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

Final Report

TNW93-11.5

SHORT-TERM FORECASTS OF FREEWAY TRAFFIC VOLUMES AND LANE OCCUPANCIES, PHASE 2

VOLUME V

by

Nancy L. Nihan
Professor and Principal Investigator

Kenneth L. Knutson
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

TECHNICAL REPORT STANDARD TITLE PAGE

1. REPORT NO.	2. GOVERNMENT ACCESSION NO.	3. RECIPIENT'S CATALOG NO.
WA-RD 288.5		

4. TITLE AND SUBTITLE
SHORT-TERM FORECASTS OF FREEWAY TRAFFIC VOLUMES AND LANE OCCUPANCIES, PHASE 2 — VOLUME V

5. REPORT DATE
June 1993

6. PERFORMING ORGANIZATION CODE

7. AUTHOR(S)
Nancy L. Nihan and Kenneth L. Knutson

8. PERFORMING ORGANIZATION REPORT NO.

9. PERFORMING ORGANIZATION NAME AND ADDRESS
Washington State Transportation Center (TRAC)
University of Washington, JD-10
University District Building; 1107 NE 45th Street, Suite 535
Seattle, Washington 98105-4631

10. WORK UNIT NO.

11. CONTRACT OR GRANT NO.
GC8719, Task 15

12. SPONSORING AGENCY NAME AND ADDRESS
Washington State Department of Transportation
Transportation Building, MS 7370
Olympia, Washington 98504-7370

13. TYPE OF REPORT AND PERIOD COVERED
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
ramp control, freeway management, traffic flow forecasts, forecast models, ramp metering

18. DISTRIBUTION STATEMENT
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	68	

DISCLAIMER

The contents of this report reflect the views of the author(s), who is responsible for the facts and accuracy of the data presented herein. This document is disseminated through the Transportation Northwest (TransNow) Regional Center under the sponsorship of the U.S. Department of Transportation UTC Grant Program and through the Washington State Department of Transportation. The U.S. Government assumes no liability for the contents or use thereof. Sponsorship for the local match portion of this research project was provided by the Washington State Department of Transportation. The contents do not necessarily reflect the views or policies of the U.S. Department of Transportation or Washington State Department of Transportation. This report does not constitute a standard, specification, or regulation.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Current levels of freeway congestion create delay, inefficient fuel consumption, and varying degrees of frustration for drivers. Adding lanes to the roadway to provide adequate service is neither financially feasible nor aesthetically desirable. Until modes of transportation other than those that use some type of a roadway exist for the public, efficient use of current highway networks will be necessary.

Ramp metering, a proven means of increasing highway efficiency, has been employed by the Washington State Department of Transportation (WSDOT) on Interstate 5 (I-5) in the Seattle area for several years. Ramp metering is one of several freeway management techniques that WSDOT employs in the Seattle-Bellevue metropolitan region. Other system technologies include induction loops embedded in the pavement to count vehicles and lane occupancies, closed-circuit television cameras for incident detection, variable message signs, and highway advisory radio. The area of freeway that is monitored includes Interstate 5 from milepost 153.5 to 186.4, Interstate 90 from milepost 1.0 to milepost 14.6, all of Interstate 405, and State Route 520 from milepost 0 to the I-405 interchange. (1)

The metering strategy until recently has been to respond to traffic, which, while helping to alleviate some operational problems on the freeway, does not prevent congestion from occurring.

A current algorithm developed by the WSDOT's Traffic Systems Management Center (TSMC) and used on one section of I-5 incorporates data from the ends of each section to forecast congestion. This process calculates "storage rates," (the number of vehicles entering less the number of vehicles leaving a section) and comparing those rates to the lane occupancy at the desired station.

When the storage rate is positive, indicating more vehicles entering a section than leaving it, and the lane occupancy exceeds 18 percent, the central computer determines that forced flow conditions have occurred. The computer then sends a message to upstream controllers in the field to reduce the metering rate by a predetermined amount until some minimum storage rate has been reached or until the conditions have improved. **(1)**

While this system is not perfect, it performs well enough to use until finer adjustments can be made. Because such a process is quite data intensive, problems can occur if data are inaccurate or missing. But, missing data are a problem in all types of transportation research, whether it be in freeway operations, as at the TSMC, in arterial control, transit facilities, or in studies encouraging pedestrian use.

The existing computer system currently has an algorithm that makes up for missing data, but this is ineffective. **As** the algorithm currently works, when a loop is not working correctly, the loop can be turned off by an operator or the computer system. When this occurs, the computer calculates the average number of vehicles in each of the remaining open lanes and adds that amount to the volumes in those lanes. For example, suppose the section of freeway in question has four lanes, one of which was a failed loop detector, and 90 vehicles are detected in those three lanes during one minute. The computer then adds 30 vehicles to the total of 90, for an estimate of 120 vehicles for that station during that particular minute. The number of vehicles traversing that data station may actually be 120, assuming a linear distribution of traffic across all lanes of a freeway is not always practical because freeways frequently include an HOV lane. The volumes in the HOV lane are usually much lower than in the general purpose lanes.

This research is part of a larger project that attempted to forecast freeway and ramp volumes, and lane occupancies for real-time use in ramp metering applications, as well as forecast data collection efforts as noted above.

This paper will look at previous efforts at developing traffic volume forecasting methods, both in real-time and off-line situations. Several models that have been developed will be discussed to provide a foundation for this and future studies. (2, 3, 4, 5, 6, 7, 8, 9) A short, noncomprehensive review of these methods is given in Appendix A which discusses state-of-the-art efforts in this subject area. The majority of the research effort has been focused on checking the validity of work done by Zhu at the University of Washington in 1990 and furthering promising aspects of that work. (9) A discussion of our reasoning can be found in Chapter Two, Research Design. Chapter Three discusses the results of this research. The final chapter contains conclusions'and recommendations.

APPENDIX A

STATE OF THE ART

The goal of this appendix is to review and discuss traffic forecasting methods developed in the past that may currently be used in some locations. These methods are not presented in any particular order for the author does not want to express a preference for any one model. The final forecasting method, however, is placed in that position because of its significance to this research.

UTCS

The Urban Traffic Control System (UTCS) has two forecasting methods. Both methods forecast demand, but in different ways: Stephanedes, et al. **(7)** described UTCS-2 as a second-generation model and UTCS-3 as a third-generation model.

UTCS-2, as the name suggests, is the older and less sophisticated model of the two. It requires substantial amounts of historical data for forecasting (though some researchers suggest only the ten most recent days will suffice), with small departures from history being taken up by current data measurements. (4) It makes forecasts in real time on the basis of these historical data and measurements taken from one location. That is, forecasts for a particular station X are based on history and measurements at station X. Predictions generally do not go beyond the next time slice, which is usually between 5 and 15 minutes. (7)

Since the main focus of this research is not on UTCS, the set of equations used has been left out of this paper. However, the prediction equation is as follows: (7)

$$V_t = m_t + \gamma(m_{t-1} - \gamma f_{t-1}) + (1-\alpha)(\Sigma_1) + \gamma(1-\alpha)(\Sigma_2)$$

where

m_t = historical volume at time t

γ = smoothing coefficient, obtained by a complex equation not shown here (see (7) for equations)

f_t = measured volume at time t

α = constant computed off-line from representative volume data for the station location

$$\Sigma_1 = \sum_{s=0}^{t-1} \alpha^s (f_{t-s-1} - m_{t-s-1})$$

$$\Sigma_2 = \sum_{s=0}^{t-2} \alpha^s (f_{t-s-2} - m_{t-s-2})$$

Additionally, the historical volume at time t is determined as the sum of cyclic volumes plus a constant obtained for an approximation of patterns for that particular location.

UTCS-3 is a third-generation model, which is more modern and simpler than UTCS-2. UTCS-3 also has the capability of forecasting more than one time period in advance. Unfortunately, according to the literature reviews of this model, it appears to be limited to only two time intervals into the future. This method also makes predictions without the use of historical data; no historical pattern is required for forecasts because only current measurements are used for forecasts. Like its predecessor, UTCS-3 only uses data obtained from one location for its predictions.

The full set of equations for this forecasting method is not included in this research; however, the demand prediction equation is as follows: (7)

$$V_{t+j} = \gamma_j f_t + (1-\gamma_j)[\mu_0 \alpha^t + (1-\alpha)\Sigma_3]$$

where

V_{t+j} = forecasted volume for time t+j at time t

γ_j = constant computed off-line from representative volume data

f_t = measured volume at time t

μ_t = exponentially smoothed volume measurement

α = smoothing coefficient

$$\Sigma_3 = \sum_{s=0}^{t-1} \alpha^s f_{t-s-1}$$

One problem with both of these methods is that forecasts are made partially or entirely on the basis of data collected at the location in question. For forecasting congestion in real time, this is a viable way to proceed; however, if the objective is to replace missing data, this method fails. It appears to break down as soon as the forecasting horizon has passed. One way to make up for this would be to use forecasted volume in the next volume prediction. This may work for a short time, but since there is no other source of information about the freeway system, there is no redundancy built into the forecasts, and this method would tend to produce large forecasting errors as time progressed.

The reason for developing another forecasting model was to overcome the shortcomings of UTCS-2. Major problems occur with UTCS-2 because of its high reliance on historical data. It is well-known that traffic volume can vary considerably, depending on several factors, such as weather, special events, etc. As discussed above, UTCS-2 does not respond to these circumstances. Therefore, the model cannot be used with different systems and locations. Additionally, the database for the historical data must be updated continually.

In theory, UTCS-3 should be an improved model. Some tests by Kreer (10) suggest that UTCS-3 is not as great an improvement over UTCS-2 as was first thought. Those findings indicate that the older model performed better, with both a lower mean absolute error and mean square error. (See Chapter Three for details.) While the effect of time lags exists in both models, the effect is greater in the third-generation model, which requires at least two time intervals to recover should a detector briefly malfunction.

Spectral Analysis

One of several formal methods of time series forecasting, the spectral analysis method was used to forecast traffic volumes at the Liverpool (U.K.) Mersey Queensway tunnel, with relatively successful results. (6) This is a math intensive method of forecasting that is difficult to understand for most laypersons. Because of these drawbacks, this method is seldom used in practice, and the forecasting equations are not presented here. The reader is directed to (6) for further reading.

Researchers took traffic to be the result of several cumulative effects, including commuters, recreational, and commercial traffic. Assuming this was true, then the basic traffic patterns would most likely only slightly vary. Each day was considered to be an ensemble of a time series. The researchers also assumed that the series was nonstationary. This method also has the property of allowing for a recursive algorithm to update constants as additional data become available.

Researchers used data in six-minute time slices for two hours in both the morning and evening commutes for 43 days, excluding weekends. While it may have seemed wise to remove (or compensate) for trends or cyclic behavior and data, this created only marginal improvements in forecasts. Researchers also found that the maximum prediction errors were in the range of 8 to 11 percent, with higher errors obtained when data compiled from averaged previous data were used in the prediction, which resulted in a 12 to 19 percent error.

The researchers felt that the results were acceptable, and indicated that most of the forecasting errors were due to faulty equipment. They believed that error rates could drop to 5 percent if improved data collection methods were available. (6)

Box-Jenkins Model

The Box-Jenkins method has been widely and successfully used for several types of forecasts, from business revenue projections to traffic forecasts. While somewhat difficult to use, this method tends to be more intuitive than spectral analysis, and for this reason is used more often.

In general, the forecasts are a function of some type of weighted moving average along with a weighted autoregressive element (though one, both, or neither may be included in the forecasting equation). The operator must determine the type and order of the process in question before proceeding with forecasts.

Typically, an autoregressive process of order p will have an autocorrelation function (ACF) that tails off, and a partial autocorrelation (PACF) that cuts off when less than two standard errors occur after lag p. On the other hand, in moving average process of order **q,** the PACF will tail off and the ACF will cut off after lag q. If both ACF and PACF tail off, the process is mixed. This indicates that some sort of transformation may be made, or that the forecasting equation includes both moving average and autoregressive terms. (3)

The general form of the Box-Jenkins technique is the ARIMA model, AutoRegressive Integrated Moving Average. If seasonality is experienced, the series needs to be modified in order to use this method. The modification may involve differencing the series (using the change in volume instead of the volume itself), or using some other transformation, such as a logarithm or square root.

Traffic volumes are often considered seasonal since there is a regular, cyclic pattern to traffic on a daily, weekly, and monthly basis. The traffic in this research is not seasonal since it occurs over a few hours and was collected in one-minute time slices.

The general form of the ARIMA model is as follows: (3)

$$\Phi_p(B)(1-B)^d(X_t-\mu) = \Theta_q(B)a_t$$

where

p, d, q = the order of ARIMA (nonnegative integers)

μ = mean of the series

$\Phi_p(B)$ = autoregressive portion of order p, equivalent to $1 - \Phi_1 B \ldots - \Phi_p B^p$, $|\Phi| < 1$

$\Theta_q(B)$ = moving average portion of order q, equivalent to $1 - \Theta_1 B \ldots - \Theta_q B^q$, $|\Theta| < 1$

a_t = random disturbances (assumed normally distributed)

B = backshift operator, such that $BX_t = X_{t-1}$

Thus, the above model is described as ARIMA of order (p,d,q). This method is unique in that it does not have a form that the user applies to the data. Instead, a data analysis must initially be done to determine the appropriate order for the model. This step, identification, is the first of three steps (estimation and checking of fit are the other two). This is often described as "letting the data speak for itself." (4) The researcher is then free to make observations and suggestions to the computer as it manipulates the model.

If the data prove to be nonstationary in the identification phase, then alteration to produce stationarity must occur. This takes place through differencing with log, or power transformations. Differencing is used if the series is nonstationary in the mean, while the other transformations are used if the series is nonstationary in the variance.

When the series is stationary, ACFs and PACFs are found to determine the order of the series, as described previously. The next step in the ARIMA process is to determine preliminary coefficients for the model via least squares. The model is then checked for fit. An effective model will have uncorrelated residuals.

Box and Pierce developed a test to determine if residuals are correlated. They found that the variable Q was approximately chi-square distributed with K-p-q degrees of freedom. The variable is defined as

$$Q = n \sum_{i=1}^{K} [r_i^2(a)]$$

where

n = number of observations

K = number of lags checked

$r_i(a)$ = residual autocorrelation for lag i

Ahmed and Cook used the Box-Jenkins technique in a study of freeway traffic data in three metropolitan areas. (2) Like the data used in our research, most of the data Ahmed and Cook used were from one-minute or shorter time slices.

They continued research done by Der, and took the data sets best described by an ARIMA (0, 1, 3) process. Der initially suggested a (1, 0, 1) process, but it has been determined that this suggests a stationarity for traffic that may not exist in all cases.

Ahmed and Cook performed further tests, comparing the ARIMA method with several other models. They used measures of effectiveness of Mean Absolute Error (MAE), and Mean Square Error (MSE), both of which are described in Chapter Three. These values for the ARIMA method were normalized to be 1, and the errors for the other models were taken as

$$\frac{\text{MOE (model of choice)}}{\text{MOE (ARIMA)}}$$

The researchers did not find any methods that minimized either MAE or MSE better than the ARIMA (0, 1, 3) method. Thus, it was felt that this model best described the traffic process.

The researchers concluded that updating the moving average coefficients might improve the model. They stressed that this would not be necessary in all circumstances, citing work by Trigg and Leach (11), which showed that consistently changing a smoothing constant resulted in larger forecasting errors than using simple exponential smoothing models.

When considering computer requirements for updating such coefficients, one must be careful in deciding when to update parameters--one possible point is when traffic patterns change (between peaks).

Another researcher praised the computational convenience of the ARIMA method as a prime consideration for on-line computer applications. (5) Eldor found that as time slices increased in size, the order of differencing increased as well. Logically this is expected to occur, since in larger time slices there is a larger variability in the volumes experienced. The researcher also suggested that, from a demand perspective, traffic responsive control systems are not necessary for conditions where there are fairly regular traffic patterns, possibly implying that such research for the Seattle area is unnecessary. He also noted that incidents alone do require immediate response in real time, so some limited control should be employed. Eldor also recommended against using systems where a cushion is employed to guard against underestimation as this cushion would tend to produce an overly conservative result. It would then be difficult, if not impossible, to know when underestimation or overestimation occurs. (4)

Pattern Recognition

The pattern recognition method is currently used on part of Interstate 5 in the Seattle area. While this method does not accurately forecast traffic volumes, it is effective in forecasting traffic congestion. The goal is to detect congestion before it occurs, thus preventing it. The model should be simple enough to allow the central computer to perform routine calculations without delay, but complete enough to allow for accurate forecasting. The system does not look for several patterns, but only one, the simple, classic "bottleneck." This greatly simplifies the system.

The central computer monitors the freeway in one-minute intervals with updates from field controllers. Congestion tends to occur when the number of vehicles entering a section exceeds the number of vehicles leaving the section (positive storage rate), and average lane occupancy across all lanes in the section exceeds 18 percent.

When the computer senses that these conditions exist, it forecasts a bottleneck, and attempts to restrict metering rates for ramps entering the section by the storage rate amount. The goal was not only to forecast the congested conditions, but also ***not*** to forecast these conditions when they did not occur, as doing so would restrict the ramp volume unnecessarily, creating more congestion rather than reducing it. (1) Though this procedure has good prospects for forecasting congestion in general, its use for determining traffic volume is minimal, since volume is typically stochastic in nature, especially on a minute-by-minute basis.

Least Squares Approach

The Least Squares Approach, which is the final method considered here, has implications for this research. Zhu and Nihan employed this method in their research. The general form of the Zhu/Nihan model is that downstream volume can be expressed as a function of the sum of fraction of upstream volumes, lagged an appropriate amount.

$$V_d = \sum_{p=p1}^{p2} k1(p)Vup(t-p) + \sum_{q=q1}^{q2} k2(q)Von(t-q) + Z(t)$$

where

$p1, p2$ = minimum and maximum travel times from upstream

$q1, q2$ = minimum and maximum travel times from on ramp

$k1$ = portion of vehicles with travel time p

$k2$ = portion of vehicles with travel time q

$Z(t)$ = error

The main challenge is to find the appropriate lags, which would indicate p1, p2, q1, and q2, and to find the appropriate coefficients k at these lags. Zhu and Nihan used spectral analysis and Fourier's transformation of the covariance to determine the lags. (Details of this proof are not given here, but may be found in reference (9).)

The above equation can be expressed in matrix form, then ordinary least squares can be used to obtain the coefficients, and updated as time progresses.

This method was used to forecast both volumes and occupancies on a section of Interstate 5, north of downtown Seattle, and an MAE of 8 percent and an MSE of 0.26 were obtained.

The conclusions suggest that off-line forecasting using only upstream volume might also be appropriate since the entrance ramp affected the volume only slightly (the coefficient was less than unity, and ramp volumes are typically less than 20.) Downstream occupancies were difficult to forecast based on upstream occupancies, thus, a model for occupancies was not feasible for the research here.

CHAPTER 2

RESEARCH DESIGN

Data Collection

To facilitate investigation of forecasting techniques, a study site was selected where data have typically been found to be reliable (see figure 2.1). Reliability refers to stations that have equipment that rarely malfunctions. This stretch of freeway has recurrent peak-hour congestion because of several entrance ramps in a short area. This is typical of what other researchers often discovered in other parts of tlie country. These factors, combined with the fact that the Washington State Department of Transportation has data stations in this area, create an area suitable for analysis.

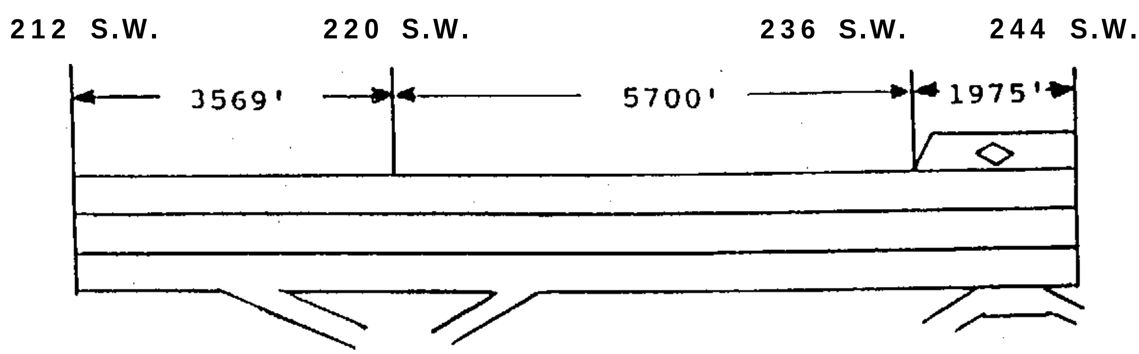

Fig. 2.1 Study Site

The section used to test the old model is 2.13 miles (11,244 ft.) in total length, encompassing 212th St. S.W. through 244th St. S.W., while the section used for the new model is 1.46 miles (7,675 ft.) in total length, encompassing one less station, 220th St. S.W. through 244th St. S.W. Both sections have two entrance ramps: one at (220th St. S.W. and the other at 236th St. S.W.) and one exit ramp. In the final 1,925 ft. of the section, just under the 236th St. overpass, an inside high occupancy vehicle (HOV) lane has been added (the left lane). This segment serves suburban commuters en route to employment centers in the Central Business District and other suburban communities.

Data were obtained electronically by WSDOT loop detectors. A loop, embedded in each lane, recorded both volume and lane occupancy. A group of loops at one location comprises a data station; at which vehicle counts are aggregated providing the data used in this study.

Of the data collected over several days, the data collected on February 11 and April 17, 1991 were selected. Two hours of data were collected in one-minute increments for each day, from 6:15 a.m. to 8:15 a.m., at the following stations: 212th St. S.W., 220th St. S.W., 236th St. S.W., and 244th St. S.W., which are on the mainline of I-5; the entrance ramp at 236th St. S.W.; and the exit ramps at 220th St. S.W. and 244th St. S.W. The weather on the first day was cloudy with light rain and good visibility; the roadway was bare and wet. The second day was partly cloudy with good visibility. (A time-series plot, as well as a listing of the data points, is given for each time series in Appendix B, Study Data and Plots.) The time series for the data set indicates that something of significance happened. When the incident log, maintained at the Traffic Systems Management Center (TSMC) was checked, it was confirmed that an incident of some type (an accident) occurred.

Calibration of Simple (Zhu/Nihan) Model

As is typically done with time series 'statistics, the first several data points were used to generate a forecasting process for the last few points in the data set. In this case, the first 90 available data points were used to obtain an equation to forecast the final 30 points.

The model developed here was originally employed to determine if Zhu and Nihan's methods could be applied to other sections of I-5. A 1.2 mile-long section south of this area was studied in that research.

In keeping with the original goals of this research, it was necessary to find time-lags of upstream stations that significantly affected the data station at 236th St. S.W.

The study area for checking Zhu and Nihan's model included the additional upstream station at 212th St. S.W. and the exit ramp at 220th, which adds 3,643 ft. (0.69 miles) to the study area of the new model described in section 2.1. We initially were reluctant to include this station, since it lengthens the study sections, and we felt it might result in conditions different from those developed by Zhu and Nihan. However, since the general geometrics of the section were the same, the station was kept in the study, and appropriate lags were determined.

To ascertain the best coefficients to use, it was necessary to first find the critical lags affecting downstream traffic volumes. To do this, we estimated travel times through the study section. These were obtained by determining the length of the section and selecting expected vehicle speeds for the time period studied. For example, since the study section is 1.46 miles long and the study period is during the morning commute, it is not unreasonable to expect a vehicle to take

$$\frac{1.46 \text{ mi.}}{30 \text{ mi./h}} = 0.049 \text{ h} = 2.92 \text{ min}$$

to traverse the section. A vehicle traveling 30 mph would take 4.2 minutes to travel from 212th St. S.W. to 244th St. S.W. Determining this assumed speed is supported by the fact that many of the volumes are around 100 veh/min, which expands to 6,000 vph. This is the assumed capacity for a three-lane freeway, such as this.

A wide range of travel times were selected (one to eight minutes for the mainline, corresponding to speeds from 11 to 88 miles per hour). Ramp travel times also ranged from one to eight minutes.

The data were input into a text file using a text editor on a personal computer, then sent through a modem to MAX, a VAX computer at the University of Washington. All statistics operations were done using MINITAB 7, a statistical software package.

The data files were read into MINITAB and were lagged the appropriate amounts. For example, if a column of data were to be 94, 103, 87, 92, 95 . . . (units of vehicles per minute), then lagging that column by one time period would result in MINITAB reading the column as *, 94, 103, 87, 92, 95, . . . where "*" represents missing or unknown data.

It was expected that the results would be of the form

$$\begin{aligned} V_{pred} = {} & a_1V_{212(t-1)} + a_2V_{212(t-2)} + \cdots + a_iV_{212(t-i)} \\ & + b_1V_{220(t-1)} + b_2V_{220(t-2)} + \cdots + b_jV_{220(t-j)} \\ & + c_1V_{on(t-1)} + c_2V_{on(t-2)} + \cdots + c_kV_{on(t-k)} \end{aligned}$$

Where V_{pred} represents the forecasted volumes at 236th; V_{212}, V_{220}, and V_{on} represent one-minute volumes at 212th St. S.W., 220th St. S.W., and the entrance ramp from 220th *St.* S.W. The parameters a, b, and c represent constant coefficients. The letters i, j, and k indicate that the number of lags for each volume need not be equal.

The results of this method can be found in Appendix ***C*** and are discussed in the following chapter.

New *Model* Development

The above method should, in theory, forecast traffic volumes relatively well because what is actually being forecasted is the critical lags. Unless the vehicles upstream exit to 220th St. S.W., the traffic volumes at upstream stations will pass the station for which a forecast is desired.

In cases where an incident, or another form of congestion occurs, it is clear that traffic volumes will tend to decrease due to a loss of capacity and natural driver curiosity about the incident. We know, from shock-wave theory, that such bottlenecks can have major effects on the bottleneck location. both upstream and downstream.

The important types of shock-waves, in this case, are the backward-forming and forward-recovery shock-waves. When an incident occurs, this is especially true. This situation can be represented by a hypothetical density contour map similàr to the one shown in Figure 2.2.

If non-recurrent or recurrent congestion occurs at point A, creating stationary and hackward-forming shock-waves, traffic backs up to some point B, where an incident, supposedly, has occurred. This creates another backward-forming shock wave and a new forward recovery wave. (11)

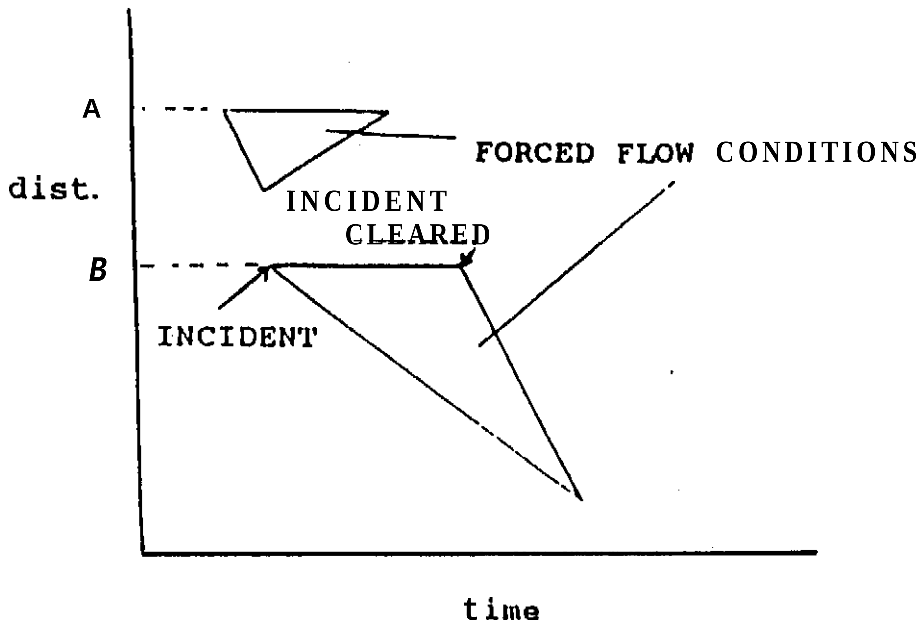

Fig. 2.2 Density Contour Map

In order to compensate for such shock waves and changes in congestion levels, it is desirable to have a model that can account for them in some way. Density contour maps serve as good indicators, but they are not always easy to understand, nor are they easy to program.

Instead, consider the model incorporating the storage rate concept, tlescrihed in Chapter One, that is currently used at the TSMC for ramp metering purposes. In the case where an incident occurs downstream of the section where a storage rate is calculated, the number of vehicles entering the section exceeds the number leaving it, and the storage rate is positive. The downstream constriction eventually will have the effect of reducing volumes entering the section until the section has storage rates approaching zero. At that point, storage rate has no effect.

In the case where reduced capacity occurs upstream of the section where a storage rate is calculated, the number of vehicles entering the section will be less than those leaving it. This, too, will have a negative effect on volumes downstream of the section, as described in the above discussion of bottlenecks downstream of a station.

Consider a generic freeway section shown:

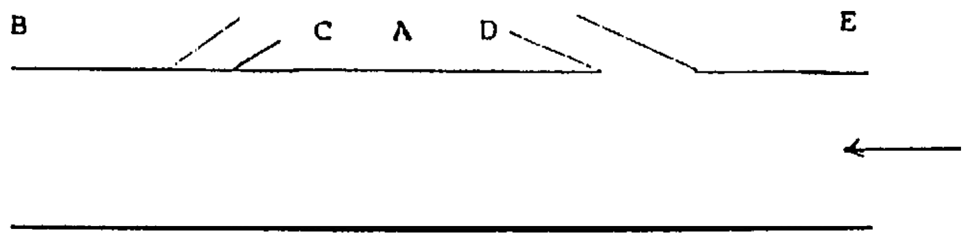

Fig. 2.3 Generic Freeway Segment

If the freeway is cut into pieces, section B-C and section D-E, then traffic volumes at point A can be viewed considering the effect of storage rates alone. When a reduction in flow occurs at B, traffic arriving at a rate q_a will back up because the flow rate departing B is q_d, and therefore, $q_a > q_d$. Suppose the shock wave will continue upstream to some point past D and **A.** Once the shock wave moves past D, the storage rate in section B-D should theoretically be zero, with the arrival flow equaling the departure flow. While the shock wave is proceeding upstream, the flow at **A** will be reduced until the shock wave moves past.

Likewise, if a bottleneck occurs at E, the arrival rate q_a will be less than the departure rate q_d at D. This is best visualized if congestion exists on the entire segment in Figure **2.3.** If a non-recurrent incident occurs at point E, at that point, the flows past E will be significantly lower than those in the rest of the section. If the little flow that does get past E does not put the freeway at, or over capacity, the downstream congestion at points A through D should disappear. This incident will also have the effect of reducing the volume at A, since this is a forward-recovery shock wave, and this only exists when there has been congestion and demand falls below the downstream bottleneck capacity. (12)

Another way to incorporate the effects of local congestion into a model is to use lane occupancy terms. The data given by the TSMC have one-minute lane occupancies to 0.1 percent. These data may not always be as accurate as 0.1 percent, especially in congested situations where vehicles have very small distance headways. This may cause some lack of clarity in the loop detectors. Nonetheless, the data presented have been used. Thus, lane occupancies will tend to indicate density that should relate to volume. Upstream occupancies have been used in addition to upstream storage rates. Downstream occupancies were not used in this model.

In all three of the above situations, a negative effect on volume occurs. For, as a downstream storage rate increases, the effect should eventually be to reduce traffic volume upstream. **As** upstream storage rate increases, traffic volume downstream should eventually decrease, and as upstream lane occupancy increases, downstream volume should decrease.

In this particular model, though the study section does not, in fact, resemble Figure 2.3, similarities do exist. The upstream station, at 220th St. S.W. can be represented by point E, and the downstream station, at 244th St. S.W. can be represented by point B. Point **A** can represent the station where forecasts are desired, the 236th St. S.W. station.

The model will use upstream storage rates calculated by $V_{220} - V_{236}$ and downstream storage rates $V_{236} - V_{244}$. Though the station where forecasts are desired is used in these calculations, this is not a problem. If the station at 236th St. S.W. malfunctions for some reason, the computer would look at the most recent mainline volume for this station and use that volume for the next forecast.

Reasoning for this may seem somewhat circular, but if forecasts have a reasonably small error, error propagation should not be a problem in off-line use. Difficulties could arise in on-line use if the central computer could not locate the last true volumes, or if it took too long to make forecasts for real-time control. If error terms are random and normally distributed, it is logical to assume that aggregating forecasted volumes (i.e., using five-, fifteen- or sixty-minute data) should result in error terms approaching zero.

The model should take the form of

$$\begin{aligned} V_{pred} = \quad & Z + a_1 V_{220(t-1)} + \ldots + a_i V_{220(t-i)} \\ & + b_1 O_{220(t-1)} + \ldots + b_j O_{220(t-j)} \\ & + c_1 SR_{up(t-1)} + \ldots + c_k SR_{up(t-k)} \\ & + d1 SR_{dn(t-1)} + \ldots + d_m SR_{dn(t-m)} \end{aligned}$$

where Z is a constant; a, b, c and d represent constant coefficients; V indicates volume; O indicates lane occupancy; SR_{up} is upstream storage rate; SR_{dn} is downstream storage rate; and i, j, k, and m are indices indicating that the number of coefficients for each variable need not be equal.

Regressions for several lags were performed on MINITAB. Final results are in Appendix D, and a discussion of these results can be found in the next chapter.

Validity Concerns

Weather and incidents can cause these models to break down and become inefficient, and may result in a model that is different from the actual circumstances. For example, if an incident occurs, it may affect the forecasting equation for days when no incident occurs. Likewise, the weather indicated on the day modeled here may not be valid for some predictions. While initially these may seem to be external validity concerns, such considerations can become internal since only one section of freeway is considered in this case.

Defective equipment, such as bad loops and bad stations, can always occur, especially in this situation where the equipment is several years old and has been subjected to weather. The effects of temperature on the equipment must also be considered. The computers at the study site have a small fan to keep them cool during the summer, and minimal heating to keep them warm during the winter. These weather effects and lack of sophisticated environmental control could contribute significantly to data errors, especially in hot weather and high humidity.

The consequence of using these instruments greatly affected this research project because the objective of the research was so closely tied to the quality of data collected. The objective was to develop a model to replace missing or bad data, so it was important to collect data known to be accurate; however, the only way to collect these data was by using the same technology known to need improvement. But good data may not exist for this research. Since the results are only as good as the original data, those data are suspect. Nevertheless, a site for the research had to be chosen, and the reason for selecting this section was that these stations consistently had reliable data, and the microcomputers at these stations rarely broke down.

CHAPTER 3

RESULTS

Introduction

In this chapter, results of the models discussed in Chapter Two are presented. A set of evaluative criteria is established, and these models are placed against criteria in order to judge them. A discussion of the results follows, including how the results were achieved and their implications.

Criteria

Two error measurements established to be effective, (7) along with two introduced here, were used to compare forecasting methods. The mean square error (MSE) penalizes large prediction errors and the mean absolute error (MAE) indicates a typical error for individual forecasts. These two standard means of judging forecasting effectiveness are defined as follows:

$$\text{MSE} = [\Sigma(\text{measured vol} - \text{forecast vol})^2]/N \text{ and}$$

$$\text{MAE} = [\Sigma\,|(\text{measured vol} - \text{forecast vol})|\,]/N$$

where N = number of predictions.

For this research, the above error terms were calculated for the data points 102 through 128 (N = 27). While the reader might expect 30 data points from 91 through 120, data collection methods employed at the TSMC for this particular day resulted in gaps in the data set at the end of each 30-minute block. These gaps, in turn, caused the first few data points to be difficult to use since the forecasting equations use lagged terms (i.e., terms using data from previous time periods). When a missing term was lagged, it resulted in an ineffective forecast. In order to compare the two models isometrically, the first three terms were dropped. Ordinarily, this would not be the case since data collected would be continuous.

Additionally, using percentages, rather than absolute numbers, was preferable. (Absolute numbers are often difficult to evaluate because it is hard to distinguish the difference in effects of a vehicle error of magnitude E on two different actual volumes: the error of 15 vehicles on actual volume of 30 vehicles is different than the error on an actual volume of 115 vehicles.) Percentages were computed as follows:

% error = 100 * (forecast vol/measured vol - 1)

When the forecast volume exceeds the actual volume, the error will be positive, indicating overprediction, and similarly, negative error rates indicate underprediction.

Another error term, Emax% was also used. This was important for developing criteria to determine when an incident or data collection error might occur. This can be used in cases where the predicted volume exceeds some pre-set maximum error. When this happens, one can conclude that some type of malfunction or incident happened.

Note that it is possible for a model to have a lower MAE in percent but a higher MAE in absolute terms. This indicates that such a model may be less sensitive and fairly consistent in its error formation. The other model would tend to have larger errors as traffic volumes decreased. One way to interpret such a discrepancy is to analyze the mean square error term. **A** desirable model should have lower error rates relative to other models in at least two of the three terms. Given the fact that the field equipment cannot collect data better than about a 10 percent error rate, the chosen models should have error rates less than 10 percent.

Percentage of error was also aggregated into five-minute increments here, since most data collection for research needs some sort of aggregation for study. If the one-minute percent error has a normal distribution with mean zero, then the five-minute error should also hover close to zero.

Forecasting Results

As discussed in Chapter Two, both forecasting techniques used the MINITAB statistics software to generate forecasting equations. The general expected format of these equations is given in that discussion. Appropriate lags were determined through numerous regression runs, taking a different combination of variables each run.

In general, runs started by using all variables at all lags, and continued by dropping the least significant variable at the end of each run until an acceptable model was developed. Though use of all variables at all lags was not always the case, it quite often was, as it was the best way to keep track of different variable combinations.

Accuracy will be somewhat limited because the data interval is one minute. Obviously, narrower time slices will tend to result in improved accuracy because of an increase in the amount of information provided. However, when such large amounts of data are used the result may be larger errors because of the data collection. Thus, the one-minute time interval limits model sensitivity.

The simpler model, which used Zhu and Nihan's dependent variable set, was found to have significant variables at 212th St. S.W. (lag 2), 220th St. S.W. (lags 1 and 2), and the entrance ramp at 220th St. S.W. (lags 1 and 3).

The forecasting equation for the simpler model is

$$V_{pred} = a_1 V_{212(t-1)} + b_1 V_{220(t-1)} + b_2 V_{220(t-2)}$$
$$+ c_1 V_{on(t-1)} + c_2 V_{on(t-3)}$$

where MINITAB determined via least squares regression the coefficients and t-ratios shown in Table 3.1.

Table 3.1, Coefficients of Simple (Zhu/Nihan Formulation) Model

coefficient	t-ratio
$a_1 = 0.186$	2.06
$b_1 = 0.281$	2.58
$b_2 = 0.491$	5.15
$c_1 = 0.456$	1.36
$c_2 = 0.598$	1.70

Note that all of the coefficients are less than unity. This should be taken as coefficient values representing percentages of vehicles. For example, in the above variables, 28 percent of the vehicles measured past 220th St. S.W. take approximately one minute to travel to the downstream section, and 49 percent of the vehicles take about two minutes to travel to the downstream station.

This model was found to obtain the following errors:

MAE = 5.18

MAE% = 9.90

MSE = 43.70

In the new (Nihan/Knutson) model, significant variables were found to be a constant, volume at 220th St. S.W. (lags 1 and 2), occupancy at 220th St. S.W. (lag I), upstream storage rate (lags 1 and 3), and downstream storage rate (lag 2). Thus, the new model gives a forecasting equation of

$$V_{pred} = Z + a_1V_{220(t-1)} + a_2V_{220(t-2)} + b_1O_{220(t-1)} + c_1SR_{up(t-1)} + c_2SR_{up(t-3)} + d_2SR_{dn(t-2)}$$

Once again, MINITAB calculated coefficients via least squares methods and obtained the coefficients shown in Table 3.2.

The results in Table 3.2 also serve as coefficients for a variation of the model, in which updated information is used. This was done to test the improved accuracy of forecasts due to new information. Table 3.2 gives coefficients for the updated model as well, since both new models begin forecasting in the same way.

Table 3.2 Coefficients of Nihan/Knutson Model

coefficient	t-ratio
$Z = 41.28$	4.16
$a_1 = 0.472$	4.10
$a2 = 0.226$	2.51
$b_1 = -0.312$	-2.90
$c_1 = -0.311$	-2.95
$c_2 = -0.153$	-1.90
$d_2 = -0.138$	-1.48

Once again, the coefficients are fractions, which should be taken as percentages of vehicles requiring a certain length of time to reach the downstream station. Exceptions to this rule exist in storage rates and lane occupancy. Analysis of other constants indicates that most of the effects on predictions lie with the constant and the upstream mainline volumes. Since most storage rates are less than about 20 vehicles per minute, the cumulative effect of the storage rate reduced the forecast by no more than 20 vehicles.

Occupancy should also have a fraction for a coefficient since, theoretically, occupancy could be up to 100 percent. In such cases, a coefficient of one combined with the high occupancy would reduce the forecast by 100 vehicles, approximately the one-minute capacity of a three-lane freeway. Higher coefficients would have a more extreme effect.

The negative lane occupancy and storage rates indicate that they will reduce volumes. The only possible exception to this would be that the occupancy would cause an increase in volume rather than a decrease. This happened in cases where conditions were below capacity, or on the left side of the flow-density curve. In this particular .case, conditions were congested, so the negative sign was within a reasonable range.

The new model resulted in the following errors:

MAE = 4.94

MAE% = 10.31

MSE = 46.08

The updated model had the following errors:

MAE = 4.33

MAE% = 8.98

MSE = 31.44

Interpretation of Results

These results are compared in Table 3.3 and in Figure 3.1. Percentages of errors are in Figures 3.2 - 3.7.

Table 3.3 Comparison of Models

	Zhu/Nihan Model	Nihan/Knutson Model	
	Old Forecast Model	**New Forecast Model**	**New Updated Forecast Model**
MAE	5.18	4.94	4.33
MAE%	9.90	10.31	8.98
MSE	43.70	46.08	31.44
Emax%	27.52	42.74	30.61

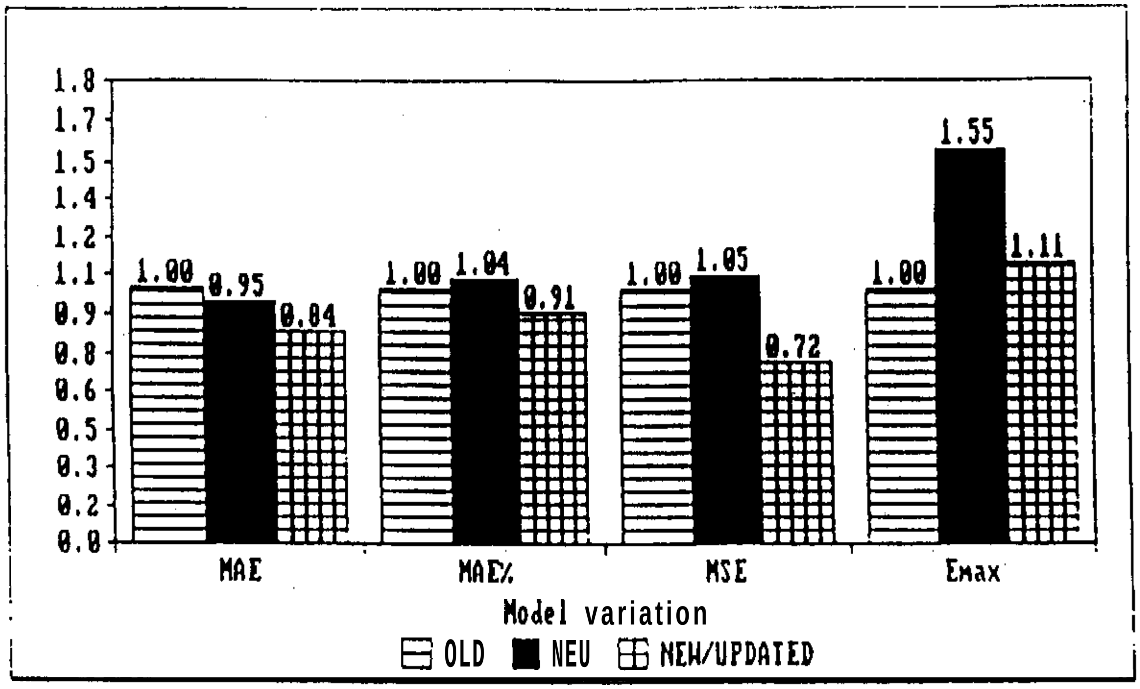

Fig. 3.1 Graphic Comparison of Model Results

It is apparent from Table 4.3 and Fig. 4.1 that the new model with updating is the better model, if a hetter model is to he determined through only the terms introduced previously. However, inspection of Figures 4.2 through 4.7 suggest that differences in new models appear to he minor.

A closer look at the plots shows that the new model seems to have a problem with overprediction, while the old model tends to underpredict. The new model with updates also seems to overpredict, and with little difference from that of the new model without updates.

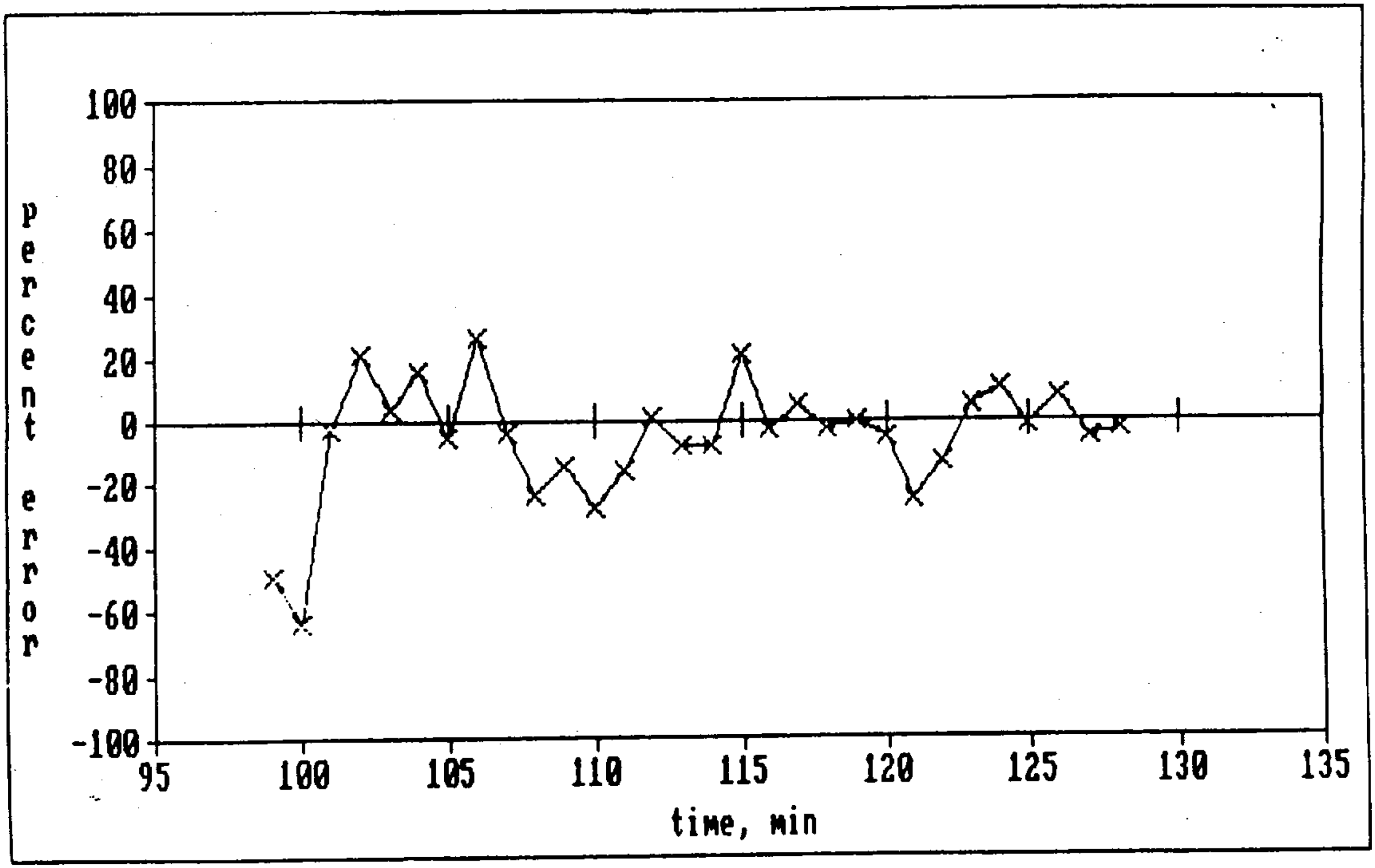

Fig. 3.2 One-Minute Error For Least Squares Method

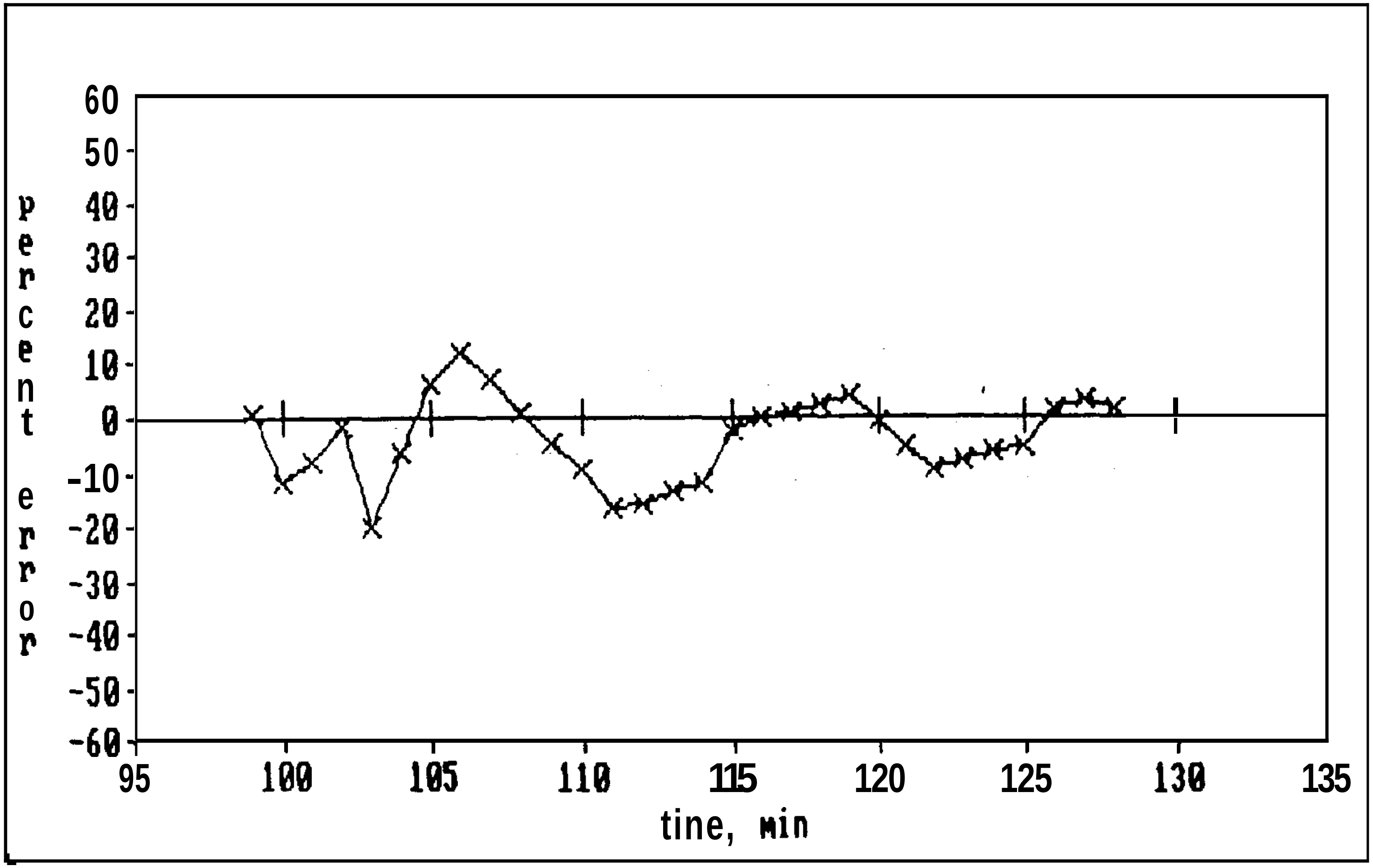

Fig. 3.3 Five-Minute Error For Least Squares Method

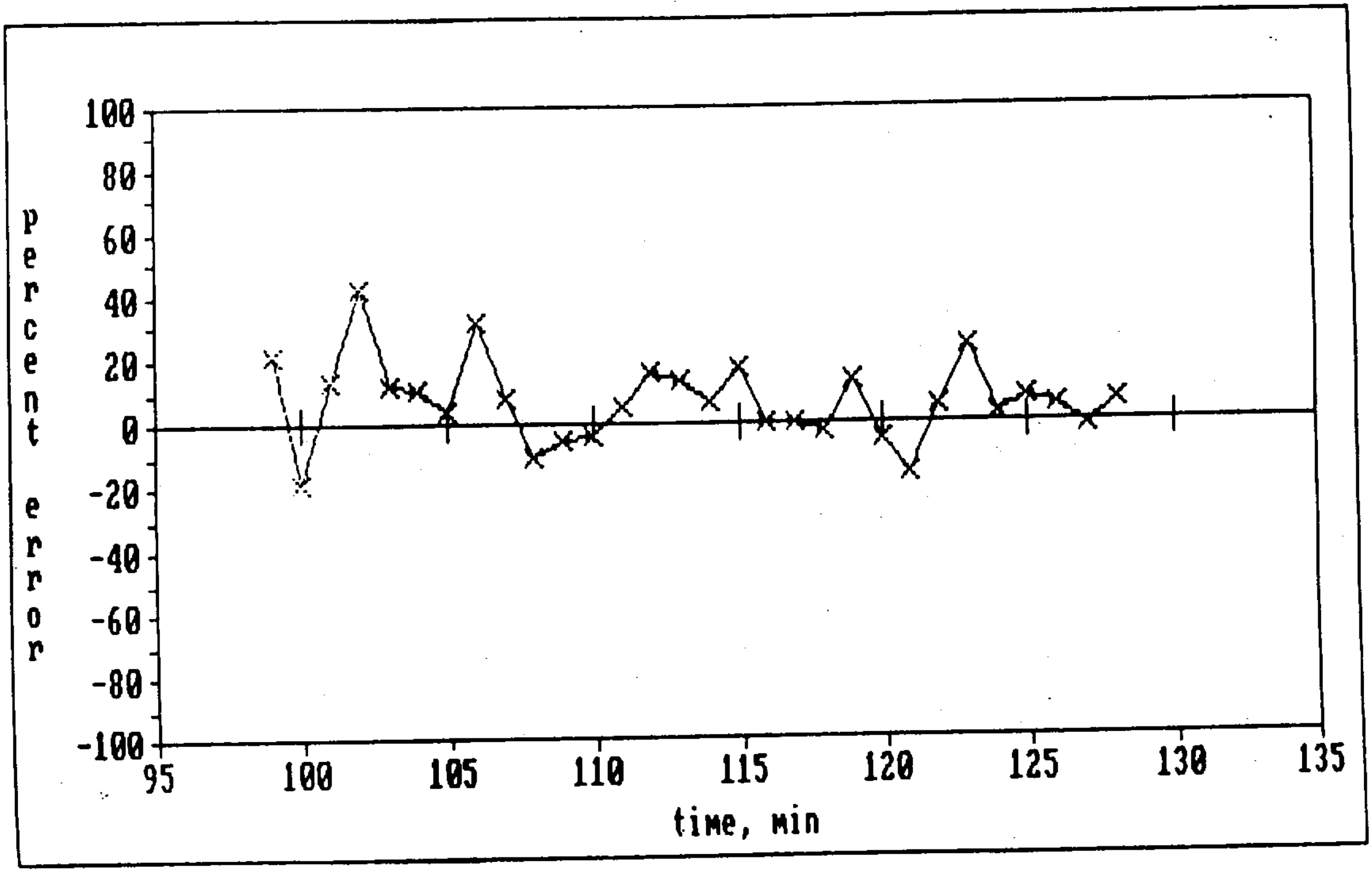

Fig. 3.4 One-Minute Error For New Method

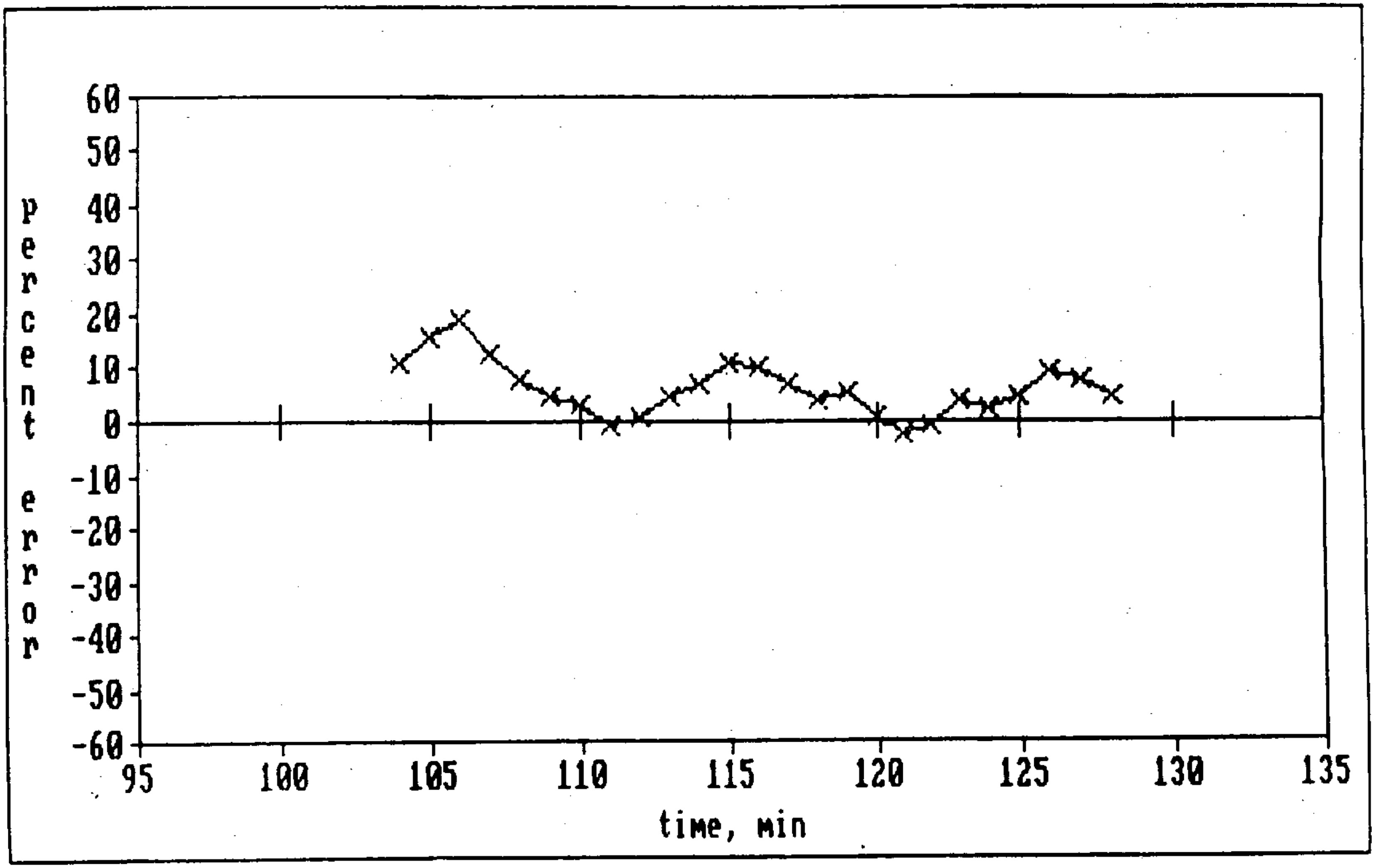

Fig. 3.5 Five-Minute Error For New Method

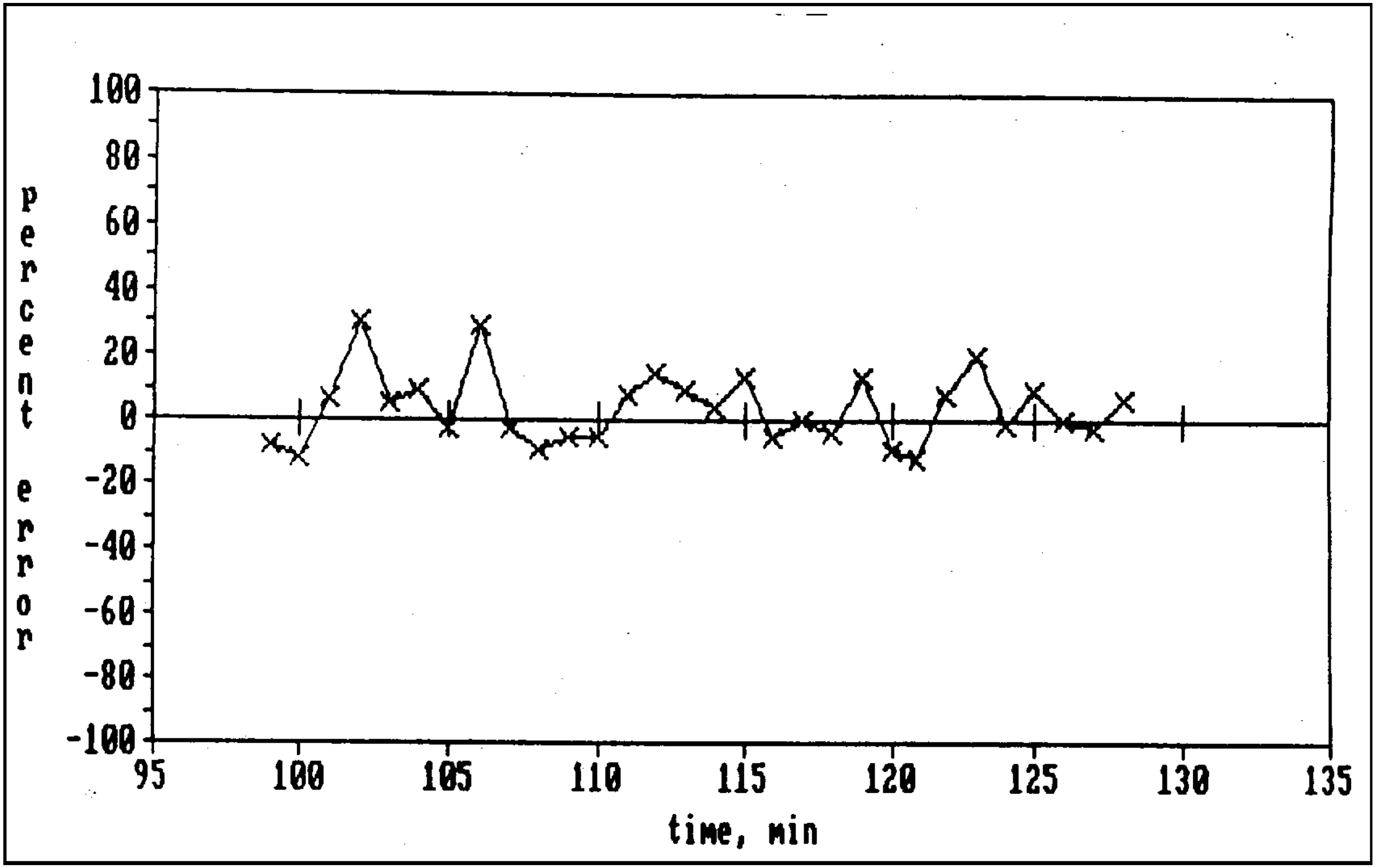

Fig. 3.6 One-Minute Error For New Method With Updates

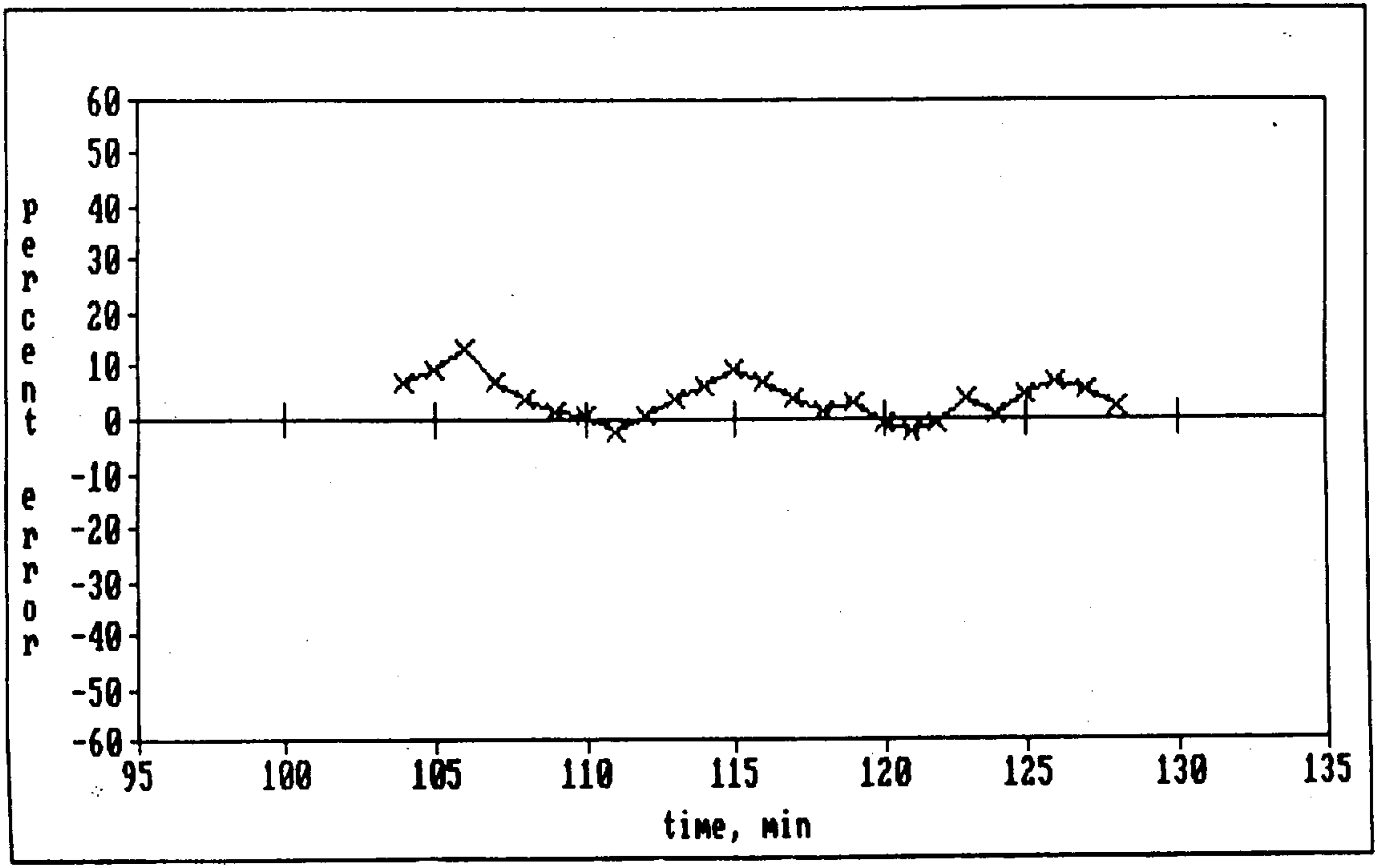

Fig. 3.7 Five-Minute Error For New Method With Updates

These prediction error results posed an interesting problem, considering an incident occurred on the day used for the analysis. Intuition suggests that due to the lack of information on downstream conditions, the old model should have forecasted lower volumes than were experienced. This lack of information should have presented a problem since congestion occurring downstream can reduce the traffic volumes upstream, as previously discussed. The overprediction calculations of the Nihan/Knutson model seemed to result from inadequate accounting of storage rates and occupancies, which in turn, provided negative influences on the volume. The updated model performed slightly better, but the overprediction suggests some systematic error in the forecasts.

The plots in Figures 3.2 -- 3.7 show that the Nihan/Knutson new models seem to recover from missing data faster than the Zhu/Nihan old model. This is because some of the data can be replaced as time passes. Since storage rates are based on previous forecasts, they provide a safety cushion for missing data. This cushion is more of a redundancy, as in a structure. By having more information in a model, each individual part loses significance and may be removed without causing significant problems; however, removing a large portion of the forecasting equation does have a negative effect.

Analysis of the five-minute error prediction results in Figures 3.3, 3.5, and 3.7 also indicates that both new models tend to overpredict traffic volume, whereas the old Zhu/Nihan model tends to underpredict it. This analysis also indicates that the assumption that error rates are normally distributed is invalid in this case. In a near perfect model, such error rates would have a zero mean, and likely be distributed so that few points fall very far from the mean.

When the data were examined and an incident was discovered, a trial run of the models was then done to see how well an incident could be detected in each. The first 60 data points were used to generate coefficients for the data lagged as in the above models. The coefficients, which are given in Table 3.4, were slightly different for these models, as might be expected. The table includes all of the models. The letters indicate where each value goes in the respective equations as described previously in section 3.3.

Table 3.4 **Coefficients** of Incident Test Models

Coefficient	Zhu/Nihan model	Nihan/Knutson model
Z	--	39.71
a_1	0.152	0.462
a_2	--	0.250
b_1	0.329	-0.312
b_2	0.576	--
c_1	0.023	-0.297
c_2	0.235	-0.194
d_1	--	-0.112

Error plots for the equations using the above coefficients can be seen in Figures 3.8 through 3.10.

Inspection of these error plots for incident detection purposes shows that all methods are equally effective in flagging an incident, indicated by the large percentage of errors (the difference between forecasted volume and actual volume). It also appears that due to the large forecasting errors a few minutes prior to time slice 85, which was the time of the incident, these models may have forecasted the incident. In detecting incidents, it seemed that as more information is fed into the model these differences in forecasted and actual volume tended to die out more quickly over time. In that respect, the old model is better for incident detection purposes. Conversely, the new models detected an incident, let the error dissipate, and then proceeded with forecasting. The forecasts will still have errors, but as time progresses, these errors should approach the mean error for the series.

Residuals for forecasts were checked for stationarity. All models had stationary residuals, so they should have valid results.

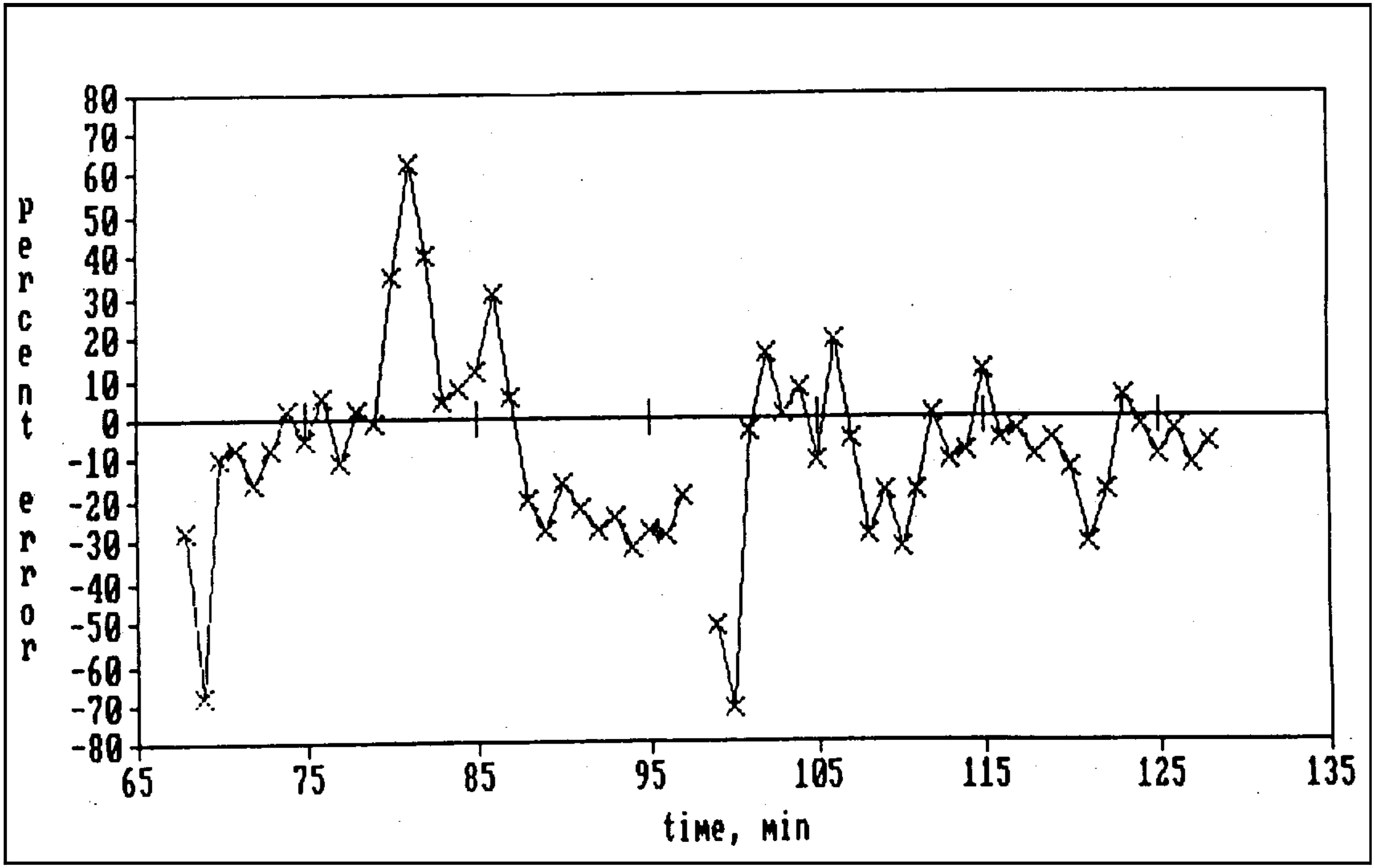

Fig. 3.8 Incident Detection Using Old Model

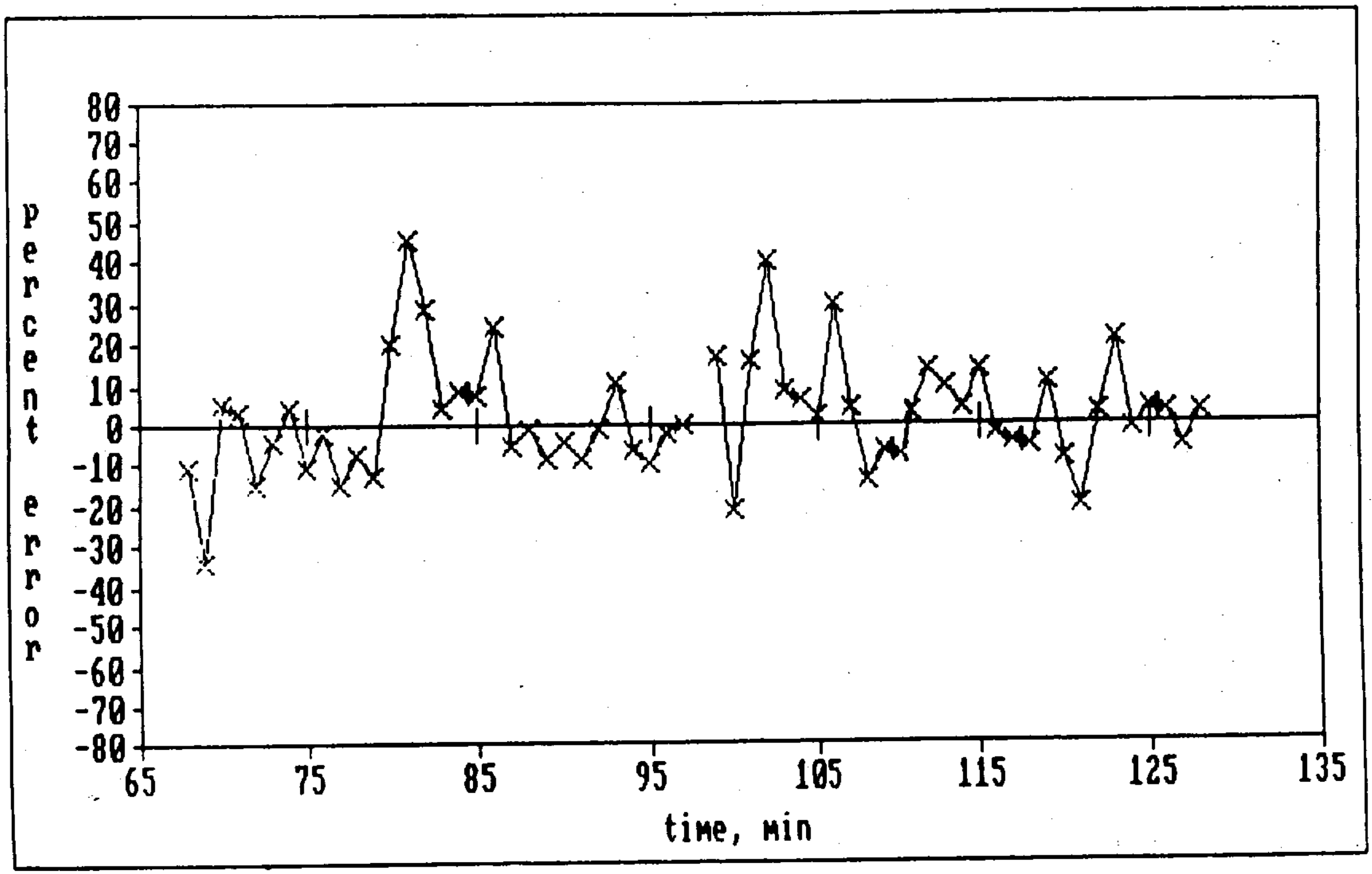

Fig. 3.9 Incident Detection Using New Model

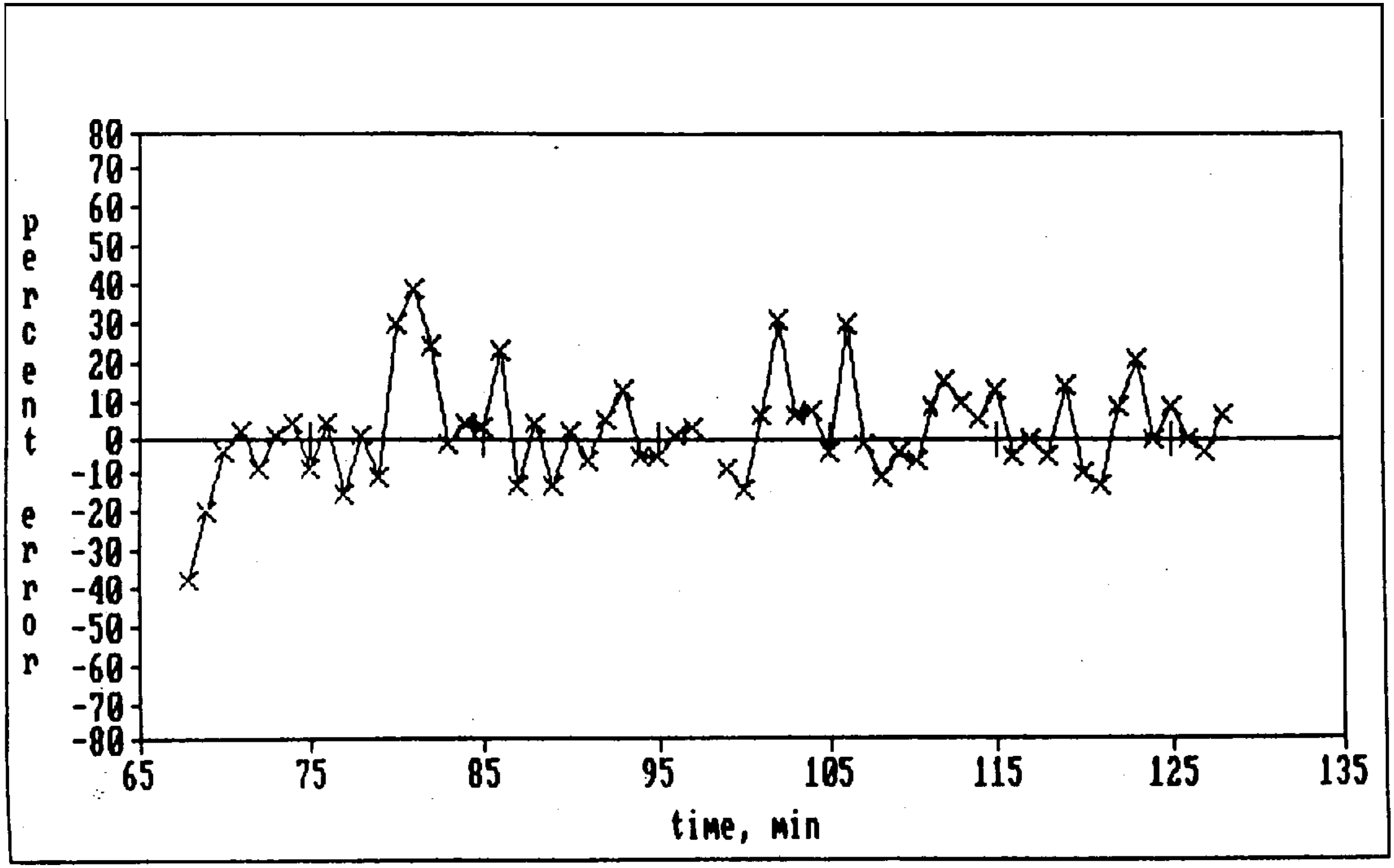

Fig. 3.10 Incident Detection Using New Updated Model

Although use recommendations are outlined in the following chapter, an initial response is noted here. The old (Zhu/Nihan) model might be selected instead of the new (Nihan/Knutson) model, because it is less data intensive. But, as was noted previously, the new model recovers more quickly, and has fewer large forecast errors when the computer has no information for one or more minutes from certain stations.

In transportation planning research, traffic volumes should be as accurate as possible. Using the old model would underpredict traffic. On a larger scale, this could result in the collection of data that do not reflect extreme conditions.

Using the new model would impose extra restrictions on traffic, such as ramp metering and HOV lanes. For example, if data were collected for input into FREQ, the results might suggest that restrictive ramp metering be employed, when the traffic conditions might not require ramp metering at all.

Validity Concerns

Internal Validity

These models were constructed from a single day's data and were checked with another day's data. Statistically, though this may seem an anomalous way to proceed, the selection of two days for analysis might have produced invalid data. At the same time, the two days selected might not have contained data that had any interesting characteristics that may be extremely important, such as an incident, or a lane closure.

Given this type of situation, the best way to proceed was to determined. One way to formulate a model is to generate several models and take some sort of an average value of the coefficients. As was discussed previously, averaging data in this way would have defeated the purpose of the research. The researchers were posed with the problem of determining an appropriate way to average coefficients, and of finding ways to combine equations that had different significant variables (i.e., ramp volume lagged two time periods in one equation and ramp volume lagged three time periods in another equation).

Moreover, since the data did not cover a whole day, the researcher realized the model would probably not be usable during parts of the day other than the morning commutes, such as weekends, off-peak hours and evening commute hours, none of which were not included in this study.

Construct Validity

Most reasons for construct validity problems were due to inadequate planning. The first models developed used storage rates that did not include on-ramps as input volumes, which is incorrect. The correct way to calculate storage rates is to use both incoming and outgoing flows. While this was not a major problem to correct, it caused an unnecessary delay in the research.

To a certain extent this project dealt with a mono-method bias, as the data were collected solely by electronic means. While some of the information could have been collected manually, this would have been exceptionally difficult (synchronizing watches and counting all vehicles in all lanes every minute for two hours), as well as expensive; lane occupancies could not have been determined without electronic devices. Mono-method biases are induced by budget restrictions, as well as the technological limitations of manually collected data.

External Validity

In attempting to construct a model for forecasting traffic volumes, it was essential that we realize the shortcomings of the model, and how applicable it is to other situations.

The setting and treatment, as previously noted, are specific to a particular time of day (the morning commute), and to certain types of drivers (commuters, who may have different characteristics than other drivers, such as vacationers, etc.)

The model was developed at only one location covering a very short length of freeway. This model is quite unlikely to work at any other site, simply based on the fact that the significant lags will probably differ from station to station. The methodology employed here may, in fact, be proven useful for other freeway sections, or other times of the day, but the variables and coefficients in the forecasting equations are specific to sites and times used for this research, and should not be blindly applied to other locations assuming that similar results may be obtained.

Data Collection

One problem discovered in the data collection process was collecting data from incorrect and irrelevant stations. The source of this problem was miscommunication between the researchers and the individuals programming the TSMC computer.

Another problem with the data collected from incorrect and irrelevant stations and data collected in the time series plot was that some data points were missing. This was due to operator error at the TSMC. The operator coding the data requests into the TSMC computer assumed that the printer had to finish the first report before starting the second one, that consecutive reports could not be requested. This is not true; consecutive reports may be scheduled with no time delay. This miscommunication resulted in a delay of at least one minute between thirty-minute reports.

Additionally, data is subject to some measurement error. Since all terms in the forecasting equation were collected by the same methods, the error might have been similar for each term. This assumption may not always be true since each controller in the field has different sensitivity settings.

When congested conditions exist, high traffic densities may cause "blurring" of the volume data to occur. This happens when vehicles get so close that they overlap in the loop detector zone. When this takes place, both vehicles occupy the detector zone, and the controller sees only one vehicle, though two are actually present. Therefore, extremely congested conditions may result in inaccurate data collection.

If both sides of the forecasting equation have some unseen error term with no serial correlation, then using the ordinary least squares regression tends to result in underestimation of the regression parameter. If the variance of the error terms is known, then consistent parameter estimates will result. (13) Unfortunately, it is difficult to obtain the results of error variance and serial correlation.

Since the new model tends to overpredict rather than underpredict, an inaccuracy other than the one described above must exist. It may be that different terms in the equation may be underpredicted to different degrees. For example, the terms causing a negative influence on the forecast may have coefficients with absolute values that are too small, while terms causing a positive influence have coefficients with absolute values that are too large.

CHAPTER 4

CONCLUSIONS AND RECOMMENDATIONS

Conclusions

Evaluation of the new (Nihan/Knutson) models and the old (Zhu/Nihan) model suggests there is little difference among them. All have similar error statistics, and all have a large number of forecasts that exceed a ten percent error (10 of *27* for the new model without updates, 7 of *27* for the new model with updates, and 11 of *27* for the old model). Additionally, applications that develop from this research must be limited to morning commutes at this particular location to maintain validity. But this research may provide an incentive to improve the existing model (for the morning commute) before researching other, more extensive models.

This research is not an accurate comparison between the old model and the new model without updates since the old model used new information to update forecasts and, thus, tended to be more accurate. The new model without updates does not use additional information to update predictions and had slightly worse forecasts. The new model with updating, on the other hand, made more accurate forecasts than the old model.

In the new method incident detection is less evident than in the old method. The "error" due to an incident, which is the difference between actual and forecasted volumes, tends to damp out in the new model as time goes on. This damping tends to mask the effects of an incident by reducing errors more quickly than the older model, making the errors more difficult to detect. Conversely, the old model tends to keep larger differences longer, providing a more apparent error flag. In that respect, the old model is better for the detection of errors. When an incident does occur, the time series plots of the error rates suggest that neither model is acceptable for anything other than detection of that incident immediately following the incident. Incidents are interesting freeway phenomena, and studying them requires a considerable amount of traffic data, but current models are unable to supply this data.

As noted previously, data collection measures have the shortcoming of being heavily dependent on other stations, which also raises the issue of the accuracy of data forecasted from other forecasted data. It is difficult to ascertain how errors propagate -- linearly, exponentially, logarithmically, or by other means.

All models are alike in that three time-lags are necessary to get the models started. Due to the comparability of the models studied in Chapter Two, the relation of storage rates to shock-waves may not affect traffic as much as first expected. In other words, since the only difference between the old model and the new models is in their lane occupancy and storage rates use, then the similar test results suggest that with respect to model accuracy, the effects of storage rates and lane occupancies are minor.

Recommendations

It is best to select a model in which data intensity is minimized, and intuitive understanding is maximized. Considering these criteria, the model developed by Zhu and Nihan seems to be the model of choice. Another advantage of this model is that it requires less computer time, which is the result of its low data-input requirement.

Yet, the new model with updates provides better forecasts. Selecting a model requires that a decision to evaluate the effects of improved forecasts versus simplicity and speed be made. Administrators at WSDOT, or other agencies that forecast traffic volumes, need to decide the value of accurate forecasts over fast, simple, and economical forecasts.

Considering the data intensity of the new model, we do not recommend that it be used for on-line applications. The new model could be used in an off-line situation to replace missing or bad data. If an on-line situation is required, the old model should be implemented, and in select locations only. These locations should be areas of high interest due to high numbers of incidents of accidents, ramp meters (current or projected), or existing research at a particular location.

If the research goal is incident detection, some sort of simple algorithm should be made so that the computer looks at actual traffic volumes and occupancies in real time, and compares these traffic volumes with forecasts. If any large difference is noted (such as exceeding Emax%, described in Chapter Three), this may be a cause for further investigation on the part of the FLOW operator.

The present is a good time to put a forecasting method into place for testing. Since a new WSDOT freeway management computer system is coming on-line soon, this offers a chance to program it now, while the current system is operational. It would then, not be necessary to take the system off-line and reduce services in order to install a forecasting method. This also allows for calibration and software debugging in a safe environment, where mistakes can be made without impacting freeway operations

Implications

Further Research Issues

Further research is necessary on this project. Both methods seem promising, but only on a local scale. Possible future research topics could include methods for coordinating forecasts for three or four stations totally independent of one another, or some stations using forecast data and others using "real" data, and perhaps another situation with all stations using forecast data. Another topic of interest might be identifying important stations such as ramp meters, locations of high traffic volume, and places of high accident and incident rates for use, rather than coordination of an entire freeway network for these types of forecasts.

Perhaps models could be simplified as well. One way to go about this would be to generate models for a few stations on the present system, and to compare coefficients with like variables. It may be that there is little variation in the coefficients, which would indicate that possible coefficient ranges could be constructed for forecasts.

Operational Issues

Given the site-specific nature of this research and other projects, a trend is developing to generate models that work in one place, with plans for expansion as good models are developed. Unfortunately, this has had the effect of making models infinitely more complicated on a large scale than they were when used individually. Research using the old model developed by Zhu and Nihan (in 1990), used spectral analysis to determine appropriate lags for the specific stations. If such a model were to be implemented on the freeways currently monitored by the TSMC, it would take several months or possibly a few years with the current computer system. In 1992, when a new computer comes on-line, this time may be reduced, but determining lags is so data intensive that it is not unreasonable to expect delays in achieving complete system coverage.

In general, it is not practical to develop models that are so site specific, due to the complex interactions among stations. For example, the old model requires input from three places: two upstream mainline stations, and an entrance ramp. If one of those stations malfunctions, the model breaks down. Likewise, the new model requires data from five places: three mainline stations, and one each entrance and exit ramp.

Given that WSDOT has over 110 data cabinets located throughout the Seattle region, this results in approximately 770 different schemes for forecasting traffic volumes (one forecast method for each day of the week, for each station). If each of these is dependent on three to five other stations for forecasts, this quickly mounts to between 2,300 and 3,800 different variables needed in memory. While this is not an inordinately large number of variables for a mini-VAX to handle, the number of calculations needed every minute can only compromise the main goals of the system, which are surveillance, control and driver information, by slowing down the result tabulations. If a system is so dependent on its individual parts, it is not difficult to imagine situations where entire systems might break down, much like a telephone line being cut, and the resultant power outage lasting several hours.

This model may not hold up very well under different conditions. Days where there is little or no congestion will certainly have different characteristics than we discovered in this research. A different model may be necessary for different conditions. If this is the case, criteria for changing from one model to another must be developed. Criteria for developing types of conditions to model must be developed as well. Since research takes place for many reasons, it seems logical that WSDOT would hope to use some form of forecasting method 24 hours a day. Different forecasting methods might be employed during the morning, evening, and lunch commutes, the periods between those times, and in the late night and early morning if useful utilization of this research is to be gained. Research on all of these times of day should be undertaken.

Policy Issues

Given the site-specific nature of this research and other projects, a trend seems to be developing to generate models that work in one place, with plans for expansion as good models are developed. Unfortunately, this has had the effect of making models infinitely more complicated on a large scale than they were when used individually. Consider the older model developed by Zhu and Nihan in 1990. Research on that model used spectral analysis to determine appropriate lags for the station in question. If such a model is to be implemented on the freeways currently monitored by the TSMC, it would take several months or possibly a few years to do so on the current computer. In 1992 when a new computer comes on-line, this time may be reduced, but the nature of determining lags is so data intensive that it is not unreasonable to expect delays in achieving complete system coverage. **As** a rule, however, simplification of models should be stressed to encourage system-wide application.

REFERENCES

[1] Jacobson, Leslie N., Kim C. Henry and Omar Mehyar, "A Real-Time Metering Algorithm for Centralized Control." *Transportation Research Record* **1232,** Urban Traffic System & Operations, Washington, D.C., Transportation Research Board, National Research Council, **1989: 17-26.**

[2] Ahmed, Mohamed S. and Allen R. Cook, "Analysis of Freeway Traffic Time-Series Data by Using Box-Jenkins Techniques." *Transportation Research Record 722* **1979: 1-9.**

[3] Box, George E.P., and Gwylim M. Jenkins *Time Series Analysis Forecasting and Control, Revised Edition.* San Francisco, Holden-Day **1976.**

[4] Eldor, Menahem, "Demand Predictors for Freeway Control Systems." Proceedings of the 7th International Symposium on Transportation and Traffic Theory. Kyoto, Japan, **14-17** August **1977: 341-358.**

[5] Levin, Moshe and Yen-Der Tsao, "On Forecasting Freeway Occupancies and Volumes." *Transportation Research Record* **773 1980: 47-49.**

[6] Nicholson, **H.** and C.D. Swam, "The Prediction of Traffic Flow Volumes Based on Spectral Analysis." *Transportation Research* **8 1974: 533-538.**

[7] Stephanedes, Yorgos J., Panos G. Michalopoulos and Roger A. Plum, "Improved Estimation of Traffic Flow for Real-Time Control." *Transportation Research Record* **795 1981: 28-39.**

[8] Trigg, D.W. and A.G. Leach, "Exponential Smoothing With an Adaptive Response Rate." *Operations Research Quarterly* **18.1** March **1967: 53-59.**

[9] Zhu, Jinde, *Short-Term Forecasts of Freeway Traffic Volumes and Lane Occupancies.* Masters Thesis, University of Washington, Seattle, **1990.**

[10] Kreer, John B. "Factors Affecting the Relative Performance of Traffic Responsive and Time-of-Day Signal Control." *Transportation Research* **10 1976: 75-81.**

[11] *Traffic Flow Theory, A Monograph.* Transportation Research Board Special Report **165,** Washington, D.C., Transportation Research Board, **1975.**

[12] May, Adolf D., *Traffic Flow Fundamentals.* Englewood Cliffs, N.J., Prentice Hall **1990.**

[13] Pindyck, R.S. and D.L. Rubenfeld, *Econometric Models and Economic Forecasts.* New York. McGraw-Hill **1976.**

BIBLIOGRAPHY

Daly, A.J., and J. de D. Ortzar, "Forecasting and Data Aggregation: Theory and Practice." *Traffic Engineering and Control* 31.12 December 1990: 632-643.

Davis, Gary A. and Nancy L. Nihan, "An Application of a Nonparametric Regression Technique in Short-Term Freeway Traffic Forecasting." *ASCE Journal of Transportation Engineering* 1989.

Davis, Gary A. and Nancy L. Nihan, "Use of Time-Series Designs to Estimate Changes in Freeway Level of Service, Despite Missing Data." *Transportation Research A* 18A.5/6 1984: 431-438.

Davis, Gary A. et al., "Adaptive Forecasting of Freeway Traffic Congestion." *Transportation Research Record* 1287 1990: 29-33.

Freedman, David, Robert Pisani and Roger Purves, *Statistics.* New York, Norton, 1978.

Gazis, Denos C. and Charles H. Knapp, "On-Line Estimation of Traffic Densities From Time-Series of Flow and Speed Data." *Transportation Science* 5.4 August 1971: 283-301.

Kalman, R.E., "A New Approach to Linear Filtering and Prediction Problems." *Journal of Basic Engineering* 82D 1960: 35-45.

Nihan, Nancy L. and Gary A. Davis, "Recursive Estimation of Origin and Destination Matrices from Input/Output Counts: *Transportation Research B* 21b.2 1987: 149-163.

Okutani, Iwao and Yorgos J. Stephanedes, "Dynamic Prediction of Traffic Volume Through Kalman Filtering Theory." *Transportation Research* B 18B.1 1984: 1-11.

Papacostas, C.S., *Fundamentals of Transportation Engineering.* Englewood Cliffs, N.J., Prentice Hall 1987.

Papageorgiou, Markos, *Applications of Automatic Control Concepts to Traffic Flow Modeling and Control.* Berlin, Springer-Verlag 1983.

Payne, Harold J., William S. Meisel and Michael D. Teener, "Ramp Control to Relieve Congestion Caused by Traffic Disturbances." *Highway Research Record* 469 1974: 52-64.

Priestley, M.B., *Spectral Analysis and Time Series.* San Diego, Academic Press 1981.

Stephanedes, Yorgos J., Eil Kwon and Panos Michalopoulos, "On-Line Diversion Prediction for Dynamic Control and Vehicle Guidance in Freeway Corridors." *Transportation Research Record* 1287 1990: 11-19.

Wattenburg, Jacob and Steven G. Godfrey, "Short Term Traffic Forecasts...and Reality: A Before-and-After Study." *ITE* 1989 *Compendium of Technical Papers* 1989: 390-393.

APPENDIX B

Study Data and Plots

Volume Series at 212th St. S.W.

106	86	103	93	106	94	86	107	102	105	108	109
117	111	104	99	98	111	97	99	99	85	87	92
84	82	100	103	107	113	*	*	*	*	*	*
106	91	86	115	83	89	94	110	104	68	102	104
87	83	96	98	105	91	89	116	111	105	105	96
65	93	102	116	97	82	*	89	78	93	105	112
109	99	100	105	111	103	93	90	73	88	97	74
97	77	74	72	59	30	31	34	48	45	47	52
40	*	58	56	58	62	64	63	54	41	61	57
53	42	41	44	61	65	65	67	56	55	42	56
59	57	55	54	53	57	62	50				

Volume Series at 220th St. S.W.

74	89	80	92	86	96	88	85	93	89	98	91
98	101	96	99	102	89	87	91	83	96	89	84
75	89	67	74	93	92	*	*	*	*	*	*
88	94	98	84	82	91	69	79	83	92	86	76
86	76	90	79	83	84	97	76	86	107	99	100
100	60	78	86	99	88	*	79	80	83	62	90
81	88	85	79	92	93	97	87	82	80	80	82
55	20	25	25	30	35	23	31	32	29	27	40
33	*	41	40	47	45	33	39	47	26	39	27
31	48	37	48	44	46	40	35	40	45	31	31
48	47	33	45	40	39	37	26				

Occupancy Series at 220th St. S.W.

13.4	15.6	14.7	17.4	19.6	16.5	14.6	14.7	15.6	15.1
17.3	19.1	19.1	21.4	26.5	28.7	29.2	26.7	29.2	30.5
22.2	28.3	26.8	20.7	15.6	16.8	10.6	12.3	17.0	19.9
*	*	*	*	*	*	15.0	15.8	17.2	15.8
15.1	18.6	14.3	23.6	26.8	26.2	26.0	17.1	21.1	28.9
24.0	19.9	18.2	22.3	26.2	22.5	23.6	24.6	21.5	26.9
28.3	15.1	16.2	20.5	24.0	27.3	*	22.6	22.7	17.9
12.3	20.3	21.1	27.7	27.9	25.8	28.7	31.5	29.5	30.9
31.6	27.3	20.1	23.3	46.3	73.9	73.6	75.5	75.1	56.9
54.2	58.8	61.9	58.7	52.7	59.1	63.0	*	62.4	66.2
56.4	57.0	73.5	61.1	60.7	65.8	70.4	66.5	57.0	52.3
66.2	50.9	58.7	60.7	64.1	71.1	66.5	49.5	70.8	67.4
49.0	54.6	60.9	56.5	62.0	62.3	65.7	65.2		

* represents missing data

Volume Time Series at 236th St. S.W.

104	97	88	99	92	97	116	98	92	99	107	101
95	106	102	100	106	105	95	97	96	100	98	95
99	96	83	108	84	85	*	*	*	*	*	*
97	91	99	109	97	94	87	8	85	101	97	100
87	102	93	86	97	91	92	98	94	90	105	103
97	104	106	90	89	86	*	93	88	94	94	101
91	95	100	91	105	93	104	77	62	66	83	82
81	61	53	43	52	45	52	49	40	52	52	48
48	*	55	51	47	42	52	45	48	42	48	50
51	53	53	47	50	55	46	54	49	50	51	53
53	54	51	49	48	53	53	48				

Volume Time Series at 244th St. S.W.

87	102	96	100	104	96	109	106	99	93	110	111
101	102	109	106	98	105	106	97	94	105	99	94
92	103	90	103	98	87	*	*	*	*	*	*
55	65	73	91	102	86	91	79	94	80	94	87
90	97	87	92	88	91	92	102	92	77	95	114
97	93	99	95	79	92	*	97	94	91	91	103
91	96	99	81	89	95	81	74	81	90	78	88
83	82	86	67	55	50	54	55	48	52	51	53
62	*	58	61	62	67	53	57	48	55	45	57
54	48	61	56	54	62	56	56	58	55	53	61
59	55	70	63	54	60	55	62				

Ramp Volume from 220th St. S.W.

11	9	12	7	9	13	4	10	14	10	12	7	3	15	6	7	9	6	6
10	8	7	7	6	6	7	6	16	6	11	*	*	*	*	*	*	11	7
7	11	9	10	8	12	6	5	11	8	9	6	5	7	6	9	10	7	10
7	11	9	8	7	14	20	17	8	*	6	9	12	8	9	9	13	12	12
12	9	7	7	8	9	14	11	11	15	11	12	13	8	8	10	6	12	12
8	6	*	7	8	8	6	10	6	7	5	6	6	5	6	6	7	6	12
7	7	8	9	9	6	5	5	18	12	7	5	6	6					

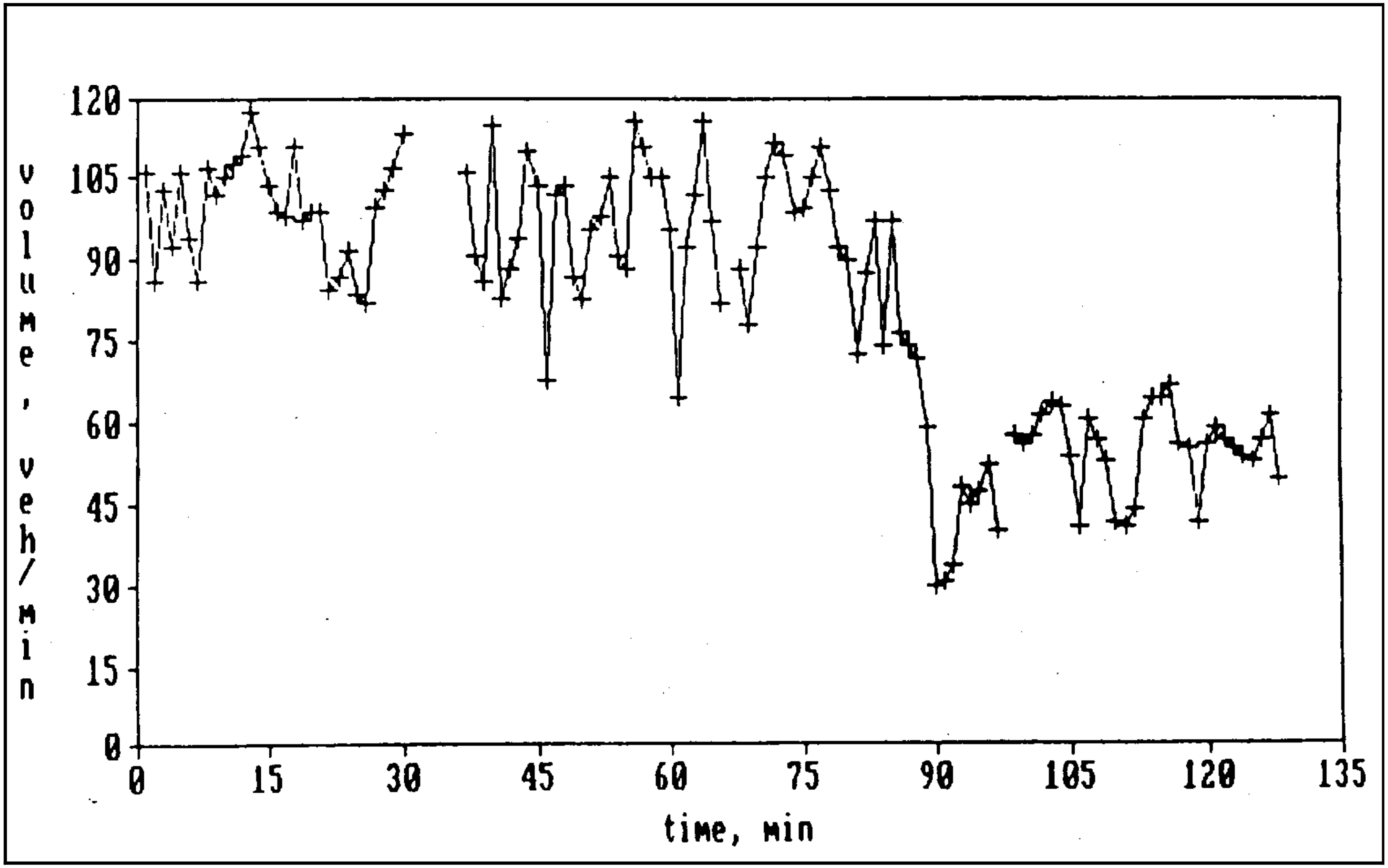

Fig. B.1 Plot of Volume Time Series at 212 S.W.

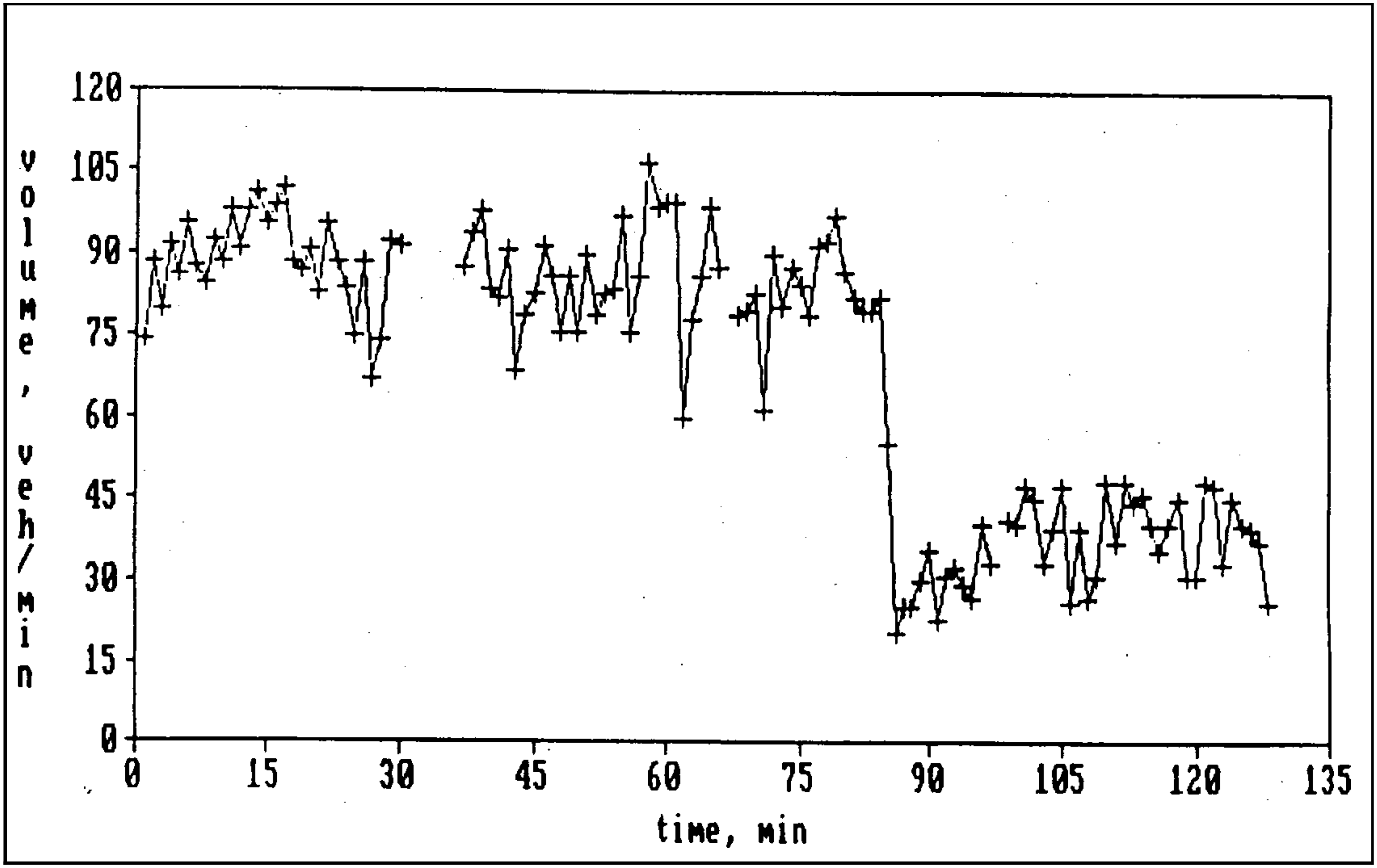

Fig. B.2 Plot of Volume Time Series at 220

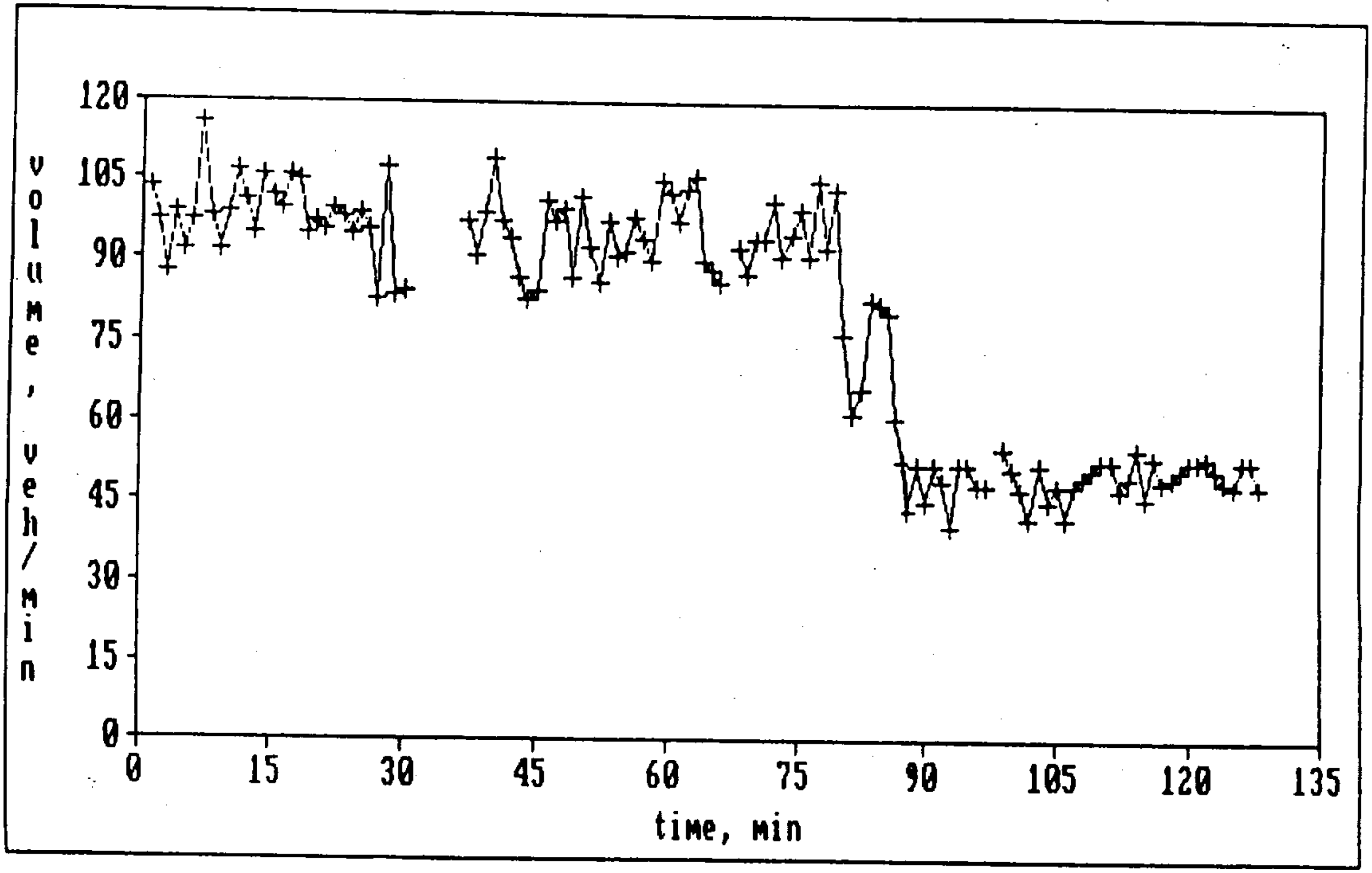

Fig. B.3 Plot of Volume Time Series at 236 S.W.

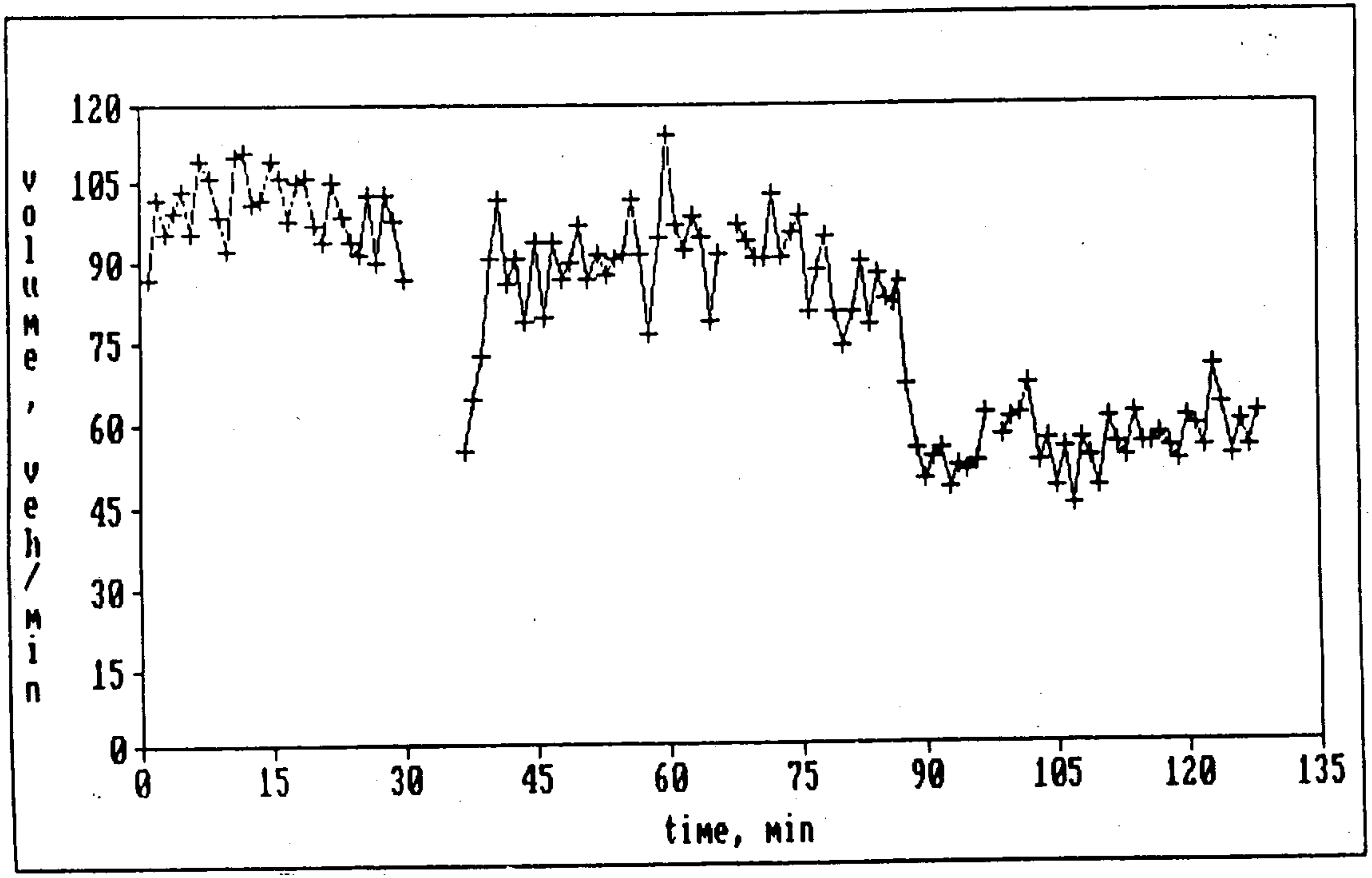

Fig. B.4 Plot of Volume Time Series at 244 S.W.

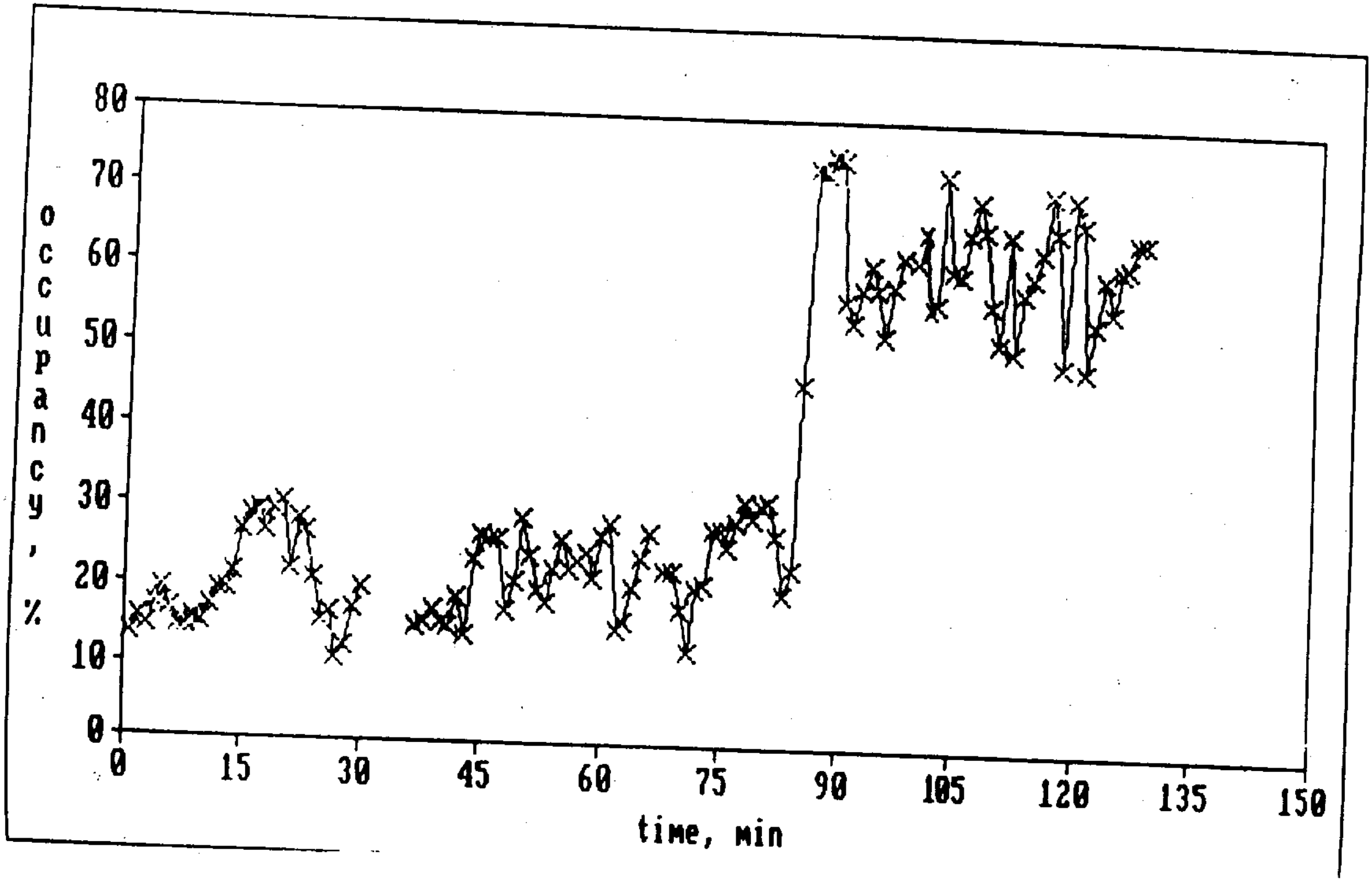

Fig. B.5 Plot of Occupancy Time Series at 220 S.W.

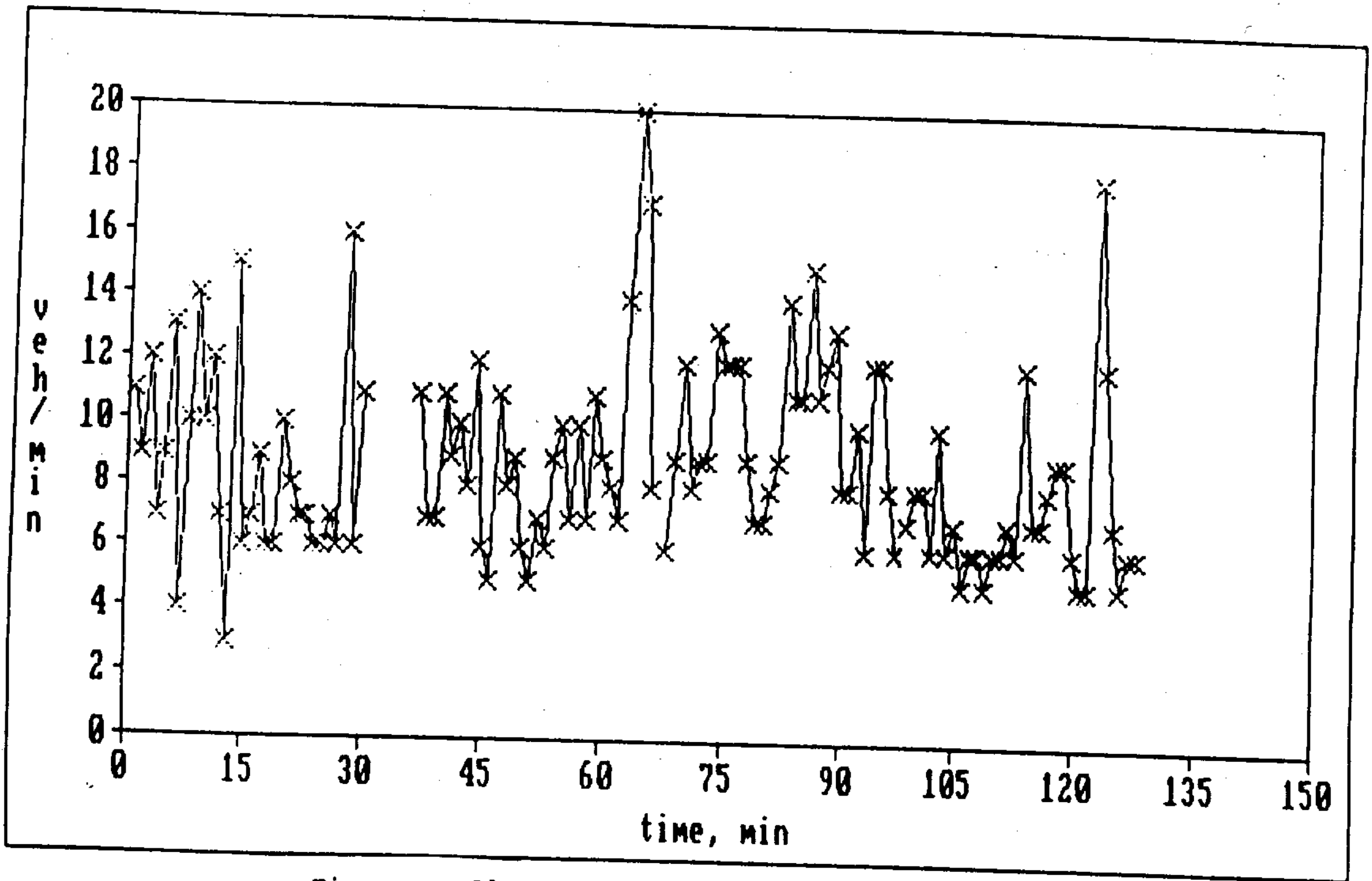

Fig. B.6 Plot of Volume Time Series at 220 S.W. Ramp

APPENDIX C

Results of Old Model Forecasts of

Volumes at 236th St. S.W.

Table C.1 Comparison of Predicted and Actual Volumes Old Model

Predicted	Actual	Error	% Error
51.07	42	9.07	21.59
54.00	52	2.00	3.85
52.22	45	7.22	16.05
45.36	48	-2.64	-5.49
53.22	42	11.22	26.71
46.28	48	-1.72	-3.59
38.25	50	-11.75	-23.50
43.79	51	-7.21	-14.14
38.42	53	-14.58	-27.52
44.86	53	-8.14	-15.35
47.48	47	0.48	1.03
46.04	50	-3.96	-7.93
50.42	55	-4.58	-8.33
55.51	46	9.51	20.67
52.67	54	-1.33	-2.46
51.91	49	2.91	5.93
48.69	50	-1.31	-2.61
50.96	51	-0.04	-0.07
49.90	53	-3.10	-5.84
39.84	53	-13.16	-24.83
46.76	54	-7.24	-13.41
53.59	51	2.59	5.08
54.13	49	5.13	10.47
47.51	48	-0.49	-1.01
57.31	53	4.31	8.13
49.89	53	-3.11	-5.87
47.05	48	-0.95	-1.99

APPENDIX D

Results of New Model Forecasts of

Volumes at 236th St. S.W.

Table D.l Comparison of Predicted and Actual Volumes New Model without Updates

Predicted	Actual	*Error*	*% Error*
59.95	42	17.95	42.74
58.31	52	6.31	12.14
49.60	45	4.60	10.21
50.14	48	2.14	4.47
55.49	42	13.49	32.13
51.68	48	3.68	7.67
45.01	50	-4.99	-9.97
48.66	51	-2.34	-4.59
50.84	53	-2.16	-4.07
56.20	53	3.20	6.04
54.59	47	7.59	16.15
56.45	50	6.45	12.91
58.80	55	3.80	6.91
53.84	46	7.84	17.03
54.11	54	0.11	0.20
48.84	49	-0.16	-0.32
48.89	50	-1.11	-2.23
57.63	51	6.63	13.00
50.45	53	-2.55	-4.81
44.65	53	-8.35	-15.75
57.19	54	3.19	5.92
62.92	51	11.92	23.37
50.60	49	1.60	3.27
52.12	48	4.12	8.58
56.11	53	3.11	5.86
52.33	53	-0.67	-1.26
51.28	48	3.28	6.83

Table D.2 Comparison of Predicted and Actual Volumes
New Model with Updates

Predicted	Actual	Error	% Error
54.86	42	12.86	30.61
55.07	52	3.07	5.91
49.11	45	4.11	9.14
46.84	48	-1.17	-2.43
54.49	42	12.49	29.74
47.07	48	-0.93	-1.94
45.40	50	-4.60	-9.20
48.65	51	-2.35	-4.60
50.32	53	-2.68	-5.06
57.32	53	4.32	8.14
53.66	47	6.66	14.16
54.86	50	4.86	9.73
57.34	55	2.34	4.26
52.38	46	6.38	13.87
51.20	54	-2.80	-5.18
49.31	49	0.31	0.63
47.75	50	-2.25	-4.50
57.94	51	6.94	13.60
48.26	53	-4.74	-8.95
46.53	53	-6.47	-12.21
58.43	54	4.43	8.20
61.17	51	10.17	19.93
48.61	49	-0.39	-0.79
52.77	48	4.77	9.94
53.22	53	0.22	0.42
51.69	53	-1.31	-2.48
51.29	48	3.29	6.84