3rd International Conference "From Scientific Computing to Computational Engineering
3rd IC-SCCE
Athens, 9-12 July, 2008

ADAPTIVE AND CLASSICAL TECHNIQUES FOR PREDICTING AND FORECASTING TIME SERIES: CRITERIA COMPARISON

Vassileios. C. Moussas[1], Stylianos. Sp. Pappas*[2], and Socrates. K. Katsikas[3]

[1] School of Technological Applications, Technological Educational Institute of Athens, 122 Ag. Spyridonas St., Egaleo 12210, Greece
[2]Department of Information and Communication Systems Engineering, University of the Aegean, Samos 83200, Greece
[3]Department of Technology Education & Digital Systems, University of Piraeus, 150 Androutsou St., Piraeus 18532, Greece
*e-mail: spappas@aegean.gr, ksiroval@otenet.gr

Keywords: Extended Kalman Filters, Mutlimodel Partition, ARMA, Order Identification, Parameter Estimation.

Abstract. *This study applies and compares classical approaches, specifically Akaike's Information Criterion (AIC), Akaike's Corrected Information Criterion (AICC) and Bayesian Information Criterion (BIC) and an extension of the well known multimodel partitioning algorithm (MMPA) to the time series prediction problem. The time series data is real and represents the monthly thunderstorm days, covering the period from 01/1980 to 12/2005. The data has been collected from the National Meteorological Authority of Hellas. Comparison shows that all the methods produce satisfactory results with the adaptive MMPA having the best performance amongst the compared algorithms.*

1 INTRODUCTION

The problem of identifying and predicting the future behavior of a time series is present for many years. Fitting an ARMA model to the time series data is an efficient way of approaching the previously mentioned tasks and is applied in many scientific fields [1-5]. An essential requirement for this procedure is the order identification and the parameter estimation of the ARMA model. Many theoretical criteria have been proposed for tackling the aforementioned problem. Some of them are the Akaike's Information Criterion (AIC), Akaike's Corrected Information Criterion (AICC), Bayesian information criterion (BIC) [6-10], and many more. Practical use of these techniques has shown that they are not always optimal and that they also suffer from deficiencies [11-14]. Additionally they work efficiently only for Gaussian data and upon asymptotic results. Finally their applicability is justified only for large samples (n). Concluding it should be stated they are two pass methods, so they cannot be used in an on line or adaptive fashion.

This paper is an extension to the work presented in [15-16] which concerns adaptive load forecasting and adaptive network anomaly respectively. The difference here is that the models used are not ARIMA but MV ARMA. The method is based on the well known adaptive multi-model partitioning theory [17-18], it is not restricted to the Gaussian case, is applicable to an on line/adaptive operation and it has been shown to be computationally efficient.

2 PROBLEM REFORMULATION AND CRITERIA PRESENTATION

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

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

where the m-dimensional vector $\boldsymbol{v}_k$ 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 x m coefficient matrices of the MV ARMA model.

Obviously the solution to the problem is the identification of predictor's order $\theta=(p, q)$ and the computation of the predictor's matrix coefficients $\{\mathbf{A}_i, \mathbf{B}_j\}$. With the completion of these tasks, someone may proceed to the prediction of the future behavior of the time series.

Assuming that the model order fitting the data is known and is equal to $\theta=(p, q)$, we can rewrite equation (1)

in standard state-space form as:

$$\mathbf{x}(k+1) = \mathbf{x}(k) \tag{2}$$

$$\mathbf{y}(k) = \mathbf{H}(k)\mathbf{x}(k) + \mathbf{v}(k) \tag{3}$$

Now assign a new variable λ such as λ=max (p, q). Then $\mathbf{x}(k)$ is an m^2 (λ+λ) × *1* vector made up from the coefficients of the matrices {$\mathbf{A}_1$, ..., $\mathbf{A}_\lambda$, $\mathbf{B}_1$, ..., $\mathbf{B}_\lambda$,}, $\mathbf{H}(k)$ is an $m \times m^2$ (λ+λ) observation history matrix of the process {$\mathbf{y}(k)$} up to time k-(λ+λ).

If the general form of the matrices $\mathbf{A}_\lambda$ and $\mathbf{B}_\lambda$ is respectively:

$$\begin{bmatrix} a_{11}^{\lambda} & \dots & a_{1m}^{\lambda} \\ \vdots & \ddots & \vdots \\ a_{m1}^{\lambda} & \cdots & a_{mm}^{\lambda} \end{bmatrix}, \begin{bmatrix} b_{11}^{\lambda} & \dots & b_{1m}^{\lambda} \\ \vdots & \ddots & \vdots \\ b_{m1}^{\lambda} & \cdots & b_{mm}^{\lambda} \end{bmatrix} \tag{4}$$

then

$$\begin{aligned} \mathbf{x}(k) \triangleq [&\alpha_{11}^{l}\, \alpha_{21}^{l} \cdots \alpha_{m1}^{l} \vdots \alpha_{12}^{l}\, \alpha_{22}^{l} \cdots \alpha_{m2}^{l} \vdots \cdots \alpha_{mm}^{l} \vdots \cdots \alpha_{mm}^{\lambda} \vdots \\ &b_{11}^{l}\, b_{21}^{l} \cdots b_{m1}^{l} \vdots b_{12}^{l}\, b_{22}^{l} \cdots b_{m2}^{l} \vdots \cdots b_{mm}^{l} \vdots \cdots b_{mm}^{\lambda}]^T \end{aligned} \tag{5}$$

$$\begin{aligned} \mathbf{H}(k) \triangleq [&y_1(k\text{-}1)\mathbf{I} \cdots y_m(k\text{-}1)\mathbf{I} \vdots \cdots \vdots y_1(k-\lambda)\mathbf{I} \cdots y_m(k\text{-}\lambda)\mathbf{I} \vdots \\ &v_1(k\text{-}1)\mathbf{I} \cdots v_m(k\text{-}1)\mathbf{I} \vdots \cdots \vdots v_1(k-\lambda)\mathbf{I} \cdots v_m(k\text{-}\lambda)\mathbf{I}\ ] \end{aligned} \tag{6}$$

where $\mathbf{I}$ is the $m \times m$ identity matrix.

In this study a comparison will be made amongst three classical criteria, AIC, AICC, BIC and the multimodel partitioning filter (MMPF).

The first three criteria attempt to identify a model that best explains the data with a minimum of free parameters based on a trade off between the fit of the model (which lowers the sum of squared residuals) and the model's complexity. In order to achieve that, they reward the goodness of fit and also include a penalty factor which is an increasing function of the number of estimated parameters. Their operation is based on the minimization of an objective function of the form:

$$\textit{Akaike Information Criterion (AIC)}\ \log\left(\left|\widehat{\mathbf{R}}_\theta\right|\right) + \frac{2(p+q)}{n} \tag{7}$$

$$\textit{Akaike's Corrected Information Criterion (AIC)}\ \ AICC = \log\left|\widehat{\mathbf{R}}_\theta\right| + \frac{2(p+q+1)n}{n-p-q-2} \tag{8}$$

$$\textit{Bayesian Information Criterion (BIC)}\ \ n\log\left(\left|\widehat{\mathbf{R}}_\theta\right|\right) + (p+q)\log(n)a \tag{9}$$

If the system model is not completely known the MMPF is one of the most widely used approaches for similar problems [19]-[26]. In our problem assume that the model uncertainty is the lack of knowledge of the model order θ. Let us further assume that the model order θ lies within a known sample space of finite cardinality, i.e. that $1 \le \theta \le M$, $\theta \in \mathfrak{I}$, where $\mathfrak{I}$ denotes the set of integers. The MMPF operates on the following discrete-time model:

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

$$\mathbf{y}(k) = \mathbf{H}(k/\theta)\,\mathbf{x}(k) + \mathbf{v}(k) \tag{11}$$

where θ is the unknown parameter - the model order in this case- $\mathbf{F}$ is the state transition matrix and $\mathbf{w}(k)$ is independent, zero mean, white noise not necessarily Gaussian with covariance $\mathbf{Q}$ which is usually set to a small positive non zero constant. The optimal *MMSE* (Minimum Mean Square Error) estimate of $\mathbf{x}(k)$ is given by:

$$\hat{\mathbf{x}}(k/k) = \sum_{j=1}^{M} \hat{\mathbf{x}}(k/k;\theta_j)\, p(\theta_j/k) \tag{12}$$

A set of models M is designed, each matching one value of the parameter vector, {(*1*,*1*), (*2*,*2*),…(*M*,*M*)}. The probabilities $p(\theta_j/k)$ for each model are set to *1*/*M*, where *M* is the cardinality of the model set.

A bank of Extended Kalman filters is then applied, one for each model, which can be run in parallel thus saving enormous computational time. At each iteration, the MMPF selects the model that corresponds to the maximum a posteriori probability as the correct one. This probability tends (asymptotically) to one, while the remaining probabilities tend to zero. The overall optimal estimate can be taken either to be the individual estimate of the elemental filter exhibiting the maximum posterior probability (MAP) [25], or the weighted average of the estimates produced by the elemental filters, as described in (12), which is the case used in this paper.

The probabilities are calculated on–line in a recursive manner as it is shown by (13) and (14).

$$p(\theta_j/k) = \frac{L(k/k;\theta_j)}{\sum_{j=1}^{M} L(k/k;\theta_j)\, p(\theta_j/k\text{-}1)}\, p(\theta_j/k\text{-}1) \tag{13}$$

$$L(k/k;\theta_j) = \left|\mathbf{P}_{\tilde{\mathbf{y}}}(k/k-1;\theta_j)\right|^{-1/2} \cdot \exp[-\frac{1}{2}\tilde{\mathbf{y}}^T(k/k-1;\theta_j)\mathbf{P}_{\tilde{\mathbf{y}}}^{-1}(k/k-1;\theta_j)\tilde{\mathbf{y}}(k/k-1;\theta_j)] \quad (14)$$

where the innovation process:

$$\tilde{\mathbf{y}}(k/k-1;\theta_j) = \mathbf{y}(k) - \mathbf{H}(k;\theta_j)\hat{\mathbf{x}}(k/k-1;\theta_j) \quad (15)$$

is a zero mean white process with covariance matrix

$$\mathbf{P}_{\tilde{y}}(k/k-1;\theta_j) = \mathbf{H}(k;\theta_j)\mathbf{P}(k/k;\theta_j)\mathbf{H}^{\mathbf{T}}(k;\theta_j) + \mathbf{R} \quad (16)$$

For equations (13)–(16) $j = 1, 2, \ldots, M$.

The formulation of the MMPF in this paper is different from the one presented in [21] and [27] because the algorithm is able in one pass to estimate the order and the coefficients of the AR and the MA component separately. This is done in the following way. The whole reformulation, equations (4) – (6) is based on the assumption that for model order $\theta=(p, q)$, a new variable is assigned such that $\lambda=\max(p, q)$. This implies that the order of both the AR and the MA component is equal. However in the last step of the proposed algorithm the coefficients of these components are calculated and if p>q then $\lambda=p$, the last $(m^2(p-q))$ MA coefficients are zero, if $q>p$ then $\lambda=q$, the last $(m^2(q-p))$ AR coefficients are zero.

3 RESULTS

All four criteria will be applied on real data representing the monthly thunderstorm days, covering the period from 01/1980 to 12/2005, fig. 1. The data has been collected from the National Meteorological Authority of Hellas. The thunderstorm, days are strongly related to the lightning and in a way determine the lightning level in an area, i.e. the number of lightning flashes to Earth. The prediction of the thunderstorm days is essential to transmission and distribution line's designers since the knowledge of the future lightning level of an area can result in a better design and consequently to the reduction of the lightning faults in lines. The data was normalized to take values from zero to one before using them as input to the four criteria. From the 312 available data points, 160 were used for the ARMA model estimation. The produced model was used for the complete series prediction. The results are presented below.

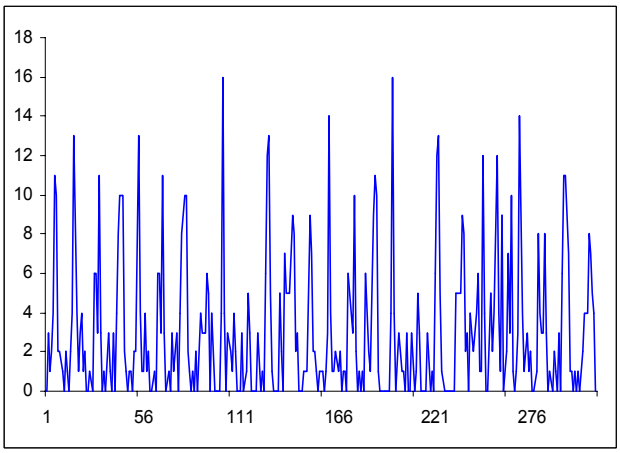

Figure 1. Monthly thunderstorm days (01/1980 – 12/2005)

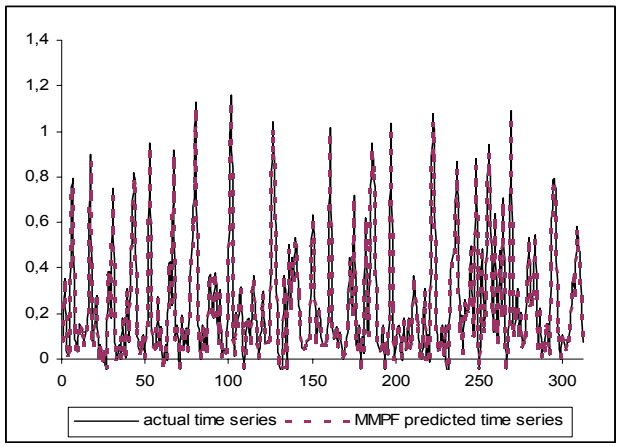

Figure 2. Actual Series and MMPF prediction

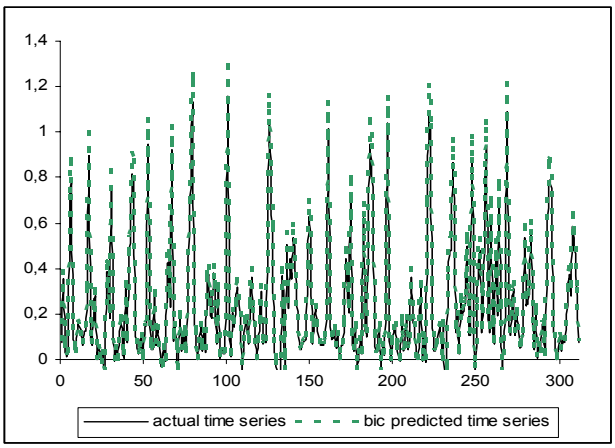

Figure 3. Actual Series and BIC prediction

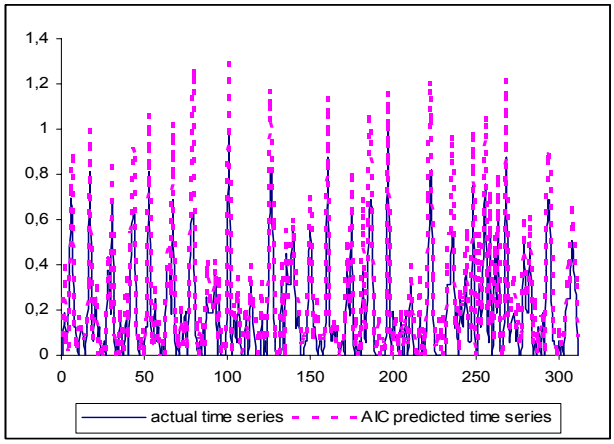

Figure 4. Actual Series and AIC prediction

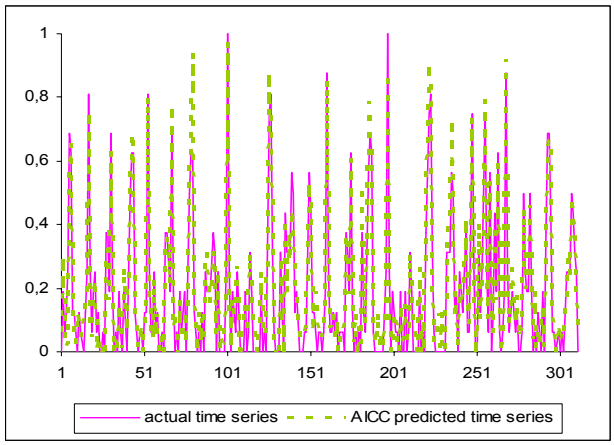

Figure 5. Actual Series and AICC prediction

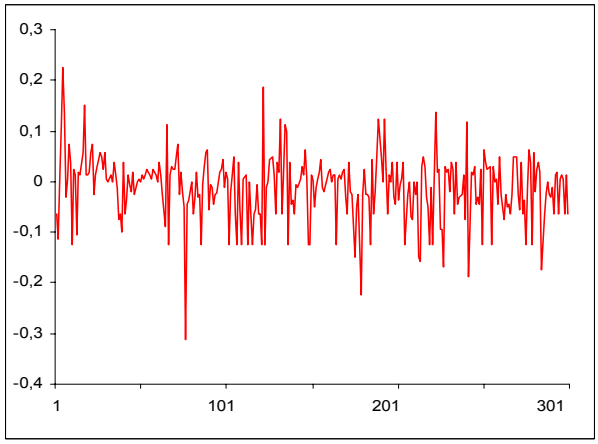

Figure 6. MMPF Actual Prediction Error

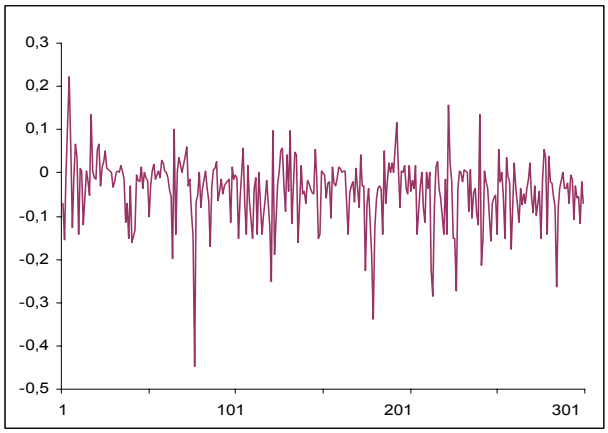

Figure 7. BIC Actual PredictionError

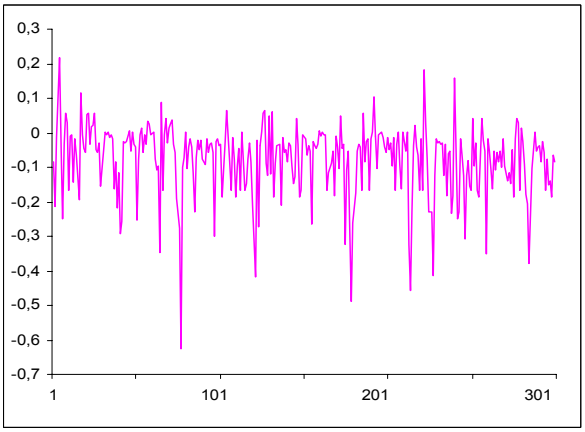

Figure 8. AIC Actual Prediction Error

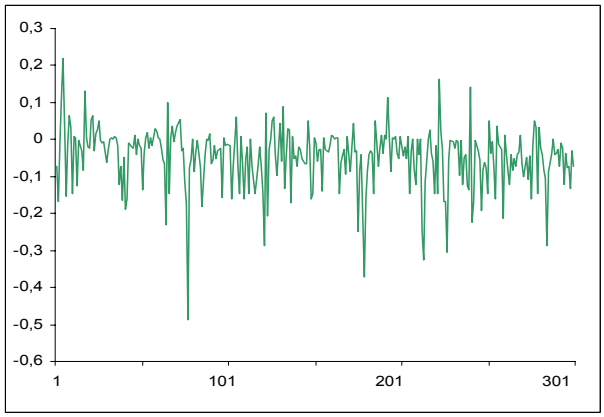

Figure 8. AICC Actual Prediction Error

4 CONCLUSIONS

Figures 2-5 show that the performance of all the criteria used can be considered to be satisfactory. However a more thorough observation indicates that the MMPF performance is slightly better than the rest. Figures 6-9 represent the actual error (|Real Value – Estimated Value|) for each point of the thunderstorm days series for each method. Obviously the error associated with MMPF is the smallest (Fig. 6). It is followed by BIC (Fig. 7), AICC (Fig. 9) and AIC (Fig. 8).

5 REFERENCES

[1] J. A. Velázquez-Muriel, C. O. S. Sorzano, J. J. Fernández and J. M. Carazo, “A method for estimating the CTF in electron microscopy based on ARMA models and parameter adjustment,” Ultramicroscopy, vol. 96, Issue 1, pp. 17-35, 2003.

[2] S. W. Dahlgren,N. H. Younan, B. J. Skinner, “Radar signal detection via ARMA modeling,” in Proc. Twenty-Seventh Southeastern Symposium on, pp. 436-440, 1995.

[3] Sheng Lu, Ki Hwan Ju, Chon KH, “A new algorithm for linear and nonlinear ARMA model parameter estimation using affine geometry [and application to blood flow/pressure data],” IEEE Transactions on Biomedical Engineering, vol. 48, Issue 10, pp. 1116-1124, 2001.

[4] Derk J, Weber S, Weber C, “Extended ARMA models for estimating price developments on day-ahead electricity markets,” Electric Power Systems Research, vol. 77, pp. 583-593, 2007.

[5] Kourosh M, Eslami HR, Kahawita R, "Parameter estimation of an ARMA model for river flow forecasting using goal programming,". Journal of Hydrology, vol. 331, pp. 293-299, 2006.
[6] H. Akaike, "Fitting autoregressive models for prediction," Ann. Inst. Stat. Math., vol. 21, pp. 243–247, 1969.
[7] C. C. Chen, R. A. Davis, P. J. Brockwell, " Order determination for multivariate autoregressive processes using reasampling methods," Journal of Multivariate Analysis, vol. 57, pp. 175-190, 1996.
[8] G. Schwarz, "Estimation of the dimension of the model," Ann. Statist., vol. 6, pp. 461-464, 1978.
[9] Rissanen, "A predictive least squares principle," IMA J. Math. Contr. Inform., vol. 3, pp. 211-222, 1986.
[10] M. Wax, "Order selection for AR models by predictive least squares," IEEE Transactions on Acoustics, Speech and Signal Processing, vol. 36, pp. 581-588, 1988.
[11] James. D. Hamilton, "State Space Models,", Handbook of econometrics, vol. IV, pp. 3041-3068, 1994.
[12] D. R. Anderson, K. P. Burnham and G. C. White , 'Comparison of Akaike information criterion and consistent Akaike's information criterion for model selection and statistical inference from capture-recapture studies', Journal of Applied Statistics, vol. 25, pp. 263 – 282, 1998.
[13] H. Lutkepohl, "Comparison of criteria for estimating the order of a vector AR process," J. Time Ser. Anal., vol. 6, pp. 35-52, 1985.
[14] Thomas M. Ludden, Stuart L. Beal, and Lewis B. Sheiner, "Comparison of the Akaike Information Criterion, the Schwarz Criterion and the F Test as Guides to Model Selection," Journal o f Pharmacokinetics and Biopharmaceutics, vol. 22, No. 5, pp. 431-445, 1994.
[15] S.Sp. Pappas, L. Ekonomou, V.C. Moussas, P. Karampelas, S.K. Katsikas , "Adaptive load forecasting of the Hellenic electric grid,", Journal of Zhejiang University SCIENCE A, Accepted, to be published.
[16] Moussas V.C.,Pappas S.S., "Adaptive Network Anomaly Detection using Bandwidth Utilisation Data," 1st International Conference on Experiments/Process/System Modeling/ Simulation/ Optimization, (1st IC-EpsMsO), Athens, 6-9 July, 2005.
[17] D. G. Lainiotis, "Partitioning: A unifying framework for adaptive systems I: Estimation", Proc. IEEE, vol. 64, pp. 1126-1143, 1976.
[18] D. G. Lainiotis, "Partitioning: A unifying framework for adaptive systems II: Control", Proc. IEEE, vol. 64, pp. 1182-1198, 1976.
[19] V. C. Moussas, S. D. Likothanassis, S. K. Katsikas, A. K. Leros, "Adaptive on-line multiple source detection," IEEE International Conference on Acoustics, Speech, and Signal Processing, (ICASSP'05), vol. 4, pp. 1029-1032, 2005.
[20] N. V. Nikitakos, A. K. Leros, S. K. Katsikas, "Towed array shape estimation using multimodel partitioning filters," IEEE Journal of Oceanic Engineering, vol. 23, no. 4, pp. 380-384, 1998.
[21] S. Sp. Pappas,A. K. Leros, S. K. Katsikas, "Joint order and parameter estimation of multivariate autoregressive models using multi-model partitioning theory," Digital Signal Processing, vol. 16, Issue 6, pp.782-795, 2006V. C. Moussas, S. Sp. Pappas, "Adaptive network anomaly detection using bandwidth utilization data," 1st Int. Conf. on Experiments/Processes/System Modelling/Simulation/Optimization, Athens, 2005.
[22] S. Sp. Pappas, L. Ekonomou, P. Karampelas, S. K. Katsikas, P. Liatsis, "Modeling of the Grounding Resistance Variation using ARMA models," Simulation Modelling, Practice and Theory, vol. 16, Issue 5, pp. 560-570, May 2008.
[23] S. K. Katsikas, A. K. Leros, D. G. Lainiotis, " Passive tracking of a maneuvering target: an adaptive approach," IEEE Transactions on Signal Processing, vol. 42, Issue 7, pp. 1820-1825, 1995.
[24] V. C. Moussas, S. K. Katsikas, D. G. Lainiotis, "Adaptive Estimation of FCG Using Nonlinear State-Space Models," Stochastic Analysis and Applications, vol. 23, Issue 4, pp. 705-722, Jul. 2005.
[25] D. G. Lainiotis, S. K. Katsikas, S. D. Likothanassis, "Adaptive deconvolution of seismic signals: performance, computational analysis, parallelism," IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 36, no. 11, pp. 1715-1734, 1988.
[26] G. Solomos and V. C. Moussas, "A Time Series Approach to Fatigue Crack Propagation," Structural Safety, vol. 9, pp. 211-226, 1991.
[27] S. Sp. Pappas,A. K. Leros, S. K. Katsikas, "Multivariate AR Model Order Estimation with Unknown Process Order," Lecture Notes in Computer Science, ISSN: 0302-9743, vol. 3746/2005, pp. 787–797.