IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Mathematical Modelling of Engineering Problems
Vol. 8, No. 3, June, 2021, pp. 393-402
Journal homepage: http://iieta.org/journals/mmep

Identifying the Optimum Forecasting Horizon to Apply the Singular Spectrum Analysis on Daily Road Traffic Volume Forecasts

Stylianos Z. Kolidakis*, George N. Botzoris

Department of Civil Engineering, Democritus University of Thrace, Section of Transportation, Kimmeria Campus, Xanthi 67100, Greece

Corresponding Author Email: skolidak@ee.duth.gr

https://doi.org/10.18280/mmep.080308

Received: 20 January 2021
Accepted: 19 May 2021

Keywords:
transport demand, road traffic forecasting, singular spectrum analysis, forecasting ability, ex-post evaluation

ABSTRACT

The paper delivers an assessment of Singular Spectrum Analysis (SSA) forecasting ability for short- and medium-term forecasting horizon, on real time traffic volume data. The key study goal is to estimate forecasting pertinency for daily traffic volume, based upon measurements at toll station. The suggested methodology is tested on real data from Moschohorion and Pelasgia Toll Station – Greece, utilizing custom developed forecasting software toolbox. Applied research results confirm an advanced forecasting ability of proposed methodology for short-term forecasting horizon against medium term forecasting horizon, when performance is compared upon the statistical criteria of the coefficient of determination R^2. The obtained results present that SSA forecasting model could provide a competent forecasting methodology for road traffic volume data.

1. INTRODUCTION

1.1 Rationale

Transportation demand forecasting is the process of estimating the amount of people or vehicles that will utilize a transport infrastructure or service over a particular time interval. It is an essential part of transportation planning, which helps determine the size of the facility, the standards of its maintenance, the personnel and equipment required, etc.

The numerous methods that have been employed in road transport demand forecasting may be classified as qualitative (such as market surveys, Delphi method, and expert opinion method), or quantitative (such as econometric, time series, and causal analysis models) [1-3]. Time series methods are used for the assessment of future traffic load volume, while past data depicting traffic of demand are available. Time series can be defined as a series of data recorded in a time order, at equal time intervals (e.g., per month, day, hour) [1].

All time series methods are reasonably accurate but are inherently sensitive to noise. Furthermore, they do not correctly capture the qualitatively heterogeneous components (such as the trend and the periodic, quasi-periodic, or structural behaviors) of which any time series is the sum. To increase the accuracy of time series forecasting, various methods have been developed to remove noise from raw data and to decompose any time series into its trend, its oscillatory components and its noise components. One of these methods is Singular Spectrum Analysis (SSA), which, for a given window length, decomposes any time series into various components that can either be trends, periodic oscillations or noise [4].

SSA is tested whether it is able to provide structural analysis of time series, which signifies that SSA is utilized to decompose original time series into the plain and understandable components such as trend, periodicities and noise residual. While trend is deducted, acknowledgement of subset time series with periodic components is the key mission for SSA to implement. Afterwards, residuals of leading components time series decomposition constitute the noise components [4].

In conclusion, the SSA techniques accomplish to decompose original time series into the components of trend, oscillation and noise which are easier to understand, analyze and forecast.

1.2 Research objectives

The scope of this paper is to deliver pragmatic proof for SSA efficiency and usefulness in modeling and identification of optimum forecasting horizon for Greek Toll Station traffic volume. The paper is focused on the following research topics:

- Evaluate and assess optimum forecasting horizon through an empirical real-life application on daily traffic volume.
- Examine whether SSA methodology is able to model and provide solid and robust forecasting ability for road traffic volume.

2. LITERATURE REVIEW

The SSA methodology has been used during last decades in many scientific areas for analysis (identification of trend, seasonality and periodic oscillations, noise detection and removal, etc.) and forecasting purposes.

Marques et al. applied SSA technique on univariate hydrological time series real data to identify important information and forecast skills [5]. The authors investigated relative research in a literature overview, discovering an extend SSA application on several scientific fields. SSA decomposition and reconstruction main stages were described in order to exploit algorithm forecasting ability in hydrological

time series. Experimental outputs suggested that SSA could extract important hydrological time series components with nonlinear behavior such as precipitation and run offs and provide an accurate forecast. Moreover, even though significant components were extracted and forecasted using SSA, a large portion of hydrological time series variability did not seem to have a suitable time structure to be forecasted.

Hassani applied SSA techniques on monthly time series concerning the accidental deaths in the USA between 1973 and 1978, and then compared the SSA with traditional forecasting methodologies [6]. The author used the SSA to split the original time series into a subgroup of trend, oscillatory time series and structureless noise time series. The results, confirmed with statistical criteria analysis such as Mean Absolute Error (MAE) and Mean Relative Absolute Error (MRAE), revealed the SSA is a robust forecasting methodology.

Alexandrov used SSA to perform trend extraction using monthly data of the unemployment level in Alaska between the years 1976 and 2006 [7]. The author underline that SSA is an interesting and robust to outliers methodology that can extracts trend of noisy time series containing oscillations of unknown period.

Hassani and Thommakos studied the SSA methodology on financial and economic time series forecasting [8]. The authors designated the SSA methodology, they stipulate certain conditions for SSA forecasting and proposed a variety of SSA applications and experimental outputs. Noteworthy strive focused on SSA forecasting aptitude and the appropriate assessment of SSA methodology basic parameters. Both the theoretical approach and the experimental outcomes proved SSA as a competent and favorable forecasting methodology of economic and financial time series.

Briceño et al. presented SSA application for electric volume forecasting purposes [9]. The authors performed a literature overview in the field of electric volume forecasting and described the main stages of the SSA: decomposition, reconstructing and forecasting. The main parameters of the SSA (window length and number of components) were selected, and case study was executed for data acquired from wind power generators in Venezuela. Comparison of the SSA with other forecasting methodologies, such as SARIMA and additive Holt-Winter, using MSE and MAPE statistical criteria, revealed that the SSA consist a reliable forecasting methodology.

Alvarez-Meza et al. proposed an automatic SSA-based methodology to split and rebuild time series [10]. A clustering process was suggested to detect input signal main components, by computing a subgroup of orthogonal basis, engaging spectrum analysis criteria. The subgroup was represented by the Discrete Fourier Transform to deduce basis vectors encoding comparable data patterns. Thus, it was feasible to identify unseen trends or periodicities into the signal. This approximation was evaluated over artificial and real-life data, indicating that proposed methodology was a robust toolbox to split time series.

Furthermore, Hassani et al. examined the possible benefits of tourism demand forecasting utilizing SSA [11]. The authors evaluated SSA forecasting ability on United States tourist arrivals monthly data from 1996 to 2012. SSA forecasting capacity was compared to a variety of other forecasting techniques such as ARIMA, exponential smoothing and Artificial Neural Networks (ANN). The experimental outcomes conclude that SSA approach accomplished significantly improved forecasts (statistically) associated to other methodologies. The key conclusion identified major SSA advantages in USA tourist arrivals forecasting and proposed SSA as a noteworthy methodology for other forecasting studies on tourism demand.

Mahmoudv et al. studied SSA feasibility to perform forecasting of the mortality rate [12]. Hyndman-Ullah benchmarking as forecasting model was assumed in order to achieve SSA forecasting ability evaluation with other forecasting techniques. The authors endorsed for SSA because of SSA ability to split time series into a subgroup of individual components, while important attributes of trend, oscillatory components and noise were identified in original time series. SSA critical parameters were indicated with rule of thumb and the gained results were verified on real data from nine cities of Europe.

Silva and Hassani applied SSA on US trade before, during and after 2008 recession forecast [13]. The authors advocated for SSA less sensitiveness to initial time series systemic breaks and consequently competent to economic values after recession modeling and forecasting. It was significant to remark that the researchers verified SSA as a propitious methodology which, when conglomerate into hybrid models such as ARIMA or ANN, could preponderate over single models with vigorous forecasting ability. Experimental results led to conclusion that SSA provided a more accurate forecasting model in comparison to other traditional forecasting techniques such as ARIMA, exponential smoothing and ANN.

Kolidakis et al. applied SSA on a hybrid forecasting model for traffic volume [4]. The main research goal was to examine whether the appliance of SSA could improve significantly the forecasting accuracy of Artificial Neural Networks (ANN) in short-term daily traffic volume forecasts crossways Greek National Highway Network. Experimental results verified the improved forecasting performance of hybrid SSA-ANN model against forecasting ability of conventional ANN forecasting model.

3. SINGULAR SPECTRUM ANALYSIS: BASIC CHARACTERISTICS, IMPLEMENTATION PRINCIPLES, AND FORECASTING PROCEDURE

3.1 Basic characteristics

Singular Spectrum Analysis (SSA) was a relatively new methodology for time series decomposition, reconstruction and forecasting. Method development has been credited to the works of Broomhead and King [14], but it was popularized by Golyandina et al. [15]. In this section, a brief description of the SSA technique for decomposition, reconstruction and forecasting is given.

SSA is confronted with a main challenge: to analyze structure of time series, which means that SSA is involved to decompose original time series into acknowledgeable components of trend, periodic oscillations and noise residuals. Analyzing and identifying systemic patterns and random noise is the main objective in time series analysis. Many time series subsets could be discerned in two key groups: time series subgroup characterizing trend and components characterizing periodicities. While trend subgroup is removed, recognition of periodic subgroup time series components is a key mission for SSA to perform. In conclusion, SSA technique accomplished

to decompose the original time series into understandable components of trend, oscillation periodicities and noise residuals (Figure 1).

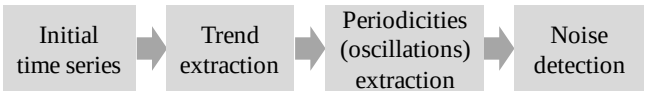

Figure 1. SSA steps for time series decomposition into simpler components

3.2 Methodology application

SSA methodology is divided into two stages: Decomposition and Reconstruction. Decomposition is also analyzed into two steps: Embedding and Singular Value Decomposition, while reconstruction is analyzed into two steps too: Grouping and Diagonal Averaging [6, 16] (Figure 2).

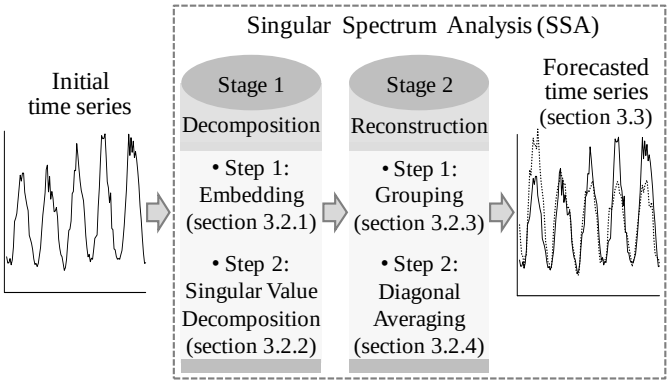

Figure 2. SSA methodology description

3.2.1 Stage 1: Decomposition, Step 1: Embedding

Embedding is called the process where a one-dimensional time series transformed into a sequence of lagged vectors of size L, by forming $K = N - L + 1$ lagged vectors:

$$X = (x_1, x_2, x_3, \ldots, x_T) \tag{1}$$

Usually, the above-mentioned vectors are referred as *L–lagged vectors*. The matrix formed after those vectors is called trajectory matrix $\boldsymbol{X} = (X_1, X_2, X_3, \ldots, X_T)$, where:

$$\boldsymbol{X} = (x_{ij})_{i,j=1}^{L,K} = \begin{pmatrix} y_1 & y_2 & y_3 & \cdots & y_K \\ y_2 & y_3 & y_4 & \cdots & y_{K+1} \\ \vdots & \vdots & \vdots & \ddots & \\ y_L & y_{L+1} & y_{L+2} & \cdots & y_T \end{pmatrix} \tag{2}$$

The resulting matrix $\boldsymbol{X}$ is a Hankel matrix, which means that all elements along the diagonal $i+j=const.$ are equal [6, 16].

3.2.2 Stage 1: Decomposition, Step 2: Singular Value Decomposition

The step consists of the Singular Value Decomposition (SVD) of the trajectory matrix $\boldsymbol{X}$ into a sum of bi-orthogonal elementary matrices of rank 1. Let (λ_1, λ_2, λ_3, …, λ_L) be the eigenvalues of $S = XX'$, arranged in decreasing order of magnitude ($\lambda_1 > \lambda_2 > \lambda_3 > \ldots > \lambda_L$) and its corresponding orthonormal system of eigenvectors $U_1, U_2, U_3, \ldots, U_L$.

By setting $d = max(i, \text{such that } \lambda_i > 0) = \text{rank X}$ and $V_i = X' U_i / \sqrt{\lambda_i}, i = 1,2,\ldots,d$, the SVD of the trajectory matrix $\boldsymbol{X}$ can be written:

$$\boldsymbol{X} = X_1 + X_2 + X_3 + \ldots + X_d \tag{3}$$

where, $X_i = \sqrt{\lambda_i} U_i V_i'$. The matrixes $\boldsymbol{X_i}$ has rank 1 and are called elementary matrixes. The selection $X_i = \sqrt{\lambda_i} U_i V_i'$ is called eigentriple (ET) of SVD. Note that in real life applications $d=L^*$, with $L^*=min\{L,K\}$ [6, 16].

3.2.3 Stage 2: Reconstruction, Step 1: Grouping

In this step, the elementary matrices X_i are separated into different groups, based on their similarity, and then the matrices within each group are summed. Let $I=\{i_1, i_2, i_3, \ldots, i_p\}$ be a group. In such a case, the matrix $\boldsymbol{X_I}$ corresponding to the group I can be defined as $X_I = X_{i_1} + X_{i_2} + \ldots + X_{i_p}$. If the set of indices J=1, 2, …, d split into m disjoint subsets I_1, I_2, …, I_m (a procedure known as eigentriple grouping) then the matrix $\boldsymbol{X}$ can be presented as:

$$\boldsymbol{X} = X_{i_1} + X_{i_2} + X_{i_3} + \ldots + X_{i_m} \tag{4}$$

3.2.4 Stage 2: Reconstruction, Step 2: Diagonal Averaging

In the final step of the SSA methodology, the matrices $\boldsymbol{I_1}, \boldsymbol{I_2}, \ldots, \boldsymbol{I_m}$ are converted into a one-dimensional time series, through a process known as diagonal averaging. Let z_{ij} be an element of an arbitrary $L \times K$ matrix $\boldsymbol{Z}$, then the k-th element of the resulting series is obtained by averaging z_{ij} over all i,j such that $i+j=k+2$. The method essentially takes the average of all the elements in all the diagonals that can be formed from the matrix $\boldsymbol{Z}$, such that $i+j=k+2$. This is also known as Hankelization of the matrix $\boldsymbol{Z}$.

Let $L* = \min\{L,K\}$, $K* = \max\{L,K\}$, $N = L - K + 1$, and $z*_{ij} = z_{ij}$ if $L < K$ and $z*_{ij} = z_{ji}$ otherwise. The diagonal averaging procedure transforms the matrix $\boldsymbol{Z}$ into the time series $(\tilde{y}_1, \tilde{y}_2, \ldots \tilde{y}_T)$ using the Eq. (5):

$$z_k = \begin{cases} \dfrac{1}{k-1} \displaystyle\sum_{m=1}^{k-1} z^*_{m,k-m+1} & 1 \le k \le L^* \\ \dfrac{1}{L^*} \displaystyle\sum_{m=1}^{L} z^*_{m,k-m+1} & L^* \le k < K \\ \dfrac{1}{N-k} \displaystyle\sum_{m=k-K^*+1}^{L} z_{m,k-m+1} & K^* \le k < N \end{cases} \tag{5}$$

The Hankelization procedure is optimal because the Hankelized matrix of $\boldsymbol{Z}$ is the closest to $\boldsymbol{Z}$ (with respect to the matrix norm) among all Hankel matrices of the same size [16]. If the Hankelization procedure is applied to the expansion in Eq. (4), then:

$$\boldsymbol{X} = \tilde{X}_{I_1} + \tilde{X}_{I_2} + \tilde{X}_{I_3} + \ldots + \tilde{X}_{I_m} \tag{6}$$

where, $\tilde{X}$ is the Hankelized matrix of $\boldsymbol{X}$. This is equivalent to the decomposition of the initial time series $Y_T = \tilde{y}_1, \tilde{y}_2, \ldots \tilde{y}_T$ into a sum of m series:

$$y_t = \sum_{k=1}^{m} \tilde{y}_t^{(k)} \quad (7)$$

where, $Y_T = \tilde{y}_1, \tilde{y}_2, \ldots \tilde{y}_T$ corresponded to the $\tilde{X}_{I_k}$ matrix [16].

3.3 Forecasting procedure

SSA can be used for time series forecasting. Given a time series $X = (x_1, x_2, x_3, \ldots, x_N)$, where N is the time series length, the following steps are implemented:

- Estimation of the window length $L<X/2$.
- Construction of the trajectory matrix $\boldsymbol{X}$ for the time series X.
- Construction of the orthogonal system of eigenvalues $U_1, U_2, U_3, \ldots, U_L$, using the Singular Value Decomposition.
- Construction of the matrix $\hat{X} = \sum_{i=1}^{L} U_i U_i^T X$.
- Construction of the Hankelized matrix of $\tilde{X} = H(\hat{X})$.
- Setting $v^2 = \pi_1^2 + \pi_2^2 + \cdots + \pi_L^2$ $(v^2 < 1)$, where π_i is the last component for vector $U_i, i = 1,2, \ldots L$.
- Determination of the vector $A = a_1, a_2, a_3, \ldots, a_{L-1}$, by using the Eq. (8):

$$A = \frac{1}{1-v^2} \sum_{i=1}^{L} \pi_i U_i \quad (8)$$

It can be proved that the last component x_L of any vector $X = (x_1, x_2, x_3, \ldots, x_L)$ is a linear combination of the first components $X = (x_1, x_2, x_3, \ldots, x_{L-1})$.

- The last step in forecasting procedure is the definition of the forecasted values $X_{N+h} = (x_1, x_2, x_3, \ldots, x_{N+h})$ as follows:

$$x_i = \begin{cases} \tilde{x}_i & \text{for } i = 1,2,\ldots N \\ \sum_{j=1}^{L-1} a_j x_{i-j} & \text{for } i = 1,2,\ldots N+h \end{cases} \quad (9)$$

where, h is the length of time series which will be forecasted.

4. CASE STUDY

A software toolbox is developed to validate SSA effectiveness and reliability for short- and medium-term road traffic forecasting, elaborating daily traffic volume from Moschohorion and Pelasgia toll stations of the Aegean Motorway (Figure 3). It is the principal north-south road connection in Greece (connecting the country's capital Athens with the country's second largest city, Thessaloniki) and part of the European route E75 which starts from Norway to Greece and runs south via Finland, Poland, Czech Rep., Slovakia, Hungary, Serbia, North Macedonia. Daily recording of traffic volume begun on April 2008 and ended on May 2014, producing 2,251 daily road traffic volume records (Figure 4).

To investigate and evaluate SSA forecasting ability for short- and medium-term horizon, the coefficient of determination R^2 is the most common forecasting ability statistical criterion [3], and is calculated by Eq. (10):

$$R^2 = 1 - \frac{\sum_{i=1}^{N} \left(y_i - \tilde{y}_i \right)^2}{\sum_{i=1}^{N} \left(y_i - \bar{y}_i \right)^2} \quad (10)$$

where, y_i are the real values of the initial time series, $\tilde{y}_i$ the estimated by the SSA values, $\bar{y}_i$ the mean value of y_i, and N the number of the time series available records (daily traffic volume data). The coefficient of determination R^2 takes values in the closed interval [0,1]. The value R^2=1 indicates that the SSA forecasting model predicts perfectly the real data of the daily traffic volume, whereas the value R^2=0 indicates that no relationship between real data and the estimated by the SSA values can be found [3].

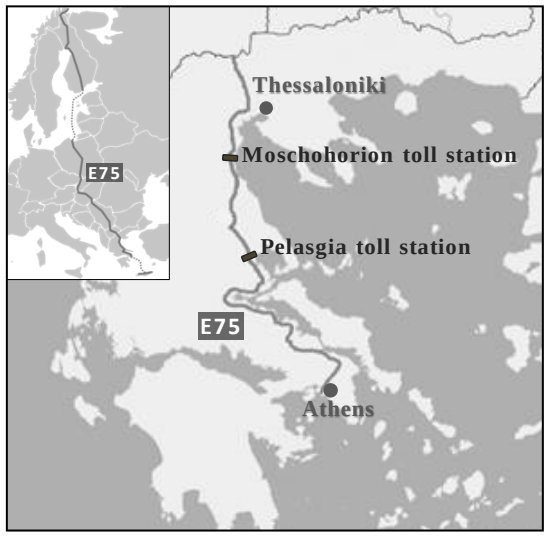

Figure 3. Moschohorion and Pelasgia toll stations, lengthwise Aegean Motorway, Greece

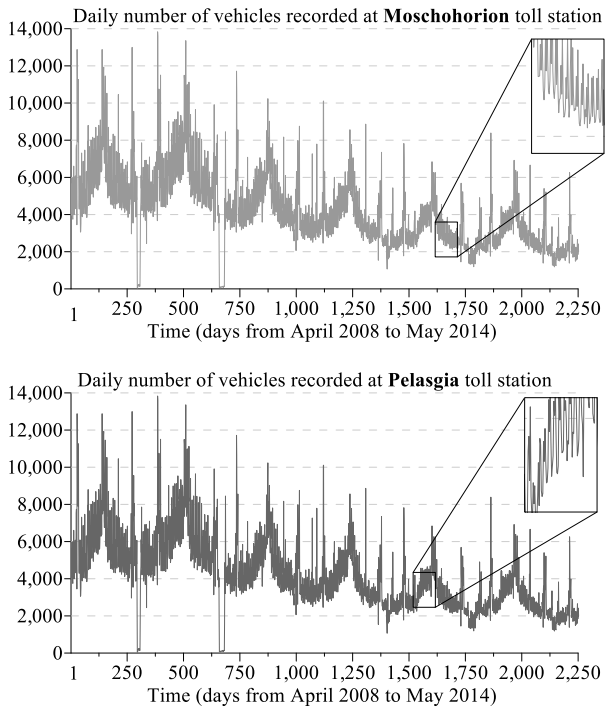

Figure 4. Daily traffic volume of Moschohorion and Pelasgia toll stations

5. APPLICATION AND RESULTS

5.1 Parameter selection

The two parameters to be selected for the SSA algorithm are the window length L and the number of elementary matrices to use for the reconstruction r. The L parameter determines the number of components that initial time series will be decomposed into and the r parameter defines the subset of critical components that will finally be used for SSA forecasting. Forecasting horizon to be investigated is assumed M=7 days and M=90 days ahead. Figure 5 indicatively illustrates the decomposition of initial time series of the daily traffic volume of Moschohorion toll station into 30 components, whereas Figure 6 gives the same information for Pelasgia toll station.

5.2 Selecting the window length, L

The window length is a significant parameter needed for the decomposition of the time series. There is currently no algorithm for selecting the window length, but many researchers have suggested as a rule of thumb choosing $L<N/2$, where N the number of available time series data [16-19]. This basic rule is derived from the fact that trajectory matrix SVD's are equivalent, due to the symmetry of left and right singular vectors. The choice of L depends on the nature of the time series data and the type of components that one is interested in extracting.

Large values for L can lead to undesired split of original time series into a subgroup of subsequent time series, where each component may mix unpleasantly with other components. This troublesome possibility may cause a poor quality for component separation, which provokes a low quality SSA, with weak ability to component separation. On the other hand, small values for L may lead to mix-up of subsequent time series components, which also deprives the SSA ability to separate initial time series into interpretable components. In any case, it is proper to run SSA several times, using different values for L, in order to estimate the optimized L value for SSA decomposition. In that perspective, six different ranges of values were simulated for L parameter, in order to identify the optimum L parameter. For both study cases (Moschohorion and Pelasgia toll station) the six different cases for L parameter were:

- L parameter range up to 20, notated as Critical-20.
- L parameter range up to 50, notated as Critical-50.
- L parameter range up to 100, notated as Critical-100.
- L parameter range up to 300, notated as Critical-300.

Different cases were investigated in order to detect processing time for different window length L values and select a balanced value between long processing time and efficient forecasting ability.

For time series data with a known period, it is recommended choosing L such that L/T is an integer [15]. For instance, if the time series data was seasonal and the period is 4, then choosing L to be multiples of 4 (4, 8, 12,...) would help capture the periodic components with periods 4. If time series had multiple periods say $T_1,T_2,...,T_n$, then L should be chosen such that L/T_i be an integer for all i=1,2,..., n, which was possible only if all $T_1,T_2,...,T_n$ were rational to each other and even then may yield $L>N/2$.

To extract only a trend component, L should be chosen large enough so that the trend was separable from other components such as the noise but not too large because large values of L mix-up the trend with other components. In conclusion, L should be chosen such that all the components from the decomposition of the time series were separable (distinct) or non-correlated.

5.3 Selecting the leading components, r

Ordinarily, a harmonic produced two eigentriples with near eigenvalues. One more helpful approach was delivered by inspecting breaking points of eigenvalue spectrum, counting on a customary noise residuals attribute of singular values slowly decreasing sequence [8]. Figure 7 presents the plot of eigenvalues, ordered by their decomposition contribution, for the Moschohorion and Pelasgia toll stations. A significant drop occurred at eigenvalue 3 and another one at eigenvalue 12. So, according to Figure 7, it is safe to assume that leading components could be 3 or 12. After 12 components, the eigenvalues decrease very slowly, thus component 13 and after are assumed to be noise.

Another typical measurement based on the contribution of each component to the variance of original time series, was assessed by Eq. (11):

$$\frac{\lambda_i}{\sum_{i=1}^{d}\lambda_i} \tag{11}$$

where, λ_i is the eigenvalue i to the corresponding eigenvector i, and d=1, 2, …, Lag.

Forecasting with SSA involved choosing only two parameters L and r. If an appropriate L had been chosen, then for an arbitrary time series, the author could choose $r<L$ such that the error in prediction was minimized [18].

For both case studies (Moschohorion and Pelasgia toll station), 95% of initial time series information was assumed to be the minimum acceptable amount of initial time series spectral content, and can define the number of critical components r when identifying the optimum forecasting horizon for the SSA (Figure 7).

Taking under consideration this fact, critical components can be estimated, for different cases of parameter L. Thus, at forecasting horizon of M=7 days and M=90 days, for Moschohorion toll station critical components r are estimated to be:

- ♦ up to 2, when initial time series is split up to 20 subgroup time series (L=20),
- ♦ up to 4, when initial time series is split up to 50 subgroup time series (L=50),
- ♦ up to 6, when initial time series is split up to 100 subgroup time series (L=100),
- ♦ up to 10, when initial time series is split up to 300 subgroup time series (L=300),

while for Pelasgia toll station, at a forecasting horizon of M=7 and M=90 days, critical components are estimated to be:

- – up to 5, when initial time series is split up to 20 subgroup time series (L=20),
- – up to 9, when initial time series is split up to 50 subgroup time series (L=50),
- – up to 13, when initial time series is split up to 100 subgroup time series (L=100),
- – up to 29, when initial time series is split up to 300 subgroup time series (L=300).

It is significant to highlight the fact that the values of the parameters L and r could differentiate in each case study (Moschohorion and Pelasgia toll station), since they depend on the nature and the characteristics of the time series that describe each case study.

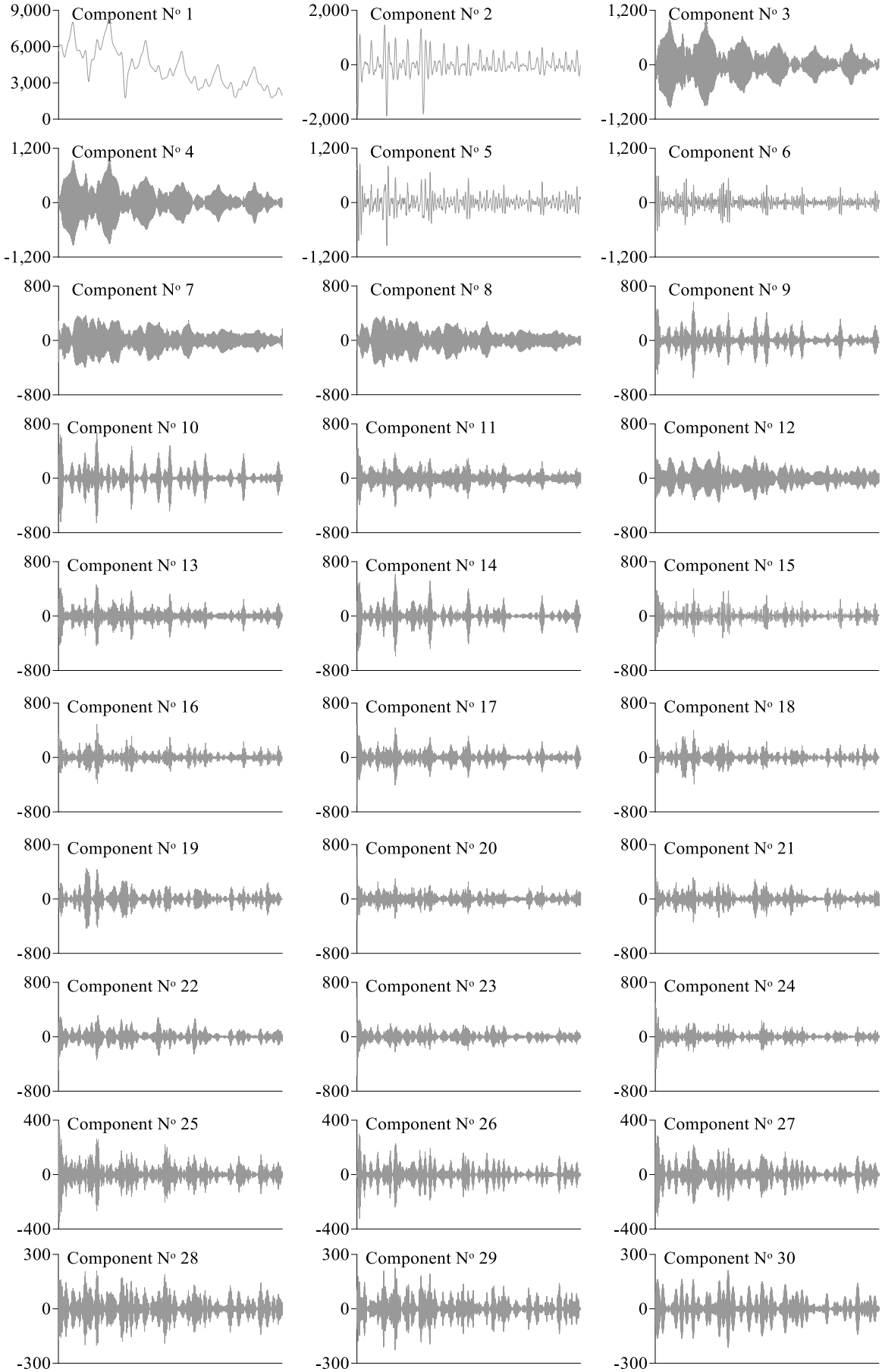

Figure 5. Principal components extracted by SSA decomposition, from Moschohorion toll station traffic volume time series

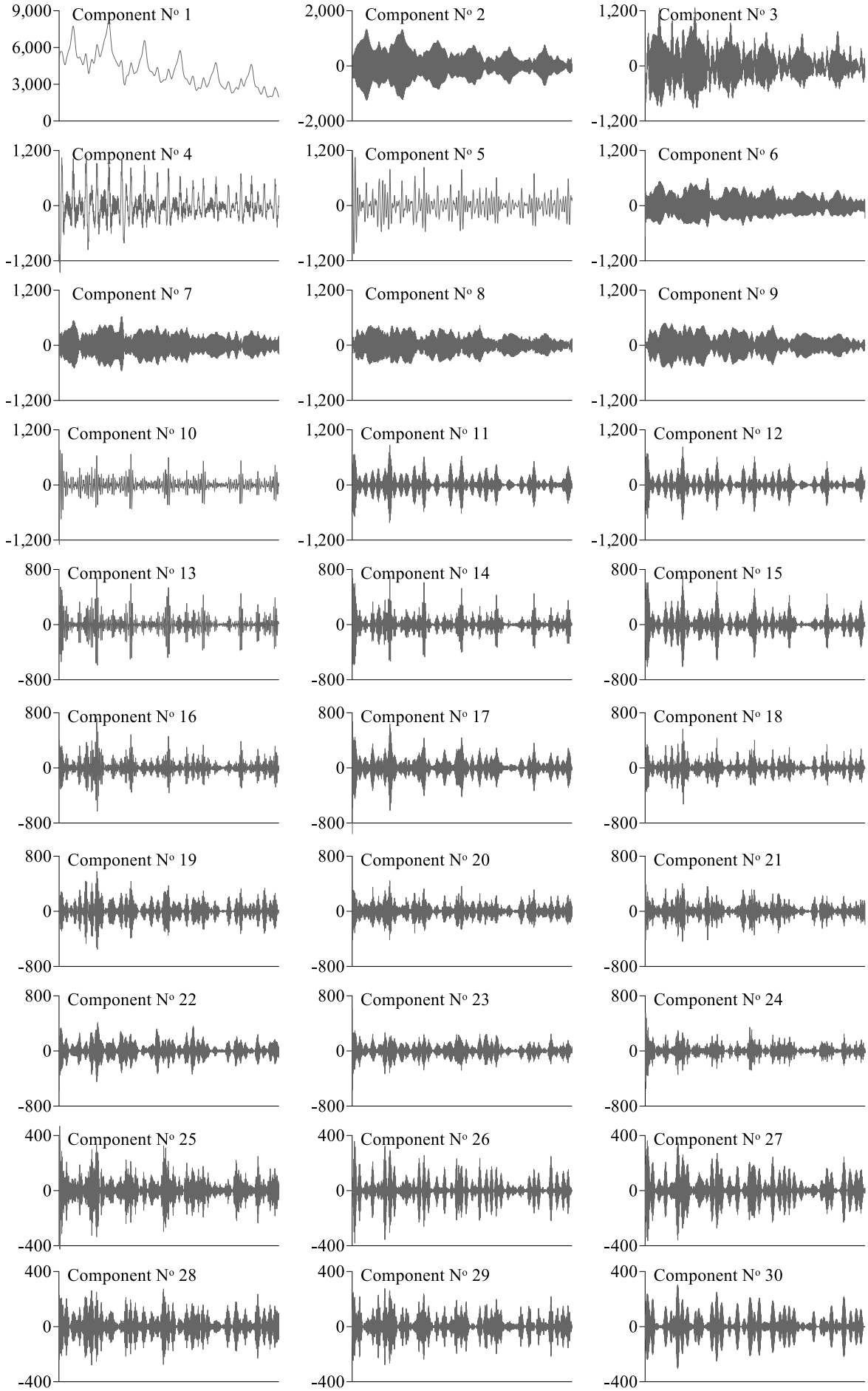

Figure 6. Principal components extracted by SSA decomposition, from Pelasgia toll station traffic volume time series

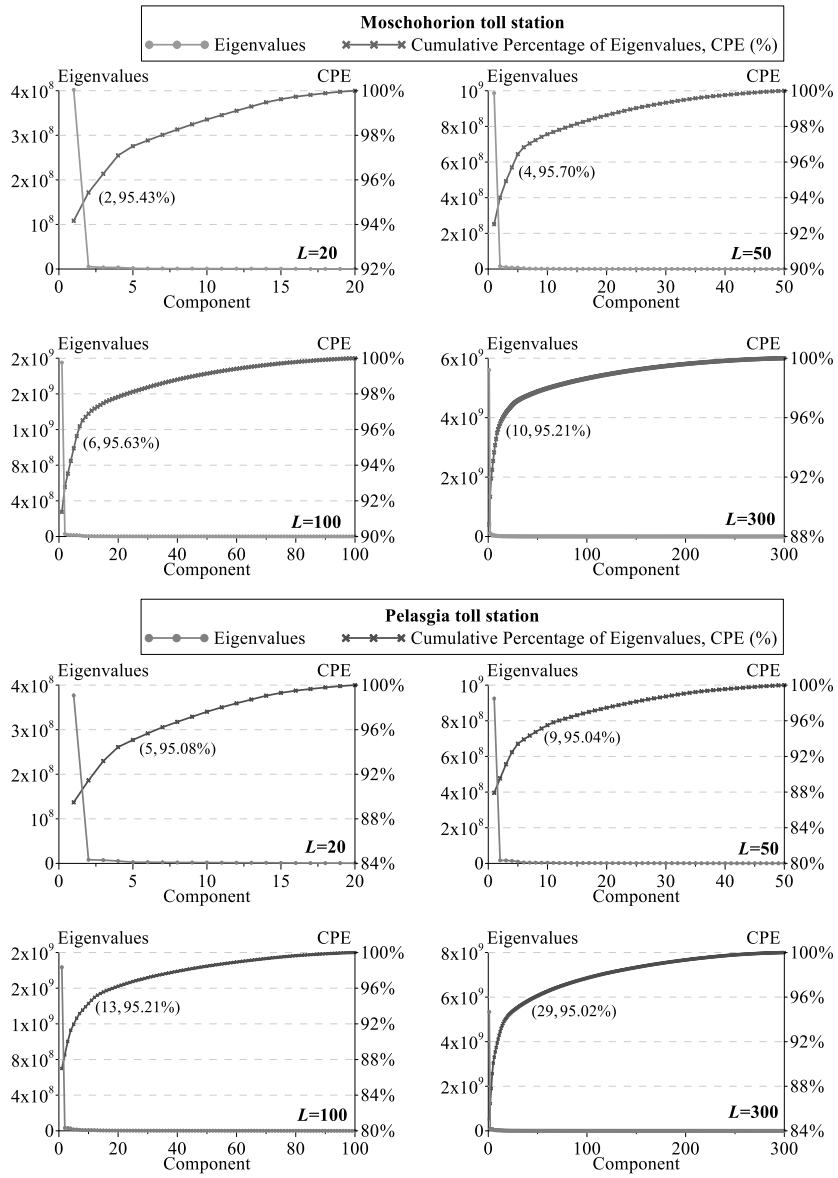

Figure 7. Eigenvalues and Cumulative Percentage of Eigenvalues (CPE) graphs for Moschohorion and Pelasgia toll stations. The values in parentheses in each graph denote the number of critical component in which the minimum acceptable time series information (at least 95%) is interpreted by SSA

5.4 Selecting optimum values for L and r

In order to identify the optimum architecture selection for L and r parameters, custom software was developed to simulate all different values calculations of forecasting window M, L and r parameters. Figure 8 illustrates the evolution of R^2 at both Moschohorion and Pelasgia toll station, for L= 20, 50, 100, and 300. In both cases, it was evident that the performance of SSA forecasting ability is degrading when forecasting horizon is extended, which mean that forecasting performance is robust for small forecasting horizon values (M=7) and weak for larger forecasting horizon values (M=90).

Moreover, a significant parameter to be highlighted is the necessary computer processing time for simulation and modeling with SSA. Obviously, the response time of a computer depends on the speed of its central processing unit and, the available memory, the used software for modeling the SSA, etc. However, it is possible to compare the relative computer processing time, setting the value 1 for the required time for SSA to forecast the future road traffic demand in the short term (M=7) when L=20, and the required time when the window length L increased (L=50, L=100, and L=300) and the forecasting horizon refers to the medium term (M=90) (Figure 9).

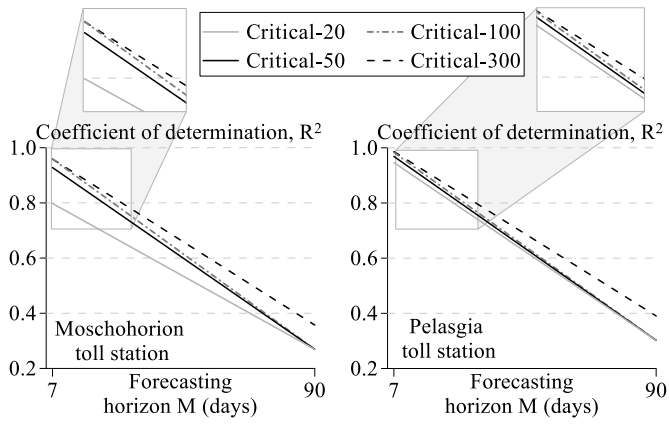

Figure 8. Evolution of the coefficient of determination R^2 for short-term (M=7 days) and medium-term (M=90 days) forecasting for different values of window length L

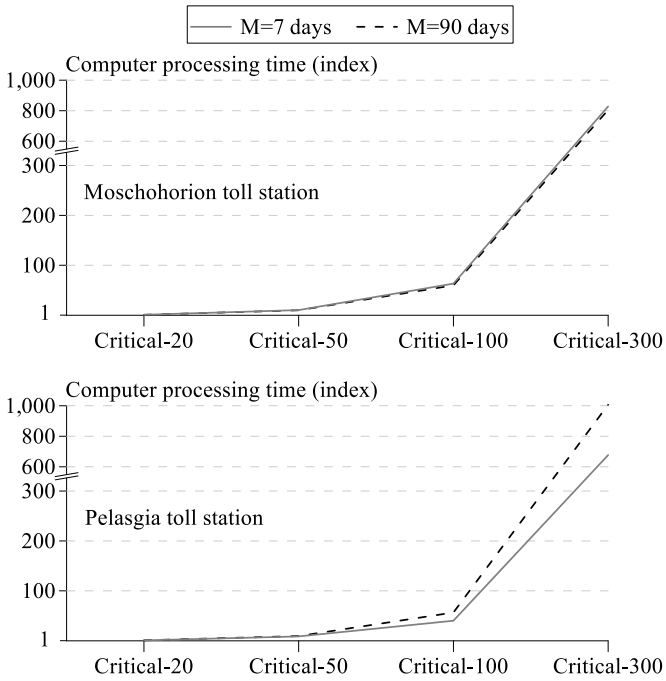

Figure 9. Computer processing time in relation to the forecasting horizon *M* and the window length *L*

From Figure 9 it is derived that small values of window length *L* (Critical-20 and Critical-50) retained processing time in small values. On the other hand, large values of window length *L* (Critical-100 and Critical-300) significantly increase the computer processing time. Having in mind that proposed forecasting methodology should be applied in real life problems, it is apparent that the window length L should increase only when there is enough available time. For example, for the case under study and using a typical desktop PC, the required processing time for *M*=7 and *L*=20 is around 12 sec, however, when *L*=300 the required processing time can reach the 5 hours.

6. CONCLUSIONS

Accurate forecasts necessity for road traffic volume is apparent for transportation assets managers, engineers and authorities. Vigorous road traffic forecasts are crucial for long-term and short-term decision-making with in planning and management, while proper management of road traffic is an important issue as the volume of traffic increases day by day. Road traffic congestion has various impacts; increased travel time leads to delays, additional fuel consumption results higher emissions, drivers stress provokes aggressive driving behaviors and in some cases accidents. Thus, road traffic volume forecasting methodologies provide a significant decision support toolbox for transportation system capacity to local and national infrastructure managing authorities.

The paper presented a relative assessment of Singular Spectrum Analysis (SSA) as a forecasting methodology, applied on Greek toll stations road traffic volume real data. The suggested methodology was tested on real data from Moschohorion and Pelasgia Toll Station (Greece), utilizing custom developed forecasting software toolbox. Experimental results revealed that SSA constitute a robust forecasting methodology for short-term forecasting horizon (7 days) but demonstrated a reduced forecasting ability for medium-term forecasts (90 days).

Moreover, the computer processing time for simulation and modeling with SSA was significantly affected by the selected window length (*L*) of the SSA methodology, since the necessary processing time from several seconds in cases of low value of *L* increased (jumped) to some hours in cases of high values of *L*, implying a certain balance arrangement of *L* in order to maintain the characteristics of initial times series, achieve a robust forecasting performance while retaining processing time at reasonable and practical level.

Forecasting road traffic volume is a very challenging engineering problem. Various factors can affect road traffic time series implying a dynamic phenomenon with stochastic real-life performance. Consequently, more systematically research should take place in order to increase forecasting accuracy for short-term and medium- or long-term forecasting horizon. Future research should be focused either on data pre-processing techniques aiming to processing time reduction and forecasting errors diminish, or to propose innovative forecasting methodologies upon the notion of combined or hybrid methodologies [19, 20].

ACKNOWLEDGMENT

The authors are stating their thankfulness to Aegean Motorways S.A. for raw traffic volume data concession.

REFERENCES

[1] Vlahogianni, E.I., Karlaftis, M.G., Golias, J.C. (2008). Temporal evolution of short-term urban traffic flow: A nonlinear dynamics approach. Computer-Aided Civil and Infrastructure Engineering, 23(7): 536-548. https://doi.org/10.1111/j.1467-8667.2008.00554.x

[2] Vlahogianni, E.I., Karlaftis, M.G., Golias, J.C. (2014). Short-term traffic forecasting: Where we are and where we're going. Transportation Research Part C: Emerging Technologies, 43(1): 3-19. https://doi.org/10.1016/j.trc.2014.01.005

[3] Profillidis, V.A., Botzoris, G.N. (2018). Modeling of Transport Demand: Analyzing, Calculating, and Forecasting Transport Demand, Elsevier, Oxford, UK. https://doi.org/10.1016/C2016-0-00793-3

[4] Kolidakis, S., Botzoris, G., Profillidis, V., Lemonakis, P. (2019). Road traffic forecasting-A hybrid approach combining artificial neural network with singular spectrum analysis. Economic Analysis and Policy, 64: 159-171. https://doi.org/10.1016/j.eap.2019.08.002

[5] Marques, C.A.F., Ferreira, J.A., Rocha, A., Castanheira, J.M., Melo-Gonçalves, P., Vaz, N., Dias, J.M. (2006). Singular spectrum analysis and forecasting of hydrological time series. Physics and Chemistry of the Earth, Parts A/B/C, 31(18): 1172-1179. https://doi.org/10.1016/j.pce.2006.02.061

[6] Hassani, H. (2007). Singular spectrum analysis: methodology and comparison. Journal of Data Science, 5(2): 239-257.

[7] Alexandrov, T. (2009). A method of trend extraction using singular spectrum analysis. RevStat-Statistical Journal, 7(1): 1-22.

[8] Hassani, H., Thomakos, T. (2010). A review on singular spectrum analysis for economic and financial time series.

Statistics and Its Interface, 3(3): 377-397. https://dx.doi.org/10.4310/SII.2010.v3.n3.a11
[9] Briceño, H., Rocco, C.M., Zio, E. (2013). Singular spectrum analysis for forecasting electric load demand. Chemical Engineering Transactions, 33: 919-924. https://doi.org/10.3303/CET1333154
[10] Álvarez-Meza, A.M., Acosta-Medina, C., Castellanos - Domínguez, G. (2013). Automatic singular spectrum analysis for time-series decomposition. European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges, Belgium, 131-136.
[11] Hassani, H., Webster, A., Silva, E.S., Heravi, S. (2015). Forecasting US tourist arrivals using optimal singular spectrum analysis. Tourism Management, 46: 322-335. http://dx.doi.org/10.1016/j.tourman.2014.07.004
[12] Mahmoudvand, R., Alehosseini, F., Rodrigues, P.C. (2015). Forecasting mortality rate by singular spectrum analysis. RevStat -Statistical Journal, 13(3): 193-206.
[13] Silva, E.S., Hassani, H. (2015). On the use of singular spectrum analysis for forecasting US trade before, during and after the 2008 recession. International Economics, 141: 34-49. https://doi.org/10.1016/j.inteco.2014.11.003
[14] Broomhead, D.S., King, P.K. (1986). Extracting qualitative dynamics from experimental data. Physica D: Nonlinear Phenomena, 20(2-3): 217-236. https://doi.org/10.1016/0167-2789(86)90031-X
[15] Golyandina, N., Nekrutkin, V., Zhigljavsky, A.A. (2001). Analysis of Time Series Structure: SSA and Related Techniques. Chapman and Hall/CRC, UK. https://doi.org/10.1201/9780367801687
[16] Hassani, H., Zhigljavsky, A. (2009). Singular spectrum analysis: methodology and application to economics data. Journal of Systems Science and Complexity, 22(3): 372-394. https://doi.org/10.1007/s11424-009-9171-9
[17] Golyandina, N., Zhigljavsky, A. (2013). Singular Spectrum Analysis for Time Series, Springer, London, UK. http://dx.doi.org/10.1007/978-3-642-34913-3
[18] Golyandina, N., Korobeynikov, A. (2014). Basic singular spectrum analysis and forecasting with R. Computational Statistics and Data Analysis, 71: 934-954. https://doi.org/10.1016/j.csda.2013.04.009
[19] Kolidakis, S., Botzoris, G., Profillidis, V., Kokkalis, A. (2020). Real-time intraday traffic volume forecasting – a hybrid application using singular spectrum analysis and artificial neural networks. Periodica Polytechnica Transportation Engineering, 48(3): 226-235. https://doi.org/10.3311/PPtr.14122
[20] Hassani, H., Yeganegi, M.R., Huang, X. (2021) Fusing nature with computational science for optimal signal extraction. Stats 2021, 4(1): 71-85. https://doi.org/10.3390/stats4010006