# Analytical performance evaluation of SAR ATR with inaccurate or estimated models

Michael D. DeVore
Department of Systems and Information Engineering
University of Virginia, P.O. Box 400747, Charlottesville VA, USA

## ABSTRACT

Hypothesis testing algorithms for automatic target recognition (ATR) are often formulated in terms of some assumed distribution family. The parameter values corresponding to a particular target class together with the distribution family constitute a model for the target's signature. In practice such models exhibit inaccuracy because of incorrect assumptions about the distribution family and/or because of errors in the assumed parameter values, which are often determined experimentally. Model inaccuracy can have a significant impact on performance predictions for target recognition systems. Such inaccuracy often causes model-based predictions that ignore the difference between assumed and actual distributions to be overly optimistic. This paper reports on research to quantify the effect of inaccurate models on performance prediction and to estimate the effect using only trained parameters. We demonstrate that for large observation vectors the class-conditional probabilities of error can be expressed as a simple function of the difference between two relative entropies. These relative entropies quantify the discrepancies between the actual and assumed distributions and can be used to express the difference between actual and predicted error rates. Focusing on the problem of ATR from synthetic aperture radar (SAR) imagery, we present estimators of the probabilities of error in both ideal and plug-in tests expressed in terms of the trained model parameters. These estimators are defined in terms of unbiased estimates for the first two moments of the sample statistic. We present an analytical treatment of these results and include demonstrations from simulated radar data.

**Keywords:** performance prediction, ATR theory, model-based ATR, model inaccuracy, synthetic aperture radar

## 1. INTRODUCTION

In statistical approaches to automatic target recognition (ATR) problems, classifier designs generally involve probability distributions over the set of observations that are inaccurately (or incompletely) known. This includes incorrect assumptions regarding the distribution family for observations and insufficient sample data to accurately estimate the distribution parameters. Such systems are often implementations of likelihood-based tests that *would* be optimal if the assumed distributions were correct, but suffer performance degradation because the design does not match the actual observations. For example, the "plug-in" rule is an approximation of an optimal classifier, in which sample data is used to estimate the unknown parameters in class-conditional probability distribution families. The parameter estimates are then substituted into a likelihood-ratio test in place of the true parameters.

A key question regarding any given classifier design is how well it will perform, both in an absolute sense and relative to how well an ideal classifier could perform. In the notation of Fukunaga,[1] we let $\mathcal{P} = \{p_1, p_2, \ldots, p_M\}$ be the set of true probability distributions for an observation $\mathbf{X}$ under each of $M$ hypotheses, which we denote $\mathcal{H}_m$, $m = 1, 2, \ldots, M$. In the absence of exact knowledge of $\mathcal{P}$, an assumed set of distributions $\tilde{\mathcal{P}} = \{\tilde{p}_1, \tilde{p}_2, \ldots, \tilde{p}_M\}$ is employed. The primary goal of performance analysis is to estimate the probability of error for a system designed for $\tilde{\mathcal{P}}$ but applied to $\mathcal{P}$. We denote this probability of error as $\varepsilon(\mathcal{P}, \tilde{\mathcal{P}})$ and refer to it as the *actual* probability of error. A secondary goal of performance analysis is to estimate the *ideal* probability of error $\varepsilon(\mathcal{P}, \mathcal{P})$, a measure of how well an ATR system could perform if enough resources were expended to more accurately determine $\mathcal{P}$.

Several sample-based methods exist for estimating the probability of error $\varepsilon(\mathcal{P}, \tilde{\mathcal{P}})$, including hold-out and cross-validation (cf. Fukunaga[1] Section 5.3). These methods divide available sample data into training and

E-mail: mdevore@virginia.edu

Algorithms for Synthetic Aperture Radar Imagery XI, edited by Edmund G. Zelnio,
Frederick D. Garber, Proceedings of SPIE Vol. 5427 (SPIE, Bellingham, WA, 2004)
0277-786X/04/$15 · doi: 10.1117/12.548352

testing sets. The training set is first used to determine $\tilde{\mathcal{P}}$, then the testing set is used to estimate $\varepsilon(\mathcal{P}, \tilde{\mathcal{P}})$ as the fraction of test samples that are incorrectly classified. These methods, particularly cross-validation, can provide very good estimates when a relatively rich set of sample data is available. If the set of sample data is small, however, the number of possible values the error estimate can assume is small. For example, in a two-class problem with 3 samples in the testing set from each class, the smallest nonzero probability of error that can be reported is $1/6 \approx 16.7\%$, which occurs when a single test sample is misclassified. This estimate is too coarse to be of use in many cases. Small sample sets are particularly likely to occur when the hypotheses include target pose. That is, it is difficult to collect a large number of training samples of a target in every possible orientation relative to the sensing device. It is more reasonable to collect sample observations from a finite set of orientations, so the number of samples in any small range is small.

In addition to the problems with small sample sets, sample-based methods do not provide a direct way to estimate the ideal error $\varepsilon(\mathcal{P}, \mathcal{P})$. One possibility is to use Monte Carlo simulation, simulating a large number of observations from $\tilde{\mathcal{P}}$ and reporting the fraction which are classified incorrectly. This is similar in spirit to the sample-based resubstitution method, in which the training and testing sets are identical. Both of these approaches produce estimates of $\varepsilon(\tilde{\mathcal{P}}, \tilde{\mathcal{P}})$, which we will refer to as the *plug-in* error. Glick[2] demonstrates that an estimate of this plug-in error will underestimate both the actual and ideal error rates on average.

The approach adopted in this paper is to find an approximate closed-form expression for probability of error that explicitly accounts for inaccuracy in $\tilde{\mathcal{P}}$. Attention is focused on the conditional probability of error between pairs of hypotheses. Approximate upper and lower bounds on the unconditional probability of error can then be obtained through appropriate combinations of these. Consider the likelihood ratio test for a vector observation $\mathbf{X}$ with independent components. If the two hypotheses are incorrectly formed as $\tilde{p}_1$ and $\tilde{p}_2$, the resulting test can be written as

$$\tilde{L} = \ln\left(\frac{\tilde{p}_1(\mathbf{X})}{\tilde{p}_2(\mathbf{X})}\right) \underset{\mathcal{H}_2}{\overset{\mathcal{H}_1}{\gtrless}} \gamma. \tag{1}$$

The expected value of $\tilde{L}$ under $\mathcal{H}_1$ is

$$\begin{aligned} \mathrm{E}[\tilde{L}|\mathcal{H}_1] &= \mathrm{E}\left[\ln\left(\frac{\tilde{p}_1(\mathbf{X})}{\tilde{p}_2(\mathbf{X})}\right)\middle|\mathcal{H}_1\right] \\ &= \mathrm{E}\left[\ln\frac{p_1(\mathbf{X})}{\tilde{p}_2(\mathbf{X})} - \ln\frac{p_1(\mathbf{X})}{\tilde{p}_1(\mathbf{X})}\middle|\mathcal{H}_1\right] \\ &= D(p_1||\tilde{p}_2) - D(p_1||\tilde{p}_1), \end{aligned} \tag{2}$$

where $D(q||r)$ is the relative entropy between the distributions $q$ and $r$. Since an error is made whenever $\tilde{L} < \gamma$, the error rate will be smaller if the true distribution of $\mathbf{X}$ under $\mathcal{H}_1$ is closer to the assumed distribution under $\mathcal{H}_1$ than the assumed distribution under $\mathcal{H}_2$, all else being equal. This expected value is not the only factor governing conditional error rates, however, because the probability that $\tilde{L} < \gamma$ also depends on higher order moments of $\tilde{L}$.

The Central Limit Theorem suggests that a sum of large numbers of independent random variables will follow an approximately Gaussian distribution. While it is often applied to sums of identically distributed terms, variants of the theorem exist which do not require identical distributions. For example, the Lindeberg-Feller theorem[3] requires the total variance of the sum to grow toward infinity while the ratio of each term's variance to the total goes toward zero. Other variants include those of Liapounov[4] and Kendall and Stuart.[5] If we assume that $\tilde{L}$ is approximately Gaussian, then to determine the conditional probability of error it is sufficient to know its conditional mean and variance.

The conditional mean and variance of $\tilde{L}$ depend not only on $\tilde{\mathcal{P}}$ but also on $\mathcal{P}$ as can be seen in (2). For these quantities to be usefully applied we must know the true conditional distributions, which we have already assumed are unknown. In this work, we focus on unbiased estimators for $\mathrm{E}[\tilde{L}|\mathcal{H}_1]$ and $\mathrm{var}(\tilde{L}|\mathcal{H}_1)$, which we use to estimate the actual conditional error rate, $\Pr[\tilde{L} < \gamma|\mathcal{H}_1]$. Further, we consider unbiased estimators for $\mathrm{E}[L|\mathcal{H}_1]$ and $\mathrm{var}(L|\mathcal{H}_1)$, which we use to estimate the probability of error in the ideal test.

The assumption throughout this paper is that the observations $\mathbf{X}$ are drawn from a zero-mean complex Gaussian distribution, the variance of which depends on the true hypothesis. A classifier is constructed by estimating these variance values from as few as 3 training samples. This model has been applied to radar returns from spread targets by many authors including Van Trees.[6] More recently, the model has been applied to high resolution radar[7] and synthetic aperture radar(SAR).[8–10] Quantitative statistical evaluation of this model for SAR imagery has been reported by DeVore and O'Sullivan.[11] Related work involving probability of error in complex Gaussian target recognition includes that of Schmid and O'Sullivan.[12] They derive an asymptotic integral expression for the conditional characteristic function of $\tilde{L}$ with relative entropy-based segmentation. This expression is exploited to establish Chernoff bounds on the conditional probabilities of error.

In Section 2 we find expressions for the conditional mean and variance of the test statistic for both the ideal (true parameters) and actual (trained parameters) tests. We present unbiased estimators for these quantities in terms of the trained parameters in Section 3 and present corresponding probability of error estimators. Behavior of these estimators is discussed in the examples of Section 4. Concluding remarks follow in Section 5.

## 2. DISTRIBUTION OF THE TEST STATISTIC

### 2.1. Known Parameters

We first consider the ideal case in which all parameters are known. The goal is to choose between one of two complex Gaussian hypotheses

$$\begin{aligned} \mathcal{H}_1 &: \mathbf{X} \sim \mathbb{C}N(\mathbf{0}, \mathrm{diag}(\sigma_1^2, \ldots, \sigma_K^2)) \\ \mathcal{H}_2 &: \mathbf{X} \sim \mathbb{C}N(\mathbf{0}, \mathrm{diag}(\tau_1^2, \ldots, \tau_K^2)). \end{aligned} \tag{3}$$

The likelihood ratio test reduces to

$$\begin{aligned} L(\mathbf{X}) &= \sum_{k=1}^{K} \left[ \left( \frac{1}{\tau_k^2} - \frac{1}{\sigma_k^2} \right) |X_k|^2 - \ln \frac{\sigma_k^2}{\tau_k^2} \right] \underset{\mathcal{H}_2}{\overset{\mathcal{H}_1}{\gtrless}} \gamma \\ &= \sum_{k=1}^{K} \left[ \frac{1}{2} \left( \frac{\sigma_k^2}{\tau_k^2} - 1 \right) \frac{2|X_k|^2}{\sigma_k^2} - \ln \frac{\sigma_k^2}{\tau_k^2} \right] \underset{\mathcal{H}_2}{\overset{\mathcal{H}_1}{\gtrless}} \gamma. \end{aligned} \tag{4}$$

The quantity $U = 2|X_k|^2/\sigma_k^2$ follows a $\chi_\nu^2$ distribution with $\nu = 2$ degrees of freedom when $\mathcal{H}_1$ is true. The probability density function for a general $\chi_\nu^2$ random variable is

$$f_{\chi_\nu^2}(u) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} e^{u/2} u^{(\nu/2)-1}. \tag{5}$$

The mean of a $\chi_\nu^2$ distributed random variable is $\nu$ and the variance is $2\nu$, so $\mathrm{E}[U|\mathcal{H}_1] = 2$ and $\mathrm{var}(U|\mathcal{H}_1) = 4$. The conditional mean and variance of the likelihood ratio under $\mathcal{H}_1$ become

$$\mathrm{E}[L|\mathcal{H}_1] = \sum_{k=1}^{K} \left( \frac{\sigma_k^2}{\tau_k^2} - \ln \frac{\sigma_k^2}{\tau_k^2} - 1 \right) \tag{6}$$

and

$$\mathrm{var}(L|\mathcal{H}_1) = \sum_{k=1}^{K} \left( \frac{\sigma_k^2}{\tau_k^2} - 1 \right)^2. \tag{7}$$

When $\mathcal{H}_1$ is true, we make a classification error if $L$ is less than $\gamma$. We can approximate the conditional probability of error given $\mathcal{H}_1$ by noting that if $K$ is large, the log-likelihood ratio will tend to a Gaussian random variable under suitable conditions as discussed in Section 1. The conditional probability of error is

$$\Pr[\mathrm{error}|\mathcal{H}_1] = \Pr[L < \gamma|\mathcal{H}_1] \approx \Phi\left( \frac{\gamma - \mathrm{E}[L|\mathcal{H}_1]}{\sqrt{\mathrm{var}(L|\mathcal{H}_1)}} \right), \tag{8}$$

where $\Phi$ denotes the cumulative distribution function for a standard Gaussian random variable. Thus when $K$ is large, the quantities $\mathrm{E}[L|\mathcal{H}_1]$ and $\mathrm{var}(L|\mathcal{H}_1)$ are parameters that completely determine the probability of error.

## 2.2. Incorrectly Assumed Parameters

If the incorrect parameters $s_k^2$ and $t_k^2$ are substituted in the log-likelihood ratio in place of $\sigma_k^2$ and $\tau_k^2$, respectively, the (now suboptimal) test becomes

$$\tilde{L}(\mathbf{X}) = \sum_{k=1}^{K} \left[ \frac{1}{2} \left( \frac{\sigma_k^2}{t_k^2} - \frac{\sigma_k^2}{s_k^2} \right) \frac{2|X_k|^2}{\sigma_k^2} - \ln \frac{s_k^2}{t_k^2} \right] \underset{\mathcal{H}_2}{\overset{\mathcal{H}_1}{\gtrless}} \gamma. \tag{9}$$

As a result, the conditional mean and variance of $\tilde{L}$ are

$$\mathrm{E}[\tilde{L}|\mathcal{H}_1] = \sum_{k=1}^{K} \left( \frac{\sigma_k^2}{t_k^2} - \frac{\sigma_k^2}{s_k^2} - \ln \frac{s_k^2}{t_k^2} \right) \tag{10}$$

and

$$\mathrm{var}(\tilde{L}|\mathcal{H}_1) = \sum_{k=1}^{K} \left( \frac{\sigma_k^2}{t_k^2} - \frac{\sigma_k^2}{s_k^2} \right)^2. \tag{11}$$

In cases where the $\sigma_k^2$ and $\tau_k^2$ are not known, we may substitute estimates computed from training data. Let $\mathbf{Y}_1, \ldots, \mathbf{Y}_{Ns}$ and $\mathbf{Z}_1, \ldots, \mathbf{Z}_{Ns}$ be training vectors collected under $\mathcal{H}_1$ and $\mathcal{H}_2$, respectively, and let $Y_{n,k}$ and $Z_{n,k}$ be the $k$th components of the $n$th training vectors. The maximum-likelihood estimates for $\sigma_k^2$ and $\tau_k^2$ become

$$S_k^2 = \frac{1}{N_s} \sum_{n=1}^{Ns} |Y_{n,k}|^2 \qquad \text{and} \qquad T_k^2 = \frac{1}{N_t} \sum_{n=1}^{Nt} |Z_{n,k}|^2. \tag{12}$$

Note that $(2/\sigma_k^2) \sum_{n=1}^{Ns} |Y_{n,k}|^2 \sim \chi^2_{2N_s}$ and a similar expression holds for the $Z_{n,k}$. As a result, the probability density functions for $S_k^2$ and $T_k^2$ can be written as

$$f_{S_k^2}(s^2) = \frac{2N_s}{\sigma_k^2} f_{\chi^2_{2N_s}} \left( \frac{2N_s}{\sigma_k^2} s^2 \right) \qquad \text{and} \qquad f_{T_k^2}(t^2) = \frac{2N_t}{\tau_k^2} f_{\chi^2_{2N_t}} \left( \frac{2N_t}{\tau_k^2} t^2 \right). \tag{13}$$

The mean and variance of $\tilde{L}$ in (10) and (11) are determined with respect to the random observation $\mathbf{X}$. With the parameters $S_k^2$ and $T_k^2$ following the distributions in (13), we seek the mean and variance of the test statistic with respect to these random quantities as well. We denote by $\hat{L}$ the test statistic with these estimates substituted for the $\sigma_k^2$ and $\tau_k^2$, and let $\hat{\lambda}_k$ denote the $k$th summand in the expression for $\hat{L}$. That is, $\hat{L} = \sum_{k=1}^{K} \hat{\lambda}_k$ where

$$\hat{\lambda}_k = \frac{1}{2} \left( \frac{\sigma_k^2}{t_k^2} - \frac{\sigma_k^2}{s_k^2} \right) \frac{2|X_k|^2}{\sigma_k^2} - \ln \frac{s_k^2}{t_k^2}. \tag{14}$$

The conditional mean of $\hat{\lambda}_k$ given $\mathcal{H}_1$ can be determined in closed form as

$$\begin{aligned} \mathrm{E}[\hat{\lambda}_k|\mathcal{H}_1] &= \int_0^\infty \int_0^\infty \int_0^\infty \left[ \frac{1}{2} \left( \frac{\sigma_k^2}{t^2} - \frac{\sigma_k^2}{s^2} \right) u - \ln \frac{s^2}{t^2} \right] f_{\chi^2_2}(u) f_{S_k^2}(s^2) f_{T_k^2}(t^2) du ds^2 dt^2 \\ &= \frac{N_t}{N_t - 1} \frac{\sigma_k^2}{\tau_k^2} - \ln \frac{\sigma_k^2}{\tau_k^2} - \frac{N_s}{N_s - 1} + \ln \frac{N_s}{N_t} + \psi(N_t) - \psi(N_s), \qquad N_s, N_t \geq 2 \end{aligned} \tag{15}$$

where $\psi(\cdot)$ denotes the logarithmic derivative of the Gamma function, commonly referred to as the *digamma* function. Note that $\lim_{r\to\infty} \psi(r) = \ln r$, so $\lim_{N_s, N_t \to \infty} \mathrm{E}[\hat{L}|\mathcal{H}_1] = \mathrm{E}[L|\mathcal{H}_1]$ as we would expect. Similarly, the conditional variance of $\hat{\lambda}_k$ given $\mathcal{H}_1$ can be determined in closed form as

$$\begin{aligned} \mathrm{var}(\hat{\lambda}_k|\mathcal{H}_1) &= \int_0^\infty \int_0^\infty \int_0^\infty \left[ \frac{1}{2} \left( \frac{\sigma_k^2}{t^2} - \frac{\sigma_k^2}{s^2} \right) u - \ln \frac{s^2}{t^2} \right]^2 f_{\chi^2_2}(u) f_{S_k^2}(s^2) f_{T_k^2}(t^2) du ds^2 dt^2 - \mathrm{E}^2[\hat{\lambda}_k|\mathcal{H}_1] \\ &= \frac{N_t^3}{(N_t-1)^2(N_t-2)} \left( \frac{\sigma_k^2}{\tau_k^2} \right)^2 - 2 \frac{N_t(N_s N_t - 1)}{(N_t-1)^2(N_t-2)} \frac{\sigma_k^2}{\tau_k^2} + \\ &\quad \frac{N_s(N_s^2 - 2N_s + 4)}{(N_s-1)^2(N_s-2)} + \psi'(N_s) + \psi'(N_t), \qquad N_s, N_t \geq 3 \end{aligned} \tag{16}$$

where $\psi'$ is the derivative of the digamma function, commonly called the *trigamma* function. The trigamma function approaches zero as its argument gets large, and the third term approaches one as $N_s, N_t \to \infty$. As a result, $\lim_{N_s,N_t\to\infty} \mathrm{var}(\hat{\lambda}_k|\mathcal{H}_1) = (\sigma_k^2/\tau_k^2)^2 - 2\sigma_k^2/\tau_k^2 + 1$. Recalling that the components of $\mathbf{X}$ are statistically independent from (3), we get the expected result $\lim_{N_s,N_t\to\infty} \mathrm{var}(\hat{L}|\mathcal{H}_1) = \mathrm{var}(L|\mathcal{H}_1)$.

Again, under appropriate conditions $\hat{L}$ approaches a Gaussian distribution for long observation vectors $\mathbf{X}$. Under this assumption, the probability of error can be approximated as before,

$$\Pr[\hat{L} < \gamma|\mathcal{H}_1] \approx \Phi\left(\frac{\gamma - \mathrm{E}[\hat{L}|\mathcal{H}_1]}{\sqrt{\mathrm{var}(\hat{L}|\mathcal{H}_1)}}\right). \tag{17}$$

As will be shown in Section 4, the difference in error rates between tests with known and estimated parameters can be drastic when the amount of training data is small.

## 3. ESTIMATING THE TEST STATISTIC DISTRIBUTION

In the previous section we derived the mean and variance of the test statistics $L$ and $\hat{L}$ that result from known and estimated parameters, respectively. While these can be useful for approximating the probability of error, they require knowledge of the parameters $\sigma_k^2$ and $\tau_k^2$. In this section we derive unbiased estimators for the quantities $\mathrm{E}[\hat{L}|\mathcal{H}_1]$, $\mathrm{var}(\hat{L}|\mathcal{H}_1)$, $\mathrm{E}[L|\mathcal{H}_1]$, and $\mathrm{var}(L|\mathcal{H}_1)$ directly in terms of the trained parameters $S_k^2$ and $T_k^2$. With these we can estimate the classification error rate we should expect given that we only have estimates of $\sigma_k^2$ and $\tau_k^2$, as well as the error rate we could expect from the ideal classifier in which these parameters were known.

### 3.1. Estimating the Actual Probability of Error

Given estimates $S_k^2$ and $T_k^2$ computed from training data, we seek unbiased estimates $\widehat{\mathrm{E}[\hat{L}|\mathcal{H}_1]}$ of $\mathrm{E}[\hat{L}|\mathcal{H}_1]$ and $\widehat{\mathrm{var}(\hat{L}|\mathcal{H}_1)}$ of $\mathrm{var}(\hat{L}|\mathcal{H}_1)$. Simply substituting the parameter estimates into (15) and (16) yields biased estimates. Unbiased estimates for these quantities are

$$\widehat{\mathrm{E}[\hat{L}|\mathcal{H}_1]} = \sum_{k=1}^{K}\left(\frac{S_k^2}{T_k^2} - \ln\frac{S_k^2}{T_k^2} - 1 - \frac{1}{N_s - 1}\right) \tag{18}$$

and

$$\widehat{\mathrm{var}(\hat{L}|\mathcal{H}_1)} = \sum_{k=1}^{K}\left[\frac{N_s N_t}{(N_s-1)(N_t-1)}\left(\frac{S_k^2}{T_k^2}\right)^2 - \frac{2(N_sN_t-1)}{(N_s-1)(N_t-1)}\frac{S_k^2}{T_k^2} + \frac{N_s(N_s^2-2N_s+4)}{(N_s-1)^2(N_s-2)} + \psi'(N_s) + \psi'(N_t)\right]. \tag{19}$$

Notice that as $N_s$ and $N_t$ grow large we expect these estimates to approach the mean and variance of the ideal log-likelihood.

With these quantities, we can estimate the conditional probability of error that will result when the test of (9) is used and the parameters have been estimated from available training data. The probability of error estimator is

$$\widehat{\Pr[\hat{L} < \gamma|\mathcal{H}_1]} = \Phi\left(\frac{\gamma - \widehat{\mathrm{E}[\hat{L}|\mathcal{H}_1]}}{\sqrt{\widehat{\mathrm{var}(\hat{L}|\mathcal{H}_1)}}}\right). \tag{20}$$

Note that while this estimate of error is a function of unbiased estimates, it will not itself be unbiased.

### 3.2. Estimating the Ideal Probability of Error

Given estimates $S_k^2$ and $T_k^2$ computed from training data, we seek unbiased estimates $\widehat{\mathrm{E}[L|\mathcal{H}_1]}$ of $\mathrm{E}[L|\mathcal{H}_1]$ and $\widehat{\mathrm{var}(L|\mathcal{H}_1)}$ of $\mathrm{var}(L|\mathcal{H}_1)$. Simply substituting the parameter estimates into (6) and (7) yields biased estimators. However, we can construct unbiased estimates as

$$\widehat{\mathrm{E}[L|\mathcal{H}_1]} = \sum_{k=1}^{K} \left[ \left( \frac{S_k^2}{T_k^2} - \ln \frac{S_k^2}{T_k^2} - 1 \right) - \left( \frac{1}{N_t} \frac{S_k^2}{T_k^2} + \ln \frac{N_s}{N_t} - \psi(N_s) + \psi(N_t) \right) \right] \tag{21}$$

and

$$\widehat{\mathrm{var}(L|\mathcal{H}_1)} = \sum_{k=1}^{K} \left[ \left( \frac{S_k^2}{T_k^2} - 1 \right)^2 - \left( \frac{N_t^2 + N_s(3N_t - 2)}{N_t^2(N_s+1)} \left( \frac{S_k^2}{T_k^2} \right)^2 - \frac{2}{N_t} \frac{S_k^2}{T_k^2} \right) \right]. \tag{22}$$

We see that these approach the true values in (6) and (7) as the amount of training data grows large. We must use the variance estimate with some care because it can become negative if the ratio $S_k^2/T_k^2$ is close to one for most values of $k$. In practice we may take the variance to be zero when the above estimator assumes a negative value.

With these quantities, we can estimate the conditional probability of error in the ideal case (if $\sigma_k^2$ and $\tau_k^2$ were known) given only the estimates $S_k^2$ and $T_k^2$. This estimator is

$$\mathrm{Pr}[\widehat{L < \gamma}|\mathcal{H}_1] = \Phi\left( \frac{\gamma - \widehat{\mathrm{E}[L|\mathcal{H}_1]}}{\sqrt{\widehat{\mathrm{var}(L|\mathcal{H}_1)}}} \right). \tag{23}$$

As in the previous section, this estimate of error is a function of unbiased estimates but will not itself be unbiased.

One potentially useful application of this error estimate is to determine the relative value of increasing the size of training sets. That is, the quantity derived above is an estimate of the optimal probability of error. The quantity derived in Section 3.1 is an estimate of conditional error given the current training data. The difference

$$\Phi\left( \frac{\gamma - \widehat{\mathrm{E}[\hat{L}|\mathcal{H}_1]}}{\sqrt{\widehat{\mathrm{var}(\hat{L}|\mathcal{H}_1)}}} \right) - \Phi\left( \frac{\gamma - \widehat{\mathrm{E}[L|\mathcal{H}_1]}}{\sqrt{\widehat{\mathrm{var}(L|\mathcal{H}_1)}}} \right) \tag{24}$$

represents an estimate of how much better the system could perform if significantly more training data were collected, since we expect the parameters $S_k^2$ and $T_k^2$ to converge to the true values. If this difference were small for a given training data set, there may be no justification for collecting more data.

## 4. EXAMPLES

In this section we look at applications of the estimators of the previous section. We first present a simple (perhaps unrealistic) experiment to demonstrate the behavior of each estimator. These estimators are shown to capture the severe performance degradation that can occur when using trained parameters. It is also shown that these estimators are superior to direct substitution of the trained parameters into the expressions of Section 2.1, which apply only when the true parameters are known. A second experiment demonstrates the estimators in a more realistic example involving simulated high resolution radar data.

### 4.1. Simple Experiment

In this experiment we address a simple scenario in which $\sigma_k^2 = 1$ and $\tau_k^2 = 2$ for $k = 1, 2, \ldots, K$, and $\gamma = 1$. We assume that the actual values of these parameters are unknown. Further, we assume it is not known that the parameters are constant under each hypothesis. The experiment is to produce estimates of $S_k^2$ and $T_k^2$ from training samples, estimate with (20) the probability of error obtainable with the actual test of (9), estimate with (23) the best-case probability of error obtainable through the ideal test of (4), and to compare these estimates with corresponding error rates obtained through Monte Carlo simulation. We also compare these estimates to the more direct approach of substituting the sample parameters directly into error expression (8). On each of one million simulation trials a new set of training data and a new test observation are randomly generated.

Figure 1 shows distribution of $\hat{L}$ conditioned on $\mathcal{H}_1$ for observation vectors of length $K = 100$ and $N_s = N_t = 3$ training samples. For this problem, $\mathrm{E}[\hat{L}|\mathcal{H}_1] \approx -5.7$ and $\mathrm{var}(\hat{L}|\mathcal{H}_1) \approx 473$. The conditional PDF given $\mathcal{H}_1$ arising from a Gaussian approximation to $\hat{L}$ is represented by the dashed line in Panel (A). The sample PDF of $\hat{L}$ generated from one million trials is represented by the solid line. The mean and variance of the sample distribution are nearly equal to their theoretical values. However, the sample distribution is clearly asymmetric and leptokurtic (sharply peaked), reflecting the fact that Central Limit Theorem arguments are only approximate for finite vector lengths. As a result, use of the Gaussian assumption causes a slight positive bias in estimates of conditional error probability.

The solid curve ("Actual") of Panel (B) represents the sample PDF of the estimates $\widehat{\mathrm{E}[\hat{L}|\mathcal{H}_1]}$ from (18). The solid vertical line shows the theoretical value $\mathrm{E}[\hat{L}|\mathcal{H}_1]$. Similarly, the dashed curve ("Ideal") shows the sample PDF of the estimates $\widehat{\mathrm{E}[L|\mathcal{H}_1]}$ from (21), and the dashed vertical line indicates the theoretical value $\mathrm{E}[L|\mathcal{H}_1]$. These results indicate that the expected values of the test statistics can be estimated well given trained parameters $S_k^2$ and $T_k^2$ for both the actual and ideal tests. The dash-dotted curve ("Plug-In") shows the distribution of mean values obtained when $S_k^2$ and $T_k^2$ are substituted into (6). This result demonstrates that substituting trained parameters into the expressions governing error rates in the known parameter case will not yield reasonable estimates for either the actual or ideal tests.

The solid curve of Panel (C) shows the sample PDF of the variance estimates $\widehat{\mathrm{var}(\hat{L}|\mathcal{H}_1)}$ from (19), and the solid vertical line shows the theoretical value $\mathrm{var}(\hat{L}|\mathcal{H}_1)$. The dashed curve near the left edge of the plot represents the sample PDF of $\widehat{\mathrm{var}(L|\mathcal{H}_1)}$ from (22), and the dashed vertical line represents the theoretical value $\mathrm{var}(L|\mathcal{H}_1)$. These results demonstrate that the test statistic variance in both the actual and ideal tests can be estimated well using the trained parameters $S_k^2$ and $T_k^2$. The dash-dotted curve shows the distribution of variance values when $S_k^2$ and $T_k^2$ are substituted into (7). This result demonstrates that substituting trained parameters into the expressions governing error rates in the known parameter case will not yield accurate estimates for either the actual or ideal tests.

The solid curve of Panel (D) shows the sample PDF of the estimated probability of error $\widehat{\Pr[\hat{L} < \gamma|\mathcal{H}_1]}$ from (20), and the solid vertical line represents the experimentally determined probability of error. The estimated probability of error averages 60%, which is slightly larger than 57% achieved in the simulation. This slight positive bias is due to the non-Gaussianity of $\hat{L}$ noted in Panel (A). The dashed curve shows the sample PDF of the estimated probability of error $\widehat{\Pr[L < \gamma|\mathcal{H}_1]}$ from (23). The curve is sharply peaked, and most of its vertical extent has been cropped from the graph in order to show details. The experimentally determined probability of error in the ideal case (with known parameters) was 0.027%. These results demonstrate that it is possible to use trained parameters to estimate the probabilities of error in both actual and ideal tests.

Behavior of the test with trained parameters is not symmetric in the two hypotheses, as is shown in Figure 2. Panel (A) shows the conditional probability of error under each of the two hypotheses using the approximation in (17) as a function of $K$, the observation vector length. We see that $\Pr[\mathrm{error}|\mathcal{H}_1]$ increases monotonically with $K$ whereas $\Pr[\mathrm{error}|\mathcal{H}_2]$ decreases. The overall probability of error, assuming equal *a priori* probabilities on the two hypotheses, initially drops below 40% but begins to rise again beyond $K \approx 100$. This is an example of the peaking phenomena often reported with trained parameters.[13]

Panel (B) demonstrates the effect of increasing the number of training samples available for observations of length $K = 100$. We see that conditional error rates under $\mathcal{H}_2$ remain quite low for as few as 4 training samples. However, 12 or more samples are required for the conditional probability of error under $\mathcal{H}_1$ to be close to the ideal rate.

## 4.2. High Resolution Radar Experiment

In this section we apply the conditional error estimators (20) and (23) to simulated high resolution radar data. The data used are of the vehicles "T1 tank" and "M1 tank" from the University Research Initiative Synthetic

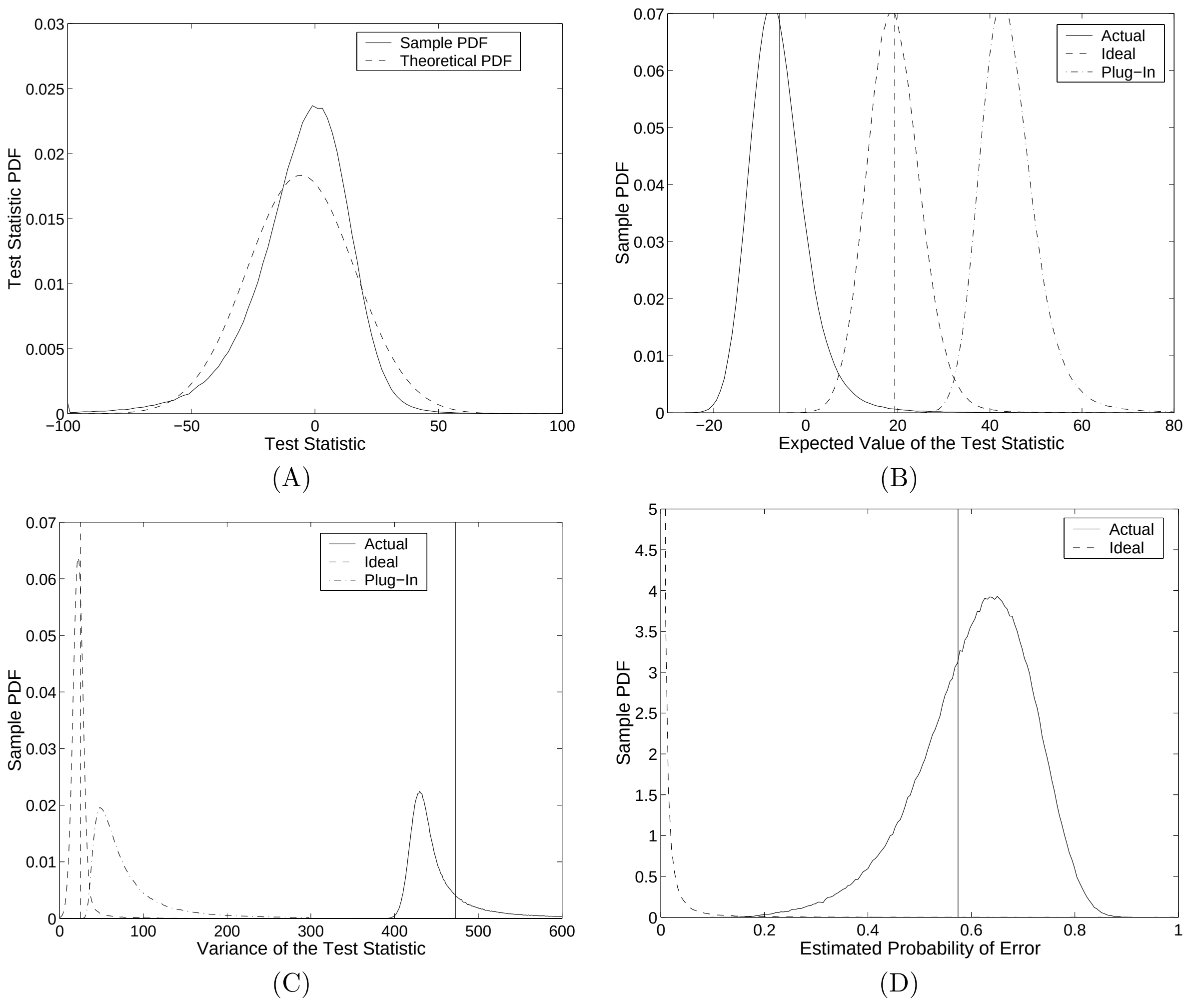
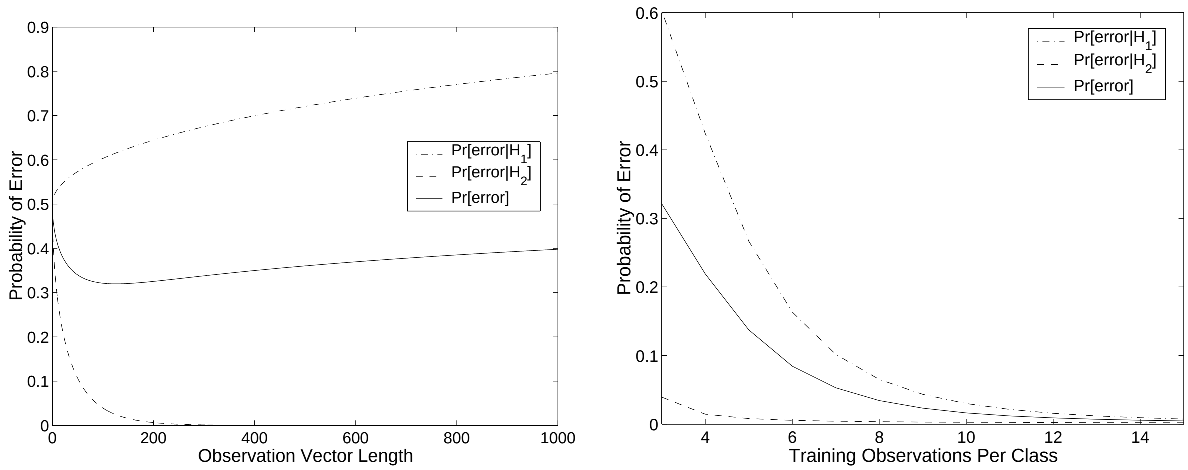
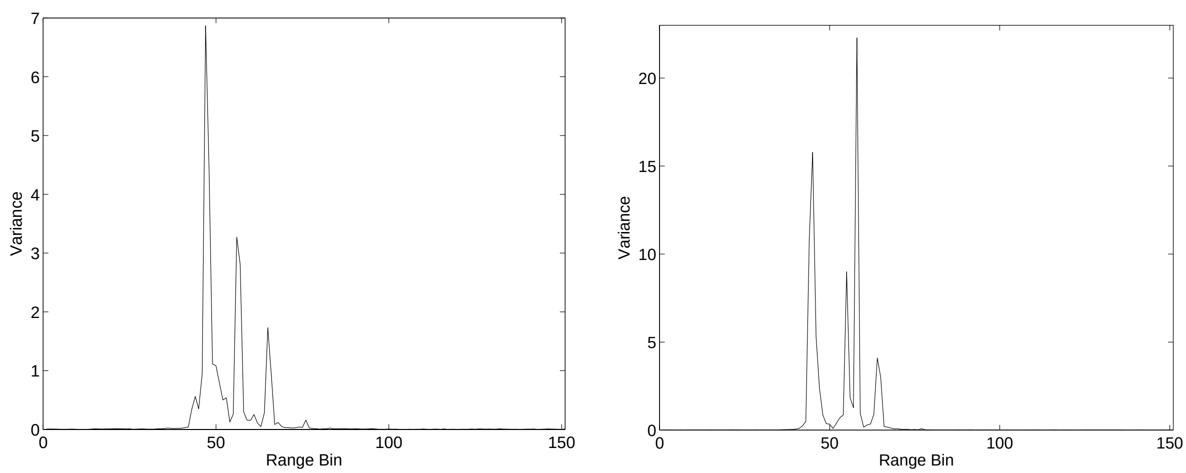



**Figure 1.** Distribution of error parameters estimated from statistics $S_k^2$ and $T_k^2$ over one million random trials. (A) Approximate and sample PDF for the test statistic $\hat{L}$ conditioned on $\mathcal{H}_1$. (B) Sample PDF of estimates $\widehat{\mathrm{E}[\hat{L}|\mathcal{H}_1]}$, $\widehat{\mathrm{E}[L|\mathcal{H}_1]}$, and those obtained by substituting ML parameter estimates directly into (6). (C) Sample PDF of estimates $\widehat{\mathrm{var}(\hat{L}|\mathcal{H}_1)}$, $\widehat{\mathrm{var}(L|\mathcal{H}_1)}$, and those obtained by substituting ML parameter estimates directly into (7). (D) Sample PDF of estimates $\mathrm{Pr}[\widehat{\hat{L} < \gamma}|\mathcal{H}_1]$ and $\mathrm{Pr}[\widehat{L < \gamma}|\mathcal{H}_1]$.



(A) (B)

**Figure 2.** Error rates in a simple classification problem. (A) Error as a function of observation vector length with three training samples. (B) Error as a function of training sample size with 100 element observation vectors.



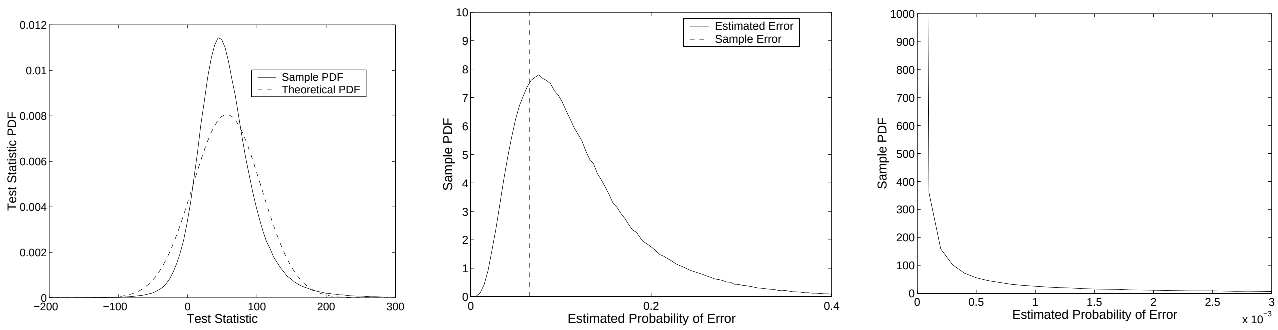
T1 Tank M1 Tank

**Figure 3.** Variance of the range profiles as a function of range bin for T1 and M1 tanks.

Dataset (URISD).[14] The data are simulated from CAD vehicle models using the signature prediction tool XPATCH. Signatures used in the experiment are vectors of $K = 151$ elements and represent UHF band measurements from a 10° elevation angle with the object nominally facing the radar. Figure 3 shows variance values estimated from 20 observations spanning a 6° range in azimuth for both vehicles.

The variance profiles of Figure 3 were taken as the "true" variance parameters $\sigma_k^2$ and $\tau_k^2$ in a test of $\mathcal{H}_1$, an assertion that an observation represents the T1 tank, versus $\mathcal{H}_2$, an assertion that an observation represents the M1 tank. In each of one million trials, $N_s = N_t = 3$ training samples were drawn, estimates $S_k^2$ and $T_k^2$ were computed, and an observation drawn according to $\mathcal{H}_1$ was classified using both the trained classifier (9) and the ideal classifier (4).

The results are summarized in Figure 4. Panel (A) shows the the sample PDF of conditional test statistic values $\hat{L}$ (solid curve) and the density obtained by a Gaussian approximation (dashed curve) with mean and variance determined from (15) and (16), respectively. Differences between the two curves is primarily due to




(A) (B) (C)

**Figure 4.** Estimated probability of error for sample data from the URISD. (A) Approximate and sample PDF for the test statistic $\hat{L}$. (B) Distribution of estimates $\Pr[\widehat{\hat{L} < \gamma}|\mathcal{H}_1]$. (C) Distribution of estimates $\Pr[\widehat{L < \gamma}|\mathcal{H}_1]$.

higher order effects. The skewness of sample $\hat{L}$ values is $\sqrt{\beta_1} = 3.35$ (as opposed to 0 for a Gaussian distribution), and the sample kurtosis is $\beta_2 = 88.18$ (as opposed to 3 for the Gaussian). By contrast, the sample mean differed from the theoretical mean of 56.57 by only 0.05%, and the sample variance differed from the theoretical variance of 2448.7 by 2%. As a result, the tails of the sample PDF are much higher than in the Gaussian approximation. This difference is due to the finite-length observation vectors employed, and we expect the Central Limit Theorem approximation to fit better with longer observations.

The solid curve in Panel (B) shows the PDF of estimated conditional probability of error $\Pr[\widehat{\hat{L} < \gamma}|\mathcal{H}_1]$, and the vertical dashed line represents the sample error rate of 6.54%. The approximate probability of error using (17) (based on true $\sigma_k^2$ and $\tau_k^2$ values) was nearly twice as high at 12.7%, and the average of the estimates $\Pr[\widehat{\hat{L} < \gamma}|\mathcal{H}_1]$ (based on trained $S_k^2$ and $T_k^2$ values) was similarly high at 11.9%. This discrepancy is due to the asymmetric and leptokurtic properties of $\hat{L}$ demonstrated in Panel (A). Nevertheless, the factor of 2 difference in estimated versus sample probability of error is better than can be obtained by substituting the trained parameters in place of the true parameters in (8). On average this more direct approach estimates a 1.2% conditional probability of error, which differs from the sample error by a factor of nearly 5.5.

Panel (C) shows the PDF of estimated conditional probability of error for the ideal case, $\Pr[\widehat{L < \gamma}|\mathcal{H}_1]$. The average estimate was 0.032%, a factor of 2 larger than the approximate error rate of 0.016% obtained from (8) using the true parameter values. In one million trials, none of the sample observations under $\mathcal{H}_1$ were misclassified by the ideal test.

## 5. CONCLUSIONS

We have considered in detail a variety of statistics related to a complex Gaussian assumption that is appropriate for radar target recognition systems. We have presented expressions for the conditional mean and variance of the test statistics that arise in both ideal tests (constructed with known parameters) and actual tests (constructed with trained parameters). These lead directly to approximations to the class-conditional probability of error of both tests. We have presented unbiased estimators of these conditional means and variances in terms of the trained parameters, and we have used these to construct estimators for the conditional probabilities of error in actual and ideal tests. These allow a designer to estimate the performance of a radar-based ATR system given the trained parameters as well as the performance that could be obtained if more training data were gathered. Finally, these estimators were demonstrated in detail on two sets of simulated data, one highly simplified and the other from simulated radar data.

The examples show that the mean and variance estimators perform quite well and that the estimated error probabilities have a positive bias. This bias was shown to result from the non-Gaussian distribution assumed by the test statistic when finite length observation vectors are used. These error estimates were contrasted with plug-in error estimates which tend to be negatively biased, and experiments showed them to be more accurate on average. The key limiting factor is the Gaussian assumption on the test statistic, and ongoing work is focused on replacing this with a better distribution family. These results are also being extended to address true distributions that come from an entirely different family from those assumed.

## ACKNOWLEDGMENTS

This work was supported by the Office of Naval Research grant N00014-03-1-0110.

## REFERENCES

1. K. Fukunaga, *Introduction to Statistical Pattern Recognition*, Academic Press, second ed., 1990.
2. N. Glick, "Sample-based classification procedures derived from density estimators," *Journal of the American Statistical Association* **67**, pp. 116–122, Mar. 1972.
3. W. Feller, *An Introduction to Probability Theory and Its Applications*, vol. II, John Wiley and Sons, Inc., 1966.
4. E. B. Manoukian, *Mathematical Nonparametric Statistics*, Gordon and Breach Science Publishers, 1986.
5. M. Kendall and A. Stuart, *The Advanced Theory of Statistics*, vol. 1, Macmillan Publishing Co., fourth ed., 1977.
6. H. L. Van Trees, *Detection, Estimation, and Modulation Theory. Radar-Sonar Signal Processing and Gaussian Signals in Noise*, vol. III, John Wiley & Sons, 1971.
7. S. P. Jacobs and J. A. O'Sullivan, "Automatic target recognition using sequences of high resolution radar range-profiles," *IEEE Transactions on Aerospace and Electronic Systems* **36**, pp. 364–382, Apr. 2000.
8. J. A. O'Sullivan, M. D. DeVore, V. Kedia, and M. I. Miller, "Automatic target recognition performance for SAR imagery using a conditionally Gaussian model," *IEEE Transactions on Aerospace and Electronic Systems* **37**, pp. 91–108, Jan. 2001.
9. M. D. DeVore and J. A. O'Sullivan, "A performance-complexity study of several approaches to automatic target recognition from SAR images," *IEEE Transactions on Aerospace and Electronic Systems* **38**, pp. 632–648, Apr. 2000.
10. M. D. DeVore and J. A. O'Sullivan, "Target-centered models and information-theoretic segmentation for automatic target recognition," *Multidimensional Systems and Signal Processing* **14**, pp. 139–159, Jan. 2003.
11. M. D. DeVore and J. A. O'Sullivan, "Quantitative statistical assessment of conditional models for synthetic aperture radar," *IEEE Transactions on Image Processing* **13**, pp. 113–125, Feb. 2003.
12. N. A. Schmid and J. A. O'Sullivan, "Thresholding method for reduction of dimensionality," *IEEE Transactions on Information Theory* **47**, pp. 2903–2920, Nov. 2001.
13. G. V. Trunk, "A problem of dimensionality: A simple example," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **PAMI-1**, pp. 29–30, July 1979.
14. Johns Hopkins University, Center for Imaging Science, "University research initiative synthetic dataset (HRR)." `http://cis.jhu.edu/data.sets/urisd`.