IVAN BOŠNJAK, D.Sc.
MIROSLAV MIKULA, D.Sc.
FRANE JELUŠIĆ, D.Sc.
Fakultet prometnih znanosti
Zagreb, Vukelićeva 4.

Science in Traffic
Original Scientific Paper
U.D.C.: 656.01:519.21
Accepted: Jan. 15, 1998
Approved: May 6, 1998

EVALUATION OF DETERMINISTIC AND STOCHASTIC COMPONENTS OF TRAFFIC COUNTS

SUMMARY

Traffic counts or statistical evidence of the traffic process are often a characteristic of time-series data. In this paper fundamental problem of estimating deterministic and stochastic components of a traffic process are considered, in the context of "generalised traffic modelling". Different methods for identification and/or elimination of the trend and seasonal components are applied for concrete traffic counts. Further investigations and applications of ARIMA models, Hilbert space formulations and state-space representations are suggested.

1. INTRODUCTION

According to the generalised traffic model introduced in [1], [2], we must identify substantial traffic quantities which can be measured or observed in space and time reference frame (space-time specification). The frequency and accuracy with which we record the chosen quantities give the "resolution level" which can range from annual or monthly statistical data to near real-time observations. From traffic observations $\{x_1, x_2, ...x_n\}$ we wish to estimate the underlying process $\{x_t\}$ in order to gain information concerning its deterministic and stochastic properties.

Different traffic flows (of vehicles, passengers, packets, cells, etc.) in traffic system can be generally described by quantity, time and space dimensions.

Most real traffic counts can be treated as a time series, which meaning a set of observations x_t, each one being recorded at specific time t. Observations are usually made at fixed time intervals, but can also be recorded continuously over a time interval. To allow the possibly stohastic nature of traffic process it is reasonable to suppose that each traffic observation x_t is a realised value of certain random variable X_t. Observed data or traffic counts can be considered as a part of realiastion of a stohastic process $\{X_t, t \in T\}$.

Application of advanced statistical modelling in the estimation of an origin-destination (OD) trip matrix from traffic counts has been considered by several researches [3]. In these studies the probabilistic properties of the observed data are considered in depth and used in the estimation of an OD matrix. A broad review of statistical estimation procedures for OD matrix, including traditional entropy-maximising estimators, Bayesian estimation etc., are discussed in reference [4].

This paper starts with generalised traffic modelling approach considered fundamental problem of estimating deterministic and stohastic components of time-series data (traffic counts). In this context we evaluate different methods for estimation and/or elimination of the trend and seasonal components from traffic data. Although the advanced statistical components and methods are used, the defined problem has to be treated as a version of a typical traffic engineering task. With satisfactory mathematical model based on generalised traffic description, it becomes possible to estimate parameters and use the fitted model to enhance the understanding of a traffic process. Once a satisfactory model has been developed (and supported by program), it may be used in a variety of ways in traffic analysis, control, prediction, design, etc.

Stationary processes play a crucial role in the mathematical analysis of time series [5]. Many observed time series are nonstationary, but, frequently such data sets can be "transformed" into a series or parts which can be reasonably modelled as realisations of the same stationary process. From the observations of a stationary time series $\{X_t\}$ we can estimate the autocovariance function $\gamma\ (_o)$ of the underlying process $\{X_t\}$ as a sample autocovariance function. The sample autocovariance (and autocorrelation functions) can be computed for any data set $\{x_1, x_2, ... x_n\}$, and the are not restricted to realisations of a stationary process.

2. DISCRIBING TRAFFIC PROCESS AS A STOCHASTIC PROCESS

The concept of stochastic process and applied stochastic system modelling are the essential part of mathematical courses at an undergraduate and graduate level for engineers and managers. In the context of generalised traffic modelling, we need to define precisely what is meant by a stochastic process and its realisations. Considerations are focused to special classes of processes which are particularly useful for modelling many of the traffic counts as a time series data.

Definition 1. A stochastic process is a family of random variables $\{X_t,\ t \in T\}$, defined on probability space $(\Omega, \zeta, P,)$.

If the time parameter T is a countable set, the process is called a discrete-time stochastic process; and if T is a continuum, the process is called a continuous time stochastic process. In time series analysis the parameter (or index) set T is a set of time points, very often : $\{1, 2, ...n\}$, $\{0, \pm 1, \pm 2, ...n\}$ or $[0, \alpha)$.

Each observation x_t is a realised value of a certain random variable X_t. The time series $\{x_t,\ t \in T_s\}$ is then a realisation of the family of random variables $\{X_t,\ t \in T_s\}$.

According to the definition of a random variable we can conclude that for each fixed $t \in T, x_t$ is in fact a function $X_t(_o\ )$ on set Ω. On the other side, for each fixed $\omega \in \Omega$, $X_o\ (\omega)$ is a function on T.

Definition 2. Realisations of a stohastic process or sample paths of the process $\{X_t, t \in T_s\}$ are the functions $\{X_0(\omega), \omega \in \Omega\}$ on T.

Stationary processes play a control role in the mathematical analysis of time series, where autocovariance function is a primary tool.

Definition 3. If $\{X_t\ , t\ \in\ T\}$ is a process such that $Var(X_t) < \alpha$ for each $t \in T$, then the autocovariance function $g_x(_{oo})$ of $\{X_t\}$ is defined by:

$$\gamma_x(r,s) = Cov(X_r, X_s) =$$
$$= E[(X_r - EX_r)(X_s - EX_s)]r, s \in T$$

Definition 4. The time series $\{X_t, t \in Z\}$ with index set Z = $\{0, \pm 1, \pm 2,...\}$ is said to be stationary if:

(i) $E\ |X_t|^2 < \alpha$ for all $t \in Z$,

(ii) $E X_t = m$ for all $t \in Z$,

(iii) $\gamma_x(r,s) = \gamma_x(r+t, s+t)$ for all $r, s, t \in Z$.

Stationary defined in Definition 4 is frequently treated as a stationary in the wide sense, weak stationary. $\{X_t, t \in Z\}$ is stationary $\gamma_x(r,s) = \gamma_x(r - s, 0)$ for all $r,s \in Z$. It is therefore reasonable to redefine the autocovariance function of a stationary process as the function of just one variable:

$\gamma_x(h) \equiv \gamma_x(h,0) = Cov\ (X_{t+h}, X_t)$ for all $t, h \in Z$.

The function $\gamma_x(_o)$ will be referred to as the autocovariance function of $\{X_t\}$ and $\gamma_x(h)$ as its value at *"lag"h*. The autocorrelation function of $\{X_t\}$ is defined analogously as the function whose value at *log h is:*

$\rho_x(h) \equiv \gamma_x(h)/\gamma_x(o) = Corr\ (X_{t+h}, X_t)$ for all $t, h \in Z$.

Definition5. The time series $\{X_t, t \in Z\}$ is said to be strictly stationary if the joint distributions of $(X_t, ... X_k)'$ and $(X_{t+h}, ... X_{k+h})'$ are the same for all positive integers k and for all $t_1, ..., t_k, h \in Z$.

From the observations $\{x_1, x_2, ...x_n\}$ of a stationary time series $\{X_t\}$ we often want to estimate the autocovariance function $\gamma(_o)$ of the underlying process $\{X_t\}$ in order to gain information about its dependence structure. The sample autocovariance function of an observed series are frequently used [5], [6].

The sample autocovariance and autocorrelation function can be computed for any data set $\{x_1, ...x_n\}$ and are not restricted to realisations of a stationary process.

Important role in the modelling of time-series data have the family of "Autoregressive Moving Average" processes (ARMA processes). For any autocovariance function $\gamma(_o)$ such that $\lim_{h \to \alpha} \gamma(h) = 0$, and for any integer $k > 0$, it is possible to find ARMA process with autocovariance function $\gamma_x(_o)$ such that $\gamma_x(h) = \gamma(h), h = 0, 1, ...k$. The linear structure of ARMA processes lead to simple and useful best linear predictions of a stationary process using observations taken at or before time n to forecast the subsequent behaviour of $\{X_t\}$.

A generalisation of the class ARMA processes are "Autoregressive-Integrated Moving Average" processes (ARIMA processes). ARIMA models incorporate a wide range of non-stationary series, i.e. processes which, after differencing finitely many times, reduce to ARMA processes. Once the data have been suitably transformed, the problem becomes one of finding a satisfactory ARMA (p,q) model for $\{X_t\}$.

3. METHODS FOR ESTIMATING AND ELIMINATING THE TREND AND SEASONAL COMPONENTS FROM TIME-SERIES DATA

We consider the usable methods for estimation and elimination of deterministic components from observed traffic counts as a time-series data showed in Table 1. Deviations from stationarity are suggested by the graph of the series itself or by the sample autocorrelation function. It is clear from the graph that time-series has strong seasonal components of period 12 (months).

Two basic methods for elimination of trend and seasonal components are:

1 "classical decomposition" of the series into a trend component, a seasonal component, and a random residual component;

2 apply difference operators repeatedly to the data $\{x_t\}$. We want to estimate and extract deterministic

components, i.e. "trend component" m_t and "seasonal component" s_t, in the hope that the residual component Y_t will turn out to be a stationary random process.

After the preliminary inspection of the concrete time-series graph (Figure 1) we adopt the method for eliminating both the trend and the seasonal components in the general decomposition model:

$X_t=m_t+s_t+Y_t$

where:

- m_t is a trend component,

Table 1 - Traffic counts with monthly "resolution level"

	1992	1993	1994	1995	1996
I	8980	7745	8157	7705	7828
II	8100	6980	7301	7456	6890
III	8920	8031	8118	7768	7786
IV	9129	8419	7865	7918	8125
V	10005	8710	9382	8630	9109
VI	10802	9505	9548	8940	9428
VII	10904	10115	10088	10071	10480
VIII	10750	9817	9615	9172	9823
IX	9705	8738	8279	8030	9102
X	9928	9122	8430	8481	9065
XI	9154	8705	8156	7870	8628
XII	8922	8675	8029	8641	9232

- s_t is a seasonal component with known period d,
- Y_t is a residual random component.

Three different methods for estimating and removing the trend and seasonal components from observed data $\{x_1,...x_{60}\}$, will be applied and evaluated. They are:

1 The Small Trend Method;
2 Moving Average Estimation;
3 Differencing at Lag d.

3.1. The Small Trend Method

For this method it will be convient to index the data by year and month. We will denote by: $x_{j,k}$, $j=1,...,5$; $k=1,...,12$ the number of traffic services in k^{th} month of the j^{th} year.

Since the trend is small in observed data, it is not unreasonable to suppose that the trend term is constant: m_j for the year j^{th}. Seasonal component is clearly 12 and we note the period d is 12. Since $\sum_{k=1}^{12} s_k = 0$ we have natural unbiased estimate:

$$\hat{m}_j = \frac{1}{12}\sum_{k=1}^{12} x_{j,k}$$

while for s_k, $k = 1, ... , 12$ we have the estimates:

$$\hat{s}_k = \frac{1}{5}\sum_{j=1}^{5}\left(x_{j,k}-\hat{m}_j\right)$$

The estimated residual component or error term for month k of the j^{th} year is:

$$\hat{Y}_{j,k} = x_{j,k} - m_{j,k} - s_{j,k} \quad , \quad j=1,...,5 \quad k=1,....,12.$$

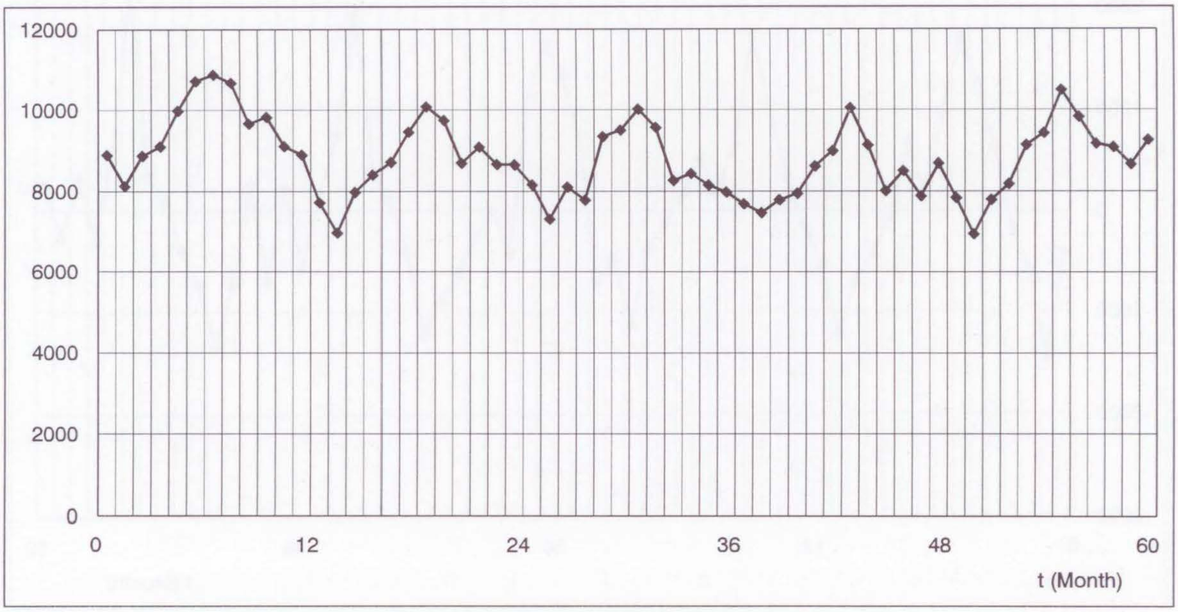

Figure 1 - Monthly traffic data

The procedure of analysing and predicting traffic counts data by introducing decomposition methods can be systematised in several steps:

1 Plot the traffic counts data.
2 Find estimates for seasonal component $\hat{s}_t$, $t=1,...,12$ for the classical decomposition model:

$$X_t = m_t + s_t + Y_t$$

where $s_t = s_{t+12}$, $\sum_{t=1}^{12} s_t = 0$, and $Ey_t = 0$.

3 Plot the deseasonalised data, $X_t - \hat{s}_t$, $t=1,...,60$.
4 Fit a parabola by least squares to the deseasonalised data and use it as estimate $\hat{m}_t$ of m_t.
5 Plot the residuals $\hat{Y}_t = \hat{X}_t - \hat{m}_t - \hat{s}_t$, t=1,...60.
6 Compute the sample autocorrelation function of the residuals $\rho(h)$, $h=0,...20$.
7 Use fitted model to predict X_t, $t > 60$.

By applying the presented method we find out detrended observations $x_{j,k} - m_j$ and deseasonalised observations: $\hat{Y}_{j,k} = x_{j,k} - \hat{m}_j - \hat{s}_k$. Monthly traffic data after substracting the trend are illustrated in Figure 2, and detrended and deseasonalised data are illustrated in Figure 3.

3.2. Moving Average Estimation Method

Moving Average Estimation Method (M2) can be preferable to previously described method since it does not rely on the assumption that m_t is nearly constant over each cycle.

This method for eliminating both trend and seasonal components, will be illustrated in general and implemented to traffic observations $\{x_1,...x_n\}$. The first step is to estimate the trend by applying a "moving average filter" specially chosen to eliminate the seasonal component and to dampen the noise. If the period d of seasonal component is even, say $d=2q$, then we use:

(xx)

$$\hat{m}_t = \left(0{,}5x_{t-q} + x_{t-q+1} + ..+ x_{t+q-1} + 0{,}5\,x_{t+q}\right)/\,d$$

$$q < t \le n-q$$

where q is a non-negative integer.

For the cases when period d is increasing, say $d=2q+1$, then we use the simple moving average:

$$\hat{m}_t = (2q+1)^{-1} \sum_{j=-q}^{q} x_{t+j},$$

$$q+1 \le t \le n-q$$

The task of second step is to estimate the seasonal component. In this step, for each $k = 1, ..., d$; we compute the average w_k of the deviations:

$$\left\{\left(x_{k+jd} - \hat{m}_{k+jd}\right): q < k + jd \le n-q\right\}$$

These average deviations do not necessarily sum to zero.

We estimate the seasonal component s_k as:

$$\hat{S}_k = w_k - d^{-1} \sum_{i=1} w_i \qquad k=1,\ldots,d$$

$$\hat{S}_k = \hat{S}_{k-d}, \quad k > d.$$

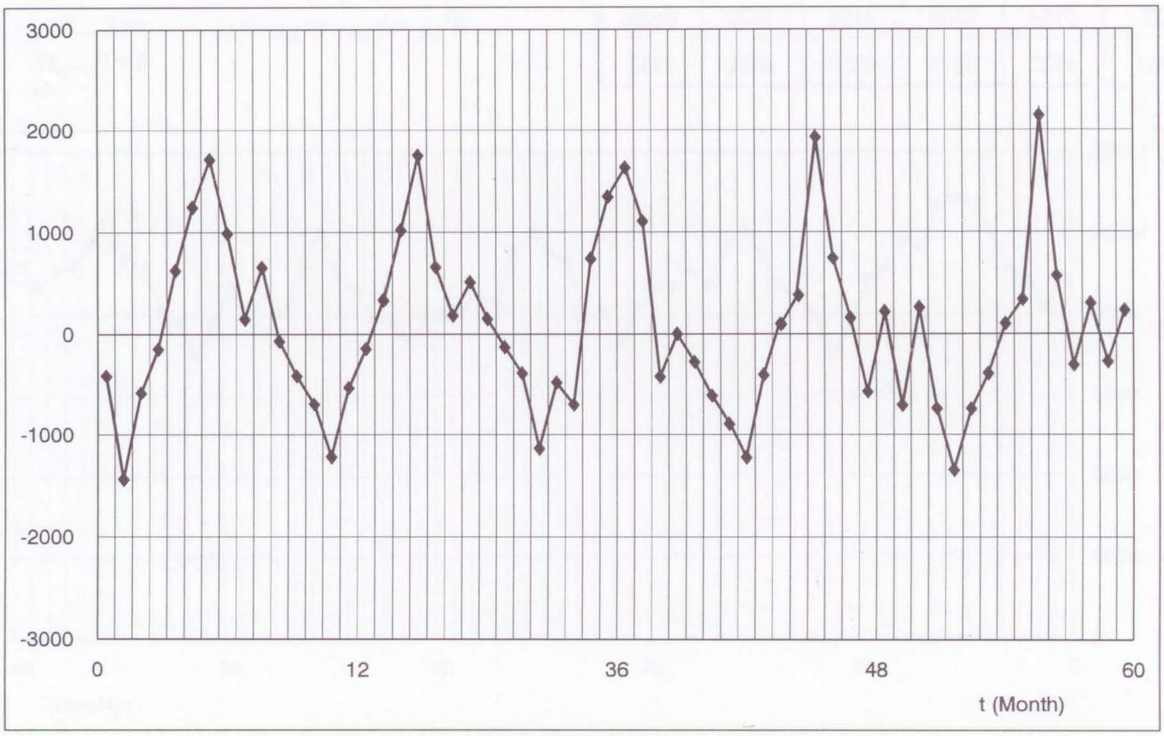

Figure 2 - Monthly traffic data after subtracting the trend

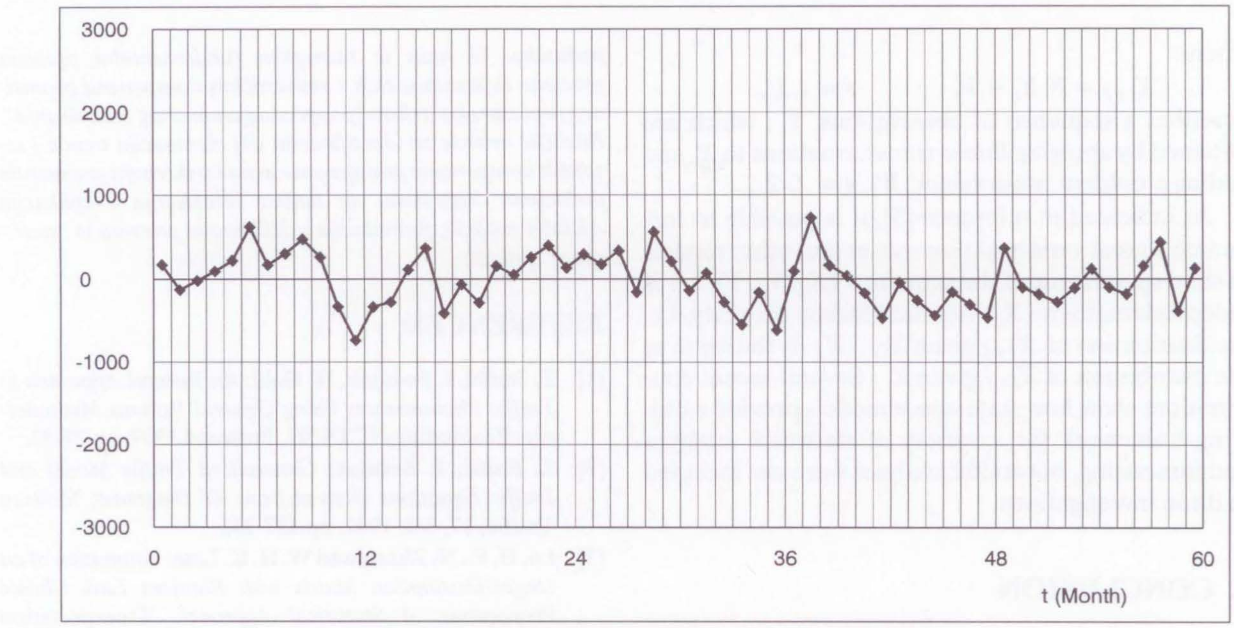

Figure 3 - The de-trended and de-seasonalised data

The deseasonalised data is then defined to be the original series with the estimated seasonal component removed:

$$d_t = x_t - \hat{d}_t \qquad t=1, ..., n.$$

Finally we re-estimate the trend from $\{d_t\}$ by:

- applying a moving average filter for non-seasonal data,
- fitting a polynomal to the series $\{d_t\}$.

Some of the developed parameter estimation program allows the options of fitting a linear or square trend $\hat{m}_t$. The estimated noise terms are then:

$$\hat{Y}_t = x_t - \hat{m}_t - \hat{s}_t \qquad t=1,...,n.$$

The results of applying Methods M1 and M2 to the traffic counts are quite similar. In this case the piecewise constant and moving average estimates of m_t are reasonably close.

3.3. Differencing at *Lag d*

The methods of differencing can be adapted to deal with seasonal component of period *d* by introducing the lag-d difference operator ∇_d defined by:

$$\nabla_d = X_t - X_{t-d} = \left(1-B^d\right)X_t$$

where *B* is the backward shift operator.

Applying the operator ∇_d to the general model

$$X_t = m_t + s_t + Y_t$$

where $\{S_t\}$ has period *d*, we obtain

$$\nabla_d X_t = m_t - m_{t-d} + Y_t - Y_{t-d}$$

which gives a decomposition of the difference $\nabla_d X_t$ into a trend component:

$$(m_t - m_{t-d})$$

and a noise term:

$$(Y_t - Y_{t-d})$$

If we apply the operator ∇ to $\nabla_{12} x_t$ and plot the resulting diferencess $\nabla \nabla_{12} x_t$, *t = 14, 15, ... 60,* we obtain the resulting graph with no apparent trend or seasonal component.

4. APPLICATIONS OF HILBERT SPACE FORMULATIONS AND STATE-SPACE REPRESENTATIONS

There are several great advantages to be gained from a Hilbert space formulation in a time series (or traffic counts) analysis. Some of these advantages are derived from our familiarity with two-of orthogonal projections in these spaces. These concepts appropriately extend to infinite-dimensional Hilbert space, have a significant contribution to the study of random variables with finite second moments, and especially in the prediction of stationary processes. In many cases, intuition gained from geometric understanding can be a valuable guide in the construction of models or algorithms.

Recent contributions with state-space representations and Kalman recursions have a strong impact on time series and related areas, especially for control of systems in relation tu fundamental concept of observability and controllability [8], [9]. The efficiency of a state-space representation lies in the simple structure of state equations which describe evolution of the state X_t of the system at time *t* (a $v \times l$ vector) in terms of a known sequence of $v \times v$ matrices F_1, F_2...and a sequence of random vectors X_1, V_1, V_2,....Equation:

$$Y_t = G_t X_t + W_t \qquad t = 1,2,...$$

where:

$$X_{t+1} = F_t X_t + V_t \qquad t = 1,2,...$$

describes a sequence of observations, Y_t, which are obtained by applying linear transformations to X_t and adding a random noise vector, W_t, $t = 1, 2, ...$

As indicated in reference [5], it is possible to formulate a great variety of time-series and other models in state-space form. If the sequence $\{X_1, V_1, V_2, ...\}$ is independent, then $\{X_t\}$ has the Markov property, i.e. the distribution of X_{t+1} given $X_t, ...X_1$ is the same as the distribution of X_{t+1} given X_t. Several recent contributions show how state-space models provide a unifying framework for a variety of statistical analyses and forecasting, but traffic analyses were not included in those investigations.

5. CONCLUSION

Fundamental problem of evaluating deterministic and stochastic components of traffic processes is investigated in several papers and doctoral dissertations. This paper has focused on evaluation of traffic counts which have characteristics of time-series data Three basic methods for identification and elimination of trend and seasonal components are applied for concrete traffic counts. Further investigations and development of generalised methodology for traffic analysis and synthesis are recommended.

SAŽETAK

EVALUACIJA DETERMINISTIČKIH I STOHASTIČKIH KOMPONENTI IZ VREMENSKOG NIZA PROMETNIH PODATAKA

Kvantitativni prometni podaci ili stohastičke evidencije prometnog procesa često su dane u obliku vremenskog niza podataka. U radu je razmotren fundamentalni problem procjene determinističkih i stohastičkih komponenti prometnog procesa u kontekstu "poopćenog prometnog modeliranja". Različite metode za identifikaciju i/ili eliminaciju trenda i sezonske komponente primijenjene su na konkretnim prometnim podacima. Sugerirana su daljna istraživanja i aplikacija ARIMA modela, formulacija u Hilbertovu prostoru te "state - space" prikazi.

REFERENCES:

[1] **Z. Radić, I. Bošnjak, H. Gold**: *An Integral Approach to Traffic Phenomenon Using General Systems Methodology*; Proceedings ICTS '97, Portorož,1997, pp.77-83.

[2] **Z. Radić, I. Bošnjak**: *Generalized Traffic Model and Traffic Equations Derived from ST Diagrams*; Modern Traffic, 17, 3-4, 1997, pp.237-241.

[3] **Lo, H. P., N. Zhang and W. H. K. Lam**: *Estimation of an Origin-Destination Matrix with Random Link Choice Proportions: A Statistical Approach*; Transportation Research-B, Vol.30, No.4, pp. 309-324, 1996.

[4] **Ortuzar, J. D. and L. G. Willumsen**: *Modelling Transport*; John Wiley and Sons, London, 1990.

[5] **Brockwell, P. J. and R. A. Davis**: *Time series-Theory and Methods*, Springer-Verlag, New York, 1995.

[6] **Box, G. E. P. and Jenkins, G. M.**: *Time Series Analysis: Forecasting and Control*, Holden-Day, San Francisco,1970.

Davids, N.M.A. and Winter, R.B.: *Stochastic Modelling and Control*, Chapman and hall, London, 1985.
Journal of Times Series Analysis
Journal of Appl. Probability

[7] **Osaki, S**: *Applied Stochastic System Modelling*, Springer-Vergal, Berlin, 1992.

Berger, M. A.: *An Introduction to Probability and Stochastic Processes*, Springer-Verlag, Berlin, 1992.

[8] **Bošnjak, I.**: *Technology of Telecommunication Traffic*, Faculty of Traffic Sciences, Zagreb, 1997.

[9] Proceedings of 8th International Conference on Road Traffic Monitoring and Control, IEE, London, 1996.