: Observed Traffic Flow on I-65



calculated *f(t)* plus the mean. If *f(t*-1) is given, then *f(t)* can be predicted as:

$$(3)\quad \hat{f}(t \mid t-1) = \bar{\phi}_1 f(t-1)$$

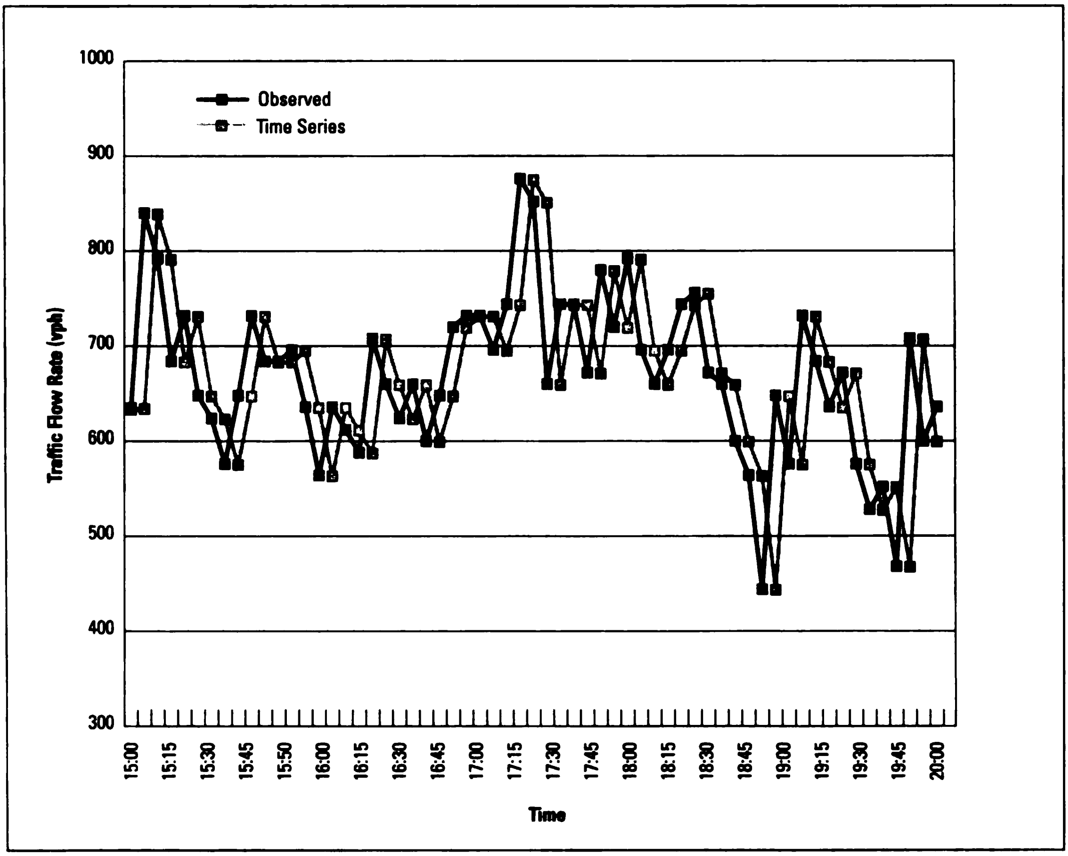
In this equation, $\bar{\phi}_1$ is the estimate of ϕ_1, and $\hat{f}(t \mid t-1)$ is the predicted value of *f(t)* based on the most recent observed traffic flow rate, *f*(*t*-1). Through this equation, predictions of traffic flow rates at the given location were calculated from 15:00 to 20:00 at five-minute intervals. For comparison, plotted in Figures 2, 3, and 4 are the predicted and observed values of the traffic flow rates.

The curves in the three figures indicate that the predicted values followed the patterns of the observed traffic flows. The accuracy of the time series predictions is reflected by the values of prediction errors. In this case, an error is the difference between the observed traffic flow rate and the traffic flow rate predicted by the time series model divided by the observed traffic flow rate, that is,

$$\text{error} = \frac{f(t) - \hat{f}(t \mid t-1)}{f(t)}.$$

The time series prediction errors expressed as percentages are listed in Table 1 for all data points during the five-hour period. There are 14 out of the 61 predictions with errors less than 5% for the driving lane, seven out of the 61 for the passing lane, and 17 out of the 61 for the total volumes of the two lanes. These error values suggest the need for improvement in the accuracy of the time series predictions.

Figure 2: Observed and Time Series Predicted Traffic Flow on Driving Lane of I-65



Traffic Flow Prediction Using Kalman Predictor

One of the applications of the control theory is to use the Kalman predictor (Bozic 1979) in recursive predictions of random processes. Random processes are often called signals because many models were originally established to systematically maximize the receipt of the desired radio transmission signals and minimize the noises (undesired signals). The noises are considered the errors of random processes. For example, a random signal model can be a first-order autoregressive process:

$$(4) \quad x(t+1) = a\,x(t) + w_t$$

where $x(t)$ and $x(t+1)$ are the values of the random signal at time t and time t+1, respectively; a is a coefficient; and w_t is the random signal error term with a mean value of 0.

The observation (or measurement) is affected by additive random error v_t:

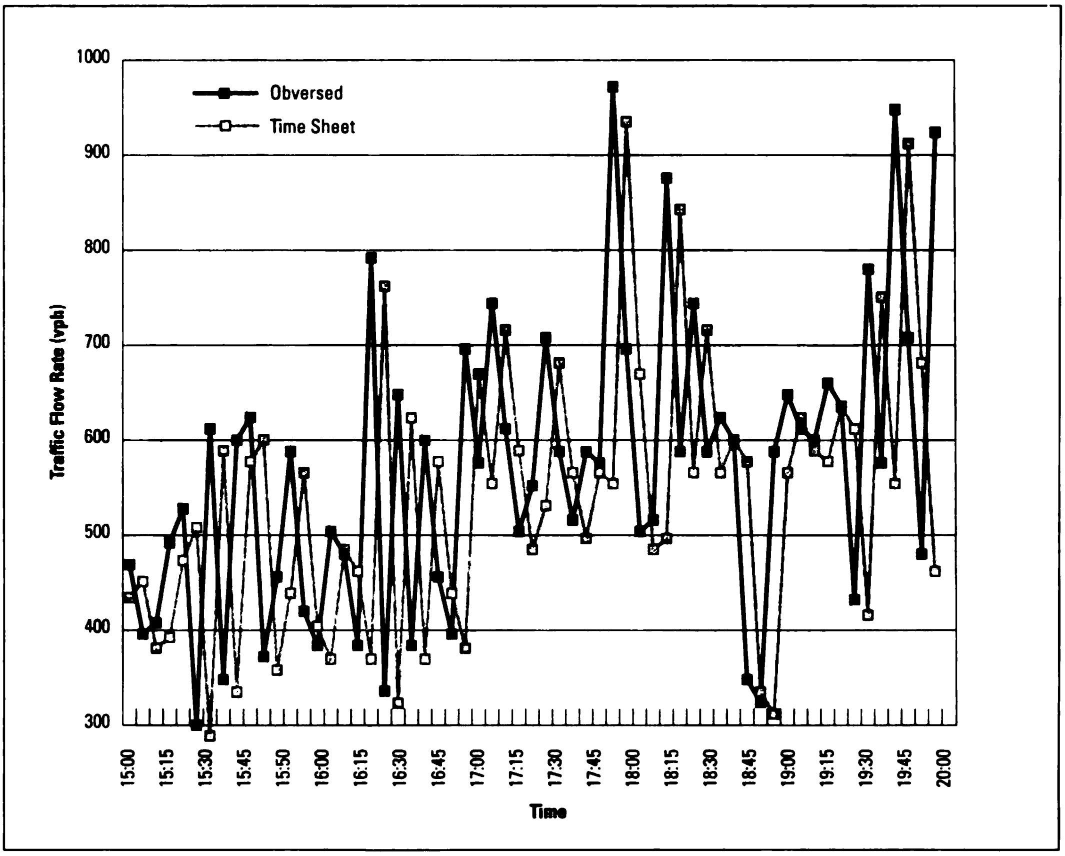
$$(5) \quad y(t) = c\,x(t) + v_t$$

where $y(t)$ is the measurement of the variable $x(t)$; c is a coefficient; and v_t is the error of the measurement with zero mean and variance σ^2_v.

The Kalman predictor for the above signal model can be expressed as follows:

Predictor equation:

Figure 3: Observed and Time Series Predicted Traffic Flow on Passing Lane of I-65



(6) $\hat{x}(t+1 \mid t) = a\,\hat{x}(t \mid t-1) + k(t)[y(t) - c\hat{x}(t \mid t-1)]$

where $\hat{x}(t / t-1)$ denotes the prediction of *x(t)* based on *x(t-1)*; and *k(t)* is the Kalman predictor gain derived through mathematically minimizing the mean-square prediction error (Bozic 1979). *k(t)* is expressed as in the following equation.

Predictor gain:

(7) $$k(t) = \frac{acp(t \mid t-1)}{c^2 p(t \mid t-1) + \sigma_v^2}$$

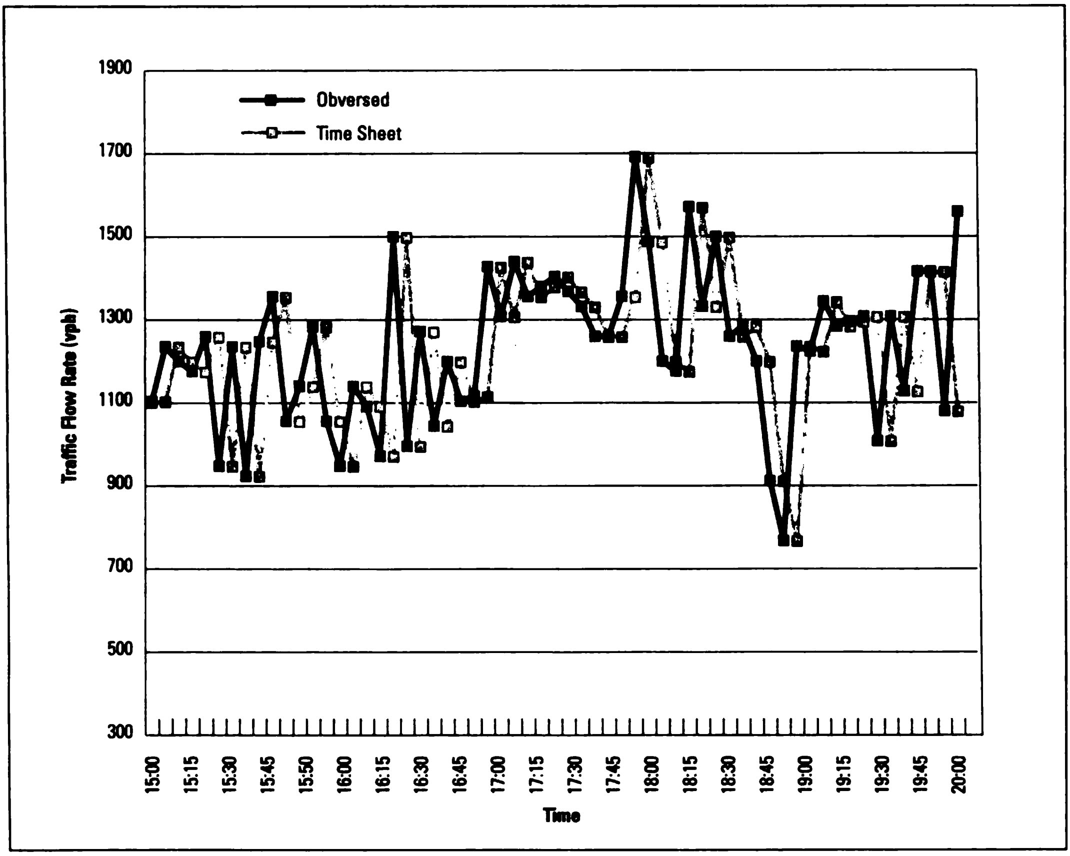
where *p(t|t-1)* is the Kalman prediction mean-square error at time t, which is also derived through mathematical manipulation (Bozic 1979). The following equation shows the prediction mean-square error at time t+1:

Prediction mean-square error:

(8) $$p(t+1 \mid t) = \frac{a}{c} k(t)\sigma_v^2 + \sigma_w^2$$

Equations 6, 7, and 8 are called one-step Kalman predictor of the signal process expressed by Equations 4 and 5. The Kalman method yields the estimate of $x(t+1)$, i.e., the signal at time t+1, given the measured data x(t) and the previous estimate $\hat{x}(t \mid t\text{-}1)$ at time t. It can be proved (Bozic 1979) that this one-step prediction estimate, denoted as $\hat{x}(t+1 \mid t)$, is an optimum estimate because the Kalman recursive prediction process minimizes the mean-square prediction error $E[x(t+1) - \hat{x}(t+1 \mid t)]$.[2]

Figure 4: Observed and Time Series Predicted Two-Lane Total Traffic Flow of I-65



The Kalman predictor has the features of recursive computation, continuous incorporation of the most recent available data, and optimum prediction. These are exactly the desirable functions for an efficient traffic flow prediction model. To use the Kalman predictor in traffic flow prediction, the AR(1) time series model as in Equation 3 can be used as the traffic flow model, that is:

(9) $f(t+1) = \phi f(t) + \varepsilon_t$

Equation 9 is the first-order autoregressive process for the traffic flow. In addition, the observation (or measurement) of the traffic flow, m(t), is affected by additive random error ν_t. In terms of traffic flow, ν_t represents the errors involved in traffic flow measurement, including traffic counter errors and human errors during data collection and data processing.

(10) $m(t) = \beta f(t) + \nu_t$

Equation 10 is related to the accuracy of the traffic data measurement devices used in data collection. The one-step Kalman recursive prediction equations can then be readily obtained from Equations 6 through 8:

Predictor equation:

(11) $\hat{f}(t+1 \mid t) = \phi \hat{f}(t \mid t-1) + k(t)[m(t) - \beta \hat{f}(t \mid t-1)]$

Table 1: Comparison of Observed and Time Series Predicted Traffic Flow Rates on I-65

Time	Driving Lane			Passing Lane			Total		
	Observed	Predicted	Error	Observed	Predicted	Error	Observed	Predicted	Error
15:00	635	633.1	**0.3%**	469	434.3	7.4%	1104	1100.1	**0.4%**
15:05	840	634	24.5%	396	451.3	-14.0%	1236	1102.1	10.8%
15:10	792	838.7	-5.9%	408	381.1	6.6%	1200	1233.8	**-2.8%**
15:15	684	790.8	-15.6%	492	392.6	20.2%	1176	1197.9	**-1.9%**
15:20	732	683	6.7%	528	473.4	10.3%	1260	1173.9	6.8%
15:25	648	730.9	-12.8%	300	508.1	-69.4%	948	1257.8	-32.7%
15:30	624	647	**-3.7%**	612	288.7	52.8%	1236	946.3	23.4%
15:35	576	623.1	-8.2%	348	588.9	-69.2%	924	1233.8	-33.5%
15:40	648	575.1	11.3%	600	334.9	44.2%	1248	922.4	26.1%
15:45	732	647	11.6%	624	577.4	7.5%	1356	1245.8	8.1%
15:50	684	730.9	-6.9%	372	600.5	-61.4%	1056	1353.6	-28.2%
15:55	684	683	**0.1%**	456	358	21.5%	1140	1054.1	7.5%
16:00	696	683	**1.9%**	588	438.8	25.4%	1284	1138	11.4%
16:05	636	694.9	-9.3%	420	565.8	-34.7%	1056	1281.7	-21.4%
16:10	564	635	-12.6%	384	404.2	-5.3%	948	1054.1	-11.2%
16:15	636	563.1	11.5%	504	369.5	26.7%	1140	946.3	17.0%
16:20	612	635	**-3.8%**	480	485	**-1.0%**	1092	1138	**-4.2%**
16:25	588	611.1	**-3.9%**	384	461.9	-20.3%	972	1090.1	-12.2%
16:30	708	587.1	17.1%	792	369.5	53.3%	1500	970.3	35.3%
16:35	660	706.9	-7.1%	336	762.1	-126.8%	996	1497.4	-50.3%
16:40	624	659	-5.6%	648	323.3	50.1%	1272	994.2	21.8%
16:45	660	623.1	5.6%	384	623.5	-62.4%	1044	1269.8	-21.6%
16:50	600	659	-9.8%	600	369.5	38.4%	1200	1042.2	13.2%
16:55	648	599.1	7.5%	456	577.4	-26.6%	1104	1197.9	-8.5%
17:00	720	647	10.1%	396	438.8	-10.8%	1116	1102.1	**1.3%**
17:05	732	718.9	**1.8%**	696	381.1	45.2%	1428	1114	22.0%
17:10	732	730.9	**0.2%**	576	669.7	-16.3%	1308	1425.5	-9.0%
17:15	696	730.9	**-5.0%**	744	554.3	25.5%	1440	1305.7	9.3%
17:20	744	694.9	6.6%	612	715.9	-17.0%	1356	1437.5	-6.0%
17:25	876	742.9	15.2%	504	588.9	-16.8%	1380	1353.6	**1.9%**
17:30	852	874.7	**-2.7%**	552	485	12.1%	1404	1377.6	**1.9%**
17:35	660	850.7	-28.9%	708	531.2	25.0%	1368	1401.5	**-2.4%**

Table 1: continued

Time	Driving Lane			Passing Lane			Total		
	Observed	Predicted	Error	Observed	Predicted	Error	Observed	Predicted	Error
17:40	744	659	11.4%	588	681.3	-15.9%	1332	1365.6	**-2.5%**
17:45	744	742.9	**0.1%**	516	565.8	-9.7%	1260	1329.6	-5.5%
17:50	672	742.9	-10.6%	588	496.5	15.6%	1260	1257.8	**0.2%**
17:55	780	671	14.0%	576	565.8	**1.8%**	1356	1257.8	7.2%
18:00	720	778.8	-8.2%	972	554.3	43.0%	1692	1353.6	20.0%
18:05	792	718.9	9.2%	696	935.3	-34.4%	1488	1689	-13.5%
18:10	696	790.8	-13.6%	504	669.7	-32.9%	1200	1485.4	-23.8%
18:15	660	694.9	-5.3%	516	485	6.0%	1176	1197.9	**-1.9%**
18:20	696	659	5.3%	876	496.5	43.3%	1572	1173.9	25.3%
18:25	744	694.9	6.6%	588	842.9	-43.4%	1332	1569.2	-17.8%
18:30	756	742.9	**1.7%**	744	565.8	24.0%	1500	1329.6	11.4%
18:35	672	754.9	-12.3%	588	715.9	-21.8%	1260	1497.4	-18.8%
18:40	660	671	**-1.7%**	624	565.8	9.3%	1284	1257.8	**2.0%**
18:45	600	659	-9.8%	600	600.5	**-0.1%**	1200	1281.7	-6.8%
18:50	564	599.1	-6.2%	348	577.4	-65.9%	912	1197.9	-31.3%
18:55	444	563.1	-26.8%	324	334.9	**-3.4%**	768	910.4	-18.5%
19:00	648	443.3	31.6%	588	311.8	47.0%	1236	766.6	38.0%
19:05	576	647	-12.3%	648	565.8	12.7%	1224	1233.8	**-0.8%**
19:10	732	575.1	21.4%	612	623.5	**-1.9%**	1344	1221.8	9.1%
19:15	684	730.9	-6.9%	600	588.9	**1.9%**	1284	1341.6	**-4.5%**
19:20	636	683	-7.4%	660	577.4	12.5%	1296	1281.7	**1.1%**
19:25	672	635	5.5%	636	635.1	**0.1%**	1308	1293.7	**1.1%**
19:30	576	671	-16.5%	432	612	-41.7%	1008	1305.7	-29.5%
19:35	528	575.1	-8.9%	780	415.7	46.7%	1308	1006.2	23.1%
19:40	552	527.2	**4.5%**	576	750.6	-30.3%	1128	1305.7	-15.8%
19:45	468	551.2	-17.8%	948	554.3	41.5%	1416	1126	20.5%
19:50	708	467.3	34.0%	708	912.2	-28.8%	1416	1413.5	**0.2%**
19:55	600	706.9	-17.8%	480	681.3	-41.9%	1080	1413.5	-30.9%
20:00	636	599.1	5.8%	924	461.9	50.0%	1560	1078.1	30.9%

Predictor gain:

$$(12) \quad k(t) = \frac{\phi \beta p(t \mid t-1)}{\beta^2 p(t \mid t-1) + \sigma_v^2}$$

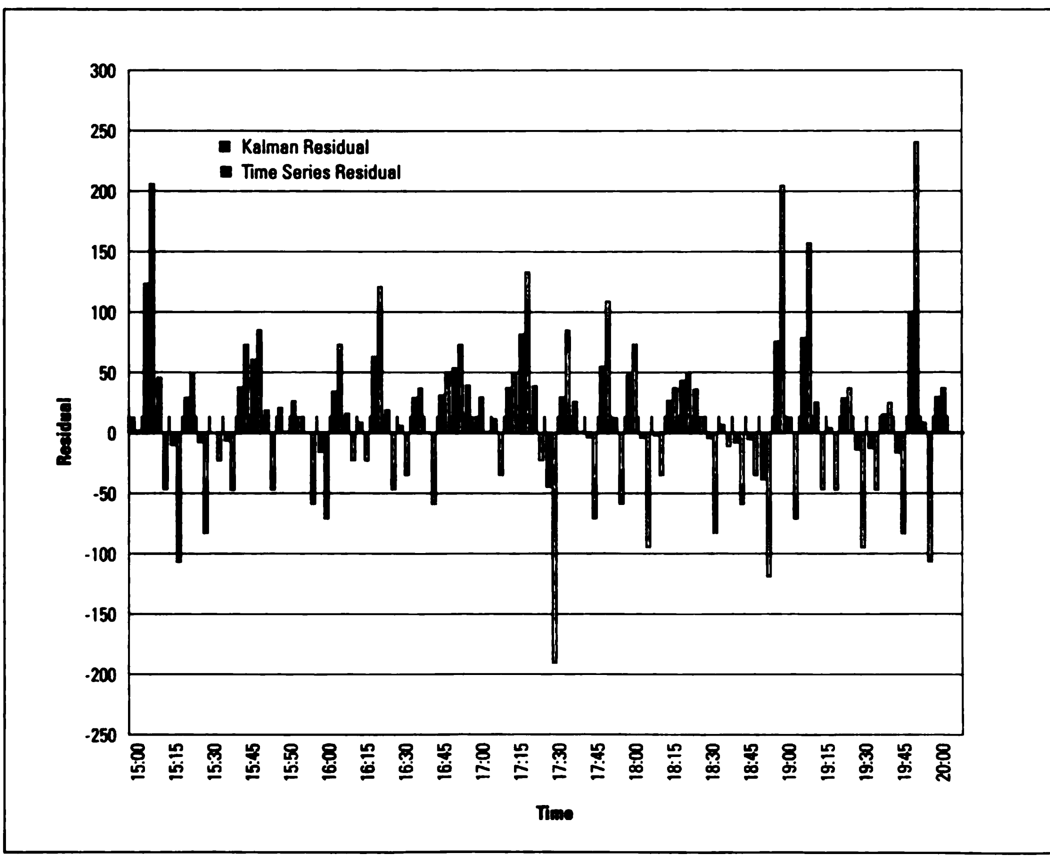
Prediction mean-square error:

$$(13) \quad p(t+1 \mid t) = \frac{\phi}{\beta} k(t) \sigma_v^2 + \sigma_\varepsilon^2$$

With Equations 9 through 13, traffic flow rate at t+1, $f(t+1)$, can be predicted as $\hat{f}(t+1 \mid t)$ for each observed data at time t, $f(t)$. Since Equation 9 is a time series model of the first order autoregressive process, this Kalman predictor model is a combination of the time series and the Kalman predictor. It was expected that this prediction model would improve the prediction accuracy over the time series model as defined in Equation 9.[1] To verify this, the Kalman predictor model was also applied to the traffic flow data described in Figure 1. The differences in the prediction accuracy of the two methods can be clearly described by plotting their corresponding residual values into the same graph, as shown in Figures 5, 6, and 7. The residual graphs distinctly show that most residuals of the Kalman predictions are considerably smaller than those of the time series predictions. Therefore, the improvement of the Kalman predictor over the time series method in traffic flow prediction is apparent.

For a quantitative comparison, the residual values of the time series and Kalman pre-

Figure 5: Residuals of Kalman and Time Series Predictions on Driving Lane



dictions are presented in Table 2. In addition, the differences between the absolute values of the time series and the Kalman residuals are also included in the table. Because there are positive and negative residuals, the use of the absolute values of the residuals is to compare the magnitudes of the residuals from the two prediction methods. The magnitude of a residual is the difference between the observed value and the predicted value. Therefore, a more accurate prediction yields a smaller magnitude of residual. If the absolute value of time series residual (TR) minus the absolute of the Kalman residual (KR) is positive, i.e., abs(TR)-abs(KR) > 0, then the magnitude of time series residual is greater than the Kalman residual, indicating the time series prediction is less accurate than the Kalman prediction.

As shown in Table 2, there are 53 positive values and eight negative values of abs(TR)-abs(KR) for the driving lane, 53 positive and eight negative ones for the passing lane, and 49 positive and 13 negative ones for the two-lane total. This indicates that 53 out of the 61 Kalman predictions are more accurate than the time series predictions for the driving lane and the passing lane, and 49 out of the 61 Kalman predictions are more accurate for the total traffic volumes of the two lanes. Table 2 also includes the statistics of the absolute values of the residuals for the predictions from the two methods. The statistics show that for

Figure 6: Residuals of Kalman and Time Series Predictions on Passing Lane

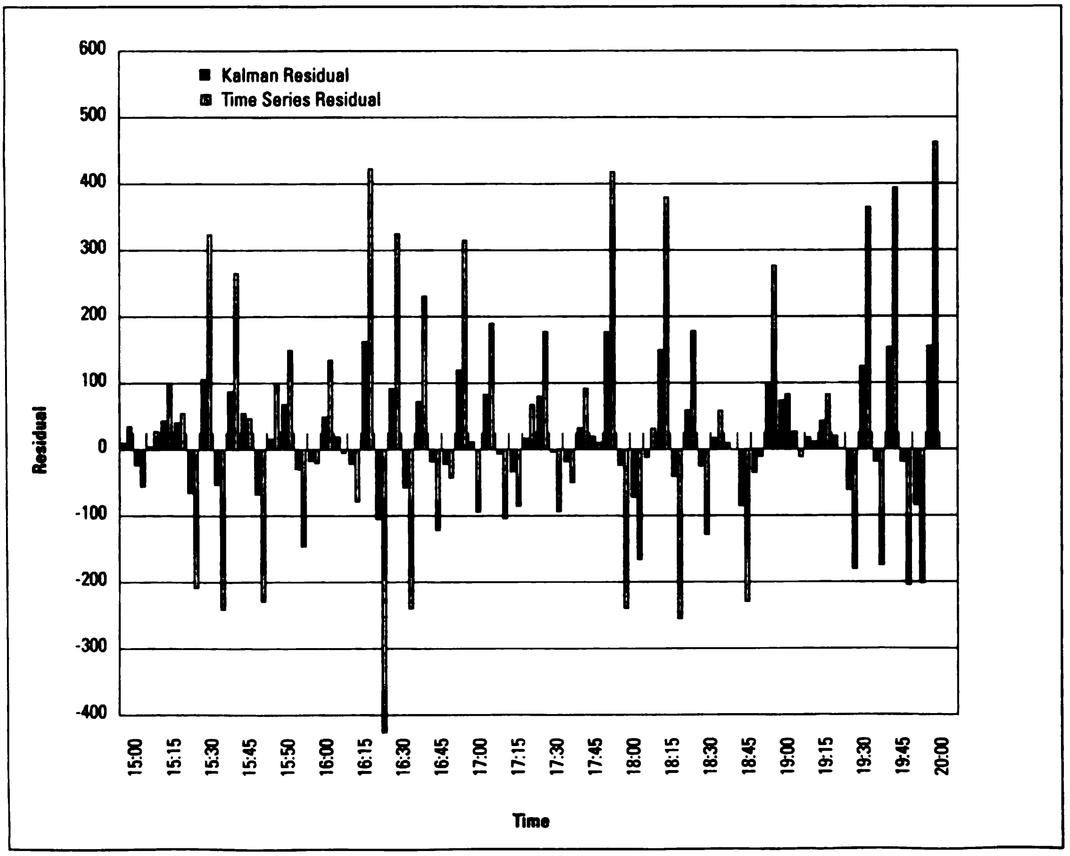
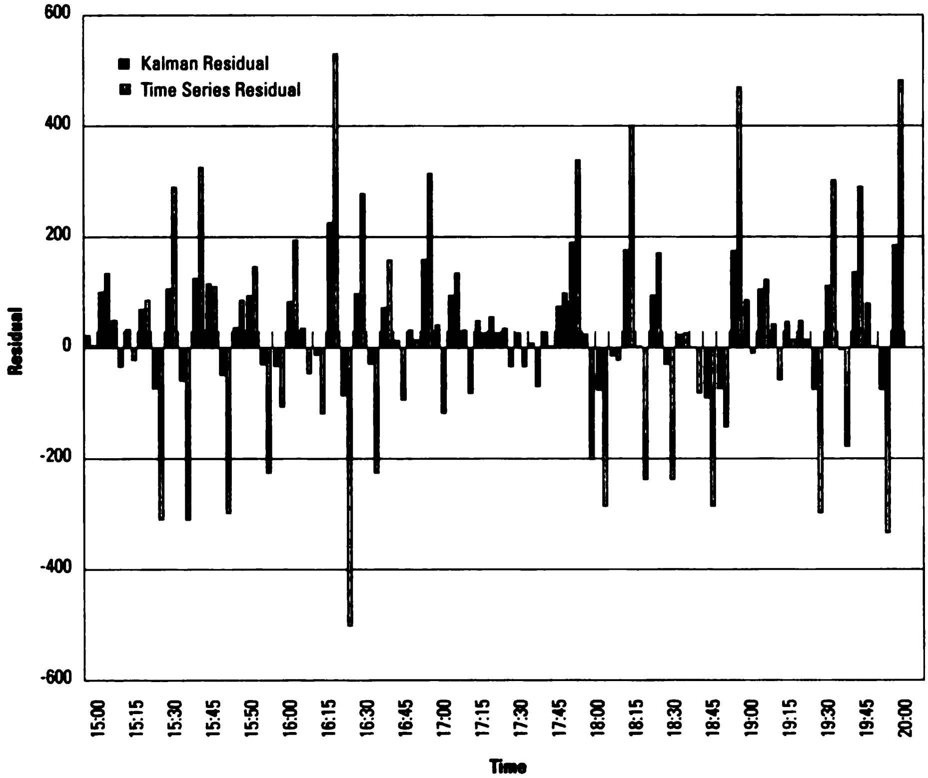


Figure 7: Residuals of Kalman and Time Series Predictions of Two-Lane Total Traffic Flow



both methods the values of means and standard deviations in the driving lane are less than those in the passing lane. This means that the predictions are more accurate for the driving lane. This is because, as shown in Figure 1, traffic flow in the passing lane had more fluctuations with time, which introduces more uncertainties and thus more errors to the predictions.

Table 2 indicates that the Kalman predictions have smaller values for the mean, standard deviation, and minimum and maximum of the absolute residual values than the time series predictions. Compared to the time series predictions, the Kalman predictions reduced the mean of the absolute residual values by (66.04-28.37)/66.04=57.0%, (161.20-54.36)/161.20=66.3%, and (173.4-69.52)/173.4=59.9%, and the standard deviation by (51.74-25.36)/51.74=51.0%, (126.5-44.36)/126.5=64.9%, and (141.0-50.92)/141.0=63.9% for the driving lane, passing lane, and two-lane total, respectively. These large reductions in the values of the mean and standard deviation represent a considerable improvement in the traffic flow predictions. Again, because of the more fluctuating nature of the passing lane traffic flow, the prediction accuracy in the driving lane is better than in the passing lane for both prediction methods. However, the reductions in the residual means and standard deviations are higher for the passing lane (66.3% vs. 57.0% for the means and 64.9% vs. 51.0%

Table 2: Comparison of Time Series and Kalman Predictions on I-65

Time	Driving Lane			Passing Lane			Total		
	Time Residual (TR)	Kalman Residual (KR)	Abs(TR) -Abs(KR)	Time Residual (TR)	Kalman Residual (KR)	Abs(TR) -Abs(KR)	Time Residual (TR)	Kalman Residual (KR)	Abs(TR) -Abs(KR)
15:00	1.9	12.7	-10.8	34.7	9.38	25.3	3.9	22.08	-18.2
15:05	206.0	123.5	82.5	-55.3	-23.25	32.1	134.0	100.22	33.7
15:10	-46.7	45.7	1.1	26.9	3.71	23.2	-33.8	49.37	-15.6
15:15	-106.8	-10.1	96.7	99.4	43.25	56.1	-21.9	33.18	-11.3
15:20	49.0	29.0	20.0	54.6	40.58	14.0	86.1	69.59	16.5
15:25	-82.9	-7.9	75.0	-208.1	-64.94	143.1	-309.8	-72.79	237.0
15:30	-23.0	0.4	22.6	323.3	105.74	217.6	289.7	106.17	183.5
15:35	-47.1	-6.5	40.6	-240.9	-52.93	188.0	-309.8	-59.42	250.4
15:40	72.9	37.8	35.1	265.1	87.34	177.8	325.6	125.09	200.5
15:45	85.0	60.7	24.3	46.6	54.62	-8.0	110.2	115.33	-5.1
15:50	-46.9	18.5	28.4	-228.5	-67.26	161.2	-297.6	-48.77	248.8
15:55	1.0	20.7	-19.6	98.0	15.46	82.6	85.9	36.13	49.7
16:00	13.0	26.3	-13.3	149.2	67.56	81.6	146.0	93.85	52.2
16:05	-58.9	0.0	58.9	-145.8	-29.61	116.2	-225.7	-29.65	196.1
16:10	-71.0	-16.0	55.1	-20.2	-17.15	3.0	-106.1	-33.13	73.0
16:15	72.9	33.9	38.9	134.5	49	85.5	193.7	82.92	110.8
16:20	-23.0	16.0	7.0	-5.0	19.06	-14.1	-46.0	35.06	10.9
16:25	-23.1	8.7	14.3	-77.9	-21.44	56.4	-118.1	-12.71	105.4
16:30	120.9	62.9	58.0	422.5	162.13	260.4	529.7	225.01	304.7
16:35	-46.9	18.8	28.1	-426.1	-104.49	321.6	-501.3	-85.66	415.7
16:40	-35.0	6.0	29.0	324.7	91.49	233.2	277.8	97.47	180.3
16:45	36.9	29.1	7.9	-239.5	-57.61	181.9	-225.7	-28.52	197.2
16:50	-59.0	0.3	58.7	230.5	71.94	158.6	157.8	72.25	85.6
16:55	48.9	31.2	17.7	-121.4	-18.15	103.2	-93.9	13.09	80.8
17:00	73.0	53.5	19.5	-42.8	-21.66	21.1	14.0	31.82	-17.9
17:05	13.1	39.4	-26.3	314.9	119.26	195.7	314.0	158.68	155.3
17:10	1.1	29.6	-28.4	-93.7	11.24	82.5	-117.5	40.8	76.7
17:15	-34.9	11.5	23.4	189.7	82.71	107.0	134.3	94.19	40.1
17:20	49.1	37.4	11.7	-103.9	-6.41	97.5	-81.5	30.98	50.5
17:25	133.1	81.6	51.5	-84.9	-33.22	51.7	26.4	48.41	-22.0

continued

Table 2: continued

Time	Driving Lane			Passing Lane			Total		
	Time Residual (TR)	Kalman Residual (KR)	Abs(TR) -Abs(KR)	Time Residual (TR)	Kalman Residual (KR)	Abs(TR) -Abs(KR)	Time Residual (TR)	Kalman Residual (KR)	Abs(TR) -Abs(KR)
17:30	-22.7	38.9	-16.2	67.0	16.63	50.4	26.4	55.48	-29.0
17:35	-190.7	-44.8	145.9	176.8	79.49	97.3	-33.5	34.73	-1.2
17:40	85.0	29.7	55.3	-93.3	-3.57	89.7	-33.6	26.16	7.4
17:45	1.1	26.1	-25.0	-49.8	-18.28	31.5	-69.6	7.85	61.8
17:50	-70.9	-3.9	66.9	91.5	32.09	59.4	2.2	28.17	-25.9
17:55	109.0	55.0	54.0	10.2	19.12	-8.9	98.2	74.11	24.1
18:00	-58.8	12.5	46.3	417.7	176.54	241.2	338.4	189.05	149.3
18:05	73.1	47.8	25.3	-239.3	-23.72	215.6	-201.0	24.1	176.9
18:10	-94.8	-4.3	90.5	-165.7	-71.56	94.2	-285.4	-75.87	209.5
18:15	-34.9	-2.1	32.9	31.0	-12.22	18.8	-21.9	-14.28	7.6
18:20	37.0	26.8	10.2	379.5	149.14	230.3	398.1	175.9	222.2
18:25	49.1	43.2	5.9	-254.9	-40.79	214.2	-237.2	2.38	234.8
18:30	13.1	36.0	-22.9	178.2	58.48	119.7	170.4	94.48	75.9
18:35	-82.9	-4.7	78.1	-127.9	-25.15	102.8	-237.3	-29.87	207.5
18:40	-11.0	6.9	4.1	58.2	16.59	41.6	26.2	23.46	2.8
18:45	-59.0	-8.1	50.9	-0.5	9.19	-8.7	-81.7	1.09	80.6
18:50	-35.1	-5.4	29.7	-229.4	-84.93	144.4	-285.9	-90.34	195.5
18:55	-119.1	-38.6	80.6	-10.9	-34.74	-23.9	-142.4	-73.33	69.1
19:00	204.7	75.6	129.1	276.2	98.52	177.7	469.4	174.08	295.3
19:05	-71.0	12.9	58.1	82.2	72.95	9.2	-9.8	85.82	-76.0
19:10	156.9	78.6	78.3	-11.5	26.22	-14.7	122.2	104.77	17.4
19:15	-46.9	25.2	21.6	11.1	17.38	-6.3	-57.6	42.62	15.0
19:20	-47.0	4.1	42.9	82.6	42.49	40.2	14.3	46.59	-32.3
19:25	37.0	28.6	8.3	0.9	19.72	-18.8	14.3	48.33	-34.0
19:30	-95.0	-14.0	81.0	-180.0	-61.1	118.9	-297.7	-75.08	222.6
19:35	-47.1	-12.9	34.2	364.3	124.18	240.1	301.8	111.28	190.5
19:40	24.8	15.2	9.6	-174.6	-18.69	155.9	-177.7	-3.45	174.2
19:45	-83.2	-16.7	66.5	393.7	152.67	241.1	290.0	136.01	154.0
19:50	240.7	98.7	142.0	-204.2	-18.89	185.3	2.5	79.79	-77.3
19:55	-106.9	8.5	98.5	-201.3	-83.84	117.4	-333.5	-75.36	258.1
20:00	36.9	29.6	7.4	462.1	154.78	307.3	481.9	184.33	297.6

Table 2: continued

Time	Driving Lane			Passing Lane			Total		
	Time Residual (TR)	Kalman Residual (KR)	Abs(TR) -Abs(KR)	Time Residual (TR)	Kalman Residual (KR)	Abs(TR) -Abs(KR)	Time Residual (TR)	Kalman Residual (KR)	Abs(TR) -Abs(KR)
	Statistics of Absolute Values of Residuals:			Statistics of Absolute Values of Residuals:			Statistics of Absolute Values of Residuals:		
		Time Series	Kalman		Time Series	Kalman		Time Series	Kalman
	Mean	66.04	28.37	Mean	161.20	54.36	Mean	173.40	69.52
	StDev	51.74	25.36	StDev	126.50	44.36	StDev	141.00	50.92
	Min	1.04	0.04	Min	0.50	3.57	Min	2.20	1.09
	Max	240.71	123.47	Max	462.10	176.54	Max	529.70	225.01

for the standard deviations). This indicates that the Kalman method can correct more errors when the input data contains higher fluctuations.

To statistically compare the predictions of the two methods, paired t-tests were performed. Since a t-test requires the data follow a normal distribution, the Anderson-Darling normality test (Minitab 1996) was used to check if the absolute values of the residuals follow a normal distribution. The normality tests indicate none of the data sets follows a normal distribution at a level of $\alpha = 0.05$. Then the data sets were transformed by square root of the absolute values of the residuals, i.e., $r'_{1i} = \sqrt{abs\ (TR)}$ and $r'_{2i} = \sqrt{abs\ (KR)}$. The Anderson-Darling normality tests on the transformed data yielded p-values greater than $\alpha = 0.05$. Therefore, the transformed data sets are normally distributed at a level of $\alpha = 0.05$ and the paired t-tests can be applied to compare them. The paired t-tests were used to test if the difference between the mean of r'_{1i} (μ_1) and the mean of r'_{2i} (μ_2) is zero or greater than zero. The hypotheses to be tested are as follows:

H_0: $\mu_1 - \mu_2 = 0$

H_a: $\mu_1 - \mu_2 > 0$

If the Type I error is controlled at $\alpha = 0.05$, then the p-value of the paired t-test can be compared to the α value according to the decision rule:

If p-value $\geq \alpha$, conclude H_0.

If p-value $< \alpha$, conclude H_a.

All of the p-values of the paired t-tests are 0.000 for the driving lane, passing lane, and two-lane total, which is less than $\alpha = 0.05$. Therefore, H_a is concluded, i.e., the mean difference in residuals is greater than zero or μ_1 is significantly greater than μ_2. This implies that the Kalman predictor in combination with the time series method provides much better predictions of traffic flow rates than the time series method.

To further compare the accuracies of the two prediction methods, the two methods were also applied to traffic flow data collected on two other freeway sections. One section was on I-69 at SR-14 and the other was on I-70 just east of SR-9. The traffic flow data on I-69 was from 17:00 to 21:00 at 10-minute intervals and on I-70 was from 6:00 to 8:00 at five-minute intervals. In the same manner as in Table 2, the residual values of the time series and Kalman predictions for I-

Table 3: Comparison of Time Series and Kalman Predictions on I-69

Time	Driving Lane			Passing Lane			Total		
	Time Residual (TR)	Kalman Residual (KR)	Abs(TR) -Abs(KR)	Time Residual (TR)	Kalman Residual (KR)	Abs(TR) -Abs(KR)	Time Residual (TR)	Kalman Residual (KR)	Abs(TR) -Abs(KR)
17:00	3.9	20.4	-16.5	30.0	13.7	16.3	4.8	34.1	-29.3
17:10	7.9	33.5	-25.5	141.3	84.6	56.7	134.4	118.1	16.3
17:20	128.0	86.1	41.9	-14.9	34.7	-19.8	92.9	120.8	-27.9
17:30	-195.1	-23.6	171.6	-178.5	-50.7	127.9	-392.8	-74.2	318.5
17:40	19.5	17.3	2.2	46.4	6.7	39.7	55.7	24.0	31.7
17:50	25.6	35.6	-10.0	-192.0	-66.5	125.5	-178.1	-31.0	147.1
18:00	-106.2	-9.6	96.6	14.6	-12.2	2.4	-94.8	-21.9	72.9
18:10	55.0	35.6	19.4	111.2	46.8	64.4	162.8	82.4	80.4
18:20	-220.6	-56.1	164.5	-243.8	-72.1	171.7	-472.5	-128.3	344.3
18:30	113.9	38.6	75.3	56.6	2.1	54.4	173.6	40.7	132.9
18:40	-83.3	-2.2	81.2	-96.7	-30.7	65.9	-179.7	-32.9	146.8
18:50	-29.9	2.2	27.7	114.9	41.0	73.9	89.6	43.2	46.4
19:00	71.9	43.9	27.9	-119.6	-25.1	94.5	-48.0	18.8	29.2
19:10	-167.6	-34.5	133.2	0.9	-2.3	-1.4	-162.2	-36.8	125.4
19:20	47.2	18.6	28.6	37.2	20.8	16.4	89.1	39.3	49.8
19:30	-144.5	-36.9	107.6	-44.9	-3.1	41.8	-186.5	-40.0	146.5
19:40	76.5	27.7	48.9	-40.8	-10.6	30.2	40.8	17.1	23.7
19:50	-6.9	20.2	-13.3	77.5	33.7	43.9	76.9	53.9	23.1
20:00	77.0	50.7	26.3	-62.6	-5.3	57.3	17.2	45.4	-28.2
20:10	-24.4	23.2	1.2	36.6	19.4	17.2	17.3	42.6	-25.3
20:20	-102.6	-18.8	83.8	-117.4	-32.6	84.9	-216.6	-51.3	165.3
20:30	22.7	13.5	9.2	3.3	-4.3	-1.0	34.5	9.2	25.3
20:40	-145.1	-40.8	104.4	-62.1	-19.7	42.5	-199.3	-60.4	138.9
20:50	-2.1	-6.8	-4.6	31.4	11.7	19.7	39.9	4.9	35.0
21:00	-62.1	-17.8	44.3	-86.7	-23.4	63.3	-139.9	-41.3	98.7
	Statistics of Absolute Values of Residuals:			Statistics of Absolute Values of Residuals:			Statistics of Absolute Values of Residuals:		
		Time Series	Kalman		Time Series	Kalman		Time Series	Kalman
	Mean	77.6	28.6	Mean	31.4	27.0	Mean	132.0	48.5
	StDev	62.5	19.6	StDev	62.1	22.8	StDev	111.4	33.1
	Min	2.1	2.2	Min	0.9	2.1	Min	4.8	4.9
	Max	220.6	86.1	Max	243.8	84.6	Max	472.5	128.3

Table 4: Comparison of Time Series and Kalman Predictions on I-70

Time	Driving Lane			Passing Lane			Total		
	Time Residual (TR)	Kalman Residual (KR)	Abs(TR) -Abs(KR)	Time Residual (TR)	Kalman Residual (KR)	Abs(TR) -Abs(KR)	Time Residual (TR)	Kalman Residual (KR)	Abs(TR) -Abs(KR)
6:00	-26.4	5.0	21.4	-54.0	0.8	53.2	-26.4	5.8	20.5
6:05	-25.3	1.9	23.4	19.9	29.3	-9.3	30.3	31.1	-0.8
6:10	-63.9	-14.4	49.5	-93.8	-21.4	72.5	-142.5	-35.8	106.7
6:15	-11.2	3.2	8.0	11.7	16.2	-4.5	52.0	19.3	32.7
6:20	74.6	43.6	31.1	-44.8	-2.0	42.8	57.2	41.6	15.6
6:25	-54.9	3.4	51.5	-31.8	0.2	31.6	-57.7	3.6	54.1
6:30	-2.1	11.1	-9.0	28.2	24.9	3.2	58.9	36.0	22.9
6:35	47.7	33.4	14.3	-52.3	-7.5	44.8	4.1	25.9	-21.8
6:40	-63.3	-4.8	58.5	57.7	31.8	25.8	16.9	27.0	-10.1
6:45	23.7	17.8	5.9	-59.9	-13.8	46.1	-41.3	4.0	37.3
6:50	-18.9	8.0	10.9	-32.8	-8.6	24.3	-31.9	-0.6	31.3
6:55	-44.7	-5.6	39.1	-2.3	7.5	-5.2	-21.7	1.9	19.7
7:00	81.9	41.2	40.7	-3.3	9.3	-6.0	97.5	50.5	47.0
7:05	-23.7	13.0	10.7	26.2	19.8	6.4	9.2	32.8	-23.5
7:10	20.7	21.0	-0.2	-41.4	-9.0	32.4	-25.9	12.0	13.9
7:15	-35.7	0.6	35.1	56.7	27.0	29.7	32.4	27.5	4.9
7:20	30.9	20.8	10.1	-18.4	-1.0	17.3	-1.0	19.8	-18.7
7:25	84.3	48.3	36.0	-22.8	-7.6	15.3	59.0	40.8	18.2
7:30	-71.0	-6.6	64.5	62.2	27.8	34.4	-8.7	21.2	-12.5
7:35	120.3	52.3	68.0	71.1	33.0	38.2	170.4	85.3	85.1
7:40	-17.6	14.8	2.8	93.1	35.9	57.2	26.5	50.7	-24.3
7:45	-35.0	-4.9	30.1	-112.9	-52.8	60.0	-212.7	-57.7	154.9
7:50	-6.2	1.1	5.2	92.7	24.7	67.9	95.0	25.8	69.2
7:55	87.9	41.3	46.6	58.1	22.8	35.4	113.8	64.1	49.8
8:00	76.6	48.7	27.9	57.1	17.7	39.4	85.6	66.4	19.2
	Statistics of Absolute Values of Residuals:			Statistics of Absolute Values of Residuals:			Statistics of Absolute Values of Residuals:		
		Time Series	Kalman		Time Series	Kalman		Time Series	Kalman
	Mean	45.9	18.7	Mean	48.2	18.1	Mean	59.1	31.5
	StDev	30.5	17.4	StDev	29.4	13.1	StDev	54.0	22.1
	Min	2.1	0.6	Min	2.3	0.2	Min	1.0	0.6
	Max	120.3	52.3	Max	112.9	52.8	Max	212.7	85.3

69 and I-70 are listed in Table 3 and Table 4, respectively.

Compared to the time series predictions, Table 3 shows that, on I-69, 20 out of the 25 Kalman predictions are more accurate for the driving lane, 22 out of the 25 Kalman predictions are more accurate for the passing lane, and 21 out of the 25 of the Kalman predictions are more accurate for the total traffic volumes of the two lanes. The percentage reductions in the means of the absolute residual values obtained by using the Kalman predictor are calculated for the driving lane, passing lane, and two-lane total, respectively, as: (77.6-28.6)/77.6=63.1%, (31.4-27.0)/31.4=14.0%, and (132.0-48.5)/132.0=63.3%. The reductions in the corresponding standard deviations are (62.5-19.6)/62.5=68.6%, (62.1-22.8)/62.1=63.3%, and (111.4-33.1)/111.4=70.3%.

Similarly, Table 4 indicates that, on I-70, 23 out of the 25 Kalman predictions are more accurate for the driving lane, 21 out of the 25 Kalman predictions are more accurate for the passing lane, and 18 out of the 25 of the Kalman predictions are more accurate for the total traffic volumes of the two lanes. The percentage reductions in the means of the absolute residual values obtained by using the Kalman predictor are calculated for the driving lane, passing lane, and two-lane total, respectively, as: (45.9-18.7)/45.9=59.3%, (48.2-18.1)/48.2=62.4%, and (59.1-31.5)/59.1=46.7%. The reductions in the corresponding standard deviations are (30.5-17.4)/30.5=43.0%, (29.4-13.1)/29.4=55.4%, and (54.0-22.1)/54.0=59.1%.

The applications of the two methods on I-69 and I-70 illustrate again that the Kalman method produced more accurate predictions than the time series method in terms of the number of more accurate predictions, and the reductions in the means and standard deviations of absolute values of residuals.

Prediction of Traffic

Congestion: Once the traffic capacity is known, the dynamic prediction of traffic flow rates discussed above constitutes a dynamic prediction of traffic congestion. As previously indicated, the average traffic capacity of four–lane freeways in Indiana is 1,767 pc/h/ln. Thus, the traffic congestion at this location can be predicted with the Kalman predictor method at each step of the prediction, according to the following criteria:

If $\hat{f}(t+1|t) < 1{,}767$ passenger cars per hour per lane, then no congestion at time t+1 is predicted;

If $\hat{f}(t+1|t) \geq 1{,}767$ passenger cars per hour, then congestion at time t+1 is predicted.

CONCLUSIONS

Capacity is defined in terms of the maximum rate of traffic flow that can be accommodated by a given traffic facility under prevailing conditions (TRB 2000). Traffic congestion occurs when traffic flow exceeds the capacity of the roadway. Consequently, during congestion vehicles travel at reduced speeds and with fluctuating traffic flow rates. Motorists endure considerably greater traffic delays under congested traffic conditions than under uncongested conditions. Based on the traffic data from the 18 Indiana WIM stations, the observed capacity values range from 1,489 to 2,006 pc/h/ln with an average value of 1,767 pc/h/ln on four-lane freeways and range from 1,463 to 2,039 pc/h/ln with an average value of 1,778 pc/h/ln on six-lane freeways.

Given the freeway capacity values, it was desired to develop methods for predicting traffic flow and congestion so that appropriate traffic control strategies could be applied to avoid traffic congestion and to reduce traffic delay. Such a method was developed in this study using the Kalman predictor in combination with the first-order autoregressive process of time series. The method provides greatly improved traffic flow predictions over using only the time series method. It predicts freeway traffic flow dynamically with each new traffic data observation. Dynamic traffic predictions with the developed model can be performed for individual lanes as well as for all the lanes of each travel direction. Therefore, the prediction model can be used as an efficient tool for traffic control. This study showed that a dynamic prediction of traffic flow rate with this prediction model would also constitute a dynamic prediction of traffic congestion if the traffic capacity was given.

Endnotes

1. The parameter values of the time series model and the Kalman predictor model are not compared because the two models are different and their parameters have different meanings. There is no basis to compare parameter values of the two models.

References

Ahmed, S., and A. Cook. "Application of Time-Series Analysis Techniques to Freeway Incident Detection." *Transportation Research Record* 841 (1982): 19-21.

Bowerman, B. L., and R.T. O'Connell. *An Applied Approach: Time Series and Forecasting.* Duxbury Press, 1979.

Bozic, S. M. *Digital and Kalman Filtering: An Introduction to Discrete-Time Filtering and Optimum Linear Estimation.* John Wiley & Sons, 1979.

Cryer, J. D. *The Time Series Analysis.* Duxbury Press, 1986.

Davis, C., N. Nihan, M. Hamed, and L. Jacobson. "Adaptive Forecasting of Freeway Traffic Congestion." *Transportation Research Record* 1287 (1990): 29-33.

Jiang, Yi. *Indiana Freeway Traffic Characteristics and Dynamic Prediction Of Freeway Traffic Flows, Final Report,* FHWA/INDOT/SPR-2121, Indiana Department of Transportation, 1999.

Lu, Jian. "Prediction of Traffic Flow by an Adaptive Prediction System." *Transportation Research Record* 1287 (1990): 54-61.

Minitab, Inc. MINITAB (Release 11) *Reference Manual,* 1996.

Okutani, I., and Y.J. Stephanedes. "Dynamic Prediction of Traffic Volume Through Kalman Filtering Theory." *Transportation Research,* Vol. 18B, No. 1 (1984): 1-11.

Persaud, B.N., and F.L.Hall. "Catastrophe Theory and Patterns in 30-Second Freeway Traffic Data—Implications for Incident Detection." *Transportation Research, Part A, General,* Volume 23A(2) (1989): 103-113.

Transportation Research Board (TRB). *Highway Capacity Manual.* National Research Council, Washington, DC, 1994.

Transportation Research Board (TRB). *Highway Capacity Manual.* National Research Council, Washington, DC, 2000.

Yi Jiang is an assistant professor in the Department of Building Construction Management at Purdue University. He holds a Ph.D. degree in Civil Engineering from Purdue University. His research interests include highway construction, bridge/pavement management systems, pavement performance modeling, and traffic engineering.