2006 IEEE International Conference on Systems, Man, and Cybernetics October 8-11, 2006, Taipei, Taiwan

Application of the ARIMA Models to Urban Roadway Travel Time Prediction – A Case Study

Daniel Billings, *Student Member, IEEE* and Jiann-Shiou Yang†, *Senior Member, IEEE*

***Abstract*—Travel time is the time required to traverse a route between any two points of interest and it is an important parameter that can be used to measure the effectiveness of transportation systems. The ability to accurately predict freeway and arterial travel times in transportation networks is a critical component for many Intelligent Transportation Systems (ITS) applications. In this paper, we focus on the application of using time series models to study the arterial travel time prediction problem for urban roadways and a section of Minnesota State Highway 194 is chosen as our case study. We use the Global Positioning System (GPS) probe vehicle method to collect data. The time series modeling is then developed, in particular, we focus on the autoregressive integrated moving average (ARIMA) model due to the non-stationary property of the data collected. The section models established for the corridor are verified via both the residual analysis and portmanteau lack-of-fit test. Finally, based on the models developed we present our prediction results. Our study indicates the potential and effectiveness of using the ARIMA modeling in the prediction of travel time. The method presented in this paper can be easily modified and applied to short-term arterial travel time prediction for other urban areas.**

I. INTRODUCTION

TRAVEL time information is very important to identify and assess operational problems along highway facilities. This information is also necessary in traffic signal timing control coordination, as input to traffic assignment algorithms, and in economic studies, etc. Since travel time is most meaningful to the motorists and also the one that can be equivalently expressed as monetary cost, travel time predictability can be used to gauge the benefits of Intelligent Transportation Systems (ITS). Therefore, development of efficient methodologies for real-time measurement and estimation of travel time has been recognized as an important component of ITS.

Travel time prediction refers to predicting and calculating the experienced travel time before a vehicle has traversed the arterial/freeway or route of interest. Travel time estimation and prediction has been an important research topic for decades. Many previous studies have been focused on predicting travel times using the time-series models [1], the artificial neural network models [2, 3], the non-parametric regression method [4], the weighted moving average and cross correlation methods [5], and the adaptive filtering techniques [6], etc. Although research on travel time estimation for freeways is very rich, research on arterial travel time estimation is quite limited. Prediction of travel time is potentially more challenging for arterials than for freeways because vehicles traveling on arterials are subject not only to queuing delay but also to signal delays as well as delays caused by vehicles entering from the cross streets.

This paper focuses on the travel time prediction problem by studying the travel time data, modeling, and diagnostic checking so that the short-term travel time can be reasonably predicted. To explore the feasibility and effectiveness of this approach, a section of the Minnesota State Highway 194 corridor is chosen as our case study. We collected the section travel time data via the Global Positioning Systems (GPS) probe vehicle technique [7]. Since the data can be considered as a collection of observations made sequentially in time and treated as a realization of a stochastic process, the travel time modeling with the time series techniques is used. Particularly, we focus on using the autoregressive integrated moving average (ARIMA) model due to the non-stationary property of the data we collected. The models established for the corridor during the afternoon rush hour are verified via both the residual analysis and portmanteau lack-of-fit test. Finally, based on the models developed we present our preliminary prediction results. Our study indicates the potential and effectiveness of using the time series modeling in the prediction of arterial travel time. Furthermore, the results presented here can be easily modified and used in short-term arterial travel time prediction for other urban areas.

II. TRAVEL TIME DATA COLLECTION

The GPS probe vehicle method involves collecting data with the aid of instrumented vehicles capable of receiving GPS signals for position and time information. The vehicle-tracking unit is based on GPS technology to monitor a vehicle's location and travel time information. And it utilizes the wireless data network to transmit data to the web server. These data includes test vehicle ID, longitude, latitude, speed, direction, time stamp, date, etc., and the time stamps data are used to calculate section travel times. The test site is a 3.7-mile Minnesota State Highway 194 corridor, one of the most heavily traveled and congested roadways in the Duluth metro area. Along this corridor, there are 10 signalized

Final manuscript submitted June 12, 2006 to The 2006 IEEE International Conference on Systems, Man, and Cybernetics. This work was supported by the Minnesota Department of Transportation (Mn/DOT) under Contract No: 08115, Work Order No: 130.

† Corresponding author (phone: 218-726-6290; fax: 218-726-7267; e-mail: jyang@d.umn.edu). D. Billings and J.-S. Yang are with the Department of Electrical and Computer Engineering, University of Minnesota, Duluth, MN 55812, USA.

intersections. From east to west, the cross streets are: Mesaba Avenue – Pecan Avenue – Arlington Avenue – Basswood Road – Anderson Road – Mall Drive – Trinity Road – Cottonwood Avenue – Maple Grove Road – Haines Road. The outbound speed limit is 30 miles per hour from Mesaba Avenue to Mall Drive and then becomes 40 miles per hour from Mall Drive to Haines Road. The inbound speed limit is 30 miles per hour for *the* entire road stretch. Only the outbound traffic is considered due to the afternoon rush hours. The test vehicle was equipped with a vehicle-tracking unit to collect weekday rush-hour (3:30-5:00 pm) data from the period of October 26, 2004 to June 24, 2005. To check whether the test vehicle enters an intersection we use the Cohen Sutherland line-clipping algorithm [8] with some modifications. We assume that the intersection is a square in space, which corresponds to the definition of window in the line-clipping algorithm. Note that the Cohen-Sutherland algorithm determines the intersection of the line joining the two points with each of the window to determine which part of that line to display. However, in our case this part is omitted because our travel space is limited to the road we travel on. We don't have to consider whether the line having zero out-code is really intersecting the window or not. The out-code is a set of binary digits that represents the position of a point with respect to that window and it defines the half-space that a point belongs to. For details about the Cohen-Sutherland algorithm, please refer to [8].

III. The ARIMA Model

A time series is a collection of observations made sequentially in time. Any data recorded over time can be considered as a time series. A time series model for the observed data, say $\{x_t\}$, is a specification of the joint distributions of a sequence of random variables $\{X_t\}$ of which $\{x_t\}$ is postulated to be a realization. The term time series can mean both the data and the process of which it is a realization. The fundamental goal of time series analysis is to understand the underlying mechanism that generates the observed data and, in turn, to forecast future values of the series. In the following, we briefly review the commonly used time series models.

Time series modeling assumes that the value of the series at time t (i.e., X_t) depends only on its previous values and on a random noise. Therefore, if this dependence of X_t on the previous p values is linear, then X_t can be represented by $X_t = \Phi_1 X_{t-1} + \Phi_2 X_{t-2} + \ldots + \Phi_p X_{t-p} + Z_t$, where $\Phi = (\Phi_1, \Phi_2, \ldots, \Phi_p)$ are the model parameters called the autoregressive (AR) coefficients and Z_t is the disturbance at time t. The process $\{Z_t\}$ is usually modeled as an independent and identically distributed (iid) white noise with zero mean and variance σ^2. That is, $E[Z_t] = 0$, $E[Z_t^2] = \sigma^2$ for all t, and $E[Z_t Z_s] = 0$ if $t \neq s$, where E[.] means the expectation. The process $\{X_t\}$ is said to be a moving average process of order q if X_t can be written as $X_t = Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + \ldots + \theta_q Z_{t-q}$, where $\theta = (\theta_1, \theta_2, \ldots, \theta_q)$ are the moving average (MA) coefficients. In the above, p and q are the orders of AR(p) model and MA(q) model, respectively. By combining the AR and MA parts, we get a mixed autoregressive moving average (ARMA) process of order (p, q). That is, $X_t - \Phi_1 X_{t-1} - \Phi_2 X_{t-2} - \ldots - \Phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + \ldots + \theta_q Z_{t-q}$, and this defines the ARMA(p,q) model. By introducing the back shift operator B, i.e., $B^i X_t = X_{t-i}$, then the ARMA (p, q) model can be simplified as $\Phi(B) X_t = \theta(B) Z_t$,where $\Phi(B) = 1 - \Phi_1 B - \Phi_2 B^2 - \ldots - \Phi_p B^p$ and $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q$. Even though in practice most time series we faced are non-stationary, the stationary ARMA model can still be generalized to incorporate a special class of non-stationary time series models. For instance, if the observed time series is non-stationary, we can difference the series with X_t replaced by $(1-B)^d X_t$ where $(1-B) X_t = X_t - X_{t-1}$, $(1-B)^2 X_t = (1-B) X_{t-1} = X_t - 2X_{t-1} + X_{t-2}$, etc. This operation is called differencing the time series. The ARMA model then becomes $(1-B)^d \Phi(B) X_t = \theta(B) Z_t$, which is called the autoregressive integrated moving average (ARIMA) model and expressed as ARIMA(p, d, q). In other words, any ARIMA(p, d, q) series can be transformed into an ARMA(p, q) series by differencing it d times and, thus, the analysis of an ARIMA process does not pose any difficulty as long as we know the number of times to difference the series. Clearly, the ARIMA process constitutes of three parts, an autoregressive part (AR), a differencing part (I), and a moving average part (MA). The differencing part is used to convert a non-stationary series into a stationary series. It removes the trend from the data.

In time series analysis, it is very important to calculate the sample autocovariance function (ACVF) and sample autocorreleation function (ACF) from the observed data of a given stationary process. The ACVF and ACF provide a useful measure of the degree of dependence among the values of a time series at different times, and for this reason they play an important role when we consider the prediction of future values of the series in terms of past and present values. To find an appropriate model for the data observed we use the correlograms. A correlogram is a graph showing the time series ACF values against the lag h. From observing a correlogram sometimes we can get important information about the time series. For example, is the series stationary? If it is stationary, then is it AR(p), MA(q) or ARMA(p, q) type? What can be the order, i.e, the values of p and q for the series? It is known that for a series that fits MA (q) model, its correlogram should show a sharp cut-off after $h > q$, that is, the ACF becomes zero if $h > q$, a special feature of MA processes. If the correlogram doesn't cut-off sharply and on the contrary, it decays either exponentially or sinusoidally or both, then it may suggest that the time series either an AR (p) or ARMA(p, q) type. In this case the correlogram doesn't provide much information about the order of the series. So, we pursue the partial correlogram (i.e., partial ACF vs. lag h) to see any additional information can be extracted to find the proper order p. It can be shown that the partial ACF of an AR(p) process "cuts off" at lag p. Note that sample correlation functions do not always resemble the true correlation functions, in particular, when the number of data observed is small. Therefore, it should always be used with caution.

Another type of ARMA order selection is based on the so-called information criteria. The idea is to balance the risks of under fitting (i.e., selecting the orders smaller than the true orders) and over fitting (i.e., selecting orders larger than true orders). This is done by minimizing a penalty function, and the two commonly used functions are: $\ln \sigma^2+2(p+q)/n$ (i.e., the Akaike's Information Criterion (AIC)) and $\ln \sigma^2+(p+q) \ln (n)/n$ (i.e., the Bayesian Information Criterion (BIC)), where σ^2 is the estimated noise variance and n is the length of the data. For details regarding AIC and BIC criteria and order selection, please refer to any standard time series analysis books (e.g., [9, 10]).

IV. Data Analysis And Modeling

To fit a time series model to data, we need to transform the raw data into a "well-behaved" form suitable for modeling. In other words, the transformed data can be modeled by a zero-mean, stationary ARMA type of process. Therefore, the section travel time data is first converted to a zero-mean time series and shown in a time plot (i.e., a graph showing the observations against time). Any non-constant or variability should be removed before modeling. The time-plot helps us determine whether the process is stationary. If not, then the series is further processed to make it stationary. Differencing is an effective way to remove trend and seasonal components in a time series. In the remainder of this paper, we use the road section # 7 (i.e., Trinity Road - Maple Grove Road section) results as an example to demonstrate our approach. For the road section # 7, the time plot of our collected travel time data is given in Fig. 1. Since this observed data sequence after removing its mean value doesn't seem to be stationary, the data is differenced once with the result shown in Fig. 2. The differenced data is then used for our modeling purpose. Figures 3 and 4 show the correlogram and partial correlogram of the differenced data. Note that in these two figures, the dotted lines, parallel to the x-axis, represent the error bounds for the data. The bounds are determined based on $\pm 2/\sqrt{n}$, where n represents the number of data points [9, 10]. If the value of the ACF and PACF lie within these lines, then we consider that they are not significantly different from zero. In other words, we can plot approximate 95% confidence limits at this value, and the observed ACF values which fall outside these limits are "significantly" different from zero at the 5% level. Based on the correlograms, it seems that an ARMA (p,q) might be adequate.

The algorithms we used for estimating ARMA parameters of Fig. 2 include: the Yule-Walker method, the Levison-Durbin algorithm, the Burg's algorithm, the Innovations algorithm, and the Hannan-Rissanen procedure [10]. Note that the Yule-Walker method and Levison-Durbin algorithm are mainly used for AR(p) model, while the Burg's algorithm works directly with the data rather than with the sample covariance function and it is asymptotically equivalent to the Yule-Walker estimates for large samples. The Innovations algorithm is used to obtain estimates of the MA(q) model, while the Hannan-Rissanen procedure is mainly used for ARMA parameter estimates. For the time series data in Fig. 1, we found that the ARIMA (6, 1, 3) model with Φ = [-0.339246, -0.778076, -0.116948, -0.0513862, 0.04, -0.158566] and θ = [-0.550511, 0.425886, -0.6653231] seems to generate the best results for the road section # 7. This is based on the residual analysis and the minimum BIC, AIC and portmanteau values produced (see Section V). Following the same approach, the rest of the seven road sections were also studied with the time series analysis applied to each data set. For detail about these models, please refer to [11].

V. Model Validation

After fitting a model to a given set of data, we need to examine the model to see if it is indeed an appropriate one. If we have a "good" model, then we should expect the residuals to be "random" and "close to zero". There are several ways of checking if a model is satisfactory. One of the commonly used approaches to diagnostic checking is to examine the residues. That is, we can treat the residues as a time series and study its properties and the correlogram of the residues (i.e., the autocorrelation coefficients of the residues at different lag h). Therefore, we check the residuals, which are generally defined as the difference between the observation and fitted value. For a good fit, the residual time series should be close to an iid zero-mean white noise. If the residues, say $\{y_1, y_2, \ldots, y_n\}$, is a realization of such an iid sequence, then about 95% of the sample autocorrelations should fall between the bound $\pm 2/\sqrt{n}$. A detailed analysis of residuals from ARMA processes can be found in [10]. To verify the models shown in Section III, we conducted residual analysis. For example, the residuals and the correlogram of the residuals from the ARIMA (6, 1, 3) model for road section # 7 are shown respectively in Fig. 5 and Fig. 6. From Fig. 6, we see that residuals roughly lie within the bounds, indicating that overall the residual time series approximates a zero mean white noise behavior. We see that ACF of the residual time series falls within the bounds except for a few ks and, thus, we have no reason to reject the hypothesis that the set of data constitutes a realization of a white noise process.

Instead of checking to see whether each sample autocorrelation coefficient falls within the bounds, it is also possible to carry out what is called a portmanteau lack-of-fit test. The portmanteau test uses a single value Q, by looking at the first k values of the residual correlogram all at once, to see if the fitted model is appropriate. The test statistic Q is defined as $Q = n\sum_{j=1}^{h}\rho^2(j)$, where n is the length of data (residues), $\rho(j)$ is the sample autocorrelation with lag j, and h is typically chosen in the range of 15 to 30. Therefore, given the residual time series $\{y_1, y_2, \ldots, y_n\}$, we reject the iid hypothesis at level α if $Q > \chi^2_{1-\alpha}(h)$, where $\chi^2_{1-\alpha}(h)$ is the $(1-\alpha)$ quantile of the chi-squared distribution with h degrees of freedom. In other words, if the fitted model is appropriate, then Q should be approximately distributed as χ^2 with h degrees of freedom. We also conducted the portmanteau test in our study. For the road section # 7's model, we found that the portmanteau statistics Q = 358.19 [11]. Since this value Q is less than $\chi^2_{0.95}$ with the given h, we accept the fitted model as adequate. The above diagnostic checking method was also used to other road sections to verify the appropriateness of these models.

VI. TRAVEL TIME PREDICTION

We use the time-series models developed to predict the travel-times for the intersections on the corridor. The travel time prediction was conducted over a two-week period from July 18 to July 22 and July 25 to July 29, 2005 during the afternoon peak hour (i.e., 3:30 - 5:00 pm). Real-time data was used together with our prediction models to perform the one-step-ahead prediction. The observed and predicted values for the eighht road sections are shown, respectively, in Figs. 6-13. In these figures, the curve marked with ♦ means the predicted travel times and the one marked ■ represents the observed values. The x-axis denotes the number of data-points (i.e. 40 next travel times) whereas the y-axis denotes the travel times in seconds. From the figures, we see that these ARIMA time series models produce reasonably good results for most of the road sections. The predicted values are within the range of our observed travel times. It does especially well for the road sections with higher allowed speed limit such as the Trinity – Maple Grove section and the Maple Grove – Haines section. Another observation is that the prediction error seemed relatively large on those road sections with a shorter distance. Lower speed limit and shorter link distance together with the relatively high cross-street traffic can all affect our prediction performance.

It is known that good travel time prediction is very challenging due to the existence of various uncertainties (e.g., weather and road conditions, traffic accidents, etc). In addition, some conditions may not be known well in advance or can be anticipated. There are many methods/techniques proposed in the literature, but to the best of our knowledge none of these methods can claim to be able to apply to all situations with high accuracy. However, based on the study we conducted, it can be seen that, in general, the autoregressive integrated moving average approach can establish a reasonably good modeling for travel time prediction purpose. This approach particularly performed well for arterials subjected to both queuing and signal delays, which sometimes cannot be handled nicely or even performed poorly by using those prediction methods developed on freeways. Our study indicates the potential and effectiveness of using the time series modeling approach in the prediction of arterial travel time.

VII. CONCLUSION

In this paper, we study the modeling and prediction of short-term arterial section travel time via the time series techniques. A section of Minnesota State Highway 194, one of the most heavily congested corridors in the Duluth area, is chosen as our case study. The GPS test vehicle technique is used to collect travel time data and these data are treated as a realization of a time series stochastic process. Time series analysis is then used for modeling of the section travel time. In particular, we focus on the ARIMA model due to the non-stationarity of the observed data. Several techniques such as the Yule-Walker method, the Levison-Durbin algorithm, the Burg's algorithm, the Innovations algorithm, and the Hannan-Rissanen procedure are explored and used to analyze appropriate models and estimate model parameter values. The model validation is then performed via both the residual analysis and the portmanteau test. Finally, the time series models are used to predict section travel time on the corridor. Our study indicates that these models can predict section travel times with reasonable accuracy. The results presented in this paper can be easily modified and used in the study of short-term arterial travel time prediction for other urban areas.

REFERENCES

[1] H. Al-Deek, M. D'Angelo, and M. Wang, "Travel time prediction with non-linear time series," *Proceedings of the 5th ASCE International Conference on Applications of Advanced Technologies in Transportation*, 1998, pp. 317-324.

[2] L. Rilett and D. Park, "Direct forecasting of freeway corridor travel times using spectral basis neural networks," *Transportation Research Record* 1617, 1999, pp. 163-170.

[3] P. V. Palacharla and P. C. Nelson, "Application of fuzzy logic and neural networks for dynamic travel time estimation. *International Transactions in Operational Research*, Vol. 6, 1999, pp. 145-160.

[4] *Transportation Research Record 1457*, 1994, pp. 166-173.

[5] D. J. Dailey, "Travel-time estimation using cross-correlation techniques," *Transportation Research, Part B*, Vol. 27, 1993, pp. 97-107.

[6] I. Okutani and Y. J. Stephanedes, "Dynamic prediction of traffic volume through Kalman filtering theory," *Transportation Research, Part B*, Vol. 18, 1984, pp. 1-11.

[7] *Travel Time Data Collection Handbook*, Texas Transportation Institute & US Federal Highway Administration, FHWA-PL-98-035, March 1998.

[8] J. Blinn, "A trip down the graphics pipeline: line clipping," *IEEE Trans. Computer Graphics and Applications*, Vol. 11, 1999, pp. 98-105.

[9] G. E. Box, G. M. Jenkins, and G. C. Reinsel, *Time Series Analysis, Forecasting and Control*, 3rd ed., Prentice Hall, Upper Saddle River, NJ, 1994.

[10] P. J. Brockwell and R. A. Davis, *Introduction to Time Series and Forecasting*, 2nd ed., Springer, New York, NY, 2000.

[11] J.-S. Yang, "A Nonlinear State Space Approach to Arterial Travel Time Prediction," Final report (MN/RC-2006-05), C: 81655, WO: 130, Minnesota Department of Transportation, February 2006.

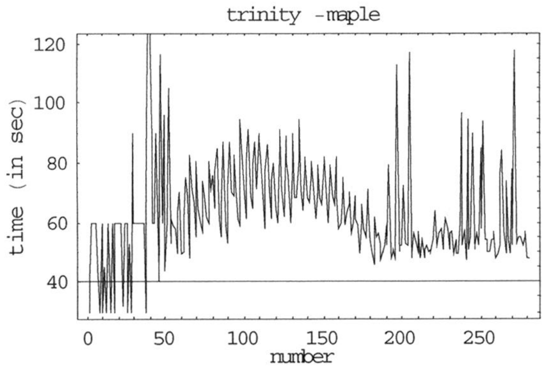

Fig. 1 Time plot of the Trinity-Maple Grove section travel time data

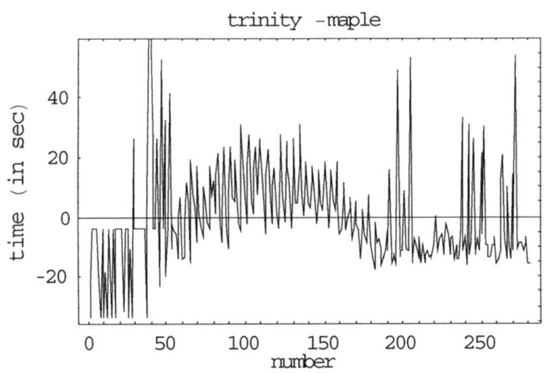

Fig. 2 Time plot of the section travel time data after removing its mean value

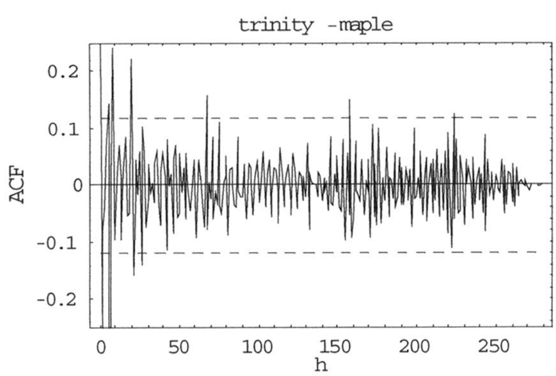

Fig. 3 The correlogram of the processed data

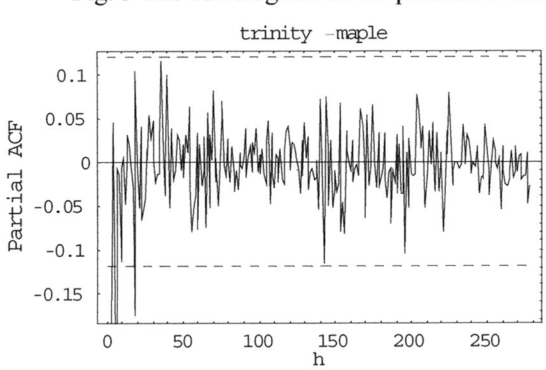

Fig. 4 The partial correlogram of the processed data

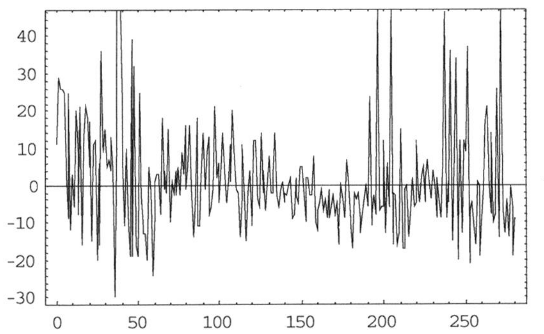

Fig. 5 The time plot of the residual time series

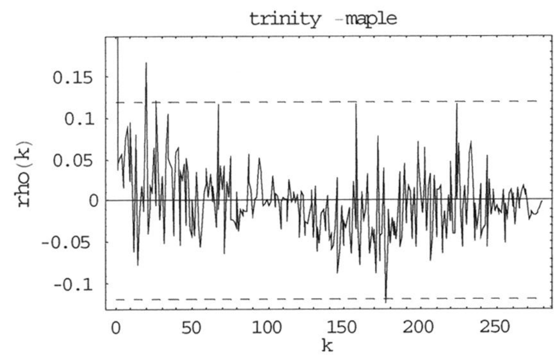

Fig. 6 The correlogram of the residual time series

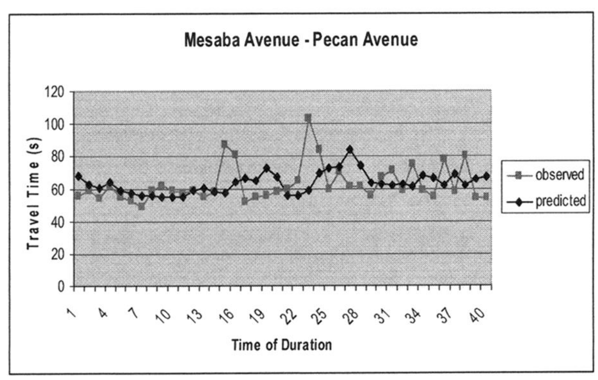

Fig. 7 Comparison of model predicted and observed travel time (section # 1).

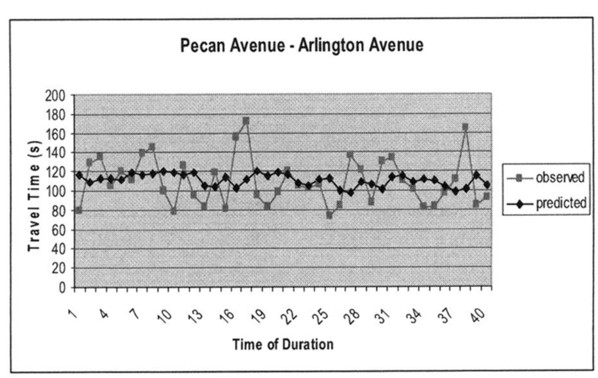

Fig. 8 Comparison of model predicted and observed travel time (section # 2).

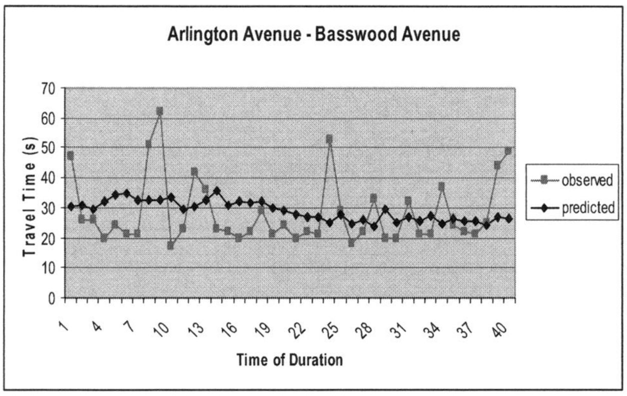

Fig. 9 Comparison of model predicted and observed travel time (section # 3).

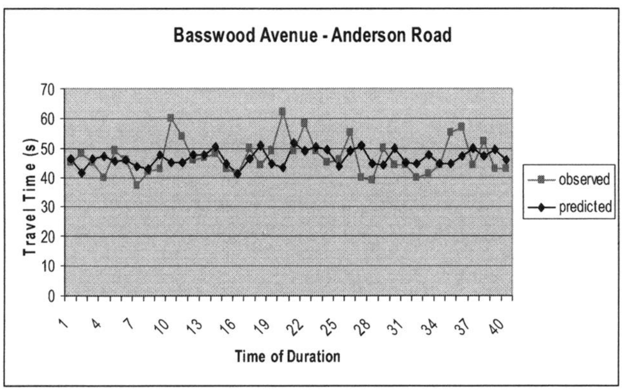

Fig. 10 Comparison of model predicted and observed travel time (section # 4).

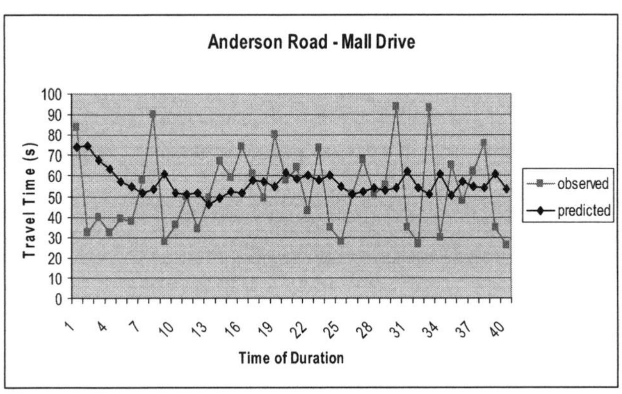

Fig. 11 Comparison of model predicted and observed travel time (section # 5).

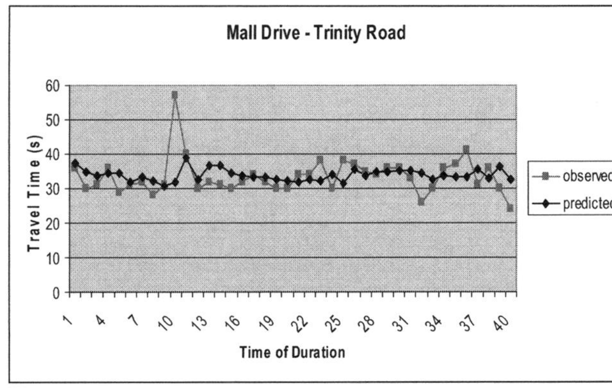

Fig. 12 Comparison of model predicted and observed travel time (section # 6).

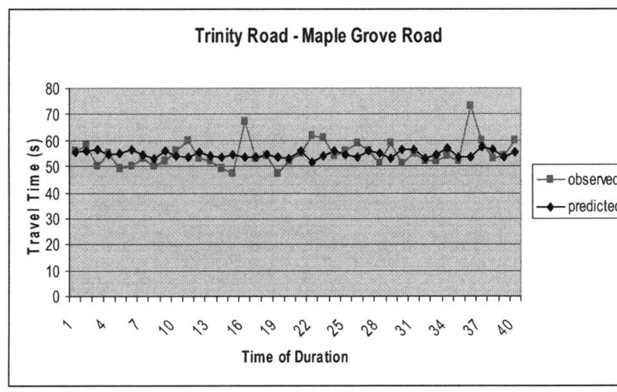

Fig. 13 Comparison of model predicted and observed travel time (section # 7).

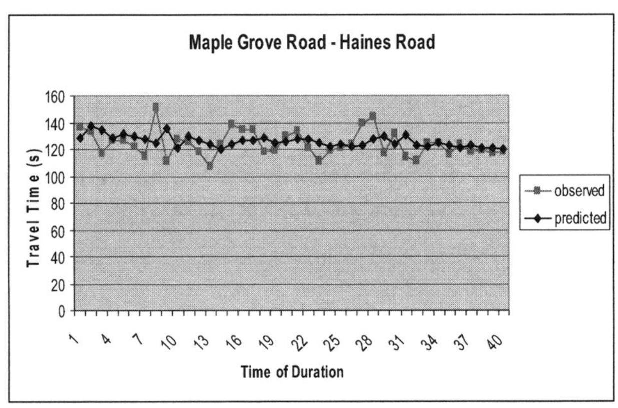

Fig. 14 Comparison of model predicted and observed travel time (section # 8).