

# Blind source separation and analysis of multispectral astronomical images

D. Nuzillard<sup>1</sup> and A. Bijaoui<sup>2</sup>

<sup>1</sup> Laboratoire d'Automatique et de Microélectronique, Université de Reims Champagne-Ardenne, Moulin de la Housse, BP. 1039, F-51687 Reims Cedex 2, France  
e-mail: Danielle.Nuzillard@univ-reims.fr

<sup>2</sup> Dpt. CERGA, UMR CNRS 6527, Observatoire de la Côte d'Azur, BP. 4229, F-06304 Nice Cedex 4, France  
e-mail: Bijaoui@obs-nice.fr

Received May 26; accepted August 31, 2000

**Abstract.** Celestial sources are studied by means of multispectral analysis, for which a pixel value can be considered as a mixture of different sources. We show that the application of blind source separation methods can display interesting features, which can help astronomers to improve their description of celestial sources. Experiments were carried out on Hubble Space Telescope images of the galaxy 3C 120. Different separation methods were applied, either based on the source correlation, or on non-Gaussian probability density functions. Each obtained source set was quantified by the mutual information between the sources. Different algorithms give quite similar decompositions which mainly correspond to real phenomena.

**Key words:** methods: data analysis — methods: statistical — techniques: image processing — galaxies: 3C 120

## 1. Multispectral analysis of astronomical images

The spectral distribution of celestial sources carries essential information on the physical processes which take place in these objects. Multispectral images are currently produced using CCD detectors and color filters which delimit well-defined spectral bands. From the flux measurements for each filter, astrophysicists compute color indexes from which they deduce the parameter values related to stars or non-stellar sources.

The color indexes allow one to classify the pixels. This is enough if the pixel intensity comes from only one source. But often pixel classification suffers from the drawback of the pixel being in fact a mixture: each pixel value results from the contribution of different objects. If we admit that

this mixing is linear, each image  $X_i$ , a mixture of physically independent sources  $S_j$ , may be written:

$$X_i = \sum a_{ij} S_j + N_i \quad (1)$$

where the matrix  $A = [a_{ij}]$  is called the mixing matrix.  $N_i$  is the noise of image  $X_i$ . This model is called *the cocktail party* in the framework of communication theory (Hyvärinen 1999). Solving Eq. (1), in which  $A$  and  $S$  are the unknowns, is the goal of all BSS methods.

The spectral and spatial energy distributions are determined by the transfer radiation equation. Non-linearities result from this transfer. Consequently a set of colored sky images cannot be exactly written as a weighted linear combination of independent sources. But since Eq. (1), being linear, is far more tractable, it is important to examine the ability of Blind Source Separation (BSS) methods to display interesting features, which could help astrophysicists to improve their description of celestial sources.

The Karhunen-Loève (KL) expansion constitutes the oldest approach to decorrelate signals. The cross-correlation matrix of the images is first computed and a singular value decomposition is carried out. The sources are computed as the weighted means of the images, the weights being the eigenvector components. This approach was applied many times for astrophysical problems: data compression (Bijaoui 1974; Pelat 1974), identification of variable phenomena (Bijaoui & Doazan 1979), visualization of a large set of images, computation of energy spectrum (Vogeley & Szalay 1996; Tegmak et al. 1997) and classification (Connolly et al. 1995).

Let us consider the probability density function (PDF) of the observed images. In the case of Gaussian PDFs uncorrelated pixel values are equivalent to independent ones, and the KL expansion is sufficient to separate sources. But generally, the PDFs are not Gaussian and a set of blind source separation methods was developed in order to take account of higher order statistics leading to what has been

**Table 1.** The characteristics of the studied images

Filter	Color	Exposure (s)
F547M	V1	$2 \times 1100$
F555W	V	$2 \times 1000$
F675W	R	$2 \times 1100$
F814W	I	$2 \times 1100$

called Independent Component Analysis (ICA). An alternative approach, based on temporal (or spatial) source correlations has also been developed (Belouchrani et al. 1997).

First we present the data set on which we have carried out the experiments. We give then a short overview of BSS methods. These applied methods provide slightly different separations. We describe some criteria to evaluate their quality. In particular, mutual information is considered. These results are discussed in the light of the physical nature of the celestial object. We conclude on the perspectives opened up by these new tools.

## 2. The data set

### 2.1. The observations

The tools we describe further were extensively tested on many sets of data in the framework of signal processing and communication. Their ability to give a pertinent solution to Eq. (1) was studied taking into account various conditions (Cardoso et al. 1999; Pajunen & Karhunen 2000). So, there is no point to give in this paper experiments on simulated data.

We chose to test BSS tools on HST images of the Seyfert radiogalaxy 3C 120. This object displays a one side radio jet, which has been optically identified (Lelièvre et al. 1994). The best HST observations were made with the *Wide Field Planetary Camera 2* (WFPC2) through the filters F547M, F555W, F675W, and F814W by J. Westphal, on July 25th 1995. The exposure characteristics are given in Table 1.

The transmission profiles of the filters are drawn in Fig. 1. The spectrum obtained by Oke et al. (1980) by a spectrophotometer is also plotted. F547M covers two strong [OIII] lines. F555W recovers F547M, the other part includes only the continuum. F675W covers  $H_\alpha$  and F814W [OI] and [SIII] lines. Linear combinations of the observed images can display specific physical phenomena. We can determine their best combinations by applying Spectral Energy Distribution (SED) methods (Bolzonella et al. 2000). With SED the photometric information is transformed into physical parameters using spectral models. But for 3C 120 we have a mixing between stellar distributions, non-thermal emission from the nucleus and light from the ionized regions, and SED does not take into account this mixing. We expect from BSS to disentangle these various phenomena.

### 2.2. The data preparation

The central WFPC2 observations correspond to images of  $800 \times 800$  pixels. The pixel size is  $0.0455'' \times 0.0455''$ . We have extracted the central part of  $256 \times 256$  pixels, corresponding to a region of  $11.6'' \times 11.6''$ .

The observations were made with a CCD receiver. The detector noise is low, and the photon noise is dominant in the galaxy region. As the noise is not stationary Gaussian, a generalized Anscombe transform (Murtagh et al. 1995) allowed us to stabilize its variance. A pixel value  $v$  is transformed by the relation:

$$t = \frac{2}{\alpha} \sqrt{\alpha v + \frac{3}{8}\alpha^2 + \sigma^2} - \alpha g \quad (2)$$

where  $v = \gamma + \alpha n$ ,  $\gamma$  is a Gaussian random variable, corresponding to the read-out noise of mean  $g$  and variance  $\sigma^2$ , while  $n$  is the number of photoevents in the pixel,  $\alpha$  is then the gain of the CCD.  $\gamma$ ,  $g$  and  $\alpha$  values are given in the HST/WFPC2 handbook.

For each resulting image, the background value is estimated and subtracted. Then the images (Fig. 2) can be considered clean enough to be processed by BSS algorithms.

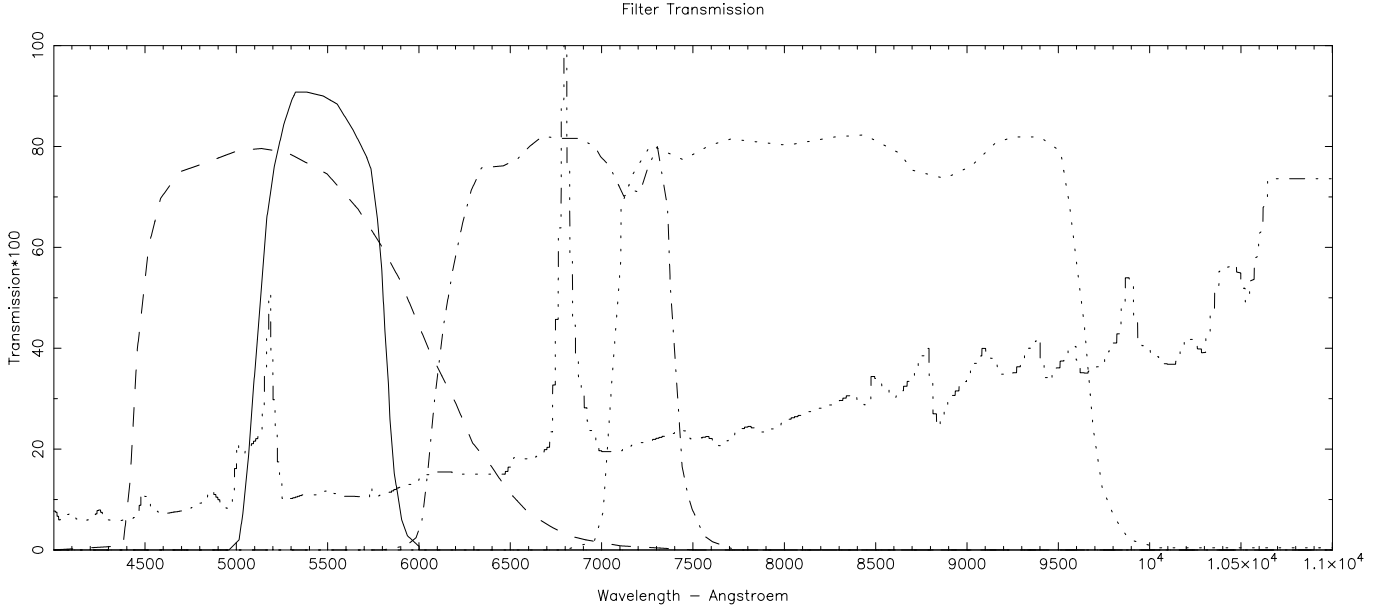
We note that the images were not linearly transformed. Roughly speaking we took their square roots (Eq. 2). By this transformation we destroyed the hypothetical linearity between the images and the sources as written in Eq. (1). We can imagine that this Eq. (1) constitutes only a working hypothesis. With the transformation given by Eq. (2), the dynamical amplitude is then reduced and the nucleus plays a smaller part in the computations. We do not search for a perfect photometric model, but we want to exhibit coherent image structures.

## 3. Introduction to BSS

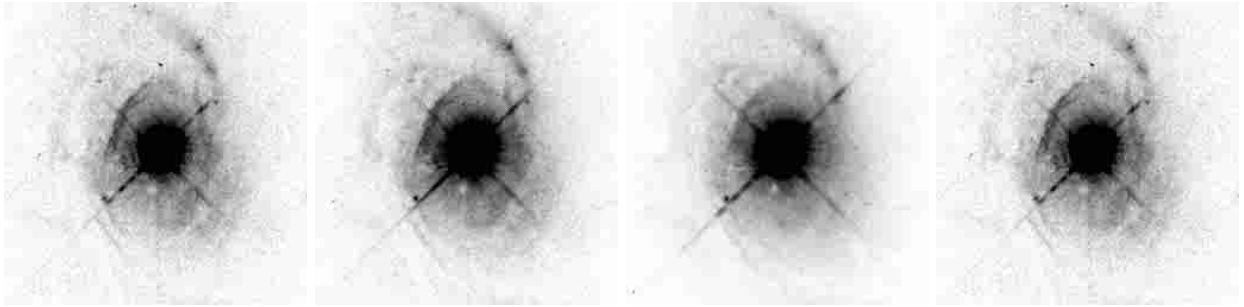
Independent component analysis (ICA) is a statistical signal processing technique that decomposes a set of  $m$  observed random or deterministic data into  $n$  independent unobserved source signals where  $m \geq n$  and estimates the mixing matrix  $A$ . The simplest model is to consider observed mixed signals  $X$  as a linear combination of unobserved source signals  $S$ , based on the mixing matrix  $A$  with added sensor noise  $N$ , as shown in Eq. (1). ICA applications can be BSS, feature extraction or blind deconvolution. Current solution techniques are based on a white noise model. From the generalized Anscombe transform, this hypothesis is satisfied in our experiments.

In the BSS context, a full identification of the mixture matrix  $A$  is impossible because the exchange of a fixed scalar factor between a given source image and the corresponding column of  $A$  does not affect the observations. For each observed image  $X_i$ :

$$X_i = \sum_j \frac{a_{ij}}{\alpha_j} \alpha_j S_j + N_i \quad (3)$$



**Fig. 1.** The four filters used: F547M (full line), F555W (dashed line), F675W (dot-dashed line) and F814W (dotted line). The galaxy spectrum in the nuclear part (Oke et al. 1980) is superimposed on these profiles (dashed-dotted line)



**Fig. 2.** HST/WFPC2 images of the radiogalaxy 3C 120 respectively obtained with the filters F547M, F555W, F675W and F814W. The images were processed in order to stabilize the noise variance

where  $\alpha_j$  is an arbitrary factor. To cater for this indeterminacy, the sources are constrained to have unit variance.  $A$  is consequently normalized. The correlation matrices are defined by:

$$R(\tau) = E[X(t + \tau) \cdot X^*(t)] = A \cdot R_S(\tau) \cdot A^H \quad (4)$$

where  $H$  denotes the complex conjugate transpose of the matrix and  $R_S(\tau)$  is the correlation matrix of the source. For unit variance,

$$R(0) = R_S(0) = I = A \cdot A^H. \quad (5)$$

The mixture matrix  $A$  can be estimated up to permutation and phase shifts. Some considerations developed below must be added to improve the knowledge of  $A$  and  $S$ .

*The Karhunen-Loève expansion.* For many decades the KL expansion has been applied for extracting the main information from a set of celestial images (see Murtagh & Heck 1987 for references). From the correlation matrix

of the observed signals, the eigenvalues are evaluated in decreasing order, and the most significant ones are kept. From the eigenvectors, orthonormal sources are obtained. The KL expansion allows us to whiten the images and it is considered as the first step of BSS. The resulting demixed images show clearly that non-independence still exists (Fig. 3).

The KL expansion is not the only transformation which leads to a diagonal correlation matrix. Any rotation of the resulting sources keeps this property, but the KL expansion is the one which maximizes the energy concentration.

*Orthogonality and rotations.* If the image PDFs are Gaussian ones, uncorrelated data mean independent ones, and then the KL expansion cannot be improved upon the basis of the PDFs. Generally the distributions are not Gaussian, and a null covariance does not mean that the sources are independent. In the space defined by the sources, any rotation keeps the orthogonality,

and the norms. But the energy related to each source is spread. We can search for what rotation an independence criterion is optimized.

The optimal rotation results from  $n(n-1)/2$  elementary rotations of angle  $\theta_{ij}$  in the plane defined by sources  $i$  and  $j$ ,  $n$  is the number of sources. This decomposition allows one to design algorithms to optimize an independence criterion.

*The independence criteria.* Modern ICA work has been inspired by neural networks (Jutten & Héroult 1991). An historical review of this new field can be found in Jutten & Taleb (2000). The ICA concept was defined by Comon (1994). The link between neural network and entropy was proposed by Bell & Sejnowsky (1995), while Amari & Cichocki (1998) introduced an algorithm based on the natural gradient (Amari 1998).

In an environment where no adaptation is needed, one prefers because of computation time and convergence properties to use batch optimization algorithms, which act on the whole set of data without interaction. Second order algorithms are based on the hypothesis of temporally (or spatially) correlated sources and allow efficient second order separation. The cross-correlation between shifted sources taken two by two is decreased, while the source correlation is increased. Other batch computations minimize or maximize contrast functions (Comon 1994) based on higher order cumulants like in JADE (Cardoso & Souloumiac 1993). They allow signals with non-Gaussian PDFs to be separated. Stochastic gradient methods are implemented in neural networks (Hyvärinen & Oja 1997). These methods will be developed below.

#### 4. The BSS experimented tools

*SOBI and its derived versions.* SOBI (Second Order Blind Identification, Belouchrani et al. 1997) is an efficient second order algorithm. It depends on the number of spatial shifts  $p$  of sources with themselves and their values  $s_i$ ,  $i \in (1, p)$ . After the data whitening, a set of  $p$  covariance matrices is computed. Cross-correlation terms are minimized, thus diagonal terms are maximized. A joint diagonalization criterion of several  $p$  covariance matrices improves its robustness (Cardoso & Souloumiac 1996). We adapted the algorithm from the temporal field to the 2D one.

In its original version (we call it SOBI1), we compute the cross-correlation matrices in the direct space: a vector signal is constructed by concatenation of the pixel rows.

For analyzing Nuclear Magnetic Resonance (NMR) spectrograms, Nuzillard (1999) modified SOBI by computing the cross-correlations of the Fourier transforms, which are easily estimated in the direct space. This can be viewed

as an alternative correlation choice. SOBI1 takes into account the correlation at short distances, while the correlations at short frequency distances play the main role in SOBI2. NMR spectrograms display narrow peaks (or lines), so that the correlation rapidly decreases, and therefore SOBI2 provides better separation. For our current data this argument is not so clear. The cross-correlation between two vectors  $A$  and  $B$  is expressed as:

SOBI1 In the direct space:

$$R_{AB}(\xi) = \frac{1}{L} \sum_k a_k \cdot b_{k-\xi}^* \quad (6)$$

where  $\xi$  is the spatial shift and  $L$  the number of samples,  $a_k$  and  $b_k$  are elements of  $A$  and  $B$ , the sampling step being 1.

SOBI2 In the Fourier space:

$$R_{AB}(n) = \sum_k a_k \cdot b_k^* \cdot e^{-2i\pi(\frac{kn}{L})} \quad (7)$$

where  $n$  is the spatial frequency shift. More precisely  $n$  corresponds to a shift of a cross-correlation computed on the Fourier transform of the vector  $A$  and  $B$ .

SOBI3 and SOBI4 are respectively SOBI2 and SOBI1 adaptations to 2D images. The algorithms take into account the correlation matrices between two images in different directions: rows, columns, diagonal, etc.

SOBI3 In the Fourier transform space:

$$R_{AB}(n, m) = \sum_k \sum_l a_{kl} \cdot b_{kl}^* \cdot e^{-2i\pi(\frac{nk}{L_1} + \frac{ml}{L_2})} \quad (8)$$

where  $n$  and  $m$  are the spatial frequency shifts,  $L_1$  and  $L_2$  the row and column numbers and  $a_{kl}$  and  $b_{kl}$  pixel values of images  $A$  and  $B$ .

SOBI4 In the direct space:

$$R_{AB}(\xi, \eta) = \frac{1}{L_1 L_2} \sum_k \sum_l a_{kl} \cdot b_{k-\xi, l-\eta}^* \quad (9)$$

where  $\xi$  and  $\eta$  are also the spatial shifts.

*JADE and FastICA algorithms.* A contrast is a mapping from the set of PDFs  $p_x$  to the real set  $\mathbb{R}$ ,  $x$  being a random vector of length  $N$  with values in  $\mathbb{R}^N$ . A contrast satisfies the following requirements:

- $f(Pp_x) = f(p_x)$ , where  $P$  is any permutation matrix;
- $f(P_\Lambda x) = f(p_x)$ ,  $\Lambda$  is diagonal and can be inverted;
- $f(p_{Ax}) \leq f(p_x)$ , if  $x$  has independent components, where  $A$  can be inverted.

In JADE (Joint Approximate Diagonalization of Eigen-matrices, Cardoso & Souloumiac 1993), the statistical independence of the sources is obtained through the joint maximization of the fourth order cumulants since these fourth order terms behave as contrast functions (Comon 1994).

In FastICA (Hyvärinen & Oja 1997) non-Gaussianity is measured by a fixed-point algorithm using an approximation of negentropy through a neural network.

FastICA was first introduced using the kurtosis as a contrast function; then it was extended for general contrast functions such as:

$$J_G(y) = |Ey\{G(y)\} - E\nu\{G(\nu)\}|^p \quad (10)$$

where  $\nu$  corresponds to the Gaussian variable which has the same mean and the same variance as  $y$ . Generally  $p$  is equal to 2.  $G(y)$  is used in its derivative form,  $g(y)$ , in the fixed point algorithm. The couple  $(G(y), g(y))$  can be chosen among the following functions:

- $G(y) = y^4, g(y) = y^3$ ;
- $G(y) = \log(\cosh(ay)), g(y) = \tanh(ay)$ ;
- $G(y) = e^{-\frac{ay^2}{2}}, g(y) = ye^{-\frac{ay^2}{2}}$ .

The algorithm works in two different ways: a symmetric solution is used for which the sources are computed simultaneously, or a deflation one for which the sources are extracted successively.

## 5. Comparison methods

Many experiments were carried out with these algorithms. For each one we got mixing and demixing matrices and image sources. Even though a visual appreciation of the resulting sources was informative, a quantitative quality measurement was needed.

*The source filters.* The spectrum at pixel location  $(k, l)$  is called  $S(k, l, \lambda)$ ,  $\lambda$  being the wavelength. Each image  $X_i(k, l)$  is observed with a filter of transmission profile  $T_i(\lambda)$ . We have:

$$X_i(k, l) = \int T_i(\lambda)S(k, l, \lambda)d\lambda. \quad (11)$$

Let us call  $c(i, j)$  the demixing coefficient of source  $j$  for image  $i$ , source  $S_j(k, l)$  is written as:

$$S_j(k, l) = \sum_i c(i, j)X_i(k, l). \quad (12)$$

Combining Eqs. (11) and (12):

$$S_j(k, l) = \int \left( \sum_i c(i, j)T_i(\lambda) \right) S(k, l, \lambda)d\lambda. \quad (13)$$

We can write:

$$S_j(k, l) = \int U_j(\lambda)S(k, l, \lambda)d\lambda \quad (14)$$

with:

$$U_j(\lambda) = \sum_i c(i, j)T_i(\lambda). \quad (15)$$

The sources can be considered as the observed images through filters  $U_j(\lambda)$ , which are called the source filters. For each BSS we display the set of its source filters.

*The energy break-down.* From each BSS, the image energy coming from a given source was evaluated. Then we can compute the energy related to each source from all the images. This criterion allows us to classify the sources by decreasing energy.

*Source visualization.* The source images are displayed in order to optimize the contrast. That allows us to compare the different sources in the best contrast conditions. This visual comparison was essential to select the best identification, but it is too qualitative.

## 6. The Source Mutual Information (SMI)

After a set of tests we noticed that the previous tools were not sufficient to select the best BSS. We introduced another test based on the mutual information between the sources.

*The mutual information.* Let us consider two sources  $S_1$  and  $S_2$ . Their entropies are defined as:

$$E(S_1) = - \sum_n p_{S_1}(n) \log_2 p_{S_1}(n) \quad (16)$$

$$E(S_2) = - \sum_n p_{S_2}(n) \log_2 p_{S_2}(n) \quad (17)$$

where  $p_{S_1}(n)$  and  $p_{S_2}(n)$  are the probabilities of the pixel value  $n$  respectively in sources  $S_1$  and  $S_2$ .

The entropy of the couple  $S_1$  and  $S_2$  is:

$$E(S_1, S_2) = - \sum_{n_1, n_2} p(n_1, n_2) \log_2 p(n_1, n_2) \quad (18)$$

where  $p(n_1, n_2)$  is the joint probability of pixel value  $n_1$  for  $S_1$  and  $n_2$  for  $S_2$ . If the sources are independent:

$$E(S_1, S_2) = E(S_1) + E(S_2). \quad (19)$$

The quantity:

$$I(S_1, S_2) = E(S_1) + E(S_2) - E(S_1, S_2) \quad (20)$$

is called the mutual information between  $S_1$  and  $S_2$ . It is the information on  $S_1$  knowing  $S_2$  and inversely (Rényi 1966). From Eqs. (16–18) and (20)  $I(S_1, S_2)$  can be written as:

$$I(S_1, S_2) = \sum_{n_1, n_2} p(n_1, n_2) \log_2 \frac{p(n_1, n_2)}{p_{S_1}(n_1)p_{S_2}(n_2)} \quad (21)$$

which is the Kullback-Leibler divergence between the joint probability and the probability obtained from the marginal probabilities (Comon 1994). If the sources are independent, the joint probability is the product of the marginal ones and this divergence is equal to 0.

Then, the mutual information of a set of  $l$  sources is defined as a generalization of Eq. (21). We have:

$$I(S_1, \dots, S_l) = \sum_{n_1, \dots, n_l} p(n_1, \dots, n_l) \log_2 \frac{p(n_1, \dots, n_l)}{\prod_{i=1, l} p_{S_i}(n_i)}. \quad (22)$$

*The observed SMI.* The observed SMI can be derived from Eq. (22). We have to extract from experimental data an available estimation of the probability  $p(n_1, n_2, \dots, n_l)$ . This probability is obtained from the number of pixels having a value  $n_1$  in source  $S_1$ ,  $n_2$  in source  $S_2$ ,  $\dots$ ,  $n_l$  in source  $S_l$ . Then we distribute the pixels among  $K^l$  cells, where  $K$  corresponds to the number of levels per source.

As  $K^l$  increases exponentially with the number of sources, the mean number of pixels per cell decreases rapidly with  $l$ , and  $E(S_1, S_2, \dots, S_l)$  is badly estimated. In order to avoid this difficulty, we notice that we do not need to compute the exact value  $I(S_1, \dots, S_l)$ , but only to say for which mixing matrix  $A$  or its inverse  $B = A^{-1}$ , the mutual information is minimum. We can write the approximation:

$$S = BX. \quad (23)$$

We obtain the sources from the images from a linear transformation, since the entropy of the source set  $E(S_1, \dots, S_n)$  is the entropy of the image set  $E(X_1, \dots, X_n)$  plus the logarithm of the Jacobian of the transformation:

$$E(S_1, \dots, S_l) = E(X_1, \dots, X_n) + \log_2 |\det B|. \quad (24)$$

We obtain:

$$I(S_1, \dots, S_l) = \sum_l E(S_l) - E(X_1, \dots, X_n) - \log_2 |\det B|. \quad (25)$$

Equation (23), and consequently Eq. (25) are true only as a limit for an increasing number of pixels. This number must be such that it allows one to get an experimental PDF with a very small sampling interval. Then it is sufficient to minimize the function (Comon 1994):

$$C = \sum_l E(S_l) - \log_2 |\det B|. \quad (26)$$

We applied Eq. (26) and the optimal set of sources was not the one retained by visual inspection. The drawback in Eq. (26) is that we never have enough pixels to validate it. We preferred to define SMI by summing the pairwise mutual information values, as explained below.

*The SMI algorithm.* The entropy depends on the coding step between two levels:

- If the step is too small the number of pixels per cell is too low and then the estimation is not available;
- If the step is too large, the entropy is too small, the PDF is smoothed and it is not sensitive to non-Gaussian features.

Then the SMI determination is achieved in four steps:

1. For each source  $i$  we determine the mean value  $m_i$  and the standard deviation  $\sigma_i$  after a  $3\sigma_i$  clipping. In this algorithm we compute iteratively these parameters and we reject the values outside the interval  $[m_i - 3\sigma_i, m_i + 3\sigma_i]$ . After a few iterations (4 to 5) the

**Table 2.** Source mutual information of the selected decompositions

Method	SMI
KL	0.771
SOBI2-16	0.581
SOBI4-8	0.805
FastICA	0.584

algorithm converges. In the case of a true Gaussian distribution, the obtained mean is correct, while the  $\sigma_i$  bias is of the order of 2%. If we are not dealing with a Gaussian distribution, these parameters define a Gaussian kernel of the PDF, and we will measure more values outside the interval  $[m_i - 3\sigma_i, m_i + 3\sigma_i]$  than for a Gaussian PDF;

2. The histogram  $H_i(k)$  of source  $i$  is determined with a cell size equal to this deviation  $\sigma_i$ . We evaluate the entropy of the source  $E_i$  by:

$$E_i = - \sum_k \frac{H_i(k)}{N} \log_2 \frac{H_i(k)}{N} \quad (27)$$

where  $N$  is the number of pixels;

3. We determine then the mutual histogram  $H_{ij}$  with the same cell size, and we compute the resulting entropy  $E_{ij}$ . The mutual information between  $i$  and  $j$  is equal to:

$$I_{ij} = E_i + E_j - E_{ij}. \quad (28)$$

This mutual information is independent of the cell size, for values smaller than  $\sigma_i$ , for a large number of pixels. A faint bias is introduced for a small number of pixels per cell. It is a possible compromise to choose a cell size equal to  $\sigma_i$ ;

4. We quantify the quality of each separation by the sum:

$$I = \sum_{i < j} I_{ij}. \quad (29)$$

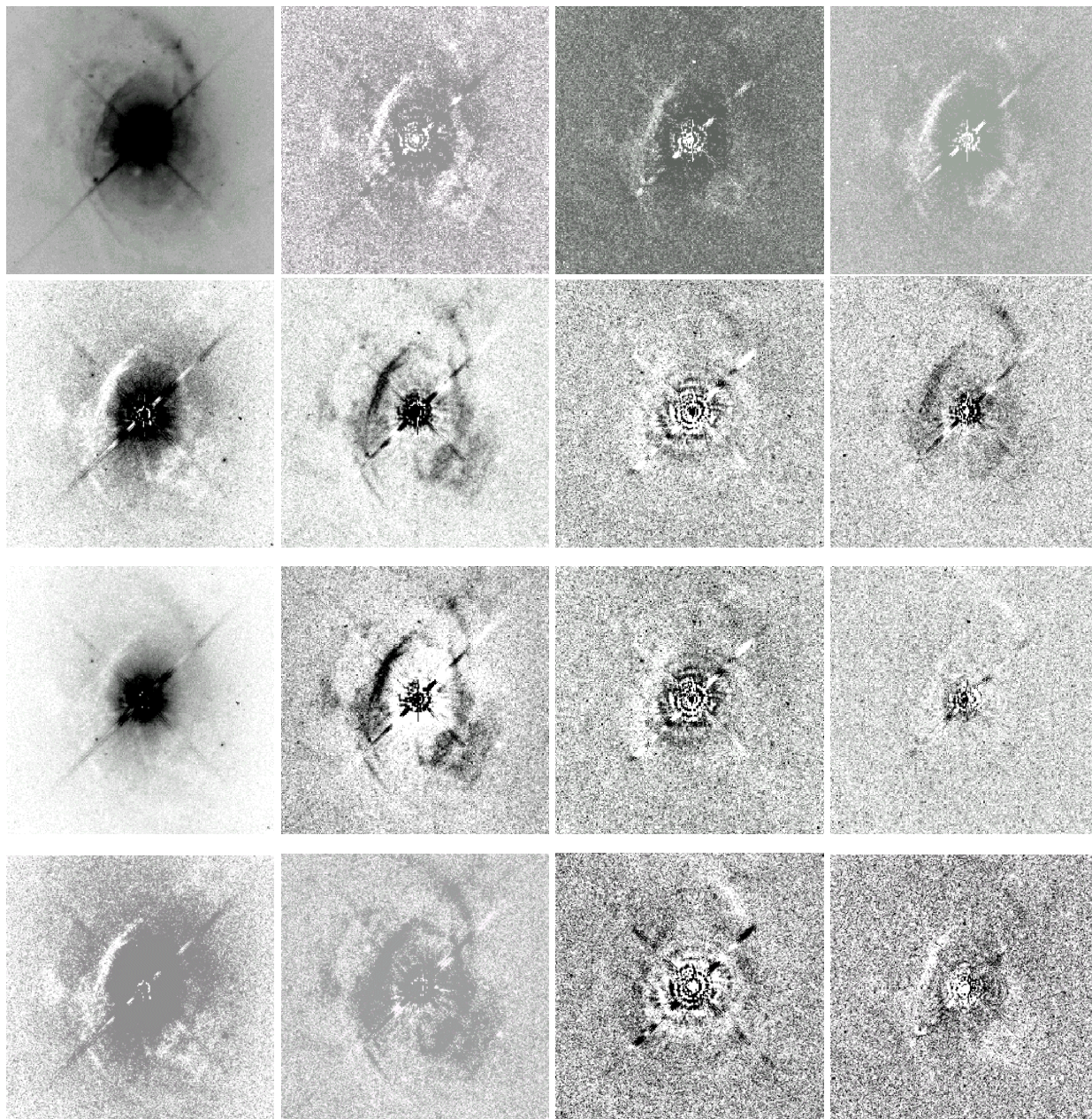
We note that  $I$  only takes into account the pixel PDF. Therefore it favors the algorithms based on PDF, such as KL, JADE and FastICA, and penalizes SOBI.

## 7. The results

We selected four decompositions for our experiments:

- KL: The basic KL expansion;
- SOBI2-16: SOBI in the 1D Fourier space, with 16 correlation matrices;
- SOBI4-8: SOBI in the 2D direct space, with 8 correlation matrices;
- FastICA: FastICA with the deflation algorithm with  $g(y) = \tanh(y)$  associated function.

Table 2 contains SMI values of each selected BSS. FastICA and SOBI2-16 give a similar SMI, significantly lower than KL and SOBI4-8 ones. In Fig. 3 the resulting sources of

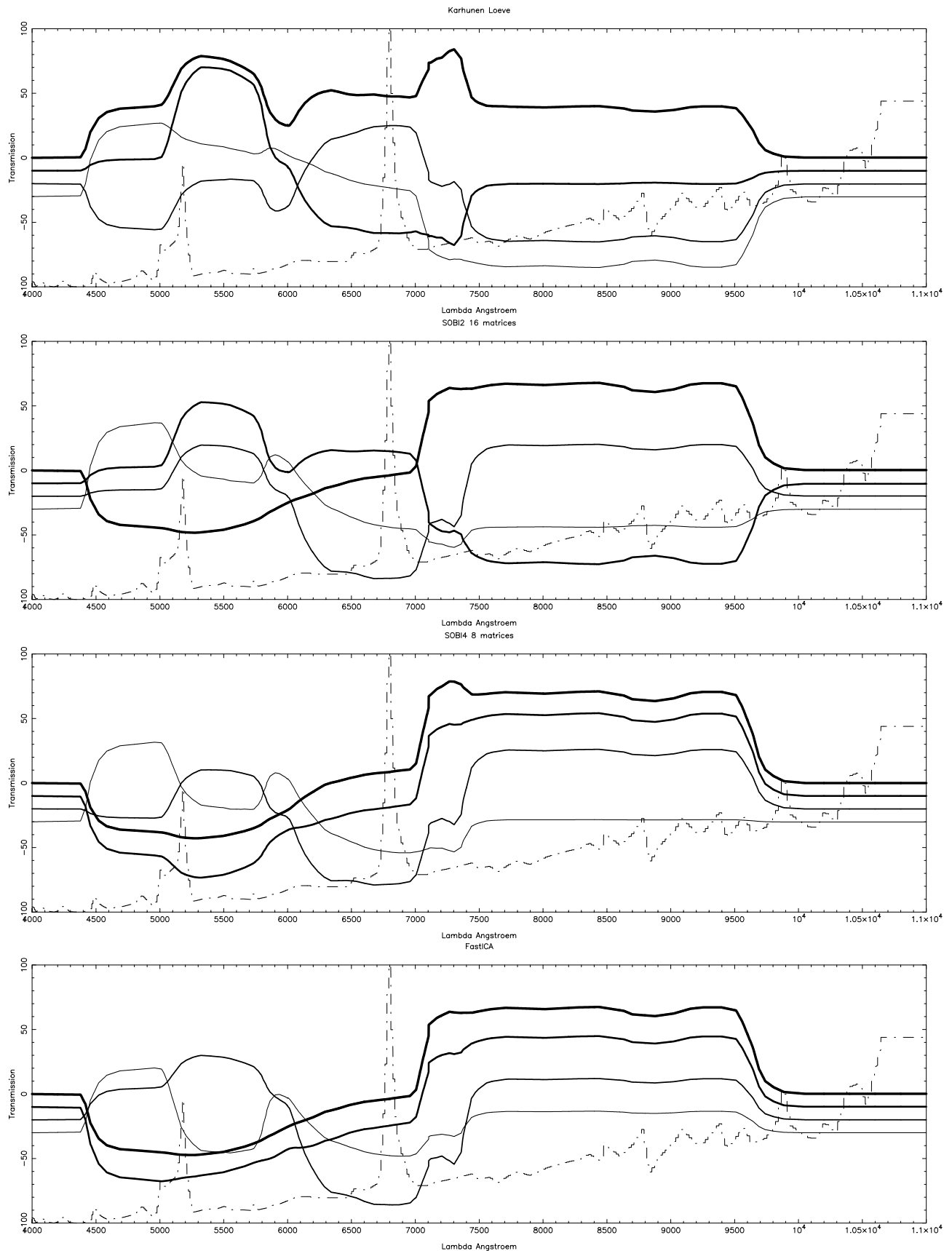


**Fig. 3.** The sources obtained from the four selected BSSs on the 3C 120 HST images. From top to bottom respectively the Karhunen-Loève expansion, SOBI2 algorithm with 16 cross-correlation matrices, SOBI4 algorithm with 8 cross-correlation matrices and FastICA algorithm with a deflation algorithm and  $g(y) = \tanh(y)$

four selected decompositions are shown. The last three decompositions provide similar sources, different from the KL expansion. In Fig. 4 the corresponding source filters are drawn. We find some similarities between the resulting filters, even if the plots are sometimes reversed due to the sign of the demixing coefficients. We note that the sources and the mixture matrix are defined with scalars  $\alpha_j$ . Even if the source variance is equal to 1, we have an uncertainty about the signs. The diagrams in Fig. 4 are built from the combination of the filter transmissions with the demixing coefficients. It is obvious that it is not possible

to obtain source filters which extract well defined spectral regions only by combinations of the original filters because we have not all the required monochromatic images. BSSs carried out the best linear image combinations, namely those having the maximum independence.

In Table 3 we indicated the percentage of energy contained in each source for each BSS. We remark that this break-down is different from one decomposition to another. The KL expansion is, of course, the decomposition that leads to the greatest energy concentration.



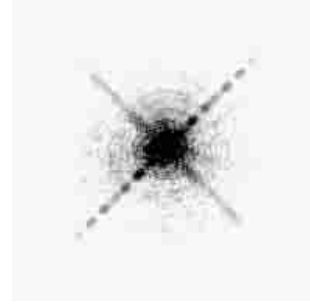
**Fig. 4.** The source filters obtained from the four selected BSSs on the 3C 120 HST images. From top to bottom respectively KL, SOBI2-16, SOBI4-8 and FastICA. The line thickness is decreasing with the source order, the thickest one corresponds to source 1. The lines are also shifted  $-10$  units on ordinate, from a source to the following one, in order to clarify the diagrams

**Table 3.** Source energy of the selected decompositions

Method	1	2	3	4
KL	0.942	0.034	0.015	0.008
SOBI2-16	0.386	0.449	0.030	0.134
SOBI4-8	0.779	0.177	0.029	0.015
FastICA	0.295	0.626	0.038	0.041

*Interpretation of the source images.* If we exclude the KL expansion, the source decomposition from the three other selected BSSs are similar, as commented on below. Each source displays interesting features:

- Source 1: This source displays mainly the central galaxy region. For SOBI4-8 decomposition, we got the maximum of energy, but for the other BSSs the energy was seriously reduced. We note a trace of structures associated with source 2. The source filters look similar. They correspond to a difference of the mean flux after 7000 Å and between 4500 to 6000 Å. The  $H_\alpha$  region is excluded, but not the [OIII] one, explaining the trace of source 2;
- Source 2: This is the most interesting source for physical insight. On the source filters we can see that the continuum after 7000 Å was extracted. The other part is at the opposite end, in a region containing [OIII] lines. For SOBI2-16,  $H_\alpha$  line plays a faint role, while it has no role for the other BSSs. This source corresponds to the ionized [OIII] regions surrounding the galaxy. It corresponds also to a large part of the energy and for two decompositions the maximum value is obtained. If we compare the source decomposition obtained with KL to the other ones, we note that for the BSSs an important part of the energy was clearly transferred from source 1 to source 2. The first KL source roughly corresponds to a mean image, and the energy from ionized regions was simply averaged. The energy from KL source 2 only corresponds to the variations from one filter to the other. For the other BSSs, the rotations allow one to partially recover the energy from the mean;
- Source 3: This corresponds clearly to a set of rings around the nucleus. Its energy is relatively very constant from one BSS to another. These rings could be seen in the KL sources, but the other BSSs display a cleaner result. Taking into account their size, these rings do not correspond to an Airy pattern associated to a simple mirror of HST size. For the three BSSs [OIII] and  $H_\alpha$  lines play the major part in the source filters, in an opposite manner. We compared the source 3 images to the point spread function (PSF) of WFPC2 at 675 nm obtained with the Tiny Tim program (Krist & Hook 2000) (Fig. 5). Even if the contrasts were not the same, the ring patterns correspond to that function. The main energy in the lines comes from  $H_\alpha$ , and this is the reason why the structure looks like its

**Fig. 5.** The point spread function at the wavelength 675 nm of the WFPC2 camera

corresponding PSF at the central position. It is amazing that BSS allows us to display a real physical phenomenon, which is due to the fact that a large part of the energy in image F675W is emitted in a spectral line in a small region, which is not resolved with the WFPC2. For this line we have the image of its PSF, which displays rings due to the central occultation of the telescope. Due to the large wavelength range, the PSF corresponding to the spectral continuum region does not display these rings. As we can see in Fig. 4, the  $H_\alpha$  line is excluded from source 1. BSSs played their role by extracting the  $H_\alpha$  PSF as a specific source;

- Source 4: This is the less informative source. For SOBI2 it looks like a smooth version of source 2. The main energy comes from the region between 4500 and 5200 Å. The influence of the emission lines is faint, but not equal to 0.

*Discussion.* BSS algorithms extracted three sources which seemed to correspond to three independent physical components. Even if the sources are orthogonal, some traces of the structures which are well-defined in each source are seen in the other sources. A physical mixture can still exist, but it is faint compared to the one in the original images. The source 4 may be interpreted as the residue of the separation.

We can note that a second order blind identification algorithm, based on correlations in a large region around a pixel, carries out quite a similar decomposition to a blind identification related to local high order statistics. That fact brings some confidence to the obtained decomposition.

This statistical analysis tool allows one to get a simple model of the galaxy 3C 120 with two components:

- *A very bright nucleus which is not resolved in these exposures.* The central pixels are saturated, a linear artifact due to the CCD blooming exists in each image. In Fig. 5 we can notice the large extension of the PSF at a given wavelength (6750 Å). This size is similar to the extension of source 1. This is because this source can be only due to an unresolved nucleus, like the source 3 as we discussed above. Sources 1 and 3 correspond to the

same physical component. Obviously we can see other features in the image sources, BSS does not provide a perfect physical model, but a decomposition which is optimal for a given statistical criterion. There is no reason that the physical reality corresponds exactly to this criterion;

- *A gaseous region surrounding the galaxy as shown in source 2.* Stellar objects can be associated with this component, but the main structures are described by the ionized [OIII] regions.

This resulting description is not new, but it cannot be seen directly in the observed images or in the KL sources. BSS failed to display the optical counterpart of the radio jet as a specific source. On the one hand, we did not process enough images allowing us to find more components. On the other hand, the radio jet does not necessarily correspond to a source with a pixel distribution independent from the ones of the other image components. Images obtained with more filters are needed to improve the decomposition.

## 8. Conclusion

In the present paper we examined the multispectral analysis of astronomical images as a particular BSS case. Even if the *cocktail party* model is not realistic for processing celestial images, we showed that the application of different BSS algorithms gives quite similar decompositions.

From the physical point of view, the resulting sources correspond mainly to real phenomena. These separations can provide useful information for astrophysicists.

We experimented with different BSSs. Some further improvements in the methods may be possible. For our application the main difficulties do not depend on these technical points but on the basic Eq. (1). The model required:

- A linear combination of physical phenomena;
- Stationary mixing;
- Stationary white noise.

Although real multispectral observations do not exactly satisfy these conditions, BSS can still be considered as an interesting exploratory tool, which can suggest a new analysis approach to the user, as we have seen above. Its use is well adapted to data mining.

We plan to apply these tools to a set of observations of typical astronomical sources, with images taken through many filters, and to process with BSS tools the several hundred of monochromatic images provided by astronomical spectroimagers.

*Acknowledgements.* The authors thank STScI for the use of the HST archive, Dr. J.F. Cardoso for the use of the original SOBI and JADE programs, Dr. A. Hyvärinen for the use of FastICA program, Dr. J.M. Nuzillard for his help in SOBI adaptations, and Dr. G. Lelièvre and G. Wlérick for comments on 3C 120 results. The authors thank also an anonymous referee for his comments that have contributed to the improvement of the final version.

## References

- Amari S., 1998, *Neural Comp.* 10, 251  
 Amari S., Cichocki A., 1998, *Proc. IEEE* 86, 2026  
 Bell T., Sejnowsky T., 1995, *Neural Comp.* 7, 1004  
 Belouchrani A., Abed-Meraim K., Cardoso J.F., Moulines E., 1997, *IEEE Trans. SP* 45, 434  
 Bijaoui A., 1974, *A&A* 30, 199  
 Bijaoui A., Doazan V., 1979, *A&A* 70, 285  
 Bolzonella M., Miralles J.M., Pelló R., 2000, *A&A* (submitted) astro-ph/0003380  
 Cardoso J.F., Souloumiac A., 1993, *IEE Proc.-F* 40, 362  
 Cardoso J.F., Souloumiac A., 1996, *SIAM J. Mat. Anal. Appl.* 17, 161  
 Cardoso J.F., Jutten Ch., Loubaton Ph., 1999, *Independent Component Analysis and Signal Separation*, INPG-LIS, Saint-Martin d'Hères, France  
 Comon P., 1994, *Signal Proc.* 36, 287  
 Connolly A.J., Szalay A.S., Bershadly M.A., Kinney A.L., Calzetti D., 1995, *AJ* 110, 1071  
 Hyvärinen A., 1999, *IEEE Trans. Neural Net.* 10, 626  
 Hyvärinen A., Oja E., 1997, *Neural Comp.* 9, 1483  
 Jutten C., Herault J., 1991, *Signal Proc.* 24, 1  
 Jutten C., Taleb A., 2000, *Independent Component Analysis and Blind Source Separation*, Pajunen P., Karhunen J. (eds.), University of Technology, Helsinki, p. 15  
 Krist J., Hook R., 2000, *The Tiny Tim User's Guide*, <http://www.stsci.edu/software/tinytim>, STScI  
 Lelièvre G., Wlérick G., Sebag J., Bijaoui A., 1994, *C. R. Acad. Sci. Paris II* 318, 905  
 Murtagh F., Heck A., 1987, in *Multivariate data analysis*. Reidel, Dordrecht, p. 13  
 Murtagh F., Starck J.L., Bijaoui A., 1995, *A&AS* 112, 179  
 Nuzillard D., 1999, in *GRETSI'99 Vannes*, p. 745  
 Oke J.B., Readhead A.C.S., Sargent W.L.W., 1980, *PASP* 92, 758  
 Pajunen P., Karhunen J., 2000, *Independent Component Analysis and Blind Signal Separation*, University of Technology, Helsinki  
 Pelat D., 1974, *A&A* 33, 321  
 Rényi A., 1966, *Calcul des Probabilités* 513. Dunod, Paris  
 Tegmak M., Taylor A.N., Heavens A.F., 1997, *ApJ* 480, 22  
 Vogeley M.S., Szalay A.S., 1996, *ApJ* 465, 34