Issues Arising from Sampling Designs and Band Selection in Discriminating Ground Reference Attributes Using Remotely Sensed Data

M. L. Labovitz
Scientific Systems Division, SASC Technologies, Inc., Lanham, MD 20706

ABSTRACT: Certain issues which relate to sampling and band selection are studied within the framework of examining spectral bands as surrogate measures of a single ground reference plane (or combinations of ground reference planes) through the development of classifers. In particular, the effect upon classification accuracy of (1) the spacings between samples forming the training set and (2) the commonly used stepwise discriminant method for band selection, are examined. The analysis was conducted on a multivariate data set of ground reference attributes for the Parker Quadrangle, Colorado (located southeast of Denver) and an MSS scene registered to it. Four attributes of the data set—land form, land use, slope, and vegetation as well as combinations of these attributes—were classified using the MSS bands. Initially, a set of spectrally distinct ground-over classes was produced for each combination of ground-plane attributes using stochastically independent observations. The purpose of this was to provide a standard of comparison for the remainder of the analysis. By standardizing on the sample sizes, the numbers of ground attribute classes, and the class composition, the effects upon ground attribute classification accuracy of the number and combination of spectral bands used and the spacing between training statistics were systematically studied. It was found that, if training samples were closely spaced, the commonly used Jackknife procedure for estimating classification accuracy was optimistically biased due to autocorrelation in the observations. This bias leads analysts to think they are doing better than they actually are. This same autocorrelation effect clearly contributes to the well known inability in remote sensing research to extend signatures developed in one location to other nearby locations. Thus, simple altering of sample spacing can account for a 25 percent improvement in the classification accuracy. In a further analysis, the results of Cover and Van Campenhout are shown to apply to remote sensing data. These two authors showed that stepwise procedures may yield suboptimal subsets of features. In a comparison between a stepwise band selection strategy and an "all-possible-subsets" band selection strategy, use of the former almost always resulted in a suboptimal set of bands averaging as much as 10 percent.

INTRODUCTION AND MOTIVATION

FROM A GENERIC or system level perspective, the investigation of remotely sensed data for the extraction of signatures and the development of thematic maps produces a host of correlative issues.

> *Utility of remote sensing data*. How "good" are the remote sensing spectral bands as surrogate measures of ground attributes, e.g., biomass mapping, crop inventorying, and lithologic mapping? If there is more than one attribute available (particularly if they are registered to one another), then the initial question has a combinatorial nature, i.e., are there individual or combinations of spectral bands which can be used as surrogate measures of individual or combinations of ground reference planes.

An answer to this question requires the investigator to develop a classifier which includes two related steps: (1) the selection of bands and (2) the estimation of training statistics. The thrust of this paper is the examination of the inadequacies of the procedures commonly used to perform these steps in a remote sensing analysis. These analysis procedures have been used by remote sensing scientists in all disciplines; consequently, they have a generic importance. Because of their close relationship to one another, the analyses of the two stages are presented in one paper, although in different sections. The background for examining each stage is as follows:

> *Optimality of bands*. What is the optimal set of bands from amongst all the spectral bands available to discern a particular phenomenon? To answer this question, many experimental situations require the researchers

PHOTOGRAMMETRIC ENGINEERING AND REMOTE SENSING,
Vol. 52, No. 2, February 1986, pp. 201–211.

0099-1112/86/5202-201$02.25/0

to develop a strategy for selecting bands from the set of bands available.

Sometimes the selection of the subset of bands is dictated by the prevailing theory. But more often some type of empirical strategy is adopted. A very popular type of empirical strategy is the stepwise strategy wherein an algorithm iteratively adds and/or removes one band at a time based upon an R^2 or F statistic criterion (Draper and Smith, 1981). A stepwise procedure will yield an unambiguous optimal subset if the bands are independent. If the assumption of independence is not true, then, as Cover and Van Campenhout (1977) have shown, a stepwise discriminant or stepwise regression may not yield the optimal subset. Further, the degree to which the set is suboptimal is not bounded. For example, under the data structure just described, the two best individual bands will very likely not be the best pair, the best pair may not form part of the best triple, and so forth. The implication of these results are that a stepwise strategy may yield as the best subset a suboptimal set whose departure from optimality is not bounded (Van Campenhout, 1980). If the collinearity of independent variables results in the behavior just described, the only protection against bias is to pursue an "all-possible-subset" strategy (Cover and Van Campenhout, 1977). For MSS such a strategy would require the research to perform an expensive but tractable $2^4-1=15$ different discriminant analyses. However, there would be a more severe impact on research using a sensor such as TM, which would require 2^7-1 or 127 discriminant analyses.

It is known that the degree of collinearity in MSS bands is high, as evidenced by the results of principal components analyses of MSS data where the bands load almost entirely on the first two components (Landgrebe, 1978). Therefore, part of this work will examine whether or not the stepwise strategy for band selection is a concern for MSS data; that is, how suboptimal is the subset of bands chosen by this strategy *vis-a-vis* an "all-possible subsets" strategy.

Construction of Training Statistics. Any empirically based analysis, whether or not it involves remotely sensed data, is data limited. This is because the parameters in any empirical model are not knowable and must be estimated by statistics. Statistics in turn are good estimates of population parameters only to the extent that the data are representative or random samples of the population from which they are drawn. In the present context, a ground cover, which may be an individual terrain attribute or combination (through anding, for example) of terrain attributes, is the population. The parameters which we wish to estimate are the mean (μ if scalar, vector $\boldsymbol{\mu}$ if multiband) and variance covariance (σ^2 if scalar, matrix Σ if multiband). Therefore, in order to obtain unbiased estimates, a random sample of pixels must be selected from each ground cover. By definition (Hogg and Craig 1978), a random sample must consist of individuals (pixels) which are identically and independently distributed. A number of authors (see Labovitz and Masuoka (1984) for summary of relevant research) have shown that (1) Landsat pixels are not independent, that the dependence is highest for adjacent pixels and approaches independence with increasing distance between samples, and 2) when calculated using contiguous pixels use of $\hat{\sigma}^2$ and $|\hat{\boldsymbol{\Sigma}}|$ as the common estimator of σ^2 and $|\Sigma|$, respectively yields an underestimation bias ($|\bullet|$ is the determinant operator). Because σ^2 and $|\Sigma|$ are measures of dispersion, their underestimation means that the population will be perceived to vary less about its centroid than it actually does. Use of these biased estimators will lead the researcher to believe that the populations are more separable than they actually are. Labovitz and Masuoka (1984) have shown that this bias results in inflated F statistics values. A third important issue then to be examined in this paper is the influence of sampling, training, and testing strategies upon the production of terrain classifications.

STUDY LOCATION

The study area from which the data were acquired is the Parker 7 ½ minute Quadrangle, Colorado. The Parker Quadrangle is located southeast of Denver, Colorado in the Great Plains Physiographic Province. The coordinates of this quadrangle are as follows:

	Latitude	Longitude
Upper Left	39°37′30″	104°52′30″
Upper Right	39°37′30″	104°45′
Lower Left	39°30′	104°52′30″
Lower Right	39°30′	104°45′

DATA SETS

The data sets used can be initially divided into two types—ground data and remote sensing data. Each of these types is, in turn, composed of four properties. The ground data properties are landform, land use, slope, and vegetation. The remote sensing properties are the four MSS bands—MSS 4 through MSS 7. Each property enters the analysis as a data plane. The data planes are arrays of 272 rows by 214 columns for 58,208 pixels with each pixel containing eight numeric value, one for each of the properties displayed in the data plane. The eight planes are composed of four ground planes and four MSS planes registered to one another. This means that the ith, jth element for each data plane represents the same ground location. We therefore may think of each ground location being represented by a vector (length 8) whose elements are the data plane attributes. Each element covers an area of about 1.1 acres (80^2 metres), the size being dictated by the ground resolution of a Landsat pixel.

ANALYSIS ASSUMPTIONS AND CONDITIONS

For this analysis, the following assumptions have been made:

- The MSS vectors for a given class i are distributed normally with mean $\boldsymbol{\mu}_i$ and covariance matrix $\boldsymbol{\Sigma}_i$.
- The classifier used is a Bayes classifier with a 0-1 loss function, i.e., a Bayes test for minimum error. The decision rule under this multiclass situation is to assign the individual to the class which maximizes the negative log likelihood ratio.
- The prior probabilities of the classes used in the anal-

ysis were estimated as the ratio of the number of occurrences of a class in the sample to the total number of observations in the sample.

The following are the conditions under which the analysis was performed:

- For a chosen combination of the ground reference planes, a new set of classes was formed by anding the numeric codes for a given pixel.
- A systematic sampling scheme is used to select samples for the training set.
- A preselected subset of the MSS values associated with each pixel is included in the estimate of the training statistics, i.e., the mean, variance-covariance and the number of observations per class.
- A preprocessing step is performed to screen out classes with insufficient samples to estimate the covariance matrix and its inverse n (per class) $\geq 2^*$ bands. The total number of observations, the number of classes, the prior probabilities, and the pixels actually used in the classification step are based on the screened set of classes.
- Several training and testing strategies are used. These included (a) training and testing on the whole image using the "C" and/or jack-knife (JAK) methods (Fukunaga, 1972); (b) training on a portion of the image and testing on the same portion using "C" and JAK; (c) training on a portion of the image and testing on other portions of the image; and (d) training on a portion of the image and testing on the same portion using "C" and JAK, and then using the same training statistics testing on other portions of the image (STAT).
- The goodness of classification criterion used is classification accuracy, (CA). The formula for CA is given in Fukunaga (1972).
- The stepwise discriminant analysis used the same algorithm as that used in the BMDP7M program (Jennrich and Sampson, 1979).

EXPERIMENTAL STRATEGY

As a first step, a spectrally distinct set of classes, estimated using independent observations, needs to be created. This set of classes and the accompanying classification accuracy serve as a background against which to compare the results. The classes used, their number, and the number of observations in each class affect the training stage, the testing stage, and the misclassification errors. Therefore, applying this information from the background set to all other analyses will remove the number of classes, their composition, and sample sizes as confounding factors.

After creating such a group of spectrally distinct classes for each combination of land cover classes, variation in CA (and TME) calculated by "C" and JAK was studied as a function of the combination of MSS bands, the combination of ground reference plans, and the spacing of samples used in the training statistics. For the individual ground reference planes an additional analysis was performed (1) to relate the CA to the testing strategy, the spacing between the training samples, and the number of bands; and 2) to examine the differences that arise in the CA when a stepwise band selection strategy versus an "all-possible-subsets" strategy is adopted.

SELECTION OF SPECTRALLY DISTINCT SETS OF LAND-COVER CLASSES

In an unpublished preliminary study of the Parker Quadrangle data base, the land-cover classes used were formed simply by "anding" the constituent ground reference planes, and no reduction in the number of classes based upon their spectral distinctiveness was attempted. For the combination of all four ground reference planes the "anding" resulted in a large number of land-cover classes (46 to 50), many of which could not be distinguished by the four MSS bands, and a very high TME resulted. Failing to filter the classes by spectral uniqueness "overwhelmed" the classifier and did not permit us to study the influence of the factors described above. Therefore, as a first step in the analyses, for each of 15 combinations of the ground reference planes, a spectrally distinct set of land-cover classes was created under the following assumptions:

- *Independence.* The degree of sample dependence and the spacing between training samples are related. This means that the set of spectrally distinct classes created by dependent observations is a function of sampling grid spacing. This situation is undesirable because spacing is one of the factors to be studied. Therefore, the initial set of candidate classes for inclusion in the spectrally distinct set was created using samples with a grid spacing of 10. This spacing follows a suggestion by Craig (1979) and still leaves enough data to perform the analysis. Sampling in such a manner has the added advantage that the specific classes and the number of observations per class found with a sample spacing of 10 are very likely to be present in samples from a denser sampling grid.
- *Number of Bands.* The membership in the spectrally unique set of classes is also a function of the number and type of bands used in the analysis. Therefore, in the development of these classes, we assumed that the number of spectrally unique groups would be maximized by using all four MSS bands. Also, by using all four bands, we would act to standardize the number of bands used in developing the unique set of classes. The analysis was conducted for each combination of ground reference planes as follows:

  (a) Sample the population five times, using a grid with a spacing of ten and starting points (1,1), (3,3), (5,5), (7,7), and (9,9).
  (b) Combine the ground reference classes to form new classes, bin the observations, and keep running totals of sums, sums of squares, and sums of cross products.
  (c) Remove from the list of candidates those classes possessing insufficient observations to produce stable estimates of the variance-covariance matrix. For the remaining classes, convert the running totals into means and variance-covariances.
  (d) Estimate the Hotelling T^2 value for each combination of two of the remaining classes.

TABLE 1. NUMBER OF SPECTRALLY DISTINCT CLASSES AND TOTAL NUMBER OF OBSERVATIONS

Combination*	Number of Classes	Number of Observations
1	5	588
2	6	563
3	4	566
4	5	586
1,2	9	369
1,3	8	485
1,4	8	550
2,3	6	295
2,4	7	476
3,4	7	535
1,2,3	6	186
1,2,4	9	317
1,3,4	6	349
2,3,4	7	273
1,2,3,4	7	188

*1 = landform, 2 = land use, 3 = percent slope, 4 = vegetation.

(e) For each pair, compute the probability that F exceeds the value of T^2 or P $(F>T_c^2) \triangleq P$.

(f) Use test due to Fisher (1954) for combining probabilities from independent tests. Under the null hypothesis (H_0), $-2 \ln P$ is distributed as χ^2 with two degrees of freedom. Therefore, for the multiple (NVALS) T^2 probabilities (see a) associated with a given pair of classes,

$$H = -2 \sum_{i=1}^{NVALS} \ln P_i$$

under H_0 is distributed as $\chi^2_{2 \cdot NVALS}$

(g) Compare H with the χ^2 critical value based upon a value adjusted for the number of pairs examined.

(h) Sort the classes by the number of samples in which they occur and, secondarily, on the number of observations per classes.

(i) Remove a class from the final list if it is not significantly different (as per g) from all classes previously included.

Table 1 shows the results of the analyses to develop sets of spectral distinct classes.

ANALYSIS OF INITIAL RESEARCH FACTORS

The previous discussion yielded three research factors. Table 2 enumerates these factors and the levels at which they exist in this experiment.

A factorial experimental design was used to relate CAs measured by the "C" and JAK methods to the combinations of the factor levels which parameterized the classifier. In these models GRC and MSC are fixed effects and INCR is a random effect. Under each combination of factor levels, two independently determined samples were derived and classified. The resulting two accuracies form the replicates for each treatment.

RESULTS

Figure 1 is a plot of mean CA calculated by both "C" and JAK for the levels of GRC (summed across MSC and INCR) and the levels of MSC (summed across GRC and INCR). From these plots, it can be seen that the program preserves the theoretical relationship between the "C" and JAK; that is, the accuracy estimated by "C" is always more optimistic than that estimated using JAK. In fact, the profiles formed by the two methods are not significantly different from one another and estimates made by either method would vary in similar a manner. Consequently, results for the remaining section will be reported for the more conservative JAK method only.

The interactions effects of the factorial model do not explain significant variation. The profiles in plots

TABLE 2. RESEARCH FACTORS AND THEIR LEVELS

Ground Reference Combination (GRC) LEVEL	MSS Combination (MSC) LEVEL	Sample Spacing (INCR) LEVEL
1. Land form (Lf)	1. MSS 4 (M4)	2
2. Land use (Lu)	2. MSS 5 (M5)	4
3. Slope (Sl)	3. MSS 6 (M6)	6
4. Vegetation (Vg)	4. MSS 7 (M7)	8
5. Lf,Lu	5. M4,M5	10
6. Lf,Sl	6. M4,M6	
7. Lf,Vg	7. M4,M7	
8. Lu,Sl	8. M5,M6	
9. Lu,Vg	9. M5,M7	
10. Sl,Vg	10. M6,M7	
11. Lf,Lu,Sl	11. M4,M5,M6	
12. Lf,Lu,Vg	12. M4,M5,M7	
13. Lf,Sl,Vg	13. M4,M6,M7	
14. Lu,Sl,Vg	14. M5,M6,M7	
15. Lf,Lu,Sl,Vg	15. M4,M5,M6,M7	

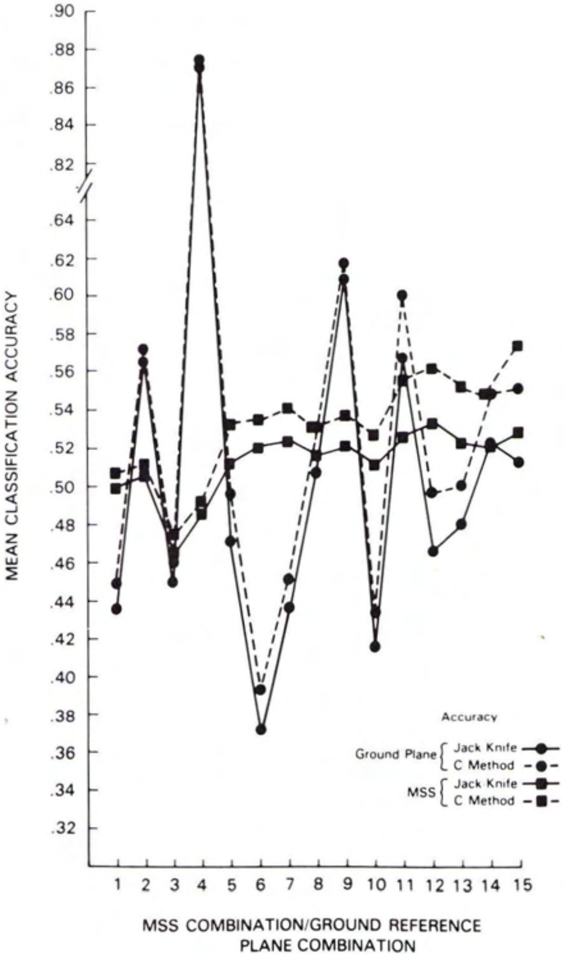

FIG. 1. Plot of classification accuracy for the combinations of ground reference planes and the combinations of MSS bands. Classification accuracies are computed by two methods (combination indicies correspond to factor levels given in text).

of the two-way interactions are parallel. Consequently, the main effects are analyzed as follows:

- *Spectral Bands (MSC).* The analysis of the MSC factor is aided by Table 3 and Figure 2. Table 3 contains three sub-tables. The columns for all three sub-tables are the same and are the levels containing 1, 2, and 3 bands with the maximum CA. The rows of the sub-tables represent levels of the second factor (either INCR or GRC) summed over the third factor or a summation over both of the remaining two factors (All). When restricted to one band, MSS5 followed by MSS7 does the best at discriminating among classes. When two bands can be used either MSS4-MSS7 or MSS5-MSS7 do the best job, and for three bands the combination of MSS4-MSS5-MSS7 is overwhelmingly the optimal set. What is interesting is the relative absence of MSS6 from among the "best" combinations. One may conclude from the dominance of MSS5 and MSS7 that the relationship between remote sensing and classes of the ground reference planes is largely a function of changes in vegetation types (cultivated and otherwise) and the presence of bare soil and barren areas. This conjecture is borne out by examination of the spectrally distinct classes. Among these classes changes in vegetation, land use (whose spectrally distinct classes largely differ by vegetation), and, to a lesser extent, land form dominate spectral distinctions.
- *Ground Attributes (GRC).* Interpreting Table 4 and Figure 3, we see that the results are very consistent. Among combinations containing single ground reference planes, the maximum CA is obtained when the vegetation classes are used. In fact, classifiers based upon the vegetation ground reference plane alone are far superior to any other combination of the four ground reference planes. This result is somewhat misleading in that of the five spectrally distinct classes, one of them—*natural grass or actively cultivated*—accounts for 85 percent of the observations. Among combinations of two ground reference planes, the pair land use and vegetation was globally optimal. It is clear that the spectrally distinct classes of land use carry information about changes in vegetation and are much more evenly distributed than the vegetation classes. Land use is also globally the ground reference plane with the second highest CA behind vegetation and, when the size dominant vegetation class is removed from the analysis, land use surpasses vegetation. Using three ground reference planes, the number of observations per class has evened out and the vegetation information the classifier uses is contained largely within the land use classes.
- *Spacing Between Samples Used in Developing Training Statistics (INCR).* Table 5 and Figure 4 show that as the spacing between samples used in developing training statistics increases, the CA decreases. This result is investigated further in the next section.

A Further Examination of Sample Spacing and Testing Strategy

The jack-knife procedure is used in order to maximize the amount of data available for the estimation of training statistics. This means that if there are a total of N observations available, the JAK procedure allows the researcher to perform an analysis of the same power that would required $2N$ observations using separate data sets for testing and training. The classification accuracy estimated by the JAK procedure should be an unbiased estimate of the CA estimated by separate training and testing, *if* the *samples* used to develop training statistics *are independent*. Recall however, that samples up to a fixed distance apart are dependent and that use of biased statistics in classification will lead analysts to think they are doing better than they are in reality. An important question then is how CA differs as a function of testing strategy (JAK versus STAT) and sample spacing.

For this analysis the data base was divided along the row dimension into three equal areas of approximately 90 rows. Training statistics were then estimated using the data in the upper third, and with these statistics CAs were calculated using JAK

TABLE 3. SUMMARY OF MSC LEVELS YIELDING THE GREATEST PERCENTAGE ACCURACY FOR LEVELS OF OTHER FACTORS.

		NUMBER OF BANDS IN MSC		
(a) INCR		1	2	3
	1	M7	M4M7	M4M5M7
	2	M5	M5M7	M4M5M7
	4	M5	M4M7	M4M5M7
	6	M5	M4M7	M4M5M7
	8	M5	M4M7	M4M5M7
	10	M5	M4M7	M4M5M7
(b) GRC		1	2	3
Lf	– 1	M5	M4M5	M4M5M7
Lu	– 2	M5	M5M6/M5M7	M5M6M7
Sl	– 3	M7	M4M7	M4M6M7
Vg	– 4	M5	M4M5	M4M5M7
Lf,Lu	– 5	M5	M5M7	M4M5M7
Lf,Sl	– 6	M5	M6M7	M4M5M7
Lf,Vg	– 7	M5	M4M5	M4M5M7
Lu,Sl	– 8	M7	M5M7	M4M5M7
Lu,Vg	– 9	M7	M5M7	M4M5M7
Sl,Vg	– 10	M7	M4M7	M4M5M7
Lf,Lu,Sl	– 11	M5	M4M7	M4M5M7
Lf,Lu,Vg	– 12	M4	M4M7	M4M5M7
Lf,Sl,Vg	– 13	M5	M4M5	M4M5M7
Lu,Sl,Vg	– 14	M7	M4M7	M4M5M7
Lf,Lu,Sl,Vg	– 15	M5	M4M7	M4M5M7
(c) All		1	2	3
		M5	M4M7	M4M5M7

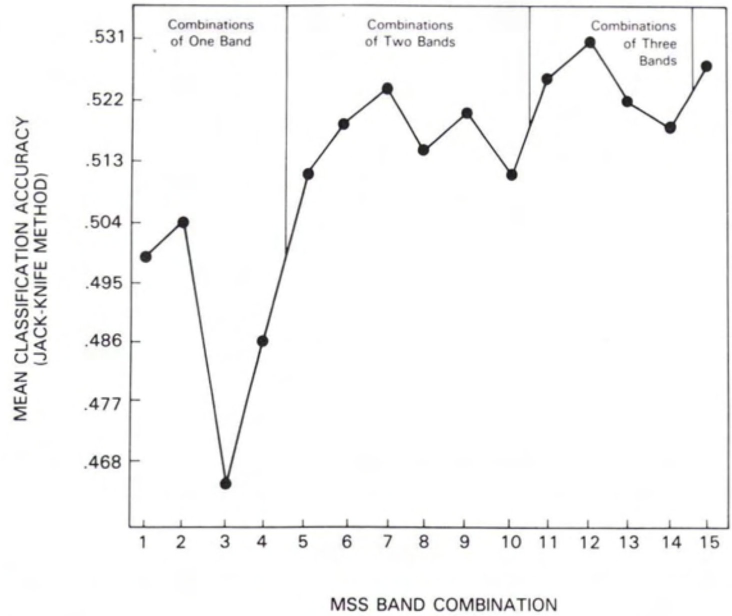

FIG. 2. Plot of classification accuracy as a function of MSS band combination (combination indicies correspond to factor levels given in text).

and separate data sets from the middle and bottom thirds of the image. This procedure was repeated using training samples with increasing sample spacing and with the added conditions:

- the spectrally distinct classes previously developed were used;
- because of the reduction in sample size for the training set, the maximum sample spacing was reduced to

TABLE 4. SUMMARY OF GRC LEVELS YIELDING THE GREATEST PERCENTAGE ACCURACY FOR LEVELS OF OTHER FACTORS.

		Number Ground Planes In GRC Levels		
(a) INCR		1	2	3
	1	Vg	Lu,Vg	Lf,Lu,Sl
	2	Vg	Lu,Vg	Lf,Lu,Sl
	4	Vg	Lu,Vg	Lf,Lu,Sl
	6	Vg	Lu,Vg	Lf,Lu,Sl
	8	Vg	Lu,Vg	Lf,Lu,Sl
	10	Vg	Lu,Vg	Lf,Lu,Sl
(b) MSC		1	2	3
M4	– 1	Vg	Lu,Vg	Lf,Lu,Sl
M5	– 2	Vg	Lu,Vg	Lf,Lu,Sl
M6	– 3	Vg	Lu,Vg	Lu,Sl,Vg
M7	– 4	Vg	Lu,Vg	Lu,Sl,Vg
M4M5	– 5	Vg	Lu,Vg	Lf,Lu,Sl
M4M6	– 6	Vg	Lu,Vg	Lf,Lu,Sl
M4M7	– 7	Vg	Lu,Vg	Lf,Lu,Sl
M5M6	– 8	Vg	Lu,Vg	Lf,Lu,Sl
M6M7	– 10	Vg	Lu,Vg	Lf,Lu,Sl
M4M5M6	– 11	Vg	Lu,Vg	Lf,Lu,Sl
M4M5M7	– 12	Vg	Lu,Vg	Lf,Lu,Sl
M4M6M7	– 13	Vg	Lu,Vg	Lf,Lu,Sl
M5M6M7	– 14	Vg	Lu,Vg	Lf,Lu,Sl
M4M5M6M7	– 15	Vg	Lu,Vg	Lf,Lu,Sl
(c) All		1	2	3
		Vg	Lu,Vg	Lf,Lu,Sl

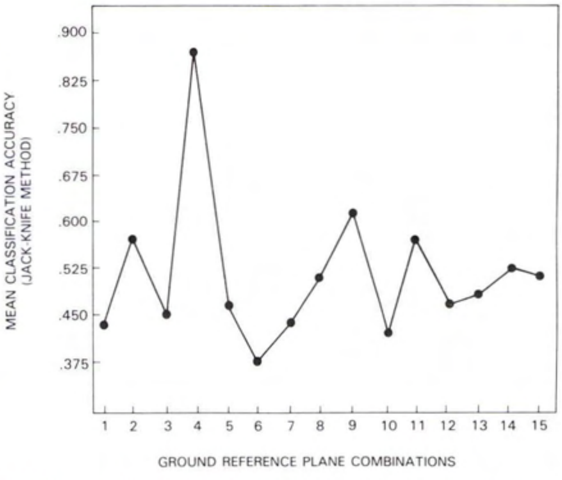

FIG. 3. Plot of classification accuracy as a function of ground reference plane combination (combination indicies correspond to factor levels given in text).

seven, the analysis was performed for GRC levels 1 to 4 (single maps) only, and the number of observations per class was divided by 3;

- while the spacing between training samples varied, the testing sets from the middle and bottom thirds of the quadrangle had randomly chosen centers and were composed of nominally contiguous pixels because they had no impact on estimating the training statistics and ordinarily a researcher would classify an entire area.

Figures 5 to 7 are plots by increment spacing of the CAvalues estimated using JAK, STAT2 (Training on the top third, testing on the middle third), and STAT3 (training on the top third, testing on the bottom third) averaged over all levels of MSS combinations for two ground reference planes (Lf and Lu) and a plot by increment spacing averaging CA over both MSC and GRC levels. In general, these plots show that the trend in CA reverses as one progresses from JAK to STAT2 to STAT3. CA decreases with increasing spacing when testing is performed using a jack-knife, and CA increases with INCR under a STAT3 strategy. The plots of CA using STAT2 appear to be a mixture of the JAK and STAT2 plots.

Several conclusions can be drawn for this analysis:

- The assumption of independence and identical distribution has not been met for the JAK and, as a clear consequence, CAs under this procedure are, in general, far more optimistic than they are under the STAT procedure which JAK is to estimate.
- It appears that, when JAK is used with smaller spacing, the effect of the local autocorrelation will yield a higher CA. In other words, the classifier is incorporating the autocorrelation information into the classification process. Samples constructed with wider spacings do not contain this information available to them and, consequently, the CA for these samples approaches that of the STAT3. The classifier is in a sense memorizing the data.
- The information contained in the autocorrelation is specific to the location of the sample. This information is an advantage when classifying in the precise area

TABLE 5. SUMMARY OF INCR LEVELS YIELDING THE GREATEST PERCENTAGE FOR LEVELS OF OTHER FACTORS.

	INCR (Spacing Between Samples Used In Developing Training Statistics)					
(a) MSC	1	2	4	6	8	10
1	1	4	3	2	5	6
2	1	3	4	2	5	6
3	1	2	3	4	5	6
4	1	2	3	4	6	5
5	1	2	3	4	5	6
6	1	2	3	4	5	6
7	1	2	3	4	5	6
8	1	2	3	4	5	6
9	1	2	3	4	5	6
10	1	2	3	4	5	6
11	1	2	3	4	5	6
12	1	2	3	4	5	6
13	1	2	3	4	5	6
14	1	2	3	4	5	6
15	1	2	5	3	4	6
(b) GRC	1	2	4	6	8	10
1	1	4	3	2	5	6
2	1	2	4	3	5	6
3	1	2	5	4	6	3
4	1	2	3	4	5	6
5	1	3	2	4	5	6
6	1	2	4	5	6	3
7	1	2	3	4	5	6
8	1	2	3	5	4	6
9	1	4	2	3	6	5
10	2	1	4	3	6	5
11	1	3	2	4	5	6
12	3	1	2	4	5	6
13	3	1	2	4	5	6
14	1	2	4	3	5	6
15	1	4	5	2	3	6
(c) All	1	2	4	6	8	10
	1	2	3	4	5	6

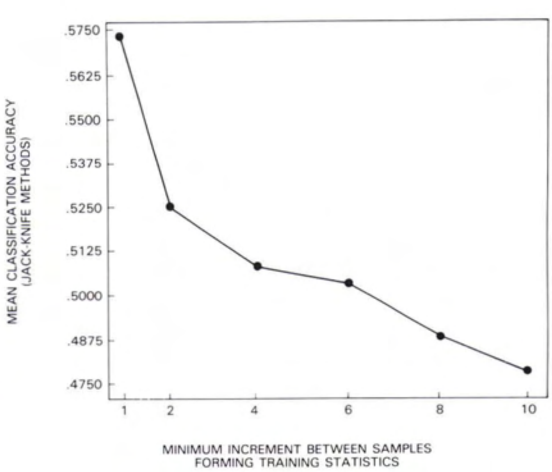

FIG. 4. Plot of classification accuracy as a function of increment spacing (in pixels) between samples forming training statistics.

from which the sample was drawn. However, the information is in error when the classifier is applied to samples from a location as close as a third of a quadrangle away. The result is that the classifier using training statistics based on closely spaced samples performs much worse than the classifier using training statistics based on widely spaced samples when the classifiers are applied to an area outside of which the training samples were drawn.

ALL-POSSIBLE-SUBSETS VERSUS STEPWISE BAND SELECTION STRATEGY

Table 6 shows the results of the analysis comparing a stepwise band selection strategy with an "all-possible-subsets" selection strategy. Thirty times a comparison was made of CAs from the best single, best pair, and best three bands as computed by both strategies. For MSC combinations formed of one band, in 23 out of 30 comparisons the stepwise procedure selected a suboptimal band with an average suboptimality in CA of eight percent. Among combinations consisting of two bands, in 25 out of 30 comparisons the stepwise procedure selected a

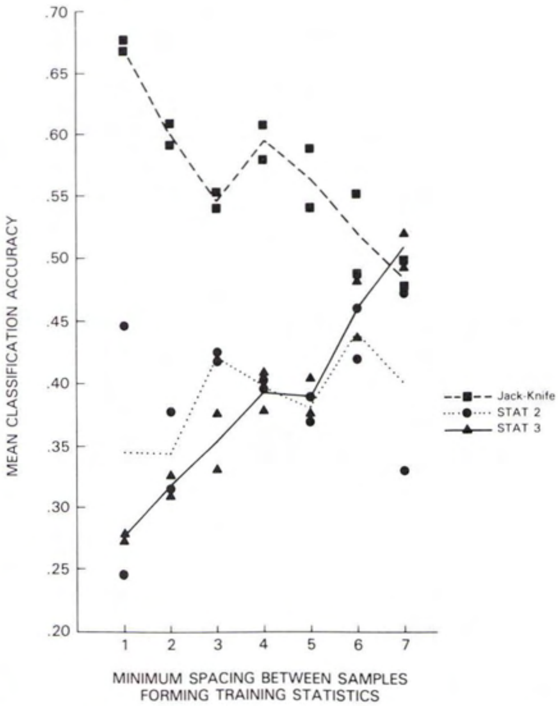

FIG. 5. Plot of classification accuracy from classifying land form as a function of spacing between samples forming training statistics, using three methods of testing.

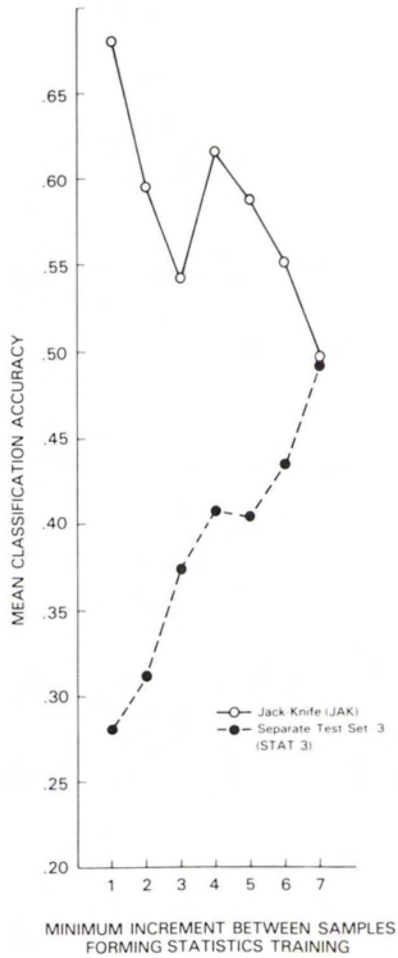

FIG. 7. Plot of classification accuracy as a function of spacing between samples forming training statistics, using two methods of testing. Classification accuracies were averaged over the four ground reference planes.

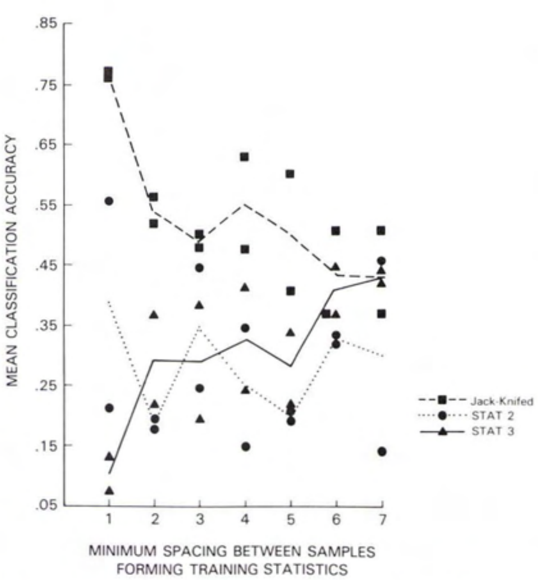

FIG. 6. Plot of classification accuracy from classifying land use as a function of spacing between samples forming training statistics, using three methods of testing.

suboptimal bands averaging a 10 percent difference in CA. Finally, among MSC combinations consisting of three bands, use of a stepwise strategy resulted in a suboptimal band selection in 27 out of 30 comparisons with an average of six percent difference in CA. All three average differences are significant at any commonly used alpha level.

SUMMARY AND CONCLUSIONS

Through this research I have tried to systematically examine limitations of MSS data. This has been pursued through analysis of the relationship between ground attribute planes and registered MSS data. The research concentrated on the answers arising from a series of questions related to performing classification. These questions dealt with the quality of MSS bands individually or in combination as

TABLE 6. PAIRED T-TEST COMPARING CLASSIFICATION ACCURACIES FROM "BEST" SUBSETS GENERATED BY "ALL-POSSIBLE-SUBSETS" SELECTION STRATEGY VERSUS STEPWISE SELECTION STRATEGY. (ALLACC = CLASSIFICATION ACCURACY FROM "ALL-POSSIBLE-SUBSETS" STRATEGY; STPAC = CLASSIFICATION ACCURACY FROM STEPWISE STRATEGY).

Combination*	Untransformed						Transformed						
	t*	P(t>t*)	DF	DIFF	STDV	N	t*	P(t>t*)	DF	DIFF	STDV	N	% of Suboptimal Comp.
Triple	4.40	0.0001	29	0.0566	0.0705	30	8.23	<0.0000	29	0.4086	0.2720	30	77
Double	5.20	<0.0000	29	0.0976	0.1028	30	7.98	<0.0000	29	0.5379	0.3693	30	83
Single	4.58	0.0001	29	0.767	0.0917	30	6.36	<0.0000	29	0.4376	0.3766	30	90

1. Untransformed Difference
2. Transformed Difference ($2^*\mathrm{SIN}^{-1}(\sqrt{\mathrm{ALLACC\text{-}STPAC}})$)

*Combination = Number of bands used in a classifier:
Triple (three bands):
Double (two bands):
Single (one band).

t* = Computed t value
P(t>t*) = Probability of t value exceeding t*
DF = Degrees of Freedom
DIFF = Mean Difference
STDV = Standard Deviation of Difference
N = Number of Comparisons

surrogate measures of combinations of ground attributes, and the impact of commonly used sampling, testing, training, and band selection strategies upon classification results.

Ground attributes maps generated by the USGS of the Parker Quadrangle, Colorado and Landsat 2 imagery of the area were examined. It was clear that the variety discerned on the ground by the human analyst who produced the ground reference maps far exceeded the discerning power of the MSS data. This difference in resolving power initially resulted in classifiers with very low accuracy and required (1) the assumption that use of all four MSS bands would yield overall the best classification accuracies and (2) a preprocessing stage of the analysis to produce a set of spectrally distinct class. Short of all four spectral bands MSS5, MSS7, and to a lesser extent MSS4 were most useful in discriminating among the spectrally distinct classes, while the combination of ground reference planes to discriminate most easily were land use and vegetation. Further, the land-use classes were dominated by changes in vegetation type or absence of vegetation. The discriminatory power of the MSS data was overwhelmingly dominated by vegetation even when other types of information were supplied to the analysis.

A further examination of testing strategies revealed that the influence of the autocorrelation on the classification results was very profound. The conclusion is that if one wishes to classify an area over which the "ground truth" is known and not to classify any where outside the area, then use contiguous pixels in the training set. However, in remote sensing, the more common research setting is to have "ground truth" information about a small area and to use this information to build training statistics and to extend these training statistics (signatures) and classifiers to areas about which little "ground truth" is known. In this second case, the use of contiguous pixels in the training set will hurt the classifier by 20–25 percent over use of widely spaced training samples. Use of widely spaced training samples will produce a more robust estimate of training statistics for the purposes of signature extension. Indeed, this result could be an explanation for the common finding of locally high classification accuracies (90 percent plus) and their rapid decline with short distances from the training area. The lower classification accuracies of widely spaced training samples therefore appear to be much better estimates of the power of MSS data for the purposes of classification. Therefore, a 20 to 25 percent increase in classification accuracy for the signature extension scheme can be achieved by just changing the sampling.

Finally, it is also clear that an average of 7 to 10 percent decrease in classification accuracy is related to the use of a stepwise band selection strategy versus an "all-possible-subsets" strategy. This difference might be more dramatic with TM data with its greater number of bands and band combinations (less likelihood that the optimal band subset will be selected by chance) and greater collinearity among bands. Alternatives to stepwise band selection should be examined unless the experimenter feels that the decrease in accuracy just described is unimportant.

REFERENCES

Cover, T. M., and J. M. Van Campenhout, 1977. On the possible orderings in the measurement selection problem, *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. SMC-7, No. 9, pp. 657–661.

Craig, R. G., 1979. Autocorrelation in Landsat Data, *Processings of the Thirteenth International Symposium on Remote Sensing of the Environment*, Ann Arbor, Michigan, pp. 1517–1526.

Fisher, R. A., 1954. *Statistical Methods for Research Worker*, 12th Edition, Oliver and Boyd, Edinburgh, 356 p.

Fukunaga, K., 1972. *Introduction to Statistical Pattern Recognition*, Academic Press, New York, 369 p.

Hogg, R. V., and A. T. Craig, 1978. *Introduction to Mathematical Statistics*, Fourth Edition, MacMillan Publishing Co., Inc., New York, 438 p.

Jennrich, R. I., and P. Sampson, 1979. P7M: Stepwise discriminant Analysis, in *BMDP Biomedical Computer Program*, P-Series, edited by W. J. Dixon and M. B. Brown, University of California Press, Berkley, CA, pp. 711–733.

Labovitz, M. L., and E. J. Masuoka, 1984. The influence of autocorrelation in signature extraction-an example from a geobotanical investigation of Cotter Basin, Montana, *International Journal of Remote Sensing*, Vol. 5, No. 2, pp. 315–332.

Landgrebe, D. A., 1978. Useful information from multispectral image data: another look, in *Remote Sensing The Quantitative Approach*, edited by P. H. Swain and S. M. Davis, McGraw-Hill International Book Company, New York, Chapter 7, pp. 336–374.

Van Campenhout, J. M., 1980. The arbitrary relation between probability of errors and measurement subsets, *Journal of the American Statistical Association*, Vol. 75, No. 369, pp. 104–109.

(Received 14 February 1985; revised and accepted 20 August 1985)

(Forum, continued from page 196)

- The simulation images were contrast enhanced on a scene by scene basis, with digital processing to expand the difference between the green and blue colors (S1, S2). Such imagery, which requires an additional step of computation, is not listed among the satellite data products.
- While the simulation imagery was produced at a scale of 1:24,000, the preliminary satellite products are at the highest magnification of 1:100,000 (SPOT Image, 1985).

These statements are not intended to detract from the excellent quality of the simulation data, and do not necessarily vitiate any of the conclusions as they apply to the satellite products. Evidently, some effort may be required by those obtaining SPOT satellite data to match the radiometric and photographic quality of the simulation data. The geometric fidelity of the satellite data should be much improved over the simulation data.

REFERENCES

Ackleson, S. G., and V. Klemas, 1985. A Comparison of SPOT Simulator Data with Landsat MSS Imagery for Delineating Water Masses in the Delaware Bay, Broadkill River, and Adjacent Wetlands, *Photo. Eng. and Rem. Sens.*, 51, pp. 1121–1129.

Bernstein, R. B., J. B. Lotspiech, H. J. Myers, H. G. Kolsky, and R. D. Lees, 1984. Analysis and Processing of Landsat-4 Sensor Data Using Advanced Image Processing Techniques and Technologies, *IEEE Trans on Geosci. and Rem. Sens.*, GE–22, pp. 192–221.

DeGloria, S. D., 1984. Evaluation of Simulated SPOT Imagery for Interpretation of Agricultural Resources in California, *Photo. Eng. and Rem. Sens.*, 51, pp. 1103–1108.

Price, J. C., 1984. Statistical Analysis of SPOT Simulation Data and Comparison with Landsat 4 Thematic Mapper Data, *SPOT Simulation Handbook*, Am. Soc. of Photogrammetry and Centre National d'Etudes Spatial, pp. 157–164.

Saint, G., and G. Weill, 1984. SPOT Simulation Methodology, Simulated versus Satellite Image Parameters, *SPOT Simulation Handbook*, Am. Soc. of Photogrammetry and Centre National d'Etudes Spatial, pp. 19–28.

SPOT Image, 1985. *Explanatory Note on Acquisition Programming and Ordering of SPOT Products*, Toulouse-Cedex, France, Product Characteristics, p. 4.

—John C. Price
Remote Sensing Research Laboratory
Agricultural Research Service, USDA
Beltsville, MD 20705