World Academy of Science, Engineering and Technology
International Journal of Mathematical and Computational Sciences
Vol:3, No:1, 2009

Time Series Forecasting Using Independent Component Analysis

Theodor D. Popescu, *Senior Member, IEEE*

***Abstract*—The paper presents a method for multivariate time series forecasting using Independent Component Analysis (ICA), as a preprocessing tool. The idea of this approach is to do the forecasting in the space of independent components (sources), and then to transform back the results to the original time series space. The forecasting can be done separately and with a different method for each component, depending on its time structure. The paper gives also a review of the main algorithms for independent component analysis in the case of instantaneous mixture models, using second and high-order statistics. The method has been applied in simulation to an artificial multivariate time series with five components, generated from three sources and a mixing matrix, randomly generated.**

***Index Terms*—Independent Component Analysis, second order statistics, simulation, time series forecasting.**

Open Science Index, Mathematical and Computational Sciences Vol:3, No:1, 2009 publications.waset.org/10199/pdf

I. Independent Component Analysis

A. Problem Formulation

INDEPENDENT Component Analysis (ICA) is a statistical and computational technique, that can be seen as an extension to Principal Component Analysis (PCA) and Factor Analysis (FA), [1]. ICA is a much more powerful technique, capable of nding the underlying factors or sources when these classic methods fail completely. The data analyzed by ICA could originate from many different kinds of application elds, including digital images, economic indicators and psychometric measurements.

The simple ICA model assumes the existence of n independent components $s_1(t), \ldots, s_n(t)$ and the observation of as many mixtures $x_1(t), \ldots, x_n(t)$, these mixtures being linear and instantaneous, i.e.

$$x_i(t) = \sum_{j=1}^{n} a_{ij} s_j(t) \tag{1}$$

for each $i = 1, n$. This is compactly represented by the mixing equation

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) \tag{2}$$

where $\mathbf{s}(t) = [s_1(t), \ldots, s_n(t)]^T$ is an $n \times 1$ column vector collecting the source components, vector $\mathbf{x}(t)$ collects the n observed variables and the square $n \times n$ "mixing matrix" $\mathbf{A}$ contains the mixture coef cients. The ICA problem consists in recovering the source vector $\mathbf{s}(t)$ using only the observed data $\mathbf{x}(t)$, the assumption of independence between the entries of

Th. D. Popescu is with National Institute for Research and Development in Informatics, 8-10 Averescu Avenue, 011455 Bucharest, Romania (e-mail:pope@ici.ro)

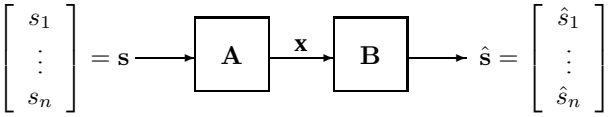


Fig. 1. Mixing and separating. Unobserved variables: $\mathbf{s}$; observations: $\mathbf{x}$; estimated source components: $\hat{\mathbf{s}}$

the input vector $\mathbf{s}(t)$ and possible some a priori information about the probability distribution of the inputs. It can be formulated as the computation of an $n \times n$ "separating matrix" $\mathbf{B}$ whose output $\hat{\mathbf{s}}(t)$

$$\hat{\mathbf{s}}(t) = \mathbf{B}\mathbf{x}(t) \tag{3}$$

is an estimate of the vector $\mathbf{s}(t)$ of the source signals (see Fig. 1).

ICA is closely related to the method Blind Source Separation (BSS) problem. A "source" means here an original component, i.e. independent component. "Blind" means that we no very little, if anything, on the mixing matrix, and make little assumptions on the source components. ICA is one method, perhaps the most widely used, for performing blind source separation.

In many applications, it would be more realistic to assume that there is some noise in the measurement data, which would mean adding a noise term in the model:

$$\begin{aligned} \mathbf{y}(t) &= \mathbf{A}\mathbf{s}(t) \\ \mathbf{x}(t) &= \mathbf{y}(t) + \mathbf{n}(t) \end{aligned} \tag{4}$$

B. Identificability of the ICA model

The identi cability of the noise-free ICA model has been treated in Comon, [2]. By imposing the following fundamental restrictions (in addition to the basic assumption of statistical independence), the identi ab ility of the model can be assured:

1) All the independent components s_i with the possible exception of one component, must be non-Gaussian.
2) The number of the observed linear mixtures m must be at least as large as the number of independent components n, i.e. $m \geq n$.
3) The matrix $\mathbf{A}$ must be of full column rank.

Usually, it is also assumed that $\mathbf{x}$ and $\mathbf{s}$ are centered. If $\mathbf{x}$ and $\mathbf{s}$ are interpreted as stochastic processes instead of simply random variables, additional restrictions are necessary. At the minimum, one has to assume that the stochastic processes are

Open Science Index, Mathematical and Computational Sciences Vol:3, No:1, 2009 publications.waset.org/10199/pdf

stationary in the strict sense. Some restriction of ergodicity with respect to the quantities estimated are also necessary.

In the ICA model of eq. (2), it is easy to see that the following ambiguities will hold:

1) We cannot determine the variances (energies) of the independent components. The reason is that, both $\mathbf{s}$ and $\mathbf{A}$ being unknown, any scalar multiplier in one of the sources s_i could always be canceled by dividing the corresponding column $\mathbf{a}_i$ in $\mathbf{A}$ by the same scalar. As a consequence we may quite as well x the magnitudes of the independent components; as they are random variables, the most natural way to do this is to assume that each has unit variance: $E[s_i^2] = 1$. Then the matrix $\mathbf{A}$ will be adapted in the ICA solution methods to take into account this restriction.
2) We cannot determine the order of the independent components. The reason is that, again both $\mathbf{s}$ and $\mathbf{A}$ being unknown, we can freely change the order of the terms in the sum (1), and call any of the independent components the rst one.

C. Algorithms for ICA

Though many papers purport to introduce "new" methods of solution, the existing framework (and solutions) for blind source separation are often the same. Speci cally, the criterion for source separation is a measure of independence, typically represented by some cost function $\mathbf{J}$. The extremum of $\mathbf{J}$, with respect to parameters of some inverse mixing processes, will correspond to more or less independent outputs. Algorithms which rely on this concept, the separation-independence equivalence, may be closed as those performing Independent Component Analysis. The problem of blind source separation is then reduced to a mathematical optimization problem, and a multitude of techniques are reported. The principal differences rest on the varieties of cost functions utilized, based on the kurtosis, mutual information, cross power-spectra, negentropy and log-likelihood. In many cases these approaches are the result of different formalisms, and can be shown to be mathematically equivalent.

When the signals are temporal coherent, it is possible to solve ICA problem using only the second-order statistics. If the signals are temporal white or have identical normalized spectral densities, without any information on a priori source distributions, the solution will need higher-order statistics. If the source signal distributions are known, the problem could be solved by maximum likelihood method.

In the next sections we present two algorithms for ICA in the case of an instantaneous mixture model (1): the rst, SOBI (Second Order Blind Identi cation) algorithm, supposes the signals temporal coherent and exploits the second-order statistics using intercovariance matrix of observations, and the second, JADE (Joint Approximate Diagonalization of Eigen-matrices), supposes the components white temporal and exploit high-order statistics, using non-linear functions.

II. SOBI Algorithm

A. Second-Order Information

The rst step of the procedure, [3] consists of prewhitening the signal part $\mathbf{y}(t)$ of the observation. This is done via a whitening matrix $\mathbf{W}$, i.e. a $n \times m$ matrix (we consider n sources and m mixtures) such that $\mathbf{W}\mathbf{y}(t)$ is spatially white. The whiteness condition is

$$\mathbf{I}_n = \mathbf{W}\mathbf{R}_y(0)\mathbf{W}^T = \mathbf{W}\mathbf{A}\mathbf{A}^T\mathbf{W}^T \tag{5}$$

where $\mathbf{I}_n$ denotes the $n \times n$ identity matrix. Equations (5) implies that $\mathbf{WA}$ is a unitary matrix: for any whitening matrix $\mathbf{W}$, it then exists a unitary matrix $\mathbf{U}$ such that $\mathbf{WA} = \mathbf{U}$. As a consequence, matrix $\mathbf{A}$ can be factored as

$$\mathbf{A} = \mathbf{W}^{\#}\mathbf{U} = \mathbf{W}^{\#}[\mathbf{u}_1, \ldots, \mathbf{u}_n] \tag{6}$$

where # denotes the pseudo-inverse and $\mathbf{U}$ is unitary. The use of second-order information - in the form of an estimate of $\mathbf{R}_y(0)$ which is used to solve for $\mathbf{W}$ in (5) - reduces the determination of the $m \times n$ mixing matrix $\mathbf{A}$ to the determination of a unitary $n \times n$ matrix $\mathbf{U}$. The whitened process $\mathbf{x}_w(t) = \mathbf{W}\mathbf{x}(t)$ still obeys a linear model:

$$\mathbf{x}_w(t) \stackrel{def}{=} \mathbf{W}\mathbf{x}(t) = \mathbf{W}(\mathbf{A}\mathbf{s}(t) + \mathbf{n}(t)) = \mathbf{U}\mathbf{s}(t) + \mathbf{W}\mathbf{n}(t) \tag{7}$$

The signal part of the whitened process now is a unitary mixture of the source signals. Note that all the information contained in the covariance is 'exhausted' after the whitening, in the sense that changing $\mathbf{U}$ in (7) to any other unitary matrix leaves unchanged the covariance of $\mathbf{x}_w(t)$.

B. Whitening Matrix Computation

This step is implemented via eigendecomposition of the sample covariance matrix $\hat{\mathbf{R}}_x(0)$. We consider here that the noise covariance is of the form $\mathbf{R}_n(0) = \sigma^2\mathbf{I}_n$. The whitening procedure is the following:

1) Estimate the covariance matrix $\hat{\mathbf{R}}_x(0)$ using T samples of the observations:

$$\hat{\mathbf{R}}_x(0) = \frac{1}{T}\sum_{t=1}^{T}\mathbf{x}(t)\mathbf{x}(t)^T, \tag{8}$$

2) Perform the eigendecomposition of the $\hat{\mathbf{R}}_x(0)$ covariance matrix

$$\hat{\mathbf{R}}_x(0) = \mathbf{H}\Delta\mathbf{H}^T \tag{9}$$

where

$$\mathbf{H} = [\mathbf{h}_1, ..., \mathbf{h}_m]$$

and

$$\Delta = diag[\lambda_1, \ldots, \lambda_m]$$

with $\lambda_i \geq \lambda_j$ for $i < j$. The number of sources can be estimated starting from the spectrum Δ [4], [5].

3) Estimate noise variance $\hat{\sigma}^2$ as the average of the $m-n$ smallest eigenvalues of Δ

$$\hat{\sigma}^2 = \frac{1}{m-n}\sum_{i=n+1}^{m}\lambda_i \tag{10}$$

4) Compute the whitening matrix $\hat{\mathbf{W}}$ as:

$$\hat{\mathbf{W}} = \Delta^{'}\mathbf{H}^{'T} \tag{11}$$

where

$$\Delta^{'} = diag[(\lambda_1 - \hat{\sigma}^2)^{-1/2}, \ldots, (\lambda_n - \hat{\sigma}^2)^{-1/2}]$$

and

$$\mathbf{H}^{'} = [\mathbf{h}_1, \ldots, \mathbf{h}_n]$$

This resulted matrix is used to obtain the whitened process

$$\hat{\mathbf{x}}_w(t) = \hat{\mathbf{W}}\mathbf{x}(t), \quad t = 1, \ldots, T \tag{12}$$

C. Intercovariance Matrix Estimation

Starting from the whitened process $\mathbf{x}_w(t)$, K intercovariance matrices of this process are computed:

$$\hat{\mathbf{R}}_w(k) = \frac{1}{T-k}\sum_{t=k+1}^{T}\mathbf{x}_w(t)\mathbf{x}_w(t-k)^T \tag{13}$$

where $1 \geq k \geq K$. The resulted matrices are of $n \times n$ dimension, and the computation effort does not depend of number of sensors, m. The value of K will be selected to realize a trade off between the statistic ef cien cy and computation effort. The value of the delays used in computation depends also on the length of the signal correlations. If we have a priori information on spectral density of sources, the value of K can be optimal chosen.

D. Joint Diagonalization

Let $\mathbf{R}_w = \{\mathbf{R}_w(k) | 1 \leq k \leq K\}$ be a set of K matrices with common size $n \times n$. A joint diagonalizer of the set $\mathbf{R}_w$ is de n ed as a unitary maximizer of the criterion

$$C(\mathbf{U}) \stackrel{def}{=} \sum_{k=1}^{K}|diag(\mathbf{U}^T\mathbf{R}_w(k)\mathbf{U})|^2 \tag{14}$$

where $|diag(\cdot)|$ is the norm of the vector build from the diagonal of the matrix argument. The problem is solved by a generalization of Jacobi technique [6], [7], [8].

E. Mixing Matrix and Source Signals Estimation

Let $\hat{\mathbf{U}} = [\hat{\mathbf{u}}_1, \ldots, \hat{\mathbf{u}}_n]$ be the unitary matrix resulted by joint diagonalization. If the objective of the blind identi cation is source separation, a brute estimation of these can be computed by:

$$\hat{\mathbf{s}}(t) = \hat{\mathbf{U}}^T\hat{\mathbf{x}}_w(t) \tag{15}$$

To estimate the mixing matrix need to inverse the effect of whitening, and the mixing matrix can be estimated by

$$\hat{\mathbf{A}} = \hat{\mathbf{W}}^{\#}\hat{\mathbf{U}} \tag{16}$$

To obtain at the output of the separator a maximum signal/noise ratio the source signals are estimated by

$$\hat{\mathbf{s}}(t) = \hat{\mathbf{A}}^T\hat{\mathbf{R}}_x(0)^{-1}\mathbf{x}(t) \tag{17}$$

F. The Algorithm

The general scheme of the SOBI algorithm (Second Order Blind Identi catio n) can now be described by the following steps:

Step 1. Form the sample covariance $\hat{\mathbf{R}}_x(0)$ and compute the whitening matrix $\hat{\mathbf{W}}$
Step 2. Whitening the data provided by the sensors:

$$\hat{\mathbf{x}}_w(t) = \hat{\mathbf{W}}\mathbf{x}(t), \quad t = 1, \ldots, T$$

Step 3. Estimate K intercovariance matrices $\hat{\mathbf{R}}_w(k)$ of $\hat{\mathbf{x}}_w(t)$ for different time delay $k = 1, \ldots, K$
Step 4. Jointly diagonalize the set of intercovariace matrices in a base $\hat{\mathbf{U}} = [\hat{\mathbf{u}}_1, \ldots, \hat{\mathbf{u}}_n]$
Step 5. Estimate the mixing matrix with

$$\hat{\mathbf{A}} = \hat{\mathbf{W}}^{\#}\hat{\mathbf{U}}$$

Step 6. Estimate the source signals by

$$\hat{\mathbf{s}}(t) = \hat{\mathbf{A}}^T\hat{\mathbf{R}}_x(0)^{-1}\mathbf{x}(t)$$

Note that at the second step of the algorithm the observation dimension is reduced to n, the source number, and the intercovariance matrices estimation is performed in a space of reduced dimension.

III. JADE Algorithm

A. Notations and Assumptions

Let be the mixing model given by eq. (4) with n sources and m observed variables and $\mathbf{A}$ a matrix $m \times n$. We discuss this approach for complex data. It exploits the fourth-order cumulants of the output. For $\mathbf{v}$ a complex d- dimensional random vector with coordinates $v_1, \ldots, v_d$ şi and nite 4th-order cumulants, we de ne a cumulant set denoted $\mathcal{Q}_v$ as:

$$\mathcal{Q}_v \stackrel{def}{=} \{Cum(v_i, v_j^*, v_k, v_l^* | 1 \leq i, j, k, l \leq d\} \tag{18}$$

For a complex stationary process $\mathbf{v}(t)$ we also denote $\mathcal{Q}_v$ rather than $\mathcal{Q}_{\mathbf{v}(t)}$, since the latter does not depend on t. We assume

1) The processes $\mathbf{n}(t), s_1(t), \ldots, s_n(t)$ are jointly stationary. The kurtosis of the p-th source is the real number

$$k_p \stackrel{def}{=} Cum\{s_p(t), s_p^*(t), s_p(t), s_p^*(t)\} \tag{19}$$

A source is to be said *kurtic* if it has a non zero kurtosis. We restrict ourselves to the case where:

Open Science Index, Mathematical and Computational Sciences Vol:3, No:1, 2009 publications.waset.org/10199/pdf

Open Science Index, Mathematical and Computational Sciences Vol:3, No:1, 2009 publications.waset.org/10199/pdf

2) There is at most one non kurtic source. The crucial assumptions blind identi catio n relies on the related to independence, exploited in the paper by assuming non Gaussian signals. More speci c ally, we assume:
3) All vectors $\mathbf{a}_1, \ldots, \mathbf{a}_n$ of matrix $\mathbf{A}$ are linearly independent but otherwise arbitrary.
4) The variables $s_1(t), \ldots, s_n(t)$ are statistically independent for each t. Under Assumptions 1-3, the matrix $\mathbf{A}$ is essentially determined from $\mathbf{R}_y \stackrel{def}{=} E\{\mathbf{y}(t)\mathbf{y}^T(t)\}$ and $\mathcal{Q}_y$. For these quantities to be consistently estimated, it is further assumed that:
5) There exist consistent estimates for $\mathbf{R}_x$ and şi $\mathcal{Q}_x$.
6) The additive noise is normally distributed and independent from the sources.
7) The additive noise is spatially white $\mathbf{R}_n = \sigma\mathbf{I}_m$ with unknown variance σ and $n < m$.

By Assumption 5, an estimate of $\mathcal{Q}_x$ also is an estimate of $\mathcal{Q}_y$ since cumulants are additive for independent variables and since high-order cumulants are zero for normaly distributed variables. By Assumption 6, an estimate of $\mathbf{R}_y$ can be clasically constructed from the eigendecomposition of an estimate of $\mathbf{R}_x$. The Assumptions 4-6 could be replaced by any other assumption set serving the same purpose: the existence of consistent estimates for $\mathbf{R}_y$ şi $\mathcal{Q}_y$.

B. Using Second-Order Information

We consider exploiting second order information by whitening the signal part $\mathbf{y}(t)$ of the observation. This is done via a *whitening matrix* $\mathbf{W}(n \times m)$ such that $\mathbf{W}\mathbf{y}(t)$ is spatially white. The whiteness condition is:

$$\mathbf{I}_n = \mathbf{W}\mathbf{R}_y\mathbf{W}^H = \mathbf{W}\mathbf{A}\mathbf{A}^H\mathbf{W}^H \tag{20}$$

where $\mathbf{I}_n$ denotes $n \times n$ identity matrix. Equation (20) implies that $\mathbf{WA}$ is a unitary matrix: for any whitening matrix $\mathbf{W}$, it then exists a unitary matrix $\mathbf{U}$ such that $\mathbf{WA} = \mathbf{U}$. As a consequence, matrix $\mathbf{A}$ can be factored as

$$\mathbf{A} = \mathbf{W}^{\#}\mathbf{U} = \mathbf{W}^{\#}[\mathbf{u}_1, \ldots, \mathbf{u}_n] \tag{21}$$

where $\mathbf{U}$ is unitary, and # denota pseudo-inverse. The use of second-order information - in the form of an estimate of $\mathbf{R}_y$ which is used to solve for $\mathbf{W}$ in (20) - reduces the determination of the $m \times n$ mixing matrix $\mathbf{A}$ to the determination of a unitary , $n \times n$ matrix $\mathbf{U}$. The whitened process $\mathbf{x}_w(t) = \mathbf{W}\mathbf{x}(t)$ still obeys a linear model:

$$\mathbf{x}_w(t) \stackrel{def}{=} \mathbf{W}\mathbf{x}(t) = \mathbf{W}(\mathbf{A}\mathbf{s}(t) + \mathbf{n}(t)) = \mathbf{U}\mathbf{s}(t) + \mathbf{W}\mathbf{n}(t) \tag{22}$$

The signal part of the whitened process now is a unitary mixture of the source signals. Note that all information contained in the covariance is 'exhausted' after the whitening, in the sense that changing $\mathbf{U}$ in (28) to any other unitary matrix leaves unchanged the covariance of $\mathbf{x}_w(t)$.

C. Determining the Unitary Factor

Two approaches for the determination of the unitary factor $\mathbf{U}$ in $\mathbf{A} = \mathbf{W}^{\#}\mathbf{U}$ have been reported. In the rst approach, $\mathbf{U}$ is computed as the diagonalizer of a $n \times n$ cumulant matrix. These 'eigenbased' techniques are computationally simple but, being based only on n^2 cumulant statistics, they may show poor statistical performance. Another approach obtains an estimate of $\mathbf{U}$ as the optimizer of some identi catio n criterion which is a function of the whole cumulant set $\mathcal{Q}_{\mathbf{x}_w}$; better performance is expected at the expense of solving an optimization problem. These approached are reviewed by Cardoso şi Souloumiac in [9].

D. The Algorithm

A blind identi catio n algorithms by Joint Approximate Diagonalization of Eigen-matrices (JADE) can be described by the following steps:

Step 1. Form the sample covariance $\hat{\mathbf{R}}_x$ and compute a whitening matrix $\hat{\mathbf{W}}$.
Step 2. Form the sample 4th-order cumulants $\hat{\mathcal{Q}}_{x_w}$, of the whitened process $\hat{\mathbf{x}}_w(t) = \hat{\mathbf{W}}\mathbf{x}(t)$; compute the n most signi cant eigenpairs $\{\hat{\lambda}_r, \hat{M}_r | 1 \le r \le n\}$.

The *cumulant matrices* are de ned as follows. To any $n \times n$ matrix $\mathbf{M}$, is associated a 'cumulant matrix', $\mathcal{Q}_{x_w}(\mathbf{M})$, de ned by

$$\mathbf{N} = \mathcal{Q}_{x_w}(\mathbf{M}) \stackrel{def}{\Longleftrightarrow} n_{ij} = \sum_{k,l=1,n} Cum(x_{w,i}, x^*_{w,j}, x_{w,k}, x^*_{w,l}) m_{lk} \tag{23}$$

for $1 \le i, j \le n$

For a any d - dimensional complex random vector $\mathbf{v}$ with 4th - order cumulants, there exist d^2 real number $\lambda_1, \ldots, \lambda_{d^2}$ and d^2 matrices $\mathbf{M}_1, \ldots, \mathbf{M}_{d^2}$, called eigenmatrices verifying, [9]:

$$\mathcal{Q}_v(\mathbf{M}_r) = \lambda_r\mathbf{M}_r, \quad Tr(\mathbf{M}_r\mathbf{M}_s^H) = \delta(r, s) \tag{24}$$

for $1 \le r, s \le d^2$
Step 3. Jointly diagonalize the set $\{\hat{\lambda}_r\hat{\mathbf{M}}_r | 1 \le r \le n\}$ by a unitary matrix $\mathbf{U}$.

For a set of s matrices with common size $n \times n$, $\mathcal{N} = \{\mathbf{N}_r | 1 \le r \le s\}$, a joint diagonalizer of the set is de ned as a unitary matrix $\mathbf{U}$ maximizer of the criterion

$$C(\mathbf{U}, \mathcal{N}) = \sum_{r=1,s} |diag(\mathbf{U}^H\mathbf{N}_r\mathbf{U}|^2 \tag{25}$$

where $|diag(.)|$ is the norm of the vector build from the diagonal of the matrix argument. When the set $\mathcal{N}$ contain only one hermitian matrix, joint diagonalization is equivalent to usual unitary diagonalization. If the set $\mathcal{N}$ cannot be exactly jointly diagonalized (this is the case when sample cumulants are processed), the unitary maximization de nes a somewhat arbitrary but quite natural 'joint approximate diagonalization'.
Step 4. An estimate of $\mathbf{A}$ is $\mathbf{A} = \mathbf{W}^{\#}\mathbf{U}$.

Open Science Index, Mathematical and Computational Sciences Vol:3, No:1, 2009 publications.waset.org/10199/pdf

Step 1 is concerned with 2nd-order statistics and is standard under the Assumption 5-6; it is implemented via eigendecomposition of $\hat{\mathbf{R}}_x$. Due to the white noise assumption, an estimate $\hat{\sigma}$ of the noise variance is the average of the $m-n$ smallest eigenvalues of $\hat{\mathbf{R}}_x$. Denote $\mu_1, \ldots, \mu_n$ the n largest eigenvalues and $\mathbf{h}_1, \ldots, \mathbf{h}_n$ the corresponding eigenvectors of $\hat{\mathbf{R}}_x$. A whitener is

$$\hat{\mathbf{W}} = [(\mu_1 - \hat{\sigma})^{-1/2}\mathbf{h}_1, \ldots, (\mu_n - \hat{\sigma})^{-1/2}\mathbf{h}_n]^H \qquad (26)$$

In **Step 2**, computation of the eigenmatrices amounts to diagonalizing a $n^2 \times n^2$ matrix made from the elements of $\mathcal{Q}_z$. A standard algorithm for eigendecomposition of o hermitian matrices can be used, but more ef cient implementations can be also be devised, by taking into account additional cumulants symmetries or the fact that only the n most signi c ant eigenpairs are needed, [**?**].

The **Step 3** is implemented by extending the single-matrix Jacobi technique to several matrices as described by Cardoso şi Souloumiac, [**?**]. Note that when $n = 2$, the Jacobi technique is not iterative: a unique Givens rotation achieves (joint) diagonalization. Also recall that joint diagonalization may be initialized with the (usual) diagonalizer of a single cumulant matrix.

In **Step 4**, the pseudo-inverse of $\hat{\mathbf{W}}$ needs not be explicitly computed: the eigendecomposition of $\hat{\mathbf{R}}_x$ may be recycled by $\hat{\mathbf{W}}^{\#} = [(\mu_1 - \hat{\sigma})^{1/2}\mathbf{h}_1, \ldots, (\mu_n - \hat{\sigma})^{-1/2}\mathbf{h}_n]$.

IV. Time Series Forecasting by ICA

The idea of multivariate time series forecasting by ICA is to estimate the independent sources from the observation data, to do the forecasting for these in the source space, and then to come back with the results in original observation space, using the estimation of the mixing matrix $\mathbf{A}$. The forecasting can be done separately, for each source, with a different method for each component, depending on its time structure. The following procedure can be used:

1) After substracting the mean of each time series, the independent components $s_j(t)$ and the mixing matrix $\mathbf{A}$ are estimated.
2) For each independent component $s_j(t)$ a suitable l-tering procedure is applied to reduce the effect of noise-smoothing for components that contain very low frequency (trend, slow cyclical variations), and high-pass ltering for components containing high frequencies and/or sudden shocks.
3) Each smoothed independent component is predicted separately, for instance using some method of autoregressive (AR) or autoregressive and moving average (ARMA) modeling. The prediction is done for a number of steps into the future.
4) The predictions for each independent component are combined by weighting them with the mixing coef- cien ts, a_{ij}, to obtain the predictions, $x_i^p(t)$ for the original time series components $x_i(t)$.

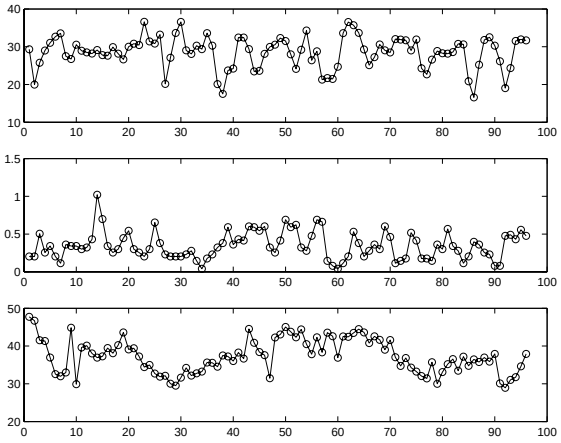


Fig. 2. Original sources

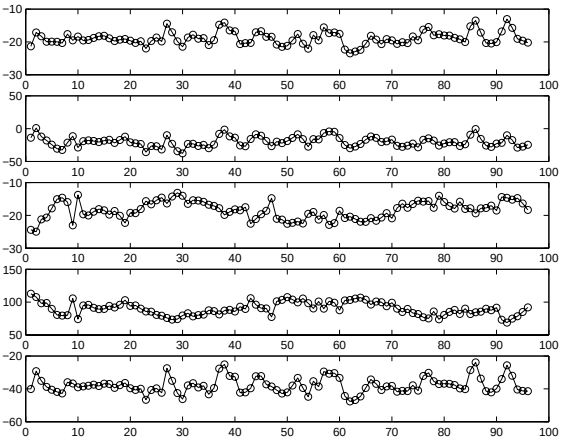


Fig. 3. Time series analyzed

V. Experimental Results

To test the method, procedure been applied to an arti cial multivariate time series with ve components, generated from three sources and a mixing matrix $\mathbf{A}$, randomly generated.

The original sources used in this application are given in Fig. 2. The mixing signals - the synthetic data - resulted are represented in Fig. 3.

The sources signals have been estimated from the mixing signals using JADE algorithm. The estimated sources are represented in Fig. 4. The estimated sources, have been assimilated to three monovariable time series and the following ARIMA models resulted for s1, s2 and s3, respectively, where ϵ_t is innovation and $Bx_t = x_t - x_{t-1}$:

$$(1 - B)s1_t = (1 - 0.402B - 0.239B^2)\epsilon_t \qquad (27)$$

with $\sigma_\epsilon^2 = 0.330$

$$s2_t = -9.066 + (1 + 0.467B - 0.188B^4)\epsilon_t \qquad (28)$$

with $\sigma_\epsilon^2 = 0.552$

$$(1 + 0.181B^3)s3_t = 1.076 + (1 + 0.472B)\epsilon_t \qquad (29)$$

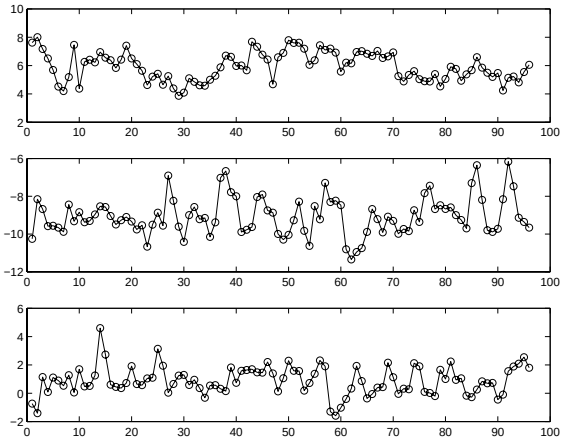


Fig. 4. Estimated sources

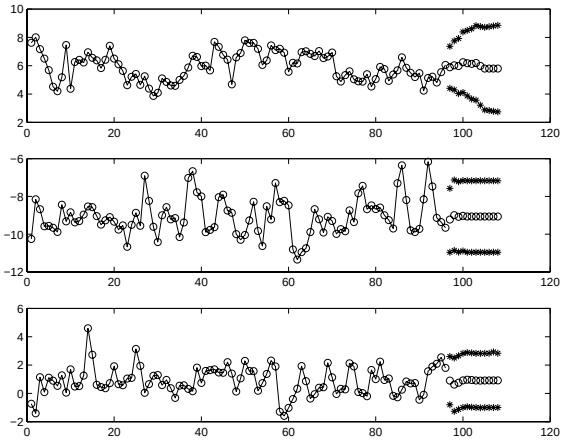


Fig. 5. Forecasting of the estimated sources

with $\sigma_\epsilon^2 = 0.571$

Starting from the resulted models we performed forecasting for the estimated sources, for a forecasting horizon of 12 steps. The con dence interval for forecasting has been chosen 95%. The resulted are presented in Fig. 5. Based on the forecasting results of the estimated sources and knowing the mixing matrix it was possible to determine the forecasting values of the multivariate time series investigated. The forecasting results with the con dence intervals are given in Fig. 6.

VI. Conclusions

The results are promising as the ICA based forecasting is expected to work better than direct forecasting. It is known that multivariate time series modeling put many problems concerning canonical representation of these series. The presented approach will need to be intensive investigated for other real time series, and the results compared with those provided in literature by other classical methods.

References

[1] A. Hyvärinen, J. Karhunen and E. Oja, *Independent Component Analysis*, John Wiley & Sons, Inc, 2001.
[2] P. Comon, "Independent Component analysis - a new concept ?", *Signal Processing*, vol. 36, pp. 287-314, 1994.

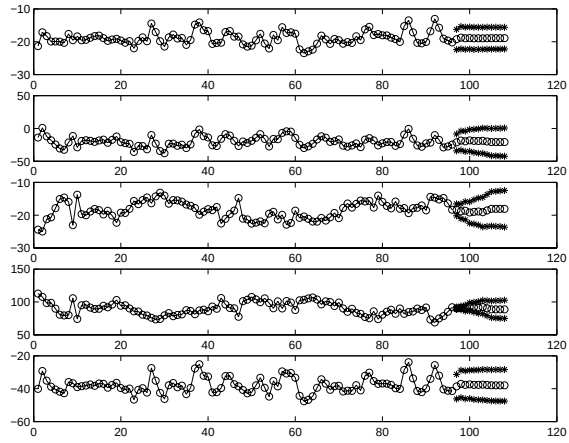


Fig. 6. Forecasting of the original time series

[3] A. Belouchrani, K. Abed Meraim, J.F. Cardoso and E. Moulines, "A blind source separation technique based on second order statistics", *IEEE Trans. on Signal Processing*, vol. 45, no. 2, pp. 434-444, 1997.
[4] M. Wax and T. Kailath, "Determining the number of signals by information theoretic criteria", *Workshop on spectral estimation II*, Florida, pp. 192-196, 1983.
[5] Y. Yin and P. Krishnaiah, "Methods for detection of the number of signals", *IEEE Trans. on ASSP*, vol. 35, no. 11, pp. 1533-1538, 1987.
[6] G. H. Golub and C.F.V. Loan, *Matrix Computation*, The John Hopkins University Press, 1989.
[7] A. Souloumiac and J.F. Cardoso, "Comparaison de methodes de separation de sources", *Proc. GRETSI*, Juan les Pines, 1991.
[8] A. Souloumiac and J.F. Cardoso, "Givens angles for simultaneous diagonalization", *SIAM J. Matrix Anal. Appl.*, 1994.
[9] J. F. Cardoso and A. Souloumiac, "Blind beamforming for non Gaussian signals, *IEE Proceedings - F*, vol. 140, no. 6, pp. 362-370, 1993.
[10] J. F. Cardoso and P. Comon, "Tensor based independent component analysis", *Proc. EU-SIPCO*, 1990.



Theodor D. Popescu was born in Rosiori de Vede, Romania, on July 4, 1949. He received M.Sc. degree and Eng.Sc.D. (PhD) degree, both in Automatic Control, from "Politehnica" University of Bucharest in 1972 and 1983, respectively. He is with National Institute for Research and Development in Informatics, Bucharest, where is senior research scientist of rs t degree and President of the Scienti c Council. The research activity was mainly oriented in the elds of system identi cation and parameter estimation, time series analysis, signal processing, change detection and diagnosis in signals and systems. He is author or co-author of more than 120 papers and of ve books.

He has been awarded with "Tudor Tanasescu" Prize of the Romanian Academy on 1995 for the works in the eld of detection of changes in dynamics of signals and systems and General Association of Engineers in Romania Prize on 2000 in the eld of Information Technology. He was granted by The Royal Society (1992), DAAD (1992, 1995, 1999), JSPS (1996, 1999), Australian Research Council(2002), CERGE-EI (2002). He is Senior Member of IEEE since 2002.

Open Science Index, Mathematical and Computational Sciences Vol:3, No:1, 2009 publications.waset.org/10199/pdf