Reprinted from

Symposium on Machine Processing of Remotely Sensed Data

June 21 - 23, 1977

The Laboratory for Applications of
Remote Sensing

Purdue University
West Lafayette
Indiana

IEEE Catalog No.
77CH1218-7 MPRSD

THE MAXIMUM LIKELIHOOD ESTIMATION OF SIGNATURE TRANSFORMATION (MLEST) ALGORITHM

S. G. THADANI
Lockheed Electronics Company, Inc.

I. INTRODUCTION

The original concept of the Large Area Crop Inventory Experiment (LACIE) called for the extensive use of signature extension (i.e., the ability to use statistics "learned" from a given LACIE training segment to classify data from one or more LACIE recognition segments located in the same general crop growing region). The signature extension effort was generally unsuccessful because of the existence of significant differences between the training and recognition segment wheat/nonwheat signatures. These differences are caused primarily by atmospheric factors such as differences in Sun elevation and haze levels over the training and recognition segments and by target-related factors such as differences in soil moisture levels and soil colors between the training and recognition segments.

It is well known[1] that the atmospheric effects mentioned above can be modeled by a positive definite diagonal affine transformation operating on the training segment signatures. However, no suitable model exists for the target-related factors. This has led to the partitioning-signature-correction approach to signature extension (i.e., the grouping of training and recognition segment pairs in order to minimize the effect of target-related factors, followed by the estimation of affine transformations for the partitioned pairs).

Various techniques have been proposed recently to estimate the optimal affine transformation with which to transform the training segment statistics before classifying the recognition segment.[1,2,3] These techniques fall into two broad categories: The first consists of techniques that use physical models for haze level and Sun angle effects to estimate the affine transformation; the second consists of techniques that attempt to match clusters from the training segment with corresponding recognition segment clusters. The matched pairs of clusters are then used to obtain an optimal affine transformation.

The material for this paper was developed under NASA Contract NAS 9-15200 and prepared for the Earth Observations Division, National Aeronautics and Space Administration, Lyndon B. Johnson Space Center, Houston, Texas.

The Maximum Likelihood Estimation of Signature Transformation (MLEST) algorithm is a method of obtaining maximum likelihood estimates (MLE) of the affine transformation. The technique allows the computation of MLE estimates for the recognition segment wheat/nonwheat *a priori* probabilities; further, the technique can easily be extended to allow the estimation of completely general nondiagonal affine transformations which possibly could model both atmospheric and target-related effects.

II. MATHEMATICAL DESCRIPTION

A. NOTATION

The following notation is used in the mathematical description of the MLEST algorithm.

$\{x\}$ = set of samples from the training segment.

$\{y\}$ = set of samples from the recognition segment.

M = number of subclasses in the training segment.

p = dimensionality of samples.

μ_i = mean vector for training segment subclass i.

Σ_i = covariance matrices for training segment subclass i.

q_i = *a priori* probabilities of training segment subclass i.

$\hat{q}_i$ = *a priori* probabilities of recognition segment subclass i.

B. ASSUMPTIONS

The MLEST algorithm is based on the following major assumptions:

1. The training and recognition segment samples are drawn from probability density functions that are mixtures of normally distributed subclasses.

2. The number of subclasses in the training segment is equal to the number of subclasses in the recognition segment. The training segment subclasses that do not exist in the recognition segment may be represented in the model by zero *a priori* probabilities.

3. The training segment subclass statistics (i.e., means and covariances) are related to the recognition segment subclass statistics by a positive definite affine transformation.

C. MATHEMATICAL DEVELOPMENT

Let $p_T(x/i)$, $i = 1,2,\cdots,M$, represent the class conditional probability density functions for the training segment subclasses. Since the training segment subclasses are assumed to be normally distributed,

$$p_T(x/i) = \frac{1}{(2\pi)^{P/2}|\Sigma_i|^{1/2}} e^{-1/2(x-\mu_i)^T\Sigma_i^{-1}(x-\mu_i)} \quad (1)$$

The overall mixture density function for the training segment is given by

$$p_T(x) = \sum_{i=1}^{M} q_i p_T(x/i) \quad (2)$$

By assumption 3, the training segment subclass statistics are related to the recognition segment subclass statistics by a positive definite affine transformation. This transformation may be represented by the (p×p) real positive definite matrix A and the (p×1) real vector B. It follows that the recognition segment subclass statistics (means and covariance matrices) are given by

$$\mu_i^* = A\mu_i + B \quad (3)$$

and $$\Sigma_i^* = A\Sigma_i A^T \quad ; \quad i = 1,2,\cdots,M \quad (4)$$

From equations (1), (2), and (3), it follows that the mixture density function for samples from the recognition segment is given by

$$\hat{p}_R(y) = \sum_{i=1}^{M} \hat{q}_i \hat{p}_R(y/i) \quad (5)$$

where

$$\hat{p}_R(y/i) = \frac{1}{(2\pi)^{P/2}|\Sigma_i^*|^{1/2}} e^{-1/2(y-\mu_i^*)^T(\Sigma_i^*)^{-1}(y-\mu_i^*)} \quad (6)$$

and $i = 1,2,\cdots,M$.

Next, suppose that one picks N statistically independent samples $y_1, y_2, \cdots, y_N$. from the recognition segment. Then the likelihood function is given by

$$\ell(y_1, y_2, \cdots, y_N) = \prod_{k=1}^{N} \hat{p}_R(y_k) \quad (7)$$

The algebra is simplified considerably if one uses the logarithm of the likelihood function

$$L = \log_e \ell = \sum_{k=1}^{N} \log_e \hat{p}_R(y_k) \quad (8)$$

It may be shown that the partial derivatives of L with respect to the matrix A, the vector B, and the *a priori* probabilities $\hat{q}_i$ are given respectively by

$$\frac{\partial L}{\partial A} = \left[\sum_{k=1}^{N} \sum_{i=1}^{M} \hat{p}_R(i/y_k)(\Sigma_i^*)^{-1}(y_k-\mu_i^*)(y_k-B)^T - (N)I_p\right](A^{-1})^T \quad (9)$$

$$\frac{\partial L}{\partial B} = \sum_{k=1}^{N} \sum_{i=1}^{M} \hat{p}_R(i/y_k)(\Sigma_i^*)^{-1}(y_k-\mu_i^*) \quad (10)$$

$$\frac{\partial L}{\partial \hat{q}_i} = \sum_{k=1}^{N} \frac{\hat{p}_R(y_k/i)}{\hat{p}_R(y_k)} \quad (11)$$

where $i = 1,2,\cdots,M$, I_p is the (p×p) identity matrix, and

$$\hat{p}_R(i/y_k) = \frac{\hat{q}_i \hat{p}_R(y_k/i)}{\hat{p}_R(y_k)} \quad (12)$$

The general MLEST algorithm obtains estimates of the (p×p) matrix A, the (p×1) vector B, and the *a priori* probabilities $\hat{q}_i$, $i = 1,2,\cdots,M$, that maximize the logarithmic likelihood function L. Estimates obtained in this manner are called MLE estimates.

In practice, the optimization indicated above is carried out by using the Davidon-Fletcher-Powell (DFP) constrained optimization program.[4] The DFP program uses equation (8) for the likelihood function and equations (9) through (11) for its partial derivatives to modify A, B, and $\hat{q}_i$, $i = 1,2,\cdots,M$, in such a manner that L is maximized. A useful feature of the DFP program is that it permits the optimization to be carried out subject to various user-input constraints. In general, these constraints are continuous differentiable functions of the parameters A, B, and $\hat{q}_i$, $i = 1,2,\cdots,M$. As an example of the use of constraints, the transformed means μ_i^*, $i = 1,2,\cdots,M$, may be restricted to be in a slab of thickness 't' that encloses the Kauth plane. Other constraints on the affine transformation may be dictated by atmospheric models.[3]

D. DISCUSSION

Experience indicates that stable convergent iterations may not be obtained if all three sets of the parameters, i.e., A, B, and $\{\hat{q}_i\}$, are iterated on simultaneously. The most stable iteration sequence appears to be as follows:

1. Iterate on the B-vector with A and $\{\hat{q}_i\}$ held constant.

2. Iterate on A and B with $\{\hat{q}_i\}$ held constant.

3. Iterate on A, B, and $\{\hat{q}_i\}$ simultaneously.

In some cases it may be desirable to replace step 3 with an iteration on $\{\hat{q}_i\}$ only with A and B held constant.

There are two approaches by which the $\{\hat{q}_i\}$ iteration may be accomplished. In the first approach, the DFP algorithm could be used to maximize L using equations (8) and (11), subject to the constraints given in equations (13) and (14) below.

$$\hat{q}_i \geq 0 \quad ; \quad i = 1,2,\cdots,M \tag{13}$$

$$\sum_{i=1}^{M} \hat{q}_i = 1 \tag{14}$$

In the second approach, the MLE estimates for the *a priori* probabilities may be obtained by a successive substitution procedure. The successive substitution equations are given below.

$$\hat{q}_i^{(\ell+1)} = \frac{1}{N}\sum_{k=1}^{N} \frac{\hat{q}_i^{(\ell)} \hat{p}_R(y_k/i)}{\hat{p}_R(y_k)} \quad ; \quad i = 1,2,\cdots,M \tag{15}$$

In the above, $\hat{q}_i^{(\ell+1)}$ denotes the value of $\hat{q}_i$ at the $(\ell+1)$th iteration. Equation (15) may be derived using the Lagrange multiplier technique to maximize L with respect to $\hat{q}_i$, subject to the constraints given by equations (13) and (14). Equation (15) is derived in appendix A.

The relative merits of the two approaches outlined above are not presently known. However, the successive substitution scheme [eq. (15)] has good convergence properties and does appear to be simpler to apply than the first approach.

III. THE MLEST PROGRAM

The MLEST program is written in Fortran V and is available on the Univac Exec 8 system at the Lyndon B. Johnson Space Center. The program may be executed either in batch mode or in demand mode. The program occupies approximately 16K words of core storage and is equipped to handle a maximum of 20 subclasses and 20 constraints for 4-channel data. The present version of the MLEST program does not include provisions for iterating on the *a priori* probabilities.

The flow chart for MLEST is illustrated in figure 1. The MLEST program consists of three major sections: the DRIVER program, the DAVIDON program, and the LOGLIK program.

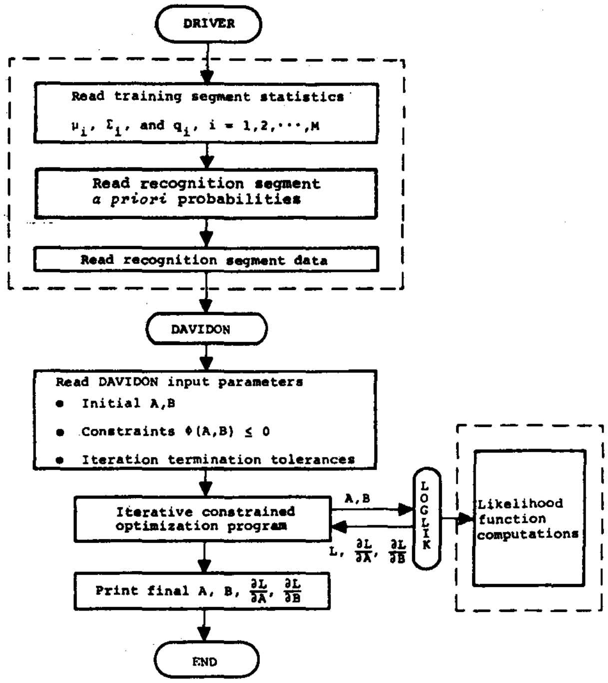

Figure 1. Flow Chart for the MLEST Program.

The DRIVER program reads the training segment statistics and the recognition segment initial *a priori* probabilities. In addition, the DRIVER reads the recognition segment data into core. The DAVIDON program performs the maximization of the logarithmic likelihood function with respect to A and B. A detailed flow chart of the DAVIDON program may be found in reference 2. The LOGLIK program computes the likelihood function and its derivatives as given in equations (8), (9), and (10).

IV. EVALUATION OF THE MLEST ALGORITHM

The MLEST algorithm was evaluated on three separate data sets. These were the simulated data set, the consecutive-day data set, and the geographical extension data set. The simulated data set consisted of four simulated training and recognition segment pairs, with the statistics for each pair related by a known affine transformation. The consecutive-day data set consisted of LACIE data gathered from seven Landsat consecutive-day acquisitions over three intensive test sites in Kansas. The consecutive-day data set served to eliminate the effects of target-related factors on differences between the training and recognition segment signatures. The geographical extension data set consisted of 27 LACIE training and recognition segment pairs from 1974-75 Kansas data, with approximately seven segment pairs per biowindow.

The evaluation procedure for all three data sets consisted of two major steps. In step 1, MLEST signature extension runs were made for each segment pair to determine MLE estimates for each A-matrix and B-vector. In all of these runs, the individual subclass *a priori* probabilities were assumed equal and held constant. The Davidon iteration was initialized with A being the identity matrix and B being the null vector. Also, A was restricted to be diagonal in all runs. In step 2, the affine transformed training segment signatures were used to classify each recognition segment (using the LACIE maximum likelihood classifier). Classification accuracies were computed for wheat/nonwheat over recognition segment training fields. Overall classification accuracies were computed for each recognition segment using the formula

$$P \text{ overall} = 0.5(P_{CW})+0.5(P_{CNW}) \tag{16}$$

where

P_{CW} = wheat classification accuracy.

P_{CNW} = nonwheat classification accuracy.

The classification results were used to estimate wheat proportions at threshold values of T = 0 and 1 percent, respectively. The classification runs described above were repeated using untransformed training segment training field statistics and recognition segment training field statistics. Henceforth, the affine transformed classification results will be referred to as MLEST results; the untransformed training segment classification results will be referred to as UT results; and the recognition segment classification results will be referred to as LOCAL results.

The simulated data set and consecutive-day data set test results have been reported elsewhere[5,6] and will not be detailed here. In essence, the MLEST algorithm successfully estimated the predetermined affine transformations for each of the four training and recognition segment pairs in the simulated data set. Also, the MLEST algorithm significantly improved upon UT classification accuracies and UT wheat proportion estimates on the consecutive-day data set.

V. GEOGRAPHICAL EXTENSION RESULTS

The MLEST program converged normally for 23 out of the 27 signature extension runs attempted. However, successful optimization iteration sequences could not be established for four segment pairs. Analysis of the data for these four segment pairs revealed that the recognition segment data were located relatively far from the modes of the corresponding initial estimates ($A = I_4$, $B = O_4$) for the training segment mixture density functions in spectral space. This resulted in floating-point underflow problems in the likelihood function computations, which in turn caused the Davidon optimization iterations to abort. The MLEST program was rerun for these four segment pairs using the following initial values for the affine transformation

$$A = I_4 \tag{17}$$

$$B = \begin{pmatrix} \mu_{T_1} - \mu_{R_1} \\ \vdots \\ \mu_{T_p} - \mu_{R_p} \end{pmatrix} \tag{18}$$

where

μ_{T_i} = mean value in channel i for the training segment.

μ_{R_i} = mean value in channel i for the recognition segment.

In other words, a mean level adjustment (MLA) was used for the initial B-vector. The reruns were successful, resulting in normal convergence for all four segment pairs.

Table 1 ennumerates the classification accuracy results obtained with the geographical extension data set.

Table 2 lists, by biowindow, the average improvement in MLEST classification accuracy over UT accuracy and the average slack between MLEST classification accuracy and LOCAL classification accuracy. The average improvement and average slack are defined below:

$$\text{Average improvement} = \text{Avg}(P_{MLEST}-P_{UT})\% \tag{19}$$

$$\text{Average slack} = \text{Avg}(P_{LOCAL}-P_{MLEST})\% \tag{20}$$

where

P_{MLEST} = MLEST classification accuracy.

P_{UT} = UT classification accuracy.

P_{LOCAL} = LOCAL classification accuracy.

Referring to tables 1 and 2, one can make the following observations.

1. The MLEST classification accuracies improved upon UT classification accuracies for a majority of the signature extension segment pairs. Improvements in overall classification accuracy are indicated for 22 of the 27 segment pairs. Improvements in the wheat/nonwheat classification accuracies are indicated for 17 of the 27 segment pairs.

Table 1. Classification Accuracy Results for the Geographical Extension Data Set.

Segment pair	Wheat accuracy, %			Nonwheat accuracy, %			Overall accuracy, %		
	UT	MLEST	LOCAL	UT	MLEST	LOCAL	UT	MLEST	LOCAL
Biowindow 1									
1854/1025	16.17	53.36	77.98	42.68	85.69	83.95	29.43	69.52	80.97
1031/1025	73.76	91.37	98.82	95.15	98.94	96.47	84.45	95.16	97.64
1176/1170	91.71	93.36	91.04	57.61	62.20	83.73	74.66	77.78	87.39
1889/1033	94.55	94.29	90.65	58.94	56.81	78.17	76.74	75.55	84.41
1169/1033	68.44	79.11	90.96	89.95	82.89	92.44	79.20	81.00	91.70
1168/1173	62.22	50.20	97.46	43.51	73.21	95.33	52.86	61.70	90.40
1174/1033	80.46	90.11	98.75	98.33	97.89	98.18	89.40	94.00	98.47
Biowindow 2									
1882/1881	94.13	95.24	87.30	65.54	70.79	89.60	79.83	83.02	88.45
1864/1025	70.89	73.39	93.07	83.22	81.20	89.71	77.06	77.29	91.39
1882/1887	0	64.29	87.30	19.63	66.00	89.60	9.81	65.15	88.45
1893/1891	53.60	53.60	82.53	68.84	70.55	84.13	61.22	62.07	83.33
1153/1875	74.12	82.43	80.48	62.69	56.02	70.80	68.51	69.22	75.64
1880/1887	.28	23.20	94.20	58.39	91.41	88.40	29.33	57.30	91.30
1178/1180	52.80	54.44	89.95	52.26	63.50	75.12	52.53	58.97	82.54
Biowindow 3									
1854/1852	50.97	47.15	80.87	84.45	85.11	79.90	67.71	66.13	80.38
1877/1875	28.41	41.47	67.97	71.99	87.38	77.92	50.20	64.43	72.95
1880/1875	22.65	74.59	75.14	40.65	74.08	93.32	31.65	74.33	84.23
1163/1165	84.21	76.61	88.30	30.22	67.14	61.46	57.22	71.87	74.88
1178/1165	66.82	75.93	91.82	52.16	52.83	89.82	59.49	64.38	90.82
1172/1181	16.61	42.08	64.04	43.49	45.15	75.90	30.05	43.61	69.97
Biowindow 4									
1859/1861	82.21	83.50	93.83	56.64	69.13	87.99	69.43	76.32	90.91
1032/1861	56.76	69.23	86.49	74.45	73.68	92.24	65.60	71.46	89.37
1031/1027	6.38	6.15	89.48	44.04	46.95	87.38	25.21	26.55	88.43
1892/1885	53.97	52.98	97.02	78.06	76.18	97.53	66.01	64.58	97.28
1883/1884	35.48	62.32	98.71	34.92	47.09	99.47	35.20	54.71	99.09
1888/1879	92.31	89.35	95.46	96.14	95.64	93.77	94.22	92.50	94.62
1176/1177	92.32	94.09	89.34	70.43	64.50	86.89	81.38	79.29	88.11

Table 2. MLEST Classification Performance Versus Biowindow for the Geographical Extension Data Set.

Criterion	Biowindow				Overall average
	1	2	3	4	
Average improvement					
Overall accuracy	9.71	13.53	14.74	4.05	10.35
Wheat accuracy	9.21	14.40	14.69	5.46	10.80
Nonwheat accuracy	10.21	12.70	14.79	2.64	9.91
Average slack					
Overall accuracy	11.75	18.30	14.75	26.06	17.82
Wheat accuracy	13.41	24.05	18.39	27.53	20.94
Nonwheat accuracy	10.09	12.55	11.33	24.59	14.76

2. The average improvement (table 2) in either overall, wheat, or nonwheat classification accuracy is approximately 10 percent. The improvements in classification accuracy are particularly striking for segment pairs 1854/1025, 1882/1887, 1880/1887, 1880/1875, and 1883/1884. The improvements in wheat classification accuracy for these segment pairs range from approximately 23 percent for segment pair 1880/1887 to approximately 64 percent for segment pair 1882/1887.

3. The degradations ($P_{UT}-P_{MLEST}$) in classification accuracy resulting from the use of MLEST are relatively insignificant. The average degradation (five segment pairs) in the overall accuracy is less than 2 percent. The average degradation in wheat classification accuracy (seven segment pairs) is less than 4 percent. The average degradation in nonwheat classification accuracy (nine segment pairs) is less than 3 percent.

4. The improvements in classification performance do appear to depend on the biowindow (table 2) in which the data were collected. Average improvements in classification accuracy are approximately 14 percent for biowindows 2 and 3, approximately 9-1/2 percent for biowindow 1, and approximately 4 percent for biowindow 4. These results are reinforced by the well-known fact that biowindows 2 and 3 provide maximum discrimination between wheat and nonwheat.

5. The MLEST classification accuracies fall short of the LOCAL accuracies. The average slack between the MLEST and LOCAL accuracies is approximately 18 percent for the overall accuracies, approximately 21 percent for wheat accuracies, and approximately 15 percent for nonwheat accuracies (table 2). However, the LOCAL classification accuracies are biased estimates since they were estimated over the same training fields that were used to train the classifier. By allowing approximately 10 percent to account for this bias, the MLEST accuracies would be within 10 percent of the "true" LOCAL accuracies.

6. MLA starting values for the B-vector were used for segment pairs 1882/1887, 1880/1875, 1877/1875, and 1883/1884. Considerable improvements may be noted in MLEST classification performance for these sites. The effect of the MLA starting values was to place the initial mixture density function in the general neighborhood of the recognition segment data. It is conjectured that the use of MLA starting values for the remainder of the signature extension data set would have resulted in better MLEST classification performance.

Table 3 lists the UT, MLEST, and LOCAL wheat proportion estimates. Table 4 lists mean absolute differences between MLEST wheat proportion estimates and LOCAL wheat proportion estimates and between UT wheat proportion estimates and LOCAL wheat proportion estimates. These mean absolute differences are averaged separately for each biowindow and collectively for the entire data set.

Referring to tables 3 and 4, one can make the following observations.

1. The MLEST proportion estimates are closer to the LOCAL proportion estimates than are the UT estimates in 14 segment pairs with 0 percent thresholding and 11 segment pairs with 1 percent thresholding.

2. The extent of improvement is erratic; however, the MLEST estimates (table 4) are closer, on the average, to the LOCAL estimates than are the UT estimates. The average absolute differences computed for each biowindow between MLEST and LOCAL and between UT and LOCAL indicate that the MLEST proportions represent improvements over UT proportions for biowindows 1, 2, and 3. The MLEST proportions represent degradations with respect to UT proportions for biowindow 4. This is reinforced by the classification accuracy results presented earlier which showed that the smallest improvement in classification accuracy using MLEST was in biowindow 4.

3. The average UT, MLEST, and LOCAL wheat proportion estimates (all 27 sites) at 0 percent thresholding are approximately equal (within 1 percent of each other). The variances of these estimates are also essentially equal. At T = 1 percent, the average MLEST and LOCAL estimates are approximately equal; however, the average UT estimate differs about 5 percent from these estimates.

4. The amount of thresholding with the MLEST classifications is significantly less than that obtained with the UT classifications. Drastic reductions in thresholding are indicated for segment pairs 1854/1025, 1168/1173, 1882/1887, 1880/1887, 1880/1875, and 1172/1181.

VI. CONCLUSIONS

On the basis of tests conducted thus far, the following conclusions can be made:

1. The use of the MLEST algorithm leads to significant improvements in classification accuracy.

2. The MLEST wheat proportion estimates are, on the average, closer to the LOCAL wheat proportion estimates than are the UT wheat proportion estimates.

3. In reference to the geographical extension results, the MLEST algorithm performs best on data from biowindows 1, 2, and 3.

4. The use of the MLEST affine transformed training segment signatures for classification drastically reduces the percentage of pixels thresholded.

These results demonstrate the viability of MLEST as a signature extension algorithm, especially when one considers that the geographical extension data quality was marginal at best. It is the author's view that the use of MLA starting

Table 3. Proportion Estimation Results for the Geographical Extension Data Set.

Segment pair	Wheat proportions, %						Pixels thresholded, %		
	Before thresholding			After thresholding					
	UT	MLEST	LOCAL	UT	MLEST	LOCAL	UT	MLEST	LOCAL
Biowindow 1									
1854/1025	27.0	38.7	47.8	10.1	37.4	47.4	59.5	4.0	0.7
1031/1025	4.7	6.5	10.1	4.4	6.3	9.8	8.8	2.5	1.3
1176/1170	61.0	58.9	44.7	59.1	58.2	44.4	7.4	2.6	1.7
1889/1033	41.6	43.6	32.8	41.3	43.4	32.7	.9	.5	.4
1169/1033	13.0	19.3	15.2	12.9	19.1	14.8	.8	1.3	2.0
1168/1173	21.8	14.1	15.2	15.4	9.8	14.6	23.7	9.6	3.0
1174/1033	38.8	49.4	49.7	37.7	48.9	48.7	1.4	1.0	3.0
Biowindow 2									
1882/1881	56.9	50.7	34.0	54.2	49.0	33.0	4.0	2.7	5.1
1864/1025	29.1	28.2	36.0	23.8	23.5	32.5	8.9	9.6	6.5
1882/1887	50.5	48.1	34.0	.1	47.9	33.0	86.9	1.3	5.1
1893/1891	16.9	19.2	45.1	16.3	18.6	41.8	1.2	.8	3.4
1153/1875	44.7	52.3	45.0	44.5	52.1	44.5	2.5	1.5	1.5
1880/1887	26.0	18.7	19.0	6.0	17.9	18.0	41.4	5.0	8.3
1178/1180	40.3	44.0	29.3	36.6	43.7	29.2	6.5	.8	2.3
Biowindow 3									
1854/1852	33.5	30.3	48.2	31.7	29.5	47.9	3.2	2.1	0.7
1877/1875	29.3	18.0	54.4	27.0	17.3	51.0	4.5	3.4	4.1
1880/1875	33.3	31.9	26.5	24.8	29.7	25.6	39.3	6.1	3.0
1163/1165	65.5	32.3	40.8	65.4	32.2	40.4	1.0	1.4	1.4
1178/1165	43.6	45.2	32.8	43.5	45.1	31.1	.4	.5	2.9
1172/1181	34.0	52.8	53.7	17.8	51.2	42.7	47.6	2.1	1.3
Biowindow 4									
1859/1861	33.6	41.5	31.2	32.5	41.0	30.3	6.8	1.4	2.5
1032/1861	39.9	45.5	39.8	39.4	44.7	39.6	2.3	2.4	.4
1031/1027	10.4	10.9	19.4	9.3	9.2	19.1	6.9	4.4	2.8
1892/1885	34.4	34.6	30.4	32.4	33.2	29.0	3.8	2.8	1.7
1883/1884	48.1	53.0	62.5	46.9	52.0	56.9	2.9	2.5	14.1
1888/1879	52.2	48.1	62.3	50.9	46.4	58.7	5.1	5.7	5.4
1176/1177	47.4	53.6	40.7	46.7	52.8	40.4	2.5	2.3	1.9

Table 4. Mean Wheat Proportion Estimate Differences Versus Biowindow for the Geographical Extension Data Set.

Biowindow	0% thresholding		1% thresholding	
	$\lvert q_L - q_{UT} \rvert$ (a)	$\lvert q_L - q_{MLEST} \rvert$ (a)	$\lvert q_L - q_{UT} \rvert$ (a)	$\lvert q_L - q_{MLEST} \rvert$ (a)
1	10.14	6.17	11.39	6.76
2	13.26	12.4	15.39	12.19
3	16.97	13.58	18.88	13.32
4	6.67	9.33	5.67	8.50
Overall	11.57	10.25	12.61	10.08

vectors, physical constraints on A and B, and the iterative equations for the *a priori* probabilities would lead to even greater improvements in the performance of the MLEST algorithm.

VII. REFERENCES

1. Methods of Extending Signatures and Training Statistics Without Ground Truth Information. Final Report ERIM 109600-16-F, May 1975.

2. Potter, J. F.: The Atmospheric Correction Program (ATCOR). Lockheed Electronics Co., Inc., Technical Memorandum LEC-8645 (Houston), Sept. 1976.

3. Finley D. R.; and Quirein, J. A.: The Optimal Signature Correction Algorithmic Routine (OSCAR) System. Lockheed Electronics Co., Inc., Technical Memorandum LEC-8086 (Houston), Feb. 1976.

4. The Davidon-Fletcher-Powell Penalty Function Method: A Generalized Iterative Technique for Solving Parameter Optimization Problems. NASA IN-75-FM-34, 1975.

5. Thadani, S. G.: A Staus Report on the Maximum Likelihood Estimation of Signature Transformation (MLEST) Algorithm. Lockheed Electronics Co., Inc., Technical Memorandum LEC-8300, Sept. 1976.

6. Potter, J. F.; Thadani, S. G.; et al.: Performance Tests of Signature Extension Algorithms. Presented at the 11th International Symposium on Remote Sensing of Environment (Ann Arbor, Michigan) April 25-29, 1977.

APPENDIX A

DERIVATION OF THE ITERATIVE MLE EQUATIONS FOR THE *A PRIORI* PROBABILITIES

The MLE estimates for the *a priori* probabilities $\hat{q}_i$, $i = 1,2,\cdots,M$, are those estimates that maximize the logarithmic likelihood function L subject to the constraint

$$\sum_{i=1}^{M} \hat{q}_i = 1 \tag{A1}$$

The MLE estimates for $\hat{q}_i = i = 1,2,\cdots,M$, may be determined by using the Lagrange multiplier technique. Define the augmented logarithmic likelihood function

$$\Phi = L+\lambda\left(\sum_{i=1}^{M}\hat{q}_i-1\right) \tag{A2}$$

where λ is the Lagrange multiplier and

$$L = \sum_{k=1}^{N}\log_e \sum_{i=1}^{M}\hat{q}_i\hat{p}_R(y_k/i) \tag{A3}$$

Now the MLE estimates $\hat{q}_i$ must satisfy

$$\frac{\partial\Phi}{\partial\hat{q}_i} = 0 \quad ; \quad i = 1,2,\cdots,M \tag{A4}$$

$$\frac{\partial\Phi}{\partial\lambda} = 0 \tag{A5}$$

Note that $\frac{\partial\Phi}{\partial\lambda} = 0$ implies $\sum_{i=1}^{M}\hat{q}_i = 1$. Also,

$$\frac{\partial\Phi}{\partial\hat{q}_i} = \sum_{k=1}^{N}\frac{1}{\hat{p}_R(y_k)}\hat{p}_R(y_k/i)+\lambda \quad ; \quad i = 1,2,\cdots,M \tag{A6}$$

where

$$\hat{p}_R(y_k) = \sum_{i=1}^{M}\hat{q}_i\hat{p}_R(y_k/i) \tag{A7}$$

Setting $\frac{\partial\Phi}{\partial\hat{q}_i} = 0$, we have

$$\sum_{k=1}^{N}\frac{1}{\hat{p}_R(y_k)}\hat{p}_R(y_k/i)+\lambda = 0 \quad ; \quad i = 1,2,\cdots,M \tag{A8}$$

Multiplying the *ith* equation by $\hat{q}_i$ and adding the resulting M equations yields

$$\left\{\sum_{k=1}^{N}\frac{1}{\hat{p}_R(y_k)}\left[\hat{q}_1\hat{p}_R(y_k/1)+\hat{q}_2\hat{p}_R(y_k/2)+\cdots+\hat{q}_M\hat{p}_R(y_k/M)\right]\right\} +\lambda(\hat{q}_1+\hat{q}_2+\cdots+\hat{q}_M) = 0 \tag{A9}$$

But

$$\sum_{i=1}^{M}\hat{q}_i = 1 \tag{A10}$$

and

$$\hat{p}_R(y_k) = \sum_{i=1}^{M}\hat{q}_i\hat{p}_R(y_k/i) \tag{A11}$$

Hence, we have

$$N+\lambda = 0 \qquad \lambda = -N \tag{A12}$$

Now let

$$\hat{p}_R(i/y_k) = \frac{\hat{q}_i \hat{p}_R(y_k/i)}{\hat{p}_R(y_k)} \quad ; \quad i = 1,2,\cdots,M \quad \text{(A13)}$$

Then

$$\frac{\hat{p}_R(y_k/i)}{\hat{p}_R(y_k)} = \frac{\hat{p}_R(i/y_k)}{\hat{q}_i} \quad ; \quad i = 1,2,\cdots,M \quad \text{(A14)}$$

Substituting equations (A14) into equation (A8) with $\lambda = -N$ yields

$$\sum_{k=1}^{N} \frac{\hat{p}_R(i/y_k)}{\hat{q}_i} - N = 0 \quad ; \quad i = 1,2,\cdots,M \quad \text{(A15)}$$

or

$$\hat{q}_i = \frac{1}{N}\sum_{k=1}^{N} \hat{p}_R(i/y_k) \quad ; \quad i = 1,2,\cdots,M \quad \text{(A16)}$$

or

$$\hat{q}_i = \frac{\hat{q}_i}{N}\sum_{k=1}^{N} \frac{\hat{p}_R(y_k/i)}{\left[\sum_{i=1}^{M} \hat{q}_i \hat{p}_R(y_k/i)\right]} \quad ; \quad i = 1,2,\cdots,M \quad \text{(A17)}$$

which is the required iterative system of equations for $\hat{q}_i$.

Suresh G. Thadani obtained his Masters degree in Electrical Engineering from the University of Houston in June 1970, and he is presently continuing his studies toward a Ph. D. at that university. Mr. Thadani has been employed as a Senior Scientist by Lockheed Electronics Company, Inc., Systems and Services Division, since February 1973 in support of the Earth Observations Division.