10259-1-X

Reprinted From

PROCEEDINGS of the SEVENTH INTERNATIONAL SYMPOSIUM on REMOTE SENSING OF ENVIRONMENT

17-21 May 1971

Sponsored by

Center for Remote Sensing Information and Analysis
Willow Run Laboratories, Institute of Science and Technology
The University of Michigan

Water Quality Office
Environmental Protection Agency

Office of Space Science and Applications
National Aeronautics and Space Administration

Federal Highway Administration
U. S. Department of Transportation

U. S. Geological Survey
U. S. Department of Interior

Air Force Cambridge Research Laboratories
Air Force Systems Command
U. S. Air Force

Agricultural Research Service
U. S. Department of Agriculture

Infrared and Optics Laboratory
Willow Run Laboratories
INSTITUTE OF SCIENCE AND TECHNOLOGY
THE UNIVERSITY OF MICHIGAN
Ann Arbor, Michigan

LARS Information Note 061771

EFFECTS OF COMPRESSION AND RANDOM NOISE ON MULTISPECTRAL DATA

P.J. READY
P.A. WINTZ
S.J. WHITSITT
D.A. LANDGREBE

The Laboratory for Applications of Remote Sensing

Purdue University, West Lafayette, Indiana

LARS Information Note 061771

EFFECTS OF COMPRESSION AND RANDOM NOISE ON MULTISPECTRAL DATA[1]

P. J. Ready, P. A. Wintz,
S. J. Whitsitt and D. A. Landgrebe

Laboratory for Applications of Remote Sensing
Purdue University
Lafayette, Indiana

ABSTRACT

Two problem areas are often encountered in the use of multispectral scanner data.

A. Large quantities of man and machine time are required to analyze and store the volumes of data gathered.

B. The quality of the data is usually reduced by the introduction of unwanted random noise.

This paper will present some effects of data compression and random noise on multispectral data, particularly as they apply to pattern recognition and picture quality.

[1]This work was supported by the National Aeronautics and Space Administration Grant Number NGL15-005-112.

I. DATA COMPRESSION

1. INTRODUCTION

The extremely large quantities of data produced by multispectral scanners have generated three interrelated problems. The first is the potentially wide bandwidth required to transmit data from a remote sensor to a data collection center, for example, in data transmissions from an earth satellite to a ground station. As the quantity of data transmitted in a given amount of time increases, so does the required bandwidth.

The second problem is the increasingly large blocks of time required for man/machine analysis of multispectral data. Even high-speed digital computers use relatively large amounts of time to process the volumes of data available.

The third problem is the actual physical storage of multispectral data. The value of gathered data never sinks to zero since one cannot always predict with certainty the future applications of the data; thus, libraries soon become unreasonably large as the quantity of stored data increases.

The application of appropriate data compression techniques to the data can significantly reduce the severity of the above three problems. Data compression reduces the quantity of data to be transmitted and thereby decreases the required transmission bandwidth. Secondly, multispectral data can be stored in the compressed state, resulting in more efficient data storage. Thirdly, analysis can be performed on the compressed data, or, as is more probable, the data can be expanded to its original state (within preset error restrictions) and then analyzed.

Since it is not possible to anticipate the needs of future users of the data, it is important that any data compression technique be information preserving. That is, the technique should not destroy more than what has been determined to be the maximum acceptable information loss. In addition, the data compression technique should be adaptive. It should be capable of efficient compression of the several different kinds of data that a multispectral scanner might encounter over changing terrain.

The extent to which multispectral data may be compressed is proportional to the degree of correlation (redundancy) that exists between data points. Highly correlated data may often be greatly compressed without substantially affecting the actual information content of the data. This is, the various users of the data, whether man or machine, are unable, within present limits, to detect any difference between analysis based on compressed data and that based on the original data. This includes, for example, classification by computer of multispectral imaged terrain.

Studies performed at the Laboratory for Applications of Remote Sensing at Purdue University have shown that multispectral scanner data is highly correlated in each of the three dimensions of the data, i.e., in the two spatial dimensions and the spectral dimension.

2. DESCRIPTION OF THE TECHNIQUE

An attempt is made to find a suitable orthogonal basis on which to expand the observed data. If a "good" basis is chosen, the number of dimensions required to represent the data will be less than for the original basis on which the data were collected. For time-discrete (sampled) data, one method of changing to another basis is by matrix transformation. [1,2,3] The technique is as follows. The original multispectral data is viewed as a vector,

$$\underset{\sim}{X} = [x_1, x_2, \ldots, x_N]^t$$

in N-space relative to the original basis. The data vector is then projected on a new, lower-dimensional ($n < N$) basis by the transformation $\underset{\sim}{T}$,

$$\underset{\sim}{Y} = \underset{\sim}{T}\,[\underset{\sim}{X} - \underset{\sim}{U}]$$

where*

$$\underset{\sim}{U} = E\{[x_1, x_2, \ldots, x_N]^t\}$$

is the mean vector of $\underset{\sim}{X}$, and $\underset{\sim}{T}$, is the n x N matrix whose n rows are the first n eigenvectors corresponding to the n largest eigenvalues of the covariance matrix $\underset{\sim}{C}$ of $\underset{\sim}{X}$. Thus, if

$$\underset{\sim}{C} = E\{(\underset{\sim}{X}-\underset{\sim}{U})(\underset{\sim}{X}-\underset{\sim}{U})^t\}$$

then the n rows of $\underset{\sim}{T}$ are the n normalized solutions to

$$\underset{\sim}{C}\underset{\sim}{\nu}_i = \lambda_i \underset{\sim}{\nu}_i \qquad i = 1,2,\ldots,n < N$$

with

$$\lambda_1 > \lambda_2 > \cdots > \lambda_n$$

The covariance matrix for $\underset{\sim}{Y}$ is given by

$$E\{\underset{\sim}{Y}\,\underset{\sim}{Y}^t\} = \begin{bmatrix} \lambda_1 & & & \mathbf{0} \\ & \lambda_2 & & \\ & & \ddots & \\ \mathbf{0} & & & \lambda_n \end{bmatrix}$$

Since the covariance matrix for Y is diagonal, the elements of the transformed data vector are uncorrelated. Each λ_i represents the variance of the i-th element of $\underset{\sim}{Y}$.

The transformed data vector $\underset{\sim}{Y}$ is projected back on the original basis by the transformation $\underset{\sim}{T}^t$,

$$\hat{\underset{\sim}{X}} = \underset{\sim}{T}^t \underset{\sim}{Y} + \underset{\sim}{U}$$

It is desirable that $\hat{X}$ be a good approximation to X. A measure of the quality or fidelity of the approximation is the mean square error (MSE),

$$\varepsilon^2(n) = E\{||\underset{\sim}{X} - \hat{\underset{\sim}{X}}||^2\}$$

The eigenvector transformation T^t is the optimum inverse transformation in that it minimizes $\varepsilon^2(n)$ for a given $n < N$. The resulting MSE is given by

$$\varepsilon^2(n) = \sum_{i=1}^{N} \lambda_i - \sum_{i=1}^{n} \lambda_i = \sum_{i=n+1}^{N} \lambda_i$$

Thus the error in projecting the original N-dimensional data onto the $n < N$ dimensional basis and then projecting back to N-dimensions is an MSE equal to the sum of the N-n discarded λ_i's. For highly correlated data the eigenvalues decrease rapidly and the error is small. The resulting data compression ratio is given by N/n.

---

* E is the expectation operator

3. EXPERIMENTAL RESULTS

Multispectral data were obtained from the sampled and quantized output of a 12-channel (.4 μm to 1.0 μm) airborne scanner flown over predominantly agricultural regions in Indiana. [4] Typical flight lines were approximately one mile wide and from five to twenty-five miles long. The number of data points available for processing from a given flight line was on the order of 5×10^6 points.

The two sources of data redundancy, spectral and spatial correlations, are examined separately.

3.1. SPECTRAL REDUNDANCY

The eigenvector transformation is applied to the 12 spectral channels by arranging the data into 12-element vectors for each ground resolution point. The twelve elements of the data vector are the 12 spectral channel intensities for that point. The covariance matrix for the 12-vector is then estimated, and the resulting eigenvectors and eigenvalues computed using the LARS IBM Model 67 digital computer. Since the required dimension of the new basis is proportional to the number of significant eigenvalues, Figure 1 indicates that only three dimensions are necessary for flight line C1. The reconstruction error in projecting the 3-vector Y back to 12 dimensions is 2%, as shown in Figure 2. Since MSE is closely related to subjective picture quality, [5] a reconstructed flight line having only a 2% MSE is practically indistinguishable from the original (see Figure 3).

The other error criteria, computer classification accuracy, is investigated by classifying flight lines subjected to several different compression ratios. The results for flight line C1, as shown in Figure 2, agree with results based on MSE. Compression ratios up to 12/3 are acceptable. This means that a typical flight line of 5×10^6 data points can be compressed using the above technique to a set of 1.25×10^6 points. When desired, the original data can be reconstructed from these stored points with negligible picture and classification degradation. If the analyst is interested only in classification of the flight line, it is not necessary to reconstruct the original data points. Each ground resolution point is now represented by the 3-vector Y rather than the original 12-vector X. Essentially all of the information (variance) regarding the particular point on the ground contained in X is now contained in Y. Thus classification using the elements of the 3-vector Y as features yields results similar to classifications based on any combination of the original 12-channels. [6] This has been tested on two flight lines, and the results are shown in Table I.

3.2. SPATIAL REDUNDANCY

The second source of data redundancy, spatial correlations, arises from the fact that adjacent ground resolution points are often quite similar. Two methods are used to take advantage of this correlation.

In the first method, the data from each spectral channel are arranged into a set of 100-element vectors consisting of sequences of 100 horizontal scan line resolution elements. The covariance matrix for this 100-element vector and the resulting eigenvalues and eigenvectors are computed. Results for flight line C1 indicate that the first 25 eigenvalues account for 96% of the total variance in the data. A compression ratio of 100/25 is then expected to yield little distortion. Classification results for this compression ratio were found to be identical to those obtained using the original data (90.1%).

The second method of spatial data compression is superior to the method just discussed in that it takes advantage of correlations existing in both the horizontal and the vertical directions. In this method, data from each spectral channel are read in as a data matrix, ten resolution points by ten resolution points. The data matrix is then converted into a 100-element vector, the first ten elements being the first row of the data matrix, the second ten elements the second row of the data matrix, and so on through 100 points. The covariance matrix and resulting eigenvectors and eigenvalues are computed, with the results for flight line C1 shown in Figure 4. Figure 5 demonstrates the correspondence between significant eigenvalues and resulting classification and picture quality, by showing the same flight line subjected to four different compression ratios. For the extreme case of 100/1 compression, the ten-by-ten data blocks are easily distinguishable.*

4. PICTURE ENHANCEMENT

If the analyst is interested in visually observing a multispectrally scanned area, he is limited to looking at one spectral channel at a time; however, some areas may not produce significant response in the particular channel he has chosen. Thus, he must compromise and select a channel that is good on the average for the chosen area he is interested in. Some areas will then be less detailed than if he had chosen the most responsive channel for that area.

This problem is reduced significantly in the following way. Since the transformed data vector $\underset{\sim}{Y}$ is of smaller dimension than the original data vector $\underset{\sim}{X}$, most of the information in the N channels of $\underset{\sim}{X}$ has been compressed into the $n < N$ "channels" of $\underset{\sim}{Y}$. The variances of the elements of $\underset{\sim}{Y}$ are arranged in decreasing order, $\lambda_1 > \lambda_2 > \cdots > \lambda_n$. Thus, the first element of $\underset{\sim}{Y}$, y_1, has the largest variance, and if the original 12 channels are at all correlated, λ_1 is greater than the variance of any of the original 12 channels. The analyst can then observe the area of interest through "channel" y_1. The effects of this enhancement are shown in Figure 6.

5. CONCLUSIONS ON DATA COMPRESSION

In order to provide the analyst with large amounts of good quality multispectral data, some means of efficiently encoding the data for transmission, analysis, and storage must be found. The eigenvector transformation discussed in this paper represents the optimum linear transformation when MSE is used as the reconstruction fidelity criteria. The transformation is ideally suited to the digital computer, and significant (100/10 spatial, 12/4 spectral) compression ratios with negligible classification and picture degradation have been achieved with LARS multispectral scanner data.

II. RANDOM NOISE

An important parameter in the design and use of remote sensing systems is the data-signal-to-noise ratio. The effects of a common type of noise on the classification of multispectral data are examined in this part of the paper. Some general problems and considerations in the classification of "noisy" data are discussed.

1. INTRODUCTION

Flight line C1 was remotely sensed by an airborne scanner optically linked to twelve sensors corresponding to twelve adjacent wavelength bands in the visible and near-infrared portions of the electromagnetic spectrum. The responses of the sensors were amplified and recorded on board as continuous waveforms. These waveforms were then sampled and uniformly quantized to 256 levels.

A signal processed in this manner is subject to the addition of noise at a number of times. For instance, thermal noise is added in the sensors, amplifiers, and ground A/D system. The effects of this noise in multispectral classification are difficult to predict.

In order to obtain a qualitative and quantitative assessment of the effects of one common type of noise in a typical classification problem, samples from C1 were perturbed by various levels of simulated, uncorrelated Gaussian noise. These "noisy" samples were then classified using a scheme described in [7] and [8]. In this procedure the samples are assumed to belong to one of M classes. Each sample is classified using a maximum likelihood decision rule assuming equal class a priori probabilities. The distribution of samples from each class is assumed to be multivariate normal.

---

*The gray scale computer printouts shown in Figures 3,5, and 6 include every other line and every other column of the flight line data.

2. EXPERIMENT

Flight line C1 covers a primarily agricultural area. Eight classes were chosen which represent bare soil and most of the crop types present: wheat, soybeans, alfalfa, corn, oats, rye, red clover. Class statistics were estimated using samples from preselected fields.

The random samples used as noise are actually random numbers generated by a power residue method on a digital computer. A typical histogram and finite-record autocorrelation (using the method of Blackman and Tukey [9]) for 15,000 samples are presented in Figure 7. The lack of any significant correlation is apparent.

The level of noise is given in terms of the number of units in one standard deviation (sigma). Six new sets of data were generated by adding noise with sigma equal to 2,5,7,10,15, and 20 units. The new data sets were quantized again with no scale changes. Classification was performed and accuracy obtained by comparing the results with known crop cover.

The best subset of four channels (wavelength bands) out of the twelve available was used. Previous experience ([8],[10]) and theoretical evidence [11] have shown that using a smaller set than the whole set of features is often desirable in terms of accuracy. The saving in processing time is apparent.

Histograms and gray-scale printouts were generated for portions of the flight line.

3. RESULTS

A photograph of one portion of C1 and the matching gray-scale printout of unperturbed data are shown in Figure 8. Figure 9 shows gray-scale printouts for the same area with noise present. Note that it is not difficult to discern visually the field patterns, even for sigma equal to fifteen units. The digitized scan line (horizontal) in the middle of this area is plotted in Figure 10 for both no noise present and sigma equal to fifteen. It seems to be much harder to pick out any general patterns in the noisy scan line.

The effects of the addition of noise on the distribution of field samples is given in Figure 11. The data spreads out for higher noise levels, as expected. The histograms are ragged due to the relatively small number of samples in one field (wheat, in this case).

Approximately 4000 samples were used in training and 13,000 in testing the classifier. The results for these samples as well as results for the two specific test classes are given in Figure 12. All of the curves have a smaller slope at either end than at the middle. Schemes designed to reduce noise by small amounts (one or two units) would tend to make little difference near the extremes of noise level.

The results for wheat and soybeans show that the effects of increasing noise will be different for each class even though performance with no noise is about the same for all classes. This is, of course, a function of the individual class distributions. This selective effect may be manipulated by the choice of features, an important point in the design of systems with a fixed set of features subject to varying levels of noise.

Figure 13 contains a printout of a portion of those samples classified as wheat for both no noise and sigma equal to fifteen.

4. CONCLUSION ON RANDOM NOISE

It is reasonable to assume that the type of data described will not be linearly separable. This implies that freedom from noise will not automatically result in good classification. Nature has a certain amount of variability of her own to spread the data out over the feature space. Hence the term "signal" in signal-to-noise ratio is difficult to define in terms meaningful for classification purposes. One would probably lump signal and noise into the class statistics and predict performance based on some "distance" measure (divergence, Bhattacharyya distance, etc.).

REFERENCES

[1] Kramer, H. P., and Mathews, M. V. 1956. A Linear Coding for Transmitting a Set of Correlated Signals. IRE Transaction on Information Theory, Vol. IT-2, pp. 41-46.

[2] Habibi, A., and Wintz, P. A. 1971. Image Coding by Linear Transformation and Block Quantization. IEEE Transactions on Communication Technology, Vol. COM-19, No. 1.

[3] Huang, J. J. Y., and Schulthesiss, P. M. 1963. Block Quantization of Correlated Gaussian Random Variables. IEEE Transactions on Communication Systems, pp. 289-296.

[4] Landgrebe, D. A. 1970. Data Analysis and Remotely Sensed Data, AIAA Earth Resources Observations and Information Systems Meeting, AIAA Paper No. 70-292.

[5] Tasto, M., and Wintz, P. A. 1970. Picture Bandwidth Compression by Adaptive Block Quantization. Technical Report TR-EE70-14, Purdue University, Lafayette, Indiana, July 1970.

[6] Fu, K. S. 1968. Sequential Methods in Pattern Recognition and Machine Learning, Adademic Press, 1968, pp. 29-54.

[7] Fu, K., Landgrebe, D., and Phillips, T. 1969. Information Processing of Remotely Sensed Data, Proceedings of the IEEE.

[8] Laboratory for Applications of Remote Sensing, (LARS) Purdue University, Research Bulletin No. 844, September 1968.

[9] Blackman, R., and Tukey, J. 1958. The Measurement of Power Spectra, Dover, New York.

[10] Fu, K., and Min, P. 1968. On Feature Selection in Multispectral Pattern Recognition, LARS Technical Report, No. TR-EE68-17.

[11] Hughes G. 1968. On the Mean Accuracy of Statistical Pattern Recognizers. IEEE Transactions on Information Theory, IT-14.

TABLE I. A COMPARISON OF CORRECT CLASSIFICATION RESULTS BASED ON ORIGINAL AND TRANSFORMED DATA

	Optimum combination of four original spectral channels	The elements of $\underset{\sim}{Y}$ with n = 3
Flight line C1	90.1%	89.5%
Flight line 24	72.7%	74.4%

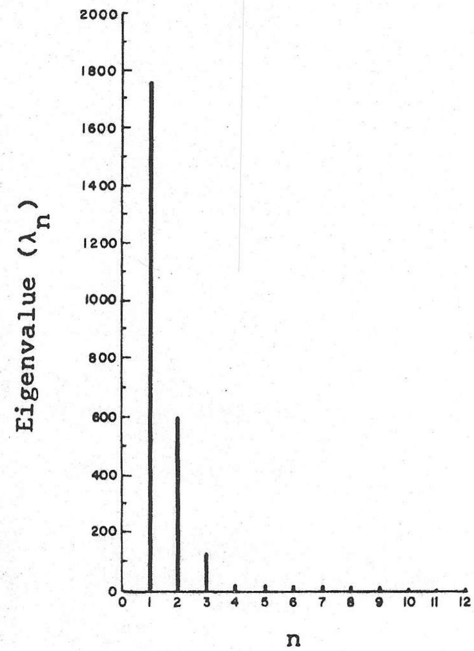


Figure 1
Twelve Eigenvalues of the Spectral Covariance Matrix for Flightline C1

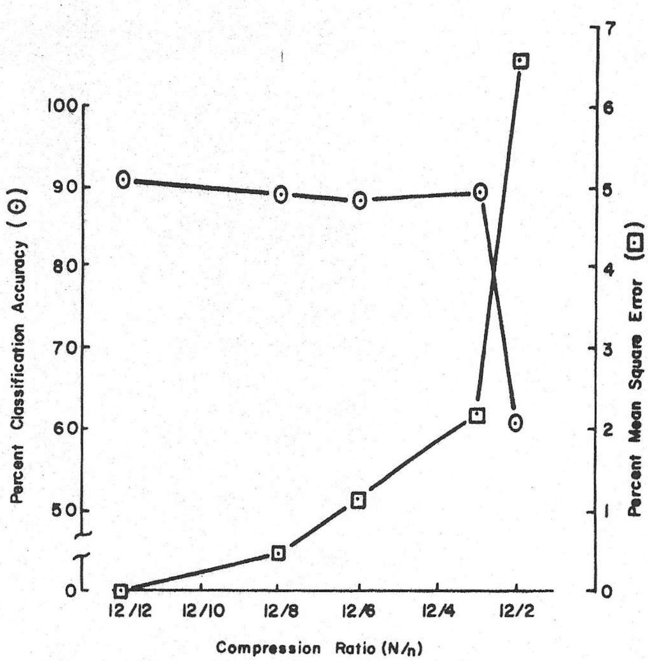
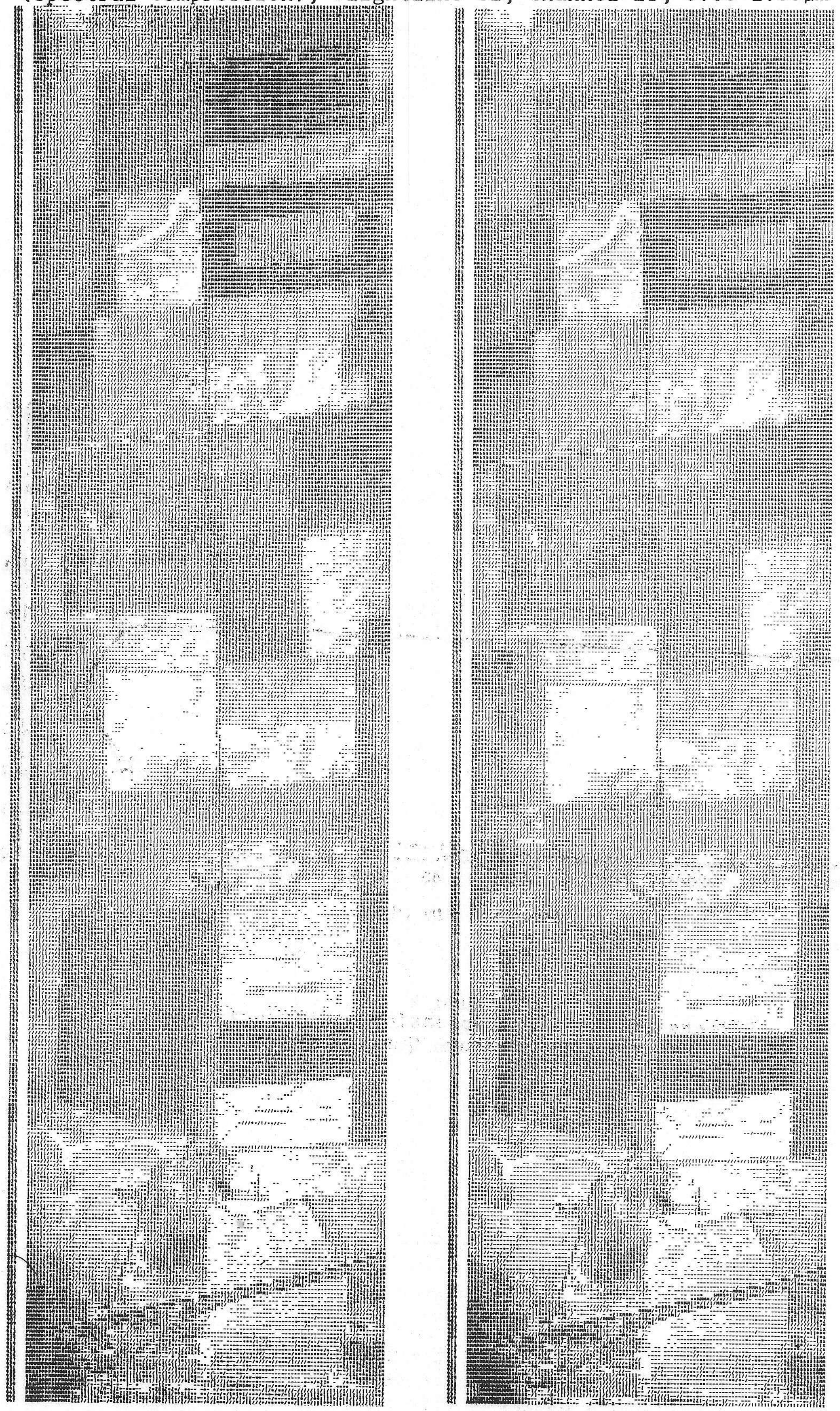


Figure 2
Effects of Spectral Compression on Classification Accuracy and Picture Quality (MSE)

Figure 3. Comparison of Original and Processed Data (Spectral Compression), Flightline C1, Channel 12, 0.80-1.00μm



Original Data Reconstructed Data

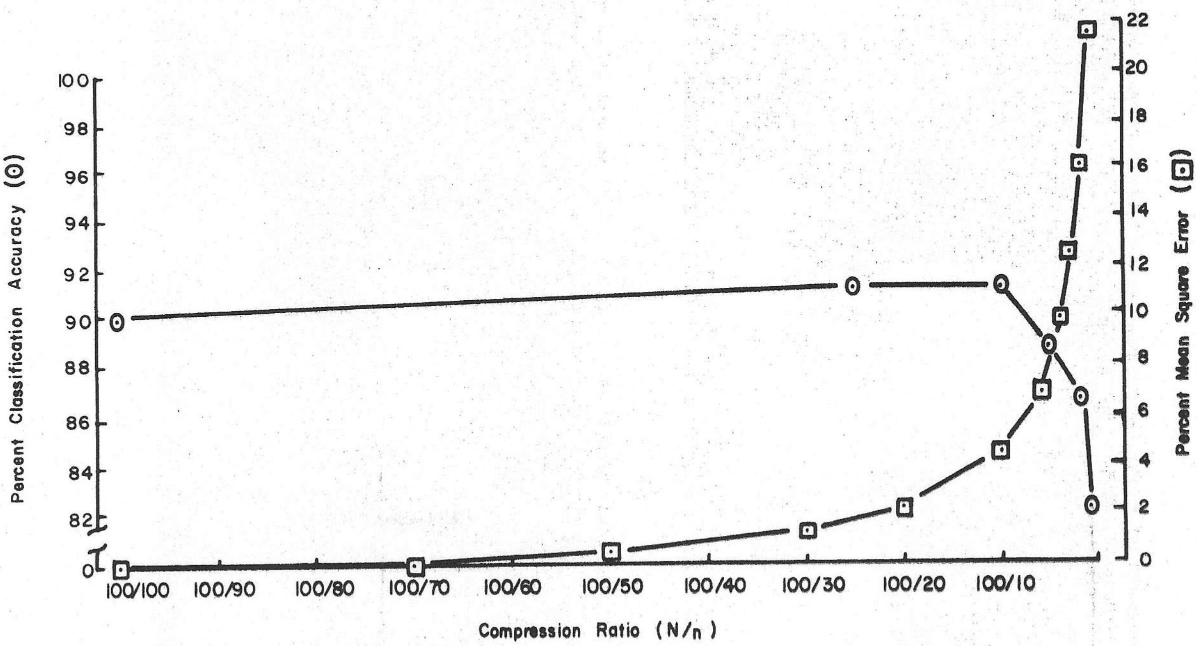
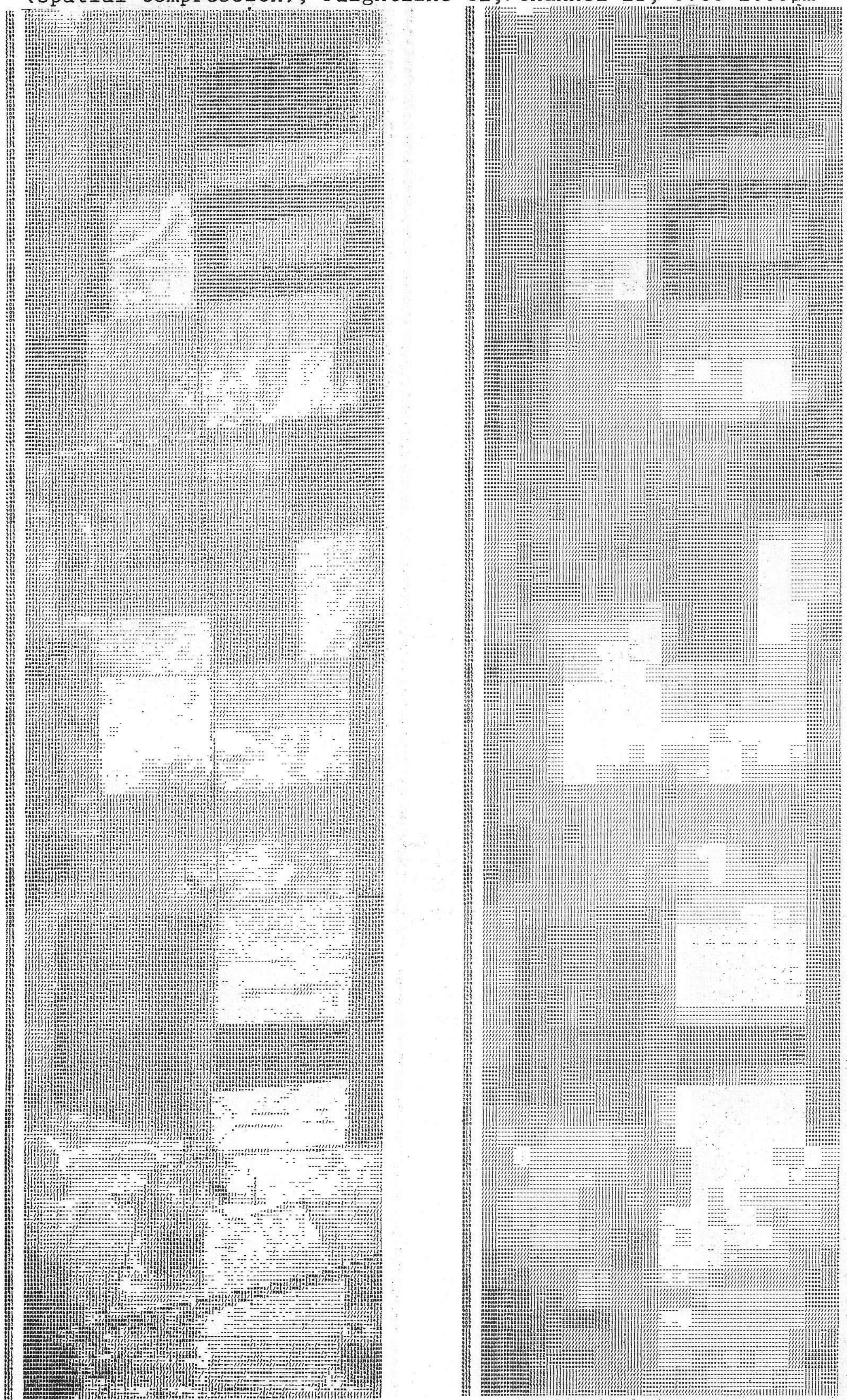


Figure 4
Effects of Spatial Compression on Classification
Accuracy and Picture Quality (MSE)

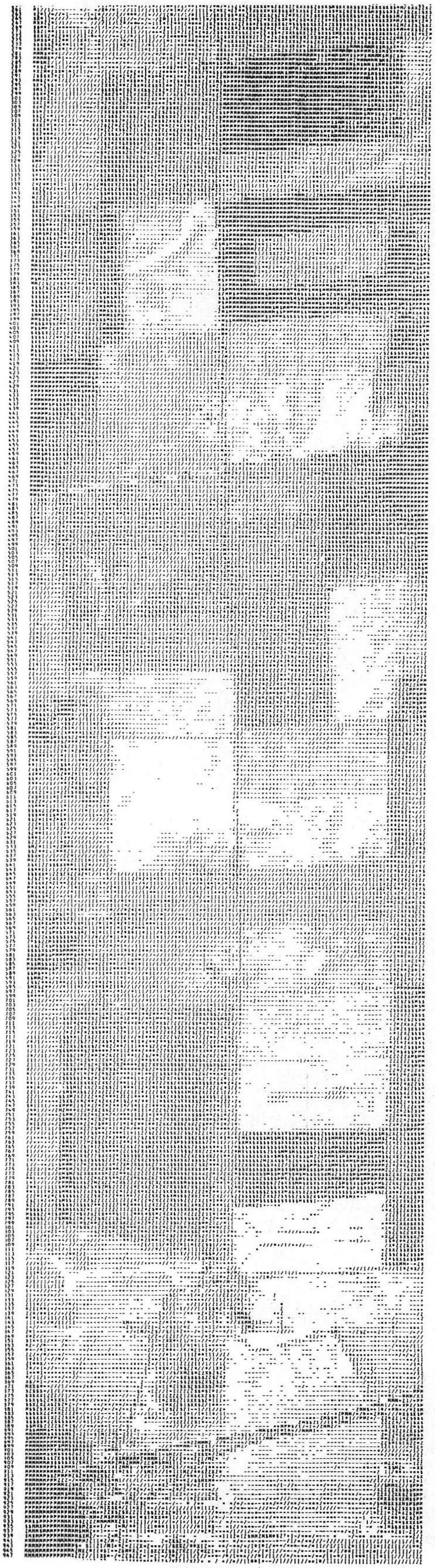
Figure 5. Comparison of Original and Reconstructed Data (Spatial Compression), Flightline C1, Channel 12, 0.80-1.00μm



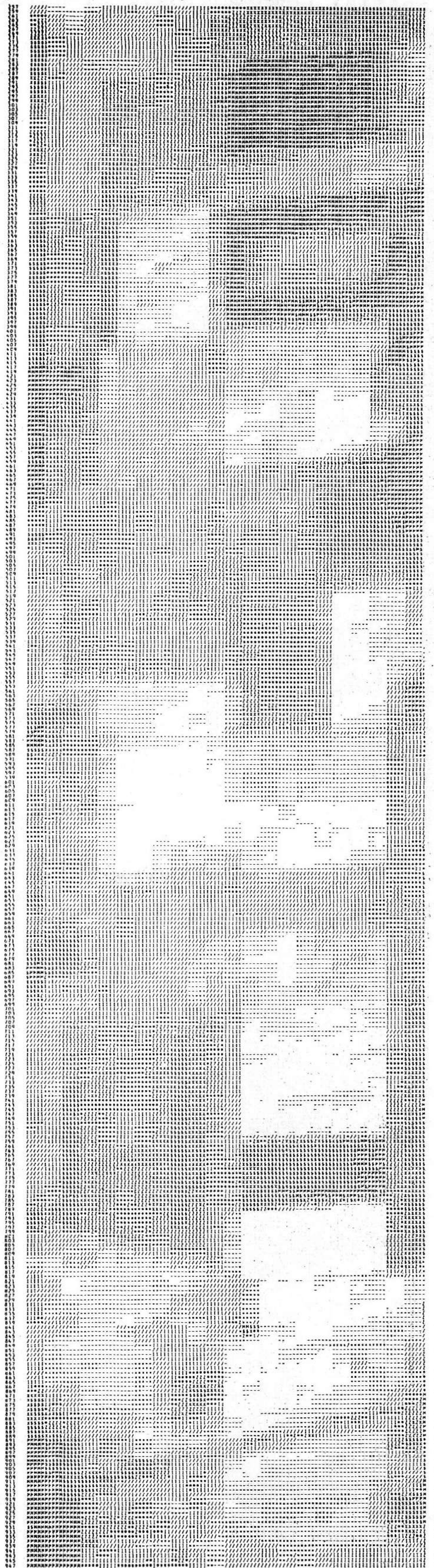
Original Data

100/1 Compression (82.1% Classification Accuracy)

Figure 5. Comparison of Original and Reconstructed Data (Spatial Compression), Flightline C1, Channel 12, 0.80-1.00μm

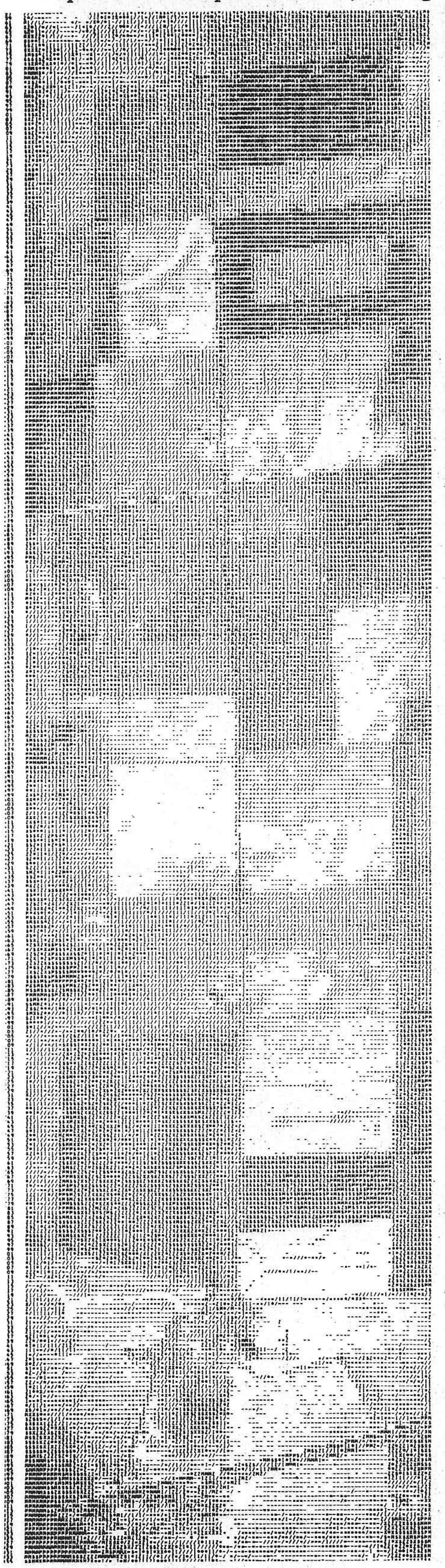


Original Data

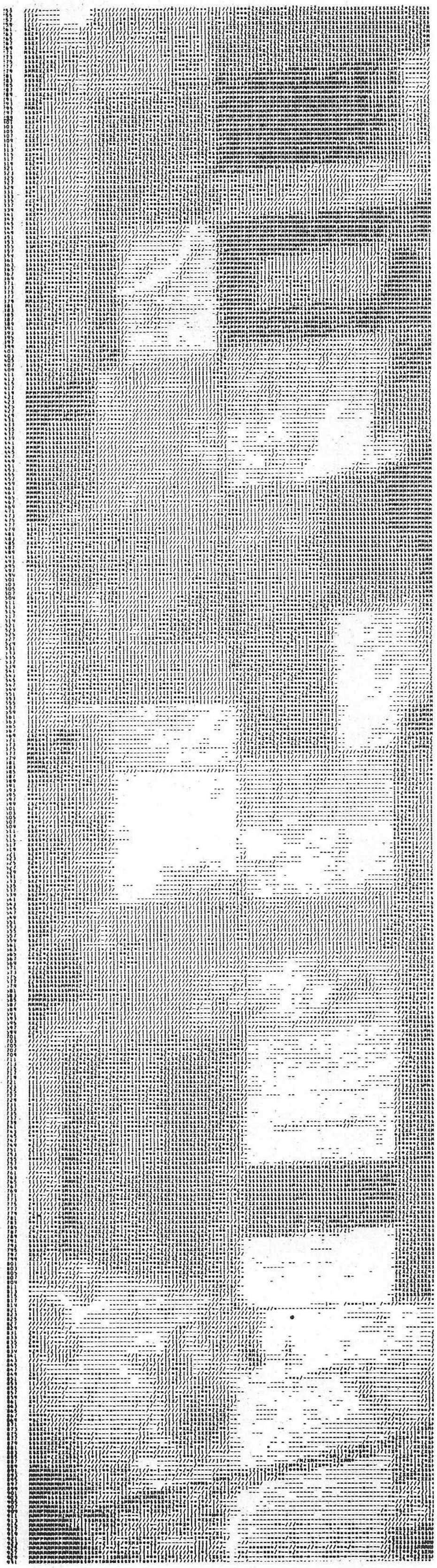


100/5 Compression
(88.7% Classification Accuracy)

Figure 5. Comparison of Original and Reconstructed Data (Spatial Compression), Flightline C1, Channel 12, 0.80-1.00μm

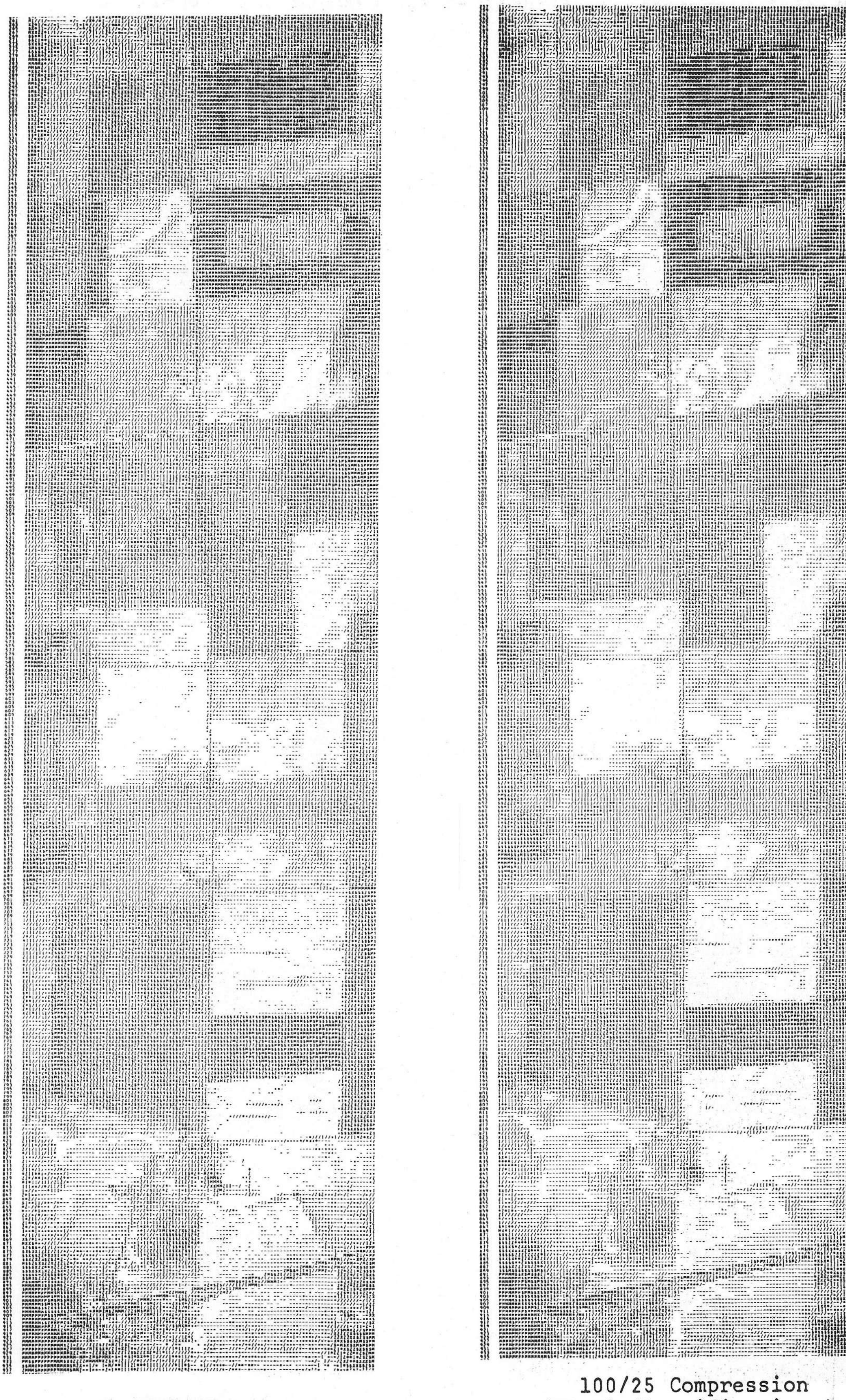


Original Data



100/10 Compression
(91.1% Classification Accuracy)

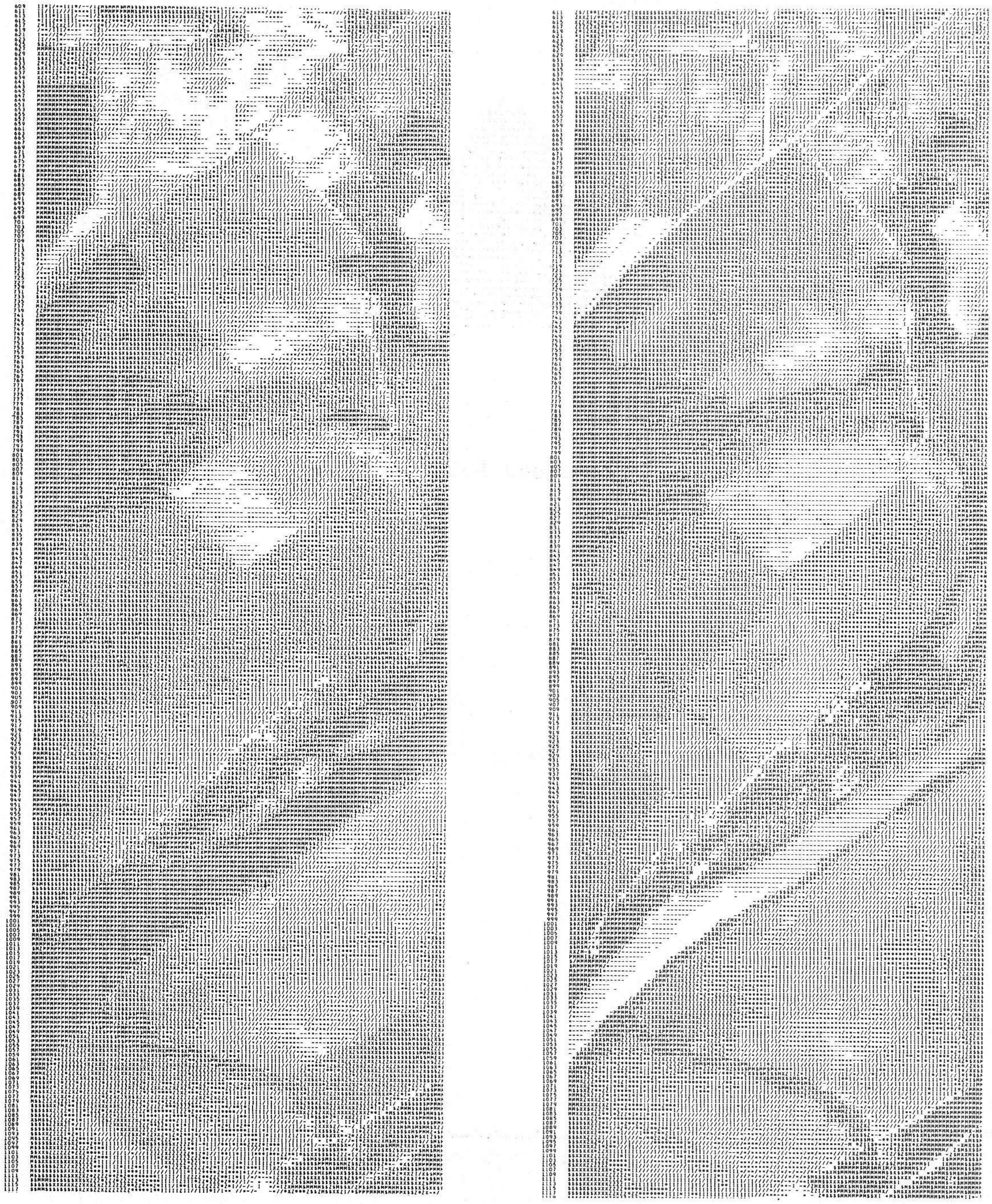
Figure 5. Comparison of Original nnd Reconstructed Data (Spatial Compression), Flightline Cl, Channel 12, 0.80-1.00μm



Original Data

100/25 Compression (90.1% Classification Accuracy

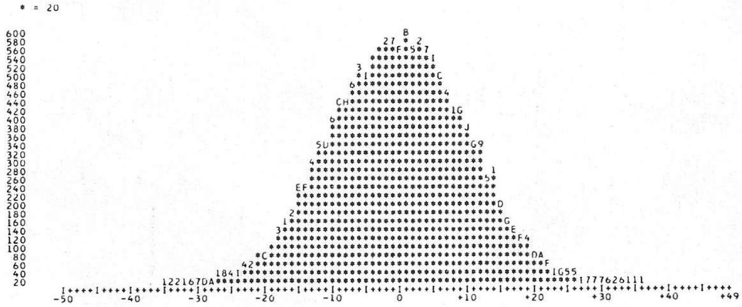
Figure 6. Comparison of Original and Processed Data (Spectral Transformation), Flightline 24, Channel 12, 0.80-1.00μm



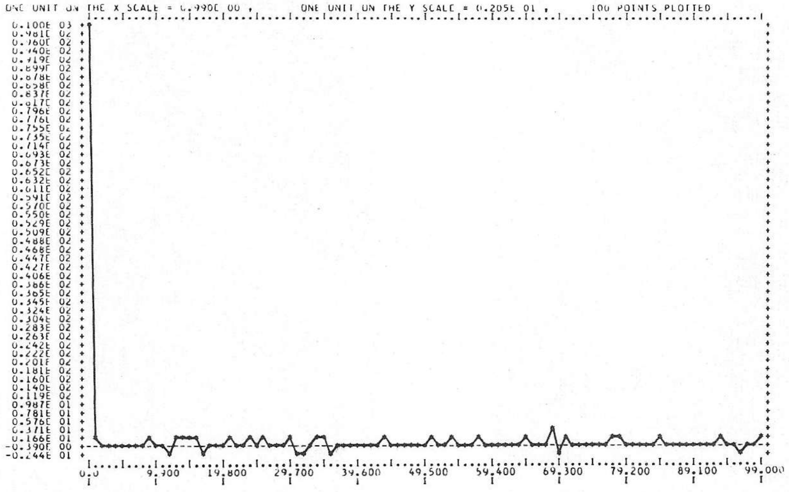
Original Data

Enhanced Data

Figure 7. Random Sample Statistics



Histogram of 15,000 Points



Autocorrelation - 100 Lags

Figure 8. Data with No Added "Noise"

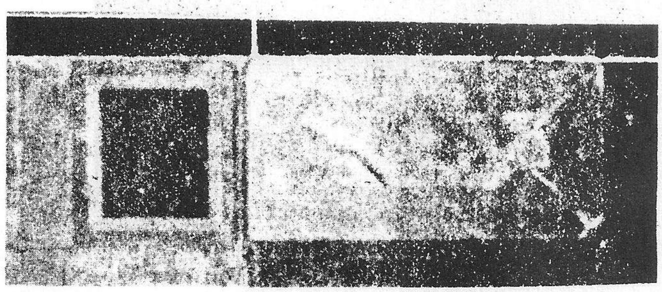


Figure 8a. Agricultural Scene from C1
(B & W Photograph)

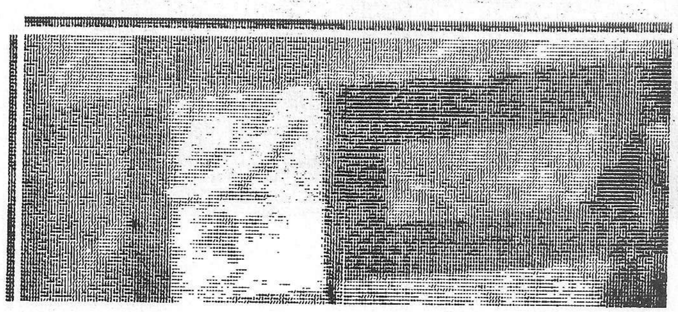


Figure 8b. Scanner Output of the Same Scene in the
0.80 to 1.00 Micrometer Band Run 66000600

Figure 9. Scanner Data from Flight Line C1 with Noise Added

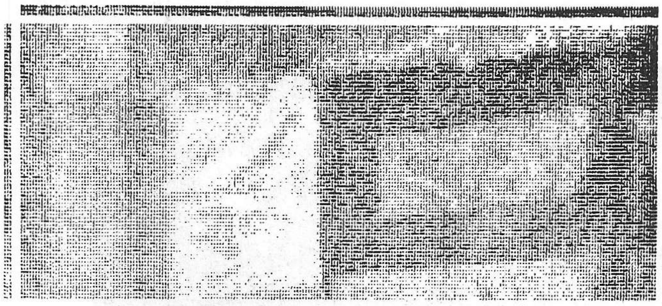


Figure 9a. Sigma = 5

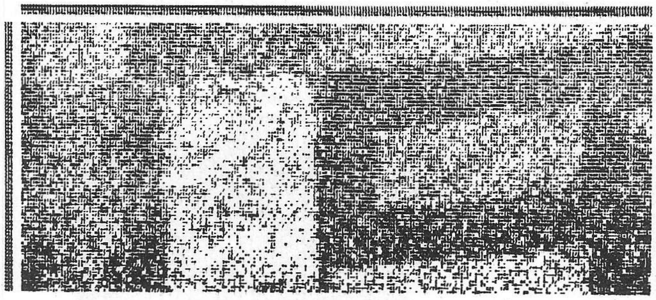
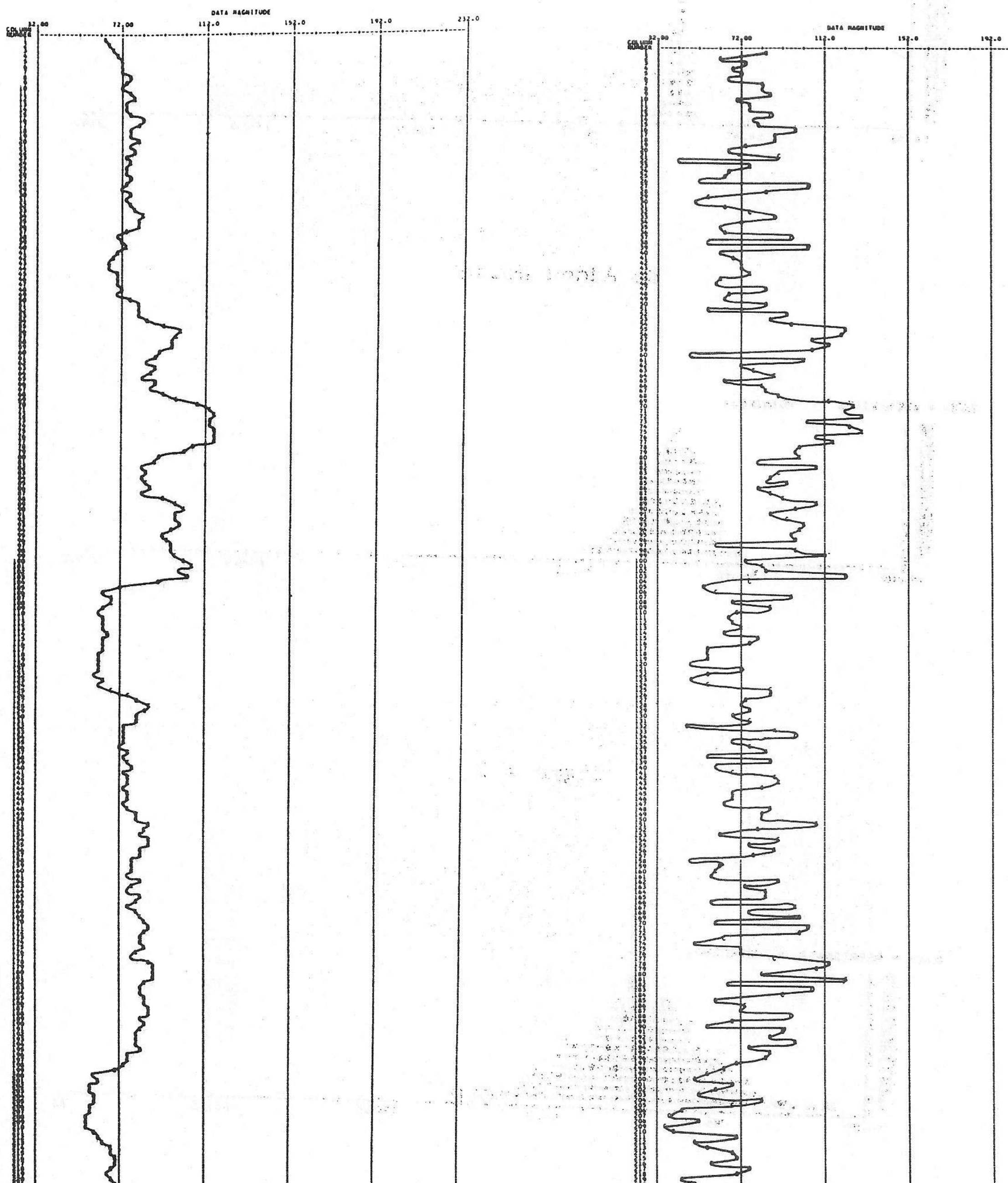


Figure 9b. Sigma = 15

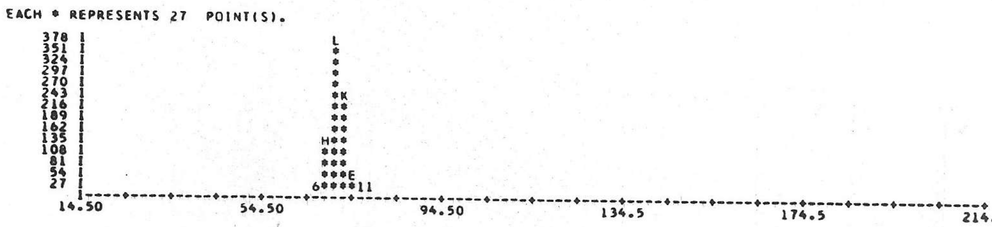
Figure 10. Plot of one scan line from flight line C1



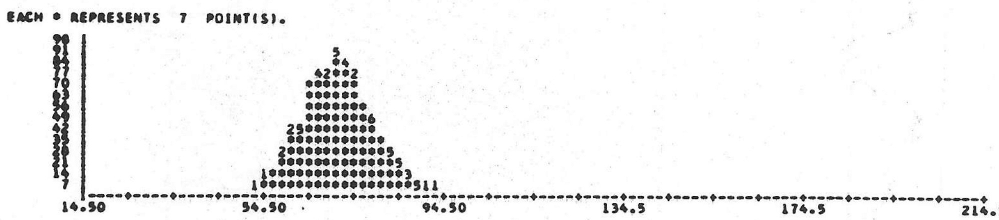
Middle Row of Figure 8b.
No Added Noise

Middle Row of Figure 9b.
Sigma = 15

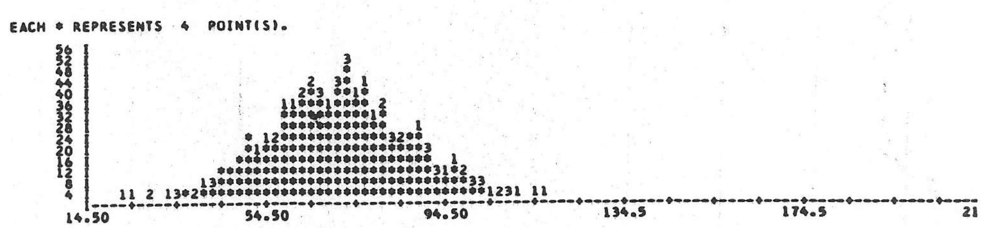
Figure 11. Histogram of One Wheat Field in the 0.40 to 0.44 Micrometer Band.



No Added Noise



Sigma = 7



Sigma = 15

Figure 12. Classification Performance vs Noise

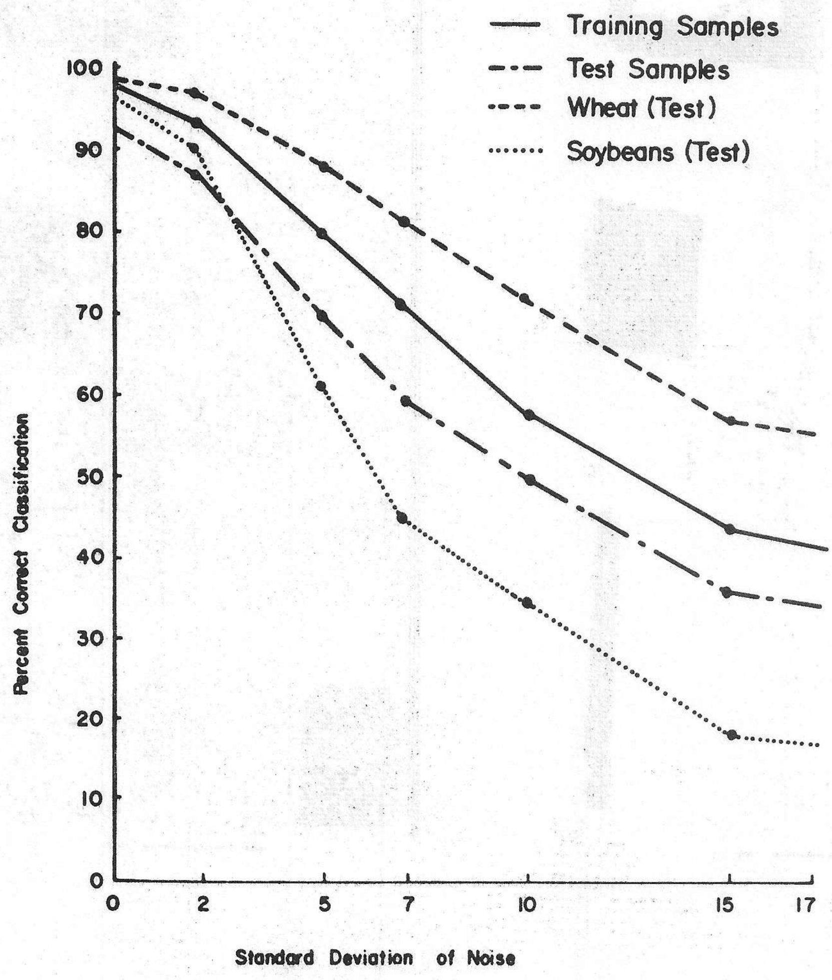
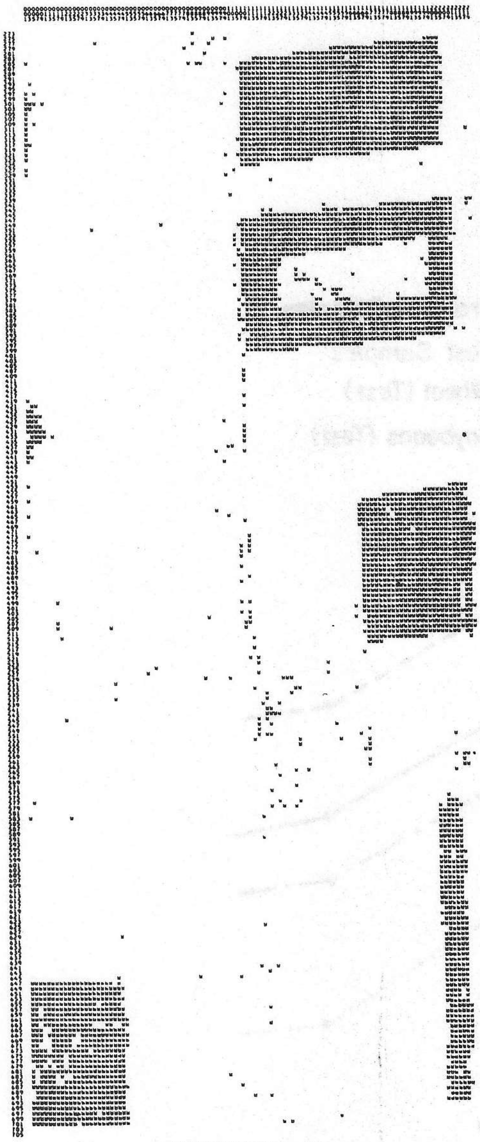
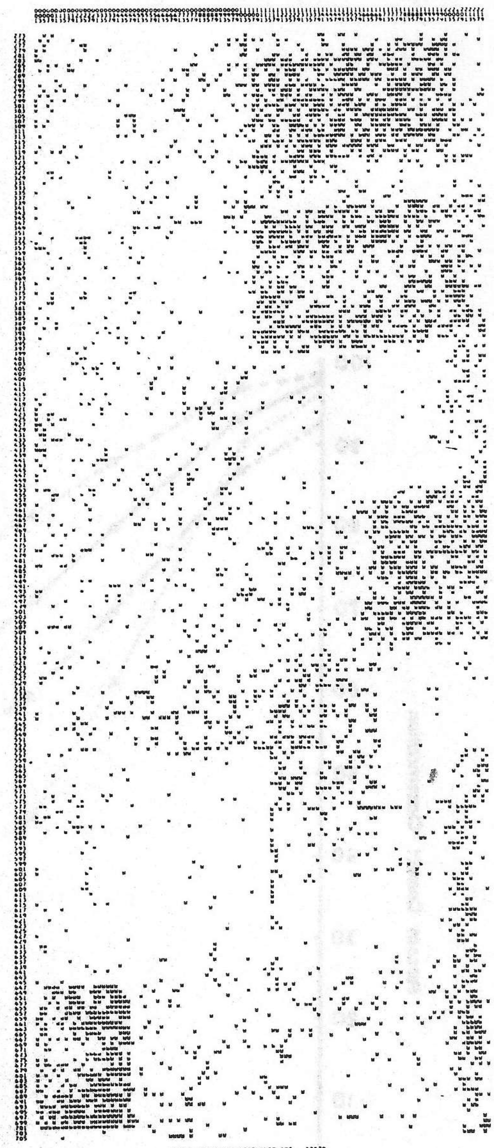


Figure 13. Classification Results for Flight Line C1: For Wheat Only



No Added Noise



Sigma = 15