SAN098-1812 C
SAND--98-1812C
CONF-981025--
RECEIVED
AUG 24 1998
OSTI

Statistical Modeling of Targets and Clutter in Single-Look Non-Polarimetric SAR Imagery

Jose S. Salazar, Don R. Hush, Mark W. Koch, Robert J. Fogler, Larry D. Hostetler
MS 0845, Sandia National Labs, P.O. Box 5800
Albuquerque, New Mexico 87185-0845

This paper presents a Generalized Logistic (*gLG*) [1] distribution as a unified model for Log-domain Synthetic Aperture Radar (SAR) data. This model stems from a special case of the G-distribution [2] known as the G^0-distribution. The G-distribution arises from a multiplicative SAR model and has the classical K-distribution as another special case. The G^0-distribution, however, can model extremely heterogeneous clutter regions that the K-distribution cannot model. This flexibility is preserved in the unified *gLG* model, which capable of modeling non-polarimetric SAR returns from clutter as well as man-made objects. Histograms of these two types of SAR returns have opposite skewness. The flexibility of the *gLG* model lies in its shape and shift parameters. The shape parameter describes the differing skewness between target and clutter data while the shift parameter compensates for movements in the mean as the shape parameter changes. A Maximum Likelihood (ML) estimate of the shape parameter gives an 'optimal' measure of the skewness of the SAR data. This measure provides a basis for an optimal target detection algorithm.

Keywords: Synthetic Aperture Radar, Automatic Target Recognition, Detection, Statistical Modeling, Generalized Logistic, Beta, G-distribution

I. Introduction

The automatic recognition of targets in SAR imagery can be computationally intensive. One approach for reducing the computational load, applies a simple detection algorithm over the entire scene. The detector locates regions that may contain objects of interest. Following the detection process, the more computationally intensive target recognition process is performed on the regions located by the detector.

The development of an optimal target detection algorithm depends on precise statistical modeling of the underlying clutter and target regions. Several models have been proposed in the past for clutter data. For example, Rayleigh [3] or Gaussian distributions are commonly used to model the backscatter associated with homogeneous regions such as bare ground surfaces, dense forest canopies or snow-covered ground. For other clutter types, such as sea surface backscatter, the Lognormal [3] and Weibull [3] have been used. In another example, the Modified Beta [4] has been proposed as a model for backscatter from different ice types.

For heterogeneous backgrounds many of these models are inadequate. Here, the K-distribution has been used extensively [5]. Also, since it was originally proposed by Jakeman and Pusey [6] to model microwave sea echoes, the K-distribution has become popular for modeling multilook [7] as well as polarimetric SAR signatures [8]. In addition, the K-distribution is attractive since it has been justified in terms of SAR backscattering processes. More recently, a new class of distributions known as the G-distribution [2] has been proposed to model SAR data. The classical K and the new G^0 distributions are special cases of this new class. However, in contrast with the K-distribution, the G^0-distribution can model extremely heterogeneous clutter, such as urban regions, that the K-distribution cannot properly model [2].

Empirical measurements indicate that histograms of many naturally occurring clutter types have different skewness from those of man-made objects. Here, the Log-domain histograms of naturally occurring clutter exhibit broad left tails whereas those of man-made objects can exhibit a range of distributions which include symmetric to right-skewed distributions. Therefore, a likely measure of the skewness or shape of these distributions can distinguish between these classes of SAR returns. In one such example, Pearson's Second Coefficient of Skewness served as basis for this measure [9].

An 'optimal' approach finds a probability density function whose parameters we can adjust and estimate to measure the shape of SAR returns from naturally occurring clutter to man-made objects. Additionally, this measure should produce reduced false alarm rates over those associated with less optimal approaches including skewness-based measures.

This paper proposes a Generalized Logistic or *gLG* [1] [10] distribution as a model for Log-domain clutter and target data. We will derive this model from the G^0-distribution. In particular, the single-look case of the G^0-distribution reduces to the Generalized Beta-Prime

-1-

MASTER

DISCLAIMER

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

(*gBP*) [1] distribution with two of its parameters set to one. We refer to this 2-parameter version as the Beta-Prime (*BP*) distribution. When Log-transformed, this yields the *gLG* distribution. This paper shows that this model provides a good fit to both Log-domain target as well as clutter data. The *gLG* distribution possesses two fixed parameters (scaling and offset) resulting from the Log transformation in addition to the Beta-Prime distribution parameters. The latter are the shape and shift parameters known as α and b respectively.

The *gLG* distribution lends itself, in part, to a Maximum Likelihood estimate of the shape parameter α. Here, we derive a closed form expression for the ML estimate of α. However, this expression is not only a function of the data samples, but also of the nuisance parameter b. Therefore, it would seem appropriate to derive the closed form expression for ML estimate for b then solve the system of two equations and two unknowns. However, a closed form expression for b does not exist. An alternate approach is to use the method-of-moments (MOM) to estimate α. However, the MOM cannot estimate the full range of α. This paper discusses one possible approach for estimating these parameters and contrasts it with a moment-based approach. Using the estimates resulting from the ML approach, we propose an 'optimal' detector whose performance can be determined analytically.

II. Classical SAR Model

SAR returns have been modeled as a distributed collection of radar scatterers each having a different amplitude and phase. The field strength from the *i-th* scatterer may be expressed as:

$$E_i = K_i E_{i_o} e^{j\phi_i} \tag{1}$$

where K_i is the system constant that accounts for radar system factors including propagation losses and antenna directivity gain, E_{i_o} is the field strength, and ϕ_i is the instantaneous phase expressed as:

$$\phi_i = \omega t - 2kR_i + \theta_i \tag{2}$$

In the above expression, θ_i is the scattering phase, R_i is the antenna to the scatterer range, ω is the carrier frequency and k is the wave number.

Assuming statistically independent scatterers, the total instantaneous field due to N scatterers can be expressed as a coherent sum of the scatterers in the usable portion of the radar beam. If we further assume that the area on the ground is small compared to the range from the radar platform to the ground, and that the antenna gain

is uniform across the area (i.e. $K_i = K$) we can write the total instantaneous field as:

$$E = K \sum_{i=1}^{N} E_{i_o} e^{j\phi_i} \tag{3}$$

III. Classical Clutter Model

For a large number of randomly range-distributed scatterers, applying the central limit describes the scattering amplitude and instantaneous phase as a complex process:

$$E e^{j\phi_i} = U + jV \tag{4}$$

where U and V are independent, normally distributed, zero-mean random variables. The SAR image formation process transforms the phase histories, $z = U + jV$, into a new random variable $w = I + jQ$, where I and Q are independent identically distributed $N(0, \sigma_G)$ random variables. The received power of each pixel in the formed image can be described by:

$$p = I^2 + Q^2 \tag{5}$$

whose statistics are described by the following exponential distribution:

$$f_P(p) = \frac{1}{2\sigma_G^2} e^{\frac{-p}{2\sigma_G^2}} \tag{6}$$

$$p, \sigma_G > 0$$

A Log-domain transformation, $g = \eta Log_e(p) + a'$, of the exponential random variable yields:

$$f_G(g) = \frac{1}{2\sigma_G^2 \eta} e^{\frac{(g-a')}{\eta}} e^{-\frac{1}{2\sigma_G^2} e^{\frac{(g-a')}{\eta}}} \tag{7}$$

$$\eta, a', \sigma_G > 0, \; g \in \Re$$

which is also known as the Gumbel distribution where η and a' represent fixed scale and bias constants. This distribution typically models Log-domain homogeneous clutter regions. Here, the parameter σ_G is a function of the back-scattering characteristics of the given homogeneous clutter type. Similarly, by applying the transformation of variables $m = \sqrt{p}$ we arrive at a Rayleigh distribution in the magnitude domain. Note that other models, such as the Weibull distribution can fit homogeneous clutter as well. This model approaches the Rayleigh distribution given an appropriate choice of parameters.

IV. Models for Man-Made Objects

A coherent sum of individual scatterers can also model man-made object radar returns. Unlike clutter, a small number of scatterers typically dominate the total instantaneous field. Therefore, the central limit theorem is not valid and the resulting scattering amplitude and instantaneous phase $Ee^{j\phi_i} = U + jV = z$ may not yield Gaussian (U, V) components.

Empirically, Log-domain probability density functions from man-made objects have broader right tails than from homogeneous clutter. From the classical model described earlier, these should fit the Gumbel distribution. Figures 1 and 2 show histograms that demonstrate this. Here, the bright and diffuse parts of a T-72 tank at approximately 230 different aspects were used for the target histogram. The clutter was histogramed from the homogeneous grassy regions of a similar set of imagery. For these examples, we used one foot resolution, single-look, HH-polarized, X-band SAR data at 17 degrees depression collected for the MSTAR[1] program.

V. Unified Model for Clutter and Targets

Equations (8), (9) and (10) show the G^0-distribution proposed by Frery, et. al. [2], for the complex, magnitude and power domains respectively. Where (8) gives the distribution of either I or Q in $w = I + jQ$ and n represents the number of looks.

$$f_{z_0}(x) = \frac{b^\alpha \Gamma(1/2 + \alpha)}{\sqrt{\pi}\,\Gamma(\alpha)(b + x^2)^{1/2+\alpha}} \tag{8}$$

$$\alpha, b > 0, \quad x \in \Re$$

$$f_M(m) = \frac{2n^n b^\alpha \Gamma(n + \alpha) m^{2n-1}}{\Gamma(n)\Gamma(\alpha)(b + nm^2)^{n+\alpha}} \tag{9}$$

$$a, b, n, m > 0$$

$$f_P(p) = \frac{n^n b^\alpha \Gamma(n + \alpha) p^{n-1}}{\Gamma(n)\Gamma(\alpha)(b + np)^{n+\alpha}} \tag{10}$$

$$\alpha, b, n, p > 0$$

For the single look case $(n = 1)$, (10) reduces to (11):

$$f_P(p) = \frac{\alpha b^\alpha}{(b + p)^{1+\alpha}} \tag{11}$$

$$\alpha, b, p > 0$$

---

where $\Gamma(\alpha + 1)/\Gamma(\alpha) = \alpha$. This form for the power domain distribution is also known as the Generalized Beta Prime distribution $gBP(b, \gamma, p, \alpha)$ [1] with parameters $\gamma = p = 1$. We refer to (11) as the Beta Prime distribution $BP(b, \alpha)$. Applying the transformation of variables $g = \eta Log_e(p) + a'$ to this expression, we obtain the $gLG(a', \eta, \eta b, \alpha)$ distribution shown in (12):

$$f_G(g) = \frac{\alpha b^\alpha}{\eta} \frac{e^{\frac{(g-a')}{\eta}}}{(b + e^{\frac{(g-a')}{\eta}})^{1+\alpha}} \tag{12}$$

$$\alpha, b, \eta > 0, \quad g, a' \in \Re$$

Here, η and a' represent the fixed scaling and offset parameters whereas α and b affect the distribution shape and translation respectively. Figures 1 and 2 demonstrate the fit achieved with the Generalized Logistic distribution using ML estimates of α and b. These estimates are discussed in more detail in section VI.
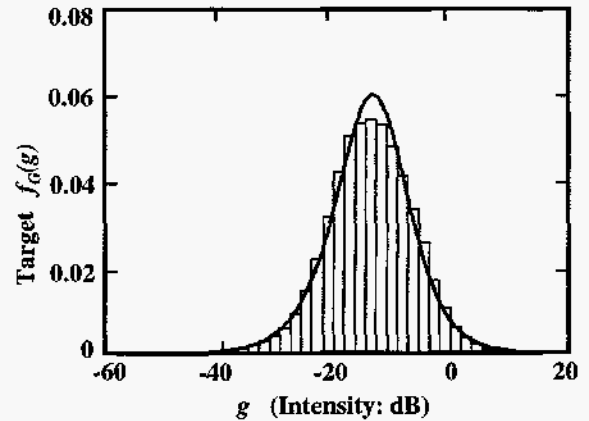

Figure 1. gLG distribution fitted to Log-domain histograms of T-72 Tanks.
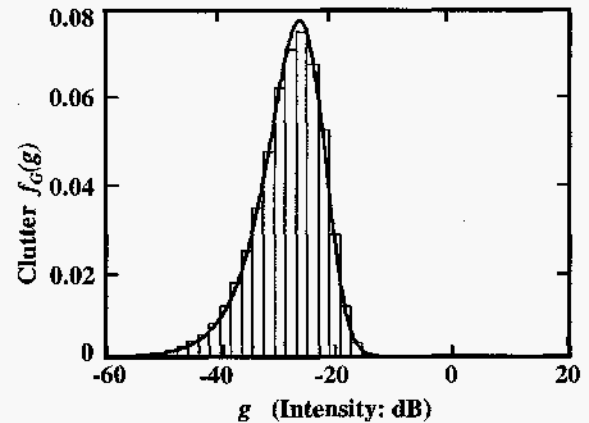

Figure 2. gLG distribution fitted to Log-domain histograms of homogeneous grassy regions.

VI. Parameter Estimation

Unfortunately, closed form parameter estimates for many popular distributions do not exist. For example, in both the Weibull and 3-parameter Lognormal [1] distributions, the ML approach yields a system of equations in which one of the parameters must be estimated numerically. One parameter estimation approach that leads to a closed form solution, is the method-of-moments. This technique has been proposed for the G^0 [2] and Modified Beta (MB) [4] distributions.

The MB distribution is similar to the gBP distribution. However, there are important differences between the two. For example, the MB distribution encompasses only three parameters while the gBP distribution has four. Also, the gBP distribution cannot be transformed to the MB distribution simply by choosing its parameters appropriately. However, an appropriate choice of the MB distribution parameters yields the BP distribution discussed in section V:

$$BP(b, \alpha) = MB(1, \alpha, 1/b) \qquad (13)$$

Given this equivalency, we decided to examine the utility of the MOM approach proposed by Maffett and Wackerman [4]. This approach proposes the following moment combinations for estimating α and b:

$$w = \mu_2 / m^2 \qquad (14)$$

$$g' = \mu_3 / (\mu_2 m) \qquad (15)$$

$$\hat{\alpha} = (4w - 3g' - 2)/(2w - g') \qquad (16)$$

$$\hat{b} = m(\hat{\alpha} - 1) \qquad (17)$$

Here, w and g', defined as the "width" and "modified skewness", are calculated from SAR data. The moments include the sample mean (m), the sample variance (μ_2), and the third central moment (μ_3). Here, w must be greater than zero since the variance and the mean squared are both greater than zero for (13). In addition, there is good reason to believe that μ_3 will also be positive for this distribution [11]. Therefore, g' should be positive as well. However, if we substitute the analytical expressions for m, μ_2, and μ_3 from (11) into (14) and (15) we observe a conflict in the resulting expressions:

$$w = \alpha/(\alpha - 2) \qquad (18)$$

$$g' = 2(\alpha + 1)/(\alpha - 3) \qquad (19)$$

Namely, if α has a value less than two, the positive condition on both (18) and (19) is violated. This result tells us that the estimates produced by (16) and (17) will be invalid for Beta Prime distributed data whose inherent shape parameter has values less than two. This is problematic if we expect SAR data of this type. Here, the MOM approach has limitations that make it inapplicable to our problem.

Alternately, using standard Maximum Likelihood estimation techniques and the marginal distribution described by (11), we can arrive at expressions that give $\hat{\alpha}$ and $\hat{b}$ in terms of the samples from the marginal distribution. We use the Beta Prime distribution instead of the Generalized Logistic, since it is slightly easier to manipulate. However, either expression should give equivalent results, since a monotonically increasing transformation moves one to the other. Equation (20) describes the result of the ML estimate on α:

$$\hat{\alpha} = \left(\frac{1}{N} \sum_{i=1}^{N} Log_e (1 + \frac{p_i}{\hat{b}}) \right)^{-1} \qquad (20)$$

However, a simple solution for the estimate of b in terms of α and the N data samples p_i is not possible as can be seen from (21):

$$f(\hat{b}) = \frac{\hat{\alpha}}{(1 + \hat{\alpha})} - \frac{1}{N} \sum_{i=1}^{N} (1 + \frac{p_i}{\hat{b}})^{-1} \qquad (21)$$

This expression results from the ML estimate on b. However, setting (21) equal to zero and solving for b would require solving an Nth degree polynomial. Instead we chose to use Newton's method to estimate b.

Table 1 shows example results for the ML and MOM approaches using data samples produced with a Monte-Carlo simulation for the Beta Prime distribution given typical parameters. These approaches were applied to 20 sets of $N=100,000$ data samples each. The average estimate is shown below:

Parameters		Avg. ML Est.		Avg. MOM Est.	
α	b	$\hat{\alpha}$	$\hat{b}$	$\hat{\alpha}$	$\hat{b}$
0.5	0.01	0.5008	0.01003	-0.3006	-3110.2
1.0	0.01	1.0012	0.01001	3.1334	0.1834
2.5	0.01	2.5121	0.01006	3.3556	0.0156
3.5	0.01	3.4927	0.00997	3.8338	0.0113
4.5	0.01	4.5057	0.01001	4.5864	0.0102

Table 1. Average parameter estimates for synthetic Beta Prime data.

Clearly, for these examples, the MOM approach is unable to estimate the parameters for Beta Prime distributed data whose α parameter is in the range discussed earlier. Outside this range, the MOM appears to be less accurate than ML. Based on these results and our analytical analysis, we conclude that ML is better at estimating the parameters for our data of interest.

In Table 2, we present example results for the ML and MOM approaches on approximately $N=100,000$ samples from the target and homogeneous grass data described in section IV. Given our earlier analysis, we expect the ML estimates to be more reliable.

	ML Estimate		MOM Estimate	
REGION	$\hat{\alpha}$	$\hat{b}$	$\hat{\alpha}$	$\hat{b}$
Target	1.177	0.0503	3.649	0.4174
Grass	5.902	0.0142	9.362	0.0241

Table 2. Parameter estimates for SAR target and homogeneous grassy regions.

Here, the ML estimates give consistent results for α when the shape of the gLG distribution changes. That is, for $0 < \alpha < 1$, $\alpha = 1$, and $\alpha > 1$ we expect right-skewed, symmetric, and left-skewed distributions respectively. The histograms in Figures 1 and 2 demonstrate this. Here, the target region histogram has a nearly symmetric shape whereas the clutter region histogram has a significantly left-skewed shape. Using these parameter estimates we can implement an 'optimal' detector as shown in Figure 3.
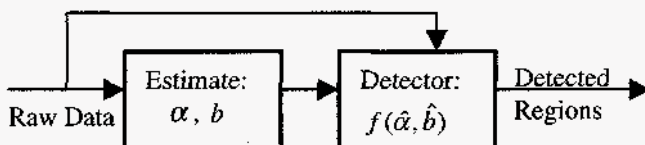

Figure 3. Conceptual detector block diagram.

VII. Conclusions

We have presented the Generalized Logistic distribution as useful for modeling both man-made as well as natural clutter. We have based this model on the multilook G^0-distribution, which can model extremely heterogeneous clutter [2]. This model and the classic K-distribution both stem from the same parent G-distribution. An attractive quality of this distribution is its justification in terms of an established multiplicative model and physical backscattering processes. For the single-look case, this distribution fits homogeneous clutter as well as target data. We can compute the shape and shift parameters from data using Maximum Likelihood estimates augmented with numerical

techniques. This estimation approach has been contrasted with a popular method-of-moments approach, which has problems in some cases. Finally, we have proposed the basis for a detector whose performance can be determined analytically.

Acknowledgement

References

[1] G. P. Patil, M. T. Boswell, and M. V. Ratnaparkhi, *Dictionary of Statistical Distributions in Scientific Work, Vol. 7*, (Fairland, MD; Intl. Co-operative Publishing House, 1984)

[2] A. C. Frery, H. J. Muller, C. C. F. Yanasse, and S. J. S. Sant'Anna, A Model for Extremely Heterogeneous Clutter, *IEEE Transactions on Geoscience and Remote Sensing*, 35(3), 1997, 648-659.

[3] F. T. Ulaby and M. C. Dobson, *Handbook of Radar Scattering Statistics for Terrain*, (Norwood, MA: Atrech House, 1989)

[4] A. L. Maffett and C. C. Wackerman, The modified beta density function as a model for synthetic aperture radar clutter statistics, *IEEE Trans. Geosci. Rem. Sensing*, 29(2), 1991, 277-283.

[5] J. K. Jao, Amplitude Distribution of Composite Terrain Radar Clutter and the K-Distribution, *IEEE Trans. Antennas & Propagation*, AP-32 (10), 1984, 1049-1061

[6] E. Jakeman and P. N. Pusey, A model for non-Rayleigh sea echo, *IEEE Trans. Antennas & Propagation*, AP-24, 1976, 806-814

[7] C. C. F. Yanasse, A. C. Frery, S. J. S. Sant'Anna, L. V. Dutra, On the Use of Multilook Amplitude K Distribution for SAR Image Analysis, *Presented at IGARS '94: Technologies, Data Analysis and Interpretation*, 4, 1994, 2173-2175

[8] S. H. Yueh, J. A. Kong, J. K. Jao, R. T. Shin, and L. M. Novak, K-distribution and polarimeteric terrain radar clutter, *J. Electromag. Waves Appl.*, 3(8), 1989, 747-768

[9] R. J. Fogler, L. D. Hostetler, D. R. Hush, SAR Clutter Suppression Using Probability Density Skewness, *IEEE Trans. Aerospace Electronic Sys.*, 30(2), 1994, 622-626

[10] S. D. Dubey, A New Derivation of the Logistic Distribution, *Naval Logistics Quarterly*, 16(1), 1969, 37-40

[11] E. S. Keeping, *Introduction to Statistical Inference*, (Princeton, NJ; D. Van Nostrand Company Inc., 1962)