FAST MAXIMUM-LIKELIHOOD SEA CLUTTER PARAMETER LEARNING FROM THE OUTPUT OF THE ENVELOPE DETECTOR

Faruk SARI[1], *Nurşen SARI*[1] , *Lamine MILI*[2]

1:TUBITAK, MRC, ITRI, Gebze, Kocaeli, TURKEY. e-mail: {faruk.sari, nursen.sari}@bte.mam.gov.tr
2: Virginia Tech, Alexandria Research Institute, Alexandria, VA, 22314, USA. e-mail: lmili@vt.edu

ABSTRACT

We develop a fast learning technique to estimate the background statistics parameters from the output of the envelope detector, the inputs of which are multi-component Gaussian Mixture (GM) distributions. We use Fisher Scoring (FS) algorithm, which is Newton based and has fast convergence properties, to solve the log-likelihood minimization problem. Experimental results are given on real radar clutter data.

1. INTRODUCTION

The understanding and modelling of radar clutter plays an essential role for radar system design and performance evaluation. Sea clutter is modelled as a stochastic process composed of random scattering from the sea surface. The radar clutter statistics vary widely and generally different types of clutter are modelled using different types of distributions. Consequently, clutter cannot be considered as homogenous. For instance, high-resolution radar systems and low resolution radars at high sea states have target-like spikes that give rise to non-Guassian heavy tailed observations [1] . It is easily seen from IPIX real radar data [2] that the probability distribution of low sea state data has shorter tails than those of the Normal distribution [3] . By contrast, the probability distribution of high sea state data has longer tails than those of the normal distribution. While traditional radar detector is designed to operate against Gaussian noise, new detection processors are required to reduce the effects of the spikes and to improve detection performance. Target detection requires the comparison of the square magnitude with a certain threshold. It is extremely important to maintain a constant false alarm rate (CFAR) when the background noise level fluctuates. Most probably, future advanced radar systems will be able to detect, identify, and estimate the parameters of a target in severe interference backgrounds [4] . Design of adaptive radar detection algorithms requires that the parameters should be learned from the operational environment [5] . Learning of the clutter parameters has been proposed, recently [6] .

In a previous work, we proposed to use a generalized distribution, which is valid for different background statistics like low and high sea states, in order to model the sea clutter [3] . A generalized GM probability density function (pdf) with zero mean was used in order to form a statistical model of the clutter for each of the in-phase (*I*) and quadrature-phase (*Q*) channels. New background pdf resembled Rician Mixtures. The derivation of the envelope detector output, which has multi component GM inputs, and the Maximum Likelihood (ML) estimation of this output had not been addressed before. In addition, the estimation of the mixtures parameters required to calculate the detection threshold in CFAR was investigated for the first time. Parameters of the background statistics were formulated as a ML problem and the log-likelihood minimization problem was solved by employing an iterative Gradient descent optimization. It is known that gradient descent algorithm has slow convergence properties. Therefore, gradient descent methods are inefficient and second-order information is essential for fast convergence.

In this work, we propose to use the Fisher Scoring (FS) optimization method to estimate the unknown sea clutter parameters, efficiently. Experimental results are presented using real sea clutter data which was collected by IPIX radar at low and high sea states.

2. CLUTTER MODELING

The pdf's of the *I* and *Q* channels, which are the inputs of the envelope detector, are denoted by *p(x)* and *p(y),* respectively.

$$p(x)=\sum_{j=1}^{J}\alpha_j\frac{1}{\left(2\pi(\sigma_x^2)_j\right)^{1/2}}\exp\left[-\frac{x^2}{2(\sigma_x^2)_j}\right] \qquad (1)$$

$$p(y)=\sum_{k=1}^{K}\beta_k\frac{1}{\left(2\pi(\sigma_y^2)_k\right)^{1/2}}\exp\left[-\frac{y^2}{2(\sigma_y^2)_k}\right] \qquad (2)$$

J and *K* denote number of components of *x* and *y*. α_j and β_k are the mixture ratios, where $\alpha_j \geq 0$, $\sum_{j=1}^{J}\alpha_j = 1$ and $\beta_k \geq 0$, $\sum_{k=1}^{K}\beta_k = 1$. $(\sigma_x^2)_j$ and $(\sigma_y^2)_k$ are the variances associated with *I* and *Q* channels, respectively. For convenience, it is assumed that $(\sigma_x^2)_j$ and $(\sigma_y^2)_k$ are ordered such that $(\sigma_x^2)_1 < (\sigma_x^2)_2 < .... < (\sigma_x^2)_J$ and $(\sigma_y^2)_1 < (\sigma_y^2)_2 < .... < (\sigma_y^2)_K$. When it is assumed that the in-phase and quadrature-phase components of clutter are independent, the output of the envelope detector is given by,

$$p(r)=\sum_{j=1}^{J}\sum_{k=1}^{K}\frac{\alpha_j\beta_k r}{\left((\sigma_x^2)_j(\sigma_y^2)_k\right)^{1/2}}\exp\left[-A_{jk}r^2\right]I_0\left(B_{jk}r^2\right) \quad (3)$$

where I_n is the n^{th} order Modified Bessel function of the first kind [3] . A_{jk} and B_{jk} are respectively defined as:

$$A_{jk}=\frac{1}{4}\left(\frac{1}{(\sigma_x^2)_j}+\frac{1}{(\sigma_y^2)_k}\right),\quad B_{jk}=\frac{1}{4}abs\left(\frac{1}{(\sigma_y^2)_k}-\frac{1}{(\sigma_x^2)_j}\right) \quad (4)$$

Note that despite some similarities, the above pdf *p(r)* is very different from the Rician Mixtures.

3. FAST ML ESTIMATION

Adaptive threshold setting problem can be specified as the estimation of the parameter vector $\phi\phi$, defined as $\phi=\left(\alpha_j,\beta_k,\sigma_x^2,\sigma_y^2\right)^T$ from the output of the envelope detector. Adaptive CFAR threshold is set according to $T=f\left(P_{fa},\alpha_j,\beta_k,\left(\sigma_x^2\right)_j,\left(\sigma_y^2\right)_k\right)$ and it is based on estimating the mixture ratios and variances. The ML-CFAR algorithm is implemented as shown in Figure 1, by setting the adaptive threshold T [7] .

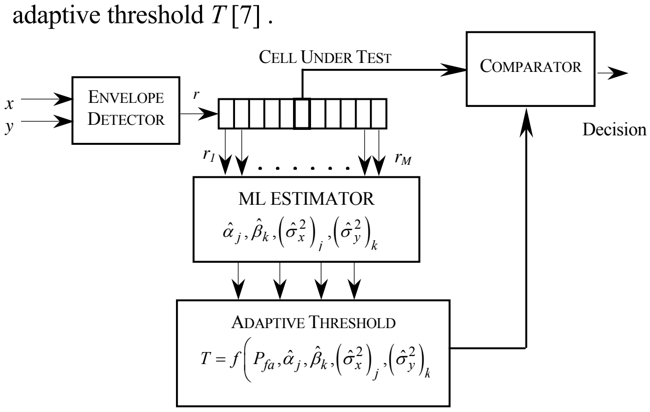

Figure 1. ML-CFAR structure.

It is well known that under some conditions, ML method results in efficient estimates for the unknown parameters of a given pdf. The ML estimator of the parameter vector ϕ is defined by $\hat{\phi}_{ML}=\arg\left\{\max_{\phi\in\Theta}L(\phi)\right\}$; here $L(\phi)$ denotes the (log)-likelihood function. The maximization problem can be expressed as a minimization problem by multiplying $L(\phi)$ by (-1). The ML estimates of unknown parameters are obtained from the M background samples. From M independent observations, the log-likelihood function is given by,

$$-L(\phi)=-\log\prod_{m=1}^{M}p(r_m;\phi)=-\sum_{m=1}^{M}\log p(r_m;\phi) \quad (5)$$

The conventional way to find the parameters involves taking the partial derivatives of the likelihood function with respect to the required parameters, and equating the results to zero. Since Eq. (5) does not have a closed form solution, numerical optimization techniques are used. Gradient descent methods are extremely inefficient especially when the number of design variables is large, so second-order information is essential for fast convergence [8] . A method to overcome this difficulty is the Newton based method where the Hessian is computed. However, generally in many practical situations, the Hessian matrix cannot be computed or stored, easily. The FS algorithm is identical to the Newton-Raphson algorithm with one exception. The FS algorithm replaces Hessian matrix by its expectation. The Fisher Information Matrix (FIM), $I_f(\phi)$ is defined by [9] ,

$$\left[I_f(\phi)\right]_{i,j}=E\left[\frac{\partial L(\phi)}{\partial\phi_i}\frac{\partial L(\phi)}{\partial\phi_j}\right] \quad (6)$$

To minimize $L(\phi)$, iterative FS algorithm can be expressed as,

$$\phi^{i+1}=\phi^i+\eta_\phi I_f^{-1}(\phi)\frac{\partial L(\phi)}{\partial\phi_i} \quad (7)$$

where, η_ϕ is the convergence parameter.

3.1. Gradient of the Log-likelihood Function

The gradient of the log-likelihood function should be calculated with respect to each unknown parameter. The required gradients are defined in the following equations and detailed formulations can be found in [3] .

$$\frac{\partial L(\phi)}{\partial\phi_i}=-\sum_{m=1}^{M}\frac{\dfrac{\partial p(R_m)}{\partial\phi_i}}{p(R_m)} \quad (8)$$

3.2. FIM of the Log-likelihood Function

The parameters $\alpha_j,\beta_k,\left(\sigma_x^2\right)_j,\left(\sigma_y^2\right)_k$ are independent from each other, so the FIM will have a block diagonal form as follows:

$$I_f(\phi)=\begin{bmatrix} I_{mn}(\underline{\alpha}) & 0 & 0 & 0 \\ 0 & I_{mn}(\underline{\beta}) & 0 & 0 \\ 0 & 0 & I_{mn}(\underline{\sigma}_x^2) & 0 \\ 0 & 0 & 0 & I_{mn}(\underline{\sigma}_y^2) \end{bmatrix} \quad (9)$$

where $\underline{\alpha},\underline{\beta},\underline{\sigma}_x^2,\underline{\sigma}_y^2$ are the vector form of the unknown parameters. The matrices $I_{mn}(\underline{\alpha})$, $I_{mn}(\underline{\beta})$, $I_{mn}(\underline{\sigma}_x^2)$ and $I_{mn}(\underline{\sigma}_y^2)$ are the corresponding FIM's and are defined as:

$$I_{mn}(\underline{\theta})=\begin{bmatrix} I_{11}(\theta) & . & \dots & I_{1n}(\theta) \\ : & : & \dots & : \\ : & : & \dots & : \\ I_{m1}(\theta) & . & \dots & I_{mn}(\theta) \end{bmatrix} \quad (10)$$

The FIM is defined in Eq. (11). Notice that the entries of the FIM need to be computed with numerical integration.

$$I_{ij}(\alpha) = E\left[\frac{\partial L(\phi)}{\partial \alpha_i}\frac{\partial L(\phi)}{\partial \alpha_j}\right] = \int_0^\infty \left(\frac{\partial L(\phi)}{\partial \alpha_i}\frac{\partial L(\phi)}{\partial \alpha_j}\right) p(r)dr \quad (11)$$

The proposed algorithm can be summarized as follows:

1. ***Initialization:*** *Begin with an initial guess of* $\alpha_j^{(0)}, \beta_k^{(0)}, \left(\sigma_x^2\right)_j^{(0)}, \left(\sigma_y^2\right)_k^{(0)}$.
2. ***FS Optimization:*** *On iteration i,*
   a. *Calculate the gradient of the log-likelihood function.*
   b. *Calculate the FIM.*
   c. *Use the FS and find the following* $\alpha_j^{(i+1)}, \beta_k^{(i+1)}, \left(\sigma_x^2\right)_j^{(i+1)}, \left(\sigma_y^2\right)_k^{(i+1)}$.
3. ***Termination :*** *Repeat the above steps until the desired accuracy is reached.*

4. EXPERIMENTAL RESULTS

Experimental results are obtained from the real sea clutter data collected at high and low sea states with McMaster University IPIX radar. For simplification, we assume that *I* and *Q* channels have the same mixture ratios and the same variances, that is: $i)\ \sigma_n^2 = \left(\sigma_x^2\right)_n = \left(\sigma_y^2\right)_n$ $ii)\ \alpha_n = \beta_n$. So it is enough to estimate σ_n^2 and α_n. Using a GM with 3 components, we have obtained satisfied fits to the IPIX radar data at low and high sea states. Estimated values are given in Table 1.

Table 1. Estimated sea clutter parameters.

		α_1	α_2	α_3	σ_1^2	σ_2^2	σ_3^2	Ite-ra-tion #
	Initial Val.	0.333	0.333	0.333	1	2	3	
Gradient	High Sea State	0.48	0.4	0.12	0.36	1.27	2.28	134
	Low Sea State	0.19	0.49	0.32	1.87	3.2	5.89	
FS	High Sea State	0.54	0.31	0.15	0.32	1.74	2.19	7
	Low Sea State	0.12	0.47	0.41	1.95	3.95	4.9	

The estimation results obtained using Gradient based and FS methods are similar but Gradient based method needs more iteration than the FS method. The pdf of the real data $p(r)$ and the estimated pdf obtained using FS method at low and high sea states are illustrated in Figure 2.

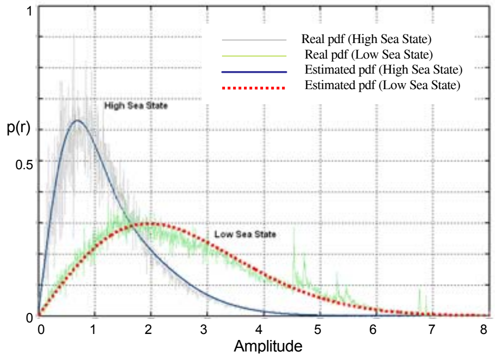

Figure 2. Real and estimated pdfs.

For validation, we compute the empirical Cumulative Distribution Function (CDF) of the real data and the estimated CDF, given as in Figure 3.

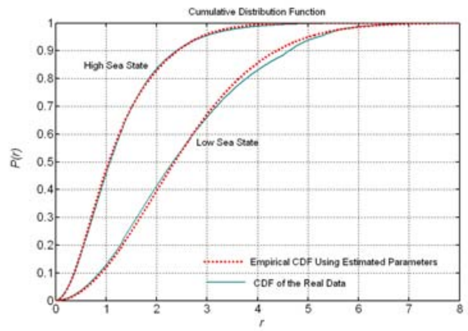

Figure 3. Estimated and real CDFs.

Kolmogorov-Smirnov (KS) goodness of fit test and Kullback Leibler (KL) distance are performed to check if the empirical distribution and the theoretical distribution are close enough. A smaller value of the KS statistic indicates a better fit of the particular distribution to the empirical data. For example, for a significance level 0.01, the KS value should be less than 0.0576 for a sample size of 800 [10] . The KL distance is a natural distance function from a true probability distribution $p = \{p_1, p_2, ..., p_n\}$, to a target probability distribution $q = \{q_1, q_2, ..., q_n\}$; and is defined to be $KL(p\|q) = \sum_{i=1}^{m} p_i \ln\left(\frac{p_i}{q_i}\right)$, where *m* is the number of levels of the variables [11] . At the end of the iterations of FS algorithm, KL distance and KS test results are given in Table 2 for low and high sea states. KS and KL plots at each iteration are given in Figure 4 and 5, respectively.

Table 2. KS test results and KL distance

	Sea States	KS	KL
Gradient	Low	0,0213	0,0409
	High	0,0174	0,0453
FS	Low	0,0281	0,0387
	High	0.0167	0,0460

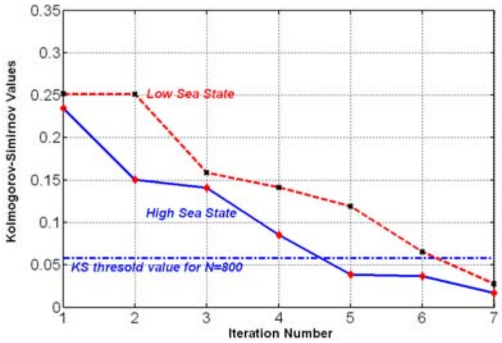

Figure 4. KS values at each iteration.

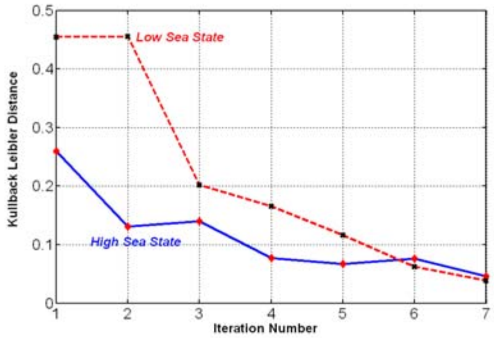

Figure 5. KL distance at each iteration.

The probability of false alarm $\left(P_{fa}\right)$ is a function of the mixture ratios, variances and threshold T. Using estimated values, $\left(P_{fa}\right)$ is calculated as for $A_{jk} > B_{jk} > 0$,

$$P_{fa} = \sum_{j=1}^{J}\sum_{k=1}^{K} \hat{\alpha}_j \hat{\beta}_k - \sum_{j=1}^{J}\sum_{k=1}^{K} \frac{\hat{\alpha}_j \hat{\beta}_k}{2\left((\hat{\sigma}_x^2)_j (\hat{\sigma}_y^2)_k\right)^{1/2}} \left(\int_0^{T^2} e^{-\left(A_{jk}\right)u} I_0\left(\left(B_{jk}\right)u\right) du \right) \tag{12}$$

Since, the integral in the above equation cannot be calculated analytically, it is evaluated numerically. $\left(P_{fa}\right)$ versus (T) plot is given in Figure 6. As expected, P_{fa} is monotonically decreasing with T.

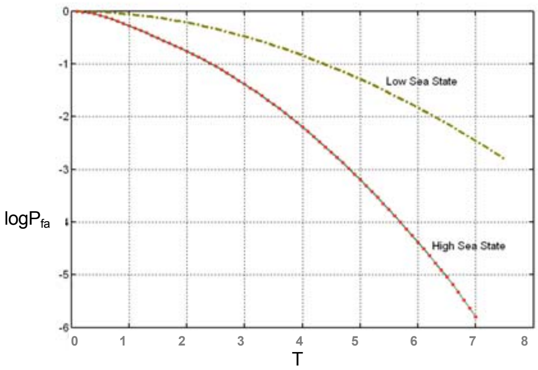

Figure 6. P_{fa} versus T using estimated parameters.

5. CONCLUSION

Multi-component GM distributions at envelope detector inputs are very convenient for modeling different types of sea clutter. The parameters of the background statistics are learned using a ML algorithm with FS optimization, efficiently. Quite satisfied fits are obtained using the ML method for estimating the parameters. It can be seen from experimental results that the empirical pdf from estimated values fits the real data histograms very well. In a few steps, FS optimization reaches the acceptable KS values while Gradient descent method needs much more iterations. Additionally, it is possible to avoid complicated Hessian equations by using only the gradient values in FS optimization. The computers with more computational power may lead to more widespread use of the GM distribution in radar practice.

Acknowledgment

We would like to thank Dr. Simon Haykin for providing the sea clutter data from IPIX radar database.

REFERENCES

[1] F. Gini, M. Greco, “Covariance Matrix Estimation for CFAR Detection in Correlated Heavy Tailed Clutter”, *Signal Processing*, vol. 82, pp.1847-1859, 2002.

[2] http://soma.ece.mcmaster.ca/ipix/

[3] F. Sarı, N. Sarı, L. Mili, “Learning of the Sea Clutter Parameters for Adaptive Detection” in Proc. Int. Conf. on Radar Systems RADAR 2004, Toulouse, France, October 18-22. 2004.

[4] P. Tsakalides, R. Raspanti, C. L. Nikias, “Angle/Doppler Estimation in Heavy-Tailed Clutter Backgrounds”, *IEEE Trans. On Aerospace and Electronic Systems,* vol. 35, no. 2, pp. 419-436, 1999.

[5] S. Haykin, R. Bakker, B. Currie, “Uncovering Nonlinear Dynamics: The Case Study of Sea Clutter”, *Proc. IEEE*, Vol. 90, no. 5, pp. 860-861, 2002.

[6] S. Haykin, “Adaptive Radar”, in *Proc. IEEE Int. Conf. On Acoustics, Speech and Signal Processing*, Orlando, Florida, USA, May 13 - 17, 2002.

[7] R.Ravid, N. Levanon, “Maximum-Likelihood CFAR for Weibull Background”, *IEE Proceedings-F*, vol. 139, no. 3, pp. 256-264,1992.

[8] F. Sarı, “Newton Type Optimization for ML Blur Identification and Restoration”, in *Proc. IEEE Int. Conf. On Acoustics, Speech and Signal Processing*, Orlando, Florida, USA, May 13 - 17, 2002, pp. 3716-3719.

[9] S.M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Printice Hall, 1993.

[10] W.J. Conover, *Practical Nonparametric Statistics*, New York: Wiley, 1980.

[11] T.M. Cover, J.A. Thomas, *Elements of Information Theory*, Wiley Interscience, New York, NY, 1991.