MIN/MAX AUTOCORRELATION FACTORS FOR MULTIVARIATE SPATIAL IMAGERY

by

Paul Switzer
Stanford University

and

A. A. Green
C.S.I.R.O. Australia

TECHNICAL REPORT NO. 6
APRIL 1984

PREPARED UNDER THE AUSPICES OF
OF
NATIONAL SCIENCE FOUNDATION
GRANT MCS 81-09584

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

MIN/MAX AUTOCORRELATION FACTORS FOR MULTIVARIATE SPATIAL IMAGERY

by

Paul Switzer
Stanford University

and

A. A. Green
C.S.I.R.O. Australia

TECHNICAL REPORT NO. 6
APRIL 1984

PREPARED UNDER THE AUSPICES OF
OF
NATIONAL SCIENCE FOUNDATION
GRANT MCS 81-09584

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

MIN/MAX AUTOCORRELATION FACTORS FOR MULTIVARIATE SPATIAL IMAGERY

Paul Switzer
Stanford Univeristy

A. A. Green
C.S.I.R.O. Australia

Abstract

Gridded multivariate data, typical of satellite and other multi-channel remote sensed data, may be linearly transformed pointwise by a procedure which tends to isolate the noise component of the data. Some properties of this MAF (min/max autocorrelation factors) procedure are explored using simple spatial stochastic structure. An example is given using ten-channel imagery.

1. Introduction and Summary

We present a method for general purpose processing of multi-channel data on a spatial grid with a view to isolating signal and noise components of the data. At each grid point x in the region of the image we have a p-variate measurement denoted $\mathbf{Z}(x)$. This measurement consists of a p-variate signal $\mathbf{S}(x)$ contaminated by p-variate noise $\epsilon(x)$, neither of which is separately observable and for which the separate statistics are not typically available. We will propose a procedure (MAF) which transforms $\mathbf{Z}(x)$ linearly to a new set of p-variates which have the following property: the low-number variates have minimal spatial autocorrelation, identified as mainly noise, and the high numbered variates have maximal spatial autocorrelation, identified as mainly signal.

The proposed procedure differs from naive principal components in that it makes explicit use of some global spatial statistics of the observable data. It also differs from filtering theory in that it does not require separate statistics for the unobservable signal and noise.

The plan of the paper is first to discuss briefly a few shortcomings of procedures based on spatial filtering (local averaging) and standard principal components. Next the MAF procedure is described together with some of its properties in the context of simple spatial models. Finally, an example is presented of ten-channel imagery in which noise separation seems to have been successfully accomplished but which did not yield as well to standard procedures.

2. Spatial Smoothing and Signal Blurring

Noise separation is conventionally attempted using spatial smoothing algorithms such as moving averages applied separately to each data channel. Such procedures will indeed reduce the variance of the noise but will inevitably create some blurring of the signal. This reduction in local signal contrast may impair subsequent image interpretation and classification.

It is possible to quantify this contrast reduction due to local averaging in the context of certain spatial stochastic models. For example, consider a univariate gaussian signal process with a smooth isotropic autocorrelation function $\rho(\Delta)$, where Δ is the spatial lag. Suppose we create a simple moving average using a circular window of radius δ. The locally averaged signal is then rescaled to have the same variance as the original signal. Let g be the mean absolute gradient of the original signal process and let g_δ be the mean absolute gradient of the locally averaged signal. Then it can be shown that

$$\frac{g}{g_\delta} = 1 - \delta^2 \left\{ \frac{\frac{1}{6}\rho^{(iv)}(0) - \frac{1}{4}|\rho''(0)|^2}{|\rho''(0)|} \right\} + 0(\delta^4).$$

Suppose further that

$$\rho(\Delta) = [1 + (a\Delta)^2]^{-2}$$

at short distances Δ, and that $\rho(\delta) = 0.8$. Then g/g_δ is $0.76 + 0(\delta^4)$, giving a 24% reduction in signal contrast for this example.

As an alternative to the blurring associated with local spatial averaging, attempts have been made (2), (3) to isolate noise in multivariate imagery by operating pointwise on the data for each pixel. The idea is to somehow use the correlation or redundancy in the information provided by different data channels. The common form of this approach is to extract principal component factors of the $p \times p$ covariance matrix Σ, computed by treating the pixels as replicates of a p-variate observation. The low numbered factors, having smallest variance, are treated as noise factors and the high numbered factors, having largest variance, are treated as signal factors.

It is not immediately obvious why such a variance criterion should distinguish signal from noise. Furthermore, no spatial properties of the image are used to define the factors, i.e., if the pixels were randomly rearranged the factor definitions would not change. But the most serious shortcoming of naive principal components factors are their lack of invariance to rescaling of the data. For example, differential stretching of data channels will alter the factor definitions, as will the reduction of the covariance matrix to a correlation matrix.

Still, application of the naive pointwise procedure does, in some examples, produce a set of factors which seem to show decreasing amounts of spatial structure. Later, we will try to provide a rationale for this outcome using a simple class of spatial stochastic models, in order to indicate when one might expect the naive procedure to succeed.

Of course if one has a full specification of the separate and joint spatial properties of both the disentangled signal and noise components of the multivariate image data, then it is possible to prescribe optimal least squares estimates of the signal itself (4). These estimates will have the form of moving averages, and they operate simultaneously on all the data channels. While such procedures are optimal pointwise for reconstructing the signal, nevertheless the spatial gradient of the estimated signal will show a downward bias in its magnitude which may impair subsequent interpretation.

3. Min/Max Autocorrelation Factors

We now propose a noise separation procedure for general purpose processing of multivariate imagery which operates pointwise to avoid the signal blurring introduced through smoothing or spatial averaging procedures. However, the pointwise operator itself is defined using primitive global spatial characteristics of the data. So, in principle, our procedure overcomes the qualitative objections to both spatial smoothing and naive pointwise data processing.

Define p orthogonal linear combinations $\mathbf{Y} = (Y_1, ..., Y_p)$ of the original multivariate observation vector $\mathbf{Z} = (Z_1, ..., Z_p)$, called MAF (min/max autocorrelation factors), with the following property: Let $Y_i(x) = \mathbf{a}_i' \mathbf{Z}(x), i = 1, ..., p$, and let $r_i(\Delta) = \text{correlation}(Y_i(x), Y_i(x + \Delta))$. Then

$$r_1(\Delta) = \min_a \text{correlation } (a' Z(x), a' Z(x + \Delta))$$
$$r_p(\Delta) = \max_a \text{correlation } (a' Z(x), a' Z(x + \Delta))$$

$$r_i(\Delta) = \min_a \text{correlation } (a' Z(x), a' Z(x + \Delta)) \qquad \text{and}$$

$$\text{correlation}(a' Z(x),\ a_j' Z(x)) = 0 \text{ for } j < i.$$

If we suppose that the image noise is weakly autocorrelated compared with the autocorrelation of the image signal, then it is reasonable to expect that the low numbered factors will be mainly noise. The MAF procedure would then provide the desired noise separation, leaving the high numbered factors relatively uncontaminated.

An important property of the MAF procedure is its invariance to linear transforms, a property not shared by naive pointwise principal components procedures or its extensions to spatial or temporal series (5). Specifically, if $\mathbf{Z}^* = \mathbf{BZ}$ where $\mathbf{B}$ is any $p \times p$ nonsingular matrix, then the MAF solution for $\mathbf{Z}^*$ is the same as the MAF solution for $\mathbf{Z}$. In particular, it is irrelevant whether or not the data have been rescaled so that each frequency band has the same range of observed values, a common practice in satellite image processing. It is also not important to know the instrument gain factors for each of the frequency bands.

The MAF procedure has other formulations which permit the use of standard multivariate routines to extract the factors. Specifically, the factors are obtained as the eigenvectors of a matrix $\Sigma_\Delta \Sigma_0^{-1}$ where

$$\Sigma_\Delta = \mathrm{cov}\{\mathbf{Z}(x) - \mathbf{Z}(x+\Delta)\}$$
$$\Sigma_0 = \mathrm{cov}\{\mathbf{Z}(x)\}.$$

This formulation is somewhat more convenient than the canonical correlation formulations where $\mathbf{Z}(x)$ and $\mathbf{Z}(x+\Delta)$ are considered to be two sets of p variables observed at each x. Since the canonical factors for the first set of p variables must be identical to the corresponding canonical factors for the second set, it will be necessary in practice to fix up the $2p \times 2p$ covariance matrix of $(\mathbf{Z}(x),\ Z(\mathbf{x}+\Delta))$ so that it has the necessary symmetry.

As a further computational convenience the eigenvectors of $\Sigma_\Delta \Sigma_0^{-1}$ are obtained in three steps.

(i) The original data $\mathbf{Z}(x)$ are linearly transformed to any $\mathbf{Z}^*(x)$ where $\mathrm{cov}\{Z^*(x)\} = I_{p\times p}$. This may be achieved by using principal components derived from the original global covariance or correlation matrix Σ_0.

(ii) Form two sets of differences of the orthogonalized data, viz. $[\mathbf{Z}^*(x) - \mathbf{Z}^*(x+\Delta')]$ and $[\mathbf{Z}^*(x) - \mathbf{Z}^*(x+\Delta'')]$ where $\Delta\prime$ is a unit horizontal shift and Δ'' is a unit vertical shift; calculate the corresponding two global covariance matrices $\Sigma^*_{\Delta'}$ and $\Sigma^*_{\Delta''}$ and pool them to form Σ^*_Δ.

(iii) Obtain the principal components corresponding to the pooled covariance matrix Σ^*. This is the MAF solution.

Finally, a grey scale map is produced for each of the orthogonal factors obtained for further interpretation; each of these should then be scaled to have constant variance. Typically, one might expect the first or first two maps to exhibit mostly noise. The example given in the next section shows this property quite clearly. If it desired to map the original variables with most of the noise suppressed, then one would map their projections onto that subspace which excludes the low numbered factors.

4. MAF and the Proportional Covariance Model

Although it seems somewhat plausible that factors with minimum spatial autocorrelation should concentrate and isolate noise, this property of MAF can be demonstrated to hold exactly in the context of a very simple class of spatial models. Suppose the signal and noise components of the observation vector $\mathbf{Z}(x)$ are uncorrelated, i.e.,

$$\mathrm{cov}\{\mathbf{Z}(x)\} \equiv \Sigma = \Sigma^S + \Sigma^\epsilon,$$

where Σ^S, Σ^ϵ are the covariance matrices, respectively, of the signal and noise. Furthermore, suppose the cross-variance of the signal and noise are each attenuating at different rates as a function of spatial lag Δ. Thus,

$$\begin{aligned} \mathrm{cov}\{\mathbf{S}(x),\ \mathbf{S}(x+\Delta)\} &= b_\Delta \cdot \Sigma^S \\ \mathrm{cov}\{\epsilon(x),\ \epsilon(x+\Delta)\} &= c_\Delta \cdot \Sigma^\epsilon. \end{aligned}$$

Without real loss of generality we may take $\Sigma = I_{p\times p}$. Define

$$\lambda_\Delta(\mathbf{a}) = \mathrm{Cov}\{\mathbf{a}'\,\mathbf{Z}(x),\ \mathbf{a}'\,\mathbf{Z}(x+\Delta)\}$$

where $\mathbf{a}'\,\mathbf{a} = 1$. Then, for the proportional covariance model of this section,

$$\begin{aligned} \mathrm{Var}\{\mathbf{a}'\,\epsilon(x)\} &= [b_\Delta - \lambda_\Delta(\mathbf{a})]/[b_\Delta - c_\Delta] \\ \mathrm{Var}\{\mathbf{a}'\,\mathbf{S}(x)\} &= [\lambda_\Delta(\mathbf{a}) - c_\Delta]/[b_\Delta - c_\Delta]. \end{aligned}$$

Therefore, the signal-to-noise ratio for the projection $\mathbf{a}'\,\mathbf{Z}(x)$ is

$$[\lambda_\Delta(\mathbf{a}) - c_\Delta]/[b_\Delta - \lambda_\Delta(\mathbf{a})].$$

The MAF maximal factor does indeed maximize $\lambda_\Delta(\mathbf{a})$ subject to $\mathbf{a}'\mathbf{a} = 1$. So, provided $b_\Delta > c_\Delta$, i.e., signal autocorrelation attenuates more slowly than noise autocorrelation, then the maximization, over projections $\mathbf{a}$, of the signal to noise ratio is equivalent to the maximal MAF. Likewise, the minimum signal-to-noise ratio is achieved for the minimal MAF.

While the proportional covariance model is hardly universal, it does motivate the MAF procedure as an optimal procedure in the class of linear factor models operating pointwise on the data. It should be noted also that the MAF, in the present context, does not depend on the spatial shift Δ, i.e., any choice of Δ gives the same set of factors, whether or not the spatial autocorrelations are isotropic. In particular, one would expect to have the same MAF using horizontal or vertical lags of any size. As a practical matter it seems that the proportional covariance model is likely to approximate the spatial structure of an image only over very short lags. Therefore, it seems prudent to use single step lags for the MAF procedure, as described in the last section and as used in the example of the next section.

Finally, a further specialization of the proportional covariance model provides insight into the occasional or frequent success of the naive principal component procedure which uses no spatial properties at all of the image. Suppose, in addition to the proportionality described above, one also has

$$\operatorname{cov}\{\epsilon(x)\} = \sigma^2 \cdot I_{p\times p}.$$

This says that all frequency bands have the same noise variance, although not necessarily the same signal variance. Using the previous notation it follows that

$$\Sigma_\Delta \Sigma_0^{-1} = b_\Delta \cdot I_{p\times p} - \sigma^2(b_\Delta - c_\Delta)\Sigma_0^{-1}.$$

Since the MAF procedure finds the eigenvectors of $\Sigma_\Delta \Sigma_0^{-1}$, then in this rather special case the MAF procedure is equivalent to finding the eigenvectors of Σ_0, i.e., the naive principal components. The spatial structure plays no role, even though both signal and noise may be spatially autocorrelated. Of course, one no longer has linear invariance; rescaling of variables, in general, will not preserve the condition of constant noise variance.

5. Example

The data of the example are ten-channel gridded values produced by the U-2 Thematic Mapper Simulator. This is a high altitude scanner which simulates the spatial and spectral band characteristics of the seven LANDSAT-D Thematic Mapper plus three additional intermediate frequency bands. The grid spacing or pixel dimensions are 28 meters square and the image contains 716 $\times$ 716 pixels. Each datum is an eight bit number providing a measurement of reflected energy integrated over a specific pixel for a given frequency band.

The area of the image is the vicinity of the Silver Bell Copper Mine in Arizona. This image was chosen merely as an illustration of the genreal purpose MAF procedure and no interpretive processing is proposed in this paper. The salient larger features of this image are mountains on the lower left, tailings ponds on the upper left, two working pits in the center and lower right, and a road on the right.

Figure 1 shows mapped values of the ten orthogonal factors with smallest to largest spatial autocorrelation. It seems that the first two factors have provided clear-cut noise separation, something that did not occur with a naive principal components analysis of the same data.

It is evident that even in the factor maps with smallest autocorrelation there is a recurring prominent spatial feature. This suggests strongly that this image be processed as two subimages using separate global spatial characteristics for each subimage. In general one should choose regions of "stationarity" of an image, which may be done in a somewhat interactive way as suggested here. The concept of stationarity in data analysis is an arbitrary one to a large extent, reflecting merely the extent over which one wishes to compute global statistics.

References

(1) Switzer, P., Geometrical Measures of the Smoothness of Random Functions, *J. Appl. Prob.* **13** (1976), 86-95.

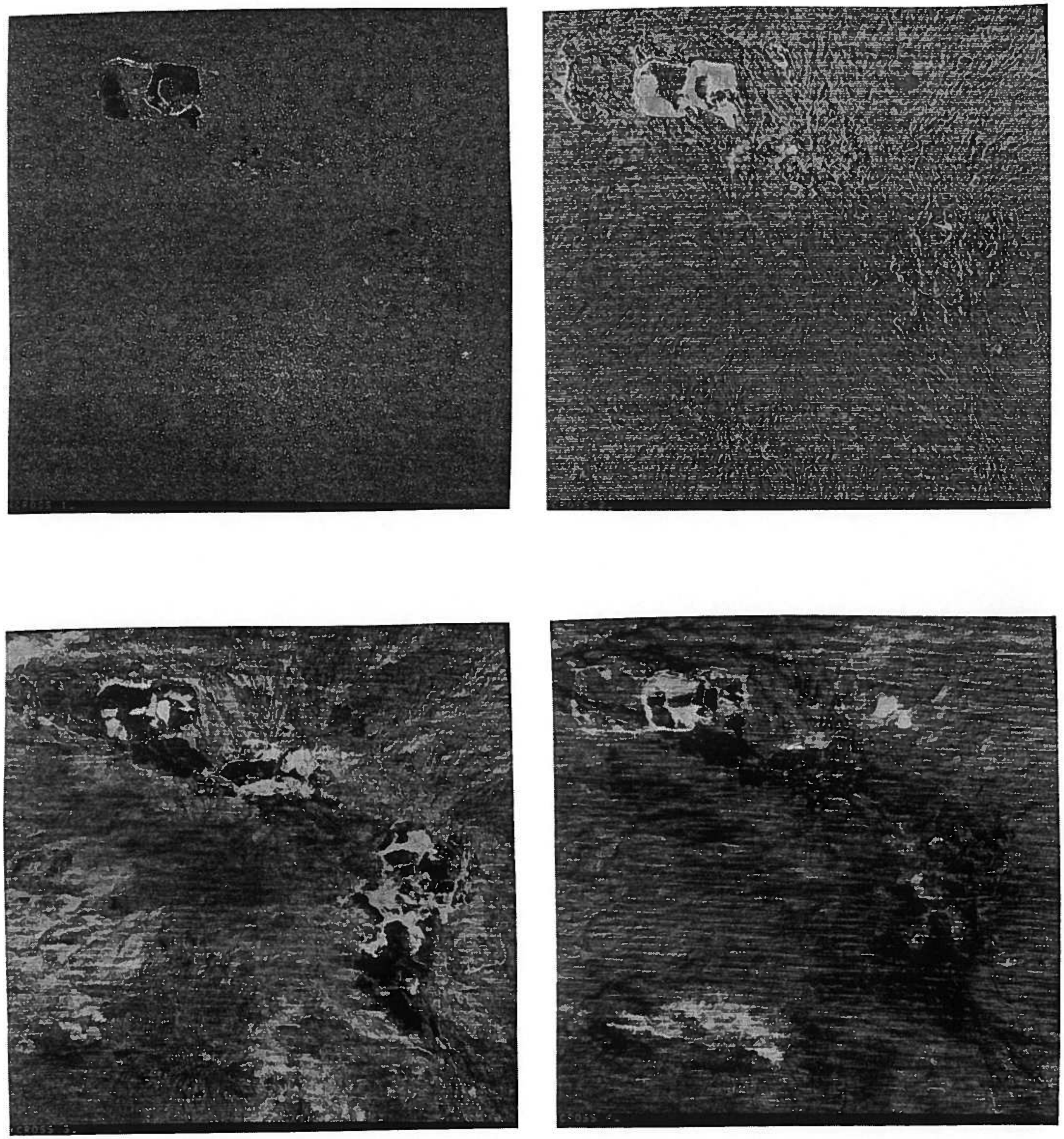
(2) Crist, E.P. and Cicone, R.C., Investigations of Thematic Mapper Data Dimensionality and Features using Field Spectrometer Data, *17th Int. Symp. on Remote Sensing of Env.*, Ann Arbor (1983).

(3) Ready, P.J. and Wintz, P.A., Information Extraction, SNR Improvement, and Data Compression in Multispectral Imagery, *IEEE Trans. on Comm.* Com-21 (1973).

(4) Hannan, E.J., **Multiple Time Series**, John Wiley and Sons, Inc. New York (1970).

(5) Brillinger, D.R., **Time Series: Data Analysis and Theory**, Holt, Rinehart, and Winston, New York (1981).

FIGURE 1. Min/Max Autocorrelation Factors: Ten channel thematic mapper simulator, Silver Bell copper mine, Arizona.



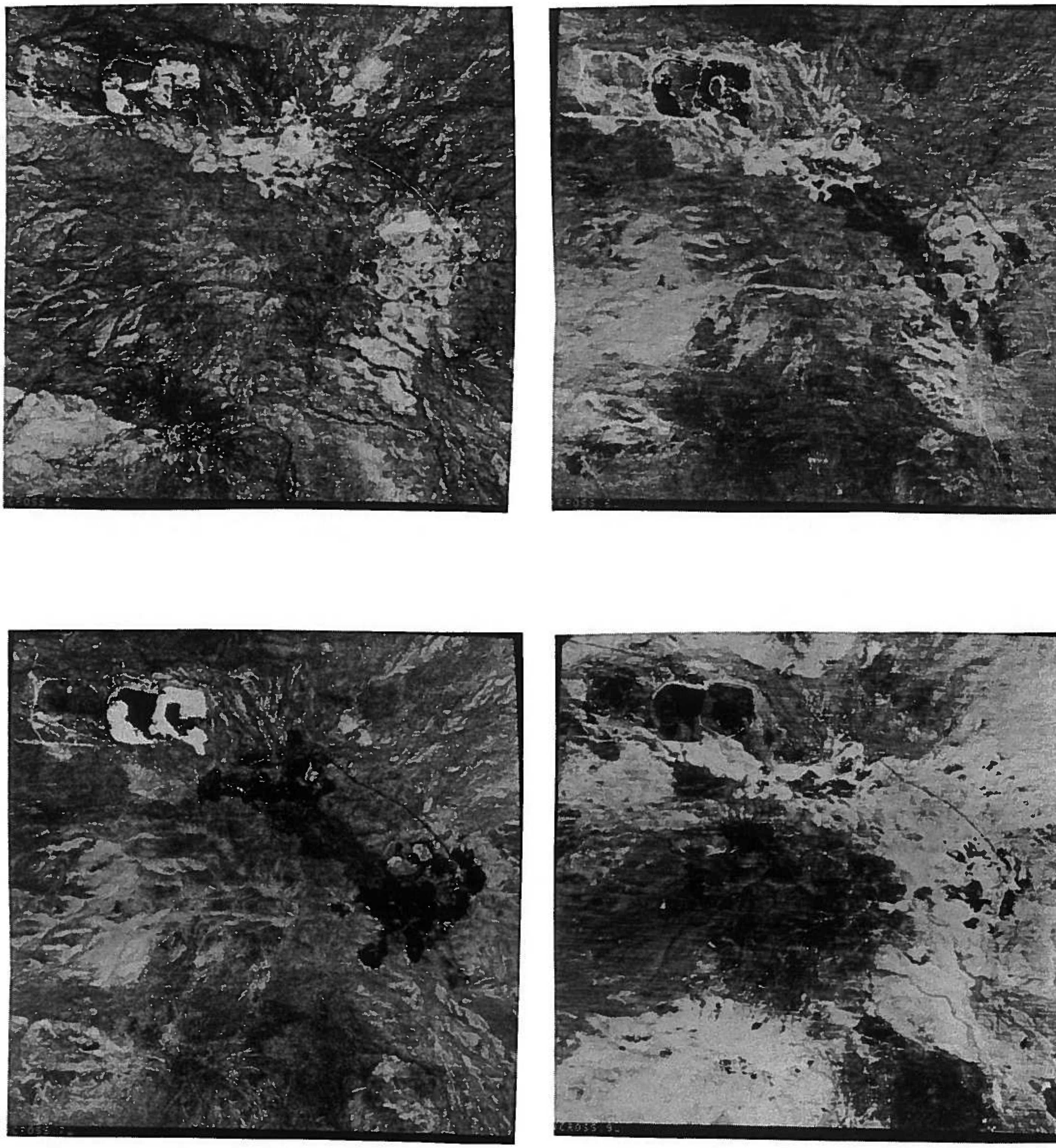


FIGURE 1. Continued