

# Quantitative Statistical Assessment of Conditional Models for Synthetic Aperture Radar

Michael D. DeVore, *Member, IEEE*, and Joseph A. O'Sullivan, *Fellow, IEEE*,

**Abstract**—Many applications of object recognition in the presence of pose uncertainty rely on statistical models—conditioned on pose—for observations. The image statistics of three-dimensional (3-D) objects are often assumed to belong to a family of distributions with unknown model parameters that vary with one or more continuous-valued pose parameters. Many methods for statistical model assessment, for example the tests of Kolmogorov–Smirnov and K. Pearson, require that all model parameters be fully specified or that sample sizes be large. Assessing pose-dependent models from a finite number of observations over a variety of poses can violate these requirements. However, a large number of small samples, corresponding to unique combinations of object, pose, and pixel location, are often available. We develop methods for model testing which assume a large number of small samples and apply them to the comparison of three models for synthetic aperture radar images of 3-D objects with varying pose. Each model is directly related to the Gaussian distribution and is assessed both in terms of goodness-of-fit and underlying model assumptions, such as independence, known mean, and homoscedasticity. Test results are presented in terms of the functional relationship between a given significance level and the percentage of samples that would fail a test at that level.

**Index Terms**—Conditional models, MSTAR, statistical hypothesis testing, statistical model assessment, synthetic aperture radar.

## I. INTRODUCTION

**R**ECOGNITION of objects that are subject to pose variation is often accomplished through the use of conditional statistical models. These models are defined in terms of parameters which are themselves functions of image contents (object class), relative pose, and pixel location. While a number of methods for quantitative assessment of statistical models exist, such as Kolmogorov–Smirnov or Pearson's  $\chi^2$  tests, these generally require either known model parameters or samples which are large enough that asymptotic distributions apply. These methods are highly applicable to studies of marginal pixel distributions where a single image may yield thousands of observations. However, the assessment of conditional image models is complicated by the fact that the distribution statistics differ with each unique combination of class, pose, and pixel

location. Further, as pose is generally continuous-valued, we are unable to obtain a large number of observations from each possible pose, and, even if pose is quantized, it may be prohibitive to obtain a sufficient number of images from each interval.

In the absence of large sample sizes, we must consider alternate methods for assessing conditional models for observed imagery and for comparing alternate model assumptions. One approach is to derive inference algorithms based on assumed models and compare their resulting performance. For example, DeVore and O'Sullivan [9] compare conditional models for synthetic aperture radar (SAR) images in terms of percent classification error as a function of model complexity. Such an approach evaluates model assumptions in light of application specific limitations, such as the model complexity one is willing to support and the amount of training data available. For example, as discussed in [9], the results may suggest a simpler but less accurate model if the training set is small.

In this paper, we consider model assessment by testing a large number of small samples, with the results aggregated for analysis of the overall conditional model. In short, an exact small-sample test is constructed for each model assumption, and the test is applied to a large number of samples. The result is a collection of observed P-values which should be uniformly distributed if the model assumptions were accurate. Deviations from uniformity can be interpreted as evidence against the model assumptions and can be used both to evaluate a given model and to compare alternate models. By comparison, methods for model selection, such as Rissanen's minimum description length [26] or Akaike's information criterion [1], seek the model from a specified set that best describes observed data, irrespective of whether any model is actually a good fit to the data. More information on statistical models for SAR and their evaluation can be found in [8].

Our focus is on complex Gaussian (Rayleigh magnitude), lognormal, and quarter-power normal conditional distribution families for SAR imagery and on common assumptions that are made to simplify modeling and manipulation, including pixel independence, zero mean, and homoscedasticity (equal variance). These analyses are performed using 6862 SAR images from ten targets and are performed separately for pixels corresponding to man-made vehicles and background clutter. These distribution families were selected because each has a history of application by researchers to radar measurements. Further, each has a direct relationship to the Gaussian distribution, which allows tests with identical powers to be applied to each and the test results can be compared directly. The underlying methodology can be applied to other distribution families as

Manuscript received November 12, 2002; revised August 18, 2003. This work was supported in part by the Office of Naval Research under Grant N00014-03-1-0110, and by the Boeing Foundation. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Thierry Blu.

M. D. DeVore is with the Systems and Information Engineering Department, University of Virginia, Charlottesville, VA 22903 USA (e-mail: mdevore@virginia.edu).

J. A. O'Sullivan is with the Electronic Systems and Signals Research Laboratory, Department of Electrical Engineering, Washington University, St. Louis, MO 63130 USA (e-mail: jao@ee.wustl.edu).

Digital Object Identifier 10.1109/TIP.2004.823825

well, with appropriate modifications. However, the resulting powers will not necessarily be equal to those of the tests presented here, so the results will not be directly comparable. In essence, tests for other distribution families will involve different test statistics and will result in a slightly different measure of deviation from model assumptions.

Many papers have addressed the evaluation of statistical models for the distribution of radar returns in which the entire target fits within a resolution cell, as opposed to the SAR imaging we consider here which can resolve one-foot square subregions on a target. Edrington [10] reports empirical distributions, approximately Rayleigh in appearance, of return magnitudes from aircraft flying toward an X-band radar. Reilly [25] considers the degrees of freedom of a  $\chi^2$  distribution estimated to fit the radar return from a variety of targets as a function of observation time. Nathanson [22, p. 171] cites a series of early experiments by the Naval Research Laboratory in which the distribution of radar return over a range of angles was measured for a variety of aircraft, and the Rayleigh distribution appeared to be a good fit for large aircraft with multiple engines. Each of these works relies heavily on graphical methods to compare empirical observations with those hypothesized.

There have been several recent assessments of the marginal distributions of pixels in SAR imagery. Shnidman [28] considers a model for pixel squared-magnitudes with a three-parameter mixture of noncentral  $\chi^2$  distributions with noncentrality parameter governed by a gamma distribution and graphically demonstrates the fit to clutter imagery. Billingsly *et al.* [3] assess the fit of Rayleigh, Weibull, lognormal, and K-distributions to pixel magnitudes in clutter data and show via the Kolmogorov–Smirnov test that none fit well over the entire range of magnitudes. Kuruoglu and Zerubia [17] graphically demonstrate the fit of  $\alpha$ -stable distributions to pixel magnitudes. These differ from the investigation of this paper in their focus on marginal distributions across all points in a scene and across all contents of the scene.

Two recent papers in particular focus on the assessment of conditional distributions for high-range resolution (HRR) radar and SAR measurements. Holt, *et al.* [14] consider the fit of a  $\chi^2$  distribution to HRR data collected from 1/16 scale targets from a wide range of azimuth angles. They group range cell data having similar estimated degrees of freedom and, after normalizing by the respective estimated scale parameters, show plots of empirical probability density functions along with best fit distribution. Kaplan [15] addresses the fit of conditional Rayleigh, Weibull, and lognormal distributions. He considers a deterministic signal in multiplicative noise and divides each observation by the estimated signal component yielding a quantity that has Rayleigh distributed magnitude if the corresponding model were correct and if the estimated signal component were exact. He indicates that, based on Kolmogorov–Smirnov tests, the Weibull and Rayleigh models are good matches across all pixels in the images and the lognormal model is a poor fit for clutter but a good fit for target pixels. However, the use of estimated parameters with the Kolmogorov–Smirnov test can yield overly optimistic results even for large sample sizes [29, p. 400].

In light of these previous efforts, our intent in this paper is to present a method for assessing conditional image models in which large numbers of small samples are available. The method is quantitative, drawing on established statistical procedures commonly applied to large sample problems. This method is consistently applied through tests of equal power to allow direct comparison of the results across model families and scene composition, such as target or clutter. Further, we construct a variety of tests, each sensitive to different model assumptions such as distribution family, independence, zero complex mean, and homoscedasticity.

In Section II, we describe three SAR image models that are the focus of this paper. In Section III, we describe the sample data on which the analyses are based and develop a framework for aggregating the large numbers of results from statistical tests to quantify the correctness of various hypotheses. In Section IV, we address issues of testing the fit of hypothesized distributions when estimated parameters are involved and apply three separate goodness-of-fit tests to the sample data. The hypothesis that pixels in a SAR image are conditionally independent given the parameters that characterize their distribution is assessed in Section V. The zero-mean hypothesis of the complex Gaussian model is assessed in Section VI. Finally, in Section VII, we consider tests of homoscedasticity and apply these to the equal variance assumption in both the lognormal and quarter-power normal models.

## II. MODELS FOR SAR IMAGES

In this section, we address stochastic models for synthetic aperture radar imagery. These models are mathematical abstractions incorporating a variety of assumptions and approximations and are intended to adequately characterize the radar observations as a function of image pixel or corresponding resolution cell, scene contents, and target pose.

### A. Complex Gaussian Model

The conditionally complex-Gaussian model for SAR imagery is based on the assumption that the region in the scene corresponding to each pixel  $i$  contains a large number of independent scattering centers producing a return with independent magnitude and phase. The complex reflections from each scatterer combine at the SAR processor where they are further combined with additive complex white Gaussian noise. The central limit theorem suggests that the return  $R_i$  in the  $i$ th pixel may be modeled well by a complex-Gaussian random variable with independent real and imaginary components of equal variance [13], [30].

The statistical properties of  $R_i$  are functions of the region in the scene contributing to pixel  $i$ , and that region is a function of the scene contents  $a$ , relative pose  $\psi$ , and the pixel location  $i$ . If the image pixels correspond to nonoverlapping regions of the scene, we can model their values as independent. We call this the nonzero-mean conditionally Gaussian model, and the probability density for SAR image  $\mathbf{R}$  can be written as

$$p(\mathbf{r}|a, \psi) = \prod_i \frac{1}{\pi \sigma_i^2(a, \psi)} e^{-\frac{|r_i - \mu_i(a, \psi)|^2}{\sigma_i^2(a, \psi)}}. \quad (1)$$

If the phase associated with each scatterer is uniformly distributed, then the mean vector can be modeled as identically zero,  $\boldsymbol{\mu}(a, \boldsymbol{\psi}) = \mathbf{0}$ . The resulting zero-mean conditionally Gaussian model has a single parameter function, and the probability density for  $\mathbf{R}$  can be written as

$$p(\mathbf{r}|a, \boldsymbol{\psi}) = \prod_i \frac{1}{\pi \sigma_i^2(a, \boldsymbol{\psi})} e^{-\frac{|r_i|^2}{\sigma_i^2(a, \boldsymbol{\psi})}}. \quad (2)$$

In this case, the magnitude  $|R_i|$  is a Rayleigh distributed random variable which leads to this model being referred to as a Rayleigh model in some literature.

### B. Lognormal Magnitude Model

Under the lognormal model, the magnitude of each pixel value in a SAR image is said to follow a lognormal distribution [24, p. 183] with parameter values that depend on the pixel location  $i$ , scene contents  $a$ , and pose  $\boldsymbol{\psi}$ . The natural logarithm of a lognormal random variable is a Gaussian random variable with mean  $\mu$  and variance  $\sigma^2$ . We exploit this relationship by transforming the received SAR image  $\mathbf{R}$  yielding an image  $\mathbf{X}$  where  $X_i = \ln |R_i|$ . Following the assumption that pixels correspond to nonoverlapping regions, the probability density function for  $\mathbf{X}$  is

$$p(\mathbf{x}|a, \boldsymbol{\psi}) = \prod_i \frac{1}{\sqrt{2\pi\sigma_i^2(a, \boldsymbol{\psi})}} e^{-\frac{(x_i - \mu_i(a, \boldsymbol{\psi}))^2}{2\sigma_i^2(a, \boldsymbol{\psi})}} \quad (3)$$

where  $\mu_i(a, \boldsymbol{\psi})$  and  $\sigma_i^2(a, \boldsymbol{\psi})$  are the mean and variance, respectively, of the  $i$ th pixel when scene contents  $a$  are viewed from pose  $\boldsymbol{\psi}$ . Some researchers have approximated the variance function  $\sigma_i^2(a, \boldsymbol{\psi})$  as constant over all pixel locations and scenes. Under this approximation, the same density function holds and we simply write  $\sigma_i^2(a, \boldsymbol{\psi}) = \sigma^2$ .

### C. Quarter-Power Normal Magnitude Model

The quarter-power normal model [11], [31] approximates the square-root of the magnitude of SAR pixel values as Gaussian random variables conditioned on pixel location  $i$ , scene contents  $a$ , and pose  $\boldsymbol{\psi}$ . The model is only approximate, as the probability is zero that the magnitude square-root takes on negative values, but all Gaussian random variables take negative values with finite probability. To make this model tractable, we transform a received SAR image  $\mathbf{R}$  yielding an image  $\mathbf{X}$ , where  $X_i = |R_i|^{1/2}$ . The pixel values  $X_i$  are then approximated as Gaussian random variables with mean and variance that are functions of  $i$ ,  $a$ , and  $\boldsymbol{\psi}$ . Following the assumption that pixels correspond to nonoverlapping regions of the scene, the probability density function becomes

$$p(\mathbf{x}|a, \boldsymbol{\psi}) = \prod_i \frac{1}{\sqrt{2\pi\sigma_i^2(a, \boldsymbol{\psi})}} e^{-\frac{(x_i - \mu_i(a, \boldsymbol{\psi}))^2}{2\sigma_i^2(a, \boldsymbol{\psi})}} \quad (4)$$

As in the lognormal model, the variance is sometimes assumed to be constant, and we denote this case by  $\sigma_i^2(a, \boldsymbol{\psi}) = \sigma^2$ .

TABLE I  
MSTAR DATASET USED IN STATISTICAL ANALYSIS. IN TOTAL, 6862 IMAGES OF TEN TARGETS AT TWO DEPRESSION ANGLES ARE USED

Target	Serial Number	15° Images	17° Images
2S1	b01	274	299
BMP-2	9563, 9566, c21	587	697
BRDM-2	E-71	263	298
BTR-60	k10yt7532	195	256
BTR-70	c71	196	233
D7	92v13015	274	299
T62	A51	273	299
T-72	132, 812, s7	582	691
ZIL131	E12	274	299
ZSU 23.4	d08	274	299

## III. CONDITIONAL MODEL ASSESSMENT

Assessment of the assumptions in a conditional model is conducted through a collection of hypothesis tests. Tests are conducted for each combination of conditioning values. We first describe the sample data and then describe the approach to hypothesis testing and aggregation of the results of multiple tests.

### A. The MSTAR Dataset

The methods of model assessment are applied to X-band SAR data collected as part of the Moving and Stationary Target Acquisition and Recognition (MSTAR) program which was supported under funding from the Defense Advanced Research Projects Agency (DARPA).<sup>1</sup> The images in the dataset were collected with HH-polarization and have 0.3-m resolution and 0.2-m pixel spacing in both range and cross range. We utilize only the publicly available data from that collection. The image set contains data for a number of military targets imaged at two or more depression angles. For each target and depression angle, the set contains SAR image data for 200–300 values of target aspect pose. Table I contains a listing of the vehicle types included and the number of images available for each type.

Each image contains a single target which, along with its shadow, lies very near the center of the image, and the remainder of the image contains background clutter. Sample log-magnitude images formed from 15° depression and 45° azimuth of the T-72, BTR-70, and D7 are shown in Fig. 1. Model fitness assessments are performed separately on vehicle and clutter pixels by restricting the tests to 20 × 20 pixel regions centered in the images and in the upper left-hand corner, respectively. The 400 pixels at the center of the images were chosen because they are consistently on the target. The pixels on clutter were chosen to represent a patch the same size and shape as the pixels on target so both sets of tests will have the same power and the results can be compared directly. Since we do not have multiple samples from any particular pose, we make the assumption that the

<sup>1</sup>The MSTAR dataset can be requested through the Sensor Data Management System (SDMS) Web page of Wright Laboratory at <http://www.mbvlab.wpafb.af.mil/public/sdms/>

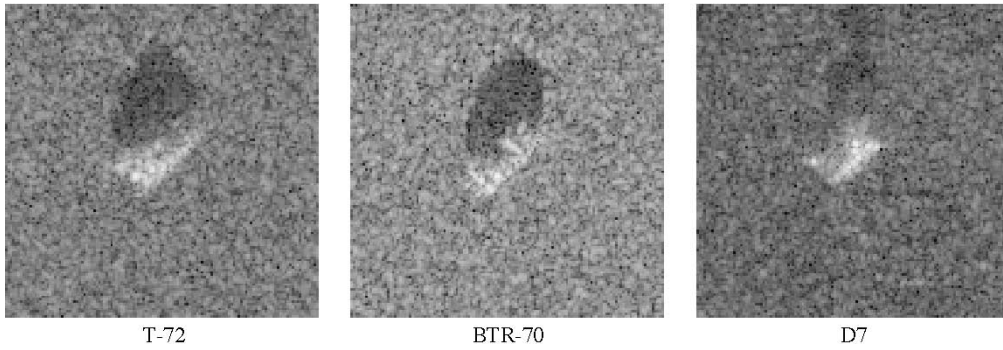


Fig. 1. Sample images from the MSTAR dataset.

change in parameter value over a  $12^\circ$  azimuth interval is small. We examine data from 60 intervals of  $12^\circ$  evenly spaced around  $360^\circ$  of azimuth. The tests are applied to the observations for each resulting combination of pixel location, azimuth interval, and vehicle class.

### B. Hypothesis Testing

We form all hypothesis tests in terms of probability distributions  $f(\mathbf{r}; \boldsymbol{\omega})$  on a collection of observed data  $\mathbf{r} = \{\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N\}$  where  $\boldsymbol{\omega}$  parameterizes the distribution. In general,  $\boldsymbol{\omega}$  can specify the family of the distribution as well as the distribution parameters. The set of all distribution parameters that are consistent with a given model assumption are denoted by  $\Omega$ . We formulate each test as a binary hypothesis testing problem

$$\mathcal{H}_0 : r \sim f(\cdot; \boldsymbol{\omega} \in \Omega) \quad \text{versus} \quad \mathcal{H}_A : r \sim f(\cdot; \boldsymbol{\omega} \in \Omega^c) \quad (5)$$

where  $\Omega^c$  denotes the complement of the parameter set  $\Omega$ .

For each test, we define a continuous-valued test statistic  $D(\mathbf{r})$  that forms the basis of selecting between the two hypotheses. We define  $D$  in such a manner that large values suggest evidence against  $\mathcal{H}_0$ . The value of  $D$  is a random variable and the statistic is chosen so that the conditional distribution  $F_D(d|\mathcal{H}_0)$  is known. Let  $D = D(\mathbf{r})$  be an observation of the test statistic, then  $1 - F_D(D|\mathcal{H}_0)$  is called the P-value of the test for the observation  $\mathbf{r}$ . The P-value for an observation is equal to the smallest significance level  $\alpha$  that will result in a rejection of  $\mathcal{H}_0$  for the given observation. Extremely small P-values are taken as evidence that the data do not represent a sample drawn from the distribution governing  $\mathcal{H}_0$ . There is no component of such a test that directly recommends  $\mathcal{H}_0$ ; however, large P-values indicate a lack of evidence to support rejecting it.

A P-value quantifies the result of a single test. We are interested in assessing models that are functions of target class, target pose, and pixel location and we conduct a separate test for each combination based on the corresponding pixel values. We therefore need a mechanism by which the various test results can be aggregated in a statistically sound manner to yield a useful overall result. Since the power of the tests, the probability of correctly rejecting  $\mathcal{H}_0$ , generally depends on the number of samples which is variable, the method of aggregation must accommodate tests of varying powers.

A solution is available in the form of the probability integral transform of continuous random variables. Let  $X$  be a continuous random variable taking values in  $\mathfrak{R}$  with cumulative distribution function  $F_X(x)$ . Let  $F_X^{-1}(p)$  denote the inverse cumulative distribution defined by

$$F_X^{-1}(p) = \inf \{x \in \mathfrak{R} : F_X(x) \geq p\}. \quad (6)$$

The probability integral transform of  $X$  yields the uniformly distributed random variable  $Y = F_X(X)$  [19, p. 58]. Thus, for a continuous-valued test statistic,  $F_D(D|\mathcal{H}_0)$  is uniformly distributed on  $[0,1]$  and the P-value  $1 - F_D(D|\mathcal{H}_0)$  is also uniform on  $[0,1]$ .

Powerful tests, tests highly effective at correctly rejecting  $\mathcal{H}_0$ , are characterized by a distribution of P-values that is sharply skewed toward zero when  $\mathcal{H}_A$  is true. In essence, most observations obtained under  $\mathcal{H}_A$  would appear to be low-probability events under  $\mathcal{H}_0$ . We can combine the results of many independent hypothesis tests by examining the empirical distribution of P-values. The empirical cumulative distribution of a sample  $z_1, z_2, \dots, z_n$  is an estimate of the cumulative distribution function given by

$$\hat{F}_Z(z) = \frac{1}{n} \sum_{i=1}^n u(z - z_i) \quad (7)$$

where  $u(z)$  is the Heaviside function. The empirical cumulative distribution of P-values shows the fraction of samples that would fail a hypothesis test with significance level  $\alpha$  as a function of  $\alpha$ . Evidence to support rejecting the model approximations and assumptions of  $\mathcal{H}_0$  can be quantified by the deviation of the empirical distribution of P-values from a uniform distribution.

## IV. GOODNESS-OF-FIT

In this section, we look at several tests which can be used to assess the evidence that a particular distribution family fits a given set of sample data. All of the models discussed in Section II involve the Gaussian distribution. There are mixed views represented in the literature as to which tests for the Gaussian distribution are the most powerful. We, therefore, employ three tests for goodness-of-fit, each of which relies on differing properties of Gaussian random variables. K. Pearson's  $\chi^2$  test, appropriate for categorical data or binned continuous-valued data, is an approximate test based on the number of random

observations in each category. It effectively tests whether the histogram of sample data is reasonable under the assumed distribution. The Kolmogorov–Smirnov test, appropriate for continuous-valued ordinal data, is based on the largest magnitude difference between the empirical cumulative distribution function and the assumed cumulative distribution. The test of R. D'Agostino and E. Pearson is a test specifically for departures from normality based on the distribution of sample skewness and kurtosis values.

We first consider the derivation of T distributed random variables with known statistics from a collection of real-valued Gaussian random variables with unknown statistics. The T distribution is often useful in making inferences on the sample mean of a population drawn from a Gaussian distribution. Our interest is in inference directly on pixel values rather than their statistics. From each observation in a sample, we generate a random variable with known distribution that is free from unknown parameters. That is, from a sample of  $n$  Gaussian random variables with unknown mean and/or variance we will generate  $n$  T distributed random variables with known parameters. Evidence that the transformed values are not T distributed can be taken as evidence that the original values were not Gaussian. This lack of unknown parameters will be useful for the  $\chi^2$  hypothesis test of Section IV-B and is necessary for the Kolmogorov–Smirnov hypothesis test of Section IV-C.

#### A. T-Conversion

A random variable  $T$  which can be represented in terms of a real, standard normal random variable  $X$ , and an independent  $\chi^2$  random variable with  $\gamma$  degrees of freedom  $\chi_\gamma^2$ , as

$$T = \frac{X}{\sqrt{\frac{\chi_\gamma^2}{\gamma}}} \quad (8)$$

is said to be T distributed with  $\gamma$  degrees of freedom. In many practical applications of the T distribution, the numerator and denominator are formed from mean and variance estimates of Gaussian random variables [4]. Motivated by these approaches, we seek to transform each observation from a sample into a T distributed random variable using an estimate of variance that is independent of the observation. In Section IV-A-1, we consider the formation of such random variables from a set of Gaussian random variables with known mean. In Section IV-A-2, we consider a similar problem with unknown mean.

1) *Known Mean:* Let  $X_1, X_2, \dots, X_n$  be independent, identically distributed, real Gaussian random variables with known mean  $\mu$  and unknown variance  $\sigma^2$ . We want to construct a T-distributed random variable from each sample  $X_i$  and an estimate of the variance. We can form an estimate of the variance that is independent of any particular  $X_i$  by defining

$$\xi_{(i)}^2 = \frac{1}{n-1} \sum_{j \neq i} (X_j - \mu)^2. \quad (9)$$

Then the ratio

$$T_i = \frac{X_i - \mu}{\xi_{(i)}} \quad (10)$$

is T distributed with  $n - 1$  degrees of freedom.

For application to the MSTAR dataset, the values at each pixel location are transformed as described above. For all 400 pixel locations in a given azimuth window, the transformation yields samples of a fully known distribution. Samples corresponding to the same pixel location are not independent, but all pixel locations yield samples with the same marginal distribution. Since each model for SAR imagery assumes that values in differing pixel locations are independent, randomly chosen collections of the transformed samples represent approximately independent observations of a T-distributed random variable.

2) *Unknown Mean:* Let  $X_1, X_2, \dots, X_n$  be independent, identically distributed, real Gaussian random variables with unknown mean  $\mu$  and unknown variance  $\sigma^2$ . We want to construct a T-distributed random variable from each sample  $X_i$ , estimates of the mean, and estimates of the variance. The random variables  $X_i$  and  $\bar{X} = (1/n) \sum_{i=1}^n X_i$  are jointly Gaussian. The difference  $X_i - \bar{X}$  has mean zero and variance  $(n - 1/n)\sigma^2$ .

The quantity  $X_i - \bar{X}$  is not independent of the usual variance estimator  $S_{n-1}^2$ . As an alternative, we use the variance estimate  $S_{(i)}^2$ , where

$$\bar{X}_{(i)} = \frac{1}{n-1} \sum_{j \neq i} X_j \quad \text{and} \quad S_{(i)}^2 = \frac{1}{n-2} \sum_{j \neq i} (X_j - \bar{X}_{(i)})^2. \quad (11)$$

Then the random variable  $T_i$  given by

$$T_i = \frac{(X_i - \bar{X})}{\sqrt{\frac{n-1}{n} S_{(i)}^2}} \quad (12)$$

is T distributed with  $n - 1$  degrees of freedom.

As in the known-mean case, the  $n$  random variables obtained in this way are not independent, so we will independently generate many such sequences with the same T distribution and samples across the sequences will be independent under the respective SAR model.

#### B. Pearson's $\chi^2$ Method

Suppose that a random experiment is conducted that has  $k$  possible outcomes,  $\omega_1, \omega_2, \dots, \omega_k$ , and that each outcome occurs with probability  $p_1, p_2, \dots, p_k$ . Given  $n$  independent trials of the random experiment, the number of outcomes of each type  $[N_1, N_2, \dots, N_k]^T$  is said to have a multinomial distribution with parameters  $n, p_1, p_2, \dots, p_k$ . It can be shown [19, p. 123] that for a multinomial random variable the test statistic

$$D^2 = \sum_{i=1}^k \frac{(N_i - np_i)^2}{np_i} \quad (13)$$

is asymptotically  $\chi^2$  distributed with  $k - 1$  degrees of freedom as  $n$  approaches infinity.

This limiting distribution can be applied to goodness-of-fit testing. Suppose that under the null-hypothesis, the data are distributed according to  $F(x)$ , a completely specified distribution of either a discrete, continuous, or mixed random variable. If the domain of possible values of the random variable  $X$  is partitioned into  $k$  regions with the probability integral over the  $i$ th

TABLE II  
COMPARISON OF EMPIRICAL CRITICAL VALUES OF SKEWNESS AND KURTOSIS  
WITH THOSE REPORTED BY ZAR FOR THE CASE OF UNKNOWN MEAN

$n$	$g_1$				$g_2$			
	$\alpha = 0.05$		$\alpha = 0.005$		$\alpha = 0.05$		$\alpha = 0.005$	
	Zar	Table	Zar	Table	Zar	Table	Zar	Table
9	1.176	1.183	1.909	1.935	N/A	2.775	N/A	5.262
10	1.125	1.132	1.846	1.874	N/A	2.633	N/A	5.186
20	0.836	0.836	1.412	1.420	1.850	1.870	4.121	4.157
100	0.396	0.395	0.646	0.643	0.889	0.878	1.820	1.820
500	0.180	0.180	0.284	0.284	0.391	0.386	0.700	0.7034

region equal to  $p_i$ , then the counting vector  $[N_1, N_2, \dots, N_k]^T$ , where  $N_i$  is the number of samples of  $X$  with values in region  $i$ , follows a multinomial distribution with parameters  $n, p_1, p_2, \dots, p_k$  under the null-hypothesis. Large values of  $D^2$  indicate that the histogram of the observed data does not match the expected histogram under  $q$  and suggest rejecting  $\mathcal{H}_0$ . The probability that  $D^2$  will exceed the observed value  $d^2$  is approximately equal to the right-tailed probability

$$\Pr[D^2 > d^2] \approx 1 - F(d^2) \quad (14)$$

where  $F$  is the cumulative distribution of a  $\chi^2$  random variable with  $k - 1$  degrees of freedom. This probability is reported as the P-value of the test. Small P-values for the  $D^2$  statistic of (13) indicate evidence that  $X$  is not distributed according to the null-hypothesis. If the distribution under the null-hypothesis is not known exactly but instead depends upon  $l$  parameters which must be estimated from the sample data, then substituting these  $l$  estimated parameters results in  $D^2$  being  $\chi^2$  distributed with  $n - l - 1$  degrees of freedom.

For continuous or mixed random variables, this implies a binning of continuous values into regions which must be arbitrarily chosen. A commonly quoted suggestion (cf. [27, p. 148]) is that none of the expected values  $np_i$  should be less than 1 and no more than 20% should be less than 5. For the experiments we consider, we choose to partition the domain of  $X$  such that each partition is equiprobable with  $np_i = 5$ . The number of MSTAR sample images in a  $12^\circ$  range azimuth are not sufficient to support a partition of more than two such bins. Therefore, the samples within an azimuth interval are first transformed via the methods discussed in Section IV-A. The transformed samples are then tested to see if they deviate from the expected T distribution.

### C. Method of Kolmogorov–Smirnov

The Kolmogorov–Smirnov test is based upon the empirical cumulative distribution function of the observed data. The method assumes a continuous distribution which must be completely specified under the null-hypothesis. The Kolmogorov statistic  $D_{KS}$  is defined as the supremum of the magnitude difference between the empirical cumulative distribution  $\hat{F}_X(x)$  and the cumulative distribution under  $\mathcal{H}_0$ ,  $F_X(x)$ . That is

$$D_{KS} = \sup_x \left| \hat{F}_X(x) - F_X(x) \right|. \quad (15)$$

This maximum distance can be written in terms of  $F(X_i)$ , a function of the random samples which is uniformly distributed for all continuous distributions under the null-hypothesis. An exact expression for the limiting distribution of  $D_{KS}$  under  $\mathcal{H}_0$  as  $n \rightarrow \infty$  was derived by Kolmogorov and can be approximated numerically as explained in [23] for  $n$  as small as 4. Large observed values  $d$  of the statistic  $D_{KS}$  suggest evidence that  $\mathcal{H}_0$  should be rejected and the probability  $\Pr[D_{KS} > d]$  is reported as the P-value of the test. Since the exact distribution of the variables must be known in order to apply this test, the samples at each pixel location are transformed as discussed in Section IV-A. The transformed samples are then tested to see if they deviate from the expected T distribution.

### D. Method of D’Agostino–Pearson

D’Agostino and Pearson [7] suggest a test for departure from normality based upon estimates of the skewness and kurtosis from the sample data. In their commentary, D’Agostino, *et al.* [6] claim that the D’Agostino–Pearson test is much superior to the  $\chi^2$  and Kolmogorov tests and further recommend that these latter not be used at all in tests for normality.

Fisher defined skewness and kurtosis in terms of cumulants as  $\gamma_1 = (\kappa_3/\kappa_2^{3/2})$  and  $\gamma_2 = (\kappa_4/\kappa_2^2)$ , respectively, where  $\kappa_i$  is the  $i$ th cumulant of the distribution. The sample Fisher skewness and kurtosis are the ratios  $g_1 = (k_3/k_2^{3/2})$  and  $g_2 = (k_4/k_2^2)$ , where  $k_i$  is the  $i$ th k-statistic defined as the unbiased symmetric estimator of  $\kappa_i$  [16].

Under the hypothesis of a Gaussian distribution, the sample skewness and kurtosis  $g_1$  and  $g_2$  can be transformed into the Gaussian random variables  $Z_{g_1}$  and  $Z_{g_2}$  by either an approximate transform as discussed by [32], or through the cumulative distributions  $F_{G_1}(g_1)$  and  $F_{G_2}(g_2)$  as  $Z_{g_1} = \Phi^{-1}(F_{G_1}(g_1))$  and  $Z_{g_2} = \Phi^{-1}(F_{G_2}(g_2))$ , where  $\Phi^{-1}(z)$  is the inverse cumulative distribution function for a standard normal random variable. The test statistic for the D’Agostino–Pearson test is defined in terms of these functions as

$$D_{DP} = [\Phi^{-1}(F_{G_1}(g_1))]^2 + [\Phi^{-1}(F_{G_2}(g_2))]^2. \quad (16)$$

Large values of the test statistic suggest evidence that  $\mathcal{H}_0$  should be rejected and the probability  $\Pr[D_{DP} > d]$  is reported as the P-value of the test. For application to SAR image data, the samples from each pixel location of each azimuth interval are used to generate a P-value for  $Z_{g_1}^2 + Z_{g_2}^2$ .

In order to determine the P-value above, we need to know the functions  $F_{G_1}$ ,  $F_{G_2}$ , and  $F_{D_{DP}}$ . Zar [32] provides an extensive table of critical values for  $g_1$  and  $g_2$  based upon normal approximations of  $F_{G_1}$  in [5] and  $F_{G_2}$  in [2]. He states that the approximation of  $F_{G_1}$  is accurate for  $n \geq 9$  and that of  $F_{G_2}$  for  $n \geq 20$ . The skewness and kurtosis estimates are not independent, so the statistic  $D_{DP}$  does not follow a  $\chi^2$  distribution, though it is a good approximation for large samples. Zar cites personal communication with D’Agostino in which he claims that a  $\chi^2$  test for  $D_{DP}$ , introduced in [7], works well for  $n \geq 20$ .

In our investigation of SAR images, however, we will need the distributions for as few as  $n = 4$  samples, so the approximations employed in those papers will not apply.

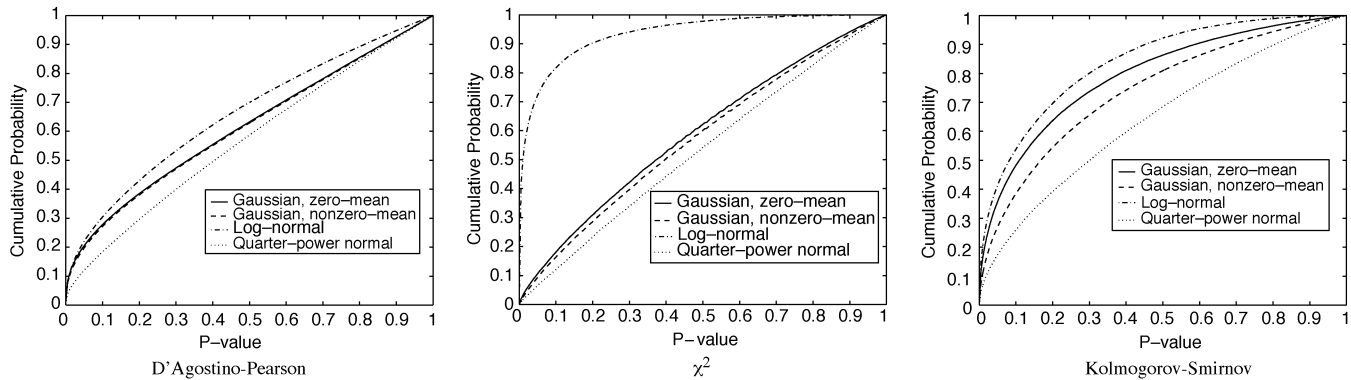


Fig. 2. Cumulative distribution of P-values for various goodness-of-fit tests applied to pixels corresponding to target regions in the scene.

Because of this, and since in each paper previously mentioned Monte-Carlo simulation was ultimately used as the standard by which the approximations were compared, we employ simulation directly to approximate the cumulative distributions. We generate empirical cumulative distributions of  $G_1$ ,  $G_2$ , and  $D_{DP}$  from 1 000 000 Gaussian samples of size  $n$  and save these in a look-up table of approximate critical values. Tables are created for both known and unknown means of the underlying Gaussian samples and for sample sizes  $n \in \{4, 5, \dots, 100, 110, \dots, 300, 325, \dots, 700\}$  using the Gaussian random number generator in Matlab version 5.3.

Table II shows a comparison of some critical values for the distributions of  $g_1$  and  $g_2$  published by Zar [32] and those obtained from our empirical distributions. The critical values in the table are the value  $F^{-1}(1 - \alpha)$ , where  $F$  is the cumulative distribution of either  $g_1$  or  $g_2$ , as indicated by the column heading, for the number of samples indicated by the row heading. The table indicates that our empirical distribution functions are quite similar to the approximations of Zar in the range where his are applicable, and it suggests that the empirical distributions are generally applicable to goodness-of-fit testing.

### E. Goodness-of-Fit Results

In this section, we present the results of applying the goodness-of-fit tests for the complex Gaussian, lognormal, and quarter-power normal models to the MSTAR dataset. For tests of the conditionally Gaussian model, which explicitly models both the real and imaginary components of pixel values, each component is tested separately and the resulting P-values are combined so that the tests of each model will have approximately the same powers. Both the zero-mean and nonzero-mean variants of the model are tested. For the lognormal and quarter-power normal models, the equal variance assumption is not treated here. The validity of equal variance assumptions are explicitly considered in Section VII.

Fig. 2 shows the cumulative distribution of P-values from pixels in the vehicle region of the images for each of the models when assessed by each of the preceding tests. Models that describe the data highly accurately should yield cumulative distributions that approximate a diagonal line through the plot center for each of the three tests. For any P-value,  $\alpha$ , the corresponding cumulative probability, indicates the fraction of samples that would fail a test with significance  $\alpha$ .

The test of D'Agostino and Pearson indicates that, for most combinations of pixel locations, vehicle, and pose, sample skewness and kurtosis values are within the range reasonably expected under each of the assumed models. There is a slight skewness toward low probability combinations suggesting that the models are not an exact fit for the data. The Kolmogorov-Smirnov test and Pearson's  $\chi^2$  test of the T-transformed sample data similarly show that many combinations of vehicle and pose yield samples that are within the range reasonably expected. The distribution of P-values for these tests is much more skewed toward low probability events, however. This suggests that some model assumptions and/or approximations are violated more often than would be otherwise expected. The distributions of P-values for the zero-mean variant of the complex Gaussian model are more skewed toward low values than for the nonzero-mean case. Because tests enforcing the zero-mean constraint are more powerful than those without, this difference in distribution cannot, by itself, be taken as evidence that the zero-mean assumption is improper. The zero-mean assumption is considered separately in Section VI.

Fig. 3 shows the distribution of P-values for the clutter region in the upper left corner of the images. Both complex Gaussian models and the quarter-power normal model exhibit distributions that are less skewed than for the pixels in the vehicle region. Since the tests were performed at equal powers, this suggests that the models may be a better fit for such clutter regions than for data drawn from the man-made vehicles. The lognormal model yields P-values that are skewed much more heavily toward low probabilities for the clutter data than for the vehicle data. The deviation from the expected uniform distribution is especially severe in the Kolmogorov-Smirnov and  $\chi^2$  tests. This suggests that the conditionally lognormal model is not a good fit for the clutter data represented in the sample data. This result is in keeping with [9] where it was noted that the presence of clutter pixels significantly reduced the performance of recognition algorithms based on a lognormal model.

With the exception of the lognormal model applied to clutter data, the majority of samples would pass a 5% significance test for all models considered. For clutter samples, nearly 90% of the samples would pass tests for the complex Gaussian and quarter-power normal models at the 5% significance levels.

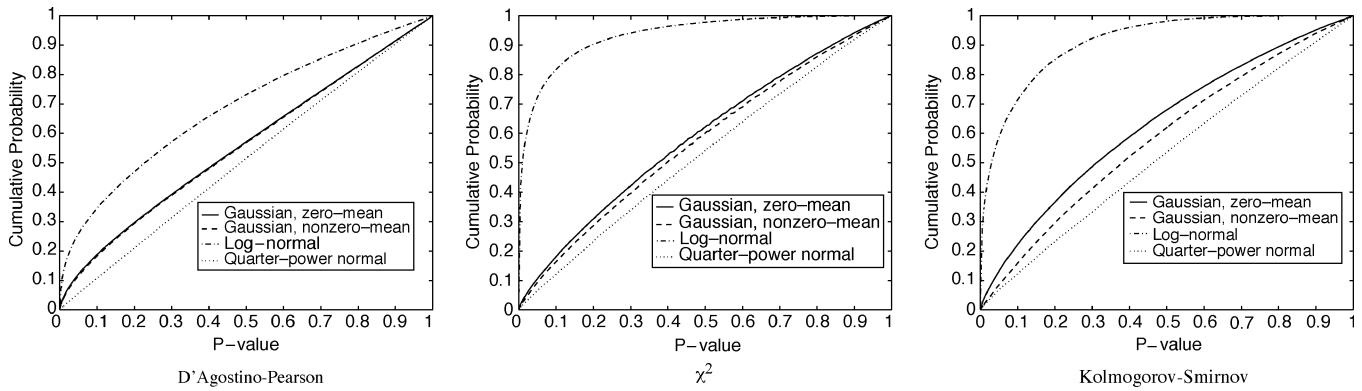


Fig. 3. Cumulative distribution of P-values for various goodness-of-fit tests applied to clutter regions in the scene.

The tests revealed that, while most samples are adequately characterized by the models, none of the models are exact matches to the available data. This is to be expected because of the approximations to mean and variance functions to accommodate continuous target pose described in Section III-A. In every test conducted the quarter-power normal model yielded a more even distribution of P-values, followed by the Gaussian and lognormal models, in turn. Since the tests were applied identically to all models, except for the more restrictive zero-mean Gaussian model, this would suggest that the quarter-power model may be a better fit to the SAR data considered. These results do not necessarily contradict the observation in [9] that recognition systems based on a complex Gaussian model consistently performed better than those based on lognormal or quarter-power models. In that paper, the assumption of homoscedasticity was imposed on the lognormal and quarter-power models. This assumption is called into question by the results of Section VII.

## V. PIXEL INDEPENDENCE

In this section, we consider the assumption of conditional independence of pixel values made in each of the three models.

### A. Test for Independence

Consider the joint density function for a pair of complex Gaussian random variables  $[X, Y]^T$ , which is

$$p(x, y) = \frac{1}{\pi^2 (1 - |\rho|^2) \sigma_x \sigma_y} \times \exp\left(-\frac{\sigma_y^2 |x|^2 + \sigma_x^2 |y|^2 - 2\sigma_x \sigma_y \operatorname{Re}\{\rho x^* y\}}{(1 - |\rho|^2) \sigma_x \sigma_y}\right). \quad (17)$$

Note that if  $Y$  and  $X$  are correlated,  $\rho \neq 0$ , then  $Y$  follows a linear regression on  $X$ . That is, conditioned on  $X$ ,  $Y$  is a complex Gaussian random variable with conditional mean given by  $\eta_2 x + \eta_1$  where  $\eta_1$  and  $\eta_2$  are linear regression coefficients. The random variables  $Y$  and  $X$  are independent, and hence uncorrelated if and only if  $\eta_2 = 0$ . We can thus formulate a test for the independence of two Gaussian random variables by testing the hypothesis  $\mathcal{H}_0 : \eta_2 = 0$  against the alternative  $\mathcal{H}_A : \eta_2 \neq 0$ .

Miller [20] addresses the general problem of least-squares linear regression with complex-valued variables and the specific problem of estimating the regression parameters of complex

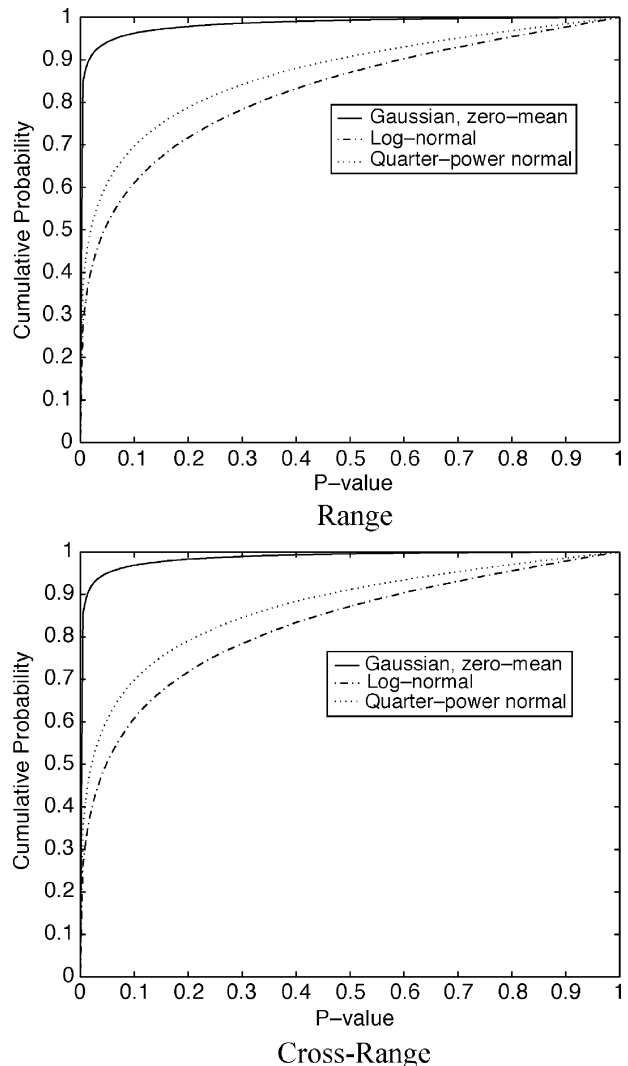


Fig. 4. Cumulative distribution of P-values for tests of correlation in the range and cross-range directions for pixels corresponding to target regions in the scene.

Gaussian random variables. Given pairs of observations  $\mathbf{y} = [y_1, y_2, \dots, y_n]^T$  and  $\mathbf{x} = [x_1, x_2, \dots, x_n]^T$ , the least-squared error estimate of  $\boldsymbol{\eta} = [\eta_1, \eta_2]^T$  satisfies

$$\hat{\boldsymbol{\eta}} = \arg \min_{\boldsymbol{\eta}} \|\mathbf{y} - H\boldsymbol{\eta}\|^2 \quad (18)$$

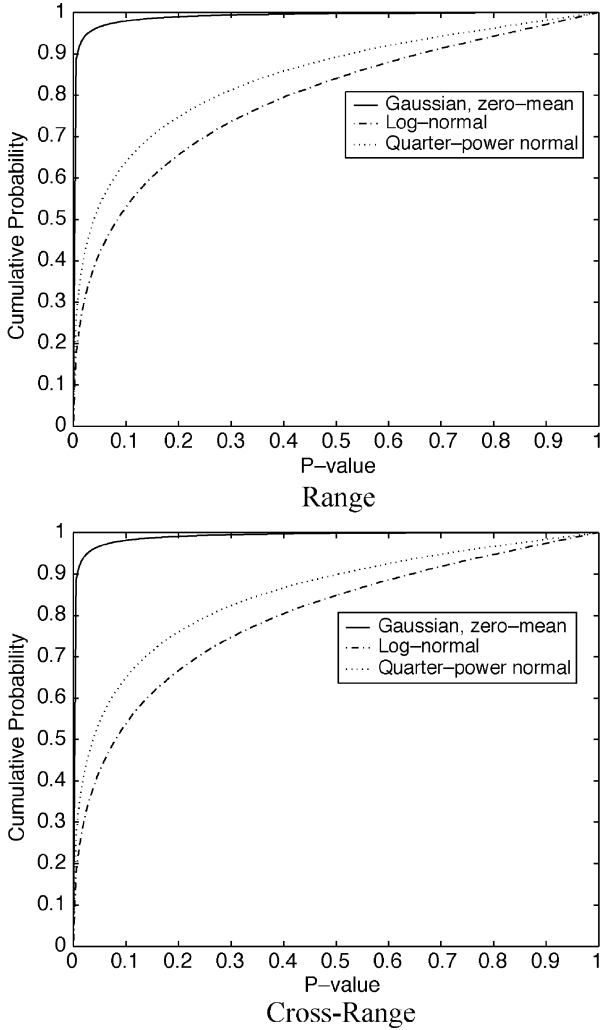


Fig. 5. Cumulative distribution of P-values for tests of correlation in the range and cross-range directions for pixels corresponding to clutter regions in the scene.

where

$$H = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}. \quad (19)$$

If not all the  $x_i$  are equal, this minimization has the solution

$$\hat{\boldsymbol{\eta}} = (H^\dagger H)^{-1} H^\dagger \mathbf{y} \quad (20)$$

where  $\dagger$  denotes the complex-conjugate transpose.

For the tests of independence, we need the distribution of the estimate  $\hat{\boldsymbol{\eta}}$  when  $\mathbf{y}$  is a complex Gaussian random vector with mean  $H\boldsymbol{\eta}$  and covariance matrix  $\sigma^2 I$ . This distribution is given by Miller [20]. For  $\mathbf{Y}$  complex Gaussian with mean  $H\boldsymbol{\eta}$  and covariance matrix  $\sigma^2 I$ ,  $\hat{\boldsymbol{\eta}}$  is complex Gaussian distributed with mean  $\boldsymbol{\eta}$  and covariance matrix  $(H^\dagger H)^{-1} \sigma^2$ . The maximum likelihood estimate for  $\sigma^2$  is given by

$$\hat{\sigma}^2 = \frac{1}{n} \|\mathbf{y} - H\hat{\boldsymbol{\eta}}\|^2. \quad (21)$$

Given samples  $x_1, x_2, \dots, x_n$  and  $y_1, y_2, \dots, y_n$  from neighboring pixel locations, we formulate the test based on the square magnitude  $|\hat{\eta}_2|^2$ . From the above,  $(\hat{\eta}_2 - \eta_2)/\xi$ , where  $\xi^2 = (H^\dagger H)_{2,2}^{-1} \sigma^2$ , is complex Gaussian with mean zero and unit variance so that the ratio  $2|\hat{\eta}_2 - \eta_2|^2/\xi^2$  follows a  $\chi^2$  distribution with 2 degrees of freedom. Additionally, Miller [20] shows that the distribution of  $\hat{\sigma}^2$  in (21) is such that the ratio  $2n\hat{\sigma}^2/\sigma^2$  is  $\chi^2$  distributed with  $2(n-2)$  degrees of freedom.

The ratio of these two  $\chi^2$  random variables, divided by their respective degrees of freedom, is independent of the variance  $\sigma^2$  and yields an F distributed random variable. That is, under the assumption of independence,  $\eta_2 = 0$  and

$$F = \frac{|\hat{\eta}_2|^2 (n-2)}{(H^\dagger H)_{2,2}^{-1} n \hat{\sigma}^2} \quad (22)$$

where the subscript 2,2 implies the lower right element of the matrix is F distributed with 2 and  $2(n-2)$  degrees of freedom. Large values of  $F$  indicate that  $\hat{\eta}_2$  is not close to zero and  $1 - F_{2,2(n-2)}(F)$ , where  $F_{2,2(n-2)}$  is the F cumulative distribution function with 2 and  $2(n-2)$  degrees of freedom, is reported as the P-value of the test.

A similar test for the independence of real-valued Gaussian random variables is straightforward and is covered in many texts on the subject (cf. [18, pp. 246–251] or [21, Chapter 11]).

### B. Independence Results

Fig. 4 shows the cumulative distribution of P-values from the independence test for adjacent pairs of pixels in the target region of the SAR imagery. The top panel shows P-values for neighboring pixels in the range direction (successive rows in the same column) and the bottom panel shows P-values for pixels in the cross-range direction (successive columns in the same row). The graphs suggest strong evidence that the independence assumption is violated by all three models, with more than half the samples failing tests with an  $\alpha = 5\%$  significance level. The distributions are similar for neighboring pixels in both directions for each model. Fig. 5 shows similar graphs produced from pixels in the clutter region of the SAR imagery. Evidence against the independence assumption is strong for the clutter region as well. That the pixel values are not independent may have been expected for the MSTAR imagery given that the resolution is larger than the pixel spacing. The severity of that assumption is made clear in these figures, particularly for the complex Gaussian model family.

Fig. 6 shows the average correlation coefficient for pairs of pixel values as a function of the distance between the pixel locations for target regions. The top panel shows average correlation in the range direction and the bottom panel shows correlation in the cross-range direction. The complex Gaussian model yields complex-valued correlation coefficients and the graphs show both real and imaginary components. The real component is a measure of correlation between the real value pairs and of the imaginary value pairs. The imaginary component of correlation is a measure of correlation between the cross components, real and imaginary, of the pixel pairs. The average correlation is seen to drop rapidly over a distance of around four

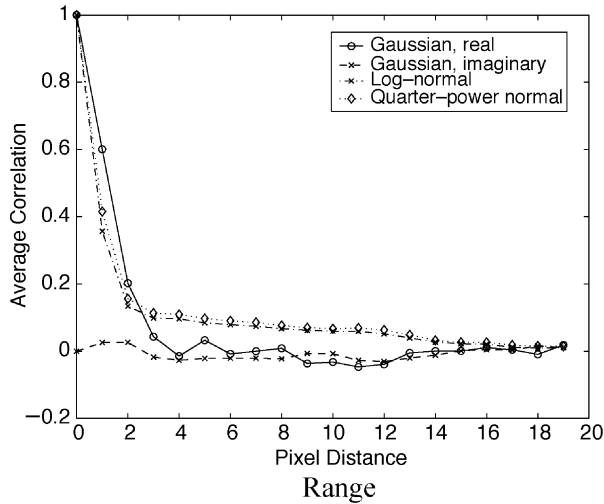


Fig. 6. Average correlation as a function of pixel distance in the range and cross-range directions for pixels corresponding to target regions in the scene.

pixels before gradually declining toward zero. The imaginary component of the average correlation in the complex Gaussian model is close to zero for all distances. Fig. 7 shows the dependence of average correlation on distance for pixels in the clutter regions. The relationship is similar to that for target regions over 1 and 2 pixel widths, but for three or more pixel widths the correlation is much closer to zero.

## VI. COMPLEX ZERO MEAN

In this section, we evaluate the zero-mean assumption of the complex Gaussian model. The test is constructed as a decision between  $\mathcal{H}_0$ , an assertion that the data are complex Gaussian with zero mean, versus  $\mathcal{H}_A$ , an assertion that they are not. Giri [12] addresses the problem of inference on the mean of complex-valued Gaussian random vectors with unknown covariance matrix. He demonstrates that the statistic

$$F = \frac{n|\bar{X} - \mu|^2}{\hat{\sigma}^2} \quad (23)$$

is an F-distributed random variable with 2 and  $2(n-1)$  degrees of freedom, where  $\bar{X}$  and  $\hat{\sigma}^2$  represent the sample mean and variance from  $n$  complex Gaussian random variables with mean

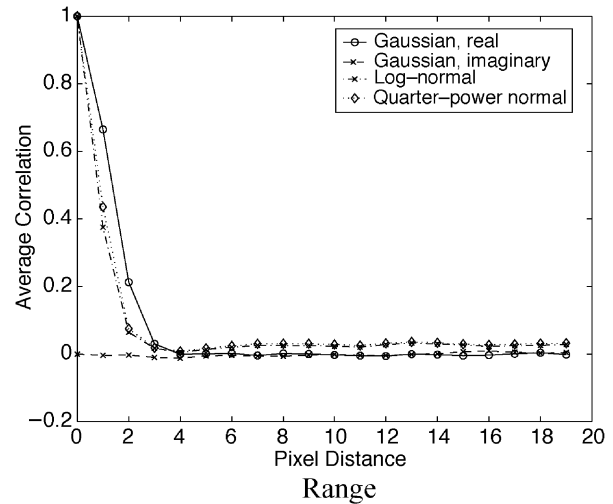


Fig. 7. Average correlation as a function of pixel distance in the range and cross-range directions for pixel corresponding to clutter regions in the scene.

$\mu$ . Note that if we let  $\mu = 0$  in the above expression, large values of  $F$  imply that the sample mean is not close to zero. We report  $1 - F_{2,2(n-1)}(F)$  as the P-value of the test, where  $F_{2,2(n-1)}$  is the cumulative distribution function for an F distributed random variable with 2 and  $2(n-1)$  degrees of freedom.

Fig. 8 shows the empirical cumulative distribution of P-values for the test of zero mean across all pixel locations for the Gaussian model for both target and clutter regions of the images. The two distributions are essentially overlapping over the entire range of P-values and demonstrate the uniform distribution we would expect if the zero-mean hypothesis were true. Thus, there is no evidence on the basis of this test that the zero-mean hypothesis is violated by the data.

## VII. HOMOSCEDASTICITY

In this section, we evaluate the assumption of homoscedasticity (equal variance) in the lognormal and quarter-power normal models. Let  $X_1, X_2, \dots, X_m$  and  $Y_1, Y_2, \dots, Y_n$  be independent samples from real-valued Gaussian random variables with unknown means  $\mu_x$  and  $\mu_y$ , respectively, and with equal variance  $\sigma_x^2 = \sigma_y^2 = \sigma^2$ . Then  $(m-1)\hat{\sigma}_x^2/\sigma^2$  is  $\chi^2$  distributed with  $m-1$  degrees of freedom, and  $(n-1)\hat{\sigma}_y^2/\sigma^2$

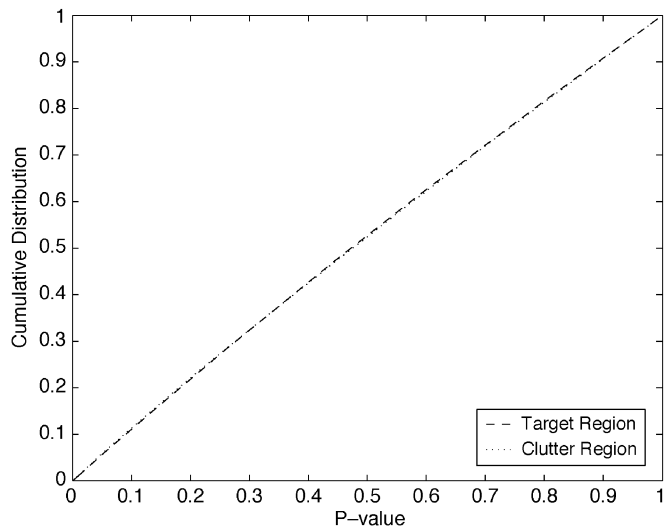


Fig. 8. Cumulative distribution of P-values for tests of zero mean.

is  $\chi^2$  distributed with  $n - 1$  degrees of freedom, where  $\hat{\sigma}_x^2$  and  $\hat{\sigma}_y^2$  are unbiased estimates of  $\sigma^2$  determined from the  $X_i$  and  $Y_i$ , respectively. The test statistic

$$F = \frac{\hat{\sigma}_x^2}{\hat{\sigma}_y^2} \quad (24)$$

is an F-distributed random variable with  $m - 1$  and  $n - 1$  degrees of freedom. Very large or small values of  $F$  indicate that the estimates for  $\sigma^2$  are dissimilar. Only one of  $F_{(m-1),(n-1)}(F)$  or  $1 - F_{(m-1),(n-1)}(F)$  is less than one-half, and it represents one-half the probability of an observation at least as extreme (the variance estimates could have reversed roles and the result would be as extreme). The P-value reported for the test is  $2 \min\{F_{(m-1),(n-1)}(F), 1 - F_{(m-1),(n-1)}(F)\}$ .

Fig. 9 shows the empirical cumulative distribution of P-values from the above test applied to randomly selected pairs of pixel locations for both models. The top panel shows the results for the central target region of the SAR images and the bottom panel shows results for the background clutter. Both sets of curves differ visibly from the expected uniform distribution and so evidence suggests that the equal variance assumption may be violated in both models. The deviation is especially severe in the quarter-power normal model for target pixels in which nearly 60% of the sample pairs would fail a test with an  $\alpha = 5\%$  significance level. The distribution of P-values for the lognormal model is very similar for both target and clutter regions.

## VIII. CONCLUSIONS

We provide a framework for quantitatively assessing conditional models of imagery across large numbers of small sample sizes through application of the probability integral transformation. The resulting methodology can be applied to a wide range of traditional statistical assessment procedures including goodness-of-fit tests, tests for independence, and inferences on sample statistics. The background and techniques for three goodness-of-fit tests are presented and transformations

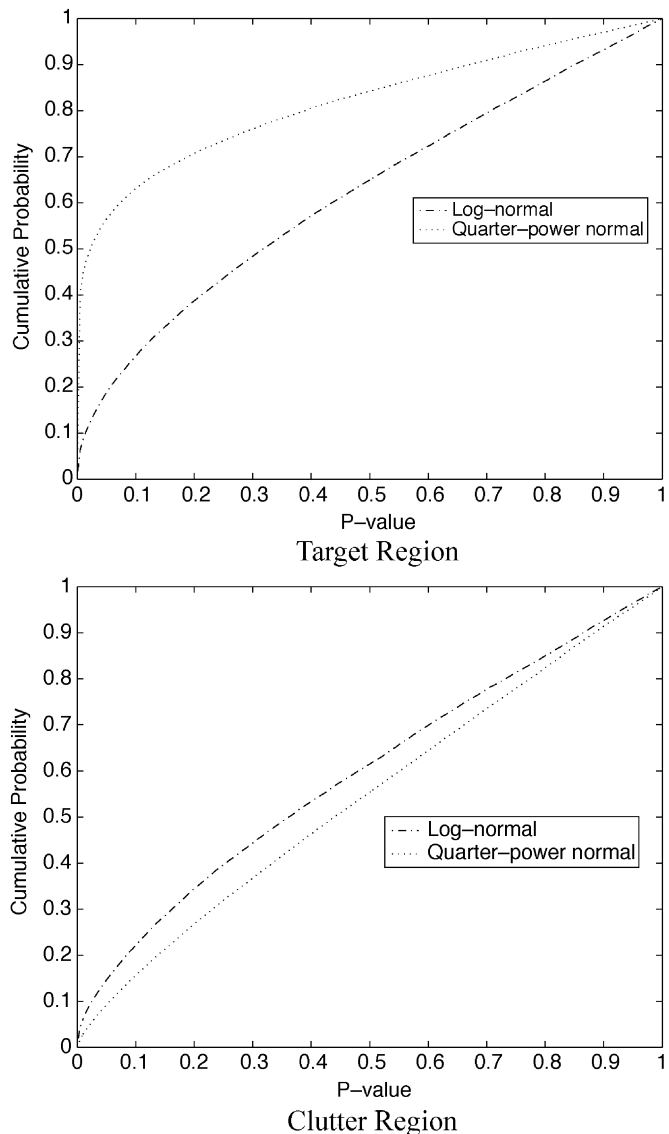


Fig. 9. Cumulative distribution of P-values for tests of homoscedasticity for both target and clutter regions.

producing samples of known distribution are derived which allow application of these tests to small samples with unknown distribution statistics. These tests are applied to three models of SAR imagery with actual data.

The three goodness-of-fit tests are sensitive to different characteristic departures from normality. The method of D'Agostino and Pearson is sensitive to the joint distribution of skewness and kurtosis estimates, and it showed both the least separation between models and the least evidence to reject the models. However, any goodness-of-fit test implicitly tests all model assumptions, in this case including pixel independence and statistics that are piecewise constant over azimuth. The comparatively large percentage of small P-values with Pearson's  $\chi^2$  and Kolmogorov-Smirnov methods, based respectively on data histograms and empirical CDFs, may be due to their increased reliance on independence in neighboring pixels or the fact that they test for the Gaussian distribution indirectly through T transformed samples. These results do not necessarily indicate

that the D'Agostino–Pearson test is less powerful in general. What is significant is that the three tests were qualitatively consistent for both object and clutter data and for all models.

All three goodness-of-fit tests suggest that none of the models is completely accurate, in the sense that more samples than would otherwise be expected result in extreme values of the respective test statistic. However, most of the samples collected from the man-made objects would pass a test having  $\alpha = 5\%$  significance level. This suggests that although none of the models is exact, they may be serviceable for particular applications. All of the tests suggest that the data are best described by the quarter-power normal model, followed by the complex Gaussian and lognormal models. Given that the quarter-power normal model is based on an approximation of the gamma distribution, one may speculate that a gamma model for SAR pixel magnitude is reasonably accurate. While the quarter-power normal and complex Gaussian models seem to describe the background clutter better than man-made objects, the ability of the lognormal model to characterize background clutter is much worse than for the man-made objects.

The same framework is also applied to assessment of pixel independence, and the results suggest strong evidence against assumptions of independence in either the range or cross-range directions. Inference tests on sample statistics suggest that the SAR data are not inconsistent with assumptions of zero mean in the Gaussian model and are inconsistent with assumptions of homoscedasticity in lognormal and quarter-power models.

## REFERENCES

- [1] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Automat. Contr.*, vol. AC-19, pp. 716–723, Dec. 1974.
- [2] F. J. Anscombe and W. J. Glynn, "Distribution of the kurtosis statistic  $b_2$  for normal samples," *Biometrika*, vol. 70, no. 1, pp. 227–234, Apr. 1983.
- [3] J. B. Billingsley, A. Farina, F. Gini, M. V. Greco, and L. Verrazzani, "Statistical analyses of measured radar ground clutter data," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 35, pp. 579–593, Apr. 1999.
- [4] H. D. Brunk, *An Introduction to Mathematical Statistics*: Ginn and Company, 1960.
- [5] R. B. D'Agostino, "Transformation to normality of the null distribution of  $g_1$ ," *Biometrika*, vol. 57, no. 3, pp. 679–681, Dec. 1971.
- [6] R. B. D'Agostino, A. Belanger, and R. B. D'Agostino Jr., "A suggestion for using powerful and informative tests of normality," *The American Statistician*, vol. 44, no. 4, pp. 316–321, Nov. 1990.
- [7] R. B. D'Agostino and E. S. Pearson, "Tests for departure from normality. Empirical results for the distributions of  $b_2$  and  $\sqrt{b_1}$ ," *Biometrika*, vol. 60, no. 3, pp. 613–622, Dec. 1973.
- [8] M. D. DeVore, "Recognition performance from synthetic aperture radar imagery subject to system resource constraints," Ph.D. dissertation, Washington Univ., St. Louis, MO, May 2001.
- [9] M. D. DeVore and J. A. O'Sullivan, "A performance-complexity study of several approaches to automatic target recognition from synthetic aperture radar images," *IEEE Trans. Aerosp. Electron. Syst.*, to be published.
- [10] T. S. Edrington, "The amplitude statistics of aircraft radar echoes," *IEEE Trans. Military Electron.*, vol. MIL-9, no. 1, pp. 10–16, Jan. 1965.
- [11] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, 2nd ed. New York: Academic, 1990, pp. 76–77.
- [12] N. Giri, "On the complex analogues of  $t_2$ - and  $r^2$ - tests," *Ann. Math. Statistics*, vol. 36, no. 2, pp. 664–670, Apr. 1965.
- [13] J. W. Goodman, "Some fundamental properties of speckle," *J. Opt. Soc. Amer.*, vol. 66, no. 11, pp. 1145–1150, Nov. 1976.
- [14] C. Holt, J. Attili, and S. Schmidt, "Validation of a  $\chi^2$  model of HRR target RCS variability and verification of the resulting ATR performance model," in *Proc. SPIE, Automatic Target Recognition XI*, vol. 4379, F. A. Sadjadi, Ed., 2001.
- [15] L. M. Kaplan, "Analysis of multiplicative speckle models for template-based SAR ATR," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 37, pp. 1424–1432, Oct. 2001.
- [16] M. Kendall and A. Stuart, *The Advanced Theory of Statistics*, 4th ed. New York: Macmillan, 1977, vol. 1.
- [17] E. E. Kuruoglu and J. Zerubia, "Modeling SAR images with a generalization of the Rayleigh distribution," in *Proc. 34th Annu. Asilomar Conf. Signals, Systems, and Computers*, vol. 1, M. B. Matthews, Ed., Oct. 2000, pp. 224–228.
- [18] A. Leon-Garcia, *Probability and Random Processes for Electrical Engineering*, 2nd ed. Reading, MA: Addison-Wesley, 1994.
- [19] E. B. Manoukian, *Mathematical Nonparametric Statistics*. New York: Gordon and Breach, 1986.
- [20] K. S. Miller, "Complex linear least squares," *SIAM Rev.*, vol. 15, no. 4, pp. 706–726, Oct. 1973.
- [21] J. S. Milton and J. C. Arnold, *Probability and Statistics in the Engineering and Computing Sciences*. New York: McGraw-Hill, 1986.
- [22] F. E. Nathanson, J. P. Reilly, and M. N. Cohen, *Radar Design Principles*, 2nd ed. New York: McGraw-Hill, 1991.
- [23] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed. Cambridge, MA: Cambridge Univ. Press, 1992.
- [24] H. R. Raemer, *Radar System Principles*. Boca Raton, FL: CRC Press, 1997.
- [25] J. P. Reilly, "On the statistical representation of targets for detection studies," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-5, pp. 560–561, May 1969.
- [26] J. Rissanen, "Stochastic complexity and modeling," *Ann. Statistics*, vol. 14, no. 3, pp. 1080–1100, 1986.
- [27] D. J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*, 2nd ed. London, U.K.: Chapman & Hall/CRC, 2000.
- [28] D. A. Shnidman, "Generalized radar clutter model," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 35, pp. 857–865, July 1999.
- [29] M. A. Stephens, "Kolmogorov–Smirnov type tests of fit," in *Encyclopedia of Statistical Sciences*, S. Kotz, N. Johnson, and C. B. Read, Eds. New York: Wiley, 1982, vol. 4, pp. 398–402.
- [30] H. L. Van Trees, *Detection, Estimation, and Modulation Theory. Radar-Sonar Signal Processing and Gaussian Signals in Noise*. New York: Wiley, 1971, vol. III.
- [31] S. W. Worrell, S. Parker, and M. L. Bryant, "Class separability assessments and MSE algorithm robustness," in *Proc. of SPIE, Algorithms for Synthetic Aperture Radar Imagery IV*, vol. 3070, E. G. Zelnio, Ed., 1997, pp. 294–304.
- [32] J. H. Zar, *Biostatistical Analysis*, 4th ed. Englewood Cliffs, NJ: Prentice-Hall, 1999.



**Michael D. DeVore** (M'01) received the B.S. degrees in electrical engineering, computer engineering, and mathematics in 1991 and the M.S. degree in electrical engineering in 1993, all from the University of Missouri at Columbia. He received the Ph.D. degree from Washington University, St. Louis, MO, in 2001.

From 1993 to 1998, he was with Amdocs, Inc., a supplier of business software to the communications industry. In 2001, he was a Visiting Assistant Professor in Electrical Engineering at Washington University, and in 2002 joined the faculty of the University of Virginia, Charlottesville, as an Assistant Professor in the Systems and Information Engineering Department. His research interests include automatic target recognition, computer vision information systems, and representations for random object shape.

Dr. DeVore is a member of SPIE and the Optical Society of America.



**Joseph A. O'Sullivan** (M'86–SM'92–F'03) was born in St. Louis, MO, in 1960. He received the B.S., M.S., and Ph.D. degrees, all in electrical engineering, from the University of Notre Dame, Notre Dame, IN, in 1982, 1984, and 1986, respectively.

In 1986, he joined the faculty of the Department of Electrical Engineering, Washington University, St. Louis, where he is currently a Professor. He has joint appointments in the Departments of Radiology and of Biomedical Engineering. He is Director of the Electronic Systems and Signals Research Laboratory and

Associate Director of the Center for Security Technologies at Washington University. He is also Chair of the Faculty Senate, Chair of the Faculty Senate Council, and Faculty Representative to the Board of Trustees at Washington University. He was Secretary of the Faculty Senate and of the Senate Council from 1995 to 1998. His research interests include information theory, information-theoretic imaging, automatic target recognition, CT imaging in the presence of known high density attenuators, information hiding, and hyperspectral imaging.

Dr. O'Sullivan was the Publications Editor for the IEEE TRANSACTIONS ON INFORMATION THEORY from 1992 to 1995, Associate Editor for Detection and Estimation, and was a Guest Associate Editor for the 2000 Special Issue on Information Theoretic Imaging. He was Chair of the St. Louis Section of the IEEE in 1994, Co-Chair of the 1999 Information Theory Workshop on Detection, Estimation, Classification, and Imaging, and Local Arrangements Chair for the 2003 IEEE Statistical Signal Processing Workshop. He is a member of Eta Kappa Nu, SPIE, SIAM, and AAAS. He was awarded an IEEE Third Millennium Medal.