NRL Report 6997

Trimmed-Mean Detector for Noncoherent Distributions

G. V. TRUNK

Radar Analysis Staff
Radar Division

December 11, 1969

NAVAL RESEARCH LABORATORY
Washington, D.C.

This document has been approved for public release and sale; its distribution is unlimited.

CONTENTS

Abstract	ii
Problem Status	ii
Authorization	ii
INTRODUCTION	1
TRIMMED-MEAN DETECTOR	1
MONTE CARLO METHOD	3
THRESHOLDS	5
DETECTION PROBABILITIES	7
COMMENTS	7
CONCLUSIONS	10
ACKNOWLEDGMENT	11
REFERENCES	11

ABSTRACT

Sea-clutter measurements using high-resolution radar indicate that clutter cross-section returns do not usually follow a Rayleigh distribution. The log-normal and contaminated-normal descriptions of sea clutter have been considered, and detection curves for these distributions have been generated for both the mean and median detectors — the median being better for the log-normal distribution and the mean being slightly better for the contaminated-normal. In this report, the trimmed-mean detector is proposed as a detector against these clutter distributions. By using a Monte Carlo technique, equations are obtained for the threshold values, and curves of probabilities of detection (P_d) are drawn for false-alarm probabilities (P_{fa}) ranging from 10^{-2} to 10^{-9}. When one compares the mean, median, and trimmed-mean detectors, the conclusion is that the trimmed-mean is the better detector against either of the two distributions considered. In fact, for the log-normal distribution, when 10 samples are available, the trimmed-mean requires 5 dB less S/N ratio than the mean in order to obtain $P_d = 0.9$ and $P_{fa} = 10^{-8}$. Unfortunately, whereas the mean and median are easily implemented, the implementation of the trimmed-mean is very difficult because of the necessity of ordering the samples.

PROBLEM STATUS

This is an interim report on one phase of the problem; work on the problem is continuing.

AUTHORIZATION

NLR Problem R02-54
Project RF 05-151-402-4010

Manuscript submitted October 2, 1969.

TRIMMED-MEAN DETECTOR FOR NONCOHERENT DISTRIBUTIONS

INTRODUCTION

Sea-clutter measurements made by NRL (1) using a high-resolution X-band radar show that clutter cross section is not Rayleigh distributed. George (2) and Trunk (3) have considered the log-normal and contaminated-normal descriptions of sea clutter, respectively, and found that both of these distributions yield better statistical models of sea clutter than the Rayleigh. Since both of these distributions have longer tails than the Rayleigh, Trunk (4) compared the mean and median detectors and showed that the median is better for the log-normal distribution and the mean is slightly better for the contaminated-normal distribution. Since the median uses only one sample while the mean uses all samples, it was thought possible that a detector which uses a sizable fraction of the available samples might be more effective than either the mean or median. Such a detector is the trimmed-mean detector.

TRIMMED-MEAN DETECTOR

Let $X_1, X_2, \ldots, X_n$ be independent samples taken from the probability density $f(X)$, and let $F(X)$ be the distribution function of $f(X)$. Let the order statistics be

$$X_{r_1} \le X_{r_2} \le X_{r_3} \le \ldots \le X_{r_n} ,$$

where X_{r_k} is called the kth-order statistic, is denoted by $y_k{}^n$, and is the value of the kth smallest observation in the sample of size n from $F(X)$. The density of $y_k{}^n$ is

$$p_{n,k}(x) = \frac{n!}{(n-k)!\,(k-1)!} F(x)^{k-1} [1 - F(x)]^{n-k} f(x) , \tag{1}$$

and the distribution of y_k^n is denoted by $P_{n,k}(x)$.

The trimmed-mean detector is given by

$$T_{n_1,n_2} = \sum_{i=n_1}^{n_2} y_i^n , \tag{2}$$

where $1 \le n_1 \le n_2 \le n$. If $n_1 = 1$ and $n_2 = n$, then T_{n_1,n_2} reduces to the mean detector; and if $n_1 = n_2 = [n/2]$, then T_{n_1,n_2} reduces to the median detector. In order to calculate the density of T_{n_1,n_2}, it is first necessary to calculate the joint density of the sequence $\{y_k{}^n\}$. Now, all joint densities can be written as

$$\begin{aligned} &p(x_1, x_2, \ldots, x_n) = \\ &\quad p(x_n | x_{n-1}, x_{n-2}, \ldots, x_1)\, p(x_{n-1} | x_{n-2}, x_{n-3}, \ldots, x_1) \ldots p(x_2 | x_1)\, p(x_1) . \end{aligned} \tag{3}$$

However, since for ordered samples

$$p(y_k^n \mid y_{k-1}^n, y_{k-2}^n, \ldots, y_1^n) = p(y_k^n \mid y_{k-1}^n) , \tag{4}$$

the joint density of the ordered samples can be written as

$$p\left(y_{n_1}^n, y_{n_1+1}^n, \ldots, y_{n_2}^n\right) = p\left(y_{n_1}^n\right) \prod_{i=n_1+1}^{n_2} p\left(y_i^n \mid y_{i-1}^n\right) \tag{5}$$

The conditional density of y_i^n is simply

$$\begin{aligned} p(y_i^n \mid y_{i-1}^n) &= p_{n,k}(y_i^n)/P_{n,k}(y_{i-1}^n) , \quad && y_i^n \geq y_{i-1}^n \\ &= 0 , && y_i^n < y_{i-1}^n \end{aligned} \tag{6}$$

Therefore, the joint density of $\{y_k^n\}$ is

$$p\left(y_{n_1}^n, y_{n_1+1}^n, \ldots, y_{n_2}^n\right) = \prod_{i=n_1}^{n_2} p_{n,k}\left(y_i^n\right) \Bigg/ \prod_{i=n_1}^{n_2-1} P_{n,k}\left(y_i^n\right) . \tag{7}$$

At this point, one usually makes the transform of variables

$$u_i = y_i^n, \quad i = n_1, \ldots, n_2 - 1 \quad \text{and} \quad T_{n_1,n_2} = \sum_{i=n_1}^{n_2} y_i^n$$

and integrates out the dummy variables $\{u_i\}$. Unfortunately, for the log-normal density,

$$f(x) = \int_0^{2\pi} \frac{x \exp\left[-\frac{2}{\sigma^2}\left(ln\sqrt{x^2 - 2xA\cos\theta + A^2}\right)^2\right] d\theta}{\pi\sqrt{2\pi\sigma^2}(x^2 - 2xA\cos\theta + A^2)} ; \tag{8}$$

and for the contaminated-normal density,

$$\begin{aligned} f(x) = \frac{x}{2\pi\sigma^2} \int_0^{2\pi} & (1-\gamma)^2 \exp\left[-(x^2 - 2xA\cos\theta + A^2)/2\sigma^2\right] \\ & + \frac{\gamma^2}{K^2} \exp\left[-(x^2 - 2xA\cos\theta + A^2)/2K^2\sigma^2\right] \\ & + \frac{\gamma(1-\gamma)}{K} \exp\left[-(x^2\cos^2\theta - 2xA\cos\theta\right. \\ & \left. + A^2 + K^2x^2\sin^2\theta)/2K^2\sigma^2\right] \\ & + \frac{\gamma(1-\gamma)}{K} \exp\left[-(x^2\sin^2\theta - 2xAK^2\cos\theta\right. \\ & \left. + A^2K^2 + K^2X^2\cos^2\theta)/2K^2\sigma^2\right] d\theta \end{aligned} \tag{9}$$

where A is the amplitude of the nonfluctuating signal, T can only be expressed in terms of an $(n_2 - n_1 + 1)$-order integral, to which there exists no known closed-form solution. Therefore, in order to calculate the probabilities of false alarm and detection, it is necessary to perform a $(n_2 - n_1 + 2)$-order numerical integration. For $(n_2 - n_1 + 2) > 3$, this is impossible for present-day computers. Therefore, instead of using this exact numerical approach, it was decided to undertake a Monte Carlo investigation of the problem.

MONTE CARLO METHOD

The basic part of the simulation is the generation of the signal-plus-noise in phase and quadrature components y_i and z_i. Gaussian noise, which is a common building block for both the log-normal and contaminated-normal densities, can be generated by the following equation:

$$g = u \sin(\theta) \tag{10}$$

where u is Rayleigh distributed and θ is uniformly distributed on $(0, 2\pi)$. Since a uniform, random-number generator is available on the computer (it generates a random number r uniformly distributed on $(0, 1)$), the generation of θ simply involves the multiplication of r by 2π. The generation of u is accomplished by the uniform, random-number generator by starting from the Rayleigh distribution

$$P(v \le u) = \int_0^u v e^{-v^2/2}\, dv \tag{11}$$

Integrating, one obtains

$$P(v \le u) = 1 - e^{-u^2/2} \tag{12}$$

or

$$e^{-u^2/2} = 1 - P(v \le u) \ . \tag{13}$$

Now, the quantity $1 - P(v \le u)$ is uniformly distributed between 0 and 1, and, consequently, can be replaced by a random number r. Hence, if one solves Eq. (13) for u,

$$u = (-2 \ln(r))^{1/2} \ , \tag{14}$$

u has a Rayleigh distribution. Therefore, the output of an envelope detector is

$$X_i = \left(y_i^2 + z_i^2\right)^{1/2} \tag{15}$$

where, for the generation of log-normal clutter,

$$y_i = \exp\{\sigma[-2 \ln(r_{i1})]^{1/2} \sin(2\pi r_{i2})\} \sin(2\pi r_{i3}) + A \tag{16}$$

and

$$z_i = \exp\{\sigma[-2 \ln(r_{i1})]^{1/2} \cos(2\pi r_{i2})\} \cos(2\pi r_{i3}) \ , \tag{17}$$

and for the generation of contaminated-normal clutter,

$$y_i = [-2 \ln (r_{i1})]^{1/2} \sin (2\pi r_{i2}) S(r_{i3}) + A , \tag{18}$$

$$z_i = [-2 \ln (r_{i1})]^{1/2} \cos (2\pi r_{i2}) S(r_{i4}) , \tag{19}$$

$$S(r_{ij}) = \begin{cases} K\sigma & r_{ij} \le \gamma \\ \sigma & r_{ij} > \gamma . \end{cases} \tag{20}$$

The sequence $\{r_{ij}\}$ consists of independent numbers uniformly distributed between 0 and 1, and A is the amplitude of the nonfluctuating signal. A single sample of the detector output is obtained by calculating n X_i's, ordering them, and then performing the summation indicated in Eq. (2).

Unfortunately, in order to calculate the required threshold for a probability of false alarm P_{fa} of 10^{-5}, $(10/P_{fa} = 10^6)$ samples are required. For $n = 10$, this calculation required 1 hour of computation time on the CDC 3800. Several Monte Carlo calculations involving 10^6 samples were conducted for the contaminated-normal and log-normal distributions with $n = 10$ and for two different trimmed means. The results are presented in Tables 1 through 4.

Table 1
Contaminated-Normal Distribution ($K = 2.25$ and $\gamma = 0.25$),
$n = 10$, $n_1 = 2$, $n_2 = 5$

T = Threshold	Number of Samples $< T$	T = Threshold	Number of Samples $< T$
5.5	892072	8.75	999630
5.75	924369	9.00	999773
6.00	947725	9.25	999852
6.25	964324	9.50	999918
6.50	976338	9.75	999942
6.75	984509	10.00	999968
7.00	989921	10.25	999979
7.25	993509	10.50	999986
7.50	995879	10.75	999992
7.75	997435	11.00	999998
8.00	998411	11.25	999998
8.25	999005	11.50	999999
8.50	999397	11.75	1000000

Table 2
Contaminated-Normal Distribution (K = 2.25 and γ = 0.25),
n = 10, n_1 = 4, n_2 = 7

T = Threshold	Number of Samples < T	T = Threshold	Number of Samples < T
7.75	906936	12.25	999700
8.00	929243	12.50	999802
8.25	946601	12.75	999863
8.50	959979	12.00	999905
8.75	970373	13.25	999939
9.00	978212	13.50	999956
9.25	983988	13.75	999967
9.50	988347	14.00	999978
9.75	991563	14.25	999986
10.00	993937	14.50	999992
10.25	995616	14.75	999994
10.50	996824	15.00	999998
10.75	997682	15.25	999999
11.00	998330	15.50	999999
11.25	998782	15.75	999999
11.50	999139	16.00	999999
11.75	999414	16.25	999999
12.00	999571	16.50	999999

THRESHOLDS

From the data presented in Tables 1 through 4, one can obtain directly very accurate thresholds for $P_{fa} \geq 10^{-5}$ and thresholds for $10^{-5} \geq P_{fa} \geq 10^{-6}$. In order to obtain thresholds for $P_{fa} < 10^{-6}$, it was decided that a curve would be fitted to the tabular data and then extrapolated to yield the lower P_{fa}'s. That is, a curve of the form

$$-\log P_{fa} = B_0 + B_1 T + B_2 T^2 \tag{21}$$

will be fitted to the tabular data by the method of least squares. If the collection of equations for the tabular data of the type appearing in Eq. (21) is written in matrix notation

$$L = TB \ , \tag{22}$$

the least-square coefficients B are given by the well-known equation

$$B = (\tilde{T}T)^{-1} \tilde{T}L \ . \tag{23}$$

Table 3
Log-Normal Distribution (σ = 6 dB) n = 10, n_1 = 2, n_2 = 5

T = Threshold	Number of Samples < T	T = Threshold	Number of Samples < T
4.00	913155	6.75	999778
4.25	945870	7.00	999875
4.50	967038	7.25	999937
4.75	980443	7.50	999965
5.00	988450	7.75	999984
5.25	993311	8.00	999990
5.50	996130	8.25	999992
5.75	997776	8.50	999996
6.00	998761	8.75	999997
6.25	999318	9.00	999998
6.50	999624	9.25	999999

Table 4
Log-Normal Distribution (σ = 6 dB), n = 10, n_1 = 4, n_2 = 7

T = Threshold	Number of Samples < T	T = Threshold	Number of Samples < T
5.75	916335	9.50	999719
6.00	939848	9.75	999802
6.25	957427	10.00	999864
6.50	969959	10.25	999915
6.75	978976	10.50	999940
7.00	985430	10.75	999958
7.25	990031	11.00	999971
7.50	993256	11.25	999978
7.75	995343	11.50	999984
8.00	996791	11.75	999992
8.25	997845	12.00	999996
8.50	998606	12.25	999996
8.75	999061	12.50	999998
9.00	999363	12.75	999999
9.25	999574	13.00	1000000

Applying Eq. (23) to the tabular data, one obtains the following equations relating the probabilities of false alarm to the corresponding thresholds:

for the contaminated-normal with $n = 10$, $n_1 = 2$, and $n_2 = 5$,

$$-\log P_{fa} = -1.691 + 0.3123\ T + 0.0308\ T^2\ ; \tag{24}$$

for the contaminated-normal with $n = 10$, $n_1 = 4$, and $n_2 = 7$,

$$-\log P_{fa} = -2.505 + 0.3681\ T + 0.0105\ T^2\ ; \tag{25}$$

for the log-normal with $n = 10$, $n_1 = 2$, and $n_2 = 5$,

$$-\log P_{fa} = -3.1735 + 1.0685\ T - 0.0070\ T^2\ ; \tag{26}$$

and for the log-normal with $n = 10$, $n_1 = 4$, and $n_2 = 7$,

$$-\log P_{fa} = -2.084 + 0.4868\ T^2 + 0.0107\ T^2 \tag{27}$$

DETECTION PROBABILITIES

To calculate the detection probabilities, the Monte Carlo procedure was used again. For each signal-to-noise ratio, a trial sample of 5000 was more than adequate to estimate probabilities of detection in the interval 0.1 to 0.99. The Monte Carlo was conducted for the four cases previously discussed, and the results appear in Figs. 1 through 4. For both clutter distributions, the signal-to-noise ratio is defined as

$$S/N = 20 \log (A/\text{median clutter value}) \tag{28}$$

where the median value equals 1 for the log-normal and equals 1.41 σ for the contaminated-normal distribution ($\gamma = 0.25$ and $K = 2.25$).

If one compares Fig. 1 with Fig. 2 and Fig. 3 with Fig. 4, one concludes that the trimmed-mean detector involving the lower ordered samples ($n_1 = 2$ and $n_2 = 5$) is more effective than the one involving the centered samples ($n_1 = 4$ and $n_2 = 7$). If one compares the trimmed-mean and mean (3) detectors for the contaminated-normal distribution, one concludes that the trimmed-mean requires about a 1-dB-lower S/N ratio than the mean. Since the mean is better than the median (4) for the contaminated-normal distribution, the trimmed-mean is the best detector for the contaminated-normal distribution. For the log-normal distribution (2), the trimmed-mean is from 1 to 5 dB better than the mean — the exact value depending on the desired P_{fa}. Since the median is only slightly better than the mean, the trimmed-mean is also the best detector for the log-normal distribution.

COMMENTS

The key step in the generation of detection probabilities was the Monte Carlo involving 10^6 samples used to calculate the required thresholds. For this Monte Carlo to be valid, the 30,000,000 random numbers required for the log-normal case (40,000,000 required for the contaminated-normal case) must be independent and uniformly distributed. To test the distribution of the random numbers, 100,000 numbers were generated. Next,

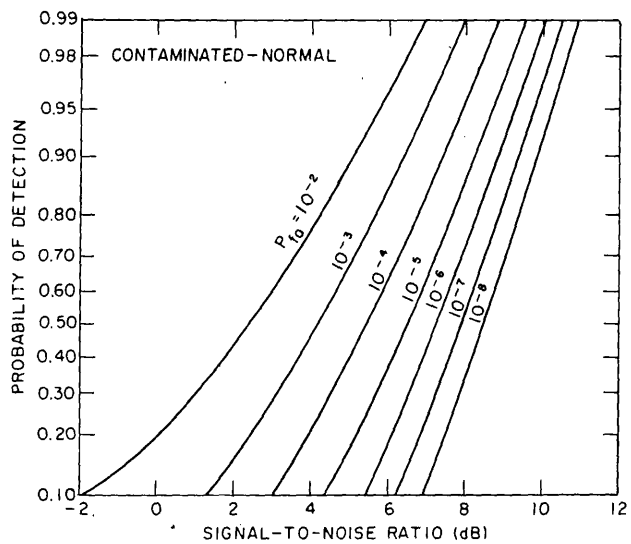

Fig. 1 - Probability of detection for the trimmed-mean detector ($n = 10$, $n_1 = 2$, and $n_2 = 5$) and the contaminated-normal distribution ($\gamma = 0.25$ and $K = 2.25$)

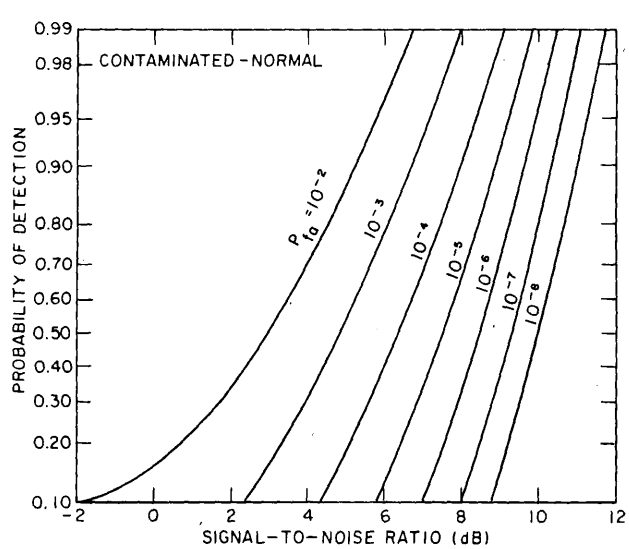

Fig. 2 - Probability of detection for the trimmed-mean detector ($n = 10$, $n_1 = 4$, and $n_2 = 7$) and the contaminated-normal distribution ($\gamma = 0.25$ and $K = 2.25$)

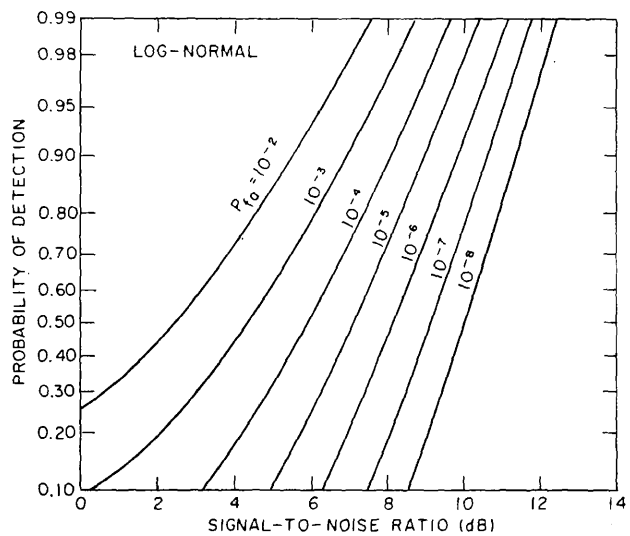

Fig. 3 - Probability of detection for the trimmed-mean detector ($n = 10$, $n_1 = 2$, and $n_2 = 5$) and the log-normal distribution ($\sigma = 6$ dB)

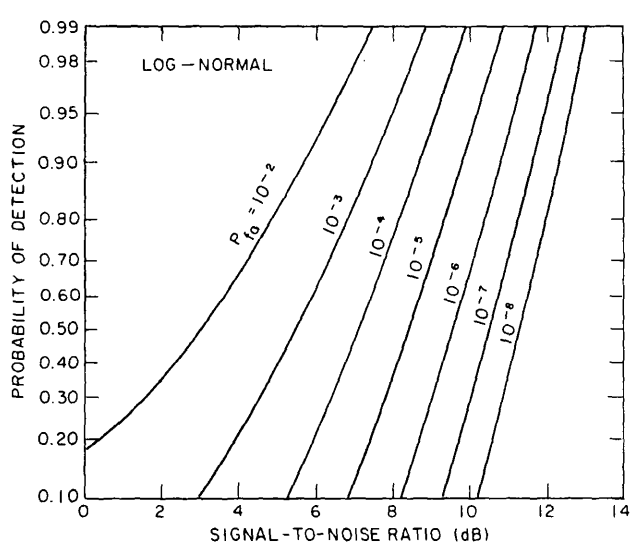

Fig. 4 - Probability of detection for the trimmed-mean detector ($n = 10$, $n_1 = 4$, and $n_2 = 7$) and the log-normal distribution ($\sigma = 6$ dB)

the Kolmogorow-Smirnov one-sample test (5) was used, with the result that the hypothesis that the random numbers were uniformly distributed was accepted very easily even at the 20% significance level. In order to check for independence, 16,384 random numbers were generated; and the fast Fourier transform was used to calculate the correlation function. The samples were found to be uncorrelated. Of course, for uniformly distributed variables, correlation implies nothing about independence. Consequently, Eq. (10) was used to generate Gaussian variables. These variables were found to be uncorrelated and hence independent. Unfortunately, because of computational difficulties, the independence of sample sizes larger than 16,384 could not be checked.

Consequently, in an attempt to reduce the possibility that the random numbers tend to repeat themselves (are dependent), the random-number generator was reinitialized after the generation of every 30,000 random numbers. To investigate this method, the Monte Carlo procedure previously described was used to calculate the median detector for the contaminated-normal distribution. The thresholds were approximated by

$$\log P_{fa} = 0.2293 + 0.1628\ T + 0.0806\ T^2 \tag{29}$$

and the detection probabilities are shown in Fig. 5. If one compares these probabilities to those previously calculated (4), one finds a maximum error of only 0.25 dB. The author is of the opinion that the trimmed-mean curves are not quite this accurate because the trimmed-mean thresholds are a more rapidly changing function of P_{fa} than the median is.

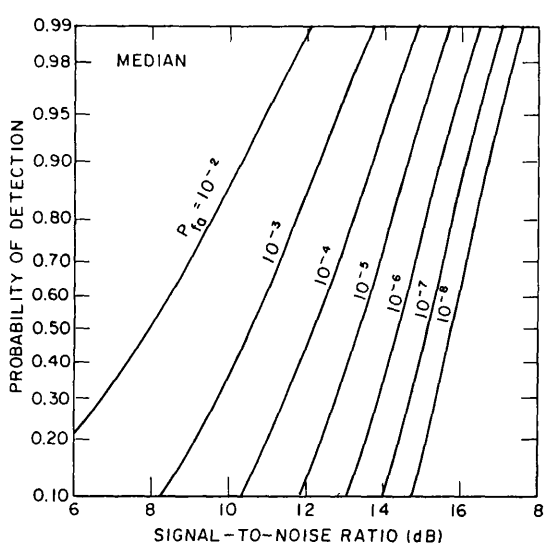

Fig. 5 - Probability of detection for the median detector (n = 3 and k = 2) and the contaminated-normal distribution (γ = 0.25 and K = 2.25)

CONCLUSIONS

Of the three detectors that have been investigated — the mean, the median, and the trimmed-mean — the trimmed-mean is the most effective detector against either the log-normal or contaminated-normal distributions. Unfortunately, whereas the mean and median detectors can be quite simply implemented, the implementation of the trimmed-mean is very difficult because of the necessity of ordering the samples.

The possibility that there may exist large errors in the trimmed-mean calculations — due either to the Monte Carlo with its associated difficulties with a large number of

independent random numbers or to the extrapolation of the Monte Carlo to very small P_{fa}'s — is very small. This is because of the accuracy of the Monte Carlo for the median detector and also because the trimmed-mean calculations show both a self-consistency that is to be expected and an intraconsistency with both the mean and median calculations.

ACKNOWLEDGMENT

The author thanks J. D. Wilson, who helped with various programming problems which arose during this study.

REFERENCES

1. Findlay, A.M., "Sea-Clutter Measurement by Radar-Return Sampling," NRL Report 6661, Feb. 12, 1968

2. George, S.F., "The Detection of Nonfluctuating Targets in Log-Normal Clutter," NRL Report 6796, Oct. 4, 1968

3. Trunk, G.V., "Noncoherent Detection of Nonfluctuating Targets in Contaminated-Normal Clutter," NRL Report 6858, Mar. 21, 1969

4. Trunk, G.V., "Median Detector for Noncoherent Distributions," NRL Report 6898, May 9, 1969

5. Siegel, S., "Nonparametric Statistics for the Behavioral Sciences," New York: McGraw-Hill, 1956

This page intentionally left blank.

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY *(Corporate author)* Naval Research Laboratory Washington, D. C. 20390	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP

3. REPORT TITLE

TRIMMED-MEAN DETECTOR FOR NONCOHERENT DISTRIBUTIONS

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*

An interim report on one phase of the problem; work on the problem is continuing.

5. AUTHOR(S) *(First name, middle initial, last name)*

Gerard V. Trunk

6. REPORT DATE December 11, 1969	7a. TOTAL NO. OF PAGES 16	7b. NO. OF REFS 5
8a. CONTRACT OR GRANT NO. NRL Problem R02-54 b. PROJECT NO. RF 05-151-402-4010 c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) NRL Report 6997 9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)*	

10. DISTRIBUTION STATEMENT

This document has been approved for public release and sale; its distribution is unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY Department of the Navy (Office of Naval Research), Washington, D.C. 20360

13. ABSTRACT

Sea-clutter measurements using high-resolution radar indicate that clutter cross-section returns do not usually follow a Rayleigh distribution. The log-normal and contaminated-normal descriptions of sea clutter have been considered, and detection curves for these distributions have been generated for both the mean and median detectors — the median being better for the log-normal distribution and the mean being slightly better for the contaminated-normal. In this report, the trimmed-mean detector is proposed as a detector against these clutter distributions. By using a Monte Carlo technique, equations are obtained for the threshold values, and curves of probabilities of detection (P_d) are drawn for false-alarm probabilities (P_{fa}) ranging from 10^{-2} to 10^{-9}. When one compares the mean, median, and trimmed-mean detectors, the conclusion is that the trimmed-mean is the better detector against either of the two distributions considered. In fact, for the log-normal distribution, when 10 samples are available, the trimmed-mean requires 5 dB less S/N ratio than the mean in order to obtain $P_d = 0.9$ and $P_{fa} = 10^{-8}$. Unfortunately, whereas the mean and median are easily implemented, the implementation of the trimmed-mean is very difficult because of the necessity of ordering the samples.

DD FORM 1 NOV 65 1473 (PAGE 1)

S/N 0101-807-6801

Security Classification

Security Classification

14. KEY WORDS	LINK A ROLE	LINK A WT	LINK B ROLE	LINK B WT	LINK C ROLE	LINK C WT
Probability of detection						
Sea clutter						
Trimmed-mean detector						
Contaminated-normal distribution						
Log-normal distribution						
Monte Carlo method						

DD FORM 1 NOV 65 1473 (BACK)
(PAGE 2)

Security Classification