07

77-10190
CR-153919
Vol. II

A FEASIBILITY ANALYSIS OF THE

EMPLOYMENT OF SATELLITE IMAGERY TO MONITOR

AND INSPECT SURFACE MINING OPERATIONS IN

WESTERN KENTUCKY

(E77-10190) A FEASIBILITY ANALYSIS OF THE EMPLOYMENT OF SATELLITE IMAGERY TO MONITOR AND INSPECT SURFACE MINING OPERATIONS IN WESTERN KENTUCKY. VOLUME 2: APPENDICES Final Report (Kentucky Dept. of Natural

N77-28555

HC A06/MF A01

Unclas

G3/43 00190

FINAL REPORT

VOLUME II - APPENDICES

Prepared for

National Aeronautics and Space Administration
Goddard Space Flight Center
Greenbelt, Maryland 20771

by

Birney R. Fish, Principal Investigator
Office of Planning and Research
Department for Natural Resources
and Environmental Protection
Commonwealth of Kentucky
Frankfort, Kentucky 40601

22640

January, 1977

TABLE OF CONTENTS

APPENDIX A

A MODEL OF ENVIRONMENTAL CONTROL SYSTEMS

In 1973 L. D. Maxim and D. E. Cullen of MATHEMATICA, Inc. presented at the 44th National ORSA meeting in San Diego a paper entitled: "A Model of Environmental Control Systems" (19). The article was written in anticipation of using the model, or a similar model, for developing an optimal inspection policy involving one or more of satellite, airplane, or surface inspections. The paper effectively illustrates the dependency of the optimal strategy on inspection costs, misclassification probabilities, and violation frequencies. This model was used as a base for the more elaborate model given in Volume I. The unabridged article is given in this appendix.

A MODEL OF ENVIRONMENT CONTROL SYSTEMS

Environmental Impact Study

1. An Environmental Problem

Environmental quality is today a major public policy issue which subsumes a complex of technical, economic, political, social, institutional and legal considerations. Broadly speaking, the goal of an environmental policy is the maintenance of the natural environment in a status which combines aesthetic values consistent with economic productivity in man's exploitation of his physical surroundings. But even if the goal were generally acceptable, the interpretation would vary widely between and among the numerous public and private interest groups in this country. In order to implement whatever interpretation was agreed upon, two activities must take place: environmental modeling and environmental control. That is, it is necessary to understand first the physical processes which determine the state of the environment, and secondly, the properties of alternative environmental control mechanisms.

This paper considers an environmental problem of national interest. Several bills have been introduced in Congress to regulate or abolish strip mining. The Hays bill, (Rep. Wayne Hays, D., Ohio), for example, would prohibit most such mining on slopes exceeding 20°. This bill was passed by the house in 1972 and has been reintroduced in 1973. The Hechler bill (Rep. Ken Hechler, D., W. Va.) would prohibit all surface mining within 6 to 18 months of passage. One state which would be greatly affected by any of these proposed laws is the Commonwealth of Kentucky.

Coal, particularly bituminous coal, is the major source of energy consumption in the United States today and is likely to remain so for many years to come. Kentucky was the largest producer of coal in the United States in 1972.[1/] The Commonwealth ranks third in recoverable bituminous coal reserves and also has large reserves of low sulphur coal. It is estimated that perhaps 70% of the total coal reserves in Eastern Kentucky are of this grade.[2/] Increasing national concern over atmospheric pollution has led many municipalities and cities to require low sulphur fuels for power generation.

It is therefore not surprising that mining is a highly important industry in Kentucky. This importance is particularly pronounced in regions such as the Appalachian area of Eastern Kentucky where coal mining ranks fourth in terms of total employment, accounting for 21% of the workforce and 27% of total wages.[3/] In 4 counties of the region, coal mining accounted for over 50% of the work force. Secondary and indirect employment in other industries, including services, transportation, trades and, to some extent, government, is highly dependent upon mining.

The public's attitude toward this industry is mixed because various forms of pollution and environmental consequences have attended coal production: sedimentation, slope failure, chemical pollution, revegetation difficulties, and aesthetic disturbance being major consequences of mining. There has been concern over these problems, and vigorous protest has been registered by environmental action groups, national newspapers, and many government agencies and coal associations.

Surface mining has been the chief focus of controversy because of its high visibility. More and more people are questioning the economic priorities of the past. Surface mines (strip and auger) accounted for 56% of total Kentucky coal production in 1971, up from about 39% in 1966.[4/] This mirrors a similar but somewhat less dramatic national trend toward surface mining. Reasons for this trend are not hard to find. National labor productivity of underground coal mines in 1967, for example, was only 15.07 tons per man-day relative to 35.17 tons per man-day and 46.48 tons per man-day for strip and auger mines respectively.[5/]

Thus, the promise and problems of surface mining have come sharply into focus. To strike a balance between economics, energy and environment is the central question facing both coal producing states and the nation at large. Such a balance will require an appropriate combination of legislation to control bad practices, of research and development to provide technology and improved operating practices, and of "enlightened self-interest" on the part of mine operators. Many of the surface mines experience a precarious existence which has precluded investment in research and development. Longitudinal studies over the period 1961-1962 suggest that perhaps 60% of the firms in Eastern Kentucky failed to survive this two year interval.[6/] Since these companies often operate on small profit margins, they also are not likely to be motivated to employ conservation practices which add to their costs but not to the price of coal. It has thus become the role of the state to enforce standards of operation upon the companies.

The Commonwealth of Kentucky has imposed several laws to reduce environmental disruption by surface mining. Historically, the laws have been enforced by inspectors who periodically visit each mine. Advances in aerial photography, however, now facilitate aerial inspection for detection of slides, revegetation failures, unauthorized mining operations and other prohibited activities.[10] Satellites, aircraft, or a combination of both may be utilized in a multi-tier system. Areas failing inspection are checked by ground inspectors. The following alternatives are considered in this paper:

1. Ground inspection
2. Satellite and ground inspection
3. Aircraft and ground inspection
4. Satellite, aircraft, and ground inspection

The objective of the work is to provide a framework for finding the most cost-effective inspection system and associated parameters.

2. Model Development

Let us assume that in an area to be inspected there are "N" sites at which coal is being surface mined. Prohibited activities are occurring at "N_1" of these sites, while at the others, the prescribed regulations are being met. The exact number and locations are, of course, unknown to the state authorities. The Commonwealth is responsible for insuring that proper mining practices are maintained and, therefore, needs to know the lowest cost way of performing the investigating activity.

It is assumed for illustrative purposes that if a man inspects a site, he will always make a correct determination of whether or not illegal activities are occurring. The cost of manned inspection, however, is high.

If inspection via remote sensing is used, either by spacecraft or aircraft, the possibility of misclassification arises. In this case there exist four possible outcomes:

(1) a site which is being mined by abusive practices is so identified (a "bad" site is correctly identified);

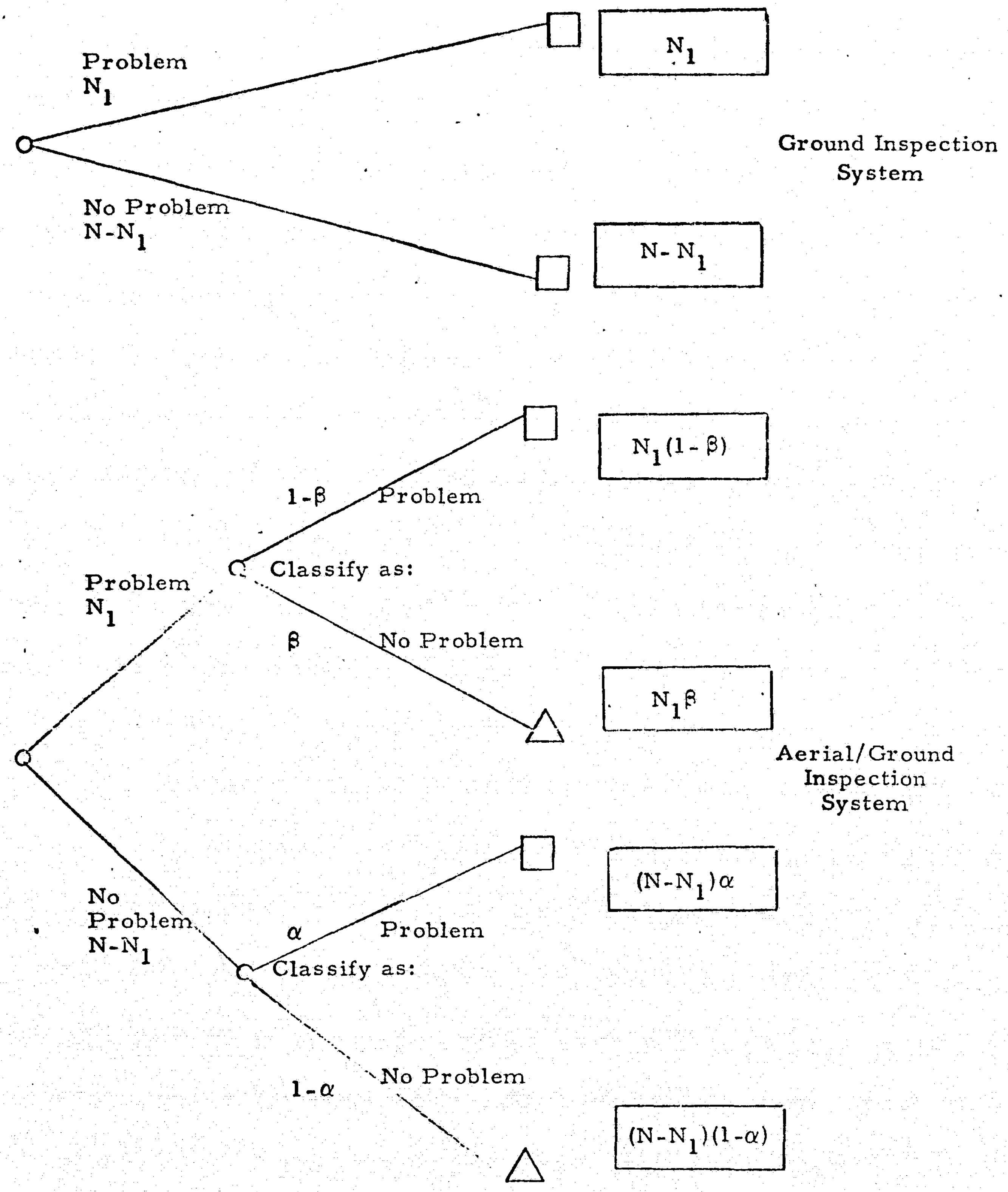
(2) a site which is being mined according to proper standards is so identified (a "good" site is correctly identified);

(3) a "bad" site is classified as "good"; and

(4) a "good" site is classified as "bad".

The last two possibilities are known as the beta "β" and alpha "α" errors respectively and are indigenous to any decision where there is less than complete information.

Figure 1 depicts the structures of inspection systems having a one-tier and a two-tier structure. The present ground inspection system has a one-tier structure. In this case all sites, whether problem areas or not, are surveyed by inspectors and consequently, are all correctly classified. Both the spacecraft/ground and aircraft/ground inspection systems have a two-tier structure. In these cases, however, there is a probability "β" that a decision rule depending on an aerial inspection will judge a problem area as a no-problem area, and a probability "α" that the rule will judge a no-problem area as a problem area. When we need to refer to quantities such as the misclassification probabilities in relation to either the satellite system or the aircraft system, we will subscript the quantities with an "s" or "a",



NOTE: Nodes represented by triangles indicate that no further inspection is conducted. Nodes represented by squares indicate manual inspection is conducted. The expected number of inspections in each category is enclosed in each bos.

Figure 1: Structures for Model Development, The One and Two-Tier Inspection System

respectively. Thus, for example, α_s will be the α error associated with the satellite inspection.

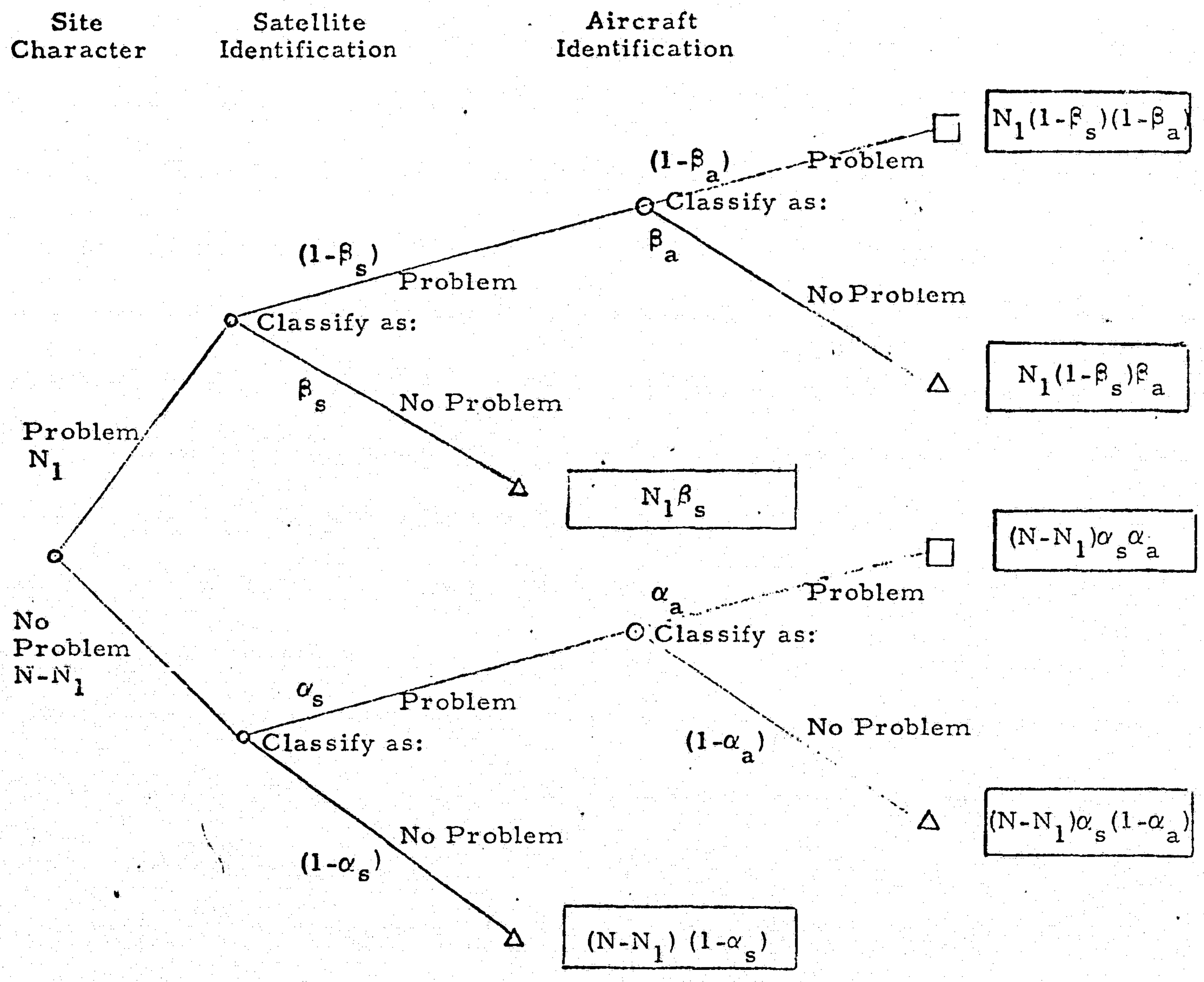
We will assume that a man is sent to check whenever the aerial inspection judges an area to be a problem area. A consequence of the alpha-error is that unnecessary manned inspection will occur $(N-N_1)\alpha$ times. The impact of the beta error is that $N_1\beta$ problem areas will go undetected. Both kinds of site misclassifications introduce associated cost penalties, the total magnitude of which is controllable through the alpha and beta risks of the remote sensing systems. The actual values of the error probabilities depend upon the technical characteristics of the remote sensing system and, consequently, the technology that is available. In the limit we might theoretically design and implement a remote sensing system that, like the ground inspection system, is subject to no errors and which would eliminate the α and β risks. Of course, the decision to implement such a system would depend on the costs, both non-recurring and recurring, that would have to be paid for such a system.

The structure of a three-tier inspection system is shown in Figure 2. The satellite/aircraft/ground inspection system has this structure. A decision rule provides that aircraft will be called in only after the spacecraft has classified an area as a problem area. The expected number of problem areas judged as no-problem areas will be

$$N_1\beta_s + N_1(1-\beta_s)\,\beta_a \qquad (1)$$

and

$$(N-N_1)\alpha_s\,\alpha_a \qquad (2)$$



NOTE: Nodes represented by triangles indicate that no further inspection is conducted. Nodes represented by squares indicate manual inspection is conducted. The expected number of inspections in each category is enclosed in each box.

Figure 2: Structure of a Three-Tier Inspection System

no problem areas will be misclassified as problem areas. Analogous to the two-tier systems, the attending misclassification penalty costs are unnecessary aircraft and manned inspection costs and the cost of undetected problem areas.

From the decision models shown in Figures 1 and 2, the cost functions presented in Figure 3 may be derived. It was assumed that satellite inspection represents a fixed cost if used; the incremental costs are regarded as zero. The aircraft inspection costs, as shown in Figure 3, may be derived directly from the structures in Figure 1 and 2, and as shown, depend on the decision model (i.e., whether or not aircraft inspection follows a determination by satellite inspection that there is a problem area). The number of "tiers," or combinations of inspection schemes, are provided for in the cost model by the binary variables X_s and X_a, their values depending on whether or not spacecraft and aircraft systems are being used, respectively. If spacecraft are used, for example, then X_s would be one. If spacecraft are not used, then X_s would be zero.

The third cost factor, "false negatives," indicates the social cost of a beta-type error. It includes the social and economic cost of nondetection and, by implication, the non-correction of a problem area. Some of the cost of nondetection results from the probability of physical damage, the value of which can be estimated. Other costs, however, are for non-market goods and activities. The values of these goods are difficult to determine and could theorectically range from zero to infinity depending upon the imputation of the social costs incurred due to misclassification.

Cost Factor	Value
1. Satellite Inspection	$C_s X_s$
2. Aircraft Inspection	$C_a \left[X_a(1-X_s)\left(N\right) + X_s X_a \left((N-N_1)\,\alpha_s + N_1(1-\beta_s) \right) \right]$
3. False Negatives	$C_p \left[X_s \left(N_1 \beta_s \right) + (1-X_s)(X_a)\left(N_1 \beta_a \right) + X_s X_a \left(N_1 (1-\beta_s)\,\beta_a \right) \right]$
4. Manual Inspection	$C_m \left[(1-X_s)(1-X_a)\left(N\right) + X_s(1-X_a)\left((N-N_1)\,\alpha_s + N_1(1-\beta_s) \right) + X_a(1-X_s)\left((N-N_1)\,\alpha_a + N_1(1-\beta_a) \right) + X_s X_a \left((N-N_1)\,\alpha_a\,\alpha_s + N_1(1-\beta_a)(1-\beta_s) \right) \right]$

WHERE

C_m = cost/site inspected manually.

C_s = cost of satellite inspection.

C_a = cost/site inspected by aircraft.

C_p = cost/problem area not detected.

α = probability "good" area is misclassified as problem area.

β = probability problem area is misclassified as good.

X_s, X_a = integer variables to denote whether satellite or aircraft inspection is used.

Figure 3: Composite Cost Function for Inspection Policies

In Figure 4, several sets of assumed values are given to the parameters discussed so far, and the alternative inspection policies are compared depending upon the values of the parameters. Policy 1 (P_1) assumes a man-only investigation and, therefore, with an assumed price of $50 per site, and a thousand sites, there is an invarient cost of $50,000 to investigate all of the locations. Policy 2 (P_2) assumes that ground investigation occurs only after it is determined by satellite that a site is a problem area. Policy 3 (P_3) assumes that ground inspection occurs only after it is determined by aircraft that a site is a problem area. Policy 4 (P_4) assumes that men are called in to investigate only after it is determined both by satellite and aircraft that a site is a problem area.

In Figure 4, the cost of implementing the four inspection plans are given under conditions of relatively high and low alpha and beta risks for aircraft and spacecraft. Holding all other parameters constant, it is seen that the costs, and consequently the choices, of the alternative inspection policies are very sensitive to the alpha and beta risks associated with aircraft and spacecraft. When the alpha risk is relatively high (20% as compared with 10%), then an increased cost would be incurred for re-inspecting sites which are, in fact, not problem areas. Also, there is a high likelihood of incurring the social cost of not detecting problem sites when the beta risk is relatively high. The asterisks in Figure 4 identify the optimal policies in each case. It is seen that even if the alpha and beta risks are relatively high, the three-tier and two-tier inspection systems are economically preferred over manual inspection only. The model demonstrates that remote sensing

DIRECT COMMON COST FACTORS PER SITE

Ground (men)	Satellite	Aircraft	Cost of Misclassification by β - type error
$C_m = 50$	$C_s = 200$	$C_a = 15$	$C_p = 2000$

	Input Factors						Cost of Survey ($1,000)			
			(%)							
Run No.	N	N_1	α_s	β_s	α_a	β_a	P_1	P_2	P_3	P_4
1	1000	10	20.	25.	5.	15.	50	15.5	20.9	11.3*
2	1000	50	20.	25.	5.	15.	50	36.6	34.5*	41.9
3	1000	100	20.	25.	5.	15.	50*	62.9	51.4	80.1
4	1000	10	10.	10.	5.	5.	50	7.6	18.9	5.4*
5	1000	50	10.	10.	5.	5.	50	17.2*	24.7	19.1
6	1000	100	10.	10.	5.	5.	50	29.2*	31.9	36.4

Costs of Optimal Choices are Denoted with an Asterisk(*)

N = Total Number of Sites

N_1 = Number of Defective Sites

α = Rate of Occurrence of α- Type Errors

β = Rate of Occurrence of β- Type Errors

$(\ )_a$ = ( ) for aircraft

$(\ )_s$ = ( ) for satellite

P_1 = Ground (men) only

P_2 = Satellite + Ground

P_3 = Aircraft + Ground

P_4 = Satellite + Aircraft + Ground

Figure 4: Illustrative Results With Simple Survey Model

systems can be useful even though they may be inaccurate. Whether this is true, in fact, will depend upon the particular application of the model and the inputs appropriate to the application. In general, if the number of problem areas, the social cost of misclassification, or the α and β errors are high, the optimal policy is ground inspection only. This results from the expectation of incurring substantial social costs for undetected problem sites. When the alpha and beta risks are relatively low and equal for aircraft and spacecraft systems, the policy P_2, a spacefraft/ground system, is preferred. This results from the fact that the spacecraft system costs are less than the aircraft system costs.

Figure 5 maps other information about the systems onto a graph in which the horizontal axis represents the parameter N_1, the number of defective areas, and the vertical axis represents the total cost of the alternative inspection programs. The value of the parameters other than N_1, are given in the top-half of Figure 4 in runs 1 through 3. The efficiency frontier that has been drawn indicates the lowest cost strategy as a function of the number of defective areas in the actual population. Any policy other than the one indicated for a given value of N_1 is inefficient from an economic standpoint. At values of N_1 less than 15, the three-tier plan, P_4, is the most cost-effective approach. Above that, up to about 39 defective areas, the man/spacecraft approach is the most cost-effective, from 40 to approximately 95, the aircraft/man plan is preferred and above 95, a man-only plan is the cost-effective approach. The shape of the efficiency frontier depends upon the value of the parameters. At the limiting case of C_p equal to infinity where no

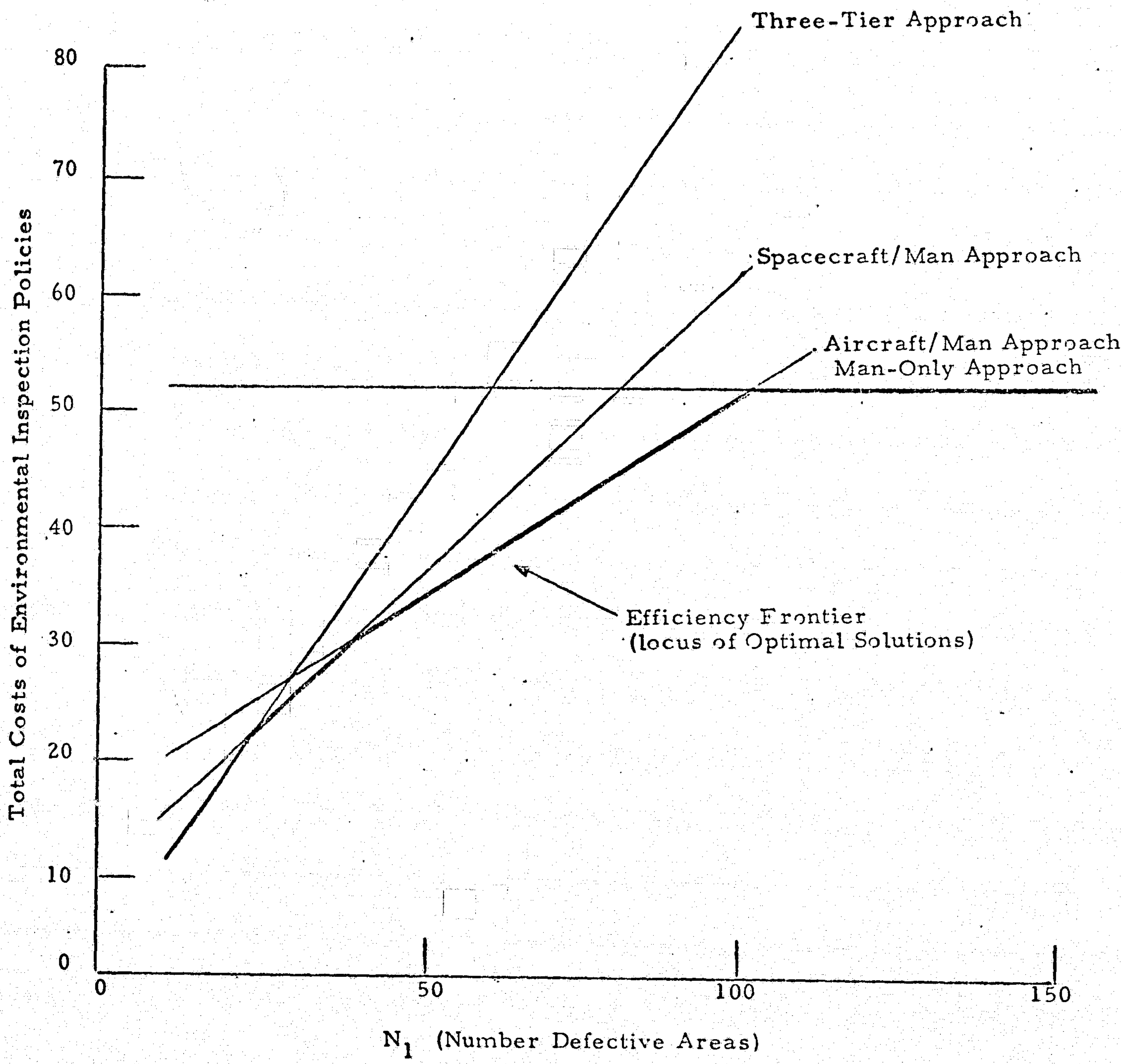


Figure 5: Efficiency Frontier For Environmental Inspection Policies

beta risks are tolerated, either a man-only system or an enhanced remote sensing system will be chosen, assuming that the technology is available to reduce β_a or β_s to zero. The choice would depend upon the relative costs of these systems.

We have seen that a simple model can be used to assess the economic impact of an important technical characteristic of remote sensing systems, the system accuracy on the selection of a cost-effective system. Another technical aspect of the remote sensing system which influences the choice of the most cost-effective inspection mode is that of system availability. This system characteristic is influenced by many factors, some of which are related to the system design and some of which are exogenous to the system, such as weather conditions. The potential impact of system availability on the choice of the economically optimum inspection mode can be determined by our model as is illustrated in Figure 6, a sample computer output. These results are based on the parameters used in run 5, shown in Figure 4.

The corner points of the cost grid map represent the four basic inspection alternatives under the assumption that the remote sensing systems are either never used or always used. For example, the man-only inspection system, having a cost of $50, 000, is represented by the grid point (aircraft, satellite) = (0, 0). In contrast, the two-tier inspection system, which calls for manual inspection of only those sites that have been classified as bad by a satellite, has a cost of $17, 200 and is represented by the grid point (aircraft, satellite) = (0, 1). By inspection of the corner points, one can readily verify that the two-tier satellite/man policy is the cheapest strategy of the four basic alternatives. Suppose, however, that we now consider the question of

N	=	1000	N1	=	50	IOPT	=	1			
ALPHA-S	=	0.100000	BETA-S	=	0.100000	ALPHA-A	=	0.050000	BETA-A	=	0.050000
COST-1	=	50.00	COST-2	=	200.00	COST-3	=	15.00	COST-4	=	2000.00

COST FOR BASIC ALTERNATIVES

MAN ONLY.	50000.
SATELLITE/MAN	17200.
AIRCRAFT/MAN.	24749.
SATELLITE/AIRCRAFT/MAN. . .	19174.

GRID MAP OF COSTS FOR MIXED POLICIES

SATELLITE. . .	0.0000	0.1000	0.2000	0.3000	0.4000	0.5000	0.6000	0.7000	0.8000	0.9000	1.0000
AIRCRAFT											
0.0000	50000.	46719.	43439.	40160.	36880.	33600.	30320.	27040.	23760.	20479.	17200.
0.1000	47475.	44467.	41459.	38451.	35443.	32436.	29428.	26420.	23412.	20405.	17397.
0.2000	44950.	42214.	39478.	36743.	34007.	31272.	28536.	25801.	23065.	20330.	17594.
0.3000	42425.	39961.	37498.	35035.	32571.	30108.	27645.	25182.	22718.	20255.	17792.
0.4000	39900.	37708.	35517.	33326.	31135.	28944.	26753.	24562.	22371.	20180.	17989.
0.5000	37375.	35456.	33537.	31618.	29699.	27781.	25862.	23943.	22024.	20106.	18187.
0.6000	34850.	33203.	31556.	29910.	28263.	26617.	24970.	23324.	21677.	20031.	18384.
0.7000	32324.	30950.	29576.	28202.	26827.	25453.	24079.	22705.	21330.	19956.	18582.
0.8000	29799.	28697.	27595.	26493.	25391.	24289.	23187.	22085.	20983.	19881.	18779.
0.9000	27274.	26445.	25615.	24785.	23955.	23126.	22296.	21466.	20636.	19807.	18977.
1.0000	24749.	24192.	23634.	23077.	22519.	21962.	21404.	20847.	20289.	19732.	19174.

Figure 6. Solutions for Partial Availabilities of Aircraft and Satellite Systems

satellite availability. If, for some reason, the satellite is available for site inspection only 80% of the time, then the actual cost of the satellite/man policy is not $17,200 but rather $23,760, corresponding to the grid point (aircraft, satellite) = (0, 0.80). The fact that the satellite is unavailable for some fraction of the site inspections markedly changes the cost of this policy and may render it a cost ineffective choice of inspection mode. For the data given in Figure 6, for example, the satellite/man inspection policy is cost effective only if satellite availability exceeds 90%. If the availability of the satellite is below this level, then the optimum inspection policy is the three-tier policy, and this remains true regardless of the availability of the aircraft remote sensing system. However, as is evident from the cost grid, the cost of implementing the three-tier system will not be $19,175 (as indicated by the grid point (aircraft, satellite) = (1, 1)) but instead will depend upon the availability actually achieved by the satellite and aircraft sensing systems. The cost model presented in Figure 4 allows for explicit consideration and evaluation of this primary technical system characteristic.

Figure 7 contains the result of a sensitivity analysis for run 2 of Figure 4 to explore the parameter ranges over which policy P_3 is optimal. For each parameter it shows the lower and upper limits and the policies which become optimal beyond the intervals. For example, the ground inspection cost can vary over a wide range from $35 to $203. Policy P_2 requires more ground inspections and, consequently, benefits more from a lower inspection cost. Conversely, policy

Base Case (Run 2):

Direct Common Cost Factors per Site

Ground (men)	Satellite	Aircraft	Cost of Misclassification by β type error
$C_m = 50$	$C_s = 200$	$C_a = 15$	$C_p = 2000$

Incidence Factors

Sites	Defective Sites
$N = 1000$	$N_1 = 50$

Misclassification Factors

Satellite		Aircraft	
$\alpha_s = 20\%$	$\beta_s = 25\%$	$\alpha_a = 05\%$	$\beta_a = 15\%$

Sensitivity Analysis:

Variable perturbed		Original Value	Range over which Policy P_3 is optimal		Optimal Policy at end of range	
Symbol	Name		Lower	Upper	Lower	Upper
C_m	Ground Cost	50	34.9091	202.8278	P_2	P_4
C_s	Satellite Cost	200	0	∞	-	-
C_a	Aircraft Cost	15	0	17.075	-	P_2
C_p	Misclassification Cost	2000	1585	4066.67	P_2	P_1
N_1	Defective Sites	50	38.9333	95.5882	P_2	P_1
α_s	Satellite α - error	20%	9.3684%	∞	P_2	-
β_s	Satellite β - error	25%	22.8718%	∞	P_2	-
α_a	Aircraft α - error	05 %	0	9.3684 %	-	P_2
β_a	Aircraft β - error	15%	0	17.1282 %	-	P_2

Figure 7: Illustrative Sensitivity Analysis

P_4 requires fewer men and suffers less from an increased cost. The satellite cost presents a different situation. Reducing the cost helps P_2 and P_4 but since at most only $200 can be saved, it is not sufficient to make either of these policies optimal. P_1 and P_3 are not dependent on the satellite cost so there is no change in their relative status and we see that P_2 is optimal over the full range of C_s. A similar review can be made for each of the other parameters, showing when and why each range limit and policy shift occurs.

3. Variations in Errors

Most systems can be altered so as to increase α-type errors while decreasing β-type errors or vice versa. In this system, the errors arise from misclassification. Changing the acceptance standards corresponds to changing the α and β errors. Hypothetical tradeoff curves for α_s, β_s and α_a, β_a are shown in Figure 8 .

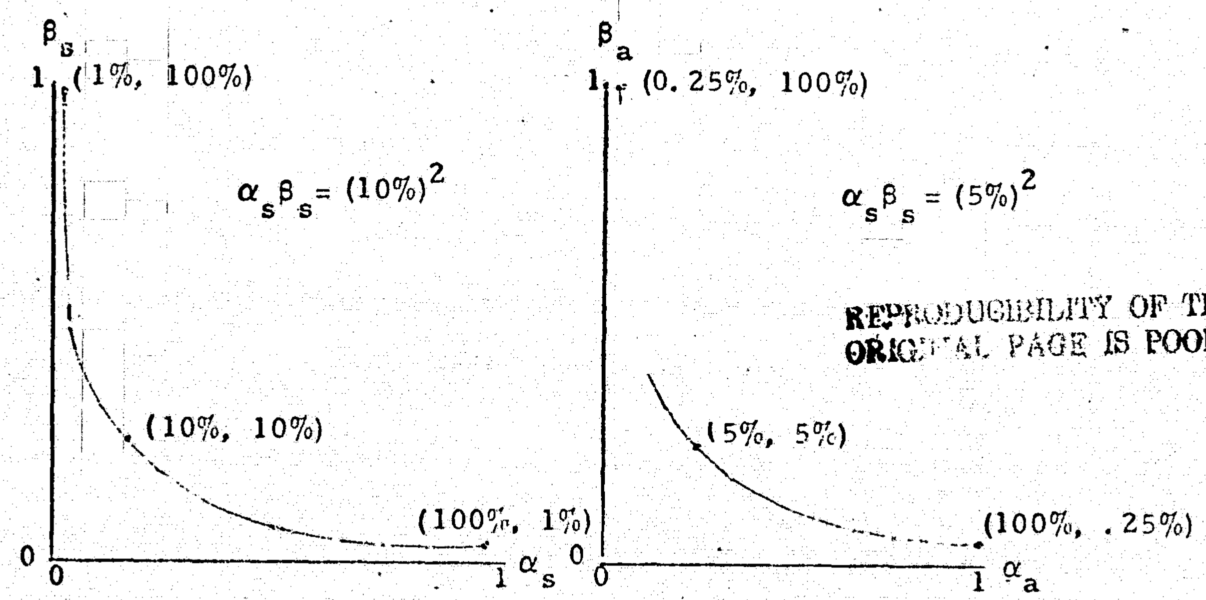


Figure 8 : Hypothetical Tradeoff Curves for Satellite and Aircraft α-type Errors Versus β-type Errors

They are hyperbolic curves with the property that reducing one probability by a factor of 50% results in doubling the other probability. The particular choice of this function is for illustrative purposes only--empirical tradeoff curves must be determined for each application.

This information combined with the earlier derived cost equations allows us to determine the optimal values of α and β to be used and consequently how to establish optimal acceptance criteria for the aircraft and satellite inspections. As an example, consider the cost expression for the satellite and man inspection system:

$$C_s + C_p N_1 \beta_s + C_m[(N-N_1)\alpha_s + N_1(1-\beta_s)] \qquad (3)$$

Using $\alpha_s \beta_s = (10\%)^2$ and rearranging terms, this expression becomes:

$$(C_s + C_m N_1) + (C_p - C_m) N_1 \beta_s + (10\%)^2 C_m (N-N_1)/\beta_s \qquad (4)$$

The optimal value of β_s is obtained by setting to zero the derivative of this expression with respect to β_s.

$$(C_p - C_m) N_1 - (10\%)^2 C_m (N-N_1)/\beta_s^2 = 0 \qquad (5)$$

Solving for β_s yields the optimal value, denoted β_s^* :

$$\beta_s^* = 10\% \, (C_m(N-N_1))^{1/2} \, ((C_p - C_m) N_1)^{-1/2} \qquad (6)$$

The corresponding value of α_s^* is:

$$\alpha_s^* = 10\%((C_p - C_m)N_1)^{1/2}(C_m(N-N_1))^{-1/2} \tag{7}$$

If these expressions result in either α_s^* or β_s^* being greater than one, then the correct solution is obtained by setting that probability to 100% and the other to 1%.

If C_p is less than C_m, the expressions for α^* and β^* become imaginary. This occurs if the penalty cost is less than the cost of manual inspection. In this case, $\beta_s^* = 100\%$, $\alpha_s^* = 1\%$ is the optimal solution.

A corresponding result for the aircraft and ground system can be obtained. In this case the expression for the cost is

$$(C_a N + C_m N_1) + (C_p - C_m)N_1\beta_a + (5\%)^2\, C_m(N-N_1)/\beta_a \tag{8}$$

The optimal values are:

$$\beta_a^* = 5\%\ (C_m(N-N_1))^{1/2}\ ((C_p - C_m)N_1)^{-1/2} \tag{9}$$

$$\alpha_a^* = 5\%\ ((C_p - C_m)N_1)^{1/2}\ (C_m(N-N_1))^{-1/2} \tag{10}$$

It will be noted that the expressions are the same except for the leading coefficients which are the square root of the constant term in the tradeoff curve.

This observation is a specified case of the general conclusion that for any two tier system, if the α and β type errors are related by the tradeoff curve $\alpha\beta = T^2$, then the optimal values are given by:

$$\beta^* = T(C_m(N-N_1))^{1/2}((C_p - C_m)N_1)^{-1/2} \quad (11)$$

$$\alpha^* = T((C_p - C_m)N_1)^{1/2}(C_m(N-N_1))^{-1/2} \quad (12)$$

If these expressions result in either α^* or β^* being greater than one, then the correct optimal solution is obtained by setting that probability to 100% and the other to T^{-2}. If C_p is less than C_m, the optimal solution is obtained by setting $\beta^* = 100\%$ and $\alpha^* = T^{-2}$.

Using the values of the parameters given in Figure 4, the optimal values for α^* and β^* for both satellite/ground and aircraft/ground systems are shown in Figure 9.

A similar analysis can be conducted for the three tier system. In this case a pair of simultaneous nonlinear equations is obtained which can be reduced to a single fourth order equation. The various cases resulting from the several roots of the equation and the interactions with the boundary conditions are too complex for presentation here but are obtained in a straight-forward manner.

Generally, the value of T can be decreased by the expenditure of more money. Increasing the time per aircraft inspection, for example, might produce such an improvement. Note that for the two tier system, the change in α^* and β^* is proportional to T. This is shown in Figure 10.

Direct common cost factors per site:

Ground (men)	Satellite	Aircraft	Cost of Misclassifications by β-type Errors
$C_m = 50$	$C_s = 200$	$C_a = 15$	$C_p = 2000$

Run No.	Input Factors				Optimal Error Factor (%)				Cost of Survey ($1,000)			Reduction in Cost (%)	
					Satellite/Grnd.		Aircrft/Grnd.						
	N	N_1	T_s * (%)	T_a *(%)	α_s	β_s	α_a	β_a	P_1	P_2	P_3	P_2	P_3
1	1000	10	$\sqrt{500} \approx 22.4$	$\sqrt{75.} \approx 8.7$	14.035	35.626	5.436	13.798	50	14.6	20.9	6.8	0.1
2	1000	50	22.4	8.7	32.036	15.607	12.408	6.045	50	33.1	29.3	9.5	15.1
3	1000	100	22.4	8.7	46.547	10.742	18.028	4.160	50	47.1	36.2	25.1	29.5
4	1000	10	10.	05.	6.276	15.933	3.138	7.962	50	6.9	18.6	9.0	1.6
5	1000	50	10.	5.	14.327	6.980	7.164	3.490	50	16.3	24.3	5.2	1.6
6	1000	100	10.	5.	20.817	4.804	10.408	12.402	50	23.9	29.4	18.0	7.9

Figure 9. Optimal α and β Type Errors

* These values of T_s and T_a correspond to those implicit in run 1 through 6 of Figure 4.

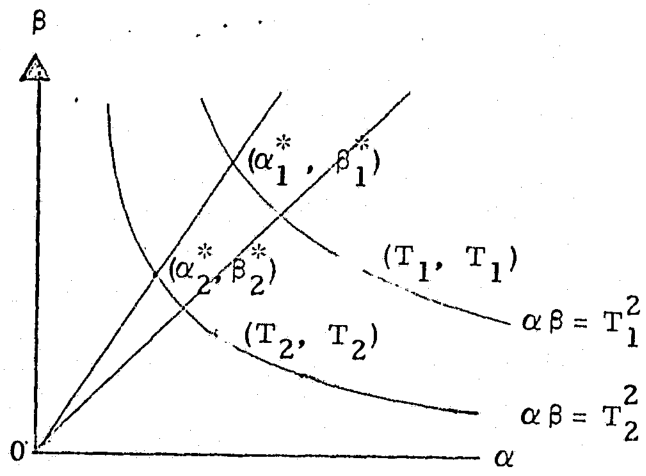


Figure 10. Change in optimal error terms for change in technological capability.

The cost of decreasing T generally rises nonlinearly as T approaches 0. Hypothetical cost curves for T are shown in Figure 11. Because changes in T often result from improvements in the technology used, these are known as technological cost curves.

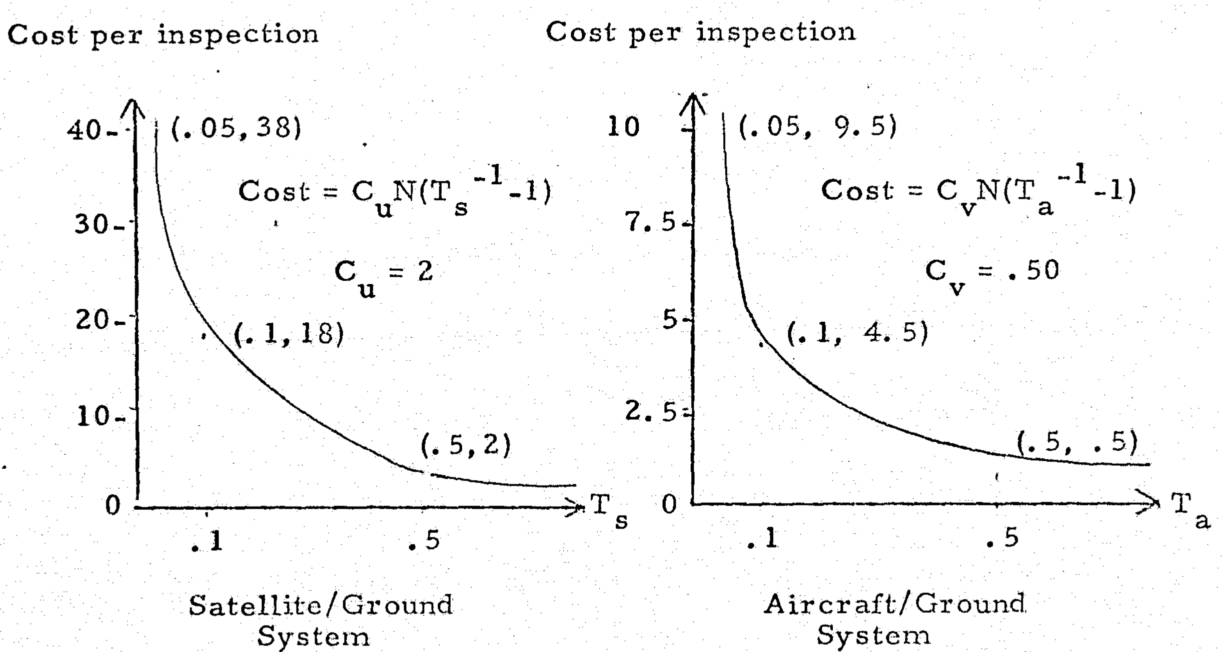


Figure 11. Hypothetical Technological Cost Curves

The cost of the satellite inspection system was given above as:

$$(C_s + C_m N_1) + (C_p - C_m) N_1 \beta_s + C_m (N - N_1) \alpha_s \tag{13}$$

where the optimal values of β_s and α_s are given by:

$$\beta_s^* = T_s (C_m (N-N_1))^{1/2} ((C_p - C_m) N_1)^{-1/2} \tag{14}$$

$$\alpha_s^* = T_s ((C_p - C_m) N_1)^{1/2} (C_m (N-N_1))^{-1/2} \tag{15}$$

Substituting and combining like terms yields the expression:

$$(C_s + C_m N_1) + 2T_s \ (C_m (C_p - C_m) (N-N_1) N_1)^{1/2} \tag{16}$$

In order to find the optimal value of T, we add the cost of technological improvement from Figure 11:

$$C_u N (T_s^{-1} - 1) \tag{17}$$

The sum of these two terms is then differentiated with respect to T_s yielding:

$$2(C_m (C_p - C_m) (N-N_1) N_1)^{1/2} = C_u N T_s^{-2} \tag{18}$$

The optimal value of T_s, denoted T_s^*, is thus:

$$T_s^* = (C_u N)^{1/2} (2C_m (C_p - C_m) (N-N_1)N_1)^{-1/4} \quad (19)$$

A similar analysis for the aircraft/ground system yields:

$$T_a = (C_v N)^{1/2} (2C_m (C_p - C_m) (N-N_1)N_1)^{-1/4} \quad (20)$$

The same approach may be used for the three tier system but is too complex for presentation here.

Using the data presented in Figure 4, the following selection of optimal T, α, and β values can be derived as shown in Figure 12 for the two two-tier systems.

4. Conclusions

It is anticipated that a model such as we have described can be very useful in determining the optimal strategy for alternative remote sensing systems since it incorporates cost, technology characteristics, econometric estimation, and public policy. The description given is for general model and individual specifications, of course, must be tailored to the application or case study to be investigated. As seen, the model is simple and yet elegant and powerful. The alpha and beta risks are technical questions and, therefore, allow us to parameterize the quality or accuracy of alternative remote sensing systems. In addition, the model allows us to parameterize the operational availability achieved by the remote sensing systems and examine the cost impact of this

Direct common cost factors per site:

Ground (men)	Satellite	Aircraft	Cost of Misclassification by β type error	Satellite Technological Cost	Aircraft Technological Cost
$C_m = 50$	$C_s = 200$	$C_a = 15$	$C_p = 2000$	$C_u = 2$	$C_v = .50$

Run No.	Input Factors		Optimal Parameters (%)						Cost of Survey ($1,000)			Reductions in Cost/Table *			
			Satellite/Ground			Aircraft/Ground						Runs 1, 2, 3		Runs 4, 5, 6	
	N	N_1	T_s	α_s	β_s	T_a	α_a	β_a	P_1	P_2	P_3	P_2	P_3	P_2	P_3
A	1000	10	17.9407	11.2604	28.5841	08.9704	05.6302	14.2921	50	11.8	21.1	44.9%	19.4%	52.4%	25.0%
B	1000	50	12.1220	17.3672	08.4609	06.0610	08.6836	04.2305	50	19.2	25.7	52.1%	25.5%	44.0%	23.8%
C	1000	100	10.3321	21.5080	04.9634	05.1661	10.7541	02.4817	50	24.6	29.7	54.6%	28.5%	41.4%	23.6%

* After adjustment for technological development cost not included there.

Figure 12. Optimal Technological and Error Parameters

important system characteristic. There are several relevant directions for further model development that are readily apparent:

- o introduction of a larger set of classification outcomes (i.e., "fuzzy results)
- o multiple inspection objectives
- o realistic cost functions for inspection techniques (e.g., fixed cost aspects)
- o dependence of alpha and beta errors upon the magnitude of a problem area
- o more realistic tradeoffs between α and β errors
- o budget constraints on inspection policies
- o more complex inspection policies (e.g., using random inspection of sites classified as no problem).

The potential of each of these factors to sharpen the analysis of, and thereby enhance, the study results may be determined by extending this model. As an illustration, a more complex ground inspection cost function is modelled in the Appendix, (A').

We wish to emphasize the important lessons that can be gleaned from this illustrative model:

1. simple models lend insight to the investigative process.
2. as our model has demonstrated, a satellite can be a cost-effective component of an information retrieval system

even though it may not be the most accurate and/or reliable component of the system.

3. model results can lead to profound changes in current eco-systems information retrieval and control practices.

APPENDIX A'

An Alternative Cost Model

The fundamental model can be extended in several ways to improve its accuracy. One such improvement can be made in representing the ground inspection costs. The agency responsible for inspection in general cannot alter its staff at will. It will in fact hire a number of inspectors for this purpose with the consequence that the cost of this staff will be fixed. To handle any additional inspections above what the staff can normally handle, the inspectors may be asked to work overtime and employees in other areas may be utilized under a part-time, temporary arrangement.

The cost relationships of this model can be defined in terms of the following parameters:

- M - the number of inspectors hired on a permanent basis
- θ - the number of inspections that can be conducted per inspector
- γ - the cost per inspector incurred in one period
- γ' - the cost per inspection for additional inspections above those that can be performed by the permanent staff.

If n inspections are required in a period, the cost is either γM or $\gamma M + \gamma'(n - \theta M)$ depending on whether n is less than or greater than θM, respectively. Mathematically this can be expressed as:

$$\text{personnel cost} = \gamma M + \gamma' \text{Max}[0, n - \theta M]$$

This is shown in Figure A-1.

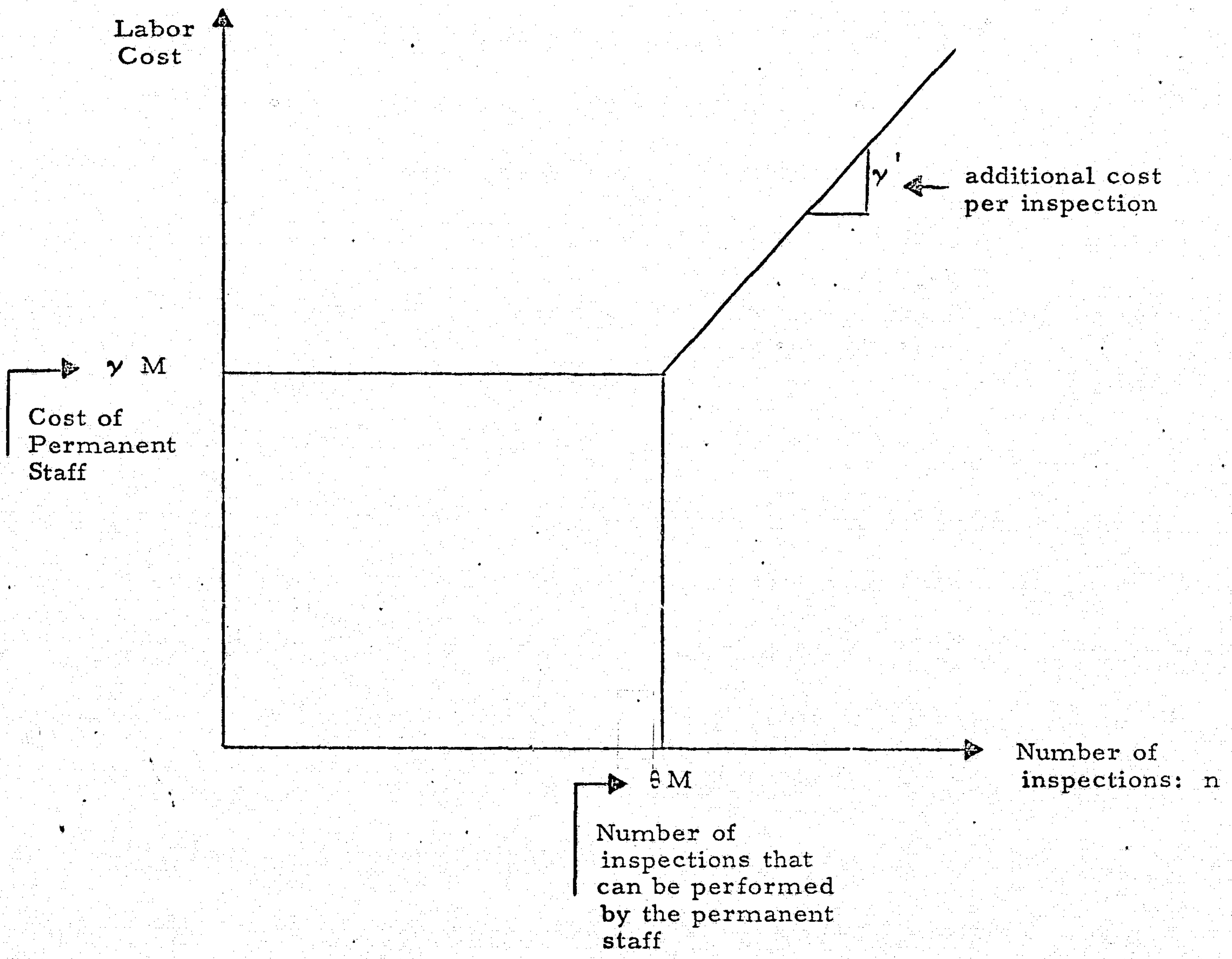


Figure A-1. Cost Model for Manual Inspections

In a two tier system, either aircraft and ground or satellite and ground, the number of inspections required is the sum of two quantities, the number of problem areas which are recognized as such, denoted n_1 and the number of non-problems which are identified as problem areas, denoted n_2. Both n_1 and n_2 are independent binomially distributed, random variables. Referring to Figure (A-2) the expected values of n_1 and n_2 are $N_1(1-\beta)$ and $(N-N_1)\alpha$ respectively. The respective variances are $N_1(1-\beta)\beta$ and $(N-N_1)\alpha(1-\alpha)$. The α and β errors are those associated with whichever two-tier system is under consideration. In practical applications, we may expect that the number of problem areas is small and that most inspections are conducted for non-problem areas. In this case, n_1 can be disregarded. For a large number of required inspections, the normal distribution provides a satisfactory approximation to the binomial distribution. Hald's[7/] inequality, $np(1-p) > 9$, provides a definition of the acceptable range for the approximation. For n_2, this is

$$(N - N_1)\alpha(1-\alpha) > 9. \qquad \text{(A1)}$$

Since $(1-\alpha)$ may be assumed to be greater than 0.5, the approximation will be valid for

$$(N - N_1)\alpha > 18. \qquad \text{(A2)}$$

The expression on the left, of course, is simply the expected number of required inspections. The use of the normal approximation permits us to develop an analytic expression for the total cost and to find the size of the permanent staff which minimizes this quantity.

The total cost, including the personnel cost, is:

$$T = C_s + (N-n_1)\, C_p + \gamma M + \gamma' \,\mathrm{Min}\,[0,\ n-\theta M] \qquad \text{(A3)}$$

The expected value of the total cost is:

$$E[T] = C_s + \beta N, C_p + \gamma M + \gamma' \int_{\theta M}^{\infty} (n-\theta M)\, p(n)\, dn \qquad \text{(A4)}$$

where p(n) represents a normal distribution with the parameters $\mu = \alpha(N-N_1)$ and $\sigma^2 = \alpha(1-\alpha)(N-N_1)$. The integral in this expression, known as the "partial expectation" does not have a closed form expression. It is tabulated in such sources as Brown,[8/] in Table D. 6.

For the values of θM in the range between the mean and the mean plus two times the standard deviation, Parker's[9/] service function approximation may be used. Mathematically this gives:

if $\mu \leq \theta M \leq \mu + 2\sigma$

then $$\int_{\theta M}^{\infty} (n-\theta M)\, p(n)\, dn = 0.45\sigma \exp(-(\theta M-\mu)/.60\sigma) \qquad \text{(A5)}$$

where $\mu = \alpha(N-N_1)$ and $\sigma = \alpha(1-\alpha)(N-N_1)$ (A6)

Substituting into the formula for the expected total cost, we obtain:

$$E[T] = C_s + \beta N_1 C_p + \gamma M + \gamma' \sigma .45 \exp(-(\theta M - \mu / .60\sigma)) \quad \text{(A7)}$$

In order to determine the optimal permanent staff, the derivative of this expression with respect to M is set to zero.

$$0 = \frac{\partial E[T]}{\partial M} = \gamma + \gamma' \sigma .45 \exp(-(\theta M - \mu / .60\sigma))(-\frac{\theta}{.60\sigma}) \quad \text{(A8)}$$

Solving for the optimal value of M, say M^*, we find:

$$M^* = \theta^{-1}[\mu - .60\sigma \, \ell n(.60\gamma/(.45\gamma'\theta))] \quad \text{(A9)}$$

Substituting for μ and σ yields:

$$M^* = \theta^{-1}[\alpha(N-N_1) - .60(\alpha(1-\alpha)(N-N_1))^{1/2} \ell n(.60\gamma/(.45\gamma'\theta))] \quad \text{(A10)}$$

If the number of problem areas identified as such, n_1, is not a negligible quantity, a different approach is required. Let us suppose that the expected value of n_1 also exceeds 18 so that the normal approximation can be used. Then since n_1 and n_2 are normally distributed, so is their sum n. The parameters of the three distributions are given in Figure A-2.

	n_1	n_2	$n=n_1+n_2$
mean	$N_1(1-\beta)$	$(N-N_1)\alpha$	$N_1(1-\beta)+(N-N_1)\alpha$
variance	$N_1\beta(1-\beta)$	$(N-N_1)\alpha(1-\alpha)$	$N_1\beta(1-\beta)+(N-N_1)\alpha(1-\alpha)$

Figure A-2. Parameters of Distributions

The preceding derivation is unchanged except for the substitution for μ and σ in the expression for M^*. The result in this case is

$$M^* = \theta^{-1}[N_1(1-\beta) + (N-N_1)\alpha -$$

$$.60\ (N_1\beta(1-\beta) + (N-N_1)\alpha(1-\alpha))^{1/2}\ \ell n(.60\gamma/(.45\gamma'\theta))] \qquad \text{(A11)}$$

To illustrate the use of this formula, the two runs with $N_1 = 50$ in Table 4 have been recalculated. The expected values of n_1 and n_2 in this case are given in Figure A-3. All are sufficiently greater

			$E[n_1]=N_1(1-\beta)$	$E[n_2]=(N-N_1)\alpha$
Satellite/ Ground System	$\alpha_s = .20$	$\beta_s = .25$	37.5	190
	$\alpha_s = .10$	$\beta_s = .10$	45.0	95
Aircraft/ Ground System	$\alpha_a = .05$	$\beta_a = .15$	42.5	47.5
	$\alpha_a = .05$	$\beta_a = .05$	47.5	47.5

Figure A-3. Expected Misclassifications for Slected Error Levels than 18 so that a normal approximation to both n_1 and n_2 is acceptable.

Each inspector can conduct 25 inspections in one period ($\theta = 25$). The cost per inspector per period is the product of the earlier cost per inspection, $C_m = 50$, and this quantity ($\gamma = 1250$). The incremental cost per inspection, assuming that these are performed on overtime, may be taken as 150% of C_m ($\gamma' = 75$).

The optimal staff of either two tier system based on these data is given by:

$$M^* = .04\mu + .00282679\,\sigma \qquad \text{(A12)}$$

In general, the value of M^* will be non-integer and must be rounded either up or down. In the results shown in Table A-4, both rounded values were checked in each case in the formula for the expected cost. Some values are out of range of the Parker approximation but not so far that a correct choice cannot be made. The optimization for the ground system must, in general, be checked in the same way, but in this example the optimal value happens to be integer.

It is noteworthy that the values of the survey costs for P_2 and P_3 are not significantly changed from those reported in Table 4. Partly this is due to the fact that satellite, aircraft and penalty factors are the dominant contributors to the cost. This also indicates that the simpler model is fairly accurate and that consequently this refinement may not be needed in many applications.

Direct common cost factor per site

Satellite	Aircraft	Cost of Misclassification by β type errors
C_s=200	C_a=15	C_p=2000

Ground cost model factors

Inspection per inspector per period	Cost per inspector per period	Incremental cost per inspection
θ = 25	γ = 1250	γ' = 75

Run No.	Input Factors						Optimal Staff*			Appx. Cost Survey($1000)		
	N	N_1	α_s	β_s	α_a	β_a	Ground	Satellite/ Ground	Aircraft/ Ground	P_1	P_2	P_3
A	1000	50	20%	25%	05%	15%	40	9	4	50	37.0	35.0
B	1000	50	10%	10%	05%	05%	40	6	4	50	17.8	25.1

* The optimal staff of a two tier system is given by: $M^* = .04\mu + .00282679\sigma$ for the given cost factors. The result is rounded up and down and the value yielding the lower cost is reported.

Figure A-4. Optimal Policies for Ground Cost Model

REFERENCES

1. Private communication R. E. Nickel, Division of Surface Mining Reclamation, Department of Natural Resources, Commonwealth of Kentucky, 1973.

2. The Issues Related to Surface Mining, Committee on Interior and Insular Affairs, U.S. Senate, December 1971, p. 24.

3. Employment and Expenditures Effect of the Kentucky Coal Industry in 1970, Spindletop Research, Lexington, Kentucky, April 1972, p. 16.

4. Annual Report, Kentucky Department of Mines and Minerals, December 31, 1972.

5. Bituminous Coal Facts, 1970, National Coal Association, Washington, D. C., 1970, p. 77.

6. The Economic Viability of the Small Underground Coal Mine in Eastern Kentucky, Spindletop Research, Lexington, Kentucky, July 1964, p. 3-4.

7. HALD, A., Statistical Theory With Engineering Applications, John Wiley & Sons, Inc., New York, 1952.

8. BROWN, R. G., Smoothing, Forecasting, and Prediction, Prentice Hall, Englewood Cliffs, New Jersey, 1963.

9. PARKER, Lawrence L., "Order Quantities and Reorder Points When The Level of Customer Service is Specified," General Memorandum No. 38, in Reference Material on Inventory Control, Arthur D. Little, Inc., Cambridge, Massachusetts, 1963.

10. DEELY, D. J., Russell, D.R. and Wobber, F. J., "Applications of Aerial and Remote Sensing to the Study of Mined Lands," paper presented at Research and Applied Technology Symposium on Mined Land Reclamation, National Coal Association, 7 March 1973, Pittsburgh, Pennsylvania.

APPENDIX B

PRELIMINARY ANALYSIS OF THE CURRENT INSPECTION PROCESS

In 1975, as part of this project MATHEMATICA [3] analyzed inspection reports for the 1971-74 interim for strip-mining permit areas in Western Kentucky. Not all inspections were included in inspection reports. For this reason, the total number of violation in the tables is low. We have assumed throughout this report that the frequencies of reported violations per inspection reports are not significantly different from the unknown frequencies of detected violations per inspection.

PRELIMINARY ANALYSIS
OF THE
CURRENT INSPECTION PROCESS:
I - INSPECTION AND
VIOLATION FREQUENCIES

A Working Paper

Submitted to

The Commonwealth of Kentucky

by

MATHMATICA, Inc.
P. O. Box 2392
Princeton, New Jersey 08540

25 August 1975

Revised 20 October 1975

Preliminary Analysis of the Current Inspection Process

The current Commonwealth of Kentucky strip mine inspection procedures call for a strip coal mine to be inspected once every two weeks. In the past three years there have been 24 inspectors assigned to the Western Kentucky Madisonville office to inspect about a hundred permitted mines which operate in this region. Currently there are 11 inspectors with each inspector assigned to about 11 mines.

From the computerized summaries of the 2760 mine inspection reports for Western Kentucky for the year 1971-1974 we have obtained the following information which characterizes the inspection situation.

1. Inspection Frequencies

Table 1 following shows data on the number of inspections by month by year for 1971-1974. Shown also in Table 1 are statistics on the number of active mines, tonnage, average weeks/mine inspection and M tons/ inspection.

> As can be seen, the average interval between inspections (calculated on the basis of a 50 week year) is significantly greater than the target value of two (by factor of roughly 3).

TABLE 1

NUMBER OF INSPECTIONS BY MONTH AND YEAR

Month	1971	1972	Year 1973	1974	1975	Total 1971-74	Average
January	49	63	56	55	126	223	55.8
February	35	59	41	55	115	190	47.5
March	53	82	55	63	-	253	63.3
April	36	77	38	56	-	207	51.8
May	63	75	46	59	-	243	60.8
June	97	89	35	58	-	279	69.8
July	74	59	27	52	-	212	53.0
August	82	70	51	56	-	259	64.8
September	75	57	42	75	-	249	62.3
October	44	40	51	68	-	203	50.8
November	62	38	46	76	-	222	55.5
December	51	36	37	96	-	220	55.0
Total	721	745	525	769	241	2760	
Average	60.1	62.1	43.8	64.1	120.5	57.5	
Number Mines*	85	71	55	90			
No. Inspections/ Mine Week	.170	.210	.191	.171	{Calculated on basis of 50-week operating year}		
Weeks/Mine Inspection	5.89	4.77	5.24	5.85			
MM Tons Produced*	31.786	33.645	31.337	28.953			
M Tons/Inspection	44.09	45.16	59.69	37.650			

*Source: U.S. Bureau of Mines

It is also of interest to analyze this data to determine relevant time trends and/or seasonal variations. Shown in the margins of Table 1 are row and column totals and appropriate mean values. Table 2 shows the complete analysis of variance of the data shown in Table 1. This analysis suggests the following conclusions:

(i) there is no significant month to month variation in inspection frequency.

(ii) year to year variations are significant at the .05 level. Nineteen seventy-three had a significantly lower inspection count than the other years. It appears that inspection frequency is keyed to the number of mines.

2. Relationship Between Violations and Inspections

When an inspection of a mine is performed, a violation (an "incident") may be reported in one of three broad categories: Method of Operation, Water Quality, or Revegetation. Each one of these main categories has several subcategories which are listed in Appendix B'. If this notice does not work, then as a last resort the State Department for Natural Resources and Environmental Protection in Frankfort may issue an order of "suspension" and request that the miner appear at hearings, at which time a spectrum of actions may be taken ranging from lifting the suspension to fines and revocation of the permit.

In Western Kentucky the following pattern of "incidents,""non-compliances" and "suspensions" existed for the years 1971-1974.

TABLE 2

ANALYSIS OF VARIANCE FOR INSPECTION FREQUENCY DATA

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F RATIO	CRITICAL F APPROXIMATE @ 95%
Row Means (Monthly Variation)	1,948	11	177.09	.723	2.13
Column Means (Yearly Variation)	3,176.25	3	1658.75	4.32	2.92
Residual	8,079.75	33	244.84		
TOTAL	13,204	47			

Sum of Squares Computations (Illustrated):

(i) Row Means $\frac{(223)^2}{4} + \frac{(190)^2}{4} + \frac{(253)^2}{4} \ldots\ldots - \frac{(2758)^2}{48} = 1,948.$

(ii) Column Means $\frac{(721)^2}{12} + \frac{(745)^2}{12} \ldots\ldots - \frac{(2758)^2}{48} = 3,176.25$

(iii) Total $49^2 + 63^2 + 56^2 + 55 + 35^2 \ldots\ldots - \frac{(2758)^2}{48} = 13,204$

(iv) Residual - By Difference

	1971	1972	1973	1974
Incidents	633	659	551	1020
Non-compliances	35	119	164	71
Non-compliances/Incidents	.055	.181	.298	.070
Suspensions	2	1	3	1
Suspensions/Non-compliances	0.057	0.008	0.018	0.014

Thus, even though the number of inspections has been relatively constant from 1971-1974*, clearly the number of reported violations increased in 1974. The ratio of non-compliances to incidents differs significantly from year to year. ($\chi^2 = 179$)

While it is not possible to tell from this data, interviews with the inspectors suggest that the reason for more incidents occurring is not that more violations are occurring but rather that the inspections have become more rigorous. On the other hand, the number of suspensions has remained small. This could possibly be due to the fact that once violations are detected they are corrected promptly.

Further insight into the current process can be gained by an examination of the relationships between violations detected and inspection frequency. If we let V be the true number of violations, $p(D)$ be the probability of detection, and v be the expected number of violations detected, it follows that

$$v = p(D)\,V. \tag{1}$$

The detection probability is a function of both technical issues (e.g., measurement devices) and operational policies (e.g., inspection frequency and thoroughness). Though detection probabilities are, of course, a function of many variables, it is generally the case that these are nonlinear with inspection effort. To illustrate, suppose that in a mining operation a given violation is detectable only for a certain length of time r (measured in fractions of a month, for example). If inspections are conducted at random instants in time* and if inspections are perfect (i.e., will always detect a violation if in progress during the inspection), then it is easy to show that the single violation detection probability, $p(D)$, is related to the monthly inspection frequency, n, by the following formula:

$$p(D) = 1 - (1 - r)^n. \tag{2}$$

(In the above equation r can also be interpreted as the single inspection detection probability.) Inspection equation (2) reveals several points:

- when $n = 0$, $p(D) = 0$;
- $p(D)$, hence v, increases as n increases, but at a decreasing rate, asymptotically approaching 1 (or V).

Figure 1 shows actual data on detections and inspection frequency by month for the years 1971-1974. Detected violations by month by year are shown in Table 3A. (A more sophisticated approach would be to compute inspections/month/mine - but the point can be made in any event.) Though substantial scatter exists, there is a clear relationship (significant at the 99% level) between violations detected and inspection frequency. This relationship will later be used to compute "corrected" violation frequencies

---

* Operational considerations may render truly random inspections impossible or more costly than fixed or scheduled inspection policies. Other inspection policies have characteristics different (and in our view poorer) from random inspection. It is beyond the scope of this paper to elaborate on these differences.

FIGURE 1. RELATION BETWEEN DETECTED VIOLATIONS AND INSPECTION FREQUENCY

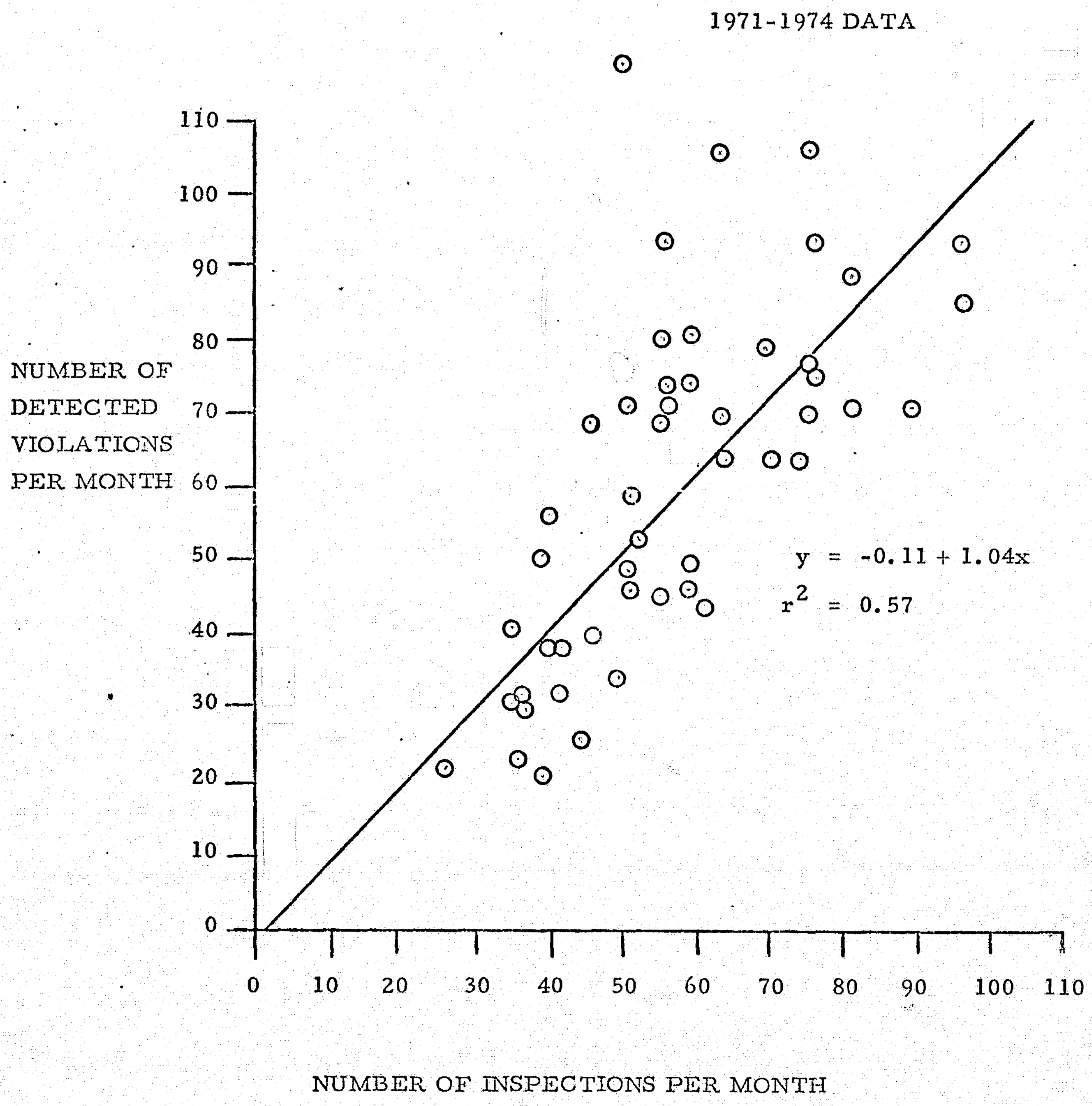


TABLE 3A. NUMBER OF VIOLATIONS BY MONTH AND YEAR

Month	1971	1972	1973	1974	Total	Average	St. Dev.
January	34	70	72	69	245	61.25	18.21
February	31	50	38	81	200	50.00	22.11
March	52	88	46	106	292	73.00	28.77
April	30	76	51	93	250	62.50	27.69
May	68	70	69	82	289	72.25	6.55
June	85	71	41	74	271	67.75	18.82
July	63	47	22	71	203	50.75	21.61
August	76	64	48	73	261	65.25	12.58
September	77	40	33	106	256	64.00	34.01
October	26	38	59	79	202	50.50	23.39
November	44	22	40	93	199	49.75	30.38
December	47	23	32	93	195	48.75	31.12
Total	633	659	551	1020			
Average	52.75	54.92	45.92	85.00			
St. Dev.	20.61	21.39	14.98	12.98			

to adjust for changes in inspection frequency. Note that v does not appear to be reaching an asymptotic value for the inspection frequencies (1971-1974) - this suggests that detection probabilities are significantly less than unity (though there are alternative explanations).

The actual counts of violations can be misleading if counts are misinterpreted as costs. This is because the counts of violations within any category depend on the refinement of violations listed under the category. For example, if vegetation violations were refined to twenty types of incidents (rather than the two types vegetation - current and vegetation - regulation used in this report), then the total number of vegetation violations might be increased tenfold. The actual cost of the violations is, however, independent of the formulation of the list of violations. A refined list of violations as used in this working paper is very useful for analysis of trends and probabilities. However, as done in this report, violations can be pooled into broad categories. The ultimate pooling is to use a single category in which a violation is defined to be one in which at least one incident occurs. Such a reduction of a multiple violation model to a single violation model is discussed in the working paper, "A Simplification of the Multiple Violation Model." Table 3 B reveals that on the average about 55% of the inspections result in an incident. This rate can be used in the cost model.

TABLE 3B. "BATTING AVERAGE" DATA

	1971	1972	1973	1974	TOTAL
Number of Inspections with at Least One Incident	399	397	274	447	1517
Total Number of Inspections	721	745	525	769	2760
Function of Inspections with at Least One Inspection	0.55	0.53	0.52	0.58	0.55

The chi-square analysis given in Table 4 shows that there are no significant year to year differences among the values in Table 3B. Again, the year to year or seasonal differences depend on the list of violations. Table 3B uses only one category (at least one incident) for a violation. However, we will show that there are indeed both seasonal and yearly differences among the aggregate number of violations. For example, vegetation incidents increased each year from 1971 to 1974, with no vegetation incidents in 1971. The aggregate yearly difference could thus be made even more dramatic if vegetation incidents were counted in twenty different ways. These results show that it is necessary to consider individual violations when analyzing trends. As a point of interest, the incidents/inspection figures are significantly higher for Western Kentucky than for Eastern Kentucky for 1972 (the only year for which such comparisons can be made). Another point of interest is that the number of incidents may represent not only "ground truth" but also changing standards in defining an incident (as reflected by the fact that no vegetation incidents were recorded in Western Kentucky in 1971).

TABLE 4

CONTINGENCY TABLE ANALYSIS

Actual frequencies from data (f_{ij})

Year	1971	1972	1973	1974	Total
Inspection with at least one incident	399	397	274	447	1517
Incident free inspections	322	348	251	322	1243
Total	721	745	525	769	2760

Expected frequencies under null hypothesis (F_{ij})

Year	1971	1972	1973	1974	Total
$\frac{(721)(1517)}{2760}$ =	396	409	289	423	1517
	325	336	236	346	1243
Total	721	745	525	769	2760

Chi-square computation:

$$\chi^2_{value} = \sum \sum \frac{(f_{ij} - F_{ij})^2}{F_{ij}} = 2.307$$

χ^2 is less than expected value of 3 and is thus insignificant.

2. Yearly and Seasonal Trends

An analysis of violation frequencies depends on the number of violations that are defined. As explained in the appendix, violation types considered in this report are listed under the following four categories: method of operation, water quality, vegetation, and discrepancies. Generally, the same trends and conclusions as given in this report will result for any sufficient refinement of violations where each violation has weights representative of the 12 types.

We first analyze the yearly trend for the aggregate violation types. The chi-square analysis for the aggregate number of incidents is given by Table 5. Because of the large number of detected incidents in 1974, the chi-square statistic has a very significant value of 96.3. A plausible explanation for the increase in 1974 has been given in the previous section.

TABLE 5

YEARLY ANALYSIS FOR AGGREGATE INCIDENTS

	1971	1972	1973	1974	Total
Observed Incidents	633	659	551	1020	2863
Expected Incidents	747.9	772.8	554.6	797.9	2863
Ratio of Observed to Expected	0.85	0.85	1.01	1.28	
Chi-square	17.65	16.76	0.08	61.8	96.3

Probably of more importance than a yearly trend is the seasonal trend. If seasonal trends exist, then adjustments in the inspection procedures can be made to increase the probability of detecting costly violations. For this reason, we have given not only a gross seasonal analysis for the aggregate incidents, but also a refined breakdown of the seasonality trend for each of the four categories.

We first examine the gross seasonality of aggregate violation counts. To do this we have calculated the relevant chi-square statistics as shown in Table 6A. The chi-square statistic was calculated under two different null hypotheses for violation counts. The first hypothesis is that violation counts are independent of either the season or the number of inspections. The chi-square value of 62.8 is very significant and thus this hypothesis must be rejected.

The second hypothesis adjusts the violation counts by the number of inspection counts. Under the second hypothesis, violation counts during any month are proportional to the number of inspections during that month but are independent of the month. The chi-square value of 27.53 reveals that again the violation counts do follow a seasonal pattern, i.e., the assumption of independence by month is invalid. This seasonal trend can be established by graphing the values of $R_i = f_i/F_i$, the rate of violation counts f_i to the average adjusted counts F_i. Results of the analyses on Table 6A are summarized by the following:

(i) Highly significant (0.01 level) seasonality in aggregate violation counts exist.

(ii) Months of greatest violation rates are January through May, with the maximum peak in April. A lower peak is attained at a point in August through October.

(iii) Months of lower violation rates bottom in July and December, with December having the lowest rate

(iv) A summary of the seasonal adjusted factors for the 12 months is shown below:

Month	% of Average	Month	% of Average
January	106	July	92
February	102	August	97
March	111	September	99
April	116	October	96
May	115	November	86
June	94	December	85

(v) A possible recommendation is that inspection frequencies be adjusted to reflect these seasonal differences.

TABLE 6A

GROSS SEASONALITY ANALYSIS

	IF NUMBER OF INCIDENTS INDEPENDENT OF NUMBER OF INSPECTIONS				IF NUMBER OF INCIDENTS ADJUSTED FOR NUMBER OF INSPECTIONS		
MONTH	AGGREGATE NUMBER OF INCIDENTS OF ALL TYPES f_i	EXPECTED NUMBER UNDER NULL HYPOTHESIS F_i	RATIO f_i/F_i	$(f_i-F_i)^2/F_i$	EXPECTED NUMBER UNDER NULL HYPOTHESIS F_i	RATIO f_i/F_i	$(f_i-F_i^2)/F_i$
JANUARY	245.	238.58	1.027	0.17	231.32	1.059	0.81
FEBRUARY	200.	238.58	0.838	6.24	197.09	1.015	0.04
MARCH	292.	238.58	1.224	11.96	262.44	1.113	3.33
APRIL	250.	238.58	1.048	0.55	214.73	1.164	5.79
MAY	289.	238.58	1.211	10.65	252.07	1.147	5.41
JUNE	271.	238.58	1.136	4.40	289.41	0.936	1.17
JULY	203.	238.58	0.851	5.31	219.91	0.923	1.30
AUGUST	261.	238.58	1.094	2.11	268.67	0.971	0.22
SEPTEMBER	256.	238.58	1.073	1.27	258.29	0.991	0.02
OCTOBER	202	238.58	0.847	5.61	210.58	0.959	0.35
NOVEMBER	199.	238.58	0.834	6.57	230.28	0.864	4.25
DECEMBER	195.	238.58	0.817	7.96	228.21	0.854	4.83
TOTAL	2863			$\chi^2_{calc} = 62.80$	2863		$\chi^2 = 27.53$

In Tables 6B, 6C, 6D, and 6E we give chi-square analysis similar to Table 6A for violations falling within single categories. While gross seasonal trends in Table 6A have been shown to be mathematically significant, in the refined analysis only the method of operation category has significant monthly differences at level .05. Vegetation is significant at level .10, while water quality is significant only at level .30 ($\chi^2_{.05} = 19.7$, $\chi^2_{.10} = 17.3$, $\chi^2_{.30} = 12.9$). Hence, it is of importance to give plausible reasons for these trends in order to establish their validity. That is, a question that should be answered is whether a particular type of incident is more likely to occur during a particular time of the year. As an aid to such a diagnostic study, we have listed the number of violations by type which occurred each month in Table 6F.

The chi-square statistic is used only for testing statistical significance and can not be used for comparing categories because the total number n of incidents falling within a category is not constant. Thus, $\chi^2 = 73.51$ for method of operation is large, both because there probably is seasonal variation and because $n = 1614$ is large. For comparison among categories χ^2/n should be used (a better statistic is the usual measure of variation given by the mean square error $s^2 = \Sigma(f_i - F)^2 / (n - 1)$). Such a comparison shows vegetation has the largest seasonal variation and water quality has the least. Both water quality and vegetation incidents peak in the spring and in the fall while method of operation incidents are consistently above average during January through June and below average the remaining six months.

TABLE 6B

SEASONALITY ANALYSIS FOR METHOD OF OPERATION

	IF NUMBER OF INCIDENTS INDEPENDENT OF NUMBER OF INSPECTIONS				IF NUMBER OF INCIDENTS ADJUSTED FOR NUMBER OF INSPECTIONS		
MONTH	AGGREGATE NUMBER OF INCIDENTS OF ALL TYPES f_i	EXPECTED NUMBER UNDER NULL HYPOTHESIS F_i	RATIO f_i/F_i	$(f_i-F_i)^2/F_i$	EXPECTED NUMBER UNDER NULL HYPOTHESIS F_i	RATIO f_i/F_i	$(f_i-F_i^2)/F_i$
JANUARY	155.	134.50	1.152	3.12	130.41	1.189	4.64
FEBRUARY	122.	134.50	0.907	1.16	111.11	1.098	1.07
MARCH	185.	134.50	1.375	18.96	147.95	1.250	9.28
APRIL	145.	134.50	1.078	0.82	121.05	1.198	4.74
MAY	157.	134.50	1.167	3.76	142.10	1.105	1.56
JUNE	164.	134.50	1.219	6.47	163.15	1.005	0.00
JULY	120.	134.50	0.892	1.56	123.97	0.968	0.13
AUGUST	144.	134.50	1.071	0.67	151.46	0.951	0.37
SEPTEMBER	138.	134.50	1.026	0.09	145.61	0.948	0.40
OCTOBER	85.	134.50	0.632	18.22	118.71	0.716	9.57
NOVEMBER	94.	134.50	0.699	12.20	129.82	0.724	9.88
DECEMBER	105.	134.50	0.781	6.47	128.65	0.816	4.35
TOTAL	1614			$\chi^2_{calc} = 73.51$			$\chi^2 = 45.99$

TABLE 6C

SEASONALITY ANALYSIS FOR WATER QUALITY

	IF NUMBER OF INCIDENTS INDEPENDENT OF NUMBER OF INSPECTIONS				IF NUMBER OF INCIDENTS ADJUSTED FOR NUMBER OF INSPECTIONS		
MONTH	AGGREGATE NUMBER OF INCIDENTS OF ALL TYPES f_i	EXPECTED NUMBER UNDER NULL HYPOTHESIS F_i	RATIO f_i/F_i	$(f_i-F_i)^2/F_i$	EXPECTED NUMBER UNDER NULL HYPOTHESIS F_i	RATIO f_i/F_i	$(f_i-F_i^2)/F_i$
JANUARY	45.	50.67	0.888	0.63	49.12	0.916	0.35
FEBRUARY	41.	50.67	0.809	1.84	41.86	0.980	0.02
MARCH	51.	50.67	1.007	0.00	55.73	0.915	0.40
APRIL	53.	50.67	1.046	0.11	45.60	1.162	1.20
MAY	62.	50.67	1.224	2.54	53.53	1.158	1.34
JUNE	49	50.67	0.967	0.05	61.46	0.797	2.53
JULY	34.	50.67	0.671	5.48	46.70	0.728	3.45
AUGUST	56.	50.67	1.105	0.56	57.06	0.982	0.02
SEPTEMBER	61.	50.67	1.204	2.11	54.85	1.112	0.69
OCTOBER	51.	50.67	1.007	0.00	44.72	1.140	0.88
NOVEMBER	60.	50.67	1.184	1.72	48.90	1.227	2.52
DECEMBER	45.	50.67	0.888	0.63	48.46	0.929	0.25
TOTAL	608			$\chi^2_{calc} = 15.68$			$\chi^2 = 13.64$

TABLE 6D

SEASONALITY ANALYSIS FOR VEGETATION

	IF NUMBER OF INCIDENTS INDEPENDENT OF NUMBER OF INSPECTIONS				IF NUMBER OF INCIDENTS ADJUSTED FOR NUMBER OF INSPECTIONS		
MONTH	AGGREGATE NUMBER OF INCIDENTS OF ALL TYPES f_i	EXPECTED NUMBER UNDER NULL HYPOTHESIS F_i	RATIO f_i/F_i	$(f_i-F_i)^2/F_i$	EXPECTED NUMBER UNDER NULL HYPOTHESIS F_i	RATIO f_i/F_i	$(f_i-F_i{}^2)/F_i$
JANUARY	9.	16.83	0.535	3.65	16.32	0.551	3.28
FEBRUARY	7.	16.83	0.416	5.74	13.91	0.503	3.43
MARCH	21.	16.83	1.248	1.03	18.52	1.134	0.33
APRIL	17.	16.83	1.010	0.00	15.15	1.122	0.23
MAY	21.	16.83	1.248	1.03	17.78	1.181	0.58
JUNE	16.	16.83	0.950	0.04	20.42	0.784	0.96
JULY	14.	16.83	0.832	0.48	15.52	0.902	0.15
AUGUST	17.	16.83	1.010	0.00	18.96	0.897	0.20
SEPTEMBER	20.	16.83	1.188	0.60	18.22	1.097	0.17
OCTOBER	26.	16.83	1.545	4.99	14.86	1.750	8.36
NOVEMBER	16.	16.83	0.950	0.04	16.25	0.985	0.00
DECEMBER	18.	16.83	1.069	0.08	16.10	1.118	0.22
TOTAL	202			$\chi^2_{calc} = 17.68$			$\chi^2 = 17.92$

TABLE 6E

SEASONALITY ANALYSIS FOR DISCREPANCIES

	IF NUMBER OF INCIDENTS INDEPENDENT OF NUMBER OF INSPECTIONS				IF NUMBER OF INCIDENTS ADJUSTED FOR NUMBER OF INSPECTIONS		
MONTH	AGGREGATE NUMBER OF INCIDENTS OF ALL TYPES f_i	EXPECTED NUMBER UNDER NULL HYPOTHESIS F_i	RATIO f_i/F_i	$(f_i-F_i)^2/F_i$	EXPECTED NUMBER UNDER NULL HYPOTHESIS F_i	RATIO f_i/F_i	$(f_i-F_i^2)/F_i$
JANUARY	36.	36.58	0.984	0.01	35.47	1.015	0.01
FEBRUARY	30.	36.58	0.820	1.18	30.22	0.993	0.00
MARCH	35.	36.58	0.957	0.07	40.24	0.870	0.68
APRIL	35.	36.58	0.957	0.07	32.92	1.063	0.13
MAY	49.	36.58	1.339	4.21	38.65	1.268	2.77
JUNE	42.	36.58	1.148	0.80	44.38	0.946	0.13
JULY	35	36.58	0.957	0.07	33.72	1.038	0.05
AUGUST	44.	36.58	1.203	1.50	41.20	1.068	0.19
SEPTEMBER	37.	36.58	1.011	0.00	39.61	0.934	0.17
OCTOBER	40.	36.58	1.093	0.32	32.29	1.239	1.84
NOVEMBER	29.	36.58	0.793	1.57	35.31	0.821	1.13
DECEMBER	27.	36.58	0.738	2.51	34.99	0.772	1.83
TOTAL	439			x^2_{calc} = 12.33			x^2 = 8.93

TABLE 6F

NUMBER OF VIOLATIONS BY TYPE BY MONTH

(1971-1974 RAW DATA)

	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
METHOD OF OPERATION	6	5	17	11	11	10	7	7	15	13	11	15
GRADING CURRENT	105	86	129	98	107	114	84	102	93	54	62	65
GRADING PLAN	38	26	34	32	33	32	20	25	20	17	12	17
ACCESS ROAD	6	5	5	4	6	8	9	10	10	1	9	8
SILT STRUCTURE	15	16	21	19	23	16	13	20	22	19	27	19
WATER QUALITY CHEMICAL	6	10	9	9	7	7	3	9	10	6	5	4
WATER QUALITY PHYSICAL	6	3	1	2	3	1	1	6	7	5	6	7
DRAINAGE PLAN	14	12	19	18	25	17	14	17	17	17	19	11
WATER IMPOUNDMENT	4	0	1	5	4	8	3	4	5	4	3	3
VEGETATION REGULATION	4	4	11	8	6	6	5	8	8	10	4	7
VEGETATION CURRENT	5	3	10	9	15	10	9	9	12	16	11	11
DISCREPANCIES	36	30	35	35	49	42	35	44	37	40	29	27
NON-VIOLATION	93	83	91	84	103	131	108	115	116	101	113	105

An analysis was made between the amount of monthly precipitation and the number of incidents to determine if such an association could account for a significant percentage of the seasonal variation. To explain this analysis, let S_i denote the average rainfall for the $i^{\underline{th}}$ month and $\overline{S}$ denote the average monthly precipitation over all months. Then the rate of precipitation for the $i^{\underline{th}}$ month above the average is defined by

$$x_i = S_i/\overline{S}$$

Let $y_i = f_i/F_i$ denote the $i^{\underline{th}}$ rate of incidents for a given category. If a linear relation exists between incidents and precipitation, then, except for random error, y_i is given by

$$y_i = a + bx_i$$

The value of b is positive if the correlation is positive, negative if the correlation is negative, and insignificant if there is no significant correlation. The total seasonal variation for incidents is

$$S_y^2 = \Sigma (y_i - \overline{y})^2$$

The total seasonal variation for precipitation is

$$S_x^2 = \Sigma (x_i - \overline{x})^2 \qquad (\overline{x} = 1)$$

The correlation R between x and y is defined by

$$R = \frac{S_{xy}}{S_x \cdot S_y} \quad \text{where } S_{xy} = \Sigma (x_i - \overline{x})(y_i - \overline{y})$$

The variation due to the linear relation between y and x is

$$S^2_{y|x} = S^2_y \cdot R^2$$

The percentage of seasonal variation accounted for by precipitation is simply $100R^2$%.

Below is listed the amount of precipitation in inches/month in Western Kentucky, averaged over the years 1931-55.

	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
Precipitation	5.10	3.69	5.31	4.30	3.77	4.09	4.17	3.55	3.10	2.50	3.35	3.92
Rate (x_i)	1.31	0.95	1.36	1.10	0.97	1.05	1.07	0.91	0.79	0.64	0.86	1.00

Table 6G summarizes the analysis of the correlation between the monthly precipitation values and the monthly rate of incidents by category. The only significant correlation that was found was in the method of operation category. The estimated linear relation for this category is

$$y_i = .29 + .70x_i$$

Figure 2 illustrates the obvious correlation between method of operation and precipitation.

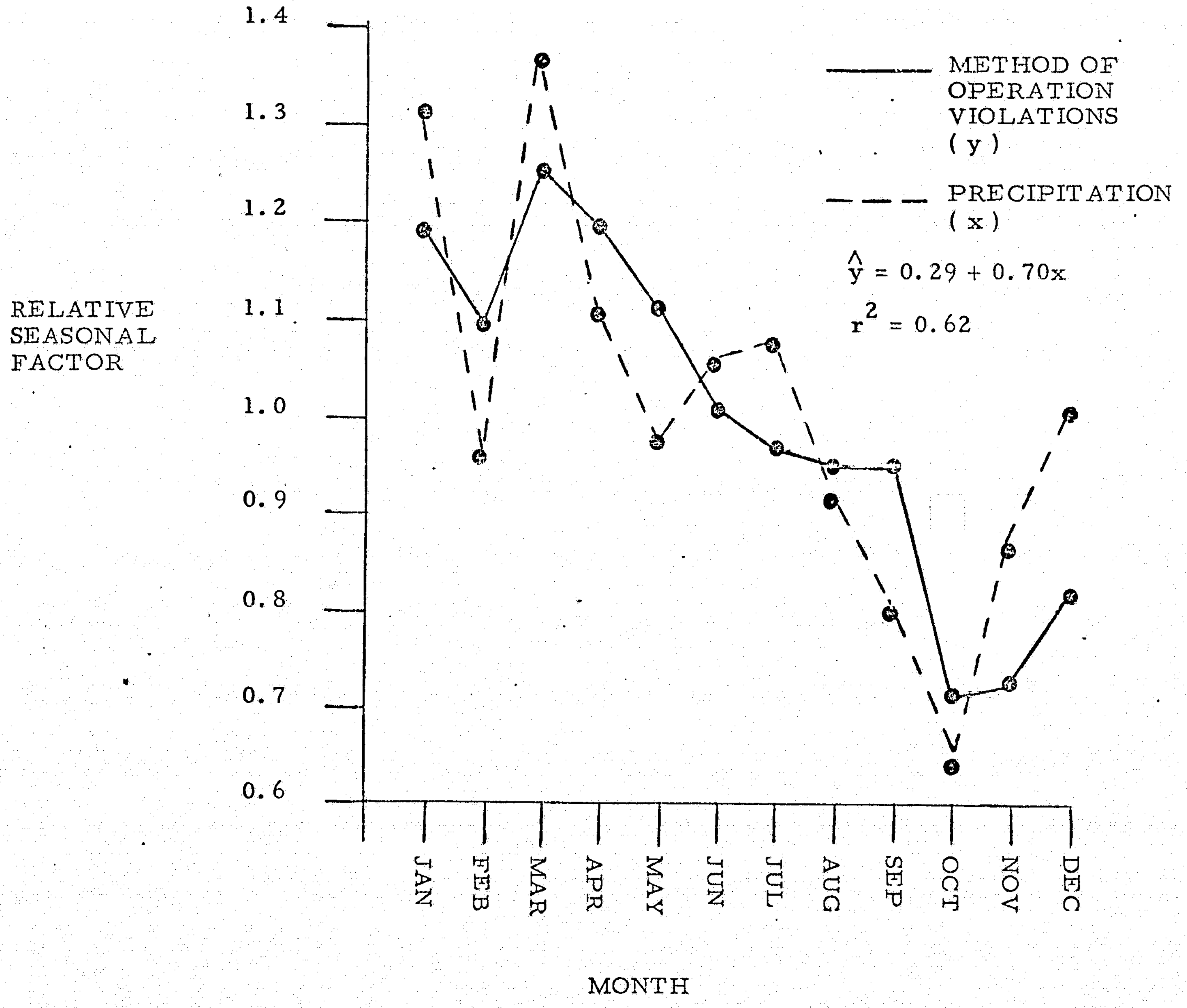
TABLE 6G

STATISTICAL ANALYSIS OF CORRELATIONS BETWEEN MONTHLY PRECIPITATION RATES AND INCIDENT RATES

	METHOD OF OPERATION	WATER QUALITY	VEGETATION
Total Monthly Variation (Sy^2)	0.355	0.270	1.167
Variation Accounted for by Precipitation ($Sy^2 \cdot R^2$)	0.221*	0.070	0.283
Correlation (R)	.789*	-.510	-.492
F-Statistic $\left(\frac{10R^2}{1 - R^2}\right)$	16.507*	3.524	3.201

* Significant at level .05 if F > 4.96

FIGURE 2. RELATION BETWEEN MONTHLY PRECIPITATION RATES AND MONTHLY RATES FOR METHOD OF OPERATION VIOLATIONS



4. Probabilities and Measures of Association Between Violations

One of the topics of this report is the frequency or probability (or marginal probability) with which a violation or incident occurs. By the true probability of an incident we mean the fraction of days that a specified incident is expected to occur. Since this is unavailable, we estimate the probability from the 2760 inspections given in the inspection reports for the interim 1971-74. A probability is estimated by the ratio of the total number of occurrences of a particular violation (at most one on any given inspection) to the total number of inspections. The values of these probability estimates depend on whether the ratio is made by counting by a specific month, year, or by counting over all 2760 inspections.

Tables 7A and 7B show the probabilities of occurrences of violations by category (each category counted at most once per inspection) by month and by year. These tables again illustrate the yearly and seasonal trends analyzed in the previous section. A conclusion not arrived at previously is that the increase in violations in 1974 by category is due to the three categories: water quality, vegetation, and discrepancies. Method of operation violations actually decreased in 1974.

Table 8 lists the probabilities of each of the twelve types of incidents, averaged over all of the inspections. The fact that there are more violations than inspections (1.0366 = 2863/2760) is consistent with the fact that several violations occur simultaneously.

The fact that some violations may be dependent on the occurrence of other violations may be an asset to aerial or satellite inspection. This is because it is possible that some violations may be easily detectable from the air or from satellites while others are not. There is less

TABLE 7A

PROBABILITY OF OCCURRENCES OF VIOLATIONS

BY CATEGORY FOR GIVEN MONTH

MONTH	METHOD OF OPERATION	WATER QUALITY	VEGETATION	DIS-CREPANCIES
JANUARY	.4933	.1570	.0224	.1614
FEBRUARY	.4842	.1474	.0211	.1579
MARCH	.5534	.1700	.0593	.1383
APRIL	.5072	.1836	.0531	.1691
MAY	.4527	.1893	.0617	.2016
JUNE	.4373	.1470	.0466	.1505
JULY	.4104	.1274	.0425	.1651
AUGUST	.4247	.1737	.0425	.1699
SEPTEMBER	.3976	.1847	.0522	.1486
OCTOBER	.3103	.2217	.0837	.1970
NOVEMBER	.3243	.2297	.0541	.1306
DECEMBER	.3727	.1773	.0500	.1227

TABLE 7B

PROBABILITY OF OCCURRENCES OF VIOLATIONS

BY CATEGORY FOR YEARS 1971-1974

YEAR	METHOD OF OPERATION	WATER QUALITY	VEGETATION	DIS-CREPANCIES
1971	.5395	.0472	.0	.1123
1972	.4846	.1154	.0054	.0859
1973	.3429	.2324	.0590	.1638
1974	.3407	.3147	.1313	.2705

TABLE 8

VIOLATION FREQUENCIES BY TYPE

TOTAL 1971-1974

VIOLATION TYPE	TOTAL VIOLATIONS 1971-1974	AVERAGE NUMBER PER INSPECTION	RELATIVE RATE (GRADING=1)
GRADING CURRENT	1099	0.3982	1.000
DISCREPANCIES	439	0.1591	0.399
GRADING TO PLAN	306	0.1109	0.278
SILT STRUCTURE	230	0.0823	0.209
DRAINAGE PLAN	200	0.0725	0.182
METHOD OF OPERATION	128	0.0464	0.116
VEGETATION CURRENT	120	0.0435	0.109
WATER QUALITY CHEMICAL	85	0.0308	0.077
ACCESS ROAD	81	0.0293	0.074
VEGETATION REGULATION	81	0.0293	0.074
WATER QUALITY PHYSICAL	48	0.0174	0.044
WATER IMPOUNDMENT	44	0.0159	0.040
	2863	1.0366	

concern about missing a specific violation if there is a high probability of detecting a different highly correlated violation. Thus, if A and B are two types of incidents which are highly correlated and A can be detected while B can not, then the inference that B has occurred might be made whenever A is detected.

For the 12 types of incidents, there are 132(=144-12) conditional probabilities of one incident given another. Thus, for simplicity, the analysis of these 144 ordered pairs is better illustrated by analyzing a single pair. The following numerical example is taken from the 12 $\times$ 12 matrices given in Tables 10 through 14.

<u>A Numerical Example</u>

In this section we provide a numerical example to illustrate the definition and computation of various quantities associated with the correlation among various violation types.

The input data for all of these computations is illustrated by Table 9A below for two violation types - (i) method of operation and (ii) drainage plan. Referring to the table, we see that of the total of 2760 inspections 34 resulted in both violation types being present, 166 detected a drainage plan violation but no method of operation violation, etc. Shown to the right of each number is a symbol which will be used subsequently. Equivalently, we may convert the data of Table 9A into a table of proportions which shows the probability of each of the various events of interest. Such a table is shown as Table 9B - to save space in this and further tables and discussions we define events A and B to represent method of operation and drainage plan violations respectively.

Table 9A. Association Between

Method of Operation and Drainage Violations

		Number of Inspections in which the Drainage Plan Was: In Violation	Not In Violation	Total
Number of Inspections in which the Method of Operation Was:	In Violation	34 n_{11}	94 n_{12}	128 $n_{1.}$
	Not In Violation	166 n_{21}	2466 n_{22}	2632 $n_{2.}$
	Total	200 $n_{.1}$	2560 $n_{.2}$	2760 $n_{..}$

Table 9B. A Probability

Matrix for Violation Types

	B	$\overline{B}$	Total
A	0.012 p_{11}	0.034 p_{12}	0.046 $p_{1.}$
$\overline{A}$	0.060 p_{21}	0.894 p_{22}	0.964 $p_{2.}$
Total	0.072 $p_{.1}$	0.928 $p_{.2}$	1.000 1

For reasons discussed in previous working papers, it is important to examine the association between events A and B. Suppose, for example, that events A and B are statistically independent. In this case the expected number of inspections resulting in the event $A \cap B$ would be given by $200 \cdot \frac{128}{2760} = 9.3$, considerably beneath the 34 cases actually observed. Similar computations for each of the other events results in the matrix of values shown below:

Table 9C. Expected Frequency

Under Null Hypothesis

	B	$\overline{B}$	Total
A	9.3	118.7	128
$\overline{A}$	190.7	2441.3	2632
Total	200	2560	2760

To test the significance of these discrepancies we compute the χ^2 statistic shown below and compare it to the appropriate critical value:

$$\chi^2 = \frac{n_{..}\left(\left|n_{11}\,n_{22} - n_{12}\,n_{21}\right| - \frac{1}{2}n_{..}\right)^2}{n_{1.}\,n_{2.}\,n_{.1}\,n_{.2}},$$

which for this example is,

$$\chi^2 = \frac{2760\left(\left|34\cdot 2466 - 166\cdot 94\right| - \frac{1}{2}\,2760\right)^2}{128\cdot 2632\cdot 200\cdot 2560} = 71.53\,.$$

The critical value (at the 95% level) for this statistic is 3.84, so the observed association is statistically highly significant - i.e., method of operation and drainage plan violations are correlated.

There are several ways in which this correlation can be estimated or illustrated. The first is by a measure termed (unfortunately since the name is not descriptive in this context) the relative risk or R statistic. This statistic is defined below:

$$R = \frac{p(B|A)}{p(B|\bar{A})}$$

and is the ratio of the conditional probabilities of event B given that event A has occurred to that in the event A has not occurred. In terms of the symbols defined in Table 9A, R is given by,

$$R = \frac{\dfrac{n_{11}}{n_{2\cdot}}}{\dfrac{n_{21}}{n_{2\cdot}}} = \frac{n_{11}n_{2\cdot}}{n_{21}n_{2\cdot}} = \frac{34 \cdot 2632}{166 \cdot 128} = 4.212 \text{ .}$$

In this case, drainage plan violations are 4.2 times more likely to occur when the mine has a method of operation violation than would be the case if no method of operation violation were noted. If A and B were independent events, R should be equal to unity.

Another descriptive measure of association is the so-called odds ratio. This is developed as follows:

(i) A measure of the relative likelihood of experiencing an outcome B when event A has occurred is,

$$\Omega_A = \frac{p(B|A)}{p(\bar{B}|A)} \quad \text{or} = \frac{n_{11}/n_{1\cdot}}{n_{12}/n_{1\cdot}} = \frac{n_{11}}{n_{12}} \text{ .}$$

In this example Ω_A is 34/94 = 0.3617, or in other words, for every inspection in which the drainage plan is in violation, there are about 1/0.3617 = 2.76 inspections when no drainage plan violation occurs given that there is a method of operation violation.

(ii) When A is absent, the odds of B's occurrence are defined as,

$$\Omega_{\bar{A}} = \frac{p(B|\bar{A})}{p(\bar{B}|\bar{A})} = \frac{n_{21}/n_{2\cdot}}{n_{22}/n_{2\cdot}} = \frac{n_{21}}{n_{22}} .$$

In this example $\Omega_{\bar{A}}$ is 166/2466 = 0.06732.

(iii) The two odds Ω_A and $\Omega_{\bar{A}}$ can be contrasted in a number of ways to provide a measure of association. The odds ratio, ω, is currently in greatest use. ω is defined as,

$$\omega = \frac{\Omega_A}{\Omega_{\bar{A}}} = \frac{n_{11}\, n_{22}}{n_{12}\, n_{21}} = \frac{(34)\ (2466)}{(166)\ (94)} = 5.373 ,$$

which, for this example, indicates that the odds of an inspection turning up a drainage plan violation are 5.4 times as likely if a method of operation violation occurs than if this is not the case. As was the case with the relative risk measure, the odds ratio is 1 if A and B are independent.

Having defined and illustrated various statistical concepts relevant to detecting, testing, and estimating association between events, we now examine the full set of inspection data.

Table 10 lists the number of times a violation A in a particular row occurs with a violation B in a particular column. For each such pair of violations a table similar to Table 9A can be constructed. Note that of the values in Table 9A, only $n_{11} = 34$ can be found in Table 10. The row and column totals $n_{1.} = 128$ and $n_{.1} = 200$ are the marginal totals given in Table 8. Since $n_{..} = 2760$ is known, all other values in Table 9A can be found by subtraction.

The probabilities $P(B|A)$ and $P(B|\bar{A})$ illustrated in Table 9B are given in Tables 11A and 11B. Large values of $P(B|A)$ and small values of $P(B|\bar{A})$ are ideal when the occurrence of A is used to identify B. The worst case is when $P(B|A) = P(B|\bar{A})$ (if $P(B|\bar{A}) > P(B|A)$ then the non-occurrence of $\bar{A}$ can be used to predict B). The chi-square statistics for the two-sided tests of $P(B|A) = P(B|\bar{A})$ are given in Table 12. Of the 66 unordered pairs off the diagonal, 29 chi-square statistics were significant at level .05. Of these 29 cases, 26 were significant at level .01. Thus, correlation among violations is widespread.

Tables 13 and 14 list only those relative risks and odds ratios for pairs that are significantly dependent. A quick overview of these tables shows that the violation types within each category tend to be more closely associated with each other than with violation types outside of the category. The two types most closely associated with each other are vegetation-current and vegetation-regulation. Discrepancies are associated with every type. This is not surprising since the detection of any violation type increases the probability of a discrepancy.

Through the use of either the odds ratios or the relative risks tables, one can determine for any violation type, that violation type which it is most closely associated with. Thus, the physical and chemical water types are both associated with silt structure. Note that although silt structure is associated with drainage plan, one can not draw the inference that drainage plan is associated with physical water quality. As in the seasonal trend analysis, it is important to make a diagnostic study to determine what, if any, causal relations exist among the pairs that are associated. The determination of a logical basis for an association supports the use of such associations for the detection of incidents. If on the other hand, associations are not necessary, such associations can not be guaranteed to exist in the future. This statement is particularly relevant if the correlation among violations is to be exploited for satellite detection purposes. If correlation is not intrinsic and miners can learn that these correlations are the "tip offs" or "signatures," the miners can rectify operating procedures in the future so as to deny these signatures In this case then, secrecy is essential.

Table 15 lists the top ten pairs of violation types that are associated according to the odds ratio measure of association. In six of the ten cases, a plausible explanation can be given for these associations. Four pairs where an association is not apparently necessary are: vegetation current and water impoundment, silt structure and vegetation current, silt structure and access road, and silt structure and vegetation regulation. It should be recognized that the estimated odds ratios for some pairs of incidents may be either much higher or lower than the true values (such as vegetation current and water impoundment) because the sample size is too small for an accurate estimate (vegetation current and water impoundment occurred together only twelve times).

TABLE 10

NUMBER OF OCCURRENCES OF TWO TYPES OF VIOLATIONS ON THE SAME INSPECTION

(DIAGONAL IS THE NUMBER OF VIOLATIONS BY TYPE)

A \ B	METHOD OF OPERATION	GRADING CURRENT	GRADING TO PLAN	ACCESS ROAD	SILT STRUCTURE	WATER QUALITY CHEMICAL	WATER QUALITY PHYSICAL	DRAINAGE PLAN	WATER IMPOUNDMENT	VEGETATION REGULATION	VEGETATION CURRENT	DISCREPANCIES
METHOD OF OPERATION	128	87	66	9	23	8	1	34	3	4	6	39
GRADING CURRENT	87	1099	274	53	90	29	15	106	28	35	62	316
GRADING TO PLAN	66	274	306	23	34	8	4	52	11	17	19	158
ACCESS ROAD	9	53	23	81	20	0	4	17	2	6	10	38
SILT STRUCTURE	23	90	34	20	230	17	12	73	2	20	31	95
WATER QUALITY CHEMICAL	8	29	8	0	17	85	4	13	1	5	8	27
WATER QUALITY PHYSICAL	1	15	4	4	12	4	48	6	2	1	3	19
DRAINAGE PLAN	34	106	52	17	73	13	6	200	6	6	12	86
WATER IMPOUNDMENT	3	28	11	2	2	1	2	6	45	4	12	22
VEGETATION REGULATION	4	35	17	6	20	5	1	6	4	82	66	40
VEGETATION CURRENT	6	62	19	10	31	8	3	12	12	66	120	68
DISCREPANCIES	39	316	158	38	95	27	19	86	22	40	68	439

TABLE 11A

MATRIX OF CONDITIONAL PROBABILITIES P(B|A) OF A VIOLATION GIVEN ANOTHER VIOLATION IS DETECTED

HERE IS THE PROBABILITY THAT THIS VIOLATION ALSO OCCURRED

A (IF THIS VIOLATION OCCURRED ON A GIVEN INSPECTION) \ B	METHOD OF OPERATION	GRADING CURRENT	GRADING TO PLAN	ACCESS ROAD	SILT STRUCTURE	WATER QUALITY CHEMICAL	WATER QUALITY PHYSICAL	DRAINAGE PLAN	WATER IMPOUNDMENT	VEGETATION REGULATION	VEGETATION CURRENT	DISCREPANCIES
METHOD OF OPERATION		0.6797**	0.5156**	0.0703*	0.1797**	0.0625	0.0078	0.2656**	0.0234	0.0313	0.0469	0.3047**
GRADING CURRENT	0.0792**		0.2493**	0.0482**	0.0819	0.0264	0.0136	0.0965**	0.0255**	0.0318	0.0564**	0.2875**
GRADING TO PLAN	0.2157**	0.8954**		0.0752**	0.1111	0.0261	0.0131	0.1699**	0.0359**	0.0556**	0.0621	0.5163**
ACCESS ROAD	0.1111*	0.6543**	0.2840**		0.2469**	0.0	0.0494	0.2099**	0.0247	0.0741*	0.1235**	0.4691**
SILT STRUCTURE	0.1000**	0.3913	0.1478	0.0870**		0.0739**	0.0522**	0.3174**	0.0087	0.0870**	0.1348**	0.4130**
WATER QUALITY CHEMICAL	0.0941	0.3412	0.0941	0.0	0.2000**		0.0471	0.1529**	0.0118	0.0588	0.0941*	0.3176**
WATER QUALITY PHYSICAL	0.0208	0.3125	0.0833	0.0833	0.2500**	0.0833		0.1250	0.0417	0.0206	0.0625	0.3958**
DRAINAGE PLAN	0.1700**	0.5300**	0.2600**	0.0850**	0.3650**	0.0650**	0.0300		0.0300	0.0300	0.0600	0.4300**
WATER IMPOUNDMENT	0.0667	0.6222**	0.2444**	0.0444	0.0444	0.0222	0.0444	0.1333		0.0889	0.2667**	0.4889**
VEGETATION REGULATION	0.0488	0.4268	0.2073**	0.0732*	0.2439**	0.0610	0.0122	0.0732	0.0488		0.8049**	0.4878**
VEGETATION CURRENT	0.0500	0.5167**	0.1583	0.0833**	0.2533**	0.0667*	0.0250	0.1000	0.1000**	0.5500**		0.5667**
DISCREPANCIES	0.0888**	0.7198**	0.3599**	0.0866**	0.2164**	0.0615**	0.0433**	0.1959**	0.0501**	0.0911**	0.1549**	
MARGINAL PROBABILITY P(B)	0.0464	0.3982	0.1109	0.0293	0.0823	0.0308	0.0174	0.0725	0.0159	0.0293	0.0435	0.1591

*SIGNIFICANTLY DIFFERENT FROM P(B) AT SIGNIFICANCE LEVEL .05.

**SIGNIFICANTLY DIFFERENT FROM P(B) AT SIGNIFICANCE LEVEL .01.

TABLE 11B

MATRIX OF CONDITIONAL PROBABILITIES P(B|$\overline{A}$) OF A VIOLATION GIVEN ANOTHER VIOLATION IS NOT DETECTED

HERE IS THE PROBABILITY THAT THIS VIOLATION OCCURRED

IF THIS VIOLATION DID NOT OCCUR ON A GIVEN INSPECTION

A \ B	METHOD OF OPERATION	GRADING CURRENT	GRADING TO PLAN	ACCESS ROAD	SILT STRUCTURE	WATER QUALITY CHEMICAL	WATER QUALITY PHYSICAL	DRAINAGE PLAN	WATER IMPOUNDMENT	VEGETATION REGULATION	VEGETATION CURRENT	DISCREPANCIES
METHOD OF OPERATION		0.3845**	0.0912**	0.0274*	0.0786**	0.0293	0.0179	0.0631**	0.0160	0.0296	0.0433	0.1520**
GRADING CURRENT	0.0247**		0.0193**	0.0169**	0.0843	0.0337	0.0199	0.0566**	0.0102**	0.0283	0.0349**	0.0741**
GRADING TO PLAN	0.0253**	0.3362**		0.0236**	0.0799	0.0314	0.0179	0.0603**	0.0139**	0.0265**	0.0412	0.1145**
ACCESS ROAD	0.0444*	0.3904**	0.1056**		0.0784**	0.0317	0.0164	0.0683**	0.0161	0.0284*	0.0411**	0.1497**
SILT STRUCTURE	0.0415**	0.3988	0.1075	0.0241**		0.0269**	0.0142**	0.0502**	0.0170	0.0245**	0.0352**	0.1360**
WATER QUALITY CHEMICAL	0.0449	0.4000	0.1114	0.0303	0.0796**		0.0164	0.0699**	0.0164	0.0288	0.0419*	0.1540**
WATER QUALITY PHYSICAL	0.0468	0.3997	0.1114	0.0284	0.0604**	0.0299		0.0715	0.0159	0.0299	0.0431	0.1549**
DRAINAGE PLAN	0.0367**	0.3879**	0.0992**	0.0250**	0.0613**	0.0281**	0.0164		0.0152	0.0297	0.0422	0.1379**
WATER IMPOUNDMENT	0.0460	0.3945**	0.1087**	0.0291	0.0840	0.0309	0.0169	0.0715		0.0287	0.0398**	0.1536**
VEGETATION REGULATION	0.0463	0.3973	0.1079**	0.0280*	0.0784**	0.0299	0.0176	0.0724	0.0153		0.0202**	0.1490**
VEGETATION CURRENT	0.0462	0.3928	0.1087	0.0269**	0.0754**	0.0292*	0.0170	0.0712	0.0125**	0.0061**		0.1405**
DISCREPANCIES	0.0383**	0.3374**	0.0638**	0.0185**	0.0582**	0.0250**	0.0125**	0.0491**	0.0099**	0.0181**	0.0224**	
MARGINAL PROBABILITY P(B)	0.0464	0.3982	0.1109	0.0293	0.0823	0.0308	0.0174	0.0725	0.0159	0.0293	0.0435	0.1591

*SIGNIFICANTLY DIFFERENT FROM P(B) AT SIGNIFICANCE LEVEL .05.

**SIGNIFICANTLY DIFFERENT FROM P(B) AT SIGNIFICANCE LEVEL .01

TABLE 12

CHI-SQUARE VALUES FOR TESTING IF THE CONDITIONAL PROBABILITY P(B|A) DIFFERS SIGNIFICANTLY FROM THE MARGINAL PROBABILITY P(B)

A \ B	METHOD OF OPERATION	GRADING CURRENT	GRADING TO PLAN	ACCESS ROAD	SILT STRUCTURE	WATER QUALITY CHEMICAL	WATER QUALITY PHYSICAL	DRAINAGE PLAN	WATER IMPOUNDMENT	VEGETATION REGULATION	VEGETATION CURRENT	DISCREPANCIES
METHOD OF OPERATION		43.16	218.78	6.47	15.02	3.47	0.25	71.53	0.09	0.03	0.00	20.16
GRADING CURRENT	43.16		352.76	21.76	0.02	0.96	1.15	15.05	8.65	0.18	6.84	223.76
GRADING TO PLAN	218.78	352.76		23.58	3.08	0.11	0.15	47.03	6.96	7.00	2.39	325.44
ACCESS ROAD	6.47	21.76	23.58		27.07	1.70	3.26	21.36	0.03	4.22	10.93	57.62
SILT STRUCTURE	15.02	0.02	3.08	27.07		14.09	15.61	219.99	0.46	26.40	47.93	118.95
WATER QUALITY CHEMICAL	3.47	0.96	0.11	1.70	14.09		2.90	7.26	0.01	1.64	4.22	15.29
WATER QUALITY PHYSICAL	0.25	1.15	0.15	3.26	15.61	2.90		1.29	0.68	0.00	0.09	18.71
DRAINAGE PLAN	71.53	15.05	47.03	21.38	219.99	7.26	1.29		1.69	0.04	1.02	116.17
WATER IMPOUNDMENT	0.09	8.65	6.96	0.03	0.46	0.01	0.68	1.69		3.67	49.47	34.74
VEGETATION REGULATION	0.03	0.18	7.00	4.22	26.40	1.64	0.00	0.04	3.67		1159.28	65.77
VEGETATION CURRENT	0.00	6.84	2.39	10.93	47.93	4.22	0.09	1.02	49.47	1159.28		152.66
DISCREPANCIES	20.16	223.76	325.44	57.62	118.95	15.29	18.71	116.17	34.74	65.77	152.66	

SIGNIFICANT AT LEVEL .05 IF $x^2 > 3.84$

VERY SIGNIFICANT AT LEVEL .01 IF $x^2 > 6.63$

TABLE 13

ODDS RATIOS FOR MEASURING THE ASSOCIATION OF TWO VIOLATION TYPES

A \ B	METHOD OF OPERATION	GRADING CURRENT	GRADING TO PLAN	ACCESS ROAD	SILT STRUCTURE	WATER QUALITY CHEMICAL	WATER QUALITY PHYSICAL	DRAINAGE PLAN	WATER IMPOUNDMENT	VEGETATION REGULATION	VEGETATION CURRENT	DISCREPANCIES
METHOD OF OPERATION	∞	3.4	10.6	2.7*	2.6			5.4				2.4
GRADING CURRENT	3.4	∞	16.9	3.0				1.8	2.5		1.7	5.0
GRADING TO PLAN	10.6	16.9	∞	3.4				3.2	2.7	2.2		8.3
ACCESS ROAD	2.7*	3.0	3.4	∞	3.9			3.6		2.7*	3.3	5.0
SILT STRUCTURE	2.6			3.9	∞	2.9	3.8	8.8		3.8	4.3	4.5
WATER QUALITY CHEMICAL					2.9	∞		2.4			2.4*	2.6
WATER QUALITY PHYSICAL					3.8		∞					3.6
DRAINAGE PLAN	5.4	1.8	3.2	3.6	8.8	2.4		∞				4.7
WATER IMPOUNDMENT		2.5	2.7						∞		8.8	5.3
VEGETATION REGULATION			2.2	2.7*	3.8					∞	200.4	5.4
VEGETATION CURRENT		1.7		3.3	4.3	2.4*			8.8	200.4	∞	8.0
DISCREPANCIES	2.4	5.0	8.3	5.0	4.5	2.6	3.6	4.7	5.3	5.4	8.0	∞

*SIGNIFICANT ONLY AT LEVEL .05

OMITTED VALUES ARE INSIGNIFICANT

UNASTERISKED VALUES ARE SIGNIFICANT AT LEVEL .01

$$\omega = \frac{\Omega_A}{\Omega_{\overline{A}}} = \frac{P(B|A)}{P(\overline{B}|A)} \frac{P(\overline{B}|\overline{A})}{P(B|\overline{A})}$$

TABLE 14

RELATIVE RISK OF THE OCCURRENCE OF VIOLATION TYPE B SPECIFIC TO THE OCCURRENCE OF TYPE A

HERE IS THE RELATIVE RISK OF THE OCCURRENCE OF THIS VIOLATION

IF THIS VIOLATION IS PRESENT ON AN INSPECTION

A \ B	METHOD OF OPERATION	GRADING CURRENT	GRADING TO PLAN	ACCESS ROAD	SILT STRUCTURE	WATER QUALITY CHEMICAL	WATER QUALITY PHYSICAL	DRAINAGE PLAN	WATER IMPOUNDMENT	VEGETATION REGULATION	VEGETATION CURRENT	DISCREPANCIES
METHOD OF OPERATION	∞	1.8	5.7	2.6*	2.3			4.2				2.0
GRADING CURRENT	3.2	∞	12.9	2.9				1.7	2.5		1.6	3.9
GRADING TO PLAN	8.5	2.7	∞	3.2				2.8	2.6	2.1		4.5
ACCESS ROAD	2.5*	1.7	2.7	∞	3.1			3.1		2.6*	3.0	3.1
SILT STRUCTURE	2.4			3.6	∞	2.7	3.7	6.3		3.5	3.8	3.0
WATER QUALITY CHEMICAL					2.5	∞		2.2			2.2*	2.1
WATER QUALITY PHYSICAL					3.1		∞					2.6
DRAINAGE PLAN	4.6	1.4	2.6	3.4	6.0	2.3		∞				3.1
WATER IMPOUNDMENT		1.6	2.2						∞		6.7	3.2
VEGETATION REGULATION			1.9	2.6*	3.1					∞	39.9	3.3
VEGETATION CURRENT		1.3		3.1	3.4	2.3*			8.0	90.8	∞	4.0
DISCREPANCIES	2.3	2.1	5.6	4.7	3.7	2.5	3.5	4.0	5.1	5.0	6.9	∞

*SIGNIFICANT ONLY AT LEVEL .05

OMITTED VALUES ARE INSIGNIFICANT

UNASTERISKED VALUES ARE SIGNIFICANT AT LEVEL .01

$$R = \frac{P(B|A)}{P(B|\overline{A})}$$

TABLE 15

TOP TEN INTERACTIONS ON BASIS ON ODDS RATIO

(DISCREPANCIES IGNORED)

ODDS RATIO	INTERACTION
200.4	Vegetation Regulation and Vegetation Current
16.9	Grading Current and Grading to Plan
10.6	Grading to Plan and Method of Operation
8.8	Silt Structure and Drainage Plan
8.8	Vegetation Current and Water Impoundment
5.4	Drainage Plan and Method of Operation
4.3	Silt Structure and Vegetation Current
3.9	Silt Structure and Access Road
3.8	Silt Structure and Physical Water Quality
3.8	Silt Structure and Vegetation Regulation

5. Conclusions

In this working paper we have given a statistical analysis to show that seasonal and yearly differences do exist for the frequencies with which violation types occur. We have also shown that significant dependencies exist among violation types. These results should be useful in developing future inspection procedures.

Some further statistical analyses can be done with the data from the 2760 inspections. For example, we have shown that the number of violations by month and year is linearly related to the number of inspections during the same months and years. We have also shown that the number of inspections shows no significant seasonal trend but does have a significant dependency on years. Since the number of mines does change from year to year, analyses should be made which examine the relationship of inspections per year per mine and violations per inspections per mine by month and year. Also, a probability model should be developed which assumes the true number of violations at any time is a variable which increases with the number of mines in operation. Because mines have variable capacities, a second approach would be to replace the number of mines with the number of tons of coal produced. Such analyses can be made by the method of analysis of covariance.

Some of the analyses given in this report will be included as input for the cost/effectiveness models of mining inspections by satellites with follow-up ground or aircraft inspections.

APPENDIX B'

Violation Categories

The eighteen (18) violations on the Data Entry Form were grouped into twelve (12) violation types. Types were formed not only on the basis of the relationships among violations but also on the basis of assumptions about what Landsat could or could not "see." Violation types were then aggregated into four (4) very broad violation categories based upon the relationships among the violation types.

To explain the group process let us examine the violations listed under "Surface Water," Jackson Turbidity Units (JTU), etc., all relate to the quality of the water discharged from the mining site. These violations could be grouped into a single violation types. However, to do so would be to ignore the fact that there are two very different components of water quality which can be resolved from the data contained on the Data Entry Form. These components are chemical water quality and physical water quality.

Iron concentration, pH, acidity, and alkalinity are measures of chemical water quality. We have used only pH and (Fe) since pH and acidity and alkalinity are, to a certain extent, redundant. Another reason for excluding acidity and alkalinity is that there was ambiguity concerning the tests for these parameters.

JTUs and the presence of settleable matter are measures of physical water quality. Both parameters were used.

By grouping the violations under "Surface Water" in this way, the inspection data could be used to determine the frequency of occurrence of these two (2) violation types and also the frequency of joint occurrence with each other and with other violation types. The frequency of joint occurrence is important since it is believed that for Landsat, physical water pollution may be "visible" while chemical water quality may be "invisible." (In fact, the existence of the former may mask the presence of the latter.) However, using the frequency of joint occurrence, it seemed possible to make inferences as to the existence of "invisible" chemical water quality violations based on "visible" physical water quality violations.

Similar reasoning was used in forming other violation types. By grouping violations in this way it was hoped that it would be possible to enhance Landsat's capability to detect "invisible" violations by detecting jointly occurring "visible" violations. Another advantage of grouping the violations into types was that it reduced the number of variables which were manipulated in the statistical analysis of the inspection data.

The violation types were further aggregated into four (4) broad categories (see Table B'2). These categories were based upon the relationships among the violation types. The water quality category, for example, includes not only the chemical and physical violation types but also the silt structures and drainage plan violation types. These were included in this category since properly designed, constructed and maintained structures are required for water treatment. The water impoundment violation type was also included in this category since unauthorized impoundments were believed to be likely sources of chemically polluted water.

TABLE B'1

ILLUSTRATIVE DATA ENTRY FORM WITH DEFINITIONS

PERMIT NUMBER: 244-72

NAME OF OPERATION: [illegible] – [illegible]

DATE OF INSPECTION:	11/1/72	1/21/73	5/9/73	8/1/73	9/21/73	10/24/73	11/8/73	10/11/73	12/3/73	2/6/74		VIOLATION TYPE
ACCESS ROAD:												ACCESS ROAD
GRADING AND BACKFILLING:												
Current												GRADING CURRENT
To Plan												GRADING TO PLAN
VEGETATION:												
SMR-RG-8						✓	✓		✓			VEGETATION REGULATION
Current						✓	✓		✓			VEGETATION CURRENT
SOIL SAMPLE:												
pH												NOT A VIOLATION
Sulfide												NOT A VIOLATION
DRAINAGE:												
Plan				✓	✓			✓				DRAINAGE PLAN
Structure												SILT STRUCTURE
Functioning												DRAINAGE PLAN
UNDERGROUND DISCHARGE pH:												NOT A VIOLATION
SURFACE WATER:												WATER QUALITY CHEMICAL:
pH						6			7.6	8.5		IF pH < 6 OR pH > 9
Fe						7			7	[illegible]		IF [Fe] > 7
JTU						[illegible]						W.Q. PHYSICAL IF JTU > 150(1)
Acidity						[illegible]				70		*
Alkalinity						[illegible]				[illegible]		*
Settleable Matter												WATER QUALITY PHYSICAL
WATER IMPOUNDMENT:	✓	✓		✓	✓	✓						WATER IMPOUNDMENT
METHOD OF OPERATION:												METHOD OF OPERATION
SILT STRUCTURE(S):												SILT STRUCTURE
DISCREPANCIES:					✓	✓	✓	✓	✓			DISCREPANCIES

NON-COMPLIANCES: (1) 10/24/73 Leakage of water with high iron
(2) 10/11/73 " " " " " "

(1) EXCEPT FOLLOWING PRECIPITATION

NOTE: "√" DENOTES A VIOLATION.
* DENOTES PARAMETER NOT USED DUE TO REDUNDANCY AND AMBIGUITY

SUSPENSIONS:

INSPECTOR(S):

REPRODUCIBILITY OF THE
ORIGINAL PAGE IS POOR

TABLE B'2

VIOLATION TYPE BY VIOLATION CATEGORIES

VIOLATION CATEGORY	VIOLATION TYPE
METHOD OF OPERATION	ACCESS ROAD GRADING CURRENT GRADING TO PLAN METHOD OF OPERATION
VEGETATION	VEGETATION REGULATION VEGETATION CURRENT
WATER QUALITY	DRAINAGE PLAN SILT STRUCTURE WATER QUALITY CHEMICAL WATER QUALITY PHYSICAL WATER IMPOUNDMENT
DISCREPANCIES	DISCREPANCIES

APPENDIX C

COMPUTER PROCESSING OPERATIONS FOR KENTUCKY STRIP MINE LANDSAT PROJECT

Diana L. Rebel
ERIM

Prepared For:

Commonwealth of Kentucky
Department for Natural Resources
and Environmental Protection
Office of Planning and Research
Frankfort, Kentucky 40601

By:

Environmental Research Institute
of Michigan
Ann Arbor, Michigan

Computer Processing Operations for Kentucky Stripmine Landsat Project

The flow of processing operations is shown in Figure 1. The beginning point for analysis was Landsat-2 CCT data in 7 track 800 bpi format. The first step in processing was to examine the data using the LIGMALS software package on the University of Michigan Amdahl 470V computer (ref 1 and 2). From this examination, a qualitative impression of data quality was obtained, level assignments determined for graymaps to be produced later, and dark levels in each band determined for later processing.

For the 30 October 1975 data (scene 2281-15465), the quality of the data in MSS channels 6 and 7 was very good. MSS-4 data had a pronounced striping pattern every sixth line. Some slight striping also existed in MSS-5. Prints of these 4-bands are included in Appendix D.

The dark level correction mentioned is an attempt to account for the additive effects of atmospheric conditions by determining what this factor is in each channel and subtracting it. In the absence of instrumentation to measure this we determined the lowest signal in each channel in an area where low reflecting objects ("blackbodies") occurred. Since the signal from a blackbody would be zero if there were no path radiance, we assumed that the difference between the signal we received from our approximations to blackbodies and zero was a measure of the path radiance. For the 30 October 1975 data, the values we determined for MSS-4 thru MSS-7 were 8, 5, 1, and 0, respectively.

The next processing steps were format conversion from Landsat format to a format compatible with ERIM computers, followed by implementation of the dark level correction and then data rotation and

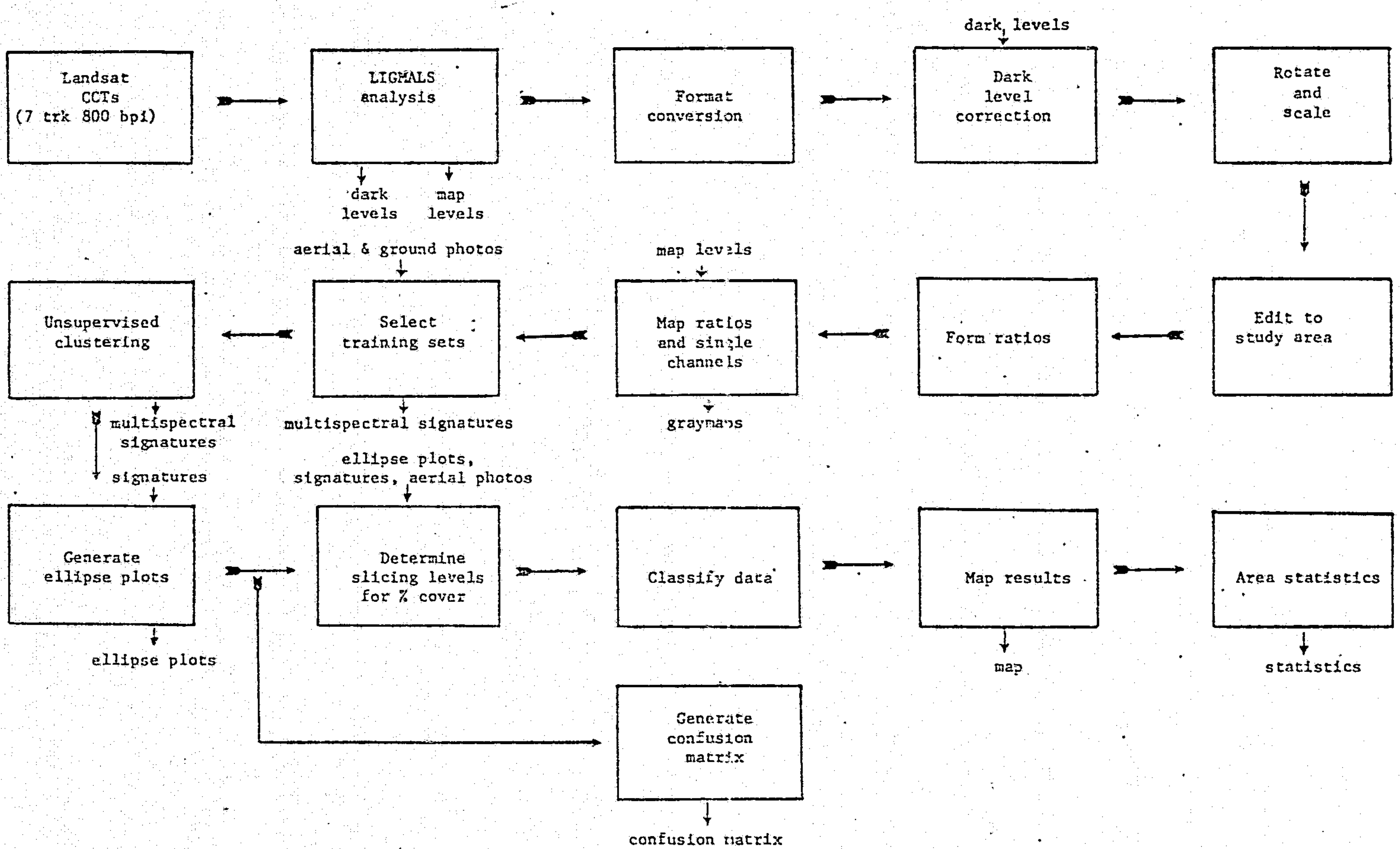


FIGURE 1. FLOW OF COMPUTER PROCESSING OPERATIONS FOR KENTUCKY STRIPMINE LANDSAT PROJECT

scaling. The rotation and scaling is needed to reduce effects of earth rotation during the time it takes to scan a scene, to rotate the data so scan lines run east-west, and to adjust the number of data points by nearest neighbor interpolation so that computer line printer maps would have a scale 1:24,000.

Following rotation and scaling, the data were edited to the study area through specification of the vertices of the study area (in Landsat line and pixel coordinates). A separate tape of the study area data was made. Then four ratio channels (MSS-5/MSS-4, MSS-5/MSS-6, MSS-7/MSS-6, MSS-7/MSS-5) were added to the four Landsat bands through further computer processing. At this time, graymaps (scale 1:24,000) were prepared of four Landsat bands and four ratios. These graymaps constituted one output product, (Appendix D).

Based on the ground and aircraft photography and the ultimate terrain classification categories desired, we selected several areas representing different types or conditions of materials to use as training sets. These areas were carefully located on an MSS-5 graymap (we attempted to avoid mixture or boundary pixels), and signatures were extracted using the STAT. program. Each multispectral signature is a statistical description of a group of data points (pixels). It contains the mean value of the signal in each channel and the covariance matrix, from which the standard deviation of the signal and the correlation between each pair of channels may be calculated. Each signature was derived using 8 channels: the four MSS channels and the ratios MSS-5/MSS-4, MSS-5/MSS-6, MSS-7/MSS-6, and MSS-7/MSS-5.

To complete the training process, we next used unsupervised clustering. Five rectangles of data were selected which appeared to contain samples of everything in the scene. The clustering algorithm was applied in two sweeps through the data: first, looking at every fourth line and every fourth point in all five rectangles, then back again looking at every pixel in all rectangles (26167 pixels looked at in total). An upward limit of 30 clusters had been specified and the two passes through the data were done in order to avoid biasing the clusters toward the materials in one rectangle. The clustering was performed on the eight channels mentioned previously and a multispectral signature was generated for each cluster. Our main objective in running CLUSTR. was to avoid missing any significant categories. In addition, clustering often produces signatures which encompass the characteristics of a class over a large area better than a few training set generated signatures. For the 30 October 1975 data twenty-two acceptable clusters were generated.

Plots of the distribution of the twenty-two cluster derived signatures in two channel hyperspace (MSS-5, red and MSS-7, IR) were compared with similar plots of the 33 training set derived signatures. This enabled us to assign names (classes) to the cluster signatures. Ellipse plots of the signatures used in the classification of this data set are shown in Figure 2. The distribution of each signature class is represented as an ellipse whose boundary is a constant probability of one X^2 distance from the mean. (In the final CLASFY. program which produced the recognition results a X^2 value of 99.99 was used.)

All but two of the signatures used in the final classification were cluster derived. Signatures for the final classification were chosen on the basis of what class they represented and their separability from other signatures representing other classes. Ellipse plots and confusion matrices (similar to Table 1) were used to help determine this separability.

We also investigated the slope/aspect situation; i.e., the differences in signal received by the sensor due to differences in irradiance. Problems arise when the same material lies on areas

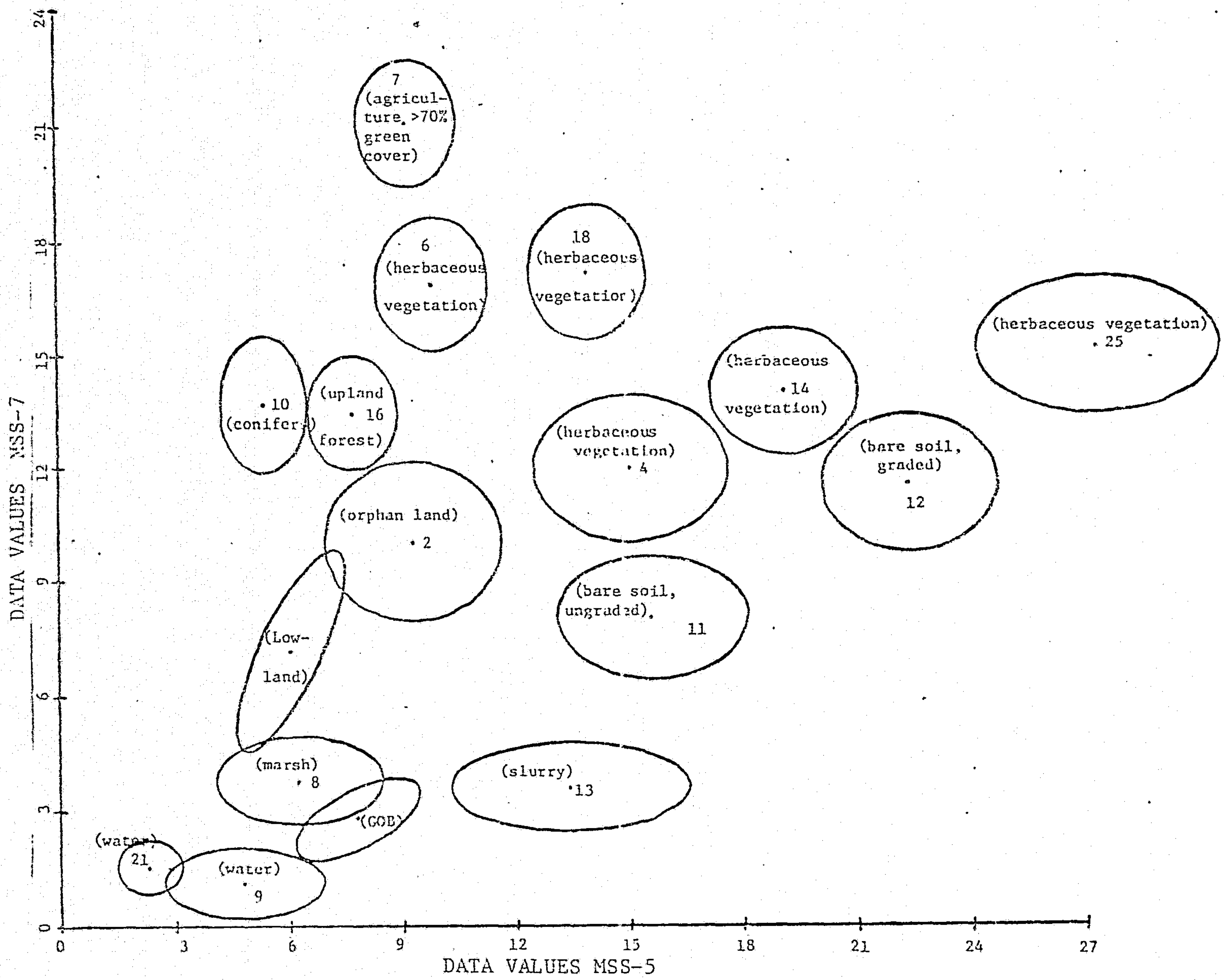


FIGURE 2. ELLIPSE PLOTS OF THE SIGNATURES USED IN THE CLASSIFICATION OF THE 30 OCTOBER 1975 DATA SET (OBSERVATION 2281-15465) IN TWO CHANNEL SPACE (MSS-5 AND MSS-7). The boundary of each ellipse represents a constant probability of one χ^2 distance from the mean.

Table 1. CONFUSION (EXPECTED-PERFORMANCE) MATRIX BASED ON THE SEVENTEEN SIGNATURES USED FOR THE FINAL CLASSIFICATION OF THE WESTERN KENTUCKY STUDY SITE. Rows represent distributions based on those signatures; columns represent the recognition classes. (Each distribution consists of 1000 points per signature taken at random and distributed according to the multivariate normal distribution specified by the signature.) Numbers are in percent and give the probability that pixels from each signature distribution will be classified into each recognition class. Dashes indicate zero percent probability. The classifier used in producing the matrix was the best linear rule classifier, that used in the final classification.

	water/21	water/9	marsh/8	gob	slurry/13	lowland forest	upland forest/16	conifers/10	orphan land/2	bare soil (ungraded)/11	bare soil (graded)/12	agriculture >70% green cover/7	herbaceous vegetation/25	herbaceous vegetation/4	herbaceous vegetation/6	herbaceous vegetation/14	herbaceous vegetation/18	unclassified
water/21	98.7	0.8	0.4	---	---	0.1	---	---	---	---	---	---	---	---	---	---	---	---
water/9	1.1	93.9	0.7	0.2	0.2	---	---	---	---	0.1	---	---	---	---	---	---	---	3.8
marsh/8	0.4	0.5	87.8	0.2	0.1	0.3	---	---	---	0.7	---	---	---	---	---	---	---	10.0
gob	---	9.6	3.3	83.1	3.7	---	---	---	---	0.3	---	---	---	---	---	---	---	---
slurry/13	---	---	---	0.2	95.8	---	---	---	---	0.4	---	---	---	---	---	---	---	3.6
lowland forest	---	---	4.8	---	---	81.8	2.4	0.4	10.1	---	---	---	---	---	0.5	---	---	---
upland forest/16	---	---	---	---	---	0.2	95.5	1.6	1.1	---	---	---	---	---	0.4	---	---	1.2
conifers/10	---	---	---	---	---	---	1.8	93.1	0.1	---	---	---	---	---	---	---	---	---
orphan land/2	---	---	0.2	---	---	0.5	1.2	---	89.6	0.2	---	---	---	1.5	0.1	---	---	6.7
bare soil (ungraded)/11	---	---	---	---	0.1	---	---	---	---	98.2	0.4	---	---	1.3	---	---	---	---
bare soil (graded)/12	---	---	---	---	---	---	---	---	---	0.3	97.3	---	1.3	---	---	1.1	---	---
agriculture >70% green cover/7	---	---	---	---	---	---	---	---	---	---	---	98.3	---	---	1.7	---	---	---
herbaceous vegetation/25	---	---	---	---	---	---	---	---	---	---	2.3	---	97.6	---	---	0.1	---	---
herbaceous vegetation/4	---	---	---	---	---	---	---	---	1.3	2.3	0.1	---	---	90.9	---	5.0	0.4	---
herbaceous vegetation/6	---	---	---	---	---	---	1.2	0.2	0.1	---	---	0.8	---	---	96.8	---	0.9	---
herbaceous vegetation/14	---	---	---	---	---	---	---	---	---	0.1	1.1	---	0.1	4.4	---	94.3	---	---
herbaceous vegetation/18	---	---	---	---	---	---	---	---	---	---	---	---	---	0.6	0.7	0.1	98.6	---

C-2

with significantly different slope and aspect such that the irradiance on the material, and hence the radiance received by Landsat is significantly different. Thus a single signature will not be able to recognize the material on all slopes and aspects, and in fact, one material with a particular slope and aspect may look like another material with a different slope and aspect.

Such problems were encountered in this area. For example, one material with a northerly-facing slope had a mean level digital count of 10.8 in MSS-5 whereas the same material on a more southerly-facing slope had a digital count value of 18.1, a 68% change in mean level. Such a situation causes grave problems in spectral recognition.

One method that has been used to ameliorate the effects of varying irradiance due to such factors as varying slope and aspect is to establish signatures using ratios of 2 spectral channels of digital data. The resulting ratios are generally less susceptible to variation in slope and aspect because the magnitude of the irradiance changes tend to be correlated between the spectral bands.

Due to the significance of the slope-aspect problem in this area we investigated the utility of ratios. For the same area for which a single red band (MSS-5) varied 68%, an MSS-7/MSS-5 ratio varied by 34%. This is not complete normalization, but it is obviously an improvement.

Unfortunately, ratioing of channels frequently causes a loss of information content and sometimes causes a loss of ability to discriminate between materials that are differentiable using individual channels of data. Under the circumstances, therefore, we felt it was best to do our classification using signatures derived from <u>both</u> individual channels and ratios of channels. The hope was that the resulting classification would embody some of the beneficial aspects of both approaches.

One of the major goals of this project was to map the location of reclaimed, regraded and re-vegetated strip mine areas where the vegetation cover had reached greater than 70%. Previous experience has indicated that an infrared/red reflectance ratio is highly effective at discriminating between various classes of green vegetation cover, generally better than ocular estimates made by observers on the ground. In addition, an IR/red reflectance ratio possesses the advantages of ratios discussed earlier. Furthermore, an IR/red reflectance ratio has been found to be a good normalizer of reflectance differences between different soil types and surface soil moistures. Therefore, we decided to use an MSS-7/MSS-5 ratio to differentiate between low (<70%) and high (>70%) green vegetation cover.

Ground photos and field notes as well as the aerial photos were used in deciding on appropriate levels for slicing the MSS-7/MSS-5 ratio. The field notes classified vegetation as being greater or less than 50% cover, and apparently referred to total (live and dead) vegetation cover. In addition, we have found ocular estimates of vegetation cover to be rather consistently too high. Therefore, our decision on an appropriate MSS-7/MSS-5 ratio slicing level was heavily dependent on the color IR aerial photos. Although it is rather difficult to estimate percent green vegetation cover on this scale of aerial photography, it did afford the advantage of a truly vertical perspective (for which % cover is defined), a synoptic view, and high sensitivity to amount of green vegetation which is a characteristic of color IR film. When we had picked training sets and computed their MSS-7/MSS-5 ratios, we compared them with the cluster signatures. The ratio we had picked to separate >70% green vegetation from <70% green vegetation (R=1.5) fell between the ratio values for two large clusters. This was a fortuitous, but beneficial result. A lower limit was also selected for the ratio so that areas with negligible green vegetation cover (<10%) would be placed in the bare soil category.

When ratio values for percentage cover had been selected the data from the test area were classified using 17 signatures and a level slice of MSS-7 for water recognition. For points classified as herbaceous vegetation an estimate of the percentage cover was made using the ratio slicing previously discussed. Results were then displayed as vegetation with 0-10, 10-70, and >70% cover rather than as individual vegetation classes.

A color coded map (Appendix D) was prepared using a computer line printer and various colored printing ribbons. At the same time area statistics were developed. These statistics, shown in Table 2, are the acreages of the various classes in the test area, as recognized by the computer. For reference, the percentage composition of the area is also shown.

The thirteen classes of the final recognition map for the 30 October 1975 data were obtained from the 17 training sets by combining the two water classes into one symbol for display, by telescoping the five herbaceous vegetation classes into 3 cover classes (as previously discussed), and by combining the 0-10% cover class with the graded bare soil class (Table 3).

As an aid to interpreting the results of classification and to understanding how areas might be misclassified, a confusion matrix was generated using the 17 final signatures and samples of data drawn from these assumed Gaussian signatures. The samples of data were classified according to the decision rule used in the classification program. The results, presented as Table 1, are not precisely indicative of the accuracy and performance of the classifier over a large area (because only data from training sets is examined), but do offer some guidance about probable kinds of errors. In Table 1 the percentage of pints in a signature class (each row represents one signature class) classified as a given signature (each column represents points classified as a given signature) is presented.

Table 2.

STATISTICS FOR THE THIRTEEN LAND USE CLASSES OBTAINED FROM RECOGNITION PROCESSING OF THE WESTERN KENTUCKY STUDY SITE. Data from Landsat-2 observation 2281-15465 obtained 30 October 1975.

CLASS	ACREAGE	PERCENT OF TOTAL AREA
water	1809.50	1.49
marsh	1930.50	1.59
lowland forest	13103.45	10.80
gob	584.87	.48
slurry	351.82	.29
upland forest	18829.97	15.52
conifers	1247.04	1.03
orphan lands	36077.90	29.75
bare soil (ungraded)	4229.63	3.49
bare soil (graded)	10377.44	8.56
>70% green herbaceous cover; probably agriculture	3780.34	3.12
10-70% green herbaceous cover	18269.76	15.06
>70% green herbaceous cover	10693.40	8.82
	121285.62	100.0

Table 3.

THE THIRTEEN CLASSES OF THE FINAL RECOGNITION MAP OF THE WESTERN KENTUCKY STUDY SITE AND THEIR DERIVATION. Data was from Landsat-2 observation 2281-15465 obtained 30 October 1975. Numbers and names refer to specific signatures. (See Figure 2 and Table 1.)

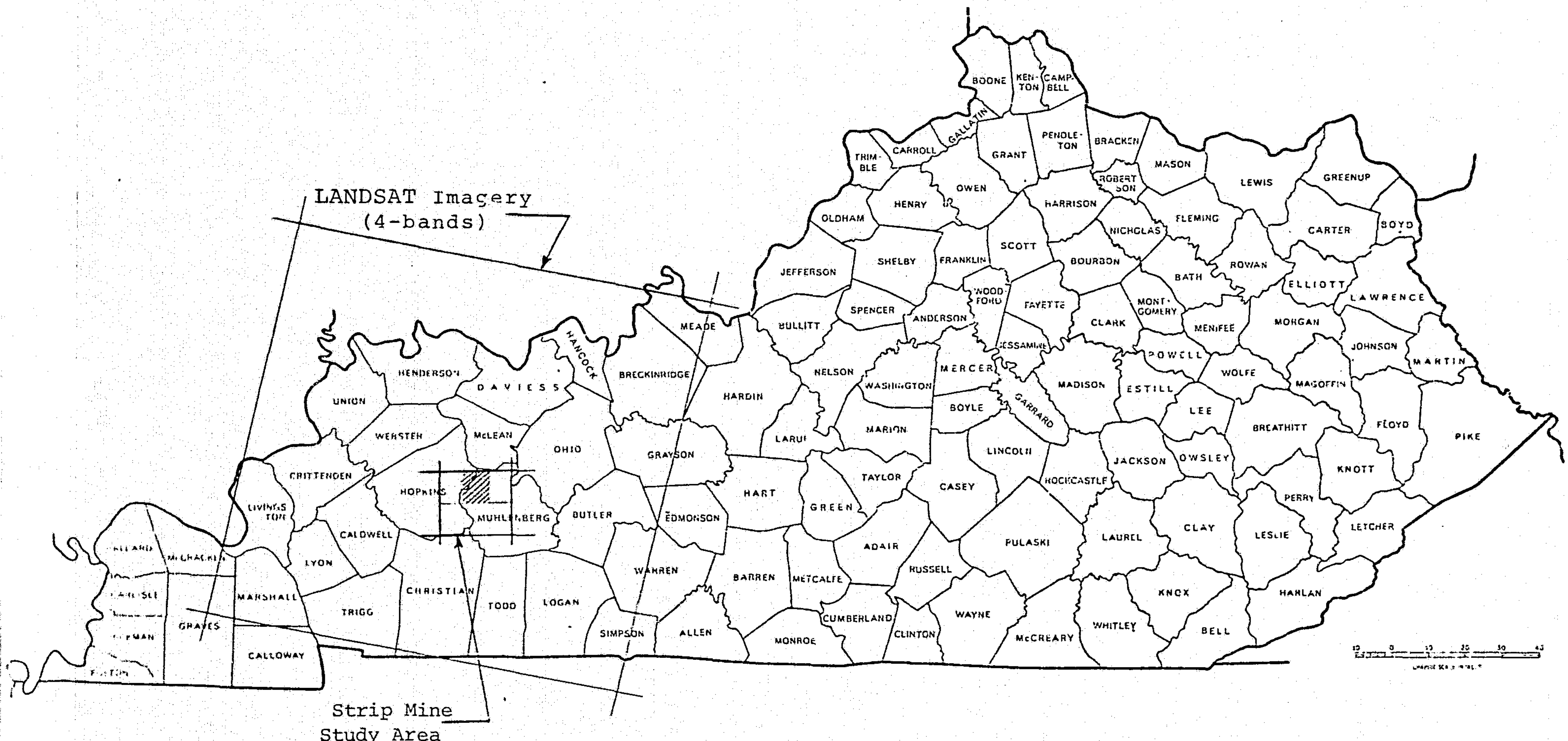
CLASS	DERIVATION
water	level slice of MSS-7; this included all data points classified under sig. no. 21 and 9 and some points classified under sig. no. 8, GOB, SLURRY and LOWLAND
"marsh" (may be shrub swamp or some other wetland type)	8
gob	GOB
slurry	13
lowland forest	LOWLAND
upland forest	16
conifers	10
orphan land	2
bare soil (ungraded)	11
bare soil (graded)	12 (and MSS-7/MSS-5 level slice of sig. no. 25, 4, 6, 14, 18)
>70% green herbaceous cover; probably agriculture	7
10-70% green herbaceous cover	MSS-7/MSS-5 level slice of sig. no. 25, 4, 6, 14, 18
>70% green herbaceous cover	MSS-7/MSS-5 level slice of sig. no. 25, 4, 6, 14, 18

REFERENCES

1. Wagner, H. L. 1976. An Interactive Digital Multispectral Data Processing Software Package. Remote Sensing of Earth Resources, Vol. V. The University of Tennessee Space Institute, Tullahoma.

2. Wagner, H. L. 1976. Feasibility Study of an Interactive System for Processing Digital Multispectral Scanner Data. Master's Thesis. The University of Michigan, Ann Arbor.

KENTUCKY


LANDSAT Imagery
(4-bands)
Strip Mine
Study Area
BOONE
KEN-TON
CAMP-BELL
GALLATIN
TRIM-BLE
CARROLL
GRANT
PENDLE-TON
BRACKEN
MASON
LEWIS
GREENUP
OWEN
HENRY
ROBERT SON
HARRISON
FLEMING
CARTER
BOYD
OLDHAM
SCOTT
NICHOLAS
JEFFERSON
SHELBY
FRANKLIN
BOURBON
BATH
ROWAN
ELLIOTT
LAWRENCE
SPENCER
ANDERSON
WOOD FORD
FAYETTE
CLARK
MONT-GOMERY
MENIFEE
MORGAN
MEADE
BULLITT
JOHNSON
HANCOCK
HENDERSON
DAVIESS
BRECKINRIDGE
NELSON
MERCER
POWELL
WOLFE
MAGOFFIN
MARTIN
UNION
HARDIN
WASHINGTON
MADISON
ESTILL
BOYLE
GARRARD
LEE
WEBSTER
McLEAN
MARION
BREATHITT
FLOYD
PIKE
OHIO
GRAYSON
LINCOLN
JACKSON
OWSLEY
CRITTENDEN
TAYLOR
KNOTT
HOPKINS
HART
CASEY
ROCKCASTLE
PERRY
LIVINGSTON
MUHLENBERG
BUTLER
EDMONSON
GREEN
CALDWELL
ADAIR
PULASKI
LAUREL
CLAY
LESLIE
LETCHER
McCRACKEN
LYON
WARREN
RUSSELL
CHRISTIAN
BARREN
METCALFE
KNOX
HARLAN
MARSHALL
TODD
LOGAN
TRIGG
WAYNE
GRAVES
CUMBERLAND
SIMPSON
ALLEN
WHITLEY
BELL
CLINTON
MONROE
McCREARY
CALLOWAY

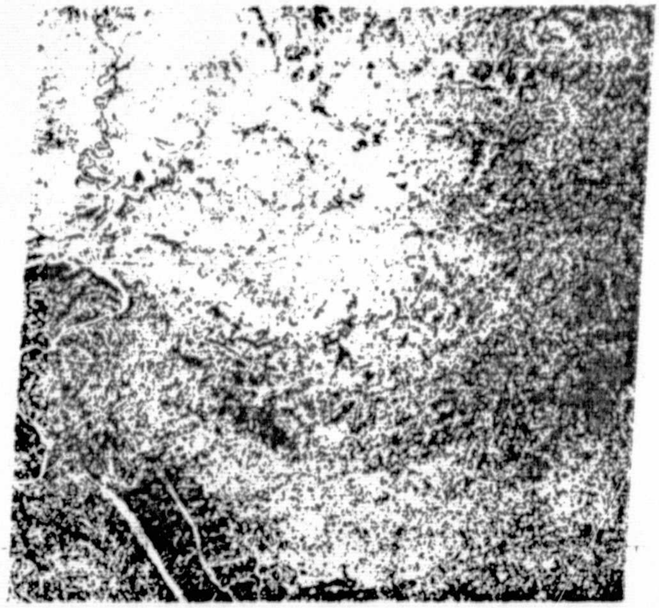


Fig. 1(a) Band 5

Fig. 1(b) Band 6

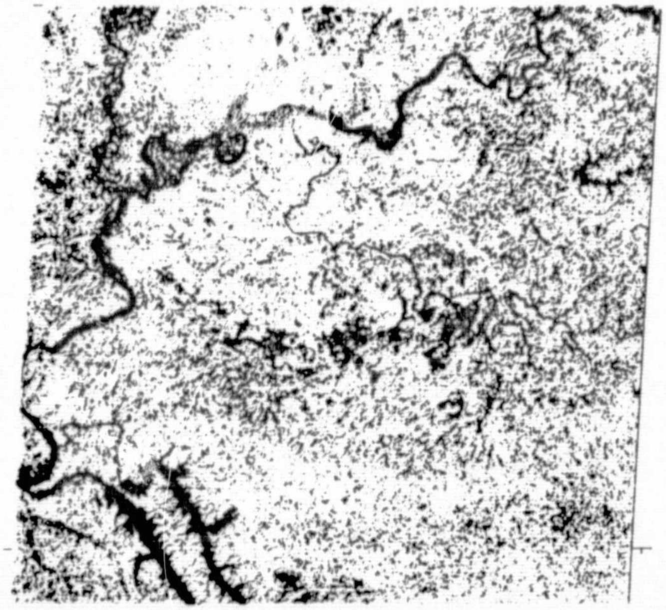


Fig. 1(c) Band 7

Figure 1
Landsat-2 imagery of Western Kentucky coal field and study area.

Figure 2

High altitude aerial, color infra-red photograph of the test area, east of Madisonville, Kentucky. Note "anchor" lake in lower center.

Figure 3
Low **altitude, color infra-red**, aerial photograph of the test area, **east** of **Madisonville, Kentucky.** Note "anchor" lake at upper left.

Figure 4
Low altitude, black and white, aerial photograph of the test area east of Madisonville, Kentucky, showing areas identified for ground truth. Note "anchor" lake in upper center.

Figure 5
(Area 109 of Fig. 4)
Strip mined and regraded to low rolling hills. Fescue and clover with mixed hardwood and pine saplings. Less than 50% ground cover. A- is an oblique view, and B- is a vertical view of the surface.

A

B

Figure 6
(Area 110 of Fig. 4)
Strip mined and regraded to rolling pasture. Fescue and alfalfa predominate. Nearly 100% ground cover with a few bare spots. A- is an oblique, and B- is a vertical view of the surface.

A

B

Figure 7
(Area 112 of Fig. 4)
Stripped and strike-off graded to long flat-topped ridges. Mixed scrub hardwoods.

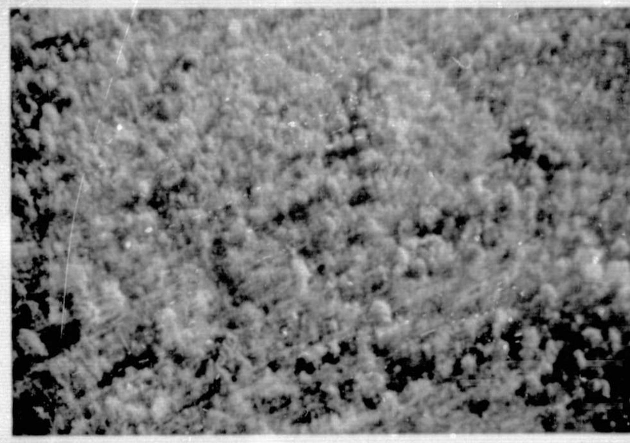


Figure 8
(Areas 113 and 114 of Fig. 4)
Unmined forested area west of "anchor" lake. Mixed lowland hardwoods with scattered shrubs and leaf litter floor.

Figure 9
(Area 115 of Fig. 4)
Unmined agricultural area southwest of "anchor" lake. Uniformly brown soybean field.

Figure 10
(Area 116 of Fig. 4)
Old slurry pond south of "anchor" lake. Fine-grained coal refuse from a coal washing facility.

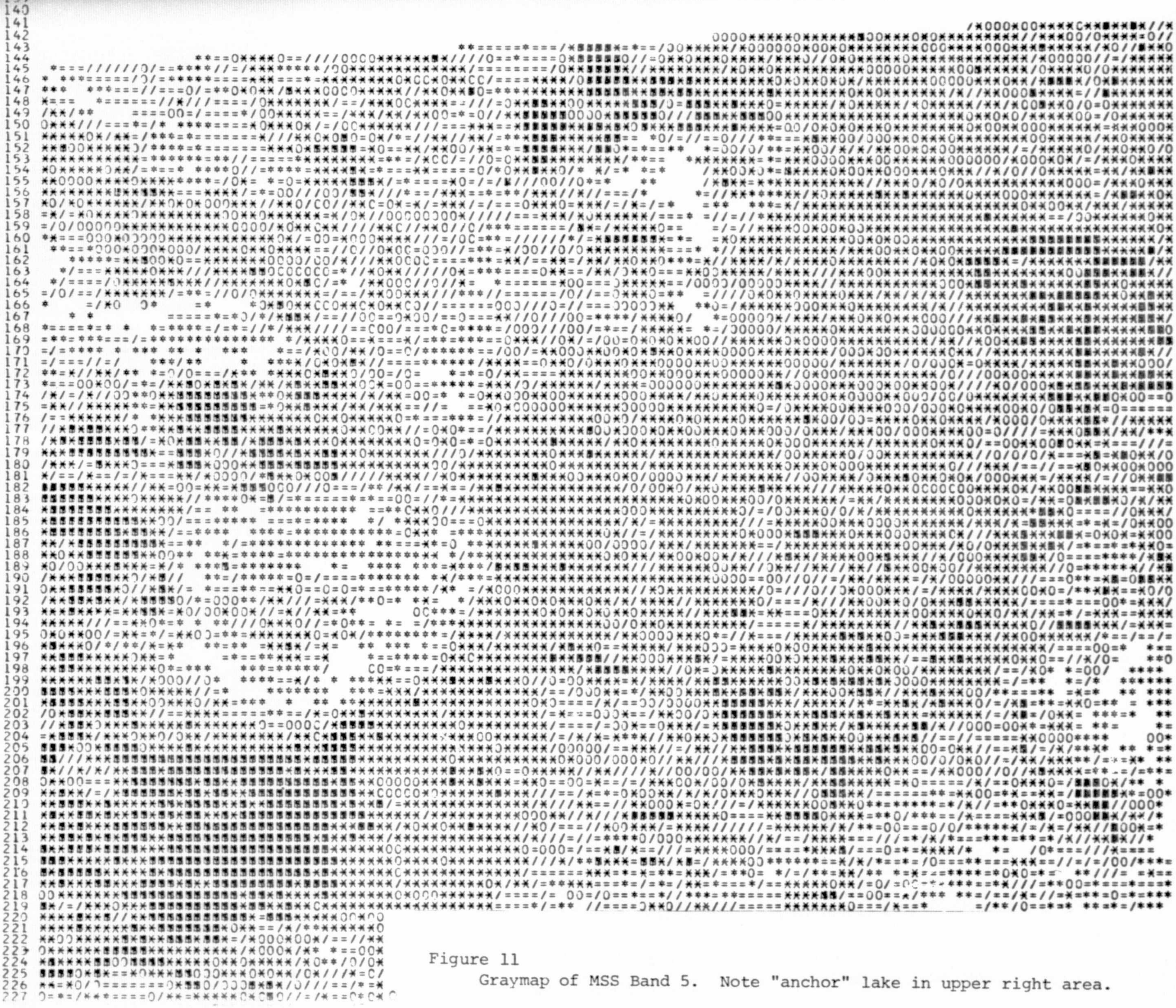


Figure 11
Graymap of MSS Band 5. Note "anchor" lake in upper right area.

Figure 12

Graymap of ratio of MSS Band 7 to MSS Band 6. Note "anchor" lake in upper right area.

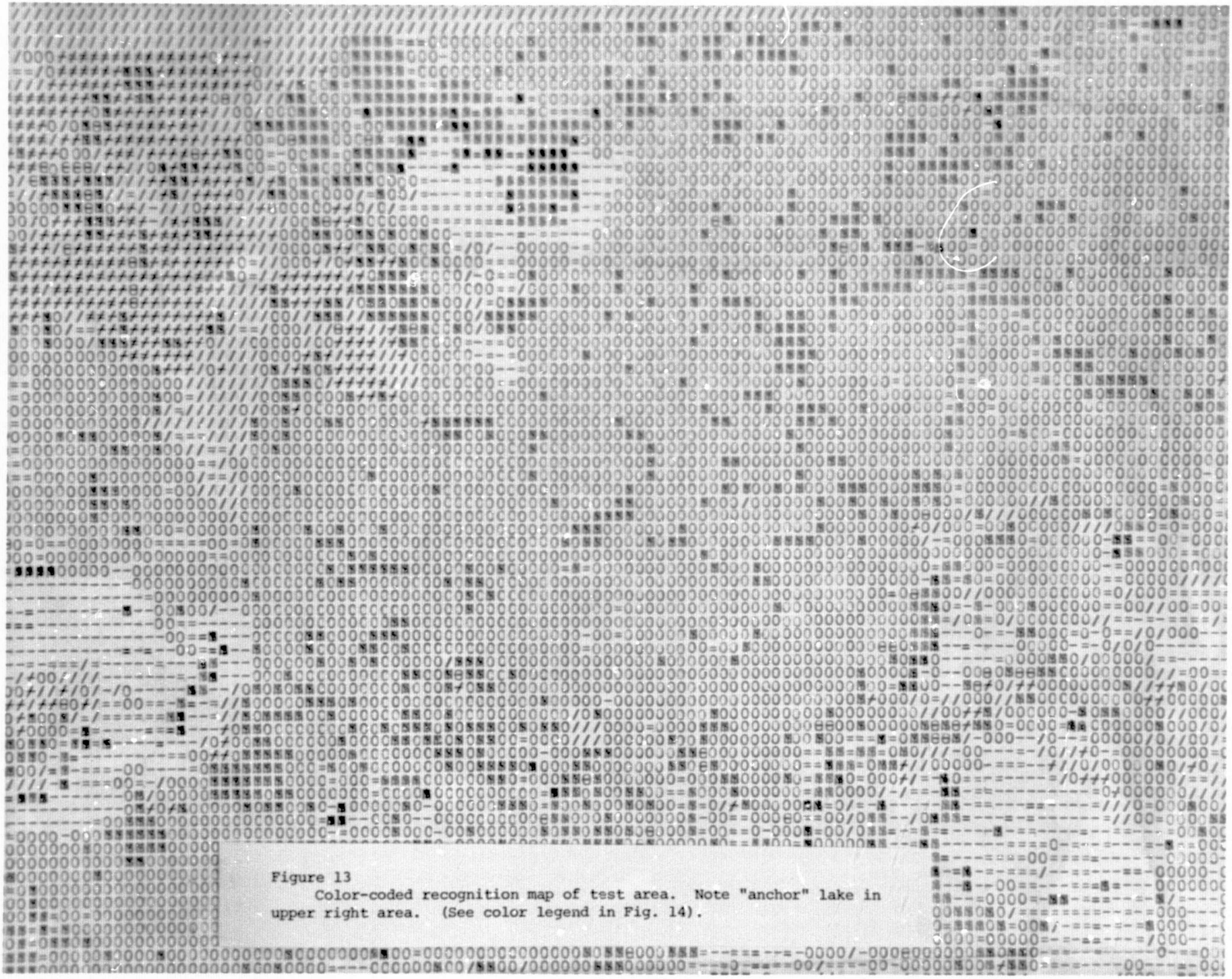


Figure 13
Color-coded recognition map of test area. Note "anchor" lake in upper right area. (See color legend in Fig. 14).

NAME	SYMBOL	NUMBER
WATER	■	477
WATER	■	0
WATER	■	0
MARSH	=	591
GOB	■	185
SLURRY	=	102
LOWLAN	O	4753
UPLAND	■	3256
CONIFR	θ	174
ORPHAN	O	9336
BARE-U	-	1478
BARE	=	332
AG-70+	≠	584
0-10	=	1933
10-70	/	5870
70-100	⌿	1679
		0
		0
NOT CLASSIFIED		0
REJECTED		0

Figure 14

Color legend for recognition map of Fig. 13. Number refers number of pixels in each category within the test strip.