ADAPTIVE M-ESTIMATORS FOR USE IN STRUCTURED AND UNSTRUCTURED ROBUST COVARIANCE ESTIMATION

Christopher L. Brown, Ramon F. Brcich, Christian Debes

Signal Processing Group, Institute of Telecommunications
Darmstadt University of Technology
Merckstrasse 25, D-64283 Darmstadt, Germany.
Email : chris.brown@ieee.org

Keywords : robust estimation, covariance, M-estimators

ABSTRACT

Covariance estimation is necessary in many applications such as source detection in array processing. Unfortunately, the sample covariance estimator is not robust. Here we investigate two broad approaches to robust covariance matrix estimation. The first is a model-free element-wise procedure, while the second is a structured approach based on pre-whitening. Both approaches utilise a robust one-dimensional scale estimator. It is the choice of this scale estimator and its effect on the overall covariance estimator that is the main purpose of this study. An adaptive M-estimator of scale is shown to have several advantages. Depending on the final comparison criterion, its use in a structured or element-wise covariance matrix estimator can lead to improved, robust performance.

1. INTRODUCTION

Numerous problems in signal processing require estimates of covariance, e.g., in array processing where the objective is to detect the number of sources impinging on an array or their directions of arrival (DOA). Unfortunately, the sample covariance estimator has poor performance when there are model deviations or outliers in the observations [1, 2].

Robust estimators protect against this, usually for only a small decrease in performance at the nominal model. Robustness is recognised as a favourable property since, in practice, it is more the norm than the exception that such disturbances exist. Several approaches have been suggested including: 1) FLOM (Fractional Lower Order Moment) estimators based on covariation [3, 4, 5], 2) nonparametric estimators using signs or ranks [2, 6], 3) Expectation maximisation (EM) applied to Gaussian mixture models [7], 4) Huber's robust M-estimators [8, 1].

The first two methods are computationally inexpensive. The iterative nature of the EM algorithm makes its complexity far greater and care must be taken to avoid local extrema. The last is arguably the method of choice but is difficult to use due to the multi-dimensional optimisation.

Here we investigate two broad approaches to covariance matrix estimation. The first is a model-free element-wise procedure, while the second is a structured approach based on pre-whitening. Both approaches utilise a robust one-dimensional scale estimator and so avoid complexity due to multi-dimensional optimisation. It is the choice of this scale estimator and its effect on the overall covariance estimator that is the main purpose of this study.

2. PROPOSED APPROACH

Consider the following signal model

$$\boldsymbol{x}(n) = A\boldsymbol{u}(n)\ , \quad n = 1, \ldots, N \tag{1}$$

where $\boldsymbol{x}(n) = [x_1(n), x_2(n), \ldots, x_M(n)]^T$ is the observation vector, $\boldsymbol{u}(n) = [u_1(n), u_2(n), \ldots, u_P(n)]^T$ is a vector of independent and identically distributed (iid) random variables and A is the $M \times P$ mixing matrix. The true covariance matrix is $C = \mathbf{E}\left[\boldsymbol{x}\boldsymbol{x}^H\right] = AA^H$ and each matrix element is $C(i,k) = \mathbf{E}\left[x_i x_k^*\right]$.

2.1. Standard element-wise estimation

The standard approach to estimation of each covariance matrix element is through the relationship

$$\hat{C}_\sigma(i,k) = \frac{1}{4}(\hat{\sigma}^2(x_i + x_k) - \hat{\sigma}^2(x_i - x_k)) \tag{2}$$

where $\hat{\sigma}(\cdot)$ is a scale estimator. If $\hat{\sigma}(\cdot)$ is the sample standard deviation, then (2) reduces to the sample covariance. However, if $\hat{\sigma}(\cdot)$ is a *robust* scale estimator, the overall covariance matrix estimate inherit robust properties.

2.2. Reconstructed covariance element-wise estimation

Huber [8] suggested that preferable to (2) is an element-wise estimator reconstructed from a robust correlation coefficient estimator and is given by

$$\hat{C}^*_\sigma(i,k) = \frac{\hat{\sigma}^2(x_i + x_k) - \hat{\sigma}^2(x_i - x_k)}{\hat{\sigma}^2(x_i + x_k) + \hat{\sigma}^2(x_i - x_k)} \hat{\sigma}(x_i)\hat{\sigma}(x_k) \ . \quad (3)$$

Using (2) and (3), we can form different estimators by using different robust scale estimators.

2.3. Pre-whitened covariance estimation

The sample spatial sign covariance matrix (SCM) was considered as a robust estimate of covariance in [2], but appeared earlier in [9] as the normalised sample covariance matrix in an adaptive radar detection problem where the clutter was K-distributed. The sample SCM is the sample covariance of the spatial sign function,

$$\boldsymbol{S}(\boldsymbol{x}) = \begin{cases} \frac{\boldsymbol{x}}{||\boldsymbol{x}||}, & \boldsymbol{x} \neq \boldsymbol{0} \\ \boldsymbol{0}, & \boldsymbol{x} = \boldsymbol{0} \end{cases},$$

of the data. This estimator was later improved in [10] to yield an iterative estimate which is a better approximate maximum likelihood solution for the covariance of K-distributed clutter than the SCM. The convergence and performance of this estimator was recently considered in [11].

When the objective is to estimate eigenvalues, empirical results [2] showed that for small samples it is better to whiten the observations using the eigenvectors of the sample SCM and then estimate the eigenvalues as the marginal variances of the transformed observations.

Using the median absolute deviation (MAD) to estimate the eigenvalues makes them scale equivariant, as does any scale equivariant estimator. In the one dimensional case, scale equivariance is assumed to be a fundamental property of scale estimators [8]. In the multidimensional case this property is still preferred. Note that due to the structured nature of the covariance matrix estimate it will be positive definite, another favourable property (c.f. the element-wise methods). Full details of the procedure can be found in the references.

2.4. Robust scale estimators

The methodologies described in the previous sections all utilise robust (one-dimensional) scale estimators. Using a different scale estimator will result in a different covariance matrix estimator. Below we briefly describe some of the scale estimators to be used with the above procedures.

2.4.1. Median absolute deviation

For a nominally Gaussian distribution, the median absolute deviation (MAD) scale estimator is

$$\hat{\sigma}_{\text{MAD}}(\boldsymbol{x}) = \frac{\text{median}(|\boldsymbol{x} - \text{median}(\boldsymbol{x})|)}{\Phi^{-1}(0.75)} \quad (4)$$

where $\Phi^{-1}(\cdot)$ is the inverse Gaussian cdf. This has been described as a ‘candidate for being the “most robust estimate of scale” ’[8].

2.4.2. M-estimators of scale

The Maximum Likelihood (ML) estimate of scale is found by solving the log-likelihood equation,

$$\sum_{n=1}^{N} \psi\left(\frac{x(n)}{\sigma}\right) = 0 \quad (5)$$

for σ where $\psi(x)$ is the scale score function associated with the density. By contrast, an M-estimator [8] replaces the nominal score function $\psi(x)$ with a similarly behaved function $\varphi(x)$ chosen to confer robustness on the estimator under deviations from the assumed density. Huber proposed a clipped quadratic score function

$$\varphi_H(x;k) = \min(x^2, k^2) - \delta = \begin{cases} x^2 - \delta, & |x| \leq k \\ k^2 - \delta, & |x| > k, \end{cases} \quad (6)$$

as it minimises the maximum relative asymptotic variance of the scale estimate in the case of a contaminated Gaussian distribution. δ is determined such that the estimator is unbiased for the nominal Gaussian distribution. The parameter k controls the sensitivity of the estimator to the contamination and should decrease as the proportion of outliers increases.

2.4.3. Adaptive M-estimators of scale

A drawback of the M-estimators is that the best value of the cut-off parameter k is dependent on the degree of contamination [12, 13]. In [14], an adaptive scheme was presented that sought to relieve this restriction. There, the score function was composed of a family of basis functions, the weights of which were chosen adaptively from the data. By using bases that were appropriate for a range of levels of contamination, the adaptive scheme was able to maintain high performance for a wider range of scenarios than the “static” M-estimators. For a full description of the adaptive algorithm, see [14].

2.4.4. FLOM based estimator

FLOM methods have impressive performance in impulsive noise [3]. They estimate the covariation of α-stable random processes – analogous to the covariance of Gaussian random variables. FLOM based measures of association were proposed in [4] for the purpose of determining DOA. The “covariation” matrices are found by

$$\hat{C}_{\text{FLOM}}(i,k;p) = \frac{\sum_{n=1}^{N} x_i(n)|x_k(n)|^{p-2}x_k^*(n)}{N} \ . \quad (7)$$

3. RESULTS

Herein, and without loss of generality, we only consider real random variables. In the results shown here, iid samples of $\boldsymbol{u}(t)$ for $M = P = 4$ and $N = 100$ were drawn from the selected distribution. The distributions were Gaussian mixtures where the nominal distribution was $\mathcal{N}(0,1)$ and the contaminating distribution was $\mathcal{N}(0,100)$. The probability of contamination took values $\varepsilon = 0$, 0.01, 0.02, 0.05, 0.1, 0.2, denoted by distributions 1,..., 6 respectively. A number of mixing matrices were considered, however, due to space limitations, results are only shown for one case

$$A(i,k) = \begin{cases} 1 & i = k \\ \frac{1}{4} & i \neq k \end{cases} .$$

Two measures of the accuracy of the resulting covariance matrix estimates are considered: 1) the Frobenius norm as a generic measure of the difference between matrices 2) the detected number of sources impinging on an array using the Minimum Description Length (MDL) criterion.

Average values for the metrics over 500 Monte Carlo runs were calculated. Selected results are shown here. The estimators follow the following naming conventions:

- Those starting with "Rec" denote they use the reconstructed element-wise procedure (see Sect 2.2). Those starting with "PWh" use the Pre-whitened covariance estimation method (Sect 2.3). Others use the standard element-wise approach (Sect 2.1).
- The adaptive M-estimator procedures ("Adapt M") use a library of clipped quadratic score functions with cut-offs of 1.5, 2 and 2.5.
- The static M-estimators ("Stat M") use clipped quadratic score functions with a cut-off of $k = 1.0$ or 2.5.

3.1. Frobenius norm

The Frobenius norm is the square root of the element-wise sum of squared differences between $\hat{C}$ and C

$$L_F(\hat{C}, C) = \sqrt{\text{trace}\{(\hat{C} - C)(\hat{C} - C)^H\}}. \tag{8}$$

From the results shown in Table 1 the following observations can be made:

- The reconstructed element-wise method showed improvement compared to the standard element-wise method except when the FLOM estimators were used. By contrast, the pre-whitened covariance estimation method did not lead to lower errors.

	Distribution					
Estimator	1	2	3	4	5	6
Sample	0.54	3.9	7.4	17	33	66
Adapt M	0.55	0.73	1.1	2.3	6.1	25
FLOM 1.5	0.67	1.1	1.9	5.1	10	20
Stat M 1.0	0.78	0.87	1	1.8	4	14
Stat M 2.5	0.54	0.85	1.4	4.9	15	44
MAD	0.91	0.96	1.1	1.8	3.9	14
Rec Adapt M	0.55	0.72	1.1	2.2	5.5	19
Rec FLOM 1.5	0.93	1.6	2.7	7.9	18	41
Rec Stat M 1.0	0.76	0.84	0.98	1.7	3.8	13
Rec Stat M 2.5	0.54	0.82	1.3	4.2	12	37
Rec MAD	0.86	0.92	1	1.6	3.7	12
SCM	2.7	2.7	2.7	2.6	2.6	2.6
PWh Adapt M	0.6	0.79	1.1	2.3	7.5	35
PWh Stat 1.0	0.74	0.85	1	1.8	4.4	19
PWh Stat 2.5	0.6	0.92	1.7	6.6	21	55
PWh MAD	0.79	0.86	1	1.7	4.1	17

Table 1. Average squared Frobenius norm of the difference between estimated and true covariance matrices.

- As expected, the sample standard deviation is good in the nominal case (distribution 1), however even small amounts of contamination results in significant degradation.
- FLOM and static M-estimator based methods rely on a design parameter that can be optimised if the degree of contamination is known – good nominal performance is traded off against increased robustness using this parameter. In general, the results shown suggest that, as expected, the adaptive M-estimator based procedure is able to maintain good nominal performance and high degrees of robustness, taking the "best of both worlds".
- The SCM estimator shows constant performance, invariant to the degree of contamination – at the cost of very poor nominal performance. PWh MAD shows better nominal performance, though not as good as the M-estimators. However it does display robust performance.

3.2. MDL

The Frobenius norm is an important theoretical measure useful in assessing the performance of a robust covariance estimator, however, the final judge of performance is how well the estimator behaves when applied to a specific engineering problem. With this in mind, the robust covariance estimators will also be assessed as an integral part of the source detection problem in array processing. There, the problem of detecting how many sources k transmit signals which impinge on a P element array from a record of N snapshots is solved by determining the multiplicity of the smallest sample eigenvalues of the data covariance. Other applications, such as the other fundamental problem of direction of arrival estimation, are possible but are not considered here.

The MDL criterion is a well-known information theoretic criterion used in model selection. It is used to estimate the number of distinct eigenvalues $k \in \{0, 1, \ldots, P-1\}$ from N iid realisations of a real $P-$variate Gaussian RV, choosing k as

$$\arg\min_k - N(P-k)\log\left(\frac{\left(\prod_{i=k+1}^{P} l_i\right)^{1/(P-k)}}{\frac{1}{P-k}\sum_{i=k+1}^{P} l_i}\right) + \frac{k(2P-k)}{2}\log N.$$

where l_i are the eigenvalues of the sample covariance matrix.

The MDL is preferred over the AIC (Akaike's Information Criterion) because it is a consistent estimator of the eigenvalue multiplicity. Other approaches exist based on a series of hypothesis tests, the sphericity test being a well known example. The common link between all these methods is that they are only functions of the data through the ratio of the geometric to arithmetic mean of the sample eigenvalues. Hence, a robust covariance estimator which performs well for the MDL can be expected to perform well for the AIC and sphericity test.

The true number of "sources" in the simulated scenario is 1. Inspecting the results in Table 2 we make the following observations:

- The pre-whitened covariance estimation procedures show noticeably superior performance to the reconstructed and sample element-wise methods for all scale estimators. This is despite the relatively poor performance seen in the previous section when using the Frobenius norm as the performance metric. This should not be completely surprising since the pre-whitening procedure is a structured estimator; i.e. one that utilizes the properties of covariance matrices, as opposed to the element-wise methods that do not exploit this structure. Therefore, since the detected model order utilizes the covariance matrix structure, it is not unreasonable that the estimators that are designed for this structure out-perform those that operate in a "model-free" manner. Further, this is confirmed by inspection of the estimated eigenvalues.

  For a qualitative evaluation of the eigenvalue estimates, consider the empirical distributions of the eigenvalue estimates as obtained from the sample covariance and adaptive M-estimators for Gaussian and Gaussian mixture $\varepsilon = 0.02$ (Distribution 3) data. Only results pertaining to the second largest eigenvalue are shown here since, in general, their behaviour is mirrored in the other eigenvalues. In addition, from an intuitive standpoint, the second largest eigenvalue in a sense represents the boundary between the signal and noise subspaces as it is the largest of the noise, i.e., multiple, eigenvalues.

  The histograms are shown in Figure 1 (sample covariance, Gaussian data), Figure 2 (adaptive M-estimator, Gaussian data), Figure 3 (sample covariance, Gaussian mixture data) and Figure 4 (adaptive M-estimator, Gaussian mixture data).

  For Gaussian data, there is no apparent difference between the estimators. This is a favourable result since in this case, the eigenvalues of the sample covariances represent the best performance.

  For Gaussian mixture data the results clearly show that the eigenvalues of the sample covariance have a much larger spread while the spread of the eigenvalues derived from the adaptive M-estimator have barely increased. It was also observed that the bias of the former estimator is much larger than that of the latter, the true value of the multiple eigenvalues being 0.56.

  Increasing ε will simply, as expected, increase the spread of the eigenvalues obtained from both the sam-

	Distribution					
Estimator	1	2	3	4	5	6
Sample	0.88	0.23	0.098	0.13	0.2	0.36
Adapt M	0.81	0.73	0.63	0.48	0.16	0.056
FLOM 1.5	0.87	0.69	0.55	0.43	0.4	0.44
Stat M 1.0	0.22	0.2	0.22	0.24	0.13	0.032
Stat M 2.5	0.85	0.78	0.67	0.21	0.048	0.046
MAD	0.092	0.082	0.082	0.086	0.048	0.036
Rec Adapt M	0.82	0.81	0.75	0.69	0.5	0.22
Rec FLOM 1.5	0.94	0.6	0.44	0.32	0.32	0.39
Rec Stat M 1.0	0.4	0.41	0.44	0.43	0.33	0.31
Rec Stat M 2.5	0.88	0.81	0.76	0.45	0.23	0.28
Rec MAD	0.23	0.26	0.26	0.27	0.2	0.13
SCM	0.94	0.94	0.94	0.95	0.95	0.93
PWh Adapt M	0.9	0.9	0.87	0.78	0.64	0.41
PWh Stat 1.0	0.66	0.7	0.66	0.62	0.64	0.45
PWh Stat 2.5	0.91	0.89	0.86	0.64	0.49	0.51
PWh MAD	0.61	0.56	0.59	0.61	0.57	0.44

Table 2. Probability of correctly detecting one source using MDL.

ple covariance and robust estimators. However, the spread of estimates obtained from the sample covariance shows a greater proportional increase compared to those obtained from the adaptive M-estimator.

Similar behaviour to that outlined above was observed for all the other robust estimators. The only notable additional trend was that the accuracy of the eigenvalue estimates decreased, i.e., the spread and bias increased, as the cut-off point of the M-estimator was increased or as the p parameter of the FLOM estimators increased.

- Static M-estimators show less variation as a function of their cut-off parameter using pre-whitening. Hence, there is also less difference between the static and adaptive M-estimator based methods.
- As noted previously, the SCM method provides excellent estimation of the subspace. Significantly, it is invariant to the degree of contamination – as should be expected from using the spatial sign of the observations.

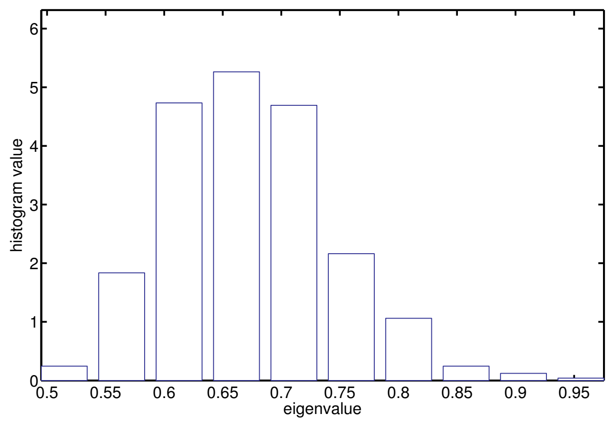

Fig. 1. Histogram of the second largest eigenvalue estimated from the sample covariance for Gaussian data (Distribution 1).

4. CONCLUSIONS

The proposed estimators, particularly those using the adaptive M-estimators, showed very encouraging performance. The advantage of the adaptive scheme over a static M-estimator is again confirmed through simulation results. In particular, it was significant to note the difference in relative performance when an element-wise method was used (leading to better element-by-element "matching" between the estimate and true value) compared to using a structured method (leading to better subspace estimation).

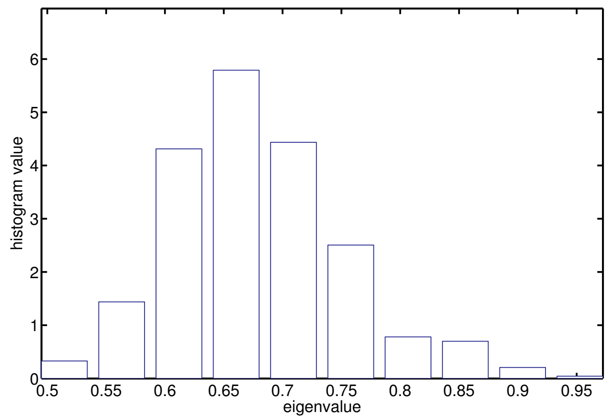

Fig. 2. Histogram of the second largest eigenvalue estimated using the adaptive M-estimator for Gaussian data (Distribution 1).

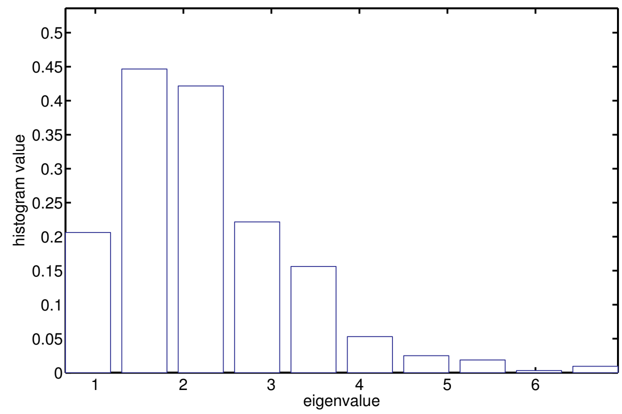

Fig. 3. Histogram of the second largest eigenvalue estimated from the sample covariance for Gaussian mixture data with $\varepsilon = 0.02$ (Distribution 3).

5. REFERENCES

[1] D. Williams and D. Johnson, "Robust estimation of structured covariance matrices," *IEEE Trans. Sig. Proc.*, vol. 41, no. 9, pp. 2891–906, Sept. 1993.

[2] S. Visuri, H. Oja, and V. Koivunen, "Subspace-based direction-of-arrival estimation using nonparametric statistics," *IEEE Trans. Sig. Proc.*, vol. 49, no. 9, pp. 2060–73, Sept. 2001.

[3] M. Shao and C. L. Nikias, "Signal processing with fractional lower order moments: Stable processes and their applications," *Proc. IEEE*, vol. 81, no. 7, pp. 986–1010, July 1993.

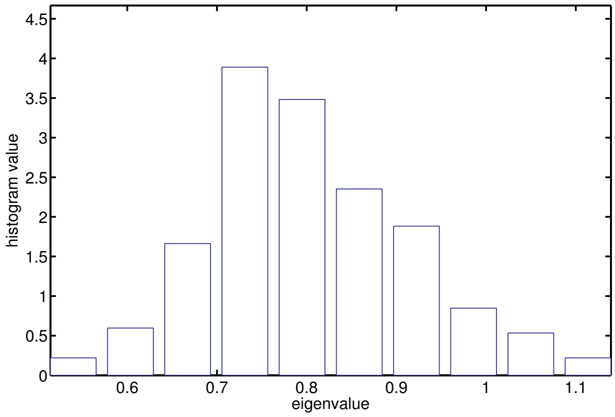

Fig. 4. Histogram of the second largest eigenvalue estimated using the adaptive M-estimator for Gaussian mixture data with $\varepsilon = 0.02$ (Distribution 3).

[4] P. Tsakalides and C. Nikias, "The robust covariation-based MUSIC (ROC-MUSIC) algorithm for bearing estimation in impulsive environments," *IEEE Trans. Sig. Proc.*, vol. 44, no. 7, pp. 1623–33, July 1996.

[5] T. Liu and J. Mendel, "A subspace-based direction finding algorithm using fractional lower order statistics," *IEEE Trans. Sig. Proc.*, vol. 49, no. 8, pp. 1605–13, Aug. 2001.

[6] M. Kendall and J. D. Gibbons, *Rank Correlation Methods*, Oxford Univ. Press, 5th edition, 1990.

[7] R. Kozick and B. Sadler, "Maximum likelihood array processing in non-Gaussian noise with Gaussian mixtures," *IEEE Trans. Sig. Proc.*, vol. 48, no. 12, pp. 3520–35, Dec. 2000.

[8] P. Huber, *Robust Statistics*, John Wiley, 1981.

[9] F. Gini, M. Greco, and L. Verrazzani, "Detection problem in mixed clutter environment as a gaussian problem by adaptive preprocessing," *Electronics Letters*, vol. 31, no. 14, pp. 1189–1190, July 1995.

[10] F. Gini and M. Greco, "Covariance matrix estimation for cfar detection in correlated heavy tailed clutter," *Signal Processing*, vol. 82, no. 12, pp. 1847–1859, December 2002.

[11] F. Pascal, P. Forster, J-P. Ovarlez, and P. Larzabal, "Theoretical analysis of an improved covariance matrix estimator in non-gaussian noise," in *Proceedings of ICASSP 2005*, 2005.

[12] R. F. Brcich and A.M. Zoubir, "Robust estimation with parametric score function estimation," in *Proc. ICASSP 2002*, Orlando, FL, USA, May 2002.

[13] C. L. Brown, R. F. Brcich, and A. Taleb, "Suboptimal robust estimation using rank score functions," in *Proc. ICASSP 2003*, Hong Kong, April 2003.

[14] R. F. Brcich, C. L. Brown, and A. M. Zoubir, "An adaptive robust estimator for scale in contaminated distributions," in *Proc. ICASSP 2004*, Montreal, Canada, May 2004.