AN AD HOC MAP EVALUATION PROCEDURE

Job Order 75-315

(NASA-CR-147746) AN AD HOC MAP EVALUATION PROCEDURE (Lockheed Electronics Co.) 35 p HC $4.00 CSCL 08B

N76-24686

Unclas
G3/43 42524

Prepared By

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

Contract NAS 9-12200

For

EARTH OBSERVATIONS DIVISION
Science and Applications Directorate

National Aeronautics and Space Administration
LYNDON B. JOHNSON SPACE CENTER
Houston, Texas

April 1976

LEC-8278

JSC-11154

AN AD HOC MAP EVALUATION PROCEDURE

Job Order 75-315

PREPARED BY

E. P. Kan, Principal Scientist

APPROVED BY

I. E. Duggan, Supervisor
Forestry Applications Section

J. E. Davis, Manager
Earth Observations Exploratory Studies
Department

Prepared By

Lockheed Electronics Company, Inc.

For

Earth Observations Division

Science and Applications Directorate

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
LYNDON B. JOHNSON SPACE CENTER
HOUSTON, TEXAS

April 1976

LEC-8278

TECHNICAL REPORT INDEX/ABSTRACT
(See instructions on reverse side.)

1. TITLE AND SUBTITLE OF DOCUMENT
AN AD HOC MAP EVALUATION PROCEDURE

2. JSC NO.
JSC-11154

3. CONTRACTOR/ORGANIZATION NAME
Lockheed Electronics Company, Inc.

4. CONTRACT OR GRANT NO.
NAS 9-12200

5. CONTRACTOR/ORIGINATOR DOCUMENT NO.
LEC-8278

6. PUBLICATION DATE (THIS ISSUE)
April 1976

7. SECURITY CLASSIFICATION
Unclassified

8. OPR (OFFICE OF PRIMARY RESPONSIBILITY)
Earth Observations Division

9. LIMITATIONS
GOVERNMENT HAS UNLIMITED RIGHTS ☐ YES ☒ NO

IF NO, STATE LIMITATIONS AND AUTHORITY

10. AUTHOR(S)
E. P. Kan

11. DOCUMENT CONTRACT REFERENCES
WORK BREAKDOWN STRUCTURE NO.
Job Order 75-315

CONTRACT EXHIBIT NO.

DRL NO. AND REVISION

DRL LINE ITEM NO.

12. HARDWARE CONFIGURATION
SYSTEM

SUBSYSTEM

MAJOR EQUIPMENT GROUP

13. ABSTRACT

An ad hoc map evaluation procedure is proposed which is most suitable for evaluating low-resolution classification maps against high-resolution ground-truth maps, such as Land Satellite maps against interpreted aircraft photographs. Commonly practiced sampling and evaluation procedures are impracticable in this context because of difficulties in registration and in comparing the samples, as experienced in the recent Tri-County Pilot Study performed in the Forestry Applications Project. This ad hoc procedure is designed to overcome these two major problems, and its practicability is discussed.

Two widely accepted parameters are estimated by the new procedure; namely, the probability of correct classification and the proportion biases. Statistical qualifications are also provided.

14. SUBJECT TERMS

Statistical evaluation

Remote sensing

Map evaluation

JSC Form 833 (Rev Sep 74) NASA-JSC

PREFACE

An ad hoc procedure was proposed, discussed, and agreed upon in a series of group meetings attended by the following personnel in the Forestry Applications Project: T. Austin, R. Dillman, E. Downes, E. Kan, A. Kerber, C. Reeves, and J. Ward. The successes and failures experienced in the evaluation process performed in the Tri-County Pilot Study provided useful insight into this ad hoc design. The invaluable reviews and comments on this paper, contributed by T. Austin and by the Forestry Applications Project Scientist, R. W. Douglass, are hereby acknowledged.

CONTENTS

TABLES

Table | Page

FIGURES

Figure | Page

PRECEDING PAGE BLANK NOT FILMED

1. INTRODUCTION

Evaluation of map accuracies by sampling and estimating the overall probability of correct classification (PCC) has been discussed in theory and practice (ref. 1). This procedure has been demonstrated (ref. 2) to be more appropriate for overall map evaluation than procedures using training class accuracy, average accuracy by class, and others which have been commonly practiced in remote sensing applications (ref. 3).

The procedure of sampling and estimating the PCC to evaluate the soils resource inventory maps prepared by the Forestry Applications Project is documented in reference 4. When interpreted soil maps at scales near 1:60 000 were checked against ground samples and against U.S. Forest Service base maps at similar scales, the procedure proved practicable and practical. (Soils resource maps were prepared by interpreting aircraft photography.) In those resource maps, landform features were normally large enough or wide enough to permit grid cell systems (refs. 1 and 5) with sizable grid cell samples [e.g., cell samples of 5 millimeters square (0.19 inch square)]. "Sizable" is a relative term, but here it is a very important concept because it implies workable, practicable, and perhaps practical.

The same basic procedure was used in the Tri-County Pilot Study (TRICPS, ref. 6) but the use of the procedure was frustrating and gave PCC estimates much lower than expected. These conclusions were based on (1) registration inaccuracy (It was almost impossible to reliably locate selected samples on the classification map and on the ground-truth map.) and (2) comparison inaccuracy [caused by the size of the sample and the majority rule in deciding the type (i.e., class or feature) to be associated with the sample].

An analysis of the experimental design of the evaluation procedure used in TRICPS revealed facts that explained the two kinds of inaccuracy. The TRICPS developed classification maps of the first Land Satellite (Landsat-1) multispectral scanner data at approximately 60 meters (197 feet) resolution [i.e., each picture element (pixel) is 60 meters square (197 feet square) on the ground]. These low resolution maps were compared with aircraft photography which was interpreted at sample locations and used as ground truth. At the 1:120 000 scales, the aircraft photography had much higher resolution than the Landsat maps. Knowing that it was impossible to use the generic Landsat pixels as sample units, a 3- by 3-pixel sample unit was used. The majority rule was also employed for assigning a class type on Landsat and photography data samples. (Details of the TRICPS evaluation procedure can be found in appendix A.) The evaluation process was further complicated by the fact that Landsat classification maps are often spotty and that narrow features such as hardwood stringers often gave rise to multiclass samples such that the majority-rule decision appeared shaky, although the majority rule is adequate for competition between only two classes.

As a result of the inaccuracies found in the evaluation procedure used in TRICPS, an ad hoc map evaluation procedure is proposed for cases when the low resolution classification maps are evaluated against the high resolution ground-truth maps. This ad hoc procedure minimizes registration inaccuracies and comparison inaccuracies. The procedure attempts to evaluate the per-pixel classification accuracy of the map by using a sample size which is small enough (2 by 2 pixels) to reflect pixel classifications and which is large enough to absorb some possible errors caused by misregistration and mixture pixels. As a result, PCC and proportion biases B are estimated. The latter measure of proportion biases is actually a byproduct of PCC calculations and is a secondary, suboptimally designed measure for the present evaluation design.

The biases provide accuracy measures for areal measurements which are prime objectives of some investigations.

Section 2 describes the procedure by using an example, thereby providing better insight into the novelty and motivation of the ad hoc design as discussed in section 3. Section 4 gives statistical qualifications of evaluation parameters and discusses the implications of the proportion bias measure and its relation to the PCC. Practical considerations in using the ad hoc procedure are discussed in section 5. Appendix B also explains the choice of the decision rule and its accompanying threshold to determine sample correct classification; the decision rule is shown to be more discriminatory than the widely used majority rule.

2. AD HOC PROCEDURE

The design of this procedure is based on three assumptions: (1) a large Landsat classification map (i.e., over 500 000 pixels) is evaluated against interpreted aircraft photography at scales of 1:60 000 or 1:120 000, (2) sample locations can be located to within one Landsat pixel (step 2 of the procedure), and (3) no dramatic classification error exists (e.g., errors as a result of consistently classifying forest as nonforest and vice versa).

2.1 PROCEDURE DESCRIPTION

The natural grid system on the classification map is the pixel grid. Assuming a total of K classes (types or features) is represented in the map, including a class of "others," these steps should be followed.

Step 1 — Randomly select M (see section 4.1) primary sampling unit (PSU) on the classification map. Each PSU is the size of 50 by 50 Landsat pixels.

Step 2 — Locate the same M PSU's on the photograph, using distinguishing features such as roads, intersections, and landmarks. Local registration error should be minimized to within 1 Landsat pixel.

Step 3 — Randomly select 10 secondary sampling units (SSU) within each PSU on the classification map; each SSU is the size of 2 by 2 Landsat pixels.

Step 4 — Using the PSU framework on the photograph as determined in step 2, locate the same SSU's on the photograph.

Step 5 — Determine the proportion of each class in each SSU on the photograph. Denote by p_{mnk} (where $m = 1,\cdots,M$; $n = 1,\cdots,10$; and $k = 1,\cdots,K$) the proportion of the k*th* class interpreted in the n*th* SSU of the m*th* PSU.

Step 6 – For each of the nine possible locations on or about the selected mn*th* SSU on the Landsat map, determine the proportions $\hat{p}_{mnk}^{(i)}$, i = 1,···,9, by counting pixels. (See section 2.2 for clarification of this step.)

Step 7 – For each of the nine locations in each mn*th* SSU, determine the error defined as:

$$E_{mn}^{(i)} \equiv \sum_{k=1}^{K} \left[p_{mnk} - \hat{p}_{mnk}^{(i)}\right]^2 \quad ; \quad i = 1,\cdots,9 \qquad (1)$$

Denote the smallest of the nine errors as E_{mn} with the corresponding proportions $\hat{p}_{mnk}$, k = 1,···,K.

Step 8 – If $E_{mn} \le 0.15$ (see appendix B for the determination of this threshold), call this mn*th* SSU correctly classified; otherwise, this mn*th* SSU is incorrectly classified.

Step 9 – Calculate PCC as

$$PCC = \frac{\text{Number of correctly classified SSU's}}{\text{Total number of SSU's}} \qquad (2)$$

In the present case, the denominator of equation (2) equals 10M.

Step 10 – Calculate the k*th* class proportion bias as

$$B_k = \left[\frac{1}{10M} \sum_{m=1}^{M} \sum_{n=1}^{10} \left(p_{mnk} - \hat{p}_{mnk}\right)^2\right]^{1/2} \qquad (3)$$

and the root-mean-square (rms) overall bias as

$$B_{rms} = \left(\frac{1}{K} \sum_{k=1}^{K} B_k^2\right)^{1/2} \qquad (4)$$

Step 11 – Using the PCC as calculated in step 9, calculate the confidence interval at specified confidence level (section 4.2). If the interval is satisfactory, stop;

otherwise, increase M and repeat steps 1 through 11 (refs. 1 and 7).

2.2 AN EXAMPLE OF THE AD HOC PROCEDURE

This subsection gives an example of the execution of steps 5 through 8 of the procedure. The other steps are self-explanatory.

Consider the generic *nth* SSU in the *mth* PSU, hereafter called the SSU, with the m and n indexes dropped, in figure 1. Part a of figure 1 shows the location of the SSU on the classification map, with an enlargement of the pixel assignment to class 1 or 2 in the 16 pixels containing the SSU. Part b of figure 1 shows the corresponding SSU on ground-truth photographs; the demarcation is interpreted to separate feature class 1 on the left from feature class 2 on the right.

Execution of step 5 of the procedure results in the estimates:

$$p_1 = 0.7 \quad ; \quad p_2 = 0.3 \tag{5}$$

Notice that the indexes m and n are dropped; indexes 1 and 2 denote the only two classes of interest.

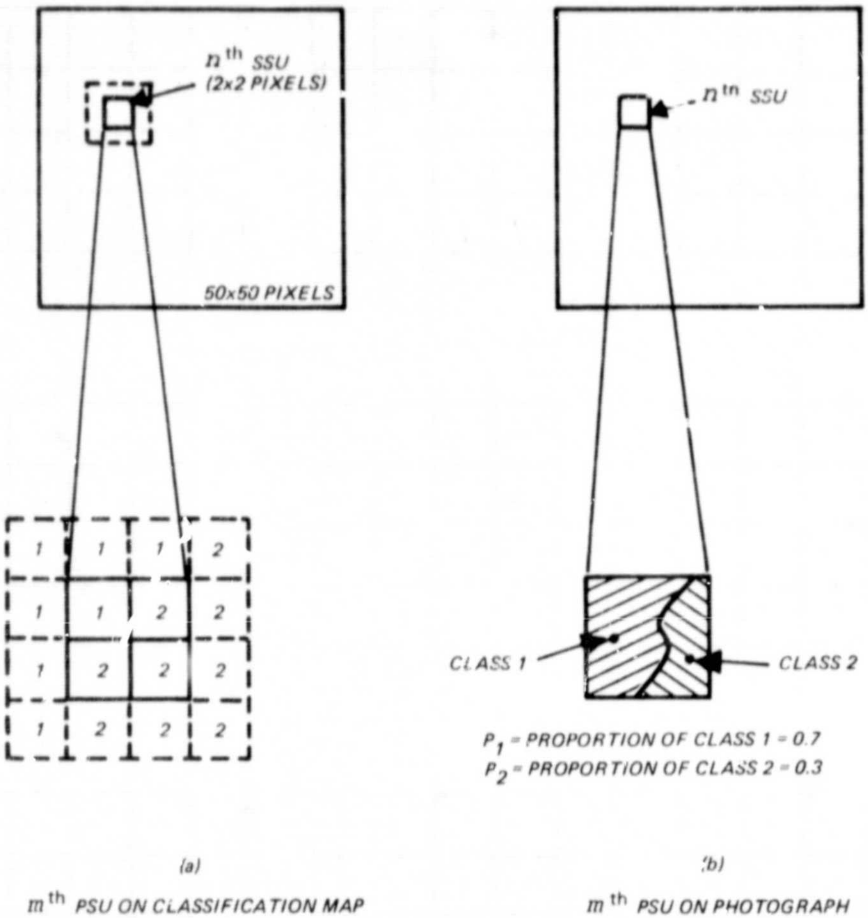
Step 6 of the procedure requires the examination of each of the nine possible locations on or about the SSU (fig. 2). Retaining only the indexes for class 1 or 2 and the superscript for the *ith* location, the proportions are estimated and tabulated in table I.

According to step 7 of the procedure and using the definition of $E_{mn}^{(i)}$ in equation (1), the errors for the nine possible locations are tabulated in the rightmost column of table I. The smallest of these errors is 0, occurring at location (b) or (d) with $\hat{p}_1 = 0.75$ and $\hat{p}_2 = 0.25$.

TABLE I.— TABULATION OF THE DIFFERENCES, E, BETWEEN THE NINE POSSIBLE SSU'S IN FIGURE 2 AND THE INTERPRETED SSU ON THE PHOTOGRAPH IN PART A OF FIGURE 1[a]

Location	$\hat{p}_1$	$\hat{p}_2$	$p_1 - \hat{p}_1$	$(p_1 - \hat{p}_1)^2$	$p_2 - \hat{p}_2$	$(p_2 - \hat{p}_2)^2$	$E = (p_1 - \hat{p}_1)^2 + (p_2 - \hat{p}_2)^2$
a	1.0	0	−0.3	0.09	0.3	0.09	0.18
b	.75	.25	−.05	0	.05	0	0
c	.25	.75	.45	.20	−.45	.20	.40
d	.75	.25	−.05	0	.05	0	.01
e	.25	.75	.45	.20	−.45	.20	.40
f	0	1.0	.7	.49	−.7	.49	.98
g	.5	.5	.2	.04	−.2	.04	.08
h	0	1.0	.7	.49	−.7	.49	.98
i	0	1.0	.7	.49	−.7	.49	.98

[a] p_1 = 0.7; p_2 = 0.3 (from part b of fig. 1).



(a) The *nth* SSU in *mth* PSU on classification map; enlarged view of pixel assignment to class 1 and 2 in SSU and surrounding pixels

(b) The same SSU in the same PSU on photograph; interpretation of SSU with assignment to class 1 or 2.

Figure 1.– An example of the procedure.

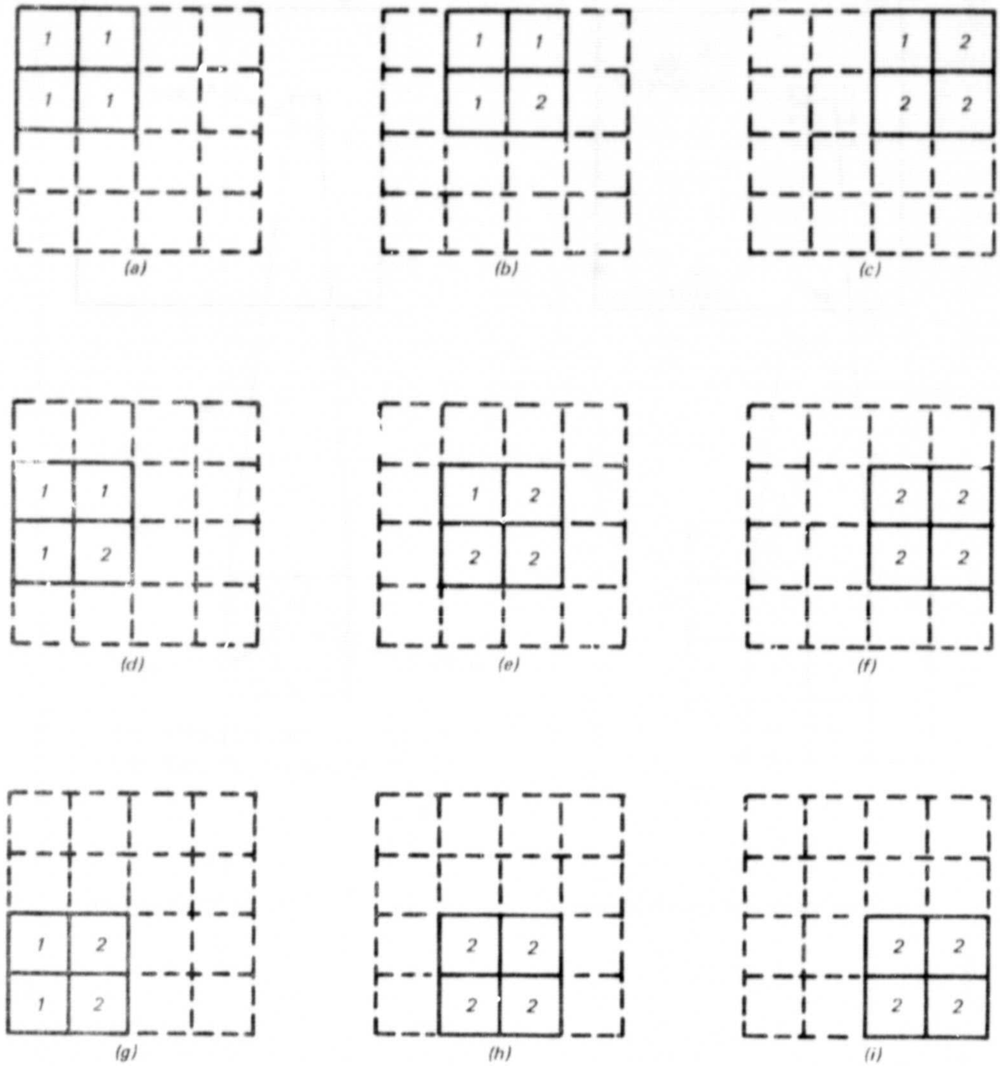


Figure 2.- The nine possible locations on or about the SSU of figure 1, with pixel assignment to class 1 or 2.

In step 8, E (E_{mn} of the n*th* SSU in the m*th* PSU) is checked against the threshold of 0.15. Since $E = 0.01 < 0.15$, this mn*th* SSU is considered to be correctly classified.

The proportion biases in this mn*th* SSU are

$$B_{mn1} = [(0.75 - 0.7)^2]^{1/2} = 0.05 \qquad (6)$$

and

$$B_{mn2} = [(0.25 - 0.3)^2]^{1/2} = 0.05 \qquad (7)$$

3. NOVELTY OF THE AD HOC PROCEDURE

As previously stated, the two main objectives of the proposed design are:

a. To minimize registration inaccuracies between a low resolution classification map and a high resolution ground-truth map; for example, a Landsat classification map versus interpreted aircraft photographs at scales of 1:60 000 and 1:120 000.

b. To minimize comparison inaccuracies, knowing that every sampling unit will likely contain more than one class of interest; in the present situation, a sample unit is a conglomerate of pixels.

The first objective is achieved by (1) having large (50 by 50 pixels) PSU's so that by identifying dominant landmarks within the PSU local registration error can be minimized to no more than 1 pixel within the PSU and (2) considering the nine SSU locations (SSU size is 2 by 2 pixels). on or about the designated location as candidates in the classification map, one of which is closest to perfect registration with the designated SSU on the ground-truth map.

The second objective is achieved by using error measures consisting of the difference in proportions of classes defined in equation (1) to determine the identity of classified SSU to ground-truth designated SSU. This comparison method is more discriminatory than using a majority rule to determine the unique SSU classification which in turn is used for comparison, as discussed in appendix B.

The present method uses the PCC concept to attempt to evaluate the per-pixel classification. This is possible because SSU's of sizes 2 by 2 pixels are small enough to reflect pixel classifications and yet large enough to absorb some possible errors caused by misregistration and mixture pixels. It is assumed that no extreme classification errors exist; for example, errors caused by consistently classifying forest as nonforest and vice versa, in which case the SSU comparison method by proportion differences will result in a totally erroneous PCC.

4. STATISTICAL QUALIFICATIONS

Three parameters are further discussed, the sample size M, the PCC and the corresponding confidence interval, and the proportion biases B.

4.1 SAMPLE SIZE M

Standard procedures (refs. 1, 2, and 7) provide suitable sample sizes M.[1] The basic assumption is that the PCC has a binomial distribution. The formulas are

$$M = 0.25(t/AE)^2 \tag{8}$$

or

$$M = PCC(1 - PCC)(t/AE)^2 \tag{9}$$

according to the availability of an estimated PCC at the beginning of the evaluation; t is the value in the table of t-distribution corresponding to a prespecified confidence level; AE is the allowable error, that is, permitted confidence interval half range.

4.2 CONFIDENCE INTERVALS OF PCC

Standard procedures (refs. 1, 2, and 7) also provide confidence intervals of PCC. The confidence interval half range is

$$t\sqrt{PCC(1 - PCC)/M} \tag{10}$$

where again t is the given value in a table of t-distribution which has been assigned a prespecified confidence level.

---

[1] In the ad hoc procedure, M PSU's with 10 SSU's in each PSU are proposed. Effectively, 10 M samples are used for evaluation. Thus, 10M should replace M in equations (8), (9), and (10).

4.3 PROPORTION BIASES AND IMPLICATIONS

The proportion biases B_k and the rms value B_{rms} in equations (3) and (4) are widely accepted evaluation parameters; however, they are not as easily understood as PCC. Whereas B_k and B_{rms} are appropriate measures for proportion accuracies, they are inadequate for measuring map accuracies although a high map accuracy (high PCC) is generally accompanied by small biases. Proportion biases for the entire map or for samples selected from the map are used frequently thereby avoiding the problem of registration inaccuracies. It is usually assumed that registration errors tend to cancel out with large samples.

To further understand the relationship between proportion biases and the PCC, figures 3 and 4 depict how biases B_k and B_{rms} vary with the PCC. [It can be shown that for a two-class case, as described, $B_1 = \left|\frac{(1 - 2q)}{q}\right|(1 - PCC)$; $B_2 = \left|\frac{(2q - 1)}{1 - q}\right|(1 - PCC)$; and $B_{rms} = \sqrt{\left(\frac{B_1^2 + B_2^2}{2}\right)}$. In these figures, a two-class map is analyzed wherein class 1 has *a priori* probability (i.e., proportion in map) q and class 2 has *a priori* probability 1 - q. Four curves are shown in each figure and labeled with the PCC value. For example, the case PCC = 0.9 assumes that the classification accuracies of classes 1 and 2 are both 0.9. (Thus, overall $PCC \equiv qp(1/1) + (1 - q)p(2/2) = p(1/1) = p(2/2)$; $p(i/i)$ denotes class i accuracy.)

As an example, when class 1 has *a priori* probability of 0.8 and PCC = 0.9, the theoretical biases can be derived from figures 3 and 4.

$$\left.\begin{aligned} B_1 &= 0.08 \\ B_2 &= 0.3 \\ B_{rms} &= 0.22 \end{aligned}\right\} \qquad (11)$$

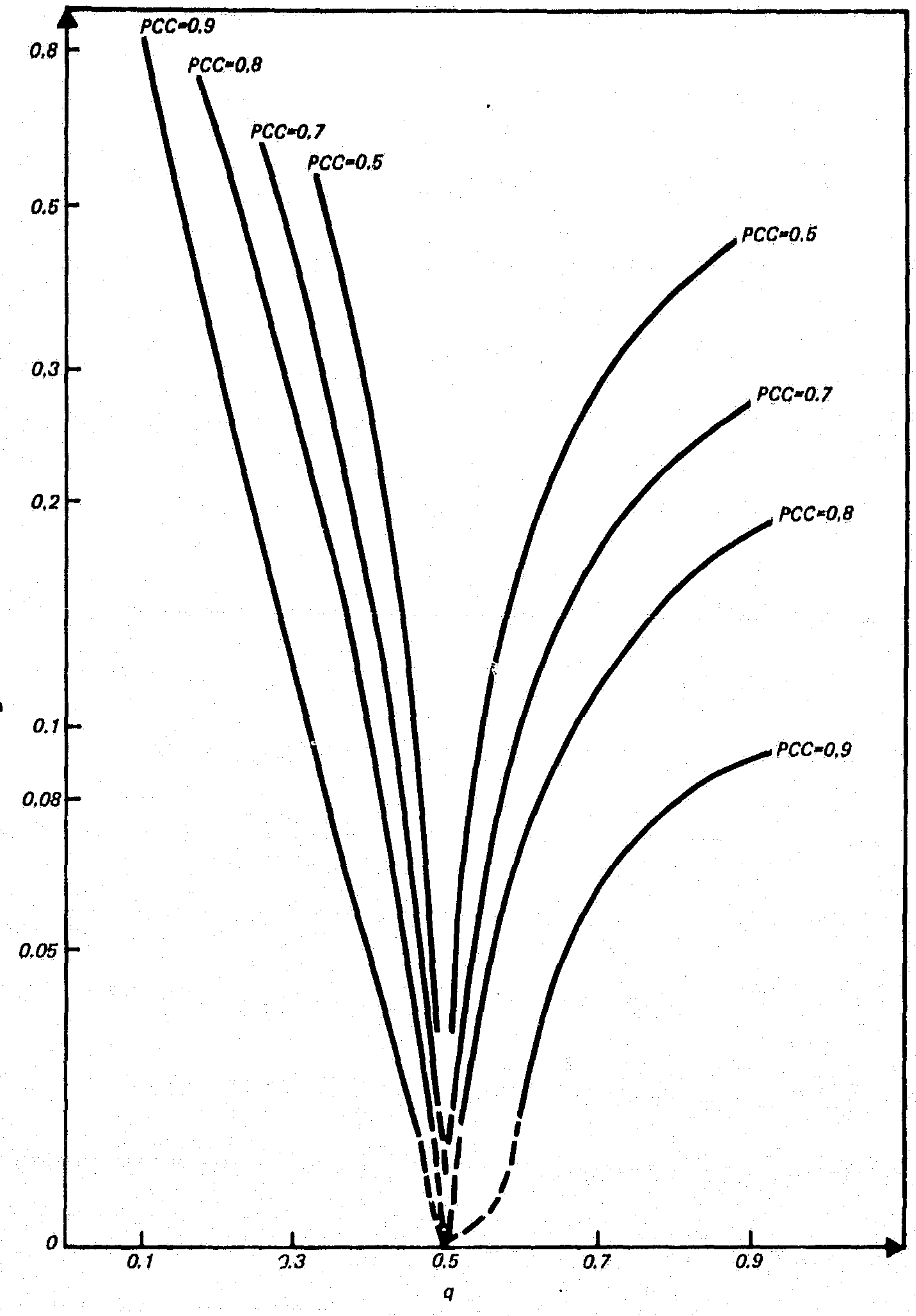


Figure 3.— Proportion bias B versus *a priori* probability q in a two-class map at various PCC's.

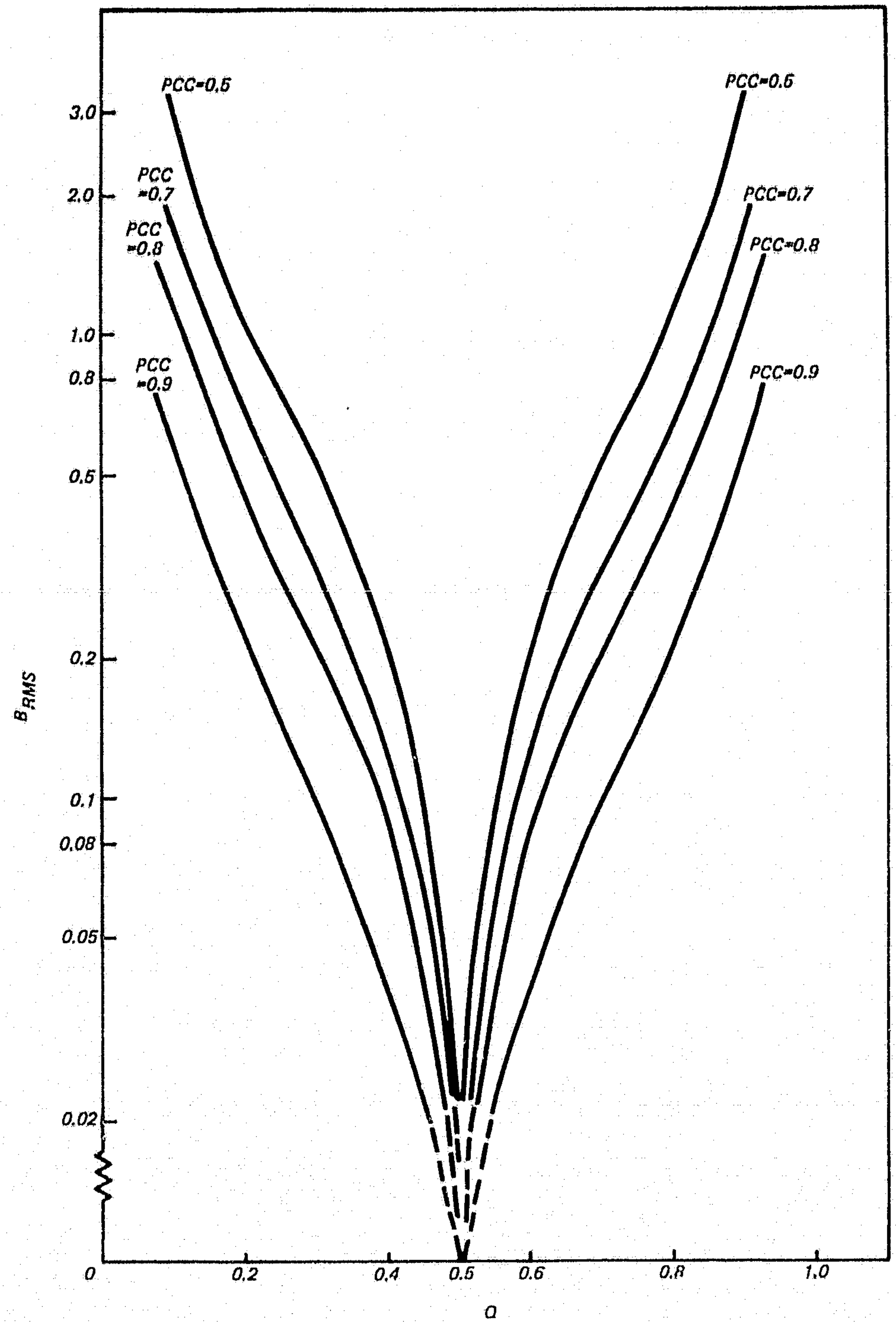


Figure 4.– Proportion rms bias B_{rms} versus *a priori* probability q in a two-class map at various PCC's.

When $q = 0.8$ and $PCC = 0.7$,

$$\left.\begin{aligned} B_1 &= 0.23 \\ B_2 &= 0.90 \\ B_{rms} &= 0.66 \end{aligned}\right\} \qquad (12)$$

It can be seen that the relationship between the biases and the PCC is very complicated even for the two-class map. No attempt has been made here to illustrate the more complicated multiple-class cases. The complex relationship between the biases and the PCC can be used to conclude that biases B_k and B_{rms} are inadequate measures for evaluating map accuracies, even though they are appropriate for measuring proportion accuracies.

Note: In the present ad hoc procedure, the proportion bias parameters are only secondary to and a byproduct of the calculation of the PCC measure; hence, their estimation is suboptimally designed in the present evaluation. The PCC is felt to be a more appropriate measure for map accuracy assessment, as discussed in the present context. However, proportion biases can be the prime evaluation parameter in investigations where the objective is areal measurement.

5. PRACTICAL CONSIDERATIONS

The main practicality factor and practicability consideration are the manipulation and registration of classification maps (Landsat maps) and ground-truth maps (interpreted aircraft photographs).

The procedure requires photographic enlargement of Landsat classification maps on positive prints or transparencies; composite classification maps are desired rather than single-class, theme prints as produced on the Gould printer of the General Electric Interactive Multispectral Image Analysis System (IMAGE 100). Aircraft photography on positive prints or transparencies should be enlarged on optical instruments such as the Zoom Transfer Scope or the Kargl reflecting projector/rectifier on which registration is performed. For easy data handling and for lesser geometric distortion, the 1:120 000 or 1:60 000 scale photographs are desired. Using such small scales, the amount of stretching on the optical instruments during registrations will be minimized or even eliminated. Table II summarizes these practical considerations.

TABLE II.— PRACTICAL CONSIDERATIONS WHEN USING THE NEW AD HOC PROCEDURE

Pixel size on classification map, meters	Approximate enlargement of classification map on film	Physical size of SSU (2 by 2 pixels)	1:120 000 scale photography		1:60 000 scale photography	
			Approximate magnification on optical equipment	FSU (50 by 50 pixels) on 9- by 9-inch frame, percent	Approximate magnification on optical equipment	FSU (50 by 50 pixels) on 9- by 9-inch frame, percent
60	90 pixels on 9-inch format (1:24 000)	(5 by 5 mm)	5X	4	2-1/2X	17
80	70 pixels on 9-inch format (1:24 000)	(6.7 by 6.7 mm)	5X	7	2-1/2X	30

6. CONCLUSION

As a result of previous experience (refs. 1, 2, 4, and 6), this ad hoc procedure was developed to evaluate low resolution classification maps (Landsat) against high resolution ground-truth maps (interpreted aircraft photographs). The data offered in this document support the value of this new design in (1) minimizing registration inaccuracies and (2) minimizing comparison inaccuracies. The ad hoc procedure provides fc the evaluation parameter of the PCC and the byproduct of proportion biases (B). The statistical implications of PCC and B are also discussed. It can be concluded that this ad hoc procedure is practical and practicable.

The proposed, novel design attempts to evaluate the per-pixel classification by estimating the PCC and by using a sufficiently small SSU size (2 by 2 pixels) to reflect pixel classifications and yet a sufficiently large SSU size to absorb some possible errors due to misregistration and mixture pixels. This ad hoc procedure is recommended for reevaluating the classification results from the TRICPS.

7. REFERENCES

1. Photointerpretation Guide for Forest Resource Inventories. (NASA SP to be published.)

2. Kan, E. P.: Multi-Class Map Accuracy Evaluation. Lockheed Electronics Co., Inc. (Houston, Texas). LEC-7936, Feb. 1976.

3. Phillips, T. L., ed.: LARSYS Version 3, Users' Manual. Laboratory for Applications of Remote Sensing, Purdue Univ. (W. Lafayette, Indiana), June 1973.

4. Weaver, J. E.; et al.: Soils Resources Inventory, Final Report for FY 76. (LEC Report to be published.)

5. Avery, T. E.: Forest Measurements. McGraw-Hill Book Co., 1967.

6. Reeves, C. A.; et al.: Tri-County Pilot Study (TRICPS), Final Report. (LEC Report to be published.)

7. Freese, F.: Elementary Statistical Methods for Foresters. USDA, Forest Service, Agriculture Handbook 317, Mar. 1974.

APPENDIX A

EVALUATION PROCEDURES USED IN TRICPS

The following is excerpted from the final report of the TRICPS. From the entire tricounty site, 100 plots were randomly selected, each plot being 10 by 10 pixels. These plots were located on the output computer classification maps which were printed by the Gould printer, reproducing one feature (versus the remaining) in any one print.

Each of the 10- by 10-pixel plots was subdivided into nine equal samples. Thus, each sample is roughly 3 by 3 pixels. The proportions of each class (softwood, hardwood, mixed, range, and "others") were counted in each of the 900 (100 × 9) samples. Each sample was then classified into its major class.

Using the Kargl reflecting projector/rectifier, the 100 randomly selected plots (900 samples) were located on the available photography at a scale of 1:120 000. The photography was interpreted and used as ground truth. Each sample was interpreted and classified by its major type.

The majority-rule classification of the computer-mapped sample was then compared to the majority-rule classification of the ground-truth sample. A calculation of accuracy followed, similar to the method stated in the text.

Notes

1. Local registration error of plots was high. Reliable registration of a 10- by 10-pixel plot on the photograph was extremely difficult because

(a) The geographic area covered by a plot was too small. Major landmarks in or surrounding the plot could not be used effectively to produce good registration.

(b) Gould printer output maps contained one class versus the remainder; using them to register to photographs was very difficult because the one-class maps (actually two classes, the one being considered and the remainder) contained insufficient details to be used for effective registration.

2. Sample classification by majority rule was shaky. The classification of an approximate 3- by 3-pixel sample by its majority type was too gross because the area covered by a sample [approximately 4.45 hectares (11 acres)] was too large and normally contained two or more types. Features like hardwood were normally narrower than the 3- by 3-pixel sample width and, therefore, tended to become misclassified when a small error existed in the registration. Furthermore, the sample comparison did not take advantage of the per-pixel classification (see section 3).

APPENDIX B

ERROR THRESHOLD FOR DECIDING CORRECT SSU CLASSIFICATION

B.1 RECAPITULATION OF THRESHOLDING PROCEDURE

For each of the nine possible locations on or about the designated SSU on classification map (see fig. 2 in text), the error measure $E^{(i)}$, $i = 1,\cdots,9$, is calculated as defined by

$$E^{(i)} \equiv \sum_{k=1}^{K} \left[p_k - \hat{p}_k^{(i)}\right]^2 \qquad \text{(B1)}$$

where p_k and $\hat{p}_k^{(i)}$ denote the proportions of the *kth* class among K classes in the ground-truth SSU and the *ith* location of the classification map SSU, respectively. (The indexes m and n corresponding to the PSU and SSU locations are dropped here.) The smallest of these $E^{(i)}$, $i = 1,\cdots,9$, is designated E; E is checked against the recommended threshold 0.15; $E > 0.15$ means that the mapped SSU is incorrectly classified and $E \leq 0.15$ means that the SSU is correctly classified.

B.2 ESTABLISHMENT OF THRESHOLD VALUE 0.15

Before determining the threshold, it is recognized that even if there were no classification error, the residual local registration error and the quantization of p_k (in the 25 percentile) will still give rise to nonzero E. Thus, a minimal amount of E must be allowed, beyond which the SSU is considered to be incorrectly classified. This residual error is illustrated by figures B1 and B2. Figure B1 is a ground-truth SSU containing classes 1 and 2 with $p_1 = 0.7$ and $p_2 = 0.3$. Assuming that a 1-pixel registration error exists between the classification-map SSU and the ground-truth SSU, the nine cases in figure B2 are possible configurations for the better SSU locations on the classification map; "better" means displaced by no more than half a pixel. The minimum $E^{(i)}$ will be equal to or less than the closest configuration from

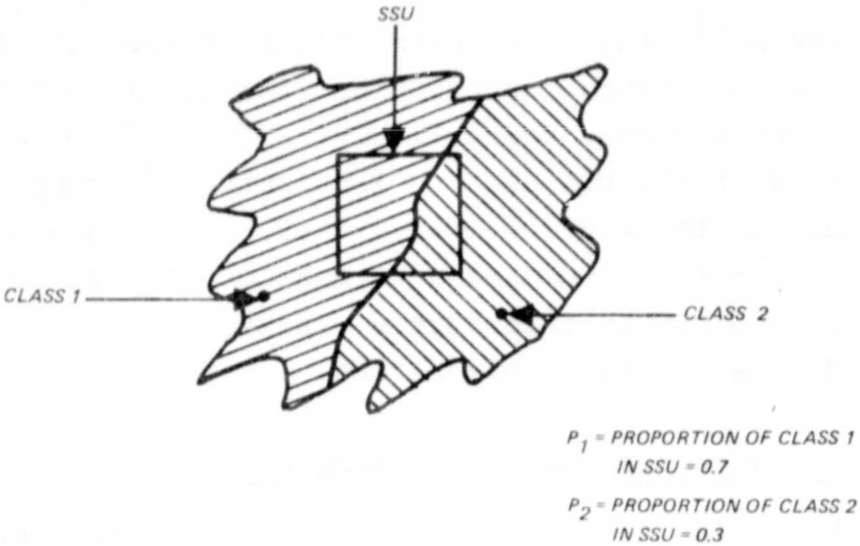


Figure B1.— Portion of ground-truth map showing location of SSU and its interpreted partition into classes 1 and 2.

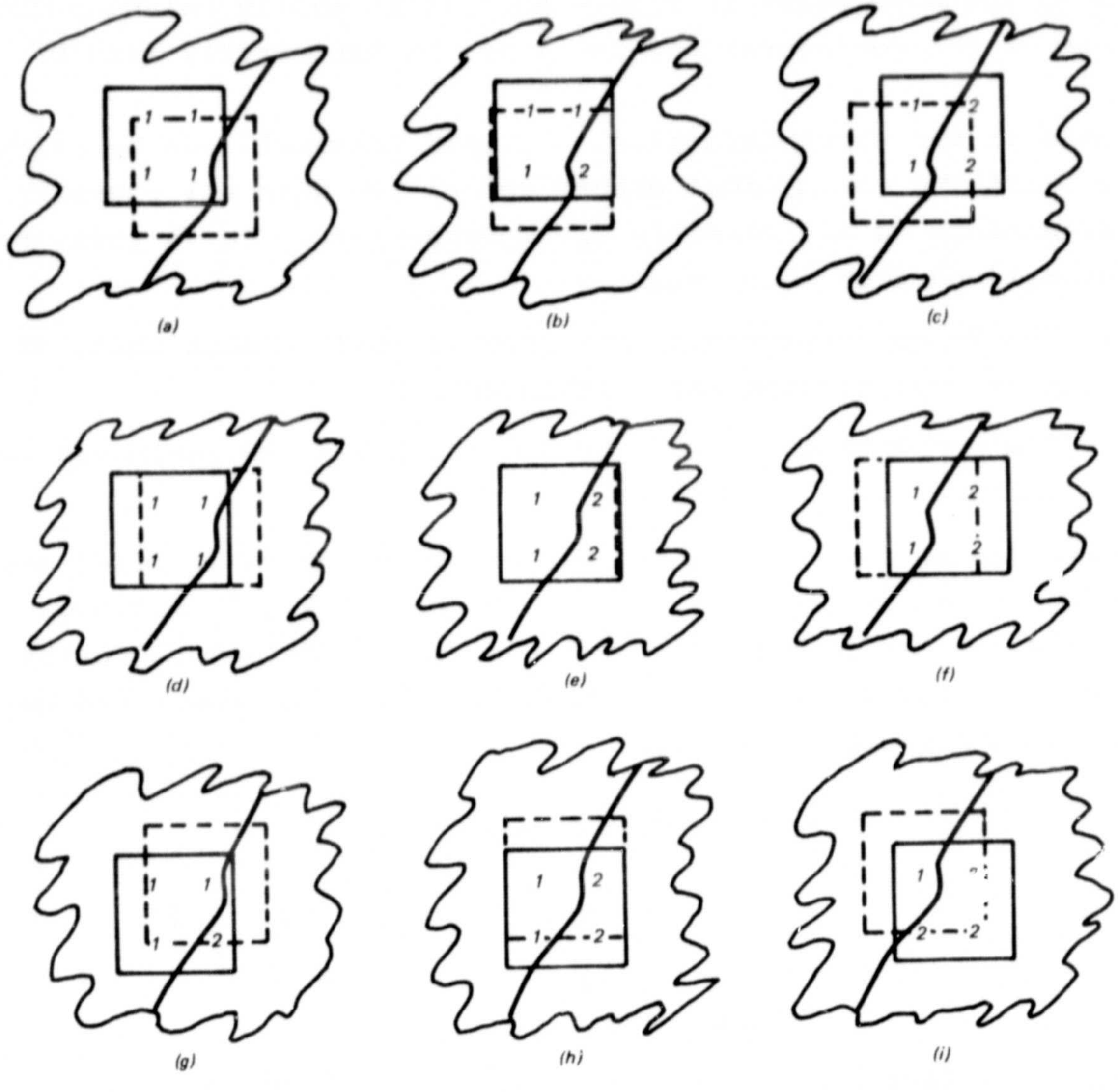


Figure B2.— Nine extreme cases, each displaced by half a pixel about the ground-truth SSU; pixel classification of the four pixels in SSU is also indicated.

among the nine cases in figure B2. It is easily calculated that none of the configurations in figure B2 has exactly zero E.

Based on the above deduction, a proper threshold can be calculated by examining the minimum errors associated with all possible pixel assignments to all possible SSU compositions. To do this, two assumptions need to be made:

a. The 2- by 2-pixel size of a SSU is small enough that, at most, two classes fall within an SSU.

b. The proportions p_k in a ground-truth SSU are estimated in increments of 10 percent.

Using assumption a, there are only five possible pixel assignment configurations; i.e., $\hat{p}_1 = 1.0$, $\hat{p}_1 = 0.75$, $\hat{p}_1 = 0.50$, $\hat{p}_1 = 0.25$, and $\hat{p}_1 = 0.0$ ($\hat{p}_2 = 1 - \hat{p}_1$ and needs not be further stated). For any assignment with $\hat{p}_1$, the error committed when the ground truth has p_1 will be

$$E = \left(p_1 - \hat{p}_1\right)^2 + \left(p_2 - \hat{p}_2\right)^2$$

$$= \left(p_1 - \hat{p}_1\right)^2 + \left[\left(1 - p_1\right) - \left(1 - \hat{p}_1\right)\right]^2$$

$$= 2\left(p_1 - \hat{p}_1\right)^2 \tag{B2}$$

Take the example of $\hat{p}_1 = 0.75$, E will be a function of p_1:

$$E = 2(p_1 - 0.75)^2 \tag{B3}$$

Figure B3 plots all the five cases; i.e., $\hat{p}_1 = 0.0, 0.25, 0.50, 0.75$, and 1.0.

Using figure B3 and assumption b, the following table can be prepared (the symmetry of the table for p_1 and below 0.5 makes it necessary to analyze only half of the full range of p_1).

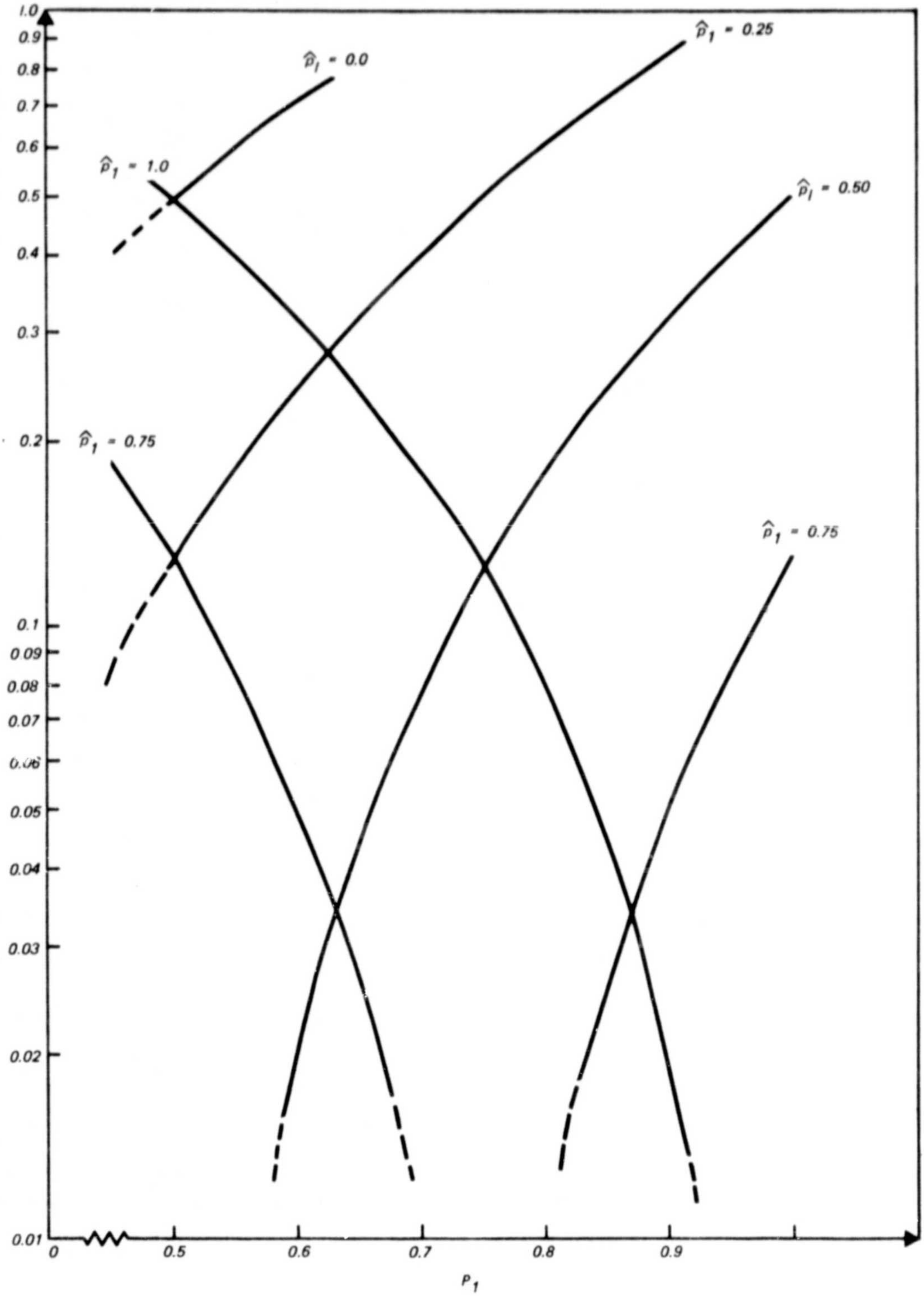


Figure B3.— Error E between ground-truth SSU configuration having p_1 and SSU pixel assignment having $\hat{p}_1$.

p_1	Those p_1 such that $E < 0.15$	Maximum proportion error, max $\lvert p_1 - \hat{p}_1 \rvert$
1.0	0.75, 1.0	0.25
.9	.75, 1.0	.15
.8	.75, 1.0	.2
.7	.5, .75	.2
.6	.5, .75	.15
.5	.25, .50, .75	.25

When a threshold of 0.15 is used, a maximum deviation of 0.15 to 0.25 in the proportion estimate $\hat{p}_1$ from the true p_1 will be tolerated, beyond which the SSU is considered incorrectly classified.

Notice that this rule is more discriminatory than the majority rule. When p_1 is 1.0, 0.9, or 0.8, a pixel configuration having $\hat{p}_1 = 0.5$ is considered correct classification by the majority rule but not by the present error threshold using 0.15. Similarly, for $p_1 = 0.7$, 0.6, or 0.5, the pixel configuration with $\hat{p}_1 = 1.0$ is considered correct classification by the majority rule but not by the threshold using 0.15. Finally, for $p_1 = 0.5$, the pixel configuration with $p_1 = 0.25$ is considered incorrect classification by the majority rule but correct classification using the 0.15 threshold. A correct classification is more plausible than incorrect classification in this case.

By experimenting with threshold values other than 0.15, similar conclusions can be drawn on the utility of the error-thresholding method. The value of 0.15 for the threshold, however, seems to allow sufficient tolerance in the proportion error between 0.15 to 0.25 without being overly lenient. An empirical study of a few cases of a three-class SSU classification produced the same conclusion.