LARS/ENTM 216

Short Course on

Advanced Topics in the Analysis of Remote Sensing Data

Purdue University
Laboratory for Applications
of Remote Sensing (LARS)
W. Lafayette, Indiana 47907 USA

ADVANCED TOPICS IN THE
ANALYSIS OF REMOTE SENSING DATA

1978

Assignment for Monday evening, April 10: <u>Multitemporal Data Analysis</u>

The availability of multipass data over a given ground scene enhances the possibility of using time-track information to assist in characterizing the various ground cover classes in the scene or to monitor changes with time in the scene. Techniques for precision registration of the data have been developed, and so the possibility of multitemporal analysis has become reality. Tomorrow we shall look in detail at a number of methods for multitemporal data analysis.

To prepare for these sessions, take some time to consider a specific remote sensing application with which you have been personally involved or are familiar. How might multitemporal analysis be used to enhance the results achievable for this application? Jot down some notes and be prepared to discuss this application in class, including:

--Definition of the remote sensing application (products needed, data available, etc.).

--Time-varying aspects of the scene. Are these likely to help in characterizing the scene?

--How would you go about incorporating the time variability in the analysis? Would additional reference data ("ground truth") be required?

--How feasible do you think it would be to apply multitemporal analysis in this context? Why? What problems would have to be overcome to use multitemporal analysis successfully? Would the benefits be likely to justify the cost?

--In this application, is "change detection" likely to be of use? What approaches can you think of for extracting change information from multitemporal/multispectral data?

In LARS Technical Reports Used
091577

TUESDAY

METHODS FOR MULTITEMPORAL ANALYSIS

Philip H. Swain

I. Introduction

II. Bayesian Formulation for the Multitemporal Problem

III. Layered Classifiers LARS ~~Info~~ Technical Report 072277

IV. Change Detection 030178

ADVANCED TOPICS IN THE
ANALYSIS OF REMOTE SENSING DATA

Assignment for Tuesday evening, April 11:

Multitemporal Analysis - Review

Look back over the notes you made in conjunction with last evening's assignment. In light of what you learned today, would you change your response? Be specific, even to the point of making a new set of notes.

Some questions you may now have concerning multitemporal analysis might not be answerable until you are faced with a real opportunity to make use of it. Others may be worth dealing with immediately. Make a list of such questions and, if you wish, raise them when Dr. Swain rejoins the group on Wednesday afternoon. Or surprise Dr. Landgrebe with them in the morning!

Information from the Spatial Domain

Since presumably you meet the prerequisite requirements of this Short Course, you have viewed a considerable amount of multispectral remote sensing imagery. If, in particular, you have seen some examples of color-composite LANDSAT imagery, you are probably aware that with your eyes you can readily identify classes of objects in the scene that probably cannot be identified by any pattern classifier that "looks" at only one pixel at a time. Examples include airports, bridges, superhighways, lakes, geological fault lines, etc. One thing which all of these classes have in common is that they are characterized at least as much by their spatial characteristics and their spatial relationships to other things in the scene as they are by their spectral properties.

Extend the list of spatially characterized classes begun above by at least six. Write down a brief description of the spatial property which characterizes each class. For instance, your list of spatial properties might include texture, shape, orientation relative to another object, proximity to another object, homogeneity, etc. Think about how a computer program might embody each of these spatial characteristics. You'll find it's a hard job if you really try to be specific!

As you did last evening, focus your attention on a specific remote sensing application with which you are familiar. Make some written notes concerning how spatial information in the scene might be relevant to this application and try to think of ways to "capture" that information by means of a computer algorithm. In the course of our discussions tomorrow, we'll make time to discuss some of these applications.

METHODS FOR

MULTITEMPORAL ANALYSIS

OF

MULTISPECTRAL REMOTE SENSING DATA

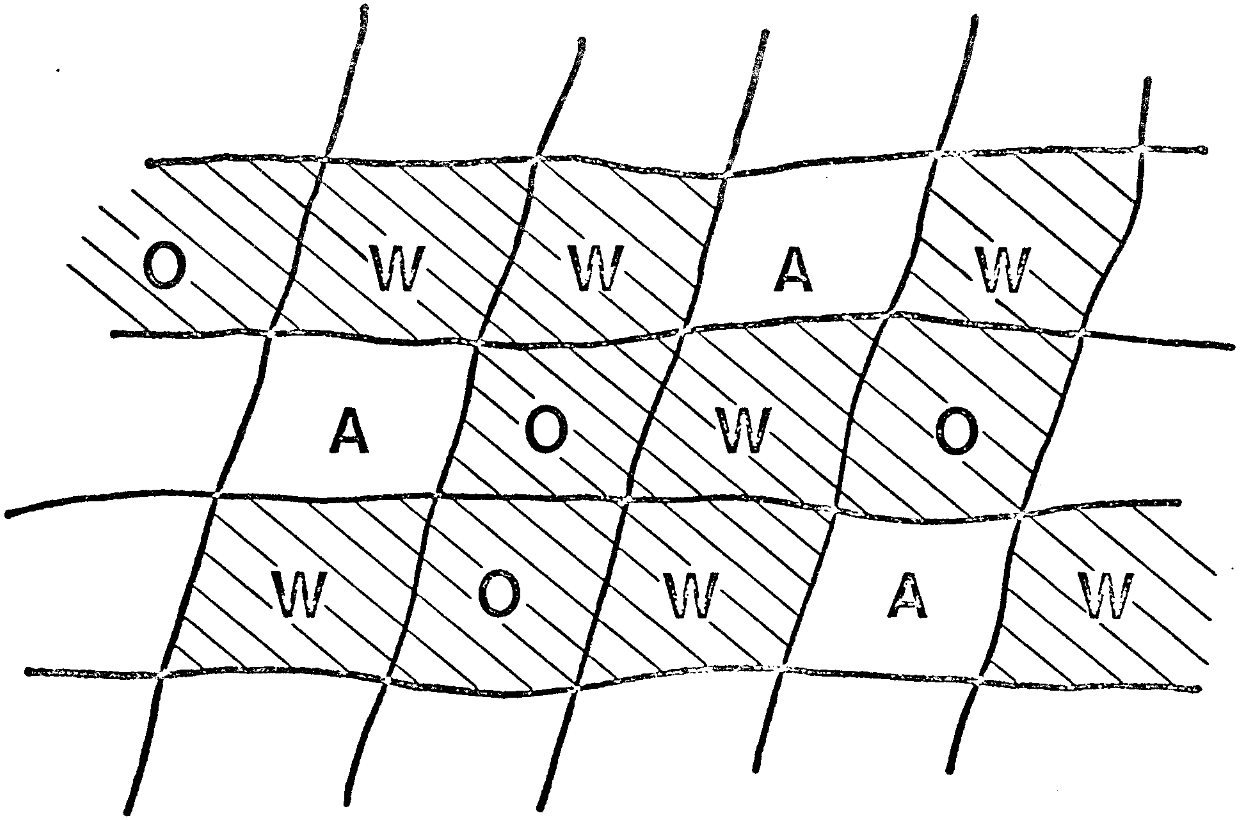
PRINCIPAL USES OF MULTITEMPORAL DATA

-- ENHANCE DISCRIMINABILITY OF INFORMATION CLASSES

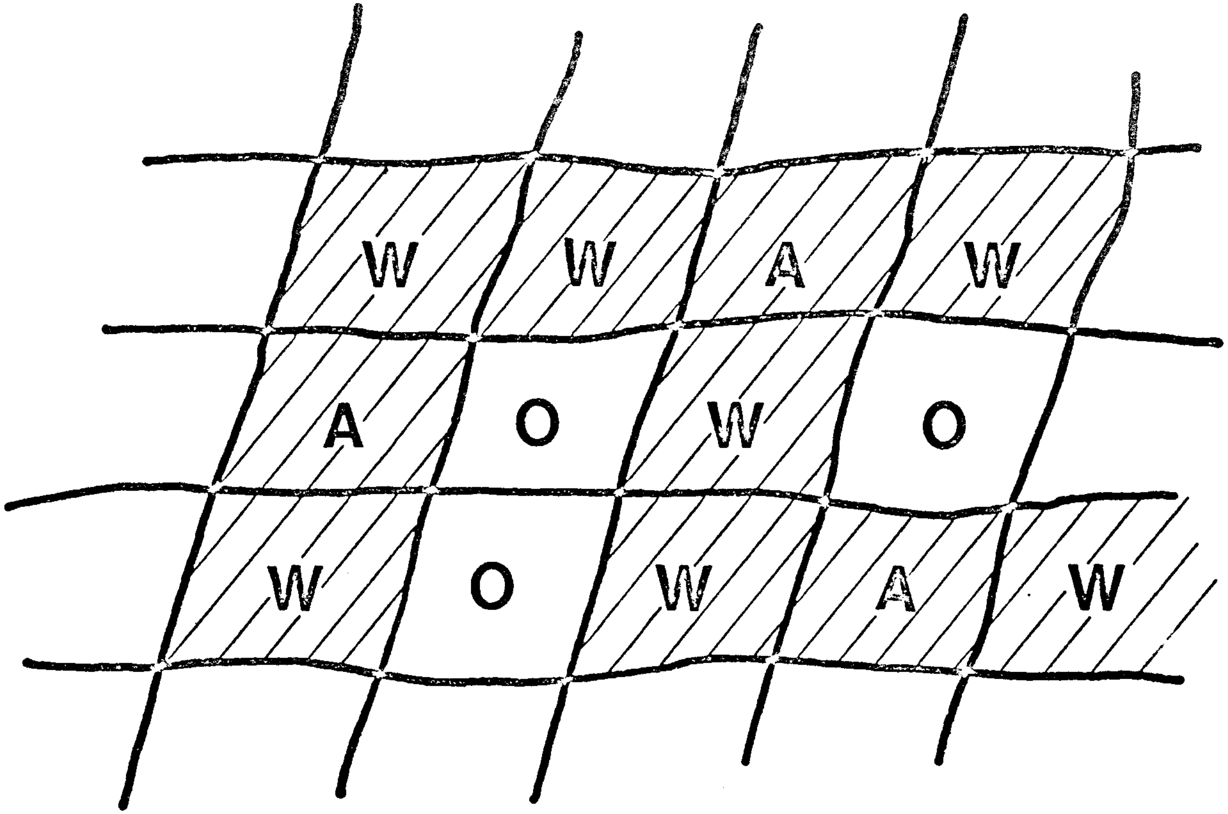
-- MONITOR CHANGE

Tu

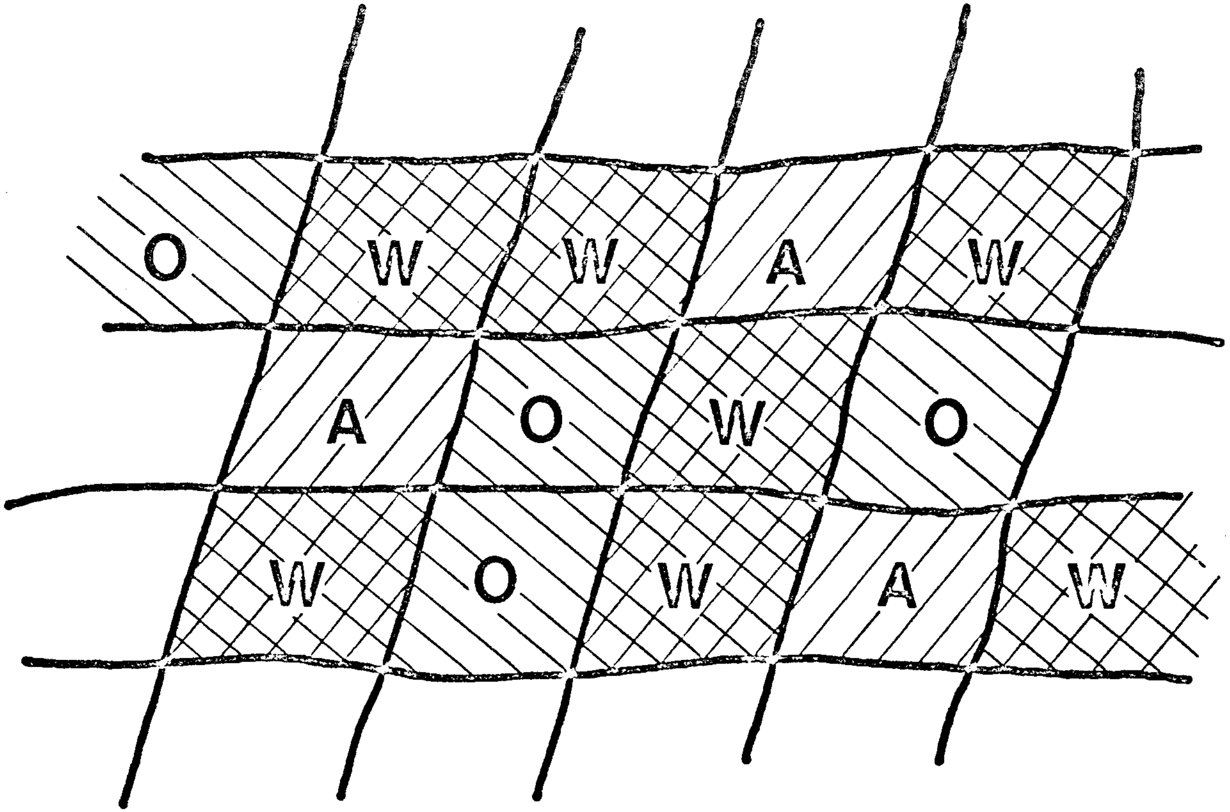
Tu-3



OPTIMAL PASS FOR WHEAT VS ALFALFA



OPTIMAL PASS FOR WHEAT VS OTHER



MULTITEMPORAL RESULT

MULTITEMPORAL ANALYSIS: APPROACHES

-- CONCATENATED ("STACKED")
VECTORS

-- MULTISTAGE CLASSIFIERS
  THE CASCADE CLASSIFIER
  THE DECISION TREE CLASSIFIER

-- CHANGE DETECTION

CONCATENATED VECTORS

$$\underline{X} = \begin{bmatrix} X(t_1) \\ X(t_2) \end{bmatrix} = \begin{bmatrix} x_{11} \\ x_{12} \\ \vdots \\ x_{1n_1} \\ x_{21} \\ x_{22} \\ \vdots \\ x_{2n_2} \end{bmatrix}$$

Mean vector & covariance matrix –

$$\underline{U} = \begin{bmatrix} U(t_1) \\ U(t_2) \end{bmatrix} \; ; \; \sum = \left[\begin{array}{c:c} \Sigma(t_1) & \begin{array}{c}\text{inter-} \\ \text{temporal} \\ \text{covariances}\end{array} \\ \hdashline \begin{array}{c}\text{inter-} \\ \text{temporal} \\ \text{covariances}\end{array} & \Sigma(t_2) \end{array}\right]$$

CONCATENATED VECTORS

ADVANTAGES:

- EXTEND MULTISPECTRAL ANALYSIS TECHNIQUES
- INTERTEMPORAL COVARIANCES MAY CONTAIN USEFUL INFORMATION

DISADVANTAGES:

- DIMENSIONALITY
- MULTIPLICITY OF CLASSES
- DEMANDS ON TRAINING DATA

THE CASCADE CLASSIFIER

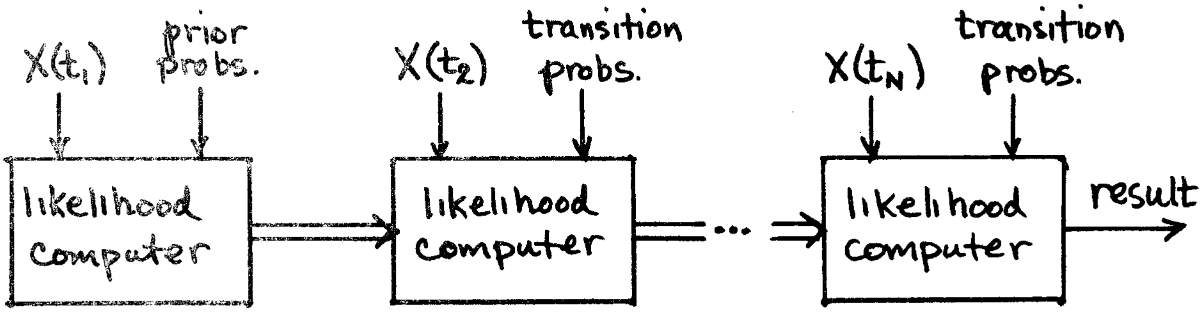
A BAYESIAN MODEL

-- ACCOMPLISHES TEMPORAL "DECOUPLING"

-- STATISTICALLY OPTIMAL

-- VERSATILE (DATA TIMES AND TYPES)

CASCADE CLASSIFIER



Bayes decision theory revisited –

Expected loss: $L_X(i) = \sum_{j=1}^{m} \lambda_{ij}\, p(\omega_j | X)$

For "0-1 loss function":

$$L_X(i) = \sum_{\substack{j=1 \\ i \neq j}}^{m} p(\omega_j | X) \quad = 1 - p(\omega_i | X)$$

Discriminant function: $p(\omega_i | X)$

or equivalently: $p(X | \omega_i)\, p(\omega_i)$

("maximum likelihood")

Bitemporal generalization –

Expected loss: $L_{X_1 X_2}(i) = \sum_{j=1}^{m} \lambda_{ij}\, p(\omega_{2j} | X_1 X_2)$

$$[\omega_{2j} = \omega_j(t_2)]$$

Maximum likelihood strategy: choose class ω_2

to maximize $p(\omega_2 | X_1, X_2)$

or, equivalently,

$$\sum_{\omega_1} p(X_1, X_2 | \omega_1, \omega_2)\, p(\omega_2 | \omega_1)\, p(\omega_1)$$

$$p(\omega_2 | X_1, X_2) = p(\omega_2, X_1, X_2) / p(X_1, X_2)$$

$$= \frac{1}{p(X_1, X_2)} \sum_{\omega_1} p(X_1, X_2, \omega_1, \omega_2)$$

$\llcorner$ "constant"

$$p(X_1, X_2, \omega_1, \omega_2) = p(X_1, X_2 | \omega_1, \omega_2)\, p(\omega_1, \omega_2)$$

$$= p(X_1, X_2 | \omega_1, \omega_2)\, p(\omega_2 | \omega_1)\, p(\omega_1)$$

Simplifying assumptions/approximations –

1. Class-conditional independence

$$p(X_1, X_2 | \omega_1, \omega_2) = p(X_1 | \omega_1, \omega_2)\, p(X_2 | \omega_1, \omega_2)$$

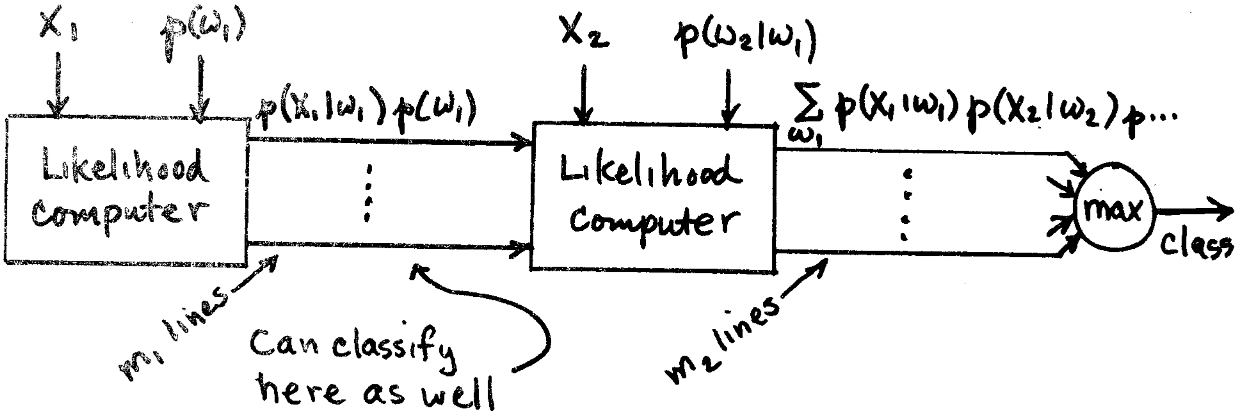
2. Sufficiency of the present class

$$p(X_1 | \omega_1, \omega_2) = p(X_1 | \omega_1)$$

$$p(X_2 | \omega_1, \omega_2) = p(X_2 | \omega_2)$$

Then select ω_2 to maximize:

$$\sum_{\omega_1} p(X_1 | \omega_1)\, p(X_2 | \omega_2)\, p(\omega_2 | \omega_1)\, p(\omega_1)$$



Some results –

Fayette County, Ill. – LANDSAT data

Corn, Soybeans, Woods, 'Other'

t_1 (6/9/73)	68 % correct	(all bands)
t_2 (7/17/73)	72%	(")
Cascade*	84 %	(")

*assumed equal priors

transition probabilities:

$$p(\omega_{2i} | \omega_{1i}) = 0.8$$

$$\sum_{i \neq j} p(\omega_{2i} | \omega_{2j}) = 0.2$$

THE DECISION TREE CLASSIFIER
(LAYERED CLASSIFIER)

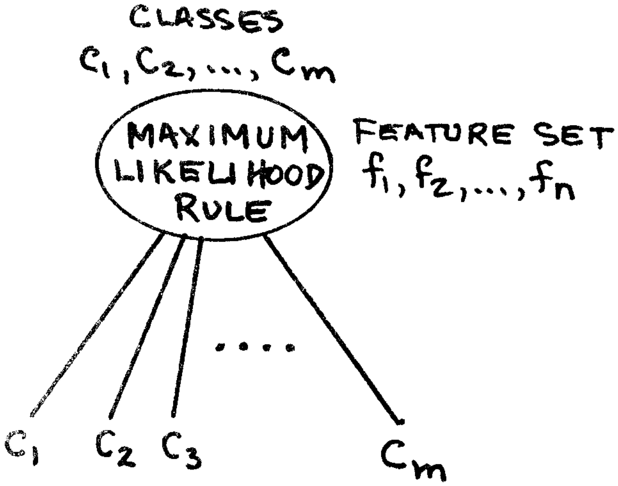
A HIERARCHICAL MODEL

-- PROVIDES GREATEST FLEXIBILITY

-- DESIGNED FOR ACCURACY AND EFFICIENCY

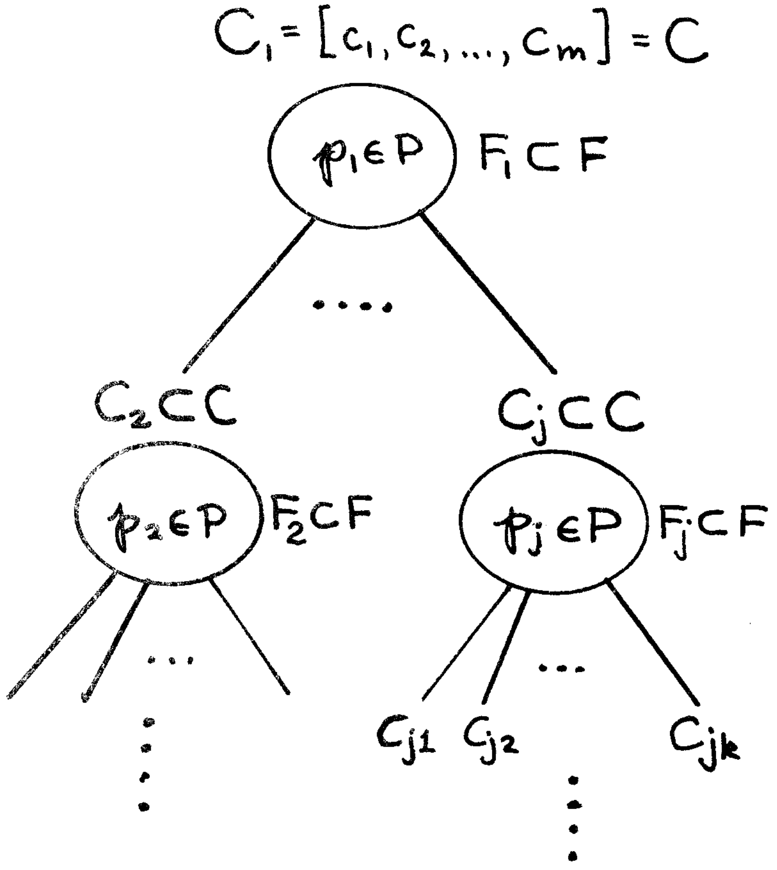
-- ACCOMPLISHES TEMPORAL "DECOUPLING"

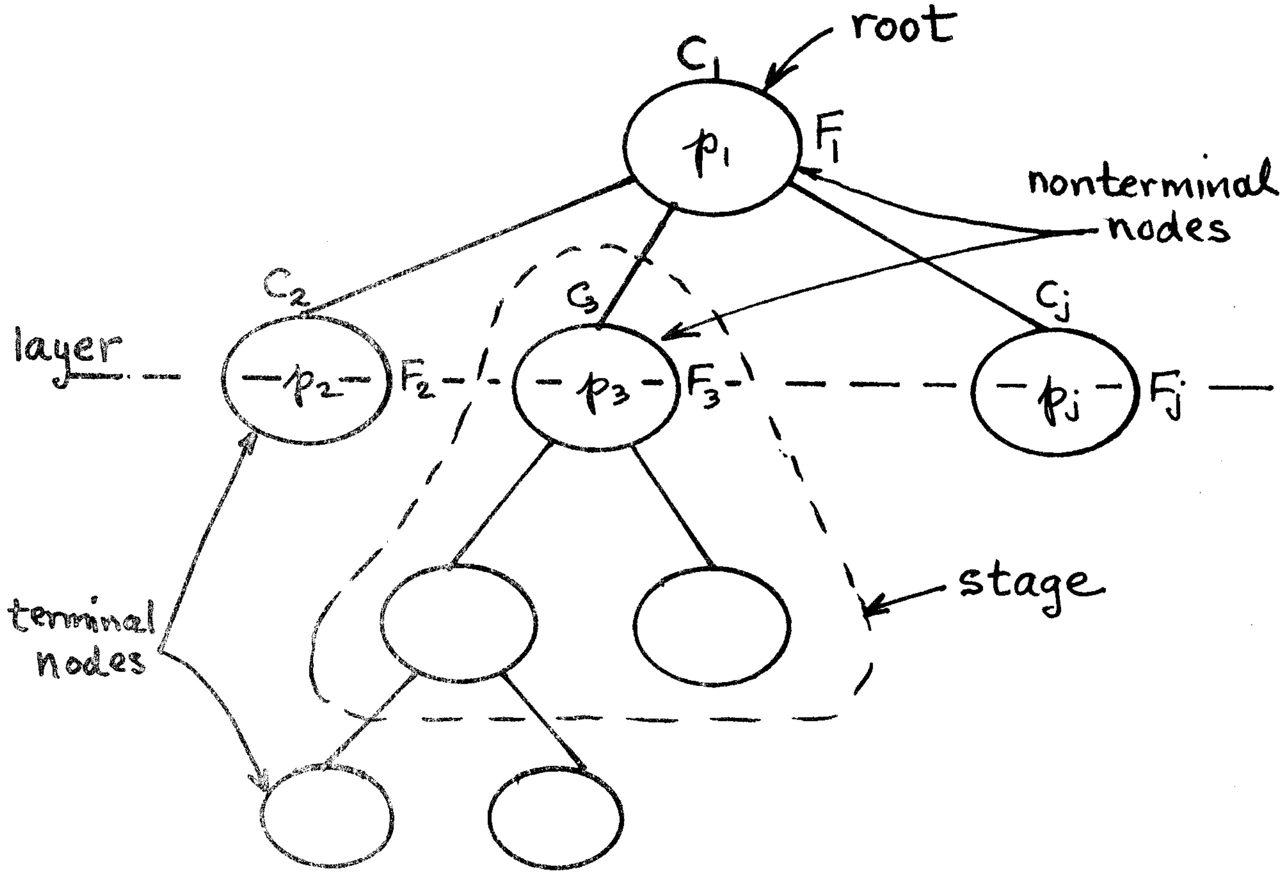
CONVENTIONAL CLASSIFIER



DECISION TREE CLASSIFIER

GENERAL MODEL



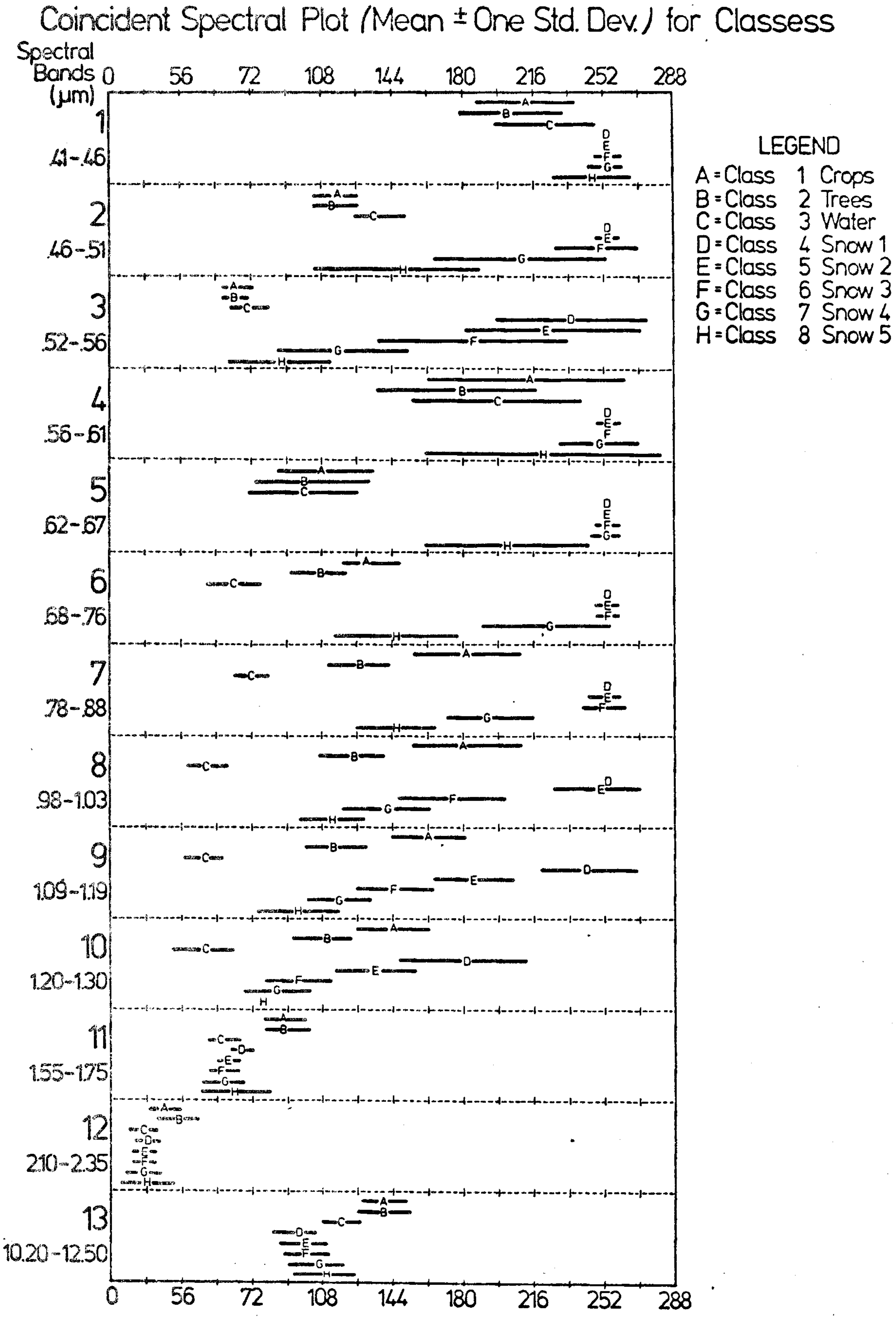

root
C1
p1
F1
nonterminal nodes
C2
C3
Cj
layer
p2
F2
p3
F3
pj
Fj
stage
terminal nodes

ADVANTAGES

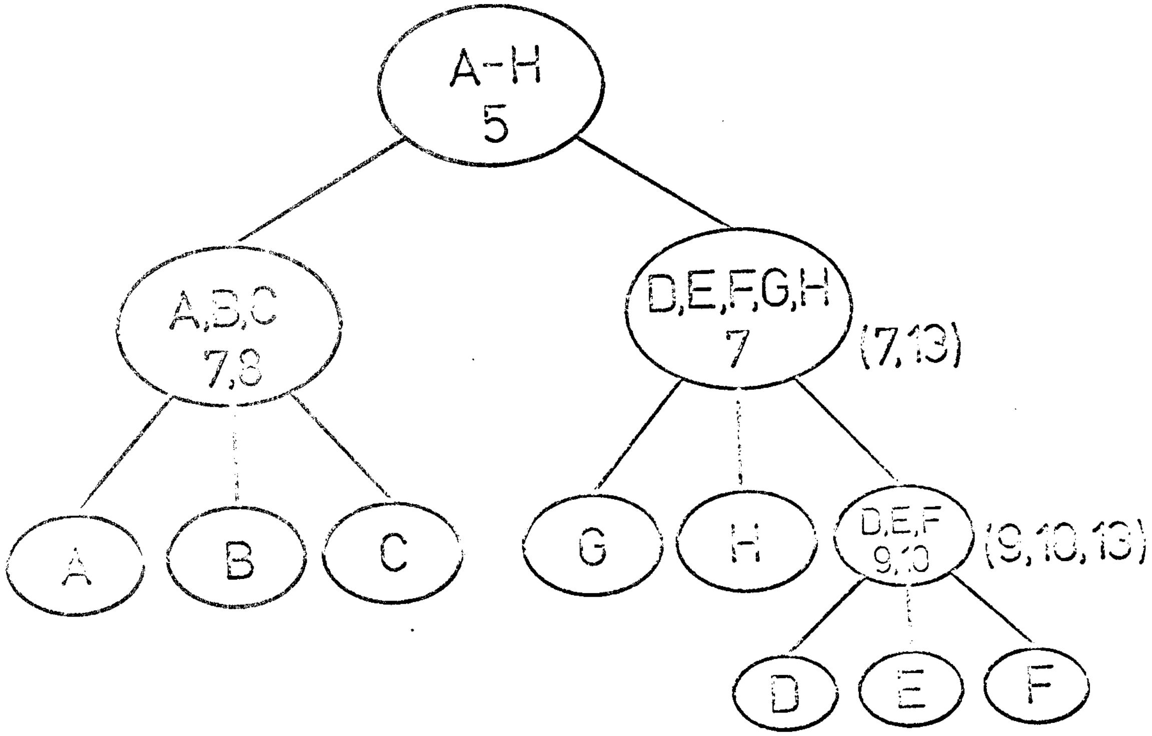
- MAY USE DIFFERENT FEATURE SUBSETS AT EACH NODE
- MAY USE DIFFERENT DECISION RULES AT EACH NODE

DISADVANTAGES

- LOGICAL OVERHEAD, BUT ....
- COMPLEX DESIGN PROBLEM, BUT ....



Coincident spectral plot for San Juan Mountains example.



Manually designed decision tree for the San Juan Mountains example (parentheses indicate final feature selection).

EXAMPLE: LANDSAT DATA, GRANT COUNTY, KANSAS

CLASSES: WHEAT, CORN, ALFALFA, FALLOW, PASTURE

MAY 9, '74	31 SPECTRAL CLASSES	62 %
JUNE 14, '74	29 SPECTRAL CLASSES	55 %

MULTITEMPORAL DECISION TREE

- M.L. NODES
- UNITEMPORAL NODES
- MANUAL DESIGN
- TWO LAYERS 71 %

RATIONALE FOR AUTOMATED APPROACH

- BETTER USE OF MULTIVARIATE INFORMATION
- NEED TO DEAL WITH COMPLEX CASES
- DESIRE FOR ANALYTICAL, REPEATABLE PROCESS
- SEARCH FOR OPTIMALITY

REQUIREMENTS (CONFLICTING)

- MAXIMIZE CLASSIFIER ACCURACY
- MINIMIZE CLASSIFICATION TIME

CANDIDATE APPROACHES

- DEVELOP VERY SOPHISTICATED SEARCH PROCEDURE
- RESTRICT THE FORM OF THE DECISION TREE

Tu

EVALUATION FUNCTION

$$E(d_i) = -T(d_i) - K \cdot \epsilon(d_i) + \sum_{j=1}^{c_i} P_{i+j} \, E(d_{i+j})$$

$T(d_i)$ = node computation time

$\epsilon(d_i)$ = node classification error

$E(d_{i+j})$ = evaluation of jth descendant of d_i

c_i = number of descendants of d_i

P_{i+j} = probability that jth descendant of d_i will be reached from d_i

K = trade-off constant, speed vs. accuracy

Lower bound (conventional stage):

$$E_0(d_i) = -T_0(d_i) - K \cdot \epsilon_0(d_i)$$

Normalized evaluation function:

$$
\begin{aligned}
E'(d_i) &= E(d_i) - E_0(d_i) \\
&= [T_0(d_i) - T(d_i)] \\
&\quad + K \cdot [\epsilon_0(d_i) - \epsilon(d_i)] \\
&\quad + \sum_{j=1}^{c_i} P_{i+j} E(d_{i+j})
\end{aligned}
$$

Normalized and bounded:

$$
\begin{aligned}
E''(d_i) &= [T_0(d_i) - T(d_i)] \\
&\quad + K \cdot [\epsilon_0(d_i) - \epsilon(d_i)] \\
&\quad + \sum_{j=1}^{c_i} P_{i+j} E_0(d_{i+j})
\end{aligned}
$$

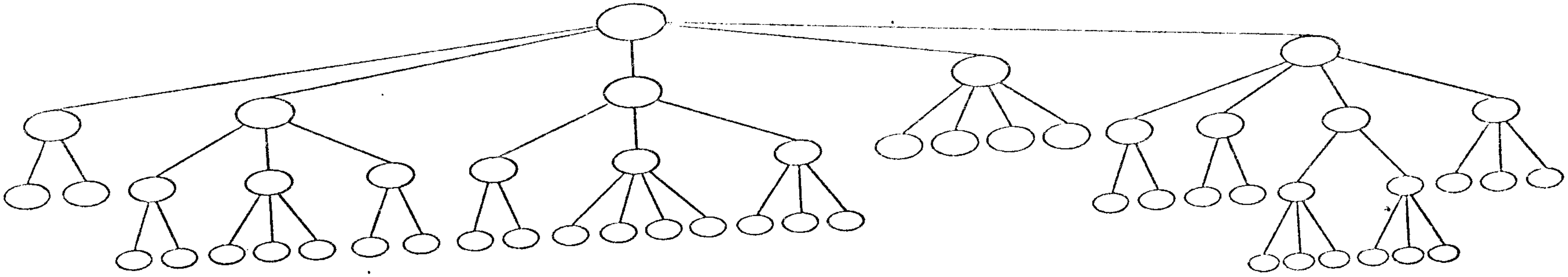
"HEURISTIC SEARCH WITH FORWARD PRUNING"

- AT EACH NODE, SAVE A LIMITED NUMBER OF BEST ALTERNATIVES
- TIME VARIES LINEARLY WITH THE NUMBER OF CLASSES, AS N(N+1) WHERE N IS THE NUMBER OF FEATURES
- USE INDIRECT MEASURE OF ERROR PROBABILITY
- IF DESCENDANTS DO NOT CORRESPOND TO DISJOINT CLASS SUBSETS, ESTIMATE THE TRANSITION PROBABILITIES BY CLASSIFYING THE TRAINING DATA

TW

Comparison of Decision Tree Classifiers and Single Stage Classifiers

Maximum Number of Features Per Node	Separability Measure	Separability Threshold	Trade Off Constant	Decision Tree Classification Time (Seconds)	Decision Tree Training Field Accuracy (%)	Single Stage Classification Time (Seconds)	Single Stage Training Accuracy (%)
4	B_T	1950	20.0,10.0	545	93.5	1574	93.7
4	D_T	1950	10.0	655	93.6		
3	B_T	1950	10.0	440	92.9	1036	93.0
3	D_T	1950	25.0	520	92.9		
2	B_T	1950	5.0	450	91.6	650	90.2
2	B_T	1850	5.0	390	90.4		
2	D_T	1850	5.0	435	92.2		



A typical machine-designed decision
tree for the Kenosha Pass example

Applications of Layered Classifiers

General Application	Example
Multi-image Analysis	Multitemporal classification Change detection
Use of Mixed Feature Types	Texture Topography Geophysical data
Recognition of Class-specific Properties	Crop stress detection Forest type mapping Water quality mapping Water temperature mapping

COMPARATIVE RESULTS

FAYETTE COUNTY, ILL. - LANDSAT - MULTITEMPORAL

CORN, SOYBEANS, WOODS, "OTHER"

t_1 (June 29, '73)	68%	235 sec.
t_2 (July 17, '73)	72%	240
LAYERED*	82%	224
CASCADE	84%	1658

* Second layer used only to resolve confusion classes

COMPARATIVE RESULTS

GRANT COUNTY, KAN. – LANDSAT – MULTITEMPORAL

WHEAT, CORN, ALFALFA, PASTURE, FALLOW

t_1 (May 9, '74)	62%	230 sec.
t_2 (July 20, '74)	55%	238 sec.
LAYERED	68%	512 sec.
CASCADE	64%	3037 sec

RESTRICTED DECISION TREE MODELS

LINEAR BINARY TREE

- FEWER STRUCTURES POSSIBLE
- COMPENSATE FOR GREATER DEPTH WITH LINEAR DECISION FUNCTION

ONE FEATURE PER NODE

- OPTIMALITY CLAIMED

MULTITEMPORAL ANALYSIS

REFERENCES

P. H. SWAIN, H. HAUSKA: THE DECISION TREE CLASSIFIER: DESIGN AND POTENTIAL, *IEEE TRANS. GEOSCIENCE ELECTRONICS*, VOL. GE-15, PP. 142-147, JULY 1977.

C. WU, D. LANDGREBE, P. SWAIN: THE DECISION TREE APPROACH TO CLASSIFICATION, *TECH. RPT. TR-EE 75-17*, SCHOOL OF ELECTRICAL ENGINEERING, PURDUE UNIVERSITY, MAY 1975.

K. C. YOU, K. S. FU: AN APPROACH TO THE DESIGN OF A LINEAR BINARY TREE CLASSIFIER, *PROC. SYMPOSIUM MACHINE PROCESSING OF REMOTELY SENSED DATA*, PURDUE UNIVERSITY 1976, PP. 3A-1 TO 3A-10 (IEEE CAT. NO. 76 CH 1103-1 MPRSD*).

A. V. KULKARNI, L. N. KANAL: AN OPTIMIZATION APPROACH TO HIERARCHICAL CLASSIFIER DESIGN, *PROC. THIRD INTERNATIONAL JOINT CONFERENCE ON PATTERN RECOGNITION*, NOVEMBER 1976, PP. 459-466 (IEEE CAT. NO. 76 CH 1140-3C*).

---

*IEEE, SINGLE COPY SALES, 445 HOES LANE, PISCATAWAY, NJ 03854

CHANGE DETECTION

AND MONITORING

- LAND USE PLANNING
- METEOROLOGICAL APPLICATIONS
- TACTICAL APPLICATIONS

Tu →

POST-CLASSIFICATION ("DIRECT")
CHANGE DETECTION

METHOD

- DIRECT COMPARISON OF INDEPENDENT CLASSIFICATIONS

ADVANTAGES

- TEMPORALLY DECOUPLED CLASSIFICATION
- USES "ORDINARY" METHODS FOR CLASSIFICATION
- CAN REPROCESS FOR CHANGE WITHOUT RECLASSIFYING

DISADVANTAGES

- NO USE OF MULTITEMPORAL INFORMATION IN CLASSIFICATION
- ERRORS ARE COMPOUNDED WITH TIME

DELTA-DATA

CHANGE DETECTION

METHOD

- CREATE "DELTA" (DIFFERENCE) DATA SET
- CLASSIFY

ADVANTAGES

- ONLY ONE CLASSIFICATION REQUIRED
- ONLY CHANGE CLASSES NEED BE CHARACTERIZED

DISADVANTAGES

- LARGE INFORMATION LOSS IN DIFFERENCING
- SENSITIVITY TO REGISTRATION ERROR

Tu

SPECTRO-TEMPORAL
CHANGE CLASSIFICATION

METHOD

- STACKED VECTOR ANALYSIS
- IDENTIFY "CHANGE" TRAINING FIELDS
- CLASSIFY AS USUAL

ADVANTAGES

- ONLY ONE CLASSIFICATION REQUIRED
- MULTITEMPORAL INFORMATION AVAILABLE FOR USE IN CLASSIFICATION

DISADVANTAGES

- LARGE NUMBER OF SPECTRO-TEMPORAL SUBCLASSES
- DIFFICULT TO FIND ADEQUATE TRAINING DATA FOR CHANGE CLASSES
- INCREASED DEMAND FOR TRAINING DATA
- COMPUTATIONAL LOAD

LAYERED CLASSIFIER

CHANGE DETECTION

METHOD

- USE DECISION TREE LOGIC FOR CLASSIFICATION AND DETECTION OF CHANGE

ADVANTAGES

- REQUIRES ONLY ONE CLASSIFICATION
- MULTITEMPORAL INFORMATION AVAILABLE FOR CLASSIFICATION
- "CHANGE CLASS" TRAINING DATA NOT NEEDED
- ALL ADVANTAGES OF LAYERED CLASSIFICATION

DISADVANTAGES

- LOGIC DESIGN MAY BE QUITE TEDIOUS

WEDNESDAY

INFORMATION FROM THE SPATIAL DOMAIN

I. Scene Segmentation and Classification of Objects - David A. Landgrebe

II. Texture - David A. Landgrebe

III. Context - Philip H. Swain

IV. Syntatic Scene Analysis - Philip H. Swain

Reference Papers for Short Course on Advanced Topics in the Analysis of Remote Sensing Data

Wednesday Session: Information from the Spatial Domain

Haralick, R.M., K. Shanmugam, and I. Dinstein. (1973) “Textural Features for Image Classification”. IEEE Transactions on Systems, Man, And Cybernetics, November 1973, Vol. SMC-#, No. 6, pp. 610-621. DOI: 10.1109/TSMC.1973.4309314.

Weszka, J.S., C.R. Dyer, and A. Rosenfeld. (1976) “A Comparative Study of Texture Measures for Terrain Classification”. IEEE Transactions on Systems, Man, and Cybernetics, April, 1976, Vol. SMC-6, No. 4, pp. 269-285. DOI: 10.1109/TSMC.1976.5408777.

Brayer, J.M., P.H. Swain, K.S. (1977) Modeling of Earth Resources Satellite Data. In: Fu K.S. (eds) Syntactic Pattern Recognition, Applications. Communication and Cybernetics, vol 14. Springer, Berlin, Heidelberg. DOI: 10.1007/978-3-642-66438-0_9

CONTEXT

IN REMOTE SENSING DATA ANALYSIS

EVIDENTLY HELPFUL

LOCAL VERSUS GLOBAL

PRESENCE INDICATORS VERSUS APPEARANCE MODIFIERS

TERMINOLOGY & NOTATION

$\Omega = \{\omega_1, \omega_2, \ldots, \omega_r\}$: r classes

$A = \{a_1, a_2, \ldots, a_s\}$: s possible actions

$L(\omega_i, a_j)$: loss incurred on deciding a_j when ω_i is the true class.

X : measurement vector

$p(X|\omega)$: class-conditional probability density function

$p(\omega)$: a priori probability of occurrance of class ω.

Bayes strategy: minimize the expected risk -

$$\frac{1}{p(X)} \sum_{i=1}^{r} L(\omega_i, a)\, p(X|\omega_i)\, p(\omega_i)$$

We

COMPOUND DECISION PROBLEM

N identical simple decision problems over an $\ell \times m = N$ frame

$$\begin{array}{ccc} \omega_{11} & \omega_{12} \cdots & \omega_{1\ell} \\ \omega_{21} & \omega_{22} \cdots & \omega_{2\ell} \\ \vdots & \vdots & \vdots \\ \omega_{m1} & \omega_{m2} \cdots & \omega_{m\ell} \end{array}$$

"frame"

$\underline{\omega} = [\omega_1, \omega_2, \ldots, \omega_N]$: vector of states

$\underline{X} = [X_1, X_2, \ldots, X_N]$: vector of measurement vectors

$\underline{a} = [a_1, a_2, \ldots, a_N]$: vector of "actions" (classifications)

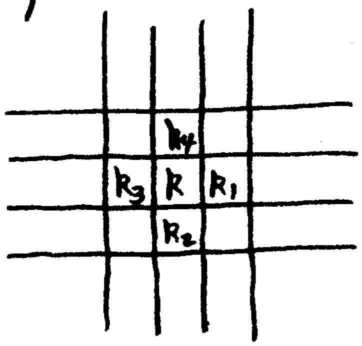
The compound decision problem: Minimize

$$L(\underline{\omega}, \underline{a}) \triangleq \frac{1}{N} \sum_{k=1}^{N} L(\omega_k, a_k)$$

Optimal strategy: Classify X_k (in k^{th} cell) to minimize

$$\sum_{\omega_k} L(\omega_k, a_k)\, p(\underline{X} \mid \omega_k)\, p(\omega_k)$$

$$p(\underline{X} \mid \omega_k) = p(\underline{X}_{NA}, \underline{X}_{AD}, X_k \mid \omega_k)$$



$$p(\underline{X} \mid \omega_k) = p(\underline{X}_{NA} \mid \underline{X}_{AD}, X_k, \omega_k)\, p(\underline{X}_{AD}, X_k \mid \omega_k)$$

* Assume contextual relationships between nonadjacent pixels are negligible.

$$p(\underline{X} \mid \omega_k) = p(\underline{X}_{NA} \mid \underline{X}_{AD})\, p(\underline{X}_{AD}, X_k \mid \omega_k)$$

$\therefore$ Minimize (choose $a_k \in A$) -

$$\sum_{\omega_k} L(\omega_k, a_k)\, p(\underline{X}_{NA} \mid \underline{X}_{AD})\, p(\underline{X}_{AD}, X_k \mid \omega_k)\, p(\omega_k)$$

or, equivalently, minimize

$$\sum_{\omega_k} \underbrace{L(\omega_k, a_k)}_{}\, \underbrace{p(\underline{X}_{AD}, X_k \mid \omega_k)\, p(\omega_k)}_{}$$

(cont.)

$$p(\underline{X}_{AD}, X_k | \omega_k) = \underset{(I)}{p(X_k | \underline{X}_{AD}, \omega_k)}\,\underset{(II)}{p(\underline{X}_{AD} | \omega_k)} \quad (*)$$

(II):

$$\begin{aligned} p(\underline{X}_{AD} | \omega_k) &= p(X_{k1}, X_{k2}, X_{k3}, X_{k4} | \omega_k) \\ &= p(X_{k1} | X_{k2}, X_{k3}, X_{k4}, \omega_k) \times \\ &\quad p(X_{k2} | X_{k3}, X_{k4}, \omega_k) \times \\ &\quad p(X_{k3} | X_{k4}, \omega_k) \times \\ &\quad p(X_{k4} | \omega_k) \end{aligned}$$

but X_k's are nonadjacent & class-conditionally independent, so

$$\begin{aligned} p(\underline{X}_{AD} | \omega_k) &= p(X_{k1} | \omega_k) \cdot p(X_{k2} | \omega_k) \cdots p(X_{k4} | \omega_k) \\ &= \prod_{i=1}^{4} p(X_{k_i} | \omega_k) \end{aligned}$$

Now, write

$$\begin{aligned} p(X_{ki} | \omega_k) &= \sum_{\omega_{ki}} p(X_{ki}, \omega_{ki} | \omega_k) \\ &= \sum p(X_{ki} | \omega_{ki}, \omega_k)\, p(\omega_{ki} | \omega_k) \\ &= \sum p(X_{ki} | \omega_{ki})\, p(\omega_{ki} | \omega_k) \end{aligned}$$

(cont.)

∴ We now have –

$$p(\underline{X}_{AD}|\omega_k) = \prod_{i=1}^{4} \sum_{\omega_{ki}} p(X_{ki}|\omega_{ki})\, p(\omega_{ki}|\omega_k)$$

(I):

$$p(X_k|\underline{X}_{AD},\omega_k) = p(X_k|\omega_k)$$

∴ (✱) – top of previous page – becomes

$$p(\underline{X}_{AD}, X_k|\omega_k) = p(X_k|\omega_k)\prod_{i=1}^{4} \sum_{\omega_{ki}} p(X_{ki}|\omega_{ki})\, p(\omega_{ki}|\omega_k)$$

Decision rule: Classify X_k as a_k to minimize

$$\sum_{\omega_k} L(\omega_k, a_k)\, p(X_k|\omega_k)\prod_{i=1}^{4} \sum_{\omega_{ki}} p(X_{ki}|\omega_{ki})\, p(\omega_{ki}|\omega_k)$$

Maximum Likelihood: Classif $X_k \in \omega_k$ to maximize

$$p(X_k|\omega_k)\, p(\omega_k)\prod_{i=1}^{4} \sum_{\omega_{ki}} p(X_{ki}|\omega_{ki})\, p(\omega_{ki}|\omega_k)$$

Reference

- J.R. Welch, K.G. Salter: A Context Algorithm for Pattern Recognition and Image Interpretation, IEEE Trans. Systems, Man and Cybernetics, Vol. SMC-1, No. 1, pp. 24-30, January 1971.

SYNTACTIC SCENE ANALYSIS

CHARACTERIZATION AND ANALYSIS
OF
SCENE STRUCTURE AND STRUCTURAL RELATIONSHIPS

We

AN ANALOGY: LANGUAGE

LETTERS
→ WORDS
→ PHRASES
→ SENTENCE

$$\underbrace{\text{THE NEW LANDSAT}}_{\text{noun phrase}}\ \underbrace{\text{HURTLED}\ \underbrace{\text{INTO}\ \underbrace{\text{A PERFECT ORBIT}}_{\text{noun phrase}}}_{\text{prepositional phrase}}}_{\text{verb phrase}}$$

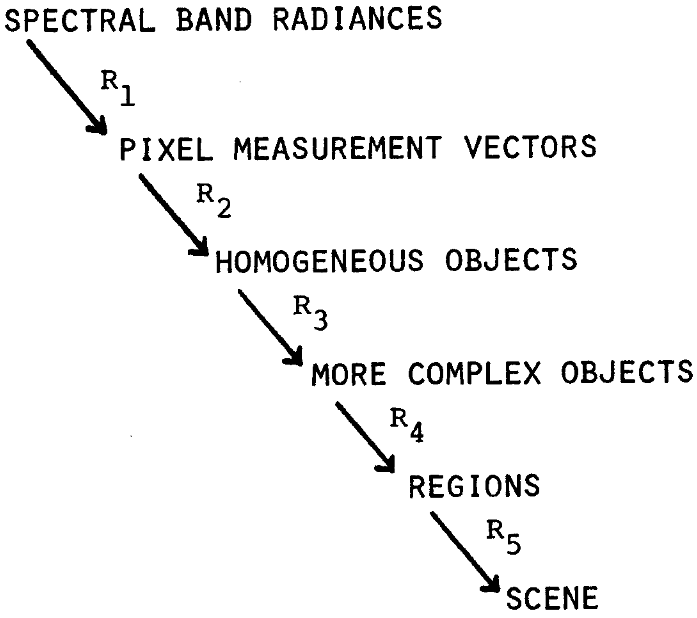
MEANING DEPENDS ON:

- THE SPECIFIC PRIMITIVES (LETTERS)
- RELATIONSHIPS AMONG PRIMITIVES (WORDS)
- RELATIONSHIPS AMONG PATTERNS OF PRIMITIVES (PHRASES)
- RELATIONSHIPS AMONG PATTERNS OF PATTERNS (PHRASES ⟶ SENTENCE)

EXAMPLE:

GENERALLY SPEAKING WOMEN ARE PRETTY
WOMEN ARE PRETTY GENERALLY SPEAKING
ARE PRETTY SPEAKING GENERALLY WOMEN
GENERALLY SPEAKING WOMEN ARE TRETPY

TYPICAL IMAGE-ORIENTED HIERARCHY:



SPECTRAL BAND RADIANCES

R_1

PIXEL MEASUREMENT VECTORS

R_2

HOMOGENEOUS OBJECTS

R_3

MORE COMPLEX OBJECTS

R_4

REGIONS

R_5

SCENE

GRAMMAR

A <u>PHRASE-STRUCTURE GRAMMAR</u> G IS

A 4-TUPLE, $G = (V_N, V_T, P, S)$, WHERE

V_N: NONTERMINALS

V_T: TERMINALS

S (IN V_N): START SYMBOL

P IS A FINITE SET OF PRODUCTIONS (REWRITING RULES), OF THE FORM $\alpha \rightarrow \beta$ WHERE α IS A STRING OF ONE OR MORE TERMINALS AND NON-TERMINALS, β IS A STRING OF TERMINALS AND NONTERMINALS.

COMMON NOTATION

V^*: SET OF ALL STRINGS OVER V, INCLUDING (STRING OF LENGTH 0)

$V^+ = V^* - \{\lambda\}$

$|x|$ = LENGTH OF STRING X

X^N = N REPETITIONS OF X

$A \underset{G}{\Longrightarrow} B$: A DERIVES B BY APPLICATION OF ONE PRODUCTION IN G

$A \overset{*}{\underset{G}{\Longrightarrow}} B$: A DERIVES B BY APPLICATION OF A SEQUENCE OF PRODUCTIONS IN G

THE LANGUAGE L GENERATED BY A GRAMMAR G IS THE SET OF STRINGS

$$L(G) = \left\{x \mid x \in V_T{}^* \text{ AND } S \overset{*}{\underset{G}{\Longrightarrow}} X\right\}$$

TYPES OF PHRASE-STRUCTURE GRAMMARS —

FORM OF PRODUCTION: $\alpha \rightarrow \beta$

CONTEXT-SENSITIVE (TYPE 1):

$$|\alpha| \leq |\beta|$$

CONTEXT-FREE (TYPE 2):

$$\alpha \in V_N$$

REGULAR (TYPE 3):

$$\left.\begin{array}{l} \alpha = A \\ \beta = aB \text{ or } \beta = a \end{array}\right\} \begin{array}{l} A \in V_N \\ B \in V_N \\ a \in V_T \end{array}$$

CONTEXT-FREE GRAMMAR G:

$V_N = \{S, A, B\}$, $V_T = \{a, b\}$,

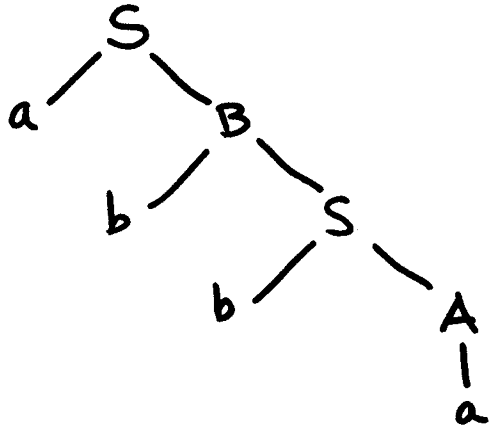
P:
(1) $S \rightarrow aB$
(2) $S \rightarrow bA$
(3) $A \rightarrow aS$
(4) $A \rightarrow bAA$
(5) $A \rightarrow a$
(6) $B \rightarrow bS$
(7) $B \rightarrow aBB$
(8) $B \rightarrow b$

$L(G) = \{X \mid X \in V_T^{+}$ and X contains an equal number of a's and b's$\}$

Typical derivation in G:

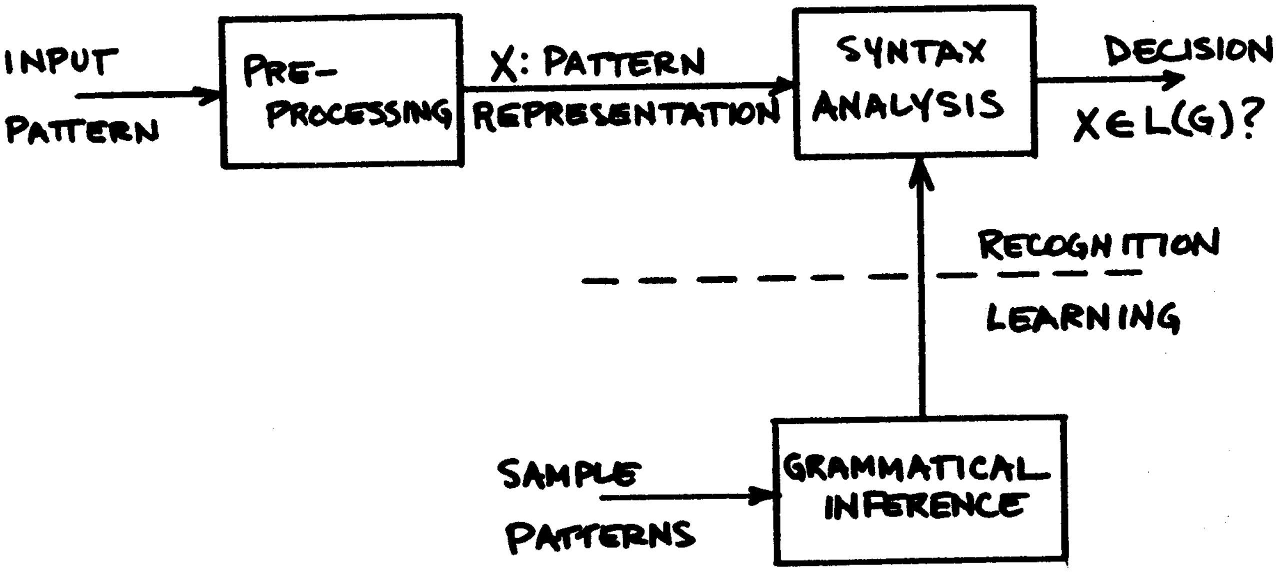
$$S \xRightarrow{(1)} aB \xRightarrow{(6)} aBS \xRightarrow{(2)} abbA \xRightarrow{(5)} abba$$

Derivation tree (same derivation):



SYNTACTIC PATTERN RECOGNITION

-SYSTEM MODEL-



SELECTION OF PATTERN PRIMITIVES

DEPENDS ON

- -- THE NATURE OF THE DATA
- -- THE APPLICATION
- -- THE IMPLEMENTATION TECHNOLOGY

REQUIREMENTS

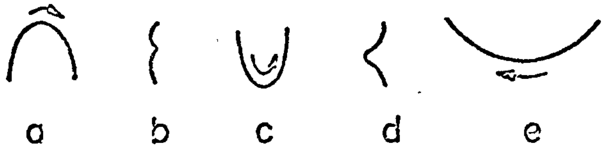
- -- BASIC PATTERN ELEMENTS
- -- EASILY RECOGNIZED

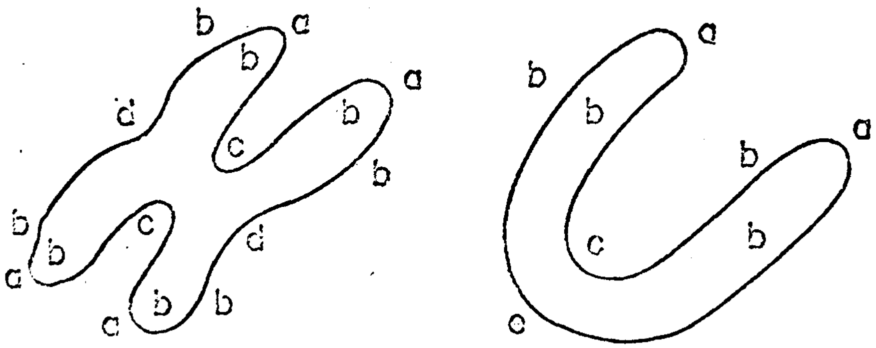
EXAMPLE

$V_T = \{a,b,c,d,e\}$

$V_N = \{S, T, \text{Bottom}, \text{Side}, \text{Armpair}, \text{Rightpart}, \text{Leftpart}, \text{Arm}\}$

P: S → Armpair · Armpair
T → Bottom · Armpair
Armpair → Side · Armpair
Armpair → Armpair · Side
Armpair → Arm · Rightpart
Armpair → Leftpart · Arm
Leftpart → Arm · c
Rightpart → c · Arm
Bottom → b · Bottom
Bottom → Bottom · b
Bottom → c
Side → b · Side
Side → Side · b
Side → b · d
Arm → b · Arm
Arm → Arm · b
Arm → a





A Grammar for Submedian and Telocentric Chromosomes

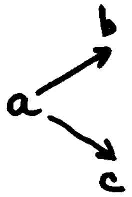
A WEB GRAMMAR:

$G = (V_N, V_T, P, S)$

where $V_N = \{A\}$, $V_T = \{a, b, c\}$, $S = \{A\}$

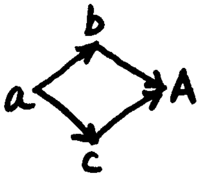
and P:

(1) A ⟶



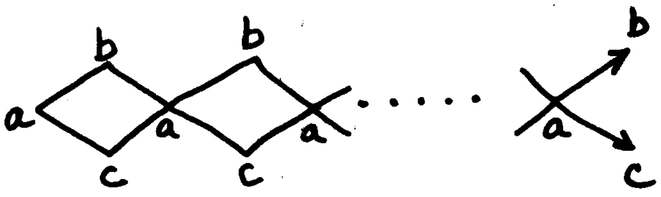
$E = \{(p, a) \mid (p, A)$ is an edge in the host web$\}$

(2) A ⟶



E = (same as above)

L(G) = the set of all webs of the form:



REFERENCES FOR SYNTACTIC PATTERN RECOGNITION

K.S. FU, P.H. SWAIN: 'ON SYNTACTIC PATTERN RECOGNITION, IN *SOFTWARE ENGINEERING*, VOL. 2, ACADEMIC PRESS, 1971.

P.H. SWAIN, K.S. FU: *PATTERN RECOGNITION* (JOURNAL OF), VOL. 4, PP. 83-100, 1972.

K.S. FU: *SYNTACTIC METHODS IN PATTERN RECOGNITION*, ACADEMIC PRESS, 1974.

K.S. FU, ED.: *SYNTACTIC PATTERN RECOGNITION, APPLICATIONS*, SPRINGER-VERLAG, 1977.

J.F. PFALTZ, A. ROSENFELD: *PROCEEDINGS FIRST INTERNATIONAL JOINT CONFERENCE ON ARTIFICIAL INTELLIGENCE*, WASHINGTON, D.C., 1969, PP. 609-619.

THURSDAY
SYSTEM DESIGN AND EVALUATION

Reference Papers for Short Course on Advanced Topics in the Analysis of Remote Sensing Data

Thursday Session: System Design and Evaluation

Bauer, M.E., M.M. Hixson, B.J. Davis, and J.B. Etheridge. (1977) "Crop Identification and Area Estimation by Computer-Aided Analysis of Landsat Data". Proceedings of 1977 Machine Processing of Remotely Sensed Data Symposium, Purdue University, June 21-23, 1977.
https://docs.lib.purdue.edu/lars_symp/188/.

SYSTEM DESIGN AND EVALUATION

MARVIN BAUER AND BARBARA DAVIS
APRIL 13, 1978

- I. SYSTEM DESIGN
  - A. SYSTEM DESIGN - KEY STEPS
  - B. DEFINING INFORMATION REQUIREMENTS
  - C. ESTABLISHING PROJECT FEASIBILITY
  - D. PROJECT PLANNING AND IMPLEMENTATION
    - 1. DATA ACQUISITION
      - A. REMOTE SENSING DATA
      - B. REFERENCE AND ANCILLARY DATA
    - 2. SAMPLE DESIGN
      - A. ENUMERATION AND SAMPLING
      - B. SAMPLING TECHNIQUES
      - C. STRATIFICATION
    - 3. DATA PREPROCESSING
      - A. SELECTION AND USE OF TECHNIQUES
    - 4. DATA ANALYSIS
      - A. RELATION OF DATA ACQUISITION, SAMPLE DESIGN AND PREPROCESSING TO ANALYSIS

- II. EVALUATION OF RESULTS
  - A. STATISTICAL TERMS
  - B. IMPACT OF OBJECTIVES
  - C. ASSESSING CLASSIFICATION AND MAP ACCURACY
    - 1. TRAINING SET EVALUATION
    - 2. TEST DATA SELECTION
    - 3. IMPACT OF REFERENCE DATA QUALITY ON ACCURACY
    - 4. CONFIDENCE INTERVALS ON ACCURACY

System Design and Evaluation (cont'd.)

D. Area and Proportion Estimation
   1. Calculating Estimates
   2. Calculating Bias and Precision of Estimates
   3. Removing Bias from Area Estimates
   4. Importance of Classification Accuracy in Area Estimation

E. Comparing Estimates and Reference Data
   1. Analysis of Variance
   2. Correlation
   3. Impact of Reference Data Quality

SYSTEM DESIGN -- KEY STEPS

1. DEFINE INFORMATION REQUIREMENTS AND OBJECTIVES
2. ESTABLISH FEASIBILITY
3. PLAN THE PROJECT
4. IMPLEMENT THE PROJECT
5. USE AND EVALUATE RESULTS

DEFINING INFORMATION REQUIREMENTS -- KEY QUESTIONS

- WHAT ARE THE EARTH SURFACE FEATURES OR COVER TYPES OF INTEREST?
- WHAT DOES THE USER WANT TO KNOW ABOUT THESE COVER TYPES? LOCATION, AREA, CONDITION?
- WHAT SIZE AREA IS INVOLVED?
- IN WHAT FORMAT ARE THE RESULTS NEEDED? MAPS, TABLES, OR BOTH?
- HOW ACCURATE MUST THE RESULTS BE?
- IS COMPLETE COVERAGE OF THE AREA REQUIRED OR WILL A SAMPLE PROVIDE THE NECESSARY INFORMATION?
- WHAT ARE THE TEMPORAL CONSIDERATIONS? TIMING AND FREQUENCY?

ESTABLISHING PROJECT FEASIBILITY

- REVIEW PREVIOUS INVESTIGATIONS AND RESULTS
- ADDITIONAL RESEARCH MAY BE REQUIRED TO DETERMINE FEASIBILITY
- FEASIBILITY STUDY SHOULD INCLUDE CONSIDERATION OF WHETHER:
  - COVER TYPES OF INTEREST HAVE DISTINGUISHABLE SPECTRAL, SPATIAL, OR TEMPORAL CHARACTERISTICS
  - SUITABLE DATA COLLECTION SYSTEMS ARE AVAILABLE
  - APPROPRIATE DATA PROCESSING AND ANALYSIS SYSTEMS ARE AVAILABLE
- PRODUCT OF FEASIBILITY STUDY SHOULD BE LIST OF OPTIONAL APPROACHES

Project Planning and Implementation

- Project planning is more specific than feasibility study
- Project planning narrows choices to a specific approach
- Decisions must be made with respect to:
  - data collection system used
  - Frequency of data collection
  - Reference data and ground observation requirements
  - Data preprocessing requirements
  - Data analysis procedures
  - Plan for results utilization and evaluation
- Implementation -- Decision making steps are put into action

DATA ACQUISITION

- SENSOR SELECTION
  1. SPECTRAL CONSIDERATIONS
     - SPECTRAL DISCRIMINABILITY OF COVER TYPES OF INTEREST
  2. SPATIAL CONSIDERATIONS
     - SIZE OF SURVEY AREA
     - COMPLETE AREA COVERAGE OR SAMPLE SEGMENT COVERAGE
     - RESULTS FORMAT: MAPS OR STATISTICS
     - SPATIAL CHARACTERISTICS OF COVER TYPES IN RELATION TO SENSOR RESOLUTION
  3. TEMPORAL CONSIDERATIONS
     - WHAT IS BEST OR REQUIRED TIME OF YEAR TO COLLECT DATA?
     - HOW FREQUENTLY IS DATA REQUIRED?
     - ARE THERE DIURNAL CONSIDERATIONS?
     - IS MULTITEMPORAL DATA REQUIRED FOR DISCRIMINATION?

- REFERENCE DATA COLLECTION
  1. GROUND OBSERVATIONS
  2. AERIAL PHOTOGRAPHY
  3. USE OF MULTISTAGE SAMPLING
- ANCILLARY DATA COMPILATION
  1. MAPS
  2. HISTORICAL DATA
- FIRST HAND EXPERIENCE IS INVALUABLE

SAMPLE DESIGN

A. ENUMERATION AND SAMPLING

- DEFINITIONS (IN A REMOTE SENSING CONTEXT)
  - ENUMERATION IS THE IDENTIFICATION OF EACH AND EVERY MEMBER OF THE POPULATION, E.G. ALL THE PIXELS IN A LANDSAT SCENE.
  - SAMPLING IS THE SELECTION AND IDENTIFICATION OF A PART (SAMPLE) OF THE POPULATION TO REPRESENT THE ENTIRE POPULATION.

- ADVANTAGES OF ENUMERATION
  - GIVES THE MOST PRECISE RESULTS; THERE IS NO VARIANCE
  - EACH INDIVIDUAL MEMBER OF THE POPULATION IS IDENTIFIED
- ADVANTAGES OF SAMPLING
  - REDUCED COST AND GREATER SPEED IN OBTAINING AND PROCESSING INFORMATION
  - GREATER SCOPE OF INVESTIGATION MAY BE POSSIBLE
  - GREATER ACCURACY MAY BE OBTAINED BY USING MORE RELIABLE PERSONNEL AND A DECREASED VOLUME OF WORK.

- Enumeration or Sampling??
  - The choice of enumeration or sampling depends upon the problem at hand.
  - Three criteria must be considered to make the choice:
    - Precision required
    - Kind of information required
    - Resources available

Th ✓

IT IS NECESSARY TO OBTAIN AN ACCEPTABLE BALANCE AMONG THESE THREE FACTORS.

SOME EXAMPLES:

- PRODUCING A MAP GENERALLY REQUIRES ENUMERATION, I.E. CLASSIFICATION OF EACH PIXEL.

- IF NO VARIANCE IS PERMISSIBLE, ENUMERATION MUST BE CARRIED OUT. SAMPLES, HOWEVER, CAN PRODUCE ESTIMATES WITH QUITE SMALL VARIANCES.

- ENUMERATION SHOULD NOT BE USED IN SELECTION OF A TRAINING SET; OTHERWISE, WHAT IS THE USE OF THE CLASSIFICATION?

B. SAMPLING TECHNIQUES

- PRINCIPAL STEPS IN A SAMPLE SURVEY
  - OBJECTIVES
  - POPULATION TO BE SAMPLED
  - DATA TO BE COLLECTED
  - DEGREE OF PRECISION REQUIRED
  - METHOD OF MEASUREMENT
  - THE SAMPLING FRAME
  - SELECTION OF THE SAMPLE
  - PRETEST
  - ORGANIZATION OF THE WORK
  - SUMMARY AND ANALYSIS OF DATA
  - INFORMATION GAINED FOR FUTURE SURVEYS

Thr

- Role of Sampling Theory

- Purpose of theory is to make sampling more efficient.
- It attempts to develop methods of sample selection and estimation that provide, at the lowest possible cost, estimates that are precise enough for the specified purpose.
- To apply this principle, we need to know for any sampling procedure under consideration, the precision and the cost to be expected.
  - Precision is judged by examining the frequency distribution generated for the estimate if the procedure is applied repeatedly to the same population.
  - A further simplification--the sample estimates are usually approximately normally distributed. Therefore, the frequency distribution is uniquely defined by the mean and standard deviation.

- PROBABILITY SAMPLING

  - ALL SAMPLING PROCEDURES FOR WHICH A THEORY HAS BEEN DEVELOPED HAVE THE FOLLOWING PROPERTIES:

    1. SET OF DISTINCT SAMPLES, S_1, $S_2, \ldots S_n$, CAN BE DEFINED.
    2. EACH POSSIBLE SAMPLE S_i HAS A KNOWN PROBABILITY OF SELECTION π_i
    3. ONE OF THE S_i IS SELECTED BY A PROCESS IN WHICH EACH S_i RECEIVES ITS APPROPRIATE PROBABILITY π_i OF SELECTION
    4. THE METHOD FOR COMPUTING THE ESTIMATE FROM THE SAMPLE MUST LEAD TO A UNIQUE ESTIMATE FROM ANY SPECIFIC SAMPLE

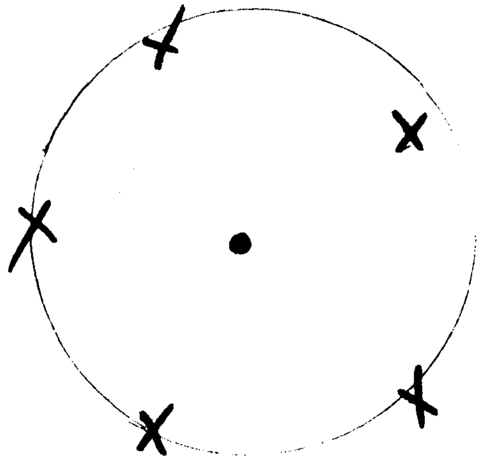
  - FOR ANY SAMPLING PROCEDURE THAT SATISFIES THESE PROPERTIES, WE CAN CALCULATE THE FREQUENCY DISTRIBUTION OF THE ESTIMATES IT GENERATES.

  - A SAMPLING THEORY CAN BE DEVELOPED FOR ANY PROCEDURE OF THIS TYPE.

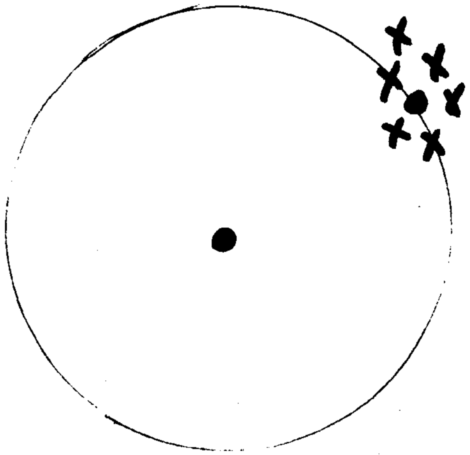
- Nonprobability Sampling
  - Is not amenable to the development of a sampling theory since no element of random selection is involved.
  - Therefore, it is not possible to determine magnitude of sampling errors and bias.
- Since there are, in actuality, almost always requirements of precision and accuracy (although this is not always recognized as such), probability sampling is highly recommended.

Th

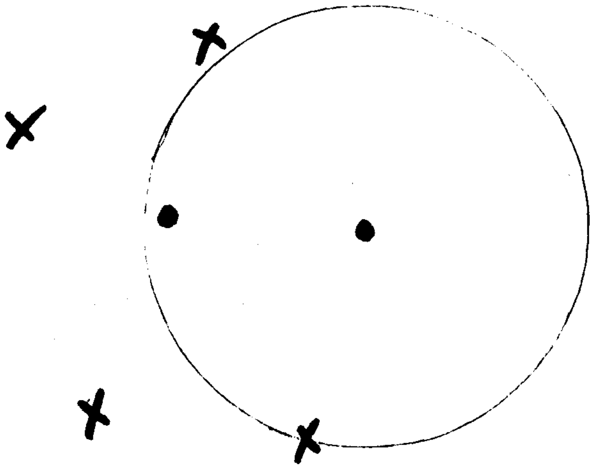
o ILLUSTRATIONS OF PRECISION AND ACCURACY



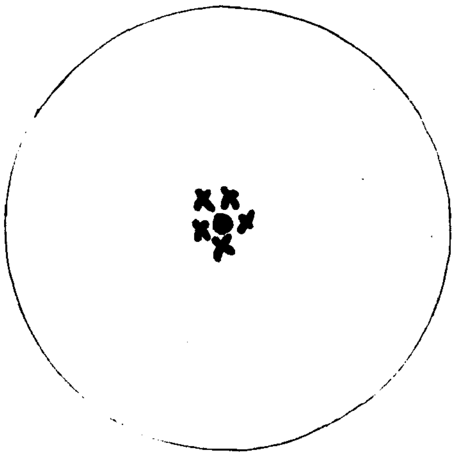
ACCURATE
NOT PRECISE



NOT ACCURATE (BIASED)
PRECISE



NOT ACCURATE
NOT PRECISE



ACCURATE
PRECISE

- THREE BASIC TYPES OF PROBABILITY SAMPLING
  - SIMPLE RANDOM SAMPLING
  - STRATIFIED RANDOM SAMPLING
  - SYSTEMATIC RANDOM SAMPLING
- PROBLEM: SUPPOSE WE WANT TO DETERMINE THE AREAS TO BE CLASSIFIED FOR MAKING AN ESTIMATE OF THE AMOUNT OF WHEAT IN KANSAS. HOW CAN THIS BE DONE USING EACH TYPE OF PROBABILITY SAMPLING?

Simple Random Sampling

(1) Divide the entire state into N blocks of a given size.

(2) Determine the number to be tabulated, say n, considering precision and cost.

(3) Randomly select n of the N areas.

(4) Classify and make a state estimate.

STRATIFIED RANDOM SAMPLING

(1) DIVIDE THE ENTIRE STATE INTO N BLOCKS OF A GIVEN SIZE.

(2) DETERMINE THE NUMBER TO BE TABULATED, SAY N, CONSIDERING PRECISION AND COST.

(3) ALLOCATE THE N SAMPLES TO COUNTIES ACCORDING TO THE HISTORICAL PROPORTION OF WHEAT.

(4) CLASSIFY AND MAKE A STATE ESTIMATE BY AGGREGATING THE COUNTY ESTIMATES.

Systematic Random Sampling

(1) Determine the sample size required to obtain acceptable precision and cost.

(2) For each county, systematically allocate the sample (i.e. randomly select a starting point and sample at a fixed interval thereafter).

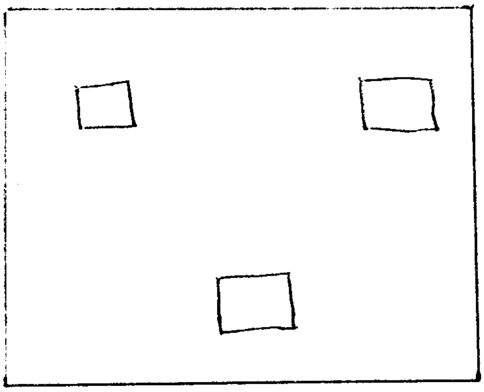
(3) Classify and make a state estimate, aggregating the county results.

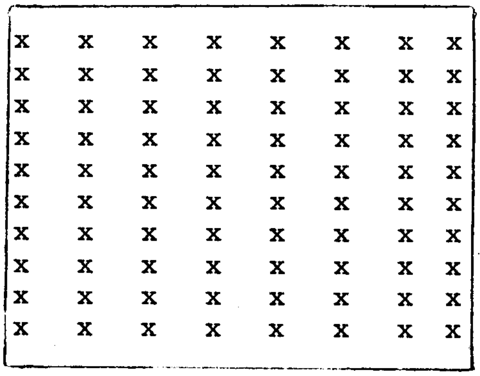
ADVANTAGES AND DISADVANTAGES

- SIMPLE RANDOM SAMPLING
  - MOST EASILY UNDERSTOOD BY PUBLIC
  - LESS CONVENIENT AND LESS PRECISE
- STRATIFIED RANDOM SAMPLING
  - ADMINISTRATIVE CONVENIENCE OF DIVIDING THE WORK BETWEEN SEVERAL INDIVIDUALS OR FIELD OFFICES
  - ABILITY TO ACHIEVE KNOWN PRECISION FOR SUBDIVISIONS OF THE POPULATION
  - INCREASED PRECISION
- SYSTEMATIC RANDOM SAMPLING
  - EASIER TO DRAW SAMPLES AND EXECUTE WITHOUT MISTAKES
  - SOMETIMES CONSIDERABLY MORE PRECISE THAN STRATIFIED RANDOM SAMPLING
  - AVOID SIX SCAN LINE NOISE BY SELECTING LINE AND COLUMN INTERVALS FOR SAMPLING AS RELATIVELY PRIME TO 6 AS POSSIBLE

SAMPLE SIZES

- SAMPLES CAN BE TAKEN EITHER AS SEGMENTS (BLOCKS OF AREA SUCH AS 5X6 NM) OR AS INDIVIDUAL PIXELS.
- SYSTEMATIC SAMPLES OF PIXELS DISTRIBUTED THROUGHOUT A COUNTY HAVE BEEN SHOWN TO PRODUCE ACCURATE AND PRECISE AREA ESTIMATES.
- PIXELS ARE NOT ALWAYS APPROPRIATE; FOR EXAMPLE, INDIVIDUAL PIXELS USED AS TRAINING FIELDS ARE INFEASIBLE.





C. STRATIFICATION

- STRATIFICATION FROM THE STATISTICAL VIEWPOINT:
  - DIVISION OF THE POPULATION OF INTEREST INTO RELATIVELY HOMOGENEOUS SUBGROUPS (STRATA).
- STRATIFICATION FROM THE REMOTE SENSING VIEWPOINT:
  - DIVISION OF THE SCENE INTO SUBGROUPS WHICH ARE RELATIVELY HOMOGENEOUS SPECTRALLY.
  - USED TO PERMIT TRAINING STATISTICS DEVELOPED FOR ONE PORTION OF THE SCENE TO BE SUCCESSFULLY USED TO CLASSIFY OTHER AREAS IN THE SCENE.
  - THE TRAINING STATISTICS MUST ADEQUATELY REPRESENT THE VARIABILITY IN THE AREA TO BE CLASSIFIED.
  - FACTORS WHICH MAY CONTRIBUTE TO THE DEFINITION OF STRATA INCLUDE SOIL TYPE, ATMOSPHERIC CONDITIONS, LAND USE, AND CROPS PRESENT.
- STRATIFICATION FROM THE SAMPLING VIEWPOINT
  - ALLOCATION OF AREAS TO BE CLASSIFIED FOR AN ESTIMATION PROBLEM IS DONE ACCORDING TO THE PROPORTION OF A CROP HISTORICALLY PRESENT.
  - THIS MAY NOT BE REALLY NECESSARY FOR ACCURATE AND PRECISE ESTIMATES IF A GOOD TRAINING SET IS FOUND.

DATA PREPROCESSING

- TECHNIQUES
  - GEOMETRIC CORRECTION
  - MULTITEMPORAL REGISTRATION
  - ANCILLARY DATA REGISTRATION
  - RADIOMETRIC CORRECTION

- BASIS FOR SELECTION AND USE OF TECHNIQUES

  GEOMETRIC CORRECTION
  - REQUIRED FOR MAP PREPARATION
  - HELPFUL FOR INTERFACING WITH DATA

  MULTITEMPORAL REGISTRATION
  - REQUIRED IF SINGLE DATE IS INADEQUATE TO ACHIEVE REQUIRED ACCURACY
  - REQUIRED FOR CHANGE DETECTION

  ANCILLARY DATA REGISTRATION
  - HELPFUL FOR INTERFACING WITH DATA
  - MAY BE USED AS A FEATURE FOR CLASSIFICATION
  - MAY BE USED TO TABULATE CLASSIFICATION RESULTS FOR SPECIFIED AREAS

  RADIOMETRIC CORRECTION
  - REQUIRED IF COMPARISONS OF RESPONSE FROM TIME TO TIME OR PLACE TO PLACE ARE NEEDED

DATA ANALYSIS

- RELATION OF DATA ACQUISITION, SAMPLE DESIGN, AND PREPROCESSING TO ANALYSIS

- THE FOLLOWING POINTS MUST BE CONSIDERED IN SELECTING AN ANALYSIS PROCEDURE:
  - SCANNER TYPE OR RESOLUTION
  - QUALITY OF MSS DATA
  - QUANTITY (VOLUME) OF DATA
  - AVAILABILITY AND QUALITY OF REFERENCE DATA
  - RESOURCES AVAILABLE IN TERMS OF PERSONNEL AND COMPUTER TIME
  - SKILL AND EXPERIENCE OF PERSONNEL
  - PREPROCESSING TECHNIQUES

- SCANNER TYPE OR RESOLUTION MAY INFLUENCE
    - METHOD OF OBTAINING TRAINING SAMPLES AND CALCULATING STATISTICS
    - CLASSIFIER
- EXAMPLES
    - SUPERVISED CLASSIFICATION IS EASIER WITH AIRCRAFT DATA THAN WITH LANDSAT DATA BECAUSE TRAINING FIELDS ARE GENERALLY LARGER AND EASIER TO FIND.
    - AN ECHO CLASSIFIER WILL GIVE GREATER BENEFITS OVER A POINT CLASSIFIER IN AREAS WITH LARGER OBJECTS.

Th ✓

- SPECIAL ANALYSIS PROCEDURES MAY BE REQUIRED TO OFFSET POOR DATA QUALITY

- EXAMPLES:

  - IF MULTITEMPORAL DATA IS AVAILABLE WHERE CLOUDS DO NOT OCCUR IN THE SAME LOCATION ON BOTH DATES, MULTITEMPORAL ANALYSIS OR THE USE OF A LAYERED CLASSIFIER CAN OVERCOME CLOUD PROBLEMS.

  - BAD DATA LINES CAN BE "REMOVED" BY THE SAME TECHNIQUES.

- THE QUANTITY (VOLUME) OF DATA HAS GREAT IMPACT ON THE COSTS OF ANALYSIS IN TERMS OF PERSONNEL AND COMPUTER TIME. THIS EFFECT CAN BE MITIGATED BY JUDICIOUS CHOICE OF ANALYSIS PROCEDURES.

- EXAMPLES:
  - SUPERVISED TRAINING ON LANDSAT DATA CAN BE MORE COSTLY AND DIFFICULT THAN UNSUPERVISED TRAINING DEPENDING ON THE AREA OF INTEREST AND THE REFERENCE DATA AVAILABLE.
  - TRAINING SAMPLES CAN BE USED TO ACCURATELY CLASSIFY LARGE AREAS IF THEY ARE GEOGRAPHICALLY DISPERSED TO REPRESENT THE SPECTRAL VARIATION IN EACH COVER TYPE.

- THE AVAILABILITY AND QUALITY OF REFERENCE DATA AFFECTS THE QUALITY OF TRAINING STATISTICS.
    - STATISTICS OBTAINED FROM A SMALL AREA WILL NOT PRODUCE AS ACCURATE A CLASSIFICATION AS THOSE OBTAINED FROM GEOGRAPHICALLY DISPERSED SAMPLES.
    - GIGO, I.E. IF THE REFERENCE DATA CONTAINS MISTAKES, THE CLASSIFICATION WILL MAKE MORE MISTAKES.
    - IF THE REFERENCE DATA IS OLDER THAN THE DATA TO BE CLASSIFIED, CHANGES IN THE SCENE WILL CAUSE ERRORS TO OCCUR.

- PREPROCESSING TECHNIQUES WITH APPROPRIATE ANALYSIS PROCEDURES CAN AID IN CLASSIFICATION BY:
  - INCREASING SPECTRAL DISCRIMINABILITY THROUGH THE USE OF MULTITEMPORAL DATA.
  - INCREASING THE AREA THAT CAN BE ACCURATELY CLASSIFIED WITH THE SAME TRAINING SET BY USING RADIOMETRIC CORRECTIONS.

DEFINITIONS OF STATISTICAL TERMS

- ACCURACY
  - LOW ERROR RATE
  - UNBIASEDNESS

LET X BE AN ESTIMATE FOUND FROM A SAMPLE. THE EXPECTED VALUE OF X IS

$$E(X) = \sum_{R} R \, P(X=R)$$

WHERE THE SUM EXTENDS OF ALL POSSIBLE VALUES R OF X.

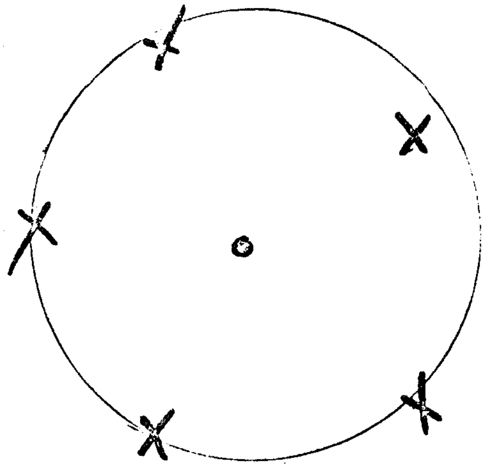
X WILL BE UNBIASED IF E(X) IS EQUAL TO THE QUANTITY BEING ESTIMATED.

- PRECISION

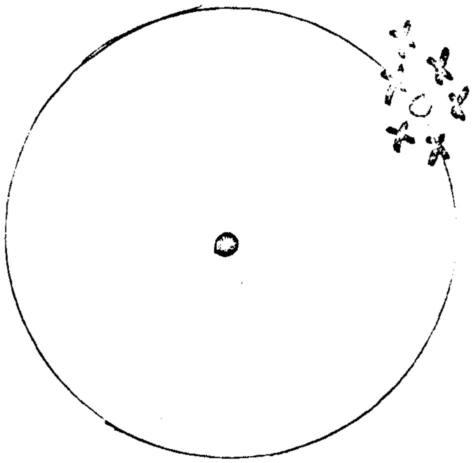
LOW VARIANCE

$$V(X) = E\left[(X-E[X])^{2}\right]$$

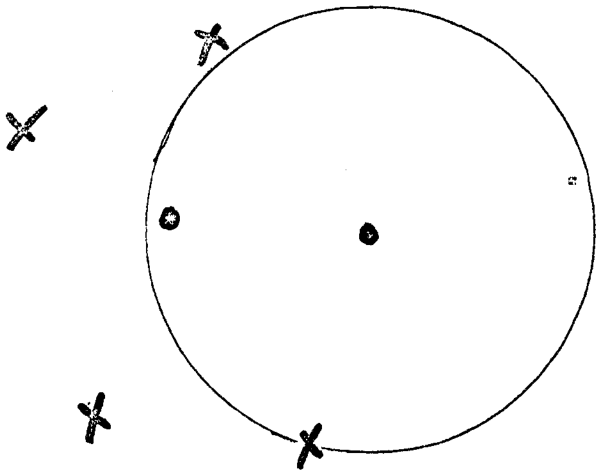
○ Illustrations of Precision and Accuracy



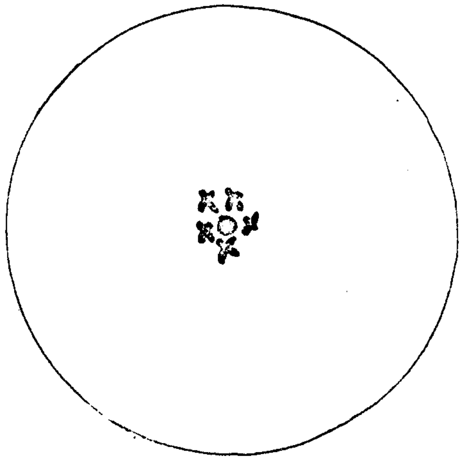
Accurate
Not Precise



Not Accurate (biased)
Precise



Not Accurate
Not Precise



Accurate
Precise

AS RELATED TO REMOTE SENSING:

- PRECISION
  - DOES NOT ALWAYS NEED TO BE CONSIDERED
  - NEEDS TO BE CONSIDERED WHEN A SAMPLE IS TAKEN FOR AREA ESTIMATE
- ACCURACY
  - ALWAYS NEEDS TO BE CONSIDERED
  - ASSESS WITH TEST FIELDS

IMPACT OF OBJECTIVE ON EVALUATION

- MAP OF COVER TYPES
  - FEW ERRORS IN IDENTIFICATION AND LOCATION
  - EVALUATE ERROR RATE
- AREA OR PROPORTION ESTIMATION
  - UNBIASED ESTIMATES
  - PRECISE ESTIMATES
  - CLASSIFICATION MUST ALSO BE ACCURATE IN THE SENSE OF "LOW ERROR RATE"

ASSESSING CLASSIFICATION AND MAP ACCURACY

- COMPLETE EVALUATION OF SCENE
  - REQUIRES "WALL-TO-WALL" "GROUND-TRUTH" WHICH IS RARELY AVAILABLE
  - IF IT IS AVAILABLE, WHY IS THE AREA BEING MAPPED?
    - AGE OF REFERENCE DATA
    - SCALE OR TYPE OF REFERENCE DATA

ASSESSING CLASSIFICATION AND MAP ACCURACY

- EVALUATION OF SAMPLE OF SCENE
  - TWO TYPES ARE COMMONLY USED:
    - TRAINING SAMPLES (FIELDS)
    - TEST SAMPLES (FIELDS)

- TRAINING SAMPLES
  - ARE USUALLY AVAILABLE
  - GIVE A BIASED ESTIMATE OF ACCURACY

ASSESSING CLASSIFICATION AND MAP ACCURACY

- TEST SAMPLES CAN FORM A BASE FOR STATISTICAL EVALUATION IF:
  - TEST SAMPLE IS OF SUFFICIENT SIZE
  - TEST SAMPLE REPRESENTS ALL VARIATION PRESENT IN AREA
  - PROBABILITY SAMPLING IS USED

METHODS OF CHOOSING TEST SAMPLES

- ANALYST SELECTED
- RANDOM SELECTION OF TEST SAMPLES
  - TEST AREAS OR SECTIONS
  - HOMOGENEOUS CELLS
  - GRID INTERSECTIONS

ANALYST SELECTED TEST FIELDS

- "HALF TRAIN, HALF TEST"
- "TYPICAL" SET
- "RANGE OF VARIATION"

ANALYST SELECTED TEST SETS HAVE INHERENT BIAS. CARE MUST BE TAKEN TO ENABLE SUCH TEST SETS TO BE USED IN CALCULATING CONFIDENCE INTERVALS AND IN STATISTICAL TESTS.

RANDOM SELECTION OF TEST SAMPLES

- TEST AREAS OR SECTIONS
  - PROCEDURE:
    1. THE AREA FOR WHICH REFERENCE DATA IS AVAILABLE IS DIVIDED INTO SUBAREAS, SAY SECTIONS.
    2. A RANDOM SAMPLE OF THE SECTIONS IS CHOSEN.
    3. ALL FIELDS IN THE SECTIONS ARE USED AS TEST FIELDS.
  - A FORM OF "CLUSTER SAMPLING"
  - EXAMPLE: CITARS

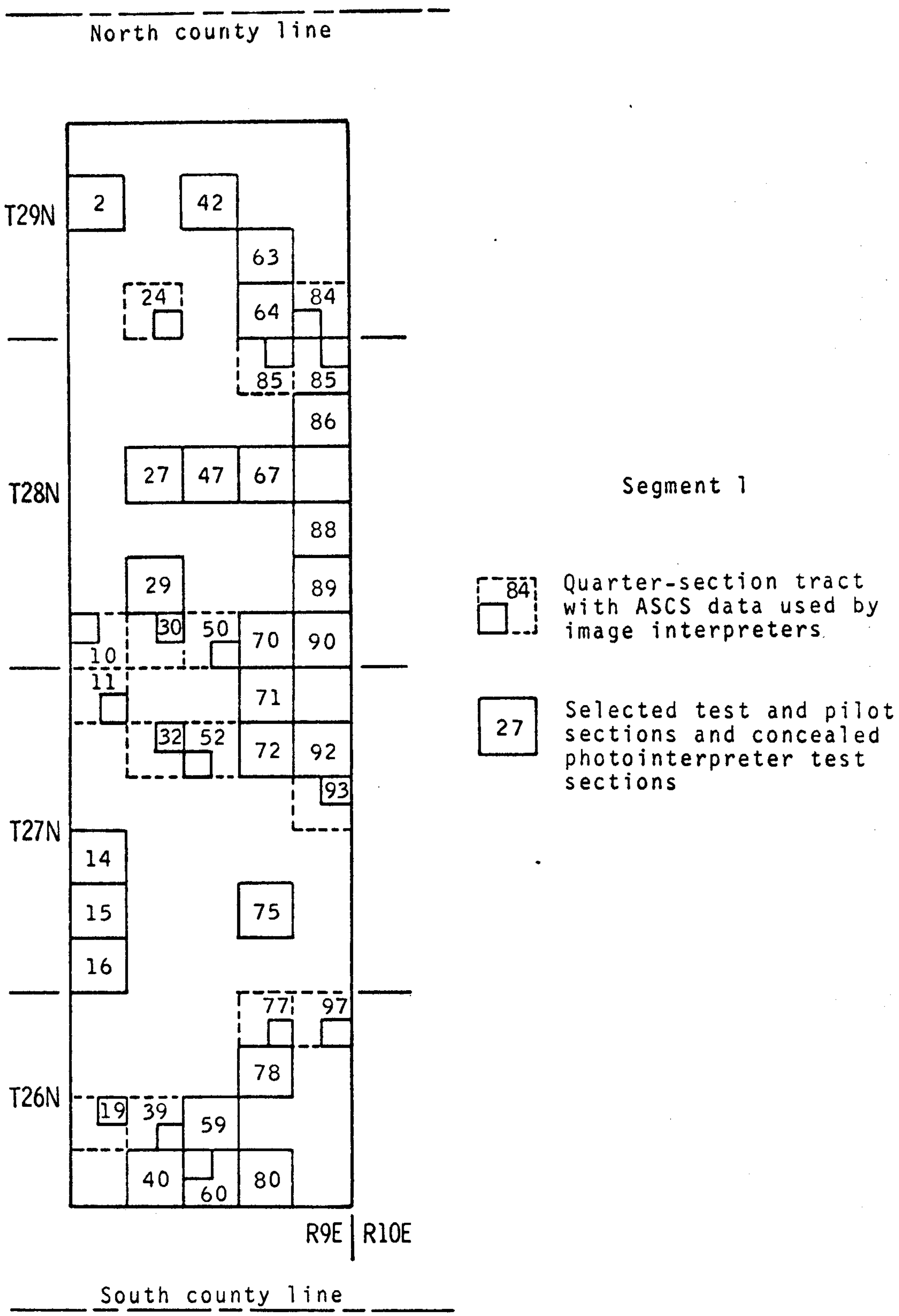


Figure D-1.— Idealized sketch of Huntington County test segment.

RANDOM SELECTION OF TEST SAMPLES

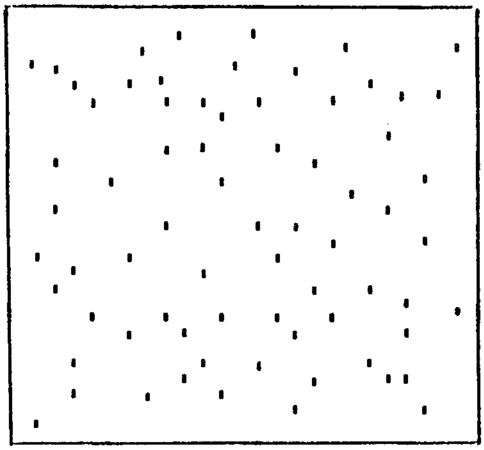
- TEST AREAS OR SECTIONS
  - ADVANTAGES:

    SAMPLING SCHEME IS EASY TO SET UP AND CARRY OUT. GROUND CHECKING OR PHOTO-INTERPRETATION CAN BE USED EFFECTIVELY.
  - DISADVANTAGES:

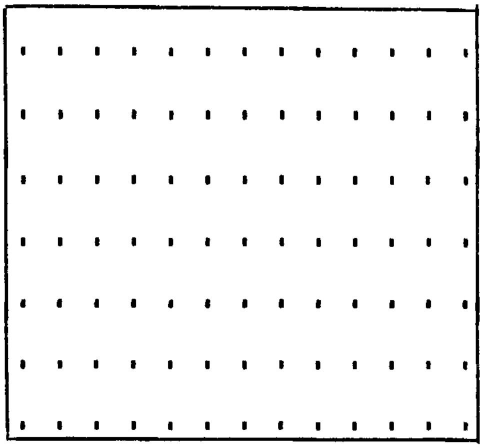
    IF THE SECTIONS ARE LARGE RELATIVE TO THE ENTIRE AREA, THE TEST SET MAY NOT BE REPRESENTATIVE.

RANDOM SELECTION OF TEST SAMPLES

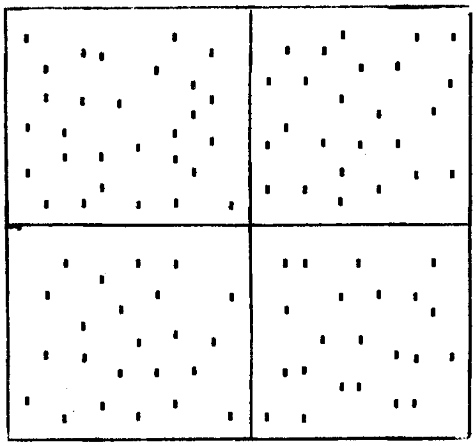
- HOMOGENEOUS CELLS
  - PROCEDURE:
    1. A GRID IS ALIGNED OVER THE ENTIRE AREA, WITH EACH BLOCK IN THE GRID AS ONE SAMPLE UNIT.
    2. A RANDOM SAMPLE OF BLOCKS IS CHOSEN.
    3. EACH BLOCK IN THE SAMPLE IS IDENTIFIED; NON-HOMOGENEOUS BLOCKS ARE REMOVED.
  - EXAMPLES: FORESTRY APPLICATIONS



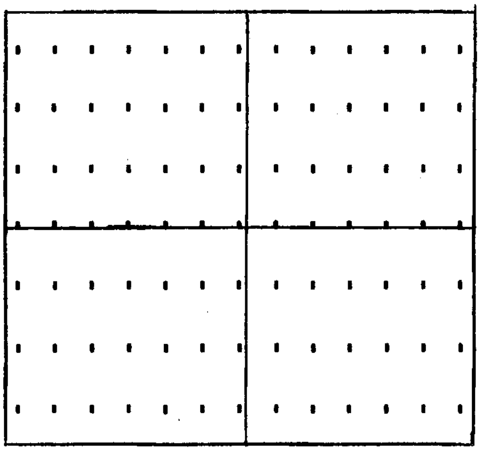
1. SIMPLE



2. SYSTEMATIC



3. STRATIFIED SIMPLE



4. STRATIFIED SYSTEMATIC

5. ANALYST SELECTED

Th2-

RANDOM SELECTION OF TEST SAMPLES

- HOMOGENEOUS CELLS
  - ADVANTAGES:

    SAMPLES ARE UNBIASED, REPRESENTATIVE, AND HOMOGENEOUS.
  - DISADVANTAGES:

    TEST SAMPLE SELECTION AND LABELLING MORE DIFFICULT AND TIME CONSUMING. THE REQUIREMENT FOR HOMOGENEITY OF A BLOCK REDUCES THE SAMPLE SIZE.

RANDOM SELECTION OF TEST SAMPLES

- GRID INTERSECTIONS
  - PROCEDURE:
    1. A GRID IS PLACED OVER THE AREA FOR WHICH REFERENCE DATA IS AVAILABLE.
    2. A RANDOM SAMPLE OF THE GRID INTERSECTIONS IS CHOSEN.
    3. EACH GRID INTERSECTION IS LABELLED, USING ADDITIONAL INFORMATION FROM THE SURROUNDING POINTS.
  - EXAMPLE: LACIE PHASE III

RANDOM SELECTION OF TEST SAMPLES

- GRID INTERSECTIONS
  - ADVANTAGES:

    GREATER RELIABILITY OF IDENTIFICATION BY USE OF SURROUNDING PIXELS. SAMPLES ARE UNBIASED AND REPRESENTATIVE. THERE IS NO RESTRICTION ON OBJECT SIZE.

  - DISADVANTAGES:

    ACTUAL NUMBER OF POINTS IN SAMPLE IS SMALL. THIS PROCEDURE CAN BE TIME CONSUMING. QUESTIONS CAN ARISE CONCERNING MIXTURE OR BOUNDARY PIXELS.

IMPACT OF REFERENCE DATA QUALITY ON ACCURACY

- INACCURATE REFERENCE DATA WILL LEAD TO ERRORS IN MEASURING CLASSIFICATION ACCURACY.
- ERRORS IN REFERENCE DATA MAY EXIST DUE TO:
  - AGE OF THE REFERENCE DATA RELATIVE TO THE MSS DATA.
  - THE METHOD OF OBTAINING "GROUND TRUTH".
    - PHOTO-INTERPRETATION
    - IMAGE INTERPRETATION, OR
    - GROUND CHECKING.

IMPACT OF REFERENCE DATA QUALITY ON ACCURACY

- THE REFERENCE DATA MAY BE FROM THE PROPER DATA AND HAVE NO ERRORS IN IDENTIFICATION BUT STILL BE OF POOR QUALITY.
- THE GEOGRAPHIC LOCATION AND DISPERSION OF REFERENCE MAY LIMIT ITS USEFULNESS IN ASSESSING ACCURACY.

TEST CLASS PERFORMANCE

		CLASSIFIED AS:		
	TOTAL	CORN	SOYBEANS	OTHER
CORN	981	853	9	119
SOYBEANS	893	4	876	13
OTHER	1397	296	93	1008

AVERAGE PERFORMANCE: 85.8

OVERALL PERFORMANCE: 83.7

PERFORMANCE: CORN 853/981 = 87.0

SOYBEANS 876/893 = 98.1

OTHER 1008/1397 = 72.2

CONFIDENCE INTERVALS FOR CLASSIFICATION ACCURACIES

ACCURACY P IS DISTRIBUTED BINOMIALLY, AS A PIXEL IS EITHER CORRECTLY OR INCORRECTLY CLASSIFIED.

$P_T = \mathrm{ARCSIN}\sqrt{P}$ IS DISTRIBUTED NORMAL WITH A STANDARD DEVIATION $S_P = \sqrt{821/N}$, WHERE N IS THE NUMBER OF OBSERVATIONS.

A 95% CONFIDENCE INTERVAL FOR P_T IS

$$(P_T - t_{\infty,0.05} S_P \, , \, P_T + t_{\infty,0.05} S_P)$$

A 95% CONFIDENCE INTERVAL FOR P IS

$$([\mathrm{SIN}(P_T - t_{\infty,0.05} S_P)]^2 . [\mathrm{SIN}(P_T + t_{\infty,0.05} S_P)]^2)$$

Example of Accuracy Confidence Interval

Overall Performance 83.7

BASED ON 3271 TEST PIXELS

$P = 0.837$

$P_T = \text{ARCSIN } \sqrt{P} = 66.19$

WITH A STANDARD DEVIATION $\sqrt{821/3271} = 0.501$

A 95% CONFIDENCE INTERVAL FOR P_T IS

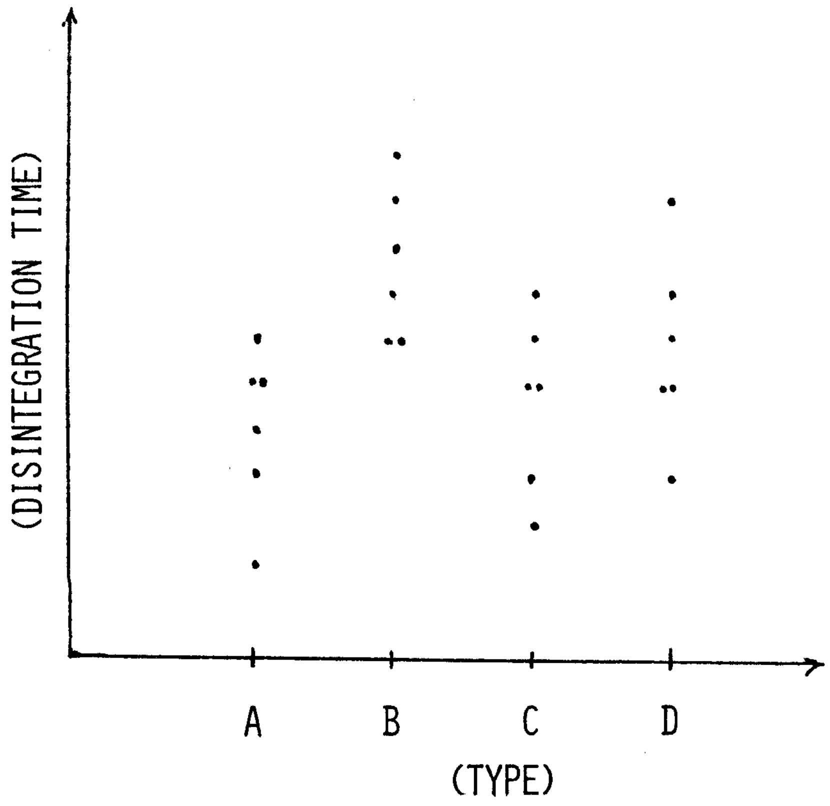
$$(66.19-(1.96)0.501, 66.19+(1.96)0.501)$$
$$(65.22 \, , \, 67.17)$$

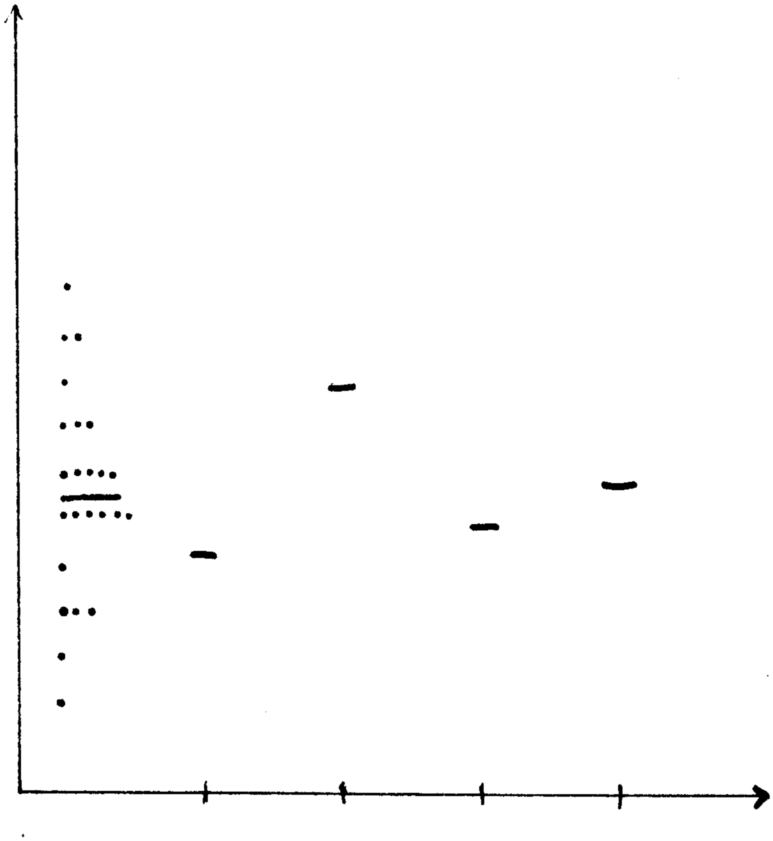
A 95% CONFIDENCE INTERVAL FOR P IS

$$(\, 82.4 \, , \, 84.9 \,)$$

ANALYSIS OF VARIANCE

- A PROCEDURE FOR PARTITIONING THE TOTAL VARIATION IN A SET OF DATA ACCORDING TO THE VARIOUS SOURCES OF VARIATION PRESENT.





SST = SSM + SSB + SSE

DF: 24 = 1 + 3 + 20

(SST-SSM) = SSB + SSE

EX: 1141.0-1027.0 = 46.5 + 67.5

114.0 = 46.5 + 67.5

ANOVA

Source	DF	SS	MS	F
Between	3	46.5	15.5	4.6
Within	20	67.5	3.4	
Total	23	114.0		

$F(3,20,.05) = 3.1$

Assumptions for ANOVA

- Normality of the observations
- Homogeneity of the variances of the sub-populations
- No restrictions on randomization in the design

April 13, 1978

The following material has been taken from:

> Landgrebe, D. A. 1976. Final Report,
> NASA Contract NAS9-14016. Laboratory
> for Applications of Remote Sensing,
> Purdue University, West Lafayette,
> Indiana 47906 (NAS9-14016, T-1039,
> MA-129TA).

Major contributors to this section were J. Berkebile, presently with the Indiana Department of Natural Resources, and R. Mroczynski, of LARS.

Evaluation of Classification Accuracy

The objective of this task was to develop procedures to statistically evaluate the accuracy of computer-assisted LANDSAT classifications. Results reported in "Analysis of Aircraft MSS Data for Timber Evaluation" (Mroczynski, et.al., 1976), and in the FAP final report for CY75 indicate the suitability of statistical evaluation of classification accuracy for large geographic areas. A necessary requirement for performing this evaluation is that the analyst have good ground reference data available. Without current ground reference data, either aerial photographs and/or detailed ground cover maps, the evaluation cannot be as meaningful.

Unfortunately, our experience has been that accurate ground information is not always available for forested test sites. Forest type maps generally indicate the age, size and merchantable status of a forest stand. These qualities may not relate to the stands spectral characteristics which are the basis for the statistical evaluation, so therefore may or may not be meaningful ground reference information.

Providing suitable ground reference information is available, a systematic sample grid can be aligned with the classification. The analyst then interprets these fields with the aid of the ground reference information. When the fields have been interpreted, various statistical tests can be applied to the results and inferences can be drawn about the classification accuracy for the entire site. This section will emphasize the statistical test procedures.

Figure 2.7-1 outlines the steps in the statistical evaluation procedure. An underlying assumption in this figure is that there exists a well-defined analysis objective. After obtaining the preliminary classification, the analyst should question how well it meets the analysis objectives. If the analyst is confident in his results there may be no need for further evaluation. Undoubtedly, the data analyst is not the ultimate user of the classification information. The statistical evaluation may be used to capture the user's confidence and therefore may be performed even if the analyst was confident in his results.

The essential steps in the evaluation are:

-- List the cover types to be tested

-- Locate potential test fields by systematic sampling with random start

-- Interpret and label the acceptable test fields

-- Perform the analysis of variance (including investigation of interaction for 2-factor Analysis of Variance) and range test calculations.

The tables which follow illustrate the Analysis of Variance procedures and are set up in the general form of 1) verbal description of step, and 2) numerical example. These tables are applicable in any situation by substituting values and performing the new evaluations based on the new values.

Analysis of Variance Tests

Remote sensing MSS data are fundamentally binomial in nature. Pixels in computer classifications are either identified correctly or incorrectly, hence their binomial distribution. As a result, the arcsin $\sqrt{p}$ transformation should be applied since, according to Steele and Torrie (1960):

> "The data can be transformed or measured on a new scale of measurement so that the transformed data are approximately normally distributed. Such transformations are also intended to make the means and variances independent, with the resulting variances homogeneous. This result is not always attained."

The importance of making the means and variances independent and the variances homogeneous is tied into the fact that these basic assumptions are made when performing an analysis of variance.

Regarding the effect of sample size upon homogeneity of variances, we recommend that at least 50 to 100 observations (one pixel = one observation) be obtained for each cover type to be tested. This removes the need for application of corrections to small sample sizes (50) as recommended by Snedecor and Cochran (1967). For sample sizes ranging from 50 to 1000, the comparisons among the percentage accuracy of cover type identification may be somewhat influenced by unequal variances. But for most studies an adjustment or weighting by the actual sample size is very seldom needed to obtain reasonably good comparisons. Especially if the range is from 100 to 500 samples, the assumption of homogeneity of variances is not usually violated enough to warrant a weighted transformation before running an ANOVA.

The main advantages of the angular (arcsin $\sqrt{p}$) transformation are that the error variance of the resulting observations (in degrees) is approximately constant, has infinity degrees of freedom (∞ df), and is equal to 821/n ("n"; sample size). The transformation is used as the unbiased estimator of the mean square error. Since sample sizes will vary among cover types, the harmonic mean (Table 2.7-7) which averages the different numbers of observations per accuracy mean, should be used according to Steele and Torrie.[3]

Two analyses of variance which will be encountered most often in LANDSAT applications are: 1) one-factor ANOVA, and 2) two-factor ANOVA. The one-factor ANOVA would be used to test for significant differences among cover types of a single classification or among overall accuracies of different classifications of the same data (e.g., wavelength band studies). The model for the one-way ANOVA (assuming transformed accuracy means) is:

$$Y_i = \mu + C_i + E(i)$$

where

Y_i = the overall accuracy (in degrees) of the i^{th} classification or cover type

μ = true overall accuracy mean

C_i = effect of the i^{th} classification or cover type

$E(i)$ = random error of the i^{th} classification or cover type

The best estimator of E(i) is assumed to 821/n (n = harmonic mean of sample sizes). This estimator is used as the mean square error (denominator, with infinity degrees of freedom) in the F test for significant variation (Tables 2.7-2 and 2.7-3).

A two-factor analysis of variance would entail testing for significant differences among different classifications of the same data set (e.g., wavelength band studies) and, at the same time, among cover types in the same data set. Thus, both factors (classification and cover type) are tested simultaneously. The model for a two-factor ANOVA is:

$$Y_{ij} = \mu + C_i + T_j + CT_{ij} + E(ij)$$

where

Y_{ij} = classification accuracy (in degrees) of the i^{th} classification for the j^{th} cover type

μ = true overall accuracy mean

C_i = effect of the i^{th} classification

T_j = effect of the j^{th} cover type

CT_{ij} = effect of the interaction between the i^{th} classification and the j^{th} cover type

E(ij) = random error, which is normally and independently distributed with mean = 0 and variance - σ^2

If the interaction effect, CT_{ij}, is found to be nonsignificant, E(ij) (which equals 821/n) provides the error mean square for the denominator of the F test. Again, it has infinity degrees freedom, thereby enabling a powerful F test.

Since interaction can occur between classifications and cover types, the interaction must be investigated to determine whether it is a "significant" source of variation. Essentially, an attempt is being made to find the best estimator of the error mean square for the F test. The interaction investigation proceeds as follows (and Table 2.7-6):

A) If (CT_{ij}/df)/821/n) is not significant at $\alpha = .25$ (F test), conclude there is no significant interaction and use 821/n for all F tests.

B) If it is significant at the .25 level, this may be due to the 821/n being too small because error other than binomial to normal is not included in 821/n. Hence, obtain the mean square for non-additivity with one degree of freedom from Anderson and McLean (1974).

   1) If the residual mean square is not significantly different at $\alpha = .25$, using 821/n as the denominator in the F, then use 821/n with infinity df for all tests.

   2) If the residual mean square is significant at $\alpha = .25$, use the residual mean square for all tests. This mean square has finite degrees freedom and provides a less powerful test than 821/n with infinity df.

The Newman-Keuls Range Test is an appropriate test for discerning which accuracy means are significantly different (Table 2.7-7). For this and preceding ANOVA's, we recommend that the level of significance be set at 90% (0.1 alpha). Thus, the tests will be quite "liberal" in the sense that if any significant differences exist, they will probably be detected. Theoretically, this means that the beta (β) error is made low (error of not detecting a significant difference when it truly exists) at the expense of raising the alpha error (error of denoting significant differences when not truly present).

Test

The one-factor analysis of variance test was applied to results obtained for the Sam Houston National Forest. Table 2.7-8 shows the calculations and the Newman Keuls Range Test for five different cover types. The conclusions drawn from this test is that at the 90% level there is no significant difference between classes one and four, or non-forest and pine. In other words, these classes would be difficult to separate, whereas the remaining classes would be more easily separable. However, the percent of the pixels correctly identified in each class is not high. One would expect better accuracy results considering the species composition of the area.

What this example typifies is a situation where sufficient ground reference data were not available for the analyst's evaluation. Given the information that was available only a limited number of systematic fields could be evaluated as indicated by the small number of pixels in each class. If a complete set of aerial photographs were available for the entire Sam Houston Test Site, the results of the statistical evaluation might be different.

REFERENCES

1. Snedocor, G. W. and Cochran, W. G. 1967. Statistical Methods, Chapter 11. Ames: Iowa State University Press, 593 pp.

2. Steele, R. G., and Torrie, J. H. 1960. Principles and Procedures of Statistics, Chapters 8 and 13. New York: McGraw-Hill, Inc., 481 pp.

3. Anderson, V. L., and McLean, R. A. 1974. Design of Experiments: A Realistic Approach, Chapter 2. New York: Marcel Dekker, Inc., 418 pp.

Th 2

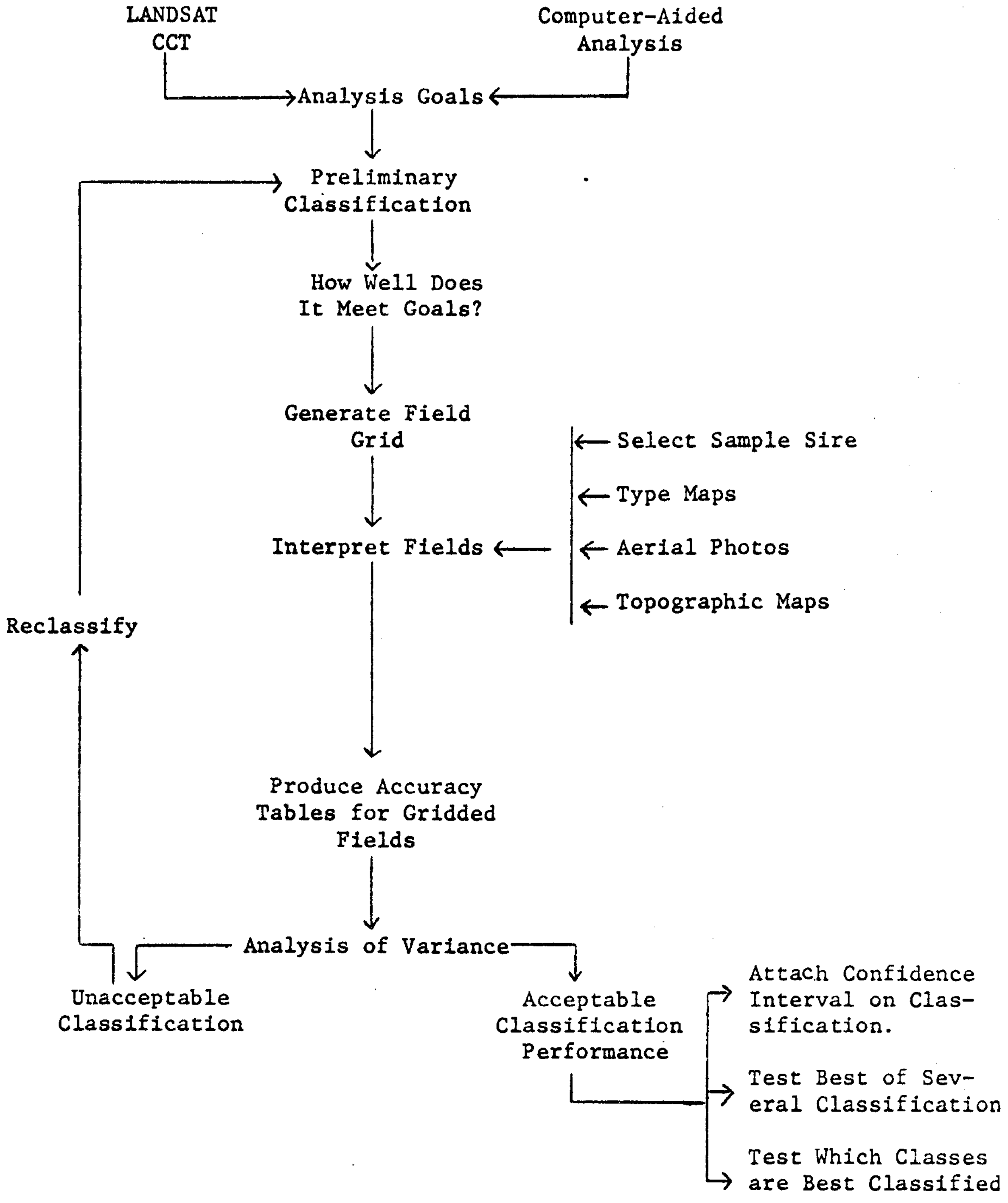


Figure 2.7-1 Steps in Implementing the Stage II Statistical Evaluation Technique

Table 2.7-2 One-factor analysis of variance and Newman-Keuls Range Test, three different cover types from the same computer classification.

PROCEDURE	EXAMPLE
Apply arcsin$\sqrt{p}^{*}$ transformation to cover type classification accuracies	COVER TYPE — NUMBER OF PIXELS — ACCURACY (%) — TRANSFORMATION Agricultural (A) — 150 — 88.7 — 70.3 Forest (F) — 70 — 81.4 — 64.5 Water (W) — 80 — 98.8 — 83.6
Calculate cover type sum of squares	$SS_T = [(70.3)^2 + (64.5)^2 + (83.6)^2]/1^{+} -$ $[(70.3 + 64.5 + 83.6)^2/3^{@}]$ $= 191.9$
Determine cover type mean square	Mean Square (i.e., MS_T) = SS_T/ (Number of means -1) $MS_T = 191.9/2 = 96.0$
Calculate F test and determine whether significant	A) $F = 96.0/[821/89.7^{**}] = 10.5$ (significant) B) Tabular $F_{2^{++},\infty} = 2.30$ (90% level, $\alpha = 0.1$)
Arrange transformed means in descending order	(W) (A) (F) 83.6 70.3 64.5
Calculate standard error of mean	$S\bar{y} = \sqrt{\text{error mean square/number obs. per mean}}$ $= \sqrt{[821/89.7]/1}$ $= 3.03$
Determine tabular ranges (Newman-Keuls Range Test)	Number-of-means range = (Studentized Range $_{(df=\infty)}{}^{*}$)($S\bar{y}$) 2 3 $R_3 = (2.902)(3.03) = 8.8$ $\alpha = 0.1$ .05=2.772 3.314 $R_2 = (2.326)(3.03) = 7.0$
Draw bars between means with ranges less than the corresponding tabular ranges	(W) (A) (F) 83.6 $\underline{70.3 \quad 64.5}$ (Hence, classification accuracy of water is significantly better than that for agriculture and forest)

*These values can be found in the Appendices of most statistics texts.

+Observations per accuracy mean = 1.

@Number of accuracy means.

**Harmonic mean = number of means/Σ(1/observations per mean)= 3/(1/150 + 1/70 + 1/80) = 89.7.

++Number of accuracy means -1 = degrees freedom.

Table 2.7-3 One-factor analysis of variance and Newman-Keuls Range Test, three different classifications of same data set.

PROCEDURE	EXAMPLE
Apply arcsin $\sqrt{p}$ transformation to overall classification accuracies	CLASSIFICATION — OVERALL ACCURACY (%) — TRANSFORMATION (DEGREES) 1 — 88.3 — 70.0 2 — 80.7 — 63.9 3 — 89.7 — 76.3
Calculate classification sum of square	$SS_c = [(70.0)^2 + (63.9)^2 + (71.3)^2]/1^* -$ $[(70.0 + 63.9 + 71.3)^2/3^+]$ $= 31.2$
Determine classification mean square	Mean Square (i.e., MS_c) = Sum square (i.e., SS_c)/(Number of means -1) $MS_c = 31.2/2 = 15.6$
Calculate F test and determine whether significant	A) $F = 15.6/ [821/300^@] = 5.7$ (significant) B) TABULAR $F_{2,\infty}^{++} = 2.30$ (90%)
Arrange transformed means in descending order	(3) (1) (2) 71.3 70.0 63.9
Calculate standard error of mean	$S\bar{y} = \sqrt{\text{error mean square/number observations per mean}}$ $= \sqrt{[821/300^@]/1}$ $= 1.65$
Determine tabular ranges	Number-of-means range = (Studentized range$)_{(df=\infty)}(S\bar{y})$ (90%) $R_3 = (2.902)(1.65) = 4.8$ $R_2 = (2.326)(1.65) = 3.8$
Draw bars between means with ranges less than the corresponding tabular ranges	(3) (1) (2) $\underline{71.3\ \ 70.0}$ 63.9 (Hence, the second classification is significantly different from the first and third.)

*Observations per accuracy mean = 1.

+Number of accuracy means.

@There are 300 test pixels (observations).

++Number of accuracy means -1 = degrees freedom.

Table 2.7-4 Arcsin-$\sqrt{\hat{p}}$ transformation of classification results. This transformation changes the binomial nature of these values (pixels are correctly or incorrectly identified) to a new scale of measurement so that the assumptions necessary for analysis of variance can be made. The transformed data should be approximately normally distributed, means and variances independent, and the resulting variances homogeneous (4).

CLASSIFICATION	COVER TYPE	NUMBER OF OBSERVATIONS	PERCENTAGE CORRECT CLASSIFICATION, $\hat{p}$	ARCSIN $\sqrt{\hat{p}}$ (degrees)
1	Agriculture (A)	150	86.7	68.6
	Forest (F)	70	85.7	67.8
	Water (W)	80	93.7	75.5
2	Agriculture	150	80.0	63.4
	Forest	70	74.2	59.5
	Water	80	87.5	69.3
3	Agriculture	150	88.6	70.3
	Forest	70	81.5	64.5
	Water	80	98.8	83.6

Table 2.7-5 Calculation of sums of squares.

<table>
<tr><th>CALCULATION CATEGORY</th><th>CALCULATION</th></tr>
<tr><td>Correction term, C.T.</td><td>$CT = (68.6 + 67.8 + \ldots + 83.6)^2/9^{*}$
$= 43{,}056.3$</td></tr>
<tr><td>Classification</td><td>$SS_c = [(68.6 + 67.8 + 75.5)^2 + (63.4 + 59.5 + 69.3)^2 + (70.3 + 64.5 + 83.6)^2]/3^{+} - C.T.$
$= 124.0$</td></tr>
<tr><td>Cover type</td><td>$SS_T = [(68.6 + 63.4 + 70.3)^2 + (67.8 + 59.5 + 64.5)^2 + (75.5 + 69.3 + 83.6)^2]/3^{@} - C.T.$
$= 236.7$</td></tr>
<tr><td>Total</td><td>$SS_{TOT} = [(68.6)^2 + (67.8)^2 + \ldots + (83.6)^2] - C.T.$
$= 43{,}456.9 - 43{,}056.3$
$= 400.6$</td></tr>
<tr><td>Interaction and/or error</td><td>$SS_{INT/ER} = SS_{TOT} - (SS_T + SS_c)$
$= 400.6 - (236.7 + 124.0)$
$= 39.9$</td></tr>
<tr><td>Set up ANOVA table</td><td>
<table>
<tr><th>source of variation</th><th>degrees of freedom (df)</th><th>sum of squares (SS)</th><th>mean square (SS/df)</th></tr>
<tr><td>Classification</td><td>(Classifications -1) = 2</td><td>124.0</td><td>62.0</td></tr>
<tr><td>Cover type</td><td>(Cover types -1) = 2</td><td>236.7</td><td>118.4</td></tr>
<tr><td>Interation and/or error</td><td>8-(2+2) = 4</td><td>39.9</td><td>10.0</td></tr>
<tr><td>TOTAL</td><td>Accuracy means-1 = 8</td><td>400.6</td><td></td></tr>
</table>
</td></tr>
</table>

* Number of means (cells).

\+ Number of cover types.

@ Number of classifications.

762-

Table 2.7-6 Investigation of Interaction.

<table>
<tr><th>PROCEDURE</th><th>EXAMPLE</th></tr>
<tr><td>A) Test for significance of interaction– STOP HERE if not significant–use 821/n for all tests</td><td>(75% level, $\alpha = .25$) F = Interaction and/or error mean square/[821/n*]
= 39.9/[821/89.7]
= 4.3 (significant)
tabular $F_{4^{+},\infty} = 1.35$</td></tr>
<tr><td>B) Obtain the mean square for Non-additivity with 1 degree of freedom</td><td>1) Produce following table:
<table>
<tr><th rowspan="2">CATEGORY</th><th colspan="3">COVER TYPES</th><th rowspan="2">ROW MEANS</th><th rowspan="2">ROW MEAN-OVERALL MEAN</th><th rowspan="2">(ROW MEAN-OVERALL MEAN)^2</th></tr>
<tr><th>AG</th><th>FOR</th><th>WAT</th></tr>
<tr><td>Cl. 1</td><td>68.6</td><td>67.8</td><td>75.5</td><td>70.63</td><td>1.46</td><td>2.13</td></tr>
<tr><td>Cl. 2</td><td>63.4</td><td>59.5</td><td>69.3</td><td>64.07</td><td>-5.10</td><td>26.01</td></tr>
<tr><td>Cl. 3</td><td>70.3</td><td>64.5</td><td>83.6</td><td>72.80</td><td>3.63</td><td>13.18</td></tr>
<tr><td>Col. $\bar{x}$</td><td>67.43</td><td>63.93</td><td>76.13</td><td colspan="2">$\bar{\bar{x}} = 69.17$</td><td>$\Sigma = 41.32$</td></tr>
<tr><td>Col. $\bar{x}$-overall $\bar{\bar{x}}$</td><td>-1.74</td><td>-5.24</td><td>6.96</td><td></td><td></td><td></td></tr>
<tr><td>(")^2</td><td>3.03</td><td>27.46</td><td>48.44</td><td colspan="2">$\Sigma = 78.93$</td><td></td></tr>
</table>
2) Calculate mean square (i.e., sum of squares/1)=
[(68.6)(1.46)(-1.74)+(67.8)(1.46)(-5.24)+(75.5)(1.46)(6.96)+ (63.4)(-5.1)(-1.74)+(59.5)(-5.1)(-5.24)+(69.3)(-5.1)(6.96)+ (70.3)(3.63)(-1.74)+(64.5)(3.63)(-5.24)+(83.6)(3.63)(6.96)]^2/ (41.32)(78.93) = 13.81</td></tr>
<tr><td>B1) Test residual mean square for significance– STOP HERE if not significant–use 821/n for all tests</td><td>1) MS_{res} = Interaction and/or Error SS – Non-additivity MS/ residual degrees freedom
=[39.9– 13.81] /3@
= 8.7
2) F = 8.7/[821/n*] (75% level)
= 8.7/9.2
= 0.94 (not significant)
tabular $F_{3^{@},\infty} = 1.37$</td></tr>
<tr><td>B2) Use residual mean square for all tests</td><td>$MS_{res} = 8.7$, df = $3^{@}$</td></tr>
</table>

*Harmonic mean = number of accuracy means/Σ(1/observations per mean) = 9/(1/150 + 1/150 + . . . + 1/80) = 89.7.

+Degrees of freedom for interaction and/or error mean square.

@Degrees of freedom for residual sum of squares = degrees freedom for interaction and/or error SS – 1.

7h2 –

Table 2.7-7 Tests for significance using best estimator of error mean square, as determined by investigation of interaction.

SOURCE OF VARIATION	DEGREES OF FREEDOM	SUM OF SQUARES	MEAN SQUARE	F TEST	TABULAR F (90%)
Classification	(Classifications- 1)= 2	124.0	62.0	(signif.) 6.7	$F_{2,\infty}$= 2.30
Cover Type	(Cover types- 1)= 2	236.7	118.4	(signif.) 12.9	$F_{2,\infty}$= 2.30
Error*	∞	——	827/89.7+ = 9.2		

---

*The residual mean square was found to be nonsignificant. If significant, the residual mean square would be used in place of 821/n for all tests.

+Harmonic mean.

Table 2.7-8 Newman-Keuls Range Test for classifications and cover types. Even if only one of two main factors is significant, the means may be kept separate for the range tests.

PROCEDURE	EXAMPLE
Arrange transformed means in descending order	(3-W)* (1-W) (3-A) (2-W) (1-A) (1-F) (3-F) (2-A) (2-F) 83.6 75.5 70.3 69.3 68.6 67.8 64.5 63.4 59.5
Calculate standard error of mean	$S\bar{y} = \sqrt{\text{error mean square/observations per cell}}$ $= \sqrt{9.2 / 1} = 3.03$
Determine tabular ranges	Number-of-means range = (Studentized range)$_{(df^{+}=\infty)}$ $(S\bar{y})$ (90% level, 0.1) R_9 = (4.037)(3.03) = 12.2 R_8 = (3.931)(3.03) = 11.9 R_7 = (3.808)(3.03) = 11.5 R_6 = (3.661)(3.03) = 11.1 R_5 = (3.478)(3.03) = 10.5 R_4 = (3.240)(3.03) = 9.8 R_3 = (2.902)(3.03) = 8.8 R_2 = (2.326)(3.03) = 7.0
Draw bars between means with ranges less than appropriate tabular ranges	83.6 75.5 70.3 69.3 68.6 67.8 64.5 63.4 59.5

---

*Accuracy mean for water of third classification.

+Same degrees freedom as the error (residual) MS. If the residual MS in Table 5 had been significant, the df for Studentized range would be 3.

Table 2.7-9

<table>
<tr><th>PROCEDURE</th><th>EXAMPLE</th></tr>
<tr><td>apply $\arcsin\sqrt{p}$</td><td>
<table>
<tr><th>COVER TYPE</th><th>NUMBER OF PIXELS</th><th>ACCURACY (%)</th><th>DEGREES</th></tr>
<tr><td>1 non-forest</td><td>144</td><td>69.4</td><td>56.42</td></tr>
<tr><td>2 hardwood</td><td>56</td><td>82.1</td><td>64.97</td></tr>
<tr><td>3 mix</td><td>72</td><td>34.7</td><td>36.09</td></tr>
<tr><td>4 pine</td><td>348</td><td>59.2</td><td>50.30</td></tr>
<tr><td>5 dense pine</td><td>180</td><td>22.2</td><td>28.11</td></tr>
</table>
</td></tr>
<tr><td>SS</td><td>$SS = [(56.42)^2 + (64.97)^2 + (36.09)^2 + (50.30)^2 + (28.11)^2]/1 - [(56.42 + \ldots + 28.11)/5]$
$= [3.83.2 + 4221.1 + 1302.5 + 2530.1 + 790.2]/1 - 11,128.8$
$= 12,027.1 - 11,128.8 = 898.3$</td></tr>
<tr><td>MS</td><td>$898.3/(5-1) = 224.6$</td></tr>
<tr><td>F Test</td><td>$F = 224.6/[821/(5/(1/144 + 1/56 + 1/72 + 1/348 + 1/180))]$
$= 224.6/[821/106.0]$
$= 29.0$ (significant)
tabular $F_{4,00} = 1.94$ (90% level, $\alpha = 0.1$)</td></tr>
<tr><td>$s_{\bar{y}}$</td><td>$s_{\bar{y}} = \sqrt{7.75/1}$
$= 2.78$</td></tr>
<tr><td>tabular ranges
(N-K)</td><td>$R_5 = (3.478)(2.78) = 9.67$
$R_4 = (3.240)(2.78) = 9.01$
$R_3 = (2.902)(2.76) = 8.07$
$R_2 = (2.326)(2.73) = 6.47$</td></tr>
<tr><td></td><td>(2) 64.97   (1) 56.42   (4) 50.30   (3) 36.09   (5) 28.11   (0.1 90%)</td></tr>
<tr><td></td><td>82.1% hardwood   69.4% non-forest   59.2% pine   34.7% mix   22.2% dense pine</td></tr>
</table>

7h3-

Purpose Of The Analysis: To find out if there is a significant difference in accuracy due to different methods of classification.

	Overall Accuracy	Transformed Value
Method 1	88.3	70.0
Method 2	80.7	63.9
Method 3	89.7	76.3

Source	DF	SS	MS	F
Method	2	31.2	15.6	5.7
Error			821/300	

where 300 is the total number of test pixels.

$F(2,\infty,.10) = 2.30$, so there is a significant effect of classification method on accuracy.

WHICH METHODS ARE DIFFERENT?

- ARRANGE THE TRANSFORMED MEANS IN DESCENDING ORDER

  76.3 70.0 63.9

- CALCULATE THE STANDARD ERROR OF THE MEAN

$$s_{\bar{Y}} = \frac{\text{ERROR MEAN SQUARE}}{\text{NO. OF OBSERVATIONS}} = 1.65$$

- DETERMINE TABULAR RANGES BY THE NEWMAN-KEULS PROCEDURE

- CONCLUSION: THE SECOND METHOD OF CLASSIFICATION IS SIGNIFICANTLY WORSE THAN THE OTHER TWO, BUT THERE IS NO SIGNIFICANT DIFFERENCE BETWEEN METHODS 1 AND 3.

ASSESSING CLASSIFICATION ACCURACY
WITHOUT TEST SAMPLES

- TRAINING SET ACCURACY
  - BIASED
  - DOES NOT MEASURE REPRESENTATIVENESS OF TRAINING

- PROBABILITY OF ERROR OVER THE AREA OF INTEREST
  - MAXIMUM LIKELIHOOD GAUSSIAN POINT CLASSIFIER CALCULATES A X^2 VALUE AFTER CLASSIFYING A POINT.

$$\text{PR}(X>C_{\alpha}|X\varepsilon\omega) = \alpha$$

IN LARSYS, A CODE IS STORED TO INDICATE THE VALUE OF α.

Th2

CALCULATION OF PROPORTION ESTIMATES
"CLASSIFY AND COUNT"

- AN ESTIMATE OF THE PROPORTION OF A CROP IN A COUNTY IS

$$\hat{P}_i = \frac{N_i}{N} = P_r(\hat{W}_i)$$

WHERE

N_i = NUMBER OF PIXELS CLASSIFIED AS CROP i

N = NUMBER OF PIXELS IN SAMPLE

$P_r(\hat{W}_i)$ = PROBABILITY THAT A PIXEL IS CLASSIFIED AS CLASS i

- THIS CAN BE REWRITTEN

$$\hat{P}_i = \sum_{j=1}^{k} P_r(\hat{W}_i | W_j)\ P_r(W_j)$$

WE REALLY WANT TO ESTIMATE $P(W_i)$, THE PROBABILITY THAT A PIXEL BELONGS TO CLASS i

"CLASSIFY AND COUNT" METHOD

- ADVANTAGES:
  - STRAIGHT FORWARD
  - IF A MAP IS REQUIRED, THIS ESTIMATE IS VERY EASY TO GET AND HAS NO SAMPLING ERROR.
- DISADVANTAGES:
  - BIASED ESTIMATE
  - ASSUMES MIXTURE PIXELS ARE CLASSIFIED INTO THE "PURE" CLASSES IN THE SAME PROPORTION AS THEY EXIST IN A MIXTURE PIXEL.

BIAS CORRECTION

- BIAS IN AREA OR PROPORTION ESTIMATES CAN BE REMOVED IF CLASSIFICATION ERROR RATES ARE KNOWN.
  - CLASSIFICATION OF KNOWN TEST FIELDS PROVIDES

$$E = \begin{array}{cc} & \begin{array}{cc} W & O \end{array} \\ \begin{array}{c} W \\ O \end{array} & \begin{pmatrix} e_{11} & e_{12} \\ e_{21} & e_{22} \end{pmatrix} \end{array}$$

  - IF P DENOTES THE VECTOR OF TRUE PROPORTION, THEN

$$\hat{P} = E^T P$$

$$\text{OR } P = (E^T)^{-1} \hat{P}$$

FOR A NUMERICAL EXAMPLE, CONSIDER CLOUD COUNTY, KANSAS

$$E = \begin{pmatrix} .85 & .15 \\ .18 & .82 \end{pmatrix}$$

$$E^T = \begin{pmatrix} .85 & .18 \\ .15 & .82 \end{pmatrix}$$

$$(E^T)^{-1} = \begin{pmatrix} 1.2239 & -.2687 \\ -.2239 & 1.2687 \end{pmatrix}$$

$$\hat{P} = \begin{pmatrix} 38.9 \\ 61.1 \end{pmatrix}$$

$$P = \begin{pmatrix} 1.2239 & -.2687 \\ -.2239 & 1.2687 \end{pmatrix} \begin{pmatrix} 38.9 \\ 61.1 \end{pmatrix} = \begin{pmatrix} 31.2 \\ 68.8 \end{pmatrix}$$

SRS HARVESTED	LANDSAT UNCORRECTED	LANDSAT CORRECTED
31.6	38.9	31.2

SOME ADDITIONAL EXAMPLES OF THE BIAS CORRECTION

COUNTY	SRS	LANDSAT UNCORRECTED	LANDSAT CORRECTED
MONTGOMERY	14.7	34.0	13.8
MCPHERSON	44.3	44.9	44.8
ALLEN	9.6	19.8	11.4

- This technique for removing bias has had mixed results.
  - Favorable: Landsat Crops - Kansas
    Northern Illinois
  - Unfavorable: CITARS
    Wabash Valley
    Landsat Crops - Indiana
- Keys to Successful Use:
  - High Overall Accuracy
  - High Individual Class Accuracy
  - Representative Test Fields
  - No Easily Confused Classes

CALCULATION OF PROPORTION ESTIMATES
"STRATIFIED AREAL ESTIMATE"

- ALL PIXELS CLASSIFIED INTO CLASS i ARE CONSIDERED TO FORM STRATUM i

- TEST SAMPLES ARE USED TO FIND

$$\alpha_{ij} = \frac{n_{ij}}{n_j}$$

WHERE n_{ij} = NUMBER OF TEST SAMPLES IN STRATUM j WHICH BELONG TO CLASS i

n_j = NUMBER OF TEST SAMPLES IN STRATUM j

- AN ESTIMATE OF THE PROPORTION OF CROP IS

$$\hat{P}_i = \sum_{j=1}^{k} \alpha_{ij} \frac{N_j}{N}$$

WHERE N_j= NUMBER OF PIXELS CLASSIFIED AS CROP j

N = NUMBER OF PIXELS IN AREA

- THIS CAN BE REWRITTEN:

$$\hat{P}_i = \sum_{j=1}^{k} P_r(W_i|\hat{W}_j)\ P_r(\hat{W}_j)$$

"STRATIFIED AREAL ESTIMATE" METHOD

- ADVANTAGES:
  - UNBIASED (IF NO ERROR IN TEST SAMPLES)
  - RELATIVELY EASY TO OBTAIN
- DISADVANTAGES:
  - MUST BE CAREFUL IN SELECTION OF TEST SAMPLES
  - ASSUMES TEST SAMPLES ARE PROPORTIONALLY ALLOCATED TO CLASSES

CALCULATION OF VARIANCE ESTIMATES

"STRATIFIED AREAL ESTIMATE"

- VARIANCE OF $\hat{P}_i$ FOR PROPORTIONAL ALLOCATION

$$\text{VAR}\ (\hat{P}_i) = \frac{1}{N}\left[\sum_{j=1}^{k} P_r(W_i|\hat{W}_j)[1-P_r(W_i|\hat{W}_j)]P_r(\hat{W}_j)\right]$$

CALCULATION OF VARIANCE ESTIMATES

- AS EACH PIXEL IS EITHER CROP i OR NOT, THE BINOMIAL DISTRIBUTION IS USED TO OBTAIN THE VARIANCE OF THE BIAS CORRECTED ESTIMATE

$$V(P_i) = \frac{P_i(1-P_i)}{N}(1-F)$$

WHERE F IS THE SAMPLING FRACTION.

- EMPIRICAL TESTS SHOWED THAT THE BINOMIAL VARIANCE IS NOT STATISTICALLY DIFFERENT FROM CALCULATED SAMPLE VARIANCES.
- IN SAMPLE FRACTIONS RANGING FROM 3 TO 50% THE STANDARD ERROR OF THE ESTIMATE RANGED FROM 0.5% TO 0.1%.

VARIANCE OF PROPORTION ESTIMATE

CLOUD COUNTY KANSAS

TOTAL	WHEAT	OTHER
21293	8283	13010

$\hat{P} = \begin{pmatrix} 38.9 \\ 61.1 \end{pmatrix}$ $\hat{P} = \begin{pmatrix} 31.2 \\ 68.8 \end{pmatrix}$

$$V(P) = \frac{(31.2)(68.8)}{21293} (.9375)$$

$$= 0.0945$$

WHEAT	31.2 ± 0.307 %
OTHER	68.8 ± 0.307 %

ASSESS THE ACCURACY OF AREA ESTIMATES BY COMPARISON TO REFERENCE DATA

- COMPUTE CORRELATIONS BETWEEN LANDSAT ESTIMATES AND REFERENCE DATA.
- FORMULATE A HYPOTHESIS ABOUT THE LANDSAT ESTIMATES AND REFERENCE DATA AND TEST THE HYPOTHESIS USING AN APPROPRIATE SIGNIFICANCE LEVEL (α-LEVEL).
- COMPUTE A CONFIDENCE INTERVAL ABOUT THE REFERENCE DATA ESTIMATE AND SEE WHETHER THE LANDSAT ESTIMATE FALLS IN THIS INTERVAL OR NOT.

Th 2

Testing Hypotheses With Reference Data

Hypotheses: $H_0: \mu_1 = \mu_2 \quad (\mu_1 - \mu_2 = 0)$

$H_1: \mu_1 = \mu_2$

Data:

Hectares of Corn

County	SRS	Landsat	Difference
Blackford	9.3	15.2	5.9
Delaware	27.2	43.9	16.7
Fayette	14.5	13.3	-1.2
Henry	29.3	23.8	-5.5
Jay	16.7	30.9	14.2
Randolph	28.1	49.0	21.0
Union	13.6	12.4	-1.2
Wayne	23.6	23.0	-0.6

Test statistic:

$$t = \frac{\bar{x}}{s/\sqrt{n}} = \frac{6.1}{9.9/\sqrt{8}} = 1.74$$

$$t_{(7,.25)} = 1.25$$

Conclusion: Reject H_0

CONFIDENCE INTERVALS FOR PROPORTIONS

SUPPOSE SRS ESTIMATE IS $\hat{P} = 36.9\%$ WITH A STANDARD ERROR OF 2%.

$$\Pr(\hat{P} - 1.96\ s < p < \hat{P} + 1.96\ s) = .95$$

$$\Pr(33.0 < p < 40.8) = .95$$

SINCE THE LANDSAT ESTIMATE $\hat{\hat{P}} = 38.9\%$ FALLS IN THIS INTERVAL, WE CAN SAY IT IS NOT SIGNIFICANTLY DIFFERENT FROM THE SRS $\hat{P}$ ESTIMATE AT THE 5% LEVEL.

FRIDAY
DATA TRANSFORMATION
AND
RESEARCH SURVEY

I. Data Transformation - Philip H. Swain

II. Research Survey - David A. Landgrebe

~~LARS Technical Rep~~ort 062076

LARS Technical Report 062076

Reference Papers for Short Course on Advanced Topics in the Analysis of Remote Sensing Data

Friday Session: Data Transformation and Research Survey

Decell, H.P. Jr, J.A. Quirein. (1973) “An Iterative Approach to the Feature Selection Problem”. Proceedings of 1973 Machine Processing of Remotely Sensed Data Symposium, Purdue University, West Lafayette, IN, October 16-18, 1973. https://docs.lib.purdue.edu/lars_symp/19/.

McMurtry, G.J. (1976) “Canonical Analysis as a Preprocessing and Feature Selection Method for Multispectral Data”. Presented at Second Workshop on Advanced Automation, University of Maryland, October 28-29, 1976. (Have pdf but not sure okay to make available online.)

DATA TRANSFORMATIONS

GEOMETRIC

- RECTIFY, ENHANCE OR CHANGE CHARACTERISTICS

RADIOMETRIC

- CALIBRATION (RELATIVE, ABSOLUTE)
- FEATURE EXTRACTION
- DIMENSIONALITY REDUCTION

EXAMPLES

- RATIOS
- LINEAR AND AFFINE TRANSFORMATIONS

RATIOS

RATIONALE

- NORMALIZATION
- EMPIRICALLY DETERMINED ENHANCEMENTS

OBJECTIVES

- REMOVE NON-TARGET INFLUENCES
- ENHANCE IDENTIFYING CHARACTERISTICS

NORMALIZATION - EXAMPLE

TARGET PROPERTIES: $p_1, p_2, \ldots, p_n$

MEASURABLE: $x_i = a(s,t)\, p_i, \quad i = 1, 2, \ldots, n$

UNWANTED COMMON GAIN FACTOR

FEATURES:

$$
\begin{aligned}
y_i &= x_i / x_{i+1} \qquad i = 1, 2, \ldots, n-1 \\
&= a p_i / a p_{i+1} \\
&= p_i / p_{i+1}
\end{aligned}
$$

depend on target <u>only</u>. (Note loss of dimensionality)

REMOVE COMMON BIAS BY USING DIFFERENCES

SIMPLIFIED RADIATION TRANSFER MODEL

$$R_{r_i} = \rho_i R_{s_i} T_i + b_i$$

where

R_{r_i} = radiation rec'd in band i

R_{s_i} = irradiation in band i

T_i = atmospheric transmissivity

b_i = backscatter & path radiance

ρ_i = object reflectivity in band i
ONLY TARGET CHARACTERISTIC

If: $R_{s_i} T_i \approx R_{s_{i+1}} T_{i+1}$ and $b_i \approx b_{i+1}$

then:

$$\frac{R_{r_1} - R_{r_2}}{R_{r_2} - R_{r_3}} = \frac{\rho_1 R_{s_1} T_1 + b_1 - \rho_2 R_{s_2} T_2 - b_2}{\rho_2 R_{s_2} T_2 + b_2 - \rho_3 R_{s_3} T_3 - b_3}$$

$$\cong \frac{\rho_1 - \rho_2}{\rho_2 - \rho_3}$$

a "purely" target characteristic

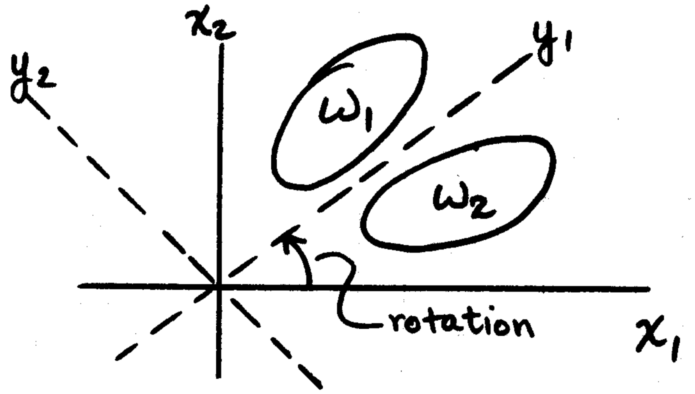
LINEAR TRANSFORMATIONS

$$Y = CX$$

$$y_1 = c_{11}x_1 + c_{12}x_2 + \cdots + c_{1n}x_n$$
$$y_2 = c_{21}x_1 + c_{22}x_2 + \cdots + c_{2n}x_n$$
$$\vdots$$
$$y_k = c_{k1}x_1 + c_{k2}x_2 + \cdots + c_{kn}x_n$$

(C is a $k \times n$ matrix)

y-axes are rotated:



PRINCIPAL COMPONENTS ANALYSIS

APPROACH

- COMPUTE THE SAMPLE COVARIANCE MATRIX FOR THE IMAGE DATA (BETWEEN CHANNELS)
- FIND THE LINEAR TRANSFORMATION $Y = CX$ SUCH THAT THE COORDINATE AXES IN THE TRANSFORMED SPACE ARE ORTHOGONAL AND THE COVARIANCE MATRIX IS DIAGONALIZED

COMPUTATION

- AN "EIGENPROBLEM"
- ORDER THE COORDINATES BY DECREASING EIGENVALUE, DELETE THE LEAST IMPORTANT

APPLICATION

- EMPHASIZES VARIABILITY AS A MEASURE OF INFORMATION CONTENT
- NOT SENSITIVE TO APPLICATION-SPECIFIC CHARACTERISTICS

AN ITERATIVE APPROACH TO FEATURE SELECTION *

IDEALIZED OBJECTIVE: DETERMINE LINEAR TRANSFORMATION $Y = BX$ FROM N-DIMENSIONAL SPACE TO K-DIMENSIONAL SPACE TO MINIMIZE PROBABILITY OF ERROR

ALTERNATIVE: OPTIMIZE AN INDIRECT MEASURE OF ERROR PROBABILITY, SUCH AS AVERAGE PAIRWISE DIVERGENCE BETWEEN CLASSES

* H. P. DECELL, J. A. QUEREIN : PROC. SYMP. MACHINE PROCESSING OF REMOTELY SENSED DATA, PURDUE UNIV., OCTOBER 1973.

Fig 7

Approach:

$$\omega_i \sim N(U_i, \Sigma_i), \quad i = 1, 2, \ldots, m$$

PMC: probability of misclassification (min.)

$$Y = BX: \quad \omega_i \sim N(BU_i, B\Sigma_i B^T)$$

$$PMC_B \geqslant PMC$$

Feature extraction ("selection"): For a given k, find a $k \times n$ matrix $\hat{B}$ of rank k such that $PMC_{\hat{B}} = \min PMC_B$

"Separability to be gained":

Let $D(i,j)$ = divergence between classes i & j in X space

$D_B(i,j)$ = divergence between classes i & j in Y space.

Then
$$D(i,j) \geq D_B(i,j)$$
or
$$D(i,j) - D_B(i,j) \geq 0$$
information is lost

measures "separation to be gained for classes i & j.

Let $D = \frac{2}{m(m-1)} \sum_{i=1}^{m-1} \sum_{j=i+1}^{m} D(i,j)$ ave. divergence

$D_B = \frac{2}{m(m-1)} \sum_{i=1}^{m-1} \sum_{j=i+1}^{m} D_B(i,j)$ B-ave. divergence

Then
$D_B \leq D$ for <u>every</u> $k \times n$ matrix B, $k = 1, 2, \ldots, n$.

<u>Theorem:</u> If $D = D_B$, then $PMC_B = PMC$

Central Theorem:

(i) Each $k \times n$ matrix of rank k maximizing D_B must satisfy

$$\left(\frac{\partial D_B}{\partial B}\right)^T = c \sum_{i=1}^{m} \left[S_i B^T - \Sigma_i B^T (B \Sigma_i B^T)^{-1} (B S_i B^T) \right] (B \Sigma_i B^T)^{-1} = (0)$$

where $c = \frac{2}{m(m-1)}$

$$S_i = \sum_{\substack{j=1 \\ j \neq i}}^{m} (\Sigma_i + \delta_{ij} \delta_{ij}^T)$$

$$\delta_{ij} = U_i - U_j$$

and $\frac{\partial D_B}{\partial B}$ represents the matrix

$$\left[\frac{\partial D_B}{\partial b_{ij}}\right] \qquad \begin{array}{l} i=1,2,\ldots,k \\ j=1,2,\ldots,n \end{array}$$

(ii) There exists a $k \times n$ matrix B of rank k which maximizes D_B.

Procedure:

1. For k = 1, 2, ... until $(D - D_B)$ is small, solve $\left(\frac{\partial D_B}{\partial B}\right) = (0)$ for $k \times n$ matrix B maximizing D_B.

2. Plot D_B / D versus k. Select k to trade computation (in classification) against D/D_B (measuring separation-to-be-gained)

CANONICAL ANALYSIS *

OBJECTIVE:

- DERIVE AN ORTHOGONAL LINEAR TRANSFORMATION TO MAXIMIZE SEPARABILITY AMONG CATEGORIES IN THE TRANSFORMED SPACE
- APPLY RESULT TO ACHIEVE DIMENSIONALITY REDUCTION WITH MINIMUM LOSS OF CLASSIFICATION ACCURACY

AN EIGENVECTOR METHOD WHICH MAXIMIZES DISCRIMINABILITY

* G. J. MC MURTRY: CANONICAL ANALYSIS AS A PREPROCESSING AND FEATURE SELECTION METHOD FOR MULTISPECTRAL DATA

Method:

1. Constraint on C: Let

$$W = \frac{1}{\left(\sum_{i=1}^{m} n_i\right) - m} \sum_{i=1}^{m} (n_i - 1) \Sigma_i$$

(within-set covariance matrix)

Require: $CWC^T = I$

Solve resulting eigenproblem.

2. Determine q based on
   - intrinsic dimensionality
   - computational demands.

3. Select components based on max. diagonal elements of CAC^T (variances of the transformed variables).

Approach:

$$\omega_i \sim N(U_i, \Sigma_i) \qquad i=1,2,\ldots,m$$

$Y = CX$, C a $q \times p$ transformation matrix

Let $\mathbb{U} = [U_1, U_2, \ldots, U_m]$

n_i = no. of observations in class i

$$N = \begin{bmatrix} n_1 & & & 0 \\ & n_2 & & \\ & & \ddots & \\ 0 & & & n_m \end{bmatrix}, \quad \underline{n} = \begin{bmatrix} n_1 \\ n_2 \\ \vdots \\ n_m \end{bmatrix}$$

$$A = \mathbb{U} N \mathbb{U}^T - \frac{1}{\sum_{i=1}^{m} n_i} (\mathbb{U}\underline{n})(\mathbb{U}\underline{n})^T$$

(<u>among</u>-classes covariance)

In the Y-space:

$$\omega_i \sim N[CU_i, C\Sigma_i C^T] \qquad i=1,2,\ldots,m$$

and the among-class covariance is CAC^T.

<u>FIND</u> matrix C to emphasize the differences among the categories.

Application:

1. Linear classifier

2. Evaluation of original channels

3. Dimensionality reduction

ITERATIVE METHOD

NUMERICAL OPTIMIZATION IS COSTLY IN TERMS OF COMPUTATION

OPTIMIZES AN ERROR-RELATED CRITERION

CANONICAL ANALYSIS

EIGENVALUE PROBLEM SOLUTION IS NONITERATIVE

OPTIMIZES A HEURISTIC GOODNESS CRITERION

LEADS TO A LINEAR CLASSIFIER

PROVIDES INFORMATION ABOUT ORIGINAL MEASUREMENTS

BOTH

USEFUL FOR DIMENSIONALITY REDUCTION

RESULT IN NORMALLY DISTRIBUTED VARIATES IF ORIGINAL DATA WAS NORMALLY DISTRIBUTED