NASA

Technical Memorandum 82064

Preliminary Evidence for the Influence of Physiography and Scale Upon the Autocorrelation Function of Remotely Sensed Data

M. L. Labovitz, D. L. Toll and R. E. Kennard

(NASA-TM-82064) PRELIMINARY EVIDENCE FOR THE INFLUENCE OF PHYSIOGRAPHY AND SCALE UPON THE AUTOCORRELATION FUNCTION OF REMOTELY SENSED DATA (NASA) 36 p HC A03/MF A01 CSCL 05B G3/43

N81-19530

Unclas
18837

DECEMBER 1980

National Aeronautics and
Space Administration

Goddard Space Flight Center
Greenbelt, Maryland 20771

TM 82064

PRELIMINARY EVIDENCE FOR THE INFLUENCE OF PHYSIOGRAPHY AND SCALE UPON THE AUTOCORRELATION FUNCTION OF REMOTELY SENSED DATA

M. L. Labovitz
Code 922, Geophysics Branch

D. L. Toll
Code 923, Earth Resources Branch

R. E. Kennard
Computer Sciences Corporation

December 1980

GODDARD SPACE FLIGHT CENTER
Greenbelt, Maryland

PRELIMINARY EVIDENCE FOR THE INFLUENCE OF PHYSIOGRAPHY AND SCALE UPON THE AUTOCORRELATION FUNCTION OF REMOTELY SENSED DATA

M. L. Labovitz
Code 922, Geophysics Branch

D. L. Toll
Code 923, Earth Resources Branch

R. E. Kennard
Computer Sciences Corporation

ABSTRACT

Previously established results of Craig (1976, 1979) and Craig and Labovitz (1980) demonstrated that Landsat data are autocorrelated and can be described by a univariate linear stochastic process known as an Auto-Regressive-Integrated-Moving-Average model of degree 1, 0, 1 or ARIMA (1, 0, 1). This model has two coefficients of interest for interpretation - ϕ_1 and θ_1. In a comparison of Landsat Thematic Mapper Simulator (TMS) data and Landsat MSS data several results were established:

(1) The form of the relatedness as described by this model is not dependent upon system look angle or pixel size.

(2) The ϕ_1 coefficient increases with decreasing pixel size and increasing topographic complexity.

(3) Changes in topography have a greater influence upon ϕ_1 than changes in land cover class.

(4) The θ_1 seems to vary with the amount of atmospheric haze.

These patterns of variation in ϕ_1 and θ_1 are potentially exploitable by the remote sensing community to yield stochastically independent sets of observations, characterize topography, and reduce the number of bytes needed to store remotely sensed data.

LIST OF FIGURES

LIST OF TABLES

PRELIMINARY EVIDENCE FOR THE INFLUENCE OF PHYSIOGRAPHY AND SCALE UPON THE AUTOCORRELATION FUNCTION OF REMOTELY SENSED DATA

INTRODUCTION

Remotely sensed data possesses an exploitable source of information in the form of the relationship between the digital count of a given pixel and those of its neighbors. The intensity and nature of this spatial relationship can be measured by examining the covariance between pixels. To introduce this concept into a remote sensing context, we will first define the covariance of any two random variable X and Y. This is given by,

$$E[(X - \mu_X)(Y - \mu_Y)] \tag{1}$$

where:

$E[\cdot]$ is the expected value operator;

μ_X is the expected value of X;

μ_Y is the expected value of Y.

This quantity is estimated by,

$$\sum_{i=1}^{n} \frac{(X_i - \bar{X})(Y_i - \bar{Y})}{n - 1} \tag{2}$$

where:

n is the number of observations;

$\bar{X}$ and $\bar{Y}$ are the arithmetic means of X and Y respectively;

X_i and Y_i are measurements of the X property and Y property on the i^{th} object.

There are several points to note.

1. The quantity (2) will be positive if the paired deviations $(\tilde{X}_i, \tilde{Y}_i)$ $[\tilde{X}_i = X_i - \bar{X}; \tilde{Y}_i = Y_i - \bar{Y}]$ have the same sign and (2) will be negative if the paired deviations have opposite signs.

2. If Y_i and $\bar{Y}$ are replaced by X_i and $\bar{X}$, (2) becomes the sample variance of X, S_X^2. Similarly, S_Y^2 is obtained by replacing X_i and $\bar{X}$ by Y_i and $\bar{Y}$.

3. The covariance between a given X and Y are dependent on the measurement scale of the variables.

In order to be able to compare pairs of variables measured on different scales, the covariance may be adjusted by dividing by the product of the square roots of the two variances. This new quantity is the familiar correlation coefficient, ρ. The estimate of ρ is r which is given by

$$\sum_{i=1}^{n} \frac{(X_i - \bar{X})(Y_i - \bar{Y})}{(\sqrt{S_X^2}\sqrt{S_Y^2})\ (n-1)}. \tag{3}$$

The range on this quantity is $-1 \leq r \leq 1$. If we once again replace Y_i, $\bar{Y}$ and S_Y^2 by X_i, $\bar{X}$ and S_X^2 (or vice versa), we have the ratio of the sample variance to itself, i.e. $r = 1$.

To put the above measures into a remote sensing context, let us define any given scan line of digital counts as a sequence $\{X_i\}$ with i being an index such that the first pixel in the scan line is $i = 1$ and neighboring pixels are consecutively indexed to the end of the scan line. Let us suppose that there are $n + 1$ pixels and replace the quantities Y_i, $\bar{Y}$ and S_Y^2 in (3) by X_{i+1}, $\bar{X}_{i+1}$ and $S_{X_i}^2$, then we have

$$\sum_{i=1}^{n} \frac{[(X_i - \bar{X}_i)(X_{i+1} - \bar{X}_{i+1})]}{\sqrt{S_{X_i}^2}\ \sqrt{S_{X_{i+1}}^2}\ (n-1)}. \tag{4}$$

If the sequence $\{X_i\}$ satisfies certain conditions, then $E[\bar{X}_i] = E[\bar{X}_{i+1}]$ and $\mathrm{Var}(X_i) = \mathrm{Var}(X_{i+1})$ [where $\mathrm{Var}(\cdot)$ is the variance operator] and (4) becomes

$$\sum_{i=1}^{n} \frac{[(X_i - \bar{X})(X_{i+1} - \bar{X})]}{S_X^2\ (n-1)}. \tag{5}$$

This expression is known as the estimate of the autocorrelation at lag 1. Note that with the exception of X_1 and X_{n+1} every element of the sequence is X_i and X_{i+1} at some point in the

summation. This expression is therefore the correlation of the sequence $\{X_i\}_{i=1}^{n}$ with itself incremented by one pixel or $\{X_i\}_{i=2}^{n+1}$. In general, we can estimate the autocorrelation at a lag k ($k = 0, 1, \ldots, n$) by

$$r_k = \sum_{i=1}^{n+1-k} \frac{[(X_i - \bar{X})(X_{i+k} - \bar{X})]}{S_X^2 (n - k)}. \tag{6}$$

The sequence $\{r_k\}_{k=0}^{n}$ is known as the estimated autocorrelation function (acf) and a plot of k versus r_k is known as the correlogram.

Clearly, if X_i and X_{i+1} are correlated, then X_i and X_{i+2} are going to have a portion of their correlation induced by their relationships to X_{i+1}. Therefore we need to define a quantity analogous to the partial correlation of conventional statistics. The quantity is called the partial autocorrelation function (pacf) and is given by the collection ϕ_{kk}'s defined as

$$\begin{aligned} \phi_{kk} &= \rho_{X_i X_{i+k} | \{X_i\}_{i=i+1}^{i+K-1}} \\ &= E[(X_i - E[X_i | X_{i+L} = x_{i+L}; L = 1, 2, \ldots k-1]) \\ &\quad (X_{i+k} - E[X_{i+k} | X_{i+L} = x_{i+L}; L = 1, 2, \ldots, k-1])] \end{aligned} \tag{7}$$

where:

> $E[\cdot | \cdot]$ is the conditional expectation of the random variable on the left of the "|" conditioned upon setting the random variables on the right of the "|" at some arbitrary but fixed values.[1]

Box and Jenkins (1970) developed a family of linear stochastic models known as the Autoregressive-Integrated-Moving-Average or ARIMA models which make use of these two functions. It can be shown (with considerable algebra which will not be done here) that member models of the ARIMA family generate specific patterns in the acf and pacf. Thus given a sample from an unknown process, the acf and pacf can be estimated and an ARIMA model fitted.

[1] $E[X_i | X_{i+1}]$ for example may be thought of as the linear regression estimate of X_i based upon X_{i+1}.

The degree of complexity of an ARIMA process is expressed as the value of three parameters known as p, d, q, where p is the order of the auto-regressive process; d is the complexity of a trend parameter; and q is the order of the moving average (p, d, q are non-negative integers). For example, if d = q = 0, than we have an Auto-Regressive process of order p, or notationally AR(p), given by,

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_p X_{t-p} + \delta + a_t$$

where:

ϕ_i are a set of coefficients (i = 1, 2, . . ., p);

$\delta = \frac{\mu_X}{(1 - \phi_1 - \phi_2 - \ldots - \phi_p)}$;

a_t is a normally and independently distributed random variable with mean 0 and variance σ_a^2 [NID(0, σ_a^2)].

In other words X_t is dependent upon the p previous values of X (X_{t-1}, X_{t-2}, . . . , X_{t-p}) plus a random perturbation generated at the present time or location. Letting p = d = 0, then we have a Moving Average process of order q, notationally MA(q) which is:

$$X_t = \mu_X = a_t = \theta_1 a_{t-1} - \theta_2 a_{t-2} - \ldots - \theta_q a_{t-q}$$

where:

$\{a_t\}$ is a q length sequence of NID (0, σ_a^2) random variables. In this model X_t is dependent upon the present and q previous random perturbations.

It has been demonstrated (Craig, 1976, 1979; Craig and Labovitz, 1980) that Landsat data are autocorrelated and that the autocorrelation function (acf) can be well approximated by the ARIMA (1, 0, 1) model of Box and Jenkins (1970). The ARIMA (1, 0, 1) model is given by,

$$\tilde{X}_t = \phi_1 \tilde{X}_{t-1} + a_t - \theta_1 a_{t-1} \qquad (8)$$

where:

$\{\tilde{X}_t\}$ is a sequence of observations indexed in time or space with each element of the sequence $\tilde{X}_t = X_t - \mu$;

$\{a_t\}$ is defined as before as a series of NID (0, σ_a^2) random variables;

ϕ_1 and θ_1 are coefficients.

Within a remote sensing context the model means that the gray scale value of a given pixel ($\tilde{X}_t$) is dependent upon the gray scale value of an adjacent pixel ($\tilde{X}_{t-1}$), a random purturbation associated with the adjacent pixel (a_{t-1}) and a random purturbation specific to the present pixel. While in this paper only scan lines indexed in the direction of scan are analyzed, Craig (1976) has found the ARIMA (1, 0, 1) model also to be applicable for sequences in the reverse direction as well as for sequences along a single element in either direction.

It was concluded by Craig and Labovitz (1980) that the value of ϕ_1 varies with some still ill defined location "effect." The authors hypothesized that the location effect is related to some combination of topography, land cover, and/or season. The θ_1 term on the other hand varied with the percent cloud cover. This study will examine the relative importance of two of these sources of variation–land cover and physiography–for the acf. A second portion of this study is motivated by NASA's intention to launch satellites processing spatial resolutions less than 80 m. Results from this investigation have implications for data analysis, data interpretation and data compression that will be covered in later sections. In summary then we will investigate:

(a) Is the ARIMA (1, 0, 1) model appropriate for remotely sensed data possessing a spatial resolution < 80 m?

(b) If (a) is true, how is spatial resolution manifested in the ARIMA (1, 0, 1) model?

(c) Can we start to determine the relative magnitudes of the contribution to the location "effect" of topography versus land cover?

DATA TYPES AND LOGIC OF EXPERIMENT

Data from three sources were used in this experiment:

(a) A Landsat 2 image, scene id 21608-1655 [nominal scene path 36, row 32 (Denver, CO,)] imaged on June 18, 1979.

(b) Two data sets acquired by the Landsat Thematic Mapper Simulator (TMS)–NS001–mounted aboard a NASA C130 aircraft.

(1) Data acquired on the plains at the eastern fringe of Denver, CO. This flight took place on June 20, 1979 starting at 1900 GMT (1 p.m. MDT) and consisted of two flight lines flown in a north-south direction, each 20.8 nautical miles (nm) in length with latitude 39° 50′N, longitude 104° 52′W and latitude and 39° 24′N, longitude 104° 42′W defining the upper left corner and lower right corner respectively of the study area.

(2) Data acquired in the Northern Rocky Mountains region of Montana. The data came from flight line number 3, flown on August 29, 1979 commencing at 1830 GMT (12:30 p.m. MDT). The flight line was 19.5 nm in length, flown in a south-to-north direction along 112° 42′W longitude, starting at 47° 00′N latitude and ending at 47° 15′N latitude. The flight line passed over Cotter Basin, Montana and hereafter will be called the Cotter Basin line. Both sets of TMS data were flown approximately 10,000 feet (3.049 km) above ground level (AGL). Since the instantaneous field-of-view of the NS001 is 2.5 milliradians, the pixels are approximately 7.5 m (25 feet) on a side. Table 1 gives the bands for which data were collected by the TMS and MSS.

Table 1
Spectral bands available for MSS and TMS data

Channel Number		Band Width (Micrometers)	
TMS	MSS	TMS	MSS
1		0.42 - 0.52	
2	4	0.52 - 0.60	0.50 - 0.60
3	5	0.63 - 0.69	0.60 - 0.70
4	6	0.76 - 0.90	0.70 - 0.80
5	7	1.00 - 1.30[1]	0.80 - 1.1
6		1.55 - 1.75	
7		2.08 - 2.35	
8		10.4 - 12.5[2]	

[1]Not available from Denver TMS
[2]Partial coverage (40%) Denver TMS

The paper is logically developed into four analyses. Because the large scan angle associated with the TMS (±50° from nadir) represents a reasonable potential source of variation in the autocorrelation, we first examine how the autocorrelation function varies with the look angle of the portion of the scan line. The importance of this analysis lies in its implications for the way we select starting points for the TMS scan lines. For example, if sequences of pixels with different look angles have different autocorrelation functions, then we must confine our scan lines to only one look angle class or randomize over all look angle classes. Otherwise, the method of selecting the position of the sequence within the scan line is unimportant with respect to these alternatives. Once a rational method of selecting TMS sequences has been determined, we will address the question of the differences in the form of the autocorrelation which are attributable to the scale at which the observations are being made (80 m for MSS versus 7.5 m, at nadir, for TMS). In the third analysis, we will examine how the acf varies with changes in the land cover. Finally, we will look at the contribution of physiography to the acf by comparing the acfs of the Denver TMS data and the Cotter Basin TMS data.

A few caveats are appropriate at this juncture. This experiment was set up as a series of analyses, instead of one large experiment, to make use of the available data. As is the case when a large design must be attacked in pieces, the confounding of some effects in other effects and the failure to detect interactions are possibilities. Confounding of effects arises from having sources of variation which covary in the experiment so that, a significant result assigned to one effect may actually represent a significant contribution from the confounded effect to the variation "explained." Interactions can only be detected when the experimental design is completely crossed or factorial, i.e., each level of each factor appears in combination with each level of all other factors. In describing this analysis, we will point out where confounding might be a consideration. Loss of information related to undetectable interactions is always a concern in exploratory research which does not employ a factorial design. However we will not be able to address this problem any further here. Such problems could only be solved by executing a much larger

factorial design, hence the reason for the title containing the words "Preliminary evidence . . . . " However, this experiment will provide useful information as an input to the selection of factors and their levels in a larger design.

The research then is presented in the order of the experiments described above, preceded by a discussion of the data reduction procedure.

EXPERIMENTAL DESIGN AND DATA REDUCTION

For each of the experiments outlined in the previous section, an experimental design and randomization procedure was constructed so that results of formal statistical hypotheses could be directly translated into conclusions about the questions motivating the experiment. The formal statistical framework being used is known as analysis of variance (ANOVA). Since the techniques falling within the purview of analysis of variance are rather complex and very extensive, the reader interested in pursuing the mechanics or philosophy beyond what is presented below is recommended to see Fisher, 1971; Scheffe, 1959; and Dayton, 1970.

Figure 1 illustrates the summarization procedure applied to the data. The elements of the population are scan lines, or portions of scan lines which represent single samples within these experiments. The information about the relatedness of pixels is contained in a scan line, but not in a usable form. Therefore, the scan line is transformed into the acf and partial autocorrelation function (pacf). The elements of these two functions are analogous both in meaning and method of computation to the conventional correlation (Pearson product moment) and partial correlation statistics. While in principle n - 1 (n = the number of pixels in the sequence) values of the acf and pacf can be calculated, typically only the first few values are significantly different from zero. Craig and Labovitz (1980) have found that the first 10 values of the acf and pacf convey all the significant information. Since the acf and pacf are diagnostic of the appropriate ARIMA model (Box and Jenkins, 1970), we can use the first 19 distinct values from the acf and pacf (the first 10 values of the acf and the second through tenth values of the pacf, the first value of

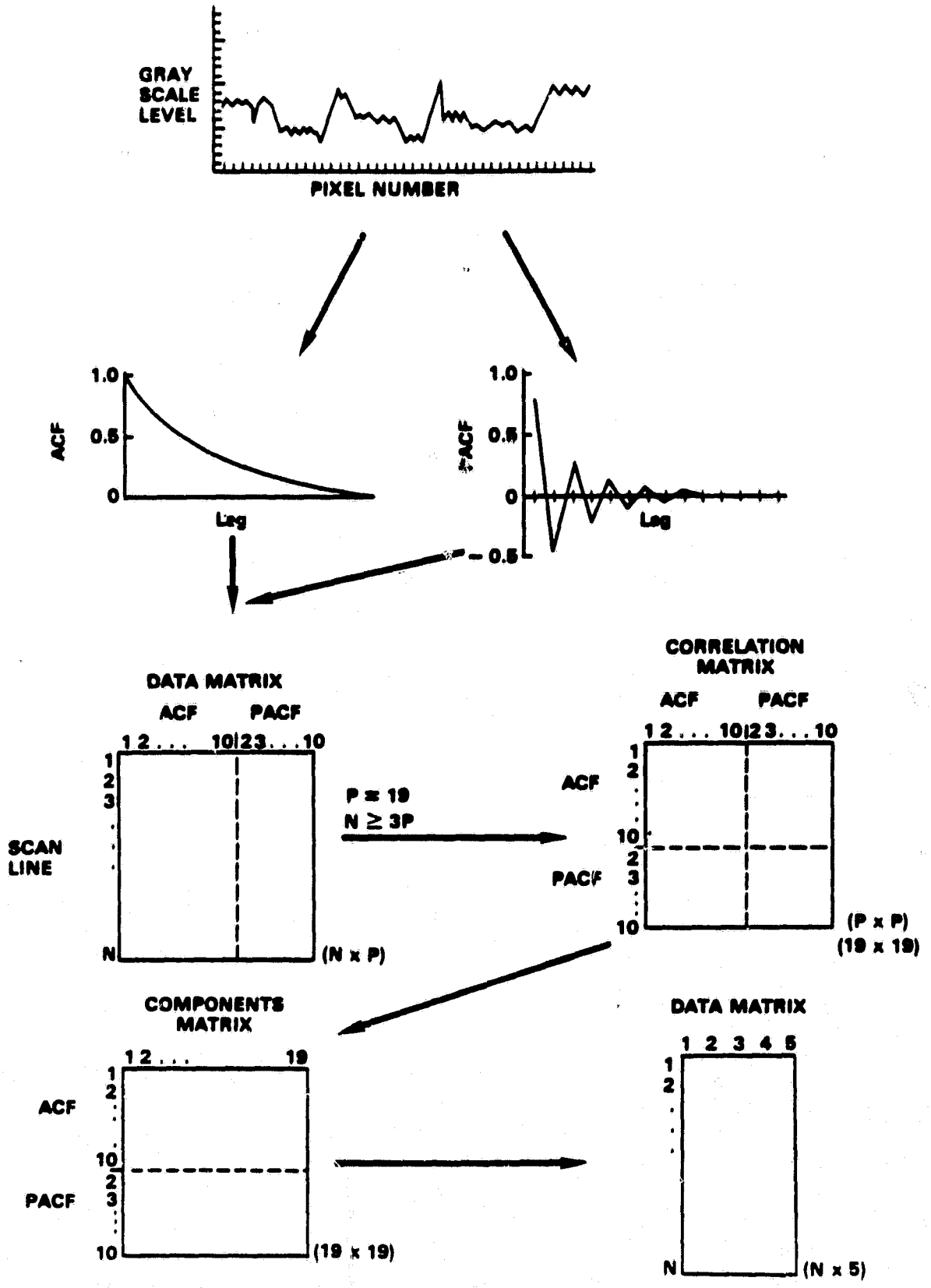


Figure 1. Procedure for transforming information about the relatedness of pixels from scan lines to composite scores.

the acf and pacf are equal). However, these values - the acf in particular - are correlated. Thus we can either analyze the 19 values using a MANOVA (Multiple - ANOVA) procedure or transform the 19 observations to a set of independent (orthogonal) observations using a principal components procedure to produce composite or factor scores. The latter procedure was chosen since Craig and Labovitz (1980) found the first two factors to be highly interpretable. The location effect dominates factor 1, while changes in the percentage of cloud cover are related to variation in factor 2 scores. In short we use a procedure which transforms the information in each sample from the intractable form of scan lines, processing upwards of 700 pixels, to a set of five independent composite scores.

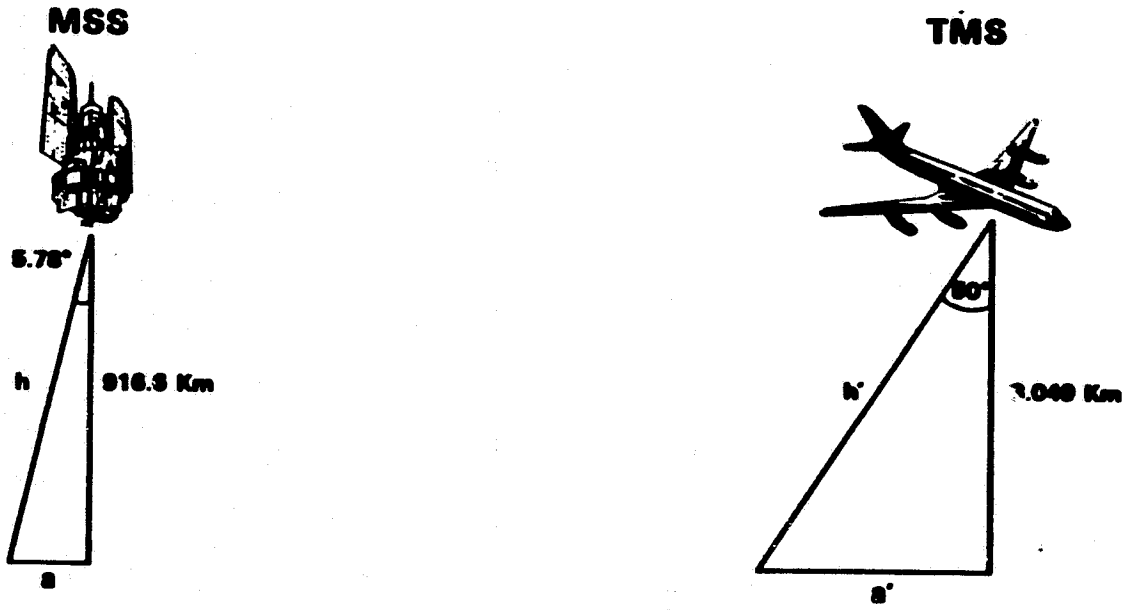
LOOK ANGLE AND ITS RELATIONSHIP TO THE ACF

A reasonable potential source of the variation between the forms of the acfs of MSS and TMS data is the widely differing scan angles possessed by the two systems. If the contribution to the variation from scan angle is significant, care must be taken in the manner in which scan lines are selected. Otherwise, look angle effects would be confounded in any hypothesis designed to test differences related to pixel size.

Using Figure 2, we can examine the effects of scan angle on pixel size. Under the assumption that Landsat 2 has a nominal altitude of 916.6km, a nadir pixel of 79m, and a scan angle of ±5.78° about nadir (NASA 1976), it is shown in Figure 2 that the width of a pixel increases by 0.40m or about 0.51 percent from the nadir pixel to either end of the scan line. On the other hand, the width of a TMS pixel, assuming an average altitude of 3.049km AGL, a nadir pixel of 7.5m, and a scan angle of ±50° about nadir, increases to 11.86m at the ends of a scan line. Thus, both the width and area of a TMS pixel increases by 58.1 percent in going from the center to the end of a scan line.

We now describe an experiment set up to test for the existence of a relation between look angle and the acf. This experimental design is a two-way factorial design with the main effects

MSS VERSUS TMS



$\frac{916.6}{h} = \cos 5.78°$

$h = 921.28$ Km

$\frac{3.048}{h'} = \cos 80°$

$h' = 4.743$ Km

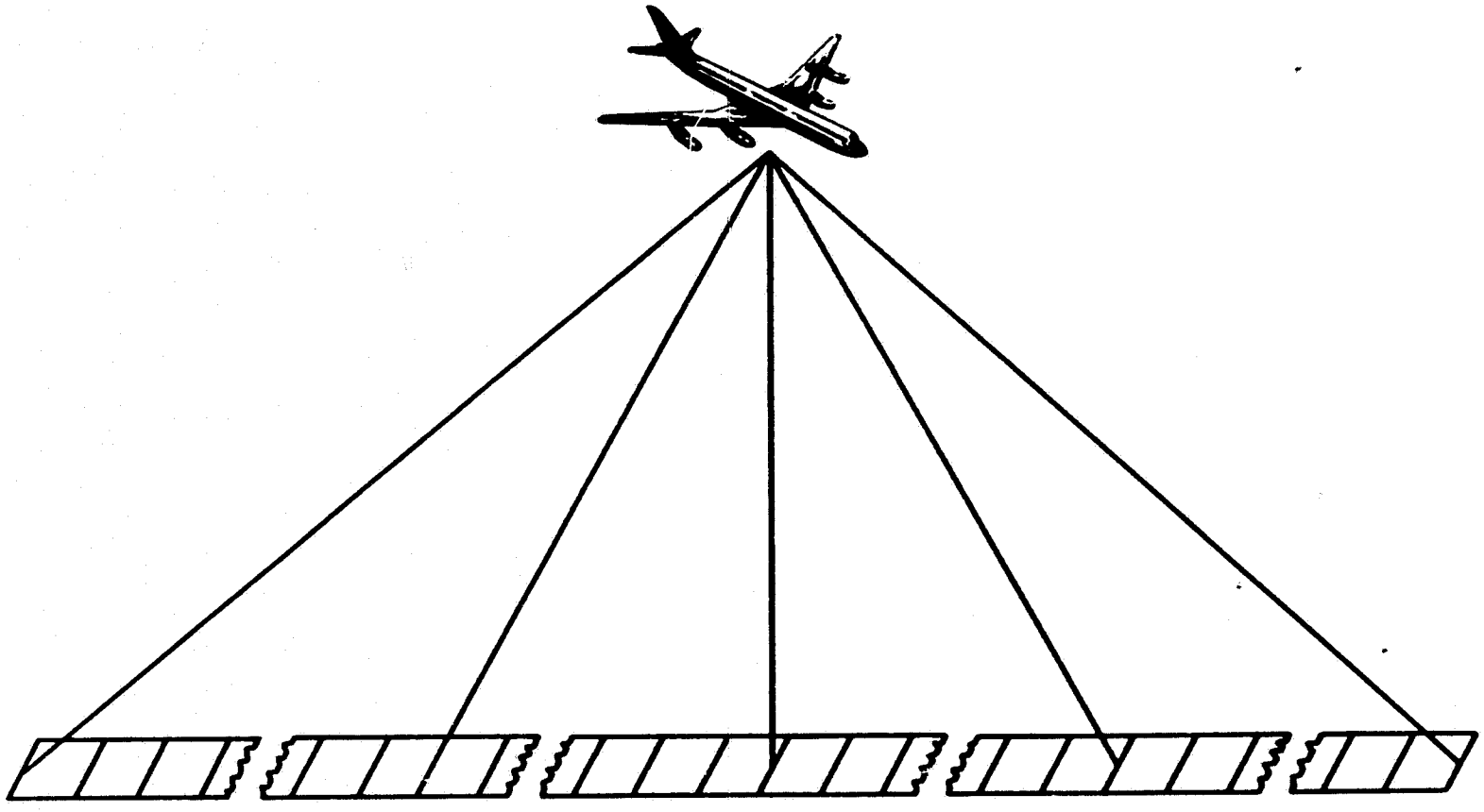
8.62×10^{-5}	IFOV (MILLIRADIANS)	2.5
79.40 m	ALONG SCAN PIXEL WIDTH AT EXTREME OF SCAN	11.86 m
79. m	WIDTH OF NADIR PIXEL	7.5 m
1.0051	RATIO OF WIDTHS OF EXTREME TO NADIR PIXELS	1.58

Figure 2. Geometry of look angle effects in TMS versus MSS (Landsat).

being channel and look angle. Two TMS channels are included in the analysis, channel 2 (0.52 - 0.60μm) and channel 6 (1.55 - 1.75μm). Selection of two of these channels allows us to examine first if an interaction between channel and look angle exists. Second, by selecting channels which are set well apart spectrally, we may test the cloud cover hypothesis suggested by Craig and Labovitz (1980). As has been previously noted, these authors found that differences in cloud cover were responsible for a significant pattern of variation in factor 2 derived from the principal component decomposition acf-pacf data matrix. The type of cloudiness being captured by this measure has been hypothesized to be general atmospheric haze. Thus, in selecting channels 2 and 6 of TMS, we have two channels which are differentially affected by atmospheric haze and backscattering. It would be reasonable then to expect a significant channel effect on factor 2 which would be attributed to the "cloudiness" of the scene. The reader should note that we are confounding cloudiness in channel, a move that is unavoidable with the data that are available. The look angle factor is divided into four levels, these are sequences of pixels collected at scan angle intervals of 50° - 25°, 25° - 0°, 0° - 25° and 25° - 50° (see Figure 3).

Since there are 19 variables, the first 10 elements of the acf and elements 2 through 10 of the pacf, 60 scan lines[2] per channel were selected at random throughout the flight lines. Thus, 120 scan lines were subset from the Computer Compatable Tapes (CCTs) using the VICAR program COPY as adapted on the Goddard IBM 360/91. One randomly chosen portion of each scan line, corresponding to a scan angle class, was used to calculate the acf and pacf. Since a TMS scan line is 700 pixels long, there are 175 pixels in each scan angle class. Figure 3 relates each scan angle class to its coding and its position on the scan line. Acfs and pacfs were calculated[3] for 15 sequences in each scan angle class. A typical acf and pacf are given in Figure 4. The exponential decline in the acf and the oscillatory behavior in the pacf are characteristic of an

[2]A common rule of thumb in choosing the number of samples is that the number should be greater than 3 times the number of variables being examined, in this case N should be greater than or equal to 3 x 19 = 58.
[3]All acf's and pacf's in this paper were calculated using a program written by Pack et al., 1972, as implemented on the IBM 370/3033 at the Pennsylvania State University.



Elements	1 - 175	176 - 350	351 - 525	526 - 700
Angle	50° - 25°	25° - 0°	0° - 25°	25° - 50°
Class	1	2	3	4

Figure 3. The design of the look angle classes for this experiment.

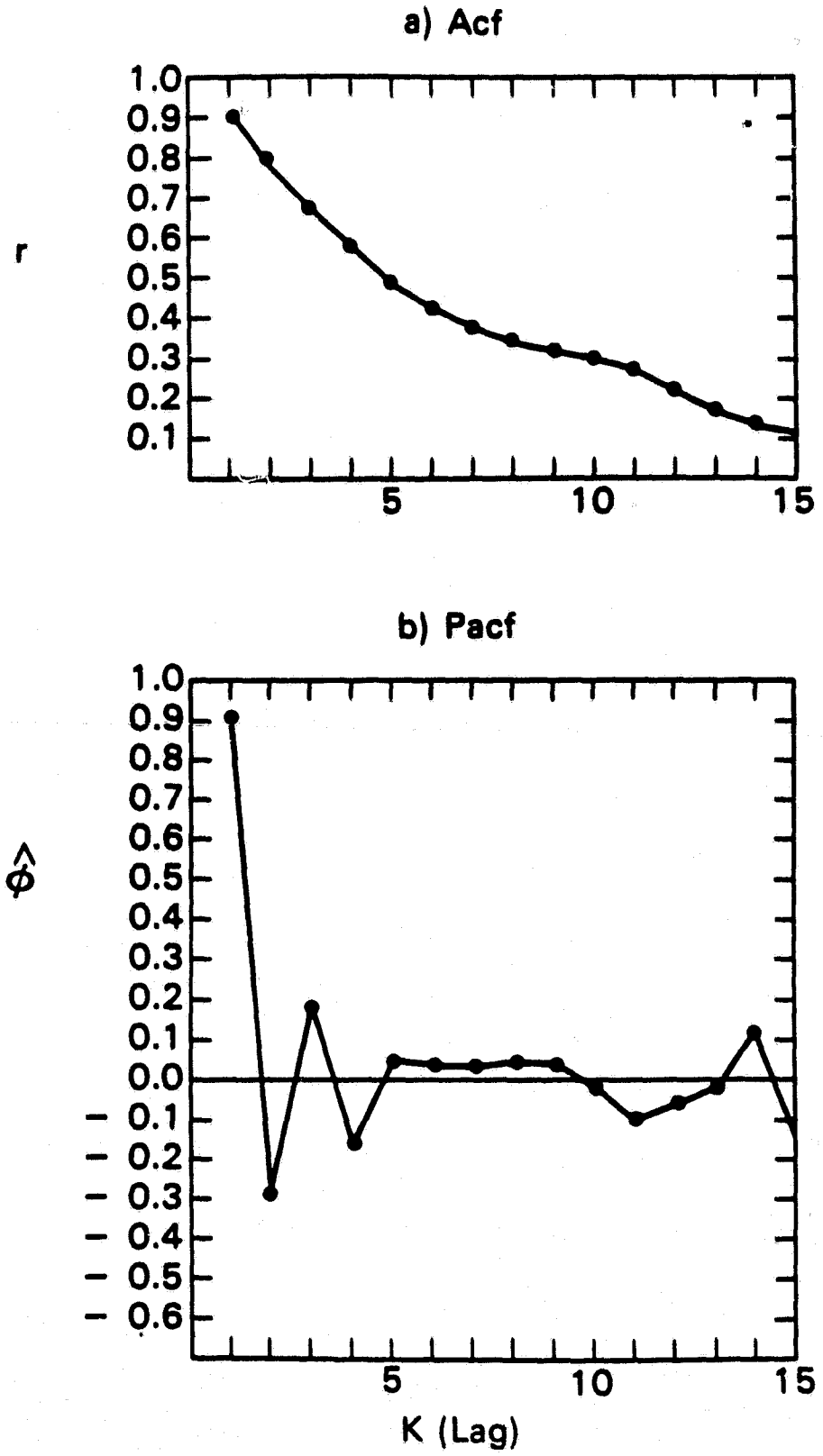


Figure 4. Typical acf and pacf from TMS data. These functions were derived from a randomly chosen scan line of length 175 from look angle class four.

ARIMA (1, 0, 1). The first 10 terms of acf and the 2nd through 10th elements of the pacf for each scan line were input into the SPSS subprogram FACTOR (Kim, 1975).

From the screen plot, in Figure 5, it was judged that the first five principal components contain the common portion of the reliable variation.[4] These composite scores were analyzed in five two-way analyses of variance using the P2V program of the BMDP series (Jennerick and Sampson, 1979a). The results for the first five composite scores are given in Table 2. Clearly there is no effect due to look angle. Further, the "haze" effect appears as previously hypothesized in the second composite score as a significant effect due to channel. No other effects appear to explain significant portions of the variation in the five composite scores.

The above analysis was repeated dividing the scan lines into 10 portions. The results were identical. We thus conclude that the look angle does not effect the acf or pacf and so sequences of pixels may be taken from any portion of the scan line.

SCALE AND THE ACF

Having demonstrated that any significant pattern of variation which discriminates TMS from Landsat is unlikely to be due to the large differences in the scan angles of the systems, we proceed to test the hypothesis about the relationship between pixel size and the acf. It should be noted from the outset that confounded within the scale effect is a system effect. However, (1) both systems are electro-mechanical scanners and (2) we may get some feel for the importance of the confounding by examining the same portion of the spectrum with each system. For example, if, despite using the same spectral bands we find a significant difference in the second composite score of the two systems, we might attribute this difference to system rather than scale differences. Therefore, scan lines of TMS channel 2 (0.52 - 0.60μm) and MSS channel 4 (0.50 - 0.60μm) were used for the analysis.

[4] A common model of factor analysis divides variation between that variation which is reliable and that which is noise. The reliable variation is in turn divided between variation which is shared by the random variables (common variation) and variation which is unique to a single variable. See Rummel, 1970 for further discussion of this model.

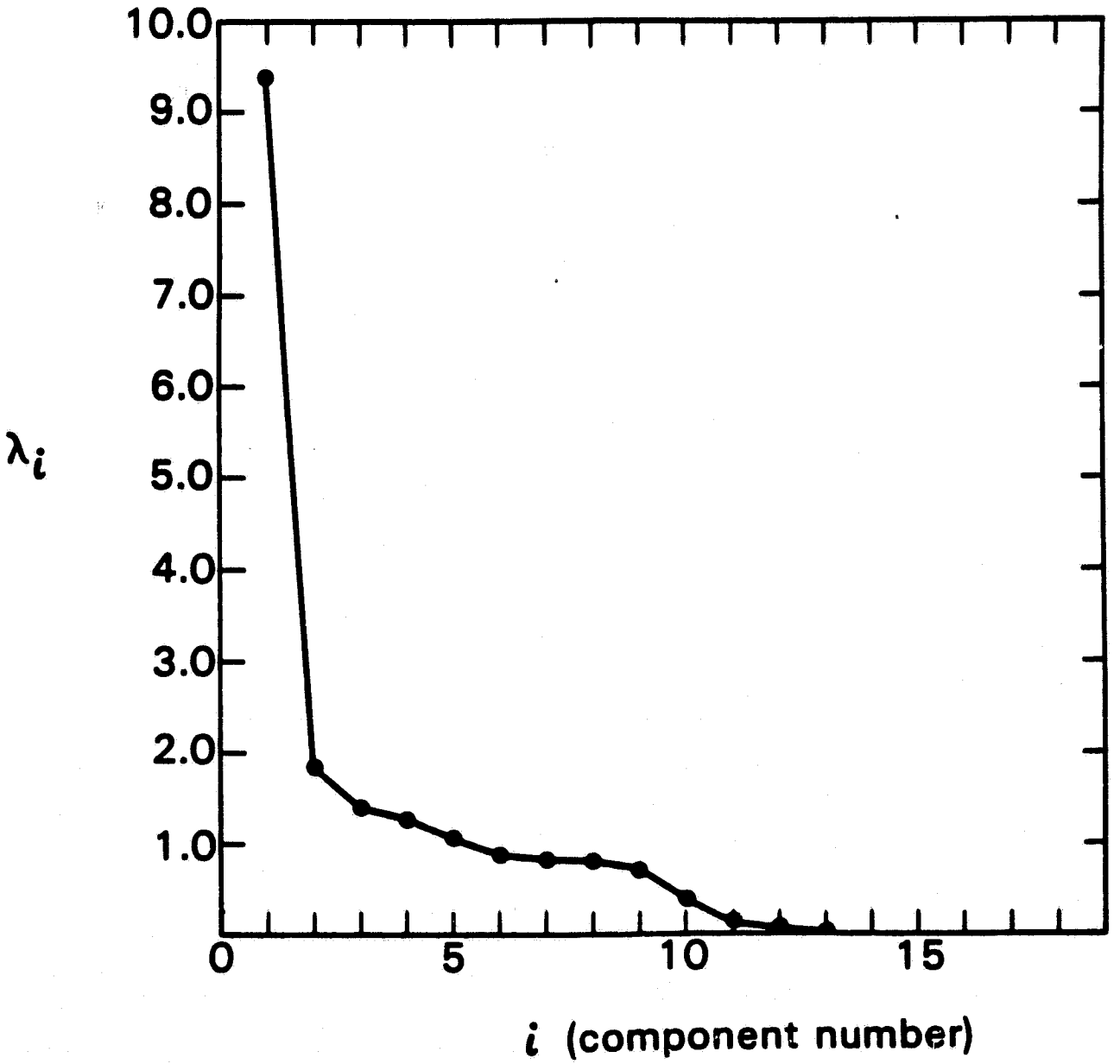


Figure 5. Screen plot generated from the eigenvalue decomposition of an acf - pacf correlation matrix. λ_i is the value of the eigenvalue associated with the i^{th} component.

The sampling and data manipulation procedures are the same as used previously. Sixty scan lines each composed of 475 pixels were selected randomly from both the TMS data and a corresponding area of the MSS data. Acfs, pacfs, and composite scores were calculated as before.

Table 3 contains the results from the one-way ANOVA's for the first two composites scores (calculations performed by the BMDP1V program [Engelman, 1979]). These analyses contain

Table 2
Two-way ANOVA (look angle by channel)
for first two composite scores

(a) FIRST COMPOSITE SCORE					
Source	Degrees of Freedom	Sum of Squares	Mean Square	F*	P(F > F*)
Angle	3	4.735	1.578	1.58	0.197
Channel	1	0.001	0.001	0.00	0.974
Angle X Channel	3	2.723	0.908	0.91	0.438
Error	112	111.540	0.996		
(b) SECOND COMPOSITE SCORE					
Source	Degrees of Freedom	Sum of Squares	Mean Square	F*	P(F > F*)
Angle	3	3.004	1.001	1.42	0.240
Channel	1	34.856	34.855	49.57	0.000
Angle X Channel	3	2.384	0.795	1.13	0.340
Error	112	78.754	0.703		

only one testable source of variation, scale, which is presented at two levels - 7.5 m and 80 m. The first composite score once again represents information in the acf, as all the terms of the acf load highly on the first component. Clearly, there is a significant pattern of variation across the first composite score. Meanwhile, there is no evidence of a "haze effect" (see composite score two, Table 3b) nor were any of the other three composite scores significant at any conventional confidence level.

The significant variation exhibited in the first composite score is specifically developed in Figure 6. The plot displays the means over 60 scan lines, of the first 10 terms of the acf for TMS-2 and MSS-4. Both mean acfs decline exponentially as is common for the acf of an ARIMA (1, 0, 1) process. However, the mean of each element of the TMS-2 derived data is significantly

Table 3
One-way ANOVA (scale factor) for first two composite scores calculated from scan lines of TMS-2 and MSS-4

(a) FIRST COMPOSITE SCORE					
Source of Variance	Degrees of Freedom	Sum of Squares	Mean Square	F*	P(F > F*)
Scale	1	23.299	23.299	28.73	0.000
Error	118	95.702	0.811		
(b) SECOND COMPOSITE SCORE					
Source of Variance	Degrees of Freedom	Sum of Squares	Mean Square	F*	P(F > F*)
Scale	1 1	0.240	0.240	0.24	0.626
Error	118	118.755	1.006		

higher than the corresponding MSS-4 derived mean value, even when we use a Bonferroni adjustment (Fisher, 1971) for the individual test confidence levels so that the group of tests has an overall α level of 0.01.

We may conclude on the basis of this analysis:

(1) It appears that the ARIMA (1, 0, 1) model is appropriate for the TMS data.

(2) The effect of scale on pixel size is to be found in the acf, that is in composite score one. From previous research (Craig and Labovitz, 1980), we believe that significant variation in the acf alone will be reflected in the ϕ_2 coefficient only.

(3) The TMS data is more highly autocorrelated than the MSS data. Initially we believe this means that the TMS data possess a greater pixel redundancy than MSS. This conjecture will need further examination in future research. However, two physical explanations (relative to the ground data), which can be accepted almost intuitively, would be supported by a hypothesis of greater redundancy generated by decreasing pixel size. These explanations would suggest that there are (1) a greater number of fields larger

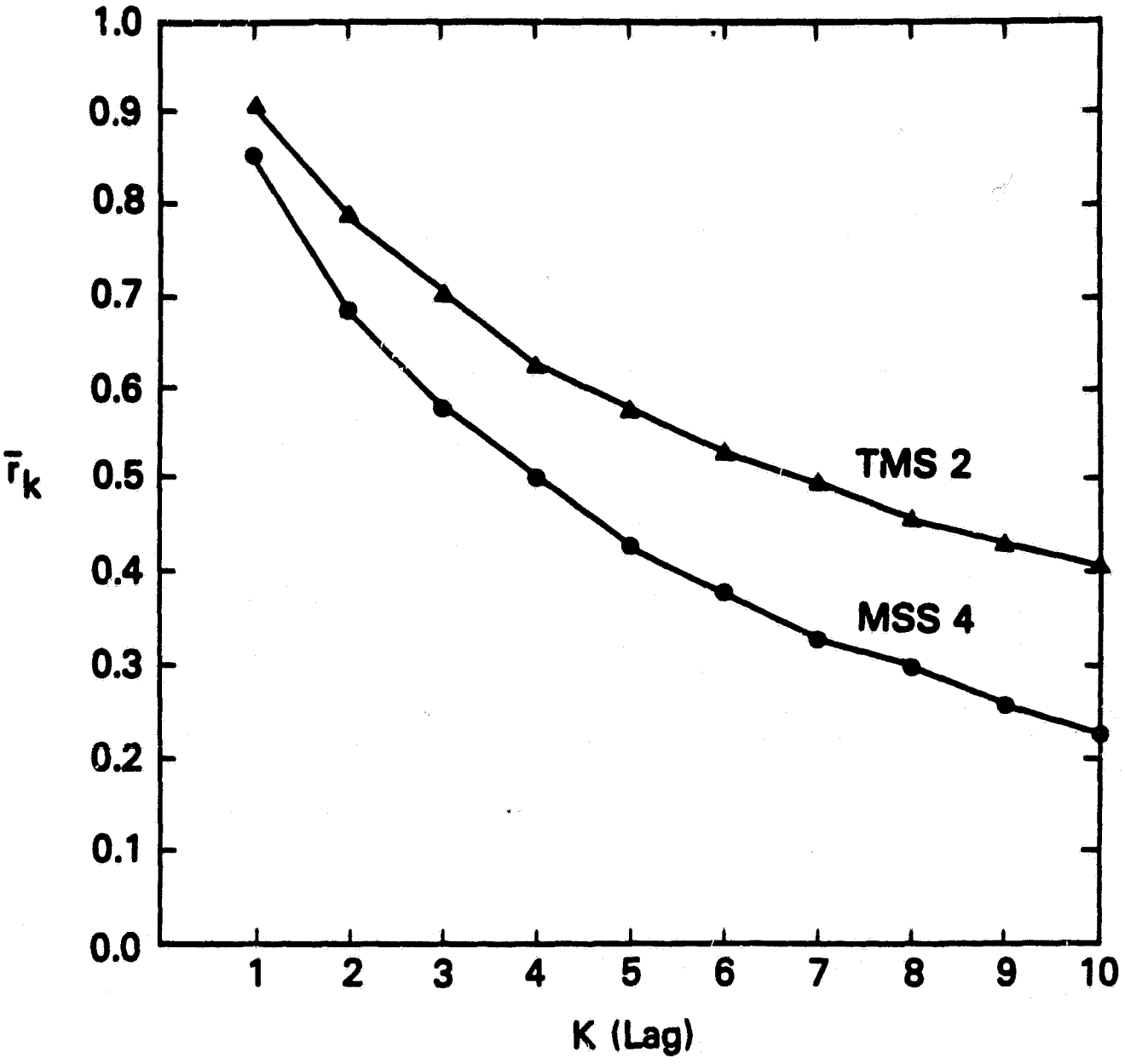


Figure 6. Mean values for the first 10 terms of the acfs derived from 60 MSS-4 scan lines and 60 TMS-2 scan lines.

than 7.5m than are larger than 80m and/or (2) that the lengths of slopes are such that more 7.5m pixels occur on one slope than 80m pixels. The next two sections will deal with the relative importance of landcover (field size differences) versus physiography (length of slopes) in the acf.

(4) The confounding of system in the scale effect, if it exists, is not of great importance.

THE INFLUENCE OF LAND COVER ON THE ACF

The effects upon the acf of four landcover classes – urban, agriculture, rangeland and a randomly selected "other" class – were examined in this experiment. The first three classes were defined on the basis of the Anderson lever I system (Anderson et al., 1976). The scan lines selected were obtained through the following procedure. Aerial photography of the Denver flight lines was photo-interpreted to an Anderson level I classification. The authors then constrained the classes examined to those which covered fairly large contiguous portions of the study area – a minimum of 300 pixels by 50 scan lines. This requirement reduced the number of classes to the aforementioned three – urban, agriculture and rangeland.

From among the contiguous areas of each of these classes, four areas (blocks) of each class were randomly selected. The blocks of the "other" class were selected by first determining the average size for blocks of the first three classes, dividing the flight lines up into areas of this mean size and randomly selecting four such areas. Randomly chosen examples of blocks from each class are shown in Figure 7. Within each block, two randomly selected scan lines were chosen for each of two channels, TMS–2 and TMS–4. In this manner, 4 (classes) x 4 (blocks) x 2 (channels) x 2 (scan lines) = 64 TMS scan lines were selected.

Two analyses were performed. The first one was the more elaborate. In this experiment, in addition to the TMS data, scan lines were randomly chosen from the TMS blocks averaged (TMSAVE) to 30 m pixels and the Landsat data of the same areas. Since the lengths of Landsat sequences from these locations were only 50 pixels, the TMS and TMSAVE sequence were reduced to this length by choosing random starting points. However, the structure of the components matrix of acfs and pacfs derived from these scan lines was unlike other matrices of components in that the acf split over the first two components. Further, there was no significant effect in the analysis of variance of the composite scores. It was concluded that scan lines 50 pixels in length were insufficient to bring out the structure, since the standard errors of the terms of the acf and pacf are dependent on the length of series.

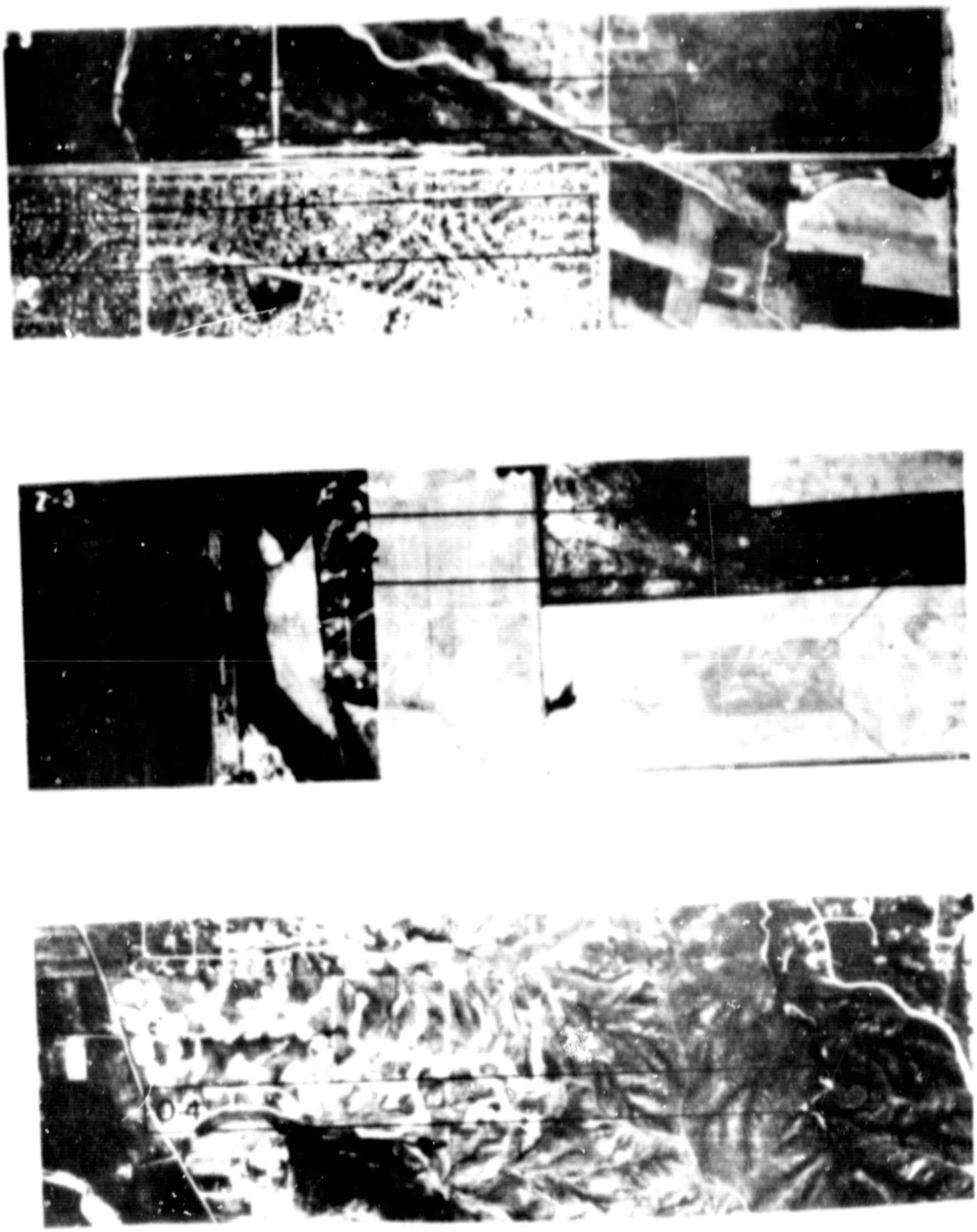


Figure 7. Randomly selected examples of land cover blocks used in this experiment.

In the second experiment, the Landsat and TMSAVE sequences were dropped from the analysis. The form of the experimental design used was a partially nested within a factorial design (Dayton, 1970). This design has a structural model given by:

$$Y^c_{ijkl} = \mu^c + \alpha^c_k + \beta^c_j + \alpha^c\beta^c_{kj} + \gamma^c_{i(j)} + \alpha^c\gamma^c_{ki(j)} + e^c_{ijkl}$$

where:

Y^c_{ijkl} is the composite score for the c^{th} component, $c = 1, 2, \ldots, 5$;

μ_c is the mean of the c^{th} composite score ($\mu^c = 0$ V c);

α^c_k is the contribution of k^{th} channel ($k = 1, 2$) to variation in the c^{th} composite score;

β^c_j is the contribution of the j^{th} land cover class ($j = 1, 2, 3, 4$) to variation in the c^{th} composite score;

$\alpha^c\beta^c_{kj}$ is the contribution of the interaction between channel and landcover class;

$\gamma^c_{i(j)}$ is the contribution of the i^{th} block ($i = 1, 2, 3, 4$) nested within the j^{th} landcover class;

$\alpha^c\gamma^c_{ki(j)}$ is the contribution of the interaction of the k^{th} channel with the i^{th} block nested within the j^{th} landcover class;

c^c_{ijkl} the error term associated with the c^{th} composite.

The scheme for the design is displayed in Table 4. Calculations for this rather complicated design were performed by the BMDP 7 program (Jenrich and Sampson, 1979b) and the results for the first two composite scores are given in Table 5. If we use the previously mentioned Bonferroni adjustment to set the overall α level for all 10 effects at 0.05, then the individual effects must be significant at the 0.05/10 = 0.005 level.

For the first composite score only the blocks effect is significant at the designated α level. The landcover effect contributes minor, if any, variation to the pattern of autocorrelation in the TMS data. Thus, the variation among blocks is greater than the variation among landcover classes. In the analysis of the second composite score the "haze" effect is once again apparent by the highly significant contribution to the variation of the channel.

Table 4
Partial nesting in a 2 x 4 factorial design

	LANDCOVER CLASS (B)															
Block (D(B))	Agriculture				Urban				Rangeland				Other			
	A1	A2	A3	A4	U1	U2	U3	U4	R1	R2	R3	R4	O1	O2	O3	O4
TMS–2 Channel (A)	S_1^c	S_2^c	S_3^c	S_4^c	S_5^c	S_6^c	S_7^c	S_8^c	S_9^c	S_{10}^c	S_{11}^c	S_{12}^c	S_{13}^c	S_{14}^c	S_{15}^c	S_{16}^c
TMS–4	S_{17}^c	S_{18}^c	S_{19}^c	S_{20}^c	S_{21}^c	S_{22}^c	S_{23}^c	S_{24}^c	S_{25}^c	S_{26}^c	S_{27}^c	S_{28}^c	S_{29}^c	S_{30}^c	S_{31}^c	S_{32}^c

S_i^c is the i^{th} sample from the c^{th} composite score; there are two composite scores for each sample.

The conclusion of this analysis is that landcover class, and by implication field size or number of boundaries, does not appear to be a major contributor to variation in the "location" effect.

THE EFFECT OF PHYSIOGRAPHY

Changes in landcover and changes in physiography were hypothesized by Craig and Labovitz (1980) as the most likely "causes" of the "location" effect. Having demonstrated that landcover is probably not a major factor, we will examine physiography by comparing data from Denver versus data from Cotter Basin. The two regions are characterized by very different physiographies. The Denver region is on the western edge of the Great Plains. The Cotter Basin on the other land is in the Northern Rocky Mountain Region, an area considerably more rugged.

Sixty scan lines 700 pixels long of TMS channel 2 were randomly selected from each location. After data reduction, five one-way ANOVA's were calculated on the composite scores. Once again the analysis of the first two scores are presented in Table 6. There is a significant variation due to physiography in the first and second composite scores (physiography is not significant in the other three scores). The pattern of variation in the first scores represents variation in the acf and hence the location effect of Craig and Labovitz (1980). This will be examined in

Table 5
Analysis of variance (partial nesting within 2 x 4 factorial design)
to test the land cover effect

(a) FIRST COMPOSITE SCORE						
Source	Error Term	Degrees of Freedom	Sum of Squares	Mean Square	F*	P(F > F*)
A	AD(B)	1	1.234	1.234	2.03	0.180
B	D(B)	3	26.526	8.842	5.00	0.018
AB	AD(B)	3	0.180	0.060	0.10	0.959
D(B)	E	12	21.232	1.769	8.67	0.000
AD(B)	E	12	7.294	0.608	2.98	0.007
E		32				
(b) SECOND COMPOSITE SCORE						
Source	Error Term	Degrees of Freedom	Sum of Squares	Mean Square	F*	P(F > F*)
A	AD(B)	1	26.237	26.237	49.28	0.000
B	D(B)	3	15.575	0.519	0.60	0.627
AB	AD(B)	3	3.046	1.015	1.91	0.182
D(B)	E	12	10.381	0.865	1.80	0.091
AD(B)	E	12	6.389	0.532	1.11	0.388
E		24				

A = Channel D(B) = Block Other terms are interactions
B = Landcover Class E = Error

greater detail below. It is unclear what produced the significant variation in the second score. It could be that the weather or the atmospheric conditions were different when the data were collected. Certainly the difference in solar angle between June (Denver) and August (Cotter Basin) is considerable and this in itself might be responsible for the variation.

Table 6
One-way ANOVA (location-physiography factor) for the first two composite scores calculated from scan lines of TMS-2

(a) FIRST COMPOSITE SCORE					
SOURCE	Degrees of Freedom	Sum of Squares	Mean Square	F*	P(F > F*)
LOCATION	1	16.347	16.347	18.97	0.000
ERROR	118	102.648	0.870		
(b) SECOND COMPOSITE SCORE					
SOURCE	Degrees of Freedom	Sum of Squares	Mean Square	F*	P(F > F*)
LOCATION	1	15.882	15.882	18.18	0.000
ERROR	118	103.113	0.874		

Returning to the information contained in the first composite score, Figure 8 is a plot of the means (over 60 observations) of the first 10 terms of the acfs derived from the Cotter Basin versus the acfs derived from the Denver data. The means are significantly different for each term at an overall α level of 0.01, with the means of the Cotter Basin acf higher than those from Denver. This implies that the Cotter Basin is more autocorrelated than the Denver data. This result suggests that the influence of slope and hence physiographic province upon the acf is considerable. Thus, we support the previously made assertions of Craig (1979) and Craig and Labovitz (1980) about the importance of slope. It will not be tested here, but we suspect that the results from the acfs arise form the Denver area having shorter length slopes than the Cotter Basin area.

SUMMARY AND CONCLUSIONS

We have tried through a series of experiments to determine the basis of the location effect of Craig and Labovitz (1980) and the applicability of the ARIMA (1, 0, 1) to data collected at different spatial resolutions. Subject to the limitations outlined in the proceeding text, we have demonstrated:

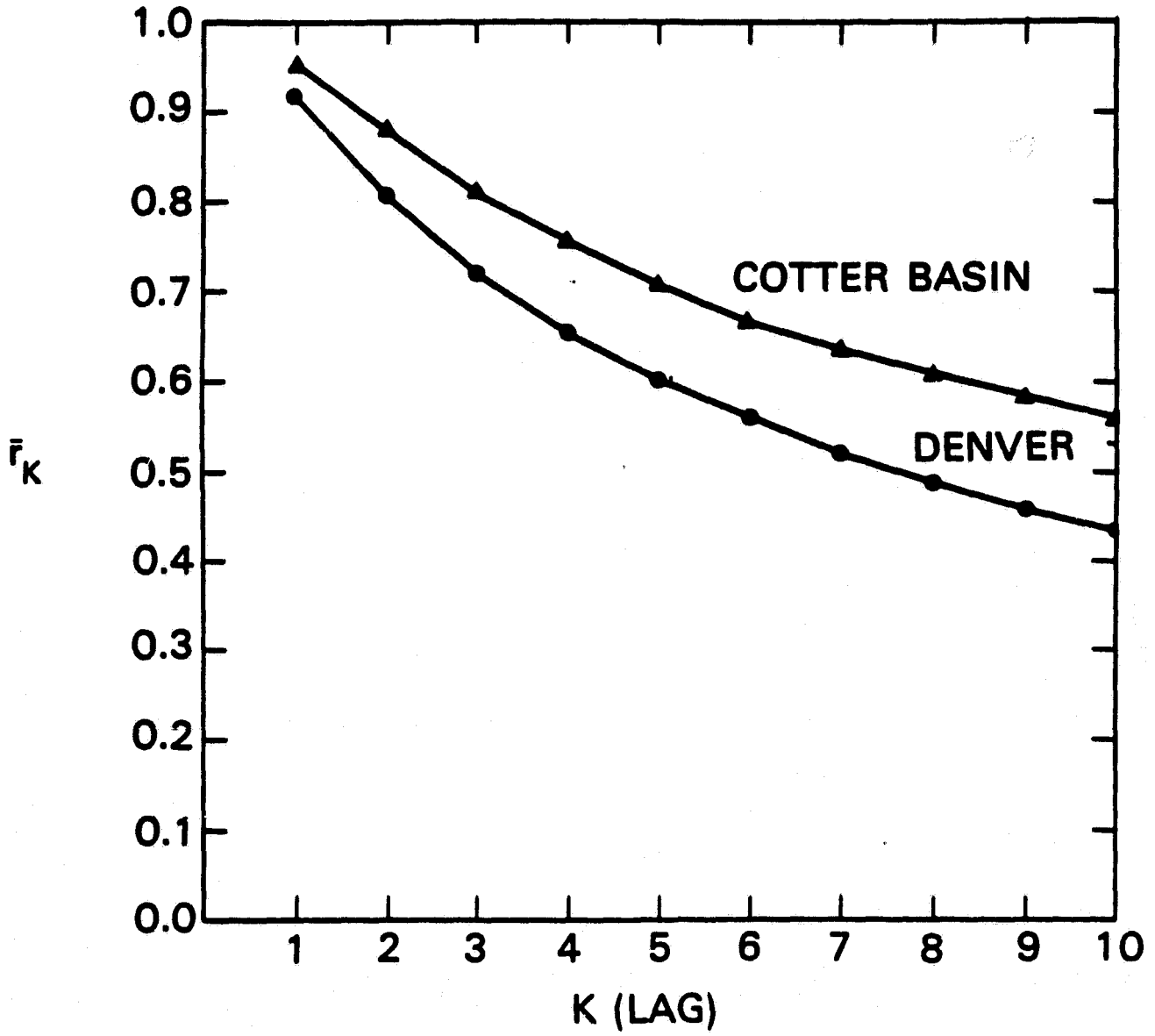


Figure 8. Mean values for the first 10 terms of the acfs derived from 60 TMS-2 scan lines over Denver and 60 TMS-2 scan lines over Cotter Basin.

(1) There is no impact of look angle on the acf;

(2) Changes in pixel size act to increase the autocorrelation of the data as the pixel size decreases, however, this change can be accommodated in the ϕ_1 coefficient and does not require abandoning the ARIMA (1, 0, 1) model;

(3) These is considerable support for the conjecture that atmospheric conditions are reflected on the second component which in turn is related to the θ_1 coefficient;

(4) Physiography is of far greater importance than land cover in explaining the location effect.

It is clear from the analyses that very useful information can be gleaned from the acf and pacf of remotely sensed data. Since the ARIMA (1, 0, 1) model is completely defined by the coefficients ϕ_1, θ_1 and $\sigma^2_{a_t}$, our results and those of Craig (previously referenced) go a long way towards describing the behavior of the model. If ϕ_1 is indeed largely controlled by physiography (holding pixel size constant), then this information can be fairly easily exploited. Since there are only 20 physiographic regions and subregions in the U.S. (Fenneman, 1938), ϕ_1 will take on a limited number of values.[5] Thus the value of ϕ_1 may readily be used to characterize the terrain being observed. Further having obtained an estimate of ϕ_1 (and θ_1) the data may be filtered by the model to yield the underlying stochastically independent process. Craig (in personal communication) has shown that the variance of a scan line and by implication the variance-covariance matrix is vastly inflated (approximately an order of magnitude depending on the values of ϕ_1 and θ_1) by the presence of an autocorrelation in the data. The filtering of the data is likely to enhance our ability to classify using remotely sensed data. Furthermore, since the variance is related to the information content of the data which in turn defines the number of bits needed to quantify the data, a decrease in the variance would theoretically allow us to code the information with fewer bits. This would allow substantial savings in the storage of Landsat and other digital imagery. We have not done very much research on looking at the feasibility of a coding scheme to exploit this information, but recommend that some attempt be made in this direction. Finally, since estimates of ϕ_1 and θ_1 are fairly easy to calculate (Nelson, 1973), both the filtering and recoding are excellent candidates for on-board satellite processing.

We wish to reiterate the need to verify our results by a more comprehensive study. This study should have a factorial design to satisfy concerns about the reliability of the results.

[5] It has been our experience as well as that of Craig that for remotely sensed data ϕ_1 is in the range 0.85 to 0.95 and θ_1 is in the range -0.35 to -0.45.

ACKNOWLEDGEMENTS

The authors wish to acknowledge the thoughtful reviews of this manuscript by several of their colleagues - Robin Bell, Lisette Dottavio, Edward Masuoka and Robert Price. The authors are particularly indebted to Richard Craig, Department of Geology, Kent State University, for his many helpful suggestions and free exchange of ideas.

REFERENCES

Anderson, J. R., Hardy, E. E., Roach, J. T. and Witmer, R. E., 1976, A Land Use and Land Cover Classification System for Use With Remote Sensor Data, U.S.G.S. Prof. Paper 964, 28 pp.

Box, G. E. P. and Jenkins, G. M., 1970, Time Series Analysis: Forecasting and Control, Holden-Day Publishing Co., San Francisco, 553 pp.

Craig, R. G., 1976, Comparison of Patterns from Earth Resources Technology Satellite Multispectral Scanner amd Glacial Drift, Northwestern Pennsylvania, Unpublished M.S. Thesis, Department of Geology, Pennsylvania State University, 378 pp.

Craig, R. G., 1979, Autocorrelation in Landsat data, Proceedings, 13th Int. Symp. on Remote Sensing of Environment, Univ. Michigan, 1517-1524.

Craig, R. G. and Labovitz, M. L., 1980, The location effect in sampling Landsat data, Proc. 14th Int. Symp. on Remote Sensing of Environment, Costa Rica, in press.

Dayton, C. M., 1970, The Design of Educational Experiments, McGraw-Hill Co., New York, 250 pp.

Engelman, L., 1979, P1V: One-way analysis of variance and covariance, in: BMDP Biomedical Computer Programs, P - Series, W. J. Dixon and M. B. Frown, eds., University of California Press, Berkeley, CA, p. 523-539.

Fenneman, N. M., 1938, Physiography of Eastern United States, McGraw-Hill Book Co., Inc., New York, 714 pp.

Fisher, R. A., 1971, The Design of Experiments, 9th ed., Hafner Press, New York, 248 pp.

Jennrich, R. and Sampson, P., 1979a, P2V: Analysis of Variance and Covariance including Repeated Measures, in: BMDP Biomedical Computer Programs, P - Series, W. J. Dixon and M. B. Brown, eds., University of California Press, Berkeley, CA, p. 540-580.

Jennrich, R. and Sampson, P., 1979b, P8B: General mixed model analysis of variance-equal cell sizes, in: BMDP Biomedical Computer Programs, P - Series, W. J. Dixon and M. B. Brown, eds., University of California Press, Berkeley, CA, p. 598.1-602.1.

Kim, J., 1975, Factor Analysis, in: Statistical Package for the Social Sciences, N. J. Nie, C. H. Hull, J. G. Jenkins, J. Stenbrenner, D. H. Bent, eds., McGraw-Hill Book Co., New York, p. 468-514.

NASA, 1976, Landsat Data Users Handbook, Document No. 76 SDS 4258, Goddard Space Flight Center, Greenbelt, MD 20771.

Nelson, C. R., 1973, Applied Time Series Analysis for Managerial Forecasting, Holden-Day, Inc., San Francisco, 231 pp.

Pack, D. J., Goodman, M. L. and Miller, R. B., 1972, Computer programs for the analysis of univariate time series using the methods of Box and Jenkins, Technical Report No. 296, Department of Statistics, University of Wisconsin, Madison.

Rummel, R. J., 1970, Applied Factor Analysis, Northwestern Univ. Press, Evanston, Ill., 617 pp.

Scheffe, H., 1959, The Analysis of Variance, John Wiley & Sons, New York.