DEPARTMENT OF MATHEMATICS
UNIVERSITY OF HOUSTON HOUSTON, TEXAS

(NASA-CR-151236) A REVIEW OF THE LEC PERFORMANCE EVALUATION OF UHMLE (Houston Univ.) 19 p HC A02/MF A01 CSCL 09B

N77-19773

Unclas
G3/61 20572

A REVIEW OF THE LEC PERFORMANCE
EVALUATION OF UHMLE
BY HENRY P. DECELL, JR
AND W. A. COBERLY
REPORT #62 JANUARY 1977

PREPARED FOR
EARTH OBSERVATION DIVISION, JSC
UNDER
CONTRACT NAS-9-15000

HOUSTON, TEXAS 77004

A REVIEW OF THE LEC PERFORMANCE EVALUATION OF UHMLE

Henry P. Decell, Jr.
Department of Mathematics
University of Houston
Houston, Texas

&

William A. Coberly
University of Tulsa
Consultant to
Department of Mathematics
University of Houston
Houston, Texas

Report #62
NAS-9-15000

January 1977

A Review of the LEC Performance Evaluation of UHMLE

In March 1976, Lockheed was directed to submit a plan [1] for comparative evaluation of several candidate signature extensions algorithms. The results of that test [2], carried out by LEC in April, were the basis for selection of two algorithms [3], OSCAR and ATCOR, for test and implementation in a sub-operational system by IBM. Four simulated (SIM) data sets and seven consecutive day (CD) data sets were used. In the following sections, two points will be addressed for each data set. 1) Analysis and evaluation of the UHMLE test. 2) Recommendations on changes in the UHMLE algorithm motivated by the test. The criterion for evaluation of each algorithm will be overall classification accuracy (Tables 8 and 9 of [2] are attached for convenience).

I. Simulated Data Test.

In previous tests carried out by the University of Houston consistently good results were observed using essentially the same data set. The poor performance of UHMLE on SIM1 and the marginal performance on SIM4 seems to contradict our previous experience. The following observation on the LEC test may explain this discrepency.

In SIM1 the iteration sequence seemed to converge before the signatures had moved into the unlabeled data region. A second run which first estimated an initial translation X + B and then applied the general UHMLE algorithm was successful. Even though translation was included in our operational algorithm delivered to JSC, the second run was not reported in the final LEC analysis.

/

Pass	Local Accuracy	1st LEC UHMLE TEST	2nd LEC UHMLE TEST w/translation option
SIM1	93.5	-21.7	-2.5
SIM2	98.6	-0.7	no trans.
SIM3	97.0	-1.0	" "
SIM4	92.8	-5.0	" "
Ave.	95.5	-7.1	-2.3
Std.		9.9	2.0

Table 1

Revised SIM test results.
Overall Accuracy Difference

The use of the translation in SIM1 would dramatically change the outlook of UHMLE in the SIM test.

The results do not suggest any modifications of the UHMLE algorithm except to re-state the need to apply the translation first.

II. Consecutive Day Test.

General: The consecutive day (CD) data set consisted of three Kansas Intensive Test Sites (ITS) outlined in [1]. From these a total of seven pairs of consecutive day passes were selected from 1973-74 LANDSAT-1 data acquisitions.

ITS	DATA SET ID	DATE TRAINING/RECOGNITION	SIZE ITS	HAZE TRAINING	RECOGNITION
Finney	F1709-8	2/1 July 74	5 × 6		
"	F1673-2	27/26 May 74	"	X	
"	F1655-4	9/8 May 74	"		
"	F1726-7	19/20 July 74	"	X	
Saline	S1455-4	21/20 Oct 73	3 × 3		
"	S1725-4	18/17 July 74	"		X
Ellis	E1726-5	12/11 June 74	3 × 3		X

Table 2

Consecutive Day Data Sets

Two UHMLE tests were run on each data set. UH/ALL uses as its unlabeled sample the rectangular area containing the selected Test/Training fields. UH/FIELDS uses the test fields only as input. The following ground areas associated with each ITS are defined for further reference.

A0 - ITS ground truth site. (Not alligned with LANDSAT ground track.)

A1 - Smallest rectangular field containing selected training field. Used as input for UH/ALL.

A2 - A0 intersect A1, used for classification area.

A3 - Designated test fields (≡ training fields within A2). Used for input to UH/FIELDS.

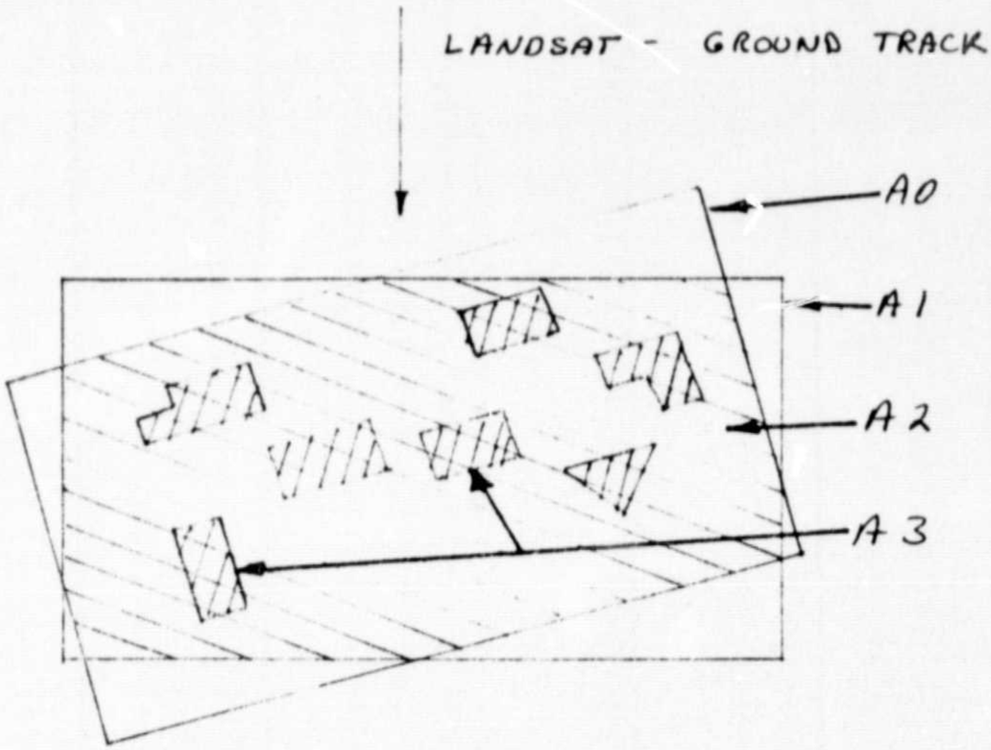

Figure 1

Ground area definitions

Proportion Estimates. UHMLE automatically estimates a proportion vector for the unlabeled input data set. These estimates are used in two ways in the Signature Extention (SE) test.

1) The UHMLE proportion estimates are used as *a priori* probabilities in the classification algorithm. Although this is not an unreasonable choice for the *a priori* probabilities, the UHMLE classification results are not comparable to those of the other candidate algorithms which used equally likely *a priori* probabilities. Moreover, in the UH/ALL test, the UHMLE proportion estimates correspond to Area A1. Area A2 was classified and only results from Area A3 were used for performance evaluation. In UH/FIELDS the unlabeled input data set and the classification region were equivalent.

2) In Tables 10-13 in [2], the estimated proportion of wheat for each algorithm is first compared to the local classification proportion estimate and then to the ground truth proportion estimate for both the SIM and CD data sets. In the CD test, the UH/ALL and UH/FIELDS are classification proportion estimates for area A2. The maximum-likelihood estimates from UHMLE (UH/ALL/MLE) correspond to area A1 . It is assumed here that the proportion estimate from local classification in Table 11 of [2] is based on A2. Hence UH/ALL/MLE is not comparable to the local standard. In Table 13 [2] the standard is ground truth. It is not clear whether or not the ground truth proportions correspond to A0 or A2 . In either case all proportion estimates listed in that table are not comparable.

Data Quality. This appears to be the most important factor in analyzing the UHMLE results. The CD data sets contained numerous data drops or "glitches." LEC was careful to choose training segments and fields so as to avoid this bad data in the computation of training statistics. However, several of the recognition segments used as input to UHMLE (in both UH/ALL and UH/FIELDS) were contaminated. This bad data effectively "captured" subclasses from both wheat and non-wheat categories and distorted means and particularly covariances in other subclasses. Only the data quality in Area A2 could be assessed from the available computer output. Further data drops, which may have been present in A1 (outside of A2), could also have an apparent degrading effect on UH/ALL test results. The implications and incidence of contaminated data is listed below in Table 3. We strongly recommend that this be the last time that this data set be used in any testing procedure.

Data Set	UH/FIELDS	UH/ALL
F 1709-8	Slight	Slight
F 1673-2	Bad	Bad
F 1655-4	Bad	Bad
F 1726-7	Bad	Bad
S 1455-4	Slight	Slight
S 1725-4	Good	Good
E 1726-5	Good	Good

Table 3

Incidence of Data Drops in CD Data Sets

Label Switching: In the UHMLE algorithm the various subclass statistics move in a quasi-independent manner to better "fit" the unlabeled data set. In this process a subclass component of the mixture model may seek out data in the unlabeled sample which is from a different category than the one assigned in the training segment. This poses no difficulty in terms of density estimation, however correct category labels are required for acreage proportion estimates. This phenomena is compounded by subclasses being "captured" by data drops, leaving unmodeled data free to be absorbed by an existing subclass. In a number of the CD tests substantially improved results are obtained if the label on a single subclass is reassigned. Interaction of the AI or DPA (at this point, prior to aggregation of acreage proportion estimates at the category level) with the view of detecting obvious category labeling errors, should be considered. This is a key point. We are simply saying that, when using UHMLE (or other algorithms), the spectral class identity extrapolated from the training segment may not be sufficient to establish crop category identity without AI interaction.

Individual CD Data Set Results. In this section each CD-data-set test is analyzed separately. Some revised results are reported along with supporting rationals.

F 1709-8 Two classes have inflated variances due to a data drop. However, both UH/ALL and UH/FIELDS do better than local classification.

F 1673-2 Very poor performance on both cases is observed. Two data drops have major effect on distorting variances and means on several subclasses. If one subclass, which is obviously mislabeled, is switched from wheat to non-wheat a substantial improvement is observed.

		LEC Test		Revised	
Local	UT	UH/FIELDS	UH/ALL	UH/FIELDS	UH/ALL
96.1	0.1	-23.7	-21.3	-3.1	-8.6

In Figure 2, the subclass means determined by UHMLE are plotted in the TACAP "brightness × green" coordinate system. Subclass W7 is clearly displaced from the other wheat subclasses. It is not unreasonable for mislabeling of this magnitude to be easily detected by an AI or DPA and corrected at the time of acreage estimation.

/

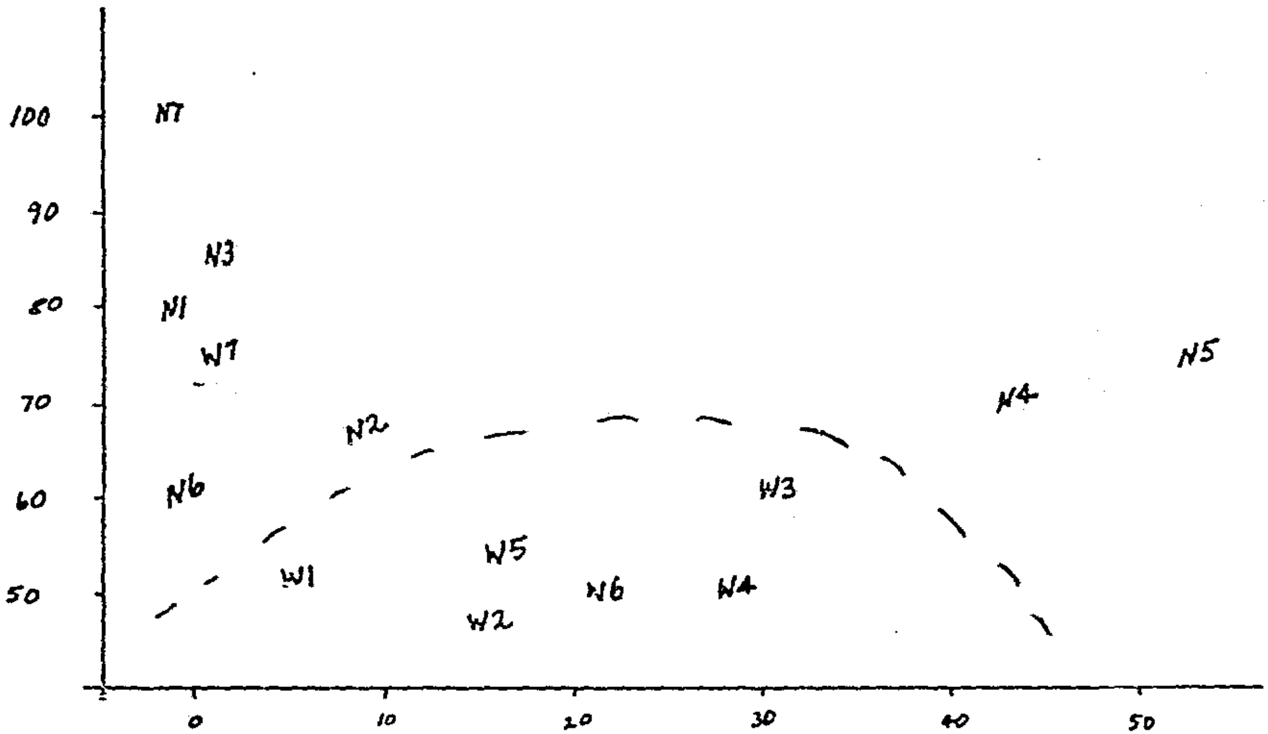

Figure 2. TACAP plot of class means.

FINNEY 1672 / UHMLE-FIELDS

F 1655-4 Again two data drops play a large role in distorting several subclass signatures in UH/ALL. One label switch again improves matters greatly. In UH/FIELDS the effects of

				Revised	
Local	UT	UH/FIELDS	UH/ALL	UH/FIELDS	UH/ALL
94.9	-3.8	-3.1	-15.0	not revised	-3.3

the data drops are not as apparent in the overall classification accuracy.

F 1726-7 Data drops substantially distort four subclasses in UH/ALL and to a lesser extent in UH/FIELDS. Even so, results are excellent (better than local classification) in UH/FIELDS. UH/ALL results are poor. No clear label switch is apparent.

S 1455-4 In this data set only four subclasses are modeled. Two subclasses are distorted by data drops, one severely in both cases. In the UH/ALL case the A1 area is much too large, introducing a large segment of extraneous data into the unlabeled sample. Further A2 is not contained in A1 (see Figure 3).

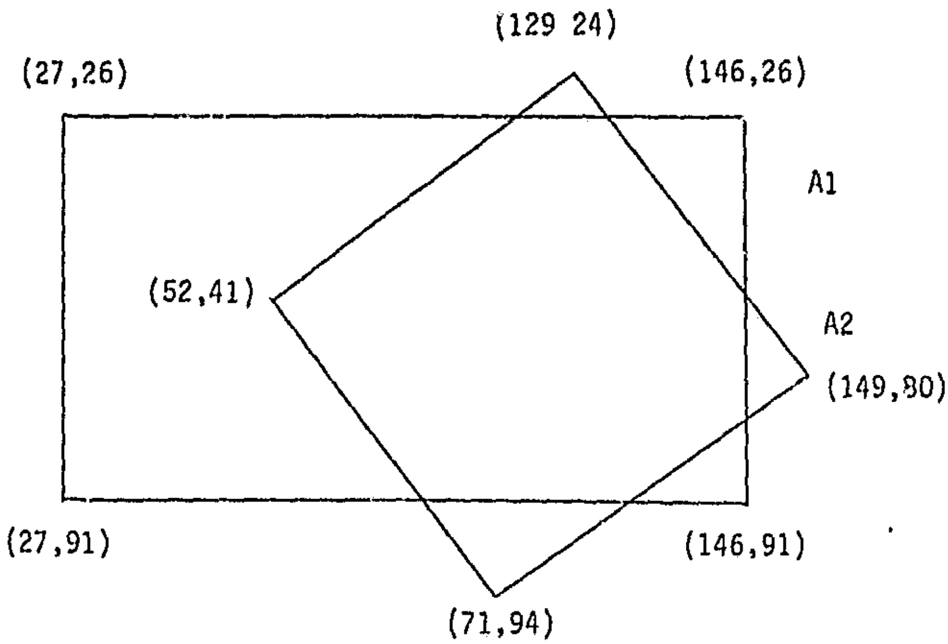

Figure 3.

Field Definition Errors in S 1455-4.

The poor data quality, errors in field definitions, and small number of subclasses render the interpretation of this test null and void. Inclusion of this test in the overall UHMLE evaluations is, therefore, meaningless.

<u>S 1725-4</u> There are no data drops or anomolies in this test.

<u>E 1726-5</u> There are no data drops. A reasonable case could be made for a label switch, however, the explanation is not as obvious as in the previous data sets and it will be omitted here. This case appears to be a reasonable test of the algorithm.

Summary of CD Test. If we introduce the three label changes (easily detected by an AI or DPA) suggested in F 1673-2 and F 1655-4 and omit the unacceptable test of S 1455-4, the performance of the algorithm is distinctly different than that reported in [2]. In light of the results presented here, the conclusions drawn by LEC in [2] concerning the relative performance of UHMLE are, at best, questionable. The original results along with the aforementioned revision and omission are listed in Table 4 below.

		LEC Original		Revised	
Data Set	Local	UH/FIELDS	UH/ALL	UH/FIELDS	UH/ALL
F 1709-8	79.5	2.7	7.3	same	same
F 1673-2	96.1	-21.3	-23.7	-3.1	-8.6
F 1655-4	94.9	-3.1	-15.0	same	-3.3
F 1726-7	80.0	0.9	-6.8	same	same
S 1455-4	86.5	-12.1	-29.5	OMIT	OMIT
S 1725-4	85.4	-4.3	0.9	same	same
E 1726-5	66.2	1.4	-7.3	same	same
Mean		-5.1	-10.6	-0.92	-2.97
Std. Dev.		8.7	13.1	2.9	6.1

Table 4.

Revised UHMLE Test Results.
Overall Classification Accuracy Differences.

We maintain that there is considerable evidence (provided, in part, by this analysis) for rejecting the original analysis and conclusions. If for no other reason, the poor data quality in five of the seven CD data sets chosen renders the LEC test results, as they pertain to UHMLE, invalid.

III. Conclusions.

Although the LANDSAT-2 data does not contain nearly the frequency of data drops observed in the LANDSAT-1 data used for this test, we clearly must incorporate a data editing scheme into the UHMLE algorithm or assume that preprocessing has deleted these pixels. There has been preliminary testing of a thresholding scheme which appears to be an adequate method when used in conjunction with an initial X + B translation.

The reassessment of labels after signature extension remains a major priority in the UHMLE signature extension algorithm. This is a small task in terms of time compared to complete local training by the AI, and appears to be a necessary AI interaction function coupled with automatic processing of recognition segments.

SUMMARY

Our comments on the SD test and on the CD test suggest that the UHMLE algorithm in particular and mixture density estimation in general should still play an important role in the solution of the signature extension problem. In another paper [4], the signature (e.g., Procedure 1) extension problem, in the context of the LACIE training procedure is reformulated. Mixture density estimation (supervised or unsupervised) will certainly play a role in the exaction of the Spectral Information Classes described in that paper. Additional work on the UHMLE algorithm, especially the details of incorporating it into the LACIE training procedure, we believe to be essential. These details are treated in the reformulation given in [4].

REFERENCES

1. Plan for Evaluating Several Signature Extension Algorithms, LEC Memorandum, April 1, 1976 Ref: 642-1877.

2. Performance Tests of Signature Extension Algorithms, LEC Ref: 642-2018 September 1976.

3. Selection of Signature Correction Algorithms for Implementation and Test by IBM. EOD Memorandum TF3/K. Baker: db: 4/26/76: 2071 April 30, 1976.

4. Henry P. Decell, Jr. and W. A. Coberly, On Signature Extension, Mathematics Department, University of Houston, December 1976 (in printing).

TABLE 8.– OVERALL ACCURACY FOR SIMULATED DATA*

[A minus sign means the algorithm was less accurate than local classification.]

Data	Local accuracy	Percentage difference between local accuracy and that obtained with various algorithms				
		R(S)	MLEST	UH fields	R(C)	UT
SIM1	93.5	0.0	-3.5	-21.7	-29.6	-99.3
SIM2	98.6	0.0	0.0	-0.7	0.0	-18.3
SIM3	97.0	0.1	0.0	-1.0	-5.2	-50.0
SIM4	92.8	-0.1	-3.2	-5.0	-2.9	-8.8
Mean	95.5	0.0	-1.7	-7.1	-9.4	-44.1
Std. dev.	2.8	0.1	1.9	9.9	13.6	40.8

*Prepared by LEC [2].

TABLE 9.— OVERALL ACCURACY FOR CONSECUTIVE DAY DATA*

[A minus sign means the algorithm was less accurate than local classification.]

Data	Local accuracy	Percentage difference between local accuracy and that obtained with various algorithms											
		R(S)	MLEST	OSCAR	REGRES	MOD R	R(C)	MOD OSCAR	ATCOR	UH fields	UT	R(S/C)	UH all
F1709-8	79.5	-5.8	-4.4	-7.0	-7.1	-7.6	-8.1	-7.8	-8.5	2.7	-8.2	-12.5	7.3
F1673-2	96.1	-2.0	-0.5	-3.2	-10.2	0.5	-1.7	-0.7	-5.0	-21.3	0.1	-1.7	-23.7
F1655-4	94.9	-3.3	-1.8	-2.1	-2.1	-2.7	-4.7	-3.0	-3.6	-3.1	-3.8	-3.8	-15.0
F1726-7	80.0	1.9	1.7	3.8	4.9	-1.9	-1.1	2.4	-5.9	0.9	-8.5	-7.1	-6.8
S1455-4	86.5	-0.2	-0.9	-3.5	-1.8	-3.2	-4.4	-2.5	0.1	-12.1	0.0	-3.5	-29.5
S1725-4	85.4	1.1	-0.5	-0.9	0.0	-3.2	-1.9	-5.0	-4.7	-4.3	-14.1	-11.0	0.9
E1726-5	66.2	-3.2	-6.[illegible]	-3.8	-3.5	-1.8	-4.1	-9.8	-2.7	1.4	-11.5	-9.8	-7.3
Mean	84.1	-1.6	-1.8	-2.4	-2.8	-2.8	-3.7	-3.8	-4.3	-5.1	-6.6	-7.1	-10.6
Std. dev.	10.2	2.7	2.6	3.3	4.9	2.5	2.4	4.2	2.7	8.7	5.5	4.2	13.1

*Prepared by LEC [2] .