SHORT-TERM FORECASTS OF FREEWAY TRAFFIC VOLUMES AND LANE OCCUPANCIES, PHASE 1

Volume IV

WA-RD 288.4
TNW 93-05.4

Final Technical Report
June 1993

TRANSNOW
Transportation Northwest
University Transportation Centers Program
Federal Region Ten

Washington State Department of Transportation
Transit, Research, and
Intermodal Planning Division
in cooperation with the
United States Department of Transportation
Federal Highway Administration

TECHNICAL REPORT STANDARD TITLE PAGE

1. REPORT NO.	2. GOVERNMENT ACCESSION NO.	3. RECIPIENT'S CATALOG NO.
WA-RD 288.4 / TNW 93-05.4		

4. TITLE AND SUBTITLE	5. REPORT DATE
SHORT-TERM FORECASTS OF FREEWAY TRAFFIC VOLUMES AND LANE OCCUPANCIES, PHASE 1 — VOLUME IV	June 1993
	6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S)	8. PERFORMING ORGANIZATION REPORT NO.
Nancy L. Nihan and Jinde Zhu	
9. PERFORMING ORGANIZATION NAME AND ADDRESS	10. WORK UNIT NO.
Washington State Transportation Center (TRAC) University of Washington, JD-10 University District Building; 1107 NE 45th Street, Suite 535 Seattle, Washington 98105-4631	11. CONTRACT OR GRANT NO. GC8719, Task 15
12. SPONSORING AGENCY NAME AND ADDRESS	13. TYPE OF REPORT AND PERIOD COVERED
Washington State Department of Transportation Transportation Building, MS 7370 Olympia, Washington 98504-7370	Final technical report
	14. SPONSORING AGENCY CODE

15. SUPPLEMENTARY NOTES

This study was conducted in cooperation with the U.S. Department of Transportation, Federal Highway Administration.

16. ABSTRACT

The current project addressed two major weak points of the existing WSDOT ramp control system. One weak point in the system is the fact that it reacts to the problem (congestion), rather than preventing the problem. The other weak point in the system is its reliance on detector data that may be in error. Both of these problems can be minimized by developing methods to accurately predict short-term traffic data. By predicting the onset of congestion early enough, the ramp metering system can act to prevent or delay occurrence of the problem. Also, if a detector has failed or is malfunctioning, the data from the detector can be estimated from short-term predictions based on neighboring detectors.

At the beginning of the current project, the researchers had hoped that the same model would provide a basis for both forecasting congestion (for predictive ramp control) and replacing erroneous data (predicting actual values). However, the best congestion or breakdown flow forecaster (the pattern recognition method) does not provide a basis for data prediction. The best method for filling in missing detector data turned out to be multivariate time series analysis.

Several pattern recognition and time series models were tested for further development. In both cases, the simpler models turned out to be the best choices, and in both cases, further model testing and development were recommended.

The research on both model types continues in follow-up studies that are expected to lead to incorporation of these models in the new TSMC computer system.

17. KEY WORDS	18. DISTRIBUTION STATEMENT
ramp control, freeway management, traffic flow forecasts, forecast models, ramp metering	No restrictions. This document is available to the public through the National Technical Information Service, Springfield, VA 22616

19. SECURITY CLASSIF. (of this report)	20. SECURITY CLASSIF. (of this page)	21. NO. OF PAGES	22. PRICE
None	None	103	

Final Technical Report
Research Project GC8719, Task 15
Freeway and Ramp Real Time Forecasting

SHORT-TERM FORECASTS OF FREEWAY TRAFFIC VOLUMES AND LANE OCCUPANCIES, PHASE 1

VOLUME IV

by

Nancy L. Nihan
Professor and Principal Investigator

Jinde Zhu
Graduate Research Assistant

Department of Civil Engineering, FX-10
University of Washington
Seattle, Washington 98195

Washington State Transportation Center (TRAC)
University of Washington, JD-10
University District Building
1107 NE 45th Street, Suite 535
Seattle, Washington 98105-4631

Washington State Department of Transportation
Technical Monitors

David Peach
State Traffic Engineer

Leslie N. Jacobson
Traffic Systems Engineer

Prepared for

Washington State Transportation Commission
Washington State Department of Transportation
Olympia, Washington 98504-7370

Transportation Northwest (TransNow)
135 More Hall, FX-10
Seattle, Washington 98195

and in cooperation with
U.S. Department of Transportation
Federal Highway Administration

June 1993

DISCLAIMER

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE: SUMMARY

Introduction

In operating a real-time, traffic-responsive control system, such as a freeway ramp control system, most of the decisions made by the system will depend on its prediction of traffic during upcoming short-term periods. Regardless of its type, the prediction of such a system must be as accurate as possible to be effective, and to allow the traffic-responsive control system to handle the prediction logic in a satisfactory manner. If traffic congestion on the freeway can be foreseen, then congestion can be avoided in one of several ways. For example, consider a short section of freeway with on-ramps and off-ramps, a lane occupancy at the section's downstream boundary greater than 18 percent, and a positive difference between inflowing and outflowing traffic (these criteria are from the Guidelines of the Traffic Systems Management Center of WSDOT), indicating that traffic flow is at or above capacity. The controller concludes that a "bottleneck" has formed and reduces the entry rates at on-ramps upstream from the section's downstream boundary. In this way, potentially heavy congestion can be avoided and operation of the freeway improved. Thus, optimal freeway traffic control can be obtained if future volume and occupancy is predictable with reasonable accuracy.

Basically, there are two classes of predictors: parametric and nonparametric. Most predictors are parametric; in urban traffic control systems, these can also be classified as second generation and third generation predictors (1-5). Recent studies have used Box-Jenkins type analyses of time series data (7,8,9,10,11,12), spectral analysis (13), Kalman filtering (14,15,16,17), automatic control concepts (18), and adaptive prediction system analysis (19), to forecast freeway traffic flow. Other researchers have used nonparametric methods to forecast freeway traffic volume (23).

This research paper deals with the prediction of freeway traffic volumes and occupancies using parametric methods, including the following steps:

1) Four existing parametric methods will be used to forecast volumes and occupancies. These models include the adaptive prediction system, the double exponential smoothing method, the exponential smoothing with adaptive response method, and the Box-Jenkins method. Each of these four models uses just one time series to forecast its own future.

2) A new model will be developed using Fourier's transformation or cross spectrum to forecast the downstream volumes (29). This model can analyze the lags between upstream or on-ramp volume time series and downstream volume series. Based on these lags, a model can be constructed which uses the upstream volumes to forecast the downstream volumes; its parameters can be obtained by the ordinary least square method and recursive method.

3) A detailed comparison of both volume and occupancy forecasting results using these different models will be made, and the interpretation of the results and the drawbacks of each model will be discussed.

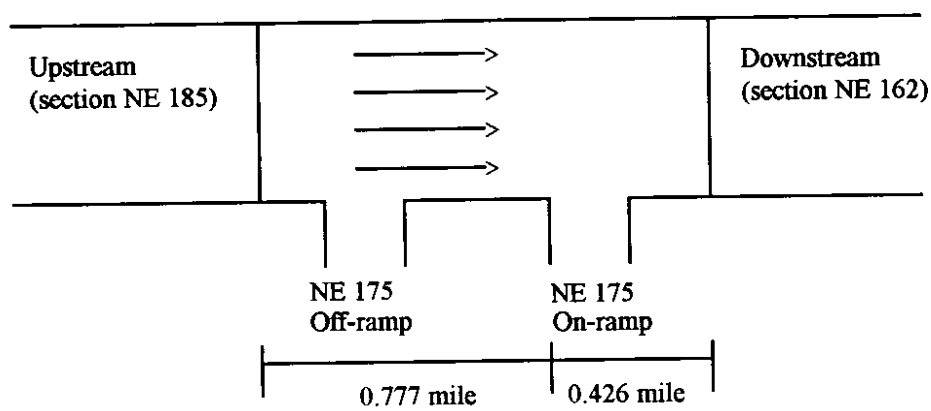
4) A conclusion will be drawn from our research and recommendations presented.

CHAPTER TWO

Research Design

To analyze and forecast volumes in this paper, a segment between the NE 185th St.section and the NE 162nd St. section on Interstate 5 in northeastern Seattle was chosen as a study site. This segment has an off-ramp and an on-ramp at NE 175th St. The distance from the upstream section to the on-ramp is 0.777 miles, and the distance from the on-ramp to the downstream section is 0.426 miles.

Figure 2.1 - Study Site



The volume time series for these sections at both the off-ramp and the on-ramp were collected on February 23, 1989, between 6:00 a.m. and 8:00 a.m. For the NE 185th St. section and the NE 162nd St. section, the volumes were aggregated over all four lanes, the data interval being one minute. There are 122 data points for each volume time series. The study data are detailed in Appendix C.

To analyze and forecast lane occupancies, a study data set, or occupancy time series, was also collected at NE 185th St. section during the same time. These 122 data points are also presented in Appendix C.

Both volume time series and occupancy time series were broken into two parts: 1) 102 data points, and 2) 20 data points. In this report, the first 102 data points will be used to build four existing models (introduced late), and the last 20 data points will be used to update the forecasting and to calculate the three criteria (introduced late) for each model.

To build the model developed in this paper for forecasting downstream volumes, the first 102 data points of the upstream time series, on-ramp volume time series, and the downstream volume time series were used; the last 20 data points of upstream and on-ramp volume time series were used to forecast the last 20 downstream volumes. These forecast and actual downstream volumes will serve to calculate the three criteria.

The four existing models used to forecast volumes and occupancies included: 1) the Box-Jenkins model, a well-known and successfully model used in traffic forecasting (26), 2) a double exponential smoothing model (27), 3) an exponential smoothing with adaptive response (28), and 4) an adaptive prediction system model introduced by Lu (19). The Box-Jenkins model and double exponential smoothing model are often used in business forecasting. Each of these models uses one or more current data points of one time series to forecast its own future. The assumption made about these models is that there exists a correlation between current and future data, and so the current data can be used to forecast its future. However, if this assumed correlation does not exist, and the future data does not occur as expected, then the prediction of these models would be incorrect.

In our study section, we know what the downstream volumes are from the upstream volumes, and the volumes at the NE 162nd St. section are from the NE 185th St. and NE 175th St. on-ramp volumes. Therefore, another forecasting model could be built by analyzing the relationship between these upstream volumes and the downstream volumes so that this model can use the upstream volumes to forecast the downstream volumes. Our objective was to forecast the future downstream volumes based on the current upstream volumes; therefore, we chose upstream sections or stations with the

thought that the current traffic volumes at these sections or stations at interval t would arrive at the downstream section no earlier than $t+1$. If the time interval is shorter, the upstream volume cannot forecast the downstream volume, as they would appear to be simultaneous. Therefore, we should analyze the vehicle travel time from the NE 185th St. section and the on-ramp NE at 175th St. to the downstream NE 162nd St. section. Since vehicles travel at varying speeds, the travel time from the upstream section to the downstream section will also vary.

A new model, which uses the Fourier's transformation to analyze the travel time or lags is developed in this report (29). After the lags are obtained, the model can be constructed, and the ordinary least square method and the recursive method will be used to obtain the parameters of the obtained model. Then, the downstream volume forecast by this model, and the three criteria listed below, can be calculated. Because of the difficulty in expressing downstream occupancies as a function of upstream occupancies, only the four existing models will be used here to forecast occupancies.

In order to compare both volume and occupancy forecasting results obtained by the different models, three criteria will be used: 1) mean relative absolute error, which indicates the error as a fraction of the measurement, 2) relative square root absolute error, which penalizes large prediction errors, and 3) maximum relative absolute error. The detailed formula is given in Chapter Five.

The computer software used to build these models includes BMDP (30) in the MAX system, which was used to build the model developed in this paper, IDA (31) in the MAX system to build time series models based on the Box-Jenkins time series analysis approach, and MINITAB in the CYBER system to calculate the three criteria and make some general calculations.

FORTRAN programs in the CYBER system were used to estimate the adaptive prediction system model, double exponential smoothing model and exponential smoothing with adaptive response model.

CHAPTER THREE

Existing Models

CHAPTER REVIEW

Four existing forecasting models will be introduced in this chapter, these are:

1. the adaptive prediction system model,
2. the double exponential smoothing model,
3. the exponential smoothing with adaptive response model, and
4. the Box-Jenkins model.

All of these models use univariate time series to forecast its own future. These models will be discussed in the following section.

ADAPTIVE PREDICTION SYSTEM MODEL

The adaptive prediction system model, introduced by Lu, can be considered a dynamic parameter estimation model; i.e., the traffic conditions predicted at time step k are the function of both past traffic conditions at time step k-1, k-2, ..., k-M (where $M<k$, and M and k are positive integers) and a set of parameters are estimated by the adaptive prediction model (19). If this adaptive prediction model is viewed as a system, then its structure is alterable in such a way that its behavior or performance improves through contact with its environment. In fact, this model is the steepest type of adaptive algorithm (24).

Assuming that the predicted future value at step k, $\hat{V}(k)$, will be the linear weighted combination of n previous values, we have

$$\hat{V}(k) = \sum_{j=0}^{n} W_{jk} \times V(k-s-j) \qquad \text{where } (k = s+1,\ s+2,\ \ldots)$$

or

$$\hat{V}(k) = \mathbf{V}_{k-s}^{\mathrm{T}} \mathbf{W}_k \tag{3.1}$$

where

$\hat{V}(k)$ is a predicted value at time k,

W_{0k}, W_{1k}, ..., W_{jk}, ..., W_{nk} are weights,

$V(k-s)$, $V(k-s-1)$, ..., $V(k-s-j)$, ..., $V(k-s-n)$ are the previous values, and where

$\mathbf{W}_k = [W_{0k}, W_{1k}, \ldots, W_{jk}, \ldots, W_{nk}]^{\mathrm{T}}$, and

$\mathbf{V}_{k-s} = [V(k-s), V(k-s-1), \ldots, V(k-s-j), \ldots, V(k-s-n)]^{\mathrm{T}}$, where

k is the time step.

After some mathematical derivation, and assuming that this time series is stationary (19), we have

$$\mathbf{W}_{k+1} = \mathbf{W}_k + 2\mu e(k)\mathbf{V}_{k-s} \tag{3.2}$$

where

$e(k) = V(k) - \hat{V}(k)$, and

μ is the step size from k to $k+1$.

Generally, a small μ and n result in good stability but slow convergence speed, so there is a trade-off between stability and response performance (19). In our study, n=10 was chosen, so that the forecasting relative mean absolute error would be minimal.

DOUBLE EXPONENTIAL SMOOTHING MODEL

The double exponential smoothing forecast can follow a linear trend. Brown has demonstrated that the steady-state response of exponential smoothing to a linear trend has a constant lag of $(1-\alpha)/\alpha$ (12). This model can be expressed as

$$\hat{V}(t+k) = a_t + b_t(k+1) \tag{3.3}$$

where

$$a_t = 2T_t^{[1]} - T_t^{[2]}$$

$$b_t = \frac{\alpha}{1-\alpha}\left(T_t^{[1]} - T_t^{[2]}\right)$$

k is the time step,

$T_t^{[1]}$ is single exponential smoothing, which is given by

$$T_t^{[1]} = \alpha V(t-1) + (1-\alpha) T_{t-1}^{[1]}$$

$T_t^{[2]}$ is double exponential smoothing, which is given by

$$T_t^{[2]} = \alpha T_t^{[1]} + (1-\alpha) T_{t-1}^{[2]}.$$

α is a constant, which we evaluated from 0.1 to 0.9 in increments of 0.1.

EXPONENTIAL SMOOTHING WITH ADAPTIVE RESPONSE

In single exponential smoothing and double exponential smoothing models, the parameter α is always a constant, so using the adaptive approach has been suggested by Trigg and Leach to adjust this smoothing constant α (28). The following is the adaptive approach proposed by Trigg and Leach:

$$\hat{V}(t) = V(t-1) + \alpha(t-1)\left[V(t-1) - \hat{V}(t-1)\right] \tag{3.4}$$

where

$$\alpha(t) = \left|\frac{\mathrm{SE}(t)}{\mathrm{SAE}(t)}\right| \qquad 0 < a < 1,$$

$$\mathrm{SE}(t) = \tau e_t + (1-\tau)\mathrm{SE}(t-1),$$

$$\mathrm{SAE}(t) = \tau|e_t| + (1-\tau)\mathrm{SAE}(t-1),$$

$$e_t = V(t) - \hat{V}(t), \text{ and}$$

τ is a constant.

Both τ and the initial α were evaluated from 0.1 to 0.9 in increments of 0.1 in our research.

BOX-JENKINS UNIVARIATE TIME SERIES MODEL

The Box-Jenkins univariate time series method (26) has been used very successfully in many fields, and has been proven very accurate in traffic forecasting (7, 8, 9, 10,11,12).

Many real-time series, unless they are seasonal, can be represented by the general class of linear models shown below. Actually, many types of time series in traffic

engineering are seasonal, i.e., monthly or weekly freeway volumes. In this study, since only short-interval and short-term forecasts and volumes covering just one or two hours are being considered, this kind of time series is not considered seasonal.

$$\Phi_p B(1-B)^d\left[V(t)-\mu\right]=\Theta_q B\cdot a_t \tag{3.5}$$

where

p, d, q are nonnegative integers,

μ is the mean of the series,

$\Phi_p B$ is the autoregressive operator of order p or

$1-\Phi_1 B-\Phi_2 B^2-\cdots-\Phi_p B^p$,

$\Theta_q B$ is the moving average operator of order q or

$1-\Theta_1 B-\Theta_2 B^2-\cdots-\Theta_q B^q$, and

a_t is random disturbance, assumed to be independently distributed as $N(0,\sigma_a^2)$.

The model in equation (3.5) is the autoregressive integrated moving average (ARIMA) model of order (p, d, q).

ARIMA models are fitted to a particular data set mainly by a three-stage iterative procedure: plots and identification, estimation, and diagnostic check. From the plots in the first stage, we can find whether the time series is nonstationary with respect to variance, or mean, or both.

If the time series is nonstationary with respect to the variance, one should use some transformation, such as log, square, square root, etc., to reduce the heteroskedascity. If the time series is nonstationary with respect to mean, then generally performing first differencing or second differencing may be enough to change the series into a stationary series. If the time series is nonstationary with respect to both variance and mean, then both transformation and differencing will need to be done to change the time series into a stationary series. After transformation and differencing, the autocorrelations and partial autocorrelations must be calculated, to identify p, d, and q by

comparing these autocorrelations and partial autocorrelations with those of a basic stochastic process. The sample autocorrelation function is given by

$$R_k = \frac{\sum_{t=0}^{n-k}[V(t)-MV][V(t+k)-MV]}{\sum_{t=0}^{n}[V(t)-MV]^2} \tag{3.6}$$

where R_k is the autocorrelation coefficient at lag k,

MV is the sample mean, and

n is the number of observations.

The autocorrelation function R_k implies how the volume at time i is correlated with the volume at time i-k; that is, how long a volume in the time series affects the state of that time series in the future.

Generally, the autocorrelation function of a moving average process of order q has a cutoff after lag q, but a partial autocorrelation function that tails off. Conversely, the partial autocorrelation function of order p will have a cutoff after lag p, but an autocorrelation function that tails off. For a mixed process, both autocorrelation function and partial autocorrelation function will tail off. Failure of the autocorrelation function to tail off rapidly suggests that differencing is necessary.

After determining p, d, and q, the parameters are estimated by nonlinear least square techniques; the last step is to check the model's goodness of fit. If the model is satisfactory, then the residuals should be white noise, or they are not correlated. The adjusted Box-Pierce test, which can be used to test whether the residuals are white noise, can be expressed as follows:

$$Q_{ad} = n(n+2)\sum_{j=1}^{m}\left[R_j(a)/(n-j)\right] \tag{3.7}$$

where

n is the number of observations

$R_j(a)$ is the residual autocorrelation for lag j, and

m is the number of calculated residual autocorrelation coefficients, about $n/4$.

Q_{ad} is distributed as a chi-square variable with (k-p-q) degrees of freedom; therefore, if this Q_{ad} is less than the tabulated chi-square value with (k-p-q) degrees of freedom at a 95 percent confidence interval, then the residuals are white noise, and the obtained model is satisfactory. If this is not the case,, we should estimate the model again, using the same three-stage iterative procedure.

To forecast occupancies, all of the above models can be used, we just need to change $\hat{V}(t)$ and $V(t)$ into $\hat{O}(t)$ and $O(t)$, where $\hat{O}(t)$ is a predicted occupancy at time t, and $O(t)$ is an actual occupancy at the same time.

In Chapter Five, Results, these four models are assessed using the actual volume and occupancy data, and the results and comparisons presented.

CHAPTER FOUR

New Model Development

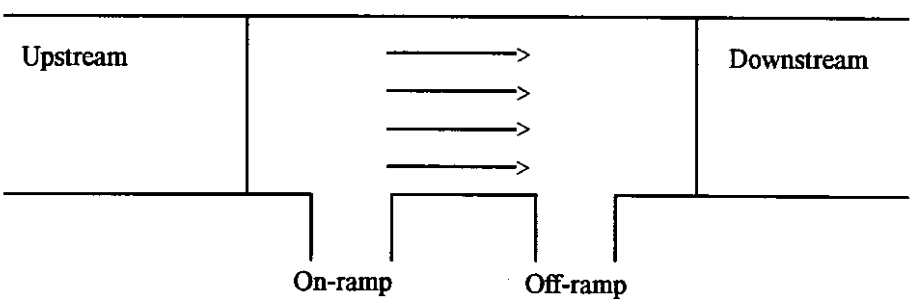
CHAPTER REVIEW

In this chapter, Fourier's transformation of the cross correlation between upstream, or on-ramp, volumes and the downstream volumes are used to analyze the relationships (lags) between origins (O), upstream and on-ramp volumes, destinations (D), and downstream and off-ramp volumes. (29) The recursive ordinary least square method is introduced to build the eventual forecasting model. The data interval requirement is also discussed this chapter.

TIME DOMAIN MODEL

First, let us look at the site example in figure 4.1.

Figure 4.1 - Site Example



Traffic flow from upstream travels straight to the downstream, or exits at the off-ramp. Similarly, traffic flow from the on-ramp travels to the downstream, or exits at the off-ramp. If no congestion exists between the upstream and the downstream sections, the inflowing volumes from the upstream and from the on-ramp should equal total outflowing volumes, which will travel through the downstream section or exit at the off-ramp. We should note that, since vehicles travel at varying speeds, the travel time from the on-ramp or upstream to the off-ramps or downstream will not be constant. Our procedure was 1)

to develop the time domain model, 2) to transform the time domain model into frequency domain model, or cross spectrum model, and 3) to analyze the downstream volume. This same procedure can also be applied to analyzing the off-ramp volume.

GENERAL ASSUMPTIONS

To derive the time domain model, we assumed that

(1) the time for all vehicles to travel from upstream to downstream is the same, $T1$ minutes,

(2) the time for all vehicles to travel from the on-ramp to downstream is also the same, $T2$ minutes,

(3) the portion of vehicles from upstream to downstream is $b_1(T1)$, and

(4) the portion of vehicles from the on-ramp to downstream is $b_2(T2)$.

If the volume of upstream is $V_{up}(t)$, and the volume of on-ramp is $V_{on}(t)$, then the volume of the downstream is

$$V_d(t) = b_1(T1)V_{up}(t-T1) + b_2(T2)V_{on}(t-T2)$$

where

$V_d(t)$ is the volume of the downstream at time t,

$V_{up}(t-T1)$ is the volume of the upstream at time $t-T1$, because of the travel time $T1$, and

$V_{on}(t-T2)$ is the volume of the on-ramp at time $t-T2$, because of the travel time $T2$.

Obviously, since vehicles travel at different speeds, the travel time for all vehicles will not be the same, as is assumed in the above models. Suppose the travel time from upstream to downstream is from a minimum $T1_{min}$ minutes to a maximum $T1_{max}$ minutes, and the travel time from the on-ramp to downstream is from $T2_{min}$ minutes to $T2_{max}$ minutes. Suppose, also, that the number of vehicles from upstream is $b_1(T1)$, where $T1$ is from $T1_{min}$ to $T1_{max}$, and the portion of vehicles from the on-ramp is $b_2(T2)$, where $T2$ is

from $T2_{\min}$ to $T2_{\max}$ In addition, if measurement errors exist, an error term, $Z(t)$, should be added to $V_d(t)$. Hence, we can obtain the following formula:

$$V_d(t) = \sum_{T1=T1_{\min}}^{T1_{\max}} b_1(T1) V_{up}(t-T1) + \sum_{T2=T2_{\min}}^{T2_{\max}} b_2(T2) V_{on}(t-T2) + Z(t)$$

For convenience, we can assume further that

$$b_1(T1) = \begin{cases} b_1(T1), & T1_{\min} < T1 < T1_{\max} \\ 0, & \text{otherwise} \end{cases}, \text{ and}$$

$$b_2(T2) = \begin{cases} b_2(T2), & T2_{\min} < T2 < T2_{\max} \\ 0, & \text{otherwise} \end{cases}.$$

Therefore, we have the following general model:

$$V_d(t) = \sum_{T1=-\infty}^{+\infty} b_1(T1) V_{up}(t-T1) + \sum_{T2=-\infty}^{+\infty} b_2(T2) V_{on}(t-T2) + Z(t). \qquad (4.1)$$

Since $V_d(t)$, $V_{up}(t-T1)$, $V_{on}(t-T2)$, $b_1(T1)$, and $b_2(T2)$ are the functions of time, t, that is this model is in the time domain, so it can be called the "time domain model."

From the above time domain model, we can see that if the ranges of $T1$ and $T2$, and the functions $b_1(T1)$ and $b_2(T2)$ can be determined, then we can use this time domain model to predict the downstream volume, assuming the expected mean of error is zero, that is

$$E[Z(t)] = 0.$$

Two problems arise here: 1) how to obtain the lags (the travel time difference from the upstream section or on-ramp to the downstream section), and 2) how to get coefficients $\mathbf{b}_1$ and $\mathbf{b}_2$ (defined in 4.4) at these different lags. From this time domain model, however, we cannot obtain any useful information about $\mathbf{b}_1(T1)$ and $\mathbf{b}_2(T2)$. In the following sections, these two problems are analyzed and solved individually.

LAG ANALYSIS AND FREQUENCY DOMAIN MODEL

From the modern time series theory, we know that the cross correlation function is a very important factor in determining the multivariate time series model. The cross correlation function can be written as follows:

$$C_{T1} = \frac{\sum_{t=1}^{n-T1}[X1(t) - MX1][X2(t+T1) - MX2]}{\sqrt{V(X1)V(X2)}} \qquad (33) \qquad (4.2)$$

where

C_{T1} is the cross correlation function between $X1$ and $X2$ at lag $T1$,

$T1$ is the lag, which is a positive integer or zero,

$X1(t)$ is a time series,

$X2(t)$ is also a time series,

$MX1$ and $V(X1)$ are the expected mean and variance of $X1$, and

$MX2$ and $V(X2)$ are the expected mean and variance of $X2$.

If C_{T1} is statistically significantly different from 0, then this means $X1$ and $X2$ are correlated at the lag $T1$. Here, "statistically significantly greater than 0" means that the t-ratio of this cross correlation should be significantly greater than 0 at a certain confidence level, say 95 percent. If there is more than one $T1$, at which the C_{T1}'s are statistically significantly different from 0, then it means that $X1$ and $X2$ are correlated at all of these lags. If there are more than two time series, the same procedures can be applied to analyze the lags between any two time series. After the determination of the lag $T1$, the multivariate time series model can be built. Our problem, determining the lags $T1$ and $T2$, is similar. If we calculate the cross correlation functions of the upstream, or on-ramp volumes, and the downstream volumes, we can determine the lags at which the cross correlation coefficients are statistically significantly greater than 0. Here, these lags mean the travel time from the upstream section, or the on-ramp, to the downstream section. If there is more than one lag at which the cross correlation coefficients differ from 0, this implies that the travel time for different vehicles is different; these non-zero cross

correlation coefficients mean that at these lags not one, the upstream volumes are correlated with the downstream volumes.

Generally, however, the cross correlation coefficients at adjacent lags are correlated, which will spread the cross correlations to a wide range of lags (22). But, in fact, we have high correlations at only a few lags. If we use these obtained lags to build the forecasting model, we cannot expect the predicted results to be very accurate.

Another alternative method of analyzing the lags is to use the Fourier's transformation of the cross covariance to determine the lags, or phase delays, between two time series at every frequency (33). Because Fourier's transformations are not correlated at different frequencies, this transformation is used in this report to analyze the lags (29).

Fourier's transformation of the covariance can be expressed as

$$F(f) = \int_{-\infty}^{+\infty} C(T1) e^{-i2\pi f \cdot T1} dT1 \tag{4.3}$$

where i is a complex notation, $i^2 = -1$, and

$e^{-i2\pi f \cdot T1} = \cos(2\pi f \cdot T1) - \sin(2\pi f \cdot T1)$.

Because $F(f)$ is the complex function of f, $F(f)$ can be expressed in polar form as follows:

$$F(f) = |F(f)| \times e^{iW(f)} \tag{4.4}$$

the real non-negative function $|F(f)|$ is called the gain, and the real function $W(f)$ is called the phase-shift. If we look at the inverse Fourier's transform given by

$$C(T1) = \int_{-\infty}^{+\infty} F(f) e^{-i2\pi f \cdot T1} df \tag{4.5}$$

We can see that $C(T1)$ may be considered to be a continuous sum of periodic components $\exp(i2\pi f \cdot T1)$ (For a detailed proof of this equation refer to the text by Don Percival. (29)). Replacing $F(f)$ in equation (4.5) by equation (4.4), we have

$$C(T1) = \int_{-\infty}^{+\infty} |F(f)| e^{-i(2\pi f \cdot T1 + W(f))} df \tag{4.6}$$

Thus, we can see that

$|F(f)|\, df$

represents the amplification of all periodic components that are inside the band (f, $f+df$) and have common phase-shift $W(f)$ at time $T1=0$.

Accordingly, the phase-delay, t_d is determined by

$$e^{-i(2\pi f \cdot T1 + W(f))} = 1$$

or

$$W(f) + 2\pi f \cdot T1 = 0$$

which gives

$$t_d = T1 = \frac{W(f)}{2\pi f} \tag{4.7}$$

where

$$-\infty < f < +\infty, \ f \neq 0. \ (34)$$

Because $C(T1)$ is real, from the equation (4.6) we have

$$C(T1) = \int_{-\infty}^{+\infty} |F(f)| \cos(2\pi f \cdot T1) df + i \int_{-\infty}^{+\infty} |F(f)| \sin(2\pi f \cdot T1) df$$

so

$$\int_{-\infty}^{+\infty} |F(f)| \sin(2\pi f \cdot T1) df = 0$$

therefore

$$C(T1) = \int_{-\infty}^{+\infty} |F(f)| \cos(2\pi f \cdot T1) df \tag{4.8}$$

From the above equation, we know that only if

$$W(f) + 2\pi f \cdot T1 = 0 \quad \text{or} \quad 2\pi k, \quad \text{where } k = \pm 1, \ \pm 2, \ \ldots, \ \pm N$$

that is

$$T1 = -\frac{W(f)}{2\pi f} \quad \text{or} \quad -\frac{2\pi k - W(f)}{2\pi f},$$

$$|F(f)| \cos(2\pi f \cdot T1) = |F(f)|.$$

Otherwise,

$$|F(f)| \cos(2\pi f \cdot T1) < |F(f)|.$$

Therefore, from equation (4.8), we can see that if for any f, $t_d = -W(f)/2\pi f$, is constant or $W(f) = \alpha \times f$, where a is constant, then obviously $C(T1)$ will reach its maximum, this means the lag or delay between these two time series is just t_d. If the range of t_d is small, then we can see that because

$$t_d \approx -\frac{W(f)}{2\pi f}$$

that is

$$2\pi f \cdot t_d + W(f) \approx 0$$

then

$$\left|\cos\left(2\pi f \cdot t_d + W(f)\right)\right| \approx 1$$

$$\left|F(f)\right|\cos(2\pi f \cdot k1) \approx \left|F(f)\right|$$

so $C(T1)$ will still be large during this range. This means the correlation between the two time series is large during this range or lag interval.

From the above discussion, it is clear that the $F(f)$'s are not correlated at these different frequencies; therefore, the phase-delay or t_d will be used to determine the lags in this paper.

MODEL DEVELOPMENT

So far, we have solved the first problem of determining the lags. Next, we will address the second problem, how to get coefficients b_1 and b_2 at the different lags.

From the general assumption made in the beginning of this chapter, we know

$$V_d(t) = \sum_{T1=T1_{\min}}^{T1_{\max}} b_1(T1) V_{up}(t - T1) + \sum_{T2=T2_{\min}}^{T2_{\max}} b_2(T2) V_{on}(t - T2) + Z(t).$$

If we arrange the above equation into the following form,

$$V_d(t) = \mathbf{V}(t)\,\mathbf{B} + Z(t) \tag{4.9}$$

where

$$\mathbf{B} = \left[b_1(T1_{\min}),\ \ldots,\ b_1(T1_{\max}),\ b_2(T2_{\min}),\ \ldots,\ b_2(T2_{\max})\right]^{\mathrm{T}} = (\mathbf{b}_1 \quad \mathbf{b}_2)^{\mathrm{T}},$$

$$\mathbf{b}_1 = \left[b_1(T1_{\min}),\ \ldots,\ b_1(T1_{\max})\right],$$

$$\mathbf{b}_2 = \left[b_2(T2_{\min}),\ \ldots,\ b_2(T2_{\max})\right],$$

$$\mathbf{V}(t) = \left[V_{up}(t - T1_{\min}),\ \ldots,\ V_{up}(t - T1_{\max}),\ V_{on}(t - T2_{\min}),\ \ldots,\ V_{on}(t - T2_{\max})\right]$$

$T1_{\min}$, $T1_{\max}$, $T2_{\min}$, and $T2_{\max}$ are as defined in section 4.2, then we can find that if we know the lags, $T1_{\min}$, $T1_{\max}$, $T2_{\min}$, and $T2_{\max}$, the above equation is just a linear equation, and **B** is a coefficient vector, so we can use the ordinary least square method to obtain these coefficients, $\hat{\mathbf{b}}_1$ and $\hat{\mathbf{b}}_2$; that is,

$$\hat{\mathbf{b}}_d = \left[\sum_t V_d(t)\mathbf{V}^{\mathrm{T}}(t)\right]\left[\sum_t \mathbf{V}^{\mathrm{T}}(t)\mathbf{V}(t)\right]^{-1} \tag{4.10}$$

where

$\left[\sum_t \mathbf{V}^{\mathrm{T}}(t)\mathbf{V}(t)\right]^{-1}$ is the inverse matrix of $\left[\sum_t \mathbf{V}^{\mathrm{T}}(t)\mathbf{V}(t)\right]$.

In the above derivation of the model, we use the ordinary least square method to obtain the coefficients, $\hat{\mathbf{b}}_1$ and $\hat{\mathbf{b}}_2$. However, it is time-consuming to calculate the inverse of the matrix; therefore, we can use an adaptive method to calculate these coefficients. If we let

$$\mathbf{A}(k) = \sum_{t-1}^{k} \mathbf{V}^{\mathrm{T}}(t)\mathbf{V}(t)$$

then we can rewrite equation (4.10) as follows:

$$\hat{\mathbf{b}}_d(k) = \hat{\mathbf{b}}_d(k-1) + \mathbf{A}^{-1}(k)\mathbf{V}^{\mathrm{T}}(k)\left[V_d(k) - \mathbf{V}(k)\hat{\mathbf{b}}_d(k-1)\right].$$

If we let

$$\mathbf{K}(k) = \mathbf{A}^{-1}(k)\mathbf{V}^{\mathrm{T}}(k)$$

$$\mathbf{Q}(k) = \mathbf{A}^{-1}(k)$$

then

$$\hat{\mathbf{b}}_d(k) = \hat{\mathbf{b}}_d(k-1) + \mathbf{K}(k)\left[V_d(k) - \mathbf{V}(k)\hat{\mathbf{b}}_d(k-1)\right] \tag{4.11}$$

$$\mathbf{K}(k) = \frac{\mathbf{Q}(k-1)\mathbf{V}^{\mathrm{T}}(k)}{1+\mathbf{V}(k)\mathbf{Q}(k-1)\mathbf{V}^{\mathrm{T}}(k)} \tag{4.12}$$

$$\mathbf{Q}(k) = \mathbf{Q}(k-1) - \frac{\mathbf{Q}(k-1)\mathbf{V}^{\mathrm{T}}(k)\mathbf{V}(k)\mathbf{Q}(k-1)}{1+\mathbf{V}(k)\mathbf{Q}(k-1)\mathbf{V}^{\mathrm{T}}(k)} \tag{4.13}$$

The derivational details can be seen in Appendix B.

Since for every k, $\hat{\mathbf{b}}_d(k)$ can be updated, the equatic (4.11)-(4.13) are called recursive least square estimates.

The recursive estimators (4.11)-(4.13) require values $\mathbf{Q}(0)$ and $\hat{\mathbf{b}}_d(0)$ to start recursion; therefore, if we have previous data, we can calculate the initial values $\mathbf{Q}(0)$ and $\hat{\mathbf{b}}_d(0)$.

DATA INTERVAL ANALYSIS

The researchers' objective here is to use the upstream and on-ramp volumes to forecast the downstream volume, i.e., using the current upstream and on-ramp volumes to forecast the downstream volume at the next interval. We will assign T to the data interval, $T1_{\min}$ (as defined before) to the minimum travel time from the upstream section to the downstream section, and $T2_{\min}$ (as defined before) to the minimum travel time from the on-ramp to the downstream section. Vehicles crossing the upstream section during the time interval (t, t+T) will begin to arrive at the downstream section at t+$T1_{\min}$; likewise, vehicles entering from the on-ramp will begin to arrive at the downstream +section at $tT2_{\min}$. If we want to forecast the downstream volumes as accurately as possible, we should choose an interval T when current upstream and on-ramp volumes just begin to arrive at the downstream section at the next interval or later, tha is,

$$T \le T1_{\min} \quad \text{and} \quad T \le T2_{\min}$$

i.e.,
$$T \le \text{minimum}\left(T1_{\min}, T2_{\min}\right). \tag{4.14}$$

Therefore, the data interval should be less than or equal to the minimum travel time from the upstream section and on-ramp.

If the interval is too small, the variation of volumes will be large, and the number of lags will be large, and it will still be difficult to forecast the downstream volumes. Therefore, we should choose the interval

$$T_{\text{optimal}} = \min(T1_{\min}, T2_{\min}). \tag{4.15}$$

CHAPTER FIVE

Results

In this chapter, volume and occupancy study data are described, and the volume forecasting results from the five models introduced and developed in the previous two chapters are presented. These results are compared to the three criteria stated subsequently. The interpretation of these results are then discussed. Finally, the occupancy forecasting results from the four existing models are given, and a comparison will be made against these three criteria.

DATA DESCRIPTION

For the purpose of this study, the traffic volume data used were collected by detectors from one I-5 freeway segment at several sections between 185th St. NE and 162nd St. NE in Seattle, between 6:00 and 8:00 a.m., on February 23, 1989. The interval of the data is 1 minute, with a total of 122 data points for each section.

The data set of 122 data points at the NE 162nd section was used in this study, the first 102 data points were used to build the four models introduced in Chapter Three, and the last 20 data points were used to update forecasting. (For the model developed in Chapter Four, the upstream and/or on-ramp volumes are needed.)

The study occupancy data were also collected at the NE 185th St. section during the same period. These 122 data points were divided as follows: the first 102 data points to build the four models, and the last 20 to update forecasting.

The adaptive prediction system model uses the previous ten to twenty data points to forecast the model's future. These previous data points to be used for forecasting are evaluated and obtained so that the forecasting relative mean error are minimal. If this number equaled ten, then the previous ten data points were used to forecast the next one, that is, ten data points from 93rd to 102nd forecast the 103rd value. The coefficients (W_t) were then updated, and these updated parameters and the ten data points from the 94th to

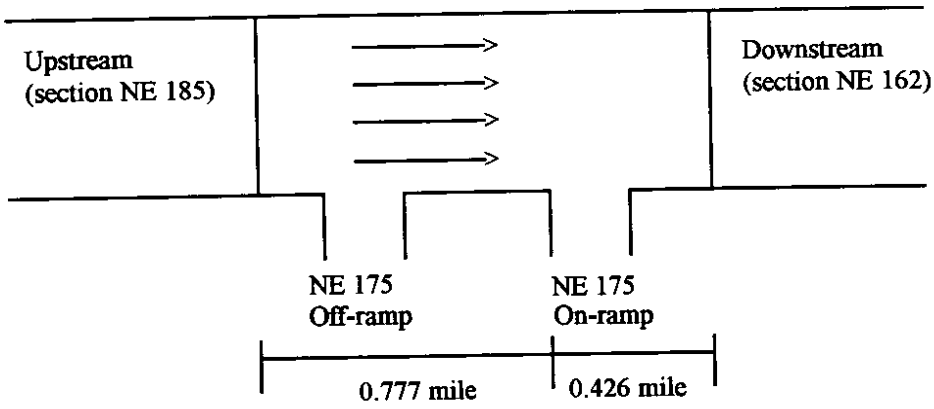
the 103rd were used to forecast the 104th value. The last 20 values are treated in a similar manner.

Using the double exponential smoothing model and the exponential smoothing with adaptive response model, the first 102 data points were used to select the point at which these two models should begin forecasting, so that the forecasting errors of the last twenty points (from 103 to 122) were at a minimum. To build the Box-Jenkins univariate time series model, the first 102 data points were analyzed, and the last twenty data points were used to update the forecasting.

As discussed in Chapter Four, the upstream volume is required to forecast the downstream volume. In order to compare the volume forecasting results of the above four models with those of the model developed in Chapter Four, the same volume data set at the NE 162nd St. section was chosen as the downstream volume. The upstream volume at the NE 185th St. section and the on-ramp volume at NE 175th St. were used as the upstream volume and the on-ramp volume (see Figure 5.1). The first 102 data points of upstream, on-ramp and downstream volumes were used to build the model; that is, the first 102 data points were used to analyze the lags and to obtain the beginning coefficients so as to forecast the last 20 data points recursively. The data interval is 1 minute, as discussed in Chapter Four; if an interval of less than 1 minute is needed, then some forecasting errors can be expected.

Appendix C shows these data sets, including upstream volumes, on-ramp volumes, downstream volumes and occupancies, and their plots.

Figure 5.1 - Study Site



CRITERIA

To evaluate both the volume and occupancy forecasting results of the above models, the following three criteria were used: (1) mean relative absolute error, which indicates the error as a fraction of the measurement, (2) relative square root absolute error, which penalizes large prediction errors, and (3) maximum relative absolute error. The corollary equations are as follows:

$$(1)\quad E_{me} = \sum_{t=103}^{122} \left| \frac{V_d(t) - \hat{V}(t)}{V_d(t)} \right| \div 20 \tag{5.1}$$

$$(2)\quad E_{sr} = \sum_{t=103}^{122} \sqrt{\left| \frac{V_d(t) - \hat{V}(t)}{V_d(t)} \right|} \div 20 \tag{5.2}$$

$$(3)\quad E_{\max} = \operatorname*{Max}_{t=103}^{122} \left| \frac{V_d(t) - \hat{V}(t)}{V_d(t)} \right| \tag{5.3}$$

where $V_d(t)$ is the actual volume at time t, and $\hat{V}(t)$ is the predicted volume at the same time t.

To evaluate occupancy results, $\hat{V}(t)$ and $V_d(t)$ should be changed into $\hat{O}(t)$ and $O(t)$, where $\hat{O}(t)$ and $O(t)$ are predicted and actual occupancies, respectively.

For a method to be superior to another, it should have at least two criteria better than those of the other method.

VOLUME FORECASTING RESULTS AND THEIR COMPARISON

In this section, the volume forecasting results of every model are described, then the results from the five models are compared.

To estimate the adaptive prediction system model, some preliminary runs were made, and it was found that n=10 and $\mu = 4 \times \mathrm{E}(-7)$, which resulted in the lowest forecasting error. The initial weight vector $\mathbf{W}_1 = (0.1, \ 0.1, \ \cdots, \ 0.1)$ was chosen; that is, the first value forecast is the average of the previous ten data. The obtained three criteria are

1) $E_{me} = 9.4$ percent,

2) $E_{sr} = 0.3$,

3) $E_{\max} = 43$ percent.

The detailed forecasting results can be found in Appendix D.

The optimal beginning point resulting in the minimum forecasting error for the double exponential smoothing model is the 85th data point, the optimal α = 0.1. The three criteria are

1) $E_{me} = 10.5$ percent,

2) $E_{sr} = 0.3$,

3) $E_{\max} = 43$ percent.

If we begin to forecast at point 100, it is found that the optimal α is still 0.1 and the three criteria are

1) $E_{me} = 10.7$ percent,

2) $E_{sr} = 0.31$,

3) $E_{\max} = 43.8$ percent.

This indicates that the two sets of criteria are almost the same; therefore, it is not necessary to store the data from point 85 to begin the forecast; only the last three data points need to be stored. Detailed results of this model can be found in Appendix E.

For the exponential smoothing with adaptive response model, the optimal beginning point for the forecast is 95, the optimal $\tau = 0.1$, and the optimal $\alpha = 0.3$. The three criteria are

1) $E_{me} = 9.8$ percent,

2) $E_{sr} = 0.28$, and

3) $E_{\max} = 39$ percent.

When we begin to forecast at point 100, the optimal $\tau = 0.4$, and the optimal $\alpha = 0.4$. The three criteria are

1) $E_{me} = 12.3$ percent,

2) $E_{sr} = 0.325$,

3) $E_{\max} = 41.6$ percent.

From the above two sets of criteria it is clear that the first one is better than the second one. Hence, we need to store more data to forecast the last 20 data points better (for a comparison of these sets of criteria, see Appendix F).

To develop the Box-Jenkins model, as was discussed above, we should use the three-stage iterative procedure. From the plot (see Appendix C) it can be seen that the first 102 data series is stationary with respect to the variance; therefore, we need not transform the data. We can also see that the time series does not need to be differenced because it is stationary with respect to the mean. From the autocorrelation function (see Appendix G), it can be seen that this time series is random because the autocorrelation coefficients are all very small, and $p=d=q=0$. Therefore the model should be

$$\hat{V}(t) = MV + a_t, \tag{5.4}$$

where

MV is the volume sample mean, and

a_t is the disturbance at time t,

which shows that the model is a random model and forecasts only the sample mean.

In the second stage, we need to estimate the coefficients of the model. Since it is a random model, we only need to estimate the sample mean, which is 109.36.

In order to accomplish the last step, the software IDA on the MAX was used to test whether the residuals were white noise or not. From the result, we found that the adjusted Box-Pierce test Q_{ad} is 14.18 when the degree of freedom is 19. Obviously, the test Q_{ad} is less than the tabulated chi-square with 19 degrees of freedom at a 95 percent confidence interval, because $Q_{ad} < 19$. This indicates that the residuals are white noise.

Therefore, the obtained model

$$\hat{V}(t) = MV + a_t$$

was sufficient. We can use this model to forecast the last 20 values. The three calculated criteria are

1) $E_{me} = 17$ percent,

2) $E_{sr} = 0.38$,

3) $E_{\max} = 58.6$ percent.

(The detailed results can be found in Appendix G.)

To build the last model, the lags must first be analyzed. For this purpose the software BMDP with the MAX system was used to calculate the phase, $W(f)$, of Fourier's transformation of cross correlation between downstream volume and upstream or on-ramp volume (see Chapter Four). Using the lag analysis method introduced in Chapter Four, we calculated the lags between the upstream and the downstream volumes, and the lags between the on-ramp and downstream volumes. The lags between the upstream and downstream volumes (see Appendix H) are 1 minute and 2 minutes because the interval of these data points is 1 minute. The results indicate that the travel time from the upstream section (NE 185th St.) to the downstream section (NE 162nd St.) is 1 minute or 2 minutes. Lags between the on-ramp (NE 175th St.) and the downstream section (NE

162nd St.) are 0 and 1 minute, indicating that travel time from the on-ramp to the downstream section is less than 1 minute since the interval is known to be 1 minute for upstream volume $T1_{min} = 1$, $T1_{max} = 2$, for the on-ramp volume $T2_{min} = 0$, $T2_{max} = 1$.

From the calculation and the discussion in section 4.5, we know that the optimal interval is less than 1 minute; however, the interval of the actual data is 1 minute, thus we can expect some error in forecasting, because we can not use $V_{on}(t)$ to forecast the simultaneous downstream volume $V_d(t)$. A comparison of the on-ramp volume to the upstream volume shows that the former is much less than the latter, so we can expect some forecasting error here as well. This error is small; therefore, we can still use these data sets to forecast the downstream volume.

Thus, the forecasting model is

$$V_d(t) = b_1(1)V_{up}(t-1) + b_1(2)V_{up}(t-2) + b_2(1)V_{on}(t-1) \tag{5.5}$$

Using the ordinary least square method, we can obtain the initial coefficients:

$$b_1(1) = 0.42$$

$$b_1(2) = 0.6$$

$$b_2(1) = 0.25.$$

t-ratios for these three coefficients are 5.72, 7.99 and 0.77 respectively. The t-ratio of $b_2(1)$ is so low because most lags between on-ramp and downstream volumes are less than 0.5, but since we used 1 as the lag, this coefficient is not statistically significant from 0. Because the downstream volumes are derived from the upstream as well as from the on-ramp, this term is kept for easy interpretation.

The obtained initial matrix $\mathbf{A}^{-1}(0)$ is

$$\mathbf{A}^{-1}(0) = \begin{vmatrix} 0.00008117 & -0.00007646 & -0.00006755 \\ -0.00007646 & 0.00007941 & -0.00003539 \\ -0.00006755 & -0.00003539 & 0.00157328 \end{vmatrix}$$

Using the recursive forecasting model, we have the three criteria:

1) $E_{me} = 8$ percent,

2) $E_{sr} = 0.26$,

3) $E_{\max} = 27.8$ percent.

The detailed results can be found in Appendix H.

The above results are individual ones; next all these results will be compared with each other.

The following table contains the three criteria for all of these models (see Figure 5.2, Criterion Comparison of Forecast Volumes by Different Models).

Table 5.1 - Criterion Comparison of Volume Forecasting Results

Model	Criterion		
	E_{me}	E_{sr}	$E_{\max}$
1	9.4 percent	0.30	43.0 percent
2	10.5 percent	0.30	43.0 percent
3	9.8 percent	0.28	39.0 percent
4	17.0 percent	0.38	58.6 percent
5	8.0 percent	0.26	27.8 percent

where

Model 1 = adaptive prediction system model,

Model 2 = double exponential smoothing model,

Model 3 = exponential smoothing with adaptive response model,

Model 4 = Box-Jenkins univariate time series model, and

Model 5 = upstream model, which was developed in Chapter Four.

From the above table it can be seen that Model 5, which uses upstream and on-ramp volumes to forecast the downstream volumes, is superior to the other models with respect to the three criteria. Its relative mean absolute error is 8 percent, the maximum relative absolute error is 27.8 percent, and the relative mean square root absolute error is 0.26. The next best results were obtained with Model 3, in which two criteria, 2 and 3,were better than those of Model 1, which obtained the third best results. The relative mean absolute error of Model 1 is better than that of Model 2, which obtained the fourth

best results. The worst results were obtained with Model 4, which has the highest relative mean absolute error of 17 percent, the highest maximum relative absolute error of 58.6 percent and the highest relative square root absolute error of 0.38. If we compare the results of Models 3 and 2, we can see that using the adaptive response a (Model 3) will improve the forecast.

If we compare the three criteria obtained from Model 5 with those from the other four models by calculating the ratios of these three criteria from Model 5 and the corresponding criteria from the other four models, then we derive the following table:

Table 5.2 - Criterion Comparison of Model 5 With Other Models

Model	Criterion		
	1	2	3
1	1.175	1.15	1.547
2	1.313	1.15	1.547
3	1.225	1.077	1.403
4	2.125	1.462	2.108

Even Model 3, the best among the four models, has a relative mean absolute error 22.5 percent higher than the corresponding errors of Model 5, a, relative mean square root absolute error 7.7 percent higher, and maximum relative absolute error 40.3 percent higher than the corresponding errors of Model 5. Models 1 and 2 have the second criterion 15 percent higher and the third criterion 54.7 percent higher than those of the same error terms, but in Model 2, criterion 1 is 31.3 percent higher than in Model 5, and and in Model 1 it is 17.5 percent higher than that of Model 5. The ratios were obtained with Model 4, the Box-Jenkins univariate time series, which has the first error 112.5 percent higher, the second error 46.2 percent and the last error 110.8 percent higher than Model 5.

From a comparison of the figures of the 20 forecast volumes vs. the actual volumes (see Figures 5.3 - 5.6), we can see that Models 1 through 3 (Figures 5.3 and 5.4) follow the trend; the forecasting performance of these three models is almost the same,

although Table 1 shows slight differences among the three criteria. Model 4 (Figure 5.4) is only able to predict the previous sample mean; when the mean changes with time it can not follow the trend. The predicted values of Model 5 (Figure 5.5) can follow the actual traffic volumes. Comparing actual volumes and the forecast volumes of Model 5 and actual the volumes and forecast volumes of Models 1 and 3 (Figures 5.5 and 5.6), it is clear that Model 5 is the best model for forecasting downstream volume.

INTERPRETATION OF THE VOLUME FORECASTING RESULTS

From the above results, Model 5 is:

$$\hat{V}(t) = b_1(1)V_{up}(t-1) + b_1(2)V_{up}(t-2) + b_2(1)V_{on}(t-1) \tag{5.6}$$

which shows that 1) the downstream volumes are derived from the upstream volumes and the on-ramp volumes; 2) vehicle travel time from the upstream to the downstream section is about 1 to 2 minutes, which can be compared with the time calculated by using the deterministic formula

$$t = \frac{\text{distance}}{\text{time}} \tag{5.7}$$

If we calculate the speed of 46 mph (corresponding to the speed at freeway level of service D (35)), the distance is 1.203 miles and the time is 1.6 minutes. At a speed of 55 mph (speed limit), the time is 1.3 minutes. (Generally, the travel time is between 1.3 and 1.6 minutes.) These travel times are between the obtained lags 1 and 2; 3) lag 1 and 2 imply that the diverse speeds of each lane result in varying travel times; and 4) because the distance from on-ramp to the downstream section is very short, 0.426 mile, the travel time is also short. Using 50 mi/hour as the speed, produces a travel time of 0.51 minute; therefore this time is between the obtained lags 0 and 1.

From the initial estimated parameters by ordinary least square (OLS), we obtain:

$$b_1(1) = 0.42$$

$$b_1(2) = 0.6$$

$$b_2(1) = 0.25.$$

$b_1(1) = 0.42$ indicates that on average 42 percent vehicles will travel about 1 minute to the downstream section. In the same way, $b_1(2) = 0.6$ indicates that an average of 60 percent of the vehicles will travel about 2 minutes from the upstream section to the downstream section. The sum of these two coefficients is 1.024, indicating that all vehicles will travel to the downstream section (if the travel time of all vehicles to the downstream section were exactly between 1 to 2 minutes, the sum of the two coefficients should be exactly 1). However, from Figure 5.1, we note that some vehicles exit at the NE 175th St. off-ramp. Compared with the volume at the NE 162nd St. section, this volume is small; therefore, the sum of these two coefficients should be less than 1. The reason that the sum is greater than 1 is due either to an error in measurement or to an inherent problem in the model that prevents the use of the simultaneous volumes to forecast downstream volumes. $b_2(1) = 0.25$ indicates that about 25 percent of the vehicles will travel to the downstream section within about 1 minute, while others will arrive at the downstream section in less than that. With a *t*-ratio = 0.7, we can see that this coefficient is not reliable, since it is not significant from 0. In Appendix C we see that 25 percent of this volume is about 2 vehicles, which is much lower, compared with the mean of the downstream volume of 109 vehicles. Omitting the third term in equation 5.6, that is, using only the upstream volumes to forecast the downstream volumes, produces the following three criteria:

1. $E_{me} = 8.2$ percent,
2. $E_{sr} = 0.265$,
3. $E_{\max} = 26$ percent.

Comparing these three criteria with 8 percent, 0.26 and 27.8 percent, we see that they are almost the same.

We can conclude that if the distance between the on-ramp to the downstream section is very short and on-ramp volume is very low compared with the downstream volume, we can ignore the on-ramp volume and consider only the upstream volume when we build a model to forecast the downstream volume.

The above results of Model 5 were obtained by the recursive method. However, if we use the initial coefficients, to forecast the last 20 volumes; that is, if the forecasting is off-line, then we have the three criteria:

1. $E_{me} = 8$ percent,
2. $E_{sr} = 0.26$
3. $E_{\max} = 27.4$ percent.

Comparing these three criteria with 8 percent, 0.26 and 27.8 percent, the results obtained by Model 5, we can see that $E_{\max}$ of this off-line forecasting is even better than that of the recursive method. When most upstream volumes travel downstream, and the speeds in different lanes do not change greatly, the main parameters b_1, b_2 will not change greatly, and the percentage of vehicles in a certain interval traveling downstream will not change greatly. (The detailed results are given in Appendix H.)

The Box-Jenkins method assumes that autocorrelation and partial autocorrelation functions of a time series will not change; that is, the form and parameters of the developed model will not change. In our situation, this is similar to the assumption that the pattern of the previous 102 data points will not change with time. Obviously, when the pattern of the time series does change (e.g., declines), this univariate time series method will fail . From Figure G.1 in Appendix G, we can see that the mean of the first 102 data points is almost the same, and these points hover around this mean; therefore, we used the Box-Jenkins and obtained the model, which can only forecast the mean. From Figure G.2 we can see that the time series declines; that is, the previous pattern changes. Model 4, the univariate time series model, is therefore the worst of the five models, as it cannot even follow the trend of the time series.

The objectives of the other three models are the same: they follow the trend of a time series. The adaptive prediction system model (Model 1), assumes that the next volume is the sum of several previous weighted volumes if the difference between this predicted volume and the actual volume is zero (that is, the predicted volume is the same

as the actual volume), then these weights remain the same. If the difference is not zero, then these weights will change so that the predicted volumes can follow that trend of the time series. The detailed formula can be found in Chapter Three.

The double exponential smoothing model (Model 2), can follow the linear trend. Brown has demonstrated that the steady-state response of exponential smoothing to a linear trend has a constant lag of (1- α)/α (27). Therefore, in our case, this model can follow the trend of the time series.

The exponential smoothing model with adaptive response (Model 3), assumes that the next volume depends on the current volume and the weighted current forecasting error, and the weight, α, will also change depending on the current forecasting error and another parameter, τ. In this way, Model 3 can follow the trend of the time series. The detailed formula can be found in Chapter Three.

OCCUPANCY FORECASTING RESULTS AND THEIR COMPARISON

In this section, the results achieved by every model will be described, then all these results compared.

To evaluate the adaptive prediction system model some preliminary runs have been made, and the researchers found that n=10 and $\mu = 4 \times \mathrm{E}(-4)$, which resulted in the lowest forecasting error. The initial weight vector $\mathbf{W}_1 = (0.1,\ 0.1,\ \cdots,\ 0.1)$ was chosen; that is, the first forecasted occupancy is the average of previous ten occupancies. The obtained three criteria are

1) $E_{me} = 11.7$ percent,

2) $E_{sr} = 0.29$,

3) $E_{\max} = 46$ percent.

The detailed forecasting results can be found in Appendix D.

The optimal beginning point resulting in the minimum forecasting error is the 82nd data point, the optimal s = 0.1 for the double exponential smoothing model. The three criteria are

1) $E_{me} = 12$ percent,

2) $E_{sr} = 0.31$,

3) $E_{\max} = 40$ percent.

If we begin to forecast at point 100, then the optimal α is 0.15 and the three calculated criteria are

1) $E_{me} = 12.9$ percent,

2) $E_{sr} = 0.33$,

3) $E_{\max} = 42.7$ percent.

These three criteria indicate that the former sets of the criteria are better than the latter, so we need to store some data to forecast the future. (The detailed comparison can be found in Appendix D.)

For the exponential smoothing with adaptive response model, the optimal beginning point for the forecast is point 99, the optimal τ=0.3 and the optimal α=0.9. The three criteria are

1) $E_{me} = 12$ percent,

2) $E_{sr} = 0.29$,

3) $E_{\max} = 53$ percent.

When we begin to forecast at point 100, the optimal τ=0.9 and the optimal α=0.8, the three criteria are

1) $E_{me} = 15.7$ percent,

2) $E_{sr} = 0.356$,

3) $E_{\max} = 63.9$ percent.

In reviewing the above two sets of criteria it is obvious that the first one is better than the second, so we need to store more data to better forecast the last 20 data points. (See Appendix F for the comparison.)

The Box-Jenkins model's volume time series plot shows that the first 102 data series is stationary with respect to the variance, so we need not transform the data (see

Appendix C). We can also see that the time series does not need to be differenced because this time series is stationary with respect to the mean. The autocorrelation and the partial autocorrelation of occupancy (see Appendix G) time series looks like the AR(1) time series, in that the first lag partial autocorrelation coefficient is significant from 0, but other partial autocorrelation coefficients are not significant from 0, so p=1, d=0, q=0. Therefore the model should be:

$$\hat{O}(t) = A_0 + A_1 O(t-1) + a_t \tag{5.8}$$

where

$\hat{O}(t)$ is a predicted occupancy at time t,

$O(t-1)$ is an actual occupancy at time t-1,

A_0 is the occupancy sample mean,

A_1 is a coefficient, and

a_t is the disturbance at time t.

This shows that the next occupancy is highly correlated with the current occupancy; that is, the next occupancy depends on the current one.

In the second stage we need to estimate the coefficients of the model, which can easily be done by running the IDA program. The obtained A_0 and A_1 are 3.1 and 0.773, their t-ratios are 13.66 and 11.9, respectively. Obviously, these two t-ratios are highly significant within a 95 percent confidence interval.

In order to accomplish the last step, the IDA software was used again. In viewing the results, we see that the adjusted Box-Pierce test, Q_{ad} is 20.25, when the degree of freedom is 18. The tabulated chi-square with 18 degrees of freedom at a 95 percent confidence interval is 28.87, because Q_{ad} < 2.87; indicating that the residuals are white noise.

Therefore, the obtained model

$$\hat{O}(t) = A_0 + A_1 O(t-1) + a_t$$

is sufficient, and we can use this model to forecast the last 20 values. The three criteria are

1) $E_{me} = 17.4$ percent,

2) $E_{sr} = 0.367$,

3) $E_{\max} = 80.5$ percent.

The detailed results can be found in Appendix G.

A comparison of these results follows in Table 5.3:

Table 5.3 - Criterion Comparison of Occupancy Forecasting Results

Model	Criterion		
	E_{me}	E_{sr}	$E_{\max}$
1	11.7 percent	0.29	46.0 percent
2	12 percent	0.31	40.0 percent
3	12 percent	0.29	53.0 percent
4	17.5 percent	0.37	80.5 percent

where

Model 1 = adaptive prediction system model,

Model 2 = double exponential smoothing model,

Model 3 = exponential smoothing with adaptive response model, and

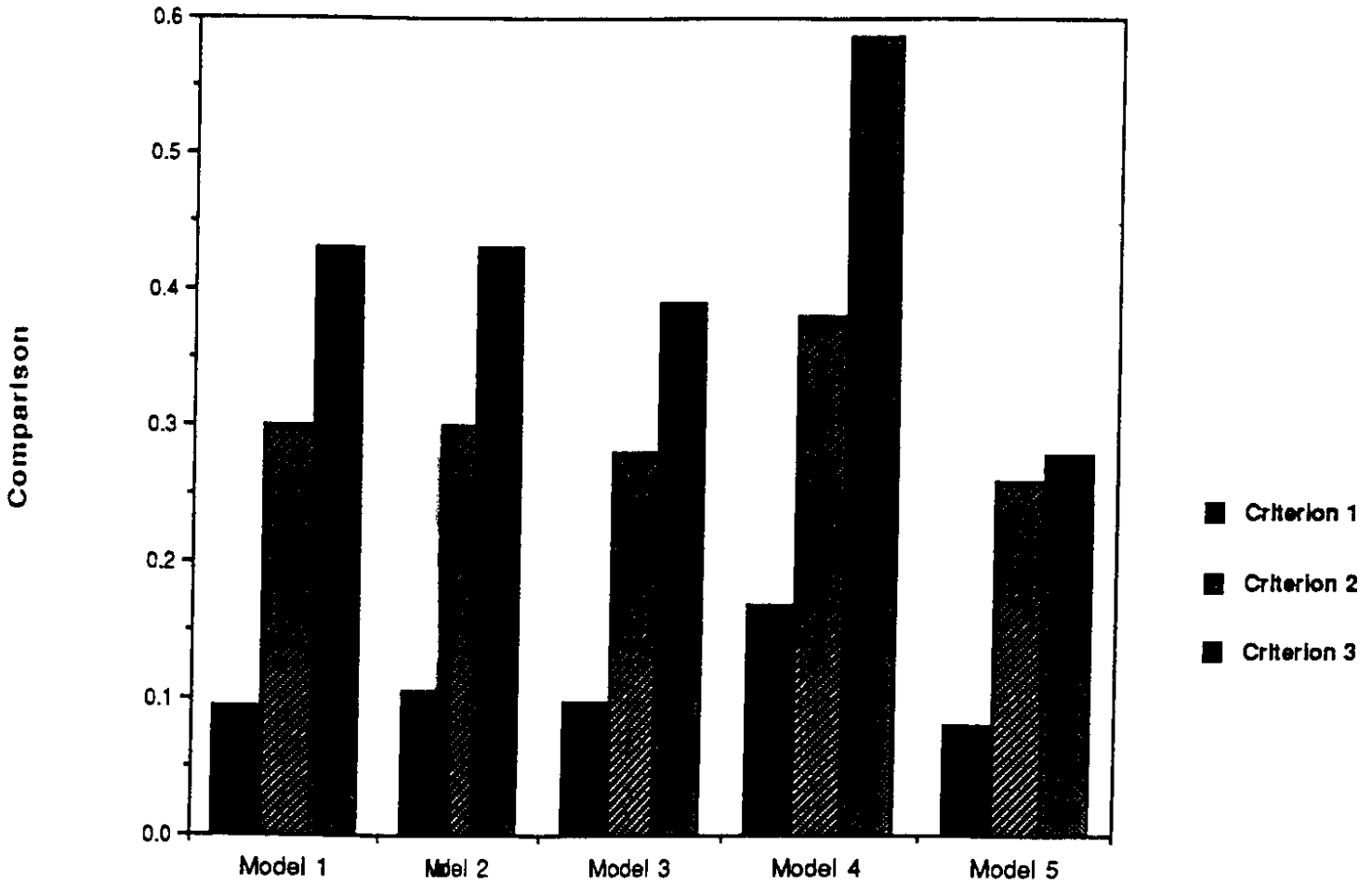
Model 4 = Box-Jenkins univariate time series model.

From the above table one can see that the criteria of Model 1, Model 2 and Model 3 are almost same and these criteria are much better than those of Model 4. Among the first three Models, the adaptive prediction system Model (Model 1), is the best one; it has the two lowest criteria. Next to that is the second Model, which has a much lower $E_{\max}$, and the third Model is in third place. These results can be seen very clearly in Figure 5.7.

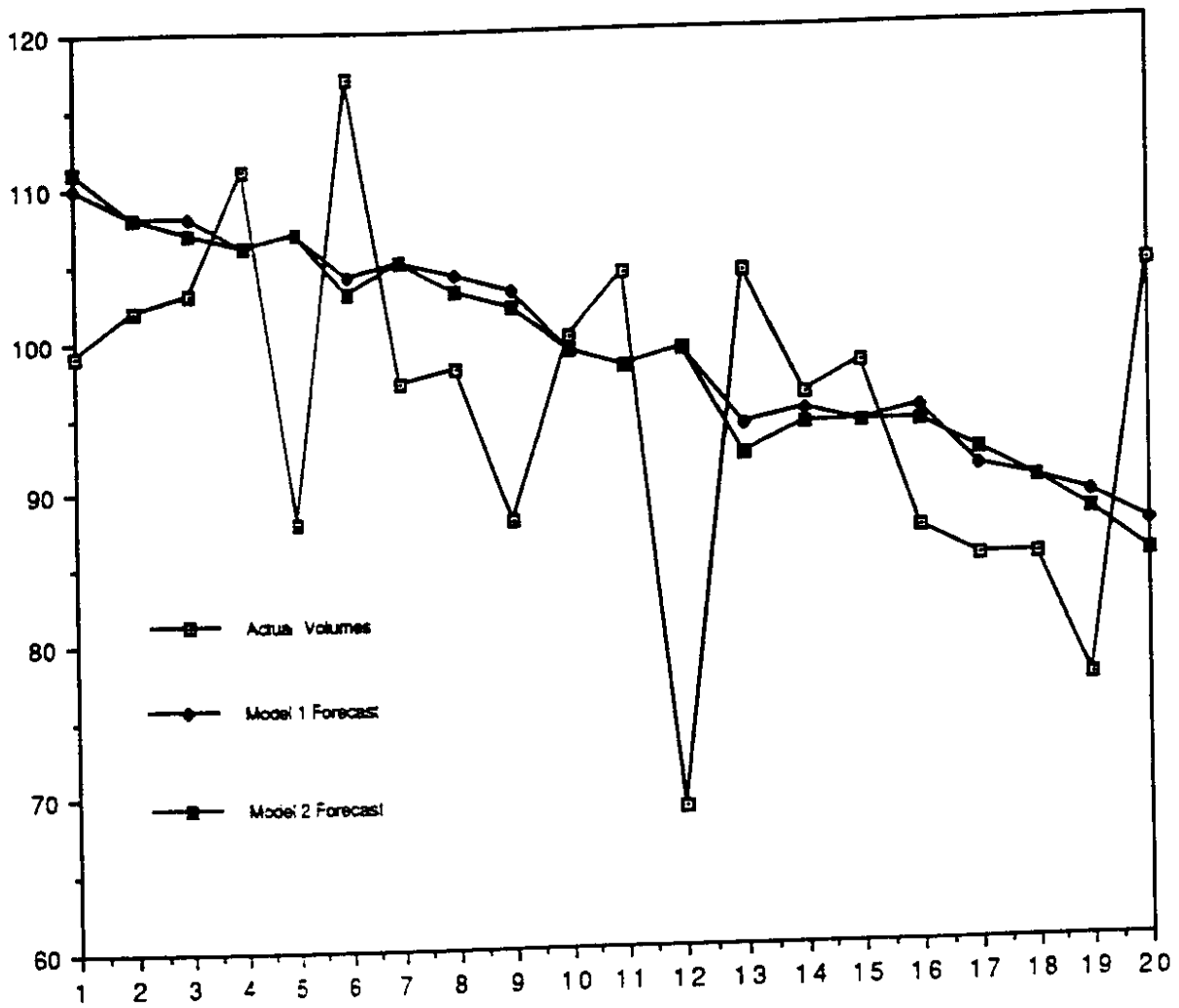
A comparison of the figures of the 20 forecast occupancies and the actual occupancies (see Figure 5.8) shows that Models 1 through 3 can follow the trend. Although in Table 5.3 a small difference can be noted among the three criteria, the forecasting performance of these three Models is almost the same,

The prediction of Model 4 (Figure 5.9) looks like it is a one lag foreward shift of the actual occupancies because the prediction of Model 4 depends mainly on current occupancy (see equation 5.8). When this pattern changes, the forecast occupancies cannot follow the trend of future occupancies.

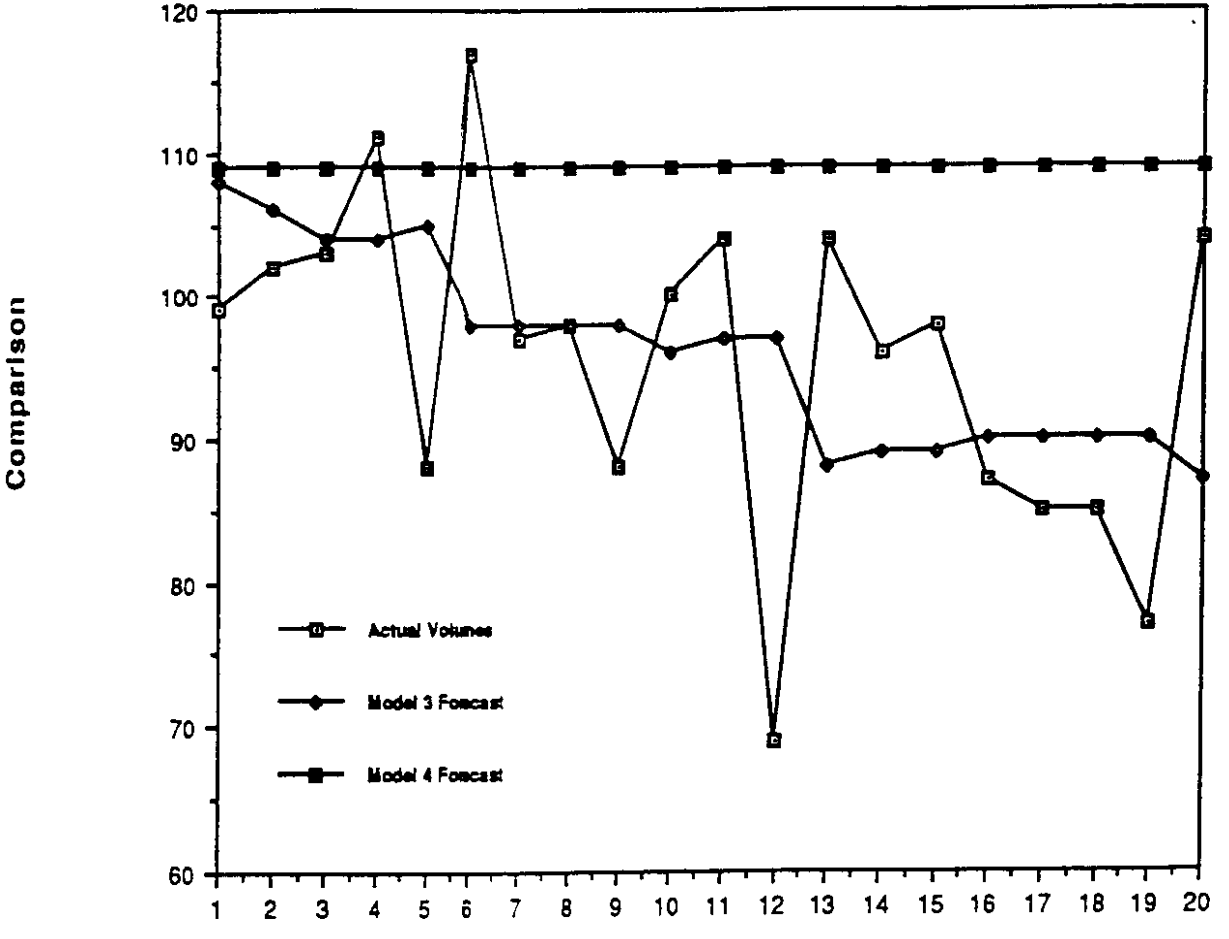
The interpretation of the occupancy forecasting results of these four Models is the same as the interpretation of the volume forecasting results of the same four Models in section 5.4. The detailed interpretation of these results can be found in Section 5.4.



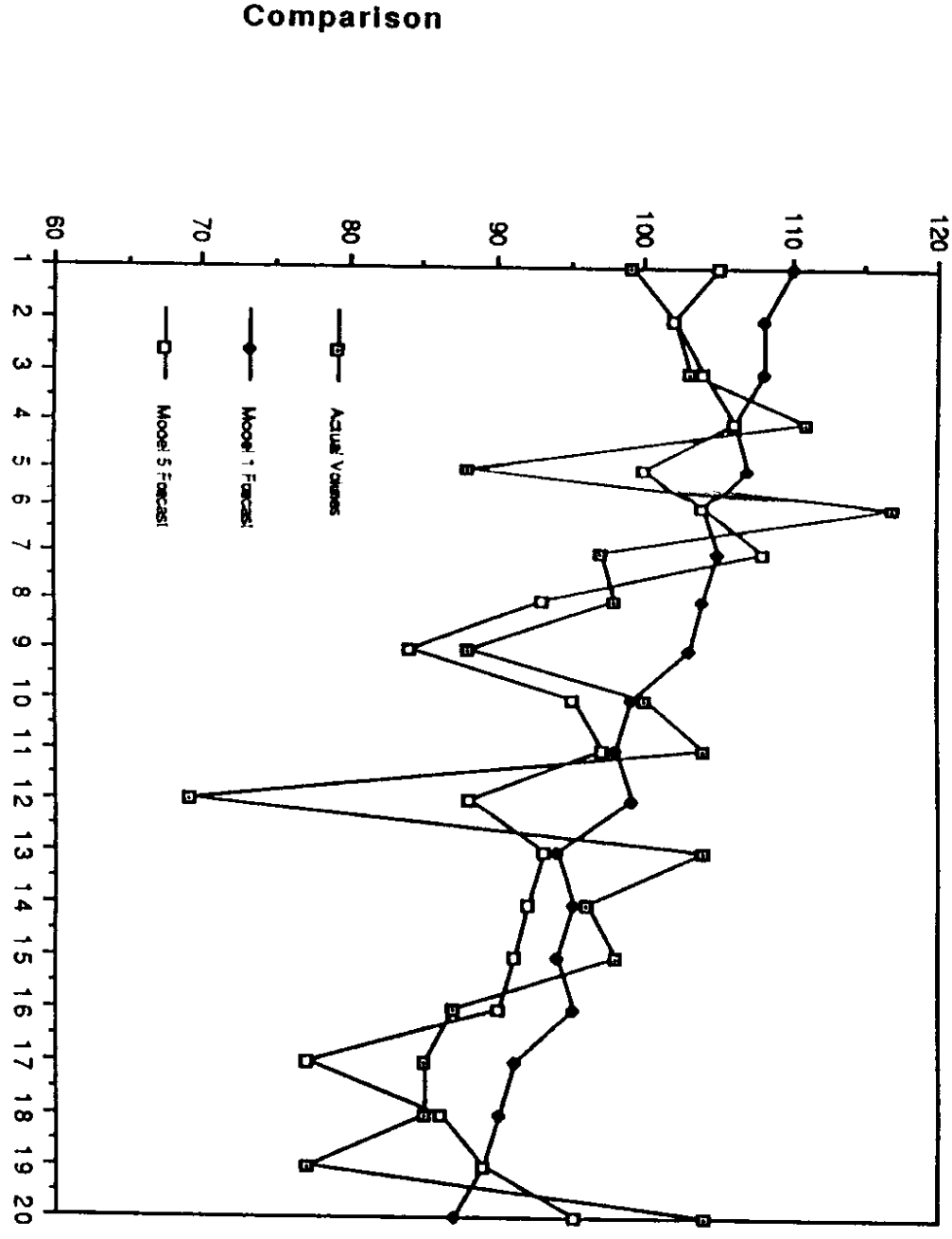
Criterion Comparison of
Forecast Volumes by Different Models



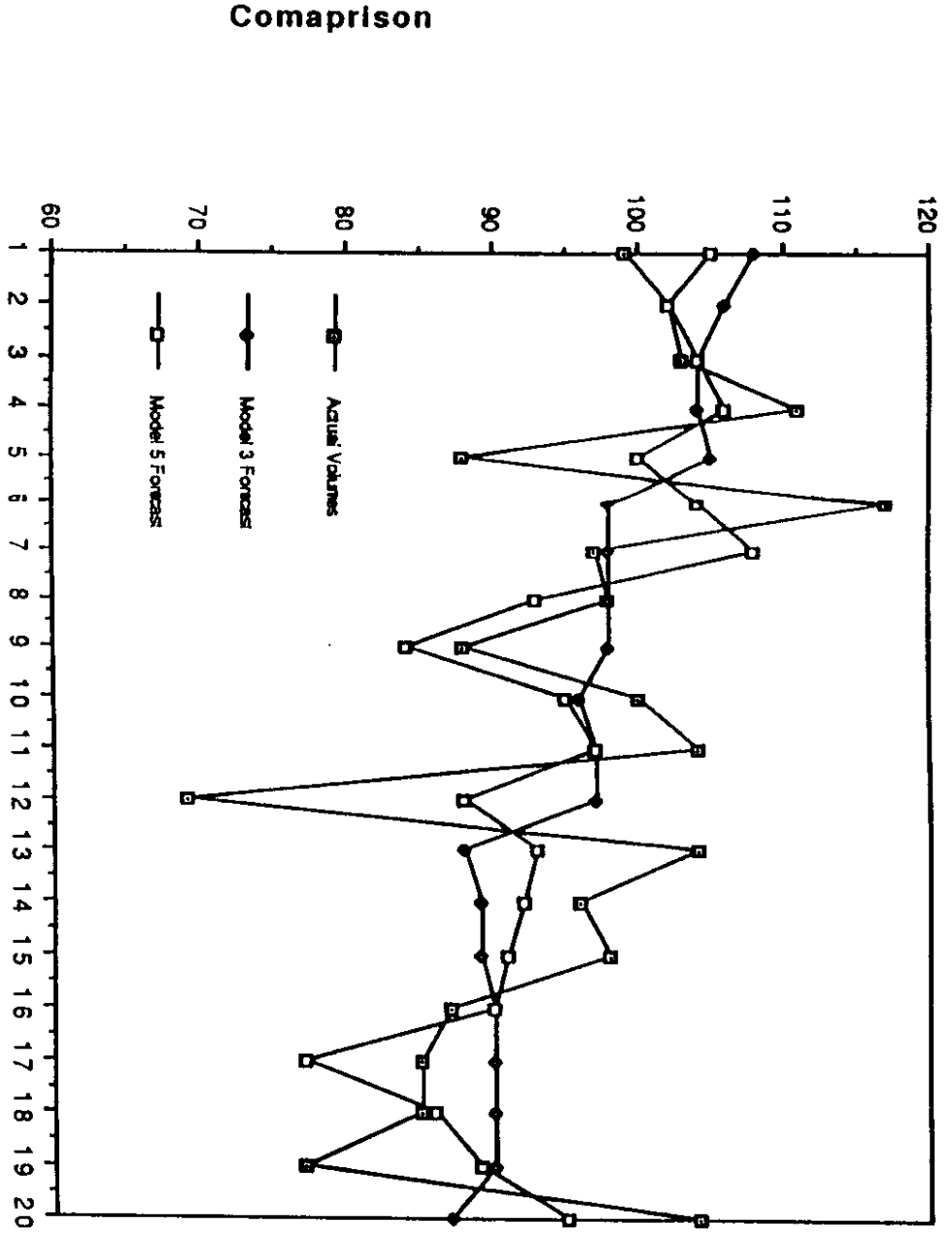
Forecast Volumes by Model 1
and Model 2 vs Actual Volumes



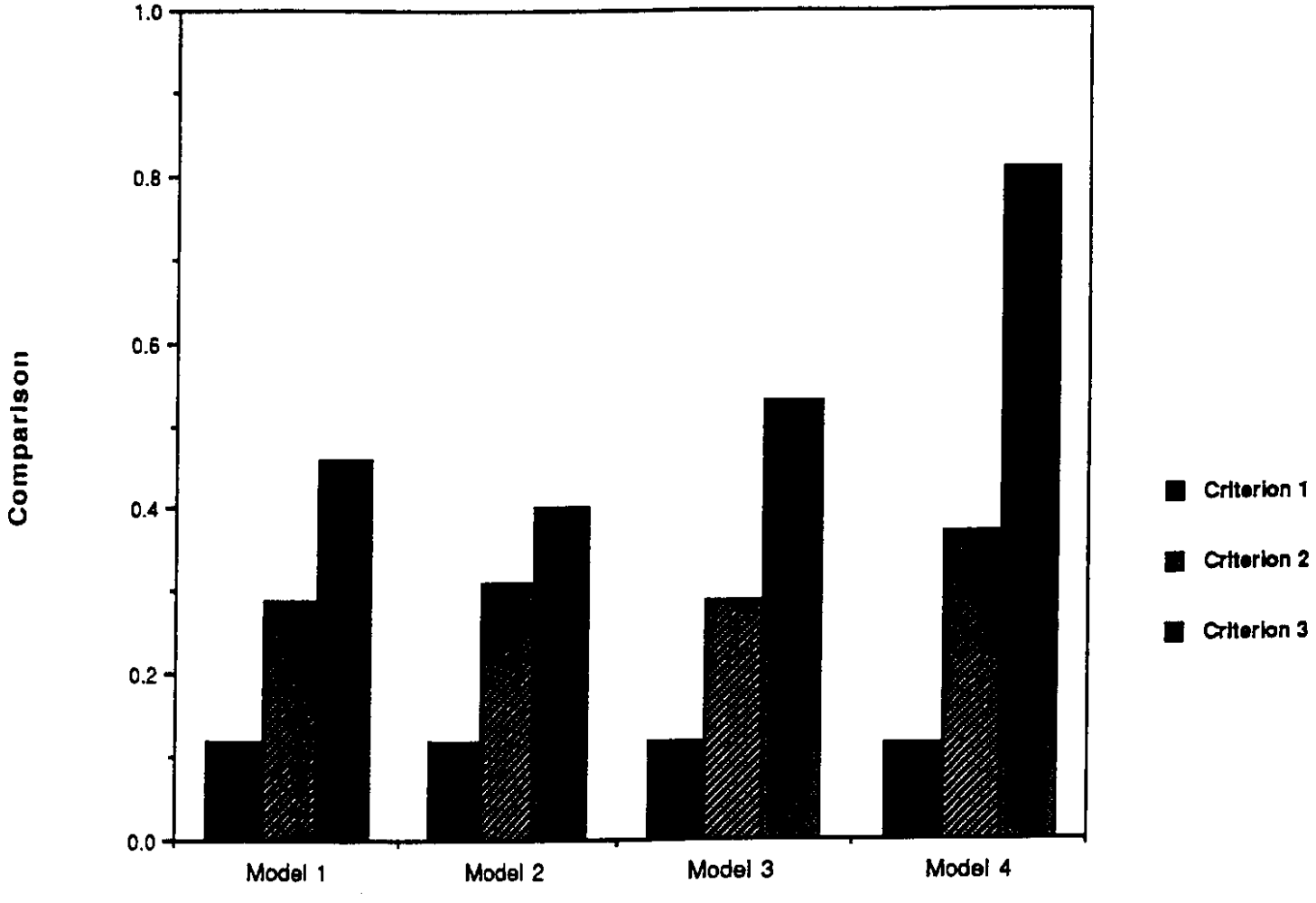
Forecast Volumes by Model 3 and Model 4 vs Actual Volumes



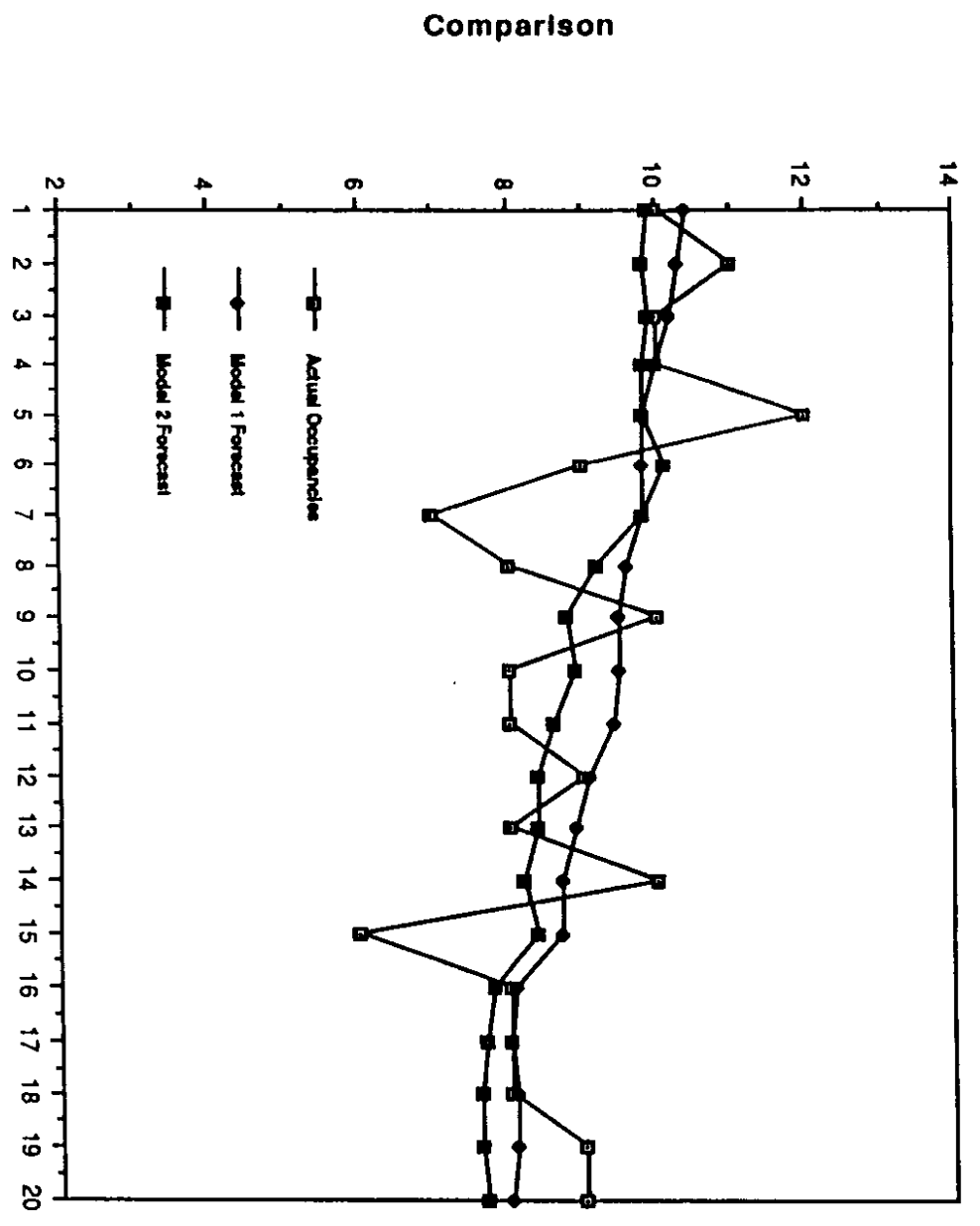
Forecast Volumes by Model 5 and Model 1 vs Actual Volumes



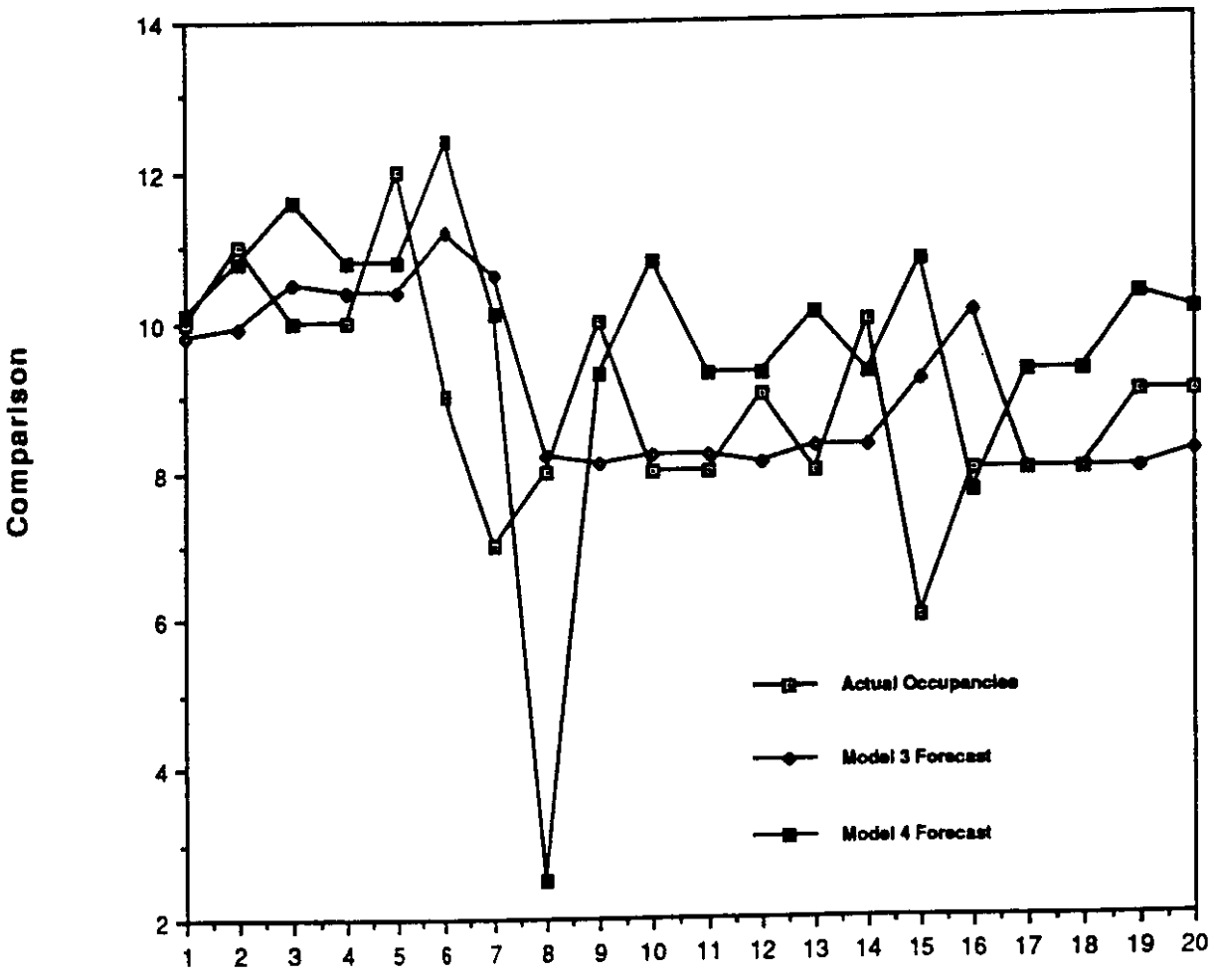
Forecast Volumes by Model 5
and Model 3 vs Actual Volumes



Criterion Comparison of
Forecast Occupancies by Different Models



Forecast Occupancies by Model 1 and Model 2 vs Actual Occupancies



Forecast Occupancies by Model 3 and Model 4 vs Actual Occupancies

CHAPTER SIX

Recommendation and Conclusion

A volume forecasting model using cross spectrum analysis to analyze the lags between upstream or on-ramp volumes and downstream volumes has been developed in this report. The ordinary least square method and the recursive method were used to obtain the parameters of the model. This developed model uses lagged upstream and lagged on-ramp volumes to forecast short-term downstream volumes.

The critical step in building this model is to analyze the lags, which can be calculated by the following equation:

$$t_d = \frac{W(f)}{2\pi f} \tag{6.1}$$

where t_d is phase delay,

$W(f)$ is phase shift of the cross spectrum, and

f is the frequency, $-\infty < f < +\infty$ and $f \neq 0$.

After the lags $T1_{\min}$, $T1_{\max}$, $T2_{\min}$, and $T2_{\max}$ (as defined in Chapter Four) have been obtained, the optimal data interval can be determined by the following equation:

$$T_{\text{optimal}} = \min(T1_{\min}, T2_{\min}) \tag{6.2}$$

and the forecasting model expressed as:

$$\hat{V}(t) = \sum_{T1=T1_{\min}}^{T1_{\max}} b_1(T1) V_{up}(t - T1) + \sum_{T2=T2_{\min}}^{T2_{\max}} b_2(T2) V_{on}(t - T2) \tag{6.3}$$

The coefficients $\hat{\mathbf{b}}_1$ and $\hat{\mathbf{b}}_2$ can be obtained by the ordinary least square method as:

$$\hat{\mathbf{b}}_d = \left[\sum_t V_d(t)\mathbf{V}^{\mathrm{T}}(t)\right]\left[\sum_t \mathbf{V}^{\mathrm{T}}(t)\mathbf{V}(t)\right]^{-1} \tag{6.4}$$

or by recursive method as:

$$\hat{\mathbf{b}}_d(k) = \hat{\mathbf{b}}_d(k-1) + \mathbf{K}(k)\left[V_d(k) - \mathbf{V}(k)\hat{\mathbf{b}}_d(k-1)\right] \tag{6.5}$$

$$\mathbf{K}(k) = \frac{\mathbf{Q}(k-1)\mathbf{V}^{\mathrm{T}}(k)}{1+\mathbf{V}(k)\mathbf{Q}(k-1)\mathbf{V}^{\mathrm{T}}(k)} \tag{6.6}$$

$$\mathbf{Q}(k) = \mathbf{Q}(k-1) - \frac{\mathbf{Q}(k-1)\mathbf{V}^{\mathrm{T}}(k)\mathbf{V}(k)\mathbf{Q}(k-1)}{1+\mathbf{V}(k)\mathbf{Q}(k-1)\mathbf{V}^{\mathrm{T}}(k)} \tag{6.7}$$

The volume forecasting results of this model and others indicate that (1) the forecast volumes can follow not only the trend of the actual volumes, but also their fluctuations; (2) the three criteria are all less when compared with those obtained by the other four existing models; (3) the forecasting results using the ordinary least square method are not much better than the results using the recursive method; and (4) if the on-ramp volumes are very small (compared with downstream volumes), and the distance between the on-ramp and the downstream section is very short (compared with the data interval), then we need not consider the on-ramp volumes when we build the forecasting model; that is, we only need the first term in the equation (6.3).

These results show that (1) the model developed in this report is the most promising of these five models for forecasting downstream volumes; (2) we should use this time series to forecast the downstream volume if we have the upstream volume time series; (3) to save computation time, off-line forecasting using just the upstream volumes is recommended because the forecasting results obtained by ordinary least square method are not much better than those obtained by recursive method, and on-ramp volumes have little effect on forecasting.

This developed model is not restricted to forecasting volumes on the freeway. It can be used to analyze any inflowing and outflowing network system, such as an Origin-Destination matrix in an urban network system, because it uses cross spectrum analysis to obtain the lags, instead of using speed to obtain the travel time. Therefore, this method alleviates the problems that could result from the implementation of traffic variables, such as route choice, signal delay, speed, etc.

It is difficult to express downstream occupancies as the function of upstream occupancies; therefore, this developed model cannot be used to forecast downstream occupancies. To forecast occupancies and to evaluate the prediction of this developed model, four other existing models were applied in this report. These models use univariate time series to forecast their own future; therefore, if only the downstream volume time series is available or if it is difficult to forecast downstream traffic variables using upstream traffic variables such as occupancies, these four models should be selected and attention should be concentrated on comparing the predictions made by these models.

The objective of the first three models is to follow the trend of actual volumes or occupancies. The first step is to obtain the optimal parameters, and then the forecasting of these models can subsequently be carried out.

The forecasting results indicate that models 1 through 3 can follow the trend; the forecasting performances of these three models are almost the same, E_{me} is about 12 percent.

However, the Box-Jenkins model, used very successfully in forecasting traffic variables, has an inherent problem: once it has been built, the form and its parameters can not be updated. Thus, if changes in the pattern of volumes or occupancies occur in the future, this model will not accurately forecast these volumes or occupancies. Therefore, when the Box-Jenkins model is used, the researchers should focus their attention on the future pattern of the time series.

REFERENCES

[1] FHWA, Urban Traffic Control System and Bus Priority System Traffic Adaptive Network Signal Timing Program: Software Description. Federal Highway Administration, Department of Transportation, Washington D.C. (1973).

[2] SRI, Improved Control Logic for Use with Computer Controlled Traffic. Stanford Research Institute, NCHRP project 3-18 (1), Final Report, Menlo Park, CA. (1977).

[3] CMT, Improved Operation of Urban Transportation Systems: Volume 3-the Development and Evaluation of a Real-Time Computerized Traffic Control Strategy. Canada Ministry of Transport, Toronto, Ontario (1976).

[4] Liberman, E. B. et al, Variable Cycle Signal Timing Program: Volume-4 Prediction Algorithms, Software and Hardware Requirements, and Logical Flow Diagrams. Federal Highway Administration, U. S. Department of Transportation, NTIS:PB241 720, Washington, D.C. (1974).

[5] Mengert, P., Brown, P. and Yuan, L., Prediction Algorithms for Urban Traffic Control. Transportation System Center, U.S. Department of Transportation, Internal Project Menmoradum, Cambridge, MA. (1979).

[6] Stephanedes, Y. J., Michalopoulos, P. G. and Plum, R.A., Improved Estimation of Traffic Flow for Real Time Control. Transpn. Res. No. 795, pp. 28-39 (1981).

[7] Ahmed, S. A. and Cook, A. R., Analysis of Freeway Traffic Time-Series Data by Using Box-Jenkins Techniques. Transpn. Res. Rec. No.72, pp. 1-9 (1979).

[8] Eldor, M., Demand Prediction for Computerized Freeway Control Systems. Proceedings of the 7th International Symposium on Transportation and Traffic Theory, Kyoto, Japan, 14-17 August 1977, pp. 341-358.

[9] Gafarian, A. V., Paul, J. and Ward, T. L., Discrete Time Series Models of a Freeway Density Process. Proceedings of the 7th International Symposium on Transportation and Traffic Theory, Kyoto, Japan, 14-17 August 1977, pp. 387-411.

[10] Levin, M. and Tsao, Y. D., On Forecasting Freeway Occupancies and Volumes. TRB, Transportation Research Record 773, pp. 47-49 (1980).

[11] Nihan, N. L. and Holmesland, K. O., Use of the Box and Jenkins Time Series Technique in Traffic Forecasting. Transpn. 9, pp. 125-143 (1980).

[12] Davis, Gary and Nancy L., Nihan, "Use of Time-Series Designs to Estimate Changes in Freeway Level of Service Despite Missing Data," Transportation Research, 18A, No. 5/6, pp. 431-438, 1984.

[13] Nicholson, H. and Swann, C. D., The Prediction of Traffic Flow Volumes Based on Spectral Analysis. Transpn. Res. 8, pp. 533-538 (1974).

[14] Gazis, D. C. and Knapp, C. H., On-line Estimate of Traffic Densities from Time-Series of Flow and Speed Data. Transpn. Sci. 5, pp. 283-301 (1971).

[15] Chang, M. F. and Gazis, D. C., Traffic Density Estimation with Consideration of Lane Changing. Transpn. Sci. 9 (4), pp. 308-320 (1975).

[16] Nahi, N. E., Freeway Traffic Data Processing. Proceedings of the IEEE 61 No. 5, pp. 537-541 (1973).

[17] Okutani, I.and Stephanedes, Y. J., Dynamic Prediction of Traffic Volume through Kalman Filtering Theory. Transpn. Res. B, Vol. 18b, No. 1, pp. 1-11 (1984).

[18] Papageorgiou, M., Applications of Automatic Control Concepts to Traffic Flow Modeling and Control. Berlin: Springer-Verlag (1983).

[19] Lu, J., "Prediction of Traffic Flow by Adaptive Prediction System," University of Texas at Austin, Unpublished (1989).

[20] Kreer, J. B., A comparison of Predictor Algorithms for Computerized Traffic Control Systems. Traffic Engng. 45 (4), pp. 51-56 (1975).

[21] Kreer, J. B., Factors Affecting the Relative Performance of Traffic Responsive and Time-of-Day Traffic Signal Control. Transpn. Res. 10, pp. 75-81 (1976).

[22] Kalman, R. E., A New Approach to Linear Filtering and Prediction Problem, J. Basic Engng. 82 D (1), pp. 35-45 (1960).

[23] Davis, Gary A. and Nancy L. Nihan, "Nonparametric Regression and Short Term Freeway Traffic Forecasting," ASCE Journal of Transportation Engineering, Vol. 117, No. 2, Mar/Apr. 1991, pp. 178-188.

[24] Widrow, B. and Stearns, S. D., Adaptive Signal Processing. Prentice-Hall, Inc., Englewood Cliffs, N. J. 07632 (1987).

[25] Passi, R. M. and Carpenter, M. J., Prediction and Frequency Tracking of Nonstationary Data with Application to the Quasi-Biennial Oscillation. Monthly Weather Review, Vol. 114, pp. 1272-1277 (1986).

[26] Box, G. E. P. and Jenkins, G. M., Time Series Analysis: Forecasting and Control. Rev. ed. Holden-Day (1976).

[27] Brown, R. G., Smoothing and Prediction of Discrete Time Series. Prentice-Hall, Englewood Cliffs, NJ. (1963).

[28] Trigg, D. W. and Leach, A. G., Exponential Smoothing with Adaptive Response Rate. Operations Research Quarterly, Vol. 18, No. 1, March 1967, pp. 53-59.

[29] Percival, D. and Walden, A., Spectral Analysis Textbook. University of Washington (1990).

[30] BMDP Statistical Software Manual, Vol. 2, University of California Press (1988).

[31] Lind, R. F., and Roberts, H.V., Interactive Data Analysis Forecasting System, Scientific Press/McGraw, 1982.

[32] Box, G. E. P. and Pierce, D. A., Distribution of Residual Autocorrelation in Autoregressive Integrated Moving Average Time Series Models. Journal of the American Statistical Association, Vol. 65, 1970, pp. 1509-1526.

[33] Jenkins, G. M. and Watts, D. G., Spectral Analysis and its Application. Holden-Day, San Francisco (1969).

[34] Robinson, E. A., Random Wavelets and Cybernetic Systems. Hanfer Publishing Company, New York (1962).

[35] Highway Capacity Manual. TRB Special Report 209. Transportation Research Board, Washington, D.C. (1985).

APPENDIX A

State of the Art

Over the past two decades, many traffic flow forecasting methods have emerged to forecast volume, or occupancy, or both (1-23). These methods fall into two categories: parametric and nonparametric. In urban traffic control systems, parametric methods are divided further into two categories: second generation and third generation methods. The former is designed for control intervals of 5-15 minutes, and the latter on a cycle-by-cycle basis. Second generation algorithms are older and typically require extensive historical data. They use current traffic measurements to correct for the traffic deviation from the average historical pattern. The second generation UTCS(UTCS-2) (1), ASCOT (2), and ASCOT-RTOP (3) all belong to this category.

Most of the third generation algorithms (4,5), were developed more recently. They make predictions based solely on current traffic measurements. The best-known method in this category, the third generation UTCS(4), requires a 'representative' data set for parameter estimation (6).

More recent studies have used Box-Jenkins type analyses of time series data (7, 8, 9, 10, 11, 12), spectral analysis (13), Kalman filtering (14, 15, 16, 17), automatic control concepts (18), and adaptive prediction system analysis (19) to forecast freeway traffic flow. Other researchers used nonparametric methods to forecast freeway traffic volume (23).

PARAMETRIC METHODS

UTCS-2

The second-generation UTCS-2, which is an urban traffic control method, predicts the next control interval (5-15 minutes) traffic volume at each detector location in real time, based solely on the measurements from the same location. The algorithm makes use

of both smoothed historical traffic data and current traffic-volume measurements from the vehicle detector.

The UTCS-2 set of equations has been presented elsewhere, but the complexity of the solution is usually not shown. By solving the difference equations of UTCS-2 (1), it can be shown that UTCS-2 results in the following demand prediction equation:

$$\hat{V}(t) = M(t) + p\left[M(t-1) - p \cdot f(t-1)\right] + (1-q)A + p(1-q)B$$

in which

$$M(t) = a(0) + \sum_{i-1}^{k}\left[a(i)\cos(2nit/N) + b(i)\sin(2nit/N)\right]$$

$$A = \sum_{s=0}^{t-1}\left\{q^s\left[f(t-s-1) - M(t-s-1)\right]\right\}$$

$$B = \sum_{s=0}^{t-2}\left\{q^s\left[f(t-s-2) - M(t-s-2)\right]\right\}$$

where

$\hat{V}(t)$ = predicted volume at time t,

$M(t)$ = historical volume at time t,

$f(t)$ = measured volume at time t,

p = constant computed off-line from representative volume data of the location in question (e.g.,for the UTCS system in Washington, D.C., p was 0.2),

q = smoothing coefficient (e.g., for the UTCS system in Washington, D.C., q was 0.9),

$a(0), a(i), b(i)$ = coefficients (computed off-line) of Fourier series approximation of historical traffic patterns for each measurement location,

k = user input parameter determining the fidelity of Fourier series approximation, which is usually the result of a trade-off between Fourier series accuracy and storage space and computation effort (in general, for more rapidly varying functions, higher values of k should be used; k-values from 6 to 20 have been used in past applications (6)),

n = number of sample points of a data set, and

N = total number of time intervals in the a data set (e.g., for 15- minute intervals, the data for a 24-hour day will consist of 96 intervals).

The UTCS-2 prediction equation is a function of

$$\hat{V}(t) = \{M(t), f(t), n, q\}$$

UTCS-3

The third-generation UTCS-3 predicts traffic volume two control intervals into the future. Like UTCS-2, UTCS-3 is also an urban traffic control method, forecasting the volume at each location in real time, based on measurements from the same location. It differs from UTCS-2 in that the prediction process relies solely on current-day measurements (no historical traffic pattern is required for prediction). By solving the difference equations of UTCS-3 (4), it can be seen that UTCS-3 results in the following demand prediction equation:

$$\hat{V}(t+j) = p \cdot f(t) + (1-p)\left[qr^t + (1-r)A\right]$$

in which

$$A = \sum_{s=0}^{t-1}\left[r^s f(t-s-1)\right]$$

where

$\hat{V}(t+j)$ = predicted volume for time $(t+j)$ at time t,

p = extrapolation constant computed off-line from a volume data of the location in question,

$f(t)$ = measured volume at time t,

q = exponentially smoothed volume measurement, also referred to as "coarse prediction of volume" , and

r = smoothing coefficient (a value of 0.95 has been used in past applications (6)).

It can be seen that the UTCS-3 prediction equation is a function of

$$\hat{V}(t) = \{f(t), q, n, r\}$$

Previous performance tests have indicated that UTCS-2 performs better than a number of existing algorithms (17). For example, it consistently performs better than UTCS-3, with a lower mean square and a lower mean absolute error (6,20,21), and a large portion of small-magnitude errors (20). UTCS-2 is not subject to an inherent time lag (as is UTCS-3), therefore, it provides reasonably good values during a vehicle detector outage, and is available as soon as detector operation is restored (17).

Time Series Model

A time series model calculates autocorrelation and partial correlation coefficients to fit the historical data and to forecast future values, assuming that these correlation values are kept nearly constant. The general formula is as follows:

$$A(B) \cdot W(t) = a_0 + C(B) a_t$$

in which

$$A(B) = x_0 + x_1 B + x_2 B^2 + \cdots + x_p B^p$$

$$C(B) = y_0 + y_1 B + y_2 B^2 + \cdots + y_q B^q$$

$x_0, \ldots, x_p, y_0, \ldots, y_q$ can be estimated by the least square mean method.

B is a backward factor, that is $B \times W(t) = W(t-1)$,

p, and q can be determined by examining the correlation coefficients,

$W(i)$ can be a time series or any order differentiated time series, $i = t, t\text{-}1, t\text{-}2, \ldots, t\text{-}p$,

a_0 is a constant coefficient, and

a_i is a noise time series, $i = t, t\text{-}1, t\text{-}2, t\text{-}q$.

This method can be expected to perform better than other simple regression methods because it analyzes the relationship among the past observed data. Well-known time series techniques were developed by Box-Jenkins (26), and predictions using this type of analyses have resulted in good accuracy of predictions (7,8,9,10,11,12).

Ahmed and Cook investigated the application of analysis techniques developed by Box and Jenkins for freeway traffic volume and occupancy time series. (7) A total of 166 data sets from three surveillance systems in Los Angeles, Minneapolis, and Detroit were used in the development of a predictor model to provide short-term forecasts of traffic data. The Los Angeles data were 20-second volumes and occupancies per lane, and the data from Minneapolis and Detroit were volumes and occupancies aggregated over all lanes at 30- and 60-second intervals, respectively. All of the data sets were represented by an autoregressive integrated moving average (ARIMA) (0,1,3) model. The moving-average parameters of the model, however, varied from location to location and over time.

For the purpose of comparing the smoothing performance of the different models, values of mean absolute error (MAE) and mean square error (MSE) of the fitted ARIMA (0,1,3) models were chosen as a basis for comparison. These values ranged from 1.3 to 6.5 for MAE, and from 2.8 to 91.41 for MSE. Results of the moving-average model indicate that both MAE and MSE increase when the number of observations (N) was increased. When N equaled five, the ratio to Box-Jenkins varied between 1.0 to 1.27 for MAE, and between 1.0 to 1.45 for MSE. Larger values of N (10-100) resulted in ratios to Box-Jenkins of between 1.0 to 2.85 for MAE, and between 1.0 and 6.86 for MSE. The best results of the double-exponential smoothing model were associated with small values

of the parameter α. For smoothing constants between 0.1 and 0.3, the ratios to Box-Jenkins ranged from 1.0 to 1.64 for MAE and from 1.0 to 1.43 for MSE. The Trigg and Leach model did not improve the forecasts. Even in the best results of this model, the ratios to Box-Jenkins varied between 1.45 and 8.20 for MAE and between 2.08 and 44.34 for MSE. The ARIMA (0,1,3) model more accurately represented the stochastic process that generated the traffic data.

The researchers also believed that a rapid adjustment in the parameter estimates--each observation interval, for example--might degrade the overall forecasting performance of the ARIMA model. They suggested that computer computational requirements should be taken into consideration when real-time updating of the model parameters is contemplated. One way to lower these requirements would be to update the parameters only occasionally, e.g., at the beginning of peak and off-peak periods.

Eldor developed short-term demand predictors for a real-time traffic-responsive freeway control system (8). Using real-life traffic counts from the Santa Monica Freeway in Los Angeles, a time-series analysis was performed which employed the Box-Jenkins technique. The analysis was applied to three typical freeway stations: an entrance ramp, an exit ramp, and a mainline station. Two types of predictors were investigated: second generation control (2-GC) predictors, using both historical and real-time data, and third generation control (3-GC) predictors, using only real-time data. After some simple models were tested, the ARIMA series (0,1,1), (0,1,0), and (0,2,1), were employed as predictors. A statistical and comparative analysis disclosed that the standard deviation of prediction error in all locations (with the 5-minute data) was approximately between five and ten vehicles per 5 minutes for the 2-GC and 3-GC predictors, according to the variance of prediction errors aggregated over all locations. The 2-GC predictor ranked first, followed by the 1-GC predictor; the 3-GC predictor was third and the random-walk model [ARIMA(0,1,0)] was last. The distribution of prediction errors according to size (in vehicles/5-minutes) served as another means for comparing the 2-GC and 3-GC

predictors, and results showed that the 2-GC predictor performed better than the 3-GC predictor in terms of percentage of underestimation. On the other hand, the 2-GC had a higher tendency to overestimate traffic with a higher proportion of small-magnitude errors (0-2 vehicles/minute). Other analyses suggested that the 1-GC predictor, namely average historical values, was the best predictor over long time intervals (10- and 15-minute) because then the aggregation cancelled out any stochastic variations, leaving the flow level as the primary changing variable. In order to provide a high response in real-time, therefore, shorter, rather than longer, time intervals for control updates were desired. The consideration of 10- or 15-minute intervals was perhaps more appropriate for fixed-time control operations with the same control strategies (stratified by time-of-day) for each day of operation.

Another researcher used the time series technique to forecast density (9). This study relied on continuous records of vehicle count in roadway sections of various lengths. These records, previously constructed from sequential aerial photographs, were sampled at uniformly spaced, discrete time segments. The time interval was based on Nyquist frequency estimates, which were obtained from the continuous process. The time series analytical techniques of Box and Jenkins were used to identify the structure of autoregressive moving average models. The results indicated that the forecast functions worked reasonably well up to roughly 5, 10, and 20 second lead times for the 92, 305, and 558 meter test sections, respectively. Each result approached the respective mean value of the process. Gafarian, Pahl and Ward discussed the applications of time series analysis in two different situations:

(1) construction of macroscopic models of large freeway systems in which the density of various sections of roadway could be simulated by using discrete time increment models, and

(2) traffic responsive control of systems such as freeways, bridges, and tunnels in which forecasts of density were desired.

They pointed out that in the design of other studies, and in the preliminary design of a particular system, the relation between section length and record length had to be considered. The correlation time of the density process increased as the physical length of the roadway section increased; thus, if density was measured over long sections, correspondingly long historical records would be required to fit models and to forecast the process. For traffic responsive control, the computational task had to be considered. However, Gafarian found this analysis, at times, resulted in unsatisfactory goodness of fit and high errors (9).

Levin and Tsao used the Box-Jenkins time-series method to analyze 20-, 40-, and 60-second occupancy and volume data collected during a morning rush period at two freeway locations of the Dan Ryan Expressway in Chicago: one on the local lanes and the other on the express lanes (10). Several autoregressive integrated-moving-average (ARIMA) models were evaluated, and the ARIMA(0,1,1) model was found to be the most statistically significant for all forecasting intervals for both volume and occupancy. A comparison between the ARIMA(0,1,1) model and the ARIMA(0,1,0) model showed significant improvement in forecasting volume. The 60-second forecasting interval was found to be the most effective interval. When compared with occupancy forecasting, volume forecasting was found to be less variable, as expressed by the ratio of the residual sum of squares to the mean of the observations. In the same way, forecasts of volumes and occupancies on the express lanes were found to be less variable than those on the local lanes.

Nihan and Holmesland explored using times series techniques for short-term traffic volume forecasts (11). A data set containing monthly volumes on a freeway segment for the years 1968 through 1976 was used to fit a time series model. The resulting model was used to forecast volumes for the year 1977, and the predicted volumes were then compared to actual volumes in 1977. With the month of December, 1976, as an origin for the forecast, the September error was 7.5 percent, while all other errors were around 5

percent or less. The results of this study imply that time series techniques could be used to develop highly accurate and inexpensive short term forecasts. The researchers also stated that the main strength of the Box and Jenkins technique was its accuracy; its second strength, and the main reason for using the model, was the ease of its application after having been specified; a third advantage was its flexibility. The last advantageous feature of the Box and Jenkins technique was the speed of detection in changes taking place in a time series, which could be used as an early warning system or as an early evaluation of impacts from changes in the environment. They recommend the use of this short term traffic forecast for the following:

1. Implementation and control of long-range plans,
2. Tactical decisions and traffic system operations,
3. Evaluation of effects of interventions, where decisions were being monitored for feedback,
4. Advanced warning in monitoring systems, and
5. Optimal control in systems which were demand responsive.

Davis and Nihan designed a time series model to estimate changes in freeway level of service despite missing data (12). They considered volume and occupancy as dependent variables, and developed the multivariable regression model for each of these variables using the Box-Jenkins time series method to check whether or not the error term was uncorrelated. This multivariable regression model could detect relatively small changes in traffic stream measures, and provided a separation of intervention effects and effects attributable to other factors. The model could also easily be extended to estimate effects from several overlapping interventions. The results from the peak 15-minute volume data showed a significant effect, which became insignificant when the dependence of the residuals was taken into account. This showed how an erroneous interpretation can result from the naive assumption of statistical independence in time series data. As the results indicate, the time series model can be used to investigate fairly subtle interactions

among the traffic stream, control policies and external factors. Since the required data were collected automatically, time series methods could be quite cost-effective, both in policy evaluation and in more theoretical research in traffic stream phenomena.

Spectral Analysis

Another branch of time series methods is spectral analysis, which was employed by Nicholson and Swann to forecast traffic flow volumes (13). If the basic pattern of traffic flow tends to repeat itself, then it is possible to consider the time series as being a part of a time series group. The problem, therefore, is to predict a nonstationary process, given a group of sample functions. It is appropriate, then, to consider that the periodic behavior can be represented in terms of the characteristic modes or functions associated with a covariance data matrix based on previous data. In this particular application, the maximum prediction errors were of the order of 8 percent and 11 percent for the morning and afternoon periods, respectively (13). In most of the cases examined, the afternoon prediction was found to be less accurate than the morning prediction. Reducing the prediction period and incorporating recent logged data would considerably improve the prediction accuracy. Predictions obtained by projecting the average values of previous data produced greater errors: 12 percent and 19 percent for the morning and afternoon periods, respectively. If the flow pattern could be described adequately in terms of a single eigenvector component, the problem of calculating several of the coefficients could be eliminated. It would then be possible to have an eigenvector length equivalent to a full day without excessive computation, thus permitting continuous prediction up to 24 hours ahead. The peak hour results indicate that application of the spectral expansion technique to the prediction of traffic flow data could be considerable. The prediction algorithms could be adapted for continuous on-line implementation to traffic control. The main disadvantage of the method was its inability to account for rapid unforeseen changes that were not reflected in the covariance matrix based on the characteristics of previous data.

Kalman Filtering

Kalman filtering is based on a theory proposed by Kalman, whose objective was to obtain the linear dynamic system specifications that resulted in the prediction, separation, or detection of a random signal (22). This method can be applied to short-term stationary or nonstationary stochastic phenomena. It has been applied very successfully to traffic systems for demand forecasting (14,17). The general formula is

$$\hat{V}(t+k) = \mathbf{W}(t) \cdot \mathbf{Q}(t)$$

where

$\hat{V}(t+k)$ is a predicted value k steps ahead at time t;

$\mathbf{W}(t) = \mathbf{W}(t-1) + A$

$\mathbf{W}(t)$ is a weight vector;

A stands for the second term, which changes step by step according to the difference between the predicted value and actual value of the last step;

$\mathbf{Q}(t)$ = the vector, composed of the previous actual time-series data of $V(t)$, and

$V(t)$ = the time series.

Gazis developed a method for estimating the number of vehicles on a section of roadway from speed and flow measurements at the entrance and exit points of that section (14). This method basically consists of three steps: 1) computing and storing the velocity and flow information; 2) computing the travel time between two points, then obtaining rough estimates of the number of vehicles; and 3) using Kalman filtering theory to filter random errors of these estimates by means of a sequential correction scheme. This algorithm was tested using data from three adjoining half-mile sections in the Lincoln Tunnel, and the exact counts were compared with those generated by the algorithm. The results indicate that 99 percent of time the error was below 10 percent. This level of accuracy had not previously been obtained using flow and speed data (14). The researcher pointed out that further improvement of the technique would require a refinement of the

travel time algorithm that reduced day-to-day variations in the bias error. The instrumentation required for this algorithm was less than that required by previously used algorithms based on recognition of patterns of car lengths. Only one detector per trap was required, compared with two in the past. The estimation algorithm could be applied to multi-lane highways, depending on the distance between traps and the frequency of lane changing. However, care must always be exercised in selecting the location of the traps.

Based on the above research, Chang and Gazis extended the Kalman filtering methodology to include explicit consideration of lane-changing on a multi-lane freeway (15). They found that the estimation error could be reduced, the reduction ranging up to 40 percent. When the roadway section was larger, resulting in an increased number of lane-changes, the error reduction was greater. Therefore, where lane-changing is frequent and when a reduction instrumentation cost is desired, it may be worthwhile to invest in proper calibration of a density estimator that includes lane-changing.

Nahi applied Kalman filtering theory to freeway traffic data processing (16). Two basic data processing problems associated with freeway traffic were formulated: estimation of traffic variables (section mean speed and density) and detection of occurrence of an incident or accident within a given section of the road. The results were very encouraging. The estimators exhibited a kind of "step response" nature which was indicative of acceptable tracking capability in the case of large variations in the real traffic situation.

Okutani and Stephanedes employed Kalman filtering theory for predicting 15-minute volumes during the day, using the traffic flow on the study section, as well as the traffic flows on other sections feeding into it (17). They developed two algorithms: one (M1) used the difference between the traffic volume on the day in question and the traffic on a day the week before as variable (rather than traffic volume on the day itself); the other (M2) considered the similarity of the traffic flow pattern from day to day. Three

prediction error indices were computed: (1) mean relative error, (2) root relative square error, and (3) maximum relative error. The results indicated that

(1) at all times all prediction models performed substantially (up to 80 percent) better than UTCS-2;

(2) as a rule, prediction model M1 outperformed prediction model M2, possibly because the latter prediction was accomplished using traffic data from a day the week before the day the study was taken;

(3) when prediction was performed using smoothed (rather than raw) data, UTCS-2 performance declined; whereas, M1 achieved its best performance under these conditions; and

(4) by increasing the number of 5-minute time intervals ahead of current time for which prediction was performed from one to nine in the models, performance was not significantly affected. Such performance is highly desirable for long-term prediction.

The researchers also pointed out that the new models seemed promising for practical applications (17). Prediction error had improved, and the computation time required was reasonable. These models could be used for predicting a variety of traffic characteristics, such as time occupancy and traffic density.

Other Parametric methods

Papageorgiou applied automatic control concepts to traffic flow modeling and control and developed a traffic flow forecasting model (18). This dynamic time-of-day control model required estimates of the on-ramp to off-ramp origin-destination matrix (18). It estimated traffic speed from measurements of volumes and lane occupancy, then forecast the volume at a downstream station based on known volumes at upstream on-ramps and mainline stations. He compared two control models: a steady-state time-of-day control model, which assumed the on-ramp to off-ramp origin-destination matrix to be

constant, and a dynamic time-of-day control model, and found that the dynamic time-of-day control model performed better than the steady-state time-of-day control model.

Other researchers employed the adaptive prediction system, the structure of which was adjustable so that its performance improved as a real-time predictor, through contact with its environment, to forecast the speed and volume of traffic flow (19). The method of steepest descent to obtain the adaptive weight vector was used here, which could be adjusted in the direction of the gradient at each step. This adaptive prediction system could be used as a real-time predictor. The performance of the predictor depends on both the characteristics of the traffic variables and the structure of the predictors. In practice, for the prediction of a specific traffic variable, an adequate number of tests should be run to obtain the optimal structure of the predictors. For the control of a traffic system, many traffic characteristics must be predicted. If these characteristics are stationary, the adaptive prediction system could be used as the predictor, even when some traffic characteristics change greatly from one level to another level. After several steps, the predicted values were almost the same as the observed data. The formula developed is similar to that used in the Kalman filtering method. The results were satisfactory, but the traffic characteristics were assumed to be stationary. In fact, when the time interval was less than 5 minutes, volume and occupancy were nonstationary (7,8,10). This method is the steepest descent type of adaptive algorithms, and was improved further when it was applied to forecast nonstationary volume and occupancy (24,25).

NONPARAMETRIC METHODS

When a parametric method is used to forecast traffic flow, it has been established that a regression model should be selected before estimating the model parameters. In this process, two problems can arise: 1) the regression model used may be wrong, or 2) not all samples to estimate the parameters may be available, (in this case, the estimated parameters will differ from the true parameter values). Therefore, although different

classes of models can be tested, the parametric method will only generate approximations of the forecasts.

Davis and Nihan suggested a nonparametric regression method, the *k*-NN method, to forecast traffic flow (23). In the *k*-NN method, we assume that we have a series of observations [$\mathbf{x}(s)$, $\mathbf{y}(s)$, s = 1, ..., n] of input/output pairs, which we call our learning sample, and an additional input measurement $\mathbf{x}(t)$, from which we want to forecast $\mathbf{y}(t)$. First, the *k*-NN method ranks the input measurements in the learning sample $\mathbf{x}(s)$ according to their distance from $\mathbf{x}(t)$. Let s_1, ..., s_k denote the indices of the k input vectors closest to $\mathbf{x}(t)$. The forecast is then computed simply as $\hat{\mathbf{y}}(t) = \frac{1}{k}\sum_i \mathbf{y}(s_i)$ (i.e. as the average of the outputs corresponding to the k nearest neighbors of $\mathbf{x}(t)$). Thus, the *k*-NN approach replaces the problem of selecting a class of models and then estimating the model parameters, which entails the problem of maintaining and sorting an adequately large learning sample.

An empirical study has shown that this method performed comparably but was not better than the linear time-series approach (23).

SUMMARY

This chapter reviewed several existing methods, mainly parametric, that have been used to forecast volume and occupancy. When choosing among these models, the following points should be considered:

1. Every parametric method has parameters that differ from place to place and over time.
2. Some methods are very simple, like the simple moving average, and some are very complicated, like the spectral method. The more complicated methods generally require additional computations.
3. The same method may have different forms, like the Box-Jenkins techniques. When these methods are used, the model should fit the particular data.

4. Time interval is also an important factor in forecasting. Generally, when the interval is below 5 minutes, the traffic variables are nonstationary, and the best way to forecast these variables is to use more complicated methods. However, when the time interval is greater than 10 minutes, simple models should be compared to complex models, so that the best model can be chosen to forecast traffic variables. Sometimes the simpler model is a better predictor.

5. So far, these methods have not been compared against the same data sets; therefore, careful selection of a model is essential.

APPENDIX B

Derivation of Recursive Model

First, we should prove that if

$$\mathbf{A}(k)=\sum_{t=1}^{k}\mathbf{V}^{\mathrm{T}}(t)\mathbf{V}(t)=\mathbf{A}(k-1)+\mathbf{V}^{\mathrm{T}}(k)\mathbf{V}(k) \tag{B.1}$$

then

$$\mathbf{A}^{-1}(k)=\mathbf{A}^{-1}(k-1)-\frac{\mathbf{A}^{-1}(k-1)\mathbf{V}^{\mathrm{T}}(k)\mathbf{V}(k)\mathbf{A}^{-1}(k-1)}{1+\mathbf{V}(k)\mathbf{A}^{-1}(k-1)\mathbf{V}^{\mathrm{T}}(k)} \tag{B.2}$$

Proof

Multiplying both sides of equation (B.1) by $\mathbf{A}^{-1}(k)$, we get

$$1=\mathbf{A}^{-1}(k)\mathbf{A}^{-1}(k-1)+\mathbf{A}^{-1}(k)\mathbf{V}^{\mathrm{T}}(k)\mathbf{V}(k)$$

then multiplying both sides by $\mathbf{A}^{-1}(k)\mathbf{V}^{\mathrm{T}}(k)$, we have

$$\mathbf{A}^{-1}(k-1)\mathbf{V}^{\mathrm{T}}(k)=\mathbf{A}^{-1}(k)\mathbf{V}^{\mathrm{T}}(k)+\mathbf{A}^{-1}(k)\mathbf{V}^{\mathrm{T}}(k)\mathbf{V}(k)\mathbf{A}^{-1}(k-1)\mathbf{V}^{\mathrm{T}}(k)$$

$$=\mathbf{A}^{-1}(k)\mathbf{V}^{\mathrm{T}}(k)\left[1+\mathbf{V}(k)\mathbf{A}^{-1}(k-1)\mathbf{V}^{\mathrm{T}}(k)\right]$$

multiplying the both sides of the above equation by

$$\left[1+\mathbf{V}(k)\mathbf{A}^{-1}(k-1)\mathbf{V}^{\mathrm{T}}(k)\right]\mathbf{V}(k)\mathbf{A}^{-1}(k-1),$$

we obtain

$$\mathbf{A}^{-1}(k-1)\mathbf{V}^{\mathrm{T}}(k)\left[1+\mathbf{V}(k)\mathbf{A}^{-1}(k-1)\mathbf{V}^{\mathrm{T}}(k)\right]^{-1}\mathbf{V}(k)\mathbf{A}^{-1}(k-1)$$

$$=\mathbf{A}^{-1}(k)\mathbf{V}^{\mathrm{T}}(k)\mathbf{V}(k)\mathbf{A}^{-1}(k-1)$$

therefore,

$$\mathbf{A}^{-1}(k-1)-\frac{\mathbf{A}^{-1}(k-1)\mathbf{V}^{\mathrm{T}}(k)\mathbf{V}(k)\mathbf{A}^{-1}(k-1)}{1+\mathbf{V}(k)\mathbf{A}^{-1}(k-1)\mathbf{V}^{\mathrm{T}}(k)}$$

$$\begin{aligned}
&= \mathbf{A}^{-1}(k-1)-\mathbf{A}^{-1}(k)\mathbf{V}^{\mathrm{T}}(k)\mathbf{V}(k)\mathbf{A}^{-1}(k-1)\\
&= \mathbf{A}^{-1}(k-1)-\mathbf{A}^{-1}(k)\left[\mathbf{A}(k)-\mathbf{A}(k-1)\right]\mathbf{A}^{-1}(k-1)\\
&= \mathbf{A}^{-1}(k-1)-\mathbf{A}^{-1}(k-1)+\mathbf{A}^{-1}(k)\\
&= \mathbf{A}^{-1}(k)
\end{aligned}$$

this results in equation (B.2).

Second, let

$$\mathbf{M}(k)=\sum_{t=1}^{k}V_d(t)\mathbf{V}^{\mathrm{T}}(t)$$

so

so

$$\mathbf{K}(k)=\mathbf{A}^{-1}(k)\mathbf{V}^{\mathrm{T}}(k) \tag{B.3}$$

From the equation (B.2), we have

$$\mathbf{K}(k)=\left[\mathbf{A}^{-1}(k-1)-\frac{\mathbf{A}^{-1}(k-1)\mathbf{V}^{\mathrm{T}}(k)\mathbf{V}(k)\mathbf{A}^{-1}(k-1)}{1+\mathbf{V}(k)\mathbf{A}^{-1}(k-1)\mathbf{V}^{\mathrm{T}}(k)}\right]\mathbf{V}^{\mathrm{T}}(k)$$

therefore

$$\mathbf{K}(k)=\frac{\mathbf{A}^{-1}(k-1)\mathbf{V}^{\mathrm{T}}(k)}{1+\mathbf{V}(k)\mathbf{A}^{-1}(k-1)\mathbf{V}^{\mathrm{T}}(k)} \tag{B.4}$$

APPENDIX C

Study Data and Plots

Volume Time Series at NE 185th St. Station

114	83	106	110	108	129	113	116	95	104	110
124	97	121	123	96	92	104	104	106	106	105
107	96	107	105	105	96	99	98	95	103	102
96	94	95	97	98	103	105	101	112	103	93
100	98	101	108	112	82	102	103	114	107	107
114	116	107	105	98	98	98	110	100	120	100
121	107	95	84	99	104	98	95	100	102	99
110	101	116	97	104	117	107	105	115	110	118
118	125	109	109	109	104	107	107	120	91	84
91	102	96	100	102	100	92	107	97	81	81
101	84	83	98	72	102	65	81	82	89	93
87										

Volume Time Series at NE 175th St. Station

15	11	4	4	17	7	8	11	15	5	8	9	10
7	6	5	6	5	6	6	6	4	6	4	6	6
3	1	6	6	15	5	5	5	5	6	5	5	6
5	7	5	5	8	5	6	6	6	5	5	6	5
7	7	5	6	5	6	5	6	6	6	9	6	9
7	8	5	7	5	7	9	8	9	10	12	5	7
10	9	6	7	12	5	9	6	9	5	7	15	7
7	10	5	5	7	6	8	5	6	7	12	7	2
10	6	15	11	4	4	11	7	9	5	13	12	10
9	8	7	6	9								

Volume Time Series at NE 162nd St. Station

112	125	108	82	116	117	106	105	110	125	112
112	108	115	111	112	109	102	98	116	99	105
105	106	106	106	107	110	105	101	110	98	109
110	107	113	114	111	107	112	105	112	104	112
99	101	110	100	105	109	105	102	109	111	111
104	107	113	108	112	112	112	103	101	115	114
109	107	111	118	106	114	102	122	97	117	104
102	122	113	109	103	105	112	127	109	120	107
111	124	121	121	114	106	119	104	107	111	109
106	118	110	99	102	103	111	88	117	97	98
88	100	104	69	104	96	98	87	85	85	77
104										

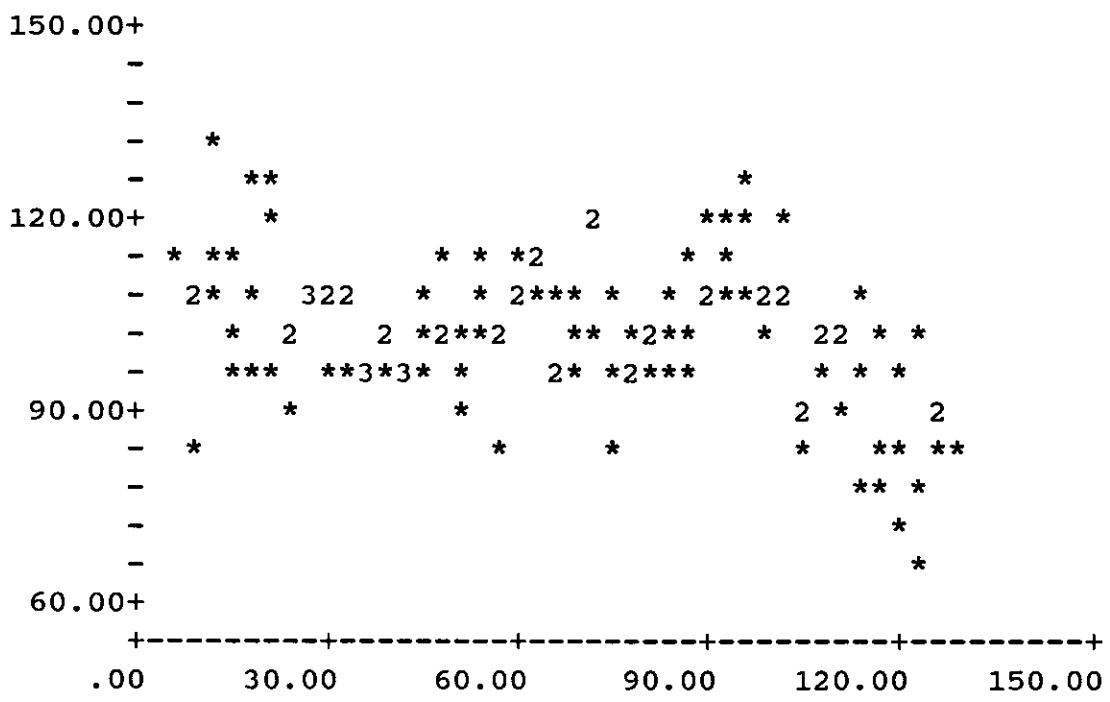


Figure C.1

Plot of Volume Time Series at NE 185th St. Station

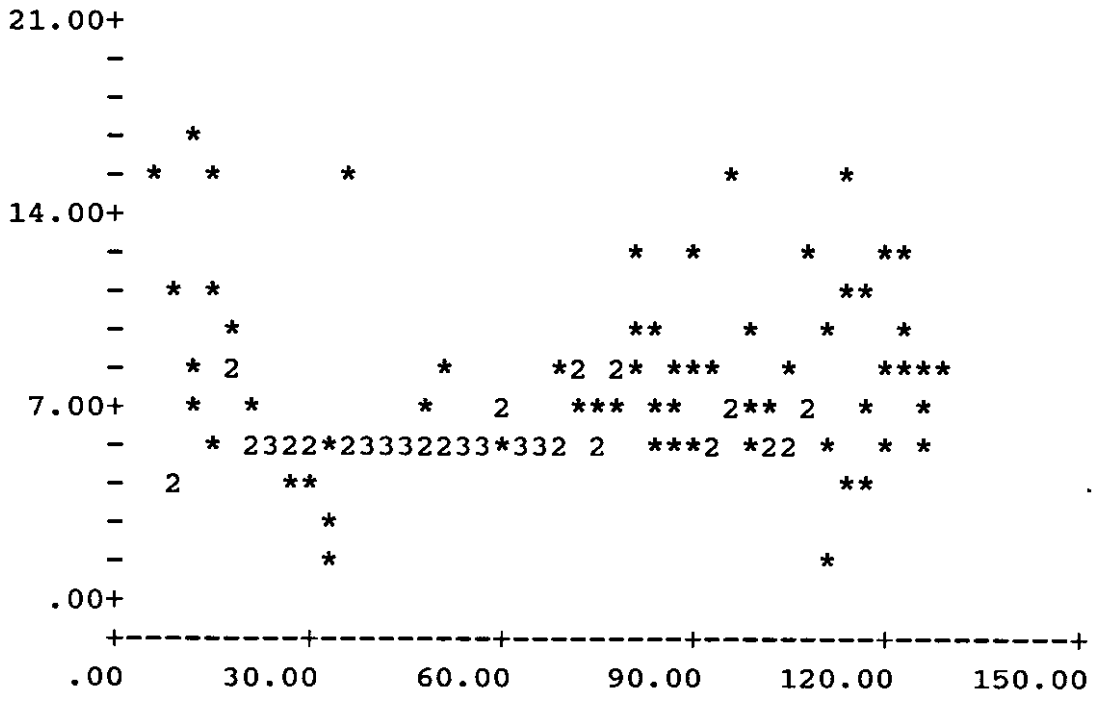


Figure C.2

Plot of Volume Time Series at NE 175th St. Station

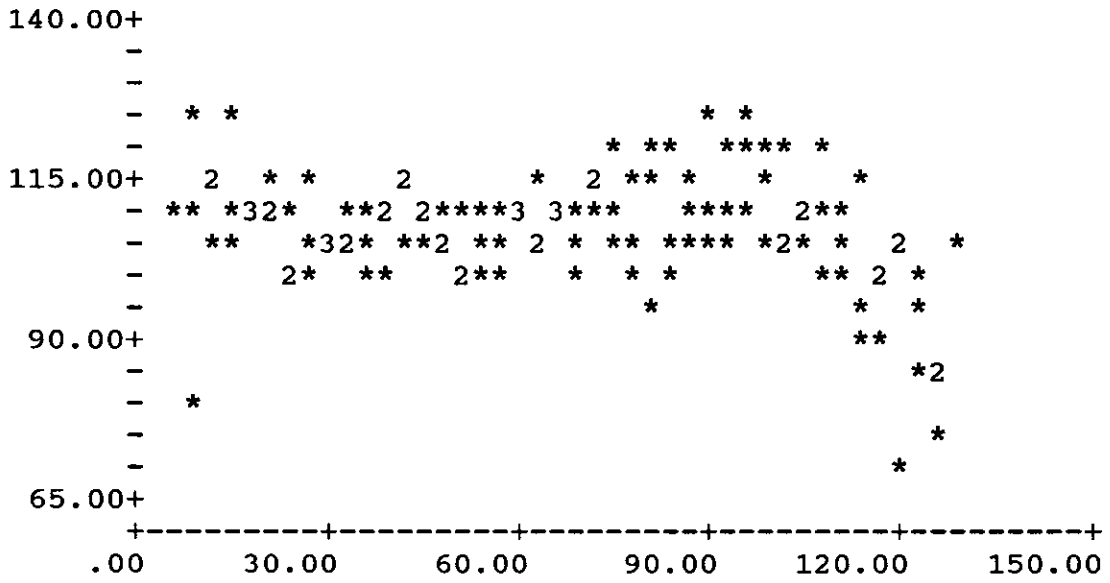


Figure C.3

Plot of Volume Time Series at NE 162nd St. Station

Occupancy Time Series at NE 185th St. Station

12.0	8.0	11.0	11.0	12.0	13.0	12.0	12.0	10.0	10.0	11.0
13.0	11.0	13.0	14.0	16.0	20.0	16.0	14.0	13.0	13.0	12.0
14.0	18.0	19.0	16.0	13.0	12.0	12.0	13.0	10.0	13.0	13.0
20.0	18.0	20.0	14.0	13.0	14.0	13.0	14.0	15.0	15.0	13.0
14.0	13.0	13.0	14.0	15.0	10.0	13.0	13.0	14.0	13.0	13.0
15.0	18.0	18.0	17.0	11.0	13.1	11.0	11.0	10.0	14.0	13.0
14.0	13.0	22.0	26.0	24.0	22.0	21.0	24.0	21.0	20.0	16.0
14.0	13.0	14.0	10.0	11.0	13.0	13.0	12.0	14.0	12.0	13.0
15.0	15.0	12.0	13.1	11.0	12.0	12.0	12.0	13.0	9.0	8.0
8.0	10.0	9.0	10.0	11.0	10.0	10.0	12.0	9.0	7.0	8.0
10.0	8.0	8.0	9.0	8.0	10.0	6.0	8.0	8.0	8.0	9.0
9.0										

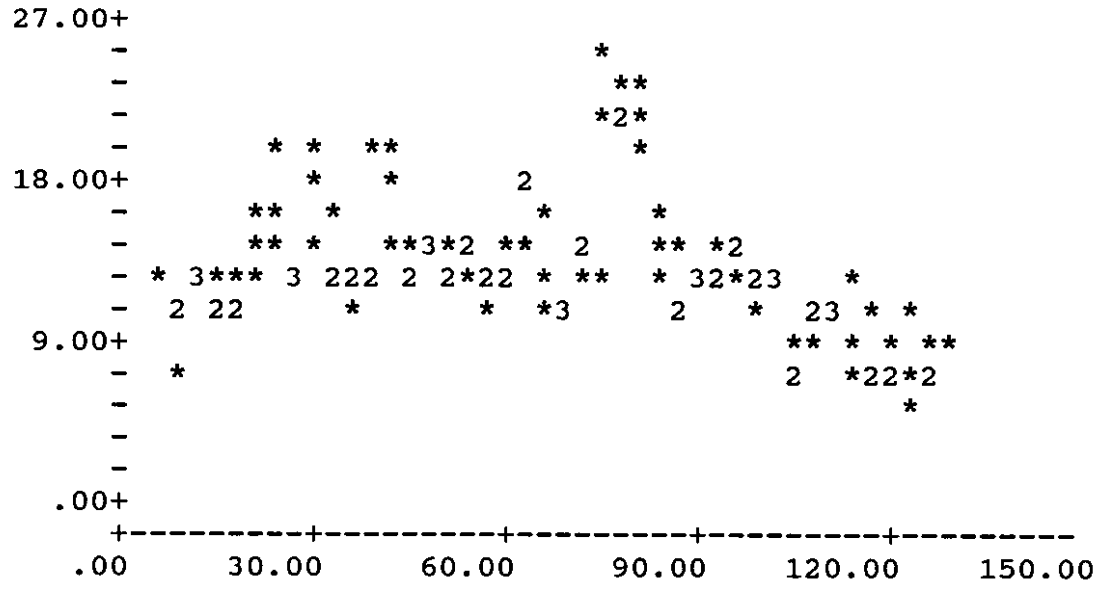


Figure C.4

Plot of Occupancy Time Series at NE 185th St. Station

APPENDIX D

Forecasting Results by Adaptive Prediction System Model

The forecast volumes by the adaptive system model are found in Table D.1 (see Chapter Three for details about the model).

Table D.1
Comparison of Predicted and Actual Volumes

Actual Volume	Predicted Volume	Differences
99	110.4	-11.4
102	108.4	-6.4
103	107.7	-4.7
111	105.9	5.1
88	106.8	-18.8
117	104.1	12.9
97	105.2	-8.2
98	103.7	-5.7
88	102.6	-14.7
100	99.1	0.9
104	98.2	5.8
69	98.9	-29.9
104	94.5	9.5
96	95.0	1.0
98	93.5	4.5
87	94.7	-7.7
85	91.5	-6.5
85	90.1	-5.1
77	88.7	-11.7
104	87.2	16.8

The above results can be obtained when N=10 and μ=0.0000004.

The three criteria for this model can be calculated as

$$(1)\quad E_{me} = \sum_{t=103}^{122} \left| \frac{V_d(t) - \hat{V}(t)}{V_d(t)} \right| \div 20 = 9.4\%$$

$$(2)\quad E_{sr} = \sum_{t=103}^{122} \sqrt{\left| \frac{V_d(t) - \hat{V}(t)}{V_d(t)} \right|} \div 20 = 0.3$$

$$(3)\quad E_{\max} = \operatorname{Max}_{t=103}^{122}\left|\frac{V_d(t)-\hat{V}(t)}{V_d(t)}\right| = 43\%$$

where E_{me}, E_{sr}, and $E_{\max}$ have be defined in Chapter Five, $V_d(t)$ is the actual volume at time t, and $\hat{V}(t)$ is the predicted volume at the same time t.

The forecast occupancies by the adaptive system model are found in Table D.2.

Table D.2
Comparison of Predicted and Actual Occupancies

Actual Occupancy	Predicted Occupancy	Differences
10	10.4	-0.4
11	10.3	0.7
10	10.2	-0.2
10	10.0	0.0
12	9.8	2.2
9	9.8	-0.8
7	9.8	-2.8
8	9.6	-1.6
10	9.5	-0.5
8	9.5	-1.5
8	9.4	-1.4
9	9.1	-0.1
8	8.9	-0.9
10	8.7	1.3
6	8.7	-2.7
8	8.1	-0.1
8	8.0	0.0
8	8.1	-0.1
9	8.1	0.9
9	8.0	1.0

The above results can be obtained when N=10 and μ=0.0004.

The three criteria for this model can be calculated as

$$(1)\quad E_{me} = \sum_{t=103}^{122}\left|\frac{O(t)-\hat{O}(t)}{O(t)}\right| \div 20 = 11.7\%$$

$$(2)\quad E_{sr} = \sum_{t=103}^{122}\sqrt{\left|\frac{O(t)-\hat{O}(t)}{O(t)}\right|} \div 20 = 0.29$$

$$(3)\quad E_{\max} = \operatorname{Max}_{t=103}^{122} \left| \frac{O(t) - \hat{O}(t)}{O(t)} \right| = 46\%$$

where E_{me}, E_{sr}, and $E_{\max}$ have be defined in Chapter Five, $O(t)$ is the actual occupancy at time t, and $\hat{O}(t)$ is the predicted occupancy at time t.

All these results were calculated by using the CYBER system.

APPENDIX E

Forecasting Results by Double Exponential Smoothing Model

The forecast volumes by the double exponential smoothing model are found in Table E.1 (see Chapter Three for details about the model).

Table E.1
Comparison of Predicted and Actual Volumes

Actual Volume	Predicted Volume	Differences
99	110.7	-11.7
102	108.3	-6.3
103	106.9	-3.9
111	105.9	5.1
88	106.6	-18.6
117	102.7	14.3
97	105.1	-8.1
98	103.2	-5.2
88	101.8	-13.8
100	98.6	1.4
104	98.4	5.6
69	98.9	-29.9
104	92.5	11.5
96	94.0	2.0
98	93.7	4.3
87	94.0	-7.0
85	92.0	-7.0
85	89.9	-4.9
77	88.2	-11.2
104	85.2	18.8

The above results were obtained when α=0.1, and N_0=95, at which point the exponential smoothing forecasting was made.

The three criteria for this model can be calculated using the results in the above table as

$$(1)\quad E_{me} = \sum_{t=103}^{122} \left| \frac{V_d(t) - \hat{V}(t)}{V_d(t)} \right| \div 20 = 10.5\%$$

$$(2)\quad E_{sr} = \sum_{t=103}^{122} \sqrt{\left| \frac{V_d(t) - \hat{V}(t)}{V_d(t)} \right|} \div 20 = 0.3$$

$$(3)\quad E_{\max} = \operatorname{Max}_{t=103}^{122}\left|\frac{V_d(t)-\hat{V}(t)}{V_d(t)}\right| = 43\%$$

where E_{me}, E_{sr}, and $E_{\max}$ have be defined in Chapter Five, $V_d(t)$ is the actual volume at time t, and $\hat{V}(t)$ is the predicted volume at the same time t.

If the double exponential smoothing model begins to forecast at point 100, the results obtained are as shown in the following table.

Table E.2
Comparison of Predicted and Actual Volumes

Actual Volume	Predicted Volume	Differences
99	112.6	-13.6
102	109.9	-7.9
103	108.2	-5.2
111	107.0	4.0
88	107.6	-19.6
117	103.4	13.6
97	105.8	-8.8
98	103.7	-5.7
88	102.3	-14.3
100	99.0	1.0
104	98.7	5.3
69	99.2	-30.2
104	92.7	11.3
96	94.1	1.9
98	93.8	4.2
87	94.0	-7.0
85	92.0	-7.0
85	89.9	-4.9
77	88.2	-11.2
104	85.1	18.9

The above results were obtained when α=0.1, and N_0=100, at which point the model begins to forecast.

The three criteria for this model can be calculated by using the results in the above table:

$$(1)\quad E_{me} = \sum_{t=103}^{122}\left|\frac{V_d(t)-\hat{V}(t)}{V_d(t)}\right| \div 20 = 10.7\%$$

$$(2)\quad E_{sr} = \sum_{t=103}^{122} \sqrt{\left|\frac{V_d(t) - \hat{V}(t)}{V_d(t)}\right|} \div 20 = 0.31$$

$$(3)\quad E_{\max} = \underset{t=103}{\overset{122}{\text{Max}}} \left|\frac{V_d(t) - \hat{V}(t)}{V_d(t)}\right| = 43.8\%$$

The forecast occupancies by the double exponential smoothing model are found in Table E.3.

Table E.3
Comparison of Predicted and Actual Occupancies

Actual Occupancy	Predicted Occupancy	Differences
10	9.9	0.1
11	9.8	1.2
10	9.9	0.1
10	9.8	0.2
12	9.8	2.2
9	10.1	-1.1
7	9.8	-2.8
8	9.2	-1.2
10	8.8	1.2
8	8.9	-0.9
8	8.6	-0.6
9	8.4	0.6
8	8.4	-0.4
10	8.2	1.8
6	8.4	-2.4
8	7.8	0.2
8	7.7	0.3
8	7.6	0.4
9	7.6	1.4
9	7.7	1.3

The above results were obtained when α=0.1, and N_0=82, at which point the forecasting was made.

The three criteria for this model can be calculated using the results in the above table:

$$(1)\quad E_{me} = \sum_{t=103}^{122} \left|\frac{O(t) - \hat{O}(t)}{O(t)}\right| \div 20 = 12\%$$

$$(2)\quad E_{sr} = \sum_{t=103}^{122} \sqrt{\left|\frac{O(t)-\hat{O}(t)}{O(t)}\right|} \div 20 = 0.31$$

$$(3)\quad E_{\max} = \operatorname*{Max}_{t=103}^{122} \left|\frac{O(t)-\hat{O}(t)}{O(t)}\right| = 40\%$$

where E_{me}, E_{sr}, and $E_{\max}$ have be defined in Chapter Five, $O(t)$ is the actual occupancy at time t, and $\hat{O}(t)$ is the predicted occupancy at time t.

If this model begins to forecast at point 100, then the results obtained are as shown in the following table.

Table E.4
Comparison of Predicted and Actual Occupancies

Actual Occupancy	Predicted Occupancy	Differences
10	10.6	-0.6
11	10.3	0.7
10	10.4	-0.4
10	10.2	-0.2
12	10.0	2.0
9	10.5	-1.5
7	10.0	3.0
8	9.0	-1.0
10	8.5	1.5
8	8.8	-0.8
8	8.4	-0.4
9	8.1	0.9
8	8.2	-0.2
10	8.0	2.0
6	8.4	-2.4
8	7.6	0.4
8	7.5	0.5
8	7.5	0.5
9	7.5	1.5
9	7.8	1.2

These results were obtained when α=0.15, and N_0=100, at which point the model begins to forecast.

The three criteria for this model can be calculated using the results in the above table:

$$(1)\quad E_{me} = \sum_{t=103}^{122} \left| \frac{O(t) - \hat{O}(t)}{O(t)} \right| \div 20 = 12.9\%$$

$$(2)\quad E_{sr} = \sum_{t=103}^{122} \sqrt{\left| \frac{O(t) - \hat{O}(t)}{O(t)} \right|} \div 20 = 0.332$$

$$(3)\quad E_{\max} = \operatorname{Max}_{t=103}^{122} \left| \frac{O(t) - \hat{O}(t)}{O(t)} \right| = 42.7\%$$

All of these results were calculated by using the CYBER system.

APPENDIX F

Forecasting Results by Exponential Smoothing with Adaptive Response Model

The forecast volumes by the exponential smoothing with adaptive response model are found in Table F.1 (see Chapter Three for details about the model).

Table F.1
Comparison of Predicted and Actual Volumes

Actual Volume	Predicted Volume	Differences
99	108.1	-9.1
102	105.6	-3.6
103	104.4	-1.4
111	103.8	7.2
88	104.6	-16.6
117	97.9	19.1
97	98.1	-1.1
98	98.1	-0.1
88	98.1	-10.1
100	96.3	3.7
104	96.6	7.4
69	96.9	-27.9
104	87.8	16.2
96	89.2	6.8
98	89.2	8.8
87	90.1	-3.1
85	89.9	-4.9
85	89.9	-4.9
77	89.6	-12.6
104	86.8	17.2

The above results were obtained when α=0.3, τ=0.1 and N_0=95, at which point the forecasting was made.

The three criteria for this model can be calculated by using the results in the above table:

$$(1)\quad E_{me} = \sum_{t=103}^{122} \left| \frac{V_d(t) - \hat{V}(t)}{V_d(t)} \right| \div 20 = 9.8\%$$

$$(2)\quad E_{sr} = \sum_{t=103}^{122} \sqrt{\left|\frac{V_d(t) - \hat{V}(t)}{V_d(t)}\right|} \div 20 = 0.28$$

$$(3)\quad E_{\max} = \underset{t=103}{\overset{122}{\text{Max}}} \left|\frac{V_d(t) - \hat{V}(t)}{V_d(t)}\right| = 39\%$$

where E_{me}, E_{sr}, and $E_{\max}$ have be defined in Chapter Five, $V_d(t)$ is the actual volume at time t, and $\hat{V}(t)$ is the predicted volume at the same time t.

If this model begins to forecast at point 100, then the results are as shown in the following table.

Table F.2
Comparison of Predicted and Actual Volumes

Actual Volume	Predicted Volume	Differences
99	110.2	-11.2
102	105.0	-3.0
103	103.2	-0.2
111	103.1	7.9
88	106.3	-18.3
117	95.5	21.5
97	102.7	-5.7
98	102.3	-4.3
88	101.7	-13.7
100	93.5	6.5
104	94.4	9.6
69	97.7	-28.7
104	81.5	22.5
96	85.3	10.7
98	89.5	8.5
87	94.2	-7.2
85	93.3	-8.3
85	91.2	-6.2
77	88.2	-11.2
104	80.1	23.9

The above results were obtained when α=0.4, τ=0.4 and N_0=100, at which time the model begins to forecast.

The three criteria for this model can be calculated by using the results in the above table:

$$(1)\quad E_{me} = \sum_{t=103}^{122}\left|\frac{V_d(t)-\hat{V}(t)}{V_d(t)}\right| \div 20 = 12.3\%$$

$$(2)\quad E_{sr} = \sum_{t=103}^{122}\sqrt{\left|\frac{V_d(t)-\hat{V}(t)}{V_d(t)}\right|} \div 20 = 0.325$$

$$(3)\quad E_{\max} = \underset{t=103}{\overset{122}{\text{Max}}}\left|\frac{V_d(t)-\hat{V}(t)}{V_d(t)}\right| = 41.6\%$$

The forecast occupancies by the exponential smoothing with adaptive response model can be found in Table F.3.

Table F.3
Comparison of Predicted and Actual Occupancies

Actual Occupancy	Predicted Occupancy	Differences
10	9.8	0.2
11	9.9	1.1
10	10.5	-0.5
10	10.4	-0.4
12	10.4	1.6
9	11.2	-2.2
7	10.6	-3.6
8	8.2	-0.2
10	8.1	1.9
8	8.2	-0.2
8	8.2	-0.2
9	8.1	0.9
8	8.3	-0.3
10	8.3	1.7
6	9.2	-3.2
8	8.1	-0.1
8	8.0	0.0
8	8.0	0.0
9	8.0	1.0
9	8.2	0.8

The above results were obtained when α=0.9, τ=0.3, and N_0=99, at which time the forecasting was made.

The three criteria for this model can be calculated using the results in the above table:

$$(1)\quad E_{me} = \sum_{t=103}^{122} \left| \frac{O(t) - \hat{O}(t)}{O(t)} \right| \div 20 = 12\%$$

$$(2)\quad E_{sr} = \sum_{t=103}^{122} \sqrt{\left| \frac{O(t) - \hat{O}(t)}{O(t)} \right|} \div 20 = 0.29$$

$$(3)\quad E_{\max} = \underset{t=103}{\overset{122}{\text{Max}}} \left| \frac{O(t) - \hat{O}(t)}{O(t)} \right| = 53\%$$

where E_{me}, E_{sr}, and $E_{\max}$ have be defined in Chapter Five, $O(t)$ is the actual occupancy at time t, and $\hat{O}(t)$ is the predicted occupancy at time t.

If this model begins to forecast at point 100, then the results are as shown in the following table.

Table F.4
Comparison of Predicted and Actual Occupancies

Actual Occupancy	Predicted Occupancy	Differences
11	9.9	1.1
10	11.0	-1.0
10	10.2	-0.2
12	10.0	2.0
9	11.9	-2.9
7	9.4	-2.4
8	7.0	1.0
10	7.6	2.4
8	9.9	-1.9
8	8.4	-0.4
9	8.0	1.0
8	8.9	-0.9
10	8.2	1.8
6	9.8	-3.8
8	6.3	1.7
8	7.4	0.6
8	7.9	0.1
9	8.0	1.0
9	9.0	0.0

The above results were obtained when α=0.8, τ=0.9, and N_0=100, at which time the model begins to forecast.

The three criteria for this model can be calculated using the results in the above table:

$$(1)\quad E_{me} = \sum_{t=103}^{122}\left|\frac{O(t)-\hat{O}(t)}{O(t)}\right| \div 20 = 15.7\%$$

$$(2)\quad E_{sr} = \sum_{t=103}^{122}\sqrt{\left|\frac{O(t)-\hat{O}(t)}{O(t)}\right|} \div 20 = 0.356$$

$$(3)\quad E_{\max} = \underset{t=103}{\overset{122}{\text{Max}}}\left|\frac{O(t)-\hat{O}(t)}{O(t)}\right| = 63.9\%$$

All these results were calculated using the CYBER system.

APPENDIX G

Forecasting Results by Box-Jenkins Time Series Model

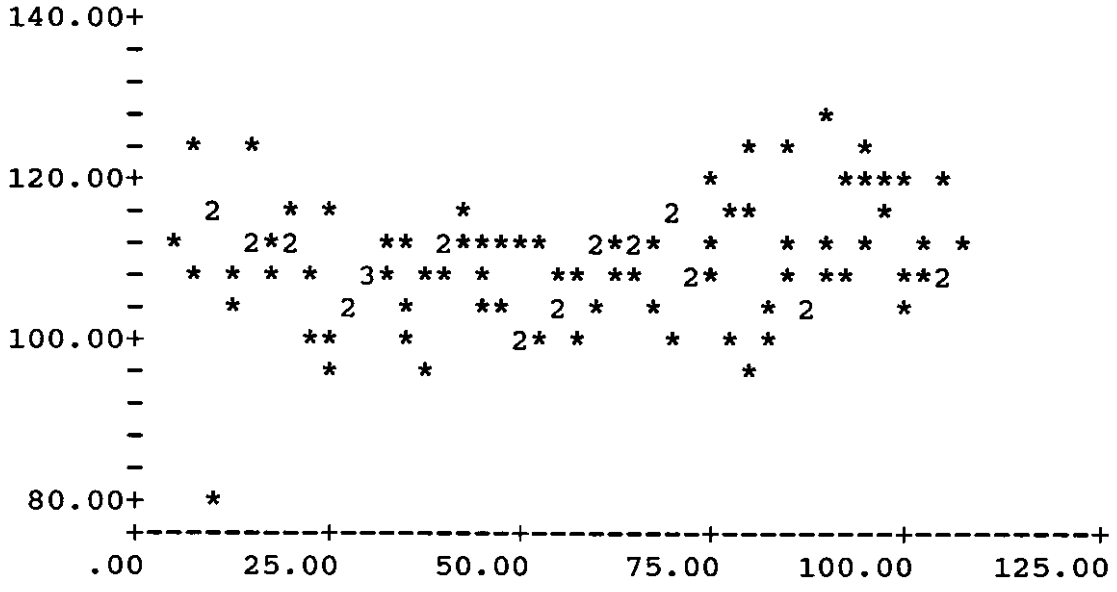


Figure G.1

Plot of the First 102 Data Points

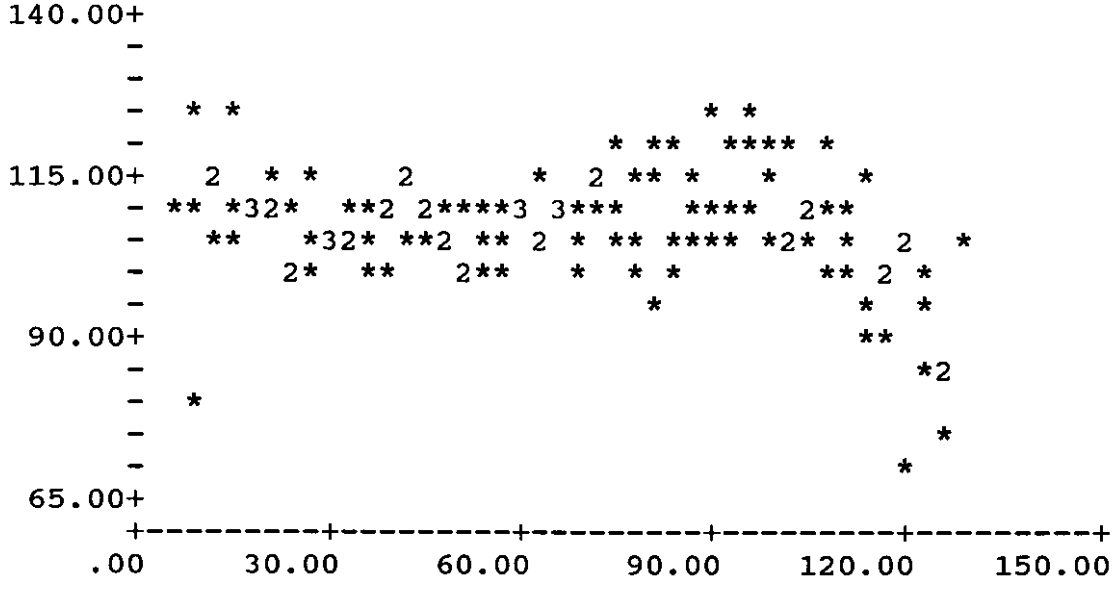


Figure G.2

Plot of the Whole Time Series (122 Data Points)

```
     -1.0  -.8  -.6  -.4  -.2   .0   .2   .4   .6   .8  1.0
        +----+----+----+----+----+----+----+----+----+----+
1    -.005                      X
2    -.003                      X
3     .076                      XXX
4     .167                      XXXXX
5     .164                      XXXXX
6     .019                      X
7    -.001                      X
8     .047                      XX
9    -.015                      X
10   -.033                     XX
11    .170                      XXXXX
12   -.112                     XXXX
13    .024                     XX
14    .093                     XXX
15    .021                     XX
16    .002                     X
17   -.041                     XX
18    .050                     XX
19    .061                     XXX
20   -.033                     XX
```

Figure G.3

Autocorrelation Function of the First 102 Data Points

The autocorrelation function indicates that the time series model is a random model.

The adjusted Box-Pierce test can be calculated as

$$Q_{ad} = N(N+2)\sum_{j=1}^{20}\frac{R_j}{(N-j)} = 14.18$$

where $N = 102$

R_j = autocorrelation coefficient at lag j.

Because the degree of freedom of this adjusted Box-Pierce test is even less than 19, the obtained model is satisfactory, that is, the time series is a random time series. Therefore, the forecasting model should be

$$\hat{V}(t) = MV + a_t$$

where $\hat{V}(t)$ = predicted volume at time t,

MV = the mean of the first 102 data points,

a_t = a disturbance at time t, the expected mean of a_t equals 0.

The obtained forecasting model indicates that this model can only forecast the mean of the time series.

Because the mean of the first 102 data points is 109.36, the three criteria are

$$(1)\quad E_{me} = \sum_{t=103}^{122} \left| \frac{V_d(t) - \hat{V}(t)}{V_d(t)} \right| \div 20 = 17\%$$

$$(2)\quad E_{sr} = \sum_{t=103}^{122} \sqrt{\left| \frac{V_d(t) - \hat{V}(t)}{V_d(t)} \right|} \div 20 = 0.38$$

$$(3)\quad E_{\max} = \operatorname{Max}_{t=103}^{122} \left| \frac{V_d(t) - \hat{V}(t)}{V_d(t)} \right| = 58.6\%$$

where E_{me}, E_{sr}, and $E_{\max}$ have be defined in Chapter Five, $V_d(t)$ is the actual volume at time t, and $\hat{V}(t)$ is the predicted volume at the same time t.

A comparison of the forecast volumes and the actual volumes can be found in Figure G.4.

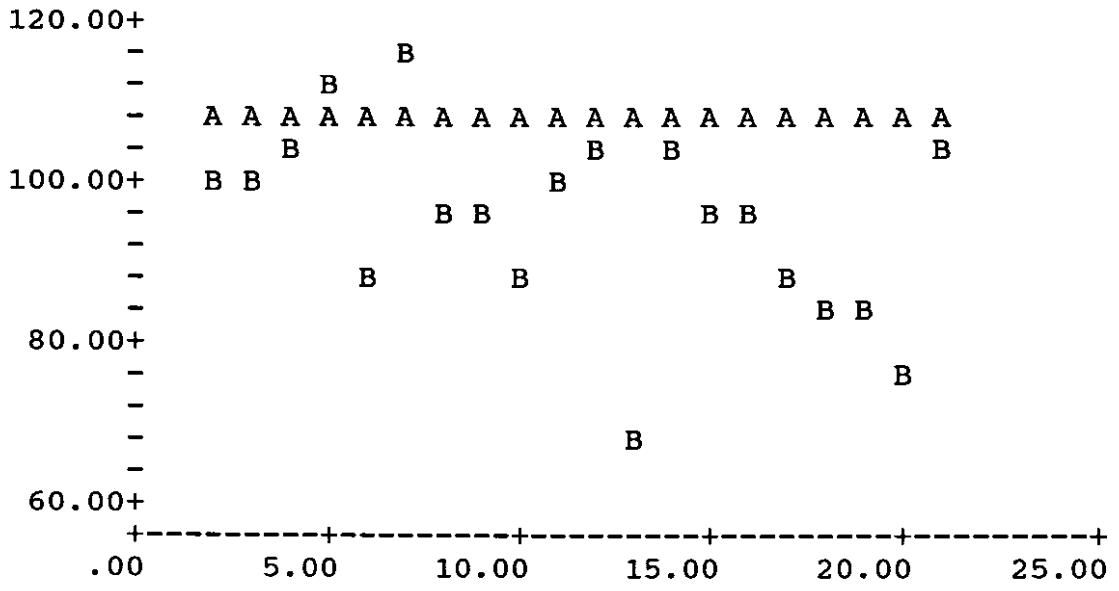


Figure G.4

Comparison of the Forecast and Actual Volumes

In the above figure, A stands for forecast volumes, B stands for actual volumes.

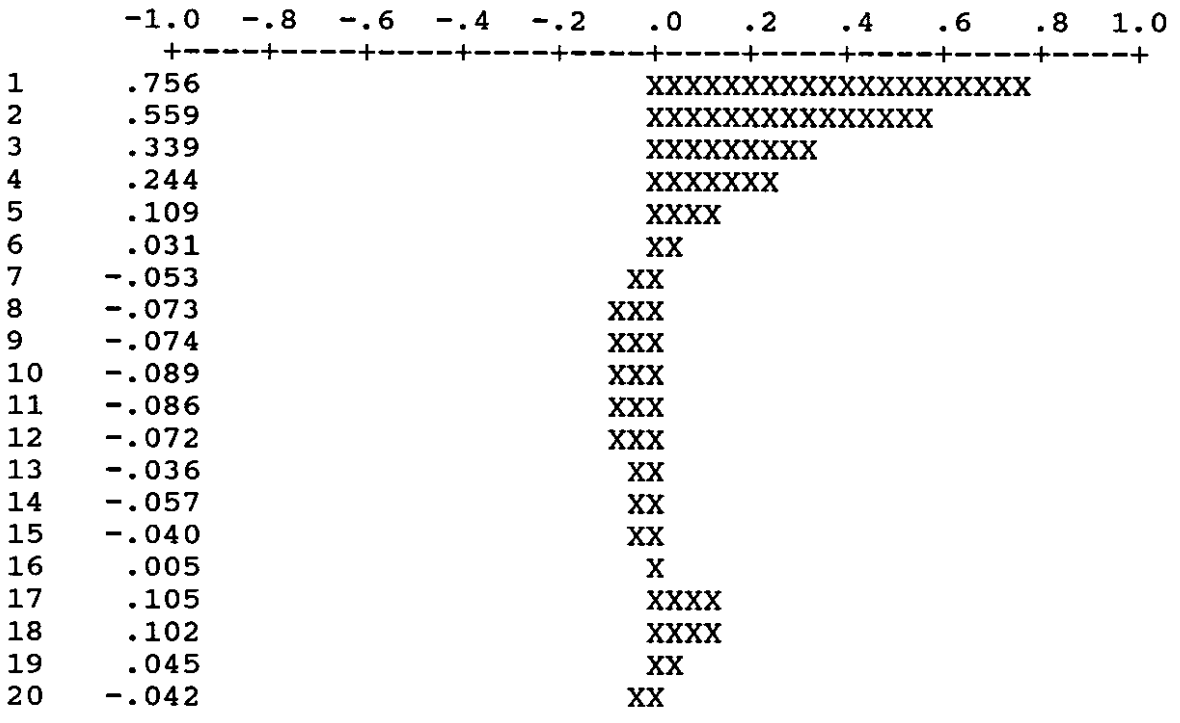


Figure G.5

Autocorrelation of Occupancy Time Series
(First 102 Data Points)

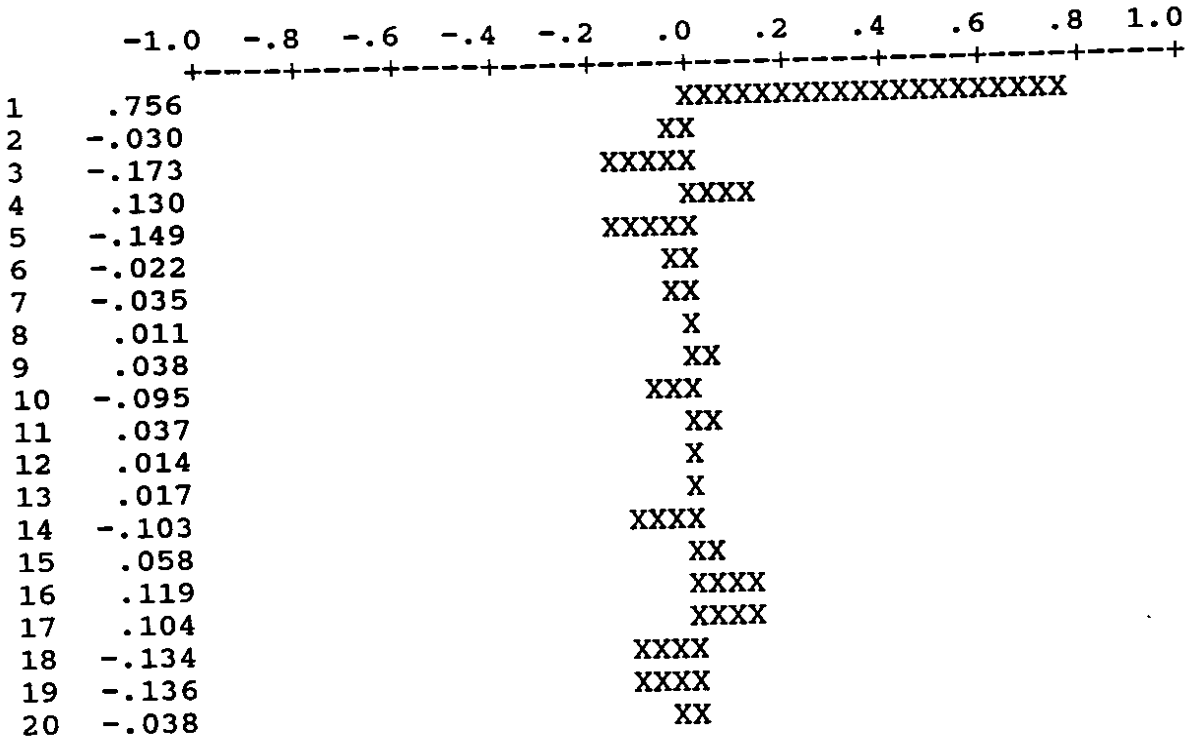


Figure G.6
Partial Autocorrelation of Occupancy Time Series
(First 102 Data Points)

The PACF plot suggests that AR(1) should be trieE. The obtained model is

$$\hat{O}(t) = A_0 + A_1 O(t-1)$$

where:

$A_0 = 3.1$, t-ratio = 13.66

$A_1 = 0.733$, t-ratio = 11.9.

Obviously, the t-ratios for these two parameters are significant within the 95 percent confidence interval.

The calculated adjusted Box-Pierce test of the residuals is Q_{ad}=20.25. When the degree of freedom is 18, the tabulated chi-square of the 95 percent confidence interval is 28.87, which is greater than Q_{ad}. Therefore, the obtained model is satifactory.

The comparison of forecast occupancies and actual occupancies can be seen in Table G.1.

Table G.1
Comparison of Predicted And Actual Occupancies

Actual Occupancy	Predicted Occupancy	Differences
10	10.1	-0.1
11	10.8	0.2
10	11.6	-1.6
10	10.8	-0.8
12	10.8	1.2
9	12.4	-3.4
7	10.1	-3.1
8	8.5	-0.5
10	9.3	0.7
8	10.8	-2.8
8	9.3	-1.3
9	9.3	-0.3
8	10.1	-2.1
10	9.3	0.7
6	10.8	-4.8
8	7.7	0.3
8	9.3	-1.3
8	9.3	-1.3
9	9.3	-0.3
9	10.1	-1.1

The three criteria for this model can be calculated using the results in the above table:

$$(1)\quad E_{me} = \sum_{t=103}^{122}\left|\frac{O(t)-\hat{O}(t)}{O(t)}\right| \div 20 = 17.46\%$$

$$(2)\quad E_{sr} = \sum_{t=103}^{122}\sqrt{\left|\frac{O(t)-\hat{O}(t)}{O(t)}\right|} \div 20 = 0.36$$

$$(3)\quad E_{\max} = \operatorname{Max}_{t=103}^{122}\left|\frac{O(t)-\hat{O}(t)}{O(t)}\right| = 80.5\%$$

where E_{me}, E_{sr}, and $E_{\max}$ are defined in Chapter Five, $O(t)$ is the actual occupancy at time t, and $\hat{O}(t)$ is the predicted occupancy at time t.

Appendix H

Calculation For Model Five

The calculation of the phase delay between the time series at NE 185th St. and the time series at NE 162nd St. is given in the following table:

Table H.1
Calculation of Phase Delay One

Frequency f	Period	Phase	Coherence	Phase Delay Phase/$2\pi f$
0.0000		0.000	0.458	
0.0098	102.0	0.042	0.444	0.68
0.0196	51.00	0.083	0.414	0.67
0.0294	34.00	0.152	0.367	0.82
0.0392	25.50	0.225	0.317	0.91
0.0490	20.40	0.284	0.260	0.92
0.0588	17.00	0.310	0.170	0.84
0.0686	14.57	0.249	0.069	0.58
0.0784	12.75	0.085	0.001	0.17
0.0882	11.33	-3.075	0.045	*
0.0980	10.20	3.045	0.133	*
0.1078	9.273	2.896	0.230	4.3
0.1176	8.500	2.695	0.285	3.6
0.1275	7.846	2.329	0.255	2.9
0.1373	7.286	1.826	0.228	2.1
0.1471	6.800	1.394	0.299	1.5
0.1569	6.375	1.180	0.383	1.2
0.1667	6.000	1.110	0.396	1.06
0.1765	5.667	1.111	0.322	1.0
0.1863	5.368	1.143	0.212	0.98
1.1961	5.100	1.201	0.096	*
0.2059	4.857	1.408	0.015	*
0.2157	4.636	-2.439	0.005	*
0.2255	4.435	-2.193	0.056	*
0.2353	4.250	-2.254	0.137	*
0.2451	4.080	-2.437	0.195	*
0.2549	3.923	-2.734	0.245	2.4
0.2647	3.778	-3.046	0.309	1.95
0.2745	3.643	2.977	0.383	1.73
0.2843	3.517	2.784	0.461	1.56
0.2941	3.400	2.674	0.490	1.45
0.3039	3.290	2.658	0.446	1.39
0.3137	3.188	2.851	0.326	1.45
0.3235	3.091	-2.995	0.277	1.62
0.3333	3.000	-2.686	0.309	1.72
0.3431	2.914	-2.626	0.319	1.7
0.3529	2.833	-2.690	0.303	1.92

0.3627	2.757	-2.748	0.274	1.55
0.3725	2.684	-2.759	0.230	1.51
0.3824	2.615	-2.644	0.180	*
0.3922	2.550	-2.340	0.145	*
0.4020	2.488	-1.797	0.115	*
0.4118	2.429	-1.146	0.134	*
0.4216	2.372	-0.721	0.157	*
0.4314	2.318	-0.374	0.143	*
0.4412	2.267	-0.086	0.121	*
0.4510	2.217	0.122	0.092	*
0.4608	2.170	0.315	0.060	*
0.4706	2.125	0.643	0.034	*
0.4804	2.082	1.132	0.016	*
0.4902	2.040	1.568	0.005	*
0.5000	2.000	3.142	0.000	*

Because coherence is similar to correlation (33), if it is very small, say 0.2, the two time series are not correlated. In the above table, if at a certain FREQUENCY the coherence is less than 0.2 and the corresponding phase delay is not be calculated, this is indicated by an asterisk (*).

From the above table we can see that the phase delay is between 0.17 and 4.3 minutes; that is, the time series at the NE 185th St. section precedes the time series at the NE 162nd St. section by 0.17 to 4.3 minutes. Most phase delays are around 1 minute (with a few exceptions), so we will use $V_{up}(t-1)$ and $V_{up}(t-2)$ to forecast the downstream volume $V_d(t)$.

The calculation of phase delay between the time series at the NE 175th St. on-ramp and the time series at the NE 162nd St. section is shown in table H.2. Phase and coherence in table H.2 were calculated by using BMDP software on MAX system (30).

Table H.2
Calculation of Phase Delay Two

Frequency f	Period	Phase	Coherence	Phase Delay Phase/2pf
0.0000		0.000	0.652	
0.0098	102.0	0.027	0.613	0.44
0.0196	51.00	0.065	0.504	0.53
0.0294	34.00	0.147	0.349	0.8
0.0392	25.50	0.330	0.201	1.34
0.0490	20.40	0.664	0.123	*
0.0588	17.00	0.959	0.084	*
0.0686	14.57	0.749	0.054	*
0.0784	12.75	0.361	0.037	*
0.0882	11.33	0.147	0.033	*
0.0980	10.20	0.233	0.029	*
0.1078	9.273	0.395	0.021	*
0.1176	8.500	0.49	0.011	*
0.1275	7.846	1.319	0.017	*
0.1373	7.286	1.074	0.035	*
0.1471	6.800	1.821	0.073	*
0.1569	6.375	0.607	0.113	*
0.1667	6.000	0.489	0.160	*
0.1765	5.667	0.458	0.197	*
0.1863	5.368	0.525	0.204	0.45
1.1961	5.100	0.769	0.183	*
0.2059	4.857	1.147	0.200	0.89
0.2157	4.636	1.464	0.273	1.08
0.2255	4.435	1.619	0.361	1.14
0.2353	4.250	1.694	0.445	1.15
0.2451	4.080	1.670	0.523	1.09
0.2549	3.923	1.609	0.581	1
0.2647	3.778	1.547	0.607	0.93
0.2745	3.643	1.493	0.584	0.87
0.2843	3.517	1.460	0.528	0.82

From the above table we can see that the phase delay is from -0.27 to 1.34 minutes, which indicates that the time series at the on-ramp at NE 175th St. will precede the time series at the NE 162nd St. (downstream) section by -0.27 to 1.34 minutes. Because -0.27 is near zero, the time series at the on-ramp at NE 175th St. precedes the time series at the NE 162nd St. section.

The calculated phase delays show that the delays are between -0.3 and 1.34 minutes; obviously, 0 and 1 are the integrals that are between the delay interval and also near the extreme values (-0.3 and 1.34). However, volume at lag 0 (simultaneous volume) cannot be used to forecast the downstream volumes, so $V_{on}(t-1)$, which is one lagged on-

ramp volume, should be used together with the upstream volumes to forecast the downstream volumes.

Therefore, the forecasting model is

$$\hat{V}(t) = b_1(1)V_{up}(t-1) + b_1(2)V_{up}(t-2) + b_2(1)V_{on}(t-1) \qquad \text{(H.1)}$$

where $b_1(1)$, $b_1(2)$, and $b_2(1)$ are coefficients.

If we use the ordinary least squares method to obtain the above coefficients, then we have

$b_1(1) = 0.42$ t-ratio = 5.72

$b_1(2) = 0.6$ t-ratio = 7.99

$b_2(1) = 0.25$ t-ratio = 0.77.

$b_1(1)$ and $b_1(2)$ are significant from 0, but $b_2(1)$ is not significant from 0. The Box-Pierce test of the residuals is Q_{ad}=27.8 when the degree of freedom is 20; compared with 31.41, the tabulated chi-square value, this Q_{ad} is within the 95 percent confidence interval, and therefore, the residuals are white noise.

The t-ratio for $b_2(1)$ is not significant from 0, so if we drop this term in equation H.1, we have

$b_1(1) = 0.43$ t-ratio = 5.84

$b_1(2) = 0.61$ t-ratio = 8.3.

The forecasting results of model H.1 can be found in table H.3.

Table H.3
Forecasting Results by OLS Method

Actual Volume	Predicted Volume	Differences
99	105.0	-6.0
102	101.8	0.2
103	103.8	-0.8
111	106.1	4.9
88	100.5	-12.5
117	104.4	12.6
97	108.1	-11.1
98	93.6	4.4
88	84.0	4.0
100	94.2	5.8
104	98.0	6.0
69	87.9	-18.9
104	92.6	11.4
96	92.6	3.4
98	89.5	8.5
87	91.3	-4.3
85	75.6	9.4
85	85.4	-0.4
77	88.7	-11.7
104	94.4	9.6

The three criteria of the above results are

$$(1)\quad E_{me} = \sum_{t=103}^{122}\left|\frac{V_d(t)-\hat{V}(t)}{V_d(t)}\right| \div 20 = 8\%$$

$$(2)\quad E_{sr} = \sum_{t=103}^{122}\sqrt{\left|\frac{V_d(t)-\hat{V}(t)}{V_d(t)}\right|} \div 20 = 0.26$$

$$(3)\quad E_{\max} = \operatorname{Max}_{t=103}^{122}\left|\frac{V_d(t)-\hat{V}(t)}{V_d(t)}\right| = 27.4\%$$

If the coefficients change (refer to Appendix B for the formula), then we have the following forecasting results.

Table H.4
Forecasting Results by Recursive Method

Actual Volume	Predicted Volume	Differences
99	105.0	-6.0
102	101.7	0.3
103	103.9	-0.9
111	105.9	5.1
88	100.5	-12.5
117	104.3	12.7
97	108.4	-11.4
98	92.8	5.2
88	83.8	4.2
100	95.0	5.0
104	97.5	6.5
69	88.2	-19.2
104	93.0	11.0
96	91.6	4.4
98	90.6	7.4
87	90.0	-3.0
85	76.8	8.2
85	85.8	-0.8
77	89.3	-12.3
104	94.6	9.4

The three criteria of the above results are

$$(1)\quad E_{me} = \sum_{t=103}^{122} \left| \frac{V_d(t) - \hat{V}(t)}{V_d(t)} \right| \div 20 = 8\%$$

$$(2)\quad E_{sr} = \sum_{t=103}^{122} \sqrt{\left| \frac{V_d(t) - \hat{V}(t)}{V_d(t)} \right|} \div 20 = 0.26$$

$$(3)\quad E_{\max} = \operatorname{Max}_{t=103}^{122} \left| \frac{V_d(t) - \hat{V}(t)}{V_d(t)} \right| = 27.8\%$$

The three varying coefficients are included in the following table.

Table H.5
Three Varying Coefficients

Last Point	$b_1(1)$	$b_1(2)$	$b_2(1)$
102	0.42	0.60	0.25
103	0.43	0.59	0.19
104	0.46	0.57	0.14
105	0.46	0.57	0.17
106	0.46	0.57	0.17
107	0.50	0.53	0.15
108	0.48	0.54	0.32
109	0.49	0.53	0.30
110	0.52	0.52	0.11
111	0.51	0.52	0.12
112	0.52	0.52	0.16
113	0.53	0.51	0.20
114	0.53	0.51	0.09
115	0.56	0.49	0.05
116	0.54	0.50	0.10
117	0.54	0.50	0.13
118	0.54	0.50	0.12
119	0.56	0.48	0.15
120	0.56	0.48	0.15
121	0.55	0.49	0.12

From the above table, we can see that $b_1(1)$, $b_1(2)$, and $b_2(1)$ vary when more data points are used to obtain these coefficients.