Image Cover Sheet

CLASSIFICATION	SYSTEM NUMBER
UNCLASSIFIED	500743

TITLE

A SEARCH PROCEDURE FOR SHIPS IN RADARSAT IMAGERY

System Number:

Patron Number:

Requester:

Notes:

DSIS Use only:

Deliver to:

National Defence / Défense nationale

A SEARCH PROCEDURE FOR SHIPS IN RADARSAT IMAGERY (U)

by

Maria T. Rey, Anastasios Drosopoulos
and Dusan Petrovic

DEFENCE RESEARCH ESTABLISHMENT OTTAWA
REPORT NO. 1305

Canada

December 1996
Ottawa

National Defence / Défense nationale

A SEARCH PROCEDURE FOR SHIPS IN RADARSAT IMAGERY (U)

by

Maria T. Rey, Anastasios Drosopoulos
Airborne Radar and Navigation Section

and

Dusan Petrovic
Thorpe Associates Ltd.

DEFENCE RESEARCH ESTABLISHMENT OTTAWA
REPORT NO. 1305

PROJECT
SEB116

December 1996
Ottawa

Abstract

This report describes the Constant False Alarm Rate, CFAR, procedure applied to the Ocean Features Workstation, OFW, based on the K distributed sea clutter model. A number of scenes from ERS-1 and RADARSAT are examined and the goodness-of-fit of the K distribution model is investigated, using both visual and standard statistical tests. Improved performance in detecting ship targets in comparison with the previous method on the OFW is observed.

Résumé

Le présent rapport expose le processus de maintien d'un taux de fausses alertes constant (TFAC) que nous avons mis en pratique à la Station de travail sur les particularités océaniques (STPO). Ce processus est fondé sur le modèle de fouillis de mer à distribution K. Nous avons examiné plusieurs scènes ERS-1 et RADARSAT pour nous pencher sur la validité de l'ajustement, en employant les tests statistiques visuel et normalisé, ce qui nous a permis d'observer, par rapport à la performance de l'ancienne méthode, une amélioration en matière de détection des navires-cibles.

Executive Summary

The Ocean Features Workstation, OFW, is a RADARSAT ground terminal, developed jointly by the Department of Fisheries and Oceans, DFO, the Canada Centre for Remote Sensing, CCRS, the Department of National Defence, DND, and the Canadian Space Agency, CSA, to facilitate the detection and tracking of ocean surface features, such as ships, ship wakes, icebergs, and wave currents. DND is primarily interested in the ship detection component, as the main objective of wide area ocean surveillance using Synthetic Aperture Radar, SAR.

This report describes the Constant False Alarm Rate, CFAR, ship detection procedure under developement for the Ocean Features Workstation, based on the K distributed sea clutter model, from SAR images. A number of scenes from ERS-1 and RADARSAT are examined and the goodness-of-fit of the sea clutter model is investigated.

The previously implemented method at the OFW was a simple thresholding within $\pm 5\sigma$ about the mean signal level. While adequate for ERS-1 type data and ship wake detection, this method suffers in performance when applied to the higher resolution RADARSAT data for the purpose of ship target detection. The CFAR technique as described in this report has improved ship target detection performance. It should however be noted that only a small number of preliminary RADARSAT scenes were available for examination at the present time, and a more general, or simpler model may be more appropriate for the tasks that the OFW will perform in the future. An on-going investigation, as more images are received, will resolve these issues.

Contents

Abstract iii

Executive Summary v

Contents vii

List of Tables ix

List of Figures xii

1 Introduction 1

2 Statistics of SAR sea scenes 3

2.1 The Clutter Distribution 3

2.2 Numerical Convergence of the CFAR 5

2.3 Parameter Estimators Based on Moments 8

2.4 Limiting Factors to the CFAR Aproach 10

2.5 RADARSAT Signal to Noise Performance 11

3 Estimation of goodness-of-fit 12

3.1 The χ^2 test 12

3.2 The Kolmogorov-Smirnov test 13

4 Data results 15

4.1 Choice of parameter estimators 15

4.2 Evaluation of Kolmogorov-Smirnov and χ^2 goodness-of-fit tests 26

5 Incidence angle dependence 28

6 Design of RADARSAT CFAR 31

6.1 Results of CFAR procedure 31

7 Conclusions and future work 34

A Simulating K distributed clutter 35

References 37

List of Tables

1 Blacknell's ν estimates using sample mean and data variance (first ν value), sample mean and mean of log (second ν value), mean of log and variance of log (third ν value) on some ERS-1 datasets. The results of the χ^2 and Kolmogorov-Smirnov tests for each case are also shown. 25

2 Results for the CFAR Technique with Varying Threshold on a variety of 31875 pixel size images. 33

List of Figures

1 Error in shape parameter estimate vs theoretical shape parameter for number of looks $L = 1, 2, 4, 8$. The solid line denotes the estimator based on the variance of the data and the dashed line on the estimator based on the log of the mean of the data. 17

2 Error in shape parameter estimate vs theoretical shape parameter for number of looks $L = 16$ and $L = 2.9$ (top graphs). The solid line denotes the estimator based on the variance of the data and the dashed line on the estimator based on the log of the mean of the data. The cross-over point as a function of L is shown at the bottom graph. 18

3 Img03dec94 with histogram and model pdf as well as data and model cdf's. . 19

4 Img04dec94 with histogram and model pdf as well as data and model cdf's. . 19

5 Img11dec91 with histogram and model pdf as well as data and model cdf's. . 20

6 Img12dec91 with histogram and model pdf as well as data and model cdf's. . 20

7 Img14dec91 with histogram and model pdf as well as data and model cdf's. . 21

8 Img15dec91 with histogram and model pdf as well as data and model cdf's. . 21

9 Img17dec91 with histogram and model pdf as well as data and model cdf's. . 22

10 Img18dec91 with histogram and model pdf as well as data and model cdf's. . 22

11 Img20dec91 with histogram and model pdf as well as data and model cdf's. . 23

12 Img21dec91 with histogram and model pdf as well as data and model cdf's. . 23

13 Img23dec91 with histogram and model pdf as well as data and model cdf's. . 24

14 Img24dec91 with histogram and model pdf as well as data and model cdf's. . 24

15 Pearson plots of kurtosis vs skewness squared for small regions of large RADARSAT images. The solid line is the theoretical model for the normalized K distribution for a variety of shape parameters ν. 27

16 Mean power level as a function of range offset (across swath) for varying window size. The last figure shows the variance in the shape parameter for the mean and mean of the log of the data estimator for the maria4 dataset. . 29

17 CFAR vs μ for a variety of fixed values of ν, L, and I_c. 30

18 Flow diagram of CFAR Ship Search Procedure. 32

1 Introduction

The ocean area of interest to the Department of National Defence (DND) is vast from a ship detection perspective. With the arrival of the new RADARSAT Synthetic Aperture Radar (SAR) sensor, which was launched on November 4, 1995, a new technology exists which potentially enables the Canadian Forces (CF) to survey large regions of ocean on an opportunistic basis[6]. SAR is an imaging sensor, generating large quantities of data which must be inspected, either by people or automatically, for the presence of ships and other ocean features of interest to the CF. Therefore the detection problem must be limited to one of minimizing the amount of SAR imagery to be searched, as well as adapting the detection criteria to the changing ocean clutter statistics. The first part, that of minimizing the amount of imagery to be searched, is achieved through the coordinated division of the area under surveillance among the three government agencies with mandates for oceam surveillance, DND, the Department of Fisheries and Oceans (DFO) and the Canada Centre for Remote Sensing (CCRS). Work on the second part, that of developing a better CFAR technique for target detection in ocean clutter, is addressed in this report.

The Department of Fisheries and Oceans (DFO), in partnership with the Canada Centre for Remote Sensing (CCRS), the Department of National Defence (DND) and the Canadian Space Agency (CSA), has funded the development of an Ocean Features Workstation, which includes as a subset of its functionality, procedures for ship wake detection in SAR imagery, which were developed by DND at the Defence Research Establishment Ottawa (DREO) [15, 16].

The Ocean Features Workstation (OFW)[17] is a Sun Sparc20 based workstation designed for the automatic detection and classification of ocean features in RADARSAT imagery[15]. To date, the principal features detected by the workstation are ships and associated wake features. These algorithms are predominantly ship wake detection algorithms with a simple CFAR for the ship detection based upon sampling of the mean and standard deviation of the ocean area being searched. The statistical assumptions on both ship and clutter parameters were deduced empirically for the SEASAT SAR sensor.

The current Ocean Features Workstation which resides in Gatineau has as its predominant ship detection algorithms the wake detection procedures developed by DREO. The fundamental flaw in this is that the wake detection algorithms are just that, wake detection algorithms, and are not intended to provide optimum ship detection capability. They are intended to be centred on a detected (suspected) ship location. The current algorithms merely apply a statistical threshold to the ocean, and then search around all pixels that pass the threshold to perform the actual wake detection. This procedure produces a great number of false alarms, and more robust ship detection and wake centering algorithms need to be explored.

To develop predominantly ship detection algorithms as opposed to wake detection ones, a different philosophy than that assumed in the DREO wake algorithms must be assumed. Experience with the visibility of ships and wakes in SEASAT imagery indicates that the feature of the highest probability of detection is the ship, not the wake. The wake can be used as ancillary information, if it exists, to enhance the detection of a ship and to provide information on course and speed.

A second, less robust technique could be applied to the ship search procedure for RADARSAT data. This would comprise improving the current threshold to simply reflect the expected signal to clutter for a ship of a given length via techniques similar to those used in Skolnik[19] and [20] or other empirical estimators, and adapted to the measured or estimated clutter background (current environmental conditions). This technique may result in a great many false alarms over a RADARSAT frame, but may be less dangerous than a CFAR based upon poor distribution estimation due to inadequate clutter statistics.

The largest potential problem with the Skolnik equation estimate is that while it is reasonable to assume that the equation holds for military vessels, at coarse resolutions, it hasn't been tested for merchant vessels, whose RCS/foot is the result of a very different superstructure, and hence the draught relationship is not necessarily valid. Also, at higher resolutions, aspect dependency and variability in superstructure are the prime considerations. Nevertheless, for RADARSAT, it is a technique which is definitely worth exploring, as if it can be validated, it will be very simple to apply.

After some debate between the various Government departments developing the OFW, it was decided to modify the wake detection algorithms to include two approaches for ship detection:

1. A Constant False Alarm Rate (CFAR) approach based on sampled ocean statistics and expected ship Radar Cross Section (RCS) and
2. a variable false alarm technique based on estimation of ambient sea state and expected target length.

This report details the development of the CFAR technique for the OFW at DREO.

2 Statistics of SAR sea scenes

CCRS has provided us with large volumes of "featureless" ERS-1 imagery, as well as currently available RADARSAT data to test out our algorithms.

Several authors[3, 21, 23] have shown the applicability of modelling the amplitude probability distribution of radar returns from the ocean via the compound K-distribution, independent of position and time. In fact, the expected SAR ocean clutter returns, under differing sea conditions, will likely vary through several classes of distributions, and this choice of compound K distribution may not ultimately be the only fit that must be tested on the data.

2.1 The Clutter Distribution

The compound K distribution model is an empirical model for sea clutter that currently enjoys wide success. It is applicable to high resolution radars that are capable of resolving fine structure on the sea surface. The received sea clutter signal by such a system is not well modelled by a Gaussian process. The K model on the other hand, successfully does so, based on the following observations:

- The time history of the envelope of a demodulated signal from an individual range cell shows fast fluctuations from pulse to pulse that is modulated by an underlying structure.
- The fast fluctuation component, or "speckle", decorrelates from pulse to pulse with frequency agile signals but is correlated for between 5 and 10 ms with fixed frequency signals. This implies a large number of scatterers within each illuminated patch. By applying the central limit theorem, the "speckle" can be approximated by the Rayleigh distribution.
- The slower modulating component is the local sea clutter mean level. This has a long temporal decorrelation period and is not affected by frequency agility. It characterizes the mean level variation of clutter, including clutter "spikes", and its intensity is found to be gamma distributed.
- An attempt to connect the K model with the scattering physics is described in [5, 10, 22] where a model of scatterer bunching is postulated, caused by a birth-death-migration process of scatterers. The gamma distribution is found in this model to be the first term in a Laguerre series expansion of the modulation pdf. The implication is that the non-Gaussian nature of sea clutter seems to be due more to scatterer bunching by the sea wave structure than anything else.

Based on the above, the overall clutter distribution, $p(\chi)$ for the amplitude χ of a resolution cell in a single-look SAR image of the ocean can be modelled as:

$$p(\chi) = \int_0^\infty p(\chi|\psi)p(\psi)d\psi, \qquad 0 \leq \chi \leq \infty \tag{1}$$

where the "speckle" component is Rayleigh distributed

$$p(\chi|\psi) = \frac{\pi\chi}{2\psi^2}\exp\left(-\frac{\pi\chi^2}{4\psi^2}\right) \tag{2}$$

and the mean value ψ has a chi or root gamma distribution

$$p(\psi) = \frac{2b^{2\nu}\psi^{2\nu-1}}{\Gamma(\nu)}\exp(-b^2\psi^2), \qquad 0 \leq \psi \leq \infty \tag{3}$$

where b is a scale parameter.

It should be noted that the above distribution parameters are very sensitive to both the operating radar system parameters and environmental conditions.

Substituting (2) and (3) into (1) and using the relation 2.3.16.1 from [14] yields

$$p(\chi) = \frac{4c}{\Gamma(\nu)}(c\chi)^\nu K_{\nu-1}(2c\chi), \qquad \nu > 0 \tag{4}$$

where $K_\nu(\chi)$ is the modified Bessel function K, $c = b\sqrt{\pi/4}$ is a scale parameter and ν is a shape parameter. For spikier clutter, visible at higher resolutions the shape parameter $\nu \to 0$. Large values of ν give shapes closer to Rayleigh with $\nu \to \infty$ representing pure thermal noise. Hou[7] shows that the scale parameter is related to the average clutter power μ through the relationship $c^2 = \nu/\mu$.

When radar imagery is being produced, an available option is to produce multiple images (looks) of the same scene, which are averaged to reduce speckle variance, while leaving the underlying mean cross-section unchanged. In so doing, the number of independent data samples that are produced is unchanged. If the number of looks L is taken into account in the determination of the K-distribution, the conditional probability density function of the speckle amplitude given in (2) changes to the conditional probability density function of the sum of L independent looks: $p(\chi_1 + \chi_2 + ... + \chi_L)$. Blacknell's formulation [3] of the K distribution, based on the power of the data and L number of looks is therefore more appropriate for our purposes.

The radar cross section y of each clutter cell is now distributed as [3]

$$p(y) = \left(\frac{\nu}{\mu}\right)^\nu \frac{y^{\nu-1}}{\Gamma(\nu)}\exp\left(-\frac{\nu y}{\mu}\right) \tag{5}$$

where μ is the mean radar cross-section of each clutter cell. The speckle component is

$$p(x|y) = \left(\frac{L}{y}\right)^L \frac{x^{L-1}}{\Gamma(L)} \exp\left(-\frac{Lx}{y}\right) \tag{6}$$

The combined radar signal power x's pdf becomes then

$$p(x) = \int_0^\infty p(x|y)p(y)dy = \frac{2}{x\Gamma(\nu)\Gamma(L)} \left(\frac{L\nu x}{\mu}\right)^{(L+\nu)/2} K_{L-\nu}\left(2\sqrt{\frac{L\nu x}{\mu}}\right) \tag{7}$$

where the symmetry relation $K_\alpha(x) = K_{-\alpha}(x)$ is employed. The relations of Blacknell's power based parameters to the amplitude ones given earlier, are:

$$\{\nu\}_{amplitude} = \{\nu\}_{power}$$

$$\{b^2\}_{amplitude} = \left\{\frac{\nu}{\mu}\right\}_{power}$$

As L tends to infinity, the distribution of the radar signal tends to the distribution of the radar cross-section, which is to be expected as the variance of the data tends to zero.

2.2 Numerical Convergence of the CFAR

Selection of an appropriate CFAR power threshold I_c requires the computation of:

$$\eta_c = \int_0^{I_c} p(x)dx \tag{8}$$

where η_c represents the required significance level and I_c is the required (user selected) threshold.

The required false alarm rate is therefore:

$$CFAR = 1 - \int_0^{I_c} p(x)dx \tag{9}$$

Using the Mathematica[24] software package, the analytical solution of (8) where $p(x)$ is given by (7) is

$$\eta_c = \frac{\nu(L\nu z/\mu)^L\Gamma(-L+\nu)\ {}_1F_2(L;1+L,1+L-\nu;L\nu z/\mu)}{L\nu\Gamma(L)\Gamma(\nu)} + \\ + \frac{L(L\nu z/\mu)^\nu\Gamma(L-\nu)\ {}_1F_2(L;1+\nu,1-L+\nu;L\nu z/\mu)}{L\nu\Gamma(L)\Gamma(\nu)} \tag{10}$$

where $z = I_c$ and the generalized hypergeometric functions are defined as[14]

$$ {}_pF_q(a_1, a_2, \ldots a_p; b_1, b_2, \ldots b_q; z) \equiv \sum_{k=0}^{\infty} \frac{(a_1)_k (a_2)_k \ldots (a_p)_k}{(b_1)_k (b_2)_k \ldots (b_q)_k} \frac{z^k}{k!} $$

and

$$ (a)_k \equiv a(a+1)\ldots(a+k-1) = \frac{\Gamma(a+k)}{\Gamma(a)} $$

is the Pochhammer symbol.

Equation (10) represents a complete analytical solution to the cumulative probability density function. The problem with this solution is that the computational accuracy required for convergence of the hypergeometric functions is very large. Another analytical solution, valid for all real L, can be derived through the following manner:

The expression to be evaluated [substituting (7) into (8)] is of the form:

$$ \eta_c = A \int_0^t x^\lambda K_\alpha(x) dx \tag{11} $$

where

$$ t = 2\sqrt{\frac{L\nu I_c}{y}} $$

$$ A = \frac{4}{2^{L+\nu}} \frac{1}{\Gamma(L)\Gamma(\nu)} \tag{12} $$

$$ \alpha = \nu - L \qquad \text{and} \qquad \lambda = \nu + L - 1 $$

and ν is an initial estimate.

This expression can be put into the generalized form:

$$ \eta_c = A[F(t) - F(0)] \tag{13} $$

From [14, vol II, pg. 47, 1.12.1-1] and the definition of a Lommel function [14, vol II, pg. 747] we can show that:

$$ \int x^\lambda K_\alpha(x) dx \equiv \frac{\lambda+\alpha-1}{(\lambda-\alpha+1)(\lambda+\alpha-1)} x^{\lambda+1} K_\alpha(x) {}_1F_2\left(1; \frac{\lambda-\alpha+3}{2}, \frac{\lambda+\alpha+1}{2}; \frac{x^2}{4}\right) + $$

$$ + \frac{1}{(\lambda-\alpha+1)(\lambda+\alpha+1)} x^{\lambda+2} K_{\alpha-1}(x) {}_1F_2\left(1; \frac{\lambda-\alpha+3}{2}, \frac{\lambda+\alpha+3}{2}; \frac{x^2}{4}\right) \tag{14} $$

The above formula can also be derived from (10). The difference in form between equations (10) and (14) is that (14) avoids the computational overflow that arises in equation (10) because of the $(-L+\nu)$ and the $(L-\nu)$ terms. Whenever these values are greater than one, one or the other converges very slowly, causing computational overflow.

To check if equation (14) is computable for all $x \geq 0$ (numerically and with required precision), we will first explore the behaviour at $x = 0$. We know from the definition of ${}_1F_2$ type hypergeometric functions that:

$$\lim_{x\to 0} {}_1F_2(a_1; b_1, b_2; x) = 1$$

Therefore, both hypergeometric functions in (14) tend to 1 and we are left with the limits:

$$\lim_{x\to 0} x^{\lambda+1} K_\alpha(x) \qquad \lim_{x\to 0} x^{\lambda+2} K_{\alpha-1}(x)$$

For small x we can use the expansion:

$$K_\alpha(x) \approx \left(\frac{x}{2}\right)^{-\alpha} \frac{(\alpha-1)!}{2} \qquad \alpha > 0$$

and the above limits equal 0 as x tends to 0 for all values $\nu > 0$ and $L > 0$. Therefore, $F(0) = 0$ in in equation (13) which then reduces to

$$\eta_c = AF(t) \tag{15}$$

where A is given in equation (12) and $F(t)$ can be computed via equation (14).

The terms of equation (14) are all positive and the function is bounded in $[0, 1]$ for $t \geq 0$. Therefore the equation is well behaved as all terms are monotone and bounded. If we examine the hypergeometric terms, which are of form

$${}_1F_2(1; a, b; x^2/4)$$

their series expansion becomes:

$$\sum_{k=0}^{\infty} \frac{(x^2/4)^k}{(a)_k (b)_k} \equiv \sum_{k=0}^{\infty} f_k \tag{16}$$

where $f_0 = 1$.

Although series (16) is known to converge and is therefore continuous and bounded, it is not computationally well behaved for all parameter values of a and b. We can determine the exact relationship between a, b and k at f_k maximum, from the following ratio:

$$\frac{f_{k+1}}{f_k} = \frac{(a)_k (b)_k}{(a)_{k+1} (b)_{k+1}} \left(\frac{x^2}{4}\right) = \frac{1}{(a+k)(b+k)} \left(\frac{x^2}{4}\right) \equiv 1 \tag{17}$$

This occurs for:

$$k^2 + (a+b)k + ab - \frac{x^2}{4} = 0$$

and since we are interested in $k > 0$, we use

$$k = \frac{-(a+b) + \sqrt{(a+b)^2 - 4ab + x^2}}{2}$$

By solving equation (17) for k, it becomes clear that depending on the choice of parameters a and b the maximum value of f_k can become enormous leading to computational overflow even with double precision arithmetic. Therefore a careful choice of a and b has to be made such that f_k starts decreasing from the beginning of the series, i.e. $k = 1$. This choice is such that

$$(a-b)^2 + x^2 = (a+b)^2 \Rightarrow x^2 = 4ab$$

where a and b are from (14) and $x \equiv t$.

$$a = \frac{\lambda - \alpha + 3}{2} \qquad b = \frac{\lambda + \alpha + 1}{2}$$

$$a = \frac{\lambda - \alpha + 3}{2} \qquad b = \frac{\lambda + \alpha + 3}{2}$$

Solving for t for both of these cases we have:

$$(\lambda + 3)^2 - \alpha^2 = t^2$$

$$(\lambda - 2)^2 - (\alpha - 1)^2 = t^2$$

These equations can then be used to determine the area in the α, λ plane that can be used to compute the Lommel function, and hence the recurrence relations of equations 11.3.3-11.3.5 in [1] can be used to recursively solve for appropriate α, λ in order to refine the ν estimate so that eq. (9) is well behaved.

2.3 Parameter Estimators Based on Moments

A standard approach to parameter estimation is to use the Maximum Likelihood (ML) solution, which provides the optimum parameter estimates in the sense that these estimates are the most probable parameter values, given the data samples used. Unfortunately for the K-distribution, the ML solution is computationally very intensive and impractical to apply. Therefore the solution must be obtained numerically or an alternative estimation scheme must be developed. Blacknell[3] looks at three approaches for parameter estimators and compares their performances vs. the ML approach. The first approach is to use the first two moments of the distribution (mean and variance) of the data and equate them with their equivalent theoretical expressions. The second approach is to use the mean and variance of the log of the data and the third one uses the mean of the data and the mean of the log of the data. All three approaches were investigated for ERS-1 data.

The nth order moment of a random variable x with a given probability density function $f(x)$ is defined as

$$m_n = \int_{-\infty}^{\infty} x^n f(x) dx$$

For a given data sequence, $\{x_i\}_{i=1}^{M}$, binned into N bins, the sample histogram can be used to approximate them via

$$\hat{m}_n = \sum_{i=1}^{N} x_i^n h(x_i) \Delta x$$

where the hat indicates an estimated value, the area under the histogram is scaled to 1 and Δx is the bin width.

The theoretical expressions for the sample mean and variance of data, assumed to come from the L-look K distribution are given by

$$\mathsf{E}[x] = \mu$$

$$\mathsf{Var}(x) = \left[\left(1 + \frac{1}{\nu}\right)\left(1 + \frac{1}{L}\right) - 1\right]\mu^2$$

where E is the expectation and Var the variance operators. Estimates for the mean and shape parameter using the first approach of Blacknell can be obtained from the equations:

$$\hat{\mu} = \frac{1}{M}\sum_{i=1}^{M} x_i \tag{18}$$

$$\left(1 + \frac{1}{\hat{\nu}}\right)\left(1 + \frac{1}{L}\right) = \frac{\frac{1}{M}\sum_{i=1}^{M} x_i^2}{\left(\frac{1}{M}\sum_{i=1}^{M} x_i\right)^2} \tag{19}$$

Therefore, if moment estimates are produced from the histogram of the sampled data from some ocean area, these estimates can be used to approximate the form of the density function and the cumulative probability for the sample area, through the sample mean and variance. This information can then be used to set a Constant False Alarm Rate (CFAR) threshold for the detection of ships in that area of ocean.

For the second parameter estimation approach, using the mean and variance of the natural log of the data, the theoretical expressions for these two quantities are[3]:

$$\mathsf{E}[\ln(x)] = \ln(\mu) + [\psi(\nu) - \ln(\nu)] + [\psi(L) - \ln(L)]$$

$$\mathsf{Var}[\ln(x)] = \psi^{(1)}(\nu) + \psi^{(1)}(L)$$

where $\psi(x)$ is the digamma function and the superscript (n) denotes the nth derivative. Estimates for the mean and order parameter are given by[3, 9] as:

$$\ln(\hat{\mu}) = \frac{1}{M}\sum_{i=1}^{M} \ln(x_i) - [\psi(\hat{\nu}) - \ln(\hat{\nu})] + [\psi(L) - \ln(L)]$$

$$\psi^{(1)}(\hat{\nu}) = \frac{1}{M}\sum_{i=1}^{M}\ln(x_i^2) - \frac{1}{M}\sum_{i=1}^{M}[\ln(x_i)]^2 - \psi^{(1)}(L)$$

Finally, the third estimation scheme considered by Blacknell uses the sample mean of the data and the sample mean of the log of the data to obtain the parameter estimates. This choice of moments was derived using an ML analysis of the L-look, K-distribution, originally derived by Oliver[11]. The theoretical expressions for these quantities, lead to an order parameter estimator that requires the solution of:

$$\hat{\mu} = \frac{1}{M}\sum_{i=1}^{M}x_i \tag{20}$$

$$\ln(\hat{\nu}) - \psi(\hat{\nu}) = \ln\left(\frac{1}{M}\sum_{i=1}^{M}x_i\right) - \frac{1}{M}\sum_{i=1}^{M}\ln(x_i) + \psi(L) - \ln(L) \tag{21}$$

for the estimate of ν.

2.4 Limiting Factors to the CFAR Aproach

Although simple in theory, in practice, in order to set the same false alarm rate, regardless of ocean background conditions, the CFAR threshold must be adaptive to the change in mean backscatter that occurs with incidence angle and the nonstationarity due to differing ocean regions. Satellite SAR swath widths are large, and there is expected to be a large change in mean backscatter, at HH polarization, relative to incidence angle for the various RADARSAT operational modes. Clearly, choice of sub swath width for true adaptive CFAR techniques using RADARSAT must be tailored in such a way as to minimize the change in mean backscatter across the subswath.

The Gatineau ground station produces "frames" of data available to the user community. Each of these frames corresponds to 64K of sample points. A rule of thumb, derived from practical experience with ERS-1 data, is that stationarity can be assumed over no more than a 10km×10km region of the ocean, in deep sea. If we make the assumption that this subregion must form the basis of the search procedure, we can probably evaluate an area of approximately 1000 samples (resolution cells) in size. Note that RADARSAT is not quite adequately sampled, and therefore there is no need to take into account the resolution cell/pixel ratio in order to avoid possible correlation problems.

Areas larger than 1000 samples in size will have an increased likelihood of exhibiting nonstationarity effects. On the other hand, nonstationarity can exist in areas considerably smaller that 1000 pixels in extent as well. This will be particularly the case in coastal regions. If CFAR techniques are to be employed, considerable methodical investigation of the parameter estimate "goodness of fit" relative to choice of sample size must be carried

out in order to determine whether such techniques are worthwhile for the ship detection case. These algorithms were coded at DREO, and a systematic investigation of the statistics of SAR sea clutter variation with sample size, look number and RADARSAT and ERS-1 subswath position was carried out.

2.5 RADARSAT Signal to Noise Performance

Another fundamental problem which must be addressed and investigated in the ship detection problem, is the target to clutter and noise performance across a RADARSAT swath. As a general trend, at HH polarization, clutter to noise ratios decrease with distance from nadir (shallower depression angles). This means that target to clutter ratios are maximized in the far swath regions, and hence ship detection performance is improved. For wakes however, which are fundamentally clutter returns modulated by the passing ship, best detection performance is obtained in the near swath, where clutter returns are highest. The implication of this is that optimum ship detection region of the RADARSAT swath is not the optimum region for performing wake detection, and vice versa. A detailed investigation using RADARSAT data should be carried out to investigate these issues, and develop empirical estimators for the decision-based processes used in the wake detection algorithms. This is beyond the scope of the present report, but will be addressed in future work.

For some RADARSAT operational modes, the swath width is wide enough that at the far edge of the swath the signal returns are likely to be predominantly noise limited, not clutter limited. Watts[23] has done an investigation of radar detection prediction performance in K-Distributed sea clutter and thermal noise, for single-look imagery. The assumption made in his paper is that in the presence of thermal noise, the speckle component of the return is effectively modified by an increase in its average power.

3 Estimation of goodness-of-fit

Regardless of the choice of estimator used in the determination of the mean and shape parameter for the K-distribution, these estimates necessitate that a test be employed to determine the goodness of fit of sample probability distribution to the estimated theoretical probability distribution fit. Two types of goodness-of-fit estimators were investigated, the χ^2 test and the Kolmogorov-Smirnov test.

3.1 The χ^2 test

The χ^2 variable is used to test how closely a set of observed frequencies corresponds to a given set of expected frequencies based upon some theoretical probability distribution.

The χ^2 statistic for the goodness of fit is given by:

$$\chi^2 = \sum_{i=1}^{N} \frac{(O_i - E_i)^2}{E_i} \tag{22}$$

where there are N histogram bins, $(N > 1)$, $E_1, ..., E_N$ are the expected frequency values in each bin, based on the assumed probability model, while $O_1, ..., O_N$ are the observed frequencies. This statistic approximates a χ^2 variable with $n = N - 1$ degrees of freedom, if the assumed probability model does not require any parameters to be estimated from the data. If it does, then the number of degrees of freedom n is decreased by the number of estimated parameters.

The χ^2 test measures the goodness of fit between the observed and theoretical values as follows: When the fit is good (O_i and E_i generally close), then the numerator in equation (22) will be generally small, and hence the value of the χ^2 test will be low. Conversely, when O_i and E_i are not similar, the value of χ^2 statistic will be large.

The K-distribution has a long tail, and in a sample set, some histogram bins may have no data samples represented. This presents a problem for the χ^2 goodness of fit test for data bins with few elements. For these situations, equation (22) is calculated over all bins with expected frequencies greater than five (a commonly applied rule-of-thumb). The remaining bins are lumped together, with lumped E_i the sum of the individual E_i's, and added as the last entry into the summation. The number of bins N is also modified appropriately. An additional two degrees of freedom is subtracted, since we use the data to estimate the two parameters, ν and μ.

The major problem with this approach is that in these cases, the number of degrees of freedom is not constant, and therefore direct comparison of goodness-of-fit results cannot be performed. To avoid this problem, the confidence level is computed, based on the χ^2 value

and the number of degrees of freedom obtained for a given fit. The confidence levels can then be used to do a comparison of the χ^2 results.

Another fundamental problem with the χ^2 test is that the results are heavily reliant, due to summation, on the main body of the distribution as opposed to the tails of the distribution. This makes it difficult to accurately determine the shape parameter, as several choices in parameter may exhibit the same goodness-of-fit results at the tail.

3.2 The Kolmogorov-Smirnov test

The χ^2 test uses the nominal statistical relationships between the sampled and theoretical distributions. A goodness-of-fit test which uses ordinal relationships on unbinned data, the Kolmogorov-Smirnov (KS) test, was also applied to the data, in an attempt to see if an ordinal test would emphasize the contributions of the tails of the distributions more reliably than the χ^2 test, which is biased by the main body of the distribution, as explained above.

The KS test is applicable to distributions that are functions of a single independent variable, where each data point can be associated with a single number. For these cases, the list of data points can be converted to an unbiased estimator $S_M(x)$ of the cumulative distribution function of the distribution from which the sample was drawn, where M events are located at values x_i, $i = 1, .., M$.

Different distribution functions give different cumulative distribution function estimates if ordered by the above procedure. However, all cumulative distribution functions agree at the smallest and largest allowable values of x (where they become respectively 0 and 1). So it is the behaviour between the two extremes that is of interest in distinguishing distributions.

The KS test looks at the maximum difference between the sampled and theoretical (assumed) cumulative distributions, by using the following statistic:

$$D = \max_{-\infty<x<\infty} |P(x) - S_M(x)|$$

where $P(x)$ and $S_M(x)$ are the relative cumulative probabilities of the theoretical and sample distributions respectively, at data value x. This test is not a binned test and does not suffer from the limitations of the χ^2 test. In fact, the KS test tends to be most sensitive around the median value, where $P(x) = 0.5$, and less sensitive at the extreme ends of the distribution where $P(x)$ is near 0 or 1. Note that the distribution median is more sensitive to long distribution tails than the mean, which is what the χ^2 test depends on.

What makes the KS statistic useful is that the distribution of D can be calculated, at least in the case of the null hypothesis (that data sets are drawn from the same distribution), to a useful approximation, thus giving the significance of the observed nonzero value of D.

From [13], the significance level of a given D value can be derived from:

$$Q_{KS}(\lambda) = 2\sum_{j=1}^{\infty}(-1)^{j-1}e^{-2j^2\lambda^2}$$

which is a monotonic function with limiting values:

$$Q_{KS}(0) = 1 \qquad \text{and} \qquad Q_{KS}(\infty) = 0$$

In terms of this function, the significance level of an observed value of D (as a disproof of the null hypothesis that the distributions are the same) is given approximately by:

$$Probability(D > observed) = Q_{KS}\left([\sqrt{M} + 0.12 + 0.11/\sqrt{M}]D\right) \tag{23}$$

where M is the number of sample points.

Note that if the assumed probability model requires the estimation of some parameters from the data there is no assurance that the statistic D follows the above probability [13]. In this case, we generate a synthetic dataset from a K model with the estimated parameters (Appendix A) and compare the two. We are then guaranteed that D follows (23).

4 Data results

The previous sections have dealt with the theory applied to this research effort. The methodology employed in the study was to evaluate the K-distribution goodness-of-fit, for ERS-1 and RADARSAT ocean data, as a function of: choice of parameter estimates, number of looks, number of histogram bins, independence of data samples and choice of goodness-of-fit estimate, in order to investigate the importance of these parameters in the choice of an appropriate CFAR approach to sampling RADARSAT data. ERS-1 was used simply to "pad up" the data set, as for the bulk of this investigation, RADARSAT data was not yet available to the user community. The final requirement is an accurate estimate of the expected ship RCS, so that the expected false alarm rate could be determined.

4.1 Choice of parameter estimators

Given that in general, estimates of the parameters of a K-distribution will only be required when the values of those parameters are not known, it is reasonable to ask which moment-based estimation scheme will give the best performance on the broadest range of possible shape parameters. Blacknell[3] performs a theoretical analysis of the errors associated with each of the parameter estimation approaches described in section 2.3. He computes the error in the mean and shape parameter estimates and compares this error with the ML parameter estimate approach for the K-distribution. As a result of this comparison, he concludes that the approach based on the calculation of the mean of the data and the mean of the log of the data, is generally the most accurate when applied to single look data, except when $\nu \geq 8$, where estimates based on the mean and variance of the data produce parameter estimates closer to the ML results. Blacknell does not do a complete evaluation of the multilook case, to see if similar results hold for multilook data parameter estimaton. Therefore, an investigation was carried out in house, to compute the error in the parameter estimate, using estimates based on the mean of the data and the mean of the log of the data, and estimates based on the mean of the data and the variance of the data, against the theoretical shape parameter for file sizes of 256×256 samples, and $L = 1, 2, 4, 8$ and 16.

The estimate using the mean and variance of the log of the data was shown by Blacknell to be generally a poorer fit to the distributions than the other two choices, therefore it was not evaluated.

Oliver[12] has calculated the normalized variance in the shape parameter to be:

$$\frac{\mathrm{Var}(\hat{\nu})}{\nu^2} = \frac{1}{M}\frac{2\nu^2 L}{(1+L)}\left(1+\frac{1}{\nu}\right)\left(\frac{1}{\nu}+\frac{1}{L}+\frac{3}{\nu L}\right)\left(\frac{1}{\nu}+\frac{1}{L}+\frac{4}{\nu L}\right) \tag{24}$$

for the estimator based on the mean and variance of the data.

For the estimator based on the mean and the log of the mean, the normalized variance in the shape parameter becomes[12]:

$$\frac{\mathrm{Var}(\hat{\nu})}{\nu^2} = \frac{[\nu\psi^{(1)}(\nu) - 1] + \nu\psi^{(1)}(L) + (1-\nu)/L}{M\nu[\nu\psi^{(1)}(\nu) - 1]^2} \qquad (25)$$

where $\psi^{(1)}$ is the first derivative of the digamma function.

Figures 1-2 show the plots of the errors in the shape parameter estimates from (24) and (25) against the theoretical shape parameter for files of 256 data samples, for several integer values of L. In order to investigate the general L case (non-integer), samples of ERS-1 data, at $L = 2.9$, were analyzed, using the two estimators in Blacknell, for comparison with the theoretical results. The error parameter estimate plot for this case is shown in Fig. 2.

It is apparent that as in the single-look case, the estimation (25) of shape parameters based on the mean and the log of the mean of the data generally produce a better estimate of the shape parameter than the mean and variance estimate (24), except for high values of shape parameter, at which there is a cross-over and the mean and variance estimate is the better one.

The cross-over point appears to be related to look number, as is evident from the diagrams. Theoretically, the cross-over point relationship can be derived by subtracting equation (25) from equation (24), and solving for ν as a function of L. Figure 2 plots this cross-over value of ν against L for $L = 1$ to 16.

The χ^2 goodness-of-fit test was calculated for the compound K distributions having each of the shape parameter estimates, and the confidence level associated with each of the goodness-of-fit values was also calculated. Table 1 shows the results for some ERS-1 images (Figures 3-14) with $L = 2.9$. As discussed above, the χ^2 values were determined by collapsing all bins with expected frequencies less than 5 into a single final bin. Starting initially from 100 bins, the number of degrees of freedom is reduced appropriately to the value shown in the table. Due to hardware quantization, the number of distinct data samples, n_d is significantly less than the total. This is taken into account when computing the maximum distance D between data and model cdf and the confidence level Q_{KS} from the Kolmogorov-Smirnov test.

An investigation of Table 1 shows a poor χ^2 confidence level, regardless of choice of parameter estimator, for all 12 files tested. Despite the good visual fit of histogram and theoretical pdf, the χ^2 value is large enough that the test rejects the hypothesis that the data are K distributed. The theoretical cross-over value of ν, where the variance estimator performs better than the log of mean, is $\nu = 19$. The results however (computed χ^2 value), indicate that $\nu \sim 10$ is probably more appropriate, although more datasets would have to be examined for a firmer conclusion. Even then, the difference is minute. This is likely due to the lack of enough variability in the shape parameter of the data. However, the theoretical

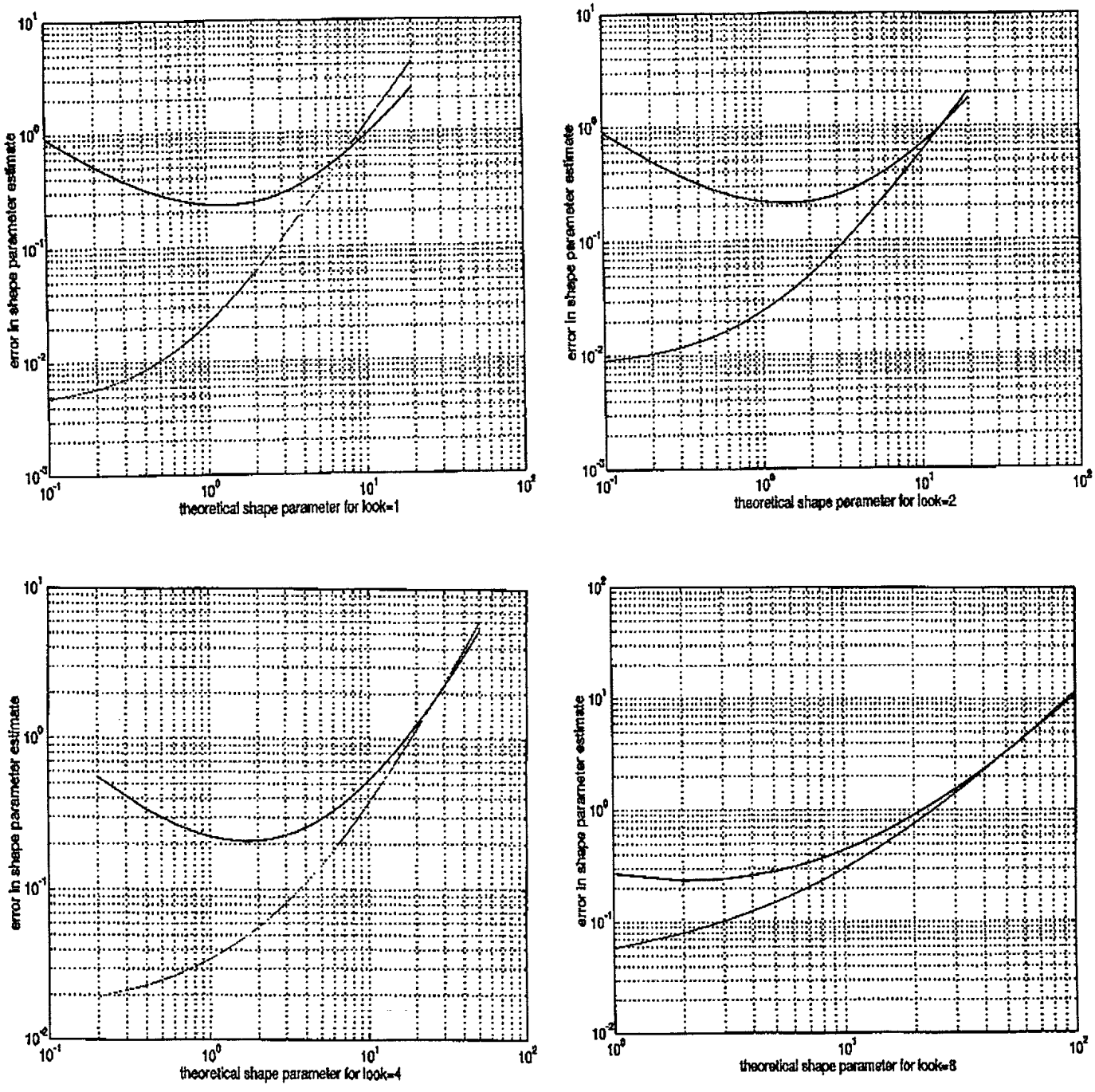

Figure 1: Error in shape parameter estimate vs theoretical shape parameter for number of looks $L = 1, 2, 4, 8$. The solid line denotes the estimator based on the variance of the data and the dashed line on the estimator based on the log of the mean of the data.

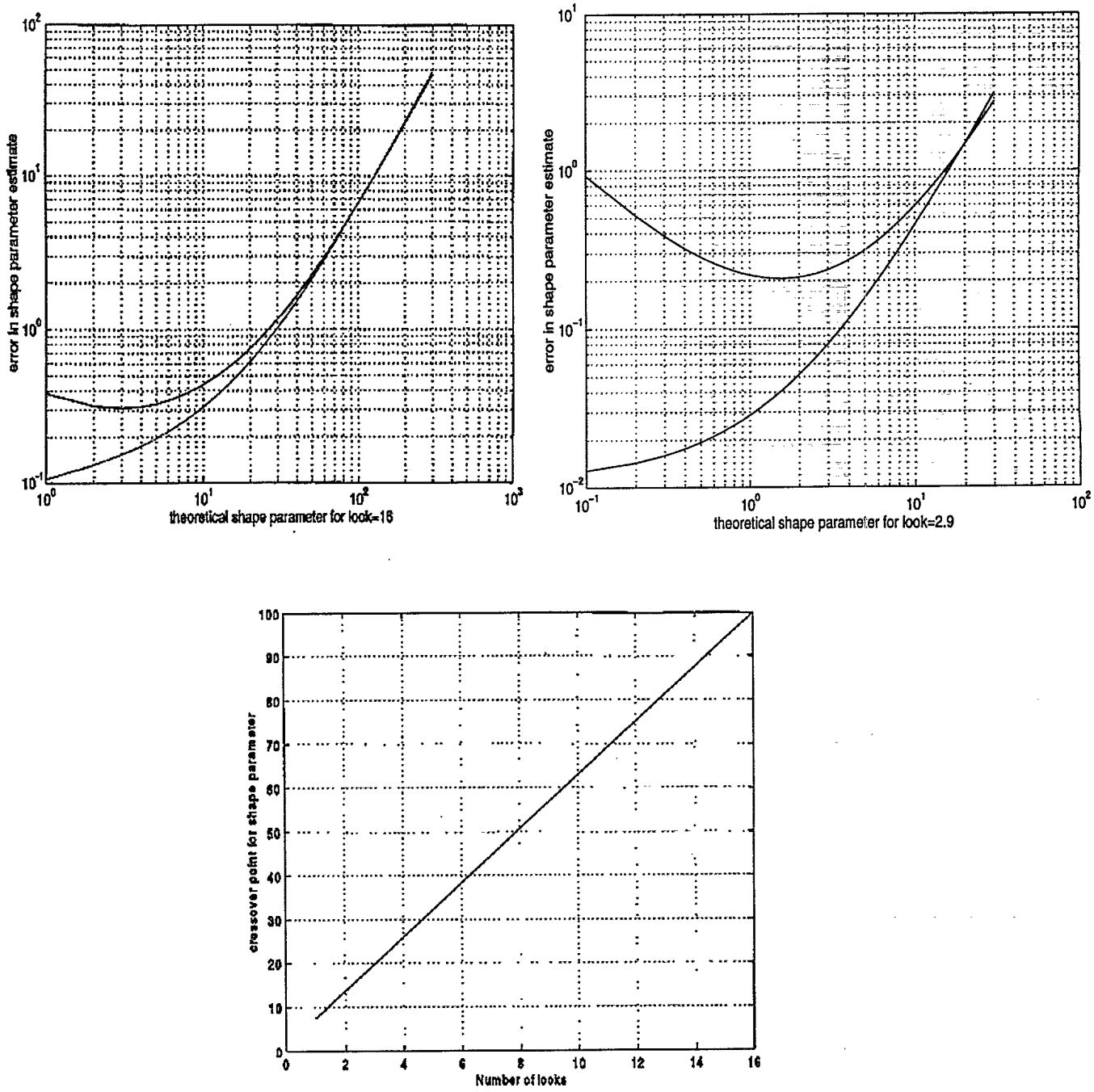

Figure 2: Error in shape parameter estimate vs theoretical shape parameter for number of looks $L = 16$ and $L = 2.9$ (top graphs). The solid line denotes the estimator based on the variance of the data and the dashed line on the estimator based on the log of the mean of the data. The cross-over point as a function of L is shown at the bottom graph.

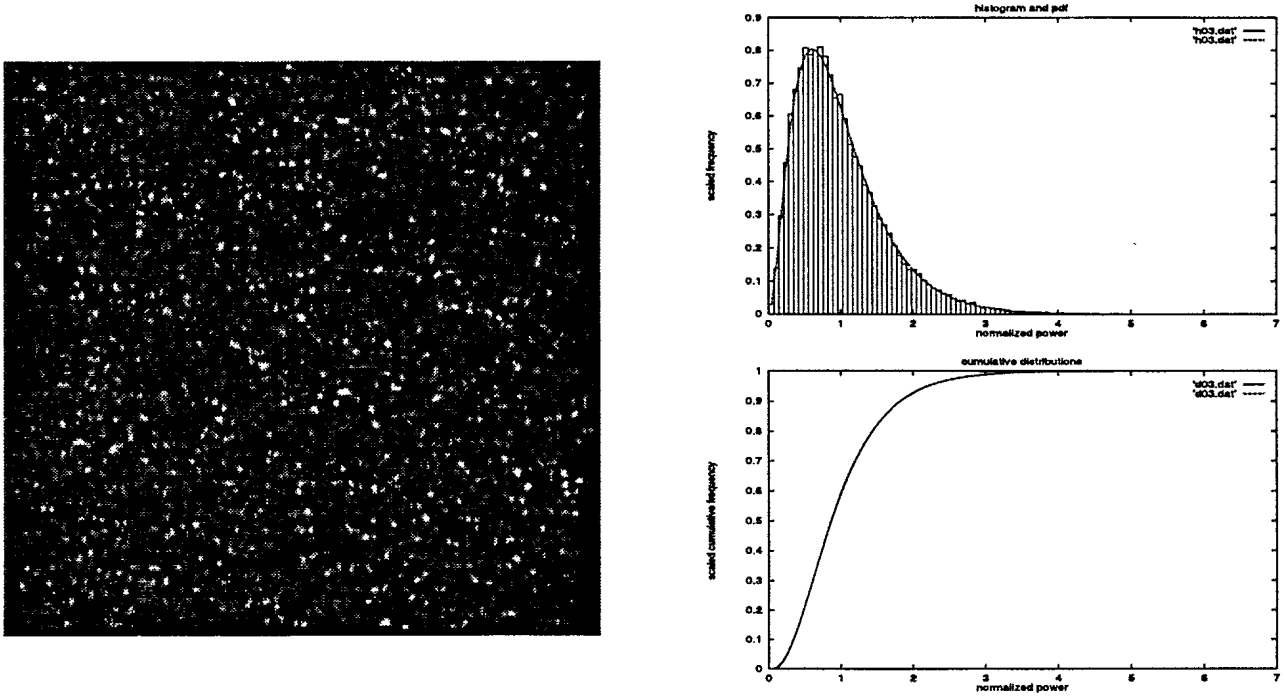

Figure 3: Img03dec94 with histogram and model pdf as well as data and model cdf's.

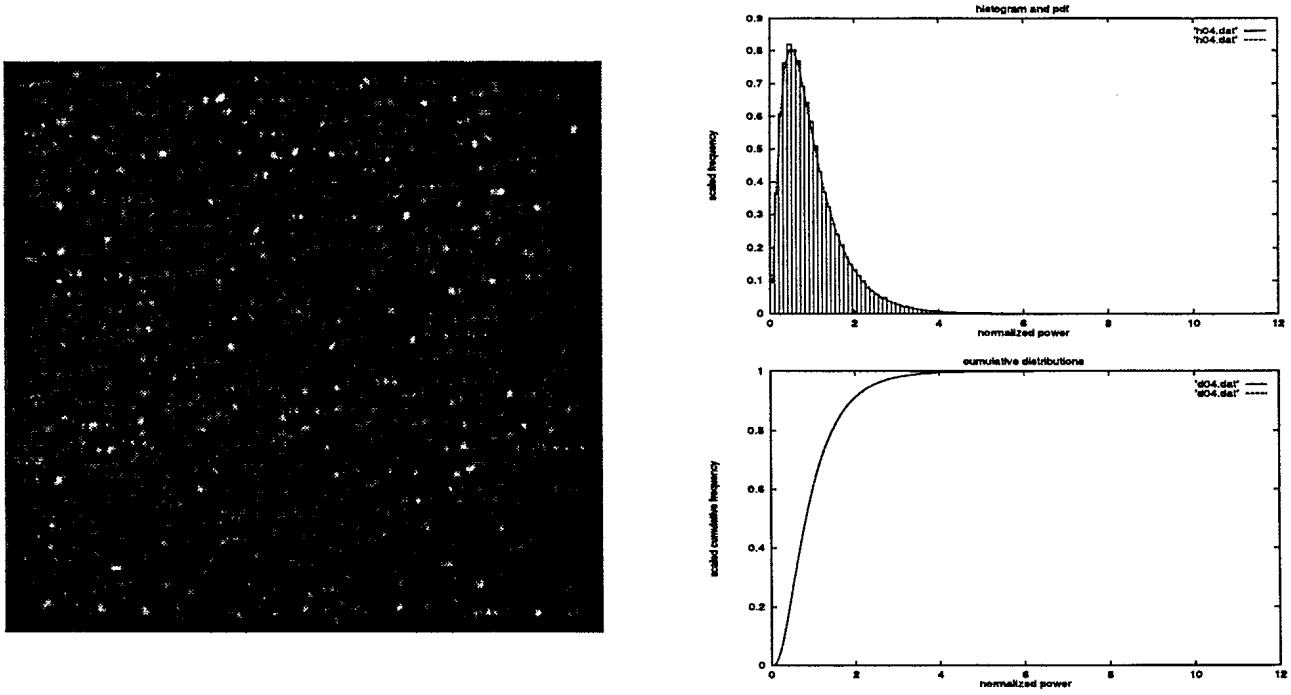

Figure 4: Img04dec94 with histogram and model pdf as well as data and model cdf's.

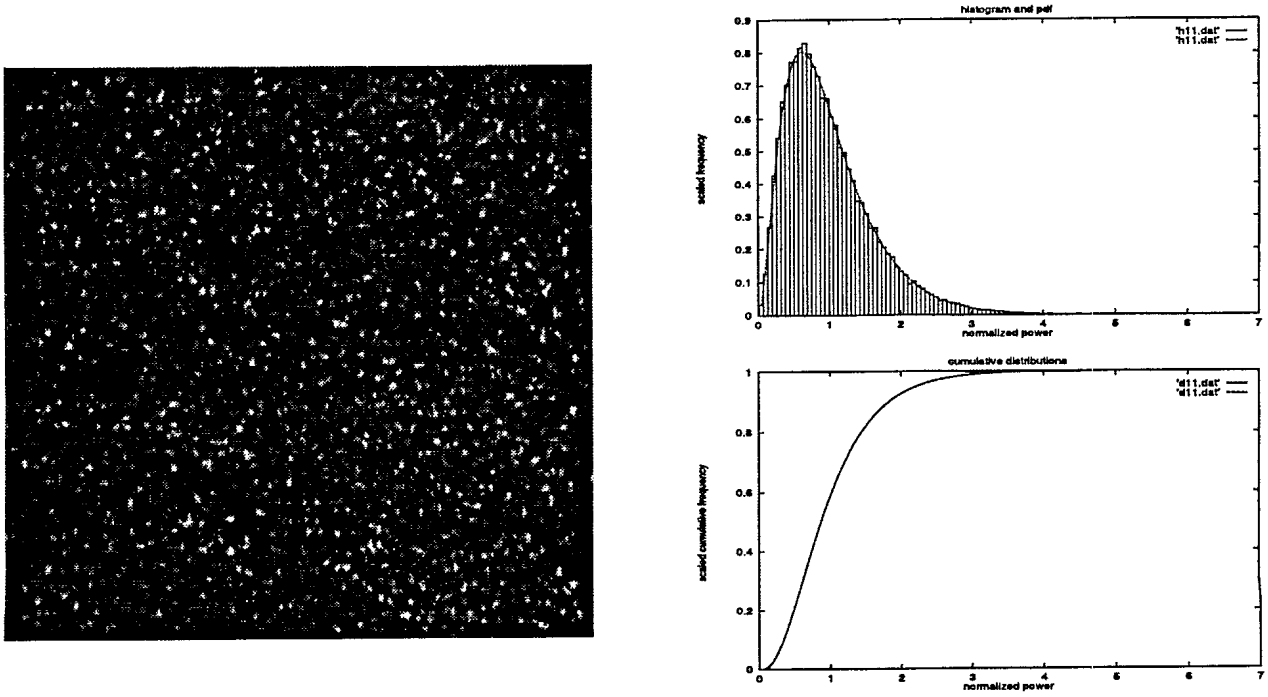

Figure 5: Img11dec91 with histogram and model pdf as well as data and model cdf's.

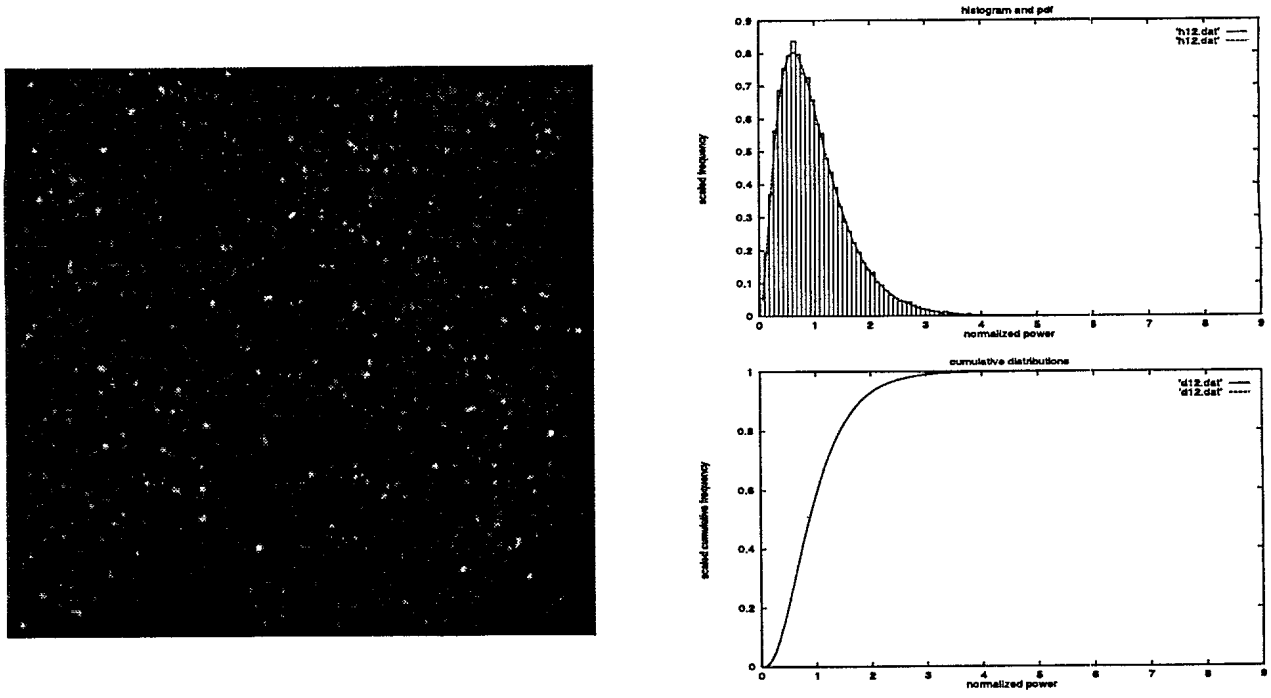

Figure 6: Img12dec91 with histogram and model pdf as well as data and model cdf's.

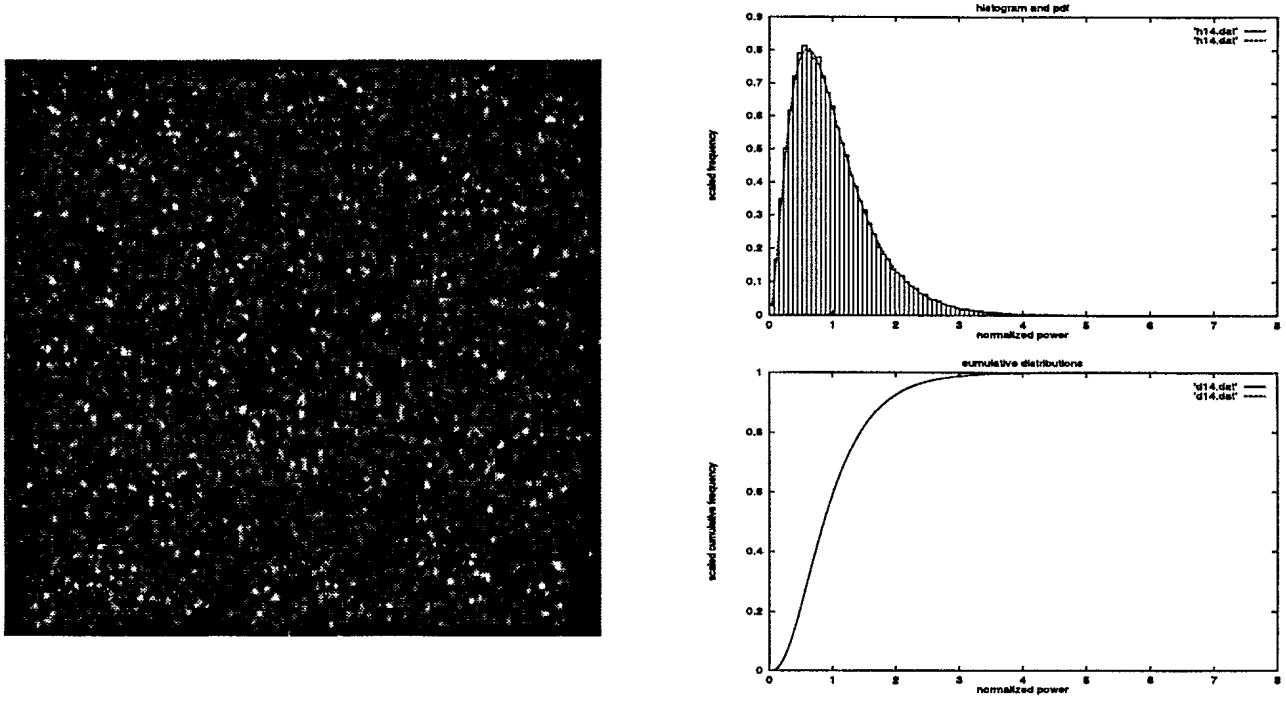

Figure 7: Img14dec91 with histogram and model pdf as well as data and model cdf's.

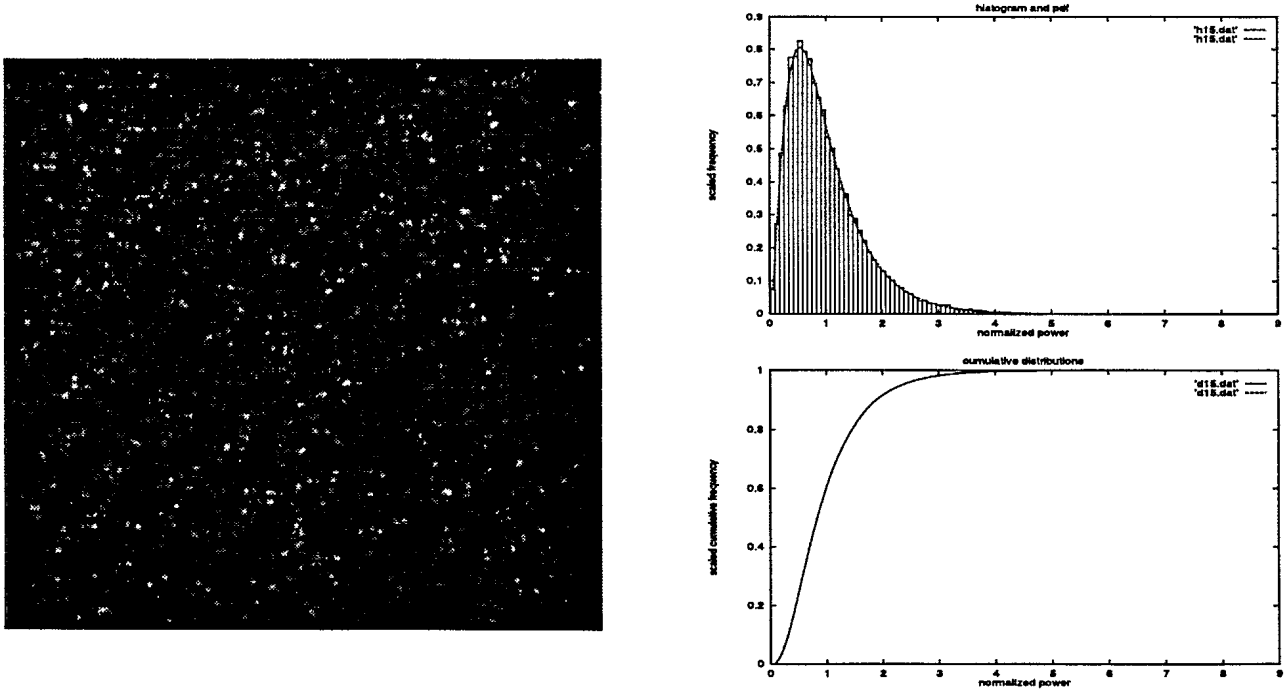

Figure 8: Img15dec91 with histogram and model pdf as well as data and model cdf's.

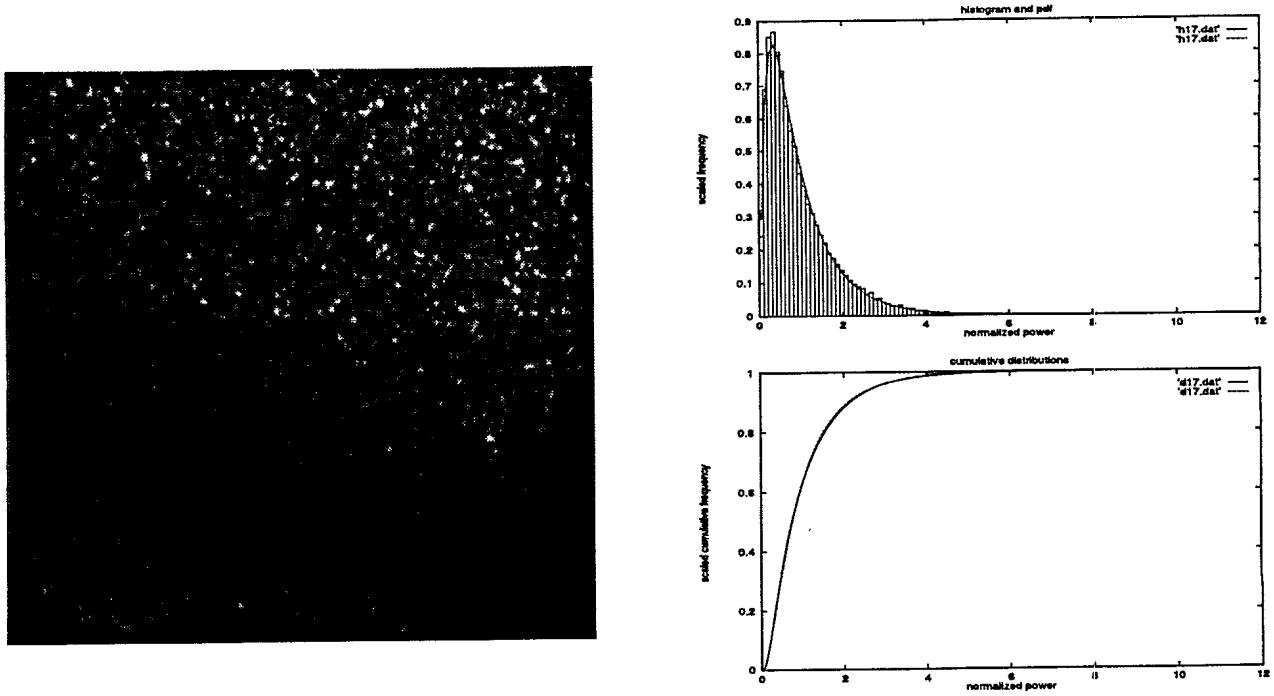

Figure 9: Img17dec91 with histogram and model pdf as well as data and model cdf's.

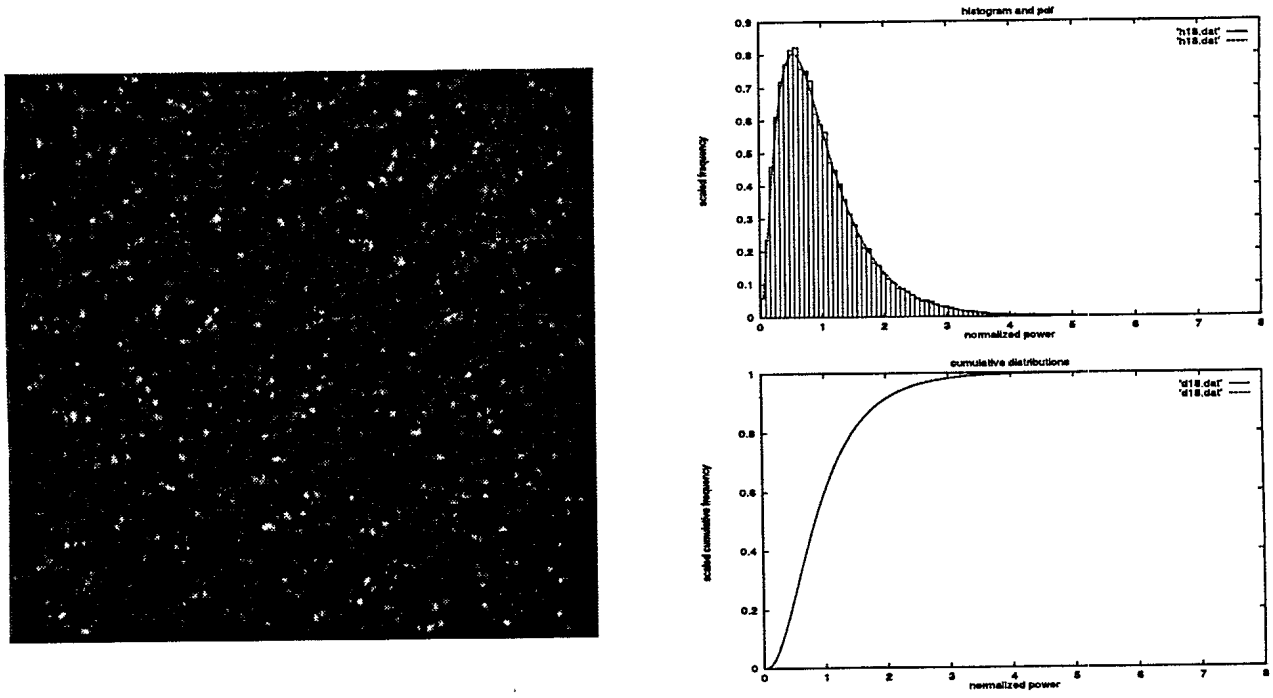

Figure 10: Img18dec91 with histogram and model pdf as well as data and model cdf's.

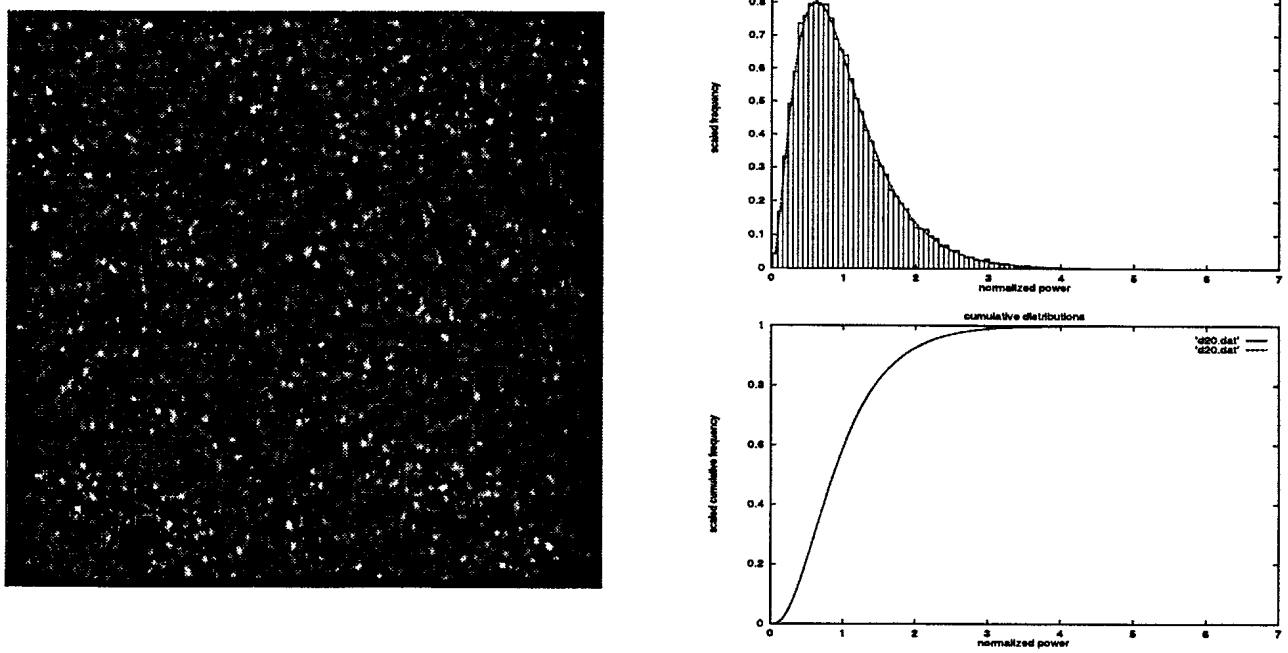

Figure 11: Img20dec91 with histogram and model pdf as well as data and model cdf's.

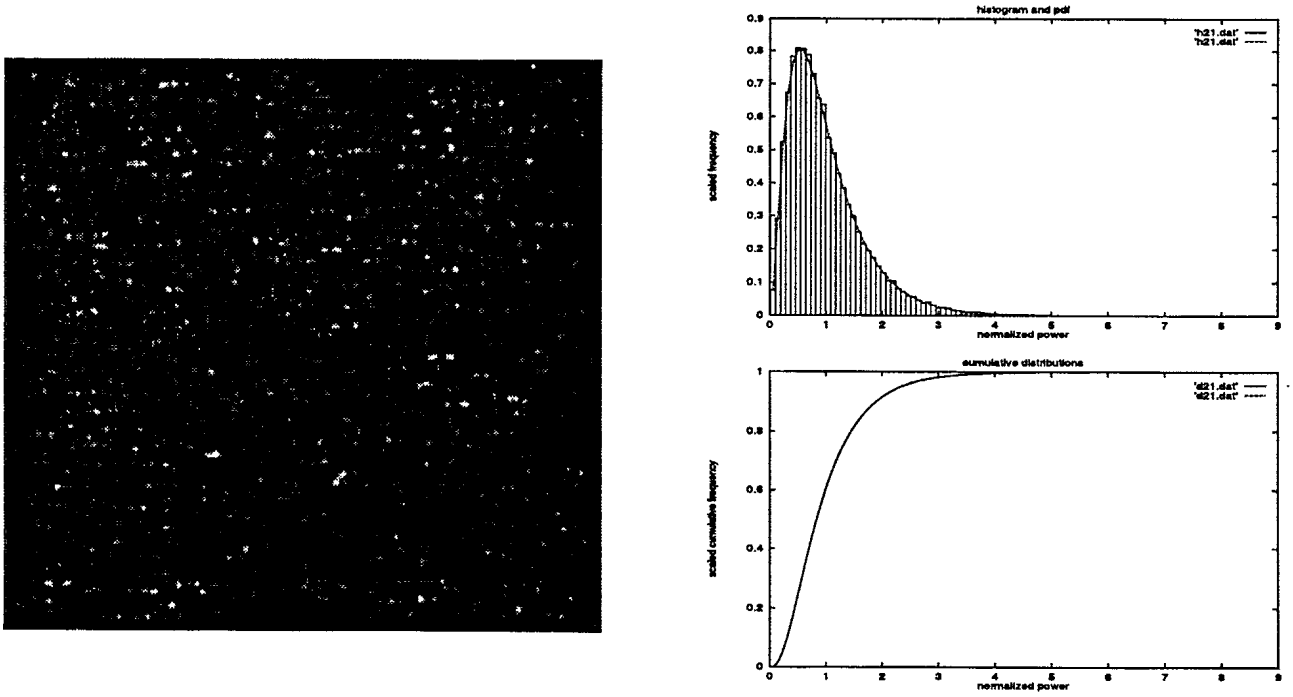

Figure 12: Img21dec91 with histogram and model pdf as well as data and model cdf's.

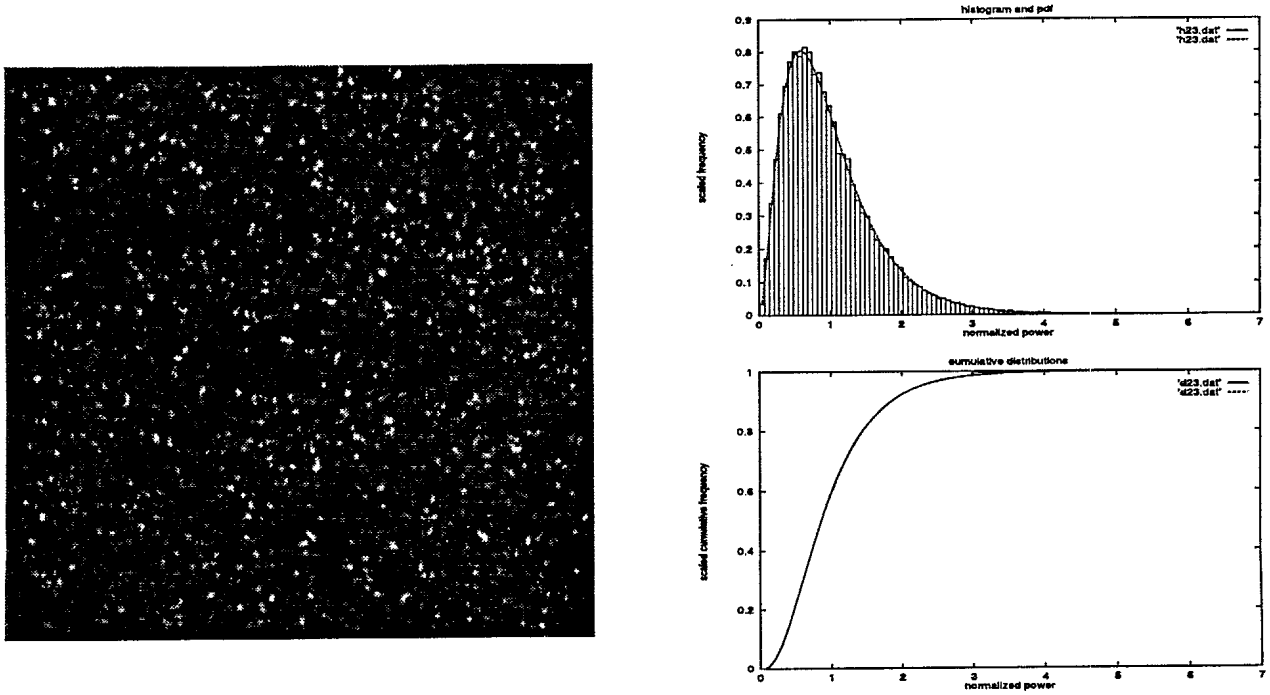

Figure 13: Img23dec91 with histogram and model pdf as well as data and model cdf's.

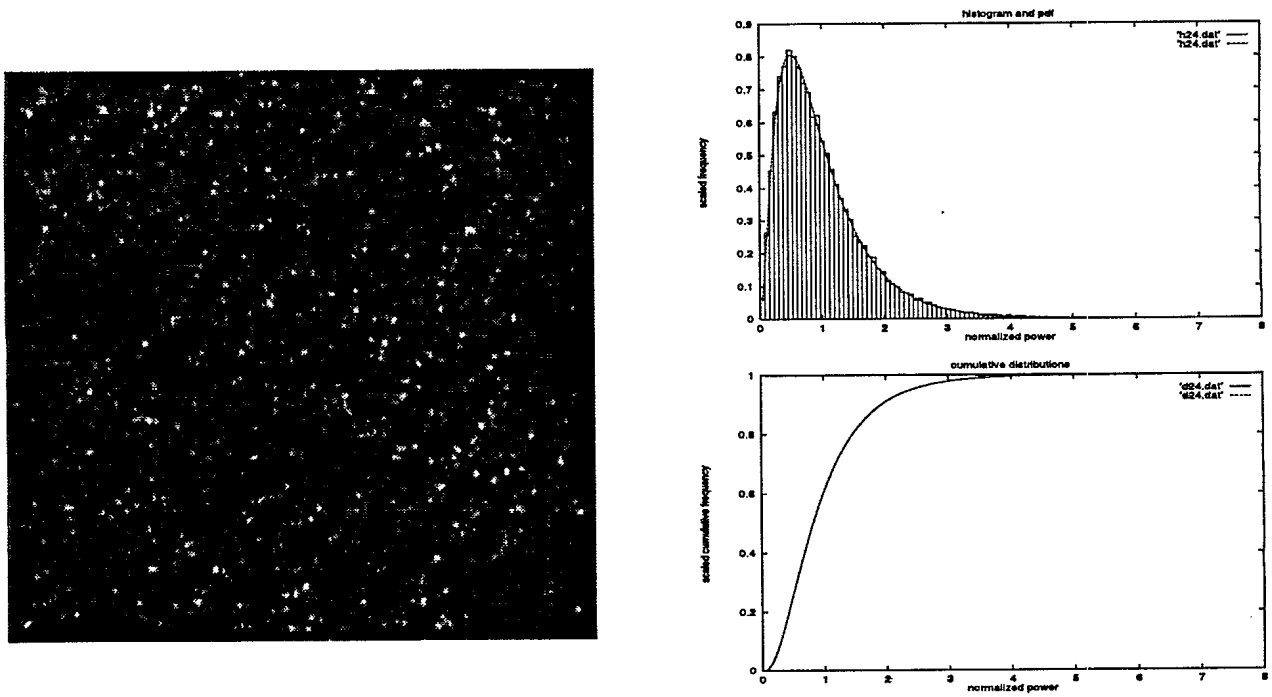

Figure 14: Img24dec91 with histogram and model pdf as well as data and model cdf's.

Image	ν	χ^2 value	df	n_d	D	Q_{KS}
	24.42	122.6	70		0.0038	1
img03dec94	29.22	130.6	69	1897	0.0049	1
	34.64	154.4	70		0.0006	1
	7.039	119.8	59		0.00776	1
img04dec91	7.740	101.2	58	1825	0.00467	1
	8.305	121.2	57		0.00557	1
	28.73	89.88	76		0.00220	1
img11dec91	30.93	91.96	75	1159	0.00187	1
	32.51	95.22	75		0.00265	1
	31.55	97.56	56		0.00467	1
img12dec91	30.78	107.7	55	1821	0.00336	1
	35.17	98.54	56		0.00389	1
	22.28	91.89	66		0.00288	1
img14dec91	24.68	96.85	65	2209	0.00263	1
	23.15	93.70	66		0.00233	1
	9.492	94.60	68		0.00443	1
img15dec91	9.859	93.95	67	1288	0.00314	1
	10.06	95.89	67		0.00248	1
	3.120	287.1	74		0.01360	0.9995
img17dec91	3.034	269.5	75	1351	0.01089	0.999998
	3.024	268.8	75		0.01058	0.999999
	10.40	109.7	70		0.00321	1
img18dec91	10.99	110.8	69	1285	0.00270	1
	10.25	106.4	71		0.00382	1
	26.70	107.0	68		0.00392	1
img20dec91	30.31	110.4	67	2095	0.00446	1
	34.05	124.1	67		0.00492	1
	9.344	95.1	63		0.00525	1
img21dec91	10.20	95.8	62	1580	0.00336	1
	10.79	104.2	61		0.00422	1
	16.34	135.5	75		0.00555	1
img23dec91	19.00	139.6	73	1608	0.00538	1
	15.70	140.6	75		0.00652	1
	7.757	112.7	77		0.00252	1
img24dec91	7.580	108.9	78	1236	0.00323	1
	7.808	113.9	77		0.00260	1

Table 1: Blacknell's ν estimates using sample mean and data variance (first ν value), sample mean and mean of log (second ν value), mean of log and variance of log (third ν value) on some ERS-1 datasets. The results of the χ^2 and Kolmogorov-Smirnov tests for each case are also shown.

curves at $L = 2.9$ (Figure 2) indicate that for a ν parameter range of about $9 - 35$, there is little difference between the two estimators. This is the range for the majority of the sampled data in our possesion. Therefore, for present algorithm development purposes, we will assume the theoretical approach for deciding on correct choice of parameter estimator, ie. a choice based on an investigation of the number of independent looks and the corresponding cross-over value in Figure 2. The equation that defines this cross-over point is approximately:

$$\nu = 6.1L + 1.25 \tag{26}$$

4.2 Evaluation of Kolmogorov-Smirnov and χ^2 goodness-of-fit tests

The χ^2 and KS tests differ fundamentally in that whereas the χ^2 test bins the data and uses the entire histogram of a distribution for comparison purposes, the KS test is an ordered statistical test, suitable for unbinned data. It is interesting to observe that when applied to the data in Figures 3-14 the fit is usually better than 0.99999. The KS test was applied to ERS-1 data files of size 256×256. The ν parameter using the mean and log of mean technique was calculated using these data files, and the value of ν was used to determine the theoretical distribution for comparison with the data results. The KS test was done using as input all distinct ordered points in the entire distribution for each of these files. The results are given in Table 1.

As an additional exercise, the Pearson plot (kurtosis vs. skewness squared) of small windows from a large RADARSAT data file is also calculated (Fig. 15). The theoretical lines for increments of one in the ν parameter are shown as well. Note that these plots simply display the data in a form that suggests possible approximate distributions that fit the observations [8]. Even though they do not quantify the goodness-of-fit of the data to the theoretical distribution, they play a supportive role in the "possible model exploration stage". A clear correlation between theoretical fits and data values for increasing values of ν can be seen, supporting our conclusion that the K-model gives a good description of sea clutter data. Formulas for the skewness and kurtosis can be found in [13, pg. 612].

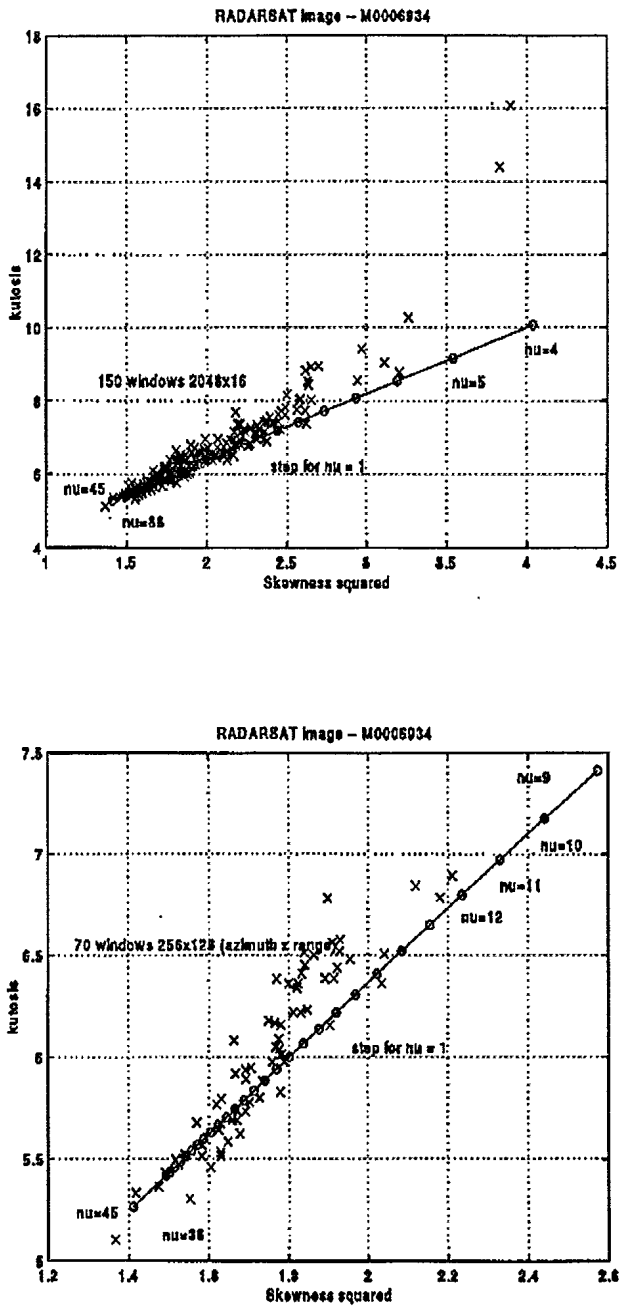

Figure 15: Pearson plots of kurtosis vs skewness squared for small regions of large RADARSAT images. The solid line is the theoretical model for the normalized K distribution for a variety of shape parameters ν.

5 Incidence angle dependence

Search procedures for ships in satellite SAR imagery must contend with large swath widths, and hence variation in mean cross-section with changing incidence angle. This aspect, and its implications on the automatic search procedure, were investigated for ERS-1 data, at VV polarization. Although similar comments will hold for the (HH-polarized) RADARSAT, the elevation pattern corrections have as yet to be fully developed for RADARSAT data, and hence this methodology was investigated on ERS-1.

Figure 16 shows the mean power level as a function of range offset (across swath) for varying window sizes. The windows are all 300 pixels in azimuth, but vary from 25 to 300 pixels in the range dimension. This process was repeated for several looks. As can be expected, the general trend in power decrease with increasing swath width (range offset) is similar across the various looks, with variance in the mean power level decreasing with increasing number of looks. The last plot in Figure 16 shows the corresponding variance in the shape parameter for the mean and mean of the log of the data estimator.

For the development of a CFAR approach, it is necessary to investigate the effect that the range offset has on the resultant mean estimate, and hence the resultant CFAR value given in equation (9). For this investigation, the CFAR is a function of four variables, three from equation (7), namely, shape parameter ν, average clutter power μ and number of looks L, as well as the target detection threshold I_c, if CFAR values are to be examined as a function of μ. Therefore equation (9) is solved iteratively for fixed values of ν, L, and I_c, as a function of μ. Figure 17 shows the resulting CFAR plots.

An inspection of the CFAR figures indicate the following trends:

1. For a given I_c, there is a dependence of false alarm rate on clutter mean, with greater deviation in false alarm for small ν parameters.

2. For a given I_c, the dependence of false alarm rate with mean estimate is sensitive to look number, with greater deviations in false alarm for a given I_c for decreasing look number.

The above discussion indicates that a careful evaluation of the mean must be made, in order to ensure CFAR robustness, particularly in single-look imagery with low shape parameter (higher sea states).

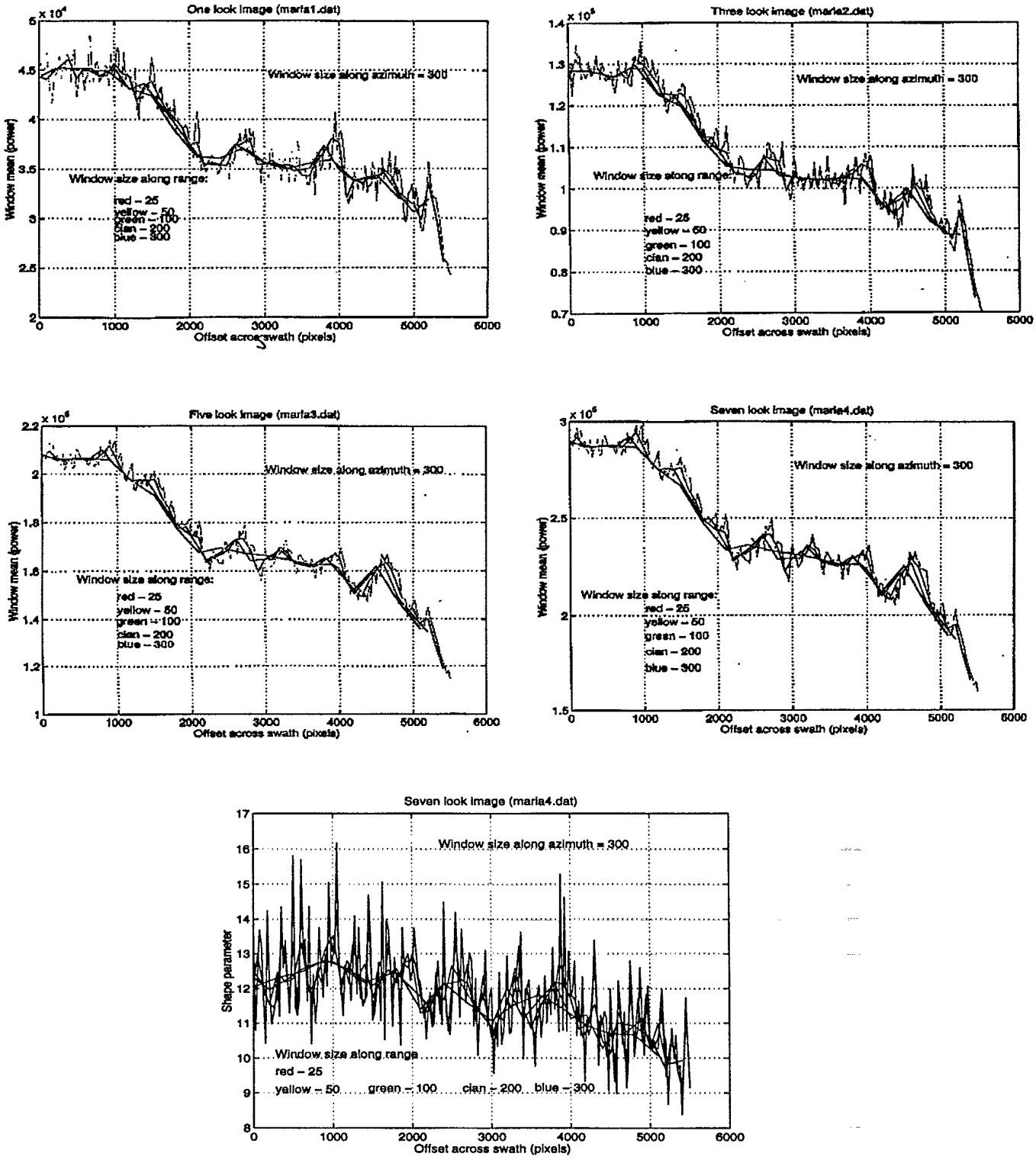

Figure 16: Mean power level as a function of range offset (across swath) for varying window size. The last figure shows the variance in the shape parameter for the mean and mean of the log of the data estimator for the maria4 dataset.

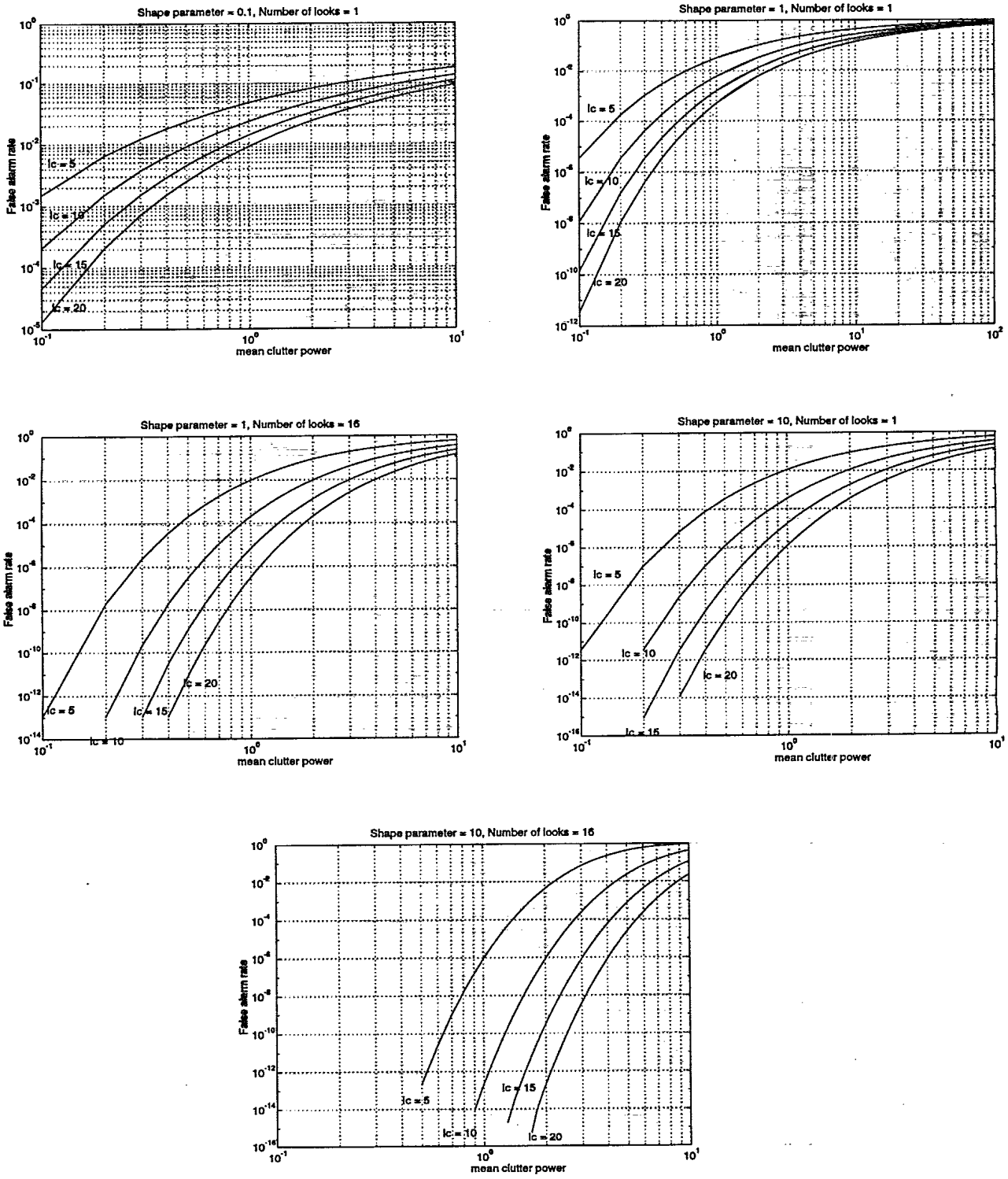

Figure 17: CFAR vs μ for a variety of fixed values of ν, L, and I_c.

6 Design of RADARSAT CFAR

Based upon investigations carried out above, it was decided to design a new ship search procedure for the OFW, and test its performance against a small number of truthed ERS-1 and RADARSAT images (with and without ships). This RADARSAT data represents to date, the only truthed RADARSAT data available, and consists of only three scenes.

The new ship search procedure first computes the mean and standard deviation of a user-selected window of RADARSAT imagery to be analyzed for the presence of ships. Both Blacknell's techniques, mean and log of the mean, and mean and variance of the data are used to estimate ν.

Equation (26) is then used to determine which ν estimate will be chosen for the ensuing steps. Once the ν parameter has been chosen, the theoretical cumulative probability density function is calculated using the threshold value in the current window. This cumulative probability value is used to compare to a user-selected CFAR input value. If the value is smaller than the user-selected CFAR value, the algorithm determines that there are no significant peaks in the window, and the window is flaged with a value of -1. If the value is higher, the algorithm computes the threshold for the required user-input significance level.

The pixels in the window are then compared with this threshold value, to find the pixels that exceed the user-selected value, unless the window has been flagged with a value of -1. Figure 18 shows the flow diagram for the CFAR detection process.

6.1 Results of CFAR procedure

The new CFAR approach was tested on 5 images of size 255×125. Two of these images contained truthed ships. The results are tabulated in Table 2.

The CFAR values are varied in to indicate the change in the number of false alarm with choice of CFAR value. It is clear that most of the truthed ships in these images pass even as the CFAR value is lowered, probably due to the considerably higher cross-section they exhibit relative to the surrounding ocean.

As little data of known ships in RADARSAT imagery exists, it is not apparent at this time what the optimal choice of CFAR should be. This will require further research as such data becomes available from the RADARSAT sensor.

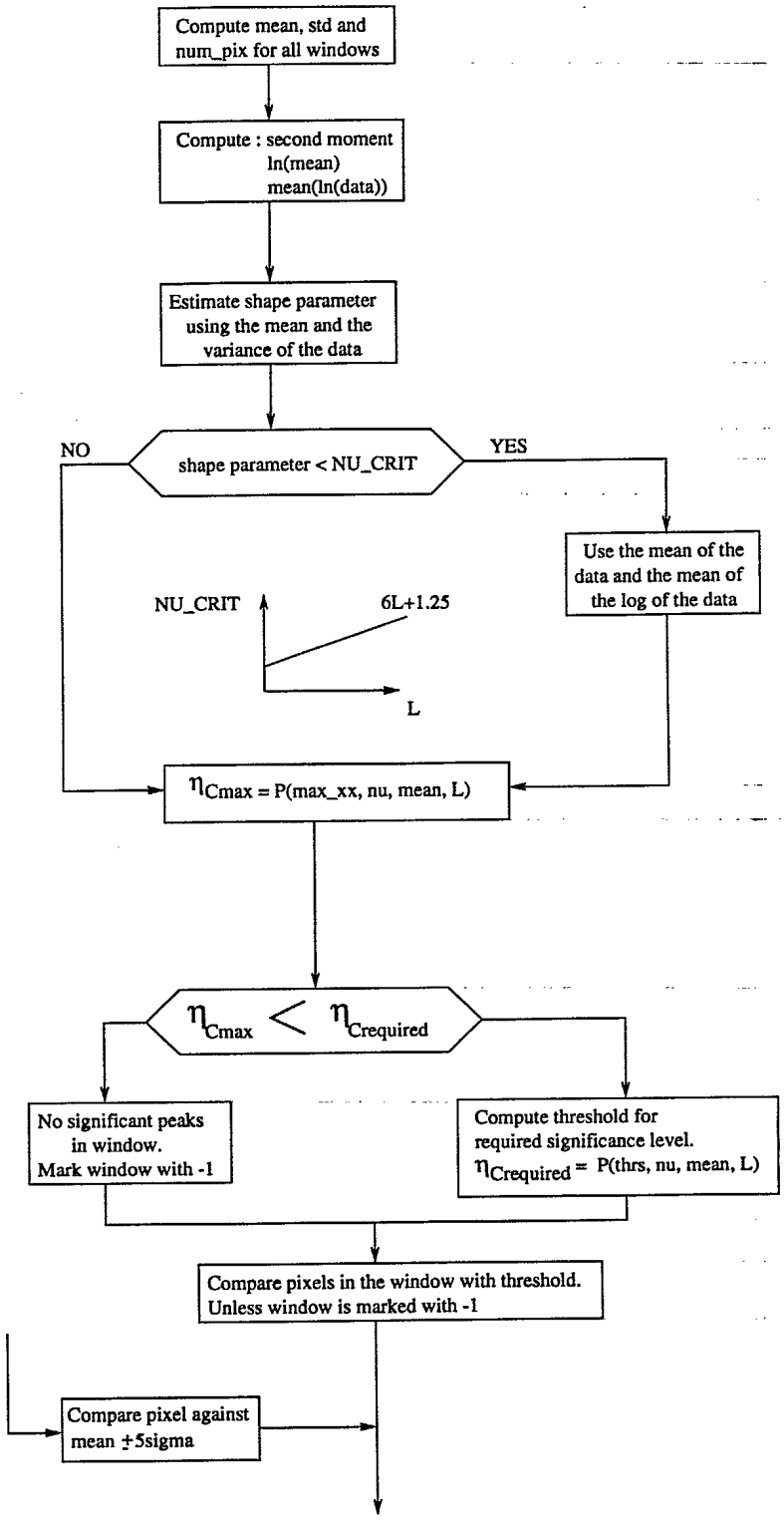

Figure 18: Flow diagram of CFAR Ship Search Procedure.

Image ID	CFAR value	peaks detected	visible ships
maria1.dat (ERS-1)	10^{-7}	6	none
	10^{-8}	0	
	10^{-9}	0	
maria7.dat (ERS-1)	10^{-7}	4	none
	10^{-8}	1	
	10^{-9}	0	
M006905 (RADARSAT)	10^{-7}	37	5
	10^{-8}	11	
	10^{-9}	8	
M006906 (RADARSAT)	10^{-7}	9	none
	10^{-8}	1	
	10^{-9}	1	
M006934 (RADARSAT)	10^{-7}	34	5 visible
	10^{-8}	11	(+ cluster of 4 small targets
	10^{-9}	4	potentially 9 in total)

Table 2: Results for the CFAR Technique with Varying Threshold on a variety of 31875 pixel size images.

7 Conclusions and future work

Currently, progress beyond this point, in terms of algorithm evaluation, is suffering predominantly from a lack of truthed RADARSAT data. It is tentatively clear, from the limited data we have investigated, that the K-distribution is a good fit to RADARSAT ocean clutter. Also, to our knowledge, and from a thorough review of literature, no one else has published the analytical cumulative distribution fuction for the L-look, K-distribution.

However, further evaluation of parameter estimation and goodness-of-fit tests with truthed RADARSAT data will continue in the next phase of the work. Alternative models will be examined to determine the optimum one for our application in terms of complexity and performance. The second technique mentioned in Section 1, based upon an investigation of ambient sea state and expected target length is definitely warranted, based upon the results of this research. Again, this second technique is particularly dependent on the need for truthed RADARSAT data of known ships and ocean clutter.

The next phase will also include an investigation of the use of ancillary information, such as use of the wake information currently produced by the OFW to confirm a peak detection. This will likely result in a considerable decrease in the false alarm rate.

A Simulating K distributed clutter

In the Kolmogorov-Smirnov test the need arises to generate K distributed clutter from

$$p(x) = \frac{2}{x\Gamma(\nu)\Gamma(L)} \left(\frac{L\nu x}{\mu}\right)^{(L+\nu)/2} K_{L-\nu}\left(2\sqrt{\frac{L\nu x}{\mu}}\right) \tag{27}$$

as per Blacknell's [3] formulation.

The most straightforward approach to achieve this is as follows:

1. Generate a gamma distributed dataset $\{u_i\}_{i=1}^{i=N}$ with pdf given by:

$$\mathcal{G}_\nu(x) = \frac{x^{\nu-1}e^{-x}}{\Gamma(\nu)} \qquad \nu > 0$$

2. Scale the above dataset as

$$y_i = \frac{\mu}{L\nu} x_i$$

to obtain another with pdf

$$\mathcal{G}_\nu(y) = \left(\frac{L\nu}{\mu}\right)^\nu \frac{y^{\nu-1}}{\Gamma(\nu)} \exp\left(-\frac{L\nu y}{\mu}\right)$$

3. Generate another, independent, gamma distributed dataset $\{z_i\}_{i=1}^{i=N}$ with pdf:

$$\mathcal{G}_L(z) = \frac{z^{L-1}e^{-z}}{\Gamma(L)}$$

4. The product $r_i = y_i z_i$ will have pdf

$$p(r) = \frac{2}{r\Gamma(\nu)\Gamma(L)} \left(\frac{L\nu r}{\mu}\right)^{(L+\nu)/2} K_{L-\nu}\left(2\sqrt{\frac{L\nu r}{\mu}}\right)$$

In essence, a good technique to generate gamma deviates is the only requirement. Use can be made of the fact that the sum of two gamma deviates is itself a gamma deviate with parameter the sum of the individual deviate parameters, i.e., $\mathcal{G}_{a+b} = \mathcal{G}_a + \mathcal{G}_b$. Parameters ν or L can then be broken into integer and fractional parts. Integer gamma deviates can be generated by taking the log of a product of uniform deviates (see [13, pg. 292]). For the fractional gamma deviates, the transformation method [13, pg. 288] can be employed, with

$$\gamma_i = F^{-1}(u_i)$$

where the u_i random variables are uniformly distributed in $[0, 1]$ and

$$F(x) = \frac{1}{\Gamma(a)} \int_0^x t^{a-1} e^{-t} dt$$

is the incomplete Γ function. This can be inverted numerically and the desired gamma deviates, γ_i, can be obtained.

This approach is more straightforward, numerically stable and simpler to implement than the ones described in Schlerer [18], Armstrong and Griffith [2] and d'Addio et.al. [4]. It can also easily be modified to generate amplitude K distributed deviates, or by suitable scaling, to include thermal noise effects as described in Watt's K distribution and thermal noise model[23].

References

[1] M. Abramowitz and I.A. Stegun. *Handbook of Mathematical Functions.* Dover, 1972.

[2] B.C. Armstrong and H.D. Griffiths. Modelling spatially correlated K-distributed clutter. *Electronics Letters*, 27:1355–1366, 1991.

[3] D. Blacknell. Comparison of Parameter Estimators for K-Distribution. *IEE Proc. Radar, Sonar, Navigation*, 141(1):45–52, Feb 1994.

[4] E. D'Addio, S. Giannatempo, and G. Galati. Generation of K-distributed random variables. *Trans. of the Soc. for Computer Simulations*, 5:159–174, 1988.

[5] Jakeman E. On the statistics of K-distributed noise. *J. Phys. A.*, 13:31–48, 1980.

[6] M.T. Rey et al. Utility of Electro-Optical and Radar Technologies for Space-Based Surveillance. *Research and Development Branch Space Studies Project Report*, July 1996.

[7] X.Y. Hou and N. Morinaga. Detecion Performance in K-Distributed and Correlated Rayleigh Clutter. *IEEE Trans. Aerospace and Electronic Systems*, AES-25(5):634–641, Sept 1989.

[8] M. Kendall, A. Stuart, and J.K. Ord. *Kendall's Advanced Theory of Statistics*, volume I: Distribution Theory. Oxford University Press, 1987.

[9] D.D. Kriethen and G.G. Hogan. Statistical Analysis of Ka-band sea clutter. In *Proc. Oceans-91*, Honolulu, Hawaii, Oct 1991.

[10] D.J. Lewinski. Nonstationary probabilistic target and clutter scattering models. *IEEE Trans. Anten. & Prop.*, AP-31:490–498, 1983.

[11] C.J. Oliver. Correlated K-distributed clutter models. *Optica Acta*, 32:1515–1547, 1985.

[12] C.J. Oliver. Optimum texture estimators for SAR Clutter. *Journal of Physics D*, 1993.

[13] W.H. Press, W.T. Vetterling, S.A. Teukolsky, and B.P. Flannery. *Numerical Recipes in C.* Cambridge University Press, 2nd edition, 1992.

[14] A.P. Prudnikov, Y.A. Brychkov, and O.I. Marichev. *Integrals and Series*, volume I: Elementary Functions and II: Special Functions. Gordon and Breach Science Publishers, 1986.

[15] M.T. Rey, J.K. Tunaley, J.T. Folinsbee, P.A. Jahans, J.A. Dixon, and M.R. Vant. Application of Radon Transform Techniques to Wake Detection in SEASAT-A SAR Images. *IEEE Trans. on Geoscience and Remote Sensing*, 28(4):553–560, July 1990.

[16] M.T. Rey, J.K. Tunaley, and T. Sibbald. Use of the Dempster-Shafer Algorithm for Detection of SAR Ship Wakes. *IEEE Trans. on Geoscience and Remote Sensing*, Sept 1991.

[17] SATLANTIC. Ocean Feature Workstation User's Guide. Technical Report 94087-11, July 1996.

[18] D.C. Schleher. Periscope detection radar. In *IEEE International Radar Conference*, pages 704–707, 1995.

[19] M.I. Skolnik. *Introduction to Radar Systems.* McGraw-Hill, 2nd edition, 1980.

[20] P. Vachon. Ship Detection by RADARSAT SAR. working paper, Canada Centre for Remote Sensing, Sept 1995.

[21] K.D. Ward. Compound Representation of High Resolution Sea Clutter. *Electronic Letters*, 17:561–563, 1981.

[22] K.D. Ward, C.J Baker, and S. Watts. Maritime surveillance radar, Part1: Radar scattering from the sea surface. *IEE Proc. F*, 137:51–62, 1990.

[23] S. Watts. Radar Detection Prediction in K-Distributed Sea Clutter and Thermal Noise. *IEEE Trans. Aerospace and Electronic Systems*, AES-28(1), Jan 1987.

[24] S. Wolfram. *Mathematica.* Addison-Wesley, 2nd edition, 1993.

DOCUMENT CONTROL DATA

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

1. ORIGINATOR (the name and address of the organization preparing the document. Organizations for whom the document was prepared, e.g. Establishment sponsoring a contractor's report, or tasking agency, are entered in section 8.)

Defence Research Establishment Ottawa

2. SECURITY CLASSIFICATION (overall security classification of the document. including special warning terms if applicable)

UNCLASSIFIED

3. TITLE (the complete document title as indicated on the title page. Its classification should be indicated by the appropriate abbreviation (S,C or U) in parentheses after the title.)

A SEARCH PROCEDURE FOR SHIPS IN RADARSAT IMAGERY (U)

4. AUTHORS (Last name, first name, middle initial)

REY, Maria, DROSOPOULOS, Tasos, PETROVIC, Dusan

5. DATE OF PUBLICATION (month and year of publication of document)

November 1996

6a. NO. OF PAGES (total containing information. Include Annexes, Appendices, etc.)

35

6b. NO. OF REFS (total cited in document)

23

7. DESCRIPTIVE NOTES (the category of the document, e.g. technical report, technical note or memorandum. If appropriate, enter the type of report, e.g. interim, progress, summary, annual or final. Give the inclusive dates when a specific reporting period is covered.)

Technical Report

8. SPONSORING ACTIVITY (the name of the department project office or laboratory sponsoring the research and development. Include the address.)

DREO

9a. PROJECT OR GRANT NO. (if appropriate, the applicable research and development project or grant number under which the document was written. Please specify whether project or grant)

5EB16

9b. CONTRACT NO. (if appropriate, the applicable number under which the document was written)

10a. ORIGINATOR'S DOCUMENT NUMBER (the official document number by which the document is identified by the originating activity. This number must be unique to this document.)

DREO REPORT 1305

10b. OTHER DOCUMENT NOS. (Any other numbers which may be assigned this document either by the originator or by the sponsor)

11. DOCUMENT AVAILABILITY (any limitations on further dissemination of the document, other than those imposed by security classification)

- (X) Unlimited distribution
- ( ) Distribution limited to defence departments and defence contractors; further distribution only as approved
- ( ) Distribution limited to defence departments and Canadian defence contractors; further distribution only as approved
- ( ) Distribution limited to government departments and agencies; further distribution only as approved
- ( ) Distribution limited to defence departments; further distribution only as approved
- ( ) Other (please specify):

12. DOCUMENT ANNOUNCEMENT (any limitation to the bibliographic announcement of this document. This will normally correspond to the Document Availabilty (11). However, where further distribution (beyond the audience specified in 11) is possible, a wider announcement audience may be selected.)

UNCLASSIFIED
SECURITY CLASSIFICATION OF FORM

13. ABSTRACT (a brief and factual summary of the document. It may also appear elsewhere in the body of the document itself. It is highly desirable that the abstract of classified documents be unclassified. Each paragraph of the abstract shall begin with an indication of the security classification of the information in the paragraph (unless the document itself is unclassified) represented as (S), (C), or (U). It is not necessary to include here abstracts in both offical languages unless the text is bilingual).

This report describes the CFAR procedure applied to the Ocean Features Workstation, OFW, based on the K distributed sea clutter model. A number of scenes from ERS-1 and RADARSAT are examined and the goodness-of-fit of the K distribution model is investgated, using both visual and standard statistical tests. Improved performance in detecting ship targets in comparison with the previous method on the OFW is observed.

14. KEYWORDS, DESCRIPTORS or IDENTIFIERS (technically meaningful terms or short phrases that characterize a document and could be helpful in cataloguing the document. They should be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location may also be included. If possible keywords should be selected from a published thesaurus. e.g. Thesaurus of Engineering and Scientific Terms (TEST) and that thesaurus-identified. If it is not possible to select indexing terms which are Unclassified, the classification of each should be indicated as with the title.)

SAR
CFAR
K-DISTRIBUTION
RADARSAT
SHIP DETECTION

500743