Component Analysis and Time Series Prediction

Eduardo Gómez-Ramírez*, Josep Ma. Martorell-Rodon†, Xavier Vilasís-Cardona‡

*Coordinación de Investigación, Universidad La Salle,
Benjamin Franklin 47, Colonia Condesa, CP 06140 México D.F. - México
egomez@ci.ulsa.mx
†Departament de Comunicacions i Teoria del Senyal,
Enginyeria i Arquitectura La Salle, U. Ramon Llull,
Pg. Bonanova 8, 08022 Barcelona.
jmmarto@salleURL.edu
‡Departament d'Electrònica, Enginyeria i Arquitectura La Salle, U. Ramon Llull,
Pg. Bonanova 8, 08022 Barcelona.
xvilasis@salleURL.edu

Abstract

The relation between component analysis (PCA and ICA) and Multi-resolution Filtering is explained and its use to improve the prediction error in a time series is shown. An example applying the previous results to forecast the Sun-spots time series is presented.

Keywords system identification, time series prediction, multi-resolution filters, component analysis.

1 Introduction

There is no need to praise the importance of time series analysis and prediction. The spread of fields where such techniques are applied, ranging from experimental physics to finance and economy through process control, is enough to point out its relevance. In this note we propose the combination of two methods in order to enhance their strong points : multi-resolution filters and independent component analysis.

In few words, multi-resolution filters consist in a recursive decomposition of a time series into several sub-series, according to sum and difference operators [1, 2]. Its use in time series prediction is based on forecasting the results of this decomposition to build up a prediction for the original series [1, 3]. This recursive generation can be cast into an adaptive spanning tree in order to choose the best possible prediction.

Component analysis is a method to separate the contribution of a number of sources on a measured signal. According to the probability distribution of those sources, we distinguish between Principal Component Analysis (PCA) [4] for Gaussian signals and Independent Component Analysis (ICA) for non-Gaussian ones [5, 6, 7, 8].

ICA has seldomly been used to analyse univariate time series [9, 10, 11]. This problem is slightly different than predicting multivariate time series [12, 13], where each component of the time series acts as different source. In the univariate case, independent component analysis has been found to extract the state variable structure of a non-linear dynamical system [9].

The note unfolds as follows. We first recall some facts about signal whitening and its application to time series. Then multi-resolution filters are sketched and component analysis described. Both methods show to be particular cases of whitening and, so, the tree expansion of multi-resolution filters is extended to the ICA decomposition. The results of applying this methodology to the prediction of the sun spots time series are then shown. Some short conclusions close the paper.

2 Whitening and time series

A common preprocessing feature in signal processing is the so-called whitening process. It just consists in transforming the original signal to obtain a new set of decorrelated components $\{\bar{Y}(k)\}$:

$$Y(k) = A\bar{Y}(k). \tag{1}$$

The form of matrix A is easily computed. In effect, consider the correlation matrix of the signal vector $Y(k)$,

$$C = \sum_k Y(k)Y(k)^T. \tag{2}$$

If we ask the sources to be decorrelated,

$$\sum_k \bar{Y}(k)\bar{Y}(k)^T = \mathbf{1},$$

we obtain

$$C = AA^T. \tag{3}$$

All solutions of equation (3) are of the form (see, for instance, reference [6])

$$A = A_{\text{CH}}R, \tag{4}$$

being A_{CH} the Cholesky decomposition [18] of C and R an N-dimensional orthogonal matrix.

All solutions can thus be parametrised using the angles of an orthogonal matrix in N dimensions, which belongs to a $N(N-1)/2$ dimensional continuous Lie group. In the case of a two component splitting, we just have a single parameter, namely the rotation angle.

In order to use the setting described above for univariate time series modelling and prediction, we need to build up a signal vector, which shall be composed the tapped delayed observations of the series,

$$Y(k)^T = (y(k), y(k-\tau), \ldots, y(k-(N-1)\tau))\,. \tag{5}$$

The choice for the correct values of τ and N in order to obtain an accurate description of the dynamical process is a matter of subtle discussion (see, for instance, reference [19]).

If the time series comes from the evolution of a deterministic dynamical system, vector $Y(k)$ in equation (5) corresponds, after Takens' theorem [20], to some state variable vector of the system.

The application of a whitening process to $Y(k)$ in equation (5) actually amounts to applying a FIR filter on the series. If we look at the Fourier transform of these filters we recover a number of high pass, low pass and band pass characteristics, where rotation angles select the bandwidths.

To illustrate this fact let us concentrate in two-dimensional filters. Without loss of generality, we can consider that the signal has unit variance (which is just an overall normalisation choice). The signal vector has the form

$$Y(k)^T = (y(k), y(k-\tau))\,, \tag{6}$$

and the correlation matrix can be parametrised as

$$C = \begin{pmatrix} 1 & x \\ x & 1 \end{pmatrix}. \tag{7}$$

The Cholesky decomposition of this matrix is then

$$A_{\text{CH}} = \begin{pmatrix} 1 & 0 \\ x & \sqrt{1-x^2} \end{pmatrix}. \tag{8}$$

Multiplying by a rotation transformation of angle θ, we get the general form for the whitening filter,

$$A = \begin{pmatrix} \cos\theta & \sin\theta \\ x\cos\theta - \sqrt{1-x^2}\sin\theta & x\sin\theta + \sqrt{1-x^2}\cos\theta \end{pmatrix}. \tag{9}$$

By inverting the whitening transformation,

$$\bar{Y}(k) = A^{-1}Y(k) = WY(k), \tag{10}$$

we obtain the filtering matrix W,

$$W = \frac{1}{\sqrt{1-x^2}} \begin{pmatrix} x\sin\theta + \sqrt{1-x^2}\cos\theta & -\sin\theta \\ -x\cos\theta + \sqrt{1-x^2}\sin\theta & \cos\theta \end{pmatrix}. \tag{11}$$

Each row of matrix W displays the coefficients of the filter applied to obtain the corresponding component of the whitened signal. Assuming the series is regularly sampled, the nature of the filter response can be seen studying their Fourier transform, in particular the squared modulus. If we denote by T the sampling time and we take $\tau = 1$, for the first filter we have

$$|w_1(\omega T)|^2 = \frac{1}{\sqrt{1-x^2}} \left(1 - x^2\cos 2\theta + x\sqrt{1-x^2}\sin 2\theta + \left(-x + x\cos 2\theta + \sqrt{1-x^2}\sin 2\theta\right)\cos\omega T\right), \tag{12}$$

and for the second,

$$|w_2(\omega T)|^2 = \frac{1}{\sqrt{1-x^2}} \left(1 + x^2\cos 2\theta - x\sqrt{1-x^2}\sin 2\theta + \left(-x - x\cos 2\theta + \sqrt{1-x^2}\sin 2\theta\right)\cos\omega T\right). \tag{13}$$

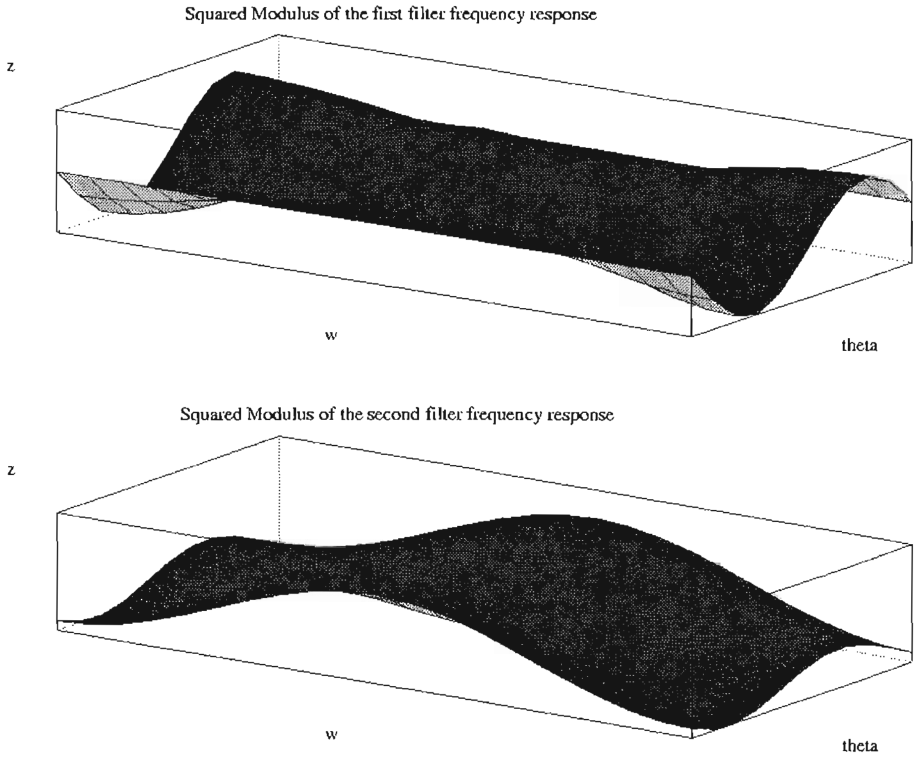

Figure 1: Modulus response of the two-dimensional whitening filters at lag 1

These functions are plotted in figure 1 where it can be appreciated how the nature of the filter (high pass, low pass) changes with the value of θ.

3 Multi-resolution filtering

Inspired in the wavelet transform techniques, a multi-resolution filter splits a time series into a number of sub-series recursively according to sum and difference operators [1, 2]. Let us unfold this description following a typical case.

Consider a time series $\{y(k)\}$. We construct two sub-series, by performing the following linear operation,

$$\begin{cases} s_1(k) = \frac{1}{2}\left(y(k) + y(k-1)\right), \\ s_2(k) = \frac{1}{2}\left(y(k) - y(k-1)\right). \end{cases} \tag{14}$$

These sum and difference operators have several interpretations. They can be viewed as an integral and a derivative or, in signal processing terms, as a low pass and high pass filters. In any case, two are the important properties of this decomposition :

1. the original series is reconstructed by summing up its terms,
$$y(k) = s_1(k) + s_2(k),$$
2. resulting series are decorrelated,
$$\sum_{k=-\infty}^{\infty} s_1(k)s_2(k) = 0.$$

These two properties justify the use of this technique in time series prediction. In effect, if we denote $\hat{s}_1(k)$ and $\hat{s}_2(k)$ two predictions of $s_1(k)$ and $s_2(k)$, a prediction for the original series, $\hat{y}(k)$, is immediately obtained,

$$\hat{y}(k) = \hat{s}_1(k) + \hat{s}_2(k). \tag{15}$$

The quadratic error of this prediction is directly the sum of the quadratic errors of the prediction of each sub-series because of the decorrelation property.

Recursively, series $s_1(k)$ and $s_2(k)$ can be decomposed following equation (14) and generate in turn new series $s_{11}(k)$, $s_{12}(k)$, $s_{21}(k)$, $s_{22}(k)$. This process can go on repeatedly. A binary tree whose root is the original series $y(k)$ is then generated. Using the same process described in equation (15), we can reconstruct a prediction for $y(k)$ with a prediction of the series in the leaves of the tree. The quality of this prediction depends, of course, on what leaves are used.

An inspection to equations (14) and (2) reveals that, up to normalisation, the filtering procedure is a whitening process.

4 Component analysis

Component Analysis is a statistical method to separate random signals using N measurements of the mixing of N decorrelated sources. If we call $Y(k)$ the measured signals and $S(k)$ its sources, we look for a linear relation,

$$Y(k) = \mathcal{U}S(k). \tag{16}$$

This transformation can be found using second order methods. They consist in finding a faithful representation of data by minimising the mean square error [5]. This is equivalent to assume that the signal follows a Gaussian probability distribution. The resulting sources S follow a Gaussian distribution and are mutually decorrelated, which amounts to be independent in this Gaussian approximation. This method is known as Principal Component Analysis [5].

For non-Gaussian distributed signals, the separation is based on the minimisation of the mutual information between signals and sources. Recall the mutual information is defined after Shannon's entropy. If the signal Y is distributed according to a probability density function $p(Y)$, the entropy is

$$H(Y) = -\int dY\; p(Y) \ln p(Y). \tag{17}$$

Once the entropy is known, the mutual information is between Y and the sources S is given by

$$I(Y,S) = H(Y) - H(S/Y), \tag{18}$$

where $H(S/Y)$ denotes the entropy for the conditional distribution of the sources with respect to the signal.

Minimising the mutual information is equivalent to look for independent sources fulfilling equation (16). This is a stronger requirement than mere decorrelation. The method is called Independent Component Analysis [5, 6, 7, 8]. Under these conditions, the mutual information (18) can be written

$$I(Y,S) = H(Y) - \sum_{i=0}^{N} H(S_i) \tag{19}$$

and is usually evaluated by the Fraser-Swinney algorithm [17].

Linear component analysis, given by equation (16), either PCA or ICA, delivers decorrelated sources and, therefore, can be considered particular cases of whitening. As such they can be applied to univariate time series analysis. In this context, Independent Component Analysis is preferred since usual time series exhibit non-Gaussian probability distribution features. Then it may be used to improve the prediction lying on the fact that predicting independent individual components may be more precise than predicting the original signal. Component predictions $\hat{S}_i(k)$ can be combined using formula (16) to deliver a prediction for the original series:

$$\hat{y}(k) = \sum_j A_{1j}\hat{S}_j(k). \quad (20)$$

In the same manner described for multi-resolution filters, a tree of series can be defined by successive splitting of the resulting components. Numerically, this can be achieved using, for instance, the FASTICA algorithm (see for instance reference [5]).

5 Adaptive Filters

Both for multi-resolution filters and component analysis rotations, a time series prediction scheme has been defined by splitting of the original univariate series $y(k)$ into a tree of components. The spanning of the tree can easily be coded into a genetic algorithm [14, 15] to explore what are the best leaves to perform the prediction. We use the final quadratic error on the original series prediction as the fitness function. The operations performed on the chromosomes shall be the standard crossover or sexual recombination, mutation and a special process we call add parents [16]. We define in this way adaptive predictive filters.

6 Results on the Sun Spots time series

These techniques are illustrated on the well known Sun Spots time series. The choice of the Wolf number for measuring the solar activity by counting the number of sun spots is based on three criteria. The first is because it is a real experimental data time series, from which any internal dynamics is unknown. The second reason is its benchmark condition in the realm of time series. The third is the recovered interest on this measurement of the solar activity boosted by the reaching of a maximum activity period. Our data set consists in 280 yearly values starting in year 1700. Our goal is to predict at lag 1 the value of the series using 5 previous values. The quality of the prediction is measured using the mean square error. Results are obtained training the networks with a training set of 220 points and two probe sets, the first one from 221 to 255 and second one from 256 to 280.

Our goal in this note is to compare predictions made with and without filter rather than comparing the final prediction with the standard benchmark results for this series (see for instance reference [21] being one of the pioneering works). Actually, best predictions are not obtained using 5 previous values.

Results comparing the performance of a second order polynomial fit for a straight prediction with no filter, using adaptive multi-resolution filters and using adaptive ICA filters is shown in table 1.

Table 2 shows the errors for the prediction made by multi-layer perceptron with architecture 5:6:1 trained with the Levenberg-Marquardt algorithm.

Despite numerical results for these methods are of the same order of magnitude, ICA-filtered errors are slightly lower in all cases. By no means should we take these results as a solid proof of the goodness of the ICA-filtering , yet they are a clear hint of its capabilities. They show the worth of undergoing further simulations to improve the genetic algorithm parameters, or even tune an automated method to select the appropriate delay for each series, not to mention the prediction of other standard benchmarks. Actually, the strong point of these adaptive methods faced to other semi-automatic prediction methods is the advantage one can take out of the selection of signals whose frequency content is strongly restricted. Nevertheless, an excessive division of the original series may result in no improvement. It is the duty of the genetic algorithm to find the appropriate equilibrium between splitting and predictability. Thus, the asset of whitening-based filtering is the information on the frequency content of the signal extracted by the predictor which may be of help in understanding any underlying phenomena. That is why we may state that these methods head in the direction of defining an automated prediction system whose mechanism may be easily interpreted.

	Training Set	Test 1	Test 2
No Filter	0.0042	0.0039	0.0104
Multi-Resolution	0.0042	0.0039	0.0103
ICA ($\tau = 11$, $d = 2$)	0.0041	0.0038	0.0102

Table 1: Mean square error for a second order polynomial fit using 5 previous values

	Training Set	Test 1	Test 2
No Filter	0.0028	0.0066	0.0699
ICA ($\tau = 11$, $d = 2$)	0.0018	0.0044	0.0111

Table 2: Mean square error for the prediction using 5 previous values

7 Conclusions

This work is the junction of two converging lines to improve time series prediction : multi-resolution filters and ICA. We have shown that both methods are related by being to ways of whitening the signal.

To improve the behaviour of multi-resolution and ICA filters, we have put forward the adaptation of the spanning tree by a genetic algorithm. This paradigm has delivered slightly better results than the straight prediction for the Sun spots time series.

In any case, work is currently in process to confirm the presented results and hypotheses by enlarging the number of benchmark series.

References

[1] S.Soltani, S.Canu, C.Boichu, Time series prediction and the wavelet transform, *International Workshop on Advanced Black Box Modelling*, Leuven, Belgium, July 1998.

[2] I.Daubechies, Ten lectures on wavelets, *CBMS-NSF Regional Conference Series on Applied Mathematics* **61**, SIAM 1992.

[3] E.Gómez-Ramírez, S.Soltani, A.González-Yunes, M.Avila-Alvarez, Improving learning process for Identification with Multi-resolution Filtering in Polynomial Artificial Neural Networks, *IASTED International Conference Intelligent Systems and Control*, October 28-30,1999 Santa Barbara, California, USA.

[4] I.T.Jolliffe, *Principal Component Analysis* (Springer Verlag, 1986).

[5] A.Hyvärinen, Survey on Independent Component Analysis *Neural Computing Surveys*, **2**, 1999, 94-128.

[6] A.Hyvärinen, Independent component analysis: algorithms and applications *Neural Networks*, **13**, 2000, 411-430.

[7] J-F.Cardoso, Blind signal separation: statistical principles, *Proceedings of the IEEE*, **9** 1998, 2009-2025.

[8] P.Comon, Independent component analysis, A new concept ? *Signal Processing* **36**, 1997, 287-314.

[9] G.Deco, B.Schürmann, Learning time series evolution by unsupervised extraction of correlations, *Phys.Rev.* **E51**, 1995, 1780-1790.

[10] R.H.Lesch, Y.Caille, D.Lowe, Component analysis in financial time series,*Proceedings IEEE/IAFE 1999 Conference on Computational Intelligence for Financial Engineering*, 183-190, (IEEE Press, 1999).

[11] E.Gómez-Ramírez, J.M.Martorell-Rodon, Xavier Vilasís-Cardona, Multi-resolution filters and component analysis for time series prediction, *Proceedings IASTED Conference on Modelling Identification and Control 2001* 848-852 (ACTA Press, 2001).

[12] J.Storck, G.Deco Nonlinear independent component analysis and multivariate time series analysis, *Physica* **D108**, 1997, 335-349.

[13] A.D.Back, A.S.Weigend, A first application of independent component analysis to extracting structure from stock returns, *Int. J. of Neural Systems*, **8(4)**, 1997, 473-484.

[14] D.E.Goldberg, *Genetic algorithms in search, optimisation and machine learning* (Addison-Wesley, 1989).

[15] D.Andre, J.Koza, *Advances in genetic programming 2* (MIT Press, 1996).

[16] E.Gómez-Ramírez, A.Poznyak, A.González Yunes, M.Avila-Alvarez, Adaptive Architecture of polynomial artificial neural networks to forecast non-linear time series, *CEC99 Special Session on Time Series Prediction*, Washington D.C. USA, 1999.

[17] A.M. Fraser, H.L.Swinney, Independent coordinates for strange attractors from mutual information , *Phys. Rev. A* **33** 1134-1140 (1986).

[18] W.H.Press, S.A.Teukolsky, W.T.Vetterling, B.P.Flannery, *Numerical Recipes in C* (Cambrige University Press, 1992).

[19] W.Liebert, H.G.Schuster Proper choice of the time delay for the analysis of chaotic time series, *Phys. Lett.* **A142** 107-111 (1989).

[20] F.Takens, Detecting strange attractors in turbulence, in *Dynamical Systems and Turbulence (Warwick 1980)*, pp 366-381, Lecture Notes in Mathematics **898**, (Springer Verlag, 1980).

[21] A.S.Weigend *et al.*, Backpropagation, weight elimination and time series prediction, in *Connectionnist Models, Proceedings of the 1990 Summer School* (Morgan Kaufmann, 1990).