AgRISTARS

A Joint Program for Agriculture and Resources Inventory Surveys Through Aerospace Remote Sensing

Supporting Research

January 1983

FINAL REPORT--ESTABLISHMENT OF A CENTER OF EXCELLENCE FOR APPLIED MATHEMATICAL AND STATISTICAL RESEARCH

Wayne A. Woodward and H. L. Gray

Center for Applied Mathematical and Statistical Research
Southern Methodist University
Dallas, Texas 75275

1. Report No. SR-63-04408	2. Government Accession No.	3. Recipient's Catalog No.

4. Title and Subtitle
FINAL REPORT--Establishment of a Center of Excellence of Applied Mathematical and Statistical Research

5. Report Date
January 1983

6. Performing Organization Code

7. Author(s)
Wayne A. Woodward, H. L. Gray

8. Performing Organization Report No.

10. Work Unit No.

9. Performing Organization Name and Address
Center for Applied Mathematical and Statistical Research
Southern Methodist University
Dallas, Texas 75275

11. Contract or Grant No.
NAS 9-16438

13. Type of Report and Period Covered
Aug. 1, 1981-
Final Report Jan. 31, 1983

12. Sponsoring Agency Name and Address
National Aeronautics and Space Administration
Johnson Space Center
Houston, Texas Tech. Monitor: G. Houston

14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract

In this report, we describe the research efforts which have been undertaken at Southern Methodist University (SMU) in support of contract NAS 9-16438. Our main purpose has been to establish a "Center of Excellence" for directing and carrying out research in the area of aerospace remote sensing in order to adequately organize and direct mathematical and statistical research in support of the AgRISTARS objectives. Our work has been concentrated in two major efforts. First, we have conducted a thorough assessment of the current state of the art (as definted by NASA and its contractors) with regard to estimation efforts in support of the crop production estimation problem. Our second major effort has been in the development of alternative generic proportion estimation techniques. This report summarizes our accomplishments during the contract period August 1, 1981 to January 31, 1983.

17. Key Words (Suggested by Author(s))
Final Report
Center of Excellence
Generic proportion estimation

18. Distribution Statement

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price*
U	U	103	

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

FINAL REPORT

Establishment of a Center of Excellence for Applied Mathematical and Statistical Research

Wayne A. Woodward and H. L. Gray

Center for Applied Mathematical and Statistical Research
Southern Methodist University
Dallas, Texas 75275

January 1983

Final Report

Establishment of a Center of Excellence for Applied Mathematical and Statistical Research

Introduction

In this report, we will describe the research efforts which have been undertaken at Southern Methodist University (SMU) in support of contract NAS 9-16438. As the title of the contract states, a first priority has been the establishment of a "Center of Excellence" for directing and carrying out research in the area of Aerospace Remote Sensing. Such a center is needed in order to adequately organize and direct mathematical and statistical research in support of the AgRISTARS objectives. We have conducted a thorough assessment of the current state of the art (as defined by NASA and its contractors) with regard to estimation efforts in support of the crop production estimation problem. In particular, we have reviewed old methods and have evaluated methods in current use.

This review and evaluation process was facilitated through seminars in which methods were presented and discussed. Among the methods reviewed in this manner were:

Proportion estimators from LACIE - analyst dependent

(i) PC estimator

(ii) Procedure 1 estimator

(iii) etc.

CLASSY/APEP

AMOEBA/HISSE

Procedure M

Spatial/Color Sequence

ERIM Profile Model

Multitemporal Profile Modeling

Others

Reviews and evaluations have been presented as lengthy written reports, such as the report in Appendix A on the multitemporal profile modeling. Other reports have been in the form of written and oral reports delivered to the project director and at workshop settings.

Our second major effort has been in the area of development of alternative generic proportion estimation techniques. Of course, there is no distinct dividing line between the efforts involved in the two tasks. For example, as we developed alternative proportion estimation

techniques, we compared these with the existing techniques. This provides us with further insight into the performance of the current procedures.

Major Reports

Our efforts have resulted in three major written reports which will be introduced in this section. These reports are included in the Appendix. During the early months of the contract, our major efforts were in the evaluation of current and former methods. At this time, G. Badhwar had introduced a procedure for modeling the multitemporal profile for a crop. It was believed that this profile (usually of "greenness") across the growing season would provide feature variables with superior discriminating power. Early results using this procedure showed that it had promise. We were asked to evaluate this procedure and make recommendations. Our report is included in Appendix A, and was presented at the January 1982 Quarterly Technical Interchange. Basically, we took a systematic look at the modeling of the greenness profile, and discussed the properties which such a model should possess. Our major concern with the early Badhwar model was that in that model, emergence date, t_0, was not a location parameter. This concern was mentioned in discussions with Dr. Badhwar in October 1981. Recent modifications of the profile model have

included emergence date as a location parameter, and we believe that our evaluations had an impact on these modifications. Various possible models for the profile are discussed in the report in Appendix A, along with results from both simulated and LANDSAT data.

In Appendix B we include a report which is a compilation of results presented at both the April and October 1982 Quarterly Technical Interchanges and at the special mini-symposium at NASA in December, 1982. These results were also presented at a special session on remote sensing at the national meetings of the American Statistical Association in Cincinnati, August 1982 and were published in the Proceedings of the Section on Survey Research Methods. This report has been distributed as Technical Report SR-62-04376, and it summarizes some of the results obtained in our second major effort, specifically, the development of alternative generic proportion estimators.

The mixture model is currently being used extensively by NASA and its contractors to obtain crop proportion estimates. CLASSY was an early result of this effort, and current investigations in this area are included in the APEP study headed by R. Heydorn. Parameter estimation in this mixture model is being accomplished using maximum likelihood (ML) techniques based upon an assumption that the underlying component distributions are normally distributed. Although

ML estimators have desirable optimality properties when the underlying assumptions are valid, they are notoriously sensitive to departures from these underlying assumptions. It is our belief that the underlying normality assumption in the case of LANDSAT data is of questionable validity. For these reasons we investigated alternatives to ML estimation which were not as sensitive to departures from the underlying assumptions. Our investigations in this area have centered around minimum distance (MD) estimation. We conducted a simulation study in which the ML and MD estimators were compared on both mixtures of normal and of non-normal components. We have shown that MD estimators are competitive with ML estimators when the components actually are normal, while they tend to be superior when the components are non-normal yet symmetric. The non-normal model used is the Student's t with 4 degrees of freedom, and similar results have recently been obtained for the double exponential. Neither of these models is extremely non-normal. Thus even when the non-normality would probably not be detectable visually, the MD estimates are better than the "optimal" ML estimates. The results of this study are given in Appendix B.

Although the results shown in Appendix B basically reflect research efforts in the area of development of generic proportion estimation techniques, they also involve an "evaluation" component. For example, we believe that the

results of our simulations provide much needed insight concerning the role of the normality assumption in the current implementations. For example, it was found that normal based estimation techniques often provide very biased estimates when the underlying distributions are actually skewed. For example, a 50-50 mixture of two chi-squared distributions will "confuse" the normal based procedures which assume that the underlying distributions are symmetric. This phencmenon is mentioned in Section 5 of the report in Appendix B. The problem of asymmetry is one of extreme concern since the variables currently being used in proportion estmation are feature variables from the profile models, and these variables have been shown to have asymmetric distributions. In Appendix C we have suggested an approach to the problem of obtaining proportion estimates when the underlying distributions are asymmetric. This report reflects material which was presented at the October 1982 Quarterly Technical Interchange and at the December 1982 mini-symposium. Briefly, instead of assuming that components are normally distributed, we have proposed that they be assumed to have Weibull distributions. This assumption is made since Weibull distributions are "flexible" in the sense that they can be either symmetric or asymmetric depending upon parameter configurations. Properties of the Weibull are summarized in Appendix C along with the proposed procedures for estimating the parameters

in a mixture of Weibulls. This procedure utilizes the MD techniques discussed in Appendix B. Although ML estimation is shown to be quite untractable in this setting, the MD estimators are relatively easy to obtain. The results in this report suggest that this Weibull assumption may prove to be a viable alternative to the procedures now in use.

Future Research Directions

In each of the reports in Appendices A-C, suggestions are made for future research. We refer the reader to those sections for a discussion of research topics which are suggested by the current results.

Other Reports

In Appendices D and E we include two other reports which were technical evaluations requested by the project directors.

APPENDIX A

A Temporal Model For Crop Classification

by

H. L. Gray and W. A. Woodward

Introduction

In a recent article G. D. Badhwar (1980) suggested a function $\rho_b(t)$ for modeling the greenness spectral profile of a crop from emergence to harvest. The function $\rho_b(t)$ is defined as follows:

$$\rho_b(t) = \rho_0 \qquad , 0 \leq t \leq t_0$$
$$\rho_b(t) = \rho_0\left(\frac{t}{t_0}\right)^{\alpha} \exp\left[-\beta(t^2-t_0^2)\right], \quad t_0 \leq t \qquad \text{(I)}$$

where

ρ_0 = Soil greenness

t_0 = Emergence date

and α and β are parameters to be estimated.

By applying the Model I to Landsat spring wheat data for LACIE segments in North Dakota and Minnesota, Badhwar demonstrated that Model I could be used to successfully estimate t_0 in these cases.

Badhwar (1979) and Badhwar, Carnes, and Austin (1981) have also applied the model in (I) to the problem of crop classification. It was demonstrated that α, β and t_0 could be used as features to correctly classify corn and soybeans. Again these methods were utilized on Landsat data, and the results were impressive on the data considered. Austin (1980), (1981) has reported on more extensive testing of these methods on LANDSAT data with the results

again being quite good. In this paper we examine (I) more closely from the perspective of a desirable mathematical model for describing crop greenness. Some shortcomings of Model I are noted and some modifications are proposed. It is shown how this modified model can be utilized for crop classification from LANDSAT data. The results are then demonstrated on some LANDSAT corn-soybean data.

Analysis

Even though a mathematical model may perform well on a selected number of data sets, it seems desirable that it also satisfy some of the more obvious physical constraints imposed by the phenomenon it seeks to explain. If this is not the case, i.e., if it does not satisfy such constraints, then it behooves the investigator to explain why such constraints can be relaxed and the model still be expected to perform its function.

Several properties which a function, $\rho(t)$ for greenness should possess are

(i) $\rho(t) = \rho_0 \qquad t \leq t_0$

(ii) $\rho(t) = \rho_T \qquad t \geq t_1$, where ρ_T is terminal greenness and t_1 is the corresponding point in time.

(iii) $\rho'(t)$ should be independent of ρ_0 after full coverage.

(iv) t_0 should be a location parameter, i.e., ρ should be a function of $t - t_0$.

Several other criteria could be listed, but the above suffice for the current discussion. The condition (iv) requires some comment. Certainly the same variety of crop planted at greatly differing times would be expected to have greenness character-

istics which differ in more ways than simple translation. However, the model in (I) is posed for crops in the same segment and as such the planting dates of the same crop are not expected to differ greatly even though it is possible. In any event it is the opinion of these authors that effects of t_0, other than location effects, have to be relegated to noise in the model or treated as producing a different classification, not necessarily generically different but labeled different, spring wheat and winter wheat for example. In any event there is no reason to believe that the model in (I) speaks to this problem. Moreover, in (I), t_0 is clearly not a location parameter. Note also that although $\rho_b(t)$ satisfies (i), it clearly fails to satisfy (ii) and (iii).

Actually Model I represents a considerable simplification of the general model suggested by Badhwar in (1980). The following definition for $\rho(t)$ makes use of that general model and the function

$$E(t;\alpha,\beta) = t^{\alpha} \exp(-\beta t^2) \quad , \tag{1}$$

demonstrated by Badhwar to be of some value in describing greenness. Let $F(t)$ be a probability distribution function such that $F(t) = 0$ for $t \leq 0$ and $F(t) = 1$ for $t \geq \lambda$. Then define

$$\rho(t) = [1-pF(t-t_0)]\rho_0 + pF(t-t_0)[\rho_1 + DE(t-t_0;\alpha,\beta)], \tag{2}$$

where

p = proportion of ground covered for $t \geq \lambda$

ρ_0 = soil greenness

ρ_1 = crop greenness at terminal greenness

t_0 = emergence date

α = greenup parameter

β = greendown parameter

D = constant

In (2) clearly

a) $\rho(t_0) = \rho_0$

b) $\rho(t) \to \rho_0 + (\rho_1 - \rho_0)p$ = terminal greenness of the pixel

c) if $p = 1$, $\rho'(t)$ is independent of ρ_0 for $t \geq \lambda$

d) t_0 is a location parameter.

Thus interpreting (b) as a satisfactory approximation to (ii), we can say that the model in (2) satisfies conditions (i)-(iv) and makes use of important aspects of the exponential function found by Badhwar as a model for greenness.

Unfortunately the model in (2) has 9 unknown parameters (assuming that the distribution F(t) has one unknown parameter). Since the data to which we intend to apply our model consists of no more than 8 acquisitions, (2) is obviously not acceptable. The problem is complicated by the fact that it is desirable to classify the data as early as possible. Therefore, from a practical point of view, one can probably only count on 4 to 6 acquisitions before a classification must be made. This clearly eliminates (2) as a practical model.

Rather than abandon (2), we will now investigate the possibility of reducing the number of unknown parameters. In the pages which follow, we will investigate the effects of the simplication we impose. Since the data to be considered includes no information for separately estimating p, the model can with no loss in generality be rewritten as

$$\rho(t) = \rho_0 + [A + BE(t-t_0;\alpha,\beta)]\, F(t-t_0) \qquad (3)$$

where

$$A = (\rho_1 - \rho_0)p, \qquad B = pD$$

and now A and B are the unknown parameters to be estimated. The model in (3), therefore, requires 8 parameters to be estimated, a reduction of 1.

For the data to be considered there is no deleterious effect in going from (2) to (3) since no data are available from which to estimate p, ρ_1, and D separately. It should be noted that (3) applies whether or not we have full crop coverage (i.e. whether or not p = 1). Of course the number of parameters in (3) is still too large to be useful.

Investigation of LANDSAT Corn-Soybean data suggests that assuming F(t) to be the distribution function associated with a uniform density over $(0,\lambda)$, yields a reasonable linear approximation to F(t). Under this assumption we have

$$F(t-t_0) = \begin{cases} 0 & t \le t_0 \\ \dfrac{t-t_0}{\lambda-t_0} & t_0 \le t \le \lambda \\ 1 & \lambda \le t \end{cases} \tag{4}$$

and (3) becomes

$$\rho(t) = \begin{cases} \rho_0 & t \le t_0 \\ \rho_0 + A\dfrac{t-t_0}{\lambda-t_0} + B\dfrac{(t-t_0)^{\alpha+1}}{\lambda-t_0} \exp\left(-\beta(t-t_0)^2\right) & t_0 \le t \le \lambda \\ \rho_0 + A + B(t-t_0)^{\alpha} \exp\left(-\beta(t-t_0)^2\right) & \lambda \le t. \end{cases} \tag{5}$$

The model in (5) represents a reduction of one parameter over (3) since the parameter λ is absorbed in the uniform distribution. From

(3), note that for $t_0 < t < \lambda$

$$\rho'(t) = AF'(t-t_0) + B[E'(t-t_0;\alpha,\beta)F(t-t_0) + E(t-t_0;\alpha,\beta)F'(t-t_0)] \quad . \tag{6}$$

But, if $\alpha > 0$,

$$E'(t-t_0;\alpha,\beta)F(t-t_0) + E(t-t_0;\alpha,\beta)F'(t-t_0)\Big|_{t\to t_0^+} = 0 \quad .$$

The left hand derivative of $\rho(t)$ at t_0 is clearly zero. Therefore, the derivative of $\rho(t)$ exists at t_0 if and only if

$$AF'(t_0) = 0 \quad , \tag{7}$$

where here $F'(t_0)$ denotes the right hand derivative at t_0. Since this seems desirable and $F'(t_0) \neq 0$, we are left with requiring $A = 0$. Since $\rho_1 \neq \rho_0$ this is clearly incorrect. However, it does not seem that taking $\rho_1 = \rho_0$ would seriously degrade the model's ability to classify since ρ_1 will probably not differ greatly from ρ_0, and ρ_1 may be nearly constant from crop to crop. Essentially this error is due to our linear approximation of $F(t)$, for if $F(t)$ were quadratic the requirement that $A = 0$ could be eliminated. Nevertheless, for the reasons mentioned above, and the fact that it results in one less parameter, we now take $A = 0$ in (3) to obtain the model

$$\rho(t) = \begin{cases} \rho_0 & t \leq t_0 \\ \rho_0 + \frac{B}{\lambda - t_0}(t-t_0)^{\alpha+1}\exp(-\beta(t-t_0)^2) & t_0 \leq t \leq \lambda \\ \rho_0 + B(t-t_0)^{\alpha}\exp(-\beta(t-t_0)^2) & t \geq \lambda \quad . \end{cases} \tag{II}$$

Note that

$$\rho'(t) = \frac{B(\alpha+1)}{\lambda - t_0}(t-t_0)^{\alpha} e^{-\beta(t-t_0)^2} + \frac{B(t-t_0)^{\alpha+2}}{\lambda-t_0}(-2\beta)e^{-\beta(t-t_0)^2} \quad \text{for } t_0 \leq t < \lambda$$

$$= \alpha B(t-t_0)^{\alpha-1} e^{-\beta(t-t_0)^2} - 2\beta B(t-t_0)^{\alpha+1} e^{-\beta(t-t_0)^2} \text{ for } \lambda < t .$$

Thus for $\rho'(t)$ to exist at λ we must have

$$B(\lambda-t_0)^{\alpha-1} e^{-\beta(\lambda-t_0)^2} = 0 . \tag{8}$$

Unfortunately this cannot occur so we must examine the model further. Since our desire is to simplify the model we do not wish to add additional terms to the model which would guarantee (8), especially for the purpose of fitting the curve in the right tail, since by that time the data will already be classified. It can be demonstrated numerically that B, λ and α play similar roles in Model II and as a result are jointly very nonrobust to errors. This is particularly true of B and α. With only a few data points it, therefore, is desirable to fix α or B in advance.

In other words, when there are only a few data points available, and there is error in the model, small differences in data values can lead to large differences in B and α. This is due to the fact that for fixed B or α a reasonable fit to the data can be obtained by varying the other. We thus let B = 1 and arrive at the following model

$$\rho(t) = \begin{cases} \rho_0 & t \leq t_0 \\ \rho_0 + \dfrac{(t-t_0)^{\alpha+1}}{\lambda-t_0} e^{-\beta(t-t_0)^2} & t_0 \leq t \leq \lambda \\ \rho_0 + (t-t_0)^{\alpha} e^{-\beta(t-t_0)^2} & \lambda \leq t , \end{cases} \tag{III}$$

where t_0, α, $\beta > 0$, $\lambda > t_0$.

Model III is a five parameter model, and therefore is a candidate for application on the data we will consider. Further simplifications of this model come to mind. For example, one might simply fix λ as some maximum value. One might also argue that attempting to fit two curves together at a point so late in time as λ, with any degree of validity, requires data past the point of interest, and hence Model III should be modified to

$$\rho(t) = \begin{cases} \rho_0 & t \leq t_0 \\ \rho_0 + B(t-t_0)^{\alpha} \exp(-\beta(t-t_0)^2), & t_0 \leq t \end{cases}, \qquad \text{(IV)}$$

where now B is again to be estimated.

Moreover, again noting that B and α play much the same role, and that a classification is desired as soon as possible, it might further be argued that $\rho(t)$ could be reduced to the four parameter model

$$\rho(t) = \begin{cases} \rho_0 & \\ \rho_0 + (t-t_0)^{\alpha} \exp(-\beta(t-t_0)^2) & t_0 \leq t \end{cases}. \qquad \text{(V)}$$

In the next section, we investigate via simulations the effects on classification of the above suggested simplifications.

Feature Selection for Classification and Simulation

Once an appropriate model has been obtained the problem of classification is not solved. The proper features to be utilized and the manner in which they are to be used must still be decided. Badhwar, Carnes, and Austin (1981) selected α, β, and t_0 as the appropriate features and utilized these in the Ho-Kashyap algorithm (essentially the linear discriminant function) to form a discriminating plane.

It seems reasonable that other "features" of the model obtained might also prove to be as useful, or more useful, than the model parameters in separating crops. One feature which will be investigated in the present report is the maximum value of the fitted curve. If t_m is the Julian date at which this maximum occurs, then $\rho(t_m)$ is the corresponding feature of interest. In addition to the maximum greenness it appears that $t_m - t_0$, i.e. the time from emergence to peak greenness, is also a feature of potential importance in the classification problem. As our investigations continue, we anticipate the examination of still other features, but in the present report we will examine only these two features in addition to the model parameters as investigated by Badhwar, Carnes, and Austin.

Performance of the Proposed Profile Models

In this paper we have discussed a general profile model which we believe is appropriate for purposes of describing the greenness of a crop across time. However, the general expression for the model is such that estimation of the parameters would be impossible given the 5-8 observations typically available from LANDSAT observations. Thus, various simplifcations of this model were proposed (Models II-V). In this section, we will discuss the results of our preliminary investigations into the performance of these models and Model I proposed by Badhwar.

Our investigations have been primarily in two areas. First, we have utilized Models I-V in order to estimate parameters and features from 1978 field data on corn and soybeans from Segment 882 in Palo Alto, Iowa. From the results of these investigations, we are able to find typical Model III parameters for corn and for soybeans. These parameters are then used

to simulate profile data from our "typical" corn model and soybean model, and investigate the performance of the various models based upon these simulations. Model III was used in the simulation since it was the most general model for which "typical" values of the parameters could be found.

As mentioned earlier, we will not primarily be investigating the models with respect to estimation of the model parameters but for the purpose of ascertaining the effect on features such as $t_m - t_0$ and $\rho(t_m)$ which may be used for classification. In Table I, the value of t_m is given for each of the models under consideration.

Table I - Julian Data (t_m) of Maximum Greenness Associated with Models I-V

Model	t_m
I	$\sqrt{\frac{\alpha}{2\beta}}$
II, III	$t_0 + \sqrt{\frac{\alpha+1}{2\beta}}$ if $\lambda > \sqrt{\frac{\alpha+1}{2\beta}}$ $t_0 + \sqrt{\frac{\alpha}{2\beta}}$ if $\lambda < \sqrt{\frac{\alpha}{2\beta}}$ $t_0 + \lambda$ if $\sqrt{\frac{\alpha}{2\beta}} < \lambda < \sqrt{\frac{\alpha+1}{2\beta}}$
IV, V	$t_0 + \sqrt{\frac{\alpha}{2\beta}}$

For each model to be considered here, the parameter estimation was accomplished using Marquardt's (1963) method for unweighted least squares estimation of nonlinear parameters.

Parameter Estimation Utilizing 1978 Data from Segment 882

In this section, we will report the parameter estimation results, based upon the utilization of Models I-V, for modeling the multitemporal

behavior of corn fields C01-C07 and soybean fields SY11 - SY17. In Figure 1, we have plotted the five models obtained for soybean field SY. It is interesting to note the various "interpretations" concerning the proper functional curve to fit to these eight points. Notice in particular the fact that t_0, the emergence date, varies considerably from model to model.

In Table II we present the results of the parameter estimation based upon Models I-V. Several observations can be made concerning the results displayed in Table II. A first observation is that parameter estimates in Model II are less stable than those in the Model III. In Model II the parameter estimates of α and B are quite variable, a behavior which was discussed earlier in this report. Based upon the results for Model II and Model III, it would appear that indeed more stable estimates of α are obtained when B is set equal to a constant (in this case 1). It should be noted that the stability of β is also affected by the inclusion of B in the model, but not to the extent that α is affected. It appears that we simply do not have a sufficient number of readings to obtain reliable estimates of 6 parameters. It should be noted that the 1978 data for segment 882 contains 8 observations. Obviously, in most situations, as many as 8 observations will not be available and hence the need to find a satisfactory reduced model is clear.

Using Model III, there is an indication that both α and β are larger for soybeans than for corn, and reasonable separation between the two crops could be made using these two parameters. Also of interest is the fact that the estimation of the maximum greenness and $t_m - t_0$ features in Model II are as stable as they are

Table II - Parameter Estimation for Corn and Soybean Field Data - Segment 882

MODEL I

		$\hat{\alpha}$	$\hat{\beta}^*_c$	$\hat{t}_0$	$\hat{max}$	$\hat{t}_m - \hat{t}_0$
CORN	CO1	15.2	1.16	146	39.9	68
	CO2	18.9	2.14	147	43.8	63
	CO3	21.5	2.42	149	52.1	62
	CO4	18.6	2.13	146	44.8	62
	CO5	14.4	1.60	134	44.2	78
	CO6	19.1	2.21	147	43.8	61
	CO7	18.6	2.15	146	43.9	62
SOYBEANS	SY11	24.9	2.52	165	42.6	57
	SY12	23.7	2.54	155	52.1	61
	SY13	21.7	2.40	152	48.4	61
	SY14	24.2	2.58	158	49.9	59
	SY15	24.9	2.67	158	52.3	58
	SY16	27.7	2.80	166	55.6	56
	SY17	26.9	2.82	163	51.5	55

*$\hat{\beta} = .0001\ \hat{\beta}_c$

PRECEDING PAGE BLANK NOT FILMED

12, 13, 14, 15, 16, 17.

Table II (continued)

MODEL II

		$\hat{\alpha}$	$\hat{\beta}_c$	$\hat{t}_0$	$\hat{\lambda}-\hat{t}_0$	$\hat{B}$	$\hat{Max}$	$\hat{t}_m-\hat{t}_0$
	C01	.89	2.08	142	125	3.9	40.6	67
	C02	-.15	1.89	158	73	153.9	43.5	48
	C03	.39	2.20	154	68	23.3	52.2	56
CORN	C04	-.23	1.91	158	63	220.7	44.9	45
	C05	1.02	1.38	141	45	1.0	44.6	61
	C06	.02	2.20	155	76	91.7	44.0	48
	C07	.93	2.76	144	123	4.8	45.5	59
	SY11	1.06	2.03	152	70	1.20	41.7	71
	SY12	1.12	2.51	151	80	1.52	51.7	65
	SY13	2.06	2.85	155	0	.03	49.3	60
SOYBEAN	SY14	1.11	2.49	152	79	1.47	49.2	65
	SY15	1.15	3.13	157	74	1.57	52.5	58
	SY16	1.14	2.78	160	71	1.43	55.3	62
	SY17	1.12	2.64	155	76	1.50	50.8	62

MODEL III

		$\hat{\alpha}$	$\hat{\beta}_c$	$\hat{t}_0$	$\hat{\lambda}-\hat{t}_0$	$\hat{Max}$	$\hat{t}_m-\hat{t}_0$
	C01	1.11	1.83	136	95	40.1	76
	C02	1.07	2.07	150	39	45.4	51
	C03	1.15	2.35	150	48	56.7	49
CORN	C04	1.07	2.08	151	23	45.5	51
	C05	1.01	1.34	141	45	44.4	61
	C06	1.07	2.01	146	45	45.9	52
	C07	1.06	2.04	149	27	44.6	51
	SY11	1.10	1.98	151	80	41.6	73
	SY12	1.24	2.75	152	79	52.2	64
	SY13	1.12	2.06	146	62	52.7	62
SOYBEAN	SY14	1.21	2.59	151	80	49.6	65
	SY15	1.26	3.10	155	76	52.5	60
	SY16	1.23	2.80	159	72	55.3	63
	SY17	1.23	2.76	154	77	51.1	64

Table II (continued)

MODEL IV

		$\hat{\alpha}$	$\hat{\beta}_c$	$\hat{t}_0$	$\hat{B}$	Max	$\hat{t}_m - \hat{t}_0$
	C01	.70	1.59	158	3.24	40.1	47
	C02	1.10	2.35	156	.93	44.5	48
	C03	2.29	3.00	146	.01	53.2	62
CORN	C04	.81	2.13	158	2.72	45.6	44
	C05	.51	1.43	158	7.10	44.7	42
	C06	1.08	2.40	155	1.03	44.6	47
	C07	.75	2.11	158	3.41	44.7	42
	SY11	3.24	3.09	148	.0002	42.5	72
	SY12	3.02	3.13	145	.0005	52.5	69
	SY13	2.98	2.85	138	.0006	49.3	72
SOYBEAN	SY14	2.97	3.41	149	.0007	50.5	66
	SY15	3.00	3.70	151	.0008	53.0	64
	SY16	2.88	3.46	156	.0010	55.6	65
	SY17	3.05	3.45	150	.0006	51.7	67

MODEL V

		$\hat{\alpha}$	$\hat{\beta}_c$	$\hat{t}_0$	Max	$\hat{t}_m - \hat{t}_0$
	C01	1.02	1.85	155	40.4	52
	C02	1.08	2.34	156	44.5	48
	C03	1.14	2.48	158	52.3	48
CORN	C04	1.09	2.41	155	45.9	48
	C05	1.05	1.86	152	45.1	53
	C06	1.09	2.40	155	44.6	48
	C07	1.09	2.46	155	45.1	47
	SY11	1.09	2.87	174	42.4	44
	SY12	1.18	3.23	170	53.2	43
	SY13	1.10	2.25	158	48.3	50
SOYBEAN	SY14	1.17	3.41	172	51.1	41
	SY15	1.20	3.69	172	53.8	40
	SY16	1.19	3.21	175	55.6	43
	SY17	1.18	3.35	172	52.1	42

in Model III. Thus, although the addition of the extra parameter B caused problems with the stability of the parameter estimation, the fitted curves were quite consistent at least with regard to the two features. From Models I and III we see that the maximum greenness is greater for soybeans than for corn, and that corn reaches its maximum greenness somewhat sooner than do soybeans. Again, reasonable separation between corn and soybeans could have been obtained based upon these two features for either Model II or Model III. The estimation of t_0 is approximately equally stable for Models II and III, with a slight indication being given that soybeans emerged later than corn. (Thus at least at this point there appears to be no negative effect in going from Model II to III.) Also of interest is the fact that the addition of the parameter B has a tremendous effect on $\lambda - t_0$, i.e. the time from emergence to maximum crop coverage (as mentioned previously, maximum crop coverage need not be <u>total</u> coverage for our model to apply). Based upon the data from Model III, it appears that time to total coverage, $\lambda - t_0$, is longer for soybeans than for corn.

The comparison between Models IV and V are similar to those between Models II and III. In particular, the inclusion of the parameter B in the model results in unstable estimates of both B and α. For these models, the general tendency is for α and β to be larger for soybeans than for corn. The emergence date, t_0, is of considerable interest. For Model IV, there does not seem to be any difference between emergence date for corn and soybeans. However, for Model V the estimate of t_0 for soybeans is approximately 170, which is significantly later than that for corn. Again, maximum

greenness and $t_m - t_0$ seem to be stable features for both models, with soybean attaining a larger value of greenness. The result of the late estimate of t_0 in Model V is to cause $t_m - t_0$ to not separate crops, whereas for Model IV this separation was apparent.

The parameter estimation using Badhwar's Model I was quite stable. Again, the tendency is for α and β to be larger for soybeans than for corn, emergence date to be later for soybeans, soybeans to attain a higher greenness, and for corn to attain its maximum greenness earlier than soybeans.

Simulations

In order to gain a better understanding of these models we have examined their performance in a simulation study. As a result of the parameter estimation study using Segment 882, we selected a typical set of corn parameters and a typical set of soybean parameters for Model III. These parameters and associated features are given in Table III.

Table III - Parameters and Features of Corn and Soybean Models (Model III)

	Corn	Soybeans
ρ_0	7.0	7.0
α	1.07	1.24
β_1^*	2.07	2.75
t_0	150.0	150.0
$\lambda - t_0$	40.0	80.0
Max	46.2	52.0
$t_m - t_0$	51.0	64.0

*$\beta = .0001\ \beta_c$

Note that although there was an indication that soybeans emerged somewhat later than corn on Segment 882, these simulations are based upon a common emergence date. One hundred realizations from each model were generated.

The simulated observations were of the form

$$\rho_s(t) = \rho(t) + \omega(t)\varepsilon(t)$$

where $\rho(t)$ is as defined in Model III and $\varepsilon(t)$ is a normal random variable with zero mean and unit variance. Note that $\varepsilon(t)$ and $\varepsilon(t')$ are independent if $t \neq t'$. In the simulation results presented here we have also taken $\omega(t) \equiv 1$.

Models I-V were applied to each realization within a set and parameter estimates and features were obtained. Summary statistics describing the results of these simulations are presented in Table IV. For each parameter we indicate the average of the parameter values obtained over the 100 realizations, the coefficient of variation in order to provide an indication of relative variability of each parameter, and lower and upper .90 content tolerance limits with 95% level of confidence. In other words there is a 95% level of confidence that 90% of parameter estimates obtained in this manner would fall between the two tolerance values given. These values will assist the reader in discerning the separability of the two crops on the basis of the given parameter. It should be noted that these tolerance limits are based upon an assumption that parameter estimates obtained in these ways will be normally distributed. This may or may not be a good assumption but nevertheless the tolerance limits given should provide crop separability information to the reader.

The results of the simulations are similar to the results

Table IV

Model I

			$\hat{\alpha}$	$\hat{\beta}_c^*$	$\hat{t}_0$	Max	$t_m - t_0$
CORN		$\bar{x}$	18.6	2.10	147.5	44.6	62.7
		CV	.04	.05	.01	.01	.03
		LTL	17.2	1.92	143.6	43.4	58.8
		UTL	20.0	2.28	151.4	45.8	66.6
SOYBEANS		$\bar{x}$	22.9	2.48	154.6	51.9	60.4
		CV	.04	.04	.01	.01	.04
		LTL	19.0	2.28	150.8	50.8	56.3
		UTL	24.6	2.68	158.4	53.0	64.5

*$\hat{\beta} = .0001\,\hat{\beta}_c$

Model II

		$\hat{\alpha}$	$\hat{\beta}_c$	$\hat{t}_0$	$\lambda - \hat{t}_0$	$\hat{B}$	Max	$t_m - t_0$
CORN	$\bar{x}$	1.11	2.02	152.1	2.82	2.05	45.6	53.2
	cv	.33	.22	.04	.61	2.26	.04	.17
	LTL	.42	1.19	141.9	0.0	0.0	41.8	36.6
	UTL	1.80	2.85	162.3	60.7	10.7	49.4	69.8
SOYBEANS	$\bar{x}$	1.13	2.69	151.3	78.3	3.82	52.1	62.7
	cv	.33	.10	.04	.11	2.17	.01	.09
	LTL	.42	2.21	141.2	61.9	0.00	50.8	51.6
	UTL	1.84	3.17	161.4	94.7	19.36	53.4	73.8

Model III

		$\hat{\alpha}$	$\hat{\beta}_c$	$\hat{t}_0$	$\lambda - \hat{t}_0$	Max	$t_m - t_0$
CORN	$\bar{x}$	1.06	2.04	152.3	28.4	45.6	51.9
	cv	.04	.19	.03	.57	.04	.12
	LTL	.98	1.32	144.5	0.0	42.5	40.6
	UTL	1.14	2.76	160.1	58.9	48.7	63.2
SOYBEANS	$\bar{x}$	1.22	2.60	149.8	75.7	52.6	65.2
	cv	.03	.10	.02	.11	.02	.05
	LTL	1.15	2.13	145.2	59.5	50.6	58.9
	UTL	1.29	3.07	154.4	91.9	54.6	71.5

Table IV (continued)

Model IV

		$\hat{\alpha}$	$\hat{\beta}_c$	$\hat{t}_0$	$\hat{B}$	$\hat{Max}$	$\hat{t}_m - \hat{t}_0$
CORN	$\bar{x}$	1.02	2.33	157.5	1.76	45.3	46.4
	cv	.18	.08	.02	2.17	.02	.10
	LTL	.68	1.97	151.4	0.00	43.9	38.0
	UTL	1.36	2.69	163.6	8.94	46.7	54.8
SOYBEANS	$\bar{x}$	2.31	3.19	152.2	.03	52.7	60.2
	cv	.12	.06	.03	3.39	.01	.07
	LTL	1.79	2.84	144.4	.00	51.5	51.8
	UTL	2.83	3.54	160.0	.21	53.9	68.6

Model V

		$\hat{\alpha}$	$\hat{\beta}_c$	$\hat{t}_0$	$\hat{Max}$	$\hat{t}_m - \hat{t}_0$
CORN	$\bar{x}$	1.08	2.41	156.9	45.5	47.5
	cv	.01	.05	.01	.01	.02
	LTL	1.06	2.19	155.1	44.2	45.5
	UTL	1.10	2.63	158.7	46.6	49.5
SOYBEANS	$\bar{x}$	1.18	3.23	168.5	53.1	42.7
	cv	.01	.05	.005	.01	.02
	LTL	1.16	2.91	167.0	51.8	40.9
	UTL	1.20	3.55	170.0	54.5	44.5

obtained from Segment 882 data. For example, the estimation of B in Models II and IV is quite unstable as indicated by the large coefficients of variation. For all five models, maximum greenness is a stable feature which seems to provide good separation between corn and soybeans. The time from emergence to maximum greenness is not quite as stable a feature as maximum greenness yet it seems to provide separation between crops for all models except Model I. The parameters α and β tend to be larger for soybeans than for corn in all models. However, as seen in the Segment 882 data, the estimation of α is not as stable in Models II and IV involving the B parameter. In Models II and IV the estimation of B is very unstable. Also of note is the fact that the estimation of $\lambda - t_0$ in Models II and III is not as stable as one would hope. In Models II and III the estimated parameters can be compared with the true parameters given in Table III which were used in the simulations. In this situation, the most difficult parameter to estimate appears to be the parameter B in Model II. Data was generated from Model III which is Model II with B=1. However fitting Model II to the data yields estimates for B of 2.05 and 3.82 for corn and soybeans respectively. In addition, $\lambda - t_0$ is seen to be difficult to estimate being significantly underestimated for corn.

Of additional interest is the estimation of t_0. There seems to be separation between crops based upon $\hat{t}_0$ for Model I and Model V. This is surprising since the true value for t_0 in the simulation model (Model III) was set at $t_0 = 150$ for both corn and soybeans. If Model III is a reasonably good approximation to the true growth model (and we believe it is) then differences in the estimation of

t_0 using models such as Models I and V may be due to adjustments which must be made in fitting a non-optimal model to a set of data. It seems that crop separation based upon t_0 must be viewed with caution. It is clear that if, for example, $t_0 = 165$ for corn and $t_0 = 150$ for soybeans and Model II were the appropriate model, then probably no separation between the two crops would be seen using Model V on the basis of $\hat{t}_0$.

A final observation will be made concerning the role of t_0 in Models I - V. Obviously in Models II-V, t_0 is a location parameter. As such, the shifting of each date in a set of observations by K will result in no change in the estimation of the other model parameters as long as the starting value for t_0 is also shifted by K. However, t_0 in Model I is not a location parameter, and it is of importance to understand the effect on the remaining parameters of Model I which result from this shift by K. In Table V we illustrate these results for K = -10, 0, 10, and 20. As an explanation of these results note for example that the 100 corn realizations which were analyzed by Models I-V in Table IV were again utilized here and the results for K = 0 are identical to those in Table IV. For K = -10, the 100 profile realizations remained unchanged yet the generated profile value for time t is now associated with time t-10, i.e. we have assumed that emergence date occurred 10 days earlier than $t_0 = 150$. The corresponding parameter estimates in Table V are those estimates obtained by applying Model I to this augmented data set. Note that the starting value for t_0 was also adjusted by -10. Similar proce-

TABLE V - The Effects of Shifts in t_0 on the Parameters of Model I

	t_0	$\hat{\alpha}$	$\hat{\beta}^*_c$	$\hat{Max}$	$\hat{t}_m - t_0$
CORN	150-10	16.8	2.10	44.6	62.5
	150	18.6	2.10	44.6	62.7
	150 + 10	20.4	2.10	44.6	62.9
	150 + 20	22.3	2.10	44.6	63.1
SOYBEANS	150-10	20.8	2.48	51.9	60.2
	150	22.9	2,48	51.9	60.4
	150+10	25.1	2.48	51.9	60.6
	150+20	27.3	2.47	51.8	61.4

$*\hat{\beta} = .0001\ \hat{\beta}_c$

dures resulted in the remaining entries in Table V. Note that α is the only parameter significantly effected by this shift in emergence date. However, based upon Table V we see that soybeans with emergence date of $t_0 = 140$ would be relatively indistinguishable from corn with emergence date $t_0 = 160$ on the basis of α. However, the separability associated with β and the "features" is still present.

There is a final observation that should be made concerning Table V and Model I. That is, from Table I, $\hat{\alpha}$ appears to be a monotonically increasing function of $\hat{t}_0$ and visa versa. The impact of this is that late emergence dates give significantly larger values of $\hat{\alpha}$ so that in this model $\hat{\alpha}$ is certainly not a reliable feature. The reverse is also true, i.e., $\hat{t}_0$ is a monotonically increasing function of $\hat{\alpha}$. Therefore large values of $\hat{\alpha}$ give large values of $\hat{t}_0$. This is obviously highly undesirable and as a result, one could not expect reliable estimates of t_0 from Model I.

The validity of this observation on actual data is born out by inspecting Table II. Note that $\hat{\alpha}$ is nearly a monotonically increasing function of $\hat{t}_0$. The pattern is also clear for soybeans, i.e. larger values of $\hat{t}_0$ tend to give larger values for $\hat{\alpha}$. Thus whether from a careful analysis of the actual data or the simulated date, $\hat{\alpha}$ from Model I by itself should not be considered a viable parameter for use in discriminating Corn and Soybeans. Moreover Model I should not be expected to produce reliable estimates of t_0.

Final Comments

We believe that the results in this paper provide important

information concerning both the development and the performance of various temporal profile models. It should be emphasized, however, that the results presented here are very preliminary in nature. Further investigation into the performance of these models is suggested in order to provide more experience with both real data and simulation. It is the opinion of the authors that the performers of "features" such as $\hat{m}ax$ and $t_{max}-t_0$ should be investigated further. From the discussion in the previous section, we definitely do not recommend using $\hat{\alpha}$ in Modal I. Further investigations should also consider the problem of separability by finding discriminating surfaces based upon the utilization of more than one parameter or feature.

REFERENCES

Austin, W. W. (1980) "Classification of Wheat: Badhwar Profile Similarity Technique", JSC-16826, October 1980.

Austin, W. W. (1981) "Classification of Corn: Badhwar Profile Similarity Technique". JSC-17113, August, 1981.

Badhwar, G. D. (1979) "A Semi-Automatic Technique for Multi-temporal Classification of a Given Crop" JSC-16381, July 31, 1979.

Badhwar, G. D. (1980) "Crop Emergence Date Determination from Spectral Data" Photogrammetric Engineering and Remote Sensing, Vol. 46, 369-377.

Badhwar, G. D., Carnes, J. G. and Austin, W. W. (1981). "Corn-Soybean Feature Extraction and Classification Using Landsat Data" - unpublished manuscript.

Ho, W. C. and Kashyap, R. L. (1965). "An Algorithm for Linear Inequalities and its Applications", IEEE Transaction on Electronic Computers, EC 14, 683-688.

Marquardt, D. W. (1963). "An Algorithm for Least-Squares Estimation of Nonlinear Parameters", J. Soc. Indust. Eng. 11, 431-441.

APPENDIX B

A COMPARISON OF MINIMUM DISTANCE AND MAXIMUM LIKELIHOOD TECHNIQUES FOR PROPORTION ESTIMATION

Wayne A. Woodward, William R. Schucany,
Hildegard Lindsey, and H. L. Gray
Center for Applied Mathematical and Statistical Research
Southern Methodist University

1. Introduction

A common objective in remote sensing is the estimation of the proportions $p_1, p_2, \ldots, p_m$ in the mixture density

$$f(x) = p_1 f_1(x) + p_2 f_2(x) + \ldots + p_m f_m(x) \qquad (1.1)$$

where m is the number of components(crops) in the mixture and for component $i, f_i(x)$ is a (possibly multivariate) density. In past practice this density has been assumed to be (multivariate) normal with X being the reflected energy in four bands of the light spectrum, certain linear combinations of these readings, or other derived "feature" variables. Generally the parameter estimation has been accomplished using maximum likelihood techniques. In this paper we examine the use of minimum distance estimation as an alternative to maximum likelihood and we will compare the performance of the two estimation techniques when dealing with mixtures of normal and of non-normal densities with varying amounts of separation. We will focus on the mixture of two univariate distributions given by

$$f(x) = pf_1(x) + (1-p)f_2(x) \qquad (1.2)$$

We are also assuming that only data from the mixture distribution are available. Other sampling schemes in which training samples from the component distributions are also available have been discussed by Hosmer(1973), Redner(1980), and Hall(1981) among others.

2. Estimation in the Mixture of Normals Model

In this section we will assume that $f_1(x)$ and $f_2(x)$ in (1.2) are normal densities with mean and variance μ_1, σ_1^2 and μ_2, σ_2^2 respectively where it is assumed that all five parameters $\mu_1, \sigma_1^2, \mu_2, \sigma_2^2$, and p are unknown. Techniques for estimating these parameters will be discussed.

(a) Maximum Likelihood

Several recent articles have dealt with the problem of obtaining the maximum likelihood estimates of μ_1, σ_1^2, μ_2, σ_2^2, and p (Hasselblad(1966), Day(1969), Wolfe(1970), Hosmer(1975), Fowlkes(1979), Lennington and Rassbach(1979), and Redner(1980).) Since the likelihood function

$$L = f(x_1)f(x_2) \ldots f(x_n) \tag{2.1}$$

where n is the sample size, is not a bounded function in this case (see Day(1969)), the objective in the maximum likelihood approach is to find a local maximum of L. This maximum is usually found by setting the partial derivatives of log(L) with respect to each of the 5 parameters equal to zero and solving the resulting set of equations, called the

likelihood equations. Since closed form solutions of these equations do not exist, they must be solved using iterative techniques. Hasselblad(1966) and Wolfe(1969) suggested that these equations be solved by taking advantage of their fixed point form. Redner(1980) and Redner and Walker(1982) have pointed out that this fixed point technique is essentially an application of the EM algorithm (see Dempster, Laird and Rubin(1977)) with the only difference being that using the EM algorithm, the estimates of σ_1^2 and σ_2^2 at step k involve the updated k^{th} step estimates of μ_1 and μ_2

Fowlkes(1979), on the other hand, maximized the likelihood function directly by utilizing a quasi-Newton method for minimizing -log(L) and found that good starting values were crucial for acceptable performance. Hosmer(1975) stated that using the likelihood equations, starting values were not a serious problem in his experience. In order to determine which of the two techniques seemed preferable in our simulation studies we replicated simulations performed by Fowlkes in which various sets of poor starting values were used to initiate the minimization procedure. We simulated realizations from the mixture utilized by Fowlkes and estimated the parameters using both direct maximization and the EM algorithm. The results of our simulations indicate that the EM algorithm approach is preferable and hence we have used this technique for obtaining MLEs in our simulations.

(b) Minimum Distance

Although ML estimation procedures are known to have certain optimality properties, their sensitivity to violations of the underlying assumptions is also recognized. The development of estimation procedures which perform well even under moderate deviations from assumptions has been a topic of major interest in recent literature. One of these robust procedures which has received recent attention is that of minimum distance(MD) estimation introduced by Wolfowitz(1957). Parr and Schucany(1980), for example, have shown that MD techniques provide robust estimators of the location parameter of a symmetric distribution. Minimum distance estimation has been used for parameter estimation in the mixture model by Choi and Bulgren(1968) and MacDonald(1971) with some success although, to our knowledge, the question of sensitivity to assumptions in this setting has not been addressed. These previous authors assumed that the parameters of the component distributions were known and that only the mixing proportion(s) was to be estimated.

In order to briefly describe minimum distance estimation, we let $X_1, X_2, \ldots, X_n$ denote a random sample from a population with distribution function F and let F_n denote the empirical distribution function, i.e. $F_n(x)=k/n$ where k is the number of observations less than or equal to x. Further, let $\mathcal{H} = \{H_\theta : \theta \in \Omega\}$ denote a family of distributions depending on the possibly vector valued

parameter θ. The MD estimate of θ is that value of θ for which the distance between F_n and H_θ is minimized. It is not necessary that $F \in H$. Of course, when a mixture of two normals is used as the projection family, H_θ becomes

$$H_\theta(x) = p \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\,\sigma_1} e^{-\frac{1}{2}\left(\frac{y-\mu_1}{\sigma_1}\right)^2} dy + (1-p) \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\,\sigma_2} e^{-\frac{1}{2}\left(\frac{y-\mu_2}{\sigma_2}\right)^2} dy.$$

Certain considerations become obvious at this point. First, we must define what we mean by the "distance" between two distributions. Several such distance measures have appeared in the literature. The reader is referred to the article by Parr and Schucany(1980) for a discussion of these measures. For our purposes we have chosen the Cramér-von Mises distance, W^2, between distribution functions G_1 and G_2 which is given by

$$W^2 = \int_{-\infty}^{\infty} [G_1(x) - G_2(x)]^2 dG_2(x).$$

In our setting a computing formula for the Cramér-von Mises distance between F_n and H_θ is given by

$$W_n^2 = \frac{1}{12n} + \sum_{i=1}^{n} \left[H_\theta(Y_i) - \frac{i-.5}{n}\right]^2,$$

where Y_i is the ith order statistic. The similarity between W_n^2 and the sum of squared differences between the empirical distribution function F_n and H_θ used by Choi and Bulgren(1968) should be noted.

Another consideration involves the minimization procedure to be employed in minimizing W_n^2. Parr and

Schucany used the IMSL quasi-Newton algorithm ZXMIN. Our comparisons have shown ,however, that the IMSL routine ZXSSQ which uses Marquardt's(1963) method for minimizing a sum of squares was significantly faster, usually taking no more than half the time required by ZXMIN. In the simulation studies reported in the next section we have used the Marquardt minimization procedure when calculating the MDE. It should be noted that minimization is subject to the constraints $\sigma_1^2 \geq 0$, $\sigma_2^2 \geq 0$, and $0 \leq p \leq 1$. Another finding which deserves mention before proceeding is that similar to the technique we have chosen for calculating the MLE, the MDE has the desirable property that it is relatively insensitive to starting values.

3. Starting Values

In order for the estimators discussed in the previous chapter to be used in practice, starting values for the iterative procedures must be provided. We have chosen to obtain starting values in this two component univariate setting using a partitioning technique which is very easy to implement. In the discussion to follow we will assume, without loss of generality, that $\mu_1 < \mu_2$. This technique involves first obtaining the initial estimate of p, denoted by p_0, and then estimating the remaining four parameters given p_0. Under the current implementation, only the 9 values .1,.2,...,.9 are allowed as possible

values for p_0. For each allowable value of p_0, the sample is divided into two subsamples :

$$Y_1, Y_2, \ldots, Y_{n_1}$$
$$Y_{n_1+1}, Y_{n_1+2}, \ldots, Y_n$$

where Y_i is the ith order statistic and n_1 is np_0 rounded to the nearest integer. The value for p_0 is that value of p for which $p(1-p)(m_1-m_2)^2$ is maximized, where m_j is the sample median of the jth subsample. The criterion used here is a robust counterpart to the classical cluster analysis procedure of selecting the clusters for which the within cluster sum-of-squares is minimized. It is easy to show, however, that the within cluster sum-of-squares is minimized in the two cluster case when $p(1-p)(\bar{x}_1-\bar{x}_2)^2$ is maximized, where $\bar{x}_j$ is the sample mean of cluster j and and $p=n_1/n$ with n_1 the number of sample values placed in cluster 1. Such a clustering is based upon a cut-point, c , for which all sample values below c are assigned to the cluster associated with population 1. It must be observed, however, that due to the overlap between the two mixture distributions, some sample points assigned to cluster 1 may be from population 2 and some observations from population 1 may be in cluster 2. The effect of this truncation of the right tail in population 1 is that the sample mean from cluster 1 is likely to underestimate μ_1 while μ_2 is likely to be overestimated. In addition σ_1^2 and σ_2^2 are likely to be underestimated by s_1^2 and s_2^2. If we

assume that the overlap between the two populations is not too severe, then the sample values in cluster 1 to the left of m_1 are relatively pure observations from population 1 in which case m_1 is a "good" estimate of the population mean in the case of symmetric distributions. This reasoning also indicates that m_1 and m_2 should provide better estimates of μ_1 and μ_2 than would $\bar{x}_1$ and $\bar{x}_2$. In order to estimate the variances of the component distributions we again will depend upon the fact that the values to the left of m_1 and to the right of m_2 are "pure" samples from populations 1 and 2 respectively. Thus, we will use only this portion of the data for estimation of the sample variances. We have used the fact that the semi-interquartile range of a standard normal distribution is .6745, to estimate σ_1^2 by

$$\sigma_1^2(0) = \left(\frac{m_1 - r_1^{(.25)}}{.6745} \right)^2 ,$$

where $r_j^{(q)}$ is the q^{th} percentile from the jth cluster, j=1,2. Similarly, $\sigma_{2(0)}^2 = [(r_2^{(.75)} - m_2)/.6745]^2$.

In the next section we will discuss the results of a major simulation investigation comparing ML and MD estimation. In these simulations the iterative techniques were initiated by the starting values as discussed in the previous paragraph. A preliminary simulation investigated the performance of the starting values described here. In this preliminary study we compared the convergence

initiated from these starting values with that when the iterative procedures are started at the true parameter values. The convergence from these two starts was almost always to the same parameter estimates, a result which held for both the MLE and MDE. For this reason and results to be shown in Section 4, we believe this starting value procedure to be adequate.

4. Simulation Results

In the previous two sections we have discussed ML and MD estimators for the parameters of the mixture of two distributions. In this section we report the results of simulations designed to compare these two estimators when the component distributions are normal and when they are non-normal. In addition we have made our comparisons under varying degrees of separation between the two distributions. All computations were performed on the CDC 6600 at Southern Methodist University.

In our comparison of the MDE and MLE we have begun by comparing their performance when the normality assumption is valid, i.e., when the component distributions actually are normal. We should mention that because of the optimality properties of the MLE we would expect that the MLE would be superior in this situation. Since in practice the validity of the normality assumption is subject to question, we are also very interested in the performance of the MDE and MLE when the component distributions are

not normal. To this end we have simulated mixtures in which the component distributions are distributed as a Student's t with 4 degrees of freedom. We simulated 500 samples of size n=100 from mixtures of normal and of t(4) components for each of the following parameter configurations:

Mixing proportion

.25

.50

.75

Variances

$\sigma_1^2 = \sigma_2^2$

$\sigma_1^2 = 2\sigma_2^2$

The nature of the mixture model also depends on the amount of separation between the two component distributions. While, for sufficient separation, the mixture model has a characteristic bimodal shape, Behboodian(1970) has shown, for example, that a sufficient condition for the mixture density (of two normal components) to be unimodal is that $|\mu_1-\mu_2| \leq 2\min(\sigma_1,\sigma_2)$. Of course, in this situation, parameter estimation is difficult.

For purposes of quantifying this separation between the components, we will define a measure of "overlap" between two distributions. Without loss of generality we

assume that population 1 is centered to the left of population 2. We define "overlap" to be the probability of misclassification using the rule:

Classify an observation x as:

population 1 if $x < x_c$

population 2 if $x \geq x_c$,

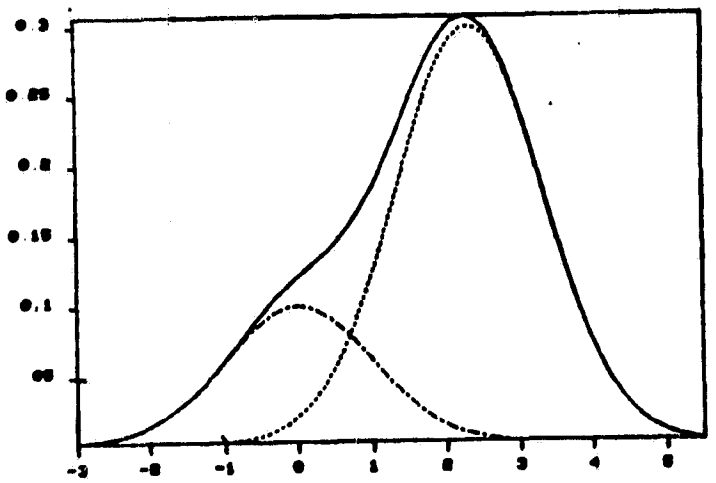
where x_c is the unique point between μ_1 and μ_2 such that

$$pf_1(x_c) = (1-p)f_2(x_c).$$

We have based our current study on "overlaps" of .03 and .10. In Figure 1 we display the mixture densities associated with normal components and $\sigma_1^2 = \sigma_2^2$. For each mixture, the scaled components $pf_1(x)$ and $(1-p)f_2(x)$ are also shown. Note that the densities for p=.75 are not displayed here since when $\sigma_1^2 = \sigma_2^2$, it follows that $f^p(x) = f^{1-p}(\mu_1+\mu_2-x)$ where $f^h(x)$ denotes the mixture density associated with a mixing proportion of h. Thus the shapes of the densities at p=.75 can be inferred from those at p=.25. Likewise, parameter estimation for p=.75 is not included in the results of the simulations when $\sigma_1^2 = \sigma_2^2$.

Although both estmation procedures provide estimates of all 5 of the parameters, only the results for the estimation of p will be shown since the mixing proportion is the parameter of primary interest. In addition, when dealing with the non-normal mixtures, the remaining parameter

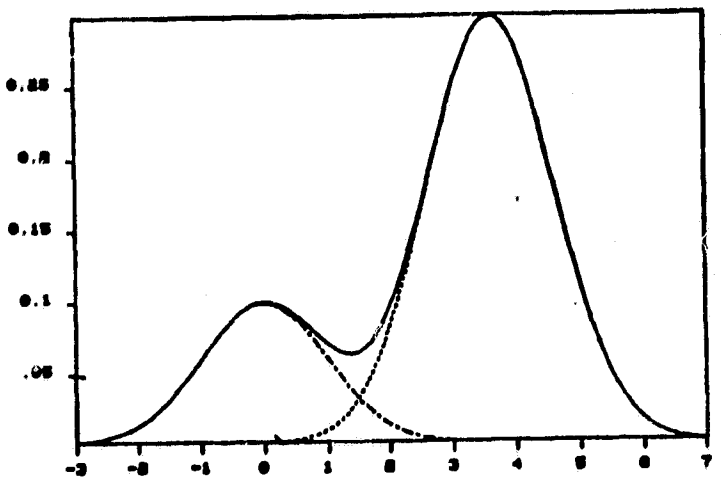
FIGURE 1 - Mixture Densities Associated with Normal Components and $\sigma_1^2 = \sigma_2^2 = 1$



p = .25

$\mu_1 = 0$, $\mu_2 = 2.32$

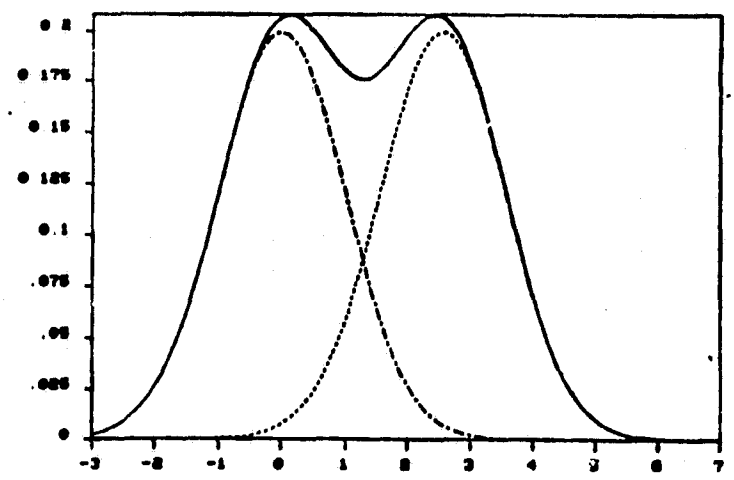
overlap = .10



p = .25

$\mu_1 = 0$, $\mu_2 = 3.6$

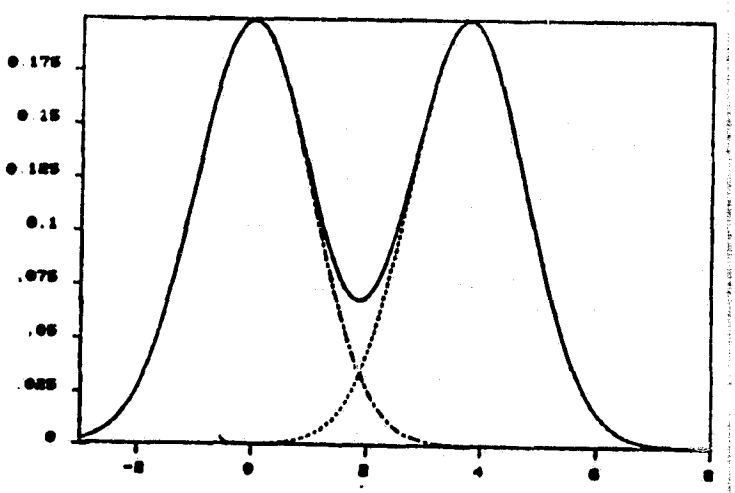
overlap = .03



p = .5

$\mu_1 = 0$, $\mu_2 = 2.56$

overlap = .10



p = .5

$\mu_1 = 0$, $\mu_2 = 3.76$

overlap = .03

——— mixture - - - - - $pf_1(x)$ ..... $(1-p)f_2(x)$

estimates often do not have a meaningful interpretation. In these simulations we have used the procedure discussed in the previous section to obtain starting values. It should be noted that although we refer to mixtures of t(4) distributions here, they are actually mixtures of distributions associated with the random variable $T'=aT+b$, where T has a t(4) distribution. These modifications are made in order to obtain the desired separation and variance ratios.

In Table 1 we show the results of the simulation comparing the performance of the MLE and MDE. In particular, let $\hat{p}_i$ denote the estimate of p for the ith sample. Then based upon the simulations, estimates of the bias and MSE are given by:

$$\hat{\text{bias}} = \frac{1}{n_s} \sum_{i=1}^{n_s} (\hat{p}_i - p)$$

$$\hat{\text{MSE}} = \frac{1}{n_s} \sum_{i=1}^{n_s} (\hat{p}_i - p)^2 ,$$

where n_s is the number of samples. It should be noted that nMSE is the quantity actually given in the table. In addition, we provide the ratio

$$E = \frac{\text{MSE(MLE)}}{\text{MSE(MDE)}}$$

as an efficiency measure.

Upon viewing the results, it can be seen, as expected, that the bias and MSE associated with the MLE were generally smaller than those for the MDE when the components were

TABLE 1

Simulation Results Comparing MLE and MDE

Sample Size = 100
Number of replications = 500

NORMAL

	Overlap = .10					Overlap = .03				
	Bias		nMSE*		E	Bias		nMSE		E
$\sigma_1^2 = \sigma_2^2$	MLE	MDE	MLE	MDE		MLE	MDE	MLE	MDE	
p = .25	.052	.125	4.26	7.80	.55	.008	.026	.54	1.09	.50
p = .50	.000	.010	3.21	3.86	.83	.000	.001	.38	.42	.90
$\sigma_1^2 = 2\sigma_2^2$										
p = .25	.002	.084	2.25	5.30	.42	.006	.027	.49	.96	.51
p = .50	-.009	.005	2.41	2.79	.86	.009	.008	.42	.44	.95
p = .75	-.086	-.137	4.87	8.36	.58	-.002	-.024	.47	1.08	.44

t(4)

	Overlap = .10					Overlap = .03				
	Bias		nMSE		E	Bias		nMSE		E
$\sigma_1^2 = \sigma_2^2$	MLE	MDE	MLE	MDE		MLE	MDE	MLE	MDE	
p = .25	.096	.104	7.35	6.18	1.19	.029	.020	.88	.44	2.00
p = .50	.015	.004	5.59	1.82	3.07	-.005	.000	.47	.27	1.74
$\sigma_1^2 = 2\sigma_2^2$										
p = .25	.061	.098	4.63	5.20	.89	.044	.029	.95	.61	1.56
p = .50	.028	.022	4.49	1.80	2.49	.010	.001	.55	.30	1.83
p = .75	-.076	-.058	7.84	3.68	2.13	-.012	-.016	.57	.36	1.58

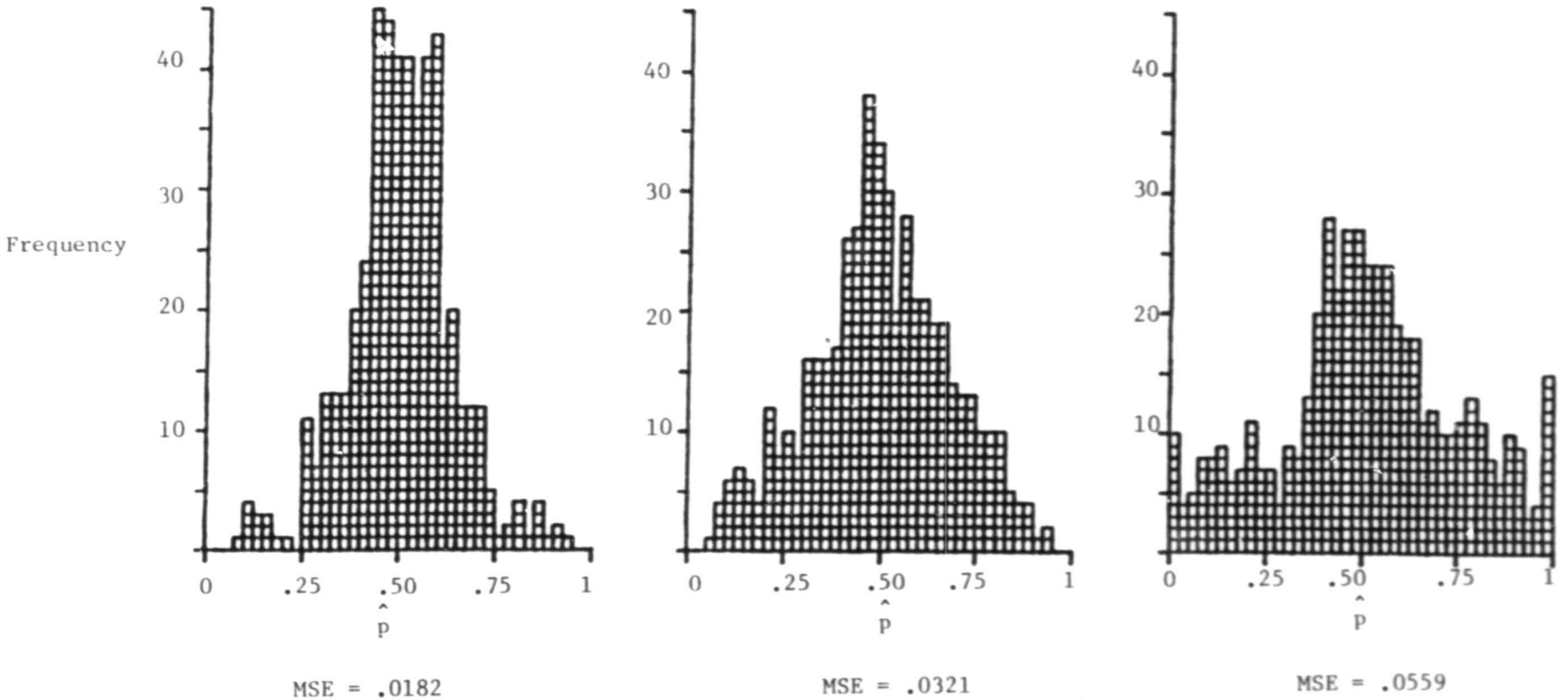
*nMSE = n times the MSE where n = sample size

normally distributed. This relationship between the estimators held for both overlaps. The MLE and MDE were quite similar at p=.5 while for p=.25 and p=.75 the superiority of the MLE is more pronounced.

For the t(4) mixtures the relationship between MDE and MLE is reversed in that the MDE generally has the smaller bias and MSE. The superiority of the MDE in this case is due in part to the heavy tails in the t(4) mixture. The MLE often interpreted an extreme observation as being the only sample value from one of the populations with all remaining observations belonging to the other. Due to the well known singularities associated with a zero variance estimate for a component distribution, Day(1969), we were concerned that the observed behavior of the MLE was due to the fact that we did not constrain the variances away from zero. However, simulation results in which equal variances were assumed (which removes the singularity) and also those which used a penalized MLE suggested by Redner(1980) were very similar to those quoted here.

Although the MSE is a widely used measure among statisticians for assessing the performance of an estimator, the practical implications, for example, of an estimator having an MSE three times larger than that for another estimator, may not be immediately apparent. Recall that each MSE quoted in Table 1 is based upon 500 estimates of p. In order to provide a better appreciation for the practical impact of differences in MSE, in Figure 2 we display histograms of the 500 estimates of p associated with three

Figure 2. Histograms of Estimates of p Based Upon 500 Samples of Size 100 from Mixtures in which p = .5



different MSEs in the table. The true value of p in each case is p= .5. It is obvious that as the MSE increases, the performance of the estimator deteriorates. Notice that the MSE for Figure 2(c) is approximately three times greater than the MSE associated with Figure 2(a), while the MSE for Figure 2(b) is aprroximately twice that for Figure 2(a). Thus, from these histograms, an intuitive feel for efficiency ratios of E=2 and E=3 can be obtained.

A very surprising result is that the starting values obtained using the procedure outlined in Section 3 produced estimators which were competitive with both the MLE and MDE. In fact, for both the normal and t(4) mixtures, the MSEs associated with the starting values were lower than those for the MDE and MLE for every parameter configuration associated with an overlap of .10. At an overlap of .03, however, the starting values estimates were generally poorer than those for the MDE and MLE.

5. Mixtures of Asymmetric Distributions

The simulation results of the previous section focus on the performance of the MLE and MDE under deviations from the assumption of normality. However, the t(4) distribution is symmetric, and recent studies have indicated that there is often a substantial asymmetry in the component distributions for variables of interest in agricultural remote sensing. A Monte Carlo examination of the performance of the MDE and MLE, assuming normal components, when in fact the component

distributions were asymmetric, was performed, and the results of this examination will be discussed in this section.

For purposes of our examination, we simulated mixtures of $\chi^2(9)$ distributions with p=.5. In these simulations the two distributions differed from each other only by a location shift. Actually the component distribution to the left is $\chi^2(9)$ while that to the right is that of a "shifted" $\chi^2(9)$ with origin no longer at 0. This shift was varied to provide overlaps of .01, .05, and .10. Since our estimation procedures involve a normality assumption, we used the means and variances of the two component $\chi^2(9)$ distributions and the true mixing proportions as our starting values. The problem of obtaining starting values from the data in this case is being examined. In Table 2 we display the results of this simulation. Only when the two component distributions were widely separated (overlap=.01) do the two procedures provide reasonable results. However, when the two chi-square distributions are not widely separated, both estimators tend to seriously underestimate p. In Figure 3 we display the three mixture distributions on which these simulations were based. We see there that it is no surprise that the estimate of p is less than .5, especially for p=.10. Both estimation procedures view this as a mixture of normals, and therefore make the reasonable interpretation that the density to the left has a smaller variance and a mixing proportion less than .5. These results point out the impact which skewed distributions can have on the proportion estimation in the

TABLE 2

Simulation Results
Mixtures of $\chi^2(9)$ Components

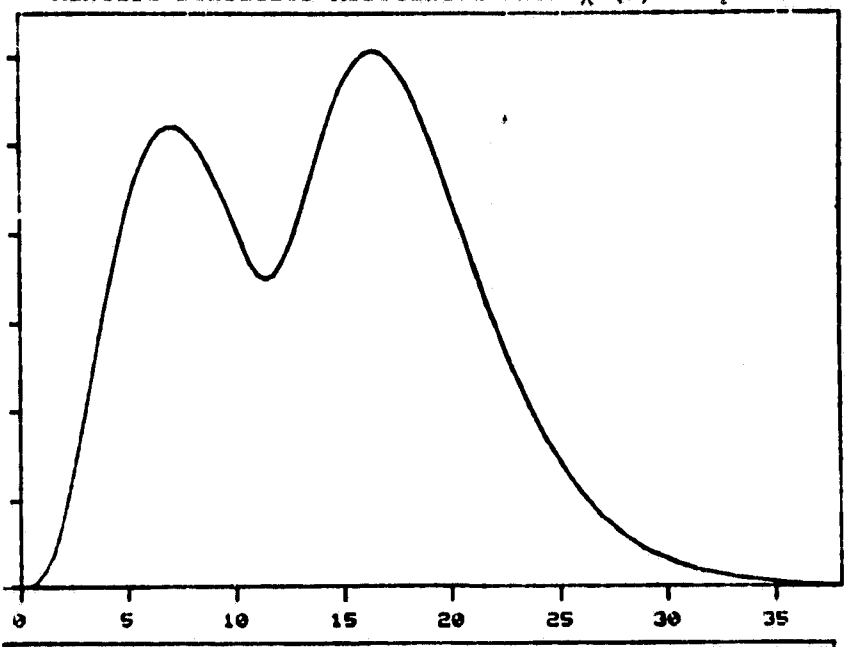
Sample Size = 100
Number of replications = 200
p = .5

		MLE			MDE		
		$\hat{p}$	Bias	nMSE	$\hat{p}$	Bias	nMSE
Overlap	.10	.28	-.22	6.8	.28	-.22	6.6
	.05	.35	-.15	2.7	.37	-.13	2.3
	.01	.47	-.03	.4	.45	-.05	.5

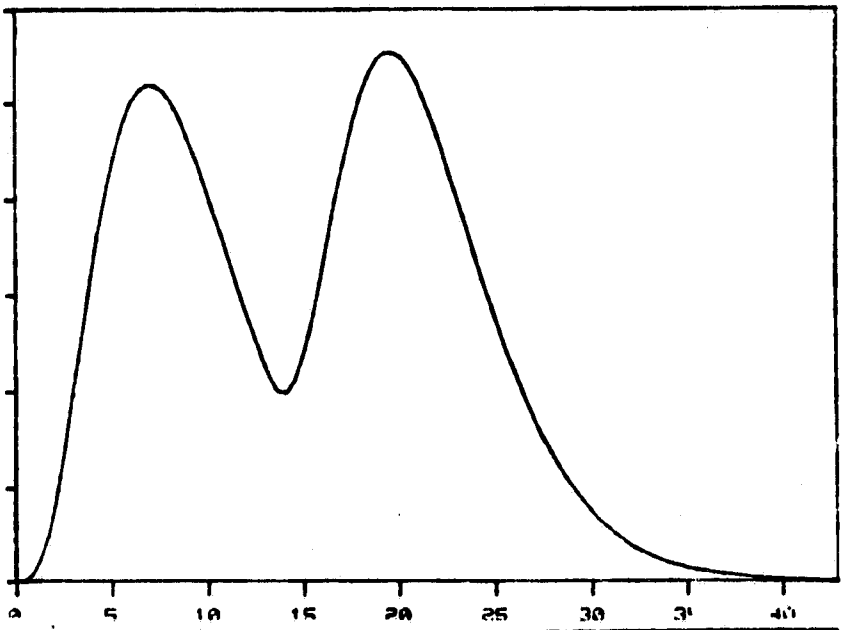
FIGURE 3

Mixture Densities Associated with $\chi^2(9)$ Components

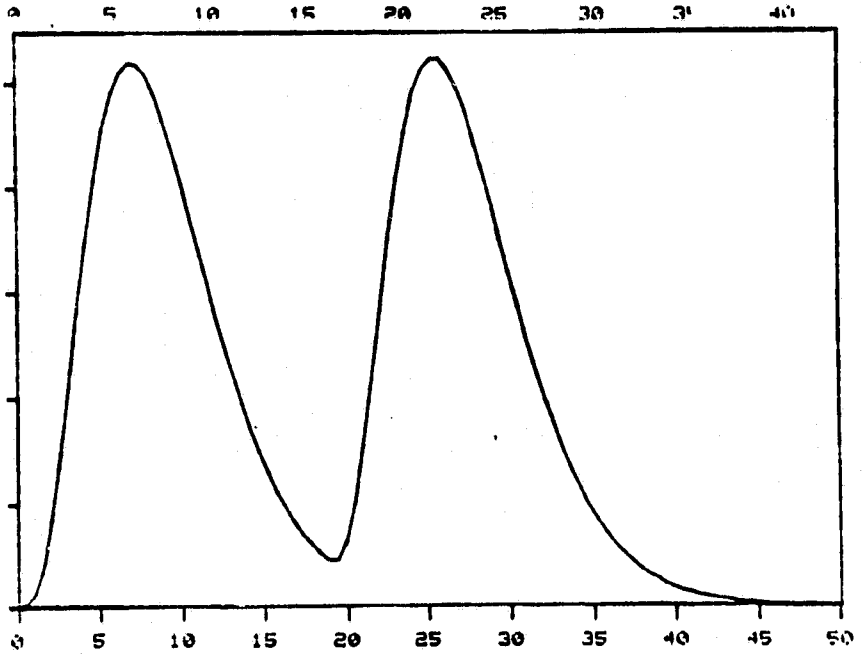
Overlap .10



Overlap .05



Overlap .01



mixture model when normal mixtures are assumed.

Current investigation into this area centers around modifying the estimation procedures by assuming that the underlying component distributions belong to some family of distributions whose members can be either symmetric or asymmetric depending on parameter configurations. At the present time, the Weibull distribution is being examined concerning its usefulness.

6. Concluding Results

We believe that the results of the preceding sections are of sufficient substance to motivate further research in the area of MD estimation in the mixture model. Our results indicate that the MDE is indeed more robust than the MLE in the sense that it is less sensitive to symmetric departures from the underlying assumption of normality of component distributions. Several areas for future investigation have already been identified in addition to the asymmetric components problem discussed in Section 5.

First, simulations similar to the ones presented here should be performed without the assumption of only two populations in the mixture. The performance of the MDE and MLE should be compared when the number of populations is known and larger than two. In addition the applicability of the MDE to the problem of estimating the number of populations also warrants investigation. We plan to examine these possibilities.

Second, the problem of applying the MDE to the multivariate setting is of interest. Preliminary indications are that such an extension will be possible.

Third, the choice of distance measure in the MDE is a topic of interest. Our results are not meant to imply that W^2 is optimal.

Finally, the MDE and MLE must ultimately be compared on real data. Several related practical considerations have not yet been investigated. For example, when applying these estimators to LANDSAT data, the number of iterations allowed must be small due to time constraints. In the simulations described here, these constraints were not imposed and iteration was allowed to continue until convergence was obtained. The performance of the MDE and MLE under convergence restrictions should be examined.

REFERENCES

1. Behboodian, J.(1970). "On the Modes of a Mixture of Two Normal Distributions," Technometrics 12, 131-139.
2. Choi,K. and Bulgren,W. G.(1968). "An Estimation Procedure for Mixtures of Distributions," JRSS-B 30, 444-460.
3. Cohen, A.C.(1967)."Estimation in Mixtures of Two Normal Distributions," Technometrics 9, 15-28.
4. Day, N. E.(1969). "Estimating the Components of a

Mixture of Normal Distributions," Biometrika 56, 463-474.

5. Dempster, A.P., Laird, N.M., and Rubin, D.B.(1977) "Maximum Likelihood from Incomplete Data via the EM Algorithm," JRSS-B 39, 1-38.
6. Everitt, B.S. and Hand, D.J.(1981). Finite Mixture Distributions, Chapman and Hall, London.
7. Fowlkes, E. B.(1979)."Some Methods for Studying the Mixture of two Normal (Lognormal) Distributions," JASA 74, 561-575.
8. Hall,P. (1981). "On the Nonparametric estimation of Mixture Proportions," JRSS-B 43 ,147-156.
9. Hasselblad, V. A.(1966)."Estimation of Parameters for a Mixture of Normal Distributions," Technometrics 8, 431-446.
10. Hosmer, D. W.(1973)."A Comparison of Iterative Maximum Likelihood Estimates of the Parameters of Two Normal Distributions Under Three Different Types of Samples," Biometrika 29, 761-770.
11. Lennington, R. K. and Rassbach, M. E.(1979). "CLASSY-An Adaptive Maximum Likelihood Clustering Algorithm," Proc. Tech. Sessions, The LACIE Symposium, vol. II, 671-690.
12. MacDonald, P.D.M.(1971). "Comment on 'An Estimation Procedure for Mixtures of Distributions' by Choi and Bulgren," JRSS-B 33, 326-329.
13. Parr, W. C. and Schucany W. R.(1980)."Minimum Distance and Robust Estimation,"JASA 75, 616-624.

14. Pearson,K.(1894)."Contribution to the Mathematical Theory of Evolution," Phil trans A 185, 71-110.

15. Redner, R.A.(1980)."Maximum Likelihood Estimation for Mixture Models," JSC-16832, October 1980.

16. Redner, R.A. and Walker, H.F.(1982). "Mixture densities, Maximum Likelihood, and the EM Algorithm," unpublished manuscript.

17. Wolfe, J.H.(1980). "Pattern Clustering by Multivariate Mixture Analysis," Multivariate Behavioral Research 5. 320-350.

18. Wolfowitz, J.(1957)."The Minimum Distance Method," Annals of Math. Statist. 28, 75-88.

APPENDIX C

PROPORTION ESTIMATION IN MIXTURES OF ASYMMETRIC DISTRIBUTIONS

Wayne A. Woodward, Richard F. Gunst,
Hildegard Lindsey, and H. L. Gray

Center for Applied Mathematical and Statistical Research
Southern Methodist University

1. Introduction

A standard approach to the estimation of crop proportions in agricultural remote sensing has been to estimate the proportions $p_1, p_2, \ldots, p_m$ in the mixture density

$$f(x) = p_1 f_1(x) + p_2 f_2(x) + \ldots + p_m f_m(x) \qquad (1.1)$$

where m is the number of components(crops) in the mixture and $f_i(x)$ is the density associated with component i. The usual procedure for estimating the parameters in the mixture model of (1.1) has been to:

(a) assume that the component distributions are normal

(b) use maximum likelihood estimation.

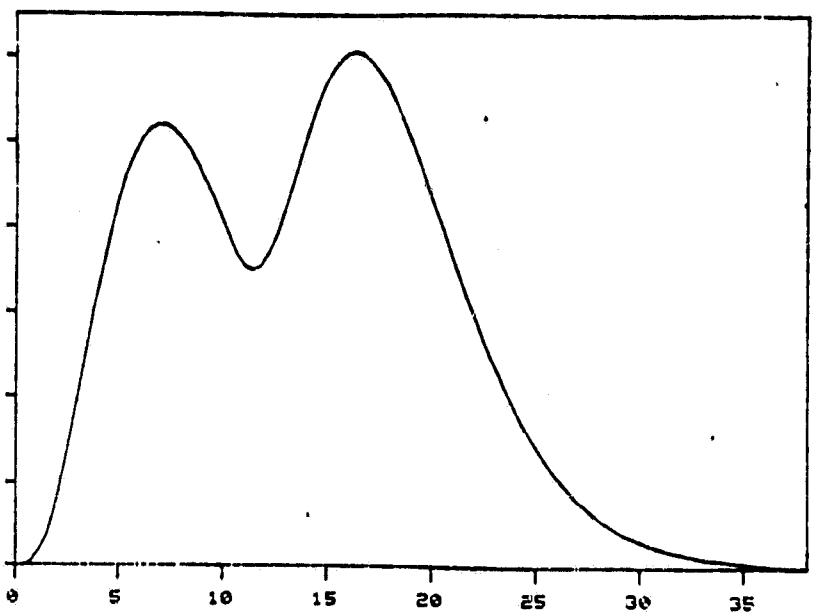
The variable X has usually been taken to be the reflected energy in the four LANDSAT bands or some linear combination of these such as greenness or brightness. Recent efforts have focused on the use of certain derived features from growth models such as g_{max} and t_{max} as variables in the mixture model. Studies have indicated that there is often a

substantial asymmetry in the distributions of these features for a given crop. Woodward et. al.(1982) have shown that asymmetry in the component distributions can cause a substantial bias in the proportion estimators when the mixture of normals model is assumed. As an example, in Figure 1 we display the mixture density associated with the mixture of two distributions. Examination of the figure reveals that if we assume that the component distributions are symmetric, then we must conclude that $p_1<p_2$ and that the component to the right has larger variance. Actually, in this mixture $p_1=p_2$ and the distribution to the left in this mixture is a $\chi^2(9)$ while the component to the right is a "shifted"$\chi^2(9)$, i.e. its left truncation point is at x=10 instead of x=0. We see that a bias will be introduced in estimating mixing proportions in this mixture if the component distributions are assummed to be symmetric, which of course is the case when the components are assumed to be normal.

In this paper we will discuss techniques for estimating the crop proportions in the presence of asymmetric component distributions. In particular the estimation procedures we will propose assume that the underlying component distributions belong to some family of distributions whose members can be either symmetric or skewed depending on parameter configurations. At the present time, the Weibull distribution is being examined concerning its usefulness in this area. The effectiveness of this technique will be

FIGURE 1

A Mixture Density



examined through simulations.

2. The Weibull Distribution

The Weibull distribution is named after the Swedish physicist Waloddi Weibull who used it to represent the distribution of the breaking strength of materials (Weibull(1939)). The distribution has been widely used in recent years in the fields of reliability and quality control. Its popularity is largely due to the flexibility which it introduces into the model due to the fact that it can be used to describe distributions which are symmetric or skewed in either direction. For these reasons we have chosen to investigate its applicability to estimation in mixtures of asymmetric components. The three-parameter Weibull density can be expressed as

$$f(x) = \frac{\gamma}{\beta}\left(\frac{x-\alpha}{\beta}\right)^{\gamma-1} e^{-\left(\frac{x-\alpha}{\beta}\right)^{\gamma}}, \quad x \geq \alpha, \quad \beta,\gamma > 0 \tag{2.1}$$

We will use the notation $X \sim W(a,b,c)$ to indicate that the random variable X has a three-parameter Weibull distribution with parameters $\alpha=a$, $\beta=b$, and $\gamma=c$. The parameter α locates the left truncation point and β serves as a scale parameter while γ determines the shape of the distribution. In Figure 2 we show Weibull densities for a fixed α and β and a range of values for γ. From the figure it is clear that the shape can vary dramatically as γ changes. In Figure 3 the

FIGURE 2

Weibull Densities with $\alpha = 0$, $\beta = 1$,

and Various Values for γ

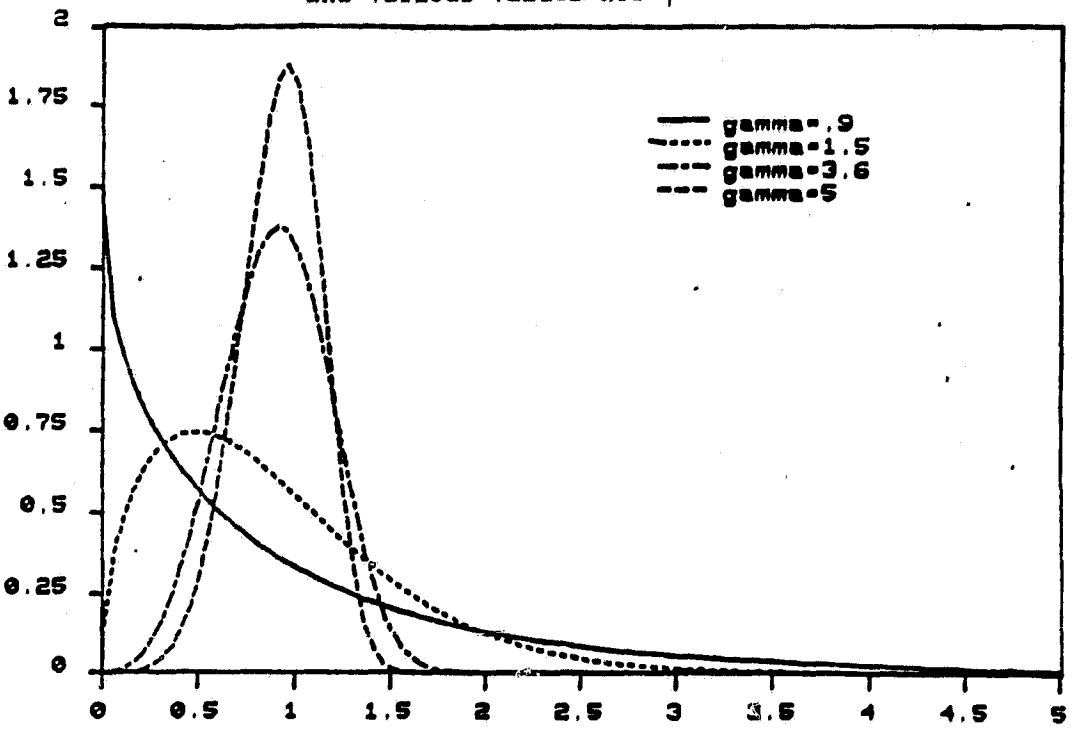
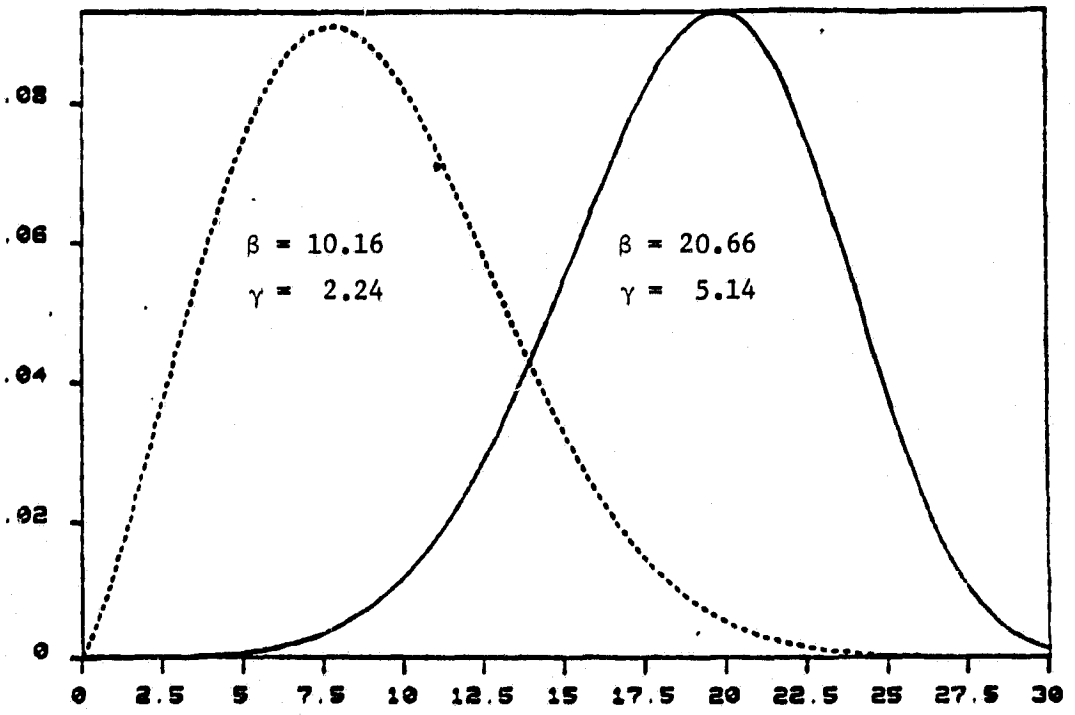


FIGURE 3

Weibull Densities with $\alpha=0$



fact that the Weibull density can be skewed to the left as well as to the right is more clearly demonstrated. For $\gamma=3.60232$ aprroximately, the standardized skewness parameter $\beta_1 = \frac{\mu_3}{\mu_1^{3/2}}$, where μ_i is the ith central moment, is zero indicating symmetry. If $\gamma<3.60232$ then the Weibull is skewed to the right, while if $\gamma>3.60232$ it is skewed to the left. The Weibull distribution is unimodal, and if $\gamma \geq 1$ the mode occurs at

$$x_m = \alpha + \beta\left(\frac{\gamma-1}{\gamma}\right)^{1/\gamma} .$$

Otherwise, when $0<\gamma<1$, the mode occurs at $x_m=\alpha$.

Dubey(1967) has studied the Weibull distribution when $\gamma=3.60232$ and has concluded that it is very similar to the normal. In particular, Dubey has shown that

$$\sup_{-3\leq v\leq 3} |F_Z(v) - F_Y(v)| = .0078 \qquad (2.3)$$

where F_Z denotes the cumulative distribution function of the random variable $Z \sim N(0,1)$ and Y is the standardized variate $Y=(X-\mu)/\sigma$ where where μ and σ^2 are the mean and variance of the Weibull variate X.

It should be noted that the Weibull distribution is often given in the literature in two parameter form in which α is assumed to be known (and usually 0). However, unless otherwise specified, when we refer to the Weibull distribution, we will be referring to the three-parameter form specified by (2.1).

The cumulative distribution function corresponding to

the three-parameter Weibull is given by the closed form expression

$$F_X(x) = 1-e^{-(\frac{x-\alpha}{\beta})^\gamma} \tag{2.4}$$

while the noncentral moments are given by

$$u_r' = \sum_{k=0}^{r} \binom{r}{k} \alpha^{r-k}\beta^k \Gamma(\frac{k}{\gamma} + 1) \tag{2.5}$$

From (2.5) it can be seen that

$$\mu = \alpha + \beta\Gamma(\frac{1}{\gamma} + 1)$$
$$\sigma^2 = \beta^2\{\Gamma(\frac{2}{\gamma} + 1) - \Gamma^2(\frac{1}{\gamma} + 1)\} . \tag{2.6}$$

The first three moments of the Weibull distribution determine the values of α, β, and γ. The method of moment estimators can be obtained using these relationships, but unfortunately the estimators do not exist in a closed form. The log-likelihood function for a random sample of n observations from the Weibull distribution is

$$\ell n(L) = n\ell n\gamma - n\gamma\ell n\beta + (\gamma-1)\sum_{i=1}^{n} \ell n(X_i-\alpha) - \frac{1}{\beta^\gamma}\sum_{i=1}^{n} (X_i-\alpha)^\gamma \tag{2.7}$$

Differentiating ln(L) yields the following likelihood equations

$$-(\gamma-1)\sum_{i=1}^{n} (X_i-\alpha)^{-1} + \frac{\gamma}{\beta^\gamma}\sum_{i=1}^{n} (X_i-\alpha)^{\gamma-1} = 0 \tag{2.8}$$

$$\beta = [\frac{1}{n}\sum_{i=1}^{n} (X_i-\alpha)^\gamma]^{1/\gamma}$$

$$\gamma = \{\sum_{i=1}^{n} [\ell n(\frac{X_i-\alpha}{\beta})][(\frac{X_i-\alpha}{\beta})^\gamma - 1]\}^{-1}$$

Let $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\gamma}$ denote the estimators obtained from the simultaneous solution of equations (2.8) to (2.10). If $0<\hat{\alpha}\leq Y_1$, where Y_i denotes the ith order statistic, these estimators are the maximum likelihood(ML) estimators for the three Weibull parameters. However, due to the restriction $x\geq\alpha$ in (2.1), if $\hat{\alpha}>Y_1$, then the MLE of α is taken to be Y_1 and β and α are estimated from (2.9) and (2.10). As in the case of method of moment estimators, the ML estimators do not have a closed form expression. For a general review of the literature on Weibull parameter estimation see Johnson and Kotz(1970).

3. Mixtures of Weibull Distributions

In order to examine the feasibility of using the Weibull as a model for the component distributions in the mixture model of (1.1), we will investigate the estimation of the parameters in the mixture of two Weibull distributions. This mixture density is given in (3.1)

$$f(x) = p\,\frac{\gamma_1}{\beta_1}\left(\frac{X-\alpha_1}{\beta_1}\right)^{\gamma_1-1} e^{-\left(\frac{X-\alpha_1}{\beta_1}\right)^{\gamma_1}} + (1-p)\,\frac{\gamma_2}{\beta_2}\left(\frac{X-\alpha_2}{\beta_2}\right)^{\gamma_2-1} e^{-\left(\frac{X-\alpha_2}{\beta_2}\right)^{\gamma_2}} \qquad (3.1)$$

where the 7 parameters p, α_1, β_1, γ_1, α_2, β_2, and γ_2 are assumed to be unknown.

Previous research in this area includes that of Kao(1959), who proposed a graphical procedure for estimating the parameters in (3.1) when one of the location parameters is assumed to be known and equal to zero. The estimation of

the 6 remaining parameters is accomplished using a graphical procedure whose applicability to our problem seems to be limited although some of his estimation rules could be automated. Rider(1961) and Falls(1970) propose estimating the parameters of a mixture of two-parameter Weibulls using the method of moments. Falls' procedure involves estimating the mixing proportion p using a graphical procedure similar to that of Kao.

Maximum likelihood estimation of the parameters of (3.1) has been discussed by Looney and Bargmann(1982). The likelihood equations obtained by differentiating the log-likelihood function ln(L)

$$\ell n(L) = \sum_{i=1}^{n} \{\ell n[pf_1(X_i) + (1-p)f_2(X_i)]\}$$

with respect to each of the 7 parameters yields the likelihood equations

$$(\gamma_j-1)\sum_{i=1}^{n} f(j|X_i)(X_i-\alpha_j)^{-1} - \frac{\gamma_j}{\beta_j^{\gamma_j}}\sum_{i=1}^{n} f(j|X_i)(X_i-\alpha_j)^{\gamma_j-1}=0, j=1,2 \qquad (3.2)$$

$$\beta_j=\{[\sum_{i=1}^{n}(X_i-\alpha_j)^{\gamma_j} f(j|X_i)]/\sum_{i=1}^{n} f(j|X_i)\}^{1/\gamma_j},\ j=1,2 \qquad (3.3)$$

$$\gamma_j=\{[\sum_{i=1}^{n}((\frac{X_i-\alpha_j}{\beta_j})^{\gamma_j}-1)\ell n(\frac{X_i-\alpha_j}{\beta_j})]/\sum_{i=1}^{n} f(j|X_i)\}^{-1}\ j=1,2 \qquad (3.4)$$

$$p = \frac{1}{n}\sum_{i=1}^{n} f(1|X_i) \qquad (3.5)$$

where $f(i|x) = p_i f_i(x)/f(x)$ with $f_i(x)$ denoting the ith component density and f(x) the mixture density. Solving this set of equations for the maximum likelihood estimators is difficult due largely to equations (3.2) which are not in fixed point form. Looney and Bargmann(1982) suggested a

procedure in which the shape parameters γ_1 and γ_2 are fixed independently at each of the values

$$\{\tfrac{1}{5}, \tfrac{1}{4}, \tfrac{1}{3}, \tfrac{1}{2}, \tfrac{2}{3}, 1, \tfrac{3}{2}, 2, 3, 4, 5\}$$

and, for each of the $(\hat{\gamma}_1, \hat{\gamma}_2)$ pairs, "preliminary" maximum likelihood estimates of the remaining 5 parameters are found. A search procedure results in selecting the $(\hat{\gamma}_1, \hat{\gamma}_2)$ pair for which ln(L) is maximized. With $\hat{\gamma}_1$ and $\hat{\gamma}_2$ fixed at these values, maximum likelihood estimation for the remaining 5 parameters is then carried through to convergence. The Looney and Bargmann procedure for solving the system of equations (3.2) - (3.5) seems overly restrictive with respect to the selection of possible values of the shape parameter, while expansion of the search procedure to allow for more shape parameter values would probably be prohibitive because of time constraints. However, solution of these likelihood equations directly appears to us to be quite untractable. For these reasons, we have investigated the use of minimum distance(MD) estimation, first introduced by Wolfowitz(1957), for estimating the 7 parameters in the mixture of Weibulls model given in (3.1). Woodward et. al.(1982) have recently studied the use of MD estimation in the mixture of normals model. These authors showed that MD estimation was easy to implement in that setting, and that MD estimators showed to be superior to ML estimators under departures from component normality. Since our use of Weibull components is due to the

flexibility which it introduces into the model rather than underlying theoretical justifications, we definitely need an estimation procedure which is robust to departures from assumptions.

The minimum distance estimator of the parameter θ (possibly vector valued) is defined to be that value of θ which minimizes the distance between H_θ and F_n where $H=\{H_\theta:\theta\in\Omega\}$ denotes a family of distributions depending on θ and F_n denotes the empirical distribution function, i.e. $F_n(x)=k/n$ where k is the number of observations less than or equal to x. The family of distributions H is referred to as the projection model, where θ in this case $\theta=(p,\alpha_1,\beta_1,\gamma_1,\alpha_2,\beta_2,\gamma_2)$, and $H_\theta(x)$ is the distribution function associated with a mixture of two Weibull components given by

$$H_\theta(x) = p[1-e^{-(\frac{X-\alpha_1}{\beta_1})^{\gamma_1}}] + (1-p)[1-e^{-(\frac{X-\alpha_2}{\beta_2})^{\gamma_2}}] . \qquad (3.6)$$

Note that in contrast to the situation in which the projection model is taken to be the mixture of two normals, $H_\theta(x)$ in (3.6) has a closed form expression. The choice of distance function to be used to measure the distance between two distributions is a topic of current interest in the field of MD estimation. Woodward et. al.(1982) used the Cramer-von Mises distance, W^2, given by

$$W^2 = \int_{-\infty}^{\infty}[G_1(x)-G_2(x)]^2 dG_2(x) \qquad (3.7)$$

where G_1 and G_2 are two distribution functions, and we have

chosen to use this distance measure in the current study. The distance between a distribution function H_θ and the empirical distribution function F_n, which is needed for calculation of the MD estimator, is given by the simplified expression

$$W_n^2 = \frac{1}{12n} + \sum_{i=1}^{n} [H_\theta(Y_i) - \frac{i-.5}{n}]^2 , \qquad (3.8)$$

where Y_i denotes the ith order statistic. Since $H_\theta(X)$ exists in closed form, the MD estimator in this case is easily obtained by using nonlinear least squares techniques to minimize (3.8). We have chosen to perform this minimization by using Marquardt's(1963) procedure.

4. Simulation Results

In Section 3 we discussed the problem of estimation in the mixture of Weibulls model. From that discussion it appears that the minimum distance techniques are preferable for estimating the parameters in a mixture of three parameter Weibulls, especially in terms of computational convenience. In this section we will discuss the results of an initial computer simulation which was designed for use in evaluating the numerical capabilities of this method. All computations were performed on the CDC 6600 at Southern Methodist University. In this section we will evaluate the performance of the MD estimation procedures discussed. Since the usual procedure is to assume that the components are normal, we will compare the Weibull based MDEs with the

normal based procedures. We have generated samples from mixtures of normal components and mixtures of $\chi^2(9)$ components. Obviously, we would expect the normal based procedures to perform better than Weibull based procedures when the mixture really is a mixture of normal components. However, if the Weibull techniques are to be useful, then they must give reasonable results in this situation since the normal assumption does appear to be a reasonable assumption in some cases. Since the Weibull with $\gamma=3.6$ is very nearly normal, there is reason to believe that Weibull procedures will perform well in this situation. We have not simulated samples from mixtures of Weibull distributions, but we plan to consider this in the future. Of course, as mentioned in the previous section, we are most interested in the performance of the Weibull based procedures when the underlying components from which we sample are not necessarily Weibulls, but are realistic representativs of the types of component distributions we see in practice.

Our simulation results are based on 200 samples of size n=200 from mixtures of normal and of $\chi^2(9)$ components. In each mixture, the variance associated with the two components are equal. In fact, the two component distributions differ from each other only by a location shift. We have simulated from mixtures having mixing proportions of .25, .50, and .75. We have simulated from mixtures with varying degrees of separation between the two component distributions. Overlap as defined by Woodward

et.al.(1982) is a quantification of this separation. It is defined as the the probability of misclassification using the rule:

Classify an observation x as:

population 1 if $x < x_c$

population 2 if $x \geq x_c$

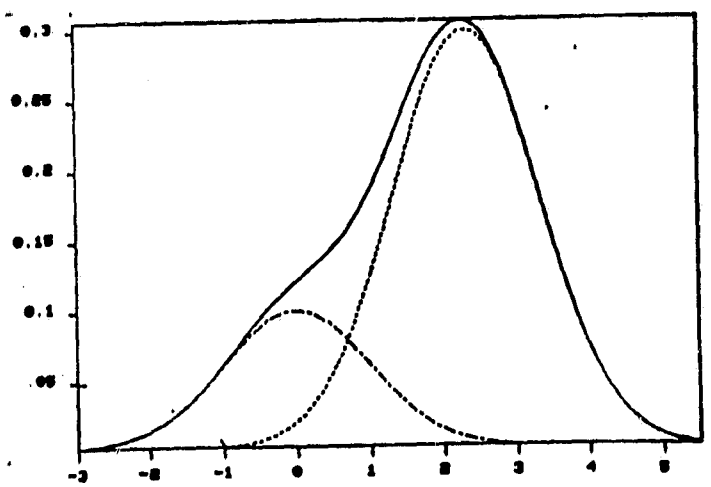
where without loss of generality, population 1 is assumed to be centered to the left of population, and where x_c is the unique point between μ_1 and μ_2 such that

$$pf_1(x_c) = (1-p)f_2(x_c) \ .$$

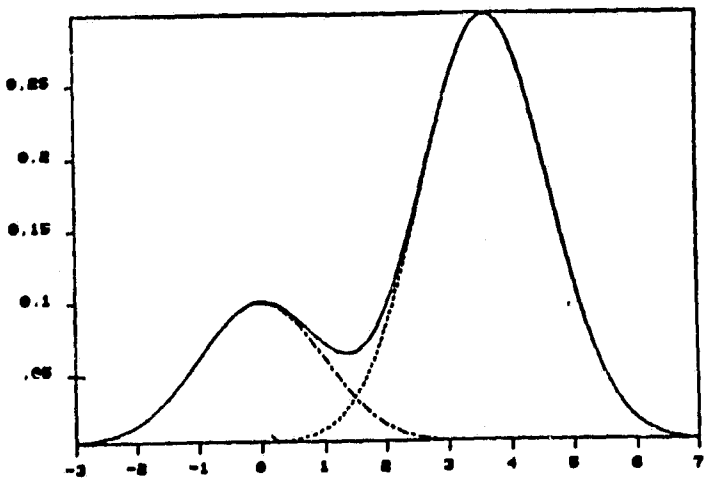
We have based our current study on "overlaps" of .03 and .10. In Figure 4 we display the mixture densities associated with normal components. For each mixture, the scaled components $pf_1(x)$ and $(1-p)f_2(x)$ are also shown. Note that the densities for p=.75 are not displayed here. Since $\sigma_1=\sigma_2$, it follows that $f^p(x)=f^{1-p}(\mu_1+\mu_2-x)$ where $f^p(x)$ denotes the mixture density associated with a mixing proportion of p. Thus the shapes of the densities at p=.75 can be inferred from those at p=.25. Likewise, parameter estimation for p=.75 is not included in the results of the simulations for the mixtures of normals. In Figure 5 we display the mixture densities associated with the mixtures of $\chi^2(9)$ components. Note that although we refer to a mixture of $\chi^2(9)$

FIGURE 4

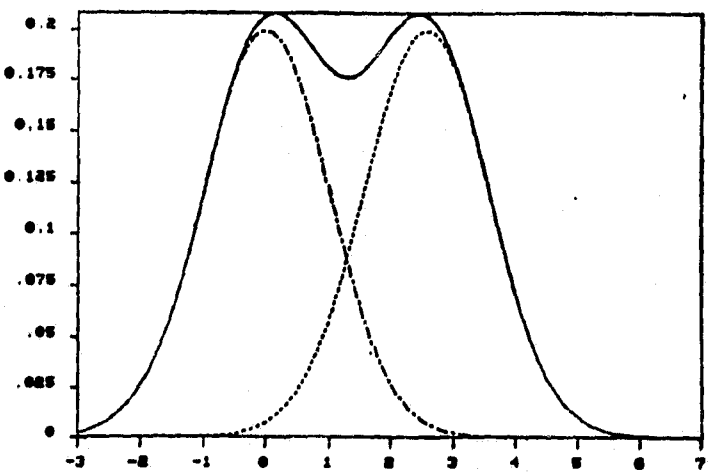
Mixture Densities with Normal Components



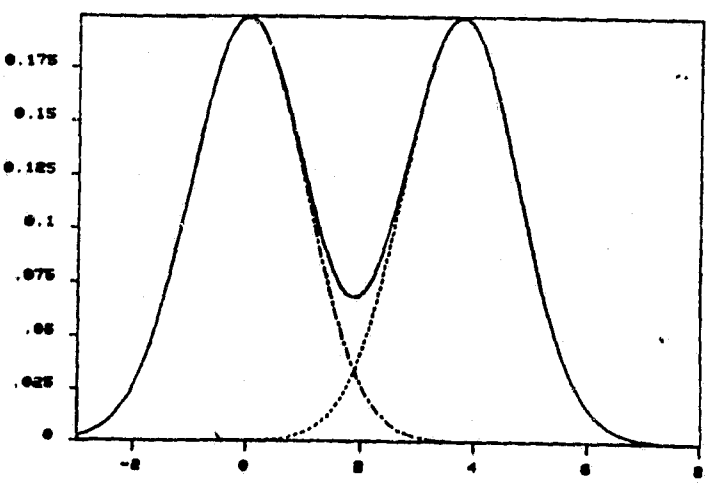
(a) p = .25, Overlap = .10



(b) p = .25, Overlap = .03



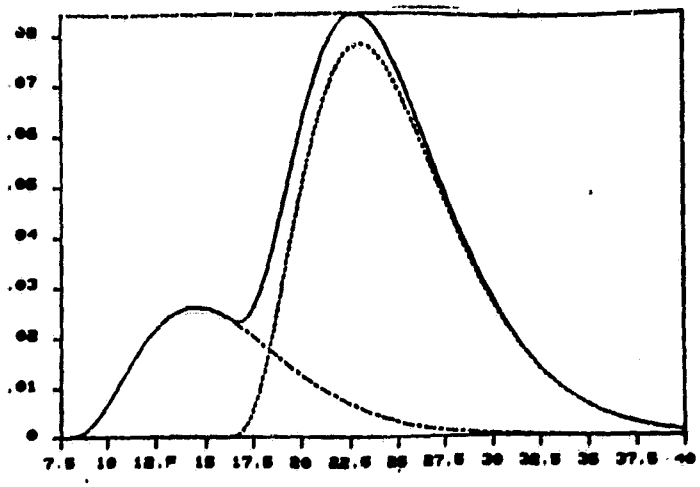
(c) p = .50, Overlap = .10



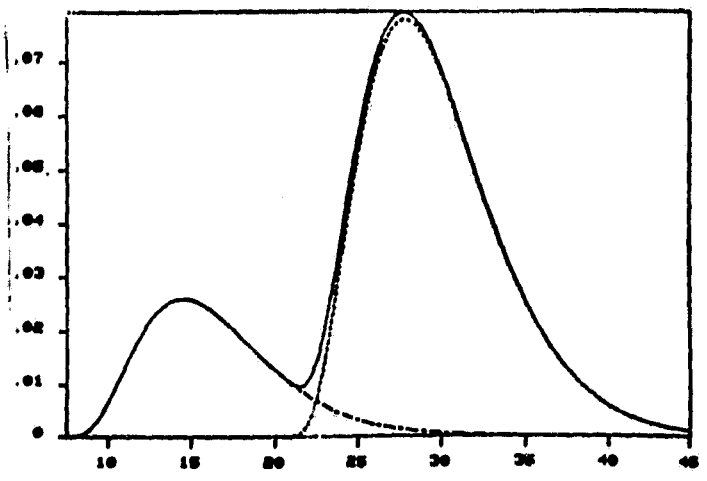
(d) p = .50, Overlap = .03

FIGURE 5

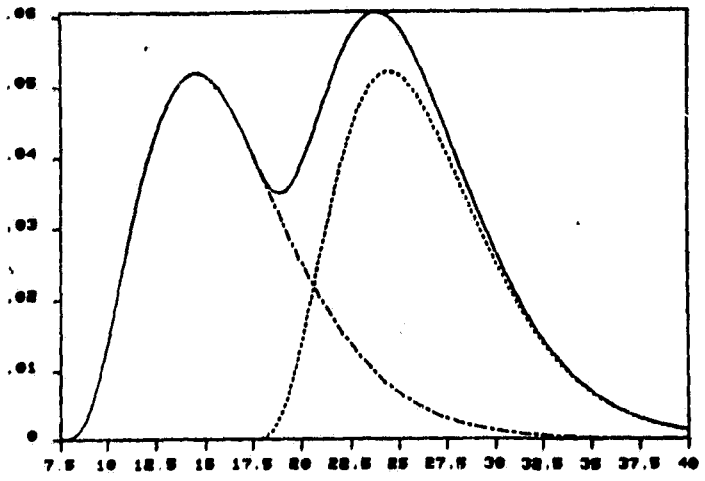
Mixtures Densities with $\chi^2(9)$ Components



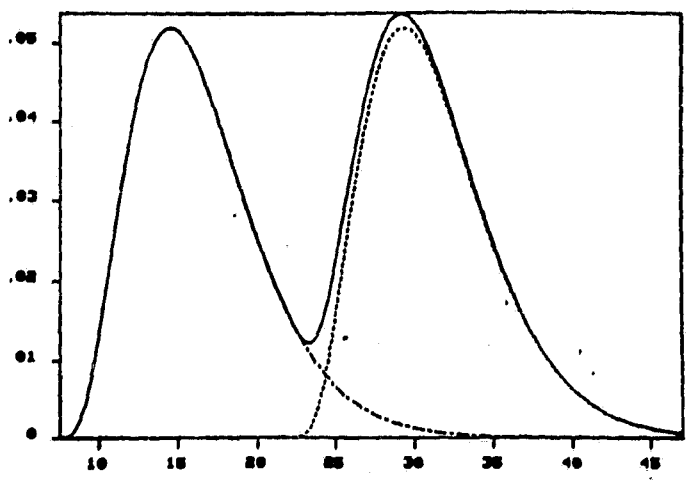
(a) p = .25, Overlap = .10



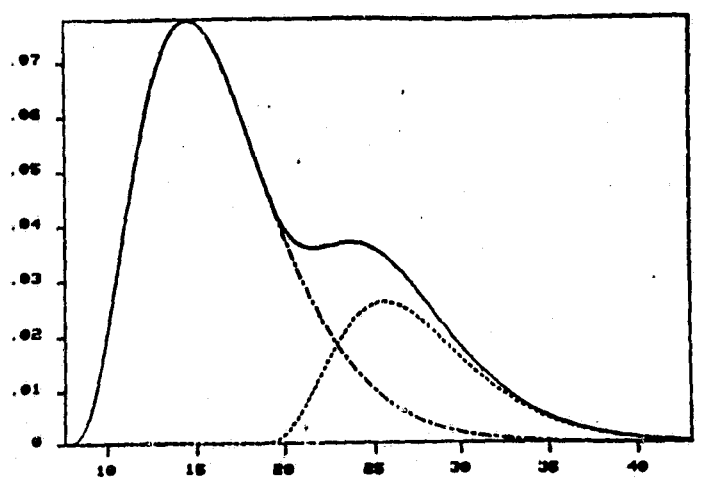
(b) p = .25, Overlap = .03



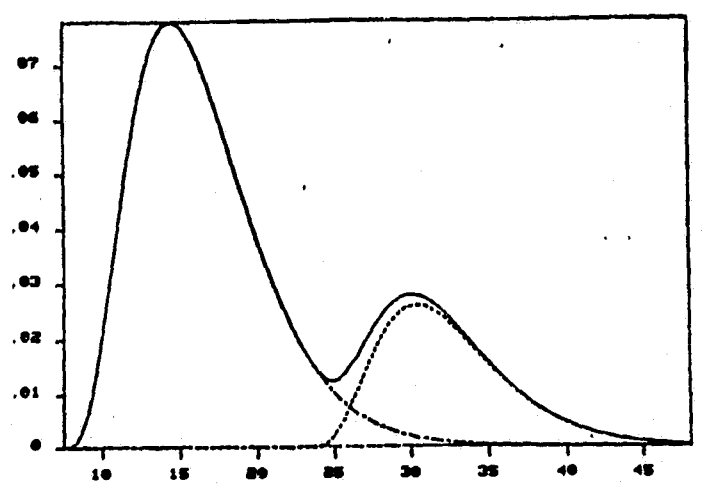
(c) p = .50, Overlap = .10



(d) p = .50, Overlap = .03



(e) p = .75, Overlap = .10



(f) p = .75, Overlap = .03

distributions here, they are actually "shifted" chi-squares, i.e. the left truncation points are different from zero.

For each of the simulated samples, three sets of parameter estimates were obtained:

(1) ML estimates based on mixture of normals model (MLEN)

(2) MD estimates based on mixture of normals model (MDEN)

(3) MD estimates based on mixture of Weibulls model(MDEW)

Although the MLEN and MDEN provide estimates of all 5 of the parameters of the mixture of normals model, and the MDEW produces estimates for all 7 parameters in the mixture of Weibulls model, only the results for the estimation of p will be shown. The mixing proportion is the parameter of primary interest, and when dealing with the "wrong-model" situations, the remaining parameter estimates often do not have a meaningful interpretation. For purposes of aiding in the discussions which follow, we will call a component model from which we actually simulated, a "simulation component model", while a component model which is assumed under a particular estimation procedure will be called an "estimation component model". Thus, a "wrong-model" situation is one in which the simulation component models are not the same as the estimation component models.

In the "correct-model" situations, i.e. using the MLEN or MDEN to estimate the parameters of a simulated mixture of normal components, the true parameter values are used as starting values for the iterative estimation procedures. In all of the other cases, there is not a "true" set of

parameters. For starting values, we have used the "true" mixing proportion, and then estimated the parameters of each component separately using a method of moments procedure. Consider a situation in which the estimation components are normal. We obtain starting values for each component by equating the first and second moments of the corresponding simulation and estimation components and using these to obtain μ_i and σ_i^2 for the normal estimation component. When the estimation components are Weibull, we have taken the approach of setting the starting value for γ at $\gamma=3.6$ for each component. Then the first two moments of the corresponding simulation and estimation components are equated to yield starting value estimates for the other two parameters. We believe that this provides a "neutral" start. If the final estimates reflect the finding of substantial skewness for one or both of the component Weibulls, this will be because of the data and not because of "skewed" starting values.

The normal component models were generated with $\mu_1=7.5$, $\sigma_1^2=\sigma_2^2=1$, and μ_2 positioned so that the desired overlap is obtained. As mentioned previously, both components in the chi-square mixtures were "shifted" chi-squares. In our simulations, the left truncation point for population 1 was always taken to be 7.5, and for population 2 it was located so that the desired overlap was obtained. In the MLEN and MDEN procedures, the natural constraints $\sigma_1^2>0$, $\sigma_2^2>0$, and $0\leq p\leq 1$ were imposed. Similarly, for the MDEW, the natural

constraints $\beta_1>0$, $\gamma_1>0$, $\beta_2>0$, $\gamma_2>0$, and $0\leq p\leq 1$ were imposed along with the constraints $\alpha_1\geq 0$ and $\alpha_2\geq 0$ which are reasonable constraints on the left-truncation point which would be imposed due to physical considerations, etc.

In Table 1 we display the results of the simulations. For a given simulation model and estimation procedure, we will obtain an estimate $\hat{p}$ of p, defined by

$$\hat{p} = \frac{1}{n_s}\sum_{i=1}^{n_s}\hat{p}_i$$

where $\hat{p}_i$ is the estimate of p for the ith sample, and n_s is the number of samples. Then based upon the simulations, estimates of the bias and MSE are given by:

$$\hat{\text{bias}} = \frac{1}{n_s}\sum_{i=1}^{n_s}(\hat{p}_i-p) = \hat{p} - p$$

$$\text{MSE} = \frac{1}{n_s}\sum_{i=1}^{n_s}(\hat{p}_i-p)^2.$$

Upon viewing the results, it can be seen that the MDEW was competitive when the component models were actually normally distributed, and it produced the best overall results for the chi-square mixtures. Of particular interest is the chi-square mixture where p=.5 and overlap=.10. This is the mixture displayed in Figure 5c and also in Figure 1 (except for location shift). When symmetric components are assumed (as with the MLEN and MDEN), a bias does occur in the estimation of p as discussed in Section 1. This behavior has been noted previously by Woodward, et.al.(1982). We see from the table that the MDEW performs substantially better

Table 1 - Simulation Results

Comparing Normal Based with
Weibull Based Estimation Procedures

Sample size = 200
Number of repititions = 200

Mixture of Normals

		Overlap = .10			Overlap = .03		
		$\hat{\rho}$	Bias	MSE	$\hat{\rho}$	Bias	MSE
p = .25	MLEN	.27	.02	.022	.25	.00	.022
	MDEN	.37	.12	.074	.26	.01	.004
	MDEW	.34	.09	.044	.30	.05	.011
p = .5	MLEN	.50	.00	.014	.50	.00	.002
	MDEN	.49	-.01	.023	.47	-.03	.002
	MDEW	.48	-.02	.019	.51	.01	.004

Mixture of $\chi^2(9)$

		Overlap = .10			Overlap = .03		
		$\hat{\rho}$	Bias	MSE	$\hat{\rho}$	Bias	MSE
p = .25	MLEN	.24	-.01	.061	.18	-.07	.006
	MDEN	.41	.16	.098	.17	-.08	.008
	MDEW	.50	.25	.122	.29	-.04	.007
p = .50	MLEN	.27	-.23	.064	.45	-.05	.011
	MDEN	.26	-.24	.061	.41	-.09	.010
	MDEW	.42	-.08	.024	.50	.00	.004
p = .75	MLEN	.50	-.25	.070	.65	-.10	.013
	MDEN	.48	-.27	.085	.64	-.11	.016
	MDEW	.62	-.13	.032	.71	.04	.005

than either of these normal based procedures on the basis of both bias and MSE. In Figure 6 we display histograms of the 200 estimates of p obtained from the three estimation procedures for the chi-square mixture shown in Figure 5c. It can be seen there, that the normal based procedures consistently estimated p to be substantially less than .5 while the estimates based on Weibull components were in general closer to the true vaue p=.5.

The one case in which the Weibull based estimates were not best, was when p=.25 with overlap=.10. This mixture is displayed in Figure 5a where it is obvious that estimation should be difficult since there is no distinct contribution due to component 1 in the mixture. Indeed, all procedures yield poor estimates as measured by the high MSEs. In Figure 7, we display histograms of the p values obtained from the three estimation procedures for this set of parameter configurations. There it can be seen that the Weibull procedure certainly gave the poorest results, with estimates being spread nearly uniformly between 0 and 1. However, the normal based procedures also had difficulty as is reflected in the histograms. In fact, there appears to be a tendency for the $\hat{p}_i$ values to be very low (approximately .10). However, $\hat{p}$ is very close to .25 for the MLEN since several of the $\hat{p}_i$ values were spread out uniformly between 0 and 1, which increased the estimate of p to near .25. However, the large MSE shown in the table for this case reflects this lack of accuracy.

FIGURE 6

Histograms of Estimates of p

Based on 200 samples of size 200 from mixture shown in Figure 5(c)

(True p = .5)

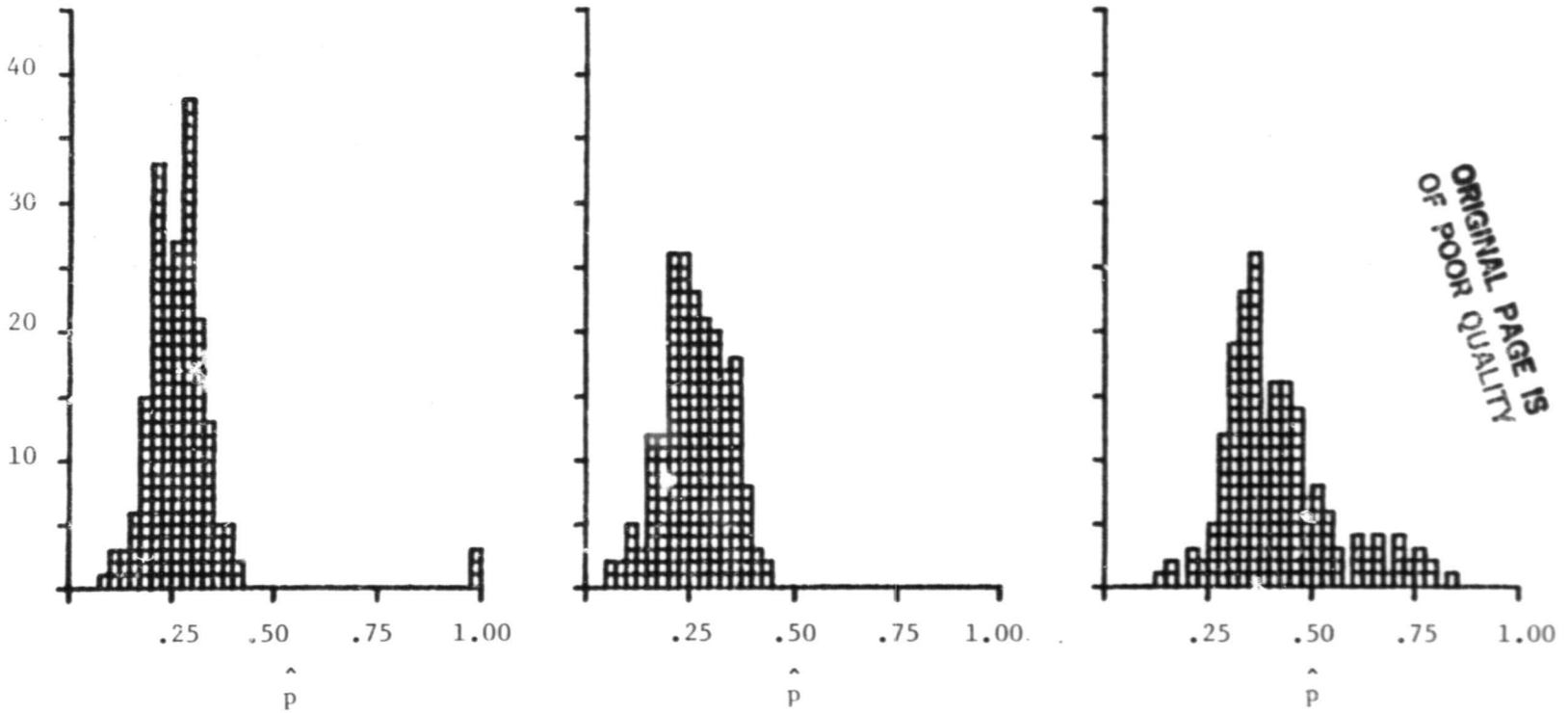
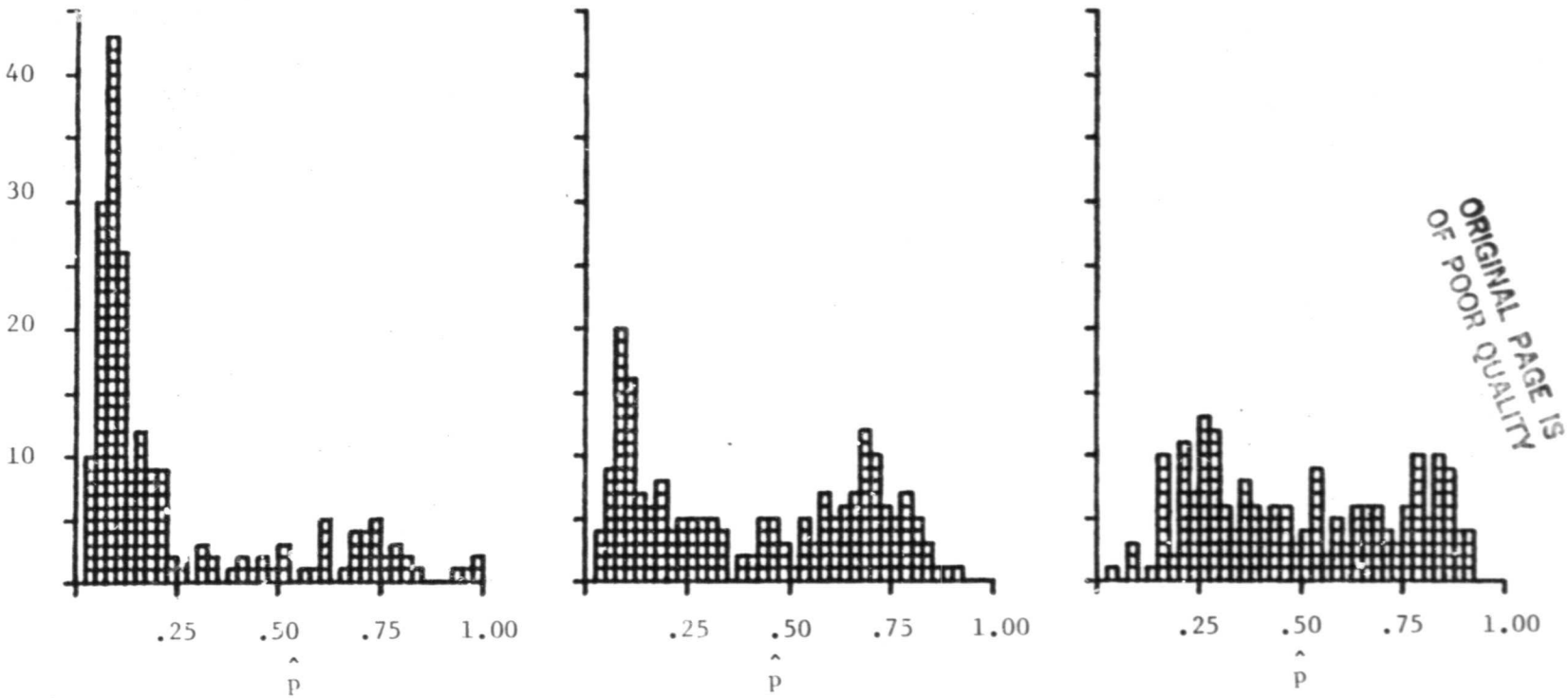


FIGURE 7

Histograms of Estimates of p

Based on 200 samples of Size 200 from mixture shown in Figure 5(a)

(True p = .25)



5. Concluding Remarks

Results in this report and in the report by Woodward, et.al.(1982) indicate that the normal based procedures perform poorly in the presence of a mixture of asymmetric distributions. In this paper we have suggested the mixture of Weibulls model as an alternative to the mixture of normals model in this situation. Results indicate that minimum distance estimation of the parameters of a mixture of Weibulls is a viable alternative to the normal-based techniques currently in use.

Before this procedure could be recommended and implemented, further research is needed. For example, the problem of how to obtain starting values for the parameters of mixtures of possibly asymmetric components has not been resolved. Also, the Weibull based procedures should be applied to LANDSAT data in order to examine their performance on the types of asymmetry which will be encountered in practice. The fact that an additional parameter has been introduced into the model for each component has caused the estimation procedures to be slower than for the normal based procedures. Further investigation concerning the practical aspects of actually implementing the procedures is needed.

References

1. Dubey, S. D. (1967). "Normal and Weibull Distributions," Naval Research Logistics Quarterly 14, 69-79.

2. Falls, L. W. (1970). "Estimation of Parameters in Compound Weibull Distributions," Technometrics 12, 399-407.

3. Kao, J. H. K. (1959). "A Graphical Estimation of Weibull Parameters in Life - testing of Electron Tubes," Technometrics 1, 389-407.

4. Johnson, N. L. and Kotz, S. (1970). Distributions in Statistics - Continuous Univariate Statistics - 2, Houghton Mifflin Company, Boston.

5. Looney, S. W. and Bargmann, R. E. (1982). "Procedures for Fitting a Mixture of Weibull Distributions to an Empirical Growth Curve," presented at National Meetings of the American Statistical Association in Cincinnati, August 1982.

6. Rider, P. R. (1961). "Estimating the Parameters of Mixed Poisson, Binomial, and Weibull Distributions by the Method of Moments," Bulletin de L'institute International de Statistique 39, 225-232.

7. Weibull, W. (1939). "A Statistical Theory of the Strength of Material," Ingeniors Vetenskaps Akademiens Handligar, Stockholm No. 151.

8. Wolfowitz, J. (1957). "The Minimum Distance Method," Annals of Mathematical Statistics 28, 75-88.

9. Woodward, W. A. , Schucany, W. R., Lindsey, H. and Gray, H. L. (1982). "A Comparison of Minimum Distance and Maximum Likelihood Techniques for Proportion Estimation," NASA Technical Report SR-62-04376, November 1982.

A P P E N D I X D

Critique of FCPF Automatic and Semi-Automatic Proportion Estimator Results

W.A. Woodward

The following discussion will concern the results given by FCPF in two recent NASA documents (1,2) concerning the performance of three new automatic and semi-automatic proportion estimation techniques. There is much overlap in these two documents although the data and conclusions in (2) represent revisions and additions to those given in (1). For this reason, the current report will concentrate mainly on the data and conclusions in (2). Before proceeding further it should be pointed out that we simply cannot draw certain inferences without access to the data itself. However, we will draw whatever conclusions we feel are warranted from the information provided.

Since we have no first hand experience with implementing any of the procedures involved, we will make no remarks concerning the implementation aspects of the various methods. Instead we will restrict ourselves to questions surrounding the quality of the proportion estimators being considered. This quality should be viewed from the perspective of how the new estimators compare with the current state-of-the-art analyst-intensive estimators as well as from a more absolute viewpoint concerning simply whether or not the new estimators meet acceptable standards.

Some useful information concerning the performance of the estimators is given in the table on page 2-22 of (2). Before continuing with the discussion of the results of this table some words of

caution should be given. First, the historical procedure is not a procedure but rather the various stages of the current P1A procedure during its evolution. Since 1979 is the only year during which the current state-of-the-art P1A technique was utilized the 1979 comparisons would be of most value from this perspective. In addition, it is our understanding that the FCPF techniques were developed using 1978 and 1979 data. We would expect procedures to perform best on segments undergoing similar weather patterns, etc. to the ones on which the techniques were based. Third, the 1976 historical data is based upon ratioed spring wheat rather than spring small grains. For this reason the 1976 comparisons will not be of much interest to us. Finally, it is impossible to tell from the table what the relationship is, for example, in 1977 among the 38 SSG4 segments, the 25 SSG3C segments, the 37 SSG3B segments and the 45 historical segments. This should be made clearer as it has a bearing on interpretation.

It is unfortunate that the mean absolute error (MAE) is not available for the historical procedures. When dealing with biased estimators, as the historical ones seem to be, the MAE and the mean squared error (MSE) are more informative measures than the standard deviation. The sample MSE is given by:

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (\hat{p}_i - p_i)^2 = \frac{1}{n} \sum_{i=1}^{n} e_i^2 \quad .$$

The MAE and MSE measure the amount of spread in the data around the ground truth whereas the standard deviation measures the amount of spread around the sample mean which may be quite

different from the ground truth. From the data on 2-22 the MAE values cannot be computed for the historical procedures. However, since MSE = $\frac{n-1}{n} S_e^2 + \bar{e}^2$, we can calculate MSE values for each procedure and each year. These MSE values are presented below:

Mean Squared Error Values for Data
of Table 2-22

	1976	1977	1978	1979
SSG4	131.0	136.0	110.0	182.3
SSG3C	99.7	147.9	131.1	354.7
SSG3B	75.2	92.5	121.3	311.9
Historical	100.9	65.7	69.4	46.1

The results of this table indicate that for 1977-1979 the MSE values for the new procedures, are two to six times as large as those for the historical procedures whereas MSE values for 1976 are similar. This reinforces the information given on page 2-22 concerning standard deviation comparisons. In comparing estimators, unbiasedness is usually not as important a criterion as MSE, i.e. the estimator with smallest MSE is usually favored regardless of the bias properties of the estimators.

Thus the historical techniques seem to be substantially better than the new procedures on the basis of MSE for 1977-1979 with results in 1979 being quite desparate. On page 2-40 the claim is made that when acquisitions can be made in windows 2, 3, and 4 (which is the optimal sampling situation) then SSG4 performs well since $\bar{e} = .04$. However, along the lines of our previous arguments, SSG4 in fact <u>does not</u> perform well in this most optimal

of situations since the associated MSE is 124. From an absolute perspective, it simply appears to us that the MSE's for the FCPF procedures are too high.

At this point, brief comment will be made concerning the bias properties of the estimators. A 90% confidence interval (based upon normality) about the true bias of a procedure is given by

$$(\bar{e} - t_{.95(n-1)} \frac{S_e}{\sqrt{n}} \ , \ \bar{e} + t_{.95(n-1)} \frac{S_e}{\sqrt{n}} \ )$$

where $t_{.95(n-1)}$ is the .95 percentage point of the t distribution with n-1 degrees of freedom. If this interval contains zero then the bias is not significantly different from zero whereas if this interval does not include zero, the bias is concluded to be significantly different from zero at the $\alpha = .10$ level of significance. It is obvious that the effect of large standard deviation or of small n is to lengthen this interval. Said another way, the result of a larger standard deviation or a smaller sample size is to decrease the power of the test i.e. decrease the probability of rejecting unbiasedness when an estimator is really biased. In the present setting the failure to reject unbiasedness in the FCPF estimators is largely a function of larger standard deviation and smaller sample size than it is due to smaller bias estimates. In fact, if for each year the FCPF estimates had the same standard deviation and sample size as that associated with the historical data, only SSG3C and SSG3B in 1978 would have yielded estimates for which the bias was not significantly different from zero. In

summary, although the sample biases were generally smaller for FCPF estimates than for historical estimates, we certainly agree with the statement on 2-3 that there is no significant difference in the biases of the two procedures. The authors were not as careful in their statement on 2-19. It should be pointed out again however that MSE is a measure of the goodness of an estimator which is appropriate for comparing estimators whether they be biased or unbiased.

The implications of the large MSE's are evident in other data presentations in (2). The small r^2 values in pages 2-23 through 2-25 are the ones associated with large MSE's with the 1979 results for SSG4, SSG3C, and SSG3B being extremely noteworthy. One would certainly be hesitant to recommend a procedure which yielded results as unrelated to ground truth as were the FCPF results in 1979. The claim on page 2-19 that the lack of "good" correlation in 1979 for SSG4 is explainable, seems to be questionable. If the outlier point is deleted, the correspondence between ground truth and SSG4 estimates is still poor. Consider a vertical line drawn through ground truth proportion .25 on the 1979 SSG4 plot. It can be seen that there is very little correlation between ground truth and $\hat{p}$ on either side of the line. The correlation which does appear is only due to the fact that SSG4 seems to do a fair job of separating low ground truth proportions from high ones. Another word of warning concerning interpretation of r^2 values is in order here. The r^2 value measures the amount of fit to the line which best fits the data. If this line is not approximately the 1-1 line, i.e. the line with slope of 1 and intercept of zero

then this fit is of little importance. The line drawn in the plots on 2-23 through 2-26 is the 1-1 line. It is clear that in some of the plots this line is not the best fitting line whose fit to the data is being measured by r^2. A more meaningful measure of fit would be one which measures departure from the data to this 1-1 line. It is easily shown that MSE is the average squared vertical distance from the data points to the 1-1 line.

We will conclude with a few additional comments. On page 2-50, data are given concerning processability rates of the procedures. In 1978 data from the second satellite were available which should have produced a higher processability rate. This increased processability rate is visible in the 3 FCPF procedures but is not visible in the 1978 historical data. Since SSG4 processability for 1976, 1977, and 1979 was approximately 12-20% lower than that for the historical, and since for 1978 the SSG4 rate was 24% higher than the historical rate, there seems to be cause for concern relating to the validity of the 1978 historical processability rate. The error characterization analyses were interesting and should indeed provide useful information concerning possible modifications of the FCPF techniques. It is not clear of course whether or not modification in the procedure will be able to improve performance.

In conclusion we feel that the results of the comparisons between the FCPF automatic and semi-automatic procedures and the historical results are not very encouraging. Although results

and cautions are in general adequately related in (2) it is our opinion that the apparent unbiasedness of the FCPF procedures resulted in excessive optimism concerning their performance. The problem with excessive variance with the FCPF procedures was mentioned but seemingly did not cause great concern possibly because of "apparently" larger bias for the historical procedures. However our analyses involving the MSE as the standard for comparing estimators indicates that indeed the FCPF procedures do not perform at the level of historical analyst-intensive techniques.

References

1. Preliminary Technical Results Review of FY81 Experiments, Volume 1, Fiscal Year 1981/82 Spring Small Grains Pilot Experiment, FC-J1-04175, JSC-17433, September 23, 1981.

2. Semi-Annual Project Management Report, Program Review Presentation to Level 1, Interagency Coordination Committee, FC-J1-04181, JSC-17438, November 1981.

A P P E N D I X E

Review of
A Crop Area Estimator Based on Changes
in the Temporal Profile of a
Vegetative Index
(Smith and Ramey)

by Wayne A. Woodward

The paper by Smith and Ramey contains some interesting ideas concerning the use of temporal data in estimating a vegetative index. I have several comments concerning the paper:

1. Although I am not extremely familiar with the Cubic Color Model of Cates, et al, I do want to make some comments about it. To me it appears that normalizing the readings from channels 1,2, and 4 based upon the segment means in each of these bands can have some undesirable effects. For example, if early in the season an acquisition is taken when a majority of the segment is bare soil, then on the CIR film it would appear mostly green in color. However, the normalizing procedure of dividing by segment means would assign what is very probably green (on CIR film) to the neutral gray position. Consequently, a pixel with relative energies of (4.99,4.98, 5.01) would be called red, and placed in the vegetative class when in fact it was represented as green on the film, and was probably nonvegetated. The same phenomenon could of course occur in reverse. In addition, the fact that (4.99,4.98,5.01) and (0,0,10)

are assigned the same "color" seems like an unfortunate loss of information.

2. The fact that the procedure does no more than provide a vegetative index, implies that it will not, of course, be able to provide proportion estimates for individual crops.

3. The shortcomings which the authors list on page 9 are quite serious. The fact that the underlying profiles are not separated by a constant violates a basic assumption in the multiple regression (or analysis of covariance) model posed. Also, although only the parameter alpha is of interest, it is likely that if estimation is a problem, estimates of alpha will suffer along with those of the betas.

4. Finally, the results of the technique as applied to segment data requires some comment. To me these results from the 10 segments seem quite unimpressive. The magnitude of the errors is unacceptably high, and the authors' statement on page 11 that the technique "apparently produced unbiased estimates" is completely unfounded. It seems that the authors believe that the impressive feature of their results is the "high" correlation of .73 between observed and expected percent

changes. Observing the 10 pairs of values upon which this correlation is based reveals that there actually does not seem to be a strong correlation between these values. In fact, when the results for segment 1658 are removed from the data set, then the correlation is only .33. (My calculations showed a correlation of .67 instead of .73 for the data shown.) Examination of the data in the table shows that both observed and expected percent change for this segment were much larger than those from other segments. This data pair thus had an inordinate influence on the correlation coefficient (sometimes called the "lollipop" effect.) If the nonparametric Spearman correlation coefficient had been used instead of the Pearson correlation which depends upon a bivariate normality assumption, the correlation using all 10 data pairs would have been only .43, again an unimpressive result.

In short, the results of this paper certainly do not convince me that the technique proposed here has any merit.