DEPARTMENT OF MATHEMATICS
UNIVERSITY OF HOUSTON HOUSTON, TEXAS

USER'S GUIDE UHMLE/RTCC VERSION

MARCH 1976

BY WILLIAM A. COBERLY

REPORT #52

PROGRAM DOCUMENTATION SIGEXT EOD-LARSYS VERSION OF UHMLE

BY WILLIAM A. COBERLY

REPORT #53

PREPARED FOR
EARTH OBSERVATION DIVISION, JSC
UNDER
CONTRACT NAS-9-12777

HOUSTON, TEXAS 77004

USERS GUIDE UHMLE/RTCC VERSION

March 1976

William A. Coberly

University of Tulsa

Consultant to

Department of Mathematics

University of Houston

Report # 52

NAS-9-12777

USERS GUIDE UHMLE/RTCC VERSION

• Step 1 Load Program tape (MLETAP = 013856).

• Step 2 EXEC PGM = ØLDTAP

INPUT

Medium	Variable	Description	Format
CARD	NCHS	No. of channels	
	KCH(I),I=1,...,NCHS	Channel list	(17I2)
CARD	KFILE CK/RT	tape file no.	(I5)
TAPE	CK/RT on file KFILE of Fortran unit 1.		

OUTPUT

DISKFILE — Statistics from CK/RT tape on Fortran unit 12

Mean vector
Covariance (symmetric) — Class 1
⋮
Mean vector
Covariance — Class NCLASS

DISKFILE — NCLASS - No. of classes (I6)

fortran unit 13

• Step 3 EXEC PGM = MLE (UHMLE)

INPUT

CARD	RUNID		(10X,A4)
DISK	M	No. of classes (unit 13)	(I6)
CARD	N	No. of channels	(10X,I/0)
CARD	L	buffer size	"
CARD	FMT(18)	format of temp. data set	(18A4)
CARD	KFILE	file no. of image data	(I5)
CARD	SAMKEY	(=1)	(10X,I10)
CARD	ISTART	start line absolute	"
CARD	ISTØP	stop line	
CARD	ISKIP	line skip (= 1 no skip)	"
CARD	JSTART	start pixel relative	"
CARD	JSTØP	stop pixel	
CARD	JSKIP	pixel skip (= 1 no skip)	"
CARD	NCHØUT	number of channels from tape	"
CARD	NCHLST( )	channel list	(10X,16I2)
CARD	NFLDS	no. of fields input	(10X,I10)
	If NFLDS = 0	skip field def cords and process all of the test site	
CARD	FID,NV,MINLIN,MAXLIN		(A8,2X,3I5)
CARD	IF(NV+1)	line coord. field 1*	(11I5)
CARD	JF(NV+1)	pixel coord.	(11I5)
	. . .		
CARD	FID,NV,MINLIN,MAXLIN		
CARD	IF(NV+1)	field NFLDS	"
CARD	JF(NV+1)		

* first field should be definition of test site all fields with FID '$$$$$$$$' are deleted from the temporary data set.

TAPE	Image data on file	KFILE	Universal
		Unit 1	
CARD	YUNIT	Temp stat unit (= 12)	I5
DISK	Stats on	YUNIT	(8F10.5)
CARD*	IFLAGA		
	IFLAGM		
	IFLAGS		
	MØDES		
	ITLIM		
	TØLA	(One Card)	
	TØLM		
	TØLS		
	EPSA		
	EPSM		
	EPSS		
	ITRPNT		(5I5,6F5.3,I5)

* This card may be repeat as many times as necessary

OUTPUT

DISK	Temporary image data set unit 23	FMT
DISK	MLE output STATS on unit 11	(8F10.5)

• Step 4 EXEC NEWTAP

INPUT

CARD	NCHS,KCH(NCHS)	(17I2)
CARD	KFILE CK/RT tape file no.	(I5)
TAPE	Old CK/RT tape on unit 1	
DISK	Stats from MLE unit 11	(8F10.5)

OUTPUT

TAPE	New CK/RT tape on unit 2	

Program Documentation

SIGEXT

EOD-LARSYS Version of UHMLE

William A. Coberly

University of Tulsa

Consultant to

Department of Mathematics

University of Houston

March 1976

Report # 53

NAS-9-12777

SIGEXT Processor

The SIGEXT processor is a version of the UHMLE program, developed at the University of Houston, which has been integrated into the Univac Exec II LARSYS system. The program uses initial signatures from a training site and unlabeled data from a recognition site and computes maximum likelihood estimates of the mixture distributions of the unlabeled sample. That is, estimates are found for the mixture proportions α_i, the mean vectors μ_i, and the covariance matrices Σ_i, for $i = 1,\ldots,m$. For a theoretical discussion of the algorithms see [1,2,3].

Two options are available:

(1) General form: There is no constraint on the movement of the subclass signatures in the iteration process. That is, there is no a priori transformation formaly assumed.

(2) Affine form: It is assumed that there exists a diagonal matrix A and a vector b defining an affine transformation $Ax + b$. Maximum likelihood estimates for A and b are found and the output signatures are computed as follows:

$$\mu_i' = A\mu_i + b, \quad \Sigma_i' = A^T\Sigma_i A, \quad i = 1,2,\ldots,m$$

For a more detailed engineering description and program documentation see [4].

The following preliminary user's guide is intended to conform to the format of [5]. Hopefully, any unresolved terms, formats, or references will be found there.

References

[1] B. Peters & H. Walker, "An Iterative Procedure for Obtaining Maximum-likelihood estimates of the parameters for a mixture of normal distributions" Report #43 University of Houston.

[2] T. McCabe & J. Solomon, "An Iterative Scheme for Computing an Affine Transformation for Signature Extension" Technical report EOD.

[3] Minter, T.S. et. al. MLEST report & documentation, LEC Technical report # .

[4] W. Coberly & L. Wiginton, UHMLE Program Description", Report #48 University of Houston.

[5] User Documentation: EØD-LARSYS LEC Document # Sept. 1975.

1. Input files

An MSS data tape in Universal or LARSYS format must be assigned to logical unit C. If initial statistics are input from tape, logical unit A must be assigned to the tape. If the statistics are input from a Module Stat Deck, then a Fastran file must be assigned to logical unit A.

2. Output files

The output statistics may be saved on file 2 of logical unit A or punched on cards.

3. Scratch files

A drum file is used as an intermediate image data file.

A temporary Fastran file, logical unit I, is required when estimating the ALPHA parameter vector only.

4. Card Input

All required input card-types are described below. The actual formats for card input are described in Section 3 of the LARSYS Users Guide (see ref [5]).

4.1 PROCESSOR CARD

The SIGEXT processor must be activated by the '$SIGEXT' processor card.

Keyword *	Function
$SIGEXT	Directs the LARSYS system executive monitor to load the signature extension processor and begins input of processor control cards.

*Briefly, the KEYWORD must begin in Column 1 and the parameter field begins in Column 11 and ends in Column 72.

4.2 SPECIAL SYSTEM DECKS

The training statistics may be input by means of the 'Module STAT deck'. The LARSYS system deck formats are described in SECTION 3 of the LARSYS Users Guide [5].

4.3 PROCESSOR CONTROL CARDS

The following control cards are the complete set recognized by the signature extension processor. All options available to the user are exercised by use of the appropriate processor control card. If a default condition is specified the control card is optional, with the processor using the default parameters if the control card is not input. If no default condition is specified, the control card is mandatory input to the processor. Ordering the sequence of processor control cards is unnecessary, with the exception of the '*END*' card and '$END*' card. The '*END*' card must follow the last processor control card the '$END*' card must follow the last field definition card defining an area to be processed.

KEYWORD	PARAMETER (S)	FUNCTION
MODULE	(DEFAULT: training subclass statistics will be read from the input file, SAVTAP)	Indicates to the processor that the training subclass statistics will be card-input. The 'MODULE STAT Deck' must immediately follow this 'MODULE' control card. See Section 3.14.1 ref [5] for further detail on this deck.

KEYWORD	PARAMETER(S)	FUNCTION
OPTION	STATS (DEFAULT: no training subclass statistics printout.	Training subclass statistics will be printed out (mean vector and covariance matrix) for each subclass.
ØPTIØN	PUNCH (DEFAULT: no stat deck is punched	Output statistics from the SIGEXT Processor will be punched on a 'MØDULE STAT DECK'.
SUBCLASS	$K_1,K_2,\ldots,K_m$ K_i are integers where $1 \leq K_i \leq 60$. m = no. of subclasses in training statistics. (DEFAULT: All the training subclasses will be used.)	The subclass numbers listed will be the set of subclasses used by the processor. This set must be a subset of the training subclasses on the SAVTAP file= $K_1,K_2,\ldots,K_M$ are integer numbers designating subclasses as they occur on the statistics file.
HED1	Any 60 alphanumeric characters beginning in col. 11. (DEFAULT: standard heading, "LYNDON B. JOHNSON SPACE CENTER')	The Processor will use these 60 characters as the first line of the page heading for every page of printout produced. (Blanks are included as 'characters')

KEYWORD	PARAMETER(S)	FUNCTIONS
ALPHA	$A_1, A_2, \ldots, A_m$ or $N*A_1, K*A_2, A_3$ where N and K are arbitrary repitition factors and A_i are decimal numbers.. m is the number of subclasses. The A_i's are automatically normalized so that $\Sigma A_i = 1$. (DEFAULT: If no ALPHA card is included then $A_i = \frac{1}{m}$ for each i = 1,...,m.)	These entries are the initial values of the mixture proportions in the MLE procedure
CHANNEL	TRAINING = $N_1, \ldots, N_n$, TEST = $L_1, \ldots, L_n$ Channels $N_1, \ldots, N_n$ are selected from the input training statistics file and channels $L_1, \ldots, L_n$ are selected from the input image data file. (DEFAULT: all channels are selected from both stat file and the image data file. Warning these files must have the same number of channels).	

KEYWORDS	PARAMETER(S)	FUNCTION
HED2	Any 60 alpha-numeric characters beginning in col. 11. (DEFAULT: standard second line, 'HOUSTON, TEXAS')	The processor will use these 60 characters as the second line of the page heading, for every page of print-out produced (blanks are included as characters')
COMMENT	Any 60 alpha-numeric characters beginning in column 11. (DEFAULT: all blanks in third line of page heading)	The processor will use these 60 characters as the third line of the page heading, for every page of print-out produced (Blanks are included as 'characters')
DATE	MM DD YY any (up to) 12 alphanumeric characters beginning in column 11. (DEFAULT: the current date, at the time of the run, will be placed in upper right-hand corner of each page of print-out).	The processor places these 12 characters in the upper right-hand corner of each page of print-out. Normally used to specify a calendar date.
END	(DEFAULT: none)	This control card terminates the input of processor control cards, and initiates the input of data card(s) defining area(s) to be processed. The field definition card(s) for the area(s) to be processed must immediately follow the '*END*' control card.

KEYWORD	PARAMETER(S)	FUNCTION
$END*	-none-	This control card terminates the input of field definition card(s) defining area(s) to be processed. This control card must immediately follow the last field definition card.

4.4 FIELD DEFINITIONS

Area(s) to be processed are communicated to the signature extension processor by using the 'field definition' data card, described in Section 3.1.3, of Ref. [5]. These cards contain the scan line and sample coordinates for the vertices of the polygon shaped area over which signature extension is performed. At least one 'field definition' card must be in the run deck, immediately following the '*END*' control card. As many 'field definition' cards as there are area(s) may be input. The areas specified on the 'field definition' card(s) must be available on the MSS data file that is input to the processor.

4.5 Iteration Control Cards

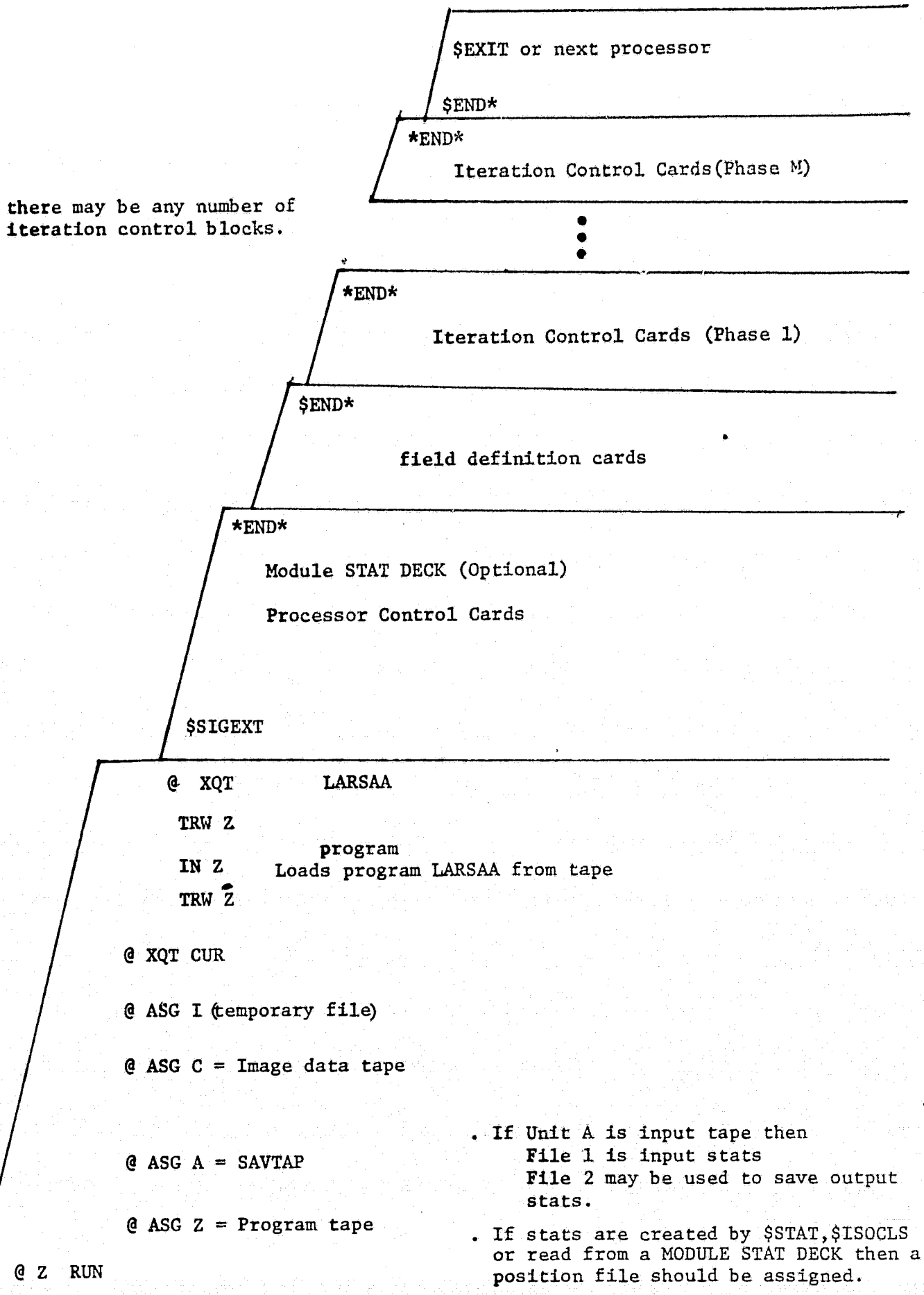
KEYWORD	PARAMETER	FUNCTION
ØPCØDE	1 or 2 (DEFAULT 1)	If 1 is chosen, the general MLE procedure is used. If 2 is chosen, then the Ax + b form is used.

KEYWORD	PARAMETER	FUNCTION
IFLGA	$I_1,\ldots,I_m$ (DEFAULT $I_j = 1$ for $j = 1,\ldots,m$)	If $I_j = 0$, then the mixture proportions for the j^{th} class is fixed for this iteration phase. If $I_j = 1$, then the mixture proportions for the j^{th} class is updated.
IFLGM	$I_1,\ldots,I_m$ (DEFAULT $I_j = 1$ for $j = 1,\ldots,m$)	If $I_j = 0$, then the mean vector for the j^{th} class is fixed for this iteration phase. Otherwise, the j^{th} mean vector is updated.
IFLGS	$I_1,\ldots,I_m$ (DEFAULT $I_j = 1$ for $j = 1,\ldots,m$)	If $I_j = 0$, then the j^{th} covariance matrix is fixed for this iteration phase. Otherwise the j^{th} covariance matrix is updated.
IFLGAM	I (DEFAULT I = 1)	If I = 0 then the diagonal matrix A is fixed for this phase of the Ax + b iteration Otherwise the matrix A is updated.
IFLGBM	I (DEFAULT I = 1)	If I = 0, then the vector b is fixed for this phase of the Ax + b iteration. Otherwise, the vector b is updated.
MØDES	1 or 2 (DEFAULT 1)	If 1 is chosen, then the full covariance matrix is used. Otherwise, only a diagonal covariance matrix is used.

EPSA	0 < X ≤ 2 real number (DEFAULT X = 1.0)	X is the step size parameter for the proportion iterations.
EPSM	0 < X ≤ 2 real number (DEFAULT X = 1.0)	X is the step size parameter for the mean vector iterations
EPSS	0 < X ≤ 2 real number (DEFAULT X = 1.0)	X is the step size parameter for the covariance matrix iteration.
EPSAM	0 < X ≤ 2 real number (DEFAULT X = 1.0)	X is the step size parameter for the A matrix iteration.
EPSBM	0 < X ≤ 2 real number (DEFAULT X = 1.0)	X is the step size parameter for the b vector iteration.
TØLA	X > 0 real number (DEFAULT X = .001)	Iteration tolerance on proportions. The maximum change in proportions from one iteration to the next over all classes is compared to TØLA in the iteration termination decision.
TØLM	X > 0 real number (DEFAULT X = .5)	Iteration tolerance on mean vectors. The maximum change in mean vector components over all channels over all classes from one iteration to the next is compared to TØLM in the iteration termination decisions.
TØLS	X > 0 real number (DEFAULT X = 1.)	Iteration tolerance on covariance matrices. The maximum change in covariance matrix elements over all channels over all classes from one iteration to the next is compared to TØLS in the iteration termination decision.

TØLAM	X > 0 real number (DEFAULT X = .05)	Iteration tolerance on the matrix A in the Ax + b option. The maximum change in the elements of the diagonal matrix A from one iteration to the next is compared to TØLAM in the iteration termination decision.
TØLBM	X > 0 real number (DEFAULT X = .05)	Iteration tolerance on the vector b in the Ax + b option. The maximum change in the elements of the vector from one iteration to the next is compared to TØLBM in the iteration termination decision.
ITRPNT	I integer (DEFAULT I = 0)	Iteration print control. If I = 0 only the results of the first and last iterations for each iteration phase are printed. Otherwise, all iterations are printed.
ITLIM	I integer (DEFAULT I = 10)	Limit on number of iterations permitted. The iteration process is terminated for a given iteration phase after ITLIM iterations have occurred regardless of tolerance comparisons
END	-none-	This card follows each set of iteration control parameters which define an iteration phase.
$END*	-none-	Control is returned to the monitor from the SIGEXT processor.

5. Deck Setup for LARSAA (SIGEXT)



6. Flow of program