CR 171624

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No. E83-10305
4. Title and Subtitle THE POSSIBLE MODIFICATIONS OF THE HISSE MODEL FOR PURE LANDSAT AGRICULTURAL DATA		5. Report Date FEBRUARY 10, 1981
	"Made available under NASA sponsorship In the interest of early and wide dissemination of Earth Resources Survey Program information and without liability for any use made thereof."	6. Performing Organization Code SR-H1-04037
7. Author(s) CHARLES PETERS		8. Performing Organization Report No 77
		10. Work Unit No.
9. Performing Organization Name and Address UNIVERSITY OF HOUSTON DEPARTMENT OF MATHEMATICS HOUSTON, TX. 77004		11. Contract or Grant No. NAS9-14689
		13. Type of Report and Period Covered TECHNICAL REPORT
12. Sponsoring Agency Name and Address NATIONAL AERONAUTICS AND SPACE ADMINISTRATION LYNDON B. JOHNSON SPACE CENTER HOUSTON, TX. 77058 TASK MONITOR: DALE BROWNE		14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract

(E83-10305) THE POSSIBLE MODIFICATIONS OF THE HISSE MODEL FOR PURE LANDSAT AGRICULTURAL DATA (Houston Univ.) 14 p HC A02/MF A01 CSCL 02C N83-27295 Unclas G3/43 00305

ORIGINAL PAGE IS OF POOR QUALITY

17. Key Words (Suggested by Author(s))	18. Distribution Statement
Maximum likelihood estimation of normal parameters, sufficient statistics, invariance.	

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price*
		13	

*For sale by the National Technical Information Service, Springfield, Virginia 22161

JSC Form 1424 (Rev Nov 75) ATTACHMENT 2 -3- NASA — J

i

Rcd OCT 0 6 1982 pm

POSSIBLE MODIFICATIONS OF THE HISSE MODEL FOR PURE LANDSAT AGRICULTURAL DATA

by

Charles Peters
Department of Mathematics
University of Houston
Houston, Texas

SUMMARY.

This report explores an idea, due to A. Feiveson, for relaxing the assumption of class conditional independence of LANDSAT spectral measurements within the same patch (field). Theoretical arguments are given which show that any significant refinement of the model beyond Feiveson's proposal will not allow the reduction, essential to HISSE, of the pure data to patch summary statistics. A slight alteration of the new model is shown to be a reasonable approximation to the model which describes pure data elements from the same patch as jointly qaussian with a covariance function which exhibits exponential decay with respect to spatial separation.

1. The Basic HISSE Model and its Modifications.

The original mathematical assumptions underlying HISSE are fully described in [7]. Briefly, they are:

a) The sampled pure pixels are organized into p patches (fields) and corresponding to each patch j, there is a set of spectral data measurements $X_j = (X_{j1}, \cdots, X_{jN_j})$, where X_{jk} is the (perhaps multitemporal) vector of spectral data from the kth pixel in the jth patch. For each patch j, there is also an unknown class designation $\theta_j \in \{1,\cdots,m\}$, where m is known.

b) The $\{(X_j,\theta_j)\}_{j=1}^{p}$ are treated as independent random variables. The θ_j have a common unknown discrete distribution Prob $[\theta_j = \ell] = \alpha_\ell > 0$, where $\sum_{\ell=1}^{m}\alpha_\ell = 1$.

c) Given that $\theta_j = \ell$, $X_{j1}, \cdots, X_{jN_j}$ are independently normally distributed with unknown mean μ_ℓ and unknown variance-covariance matrix Ω_ℓ.

A proposed modification due to A. Feiveson [3], introduces one additional matrix parameter for each class. Assumption (c) is changed to

c') Given that $\theta_j = \ell$, $X_{jk} = \mu_\ell + e_j + d_{jk}$, where $E(e_j) = E(d_{jk}) = 0$, var $(e_j) = \Sigma_\ell$, var $(d_{jk}) = \psi_\ell$ and the e_j's and d_{jk}'s are independent normal random variables. Thus the elements $X_{j1}, \cdots, X_{jN_j}$ of X_j are jointly normal with marginal distributions $X_{jk} \sim N(\mu_\ell, \Sigma_\ell + \psi_\ell)$, and constant within-patch covariance $\mathrm{cov}(X_{jk}, X_{ji}) = \Sigma_\ell$, for $k \neq i$.

Notice that the original assumption (c) is a limiting case of (c') obtained by allowing $\Sigma_\ell = 0$.

For reasons discussed later, we will alter (c') to

(c") The constant within-patch covariance for elements of the jth patch is $\text{cov}\,(X_{jk}, X_{ji}) = \frac{1}{N_j}\Sigma_\ell$.

The effect of (c") is that data elements from large patches are considered more weakly correlated than those from small patches. Assumption (c') is perhaps more appropriate if the correlation between pixels of the same patch is really independent of their spatial separation, while (c") is better if the correlation falls off rapidly with spatial separation, on account of the preponderance of spatially distant pairs in larger patches. Calculations are presented in Section 4 to suggest that (c") is a reasonable approximation to the average covariance between pairs when the correlation decreases exponentially with spatial separation. In Section 3 theoretical arguments are given which severely restrict the covariance models for which the patch mean vector and scatter matrix are sufficient statistics without, however, eliminating (c') and (c"). This is an important consideration, since procedures like HISSE are feasible only if the spectral information in patches can be summarized in a small number of statistics.

2. Numerical Procedures for the Alternative Covariance Models.

The likelihood function and iterative procedure for the current version of HISSE are given in [7] and will not be repeated here. For covariance models (c') and (c"). The likelihood functions is

$$L = \sum_{j=1}^{p} \log \sum_{\ell=1}^{m} \alpha_\ell f_\ell(X_j) = \sum_{j=1}^{p} \log f(X_j)$$

ORIGINAL PAGE IS
OF POOR QUALITY

where the model (c')

$$f_\ell(X_j) = |\psi_\ell|^{\frac{-N_j-1}{2}} |\psi_\ell + N_j\Sigma_\ell|^{-\frac{1}{2}} \exp[-\tfrac{1}{2} Q_\ell'(X_j)]$$

and $$Q_\ell'(X_j) = \mathrm{tr}\psi_\ell^{-1}S_j + N_j(m_j-\mu_\ell)^T(\psi_\ell+N_j\Sigma_\ell)^{-1}(m_j-\mu_\ell),$$

while for model (c")

$$f_\ell(X_j) = |\psi_\ell|^{\frac{-N_j-1}{2}} |\psi_\ell + \Sigma_\ell|^{-\frac{1}{2}} \exp[-\tfrac{1}{2} Q_\ell''(X_j)]$$

and $$Q_\ell''(X_j) = \mathrm{tr}\psi_\ell^{-1}S_j + N_j(m_j-\mu_\ell)^T(\psi_\ell+\Sigma_\ell)^{-1}(m_j-\mu_\ell).$$

In both these expressions m_j and S_j are, respectively the patch mean and scatter

$$m_j = \frac{1}{N_j}\sum_{k=1}^{N_j} X_{jk}$$

$$S_j = \sum_{k=1}^{N_j}(X_{jk}-m_j)(X_{jk}-m_j)^T.$$

Thus for both of these covariance models the patch mean and scatter are jointly sufficient.

The unconstrained likelihood equations for model (c") have the form

(1.1) $$\alpha_\ell = \frac{1}{p}\sum_{j=1}^{p} \frac{\alpha_\ell f_\ell(X_j)}{f(X_j)}$$

(1.2) $$\mu_\ell = \sum_{j=1}^{p} N_j \frac{f_\ell(X_j)}{f(X_j)} m_j \Big/ \sum_{j=1}^{p} N_j \frac{f_\ell(X_j)}{f(X_j)}$$

ORIGINAL PAGE IS
OF POOR QUALITY

(1.3) $$\psi_\ell = \sum_{j=1}^{p} \frac{f_\ell(X_j)}{f(X_j)} S_j / \sum_{j=1}^{p} (N_j-1) \frac{f_\ell(X_j)}{f(X_j)}$$

(1.4) $$\Omega_\ell = \sum_{j=1}^{p} \frac{f_\ell(X_j)}{f(X_j)} N_j(m_j-\mu_\ell)(m_j-\mu_\ell)^T / \sum_{j=1}^{p} \frac{f_\ell(X_j)}{f(X_j)}$$

where the new parameter Ω_ℓ is defined as $\Sigma_\ell + \psi_\ell$.

The expressions on the right of equations (1.1) - (1.4) are appealing in that they are averages of quantities whose expectations, given $\theta_j = \ell$, are the parameters on the left. In addition, the successive substitutions scheme suggested by equations (1.1) - (1.4) is a slight variation of the generalized E-M procedure of Dempster, Laird, and Rubin [2]. For covariance model (c'), the likelihood equations do not suggest a natural iterative procedure and it appears that the generalized E-M procedure has no simple formulation.

To be consistent with the original interpretation of the parameter Σ_ℓ as a variance-covariance matrix, it is necessary to maximize the likelihood subject to the additional inequality constraint $\Omega_\ell \geq \psi_\ell$. Since a solution of equations (1.1) - (1.4) need not satisfy this constraint, maximizing the likelihood subject to $\Omega_\ell \geq \psi_\ell$ requires a much more complicated numerical procedure. The condition $\Omega_\ell \geq \psi_\ell$ is equivalent to a set of scalar inequality and nonlinear equality constraints, and numerical procedures for such problems are generally very slow to converge. The unconstrained maximum likelihood procedure is appropriate if as in (c") we merely assume that cov (X_{ji}, X_{jk}) is the same for all i and k, without introducing random variables c_j and d_{jk}.

3. Covariance Models for which patch mean and scatter are sufficient.

Let $X = (X_1|\cdots|X_N)_{n\times N}$ be a matrix whose columns are jointly normally distributed n-vectors. We are interested in characterizing those families of distributions of X for which the statistic (m,S) is sufficient, where $m = X_1 + \cdots + X_N$ and $S = X_1X_1^T + \cdots + X_NX_N^T$. We begin by recalling the following definitions [4, p. 32].

Definition: Let G be a group of homeomorphisms on $\mathbb{R}^n$. A function T defined on $\mathbb{R}^n$ is invariant under G if $T(gx) = T(x)$ for all $x \in \mathbb{R}^n$, $g \in G$. T is a maximal invariant of G if T is invariant and $T(x) = T(y)$ implies that there is a $g \in G$ such that $y = gx$. A measure λ is invariant under G if $\lambda g = \lambda$ for all $g \in G$, where $\lambda g(E) = \lambda(g(E))$.

Lemma 1: Let elements of $\mathbb{R}^{nN}$ be represented as $x = (x_1|\cdots|x_N)$ and let $e^T = (1,1,\cdots,1)_{1\times N}$. For each $N \times N$ real orthogonal matrix u satisfying $ue = e$, let $g_u(x) = xu$. Then $T(x) = (m,S) = (xe,xx^T)$ is a maximal invariant of the group $G = \{g_u\}$.

Proof: $T(g_ux) = (xue, xu(xu)^T) = (xe, xx^T) = T(x)$. Thus T is invariant. Suppose that $T(x) = T(y)$ so that $xe = ye$ and $xx^T = yy^T$. If $x^{(i)}$ and $y^{(i)}$ denote the ith rows of x and y then $x^{(i)}x^{(j)T} = y^{(i)}y^{(j)T}$ and $x^{(i)}e = y^{(i)}e$ for all i and j. This implies that corresponding rows of x and y have the same Euclidean norm and form the same angle with the vector e^T. In addition, the rows of x describe the same set of angles in $\mathbb{R}^N$ as do the corresponding rows of y. Thus, by carrying out parallel Gram-Schmidt procedures on $\{e^T, x^{(1)}, \cdots, x^{(n)}\}$ and $\{e^T, y^{(1)}, \cdots, y^{(n)}\}$, it is easy to construct an orthogonal matrix u such that $e^Tu = e^T$ and $x^{(i)}u = y^{(i)}$ for each i; that is, such that $y = g_ux$. Therefore T is a maximal invariant.

Example: Any linear function T defined on $\mathbb{R}^n$ is a maximal invariant under the group of translations by elements of the kernel of T. In fact, most of the results in [6] characterizing linear sufficient statistics depend only on this aspect of linearity.

If T is a maximal invariant then any invariant function on $\mathbb{R}^n$ is a function of $T(x)$. Moreover, a function $h \circ T$ on $\mathbb{R}^n$ is a maximal invariant if and only if h is one to one on the range of T. In the theorems which follow we shall require that T be a continuous open mapping, in addition to being a maximal invariant. The following lemma shows that to some extent T may be chosen for convenience, with affecting the property of openness.

Lemma 2: Let V be an open subset of $\mathbb{R}^n$, let G be a group of homeomorphisms from V to V and let T_1 and T_2 be continuous maximal invariants of G defined on V with values in $\mathbb{R}^m$. If T_1 is an open mapping then so is T_2.

Proof: Since T_2 and T_1 are maximal invariants, there is a one to one function $h:T_1(V) \to T_2(V)$ such that $T_2 = hT_1$. Since $h^{-1} = T_1T_2^{-1}$ on $T_2(V)$, T_2 is continuous and T_1 is open, h is continuous. By the Brouwer invariance of domain theorem [8, p. 3] h is an open mapping. Therefore, T_2 is also open.

Theorem 1: Let V. be an open subset of $\mathbb{R}^n$, let $\mathcal{M}$ be a homogeneous collection of finite Borel measures on $\mathbb{R}^n$, and let λ be a fixed element of $\mathcal{M}$. Suppose that $\lambda(V^c) = 0$ and $\lambda(U) > 0$ for each nonempty open subset U of V. Let G be a group of homeomorphisms from V to V such that $\lambda(gB) = 0$ whenever $\lambda(B) = 0$ and $g \in G$. Suppose that f_μ is a continuous representative of $\frac{d\mu}{d\lambda}$ for each $\mu \in \mathcal{M}$ and that $T:V \to \mathbb{R}^m$ is a continuous open maximal invariant of

G. Then T is a sufficient statistic for $\mathcal{M}$ if and only if each f_μ is invariant under G.

Proof: Suppose that T is sufficient. Then for each $\mu \in \mathcal{M}$ there exists a Borel measureable function k_μ such that $k_\mu \cdot T$ is a version of $d\mu/d\lambda$, [1].

Let $\mu \in \mathcal{M}$ and $g \in G$ be fixed. The set

$$U = \{x \in V \mid f_\mu(x) \neq f_\mu(gx)\}$$

is an open subset of $B \cup g^{-1}(B)$, where

$$B = \{x \in V \mid f_\mu(x) \neq k_\mu(T(x))\}.$$

Since $\lambda(B) = 0$, $\lambda(g^{-1}(B)) = 0$ and $\lambda(U) = 0$. Therefore, U is empty and it follows that f_μ is invariant. Conversely, if each f_μ is invariant, then for each $\mu \in \mathcal{M}$ there exists a function h_μ such that $f_\mu = h_\mu \cdot T$. Since f_μ is continuous and T is open, h_μ is continuous on $T(V)$. Therefore, by [1, Corollary 6.1] T is sufficient.

Corollary 1.1: Given the hypotheses of Theorem 1, if λ is invariant then T is sufficient if and only if each $\mu \in \mathcal{M}$ is invariant.

Proof: In general, a density with respect to λ of μg is $f_{\mu g} = (f_\mu \circ g)h$, where h is a version of $d\lambda g/d\lambda$. If λ is invariant, then we can take $h = 1$ to obtain $f_{\mu g} = f_\mu \circ g$ as a unique continuous density of μg, for each μ, g. By Theorem 1, T is sufficient if and only if $f_{\mu g} = f_\mu$, which is equivalent to $\mu g = \mu$.

Suppose that $\mu g \in \mathcal{M}$ for each $\mu \in \mathcal{M}$, $g \in G$ and that θ is an r-dimensional parameterization of $\mathcal{M}$; i.e., a one to one function from $\mathcal{M}$ onto $\Omega = \theta(\mathcal{M}) \subset \mathbb{R}^r$. Then there is a homomorphism $g \to \bar{g}$ from G onto a group $\bar{G}$ of transformations on Ω defined by $\bar{g}(\theta_0) = \theta(\theta^{-1}(\theta_0)g)$. The following corollary is clear.

Corollary 1.2. Given the hypothesis of Theorem 1, if λ is invariant then T is sufficient iff $\bar{G}$ is the trivial group consisting only of the identity mapping on Ω.

To apply these results to the characterization problem at hand, let $X = (X_1|\cdots|X_N)$ be a random $n \times N$ matrix having one of a given family of normal distributions and let $X^{(i)}$ denote the ith row of X. We think of $X_1, \cdots, X_N$ as being the observed random vector, but at various times wish to consider the parameters

$$\mu_i = E(X_i)$$

$$\mu^{(i)} = E(X^{(i)})$$

$$C_{ij} = \text{cov}(X_i, X_j)$$

$$R^{(ij)} = \text{cov}(X^{(i)}, X^{(j)}).$$

For the open set V of Theorem 1, we take the set of regular points of $T(x) = (xe, xx^T)$; that is, the set of points x at which $T'(x)$ is surjective $T'(x)$ is surjective if the matrix $\binom{e^T}{x}$ has rank $n + 1$, which is almost certainly true for any of the probabilities under consideration as soon as $N \geq n + 1$. Clearly any of the mappings g_U of lemma 1 is a homeomorphism from V onto itself and T is a continuous open mapping on V. $\mathcal{M}$ will be the given set of nN-variate normal probability measures. The invariant measure λ of Corollary 1.2 will be that given by $\mu_i = 0$, $C_{ij} = 0$ if $i \neq j$, $C_{ii} = I_{n\times n}$. If λ is not already a member of $\mathcal{M}$, it may be added without affecting the sufficiency of T for $\mathcal{M}$. According to Corollary 1.2, and lemma 1, T is sufficient for $\mathcal{M}$ if and only if

(2.1) $$\mu^{(i)}u = \mu^{(i)}$$

and

(2.2) $$u \; R^{(i,j)}u^T = R^{(i,j)}$$

for all i,j and $u \in U = \{N \times N$ orthogonal matrices u such that $ue = e\}$.

Now, (2.1) holds if and only if each $\mu^{(i)} = \lambda_i e^T$ for some scalar λ_i, which is equivalent to $\mu_1 = \cdots = \mu_N$. In (2.2) U may be replaced by the larger set $U' = \{N \times N$ orthogonal matrices such that $ue = \pm e\}$. Let $P = \frac{1}{N} ee^T$ and $Q = I - P$. Then U' is the set of all orthogonal matrices which commute with P, and (2.2) states that each $R^{(i,j)}$ commutes with each $u \in U'$. Let w be an orthogonal matrix such that

$$wPw^T = \left[\begin{array}{c|c} 1 & 0_{1\times(N-1)} \\ \hline 0_{(N-1)\times 1} & 0_{(N-1)\times(N-1)} \end{array}\right].$$

Then U' is the set of all orthogonal matrices u such that wuw^T commutes with wPw^T and (2.2) holds iff $wR^{(i,j)}w^T$ commutes with wuw^T for each $u \in U'$. Elementary calculations show that wuw^T must be of the form

$$wuw^T = \left[\begin{array}{c|c} 1 & 0 \\ \hline 0 & v \end{array}\right]$$

where v is $(N-1)\times(N-1)$ orthogonal, and that for some scalars $\lambda_1^{(i,j)}$, $\lambda_2^{(i,j)}$,

$$wR^{(i,j)}w^T = \left[\begin{array}{c|c} \lambda_1^{(i,j)} & 0 \\ \hline 0 & \lambda_2^{(i,j)}I \end{array}\right].$$

If follows that (2.2) is true iff each $R^{(i,j)}$ is a linear combination of P and Q. Therefore, (2.2) holds if and only if each $R^{(i,j)}$ has constant diagonal elements and constant off diagonal elements, which may depend on i and j. Thus, there are matrices $A = (a_{ij})$ and $B = b(_{ij})$ such that

$$\operatorname{cov}(X_{ik},X_{j\ell}) = \begin{cases} a_{ij} & \text{if } k = \ell \\ b_{ij} & \text{if } k \neq \ell \end{cases}$$

That is,

$$\operatorname{var}(X_k) = A \qquad \text{for all } k$$

and

$$\operatorname{cov}(X_k,X_\ell) = B \qquad \text{if } k \neq \ell .$$

Consequently, A and B are symmetric and we have established

Theorem 2: Let $X_1, \cdots, X_N$ be jointly normally distributed n-vectors whose joint distribution is a number of a family $\mathcal{M}$. Then the mean and scatter matrix of the X_i's are sufficient for $\mathcal{M}$ if and only if for each member of $\mathcal{M}$, (a) the X_i's are identically distributed, and (b) $\operatorname{cov}(X_i,X_j)$ is independent of i and j.

4. Conclusion:

As we mentioned in Section 1 if one thinks of a patch as an approximation to a field then it is difficult to understand how the within-patch covariance of spectral measurements from a given patch could be constant but dependent on the patch size as in (c"). According to the results of Section 3, there is no more sophisticated covariance model whose parameters can be estimated with optimum efficiency using only the patch means and scatters; however, there may be more realistic covariance models which are well approximated by (c') or (c"). For example, suppose that a patch is rectangular in shape with multidimensional spectral information $\{X_{ij} | i = 1\cdots r;\ j = 1\cdots c\}$ where i and j denote the spatial line and column number of the pixel producing X_{ij}. Suppose further that the correlation of two observations X_{ij} and $X_{k\ell}$ decays exponentially with their spatial separation; that is,

$$\operatorname{cov}(X_{ij}, X_{k\ell}) = \Omega^{\frac{1}{2}} A^{|i-k|} B^{|j-\ell|} \Omega^{\frac{1}{2}},$$

where Ω is their common variance matrix and A and B are symmetric commuting matrices of spectral radius less than 1. Let Σ be the average covariance over all pairs of distinct pixels. Then a simple calculation shows that for large r and s (large patch size) $rs\Sigma$ is nearly $4\Omega^{\frac{1}{2}} A(I-A)^{-1}B(I-B)^{-1}\Omega^{\frac{1}{2}}$, so that Σ is nearly inversely proportional to the patch size, as is required by (c"). If A and B are positive semidefinite, so that $z^T X_{ij}$ is always positively correlated with $z^T X_{k\ell}$ for any z, then the expression just given is an upper bound for the average within-patch covariance for any patch size. Therefore, the effect of approximating the exponential covariance model with the constant covariance model (c") may be predictable, and not serious.

References

1. Bahadur, R.R. (1954); Sufficiency and statistical decision functions. Ann. Math. Statist. 25, 423-463.

2. Dempster, A.P., Laird, N.M., and Rubin, D.B. (1977); Maximum likelihood from incomplete data via the EM algorithm. J. Roy. Statist. Soc. Ser. B 39, 1-38.

3. Feiveson, A.; private communication.

4. Giri, N.C. (1977); Multivariate Statistical Inference, Academic Press, New York.

5. Hall, W.J., Wijsman, R.A., and Ghosh, J.K. (1965); The relationship between sufficiency and invariance with application in sequential analysis, Ann. Math. Statist. 36, 575-614.

6. Peters, B.C., Decell, H.P., and Redner, R.A. (1978); Characterizations of linear sufficient statistics. Sankhya, Ser. A, 40, 303-309.

7. Peters, B.C., and Kampe, F. (1980); Numerical trials of HISSE. Report No. 75, Department of Mathematics, University of Houston, Contract NAS9-14689, August 1980.

8. Spivak, M. (1979); Differential Geometry, Vol. I, 2nd Ed., Publish or Perish Inc., Berkeley, Ca.