Proceedings of The Institute of Acoustics

# HIGH RESOLUTION METHODS WITH NOISE CORRELATED SENSORS OUTPUTS. A RELATIVE ENTROPY APPROACH, METHOD AND RESULTS

J.P Le Cadre

(1) GERDSM Le Brusc 83140 SIX-FOURS-Les-Plages FRANCE
(2) CEPHAG SAINT-MARTIN d'HERES BP 48 CEDEX FRANCE

## INTRODUCTION

The problem of estimation of noise correlation between sensors (NCS) occurs very often in the area of spatial signal processing and specially for application of high resolution methods.

High resolution methods are, indeed, based upon separation (by linear algebra methods) of the outputs of an array into a coherent part (corresponding to point sources) and an incoherent part (related to ambient noise). Hypotheses (corresponding to an a priori information) about structure of ambient noise cross-spectral matrix (CSM) is fundamental for the use of high resolution methods.

The lack of efficient and simple methods for estimation of parameters defining NCS often leads to consider ambient noise as decorrelated (sensor to sensor). This hypothesis is not without danger and may lead to damage strongly performances of high resolution methods.

One considers, in the following, that ambient noise is the sum of all non coherent signals received by the array, ie traffic noise, surface noise, flow noise, etc...

Some methods for estimation of NCS problem have yet been proposed. Among these one may mention :

1) Use of the likelihood as a function of eigenvalues of C.S.M. of whitened outputs [4 ].

In that method one considers influence of variations of noise correlation parameters on the eigenvalues of whitened cross-spectral matrix. Unfortunately, eigenvalues of whitened CSM are not easily related to the parameters of NCS in the general case. Iterative algorithms are, then, not easily efficients. An heuristic criterion [ 4] have been proposed and gives interesting results for usual modelisation of NCS (Spherical noise, surface noise, etc...).

2) Another interesting method is to use high resolution methods behind beamformers [1 ].

The approach which is here proposed is completely different of the preceedings. One wants describe parameters of NCS in a general way.

For the resolution of that problem a priori information is however needed, but must be as little as possible. One uses, indeed, the following a priori information : the number of parameters defining NCS a "little" in respect to the number of sensors (of the array). Thus information on spatial structure of noise received by the array is very redundant.

One is , thus, led to define a functional depending only on sensors CSM and ambient noise CSM. That functional will be named, in the following, relative entropy functional (R.E.F.). R.E.F. will be maximized by iterative methods

HIGH RESOLUTION METHODS WITH NOISE CORRELATED SENSORS OUTPUTS. A RELATIVE ENTROPY APPROACH, METHOD AN RESULTS

(Gradient's like) relatively to the parameters of NCS.

One obtains by this way an estimation of NCS parameters which will be more acurate as sensors number increase relatively to number of parameters defining NCS. These estimates will be used for whitening of CSM outputs.

Firstly, one presents the relative entropy functional and the accuracy of the method. Then one considers an iterative algorithm for maximization of R.E.F., convergence is studied. Lastly one presents some results of simulation and data performed at sea.

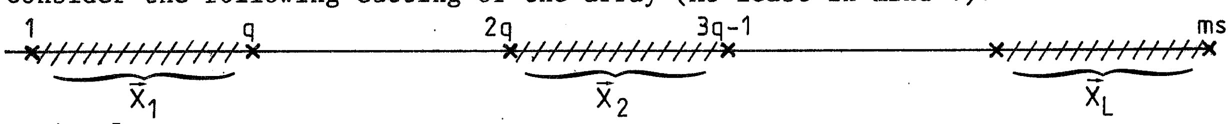
Some ideas presented here have yet been proposed by D.R. Farrier in a somewhat different context [ 2 ], [ 3 ] and specially the introduction of the R.E.F.

RELATIVE ENTROPY FUNCTIONAL
FORMULATION OF PROBLEM AND DERIVATION

Let be an array (linear) constituted by $n_s$ sensors, and B the CSM of ambient noise. B is assumed a Toeplitz matrix and will be, consequently, defined by its first row.

One supposes moreover that noise is decorrelated beyond q sensors, therefore B is completely described by q parameters. (The first q coefficients of the first row of B).

Consider the following cutting of the array (At least in mind !).




$\{ \vec{X}_i \}_1^L$ are the outputs vectors at a given frequency (omitted).

$$\vec{X}_1 = \vec{S}_1 + \vec{B}_1 \; ; \; \vec{X}_2 = \vec{S}_2 + \vec{B}_2 \; ; \; \dots \; ; \; \vec{X}_L = \vec{S}_L + \vec{B}_L \tag{1}$$

( $\vec{S}_i$ = signal part , $\vec{B}_i$ = noise part)

Then :

$$E\,(\vec{X}_1 \cdot \vec{X}_1^{\,*}) = Rq \quad (q,q) \text{ matrix}$$
$$E\,(\vec{B}_i \cdot \vec{B}_i^{\,*}) = B \quad (q,q) \tag{2}$$

Moreover vectors $\vec{B}_i$ and $\vec{B}_j$ are assumed decorrelated $i \neq j$ (because spacing between sensors corresponding to $B_i$ and $B_j$ is superior to q).

One considers then the vector (q. (L+1), 1) defined by :

$$\vec{\mathfrak{X}}^{\,t} = (\vec{X}_1 , \vec{B}_1 , \vec{B}_2 , \dots , \vec{B}_L)$$

Supposing that is a gaussian vector, the entropy corresponding to is given by the formula :

$$H\,(\vec{\mathfrak{X}}) = \text{Log det} \begin{bmatrix} Rq & B & & 0 \\ B & B & & \\ & & \ddots & \\ 0 & & & B \end{bmatrix} \tag{3}$$

Computation of H ( $\bar{\mathfrak{X}}$ ) becomes very expensive if L becomes great. This for computation of (3) one uses the following classical result for determinant of block matrices.

$$\det \begin{pmatrix} A & B \\ C & D \end{pmatrix} = \det A \,.\, \det (D - C.A^{-1}.B) \qquad (4)$$

In our case one considers the following block decomposition :

$$H(\vec{\mathfrak{X}}) = \text{Log} \det \left(\begin{array}{c|cccc} Rq & B & 0 & & 0 \\ \hline B & B & & & 0 \\ 0 & & \ddots & & \\ \cdot & & 0 & \ddots & \\ 0 & & & & B \end{array}\right)$$

Now using (4), one obtains :

$$H(\vec{\mathfrak{X}}) = \text{Log} \left\{ \det Rq.\det \left( \begin{pmatrix} B & & 0 \\ & \ddots & \\ 0 & & B \end{pmatrix} - \begin{pmatrix} B \\ 0 \\ \vdots \\ 0 \end{pmatrix} Rq^{-1} \,.\, (B\ 0 \ldots 0) \right) \right\}$$

Now :

$$\begin{pmatrix} B \\ 0 \\ \vdots \\ 0 \end{pmatrix} Rq^{-1} \; (B\ 0 \ldots 0) = \begin{pmatrix} B\,Rq^{-1}\,B & 0 & \ldots & 0 \\ 0 & 0 & & \\ \cdot & & \ddots & \\ 0 & & & 0 \end{pmatrix}$$

Therefore :

$$H(\vec{\mathfrak{X}}) = \text{Log} \left\{ \det Rq \,.\, \det \begin{pmatrix} B - B \,.\, Rq^{-1}\, B & & 0 \\ & B & \\ 0 & & \ddots \; B \end{pmatrix} \right\}$$

$$= \text{Log} \left\{ \det Rq \,.\, (\det B)^{L} \det (B - B \,.\, Rq^{-1} .\, B) \right\}$$

Now : $\det (B - B \,.\, Rq^{-1}.\, B) = \det B \,.\, \det (Id - Rq^{-1}\, B)$

$$= \det B \det Rq^{-1} \det (Rq - B)$$

One obtains therefore :

$$H(\vec{\mathfrak{X}}) = \text{Log} \det (Rq - B) + (L + 1) \,.\, \text{Log} \det B \qquad (5)$$

which is the formulation of the R.E.F. using only (q,q) matrices. The (L+1) factor (about $n_s/2q$) represent redundancy of information about noise received by the array.

The preceeding formula uses only the first q sensors, one may consider the functional obtained by averaging functionals related to a given choice of q sensors. Indeed, one prefers use of the Toeplitz matrix Rq which is obtained by orthogonal projection of R onto the Toeplitz subspace, ie :

$$\hat{Rq}\,(1,i) = \frac{1}{n_s - i + 1} \sum_{j=1}^{n_s} R\,(j - i + 1,\, j)$$

HIGH RESOLUTION METHODS WITH NOISE CORRELATED SENSORS OUTPUTS. A RELATIVE ENTROPY APPROACH, METHOD AN RESULTS

In that follows one considers the R.E.F. defined by :

$$Hq\ (B_1\ , \ldots, Bp) = \text{Log det}\ (\hat{R}q - B) + (L + 1)\ \text{Log det}\ B \qquad (6)$$

where $B = B\ (\beta_1, \beta_2, \ldots, \beta_p)$ $\qquad p < q$

$$B = \begin{pmatrix} \beta_1 & \beta_2 & \cdots & \beta_p & & 0 \\ \beta_2 & \ddots & & & \ddots & \\ \vdots & & & & & \beta_p \\ \beta_p & & & & & \\ & \ddots & & & & \\ 0 & & \beta_p & & & \beta_1 \end{pmatrix} \qquad B = \sum_{i=1}^{p} \beta_i\ U_i \quad \text{(real case)}$$

$\{ U_i \}$ basis of Toeplitz subspace

$$U_i = \begin{pmatrix} 1 & & 1 & & 0 \\ & \ddots & 0 & \ddots & \\ 0 & & 1 & & 1 \end{pmatrix}$$

Obviously another approachs of computation of R.E.F. are possibles.

DECORRELATED NOISE

Now one assumes that the noise received by the array is decorrelated (sensor to sensor). Then computation of the R.E.F. is very simple in this case and permit to connect the result with the eigenvalues of R.

Noise is assumed decorrelated, therefore $B = \lambda\ Id$.
Consider the eigensystem decomposition of R, let be :

$$R = \sum_{i=1}^{q} \alpha_i\ \vec{V}_i \cdot \vec{V}_i^{\,*} \quad , \vec{V}_i \perp \vec{V}_j\ (i \neq j)\ , \ \| \vec{V}_i \| = 1 \qquad \forall i \qquad (7)$$

with $\alpha_i \geqslant \alpha_2 \geqslant \ldots \geqslant \alpha_q$

Then :

$$H = \sum_{i=1}^{q} \text{Log}\ (\alpha_i - \lambda) + L \cdot q\ \text{Log}\ \lambda \qquad (8)$$

and :

$$\frac{\partial H}{\partial \lambda} = - \sum_{i=1}^{q} \frac{1}{\alpha_i - \lambda} + \frac{L \cdot q}{\lambda} = f\ (\lambda)$$

$f\ (\lambda)$ may also be written :

$$f\ (\lambda) = \frac{L \cdot q}{\lambda} + \sum_{i=1}^{q} \frac{1}{(\lambda - \alpha_q) + (\alpha_q - \alpha_i)}$$

If $\lambda = \alpha_q\ \left( \frac{L}{L + 1} \right)$, then :

$$f\ (\lambda) = (L + 1) \cdot \frac{q}{\alpha_q} + \sum_{i=1}^{q} \frac{1}{- \frac{1}{(L+1)}\ \alpha_q + (\alpha_q - \alpha_i)}$$

HIGH RESOLUTION METHODS WITH NOISE CORRELATED SENSORS OUTPUTS. A RELATIVE ENTROPY APPROACH, METHOD AN RESULTS

Now : $\alpha_q - \alpha_i \leqslant 0 \quad (i = 1, 2, \ldots, q)$
Therefore :

$$\left( - \frac{1}{(L+1)} \cdot \alpha_q + (\alpha_q - \alpha_i) \right)^{-1} \geqslant - (L+1) \cdot \frac{1}{\alpha_q}$$

Thus : $f\left( \frac{L}{L+1} \cdot \alpha_q \right) \geqslant 0$

Now, if $\lambda$ tends towards $\alpha q$ , $f(\lambda)$ tends towards $-\infty$. H being a differentiable and concave function in the interval $]0, \alpha q[$ one can deduces from that preceeds that H as only one maximum in $]0, \alpha_q[$ and this maximum $\hat{\lambda}$ is between $\alpha_q \cdot \left( \frac{L}{L+1} \right)$ and $\alpha_q$ .

$$\alpha_q \left( \frac{L}{L+1} \right) \leqslant \hat{\lambda} \leqslant \alpha_q \tag{9}$$

Obviously, if L becomes "great", $\hat{\lambda}$ tends to $\alpha_q$ . If the noise is correlated the preceeding approach is not valid and one must considers iterative methods.

## ITERATIVES METHODS FOR MAXIMIZATION OF H

### Study of convergence

$$\left( B = \sum_{i=1}^{p} \beta_i U_i \right)$$

$$\frac{\partial H}{\partial \beta_i} (R, B) = - \operatorname{tr} \left[ (R - B)^{-1} \cdot U_i \right] + L \operatorname{tr} \left[ B^{-1} \cdot U_i \right] \tag{10}$$

and : (11)

$$\frac{\partial^2 H}{\partial \beta_i \partial \beta_j} (R,B) = - \operatorname{tr} \left[ (R-B)^{-1} \cdot U_j \cdot (R-B)^{-1} \cdot U_i \right] + L \operatorname{tr} \left[ B^{-1} \cdot U_j \cdot B^{-1} \cdot U_i \right]$$

(Using $\frac{\partial}{\partial \beta} \text{Log det } R(\beta) = - \operatorname{tr} \left( R^{-1}(\beta) \cdot \frac{\partial R}{\partial B} \right)$ and $\frac{\partial R^{-1}}{\partial \beta} = - R^{-1} \frac{\partial R}{\partial \beta} R^{-1}$ )

Let be X some vector of $R^p$ ($X^t = (x_1, \ldots, x)$). Then :

$$\vec{X}^t H_2 \vec{X} = \sum_{i,j} X_i \frac{\partial^2 H}{\partial X_i \partial X_j} x_j$$

$H_2$ being the Hessian matrix. From (11), one obtains :

$$\begin{aligned} \vec{X}^t H_2 \vec{X} = & - \operatorname{tr} \left[ (R-B)^{-1} \cdot \left( \sum_{i=1}^{p} \beta_i U_i \right) . (R-B)^{-1} \cdot \left( \sum_{j=1}^{p} \beta_j U_j \right) \right] \\ & - \operatorname{tr} \left[ B^{-1} \cdot \left( \sum_{i=1}^{p} \beta_i U_i \right) . B^{-1} \cdot \left( \sum_{j=1}^{p} \beta_j U_j \right) \right] \end{aligned} \tag{12}$$

Each term of (12) may be written :

$- \operatorname{tr} (A \cdot C \cdot A \cdot C) \quad (A = (R-B)^{-1} \text{ or } B^{-1})$

A being supposed definite positive, may be decomposed in Choleski factors, ie $A = T \cdot T^*$, therefore :

$$\begin{aligned} - \operatorname{tr} [A \cdot C \cdot A \cdot C] &= - \operatorname{tr} [T \cdot T^* C \cdot T \cdot T^* C] \\ &= - \operatorname{tr} [ (T^* C T) (T^* CT)] = - \| T^* C T \|^2 \end{aligned}$$

Finally, one obtains that $H_2$ is definite negative and consequently H is (relatively to the $\beta_i$'s) a concave functional on the set of parametrised matrices B ($B = \sum_i \beta_i U_i$) such that B and (R-B) are definite positives matrices.

One can thus deduces from the above that gradient's methods (optimal step) will be converge on this subset.

Computation of optimal step

Computation of the gradient vector is straight forward, one has :

$$G_k(i) = - \operatorname{tr} [ (R-B_k)^{-1} U_i] + L \operatorname{tr} [B_k^{-1} U_i] \quad . \tag{13}$$

A gradient method is written as :

$$\vec{X}_{k+1} = \vec{X}_k - \rho_k \cdot \vec{G}_k \tag{14}$$

with : $\vec{X}_k^t = (\beta_1^k , \ldots, \beta_p^k)$

$\rho_k$ is the step of the algorithm, one wants to determine the optimal step for ensure convergence and principally satisfy constraints (R-B) and B definite positives.

The matricial translation of (13) and (14) is :

$$B_{k+1} = B_k - \rho_k \cdot D_k \tag{15}$$

where : $B_k = \sum_{i=1}^{p} \beta_i^k \cdot U_i$

$D_k = \sum_{i=1}^{p} G_k (i) \cdot U_i$

One wants obtain an explicit formulation of H $(R,B_{k+1})$ in function of $\rho$ . $B_k$ and $(R-B_k)$ being positive definites one may uses a Choleski decomposition of them, ie :

$$B_k = T_k \cdot T_k^* \text{ and } R - B_k = S_k \cdot S_k^* \tag{16}$$

Therefore :

$$\operatorname{Log} \det (R - B_k + \rho D_k) = \operatorname{Log} \det (S_k \cdot S_k^* + \rho D_k)$$

$$= \operatorname{Log} \det [S_k (Id + \rho S_k^{-1} D_k S_k^{-1*} ) S_k^* ]$$

$$= \operatorname{Log} \det (R - B_k) + \log \det (Id + \rho S_k^{-1} D_k S_k^{-1*}) \quad .$$

By the same way :

$$\operatorname{Log} \det (B_k - \rho D_k) = \operatorname{Log} \det B_k + \operatorname{Log} \det (Id - \rho T_k^{-1} \cdot D_k T_k^{-1*} )$$

Finally :

$$H (\rho) = \operatorname{Log} \det (Id + \rho S_k^{-1} D_k S_k^{-1*}) + \operatorname{Log} \det (Id - \rho T_k^{-1} D_k T_k^{-1*}) \tag{17}$$

$$+ \text{ constant term } (\rho) \quad .$$

HIGH RESOLUTION METHODS WITH NOISE CORRELATED SENSORS OUTPUTS. A RELATIVE ENTROPY APPROACH, METHOD AN RESULTS

Matrices $S_k^{-1} D_k S_k^{-1*}$ and $T_k^{-1} D_k T_k^{-1*}$ being hermitians, they are diagonalisables. Let $\{ \lambda_i^k \}_{i=1}^q$ and $\{ \mu_i^k \}_{i=1}^q$ be their respective eigenvalues, then :

$$H(\rho) = \sum_{i=1}^q \text{Log}(1 + \rho\, \lambda_i^k) + \sum_{i=1}^q \text{Log}(1 - \rho\, \mu_i^k) + \text{constant}(\rho) \qquad (18)$$

Constraints $(R - B_{k+1})$ and $(B_{k+1})$ definite positives are translated in :

$$\begin{aligned} 1 - \rho\, \lambda_i^k &\geqslant 0 \qquad i = 1, \dots, q \\ 1 - \rho\, \mu_i^k &\geqslant 0 \qquad i = 1, \dots, q \end{aligned} \qquad (19)$$

$\rho_k$ will be therefore determined by maximization of $H(\rho)$ given by (18) under constraints (19) by an unidimensionnal Newton method initialised at 0.

Implementation of algorithm of maximization of H

The price paid for computation of optimal step using formulas (18, (19) is not neglegeable but its great advantage is to explicit constraints of positivity on B and R-B. In that meaning, it is a very simple version of a projected gradient method.

In practice, one uses directly the algorithm defined by (13), (15), (18), (19), $G_k$ being calculated by formula (13) for the first iterations followed by a conjugate gradient method. Extension to complex parameters $\beta_i$ is straightforward.

## SIMULATION AND RESULTS

Signals received by the array are simulated as :

$$\vec{X} = \sum_{i=1}^n \sigma_i\, \vec{D}_{\theta i} + \vec{B} \qquad (\vec{D}_{\theta i} : \text{steering vector}) \qquad (20)$$

where $\sigma_i$ is $\mathcal{N}(0, \rho_i)$ and $\vec{B}$ is a gaussian vector of given covariance matrix B ($\vec{B}$ is obtained by Choleski decomposition of B).

At the end of running (of the algorithm) one obtains an estimate of B (named $\hat{B}_f$). One tries to determine the acuraçy of the algorithm by computation of cosine $(\hat{B}_f, B)$ defined by :

$$\cos(\hat{B}_f, B) = \frac{\text{tr}(\hat{B}_f^* \cdot B)}{\text{tr}(\hat{B}_f^* \cdot \hat{B}_f)\ \text{tr}(B^* B)} \qquad (21)$$

When $\hat{B}_f$ is colinear to B then $\cos(\hat{B}_f, B) = 1$, so $\cos(\hat{B}_f, B)$ must be as greater as possible ($\leqslant 1$). Then $\hat{B}_f$ is decomposed in triangular factors $\hat{B}_f = T_f \cdot T_f^*$ and R the C.S.M. of sensors outputs is whitened by the formula :

$$R_W = T_f^{-1} \cdot R \cdot T_f^{-1*} \qquad (22)$$

The Pisarenko method is applied concurrently to R and $R_W$ by formulas :

$$f_1(\theta) = \Big\{ \sum_{i=n_s-n}^{n_s} \big| \vec{U}_i^* \cdot \vec{D}_\theta \big|^2 \Big\}^{-1} \ ; \ f_2(\theta) \Big\{ \sum_{i=n_s-n}^{n_s} \big| \vec{V}_i^* \cdot T_f^{-1} \vec{D}_\theta \big|^2 \Big\}^{-1}$$

$\{ \vec{U}_i \}$ , $\{ \vec{V}_i \}$ being eigenvectors associated respectively to R and $R_W$.

HIGH RESOLUTION METHODS WITH NOISE CORRELATED SENSORS OUTPUTS. A RELATIVE ENTROPY APPROACH, METHOD AN RESULTS

One presents in figure 1 and 2 results corresponding to $f_1(\theta)$ and $f_2(\theta)$. In this case $q = 4$ $\beta_1 = 15.$, $\beta_2 = -4.$, $\beta_3 = 1.5$, $\beta_4 = 0.0$ with our algorithm $\hat{\beta}_1 = 12.6$, $\hat{\beta}_2 = -3.8$, $\hat{\beta}_3 = 1.0$, $\hat{\beta}_4 = 0.7$ and the average of cos $(B_f, B) = 1.0$ (0.999).

This result for estimation of the $\{\beta_i\}$ is good but it obtains for almost all choices fo $\{\beta_i\}$. Each maximization of the R.E.F. needs about twenty iterations or less.

One may also remark that sources are much more perceptibles on figure 2 rather than on figure 1. But in fact another great problemn is the existence of numerous spurious sources when NCS parameters are not or bad estimated. In another hand the misadjustement of p is not important.

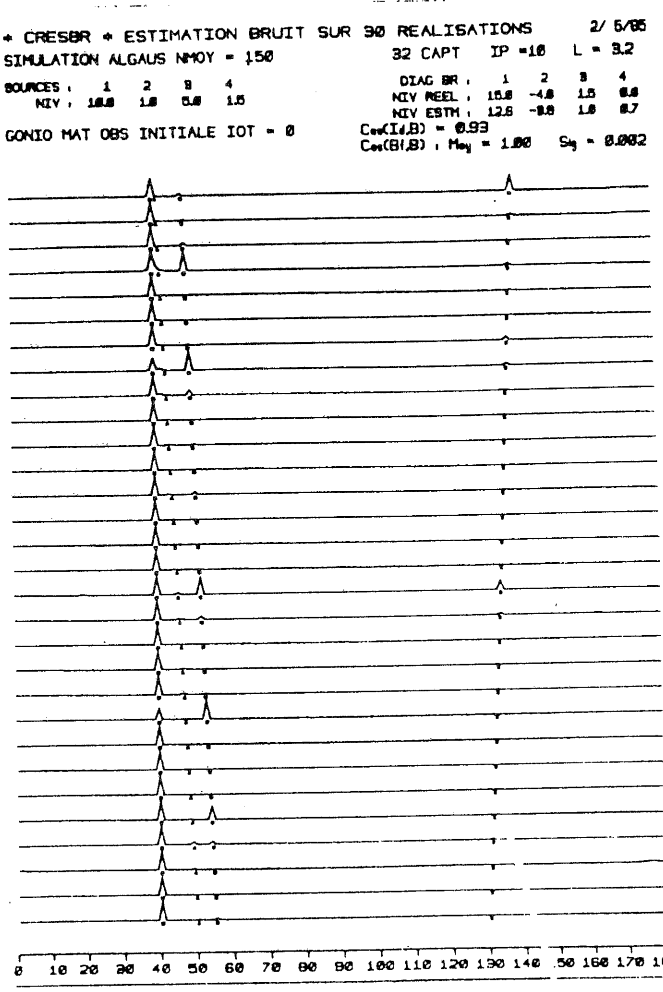



Fig. 1

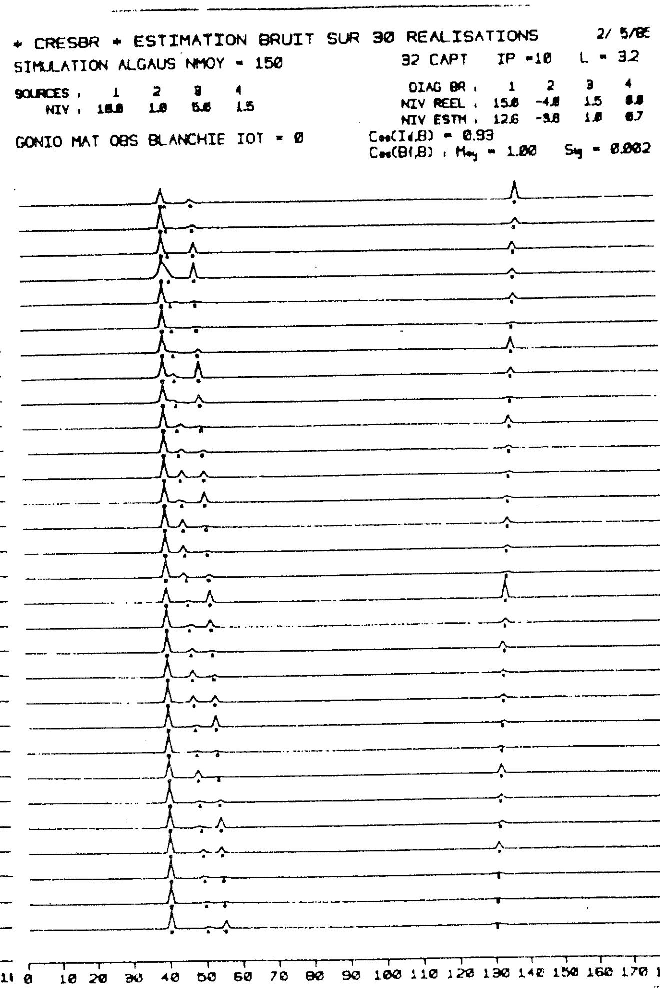



Fig. 2 (whitened)

Results for data performed at sea are illustrated by figures 3 and 4. One may remark that the use of Pisarenko method on the estimated CSM leads to many spurious sources. The algorithm of NCS estimation leads to suppress these spurious sources and to strongly enhances the true sources.

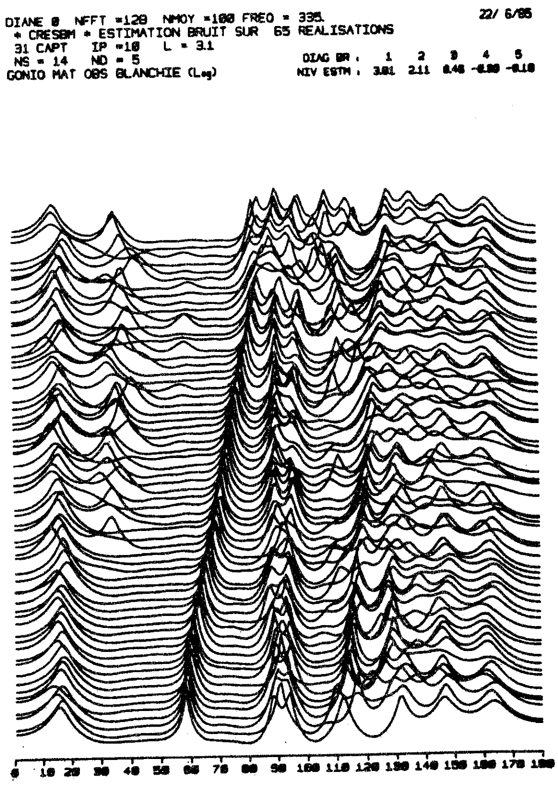



Fig. 4 (Whitened)

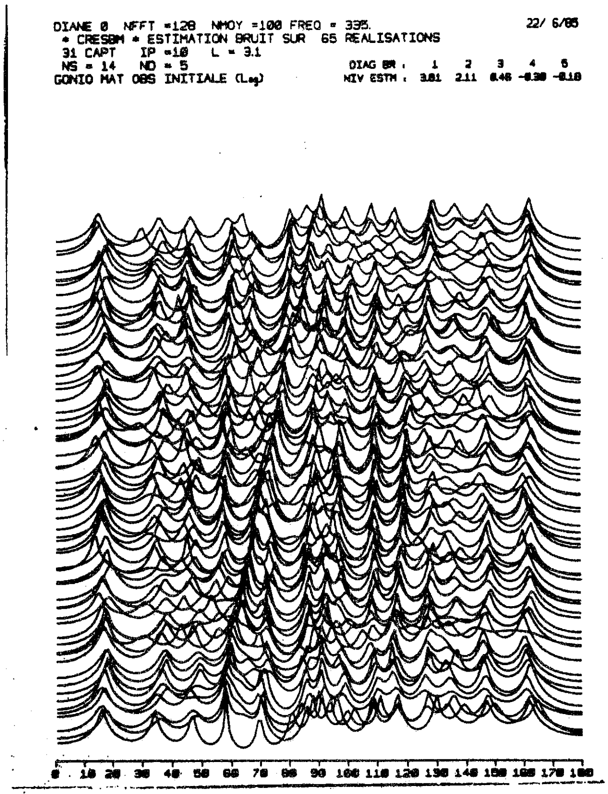



Fig. 3

Conclusion :

The method which is presented here is very suitable for simulation and data at sea, and provides a very convenient estimation of NCS parameters.

Furthermore it is robust and not very expensive in computation time.

Consequently that method leads to an important improvement of high resolution method.

REFERENCES :

[1] G. Bienvenu and L. Kopp, 'New approaches in the adaptive array processing field', NATO ASI on adaptive methods in underwater acoustics, Luneburg, D. Reidel 1985.

[2] D.R Farrier, 'Extension of maximum entropy spectral analysis. Part 1 : the inclusion of prior information', I.E.E. Proc, vol 131, Part F, n° 1, Feb 1984.

[3] D.R Farrier, 'Extension of maximum entropy spectral analysis. Part 2 interpretation and application of the method', I.E.E. Proc, vol 131, Part F, n° 1, Feb 1984.

[4] L. Kopp and G. Bienvenu, 'Multiple detection using eigenvalues when the noise spatial coherence is partially unknown', NATO ASI on adaptive methods in underwater acoustics, Luneburg, D. Reidel 1985.