Short-Term Prediction of Traffic Volume in Urban Arterials

By Mohammad M. Hamed,[1] Hashem R. Al-Masaeid,[2] and Zahi M. Bani Said[3]

(Reviewed by the Urban Transportation Division)

Abstract: This paper attempts to develop time-series models for forecasting traffic volume in urban arterials. The Box-Jenkins approach is used to estimate the time-series models. A 1-min data set representing traffic volume on five major urban arterials were available to estimate time-series models. The Box-Jenkins auto-regressive integrated moving average (ARIMA) model of order (0, 1, 1) turned out to be the most adequate model in reproducing all original time series. The developed model is easy to understand and implement. Further, the model is computationally tractable, and only requires the storage of the last forecasted error and current traffic observation.

INTRODUCTION

Urban traffic congestion is a daily occurrence in most active city centers worldwide. Urban arterials carry a significant portion of the daily traffic. As social and economic activities become more complex and larger in scale, the demand for transportation services and facilities increases rapidly. This has meant significant delays and unbearable congestion on both urban arterials and freeways. Therefore, it is not surprising to see a great deal of research on the management and control of the traffic on these transportation facilities.

The overall purpose of any management and control strategy is to provide smooth traffic flow in urban areas. Among the assumptions embedded in the transportation systems management (TSM) action, and identified by Jones and Sullivan (1978), is the notion that it is desirable to manage transportation demand as an alternative to increasing the supply of urban transportation facilities. In building up any management and control system, traffic conditions on urban arterials and freeways have to be converted to information-oriented data. This data can be used effectively to, if needed, develop short-term TSM actions.

Traffic characteristics such as traffic volume, speed, and density are of interest to the users of traffic-control systems. These variables are usually considered as measures of the performance of urban arterials. A very important requirement in nearly all transportation planning and design strategies is the knowledge of spatial and temporal distribution of traffic. In a traffic-control system of an arterial network, one wishes to predict future traffic characteristics in order to consider the proper management and control strategy. The implementation of an arterial traffic-control system requires a means to acquiring spatial and temporal distribution of traffic.

Forecasting methods are generally classified into qualitative or quantitative techniques. Forecasts under the quantitative methods are based on statistical models that can be either deterministic or probabilistic (stochastic). Quantitative forecasting can be applied when three conditions exist (1) availability of past information; (2) the ability to quantify historical data in a numerical form; and (3) the assumption that some aspects of the past pattern will continue into the future. An additional dimension to classifying quantitative forecasting methods is to consider the underlying model involved. There are two major types of forecasting models: time-series and causal models. In the first type, prediction of the future is based on past values of a variable and the past error term. The objective of these time-series models is to identify the pattern in the historical data and extrapolate that pattern into the future.

There are a great deal of studies primarily concerned with the forecasting of traffic volume. Different methodologies and techniques have been used for this purpose, particularly in the last decade. These include the Kalman filtering models (Okutani and Stephanedes 1984); prediction error minimization and maximum likelihood models (Nihan and Davis 1989), time-series models (Ahmed and Cook 1979; Kyte et al., unpublished paper, 1989), and spectral models (Nicholson and Swann 1979). Davis et al. (1990) and Jian (1990) used the adaptive prediction system to predict freeway traffic congestion and hourly traffic flow. In Jian (1990), the developed adaptive prediction system was applied to real traffic-flow data collected from a highway network. However, the counting interval was 1 hr, which is certainly a very large sampling interval if different management and control strategies were to be designed and implemented.

The availability of automatic presence detectors and other technologically advanced instruments are effective tools for providing traffic-management centers with huge amounts of traffic-flow data at a relatively adequate cost. Since time-series analysis gains more credibility with a larger amount of data, the analysis is a cost-effective tool for predicting future traffic volume.

With the limited financial and technical resources in developing countries, it becomes necessary to develop models capable of supplying a quick and accurate short-term prediction of traffic volume. Simultaneously, these models should be easy to implement. Therefore, the objective of this paper is to develop inexpensive and computationally tractable time-series models capable of predicting future traffic-volume values on urban arterials. The developed models are expected to be capable of predicting traffic volume 1 min ahead of time.

Model Formulation

Consider the function $Z_i(t)$, which represents the historical performance of traffic flow on a given arterial i; the idea is to be able to predict the future value of $Z_i(t)$, given historical information on Z_i. The time-ordered sequence of observa-

[1]Asst. Prof., Dept. of Civ. Engrg., Jordan Univ. of Sci. and Technol., Irbid P.O. Box 3030, Jordan.

[2]Asst. Prof., Dept. of Civ. Engrg., Jordan Univ. of Sci. and Technol., Irbid P.O. Box 3030, Jordan.

[3]Grad. Student, Dept. of Civ. Engrg., Jordan Univ. of Sci. and Technol., Irbid P.O. Box 3030, Jordan.

Note. Discussion open until November 1, 1995. To extend the closing date one month, a written request must be filed with the ASCE Manager of Journals. The manuscript for this paper was submitted for review and possible publication on October 20, 1993. This paper is part of the ***Journal of Transportation Engineering***, Vol. 121, No. 3, May/June, 1995. ©ASCE, ISSN 0733-947X/95/0003-0249–0254/$2.00 + $.25 per page. Paper No. 7198.

J. Transp. Eng. 1995.121:249-254.

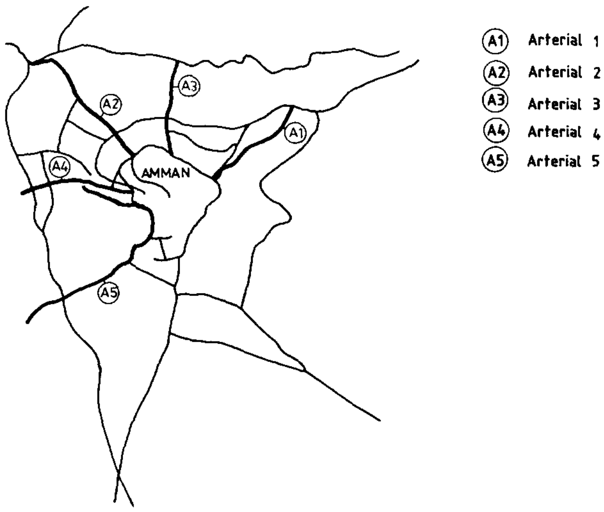

FIG. 1. Location of Study Sites

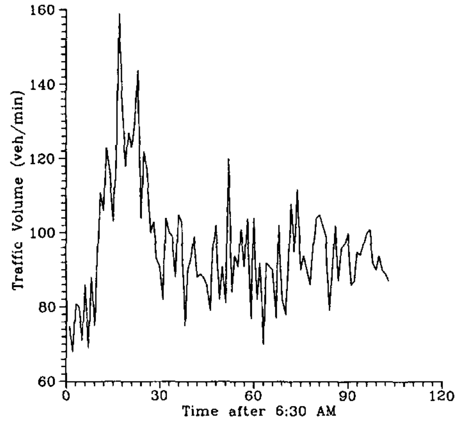

FIG. 2. Observed Traffic Volume on Arterial 1

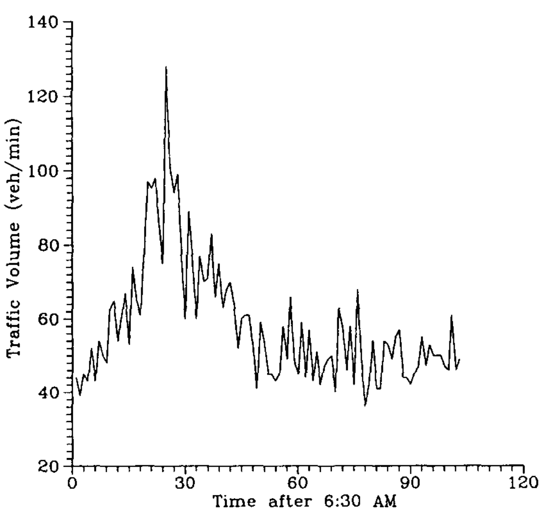

FIG. 3. Observed Traffic Volume on Arterial 2

tions of traffic flow $Z_{i1}, Z_{i2}, Z_{i3}, \ldots, Z_{in}$ are considered realizations of the random variables $Z_{i1}, Z_{i2}, z_{i3}, \ldots, Z_{in}$. These variables have characteristics of dependency. As such, only historical values (time-series data) of traffic flow on the arterial will be available to carry out the prediction of future values. Depending on the structure of the collected data, a general mathematical model to predict future values can take the following form:

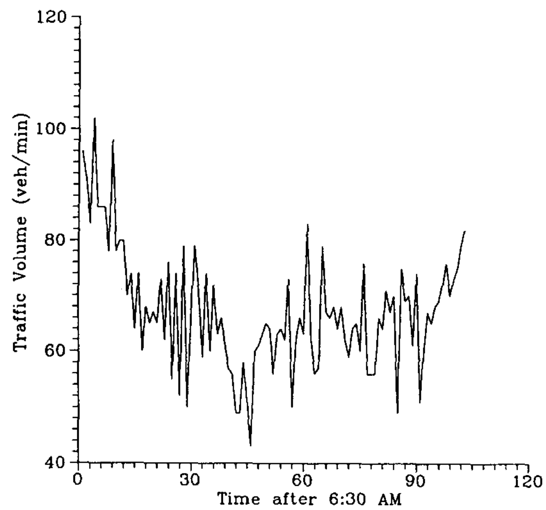

FIG. 4. Observed Traffic Volume on Arterial 3

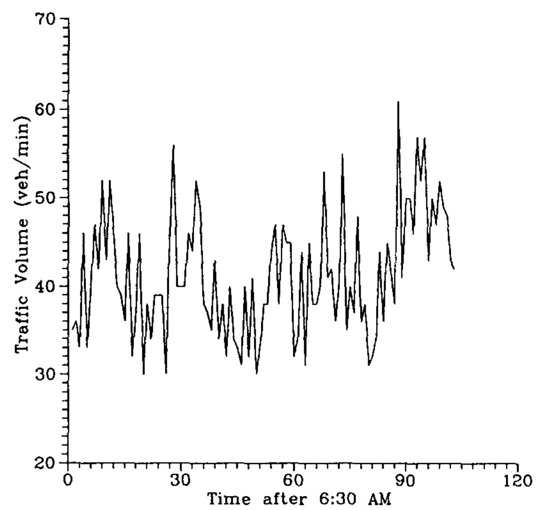

FIG. 5. Observed Traffic Volume on Arterial 4

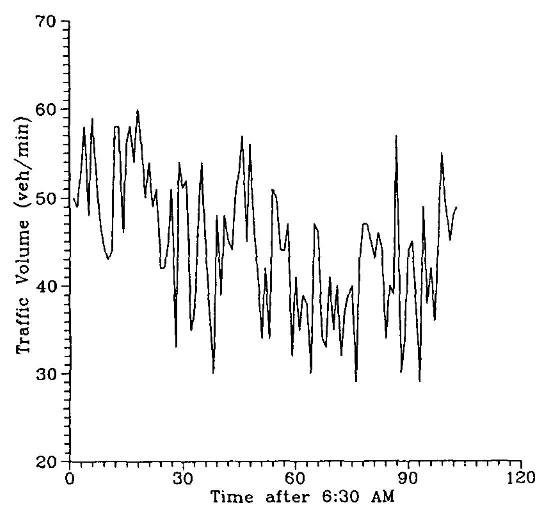

FIG. 6. Observed Traffic Volume on Arterial 5

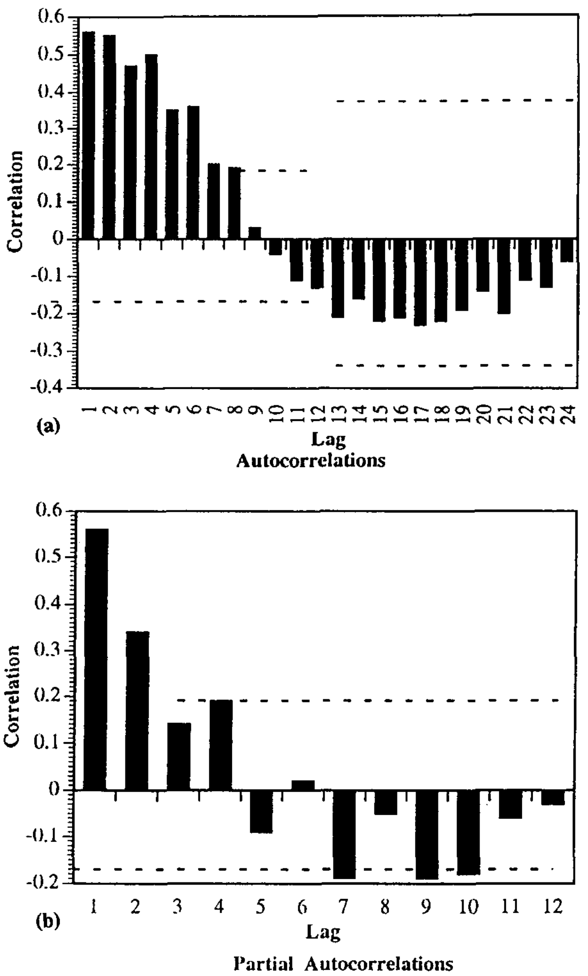

FIG. 7. Autocorrelation and Partial Autocorrelations of Traffic-Volume Series for Arterial 1

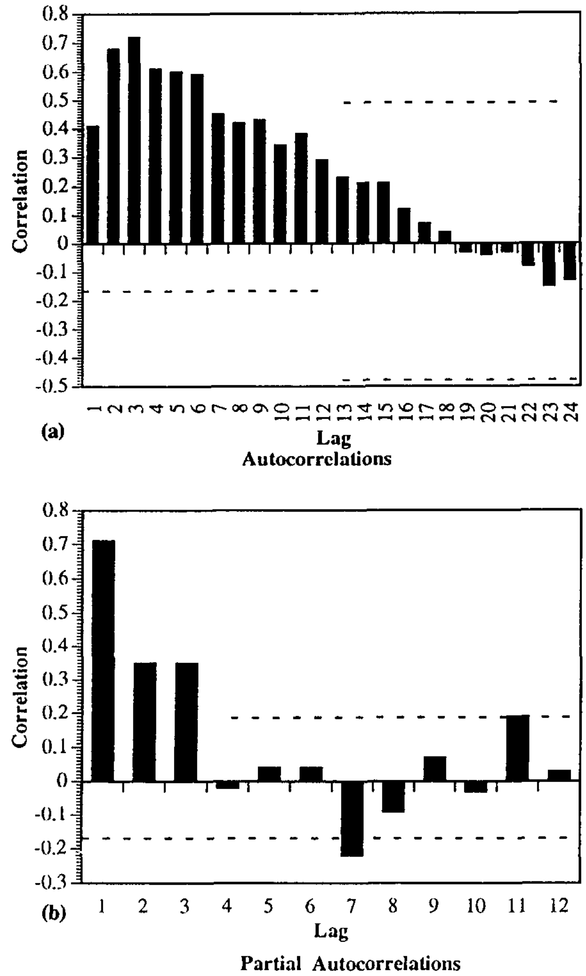

FIG. 8. Autocorrelations and Partial Autocorrelations of Traffic-Volume Series for Arterial 2

$$Z_i(t) = \beta_0 + \beta_1 Z_i(t-1) + \beta_2 Z_i(t-2) + \cdots + \beta_n Z_i(t-n) + \xi_i \tag{1}$$

where $Z_i(t)$ = future value of traffic volume at time of day t for the ith arterial; β = estimable coefficients; and ξ_i = error term assumed to be normally distributed.

The structure and pattern of the time series should be identified in order to select an appropriate smoothing method. One of the main characteristics of time-series data, important to the analyst, is the stationarity of the time series. Let Z_t represent a nonseasonal time series of traffic volume observations taken at equally spaced time intervals. Now, Z_t is either stationary or reducible to a stationary form (M_t) by computing the difference for some integer number of times D, such that

$$M_t = (1 - B)DZ_t \tag{2}$$

where B = back shift operator defined as $BZ_t = Z_{t-1}$. Mathematically, a stationary time series is one in which the probability distribution of any $(n + 1)$ observation $(M_t, \ldots, M_{t-n})$ is invariant with respect to t. As such, any set of traffic-volume observation from a stationary series will have the same mean value.

A number of approaches to model time-series data exist in the literature. One commonly used technique is the Box-Jenkins approach. Box and Jenkins (1976) incorporated the relevant information required to understand and use univariate time-series autoregressive integrated moving average (ARIMA) models. Their approach consists mainly of three phases: (1) identification; (2) estimation; and (3) application (forecasting). In fact, the Box and Jenkins (1976) technique has given accurate future forecasts [see Nihan and Holmesland (1980) and Ahmed (1989)]. Ahmed's study (1989) showed that the ARIMA (0, 1, 3) model yielded the minimum mean-square error forecasts (traffic volume and density) when compared with the double exponential model, the moving-average model, and the adaptive-exponential model.

In this paper, the Box-Jenkins approach will be used to develop a forecasting model based on time-series traffic volume, through the use of 1-min data. A total of five time series, representing more than 500 min of traffic-volume observations, will be used in the development and evaluation of forecasting models.

The general model for representing a wide class of nonstationary time series is the ARIMA (P, D, Q), which can be expressed as

$$\phi_P(B)(1 - B)^D(Z_t - \mu) = \theta_Q(B)e_t \tag{3}$$

where $\phi_P(B)$ = autoregressive operator of order P; $\theta_Q(B)$ = moving-average operator of order Q, $(1 - B)^D$ = Dth differencing of the original time series; μ = mean of the original time series; e_t = random error assumed to be normally distributed with zero-mean and known variance; B = back-shift operator; and ϕ and θ = estimable coefficients.

The ARIMA models are fitted to a particular data set by a three-stage iterative process: identification, estimation, and diagnostic checking. In the first step, the values of P, D, and Q are determined by inspecting both the autocorrelation and

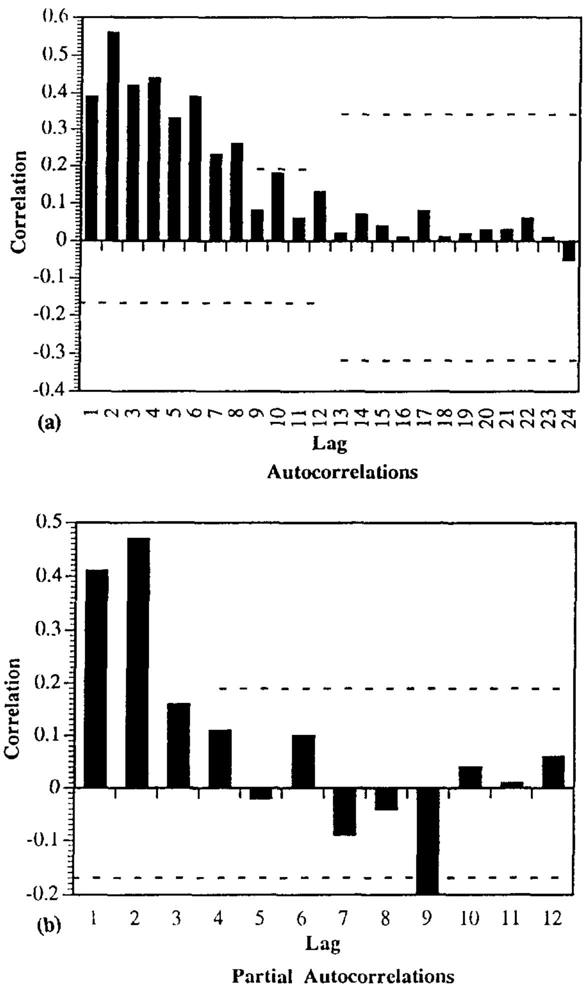

FIG. 9. Autocorrelations and Partial Autocorrelations of Traffic-Volume Series for Arterial 3

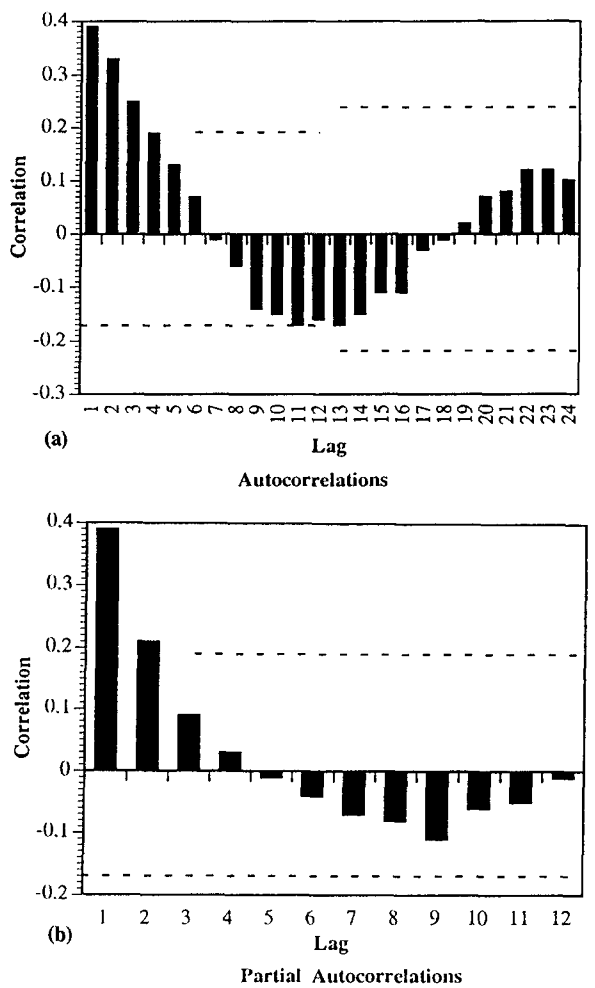

FIG. 10. Autocorrelations and Partial Autocorrelations of Traffic-Volume Series for Arterial 4

partial-autocorrelation functions of the original time series. Once the values of P, D, and Q have been determined, the autoregressive and moving-average parameters are estimated using the nonlinear least-squares technique. If the form of the chosen time-series model is satisfactory, the resulting residuals should be uncorrelated random deviations.

The aforementioned general ARIMA (P, D, Q) model does not address the seasonality part. However, in the short-term forecasting of traffic volume, the seasonality does not need to be considered for obvious reasons.

Empirical Settings

Five urban arterials in Amman, Jordan's capital, were chosen for data collection. A 1-min traffic-volume data on each arterial were collected during the peak morning period (6:30 to 8:15 a.m.). Fig. 1 shows the locations of these arterials. These arterials carry a large proportion of the daily traffic in Amman. In fact, they are considered the main entrance to the city (Bani Said 1992). Figs. 2–6 represent traffic-volume counts on the selected urban arterials. Generally, the traffic-volume plots show nonstationarity in the mean (the mean traffic volume changes with time). Further evidence of the nonstationarity can be clearly seen in the autocorrelation and partial-correlation plots (discussed in later sections). As such, regular differencing is suggested to achieve stationarity.

ESTIMATION RESULTS

The autocorrelation and partial-autocorrelation plots for all arterials indicate the nonstationarity in the mean (see Figs. 7–11). Subjecting all original time series to a regular differencing of order one ($D = 1$) was enough to cause nonstationarity. No higher order differencing was needed. Generally, autocorrelation of the differenced time series showed a significant value at lag one but the rest were not significantly different from zero, which indicates that the time series of first difference transformed the original traffic-volume data into a stationary form.

Several ARIMA models were evaluated for each arterial. The ARIMA model of order (0, 1, 1) was found to be the most statistically significant for forecasting traffic volume. As such, traffic observations can be represented as

$$(1 - B)(Z_t - \mu) = (1 - \theta)e_t, \quad |\theta| \leq 1 \qquad (4)$$

Table 1 shows the time-series parameter estimates. All parameter estimates were significantly different from zero at the 0.05 level of significance. Diagnostic checking was carried out for each model by inspecting the mean and the autocorrelation of the residual series. It turned out that the values of the autocorrelation are not significant at all lags and showed no discernible pattern. This indicates that the residuals are uncorrelated. Further, the Q test values for each estimated time-series model were compared with the tabulated chi-square values. Comparisons indicate that the residuals are white noise at the 0.05 level of significance. Further, there was no evidence of cross correlation between the residuals and the differenced time series. The assumption of independence between the differenced series and e_t was satisfied. Thus, one

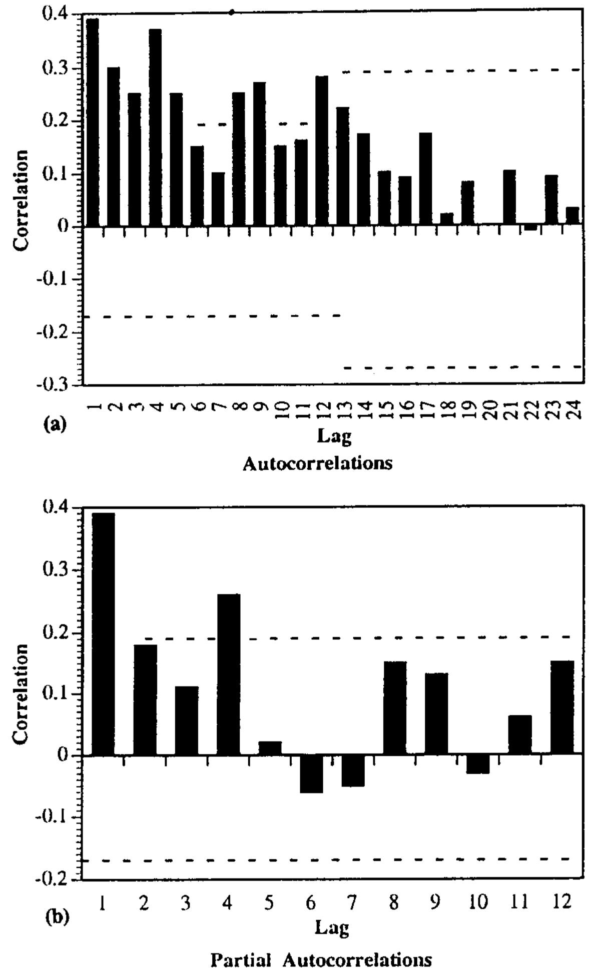

FIG. 11. Autocorrelations and Partial Autocorrelations of Traffic-Volume Series for Arterial 5

TABLE 1. Summary of Box-Jenkins ARIMA Model Estimates

Parameter type (1)	Order of parameter (2)	Degree of differencing (3)	Coefficient (4)	T-test (5)
		(*a*) Arterial 1		
Moving average	1	1	0.60	7.19
		(*b*) Arterial 2		
Moving average	1	1	0.60	7.62
		(*c*) Arterial 3		
Moving average	1	1	0.74	11.12
		(*d*) Arterial 4		
Moving average	1	1	0.75	11.89
		(*e*) Arterial 5		
Moving average	1	1	0.84	15.46

might conclude that the ARIMA model (0, 1, 1) is an adequate model.

To make the aforementioned estimated model operational, let $\tilde{Z}_{t-1}(1)$ be the predictor of Z_t up to time period $(t-1)$. This predictor can be written as

$$\tilde{Z}_{t-1}(1) = Z_{t-1} - \theta\delta_{t-1} \tag{5}$$

with the forecasted error at time period $(t-1)$ given by

$$\delta_{t-1}(1) = Z_t - \tilde{Z}_{t-1}(1) \tag{6}$$

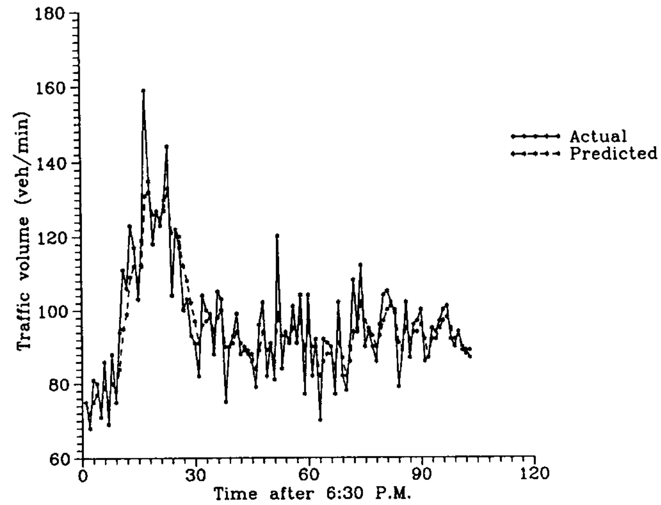

FIG. 12. Actual and Predicted Traffic Volume for Arterial 1

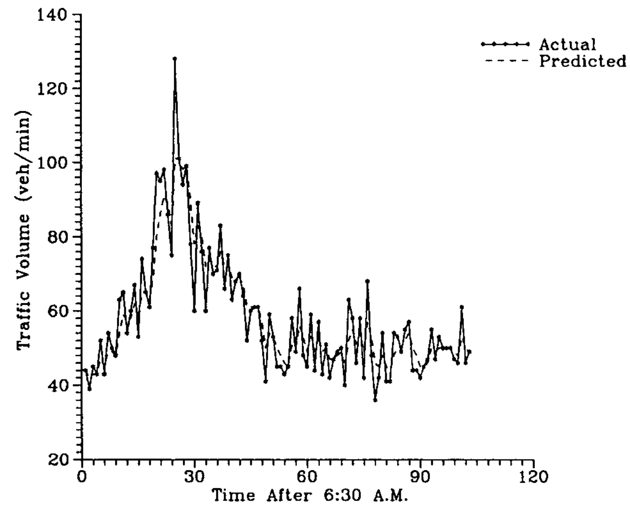

FIG. 13. Actual and Predicted Traffic Volume for Arterial 2

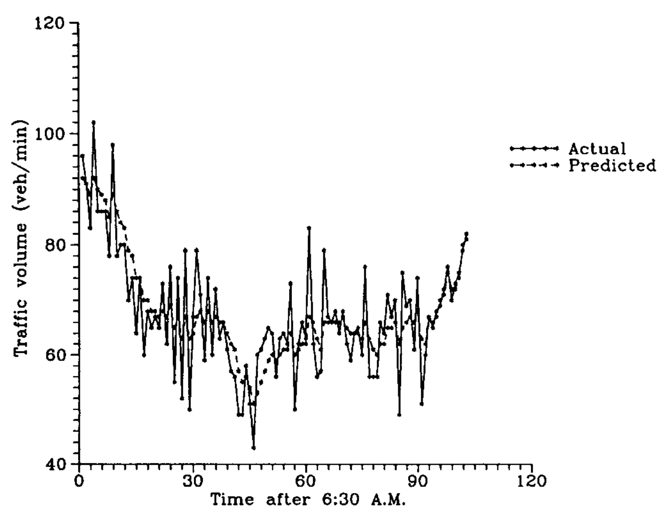

FIG. 14. Actual and Predicted Traffic Volume for Arterial 3

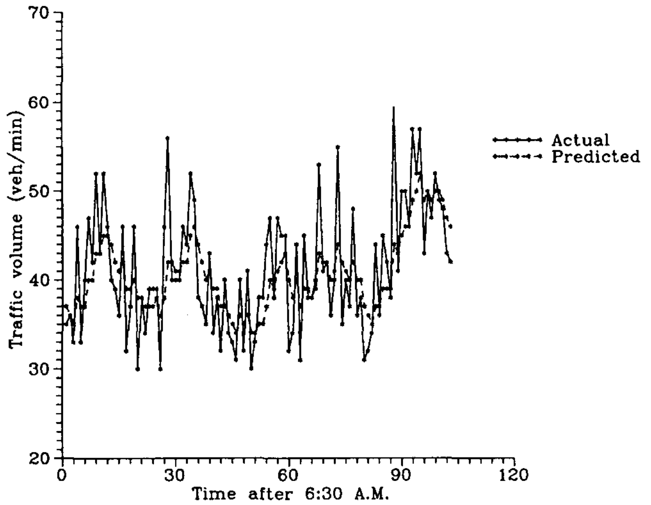

FIG. 15. Actual and Predicted Traffic Volume for Arterial 4

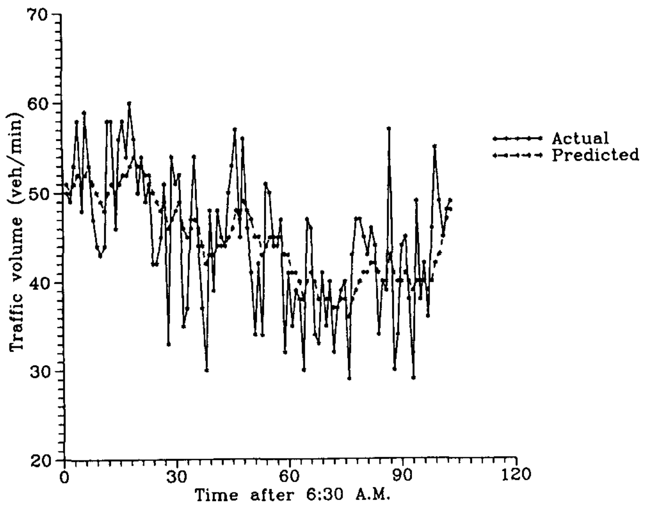

FIG. 16. Actual and Predicted Traffic Volume for Arterial 5

The forecasted value for Z_{t+1} at time period t can then be written as

$$\bar{Z}_t(1) = Z_t - \theta\delta_{t-1}(1) \quad (7)$$

As such, (7) requires the storage of the last forecasted error and the current traffic-flow observations.

Figs. 12–16 represent the actual and predicted traffic volume for the five selected urban arterials. Generally speaking, the ARIMA model (0, 1, 1) seems to adequately reproduce the observed traffic volume in all arterials.

CONCLUDING REMARKS

The main objective of this paper was to develop time-series models for the short-term prediction of traffic flow in urban arterials. A 1-min data set representing traffic volume on five selected arterials was available to develop the time-series models. The five arterials were located in Amman, Jordan's capital. These arterials were selected on the basis that they carry a large proportion of the daily traffic volume and are also considered the main entrances to the capital city.

The Box-Jenkins (1976) approach was used in this paper to develop time-series models. Each arterial's traffic volume was analyzed by the Box-Jenkins technique to obtain a model suitable for the short-term prediction of traffic volume. The ARIMA model (0, 1, 1) turned out to be the most adequate model capable of representing the actual traffic volume on all urban arterials. The model can be easily implemented, is computationally tractable, and only requires the storage of the last forecasted error and current traffic observation.

ACKNOWLEDGMENTS

The writers wish to acknowledge the Department of Civil Engineering at Jordan University of Science and Technology for making their facilities available to them. Appreciation is also extended to anonymous reviewers for their instructive and helpful comments.

APPENDIX. REFERENCES

Ahmed, S., and Cook, A. (1982). "Application of time-series analysis techniques to freeway incident detection." *Transp. Res. Rec. 841*, Transp. Res. Board, Washington, D.C., 19–21.

Ahmed, S. (1989). "Operational filtering of traffic data." *J. of Forecasting*, 8(1), 19–32.

Bani Said, Z. M. (1992). "Short-term forecasting of traffic volume in Amman's Urban Arterials," ME thesis, Dept. of Civ. Engrg., Jordan Univ. of Sci. and Technol., Irbid P.O. Box 3030, Jordan.

Box, G., and Jenkins, G. (1976). *Time-series analysis: forecasting and control*. Holden-Day, San Francisco, Calif.

Davis, A. G., Nihan, N. L., Hamed, M. M., and Jacobson, L. N. (1990). "Adaptive forecasting of freeway traffic congestion." *Transp. Res. Rec. 1287*, Transp. Res. Board, Washington, D.C., 29–33.

Jian, L. (1990). "Prediction of traffic flow by an adaptive prediction system." *Transp. Res. Rec. 1287*, Transp. Res. Board, Washington, D.C., 54–61.

Jones, D. W., and Sullivan, F. C. (1978). "TSM: tinkering superficially at the margin?" *J. Transp. Engrg.*, ASCE, Vol. 104, 817–834.

Nicholson, H., and Swann, C. D. (1974). "The prediction of traffic flow volumes based on spectral analysis." *Transp. Res.*, Vol. 8, 533–538.

Nihan, N. L., and Davis, G. A. (1989). "Application of prediction-error minimization and maximum likelihood to estimate intersection O-D matrices from traffic counts." *Transp. Sci.*, 23(2), 77–90.

Nihan, N. L., and Holmesland, K. O. (1980). "Use of the Box Jenkins time series techniques in traffic forecasting." *Transp.*, Vol. 9, 125–143.

Okutani, I., and Stephanedes, J. (1984). "Dynamic prediction of traffic volume through Kalman filtering theory." *Transp. Res.*, 18B(1), 1–11.

Stephanedes, Y. J., Michalopoulos, P., and Plum, R. (1981). "Improved estimation of traffic flow for real-time control." *Transp. Res. Rec. 795*, Transp. Res. Board, Washington, D.C., 28–39.

J. Transp. Eng. 1995.121:249-254.