Electricity Demand Loads Modeling using AutoRegressive Moving Average (ARMA) models

Article *in* Energy · September 2008

DOI: 10.1016/j.energy.2008.05.008

Energy 33 (2008) 1353–1360

Contents lists available at ScienceDirect

Energy

journal homepage: www.elsevier.com/locate/energy

Electricity demand loads modeling using AutoRegressive Moving Average (ARMA) models

S.Sp. Pappas [a], L. Ekonomou [b,*], D.Ch. Karamousantas [c], G.E. Chatzarakis [b], S.K. Katsikas [d], P. Liatsis [e]

[a] Department of Information and Communication Systems Engineering, University of the Aegean, Karlovassi, 83 200 Samos, Greece
[b] Department of Electrical Engineering Educators, ASPETE—School of Pedagogical and Technological Education, N. Heraklion, 141 21 Athens, Greece
[c] Technological Educational Institute of Kalamata, Antikalamos, 24100 Kalamata, Greece
[d] Department of Technology Education & Digital Systems, University of Piraeus, 150 Androutsou Srt., 18 532 Piraeus, Greece
[e] Division of Electrical Electronic and Information Engineering, School of Engineering and Mathematical Sciences, Information and Biomedical Engineering Centre, City University, Northampton Square, London EC1V 0HB, UK

ARTICLE INFO

Article history:
Received 10 March 2008

Keywords:
Adaptive multi-model filtering
Electricity demand load
ARMA
Order selection
Parameter estimation
Kalman filter

ABSTRACT

This study addresses the problem of modeling the electricity demand loads in Greece. The provided actual load data is deseasonilized and an AutoRegressive Moving Average (ARMA) model is fitted on the data off-line, using the Akaike Corrected Information Criterion (AICC). The developed model fits the data in a successful manner. Difficulties occur when the provided data includes noise or errors and also when an on-line/adaptive modeling is required. In both cases and under the assumption that the provided data can be represented by an ARMA model, simultaneous order and parameter estimation of ARMA models under the presence of noise are performed. The produced results indicate that the proposed method, which is based on the multi-model partitioning theory, tackles successfully the studied problem. For validation purposes the produced results are compared with three other established order selection criteria, namely AICC, Akaike's Information Criterion (AIC) and Schwarz's Bayesian Information Criterion (BIC). The developed model could be useful in the studies that concern electricity consumption and electricity prices forecasts.

1. Introduction

A significant issue for electricity markets is the management of demand load variability. In addition to placing stress on the physical transmission network, variability in electricity consumption is important for the financial aspects of the system since it influences everything from the current spot price to long-term investment decisions in base of expanding the existing plants. Load forecast has become increasingly important since the rise of the competitive energy markets. Many forecast models have been proposed and implemented in this field including linear regression [1] and econometric models [2], neuro-fuzzy models [3] and data mining procedures [4]. Artificial intelligent techniques have been also applied [5,6], while simple AutoRegressive (AR) [7], AutoRegressive Integrated Moving Average (ARIMA) [8] and AutoRegressive Moving Average (ARMA) [9–12] models have presented very good forecasted electricity demand load and electricity prices results.

The aim of this paper is not to add yet another ARMA model selection criterion to the rich literature in this area, rather to focus firstly on whether the electricity loads in the Hellenic power market can be modeled by an ARMA process and secondly on a comparison of ARMA model order selection criteria under the presence of noise. The current study also presents a new method for multivariate (MV) ARMA model order selection and parameter estimation based on the adaptive multi-model partitioning theory [13–15]. The proposed method which is not restricted to the Gaussian case, it is applicable to on-line/adaptive operation and it is computationally efficient and identifies very fast the correct model order.

2. Data preparation

The actual load data used in this work has been provided by the Hellenic Public Power Corporation S.A. [16], being a time series that contains the daily electricity demand load covering the period from January 1, 2004 to December 31, 2005. Figs. 1 and 2 clearly indicate the weekly and annual seasonality of the provided data. The first necessary step that must be done is to remove this feature. This is achieved by applying to the data a new technique

* Corresponding author. Tel.: +30 697 270 2218; fax: +30 210 772 3504.
E-mail address: leekonom@gmail.com (L. Ekonomou).

0360-5442/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.energy.2008.05.008

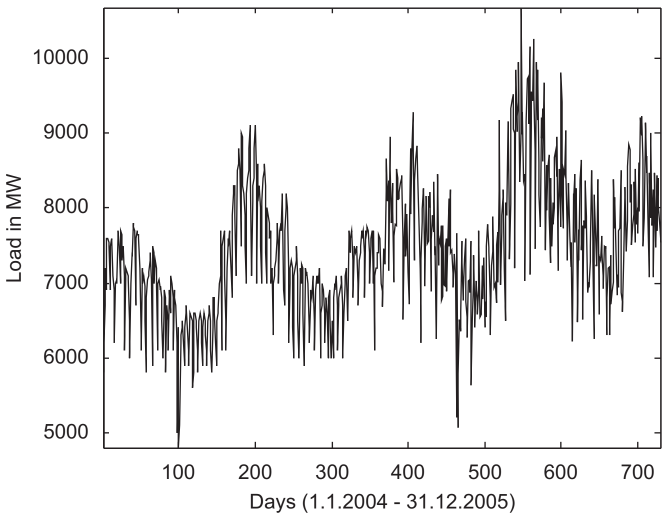

Fig. 1. Hellenic power market daily system-wide load from January 1, 2004 to December 31, 2005. The annual seasonality, the weekly seasonality and the increase in the load demand are clearly visible.

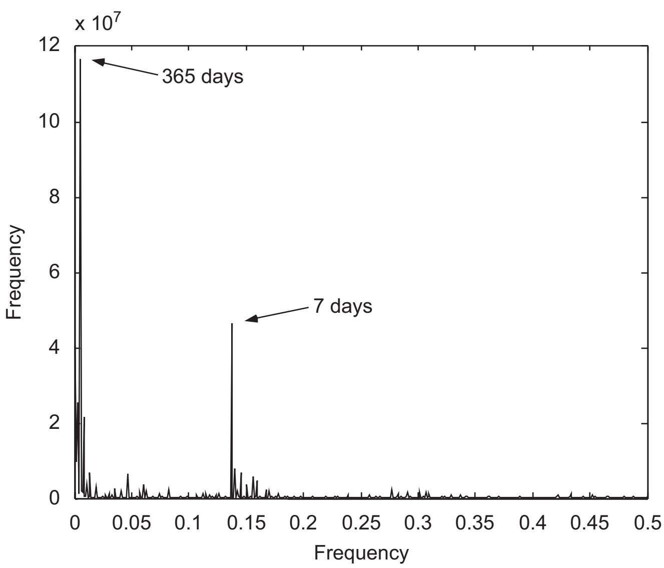

Fig. 2. Periodogram of the Hellenic power market daily system-wide load from January 1, 2004 to December 31, 2005. The annual and weekly frequencies are clearly visible.

analytically introduced by Nowicka-Zagrajek and Weron [10,11] and is briefly described as follows:

- Assume a vector of daily loads: $\{x_1, \ldots, x_{730}\}$
- Apply a moving average filter for the elimination of the weekly component, $\hat{m}_t = (1/7)(x_{t-3} + \cdots + x_{t+3})$ where $t = 4, 5, \ldots, 727$
- Estimation of the Seasonal component, $\hat{s}_k = w_k - (1/7)\sum_{i=1}^{7} w_i$, where $k = 1, 2, \ldots, 7$ and $\hat{s}_k = \hat{s}_{k-7}$ for $k > 7$, w_k is the average of the deviations $\{(x_{k+7j} - \hat{m}_{k+7j}), 3 < k+7j \leqslant 727\}$
- Definition of the deseasonilized data, $d_t = x_t - \hat{s}_t$, for $t = 1, \ldots, 730$
- Remove the trend from the $\{d_t\}$ by taking the logarithmic returns, $r_t = \log(d_{t+1}/d_t)$, $t = 1, \ldots, 730$
- Remove annual seasonality by assigning $\overline{R}_t = (1/25)\sum_{i=0}^{24} R_{t+1}$, $t = 1, \ldots, 730$
  - (i) Calculation of the 25-day rolling volatility, $v_t = \sqrt{(1/24)\sum_{i=0}^{24}(R_{t+1} - \overline{R}_t)^2}$, the vector of returns $\{R_t\}$ is such that $R_1 = R_2 = \cdots = R_{12} = r_1$, $R_{12+t} = r_t$ for $t = 1, \ldots, 730$ and $R_{743} = R_{744} = \cdots = R_{754} = r_{730}$
  - (ii) Calculation of the average volatility for 1 year, $\overline{v}_t = (v_t^{2004} + v_t^{2004}/2)$
  - (iii) Smooth the volatility by taking a 25-day moving average of $\overline{v}_t$
  - (iv) Rescale the returns by dividing them by a smoothed the annual volatility.

Furthermore, the Modeling and Forecasting Electricity (MFE) loads and prices toolbox for Matlab® has been used in this study provided by Weron [17].

3. Modeling with ARMA processes

The resulting time series (Figs. 3 and 4) show no apparent trend or seasonality. It can be considered as a realization of a stationary process. In addition to that, both the Auto Correlation

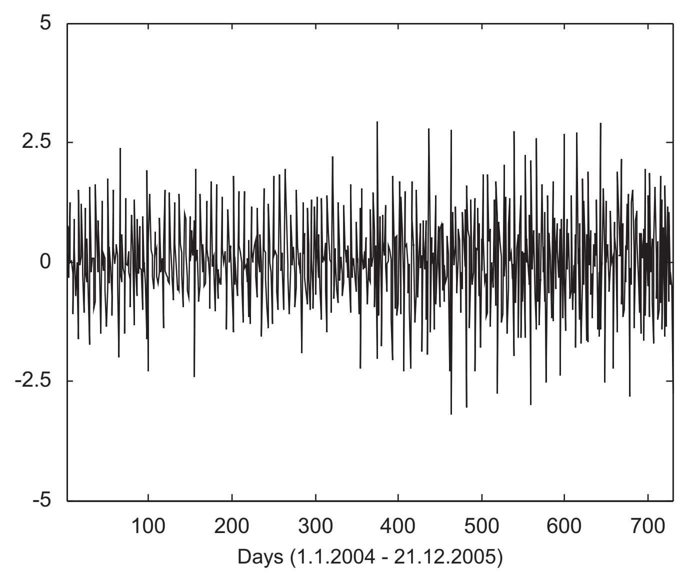

Fig. 3. Load returns after removal of the weekly and seasonal cycles.

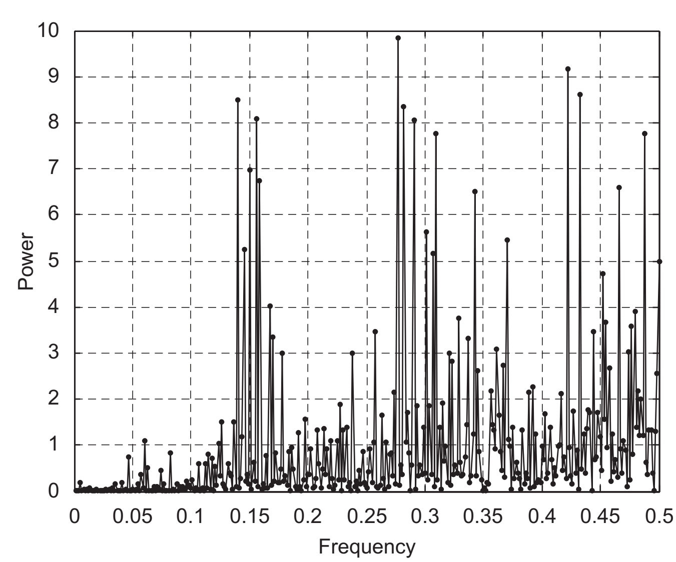

Fig. 4. Periodogram of the load returns after removal of the weekly and annual cycles. No dominating frequency can be observed.

Function (ACF) and Partial Auto Correlation Function (PACF) tend very fast to zero (Figs. 5 and 6) showing that the deseasonilized data returns can be modeled by an AR or an ARMA model.

An m-variate ARMA model of order (p, q) [ARMA (p, q)] for a stationary time series of vectors $\mathbf{y}$ observed at equally spaced instants $k = 1, 2, n$ is defined as:

$$\mathbf{y}_k = \sum_{i=1}^{p} \mathbf{A}_i \mathbf{y}_{k-i} + \sum_{j=1}^{q} \mathbf{B}_j \mathbf{v}_{k-j} + \mathbf{v}_k, \quad E[\mathbf{v}_k \mathbf{v}_k^{\mathrm{T}}] = \mathrm{R}, \tag{1}$$

where the m-dimensional vector $\mathbf{v}_k$ is uncorrelated random noise, not necessarily Gaussian, with zero mean and covariance matrix $\mathbf{R}$, $\theta = (p, q)$ is the order of the predictor and $\mathbf{A}_1, \ldots, \mathbf{A}_p, \mathbf{B}_1, \ldots, \mathbf{B}_q$ are the $m \times m$ coefficient matrices of the MV ARMA model. It is obvious that the problem is two-fold. The first and probably the most important task is the successful determination of the predictor's order $\theta = (p, q)$ and the second task is the estimation of the predictor's matrix coefficients $\{\mathbf{A}_i, \mathbf{B}_j\}$.

Determining the order of the ARMA process is usually the most important part of the problem. Over the past years, substantial literature has been produced for this problem and various different criteria, such as Akaike's, Rissanen's, Schwarz's, Wax's [18–22], have been proposed to implement the order selection process. Using the real data provided *off-line* ARMA model order, $\theta = (p, q)$ identification and parameter estimation was accomplished by minimizing the Akaike's Corrected Information Criterion (AICC):

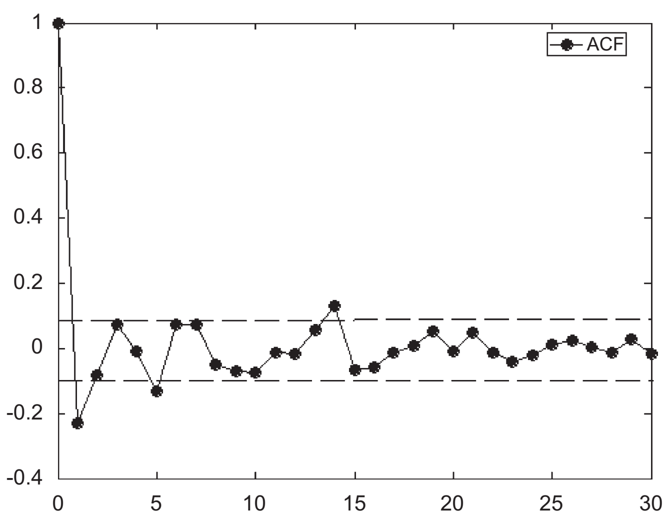

Fig. 5. The ACF for the mean corrected deseasonilized load returns. Dashed lines represent the bounds of $\pm 1.96/\sqrt{730}$, i.e. the 95% confidence intervals of Gaussian white noise.

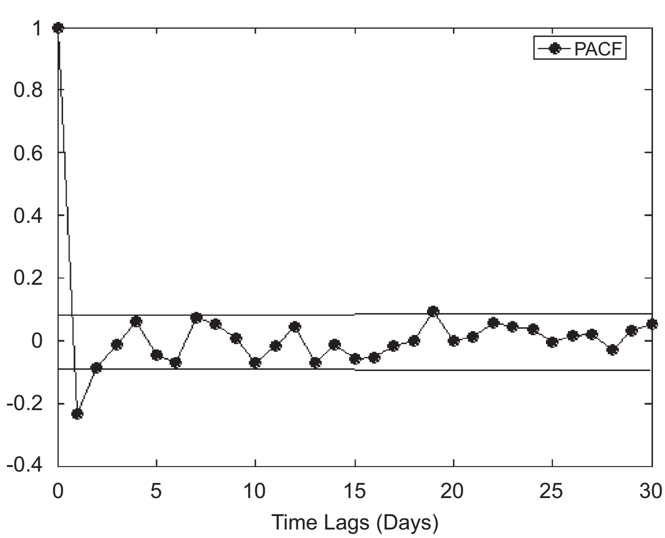

Fig. 6. The PACF for the mean deseasonilized load returns. Solid lines represent the bounds of $\pm 1.96/\sqrt{730}$, i.e. the 95% confidence intervals of Gaussian white noise.

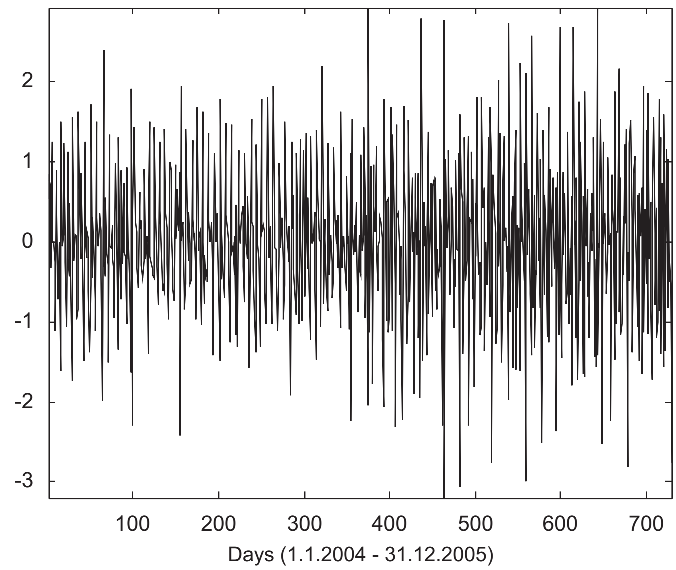

Fig. 7. Residuals obtained from the ARMA(2, 6) model.

$$\mathrm{AICC} = \log|\hat{\mathbf{R}}_\theta| + \frac{2(p+q+1)n}{n-p-q-2}, \tag{2}$$

where $n = 730$ is the sample size and $\theta\ (p, q)$ is the model order and $\hat{\mathbf{R}}_\theta$ is a maximum likelihood estimate of $\mathbf{R}$ under the assumption that ARMA(θ) is the correct model order [23].

The optimization procedure led us to the following ARMA (2, 6) model with parameters:

$$\mathbf{y}_k = \sum_{i=1}^{2} \mathbf{A}_i \mathbf{y}_{k-i} + \sum_{j=1}^{6} \mathbf{B}_j \mathbf{v}_{k-j} + \mathbf{v}_k, \tag{3}$$

where $\mathbf{A}_1 = [-0.3153]$, $\mathbf{A}_2 = [0.0287]$, $\mathbf{B}_1 = [-0.1363]$, $\mathbf{B}_2 = [0.1985]$, $\mathbf{B}_3 = [0.1604]$, $\mathbf{B}_4 = [0.4057]$, $\mathbf{B}_5 = [0.4211]$, $\mathbf{B}_6 = [0.2417]$. The value of the AICC criterion obtained for this model was AICC = 1821.362. It must be mentioned that the statistical significance of all estimated parameters was 5%.

The residuals obtained from the ARMA(2, 6) model fit satisfactorily the deseasonilized load returns (Fig. 7). The graph shows no indication of a non-zero or non-constant variance. In addition to that, both ACF and PACF fall between the calculated bounds proving that there is no correlation in the series (Figs. 8 and 9). Consequently, there is no reason to reject the fitted model on the basis of the autocorrelation or partial autocorrelation function.

4. Problem reformulation using multi-model partitioning filter (MMPF)

The model used in this study is a univariate model. The new algorithm for ARMA model order and parameter estimation proposed concerns the general MV case. The method is easily transformed to univariate by just substituting the matrix coefficients $\{\mathbf{A}_i, \mathbf{B}_j\}$ with real numbers, by setting dimensionality $m = 1$ and re-writing Eqs. (4)–(11) appropriately without the vector notation (bold style).

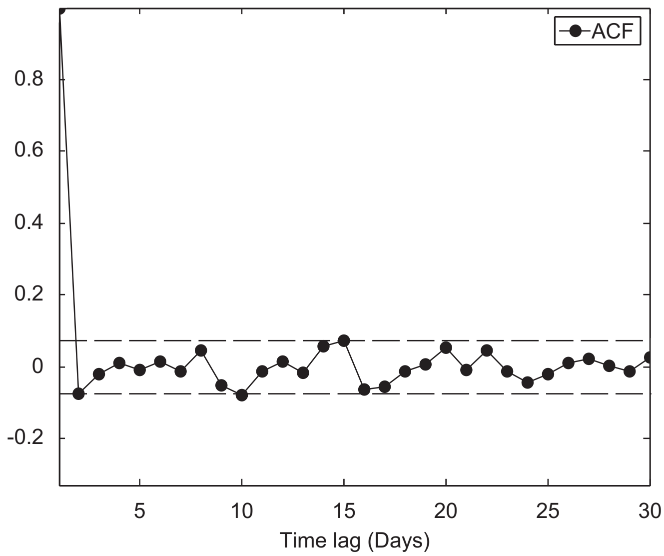

Fig. 8. The ACF for the obtained ARMA(2, 6) model. Dashed lines represent the bounds of $\pm 1.96/\sqrt{730}$, i.e. the 95% confidence intervals of Gaussian white noise.

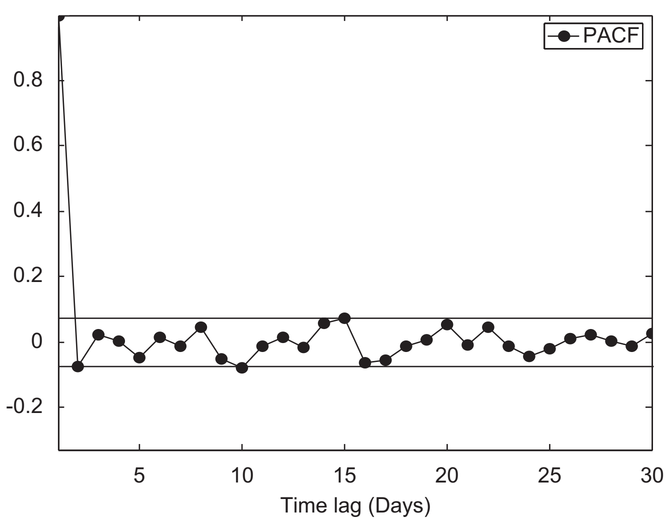

Fig. 9. The PACF for the obtained ARMA(2, 6) model. Solid lines represent the bounds of $\pm 1.96/\sqrt{730}$, i.e. 95% confidence intervals of Gaussian white noise.

Assuming that the model order fitting the data is known and is equal to $\theta = (p, q)$, Eq. (1) can be written in standard state-space form as:

$$\mathbf{x}(k+1) = \mathbf{x}(k), \tag{4}$$

$$\mathbf{y}(k) = \mathbf{H}(k)\mathbf{x}(k) + \mathbf{v}(k), \tag{5}$$

where $\mathbf{x}(k)$ is an $m^2(p+q) \times 1$ vector made up from the coefficients of the matrices $\{\mathbf{A}_1, \ldots, \mathbf{A}_p, \mathbf{B}_1, \ldots, \mathbf{B}_q\}$, and $\mathbf{H}(k)$ is an $m \times m^2(p+q)$ observation history matrix of the process $\{\mathbf{y}(k)\}$ up to time $k-(p+q)$.

Assuming that the general forms of the matrices $\mathbf{A}_p$ and $\mathbf{B}_q$ are as follows:

$$\begin{bmatrix} a^p_{11} & \ldots & a^p_{1m} \\ \vdots & \ddots & \vdots \\ a^p_{m1} & \cdots & a^p_{mm} \end{bmatrix}, \tag{6}$$

$$\begin{bmatrix} b^q_{11} & \ldots & b^q_{1m} \\ \vdots & \ddots & \vdots \\ b^q_{m1} & \cdots & b^q_{mm} \end{bmatrix}, \tag{7}$$

$$\begin{aligned}\mathbf{x}(k) \triangleq [&\alpha^1_{11}\alpha^1_{21}\cdots\alpha^1_{m1} \vdots \alpha^1_{12}\alpha^1_{22}\cdots\alpha^1_{m2} \vdots \cdots \alpha^1_{mm} \vdots \cdots \alpha^p_{mm} \vdots \\ &b^1_{11}b^1_{21}\cdots b^1_{m1} \vdots b^1_{12}b^1_{22}\cdots b^1_{m2} \vdots \cdots b^1_{mm} \vdots \cdots b^q_{mm}]^{\mathrm{T}},\end{aligned} \tag{8}$$

$$\begin{aligned}\mathbf{H}(k) \triangleq [&y_1(k-1)I \cdots y_m(k-1)I \vdots \cdots \vdots y_1(k-p)I \cdots \\ &y_m(k-p)I \vdots v_1(k-1)I \cdots v_m(k-1)I \vdots \cdots \vdots \\ &v_1(k-q)I \cdots v_m(k-q)I],\end{aligned} \tag{9}$$

where $\mathbf{I}$ is the $m \times m$ identity matrix and $\theta = (p, q)$, is the model order.

In case that the system model and its statistics were completely known, the Kalman filter (KF) in its various forms would be the optimal estimator in the minimum variance sense. Moreover, in case that the prediction coefficients are subject to random perturbations Eq. (4) becomes:

$$\mathbf{x}(k+1) = \mathbf{x}(k) + \mathbf{w}(k), \tag{10}$$

where $\mathbf{v}(k)$, $\mathbf{w}(k)$ are independent, zero-mean, white processes, not necessarily Gaussian.

The form of $\mathbf{w}(k)$ is:

$$\begin{aligned}\mathbf{w}(k) \triangleq [&w^1_{11}w^1_{21}\cdots w^1_{m1} \vdots w^1_{12}w^1_{22}\cdots w^1_{m2} \vdots \cdots w^1_{mm} \vdots \cdots w^p_{mm} \vdots \\ &w^1_{11}w^1_{21}\cdots w^1_{m1} \vdots w^1_{12}w^1_{22}\cdots w^1_{m2} \vdots \cdots w^1_{mm} \vdots \cdots w^q_{mm}]^{\mathrm{T}}.\end{aligned} \tag{11}$$

A complete system description requires the value assignments of the variances of the random processes $\mathbf{w}(k)$ and $\mathbf{v}(k)$. Adopting the usual assumption that $\mathbf{w}(k)$ and $\mathbf{v}(k)$ at least wide sense stationary processes, hence their variances, $\mathbf{Q}$ and $\mathbf{R}$, respectively, are time invariant. To obtain these values is not always trivial. If $\mathbf{Q}$ and $\mathbf{R}$ are not known they can be estimated by using a method such as the one described in Ref. [24]. In the case of coefficients constant in time, or slowly varying, $\mathbf{Q}$ is assumed to be zero.

It is also necessary to assume an a priori mean and variance for each $\{\mathbf{A}_i, \mathbf{B}_j\}$. The a priori mean of the $\mathbf{A}_i(\mathbf{0})$'s and $\mathbf{B}_j(\mathbf{0})$'s can be set to zero if no knowledge about their values is available before any measurements are taken (the most likely case). On the other hand, the usual choice of the initial variance of the $\mathbf{A}_i$'s and $\mathbf{B}_j$'s, denoted by $\mathbf{P}_0$ is $\mathbf{P}_0 = n\mathbf{I}$, where n is a large integer.

Considering the case where the system model is not completely known the adaptive MMPF is one of the most widely used approaches for similar problems. This approach was introduced by Lainiotis [13–15] and summarizes the parametric model uncertainty into an unknown, finite dimensional parameter vector whose values are assumed to lie within a known set of finite cardinality. A non-exhaustive list of the reformulation, extension and application of the MMPF approach as well as its application to a variety of problems can be found in Refs. [25–32]. In the studied problem in this paper, it is assumed that the model uncertainty is the lack of knowledge of the model order θ. It is also assumed that the model order θ lies within a known set of finite cardinality: $1 \leqslant \theta \leqslant M$, where $\theta = (p, q)$, is the model order.

The MMPF operates on the following discrete model:

$$\mathbf{x}(k+1) = \mathbf{F}(k+1, k/\theta)\mathbf{x}(k) + \mathbf{w}(k), \tag{12}$$

$$\mathbf{y}(k) = \mathbf{H}(k/\theta)\mathbf{x}(k) + \mathbf{v}(k), \tag{13}$$

where $\theta = (p, q)$ is the unknown parameter, the model order in this case. A block diagram of the MMPF is presented in Fig. 10.

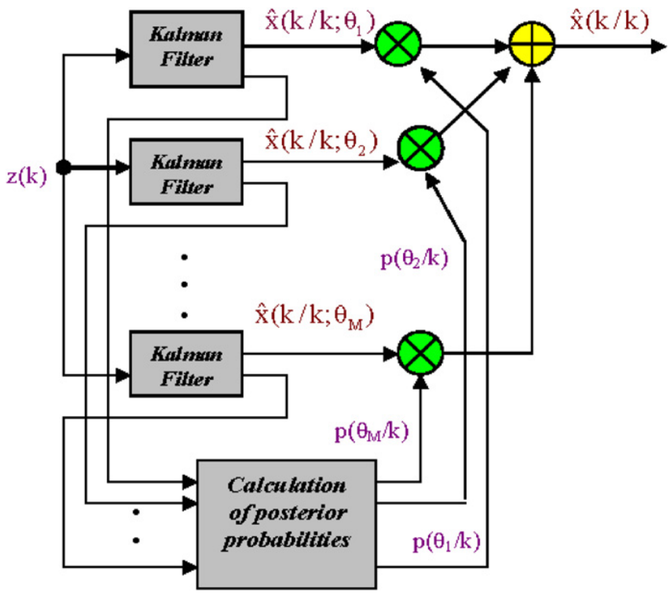

Fig. 10. MMPF block diagram.

In the Gaussian case the optimal MMSE estimate of $\mathbf{x}(k)$ is given by:

$$\hat{\mathbf{x}}\frac{k}{k} = \sum_{j=1}^{M} \hat{\mathbf{x}}\left(\frac{k}{k};\theta_j\right) p\left(\frac{\theta_j}{k}\right), \tag{14}$$

finite set of models is designed, each matching one value of the parameter vector. In the case that the prior probabilities $p(\theta_j/k))$ for each model are already known, these are assigned to each model. In the absence of any prior knowledge, these are set to $p(\theta_j/k)) = 1/M$ where M is the cardinality of the model set.

A bank of conventional elemental filters (non-adaptive, e.g. Kalman) is then applied, one for each model, which can be run in parallel. At each iteration, the MMPF selects the model which corresponds to the maximum posteriori (MAP) probability as the correct one. This probability tends to 1, while the others tend to 0. The overall optimal estimate can be taken either to be the individual estimate of the elemental filter exhibiting the highest posterior probability, called the MAP estimate [30], the case used in this paper, or the weighted average of the estimates produced by the elemental filters, as described in Eq. (14). The weights are determined by the posterior probability that each model in the model set is in fact the true model. The posterior probabilities are calculated on-line in a recursive manner as follows:

$$p(\theta_j/k) = \frac{L(k/k;\theta_j)}{\sum_{j=1}^{M} L(k/k;\theta_j)p(\theta_j/k-1)} p(\theta_j/k-1), \tag{15}$$

$$\begin{aligned} L\left(\frac{k}{k};\theta_j\right) = {} & \left|\mathbf{P}_{\tilde{y}}\left(\frac{k}{k-1};\theta_j\right)\right|^{-1/2} \\ & \times \exp\left[-\frac{1}{2}\tilde{\mathbf{y}}^{\mathrm{T}}\left(\frac{k}{k-1};\theta_j\right)\right. \\ & \left.\times \mathbf{P}_{\tilde{y}}^{-1}\left(\frac{k}{k-1};\theta_j\right)\tilde{\mathbf{y}}\left(\frac{k}{k-1};\theta_j\right)\right], \end{aligned} \tag{16}$$

where the innovation process:

$$\tilde{\mathbf{y}}\left(\frac{k}{k-1};\theta_j\right) = \mathbf{y}(k) - \mathbf{H}(k;\theta_j)\hat{\mathbf{x}}\left(\frac{k}{k-1};\theta_j\right), \tag{17}$$

is a zero mean white process with covariance matrix:

$$\mathbf{P}_{\tilde{y}}\left(\frac{k}{k-1};\theta_j\right) = \mathbf{H}(k;\theta_j)\mathbf{P}\left(\frac{k}{k};\theta_j\right)\mathbf{H}^{\mathrm{T}}(k;\theta_j) + \mathbf{R}. \tag{18}$$

It must be mentioned that in Eqs. (9)–(18) the value of $j = 1, 2, \ldots, M$. An important feature of the MMPF is that all the Kalman filters needed to be implemented can be independently realized. This enables to implement them in parallel, saving an enormous computational time [30]. Eqs. (14) and (15) refer to the current case where the sample space is naturally discrete. However, in real world applications, θ's probability density function (pdf) is continuous and an infinite number of Kalman filters have to be applied for the exact realization of the optimal estimator. The usual approximation considered to overcome this difficulty is to approximate θ's pdf by a finite sum. Many discretization strategies have been proposed in the literature and some of them are presented in Refs. [33,34]. When the true parameter value lies outside the assumed sample space, the adaptive estimator converges to the value that in the sample space which is closer (i.e. minimizes the Kullback Information Measure) to the true value [35]. This means that the value of the unknown parameter cannot be exactly defined. The application of hybrid techniques that combine the MMPF with Genetic Algorithms are able to overcome this difficulty [27,36].

5. Application

The problem of ARMA modeling is much more difficult when an adaptive on-line procedure is required and when noise is present. The earlier mentioned criteria, Akaike's, Rissanen's, Schwarz's, and Wax's, [18–22], are not always optimal and are also known to suffer from deficiencies; for example, Akaike's information criterion (AIC) suffers from over fit [37]. Also their performance depends on the assumption that the data is Gaussian and upon asymptotic results. In addition to this, their applicability is justified only for large samples; furthermore, they are two pass methods, so they cannot be used in an on-line/adaptive fashion.

The actual demand load data provided (Fig. 3) is correlated with a significant amount of noise with covariance $R = [1.25]$. As it can be seen from Fig. 11 (bottom panel), the new data is obviously different than the original one (upper panel). A comparison amongst the following criteria will be performed.

- AICC, see Eq. (2)
- AIC

$$\log(|\hat{\mathbf{R}}_\theta|) + \frac{2(p+q)}{n} \tag{19}$$

- Bayesian Information Criterion (BIC)

$$n\log(|\hat{\mathbf{R}}_\theta|) + (p+q)\log(n) \tag{20}$$

- MMPF

Fig. 12 shows that the proposed MMPF finds the correct model order ($\theta = 8(p+q)$) very fast in only 21 steps and using only 50 samples as an initial data set. Convergence is taken to occur when the posterior probability of the model exceeds 0.9. Note that the abscissa in Fig. 12 begins from the 10th time instant. This is because of the initial assumption that the unknown model order θ lies between [1, M], were $M = 10$. Consequently, the first M samples of the data set are used for algorithm initialization. Fig. 13 shows that the MMPF is consistent and 100% successful in correct model order identification for all three data sets. The only criterion that meets its performance is BIC but only for

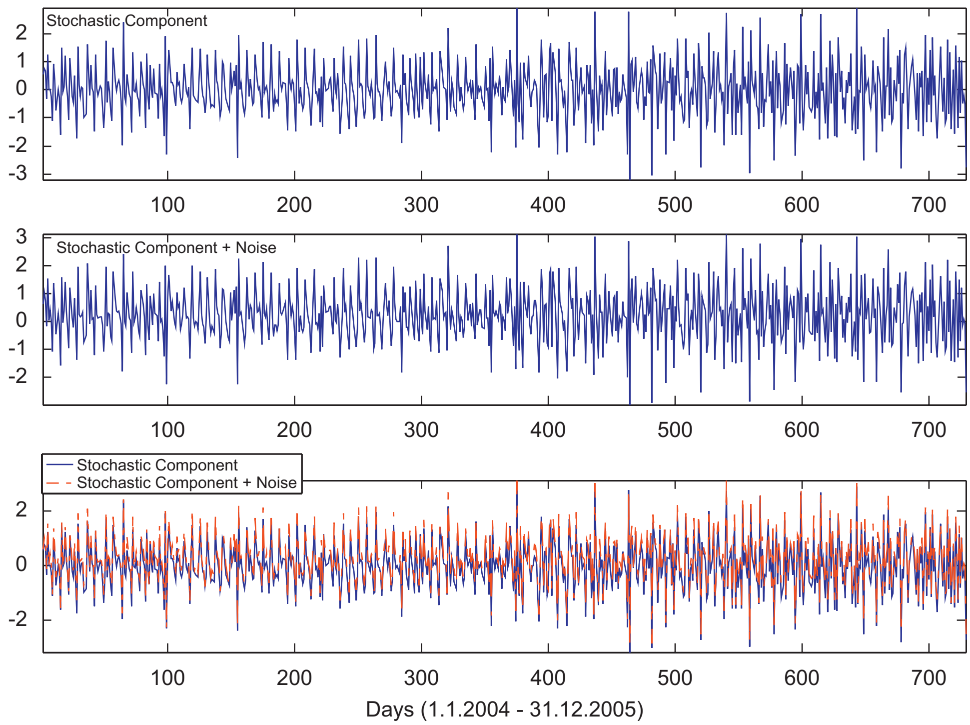

Fig. 11. Noisy set of data used for the criteria comparison.

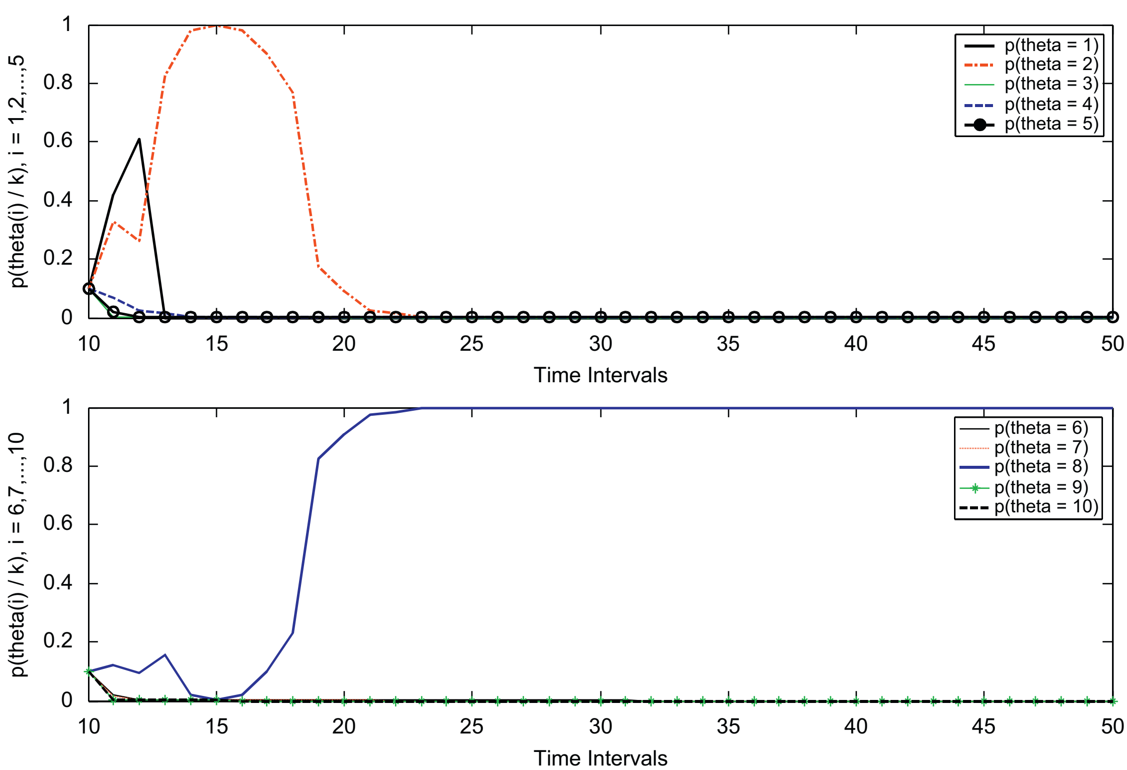

Fig. 12. Posterior probability sequence (sample size 50).

the largest data set (300 samples). BIC is also the approach that requires the less computational time, after the proposed one (MMPF), with AICC to follow and finally the AIC approach which is the slowest of all.

6. Error analysis

Table 1 shows the estimated parameters with statistical significance 5%, using a noisy input data set (Fig. 11) and Table 2 the associated RMSE for each criterion. The later is given by

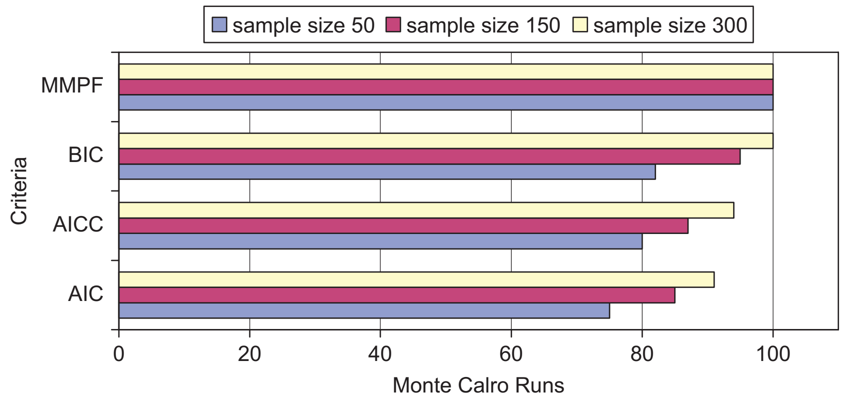

Fig. 13. Criteria comparison—number of correct model order identifications out of 100 Monte Carlo runs.

Table 1
True parameters, estimated parameters (mean values)

Index	True parameters	MMPF estimated parameters	AICC estimated parameters	AIC estimated parameters	BIC estimated parameters
A_1	−0.3153	−0.3636	−0.3765	−0.5765	−0.3656
A_2	0.0287	0.0286	0.0307	0.0472	0.0297
B_1	−0.1363	−0.1297	−0.1243	−0.1103	−0.1143
B_2	0.1985	0.1902	0.2075	0.1875	0.2175
B_3	0.1604	0.1656	0.1565	0.1708	0.1506
B_4	0.4057	0.4103	0.3891	0.3621	0.3835
B_5	0.4211	0.4192	0.4183	0.4313	0.4134
B_6	0.2417	0.2429	0.2391	0.2512	0.2342

Table 2
Root mean square error (RMSE) associated to each criterion

Index	True parameters	MMPF RMSE	AICC RMSE	AIC RMSE	BIC RMSE
A_1	−0.3153	0.1223	0.1302	0.1403	0.1357
A_2	0.0287	0.0692	0.1029	0.1215	0.1109
B_1	−0.1363	0.1016	0.1082	0.1224	0.1132
B_2	0.1985	0.0653	0.0869	0.1162	0.1093
B_3	0.1604	0.0972	0.1161	0.1295	0.1225
B_4	0.4057	0.1002	0.1249	0.1337	0.1281
B_5	0.4211	0.0142	0.0469	0.1036	0.0542
B_6	0.2417	0.0361	0.0915	0.1007	0.0973

the formula:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{m}(\hat{x}_i - x_i)}{m}}, \qquad (21)$$

where $\hat{x}_i$ is the predicted estimate of the proposed method and x_i the true value.

It must be mentioned that the initial input data set was a noisy set (obtained from the behavior of the true ARMA model corrupted with noise) and the result was to obtain an estimate model very close to the real one.

7. Conclusions

The proposed method (MMPF) successfully selects the correct model order in very few steps and identifies very accurately the ARMA parameters. Comparison with other established order selection criteria (AIC, AICC and BIC) show that the method needs the shortest data set for successful order identification and accurate parameter estimation whereas the other criteria require longer data in order either to achieve the same performance (BIC for the 300 data set) or to attain a performance greater than 90%. As a further step to this study would be the optimization of the algorithm in order to be able to identify the order of the AR component (p) and MA component (q) separately and to apply the algorithm for the on-line prediction of the load. This work can be useful in the studies that concern electricity consumption and electricity prices forecasts giving the possibility to the electricity providers, retailers and regulatory authorities to supply uninterrupted energy at a low cost.

Acknowledgments

This paper is dedicated to the memory of Prof. Dimitrios. G. Lainiotis, the founder of the Multi-Model Partitioning Theory, who suddenly passed away in 2006. The authors would also like to express their gratitude to the Hellenic Public Power Corporation S.A. for the supply of various data.

References

[1] Mohamed Z, Bodger P. Forecasting electricity consumption in New Zealand using economic and demographic variables. Energy 2005;30:1833–43.
[2] Yang M, Yu X. China's rural electricity market-a quantitative analysis. Energy 2004;29:961–77.
[3] Padmakumari K, Mohandas KP, Thiruvengadam S. Long-term distribution forecasting using neuro-fuzzy computations. Elect Power Energy Syst 1999;21:315–22.
[4] Wu HC, Lu CN. A data mining approach for spatial modeling in small area load forecast. IEEE Trans Power Syst 2002;17:516–21.
[5] Ozturk HK, Ceylan H, Canyurt OE, Hepbasli A. Electricity estimation using genetic algorithm approach: a case study of Turkey. Energy 2005;30:1003–12.
[6] Kermanshahi B, Iwamiya H. Up to the year 2020 load forecasting using neural nets. Elect Power Energy Syst 2002;17:789–97.
[7] Haida T, Muto S. Regression based peak load forecasting using a transformation technique. IEEE Trans Power Syst 1994;9(4):1788–94.
[8] Contreras J, Espinola R, Nogales FJ, Conejo AJ. ARIMA models to predict next-day electricity prices. IEEE Trans Power Syst 2003;18(3):1014–20.
[9] Nogales FJ, Contreras J, Conejo AJ, Espinola R. Forecasting next-day electricity prices by time series models. IEEE Trans Power Syst 2002;17(2):342–8.
[10] Weron R, Kozlowska B, Nowicka-Zagrajek J. Modeling electricity loads in California: a continuous-time approach. Physica A: Stat Mech Appl 2001; 299(1–2):344–50.
[11] Nowicka-Zagrajek J, Weron R. Modeling electricity loads in California: ARMA models with hyperbolic noise. Signal Process 2002;82(12):1903–15.
[12] Swider DJ, Weber C. Extended ARMA models for estimating price developments on day-ahead electricity markets. Elect Power Syst Res 2007;77: 583–93.

[13] Lainiotis DG. Optimal adaptive estimation: structure and parameter adaptation. IEEE Trans Autom Control 1971;AC-16:160–70.
[14] Lainiotis DG. Partitioning: a unifying framework for adaptive systems I: Estimation. Proc IEEE 1976;64(8):1126–43.
[15] Lainiotis DG. Partitioning: a unifying framework for adaptive systems II: Control. Proc IEEE 1976;64(8):1182–98.
[16] PPC S.A. Annual electrical energy's statistical and economical data. Athens: Hellenic Public Power Corporation S.A.; 2006.
[17] Weron R. 〈http://www.im.pwr.wroc.pl/~rweron/MFE.html〉; toolbox downloads; 05.11.2007.
[18] Akaike H. Fitting autoregressive models for prediction. Ann Inst Stat Math 1969;21:243–7.
[19] Rissanen J. Modeling by shortest data description. Automatica 1978; 14:465–71.
[20] Rissanen J. A predictive least squares principle. IMA J Math Contr Inform 1986;3:211–22.
[21] Schwarz G. Estimation of the dimension of the model. Ann Stat 1978;6:461–4.
[22] Wax M. Order selection for AR models by predictive least squares. IEEE Trans Acoust Speech Signal Process 1988;36:581–8.
[23] Chen CC, Davis RA, Brockwell PJ. Order determination for multivariate autoregressive processes using reasampling methods. J Multivariate Anal 1996;57:175–90.
[24] Sage AP, Husa GW. Adaptive filtering with unknown prior statistics. In: Proc. of joint. automatic control conference, Boulder, Colorado, 1969. p. 760–9.
[25] Watanabe K. Adaptive estimation and control: partitioning approach. Englewood Cliffs, NJ: Prentice-Hall; 1992.
[26] Pappas SSp, Leros AK, Katsikas SK. Joint order and parameter estimation of multivariate autoregressive models using multi-model partitioning theory. Digit Signal Process 2006;16(6):782–95.
[27] Katsikas SK, Likothanassis SD, Beligiannis GN, Berkeris KG, Fotakis DA. Genetically determined variable structure multiple model estimation. IEEE Trans Signal Process 2001;49(10):2253–61.
[28] Lainiotis DG, Papaparaskeva P. A partitioned adaptive approach to nonlinear channel equalization. IEEE Trans Commun 1998;46(10):1325–36.
[29] Moussas VC, Likothanassis SD, Katsikas SK, Leros AK. Adaptive on-line multiple source detection. IEEE Int Conf Acoust Speech Signal Process (ICASSP'05) 2005;4:1029–32.
[30] Lainiotis DG, Katsikas SK, Likothanassis SD. Adaptive deconvolution of seismic signals: performance, computational analysis, parallelism. IEEE Trans Acoust Speech Signal Process 1988;36(11):1715–34.
[31] Nikitakos NV, Leros AK, Katsikas SK. Towed array shape estimation using multimodel partitioning filters. IEEE J Oceanic Eng 1998;23(4):380–4.
[32] Moussas VC, Pappas SSp. Adaptive network anomaly detection using bandwidth utilization data. In: 1st international conference on experiments/processes/system modelling/simulation/optimization, Athens, 2005.
[33] Sengbush RL, Lainiotis DG. Simplified parameter quantization procedure for adaptive estimation. IEEE Trans Autom Control 1969;AC-14:424–5.
[34] Anderson BDO, Brinsmead TS, De Bruyne F, Hespanha J, Liberzon D, Morse AS. Multiple model adaptive control, Part 1: finite controller coverings. Int J Robust Nonlinear Contr 2000;10:909–29.
[35] Hawks RM, Moore JB. Performance of Bayesian parameter estimators for linear signal models. IEEE Trans Autom Control 1976;AC-21:523–7.
[36] Beligiannis G, Skarlas L, Likothanassis S. A generic applied evolutionary hybrid technique. IEEE Signal Process Mag 2004;21(3):28–38.
[37] Lutkepohl H. Comparison of criteria for estimating the order of a vector AR process. J Time Ser Anal 1985;6:35–52.

Stylianos Sp. Pappas was born in Athens, Greece on June 22, 1974. He received a Bachelor of Engineering (Hons.) in Electrical Engineering and Electronics in 1997 and a Master of Science in Advanced Control in 1998 from University of Manchester Institute of Science and Technology (U.M.I.S.T) in United Kingdom. In 2008, he graduated with a Ph.D. from the Department of Information and Communication Systems Engineering, University of the Aegean, in Karlovassi of Samos in Greece. His research interest is in the area of Partitioning Theory & their applications, evolutionary algorithms and control design techniques.

Lambros Ekonomou was born on January 9, 1976 in Athens, Greece. He received a Bachelor of Engineering (Hons.) in Electrical Engineering and Electronics in 1997 and a Master of Science in Advanced Control in 1998 from University of Manchester Institute of Science and Technology (U.M.I.S.T.) in United Kingdom. In 2006, he graduated with a Ph.D. in High Voltage Engineering from the National Technical University of Athens (N.T.U.A.) in Greece. He has worked with various companies including Hellenic Public Power Corporation S.A. and Hellenic Aerospace Industry S.A., while in 2008 he was appointed Assistant Professor in ASPETE. His research interests concern high voltage engineering, transmission and distribution lines, lightning performance and protection, stability analysis, control design, reliability, electrical drives and artificial neural networks.

Dimitrios Ch. Karamousantas was born in Kalapodi, Greece, on August 1, 1957. He received the Diploma in Engineering and the Ph.D. degrees in Mechanical Engineering from the Institute Polytechnic CLUJ-NAPOCA, Romania, in 1984 and 1988, respectively. Currently, he is an Associate Professor in the Technological Educational Institute of Kalamata, Greece.

George E. Chatzarakis was born in Serres, Greece, on May 20, 1961. He received the Dipl. Eng. and the Ph.D. degrees in Electrical Engineering from the National Technical University of Athens (NTUA) in 1986 and 1990, respectively. Currently, he is a Professor in the School of Pedagogical and Technological Education (ASPETE) in Athens, Greece.

Sokratis K. Katsikas was born in Athens, Greece, in 1960. He received the Diploma in Electrical Engineering from the University of Patras, Patras, Greece in 1982, the Master of Science in Electrical & Computer Engineering degree from the University of Massachusetts at Amherst, Amherst, USA, in 1984 and the Ph.D. in Computer Engineering & Informatics from the University of Patras, Patras, Greece in 1987. In 1990, he joined the University of the Aegean, Greece, where he served as Professor of the Department of Information & Communication Systems Engineering and as the Rector. In 2007, he joined the Department of Technology Education and Digital Systems of the University of Piraeus, as a Professor. His research interests lie in the areas of information and communication systems security and of estimation theory and its applications. He has authored or co-authored more than 150 journal publications, book chapters and conference proceedings publications in these areas. He is serving on the editorial board of several scientific journals, he has authored/edited 20 books and has served on/chaired the technical programme committee of numerous international conferences.

Panos Liatsis graduated with a Diploma in Engineering in Electrical Engineering from the Democritos University of Thrace, Greece and a Ph.D. in Computer Vision and Neural Networks from the Control Systems Centre at UMIST. In April 1994, he was appointed a Lecturer in the Control Systems Centre at UMIST, while he worked with various industrial partners including British Aerospace, Lucas Industries, and TRW Automotive. In November 2003, he moved to the School of Engineering & Mathematical Sciences at City University, where he is currently a Senior Lecturer and Director of the Information and Biomedical Engineering Centre (IBEC). He is a member of the EPSRC Peer Review College, the IEEE and a European Engineer (Eur Ing). He is a member of various International Conference Programme Committees including the International Conference on Video Processing & Multimedia Communications, EURASIP Conference on Video & Image Processing and the International Workshop on Systems, Signals and Image Processing. His main research interests are neural networks, genetic algorithms, computer vision and pattern recognition. He has published over 80 scientific papers in international journals and conferences and edited two international conference proceedings.