computing@tanet.edu.te.ua
www.tanet.edu.te.ua/computing

ISSN 1727-6209
International Scientific
Journal of Computing

TIME SERIES PREDICTION USING ICA ALGORITHMS

Juan M. Górriz [1], Carlos G. Puntonet [2], Moisés Salmerón [2], E.W. Lang [3]

1) Dpto. Ingenierýá de Sistemas y Automática, Tec. Electrónica y Electrónica,
Universidad de Cádiz (Spain).
2) Dpto. Arquitectura y Tecnologýá de Computadores, Universidad de Granada (Spain)
3) Institute of Biophysics, University of Regensburg (Germany), juanmanuel.gorriz@uca.es

Abstract: *In this paper we propose a new method for volatile time series forecasting using Independent Component Analysis (ICA) algorithms and Savitzky-Golay filtering as preprocessing tools. The preprocessed data will be introduce in a based radial basis functions (RBF) Artificial Neural Network (ANN) and the prediction result will be compared with the one we get without these preprocessing tools or the classical Principal Component Analysis (PCA) tool.*

Keywords: - *Independent Component Analysis (ICA), Time Series Analysis, Neural Networks, Signal Processing*

1. INTRODUCTION

Different techniques have been developed in order to forecast time series using data from the stock. There also exist numerous forecasting applications like those ones analyzed in [19]: signal statistical preprocessing and communications, industrial control processing, Econometrics, Meteorology, Physics, Biology, Medicine, Oceanography, Seismology, Astronomy y Psychology.

A possible solution to this problem was described by Box and Jenkins [8]. They developed a time-series forecasting analysis technique based in linear systems. Basically the procedure consisted in suppressing the nonseasonality of the series, parameters analysis, which measure time-series data correlation, and model selection which best fitted the data collected (some specific order ARIMA model). But in real systems non-linear and stochastic phenomena crop up, thus the series dynamics cannot be described exactly using those classical models. ANNs have improved results in forecasting, detecting the non-linear nature of the data. ANNs based in RBFs allow a better forecasting adjustment; they implement local approximations to non-linear functions, minimizing the mean square error to achieve the adjustment of neural parameters. Platt´salgorithm [18], RAN (Resource Allocating Network), consisted in the control of the neural network's size, reducing the computational cost associated to the calculus of the optimum weights in perceptrons networks.

Matrix decomposition techniques have been used as an improvement of Platt model [24] with the aim of taking the most relevant data in the input space, for the sake of avoiding the processing of non-relevant information (NAPAPRED "Neural model with Automatic Parameter Adjustment for PREDiction"). NAPA-PRED also includes neural pruning [25].

The next step was to include the exogenous information to these models. There are some choices in order to do that; we can use the forecasting model used in [11] which gives good results but with computational time and complexity cost; Principal Component Analysis (PCA) is a well-established tool in Finance. It was already proved [24] that prediction results can be improved using the PCA technique. This method linear transform the observed signal into principal components which are uncorrelated (features), giving projections of the data in the direction of the maximum variance [17]. PCA algorithms use only second order statistical information; Finally, in [4] we can discover interesting structure in finance using the new signal-processing tool Independent Component Analysis (ICA). ICA finds statistically independent components using higher order statistical information for blind source separation ([5], [14]). This new technique may use Entropy (Bell and Sejnowski 1995, [7]), Contrast functions based on Information Theory (Comon 1994, [9]), Mutual Information (Amari, Cichocki y Yang 1996, [3]) or geometric considerations in data distribution spaces (Carlos G. Puntonet 1994 [21],[27], [1], [22], [2]), etc. Forecasting and analyzing financial time series using ICA can contributes to a better understanding and prediction of financial markets ([23],[6]).

2. BASIC ICA

ICA has been used as a solution of the blind source separation problem [15] denoting the process of taking a set of measured signal in a vector, **x**, and extracting from them a new set of statistically independent components (ICs) in a vector **y**. In the basic ICA each component of the vector **x** is a linear instantaneous mixture of independent source signals in a vector **s** with some unknown deterministic mixing coefficients:

$$x_i = \sum_{i=1}^{N} a_{ij} s_j \tag{1}$$

Due to the nature of the mixing model we are able to estimate the original sources $\tilde{s}_i$ and the de-mixing weights b_{ij} applying i.e. ICA algorithms based on higher order statistics like cumulants.

$$\tilde{s}_i = \sum_{i=1}^{N} b_{ij} x_j \tag{2}$$

Using vector-matrix notation and defining a time series vector $\boldsymbol{x} = (x_1, \ldots, x_n)^T$, $\boldsymbol{s}$, $\tilde{s}$ and the matrix $\mathbf{A} = \{a_{ij}\}$ and $\mathbf{B} = \{b_{ij}\}$ we can write the overall process as:

$$\tilde{s} = Bx = BAs = Gs \tag{3}$$

where we define **G** as the overall transfer matrix. The estimated original sources will be, under some conditions included in Darmois-Skitovich theorem (chapter 1 in [10]), a permuted and scaled version of the original ones. Thus, in general, it is only possible to find **G** such that **G** = **PD** where **P** is a permutation matrix and **D** is a diagonal scaling matrix.

In Financial time series this model (equation 1) can be applied to the stock series where there are some underlying factors like seasonal variations or economic events that affect the stock time series simultaneously and can be assumed to be quite independent [16].

3. PREPROCESSING TIME SERIES WITH ICA+FILTERING

The main goal, in the preprocessing step, is to find nonvolatile time series including exogenous information i.e. fi- nancial time series, easier to predict using ANNs based on RBFs. This is due to smoothed nature of the kernel functions used in regression over multidimensional domains [13]. We propose the following Preprocessing Steps

- After Whitening the set of time series $\{x_i\}_{i=1}^{n}$ (subtract the mean of each time series and removing the second order statistic effect or covariance matrix diagonalization process)
- We apply some ICA algorithm to estimate the original sources *si* and the mixing matrix **A** in equation 1. Each IC has information of the stock set weighted by the components of the mixing matrix. In particular, we use an equivariant robust ICA algorithm based in cumulants proposed in [10]. The de-mixing matrix is calculated according the following iteration:

$$\mathbf{B}^{(n+1)} = \mathbf{B}^{(n)} + \mu^{(n)} (C_{s,s}^{1,\beta} S_s^{\beta} - I)\, \mathbf{B}^{(n)} \tag{4}$$

where **I** is the identity matrix, $C_{s,s}^{1,\beta}$ is the β +1 order cumulant of the sources (we chose β = 3 in simulations) and S_s^{β} = *diag*(*sign*(*diag*($C_{s,s}^{1,\beta}$))).

Once convergence, which is related with crosscumulants[1] absolute value, is reached, we estimate the mixing matrix inverting **B**.

• Filtering.

1. We neglect non-relevant components in the mixing matrix **A** according to their absolute value. We consider the rows **A***i* in matrix **A** as vectors and calculate the mean Frobenius norm[2] of each one. Only the components bigger than mean Frobenius norm will be considered. This is the principal preprocessing step using PCA tool, in this case is not enough.

$$\tilde{A} = Z \cdot A \tag{5}$$

where $\{\mathbf{Z}\}_{ij} = \left[\{A\}_{ij} > \frac{\|A_i\|_{Fr}}{n} \right]$

2. We apply a low band pass filter to the ICs. We choose the well-adapted for data smoothing Savitsky-Golay smoothing filter [26] for two reasons: *a*) ours is a real-time application for which we must process a continuous data stream and wish to output filtered values at the same rate we receive raw data and *b*) the quantity of data to be processed is so large that we just can afford only a very small number of floating operations on each data point thus computational cost in frequency domain for

[1] cumulants between different sources

[2] Given $x \in R^n$, its Frobenius norm is $\|x\|_{Fr} \equiv \sqrt{\sum_{i=1}^{n} x_i^2}$

high dimensional data is avoided even the modest-sized FFT. This filter is also call Least-Squares [12] or DISPO [29]. These filters derive from a particular formulation of the data smoothing problem in the time domain and their goal is to find filter coefficients c_n in the expression:

$$\bar{s}_i = \sum_{n=-n_L}^{n_R} c_n s_{i+n} \qquad (6)$$

where $\{s_{i+n}\}$ represent the values for the ICs in a window of length $n_L + n_R + 1$ centered on i and $\tilde{s}_i$ is the filter output (the smoothed ICs), preserving higher moments [20].

For each point s_i we least-squares fit a m order polynomial for all $n_L + n_R + 1$ points in the moving window and then set $\tilde{s}_i$ to the value of that polynomial at position i. As shown in [20] there are a set of coefficients for which equation 6 accomplishes the process of polynomial leastsquares fitting inside a moving window:

$$c_n = \{(M^T \cdot M)^{-1}(M^T \cdot e_n)\}_0 = = \sum_{j=0}^{m} \{(M^T \cdot M)^{-1}\}_{0j} \cdot n^j \qquad (7)$$

where $\{\mathbf{M}\}_{ij} = i^j$, $i = -n_L, \ldots, n_R, j = 0, \ldots, m$, and $\mathbf{e}_n$ is the unit vector with $-n_L < n < n_R$.

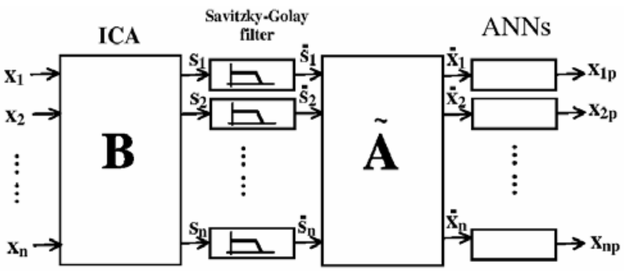

Fig. 1 – Schematic representation of prediction and filtering process.

- Reconstructing the original signals using the smoothed ICs and filtered $\tilde{A}$ matrix we get a less high frequency variance version including exogenous in- fluence of the old ones. We can write using equations 5 and 4.

$$x = \tilde{A} \cdot \bar{s} \qquad (8)$$

4. TIME SERIES FORECASTING MODEL

We use an ANN using RBFs to forecast a series *xi* from the Stock Exchange building a forecasting function **P** with the help of NAPA-PRED+LEC algorithm [23], for one of the set of signals $\{x_1, \ldots, x_n\}$. As shown in [24] the individual forecasting function can be expressed in term of RBFs [28] as:

$$F(x) = \sum_{i=1}^{N} f_i(x) = \sum_{i=1}^{N} h_i \exp\{\frac{\|x - c_i\|}{r_i^2}\} \qquad (9)$$

where $\boldsymbol{x}$ is a p-dimensional vector input at time **t**, **N** is the number of neurons (RBFs) , $\mathbf{f}_i$ is the output for each neuron i-th , $\mathbf{c}_i$ is the centers of i-th neuron which controls the situation of local space of this cell and $\mathbf{r}_i$ is the radius of the i-th neuron. The global output is a linear combination of the individual output for each neuron with the weight of $\mathbf{h}_i$. Thus we are using a method for moving beyond the linearity where the core idea is to augment/replace the vector input **x** with additional variables, which are transformations of **x**, and then use linear models in this new space of derived input features. RBFs are one of the most popular kernel methods for regression over the domain R^n and consist on fitting a different but simple model at each query point $\mathbf{c}_i$ using those observations close to this target point in order to get a smoothed function. This localization is achieved via a weighting function or kernel $\mathbf{f}_i$.

The preprocessing step suggested in section 3 is necessary due to the dynamic of the series [23] and it will be shown that results improve sensitively. Thus we will use as input series the ones we got in equation 8.

5. SIMULATIONS

We work with indexes of different Spanish banks and companies during the same period to investigate the effectiveness of ICA techniques for financial time series (top figure in Fig. 2). We have specifically focussed on the IBEX35 from Spanish stock, which we consider the most representative sample of the Spanish stock movements, using closing prices in 2000.

We considered the closing prices of Bankinter for prediction and 8 more stock of different Spanish companies (ACS, Aguas de Barcelona, Banco Popular, Banco Santander, BBVA, Dragados, Carrefour and Amadeus). Each time series includes 200 points corresponding to selling days (quoting days).

We performed ICA on the Stock returns using the ICA algorithm presented in section 3 assuming that the number of stocks equals the number of sources supplied to the mixing model. This algorithm whiten the raw data as the first step. The ICs are shown in the middle-top part of Fig. 2. These ICs represents independent and different underlying factors like seasonal variations or economic events that affect the stock time series simultaneously. Via the rows of **A** we can reconstruct the original signals with the help of these ICs i.e. Bankinter stock after we preprocess the raw data:

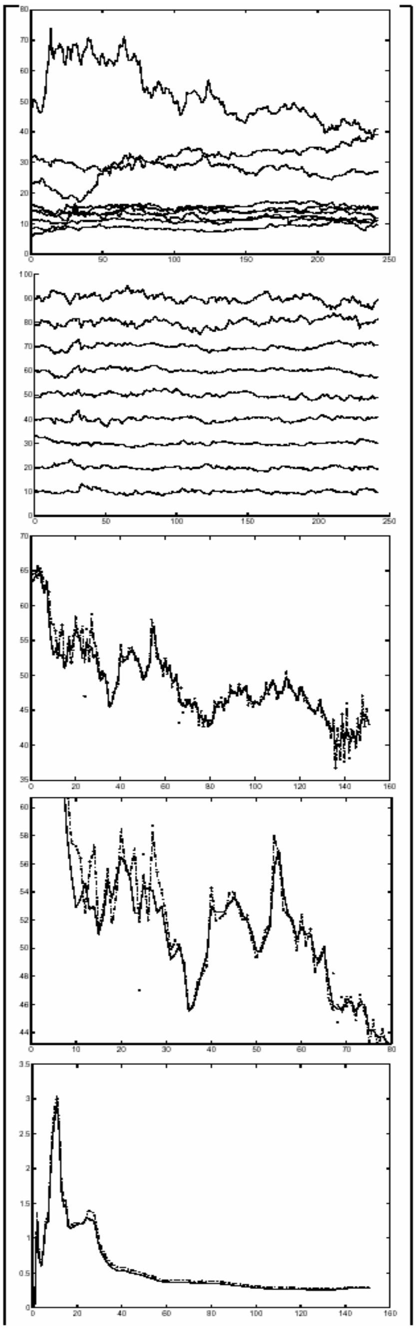

Fig. 2 – Simulations. Top: set of Stock Series. Middle-top: ICs for the stock series. Middle: Real Series(line; predicted Series with ICA+SG(dash-dotted); predicted Series without preprocessing (dotted); Middle-bottom: zoomed version of previous figure. Bottom: NRMSE evolution for ANN with ICA+SG (line) and without (dotted line).

- Frobenius Filtering: the original mixing matrix[3]:

$$A = \begin{pmatrix} \ddots & \vdots & \vdots & \vdots & \vdots \\ \dots & 0.33 & -0.23 & 0.17 & 0.04 \\ \dots & -0.28 & 1.95 & -0.33 & 4.70 \\ \dots & 0.33 & -0.19 & 0.05 & -0.23 \\ \dots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}$$

is transformed to:

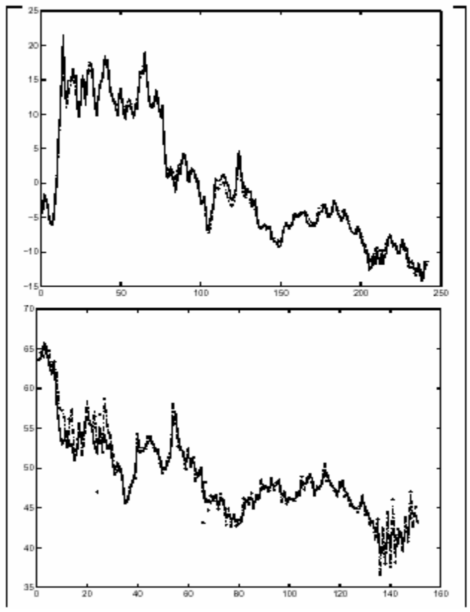

Fig. 3 – Simulations (continued). Top: real Series from ICA reconstruction (scaled old version)(line) and Preprocessed Real Series (dotted line). Bottom: same figure, zoomed.

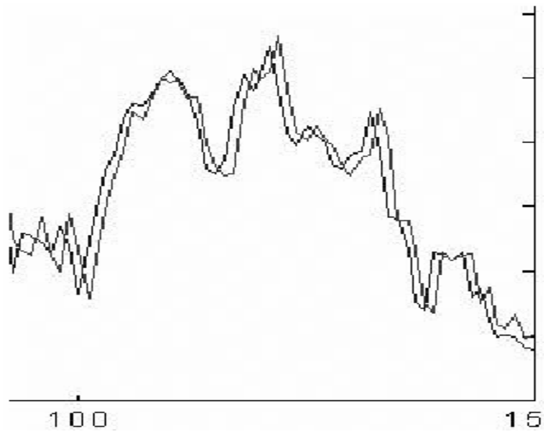

Fig. 4 – Delay problem in ANNs

$$\overline{A} = \begin{pmatrix} \ddots & \vdots & \vdots & \vdots & \vdots \\ \dots & 0.33 & -0.23 & 0.17 & 0.04 \\ \dots & 0 & 1.95 & 0 & 4.70 \\ \dots & 0.33 & -0.19 & 0.05 & -0.23 \\ \dots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}$$

thus we neglect the influence of two ICs on the original 5^{th} stock. Thus only a few ICs

[3] We show and select the relevant part of the row corresponding to the Bankinter Stock.

contribute to most of the movements in the stock returns and each IC contributes to a level change depending its amplitude transient [4].

- We do a polynomial fit in the ICs using the library supported by *MatLab* and the reconstruction of the selected stock (see Fig. 3) to supply the ANN.

In the three last figures in Fig. 2 we show the results we got using our ANN with the above mentioned algorithm. We can say that prediction is better with the preprocessing step avoiding the disturbing peaks or convergence problems in prediction. As is shown in the lowermost figure, the NRMSE is always lower using the techniques we discussed in section 4.

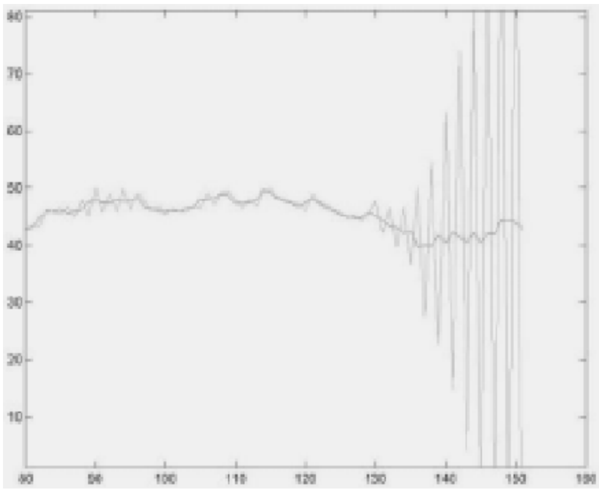

Fig. 5 – Curse of dimensionality.

Finally with these models we avoid the curse of dimensionality or difficulties associated with the feasibility of density estimation in many dimensions presented in AR or ANNS models [23] with high number of inputs as shown in fig. 5 and the delay problem presented in non relevant time periods of prediction (Fig. 4).

6. CONCLUSIONS

In this paper we showed that prediction results can be improved with the help of techniques like ICA. ICA decompose a set of 9 returns from the stock into independent components which fall in two categories: *a*) large components responsible of the major changes in level prices and *b*) small fluctuations. Smoothing this components and neglecting the non-relevant ones we can reconstruct a new version of the Stock easier to predict. Moreover we describe a new filtering method to volatile time series that are supplied to ANNs in real-time applications.

7. REFERENCES

[1] C.G. Puntonet A. Mansour, N. Ohnishi. Blind multiuser separation of instantaneous mixture algorithm based on geometrical concepts, Signal Processing 82 (2002). pp. 1155–1175.

[2] M. Rodrguez Alvarez, C.G. Puntonet, I. Rojas. Separation of sources based on the partitioning of the space of observations, Lecture Notes in Computer Science 2085 (2001). pp. 762–769.

[3] S. Amari, A. Cichocki, H. Yang. A new learning algorithm for blind source separation, Advances in Neural Information Processing Systems, MIT Press 8 (1996). pp. 757–763.

[4] A. D. Back, A. S. Weigend. Discovering structure in finance using independent component analysis, Computational Finance (1997).

[5] Andrew D. Back, Thomas P. Trappenberg. Selecting inputs for modelling using normalized higher order statistics and independent component analysis, IEEE Transactions on Neural Networks 12 (2001).

[6] Andrew D. Back, A. S. Weigend. Discovering structure in finance using independent component analysis, 5th Computational Finance 1997 (1997).

[7] A. J. Bell, T. J. Sejnowski. An information maximization approach to blind separation and blind deconvolution, Neural Computation 7 (1995), pp. 1129 – 1159.

[8] G.E.P. Box, G.M. Jenkins, G.C. Reinsel. Time series analysis. forecasting and control. Prentice Hall, 1994.

[9] P. Comon. Independent component analysis: A new concept?, Signal Processing 36 (1994). pp. 287–314.

[10] S. Cruces. An unified view of bss algorithms(in spanish). University of Vigo, Spain, 1999.

[11] J.M. Gorriz, J.J.G. dela Rosa, Carlos G. Puntonet, M. Salmeron Campos, New model for time-series forecasting using rbf's and exogenous data, In Press (2003).

[12] R.W. Hamming. Digital filters, 2a ed., Prentice Hall, 1983.

[13] T. Hastie, R. Tibshirani, J. Friedman, The elements of statistical learning, Springer, 2000.

[14] A. Hyvarinen, E. Oja. Independent component analysis: Algorithms and applications, Neural Networks 13 (2000). pp. 411–430.

[15] A. Hyvrinen, E. Oja. Independent component analysis: algorithms and applications, Neural Networks 1 (2000). pp. 411–430.

[16] K. Kiviluoto, E. Oja, Independent component analysis for paralell financial time

series. Proc. in ICONIP98 1 (1998), pp. 895–898.

[17] T. Masters. Neural, novel and hybrid algorithms for time series analysis prediction. John Miley & Sons, 1995.

[18] J. Platt. A resource-allocating network for function interpolation, Neural Computation 3 (1991). pp. 213–225.

[19] D.S.G. Pollock. A handbook of time series analysis, signal processing and dynamics. Academic Press, 1999.

[20] W. H. Press, S. A. Teukolsky, W. T. Vertterling, B. P. Flannery. Numerical recipes in c++, 2ª ed. Cambridge University Press, 2002.

[21] C.G. Puntonet. Nuevos Algoritmos de Separacin de Fuentes en Medios Lineales. Ph.D. thesis, University of Granada, Departamento de Arquitectura y Tecnologia de Computadores, 1994.

[22] C.G. Puntonet, Ali Mansour. Blind separation of sources using density estimation and simulated annealing, IEICE Transactions on Fundamental of Electronics Communications and Computer Sciences E84-A (2001).

[23] J. M. Gorriz Saez. Prediccion y Tecnicas de Separacion de Senales. Ph.D. thesis, University of Cadiz, Departamento de Ing. de Sistemas y Aut. Tec. Eleectronica y Electronica, 2003.

[24] M. Salmern-Campos. Prediccin de Series Temporales con Redes Neuronales de Funciones Radiales y Tecnicas de Descomposicion Matricial. Ph.D. thesis, University of Granada, Departamento de Arquitectura y Tecnologýa de Computadores, 2001.

[25] Moises Salmeron, Julio Ortega, Carlos G. Puntonet, Alberto Prieto. Improved ran sequential prediction using orthogonal techniques, Neurocomputing 41 (2001). pp. 153–172.

[26] A. Savitzky, M.J.E. Golay. Analytical Chemestry 36 (1964). pp. 1627–1639.

[27] F. J Theis, A. Jung, E.W. Lang, C.G. Puntonet, Multiple recovery subspace projection algorithm employed in geometric ica, in press on Neural Computation (2001).

[28] J. Moody, C. J. Darken. Fast learning in networks of locally-tuned processing units, Neural Computation 1 (1989). pp. 284–294.

[29] H. Ziegler. Applied Spectroscopy 35 (1981). pp. 88-92.

Juan Manuel Górriz Sáez *was born in Almerнa, Spain, on May 27, 1976. He received two B.Sc. degrees in 2000 from the university of Granada, Spain and his Ph.D. degree in 2003 from the University of Cádiz, Spain. These degrees are in physics and electroninc engineering.*

He has been Visiting Researcher at the “CPR_FEC” (Munich, Germany), at the Infineon Tech. AG. Currently, he is an Assistant Professor at the department of "Electronics" at the University of Cádiz, Spain. His research interests lie in the fields of signal processing, statistical Learning Theory, artificial neural networks and optimization methods.

Carlos G. Puntonet *was born in Barcelona, Spain, on August 11, 1960. He received the B.Sc. degree in 1982, the M.Sc. degree in 1986 and his Ph.D. degree in 1994, all from the University of Granada, Spain. These degrees are in electronics physics.*

He has been Visiting Researcher at the УLaboratorie de Traitement dTImages et Reconnaissance de FormesФ (INPG, Grenoble, France), at the Institute of Biophysics (Regensburg, Germany) and at the Institute of Physical and Chemical Research (RIKEN, Nagoya, Japan). Currently, he is an Associate Professor at the department of "Architecture and Computer Technology" at the University of Granada, Spain. His research interests lie in the fields of signal processing, independent component analysis and blind separation of sources, artificial neural networks and optimization methods.

Moisés Salmerón *received his B.Sc. degree in 1994, his M.Sc. degree in 1997, and his PhD in 2001, all in Computer Science from the University of Granada, Spain, and is completing a second M.Sc. in Statistics from the same University. His last-year engineering project was about the VLSI design of testable neural networks. In 1997 he joined the CASIP (Circuits and Systems for Information Processing) research group at the Department of Computer Architecture and Computer Technology (University of Granada). His PhD thesis (which obtained one of the first spanish European Doctorate Mentions in Informatics) dealed with hybridization of neural networks and statistical techniques for time series forecasting. He has published over 40 papers both in specialized journals and conferences. Also, he is associate member of the IEEE and student (graduate) member of SIAM and the ACM, and has contributed as judge and reviewer for the IEEE CSIDC engineering competition and for several scientific journals. He* *currently teaches Computer Technology, Design of Microelectronic Circuits and a doctorate course dealing with Statistical Theory of Adaptive Systems. His research interests lie in the fields of Support Vector Machines (SVMs), neural networks, theory of adaptive systems and statistical modelling for time series prediction.*

Elmar W. Lang *was born in Regensburg, Germany in July 19, 1951. He obtained a degree in Physics in 1977 and received his PhD in 1980. In 1988 he finished his Habilitation in Biophysics and became apl.Professor in 1994 at the University of Regensburg. His research interests include statistical signal processing, independent component analysis and blind source separation and medical image analysis.*