N O T I C E

THIS DOCUMENT HAS BEEN REPRODUCED FROM MICROFICHE. ALTHOUGH IT IS RECOGNIZED THAT CERTAIN PORTIONS ARE ILLEGIBLE, IT IS BEING RELEASED IN THE INTEREST OF MAKING AVAILABLE AS MUCH INFORMATION AS POSSIBLE

CROP AREA ESTIMATION BASED ON REMOTELY-SENSED DATA WITH AN ACCURATE BUT COSTLY SUBSAMPLE

(NASA-CR-176017) CROP AREA ESTIMATION BASED ON REMOTELY-SENSED DATA WITH AN ACCURATE BUT COSTLY SUBSAMPLE Final Report, 1 Jan. 1984 - 31 Aug. 1985 (Southern Methodist Univ.) 33 p HC A03/MF A01 CSCL 02C G3/43

N86-13814

Unclas
04833

FINAL REPORT

ON

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

COOPERATIVE AGREEMENT NO. NCC 9-9

1 January 1984 to 31 August 1985

Richard F. Gunst
Department of Statistics
Southern Methodist University
Dallas, Texas 75275

ABSTRACT

This report documents research activities which were conducted under the auspices of National Aeronautics and Space Administration Cooperative Agreement NCC 9-9. During this contract period research efforts were concentrated in two primary areas. The first area was an investigation of the use of measurement error models as alternatives to least squares regression estimators of crop production or timber biomass. The second primary area of investigation was on the estimation of the mixing proportion of two-component mixture models. This report documents publications, technical reports, submitted manuscripts, and oral presentations which occurred as a result of these research efforts. Possible future areas of fruitful research are mentioned.

Table of Contents

I. MAJOR RESEARCH ACTIVITIES

Research supported by this cooperative agreement primarily focussed on two major topics: regression estimation using models in which both the response and the predictor variables are subject to measurement errors and the estimation of parameters in mixture models. The first topic was investigated as a means of improving satellite remote-sensing estimates of crop production and timber biomass by combining satellite estimates with ground truth measurements. The latter topic was explored in an effort to improve the estimation of parameters in models of crop or vegetation proportions when there are two or more different types of vegetation in a segment.

Several research publications resulted from the research on these and other topics. Six publications appeared in referred scientific journals and another three manuscripts were published as technical reports or proceedings papers. In addition, four manuscripts are currently submitted for publication, the last of which is included in this report as Appendix C. A complete list of publications and submitted manuscripts is given in Appendix A. Ten oral presentations, listed in Appendix B, were given to further disseminate the results of this research.

A summary of the major topics of research which were undertaken through the support of this cooperative agreement are presented in the next three sections.

A. MEASUREMENT ERROR MODELS

The primary focus of the research conducted under the auspices of this cooperative agreement was on the investigation of techniques for improving the use of satellite remote-sensing estimates of crop proportions and timber biomass. Specific attention was devoted to the combining of relatively inexpensive but imprecise satellite estimates with relatively expensive but highly precise ground-truth estimates. Although least squares regression estimates had been used for some time in this effort, it was recognized that both the satellite estimates and the ground truth estimates are subject to measurement error, thereby invalidating one of the key assumptions needed for the use of least squares estimation. Thus intense research activities were directed toward the investigation of regression estimation with measurement error models.

Denote a true (i.e., error-free) ground-truth measurement by Y and the corresponding error-free satellite measurement by X. Assume that an adequate representation of the relationship between the two measurements is given by a linear model of the form

$$Y = \alpha + \beta X.$$

Because of errors of measurement (e.g., registration errors, irregularity shaped fields, etc.), the true ground-truth and satellite measurements are not observed. Rather, one observes

$$x = X + u \quad \text{and} \quad y = Y + \nu \ ,$$

where u and ν represent the measurement errors. In this framework the least squares estimators of α and β are biased since an underlying assumption which is necessary for unbiased least squares estimation is that the observable predictor variable x is measured without error.

To date, research has concentrated on structural measurement error models; i.e., the true predictor variable X is stochastic. It is well known that if X, u, and v are normally distributed there does not exist consistent estimators of α and β unless (a) one or more of the model parameters ir known, (b) replicated observations are available , or (c) measurements are taken on additional (instrumental) variables which are correlated with the true predictor variable X but not with the errors u and v.

Assuming independent normal distributions for X, u, and v with $\lambda = var(v)/var(u)$ known, an investigation was conducted into the effects of sample size on the precision of the maximum likelihood estimator of β and on the consequences of selecting an erroneous value for λ. The results of this research were reported in publication 1(b) (Appendix A) and technical reports 2(a) and 2(c). A parallel investigation when the errors u and v are correlated is reported in publication 1(d). A summary of many of the properties of these estimators and a comparison with least squares is reported in the manuscript 3(b).

B. MIXTURE MODELS

Mixture models are used to probabilistically characterize the occurrence of spectral measurements from segments in which two or more crops or other vegetation are present. If x denotes a (possibly vector-valued) spectral measurement from a segment in which two crops are present, the probability density function $f_\theta(x)$ can be expressed as

$$f_\theta(x) = pg_1(x;\phi_1) + (1-p)g_2(x;\theta_2)$$

where $g_j(x;\theta_j)$ is the density function for x in crop j, $\theta = (p,\phi_1,\phi_2)'$ is a vector of model parameters (with θ_j possibly vector-valued), and p is the proportion of crop 1 in the segment.

Estimation of θ presents challenging and, to date, many unsolved problems. During the course of this cooperative agreement progress was achieved in the estimation of the mixing proportion p. The manuscript 3(a) examines maximum likelihood and minimum distance estimation of the mixing proportion p when the two component distributions $g_j(x_j;\phi_j)$ are represented by three-parameter Weibull distributions. The three-parameter Weibull was used because of the variety of shapes it can have by specifying values of the parameters. In manuscript 3(a) the distance measure used was the Cramer-von-Mises distance W^2, whereas in manuscript 3(d) (see also Appendix C) the Hellinger distance was used with normal component distributions.

C. OTHER TOPICS

Additional research which was completed during the duration of this cooperative agreement is reported in Appendix A. Several papers were published (1(a), 1(e), 1(f), 2(b)) or submitted for publication (3(c)) on regression was with collinear variables. One article (1(c)) was published on the use of quadratic forms in screening procedures.

II. PROSPECTIVE FUTURE RESEARCH

The research outlined in this report has not only led to the publication of several scholarly articles, it has also identified topics which offer great potential for further productive research efforts. Some of the general areas of possible future research activities are now briefly outlined.

Many important problems remain unresolved when estimating the parameters of measurement error models. While there is theoretically no problem with estimating the parameters of measurement error models when the true predictor variable X follows some nonnormal probability distribution, little work has been conducted on the implementation of maximum likelihood (or minimum distance) estimation in this setting. Likewise, no work has been done on evaluating the effects of sample size on such estimation procedures.

Another potential area of research is the extension of the results reported above to regression models having more than one predictor variable. Questions relating to the choice of error variance ratios and the consequences of misspecifying these ratios require theoretical and simulation investigations. Again, the difficulty with implementing estimation procedures other than least squares when the predictor variables are nonnormal must be addressed.

Associated with the estimation of the parameters of measurement error models is the use of fitted models for prediction and calibration. Only a few published articles have appeared on prediction and calibration with measurement error models.

The other major topic which was investigated during the course of this cooperative agreement, mixture model estimation, likewise poses many problems for potentially fruitful future research. In the current research only the estimation of the mixing proportion p was studied. Much work remains before acceptable estimation of all the model parameters can be achieved. So too, estimation procedures for three or more component distributions need to be developed.

A great deal of work remains to be done on the selection of component distributions for use with the mixture model. Mi imum distance estimation does not necessitate that the component distributions be the "true" ones in order to satisfactorily estimate crop proportions. During the investigation of the Weibull components it was discovered that several widely differing sets of parameters could yield mixture distributions which were virtually identical. It would be extremely useful to identify component families of distributions for which parameter estimation is computationally efficient and which are "stable". The stability of estimates would require that small perturbations of the data would not result in widely differing parameter estimates and that radically different parameter choices could not produce virtually identical mixtures.

III. ORAL PRESENTATIONS AND OTHER ACTIVITIES

Appendix B lists ten oral presentations of the research conducted during this cooperative agreement. These oral presentations permitted rapid dissemination of the major accomplishments of this research. Presentations (8), (9), and (10) will be acknowledged as "Outstanding Contributed Paper Presentations" by the Section on Physical and Engineering Sciences of the American Statistical Association (ASA) during the 1986 Annual ASA Meetings next August.

Funding from this cooperative agreement was used for partial support of the principal investigator during the summer months of 1984-1985. Also partially supported during the summer months of 1985 was Professor Wayne A. Woodward, who led the mixture model research. Three advanced graduate students were also partially supported during the duration of this cooperative agreement: Mr. Kelly Cunningham, Miss Miriam Reilman and Dr. Many Lakshminarayanan. Dr. Lakshminarayanan completed his doctoral degree requirements while supported by funds from this contract.

IV. APPENDICES

A. COMPLETED RESEARCH

B. ORAL PRESENTATIONS

C. ADDITIONAL WORK

A. COMPLETED RESEARCH

1. Publications in Refereed Journals

    (a) "Toward a Balanced Assessment of Collinearity Diagnostics," The American Statistician, 38 (1984), 79-82.

    (b) "Estimation of Parameters in Linear Structural Relationships: Sensitivity to the Choice of the Ratio of Error Variances," Biometrika, 74 (1984), 569-573.

    (c) "Screening Procedures Using Quadratic Forms," Communications in Statistics, A14 (1984), 1393-1404.

    (d) "Structural Model Estimation with Correlated Measurement Errors," Biometrika, 72 (1985) to appear.

    (e) "Outlier-Induced Collinearities", Technometrics, 27 (1985), to appear.

    (f) "Selecting Principal Components in Regression," Statistics and Probability Letters 3 (1985), 299-301.

2. Technical Report and Proceedings Papers

    (a) "Sensitivity of Errors-in-Variables Estimators to the Specification of the Ratio of Error Variances," Technical Report, NASA Johnson Space Center, Houston, TX (1983).

    (b) "Regression Diagnostics and Approximate Inference Procedures for Penalized Least Squares Estimators," Department of Statistics Technical Report No. 181, SMU, Dallas, TX (1983).

    (c) "Exploring the Use of Linear Structural Models to Improve Remote-Sensing Agricultural Estimates, Proceedings of the NASA Symposium on Mathematical Pattern Recognition and Image Analysis, NASA Johnson Space Center, Houston, TX (1984).

3. Research Articles Submitted for Publication

   (a) "Estimating Mixture Proportions for Component Weibull Distributions."

   (b) "Stochastic Regression with Errors in Both Variables".

   (c) "Diagnostics for Penalized Least Squares Estimators"

   (d) "Minimum Hellinger Distance Estimation of Mixture Model Parameters: A Re-Examination"

B. ORAL PRESENTATIONS

1. "Exploring the Use of Linear Structural Models to Improve Remote-Sensing Agricultural Estimates," NASA Symposium on Mathematical Pattern Recognition and Image Analysis, June 6-8, 1984, Johnson Space Center, Houston, Tx.

2. "Effects of Misspecifying the Error Variance Ratio in Linear Structural Relationships." Joint Annual Meetings of the American Statistical Association and the Biometric Society, Philadelphia, PA, August 13-16, 1984.

3. "Collinearity Assessment with Errors-in-Variables Models." Joint Annual Meetings of the American Statistical Association and the Biometric Society, Philadelphia, PA, August 13-16, 1984.

4. "Regression with Collinear Predictor Variables: Implications for Causal Inference", Department of Quantitative Business Analysis, Louisiana State University, October 18, 1984; Department of Mathematics, Northern Arizona University, November 9, 1984.

5. "Regression Estimation with Linear Structural Models," Division of Mathematical Sciences, University of Texas at Dallas, February 14, 1985.

6. "Regression Models when Both Variables are Subject to Measurement Errors". Spring Regional Meeting of the Biometric Society (ENAR), Raleigh, NC, March 25-27, 1985.

7. "Measurement Error Models," Mathematics Department, General Motors Research Laboratories, July 22, 1985.

8. "Replication and Instrumental Variables Estimators for Linear Structural Models." Annual Meetings of the American Statistical Association and the Biometric Society, Las Vegas, NV, August 5-8, 1985.

9. "Regression Estimation with Controlled Observations". Annual Meetings of the American Statistical Association and the Biometric Society, Las Vegas, NV, August 5-8, 1985.

10. "Maximum Likelihood as Least Squares in Structural Model Estimation." Annual Meetings of the American Statistical Association and the Biometric Society, Las Vegas, NV, August 5-8, 1985.

C. ADDITIONAL WORK

"Minimum Hellinger Distance Estimators
of Mixture Model Parameter:
A Re-Examination"

MINIMUM HELLINGER DISTANCE ESTIMATION
OF MIXTURE MODEL PARAMETERS
A RE-EXAMINATION

Wayne A. Woodward

1. Introduction

The problem of using minimum Hellinger distance estimation suggested by Beran (1977) for purposes of estimating mixture model parameters was initially discussed by Woodward and Eslinger (1983). The use of the minimum Hellinger distance estimator (MHDE) is intuitively appealing due to the fact that it asymptotically efficient and asymptotically normal in various settings (see Beran (1977), Stather (1981)) yet it has been shown to be robust to departrues from normality (see Eslinger (1984)). It was believed that its use in the mixture of normals setting often assumed in crop proportion estimation could provide efficient estimates when the model is correct along with robustness to departures from normality. Woodward, et. al (1982, 1983, 1984) studied the use of minimum distance estimation of the mixture model parameters using Cramér-von Mises distance and found that the Cramér-von Mises distance estimator (MCVMDE) provided results superior to the maximum likelihood estimator

(MLE) under departrues from component normality but yielded decidely poorer results than the MLE when the assumption of component normality was true.

Woodward and Eslinger (1983) showed empirically that the MHDE does provide estimators comparable to the MLE under component normality which at the same time show some robustness to departures from normality. The MHDE was not, however, as robust as the MCVMDE, a result also shown by Eslinger (1984) for the two parameter normal. These results are to be expected since in essence the MHDE is a compromise between the very robust estimator which is not efficient at the true model and the efficient estimator such as the MLE which is not robust. However, Woodward and Eslinger (1983) encountered problems in implementing the MHDE in the mixture setting. In particular the implementation of the MHDE used by those authors tended to be extremely sensitive to starting values which resulted in "failure to converge" problems an unacceptable high percentage of the time. In particular for mixtures with overlap, as defined by Woodward et al. (1984) of .1 the MHDE converged on the average about 82% of the time and converged 88% of the time for .03 overlap.

In this report we re-examine the use of the MHDE in the mixture of normals setting. We empirically examine the effect of the selection of the smoothing parameter h used in the kernel density estimation component of the MHDE, and we examine the use of an alternative maximization scheme. In this report as in the earlier reports by this author, we will be concerned only with the estimation of the mixture proportion.

In Section 2 we provide a brief description of the MHDE in the mixture-of-normals setting as implemented by Woodward and Eslinger (1983). Section 3 provides details concerning new approaches designed to improve the convergence of the MHDE. In Section 4 we give details of simulations examining the new techniques.

2. The Minimum Hellinger Distance Estimator

Let $\mathcal{F}_\theta = \{F_\theta : \theta\varepsilon\Theta\}$ denote a family of distributions which will be referred to as the projection family. These distributions depend on the possibly vector valued parameter θ. We will assume for our purposes here that the distributions in $\mathcal{F}_\theta$ are absolutely continuous. In our case we use the mixture of normals projection model

$$f_\theta(x) = \frac{p}{\sqrt{2\pi}\,\sigma_1} e^{-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2} + \frac{(1-p)}{\sqrt{2\pi}\,\sigma_2} e^{-\frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2} \qquad (2.1)$$

where $\underline{\theta} = (\mu_1,\sigma_1,\mu_2,\sigma_2,p)'$ are all assumed to be unknown. An MHDE of θ is a value $\hat{\theta}$ which minimizes a "distance" between the data distribution (whose model is unknown) and the projection model. The distance measure used in MHD estimation is the Hellinger distance. The Hellinger distance between two absolutely continuous distributions is defined to be $||f^{1/2}-g^{1/2}||$ where f and g denote the corresponding density functions and $||\cdot||$ denotes the L^2 norm, i.e.

$$||f^{1/2}-g^{1/2}|| = [\int(f^{1/2}-g^{1/2})^2dx]^{1/2} \qquad (2.2)$$

where integration is with respect to Lebesgue measure on the real line. The MHD estimator $\hat{\theta}_H$ of θ is defined as a value of θ which minimizes $||f_\theta^{1/2}-\hat{g}_n^{1/2}||$ where $\hat{g}_n$ is a nonparametric density estimator. We will employ a kernel density estimator

$$\hat{g}_n(x) = \frac{1}{nh_n} \sum_{i=1}^{p} w\left(\frac{X-X_i}{h_n}\right)$$

based on the Epanechnikov kernel $w(x) = .75(1-x^2)$ for $|x|\le 1$. It should be noted that minimizing $||f_\theta^{1/2}-\hat{g}_n^{1/2}||$ is equivalent to maximizing

$$\int f_\theta^{1/2}\hat{g}_n^{1/2}dx. \tag{2.3}$$

In the earlier report Woodward and Eslinger (1983) examined the use of the MHDE in this setting by maximizing (2.3) using Newton's method. This recursive method involved the calculation of $\frac{\partial f_\theta^{1/2}}{\partial\theta}$ and $\frac{\partial^2 f_\theta^{1/2}}{\partial\theta^2}$, the forms of which are given in the Appendix of that report.

3. New Implementations

In the earlier report the smoothing parameter h_n in the kernel density estimator was taken to be $c_n s_0$ where $c_n = 2.16n^{-.271}$ and s_0 was set to the starting value estimate of the component with the larger mixing proportion, i.e.

$$s_0 = \hat{\sigma}_1(0) = [(\hat{\mu}_1(0) - r_1^{(.25)})/.6745]^2$$

if $p \ge .5$ and

$$s_0 = \hat{\sigma}_2(0) = [(r_2^{(.75)} - \hat{\mu}_2(0)/.6745]^2$$

if $p < .5$ where $\hat{\mu}_j(0)$ is the initial estimate of μ_j and $r_j^{(q)}$ is the qth percentile from the jth cluster, $j=1,2$. We denote this value of h_n as $h_n^{(1)}$. This value of h_n was chosen since $h_n = c_n s_n$ with s_n the median absolute deviation, had been shown, Eslinger (1983) to be the optimal h_n when using a two-parameter normal projection model. In this report we examine the impact of two other choices for h_n. The first modification to the earlier work was to take $h_n = c_n s_m$ where c_n is as before and s_m is an initial estimate of the mixture standard deviation given by

$$s_m = [\hat{p}(0)(\hat{\sigma}_1^2(0) + d_1^2) + (1-\hat{p}(0))(\hat{\sigma}_2^2(0) + d_2^2)]^{1/2}$$

where $d_1 = \hat{\mu}_1(0) - \hat{\mu}_m(0)$, $d_2 = \hat{\mu}_2(0) - \hat{\mu}_m(0)$ with

$\hat{\mu}_m(0) = \hat{p}(0)\hat{\mu}_1(0) + (1-\hat{p}(0))\hat{\mu}_2(0)$.

This resulted in larger values for h_n, especially in the cases in which there is substantial separation between the two components.

Parzen (1962) has found the h_n which minimizes the integrated mean square error between a kernel density estimator and the true density f. His result shows that the h_n optimal in this sense is $h_n = \alpha(w)\beta(f)n^{-1/5}$ where

$$\alpha(w) = \frac{[\int w^2(y)dy]^{1/5}}{[\int w(y)y^2 dy]^{2/5}} \tag{3.1}$$

and

$$\beta(f) = [\int[\frac{\partial^2 f(x)}{\partial x^2}]^2 dx]^{-1/5} .$$

For the Epanechnikov kernel, $\alpha(w) = 1.71877$. In the case in which f is a mixture of normals as in (2.1), $[\partial^2 f(x)/\partial x^2]^2$ is given by

$$\left(\frac{\partial^2 f(x)}{\partial x^2}\right)^2 = \frac{p^2}{2\sigma_1^5\sqrt{\pi}}\, n(\mu_1,\sigma_1/\sqrt{2})(z_1-1)^2$$

$$+ \frac{(1-p)^2}{2\sigma_2^5\sqrt{\pi}}\, n(\mu_2,\sigma_2/\sqrt{2})(z_2-1)^2$$

$$+ \frac{2p(1-p}{2\pi\,\sigma_1^3\sigma_2^3}\, e^{-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2} e^{-\frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2} (z_1-1)(z_2-1)$$

where $z_1 = \frac{x-\mu_1}{\sigma_1}$, $z_2 = \frac{x-\mu_2}{\sigma_2}$ and $n(\mu,\sigma)$ is the normal density with mean μ and standard deviation σ. We examined the use of $h_n^{(2)} = 1.71877\beta(f)n^{-1/5}$ where $\beta(f)$ was approximated numerically using IMSL's numerical integration routine DCADRE.

As another strategy for improving convergence of the MHDE, an alternative maximization technique was considered. Recall that the MHDE is defined to be a value $\hat{\theta}$ which minimizes the integral

$$I = \int(f_\theta^{1/2} - \hat{g}_n^{1/2})^2 dx.$$

This integral can be approximated using the trapezoidal rule by

$$\hat{I} = \Delta t_i \sum_{i=1}^{k} a_i (f_\theta^{1/2}(t_i) - \hat{g}_n^{1/2}(t_i))^2 \qquad (3.2)$$

where $a_1 = a_k = 1/2$ and $a_i = 1$ for $i=2,3,\ldots,k-1$ for a partition $t_1,t_2,\ldots,t_k$ of $[a,b]$ a finite interval, in our case taken to be the support of $\hat{g}_n$, i.e., $a = Y_1 - h_n$ and $b = Y_n + h_n$ where Y_i

denotes the ith order statistic. The procedure employed was to minimize the sum-of-squares in (3.2) using IMSL routine ZXSSQ which employs the Marquardt-Levenburg algorithm (1963).

4. Simulation Results

In this section we describe the results of simulations designed to test the procedures described in the previous section. Both the Newton and Marquardt-Levenberg algorithms were examined using $h_n^{(1)}$ and $h_n^{(2)}$. Simulations involve mixtures of normal and of non-normal components using the parameter configurations employed by Woodward and Eslinger (1983). In particular, we use mixing proportions .25, .50, and .75 and "overlap" as previously defined by Woodward, et al. (1984) of .03 and .10. As in the previous report we consider cases in which σ_1/σ_2 equals 1 and $\sqrt{2}$. In the present study we consider mixtures of normal and t(4) components while some results are given for mixtures t(2) components. For each set of configurations considered, 500 samples of size n=100 were generated from the corresponding mixture distribution. Simulations were performed on the IBM 3081-D24 computer at Southern Methodist University. Starting values were obtained as discussed by Woodward, et al. (1984). For each sample simulated, the MLE and MCVMDE were obtained along with several MHDE estimators. MHD estimators employing Marquardt's procedure for minimizing (3.2) will be denoted MHDEM(i) where i=1 or 2 depending on whether $h_n^{(1)}$ or $h_n^{(2)}$ were used in the density estimation. Likewise MHDEN(1) denotes the estimator using $h_n^{(1)}$ and employing the Newton's method for maximizing (2.3). The estimator MHDEN(2) was examined for selected configurations and was seen to not perform as well as MHDEN(1). Also shown are results labeled MHDEN'(1). These correspond to the MHDEN(1) type estimates obtained by using starting values for σ_1 and σ_2 which are smaller by a factor of 1.2 than those obtained by the straightforward starting values given by Woodward et al. (1984). Woodward and Eslinger (1983) showed that in the MHDE setting studied there, these smaller starting values produced better results. For means of comparison we have denoted by MHDE* in the tables the corresponding values obtained by Woodward and Eslinger (1983). These estimators were not obtained on the same sequence of samples as the current simulations. In Table 1 we present results for simulated mixtures of normal components, in Table 2 we show the corresponding results for simulated mixtures of t(4) components while in Table 4 we show a few results for simulated t(2) mixtures. Simulation based estimates of the bias and MSE associated with the various estimators are given by

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

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

where n_s denotes the number of samples (500 in our case) and $\hat{p}_i$ denotes an estimate of p for the ith sample. As in the earlier reports $n\widehat{\text{MSE}}$ is given in the tables where n is the size of each individual sample (100 in our case). We provide empirical measures of the relative efficiencies of the various estimators with the MLE, by

$$\hat{E} = \frac{\hat{MSE}\ (MLE)}{\hat{MSE}\ (estimator)} \quad .$$

An approximate standard error of a tabled n$\hat{MSE}$ is (.0632) (n$\hat{MSE}$). Also listed is the percentage of the 500 samples for which that estimator converged to "reasonable" values. For a discussion of "reasonable" values see Woodward, et al. (1984). If such convergence was not obtained, for purposes of the simulation study the estimate of p was taken to be the starting value.

Examination of the tables shows that indeed the MHDE does appear to behave as expected, i.e.provides fairly efficient estimators under component normality (as evidenced by $\hat{E}$ values near 1) along with estimates more robust than the MLE for simulations of mixtures of non-normal components (as evidenced by $\hat{E} > 1$).

The percentage converging information is summarized in Table 3 where the value tabled for a given overlap is the average percent convergence obtained for the 10 configurations of normal and t(4) components in Tables 1 and 2 for that overlap. All of the techniques proposed in this report produced estimators with higher rates of convergence than the estimators in the earlier study, especially for the .03 overlap. It is very clear from the table that the Marquardt based estimators are far superior to those using Newton's method in terms of percentage convergence obtained. It should be noted that convergence was almost always obtained in these settings by the MLE and MCVMDE.

Concerning MSE performance, initial examination shows that MHDEN(1) provides smaller MSE's than the other MHDE techniques for .10 overlap yet MHDEM(2) seems to provide the best estimates at the .03 overlap. However, due to the fact that in the .10 overlap case the starting values tend to outperform all other estimators and that about 16% of the "MHDEN(1)" results are actually starting values, this has a tendency to deflate the $n\hat{MSE}$ for MHDEN(1). As in the earlier study, the MHDEN'(1) (using the scaled starting values) seem to perform better than MHDEN(1).

It should be pointed out that about twice as much time is required to produce MHDEM(i) estimates as the MHDEN(i) estimates. It was mentioned by Woodward and Eslinger (1983) that the time required for the MHDEN was comparable to that for the Cramér-von Mises estimator.

In Table 4 we show results for simulated t(2) components for configurations with $\sigma_1/\sigma_2 = 1$. It is seen that again, in this extremely heavy tailed departure from component normality, the MHDE produced results markedly better than the MLE yet usually not as robust as the MCVMDE. Convergence in this setting was sometimes a problem for the MLE procedure which used the EM algorithm. However, the MHDEM(2) estimates converged at an extremely high rate.

5. Concluding Remarks

As a result of the present study it seems that the MHDE could indeed be a useful estimator of the mixing proportion of a two component mixture. Using the Marquardt procedure, the convergence problems no longer exist although computer time is doubled. Work should be done in order to determine whether or not the time required to calculate the MHDEM(i) estimates can be decreased.

Parameter estimates using the MHDE conform to predictable patterns, i.e. the estimates are more efficient than the Cramér-von Mises estimates under component normality, yet are not as robust as the Cramér-von Mises results. Because of the increased percentage of convergence, the results obtained here provide a clearer picture of actual estimator performance over the results given earlier by Woodward and Eslinger (1983).

Whether the MHDE could be successfully used as an alternative to maximum likelihood estimation of a crop proportion based on remote sensing data remains to be determined. However, the results shown here imply that its successfully application is a possibility.

REFERENCES

Beran, R. (1977). "Minimum Hellinger Distance Estimates for Parametric Models," Annals of Statistics 5, 445-463.

Eslinger, P.W. (1983). "Minimum Hellinger Distance Estimation," Ph.D. Dissertation, Statistics Department, Southern Methodist University, Dallas, Texas.

Marquardt, D. W. (1963). "An Algorithm for Least-Squares Estimation of Nonlinear Parameters," Journal of the Society for Industrial Engineers 11, 431-441.

Parzen, E. (1962). "On the Estimation of Probability Density Function and Mode," Annals of Mathematical Statistics 33, 1065-1076.

Stather, C. R. (1981), "Robust Statistical Inference Using Hellinger-Distance Methods", Ph.D. Dissertation, Department of Mathematical Science, School of Physical Science, LaTrobe University, Australia.

Woodward, W. A. and Eslinger, P. W. (1983). "Minimum Hellinger Distance Estimation of Mixture Model Parameters," NASA Technical Report SR-63-04433, July 1983.

Woodward, W. A., Parr, W. C., Schucany, W. R. and Gray, H. L. (1983). "Minimum Distance Estimation of Mixture Model Parameters-Asymptotic Results and Simulation Comparisons with Maximum Likelihood," NASA Technical Report SR-63-04427, June 1983.

Woodward, W. A., Parr, W. C., Schucany, W. R., and Lindsey, H. (1984). "A Comparison of Minimum Distance and Maximum Likelihood Estimation of a Mixture Proportion," Journal of American Statistical Association 79, 590-598.

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.

Table 1. Simulation Results for Mixtures of Normal Components

Sample Size = 100
Number of Realizations = 500

		Overlap = .10				Overlap = .03			
	σ_1/σ_2	% Convergence	$\hat{\text{Bias}}$	$n\hat{\text{MSE}}$	$\hat{E}$	% Convergence	$\hat{\text{Bias}}$	$n\hat{\text{MSE}}$	$\hat{E}$
P	Starts		.089	2.254			.051	.882	
	MLE	100.0	.063	5.003		100.0	.008	.449	
	MCVMDE	100.0	.142	8.944	.56	100.0	.028	1.029	.44
.25	MHDEN(1)	75.8	.100	4.255	1.18	92.4	.014	.604	.74
	MHDEN'(1)	78.6	.103	4.704	1.06	96.6	.011	.498	.90
	MHDEM(1)	99.6	.073	5.150	.97	100.0	.012	.557	.81
	MHDEM(2)	99.6	.091	6.301	.79	100.0	.014	.556	.81
	MHDE*	77.0	.083	3.848		83.6	.018	.522	
	Starts		.001	1.212			.000	.656	
.50	MLE	100.0	.007	3.158		100.0	.004	.412	
	MCVMDE	100.0	-.009	3.683	.86	100.0	.004	.440	.94
	MHDEN(1)	93.0	.003	2.213	1.43	98.8	.006	.495	.83
	MHDEN'(1)	97.6	.005	2.318	1.36	99.6	.005	.487	.85
	MHDEM(1)	99.8	.001	2.882	1.10	100.0	.005	.476	.87
	MHDEM(2)	99.8	.004	3.343	.94	100.0	.005	.406	1.01
	MHDE*	89.8	-.004	2.407		94.6	.003	.487	

Table 1. (Continued)

P	σ_1/σ_2	Overlap = .10 % Convergence	Biâs	nMŜE	Ê	Overlap = .03 % Convergence	Biâs	nMŜE	Ê
	Starts		-.011	.910			.012	.578	
	MLE	100.0	-.005	2.117		100.0	.005	.479	
	MCVMDE	100.0	.080	5.228	.40	100.0	.019	.831	.58
.25	MHDEN(1)	78.0	.000	1.538	1.38	91.0	-.001	.404	1.19
	MHDEN'(1)	81.6	.002	1.527	1.39	92.6	-.003	.432	1.11
	MHDEM(1)	99.8	.006	2.009	1.05	100.0	.003	.444	1.08
	MHDEM(2)	99.8	.028	2.969	.71	100.0	.011	.561	.85
	MHDE*	76.6	.001	1.483		87.4	.009	.390	
	Starts		.092	1.966			-.057	.936	
	MLE	100.0	-.014	2.584		100.0	-.002	.402	
	MCVMDE	100.0	.005	2.951	.88	100.0	.000	.393	1.02
	MHDEN(1)	92.4	-.023	1.784	1.45	99.2	-.007	.472	.85
.50	MHDEN'(1)	96.0	-.022	1.799	1.44	99.2	-.007	.469	.86
	MHDEM(1)	100.0	.001	2.257	1.14	100.0	-.003	.428	.94
	MHDEM(2)	100.0	.004	2.623	.99	100.0	-.002	.388	1.04
	MHDE*	92.0	-.033	1.963		96.0	-.011	.556	
	Starts		-.165	4.294			-.091	1.426	
	MLE	100.0	-.077	4.993		100.0	-.002	.512	
	MCVMDE	99.8	-.119	7.742	.64	100.0	-.022	1.020	.50
	MHDEN(1)	84.8	-.103	4.369	1.14	93.8	-.011	.506	1.01
.75	MHDEN'(1)	89.4	-.086	4.235	1.18	95.8	-.006	.438	1.17
	MHDFM(1)	99.0	-.075	5.022	.99	100.0	-.005	.440	1.16
	MHDEM(2)	99.6	-.076	5.743	.87	100.0	-.005	.501	1.02
	MHDE*	81.0	-.120	4.711		82.0	-.039	.978	

Table 2 - Simulation Results for Mixtures of t(4) Components

Sample Size = 100
Number of Replications = 500

P	σ_1/σ_2	Overlap = .10 % Convergence	Bias	nMSE	$\hat{E}$	Overlap = .03 % Convergence	Bias	nMSE	$\hat{E}$
	Starts		.068	1.418			.078	1.102	
	MLE	98.4	.069	5.725		99.6	.035	.823	
.25	MCVMDE	100.0	.066	4.144	1.38	100.0	.023	.428	1.92
	MHDEN(1)	74.2	.058	3.264	1.75	95.4	.015	.451	1.82
	MHDEN'(1)	78.2	.033	2.780	2.06	97.8	.013	.442	1.86
	MHDEM(1)	99.0	.037	4.460	1.28	100.0	.016	.576	1.43
	MHDEM(2)	99.4	.043	4.218	1.36	100.0	.020	.376	2.19
	MHDE*	79.0	.051	2.476		82.8	.029	.580	
	Starts		-.002	1.092			-.001	.554	
	MLE	99.4	.024	5.457		99.6	.003	.473	
	MCVMDE	100.0	.003	1.855	2.94	100.0	.001	.301	1.57
.50	MHDEN(1)	86.8	.015	3.525	1.55	99.0	-.001	.503	.94
	MHDEN'(1)	89.8	.013	3.326	1.64	98.2	.001	.485	.98
	MHDEM(1)	100.0	.011	5.296	1.03	100.0	.000	.462	1.02
	MHDEM(2)	99.6	.006	3.646	1.50	100.0	.000	.307	1.54
	MHDE*	87.4	.002	1.746		97.6	.004	.333	

Table 2 (continued)

		Overlap = .10				Overlap = .03			
P	σ_1/σ_2	% Convergence	Biâs	nMŜE	Ê	% Convergence	Biâs	nMŜE	Ê
	Starts		.006	.890			.038	.582	
	MLE	100.0	.090	5.080		99.6	.046	1.003	
	MCVMDE	100.0	.095	4.968	1.02	100.0	.031	.652	1.54
.25	MHDEN(1)	76.6	.017	2.210	2.30	91.6	.021	.534	1.88
	MHDEN'(1)	77.0	.004	2.008	2.53	92.2	.013	.539	1.86
	MHDEM(1)	100.0	.057	3.846	1.32	100.0	.021	.599	1.67
	MHDEM(2)	100.0	.091	4.676	1.09	100.0	.025	.566	1.77
	MHDE*	76.8	.024	2.038		82.2	.018	.456	
	Starts		-.082	1.698			-.046	.776	
	MLE	99.4	.024	5.030		99.8	.009	.456	
	MCVMDE	100.0	.025	1.978	2.54	100.0	.000	.254	1.80
.50	MHDEN(1)	86.6	-.003	3.189	1.58	99.0	.006	.418	1.09
	MHDEN'(1)	88.0	.003	3.100	1.62	99.2	.006	.382	1.19
	MHDEM(1)	100.0	.043	4.985	1.01	100.0	.008	.413	1.10
	MHDEM(2)	100.0	.046	3.486	1.44	100.0	-.002	.249	1.83
	MHDE*	84.6	-.049	2.316		97.0	-.008	.320	
	Starts		-.139	3.106			-.109	1.678	
	MLE	99.0	-.078	8.483		99.8	-.014	.519	
	MCVMDE	100.0	-.055	4.045	2.10	99.8	-.019	.396	1.31
.75	MHDEN(1)	76.2	-.043	3.918	2.17	98.0	-.008	.592	.88
	MHDEN'(1)	80.0	-.042	4.501	2.09	97.2	-.008	.592	.88
	MHDEM(1)	98.2	-.028	5.765	1.47	100.0	-.003	.383	1.36
	MHDEM(2)	99.4	-.036	5.091	1.67	100.0	-.014	.294	1.77
	MHDE*	76.8	-.099	4.241		78.2	-.049	1.053	

Table 3 - Percentage Convergence Obtained for Normal and t(4) Components

Overlap

	.10	.03
MHDEN(1)	82.4	95.8
MHDEN'(1)	85.6	96.8
MHDEM(1)	99.5	100.0
MHDEM(2)	99.7	100.0
MHDE*	82.1	88.1

Table 4 - Simulation Results for Mixtures of t(2) Components

Sample Size = 100
Number of Replications = 500

		Overlap = .10				Overlap = .03			
P	σ_1/σ_2	% Convergence	Biâs	nMŜE	Ê	% Convergence	Biâs	nMŜE	Ê
	Starts		.067	1.934			.095	1.342	
	MLE	89.0	.097	7.962		90.0	.069	1.555	
.25	MCVMDE	99.4	.079	3.745	2.13	100.0	.024	.328	4.74
	MHDEN(1)	71.0	.045	3.902	2.04	96.2	.022	.527	2.95
	MHDEM(2)	99.4	.097	5.782	1.38	100.0	.016	.289	5.38
	Starts		-.004	1.452			.001	.420	
	MLE	85.8	-.003	10.016		88.2	.004	.843	
.50	MCVMDE	99.2	.006	1.172	8.55	99.8	-.002	.282	2.99
	MHDEN(1)	73.8	-.005	4.607	2.17	99.2	-.001	.411	2.05
	MHDEM(2)	99.6	-.022	4.161	2.41	99.8	-.002	.286	2.95