PERGAMON

Transportation Research Part C 11 (2003) 121–135

TRANSPORTATION
RESEARCH
PART C

www.elsevier.com/locate/trc

A multivariate state space approach for urban traffic flow modeling and prediction

Anthony Stathopoulos *, Matthew G. Karlaftis

Department of Transportation Planning and Engineering, School of Civil Engineering, National Technical University of Athens, 5, Iroon Polytechniou Str., Zografou Campus, Athens 157 73, Greece

Received 5 April 2001; accepted 29 April 2002

Abstract

Urban traffic congestion is one of the most severe problems of everyday life in Metropolitan areas. In an effort to deal with this problem, intelligent transportation systems (ITS) technologies have concentrated in recent years on dealing with urban congestion. One of the most critical aspects of ITS success is the provision of accurate real-time information and short-term predictions of traffic parameters such as traffic volumes, travel speeds and occupancies. The present paper concentrates on developing flexible and explicitly multivariate time-series state space models using core urban area loop detector data. Using 3-min volume measurements from urban arterial streets near downtown Athens, models were developed that feed on data from upstream detectors to improve on the predictions of downstream locations. The results clearly suggest that different model specifications are appropriate for different time periods of the day. Further, it also appears that the use of multivariate state space models improves on the prediction accuracy over univariate time series ones.

Keywords: Traffic flow; Multivariate time series; Short-term predictions

1. Introduction

Urban traffic congestion is a problem that adversely and significantly affects all aspects of the quality of life in, particularly larger, cities. As such, it comes as little surprise that intelligent transportation systems (ITS) technologies make a conscious effort toward dealing with congestion in urban areas. Advanced traffic control technologies may lead to more efficient use of existing

* Corresponding author. Tel.: +30-10-7721288; fax: +30-10-7722404.
E-mail addresses: astath@transport.ntua.gr (A. Stathopoulos), mgk@central.ntua.gr (M.G. Karlaftis).

doi:10.1016/S0968-090X(03)00004-4

road network systems resulting in reduced traffic congestion, delays, emissions, energy consumption and improved safety. The success of these strategies depends, to a large extent, on the quality and accuracy of the information given to the drivers. ITS technologies, with advanced traveler information systems (ATIS), travel advisories, variable message signs (VMS) and others, attempt to relieve congestion and decrease travel time by assisting drivers on selecting routes, departure times, and even mode of travel.

For the above selection to be effective, accurate real-time information and *short-term predictions* of traffic parameters such as traffic volumes, travel speeds, occupancies, etc., are needed. It is also important to note that the short-term prediction of traffic parameters, occupancy in particular, could be used to predict travel times along pre-specified corridors; accurate provision of predicted travel time information could be of paramount importance to commuters' travel mode and route selection and to the success of ITS systems in general. Further, a key component of the daily planning and operation of an urban area traffic control center is the optimal regulation of road traffic on all types of streets; for this task, short-term forecasting of traffic parameters is essential.

The need for, to the extent possible, accurate traffic parameter prediction has long been recognized in the international scientific literature. Since the early 1980s, researchers have used an extensive variety of specifications to model traffic characteristics and produce short-term forecasts (a review of many of the earlier methods used in short-term traffic flow prediction have been reviewed in Van Arem et al., 1997). Specifications have ranged from Kalman filtering (Okutani and Stephanedes, 1984; Whittaker et al., 1997), exponential filtering (Ross, 1982), nonparametric statistical methods (Davis and Nihan, 1991; Smith et al., 2000), spectral and cross-spectral analyses (Stathopoulos and Karlaftis, 2001a,b) and sequential learning (Chen and Grant-Muller, 2001), to cusp catastrophe theory (Pushkar et al., 1995). Nevertheless, a large amount of literature has been concerned with predictions from pure time-series models ranging from ARIMA to dynamic generalized linear models (Chang and Miaou, 1999; Hamed et al., 1995; Lee and Fambro, 1999; Lingras et al., 2000; Williams, 2001), and neural networks (Chang et al., 2000; Fu and Rilett, 2000; Park et al., 1999; Park and Rilett, 1998; Rilett and Park, 1999; Smith and Demetsky, 1994; Van Der Voort et al., 1996; Yasdi, 1999; Zhang, 2000; Zhang et al., 2000; Dia, 2001). Interestingly, Smith and Demetsky (1997) in a comparison of historical average, time-series, neural network, and nonparametric regression models found that the nonparametric regression model significantly outperformed the other models and was also easier to implement.

With few exceptions, previous work has been characterized by two important issues; first, data used for model calibration and prediction has been, mostly, from (urban) *freeways* and, second, models have been univariate in nature in that they use historical data from a given location to predict its future behavior. But, with ITS's potential applications and benefits in *urban arterials and streets*, models to forecast short-term traffic in these conditions are needed. Further, the setup of urban traffic data collection systems (regardless of whether the system is based on loops, cameras, microwave systems or a combination) most frequently allows for the exploitation of multivariate time-series techniques that may provide increased prediction accuracy.

From the discussion above it becomes clear that there is a need, both a practical and a theoretical one, to combine urban street data with multivariate techniques and examine the potential for accurate flow predictions in the context of urban streets. As such, the goal of this paper is to model and predict flow at an *urban signalized arterial*. To achieve this, data from an urban arterial street near the center of Athens, Greece, are used; these data are modeled using a flexible and explicitly multivariate time-series methodology. The remainder of this paper is organized as follows. The Section 2 is a brief overview of the state space approach to time series modeling that explicitly allows for jointly considering data from different detectors. Section 3 describes the data collection process used in this study and Section 4 describes the model estimation process. Finally, Section 5 summarizes the findings of the paper.

2. The state space approach to time series modeling

The historical background of the state space methodology is a long one. Its roots reach back to the development of the Wiener filter and engineering state vector models. The early period of state space models culminated with the invention of the Kalman filter by Kalman (1960). State space models apply to multiple input–multiple output systems of stochastic variables which, unlike other time series techniques such as ARIMA, do *not* require the variables to be stationary. Further, as Harvey (1984) showed, using the state space estimation method produces equations of nearly the same form for both stationary and nonstationary data.

Before proceeding with the theory, it is important to note that the "state space model" and the more widely known "Kalman Filter model" refer to the same basic underlying theory. Generally, the term state space refers to the model and the term Kalman filter refers to the estimation of the state, as will be shown in the next paragraph. There are also few differences between ARMAX models and the state space model. Most times, the process of estimating a state space models begins by estimating its ARMAX equivalent in order to capture the statistically significant AR and MA dimensions. Nevertheless, the state space approach has advantages over the more widely used family of ARMA (ARIMA, ARMAX, etc.) models as Durbin (2000) notes: (i) it is based on the analysis of the model structure and as such the various components that make up the series (trend, seasonal variation, cycle, etc.) together with the effects of explanatory variables and interventions are modeled separately before being integrated in the final model, (ii) it is very flexible in that changes in the structure of the modeled system over time can be explicitly modeled, (iii) multivariate observations can be handled as straightforward extensions of univariate theory unlike ARIMA models, (iv) it is rather simple to deal with missing observations, (v) the Markovian nature of state space models enables increasingly large models to be handled effectively without disproportionate increases in computational efficiency and (vi) it is straightforward to predict future values by projecting the Kalman filter forward into the future and obtaining their estimated values and standard errors.

On the other hand, the mathematical theory of state space is quite complex. To understand it fully on a mathematical level requires a good background in probability theory, matrix theory, multivariate statistics and the theory of Hilbert space. But, the state space model is so useful, especially for multivariate time series processes, that it deserves wider attention in the transportation literature. Here we note that much of the theory behind state space modeling is beyond the scope of this paper and, as such, only the very essential parts of the formulation are discussed; nevertheless, readers interested in the finer mathematical details of state space modeling should refer to Shumway and Stoffer (2000), Durbin (2000), and Whittaker et al. (1997) for a transportation application.

The concept of the *state* of a system emerged from physical science. There the state consists of a vector that includes all the information about the system that carries over into the future; that is, the state vector consists of a linearly independent set of linear combinations from the past that are correlated with future endogenous variables. To construct such a vector, the simplest and most straightforward way is through the one-step predictor, then the two-step predictor, and so forth, so long as the predictors extracted in this way are linearly independent. The set of predictors extracted in this way are, by construction, a state vector. They contain all the relevant information from the past, since all predictors of the future, for any time horizon, are linearly dependent upon them. It is also a minimal set, since they are linearly independent. The *space* spanned by the selected predictors is called the *predictor space*. The state space model for an r-variate (multivariate) time series $\{y_t,\ t = 1,\dots,n\}$ is written

$$\begin{aligned} y_t &= M_t x_t + D z_t + u_t, \\ x_t &= \Phi x_{t-1} + w_t. \end{aligned} \tag{1}$$

These two equations are called the *observation equation* and the *state equation* respectively. x_t, a p-vector, is the (observed) state of the process at time t. The evolution of x_t is governed by the state equation: the state at time t evolves by autoregression on the previous state and is further modified by the innovation term w_t. The initial state x_0 is taken to have mean μ and nonsingular covariance matrix Σ. The sequence $\{w_t,\ t = 1,\dots,n\}$ consists of independent and identically distributed random p-vectors. The first s elements of w_t have mean zero and nonsingular covariance matrix Q; the last $(p-s)$ elements of w_t are identically zero. The unobserved (latent) data vector, y_t, is a possibly time-dependent linear transformation of x_t with the addition of a regression on the k-vector z_t of known exogenous regressors and observational noise modeled by u_t. The sequence $\{u_t,\ t = 1,\dots,n\}$ consists of independent and identically distributed random r-vectors. The first m elements of u_t have mean zero and nonsingular covariance matrix R; the last $(r-m)$ elements of u_t are identically zero. Furthermore, u_t and w_t are mutually independent of each other and of x_0 for all t and t'.

Assuming that the random vectors x_0, u_t and w_t have multivariate normal distributions, the log-likelihood, given $x_0, x_1,\dots,x_n$, $y_1,\dots,y_n$, $z_1,\dots,z_n$, is, up to a constant,

$$\begin{aligned} \log L = {} & -\frac{1}{2}\log|\Sigma| - \frac{1}{2}(x_0-\mu)^{\mathrm{T}}\Sigma^{-1}(x_0-\mu) - \frac{1}{2}n\log|Q| \\ & -\frac{1}{2}\sum_{t=1}^{n}(x_t-\Phi x_{t-1})^{\mathrm{T}}P^{\mathrm{T}}Q^{-1}P(x_t-\Phi x_{t-1}) - \frac{1}{2}n\log|R| \\ & -\frac{1}{2}\sum_{t=1}^{n}(y_t-M_t x_t-Dz_t)^{\mathrm{T}}S^{\mathrm{T}}R^{-1}S(y_t-Mx_t-Dz_t). \end{aligned} \tag{2}$$

The objective of the state space process is to find the set of μ, Φ, D, Q and R that maximize $\log L$. The essential parameters of the traffic model to be developed consists of the following components: an observed state vector x_t that holds one variable per loop examined (traffic flow) and is ordered by corresponding loop number, starting from L108 and moving to L12 (Fig. 1), and an unobserved data vector, y_t, that is a time-dependent linear transformation of x_t with the addition

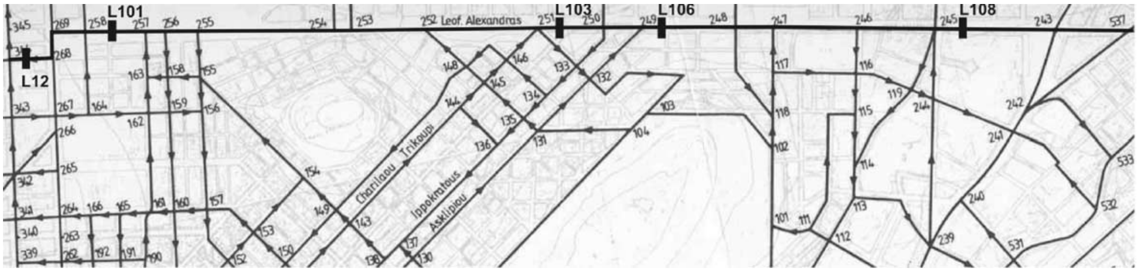

Fig. 1. Layout of loop detector placement on the urban signalized arterial under study.

of a vector z_t of known exogenous regressors (such as roadway capacity, examined section length, signalization plans, weather, etc.). [1] It should be noted that in the context of this study, the term multivariate is used from a spatial standpoint (multiple detector locations) as opposed to a traffic one (such as volume, speed and occupancy from a single detector).

Finally, the assumption of multivariate normality for the random vectors x_0, u_t and w_t of Eq. (1) was examined for the models developed. Of course, it should be noted that the investigation of multivariate normality is not as straightforward as assessing univariate normality. It has proved very difficult to construct a test for overall test of joint normality in more than two dimensions because of the large number of things that can go wrong (Johnson and Wichern, 1992). One thing that can be done is to check for the normality of each variable distribution. [2] In general, we could not reject the null hypothesis of univariate normality for most of the variables. Having satisfied multiple univariate normality, we proceeded as though multivariate normality conditions are met.

3. The collection of urban traffic data

3.1. Data source and analysis

Providing users with accurate information regarding travel time is considered to be an important element for travelers' route choice decisions, and can lead to an improvement in network efficiency. Toward meeting the goal of reducing congestion, decreasing travel times, and giving drivers the ability to make better route and departure time choices, a dynamic traffic map and ATIS for the central Athens area has been operating since 1996 (NTUA, 1996). As a first attempt

[1] It is important to note that while in this application we only consider flow as the traffic parameter to be examined, the state space models allows, by straightforward extension, to include more characteristics in the unobserved state vector (such as occupancy, speed, etc.).

[2] This is a necessary but not sufficient test for multivariate normality. In general, we recognize that marginal univariate normality is not sufficient for joint normality. Nevertheless, as Stevens (1992) notes, for most practical work one-dimensional investigations of normality are ordinarily sufficient. Further, data sets that are normal in lower dimensional representations but non-normal in higher dimensions are very infrequent in practice (Johnson and Wichern, 1992).

to meet driver information needs, travel time information is given over the internet updated every 15 min. To do this, six origin points have been defined at the major city entrances, from which a total of 17 possible (commonly used) routes have been considered. Speeds and travel times along these 17 routes are being estimated using *flow* (*volume*) and *occupancy* data collected directly from the controllers installed at 144 locations of the down-town urban network.

3.2. The data analyzed

In an effort to reduce the size of the data-set to be processed without losing much essential information, a new data-set was created containing flow measurements for the period between the 3rd of January 2000 and 31st of May, 2000. From previous analyses it has been shown that while flows do not vary between weekdays, they vary significantly between weekdays and weekends and for different time periods of the day (Stathopoulos and Karlaftis, 2001a). As such, Saturday and Sunday data were excluded from further analysis; this yielded a data set for 106 days containing 50 880 3-min flow observations to be used in the state space analysis. Further, to make more practical sense of the results, and after preliminary testing indicated this to be a statistically acceptable approach, data were further grouped into six separate time periods (period 1: midnight–6:30 am, period 2: 6:30–10:00, period 3: 10:00–13:30, period 4: 13:30–17:00, period 5: 17:00–20:30, period 6: 20:30–midnight). As previously discussed, there are 144 loop locations whose data are available for analysis. Obviously, it is impossible to analyze the data for all detectors in this paper. To demonstrate the abilities of the chosen methodology, five detectors were chosen for further analysis: loop detectors 108 (L108), 106 (L106), 103 (L103), 101 (L103), and 12 (L12); all five locations are along an important 3-lane per direction signalized arterial on the periphery of the core area of the city (Alexandras Avenue) and are *sequential*; that is, L12 is located downstream of (in order) L101, L103, L106, L108 (Fig. 1). This sequential ordering of the loop detector placements yields a great opportunity to explore the modeling and prediction capabilities of the state space approach in a *multivariate* setting.

Prior to model estimation, all signalized intersections of the chosen arterial were scrutinized for traffic flow dependencies due to cyclic flow phenomena and platooning. In practice, the flow profile during any single signalization cycle will vary from the average due to the random behavior of individual vehicles. The profile of traffic entering a link will be displaced in time and modified during the journey along the link due to the different speeds of the individual vehicles and, as a result, platoons of vehicles will be partly dispersed. In order to test for this assumption, a simulation was conducted using actual signalization data for the entire corridor being studied. Measures of the mean modulus of error (MME) were derived that helped in the assessment of how far the profile of the arrival flow at the stopline deviates from the mean value. This measure is expressed as a number in the range 0.0–2.0 (Vincent et al., 1980). A uniform flow has a MME of 0.0 while a high MME would indicate a link on which the flow was strongly bunched or platooned. The derived values for the sequence of intersections measured by loop detectors L108 through to L12 are 0.58, 0.72, 0.85, 0.61 and 0.84. These values indicate the relevant efficiency of the existing signal coordination plans and, despite the existence of substantial inflows and outflows of traffic, the roadway retains a relative smooth operation suitable for flow estimation testing and prediction.

4. Model estimation

4.1. Preliminary investigation

As a first step in the model building process, we examined the traffic patterns for different days for the various loop locations under consideration. Figs. 2–4 show the observed traffic volumes at

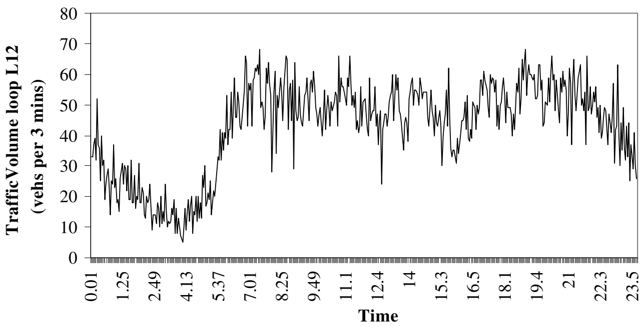

Fig. 2. Observed traffic volumes for a typical 24-h period (Loop L12).

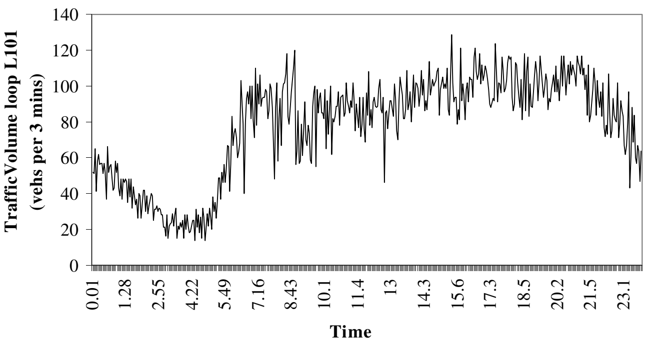

Fig. 3. Observed traffic volumes for a typical 24-h period (Loop L101).

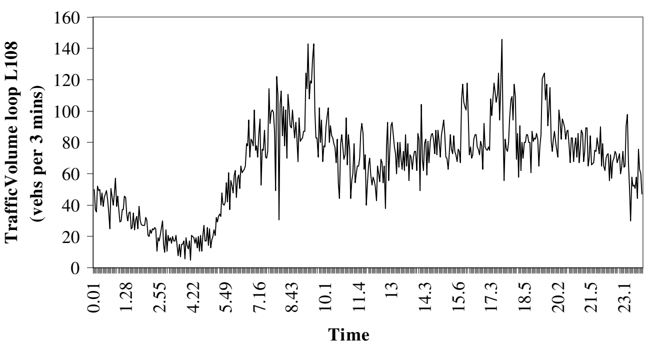

Fig. 4. Observed traffic volumes for a typical 24-h period (Loop L108).

three locations for a typical 24-h period. It is rather interesting to note that the *general* pattern is very similar for detector locations L101 and L12 that are less than 250 m apart, and quite similar even for detector locations such as L108 and L12 that are a distance of more than 2500 m apart. This comparison reinforces the idea that a joint multivariate consideration of a number of detectors may improve on both the modeling and prediction aspects.

Although, as noted in the Section 2, stationarity is *not* a requirement for state space models, we tested for it in order to compare the results of the developed models with those from the ARIMA approach. In general, a stationary time series has the same *statistical* behavior at each point in time. Forecasts based on nonstationary series usually exhibit large errors, and as such series are often differenced to obtain stationarity before the data can be analyzed. As can be seen from Fig. 5, the autocorrelation function (ACF) for L101 demonstrates considerably slow decay indicating the lack of stationarity. Interestingly, using only first order differencing, results in an ACF and a partial autocorrelation function (PACF) which tail-off exponentially (Fig. 6). Series whose ACF and PACF both tail-off exponentially indicate a combined autoregressive-moving average (ARIMA) process (Shumway and Stoffer, 2000). Nevertheless, because of the importance of correct differencing in the development of robust ARIMA models, we also tested for stationarity using the augmented Dickey Fuller (ADF) test. [3] At the 99% significance level, we rejected the null hypothesis of nonstationarity *after* first differencing was performed.

A further preliminary investigation was done to examine some of the basic autoregressive and cross-correlation characteristics of the time series for the various loop locations (Stathopoulos and Karlaftis, 2001b). The results showed that the series for all detector locations demonstrated strong positive autocorrelation of lags 1–3, and variable cross-correlations that depended on the time period of day examined. For example, the results showed that, when considering the pooled data, L12 was strongly cross-correlated with lags 2 and 3 of L101, lags 5 and 6 of L103, lags 8 and 9 of L106, and lags 12 and 13 of L108. This investigation, which closely follows the geometry of the test network, was the starting point for the development of the state space models presented in the next subsection.

4.2. Parameter estimation

Based on the results of the preliminary investigation, state space models were developed for both the pooled data (data from all time periods combined), and the data from the various periods separately. Since Figs. 5 and 6 indicate the presence of both an autoregressive and a moving average type process, the models developed "allowed" for an autoregressive and a moving average order of up to three lags. This means that the initial model specifications contained both autoregressive and moving average terms of up to, and including, third-order lags. Of course, through a thorough iterative search process for the "best" model specification, in terms of parameter

[3] Several tests for nonstationarity (unit root tests) have been devised. Experiments conducted by Granger and Engle (1984) indicate that the most satisfactory of these tests is the augmented Dickey Fuller test. In this test, the null hypothesis is that the variable under consideration, say, is not stationary, i.e. it requires at least one differencing. The alternative hypothesis is that the time series is already stationary. To test this hypothesis the following regression model is estimated: $\nabla y_t = \varphi y_{t-1} + \sum_{j=1}^{p} \gamma_j \nabla y_{t-1} + \varepsilon_t$, where the number of terms p is the minimum required to render ε_t into white noise. The null hypothesis is that $\varphi = 0$.

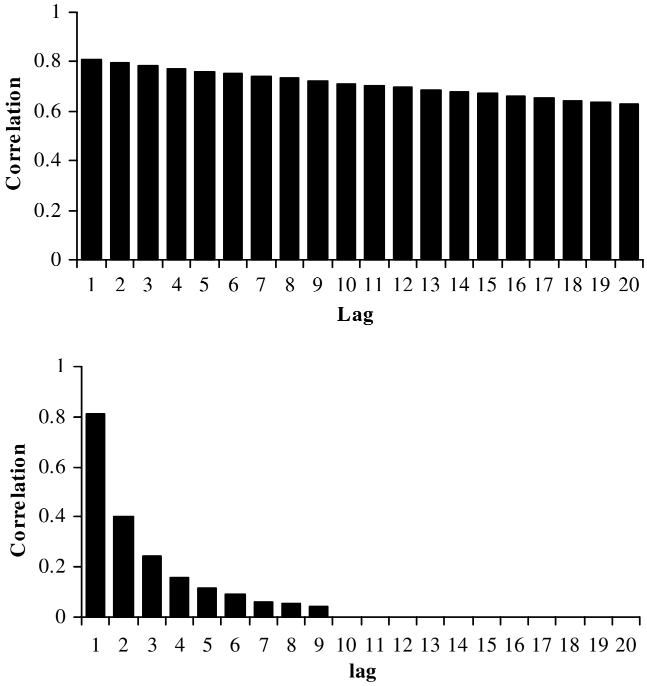

Fig. 5. Traffic volume autocorrelation and partial autocorrelation functions for L101.

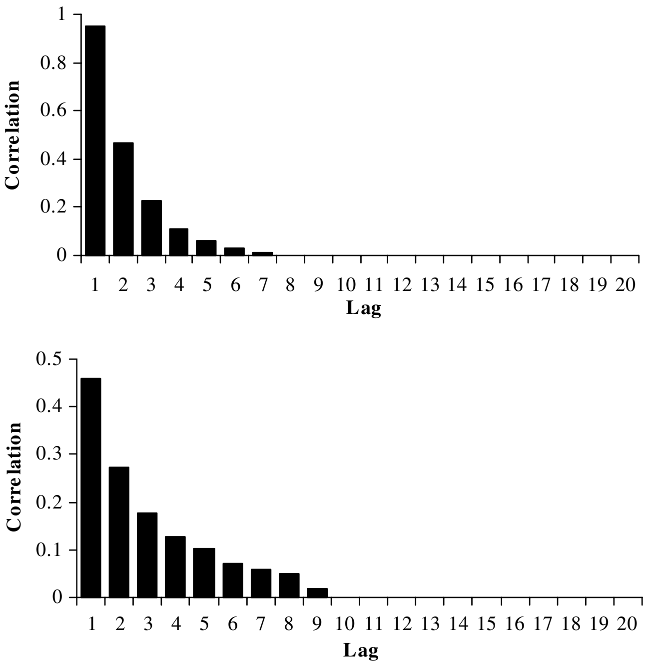

Fig. 6. Traffic volume autocorrelation and partial autocorrelation functions, after first differencing, for L103.

Table 1
Summary of the state space model autoregressive parameter order (Autoregressive Lags; All models include an MA(3) term)

Predicted		Order of parameter (lags)				
		L12	L101	L103	L106	L108
Location 12	Pooled[a]	1,2,3	2,3	5,6	8,9	12,13
	Period 1	1,2,3	2,3	5	8,9	12,13
	Period 2	1,2,3	–	6	8	12
	Period 3	1,2,3	2	–	9	–
	Period 4	1,2,3	3	6	–	13
	Period 5	1,2,3	2,3	–	9	–
	Period 6	1,2,3	2,3	6	9	12
Location 101	Pooled[a]		1,2,3	2,3	5,6	8,9
	Period 1		1,2,3	2	5,6	8,9
	Period 2		1,2,3	2,3	–	–
	Period 3		1,2,3		–	–
	Period 4		1,2,3	2	5	8,9
	Period 5		1,2,3	2	6	–
	Period 6		1,2,3	2	–	9
Location 103	Pooled[a]			1,2,3	2,3	5,6,7
	Period 1			1,2,3	–	5,6
	Period 2			1,2,3	–	7
	Period 3			1,2,3	–	5
	Period 4			1,2,3	3	–
	Period 5			1,2,3	3	–
	Period 6			1,2,3	2,3	5
Location 106	Pooled[a]				1,2,3	2,3
	Period 1				1,2,3	2,3
	Period 2				1,2,3	2
	Period 3				1,2,3	2,3
	Period 4				1,2,3	2
	Period 5				1,2,3	2
	Period 6				1,2,3	2

[a] Pooled: data from all periods (1–6) combined.

significance and predictive power, the final models comprise of a subset of the full specification. The results presented in Table 1 show the order of the significant (95%) autoregressive lags; the moving average part of the process is not presented, but the predictions obtained were based on the full model specification (including the moving average term).

From Table 1 it can be seen that the "best" specification for the pooled data from L12 follow a third order autoregressive-moving average process, while they are also significantly influenced from data from L101, L103, L106 and L108 at varying, because of the distance between the detector locations, lags. L101 in turn also follows an autoregressive-moving average process influenced by the data from the other locations except L12 (since we are examining flows from L101 *to* L12. The results for all locations appear in Table 1). Interestingly, while the pooled data consistently demonstrate the process described, when the models for the separate periods of the day were examined, the significant cross-correlation lags varied; this is a very interesting finding that can be mainly attributed to the difference in the underlying traffic and OD patterns during the different time periods of the day. Interestingly, the state space model for period 3 of L101 "collapsed" to an ARIMA (3,3) process. This implies that one of the most important strengths of the state space approach is its ability to model a wide variety of time series processes which may vary from a rather complicated mixed ARIMA and cross-correlation process to a simpler ARIMA or AR model. Finally, since L108 was considered as the "entry" point for the area examined we only considered the ARIMA (3,3) process as appropriate.

4.3. Predictions

As already mentioned, the main purpose of the state space model developed was to obtain short-term traffic volume predictions for the corridor examined. To obtain predictions, the original data set was split into two distinct subsets: the model development and the testing subsets; the first contained roughly 70% of the observations and the second the remaining 30% (to obtain this split we followed the more general rule-of-thumb that 2/3 of the data should be used for estimation and 1/3 for testing). Once a model was developed using the data from the first set, forecasts were generated and were then compared to the actual values from the testing set (technically this means that the predictions presented here are not in-sample; that is, different sets were used for estimation and testing. This approach yields inferior fit but more realistic forecasts than if the same data were used for both steps).

Fig. 7 presents a comparison of the predictions obtained from the state space model with those of the ARIMA model for each detector location. The predictions for L108 are the same since, as previously discussed, the state space model "collapsed" into an ARIMA model. It is clear that the predictions obtained from the state space models are superior to those obtained from the ARIMA models; the highest difference was demonstrated for the distant L12 where the state space model yields a mean absolute percent error (MAPE) of 12% compared to a 20% MAPE value from the ARIMA model, possibly due to the accumulated gain in traffic volume information from the previous detector locations. The differences in the models for the other detectors are important, but not as large. Interestingly, a series of *t-tests* indicated statistically significant differences between the prediction results obtained with the two methods for detectors L103, L101 and L12, and statistically no significant difference for L106.

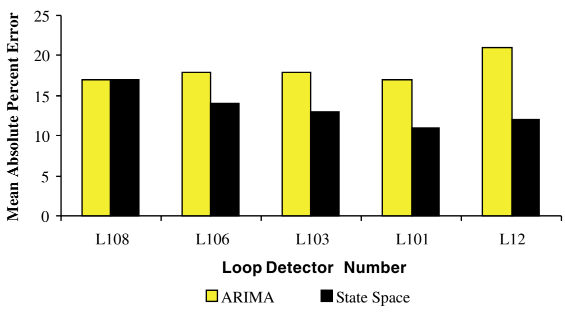

Fig. 7. Mean absolute error rate comparison between ARIMA and state space models.

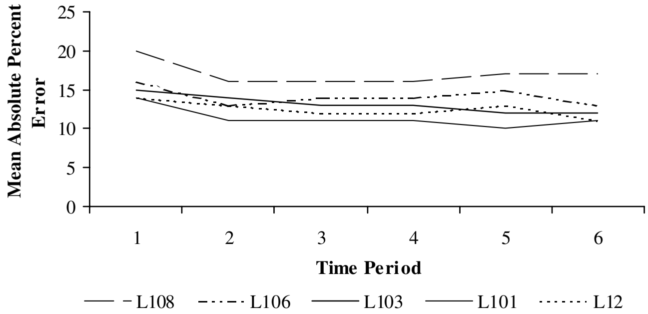

Fig. 8. Mean absolute percent error rate comparison for different time periods from the state space model.

Finally, Fig. 8 presents a comparison of MAPE values for the different detectors examined per time period. The MAPE values are quite similar, and statistically *not* significantly different, for the different time periods within detectors, with the exception of the first period (midnight–6:30 am), were MAPE values are relatively higher. As a final note, it should be mentioned that the MAPE values reported here are rather high when compared to values reported in other studies. This is, most probably, a reflection of the data available and the context of this study rather than a reflection of the methodology. It appears that, despite its potential usefulness, traffic flow in signalized urban arterials cannot be predicted, at least in the short-run, with as much accuracy as flow in urban freeways. This, of course, presents many methodological and model development challenges that need to be addressed in the future.

4.4. A caveat

In the study of traffic the problem is often one of locating queues, and thus bottlenecks in the transportation system. In detectorized networks, measurements of occupancy (the fraction of time vehicles are over a detector) are used to assess the concentration of traffic along roadways, a measurement that indicates the onset of congestion. The joint observation of actual traffic counts and occupancies over a link of the signalized roadway is shown in Fig. 9. The apparent trend of

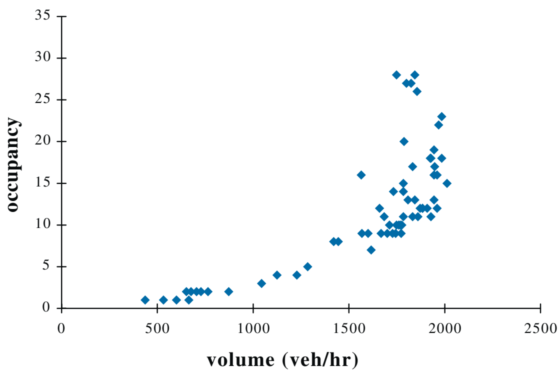

Fig. 9. Joint observations of actual traffic counts and occupancies over a link of the signalized roadway examined.

the scatter diagram is characteristic of the volume–occupancy relationship and encompasses many observed traffic regimes. Counts were aggregated over 15 min intervals to decrease the fluctuations of the scatter diagram. Observation fluctuate about the "optimum" link flow indicating that the accuracy of traffic predictions discussed in this paper could be improved by jointly examining traffic volumes and occupancies. This discussion also suggests that extreme values (of either volume or concentration) could be better analyzed in a meaningful traffic engineering way.

5. Conclusions

It is widely accepted that ITS technologies may lead to more efficient use of existing road network systems resulting in reduced traffic congestion, delays, emissions, energy consumption and improved safety. The success of these strategies depends on short-term traffic prediction for parameters such as traffic volumes, travel speeds and occupancy. The field of short-term prediction of traffic characteristics has attracted considerable interest in the literature that, with few exceptions, has been characterized by the univariate examination of freeway data. The present paper took a somewhat different approach to examining traffic volume data.

First, 3 min data from urban arterials streets near downtown Athens were used to estimate explicitly multivariate time series models. The models developed fall in the flexible family of state space approaches. Among the important advantages of this approach are both their explicit multivariate nature, which allows for data from different loop detectors to be jointly considered, and their ability to model a wide variety of univariate models, such as ARIMA, as special cases. Indeed, in some of the cases examined, the best fit was provided by ARIMA models that turn out to be quite useful especially for points of entry into the study area. The results of the models developed clearly suggest that, at least in the case of Athens, different specifications are appropriate for different time periods. Further, it also appears that the use of multivariate state space models holds promise in the urban roadway system.

It should be noted that the multivariate modeling of flow, speed and occupancy data in urban areas is a fairly complex and tedious process. Data from different detectors are not only highly correlated among themselves but are also related to prevailing traffic conditions which tend to exhibit high short-term fluctuation. In addition, during large parts of the day, traffic is highly congested, approaching unstable conditions. At such conditions the usual statistical modeling procedures may suffer from an inherent inability to describe traffic characteristics and as a result produce accurate short-term predictions. It is deemed necessary to develop new approaches with the ability to capture traffic behavior at their boundary conditions.

References

Chang, S.C., Kim, S.J., Ahn, B.H., 2000. Traffic-flow forecasting using time series analysis and artificial neural network: the application of judgmental adjustment. Presented in the 3rd IEEE International Conference on Intelligent Transportation Systems.

Chang, J.L., Miaou, S.P., 1999. Real-time prediction of traffic flows using dynamic generalized linear models. Transportation Research Record 1678, 168–178.

Chen, H., Grant-Muller, S., 2001. Use of sequential learning for short-term traffic flow forecasting. Transportation Research Part C 9, 319–336.

Davis, G.A., Nihan, N.L., 1991. Nonparametric regression and short-term freeway traffic forecasting. ASCE Journal of Transportation Engineering 117 (2), 178–188.

Dia, H., 2001. An object-oriented neural network approach to short-term traffic forecasting. European Journal of Operational Research 131, 253–261.

Durbin, J., 2000. The Foreman lecture: the state space approach to time series analysis and its potential for official statistics. Australian and New Zealand Journal of Statistics 42 (1), 1–23.

Fu, L., Rilett, L.R., 2000. Estimation of time-dependent, stochastic route travel times using artificial neural networks. Transportation Planning and Technology 24 (1), 25–36.

Granger, C.W., Engle, P., 1984. Dynamic model specification with equilibrium constraints: co-integration and error correction. Draft manuscript.

Hamed, M.M., Al-Masaeid, H.R., Said, Z.M.B., 1995. Short-term prediction of traffic volume in urban arterials. ASCE Journal of Transportation Engineering 121 (3), 249–254.

Harvey, A.C., 1984. A unified view of statistical forecasting procedures. Journal of Forecasting 3 (3), 245–283.

Johnson, R.A., Wichern, D.W., 1992. Applied Multivariate Statistical Analysis. Prentice Hall, Englewood Cliffs, NJ.

Kalman, R.E., 1960. A new approach to linear filtering and prediction problems. Journal of Basic Engineering, Transactions of the ASME 82D, 95–108.

Lee, S., Fambro, D.B., 1999. Application of subset autoregressive integrated moving average model for short-term freeway traffic volume forecasting. Transportation Research Record 1678, 179–188.

Lingras, P., Sharma, S.C., Osborne, P., 2000. Traffic volume time-series analysis according to the type of road use. Computer-Aided Civil and Infr a structure Engineering 15 (5), 365–373.

National Technical University of Athens, Department of Transportation Planning and Engineering, 1996. Available at: http://www/transport.ntua.gr/map.

Okutani, I., Stephanedes, Y.J., 1984. Dynamic prediction of traffic volume through Kalman filtering theory. Transportation Research part B 18 (1), 1–11.

Park, D., Rilett, L.R., 1998. Forecasting multiple-period freeway link travel times using modular neural networks. Transportation Research Record 1617, 63–70.

Park, D., Rilett, L.R., Han, G., 1999. Spectral basis neural networks for real-time travel time forecasting. ASCE Journal of Transportation Engineering 125 (6), 515–523.

Pushkar, A., Hall, F.L., Acha-Daza, J.A., 1995. Estimation of speeds from single-loop freeway flow and occupancy data using cusp catastrophe theory. Transportation Research Record 1457, 149–157.

Rilett, L., Park, D., 1999. Direct forecasting of freeway travel times using spectral basis neural networks. Presented at the 1999 Transportation Research Board Annual Meeting, Washington, DC.

Ross, P., 1982. Exponential filtering of traffic data. Transportation Research Record 869, 43–49.

Shumway, R.H., Stoffer, D.S., 2000. Time Series Analysis and Its Applications. Springer, New York, NY.

Smith, B.L., Demetsky, M.J., 1994. Short-term traffic flow prediction: neural network approach. Transportation Research Record 1453, 98–104.

Smith, B.L., Demetsky, M.J., 1997. Traffic flow forecasting: comparison of modeling approaches. ASCE Journal of Transportation Engineering 123 (4), 261–266.

Smith, B.L., Williams, B.M., Oswald, R.K., 2000. Parametric and nonparametric traffic volume forecasting. Presented at the 2000 Transportation Research Board Annual Meeting, Washington, DC.

Stathopoulos, A., Karlaftis, M.G., 2001a. Temporal and spatial variations of real-time traffic data in urban areas. Transportation Research Record, forthcoming.

Stathopoulos, A., Karlaftis, M.G., 2001b. Spectral and cross-spectral analysis of urban traffic flows. To be Presented in the 4th IEEE International Conference on Intelligent Transportation Systems.

Stevens, J.P., 1992. Applied Multivariate Analysis for the Social Sciences. Lawrence Erlbaum, Hillsdale, NJ.

Van Arem, B., Kirby, H.R., Van Der Vlist, M.J.M., Whittaker, J.C., 1997. Recent advances in the field of short-term traffic forecasting. International Journal of Forecasting 13, 1–12.

Van Der Voort, M., Dougherty, M., Watson, S., 1996. Combining Kohonen maps with ARIMA time series models to forecast traffic flow. Transportation Research Part C 4 (5), 307–318.

Vincent, R.A., Mitchell, A.I., Robertson, D.I., 1980. User Guide to TRANSYT v.8, TRRL LR888, London, UK.
Whittaker, J., Garside, S., Lindveld, K., 1997. Tracking and predicting a network traffic process. International Journal of Forecasting 13, 51–61.
Williams, B.M., 2001. Multivariate vehicular traffic flow prediction: an evaluation of ARIMAX modeling, Presented at the 2001 Transportation Research Board Annual Meeting, Washington, DC.
Yasdi, R., 1999. Prediction of road traffic using a neural network approach. Neural Computing and Applications 8, 135–142.
Zhang, H.M., 2000. Recursive prediction of traffic conditions with neural network models. ASCE Journal of Transportation Engineering 126 (6), 472–481.
Zhang, H., Ritchie, S.G., Lo, Z.P., 2000. Macroscopic modeling of freeway traffic using an artificial neural network. Transportation Research Record 1588, 110–119.