Purdue University
Purdue e-Pubs

LARS Symposia | Laboratory for Applications of Remote Sensing

1-1-1976

Estimation of the Probability of Error without Ground Truth and Known A Priori Probabilities

K. A. Havens

T. C. Minter

S. G. Thadani

Follow this and additional works at: http://docs.lib.purdue.edu/lars_symp

Havens, K. A.; Minter, T. C.; and Thadani, S. G., "Estimation of the Probability of Error without Ground Truth and Known A Priori Probabilities" (1976). *LARS Symposia*. Paper 134.
http://docs.lib.purdue.edu/lars_symp/134

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.

Reprinted from

Symposium on Machine Processing of Remotely Sensed Data

June 29 - July 1, 1976

The Laboratory for Applications of
Remote Sensing

Purdue University
West Lafayette
Indiana

IEEE Catalog No.
76CH1103-1 MPRSD

Copyright © 1976 IEEE
The Institute of Electrical and Electronics Engineers, Inc.

Copyright © 2004 IEEE. This material is provided with permission of the IEEE. Such permission of the IEEE does not in any way imply IEEE endorsement of any of the products or services of the Purdue Research Foundation/University. Internal or personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution must be obtained from the IEEE by writing to pubs-permissions@ieee.org.

By choosing to view this document, you agree to all provisions of the copyright laws protecting it.

ESTIMATION OF THE PROBABILITY OF ERROR WITHOUT GROUND TRUTH AND KNOWN *A PRIORI* PROBABILITIES*

K. A. Havens, T. C. Minter and
S. G. Thadani

Lockheed Electronics Company, Inc./Aerospace Systems Division, Houston, Texas

I. ABSTRACT

The probability of error or, alternatively, the probability of correct classification (PCC) is an important criterion in analyzing the performance of a classifier. Labeled samples (those with ground truth) are usually employed to evaluate the performance of a classifier. Occasionally, the numbers of labeled samples are inadequate, or no labeled samples are available to evaluate a classifier's performance; for example, when crop signatures from one area from which ground truth is available are used to classify another area from which no ground truth is available. This paper reports the results of an experiment to estimate the probability of error using unlabeled test samples (i.e., without the aid of ground truth).

II. INTRODUCTION

This paper presents the results of an experiment to estimate the probability of error using unlabeled samples. Two procedures, along with the test results of each, are presented. The first procedure estimates the probability of error analytically, using the posterior density function. (The analytical estimate is shown to be unbiased.) The second labels fields (for use in estimating the probability of error) simply by noting the class into which most of the field picture elements (pixels) were classified by the classifier (called the *majority rule method*). Empirical results from both procedures are presented, and an analysis of variance from which conclusions are drawn is performed.

*The material in this paper was developed under NASA Contract NAS 9-12200 and prepared for the Earth Observations Division, NASA/JSC, Houston, Texas.

III. DESCRIPTION OF THE TWO PROCEDURES

A. Estimating the Probability of Error Without Ground Truth

Let X represent an unlabeled sample which is a random, independent, n-dimensional measurement vector belonging to one of m classes. These classes are assumed to be normally distributed with *a priori* probabilities given by q_i and the conditional probability density function given by $p(X/i)$, $i = 1,\cdots,m$. Then, the overall mixture density is denoted as

$$p(X) = \sum_{i=1}^{m} q_i p(X/i) \tag{1}$$

The Bayes decision rule defines a Bayes region R_i for each class i.

$$R_i = [X | q_i p(X/i) \geq q_k p(X/k)] \; ; \quad k = 1,\cdots,m \tag{2}$$

Thus, the probability of error is defined as

$$Pr(Er) = \sum_{i=1}^{m} \sum_{\substack{j=1 \\ i \neq j}}^{m} q_i Pr(j/i) \tag{3}$$

where $Pr(j/i)$ is the probability of misclassifying a sample from class i into class j.

$$Pr(j/i) = \int_{R_j} p(X/i)\,dX \tag{4}$$

Multiplying $p(j/i)$ by $\frac{p(X)}{p(X)}$ in equation (4),

$$Pr(j/i) = \int_{R_j} \left[\frac{p(X/i)}{p(X)}\right][p(X)\,dX] \tag{5}$$

Let

$$f(i/X) = \frac{p(X/i)}{p(X)} \tag{6}$$

and, if $\frac{p(X/i)}{p(X)}$ is replaced by $f(i/X)$, the result is

$$Pr(j/i) = \int_{R_j} f(i/X)p(X)dX \tag{7}$$

From equation (7), it is shown that the probability of classifying a pixel from class i into class j is the "expected value" of $f(i/X)$ over the Bayes region R_j; that is,

$$Pr(j/i) = E_{R_j}[f(i/X)] \tag{8}$$

Likewise, it can be shown that the PCC for class i, $Pr(i/i)$, is

$$Pr(i/i) = E_{R_i}[f(i/X)] \tag{9}$$

These expressions for $Pr(j/i)$ and $Pr(i/i)$ have two useful properties [in the following, $\hat{P}r(Er)$ will represent the empirical estimate of $Pr(Er)$]:

1. $Pr(j/i)$ and $Pr(i/i)$ can be estimated using unlabeled samples.

2. An upper bound is known for the variance of $\hat{P}r(Er)$ in the two-class case.

To implement this analytical procedure, an assumption is made which slightly amends the computational procedure. It is assumed that the proportion of pixels classified as wheat compared to the total number of pixels is a good estimate of the *a priori* PCC as wheat; that is,

$$q_w = \frac{\text{number of pixels classified as wheat } (N_w)}{\text{total number of pixels } (N_T)} \tag{10}$$

Using equations (6) and (9),

$$\hat{P}r(w/w) = \frac{1}{N_T}\sum_{k=1}^{N_w}\frac{p(X/w)}{p(X)}$$

$$= \frac{1}{q_w N_T}\sum_{k=1}^{N_w}\frac{q_w p(X/w)}{p(X)}$$

$$= \frac{1}{N_w}\sum_{k=1}^{N_w}\frac{q_w p(X/w)}{p(X)} \tag{11}$$

Unbiased Estimates of $p(j/i)$ and $p(i/i)$. By definition,

$$p(j/i) = \int_{R_j} f(i/X)p(X)dX$$

$$= \int_R \left[f(i/X)I_{R_j}(X)\right]p(X)dX \tag{12}$$

and

$$p(i/i) = \int_{R_i} f(i/X)p(X)dX$$

$$= \int_R \left[f(i/X)I_{R_i}(X)\right]p(X)dX \tag{13}$$

where

$$I_{R_j}(X) = \begin{cases} 1 & \text{if } X\varepsilon R_j \\ 0 & \text{otherwise} \end{cases}$$

$$I_{R_i}(X) = \begin{cases} 1 & \text{if } X\varepsilon R_i \\ 0 & \text{otherwise} \end{cases}$$

and $f(i/X)$ and $p(X)$ are completely known functions.

If $\{X_1,\cdots,X_N\}$ denotes a sample from a population with population density function $p(X)$ and if $X_1^{(j)},\cdots,X_{N_j}^{(j)}$ denote observations falling in R_j, then

$$N_1 + \cdots + N_m = N \tag{14}$$

where $[R_1,R_2,\cdots,R_m]$ is the partition and N_j denotes the number of observations falling in R_j.

$$\hat{p}(j/i) = \frac{1}{N}\sum_{\alpha=1}^{N} f(i/X_\alpha)I_{R_j}(X_\alpha)$$

$$= \frac{1}{N}\sum_{\ell=1}^{N_j} f\left[i/X_\ell^{(j)}\right] \tag{15}$$

and

$$\hat{p}(i/i) = \frac{1}{N}\sum_{\alpha=1}^{N} f(i/X_\alpha)I_{R_i}(X_\alpha)$$

$$= \frac{1}{N}\sum_{\ell=1}^{N_i} f\left[i/X_\ell^{(i)}\right] \tag{16}$$

Thus, $\hat{p}(j/i)$ and $\hat{p}(i/i)$ are unbiased estimates of $p(j/i)$ and $p(i/i)$ since X_α's are

identically distributed random variables with population density function p(X); and hence

$$E[\hat{p}(j/i)] = \frac{1}{N}\sum_{\alpha=1}^{N} E\left[f(i/X_\alpha)I_{R_j}(X_\alpha)\right]$$

$$= E\left[f(i/X)I_{R_j}(X)\right]$$

$$= p(j/i) \qquad (17)$$

and

$$E[\hat{p}(i/i)] = \frac{1}{N}\sum_{\alpha=1}^{N} E\left[f(i/X_\alpha)I_{R_i}(X_\alpha)\right]$$

$$= E\left[f(i/X)I_{R_i}(X)\right]$$

$$= p(i/i) \qquad (18)$$

Variance of $\hat{P}r(Er)$. For clarity of presentation, an upper bound on the variance of $\hat{P}r(Er)$ for a two-class case will be discussed[1]. First, some terms will be defined. The mixture density for two classes is given by

$$p(X) = q_1 p(X/1) + q_2 p(X/2) \qquad (19)$$

and

$$p(i/X) = \frac{q_i p(X/i)}{p(X)} \qquad (20)$$

denotes the posterior probability.

When X is classified according to the Bayes decision rule, the conditional probability of error is

$$r(X) = \min\ [p(1/X), p(2/X] \qquad (21)$$

When the expectation of r(X) is taken with respect to the random vector X, the Bayes error is given by

$$Pr(Er) = E[r(X)] \qquad (22)$$

and the Bayes error can be estimated by the sample mean of $r(X_i)$ for N_t test samples as

$$\hat{P}r(Er) = \frac{1}{N_T}\sum_{i=1}^{N_T} r(X_i) \qquad (23)$$

Thus, the variance of $\hat{P}r(Er)$ is

$$var[\hat{P}r(Er)] = \frac{var[r(X)]}{N_T} \qquad (24)$$

Since $\quad 0 \le r(X) \le \frac{1}{2} \qquad (25)$

then $\quad r^2(X) \le \frac{1}{2}r(X) \qquad (26)$

and $\quad E[r^2(X)] \le \frac{1}{2}E[r(X)] = \frac{1}{2}Pr(Er) \qquad (27)$

Therefore, a bound on var[r(X)] is

$$var[r(X)] = E[r^2(X)] - Pr^2(Er) \qquad (28)$$

$$\le \frac{1}{2}Pr(Er) - Pr^2(Er) \qquad (29)$$

$$= Pr(Er)[1 - Pr(Er)] - \frac{1}{2}Pr(Er) \qquad (30)$$

Thus,

$$var[\hat{P}r(Er)] \le \frac{1}{N_T}\left\{Pr(Er)[1 - Pr(Er)] - \frac{1}{2}Pr(Er)\right\} \qquad (31)$$

On the other hand, if class identification were available for the N_T test samples and an empirical error count were made, the error count would also give an unbiased estimate $\hat{P}r'(Er)$ of the Bayes error. The variance of this estimate is known to be[2]

$$var[\hat{P}r'(Er)] = \frac{Pr(Er)[1 - Pr(Er)]}{N_T} \qquad (32)$$

Thus, a reduction in variance can be achieved by using $\hat{P}r(Er)$.

B. Majority Rule Field Identification Procedure

The second estimate of the probability of error will be computed after the classification procedure is completed. The test fields are labeled wheat or nonwheat to agree with the classification of the majority of pixels in the particular field. This estimate will be called the *majority rule field identification procedure.*

Let N_i be the total number of pixels in field i, where $i = 1,\cdots,m$, and let N_w be the total number of pixels in all of those fields identified as wheat fields from the majority pixel count; N_{cw_i} will be the total number of pixels classified as wheat in field i. The probability of correctly classifying wheat will be estimated by

$$\hat{P}r(w/w) = \frac{N_1}{N_w}Pr_1(w/w) + \frac{N_2}{N_w}Pr_2(w/w) + \cdots + \frac{N_m}{N_w}Pr_m(w/w) \qquad (33)$$

where the probability of correctly classifying wheat in field i is given by

$$\hat{Pr}_i(w/w) = \frac{N_{cw_i}}{N_i} \qquad (34)$$

Thus,

$$Pr(w/w) = \frac{1}{N_w} \sum_{i=1}^{m} N_{cw_i} \qquad (35)$$

IV. DESCRIPTION OF THE EXPERIMENT

The data used in this experiment are multispectral scanner (MSS) data taken from the Earth Resources Technology Satellite (ERTS-1). [The ERTS-1 was renamed the first Land Satellite (Landsat-1) in January 1975.] The intensive test sites (ITS's) are 9- by 11-kilometer (5- by 6-nm) tracts in Morton and Finney Counties, Kansas, for which ground truth is available.

The test fields were chosen by randomly selecting 100 fields for each county from the ground truth information. Thirty training fields were chosen for each county, first by clustering the ITS into 30 clusters and then by selecting a representative number (10 wheat, 20 nonwheat) of fields of homogeneous clusters.

The classification procedures were applied to the test sites, and a paired t-test was used in evaluating the performance of the analytical and pixel-count procedures. In both cases, this test used the difference between the estimated PCC and the true PCC (TPCC). Additionally, analysis-of-variance tables were done for the PCC, the estimated probability of classification as wheat* (PCW), and the estimated probability of classification as "other" (PCO), as calculated from the analytical procedure.

V. RESULTS OF THE EXPERIMENT

By means of the analytical procedure, the random test fields in each county were classified using various combinations of ERTS-1 passes. The PCC was approximated by systematic samplings of the entire test area using the analytic estimate:

$$PCC = q_w p(w/w) + q_o p(o/o) \qquad (36)$$

To compare estimated PCC with the TPCC, a paired t-test was performed and is presented in table 1. Each entry in the column "mean PCC" represents the average of 30 estimates of PCC, each arising from a separate systematic sample from the same classification run. The rows of this table refer to the biological phase or phases and site used in that classification run. The t-test indicates that the PCC is a biased estimator. Upon further examination, the estimates of PCC were found to be 6.0 percentage points too high. The tolerance ($\pm t_{.05}\sigma$) of the bias was calculated to the ±1.5 percentage points.

Figure 1 compares the computed probability with the TPCC and shows the theoretical line on which the points should be found. Again, the PCC is biased toward a slightly higher estimate.

All approximations were tested further by an analysis of variance on the differences between the estimates and true values.

For the PCW, no significant differences were found among the sites or phases in the performance of the estimate. However, the mean F-value was significant. The bias was found to be 11 percentage points higher than the true value. The tolerance ($\pm t_{.05}\sigma$) of the bias was calculated to be ±6.6 percentage points. This supports the conclusion that the PCW, on the average and over all phases, overestimates the true values.

The PCO was found to be unbiased over sites, phases, or means.

For the PCC, significant differences were encountered in the means and sites. The average overestimation for Morton County was found to be 8.0 percentage points and for Finney County 4.0 percentage points. These corroborate the fact that the bias was 6.0 percentage points high over both sites. (It should be noted that only two sites were used, and the inferences made here must be confined to these particular sites.) Thus, the PCC is a consistent overestimator of the true value.

The results of using the pixel-count procedure to compare the estimated with the true values of the PCW and PCO are given in table 2. The statistical analysis of the pixel-count procedure includes the Student t-statistic for the PCW and PCO for all test combinations of passes. Most of the statistics indicated significance at the 5-percent level with no apparent dependence on the passes used. At this point the results from table 2 were graphed. From figure 2, the pixel-count estimate of the PCO seems to be insensitive to the TPCO. However, the range of true values is not wide enough to draw definite conclusions.

*In this paper, PCW and PCO will denote the probability of correctly classifying wheat and "other," respectively, as calculated by either the analytical procedure or the majority rule procedure.

From figure 3, the estimate of the PCW becomes more inaccurate as the TPCW takes values less than 90 percent.

VI. REFERENCES

1. Fukunaga, K.; and Kessell, D. L. "Nonparametric Bayes Error Estimation Using Unclassified Samples." *IEEE Transactions on Information Theory*, Vol. IT-19, July 1973.

2. Fukunaga, K. *Introduction to Statistical Pattern Recognition*. Academic Press (New York), 1972, p. 146.

Table 1. Computational Results of Analytical Procedure with Paired T-Test on PCC

Kansas ITS	Biological phase	Mean PCC	TPCC	Δ
Morton County	1	.864786	.751103	.113683
	2	.939788	.890700	.049088
	3	.759938	.684564	.075374
	4	.885723	.819559	.066164
	1,2	.974095	.897987	.076108
	1,2,3	.972914	.889166	.083748
	1,2,3,4	.985543	.885331	.100212
Finney County	1	.801935	.787341	.014594
	2	.914815	.887302	.027513
	3	.968512	.933616	.034896
	4	.946467	.901196	.045271
	1,2	.954650	.902354	.052296
	1,2,3	.988710	.940178	.048532
	1,2,3,4	.991599	.934002	.057597

Mean = 0.060363
Variance = 0.0007596
Standard deviation = 0.007366
T-statistic = 8.1948
Bias = 6.04
Tolerance of bias = ±1.51

Table 2. Computational Results for Pixel-Count Procedure

Pass	PCC	TPCC	PCW	TPCW	PCO	TPCO
Morton County						
1	.864786	.751103	.745967	.385343	.889899	.926767
2	.939788	.890700	.951765	.888889	.933004	.891570
3	.759938	.684564	.663261	.303191	.782292	.867726
4	.885723	.819559	.900269	.759462	.831856	.848425
1,2	.974095	.897987	.977242	.930260	.972150	.882865
1,2,3	.972914	.889166	.975404	.937943	.971287	.865739
1,2,3,4	.985543	.885331	.983918	.968085	.986713	.845586
Finney County						
1	.801935	.787341	.676597	.544186	.840346	.867934
2	.914815	.887302	.826309	.730232	.940840	.939363
3	.968512	.933616	.942830	.888372	.977528	.948612
4	.946467	.901196	.932542	.934884	.952884	.890031
1,2	.954650	.902354	.904808	.756589	.969130	.950668
1,2,3	.988710	.940178	.977139	.889922	.992654	.956834
1,2,3,4	.991599	.934002	.988017	.936434	.993025	.933196

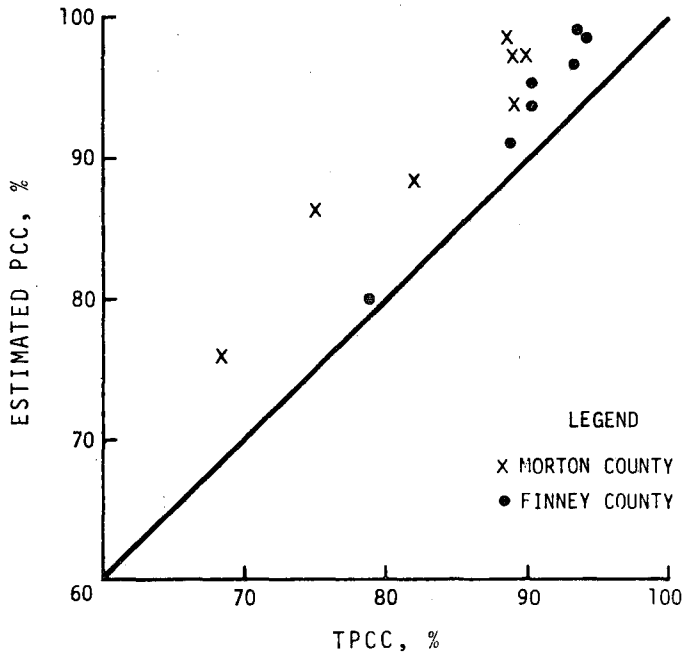

Figure 1. Comparison of Estimated Versus True Probability of Correct Classification for Analytical Procedure.

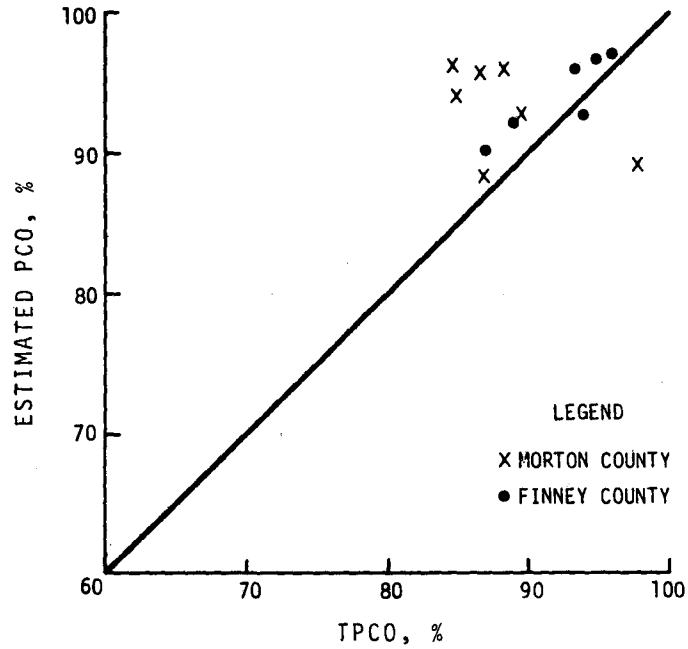

Figure 2. Comparison of the Pixel-Count Estimate Versus the True Probability of Classification as "Other."

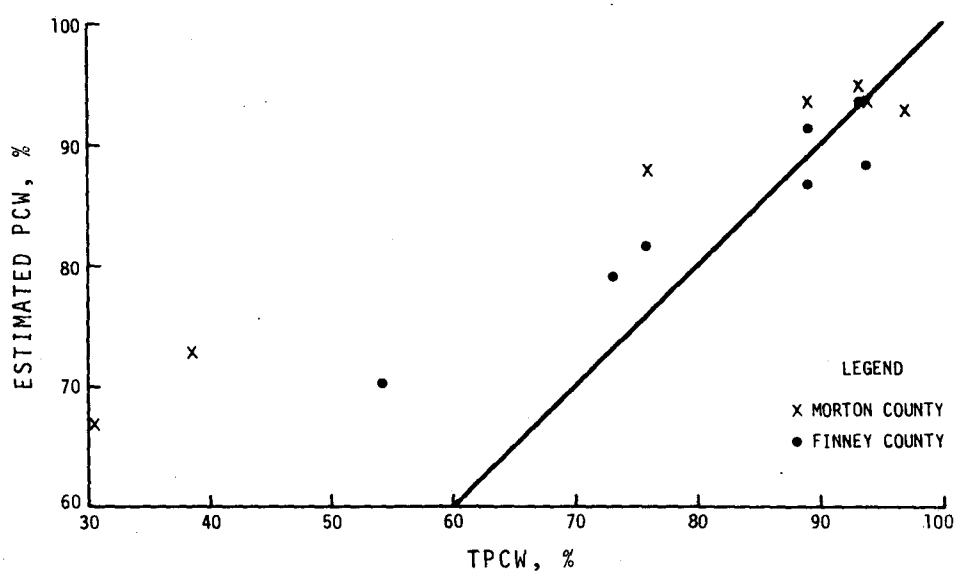

Figure 3. Comparison of the Pixel-Count Estimate Versus the True Probability of Classification as Wheat.