United States
Department of
Agriculture

National
Agricultural
Statistics
Service

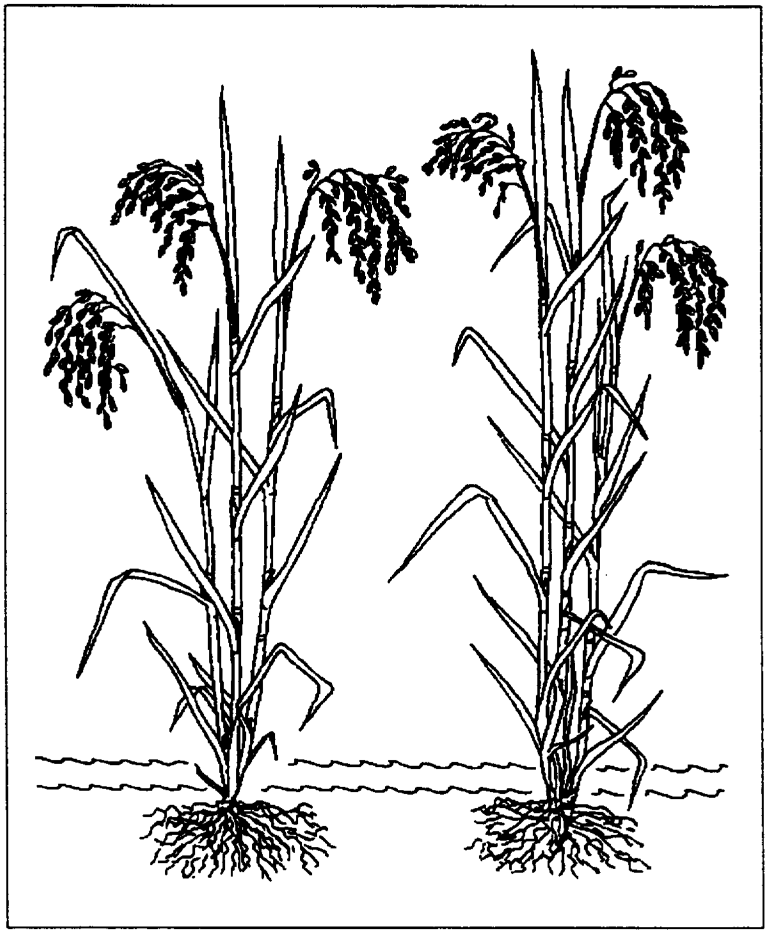
Research Division

STB Research Report
Number STB-94-02

December 1994

Application Of Satellite Data To Crop Area Estimation At The County Level

Michael E. Bellow



APPLICATION OF SATELLITE DATA TO CROP AREA ESTIMATION AT THE COUNTY LEVEL, by Michael E. Bellow, Research Division, National Agricultural Statistics Service, U.S. Department of Agriculture, Washington, D.C., December 1994. NASS Research Report No. STB-94-02.

ABSTRACT

This report describes the use of earth resources satellite data by the National Agricultural Statistics Service (NASS) to estimate crop planted area at the county or small domain level. The Battese-Fuller random effects model has recently been applied to obtain acreage indications submitted to NASS State Statistical Offices in the Mississippi Delta Region for input into their setting of official county estimates. This method extends the regression approach used for crop area estimation at the State level by incorporating an additional term that accounts for county effects. An alternative method, known as pixel count estimation, uses overall counts of satellite pixels classified to different crops and ground cover types within a county. The two methods are described and compared for several crops using Landsat Thematic Mapper (TM) data from Iowa (1988), Mississippi (1991-92) and Louisiana (1992). The satellite based estimates are compared with corresponding NASS official estimates.

KEY WORDS

Regression, Battese-Fuller, ratio, county effect

ACKNOWLEDGEMENTS

The author thanks Mitch Graham, Martin Ozga and Bill Mason for their assistance with data processing and analysis. Appreciation is due to the State Statistical Office staff members, NASDA enumerators and Remote Sensing Section support staff who performed various tasks associated with collection and preparation of the data. Thanks to Mike Craig, Paul Cook and Bill Iwig for their review and helpful suggestions. Finally, the author thanks Wendy Waters for her perserverance in the word processing of this report.

TABLE OF CONTENTS

SUMMARY iii

INTRODUCTION 1

BATTESE-FULLER ESTIMATION 3

PIXEL COUNT ESTIMATORS 6

EMPIRICAL EVALUATION 8

DISCUSSION AND RECOMMENDATIONS 17

REFERENCES 18

APPENDIX A: 20
HUDDLESTON-RAY AND CARDENAS ESTIMATORS

APPENDIX B: 21
ESTIMATION OF BATTESE-FULLER VARIANCE COMPONENTS

APPENDIX C: 21
DERIVATION OF VARIANCE FORMULA FOR SYNTHETIC ESTIMATOR

APPENDIX D: 22
DERIVATION OF VARIANCE FORMULA FOR COMBINED RATIO ESTIMATOR

APPENDIX E: 23
DISTRIBUTION-FREE PROCEDURES FOR COMPARISON OF TREATMENTS

SUMMARY

The focus of this report is the use of data from earth resources satellites to improve estimation of crop planted area at the county or small domain level. Large scale sample surveys conducted by the National Agricultural Statistics Service (NASS) for national and State level estimation are often inadequate for estimation at the county level due to insufficient sample sizes within counties. This fact has led to investigation and usage of auxiliary data sources such as list frame control data, previous year estimates and satellite data. NASS has studied the application of satellite data to crop area estimation at the county level since the mid 1970's.

The Battese-Fuller model, a random effects model, is currently used by NASS for small domain crop area estimation with combined survey and satellite data. The approach is based on the regression methodology used for estimation at the State level, but incorporates an additional term that models county effects. Alternative estimators based on overall counts of classified satellite pixels have recently begun to receive consideration. Two such estimators are the raw pixel count estimator (RPCE) and the combined ratio estimator (CRE).

An empirical study using data from Iowa, Mississippi and Louisiana compared the Battese-Fuller estimator (BFE) with the RPCE, CRE and a survey based synthetic estimator (SYN). Distribution-free test procedures were used to evaluate estimator accuracy. Overall, the BFE appeared to be the most accurate of the four estimators, but the CRE had lower variance. The performance of estimators was influenced by the specific crop and region. The BFE and CRE tended to be biased downward and the RPCE biased upward.

INTRODUCTION

The National Agricultural Statistics Service has published county level estimates of crop acreage, crop production, crop yield and livestock inventories since 1917. These estimates are used by government agencies for local economic decision making and by agribusinesses for marketing and sales purposes. The primary source of data for agricultural commodity estimates has always been surveys of farmers, ranchers and agribusinesses who voluntarily provide information on a confidential basis. However, surveys designed and conducted at the national and State levels are often inadequate for producing reliable information at the county or small domain level. The major obstacle to obtaining accurate small area estimates from large scale sample surveys is the fact that such areas usually contain very few sample units. Therefore, supplementary data sources such as NASS list frame control data, previous year estimates and Census of Agriculture data are often used to improve county estimates. In addition, special supplemental surveys are conducted by NASS State Statistical Offices (SSO's) for county estimation purposes. Earth resources satellite data, e.g., from the U.S. Landsat or French SPOT satellites, represents a useful ancillary data source for county level estimation of area planted in a crop. The potential for improved estimation accuracy using satellite data is based on the fact that, with adequate coverage, all area within a county can be classified to crops and other ground cover types. The accuracy of the estimates depends to a large degree upon how accurately the satellite data are classified to each crop.

Beginning in 1972, NASS has pursued the use of satellite data to improve crop area estimation in general. From 1978-87, the Agency used Landsat Multispectral Scanner (MSS) data to provide timely State level estimates of major crops in eight midwestern and south central States to NASS's Agricultural Statistics Board (ASB). From 1991-93, NASS used Landsat Thematic Mapper (TM) data to estimate cotton, rice and soybean planted acreage in the Mississippi Delta region. The States involved were Arkansas (1991-93), Louisiana (1992) and Mississippi (1991-92). Estimation at the State or region level is accomplished using a first order regression model relating ground survey data to satellite data. A limitation of the use of satellite data is difficulty obtaining cloud free imagery for a given region, in particular the Delta area which gets frequent rain.

The basic element of satellite spectral data is the set of measurements taken by a sensor of a square area on the earth's surface. The sensor measures the amount of radiant energy reflected from the surface in selected bands of the electromagnetic spectrum. The individual imaged areas, known as pixels, are arrayed along east-west rows within the north-to-south pass (swath) of the satellite. For purposes of easy storage, the data within a swath are subdivided into overlapping square blocks called scenes. The Landsat satellites image a given point on the earth's surface once every 16 days. The Landsat Multispectral Scanner (MSS) sensor contains four spectral bands with a spatial resolution of 80 meters. The more advanced Landsat TM sensor has seven bands (three visible and four infrared) with 30 meter resolution. The French SPOT Multispectral Scanner has three bands (two visible and one infrared) with 20 meter resolution.

The remote sensing approach to crop area estimation at the State, region and county levels depends upon the area frame portion of NASS's annual June Agricultural Survey (JAS). NASS's Area Frame Section divides the area of each State except Alaska into land use strata depending on degree of cultivation and other factors (Cotter and Nealon, 1987; Bush and House, 1993). During the JAS, enumerators interview farm operators within randomly selected area sampling units called segments and record the responses on questionnaire forms. At State Statistical Offices, the completed questionnaires are compiled and edited. The summarized results are used to generate State level estimates of crop acreage with measurable precision. These estimates are transmitted to the Agricultural Statistics Board in Washington, D.C., which is responsible for setting official estimates.

The number of satellite scenes required to cover a region of interest within a State depends upon the image area of the satellite. One cannot always have the same image dates for all scenes due to satellite overpass schedule, cloud cover and image quality factors. Consequently, following the JAS and scene acquisition, NASS's Remote Sensing Section (RSS) divides a State into smaller areas called analysis districts. An analysis district is either a collection of counties and parts of counties completely contained in one or more scenes having the same image date or an area for which usable satellite data are not available. State level (first domain) crop area estimates are obtained by summing all analysis district level (second domain) estimates within the State. County level (third domain) estimates can be computed using one of the methods described in this report. The Battese-Fuller approach is currently favored.

All satellite based county crop area estimators use stratum level or overall counts of pixels within a county classified to specific crops. Three regression based small domain estimation methods have been applied or considered by RSS. From 1976-82, the Huddleston-Ray estimator (Huddleston and Ray, 1976) was used. In 1978, the Cardenas family of estimators (Cardenas, Blanchard and Craig, 1978) was considered but not implemented. Appendix A provides a mathematical description of the Huddleston-Ray and Cardenas estimators. From 1982-87, the Battese-Fuller estimator was applied for county level estimation of major crops in the central United States using Landsat Multispectral Scanner (MSS) data. The same method was used to calculate county level estimates of rice, cotton, soybean and sugar cane acreage in the Mississippi Delta region in 1991-93 with Landsat TM data. Recently, non-regression estimators based on overall counts of classified pixels have begun to be studied. Two such estimators are discussed later.

Most data processing associated with satellite based crop area estimation is done using PEDITOR, a special purpose software system developed at NASS (Ozga, Mason and Craig, 1992). PEDITOR is written mainly in PASCAL and maintained on a Microvax 3500 computer, with many modules that also run on personal computers. A Cray supercomputer has been used for computationally intensive tasks.

For each analysis district having usable satellite coverage, a separate regression estimator (Cochran, 1977) is applied to compute crop area estimates that are more precise than the direct expansion estimates obtained from JAS data alone. Allen (1990) provides a detailed description of the methodology involved. The steps required are as follows:

1. Register each satellite scene to a map base (Cook, 1982).
2. Label each pixel within the sample segments to a crop or other ground cover type using JAS reported data.
3. Cluster sets of pixels corresponding to distinct cover types to create cover signatures, represented by discriminant functions (Bellow and Ozga, 1991).
4. Classify all pixels within the sample segments to cover types using cover signatures.
5. Develop regression relationships between JAS reported crop area (dependent variable) and corresponding counts of pixels classified to a given crop (regressor variable) on a per stratum basis.
6. Classify all pixels within the analysis district to cover types.
7. Generate stratum level crop area estimates by substituting analysis district level classified pixel averages into regression equations.
8. Sum stratum level estimates to obtain overall analysis district level crop area estimates.

The procedure uses direct expansion estimation in analysis districts lacking adequate satellite coverage. The final State level estimate for each crop is a composite of the regression and direct expansion estimates within the State.

In many States, counties typically contain fewer than five sampled JAS segments and may contain no segments at all. This fact makes it generally infeasible to define analysis districts to be individual counties and use the above procedure to obtain county level estimates. Instead, indirect estimators that utilize information from outside a county have been studied and applied.

While the word "county" is used in the upcoming discussion, the concepts apply to any small domain, i.e., an area for which insufficient ground survey information is available.

BATTESE-FULLER ESTIMATION

The Battese-Fuller approach to crop area estimation at the county level is based upon the regression methodology used for State level estimation (Allen, 1990; Graham, 1993). The Battese-Fuller estimator (BFE) uses the analysis district level regression, but invokes an additional term that accounts for county effects.

The Battese-Fuller model was first developed in the general framework of linear models with nested error structure (Fuller and Battese, 1973), and later applied to county crop area estimation (Battese, Harter and Fuller, 1988). As mentioned earlier, a set of counties and subcounties covered by one or more satellite scenes of the same date forms an analysis district, within which stratum level regression relationships between survey reported crop area and counts of classified pixels are developed. The Battese-Fuller model assumes that segments grouped by county have the same slope parameter as the analysis district, but a different intercept is required. The model can be applied within an analysis district for all strata where classification and regression have been done. The analyst computes stratum level Battese-Fuller crop area estimates for all counties and subcounties within the boundaries of each analysis district. For land use strata where regression cannot be done due to lack of adequate satellite coverage or too few segments, a domain indirect synthetic estimator that depends only on the ground survey data is used.

For a given analysis district, the strata where regression is done are referred to as regression strata and the remaining ones as synthetic strata. For convenience, the regression strata are labeled $h=1,\ldots,H_r$ and the synthetic strata $h=H_r+1,\ldots,H$, where H_r is the number of regression strata and H is the total number of strata in the analysis district. If any given county is partially contained in the analysis district, then the estimation formulas given below apply to the included portion.

For each sample segment within a given stratum h in county c, the Battese-Fuller model specifies the following relation:

$$y_{hci} = \beta_{0h} + \beta_{1h} x_{hci} + \nu_{hc} + \epsilon_{hci}$$

$$= \beta_{0h} + \beta_{1h} x_{hci} + \omega_{hci}, \quad i = 1,\ldots,n_{hc}$$

where:

n_{hc} = number of sample segments in stratum h, county c

y_{hci} = reported area in crop of interest in stratum h, county c, sample segment i

x_{hci} = number of pixels classified to crop of interest in stratum h, county c, sample segment i

ν_{hc} = county effect in stratum h, county c

ϵ_{hci} = random error in stratum h, county c, sample segment i

ω_{hci} = total error in stratum h, county c, sample segment i

β_{0h}, β_{1h} = analysis district level regression parameters for stratum h

This formulation is recognized as a random effects model. The county effect ν_{hc} and random error ϵ_{hci} are assumed to be independent and normal, with mean zero and variances σ^2_{vh} and σ^2_{eh}, respectively. The total error has covariance structure:

$$\begin{aligned} \mathrm{cov}(\omega_{hci}, \omega_{hc'i'}) &= 0 && \text{if } c \neq c' \\ &= \sigma_{vh}^2 && \text{if } c = c',\ i \neq i' \\ &= \sigma_{vh}^2 + \sigma_{eh}^2, && \text{if } c = c',\ i = i' \end{aligned}$$

The parameter σ_{vh}^2 is both a within county covariance and a between county component of the variance of any residual, while σ_{eh}^2 is the within county variance component for stratum h. The county mean residuals are observable and given by:

$$\bar{u}_{hc.} = \bar{y}_{hc.} - \hat{\beta}_{0h} - \hat{\beta}_{1h} \bar{x}_{hc.}$$

where:

$$\bar{y}_{hc.} = (1/n_{hc}) \sum_{i=1}^{n_{hc}} y_{hci}$$

$$\bar{x}_{hc.} = (1/n_{hc}) \sum_{i=1}^{n_{hc}} x_{hci}$$

$\hat{\beta}_{0h}, \hat{\beta}_{1h}$ = least squares regression parameter estimators for stratum h

For a given county, the stratum level mean crop area per population unit (segment) is estimated by:

$$\bar{y}^{(BF)}_{hc.} = \hat{\beta}_{0h} + \hat{\beta}_{1h} \bar{X}_{hc} + \delta_{hc} \bar{u}_{hc.}$$

where:

$\bar{X}_{hc}$ = mean number of pixels per population unit classified to crop in stratum h, county c

$$0 \leq \delta_{hc} \leq 1$$

The mean square error of this estimator is:

$$MSE\,(\bar{y}^{(BF)}_{hc.}) \;=\; (1-\delta_{hc})^2\sigma_{vh}^2 + \delta_{hc}^2(\sigma_{eh}^2/n_{hc})$$

The (unadjusted) stratum level estimator of total crop area in the county is:

$$\hat{T}^{(uBF)}_{hc} = N_{hc}[\hat{\beta}_{0h} + \hat{\beta}_{1h}\bar{X}_{hc} + \delta_{hc}\bar{u}_{hc.}]$$

where:

N_{hc} = number of population units in stratum h, county c

The range of allowed values of δ_{hc} defines a family of Battese-Fuller estimators. If $\delta_{hc}=0$, then the estimate lies on the analysis district regression line for the stratum. The value that minimizes the mean square error for stratum h in county i is (Walker and Sigman, 1982):

$$\delta^{*}_{hc} \;=\; n_{hc}\sigma_{vh}^2/(n_{hc}\sigma_{vh}^2 + \sigma_{eh}^2)$$

In general, the variance components σ_{vh}^2 and σ_{eh}^2 are unknown and must be estimated. The estimators given in Appendix B represent a special case of the more general unbiased estimators derived by Fuller and Battese (1973), using the "fitting-of-constants" method. They require that a stratum contain at least two sample segments within the county in question. If there are fewer than two segments, then δ_{hc} is set to zero in the computation of the county estimate.

The unadjusted estimates of county totals generally do not sum to the corresponding analysis district totals obtained from large scale estimation. In order to get agreement, adjustment terms must be added to the estimates. The formula for the adjusted Battese-Fuller estimator is:

$$\hat{T}^{(aBF)}_{hc} \;=\; \hat{T}^{(uBF)}_{hc} - (N_{hc}/N_h)\sum_{c=1}^{C}\delta_{hc}\bar{u}_{hc.}$$

where:

N_h = number of population units in stratum h

The adjusted Battese-Fuller estimator of total crop area in the regression strata of county c is:

$$\hat{T}^{(aBF)}_{c} \;=\; \sum_{h=1}^{H_r}\hat{T}^{(aBF)}_{hc}$$

Estimation of the variance of the BFE is discussed by Walker and Sigman (1982). Their estimator of mean square error is used to derive the variance estimator, but is known to have a downward bias due to estimation of the variance components. This bias can be significant and a correction due to Prasad and Rao (1990) may be implemented in the future.

The presence of a county main effect across strata introduces cross strata covariance and requires revisions in both the mean square error formula and the choice of an optimal set of multipliers for the county mean residuals. Walker and Sigman (1982) developed an extension of the above model that requires estimation of a vector of county effects by strata.

As mentioned previously, synthetic estimation is done in strata where regression is not viable. A given county usually contains few segments in a given stratum, so the stratum level mean crop acreage per segment for the analysis district is used to compute the synthetic estimates. In synthetic stratum h, the crop area in county c is estimated by:

$$\hat{T}^{(SYN)}_{hc} \;=\; N_{hc}\bar{y}_{h..}$$

where:

$\bar{y}_{h..}$ = mean reported crop area per sample segment in stratum h

The domain indirect synthetic estimator of total crop area in the synthetic strata of county c is then:

$$\hat{T}^{(SYN)}_{c} = \sum_{h=H_r+1}^{H} N_{hc}\, \bar{y}_{h..} \qquad (3.1)$$

An estimator of the variance, derived in Appendix C, is the following:

$$v(\hat{T}^{(SYN)}_{c}) = \sum_{h=H_r+1}^{H} N_{hc}^{2} s_{yh}^{2} (N_h - n_h)/N_h n_h$$

where:

$$s_{yh}^{2} = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} \sum_{c=1}^{C} (y_{hci} - \bar{y}_{h..})^2$$

n_h = number of sample segments in stratum h

Synthetic estimation's use of prorated district level averages to estimate county totals ignores county effects. Therefore, the synthetic component of a county estimate can have a significant bias. The bias is reduced if a synthetic district has homogeneous agricultural intensity for the crop of interest.

The final county estimate is the sum of the regression and synthetic components:

$$\hat{T}_{c} = \hat{T}^{(aBF)}_{c} + \hat{T}^{(SYN)}_{c}$$

The estimated variance of the final county estimates is computed by summing the variance estimates of the regression and synthetic components.

Walker and Sigman (1982) studied the Battese-Fuller model using Landsat MSS data for a six county region in eastern South Dakota. At that time, NASS was using the Huddleston-Ray estimator (Appendix A). They found a modest lack of fit of the model, with larger model departure corresponding to low correlations between classified pixel counts and ground survey observations. The county effect parameter was found to be highly significant for corn, the most prevalent of the four crops considered in the research study. Furthermore, this effect manifested itself within several strata but was negligible across strata. The study nonetheless showed robustness of the Battese-Fuller estimators against departure from certain model assumptions. Two members of the Battese-Fuller family satisfied the criterion for small relative root mean square error; i.e., less than 20 percent of the estimate was due to root mean square error. These estimators were the ones that minimized mean square error and bias, respectively, under the model assumptions. However, the Battese-Fuller estimate closest to the Huddleston-Ray estimate was far less satisfactory, failing to meet the desired upper limits for mean square error and bias. This study provided the justification for adopting the Battese-Fuller estimator as a replacement for the Huddleston-Ray estimator.

PIXEL COUNT ESTIMATORS

As improved satellite sensors enable higher classification accuracy, the overall (across strata) count of pixels within an area classified to a given crop has become a more interesting number. The overall pixel count represents a "census" of pixels covering the area in question and therefore is not subject to sampling error. However, there is a nonsampling error due to pixel misclassification. As a result, the estimator defined as the overall pixel count converted to area units can have a significant bias. Bias reduction is achieved by using an adjustment factor based on sample level information. Although a pixel count estimator could be a function of counts of pixels classified to many different cover types, this discussion is restricted to estimators based on the number of pixels classified to the crop of interest only. A general expression for such an estimator is:

$$\hat{T}_c = \eta X_c$$

where:

X_c = number of pixels classified to crop of interest in county c

η = adjustment term

The adjustment term could be a function of the sample level classification data. The choice of adjustment term determines the specific estimator to be used. The raw pixel count estimator (RPCE), alluded to above, is obtained by setting the adjustment term to the area on the ground corresponding to a single pixel:

$$\hat{T}_c^{(RPC)} = \lambda X_c$$

where λ is the conversion factor (area units per pixel) for the satellite sensor in use.

By considering only the specific pixel classification obtained and not the superpopulation consisting of all possible classifications for a given data set, one can assume that the RPCE has zero variance. The bias depends upon the difference between the theoretical commission error (probability of a pixel from another cover type being classified to the crop of interest) and omission error (probability of a pixel from the crop of interest being classified to another cover type).

The combined ratio estimator (CRE) is based on the estimator of the same name described in Cochran (1977). This estimator is conceptually simple, uses stratum level information to compute the adjustment term and has a simple formula for estimating the variance. The CRE can be expressed as follows:

$$\hat{T}_c^{(CR)} = [(\sum_{h=1}^{H} N_h \bar{y}_{h..})/(\sum_{h=1}^{H} N_h \bar{x}_{h..})]X_c$$

$$= \hat{R}X_c$$

where:

$\bar{x}_{h..}$ = mean number of pixels per sample segment classified to crop in stratum h

Since the adjustment term is based on sample level information, this estimator has a positive sampling error. An estimator for the variance of the CRE, valid for large samples, is the following (see Appendix D for derivation):

$$v[\hat{T}_c^{(CR)}] = [X_c/X]^2 \sum_{h=1}^{H} [(N_h^2(1-f_h)/n_h](s_{yh}^2 + \hat{R}^2 s_{xh}^2 - 2\hat{R}s_{xyh})$$

where:

$$s_{xh}^2 = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} \sum_{c=1}^{C} (x_{hci} - \bar{x}_{h..})^2$$

$$s_{xyh} = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} \sum_{c=1}^{C} (x_{hci} - \bar{x}_{h..})(y_{hci} - \bar{y}_{h..})$$

$$f_h = n_h/N_h$$

x = total number of pixels classified to crop in analysis district

EMPIRICAL EVALUATION

The performance of the satellite based county crop area estimators was evaluated using data from Iowa, Louisiana and Mississippi. The Iowa data, from 1988, had been used previously for a sensor comparison study (Bellow, 1991). The Mississippi and Louisiana data came from NASS's 1991-1992 operational projects in the Mississippi Delta region (Craig, 1993). Table 1 gives information about the four regions used in the study. The number of scene dates for a given region indicates whether the analysis for that region was unitemporal (one date) or multitemporal (two dates).

Region A is a crop reporting district in western Iowa with a high concentration of corn and soybeans. Parts of Calhoun, Crawford and Ida counties were outside the TM scene used. Region B is comprised of two contiguous crop reporting districts in northwest Mississippi which accounted for most of the rice and cotton produced in the State in 1991 and 1992. The same area sampling frame was in effect in both years, allowing for a direct year-to-year comparison of county estimates. The slight increase in number of segments from 1991 to 1992 was due to NASS's annual rotation of segments into and out of a State's area sample. All areas except for a small part of Yazoo county were covered by the TM scenes. Region C is a crop reporting district in southwest Louisiana that accounted for 56 percent of the State's rice production in 1992. Region D is a crop reporting district in south central Louisiana that accounted for 56 percent of the State's sugar cane production in 1992. Part of Vermilion parish was outside the TM scenes.

Table 1: Regions in Empirical Study

			Number of				
Region	Location	Year	Counties	Strata	Segments	Scenes	Scene Dates
A	W Iowa	1988	9	2	30	1	7/25
B	NW Mississippi	1991	12	6	79	4	4/1, 8/23
		1992	12	6	82	4	5/5, 7/24
C	SW Louisiana	1992	7	5	51	2	4/26, 8/16
D	SC Louisiana	1992	6	4	21	1	5/5

In each region, all available TM spectral bands were used. For Iowa, the analysis used all 30 segments, with 28 coming from stratum 14 (agricultural) and the remaining two from stratum 30 (agri-urban). Data from the segments in stratum 14 were used for regression and Battese-Fuller estimation. For the BFE, CRE and RPCE, synthetic estimation was used within stratum 14 outside the scene. For the BFE, synthetic estimation was used in stratum 30 for all areas.

In Mississippi, Battese-Fuller estimation was used for cotton in strata 11 (75 - 100% cultivated), 12 (51 - 75% cultivated), 20 (15 - 50%) cultivated) and 40 (0 - 15% cultivated). For rice, the BFE was computed only in strata 11 and 12 due to an insufficient number of segments with positive reported rice acreage in the other strata. Synthetic estimation was used in the remaining strata for each crop. In Louisiana, Battese-Fuller estimation was used in strata 13 (50 - 100% cultivated) and 20 (15 - 50% cultivated) for both rice and sugar cane.

Tables 2-5 give the computed values of the satellite based BFE, CRE and RPCE for Iowa, Mississippi and Louisiana, respectively. For comparison, values of the pure synthetic estimator (SYN) are also shown. This survey based estimator uses synthetic estimation in all strata and is defined by equation (3.1) with $H_r=0$. Estimated standard deviations are shown for the SYN, BFE and CRE. The official county or parish acreage estimates (OFF) issued by NASS's Iowa, Mississippi and Louisiana State Statistical Offices are also shown. These published estimates are benchmark data against which the accuracy of satellite based estimates can be assessed. Rice figures are given only for 1992 in Issaquena and Yazoo counties since Mississippi did not issue official 1991 rice estimates for those two counties.

Table 2: County Estimates for Iowa 1988 (1000 Acres)

CORN:

County	*OFF*	*SYN*	*SD*	*BFE*	*SD*	*CRE*	*SD*	*RPCE*
Audubon	100.0	112.4	6.5	92.2	3.2	93.6	2.1	100.6
Calhoun*	133.0	144.9	8.3	133.2	3.9	134.4	2.9	144.2
Carroll	141.0	146.2	8.4	141.4	4.5	142.1	3.1	152.6
Crawford*	147.0	183.2	10.6	152.7	4.7	155.1	3.2	164.9
Greene	125.0	145.9	8.4	130.0	3.9	132.8	2.9	142.7
Guthrie	98.0	151.3	8.7	106.3	5.2	107.8	2.4	115.8
Ida*	112.0	111.4	6.4	107.0	4.0	107.0	3.8	110.3
Sac	136.0	148.1	8.5	138.3	4.0	139.6	3.1	150.0
Shelby	155.0	149.4	8.6	140.7	4.0	141.5	3.1	152.1

SOYBEANS:

County	*OFF*	*SYN*	*SD*	*BFE*	*SD*	*CRE*	*SD*	*RPCE*
Audubon	70.7	74.0	7.5	69.9	4.6	70.4	2.1	74.8
Calhoun*	150.0	95.4	9.6	145.0	5.8	136.9	4.0	145.2
Carroll	117.0	96.1	9.7	106.7	9.7	106.4	3.1	113.0
Crawford*	106.0	120.4	12.1	106.9	5.8	108.1	3.1	113.8
Greene	143.0	96.1	9.7	117.5	5.4	109.6	3.2	116.3
Guthrie	77.5	99.5	10.0	64.4	7.0	78.8	2.3	83.7
Ida*	75.2	73.3	7.4	76.4	5.3	76.1	4.3	78.2
Sac	124.0	97.3	9.8	112.9	5.5	108.8	3.2	115.5
Shelby	94.9	98.3	9.9	81.0	6.0	91.1	2.7	96.7

* - incomplete satellite coverage

Table 3: County Estimates for Mississippi 1991 (1000 Acres)

COTTON:

County	*OFF*	*SYN*	*SD*	*BFE*	*SD*	*CRE*	*SD*	*RPCE*
Bolivar	65.5	106.2	15.4	61.6	6.1	60.6	3.9	80.6
Coahoma	105.7	59.2	8.4	88.3	4.2	82.6	5.2	109.8
Humphreys	61.6	53.2	7.2	57.3	3.4	54.2	3.4	72.1
Issaquena	38.0	42.6	8.6	34.6	3.9	27.5	1.8	36.6
Leflore	79.2	68.8	9.6	87.8	3.5	83.4	5.3	111.0
Quitman	31.0	48.1	7.2	46.4	4.0	44.5	2.8	59.3
Sharkey	47.0	43.2	6.9	48.6	3.4	42.5	2.7	56.6
Sunflower	100.0	95.6	15.0	79.3	5.5	73.9	4.7	98.3
Tallahatchie	64.2	68.9	10.5	67.9	4.9	60.3	3.8	80.3
Tunica	45.6	47.1	6.9	38.0	2.5	36.5	2.3	48.6
Washington	95.7	84.4	11.6	102.4	4.0	93.2	5.9	124.1
Yazoo*	94.5	89.3	23.4	93.9	7.5	82.6	5.2	109.6

RICE:

County	*OFF*	*SYN*	*SD*	*BFE*	*SD*	*CRE*	*SD*	*RPCE*
Bolivar	74.0	50.8	11.9	66.2	3.6	66.9	6.1	60.9
Coahoma	15.8	20.3	4.7	10.4	2.5	10.7	1.0	9.7
Humphreys	3.6	22.8	5.2	7.1	2.3	4.7	0.4	4.3
Leflore	16.6	30.7	7.1	19.4	3.6	17.3	1.6	15.8
Quitman	9.6	24.4	5.6	9.3	2.8	6.2	0.6	5.6
Sharkey	5.0	18.0	4.1	7.8	1.7	6.5	0.6	5.9
Sunflower	36.0	51.1	12.0	37.8	3.5	36.7	3.4	33.4
Tallahatchie	9.6	20.9	5.1	8.5	3.0	8.1	0.7	7.4
Tunica	17.5	17.6	4.3	9.9	2.6	13.0	1.2	11.9
Washington	30.5	39.6	9.0	22.6	3.5	28.0	2.6	25.4

* - incomplete satellite coverage

Table 4: County Estimates for Mississippi 1992 (1000 Acres)

COTTON:

County	*OFF*	*SYN*	*SD*	*BFE*	*SD*	*CRE*	*SD*	*RPCE*
Bolivar	66.0	117.8	15.5	81.6	7.1	75.8	4.1	95.3
Coahoma	106.2	61.1	8.6	70.5	6.1	72.6	4.5	79.6
Humphreys	63.5	58.2	7.6	43.4	4.8	44.9	3.0	58.5
Issaquena	38.2	40.6	6.9	35.8	3.0	34.5	2.3	45.0
Leflore	95.0	74.8	9.8	86.0	5.5	82.1	5.5	107.1
Quitman	34.0	52.8	7.3	40.2	7.4	37.9	2.4	41.6
Sharkey	53.0	44.2	5.9	54.2	3.9	50.4	3.4	65.7
Sunflower	103.3	109.6	15.3	62.9	7.8	62.8	4.0	81.1
Tallahatchie	73.0	69.7	10.7	57.3	6.6	58.5	3.3	65.3
Tunica	44.0	47.8	6.6	31.9	4.3	33.6	2.1	36.9
Washington	95.0	92.4	11.8	92.8	6.1	88.2	5.9	115.0
Yazoo*	98.5	82.8	23.5	88.0	4.1	83.3	5.7	106.7

RICE:

County	*OFF*	*SYN*	*SD*	*BFE*	*SD*	*CRE*	*SD*	*RPCE*
Bolivar	82.0	31.4	7.1	50.1	8.1	49.8	5.0	69.5
Coahoma	18.0	15.5	4.0	6.5	2.5	6.4	0.2	7.7
Humphreys	7.6	15.8	3.6	12.2	6.2	14.2	1.6	20.1
Issaquena	4.1	8.8	2.4	7.2	4.5	17.0	2.0	24.2
Leflore	18.0	21.2	5.0	17.2	7.9	24.4	2.8	34.7
Quitman	18.5	15.8	3.8	6.6	2.6	3.8	0.1	4.6
Sharkey	10.6	11.7	2.6	12.0	5.6	12.8	1.5	18.2
Sunflower	44.0	30.5	7.0	37.4	7.8	35.0	3.7	49.1
Tallahatchie	14.0	17.1	4.8	7.3	2.9	5.9	0.2	7.4
Tunica	19.5	12.9	3.3	7.7	2.7	8.9	0.2	10.7
Washington	30.5	25.7	5.7	28.1	8.0	39.7	4.6	56.4
Yazoo*	1.2	13.1	8.5	1.0	2.1	8.3	1.2	11.4

* - incomplete satellite coverage

Table 5: Parish Estimates for Louisiana 1992 (1000 Acres)

RICE:

County	*OFF*	*SYN*	*SD*	*BFE*	*SD*	*CRE*	*SD*	*RPCE*
Acadia	96.0	53.9	8.7	87.5	1.7	80.4	3.0	92.3
Allen	28.0	11.8	1.9	16.1	0.4	20.3	0.8	24.7
Beauregard	4.0	7.0	1.4	3.6	0.5	8.3	0.3	10.1
Calcasieu	36.0	46.6	7.4	28.7	1.1	34.9	1.4	42.4
Cameron	15.0	12.3	2.0	11.8	0.3	61.8	2.6	75.1
Jeff. Davis	97.0	56.5	9.3	86.8	1.3	83.2	3.5	101.1
Vermilion[a]	102.0	61.1	10.0	96.8	3.5	149.6	9.9	152.4

SUGAR CANE:

County	*OFF*[b]	*SYN*	*SD*	*BFE*	*SD*	*CRE*	*SD*	*RPCE*
Assumption	35.0	7.7	1.8	24.4	2.0	21.5	2.2	29.6
Iberia	53.7	12.8	3.0	54.8	2.0	45.6	4.6	63.0
Iberville	31.4	9.2	2.1	28.6	1.3	23.7	2.4	32.7
Lafayette	8.9	13.1	3.0	12.3	1.8	10.3	1.0	14.2
St. Martin	29.5	15.2	3.6	21.3	2.6	25.5	2.6	35.2
St. Mary	42.0	11.2	2.6	51.4	1.5	39.4	4.0	54.4

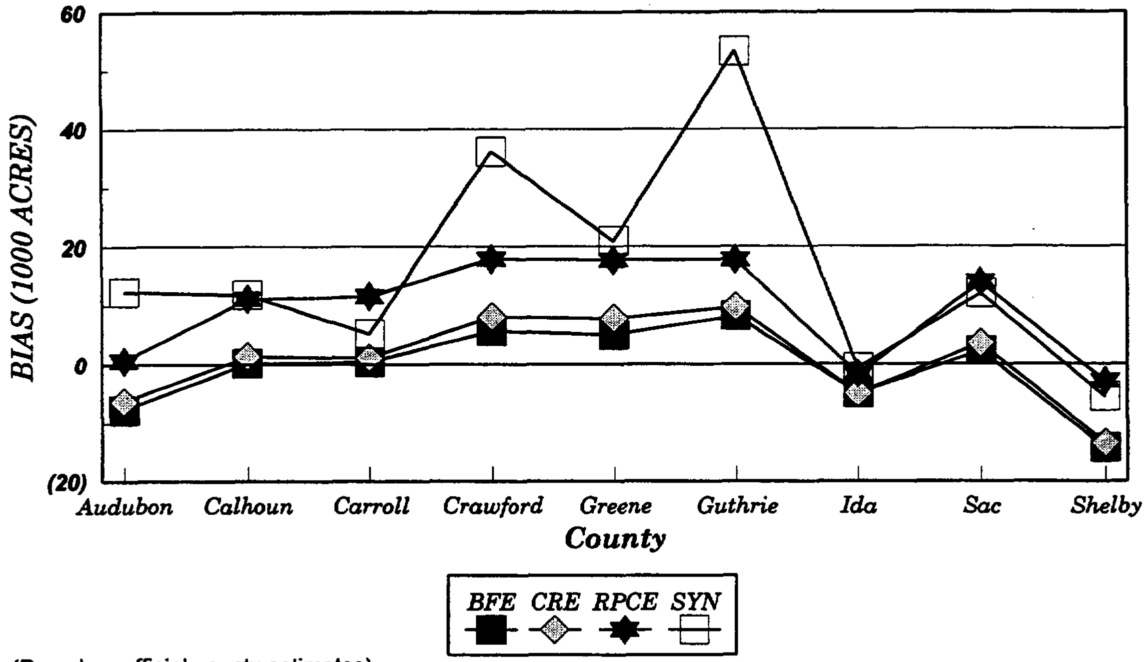
a - incomplete satellite coverage
b - harvested area

Figures 1 and 2 show measured bias of the BFE, CRE, RPCE and SYN for corn and soybeans in Region A. The bias measurements are simply the county level differences between each crop area estimate and the official estimate.

To determine if any of the four estimators were significantly different from the official county estimates for these data sets, a distribution-free Friedman test was applied (Hollander and Wolfe, 1973). Appendix E gives a description of the procedure. The subjects were counties in a given data set, while the treatments were the five types of estimate (OFF, BFE, RPCE, CRE, SYN).

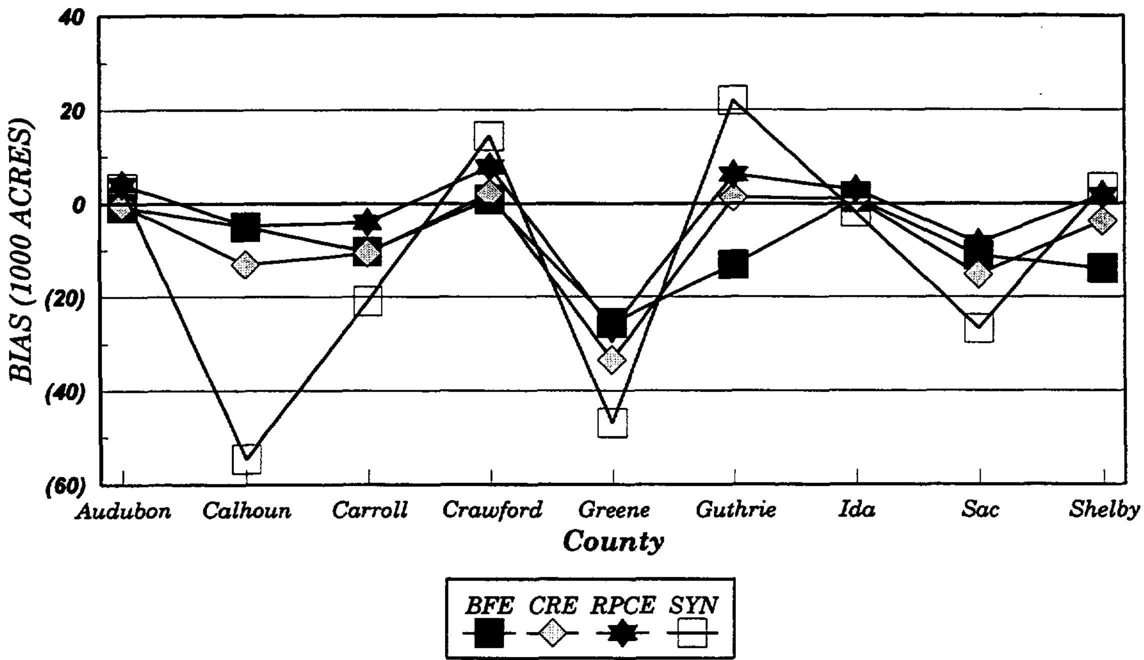
The null hypothesis states that the treatment effects are equal, i.e., the five types of estimate are not significantly different. This hypothesis is rejected if the test statistic S is sufficiently large. Table 6 gives the rank sums and values of the Friedman S statistic for each of the eight data sets. Upper bounds on p-values, computed using a χ^2_{k-1} approximation to the null distribution of S, are also shown. The upper bounds show that the null hypothesis of equal treatment effects is rejected at the $\alpha = .1$ level for all eight data sets.

Figure 1: Measured Bias For Iowa 1988 Corn



(Based on official county estimates)

Figure 2: Measured Bias For Iowa 1988 Soybeans



(Based on official county estimates)

Table 6: Rank Sums and Friedman S Values

			Rank Sums						
State	*Year*	*Crop*	*OFF*	*BFE*	*RPCE*	*CRE*	*SYN*	*S*	*p-value*
IA	1988	Corn	19	15	37	24	40	21.6	<.005
IA	1988	Soybeans	31	22	37	21	24	8.27	<.1
MS	1991	Cotton	38	35	56	18	33	24.6	<.005
MS	1991	Rice	31	29	15	29	46	19.36	<.005
MS	1992	Cotton	43	28	50	22	37	16.87	<.005
MS	1992	Rice	38	25	49	31	37	10.67	<.05
LA	1992	Rice	25	13	32	21	14	14.29	<.01
LA	1992	Sugar Cane	20	19	29	13	9	15.47	<.005

Since the Friedman tests concluded significant differences for all eight data sets, distribution-free multiple comparison tests were performed to determine which estimators, if any, were significantly different from the official estimates. The treatments vs. control procedure described in Appendix E was used, with the official estimates playing the role of control treatment. Table 7 gives the absolute differences between rank sums. The critical test values M(.05) and M(.01) at the .05 and .01 levels, respectively, are also shown.

Table 7: Rank Sum Absolute Differences (Four Estimators vs. OFF) and Critical Test Values

			Rank Sum Absolute Differences					
State	*Year*	*Crop*	*BFE*	*RPCE*	*CRE*	*SYN*	*M(.05)*	*M(.01)*
IA	1988	Corn	4	18	5	21	16.4	20.1
IA	1988	Soybeans	9	6	10	7	16.4	20.1
MS	1991	Cotton	3	18	20	5	18.9	23.2
MS	1991	Rice	2	16	2	15	17.3	21.2
MS	1992	Cotton	15	7	21	6	18.9	23.2
MS	1992	Rice	13	11	7	1	18.9	23.2
LA	1992	Rice	12	7	4	11	14.4	17.7
LA	1992	Sugar Cane	1	9	7	11	13.4	16.4

The table shows that the RPCE was significantly different from OFF at the .05 level for Iowa 1988 corn, while the CRE was significantly different at the .05 level for Mississippi cotton in both 1991 and 1992. The survey based estimator SYN was significantly different from OFF at the .01 level for Iowa 1988 corn. Of the four estimators, the BFE was the only one found not significantly different from OFF for all eight data sets.

To quantify the differences between estimators and gain some insight into bias, Doksum contrast estimates were generated. These contrasts between treatments were computed using medians of pairwise differences between estimates. The procedure is described in Appendix E. For each data set, Table 8 gives the contrast estimates corresponding to differences between treatment effects of the four estimators and the treatment effect of the official estimates. The signs of the contrast estimates indicate direction of bias, i.e., positive values signify overestimation and negative values underestimation. The magnitudes of the contrast estimates generate an accuracy ranking of the four estimators for a given data set. These ranks are shown in parentheses in the table.

Table 8: Doksum Contrast Estimates - Four Estimators vs. OFF (rankings of magnitudes in parentheses)

			Contrast Estimates							
State	*Year*	*Crop*	*BFE*		*RPCE*		*CRE*		*SYN*	
IA	1988	Corn	.34	(1)	11.4	(3)	1.58	(2)	12.18	(4)
IA	1988	Soybeans	-5.2	(3)	.62	(1)	-4.24	(2)	-5.38	(4)
MS	1991	Cotton	-1.92	(1)	12.2	(4)	-7.26	(3)	-2.47	(2)
MS	1991	Rice	-.66	(1)	-2.88	(3)	-1.55	(2)	11.24	(4)
MS	1992	Cotton	-9.26	(3)	3.89	(2)	-10.74	(4)	-1.04	(1)
MS	1992	Rice	-4.05	(3)	4.95	(4)	-1.93	(2)	-.77	(1)
LA	1992	Rice	-7.98	(2)	9.44	(3)	.82	(1)	-20.78	(4)
LA	1992	Sugar Cane	-1.9	(1)	3.25	(2)	-6.18	(3)	-21.12	(4)

The table shows that the contrast estimate of the BFE had the lowest magnitude for four of the eight data sets, and the second lowest magnitude for one other data set. The CRE had the lowest magnitude contrast estimate for one data set and the second lowest for four others. SYN had the highest magnitude contrast estimate for five data sets. These results provide further evidence that none of the other three estimators is more accurate that the BFE. The contrast estimate for the BFE was negative for seven data sets and that of SYN for six, while the contrast estimate for the RPCE was positive for seven data sets. These observations suggest negative bias tendencies for the BFE and SYN, and a positive bias tendency for the RPCE.

Tables 9-12 give four measures of estimator accuracy for all data sets, computed based on the final official figures. The measures are mean deviation from official estimates (MD), root mean square deviation (RMSD), mean absolute deviation (MAD) and largest absolute deviation (LAD).

Table 9: Iowa 1988 Estimator Accuracy (1000 Acres)

	Corn				*Soybeans*			
EST	*MD*	*RMSD*	*MAD*	*LAD*	*MD*	*RMSD*	*MAD*	*LAD*
BFE	-0.6	6.8	5.4	14.3	-8.6	11.9	9.1	25.5
RPCE	9.6	12.6	10.6	17.9	-2.3	10.3	7.4	26.7
CRE	0.8	7.4	6.3	13.5	-8.0	13.5	9.0	33.4
SYN	16.2	23.8	17.6	53.3	-12.0	28.0	21.6	54.6

Table 10: Mississippi 1991 Estimator Accuracy (1000 Acres)

	Cotton				Rice			
EST	*MD*	*RMSD*	*MAD*	*LAD*	*MD*	*RMSD*	*MAD*	*LAD*
BFE	-1.8	10.0	7.8	20.7	1.9	4.9	4.1	7.9
RPCE	13.2	17.2	13.8	31.8	-3.8	5.4	4.1	13.1
CRE	-7.2	12.5	10.1	26.1	-2.0	3.5	2.8	7.1
SYN	-1.8	19.4	13.2	46.5	7.8	14.0	12.4	23.2

Table 11: Mississippi 1992 Estimator Accuracy (1000 Acres)

	Cotton				Rice			
EST	*MD*	*RMSD*	*MAD*	*LAD*	*MD*	*RMSD*	*MAD*	*LAD*
BFE	-10.4	18.7	14.3	40.4	-6.2	11.4	7.7	31.9
RPCE	2.3	16.0	13.8	29.3	3.8	13.8	12.5	25.9
CRE	-12.1	18.3	14.4	40.5	-3.5	13.0	10.9	32.2
SYN	-1.5	22.2	15.3	51.8	-4.0	16.0	9.4	50.6

Table 12: Louisiana 1992 Estimator Accuracy (1000 Acres)

	Sugar Cane				Rice			
EST	*MD*	*RMSD*	*MAD*	*LAD*	*MD*	*RMSD*	*MAD*	*LAD*
BFE	-1.3	6.9	5.9	10.6	-6.7	7.6	6.7	11.9
RPCE	4.8	7.4	6.6	12.4	17.2	29.9	19.2	60.1
CRE	-5.7	7.4	6.2	13.5	8.6	26.6	19.6	47.6
SYN	-21.9	26.1	23.3	40.9	-18.4	28.0	22.3	42.1

These figures tend to support the previous conjecture that none of the other three estimators is more accurate than the BFE. For example, the root mean square deviation of the BFE was lower than that of the other three estimators for five of the eight data sets. The results also suggest region and crop specific aspects to performance of the three satellite based estimators. To illustrate, the CRE showed much lower values of RMSD and MAD for rice in Mississippi than for the same crop in Louisiana, while both the BFE and CRE showed noticeably lower values of the same two measures for corn in Iowa than for soybeans in the same State.

Table 5 shows that in the Louisiana parishes of Cameron and Vermilion, the RPCE and CRE severely overestimated the official rice acreage while the BFE was much more accurate. The reason for this anomaly is the fact that in stratum 40 (less than 15 percent cultivated) some wetland area was misclassified as rice in both parishes. Since the RPCE and CRE are computed from the pixel count in all strata combined, they were both grossly high for rice. The Battese-Fuller approach used synthetic estimation in stratum 40 and thus circumvented the difficulty. This situation indicates that one must use caution when applying pixel based estimation. Table 8 shows that the CRE had the lowest magnitude contrast estimate for the same data set. This result is misleading since the medians of paired differences were not greatly influenced by the extreme outliers. Table 12 reflects the situation more accurately, showing the BFE with a root mean square deviation of 7,600 acres compared with 29,900 acres for the RPCE and 26,600 acres for the CRE.

The mean deviation of the BFE was negative for seven of the eight data sets, supporting the earlier conjecture of negative bias. Table 5 shows that for both rice in Region C and sugar cane in Region D, the BFE had a lower variance than the CRE in most parishes. However, that was not the case for the Region A and B data sets.

The RPCE's mean deviation was positive for six of the eight data sets, supporting a positive bias. Surprisingly, this estimator showed the lowest RMSD and MAD for soybeans in Region A and cotton in Region B (1992).

There is insufficient evidence to make any statement regarding direction of bias for the CRE. From Table 1, the CRE had lower variance than the BFE for both corn and soybeans in all Region A counties. Furthermore, in most counties the CRE showed lower variance than the BFE for rice and cotton in Region B (1991 - 92).

The mean deviation of SYN was negative for six data sets, supporting negative bias. For both cotton and rice in Region B (1991), SYN had higher variance than the BFE and CRE in each county. SYN did display lower variance than the BFE in seven of twelve counties for rice in Region B (1992).

DISCUSSION AND RECOMMENDATIONS

Based on the results of the empirical study, the Battese-Fuller method should continue to be used for county level crop area estimation with satellite data. However, research on other estimators will continue. Estimators not considered in this report, such as an indirect separate ratio estimator, could be investigated.

The future usage of satellite data for county crop area estimation at NASS must be evaluated in the context of the Agency's overall remote sensing and county estimation programs. Remote sensing research will continue to focus on identifying new geographic areas and crops where the methodology would be beneficial. In addition, benefits of other sources of remotely sensed data, such as radar satellites, will be studied as data become available. From 1991-93, Landsat TM data were used to produce State and county level crop area estimates in the Delta region. In 1993, satellite data were used only for Arkansas due to budgetary constraints. The Remote Sensing Section is using both French SPOT multispectral data and Landsat TM data for crop area estimation in Arkansas in 1994. This effort enables a comparison to be made between utility of the two sensors for this application. As the prototype field office for remote sensing, the Arkansas SSO is receiving special computing capabilities and training related to processing of earth resources satellite data. An empirical study comparing Landsat based Arkansas county estimates with various county indications used by the SSO is planned. There are 3-year (1991-93) time series available for remote sensing estimates and the other data sources. Furthermore, in 1995 there will be no full State project in Arkansas. Instead, NASS plans to perform crop acreage verification and crop condition/yield assessment using extra area frame sample segments in Craighead County. Multiple satellite data sources and acquisition dates will be utilized.

NASS State Statistical Offices rely on a number of data series to help set official county estimates of crop acreage and production (Iwig, 1993). Fairly reliable administrative data sources are available for many commodities. The new NASS County Estimate System, developed in an effort to standardize county estimation methods across States, uses a combination of scaling and compositing techniques to provide a county level total estimate for any agricultural commodity (Bass et al., 1989). Separate estimates that can be composited together include previous year official estimates, current year direct expansion and ratio estimates, Agricultural Stabilization and Conservation Service (ASCS) figures and satellite based estimates. The SSO' s conduct a large non-probability county estimates survey that also provides updated control data for NASS's list sampling frame. This survey is an integral part of the Agency's overall survey program and will continue in some form in the foreseeable future.

There is an ongoing NASS cooperative research project in small area estimation not involving satellite data. That project is evaluating various mixed effects models for estimating county level crop production from the State-wide non-probability sample of farms (Stasny, Goel and Rumsey, 1991). A current research effort compares pixel estimators with stratum level regression estimators for State and region level crop area estimation (Bellow, 1994). In addition to large domain versions of the RPCE and CRE, combined regression and separate ratio estimators are considered.

A byproduct of satellite data that is useful to State Statistical Offices is a set of color coded land use maps at the county level. These maps provide a pictorial view of the distribution of crops within each county, based on the satellite pixel classifications. Recently, a capability to display roads and water bodies on the maps using digital line graph (DLG) data has been added. The Arkansas SSO provides feedback to NASS's Remote Sensing Section concerning the satellite based county estimates and maps.

REFERENCES

Allen, J. D. (1990), "A Look at the Remote Sensing Applications Program of the National Agricultural Statistics Service," Journal of Official Statistics, 6(4), 393-409.

Bass, J., B. Guinn, B. Klugh, C. Ruckman, J. Thorson and J. Waldrop (1989), Report of the Task Group for Review and Recommendations on County Estimates, National Agricultural Statistics Service, U.S. Department of Agriculture.

Battese, G.E., R.M. Harter and W.A. Fuller (1988), "An Error-Components Model for Prediction of County Crop Areas using Survey and Satellite Data," Journal of the American Statistical Association, 83(401), 28-36.

Bellow, M.E. (1991), Comparison of Sensors for Corn and Soybean Planted Area Estimation, U.S. Department of Agriculture , NASS Research Report No. SRB-91-02.

Bellow, M.E. (1994), "Large Domain Satellite Based Estimators of Crop Planted Area," to appear in American Statistical Association 1994 Proceedings of the Section on Survey Research Methods.

Bellow, M.E. and M. Ozga (1991), "Evaluation of Clustering Techniques for Crop Area Estimation using Remotely Sensed Data," American Statistical Association 1991 Proceedings of the Section on Survey Research Methods, pp. 466-471.

Bush, J. and C. House, (1993), "The Area Frame: A Sampling Base for Establishment Surveys," Proceedings of the International Conference on Establishment Surveys, pp. 335-344.

Cardenas, M., M.M. Blanchard and M.E. Craig (1978), On the Development of Small Area Estimators Using LANDSAT Data as Auxiliary Information, Economics, Statistics and Cooperatives Services, U.S. Department of Agriculture.

Cochran, W.G. (1977), Sampling Techniques, New York, N.Y.: John Wiley & Sons, pp. 26, 150 -204.

Cook, P.W. (1982), Landsat Registration Methodology Used by U.S. Department of Agriculture's Statistical Reporting Service, 1972-1982, Statistical Reporting Service, U.S. Department of Agriculture.

Cotter, J. and J. Nealon, (1987), Area Frame Design for Agricultural Surveys, National Agricultural Statistics Service, U.S. Department of Agriculture.

Craig, M.E. (1993), "The NASS Delta Project: 1991-92 Rice and Cotton Acreage," ACSM/ASPRS Annual Convention and Exposition.

Fuller, W.A. and G.E. Battese (1973), "Transformations for Estimation of Linear Models with Nested-Error Structure," Journal of the American Statistical Association, 68(343), 626-632.

Graham, M.L. (1993), "State Level Crop Area Estimation using Satellite Data in a Regression Estimator," Proceedings of the International Conference on Establishment Surveys, pp. 802-806.

Hollander, M. and D.A. Wolfe (1973), Nonparametric Statistical Methods, New York, N.Y.: John Wiley & Sons, pp. 138- 162.

Huddleston, H.F. and R. Ray (1976), "A New Approach to Small Area Crop Acreage Estimation," Proceedings of the Annual Meeting of the American Agricultural Economics Association.

Iwig, W.C. (1993), "The National Agricultural Statistics Service County Estimates Program," Indirect Estimators in Federal Programs, Statistical Policy Working Paper 21, Federal Committee on Statistical Methodology, Subcommittee on Small Area Estimation, Ch. 7.

Ozga, M, W.W. Mason and M.E. Craig (1992), "PEDITOR - Current Status and Improvements," Proceedings of the ASPRS/ACSM/RT 92 Convention, pp. 175-183.

Prasad, N.G.N and J.N.K. Rao (1990), "The Estimation of the Mean Squared Error of Small-Area Estimators," Journal of the American Statistical Association, 85(409), 163-171.

Stasny, E.A., P.K. Goel and D.J. Rumsey (1991), "County Estimates of Wheat Production," Survey Methodology, 17(2), 211-225.

U.S. Department of Agriculture (1993), Area Frame Design Information, Area Frame Section, NASS.

Walker, G. and R. Sigman (1982), The Use of LANDSAT for County Estimates of Crop Areas - Evaluation of the Huddleston-Ray and Battese-Fuller Estimators, U.S. Department of Agriculture, SRS Staff Report No. AGES 820909.

APPENDIX A: HUDDLESTON-RAY AND CARDENAS ESTIMATORS

The Huddleston-Ray estimator (Huddleston and Ray, 1976) replaces the classified pixel average for the analysis district with the classified pixel average for a county when estimating the county mean crop area per population unit. Within an analysis district, the overall mean crop area in regression stratum h is estimated by:

$$\bar{y}_{h..}^{(HR)} = \bar{y}_{h..} + \hat{\beta}_{1h}(\bar{X}_h - \bar{x}_{h..})$$

and the stratum level mean crop area for county c is estimated by:

$$\bar{y}_{hc.}^{(HR)} = \bar{y}_{h..} + \hat{\beta}_{1h}(\bar{X}_{hc} - \bar{x}_{h..})$$

where:

$\bar{y}_{h..}$ = mean reported crop area per sample segment in stratum h

$\bar{x}_{h..}$ = mean number of pixels per sample segment classified to crop in stratum h

$\bar{X}_h$ = mean number of pixels per population unit classified to crop in stratum h

$\bar{X}_{hc}$ = mean number of pixels per population unit classified to crop in stratum h, county c

$\hat{\beta}_{1h}$ = least squares regression slope parameter estimator

The Huddleston-Ray estimator of total crop area in the regression strata of county c is then:

$$\hat{T}^{(HR)}_c = \sum_{h=1}^{H_r} N_{hc}[\bar{y}_{h..} + \hat{\beta}_{1h}(\bar{X}_{hc} - \bar{x}_{h..})]$$

where:

N_{hc} = number of population units in stratum h, county c

H_r = number of regression strata

The Cardenas family of estimators (Cardenas, Blanchard and Craig, 1978) uses the stratum level differences between mean number of pixels classified to the crop of interest in the county and analysis district, respectively, to adjust the mean reported crop area per sample segment. Within regression stratum h, the estimate of mean crop area per population unit for county c is:

$$\bar{y}_{hc.}^{(CAR)} = \bar{y}_{h..} + B_h(\bar{X}_{hc} - \bar{X}_{h..})$$

where:

B_h = parameter relating classified pixel counts to reported crop area

The estimate of total crop area in the regression strata of county c is:

$$\hat{T}_c^{(CAR)} = \sum_{h=1}^{H_r} N_{hc}[\bar{y}_{h..} + B_h(\bar{X}_{hc} - \bar{X}_h)]$$

There are three alternative estimators of B_h:

1. Ratio estimator

$$\hat{B}_h = \bar{y}_{h..}/\bar{X}_h$$

2. Separate regression estimator

$$\hat{B}_h = \frac{N_h \sum_{c=1}^{C} n_{hc}(\bar{X}_{hc} - \bar{X}_h)\bar{y}_{hc.}}{n_h \sum_{c=1}^{C} N_{hc}(\bar{X}_{hc} - \bar{X}_h)^2}$$

3. Combined regression estimator

$$\hat{B}_h = \frac{\sum_{h=1}^{H_r} (N_h^2/n_h) \sum_{c=1}^{C} n_{hc}(\bar{X}_{hc} - \bar{X}_h)\bar{y}_{hc.}}{\sum_{h=1}^{H_r} N_h \sum_{c=1}^{C} N_{hc}(\bar{X}_{hc} - \bar{X}_h)^2}$$

where:

N_h = number of population units in stratum h

n_h = number of sample segments in stratum h

n_{hc} = number of sample segments in stratum h, county c

$\bar{y}_{hc.}$ = mean reported crop area per sample segment in stratum h, county c

The combined regression estimator is applicable only if the B_h's are assumed to be constant across strata.

APPENDIX B: ESTIMATION OF BATTESE-FULLER VARIANCE COMPONENTS

The estimators of the Battese-Fuller variance components at the analysis district level represent a special case of the more general unbiased estimators derived by Fuller and Battese (1973). The estimators are given by:

$$\hat{\sigma}_{eh}^2 = [1/(n_h - C - 1)] \sum_{c=1}^{C} \sum_{i=1}^{n_{hc}} [y_{hci} - \bar{y}_{hc.} - \hat{\alpha}_h(x_{hci} - \bar{x}_{hc.})]^2$$

$$\hat{\sigma}_{vh}^2 = \max[0, (s_{uh}^2 - (n_h - 2)\hat{\sigma}_{eh}^2)/(n_h - T_h)]$$

where:

$$\hat{\alpha}_h = \frac{\sum_{c=1}^{C} \sum_{i=1}^{n_{hc}} (x_{hci} - \bar{x}_{hc.})(y_{hci} - \bar{y}_{hc.})}{\sum_{c=1}^{C} \sum_{i=1}^{n_{hc}} (x_{hci} - \bar{x}_{hc.})^2}$$

$$s_{uh}^2 = \sum_{c=1}^{C} \sum_{i=1}^{n_{hc}} (y_{hci} - \hat{\beta}_{0h} - \hat{\beta}_{1h} x_{hci})^2$$

$$T_h = \frac{n_h \sum_{c=1}^{C} n_{hc.}^2 \bar{x}_{hc.}^2 + (\sum_{c=1}^{C} n_{hc}^2)(\sum_{c=1}^{C} \sum_{i=1}^{n_{hc}} x_{hci}^2) - 2 n_h \bar{x}_{h..} \sum_{c=1}^{C} n_{hc}^2 \bar{x}_{hc.}}{(n_h \sum_{c=1}^{C} \sum_{i=1}^{n_{hc}} x_{hci}^2) - n_h^2 \bar{x}_{h..}^2}$$

(The Section on Battese-Fuller estimation defines all remaining terms used above).

The value of δ_{hc} that minimizes the mean square of the Battese-Fuller estimator can be estimated by:

$$\hat{\delta}^*_{hc} = n_{hc}\hat{\sigma}_{vh}^2/(n_{hc}\hat{\sigma}_{vh}^2 + \hat{\sigma}_{eh}^2)$$

The stratum level mean square error of the estimator of mean crop area is estimated by:

$$mse(\bar{y}_{(BF),hc.}) = (1 - \hat{\delta}^*_{hc})^2 \hat{\sigma}_{vh}^2 + \hat{\delta}^{*2}_{hc}(\hat{\sigma}_{eh}^2/n_{hc})$$

Walker and Sigman (1982) provide expressions for the mean square error and mean square conditional bias of the stratum level Battese-Fuller estimator. Separate formulas are required depending upon whether the regression parameters are known or estimated. Variance estimators are derived from these formulas.

APPENDIX C: DERIVATION OF VARIANCE FORMULA FOR SYNTHETIC ESTIMATOR

In the Section on Battese-Fuller estimation, the strata where synthetic estimation was done were labeled $h = H_r + 1, \ldots, H$, where H_r is the number of regression strata. The domain indirect synthetic estimator of area planted in a given crop in the synthetic strata of county c is:

$$\hat{T}^{(SYN)}_{c} = \sum_{h=H_r+1}^{H} N_{hc}\bar{y}_{h..}$$

The pure synthetic estimator corresponds to $H_r = 0$. The true variance of the estimator for a given stratum can be expressed as:

$$V(\hat{T}^{(SYN)}_{hc.}) = V(N_{hc}\bar{y}_{h..})$$

$$= N_{hc}^{2}V(\bar{y}_{h..})$$

From Cochran (1977, p. 26), an estimator of the variance of the stratum level mean crop area per sample segment is:

$$v(\bar{y}_{h..}) = s_{yh}^{2}(N_h - n_h)/N_h n_h$$

where:

$$s_{yh}^{2} = \frac{1}{n_h - 1}\sum_{i=1}^{n_h}\sum_{c=1}^{C}(y_{hci} - \bar{y}_{h..})^2$$

Hence the variance of the stratum level synthetic estimator can be estimated by:

$$v\ [\hat{T}^{(SYN)}_{hc.}] = N_{hc}^{2}v(\bar{y}_{h..})$$

$$= N_{hc}^{2}s_{yh}^{2}(N_h - n_h)/N_h n_h$$

Summing over synthetic strata gives the result:

$$v[\hat{T}^{(SYN)}_{hc.}] = \sum_{h=H_r+1}^{H} N_{hc}^{2}s_{yh}^{2}(N_h - n_h)/N_h n_h$$

APPENDIX D: DERIVATION OF VARIANCE FORMULA FOR COMBINED RATIO ESTIMATOR

Cochran (1977, p. 166) gives an approximate formula for the variance of a combined ratio estimator of the form:

$$\hat{T}^{(CR)} = [(\sum_{h=1}^{H} N_h\bar{y}_{h.})/(\sum_{h=1}^{H} N_h\bar{x}_{h.})]X$$

where:

$\bar{y}_{h.}$ = sample mean of main variable in stratum h

$\bar{x}_{h.}$ = sample mean of auxiliary variable in stratum h

X = overall population total of auxiliary variable

For the combined ratio pixel count estimator of small domain crop area, X is replaced by X_c, the population total of the auxiliary variable in county c. Cochran's approximate formula, valid for large samples, is as follows:

$$V(\hat{T}^{(CR)}) = \sum_{h=1}^{H}[(N_h^{2}(1-f_h)/n_h](S_{yh}^{2} + R^{2}S_{xh}^{2} - 2RS_{xyh})$$

where:

$$f_h = n_h/N_h$$

$$S_{xh}^{2} = \frac{1}{N_h - 1}\sum_{i=1}^{N_h}(x_{hi} - \bar{X}_h)^2$$

$$S_{yh}^{2} = \frac{1}{N_h - 1}\sum_{i=1}^{N_h}(y_{hi} - \bar{Y}_h)^2$$

$$S_{xyh} = \frac{1}{N_h - 1}\sum_{i=1}^{N_h}(x_{hi} - \bar{X}_h)(y_{hi} - \bar{Y}_h)$$

$$R = Y/X$$

x_{hi} = true value of auxiliary variable in stratum h, population unit i

y_{hi} = true value of main variable in stratum h, population unit i

Y = overall population total of main variable

$\bar{X}_h$ = population mean of auxiliary variable in stratum h

$\bar{Y}_h$ = population mean of main variable in stratum h

Replacing the true variances by their sample estimators gives an estimator for the variance of the domain direct combined ratio estimator:

$$v(\hat{T}^{(CR)}) = \sum_{h=1}^{H} [(N_h^2(1-f_h)/n_h](s_{yh}^2 + \hat{R}^2 s_{xh}^2 - 2\hat{R}s_{xyh})$$

where s_{yh}^2, s_{xh}^2, and s_{xyh} are defined in the Section on pixel count estimators. The domain indirect combined ratio estimator is obtained by multiplying the domain direct estimator by (X_c/X), the county-to-region ratio of pixels classified to the crop of interest. Since X_c and X are known, this ratio is a constant and the variance of the CRE is estimated by:

$$v(\hat{T}_c^{(CR)}) = [X_c/X]^2 \sum_{h=1}^{H} [(N_h^2(1-f_h)/n_h](s_{yh}^2 + \hat{R}^2 s_{xh}^2 - 2\hat{R}s_{xyh})$$

APPENDIX E: DISTRIBUTION-FREE PROCEDURES FOR COMPARISON OF TREATMENTS

I. Friedman Rank Sum Test

For a two-way layout of n subjects and k treatments, the Friedman rank sum test assumes the following model:

$$X_{ij} = \omega + \beta_i + \tau_j + e_{ij}, \quad i=1,\ldots,n;\ j=1,\ldots,k$$

where X_{ij} is the data value for subject i and treatment j, ω is the unknown overall mean, β_i is the subject i effect, τ_j is the unknown treatment j effect and e_{ij} is a random error. The τ_j's are assumed to sum to zero. The null hypothesis of the Friedman rank sum test is:

$$H_0: \tau_1 = \tau_2 = \ldots = \tau_k \quad \text{(equal treatment effects)}$$

Each subject's k data values are ranked from smallest to largest and the resulting ranks are averaged within treatments. Friedman's S statistic is defined as:

$$S = \frac{12n}{k(k+1)} \sum_{j=1}^{k} (\bar{R}_j - \bar{R}_{..})^2$$

where:

R_j = sum of ranks for treatment j
$\bar{R}_j$ = R_j/n(average rank for treatment j)
$\bar{R}_{..}$ = (k+1)/2(average within-subject rank)

Table E1 shows the estimates, within-subject rankings and rank sums for the Iowa 1988 corn data set.

Table E1: Estimates and Within-Block Ranks (in Parentheses) for Iowa 1988 Corn

COUNTY	*OFF*		*BFE*		*RPCE*		*CRE*		*SYN*	
Audubon	100	(3)	92.2	(1)	100.6	(4)	93.6	(2)	112.2	(5)
Calhoun	133	(1)	133.2	(2)	144.2	(4)	134.4	(3)	144.9	(5)
Carroll	141	(1)	141.4	(2)	152.6	(5)	142.1	(3)	146.2	(4)
Crawford	147	(1)	152.7	(2)	164.9	(4)	155.1	(3)	183.2	(5)
Greene	125	(1)	130.0	(2)	142.7	(4)	132.8	(3)	145.9	(5)
Guthrie	98	(1)	106.3	(2)	115.8	(4)	107.8	(3)	151.3	(5)
Ida	112	(5)	107.0	(1)	110.3	(3)	107.1	(2)	111.4	(4)
Sac	136	(1)	138.3	(2)	150.0	(5)	139.6	(3)	148.1	(4)
Shelby	155	(5)	140.7	(1)	152.1	(4)	141.5	(2)	149.4	(3)
Rank Sums		19		15		37		24		40

For large n, the null distribution of S can be approximated by the χ^2_{k-1} distribution. The null hypothesis is rejected at level α if S is larger than the upper α percentile point of the χ^2_{k-1} distribution.

II. Multiple Comparison Test for Treatments vs. Control

The multiple comparison test of treatments vs. control computes the absolute differences between the rank sums of the k-1 treatments being evaluated (treatments 2, ..., k) and one control treatment (treatment 1). An approximate two-sided test is to conclude that a given treatment u is different from the control treatment at the α level of significance if:

$$|R_u - R_1| > |m|(\alpha, k-1, .5)[nk(k+1)/6]^{1/2}$$

where R_u is the Friedman rank sum for treatment u (u=1, ...,k) and |m| (α, k-1,.5) is the upper α percentile point of the distribution of the maximum absolute value of k-1 standard normal random variables with common correlation 1/2. Table A.14 in Hollander and Wolfe (1973) gives critical values of this distribution for α = .01 and .05.

III. Doksum Estimates of Treatment Effects and Differences

Doksum estimates of treatment effects and pairwise differences between treatment effects are computed as follows. The first step is to generate k tables, with table j containing the within-subject differences $D^i_{jp} = X_{ij} - X_{ip}$ (i=1,...,n; p=1,...,k). For each treatment, the median of pairwise differences is then computed:

$$Z_{jp} = median\ (D^1_{jp}, D^2_{jp}, \ldots, D^n_{jp})$$

Table E2 gives the pairwise differences and median pairwise differences associated with the official estimates for the Iowa 1988 corn data set.

Table E2: Pairwise Differences from Official Estimates (Iowa 1988 Corn)

	Difference from OFF									
County	*OFF*	*(1)*	*BFE*	*(2)*	*RPCE*	*(3)*	*CRE*	*(4)*	*SYN*	*(5)*
Audubon	0.0		7.8		-0.6		6.4		-12.4	
Calhoun	0.0		-0.2		-11.2		-1.4		-11.9	
Carroll	0.0		-0.4		-11.6		-1.1		-5.2	
Crawford	0.0		-5.7		-17.9		-8.1		-36.2	
Greene	0.0		-5.0		-17.7		-7.8		-20.9	
Guthrie	0.0		-8.3		-17.8		-9.8		-53.3	
Ida	0.0		5.0		1.7		5.0		0.6	
Sac	0.0		-2.3		-14.0		-3.6		-12.1	
Shelby	0.0		14.3		2.9		13.5		5.6	
Median	0.0		-0.4		-11.6		-1.4		-12.1	

Estimates of treatment effects are then generated by averaging the median differences associated with each treatment:

$$\hat{\tau}_j = (1/k)\sum_{p=1}^{k} Z_{jp},\ j=1,\ldots,k$$

The contrasts or differences between treatment effects are estimated by:

$$\hat{\theta}_{jp} = \hat{\tau}_j - \hat{\tau}_p, \; j=1,\ldots,k; \; p=1,\ldots,k$$

Table E3 gives the estimated treatment effects and contrasts for the Iowa 1988 corn data set.

Table E3: Estimated Treatment Effects and Constrasts for Iowa 1988 Corn

	Contrasts					*Treatment*
Treatment	*OFF (1)*	*BFE (2)*	*RPCE (3)*	*CRE (4)*	*SYN (5)*	*Effect*
OFF (1)	0.0	-0.34	-11.4	-1.58	-12.18	-5.1
BFE (2)	0.34	0.0	-11.06	-1.24	-11.84	-4.76
RPCE (3)	11.4	11.06	0.0	9.82	-0.78	6.3
CRE (4)	1.58	1.24	-9.82	0.0	-10.6	-3.52
SYN (5)	12.18	11.84	0.78	10.6	0.0	7.08