

# Robust Estimation of Cyclic Correlation in Contaminated Gaussian Noise

Thomas E. Biedka<sup>†‡</sup>, Lamine Mili<sup>‡</sup>, and Jeffrey H. Reed<sup>†</sup>

<sup>†</sup> E-Systems, Inc.

Greenville Division  
Greenville, TX 75403

<sup>‡</sup> Virginia Tech

Dept. of Electrical Engineering  
Blacksburg, VA 24061

## Abstract

This paper considers the effect of non-gaussian noise on the conventional estimate of cyclic correlation. It is shown that noise having a distribution function with heavier tails than the gaussian slows the convergence of the estimate to the expected value. Alternative estimators are proposed based on the statistical concepts of robustness. These alternative estimators are shown via Monte Carlo simulation to perform well in both gaussian and non-gaussian noise. Another contribution of this paper is the generalization of some robust estimators to complex (versus real) data.

## 1 Introduction

A signal is said to exhibit second-order cyclostationarity if it is correlated with a frequency shifted (and possibly time-delayed and/or conjugated) version of itself. This correlation will occur only for a set of discrete frequency shifts  $\{\alpha_k\}$  known as the cycle frequencies of the signal. Many communication signals, such as AM, PSK, and QAM, are second order cyclostationary. The cycle frequencies are typically equal to integer multiples of the symbol rate, twice the carrier frequency, and sums and differences of these frequencies. Typical applications where cyclostationarity can be exploited include signal detection, classification, interference cancellation, etc. [1]. The performance improvement over more conventional methods arises from the fact that the noise and interference present in the data is not correlated with a frequency shifted version of itself. This of course assumes that the noise and interference is not cyclostationary, or exhibits cyclostationarity at cycle frequencies other than those of the signal of interest.

Many algorithms that exploit cyclostationarity use the sample cyclic correlation. This statistic can be viewed as the sample cross-correlation between a sequence and a frequency shifted version of that same sequence. This statistic is shown here to perform sub-optimally if the noise present in the observed data is not gaussian. To be specific, the sample cyclic correlation performs poorly if the noise has longer tails than the gaussian. Noise present in communications systems is often assumed to be gaussian. This assumption is valid, for example, when receiver thermal noise is dominant. In other cases the noise is clearly *not* gaussian. Near urban areas, ignition systems can

cause impulse noise in the VHF band. In radar systems, glint noise has been shown to be impulsive. In the ELF band, the noise is well known to be very impulsive. Because the exact distribution of the noise is rarely known, it is important to develop estimators which perform well over a range of noise distributions. This is a basic concept of robust statistics.

The maximum likelihood estimator (MLE) has the minimum possible asymptotic variance for a given probability density function (PDF). In contrast, a robust estimator accepts some small increase in variance at one PDF in order to have lower variance at other PDFs. There are many different definitions of what constitutes a 'robust' estimator. For the problem considered here, we will consider *variance* robustness. That is, we seek estimators of cyclic correlation which have relatively low variance over a broad range of distributions. The concepts of robustness have received an enormous amount of attention in the statistics literature, e.g., [2, 3, 4]. There has also been considerable attention given to robust signal processing methods (e.g., see the survey by Kassam and Poor [5]).

The problem of detecting a cyclostationary signal in non-gaussian noise has been previously considered by Izzo et al. [6]. Their work considers the problem of locally optimal detection in specific non-gaussian noise distributions. The resulting methods require utilization of specific estimators with specific noise distributions. This implies a priori knowledge of the noise distribution, which is very difficult to obtain.

The robust methods proposed here will be shown to have an intuitively appealing form in that they place low weight on samples with large magnitude. However, we will motivate these methods from a more theoretical perspective. We first give a very brief overview of robust estimation, then show how a robust estimate of scale can be modified to become a robust estimate of cyclic correlation. In so doing we also introduce robust estimates of scale for complex sequences. Then Monte Carlo simulation results are presented that illustrate the performance of the conventional and robust estimators in both gaussian and non-gaussian noise.

## 2 Overview of Robust Statistics

Although we are working with communication signals, which are zero mean, it is instructive to consider robust estimation of location. The location of a sym-

metric PDF is equal to the mean, the median, and the mode. An  $M$ -estimator of location is the solution  $\theta$  of

$$\sum_{n=0}^{N-1} \psi(x_n - \theta) = 0 \quad (1)$$

where  $\psi$  is called the score function. The MLE at a particular PDF  $f(x)$  is obtained by taking  $\psi$  proportional to the derivative of the log-likelihood:  $\psi(x) = -(f'/f)(x)$ . The MLE at the gaussian distribution is not robust because the resulting score function  $\psi(x) = x$  is not bounded. Thus outliers in the data will not be downweighted. Modifying this  $\psi$  function to be bounded yields *Huber's  $\psi$  function*:

$$\psi(x) = \begin{cases} x & \text{for } |x| < a \\ ax/|x| & \text{otherwise} \end{cases} \quad (2)$$

If  $a$  is relatively small, the estimator will behave more like the sample median and will be more robust. If  $a$  is relatively large, the estimator will behave more like the sample mean and will have lower variance at the gaussian distribution.

The variance of a robust estimator is usually considered relative to the variance of the MLE. This is the asymptotic relative efficiency (ARE). Denote the MLE by  $\theta_{ML}$  and the robust estimator by  $\theta_1$ . Then the ARE of  $\theta_{ML}$  and  $\theta_1$  is defined as

$$e_{\theta_1, \theta_{ML}} \triangleq \lim_{N \rightarrow \infty} \frac{\text{var}(\sqrt{N} \theta_{ML})}{\text{var}(\sqrt{N} \theta_1)} \quad (3)$$

where  $\text{var}(\cdot)$  denotes variance at some PDF. For example, take  $\theta_1$  to be the sample median and  $\theta_{ML}$  to be the sample mean. At the gaussian distribution, the ARE of the sample median with respect to the sample mean is  $0.637 < 1$ . Thus the sample mean is said to be more *efficient*. However, at the Laplacian distribution, the ARE of the sample median w.r.t. the sample mean is  $2 > 1$ . Thus the sample median is more efficient than the sample mean at the Laplacian distribution. The efficiency of Huber's  $M$ -estimator at the gaussian distribution can be increased by increasing the parameter  $a$  in (2). However, increasing  $a$  results in higher variance at distributions with longer tails. This demonstrates the tradeoff between an estimator's robustness and its efficiency.

Huber's  $\psi$  function is only one of many proposed  $\psi$  functions. Others may be used depending on the desired behavior. For example, Huber's  $\psi$  function does not reduce the influence of extreme outliers to zero. Therefore it may be desirable to use a *redescending*  $\psi$  function. These functions have the property that  $\psi(x) = 0$  for  $x > a$ . One example of a redescending  $\psi$  function is the Huber-type skipped mean<sup>1</sup>  $\psi$ , which is identical to (2) except that  $\psi(x) = 0$  for  $|x| > a$ . Another example is Tukey's biweight function, where

<sup>1</sup>for brevity, we will refer to this as the 'skipped'  $\psi$  function.

$\psi(x) = x(a^2 - |x|^2)^2$  for  $|x| < a$ , and  $\psi(x) = 0$  otherwise. A key feature of the  $\psi$  functions described here is  $\psi(x) \simeq x$  for small  $x$ .

This discussion of robust estimation of location has been included to give background information and also to demonstrate the importance of the  $\psi$  function. For the application considered here we can safely assume that the location parameter  $\theta = 0$ . Thus we need only consider estimation of the scale parameter.

The MLE of scale is found by: 1) Taking the partial derivative of the log-likelihood function with respect to the scale parameter  $\sigma$ ; 2) setting this to zero; and 3) solving for  $\sigma$ . One robust method for estimating scale is Huber's so-called 'Proposal 2'. For a zero-mean sequence, this requires solving for  $\sigma$  such that

$$\sum_{n=0}^{N-1} \psi^2\left(\frac{x_n}{\sigma}\right) = Nc. \quad (4)$$

The tuning constant  $c$  is selected so that the estimator is Fisher consistent at the gaussian distribution, i.e., so that  $\sigma \rightarrow \sigma_{\text{true}}$  as  $N \rightarrow \infty$  at the gaussian distribution. The constant  $c$  is dependent on the  $\psi$  function, which in turn is selected based on the conflicting goals of high efficiency and robustness. One method of solving (4) proposed by Huber is to initialize  $\sigma$  with a very robust (but perhaps inefficient) estimate of scale. At the  $k$ th iteration, the new estimate is

$$\sigma_{k+1}^2 = \frac{\sigma_k^2}{Nc} \sum_{n=0}^{N-1} \psi^2\left(\frac{x_n}{\sigma_k}\right) \quad (5)$$

To save computations, a *one-step* estimator is often used. Such an estimator applies only one iteration of the iterative procedure used to solve for the estimate. It is often the case that such estimators are both efficient and robust. The robust initial estimate  $\sigma_0$  is often taken to be the Median Absolute Deviation from the median (MAD), defined as

$$\text{MAD} \triangleq \text{med}_n |x_n - \text{med}_i x_i| \quad (6)$$

where  $\text{med}(\cdot)$  is the sample median. The MAD is extremely robust because it is based on the sample median. For large  $N$  the MAD is scaled by 1.4826 for Fisher consistency at the gaussian distribution.

### 3 Robust Cyclic Correlation

A commonly used estimate of the cyclic correlation is the sample cyclic correlation, defined as

$$\hat{R}_{xx}^{\alpha}(\tau) \triangleq \sum_{n=0}^{N-1} x_n x_{n-\tau}^* e^{-j2\pi\alpha n}. \quad (7)$$

For infinite collect time, noise and interference not exhibiting cyclostationarity at the selected cycle frequency will not contribute to  $\hat{R}_{xx}^{\alpha}(\tau)$ . In practice, the collect time  $N$  must be large enough so that  $\hat{R}_{xx}^{\alpha}(\tau)$  is dominated by contributions from the desired signal

in order for cyclostationarity exploiting algorithms to behave as expected.

The robust methods presented so far are only valid for *real* data. Modifications to the existing methods are required since we will assume that the data is in a complex baseband representation. To the authors' best knowledge, generalization of robust statistics to complex data has not been previously described. We first generalize the MAD to complex data. For complex, zero mean data, the complex MAD is given by

$$\text{CMAD} \triangleq \text{med} |x_n|. \quad (8)$$

The CMAD must be multiplied by  $1/\sqrt{\ln(2)} = 1.2011$  for Fisher consistency, since the expected value of (8) is  $\sqrt{\ln(2)}$  if  $x$  is a complex, zero mean, unit variance gaussian random variable.

In general the  $\psi$  function is only defined for real signals. However, the  $\psi$  functions as defined here are valid for both real and complex data. For  $x = re^{j\theta}$ , where  $r = x/|x|$ ,  $\psi(x) = \psi(re^{j\theta}) = \psi(r)e^{j\theta}$ . Thus only the magnitude is affected; the phase of each sample is not altered. Also note that  $\psi^*(x) = \psi(x^*) = \psi(r)e^{-j\theta}$ . We can define a robust one-step estimate of scale for a complex sequence as

$$\sigma_1^2 = \frac{\sigma_0^2}{Nc} \sum_{n=0}^{N-1} \psi\left(\frac{x_n}{\sigma_0}\right) \psi\left(\frac{x_n^*}{\sigma_0}\right) \quad (9)$$

where the robust initial estimate of scale  $\sigma_0$  is the CMAD. The tuning constant  $c$  in (9) is included for Fisher consistency. For  $\psi(x) = x$ , (9) reduces to the more familiar sample variance. We can now define a robust one-step estimate of the cyclic correlation as

$$\hat{R}_{xx}^\alpha(\tau) = \frac{\sigma_0^2}{Nc} \sum_{n=0}^{N-1} \psi\left(\frac{x_n}{\sigma_0}\right) \psi\left(\frac{x_{n-\tau}^*}{\sigma_0}\right) e^{-j2\pi\alpha n}. \quad (10)$$

If the lag parameter  $\tau$  is set to zero,

$$\hat{R}_{xx}^\alpha(0) = \frac{\sigma_x^2}{Nc} \sum_{n=0}^{N-1} \psi^2\left(\frac{r}{\sigma_x}\right) e^{-j2\pi\alpha n}. \quad (11)$$

A key point is that both (10) and (11) are explicit functions of the data  $x$ .

This method for obtaining a robust estimate of the cyclic correlation can be motivated from a more intuitive perspective. Passing a second-order cyclostationary signal through a quadratic non-linearity causes regeneration of spectral lines. However, a quadratic non-linearity is a poor choice in impulsive noise since large values are greatly amplified. If we know that the noise is impulsive in nature, it makes sense to use some method to limit the effect of large values. We can use a hard rejection non-linearity, where samples above a certain threshold are set to zero. We can also use a soft-limiter non-linearity, such as the Huber  $\psi$ . The robust  $\psi$  functions make good choices for the non-linearities since they bound the contribution

of large samples and  $\psi(x) \simeq x$  for small  $x$ . The trick is to properly select the threshold to get good performance. The CMAD provides a robust method to set the threshold for various non-linearities.

A short aside is needed here on using the CMAD to estimate scale rather than the conventional sample variance. It is well known that the sample variance is not robust to outliers. The presence of outliers typically inflates the value of the sample variance. Hence a conventional trimming rule based on the sample variance may not trim abnormally large values. Thus it is advisable to use a robust estimate of scale when using a trimming approach.

One way to make any estimator more robust is to remove a fixed fraction  $\beta$  of the samples having the largest magnitude. The trimmed mean is a good example. Define the frequency-shifted lag-product waveform  $z_n$  as  $x_n x_{n-\tau}^* e^{-j2\pi\alpha n}$ . Then define the ordered samples  $z_{(n)}$  by sorting  $z_n$  from smallest to largest magnitude so that  $|z_{(0)}| \leq |z_{(1)}| \leq \dots \leq |z_{(N-1)}|$ . The *trimmed* cyclic correlation is then defined as

$$\hat{R}_{xx}^\alpha(\tau) \triangleq \frac{1}{N - \beta N} \sum_{n=0}^{N - \beta N - 1} z_{(n)} \quad (12)$$

where  $\beta$  is the fraction of samples to be trimmed. The trimmed cyclic correlation is simple to implement and is also highly effective in removing outliers. However, there are some drawbacks to this approach. First, by always removing a fixed amount of data there is a loss of efficiency when the noise is gaussian. Second, this method is robust only if the percentage trimmed is greater than the percentage of outliers. If the number of outliers exceeds the percentage trimmed, the robustness is lost. The trimmed cyclic correlation must also be scaled so that it is Fisher consistent.

#### 4 Performance Comparison

The problem of estimating the baud timing of a QPSK signal will be used to compare the performance of the conventional and proposed robust methods. Ideally the phase of the cyclic correlation will be equal to the baud timing phase when the cycle frequency is set to the baud rate. In the presence of noise, the estimated baud timing will be a random variable with a certain mean and standard deviation. Therefore the standard deviation of the phase of the cyclic correlation is used as the performance criterion. Performance in both gaussian and non-gaussian noise is examined. The non-gaussian noise is drawn from a gaussian mixture process, also known as a contaminated gaussian process. This process has a PDF given by

$$f(x) = (1 - \epsilon)\phi(x/\sigma_1) + \epsilon\phi(x/\sigma_2) \quad (13)$$

where  $\phi(x)$  is the gaussian PDF with mean zero and unit variance,  $\epsilon < 0.5$  controls the degree of contamination, and  $\sigma_2/\sigma_1$  controls the relative strength of the contamination. For results presented here,  $\epsilon$  is fixed at 0.1. When  $\sigma_2 > \sigma_1$ , (13) has longer tails than the gaussian. Since we are working with complex signals, the noise is also complex, and in addition

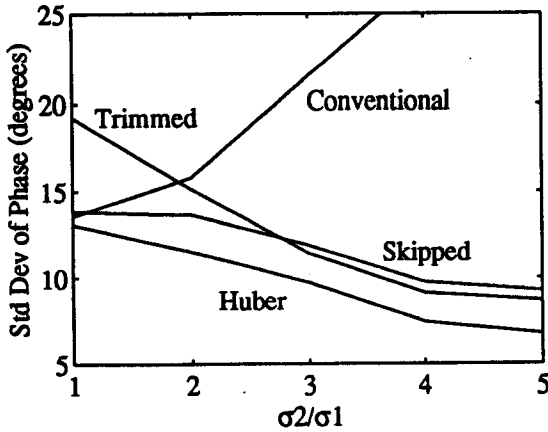


Figure 1: Standard deviation of the baud timing estimate versus relative contamination strength  $\sigma_2/\sigma_1$ .

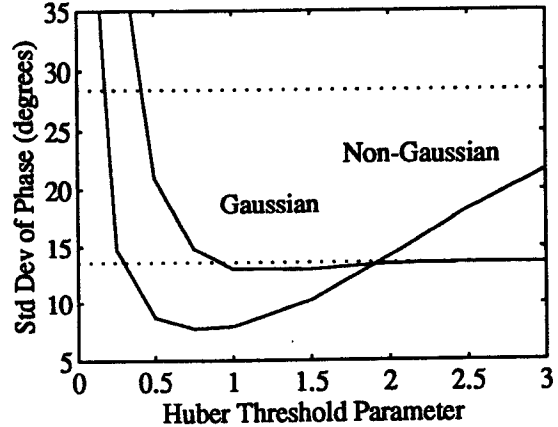


Figure 2: Performance of the Huber estimator (solid lines) as a function of the threshold parameter 'a'. Dotted lines correspond to the conventional method.

it is circularly symmetric. The QPSK signal has a baud rate of 0.25 relative to the sample rate of unity, and has Nyquist zero-ISI pulse shaping with 100% excess bandwidth. The signal is generated with a known baud timing of  $0^\circ$  and has zero carrier frequency offset. The cyclic correlation is computed with  $\alpha$  equal to the baud rate and  $\tau$  set to zero. All results are based on 1000 independent trials at 0 dB SNR.

We first evaluate the efficiency and robustness of the estimators by comparing performance as a function of contamination strength. Figure 1 shows the standard deviation of the angle of the cyclic correlation as a function of  $\sigma_2/\sigma_1$  for both the conventional and robust methods, based on  $N = 1024$ . The Huber method uses  $a = 1$ , the skipped method uses  $a = 2$ , and the trimmed method removes the 5% of the samples having the largest magnitude. We compare the performance of the estimators when the noise is purely gaussian ( $\sigma_2/\sigma_1 = 1$ ) to examine efficiency. Figure 1 shows that the conventional, Huber, and skipped methods perform nearly the same. This indicates that the Huber and skipped methods are efficient at the gaussian distribution. In contrast, the trimmed method performs much worse than the other methods. Thus we say that the trimmed method is less efficient than the Huber or skipped methods.

We compare performance as  $\sigma_2/\sigma_1$  increases to examine the robustness of the estimators. Figure 1 shows that the performance of the conventional method degrades rapidly as the tails of the noise become longer. There is noticeable degradation even for relatively small deviation from gaussian. In contrast, all of the robust methods proposed here perform much better than the conventional method in non-gaussian noise. The Huber method appears to have the best overall performance for this particular integration time and SNR. Figure 1 shows that we have achieved our main goal of developing estimators of cyclic correlation that work well in both gaussian noise and noise with heavier tails.

Note that the performance of the robust methods

improves as the noise becomes more impulsive. The improvement of the trimmed method is particularly noticeable. This can be explained as follows. Consider a case where the noise is gaussian except for a single outlier. As the power of the outlier is increased, the power of the gaussian noise must be reduced so that the SNR in the data remains 0 dB. In the limit, all of the noise power will be contained in the outlier and the rest of the data will be noise-free. Therefore if the outlier is removed, the cyclic correlation can be very accurately estimated.

We might expect the conventional method to perform better than any other method when the noise is gaussian. However, the results in Figure 1 seem to show that the Huber estimator has slightly lower phase variance, even when the noise is gaussian. This behavior has appeared consistently in simulations. The following explanation is offered. The performance of an estimator depends on the PDF of the data. The data we are working with contains gaussian noise and a QPSK signal. Even though the noise is gaussian, the data is *not*. Since the QPSK signal and the noise have equal power, the real part of the data follows a gaussian mixture PDF approximately given by

$$f(x) = 0.5 \phi(x - 1) + 0.5 \phi(x + 1) \quad (14)$$

where  $\phi(x)$  is the gaussian PDF. The imaginary part of the data has the same distribution as the real part. This gaussian mixture PDF has longer tails than the gaussian. Thus it is quite plausible that the Huber estimator would perform better than the conventional method. Note that if the SNR were lower, the data would be closer to gaussian, and this small performance difference would likely not be apparent.

It has been stated that the parameter  $a$  in the Huber score function controls the relative efficiency and robustness of the estimator. Small values of  $a$  yield greater robustness at the cost of lower efficiency. While this is true in general, special considerations must be taken into account when exploiting cyclosta-

tionarity. If  $a$  is set to zero in the Huber  $\psi$  function, we have  $\psi(x) = x/|x| = e^{j\phi}$ . If the lag parameter  $\tau$  is set to zero we have, from (11),  $R_{\tau x}^\alpha(0) = 0 \forall \alpha \neq 0$ . Thus we expect the performance of the robust cyclic correlation given by (11) to degrade as  $a$  becomes small. This behavior is illustrated in Figure 2. This figure also shows why  $a = 1$  was used in the Huber method for the results in Figure 1. Results in both figures are based on 256 symbols. However, the Huber  $\psi$  function can be used with a very small  $a$  if the lag parameter  $\tau$  is set to a non-zero value. Simulation results not shown here reveal that setting  $\tau$  equal to one-half of the baud period yields good performance with small  $a$ . This makes sense, since the Huber  $\psi$  with a small  $a$  forces the signal to have a nearly constant envelope. For constant envelope PSK, setting  $\tau$  equal to half the baud period yields the strongest spectral line.

By definition, a signal is second order cyclostationary if passing the signal through a second order non-linearity causes regeneration of spectral lines. Because the robust estimators use a non-linearity which is not strictly second order, it is not clear that these estimators solely exploit second order cyclostationarity. This is mainly an issue if, for example, the threshold parameter  $a$  in the Huber estimator is very small. A large value of  $a$  yields an estimator that is very similar to the conventional cyclic correlation. It is important to note that the non-linear  $\psi$  functions used here do not themselves appear to regenerate spectral lines.

Another important issue is the performance of these methods as a function of integration time. This is illustrated in Figure 3. Here the noise is non-gaussian with  $\epsilon = 0.1$  and  $\sigma_2/\sigma_1 = 4$ . As expected, the conventional cyclic correlation performs well as  $N$  becomes large, even though the noise is not gaussian. However, the robust methods always have lower variance.

## 5 Summary and Conclusions

This paper has introduced the statistical concepts of robust estimation to the exploitation of cyclostationarity. In some ways the robust estimators of cyclic correlation proposed here behave as expected. They perform well if the noise is gaussian or if the noise is more impulsive. However, some subtleties arise. In order to obtain robust estimators, the contribution of large samples must be downweighted. In the context of exploiting cyclostationarity, this amounts to passing the data through a non-linearity before computing the conventional cyclic correlation. These non-linearities do not appear to regenerate spectral lines. However, if the non-linearity is strong, it can affect the cyclostationarity properties of the desired signal. For example, a robust estimator may require a different choice for the lag parameter  $\tau$  than would normally be used with the sample cyclic correlation. By processing the data with a non-linearity before computing the cyclic correlation, we are not using a simple second order non-linearity to regenerate spectral lines. Thus, strictly speaking, robust estimators of cyclic correlation may not solely exploit second order cyclostationarity. These points will be investigated further.

Another area which requires further investigation is the incorporation of robust estimates of cyclic cor-

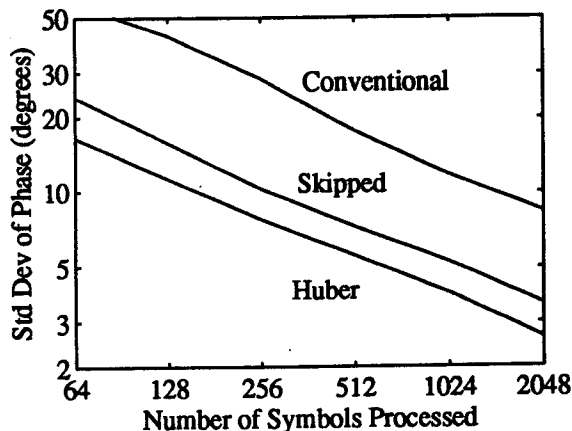


Figure 3: Performance of the conventional and robust methods versus collect time in non-gaussian noise.

relation in existing algorithms. Preliminary analysis not included here indicates that, in some cases, the conventional estimate can be simply replaced by the robust estimate. This includes array processing techniques which use the cyclic cross correlation. Incorporation of robust estimators should not be a problem because the robust estimators modify only the magnitude, not the phase, of complex data.

As a final note, it is believed that robust estimators of higher order cyclostationarity may prove useful. Higher order statistics are even more sensitive to deviations from gaussian than are second order statistics, and could therefore benefit from robust estimation.

## Acknowledgements

The authors wish to thank E-Systems Greenville Division for supporting publication of this work.

## References

- [1] W.A. Gardner. *Statistical Spectral Analysis: A Nonprobabilistic Theory*. Prentice Hall, Englewood Cliffs, NJ, 1988.
- [2] P.J. Huber. *Robust Statistical Procedures*. SIAM Press, Philadelphia, PA, 1977.
- [3] F.R. Hampel, E.M. Ronchetti, P.J. Rousseeuw, and W.A. Stahel. *Robust Statistics: The Approach based on Influence Functions*. Wiley & Sons, New York, 1986.
- [4] R.G. Staudte and S.J. Sheather. *Robust Estimation and Testing*. Wiley & Sons, New York, 1990.
- [5] S.A. Kassam and H.V. Poor. Robust techniques for signal processing: A survey. *Proc. of the IEEE*, 73(3):433-481, March 1985.
- [6] L. Izzo, L. Paura, and M. Tanda. Signal interception in non-gaussian noise. *IEEE Trans. Commun.*, 40(6):1030-1037, June 1992.